



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS
CENTRO DE CIENCIAS GENÓMICAS

CARACTERIZACIÓN FUNCIONAL DE LOS GENES INVOLUCRADOS EN LA
RESISTENCIA A METALES PESADOS EN LA BACTERIA *Rhizobium etli* CFN42

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS

PRESENTA:
CIRO ALBERTO CUBILLAS RAMÍREZ

DIRECTOR DE TESIS
DR. ALEJANDRO GARCÍA DE LOS SANTOS
CENTRO DE CIENCIAS GENÓMICAS

COMITÉ TUTOR
DR. ING. CHRISTIAN SOHLENKAMP
CENTRO DE CIENCIAS GENÓMICAS
DR. MARIO SOBERÓN CHÁVEZ
INSTITUTO DE BIOTECNOLOGÍA

MÉXICO, D. F. FEBRERO DE 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

El presente trabajo se desarrolló en el Centro de Ciencias Genómicas de la UNAM como parte de los estudios del Programa de Doctorado en Ciencias Biomédicas en el Programa de Ingeniería Genómica, bajo la dirección del Dr. Alejandro García de los Santos.

Durante la realización de este trabajo se contó con el apoyo económico de una beca del Consejo Nacional de Ciencia y Tecnología y del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica-UNAM IN201112 y al Programa de Apoyo para Estudios de Posgrado.

AGRADECIMIENTOS

A mis padres, mi hermana y sobrinos por ser mi fuente más importante de inspiración, siempre.

A Alejandro, por su apoyo incondicional en este proyecto al apostar por lo nuevo a pesar del alto riesgo y la crítica.

A Pablo Vinuesa, por sus clases privadas de filogenética y análisis de secuencias.

A mis amigos que me acompañaron todo este tiempo: a Yalbi, a Arlet, a Casorla, a Héctor, a Marco y a César por todos los buenos ratos, las largas charlas.

Sin duda a toda la gente que hizo amena la estancia en el laboratorio, en especial a Laura, a Fabiola, Eunice y Roger, por los buenos momentos que pasamos.

Al resto de investigadores del laboratorio, Susi Brom, Pablo Vinuesa y David Romero, por su apoyo y crítica a lo largo del proyecto y sobre todo, por compartir su visión de la investigación en México.

ÍNDICE

LISTA DE FIGURAS.....	6
PRESENTACIONES EN CONGRESOS.....	7
RESUMEN.....	8
ABSTRACT.....	9
CAPITULO 1. INTRODUCCIÓN.....	10
1.1. Metales de transición en metaloproteínas.....	10
1.2. Distribución celular de metales en los seres vivos.....	12
1.3. Sistemas de homeostasis de metales de transición en bacteria.....	14
1.4. Proteínas de expulsión de metales en bacteria.....	15
1.5. Proteínas CDF en los seres vivos: estructura y función.....	17
1.6. Inferencia filogenómica y clasificación funcional de las proteínas CDF.....	18
1.7. Generalidades de los rizobios.....	21
CAPITULO 2. ANTECEDENTES.....	22
2.1. Generalidades de <i>Rhizobium etli</i> CFN42.....	22
2.2. Resistencia a metales de transición en rizobios.....	22
HIPÓTESIS.....	24
OBJETIVOS.....	25
Objetivo general.....	25
Objetivos particulares.....	25
RESULTADOS.....	26
CAPITULO 3. IDENTIFICACIÓN Y CONTRIBUCIÓN DE LAS PROTEÍNAS CDF Y SU PAPEL EN LA RESISTENCIA A METALES EN <i>R. etli</i> CFN42.....	26
3.1. Toxicidad de metales en <i>R. etli</i> CFN42.....	26
3.2. Búsqueda de proteínas CDF homólogas codificadas en el genoma de <i>R. etli</i> CFN42.....	27
2.4. Sensibilidad a Ni ²⁺ de la cepa <i>Rhizobium etli</i> CFNX185.....	29

CAPITULO 4. Phylogenomic analysis of Cation Diffusion Facilitator proteins uncovers Ni²⁺/Co²⁺ transporters.....	31
CAPITULO 5. The <i>emnA</i> gene from <i>Rhizobium etli</i> encodes for a bacterial cation diffusion facilitator protein involved in Mn²⁺ resistance (in preparation)	32
CAPITULO 6. Hipertolerancia a Cu²⁺ de la cepa <i>Rhizobium etli</i> CFNX185.....	41
6.1. Identificación de una mutante Tn5 afectada en la homeostasis de Cu ²⁺	43
CAPITULO 7. CONCLUSIONES.....	47
CAPITULO 8. HOMEOSTASIS DE METALES DE TRANSICION EN <i>Rhizobium etli</i> CFN42: PERSPECTIVAS.....	48
REFERENCIAS.....	49

ÍNDICE DE FIGURAS

Figura 1. Distribución de los 17 metales pesados con mayor influencia biológica en la tabla periódica. (p. 10)

Figura 2. Configuraciones electrónicas de metales de transición comparadas con la de otros metales. (p. 11)

Figura 3. Representación de los sistemas de transporte (ingreso y expulsión) de metales que están presentes en *E. coli*. (p. 15)

Figura 4. Estructuras cristalográficas de CopA y CusA. (p. 16)

Figura 5. Estructura cristalográfica de YiiP y la presencia de 3 sitios de unión a metal. (p. 17)

Figura 6. Diagrama de flujo describiendo los pasos de la inferencia filogenómica. (p. 19)

Figura 7. Las relaciones filogenéticas entre las proteínas CDF están determinadas por el sustrato que es reconocido, las cuales dan lugar a grupos de transportadores de Zn-, Fe/Zn- y Mn, pero no de Ni, Co o Cu. (p. 20)

Figura 8. Citotoxicidad de metales en *R. etli* crecida en medio mínimo con diferentes concentraciones de metales. (p. 26)

Figura 9. Alineamiento múltiple de secuencias de proteínas CDF de *E. coli* contra las CDF putativas de *R. etli*. (p. 28)

Figura 10. El gene *nepA* (PE218) está involucrado con la tolerancia a Ni²⁺ en *Rhizobium etli* CFN42. (p. 30)

Figura 11. The alignment of characterized clade VI CDF proteins shows they have identical MBS A, B and C in spite of having different metal substrates. (p. 35)

Figura 12. Phylogenetic tree of CDF proteins from group VI reveals rhizobial EmnA is part of its own subgroup. (p. 36)

Figura 13. The *R. etli emnA* gene is metal responsive to Fe²⁺, Zn²⁺, Cd²⁺, and Mn²⁺. (p. 37)

Figura 14. *R. etli emnA* gene is essential for Mn²⁺ resistance. (p. 38)

Figura 15. Efecto de las deleciones del plásmido p42e en la tolerancia a cobre de *R. etli* CFN42. (p. 42)

Figura 16. Comparación de fenotipos de la mutante cur201 con la cepa parental CFN42. (p. 43)

Figura 17. Análisis de complementación de la mutante cur201. (p. 45)

PRESENTACIONES EN CONGRESOS

- 2013 Gordon Research Conference: Cell Biology of Metals, Newport, RI. USA. (July 28 – August 2). “Phylogenomic inference of Cation Diffusion Facilitator (CDF) superfamily uncovers novel prokaryotic Ni²⁺ and Co²⁺ transporters”. **Ciro A. Cubillas**, Vinuesa P, Tabche M.L., and García-de los Santos A.
- 2011 Gordon Research Conference: Cell Biology of Metals, Newport, RI. USA. (July 31 – August 5). “The Novel CDF protein NepA, essential for nickel resistance in *Rhizobium etli* CFN42”. **Ciro A. Cubillas**, Vinuesa P, and García-de los Santos A.
- 2010 Bioinorganic Chemistry Gordon Research Seminar, Ventura CA, USA. (February, 4-7) “Dissecting the genetic basis of nickel resistance in *Rhizobium etli*: Key role of a cation diffusion facilitator protein”. **Ciro A. Cubillas** and García-de los Santos A.
- 2009 XVI International Congress on Nitrogen Fixation, Montana, Estados Unidos. (June, 14-19) “Functional characterization of *Rhizobium etli* plasmid-encoded genes required for Nickel and Cobalt resistance”. **Ciro A. Cubillas**, Miguel A. Ramírez y Alejandro García-de los Santos.

RESUMEN

La homeostasis celular de metales es un conjunto de procesos donde proteínas de ingreso, tráfico y expulsión actúan concertadamente para mantener los requerimientos celulares o disminuir el estrés generado por su exceso. Las proteínas de expulsión, componentes críticos de este sistema homeostático, juegan un papel clave en conferir resistencia a metales.

La mayoría de las proteínas de expulsión de metales que han sido descritas en bacterias, pertenecen a las familias de Resistencia-Nodulación-División (RND), de Difusión Facilitada de Cationes (CDF) y las ATPasas Tipo P_{IB}. Uno de los grandes retos en el estudio de estas proteínas es identificar los motivos involucrados con la especificidad al metal, que permitan predecir el metal transportado por las proteínas de expulsión codificadas en un genoma. En este sentido, una de las contribuciones más recientes (1), fue la clasificación funcional de las proteínas de la familia CDF, inferida a partir de una filogenia Neighbor-Joining que clasificó 273 proteínas CDF (putativas y caracterizadas) en tres clados definidos por el sustrato: Zn, Fe/Zn y Mn. Esos estudios fueron incapaces de discernir clados para transportadores de Ni, Co, Cd y Cu, a pesar de que existen en esta familia.

A fin de mejorar y actualizar la clasificación funcional de las proteínas CDF y contribuir a identificar los residuos involucrados con la especificidad del metal, esta tesis propone una nueva clasificación funcional de los transportadores CDF inferida a partir de un enfoque filogenómico. Nuestro estudio propone una clasificación filogenética de las proteínas CDF en 18 grupos. Este análisis predijo nuevos grupos de transportadores de Ni²⁺, Co²⁺ y Mn²⁺ los cuales fueron validados experimentalmente caracterizando el fenotipo de mutantes en los genes que codifican para las proteínas CDF NepA (Ni²⁺/Co²⁺), CepA (Co²⁺) y EmnA (Mn²⁺) de *Rhizobium etli* CFN42, una bacteria modelo en estudios de fijación simbiótica de nitrógeno. También encontramos residuos clado-específicos cuyas mutaciones producen una proteína NepA no funcional.

Como parte de la caracterización de los mecanismos de resistencia a metales en *R. etli* CFN42, reportamos el análisis fenotípico de mutantes el cual nos permitió identificar dos loci relacionados con la homeostasis de Cu²⁺, diferentes al mecanismo de resistencia previamente reportado dependiente de ActP (una ATPasas Tipo P_{IB}).

ABSTRACT

The cellular metal homeostasis is a process where uptake, traffic and efflux proteins act concertedly either to maintain cellular metal requirements or to decrease the stress generated by their excess. The efflux proteins, critical components of this homeostatic system, play a key role in conferring metal resistance.

Most of the bacterial metal efflux proteins that have been described in bacteria belong to the Resistance-Nodulation-Division (RND), Cation Diffusion facilitator (CDF) and P_{IB}-type ATPases. One of the major research challenge in the area of metal transporters is identifying the motifs involved with metal specificity that allow to predict the metal transported by efflux proteins encoded in the genomes. In this way, one of the most recent contributions (1) was the functional classification of proteins from the CDF family, inferred from a Neighbor-Joining phylogeny which distributed 273 CDF proteins (characterized and putative) into three substrate-defined groups: Zn²⁺, Fe²⁺/Zn²⁺ and Mn²⁺. These studies were unable to discern substrate-defined clades for Ni²⁺, Co²⁺, Cd²⁺ and Cu²⁺ transporters, despite their existence in this family.

In order to improve and update the functional classification of CDF proteins and contribute to identify residues involved with metal specificity, this thesis propose a novel functional classification of the CDF transporters inferred from a phylogenomic approach. Our study proposes a filogenetic classification of CDF proteins into 18 groups. This analysis predicted novel groups for Ni²⁺, Co²⁺ and Mn²⁺ which were validated experimentally through the phenotypic characterization of *nepA* (Ni²⁺), *cepA* (Co²⁺) and *emnA* (Mn²⁺) mutants derived from the nitrogen-fixing bacteria *Rhizobium etli* CFN42.

We also found that the NepA ability to confer metal resistance is dependent on clade-specific residues whose mutations produce a non-functional protein.

As a part of the metal resistance mechanisms in *R. etli* CFN42 we report the phenotypic analyses of mutants, which allowed the identification of two different loci involved in Cu²⁺ homeostasis, different to those previously dependent of the ActP protein (P_{IB}-type ATPase) in this bacterium.

CAPITULO 1. INTRODUCCIÓN

1.1. Metales de transición en metaloproteínas

En la naturaleza existen 90 elementos químicos; de ellos, 21 no son metales, 16 son metales ligeros (densidad <5g cm⁻³) y los 53 restantes son metales pesados (densidad >5g cm⁻³). Por su abundancia y solubilidad se consideran al Hierro (Fe), Cobre (Cu), Zinc (Zn), Cobalto (Co), Molibdeno (Mo), Manganeso (Mn), Níquel (Ni), Vanadio (V), Tungsteno (W), Cromo (Cr), Arsénico (As), Plata (Ag), Antimonio (Sb), Cadmio (Cd), Mercurio (Hg), Plomo (Pb) y Uranio (U) como los metales pesados biológicamente relevantes (2) (Fig. 1). De ellos, el As, Ag, Cd, Hg y Pb tienen una actividad biológica muy limitada por lo que son considerados básicamente como contaminantes tóxicos (3).

Aunque práctica, la clasificación de los elementos con base en su densidad dice muy poco sobre su influencia biológica sobre los seres vivos; la cual, está determinada esencialmente por su comportamiento químico cuando interactúan con otras moléculas.

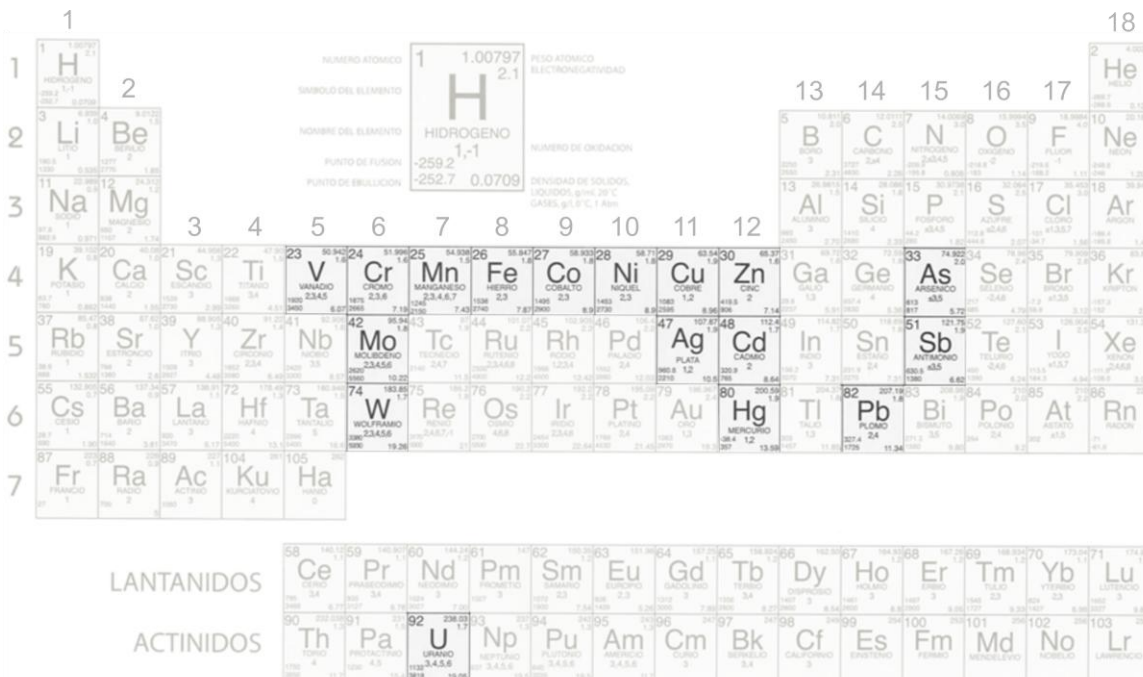


Fig. 1 Distribución de los 17 metales pesados con mayor influencia biológica (resaltados) en la tabla periódica según (2).

Los metales de transición son 38 elementos de la tabla periódica que están comprendidos entre las familias 3 y 12 a partir del cuarto periodo. Desde una perspectiva química, los metales de transición se caracterizan por poseer orbitales *d* incompletos en la última capa de valencia (Fig. 2). Esta característica les permite formar enlaces de coordinación con ciertos elementos como O, N y S, principalmente, lo que facilita su interacción con los aminoácidos que forman parte de las metaloproteínas.

Los aminoácidos que con mayor frecuencia interactúan con metales en las metaloproteínas son: glutamina, ácido glutámico, ácido aspártico, histidina, cisteína, metionina, asparagina, serina, treonina y tirosina (4-5). Por su importancia biológica, el Fe, Mn, Co, Ni y Cu son los metales de transición de mayor relevancia en la fisiología celular (2).

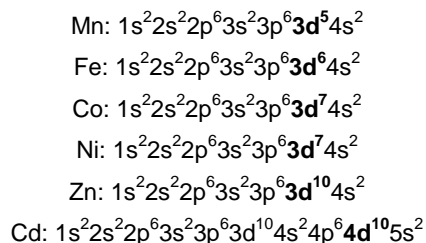


Fig. 2. Configuraciones electrónicas de metales de transición comparadas con la de otros metales, poniendo énfasis en sus orbitales *d* incompletos en la capa de valencia. Nótese que el Zn y el Cd poseen orbitales *d* completos.

1.2. Distribución celular de metales en los seres vivos

La abundancia y disponibilidad de metales, determinadas ambas por la biogeoquímica terrestre, así como los requerimientos celulares de complejos enzimáticos dependientes de metales han favorecido que las células presenten un gran diversidad de metaloproteínas (6). El gran evento de oxidación, manifestado por un incremento de los niveles de O₂ atmosférico como resultado del surgimiento de seres vivos con sistemas fotosintéticos oxigénicos, permitió que complejos minerales ricos en Zn y Cu liberaran su contenido metálico, mientras que restringía la disponibilidad de Fe debido a un aumento en su insolubilidad; siendo quizá el evento que ha tenido los mayores efectos en la evolución de metaloproteínas de Zn y Cu. Al mismo tiempo sometió a los seres vivos a nuevos elementos químicos capaces de generar estrés oxidante dependiente de oxido-reducción metálica, la cual genera iones superóxido y peróxido de hidrógeno como productos de las reacciones tipo Fenton.

Muchos organismos requieren de metales de transición tales como: Mn, Fe, Co, Ni, Cu, y Zn como cofactores catalíticos de enzimas clave para la sobrevivencia, incluyendo citocromos, proteínas de respuesta a estrés oxidativo como catalasa, superóxido dismutasa, hidrogenasas, nitrito reductasas y enzimas del ciclo de Krebs como aconitasa y fumarasa. Dichos metales son también componentes estructurales de metaloproteínas, como ocurre por ej., con los dedos de Zn.

El comportamiento termodinámico de los metales de transición con sus ligandos es descrito por las series Irving-Williams, según la cual la afinidad de los metales de transición hacia las proteínas incrementa en el orden Mn<Fe<Co<Ni<Cu. Para funcionar adecuadamente, las metaloproteínas requieren de mecanismos que les permitan acceder o discriminar el metal correcto entre el conjunto de metales que están siendo movilizados en la célula y que se encuentran en un mismo compartimento (ej. citoplasma). Es por esta razón que las células requieren de un complejo sistema de homeostasis que module los niveles de dichos metales y que sea capaz de distinguir selectivamente entre iones.

Los metales de transición más comúnmente encontrados en células son Fe, Cu, Mn, Co, Ni (7). En células de *Escherichia coli* crecidas en medio rico, las concentraciones más altas (en moles de un ion dado por volumen celular) son para Fe, Zn y Cu, seguidos por el Mn, Co y Ni (7). Bajo condiciones anaeróbicas el contenido de Cu intracelular en *E. coli* aumenta en comparación al crecimiento aeróbico (8). Así, las concentraciones celulares de diferentes metales pueden variar según las condiciones ambientales.

Las concentraciones de metales también pueden variar entre organismos. En células de *Streptococcus pneumoniae* crecidas en medio rico en condiciones de microaerobiosis, los contenidos de Mn, Fe y Zn son prácticamente equivalentes (9), a diferencia de *E. coli* donde el contenido de Fe es, al menos, 5 veces mayor que el de Mn y ~30% más que el de Zn (7).

La búsqueda de dominios de unión a metales utilizando métodos de análisis bioinformático en las proteínas codificadas por genoma estimó que entre un cuarto y un tercio del total de proteínas celulares podrían contener metales (10), mientras que en la Protein Data Bank (PDB) ~40% del total de proteínas contiene algún metal (11).

Este abundante uso de metales en las metaloproteínas de los seres vivos actuales ha dejado huellas en sus proteomas. Se ha estimado que la proporción de dominios de unión a metal (secuencias de aminoácidos que potencialmente unen metales) varía entre los genomas eucariontes, de archaeas y de bacterias, según el metal del que se trate y donde los dominios asociados a un metal escalan proporcionalmente según el tamaño del genoma (12). De este modo, los dominios de unión a metal en Archaea son proporcionalmente más abundantes para Ni y Co que para Zn, mientras que para los genomas eucariontes dichos dominios son mayores para Zn que para Ni y Co (12-13); notablemente, hay una relación entre el número de dominios de unión a metales y el tamaño del genoma, siendo los metaloproteomas eucariontes exponencialmente más grandes con relación a los procariontes(12-13). La proporción de dominios para Fe es mayor para genomas de archaea y bacterias comparados con los genomas eucariontes (13).

1.3. Sistemas de homeostasis de metales de transición en bacteria

La homeostasis de metales es el conjunto de procesos que median el ingreso, tráfico y expulsión de estos, hacia y desde el citoplasma celular (10). Los componentes clave en estos procesos están conformados por un conjunto de múltiples transportadores (de ingreso y expulsión), que permiten la movilización de dichos metales hacia y desde las células; de proteínas de tráfico (denominadas metalochaperonas), que trasladan el metal entre metaloproteínas así como también proteínas sensoras, que modulan la expresión de los genes codificantes para dichas metaloproteínas.

El paradigma de homeostasis de metales de transición en bacteria está representado por *E. coli*. Para esta bacteria, se conocen los mecanismos de ingreso, expulsión tráfico y regulación para 6 metales comunes en metaloproteínas, excepto para Cu, para el cual no se conocen con detalle los mecanismos de ingreso (Tabla 1) (Fig. 3).

Tabla 1. Mecanismos conocidos de homeostasis de metales de transición en *Escherichia coli*.

Mecanismo	Mn	Fe
Ingreso	MntH (14)	FecA-E, FeoAB, FepA-E, EfeUOB, FhuA-E
Regulador Ingreso	MntR, Fur (14)	Fur, <i>ryhB</i> (15-17)
Expulsión	MntP (18)	FieF (YiiP) (19)
Regulador Expulsión	Fur(?), MntR, MntS(18)	Desconocido
Mecanismo	Co	Cu
Ingreso	BtuCD-F (20)	Desconocido
Regulador Ingreso	BtuR (21)	Desconocido
Expulsión	RcnAB (22-23)	CopA(24), CusABC(25-26)
Regulador Expulsión	RcnR (27-28)	CueR, CusRS (29-30)
Mecanismo	Ni	Zn
Ingreso	NikABCDE(31)	ZnuABC (32), ZupT (33)
Regulador Ingreso	NikR (34-35)	Zur (32, 36)
Expulsión	RcnAB (22-23)	ZntA, ZitB (37-38)
Regulador Expulsión	RcnR (27-28)	ZntR (7, 39)

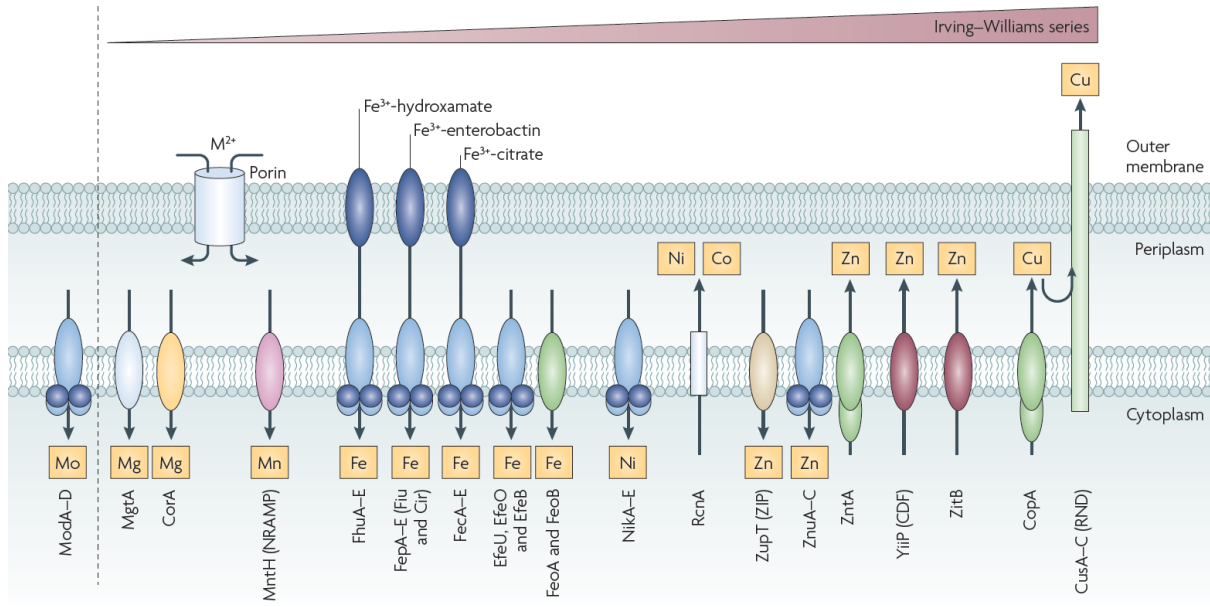


Fig. 3 Representación de los sistemas de transporte (ingreso y expulsión) de metales que están presentes en *Escherichia coli* [Tomado de (10)]

1.4. Proteínas de expulsión de metales en Bacteria

Ya que la resistencia a metales depende de proteínas de expulsión, en este trabajo hacemos énfasis en ellas, sin describir a las proteínas de ingreso. Cuando las concentraciones intracelulares de metales se elevan, se vuelve necesaria la presencia de proteínas de expulsión; las cuales son críticas para mantener las concentraciones a niveles fisiológicamente aceptables en estos casos. Hasta ahora, en Bacteria han sido descritas proteínas de expulsión de metales que pertenecen al menos a 7 familias distintas de transportadores, denominadas: Resistencia-Nodulación y División (RND, por sus siglas en inglés), Difusión facilitada de Cationes (CDF, por sus siglas en inglés), de la familia de facilitadores Principales (MFS, por sus siglas en inglés), transportadores de Ni y Co (NiCoT), transportadores del grupo YebN/MntP, ATPAsas Tipo P_{IB} y transportadores del grupo RcnA/MrdH; aunque únicamente 3 de ellas, las proteínas RND, ATPAsas Tipo P_{IB} y las CDF se han estudiado con detalle. Actualmente, proteínas representativas de cada una de estas 3 familias han sido cristalizadas: YjiP y CzrB para la familia CDF (40-41), CopA para las ATPAsas Tipo IB (42) y CusA para las proteínas RND (43).

Las estructuras de CusA y de CopA como ejemplos representativos de las familias RND y ATPasas tipo P_{1B} (Fig. 4), respectivamente (siendo Cu sustrato de ambas), revelan que el transporte mediado por metioninas, que forma parte del sitio activo de cada una de estas proteínas, ha convergido en este par de familias. La unión del Cu en el sitio activo de cada una de las proteínas induce un cambio conformacional en la estructura proteica que eventualmente desencadena el transporte. La energía requerida para esta actividad procede de la hidrólisis de ATP, en el caso de las ATPasas tipo P_{1B}, esto es, transporte activo, mientras que en el caso de las proteínas RND es dependiente de un gradiente quimiosmótico, por antiporte (ingreso de un protón, expulsión de un catión) el transporte en ambos casos se da contra un gradiente de concentración.

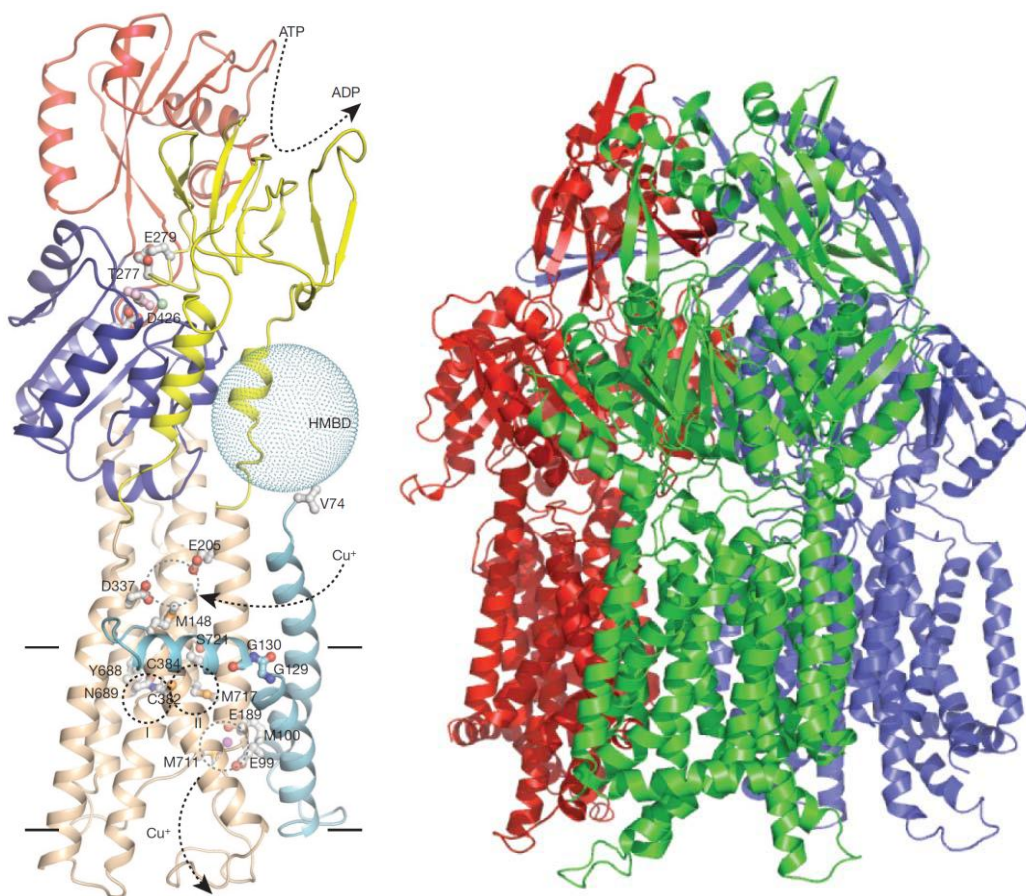


Fig 4. Estructuras cristalográficas de CopA (izq.) (42) y CusA (der.) (43) en la forma apo (libre de metal). Ambas estructuras revelan un requerimiento de metioninas que facilitan el transporte, ya sea en el sitio activo o a lo largo de la ruta de tráfico del ion a lo largo de la proteína.

1.5. Proteínas CDF en los seres vivos: estructura y función

Las proteínas CDF caracterizadas están constituidas de 6 dominios transmembranales con un mecanismo de acción dependiente de un gradiente de protones K^+ o H^+ conservado de bacterias a eucariontes (44-45). Esta familia de transportadores comprende alrededor de 45 proteínas caracterizadas procedentes de bacterias y eucariontes, cuyos sustratos son únicamente metales. Los sustratos conocidos en esta familia de transportadores incluyen: Zn^{2+} , Co^{2+} , Cd^{2+} , Mn^{2+} , Fe^{2+} , y Ni^{2+} . Las bases moleculares de la especificidad no se entienden del todo, sin embargo, existe evidencia de que las diferencias de secuencia al interior de esta familia, y no su regulación, son las responsables del reconocimiento diferencial (1).

A partir de la estructura cristalográfica de la proteína YiiP de *Escherichia coli*, involucrada con el transporte de Zn^{2+} y Cd^{2+} (Fig. 5) se definieron 3 sitios de unión al metal denominados A, B y C (40-41) involucrados con el transporte iónico. En esta proteína, el sitio A comprende 4 residuos que coordinan a un átomo de Zn^{2+} y/o Cd^{2+} distribuidos en los dominios transmembranales 2 y 5 y es crucial para mediar el proceso de traslocación. El sitio B facilita la formación de homodímeros actuando como pivote, mientras que el C interviene en la regulación de la actividad de transporte influyendo en la velocidad de recambio (41).

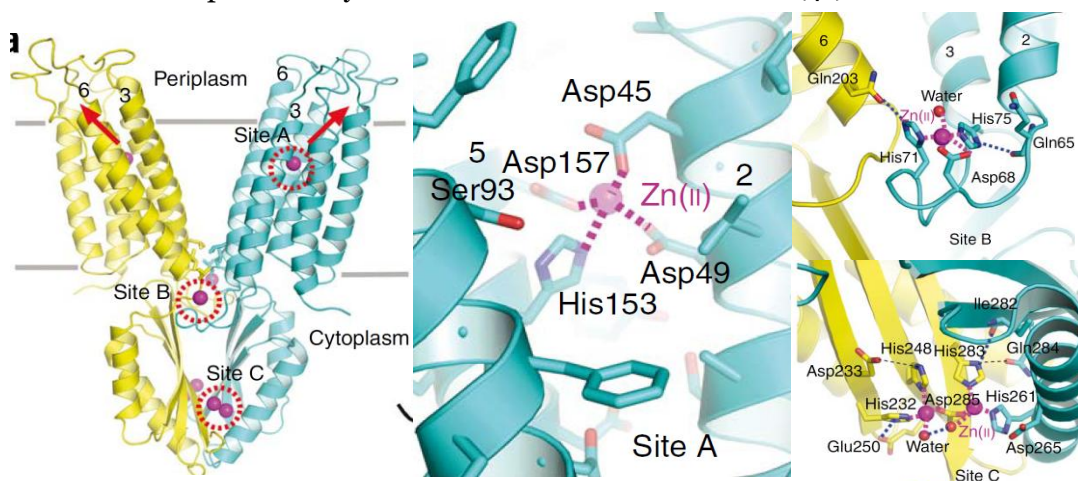


Fig. 5 Estructura cristalográfica de YiiP y la presencia de 3 sitios de unión a metal. La estructura a 2.9 Å de resolución de YiiP (izq.) reveló 3 sitios de unión, denominados A (centro), B y C (der.). El sitio A es el sitio activo, involucrado en la traslocación de Zn desde un compartimento al otro [Tomado de (41)].

1.6. Inferencia filogenómica y clasificación funcional de las proteínas CDF

De las tres familias de transportadores previamente mencionados, la familia CDF es una de las más ampliamente distribuidas. La base de datos Pfam (versión 27.0, Marzo 2013) incluye más de 12000 proteínas CDF putativas que provienen de eucariontes, bacterias y arqueas. Pese a que los únicos sustratos conocidos para las proteínas CDF son metales, principalmente Zn^{2+}/Cd^{2+} , Fe^{2+} y Mn^{2+} , los programas de búsqueda de homología automatizados (como BLAST) aún son insuficientes para predecir con precisión el (los) sustratos específicos para estas proteínas a partir de su secuencia primaria de aminoácidos. Para incrementar la precisión de estas predicciones funcionales entre conjuntos de secuencias homólogas se han desarrollado conceptos integrados de análisis, siendo uno de ellos la inferencia filogenómica.

La inferencia filogenómica de funciones es un concepto desarrollado para predecir la función de una proteína sin caracterizar considerando su relación evolutiva (inferida a partir de una filogenia) con las proteínas de la familia a la que pertenece, para las que en algunos casos ya existen ortólogos funcionalmente caracterizados, gracias a los cuales se infiere su función al pertenecer al mismo clado (46-47). La inferencia filogenómica es una herramienta clave en la curación de las funciones predichas por los métodos de comparación de secuencias “high-throughput” como BLAST, los cuales no contemplan la evolución de funciones por duplicación de genes, recombinación de dominios y cambios de función en ortólogos de especies distantes. No menos importante es el hecho de que los métodos de comparación de secuencias han propiciado la propagación de errores en la anotación de genomas (48).

La inferencia filogenómica de funciones se desarrolla en una serie de pasos previamente definidos (46-47) (Fig. 6):

- 1) Colecta de secuencias
- 2) Generación de un alineamiento múltiple de secuencias
- 3) Examinación/edición del alineamiento

4) Enmascaramiento de secuencias, consistente en remover aquellas secuencias muy divergentes, o carentes de regiones estructurales clave en la función; inferida a partir de las secuencias caracterizadas. El alineamiento obtenido en el paso 3 es analizado a detalle para eliminar columnas que no son confiables, por ejemplo debido a la presencia de muchos gaps, que podrían influir en la reconstrucción filogenética.

5) Construcción de un árbol filogenético de las proteínas a analizar. Ya que el análisis filogenético contempla el uso de programas que incorporan modelos de sustitución de aminoácidos, es importante definir cual es el mejor modelo para la familia de proteínas en consideración.

6) Identificar sub-árboles con soporte estadístico.

7) Sobrelapar topologías con la información experimental.

8) Etiquetado de nodos indicando duplicación o especiación.

9) Inferencia funcional de las proteínas CDF de clados con proteínas caracterizadas.

A diferencia de la filogenética, la cual establece la relación ancestro-descendencia entre un conjunto de proteínas, en este caso; la filogenómica incorpora en la filogenética aquella información experimental relevante, que en conjunto permite inferir la función, o como ocurre en el caso de las proteínas CDF, predecir el sustrato metálico transportado.

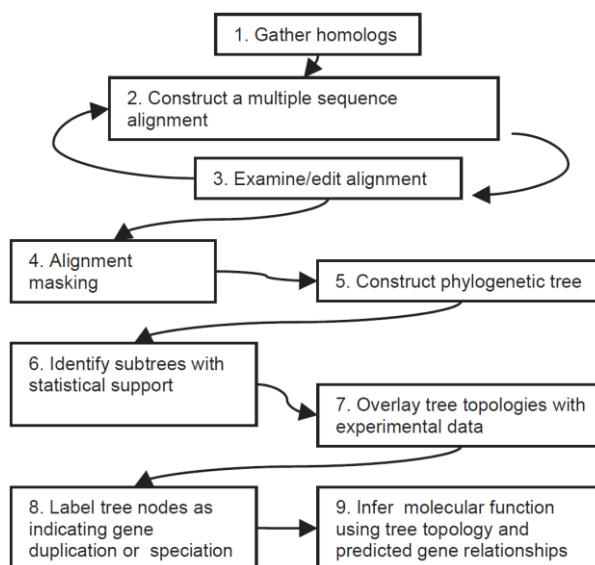


Fig. 6 Diagrama de flujo describiendo los pasos de la inferencia filogenómica. Tomado de (48).

Los diferentes esfuerzos por clasificar a las proteínas CDF con base en un análisis filogenético establecen una correlación entre la secuencia de las proteínas CDF y sus sustratos (1, 49). El análisis filogenético más reciente de las proteínas CDF basado en el método Neighbor-Joining (Fig. 7) (1) propone una clasificación funcional en tres clados: Zn^{2+} , Fe^{2+}/Zn^{2+} y Mn^{2+} . Sin embargo, este análisis no contó con la suficiente resolución para separar en clados independientes transportadores de Ni^{2+} y Co^{2+} , dos de los posibles sustratos de las proteínas CDF y para los cuales ya existen proteínas funcionalmente caracterizadas.

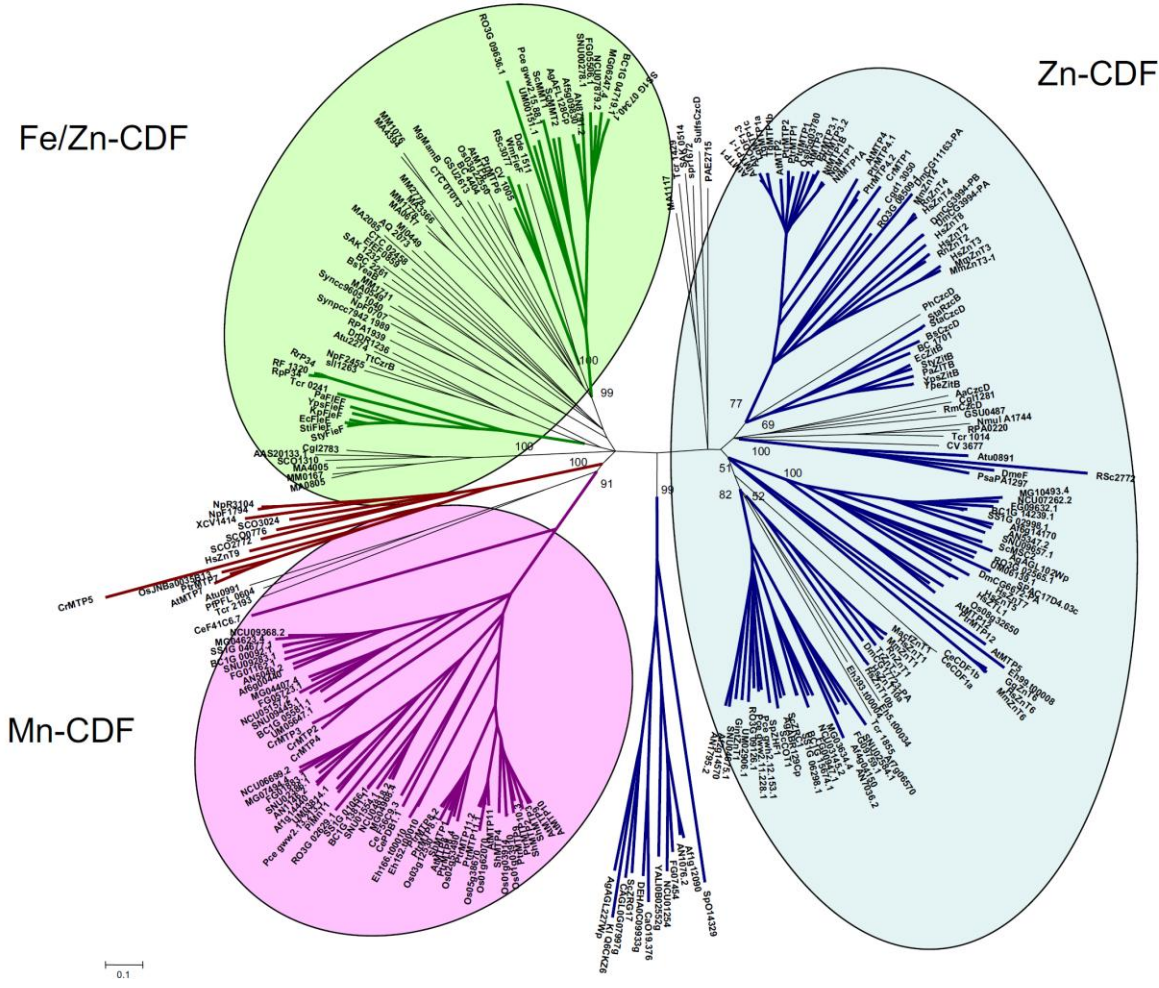


Fig. 7 Las relaciones filogenéticas entre las proteínas CDF están determinadas por el sustrato que es reconocido, las cuales dan lugar a grupos de transportadores de Zn-, Fe/Zn- y Mn, pero no de Ni, Co o Cu [tomado de (1)].

Es importante señalar que de 2007 a la fecha han sido caracterizadas funcionalmente varias proteínas CDF, entre ellas algunas involucradas con el transporte de Mn^{2+} en bacterias, que podrían derivar en la formación de nuevos clados independientes respecto a los ya conocidos. Adicionalmente, este estudio no incluyó las proteínas de rizobios secuenciados a esa fecha, grupo bacteriano de nuestro interés (ver explicación más adelante). Por lo anterior, consideramos importante actualizar el número de proteínas CDF caracterizadas funcionalmente, incluir las proteínas CDF de rizobios y mejorar la clasificación de las proteínas de la familia CDF mediante inferencia filogenómica utilizando el método de máxima verosimilitud para construir la filogenia de la familia de proteínas CDF.

1.7. Generalidades de los rizobios

Rizobios es el nombre común con el que se denomina a un grupo de bacterias Gram-negativas, α -proteobacterias pertenecientes al orden Rhizobiales, dentro del cual se encuentran los géneros *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Allorhizobium*, *Azorhizobium* y *Bradyrhizobium* (<http://www.rhizobia.co.nz/taxonomy/rhizobia>, última actualización mayo 2013). Su importancia agrícola se debe a la capacidad para establecer asociaciones mutualistas principalmente con raíces de plantas leguminosas. La asociación mutualista mejor estudiada es la fijación simbiótica de nitrógeno, pero también pueden vivir como endófitos en las raíces de plantas no leguminosas (50). Durante la fijación simbiótica de nitrógeno, los rizobios inducen en su planta hospedera la formación de una estructura denominada nódulo, dentro del cual las bacterias sintetizan la enzima nitrogenasa, responsable de la reducción del nitrógeno atmosférico en amoníaco. El amoníaco es transferido a las células vegetales, para que la planta lo utilice para su crecimiento. A cambio, las bacterias reciben de la planta ácidos di-carboxílicos y glutamato, los cuales permiten mantener activo el ciclo de los ácidos tricarboxílicos y producir suficiente ATP para el funcionamiento de la nitrogenasa (51). Cabe mencionar que los rizobios no son simbioses obligados, es decir también se les encuentra en suelo y en la rizósfera, la porción de suelo influido por las raíces.

CAPITULO 2. ANTECEDENTES

2.1. Generalidades de *Rhizobium etli* CFN42

R. etli es la bacteria más abundante en suelos de América Latina donde se siembra frijol, la cepa mejor caracterizada de esta especie es la CFN42, aislada de nódulos de raíces de frijol cultivado en Guanajuato, México (52). Su genoma, totalmente secuenciado en el Centro de Ciencias Genómicas (CCG) de la UNAM, está constituido por un cromosoma circular de 4.3 Mb y seis plásmidos denominados p42a (194.2 kb), p42b (184.3 kb), p42c (250.9 kb), p42d (371.2 kb) p42e (505.3 kb) y p42f (642.5 kb)(53). En su genoma se encuentran codificados numerosos transportadores los cuales se consideran parte esencial de la capacidad de esta bacteria para adaptarse a la vida simbiótica, dentro de la planta hospedera, y saprofitica en la rizósfera. Se desconocen cuáles de ellos conforman el transportoma de iones metálicos divalentes, es decir, las proteínas que regulan el ingreso, tráfico y expulsión de estos micronutrientes o iones tóxicos.

2.2. Resistencia a metales de transición en rizobios

Una condición de estrés ambiental que enfrentan las bacterias de suelo es la presencia de elevadas concentraciones de metales debido a la aplicación de fertilizantes y pesticidas, al riego con aguas negras, así como por contaminantes derivados de las actividades mineras y metalúrgicas (54). Experimentos de campo han demostrado que niveles tóxicos de metales disminuyen la biomasa de rizobios del suelo y propician la gradual disminución de la diversidad genética de la población (55-57).

Debido a la importancia agrícola de estas bacterias y a su inminente exposición a altas concentraciones de metales en su hábitat, es necesario conocer la resistencia o susceptibilidad de diferentes especies y cepas y caracterizar los mecanismos moleculares involucrados con la homeóstasis de metales, lo cual permitirá no sólo tener un conocimiento básico en estas bacterias sino también planear el diseño o selección de cepas resistentes.

A la fecha se han secuenciado los genomas de 98 rizobios, incluyendo aquellos capaces de establecer simbiosis con plantas leguminosas, pero los inventarios de sus genes y proteínas no permiten determinar *a priori* su resistencia o susceptibilidad a metales ni los mecanismos que les permiten contender con concentraciones tóxicas de los mismos. El sistema de homeostasis de hierro es el mejor caracterizado en rizobios (58-60). Para otros metales existen pocos reportes acerca de la caracterización de mutantes sensibles a metales. La primera de ellas es una mutante de *R. leguminosarum* sensible a Cu^{2+} que lleva un transposón en el gene *actP* (61). La segunda es una mutante de *Sinorhizobium meliloti* sensible a Cd^{2+} y Zn^{2+} que lleva un transposón en el gene SMc04128 (62). Ambos genes codifican para ATPasas de tipo P_{IB} putativas. La tercera es una mutante de *Bradyrhizobium sp* sensible a Ni^{2+} por interrupción del gene *cnrA* el cual codifica para una posible proteína de expulsión perteneciente a la familia RND (63). El operón *copARZ* involucrado con la resistencia a Cu^{2+} fue caracterizado en *Agrobacterium tumefaciens*, un patógeno de plantas filogenéticamente relacionado con el género *Rhizobium* (64). Otro componente celular de rizobios involucrado en la resistencia a metales son los polisacáridos de superficie (62, 65-66). Las proteínas de la familia CDF no han sido caracterizadas en rizobios.

Las rizobios producen tres tipos de polisacáridos, exopolisacáridos (EPS) los cuales son secretados al medio o permanecen adheridos a la membrana externa formando un polisacárido capsular. Lipopolisacáridos (LPS) los cuales forman parte de la capa exterior de la membrana externa y están anclados al lípido A. Debido a que ambos tipos de polisacáridos están cargados negativamente y los iones metálicos lo están positivamente, la interacción electrostática polisacárido-metal es capaz de inmovilizar químicamente a los últimos.

Los β -(1,2) glucanos cíclicos son otro tipo de polisacáridos que se concentran en el espacio periplásmico en respuesta a cambios de osmolaridad en el ambiente (67). De forma análoga a los EPS y LPS, estos polisacáridos interactúan con el metal en un bolsillo hidrofóbico, negativamente cargado, lo que permite inactivar a los iones metálicos, ayudando a las células a contender contra ese estrés.

HIPÓTESIS

La resistencia a metales de transición depende de proteínas de expulsión de algunas de las familias RND, CDF, MFS, NiCoT, transportadores del grupo YebN/MntP, ATPAsas Tipo P_{IB} y/o transportadores del grupo RcnA/MrdH, codificadas en el genoma de *Rhizobium etli*.

OBJETIVOS

OBJETIVO GENERAL

Identificar mecanismos de resistencia para Ni²⁺, Co²⁺, Mn²⁺ y Cu²⁺ en *Rhizobium etli* CFN42.

OBJETIVOS PARTICULARES

- a) Proponer una nueva clasificación funcional de los transportadores CDF mediante un análisis filogenómico, que permita predecir con mayor precisión la especificidad de metales de proteínas CDF putativas.
- b) Predecir y determinar experimentalmente la especificidad de sustrato de las proteínas CDF de *R. etli* CFN42, a fin de validar el análisis filogenómico.
- c) Determinar que metales inducen la expresión de los genes que codifican para las proteínas CDF de *R. etli* CFN42.
- d) Identificar y mutar aminoácidos clado-específicos potencialmente involucrados con la especificidad de sustrato.

RESULTADOS

CAPITULO 3. IDENTIFICACIÓN Y CONTRIBUCIÓN DE LAS PROTEÍNAS CDF Y SU PAPEL EN LA RESISTENCIA A METALES EN *R. etli* CFN42

3.1 Toxicidad de metales en *R. etli* CFN42

Como parte inicial del trabajo experimental de esta tesis intentamos responder una cuestión básica que no había sido previamente estudiada, siendo además clave para iniciar este proyecto: cuál es el nivel de resistencia de *R. etli* a distintos metales. Para responder esta cuestión se establecieron condiciones de ensayo en placas de medio mínimo (MM) conteniendo o no Zn, Co, Cd, Fe, Mn, Ni y Cu, partiendo de la premisa que a menor concentración requerida para inhibir el crecimiento mayor toxicidad y viceversa. A partir de este análisis determinamos que el metal más tóxico para *R. etli* CFN42 es Cu (20 μ M), el menos tóxico Mn (35 mM) y que el gradiente de toxicidad de los metales incrementaba de forma paralela a como se comportan los metales según las series Irving Williams (Fig. 8).

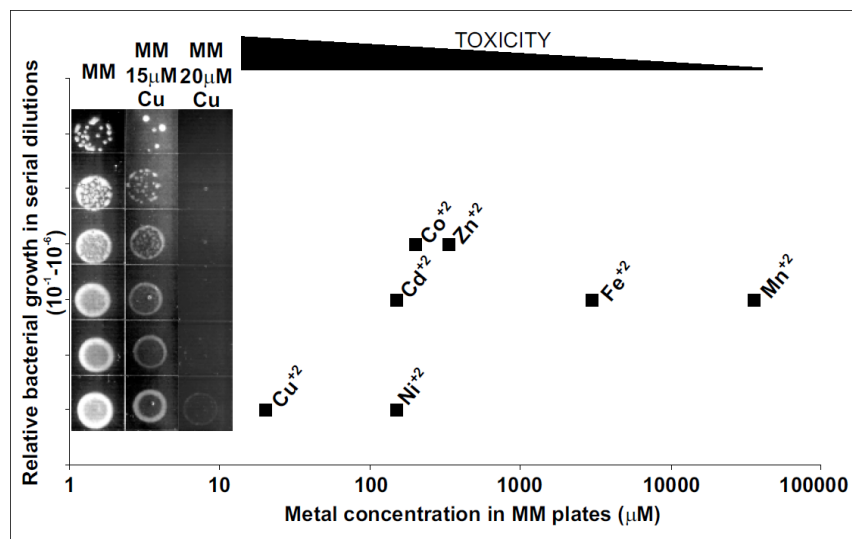


Fig. 8. Citotoxicidad de metales en *R. etli* crecida en medio mínimo con diferentes concentraciones de metales. La placa en el lado izquierdo muestra el crecimiento en diluciones seriadas sin metal (MM), la concentración mínima inhibitoria (15 μ M de Cu) y la concentración totalmente inhibitoria de 20 μ M de Cu, como un ejemplo. Lo mismo fue hecho para el resto de metales y la concentración mínima inhibitoria está representada por los cuadros.

3.2. Búsqueda de proteínas CDF homólogas codificadas en el genoma de *R. etli* CFN42

Ya que las proteínas CDF de rizobios no han sido funcionalmente caracterizadas, exploramos cuántas proteínas CDF estaban codificadas en el genoma de *R. etli* y cuáles eran sus posibles sustratos. Para hacerlo, llevamos a cabo una búsqueda de proteínas homólogas a las proteínas CDF de *E. coli* ZitB y YiiP involucradas con la resistencia a Zn/Cd y Fe, respectivamente (incluidas en la Tabla 1), codificadas en el genoma de *R. etli* mediante BLASTP. Como resultado de esta búsqueda (Fig. 9) encontramos 4 genes codificantes para proteínas CDF potencialmente involucrados con la homeostasis de Zn²⁺, Cd²⁺ y Fe²⁺; aunque dado que los sustratos de estos transportadores aún no pueden predecirse a partir de la secuencia primaria de aminoácidos, estas proteínas CDF podrían conferir resistencia adicionalmente a Co²⁺, Ni²⁺ y Mn²⁺.

Los genes codificantes para proteínas CDF son: *RHE_CHO3072* (*emnA*), *RHE_CHO1219* (*cepA*), *RHE_PD00193* (*PD193*) y *RHE_PE00218* (*nepA*); algunos se encuentran formando parte de posibles operones como ocurre entre los genes *nepA-RHE_PE00217* y *RHE_CHO1218-cepA* y que son comparables a algunos operones involucrados con la expulsión de Zn²⁺, Cd²⁺, Ni²⁺, Co²⁺ previamente descritos (68-71), localizados en el plásmido p42e y en el cromosoma de *R. etli* CFN42. La anotación del genoma los señala como bombas de expulsión de cationes, *nepA* incluso estaba anotado como una proteína de expulsión de Co²⁺, Zn²⁺ y Cd²⁺; sin embargo, no existe evidencia experimental reportada para ninguna CDF de *R. etli* ni de otros rizobios.

Adicionalmente, todas las proteínas CDF putativas de *R. etli* se encuentran en la base de datos de Pfam (72) para esta familia (numero de acceso PF01545), y en la base de transportadores TransportDB (<http://www.membranetransport.org/>), lo cual confirma nuestros análisis comparativos de secuencias. En ambas bases de datos se hallaron las mismas 4 proteínas CDF putativas identificadas por nosotros las cuales poseen las firmas de secuencia características de la familia en su secuencia primaria de aminoácidos (Fig. 9).

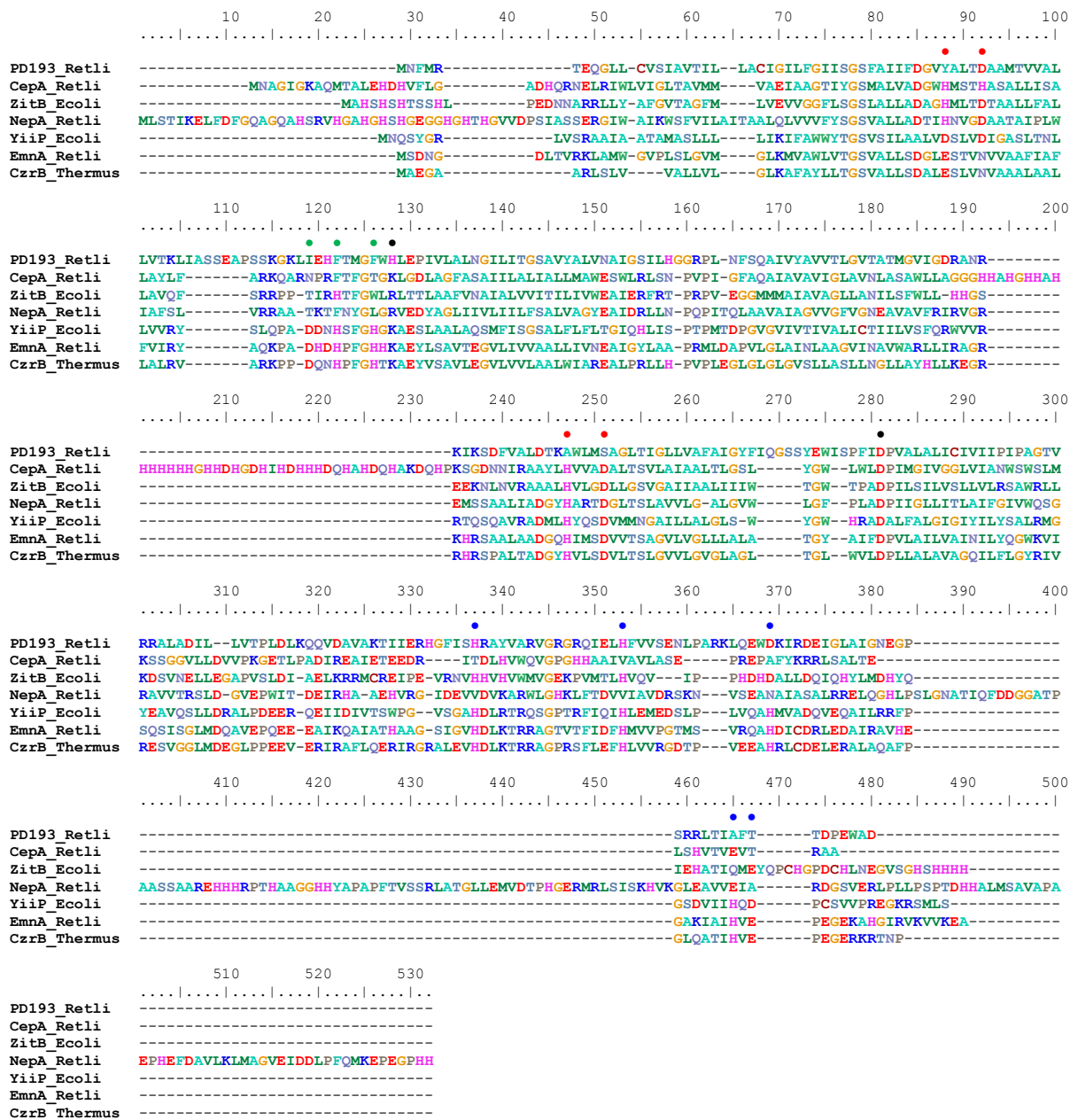


Fig. 9. Alineamiento múltiple de secuencias de proteínas CDF de *Escherichia coli* contra las CDF putativas de *Rhizobium etli*. Tanto ZitB, YiiP de *Escherichia coli* como CzrB de *Thermus thermophilus*, han sido funcionalmente caracterizadas (40, 73-74). Las proteínas de *Rhizobium etli* se identificaron mediante BLASTP usando como templatado de búsqueda las proteínas de *E. coli* (e-value ≤ 0.001 , identidad $>40\%$, cobertura del alineamiento $>80\%$ de la secuencia). Estos datos fueron validados buscando desde la base de datos de Pfam (accession PFO1545, familia de las proteínas CDF) usando el identificador RHIEC, específico del genoma de *R. etli* CFN42. Los sitios de unión al metal A, B, C están identificados por los puntos rojo, verde y azul respectivamente, mientras que los puentes salinos Lys⁷⁷-Asp²⁰⁷ se representan en color negro.

3.3. Sensibilidad a Ni²⁺ de la cepa *Rhizobium etli* CFNX185

La cepa CFNX185 es una mutante, derivada de la cepa CFN42, (42) que sufrió una deleción espontánea la cual carece de ~200 genes (genes RHE_PE00074-PE00265) incluidos los genes *RHE_PE00215-nepA*, codificantes para un homólogo de RcnA de *E. coli* involucrado con la expulsión de Ni y Co (22) (22% identidad, 80% cobertura, e-value=3X10⁻¹²) y para una proteína CDF (Fig. 9), respectivamente. La cepa CFNX185, nos permitió evaluar de forma indirecta si dicha región estaba involucrada con la resistencia a alguno de los metales que habíamos predicho anteriormente. Como se muestra en la Fig. 10B, al comparar el crecimiento de la cepa silvestre CFN42 con la cepa mutante CFNX185 en presencia de Ni²⁺, esta última mostró un crecimiento disminuido en al menos 6 unidades logarítmicas con respecto a la cepa silvestre.

De la búsqueda de clonas complementantes de la resistencia a Ni²⁺ en una librería genómica de *R. etli* se obtuvo un cósmido (denominado pL313) conteniendo un inserto de ~25kb de DNA de *R. etli*, a partir del cual se subclonó una banda de 7.1kb (Fig. 10A) que portaba, entre otros, los genes *RHE_PE00215* y *nepA* y que permitía recuperar el crecimiento en presencia de Ni²⁺. De esta banda de 7.1kb se subclonaron dos fragmentos: una banda *EcoRI-BamHI* de 4.9kb, que contenía el bloque *nepA-RHE_PE00214*; mientras que el segundo, *EcoRI-PstI* de 3.7kb contenía los genes *RHE_PE00217-RHE_PE00214* pero no a *nepA* codificante para la proteína CDF. De este modo podríamos discriminar cual gene estaba involucrado con la resistencia a Ni²⁺, si es que alguno de ellos lo estaba.

El análisis de complementaciones mostró que solamente el fragmento de 4.9kb y no el de 3.7kb complementaban la resistencia (Fig.10). A partir de estos datos, y teniendo en mente que el gene *nepA* codifica para una bomba de expulsión de metales, fue clonado y movilizó al fondo génico de la mutante CFNX185. Como resultado se determinó que el gene *nepA*, es el responsable de la resistencia a Ni²⁺ en *R. etli* (Fig. 10). Este hallazgo es de gran relevancia ya que no habían sido descritas proteínas CDF confiriendo resistencia específica a Ni²⁺ en bacterias.

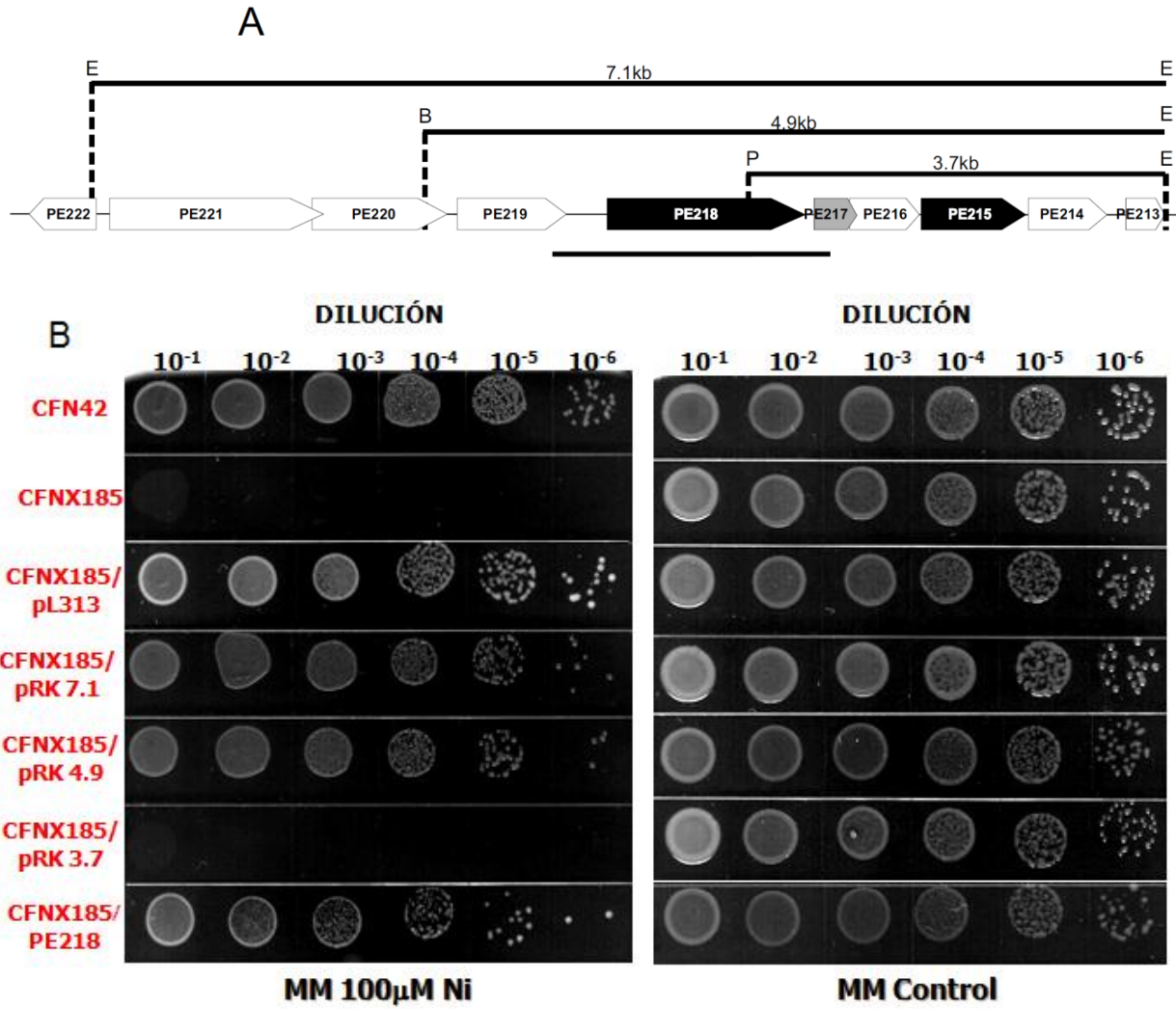


Fig. 10 El gene *nepA* (PE218) está involucrado con la tolerancia a Ni²⁺ en *Rhizobium etli* CFN42. A) Mapa genético de los genes codificantes para transportadores presentes en el plásmido p42e, PE218 (*nepA*). B) Crecimiento de cepas de *R. etli* con fragmentos que llevan genes codificantes para diferentes transportadores en placas de MM con y sin 100 μ M Ni²⁺ (diluciones sucesivas 1/10). Puede verse la complementación del crecimiento en presencia del cósmido pL313, la banda de 4.9kb y con el gene *nepA* pero no con la banda de 3.7kb.

CAPITULO 4. Phylogenomic analysis of Cation Diffusion Facilitator proteins uncovers Ni²⁺/Co²⁺ transporters

Con base en los principios previamente propuestos (46), se obtuvo una nueva clasificación funcional de la familia de transportadores CDF usando inferencia filogenómica. A partir de la filogenia se hizo una predicción global de la especificidad de sustrato de las proteínas CDF de los rizobios, se definieron nuevos clados conteniendo transportadores de Zn²⁺, Fe²⁺, Cd²⁺, Mn²⁺, Ni²⁺ y Co²⁺ y se identificaron residuos grupo específicos de algunos de los nuevos clados descubiertos. Los resultados de este análisis se reportan en el artículo anexo:

Cubillas C., Vinuesa P., Tabche ML., and García-de los Santos Alejandro. Phylogenomic analysis of Cation Diffusion Facilitator proteins uncovers Ni²⁺/Co²⁺ transporters. *Metallomics*. 5: 1634

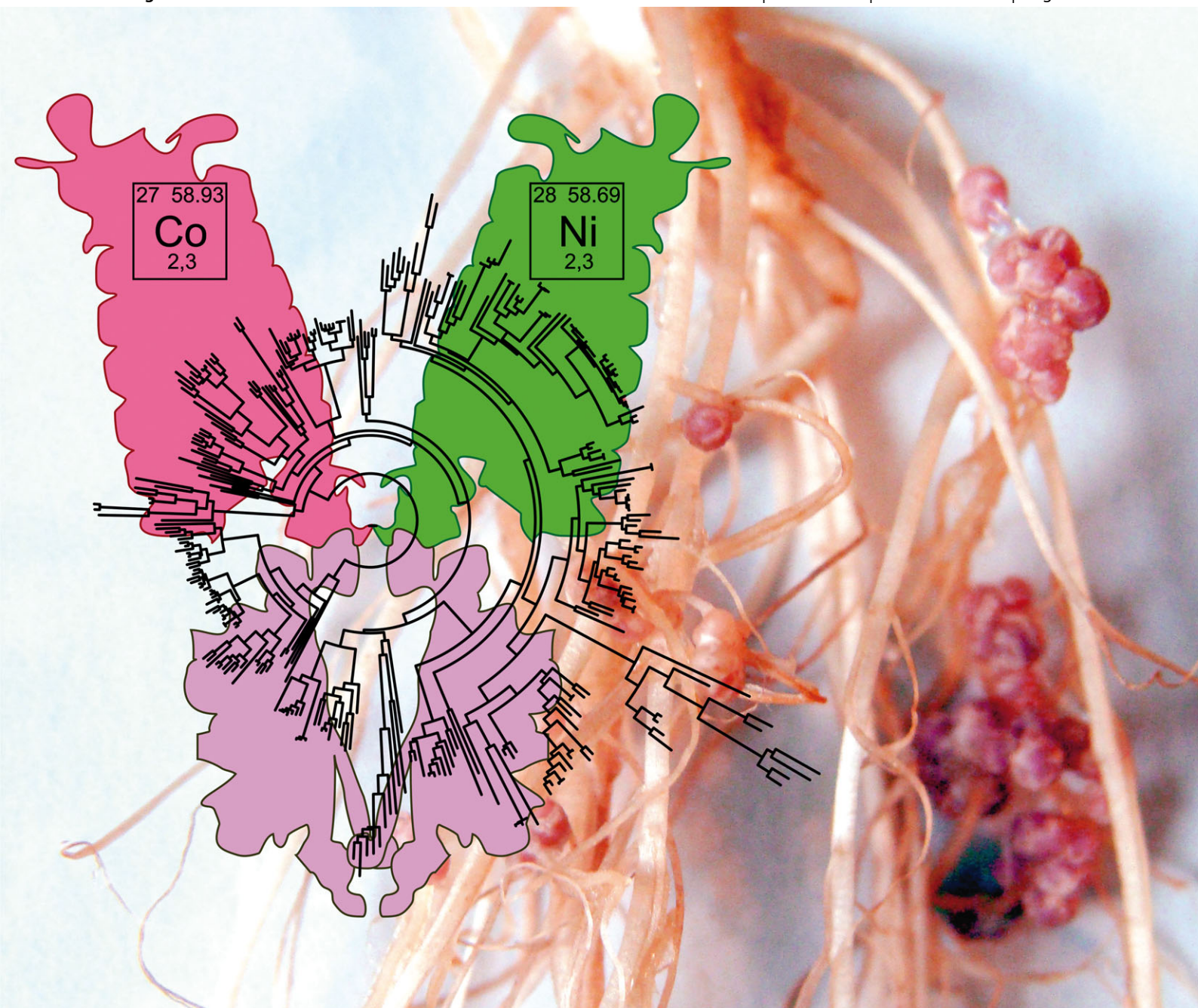
Además, se incluye una sección de resultados adicionales en donde se muestran resultados complementarios para entender mejor los procesos relacionados con la homeostasis de metales de transición en *Rhizobium etli* CFN42.

Metallomics

Integrated biometal science

www.rsc.org/metallomics

Volume 5 | Number 12 | December 2013 | Pages 1577–1702



ISSN 1756-5901

RSC Publishing

PAPER

Alejandro García-de los Santos *et al.*
Phylogenomic analysis of Cation
Diffusion Facilitator proteins uncovers
Ni²⁺/Co²⁺ transporters

Indexed in
MEDLINE!



1756-5901(2013)5:12;1-K

PAPER

Phylogenomic analysis of Cation Diffusion Facilitator proteins uncovers Ni²⁺/Co²⁺ transporters†

Cite this: *Metallomics*, 2013, 5, 1634

Ciro Cubillas,^{‡a} Pablo Vinuesa,^a Maria Luisa Tabche^b and Alejandro García-de los Santos^{‡*a}

The ubiquitous Cation Diffusion Facilitator proteins (CDF) play a key role in maintaining the cellular homeostasis of essential metal ions. Previous neighbor-joining phylogenetic analysis classified CDF proteins into three substrate-defined groups: Zn²⁺, Fe²⁺/Zn²⁺ and Mn²⁺. These studies were unable to discern substrate-defined clades for Ni²⁺, Co²⁺, Cd²⁺ and Cu²⁺ transporters, despite their existence in this family. In this study we improved the accuracy of this previous functional classification using a phylogenomic approach based on a thorough maximum-likelihood phylogeny and the inclusion of recently characterized CDF transporters. The inference of CDF protein function predicted novel clades for Zn²⁺, Fe²⁺, Cd²⁺ and Mn²⁺. The Ni²⁺/Co²⁺ and Co²⁺ substrate specificities of two clades containing uncharacterized proteins were defined through the functional characterization of *nepA* and *cepA* metal inducible genes which independently conferred Ni²⁺ and Co²⁺ resistances to *Rhizobium etli* CFN42 and increased, respectively, Ni²⁺/Co²⁺ and Co²⁺ resistances to *Escherichia coli*. Neither NepA nor CepA confer Zn²⁺, Fe²⁺ and Mn²⁺ resistances. The ability of NepA to confer Ni²⁺/Co²⁺ resistance is dependent on clade-specific residues Asn⁸⁸ and Arg¹⁹⁷ whose mutations produce a non-functional protein.

Received 25th July 2013,
Accepted 6th September 2013

DOI: 10.1039/c3mt00204g

www.rsc.org/metallomics

Introduction

Transporters of the Cation Diffusion Facilitator family (CDF, also known as cation-efflux or SLC30, TC 2.A.4.)¹ are distributed across the three domains of life. They maintain the cellular homeostasis of metal ions mainly Mn²⁺, Fe²⁺, Zn²⁺, and Cd²⁺ through extrusion, by means of a proton-dependent antiport mechanism conserved from bacterial to mammalian CDFs.^{2–4}

Detailed structural studies of the complete *Escherichia coli* Zn²⁺/Cd²⁺ transporter YiiP protein and of the Zn-bound *Thermus thermophilus* CzrB cytoplasmic domain revealed the presence of six transmembrane domains containing three metal binding sites (MBS) involved in metal transport (site A), stabilization of the Y-shaped homodimer (site B) as well as a metallochaperone-like fold where site C (involved in the turnover rate) is located.^{5,6}

One of the most important challenges in metalloprotein research is to establish structure–function relationships by

defining group-specific sequence motifs involved in metal binding, metal specificity or protein structure. Phylogenomic inference of protein function involving selection of homologs (with or without known function), multiple sequence alignment (MSA) and phylogenetic tree construction is a reliable approach which provides not only the evolutionary history of functional changes in a protein family, but also a rational basis for identifying molecular determinants of substrate specificity and characterizing protein structure.^{7,8} Thus, the estimation of a highly resolved phylogeny of the CDF family could represent a key approach to achieve a functional classification of CDF transporters into substrate-defined groups.

The pioneering phylogenetic study classified CDF proteins according to substrate specificity rather than species taxonomy and proposed that homologous proteins grouped into a clade are expected to share the same set of target metals. This study classified CDF proteins into three substrate-defined groups: Zn²⁺, Fe²⁺/Zn²⁺ and Mn²⁺ (ref. 9) but did not discern substrate-defined clades for Ni²⁺, Co²⁺, Cd²⁺ and Cu²⁺ transporters even though several CDF conferring resistance to these metals have been described.^{10–12} The aim of our study was to update and improve the phylogeny of CDF proteins in order to identify additional clades containing proteins with potentially new metal specificities using a phylogenomic approach. The inferred maximum-likelihood (ML) tree sorted CDF proteins into eighteen clades, contrasting with the most recent reported

^a Programa de Ingeniería Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apdo. Postal 565-A, Cuernavaca, Morelos, México. E-mail: alex@ccg.unam.mx; Fax: +52 777 317 5581; Tel: +52 55 5622 7691

^b Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. Fax: +52 777 317 2388; Tel: +53 777 329 1600

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3mt00204g

‡ These authors contributed equally to this work.

neighbor-joining (NJ) phylogeny that classified CDF proteins into just three major groups: Zn-, Fe/Zn- and Mn-CDFs.⁹ We found a correlation between Mn²⁺, Fe²⁺, Co²⁺, Ni²⁺, Cd²⁺ and Zn²⁺ substrates of CDF proteins and their corresponding phylogenetic clades. We propose the Ni²⁺ and Co²⁺ specificities for group III, based on the functional characterization of NepA (formerly RHE_PE00218); this protein confers Ni²⁺ resistance to the symbiotic nitrogen fixing bacteria *Rhizobium etli* CFN42, whose metal resistance mechanisms have been analyzed in our laboratory to determine its metal homeostasis machinery. The putative Co²⁺ specificity of group XII, deduced by the presence of *Cupriavidus metallidurans* CmDmeF, was supported by the functional characterization of its ortholog CepA (formerly RHE_CH01219), which specifically confers Co²⁺ resistance to *R. etli* CFN42. We also found that the NepA ability to confer Ni²⁺/Co²⁺ resistance is dependent on clade-specific Asn⁸⁸ and Arg¹⁹⁷ residues whose mutations produce a non-functional protein.

Results

Phylogenetic analysis classifies CDF proteins into 18 groups and uncovers novel clades with undefined substrate specificity

The data set analyzed in this study consisted of the previously reported 273 CDF homologs⁹ plus 56 not previously analyzed CDF proteins. These 329 proteins were collected from 116 species: 64 bacteria (119 proteins), 8 archaea (21 proteins) and 44 eukarya (189 proteins) (Table S1, ESI†). To define the most representative domain in our CDF sample containing proteins ranging in length from 209 to 991 amino acids, the 329 proteins were aligned with the Pfam Cation_efflux (PF01545.16) Hidden Markov Model domain which contains structural and functionally relevant residues (Fig. S1, ESI†). From the resultant alignment 11 proteins were removed because they lack of key structural elements such as the metal binding site (MBS) (indicated in Table S1, ESI†). The final 318-protein alignment, subsequently named the Pfam HMM-CDF domain, contained 22 putative CDF proteins from members of the order *Rhizobiales* which have not been previously analyzed as well as 44 characterized CDF proteins (Table S1, ESI†). Since the accuracy of the alignment is a critical step for the phylogenetic analysis, the Pfam HMM-CDF domain was additionally aligned with high-ranked alignment methods (MAFFT, MUSCLE, MSA-probs, Clustal Omega and ClustalW) and evaluated using MUMSA.¹³ For comparison purposes, we used ClustalW¹⁴ which was the same previously reported aligner.⁹ The aligners used in this study gave better alignment accuracy than ClustalW with exception of MSAprobs (Table S2, ESI†). The selection of the most accurate alignment is a very important difference between our study and Montanini's work because it will affect the inferred phylogeny. The best-scoring alignment was obtained with the ClustalO (Table S2, ESI†) and used for the ML tree search. Intensive tree searching was performed on this alignment using PhyML,¹⁵ as detailed in Methods.

The best phylogenetic tree was constituted by 18 clades (Fig. 1A), 16 of them are supported by *p*-values ≥ 0.90 whereas clades II and XI are supported by *p*-values = 0.80 and 0.88 respectively.

The substrate specificity for each clade was predicted by the presence of the 44 characterized CDF proteins (Table S1, ESI†) in different clades. Twelve clades had at least one characterized member (clades I, II, IV–XIII), 6 clades comprised only uncharacterized CDFs (III, XIV–XVIII clades) and 8 clades (V–XI and XIII) had at least one Zn²⁺ transporter, making Zn²⁺ transport polyphyletic and the most widespread feature of the CDF protein family (Fig. 1A). Characterized proteins in clades I and IV from bacteria and plants, respectively, transport Mn²⁺, while members of clade II share Fe²⁺ as a substrate. Group III had not been identified in previous phylogenies; it only comprises uncharacterized proteins, being unique because of the inclusion of archaeal CDF members. Archaeal and mycobacterial proteins predominate in this group; the putative *R. etli* CFN42 CDF protein NepA (RHE_PE00218) was also included in this clade as well as its only two orthologs found in the order *Rhizobiales*: CIATCH3409 from *R. phaseoli* CIAT652 and RL1175 from *R. leguminosarum* 3841 (Table S3, ESI†).

The characterized protein of group XII is CmDmeF, involved in Zn²⁺/Cd²⁺/Fe²⁺/Co²⁺/Ni²⁺ resistances in *C. metallidurans*.¹¹ Our phylogeny grouped CmDmeF together with the uncharacterized CDF protein of *R. etli* CepA (RHE_CH01219). The broad spectrum of metals recognized by CmDmeF precludes the assignment of the metal specificity to homologs from the same clade, which may also be of a broad spectrum.

The *R. etli* protein RHE_CH03072 was located in clade VI, together with Mn²⁺, Zn²⁺/Cd²⁺ and Fe²⁺ characterized transporters, whereas RHE_PD00193 was included in clade XVII with uncharacterized transporters. Both putative CDF proteins were not further analyzed.

Clades III and XII contain proteins involved in Ni²⁺ and Co²⁺ resistances

It has been previously reported that the expression of *Mycobacterium tuberculosis* *Rv2025c* gene, whose product is grouped into clade III, is specifically de-repressed by Ni²⁺ and Co²⁺ but not by Fe²⁺, Mn²⁺, Zn²⁺, Cd²⁺ or Cu²⁺ and controlled by the high-affinity Ni²⁺/Co²⁺ sensor KmtR.¹⁶ *Rv2025c* is unable to complement the Zn²⁺ sensitivity of the *M. smegmatis* *zitA* mutant¹⁷ and *M. smegmatis* expressing *Rv2025c in trans* does not significantly increase Zn²⁺ efflux compared to control cells without *Rv2025c*.¹⁸ These data suggest that *Rv2025c* may use Ni²⁺ and Co²⁺ as substrates, but not Zn²⁺.

A synteny analysis of archaeal genes encoding protein members of clade III revealed that the putative CDF proteins Mpal_2329, Mbar_A1220, and Mbar_A2147 are encoded next to the previously reported CbiMNQO-type Co²⁺ transport systems¹⁹ (Fig. 1B). In fact, the *Mpal_2330-Mpal_2333* and *Mbar_A1216-Mbar_A1219* genes encoding for putative CbiMNQO proteins overlap with the putative CDF-encoding genes *Mpal_2329* and *Mbar_A1220*, respectively. Thus, the phylogeny, gene context analysis and transcriptional activity of *Rv2025c* under Ni²⁺ and Co²⁺ stress, strongly suggests that clade III groups Ni²⁺/Co²⁺ transporters.

To validate the above hypothesis we generated a recombination-based mutation in the *nepA* gene (RHE_PE00218) encoding a putative *R. etli* CDF protein grouped into clade III. The mutant was tested for sensitivity to 100 μ M Co²⁺, 100 μ M Ni²⁺, 200 μ M

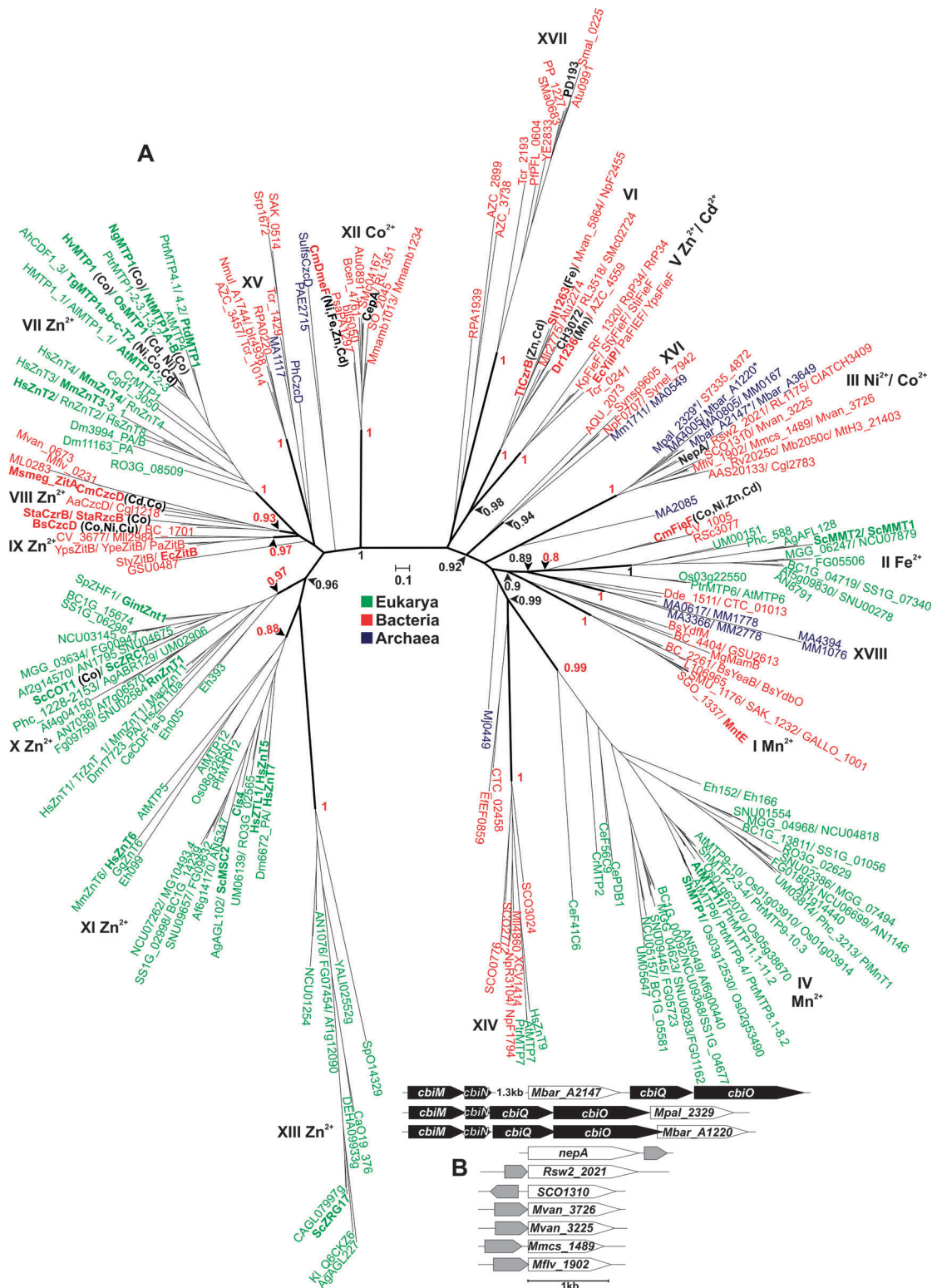


Fig. 1 Phylogenetic tree of CDF protein family and gene synteny of proteins grouped into clade III. (A) Unrooted, ML-based tree of the CDF family generated from characterized proteins (in bold) and their uncharacterized homologues. A key number as well as the shared metal substrate(s) by proteins per clade are indicated. Additional metal substrate(s) for characterized proteins with one or more non-shared substrates compared to the other proteins from the same clade are in black parentheses. Shimodaira–Hasegawa-like *p*-values for bipartitions indicating substrate-defined clades are shown in red, *p*-values (only ≥ 0.90) for bipartitions subtending two or more substrate-defined clades are marked in black. The scale bar indicates the expected number of amino acid substitutions per site under the LG + G + f model. Colors represent proteins from *eukarya* (green), *bacteria* (red) and *archaea* (blue). Putative CDF proteins from *R. etli* are located in clusters III (NepA), VI (CH3072), XII (CepA) and XVII (PD193) (black bold). Proteins with asterisks in group III may be part of putative Co^{2+} transport systems (see the text for details). (B) Proteins belonging to clade III located next to metalloregulators of the RcnR/CsoR and Arsr/SmtB families (in gray) or close to putative CbiMNQ-type Co^{2+} transporters (in black) are shown.

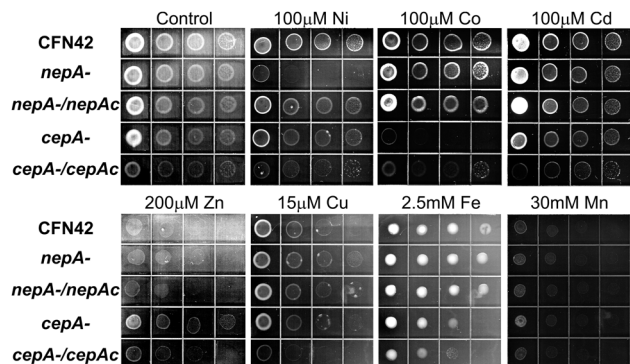


Fig. 2 *R. etli* *nepA* and *cepA* genes are essential for Ni^{2+} and Co^{2+} resistances. Cation sensitivity plate assay showing the growth of wild type CFN42, *nepA*⁻ and *cepA*⁻ mutants, and their derivatives complemented with the wild type genes *nepA*⁻/*nepAc*, and *cepA*⁻/*cepAc* in the presence or absence (control) of metals. Overnight cultures (OD₆₂₀ = 0.7) of each strain were serially diluted (10^{-1} – 10^{-4} , left to right) and 20 μl spotted onto MM plates with or without the indicated metals.

Zn^{2+} , 100 μM Cd^{2+} , 2.5 mM Fe^{2+} , and 15 μM Cu^{2+} , which were determined to be sub-lethal concentrations for *R. etli* CFN42. Fig. 2 shows growth inhibition of the *nepA* mutant only in the presence of 100 μM of Ni^{2+} , indicating that this gene is essential only for Ni^{2+} resistance, supporting the hypothesis that clade III groups Ni^{2+} transporters.

Substrate specificity for clade XII proteins is difficult to predict. *C. metallidurans* CmDmeF, the only characterized member of this clade, which is essential for Co^{2+} detoxification, also mediates resistance to a wide variety of divalent cations such as Zn^{2+} / Cd^{2+} / Fe^{2+} / Ni^{2+} .¹¹ In addition, phylogenetic analyses positioned CmDmeF together with the Zn^{2+} -CDF transporters ZnTs and ZitB.^{9,20} To understand the metal specificity of group XII, we determined the role of CepA (RHE_CH01219) in *R. etli* CFN42 metal homeostasis. For this purpose, we constructed a second recombination-based mutant in the *cepA* gene which showed growth inhibition only in the presence of 100 μM Co^{2+} , but was resistant to 100 μM Ni^{2+} and other metals (Fig. 2), indicating that this gene is only essential for Co^{2+} resistance. These findings support the hypothesis that CDF proteins grouped into clade XII share Co^{2+} as a substrate.

The *nepA* and *cepA* mutants, complemented with their respective genes, recovered the wild type levels of tolerance toward Ni^{2+} and Co^{2+} , confirming their essential role in resistance to these metals, discarding possible polar effects of these mutations on downstream genes.

The *nepA* and *cepA* gene expression in *R. etli* CFN42 is differentially induced by Ni^{2+} , Co^{2+} and Cd^{2+} but not by Zn^{2+}

As a part of the functional characterization of *R. etli* CDF proteins belonging to groups III and XII, we explored if the CDF-encoding genes are metal responsive by measuring the influence of Ni^{2+} , Co^{2+} , Zn^{2+} and Cd^{2+} ions on gene expression (Fig. 3) through quantitative reverse transcription-PCR (qRT-PCR). The increased relative expression of *nepA* gene averaged 9.2-, 4.7- and 11.9-fold by the addition of Ni^{2+} , Co^{2+} , and Cd^{2+} , respectively (Fig. 3).

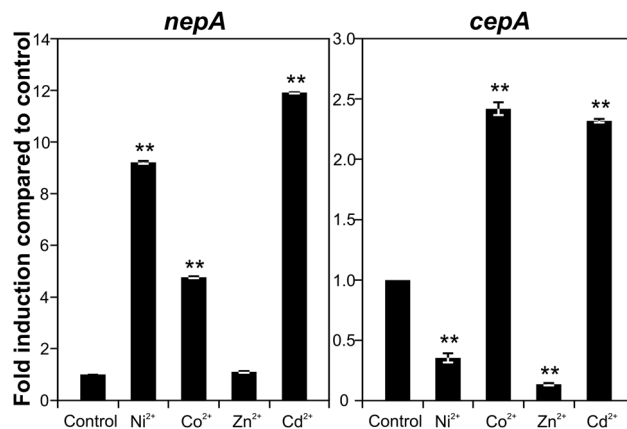


Fig. 3 The *R. etli* *nepA* and *cepA* CDF-encoding genes are differentially induced by Ni^{2+} , Co^{2+} and Cd^{2+} but not by Zn^{2+} . The data represent the fold induction defined as the ratio between the mRNA levels of *nepA* and *cepA* genes in the presence of 1 mM of NiCl_2 , ZnCl_2 , CdCl_2 and CoCl_2 and mRNA levels in the absence of metals (control). Both were normalized to *hisCd* mRNA levels as published⁵⁶ (mean \pm s.d., $n = 4$) (t-test, ** $p \leq 0.01$).

The *cepA* gene increased its relative expression by the addition of Co^{2+} (2.4-fold) and Cd^{2+} (2.3-fold), whereas no significant increase was observed by the addition of Ni^{2+} or Zn^{2+} (Fig. 3). Since *cepA* has a 6.85-fold higher expression in Co^{2+} compared to Ni^{2+} , while *nepA* has 1.93-fold higher expression in Ni^{2+} compared to Co^{2+} , these results indicate that Ni^{2+} is a better inducer than Co^{2+} for *nepA* but Co^{2+} is better than Ni^{2+} for *cepA*. Interestingly, Zn^{2+} is not an inducer for these CDF-encoding genes; instead, down-regulates *cepA* expression as Ni^{2+} does.

The heterologous expression of NepA and CepA in *E. coli* confirms their specialized role in Ni^{2+} and Co^{2+} resistances

Since Ni^{2+} and Co^{2+} have similar physicochemical properties, we decided to determine whether NepA and CepA are able to recognize both metals, as happens in other non-CDF proteins such as RcnA or NccA.^{21,22} For this, the resistance to Ni^{2+} , Co^{2+} , Zn^{2+} and Cd^{2+} of *E. coli* MC4100 cells²¹ expressing the *R. etli* *nepA* and *cepA* genes in *trans* was analyzed. Fig. 4A shows that the presence of the wild type *nepA* gene in *E. coli* increases 10 000-fold the number of CFUs found in the presence of Ni^{2+} and Co^{2+} as compared to cells carrying the empty vector. In contrast, Zn^{2+} and Cd^{2+} resistances were not significantly increased under comparable conditions, despite the fact that NepA has a putative metal binding site A (HD–HD residue composition) identical to 21 Zn-CDF transporters (Fig. S2, ESI[†]). On the other hand, the wild type *cepA* gene in *E. coli* increases 10 000-fold the number of CFUs found in the presence of Co^{2+} but not of Ni^{2+} compared to cells carrying the empty vector (Fig. 4A). As *nepA*, *cepA* does not increase Zn^{2+} or Cd^{2+} tolerances in *E. coli*. These results indicate that: (i) NepA recognizes Co^{2+} and Ni^{2+} in *E. coli*, which is not observed in *R. etli*, (ii) in contrast to its CmDmeF ortholog, CepA has a restricted Co^{2+} specificity both in *E. coli* and in *R. etli* and (iii) neither NepA nor CepA increase Zn^{2+} or Cd^{2+} tolerances in *E. coli* or *R. etli* under our assay conditions.

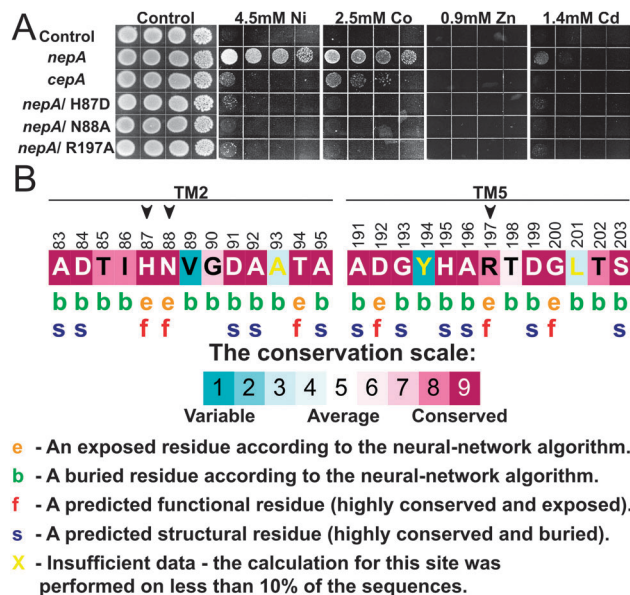


Fig. 4 The *nepA* and *cepA* genes increase $\text{Ni}^{2+}/\text{Co}^{2+}$ and Co^{2+} resistances in *E. coli*, respectively. (A) Metal resistance assay plate of *E. coli* cells transformed with the empty plasmid PCR2.1-TOPO (control), the plasmids containing the wild type *nepA* and *cepA* genes, and *nepA* with point mutations in residues H87 (*nepA/H87D*), N88 (*nepA/N88A*) and R197 (*nepA/R197A*). Dilutions 10^{-1} to 10^{-4} are shown (left to right). *E. coli* cells harboring the mutant *nepA* genes showed reduced Ni^{2+} and Co^{2+} resistances. (B) A ConSurf analysis reveals three highly conserved amino acid positions: H87, N88 and R197 (black arrows), near the putative MBS A (residues H87, D91, H195 and D199) of clade III proteins, whose mutations disrupt NepA function. The numbering corresponds to the NepA protein sequence.

The highly conserved H87, N88 and R197 residues are required for NepA function

Since amino acids constituting site A of CzcD^{H49} (ref. 23), YiiP^{D45} (ref. 24) and ZitB^{H53} (ref. 25) play a crucial role in metal binding and transport, the question arose whether the homologous residue, NepA^{H87} (Fig. S2, ESI[†]), is also functionally critical. Fig. 4A shows that a NepA^{H87D} mutation suppresses the ability of this protein to confer Ni^{2+} and Co^{2+} resistances to *E. coli*, when compared to the wild type gene. This result suggests that H87 residue could be part of the NepA metal binding site A. Although the H87D substitution in NepA produced a DDHD residue composition identical to the YiiP metal binding site A, the modified protein was unable to increase Zn^{2+} or Cd^{2+} resistances, compared to the empty vector.

To reinforce the hypothesis that proteins grouped into clade III share both Ni^{2+} and Co^{2+} as substrates, and bearing in mind the higher number of coordinating ligands in Ni^{2+} and Co^{2+} metalloproteins, we focused on the identification of group-specific residues which potentially act as ligand donors, and are adjacent to the putative NepA metal binding site A. Since it was recently shown that a L91M mutation in AtMTP1 (clade VII) extends the selectivity of this Zn^{2+} -transporting CDF to Co^{2+} (ref. 26), we predicted that either Co^{2+} or Ni^{2+} specificity could be determined by residues in the homologous positions in NepA and conserved among clade III proteins. A ConSurf-analysis²⁷ of amino acids adjacent to the putative site A of group III proteins

revealed not only the N88 as a conserved residue, but also R197 both presenting potential N-ligand donors (Fig. 4B) which are group III-specific (Fig. S3, ESI[†]). To establish their role in protein function, we examined the effect of N88A and R197A mutations on the ability of NepA to increase *E. coli* Ni^{2+} and Co^{2+} resistances. We found that each of these two point mutations suppressed the ability of NepA to increase Ni^{2+} and Co^{2+} *E. coli* resistances (Fig. 4A). This result suggests that N88 and R197 residues might play a key role in protein function reinforcing our hypothesis that proteins grouped into clade III share Co^{2+} as a substrate. However, we cannot discount the possibility that these mutant proteins may have unspecific structural alterations that affect their proper function.

This result not only suggests that N88 and R197 residues play a key role in protein function but also reinforces our hypothesis that proteins grouped into clade III share Co^{2+} as a substrate.

Currently, it is not known if the NepA N-terminal His-stretch or the CepA centrally-located His-stretch which are conserved in clades III and XII (Fig. S3, ESI[†]), respectively; are functionally required to recognize their metal substrates.

Discussion

In this study we used a phylogenomic approach^{7,8} to improve the accuracy of the functional classification of CDF proteins by finding novel groups of CDF transporters. This new classification increases our capabilities to predict the substrates which are recognized by groups of ortholog proteins belonging to any given clade. Our phylogeny revealed several important improvements over the most recently reported NJ phylogenetic analysis of CDF transporters.⁹ Below we discuss how these improvements are supported.

The ML-based phylogeny expands CDF classification into 18 independent clades

Based on the occurrence of characterized proteins, our analysis predicted the substrate specificity for 12 clades. The six clades with Zn^{2+} specificity (VII–XI and XIII) and one clade with Mn^{2+} containing eukaryotic transporters (IV) agree with Zn -CDF and Mn -CDF groups described in previous phylogenies. In contrast with the previously proposed Fe/Zn -CDF group,⁹ our phylogeny separates the transporters grouped in this sole cluster into independent clades with $\text{Zn}^{2+}/\text{Cd}^{2+}$ (VI), $\text{Co}^{2+}/\text{Ni}^{2+}$ (III and XII), Fe^{2+} (II), and $\text{Zn}^{2+}/\text{Cd}^{2+}/\text{Fe}^{2+}/\text{Mn}^{2+}$ (VI) specificities. Prokaryotic Mn^{2+} -CDF transporters, not deduced from previous phylogenies, were grouped into clades I and VI separately of eukaryotic Mn^{2+} transporters (in clade IV). Five clades containing only uncharacterized CDFs were also identified (XIV–XVIII).

The novel clade III groups Ni^{2+} and Co^{2+} transporters from archaea, actinobacteria and rhizobia

The involvement of *R. etli* NepA in Ni^{2+} tolerance, the enhanced Ni^{2+} and Co^{2+} resistances of *E. coli* expressing NepA, the transcriptional activation of *Rv2025c*¹⁶ and *nepA* genes by Ni^{2+} and Co^{2+} , as well as the presence of potential operons containing

CDF NepA orthologs overlapping with putative archaeal CbiMNQO-type Co^{2+} transport proteins support the hypothesis that proteins grouped into clade III share Ni^{2+} and Co^{2+} as substrates. Notably, neither *Rv2025*⁷ nor *nepA* genes are induced by Zn^{2+} (ref. 16). Furthermore, *nepA* does not confer resistance to Zn^{2+} in *R. etli* or *E. coli*, while *Rv2025c* does not complement the Zn^{2+} sensitivity of the *M. smegmatis zitA* mutant,¹⁷ suggesting a primary role in Ni^{2+} and Co^{2+} transport. Some of the NepA uncharacterized orthologs (Cgl2783, SCO1310, AAS20133, MA4005, MM0167, MA0805) were previously included in the Fe/Zn-CDF clade;⁹ however, because group III *Rv2025c* and NepA transporters are not involved in Zn^{2+} tolerance, we propose this group as a *bona fide* Ni^{2+} and Co^{2+} clade. The previously known $\text{Ni}^{2+}/\text{Zn}^{2+}$ and $\text{Co}^{2+}/\text{Zn}^{2+}$ transporters such as *Thlaspi goesingense* TgMTP1t2,¹² *Bacillus subtilis* Czcd,¹⁰ *Saccharomyces cerevisiae* COT1²⁸ or *C. metallidurans* Czcd²³ are polyphyletic and do not belong to group III or XII. Currently, we cannot distinguish if Ni^{2+} transport evolved more than once in these clades, or if Ni^{2+} or Co^{2+} specificity was derived from Zn^{2+} transporters. The latter hypothesis is supported by the observation that Zn^{2+} transporters are the dominant proteins found in the clades that harbor these $\text{Ni}^{2+}/\text{Co}^{2+}$ transporters. The metal specificity shared among members of this clade is also supported by the discovery of three highly conserved residues required for NepA function (Fig. 4).

Discovery of prokaryotic Mn-CDF clades

The Mn-CDFs reported in previous phylogenetic studies comprised exclusively proteins from plants and fungi.⁹ In our data set we included the recently discovered bacterial Mn-CDF MntE (clade I) from *Streptococcus pneumoniae*,²⁹ Dr1236 (clade VI) from *Deinococcus radiodurans*³⁰ and some of their homologs. In our phylogeny they were recovered in two independent clades (I and VI). This clustering is consistent with a polyphyletic origin of Mn-CDF proteins.

The novel clade XII groups transporters sharing Co^{2+} as a substrate

Our phylogeny separates the broad metal specificity transporter *C. metallidurans* CmDmeF,¹¹ conferring $\text{Zn}^{2+}/\text{Co}^{2+}/\text{Cd}^{2+}/\text{Ni}^{2+}/\text{Fe}^{2+}$ resistance into the new and independent clade XII. This protein was previously grouped with ZitB and ZnT Zn^{2+} transporters.^{9,20} The Co^{2+} specificity of clade XII is supported by the reported central role of CmDmeF in Co^{2+} homeostasis,¹¹ as well as the functional characterization of its *R. etli* ortholog CepA (this study), which only confers Co^{2+} resistance to this organism.

Rationalized arrangement of Fe^{2+} and Zn^{2+} -CDF transporters

The Fe/Zn group⁹ was defined based on the presence of Fe^{2+} -transporting CDFs CmFieF/ScMMT1-2^{11,31} and the $\text{Zn}^{2+}/\text{Cd}^{2+}$ transporter EcYiiP³² in the same clade. Our phylogeny separates EcYiiP in clade V and CmFieF/ScMMT1-2 in clade II. This phylogenetic divergence is congruent with *in vitro* EcYiiP studies showing that this protein is unable to mobilize Fe^{2+} and, instead, transports Zn^{2+} or Cd^{2+} (ref. 33). A key functional difference between CmFieF and EcYiiP is that CmFieF is able to increase Fe^{2+} tolerance in the heterologous model *E. coli* $\Delta\text{YiiP}/\Delta\text{Fur}$

double mutant¹¹ but EcYiiP is unable to complement Fe^{2+} uptake in the *Synechocystis* sp. Sll1263 mutant.³⁴ In addition, disruption of *C. metallidurans* *fieF* has only a marginal influence on Zn^{2+} tolerance¹¹ while ScMMT1-2 is clearly involved in Fe^{2+} mitochondrial transport.³¹ In the current state, we do not know whether CmFieF gained Co, Ni, Zn, Cd specificities or if ScMMT1-2 lost these 4 specificities; however, we assume their common ancestor recognized Fe, the shared substrate. This fact supports clade II as a Fe-CDF group.

The novel clade VI groups proteins with $\text{Fe}^{2+}/\text{Zn}^{2+}/\text{Cd}^{2+}/\text{Mn}^{2+}$ specificities

Two CDF transporters belonging to clade VI were recently characterized: Sll1263 from *Synechocystis* sp.³⁴ and Dr1236 from *Deinococcus radiodurans*,³⁰ they are involved in Fe and Mn transport, respectively. The independence of these transporters from CmFieF (Fe, clade II) and MntE (Mn, clade I) suggests Fe^{2+} and Mn^{2+} recognition by CDFs evolved at least twice. Clade VI also includes *T. thermophilus* CzrB which increases *E. coli* $\text{Zn}^{2+}/\text{Cd}^{2+}$ resistance³⁵ and binds Zn at its cytoplasmic domain.⁶ Neither Sll1263 nor Dr1236 are involved in Zn tolerance but the ability of TtCzrB to bind or transport Mn or Fe is unknown.^{6,35} The characterization of additional members of this clade will provide new insights into the predominant specificities, if any, in this group of proteins.

Absence of single Cd^{2+} and Cu^{2+} CDF clades

The examination of all known Cd-CDF proteins present in our data set revealed that they are also Zn-CDF proteins. Most of them are grouped into clades V, VI and VIII. This distribution is probably due to the strong preference shared by both metals for tetrahedral coordination. Using site-directed mutagenesis, exclusive Cd^{2+} transport could not be generated from the $\text{Zn}^{2+}/\text{Cd}^{2+}$ transporter YiiP; single Zn^{2+} transport could be obtained from YiiP by decreasing its affinity for Cd^{2+} , but Cd^{2+} transport could be achieved by increasing its affinity for the Zn^{2+} transporter ZnT5.³⁶

On the other hand, no Cu-CDF clades could be directly predicted from our phylogeny. Although most transporters with known Cu specificity belong to the P_{IB} -type ATPases protein family,³⁷ there is one reported CDF protein (Czcd) that is related to Cu^{2+} -tolerance in *Bacillus subtilis*.¹⁰ In our phylogeny this protein was grouped together with canonical Zn-CDF transporters such as CmCzcd in clade VIII. Perhaps proteins with separate Cd^{2+} and Cu^{2+} specificities might be present in some of the five clades containing only uncharacterized CDF proteins (XIV–XVIII).

Group specific His-rich stretches influence CDF protein functional divergence of metal specificity

It has been suggested that N-terminal, or centrally-located H-rich sequence regions present in CDF proteins such as plant CDFs from group VII,^{12,38,39} and proteins from clades XII and III (Fig. S3, ESI†) could play a relevant role in protein function, narrowing or broadening metal selectivity. In fact, an AtMTP1 chimera containing the HvMTP1 His-stretch (both proteins

located in clade VII) modified the AtMTP1 substrate specificity from Zn^{2+} to $\text{Zn}^{2+}/\text{Co}^{2+}$, gaining Co^{2+} recognition.³⁸ In addition, the removal of the VTVIT motif located in the AtMTP1 His-stretch allowed acquisition of Co^{2+} specificity to this Zn^{2+} transporter.³⁸ Since all clade VII proteins have His-stretches, it is difficult to ascertain whether they evolved from a common $\text{Zn}^{2+}/\text{Co}^{2+}/\text{Ni}^{2+}$ ancestor transporter that differentially lost $\text{Co}^{2+}/\text{Ni}^{2+}$ recognition (as in AtMTP1) or if the ancestor was a Zn^{2+} transporter which differentially gained $\text{Co}^{2+}/\text{Ni}^{2+}$ recognition (as in HvMTP1, NgMTP1, NtMTP1A/B). Whatever the scenario, it is clear that Zn^{2+} is the ancestral metal specificity shared by all members of this clade. These His-rich regions have been suggested to act as metal chaperones which may increase the local metal concentration above the threshold required for selective transport, thus forcing subsequent non-target metal translocation.^{38–40} Alternatively, they could modulate protein function acting as cytoplasmic sensors of metal concentrations probably working as buffering pockets.⁴⁰ The presence of a C-terminal Leu zipper motif in PtdMTP1 and in its orthologs (clade VII), ShMTP1-4/AtMTP8-10 (clade IV) has also been related to functional activity and metal selectivity of CDFs.^{26,41}

Methods

Dataset and classification of CDF proteins

The data set analyzed in this study contained 329 homolog proteins, 273 of them represent the complete data set reported in a previous phylogeny.⁹ As part of our interest in characterizing the metal homeostasis in rhizobia, the data set also contains 23 putative CDF proteins from six members of the order *Rhizobiales*, which have not been previously analyzed. Based on experimental evidence, we considered 44 CDF proteins as functionally characterized, with their substrates being known (Table S1, ESI[†]). These proteins allowed the correlation of CDF groups with substrate specificity.

Programs and parameters for multiple sequence-alignment of the CDF protein signature blocks

The strong evolutionary divergence of the compiled data set is clearly reflected in its sequence-length heterogeneity, which ranges from 209 to 991 amino acids. An alignment of the more conserved regions in the dataset was obtained with hmalign from the HMMER 3.0 package,⁴² in combination with the HMMER3/b cation efflux family HMM model PF01545.16, downloaded from Pfam.⁴³ The resulting alignment was named the Pfam HMM-CDF domain. In an attempt to further improve this alignment, it was de-gaped and subjected to standard global multiple sequence alignment (MSA) with some of the best ranking software in recent comparative benchmark analyses,⁴⁴ we chose Clustal Omega v1.1.0,⁴⁵ MAFFT v6.935,⁴⁶ MSAProbs v0.9.7,⁴⁷ ClustalW v.2.1¹⁴ and MUSCLE v3.8.31.⁴⁸ All programs were run using accuracy-oriented settings.

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile–profile techniques to generate alignments. We took advantage of this capacity to perform a model-based global multiple sequence alignment.

For this purpose we used the above mentioned CDF-HMM (acc. no. PF01545) to perform the alignment (using `-full-iter -full -iter 5` option).

We ran MAFFT using the accuracy-oriented method L-INS-i (an iterative refinement method incorporating local pairwise alignment information).

MSAProbs is recently released alignment software.⁴⁷ It implements a combination of pair HMMs and partition functions to calculate posterior probabilities. It also incorporates weighted probabilistic consistency transformation and weighted profile–profile alignment strategies to improve alignment accuracy. It was run using `-ir 100`. ClustalW was run using default settings.

Finally, we also used MUSCLE, which implements fast distance estimation using *k*mer counting, progressive alignment using a profile function called the log-expectation score, and refinement using tree-dependent restricted partitioning. MUSCLE was run with `maxiters = 32` and `maxtrees = 5`, with 3 refinement rounds.

The quality of the alignments containing 318 CDF sequences produced by the different software and alignment strategies described above was evaluated using MUMSA¹³ (Table S2, ESI[†]). The best-scoring alignments were obtained with ClustalO and MAFFT L-INS-i. We chose the former one for phylogenetic analysis.

ML-tree search strategies

Only heuristic methods are capable of estimating a phylogeny with 318 terminals. Due to the hill-climbing nature of the branch-swapping algorithms, tree searches get stuck in local optima. To improve phylogeny estimation, we used 100 random seed trees in addition to a BioNJ tree to start 101 searches. The random trees were generated by an *ad hoc* bioperl script. Tree searching under the maximum-likelihood criterion was performed with PhyML v3.0.¹⁵ The best-fitting substitution matrix was selected using ProtTest3⁴⁹ and a non-redundant subset of the original data set consisting of 102 proteins with $\leq 55\%$ identity generated with CD-hit.⁵⁰ The selected model was LG + G + f (AIC weight = 1). This is the recently derived LG substitution matrix,⁵¹ the gamma distribution to model among-site rate heterogeneity and estimating amino acid frequencies empirically from the data. The “BEST” option for branch swapping was selected, which keeps the best-scoring tree of either NNI or TBR moves in each iteration. The bipartition support was evaluated by the Shimodaira–Hasegawa-like approximate LRT test.⁵² The tree searches were performed on a 13 node compute cluster, each node having 12×2.16 GHz x86_64 CPUs with 16GB RAM. The best tree from each search was identified as the one with the highest log likelihood score.

Identification of conserved amino acids in the CDF protein domain

Based on the ClustalO alignment, we analyzed the position-specific estimated evolutionary rate of amino acids present in the putative site A and its neighborhood for clade III proteins, and compared them with known Zn^{2+} -CDF transporters from the ones of those from clades V, VII, X, XI and a subset of

21 characterized Zn²⁺-specific CDF (Table S1, ESI[†]) using ConSurf.²⁷

Bacterial strains, media and conditions

The strains are listed in Table S4 (ESI[†]). *Rhizobium etli* strains were grown at 30 °C in rich PY medium and minimal medium (MM),⁵³ while *E. coli* strains were grown in Luria-Bertani medium at 37 °C. Antibiotics for *R. etli* were added at the following concentrations (μg ml⁻¹): nalidixic acid, 20; streptomycin, 100; gentamicin, 5; and tetracycline, 3. For *E. coli* the antibiotic concentrations were (μg ml⁻¹): kanamycin, 30; gentamicin and tetracycline 10.

Cation sensitivity assays

Metal sensitivity was determined using a plate assay as follows: 50 mM stock solutions of Co, Ni, Fe, Mn, Cu, Zn and Cd chloride salts (Sigma-Aldrich, St Louis, MO) were prepared in Milli-Q water, filter sterilized and added at increasing concentrations to solid (1.5% wt/vol agar) MM. The *R. etli* overnight cultures were adjusted to OD₆₂₀ = 0.7, washed twice with MgSO₄ 10 mM, serially diluted (10⁻¹–10⁻⁴) and spotted (20 μl) on solid MM with or without metal ions as controls. Growth was recorded after 7 days of incubation at 30 °C. The MICs of metals able to reduce, in at least one log-unit, the growth (CFU) of *R. etli* wild type and mutant strains were 100 μM Ni²⁺, 100 μM Co²⁺, 200 μM Zn²⁺, 100 μM Cd²⁺, 15 μM Cu²⁺, 2.5 mM Fe²⁺ and 30 mM Mn²⁺.

Genetic manipulations

R. etli mutants were generated by recombination-based vector integration mutagenesis. Internal 0.3–0.5 kbp DNA fragments of *cepA* and *nepA* genes were amplified by PCR with primers C/D, E/F, respectively (Table S5, ESI[†]), cloned into suicide plasmid pPDGm⁵⁴ and mobilized into *R. etli* CFN42 by triparental mating. Disruption of target gene by single crossover was confirmed by Southern blot hybridization.⁵⁵ To complement the *nepA* and *cepA* mutants, full-length genes were amplified by PCR using primers I/J, M/N, respectively, cloned into pCR 2.1 TOPO, subcloned into pBBR1MMCS-3 (*cepA*) or pRK7813 (*nepA*) and mobilized by conjugation to complement the respective mutants. The expression of the cloned genes was under the control of the vector's *lacZ* promoter. For site-directed mutations at H87, N88 and R197 conserved residues in NepA site A the QuickChange multi site-directed mutagenesis kit (Agilent) was used along with primers H87D-P1, N88A-P1 and R197A-P1. The fidelity of single site mutations was verified by DNA sequencing.

Metal-dependent expression of CDF-encoding genes in *R. etli* CFN42

10 ml of mid-log (OD₆₂₀ = 0.45–0.55) growing cells were exposed for 30 min to 1 mM of NiCl₂, ZnCl₂, CdCl₂ and CoCl₂ in PY rich medium. Control cells unexposed to metals were also included. Then, mRNA was extracted by using the TriPure isolation reagent (Roche). The total RNA (DNA free) was reverse transcribed to cDNA by using ReverAid H minus FirstStrand cDNA Synthesis (Fermentas). Quantitative real-time PCR was performed on PCR

System 3700 (Applied Biosystems) using Maxima Syber Green/ROX qPCR master Mix (Fermentas). The *nepA*, *cepA* and *hisCd* genes were amplified by using primers R/S, T/U and Y/Z, respectively. Their expression levels in the presence and absence of metals were normalized to the expression level of house-keeping *hisCd* gene. The data represent averages of four independent experiments with three technical replicates each. The fold change in gene expression was calculated using the ΔΔC_T method.⁵⁶

Heterologous complementation of *E. coli* cells with *R. etli* CDF-encoding genes

The influence of *nepA* and *cepA* genes on metal resistance was determined by transforming chemically competent *E. coli* MC4100 cells²¹ with plasmid pCR2.1 TOPO harboring *cepA*, *nepA* and their respective point mutations as well as the empty vector as control. The metal resistances of the resultant transformed clones were tested by using a LB plate assay (Fig. 4). 12 h grown cells were adjusted to OD₅₅₀ = 1, washed twice with MgSO₄ 10 mM, serially diluted (10⁻¹–10⁻⁴), spotted (15 μl) on plates and incubated 24 h at 37 °C.

Conclusions

The combined use of evolutionary and experimental information on a phylogenomics framework allowed the inference of a new functional classification for the CDF protein family where Ni, Co, Fe, Mn, Zn and Cd could be associated with highly supported clades.

The ML-based phylogeny assorted the CDF proteins in 18 well supported phylogenetic clades, in contrast with the previously reported three-group classification, implying that any given metal could be present in two or more clades.

This thorough phylogenetic analysis enabled the discovery of two novel clades for Ni²⁺/Co²⁺ and Co²⁺ transporters, experimentally supported by the functional characterization of *nepA* and *cepA* genes of *R. etli* CFN42. New clades for Zn²⁺, Fe²⁺, Cd²⁺ and Mn²⁺ were also inferred.

The presence of Mn-CDF transporters in bacterial genomes plus those found in eukaryotic organisms support the idea of a polyphyletic evolution of this group of transporters.

The comparison of the position-specific evolutionary rate through ConSurf in Zn vs. Ni/Co CDF-transporters allowed testing of a hypothesis about the association among highly conserved residues, protein function, and metal recognition. Two conserved position-specific residues were found in the novel group of Ni/Co-CDF transporters located in clade III that are relevant for protein function.

Acknowledgements

We gratefully acknowledge David Romero and Susana Brom for critically reviewing the manuscript, Barbara Montanini for sharing her original CDF dataset with us, and Laura Cervantes, Araceli Dávalos, Jose Ocelotl and Elena Lounejeva (Instituto de Geología, UNAM) for their skillful technical assistance. We also

acknowledge Universidad Nacional Autónoma de México where this research was done and Programa de Doctorado en Ciencias Biomédicas where C.C. studied his PhD, this article is part of his PhD thesis. We thank Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT Grant number IN201112 to A.G.S) for supporting this research, Consejo Nacional de Ciencia y Tecnología (México) for the PhD scholarship given to C.C. (CVU 269108) and Ciencia Básica Grant number 179133 to P.V., for the bioinformatics support.

Notes and references

- 1 T. Paulsen and M. H. Saier, Jr., *J. Membr. Biol.*, 1997, **156**, 99–103.
- 2 Y. Chao and D. Fu, *J. Biol. Chem.*, 2004, **279**, 12043–12050.
- 3 A. Guffanti, Y. Wei, S. V. Rood and T. A. Krulwich, *Mol. Microbiol.*, 2002, **45**, 145–153.
- 4 E. Ohana, E. Hoch, C. Keasar, T. Kambe, O. Yifrach, M. Hershfinkel and I. Sekler, *J. Biol. Chem.*, 2009, **284**, 17677–17686.
- 5 M. Lu, J. Chai and D. Fu, *Nat. Struct. Mol. Biol.*, 2009, **16**, 1063–1067.
- 6 V. Cherezov, N. Hofer, D. M. Szebenyi, O. Kolaj, J. G. Wall, R. Gillilan, V. Srinivasan, C. P. Jaroniec and M. Caffrey, *Structure*, 2008, **16**, 1378–1388.
- 7 J. A. Eisen, *Genome Res.*, 1998, **8**, 163–167.
- 8 D. Brown and K. Sjolander, *PLoS Comput. Biol.*, 2006, **2**, e77.
- 9 B. Montanini, D. Blaudez, S. Jeandroz, D. Sanders and M. Chalot, *BMC Genomics*, 2007, **8**, 107.
- 10 C. M. Moore, A. Gaballa, M. Hui, R. W. Ye and J. D. Helmann, *Mol. Microbiol.*, 2005, **57**, 27–40.
- 11 D. Munkelt, G. Grass and D. H. Nies, *J. Bacteriol.*, 2004, **186**, 8036–8043.
- 12 M. W. Persans, K. Nieman and D. E. Salt, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 9995–10000.
- 13 T. Lassmann and E. L. Sonnhammer, *Nucleic Acids Res.*, 2005, **33**, 7120–7128.
- 14 M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, 2947–2948.
- 15 S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel, *Syst. Biol.*, 2010, **59**, 307–321.
- 16 D. R. Campbell, K. E. Chapman, K. J. Waldron, S. Tottey, S. Kendall, G. Cavallaro, C. Andreini, J. Hinds, N. G. Stoker, N. J. Robinson and J. S. Cavet, *J. Biol. Chem.*, 2007, **282**, 32298–32310.
- 17 A. Grover and R. Sharma, *J. Bacteriol.*, 2006, **188**, 7026–7032.
- 18 G. Riccardi, A. Milano, M. R. Pasca and D. H. Nies, *FEMS Microbiol. Lett.*, 2008, **287**, 1–7.
- 19 D. A. Rodionov, P. Hebbeln, M. S. Gelfand and T. Eitinger, *J. Bacteriol.*, 2006, **188**, 317–327.
- 20 D. H. Nies, *FEMS Microbiol. Rev.*, 2003, **27**, 313–339.
- 21 A. Rodrigue, G. Effantin and M. A. Mandrand-Berthelot, *J. Bacteriol.*, 2005, **187**, 2912–2916.
- 22 T. Schmidt and H. G. Schlegel, *J. Bacteriol.*, 1994, **176**, 7045–7054.
- 23 A. Anton, A. Weltrowski, C. J. Haney, S. Franke, G. Grass, C. Rensing and D. H. Nies, *J. Bacteriol.*, 2004, **186**, 7499–7507.
- 24 Y. Wei and D. Fu, *J. Biol. Chem.*, 2006, **281**, 23492–23502.
- 25 S. M. Lee, G. Grass, C. J. Haney, B. Fan, B. P. Rosen, A. Anton, D. H. Nies and C. Rensing, *FEMS Microbiol. Lett.*, 2002, **215**, 273–278.
- 26 M. Kawachi, Y. Kobae, S. Kogawa, T. Mimura, U. Kramer and M. Maeshima, *FEBS J.*, 2012, **279**, 2339–2356.
- 27 F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz and N. Ben-Tal, *Bioinformatics*, 2003, **19**, 163–164.
- 28 D. S. Conklin, J. A. McMaster, M. R. Culbertson and C. Kung, *Mol. Cell. Biol.*, 1992, **12**, 3678–3688.
- 29 J. W. Rosch, G. Gao, G. Ridout, Y. D. Wang and E. I. Tuomanen, *Mol. Microbiol.*, 2009, **72**, 12–25.
- 30 H. Sun, G. Xu, H. Zhan, H. Chen, Z. Sun, B. Tian and Y. Hua, *BMC Microbiol.*, 2010, **10**, 319.
- 31 L. Li and J. Kaplan, *J. Biol. Chem.*, 1997, **272**, 28485–28493.
- 32 Y. Chao and D. Fu, *J. Biol. Chem.*, 2004, **279**, 17173–17180.
- 33 Y. Wei and D. Fu, *J. Biol. Chem.*, 2005, **280**, 33716–33724.
- 34 H. B. Jiang, W. J. Lou, H. Y. Du, N. M. Price and B. S. Qiu, *Plant Cell Physiol.*, 2012, **53**, 1404–1417.
- 35 S. Spada, J. T. Pembroke and J. G. Wall, *Extremophiles*, 2002, **6**, 301–308.
- 36 E. Hoch, W. Lin, J. Chai, M. Hershfinkel, D. Fu and I. Sekler, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 7202–7207.
- 37 M. Gonzalez-Guerrero, D. Raimunda, X. Cheng and J. M. Arguello, *Mol. Microbiol.*, 2010, **78**, 1246–1258.
- 38 D. Podar, J. Scherer, Z. Noordally, P. Herzyk, D. Nies and D. Sanders, *J. Biol. Chem.*, 2012, **287**, 3185–3196.
- 39 Y. Shingu, T. Kudo, S. Ohsato, M. Kimura, Y. Ono, I. Yamaguchi and H. Hamamoto, *Biochem. Biophys. Res. Commun.*, 2005, **331**, 675–680.
- 40 M. Kawachi, Y. Kobae, T. Mimura and M. Maeshima, *J. Biol. Chem.*, 2008, **283**, 8374–8383.
- 41 D. Blaudez, A. Kohler, F. Martin, D. Sanders and M. Chalot, *Plant Cell*, 2003, **15**, 2911–2928.
- 42 S. R. Eddy, *Genome Inf.*, 2009, **23**, 205–211.
- 43 R. D. Finn, J. Tate, J. Mistry, P. C. Coghill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman, *Nucleic Acids Res.*, 2008, **36**, D281–D288.
- 44 J. D. Thompson, B. Linard, O. Lecompte and O. Poch, *PLoS One*, 2010, **6**, e18093.
- 45 F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins, *Mol. Syst. Biol.*, 2011, **7**, 539.
- 46 K. Katoh and H. Toh, *Briefings Bioinf.*, 2008, **9**, 286–298.
- 47 Y. Liu, B. Schmidt and D. L. Maskell, *Bioinformatics*, 2010, **26**, 1958–1964.
- 48 R. C. Edgar, *Nucleic Acids Res.*, 2004, **32**, 1792–1797.
- 49 D. Darriba, G. L. Taboada, R. Doallo and D. Posada, *Bioinformatics*, 2011, **27**, 1164–1165.

- 50 W. Li and A. Godzik, *Bioinformatics*, 2006, **22**, 1658–1659.
- 51 S. Q. Le and O. Gascuel, *Mol. Biol. Evol.*, 2008, **25**, 1307–1320.
- 52 M. Anisimova and O. Gascuel, *Syst. Biol.*, 2006, **55**, 539–552.
- 53 S. Encarnacion, M. Dunn, K. Willms and J. Mora, *J. Bacteriol.*, 1995, **177**, 3058–3066.
- 54 J. M. Martinez-Salazar and D. Romero, *Gene*, 2000, **243**, 125–131.
- 55 M. Flores, V. Gonzalez, S. Brom, E. Martinez, D. Pinero, D. Romero, G. Davila and R. Palacios, *J. Bacteriol.*, 1987, **169**, 5782–5788.
- 56 E. Salazar, J. J. Diaz-Mejia, G. Moreno-Hagelsieb, G. Martinez-Batallar, Y. Mora, J. Mora and S. Encarnacion, *Appl. Environ. Microbiol.*, 2010, **76**, 4510–4520.

CAPITULO 5. The *emnA* gene from *Rhizobium etli* encodes for a bacterial cation diffusion facilitator protein involved in Mn²⁺ resistance (in preparation)

Introduction

Rhizobial Mn²⁺ homeostasis comprises an aggregate of uptake, efflux, traffic and metalloregulatory proteins. In *Sinorhizobium meliloti*, *Rhizobium leguminosarum* and *Bradyrhizobium japonicum* the Mn²⁺ uptake systems have been shown to be dependent of *sitABC* (75-76) and *mntH* (77), respectively; whose expression is regulated by Mur (76, 78) and Fur (77) under low Mn²⁺ concentrations. For *Rhizobium leguminosarum*, also are genome-encoded uncharacterized *Escherichia coli* MntH orthologs (76).

In groups other than rhizobia the information about Mn efflux mechanism is still limited. The other reported non-CDF dependent Mn²⁺ efflux systems correspond to *Neisseria meningitidis* MntX (79), *Xylella fastidiosa* YebN (80) and *Escherichia coli* MntP efflux protein (18) being all homologues to each other and belonging to the same family (PF02659). Searches for MntX/YebN/MntP homologues encoded in the *R. etli* genome were unsuccessful, supporting the hypothesis that other transporters could be involved in Mn²⁺ tolerance.

Recently, we reported the functional classification of the *R. etli* CFN42 CDF transporters using phylogenomic inference (81). According with this study, in the genome of this organism are encoded four CDF efflux proteins, two of them, NepA (RHE_PE00218) and CepA (RHE_CH01219) independently confer Ni²⁺ and Co²⁺ resistance, respectively. The third putative CDF transporter, RHE_PD00193, was classified in the group XVII of unknown function. We also described the existence of monophyletic bacterial Mn-CDF clades; in one of them, clade VI, where was located the fourth putative CDF protein RHE_CH03072 and its rhizobial orthologs. The highly-diverse metal specificity clade VI also contains the characterized Zn²⁺/Cd²⁺ TtCzrB

from *Thermus Thermophilus*; the Mn²⁺ transporter Dr1236 from *Deinococcus radiodurans* and the Fe²⁺ transporter SII1263 from *Synechocystis* sp. It has been previously shown that Zn²⁺/ Cd²⁺ resistance relies on *SMc04128*-encoded P_{IB}-type ATPase in *S. meliloti* (62) whose ortholog in *R. etli* is *RHE_CH03719*. Since Mn²⁺ or Fe²⁺ resistance has not been explored in any rhizobia species, we decided to explore if the RHE_CH03072 protein was involved in tolerance to these metals in *R. etli*.

Here we report that RHE_CH03072 as an essential gene required for Mn²⁺ resistance in *R. etli*. The RHE_CH03072 gene is a distant MntE homolog (*Streptococcus pneumoniae*, clade I), but a close DR1236 homolog (*Deinococcus radiodurans*) (81), the only known bacterial Mn-CDFs proteins. The *RHE_CH03072* gene it is metal inducible and its homologues occur in multiple rhizobial genomes. We propose *emnA* (efflux of Mn²⁺) as new denomination for *RHE_CH03072*.

RESULTS

The RHE_CH03072 (EmnA) protein shares identical metal binding sites as Fe²⁺, Mn²⁺, Zn²⁺/Cd²⁺ CDF transporters. For clade VI proteins in general (81), and for RHE_CH03072 in particular, substrate prediction is difficult since characterized TtCzrB (Zn²⁺/ Cd²⁺) (40, 82), DR1236 (Mn²⁺) (83) and Sll1263 (Fe²⁺) (84) proteins share very close identity values to each other: RHE_CH03072 /DR1236 (55%, 89% coverage), RHE_CH03072/Sll1263 (44%, 90% coverage), and RHE_CH03072 /TtCzrB (51%, 95% coverage), all they have different substrates and they belong to the same phylogenetic clade.

Accordingly to YiiP and TtCzrB crystal structures (40-41), metals (Zn²⁺ or Cd²⁺) are bound in three different metal binding sites (MBS) denominated A, B or C, where site A is the active one, B is involved in dimer formation and site C is implicated in turnover rate. The YiiP site A is involved in metal translocation, metal discrimination and has a residue composition of DD-HD (Fig 1); so in order to explain the observed differences in the recognized metal substrates we expected that clade VI proteins would have pronounced differences to each other and particularly in their putative MBSs. A multiple sequence alignment of these proteins did not reveal differences at MBS A, B or C among TtCzrB, Sll1236, Dr1263 and RHE_CH03072 proteins when are compared to each other (Fig. 11). This alignment contrast with that previously reported (83) where the DXXXD commonly found in plant Mn-CDF proteins were assumed to be present in their bacterial counterparts. In addition, this alignment gives some insights about the composition of putative MBS A for all clade VI proteins under analysis, which is ENHD, differing from the DDHD composition of YiiP-like (clade V) proteins (81).

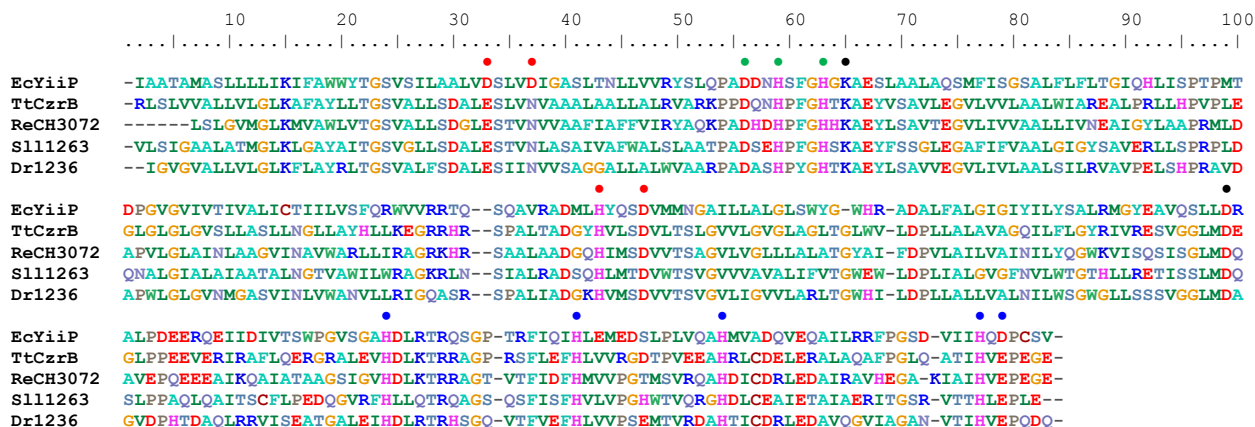


Fig. 11. The alignment of characterized clade VI CDF proteins shows they have identical MBS A, B and C in spite of having different metal substrates. The red, blue, green and black dots indicate the MBS A, B, C and the interlocked Lys⁷⁷-Asp²⁰⁷ salt bridge, respectively as reported for *E. coli* YiiP (PDB 3H90) and *T. termophilus* CzrB (PDB 3BYR) Zn-bound protein structures. This representative alignment compares the conserved HMM CDF-domains of proteins under analysis as we published (81).

Phylogenetic analysis of EmnA and its homologues reveals that is part of its own subfamily and uncovers two additional subgroups of uncharacterized CDF proteins. Recently we determined that clade VI, where EmnA belongs, has the most variable substrates in the CDF family (81), where the 3 characterized proteins recognized different metals. The absence of variation in MBS ABC and the observed closed BLASTP values among clade VI proteins, led us to investigate protein divergence by using a phylogenetic approach. For this, we reconstruct a phylogeny of 197 CDF homologues (see methods) in order to establish whether EmnA was close to any given characterized protein or if it formed part of its own subfamily of uncharacterized proteins. As a result, this analysis allowed to uncover that EmnA was part of its own subfamily (p -value=0.99) (Fig. 12) and hence its substrate may not predicted based on the sequence identity

values when compared to the characterized proteins. Interestingly, our phylogeny showed that DR1236 and EmnA groups are monophyletic to each other (p -value=0.99) suggesting that both subgroups of proteins could share Mn^{2+} as substrate. In addition, it was observed that the other three previously known CDFs are divided into three subfamilies, one protein by subfamily. Two additional uncharacterized subgroups were also identified but their substrates are currently unknown (Fig. 12).

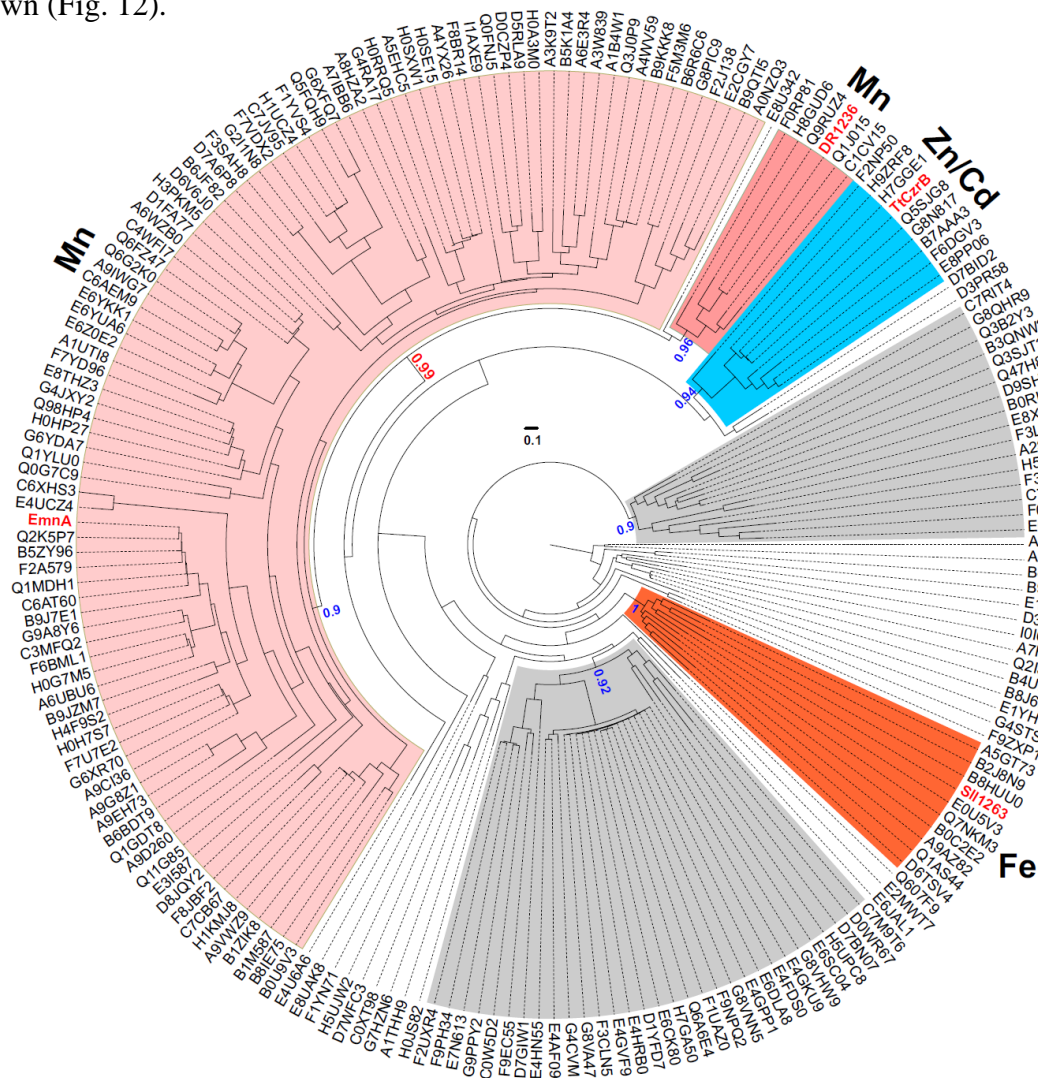


Fig. 12. Phylogenetic tree of CDF proteins from group VI reveals rhizobial EmnA is part of its own subgroup. Maximum-likelihood based tree of group VI proteins reveals the existence of multiple subfamilies accordingly to substrate. Shimodira-Hasegawa-like p -values for bipartitions indicating substrate-defined clades (p -values ≥ 0.9) are shown in blue. The p -value in red indicates

the monophyletic origin of EmnA and DR1236. The scale bar indicates the expected number of amino acid substitutions per site under the LG model.

The *emnA* gene is responsive to divalent metals. To determine metal-responsiveness of *emnA* gene in *R. etli* cells we measured the expression values of Zn²⁺, Cd²⁺, Fe²⁺ and Mn²⁺-exposed cells in MM by using qRT-PCR. Under our tested conditions, *emnA* gene was responsive to all tested metals, having Zn and Cd the strongest and weakest responses, respectively (Fig. 13).

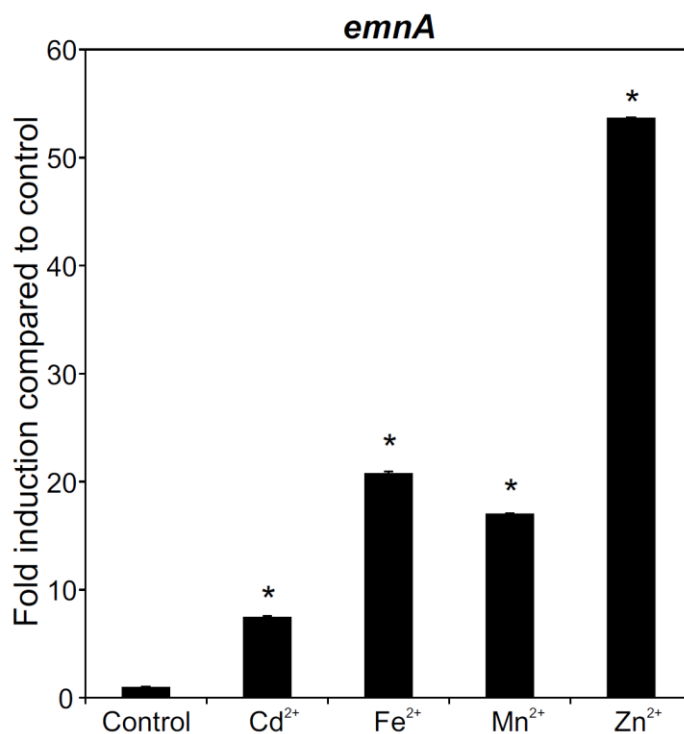


Fig. 13. The *R. etli emnA* gene is metal responsive to Fe²⁺, Zn²⁺, Cd²⁺, and Mn²⁺. The data represent the fold induction, defined as the ratio between the mRNA levels of *emnA* gene in the presence of 5 mM of FeCl₃, ZnCl₂, CdCl₂ and MnCl₂ and mRNA levels in the absence of metals (control). Both were normalized to *hisCd* mRNA levels as published (81) (mean ± s.d., n = 3) (*t*-test, **P≤0.01).

EmnA is essential for Mn²⁺ resistance in *R. etli*. To directly validate what metal(s) were the actual substrates of RHE_CH03072, we obtained a mutant in this gene by vector-integration mutagenesis and tested its growing abilities in minimal medium (MM) plates containing metals. Our results indicate this mutant is only sensitive to Mn²⁺ but it was as resistant as the wild type in the presence of other metal ions such as Zn²⁺, Cd²⁺, Cu²⁺, Ni²⁺ or Co²⁺. In addition, this mutant showed a one-log unit inhibition in the presence of Fe²⁺ (Fig. 4). The fact that this gene is a critical component of Mn²⁺ tolerance in *R. etli* prompted us to rename it as *emnA* (efflux of Mn protein A).

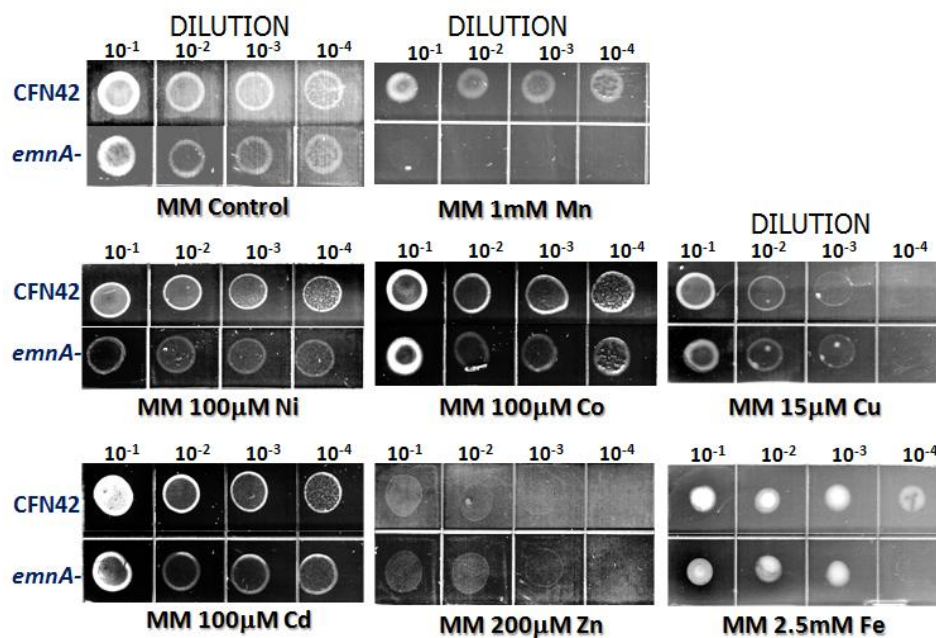


Fig. 14. *R. etli emnA* gene is essential for Mn²⁺ resistance. Cation sensitivity plate assay showing the growth of wild type CFN42, *emnA*-, and its derivative complemented with the wild type gene in the presence or absence (control) of metals, as indicated. Overnight cultures (OD₆₂₀=0.7) of each strain were serially diluted (10⁻¹–10⁻⁴ left to right) and 20 μl spotted onto MM plates with or without the indicated metals.

DISCUSSION

Our recent work describing Ni²⁺ and Co²⁺-CDF transporters also uncovered the presence of Mn-CDF clades in bacteria (81). Here, we analyzed one bacterial Mn-CDF clade and show that Mn, Zn/Cd and Fe-CDF proteins from the same group share identical MBS A, B or C suggesting alternative metal binding residues to discriminate between ions, since all of them are identical. Simultaneously, clade VI proteins differ in their MBS compositions when compared to other bacterial Mn-, Zn/Cd-CDFs, such as MntE or YiiP, in spite they share Mn²⁺, Zn²⁺ and Cd²⁺ as substrates. Since our phylogeny uncovered new subgroups in group VI proteins we cannot discard the recognition of other substrates by them. This phenomenon is similar to that previously observed by us for Ni²⁺- and Zn²⁺-CDFs which also shared the same MBS A but still have different substrates.

Based on the observed group-metal substrates relationships in this and previous phylogenies (1, 49, 81, 85), we can figure that metal selectivity variations in CDF proteins occur at 3 different levels. 1) Metal specificity among clades. The evolutionary divergence among CDF clades is the most evident feature observed in all reported CDF phylogenies. In this case, variations are seen along over all the protein sequences. In this level we found the MntE vs EmnA protein groups CepA, NepA, ScMMT1/2 or ZnT clades, whose identity values to each other are around 30-35%. 2) Metal specificity variations among CDF proteins from the same clade. In this class we found two different examples, a) monophyletic clades such as clade VI studied here, and proteins from the clade X which recognize Co²⁺ and Zn²⁺ as COT1 and ZRC1 (81) in spite of having identical MBS A, B and C all they have different metal specificities, in these cases the molecular basis of selectivity are completely unknown; b) the clade VII proteins (81) HvMTP1 and AtMTP1 which share identical MBS A but discriminate between Zn²⁺ and Co²⁺ by the use of short non-conserved motifs like His-rich tracts or the presence of the VTVTT motif in AtMTP1 whose removal allows

Co²⁺ recognition (86). Proteins of above examples have around >50% identity 3) Multispecificity or promiscuity. This category encompasses all those proteins having 2 or more substrates. There are several examples of this: CmDmeF, CmFieF, BsCzcD, CzcD, OsMTP1; however, nothing is known about how metal selectivity for chemically different metals (i.e. Fe²⁺ and Zn²⁺) is achieved in the same protein. Phylogenetic and multiple sequence alignment analyses coupled to biochemical and genetic approaches of every kind of variations in this family of proteins could help in determining the molecular basis of metal selectivity in CDF proteins.

Rhizobial Mn²⁺ homeostasis has been focused mainly on metal uptake, several systems are currently known as responsible for transport into *B. japonicum* (77), *S. meliloti* (75) and *R. leguminosarum* (76) cells; however, the opposite process was unknown. The characterization of EmnA offer insights about the Mn²⁺ efflux mechanism in this bacterial group. It is particularly interesting that EmnA orthologs are present in almost all analyzed rhizobial genomes by exception of *B. japonicum*.

The EmnA subgroup contained 50% of all proteins under our analysis and it is monophyletic relative to DR1236 which suggest that Mn²⁺ transport is probably the most predominant metal substrate for clade VI proteins. Unfortunately, there is not published information about TtCzrB specificities beyond Zn²⁺ or Cd²⁺, since other metals were not tested. The TtCzrB orthologs are restricted to *Thermus* species while EmnA/DR1236 ones are widespread in at least 3 different genus.

CAPITULO 6. Hipertolerancia a Cu²⁺ de la cepa *Rhizobium etli* CFNX185

Como parte de la identificación de los mecanismos de homeostasis de metales en *R. etli* CFN42, se analizó el crecimiento de la colección de mutantes de *R. etli* curadas o con deleciones de plásmidos previamente reportadas (87). Descubrimos que la mutante CFNX185, cuyo plásmido p42e carece de ~200 kb, mostraba un aumento en la resistencia a Cu²⁺ (Fig. 9). Ya que mutantes con características de hipertolerancia a Cu²⁺ no habían sido descritas desde hace más de 30 años (88) y que este fenotipo no había sido reportado en rizobios, decidimos investigar qué gene(s), al estar ausentes en la mutante CFNX185, provocan un aumento en la tolerancia a cobre.

La búsqueda de dicho gene(s) se vio facilitado por 2 razones: 1) El aislamiento de un cósmido, proveniente de una librería genómica de *R. etli* CFN42, conteniendo aproximadamente 20 Kb del plásmido p42e, que complementó la resistencia a Ni²⁺ (por contener al gene *nepA*) pero no disminuyó la resistencia a Cu²⁺ de la cepa CFNX185 al nivel de la cepa silvestre CFN42, indicativo de que los fenotipos de Ni²⁺ y Cu²⁺ eran independientes entre sí (Fig. 15A, *pe*Δ/*pL313*). 2) Mutantes con deleciones del plásmido p42e de menor tamaño que permitieron localizar una región con menor cantidad de genes que nos aproximara a encontrar el gene responsable de la observada hipertolerancia a Cu²⁺. Los fenotipos de estas dos deleciones: *pe*Δ11 y *pe*Δ10 (89) en 20 μM Cu²⁺, mostró que la cepa *pe*Δ10 tiene un fenotipo comparable al de la CFNX185 (Fig 15B).

Ya que la cepa *pe*Δ10 aun carece de ~100 genes, el análisis individual de cada uno de ellos era complejo, por lo que a partir de esta mutante se construyeron un conjunto de deleciones (Fig. 15A) cuyo propósito final era reducir a una región mínima (de 25 o menos genes) (Fig. 15C) el fenotipo de hipertolerancia a Cu²⁺.

La localización, identificación y caracterización del (los) gene (s) responsables de tal hiperresistencia se están investigando actualmente en nuestro laboratorio.

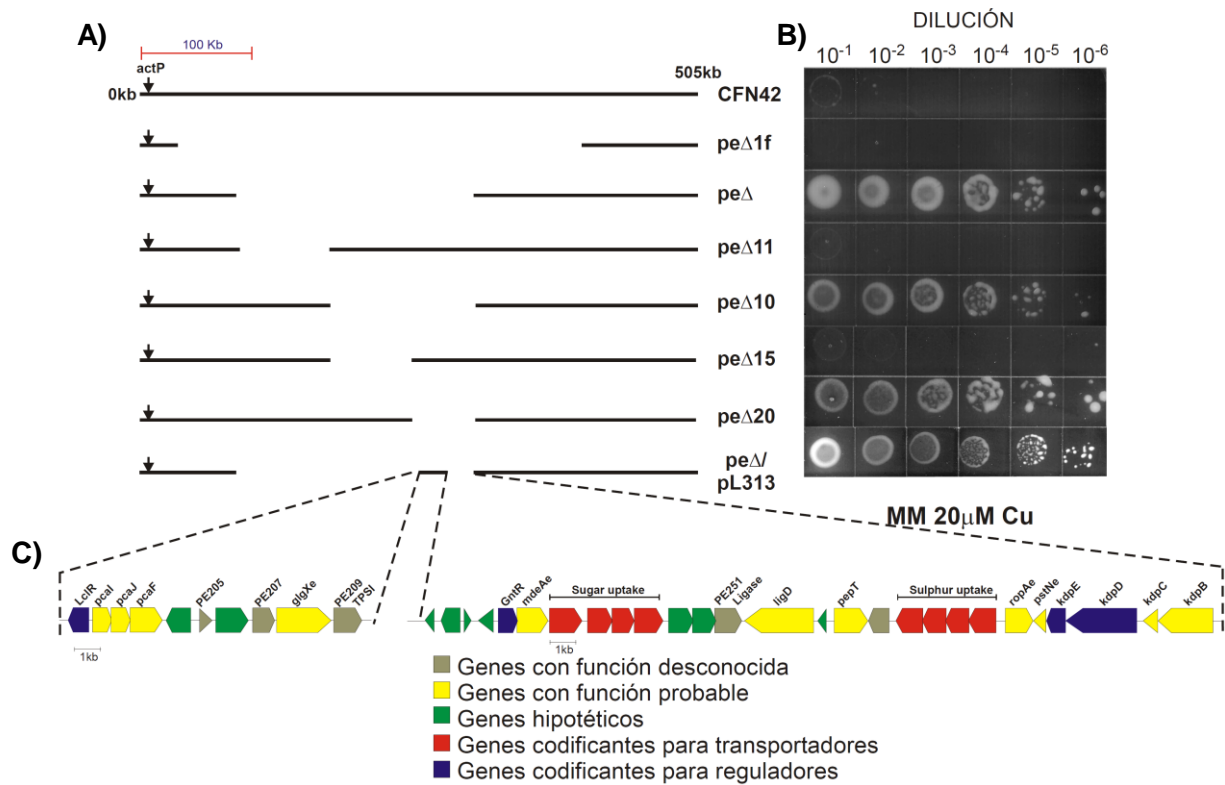


Fig. 15. Efecto de las deleciones del plásmido p42e en la tolerancia a cobre de *R. etli* CFN42. A) Representación lineal de las 505kb del plásmido p42e (tal como está en la cepa CFN42) y sus deleciones representadas como espacios en blanco. B) Efecto de la deleción sobre el crecimiento en presencia de cobre. Los controles de MM sin metal crecen en diluciones 10⁻⁶ (no mostrado). C) Región mínima responsable de la hipertolerancia a Cu²⁺, inferida a partir del análisis de deleciones y de los genes contenidos en el cósmido pL313 aislado de la librería genómica de *R. etli* CFN42 (conteniendo ~25kb de p42e, que complementan la resistencia a Ni²⁺ pero no afecta la resistencia a Cu²⁺).

6.1. Identificación de una mutante Tn5 afectada en la homeostasis de Cu²⁺

Paralelamente a la estrategia de deleciones en p42e, se efectuó una mutagénesis con el transposón Tn5 en el fondo de la cepa CFN42, cuyo propósito era obtener mutantes hipertolerantes a Cu²⁺. Se utilizó medio mínimo 20μM de Cu²⁺ como medio de selección positiva, concentración totalmente inhibitoria del crecimiento de la cepa silvestre CFN42.

Se obtuvo una mutante que denominamos cur201, cuya DO₆₂₀ después de 16hrs (tiempo requerido para que la cepa CFN42 alcance su tope máximo de crecimiento exponencial) es ~10 veces menor al de la cepa silvestre (Fig. 16A) y muestra una alta resistencia a Cu²⁺ semejante al de la mutante CFNX185 (Fig. 16B).

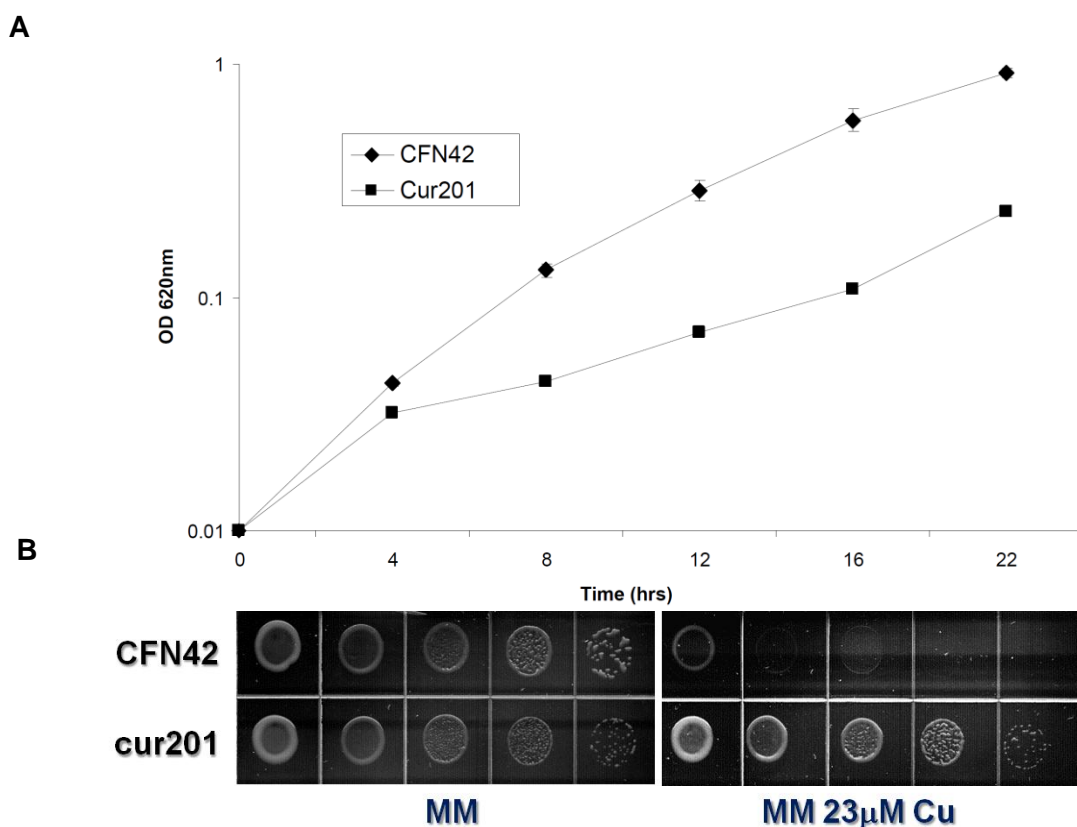


Fig. 16. Comparación de fenotipos de la mutante cur201 con la cepa parental CFN42. (A) Crecimiento de la cepa mutante *R. etli* cur201 comparado con el de la cepa silvestre CFN42 en medio rico PY (promedio de 3 replicas). (B) Crecimiento de la mutante cur201 en presencia de Cu. Se muestran las diluciones 10⁻¹ – 10⁻⁵ (de izquierda a derecha).

Para identificar el (los) genes afectados por la mutación en la cepa *cur201* se siguieron dos estrategias: 1) clonar y secuenciar un fragmento EcoRI que contuviese una sección del genoma, el Tn5 completo y que fuese susceptible a ser clonado y secuenciado y 2) complementar esta mutante con el banco genómico de *R. etli*, en cuyo caso se secuenciarían fragmentos derivados de los cósmidos identificados obtenidos por clonación en los vectores adecuados para tal fin.

La estrategia de complementación fue la primera en dar resultados. Debido a que el crecimiento deficiente de la mutante *cur201* en medio rico y la hiperresistencia a Cu^{2+} parecían estar asociados, buscamos clonas complementadas a partir del banco genómico que mostraran un crecimiento más rápido en PY. De aquellas que lo lograron (4 clonas), establecimos si contenían un cósmido extra y si recuperaban el fenotipo de la cepa CFN42 en Cu^{2+} . Con esta estrategia determinamos que las 4 clonas que denominamos *cur201/pL1-pL4*, mostraban crecimiento silvestre en PY (no mostrado), crecimiento silvestre en MM con $20\mu\text{M Cu}^{2+}$ (Fig. 17B) así como la presencia del cósmido extra observable mediante geles tipo eckhardt (Fig. 17C).

Los cósmidos pL1-4 se movilizaron hacia la cepa *E. coli* DH5 α desde cada una de ellas, se purificaron y se estableció el patrón de restricción con EcoRI. Dicho patrón reveló que cada uno de los cuatro cósmidos compartían exactamente las mismas bandas EcoRI entre sí (Fig. 17D), por lo que asumimos que se trataba del mismo cósmido en los 4 casos.

Para identificar la región involucrada con la complementación decidimos subclonar algunas bandas PstI hacia pBC-SK para secuenciar un fragmento del DNA del cósmido que nos permitiera posicionarnos en el genoma. La secuenciación de 4 fragmentos entre 0.2-2kb, el análisis de restricción de los cósmidos, mas la hibridación del cósmido pL1 en un gel tipo Eckhardt nos permitió identificar la región que comprende los primeros 25 genes del plásmido p42f (Fig. 17A).

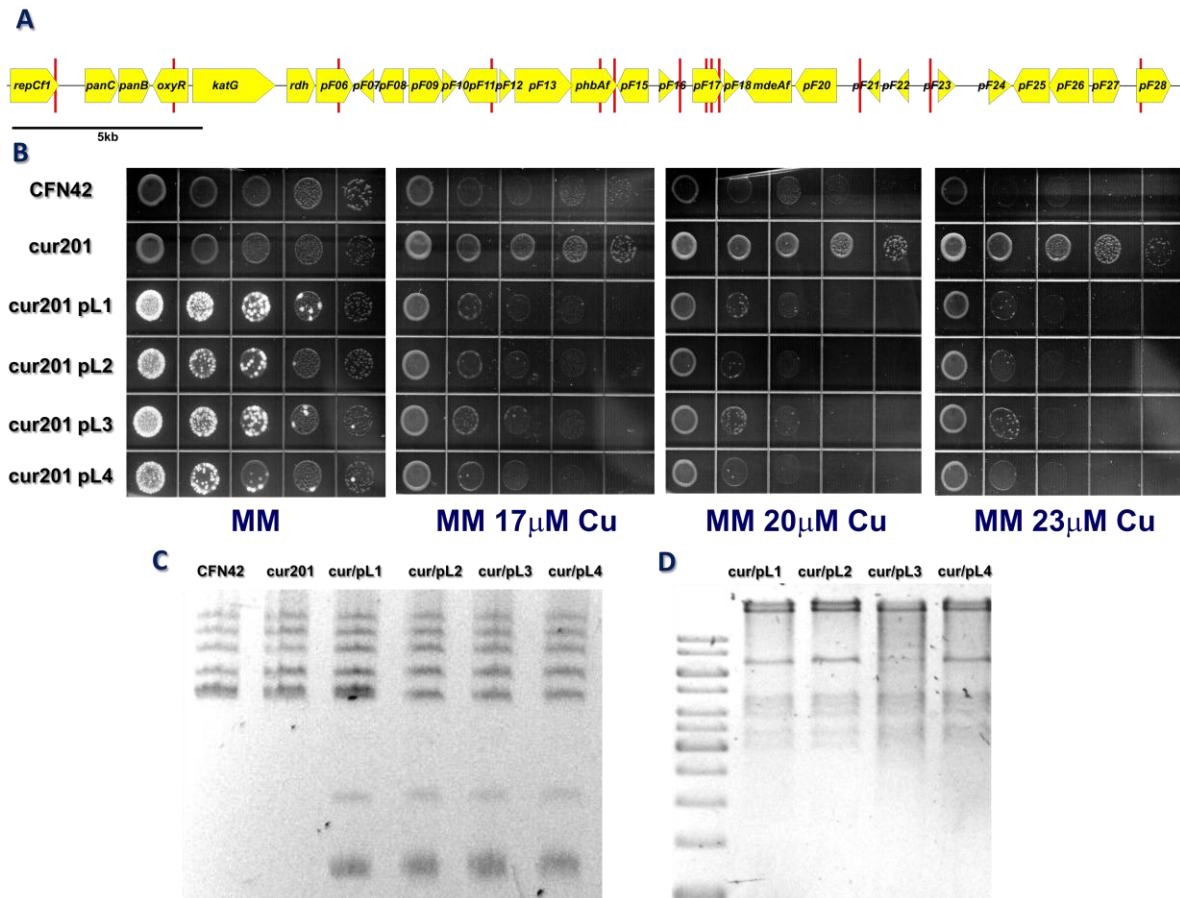


Fig. 17 Análisis de complementación de la mutante *cur201*. A) Genes contenidos en las primeras 25kb del plásmido p42f, indicando los sitios de restricción *EcoRI* (líneas rojo). B) Crecimiento de las cepas silvestre, mutante *cur201* y mutantes complementadas con los cósmidos pL1-pL4 en diferentes concentraciones de Cu^{2+} . Se muestran las diluciones 10^{-1} - 10^{-5} . C) Geles tipo Eckhardt indicando la presencia de los cósmidos pL1-pL4 en las cepas usadas en B. D) Análisis de restricción de los cósmidos pL1-pL4 obtenidos de cada una de las mutantes *cur201* complementadas (nótese el patrón idéntico de restricción)

El análisis de restricción *EcoRI* del cósmido pL1 (Fig. 17D) mostró que era idéntico al pCos24 previamente reportado (90). Entre los genes contenidos en este cósmido se encontraban *panC* y *panB* involucrados con la síntesis de pantotenato (Fig. 17A). Mutantes en los genes *panC*- y *panB*- no crecen en MM debido a su incapacidad para sintetizar pantotenato, una vitamina esencial, pero capaces de crecer en medio rico (91).

El crecimiento deficiente de la mutante *cur201* como el de las mutantes *panCB* (91) y la complementación de ambas con cósmidos idénticos, sugiere una relación entre todas ellas. La

diferencia más notable entre la mutante *cur201* y las mutantes *panCB* es que éstas crecían a niveles silvestres en medio rico, mientras que la mutante *cur201* no lo hace [(91), este trabajo].

Respecto a la estrategia experimental a seguir con el análisis de esta segunda mutante, se planea movilizar bandas *EcoRI* derivadas de *pL1* para determinar cuál de ellas complementa los niveles de resistencia silvestre a Cu^{2+} .

El Cu^{2+} libre intracelular afecta el ensamble de proteínas que poseen grupos Fe-S, como resultado, el Cu reemplaza al Fe y el Fe libre promueve la generación de especies reactivas de oxígeno por reacción de fenton, las cuales dañan membranas, DNA y proteínas (92-93). Ya que entre los genes presentes en *pL1* se encuentra el gene *oxyR*, un sensor de estrés oxidante que modula la expresión génica de un alto número de genes, su ausencia en la mutante *cur201* podría disparar la expresión de proteínas de expulsión a Cu^{2+} , lo que explicaría la hipertolerancia a Cu^{2+} observada en esta mutante.

Los resultados obtenidos del cósmido *pL1*, la complementación de la mutante *cur201* con él y su relación con la homeostasis de Cu^{2+} indican que al menos hay dos loci involucrados con la hiperresistencia a Cu^{2+} , uno en el plásmido *p42e* y otro en el *p42f*. Actualmente estamos investigando la relación de *oxyR* y la hipertolerancia a Cu^{2+} en nuestro laboratorio.

CAPITULO 7. CONCLUSIONES

Como resultado de las investigaciones desarrolladas a lo largo de esta tesis doctoral, he llegado a las siguientes conclusiones:

El análisis filogenómico de las proteínas CDF provee de una estrategia para identificar y predecir funciones no solo al interior de esta familia de transportadores, también en la identificación de residuos clado-específicos potencialmente involucrados con la selectividad de sustrato.

La resistencia a metales en *R. etli* involucra múltiples transportadores tipo CDF con sustratos asociados a clados no previamente caracterizados e independientes, como es el caso de NepA, CepA y EmnA para Ni²⁺, Co²⁺ y Mn²⁺, respectivamente.

La hipertolerancia a Cu²⁺ dependiente de pérdida de genes en *R. etli* es un fenotipo interesante para identificar funciones relacionadas con la homeóstasis de cobre, pero independientes de proteínas de expulsión.

En suma, los resultados de esta tesis contribuyen a entender los mecanismos de resistencia a Ni²⁺, Co²⁺, Mn²⁺ y Cu²⁺ en *Rhizobium etli* cepa CFN42 y la contribución de las proteínas CDF en este proceso. Esto representa un avance en el área de la homeostasis de metales en *Rhizobiaceae*, ya que no había estudios de CDF en ningún otro rhizobio.

CAPITULO 8. HOMEOSTASIS DE METALES DE TRANSICION EN *Rhizobium etli*: PERSPECTIVAS.

La búsqueda de mecanismos de resistencia en *Rhizobium etli*, nos permitió identificar a miembros de la familia de transportadores CDF como responsables de la resistencia a 3 metales distintos: Ni²⁺, Co²⁺ y Mn²⁺. El enfoque filogenómico utilizado ha permitido establecer relaciones funcionales entre grupos de proteínas ortólogas y ha facilitado la identificación de residuos grupo-específicos involucrados con la función proteica.

En este sentido, es claro que aún quedan por describir los mecanismos moleculares de selectividad que expliquen por qué unas proteínas reconocen Ni²⁺ y Co²⁺ pero no Zn²⁺ y viceversa. Dichos mecanismos, actualmente desconocidos, podrían implicar el uso de aminoácidos que no solo sean capaces de unir metales, como cisteínas, histidinas o ácidos aspártico y glutámico, sino que sean conservados entre proteínas reconociendo los mismos metales. Cuáles y cómo estos residuos afectan las cinéticas de transporte de una proteína CDF dada, es un tema emergente de investigación.

Por otro lado, ya que la homeostasis de un metal tiene influencia sobre otros en un mismo citoplasma, aún es necesario responder el cómo la expresión de los genes codificantes para CDF es regulada y cuál es su influencia en la movilización intracelular de iones que no son blancos. La identificación de metalreguladores, sus regiones de unión a DNA y sus ligandos son asuntos pendientes en el estudio de la homeostasis de metales en *R. etli*.

En *R. etli* el mecanismo de resistencia a Cu²⁺ fue previamente reportado en nuestro laboratorio (89) y mostró ser dependiente de ActP, una ATPasa Tipo P_{IB}; sin embargo, la mutante obtenida en el gene codificante es sensible a Cu. Nuestras observaciones con respecto a la hipertolerancia a Cu²⁺, no involucran directamente a ActP, por lo que es posible que los fenotipos observados en las mutantes CFNX185 y cur201 nos den indicios sobre nuevas rutas de tráfico intracelular de Cu²⁺ en *R. etli* CFN42.

REFERENCIAS

1. Montanini B, Blaudez D, Jeandroz S, Sanders D, & Chalot M (2007) Phylogenetic and functional analysis of the Cation Diffusion Facilitator (CDF) family: improved signature and prediction of substrate specificity. *BMC Genomics* 8:107.
2. Nies DH (1999) Microbial heavy-metal resistance. *Appl Microbiol Biotechnol* 51(6):730-750.
3. He ZLL, Yang XE, & Stoffella PJ (2005) Trace elements in agroecosystems and impacts on the environment. *J Trace Elem Med Bio* 19(2-3):125-140.
4. Harding MM (2004) The architecture of metal coordination groups in proteins. *Acta Crystallogr D Biol Crystallogr* 60(Pt 5):849-859.
5. Dokmanic I, Sikic M, & Tomic S (2008) Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. *Acta Crystallogr D Biol Crystallogr* 64(Pt 3):257-263.
6. Anbar AD (2008) Oceans. Elements and evolution. *Science* 322(5907):1481-1483.
7. Outten CE & O'Halloran TV (2001) Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* 292(5526):2488-2492.
8. Outten FW, Huffman DL, Hale JA, & O'Halloran TV (2001) The independent *cue* and *cus* systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. *J Biol Chem* 276(33):30670-30677.
9. Jacobsen FE, Kazmierczak KM, Lisher JP, Winkler ME, & Giedroc DP (2011) Interplay between manganese and zinc homeostasis in the human pathogen *Streptococcus pneumoniae*. *Metallomics* 3(1):38-41.
10. Waldron KJ & Robinson NJ (2009) How do bacterial cells ensure that metalloproteins get the correct metal? *Nat Rev Microbiol* 7(1):25-35.
11. Harding MM, Nowicki MW, & Walkinshaw MD (2010) Metals in protein structures: a review of their principal features. *Crystallogr Rev* 16(4):247-302.
12. Dupont CL, Butcher A, Valas RE, Bourne PE, & Caetano-Anolles G (2010) History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc Natl Acad Sci U S A* 107(23):10567-10572.
13. Dupont CL, Yang S, Palenik B, & Bourne PE (2006) Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proc Natl Acad Sci U S A* 103(47):17822-17827.
14. Patzer SI & Hantke K (2001) Dual repression by Fe²⁺-Fur and Mn²⁺-MntR of the *mntH* gene, encoding an NRAMP-like Mn²⁺ transporter in *Escherichia coli*. *J Bacteriol* 183(16):4806-4813.
15. Bagg A & Neilands JB (1987) Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* 26(17):5471-5477.
16. Masse E & Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* 99(7):4620-4625.
17. Wee S, et al. (1988) Expression, isolation and properties of Fur (ferric uptake regulation) protein of *Escherichia coli* K 12. *Biol Met* 1(1):62-68.
18. Waters LS, Sandoval M, & Storz G (2011) The *Escherichia coli* MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. *J Bacteriol* 193(21):5887-5897.
19. Grass G, et al. (2005) FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. *Arch Microbiol* 183(1):9-18.
20. Korkhov VM, Mireku SA, & Locher KP (2012) Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F. *Nature* 490(7420):367-372.

21. Lundrigan MD & Kadner RJ (1989) Altered cobalamin metabolism in *Escherichia coli* *btuR* mutants affects *btuB* gene regulation. *J Bacteriol* 171(1):154-161.
22. Rodrigue A, Effantin G, & Mandrand-Berthelot MA (2005) Identification of *rcnA* (*yohM*), a nickel and cobalt resistance gene in *Escherichia coli*. *J Bacteriol* 187(8):2912-2916.
23. Bleriot C, Effantin G, Lagarde F, Mandrand-Berthelot MA, & Rodrigue A (2011) RcnB is a periplasmic protein essential for maintaining intracellular Ni and Co concentrations in *Escherichia coli*. *J Bacteriol* 193(15):3785-3793.
24. Rensing C, Fan B, Sharma R, Mitra B, & Rosen BP (2000) CopA: An *Escherichia coli* Cu(I)-translocating P-type ATPase. *Proc Natl Acad Sci U S A* 97(2):652-656.
25. Franke S, Grass G, Rensing C, & Nies DH (2003) Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. *J Bacteriol* 185(13):3804-3812.
26. Mealman TD, Blackburn NJ, & McEvoy MM (2012) Metal export by CusCFBA, the periplasmic Cu(I)/Ag(I) transport system of *Escherichia coli*. *Curr Top Membr* 69:163-196.
27. Iwig JS, Rowe JL, & Chivers PT (2006) Nickel homeostasis in *Escherichia coli* - the *rcnR-rcnA* efflux pathway and its linkage to NikR function. *Mol Microbiol* 62(1):252-262.
28. Iwig JS, Leitch S, Herbst RW, Maroney MJ, & Chivers PT (2008) Ni(II) and Co(II) sensing by *Escherichia coli* RcnR. *J Am Chem Soc* 130(24):7592-7606.
29. Outten FW, Outten CE, Hale J, & O'Halloran TV (2000) Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, *cueR*. *J Biol Chem* 275(40):31024-31029.
30. Munson GP, Lam DL, Outten FW, & O'Halloran TV (2000) Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *J Bacteriol* 182(20):5864-5871.
31. Navarro C, Wu LF, & Mandrand-Berthelot MA (1993) The *nik* operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. *Mol Microbiol* 9(6):1181-1191.
32. Patzer SI & Hantke K (1998) The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. *Mol Microbiol* 28(6):1199-1210.
33. Grass G, Wong MD, Rosen BP, Smith RL, & Rensing C (2002) ZupT is a Zn(II) uptake system in *Escherichia coli*. *J Bacteriol* 184(3):864-866.
34. De Pina K, Desjardin V, Mandrand-Berthelot MA, Giordano G, & Wu LF (1999) Isolation and characterization of the *nikR* gene encoding a nickel-responsive regulator in *Escherichia coli*. *J Bacteriol* 181(2):670-674.
35. Bloom SL & Zamble DB (2004) Metal-selective DNA-binding response of *Escherichia coli* NikR. *Biochemistry* 43(31):10029-10038.
36. Patzer SI & Hantke K (2000) The zinc-responsive regulator Zur and its control of the *znu* gene cluster encoding the ZnuABC zinc uptake system in *Escherichia coli*. *J Biol Chem* 275(32):24321-24332.
37. Rensing C, Mitra B, & Rosen BP (1997) The *zntA* gene of *Escherichia coli* encodes a Zn(II)-translocating P-type ATPase. *Proc Natl Acad Sci U S A* 94(26):14326-14331.
38. Grass G, *et al.* (2001) ZitB (YbgR), a member of the cation diffusion facilitator family, is an additional zinc transporter in *Escherichia coli*. *J Bacteriol* 183(15):4664-4667.
39. Outten CE, Outten FW, & O'Halloran TV (1999) DNA distortion mechanism for transcriptional activation by ZntR, a Zn(II)-responsive MerR homologue in *Escherichia coli*. *J Biol Chem* 274(53):37517-37524.
40. Cherezov V, *et al.* (2008) Insights into the mode of action of a putative zinc transporter CzcB in *Thermus thermophilus*. *Structure* 16(9):1378-1388.
41. Lu M, Chai J, & Fu D (2009) Structural basis for autoregulation of the zinc transporter YiiP. *Nat Struct Mol Biol* 16(10):1063-1067.
42. Gourdon P, *et al.* (2011) Crystal structure of a copper-transporting P_{1B}-type ATPase. *Nature* 475(7354):59-64.

43. Long F, *et al.* (2010) Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature* 467(7314):484-488.
44. Guffanti AA, Wei Y, Rood SV, & Krulwich TA (2002) An antiport mechanism for a member of the cation diffusion facilitator family: divalent cations efflux in exchange for K⁺ and H⁺. *Mol Microbiol* 45(1):145-153.
45. Ohana E, *et al.* (2009) Identification of the Zn²⁺ binding site and mode of operation of a mammalian Zn²⁺ transporter. *J Biol Chem* 284(26):17677-17686.
46. Brown D & Sjolander K (2006) Functional classification using phylogenomic inference. *PLoS Comput Biol* 2(6):e77.
47. Eisen JA (1998) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* 8(3):163-167.
48. Sjolander K (2004) Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* 20(2):170-179.
49. Gustin JL, Zanis MJ, & Salt DE (2011) Structure and evolution of the plant cation diffusion facilitator family of ion transporters. *BMC Evol Biol* 11:76.
50. Rosenblueth M & Martinez-Romero E (2004) *Rhizobium etli* maize populations and their competitiveness for root colonization. *Archives of Microbiology* 181(5):337-344.
51. Kondorosi E, Mergaert P, & Kereszt A (2013) A paradigm for endosymbiotic life: cell differentiation of *Rhizobium* bacteria provoked by host plant factors. *Annu Rev Microbiol* 67:611-628.
52. Martinez-Romero E (2003) Diversity of *Rhizobium-Phaseolus vulgaris* symbiosis: overview and perspectives. *Plant Soil* 252(1):11-23.
53. Gonzalez V, *et al.* (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci U S A* 103(10):3834-3839.
54. Mireles A, *et al.* (2004) Heavy metal accumulation in plants and soil irrigated with wastewater from Mexico city. *Nucl Instrum Meth B* 219:187-190.
55. Giller KE, Witter E, & McGrath SP (2009) Heavy metals and soil microbes. *Soil Biol Biochem* 41(10):2031-2037.
56. Laguerre G, *et al.* (2006) Response of rhizobial populations to moderate copper stress applied to an agricultural soil. *Microbial Ecology* 52(3):426-435.
57. Hirsch PR, Jones MJ, Mcgrath SP, & Giller KE (1993) Heavy-Metals from Past Applications of Sewage-Sludge Decrease the Genetic Diversity of *Rhizobium leguminosarum* biovar Trifolii Populations. *Soil Biol Biochem* 25(11):1485-1490.
58. Puri S, Hohle TH, & O'Brian MR (2010) Control of bacterial iron homeostasis by manganese. *P Natl Acad Sci USA* 107(23):10691-10695.
59. Todd JD, Sawers G, Rodionov DA, & Johnston AWB (2006) The *Rhizobium leguminosarum* regulator IrrA affects the transcription of a wide range of genes in response to Fe availability. *Mol Genet Genomics* 275(6):564-577.
60. Carter RA, *et al.* (2002) The *vbs* genes that direct synthesis of the siderophore vicibactin in *Rhizobium leguminosarum*: their expression in other genera requires ECF sigma factor Rpol. *Molecular Microbiology* 44(5):1153-1166.
61. Reeve WG, Tiwari RP, Kale NB, Dilworth MJ, & Glenn AR (2002) ActP controls copper homeostasis in *Rhizobium leguminosarum* bv. viciae and *Sinorhizobium meliloti* preventing low pH-induced copper toxicity. *Mol Microbiol* 43(4):981-991.
62. Rossbach S, *et al.* (2008) Response of *Sinorhizobium meliloti* to elevated concentrations of cadmium and zinc. *Appl Environ Microbiol* 74(13):4218-4221.
63. Chaintreuil C, *et al.* (2007) Nickel resistance determinants in bradyrhizobium strains from nodules of the endemic New Caledonia legume *Serianthes calycina*. *Appl Environ Microbiol* 73(24):8018-8022.
64. Nawapan S, *et al.* (2009) Functional and expression analyses of the *cop* operon, required for copper resistance in *Agrobacterium tumefaciens*. *J Bacteriol* 191(16):5159-5168.

65. Foster LJR, Moy YP, & Rogers PL (2000) Metal binding capabilities of *Rhizobium etli* and its extracellular polymeric substances. *Biotechnol Lett* 22(22):1757-1760.
66. Zhengwei Z, Fang W, Lee HY, & Yang Z (2005) Responses of *Azorhizobium caulinodans* to cadmium stress. *FEMS Microbiol Ecol* 54(3):455-461.
67. Skorupska A, Janczarek M, Marczak M, Mazur A, & Krol J (2006) Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb Cell Fact* 5.
68. Marrero J, Auling G, Coto O, & Nies DH (2007) High-level resistance to cobalt and nickel but probably no transenvelope efflux: Metal resistance in the Cuban *Serratia marcescens* strain C-1. *Microb Ecol* 53(1):123-133.
69. Grass G, et al. (2001) NreB from *Achromobacter xylosoxidans* 31A Is a nickel-induced transporter conferring nickel resistance. *J Bacteriol* 183(9):2803-2807.
70. Haritha A, et al. (2009) MrdH, a novel metal resistance determinant of *Pseudomonas putida* KT2440, is flanked by metal-inducible mobile genetic elements. *J Bacteriol* 191(19):5976-5987.
71. Tian J, et al. (2007) Nickel-resistant determinant from *Leptospirillum ferriphilum*. *Appl Environ Microbiol* 73(7):2364-2368.
72. Finn RD, et al. (2008) The Pfam protein families database. *Nucleic Acids Res.* 36:D281-288.
73. Chao Y & Fu D (2004) Thermodynamic studies of the mechanism of metal binding to the *Escherichia coli* zinc transporter YiiP. *J Biol Chem* 279(17):17173-17180.
74. Chao Y & Fu D (2004) Kinetic study of the antiport mechanism of an *Escherichia coli* zinc transporter, ZitB. *J Biol Chem* 279(13):12043-12050.
75. Davies BW & Walker GC (2007) Disruption of *sitA* compromises *Sinorhizobium meliloti* for manganese uptake required for protection against oxidative stress. *J Bacteriol* 189(5):2101-2109.
76. Diaz-Mireles E, et al. (2004) The Fur-like protein Mur of *Rhizobium leguminosarum* is a Mn²⁺-responsive transcriptional regulator. *Microbiology* 150(Pt 5):1447-1456.
77. Hohle TH & O'Brian MR (2009) The *mntH* gene encodes the major Mn²⁺ transporter in *Bradyrhizobium japonicum* and is regulated by manganese via the Fur protein. *Mol Microbiol* 72(2):399-409.
78. Chao TC, Becker A, Buhrmester J, Puhler A, & Weidner S (2004) The *Sinorhizobium meliloti* *fur* gene regulates, with dependence on Mn(II), transcription of the *sitABCD* operon, encoding a metal-type transporter. *J Bacteriol* 186(11):3609-3620.
79. Veyrier FJ, Boneca IG, Cellier MF, & Taha MK (2011) A novel metal transporter mediating manganese export (MntX) regulates the Mn to Fe intracellular ratio and *Neisseria meningitidis* virulence. *PLoS Pathog* 7(9):e1002261.
80. Li C, Tao J, Mao D, & He C (2011) A novel manganese efflux system, YebN, is required for virulence by *Xanthomonas oryzae* pv. *oryzae*. *PLoS One* 6(7):e21983.
81. Cubillas C, Vinuesa P, Tabche ML, & Garcia-de Los Santos A (2013) Phylogenomic analysis of Cation Diffusion Facilitator proteins uncovers Ni²⁺/Co²⁺ transporters. *Metallomics* 5(12):1634-1643.
82. Spada S, Pembroke JT, & Wall JG (2002) Isolation of a novel *Thermus thermophilus* metal efflux protein that improves *Escherichia coli* growth under stress conditions. *Extremophiles* 6(4):301-308.
83. Sun H, et al. (2010) Identification and evaluation of the role of the manganese efflux protein in *Deinococcus radiodurans*. *BMC Microbiol* 10:319.
84. Jiang HB, Lou WJ, Du HY, Price NM, & Qiu BS (2012) Sll1263, a unique cation diffusion facilitator protein that promotes iron uptake in the cyanobacterium *Synechocystis* sp. Strain PCC 6803. *Plant Cell Physiol* 53(8):1404-1417.
85. Nies DH (2003) Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* 27(2-3):313-339.
86. Podar D, et al. (2012) Metal selectivity determinants in a family of transition metal transporters. *J Biol Chem* 287(5):3185-3196.

87. Brom S, *et al.* (1992) Different plasmids of *Rhizobium leguminosarum* bv. phaseoli are required for optimal symbiotic performance. *J Bacteriol* 174(16):5183-5189.
88. Lutkenhaus JF (1977) Role of a major outer membrane protein in *Escherichia coli*. *J Bacteriol* 131(2):631-637.
89. Landeta C, *et al.* (2011) Plasmids with a chromosome-like role in rhizobia. *J Bacteriol* 193(6):1317-1326.
90. Garcia-de Los Santos A, *et al.* (2008) Requirement of a plasmid-encoded catalase for survival of *Rhizobium etli* CFN42 in a polyphenol-rich environment. *Appl Environ Microbiol* 74(8):2398-2403.
91. Villasenor T, *et al.* (2011) Housekeeping genes essential for pantothenate biosynthesis are plasmid-encoded in *Rhizobium etli* and *Rhizobium leguminosarum*. *BMC Microbiol* 11:66.
92. Macomber L & Imlay JA (2009) The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc Natl Acad Sci U S A* 106(20):8344-8349.
93. Macomber L, Rensing C, & Imlay JA (2007) Intracellular copper does not catalyze the formation of oxidative DNA damage in *Escherichia coli*. *J Bacteriol* 189(5):1616-1626.