

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

CARACTERIZACIÓN FUNCIONAL DE LOS ORÍGENES DE REPLICACIÓN DEL PLÁSMIDO SIMBIÓTICO DE *Rhizobium tropici* CFN299

Т]	E		S		Ι	S
QUE	PAR	A OI	BTE	ENER	EL	TÍTULO) DE:
B	Ι	Ó		L	0	G	0
Р	R	E	S	Ε	N	Т	A :

NOÉ BECERRA LOBATO

DIRECTOR DE TESIS: Dr. ISMAEL HERNÁNDEZ LUCAS



2009



Universidad Nacional Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Hoja de Datos del Jurado

- Datos del alumno Becerra Lobato Noé 56 43 28 25 Universidad Nacional Autónoma de México Facultad de Ciencias Biología 095359382
- 2. Datos del tutor Dr Ismael Hernández Lucas
- Datos del sinodal 1 Dr Javier Izquierdo Sánchez
- Datos del sinodal 2
 Dra
 Laura
 Kawasaki
 Watanabe
- 5. Datos del sinodal 3 M en C Alfonso José Vilchis Peluyera
- Datos del sinodal 4 M en C María de los Angeles Aguilar Santamaría
- Datos del trabajo escrito Caracterización Funcional de los Orígenes de Replicación del Plásmido Simbiótico de *Rhizobium tropici* CFN299 p64 2009

Agradecimientos

A mi tutor, el Dr. Ismael Hernández Lucas por todo el apoyo y la confianza que tuvo en mi.

A mis sinodales por el tiempo dedicado a la revisión de la tesis y por sus valiosos consejos y recomendaciones.

A mis amigos y compañeros del laboratorio por hacer un lugar de respeto y agradable para trabajar. En especial a Ismael, Aurora, Augusto y a Tono por su amistad, ayuda técnica, apoyo y consejos recibidos.

A la Química María Luisa Arias Mendoza por el apoyo que tuve durante la realización de este trabajo.

Al Dr. Javier Izquierdo por sus sugerencias para este trabajo.

Al Dr. Jesús Caballero por sus valiosos consejos.

A la Dra. Esperanza Martínez Romero por muchos conocimientos transmitidos.

A la M. en C. María de los Ángeles Aguilar Santamaría por ser una excelente profesora y recibir siempre consejos muy valiosos de su parte.

A Deyanira y a Nancy por su amistad y contribuciones que me ayudaron a realizar parte de este trabajo.

A mis amigos y a las personas que me quieren porque de alguna forma siempre están conmigo incondicionalmente.

A mi madre, Por todo lo que he aprendido de ella, por su ejemplo, su comprensión y apoyo incondicional, por estar siempre conmigo.

A mis hermanos, Alfredo, Myriam y Maricruz por su compañía, amistad y apoyo, además de lo que me han enseñado cada uno de ellos.

> A mi tía Adelina, Que siempre ha estado al pendiente de mí.

INDICE

RESUMEN	3
I. INTRODUCCIÓN	4
1.1. Fijación biológica del nitrógeno	4
1.2. Interacción Rhizobium-leguminosa	4
1.3. Plásmidos en <i>Rhizobium</i>	7
1.4. Orígenes de replicación en plásmidos de Rhizobium	8
1.4.1. Orígenes de replicación de la familia repABC en Rhizobium	8
1.4.1.1. Sistema de partición en replicones <i>repABC</i>	10
1.4.1.2. RepC en plásmidos repABC	12
1.4.1.3. Regulación de plásmidos <i>repABC</i>	12
1.4.2. Orígenes de replicación de la familia repC en Rhizobium	15
1.4.3. Orígenes de replicación de la familia repAC en Rhizobium	16
1.5. Rhizobium tropici	17
1.6. Estructura genómica de <i>R. tropici</i> CFN299	18
	19
III. OBJETIVOS	21
IV. MATERIAL Y MÉTODOS	22
4.1. Cepas y plásmidos utilizados	22
4.2. Medios y condiciones de cultivo	23
4.3. Transformación bacteriana	23
4.4. Conjugación bacteriana	23
4.5. Purificación de DNA	23
4.6. Southern blot	24
4.7. Amplificación por PCR y clonación	24
4.8. Perfil de plásmidos	24
4.9. Secuenciación	24
4.10. Análisis de Secuencias	25

V. RESULTADOS	26
V. I. Caracterización del gen <i>repC</i>	
5.1.1. Aislamiento del replicador tipo <i>repC</i>	
5.1.2. Clonación de la secuencia amplificada <i>repC</i> en el plásmido pCR2.1	27
5.1.3. Clonación del gen <i>repC</i> en el plásmido pUX19	27
5.1.4. Transferencia de A780 (pUX19-repC) a A. tumefaciens	
5.1.5. Clonación in vivo del gen repC	
5.1.6. Obtención y análisis de la secuencia del gen repC	
5.1.7. Transferencia del gen repC completo a A. tumefaciens	32
V. II. Caracterización del operón <i>repABC</i>	33
5.2.1. Aislamiento de <i>repBC</i>	33
5.2.2. Clonación de <i>repBC</i> en el plásmido pUX19	33
5.2.3. Transferencia de A610 (pUX19-repBC) a A. tumefaciens	34
5.2.4. Detección del operón <i>repABC</i> por Southern blot	34
5.2.5. Obtención y clonación del operón repABC	35
5.2.6. Recuperación del plásmido (pUX19-repABC) de A. tumefaciens	36
5.2.7. Región mínima replicable del operón <i>repABC</i>	36
5.2.8. Obtención y análisis de la secuencia del operón <i>repABC</i>	38
5.2.9. Secuencia del operón <i>repABC</i>	40
V. III. Presencia de repC y repABC en R. tropici CFN299 y CFN299-10	41
5.3.1. Transferencia de repC (A789) y repABC (A627) a R. tropici CFN299 y CFN299-10	41
5.3.2. PCRs de <i>repC</i> y <i>repABC</i> de las cepas CFN299 y CFN299-10	41
5.3.3. Detección de los replicadores de las cepas CFN299 y CFN299-10 por Southern blot.	43
VI. DISCUSIÓN	44
6.1. Gen <i>repC</i>	44
6.2. Operón <i>repABC</i>	48
VII. CONCLUSIÓN	51
VIII. APÉNDICE	52
8.1 Medios de cultivo	52
8.2 Protocolos	53
IX. REFERENCIAS	59

RESUMEN

El género *Rhizobium* establece asociaciones simbióticas con plantas leguminosas, las bacterias infectan las raíces e inducen en ellas nódulos fijadores de nitrógeno. Este proceso permite la fijación biológica del nitrógeno la cual es de gran importancia para los ecosistemas y para la agricultura debido a que representa la mayor fuente de abastecimiento de nitrógeno.

La bacteria *R. tropici* CFN299 habitante del suelo, es capaz de establecer relaciones simbióticas con múltiples leguminosas, posee una estructura genómica conformada por un cromosoma aproximado de 3000 kb, un megaplásmido de 1500 kb, un plásmido simbiótico (pSym) de 600 kb y 2 plasmidos de menor tamaño; (pB) de 240 kb y (pA) de 180 kb. El pSym posee una alta densidad de genes involucrados en la vida libre y simbiótica de *R. tropici* CFN299. En el presente trabajo se identificaron 2 orígenes de replicación en el pSym de *R. tropici* CFN299, uno perteneciente a la familia *repC* el cual posee una identidad de 63% con el gen *repC* del plásmido pRmeGR4a de *S. meliloti* GR4 y el segundo origen identificado se encuentra formando un operón y pertenece a la familia *repABC*, el gen *repC* de este operón posee una identidad de 76% con el gen *repC* del plásmido pSmeSM11a de *S. meliloti* SM11.

Los 2 orígenes de replicación del pSym de *R. tropici* CFN299 fueron clonados y secuenciados, sin embargo solamente el operón *repABC* fue capaz de replicarse autónomamente en *A. tumefaciens*. La región mínima de replicación del operón *repABC* es de 5.5 kb. Diversas evidencias experimentales, apoyan la hipótesis de que el origen de replicación funcional del plásmido simbiótico de *R. tropici* CFN299 corresponde al operón *repABC*.

I. INTRODUCCIÓN

1.1. Fijación biológica del nitrógeno

El nitrógeno es muy abundante en la atmósfera, sin embargo, las plantas no pueden utilizarlo en su forma elemental y tienen que obtenerlo del suelo principalmente en forma de nitratos o amonio. La fijación biológica de nitrógeno es un proceso clave en la biosfera, por el cual microorganismos portadores de la enzima nitrogenasa convierten el nitrógeno gaseoso en nitrógeno combinado. El grupo de bacterias al que se conoce como rizobios inducen en las raíces o en el tallo de las leguminosas la formación de nódulos, dentro de los cuales el nitrógeno gaseoso es reducido a amonio (Wang et al., 2001). El grupo de rizobios taxonómicamente pertenecen al orden Rhizobiales, dividido en cuatro familias y perteneciente a la subdivisión alfa de las proteobacterias (Spaink, 2000). En esta simbiosis, la planta huésped obtiene compuestos nitrogenados por parte del simbionte y la bacteria obtiene por parte de la planta fuentes de carbono y un ambiente favorable para fijar nitrógeno. Esta simbiosis contribuye con una parte considerable del nitrógeno combinado en la tierra y permite a las plantas leguminosas crecer en algunos casos sin fertilizantes nitrogenados. Actualmente se reconocen seis géneros dentro del grupo de de los rizobios: Allorhizobium, Azorhizobium, Bradyrhizobium, Mesorhizobium, Rhizobium y Sinorhizobium, con un número variable de especies descritas (Wang et al., 2001).

1.2. Interacción Rhizobium-leguminosa

Las leguminosas muestran una amplia diversidad ecológica y morfológica, existen desde formas herbáceas hasta árboles tropicales. Muchas leguminosas son noduladas por bacterias del género *Rhizobium* las cuales forman colonias que miden 0.5-1.0 x 1.2-3.0 mm. Se mueven por medio de 1-6 flagelos que pueden ser peritricales o subpolares. Las colonias generalmente son blancas o color beige, circulares, convexas, semitranslúcidas u opacas y mucilaginosas. Son quimio-organotróficas, utilizan una gran variedad de carbohidratos y ácidos orgánicos. Algunas cepas requieren biotina, ácido nicotínico, pantotenato o tiamina como factores de crecimiento. En el género *Rhizobium* se incluye gran número de especies entre las que se encuentran: *R. leguminosarum, R. etli, R. galegae, R. gallicum, R. giardinii, R. hainanense, R. huautlense, R. mongolense* y *R. tropici*. Estas nodulan diferentes leguminosas en zonas templadas o tropicales (Martínez-Romero y Caballero-Mellado, 1996).

Las plantas leguminosas secretan compuestos específicos que atraen a las bacterias. Entre estos compuestos se encuentran flavonoides y en respuesta a ellos, *Rhizobium* activa una serie de genes implicados en la nodulación. El primer paso en la formación de los nódulos es la adherencia de la bacteria a la planta. En la superficie de *Rhizobium* se localiza una proteína específica a la adherencia, la ricadesina, la cual es una proteína que se une al calcio y puede actuar captando complejos de calcio. Otras proteínas como las lectinas, también participan en la adherencia planta-bacteria. Las lectinas han sido identificadas en los extremos de los pelos radicales (Wang *et al.*, 2001).

Después de la unión del simbionte, los pelos radicales se enroscan debido a la acción de sustancias específicas secretadas por la bacteria que se conocen como factores Nod. Algunos pelos radicales se enroscan hasta 360 °. La bacteria penetra entonces en el pelo radical e induce la formación, por parte de la planta, de un tubo de composición similar a la pared celular, conocido como canal de infección. A continuación, la infección alcanza a las células de la raíz adyacentes a los pelos radicales, y los factores Nod estimulan la división de las células vegetales produciendo finalmente el nódulo. Las bacterias son liberadas desde el hilo de infección al citoplasma de las células vegetales por un mecanismo similar al de endocitosis. Las bacterias quedan separadas del citoplasma por una membrana derivada de la planta hospedadora llamada membrana peribacteroidal (MPB). A continuación hay una división continua y sincronizada de las bacterias rodeadas de MPBs. Al cesar la división las bacterias se transforman en unas formaciones ramificadas, hinchadas y deformes, llamadas bacteroides. Estos quedan rodeados, individualmente o en pequeños grupos por la MPB. A la estructura que contiene estos grupos de bacteroides rodeados por la MPB se les llama simbiosomas. Los bacteroides pueden llegar a ser hasta 40 veces más grandes que las células a partir de las que se desarrollaron, y hasta varios miles se encuentran en una sola célula vegetal. La fijación de nitrógeno no se inicia hasta que se han formado los bacteroides, en los cuales se realiza la fijación simbiótica de nitrógeno. En el proceso de fijación de nitrógeno, la enzima nitrogenasa cataliza la reacción (Wang et al., 2001):

 $N_2 + 8H^+ + 8e^- + 16 Mg-ATP \longrightarrow 2NH_3 + H_2 + 16 Mg-ADP + 16Pi$

La nitrogenasa es una proteína de gran tamaño que consiste de dos componentes, la proteína homodimérica que contiene Fe y es codificada por *nifH*, y la proteína tetramérica que contiene Fe y molibdeno (Mo), codificada por los genes *nifD* y *nifK*. La nitrogenasa de los nódulos radiculares posee características similares a la enzima de las bacterias fijadoras de nitrógeno en vida libre, incluyendo la sensibilidad al oxígeno y la capacidad de reducir acetileno y N₂. La formación de H₂ es parte del mecanismo de la nitrogenasa pero representa una pérdida significativa de energía. En algunas especies de *Rhizobium* existe una hidrogenasa codificada por los genes *hup* que es capaz de reciclar el H₂ formado por la nitrogenasa, que resulta en un uso más eficiente de la energía (Rey *et al.,* 1993). Es por ello, que se ha impulsado la transferencia de los genes *hup* a cepas de *Rhizobium* para la inoculación en el campo (Báscones *et al.,* 2000).

Los bacteroides dependen totalmente de la planta para obtener la energía necesaria para la fijación de nitrógeno. Los principales compuestos orgánicos transportados al interior de los bacteroides a través de la membrana peribacteroidal son los intermediarios del ciclo del ácido cítrico, en particular los ácidos de cuatro carbonos: succinato, malato y fumarato. Estos ácidos son utilizados como donadores de electrones para la producción de ATP, y tras su conversión en piruvato, como última fuente de electrones para la reducción del N₂. El primer producto estable que se obtiene de la fijación de N₂ es el amonio, y varias pruebas indican que la asimilación del amonio para formar compuestos de nitrógeno orgánico en los nódulos radicales lo lleva principalmente la planta (Wang *et al.*, 2001). También se ha reportado que el nitrógeno puede ser transferido a la planta en forma de alanina (Waters *et al.*, 1998).

Durante el proceso de simbiosis la planta también expresa proteínas específicas del nódulo a las que se llama nodulinas. Entre ellas, la leghemoglobina tiene la función de aportar O_2 a los bacteroides y de controlar los niveles de oxígeno. La leghemoglobina se localiza en el citosol de las células de la planta infectada por bacteroides y es la que confiere el típico color rojo o rosado de los nódulos funcionales (Wang *et al.*, 2001).

La relación procarionte-eucarionte constituye un modelo de estudio para descifrar el funcionamiento bacteriano en vida libre y simbiosis. En *Rhizobium* se encuentran plásmidos en su estructura genómica. Algunos de estos han sido estudiados, por ejemplo, los de las cepas *R*. spp. NGR234 (Freiberg *et al.*, 1997) y *Sinorhizobium meliloti* 1021 (Galibert *et al.*, 2001) han sido completamente secuenciados. La secuencia nucleotídica de estos replicones ha revelado características interesantes entre las que destacan la presencia de múltiples elementos de inserción, tRNA para la síntesis de arginina así como genes que codifican para las proteínas MinCDE (Finan *et al.*, 2001) involucradas en el proceso de división celular. La presencia de estos genes demuestra que algunos plásmidos de *Rhizobium* son esenciales para la supervivencia de la bacteria.

1.3. Plásmidos en *Rhizobium*

Un plásmido es una molécula de DNA extracromosomico circular o lineal que se replica y se transcribe independientemente del DNA cromosómico. Estos replicones han sido encontrados en las tres formas representativas de vida nombradas dominios; *Archaea, Bacteria y Eukarya* (Woese *et al.*, 1990) y pueden incorporar o liberar genes por recombinación o transposición favoreciendo cambios genéticos en poblaciones de bacterias (Del Solar *et al.*, 1998). En la naturaleza existen una gran cantidad de bacterias que interactúan con células eucariontes, por ejemplo la mayoría de especies de *Agrobacterium* son patógenos para las plantas mientras miembros del genero *Rhizobium* fijan nitrógeno en simbiosis con leguminosas. A finales de los años 60 y a principios de los 70 quedó claro que ambos géneros poseen plásmidos de bajo número de copias y que la mayoría de los genes involucrados en las interacciones planta-microbio se encuentran en tales plásmidos (Higashi, 1967; Escobar y Dandekar, 2003).

El género *Rhizobium*, es muy conocido entre las bacterias gram negativas por su enorme cantidad de información genética extracromosomal. Esta información se encuentra en plásmidos que varían en tamaño, desde 150 hasta 1500 kb y son componentes comunes de *Rhizobium*. Gran parte de la información involucrada en simbiosis se encuentra en los plásmidos simbióticos (pSym) que pueden coexistir con otros, llamados plásmidos crípticos (García de los Santos *et al.*, 1996). Los plásmidos pueden comprender hasta una tercera parte del genoma de *Rhizobium* como es el caso de *Rhizobium etli* CFN42 (González *et al.*, 2006).

El estudio de los plásmidos en *Rhizobium* ha permitido describir diferentes funciones como son: fijación de nitrógeno, competencia para nodular, genes requeridos para la síntesis de lipopolisacáridos y otros genes catabólicos (Brom *et al.*, 2000). Estudios de genética de poblaciones en *Rhizobium*, basados en características cromosómicas y plasmídicas sugieren que la transferencia horizontal genética ocurre en poblaciones naturales (Schofield *et al.*, 1987).

1.4. Orígenes de replicación en plásmidos de Rhizobium

En las bacterias del género *Rhizobium* se han localizado 3 familias de replicadores (*repABC*, *repC* y *repAC*). Un replicador es una región esencial la cual contiene los genes o loci involucrados en la replicación y control del plásmido (Del Solar *et al.,* 1998).

1.4.1. Orígenes de replicación de la familia repABC en Rhizobium

La familia repABC consiste en un operón que se caracteriza por la presencia de 3 genes: repA repB y repC y por la presencia de un gen que codifica un pequeño RNA antisentido que se transcribe en dirección opuesta al operón, este RNA contra-transcrito (ctRNA) no se traduce y esta conservado ampliamente entre los genes repB y repC (Fig. 1.1) (Venkova-Canova *et al.*, 2004). Los replicadores repABC se encuentran ampliamente distribuidos en las α proteobacterias.



Figura 1.1. Representación de la región de replicación y partición de *repABC*. En flechas se observan los 3 genes que constituyen el operón: *repA*, *repB* y *repC*. P indica la posición de la región promotora del operón, en gris se muestra la secuencia que codifica para el ctRNA.

Este tipo de replicadores se ha encontrado en plásmidos y cromosomas en al menos 19 géneros de α proteobacterias (Tabla 1.1) (Cevallos *et al.*, 2008). Algunas bacterias poseen varios plásmidos, cada uno conteniendo un operón *repABC*, además se ha encontrado que algunos plásmidos contienen 2 operones *repABC* funcionales. Un ejemplo de esta peculiaridad son *R. etli* CFN42 y *R. leguminosarum* 3841, las cuales contienen 6 plásmidos, todos pertenecientes a la familia *repABC*. Hay 2 replicones en *R. etli* CFN42 y uno en *R. leguminosarum* 3841 contienen 2 operones *repABC*, esto sugiere que diferentes plásmidos con *repABC* pertenecen a diferentes grupos de incompatibilidad (Young *et al.*, 2006; González *et al.*, 2006). La incompatibilidad ocurre cuando 2 o más plásmidos no pueden coexistir en la misma célula como replicones independientes en ausencia de presiones selectivas o como consecuencia de un sistema de replicación o de partición similar.

Los productos de los genes *repA* y *repB* están involucrados en la partición y regulación de número de copias del plásmido, *repC* es esencial para la replicación y este es considerado como la proteína iniciadora (Bartosik *et al.*, 1998; Ramírez-Romero *et al.*, 2000; Tabata *et al.*, 1989).

Tabla 1.1. Lista de es	pecies bacterianas o	ue co	ntienen opero	nes repABC en sus replice	ones
Сера	Replicón	n	tamaño	Referencia	GenBank
	A CD3/01		202 5 171	0	
Acidiphilium cryptum JF5	pACRY01	1	203.5 Kb	Sin publicar	NC_009467
Agrobacterium rhizogenes K 599	pRIA40	1	185.4 Kb	Mankin <i>et al.</i> 2007	EU186381
Agrobacterium rhizogenes MAFF03-01724	pRi2039	1	217 5 Kb	Moriguchi et al. 2007	NC 002575
Agrobacterium tumefaciens C58	ChrL*1	1	2.07 Mb	Goodner <i>et al.</i> 2001	NC_003305
-8	pATC58	1	542.7 Kb	,	NC 003064
	pTiC58	1	214.2 Kb		NC_003065
Agrobacterium tumefaciens	pTiB6S3	1	nd	Tabata et al., 1989	M24529
Agrobacterium tumefaciens R10	pTiR10	1	194.1 Kb	Zhu et al., 2000	AF242881
Agrobacterium tumefaciens MAFF301001	pTi-SAKURA	1	206.4 Kb	Suzuki et al., 1998	NC_002147
Bradyrhizobium sp. BTA11	pBBta01	1	228.8 Kb	Giraud <i>et al.</i> , 2007	NC_009475
Brucella abortus by 1 str 9–941	ChrII	1	1.16 MD	Chain at al. 2005	NC_006933
Brucella melitensis 16 M	ChrII	1	1.15 Mb	Del Vecchio <i>at al.</i> 2003	NC_003318
Brucella suis 1330	ChrII	1	1.2 Mb	Paulsen et al 2002	NC_004311
Brucella ovis ATCC 25840	ChrII	1	1.16 MB	Sin publicar	NC 009504
Mesorhizobium sp. BNC1	p1	2	343.9 Kb	Sin publicar	NC_008242
-	p2	1	131.2 Kb	-	NC_008243
	p3	1	47.5 Kb		NC_008244
Mesorhizobium loti MAFF303099	pMLa	1	351.9 Kb	Kaneko et al., 2000	NC_002679
	pMLb	1	208.3 Kb		NC_002682
Nitrobacter hamburgensis X14	pl	1	294.8 Kb	Sin publicar	NC_007959
	p2	1	188.3 Kb		NC_007961
Occurring la hatermain UTCC2507	p3	1	121.4 K0	Sin muhlimm	NC_00/961
Oceanicola baisensis HTCC2597	Obat HTCC2597a	1	nd	Sin publicar	NZ_AAMO01000009
Ochrobactrum anthroni ATCC 49188	ChrII	1	1 89 Mb	Sin publicar	NC 009668
Senrobaen um unin opt ATEC 49188	pOANT01	1	170 3 Kh	Shi publicai	NC_009669
	pOANT02	1	101 4 Kb		NC_009670
	pOANT03	2	93.5 Kb		NC 009671
Oligotropha carboxidovorans	pHCG3	1	133 Kb	Fuhrmann et al., 2003	NC_005873
Paracoccus versutus UW1	pTAV320	1	nd	Bartosik et al., 1998	U60522
Rhizobium etli CFN42	p42a	2	194.2 Kb	Gonzáles et al., 2006	NC_007762
	p42b	1	184.3 Kb		NC_007763
	p42c	1	250.9 Kb		NC_007764
	p42d	1	371.2 Kb	González et al., 2003	NC_004041
	p42e	1	505.3 Kb	González et al., 2006	NC_007765
Dhisshim Isaamin samma hussisis 2041	p42f	2	642.5 Kb	Norma at al. 2006	NC_00/766
Knizobium leguminosarum bv viciae 3841	pkL/	2	131.3 KD	Young et al., 2006	NC_008382
	nRL9	1	352.7 Kh		NC_008379
	nRL10	1	488 1 Kb		NC_008381
	pRL11	1	684 2 Kb		NC 008384
	pRL12	1	870 Kb		NC 008378
Rhizobium sp. NGR234	pNGR234a	1	536.1 Kb	Freiberg et al., 1997	NC_000914
Rhodobacter sphaeroides 2.4.1.	pB	1	114.1 Kb	Sin publicar	NC_007488
	pD	1	100.8 Kb		NC_007490
Roseovarius nubinhibens ISM	RnuISM	1	nd	Sin publicar	NZ_AALY01000005
Roseovarius sp. 217	Ros217a	1	nd	Sin publicar	NZ_AAMV01000007
	Ros217b	1	nd		NZ_AAMV01000021
Researching on HTCC2601	R0S21/C Bas HTCC2601a	1	nd	Sin publicar	NZ_AAMV01000023
Roseovarius sp. 111002001	Ros HTCC2601a	1	nd	Sili publicai	NZ_AATQ01000004
	Ros HTCC2601c	1	nd		NZ_AATO01000027
Roseovarius sp. TM1035	RosTM1035	1	nd	Sin publicar	NZ ABCL01000010
Ruegeria sp. PR1b	pSD25	1	148.6 Kb	Zhong et al., 2003	NC 004574
Sagittula stellata E-37	SsteE37	1	nd	Sin publicar	NZ_AAYA01000007
Sinorhizobium meliloti SM11	pSymA	1	nd	Stiens	Pers. Comm.
	pSymB		nd		
	pSmeSM11a	1	144.1 Kb	Stiens et al., 2006	DQ145546
	pSmeSM11b	2	181.2 Kb	Stiens <i>et al.</i> , 2007	EF066650
Sinornizobium meliloti 1021	pSymA	1	1.35 MD	Einen et al., 2001	NC_003037
Sinorhizohium malilati MBAQ	pSymB pMBA9a	1	108.5 KD	Finan et al., 2001 Watson v Have 2006	AV01/873
Sinorhizobium medicae WSM/19	pMBA9a nSMED01	1	1.5 Mb	Sin publicar	NC 009620
Sinorm200ium meuteue w Sivi41)	nSMED02	1	1.3 Mb	Shi publicai	NC_009621
	pSMED02	1	219.3 Kb		NC 009622
Stappia aggregatal AM 12614	Sta IAM12614a	1	nd	Sin publicar	NZ AAUW01000020
	Sta IAM12614b	1	nd	1	NZ_AAUW01000021
Sulfitobacter sp. NAS-14.1	Sulfit NAS141a	1	nd	Sin publicar	NZ_AALZ01000014
	Sulfit NAS141b	1	nd		NZ_AALZ01000012
	Sulfit NAS141c	1	nd		NZ_AALZ01000013
	Sulfit NAS141d	1	nd		NZ_AALZ01000015
	Sulfit NAS141e	1	nd	<u> </u>	NZ_AALZ01000011
Sulfitobacter sp. EE-36	SulfitEE36	1	nd	Sin publicar	NZ_AALV01000012
Auninobacier autoiropnicus Py2	PACI01	1	310.1 KD	Sin publicar	NC_009/1/

n= número de operones *repABC* en cada replicón **nd**= tamaño de replicón no determinado

Tomado de M. A. Cevallos et al. Plasmid 60 (2008) 19-37

Entre los operones *repABC* hay algunas diferencias como las que involucran elementos de regulación transcripcional, diferencias relacionadas al número y posición de sitios *parS* y la presencia de minigenes que codifican péptidos dentro del operón *repABC*. Estas diferencias pueden observarse en la Fig. 1.2 (Cevallos *et al.*, 2008).



Figura 1.2. Las flechas blancas representan genes repA y repB que codifican proteínas involucradas en la partición y regulación transcripcional negativa del operón. Los círculos blancos muestran los sitios *parS*. El circulo negro muestra la caja *vir*. Las flechas sombreadas en la región río arriba de *repC* indican la posición del gen que codifica para el ctRNA y adyacente a estos en dirección opuesta se muestra la posición de los genes repC que codifican proteínas para replicación. La flecha negra representa el gen repD. Los cuadros muestran la posición de los promotores y los triángulos la posición de cajas *tra*. Tomado de M. A. Cevallos *et al.* 2008.

1.4.1.1. Sistema de partición en replicones *repABC*

El sistema de partición frecuentemente se encuentra en módulos o cassettes, los cuales consisten típicamente de 2 proteínas (ParA y ParB), y un sitio de partición (*parS*) parecido a los centrómeros. La secuencia *parS* es específica en cada especie y es incompatible con otros plásmidos con un sitio idéntico o similar (Funnell, 2005). La proteína ParB se une a la secuencia *parS* y a la proteína ParA. En el sistema de partición de algunos plásmidos como el P1, ParA, ParB y *parS* corresponden a una ATPasa, una proteína de unión al centrómero y al sitio centrómero-partición respectivamente (Funnell, 2005). En general, el sistema de partición de plásmidos puede ser dividido en 2 familias, basandose en la homología de la secuencia de su ATPasa (Gerdes *et al.*, 2000).

El primer paso en la partición es la formación de un complejo en la secuencia *parS*. La proteína ParB al unirse a *parS* crea una estructura que segrega las copias de plásmidos a las células hijas. Las proteínas IHF también se unen a este sitio para formar parte de la estructura. El complejo de ParB e IHF con *parS* es llamado complejo de partición. El complejo formado es mediado por ParA que es una ATPasa oscilante que es capáz de formar estructuras helicoidales, esta ATPasa posiciona al plásmido en el centro del nucleoide de la bacteria, posteriormente separa y segrega a los plásmidos a las

células hijas (Ebersbach y Gerdes, 2004). Las evidencias bioquímicas y genéticas en cepas silvestres y en las mutantes ParA demuestran que la energía de unión e hidrólisis del ATP es necesaria en ParA para interaccionar productivamente con ParB en el complejo de partición (Bouet y Funnell, 1999). También existe el sistema ParM del plasmido R1, en el cual la ATPasa (ParM) parecida a actina es codificada por *par*, locus que tambien codifica ParR y el sitio *parC* parecido al centromero en el cual se une ParR (Jensen *et al.*, 1998). De los 2 sistemas (ParA y ParM) *repAB* es similar a ParA.

Los replicones *repABC* tienen un número muy bajo de copias por lo que requieren un mecanismo eficiente de segregación para asegurar su propagación estable. RepA y RepB son parte de la maquinaria de segregación (Ramírez-Romero *et al.*, 2000). Los productos de los genes *repA* y *repB* del pSym de *R. etli* están involucrados en la partición del pSym y en la regulación del número de copias del plásmido (Cevallos *et al.*, 2002).

Se ha demostrado que las proteínas RepA y RepB, y la secuencia *parS* son determinantes para la segregación en los plásmidos: pTiB6S3 de *A. tumefaciens*, pTAV320 de *Paracoccus versutus*, pCFN42d de *R. etli* y pSymA de *S. meliloti*. Los análisis genéticos y moleculares muestran que inserciones, mutaciones o deleciones en *repA* o *repB* disminuyen la estabilidad de los plásmidos (Tabata *et al.*, 1989; Bartosik *et al*, 1998; Ramírez-Romero *et al.*, 2000). Los loci *parS* de los plásmidos mencionados anteriormente han sido identificados. Su secuencia consenso consiste de una o más copias palindrómicas de 16 pb (**GTTNNCNGCNGNNAAC**) y cumplen con 3 requisitos: primero, son esenciales para la estabilidad de los plásmidos; segundo, son sitios de unión para RepB y finalmente son incompatibles con su respectivo plásmido parental (Bartosik *et al*, 1998; Soberón *et al.*, 2004).

El número de sitios *par* presentes en el locus *parS* de replicones *repABC* varía ampliamente. Algunos plásmidos como el pCFN42d contienen solo una secuencia consenso *par*, mientras algunos plásmidos como el pSymA de *S. meliloti* contienen 6 (Bartosik *et al*, 2001; Soberón *et al.*, 2004). Sin embargo sólo una secuencia consenso es necesaria para una adecuada partición del plásmido.

1.4.1.2. RepC en plásmidos *repABC*

La proteína RepC, codificada por el operón *repABC* tiene un papel limitante para la replicación. Las mutaciones, deleciones e inserciones en este gen anulan la capacidad de replicación de los plásmidos (Tabata *et al.*, 1989; Ramírez-Romero *et al.*, 2000). La secuencia RepC no tiene homología con otra función asignada. La mayoría de los genes *repC* están presentes en los operones *repABC*, y en todos los casos que han sido clonados esos operones, replican el plásmido en el que han clonado dentro de la bacteria parental, siendo vectores no replicables para la cepa en donde se introducen (Bartosik *et al.*, 1998; Turner y Young, 1995; Cevallos *et al.*, 2002).

La primera evidencia de que repC es la región mínima esencial y suficiente para la replicación fue obtenida de la inserción del gen repC del plásmido pTAV320 de *P. versutus* en un vector suicida bajo el control de un promotor, se obtuvieron transconjugantes de *P. versutus* (Bartosik *et al.*, 1998), esto indica que el gen repC es esencial para la replicación.

1.4.1.3. Regulación de los plásmidos *repABC*

Los operones repABC poseen mecanismos de regulación para asegurar la producción y cantidad apropiada de la proteína iniciadora y de los productos de la partición para una replicación estable y para segregar los plásmidos recién replicados dentro de las células hijas. Este sistema incluye mecanismos de regulación transcripcionales y post-transcripcionales. Algunos son comunes en casi todos los plásmidos de la familia repABC mientras que otros sólo se encuentran en pocos miembros de la familia. La replicación de los plásmidos por repABC es influída por mecanismos de regulación positiva y negativa, repA o repA con repB son represores de la transcripción del operón (Pappas y Winans, 2003; Ramírez-Romero *et al.*, 2001). El gen entre repB y repC que codifica para un ctRNA actúa como un regulador post-transcripcional negativo de la expresión del gen repC, pero también es requerido para la replicación del plásmidos (Venkova-Canova *et al.*, 2004). El número de copias de algunos plásmidos repABC es regulado positivamente por un mecanismo de respuesta en coro mediado por TraR y su autoinductor asociado (Li y Farrand, 2000).

Algunos replicadores repABC se transcriben de un operón traI-trb que se localiza en la región río arriba de repABC y otros no tienen genes en común con traI-trb. Un ejemplo de este tipo de operones se localiza en el plásmido pTiC58 de *Agrobacterium* el cual tiene una sofisticada regulación. Su número de copias puede ser influido positiva y negativamente por factores externos que incluyen sistemas de respuesta en coro y factores liberados por el hospedero. En contraste los que no se relacionan con un operón *traI-trb* como el plásmido pCFN42d de *R. etli* parecen tener un mecanismo más simple de regulación.

Los plásmidos Ti de Agrobacterium modulan su número de copias y su transferencia conjugativa en respuesta a su densidad de población y por la presencia de señales químicas generadas por la planta hospedera. Los operones repABC de los plásmidos pTi, involucrados en la segregación y partición del plásmido, se transcriben en dirección opuesta de los genes tral-trb. La región intergénica entre estos 2 genes cuenta con 4 promotores y secuencias involucradas en su regulación transcripcional (Fig. 1.2) (Pappas y Winans, 2003; Li y Farrand, 2000). La actividad de los promotores P1, P2 y P3, y la transcripción del operón traI-trb es influída positivamente por TraR, homólogo a LuxR y tiene un papel central en el sistema de regulación por respuesta en coro en Agrobacterium. El dímero TraR ejerce su efecto en presencia de N-3oxooctanoyl-Lhomoserina lactona (3-O-C₈–AHL), un autoinductor sintetizado por Tral, y este involucra su unión a las cajas tra II y III (Fig. 1.2). La adición exógena de (3-O-C₈-AHL) activa los promotores P1, P2 y P3 y produce un aumento en el número de copias del plásmido, probablemente aumentando la concentración intracelular de RepC, la proteína iniciadora. Este incremento en el número de copias del plásmido ha sido asociado con un aumento en la tumorigénesis (Pappas y Winans, 2003).

Otros genes reguladores que contribuyen al control del número de copias de los plásmidos son los sensores transmembranales como la cinasa VirA y su regulador responsable VirG que constituyen un sistema de 2 componentes que coordinan la expresión de los genes Ti *vir* y modulan la actividad transcripcional del operón *repABC*, positiva y negativamente. En un medio extracelular ácido, VirA puede detectar varios factores, incluyendo compuestos fenólicos y algunos monosacáridos liberados por la planta hospedera (Zhu *et al.*, 2000). Fosfo-VirG, la forma activa de la proteína se une a la secuencia de DNA llamada caja *vir*. La adición de acetosiringona, un componente fenólico usado como inductor, provoca un incremento en el número de copias del los plásmidos pTi58 y pTiR10, este efecto se anula en mutantes carentes de *virA* o *virG*. En el plásmido

acetosiringona aumenta sólo la actividad del promotor P4 (Cho y Winans, 2005) que es capaz de mantener la replicación en ausencia del autoinductor y de los factores liberados por el hospedero (Pappas y Winans, 2003). *Agrobacterium* en vida libre contiene una copia del plásmido Ti por cromosoma y esta copia aumenta por un incremento en su población provocando un mecanismo de respuesta en coro y también aumenta por factores liberados por la planta hospedera.

Se ha observado que 2 pSym de *Rhizobium* (pRL1JI y pNGR234a) tienen semejanzas con los plásmidos Ti, ambos contienen un operón *traI-trb* adyacente a *repABC* que es transcrito en dirección contraria. El plásmido pRL1JI responde al mecanismo de respuesta en coro, mientras que el plásmido pNGR234a responde a factores liberados por el hospedero. En el plásmido pNGR234a la adición de daidzeína, un flavonoide que induce los genes *nod* en esta cepa también induce la transcripción de su operón *repABC*. Este mecanismo se desconoce aún, ya que la región promotora de *repABC* no contiene cajas nod, sitios al cual NodD, se une al DNA (Perret *et al.*, 1999).

No todos los plásmidos contienen un operón *traI-trb* o son regulados por el mecanismo de respuesta en coro. El pSym de *R. etli* (pCFN42d) contiene un pseudogen *traI* en la región río arriba del operón *repABC*, pero carece de genes *trb*. Los análisis genéticos y moleculares muestran que este operón contiene sólo un promotor en la región río arriba de *repA* y es activado por SigA, el factor sigma de expresión constitutiva de *Rhizobium* (Ramírez-Romero *et al.*, 2001). Esto sugiere que en la transcripción de *repABC* de pCFN42d el promotor no es influenciado por señales externas.

Detallados análisis de pCFN42d y pTiR10 revelan que RepA y RepB son reguladores negativos de la transcripción del operón *repABC*, RepA se une a una región operadora a pocas bases río abajo de *repABC*, estimulada por RepB. RepA, una ATPasa débil, se une a la región operadora en presencia de ATP o ADP. El centro de la región operadora consiste de una secuencia de un palíndromo imperfecto, deleciones en esta región da como resultado una expresión constitutiva. (Ramírez-Romero *et al.*, 2001).

Una característica universal de los operones repABC es la presencia de una secuencia intergénica entre los genes repB y repC, la cual provoca incompatibilidad con su plásmido parental (Ramírez-Romero *et al.*, 2000) y tiene una alta similitud con regiones

que codifican para un ctRNA. Los promotores para los genes que codifican los RNAs antisentido son conservados en secuencia y tienen similitudes con la secuencia consenso del promotor σ70 de *E. coli* (Venkanova-Canova *et al.*, 2004). La secuencia consenso del promotor ctRNA es <u>TTGACAGTGATTCGTGGAAATGTGATTCT</u> en la que se observa subrayadas las regiones -10 y -35.

La estructura secundaria predicha del ctRNA es un tallo asa con una cola rica en uracilo en la terminación 3' pareciéndose a los terminadores transcripcionales independientes de Rho. Las regiones intergénicas entre repB y repC que contienen mutaciones que anulan la actividad promotora del gen que codifica para el ctRNA, son incapaces de ejercer incompatibilidad con su plásmido parental. Se cree que el ctRNA podría estar modulando los niveles de RepC y por lo tanto el número de copias de los plásmidos. Chai y Winans (2005) describe un modelo en el que propone que el mRNA de repABC tiene 2 estructuras secundarias alternativas, en las cuales el ctRNA está unido, o no unido con su secuencia complementaria. En ausencia del ctRNA la sección del mRNA de *repABC* que incluye la región intergénica entre *repB* y *repC* forma una estructura en forma de tallo asa, donde solamente el codón de inicio de repC y la secuencia predicha del sitio de unión al ribosoma (Shine-Dalgarno) quedan en la cadena sencilla (Fig. 1.3 C). En esta conformación la traducción de *repC* es permitida. En la secuencia del mRNA de repABC cuando se encuentra presente el ctRNA se forma un tallo asa en la región río arriba de repC cerca del codón de inicio ocluyendo el Shine-Dalgarno (Fig. 1.3 A), de este modo se suprime la traducción de repC.

Figura 1.3. A) estructura secundaria del ctRNA. **B)** Heterodúplex de mRNA ctRNA-*repABC*, la secuencia del Shine-Dalgarno es ocluida por la estructura del tallo asa, atenuando la traducción de *repC*. **C)** En ausencia del ctRNA, el mRNA de *repABC* adopta estructura alternativa en la cual la secuencia del Shine-Dalgarno es totalmente accesible al ribosoma. Tomado de M. A. Cevallos *et al.* 2008.



1.4.2. Orígenes de replicación de la familia repC en Rhizobium

La segunda familia de replicadores son los repC. Estos están relacionados evolutivamente con los replicadores de la familia repABC. El gen repC codifica la proteína iniciadora, pero en contraste con la organización repABC no se encuentran asociados con los genes repA y repB (Mercado-Blanco y Olivares, 1994). El primer miembro de esta familia fue aislado de pRmeGR4a que es un plásmido críptico auto-transferible de *S. meliloti* GR4, y después se observó que este plásmido se encuentra en poblaciones naturales de *S. meliloti*, *Sinorhizobium fredii* y *R. tropici* (Burgos *et al.*, 1996; Mercado-Blanco y Olivares, 1993; Villadas *et al.*, 1995). De esta familia de replicadores se ha estudiado poco acerca de sus mecanismos de regulación.

Se ha demostrado, que en el plásmido pRmeGR4a de *S. meliloti* se localiza un gen en la región río arriba de la secuencia repC que codifica para un ctRNA. Este tiene un papel fundamental como un regulador negativo de la expresión del gen repC y es un sitio de incompatibilidad para este plásmido. Esta región nombrada *inc1* (Izquierdo *et al.*, 2005), tiene similitud con los genes que codifican para los ctRNAs localizados entre repB y repC de los replicadores de la familia repABC (Venkova-Canova *et al.*, 2004) y es responsable de incompatibilidad.

En el plásmido pRmeGR4a la región mínima replicable consiste de 2 genes, uno que codifica para la proteína RepC y el otro para un ctRNA que sólo interactúa con repC ya que no modula los productos codificados en el plásmido, además de que tiene una estructura parecida a un tallo asa, por lo que actúa como un terminador transcripcional independiente de Rho regulando negativamente el promotor de repC (Izquierdo *et al.*, 2005).

Las regiones promotoras de los genes que codifican para los ctRNAs de los replicadores repABC son parecidas a la región promotora del gen que codifica el ctRNA del gen repC del plásmido pRmeGR4a de *S. meliloti*. La diferencia entre ambos replicadores consiste en la presencia de un sistema activo de partición codificado por los genes repA y repB y que además ejercen una regulación negativa de la transcripción del operón.

1.4.3. Orígenes de replicación de la familia repAC en Rhizobium

La familia repAC contiene un sistema de replicación que consiste solamente de repA y un gen parecido a repC, se ha encontrado al menos en el plásmido pRL7jl de *R*. *leguminosarum* (Young *et al.*, 2006). El gen repC ha sido encontrado en estrecha asociación con repA, pero no con repB o algún RNA antisentido y estos genes son capaces de mantener su replicación en *Rhizobium* (Pérez-Segura y Cevallos, datos sin publicar).

1.5. *Rhizobium tropici*

Se llamada *tropici* por crecer cerca de y en el trópico, nativa de Brasil y México pero también ha sido aislada de nódulos de *P. vulgaris* de suelos ácidos en Kenya, y en suelos arenosos del suroeste de Francia. Es una bacteria aeróbica, gram-negativa, la cual forma colonias circulares, convexas, semitranslúcidas y usualmente crece de 2 a 4 mm de diámetro durante 2 a 4 días en medio de levadura peptona (PY). Producen ácido en medio de levadura malta (YM). No producen 3-cetolactosa sin embargo crecen en medio mínimo (MM) conteniendo lactosa y son resistentes al ácido nalidíxico (Martínez-Romero *et al.*, 1991).

Rhizobium tropici, bacteria con un amplio rango de nodulación, es capaz de sobrevivir en suelos con pH ácido y metales pesados. La cepa *R. tropici* PRF81 se ha usado como inoculante en suelos brasileños (Hungria y Vargas, 2000; Mostasso *et al.*, 2002). Esta cepa no tiene ninguna modificación genética, sin embargo es capaz de mejorar la productividad agrícola en suelos cuyo pH sea <5.0, presenten toxicidad por aluminio, bajo contenido de fósforo, insuficiente agua y altas temperaturas ambientales. *R. tropici* se divide en dos subgrupos (Tabla 1.2).

Tabla 1.2. Subgrupos de Rhizobium tropici.					
Rhizobium tropici					
Subgrupo A CFN299	Subgrupo B CIAT 899				
Crecimiento 30°C	Crecimiento 30 a 40°C				
Colonias en YM blancas opacas	Colonias en YM translúcidas				
No móvil en agar 0.3%	Móvil en agar 0.3%				
Requiere Ca^{\dagger} en PY	No requiere Ca^{+} en PY				
Susceptibles a antibióticos y metales pesados	Resistente a Ni, Pb, Co, Cu, Ag y Cr				
Resistente al ácido nadilixico	Resistente a cloranfenicol, carbenicilina, espectinomicina y rifampicina				
No produce clorosis	Produce clorosis				

R. tropici establece asociaciones simbióticas con múltiples leguminosas (Hernández-Lucas *et al.*, 1995b.) (Tabla 1.3), posee una gran versatilidad metabólica para la utilización de diferentes fuentes de carbono tanto en vida libre como en simbiosis.

	C	aracterísticas de nódulos formados	con:
		R. tropici	
	CIAT899	UMR1173	CFN299
Albizia lebbeck	N + F	N + F	N + F
Cajanus cajan	Ν	Ν	Ν
Canavalia ansiformis	N	Ν	Ν
Clianthus formosus	F	Ν	F
Crotalaria sericia	N ^s + F	N ^s + F	N ^s + F
Cyamopsis tetragonoloba	0	0	0
Desmanthus illinoensis	F	0	F
Desmosium canadense	F	0	F
Flemingia congesta	0	0	0
Gliricidia maculata	F	F	F
Glycine max	Ν	Ν	Ν
Indigofera tinctoria	N ^s	N ^s + F	N ^s + F
Leucaena leucocephala	F	F	F
Lotus corniculatus	F	0	F
Macroptilium atropurpureum	N ^s + F	N ^s + F	N ^s + F
Phaseolus angularis	N ^s	N ^s	N ^s + F
Phaseolus vulgaris	F	F	F
Sesbania exaltata	F	0	F
Tephrosia vogelii	Ν	Ν	Ν
Vigna umbellata	N ^s + F	N ^s + F	N ^s + F
Vigna unguiculata	N + F	N + F	N + F
Vigna vexillatta	N ^s + F	N ^s + F	F

Tabla 1.3. Rango de hospedaje de Rhizobium tropici capaces de nodular fríjol

N^s, nódulos carentes de leghemoglobina y conteniendo adicionalmente células senescentes oscurecidas Tomado de Hernández-Lucas *et al.* 1995b. N, nódulos carentes de leghemoglobina

F, nódulos con leghemoglobina

0, sin nódulos

1.6. Estructura genómica de R. tropici CFN299

La estructura genómica de R. tropici CFN299 se encuentra constituida por un cromosoma de aproximadamente 3000 kb, un megaplásmido de 1500 kb, un plásmido simbiótico (pSym) de 600 kb y dos plásmidos: el plásmido A (pA) de 180 y el plásmido B (pB) de 240 kb (Geniaux et al., 1995). Se ha observado que exoF se encuentra en el megaplásmido, en el pA se localizan genes para el transporte de compuestos exudados por la raíz del fríjol (Rosenblueth et al., 1998) y el pB parece estar relacionado con la competitividad de la bacteria (Martínez-Romero y Rosenblueth, 1990). Para ampliar los conocimientos sobre el pSym de R. tropici CFN299 se ha desarrollado un sistema de clonación en vivo para aislar plásmidos de Rhizobium en E. coli. Con esta estrategia se logró clonar fragmentos (300-550 kb) del pSym de R. tropici CFN299 (Hernández-Lucas et al., 2002). Los fragmentos clonados se encuentran sobrelapados y cubren la mayor parte de este replicón. También se ha secuenciado parcialmente un fragmento de 300 kb del pSym de R. tropici CFN299. El análisis de secuencia revela que este plásmido contiene una alta densidad de genes codificables entre los que se encuentran genes de nodulación, fijación de nitrógeno, transferencia, replicación, quimiotaxis, metabolismo primario, reguladores, transportadores así como un número discreto de secuencias de inserción.

II. ANTECEDENTES

La estructura genómica de *R. tropici* CFN299 presenta un plásmido simbiótico de 600 kb. En estudios previos se ha analizado el papel de algunos de sus genes, entre los cuales algunos codifican para enzimas de rutas metabólicas tales como la citrato sintasa (Hernández-Lucas *et al.*, 1995a) y la isocitrato liasa, enzimas que participan en el ciclo de los ácidos tricarboxílicos y en el ciclo de glioxilato. Las mutantes en el gen de citrato sintasa afectan el proceso simbiótico, sin embargo, las mutantes en los genes del ciclo del glioxilato no afectan el desarrollo de la relación entre *R. tropici* CFN299 y la leguminosa (Ramírez-Trujillo *et al.*, 2007).

Se ha establecido que el pSym de *R. tropici* CFN299 al ser transferido a otras bacterias como Agrobacterium tumefaciens (que no cuenta con el plásmido Ti que le confiere la patogenicidad) adquiere la habilidad de nodular y fijar nitrógeno; sin embargo dicha habilidad es muy baja, 10 al 25 % por planta comparado con la cepa donadora R. tropici CFN299 (Martínez-Romero et al., 1987). Otra bacteria a la cual se le transfirió este plásmido simbiótico es Ensifer adherens (actualmente conocido como Sinorhizobium adherens) la cual adquirió las mismas características de simbionte mutualista con una fijación biológica de nitrógeno baja con respecto a R. tropici CFN299 (Rogel et al., 2001). Si el pSym es transferido a patógenos como Brucella melitensis esta bacteria es capaz de inducir esporádicamente pseudonódulos incapaces de fijar nitrógeno. Otra de las características del pSym es que no se ha podido curar, de las 600 kb que lo constituyen sólo se han podido deletar 300 kb las cuales comprenden los genes para nodular y fijar nitrógeno. Mavingui y colaboradores (1997), señalan que 30 kb del pSym de R. tropici CFN299 intervienen no solo en la capacidad de nodular M. atropurpureum, sino que le otorga a la bacteria una mayor competitividad en comparación a la cepa silvestre. Cabe señalar que el aumento de número de copias de este fragmento es suficiente para hacer más eficiente a la bacteria en nodulación. Por todo lo anterior, el plásmido simbiótico de R. tropici CFN299 no sólo tiene un papel determinante en la relación simbiótica con las leguminosas, sino que tiene características únicas que otros plásmidos simbióticos no poseen.

El presente trabajo tiene como interés identificar cúal es el replicador básico del pSym de *R. tropici* CFN299, ya que estudios realizados previamente indican que posee un gen del tipo *repC* (Burgos *et al.*, 1996), pero al secuenciar parcialmente un fragmento de 390 kb del pSym (Hernández-Lucas *et al.*, 2002) se identificaron secuencias que indican la presencia de 2 replicadores, uno del tipo *repABC* y otro de

tipo *repC*. Para resolver nuestra pregunta se aislaron ambos replicadores, se clonaron y se evaluó su funcionalidad en *A. tumefaciens*.

III. OBJETIVOS

- Aislar el origen de replicación básico del plásmido simbiótico (pSym) de *Rhizobium tropici* CFN299.
- Aislar, clonar y secuenciar los orígenes de replicación, (operón *repABC* y el gen *repC*) del pSym de la cepa *R. tropici* CFN299.
- Determinar la funcionalidad de cada uno de los replicadores.

IV. MATERIAL Y MÉTODOS

4.1. Cepas y plásmidos utilizados

Las cepas y los plásmidos empleados en este trabajo se enlistan en la tabla 4.1.

Tabla 4.1 Cepas bacterianas y plásmidos						
Cepa o plásmido	Características relevantes	Referencia				
Cepas						
R. tropici						
CFN299	Cepa Silvestre (Nal ^r)	Martínez-Romero <i>et al.</i> , 1991				
CFN299-10	CFN299 Δ pSym	Martínez-Romero y Rosenblueth, 1990				
RH271	R. tropici CFN299 + A780 integrado en pSym	Este trabajo				
A. tumefaciens						
GMI 9023	C-8 curada de plásmidos nativos (Rif')	Rosenberg y Huguet, 1984				
E. coli						
DH5a	F ⁻ Φ80 <i>dlac</i> Z∆M15 <i>end</i> A1 <i>rec</i> A1	Gibco BRL				
	hdsR17(r _k -m _k) thi-1supE44λ ⁻ gyrA96					
HB101	SupE44hsdS20 (rb ^m b) recA13 ara-14	Gibco BRL				
	proA2/acY1ga/K2 rpsL20 xyl-5 mtl-1					
Plásmidos						
pCR2.1	Vector de clonación para productos de PCR, Km ^r Amp ^r	Invitrogen				
pRK 2013	Replicón ColE1 con región de transferencia RK2 Km ^r Nm ^r	Figurski y Helinski, 1979				
pUX19	Derivativo de pUK21 con <i>oriT</i> de pSUP202. Km ^r	Lies y Roberts [*]				
	Vector suicida para A. tumefaciens y R. tropici					
pBAC A303	390 kb pSym <i>R. tropici</i> CFN299 en pBAC Cm ^r sp ^R Km ^R	Hernández-Lucas <i>et al.</i> , 2002				
A610	pUX19 con fragmento <i>Sal1</i> y <i>HindIII</i> de 1807 pb conteniendo <i>repC</i> y <i>repB</i> incompleto, Km ^r	Este trabajo				
A627	pUX19 con fragmento <i>BamHI</i> de 7 kb conteniendo operón <i>renABC</i> , derivativo de pBAC A303 Km ^r	Este trabajo				
A763	pCR2.1 con PCR de 2591 pb conteniendo el gen	Este trabajo				
A766	pUX19 con fragmento <i>Pst</i> I- <i>BamHI</i> de 6 kb conteniendo <i>repABC</i> sin 368 pb iniciales de <i>repA</i> , derivativo de pBAC A303 Km ^r	Este trabajo				
A780	pUX19 con 2591 pb, conteniendo <i>repC</i> sin codón de paro, derivativo de A763, Km ^r	Este trabajo				
A789	pUX19 conteniendo gen $repC$ completo, clonado in vivo, derivativo de RH271 Km ^r	Este trabajo				
A802	pUX19 con fragmento <i>BamHI-MIu</i> I de 5.5 kb conteniendo el operón <i>repABC</i> , derivativo de pBAC A303 Km ^r	Este trabajo				

Datos sin publicar

4.2. Medios y condiciones de cultivo

Las cepas de *R. tropici* CFN299 fueron crecidas en medio PY (Noel *et. al.*, 1984) a 30 °C, las cepas de *E. coli* y *A. tumefaciens* en medio LB (Sambrook *et. al.*, 1989) a 37 °C. Para medios sólidos se agregó 15 g de agar por litro. Cuando fue necesario los medios fueron suplementados con antibióticos (Tabla 4.2).

Antibiótico	Solución concentrada (mg/mL)	solvente	Concentración para <i>E. coli.</i> (µg/mL)	Concentración para <i>R. tropici.</i> (μg/mL)	Concentración para A. tumefaciens. (µg/mL)
Ampicilina (Amp)	100	H ₂ O	100		
Gentamicina (Gm)	30	H_2O	20	20	
Estreptomicina (sm)	50	H ₂ O	25	50	
Espectinomicina (Sp)	50	H ₂ O	25	50	
Ácido nalidixico (Nal)	30	NaOH 0.1M	20	20	
Tetraciclina (Tc)	10	CH ₃ CH ₂ OH al 50%	10	10	
Rifampicina (Rif)	10	CH ₃ OH	10		50
Kanamicina (Km)	50	H ₂ O	40	40	

Tabla 4.2. Antibióticos utilizados para la selección de cepas

4.3. Transformación bacteriana

Las construcciones obtenidas (vectores con genes clonados) se transformaron en células competentes *E. coli* DH5α y *E. coli* HB101 (Gibco BRL) (ver apéndice II).

4.4. Conjugación bacteriana

Para los experimentos de conjugación se utilizó como ayudador el plásmido pRK2013 (Figurski y Helinski, 1979) y *A. tumefaciens* GMI 9023 (Rosenberg y Huguet, 1984) como receptora de los replicadores. Como control positivo se utilizó el gen *repC* del plásmido pRmeGR4a clonado en pUX19 y como control negativo el plásmido pUX19 (ver apéndice II).

4.5. Purificación de DNA

El DNA total (cromosómico y plasmídico) de las cepas se aisló mediante un estuche comercial (Amersham Biosciences, GenomicPrep cells and tissue DNA isolation) y los plásmidos fueron purificados con un estuche High pure plasmid isolation (Roche Applied Science, Germany).

4.6. Southern blot

Para la hibridación, los DNAs fueron purificados y digeridos con enzimas de restricción, después fueron transferidos de geles de agarosa a membranas de nylon. Las sondas fueron marcadas con ³²P. El experimento se realizó bajo condiciones de alta estringencia (Southern, 1975; Sambrook *et al.* 1989) (ver apéndice II).

4.7. Amplificación por PCR y clonación

Los oligonucleótidos usados en este trabajo se diseñaron usando el programa OLIGO 2.1 y fueron sintetizados por la unidad de síntesis del Instituto de Biotecnología, UNAM (tabla 4.3.). Para las reacciones de PCR se utilizó el estuche GeneAmp XL PCR (Roche Applied Biosystems) y se amplificaron en un termociclador GeneAmp PCR system 9700 (Roche Applied Biosystems). Como vectores de clonación se utilizaron los plásmidos pUX19 (Lies y Roberts, datos sin publicar) y pCR 2.1 (invitrogen) para clonar los PCRs.

Tabla 4.3 Oligonucleótidos usados en este trabajo							
Oligonucleótido	Secuencia	Tm					
repABC							
Upper Primer REP1	[5'-GCG GAA AAT TAC ATC TCG ATC A-3']	62°C					
Lower Primer REP2	[5´-TAG AGG CAA TTA GGG TAC GTG G-3´]	66°C					
Lower Primer REPABC4	[5´-AAT GGC TGG GCA GTT GTA GAA T-3´]	64°C					
repC							
Upper Primer REP4	[5'-AGC AAA TGA AAA ATA TCG TGT A-3']	56°C					
Lower Primer REP5	[5'-ATC CGC CAG ACC GAA GAC CGA G-3']	72°C					
Upper Primer REPC7	[5'-TTC TGG AAA AAG CAA ATC GG-3']	56°C					

4.8. Perfil de plásmidos

Para visualizar los plásmidos de alto peso molecular se realizaron geles de agarosa modificados de Eckhardt (Hynes y McGregor, 1990).

4.9. Secuenciación

Las secuencias se obtuvieron en un secuenciador automático (Perkin Elmer/Applied Biosystems 377-18), utilizando los oligonucleótidos universal y reverso (m13/pUC -40 FORWARD 5'-GTT TTC CCA GTC ACG TTG TA-3' y m13/pUC REVERSE 5'-TTG TGA GCG GAT AAC AAT TTC-3').

4.10. Análisis de Secuencias

Las secuencias se analizaron con el programa BLAST del Nacional Center for Biotecnology Information (NCBI) (<u>http://www.ncbi.nlm.nih.gov/blast/Blast.cqi</u>).

V. RESULTADOS

Los experimentos previos revelan que el pSym de *R. tropici* CFN299 posee un replicador *repC* (Burgos *et al.*, 1996); sin embargo, los resultados en nuestro laboratorio revelan la presencia de 2 replicadores, uno *repC* y otro *repABC*. Es de nuestro interes determinar la funcionalidad de cada uno de estos replicadores; para establecer cúal es el replicador funcional en el pSym de *R. tropici* CFN299 por lo que se aislaron y evaluaron ambos replicadores.

V. I. Caracterización del gen repC

5.1.1. Aislamiento del replicador tipo *repC*

El análisis de secuencia de un fragmento de 390 kb del pSym de *R. tropici* CFN299 reveló la presencia de un fragmento de DNA, homólogo al gen *repC* del plásmido pRmeGR4a de *S. meliloti* (Mercado-Blanco y Olivares, 1994; Izquierdo *et al.*, 2005). Basándose en la secuencia analizada de *R. tropici* CFN299 se diseñaron oligonucleótidos y se amplificó parte del gen *repC* de esta cepa por PCR, usando como templado DNA del pBAC de la cepa A303 (390 kb del pSym de *R. tropici* CFN299) y los oligonucleótidos REP4 y REP5. Se obtuvo un fragmento de 2591 pares de bases (pb) (Fig. 5.1).



Figura 5.1. Gel de agarosa mostrando la amplificación por PCR del gen *repC* del pSym de *R. tropici* CFN299, usando como templado DNA del plásmido A303. Se usaron los oligonucleótidos REP4 y REP5. Carril 1 se observa el marcador de DNA 1Kb plus DNA ladder y carril 2 muestra el tamaño del replicador amplificado de un tamaño aproximado de 2591 pb. El Tamaño aproximado se obtuvo mediante un analisis de secuencia

5.1.2. Clonación de la secuencia amplificada *repC* en el plásmido pCR 2.1

El plásmido pCR 2.1 tiene resistencia a kanamicina (Km) y Ampicilina (Amp) con selección positiva (lacZ), esto permite la generación de clonas blancas y azules. Cuando se interrumpe el gen lacZ como resultado se obtienen clonas blancas que sugiere la integración en el plásmido del gen de interés. Se utilizó este plásmido debido a que no se requieren sitios de restricción específicos para clonar fragmentos sino que el vector posee extremos de timina, los cuales se unieron a los extremos de adenina del PCR repC que fueron sintetizados por acción de la Taq polimerasa. Se seleccionaron las clonas blancas las cuales interrumpieron el gen lacZ. Así se obtuvo la construcción A763 (pCR 2.1- repC). (Fig. 5.2 A). Se hicieron digestiones con enzimas de restricción de la construcción para verificar que repC estuviera clonado (Fig. 5.2 B).



Figura 5.2. A) Construcción A763, obtenida a partir del PCR *repC* clonado en el plásmido pCR 2.1. **B)** Gel de agarosa mostrando en el carril 1 marcador 1Kb plus DNA ladder, carril 2, digestión de A763 con *Bam*HI y *Xba*I, se observan 2 fragmentos, uno del plásmido pCR 2.1 de 3.9 kb y otro del *repC* de 2591.

5.1.3. Clonación del gen repC en el plásmido pUX19

El vector pUX19 es derivado de pUK21 (Viera y Messing, 1991), con un replicador *oriV* (origen de replicación) derivado del plásmido pMB1 y un *oriT* (origen de transferencia conjugativa) de pSUP202 (Simon *et al.*, 1983). Como marcador de selección posee resistencia a Km y selección positiva (*lacZ*), plásmido construido por Lies y Roberts (Datos sin publicar). Es un vector incapaz de replicarse en *A. tumefaciens* y *R. tropici*, por esta razón se utilizó pUX19 para clonar los replicadores del pSym, ya que al ser transferido a *A. tumefaciens* sólo tendremos replicación del plásmido si el origen de replicación es funcional.

La secuencia repC contenida en la construcción A763 (pCR 2.1-repC) se clonó en el plásmido pUX19. Se obtuvo la construcción A780 (pUX19-repC) (Fig. 5.3) que se transformó a DH5 α para posteriormente realizar cruzas con *A. tumefaciens*.



Figura 5.3. La construcción A763 se digirió con las enzimas de restricción *Bam*HI y *Xba*I separando así *repC*, que se ligó en el vector de clonación pUX19 digerido también con *Bam*HI y *Xba*I , obteniendo la construcción A780 pUX19- *repC* con un tamaño aproximado de 6458 pb.

El análisis de secuencia de la clona A780 reveló la presencia de 2591 pb del DNA del pSym de *R. tropici* CFN299, el cual codifica para el gen repC. Sin embargo esta construcción no posee los 14 aminoácidos del extremo carboxilo terminal de repC (Fig. 5.4).



Figura 5.4. Fragmento *XbaI-Bam*HI clonado en pUX19 para obtener la construcción A780. La secuencia *repC* clonada codifica para 418 aminoácidos, en gris se muestran los 14 aminoácidos faltantes, se esquematiza el codón de inicio (ATG) y término (TAA) además de 1337 pb del promotor probable.

5.1.4. Transferencia de A780 (pUX19-repC) a A. tumefaciens

El plásmido de la cepa A780 en *E. coli* DH5 α (cepa donadora) se transfirió a *A. tumefaciens* (cepa receptora) en presencia de pRK2013 como ayudador. Este experimento se realizó para verificar si la secuencia *repC* clonada es capaz de replicarse autónomamente. No se obtuvieron transconjugantes, por lo tanto la secuencia *repC* que codifica para 418 aminoácidos de 432 del gen *repC* no se replica en *A. tumefaciens*.

5.1.5. Clonación en vivo del gen repC

Debido al resultado anterior, en el cual se muestra que repC es incapaz de replicarse en *A. tumefaciens*, decidimos obtener el gen repC completo, desde la región promotora hasta el codón de paro y así probar su funcionalidad. Para tal experimento se purificó el plásmido de la cepa A780 contenido en *E. coli* DH5 α y se transformó a *E. coli* HB101 para obtener la cepa sensible a ácido nalidixico y resistente a Km. Esta cepa se conjugó con *R. tropici* CFN299 mediante el ayudador pRK2013. A las transconjugantes obtenidas RH271 (*R. tropici* CFN299+pUX19-*repC* integrado en el pSym) resistentes a ácido nalidixico y Km se les purificó el DNA.

Teóricamente el plásmido pUX19-*repC* recombinó en el pSym de *R. tropici* CFN299 (Fig. 5.5 A y B), un análisis de restricción de pUX19-*repC* reveló 4 sitios únicos de restricción, *Bam*HI, *Nde*I, *Sac*I y *Nco*I (Fig. 5.5 B), por lo que supusimos que el gen *repC* del pSym de *R. tropici* cepa silvestre, tampoco podría ser cortado por ninguna de las 4 enzimas antes mencionadas.

Para obtener el gen repC de *R. tropici* CFN299 cepa silvestre se digirió el DNA total de la cepa RH271 (*R. tropici* CFN299 + pUX19-repC integrado en el pSym) con la enzima *Bam*HI (Fig. 5.5 C), posteriormente se religó y se transformó en células competentes *E. coli* DH5 α ; se obtuvieron 12 colonias positivas resistentes a Km. El análisis de restricción de una de estas clonas reveló que se clonó *in vivo* el gen repC completo del pSym de *R. tropici* CFN299 cepa silvestre en un fragmento *Bam*HI de aproximadamente 10 kb, este fragmento incluye el plásmido pUX19 junto con el gen repC, cepa A789 (Fig. 5.5 D). Se obtuvo la secuencia del gen repC mediante la secuenciación de esta clona y con las secuencias parciales anteriormente obtenidas (Fig. 5.6).

En el siguiente esquema se representa la metodología para la obtención del gen *repC* in vivo:



D)



Figura 5.5.

A) Construcción A780 la cual posee el gen repC clonado sin 14 a.a en el extremo 3'. B) Integración del plásmido pUX19-repC en el pSym de *R. tropici* cepa silvestre mediante recombinación homóloga, la recombinación originó la cepa RH271. Se muestran los sitios únicos de restricción de pUX19 que no cortan ninguno de los 2 repC. C) DNA de RH271 digerido con *BamH*I originó un fragmento que contiene el plásmido pUX19 con su origen de transferencia (*ori T*), su origen de replicación (*oriV*), la resistencia a Km y el gen repC completo del pSym de *R. tropici* CFN299 cepa silvestre dejando fuera el gen repC anteriormente clonado que le faltaba el codón de paro. D) La ligación del DNA dio como resultado la clonación *in vivo* del gen repC cepa silvestre en el plásmido pUX19, construcción A789.

5.1.6. Obtención y análisis de la secuencia del gen repC

La clona A789 fue secuenciada y junto con la secuencia del contig obtenido anteriormente, se obtuvo una secuencia sobrelapada con el gen *repC* completo (Fig. 5.6).

AGATCGGCTC CTCCAGATCG TGTTTTTTTG AGGTCACAAG TCGGGACGGG CGATCGGGTT ATCCAGCGCT CTCCACGCGT TACGCTTACT 91 GAAAATTGCA TACTCAGCTC AAAAAATGAG TCGACTGACA TAATGCTTTC CATTTTTTTG CCGAAAAAAAC GCAAAATATA TTTGGCTCAC 181 CTTCAAGCGT TCTCACTAAA GGTATCCTAT GAAATACCAG CTCGATTCAA CTGATTTCAG TACAAATCTT TAGAATTTCA GCCGACCCGG 271 ANAGCTGTCC AGATTGCGAA ACACGTGCCA AATCGGCGCG TACGGCAATT TACAACGCCC TATGAGAACA GCATATTGCG TCGCGCCGCC 361 GCTATCTATG GTTGGTATCA CAGCTCTCCC TTCTCCCTCA TCTCTGTGGG CACGAGATCG CTTCAAATAG GAGGCTAAAG CAAATGAAAA 451 ATATCGTGTA CGTGTTAGAG TCGGATACCG CCCGTTTTGG GGCGGTTGAG CTTCAGGATG GAAGCATTGC TCCGATCTAT GACTTGATTG 541 GTAGCAGGCT AATCGAGATG ATCCGCTTTG ACGATATGCA CTCGCTGTTC GTAGATGAAG AAGCGTTGCG TGATGGGCTA ACCGCCTTTA 631 CAATCTTTGA TGGCTGTCCA AAGCCTCTGG CGGGGAAAAT TGTCCTTGTG GGAGGCGACT GCGGACAACC TTATCGTGCC CCATTGATCA 721 CGATTGCGGA TGCGGCGGCA CGCTTCCGAT GTTGCCGGCC GGTGCTCGAT CCAATCTTTG CCACGGCCGG CGATATGACG CCAAAGGGTC 811 TTATAGTTGC CGGGGCACTG GCAAGCTTGC AAGTCCGCAT TGATCGTCGC TCTCCCGTGC TCGAAGAGAG CGAAAATCGC GGTTTGATAC 901 AATAAATCCG ACCGCCACGC TCGCGAAAGA TCGCGCGGGC ATGGCGGCAA AAGCTTAGGT ATTCGGTTTC AACTCACTCA CTCATTTGCC 991 AGGCGGCAAC CCCTTAATGA ATTTGCTGCG GTGGCGCTGC CGCTAAATAG GCTGGCAGAT CTCAGGGAAC TCCGGAAAAC GCGCGCACCG 1081 TTTGCAGGTC TGAGATGGCG ATCCAACGCA CCTCGGGTAA TCATTGCTCC AGGAGATCAG ATCTCAAAAC CCAAAATGAG GGGATGGAGT 1171 TTACGTCGAC CACGCACGGC ACAATCCTCC CGCCTCTTGT GATAGCCCCC GGGTCGATGA TGACCAGCTG GGAAAGCGGG CTGTCGCTAT 1261 TCATCCAGCC ATCCACGTTT CGCACGCTGA TCCTCAACGT CAAGTACGGG AGGAGCATTG CGTTTTCGGT CAAGCGTCAT TGGCTCCATG 1351 TCCGAGCATC GGCCAATTTG AGAATTTGT CCGTAAGGCG AACCAAATTT GATCGGTCGA AACGGTATCG AAATGGCATT TAGACTTTCA 1441 AAACAATTAA TAAAATCAAT GCGTTAGTTT AGTCACTATC CCGCTGCGAC CTCAAGTAGC AGCTCAAACT CGTACATATC TTGCGCCAAA 1531 TTGATTTGTT GGGATATGCT TTTCGCGTAT CGACCGCAAA CATCCCGACG CGTGCGTCGG GTCTGAAATG GAGAGAAAAA ACCGGCCAGA 1621 AAAAGAAAAA CCGCTCCCGA AGCAATCCGG AAGCGGGTCT GTAAGAGAAT AAATTAGCCA AAACCTTTGT CGCAGACCCT TCCTTCGTTG 1711 TCAACAAAAA ACATCACTTT GGTGAACGGC GGAGCTTTGC CTGGTCCCGA ACGGAGACGA GGACATGGAA GATATGACCA CCTGCCAACG мЕ **D M T T** CQR 1801 GCCGGTCGGC CGTAAGATAA CGCCCCAAAG GGCTGAATTC CGGCGCTGGG CAAAATCGGC AGAAATCGGA ATGGTGACGC GTGCCCAATT PVG RKIT PQR AEF RRWA KSA EIG MVTR AQL 1891 GECTGTGCTT GCTCAAAAATC TGACCTGCAC AGGCCTGATA AATGGAACGG AAAGCCATCT GTTGCTAATA TTGATCAACA CCGCCCCCGC A V L A O N L T C T G L I N G T E SHL LLI LINT APA 1981 CGATACTTTC GATACGTCGG GGCGGCCAAT CGTGTTCAAA TCGAACCGAC AATTGGCATT TGAGGTCGGG CGTTCGGCCG GCCGGGTCAG DTF DTSG RPI VFK SNRQ LAF EVG RSAG R V S 2071 TCGAATGCTT TCCAAGCTTT TCGATCTGGG TTTGATCACG ATGCAGGACA GTGGAAACTA TAAGCGATAT CCCGTCAGAA ACAGCGATGG DLG LIT MQDS GNY RML SKLF KRY P VRN S D 2161 CGTCATAACG GATGGATGCG GGGTCGATCT TCGAATATTG GTCCTGCGAT ATGGCGAGCT TACTCGCCTC GTAAAAGACA TCCGAGCGGA V D L R I L VLRY GEL TRL VIT DGCG νκρι RAE 2251 GAAGAGGGCG ACCGAAGCTG CCTTGCGACG GTACCGAGGC GCCTTGCGCA ATATCCGTTG TGCCCTCGCT GCTGTCGCTG ATGTAAGTGC K R A T E A A L R R Y R G A L R N I R C A L A A V A D VSA 2341 AACAATGCGC CAACGCATTG CCGACCGGGT CGAGCGCATC AGCAATTATA TCGGAATTGC TGCCAAAGCA TCCAACCGGA TGCTGCGTCG TMR Q R I A D R V E R I S N Y I GIAA KΑ SNRM LRR 2431 TGCGACTGTC CTTCTTGAAT GGTCTATGGA CCGACTGATG CGGCTTGACC CGCGACCTCA AGACCAACTC GGCGCAGGGA AAACGACATG LLEW SMD R L M R L D P а т v R P Q D Q L GAGK ттс 2521 CGTGCCTGTC GAAAACGACA TGCAGAAACA GACTACACCT CCTTATCAAT CAAAACGTAG TAGTAATGTT CGAGAGGGTT CGGCTCATGT ENDM **ΟΚΟ ΤΤΡ** PYQS KRS S N V REGS АНV 2611 CGGACAATAC AGTTTTCCAA AAGCCAGCGC CGCCAGCAGC TCGGCTTTTG AAGAAACCGC GACCCTTTGG CGAGGCTCAG TCAGCGAGCG GQY SFPK ASA ASS SAFE ЕТА TLW RGSV SER 2701 AGAACCTCAG CGGCAGGCTC TAATCGCGCT TAAGGATGTC TTGCGGGCGG CCCCGGCCCT CGCGGAGCTT GGATGCACTC CCAAGAGCTT E P Q R Q A L I A L K D V LRAA PALAEL GCTP KSF 2791 CCAGGAGTTT GTTCGGCTGA CTCCGATGCT GTGCCGCTTC GTCGGGATCT CAGAGGATGC CCGCCGTGGC GCCGTGGAGG CAATGGGCGA V G I S E D A R R G A V E A M G E VRLT PML CRF QEF 2881 ACAGCAAGCA GCGGTGGCGG TGGCAGTTAT TCTGGAAAAA GCAAATCGGC AGGAGGTCAG TTCGCCCGGC GGATATCTGC GCGCAATGGC O O A A V A V A V I L E K A N R O E V S S P G GYLR AMA 2971 CGAGCGCGCA GCCGCGGGCA GATTGTATCT TACCCGCTCG GTCTTCGGTC TGGCGGCTCG CCGCGACCTT GAGCAAAGCA CTCGAATGGC ERA AAGR LYL TRS VFGL A A R R D L E Q S T 3061 CGGGCAAACT TAACTGGTGG GTCACGCTCT TAAGCGGTGC AGCGTTCTAG TCGACACTGT TCATGCGCCA CGATCAAAGT GAGCGACGAT G Q T 3151 GTCGCGCGAT GCGGTGACCG ACCGGAAAAC GAGTGAGTGT GAGGTCCAGG ATTCTATAGT TGTGGCGCGTG GTTCCTTTAA AAAGATGATC 3241 ATGAAAACCA CTTTGCGGCG AGATCGCGGT GAAGACCGAC GCTTTCCGTT GCCGGAGTGC CTACCAGACA GCCAAGCTTA CCTGCGATTG 3331 ATGTCTCATT AGCTCCCGTT ATCTTGAGGC ACCATACGCT TCAGCAGCCG AACGAGATCG CGCTTTGCCT CGCTCGGAAG GGTTCTGATA 3421 ATTCGGTCCA GAGTGAGACG ATCCTCGGCT TCCTCAGGCG TGCGCCCCCCA AAGATGGGGC GCCGCTTCAA AGAGCATATC TATGGGCATA 3511 AAGCCGAGAA TTTCACAGAG ATGGATCATT CGGGTGACCG TCATCCTTGA AAACGCCCGC TCGTAACGTC CGTACACGGG TTGTCGACAG 3601 ACCGAGCATT GGCGCCAAGAT CCGCGCGAGC GAGGCCCTCG GCCTCGCGCG TTTTTTTAA AAATGCGCTG GATCAGTTCT TCCATATCTC 3691 CTAGCGTCGC AACTTTCCCG TCAAAACCGG GCTTGCGGTA GATCGGCTTG TCGACTTCCG GATCGGTGGG TATCTACGAG G

Figura 5.6. Secuencia nucleotídica de 3771 pb la cual incluye *repC*, en rojo se muestra el codón de inicio y el codón de paro. La secuencia sombreada corresponde al gen que codifica para un ctRNA.

La secuencia obtenida de 3771 pb se comparó en el genBank, utilizando el programa blastx (<u>http://www.ncbi.nlm.nih.gov/Blast.cqi</u>). Se identificó el gen *repC* de *R. tropici* CFN299 el cual posee una identidad del 63% a nivel de aminoácidos con el origen de replicación *repC* del plásmido pRmeGR4a de *S. meliloti* GR4 (Burgos *et al.* 1996). Se localizó un ORF que corresponde a un gen *psiB* el cual tiene una identidad del 69% con el de *R. etli* CIAT652. También se identificó una región intergénica (*inc1*) como la localizada en el plásmido pRmeGR4a de *S. meliloti* (Izquierdo *et al.*, 2005) (Fig. 5.7).



Figura 5.7. Mapa físico de la secuencia obtenida, se muestra un gen psiB en un ORF de la base 444 a la 903, posteriormente se muestra la probable región promotora de repC entre 903 y 1774 pb, entre esta secuencia en rojo se muestra el gen que codifica para el ctRNA. Finalmente se muestra el gen repC en la secuencia de 1775 a 3071.

5.1.7. Transferencia del gen repC completo a A. tumefaciens

Al obtener la región regulatoria, así como la proteína RepC completa (A789) se transfirió a *A. tumefaciens* en presencia de pRK2013 para verificar la funcionalidad del gen *repC*. No se obtuvieron transconjugantes *repC*, del pSym de *R. tropici* CFN299 no se replica en *A. tumefaciens*.

V. II. Caracterización del operón repABC

5.2.1. Aislamiento de *repBC*

El análisis de secuencia de un fragmento de 390 kb del pSym de *R. tropici* CFN299, contig 177 (Hernández-Lucas *et al* 2002), revela la presencia de un fragmento de DNA (3450 pb) el cual contiene 2 ORFs con homología a los genes *repB* y *repC* de otros plásmidos que contienen un operón *repABC*. Tomando en cuenta que *repB* y *repC* se encuentran contiguos se diseñaron los oligonucleótidos REP1 y REP2 (tabla 4.3). Se amplificó su secuencia nucleotídica por PCR, utilizando como templado DNA del pBAC de la cepa A303 y se obtuvo un fragmento de 3249 pb (Fig. 5.8).



Figura 5.8. A) Gel de agarosa mostrando en carril 1 el tamaño del producto de PCR *repBC* (3249 pb), carril 2 muestra control negativo (reacción para PCR sin templado) y carril 3 marcador de peso molecular 1Kb plus DNA ladder. B) Mapa físico del PCR, se muestra que el fragmento amplificado contiene los genes *repB* y *repC* con los sitios de restricción *Sal* I y *Hind* III.

5.2.2. Clonación de *repBC* en el plásmido pUX19

El producto de PCR de 3249 pb se digirió con las enzimas de restricción *Sal*I y *Hind*III, se obtuvo un fragmento de 1807 pb (Fig. 5.9 A), que contiene parte del gen *repB* y el gen *repC* completo. Este fragmento se clonó en el plásmido pUX19 (Fig. 5.9 B). El plásmido obtenido (construcción pUX19-*repBC*) al cual le falta un fragmento de *repB* se transformó en *E. coli* DH5 α obteniendo así la construcción A610 (Fig. 5.9 C).



Figura 5.9. A) Gel de agarosa mostrando en carril 1 marcador de peso molecular 1Kb plus DNA ladder, carril 2 se muestra el producto del PCR de *repBC* (3249 pb) y en carril 3 se observa el mismo producto digerido con las enzimas *Sal* I y *Hind* III, se obtuvieron 3 fragmentos de aproximadamente 1807, de 816 y de 626 pb. **B)** Se muestra la forma en que fue clonado el producto de PCR digerido con *Sal* I y *Hind* III en el plásmido pUX19, **C)** construcción A610.

5.2.3. Transferencia de A610 (pUX19-repBC) a A. tumefaciens

Se transfirió A610 a *A. tumefaciens* utilizando como ayudador el plásmido pRK2013. No se obtuvieron transconjugantes, esto indica que *repBC* no se replica autónomamente.

5.2.4. Detección del operón *repABC* por Southern blot

Tomando en cuenta el resultado en el cual la construcción A610 es incapaz de replicarse en *A. tumefaciens* decidimos obtener los 3 genes completos: *repA*, *repB* y *repC*. Para localizar en el pSym el operón *repABC* se realizó un analisis de restricción del contig 177 que contiene la secuencia de los genes *repC* y *repB*. El análisis reveló que la secuencia no posee sitios de restricción para las enzimas *Bam*HI y *Pst*I. Debido a lo anterior se realizó un Southern blot, utilizando el producto de PCR repBC de 3249 pb (Fig. 5.8) como sonda para hibridar con el DNA del pBAC de la cepa A303, digerido con *Bam*HI y *Pst*I. Los resultados del Southern blot, mostraron que la sonda repBC hibridó en un fragmento *Pst*I de aproximadamente 10 kb y en una banda *Bam*HI de 7 kb aproximadamente (Fig. 5.10).





5.2.5. Obtención y clonación del operón repABC

En el experimento anterior, la banda *Bam*HI de 7 kb del Southern blot nos indica que el operón posiblemente no es cortado por la enzima BamHI y se ubica dentro de este tamaño. Tomando en cuenta este resultado se hicieron restricciones del pBAC de la cepa A303 con la enzima *Bam*HI, los fragmentos obtenidos se ligaron en el plásmido pUX19 digerido con BamHI, estas ligaciones se transformaron en E. coli DH5a. Se obtuvieron 500 clonas positivas lacZ con las cuales se hicieron 5 grupos de 100 clonas cada uno, los grupos de las clonas se crecieron en medio líquido para realizar 5 cruzas a A. tumefaciens, utilizando como ayudador el plásmido pRK2013. De las 5 cruzas, solo en 3 se obtuvieron transconjugantes en placas de medio selectivo, siendo en total 43 transconjugantes en la dilución 10⁻³, esto equivalente a 4.3 X 10⁴, teniendo en promedio 1.43 X 10⁴ unidades formadoras de colonias (UFC) por mililitro de muestra original en cada una de las 3 cruzas. Al obtener transconjugantes es debido a que obtuvimos el replicador repABC funcional clonado en pUX19 en un fragmento BamHI de 7 kb. Para verificar si las clonas obtuvieron el plásmido se tomaron 30 transconjugantes con las que se realizó un experimento tipo Eckhardt. El experimento mostró que A. tumefaciens adquirió un plásmido, lo cual significa que el operón *repABC* se replica autónomamente.

5.2.6. Recuperación del plásmido (pUX19-repABC) de A. tumefaciens

Para obtener de *A. tumefaciens* el plásmido pUX19 con el operón *repABC* clonado, se conjugó esta cepa a *E. coli*. Se obtuvieron 138 transconjugantes resistentes a Km lo que indica que tenemos pUX19 con *repABC*. Se tomaron 4 transconjugantes, se les purificó el plásmido y se digirieron con *Bam*HI. En los 4 plásmidos se identificó el fragmento esperado *Bam*HI de 7 kb conteniendo el operón *repABC*. También se identificó otro fragmento *Bam*HI clonado de aproximadamente 2 kb (Fig. 5.11). Se tomó uno de estos plásmidos y se le secuenció el fragmento clonado para asegurarnos que tuviera el operón *repABC*, el plásmido resultó positivo y se le nombró A627 (pUX19-*repABC* en fragmento *Bam*HI de 7 kb) (Fig. 5.14 A).



Figura 5.11. Gel de agarosa mostrando digestión de 4 plásmidos recuperados de A. tumefaciens con la enzima BamHI. Carril 1 marcador de peso molecular 1Kb plus DNA ladder, carriles del 2 al 5 se observa el fragmento aproximado de 7 kb que contiene repABC. En las 4 diaestiones aparece el plásmido pux19 (3.8 kb). En todas las digestiones se observó otro fragmento BamHI clonado de 2 kb.

5.2.7. Región mínima replicable del operón *repABC*

Para obtener la región mínima replicable se realizaron 2 experimentos con el plásmido de la cepa A627.

En el primer experimento se digirió la construcción A627 con *Pst*I y *Bam*HI ya que mediante análisis de secuencia se observó que ambas enzimas generan un fragmento aproximado de 5.8 Kb, el cual contiene el gen *repA* interrumpido por *Pst*I en la región río arriba y los genes *repB* y *repC* completos (Fig. 5.14 B). Este fragmento con extremos *Pst*I-*Bam*HI se clonó en pUX19. Posteriormente el plásmido obtenido se transformó en *E. coli* DH5 α , se purificó y se obtuvo la construcción A766 la cual se verificó por restricción (Fig. 5.12).



Figura 5.12. Gel de agarosa que muestra en carril 1 la digestión del plásmido A627 con las enzimas *PstI-Bam*HI, se muestra el fragmento aproximado de 5.8 kb del operón y el fragmento de 3.8 kb que corresponde al plásmido pUX19.

Al obtener la construcción A766, se analizó si *repABC* con *repA* interrumpido por *Pst*I es funcional, por lo que se conjugó a *A. tumefaciens* utilizando como ayudador el plásmido pRK2013. No se obtuvieron transconjugantes, el fragmento *Pst*I-*Bam*HI (Fig. 5.14 B) clonado en pUX19 no se replica en *A. tumefaciens*.

Tomando en cuenta el resultado anterior, se realizó un segundo experimento para obtener el operón completo en el que se utilizó MluI. Al analizar la construcción A627 se observó que esta enzima no interrumpe ningún gen del operón repABC (Fig. 5.14 C). Al digerir se obtuvieron 2 fragmentos MluI, uno aproximado de 3.5 kb de la región río abajo del operón sin afectarlo y el otro aproximadamente de 9.5 kb que contiene al plásmido pUX19 y al operón repABC completo. Al religar la digestión solamente se ligó el fragmento de 9.5 kb perdiéndose el fragmento MluI de 3.5 kb. Se obtuvo así el operón repABC clonado en el plásmido pUX19 que se transformó en *E. coli* DH5 α . Posteriormente se purificó obteniéndose la construcción A802 que fue evaluada por restricción (Fig. 5.13). El análisis de secuencia de la construcción reveló que el operón se encuentra en un fragmento *Bam*HI-*Mlu*I de aproximadamente de 5.5 kb.



Figura 5.13. Gel de agarosa mostrando en carril 1 marcador de peso molecular 1Kb DNA ladder, carril 2 digestión del plásmido A627 con la enzima *Mlu*l, se obtienen 2 fragmentos *Mlu*l, uno de 3.5 kb y otro de 9.5 kb aproximadamente que contiene el plásmido pUX19 junto con el operón *repABC* completo. Se transfirió la construcción A802 a *A. tumefaciens* utilizando como ayudador el plásmido pRK2013, se obtuvieron transconjugantes en placas de medio selectivo, teniendo 3.8 X 10^4 (UFC) por mililitro de muestra original. Se obtuvo una región mínima replicable del operón *repABC* en un fragmento *Bam*HI-*Mlu*I de 5.5 kb clonado en pUX19, el operón *repABC* se replica autónomamente y es funcional en el pSym de *R. tropici* CFN299.



Figura 5.14. A) Construcción A627, se observa el operón *repABC* en un fragmento *Bam*HI de 7kb, adyacente a él, también hay otro fragmento *Bam*HI de 2kb, en azul secuencia de pUX19. **B)** Construcción A766 originada al digerir A627 con *Pst*I y *Bam*HI, se obtiene un fragmento aproximado de 5.8 kb, *repA* es interrumpido por *Pst*I. **C)** Construcción A802 originada al digerir A627 con *Mlu*I, al religar se obtiene un fragmento clonado de 5.5 kb el cual contiene el operón *repABC* completo.

5.2.8. Obtención y análisis de la secuencia del operón repABC

La construcción A802 se secuenció mediante oligos sobrelapados. En total se secuenciaron 5885 pb. La secuencia obtenida se comparó en el genBank utilizando el programa blastx (<u>http://www.ncbi.nlm.nih.gov/Blast.cgi</u>). Se observó una identidad con *repA* de *R. etli* CFN42 de 76%, con *repB* de *R. etli* CFN42 de 74% y con *repC* del plásmido pSmeSM11a de *Sinorhizobium meliloti* SM11 de 76%. Además adyacente a *repA* pero en dirección opuesta, se identificó una secuencia parcial del gen *traI*, con una identidad de 64% con el de *A. tumefaciens* y de 63% con el de *Sinorhizobium meliloti* (Fig. 5.15).



Figura 5.15. 4 secuencias parciales (**A**, **B**, **C** y **D**), correspondientes al operón *repABC* se sobrelaparon para obtener la secuencia completa del operón *repABC* del pSym de *R. tropici* CFN299. **E**) El fragmento clonado en pUX19 abarca del sitio *Bam*HI al sitio *Mlu*I de la secuencia.

5.2.9. Secuencia del operón repABC

1	ACGGCCGNTG	ANTTGTAATA	CGACTCACTA	TAGGGCGAAT	TGGGGATCGA	TCCACTAGTT	CTAGAGCGGC	CGCCACGGCG	ATATCGGATC	CGCTCGAATC
201	GAAGATCGGT TCCCCCAACA	TCGACCAACG	AGCTCCGNGT TTGTGTCCAC	ACCCATTIGC	GATGCACCAT	TCGATGATGC	C'I'GCGAACA'I' ≱TGGGCGTTT	GGTCAAAGTG AGGTGTCCAT	GCCTCGTGGA	TGACGAAAAG
301	ACATGGGTCA	CCATCGTCGG	ACCCATGGCC	GGCANGAGTC	TAGCGCATCC	TGCCACACGA	GCGCTGTCCG	ATACGGCGAG	TATGTAAGTC	GGCTGGAGAG
401	CGTCGAAGAC	GTCCGCTTCC	CGGCCGTCAA	GAACACCGAC	CTCCCAACCC	AACCGATTTG	AAAACACTTG	AGCTCNAAGT	CNATGATGTG	CTTCGAGAAG
501	GTGCTGTGCA	GGAAGCCCAT	GTCTGGTCGA	TATCGCAACA	ACCTTCATAC	TGATCTCCGT	CNCTTANATT	TATCGACAAA	GATCAGCACG	GATTANAATG
601	GCTGGGCAGT	TGTAGAATCC	TCAGGGATTT	TGGTGAGTCG	CGCGCGGTTT	TGGGCGACAA	CGCGCTGATC	TGATCTCGCG	TGTTATTCCC	TTAGTTGTAG
801	AGGGGGCGTCC	CGGCTATTAA	CAGTTTATTA	GCCGCATTTT	CCTTGCCGGC	CCGCGAGAAT	CGGACATAAT	AACGGTTCAG	TGGCAATATC	GGGCCGAAAA
901	CCGTTATCTG	AAAGATTTCT	GACAATGAAC	GCTAATTCGA	TGGCTTCGCT	TAAGATGCCG	ATGCACTTTG	AAGATCAAAT	CCTAGAACAG	GGTGATTTGA
1001	TCTCGAAAAA	GCTGCACCTA	CTTAGCGTAC	AACGATTCCC	GCCCAACGCG	AAAAAGACAT	TGCGAAGCTT	CTCGCTGGCG	GAGGTCGCGA	CCTATGTCGG
1101	CGTATCGCAG	AGCACATTGA	AAAAGCTGCA	TCTGGACGGA	AAGGGCCCCT	CCCCTCAAAC	ATCGTCCTCC	GGTCGTCGCT	CCTACACCGC	CGAACAGATG
1201	TTGGAACTGC	GTCAATATCT	GGGATACGCA	CGGTCGATCC	GACAACCGCA	TGTATGTTCC	TACCTCCCAT	CGGCGGTGAA	AAGCTGCACT	GCAGGTGCTT
1401	TCGACCCTCA	AGCATCCCTT	TCGTCTCTGC	ACGGCTTCCA	ACCGGAGCTC	GATCAGATCG	CGTCCATCTA	TGATGCCATT	CGCTATGATG	GTGCGAGAAA
1501	GCCGATCTCG	GAGATCATCC	AGACCACTAA	TTTTCCTGGA	TTGGACATTG	TCCCTGCAAA	CCTCGATTTG	CAGGAGTATG	AATATGACAC	GCCGCTTGCG
1601	ATGTCTGATA	AGTCTTCTAA	TGAAGGCAAA	ACGTTTTTCA	CCCGCATATC	GCGGGGCTTTG	GCTGAGGTCG	ACGATCGTTA	TGATGTCGTC	GTCGTCGACT
1701	GTCCACCGCA	GCTCGGATAT	TTGACGCTGA	CCGCACTGAC	CGCAGCTACA	AGTGTTTTGA	TAACAATCCA	TCCCGCAGAT	GCTCGACGTG	ATGTCGATGG
1801	GGTCAGTTCC	TTCTGATGCT	TGGTGGAATC	CTAAAAACCC	GATCCGTGCC	GCTGGTGCGG	GAAGTGAACC	TCTCCTGGTA	CAGATACTTG	ATTTACGCGC
2001	TTTCAGATGC	CGGAIGIICC	AAGCAAACTC	TGTACGAAGT	CGATCGGGCA	TCCATGAATA	GAGGIICAIAC	TGACCGTGCG	ATGGAATCGT	TGGATGCAGT
2101	GAATGCTGAA	ATTGCCGGCC	TCGTTCACGA	GTCATGGGGA	CGTCGATTTG	TCAGCTGACA	AAAACGAACG	TTTTATGCGG	AAAAACAAGC	AGCTAACGGC
2201	TGCTCGGAGC	AAATGAG <mark>ATG</mark>	GCACGGAAAA	ATCTGCTCGA	GGGACTTGCT	GACCTGCATG	AAGATAGCGG	CACGGCACCT	GCATATCCAA	TGCGAGGCGC
2301	AGGCAAATCT	CTCGTGCGTT	CGCTGGACGA	GCTGGCAAAG	CAGGCTGACA	AATATCTCGA	AGGCGAGGCA	ATTATCGAAC	TTGATCCGGC	ACAAATTGAA
2401	GGCTCCTTCG	TAAAGGATCG	GATATCAGAC	GACGACGGCG	AGTATCAAGA	TCTGCTAGAG	GCAATTAGGG	TATCGTGGGC	AGGATACGCC	TATTCTCGTG
2501	ACCONTRACTA	CAAAGGTUGA	CATCTCAT	CAGGTTGTCT	CCADAATTCC	ACGGGTAAGG	GCAGCAAGGG	TATCCAAACC	AAAAGTCCGG	GCTATTGTCA
2701	GGATGATCTT	GGTTACGGCC	GAGAGGTGAT	TTCCTTGGCA	TTGTCCGCGA	ATGCGGCGCGA	GGTCTCGAAG	ATGATGACGG	TGGCAGAGCG	CATCTCGGCT
2801	AACGTCATAG	AAAAAATCGG	GTCGGCTCCC	GGTGTGGGGCC	GAGAGCGCTG	GATAGAGCTT	TCACTTCTAG	TCGGAAAATT	CGCCAATAGC	GGCAAAGTGT
2901	CAGAAGTTTT	GAGCACGGAC	GGCTTCGACG	CATTGCCGTC	CGATGAACGC	TTTTCATCTC	TGTACAAGGC	GCTCAACAGA	ACAGCGCGGC	CCGTGAGGAA
3001	GGCGGAGGTT	TCGACTCAAA	AGGCGAAGTG	GCAGCCCCGG	GACAAGGCAG	TGCAGGCCGA	AATCAAGAAT	ACGGGCAAAG	CTTTCTCGCT	CTCGATGAAG
3101	GCGAAGAATG	CTGGCCGTTT	CGGAGAATAT	CTTTCCCGCA	ATTTGGATGC	GCTCTACGAG	CAATTCCTGA	AGGAGGGAAG	CAAAGGAGAC	TGAAGCCGCA
3201	AAAGAAAAAG	GCCCCCAAAC	GACCAAGTCG	IGGAAGCCCI	TCTCGTATTC	CCTAAGCAAG	ATTGAGAATC	GCATTICCAC	GAATCCCTGT	CAA GAGTCGT
220I	AACCGTCGAT	TIGGCGGGGG	GATCTCTTTT	GUUTUAATTA	AGGTGAGAGA	M Q	I G S V	TGACGACGCC T T P	F G R	R P I S
3401	CCCTTCCCA	GGTAAGAGGG	Сазатсазаз	CCCCTCACAT	GCGGCCAGGT	AAGACGGCCG	асааатссаа	CATCTTTCCC	GATECCTCCC	ACCCCACCCC
9401	LAQ	V R G	QMKT	A E M	R P G	K T A D	K W K	I F R	D A S E	A R A
3501	AGTGCTCGGC	ATTCAGGACC	GCGCCCTTGC	CGTTCTTGAT	GCGCTCTTGA	CCTTCTATCC	CGACAATGAG	CTTTCGGAAG	AGCGCGGCCT	GGTGGTGTTC
	VLG	торк	ALA	νьр	ALLT	FYP	DNE	LSEE	ксь	V V F
3601	CCATCAAATG	AACAGCTTTC	CGTTCGAGCG	CATGGTATTG	CTGGCACGAC	ATTGCGCCGC	CACCTTGCTG	CACTCGTCGA	TGCTGGACTG	ATCCAGCGCA
3601	CCATCAAATG PSNE	AACAGCTTTC Q L S	CGTTCGAGCG V R A	CATGGTATTG H G I A	CTGGCACGAC G T T	ATTGCGCCGC L R R	CACCTTGCTG H L A A	CACTCGTCGA L V D	TGCTGGACTG A G L	ATCCAGCGCA I Q R K
3601	CCATCAAATG PSNE	AACAGCTTTC Q L S	CGTTCGAGCG V R A	CATGGTATTG H G I A	CTGGCACGAC G T T	ATTGCGCCGC L R R	CACCTTGCTG H L A A	CACTCGTCGA L V D	TGCTGGACTG A G L	ATCCAGCGCA I Q R K
3601 3701	CCATCAAATG P S N E AAGACAGTGC	AACAGCTTTC Q L S TAATGGCAAA	CGTTCGAGCG V R A CGATACGCCC	CATGGTATTG H G I A GAAAGGACAA	CTGGCACGAC G T T TGGCGGCCAG	ATTGCGCCGC L R R ATCGAAGATG	CACCTTGCTG H L A A CGTTCGGCTT	CACTCGTCGA L V D CAGCCTCGCG	TGCTGGACTG A G L CCGCTGCTTG	ATCCAGCGCA I Q R K CTCGCGCCGA
3601 3701	CCATCAAATG P S N E AAGACAGTGC D S A	AACAGCTTTC Q L S TAATGGCAAA N G K	CGTTCGAGCG V R A CGATACGCCC R Y A R	CATGGTATTG H G I A GAAAGGACAA K D N	CTGGCACGAC G T T TGGCGGCCAG G G Q	ATTGCGCCGC L R R ATCGAAGATG I E D A	CACCTTGCTG H L A A CGTTCGGCTT F G F	CACTCGTCGA L V D CAGCCTCGCG S L A	TGCTGGACTG A G L CCGCTGCTTG P L L A	ATCCAGCGCA I Q R K CTCGCGCCGA R A E
3601 3701 3801	CCATCAAATG P S N E AAGACAGTGC D S A AGAGCTTGCG	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG	CACTCGTCGA L V D CAGCCTCGCG S L A ACGATCTGCC	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG
3601 3701 3801	CCATCAAATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L	CACTCGTCGA L V D CAGCCTCGCG S L A ACGATCTGCC T I C R	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V	ATCCAGCGCA I Q R K CTCGCGCCCGA R A E TCGTAAGCTG R K L
3601 3701 3801	CCATCAAATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L	CACTCGTCGA L V D CAGCCTCGCG S L A ACGATCTGCC T I C R	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG R K L
3601 3701 3801 3901	CCATCAAATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCTG	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG	CTGGCACGAC G T T TGGCGGCCAG G Q CGGCACTTCC R H F R AAAAGCTGGA	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAGCTGC	CACTCGTCGA L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG R K L CCAACCCTCA
3601 3701 3801 3901	CCATCAAATG PSNE AAGACAGTGC DSA AGAGCTTGCG ELA CTGACGGCTG LTAA	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E	CTGGCACGAC G T T TGGCGGCCAG G Q CGGCACTTCC R H F R AAAAGCTGGA K L E	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y	$\begin{array}{c c} CACCTTGCTG\\ H & L & A & A \\ \\ CGTTCGGCTT\\ F & G & F \\ \\ GGAAGCACTG\\ E & A & L \\ \\ \\ GTGCAGCTGC\\ V & Q & L & L \end{array}$	CACTCGTCGA L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K
3601 3701 3801 3901	CCATCANATG PSNE ANGACAGTGC DSA AGAGCTTGCG ELA CTGACGGCTG LTAA	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCCTCCTC	CGTTCGAGGG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AARAGCTGGA K L E	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAAGCTGC V Q L L	CACTCGTCGT L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGCCAT A G I	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG R K L CCCAACCCTCA P T L K ATLOCCATCA
3601 3701 3801 3901 4001	CCATCAAATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCTG L T A A AAGATATTGA D I E	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGCCTCC S V L	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCCTTCT G A S GACGAATGC D E M R	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGGG M L R	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGGAGTC E E I	ATTGCGCCCCC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAGCTGC V Q L L TGGAATTTCA E F O	CACTCGTCGC L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATTCT E D S	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAABABACAG O K T D	ATCCASCGCA I Q R K CTCSCCCA R A E TCGTAAGCTC R K L CCAACCCTCA P T L K ATACCASTA
3601 3701 3801 3901 4001	CCATCAAATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCTG L T A A AAGATATTGA D I E	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L	CGATTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R	CTGGCACGAC G T T TGGCGGCCAG G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L	$\begin{array}{c} \text{CACCTTGCTG}\\ \text{H} \text{L} \text{A} \text{A} \\\\ \text{CGTTCGGCTT} \\ \text{F} \text{G} \text{F} \\\\ \text{GGAAGCACTG} \\ \text{E} \text{A} \text{L} \\\\ \text{GTGCAGCTGC} \\\\ \text{V} \text{Q} \text{L} \text{L} \\\\ \text{TGGAATTCA} \\ \text{E} \text{F} \text{Q} \end{array}$	CACTCGTCGC L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATTCT E D S	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CARAARACAG Q K T D	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D
3601 3701 3801 3901 4001 4101	CCATCANATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCTG L T A A AAGATATTGA D I E TGTCCAAATT	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA	CGATTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCCTTCT G A S GACGAAATGC D E M R TACAGAATTC	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R AAAACCCGAA	CTGGCACGAC G T T TGGCGGCCAG G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I TCCAGCAATG	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC	$\begin{array}{c} \text{CACCTTGCTG}\\ \text{H} \text{L} \text{A} \text{A} \\\\ \text{CGTTCGGCTT} \\ \text{F} \text{G} \text{F} \\\\ \text{GGAAGCACTG} \\ \text{E} \text{A} \text{L} \\\\ \text{GTGCAGCTGC} \\ \text{V} \text{Q} \text{L} \text{L} \\\\ \text{GTGCAGCTGC} \\ \text{V} \text{Q} \text{L} \text{L} \\\\ \text{TGGAATTTCA} \\ \text{E} \text{F} \text{Q} \\\\ \text{TCGCTCTGAA} \end{array}$	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATTCT E D S ACAGAGCAGG	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAAACAG Q K T D GGGCGAAATC	ATCCASCGCA I Q R K CTCSCCCA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC
3601 3701 3801 3901 4001 4101	CCATCAAATG PSNE AAGACAGTGC DSA AGAGCTTGCG ELAA CTGACGCTG LTAA AAGATATTGA DIE TGCCCAAATT VQI	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R AAAACCCGAA K P E	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I TCCAGCAATG S S N E	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P	$\begin{array}{cccc} \text{CACCTTGCTG}\\ \text{H} & \text{L} & \text{A} & \text{A} \\ \text{CGTTCGGCTT}\\ \text{F} & \text{G} & \text{F} \\ \text{GGAAGCACTG}\\ \text{E} & \text{A} & \text{L} \\ \text{GTGCAGCTGC} \\ \text{V} & \text{Q} & \text{L} & \text{L} \\ \text{TGGAATTTCA} \\ \text{E} & \text{F} & \text{Q} \\ \text{TCGCTCTGAA} \\ \text{R} & \text{S} & \text{E} \end{array}$	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R T GGAAGATTCT E D S ACAGAGCAGG T E Q G	$\begin{array}{ccc} TGCTGGACTG\\ A & G & L\\ CCGCTGCTGTG\\ P & L & L\\ GTCGTGACGT\\ R & D & V\\ CCCCCGCTCC\\ P & R & S\\ CAAAAAACAG\\ Q & K & T & D\\ GGGCGAAATC\\ A & K & S \end{array}$	$I = \begin{bmatrix} Q & R \\ Q & R \end{bmatrix} K$ $CTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG$
3601 3701 3801 3901 4001 4101	CCATCANATG PSNE ANGACAGTGC DSA AGAGCTGCG ELA CTGACGGCTG LTA AGATATTGA DLE TGTCCANATT VQL	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I	CGGTCCGACGCCC V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGG G D W E GAATGTTGCG M L R AAAACCCGAA K P E	CTGGCACGAC G T T TGGCGGCCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I TCCAGCAATG S S N E	ATTGCGCCCC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTGAACC F E P	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAGCTGC V Q L L TGGAATTTCA E F Q TCGCTCTGAA R S E	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATCT E D S ACAGAGCAGG T E Q G	TGCTGGACTG A G L CCCCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAAACAG Q K T D GGGCGAAATC A K S	ATCCAGCGCA I Q R K CTCGCGCCGA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC N P D
3601 3701 3801 3901 4001 4101 4201	CCATCANATG P S N E ARGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCTG L T A A AGATATTGA D I E TGTCCAAATT V Q I ATTGGCGCAA	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCC P F D	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGATTC	CATGGTATTG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R AAAACCCGAA K P E TTTCCACTGG	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I TCCAGCAATG S S N E GAATGGCTTT M V L	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAGCTGC V Q L L TGGAATTTCA E F Q TCGCTCTGAA R S E CCTGGAGTTT F S F	CACTCGTCGC L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATTCT E D S ACAGAGCAGG T E Q G CGAATACGA C	TGCTGGACTG A G L CCCCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAAACAG Q K T D GGCGCAAATC A K S GCCGGCGGGG	ATCCAGCGCA I Q R K CTCGCCGCA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC N P D GCGGTCGGGA A V G S
3601 3701 3801 3901 4001 4101 4201	CCATCANATG P S N E AAGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCGG L T A A AAGATATTGA D I E TGTCCAAATT V Q I ATTGGCGCGAA I G R M	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCC R E P	CGATTCGAGCG V R A CGATAACGCCC R Y A R AGCAGATAGC Q I A GGGCGCCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGGCT I K A	CATGGTATTG H G I A GARAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R AAAACCCGAA K P E TTTCCACTGG F P L G	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGACATC E E I TCCAGCAATG S S N E GAATGGTCTT M V L	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C	$\begin{array}{ccc} \text{CACCTTGCTG}\\ \text{H} & \text{L} & \text{A} & \text{A} \\\\ \text{CGTTCGGCTT}\\ \text{F} & \text{G} & \text{F} \\\\ \text{GGAAGCACTG}\\ \text{E} & \text{A} & \text{L} \\\\ \text{GTGCAGCTGC} \\\\ \text{V} & \text{Q} & \text{L} & \text{L} \\\\ \text{TGGAATTCA}\\ \text{E} & \text{F} & \text{Q} \\\\ \text{TCGCTCTGAA}\\ \text{R} & \text{S} & \text{E} \\\\ \text{CCTGAGATTT}\\ \text{P} & \text{E} & \text{I} & \text{S} \end{array}$	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATTCT E D S ACAGAGCAGG T E Q G CGAATTACGG N Y G	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CANANACAG Q K T D GGGCGANATC A K S GCCGGGCGGG P G G	ATCCAGCGCA CTCCGCCGA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC N P D GCGGTCGGGA A V G S
3601 3701 3801 3901 4001 4101 4201 4301	$\begin{array}{c} \text{CCATCAAATG}\\ \textbf{P} & \textbf{S} & \textbf{N} & \textbf{E}\\\\ \text{AAGACAGTGC}\\ \textbf{D} & \textbf{S} & \textbf{A}\\\\ \text{AGAGCTTCCG}\\ \textbf{E} & \textbf{L} & \textbf{A}\\\\ \text{CTGACGGCTG}\\ \textbf{L} & \textbf{T} & \textbf{A} & \textbf{A}\\\\ \text{AAGATATTGA}\\ \textbf{D} & \textbf{I} & \textbf{E}\\\\ \text{TGTCCAAATT}\\ \textbf{V} & \textbf{Q} & \textbf{I}\\\\ \text{ATTGGGCGCAA}\\ \textbf{I} & \textbf{G} & \textbf{R} & \textbf{M}\\\\ \text{GTTGGCGCGCA}\end{array}$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A $QCGATAGAAGAI E EATCCGTCCTCS V LGGACGTCACAG R H ITGCGAGAGCCR E PTCTCATGTCA$	CGTTCGAGCC V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGCCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGGCT I K A GCCGCCGGTGG	CATGGTATIG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R AAAACCCGAA K P E TTTCCACTGG F P L G TCGTGCGATC	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCGGCGGA K L E TGAAGAGATC E E I TCCAGCAATG S S N E GAATGGTCTT M V L GATGCTGGCC	ATTGCGCCGC L R R ATCGAAGATG I E D A GCACGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA	$\begin{array}{c c} \text{CACCTTGCTG}\\ \textbf{H} \textbf{L} \textbf{A} \textbf{A} \\ \text{CGTTCGGCTT}\\ \textbf{F} \textbf{G} \textbf{F} \\ \text{GGAAGCACTG} \\ \textbf{E} \textbf{A} \textbf{L} \\ \text{GTGCAGCTGC} \\ \textbf{V} \textbf{Q} \textbf{L} \textbf{L} \\ \text{TGGAATTTCA} \\ \textbf{E} \textbf{F} \textbf{Q} \\ \text{TCGCTCTGAA} \\ \textbf{R} \textbf{S} \textbf{E} \\ \text{CCTGAGATTT} \\ \textbf{P} \textbf{E} \textbf{I} \textbf{S} \\ \text{GTGCCTACCA} \end{array}$	$\begin{array}{c c} CACTCGTCGCG\\ L & V & D\\ CAGCCTCGCG\\ S & L & A\\ ACGATCGCCT\\ T & I & C & R\\ TCGCAGGCAT\\ A & G & I\\ GGAAGATTCT\\ E & D & S\\ ACAGAGCAGGC\\ T & E & Q & G\\ CGAATTACGG\\ N & Y & G\\ GAATGCCTGT\\ \end{array}$	$\begin{array}{cccc} TGCTGGACTG\\ A & G & L\\ CCGCTGCTGTG\\ P & L & L\\ GTCGTGACGT\\ R & D & V\\ CCCCCCCCCC\\ P & R & S\\ CAAAAACAG\\ Q & K & T & D\\ GGGCGAAATC\\ A & K & S\\ GCCGGCCGGC\\ P & G & G\\ GAGGCACTCG\\ \end{array}$	$ \begin{array}{c c} I & Q & R & K \\ I & Q & R & K \\ C & C & C & C & C \\ R & A & E \\ C & C & A & C \\ R & K & L \\ C & A & C & C & C \\ R & T & V & L \\ C & A & C & C & A \\ R & A & C & C \\ R & C & C \\ R & C & C & C \\ R & $
3601 3701 3801 3901 4001 4101 4201 4301	CCATCANATG PSNE ANGACAGTGC DSNE AGGGCTTGCG ELA ACACGCTGC LTA ANGATATTGA DIE TGCCCANATT VQI ATTGGGCGAA IGCCCGA WRD	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAAGAGCC R E P TCTCATGTCA L M S	CGTTCGAGCG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGCCCTTCT G A S GACGAAATGC D E M R TACAGAAATGC Q N S GATAAAGGCT I K A GCCGCCGGTGG A A V V	CATGGTATG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGG G D W E GAATGTTGCG M L R AAAACCGAA K P E TTTCCACCGAA F P L G TCCTGCGCATC V R S	$\begin{array}{ccc} CTGGCACGAC\\ G & T & T\\ TGGCGGCCAG\\ G & Q\\ CGGCACTTCC\\ R & H & F & R\\ AAAGCTGGA\\ K & L & E\\ TGAAGGACTCGA\\ K & L & E\\ TGAAGGACTCGA S & S & N & E\\ GAATGGTCTT\\ M & V & L\\ GATGCTGGCC\\ M & L & A\\ \end{array}$	ATTGCGCCCCC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S	$\begin{array}{cccc} \text{CACCTTGCGGCTT} \\ \text{F} & \text{L} & \text{A} & \text{A} \\ \text{CGTTCGGCTT} \\ \text{F} & \text{G} & \text{F} \\ \text{GGAAGCACTG} \\ \text{E} & \text{A} & \text{L} \\ \text{GTGCAGCTGC} \\ \text{V} & \text{Q} & \text{L} & \text{L} \\ \text{GTGCAGCTGC} \\ \text{CGGAATTTCA} \\ \text{E} & \text{F} & \text{Q} \\ \text{TCGCTCTGAA} \\ \text{R} & \text{S} & \text{E} \\ \text{CCTGAGATTT} \\ \text{P} & \text{E} & \text{I} & \text{S} \\ \text{GTGCCTACCA} \\ \text{A} & \text{Y} & \text{Q} \end{array}$	$\begin{array}{c c} CACTCGTCGCG\\ L & V & D\\ CAGCCTCGCGCGCT\\ T & I & C & R\\ T & C & C & R\\ T & C & C & CCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} TGCTGGACTG \\ A & G & L \\ CCGCTGCTTG \\ P & L & L \\ GTCGTGACGT \\ R & D & V \\ CCCCCGCTCC \\ P & R & S \\ CAAAAAACAG \\ Q & K & T & D \\ GGCCGAAATC \\ A & K & S \\ GCCGGGCGAAATC \\ A & K & S \\ GCCGGCGGG \\ P & G & G \\ GAGGCACTCC \\ E & A & L & G \end{array}$	$I = \begin{bmatrix} Q & R \\ R & A \end{bmatrix}$ $C = \begin{bmatrix} C \\ C$
3601 3701 3801 3901 4001 4101 4201 4301	CCATCANATG PSNE ANGACAGTGC DSA AGAGCTGCG ELA CTGACGGCGG LTA AGATATTGA DLE TGTCCANATT VQL ATTGGGCGAA IGR MGTTGGCGCGA WRD	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCC R E P TCTCATGTCA L M S	CGATTCGACG V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGCCT I K A GCCCCCGTGG A A V V	CATGGTATG H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGG G D W E GAATGTTGG M L R AAAACCGAA K P E TTTCCACTGG F P L G TCGTGCGATC V R S	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGGACATC E E I TCCAGCAATG S S N E GAATGGCTCTT M V L GATGCTGGCCC M L A	ATTGCGCCCCC L R R ATCGAAGATG I E D A S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAGCTGC V Q L L TGGAATTTCA E F Q TCGCTCTGAA R S E CCTGAGATTT P E I S GTGCCTACCA A Y Q	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCATG A G I GGAAGGATCT E D S ACAGGACAGG T E Q G CGAATTACGG N Y G GAATGCCTGT N A C	$\begin{array}{c c} TGCTGGACTG \\ A & G & L \\ CCGCTGCTTG \\ P & L & L & A \\ GTCGTGACGT \\ R & D & V \\ CCCCCGCTCC \\ P & R & S \\ CAABAAACAG \\ Q & K & T & D \\ GGCGGAAATC \\ A & K & S \\ GCCGGCGGGG \\ P & G & G \\ GAGGCACTCG \\ E & A & L & G \\ \end{array}$	ATCCASCGCA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC N P D GCGGTCCGGA A V G S GTGCCGAGAA A E N
3601 3701 3801 3901 4001 4201 4201 4301	CCATCANATG PSNNE AGAGCAGTGC DSNA AGAGCTGCGC CTGACGGCTG LTAA ACTGACGGCTG LTAA A AGATATTGA DIE TGTCGCGCAA IGRMA GTTGGCGCGAA GTTGGCGCGAA CAA	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCC R E P TCTCATGTCA L M S GTGATGGCCT	CGATACGCCC R Y A R AGCGATACGCCC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGCT I K A GCCCCCGCTCG A A V V GCATCCTCA	CATGGTATG H G I A GAAAGGACAA K D N AGCCGACGGCGC A E R GGCGATTGG G D W E GAATGTTGG M L R AAACCCGAA K P E TTTCCACTGG F P L G TCGTGCGATC V R S AAGGCAGGG	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGACTC E E I TCCAGCAATG S S N E GAATGGCTCTT M V L GATGCTGGCC M L A CACACCAACA	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S C GGCGC	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG E A L GTGCAGCTGC V Q L L TGGAATTTCA E F Q TCGCTCTGAA R S E CCTGAGATTT P E I S GTGCCTACCA A T Y Q	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGGATTCT E D S ACAGAGCAGG T E Q G CGAATTACGG N Y G GAATGCCTGT N C C GACTCACCC	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAAACAG Q K T D GGGCGAAATC A K S GCCGGCCGGG P G G GAGGCACTCG E A A C G	ATCCACCOCA R A E TCCTACCCCA R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC N P D GCGGTCCGGA A V G S GTGCCCAGAA A C CCAGACA
3601 3701 3801 3901 4001 4201 4201 4301	$\begin{array}{c c} CCATCANATG \\ P & S & N & E \\ \\ AAGACATGC \\ D & S & A \\ \\ AGACTTCCG \\ E & L & A \\ \\ CTGACGCCTG \\ L & T & A & A \\ \\ CTGACGCCCG \\ L & T & A & A \\ \\ AAGATATTGA \\ D & I & E \\ \\ TGTCCAAATT \\ V & Q & I \\ \\ ATTGGCCCAA \\ I & G & R & M \\ \\ GTTGGCCCCGA \\ W & R & D \\ \\ TGCTCGCGCCC \\ A & A & A \end{array}$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCCC R E P TCTCATGTCA L M S GTGATGGCCT V M A C	CGTTCGACCC V R A CGATACGCCC R Y A R AGCAGATAGC Q I A GGCCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGGCT I K A GCCCGCGGTGG A A V V GCATCCTCGA I L E	$ \begin{array}{c c} CATGGTATTG\\ H & G & I & A \\ K & D & N \\ AGCCGAGCGC \\ A & E & R \\ GGCGATTGGG \\ G & D & W & E \\ GAATGTTGCG \\ M & L & R \\ AAAACCCGAA \\ K & P & E \\ TTTCCACTGG \\ F & P & L & G \\ TCGTGCGATC \\ V & R & S \\ AAGGCAGGG \\ R & A & G \\ \end{array} $	CTGGCACGAC G T T TGGCGGCCAG G G Q CGGCACTTCC R H F R ARAAGCTGGA K L E TGAAGAGACATC E E I TCCAGCAATG S S N E GAATGGTCTT M V L GATGCTGGCC M L A CACATCAACA H I N N	ATTGCGCCGC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGGCGGG A G G	$\begin{array}{ccc} \text{CACCTTGCTG}\\ \text{H} & \text{L} & \text{A} & \text{A} \\\\ \text{CGTTCGGCTT}\\ \text{F} & \text{G} & \text{F} \\\\ \text{GGAAGCACTG}\\ \text{CGCCCTGC} \\\\ \text{V} & \text{Q} & \text{L} & \text{L} \\\\ \text{TGGAATTTCA}\\ \text{R} & \text{S} & \text{E} \\\\ \text{TCGCTCTGAA}\\ \text{R} & \text{S} & \text{E} \\\\ \text{CCTGAGATTT}\\ \text{P} & \text{E} & \text{I} & \text{S} \\\\ \text{GTGCCTACCA}\\ \text{A} & \text{Y} & \text{Q} \\\\ \text{TTATTTGCGT}\\ \text{Y} & \text{L} & \text{R} \end{array}$	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGATTCT E D S ACAGACCAGG T E Q G CGAATTACGG N Y G GAATGCCTGT N A C GACCTCACCC D L T R	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CANANACAG Q K T D GGGCGANATC A K S GCCGGGCGGGG P G G GAGGCACTCG E A L G GCAAGGCAGA K A E	$\begin{array}{c c c c c c c c c c c c c c c c c c c $
3601 3701 3801 3901 4001 4101 4201 4301 4401	$\begin{array}{c} CCATCAAAATG \\ P & S & N & E \\ AAGACAGTGC \\ D & S & A \\ AGAGCTTGC \\ E & L & A \\ CTGACGGCTG \\ L & T & A & A \\ AAGATATTG \\ D & I & E \\ TGTCCAAATT \\ V & Q & I \\ ATTGGGCGCAA \\ I & G & R & M \\ GTTGGCGCGCA \\ W & R & D \\ TGCTCCGGCC \\ A & A & A \\ TTTTCACTTG \\ \end{array}$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCC R E P TCTCATGTCA L M S GTGATGGCCT V M A C GGCCGATGTT	$\begin{array}{c c} CGTTCGACGCCC\\ V & R & A \\ CGATACGCCCC\\ C & Y & A & R \\ AGCAGATAGC \\ Q & I & A \\ GGCCGCTTCT\\ G & A & S \\ GACGAAATGC \\ D & E & M & R \\ TACAGAATTC \\ Q & N & S \\ GATAAAGGCT \\ I & K & A \\ GCCGCCGCGTGG \\ A & A & V & V \\ GCATCCTCGA \\ I & L & E \\ GATGGCGCCTG \\ \end{array}$	$ \begin{array}{c c} CATGGTATTG\\ CATGGTATTGGG CATGGG CGC CGC CGC CGC CGC CGC CGC CGC CGC$	$\begin{array}{c c} CTGGCCAGGCCAG\\ G & T & T\\ TGGCGGCCCAG\\ G & G & Q\\ CGGCACTTCC\\ R & H & F & R\\ ARAAGCTGGA\\ K & L & E\\ TGAAGGATC\\ E & E & I\\ TCCAGCAATG\\ S & S & N & E\\ GAATGGTCCT\\ M & V & L\\ GATGCTGGCC\\ M & L & A\\ CACATCAACA\\ H & I & N \\ ATGGTTCGCC\\ \end{array}$	ATTGCGCCCCC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGGCCGG A G G	$\begin{array}{c c} CACCTTGCTG\\ H & L & A & A \\ CGTTCGGCTT \\ F & G & F \\ GGAAGCACTG \\ E & A & L \\ \\ GTGCAGCTGC \\ TGGAATTTC \\ F & F & Q \\ \\ TCGCTCTGAA \\ R & S & E \\ \\ CCTGAGATTT \\ P & E & I & S \\ \\ GTGCCTACCA \\ A & Y & Q \\ \\ \\ TTATTTGCGT \\ Y & L & R \\ \end{array}$	$\begin{array}{c c} CACTCGCG\\ L & V & D\\ CAGCCTCGCG\\ S & L & A\\ ACGATCTGCC\\ T & I & C & R\\ TCGCAGGCAT\\ A & G & I\\ GGAAGATTCT\\ E & D & S\\ ACAGAGCCCCGC\\ T & E & Q & G\\ CGAATTCCGCACAGC\\ N & Y & G\\ GAATGCCTGCAAAAGCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} TGCTGGACTG \\ A & G & L \\ CCGCTGCTGTG \\ P & L & L \\ GTCGTGACGT \\ R & D & V \\ CCCCCGCTCC \\ P & R & S \\ CAAAAAACAG \\ Q & K & T & D \\ GGGCGAAATC \\ A & K & S \\ GCCGGGCCGGG \\ GAGGCACTCG \\ E & A & L & G \\ GCAAGGCACTCG \\ K & A & E \\ CTGATGCTGT \\ \end{array}$	$ \begin{array}{c c} I & Q & R & K \\ I & Q & R & K \\ C & C & C & C & C & C \\ R & A & C & C & C & C \\ C & C & A & C & C & C \\ P & T & L & L & K \\ C & C & A & C & C & C \\ P & T & L & L & K \\ C & C & C & A & C & C \\ C & A & C & C & C & C \\ C & C & C & C & C & C$
3601 3701 3801 4001 4101 4201 4301 4401 4501	$\begin{array}{c c} CCATCAAAATG \\ P & S & N & E \\ \hline AAGACAGTGC \\ D & S & A \\ \hline AGGAGCTGCC \\ E & L & A \\ \hline CTGACGCCGC \\ L & T & A \\ \hline CTGACGCCCG \\ L & T & A \\ \hline TGCCCCAAATT \\ V & Q & I \\ \hline TGCCCGCCGA \\ I & G & R & M \\ \hline GTTGGCCGCCG \\ A & A & A \\ \hline TTTTCCACTG \\ F & S & L & G \end{array}$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCCTC R E P TCCCATGCACA C R H I TGCGAGAGCCC T M S GTGATGGCCT V M A C GGCCGATGTT P M L	$\begin{array}{c c} CGTTCGACGCCC\\ V & R & A \\ CGATACGCCC\\ R & Y & A & R \\ AGCAGATAGC\\ Q & I & A \\ GGCCGCTTCT\\ G & A & S \\ GACGAATGC\\ Q & N & S \\ GACGAATGC\\ Q & N & S \\ GATAAGGCT\\ I & K & A \\ GCCGCCGTGG\\ A & A & V & V \\ GCATCCCCGA\\ I & L & E \\ \end{array}$	CATGGACGAC H G I A GAAAGGACAA K D N AGCCGACGCC A E R GGCGATTGGG G D W E GGAATGTTGCG M L R AAAACCCGAA K P E TTTCCACTGGGATC V R S AAGGCCACGAC R A G ATGCGAACGA M R T N	CTGGCCGCCCCG G G Q CGGCCCCGC C G Q CGGCCCCCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I TCCCAGCAATG S S N E GAATGGTCTT M V L GAATGGTCGCC C A N N ATGGTTCGCC	ATTECCECCCC L R R ATCCAAGATG I E D A GCAGCCCCAA S A K GGGCTACTACC G Y Y ATCCAACCAACC I N Q L AATTTGACCC F E P GCAGGCCTGC Q A C GTCAGTCCGA Y S P S ATGCCGGCGGG A G G AGGCCCGGAGA G R R	$\begin{array}{c c} CACCTTGCTGH L A ACGTTCGGCTTF G FGGAAGCACTGE A LGTGCAGCTGCV Q L LTGGAATTTCAE F QTCGCTCTGAAR S ECCCTGAGATTTP E I SGTGCCTACGAGA Y V$	CACTCGCCGC S L A ACGATCTGCC T I C R TCGCAGGCATT C GGAAGATTCT E D S ACGGACTCGCCGC T E Q G CGAATGCCCGC N X G GAAGCCCGCT GACCCCCCCC D L T R CTGCAAAAG	TGCTGGACTG A G L CCCCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAAACAG Q K T D GGGCGAAATC A K S GCCGGGCGGG P G G GAGGCACTCG E A L G GCAAGCAGCAGA K A E CTGATGCTGT	$\begin{array}{c c c c c c c c c c c c c c c c c c c $
3601 3701 3801 3901 4001 4101 4201 4301 4401	CCATCANATG PSNE ANGACAGTGC DSNE CTGACGGCTGC CTGACGGCTG CTGACGGCTG CTGACGGCCA TTGCCAAATT VQI ATTGGCCAAAT GTGCCCGAA TGCTGGCCGA WRD GTGGCGCGCA WRD TGCTGGCCCGA CAAAAAA TTTCCCTTG FSLG	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCCTC R E P TCCCATGTCA L M S GTGATGGCCTT V M A C GGCCGATGTT P M L	CGATECGATACGCCC R Y A R AGCAGATAGC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATGC Q N S GATAAGGCTT I K A GCCCCGCGTGG A A V V GCATCCCCA I L E GATGGCCTG M A L	CATGGATGGA H G I A GAAAGGACAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTTGG M L R AAACCCGAA K P E TTTCCACTGG F P L G TCGTGCGATC V R S AAGGCAGGA R A G ATGCGAACGA	CTGGCCAGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGGAGATC E E I TCCAGCAATG S S N E GAATGGTCTT M V L GATGCTGGCC M L A CACATCAACA H I N N ATGGTCGCC G S P	ATTGCGCCCC L R R ATCGAAGATG I E D A S A K GGCCCCCAA S A K GGCTACTAC G Y Y ATCAACCAAC I N Q L AATTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGCCGGC A G G AGCCCGGGCA G R R	$\begin{array}{c c} CACCTTGCTGH L A ACGTTCGGCTTF G FGGAAGCACTGCV Q L LTGGAATTTCAE F QTCGCTCTGAAR S ECCTGAGATTTP E I SGTGCCTACCAAA Y QTTATTGCGTY L RGCATCATCAGA S *$	CACTCGTCGCG L V D CAGCCTCGCG S L A ACGATCTGCC T I C R TCGCAGGCAT A G I GGAAGGATCT E D S ACAGGCAGG T E Q G ACAGGCAGG N Y G GAATACGG N Y G GAATACGG N Z G GAATACCG D L T R CTGCAAAAAG	TGCTGGACTG A G L CCCCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAACAG Q K T D GGCGGAAATC A K S GCCGGCGGGG P G G GAGGCACTCG E A L G GCAAGGCAGA K A E CTGATGCTGT	ATCCASCGCA R A E TCGTAAGCTG R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC A V G S GTGCCGAGAA A V G S GTGCCGAGAA C G E TTGGAGACCG
3601 3701 3801 3901 4001 4101 4201 4301 4401 4501	CCATCANATG P S N E ANGACAGTGC D S A AGAGCTTGCG E L A CTGACGGCGG L T A A AGATATTGA D I E TGCCCAAATT V Q I ATTGGCGCGAA I G R M GTTGGCGCGA W R D TGCTGCGCGCC A A A TTTTCACTG F S L G CAGCAGAGCGG	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGCCATC R H I TGCGAGAGCC R H I TGCGAGAGCC R P TCTCATGTCA G GCCGATGTT P M L AGATCCTGCC	CGATACGCCC R Y A R AGCGATACC Q I A GGGCGCTTCT G A S GACGAAATGC D E M R TACAGAATTC Q N S GATAAAGGCT I K A GCCCCCGCGCGC A A V V GCATCCTCGA I L E GATGGCCCCTG M A L TTGTGGTCTC	CATGGTATGG H G I A GAAAGGACAA K D N AGCCGACGGC A E R GGCGATTGGG G D W E GAATGTTGCG M L R AAACCCGAA K P E TTTCCACTGG F P L G TCGTGCGATCGA R A G AGGCAGGG R A G CCGAGTATCG	CTGGCACGGC G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGAATC E E I TCCAGCAATG S S N E GAATGGTCTT M V L GATGCTGGCC M L A CACATCAACA H I N N ATGGTTCGCC G S P CCGATAACCT	ATTECGCCCCC L R R ATCGAAGATG I E D A GCAGCCCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGGCGG A G G AGGCCGGAGA G R R	CACCTTGCTG H L A A CGTTCGGCTT F G F GGAAGCACTG C A L CGTGCAGCTGC V Q L L TGGAATTTCA E F Q TCGCTCTGAA R S E CTGGAGATTT P E I S GTGCCTACCA A X V Q CACTCATCAG A S *	CACTCGCCGCG S L A ACGATCTCCC T I C R TCGCAGGCAT A G I GGAAGATCT E D S ACAGAGCAGG T E Q G CGAATTACGG N Y G GAATGCCTGT N C C GACTCACCCC D L T R CTGCAAAACAG	TGCTGGACTG A G L CCGCTGCTTG P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAAACAG Q K T D GGGCGAAATC A K S GCCGGCCGGC GAGGCACTCG E A L G GCAAGGCAGA K A E CTGATGCTGT	ATCCACCOCA R A E TCCTACCCCA R K L CCAACCCTCA P T L K ATACCAATGA T N D GAACCCAGAC N P D GCGCTCCGGA A V G S GTGCCCGAGAC R G E TTGGAGACCG GGGCCCGTAAC
3601 3701 3801 3901 4001 4101 4201 4301 4401 4501	CCATCAAAG PSN AGACAGTGC DSN AGAGCTTGC ELA TAC CTGACGCCTG LTA AGATATTGA DIE CTGCCCAAAT COCCAAGCTG CCCCAAGCTG CCCCAAGCTG	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGACCC R E P TCTCATGTCA L M S GTGATGGCCT V M A C GGCCGATGTT P M L AGATCCTGCC	$\begin{array}{c c} CGTTCGACCC\\ V & R & A \\\\ CGATACGCCC\\ R & Y & A & R \\\\ AGCAGATAGC\\ Q & I & A \\\\ GGCGCGCTTCT\\ G & A & C \\\\ GACGAAATGC\\ D & E & M & R \\\\ GACGAAATGC\\ Q & N & S \\\\ GATAAAGGCT\\ I & K & A \\\\ GCCGCCGCGCTGG\\ A & A & V & V \\\\ GATGCCCCCGGA\\ I & L & E \\\\ GATGCCCCTCCG\\ M & A & L \\\\ TTGTGGTCCCTCTAGCCTTTTC \\\\ ACCCTTCTG \\\\ A & C & C \\\\ CACCTTCTC \\\\ A & C & C \\\\ CACCTTCTC \\\\ A & C & C \\\\ C C \\\\ C & C \\\\ $	$\begin{array}{c c} CATGGTATTG\\ CATGGTATGGG\\ CATGGTGTGGG\\ CATGGTGTGGG\\ CATGGTGTGGG\\ CATGGTGTGGG\\ CATGGTGTGGG\\ M L R \\ AAAACCCGAA\\ K P E \\ TTTCCACTGG\\ F P L G \\ TCGGTGGGATGGG\\ R A G \\ AAGGGCAGGGG\\ R A G \\ ATGCGAATCA \\ M R T N \\ CCGAGTATCG \\ AACGGAATCA \\ AACGGCAATCA \\ ATCGGAATCG \\ ATCGGAATCA \\ ATCGCAATCA \\ ATCGGAATCA \\ ATCGCCACATCA \\ ATCGCCCC \\ ACCGAATCA \\ ATCGCCCCC \\ ACCGAATCA \\ ACCGAATCA \\ ACCGCCCCC \\ ACCGAATCA \\ ACCGCCCCCC \\ ACCGAATCA \\ ACCGCCCCCCC \\ ACCGAATCA \\ ACCGCCCCCCCCCC \\ ACCGAATCA \\ ACCGCCCCCCCC \\ ACCGAATCA \\ ACCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CTGGCCAGGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGAGATC E E I TCCAGCAATG S S N E GAATGGTCGCC M L A CACATCAACA H I N N ATGGTTCGCC G S P CCGATAAGCT	ATTGCGCCCCC L R R ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTGACCAAC I N Q L AATTGACCAAC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGGCCGGAGA G R R ATCCAATCTAA CCCCCCCCCC	$\begin{array}{c c} CACCTTGCTGH L A ACGTCCGCTTF G FGGAAGCACTGE A LGGTGCAGCTGCV Q L LTGGAATTTCAE F QTCGCTCTGAAR S ECCTGAGATTTP E I SGTGCCTACCAA Y QL RGCATCATGAGA S *$	$\begin{array}{c c} CACTCGTCGCG\\ L & V & D\\ CAGCCTCGCGCGT\\ T & C & R\\ CCGCAGGCAT\\ R & G & I\\ GGAAGATTCT\\ E & D & T\\ CGCAGGCATCGCCCGT\\ T & E & Q & G\\ CGAATGCCTCT\\ N & A & C\\ GAATGCCTCCCCCD & L & T & R\\ CTGCAAAACGCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} TGCTGGACTG \\ A & G & L \\ \\ CCCGCTGCTGT \\ P & L & A \\ \\ GTCGTGACGT \\ R & D & V \\ \\ \\ CCCCCGCTCC \\ P & R & S \\ \\ \\ CCCGGCCGAAAAACAG \\ Q & K & T & D \\ \\ \\ GGCGGCAAAAACAG \\ Q & K & T & D \\ \\ \\ GGCGGCGAAAACAG \\ Q & K & T & D \\ \\ \\ GGCGGCGAAAAACAG \\ Q & K & T & D \\ \\ \\ GGCGGCGCACAC \\ \\ CAAAAAAACAG \\ \\ \\ CTGATGCTCCAAAAAAACAG \\ \\ \\ \\ \\ CTGATGCTTCCAAAAAAACAG \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$	$\begin{array}{c c} I & Q & R & K \\ I & Q & R & K \\ C & C & C & C & C & C \\ R & A & E \\ C & C & A & C & C & C \\ R & K & L \\ C & C & A & C & C & C \\ R & K & L \\ C & C & A & C & C \\ R & K & L \\ C & C & C & C & C \\ R & K & L \\ C & C & C & C & C \\ R & K & L \\ C & C & C & C & C \\ R & K & L \\ C & C & C & C & C \\ R & K & L \\ C & C & C & C & C \\ R & C$
3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4701 4801 4701	CCATCANAGE PSN AGACAGTGC DSN AGAGCTTGCG PL T CTGACGCCTG L T CTGACGCCTG L T CTGACGCCCA C CGCCCCAAGCTG CCCGGACTGG CCCGGACTGG	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCCC R E P TCTCATGTCA L M S GTGATGGCCTT V M A C GGCCGATGTT P M L AGATCCTGGC	$\begin{array}{c c} CGTTCGACGCCC\\ V & R & A \\ \\ CGATACGCCC\\ R & Y & R \\ \\ AGCAGTTC\\ Q & I & A \\ \\ GGCGCCTTCT\\ GACGAATTC\\ Q & N & S \\ \\ GACGAATTC\\ Q & N & S \\ \\ GATAAGGCT\\ I & L & E \\ \\ GATGGCCCTCG\\ R & A & V \\ \\ GCATCCTCCG\\ M & A & L \\ \\ \end{array}$	CATGGTATG G A G I A GAAAGGCAA K D N AGCCGAGCGC A E R GGCGATTGGG G D W E GAATGTGCG M L R AAAACCGAA K P E TTTCCACTGG F P L G TCGTGCGATC V R S AAGGCCACGA M R T N CCGAGTATCG AACGAATCA	$\begin{array}{c c} CTGGCCAGGCCAG\\ G & T & T\\ TGGCGGCCCAG\\ G & G & Q\\ CGGCACTCCC\\ R & H & F & R\\ AAAAGCTGGA\\ K & L & E\\ TGAAGGATCGGA\\ K & L & E\\ TGAAGGATCGGACT\\ TCCCAGCAATGGCCT\\ M & V & L\\ GAATGGTCGCCC\\ G & S & P\\ CCGATAACCT\\ CGGCACT\\ TCCGGCTGT\\ TTTTTCCGCCCC\\ TTTTTTTCCGCCCC\\ TTTTTTTCCGCCCC\\ TTTTTTTTCCGCCCC\\ TTTTTTTCCGCCCC\\ TTTTTTTCCCGCCCTC\\ TTTTTTCCCGCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTTCCCGCCCCC\\ TTTTTTTCCCGCCCCC\\ TTTTTTTCCCGCCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTTCCCGCCCCC\\ TTTTTTCCCGCCCCC\\ TTTTTTCCCCCCC\\ TTTTTTCCCCCCCC\\ TTTTTTCCCCCCC\\ TTTTTTCCCCCCC\\ TTTTTTCCCCCC\\ TTTTTCCCCCCC\\ TTTTTCCCCCCC\\ TTTTTCCCCCCCC$	ATTGCGCCCCA ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTGACCAC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGCCGC A G G AGGCCGGAGA ACTGGACA	$\begin{array}{c c} CACCTTGCTGH L A ACGTTCGGCTTF G FGGAAGCACTGE A LGTGCAGCTGCV Q L LTGGAATTTCAE F QTCGCTCTGAAR S ECCTGAGATTTP E I SGTGCCTACCAA Y QTTATTTGCGTY L RGCATCATGAGA S *$	CACCORECTOR L V D CASCCTCCCC S L A ACGATCTCCC T I C R TCCCCAGCAT CCCCAGCATTCT CCCCACGCATCC T E Q G CGAATACCG CCCCCCCCCT N Z C CCCCAAAACC CCCCCAAAACCC CCCCCACCCCCCC	TGCTGCTGC A G L CCCCTGCTGC P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CCAAAAACAG Q K T D GGCCGAAATC A K S GCCGGCGGG P G G GAGGCACTCG CTGATGCTGCT CATATTCGAA	$\begin{array}{c c c c c c } \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} $
3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4501	$\begin{array}{c c} CECTCCATCAAATG\\ P & S & N & E \\\\ AAGACACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGCCC R E P TCTCATGTCA GTGATGGCCTTGTCA AGATCCTGGC AACAGCTGTG ATAACATCAT	CGATCCGATACGCCC R Y A R AGCAGATACC Q I A GGGCCGCTCT G A S GGCCGCTCT Q N S GACGAATCC Q N S GATAAGCCT I K A GCCCCCGTCG GATCCCCCGA GATCCCCCGA GATCCCCCGA A L V CACCCCCCCC GATCCCCCCCC M A L	CATGGATCGA CATCGACGA CAAAGCAAAGCAA K D N AGCCGACGCCGC A E R AGCCGACGCGC M R C CCGAGTATCG AACGAATCA	CTGGCCAGCCAG G G Q CGGCACTTCC R H F R AAAAGCTGGA K L E TGAAGGAGATC E E I TCCAGCAATG S S N E GAATGGTCTT M V L GATGCTCGCCC G S P CCCGATAAGCT TCTCCGGCTGT	ATTECGACCGC L R R ATCGAAGATG I E D A GCAGCCCCAA S A K GGGCTACTACC G Y Y ATCAACCAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA V S P S ATGCGGCCGGCG A G G AGGCCCGCGGCG A G G AGCCCCTATG	$\begin{array}{c c} \text{CACCTUSCUSCUT}\\ \textbf{H} & \textbf{L} & \textbf{A} & \textbf{A} \\ CGTUSCUSCUSCUSCUSCUSCUSCUSCUSCUSCUSCUSCUSCU$	CACCUCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TGCTGGATGGATCG TCTGGATGACAGA TCGCCCGCTCC TCGCTGACGT TCGCTCGATGC TTTCGGATCC TTTCGGATCC TTTCGGATCC TTTCCGATGC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCGATCC TTTCCCC TTCCCCC TTCCCCC TTTCCCCCC TTTCCCCCC TTTCCCCCC TTTCCCCCC TTTCCCCCC TTCCCCCCCC	$\begin{array}{c c c c c } ATCCCACCACCACCACCACCACCACCACCACCACCACCAC$
3601 3701 3801 3901 4001 4101 4201 4301 4401 4401 4401 4401 4401 501	CCATCANCG P S N P AGACACTCC P Z N AGACCTCC P Z N AGACCTCC P Z N AGACATATCA D I P AACATATCA D I P AACATATCA CCCACCCC A A A A CTCCCCACCCC CCCCACCCCC CGACCCCCCCCC	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCCCCC Q L A Q CGATAGAGAG I E E ATCCGTCCTCC R H I GGACGTCACA R H I TCCGAGAGCCC R F TCCATGTCA CGATGCCTCG GGCCGATGTT P M L AGATCCTGGC TTTATTGATC GAACGCCTACA	CGATCCGATACCCCC R Y A R CGATACCCCC Q I A GGCCCTTT G A S GGCCGCTCT C A S GACCAATCC Q N S GACCAATCC Q N S GACCGCCCCCG A A V V GCATCCCCGA GATCCCCCGA TTGTGCCTTCG GCATCTCTG GCATTCCACCG GATTCCACCG GATTCCACCG	CATGGATTCGACCAA	CTGGCCAGGCCAG G G Q CGGCACTTCC R H F R ANAGCTGGA K L E TGAAGGAGATC E E I TCCAGCAATG GATGCTGGCCTM M V L GATGCTGGCCCM H I N N ATGGTTCGCCC G S P CCGATAGCT TTTTACCGC	ATTECGACCCC L R R ATCCGACGACAA S A K GCACCCCAA S A K GGCCTACTAC G Y Y ATCAACCAAC I N Q L AATTTGAACC F E P GCAGGCCTGC Q A C GTCAGTCCGA A G G AGGCCCGACG A G G ACCCCCCCCGCG CAATCTAA CAATCTAA CAATCTAA CAATCTAA CACCCCCCCCCC	CACCTTGCGCTT F G F GGAGGCACTG E A L GTGCAGCTGC V Q L L TGGAATTTCA E F Q TGGCATCTGAA R S E CCTGAGATTT P E I S GTGCCTACCA A Y Q TTATTTGCGT Y L R GCATCATCAG A S *	CACCECCECC S L A ACCATCTECC T I C R TCCCACECCT A G I CCCCCACCA A G I CCCCCCCCCC N Z G CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TGCTGGATGGACTG A G L CCGCTGCTGTG P L L A GTCGGGACGGACTG Q K T D GGCGGAATCG A K S GCCGGCGGGG P G G GAGGCACTGG CATATTCGAA CTGATGCTGT CATATTCGAA CTGCTTTGGGAC CGAAATTCGAA	ATCCACCOCC R R R CCACCCCA R R C CCAACCTCA P T L R ATACCAACA T CCAACAACA ATACCAACAA T N D CCACCAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAA ATACAACAA ATACAACAA ATACAA ATACAACAA ATACAACAA ATACAACAA ATACAA ATACAACAA ATACAACAA ATACAACAA ATACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAA ATACAACAACAA ATACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAA ATACAACAACAACAAA ATACAACAACAAA ATACAACAACAAA ATACAACAACAAACAA ATACAACAAACAAA ATACAACAACAAACAAAAAAAAAA
3601 3701 3801 3901 4001 4101 4201 4301 4401 4401 4701 5001 5001	CCATCANAG PSN AGACACTCC CCACCTCC CCACCTCC CCACCTCC CCACCTCC CCACCTCC CCACCTCC CCACCTCC CCACCTCCC CCCCACTCC CCCCCCCCCC	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTCC S V L GGACGTCACA G R H I TGCGAGAGACCCT R E P TCTCATGTCA L M S GTGATGGCCTC V M A C GGCCGATGTT P M L AGATCCTGGC AAGAGCTGGG ATTAGTTGATC GAACCAAAG	$\begin{array}{c c} CGTTCGACCC\\ V & R & A \\\\ CGATACCCC\\ R & Y & R \\\\ AGCAGATC\\ Q & I & A \\\\ GGCCCCTTCT\\ G & A & S \\\\ GATAAAGATTC\\ Q & N & S \\\\ GATAAAGCT\\ I & K & A \\\\ GCCCCCCCCGACCTCGA\\ I & L & E \\\\ GATGCCCTCTGACCTTCC \\\\ AA & V & V \\\\ GATCCTTCCC \\\\ AA & C \\\\ AA & C \\\\ CCCCCTCTGACCCTCCGACCCCCCACCCCCCCCCCCCCC$	CATCGCCCCA	$\begin{array}{c c} CTGGCCAGGCCAG\\ G & G & Q\\ CGGCACTTCC\\ R & H & F & R\\ AAAAGCTGGA\\ K & L & E\\ TGAAGAGCTGGA\\ K & L & E\\ TGAAGAGCTGGA\\ K & L & E\\ TGAAGGCTGGC\\ S & S & N & E\\ GAATGGTCGCCC\\ M & L & A\\ CACATCACCCC\\ G & S & P\\ CCGATAAGCT\\ TCTGGGCTTTTTTTCCCGCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} \text{ATTGCGCCCCC}\\ \textbf{L} & \textbf{R} & \textbf{R} \\ \text{ATCGAAGATG}\\ \textbf{I} & \textbf{E} & \textbf{D} & \textbf{A} \\ \text{GCACCCCAAC}\\ \textbf{S} & \textbf{A} & \textbf{K} \\ \text{GGGCTACTAC}\\ \textbf{G} & \textbf{Y} & \textbf{Y} \\ \text{ATCAACCAAC}\\ \textbf{I} & \textbf{N} & \textbf{Q} & \textbf{L} \\ \text{AATTGCACCAAC}\\ \textbf{I} & \textbf{N} & \textbf{Q} & \textbf{L} \\ \text{AATTGCACCCAAC}\\ \textbf{G} & \textbf{K} & \textbf{R} \\ \text{GCAGGCCTGC}\\ \textbf{Q} & \textbf{A} & \textbf{C} \\ \text{GTCAGTCCGAAC}\\ \textbf{G} & \textbf{R} & \textbf{R} \\ \text{ATCCATCTAA}\\ \text{CGCCCCGCACA}\\ \text{CGCAATTTAA}\\ \text{ACCCCGCACC}\\ \text{AGCGCCTATC} \\ \text{AACCCGGCCTTC} \\ \text{AGCCCCGACAC} \\ \text{AGCGCCTTCC} \\ \text{ACCCGCTCTC} \\ \text{ACCCCCCCCCCC} \\ \text{ACCCCCCCCCCCC} \\ \text{ACCTGCTTAA} \\ \text{AACCCCCCCCCCCCCCC} \\ \text{ACCCCCCCCCCCCCCCC} \\ \text{ACCCCCCCCCCCCCCCC} \\ \text{ACCCCCCCCCCCCCCCCCCCCC} \\ ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c} CACCTTGCTGC H L A A CGTCCGCTT F G F GGAACCTGC F A L GGAACCTGC V Q L L TGGAATTCA F P Q TCGCTCTGAA R S E CCTGAGATTT P E I S GTGCCTACGA A S * AGTGGGGTCG GAACAACGGC ACACCCAACGCC ACACCCGAAC CTACTCTACGA$	CACCUCCUCUCUCUCUCUCUCUCUCUCUCUCUCUCUCUC	$\begin{array}{c c} TGCTGGACGG\\ A & G & L \\ \\ CCGCTGCTGTGACGT\\ P & L & A \\ \\ GTCGTGACGTCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	$\begin{array}{c c c c c } & Q & Q & Q & Q & Q & Q & Q & Q & Q & $
3601 3701 3801 3901 4001 4101 4201 4301 4301 4401 4501 4701 4801 5201 5201	$\begin{array}{c c} CEATCEATACTACTACTACTACTACTACTACTACTACTACTACTAC$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCGCAC Q L A Q CGATAGAAGA I E E ATCCGTCCTC S V L GGACGTCACA G R H I TGCGAGAGACCT C R E P TCTCATGTCA G GCGAGAGCCT V M A C GGCCGATGTT P M L AGATCCTCGC AAGACCTGG TTATGATCA GAACCCAAAG ATTGGCCTCA	$\begin{array}{c c} CGTTCGACCC\\ V & R & A \\ CGATACCCCC\\ CGATACCC\\ C & I & A \\ RGCAGTACC\\ Q & I & A \\ GGCGCTCTC\\ G & A & S \\ GACGAAATC\\ GACGCCCCCC\\ GATACCCCCCC\\ A & A & V \\ GATGGCCCCCCC\\ A & A & L \\ CCCCCCCCCCC\\ GCAGTCTCCC\\ GCAGTCTCCC\\ GAGCCCCTCCC\\ GAGCCCCCCCC\\ GATCCCCCCCCC\\ GACCCCCCCCCC\\ GACCCCCCCCCC\\ GACCCCCCCCCC\\ GACCCCCCCCCCC\\ GACCCCCCCCCCC\\ GACCCCCCCCCCCCC\\ GACCCCCCCCCCCCCCCCCCCC\\ GACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC$	CATEGACAA	$\begin{array}{c c} CTGGCCACGCCACGCCACGCCCCCGCCACTCCCGCACTCCCGCACTCCCGCACTCCCCCACGCCCCCCACGCCCCCACGCCCCCCCC$	ATTGCGCCCCCA ATCGAAGATG I E D A GCAGCGCCAA S A K GGGCTACTAC G Y Y ATCCAACCAAC I N Q L AATTGCACCAAC I N Q L AATTGCACCCAC G Z CCCCCCCCCC GCAGCCCCGCGC A G G ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	$\begin{array}{c c} CACCTTGCTGH L A ACGTCCGCTTF G FGGAAGCACTGC A LGTGCAGCTGCV Q L LGTGCAGCTGCV Q L LTGGAATTTCAF F QTCCCTCGAACCCTGAGATTTP E I SGTGCCTCCCAA Y QTTATTTGCGTY L RGCATCACCGCGCCTCACCGAACACCCCGAACACCCCGAACTACTCAC$	$\begin{array}{c c} CACCOUNT Control Cont$	TGCTGATGCGAC CCCCCCGCTCC P L L A GTCGTGACGT R D V CCCCCGCTCC P R S CAAAAACAG Q K T D GGCGAAATC A K S GCCGGCCGGC P G G GAGCCACTCC E A L G CTGATGCTTCG CTTGATCCTTCGAT CTTTGCGAC CTTTTGCGAC CTTTTGCGAC CTTTTGCGAC CTTTTGCGAC	$\begin{array}{c c c c c } \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} \\ \mathbf{X} & \mathbf{X} \\ \mathbf{X} \\ \mathbf{X} & \mathbf{X} \\ $
3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4501 5001 5001 5001	$\begin{array}{c c} CCACCCACCCCCCCCCC$	AACAGCTTTC Q L S TAATGGCAAA N G K CAGCTCCCCCC Q L A Q CGATAGAAGA I E E ATCCCTCCTC S V L GGACGTCCTC R E P TCTCATGTCA G R H I TGCGAGGCCTTT P M L AGATCCTCGC AAGAGCTGTG TTATGATCATCAT AAGAAGTTGG CCATTTGCC	CGATCCGATCG V R A CGATACGCCC R Y A R GGCCGCTCT G A S GGCCGCTCT G A S GACGAAATGC Q N S GATAAGGCT I K A GCCGCCGCGCGC A A V V GCATCCTCG GAAGTTACG GAGTCTTGCT CCGCGCCCTCG A A C	CATGGACAGA	CTGACAGACT CGCCACTCC CGCCACTCC CGCCACTCC CGCCACTCC CGCCACTCC CGACGCACTCC CGACGCACTCC CGACGCACTC CCCCACTCC CGACGCCCCC CACTTACTT ACCTCCCCC	ATTGCGCCCCC L R R ATCGAAGATG I E D A GCAGCCCCAA S A K GGGCTACTAC G Y Y ATCAACCAAC I N Q L AATTGACC F E P GCAGGCCTGC Q A C GTCAGTCCGA G R R ATCCAACTAA ACTCGACGACA ACTCGACTAT GCAGGCTATC CCCCCGCTT AGCAACTAA	$\begin{array}{c c} CACCTUSCTGH L A A CCGTUSCGCTTF G FGGAAGCACTGC A C CC C C C C CC C C C C C CC C C C $	CACCONCICCA L V D CASCCTCCCC S L A CGATCTCCCC GGACCACC GGACCCCCC GACCCCCCCCCC	TGCTGGACGG A G CCCCCTGCTG P L A G GTCGTGACGT R D V CCCCCCGCTCC P R Q K GGCCGAAACCAG P G GGCCGGAATCC P G GAGGCCACTCG E A GCTGATGCTGT CATATTCCGAA CTGATGCTGT CATATTCCGAA CTTCTGGATGCCATCG CTTCTGGATCCTTCGG TTTTGACCC CATATCCGAAGGAC CTTCTGGATGCCATTCAT CTTCTGGATGCCATTCAT CTTCTGGATGCCATTCAT CTTTTGACGCC CTTTTGAGCGAC	$\begin{array}{c c c c c c } \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} \\ \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} & \mathbf{X} $

5701 TCGGCGCTCG GGCAACCTT GAAGCCTCC ATCCCTTCGC GAACATCG GGGGAACGG CTGGCGATG TTGGCGATG CGGGGAACCG TTCATAGAAC 5801 TTTGCCTTGC GGATATAAAA CGCAACGCGC GTGAGAGTCG TCTGCTGCCG ATCGAGAATG GTTTGCAAGA AGAAGGGCGA AGTTG

Figura 5.16. Fragmento que se obtuvo mediante secuencias sobrelapadas, el cual contiene al operón repABC y adyacente a éste, la secuencia 1 a 927 una secuencia parcial del gen traI. En gris se muestra el gen repA, en azul el gen repB y en negritas el gen repC con su secuencia de aminoácidos, en rojo se muestran los codones de inicio y término. La secuencia sombreada codifica para un ctRNA. Secuencia subrayada indica posible sitio de partición (*parS*).

V. III. Presencia de repC y repABC en R. tropici CFN299 y CFN299-10

5.3.1. Transferencia de repC (A789) y repABC (A627) a R. tropici CFN299 y CFN299-10

Para estos experimentos se realizaron conjugaciones en las que se transfirieron las construcciones A627 (pUX19-*repABC*) y A789 (pUX19-*repC*) a *R. tropici* CFN299 y *R. tropici* CFN299-10 mediante el ayudador 2013. La cepa CFN299-10 tiene una deleción en el pSym (Martínez-Romero y Rosenblueth, 1990) y se utilizó para saber si en la parte deletada se perdió uno de los 2 replicadores. Los resultados fueron los siguientes: en la transferencia de A627 a *R. tropici* CFN299 se obtuvieron 16.4 x 10⁴ UFC por mililitro de muestra original, de A627 a *R. tropici* CFN299-10 se obtuvieron 3.9 x 10⁴ UFC por mililitro. En la transferencia de A789 a *R. tropici* CFN299 se obtuvieron 3.9 x 10⁴ UFC por mililitro, mientras que de A789 a *R. tropici* CFN299-10 no se obtuvo ninguna transconjugante.

Mediante geles tipo Eckhard se mostró que el operón *repABC* contenido en la construcción A627 con resistencia a Km recombinó en el genoma de *R. tropici* CFN299 y *R. tropici* CFN299-10, esto indica que *repABC* se encuentra en las 2 cepas. El gen *repC* contenido en la construcción A789 recombinó sólo en *R. tropici* CFN299 y no en CFN299-10 por lo que en esta cepa no existe una secuencia *repC* homóloga para recombinar, lo cual significa que no tiene el gen *repC*. Con este experimento, se demuestra que en el pSym deletado de *R. tropici* CFN299-10 únicamente se encuentra el replicador *repABC* y que este replica autonomamente al pSym. Cuando se transfirieron las construcciones A627 y A789 a *R. tropici* CFN299 no se logro curar la cepa del pSym mediante incompatibilidad de los replicadores. En otros trabajos tampoco se ha logrado curar a *R. tropici* CFN299-10 (Martínez-Romero y Rosenblueth, 1990).

5.3.2. PCRs de *repC* y *repABC* de las cepas CFN299 y CFN299-10

Tomando en cuenta los resultados anteriores, se amplificaron los replicadores de CFN299 y CFN299-10 para asegurarnos que en la parte deletada del pSym de *R. tropici* CFN299-10 sólo existe *repABC*, se obtuvieron los siguientes resultados:

Sólo con el DNA de *R. tropici* CFN299 se obtuvieron amplificaciones de los 2 replicadores, *repABC* y *repC* del pSym (Figs. 5.17 y 5.18). Con el DNA de *R. tropici* CFN299-10 no se amplificó la secuencia correspondiente al gen *repC*, sólo se amplificó *repABC* (Fig. 5.18).



Figura 5.17. Gel de agarosa con muestras de PCRs de *repABC* usando como templado DNA de las cepas de *R. tropici* CFN299 y CFN299-10, y los oligonucleótidos REP1 y REP2 los cuales amplifican *repBC*. Carril 1 marcador de peso molecular 1Kb plus DNA ladder. Carril 2 control positivo. Carril 3 y 4, con DNA de la cepa CFN299 como templado se amplifica *repBC*. Carril 5 y 6, también se amplifica *repBC* con DNA la cepa CFN299-10.



Figura 5.18. Gel de agarosa con muestras de PCRs del gen repC usando como templado DNA de las cepas de *R. tropici* CFN299 y CFN299-10. Carril 1 marcador de peso molecular 1Kb plus DNA ladder. Carril 2 y 3, con DNA de la cepa CFN299 como templado se amplifica el gen repC. Carril 4 y 5, no se amplifica el gen repC con DNA de la cepa CFN299-10.

5.3.3. Detección de los replicadores de las cepas CFN299 y CFN299-10 por Southern blot

Para tener más certeza en los resultados obtenidos en las conjugaciones y por PCR, se realizó un experimento tipo Southern blot para localizar los replicadores de *R. tropici* CFN299 y *R. tropici* CFN299-10 (Fig. 5.19). Nuevamente se encontró el operón *repABC* en ambas cepas y el gen *repC* sólo en *R. tropici* CFN299. Este experimento confirma que el replicador funcional en el pSym de *R. tropici* CFN299 es el operón *repABC*.



Figura 5.19. Southern blot. **A)** Carril 1, DNA total de *R. tropici* CFN299-10 y carril 2, DNA total de *R. tropici* CFN299 digeridos con *Eco*RI, como sonda se utilizó A610 (*repBC*) digerido con *Eco*RI-*Hin*dIII, *repABC* se encuentra en las 2 cepas. Se muestran 2 bandas por que el operón es cortado por *Eco*RI. **B)** Carril 1 DNA CFN299-10 y carril 2 DNA CFN299, digeridos con *Bam*HI, utilizando como sonda A780 (*repC*) digerido con *Eco*RI, sólo hay señal en la cepa CFN299 lo que indica que el gen *repC* no se localiza en la cepa CFN299-10.

VI. DISCUSIÓN

6.1. Gen *repC*

El pSym de *R. tropici* CFN299 contiene 2 replicadores diferentes, uno pertenece a la familia *repC* y el otro a la familia *repABC*. También se ha encontrado que algunos plásmidos contienen 2 replicadores *repABC* funcionales como es el caso del plasmido pCFN42a (Cevallos *et al.*, 2008). El replicador *repC* se localiza también en plásmidos de *S. meliloti* y S. *fredii* (Burgos *et al.*, 1996; Villadas *et al.*, 1995).

En experimentos previos se ha demostrado que el gen repC es esencial para la replicación en el plásmido pRmeGR4a de *S. meliloti* GR4 (Mercado-Blanco y Olivares, 1994; Izquierdo *et al.*, 2005). Sin embargo, cuando la construcción A789 conteniendo el gen repC desde el codón de inicio hasta el codón de paro, con una posible secuencia promotora completa, se conjuga a *A. tumefaciens* no se produjo ninguna transconjugante, lo cual indica que el gen repC posiblemente no replica el pSym de *R. tropici* CFN299. Esto también se demostró cuando no se localizó el gen repC en el pSym de *R. tropici* CFN299-10 y sin embargo en esta cepa el pSym se replica mediante el operón repABC.

El replicador básico del plásmido pRmeGR4a de *S. meliloti* GR4 consiste solamente de 2 genes; uno que codifica para la proteína iniciadora RepC y el otro que codifica un pequeño RNA antisentido que se transcribe en dirección opuesta a repC. El producto del RNA antisentido, un RNA contra-transcrito (ctRNA) es complementario al transcrito del mRNA de repC. El ctRNA tiene una influencia negativa sobre la actividad del promotor del gen repC y forma una estructura en forma de tallo asa que funciona como un terminador independiente de *Rho* (Izquierdo *et al.*, 2005). Esta estructura de replicación se localiza también en la región promotora del gen repC del pSym de *R. tropici* CFN299.

La secuencia aislada del pSym de *R. tropici* CFN299 que contiene el gen *repC* se ubica en un tamaño de 3771 pb. En este tamaño también se localizaron 3 genes más. En dirección 5' a 3' primero se localiza un gen *psiB* que codifica para una proteína PsiB, en la región río abajo de éste, una secuencia aún desconocida de 335 pb (Fig. 6.1) y posteriormente un gen que codifica para un pequeño ctRNA. En la región río debajo de estos 3 genes se localiza la secuencia del gen *repC*. Esta secuencia fue clonada en el plásmido pUX19.



Figura 6.1. Se Muestra la secuencia *repC* obtenida de *R. tropici* CFN299 que fue comparada en la base de datos. Las secuencias debajo del número 1 corresponden a un gen *psiB*, la secuencia debajo de 2 a una secuencia desconocida encontrada sólo en *R. tropici* CFN299 y las secuencias debajo del número 3 a un gen *repC*.

La proteína PsiB previene la inducción de la respuesta SOS tras un proceso de conjugación en bacterias receptoras. Dicha acción anti-SOS inhibe la activación de la proteína RecA, la cual a su vez previene la actividad de LexA, por lo que el producto de *psiB* promueve de esta manera la instalación de un replicón en la nueva célula a donde fue transferido. Así el pSym podría ser trasferido a otras bacterias.

En el análisis de la secuencia clonada se observa que el grado de identidad en secuencia de aminoácidos de *repC* de *R. tropici* CFN299 con el gen *repC* del plásmido pRmeGR4a de *S. meliloti* GR4 es de 63%. En la región promotora del gen *repC* de *R. tropici* CFN299 se localizó una secuencia que codifica un RNA antisentido y esta secuencia tiene una identidad con el RNA antisentido de de *S. meliloti* GR4 del 81%, sin embargo con el resto de la región promotora no existe ninguna identidad entre ambas secuencias (Fig. 6.2).



Figura 6.2. Comparación entre genes repC de *R. tropici* CFN299 y *R. meliloti.* Los genes que codifican para los ctRNAs se representan en rectángulos, se observa que en la región promotora sólo hay identidad desde el ctRNA hasta el inicio del gen repC y ésta es del 81%. Los genes repC comparten una similitud del 63% a nivel de aminoácidos.

Al no obtener replicación de la construcción A789 y tomando en cuenta resultados de trabajos previos podemos deducir las posibles causas por las que repC no se replica en *A. tumefaciens*:

- a) Posiblemente psiB y la región desconocida interfieran en la expresión del gen repC y tengan una función de regulación negativa sobre el replicador.
- b) Se regule de otra manera por elementos regulatorios distintos a los presentes en *repC* de *S. meliloti* GR4.
- c) Que el ctRNA funcione de manera diferente a *S. meliloti* GR4.

Para tener una posible respuesta, de por qué repC no es funcional se deben responder los puntos anteriores para tener un acercamiento al funcionamiento del replicador.

En el inciso a) sería necesario realizar diversos recortes de la región promotora y analizar su expresión mediante fusiones transcripcionales. Los recortes nos indicarían cuales regiones río arriba de la proteína RepC están involucradas en la represión del replicador *repC*. También sería necesario eliminar específicamente el gen *psiB* y la secuencia sin homología, ya que así se tendría una región como la reportada para *S. meliloti* GR4, en la cual la región mínima replicable está conformada por *repC*, su secuencia promotora y dentro de esta secuencia se localiza el gen que codifica para el ctRNA (Izquierdo *et al.*, 2005).

En el caso b) al realizar una comparación del promotor repC de *R. tropici* CFN299 en la base de datos, el mayor grado de identidad fue con el promotor del gen repC del plásmido pRmeGR4a, sin embargo la identidad se restringe al gen para el ctRNA de *S. meliloti* (Fig. 6.2). La parte restante del promotor no tiene similitud alguna. Esto indica que la región promotora del gen repC R. tropici CFN299 no es similar a la del gen repC del plásmido pRmeGR4a, por lo que se puede sugerir que se regulan de manera diferente.

Una solución a este problema sería clonarle al gen *repC* de *R. tropici* CFN299 un promotor funcional como el de *S. meliloti* ya que ambos promotores contienen un gen que codifica un ctRNA y se tiene la idea de que sea probablemente requerido para la formación de un complejo de iniciación (Izquierdo *et al.*, 2005), así se tendría una posible respuesta acerca de si el promotor es funcional o no, o quizás la región regulatoria posea más elementos de regulación negativa que *S. meliloti*.

En c) al comparar el gen para el ctRNA que se localiza en la región promotora del gen repC de *R. tropici* CFN299 con otros 7 homólogos, 5 pertenecientes a repABC y 2 a repC se encontró que 1 nucleótido en repC de *R. tropici* CFN299 es diferente a los

que se encuentran conservados en las otras 7 secuencias, éste se muestra en rojo (Fig. 6.3).

Esta comparación nos podría indicar que al ser diferente el gen para el ctRNA en un nucleótido conservado en los otros, posiblemente funcione de manera distinta al de pRmeGR4a.

1 pRmeGR4b	WGACAGGCAGGCAGGACGACCAGGAAAGGATICGCCTTATTTATCTACT-TTGAGCCCCCCTTGCGGACGGCGCCGCGCGCGCGCGCGCG
2 repC tropici	TTGACAAGGAAGGAAGGGTCTGCGACAAAAGGTTTTGGCTAATTTATTCTCTTACAGACCCGCTTCCGGATGCCTTCGGGAGCGGTTTTTCTTTTT-
3 pRmeGR4a	TTGACAGGGAAGGGAGGGTCTGCGAGAAAGGATTCGGCTTATTTAT
4 repABC tropici	TTGACAGGGATTCCTGGAAATGCGATTCTCAATCTTGCTTAGGGAATACGAGAAGCCTTTCCACGACGACTTG-GTCGTTTGGGGGGCCTTTTTCTTTTC
5 pRetCFN42d	TTGACGAGGATTCCAGGAAATGCGATTCTGTTCTTC-CTG-CTACAGAG-ACAGA-AGGCGTTCCACGACGACGACGTCGAGGGAGGTCTTTTCTTTTCCGG
6 pSymB	TTGACACTGATTCCCCGGAAGTCCGATTCTCTAGTTC-CTACAC-G-AGAGAAGCGTTTCCCGCAGCCCTACTCGCAAGCCTTTTTCTTTTCC-
7 pTiC58	TTGACTACGATTCCTGGAAATGCGATTCTCTTAGCTGCGAGACTACAGAGA-AGCTGTTCCCGGAACGCCGTTTCGGGGGCCTTTTTCTTTTC
8 pSymA	<u>TTGAC</u> TGT <mark>GA</mark> TTC <mark>GTGGAAA<u>TGCGA</u>TTCTCGATCTTGCATCAAATCTC<mark>A</mark>GGAA<mark>GGC</mark>TTCCACGGC-TTGTTCGTT-GGGGCCCTTTTTC</mark> CT <u>TTT</u> CGT

Figura 6.3. Comparación de algunos genes que codifican para un ctRNA de replicadores repC y repABC. De la secuencia 1 a 3 se observan los pertenecientes a la familia repC y de la secuencia 4 a 8 se muestran los pertenecientes a la familia repABC. En la secuencia 2 se muestra el gen para el ctRNA de *R. tropici* CFN299 que se localiza en la región río arriba del gen repC del pSym, en esta secuencia se muestra en rojo un nucleótido distinto a todos las demás que se conservan completamente. También se muestra en la secuencia 4 el gen para el ctRNA del operón repABC del pSym de *R. tropici* CFN299. La secuencias 1 y 3 corresponden a *S. meliloti* GR4, las secuencias 6 y 8 a *S. meliloti*, la 5 a *R. etil* y la 7 a *A. tumefaciens*.

Al haber un nucleótido distinto existiría la probabilidad de que el gen que codifica para el ctRNA pudiera estar mutado en repC de *R. tropici* CFN299 y como es esencial para la replicación de repC (Izquierdo *et al.*, 2005), posiblemente esta sea una razón por la cual no se replica.

Para saber si el gen que codifica para el ctRNA está mutado puede hacerse un experimento en el que se cambie el nucleótido que se encuentran en las otras 7 secuencias, o clonar la región promotora del gen *repC* de *R. tropici* CFN299 al gen *repC* de *S. meliloti* y observar si este gen que sabemos es funcional se expresa, en el caso contrario sabremos que el ctRNA o el promotor está impidiendo la replicación de *repC* de *R. tropici* CFN299.

Al existir dos moléculas silvestres de ctRNA se observa incompatibilidad sin embargo cuando uno de los dos genes que codifican para el ctRNA se encuentra mutado no existe incompatibilidad (Izquierdo *et al.*, 2005). En *R. tropici* CFN299 no hubo desplazamiento del pSym cuando se transfirió la construcción A789, debido a que el gen que codifica para el ctRNA no se está expresando, razón por la cual A789 recombinó en el pSym.

Como se argumentó anteriormente, puede ser el caso de que la proteína RepC de *R. tropici* CFN299 no sea funcional, pero también podría ser que en ausencia de

repABC este origen se replique. Posiblemente *repC* se podría expresar en algún momento de la vida libre o simbiótica de *R. tropici* CFN299.

6.2. Operón *repABC*

En este trabajo se identificó y se obtuvo el operón repABC que es el replicador básico de *R. tropici* CFN299. El gen repC del operón es esencial para la replicación del pSym, sin embargo este gen al clonarlo con una parte del gen repB (construcción A610) no replicó el plásmido en el cual fue clonado. También al clonar el operón con el gen repA interrumpido en la región río arriba (A766) se obtuvo el mismo resultado. Esto indica que los 3 genes están organizados en un operón y que repC no fue capaz de replicarse debido a que no se encontraba el promotor del operón.

El operón funcional *repABC* se obtuvo mediante deleciones producidas por reacciones de digestión en un fragmento mínimo *BamH*I-*Mlu*I de aproximadamente 5.5 kb. Se han reportado fragmentos replicables mínimos estables desde 4.2 kb como el del plásmido pTiB6S3 (Tabata *et al.*, 1989), 4.3 kb como el del pSym p42 de *R. etli* (Ramírez-Romero *et al.*, 2000) hasta 5 kb como el pRL8JI de *R. leguminosarum* (Turner y Young, 1995), lo cual indica que se puede obtener un tamaño menor del operón *repABC* de *R. tropici* CFN299.

El operón *repABC* de *R. tropici* CFN299 contiene al igual que los replicadores de la familia *repC* un gen que codifica para un ctRNA localizado entre *repB* y *repC* el cual se conserva en ambas familias y funciona como sitio de incompatibilidad (Fig. 6.2). También en el operón *repABC* de *R. tropici* CFN299 se localizó otro sitio de incompatibilidad llamado *parS*.

Al transferir la construcción A627, la cual contiene al operón completo y funcional a *R. tropici* CFN299, no se provocó ninguna incompatibilidad con el pSym, esto posiblemente se deba a la acción del mecanismo de recombinación de *R. tropici* CFN299 el cual la evita. Otra posibilidad es que el pSym tenga genes esenciales para la vida libre de *R. tropici* CFN299 y no sea posible deletarlo por completo y como se ha mencionado algunos plásmidos de *Rhizobium* son esenciales en la vida libre de bacterias fijadoras de nitrógeno. Sin embargo, se requiere realizar experimentos en los cuales se inactive el gen *repC* del operón, este experimento nos indicaría además si solamente *repABC* es funcional en el pSym o si bajo estas condiciones el replicador *repC* podría ser funcional.

Discusión

Cuando se conjugó la construcción A627 a *R. tropici* CFN299 recombinó con el pSym formando un cointegrado estable. Este mismo resultado se ha observado con el plásmido pCFN42d de *R. etli* CFN42 el cual forma un cointegrado cuando se le transfiere una construcción conteniendo el replicón de este plásmido. Sólo se puede ejercer incompatibilidad en 2 condiciones: cuando se introduce la construcción en una cepa de *R. etli recA* menos, la cual evita la recombinación y cuando además del replicador básico se incluye en la construcción un sitio de acción (ris) un sitio específico para la recombinasa RinQ en la misma orientación del plásmido parental (Quintero *et al., 2002*). Realizando experimentos, en los cuales se incluyan las 2 condiciones anteriores mencionadas podría obtenerse una cepa de *R. tropici* CFN299 con el pSym curado, y por lo tanto demostrar si el pSym es esencial para la vida libre de la bacteria.

Existe otra razón por la cual se formó un cointegrado de la construcción A627 con su plásmido parental, quizás es debido a que existen genes que tienen influencia en la incompatibilidad y estabilidad del plásmido, y estos no se incluyan en la construcción. Esto se ha observado en los plásmidos pCFN42d y pTi-SAKURA en donde existen elementos genéticos que tienen propiedades para su incompatibilidad y estabilidad y estabilidad y estabilidad.

Actualmente con las secuencias de los genomas de algunos plásmidos depositadas en el GeneBank se puede observar qué genes están cercanos al operón *repABC* y cuáles de ellos se relacionan con un operón *tral-trb*, esto nos permitirá saber que plásmidos están regulados por un mecanismo de respuesta en coro. El operón *repABC* del plásmido Ti de *Agrobacterium* es regulado positivamente por el operón *tral-trb* mediante el mecanismo de respuesta en coro (Li y Farrand, 2000). En el pSym de *R. tropici* CFN299 se encontró un gen *tral* adyacente a *repA* en dirección opuesta con una identidad del 64% con el de *Agrobacterium*, pero mientras no se tenga una región más grande secuenciada donde se localice un gen *trb* no se sabrá si el replicador es regulado por el mecanismo de respuesta en coro o si la secuencia con identidad a *tral* es un pseudogen como es el caso del plásmido pCFN42d que solamente contiene una secuencia con similitud a *tral* pero carece de *trb*. En términos evolutivos, la localización de una parte de la secuencia del operón *tral-trb* cercana al operón *repABC* en algún plásmido significa que estos plásmidos derivan de un mismo ancestro común.

Por otra parte el pSym de *R. tropici* CFN299 podría ser un bireplicón causado por la cointegración de un plásmido entrante con un plásmido residente, pues se ha

reportado que la formación de cointegrados entre plásmidos incompatibles de la familia *Rhizobeaceae* es frecuente (Ramírez-Romero *et al.*, 1997) y estos cointegrados son estables. En este caso uno de los 2 replicadores puede cambiar acumulando mutaciones y perder su funcionalidad sin arriesgar otras funciones codificadas en el plásmido (Cevallos *et al.*, 2002).

Los datos obtenidos en el presente trabajo son útiles para el estudio del funcionamiento del pSym de la cepa de *R. tropici* CIAT899 ya que se localizaron genes ortólogos para *repC* y *repABC* en esta cepa.

VII. CONCLUSIÓN

- Se identificaron 2 orígenes de replicación en el pSym de *R. tropici* CFN299, uno de la familia *repABC* y el otro perteneciente a *repC*.
- Se aislaron, secuenciaron y clonaron ambos replicadores del pSym de *R. tropici* CFN299
- Sólo el replicador de la familia *repABC* fue capaz de replicar autónomamente en *A. tumefaciens*
- Los genes *repA*, *repB* y *repC* se encuentran formando un operón.
- *repC* y *repABC* no producen incompatibilidad con el pSym.

VIII. APÉNDICE

8.1. Medios de cultivo

MEDIO PY (Noel <i>et. al.,</i> 1984)						
Peptona de caseína 5 g						
Extracto de levadura 3 g						
CaCl ₂	1 g					
Agar (en medios sólidos)	15 g					
H ₂ O	Para 1 litro					
Calcio 1M	(1ml/100ml de medio estéril)					

MEDIO LB (Sambrook et. al., 1989)	
Peptona de caseína 10) g
Extracto de levadura 5 g	
NaCl	10 g
Agar (en medios sólidos	s) 15 g
H ₂ O	Para 1 litro

8.2. Protocolos

PURIFICACIÓN DEL DNA (Amersham biosciences, GenomicPrep cells)

Lisis celular

- Agregar 1 mL de células en suspensión, crecidas toda la noche a un tubo de 1.5 mL.
- 2. Centrifugar a 14000 rpm 1 min hasta obtener un pellet. Remover el sobrenadante.
- 3. Adicionar 600 µL de solución para lisar células y pipetear suavemente hasta resuspender.
- 4. Incubar la muestra a 80 °C por 10 min para lisar las células. Enfriar la muestra a temperatura ambiente.

Tratamiento RNasa

- 5. Adicionar 3 µL de solución RNasa A las células lisadas.
- 6. Vortexear la muestra.

Precipitación de proteínas

- 7. Enfriar la muestra a temperatura ambiente.
- 8. Adicionar 200 µL de solución para precipitar proteínas.
- Vortexear a alta velocidad 20 seg para mezclar uniformemente la solución para precipitar proteínas con las células lisadas.
- 10. Centrifugar a 14000rpm por 10 min.

Precipitación del DNA

- Tomar cuidadosamente el sobrenadante que contiene el DNA en un tubo de 1.5 mL que contenga 600 μL de isopropanol al 100%.
- 12. Vortexear la muestra.
- 13. Centrifugar a 14000rpm por 10 min, el DNA debe verse como un pequeño pellet blanco.
- 14. Retirar el sobrenadante y secar el tubo en papel secante. Adicionar 600 μL de etanol al 70%. Invertir el tubo varias veces para lavar el pellet de DNA.
- 15. Centrifugar a 14000rpm 10 min. Cuidadosamente retirar el etanol sin desalojar el pellet de DNA.
- 16. Drenar el tubo en papel secante y secar en savant 15 min.

Hidratación del DNA

- 17. Adicionar 50 μ L de H₂O al pellet de DNA para hidratarlo.
- Calentar a 65 °C por una hora, vortexear para dispersar el DNA y dejar el tubo de DNA hidratando toda la noche a temperatura ambiente.

19. Guardar de 2-8 °C.

PURIFICACIÓN DE PLÁSMIDOS (Roche Applied Science, Germany)

- Centrifugar 1.5 mL de medio liquido con bacterias *E. coli* en tubos eppendorf durante 1 min, retirar el sobrenadante, agregar nuevamente 1.5 mL de medio con bacterias, repetir 3 veces.
- Resuspender el pellet en 250 µL de solución de suspensión buffer/RNasa y vortexear fuertemente.
- 3. Adicionar 250 μ L de lisis buffer, vortexear 1 vez e incubar 5 min a temperatura ambiente.
- 4. Adicionar 350 μL de chilled binding buffer, vortexear e incubar los tubos en hielo durante 5 min.
- 5. Centrifugar 10 min a 14000rpm.
- 6. Insertar un tubo High Pure Filter dentro de un tubo colector por muestra.
- 7. Transferir el sobrenadante sobre el tubo con filtro.
- 8. Centrifugar 1 min a 14000rpm.
- Desconectar el tubo con filtro y desechar el líquido filtrado, conectar de nuevo el tuvo.
- 10. Adicionar 700 µL de Wash buffer II y centrifugar a 14000rpm 1 min.
- 11. Desconectar el tubo con filtro, desechar el líquido filtrado, reconectar, centrifugar 1 min para remover los residuos del Wash buffer II y desechar el liquido filtrado
- 12. Insertar el tubo con filtro dentro de un tubo eppendorf adicionar 100 μ L de H₂O en el tubo con filtro y centrifugar 1 min a 14000rpm, obteniendo finalmente así los plásmidos en el tubo eppendorf.

AMPLIFICACIÓN POR PCR

Para cada muestra a amplificar por PCR se preparan 2 reacciones

Reacción II

2.8 μ L H₂O, 6 μ L Buffer II, 4 μ L dNTPs (0.8mM), 2.2 μ L Mg (1.1mM), 2.5 μ L de cada oligonucleótido, upper primer y lower primer (10 pmol/ μ L).

Reacción I

19 μ L H₂O y 0.5 μ L de DNA polimerasa (0.02 unidades/ μ L).

En un tubo para PCR agregar 2 µL de DNA como templado para un volumen final de

50 μ L, adicionar los 28.5 μ L de reacción del tubo II y después los 20 μ L de reacción del tubo I. Colocar en termociclador sin mezclar las muestras.

TRANSFORMACIÓN BACTERIANA

- **1.** Colocar la ligación en un tubo eppendorf que contiene 100 μL de células competentes, dejar en hielo 45 min.
- 2. Poner a 42 °C por 2 min.
- **3.** Colocar en hielo por 2 min.
- **4.** Transferir la transformación a un tubo de ensaye con 1 mL de medio LB e incubar a 37 °C por una hora en agitación.
- 5. Platear 100 μL de la transformación en cajas petri con los antibióticos necesarios para la selección de cada construcción y cepa, IPTG 24 μg/mL y Xgal 24 μg/mL.
- 6. Incubar toda la noche a 37 °C.
- 7. Seleccionar colonias recombinantes.

CONJUGACIÓN BACTERIANA

- **1.** Crecer las cepas (donadora, ayudadora y receptora) en 5 ml de medio líquido toda la noche a 30°C.
- **2.** Tomar 1.5 mL de cepa receptora y 0.75 mL de las cepa donadora y ayudadora en tubos por separado.
- 3. Centrifugar los tubos a 14000 rpm durante un minuto y retirar el sobrenadante.
- 4. Agregar a los tubos 1 mL de NaCl 0.85%, vortexear y resuspender.
- 5. Centrifugar a 14000 rpm durante un minuto y retirar el sobrenadante.
- 6. Resuspender en 150 μ L de NaCl 0.85%.
- **7.** Tomar 50 μ L de la cepa donadora, 50 μ L de la cepa ayudadora y 100 μ L de la cepa receptora y mezclar en tubos eppendorf (cruzas).
- 8. Colocar la mezcla en medio LB sólido.
- 9. Como control crecer en medio LB sólido 50 µL de cada cepa por separado.
- 10. Incubar todas las cajas con medio sólido a 30° C por 24 horas.
- De las cruzas crecidas tomarlas con un palillo y colocarlas en 1 mL de NaCl 0.85%
- **12.** Vortexear las cruzas y hacer diluciones de 10° a 10^{-3} .
- **13.** Platear 100 µL de cada dilución y controles en cajas con medio líquido y antibioticos.

14. Incubar las cajas a 30° C.

SOUTHERN BLOT

Obtención del gel

- 1. Purificar DNA cepa A303.
- 2. Digerir DNA con enzima de restricción BamHI.
- 3. Correr en gel de agarosa al 0.85% en tris acetatos 1X.
- 4. Teñir con bromuro de etidio y visualizar en transluminador de luz UV.

Depuración del gel

- Sumergir el gel en HCI 0.25 N (20.8 mL de HCl concentrado en 1 lt de H₂O) 15 min a temperatura ambiente dos veces.
- 6. Lavar con H₂O.
- Desnaturalizar sumergiendo el gel en una solución 0.5 N de NaOH (20 g en 1 lt de H₂O) y 1.5 M de NaCl (87.75 g en 1 lt de H₂O) 45 min a temperatura ambiente.
- 8. Lavar con H₂O.
- Sumergir en SSC 6X por 10 min a temperatura ambiente (SSC 20X: NaCl 175.2 g, citrato de sodio 84.2g, adicionar H₂O hasta 1 lt.
- **10.** Transferir el DNA digerido del gel a una membrana de nylon. El gel se coloca en un sistema de transferencia como el de la imagen. Transferir durante 24 horas y cambiar frecuentemente el papel absorbente.

Transferencia de DNA a membrana



11. Marcar los carriles en la membrana de nylon y lavar membrana con SSC 2X

por 10 min.

- **12.** Secar a 65 °C 20 min.
- **13.** Fijar DNA con luz UV 5 min.
- 14. Prehibridizar membrana con solución de hibridación (Rapad-hyb buffer Amersham Pharmacia Biotech) por lo menos 15 min antes de colocar la sonda marcada con ³²P radioactivo.

Marcaje radioactivo

- **1.** De la sonda (*repBC*) tomar 5 μ L y agregarlos a tubos eppendorf.
- 2. agregar al tubo 40 µL de EDTA 10:1 pH8.
- 3. vortexear y centrifugar solo para bajar la muestra.
- 4. Calentar a 95 °C por 10 min para desnaturalizar la muestra.
- 5. Enfriar en hielo por 5 min.
- 6. Centrifugar poco solo para bajar la muestra.
- Pasar los 45 µL de solución a los tubos de reacción Rediprime (Solución de dATP, dGTP y dTTP, enzima klenow libre de exonucleasa y oligonucleótidos de 6 nucleótidos).
- 8. Mezclar.
- **9.** Adicionar 5 µL de ³²P y mezclar hasta obtener un color púrpura.
- **10.** Centrifugar.
- 11. Incubar 2 horas a 37 °C.
- **12.** Adicionar 600 μL de Magic probe (acetato de amonio 2.5 M, DNA de esperma de salmón 0.6 mg/mL, EDTA 20 mM).
- **13.** Adicionar 1 mL de isopropanol al 100%.
- 14. Mezclar con vortex.
- **15.** Centrifugar 10 min y tirar sobrenadante.
- **16.** Resuspender en 100 µL de NaOH 0.1 M y vortexear mucho.
- **17.** Contar cuentas por minuto en un contador de cuentas radioactivas (probe count).
- 18. Colocar los 100 µL en un vial con 2 mL de solución para prehibridizar.
- **19.** Adicionar sondas ya marcadas a las membranas prehibridizadas y dejar toda la noche hibridizando.
- 20. Lavar membrana con solución SSC 2X y SDS 0.1% 10 min, SSC 1X y SDS 0.1% 10 min, SSC 0.5X y SDS 0.1% 10 min, SSC 0.1X y SDS 0.1% 10 min, SSC 0.1X y SDS 0.1% 10 min a 65 °C y finalmente SSC 0.1X sin SDS hasta

eliminar el detergente.

- **21.** Secar a 65 °C.
- 22. Exponer membrana a una película Kodak.
- 23. Revelar.

IX. REFERENCIAS

- Barran LR, Ritchot N, Bromfield ES: *Sinorhizobium meliloti* plasmid pRm1132f replicates by a rolling-circle mechanism. *J Bacteriol* 2001, 183:2704–2708.
- Bartosik D, Baj J, Wlodarczyk M: **Molecular and functional analysis of pTAV320, a repABC-type replicon of the** *Paracoccus versutus* **composite plasmid pTAV1.** *Microbiology* 1998,**144:**3149–3157.
- Bartosik D, Szymanik M, Wysocka E: Identification of the partitioning site within the repABC-type replicon of the composite *Paracoccus versutus* plasmid pTAV1. *J Bacteriol* 2001, **21**:6234–6243.
- Báscones E, Imperial J, Ruiz-Argüeso T, y Palacios J.M: Generation of new hydrogen-recycling *Rhizobiaceae* strains by introduction of a novel *hup* minitransposon. *App Env Microbiol* 2000, **66**: 4292-4299.
- Bouet JY, Funnell BE: P1 ParA interacts with the P1 partition complex at *parS* and an ATP-ADP switch controls ParA activities. *EMBO J* 1999, **18**:1415-1424.
- Brom S, García de los Santos A, Cervantes L, Palacios R, Romero D: In *Rhizobium etli* symbiotic plasmid transfer, nodulation competitivity and cellular growth require interaction among different replicons. *Plasmid* 2000, **44**:34-43.
- Burgos PA, Velázquez E, Toro N: Identification and distribution of plasmid-Tipe A replicator region in Rhizobia. *Mol Plant Microbe Interact* 1996, **9:**843-846.
- Cevallos MA, Porta H, Izquierdo J, Tun-Garrido C, García de los Santos A, Dávila G, Brom S: *Rhizobium etli* CFN42 contains at least three plasmids of the repABC family: a structural and evolutionary analysis. *Plasmid* 2002, **48**:104–116.
- Cevallos MA, Cervantes-Rivera R, Gutiérrez-Ríos RM: **The** *repABC* plasmid family. *Plasmid* 2008, **60:**19-37.
- Chai Y, Winans SC: A small antisense RNA downregulates expression of an essential replicase protein of an Agrobacterium tumefaciens Ti plasmid. *Mol Microbiol* 2005, **56:**1574-1585.
- Cho H, Winans SC: VirA and VirG activate the Ti plasmid repABC operon, elevating plasmid copy number in response to wound-released chemical signals. *Proc Natl Acad Sci USA* 2005, **102:**14843–14848.
- Del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R: **Replication and Control of circular bacterial plasmids.** *Microbiol* and *Mol Biol Rev* 1998, **62:**434-464.
- Ebersbach G, Gerdes K: Bacterial mitosis: partitioning protein ParA oscillates in spiral- shaped structures and positions plasmids at mid-cell. *Mol Microbiol* 2004, **52:**385-398.

- Escobar MA, Dandekar AM: *Agrobacterium tumefaciens* as an agent of disease. *Trends Plant Sci* 2003, 8:380–386.
- Figurski DH, Helinski DR: **Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans.** *Proc Natl Acad Sci USA* 1979, **76:**1648-1652.
- Finan TM, Weidner S, Wong K, Buhrmester J, Chain P, Vorhölter FJ, Hernandez-Lucas I, Becker A, Cowie A, Gouzy J, Golding B, Pühler A: The complete sequence of the 1683-kb pSymB megaplasmid from the N₂-fixin endosymbiont Sinorhizobium meliloti. Proc Natl Acad Sci USA 2001, 98:9889-9894.
- Freiberg C, Fellay R, Bairoch A, Broughton J, Rosenthal A, Perret X: Molecular basis of symbiosis between *R. tropici* and legumes. *Nature* 1997, **387**:394-401.
- Funnell BE: Partition-mediated plasmid pairing. Plasmid 2005, 53:119-125.
- Galibert F, Finan TM, Long SR, Pühler A, Abola A, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe M, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dréano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thébault P, Vandenbol M, Vorhölter FJ, Weidner S, Wells DH, Wong K, Chen Yeh K, Batut J: The composite genome of the legume symbiont Sinorhizobium meliloti. Science 2001, 293:668-672.
- García de los Santos A, Brom S, Romero D: Rhizobium plasmids in bacteria-legume interactions. *World J Microbiol Biotechnol* 1996, **12**:119-125.
- Geniaux E, Flores M, Palacios R, Martínez E: **Presence of megaplasmids in** *Rhizobium tropici* and furhter evidence of differences between the two *R. tropici* subtypes. *Int J Sist Bacteriol* 1995, **47:**392- 394.
- Gerdes K, Moller-Jensen J, Jensen RB: **Plasmid and cromosome partition: surprises** from phylogeny. *Mol Microbiol* 2000, **37:**455–466.
- González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci USA* 2006, 10:3834-3839.
- González V, Bustos P, Ramírez-Romero MA, Medrano-Soto A, Salgado H, Hernández González I, Hernández-Celis JC, Quintero V, Moreno-Hagelsieb G, Girar L, Rodríguez O, Flores M, Cevallos MA, Collado-Vides J, Romero D, Dávila G: The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biology* 2003, 4: R 36

- Hernández-Lucas I, Mavingui P, Finan T, Chain P, Martínez-Romero E: In vivo cloning strategy for *Rhizobium* plasmids. *Biotechniques* 2002, **33**:728-788.
- Hernández- Lucas I, Pardo MA, Segovia L, Miranda J, Martínez-Romero E: *Rhizobium tropici* cromosomal citrate synthase gene. *Appl Environ Microbiol* 1995a, 61:3992-3997.
- Hernández- Lucas I, Segovia L, Martínez-Romero E, Pueppke SG: **Phylogenetic** relationships and host range of *Rhizobium spp.* that nodulate *Phaseolus vulgaris L. Appl Environ Microbiol* 1995b, **61:**2775-2779.
- Higashi S: Transfer of clover infectivity of *R. trifolli* to *R. phaseoli* as mediated by an episomic factor. *J Gen Appl Microbiol* 1967, **13**:391–403.
- Hungria M, Vargas MAT: Environmental factors affecting N₂ fixation in grain legumes in the tropics, with an emphasis on Brazil. *Field Crops Res* 2000, **65:**151-164.
- Hynes MF, McGregor NF: Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen fixing nodules by *Rhizobium leguminosarum*. *Mol Microbiol* 1990, **4:**567–574.
- Izquierdo J, Venkova-Canova T, Ramírez-Romero MA, Téllez-Sosa J, Hernández-Lucas I, Sanjuán J, Cevallos MA: An antisense RNA plays a central role in the replication control of a *repC* plasmid. *Plasmid* 2005, **54:**259–277.
- Jensen RB, Lurz R, Gerdes K: Mechanism of DNA segregation in prokaryotes: replicon pairing by parC of plasmid R1. Proc Natl Acad Sci USA 1998, 95:8550-8555.
- Li P, Farrand SK: The replicator of the nopaline-type Ti plasmid pTiC58 is a member of the *repABC* family and is influenced by the TraR-dependent quorumsensing system. J Bacteriol 2000, 182:179–188.
- Khan SA: Rolling-circle replication of bacterial plasmids. *Microbiol Mol Biol Rev* 1997, 61:442-455.
- Martínez-Romero E, Caballero-Mellado J: *Rhizobium* phylogenies and bacterial genetic diversity. *Crit Rev Plant Sci* 1996, **15:**113-140.
- Martínez-Romero E, Palacios R, Sánchez F: Nitrogen-fixing nodules induced by *Agrobacterium tumefaciens* harboring *Rhizobium phaseoli* plasmids. J *Bacteriol* 1987, 6:2828–2834.
- Martínez-Romero E, Rosenblueth M: Increased bean (*Phaseolus Vulgaris* L.) nodulation competitiveness of genetically modified *Rhizobium* strains. *App Env Microbiol* 1990, **56:**2384-2388.
- Martínez-Romero E, Segovia L, Mercante FM, Franco AA, Graham P, Pardo MA: *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris L.* beans and *Leucaena sp* trees. *Int J Sist Bacteriol* 1991, **41:**417-426.

- Mavingui P, Flores M, Romero D, Martínez-Romero E, Palacios R: Generation of *Rhizobium* strains with improved symbiotic properties by random DNA amplification (RDA). *Nature Biotechnol* 1997, **15**:564-569.
- Mercado-Blanco J, Olivares J: **The large nonsymbiotic plasmid pRmeGR4a of** *Rhizobium meliloti* **GR4 encodes a protein involved in replication that has homology with the RepC protein of** *Agrobacterium* **plasmids.** *Plasmid* 1994, **32:**75–79.
- Mercado-Blanco J, Olivares J: Stability and transmissibility of the cryptic plasmids of *Rhizobium meliloti* GR4. *Arch Microbiol* 1993, 160:477–485.
- Mostasso L, Mostasso FL, Díaz BG, Vargas MAT, Hungria M: Selection of bean (*Phaseolus vulgaris* L.) rhizobial strains for the brazilian cerrados. *Field Crops Res* 2002, **73:**121.
- Noel KD, Sánchez A, Fernández L, Leemans J, Cevallos MA: *Rhizobium phaseoli* symbiotic mutans with transposon Tn5 insertions. *J Bacteriol* 1984,158:148-155.
- Pappas KM, Winans SC: The RepA and RepB autorepressors and TraR play opposing roles in the regulation of a Ti plasmid *repABC* operon. *Mol Microbiol* 2003, **49:**441–455.
- Perret X, Freiberg C, Rosenthal A, Broughton WJ, Feilla R: **Highresolution transcriptional analysis of the symbiotic plasmid of Rhizobium sp. NGR234.** *Mol Microbiol* 1999, **32:**415–425.
- Quintero V, Cevallos MA, Dávila G: A site-specific recombinase (RinQ) is required to exert incompatibility towards the symbiotic plasmid of *Rhizobium etli*. *Mol Microbiol* 2002, **46**:1023-1032.
- Ramírez-Romero MA, Bustos P, Girard L, Rodríguez O, Cevallos MA, Dávila G: Sequense, localization and characteristics of the replicator region of the symbiotic plasmid of *Rhizobium etli*. *Microbiology* 1997, **143**:2825-2831.
- Ramírez-Romero MA, Soberón N, Pérez-Oseguera A, Téllez-Sosa J, Cevallos MA: Structural elements required for replication and incompatibility of the *Rhizobium etli* symbiotic plasmid. *J Bacteriol* 2000, **182:**3117–3124.
- Ramírez-Romero MA, Téllez-Sosa J, Barrios H, Pérez-Oseguera A, Rosas V, Cevallos MA: **RepA negatively autoregulates the transcription of the** *repABC* operon of **the** *Rhizobium etli* symbiotic plasmid basic replicon. *Mol Microbiol* 2001, **42:**195–204.
- Ramírez-Trujillo JA, Encarnación S, Salazar E, García de los Santos A, Dunn MF, Emerich DW, Calva E, Hernández-Lucas I: **Functional characterization of** *Sinorhizobium meliloti* acetate metabolism genes *aceA*, *SMc00767* and *glcB*. *J Bacteriol* 2007,**189:**5875-5884.

- Rey L, Murillo J, Hernando Y, Hidalgo E, Cabrera E, Imperial J, Ruiz-Argüeso: **Molecular** analysis of a microaerobically induced operon required for hydrogenase in *Rhizobium leguminosarum biovar viciae*. *Mol Microbiol* 1993, **8**:471-481.
- Rogel MA, Hernández-Lucas I, Kuykendall D, Balkwill D, Martínez-Romero E: Nitrogenfixing nodules with *Ensifer adhaerens* harboring *Rhizobium tropici* symbiotic plasmids. *Appl Environ Microbiol* 2001, **67:**3264-3268.
- Rosenberg C, Huguet T: The pAtC58 plasmid of Agrobacterium tumefaciens is not essential for tumour induction. *Mol Gen Genet* 1984, **196:**533-536.
- Rosenblueth M, Hynes F, Martínez-Romero E: *Rhizobium tropici teu* genes involved in specific uptake of *Phaseolus vulgaris* bean-exudate compounds. *Mol Gen Genet* 1998, **258**:587-598.
- Sambrook J, Fritsch EF, Maniatis TA: 1989. **Molecular cloning: a laboratory manual.** 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Simon R, Priefer U, Pühler A: A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. Bio/Technology 1983, 1:784–791.
- Schofield PR, Gibson AH, Dudman FW, Watson JM: Evidence for genetic exchange and recombination of *Rhizobium* symbiotic plasmids in a soil population. *Appl Environ Microbiol* 1987, **53**:2942-2947.
- Soberón N, Venkova-Canova T, Ramírez-Romero MA, Tellez-Sosa J, Cevallos MA: Incompatibility and the partitioning site of the *repABC* basic replicon of the symbiotic plasmid from *Rhizobium etli*. *Plasmid* 2004, **51**:203–216.
- Southern EM: Detection of specific sequens among DNA fragments separated by gel electrophoresis. J Mol Biol 1975, 98:503-517.
- Spaink HP: Root nodulation and infection factors produced by Rhizobial bacteria. Annual Rev in Microbiol 2000, **54:**257-288.
- Tabata S, Hooykaas PJ, Oka A: Sequence determination and characterization of the replicator region in the tumorinducing plasmid pTiB6S3. *J Bacteriol* 1989, 171:1665–1672.
- Turner SL, Young JPW: The replicator region of the *Rhizobium leguminosarum* cryptic plasmid pRL8JI. *FEMS Microbiol Lett* 1995, **133:**53–58.
- Venkova-Canova T, Soberón NE, Ramírez-Romero MA, Cevallos MA: **Two discrete** elements are required for the replication of a *repABC* plasmid: an antisense RNA and a stem–loop structure. *Mol Microbiol* 2004, **54:**1431–1444.
- Villadas PJ, Velázquez E, Martínez-Molina E, Toro N: Identification of nodule-dominant *Rhizobium meliloti* strains carrying pRmeGR4b type plasmid within indigenous soil populations by PCR using primers derived from specific DNA sequences. *FEMS Microbiol Ecol* 1995,17:161–168.

- Viera J, Messing J: New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene* 1991, 100:189-194.
- Wang E, Martínez-Romero J, López-Lara I: *Rhizobium* y su destacada simbiosis con plantas. (http://www.biblioweb.dgsca.unam.mx/libros/microbios/Cap8/) 2001.
- Waters JK, Hughes BL, Purcell CL, Gerhardt KO, Mawhinney TP, Emerich DW: Alanine, not ammonia, is excreted from N₂-fixing soybean nodule bacteroids. *Proc Natl Acad Sci USA* 1998, **95**:12038-12042.
- Woese CR, Kandler O, Wheelis ML: Towards a natural system of organisms. Proposal for the domains Archaea, Bacteria and Eukarya. Proc Natl Acad Sci USA 1990, 87:4576-4579.
- Young JW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson ARJ, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach A, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J: The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* 2006, **7**:R34.
- Zhu J, Oger PM, Schrammeijer B, Hooykaas PJ, Farrand SK, Winans SC: **The bases of** crown gall tumorigenesis. *J Bacteriol* 2000, **182**:3885–3895.