Facultat de Ciències Biològiques

Departament de Bioquímica i Biologia Molecular

Programa de Doctorado en Biotecnología

Characterization of the mutualistic endosymbiosis between intracellular bacteria and mealybugs (Hemiptera: Pseudococcidae)

Memoria presentada por D. Sergio López Madrigal, candidato al grado de Doctor por la Universitat de València.

Co-directores:

Dr. Andrés Moya Simarro

Dra. M. Rosario Gil García

Valencia, 2015

El Dr. Andrés Moya Simarro, Catedrático del Departamento de Genética de la Facultat de Ciències Biològiques (Universitat de València) y la Dra. M. Rosario Gil García, Profesora Titular del Departamento de Genética de la Facultat de Ciències Biològiques (Universitat de València)

AUTORIZAN la presentación de la memoria "Characterization of the mutualistic endosymbiosis between intracellular bacteria and mealybugs (Hemiptera: Pseudococcidae)" y CERTIFICAN que los resultados que incluye fueron obtenidos bajo su co-dirección en el Institut Cavanilles de Biodiversitat i Biologia Evolutiva por D. Sergio López Madrigal.

Y para que conste, firman el presente certificado.

En Paterna, a de de 2015.

Dr. Andrés Moya Simarro Dra. M. Rosario Gil García

Agradecimientos

La realización de una Tesis Doctoral es un proceso largo y difícil. Requiere un gran esfuerzo por parte del doctorando y, por supuesto, la colaboración de muchas personas. Es por eso que, si bien no forma parte de la composición formal del documento, no creo que nadie pueda escribir una tesis sin empezar agradeciendo la colaboración, los consejos, la compañía o el aliento aportado por tantas personas a lo largo del camino.

Agradezco a amigos, compañeros de clase y profesores de la Facultat de Ciències Biològiques de la Universitat de València, que hicieron de mis años como estudiante de Licenciatura y Master una experiencia tan agradable y enriquecedora. A quienes alimentaron, sin saciarlas, mis ganas de saber, de estudiar, de descubrir y que me empujaron a emprender este camino con la satisfacción de sentir que hacía exactamente aquello que quería hacer.

Agradezco a mis co-directores, el Dr. Andrés Moya y la Dra. Rosario Gil, por la confianza depositada y por haberme ofrecido la oportunidad de empezar mi andadura en el mundo de la investigación científica trabajando con ellos en el Institut Cavanilles de Biodiversitat i Biologia Evolutiva (ICBiBE). Quiero hacer extensivo este agradecimiento a la Dra. Amparo Latorre, por su generosidad y su tiempo. Han sido muchas las personas del ICBiBE que han hecho de estos años -con su colaboración, su compañía y su amistad- un tiempo inolvidable. A todas ellas, muchas gracias de todo corazón. Gracias también al Dr. Ferran Garcia, a la Dra. Antonia Soto y al Dr. Aleixandre Beltrà del Instituto Agroforestal Mediterráneo (Universitat Politècnica de València), a D. Alberto García del Centro de Sanidad Vegetal (Generalitat Valenciana) y a los miembros del Servei de Genòmica del SCSIE (Universitat de València). Sin vuestra amable colaboración este trabajo tampoco habría sido posible.

Quiero agradecer al Dr. Abdelaziz Heddi y a los miembros de La Unidad BF2I (Biologie Fonctionnelle, Insectes et Interaction), por acogerme tan calurosamente durante el frío invierno lionés y poner a mi disposición su tiempo, su experiencia y sus recursos. Mi paso por Lyon no pudo ser más gratificante: alcancé con éxito el objetivo de mi estancia de investigación y regresé a España con un buen número de nuevos y buenos amigos. *Merci à tous!*

Me gustaría cerrar este apartado dando las gracias a mi familia. No solo por dejarme ser, desde una perspectiva biológica, sino por permitirme crecer en un ambiente en el que *trabajo*, *esfuerzo* y *educación* no son solo palabras, sino sólidos principios para la vida. Quiero agradecer muy especialmente a mis padres, sin cuya dedicación y cariño -pero también sin cuyo ejemplo- no habría podido llegar hasta este lugar, hasta este momento. Y a Andrea, claro: por la inspiración, por su comprensión infinita y por tanto aliento.

Table of Contents

1. RESUMEN EN CASTELLANO

1.1. Introducción

Simbiosis, del griego *sym* “con” y *biosis* “vivir”, hace referencia a la asociación estable entre individuos de dos o más especies (simbiontes) que muestran interdependencia a cualquier nivel biológico. En función del efecto sobre la eficacia biológica de los simbiontes implicados, distinguimos tres subtipos principales de simbiosis: (1) **comensalismo**, cuando uno de los simbiontes se beneficia de la asociación sin causar perjuicio al otro; (2) **parasitismo**, si el beneficio de uno de los simbiontes provoca una caída en la eficacia del otro; (3) **mutualismo**, cuando la asociación beneficia a ambos simbiontes. Por otra parte, hablamos de **ectosimbiontes** cuando uno de los simbiontes coloniza la superficie corporal de su hospedador o **endosimbiontes** si mantiene un estilo de vida intracelular. Finalmente, la simbiosis puede considerarse **facultativa** u **obligatoria**, según la capacidad de los simbiontes para sobrevivir al margen de dicha asociación.

La ubiquidad de las asociaciones simbióticas en las ramas principales del árbol de la vida evidencia la relevancia global de la simbiosis en la evolución de la vida (Figura 2, página 15) (Moya et al., 2008; McFall-Ngai, 2008; Moya and Peretó, 2011). Las asociaciones simbióticas protagonizadas por procariotas son especialmente abundantes, aunque aquellas que se establecen entre procariotas y eucariotas son las más estudiadas. La asociación con simbiontes bacterianos ha proporcionado a los organismos eucariotas, funcionalmente limitados, una amplia gama de habilidades metabólicas (Schink, 1997; Minic and Hervé, 2004; Zientz et al., 2004; Stewart et al., 2005; Kneip et al., 2007). La simbiosis con bacterias intracelulares se considera un factor clave en el éxito adaptativo de los insectos, el linaje animal más diverso conocido (Mora et al., 2011). Según su relevancia respecto de la supervivencia del hospedador, los simbiontes bacterianos se clasifican en dos grandes categorías: endosimbiontes obligatorios o primarios (P-endosimbiontes) y simbiontes facultativos o secundarios (S-simbiontes) (Baumann, 2005).

Los P-endosimbiontes son esenciales para la supervivencia y la reproducción del insecto y, por tanto, están fijados en la población del hospedador (Houk and Griffiths, 1980). Se trata de bacterias intracelulares, no cultivables, que habitan en el citoplasma de células especializadas (bacteriocitos), frecuentemente agregadas en forma de órgano (bacterioma). Los P-endosimbiontes se transmiten verticalmente a la descendencia por vía materna, de modo que bacteria e insecto coevolucionan (Buchner, 1965; Baumann, 2005). Muchos P-endosimbiontes complementan las dietas restringidas de sus hospedadores, observándose con frecuencia un nexo entre la composición de la dieta de éstos y el papel del P-endosimbionte correspondiente. Por ejemplo, los P-endosimbiontes de los insectos chupadores de savia vegetal, muy pobre en compuestos nitrogenados, suelen complementar la dieta de su hospedador con amino ácidos esenciales y vitaminas (Nogge, 1981; Wicker and Nardon, 1982; Prosser and Douglas, 1992; Shigenobu et al., 2000; Tamas et al., 2002; van Ham et al., 2003; Pérez-Brocal et al., 2006; Takiya et al., 2006; McCutcheon et al., 2009a; Lamelas et al., 2011; Degnan et al., 2011; Urban and Cryan, 2012). No obstante, algunos P-endosimbiontes intracelulares aportan ventajas

adaptativas también en insectos omnívoros (Gil et al., 2003; Davidson et al., 2003; Wolschin et al., 2004; Feldhaar et al., 2007; López-Sánchez et al., 2009). En cualquier caso, los P-endosimbiontes generalmente sufren una drástica reducción en el tamaño de sus genomas y un fuerte incremento del contenido en AT de éstos. La reducción genómica comporta la pérdida de una amplia gama de funciones génicas innecesarias y/o redundantes vinculada a la riqueza y estabilidad del nicho intracelular (Moya et al., 2008). Por otra parte, el enriquecimiento genómico en AT se ha relacionado con el coste energético diferencial de la biosíntesis de nucleosidos trifosfato (Rocha and Danchin, 2002) y/o con el sesgo mutacional GC→AT y la pérdida de sistemas de reparación de ADN asociada a la evolución reductiva (Lind and Andersson, 2008; Hershberg and Petrov, 2010).

Por el contrario, los S-simbiontes no son esenciales para la supervivencia y reproducción del hospedador. No son necesariamente bacterias intracelulares ni, de serlo, se encuentran exclusivamente en el citoplasma de los bacteriocitos (Dale and Welburn, 2001; Moran et al., 2005; Brumin et al., 2011). Puesto que pueden ser transferidos horizontalmente entre especies hospedadoras, los S-simbiontes y los insectos suelen presentar incoherencia filogenética (Russell et al., 2003; Nováková et al., 2009). En caso de que, eventualmente, el hospedador requiriera funciones esenciales codificadas por el S-simbionte, tal asociación pasaría a ser necesaria y permanente. El P-simbionte original y el S-simbionte pasan entonces a ser simbiontes co-primarios y el sistema se enfrenta a dos posibles soluciones evolutivas: complementación y reemplazamiento. El primer caso se da cuando las necesidades metabólicos del sistema implican la participación de dos o más endosimbiontes co-primarios (Takiya et al., 2006; McCutcheon and Moran, 2007; Gosalbes et al., 2008; McCutcheon et al., 2009a; McCutcheon et al., 2009b; McCutcheon and Moran, 2010; Lamelas et al., 2011; Urban and Cryan, 2012; Morris et al., 2012; Sachs and Hollowell, 2012). Por el contrario, el reemplazamiento tiene lugar cuando la evolución reductiva afecta principalmente a uno de los endosimbiontes co-primarios, facilitando su extinción (Lefèvre et al., 2004; Moya et al., 2009; Koga et al., 2013).

Las cochinillas algodonosas (Hemiptera: Pseudococcidae), cuyo nombre hace referencia a las secreciones algodonosas que cubren el cuerpo de ninfas y hembras adultas, son una plaga común de cultivos agrícolas y plantas ornamentales (Hardy et al., 2008; ScaleNet: http://www.sel.barc.usda.gov/scalenet/scalenet.htm). Las hembras adultas, con 1-3 mm de longitud, presentan una forma ovoide similar a la de los estadios ninfales precedentes. Por el contrario, los machos son insectos alados significativamente más pequeños (Figura 4, página 25). Como en el caso de otros insectos, temperatura (Narai and Murai, 2002; Goldasteh et al., 2009) y plantas hospedadoras (Ben-Dov, 1994; Malleshaiah et al., 2000; Laflin and Parella, 2004; Polat et al., 2007) tienen gran influencia sobre los parámetros de historia de vida de las cochinillas algodonosas. La plasticidad fenotípica y el alto grado de similitud morfológica entre taxones ha sido un obstáculo para el desarrollo de la sistemática de estos insectos, originalmente basada en caracteres cuticulares microscópicos (Cox, 1983; Millar, 2002; Watson and Kubiriba, 2005). Solo recientemente se han identificado diversos marcadores moleculares (Beuning et al., 1999; Thao et al., 2002; Downie and Gullan, 2004; Baumann et al., 2005; Hardy et al., 2007; Hardy et

al., 2008; Rung et al., 2008) de gran utilidad para la identificación a nivel de especie (Beuning et al., 1999; Demontis et al., 2007; Rung et al., 2008; Saccaggi et al., 2008; Daane et al., 2011) y/o para el estudio de las relaciones evolutivas entre éstas (Downie and Gullan, 2004; Hardy et al., 2008). Análisis filogenéticos basados en datos morfológicos y/o moleculares sugieren que la familia Pseudococcidae está compuesta por las subfamilias Pseudococcinae y Phenacoccinae, revelando además la polifilia de los géneros *Dysmicoccus* (Ferris) y *Pseudococcus* (Westwood) en la primera y del género *Phenacoccus* (Cockerell) en la última (Hardy et al., 2008).

Al igual que otros chupadores de savia pertenecientes al suborden Sternorrhyncha (áfidos, psílidos, moscas blancas y otros insectos escama; Gullan and Martin, 2009), las cochinillas algodonosas mantienen simbiosis mutualista con bacterias intracelulares, que albergan en un bacterioma oval de localización abdominal (Fukatsu and Nikoh, 2000; von Dohlen et al., 2001; Gatehouse et al., 2012) y que complementan su dieta nutricionalmente desequilibrada (Köhler and Schwartz, 1962; Kono et al., 2008). En este contexto, las especies de las subfamilias Phenacoccinae y Pseudococcinae presentan claras diferencias (Buchner, 1965; Tremblay, 1989; Hardy et al., 2008; Koga et al., 2012). La betaproteobacteria "*Candidatus* Tremblaya phenacola" es el P-endosimbionte de cochinillas de la subfamilia Phenacoccinae, salvo en (al menos) las especies del género *Rastrococcus* y la tribu *Rhizoecini*, donde ha sido reemplazada por endosimbiontes del *phylum* Bacteroidetes (Figura 11, página 38) (Gruwell et al., 2010). Por el contrario, la betaproteobacteria "*Candidatus* Tremblaya princeps", clado hermano de "*Ca*. Tremblaya phenacola", es el P-endosimbionte de las cochinillas de la subfamilia Pseudococcinae (Munson et al., 1992; Thao et al., 2002; Downie and Gullan, 2005; Baumann and Baumann, 2005). A diferencia de "*Ca*. Tremblaya phenacola", "*Ca*. Tremblaya princeps" no está sola en los bacteriocitos de la mayor parte de cochinillas de la subfamilia Pseudococcinae (Kantheti et al., 1996; Fukatsu and Nikoh, 2000), sino que ha sufrido múltiples eventos de infección por parte de diferentes gammaproteobacterias, dando lugar a consorcios bacterianos cuya organización estructural no tiene precedentes (Figura 6, página 29) (von Dohlen et al., 2001; Thao et al., 2002; Kono et al., 2008; Gatehouse et al., 2012; Koga et al., 2012). La abundancia relativa de beta y gamma-endosimbiontes en el bacterioma, su dinámica poblacional a lo largo del ciclo biológico del insecto hospedador (Kono et al., 2008) y su transmisión coordinada a la siguiente generación (von Dohlen et al., 2001; Thao et al., 2002) sugieren que ambos miembros del consorcio están íntimamente relacionados, cualquiera que sea la filiación filogenética del gamma-endosimbionte correspondiente. Finalmente, la caracterización genómica parcial de la cepa de "*Ca*. Tremblaya princeps" perteneciente a la cochinilla de la piña *Dysmicoccus brevipes* (Cockerell) reveló características genómicas atípicas para un P-endosimbionte con 100-200 millones de años de antigüedad (alto contenido en GC, baja densidad génica, *loci* parálogos persistentes e idénticos) (Baumann et al., 2002).

1.2. Objetivos

Con el fín de inferir las bases moleculares de la endosimbiosis mutualista en cochinillas algodonosas, nos planteamos dos objetivos principales:

1. Secuenciación y ensamblaje del genoma completo del beta- y el gamma-endosimbionte de *Planococcus citri* (Risso), la cochinilla algodonosa de los cítricos.
2. Anotación y análisis functional de los genomas de los miembros de dicho consorcio endosimbionte.

Como consecuencia de los resultados obtenidos, nos planteamos objetivos adicionales:

3. Estudio de la comunicación molecular entre los miembros del consorcio endosimbionte en *P. citri*.
4. Análisis de la biosíntesis de aminoácidos esenciales en especies adicionales de cochinillas algodonosas pertenecientes a las subfamilias Pseudococcinae y Phenacoccinae.
5. Exploración de las bases moleculares del nexo entre la endosimbiosis anidada y la evolución reductiva atípica de "*Ca*. Tremblaya princeps" en cochinillas de la subfamilia Pseudococcinae.

1.3. Metodología

Caracterización genómica: secuenciación, ensamblaje, anotación y análisis funcional

Llevamos a cabo una extracción de ADN total ($_T$DNA) (CTAB; Ausubel, 1999) a partir de 20-30 hembras adultas evisceradas de la cepa PCVAL de *P. citri*. Los insectos procedían de una población de laboratorio generada a partir de una muestra tomada en el Jardín Botánico de la Universitat de València (Valencia, España). Pirosecuenciamos media placa de *single-ends* y un cuarto de placa de *paired-ends* con la plataforma GS-FLX (454 Life Sequencing Inc., Branford, CT) disponible en el Área de Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO, Valencia). Las lecturas fueron filtradas (Galaxy; http://galaxyproject.org) antes de llevar a cabo el ensamblaje automático (MIRA; http://sourceforge.net/projects/mira-assembler). Los *contigs* resultantes fueron manualmente editados con el programa Gap4 (Staden package; Staden et al., 1998) y los *gaps* residuales del genoma del gamma-endosimbionte de *P. citri* cerrados mediante PCR y secuenciación ABI en el servicio de secuenciación de la Universitat de València. El programa OriginX (Worning et al., 2006) se usó para determinar el *ori*C putativo de ambos genomas. Los genes de RNA fueron anotados mediante los programas ARAGORN (Laslett y Canback, 2004), tRNAscan (Lowe y Eddy, 1997) y Rfam (Gardner et al., 2009). BASys (Bacterial Annotation

System; Van Domselaar et al., 2005) y RAST (Aziz et al., 2008) se usaron para anotar los genes codificantes. Unipro UGENE (Okonechnikov et al., 2012) se aplicó para detectar secuencias repetitivas con una longitud mínima de 20 pb, mientras que Phobos (Kraemer et al., 2009) se usó para estudiar la abundancia relativa de microsatélites. Artemis (Carver et al., 2008) y MEGA (Tamura et al., 2011) se usaron para describir parámetros genómicos generales, caracterizar las secuencias repetitivas y calcular el sesgo en el uso de codones. El análisis funcional se llevó a cabo mediante los programas Blast2Go (http://www.blast2go.com) y KAAS (Moriya et al., 2007), con el apoyo de las bases de datos BioCyc (http://biocyc.org), KEEG (http://www.genome.jp/kegg) y BRENDA (http://www.brenda-enzymes.info). Los programas PRED-TAT (Bagos et al., 2010) y SignalP (Petersen et al., 2011) se utilizaron para escanear las proteínas codificadas por el gamma-endosimbionte de *P. citri* en busca de péptidos señal (SP).

Cribado genético

El análisis funcional de los genomas completos del beta- y el gamma-endosimbionte de *P citri* reveló la presencia de un gran número de genes implicados en rutas de recombinación homóloga (HR) o de biosíntesis de aminoácidos esenciales, que podrían haber sido determinantes en la evolución reductiva atípica de "*Ca*. Tremblaya princeps". Con objeto de profundizar en dicha hipótesis, estudiamos la presencia de algunos de estos genes en los sistemas endosimbiontes de diversas especies de pseudocóccidos pertenecientes a la subfamilia Pseudococcinae (*Dysmicoccus boninsis* Kuwana, *Planococcus ficus* Signoret, *Pseudococcus longispinus* Targioni-Tozzeti y *Pseudococcus viburni* Signoret) y Phenacoccinae (*Phenacoccus peruvianus* Granara de Willink y *Phenacoccus madeirensis* Green), respectivamente. Los genes de HR estudiados son *recA* (que codifica la proteína de recombinación RecA), *rec*G (que codifica la helicasa RecG dependiente de ATP, EC:3.6.4.12), *ruv*A (que codifica la helicasa RuvA, EC:3.6.4.12), *ruvB* (que codifica la helicasa RuvB, EC:3.6.4.12), *ruvC* (que codifica la endodeoxiribonucleasa RuvC, EC:3.1.22.4) y *priA* (que codifica la proteina primosomal N', EC:3.6.4.-). En cuanto a los genes implicados en rutas de biosíntesis de aminoácidos esenciales, los *loci* estudiados son *arg*H (que codifica la argininosuccinato liasa, EC:4.3.2.1), *ilv*D (que codifica la dihydroxiácido dehidratasa, EC:4.2.1.9), *leu*B (que codifica la 3-isopropilmalato deshidrogenasa, EC:1.1.1.85), *met*E (que codifica la metionina sintasa independiente de cobalamina, EC:2.1.1.14), *thr*C (que codifica la treonina sintasa, EC:4.2.3.1) y *trpB* (que codifica la subunidad beta de la triptófano sintasa, EC:4.2.1.20). Para todos los casos, el alineamiento múltiple de genes homólogos procedentes de beta- y gamma-proteobacterias (Tabla S1 en Capítulo 4, página 116) se llevó a cabo con ClustalW (Larkin et al., 2007), permitiendo la identificación de motivos altamente conservados sobre los que fueron diseñados los cebadores degenerados correspondientes (Tabla 1 en Capítulo 4, página 98; Tabla S1 en Capítulo 5, página 144). Las amplificaciones (utilizando el KAPATaq DNA Polymerase Kit, Kapa Biosystems) se realizaron a partir de extracciones de ᴛDNA de los insectos hospedadores (JETFLEX Genomic DNA Purification Kit, GENOMED), utilizando 50–60 μmol de cada cebador/50 μl de reacción y

una temperatura de hibridación de 52ºC. En el caso de los genes de biosíntesis de aminoácidos esenciales, se usaron también temperaturas de hibridación de 56ºC y los amplicones obtenidos fueron clonados (pGEM-T Easy Vector System I Kit, Promega) antes de ser secuenciados. La filiación taxonómica de los haplotipos muestrados fue establecida mediante búsquedas de BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi/; Altschul et al., 1997).

Localización *in situ* de macromoléculas

Llevamos a cabo la detección *in situ* de 16S rRNA de "*Ca*. Tremblaya phenacola" de *P. peruvianus* y de los miembros del consorcio endosimbionte en *P. citri* mediante FISH, así como la inmunotinción de la chaperonina GroEL y el canal de membrana MscL del gamma-endosimbionte de *P. citri.* En todos los casos, hembras adultas del insecto hospedador fueron decapitadas y fijadas en paraformaldehido al 4% antes de ser incluidas en bloques de parafina.

Las secciones de *P. peruvianus* fueron hibridadas con la sonda universal Cy5-EUB338 (Amann et al., 1990) o la sonda específica Cy3-TphPPER1290 (5'-CCGCAATTCGTACTGAGGTTAGG-3'), diseñada sobre el gen 16S rRNA de "*Ca*. Tremblaya phenacola". Las secciones de *P. citri* fueron hibridadas con las sondas TAMRA-b91 y 6FAM-g630 (von Dohlen et al., 2001). Para la inmunodetección de la chaperonina GroEL del gamma-endosimbionte de *P. citri*, se usaron anticuerpos policlonales de conejo contra la proteína homóloga en *Buchnera aphidicola* APS, gamma-endosimbionte del áfido *Acyrthosiphon pisum* (Harris) (Hara et al., 1990). Por otra parte, la inmunodetección del canal de membrana MscL se llevó a cabo utilizando anticuerpos policlonales de conejo generados por Covalab (Villeurbanne, France) sobre epítopos periplásmicos (C-KQFSWVLKPAQGNR, residuos 55–68) y citoplásmicos (C-HNKEEEETPNELSKQS, residuos 103–118) de dicha proteína en el gamma-endosimbionte de *P. citri* (Moe et al., 1998). El marcaje primario fue revelado mediante anticuerpos secundarios de asno anti-conejo marcados con Alexa fluor 488. La tinción con DAPI (4 ',6-diamino-2-fenilindol) permitió visualizar el genoma nuclear del insecto hospedador en todos los casos.

Las muestras fueron observadas en microscopios de epifluorescencia (Olympus IX81; Nikon Eclipse 80i). Los programas Cell F (AnalySIS) y NIS-Elements BR se utilizaron para la captura y el procesado de las imágenes.

Caracterización molecular de bacterias endosimbiontes

Los genes 16S rRNA de "*Ca*. Tremblaya phenacola" de *P. peruvianus* y *P. madeirensis*, del beta- y el gamma-endosimbionte de *D. boninsis* y del gamma-endosimbionte de *P. longispinus* fueron amplificados (KAPATaq DNA Polymerase Kit, Kapa Biosystems) con los cebadores universales 16S-up y 16S-down (van Ham et al., 1997). Los amplicones obtenidos fueron clonados con pGEM-T Easy Vector System I Kit (Promega) y al menos 25 clones/insecto hospedador fueron secuenciados (BigDye Terminator v3.1 Cycle Sequencing Kit, Applied

Byosistems) con los cebadores T7 y SP6. Por otra parte, los cebadores leuA, prs5/6 y U16S (Baumann et al., 2002) se usaron para amplificar los fragmentos parálogos *leuA-rrs*1 y *prs-rrs*2 de los genomas de "*Ca*. Tremblaya princeps" de *D. boninsis*, *P. longispinus*, *P. viburni* y *P. ficus*. Dichos cebadores, así como OR-leuAR2 (5´-TCAGTMATTAHGGCWACCTGCAC-3´), OR-prsR2 (5´-AATAGCYAAGCGGGTCAAGGC-3´) y OR-UF2 (5´-TGGCGCATGCTGTATGAGTTC-3´) sirvieron para secuenciar tales fragmentos génomicos.

La caracterización filogenética de las cepas de "*Ca*. Tremblaya phenacola" de *P. peruvianus* y *P. madeirensis*, el gamma-endosimbionte de *P. longispinus* y los miembros del consorcio endosimbionte de *D. boninsis* se llevó a cabo mediante la reconstrucción de árboles basados en el gen 16S rRNA o secuencias parciales de genes codificantes (*trp*B, *arg*H; ver apartado **Cribado genético**). Los alineamientos múltiples se realizaron con ClustalW (Larkin et al., 2007). El modelo evolutivo GTR+I+G fue inferido mediante JModelTest (Guindon and Gascuel, 2003; Darriba et al., 2012). Todas las recontrucciones filogenéticas fueron realizadas por los métodos de Máxima Verosimilitud (ML), Máxima Parsimonia (MP) e Inferencia Bayesiana (BI), usando los programas RAxML (Stamatakis, 2006), DNAPARS del paquete PHYLIP (Felsenstein, 2005) y MrBayes (Ronquist et al., 2012), respectivamente. Las figuras resultantes de los análisis filogenéticos se prepararon con los programas FigTree (http://tree.bio.ed.ac.uk/software/figtree/) e Inkscape (https://inkscape.org/es/).

1.4. Conclusiones

1. La secuenciación y caracterización del genoma completo de "*Ca*. Tremblaya princeps" PCVAL confirmó la presencia de características atípicas para un P-endosimbionte (alto contenido en GC, baja densidad génica, evolución concertada de fragmentos genómicos parálogos). Por otro lado, el genoma de "*Ca*. Tremblaya princeps" es uno de los genomas procariotas más reducidos conocidos y el genoma bacteriano con menor número de genes descrito hasta el momento.
2. Ni la abundancia de pseudogenes ni la expansión de repeticiones en tandem explican la baja densidad génica en el genoma de "*Ca*. Tremblaya princeps".
3. La comparación entre "*Ca*. Tremblaya princeps" PCVAL y "*Ca*. Tremblaya phenacola" PAVE, P-endosimbionte de la cochinilla de la avena *Phenacoccus avenae* (Borchsenius), sugiere que la duplicación del operon ribosomal y su contexto genómico inmediato tuvo lugar en el ancestro común de ambas especies de "*Ca*. Tremblaya", antes de la divergencia de las subfamilias Pseudococcinae y Phenacoccinae.
4. La secuenciación y caracterización del genoma completo del gamma-endosimbionte de *P. citri*, "*Candidatus* Moranella endobia", reveló otro evento de duplicación genómica parcial sometida a evolución concertada.

5. La secuenciación y caracterización de las regions *leu*A-*rrs*1 y *prs*-*rrs*2 del genoma de "*Ca*. Tremblaya princeps" de *P. ficus* confirmó la incidencia reciente de eventos de HR en "*Ca*. Tremblaya princeps" del clado E.
6. La recuperación del potencial de HR mediante la internalización recurrente de gamma-endosimbiontes podría explicar, al menos en parte, la asociación entre la endosimbiosis anidada y la evolución reductiva atípica de "*Ca*. Tremblaya princeps". La detección de genes implicados en rutas de HR en el gamma-endosimbionte de *P. citri* y *P. ficus*, pero también en los gamma-endosimbiontes de *P. longispinus* y *P. viburni*, es coherente con esta hipótesis.
7. La caracterización de secuencias repetitivas en los genomas de "*Ca*. Tremblaya princeps" y "*Ca*. Moranella endobia" sugiere mayor susceptibilidad relativa del primero a la maquinaria de HR codificada por el último.
8. El análisis funcional de ambos genomas sugiere que (1) "*Ca*. Moranella endobia" es el único miembro del consorcio capaz de generar energía y que (2) "*Ca*. Tremblaya princeps" necesita importar un amplio número de productos génicos codificados exclusivamente en el genoma de "*Ca*. Moranella endobia". Así, la complementación entre ambas bacterias no afectaría solo al metabolismo intermediario, sino que sería esencial a nivel informacional. Ambos endosimbiontes formarían parte de una entidad biológica sin precedentes.
9. Queda por aclarar la forma en la que se produce la exportación de proteínas esenciales hacia "*Ca*. Tremblaya princeps", a pesar de que "*Ca*. Moranella endobia" codifica un sistema de translocación Sec putativamente functional. La conservación del canal mecanosensible MscL y la inmunotinción de proteínas sugiere que la comunicación molecular estaría mediada, al menos en parte, por estímulos osmóticos y apunta a un patrón esporádico de exportación.
10. Los miembros del consorcio microbiano colaboran estrechamente en la síntesis de aminoácidos esenciales, como es habitual en endosimbiontes de insectos chupadores de savia. El suplementado de la dieta del hospedador con aminoácidos esenciales y la especialización al respecto de los miembros del consorcio endosimbionte es un fenómeno general en cochinillas de la subfamilia Pseudococcinae.
11. Al menos un evento de transferencia génica horizontal (HGT) habría influido en la evolución de la biosíntesis de aminoácidos esenciales en "*Ca*. Tremblaya phenacola" de *P. peruvianus* y *P. madeirensis*.

2. INTRODUCTION

2.1. Symbiosis

2.1.1. Definition and classification

Symbiosis (which etymologically derives from Greek sym "together" and biosis "living") is defined as a long-term association involving individuals from two or more species who depend on each other at any biological level (i.e., genetic, metabolic or behavioural).

The term was first proposed by the botanist Anton de Bary in his work "*Die Erscheinungen des symbiose*" (1879), based on his observations on the lichen biology. In the de Bary's definition, short-term associations are not considered symbiotic while no judgement about the effects of the biological interaction on the fitness of the participants is included. In the most extreme cases, both members can constitute a new living being displaying its own physiological properties. Although pioneer, such definition of symbiosis is not universally accepted. Since the birth of the term, several authors have restricted the use of "symbiosis" for those cases where interacting organisms take profit from their interaction (Saffo, 1992), or in the most extreme sense, only when new metabolic capabilities arise in a certain organism through the association with another one belonging to a different species (Douglas, 1994). The debate between those supporting the more restricted definition of symbiosis and those defending the broad sense concept, closely related to the original definition by de Bary, is still open (Paracer and Vernon, 2000).

Under the broad sense conception of symbiosis, three main subtypes of association can be distinguished depending on its effect on the fitness of interacting organisms (symbionts). Thereby, if one of the members benefits from the other without causing neither harm nor benefit to the partner, we talk about **commensalism**. In contrast, we talk about **parasitism** when one of the members is beneficed by causing a decrease in the fitness of the partner. Finally, we speak about **mutualism** when both interacting organisms get benefited from the relationship. Strikingly, such three terms were described by Pierre-Joseph van Beneden in the book "*Les commensaux et les parasites*" (1876) even before the introduction of the term "symbiosis". In any case, accurately discriminating among these ecological scenarios is often difficult (Yule et al., 2013; Garcia and Gerardo, 2014).

On the other hand, symbionts can be differentially classified according to the relative localization of the partners and the degree of the symbiotic association. Thus, they are known as **ectosymbionts** when living on the surface of the host's body. This includes internal surfaces such as the digestive tube lining and the ducts of glands. In contrast, they are kown as **endosymbionts** when living within the host's cells. Additionally, they can be considered **facultative** or **obligate** depending on its ability to survive, or not, outside of the symbiotic consortium.

2.1.2. Symbiosis as a major evolutionary force

Historically, the role of the symbiosis in evolution has been controversial. However, there is a common agreement in this context: symbiosis was essential in the origin and early evolution of the eukaryotic cell. Thus, the first eukaryote may have appeared through a symbiotic association involving a bacterium and an archea (de Duve, 2007). Furthermore, some of the organelles of the eukaryotic cell were the result of the subsequent invasion of the primitive eukaryotic cell by several prokaryotes, according to the Serial Endosymbiotic Theory (SET) proposed by Lynn Margulis (Margulis, 1993 and references therein) (Figure 1). Thus, the mitochondrion may be originated from an ancestral alphaproteobacterium in one unique event of symbiosis, while the chloroplast evolved from an ancestral cyanobacterium that was engulfed and retained by the host cell. Despite the initial controversy, the evolutionary origin of mitochondria and chloroplasts is nowadays well supported by physiological, morphological and molecular data (Gray and Spencer, 1996; Martin et al., 1998). Symbiosis has shaped the evolution of life in many other ways. Thereby, the evolution of the immune system of complex eukaryotes has been deeply influenced by pathogenic interactions with other organisms while the mutualistic relationships that both plants and animals have established with prokaryotes and fungi would had furnished them with new metabolic capabilities, allowing the colonization of other niches that otherwise would be inaccessible.

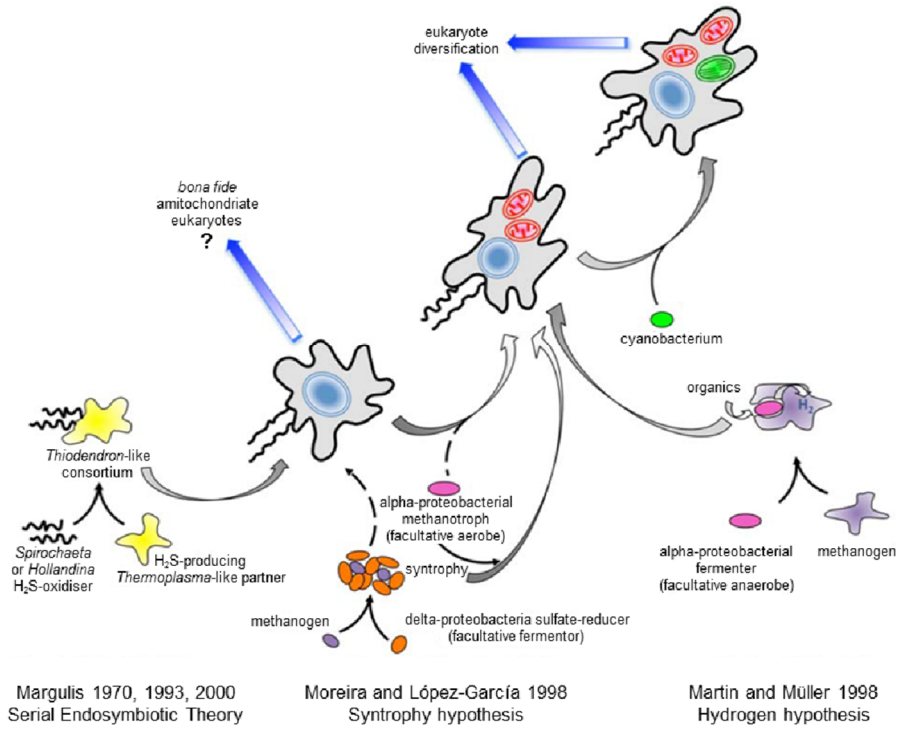


Figure 1. Symbiogenesis as the origin of the eukaryotic cell and its organelles (Moya and Peretó, 2011; adapted from Latorre et al., 2011).

The huge variety of symbiotic associations described along the three domains of life (Figure 2) is undoubtedly an indicator of the evolutionary relevance of symbiosis (Moya et al., 2008; McFall-Ngai, 2008; Moya and Peretó, 2011). Many examples of symbiotic associations between eukaryotes have been already described. The most renowned case is that of lichens, which are associations between a fungus and an algae, but there are also associations of fungi with protists, animals, plants or other fungi (Paracer and Ahmadjian 2000; Moya et al. 2008). Nevertheless, symbiotic associations involving (at least) one prokaryotic microorganism are especially abundant.

Prokaryotes possess an impressive set of metabolic capabilities that allowed them to colonize a broad range of ecosystems and environmental conditions. Thus, they are good candidates to colonize the surfaces and inner spaces of other organisms. In coherence, many prokaryotes live physically attached to other living beings. Biofilm formation is a clear example of an association between different species of prokaryotes (Hall-Stoodley et al., 2004). However, the symbiotic associations between eukaryotes and prokaryotes are the most studied. There have been described symbiotic associations with prokaryotes in practically all branches of the eukaryotic tree of life (Figure 2). Eukaryotes, and especially animals, have limited metabolic skills, and they have acquired several metabolic capabilities like nitrogen fixation (Kneip et al., 2007), methanogenesis (Schink, 1997), chemolithoautotrophy (Stewart et al., 2005), nitrogen assimilation (Minic and Hervé, 2004), and biosynthesis of several nutrients lacking in the diet (Zientz et al., 2004) through the association with bacterial symbionts.

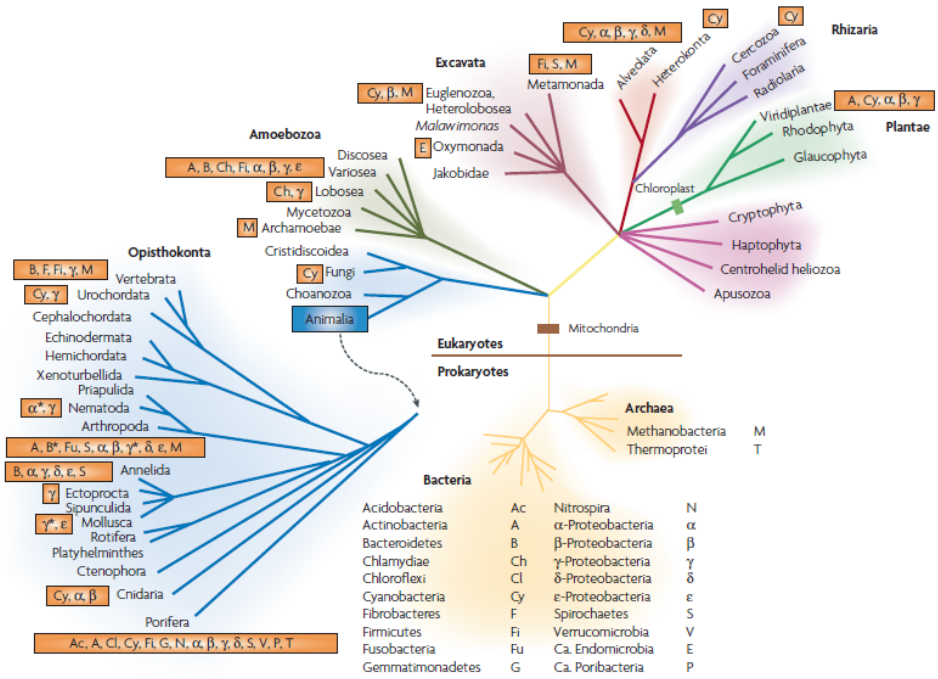


Figure 2. Phylogenetic distribution of eukaryote-prokaryote symbioses. Orange boxes represent the phyla of the symbionts. Asterisks indicate availability of symbionts genomes by 2008 (Moya et al., 2008).

2.1.3. Bacterial symbionts of insects

Symbiotic association with microorganisms are ubiquitous among many groups of insects. Such microorganisms (i.e., bacteria, fungi and protozoans) can display very diverse location, such as the gut lumen, ceca connected to the gut, specialized gut epithelial cells, the hemocoel or highly developed symbiotic organs called mycetomes/bacteriomes, typically associated with the midgut in the insect body cavity (Buchner, 1965; Tanada and Kaya, 1993). Bacterial symbiosis with insects has been particularly well studied. Paul Buchner first catalogued a large amount of these associations in his book "*Endosymbiosis of animals with plant microorganisms*" (1965). After their appeareance in the Devonian (350-400 million years ago), insects become the most diverse animal lineage, with 1 million described species wordwide (Mora et al., 2011). Association with endosymbiotic bacteria, which particularly affect insects from orders Blattaria, Curculionidae and Hemiptera (Dasch et al., 1984), has been considered as a key factor for the evolutionary success of insects. Many mutualistic bacteria complement their host's unbalanced diets (e.g., plant sap, cereals or blood), allowing them to colonize nutritionally poor niches. Symbionts are classified into two major categories, attending to their dispensability for their host survival: primary (P-) or obligate endosymbionts and secondary (S-) or facultative symbionts (Baumann, 2005).

P-endosymbionts are essential for the survival and reproduction of the host, being therefore fixed in the host population (Houk and Griffiths, 1980). They are typically endocellular bacteria living inside specialized polyploid cells (bacteriocytes) that use to aggregate forming a bacteriome. Furthermore, they are vertically transmitted from mothers to offspring (Buchner, 1965; Baumann, 2005), so that both insect hosts and their obligate endosymbionts traditionally show congruent phylogenies due to coevolution. Such a complex evolutionary pattern, typically found when unrelated organisms maintain tight ecological interactions, has already been revealed for many insect lineages such as aphids (Munson et al., 1991), psyllids (Thao et al., 2000), whiteflies (Thao and Baumann, 2004), leafhoppers (Moran et al., 2003), cockroaches (Lo et al., 2003), tsetse flies (Chen et al., 1999), carpenter ants (Sauer et al., 2000), weevils (Conord et al., 2008), lice (Allen et al., 2007) and scale insects (Rosenblueth et al., 2012), among others. Since P-endosymbionts are highly adapted to the intracellular environment, these microorganisms have remained generally unculturable. Thus, PCR, DNA sequencing and molecular phylogenetic analyses are frequent approaches for their characterization (Murray and Schleifer, 1994).

On the other hand, facultative S-symbionts are nonessential for host survival and reproduction, so they usually show partial infection in host populations. They are not necessarily endocellular bacteria. In contrast, they can be found both intra- or extracellularly, in the bacteriocytes or close to the bacteriome, in other cell types or free in the host haemolymph (Dale and Welburn, 2001; Moran et al., 2005; Brumin et al., 2011). Moreover, they usually do not show congruent phylogenies with their hosts, since they can be laterally transferred between host species (Russell et al., 2003; Nováková et al., 2009).

2.1.3.1. Evolutionary dynamics of bacterial symbiosis in insects

The infection of the eukaryotic host by a free-living bacterium is the first step towards the establishment of an obligate endosymbiosis. In such case, the host develops specialized cells to harbor the bacterium in a stable and chemically enriched niche, while, in turn, the bacterium provides essential nutrients to the host (Moya et al., 2009). Eventual presence of facultative S-symbionts is not infrequent, so that a complex biological interplay can be established among the P-endosymbiont, S-symbiont(s) and insect hosts. A new stable association occurs if the newly established symbiont supplies an additional benefit to the consortium. In this context, two evolutionary outcomes are possible: complementation and replacement.

Complementation occurs when two or more bacteria, known as co-primary endosymbionts, are required to fulfill the consortium metabolic needs. Such a concept perfectly fits into the Black Queen Hypothesis (Morris et al., 2012), a novel theory on reductive evolution that includes two main assumptions. (1) Due to the nature of the diffusion gradients and the bacterial needs for permeable membranes, bacterial functions are often leaky, so that cells produce resources that unavoidably benefit other cells from the local bacterial community. (2) Since the maintenance of pathways for the production of available products has an associated detrimental cost, the receiver cells will tend to delete redundant pathways, thus developing dependency upon such cellular interaction. In this scenario, the authors suggested that initial dependencies might favor the arising of additional ones, so that simpler initial interactions would evolve into more complex associations based upon costly resources exchange. Such a new concept suggest that many cooperative interactions would frequently be formed within bacterial communities while leaving evident genomic signature on interacting bacteria throughout the complementary loss of shared diffusible functions and/or metabolites (Sachs and Hollowell, 2012). This testable prediction has been vastly confirmed, since genome streamlining (deleting costly genetic pathways producing transferable biological products) is common among parasites and symbionts (e.g., dual symbiosis found in many insects) (Wu et al., 2006; McCutcheon and Moran, 2007; Gosalbes et al., 2008; McCutcheon et al., 2009a; McCutcheon et al., 2009b; McCutcheon and Moran, 2010; Lamelas et al., 2011). In contrast, replacement occurs when reductive evolution affects mostly one of the co-existing bacterial symbionts, thus leading to its extinction (Moya et al., 2009). Replacements appear to allow the host to avoid dependence on an ancient, degraded and potentially inefficient symbiont after long coevolutionary periods (Koga et al., 2013).

Several examples of replacement of ancestral endosymbionts are known. This is the case of the replacement of "*Candidatus* Nardonella" by "*Candidatus* Sodalis pierantonius" in grain weevils (Coleoptera: Curculionoidea) (Lefevre et al., 2004; Oakeson et al., 2014). Both complementation and replacements are common among sap-feeding insects from suborder Auchenorrhyncha. The consortium involving "*Candidatus* Sulcia muelleri" and "*Candidatus* Zinderia insecticola" was apparently present in the common ancestor of Auchenorrhyncha (Koga et al., 2013), and can still be found in some of its extant descendants. In other members of the suborder, new consortia have been reported. Most of them involve "*Ca.* Sulcia muelleri", but "*Ca.* Zinderia insecticola" has

been replaced by a variety of newly incorporated endosymbionts, such as "*Candidatus* Baumannia cicadellinicola" in most sharpshooters (Takiya et al., 2006), "*Candidatus* Hodgkinia cicadicola" in cicadas (McCutcheon et al., 2009a), and "*Candidatus* Vidania fulgoroideae" in planthoppers (Urban and Cryan, 2012). Transitional stages in the evolutionary succession of symbioses, have also been described. Thus, *Aphrophora quadrinotata* (Say) retains "*Ca.* Sulcia muelleri" and "*Ca*. Zinderia insecticola", while it has already acquired a *Sodalis*-like symbiont (Koga et al., 2013). Such intermediate stage of coexistence might help to overcome barriers to the replacement of ancient P-endosymbionts by novel S-symbionts. Furthermore, the potential for future bacterial replacement has been empirically tested in some systems. Thus, a novel symbiotic system in which the facultative "*Candidatus* Serratia symbiotica" both physiologically and cytologically replaces *Buchnera aphidicola*, the P-endosymbiont of aphids (Hemiptera: Aphididae), has been empirically stablished under laboratory conditions (Koga et al., 2003, Moran et al., 2005). Replacements do not occur only between endocellular bacteria, but also between gut microbiota and endocellular bacteria, or even between more distantly related microorganisms. Examples of each type include the replacement of *Blattabacterium* in most termite species (Blattodea: Isoptera) except in the basal lineage *Mastotermes darwiniensis* Froggatt (Sabree et al., 2012), and the replacement of *B. aphidicola* by a fungal species in some aphids from the tribe Cerataphidini (Fukatsu and Ishikawa, 1996).

2.1.3.2. Functional relevance of primary endosymbionts

Even though most of them display very uniform nutritional requirements (Dadd, 1985), insects have colonized a broad range of ecological niches, thus feeding on a number of different diets. Examples range from generalists (i.e., omnivores or scavengers) to very specialized groups (Slansky and Rodriguez, 1987). In many cases, their diet cannot satisfy their nutritional requirements, so that mutualistic associations with endosymbiotic bacteria become essential, and a clear link between the diet composition and the role of the bacterial endosymbionts is usually observed (Table 1, page 21). For example, the hemipteran insects have needle-like mouthparts and feed exclusively on plant sap. Although enriched in carbohydrates (mainly sucrose), phloem sap contains very small amounts of lipids, proteins or vitamins, while even carbohydrates are absent from xylem sap (Douglas, 1993; Sandström and Pettersson, 1994; Dinant et al., 2010). In phloem-feeding insects, most lipids can be synthesized from sap carbohydrates. This is not the case of proteins and vitamins, due to great deficiencies in nitrogenous precursors. Thus, supplementation of deficient diets through the provision of essential amino acids and vitamins is frequently the role of the bacterial P-endosymbionts, as early suggested by empirical studies on aposymbiotic insects (Nogge, 1981; Wicker and Nardon, 1982; Prosser and Douglas, 1992). In coherence, the genome sequencing of several *B. aphidicola* strains revealed that most of them present complete biosynthetic pathways for essential amino acids (Shigenobu et al., 2000; Tamas et al., 2002; van Ham et al., 2003; Pérez-Brocal et al., 2006; Degnan et al., 2011). Nutritional supplementation can be performed by a single P-endosymbiont or by a bacterial consortium. This is the case of the cedar aphid *Cinara cedri* (Mimeur), where *B. aphidicola* and "*Ca*. Serratia symbiotica" coexist in the bacteriome and collaborate in the biosynthesis of tryptophan and lysine

(Lamelas et al., 2011). Collaboration in essential amino acids supply is widely extended in sap-feeding insects from Auchenorrhyncha in which bacterial consortia involving "*Ca.* Sulcia muelleri" and several anternative co-primary endosymbionts (e.g., "*Candidatus* Baumannia cicadellinicola", "*Candidatus* Hodgkinia cicadicola", "*Candidatus* Vidania fulgoroideae"; Takiya et al., 2006; McCutcheon et al., 2009a; Urban and Cryan, 2012) have been reported. In the cases of insects that feed on blood of vertebrates, endosymbionts are mainly devoted to the production of cofactors and vitamins. This is the main role of *Wigglesworthia glossinidia* in tsetse flies (Diptera: Glossinidae) (Nogge, 1981; Akman et al., 2002), or "*Candidatus* Riesia pediculicola" in the human body louse *Pediculus humanus* (Haeckel) (Phthiraptera: Pediculidae) (Kirkness et al. 2010).

Although usually linked to restricted diets, mutualistic intracellular bacteria are also found in omnivorous insects (Table 1, page 21). This is the case of "*Candidatus* Blochmannia", the P-endosymbiont of carpenter ants from the genus *Camponotus* (Mayr) (Hymenoptera: Formicidae), which supply its host with an additional source of essential amino acids (Gil et al., 2003; Feldhaar et al., 2007) that might allow it to exploit alternative niches (Davidson et al., 2003), and help it to overcome certain developmental stages of its biological cycle (Wolschin et al., 2004). "*Ca*. Blochmannia" is involved in nitrogen recycling (Fedhaar H et al., 2007). Strikingly, *Blattabacterium cuenotti* displays a similar role in cockroaches (Blattodea), including nutrient supply and nitrogen recycling, which explains the extraordinary shift from uricotely (typically found among terrestrial arthropods) to ammoniotely of their host (López-Sánchez et al., 2009; Patiño-Navarrete et al., 2014). Both "*Ca*. Blochmannia" and *Blattabacterium*, represent a remarkable example of evolutionary convergence during the symbiotic process (López-Sánchez et al., 2009).

2.1.3.3. Functional relevance of secondary symbionts

Although their transfer mechanisms in nature are not well understood, S-symbionts can be experimentally cured or transferred among insects, allowing empirical approaches to the study of their effects on the eukaryotic host (Dale and Welburn, 2001; Montllor et al., 2002; Oliver et al., 2003; Tsuchida et al., 2004). Their presence influences host biology, both directly or through interactions with coexisting obligate endosymbionts. Thus, many S-symbionts are known to play an important role in determining ecological traits of their host (Table 1, page 21), which explains their persistence and spread in host populations. Therefore, symbiont dynamics strongly depends on the ecological conditions affecting a certain host population (Himler et al., 2011; Pan et al., 2012).

Reproductive alterations, such as cytoplasmic incompatibility (CI), parthenogenesis induction in haplodiploid species (PI), feminization of genetic males or male-killing, are very common among microbe-infected insects (Hunter et al., 2003) and cause highly skewed sex ratios favoring female production. Such sex ratio distortion (SRD) is generally very adaptive, since S-symbionts are typically inherited in a maternal manner. Collectively known as reproductive parasites, these bacteria are widely distributed among arthropods (Duron et al., 2008). The alphaproteobacterium *Wolbachia pipientis* is considered the master reproduction manipulator microbe, but it is not alone (Duron et al., 2008; Serbus et al., 2008; Zug and Hammerstein, 2012). A wide variety of bacteria

have demonstrated to effectively cause one or more of the previously mentioned reproductive alterations. For instance, Bacteroidetes and *Cardinium* symbionts are known to produce CI in the parasitoid wasp *Encarsia pergandiella* (Howard) (Hunter et al., 2003) and the spider mite *Eotetranychus suginamensis* (Yokoyama) (Gotoh et al., 2007), respectively. On the other hand, selective killing of male embryos by *Rickettsia* (Lawson et al., 2001; Majerus and Majerus, 2010), Flavobacteria (Hurst et al., 1997; Hurst et al., 1999), and the gammaproteobacteria *Arsenophonus nasoniae* (Gherna et al., 1991) has been observed in the buprestid leaf-mining beetle *Brachys tessellatus* (Fabricius), the Japanese ladybird *Propylea japonica* (Thunberg), the ladybirds beetles *Adonia variegata* (Goeze) and *Coleomegilla maculata* (De Geer), or the parasitoid wasp *Nasonia vitripennis* (Walker).

Moreover, infection by some S-symbionts appears to enrich insect genotype, providing them with new adaptive traits. Thus, they are known to broad the food plant range of the host, as well as to influence the host resistance to both biological and physical environmental factors. This is the case of the gammaproteobacteria "*Candidatus* Regiella insecticola", which facilitates feeding of pea aphid *Acyrthosiphon pisum* Harris on *Trifolium repens* (Fabaceae) (Tsuchida et al., 2004) and confers protection against the infection by the fungi *Pandora neoaphidis* (Ferrari et al., 2004). On the other hand, the gammaproteobacteria "*Ca*. Serratia symbiotica" and "*Candidatus* Hamiltonella defensa" reduce pea aphid vulnerability to the parasitoid wasps *Aphidius ervi* (Haliday) and *Aphidius eadyi* (Stary, Gonzalez & Hall) (Oliver et al., 2003; Ferrari et al., 2004; Moran et al., 2005; Degnan et al., 2009). Furthermore, "*Ca*. Serratia symbiotica" is known to enhance *A. pisum* thermotolerance at early developmental stages, apparently in a specific manner (Montllor et al., 2002; Moran et al., 2005). In contrast, the thermotolerance induced in the whitefly *Bemisia tabaci* (Gennadius) by a *Rickettsia* S-symbiont is considered as a side effect, since the presence of this microbe induces the expression of host thermotolerance-related genes under normal temperature conditions (Brumin et al., 2011).

Table 1. Endosymbiotic bacteria of insects. Bacterial roles as primary (P), secondary (S) or co-primary (CP) endosymbionts are indicated.

Host	Diet	Symbionts (Type)	Phyla	Metabolic capability
Pea aphid (*Acyrthociphon pisum*)	Phloem	*Buchnera aphidicola* (P)	Gammaproteobacteria	Essential amino acids & vitamins
		"*Ca.* Hamiltonella defensa" (S)	Gammaproteobacteria	Protection against parasitoids
		"*Ca.* Regiella insecticola" (S)	Gammaproteobacteria	Protection against pathogenic fungi
		"*Ca.* Serratia symbiotica" (S)	Gammaproteobacteria	Protection against heat stress
Cedar aphid (*Cinara cedri*)	Phloem	*Buchnera aphidicola* (CP)	Gammaproteobacteria	Essential amino acids & vitamins
		"*Ca.* Serratia symbiotica" (CP)	Gammaproteobacteria	Tryptophan & vitamins
Sharpshooters (*Homalodisca vitripennis*)	Xylem	"*Ca.* Sulcia muelleri" (CP)	Bacteroidete	Essential amino acids
		"*Ca.* Baumannia cicadellinicola" (CP)	Gammaproteobacteria	Histidine, methionine, cofactors & vitamins
Cicadas (*Diceroprocta semicinta*)	Xylem	"*Ca.* Sulcia muelleri" (CP)	Bacteroidete	Essential amino acids
		"*Ca.* Hodgkinia cicadicola" (CP)	Alphaproteobacteria	Histidine & methionine
Spittlebugs (*Clastoptera arizonana*)	Xylem	"*Ca.* Sulcia muelleri" (CP)	Bacteroidete	Essential amino acids
		"*Ca.* Zinderia insecticola" (CP)	Betaproteobacteria	Histidine, methionine & tryptophan
Tsetse fly (*Glossina brevipalpis*)	Blood	*Wigglesworthia glossinidia* (P)	Gammaproteobacteria	B-family vitamins
		Sodalis glossinidius (S)	Gammaproteobacteria	Immunity
Louse (*Pediculus humanus*)	Blood	"*Ca.* Riesia pediculicola"	Gammaproteobacteria	B-family vitamins
Rice weevil (*Sytophilus oryzae*)	Grain	"*Ca.* Sodalis pierantonius"	Gammaproteobacteria	Essential amino acids & vitamins
Ants (*Camponotus* spp.)	Omnivorous	*Blochmannia* spp. (P)	Gammaproteobacteria	Nitrogen metabolism
Cockroaches and termites (Order Blattodea)	Omnivorous	*Blattabacterium* spp. (P)	Bacteroidete	Nitrogen metabolism
Arthropods	-	*Wolbachia* spp. (S)	Alphaproteobacteria	Host reproductive parasite

2.1.4. Genomic changes in endosymbiotic bacteria

Free-living bacteria undergo dramatic genomic changes during the transition to an endocellular lifestyle. The most prominent ones are a drastic reduction of their genomes, which typically include small sets of genes, and a significant increase in their AT-content (Baumann, 2005; McCutcheon and Moran, 2012).

Two main factors are thought to drive genome shrinkage in these bacteria: first, the intracellular environment where such bacteria reside, and second, the particular population dynamics of endosymbionts (Moya et al., 2008). Living inside a eukaryotic cell renders several genes unnecessary or redundant, because they are involved in functions that can be performed by the host. Thus, the eventual loss of these genes has no effect on bacterial fitness and, therefore, the pressure of purifying selection over them is relaxed. In consequence, values of genetic essentiality in endosymbiotic networks are typically higher than those observed in free-living bacteria, whose metabolic networks are much more redundant (Feist et al., 2007; Thomas et al., 2009; González-Domenech et al., 2012). On the other hand, the strict vertical transmission of few endosymbionts from mother to offspring reduces the effective population size, increasing the effects of random genetic drift on endosymbiont evolutionary changes (Moran, 1996). The impact of vertical transmission in shaping the genomes of obligate symbionts is supported by the fact that similar genomic features are found in "*Candidatus* Ishikawaella capsulata", extracellular symbiont of stinkbugs from the family Plataspidae (Hosokawa et al., 2006). Thus, these two major factors facilitate the fixation of slightly deleterious mutations in non-essential genes, causing their inactivation and subsequent loss. Genes involved in DNA repair, recombination and uptake are typically affected by this process. Moreover, the early loss of these genes further increases the mutation rate, and prevents the genetic exchange by means of homologous recombination (HR). Additionally, the relative isolation of these bacterial populations usually hinder the possibility of gaining new genetic material through horizontal gene transfer (HGT) events, ruling out the potential compensation of the genetic losses. Finally, eventual acquisition of functional bacterial genes by the genome of the insect host is thought to facilitate the extreme genome reduction of long-term intracellular endosymbionts (Sloan et al., 2014).

The genome reduction process is apparently divided in two main phases. First, soon after the establishment of the symbiosis and favoured by the relaxation of purifying selection, there is a huge proliferation of mobile elements, such as insertion sequences (IS) (Moran and Plague, 2004; Gil et al., 2008). The accumulation of IS, in turn, enhances the reductive process by increasing the intrachromosomal HR events, which induces genome rearrangements and the loss of large genomic fragments (Parkhill et al., 2003). Moreover, transposition itself can inactivate individual genes that would be degraded afterwards (Gil et al., 2008). During the second phase, genome reduction continues through the pseudogenization and loss of individual genes scattered throughout the genome (Silva et al., 2003). At this point, genomes are completely stabilized and there are not more rearrangements (Tamas et al., 2002), mostly due to the elimination of both

mobile elements and genes involved in DNA repair and recombination (Moya et al., 2008) (Figure 3).

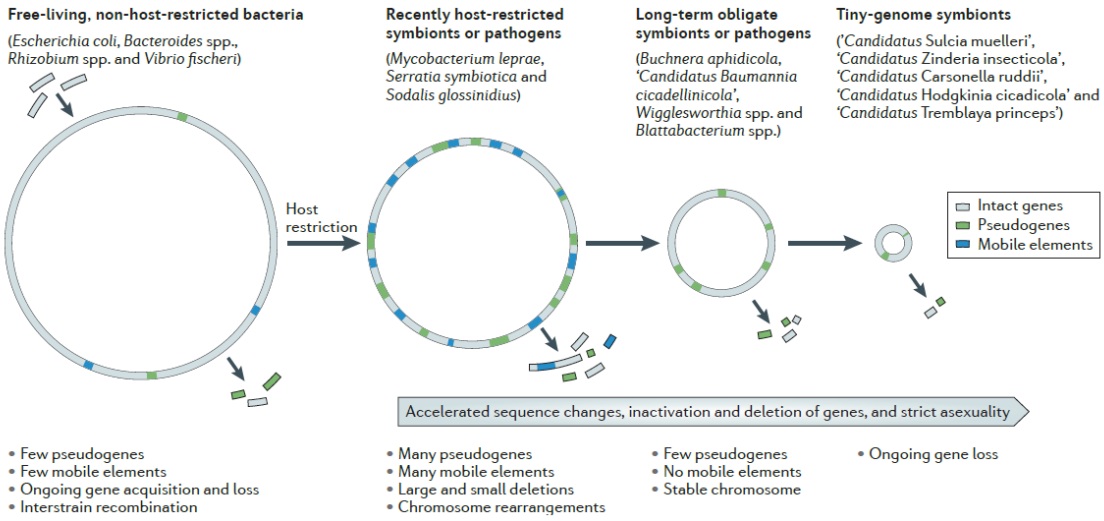


Figure 3. Genome reduction process in obligate endosymbiotic bacteria (McCutcheon and Moran, 2012).

Two hypothesis have been proposed in orden to explain the high AT-content typically shown by the genome of long-term endosymbionts. One of them takes into account the higher energetic cost of the biosynthesis of GTP and CTP nucleotides compared to the cost of producing ATP or TTP (Rocha and Danchin, 2002). The other one suggests this bias to result from the combination of mutational pressure from G/C to A/T and the loss of DNA repair systems (Lind and Andersson, 2008; Hershberg and Petrov, 2010). Whatever the case, this bias usually has a strong impact on both codon-usage and the amino acid composition of the proteome, which in turn decreases the thermal stability of the endosymbiont proteins (Rispe et al., 2004). In this context, the overexpression of chaperonins such as GroEL are thought to compensate the altered structure of many of these proteins (Fares et al., 2004).

2.2. Mealybugs

Mealybugs (Hemiptera: Pseudococcidae) receive such a name because of the powdery secretions covering the bodies of immature stages and adult females. They are mostly phloem-sucking plant parasites (Hardy et al., 2008), being common pests of a wide range of fruit crops, field crops and ornamentals. They feed on a great diversity of plant hosts, and some species are extremely polyphagous. For instance, the long tailed mealybug *Pseudococcus longispinus* (Targioni-Tozzeti) has been reported from more than 85 families of vascular plants (ScaleNet: http://www.sel.barc.usda.gov/scalenet/scalenet.htm). Economic damage caused by mealybugs is

not only due to plant feeding or fruits contamination, but also to viruses transmission. For example, several mealybugs species (including *P. longispinus* but also the obscure mealybug *Pseudococcus viburni* Signoret, the vine mealybug *Planococcus ficus* Signoret and the citrus mealybug *Planococcus citri* Risso) are known to act as vectors for the transmission of a set of viruses species collectively known as grapevine leafroll-associated viruses (GLRaV) (Tsai et al., 2010; Daane et al., 2011; Kol-Maimon et al., 2014). Thus, they are considered as a dangerous pest even at low population densities. Moreover, the wax-like coating of their bodies is thought to favor the development of mold fungi, which interferes with plant photosynthesis and decreases ornamental plants quality, while allows them to exhibit high levels of resistance to insecticides (Copland et al., 1985; Walker, 2000).

Due to their small size and cryptic behaviour, scale insects often remain undetected during quarantine inspections, thus typically being invasive pests. Consistently, mealybugs are world-wide distributed, representing the third most commom family of alien insects in Europe, just behind aphids and armored scale insects (Hemiptera: Diaspididae). About 37 mealybugs species have been newly stablished in our continent, being up to a quarter of the mealybug fauna currently present in Europe (Roques et al., 2009).

2.2.1. Life cycle and sex determination

Mealybugs' life cycle includes nymphs of first, second, and third-instars in females. In contrast, male immature stages include only first and second-instars nymphs, as well as prepupae and pupae instars (Figure 4, page 25). Adult females (1-3 mm in length) are ovoid, resembling the nynphal stages. Adult males are significantly smaller, winged, non-feeding and short-lived. Temperature has profound effects on insect life history parameters such as developing time, survival, longevity and reproduction. Males and females successfully develop into adults at temperatures ranging from 15 to 32ºC and from 18ºC to 32ºC, respectively, reaching an optimum at about 25ºC (Narai and Murai, 2002; Goldasteh et al., 2009).

Fitness of herbivorous insects is typically affected by the components of host plant quality, such as carbon, nitrogen and defensive metabolites (Awmack and Leather, 2002). As expected, this is also the case of mealybugs, according to studies on citrus, coffee, pumpkin and ornamental plants such as *Rosa hybrida* (Rosaceae), *Nerium oleander* (Apocynaceae), *Syngonium podophyllum* (Araceae), *Schefflera arbicola* (Araliaceae) and *Kalabanchoe blossfeldiana* (Crassulaceae) (Ben-Dov, 1994; Malleshaiah et al., 2000; Laflin and Parella, 2004; Polat et al., 2007). Taking advantage of their typical polyphagia, several food sources have been succesfully used in order to rear mealybugs under laboratory conditions (e.g., pumpkin fruits, bleached potato sprouts, potted citrus, and broad bean sprouts).

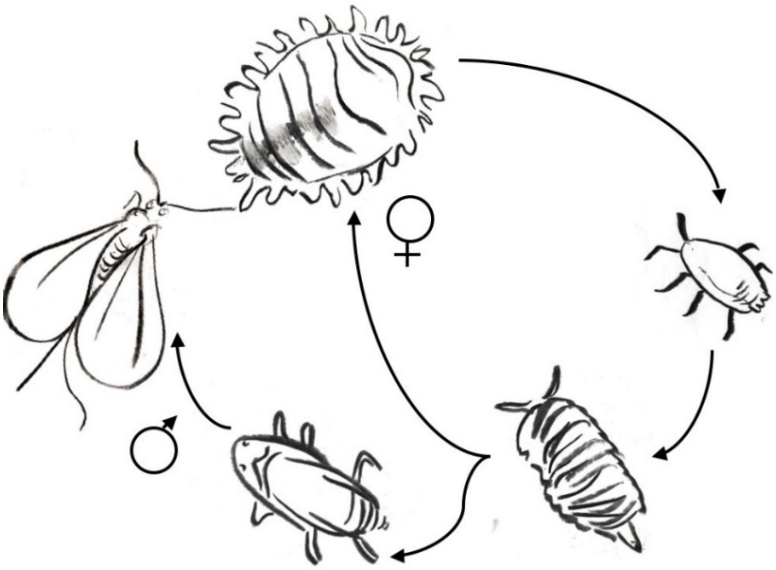


Figure 4. Life cycle of *Planococcus citri*. After hatching, females go through three inmature nynphal instars before maturity. In contrast, males go through four pre-adult stages, including prepupa and pupa.

Both physical (e.g., extreme temperatures; Goldasteh et al., 2009) and biological (e.g., host plant species, female aging; Polat et al., 2007, Buglia et al., 2009) factors are known to cause significant changes in the sex ratio, likely because sex determination mechanisms appear to be directly related to genomic imprinting. The genome of females from most mealybug species is organized into five pairs of functional chromosomes, without identifiable sex ones (Lecanoid system). In contrast, male mealybugs behave functionally as haploids (parahaploidy), since most genetic material contributed by the sperm is early inactivated in male embryos. Heterochromatized chromosomes do not proceed through spermatogenesis, so that only the maternally inherited euchromatic genome is transmitted to the next generation via males. Thus, at the end of meiosis, a quadrinucleate syncytium produces only two vital spermatids due to the elimination of such heterochromatized chromosomes. Both spermatids differ regarding their content in Heterochromatic Protein 1 (HP1) and the methylation level of the histone H3 at the lysine 9 (H3K9me3), which is a specific binding-site for several chromodomain proteins such as HP1, and is involved en gene silencing via heterochromatin formation (Buglia and Ferraro, 2004). Furthermore, HP1 is known to interact with the protein Su(Var)3-9, which has a specific H3-K9 histone methyltransferase activity. Therefore, differential protein content of spermatids appears to be responsible for male formation (through heterochromatization) in normal mating events. In coherence, the ooplasm of gametes from aged females, who significantly produce more sons than daughters, is known to be enriched in heterochromatic proteins (i.e., HP1 and Su(Var)3-9) (Buglia et al., 2009).

2.2.2. Mealybugs management and systematics

The high morphological similarity exhibited by mealybug taxa has traditionally been a major problem for the management of these pests, as well as for the study of their systematics and population biology. Species identification originally relied on microscopic cuticular features only (Figure 5). Therefore, it was time-consuming, required a high level of expertise and could not always distinguish between taxa, particularly if they were closely related (i.e., cryptic species). Moreover, since their similarity is higher in immature developmental stages, identification keys are usually available for pre-oviposition adult females (Millar, 2002; Watson and Kubiriba, 2005). Thus, when detected, immature mealybugs must be reared to adults before proper identification, a process that may take up to 6 weeks (Beuning et al., 1999). Additionally, microscopic cuticular features can show phenotypic plasticity, being also affected by environmental conditions (e.g., rearing temperature; Cox, 1983). In this context, mealybug management was strongly challenged by frequent species misidentification, which decreased the efficiency of crop protection methods and increased pesticide use.

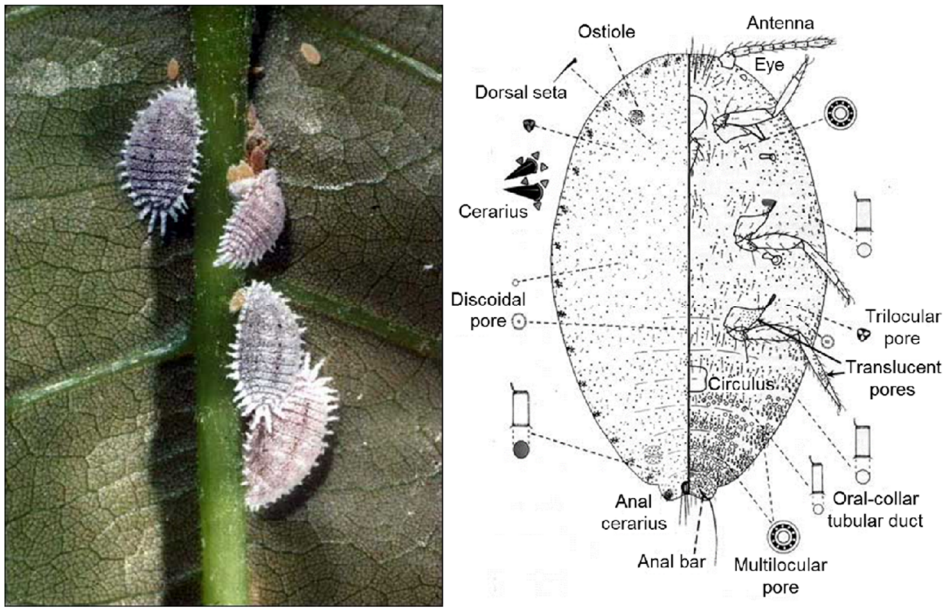


Figure 5. General characteristics of *Planococcus citri*. (Left) Adult females and crawlers; (Right) Morphological features traditionally used for *P. citri* taxonomy. Source: http://www.sel.barc.usda.gov/

Such situation motivated studies on the molecular characterization of mealybugs, in order to avoid the effect of sex, developmental stage or environmental conditions on their identification. Various DNA markers have been identified to date, including *loci* from the nuclear genome (18S rDNA, 28S rDNA, elongation factor 1 alpha, transcribed spacers I and II, triose phosphate isomerase and dynamin; Beuning et al., 1999; Downie and Gullan, 2004; Hardy et al., 2007; Hardy et al., 2008), the mithocondrial genome (cytochrome oxidase subunit I, cytochrome oxidase

subunit II, 16SrDNA; Thao et al., 2002; Baumann et al., 2005; Rung et al., 2008), and the genome of the betaproteobacterial endosymbiont of Pseudococcinae mealybugs (16S rDNA, 23S rDNA, 16S-23S rDNA and *rps*15-16S rDNA; Thao et al., 2002; Baumann et al., 2005). In the last decades, diverse molecular approaches have been used for mealybugs identification, including species-specific PCR (simple or multiplex) (Beuning et al., 1999; Demontis et al., 2007; Rung et al., 2008; Saccaggi et al., 2008; Daane et al., 2011), PCR plus RFLP (Cavalieri et al., 2008) or RAPD (Demontis et al., 2007).

Some of these *loci* have also been used for phylogenetic analyses, providing knowledge on mealybug major clades and their evolutionary relationships. This unexplored field was first addressed by Downie and Gullan on 64 mealybugs species by using partial DNA sequences from 3 nuclear genes, EF-1a (elongation factor 1 alpha), 18S rDNA and 28S rDNA. Their study revealed the existence of subfamilies Pseudococcinae and Phenacoccinae whithin the family Pseudococcidae (Downie and Gullan, 2004). Furthermore, they recognized the tribes Pseudococcini, Planococcini and Trabutinini within the Pseudococcinae. Such concepts were posteriorly refined by Hardy and coworkers, who added 33 new mealybug species and used a 2170 characters dataset, including both molecular and morphological features (i.e., characters from adult females, adult males and first-instar nynphs). Thus, they provided a more useful subfamily-level classification by identifying family-level diagnostic morphological features and presenting a list of all genera included in each subfamily (Hardy et al., 2008). According to their results, Pseudococcidae includes two primary clades: Pseudococcinae (already stated by Downie and Gullan in 2004), and Phenacoccinae. Regarding Pseudococcinae, statistical support was found only for subclades Trabutinini, Pseudococcini and the informal *Ferrisia* (Fullaway) group. Genus *Maconellicoccus* (Ezzat) forms a sister clade with respect to all other Pseudococcinae, as previously hypothesized by Downie and Gullan, while monophyly is not recovered either for the genera *Dysmicoccus* (Ferris) or *Pseudococcus* (Westwood) (Downie and Gullan, 2004; Hardy et al., 2008; Malausa et al., 2011). On the other hand, Hardy and coworkers redefined subfamily Phenacoccinae by including the hypogaeic mealybugs as the well supported monophyletic tribe Rhizoecini, which contains some species from genus *Phenacoccus* (i.e., the Madeira mealybug *Phenacoccus madeirensis* Green and the solanum mealybug *Phenacoccus solani* Ferris). Thus, monophyly of genus *Phenacoccus* is not supported either (Hardy et al., 2008).

2.3. Intracellular bacteria and mealybugs

Like other hemipteran phloem-feeding insects belonging to the suborder Sternorrhyncha (i.e., others scale insects, aphids, psyllids and whiteflies; Gullan and Martin, 2009), mealybugs rely on bacterial endosymbionts to complement their nutritionally poor diets. Although pseudococcids were originally thought to be monosymbiotic, early microscopic studies by Paul Buchner revealed an intriguing diversity among different lineages, which are particularly pronounced among Phenacoccinae species (Buchner, 1965). Furthermore, striking morphological variations known

as “symbiont cycles”, “infection forms”, “vegetative forms” or “hungry forms” were described even for single bacterium species depending on the host stage, transmission to the offspring, and environmental factors (Buchner, 1965; Tremblay, 1989). Bacterial endosymbionts showed to be located in a large oval bacteriome, enveloped by an epithelial layer and placed in the mealybugs abdomen (Fukatsu and Nikoh, 2000; von Dohlen et al., 2001). Bacteriomes are about 300 to 500 µm by 150 to 200 µm, including up to 100 bacteriocytes, approximately (Gatehouse et al., 2012). Interestingly, while endosymbionts were freely distributed in the cytoplasm of bacteriocytes from analyzed Phenacoccinae genera, those from Pseudococcinae appeared included in a number of cytoplasmic vesicle-like structures which were named as “mucus packets” or “mucous spherules” (Buchner, 1965; Tremblay, 1989). Moreover, such atypical symbiotic spheres (defined by three unit membranes) were structurally different to any known metazoan organelle or vesicle (Louis and Giannotti, 1974; von Dohlen et al., 2001). Thus, an unprecedented organization of the endosymbiotic system was early described for the analyzed Pseudococcinae mealybugs.

2.3.1. Nested endosymbiosis in Pseudococcinae mealybugs

Early culture-independent approaches to the description of the P-endosymbiont from Pseudococcinae mealybugs yielded contradictory results. Thus, while Munson and coworkers found evidences for the presence of a monophyletic betaproteobacterium in *P. longispinus*, the grape mealybug *Pseudococcus maritimus* (Ehrhorn), and the grey pineapple mealybug *Dysmicoccus neobrevipes* (Beardsley), only a gammaproteobacterium was found by Kantheti and colleagues in the cacao mealybug *Planococcus lilacinus* (Cockerell) (Munson et al., 1992; Kantheti et al., 1996). Such apparently discrepant results were reconcilied by Fukatsu and Nikoh, who revealed the co-localization of both a beta- and a gammaproteobacterial endosymbiont inside the “mucus spherules” of the bacteriocytes from the bamboo pseudococcid *Antonina crawii* (Cockerell) (Fukatsu and Nikoh, 2000). Moreover, von Dohlen and colleagues demonstrated on *P. citri* that these bacteria were not actually included in the cytoplasmic spherules, but that these vesicle-like structures were, in fact, the cells of the betaproteobacterial endosymbiont of the Pseudococcinae mealybugs (Figure 6, page 29). Approximately 8-10 betaproteobacterial cells (carrying 10-20 rod-shaped gammaproteobacteria) are found per bacteriocyte, while cell sizes of both endosymbionts (10-20 µm for the the betaproteobacterium and 3-6 µm for the gammaproteobacteria) where in the range of known eubacterial cells. Thus, they reported an striking case of intracellular nested endosymbiosis, which has been afterwards described for many others Pseudococcinae species such as the Japenese mealybug *Planococcus kraunhiae* (Kuwana), the comstock mealybug *Pseudococcus comstocki* (Kuwana) (Kono et al., 2008), *P. longispinus*, the citrophilus mealybug *Pseudococcus calceolariae* (Maskell), *P. viburni* (Gatehouse et al., 2012) and the azalea mealybug *Crisicoccus azaleae* (Tinsley) (Koga et al., 2012). In coherence with what is recurrently found among P-endosymbionts of insects, the three-membrane bilayers defining the cytoplasm of the betaproteobacterium would correspond to the

typical gram-negative bacteria double membrane enclosed by a third, host-derived membrane (Douglas, 1989; von Dohlen et al., 2001).

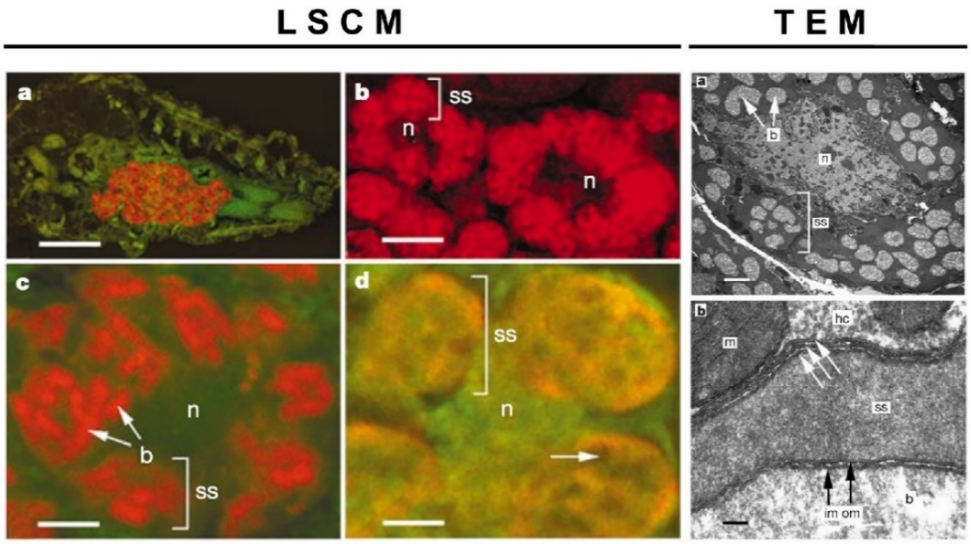


Figure 6. Nested endosymbiosis of *Planococcus citri*. (LSCM) Laser-scanning confocal microscopy. Scale bars: 0.42 mm (**a**), 10.5 µm (**b**), 4.5 µm (**c**), 4.7 µm (**d**); (TEM) Transmission electron microscopy. Scale bars: 2.33 µm (**a**) and 0.07 µm (**b**). b: Bacteria; hc: Host cell cytoplasm; im: Inner membrane; m: Mitochondrion; n: Nucleus; om: Outer membrane; ss: Symbiotic sphere. Adapted from von Dohlen et al., 2001.

The monophyletic origin of the betaproteobacterial endosymbiot was further studied by extending the limitted species sampling early performed by Munson and colleagues (Munson et al., 1992; Thao et al., 2002). Thao and coworkers revealed that endosymbiont phylogeny included (at least) two major monophyletic clades which appear to match with the presence/absence pattern of inner gammaproteobacterial endosymbionts among Pseudococcinae lineages. Up to five different subclades (A to E) were recognized in the biggest cluster. Moreover, they confirmed cospeciation of betaproteobacterial endosymbionts and insect hosts through the congruence between the topology of major betaproteobacterial clades and subclades (A, C, E and F) and that obtained with the analysis of mitochondrial sequences from a set of representative Pseudococcinae species, including *P. citri*, the pineapple mealybug *Dysmicoccus brevipes* (Cockerell), the wattle mealybug *Melanococcus albizziae* (Maskell) and the pink hibiscus mealybug *Maconellicoccus hirsutus* (Green) (Thao et al., 2002). Cospeciation has been recurrently shown, either using mitochondrial or nuclear *loci* of the insect hosts (Figure 7, page 30) (Downie and Gullan, 2005; Baumann and Baumann, 2005). Thus, the common ancestor of current Pseucococcinae species would have been infected by an ancestor of the current betaproteobacterial endosymbiont, at least 100-200 million years ago (Thao et al., 2002; Hardy et al., 2008). Authors named it as "*Candidatus* Tremblaya princeps". Genus name (Trem.bla'ya. N. L. fem. n.) refers to the Italian entomologist Ermenegildo Tremblay, who made great

contributions to the knowledge of endosymbionts of plant sap-sucking insects (Iaccarino and Tremblay, 1970; Tremblay, 1989; Tremblay, 1990; Tremblay, 1997), while species name (prin`ceps. N. adj., first in rank) refers to the role of Pseudoccinae primary endosymbiont as the first described species of the genus "*Ca*. Tremblaya" (Thao et al., 2002). The beta-endosymbiont of *D. brevipes*, 65 kb of whose genome were early sequenced and analyzed, was proposed as the type strain (Baumann et al., 2002; Thao et al., 2002).

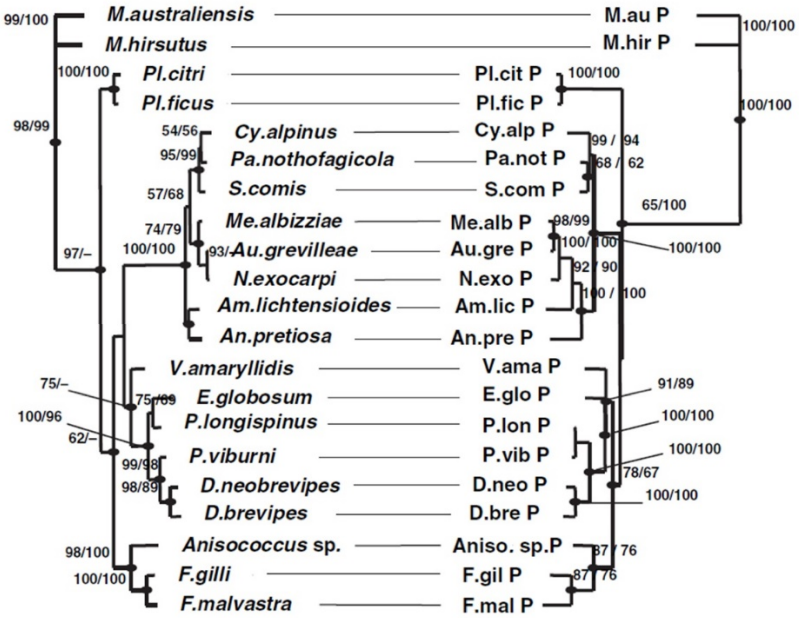


Figure 7. Co-evolution of Pseudococcinae mealybugs and their betaproteobacterial endosymbionts. ML/MP trees inferred for both the host insects (left) and their "*Ca*. Tremblaya princeps" strains (right) are depicted (Downie and Gullan, 2005).

In contrast to the case of "*Ca*. Tremblaya princeps", gammaproteobacterial endosymbionts have been found in most, but not all tested Pseudococcinae species (being absent in the genus *Maconellicoccus* and the *Ferrisia* group). Moreover, they showed a polyphyletic origin along the gammaproteobacteria phylogeny, suggesting that they come from at least four independent infection events of "*Ca*. Tremblaya princeps" (Figure 8, page 31) (Thao et al., 2002). This fact likely explains (at least partially) the morphological diversity early noticed among mealybugs endosymbionts (Buchner, 1965; Iaccarino and Tremblay, 1970; Tremblay, 1989). However, adequacy of such gammaproteobacteria to the concept of secondary symbionts is only partial, since built clusters were coherent with those observed for "*Ca*. Tremblaya princeps" subclades (Thao et al., 2002). Thus, phylogenetic analyses suggest that after the independent infections of "*Ca*. Tremblaya princeps" by the corresponding gammaproteobacterial symbionts, both organisms have been transmitted together from mothers to offspring (Figure 9, page 32). This is coherent with the cytological details of the nested endosymbiotic system transmission to the next

generation, where “symbiotic spheres” are transferred from gravid females to the embryos as a unit. So, the complete nested endosymbiotic system is released upon the host bacteriocytes breakage. They cluster around the nutrient plasma cord that connect the nurse cells to the anterior pole of the oocyte, eventually penetrating the oocyte. Once embryogenesis is in progress, symbiotic spheres migrate to the embryo host cells, which also move from the middle of the embryo in order to engulf them (von Dohlen et al., 2001).

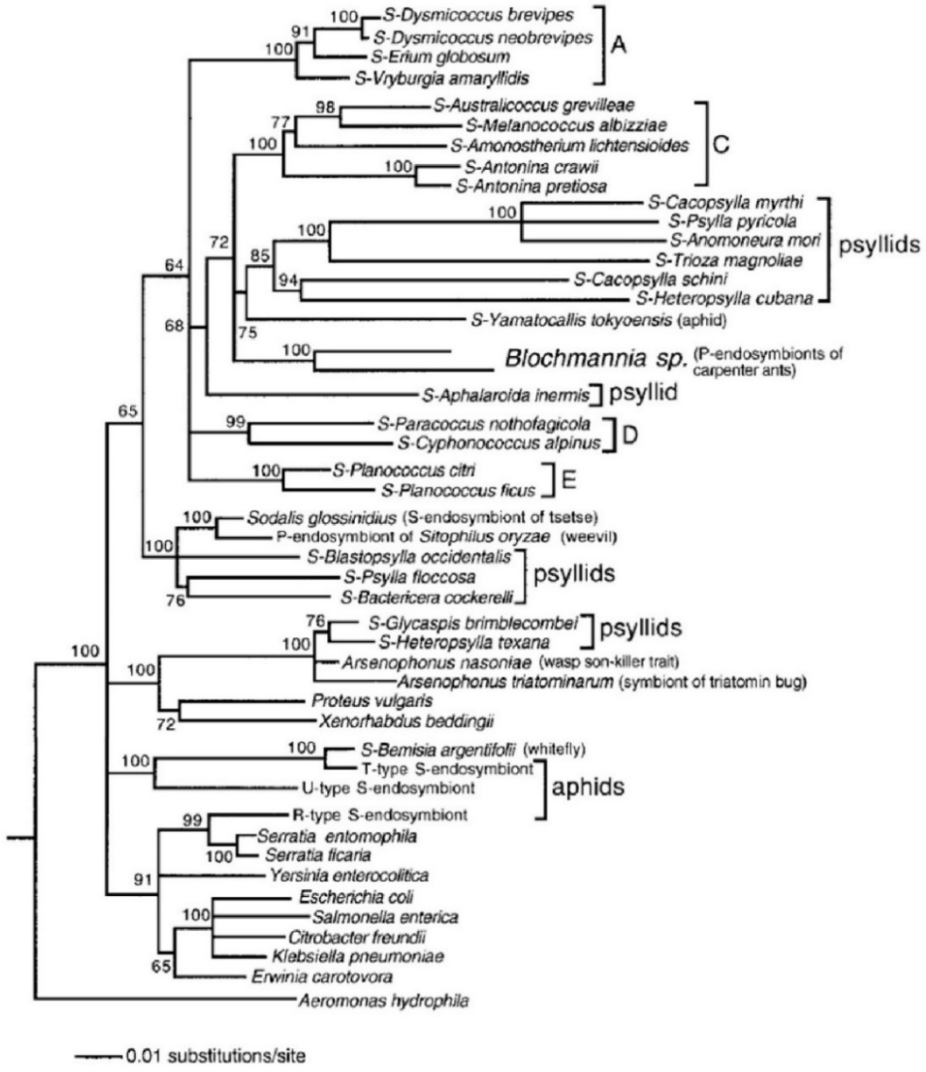


Figure 8. Evolutionary origins of mealybug gammaproteobacterial symbionts. Independent acquisition of gammaproteobacterial symbionts ocurred several times, according to the neighbor joining inference (Thao et al., 2002).

Early experiments suggested the existence of tight interactions both at symbiont-symbiont and host-symbiont levels. Thus, pseudococcids ingestion of antibiotics or exposure to high temperatures causes the degeneration of the whole endosymbiotic system and, subsequently, the death of the host insect (Köhler and Schwartz, 1962). Nested organization is known to be stable (i.e., neither gammaproteobacteria have been detected outside “*Ca*. Tremblaya princeps” cells nor empty betaproteobacterial cells have been reported; von Dohlen et al., 2001), while both

beta and gammaproteobacterial endosymbionts exhibit similar population dynamics along the host insect life cycle (Kono et al., 2008). They increase during nynphal and female development, reach a peak in actively reproducing females and decline in old females, once they had ceased reproduction. In contrast, the complete endosymbiotic system degenerates along male developmental pathway, being almost lost in adult males (Kantheti et al., 1996; Kono et al., 2008). Since males neither feed nor grow along its life cycle (they mostly molt and metamorphose), this population dynamics further confirmed the putatively essential nutritional role of these bacteria, as usually found among endosymbionts of sap-feeding insects. So, in spite of the polyphyletic evolutionary origin of the gamma-endosymbionts and their apparent absence in some Pseudococcinae lineages, its relative abundance in the bacteriome and the intimate endocellular symbiotic association established with "*Ca*. Tremblaya princeps" cells suggest such bacteria to be as relevant as the betaproteobacterial endosymbiont for insect host survival (Kono et al., 2008; Fukatsu and Nikoh, 2000).

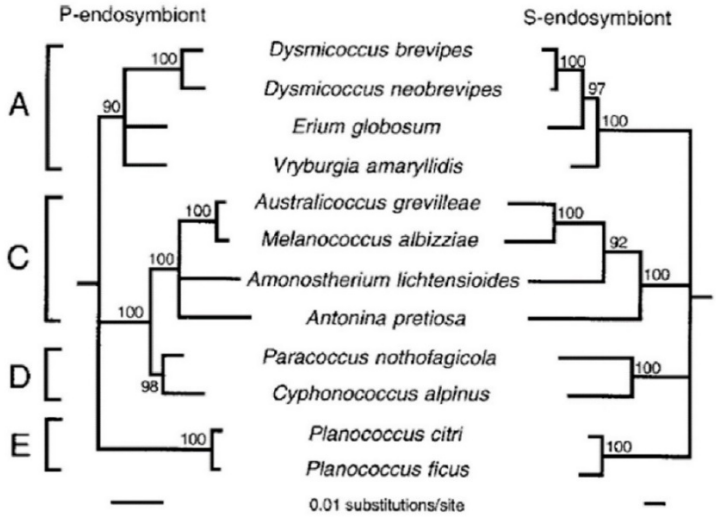


Figure 9. Co-evolution of mealybug endosymbionts. ML trees show congruent phylogeny for betaproteobacteria (P) and gammaproteobacteria (S) endosymbionts (Thao et al., 2002).

2.3.1.1. Atypical molecular characteristics for a long-term bacterial endosymbiont

Mealybug endosymbiosis has been traditionally considered as atypical. The role as P-endosymbiont of the betaproteobacterium "*Ca*. Tremblaya princeps" itself was originally considered as a rarity, since many early-described insect P-endosymbionts belong to the Class Gammaproteobacteria (e.g., aphids, psyllids, whiteflies, tsetse flies, ants, weevils) (Munson et al., 1992). Additionaly, the nested endosymbiosis found in most Pseudococcinae mealybugs represents the only case of prokaryote-prokaryote endocellular symbiosis ever described in nature (von Dohlen et al., 2001). Moreover, as early revealed by the partial genomic characterization of "*Ca*. Tremblaya princeps" strain from *D. brevipes*, the "*Ca*. Tremblaya princeps" genome presents some molecular and structural properties that distinguish it from most

other long-term P-endosymbionts, including high GC-content, low gene density, maintenance of duplicated paralogous *loci* and genome structural instability (Baumann et al., 2002; McCutcheon and von Dohlen, 2011).

In contrast with the typical genomic AT-enrichment found in most insects P-endosymbionts, "*Ca*. Tremblaya princeps" displays a high GC-content (e.g., 57.1%, according to the partial genomic exploration of the bacterial strain from *D. brevipes*) (Munson et al., 1992; Baumann et al., 2002). Moreover, no differences regarding GC-content could be found either between structural genes and intergenic regions (IGRs), or between *loci* coding for highly and poorly conserved proteins. In coherence, amino acids encoded by GC-enriched codons (i.e., alanine, glycine, arginine and proline) were estimated to represent about one-third of the proteome's residues, revealing the impact of the GC-content on proteins composition (Baumann et al., 2002). Although an increase in gene density tends to go together with the genome reduction in long-term endosymbionts, low gene density was early reported for "*Ca*. Tremblaya princeps". Thus, a gene density of 81.6% was estimated for "*Ca*. Tremblaya princeps" of *D. brevipes*, while such data calculation on the complete genome of "*Ca*. Tremblaya princeps" strains PCIT and PCVAL from *P. citri* exceeded the 70%, with an average IGRs length of 389-467 bp (Baumann et al., 2002; McCutcheon and von Dohlen, 2011; López-Madrigal et al., 2011; López-Madrigal et al., 2013a). Some authors stated that such atypical trait would be mainly a consequence of the unexpectedly large number of annotated pseudogenes (McCutcheon and von Dohlen, 2011; Husnik et al., 2013). However, we found that low gene density does not affect "*Ca*. Tremblaya princeps" genome homogeneously, and that pseudogenes presence cannot fully explain it (López-Madrigal et al., submitted). In spite of the low gene density, several identical paralogous *loci* can be found in "*Ca*. Tremblaya princeps", as a result of a single partial genome duplication that affected the complete ribosomal operons and its genomic context. The duplication includes the 3'-end of the gene *leu*A (coding for the alpha-isopropylmalate synthase, EC 2.3.3.13), the complete *rps*O gene (coding for the ribosomal protein S15) and the 5'-end of the gene *rsm*H (encoding 16S rRNA m^4C1402 methyltransferase, EC 2.1.1.199), at least. Baumann and coworkers found evidences for such partial genome duplication in several "*Ca*. Tremblaya princeps" strains, including representatives of the clades A, C, E and F (Thao et al., 2002), suggesting that this event would be ancestral in the "*Ca*. Tremblaya princeps" lineage (Baumann et al., 2002). Our comparative genomics analysis between "*Ca*. Tremblaya princeps" from *P. citri* and "*Ca*ndidatus Tremblaya phenacola" from the oat mealybug *Phenacoccus avenae* (Borchsenius) revealed that the partial genome duplication affecting the ribosomal operon occured before the evolutionary split of subfamilies Pseudococcinae and Phenacoccinae (López-Madrigal et al., submitted). Furthermore, duplicated fragments appear to be evolving under concerted evolution (CE), since they remain identical (e.g., 5,689 bp in "*Ca*. Tremblaya princeps" from *D. brevipes,* and 5,702 bp in strains PCIT and PCVAL from *P. citri*) after a long time of independent evolution. In our studies, we showed CE to occur also in "*Ca*. Tremblaya princeps" strains from *P. ficus*, *P. viburni*, *P. longispinus* and the gray sugarcane mealybug *Dysmicoccus boninsis* (Kuwana). CE is a universal biological phenomenon that leads to the homogenization of the polymorphisms independently

accumulated by DNA sequences belonging to a given repetitive family. In such a case, greater sequence similarity is found within a species than between species, suggesting that the corresponding repetitive DNA sequences, whether coding or noncoding, are not evolving independently but in a concerted fashion (Liao 1999; Innan and Kondrashov, 2010; Hastings, 2010). It is a non-reciprocal recombination process, which is thought to be directed by the action of proteins involved in DNA recombination, replication and repair. Although no HR genes have been annotated in the "*Ca*. Tremblaya princeps" genome, conserved identity of these paralogous *loci* is not the only molecular trait evidencing the occurrence of HR events (Baumann et al., 2002; McCutcheon and von Dohlen, 2011; López-Madrigal et al., 2011). While highly reduced genomes typically display an extreme structural stability, due to early loss of recombinogenic activities and the scarcity of repetitive sequences (Tamas et al., 2002; Silva et al., 2003), McCutcheon and von Dohlen (2011) showed the surprising existence of structural instability at the population level for the strain PCIT of *P. citri*. The authors reported an inversion polymorphism affecting a 7,032 bp region flanked by 71-bp inverted repeats, so that both orientations can be found even within a single *P. citri* individual.

Since genomic AT-enrichment, high gene density, loss of redundant or unnecessary genetic functions and structural stability are common features of long-term bacterial endosymbionts, the peculiarities of the "*Ca*. Tremblaya princeps" genome indicate that reductive evolution in this organism is not acting in the expected manner. The unconventional reductive genome evolution of "*Ca*. Tremblaya princeps" has been associated with the nested gammaproteobacterial endosymbionts found in most Pseudococcinae mealybugs (Husnik et al., 2013). We have suggested that HR genes, recurrently acquired through the independent internalization of several gammaproteobacterial endosymbionts, could be on the root of such association. The analysis of DNA repeats along the "*Ca*. Tremblaya princeps" PCVAL genome suggested it to be more sensitive to the action of a functional HR machinery than its gammaproteobacterial partner (López-Madrigal et al., submitted). In spite of this uncommon genomic features for an endosymbiont, the genome of "*Ca*. Tremblaya princeps" strains PCIT and PCVAL is only 139 kb in length (López-Madrigal et al., 2011; McCutcheon and von Dohlen, 2011). It is the second most reduced prokaryote genome known so far, just behind the betaproteobacteria "*Candidatus* Nasuia deltocephalinicola" (112 kb), the co-primary endosymbiont of the aster leafhooper *Macrosteles quadrilineatus* (Forbes) (Auchenorryncha: Cicadellidae) (McCutcheon and von Dohlen, 2011; Bennett and Moran, 2013). Moreover, although variable, inferred rates of nucleotide substitution among "*Ca*. Tremblaya princeps" lineages appears to be significantly faster than those of free-living betaproteobacteria (from 3.4 to more than 7-fold faster) (Gruwell et al., 2010; Koga et al., 2012).

2.3.1.2. Mealybugs nested endosymbiosis: a tripartite relationship

The complete genome sequence of the bacterial endosymbionts from *P. citri* allowed to postulate the biosynthesis of essential amino acids as the main output of the nested endosymbiosis in mealybugs, as already stated for other phloem-feeding insect groups in the

suborder Sternorrhyncha (e.g., aphids, psyllids, whiteflies) (López-Madrigal et al., 2011; McCutcheon and von Dohlen, 2011). We reported the recurrent maintenance of genes involved in essential amino acids biosynthesis in endosymbionts from several Pseudococcinae mealybugs (*P. longispinus*, *P. viburni* and *D. boninsis*), further confirming this point (López-Madrigal et al., 2014). The endocellular gammaproteobacteria found in mealybugs belonging to the genus *Planococcus* was named "*Candidatus* Moranella endobia". Genus name *Moranella* refers to the American evolutionary biologist Nancy A. Moran, while *endobia* (endo = inside, bia = living; femenin form) comes from the surprising characteristic of living exclusively inside "*Ca*. Tremblaya princeps" cells (von Dohlen et al., 2001). "*Ca*. Tremblaya princeps" from *P. citri* possess, at least, 29 genes involved in essential amino acids biosynthesis (López-Madrigal et al., 2011; McCutcheon and von Dohlen, 2011). Even though it represents more than 20% of the total length of this tiny genome, none of the corresponding biosynthetic pathways is complete (Figure 10, page 36). Only the pathways for the biosynthesis of leucine and valine are nearly complete, except for the absence of the small subunit of the acetohydroxybutanoate synthase/acetolactate synthase (EC 2.2.1.6; encoded by *ilv*N in other bacteria) and the branched-chain amino acid aminotransferase (BCAT, EC 2.6.1.42, encoded by *ilv*E), being the latter recurrently lost in many other endosymbiotic bacteria. *ilv*N is not essential for the proper functioning of the pathway, since the large subunit of acetohydroxybutanoate synthase/acetolactate synthase (EC 2.2.1.6; encoded by *ilv*B) contains the active sites for both enzymatic activities and is known to be functional in absence of IlvN (Weinstock et al., 1992). In contrast, the host insect would be in charge of the BCAT activity (the corresponding gene has already been found in fully sequenced insect genomes; Wilson et al., 2010), as suggested by its significant upregulation in *P. citri* bacteriocytes (Husnik et al., 2013). Regarding methionine biosynthesis, cooperation between "*Ca*. Tremblaya princeps" and the host insect, which appears to supply both cystathionine gamma-lyase (CGL) and cystathionine beta-lyase (CBL) activities, has also been stablished. On the other hand, the genome of the corresponding "*Ca*. Moranella endobia" strain (538 kb) encodes 16 proteins involved in the essential amino acids biosynthesis, most of which do not present functional homologs in "*Ca*. Tremblaya princeps" (McCutcheon and von Dohlen, 2011; López-Madrigal et al., 2013a; Husnik et al., 2013). Thus, a complex patchwork of gene products coming from both bacteria would be necessary in order to accomplish the biosynthesis of tryptophan and threonine (Figure 10, page 36). A similar evolutionary scenario can be found in the nested consortia of several Pseusococcinae mealybugs, where each endosymbiont seems functionally specialized, and genetic redundancy appears to be scarce (López-Madrigal et al., 2014). This intricate metabolic cooperation is unprecedented in multispecies symbiotic communities described in other insects, where complete or near-complete pathways for the synthesis of a certain nutrient is exclusively encoded by one of the members of the consortium (Wu et al., 2006; McCutcheon et al., 2009). Moreover, the biosynthesis of phenylalanine, arginine and isoleucine implies the participation of both endosymbionts and also the host insect through the supply of several enzymatic activities (i.e., aspartate aminotransferase, AAT, EC 2.6.1.1; ornithine aminotransferase, OAT, EC 2.6.1.13; cystathionine gamma-lyase, CGL, EC 4.4.1.1; cystathionine

beta-lyase, CBL, EC 4.4.1.8; threonine dehydratase, TDH, EC 4.3.1.19) (McCutcheon and von Dohlen, 2011). Since only TDH was not found to be upregulated in *P. citri* bacteriocytes, the authors suggested that the CGL activity guarantees the availability of 2-oxobutanoate required for the biosynthesis of isoleucine (McCutcheon and von Dohlen, 2011; Husnik et al., 2013). All three genomes are also involved in the biosynthesis of lysine. However, in this case, the genes *dap*F and *lys*A, two of the four genes necessary to complete the pathway that are missing in the endosymbionts, have been found in the *P. citri* nuclear genome as a result of ancestral HGT events from additional symbiotic bacteria. The same applies to the prokaryotic gene *cys*K, wich participates in the biosynthesis of methionine and isoleucine through the production of the substrate for CGL activity. Similar supplementation strategies for genes missing in the biosynthetic pathways of lysine (i.e., *arg*D, *dap*E) or histidine (i.e., *his*C, *his*D) cannot be ruled out. Thus, the authors suggested the direct participation of the host in the biosynthesis of phenylalanine, leucine, valine, methionine, isoleucine, lysine and, possibly, histidine.

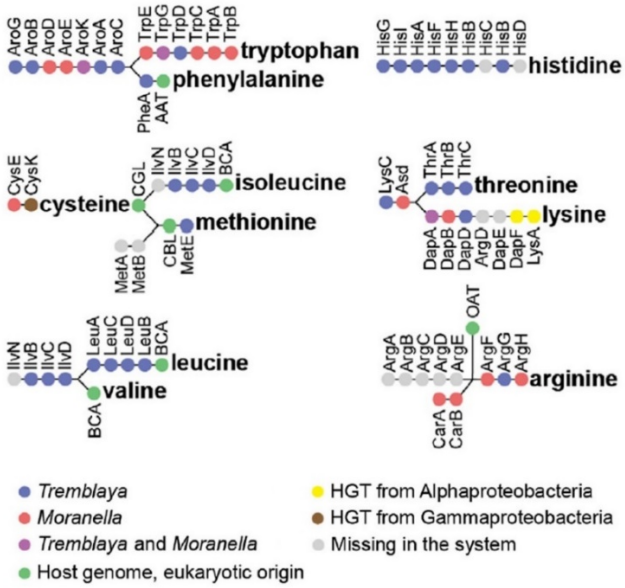


Figure 10. Essential amino acids biosynthesis in the *P. citri* tripartite symbiosis (adapted from Husnik et al., 2013).

A total of 22 functional prokaryotic genes have been already found in the host nuclear genome, likely as a consequence of multiple ancestral HGT events derived from transient infections of *P. citri* ancestors by a diverse set of S-symbionts (Husnik et al., 2013). Three of them could be only identified as members of family Enterobacteriaceae, while two of them did not cluster with bacteria currently known to act as S-symbionts. Regarding the rest (17), they showed to come from Bacteroidetes (2), Alphaproteobacteria (10) and Gammaproteobacteria (5) closely related to the bacterial genera *Wolbachia*, *Rickettsia* (Alphaproteobacteria), *Cardinium* (Bacteroidetes), *Serratia*, *Sodalis* and *Arsenophonus* (Gammaproteobacteria). As above mentioned, only three of

them (i.e., *dap*F, *lys*A and *cys*K) are devoted to fulfill missing gaps in the biosynthetic pathways of the essential amino acids lysine and methionine, while five of them appear to complement "*Ca*. Moranella endobia" incomplete pathways for the biosynthesis of riboflavin (*rib*A and *rib*D) and biotin (*bio*A, *bio*B and *bio*D). Interestingly, nine of them are involved in peptidoglycan metabolism, including both biosynthesis (*mur*A, *mur*B, *mur*C, *mur*D, *mur*E, *mur*F and *ddl*B) and recycling (*mlt*D and *ami*D), and might be involved in the control of the endosymbiotic system dynamics by the host insect. The large number of peptidoglycan production/recycling genes found in the *P. citri*'s nuclear genome appears to support that lysis of "*Ca*. Moranella endobia" cells would be the mechanism allowing the acquision of required essential gene products by "*Ca*. Tremblaya princeps" (McCutcheon and von Dohlen, 2011; Koga et al., 2013). This is coherent with the virtual incapability of the Sec translocon machinery to deal with the broad molecular exchange predicted to occur between both endosymbionts (López-Madrigal et al., 2013b). Moreover, it seems supported by early microscopic observations from Paul Buchner, who described at least to morphological forms of the gammaproteobacterial endocellular symbiont, apparently synchronized within a bacteriocyte and depending on the insect life stage (i.e., the "reproductive form", in which small sized cells were in the dividing process; and the "degenerative phase", in which cells were unevenly shaped and elongated). More recently, a subset of amorphous, apparently degenerating, gammaproteobacterial cells have been reported in the bacteriocytes of the Pseudoccocinae mealybug *C. azaleae* (Koga et al., 2013), while the uncoupling of "*Ca*. Tremblaya princeps" and its endocellular gammaproteobacterial endosymbiont, specially during males development, was previously demonstrated in the Pseudococcinae mealybugs *P. kraunhiae* and *P. comstocki* (Kono et al., 2008). Thus, authors have suggested that the insect host would be able to regulate the stability of "*Ca*. Moranella endobia" cell wall by differentially controlling the expression of the horizontally acquired genes involved in peptidoglycan metabolism. Upregulation of genes *murABCDE* would theoretically increase the integrity of the "*Ca*. Moranella endobia" cellular wall, while the upregulation of genes *mltD* and *amiD* would reduce its strength (Husnik et al., 2013). However, our immunostaining assays on the "*Ca*. Moranella endobia" proteins GroEL and MscL indicated that both proteins are mostly restricted to "*Ca*. Moranella endobia" cells. Based on these results we have proposed an alternative osmotically-driven mechanism for the molecular communication between both endosymbionts, where the regulation of the cell wall strength by *P. citri* could also play an important role (López-Madrigal et al., 2013b).

2.3.2. Conventional symbiosis of Phenacoccinae mealybugs

In contrast to the case of Pseudococcinae mealybugs, no "mucous spherules" envolving bacterial symbionts were detected during the exploration of bacteriocytes from mealybugs belonging to subfamily Phenacoccinae (genera *Phenacoccus*, *Heliococcus*, *Centrococcus*, *Ripersia* and *Eumyrmococcus*), suggesting the existence of drastic differences between both subfamilies regarding the structural organization of their endosymbiotic systems (Buchner, 1965;

Tremblay, 1989; Hardy et al., 2008). Bacterial endosymbionts of 20 Phenacoccinae mealybug species have been recently analyzed by Gruwell and coworkers, who applied only phylogenetic criteria in order to characterize the corresponding P-endosymbionts (Hardy et al., 2008; Gruwell et al., 2010). Most of the analyzed species (i.e., those belonging to genera *Phenacoccus*, *Heliococcus*, *Heterococcus*, *Mirococcus*, *Oxyacanthus* and *Peliococcus*) harbor a betaproteobacterial P-endosymbiont forming a sister clade to "*Ca*. Tremblaya princeps". The authors named this new species "*Ca*. Tremblaya phenacola". It is the P-endosymbiont of Phenacoccinae mealybugs, with the exceptions of (at least) species belonging to the genus *Rastrococcus* and those from the tribe *Rhizoecini* , where it appears to have been replaced by symbiotic bacteria from the phylum Bacteroidetes (Figure 11) (Gruwell et al. 2010). The endosymbiont found in *Rhizoecini* has been named "*Candidatus* Brownia rhizoecola" (Gruwell et al. 2010).

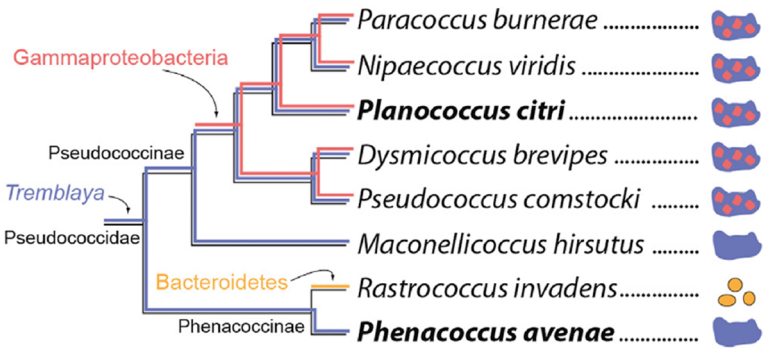


Figure 11. Cladogram of selected mealybug species and their bacterial endosymbionts. Lineages where "*Ca*. Tremblaya" (blue) coevolves with diverse gammaproteobacterial partners (red) or has been replaced by Bacteroidetes (yellow) are indicated (Husnik et al., 2013).

Both Fluorescence *In Situ* Hybridization (FISH) and Transmission Electron Microscopy (TEM) analysis on *Phenacoccus azaleae* (Kuwana) confirmed that Phenacoccinae mealybugs contain a simpler endosymbiotic system in which "*Ca*. Tremblaya phenacola" is located inside the host bacteriocytes without any gammaproteobacterial associates (Koga et al., 2012). We obtained identical results when analyzing the microbiota associated to the bougainvillea mealybug *Phenacoccus peruvianus* (Granara de Willink), further confirming this point (López-Madrigal et al., 2014). No evidences for additional endosymbionts were found when the complete genome of the "*Ca*. Tremblaya phenacola" PAVE, from *P. avenae*, was sequenced (Husnik et al., 2013). Moreover, a comparative analysis using the 16S rRNA gene advanced that the "*Ca*. Tremblaya phenacola" genome would be more AT-enriched than that of "*Ca*. Tremblaya princeps" (Baumann et al., 2002; McCutcheon and von Dohlen, 2011; Koga et al., 2012; Husnik et al., 2013), while it displays a 1.5 higher evolutionary rate. It has been suggested that such differences, both in GC-

content and evolutionary rate, could be related to the presence/absence of endocellular gammaproteobacterial endosymbionts (Koga et al., 2012). Moreover, "*Ca*. Tremblaya phenacola" PAVE possess also a tiny genome (171 kp) but with high gene density (93.5%), including very few pseudogenes, which is typical among long-term bacterial endosymbionts (Husnik et al., 2013). Our studies showed that, in contrast to "*Ca*. Tremblaya princeps" PCVAL, reductive evolution has homogeniously affected the genome of "*Ca*. Tremblaya phenacola" PAVE (López-Madrigal et al., submitted). Both "*Ca*. Tremblaya" genomes are highly collinear, although "*Ca*. Tremblaya phenacola" PAVE has suffered one single chromosomic inversion and contains a small plasmid coding for two of its ribosomal proteins. In spite of these molecular and structural genomic differences, both "*Ca*. Tremblaya" species are expected to display very similar nutritional roles, since the composition of the natural diets of mealybugs from subfamilies Pseudococcinae and Phenacoccinae is similar or even identical. This is the case of Japan populations of the Pseudococcinae mealybug *C. azalea*e and the Phenacoccinae mealybug *P. azaleae*. Both of them feed preferentially on the giant azalea *Rhododendron pulchrum* (Ericaceae), frequently forming mixed colonies in the host plant (Koga et al., 2012). Recurrent detection of *loci* involved in essential amino acids biosynthesis in the endosymbiotic systems of the Phenacoccinae *P. peruvianus* and *P. madeirensis* also support this idea (López-Madrigal et al., 2014). In coherence with its role as *P. avenae* sole endosymbiont, the genome of "*Ca*. Tremblaya phenacola" PAVE contains the complete set of genes collectively retained by "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" for essential amino acids biosynthesis in the dual symbiosis of *P. citri*. Furthermore, both endosymbiotic systems lack the same genes from the biosynthetic pathways of histidine (*argD*, *dapE*, *dapF* and *lysA*) and lysine (*hisC* and *hisD*), suggesting that the required intermediates and/or enzymes would be available in the bacteriocytes of both insects (Husnik et al., 2013).

2.3.3. Secondary symbionts in mealybugs

Although originally classified as S-symbionts, gammaproteobacteria inhabiting mealybug's bacteriocytes appear to be essential for their corresponding host survival. Thus, they should be considered co-primary endosymbionts, as already suggested by cytological, phylogenetic and functional evidences (von Dohlen et al., 2001; Thao et al., 2002; McCutcheon and von Dohlen, 2011). In contrast to other hemipteran insects, which usually harbor facultative bacteria (Moran et al. 2005; Pan et al. 2012), the description of real S-symbionts have been extremely rare among mealybugs. In fact, culture-independent approaches to their microbiome description typically reveal only the presence of one or two bacterial species belonging to the classes Gammaproteobacteria and/or Betaproteobacteria (Munson et al., 1992; Kantheti et al., 1996; Thao et al., 2002; Kono et al., 2008; Gruwell et al., 2010, Gatehouse et al., 2012; Koga et al., 2013). Only two cases of real S-symbionts have been reported to date, involving members of both mealybug subfamilies. The first described S-symbiont was a *Spiroplasma* identified by Fukatsu and Nikoh (2000) in the Pseudococcinae *A. crawii*, in addition to the typical co-primary beta- and

gamma-endosymbionts. Sequence information indicated that it is closely related to *Spiroplasma* spp. from ladybird beetles that, interestingly, are important predators of coccids. No conspicuous structure was detected for this S-symbiont, which showed intracellular location along various tissues such as gut, fat bodies and epithelia. In contrast to the co-primary bacteria, they were only occasionally detected in developing eggs, without exhibiting any specific localization. Moreover, diagnostic PCRs on four additional pseudococcid species (i.e., *Dysmicoccus wistariae* Green, *P. kraunhiae*, *P. citri* and *Pseudococcus citriculus* Green) failed to reveal the presence of *Spiroplasma*, suggesting a parasitic/commensal association with *A. crawii*.

In addition, Singh and coworkers (2013) reported the presence of a *Rickettsia* symbiont while surveying the bacterial biodiversity in the Phenacoccinae *Phenacoccus solenopsis* (Tinsley) by using genus-specific primers for the common S-symbiotic bacteria *Rickettsia*, *Cardinium*, *Wolbachia* and *Arsenophonus*. The newly identified S-symbiont showed a maximum similarity hit with the *Rickettsia* symbiont of an Indian population of the whitefly *B. tabaci*, while being clearly different from *Rickettsia* symbionts reported in other insects, including Israelian whiteflies. In contrast to the case of *Spiroplasma* in *A. crawii*, the authors showed that both the P-endosymbiont "*Ca*. Tremblaya phenacola" and *Rickettsia* co-localize in the host bacteriome, although no further information regarding their organization inside the bacteriocytes cytoplasm was supplied. Nevertheless, signals from both bacteria appear to be restricted to a minor insect body area, instead of the huge bacteriome typically found in mealybugs (Fukatsu and Nikoh, 2000; von Dohlen et al., 2001; Kono et al., 2008; Koga et al., 2013). Moreover, the pictures shown in the article present high levels of autofluorescence, which jeopardizes the accuracy of the conclusions. In any case, *Rickettsia* was found in only 4 out of the 10 samples of *P. solenopsis* analyzed.

Scarcity of S-symbionts is coherent with the very low levels of diversity typically found in mealybug microbial populations, despite differences among regions and times of mealybug collection (Franke-Whittle et al., 2004). Thus, no other bacteria (except for essential endosymbionts) were detected either in the Pseudococcinae *P. citri* or in the Phenacoccinae *P. peruvianus* when analyzed by FISH with 16S rRNA universal probes (von Dohlen et al., 2001; López-Madrigal et al., 2014). Even their external surface appears to be hardly habitable by environmental microbes. The acidic honeydew of the pink sugar cane mealybug *Saccharococcus sacchari* (Cockerell) is known to support an atypical epiphytic microbiota dominated by acetic acid bacteria (Alphaproteobacteria), including the sporadically present species *Gluconacetobacter diazotrophicus*, *Gluconacetobacter sacchari* and *Gluconacetobacter liquefaciens* (Ashbolt and Inkerman, 1990; Franke et al., 2000; Franke-Whittle et al., 2004). Acidophilic yeast species have been also isolated. The high carbon/nitrogen ratio of mealybug's exudate, the microbial production of ketogluconic acids and gamma-pyrones and the associated lowering of pH would enhance the selection against most microorganisms, including the mealybug parasitic fungus *Aspergillus parasiticus* (Ashbolt and Inkerman, 1990).

3. PUBLICATIONS

CHAPTER 1

Status: Published. October, 2011.

Citation: López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. J. Bacteriol. 193:5587–5588.

Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status

Sergio López-Madrigal,[1,2] Amparo Latorre,[1,2,3] Manuel Porcar,[4] Andrés Moya,[1,2,3] and Rosario Gil[1,2]*

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado Postal 22085, 46071 València, Spain[1]; Departament de Genètica, Universitat de València, Dr. Moliner, 50, 46100 Burjassot (València), Spain[2]; Área de Genómica y Salud, Centro Superior de Investigación en Salud Pública (CSISP), Avenida de Cataluña 21, 46020 Valencia, Spain[3]; and Fundació General de la Universitat de València. Apartado Postal 22085, 46071 València, Spain[4].

*Corresponding author. Mailing address: Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado Postal 22085, 46071 València, Spain. Phone: 34 963543824. Fax: 34963543670. E-mail: rosario.gil@uv.es

ABSTRACT

The sequence of the genome of "*Candidatus* Tremblaya princeps" strain PCVAL, the primary endosymbiont of the citrus mealybug *Planococcus citri*, has been determined. "*Ca*. Tremblaya princeps" presents an unusual nested endosymbiosis and harbors a gammaproteobacterial symbiont within its cytoplasm in all analyzed mealybugs. The genome sequence reveals that "*Ca*. Tremblaya princeps" cannot be considered an independent organism but that the consortium with its gammaproteobacterial symbiotic associate represents a new composite living being.

GENOME ANNOUNCEMENT

Obligate intracellular bacteria are characterized by their genome reduction due to the loss of genes rendered unnecessary in their protected environments. However, they are still able to fulfill the essential cellular functions that allow them to replicate and synthesize their own proteins (7). Mealybugs contain a primary (P) endosymbiont, "*Candidatus* Tremblaya princeps", which harbors a secondary (S) symbiont in its cytoplasm; this is the first described case of a nested bacterial symbiosis (10, 12).

We report the complete genome of "*Ca*. Tremblaya princeps" from the citrus mealybug *Planococcus citri* (Risso, 1813) isolated from cultivated plants in Valencia, Spain. Bacterium-enriched genomic DNA was extracted from adult females by visceral dissection. This genomic DNA was sequenced through a whole-genome shotgun strategy using Roche 454 GS-FLX Titanium pyrosequencing. A total of 19,693 quality reads were assembled and analyzed using the MIRA (http://sourceforge.net/projects/mira-assembler/) and Staden Package software programs

(8), providing 41.4X coverage. Screening for RNA genes was performed with ARGORN (4), tRNAscan (5), and Rfam (3), and prediction and annotation of protein-coding genes were performed through BASys (Bacterial Annotation System) (11) and BLAST searches (1). The complete genome is 138,931 kb in length and has a 58.8% G+C content. The genome contains only 110 functional protein-coding genes, 43 of which correspond to ribosomal proteins, and 23 pseudogenes. It contains two 16S-23S rRNA operons, a common feature of other mealybug P-endosymbionts (2). However, only 8 functional tRNA genes are present, no functional aminoacyltRNA synthetases have been identified, and the synthesis of some ribosomal proteins and the modification, maturation, and function of ribosomes must depend on the S-symbiont.

Additionally, genes coding for protein translocation machinery, synthesis of nucleotides and cofactors, energy production, transport, and cell envelope biogenesis are absolutely absent, while only part of the replication machinery is preserved. In contrast, "*Ca*. Tremblaya princeps" has retained the ability to synthesize most of the amino acids that are essential for its phloem-feeding insect host. Thus, the genome retains the genes for the synthesis of amino acids of the branched family (valine, leucine, and isoleucine), histidine, threonine, and phenylalanine, and the final step in the synthesis of methionine from homocysteine that, as in other endosymbionts previously analyzed, must be provided by the host (13).

This genome represents another example of the difficulties of maintaining the distinction between autonomous bacterial life and organelles (9). Considering life the emerging property of a system that simultaneously displays homeostasis, self-reproduction, and evolution (6), "*Ca*. Tremblaya princeps" cannot be considered a living organism, although it is not like any organelle previously described. However, the composite organism made up of "*Ca*. Tremblaya princeps" and its so-called S-symbiont (genome sequencing is in progress) might resemble a eukaryotic cell, with a compartimentalized living scheme. Similarly to mitochondria, "*Ca*. Tremblaya princeps" retains only some of the genes necessary for symbiotic functions (synthesis of essential amino acids) and vital functions but needs the recruitment of the gene products encoded by the S-symbiont to be exported to the cytoplasm of "*Ca*. Tremblaya princeps", where most translation must take place.

Nucleotide sequence accession number. The complete genome sequence of "*Ca*. Tremblaya princeps" PCVAL was deposited in GenBank under accession number CP002918.

ACKNOWLEDGMENTS

Financial support for this study was provided by grant BFU2009-12895-C02-01/BMC (Ministerio de Ciencia e Innovación, Spain) to A. Latorre and by grant Prometeo/2009/092 (Conselleria d'Educació, Generalitat Valenciana, Spain) to A. Moya. S. López-Madrigal is a recipient of a fellowship from the Ministerio de Educación (Spain).

REFERENCES

1. Altschul, S. F., et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

2. Baumann, L., M. L. Thao, J. M. Hess, M. W. Johnson, and P. Baumann. 2002. The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. Appl. Environ. Microbiol. 68:3198–3205.

3. Gardner, P. P., et al. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res. 37:D136–D140.

4. Laslett, D., and B. Canback. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11–16.

5. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

6. Luisi, P. L. 2002. Toward the engineering of minimal living cells. Anat. Rec. 268:208–214.

7. Moya, A., J. Pereto, R. Gil, and A. Latorre. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat. Rev. Genet. 9:218–229.

8. Staden, R., K. F. Beal, and J. K. Bonfield. 2000. The Staden package, 1998. Methods Mol. Biol. 132:115–130.

9. Tamames, J., et al. 2007. The frontier between cell and organelle: genome analysis of *Candidatus* Carsonella ruddii. BMC Evol. Biol. 7:181.

10. Thao, M. L., P. J. Gullan, and P. Baumann. 2002. Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl. Environ. Microbiol. 68:3190–3197.

11. Van Domselaar, G. H., et al. 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 33:W455–W459.

12. von Dohlen, C. D., S. Kohler, S. T. Alsop, and W. R. McManus. 2001. Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. Nature 412:433–436.

13. Wilson, A. C., et al. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. Insect Mol. Biol. 19 (Suppl 2):249–258.

CHAPTER 2

Status: Published. April 1, 2013.

Citation: López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2013. Mealybugs nested endosymbiosis: going into the “matryoshka” system in *Planococcus citri* in depth. BMC Microbiol. 13:74.

Mealybugs nested endosymbiosis: going into the 'matryoshka' system in *Planococcus citri* in depth

Sergio López-Madrigal[1], Amparo Latorre[1,2], Manuel Porcar[1,3], Andrés Moya[1,2] and Rosario Gil[1]*

1 Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado Postal 22085, València 46071, Spain. 2 Área de Genómica y Salud, Centro Superior de Investigación en Salud Pública (CSISP), Avenida de Cataluña 21, Valencia 46020, Spain. 3 Fundació General de la Universitat de València, Apartado Postal 22085, València 46071, Spain.

*Correspondence: rosario.gil@uv.es

ABSTRACT

Background: In all branches of life there are plenty of symbiotic associations. Insects are particularly well suited to establishing intracellular symbiosis with bacteria, providing them with metabolic capabilities they lack. Essential primary endosymbionts can coexist with facultative secondary symbionts which can, eventually, establish metabolic complementation with the primary endosymbiont, becoming a co-primary. Usually, both endosymbionts maintain their cellular identity. An exception is the endosymbiosis found in mealybugs of the subfamily Pseudoccinae, such as *Planococcus citri*, with *Moranella endobia* located inside *Tremblaya princeps*.

Results: We report the genome sequencing of *M. endobia* str. PCVAL and the comparative genomic analyses of the genomes of strains PCVAL and PCIT of both consortium partners. A comprehensive analysis of their functional capabilities and interactions reveals their functional coupling, with many cases of metabolic and informational complementation. Using comparative genomics, we confirm that both genomes have undergone a reductive evolution, although with some unusual genomic features as a consequence of coevolving in an exceptional compartimentalized organization.

Conclusions: *M. endobia* seems to be responsible for the biosynthesis of most cellular components and energy provision, and controls most informational processes for the consortium, while *T. princeps* appears to be a mere factory for amino acid synthesis, and translating proteins, using the precursors provided by *M. endobia*. In this scenario, we propose that both entities should be considered part of a composite organism whose compartmentalized scheme (somehow) resembles a eukaryotic cell.

Keywords: Nested endosymbiosis, *Planococcus citri*, *Moranella endobia*, *Tremblaya princeps*, functional complementation

BACKGROUND

Symbiosis is a widespread natural phenomenon that has been postulated as one of the main sources of evolutionary innovation [1,2], and it is an example of compositional evolution involving the combination of systems of independent genetic material [3]. Many insects have established mutualistic symbiotic relationships, particularly with intracellular bacteria that inhabit specialized cells of the animal host (bacteriocytes). In most insect-bacteria endosymbioses described to date, host insects have unbalanced diets, poor in essential nutrients that are supplemented by their endosymbionts. Attending to their dispensability for host survival, we distinguish between primary (P) or obligate, and secondary (S) or facultative endosymbionts. P-endosymbionts are essential for host fitness and reproduction, and maternally transmitted through generations, while S-symbionts are not essential and can experience horizontal transfer. The genomes of P-endosymbionts usually exhibit an increase in their A+T content and undergo great size reduction, among other changes. The main evolutionary forces accounting for these features are relaxation of purifying selection on genes rendered unnecessary in the enriched intracellular environment, and random genetic drift due to a strong population bottlenecking throughout intergenerational transmission of the bacteria [4]. P and S symbionts can coexist in the same host. When an S-symbiont is also present, the irreversible genomic degenerative process could lead to the loss of some P-endosymbiont metabolic capabilities needed by the host. In this situation, two outcomes are possible: the host insect can recruit those functions from the S-symbiont, which then becomes a co-primary endosymbiont, establishing metabolic complementation with the former P-endosymbiont to fulfill the host needs or [5-8]; alternatively, the S-symbiont may replace its neighbor [9].

Mealybugs (Hemiptera: Sternorrhyncha: Pseudoccidae) form one of the largest families of scale insects, including many agricultural pest species that cause direct crops damage or vector plant diseases while feeding on sap [10]. All mealybug species analyzed so far possess P-endosymbionts. Two subfamilies have been identified, Phenacoccinae and Pseudococcinae [11], the latter having been studied in greater depth, all of which live in symbiosis with the β-proteobacterium "*Candidatus* Tremblaya princeps" (*T. princeps* from now on, for the sake of simplicity). Universal presence, along with the cocladogenesis of endosymbionts and host insects, led to *T. princeps* being considered the mealybug P-endosymbiont [12]. However, recently, other P-endosymbionts from the β-proteobacteria and Bacteroidetes groups have been identified in the subfamily Phenacoccinae [13]. Most genera of the subfamily Pseudococcinae also harbor additional γ-proteobacteria endosymbionts that, due to their discontinuous presence and polyphyletic origin, have been considered as S-symbionts [14]. An unprecedented structural organization of the endosymbionts of the citrus mealybug *Planococcus citri* was revealed by von Dohlen and coworkers [15]: each *T. princeps* cell harbors several S-endosymbiont cells, being the first known case of prokaryote-prokaryote endocelullar symbiosis. The S-endosymbiont has recently been named "*Candidatus* Moranella endobia" (*M. endobia* from now on) [16]. The dynamics of both endosymbiont populations throughout the insect life-cycle and their differential

behavior depending on host sex [17] suggest that both play an important role in their hosts' nutritional and reproductive physiology, putting into question the secondary role *of M. endobia*.

The sequencing of two fragments of the genome of *T. princeps* from the pineapple mealybug, *Dysmicoccus brevipes* [18], showed a set of unexpected genomic features compared with that found in most P-endosymbiont reduced genomes. This species presents a rather high genomic G+C content – a rare condition among P-endosymbionts with the only known exception being "*Candidatus* Hodgkinia cicadicola" (P-endosymbiont of the cicada *Diceroprocta semicincta* [7]) – , a partial genomic duplication including the ribosomal operon and neighbor genes, and low gene density. All other sequenced genomes from endosymbionts having a long relationship with their host maintain a single set of rRNA genes, therefore these data suggested an unprecedented complexity for this P-endosymbiont genome, an unexpected finding for a long co-evolutionary process, as already elucidated for this symbiotic system [18]. However, the recent sequencing of two strains of *T. princeps* from *P. citri* (PCIT and PCVAL) has shown that it is, in fact, the smallest (139 kb) and most simplified bacterial genome described to date [16,19]. Functional analysis reveals that the genetic repertoire of *T. princeps* is unable to sustain cellular life, according to Gil et al. (2004) [20], and that it entirely depends on *M. endobia* for many essential functions. Even though most of its genome is occupied by ribosomal genes and genes involved in the biosynthesis of essential amino acids, *T. princeps* likely depends on its symbiotic consortium partner to build its own ribosomes and for amino acid production [16,19].

The work published by McCutcheon and von Dohlen [16] mainly focused on the analysis of the *T. princeps* genome and detangling the amino acid biosynthetic pathways in which all three partners (*T. princeps*, *M. endobia* and the host) appear to be involved. However, the characteristics and functionality of the *M. endobia* genome, as well as other possible modes of complementation between the two endosymbionts, have remained largely unexplored. In this work we present a comprehensive analysis of the predicted consortium functional capabilities and interactions, thus offering new insights into how this bacterial consortium may function internally. Additionally, we have performed a comparative analysis of both endosymbiont genomes in two *P. citri* strains, PCIT [16] and PCVAL ([19] and this work). Our analysis suggests that both genomes have undergone reductive evolution, albeit with some unusual genomic features, probably as a consequence of their unprecedented compartmentalized organization.

RESULTS AND DISCUSSION

Main features and genomic variability between two strains of *P. citri* nested endosymbionts

The main molecular features of the genomes of *T. princeps* str. PCVAL [19] and PCIT [16], and *M. endobia* str. PCVAL (this work) and PCIT [16] are summarized in Table 1. It is worth mentioning that differences in CDS numbers and coding density between both strains are due to differences in the annotation criteria used, since the number of polymorphisms detected between

the two sequenced strains of *T. princeps* and *M. endobia* is minimal (see Table S1 for a list of annotation differences in CDS and tRNA genes).

Table 1. Main genomic features of the two strains of the *P. citri* endosymbiotic consortium already sequenced.

	***T. princeps* PCVAL**	***T. princeps* PCIT**	***M. endobia* PCVAL**	***M. endobia* PCIT**
GenBank accession number	CP002918	CP002244	CP003881	CP002243
Genome size (bp)	138931	138927	538203	538294
Total gene number	130	136	458	452
CDSs	116	121	411	406
rRNAs	6	6	5	5
tRNAs	7	8	41	41
Small RNA genes	1	1	1	0
Pseudogenes	19(CDS) 6(tRNA)	19(CDS) 4(tRNA)	25(CDS)	23(CDS)
Overall gene density (%)	71.2	72.9	79.3	79.0
Average ORF length (bp)	775	760	1012	1022
Average IGRs (bp)	466.8	389.0	260.3	268.0
G+C content (%)	58.8	58.8	43.5	43.5
genes	58.6	58.5	45.5	45.4
pseudogenes	58.8	59.9	43.6	44.7
IGR	59.4	59.5	36.0	36.2

Both consortium partners lack a canonical *ori*C, which is consistent with the absence of *dna*A, similarly to many other reduced endosymbiont genomes already sequenced (e.g., *Blochmannia floridanus* [21], *Wigglesworthia glossinidia* [22], *Carsonella rudii* [23], *Hodgkinia cidadicola* [24], *Zinderia insecticola* [8], and *Sulcia muelleri* [25]). This has been considered an indication that the endosymbionts rely on their host for the control of their own replication [21]. Another shared genomic characteristic of both endosymbionts is their low gene density (already noticed in [16] for *T. princeps*) and the large average length of the intergenic regions, in which no traces of homology with coding regions of other bacteria can be found. Although these traits are unusual in bacterial endosymbionts, they have also been described for *Serratia symbiotica* SCc, the co-primary endosymbiont of *Buchnera aphidicola* in the aphid *Cinara cedri* [5]. This non-coding DNA is probably the remnant of ancient pseudogenes that are gradually being eroded [26].

Another remarkable feature, compared with other endosymbiotic systems, is that both *T. princeps* and *M. endobia* display one partial genomic duplication event involving the ribosomal operon (Figure 1). The duplication in *T. princeps* has been described in other mealybugs [18], and it affects the rRNA genes (*rrs*A, *rrl*A and *rrf*A) plus *rps*O (encoding ribosomal protein S15). Ribosomal genes and *loci* from its closest genomic context (*acp*S and partial *pdx*J) are also duplicated in *M. endobia* but, unlike in *T. princeps*, the two copies of the *M. endobia* ribosomal operon have not remained intact. Comparative synteny among several γ-proteobacteria species suggests that the additional copy was inserted in the lagging strand, while the original copy

suffered the losses. Thus, although 4 kb of the duplicated region (positions 109,083-113,105 and 343,701-347,723 for the copies in the direct and lagging strand, respectively) seem to be under concerted evolution (both regions are identical in both genomes), the original copies of *rrs*A, *trn*I and *trn*A have been lost.

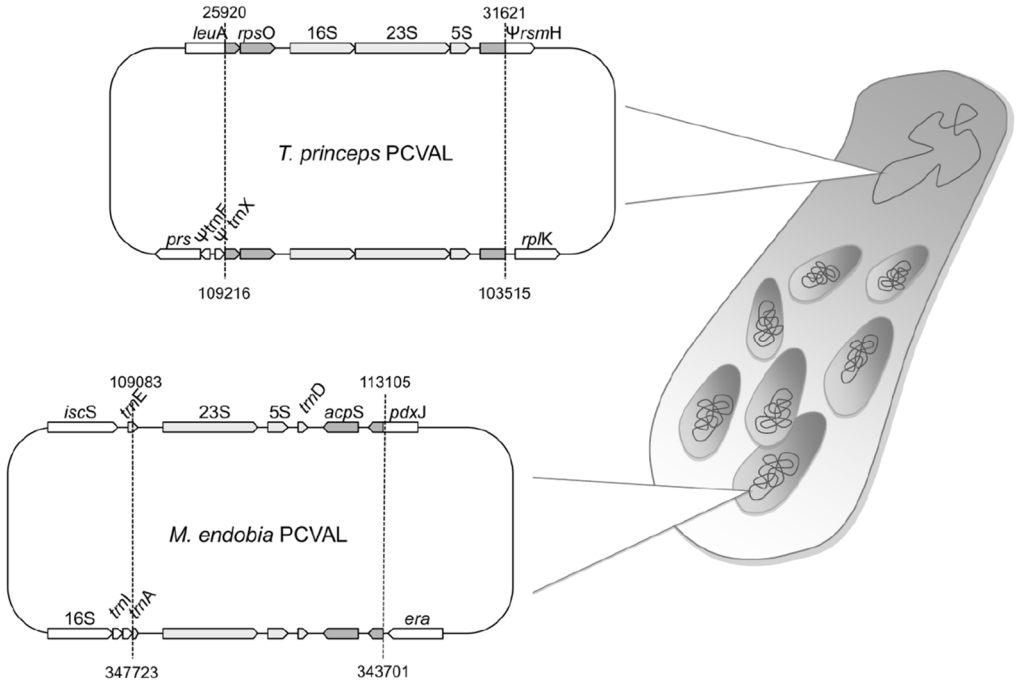


Figure 1. Endosymbionts partial genome duplications. Duplicated regions evolving under concerted evolution in *T. princeps* and *M. endobia* are represented. Only affected genes (grey arrows: coding genes; light grey arrows: RNA genes) and their closest neighbors (white arrows) are depicted. Numbers indicate the location of these duplicated regions in the corresponding genomes.

The reductive process affecting both genomes has led to the loss of most regulatory functions. Thus, they lack most regulatory genes and some genes have lost regulatory domains. This is the case of *met*L and *adk* from *T. princeps*, which have lost the regulatory 'ATC' domain, or the loss of the 'HTH' domain of *bir*A, the 'PNPase C' domain of *rne* and the 'DEAD box A' of *dead* in the case of *M. endobia*. Additionally, many other genes have been shortened due to frameshifts or the presence of premature stop codons, in comparison with their orthologs in free-living relatives (e.g. *ssp*B, *rpl*Q, *rpl*O and *aro*C in *T. princeps*; *thi*C, *ybg*I, *yac*G, *ygb*Q, *fts*L, *fts*Y and *til*S in *M. endobia*). In some cases, the shortening removes some non-essential protein domains completely (e.g., *eng*A, *rpo*A and *rpo*D in *T. princeps*; *sec*A, *ace*F, *yeb*A and *met*G in *M. endobia*). The loss of the 'anticodon binding domain of tRNA' and 'putative tRNA binding domain' of *met*G, encoding methionyl-tRNA synthetase is common to other endosymbionts with reduced genomes.

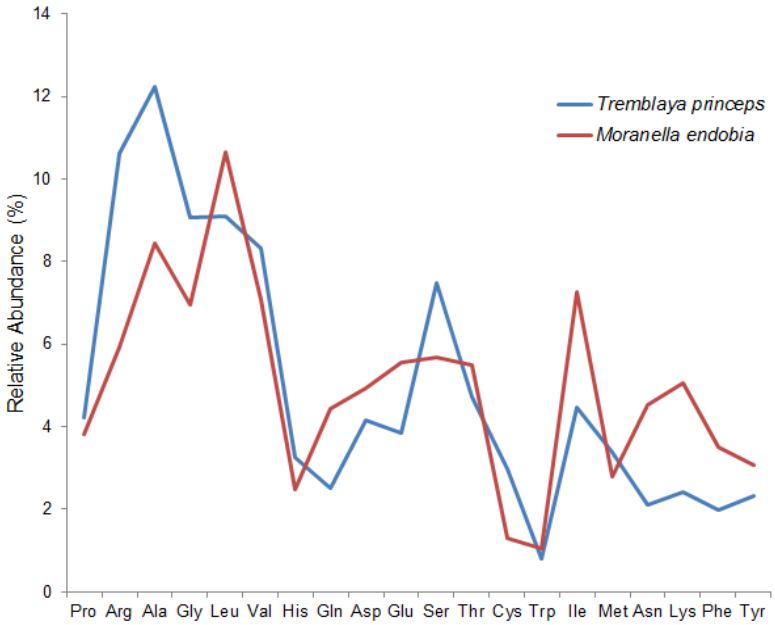


Figure 2. Amino acid content profiles for *T. princeps* and *M. endobia* proteomes. Amino acids are ranked from left to right according to the GC-richness of the corresponding codons (see Additional data file 2).

Finally, even though both genomes have an unusually high G+C content compared with most bacterial endosymbionts, at least *M. endobia* seems to be suffering the AT mutational bias typical of bacterial genomes [27,28]. This conclusion is drawn from the analysis of the nucleotide composition of genes, pseudogenes and IGRs (Table 1), as well as the preferential use of AT-rich codons (Table S2) including a high incidence of the TAA stop codon (56.44%). Since both genomes seem to rely on the DNA replication and repair machinery of *M. endobia* (see next section), both genomes could be expected to undergo a similar trend towards an increase in AT content. However, this trend is undetectable in *T. princeps*, where the G+C content of pseudogenes and IGRs do not differ from that of the genes (Table 1). The differences in G+C content between both genomes could be due to a higher ancestral G+C content plus a slower evolutionary rate for *T. princeps*, due to its extreme genome reduction, and the biology of the system (i.e., a lower replication rate, since each *T. princeps* cell retains several M. endobia cells). In fact, the codon usage bias (Table S2) and differences in the amino acidic composition between both endosymbiont proteomes (Figure 2) reflect their differences in G+C content. Thus, *T. princeps* proteins are rich in amino acids encoded by GC-rich codons (Arg, Ala, Gly, Leu, Val and Ser represent 56.82% of the total, whereas Phe and Tyr are scarce), while *M. endobia* has a weaker amino acid composition bias (Table S2).

T. princeps genome comparison

The genome alignment of both *T. princeps* strains showed a high degree of identity at the sequence level (99.98%, being 138,903 bp identical), which is coherent with their evolutionary proximity and extreme genome reduction. Although we also detected the 7,032-bp region flanked

by 71-bp inverted repeats described in the strain PCIT [16], we only found one orientation in the population used for genome sequencing.

The genome of strain PCVAL only differs in 4 nucleotides in length from strain PCIT [16], involving five short indel events of one (4 cases) or two nucleotides (1 case). Additionally, 23 nucleotide substitutions were detected. Transitions represent 43.5% (10/23) of the total substitutions. Although the number of mutations is too small to be representative and, therefore, it is difficult to draw clear conclusions, it is noteworthy that all indels plus 87% of the detected substitutions between both strains are located in the coding fraction of the genome, in spite of its low coding density. One of the detected indels affects the start codon of *aro*C, involved in the biosynthetic pathway of aromatic amino acids, which is then changed to a GTG start codon. Two other short indels yield the loss (AT) and recovery (T) of the open reading frame of *ilv*D, needed for the synthesis of isoleucine and valine. The non-inactivating character of these mutations on genes involved in biosynthetic pathways of essential amino acids without an ortholog in the genome of *M. endobia*, corroborates their importance for the bacterial partnership. The other two indels, as well as 20 out of 23 of the observed substitutions, were located at the 3' end of *rpl*Q, which suggests that this region could be a mutational hot-spot. To confirm this point, we analyzed the original *P. citri* DNA samples used in the genome sequencing experiments by PCR amplification of the *rpl*Q and flanking ITS regions, as well as new DNA samples obtained from individual insects cultivated in Almassora (Spain) and from environmental colonies collected in Murcia (Spain). Although all three samples were obtained from different plant hosts and separated by more than 300 Km, they were identical. Since we have no direct availability of the PCIT strain, it is feasible that the Spanish and American populations differ.

M. endobia genomes comparison

The alignment of both genomes of *M. endobia* showed that the genome of strain PCVAL is 91 nucleotides shorter than that of PCIT, and allowed the identification of 262 substitutions. Among them, 90.1% were G/C↔A/T changes, with only 18 A↔T changes and 8 G↔C changes, which is additional indirect evidence of the mutational bias towards A/T already observed in the codon usage analysis (Table S2). As expected for a neutral process, the mutational bias affected both strains equally, being the changes G/C↔A/T evenly distributed (50.4% A/T in strain PCIT and 49.5% in PCVAL). Regarding the genome distribution of the polymorphisms, 47% of them (123) map onto IGRs, and 4.5% (12) onto 10 pseudogenes. The 139 substitutions detected in the coding fraction affect only 111 out of the 406 orthologous genes. Among these substitutions, 77 are synonymous ($dS = 0.0011 \pm 0.0001$), and 62 nonsynonymous ($dN = 0.0005 \pm 0.0000$), with a $\omega = 0.44$, suggesting the action of purifying selection. It is worth noticing that about 75% of them affect functional domains, suggesting that many putatively functional genes accumulate mutations, which also justifies the maintenance of a minimal set of molecular chaperones to help in the proper folding of the encoded proteins.

Additionally, 60 indels were detected between both *M. endobia* strains, with a mean size of 5.4 nucleotides, although there is a great variance, between 1 and 75 nucleotides. Results

showed 58.3% (35/60) of the indels affect homopolymers of A (18/35), T (11/35) and, less frequently, G (3/35) and C (3/35), which is consistent with the higher proportion of A and T homopolymers. This fact may be related with the above-mentioned A/T mutational bias. Although artifacts due to sequencing errors cannot be ruled out, given that PCVAL genomes were assembled based on 454 sequencing data, there are several pieces of evidence that indicate that the observed indels may be real. First, although homopolymers can be found both in coding and non-coding regions, most indels affect the non-coding parts of the genome. Second, even when A/T homopolymers are quite abundant in the *M. endobia* genome (844 cases equal to or bigger than 6 nucleotides), only a small fraction of them are affected by indels (29 cases, representing 3.4%). Finally, the coverage of the affected regions was always higher than 27X, and the PCVAL reads polymorphism was almost null. Most of the remaining indels affect microsatellites of 2 to 8 nucleotides with a small number of copies. Forty-seven indels (78.3%) map onto intergenic regions, pseudogenes (2 in Ψ*pdx*B, 1 in Ψ*prf*C) or the nonfunctional part of shortened genes (*dna*X), and only 13 indels (21.7%) map onto coding regions. Most of these are located on the 3' end of the affected gene, causing enlargement or shortening of the ORFs compared with the orthologous gene in other γ-proteobacteria. Thus, *gly*Q (involved in translation) and *pts*I (participating in the incorporation of sugars to the intermediary metabolism) are enlarged in strain PCVAL, while *rpp*H (involved in RNA catabolism) is shortened in this strain without affecting described functional domains. Conversely, the shortening of *fis* (encoding a bacterial regulatory protein) in PCVAL, and of *yic*C (unknown function) and *pan*C (involved in the metabolism of cofactors and vitamins, a function that is incomplete in *M. endobia*) in PCIT, affect some functional domains, although their activity might not be compromised. Finally, amino acid losses without frameshift were observed in PCVAL (relative to PCIT) for the *loci hol*C (encoding subunit chi of DNA polymerase III), *rlu*B (involved in ribosome maturation), *sur*A (encoding a chaperone involved in proper folding of external membrane proteins), and *pit*A (encoding an inorganic phosphate transporter). None of the corresponding functional domains were affected in the first two cases, while the indel polymorphisms mapped inside the 'PPIC-type PPIASE' domain in *sur*A, which appears to be dispensable for the chaperone qualities of the protein [29]. Therefore, it seems that most (if not all) changes that could affect the functions of the encoded proteins have been removed by the action of purifying selection.

Functional analysis of the nested consortium

Most endosymbiotic systems analyzed to date at the genomic level have a nutritional basis, and many of them involve the biosynthesis of essential amino acids that are in short supply in the host diet. The metabolic pathways leading to amino acid biosynthesis in the *T. princeps-M. endobia* consortium found in *P. citri* were recently analyzed in detail by McCutcheon and von Dohlen [16] and, therefore, they will not be dealt with in this study. These authors also stated that *T. princeps* is unable to perform DNA replication, recombination or repair by itself, and the same applies to translation. They speculate that a passive mechanism such as cell lysis could provide *T. princeps* with the needed gene products from *M. endobia*. Our present work provides a detailed analysis of the *M. endobia* functional capabilities, based on a functional analysis of its genome,

regarding informational functions or other intermediate metabolism pathways beyond amino acids biosynthesis. In the following sections these functional capabilities will be analyzed in a comprehensive manner, considering both endosymbiotic partners, in order to identify putative additional levels of complementation between them.

DNA repair and recombination

Contrary to what is found in bacterial endosymbionts with similarly reduced genomes, *M. endobia* has quite a complete set of genes for DNA repair and recombination, while none were annotated in the *T. princeps* genome [16,19]. Although it has lost the nucleotide excision repair genes (only *uvr*D is present), *M. endobia* retains a base excision repair system (the DNA glycosylases encoded by *mut*M and *ung* plus *xth*, the gene encoding exonuclease III, involved in the repair of sites where damaged bases have been removed). The mismatch repair system is also almost complete, since only *mut*H, encoding the endonuclease needed in this process to cleave the unmethylated strand, has been lost. Additionally, *M. endobia* also retains almost the entire molecular machinery for homologous recombination (*rec*ABCGJ, *ruv*ABC, *pri*AB), which could be responsible for the concerted evolution of the duplications in both genomes. In the absence of *rec*D, the RecBC enzyme can still promote recombination, since it retains helicase and RecA loading activity. The missing exonuclease V activity can be replaced by other exonucleases with ssDNA degradation activity in the 5'→3' sense, such of RecJ [30], which has been preserved. The final step in homologous recombination requires the reloading of origin-independent replication machinery. Two replisome reloading systems have been described in *E. coli*, one of which requires the participation of PriA, PriB and DnaT [31], and it appears that helicase DnaB loading and unwinding of a replication fork is dependent upon the activities of DnaT and DnaC, among other restart proteins. These last two proteins are the only two elements of the replisome that are not encoded in the *M. endobia* genome. However, mutations in *dna*C which have the ability to bypass such requirements in the loading of DnaB have been described [32], and *dna*C is also absent in other reduced genomes that have been characterized (e.g. *Blochmannia floridanus* [21], *Wigglesworthia glossinidia* [22] or *Mycoplasma genitalium* [33]). Additionally, the role of DnaT in primosome assembly has not been fully elucidated [34]. Therefore, it cannot be ruled out that *dna*T is not essential for the functioning of the homologous recombination system in this bacterial consortium.

RNA metabolism

Even though most genes present in the *T. princeps* genome are involved in RNA metabolism (78 out of 116 genes, occupying 35% of its genome length and 49% of its coding capacity) [16,19], it still seems to depend on *M. endobia* for transcription and translation. Thus, *T. princeps* encodes every essential subunit of the core RNA polymerase (*rpo*BCA) and a single sigma factor (*rpo*D), but no other genes involved in the basic transcription machinery or in RNA processing and degradation are present in its genome. On the other hand, *M. endobia* possesses a minimal but yet complete transcription machinery [35] plus some additional genes, including the ones encoding the ω subunit of the RNA polymerase (*rpo*Z), the sigma-32 factor (*rpo*H), and the

transcription factor Rho. It also presents several genes involved in the processing and degradation of functional RNAs, i.e. *pnp*, *rnc* (processing of rRNA and regulatory antisense RNAs), *hfq* (RNA chaperone), *rne*, *orn*, *rnr* (rRNA maturation and mRNA regulation in stationary phase), and *rpp*H (mRNA degradation). It is surprising that the small genome of *M. endobia* has also retained several transcriptional regulators, the functions of which are not yet fully understood, and which are absent in other endosymbionts with reduced genomes. These include CspB and CspC (predicted DNA-binding transcriptional regulators under stress conditions), and NusB, which is required in *E. coli* for proper transcription of rRNA genes, avoiding premature termination [36]. *cpx*R, encoding the cytoplasmic response regulator of the two-component signal transduction system Cpx, the stress response system that mediates adaptation to envelope protein misfolding [37], is also preserved, while the companion sensor kinase *cpx*A appears to be a pseudogene. This might be an indication of a constitutive activation of the regulatory protein.

Regarding translation, an extremely complex case of putative complementation between both bacteria is predicted, which would represent the first case ever described for this function. Thus, only *M. endobia* presents the genes *fmt* and *def*, responsible for the synthesis of formil-methionil-tRNA and methionine deformilation, respectively, and a minimal set of genes for tRNA maturation and modification [35], as well as a complete set of aminoacyl-tRNA synthetases. Additionally, it codes for more than 80% of the tRNA genes annotated in both genomes and, therefore, is supposed to be the source of these tRNAs for the whole consortium. Comparative analysis with other endosymbiotic or free-living bacteria reveals a significant overload of tRNA genes in *M. endobia* in relation with its translational requirements (Figure 3). It should be noted that *M. endobia* has multiple tRNAs *loci* for codons that are more frequently represented in *T. princeps* than in itself (Table S2 and TableS3), due to their different G+C content. On the other hand, *T. princeps* has only retained tRNA genes with the anticodon complementary to its most frequently used codons for alanine (GCA) and lysine (AAG). Surprisingly, it has two copies (plus a pseudogene) of the last one, a quite unusual situation for such a reduced genome, while this tRNA is missing in the *M. endobia* genome. This fact might be an indication that *T. princeps* is providing this tRNA to its nested endosymbiont, whose absolute requirements for this tRNA are considerably larger (2032 codons).

Finally, as it was already stated, ribosomes are the best preserved molecular machinery in *T. princeps* [16,19]. In addition to two copies of the ribosomal 23S-16S operon, it encodes 49 out of 56 ribosomal proteins needed to make a complete ribosome. On the other hand, *M. endobia* has also retained a full set of ribosomal proteins and also presents two copies of the 23S and 5S rRNA genes. The high redundancy of rRNA and ribosomal protein genes might indicate that ribosomes from both members of the consortium are not exchangeable, or that redundancy is needed to achieve proper levels of ribosomal components for cell functioning. Both genomes encode the tmRNA, a molecule needed to solve problems that arise during translation while only *M. endobia* encodes ribosome maturation proteins and translational factors.

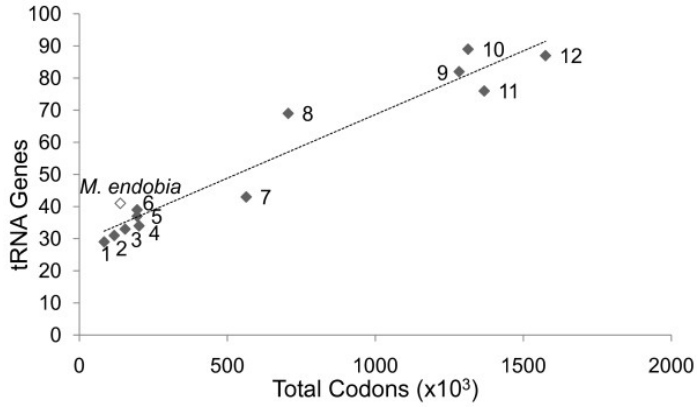


Figure 3. Correlation between tRNA genes content and translational requirements. Selected genomes with variable translational requirements are taken into account: *Sulcia muelleri* CARI (1), *Buchnera aphidicola* BCc (2), *Moranella endobia* PCVAL (white), *Riesia pediculicola* (3), *Blatabacterium* sp. Bge (4), *Blochmania floridanus* (5), *Baumania cicadicolla* (6), *Hamiltonella defensa* (7), *Sodalis glossinidius* (8), *Yersinia enterocolitica* subsp. Enterocolitica 8081 (9), *Escherichia coli* str. K-12 MG1655 (10), *Dickeya dadantii* Ech586 (11), and *Serratia* sp. AS9 (12). A high correlation between both parameters was observed when every genome except *M. endobia* were included (R^2 = 0.94), as well as when only endosymbionts except *M. endobia* were considered (R^2 = 0.77). Inclusion of *M. endobia* among endosymbionts caused a drastic diminution of the coefficient (R^2 = 0.33).

Protein processing, folding and secretion

As compared with their orthologs in free-living relatives, both endosymbionts have retained at least a minimal set of chaperones [35] required for the proper folding of functional proteins in both members of the consortium. This is consistent with the presence of proteins accumulating non-synonymous substitutions. Some proteins can also be exported across the inner and outer membranes via typical gram-negative secretion systems (reviewed in [38]) encoded exclusively in the *M. endobia* genome. As other endosymbionts with similarly reduced genomes, *M. endobia* has retained a fully functional Sec translocation complex [16]. It also encodes Ffh, which together with 4.5S RNA forms the signal recognition particle (SRP), needed to bind the signal sequence of the proteins targeted for secretion through this system and to drive them to FtsY, the SRP receptor. Although in other endosymbionts there is an alternative system to assist proteins in their secretion, in which the proteins are recognized by the SecB chaperone after translation, this system cannot be functional in this consortium, because *sec*B appears to be a pseudogene [16].

Intermediate metabolism

T. princeps has almost null metabolic capacities, except for the production of essential amino acids, as described elsewhere [16]. Only *M. endobia* encodes a phosphotransferase system (PTS) for the uptake of hexose as carbon source, and it is predicted to perform glycolysis, transform pyruvate into acetate, and use it to feed the pathway for fatty acids biosynthesis, similarly to that described for *B. aphidicola* BCc, with highly reduced metabolic capabilities [39]. However, the pentose phosphate pathway appears to be incomplete, since only *zwf*, *pgl* and *tkt*

have been preserved, while *tal*A appears to be a pseudogene. Interestingly, *T. princeps* has retained a transaldolase TalB, which along with transketolase (Tkt) creates a reversible link between the pentose phosphate pathway and glycolysis, revealing another possible case of metabolic complementation between both bacteria.

Regarding the tricarboxylic acid (TCA) cycle, only *mdh* (encoding malate dehydrogenase) has been preserved in *T. princeps*, while *M. endobia* has retained only the genes that encode succinyl-CoA synthetase. This is the only step that has been maintained in *S. symbiotica* SCc [5], where the authors indicate that it must have been retained because it is necessary for lysine biosynthesis. Nevertheless, this cannot be the case in this consortium, since lysine biosynthesis cannot be accomplished.

As in other endosymbionts, NAD^+ can be regenerated by the action of the NADH-quinone oxidoreductase encoded by the *nuo* operon. But, in the absence of ATP synthase coupled to the electron transport chain, the whole consortium relies on substrate-level phosphorylation as a source of ATP. Acetyl-CoA can also be a source of ATP thanks to the presence of the genes *ack*A and *pta*.

The consortium also shares with other endosymbiotic bacteria with reduced genomes the incapability to synthesize nucleotides *de novo*. *T. princeps* has completely lost all genes involved in this function, while *M. endobia* retains a metabolic capacity similar to *B. aphidicola* BCc [39]. All triphosphate nucleotides could be obtained by phosphorylation from diphosphate nucleotides via pyruvate kinase A (*pyk*A), while deoxynucleotides could be obtained via ribonucleoside diphosphate reductase 1 (whose subunits are encoded by *nrd*A and *nrd*B). The only preserved diphosphate kinase is adenylate kinase (*adk*), while cytidylate kinase appears to be a pseudogene. Although it has been described that at least one purine and one pyrimidine kinase are needed to phosphorylate all dinucleotides, the fact that Adk is the same kinase that has been preserved in *B. aphidicola* BCc might be an indication that, in endosymbiotic bacteria, this enzyme can act on both nucleotide types. The presence of *dut* guarantees that the thymidylate nucleotides can also be synthesized using dUTP as a primary source.

The endosymbiotic system has almost completely lost the ability to synthesize vitamins and cofactors. Yet, the importance of the [Fe-S] clusters in this consortium is revealed by the presence of complete machinery for the assembly of such components, a complex system that is not fully preserved in other reduced genomes of endosymbiotic bacteria. The [Fe-S] clusters are one of the most ubiquitous and functionally versatile prosthetic groups in nature [40]. Although it is known that these clusters can spontaneously be assembled from the required components under the proper conditions, it is not an efficient procedure *in vivo* [41]. In *E. coli*, their assembly requires a complex machinery and it is achieved by two sets of proteins, the Suf (*suf*ABCDSE) and the Isc (*isc*SUA) systems. Both members of the consortium are involved in the maintenance of this machinery, revealing another possible case of metabolic complementation. The complete suf operon is present in the genome of *M. endobia*. Regarding the Isc system, both partners of the consortium retain *isc*S, and *T. princeps* also encodes *isc*U, while they both lack *isc*A. However,

IscA belongs to the HesB family of proteins, and a *hes*B gene has been identified in *T. princeps*. Additionally, ErpA, an A-type iron-sulfur protein that can bind both [2Fe-2S] and [4Fe-4S] clusters, is present in *M. endobia*.

The cell envelope structure is usually highly simplified in Gram-negative endosymbiotic bacteria, which lack most (if not all) of the genes needed for the biosynthesis of murein and lipopolysaccharides, and these two bacteria are not an exception. In fact, *T. princeps* has lost all the genes involved in these functions, and *M. endobia* has also lost many pathways, although it still retains some peptidoglycan synthetases and hydrolases needed for septum formation during cell division. It is noteworthy that this is the first analyzed case of an endosymbiont with a highly reduced genome that retains the ability to synthesize lipid IV_A, the biosynthetic precursor of lipopolysaccharydes.

Cellular transport

Only *M. endobia* has preserved genes related to celular transport, which must ensure proper exchange of metabolites with the host cell and between both endosymbionts. Many nutrients pass the outer membrane of Gram-negative bacteria via a family of integral outer membrane proteins (OMPs). The only OMP encoded in the consortium genomes is OmpF, the protein that forms osmotically regulated pores for the passage of small solutes such as sugars, ions and amino acids, with a preference for cationic molecules. Its proper functioning might be essential for the system, since *bam*A (*yae*T) and *bam*D (*yfi*O), coding for the essential components of the assembly machinery of beta-barrel OMPs, as well as *bam*B (*yfg*L), the gene encoding an additional lipoprotein of the system, have been preserved [42]. Additionally, it also retained the two chaperones Skp and SurA, which prevent folding and aggregation of OMPs in the periplasm during passage through the Sec translocon, and assist in their folding once they reach the assembly machinery in the outer membrane, respectively. Although DegP, the protease and chaperone identified to be involved in the degradation of misfolded OMPs, is not present, *M. endobia* encodes DegQ, another periplasmic protease which exhibits functional overlap with its homolog DegP [43,44].

Only a limited set of active transporters are encoded in the *M. endobia* genome. Those include a phosphotransferase system for the transport of hexoses, ABC transporters for zinc, glutathione, lipopolysaccharides and lipid A, as well as a low-affinity inorganic phosphate transporter. Additionally, the *M. endobia* genome also codes for two channels associated with osmotic stress response, MscL and YbaL, which are absent in all Sternorrhyncha endosymbiont genomes sequenced so far. It is worth mentioning that, in addition to low molecular weight molecules, such as ions, metabolites and osmoprotectants, MscL is reported to be involved in the excretion of some small cytoplasmic proteins [45-47]. Therefore, it cannot be ruled out that the preservation of this mechanosensitive channel is an essential part of this peculiar endosymbiont nested system. MscL might be involved in the exchange of molecules between the two bacteria.

CONCLUSIONS

The detailed analysis of the functional capabilities of the two components of the nested endosymbiosis in *P. citri* suggests the existence of an intricate case of complementation, involving not only metabolic but also informational functions. Thus, despite the fact that *M. endobia* resembles *B. aphidicola* BCc [39], another endosymbiont with a highly reduced genome, in many functions such as transport, biosynthesis of cellular envelope and nucleotides, and its incapability to synthesize ATP coupled to the electron transport chain, it possesses particular characteristics that might be related to its coevolution with *T. princeps*. While complementation for amino acid biosynthesis has been described in other endosymbiotic systems, this is the first case in which all energy sources appear to be provided only by one of the partners, similarly to what happens in the eukaryotic cell, where the mitochondria is in charge of this function. Additionally, two genes encoding channels associated with osmotic stress response (*msc*L and *yba*L) have been preserved in its genome. The fact that this kind of molecule has not been identified in other P-endosymbionts with reduced genomes might indicate their connection with special requirements of nested endosymbiosis, and might be involved in the exchange of molecules between both partners.

On the other hand, *T. princeps* does not resemble any known organelle, but it would not be reasonable to consider it, in a strict sense, as a living organism, since it has lost many essential genes involved in informational functions, as well as most metabolic pathways except for the ability to synthesize most essential amino acids, some of which require the cooperation of *M. endobia* and the host [16]. *T. princeps* retains most, but not all, of the translation machinery, for which it also seems to depend on *M. endobia*, even though almost half of its coding capacity is devoted to this function [16,19]. Additionally, it is unable to replicate on its own, although one can hypothesize that composite DNA and RNA polymerases (made of subunits encoded in both genomes) perform this function. *T. princeps* appears to be completely dependent on *M. endobia* for the synthesis of ATP, nucleotides or its cellular envelope, but still retains a complete set of molecular chaperones and proteins needed for the synthesis of [Fe-S] clusters.

Another intriguing fact revealed by our analysis is the overrepresentation of tRNAs genes in the *M. endobia* genome. This fact, together with the duplication in the rRNA operon in both genomes, appears to indicate an important translational activity in which both endosymbionts seem to be engaged. However, it lacks tRNA-Lys (AAG) which, surprisingly, has two functional copies in the small genome of *T. princeps*. This might be an indication that there is a mutual exchange of molecules between both compartments, although further studies are required to demonstrate this.

Nature is prolific in instances of symbiotic cooperation to give rise to new organisms, and new discoveries are always possible. Taking into consideration the deduced exceptional complementation inferred for this endosymbiotic system, we propose that *T. princeps* and *M.*

endobia should be considered part of a new composite organism rather than a bacterial consortium.

METHODS

Insect sample collection and DNA extraction

A population of *P. citri* from an initial sample obtained from a Cactaceae at the Botanical Garden of the Universitat de València (Valencia, Spain) was reared in the laboratory at room temperature, fed on fresh pumpkins and used for genome sequencing. Two other populations of *P. citri* were used for additional experiments. One of them was obtained from a melon field in Murcia (Spain), the second one from a cultured population reared on germinated potatoes at the "Centro de Sanidad Vegetal" (Generalitat Valenciana) in Almassora (Castelló, Spain).

Total DNA enriched in bacterial endosymbionts was extracted from viscera of 20–30 adult female insects in sterile conditions and mechanically homogenized. In order to reduce insect DNA contamination, the samples were subjected to consecutive centrifugations at 1150 g and 1300 g for 10 minutes, and genomic DNA was obtained from the supernatant following a CTAB (Cetyltrimethylammonium bromide) extraction method [48].

Genome sequencing and assembly

The purified genomic DNA was shotgun sequenced using 454/Roche GS-FLX Titanium technology at the Genomics and Health area of the Public Health Research Center (CSISP, Generalitat Valenciana). One half-plate single-ends, and one-fourth plate paired-ends (3 kb of fragment size) sequencing experiments were performed, yielding a total of 1.3 million reads. Sequences of eukaryotic origin were eliminated after a taxonomic assignation process by Galaxy [49]. Filtered reads were automatically assembled by MIRA [50] and the resulting contigs were manually edited with the Gap4 program from the Staden package software [51]. The remaining gaps in the genome of *M. endobia* str. PCVAL were closed by ABI sequencing of PCR products obtained with designed primers, at the sequencing facility of the Universitat de València. Potential *ori*C on both genomes were sought with the OriginX program [52].

Total DNA samples obtained from the *P. citri* populations from Murcia and Almassora were used to further analyze the *rpl*Q region from the *T. princeps* genome. The region comprised between genes *rpo*A and *aro*K was amplified and sequenced using the primers rpoA-F (5'-TGCCAGGCCTAGTGCTAAACATCA-3') and aroK-R (5'-TGTCGCCAGGACTGCTATCAATGT-3').

Gene annotation and functional analysis

ARAGORN [53], tRNAscan [54], and Rfam [55] software packages were used for RNA genes prediction. Coding genes were annotated by BASys (Bacterial Annotation System, [56], RAST [57] and refined by BLAST searches [58]. Finally, functional domain studies in Pfam database [59] were performed when coding-genes functionality assessment was required. Artemis [60] and

MEGA5 [61] programs were used for genome statistics calculation and codon usage analysis. Metabolic capabilities were analyzed with Blast2Go [62] and KAAS [63] programs. Functional information from the BioCyc [64], KEEG [65] and BRENDA [66] databases were also used in this context. Genome alignments were performed using MAFFT [67].

Annotated ORFs were considered as functional genes following two non-exclusionary criteria: the conservation of at least 80% of the sequence length of the closest orthologs found by BLAST in non-redundant databases, and/or the maintenance of the essential functional domains detected by Pfam [59].

Accession Numbers. The genome sequence of *M. endobia* strain PCVAL has been deposited at the GenBank (accession number CP003881). The GenBank accession numbers of the other three genome sequences used in this study are as follows: "*Ca*. Tremblaya princeps" str. PCVAL, CP002918; "*Ca*. Tremblaya princeps" str. PCIT, CP002244; *M. endobia* strain PCIT, CP002243.

AUTHORS' CONTRIBUTIONS

SLM and MP reared and sampled the insects, and performed the DNA extractions. SLM performed the *M. endobia* genome assembly and annotation, and the comparative analyses. SLM and RG performed the functional analysis and prepared figures and tables. RG, AL and AM designed and coordinated the study, and drafted and conducted the manuscript writing. All authors participated on the discussion, reading and approval of the final manuscript.

ACKNOWLEDGEMENTS

We thank Dr. Ferran Garcia (Universitat Politècnica de Valencia, Spain) and Alberto García (Centro de Sanidad Vegetal, Generalitat Valenciana, Almassora, Spain) for providing mealybug samples. Financial support was provided by grants BFU2009-12895-C02-01/BMC (Ministerio de Ciencia e Innovación, Spain) and BFU2012-39816-C02-01 (Ministerio de Economía y Competitividad, Spain) to A. Latorre and by grant Prometeo/2009/092 (Conselleria d'Educació, Generalitat Valenciana, Spain) to A. Moya. S. López-Madrigal is a recipient of a fellowship from the Ministerio de Educación (Spain).

REFERENCES

1. Moya A, Pereto J, Gil R, Latorre A: Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* 2008, 9:218–229.

2. McCutcheon JP, Moran NA: Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 2011, 10:13–26.

3. Watson RA: *The impact of sex, symbiosis and modularity on the gradualist framework of evolution*. Cambridge (Massachusetts): The MIT Press; 2006.

4. Gil R, Latorre A, Moya A: Evolution of prokaryote-animal symbiosis from a genomics perspective. In *(Endo)symbiotic Methanogenic Archaea*. Edited by Hackstein JHP. Berlin Heidelberg: Springer; 2010:207–233 [Steinbüchel A (Series Editor): *Microbiology Monographs*, vol. 19].

5. Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, Latorre A: *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet* 2011, 7:e1002357.

6. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, et al: Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 2006, 4:e188.

7. McCutcheon JP, McDonald BR, Moran NA: Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* 2009, 5:e1000565.

8. McCutcheon JP, Moran NA: Functional convergence in reduced genomes of bacterial symbionts spanning 200 MY of evolution. *Genome Biol Evol* 2010, 2:708–718.

9. Lefevre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A: Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial replacement. *Mol Biol Evol* 2004, 21:965–973.

10. *ScaleNet*. http://www.sel.barc.usda.gov/scalenet/scalenet.htm.

11. Hardy NB, Gullan PJ, Hodgson CJ: A subfamily-level classification of mealybugs (Hemiptera: Pseudococcidae) based on integrated molecular and morphological data. *Syst Entomol* 2008, 33:51–71.

12. Munson MA, Baumann P, Moran NA: Phylogenetic relationships of the endosymbionts of mealybugs (Homoptera: Pseudococcidae) based on 16S rDNA sequences. *Mol Phylogenet Evol* 1992, 1:26–30.

13. Gruwell ME, Hardy NB, Gullan PJ, Dittmar K: Evolutionary relationships among primary endosymbionts of the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). *Appl Environ Microbiol* 2010, 76:7521–7525.

14. Thao ML, Gullan PJ, Baumann P: Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Appl Environ Microbiol* 2002, 68:3190–3197.

15. Von Dohlen CD, Kohler S, Alsop ST, McManus WR: Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature* 2001, 412:433–436.

16. McCutcheon JP, Von Dohlen CD: An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 2011, 21:1366–1372.

17. Kono M, Koga R, Shimada M, Fukatsu T: Infection dynamics of coexisting beta and gammaproteobacteria in the nested endosymbiotic system of mealybugs. *Appl Environ Microbiol* 2008, 74:4175–4184.

18. Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P: The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. *Appl Environ Microbiol* 2002, 68:3198–3205.

19. Lopez-Madrigal S, Latorre A, Porcar M, Moya A, Gil R: Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. *J Bacteriol* 2011, 193:5587–5588.

20. Gil R, Latorre A, Moya A: Bacterial endosymbionts of insects: insights from comparative genomics. *Environ Microbiol* 2004, 6:1109–1122.

21. Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Holldobler B, Van Ham RCHJ, Gross R, Moya A: The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc Nat Acad Sci USA* 2003, 100:9388–9393.

22. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S: Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 2002, 32:402–407.

23. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M: The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 2006, 314:267.

24. McCutcheon JP, McDonald BR, Moran NA: Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Nat Acad Sci USA* 2009, 106:15394–15399.

25. McCutcheon JP, Moran NA: Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Nat Acad Sci USA* 2007, 104:19392–19397.

26. Gil R, Latorre A: Factors behind junk DNA in bacteria. *Genes* 2012, 3:634–650.

27. Hershberg R, Petrov DA: Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* 2010, 6:e1001115.

28. Hildebrand F, Meyer A, Eyre-Walker A: Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 2010, 6:e1001107.

29. Behrens S, Maier R, De Cock H, Schmid FX, Gross CA: The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J* 2001, 20:285–294.

30. Dermic D: Functions of multiple exonucleases are essential for cell viability, DNA repair and homologous recombination in *rec*D mutants of *Escherichia coli*. *Genetics* 2006, 172:2057–2069.

31. Heller RC, Marians KJ: The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. *Mol Cell* 2005, 17:733–743.

32. Xu L, Marians KJ: Purification and characterization of DnaC810, a primosomal protein capable of bypassing PriA function. *J Biol Chem* 2000, 275:8196–8205.

33. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA III, Venter JC: The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995, 270:397–403.

34. Lopper M, Boonsombat R, Sandler SJ, Keck JL: A hand-off mechanism for primosome assembly in replication restart. *Mol Cell* 2007, 26:781–793.

35. Gil R, Silva FJ, Pereto J, Moya A: Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* 2004, 68:518–537.

36. Quan S, Zhang N, French S, Squires CL: Transcriptional polarity in rRNA operons of *Escherichia coli nus*A and *nus*B mutant strains. *J Bacteriol* 2005, 187:1632–1638.

37. Price NL, Raivio TL: Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J Bacteriol* 2009, 191:1798–1815.

38. Tseng TT, Tyler BM, Setubal JC: Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* 2009, 9(Suppl 1):S2.

39. Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A: A small microbial genome: the end of a long symbiotic relationship? *Science* 2006, 314:312–313.

40. Johnson DC, Dean DR, Smith AD, Johnson MK: Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* 2005, 74:247–281.

41. Beinert H: Iron-sulfur proteins: ancient structures, still full of surprises. *J Biol Inorg Chem* 2000, 5:2–15.

42. Malinverni JC, Werner J, Kim S, Sklar JG, Kahne D, Misra R, Silhavy TJ: YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Mol Microbiol* 2006, 61:151–164.

43. Singh N, Kuppili RR, Bose K: The structural basis of mode of activation and functional diversity: a case study with HtrA family of serine proteases. *Arch Biochem Biophys* 2011, 516:85 96.

44. Sawa J, Malet H, Krojer T, Canellas F, Ehrmann M, Clausen T: Molecular adaptation of the DegQ protease to exert protein quality control in the bacterial cell envelope. *J Biol Chem* 2011, 286:30680–30690.

45. Ajouz B, Berrier C, Garrigues A, Besnard M, Ghazi A: Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells. *J Biol Chem* 1998, 273:26670–26674.

46. Berrier C, Garrigues A, Richarme G, Ghazi A: Elongation factor Tu and DnaK are transferred from the cytoplasm to the periplasm of *Escherichia coli* during osmotic downshock presumably via the mechanosensitive channel mscL. *J Bacteriol* 2000, 182:248–251.

47. van den Bogaart G, Krasnikov V, Poolman B: Dual-color fluorescence-burst analysis to probe protein efflux through the mechanosensitive cannel MscL. *Biophys J* 2007, 92:1233–1240.

48. Ausubel F: *Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology*. New York: Wiley 1999.

49. *Galaxy*. http://galaxyproject.org.

50. *MIRA*. http://sourceforge.net/projects/mira-assembler.

51. Staden R, Beal KF, Bonfield JK: The Staden package, 1998. *Methods Mol Biol* 2000, 132:115–130.

52. Worning P, Jensen LJ, Hallin PF, Staerfeldt HH, Ussery DW: Origin of replication in circular prokaryotic chromosomes. *Environ Microbiol* 2006, 8:353–361.

53. Laslett D, Canback B: ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 2004, 32:11–16.

54. Lowe TM, Eddy SR: tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997, 25:955–964.

55. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A: Rfam: updates to the RNA families database. *Nucleic Acids Res* 2009, 37(Database issue):D136–D140.

56. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS: BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res* 2005, 33(Web Server issue): W455–W459.

57. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O: The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008, 9:75.

58. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389–3402.

59. *Pfam*. http://pfam.sanger.ac.uk.

60. Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream MA: Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 2008, 24:2672–2676.

61. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28:2731–2739.

62. *Blast2Go*. http://www.blast2go.com.

63. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: KAAS: an automatic genome annotation and pathway reconstruction server. *Web Server issue* 2007, 35:W182–W185.

64. *BioCyc*. http://biocyc.org.

65. *KEEG*. http://www.genome.jp/kegg.

SUPPLEMENTARY MATERIAL

Table S1. Differences in gene annotation between strains PCIT and PCVAL for *T. princeps* and *M. endobia*. Gene names refer to the annotation of the PCVAL strain. For those genes duplicated, or encoding hypothetical or unknown proteins, the locus tag is indicated. Gene names or locus tags for the PCIT strain are indicated into brackets when necessary. (+) functional gene; (-) missing gene; (Ψ) pseudogene.

Species	Gene	Product	Status		Comments
			PCIT	PCVAL	
T. princeps	(*rplU*)	50S ribosomal subunit protein L21	+	-	No significant blastx hits against nr-database
	(*aceF*)	E2 component of pyruvate dehydrogenase	Ψ	-	No significant blastx hits against nr-database
	(TPPCIT_152)	Unknown protein	Ψ	-	No significant blastx hits against nr-database
	(TPPCIT_150)	Hypothetical protein	Ψ	-	No significant blastx hits against nr-database
	(*lpd)*	Lipoamide dehydrogenase	Ψ	-	No significant blastx hits against nr-database
	*leu*A/TCP_127	α-isopropylmalate synthase	-	Ψ	Previously annotated as part of *T. princeps* partial genomic duplication [18]
	*ace*E	Pyruvate dehydrogenase	-	Ψ	*Locus* identity confirmed by blastx against nr-database
	TCP_012	Hypothetical protein	+	Ψ	One frameshift and an early stop codon with respect to the originally annotated *locus* [18]
	*smp*B	SsrA-binding protein	+	Ψ	Truncated at both extremes, 50% shorter than its closest orthologs. No detection of the corresponding functional domain by Pfam.
	*trp*G	Anthranilate synthase component I	+	Ψ	No detectable similarities against N-terminal extreme of its closest orthologs. Less than 40% of the corresponding functional domain recognized by Pfam.
	TCP_134	Hypothetical protein	+	Ψ	Several frameshifts with respect to the originally annotated *locus* [18]
	tRNA-Lys (TPPCIT_098)	Anticodon=CUU	+	Ψ	Annotated as a pseudogene by tRNAscan-SE
	*grp*E	Heat shock protein	Ψ	+	Retains more than 50% of its closest orthologs; almost 65% of the corresponding functional domain detected by Pfam
	*rpo*D	RNA polymerase, sigma 70 subunit	Ψ	+	Although evidently reduced, 3 (r2, r3 y r4) out of the 5 functional domains in its closest orthologs recognized by Pfam. The lost r1.1 domain has a regulatory role.
	tRNA-Glu	Anticodon=UUC	-	Ψ	Pseudogene predicted by tRNAscan-SE
M. endobia	*ygb*Q (MEPCIT_471)	Essential cell division protein FtsB	Ψ	+	Shortened, 65% of the corresponding functional domain recognized by Pfam.
	MPC_123	Hypothetical protein	Ψ	+	More than 60% gene length retained, showing more than 80% identity with its closest ortholog.
	*ssr*A (MEPCIT_479)	tmRNA (hypothetical protein)	+	+	Different annotation: tmRNA gene identified by Rfam in PCVAL. Hypothetical protein in PCIT.
	*yac*G (MEPCIT_480)	Zinc-binding protein	Ψ	+	Retains 70% length of its ortholog in *S. glossinidius*, with 70% identity. The corresponding functional domain recognized by Pfam
	*ibg*I (MEPCIT_466)	hydrolase-oxidase protein	Ψ	+	More than 72% gene length retained, showing more than 73% identity with its ortholog in *S. glossinidius*. Most of the corresponding functional domain recognized by Pfam
	MPC_265	Hypothetical protein	-	Ψ	Annotation based on blastx and synteny analysis, compared with *S. glossinidius* (SG0460)
	*pdx*J/MPC_306	Pyridoxine 5'-phosphate synthase	-	Ψ	Pseudogene involved in the partial genomic duplication

Table S2. Codon usage bias in *T. princeps* PCVAL and *M. endobia* PCVAL. Codon frequencies resulted significantly biased (p-value= 0.01) for all amino acids in *T. princeps*. The same applies to *M. endobia* except for cysteine. In black, frequency of the most used codon for the corresponding amino acid in both species.

		T. princeps		***M. endobia***	
AA	**Codons**	**Abundance**	**%**	**Abundance**	**%**
Pro	CCA	454	36.09	1953	37.08
	CCT	291	23.13	1245	23.64
	CCG	274	21.78	1514	28.75
	CCC	239	18.99	555	10.54
Arg	AGG	1340	42.27	372	4.53
	CGC	747	23.57	2677	32.62
	CGG	331	10.44	818	9.96
	CGT	289	9.11	3005	36.60
	AGA	253	7.98	552	6.72
	CGA	210	6.63	785	9.56
Ala	GCA	1139	31.15	3255	27.95
	GCC	897	24.53	2505	21.49
	GCG	857	23.44	2349	20.16
	GCT	764	20.89	3543	30.40
Gly	GGC	1316	48.63	3107	32.35
	GGG	526	19.44	898	9.35
	GGA	450	16.63	1473	15.34
	GGT	414	15.30	4125	42.96
Leu	CTG	907	33.30	3364	22.85
	CTA	613	22.50	3094	21.01
	CTT	482	17.70	2070	14.06
	CTC	459	16.85	659	4.47
	TTG	172	6.31	2720	18.47
	TTA	91	3.34	2818	19.15
Val	GTG	895	35.76	2503	25.52
	GTA	595	23.77	3005	30.63
	GTT	563	22.49	2987	30.46
	GTC	450	17.98	1313	13.39
His	CAC	604	62.01	1103	32.12
	CAT	370	37.99	2330	67.88
Gln	CAG	610	81.33	3320	54.15
	CAA	140	18.67	2808	45.85
Asp	GAC	770	62.00	2027	29.63
	GAT	472	38.00	4815	70.37
Glu	GAG	933	81.20	2217	28.77
	GAA	216	18.80	5490	71.23
Ser	AGC	686	30.71	1753	22.28
	TCG	393	17.60	913	11.61
	TCA	384	17.12	1297	16.49
	TCC	366	16.38	910	11.57
	TCT	286	12.80	1451	18.47
	AGT	119	5.33	1541	19.59
Thr	ACG	434	30.80	1026	13.52
	ACC	366	25.98	2501	32.96
	ACA	340	24.13	1578	20.83
	ACT	269	19.10	2481	32.69
Cys	TGC	747	84.31	896	50.20
	TGT	139	15.69	889	49.80
Trp	TGG	235	100	1469	100
Ile	ATA	711	53.10	2444	24.33
	ATC	354	26.44	2781	27.70
	ATT	274	20.46	4818	47.97
Met	ATG	981	100	3793	100
Asn	AAC	473	74.96	2542	40.70
	AAT	158	25.04	3702	59.30
Lys	AAG	602	83.50	2032	29.10
	AAA	119	16.50	4951	70.90
Phe	TTC	485	81.93	1699	34.99
	TTT	107	18.07	3154	65.01
Tyr	TAC	499	72.11	1444	33.97
	TAT	193	27.89	2805	66.03

Table S3. Aminoacyl tRNA synthetases and tRNA genes detected in the *T. princeps* and *M. endobia* genomes. (+) annotated gene; (-) absent gene; (Ψ) pseudogene; (N) number of tRNA isoacceptors detected.

	T. princeps			***M. endobia***		
tRNA	Anticodons	N	Aminoacyl tRNA synthetase	Anticodons	N	Aminoacyl tRNA synthetase
tRNA-Ala	UGC	1	-	GGC, UGC	2	+
tRNA-Gly	-	-	-	GCC, UCC	2	+
tRNA-Pro	-	-	-	UGG, CGG, GGG	3	+
tRNA-Thr	-	-	-	GGU, UGU, CGU	3	+
tRNA-Val	-	-	-	UAC, GAC	2	+
tRNA-Ser	UGA	1	-	GGA, CGA, UGA, GCU	4	+
tRNA-Arg	ACG	1	Ψ	ACG, CCG, CCU, UCU	4	+
tRNA-Leu	-	-	-	UAG, CAG, GAG, UAA, CAA	5	+
tRNA-Phe	GAA	Ψ	-	GAA	1	+
tRNA-Asn	-	-	-	GUU	1	+
tRNA-Lys	CUU	2, Ψ	-	UUU	1	+
tRNA-Asp	-	-	-	GUC	2	+
tRNA-Glu	UUC	Ψ	-	UUC	1	+
tRNA-His	-	-	-	GUG	1	+
tRNA-Gln	-	-	-	UUG, CUG	2	+
tRNA-Ile	GAU	Ψ	-	GAU	1	+
tRNA-Met	CAU	1	-	CAU	3	+
tRNA-Tyr	-		-	GUA	1	+
tRNA-Cys	-		Ψ	GCA	1	+
tRNA-Trp	CCA	1	-	CCA	1	+

CHAPTER 3

Status: Published. October 21, 2013.

Citation: López-Madrigal S, Balmand S, Latorre A, Heddi A, Moya A, Gil R. 2013. How does *Tremblaya princeps* get essential proteins from its nested partner *Moranella endobia* in the mealybug *Planoccocus citri*? PLoS One 8:e77307.

How does *Tremblaya princeps* get essential proteins from its nested partner *Moranella endobia* in the mealybug *Planoccocus citri*?

Sergio López-Madrigal[1], Séverine Balmand[2], Amparo Latorre[1,3], Abdelaziz Heddi[2], Andrés Moya[1,3], Rosario Gil[1]*

1 Institut Cavanilles de Biodiversitat y Biologia Evolutiva, Universitat de València, Paterna (València), Spain. 2 UMR203 BF2I, Biologie Fonctionnelle Insectes et Interactions, INSA-Lyon, INRA, Villeurbanne, France. 3 Área de Genómica y Salud, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana FISABIO–Salud Pública, València, Spain.

ABSTRACT

Many insects maintain intracellular mutualistic symbiosis with a wide range of bacteria which are considered essential for their survival (primary or P-endosymbiont) and typically suffer drastic genome degradation. Progressive loss of P-endosymbiont metabolic capabilities could lead to the recruitment of co-existent facultative endosymbiont (secondary or S-endosymbiont), thus adding more complexity to the symbiotic system. *Planococcus citri*, among other mealybug species, harbors an unconventional nested endosymbiotic system where every *Tremblaya princeps* cell (β-proteobacterium) harbors many *Moranella endobia* cells (γ-proteobacterium). In this system, *T. princeps* possess one of the smallest prokaryote genome known so far. This extreme genome reduction suggests the supply of many metabolites and essential gene products by *M. endobia*. Although sporadic cell lysis is plausible, the bacterial participation on the regulation of the predicted molecular exchange (at least to some extent) cannot be excluded. Although the comprehensive analysis of the protein translocation ability of *M. endobia* PCVAL rules out the existence of specific mechanisms for the exportation of proteins from *M. endobia* to *T. princeps*, immunolocation of two *M. endobia* proteins points towards a non-massive but controlled protein provision. We propose a sporadic pattern for the predicted protein exportation events, which could be putatively controlled by the host and/or mediated by local osmotic stress.

INTRODUCTION

Symbiosis is a natural widespread phenomenon that has been postulated to be a key factor for the evolutionary success of insects, many of which maintain mutualistic symbiotic relationships with intracellular bacteria. These endosymbiotic bacteria inhabit specialized host cells (bacteriocytes) and complement their normally unbalanced diets [1,2,3]. According to their dispensability for insect survival, they are classified as primary (P) or obligate endosymbionts,

and secondary (S) or facultative symbionts, respectively. Thus, while S-symbionts can be horizontally transferred, are not necessarily present in every individual of a certain host species and can be placed outside bacteriocytes [4], P-endosymbionts are only vertically transmitted (from mothers to offspring). Strong incidence of genetic drift, together with relaxation of purifying selection on genes rendered unnecessary in the intracellular environment, lead P-endosymbiont genomes to undergo a huge size reduction. Eventually, if an S-symbiont is present, interactions among both bacteria and the eukaryotic host would take place, and new genes will become redundant. Thus, the P-endosymbiont might lose genes involved in the provision of metabolic capabilities required by the host, which can still be recruited from the co-existing S-symbiont (then becoming co-primary) [5]. Ongoing degeneration of both bacterial genomes could eventually cause a reciprocal metabolic complementation, adding more complexity to this ecological system [6,7,8,9].

Many studied mealybug species from the subfamily Pseudococcinae harbor an unusual nested endosymbiotic organization [10,11] in which each cell of the β-proteobacteria *Candidatus* Tremblaya princeps (*T. princeps* from now on, for the sake of simplicity) harbors several cells of another endosymbiont belonging to different bacterial clades depending on the host species. Both members of the consortium seem to be closely involved in the nutritional and reproductive physiology of their hosts [12]. In the mealybug *Planococcus citri*, *T. princeps* harbors the γ-proteobacterium *Candidatus* Moranella endobia (*M. endobia* from now on). Although *T. princeps* was originally considered as the P-endosymbiont according to phylogenetic criteria [13], recent complete genome sequencing of the two endosymbionts from two *P. citri* strains (PCIT and PCVAL) showed that both bacteria are functionally co-primaries [14,15,16], and display an unprecedented level of metabolic complementation between them. *T. princeps* possess the second smallest prokaryote genome described so far, most of which is devoted to the production of nearly complete ribosomes, with almost null metabolic capabilities except for the assembly of [Fe-S] clusters and the ability to partially synthesize some essential amino acids. Not only a huge range of metabolites but also proteins and tRNAs are supposed to be transferred from *M. endobia* to *T. princeps* in order to perform even essential informational functions, i.e., replication, transcription and translation.

Despite all mentioned *in silico* predictions, the way *M. endobia* proteins are recruited by *T. princeps* remains unknown. A recent survey of the *P. citri* nuclear genome led to the discovery that several genes of bacterial origin (neither from *T. princeps* nor *M. endobia*), involved in peptidoglycan production and recycling, have been acquired by the host by horizontal gene transfer, and are being expressed in bacteriocytes [17]. The authors propose that these genes might be involved in the release of the *M. endobia* cytoplasmic content by cell lysis. However, the existence of controlled mechanisms for specific macromolecules exportation from *M. endobia* to *T. princeps* cannot be ruled out.

The Sec machinery is the most generally employed mechanism for protein translocation across the inner membrane in Gram-negative bacteria, including endosymbionts. Sec-dependent

secretory proteins can be exported, periplasmic and outer membrane proteins. They are synthesized at the cell cytoplasm as precursor macromolecules, carrying cleavable amino-terminal signal peptide (SP) sequences. Although *M. endobia* encodes an apparently functional Sec translocation complex, proteins with SP appear to be scarce in its proteome. McCutcheon and von Dohlen [14] have roughly reported that only 27 proteins contain SP sequences in *M. endobia* PCIT. Nevertheless, protein exportation might still take place through an abnormally permeable Sec translocation complex. Alternatively, the proteins can be exported through a non-specific transport mechanism, such as the permissive MscL membrane channel. The *msc*L gene has been preserved in the *M. endobia* genome [14,16] while it is absent in all other known endosymbionts with reduced genomes. MscL forms a mechanosensitive channel which acts as a pressure release valve allowing solutes to exit the cell through a large pore in response to environmental osmotic downshift [18,19,20], and passage of small macromolecules through it has been described [21,22,23].

In order to better understand the mechanisms behind the provision of essential *M. endobia* proteins to the *T. princeps* cytoplasm, we have explored *in silico* the potential of the *M. endobia* Sec translocon machinery to participate on the process, and applied inmunohistochemistry assays with polyclonal antibodies to reveal the location of two *M. endobia* proteins throughout the nested endosymbiotic system: the channel protein MscL, only encoded in the *M. endobia* genome, and the chaperone Hsp60 (GroEL), a highly expressed protein in endosymbionts [1] that is also encoded in the *T. princeps* genome. Our results show the lack of massive and constitutive protein traffic from *M. endobia* to the *T. princeps* cytoplasm. Thus, both *in silico* analysis and experimental evidences support a model were *M. endobia* proteins would mostly be retained in its cell, allowing the controlled passage of needed macromolecules and metabolites through a highly permissive cell wall, whose strength could be controlled by the insect host, and only sporadically releasing its cytoplasmic content by cell lysis.

RESULTS AND DISCUSSION

The Sec protein secretion pathway in *M. endobia*, under scrutiny

The dramatic reduction of the *T. princeps* genome implies the need for the recruitment of a wide range of proteins, whose unique possible source is the *M. endobia* cytoplasm, for the performance of essential cellular functions [14,15,16]. Although a simplified but likely functional Sec translocon is encoded by the *M. endobia* genome [14,16], its potential implication in both endosymbionts molecular communication has not been analyzed in detail.

***Signal sequences prediction along M. endobia proteome*.** To determine if *M. endobia* proteins that are essential for *T. princeps* could be exported by this general secretion system, we performed a comprehensive scanning of the 411 coding sequences from *M. endobia* str. PCVAL, using the PRED-TAT [24] and SignalPv4.1 softwares [25]. A putative SP sequence was predicted

for 33 proteins at least by one of the softwares (Table 1), a fairly similar number to what has been mentioned for strain PCIT [14]. In 18 cases, the same program also detected a SP sequence at the *E. coli* and/or *S. glossinidius* homolog, although such motif was not always described in the literature, based on the information available at EcoCyc v16.5 [26]. In some other cases, SP has only been predicted in the *M. endobia* homolog. Several of such proteins are known to maintain strict cytoplasmic distribution in other bacteria. They could be false positives or may represent cases of the gain of SP mimicking sequences by traditionally non-secreted proteins and, therefore, reflect possible adaptations to the nested endosymbiosis. In any case, most proteins for which an SP sequence has been predicted are usually located in the cell envelope and, only a small amount of cytoplasmic proteins needed to perform essential functions have been identified in this analysis. Even more, the two ribosomal proteins detected are also encoded by the *T. princeps* genome. Therefore, it is not plausible that a canonical Sec translocon could be used for the provision of essential proteins to *T. princeps*.

***Permeability of the Sec translocon machinery in M. endobia*.** A variety of genetic and biochemical approaches led to the molecular characterization of the Sec translocon machinery in *E. coli* through the description of a group of dominant mutations allowing the exportation of signal sequence-defective precursors or even signal sequence-less proteins, collectively called *prl* alleles [27,28] These mutations are able to expand the repertoire of secretory proteins. To date, well characterized *E. coli prl* phenotypes have been linked to mutations on a total of 37 codons at genes *sec*A (*prlD* mutants) [29,30], *sec*E (*prlG* mutants), *sec*G (*prlH* mutants) [31] and *sec*Y (*prlA* mutants) [28]. In order to evaluate the Sec translocon permeability in *M. endobia*, we performed comparative analysis between the products of these orthologous genes in *M. endobia* and *E. coli*. Identity levels range from 89.5% for SecY, 77.4% for SecA, 64.3% for SecE to 49.1% for SecG (unambiguous alignment of 440, 839, 126 and 110 amino acids, respectively). Nevertheless, among the 37 amino acids involved in *prl* mutations in *E. coli*, the vast majority (33) was conserved in *M. endobia* (Table S1). The four observed amino acid substitutions might have a small impact in protein function, since amino acid polarity and molecular weight is almost unaffected. Additionally, the impact of some of the mutations detected in *E. coli* in two of the cases are not clear, because they have been found in double mutants, and the additional change detected has been linked to a *prl* phenotype by itself. This is the case of the A277E change in *sec*Y observed in mutant *prlA7*, and E148K change in *sec*A observed in mutant *prlD21*.

According to our results, no clear evidences for abnormal Sec translocon permeability are observable in *M. endobia*, suggesting that SP sequences would be still necessary for *M. endobia* proteins exportation across this machinery. However, since overexpression of translocation machinery components such as SecA had also been linked to *prl* phenotypes [29], and analyzed residues are probably just a fraction of those potentially relevant for proper functioning, we cannot rule out such a possibility. Experimental approaches should be necessary in order to get a more realistic view on the molecular communication between both nested endosymbiosis members through this general protein secretion system.

Table 1. *M. endobia* proteins potentially harboring SP sequences.

Gene	Protein product	Cellular function	Program	SP	CS	Eco	Sgl	Location
skp	Chaperone	DegP/Skp folding pathway	SignalP	1..30	[AQA-ND]	+	+	P, C
*deg*Q	Serine endoprotease	DegP/Skp folding pathway	PRED-TAT	1..33	[ARA-RP]	+	+	P
*yra*P	Uncharacterized protein	DegP/Skp folding pathway	PRED-TAT	1..22	[VGA-MV]	+	+	P
*sur*A	Chaperone	OMP biogenesis	SignalP	1..20	[TLA-MS]	+	+	P
*znu*A	High-affinity zinc uptake system protein	ABC transporter	PRED-TAT	1..36	[AQA-AL]	+	+	P
*lpt*A	LPS export system protein, LptA subunit	ABC transporter	PRED-TAT SignalP	1..34	[ALA-LT]	+	+	P
*lpt*D	LPS transport and assembly complex, LptD subunit	LPS transport and assembly	PRED-TAT SignalP	1..25	[ARA-AL]	+	+	IM, OM
*lpt*E	LPS transport and assembly complex, LptE subunit	LPS transport and assembly	PRED-TAT	1..24	[ATA-AT]	+	+	IM, OM
*bam*A	Outer membrane protein assembly factor	OMP biogenesis	PRED-TAT SignalP	1..19	[SRA-DE]	+	+	IM, OM
*bam*D	Lipoprotein	OMP biogenesis	PRED-TAT SignalP	1..18	[VLA-DC]	+	+	IM, OM
*omp*F	Outer membrane protein F	Pore	PRED-TAT SignalP	1..21	[ARA-TE]	+	+	IM, OM
*dsb*B	Disulfide bond formation protein B	Protein modification	PRED-TAT	1..24	[AFA-LE]	-	-	IM
*min*D	Septum site-determining protein	Cell shape and division	PRED-TAT	1..21	[SSA-SI]	+	-	IM
*ygb*Q	Hypothetical protein	Cell shape and division	PRED-TAT	1..18	[QYA-LW]	-	+	IM
*nuo*A	NADH-quinone oxidoreductase subunit A	Electron transport	PRED-TAT	1..41	[AQA-RT]	-	-	IM
*nuo*L	NADH-quinone oxidoreductase subunit L	Electron transport	PRED-TAT	1..26	[RWS-EN]	+*	+	IM
*htp*X	Protease	Poorly characterized	PRED-TAT	1..29	[IQS-SS]	+*	+	IM
*ybg*T	Hypothetical protein	Poorly characterized	PRED-TAT	1..23	[ALA-IE]	+	-	IM, OM
*lpo*B	Lipoprotein, penicillin-binding protein activator	Peptidoglycan biosynthesis	PRED-TAT	1..25	[PQA-NI]	+	+	IM, OM
*mlt*C	Membrane-bound lytic murein transglycosylase C	Peptidoglycan biosynthesis	PRED-TAT	1..23	[THG-KE]	+	+	IM, OM
*lpd*A	Dihydrolipoyl dehydrogenase	Pyruvate and TCA metabolism	PRED-TAT	1..22	[SAA-FR]	-	-	IM, C
*dap*B	Dihydrodipicolinate reductase	Amino acids biosynthesis	PRED-TAT	1..26	[IQA-VT]	-	-	C
*fab*B	3-oxoacyl-[acyl-carrier-protein] reductase	Fatty acid biosynthesis	PRED-TAT	1..22	[AIA-ET]	-	-	C
*fab*G	3-oxoacyl-[acyl-carrier-protein] reductase	Fatty acid biosynthesis	PRED-TAT	1..22	[AIA-ET]	-	-	C
*ser*C	Phosphoserine aminotransferase	Vitamin B6 metabolism	PRED-TAT	1..22	[SQA-QQ]	-	-	C
*erp*A	Iron-sulfur cluster insertion protein erpA	FeS clusters biosynthesis	PRED-TAT	1..25	[LIA-AE]	-	-	C
*trx*B	Thioredoxin reductase	Electron transport	PRED-TAT	1..28	[ARA-NL]	-	-	C
*arg*S	Arginyl-tRNA synthetase	Translation	PRED-TAT	1..26	[CEA-QV]	-	-	C
*dea*D	Cold-shock DEAD box protein A	Translation	PRED-TAT	1..25	[LQA-LT]	-	-	C
*rpl*K	50S ribosomal protein L11	Translation	PRED-TAT	1..29	[ALG-QQ]	-	-	C
*rpl*L	50S ribosomal protein L7/L12	Translation	PRED-TAT	27..49	[AEA-AE]	-	-	C
*tus*B	tRNA 2-thiouridine synthesizing protein	Translation	PRED-TAT	1..21	[RSA-QT]	-	-	C
*yje*E	ATP-binding protein	Poorly characterized	PRED-TAT	1..23	[VAA-AC]	-	-	C

SP coordinates, the more likely cleavage site (CS) for each protein, detection in *E. coli* (Eco) and *S. glossinidius* (Sgl), and subcellular location are indicated. (C) cytoplasm; (IM) inner membrane; (OM) outer membrane; (P) periplasm.*SP not described in EcoCyc v16.5

Search for non-specific protein efflux mechanisms

Since only a short set of *M. endobia* membrane proteins were shown to harbor secretory SPs (Table 1), and no evident permeability tolerance is expected for its Sec translocon (Table S1), we decided to explore the potential relevance of passive and non-specific transport mechanisms, such as mechanosensitive channels. MscL is one of several mechanosensitive ion channels that have been characterized in *E. coli*. However, it is the only one detected in *M. endobia*. An extensive search in the genomic databases showed that even this one is absent in all insect endosymbionts analyzed to date, including *T. princeps*.

MscL is known to release osmotic metabolites and ions in response of osmotic downshock, preventing cell lysis [18,32]. It has also been empirically proven to allow the passage of small proteins up to 6.5 kDa [21], and there is some controversy regarding the possibility that thioredoxin (11.5 kDa) can go through it [21,22]. Although other authors indicate that the 41-kDa chaperone DnaK [23,33], the 52-kDa elongation factor Tu [23,34], or even the 142-kDa enterobactin synthase EntF [35] can be released through MscL pores in osmotically stressed *E. coli* cells, the passage of these proteins through the channel has never been demonstrated, and it seems unlikely due to their large sizes [21]. However, it has been proven that *E. coli* cells subject to osmotic shock release up to 10% of their protein content, including periplasmic and cytoplasmic components under 100 kDa in a McsL-independent manner [36]. These authors suggest that osmotic shock might cause transient perforation of the plasma membrane, so that proteins can exit the damaged bacterial envelope pushed through a molecular sieve, whose limiting size is determined by the peptidoglycan mesh that surrounds the cell.

In silico predictions based on the *M. endobia* gene content and the discovery of several genes involved in peptidoglycan biosynthesis in the host nuclear genome indicate that the strength of the murein sacculus might be controlled by the host [17]. It is tempting to speculate that, if the above described mechanism is active in this bacterium, the peptidoglycan mesh could allow the passage of even larger proteins in a controlled manner. Furthermore, the outer membrane of *M. endobia* contains lipid IV_A, lacking any secondary acyl chains and Kdo (2-keto 3-deoxy-D-mannooctulosonate), instead of the usual lipopolysaccharydes (LPS) typically observed in gram-negative bacteria. This particular composition could probably lead to a more fluid and permeable outer membrane [37], which might also facilitate the efflux of these large proteins. In this hypothetical scenario, the higher metabolic activity of *M. endobia*, compared with *T. princeps* would generate osmotic pressure on the *M. endobia* membrane, which could lead to the release of some cytoplasmic proteins through the unspecific mechanism above described until it is alleviated by the action of McsL. The channel could also participate in the extrusion of small peptides. Thus, the mechanism of protein passage would depend on the corresponding protein dimensions.

Among the gene products that should be supplied by *M. endobia* to allow *T. princeps* functional viability, those involved in DNA replication and recombination, potentially responsible for the

concerted evolution noticed on *T. princeps* genome [16], are especially relevant. As seen on Table 2, most of them are below 100 kDa of molecular weight and, therefore, could go through the *M. endobia* envelope by the proposed mechanism.

Table 2. *M. endobia* genes potentially involved in concerted evolution in *T. princeps,* ordered by their product sizes.

Gene	Product	PL (aa)	MW (kDa)
priB	Primosomal replication protein N	108	11.9
ruvC	Crossover junction endodeoxyribonuclease RuvC	164	17.9
ssb	Single-stranded DNA-binding protein	186	20.6
ruvA	Holliday junction ATP-dependent DNA helicase RuvA	206	22.9
ruvB	Holliday junction ATP-dependent DNA helicase RuvB	334	36.9
recA	Protein RecA	353	38.3
recJ	Single-stranded-DNA-specific exonuclease RecJ	579	63.1
ligA	DNA ligase	676	75.8
dnaG	DNA primase	582	66.6
recG	ATP-dependent DNA helicase RecG	695	78.1
priA	Primosomal protein N'	716	81.5
polA	DNA polymerase I	939	105.7
recC	Exodeoxyribonuclease V gamma chain	1101	127.0
recB	Exodeoxyribonuclease V beta chain	1181	134.6

PL: Predicted protein length; MW: Predicted protein molecular weight.

Testing the controlled protein efflux mechanism *versus* the cell lysis hypotheses

Husnik and coworkers [17] suggested that the presence of expressed genes of bacterial origin involved in peptidoglycan biosynthesis and recycling in the host genome might be an indication that cell lysis is the mechanism used to provide *M. endobia* proteins and metabolites to *T. princeps*. If the above proposed scenario of controlled efflux of *M. endobia* proteins towards *T. princeps* were correct, only small quantities of *M. endobia* proteins would be present in *T. princeps* at a given time, contrary to what would be expected if they are released by cell lysis. In order to test this prediction, we performed experimental immunolocation studies to determine the spatial distribution of two proteins encoded by the *M. endobia* genome, GroEL and MscL, across the nested endosymbiotic system.

GroEL is an essential protein required for the proper folding of a wide range of cytosolic proteins [38] It was selected because it has been proven to be highly expressed in endosymbiotic systems, where it is proposed to alleviate the destabilizing effects of nonsynonimous mutations during protein folding [1,39,40]. In our experiments (Figure 1), fluorescent signal aggregated into bacilliform bodies (Figure 1, panels F, G), following a pattern that accurately mimics the one obtained when specific *M. endobia* detection is performed by Fluorescence *In Situ* Hybridization (FISH) experiments (Figure 1, panels C, D), thus proving that the protein is confined inside *M. endobia* cells. Only in a few cases, some GroEL staining was detected at the *T. princeps* inner membrane surface. Although there is also a *gro*EL gene in *T. princeps*, only the *M. endobia gro*EL product could be detected. Even though GroEL is a highly conserved protein, this is not a

completely unexpected result. First, the polyclonal antibodies used to detect the protein were generated against GroEL from *Buchnera aphidicola* APS [41], the endosymbiont of the pea aphid. Both *B. aphidicola* and *M. endobia* are γ-proteobacteria, whereas *T. princeps* is a β-proteobacteria. Additionally, differences between GroEL proteins from *M. endobia* and *T. princeps* are pronounced in terms of inmuno-histochemical detection, since most identical residues are located in short motives that are unlikely immunogenic (Figure S1). Thus, only half of the conserved residues map on potential minimal epitopes (5–8 residues), and only 38% of them map on amino acidic motives over 10 residues in size.

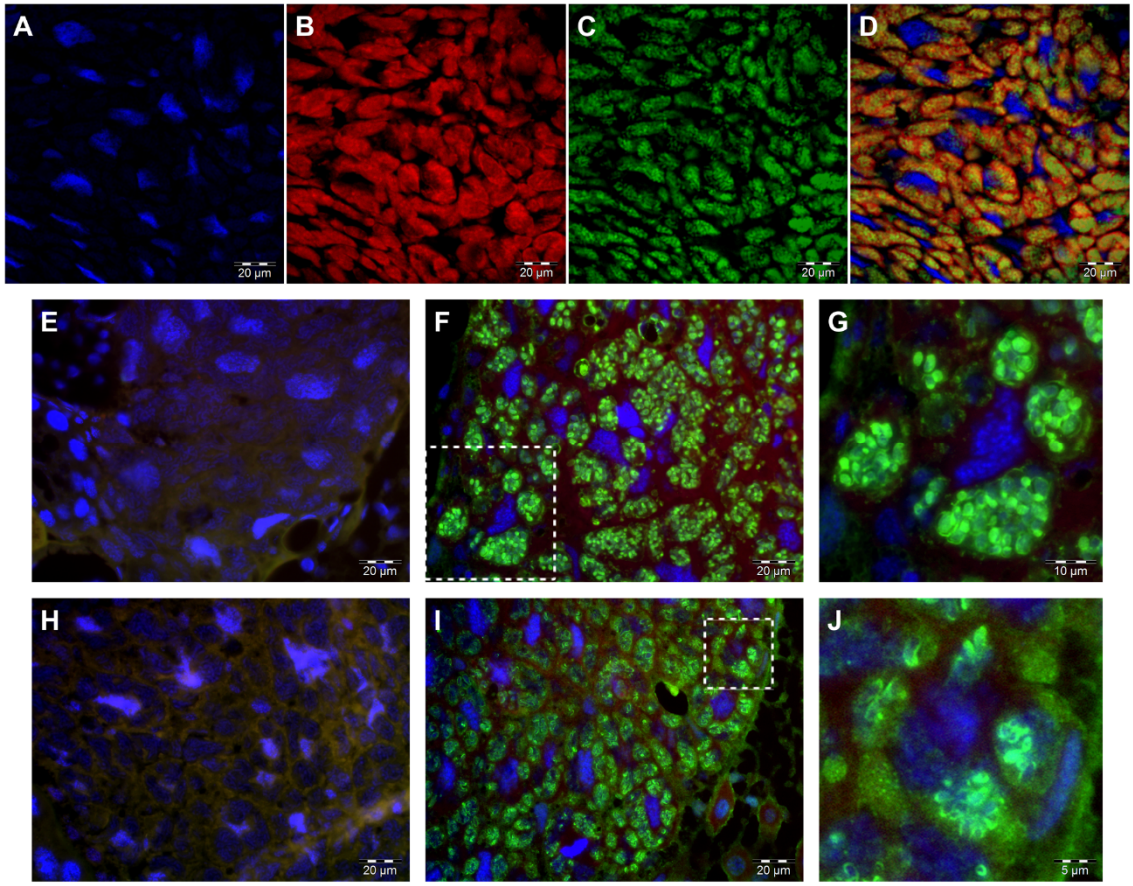


Figure 1. Microscopic analysis of *P. citri* PCVAL bacteriocytes showing the distribution of *M. endobia*'s GroEL and MscL. (A-D) FISH detection of *M. endobia* and *T. princeps* cells in *P. citri* bacteriocytes. Bacteriocyte nuclear genome is stained with DAPI (A, blue), *T. princeps* with probe b91 (B, red), *M. endobia* with probe g630 (C, green). Panel D shows the combination of all staining. (E-J) Immuno-histochemistry detection of GroEL (E-G) and MscL (H-J). Bacteriocyte nuclear genome is stained with DAPI (blue). Images G and J are amplifications of interesting details (in squares) from pictures F and H, respectively. No specific fluorescent signal was observed when serum of unimmunized rabbits was used (negative controls, E and H).

Location of MscL was approached using polyclonal antibodies specifically generated against the *M. endobia* protein. Again, the protein was only present in detectable levels in the *M. endobia*

cells (Figure 1, panels I, J). In this case, and in coherence with its expected subcellular location, the shiner signal was recurrently detected only in *M. endobia* cell membranes. A few degenerating cells show some MscL staining in the *T. princeps* cytoplasm only, which suggests that the protein is not performing its functional role at the membrane in this bacterium. Thus, sporadic cell lysis might occur, but the pattern of distribution of both GroEL and MscL appears to indicate that there is not a massive supply of *M. endobia* proteins to the *T. princeps* cytoplasm, as it would be expected if such supply were accomplished mainly by cell lysis.

CONCLUSIONS

The results of immuno-histochemical assays do not provide any evidence supporting constitutive and massive supply of *M. endobia* proteins to *T. princeps*. On the contrary, they indicate that (at least) the tested proteins accumulate only in *M. endobia* cells. A comprehensive scanning of *M. endobia* proteins that could potentially be targeted for exportation through the conserved Sec translocon machinery did not explain the intense protein traffic predicted to occur between both endosymbiotic bacteria. Neither putatively exported-protein adaptations, nor already described changes allowing abnormal permeability of the Sec translocon machinery for proteins without SP could be found. Although controlled and/or sporadic cell lysis can occur, the role of passive and non-specific communication gates, such as transient perforation of the plasma membrane and membrane MscL channels, both controlled by osmotic stress, can provide alternative explanations for the *M. endobia*-*T. princeps* molecular communication. Putatively recurrent osmotic stress events could result from the extremely biased distribution of metabolic capabilities between both endosymbionts. The unusual composition of the outer membrane lipopolysaccharides, in combination with the modulation of the strength of the peptidoglycan mesh controlled by the host, might also help on the provision of essential proteins from *M. endobia* to the *T. princeps* cytoplasm.

MATERIALS AND METHODS

Insects collection and sample preparation

Adult females were sampled from a laboratory population of *P. citri* which was reared on pumpkins at room temperature. Insects were decapitated and placed into 4% paraformaldehyde for fixation. Samples were stored at 4ºC in phosphate buffered saline (PBS) with 0.05% azide until their preparation for paraffin inclusion. To do so, samples were dehydrated through a graded ethanol series, from ethanol 70º to absolute ethanol, and washed twice in butanol at room temperature for 30 minutes. Then, they were embedded in paraffin and cut on a microtome at 3–5 μm thick sections that were placed on poly-lysine coated slides, air dried and kept at 4ºC until experiments performance. Prior usage, paraffin sections were dewaxed in two methylcyclohexan baths followed by two absolute ethanol baths of 10 minutes each.

Fluorescence *In Situ* Hybridization (FISH), immunostaining and microscopy

Samples devoted to FISH were coated with a few drops of 70% acetic acid while incubated in a 60ºC hotplate during 1 minute, in order to permeabilize cellular membranes. Once rinsed with PBS, they were dehydrated again through a graded ethanol series and air-dried. Then, samples were incubated in hydrochloride acid 0.01N with pepsin 0.1 mg/ml during 10 minutes in a 37ºC waterbath for deproteinization, rinsed again with PBS, and dehydrated in ethanol. Prehybridization was performed for 30 minutes at 45ºC in a buffer composed of 79% hybridization buffer (NaCl 0.9M; Tris 20 mM; EDTA 5 mM, in water, pH 7.2), 20% Denhardt solution (Ficoll 5 g; PVP 5 g; Bovine Serum Albumin 5 g in 500 ml water) and 1% SDS. Sections were subsequently coated by 100 ml of prehybridization buffer plus 1 mg of the corresponding labeled probe and incubated 3 hours at 45ºC. *T. princeps* and *M. endobia* specific detection was performed with probe b91 (5'-GCCTTAGCCCGTGCTGCCGTAC-3', TAMRA labeled) and probe g630 (5'-CGAGACTCTAGCCTATCAGTTTC-3', 6FAM labeled), respectively [10]. In order to preserve fluorescent signal, slides were kept in dark from this point on. Then, they were rinsed twice in PBS with SDS 0.1% at 45ºC, and at room temperature in PBS and water. Once completely airdried, slides were mounted with an aqueous mounting media made of Gel Mount and 3 μg/ml DAPI.

Samples devoted to immunostaining were rehydrated though an ethanol gradient to PBS, permeabilized with Triton-100 0.1% for 10 minutes at room temperature, washed with PBS, and incubated with 1% bovine serum albumin (BSA) in PBS during 30 minutes prior to primary antibody incubation at 4ºC overnight. Detection of GroEL was performed with rabbit polyclonal serum raised against the *B. aphidicola* APS homologous protein [41]. A rabbit polyclonal serum against MscL *M. endobia* was generated by Covalab (Villeurbanne, France), using the peptides C-KQFSWVLKPAQGNR (amino acids 55–68) and C-HNKEEEETPNELSKQS (amino acids 103–118) that correspond to a periplasmic and a cytoplasmic epitope respectively, according to an alignment with the *E. coli* homolog protein [42]. MscL antiserum was immunopurified using an agarose column coupled with the peptides used for immunization. Unimmunized rabbit serum was used as negative control. Antisera were diluted at 1:500 (GroEL) or 1:200 (MscL) in PBS with 0.1% BSA for primary antiserum incubation. Then, sections were washed with PBS containing 0.2% Tween, and incubated for 1 hour with donkey anti-rabbit secondary antibodies (1:600) labeled with Alexa fluor 488 for primary antibodies detection. In order to preserve secondary antibody fluorescence, subsequent PBS-Tween washing and rinses with both PBS and tap water were conducted in darkness. The samples were completely air-dried and mounted as previously described for FISH experiments.

Both FISH and immunostaining slides were observed under an epifluorescence microscope (Olympus IX81) using filters HQ535/50, D470/40 and HQ610/75 for green (FISH with probe g630 and immunostaining), blue (DAPI) and red (FISH with probe b91) signals. Cell F Software (AnalySIS) was used for image capturing and processing.

Sequence data analysis

SP screening was performed using both PRED-TAT [24] and SignalPv4.1 [25] softwares. *M. endobia* PCVAL complete proteome (CP003881) was scanned in order to detect SP motifs at N-terminal region of its 411 predicted CDS. In order to increase prediction accuracy, *E. coli* and *S. glossinidius* homologs to *M. endobia* candidate proteins were also scanned with both softwares. SP predictions were additionally contrasted with information on the corresponding *E. coli* proteins available at EcoCyc v16.5 (http://ecocyc.org/) [26]. Protein molecular weights were estimated with the Compute pI/Mw tool from ExPASy (http://web.expasy.org/compute_pi/) [43]. Pairwise alignments were performed by ClustalW [44].

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: RG SLM AL AH AM. Performed the experiments: SLM SB. Analyzed the data: SLM RG AH. Contributed reagents/materials/analysis tools: AL AM AH RG SLM. Wrote the paper: SLM RG AL.

ACKNOWLEDGMENTS

We thank Vanesa Martínez and Jorge Fuentes for their valuable advice regarding insect samples preparation for microscopy.

FUNDING

Financial support was provided by grants BFU2012-39816-C02-01 (Ministerio de Economía y Competitividad, Spain; http://www.mineco.gob.es/) to A. Latorre, Prometeo/2009/092 (Conselleria d'Educació, Generalitat Valenciana, Spain; http://www.cece.gva.es/es/) to A. Moya, and ANR-2010-BLAN-170101 (ImmunSymbArt; http://www.agence-nationale-recherche.fr/en/) to A. Heddi. S. López-Madrigal is a recipient of a fellowship from the Ministerio de Educación (Spain; http://www.mecd.gob.es/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Baumann P (2005) Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol 59: 155-189.

2. Moya A, Pereto J, Gil R, Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9: 218-229.

3. McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10: 13-26.

4. Russell JA, Latorre A, Sabater-Muñoz B, Moya A, Moran NA (2003) Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol Ecol 12: 1061-1075.

5. Gil R, Latorre A, Moya A (2010) Evolution of prokaryote-animal symbiosis from a genomics perspective. In: Hackstein JHP, editor. (Endo)symbiotic methanogenic archaea: Springer-Verlag. pp. 207-233.

6. Wu D, Daugherty SC, van Aken SE, Pai GH, Watkins KL, et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol 4: e188.

7. McCutcheon JP, McDonald BR, Moran NA (2009) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci USA 106: 15394-15399.

8. McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 MY of evolution. Genome Biol Evol 2: 708-718.

9. Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, et al. (2011) *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. PLoS Genet 7: e1002357.

10. von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. Nature 412: 433-436.

11. Gatehouse LN, Sutherland P, Forgie SA, Kaji R, Christeller JT Molecular and histological characterization of primary (betaproteobacteria) and secondary (gammaproteobacteria) endosymbionts of three mealybug species. Appl Environ Microbiol 78: 1187-1197.

12. Kono M, Koga R, Shimada M, Fukatsu T (2008) Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. Appl Environ Microbiol 74: 4175-4184.

13. Thao ML, Gullan PJ, Baumann P (2002) Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl Environ Microbiol 68: 3190-3197.

14. McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr Biol 21: 1366-1372.

15. Lopez-Madrigal S, Latorre A, Porcar M, Moya A, Gil R (2011) Complete genome sequence of '*Candidatus* Tremblaya princeps' strain PCVAL, an intriguing translational machine below the living-cell status. J Bacteriol 193: 5587-5588.

16. Lopez-Madrigal S, Latorre A, Porcar M, Moya A, Gil R (2013) Mealybugs nested endosymbiosis: going into the 'matryoshka' system in *Planococcus citri* in depth. BMC Microbiol 13: 74.1-74.12.

17. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, et al. (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153: 1567-1578.

18. Sukharev S (1999) Mechanosensitive channels in bacteria as membrane tension reporters. FASEB J 13 Suppl: S55-61.

19. Oakley AJ, Martinac B, Wilce MC (1999) Structure and function of the bacterial mechanosensitive channel of large conductance. Protein Sci 8: 1915-1921.

20. Pop OI, Soprova Z, Koningstein G, Scheffers DJ, van Ulsen P, et al. (2009) YidC is required for the assembly of the MscL homopentameric pore. FEBS J 276: 4891-4899.

21. van den Bogaart G, Krasnikov V, Poolman B (2007) Dual-color fluorescence-burst analysis to probe protein efflux through the mechanosensitive channel MscL. Biophys J 92: 1233-1240.

22. Ajouz B, Berrier C, Garrigues A, Besnard M, Ghazi A (1998) Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells. J Biol Chem 273: 26670-26674.

23. Berrier C, Garrigues A, Richarme G, Ghazi A (2000) Elongation factor Tu and DnaK are transferred from the cytoplasm to the periplasm of *Escherichia coli* during osmotic downshock presumably via the mechanosensitive channel *msc*L. J Bacteriol 182: 248-251.

24. Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD (2010) Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 26: 2811-2817.

25. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.

26. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, et al. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res 39: D583-590.

27. Khatib K, Belin D (2002) A novel class of *sec*A alleles that exert a signal-sequence-dependent effect on protein export in *Escherichia coli*. Genetics 162: 1031-1043.

28. Smith MA, Clemons WM, Jr., DeMars CJ, Flower AM (2005) Modeling the effects of *prl* mutations on the *Escherichia coli* SecY complex. J Bacteriol 187: 6454-6465.

29. Fikes JD, Bassford PJ, Jr. (1989) Novel *sec*A alleles improve export of maltose-binding protein synthesized with a defective signal peptide. J Bacteriol 171: 402-409.

30. Huie JL, Silhavy TJ (1995) Suppression of signal sequence defects and azide resistance in *Escherichia coli* commonly result from the same mutations in *sec*A. J Bacteriol 177: 3518-3526.

31. Bost S, Belin D (1995) A new genetic selection identifies essential residues in SecG, a component of the *Escherichia coli* protein export machinery. EMBO J 14: 4412-4421.

32. Berrier C, Coulombe A, Szabo I, Zoratti M, Ghazi A (1992) Gadolinium ion inhibits loss of metabolites induced by osmotic shock and large stretch-activated channels in bacteria. Eur J Biochem 206: 559-565.

33. el Yaagoubi A, Kohiyama M, Richarme G (1994) Localization of DnaK (chaperone 70) from *Escherichia coli* in an osmotic-shock-sensitive compartment of the cytoplasm. J Bacteriol 176: 7074-7078.

34. Jacobson GR, Rosenbusch JP (1976) Abundance and membrane association of elongation factor Tu in *E. coli*. Nature 261: 23-26.

35. Hantash FM, Ammerlaan M, Earhart CF (1997) Enterobactin synthase polypeptides of *Escherichia coli* are present in an osmotic-shock-sensitive cytoplasmic locality. Microbiology 143 (Pt 1): 147-156.

36. Vazquez-Laslop N, Lee H, Hu R, Neyfakh AA (2001) Molecular sieve mechanism of selective release of cytoplasmic proteins by osmotically shocked *Escherichia coli*. J Bacteriol 183: 2399-2404.

37. Meredith TC, Aggarwal P, Mamat U, Lindner B, Woodard RW (2006) Redefining the requisite lipopolysaccharide structure in *Escherichia coli*. ACS Chem Biol 1: 33-42.

38. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 122: 209-220.

39. Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E (2002) Endosymbiotic bacteria: *gro*EL buffers against deleterious mutations. Nature 417: 398.

40. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668-673.

41. Hara E, Fukatsu T, Ishikawa H (1990) Characterization of symbionin with anti-symbionin antiserum. Insect Biochemistry 20: 429-436.

42. Moe PC, Blount P, Kung C (1998) Functional and structural conservation in the mechanosensitive channel MscL implicates elements crucial for mechanosensation. Mol Microbiol 28: 583-592.

43. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. The Proteomics protocols handbook: Humana Press. pp. 571-607.

44. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

SUPPLEMENTARY MATERIAL

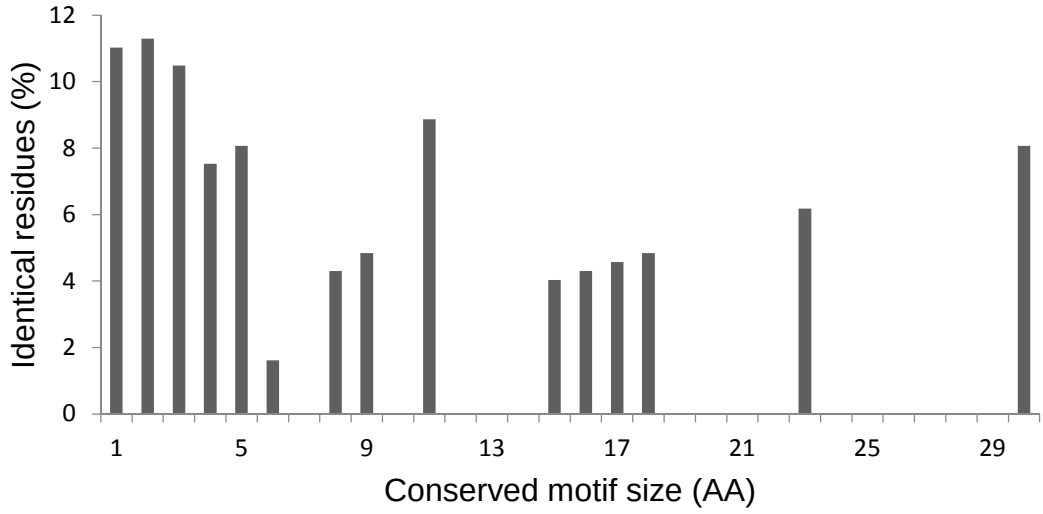


Figure S1. Distribution of both *T. princeps* and *M. endobia* GroEL identical residues. Conservation clusters with variable sizes along GroEL alignment were taken into account in order to evaluate protein similarities in terms of immuno-histochemistry detection. Identical orthologous residues got distributed among clusters with 1–6, 8, 9, 11, 15–18, 23 and 30 residues in length.

Table S1. Analysis of known critical residues at proteins of the Sec translocon in *M. endobia*. Well characterized mutational changes known to yield *prl* phenotypes in *E. coli* have been considered. Studied residues are ordered according to their position in the corresponding proteins, from N to C-terminus. Superindex denotes co-existing mutations in double mutant strains (*prlA4*, *prlA6*, *prlA7* and *prlA11* for SecY, and *prlD21* for SecA) and the deletion of two adjacent codons (*prlG8* for SecE).

Gene Product	Amino acid	Mutant alleles	Mutation details	*M. endobia* status	*E. coli* codon	*M. endobia* codon
SecY	37	*prlA8911*	Ser → Phe	38Ser	TCT	TCG
	64	*prlA300*	Phe → Cys	65Phe	TTT	TTT
	65	*prlA8914*	Asn → Tyr	66Asn	AAC	AAT
	67	*prlA3*	Phe → Cys	68Phe	TTC	TTC
		prlA666	Phe → Ser			
	68	*prlA726*	Ser → Pro	69Ser	TCT	TCT
		prlA799	Ser → Leu			
		prlA8913	Ser → Phe			
	69	*prlA9*	Gly → Asp	70Gly	GGT	GGT
		prlA205	Gly → Cys			
	71	*prlA302*	Ala → Asp	72Ala	GCT	GCA
	73	*prlA306*	ΔSer	74Ser	AGC	AGC
	90	*prlA304*	Ile → Asn	91Ile	ATC	ATC
	188	*prlA6*	Ser → Leu[a]	189Ser	TCA	TCC
	191	*prlA200*	Ile → Ser	192Ile	ATC	ATT
	274	*prlA1*	Val → Gly	276Val	GTA	GTT
	277	*prlA7*	Ala → Glu[b]	279Thr	GCA	ACA
	278	*prlA202*	Ile → Ser	280Ile	ATC	ATT
		prlA208	Ile → Asn			
		prlA303	Ile → Thr			
	282	*prlA401*	Ser → Arg	284Ser	AGT	AGC
	286	*prlA4*	Phe → Tyr[c]	288Phe	TTC	TTT
	407	*prlA7, 11* and *301*	Leu → Arg[b,d]	409Leu	CTT	TTG
	408	*prlA4* and *6*	Ile → Asn[a,c]	410Ile	ATC	ATA
	411	*prlA11*	Val → Gly[d]	413Val	GTC	GTG
SecE	105	*prlG2*	Ser → Pro	106Ser	TCA	TCA
	108	*prlG1*	Leu → Arg	109Leu	CTG	CTG
	116	*prlG8*	ΔVal[e]	117Val	GTT	GTT
	117	*prlG8*	ΔArg[e]	118Arg	CGC	CGT
	120	*prlG3*	Ser → Phe	121Ser	TCC	TCA
SecG	25	*prlH2*	Gly → Asp	25Gly	GGT	GGT
	26	*prlH1*	Lys → Glu	26Lys	AAA	AAA
	50	*prlH3*	Asn → Tyr	50Thr	AAC	ACT
		prlH4	Asn → Ile			
	57	*prlH6*	Ala → Glu	57Ala	GCG	GCA
	65	*prlH5*	Ile → Thr	65Val	ATC	GTG
SecA	111	*prlD4*	Thr → Asn	111Thr	ACC	ACT
	134	*prlD22*	Tyr → Cys	134Tyr	TAC	TAC
		azi-17	Tyr → Asn			
		prlD21 and *23*	Tyr → Ser[f]			
	148	*prlD21*	Glu → Lys[f]	148Asp	GAA	GAT
	373	*prlD5*	Ala → Val	373Ala	GCT	GCG
	484	*prlD43*	His → Gln	484His	CAC	CAC
	488	*prlD2* and *3*	Ala → Val	488Ala	GCG	GCA
	507	*prlD20, 24, 26-28, 31*	Ala → Val	507Ala	GGC	GGC
	645	*azi-4*	Leu → Gln	645Leu	CTG	CTA

[a]Mutations in double mutant *prl*A6; [b]Mutations in double mutant *prl*A7; [c]Mutations in double mutant *prl*A4; [d]Mutations in double mutant *prl*A11; [e]Mutations in *prl*G8; [f]Mutations in double mutant *prl*D21

.

CHAPTER 4

Status: Published. August 26, 2014.

Citation: López-Madrigal S, Beltrà A, Resurrección S, Soto A, Latorre A, Moya A, Gil R. 2014. Molecular evidence for ongoing complementarity and horizontal gene transfer in endosymbiotic systems of mealybugs. Front. Microbiol. 5:449.

Molecular evidence for ongoing complementarity and horizontal gene transfer in endosymbiotic systems of mealybugs

Sergio López-Madrigal,[1] Aleixandre Beltrà,[2] Serena Resurrección,[1] Antonia Soto,[2] Amparo Latorre,[1,3] Andrés Moya,[1,3] and Rosario Gil[1,*]

[1] Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Valencia, Spain

[2] Instituto Agroforestal Mediterráneo, Universitat Politècnica de València, Valencia, Spain

[3] Área de Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO)–Salud Pública, Valencia, Spain

ABSTRACT

Intracellular bacterial supply of essential amino acids is common among sap-feeding insects, thus complementing the scarcity of nitrogenous compounds in plant phloem. This is also the role of the two mealybug endosymbiotic systems whose genomes have been sequenced. In the nested endosymbiotic system from *Planococcus citri* (Pseudococcinae), "*Candidatus* Tremblaya princeps" and "*Candidatus* Moranella endobia" cooperate to synthesize essential amino acids, while in *Phenacoccus avenae* (Phenacoccinae) this function is performed by its single endosymbiont "*Candidatus* Tremblaya phenacola." However, little is known regarding the evolution of essential amino acid supplementation strategies in other mealybug systems. To address this knowledge gap, we screened for the presence of six selected *loci* involved in essential amino acid biosynthesis in five additional mealybug species. We found evidence of ongoing complementarity among endosymbionts from insects of subfamily Pseudococcinae, as well as horizontal gene transfer affecting endosymbionts from insects of family Phenacoccinae, providing a more comprehensive picture of the evolutionary history of these endosymbiotic systems. Additionally, we report two diagnostic motifs to help identify invasive mealybug species.

Keywords: mealybugs, endosymbiosis, "*Candidatus* Tremblaya princeps", "*Candidatus* Tremblaya phenacola", amino acid biosynthesis, horizontal gene transfer

INTRODUCTION

The establishment of permanent intracellular symbioses with bacteria has played a key role in insect evolution. Endosymbionts are located in specialized eukaryote cells (bacteriocytes) and complement the insect's limited heterotrophic metabolism with metabolic pathways for the biosynthesis of essential amino acids, fatty acids and/or vitamins (Baumann, 2005). Population dynamics imposed by their lifestyle, together with the stability and nutritional richness of the intracellular environment, trigger a number of genomic changes in the prokaryote symbiont. Such

evolutionary changes include drastic genome size reduction, due to the loss of genes rendered unnecessary for the association (i.e., those that are superfluous in a protected and stable intracellular niche or whose function can be provided by the host), and an increase in AT content in most analyzed cases (Baumann, 2005; Moran et al., 2008; Moya et al., 2008; Gil et al., 2010). Eventually, if a second bacterium joins the association, both bacteria coevolve and the ongoing reductive genome process affects both of them, leading to two possible outcomes: either both bacteria become essential for the fitness of the association (complementation), or one bacterium undergoes an extreme genome degenerative process, which may end up in its extinction, while the remaining bacterium continues the reductive process alone (replacement) (Moya et al., 2009).

As in other phloem-feeding insects, mealybugs rely on their endosymbionts for the provision of essential amino acids (Baumann, 2005). This fact is supported by the recent genome sequencing of endosymbionts from two mealybug species belonging to subfamilies Pseudococcinae and Phenacoccinae (Lopez-Madrigal et al., 2011; McCutcheon and von Dohlen, 2011; Husnik et al., 2013). Mealybugs from these subfamilies present an intricate variety of endosymbiotic relationships that reflect both complementation and replacement events. Phylogenetic studies suggest that a betaproteobacterial ancestor of "*Ca*. Tremblaya" infected a mealybug ancestor before the split of subfamilies Phenacoccinae and Pseudococcinae (Hardy et al., 2008). Later on, except for the *Ferrisia* and *Maconellicoccus* clades (where no additional endosymbiont has been reported), the ancestor of "*Ca*. Tremblaya princeps" was infected multiple times by different gammaproteobacteria, establishing a diversity of stable nested endosymbiotic consortia, with each "*Ca*. Tremblaya princeps" cell containing several cells of the corresponding gammaproteobacterium (Thao et al., 2002; Gatehouse et al., 2012). By contrast, "*Ca*. Tremblaya phenacola" remained alone in subfamily Phenacoccinae, except in several clades (the tribe *Rhizoecini* and the genus *Rastrococcus*) where it was replaced by different *Bacteroidetes* (Gruwell et al., 2010; Husnik et al., 2013).

The most widely studied mealybug endosymbiotic system belongs to *Planococcus citri* (Risso), where "*Ca*. Tremblaya princeps" harbors "*Ca*. Moranella endobia" (McCutcheon and von Dohlen, 2011), with a tight relationship between the nested endosymbiosis dynamics and the insect life-cycle (von Dohlen et al., 2001; Kono et al., 2008). Independent genomic studies on the endosymbiotic consortium of two *P. citri* strains (PCIT and PCVAL) revealed their entangled metabolic complementation for the biosynthesis of some essential amino acids (McCutcheon and von Dohlen, 2011; Lopez-Madrigal et al., 2013), and necessary participation of the insect host (Husnik et al., 2013). Husnik and coworkers also reported horizontal gene transfer (HGT) of some genes involved in these pathways from diverse bacteria to the insect nuclear genome. Even though "*Ca*. Tremblaya princeps" from *P. citri* exhibits one of the smallest prokaryote genomes known so far (139 kb), with an extremely reduced gene set, it is the source for at least 29 enzymes needed to synthesize several essential amino acids (Lopez-Madrigal et al., 2011; McCutcheon and von Dohlen, 2011). Functional redundancy in both endosymbiotic partners was observed only for *dap*A (involved in lysine biosynthesis) and *aro*K (involved in phenylalanine and tryptophan biosynthesis), an indication that each bacterium has adopted a specific role in essential amino

acid provision. Recent sequencing of "*Ca*. Tremblaya phenacola" PAVE, the sole endosymbiont of *Phenacoccus avenae* (Borchsenius), revealed a remarkable case of evolutionary convergence, since it has preserved exactly the same set of genes, collectively retained by "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" in *P. citri*, for supplying essential amino acids to the host (Husnik et al., 2013).

In order to study the evolution of essential amino acids provisioning within unexplored "*Ca*. Tremblaya" lineages, we have performed a genetic screening on several mealybug species, including members of both subfamily Pseudococcinae (*Dysmicoccus boninsis* Kuwana, *Pseudococcus viburni* Signoret and *Pseudococcus longispinus* Targioni-Tozzetti) and subfamily Phenacoccinae (*Phenacoccus peruvianus* Granara de Willink and *Phenacoccus madeirensis* Green). We searched for the presence of genes *arg*H, *ilv*D, *leu*B, *met*E, *thr*C, and *trp*B, encoding the enzymes that participate in the last steps performed by the *P. citri* and *P. avenae* endosymbiotic systems in the pathways for the biosynthesis of arginine, branched amino acids, methionine, threonine and tryptophan, respectively. We have also characterized the endosymbionts present in these mealybug species at the molecular and phylogenetic level. Our results reveal differences among different clades, both at the molecular and functional levels, thus providing a more complete picture on the complex evolutionary history of the two "*Ca*. Tremblaya" lineages.

MATERIALS AND METHODS

Insect sample collection and DNA extraction

Insects belonging to the species *D. boninsis*, *P. viburni*, *P. longispinus*, *P. peruvianus*, and *P. madeirensis* were field collected in Valencia (Spain) and stored in absolute ethanol at -20°C. Total DNA ($_T$DNA) extractions were performed with JETFLEX Genomic DNA Purification Kit (GENOMED) on 5–8 adult females.

DNA amplification, sequencing and analysis

PCR amplifications were performed on insect $_T$DNA, with appropriate primer pairs (see Section Gene Screening), using 50–60 µmol of each primer/50 µl reaction, with KAPATaq DNA Polymerase Kit (Kapa Biosystems). The thermal cycling protocol was as follows: an initial denaturation at 95°C for 5 min, followed by 35 cycles of 50 s at 95°C, 40 s at 56°C (or 52°C when indicated), and 2 min at 72°C, plus a final extension step of 7 min at 72°C. When needed, amplicons were cloned using pGEM-T Easy Vector System I Kit (Promega). ABI sequencing was performed, using specific or vector primers T7 and SP6, at the sequencing facility of the Universitat de València. Sequencing reads were quality surveyed and assembled with Staden Package (http://staden.sourceforge.net/; Staden et al., 2000). Artemis software was used for sequence data management (http://www.sanger.ac.uk/resources/software/artemis/; Rutherford

et al., 2000). Multiple alignments were performed with ClustalW (Larkin et al., 2007). MEGA5 was used for the calculation of both p-distance and nucleotide composition.

Gene screening

Complete sequences of *arg*H (encoding argininosuccinate lyase, EC 4.3.2.1, involved in arginine biosyntesis), *ilv*D (encoding dihydroxy-acid dehydratase, EC 4.2.1.9, involved in isoleucine and valine biosynthesis), *leu*B (encoding 3-isopropylmalate dehydrogenase, EC 1.1.1.85, involved in leucine biosynthesis), *met*E (encoding cobalamin-independent homocysteine transmethylase, EC 2.1.1.14, involved in methionine biosynthesis), *thr*C (encoding threonine synthase, EC 4.2.3.1, involved in threonine biosynthesis), and *trpB* (encoding the beta subunit of tryptophan synthase, EC 4.2.1.20, involved in tryptophan biosynthesis), were retrieved from GenBank for a set of selected beta and gammaproteobacteria (Table S1 in Supplementary Material). Multiple alignments were performed in ClustalW to allow the design of degenerate primers (Table 1) to amplify the corresponding gene by PCR, using both 52°C and 56°C as annealing temperature (Ta) for all primer pairs. ᴛDNA from *P. citri* was used as a positive control. At least 10 clones for each amplicon were sequenced. BLAST searches against the non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi/; Altschul et al., 1997) were performed in order to identify the putative taxonomic origin of the sequences obtained.

Table 1. Degenerate primers used in the gene screening.

Gene	Primer	Sequence (5´→3´)	Degeneracy* (%)	Conserved Motif
argH	argH-F	AAYGAYCARRTNGCNACNGA	35	NDQ(V/I)ATD
	argH-R	TCNGGRTTYTTYTTYTGNGGCAT	26	MPQKKNPD
ilvD	ilvD-F	ATGTWYACVGCNAAYWCNATG	33	M(F/Y)TAN(S/T)M
	ilvD-R	GARAARAANCKVCCRTCNGT	35	SGGT(S/W)G
leuB	leuB-F	GGNGAYGGHATHGGBCCBGA	30	GDGIGPE
	leuB-R	TCVGAVARDATRTCGCCRAA	30	FGDILSD
metE	metE-F	AACTAYCACTAYMTVGTVCCNGA	26	NYHY(M/I/L)VPE
	metE-R	CCRCARTCHGGRTTNAYCCA	30	W(V/I)NPDCG
thrC	thrC-F	GCDACNTCNGGBGAYACNGG	30	ATSGDTG
	thrC-R	TYNCCRAARTTNCCNSWNGG	45	PSGNFG(D/N)
trpB	trpB-F	GTNHTNGGNCARGCNYTNYTNGC	43	V(L/I)GQALLA
	trpB-R	TCNCCNCGNCCNGANAGRTT	30	NLSGRGD

*Primer degeneracy was measured as the proportion of ambiguous sites.

ᴛDNA from *D. boninsis*, *P. longispinus*, *P. peruvianus*, and *P. madeirensis* were used as templates for PCR amplification using 16S rDNA universal primers (Table S2 in Supplementary Material; van Ham et al., 1997). The amplicons were cloned as above mentioned, and at least 25 clones were sequenced for each species. The newly found sequences have been deposited in the GenBank database (see Table S3 in Supplementary Material). Samples from *P. peruvianus* and *P. madeirensis* were also analyzed by PCR using gammaproteobacteria-specific 16S rDNA primers (Mühling et al., 2008).

In order to determine the location of the *trp*B gene in *P. peruvianus*, we dissected three adult females to separate the head from the rest of the body, and extracted ᴛDNA from both samples. Universal primers were used to amplify 18S rDNA from the host genome (Ta = 52°C), as PCR positive control (Table S2 in Supplementary Material; Littlewood and Olson, 2001). We designed two sets of specific primers to amplify 16S rDNA from "*Ca*. Tremblaya phenacola" and the *trp*B sequence found in *P. peruvianus*, respectively. All PCR products were sequenced using the same primers to confirm their identity.

A diagnostic screening by restriction enzyme analysis was performed on the 16S rRNA genes amplified from *P. longispinus*, in order to check for the putative presence of other gammaproteobacterial haplotypes previously identified in this species (Duron et al., 2008; Rosenblueth et al., 2012). We designed a pair of primers in a conserved region of the 16S rDNA sequences of the three gammaproteobacterial haplotypes and "*Ca*. Tremblaya princeps" from *P. longispinus* (Table S2 in Supplementary Material). RFLP-up and RFLP-down amplify the region from sites 516 to 1075 in the *E. coli* K-12 substr. MG1655 homolog. PCR products were digested with the enzyme RsaI (Roche). Restriction digest products were run on an agarose gel, stained in ethidium bromide, and visualized with UV light.

Fluorescence *In Situ* Hybridization (FISH)

Field collected *P. peruvianus* adult females were decapitated and placed in 4% paraformaldehyde for fixation. Samples were stored at 4°C in phosphate buffered saline (PBS) with 0.05% azide until preparation for paraffin inclusion. To do so, samples were dehydrated through a graded ethanol series, from 70 to 100% ethanol, and finally washed twice in xylene at room temperature for 30 min. Then, they were embedded in paraffin and cut on a microtome into 5 μm thick sections, placed on poly-lysine coated slides, air dried, and kept at 4°C. Prior to usage, paraffin sections were dewaxed in two xylene baths, followed by two absolute ethanol baths of 10 min each.

In order to permeabilize cellular membranes, samples were coated with a few drops of 70% acetic acid while incubated on a 60°C hotplate for 1 min. Once rinsed with PBS, they were dehydrated again through a graded ethanol series and air-dried. Slides were subsequently coated with 100 μl of hybridization buffer (Tris 20 mM pH 8, NaCl 0.9 M, SDS 0.01% and formamide 30%) plus 100 ng of the 16S rDNA universal probe Cy5-EUB338 (Amann et al., 1990) or the specific probe Cy3-TphPPER1290 (5'-CCGCAATTCGTACTGAGGTTAGG-3'), designed on "*Ca*. Tremblaya phenacola" PPER 16S rRNA gene, and incubated for 3 h at 45°C. To confirm hybridization signal specificity, the following control experiments were performed: a no-probe control, an RNase digestion control (slides treated for 30 min with RNase A prior to hybridization), and a competitive suppression control with excess unlabeled probe (80 ng/μl of hybridization buffer) (Fukatsu et al., 1998). In order to preserve fluorescent signal, slides were kept in dark from this point on. After hybridization, they were coated with 1 μg/mL DAPI for 10 min. Four 10-min washes were performed with washing buffer (Tris 20 mM pH 8, NaCl 112 mM, SDS 0.01% and

EDTA 5 mM) at 48°C. Finally, samples were rinsed twice with milliQ water for 5 min at room temperature. Once completely air-dried, slides were mounted with Fluoromount-G (Southern Biotech) and kept at 4°C overnight.

Slides were observed under an epifluorescence microscope (Nikon Eclipse 80i). Nikon DS-Qi1Mc digital camera and NIS-Elements BR 3.0 software were used for image capturing and processing, respectively.

Phylogenetic analyses

Nucleotide sequences used for the phylogenetic analysis were retrieved from GenBank or obtained in this work. The complete list of sequences, from selected alpha, beta and gammaproteobacteria (including both endosymbiotic and free-living species) is presented in Table S3 in Supplementary Material.

Phylogenetic reconstructions were carried out by Maximun Likelihood (ML), Maximum Parsimony (MP) and Bayesian methods, in RAxML (Stamatakis, 2006), DNAPARS from PHYLIP v3.69 package (Felsenstein, 2005), and MrBayes 3.2 (Ronquist et al., 2012), respectively. According to JModelTest (Guindon and Gascuel, 2003; Darriba et al., 2012), we applied a separate general time-reversible evolutionary model with gamma-distributed rates and a proportion of invariant sites (GTR+I+G) in phylogenetic reconstructions by ML and Bayesian methods. In ML and MP reconstructions, bootstrap analyses were performed with 1000 replications. In Bayesian reconstructions, phylogenetic trees were generated from two runs of 200,000 generations for 16S rDNA and two runs of 500,000 generations for *trp*B and *trp*B-*arg*H. Likelihood settings were set to nst=6, rates=invgamma and ngammacat=4. Sampling was performed every 100 generations. First 2300, 3200, and 3500 generations were discarded as "burn in" for runs on *trp*B-*arg*H, 16S rDNA and *trp*B molecular data, respectively. Figures on phylogenetic analysis were prepared with FigTree v1.4.0 software (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

Screening of genes involved in essential amino acids biosynthesis

The putative capability for essential amino acid biosynthesis among the analyzed species was evaluated by a PCR screening of selected genes involved in the last step usually performed by the mealybug endosymbiotic systems in the biosynthetic pathways of most essential amino acids. We did not attempt to detect genes involved in the biosynthesis of histidine and phenylalanine given the impracticality of designing reliable degenerate primers on genes *his*B and *phe*A, due to their considerably smaller length and/or conservation level. The results of this genetic screening are shown in Figure 1.

*arg*H of gammaproteobacterial origin, according to BLAST results, was found in all analyzed Pseudococcinae lineages. Although "*Ca*. Tremblaya princeps" from *P. citri* also contains an *arg*H homolog, it is pseudogenized (Lopez-Madrigal et al., 2011; McCutcheon and von Dohlen, 2011). *Locus* degeneration is evidenced by two deletions, involving 6 and 57 nucleotides, affecting a highly conserved region in analyzed betaproteobacteria and *E. coli* (gammaproteobacterium), as well as by an inactivating frameshift caused by a single cytosine deletion (see Supplementary Material). An *arg*H homolog of betaproteobacterial origin was also detected in *P. peruvianus*, but not in the *P. madeirensis* sample.

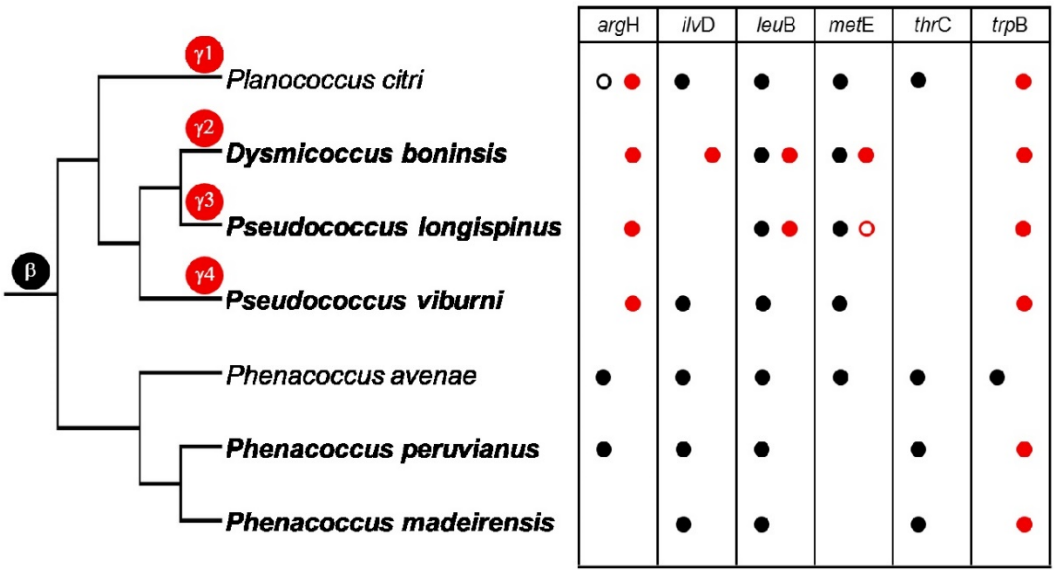


Figure 1. Genetic screening of selected *loci* involved in essential amino acid biosynthesis. The taxonomic assignation of the amplified sequences is indicated by black (betaproteobacteria) and red (gammaproteobacteria) circles. Empty circles represent pseudogenes. The species analyzed in this work appear in bold. *P. citri* and *P. avenae* endosymbionts are shown for comparison. Cladogram topology represents the evolutionary relationships between insect lineages (based on Hardy et al., 2008; Gruwell et al., 2010; this work). Marks on the branches indicate the infection by the ancestor of "*Ca*. Tremblaya" (β) and different lineages of gammaproteobacteria. γ1 corresponds to "*Ca*. Moranella endobia." γ2–4 correspond to non-monophyletic gammaproteobacteria, based in our phylogenetic analyses.

Regarding *ilv*D, *leu*B and *met*E, *P. viburni* resembles *P. citri*, since they both contain orthologs solely of betaproteobacterial origin. In contrast, only an *ilv*D gene of gammaproteobacterial origin was detected in *D. boninsis*, while no functional homolog was detected in *P. longispinus*. *Loci* *leu*B and *met*E were redundant both in *D. boninsis* and *P. longispinus*, although the gammaproteobacterial *met*E homolog in *P. longispinus* is apparently inactivated due to a non-sense mutation (TGG→TAG) affecting a highly conserved residue (W140 in the *E. coli* homolog protein). Nevertheless, all the other important residues for protein functioning examined are still preserved (Table S4 in Supplementary Material). MetE key residues identified in *E. coli* are

preserved in *Burkholderia*, the closest free-living betaproteobacterial relative of "*Ca*. Tremblaya," and they are also present in all homolog proteins of the different "*Ca*. Tremblaya princeps" strains analyzed. Only a non-synonymous Ile→Val change was observed in all of them. Nonetheless, this change is not expected to have functional consequences, given the similarities in size and polarity of both amino acids. However, this gene was not detected in the two Phenococcinae analyzed in this work. In the case of *thr*C, we could not detect this gene in other "*Ca*. Tremblaya princeps" except that from *P. citri*, while it was identified in the two sampled "*Ca*. Tremblaya phenacola". Finally, only a gammaproteobacterial homolog of *trp*B was found in all analyzed mealybugs.

As expected for mealybugs of subfamily Phenacoccinae, most of the amplified genes are of betaproteobacterial origin. The genetic screening performed on samples from *P. peruvianus* and *P. madeirensis* allowed the amplification of five and four out of the six screened *loci*, respectively (Figure 1). This is consistent with previous descriptions of Phenacoccinae mealybug endosymbiotic systems (Husnik et al., 2013; Koga et al., 2013), suggesting that "*Ca*. Tremblaya phenacola" is the only endosymbiont in these species too. Unexpectedly, and in contrast with its recently described homolog in "*Ca*. Tremblaya phenacola" PAVE (Husnik et al., 2013), the *trp*B homologs identified in both Phenacoccus samples have best similarity hits with gammaproteobacterial proteins. Both sequences are highly similar but not identical (p-distance=0.117), as expected for very closely-related orthologs. Additionally, the analysis of their nucleotide composition showed an AT-accumulation at codon degenerated positions (AT_{N3}=80.2%, contrasting with AT_{N1}=54.4% and AT_{N2}=58%), which is common among obligate endosymbionts. Both facts appear to discard possible DNA contamination. To confirm the gammaproteobacterial origin of such sequences we performed a phylogenetic analysis using 771 unambiguously aligned sites of the *trp*B gene from 33 different prokaryote lineages including free-living and endosymbiotic bacteria from classes Beta and Gammaproteobacteria (Figure 2). Due to the short length of the aligned sequences, the topology, in some cases, does not reproduce the natural clades. Nevertheless, the sequences obtained from *P. peruvianus* and *P. madeirensis* are clearly located in the gammaproteobacterial clade.

Determination of the location of the *trp*B gene in *Phenacoccus*

The amplification of a *trp*B gene of gammaproteobacterial origin might indicate that a second symbiont is also present in the two Phenacoccus species under study. To address this issue, we performed a 16S rRNA gene amplification on *P. peruvianus* and *P. madeirensis*. The PCR products were cloned, and 25 clones were sequenced, yielding a single sequence of betaproteobacterial origin from each mealybug species (1477 and 1467 bp, respectively). In order to search for the putative presence of low-density gammaproteobacteria in the tested samples, we performed a second PCR screening with gamma-specific primers. $_T$DNA from the Pseudococcinae species *P. citri*, *D. boninsis*, *P. viburni*, and *P. longispinus*, where gammaproteobacterial endosymbionts had been detected, were used as positive controls. No gammaproteobacteria could be detected on the two *Phenacoccus* samples (Figure 3)

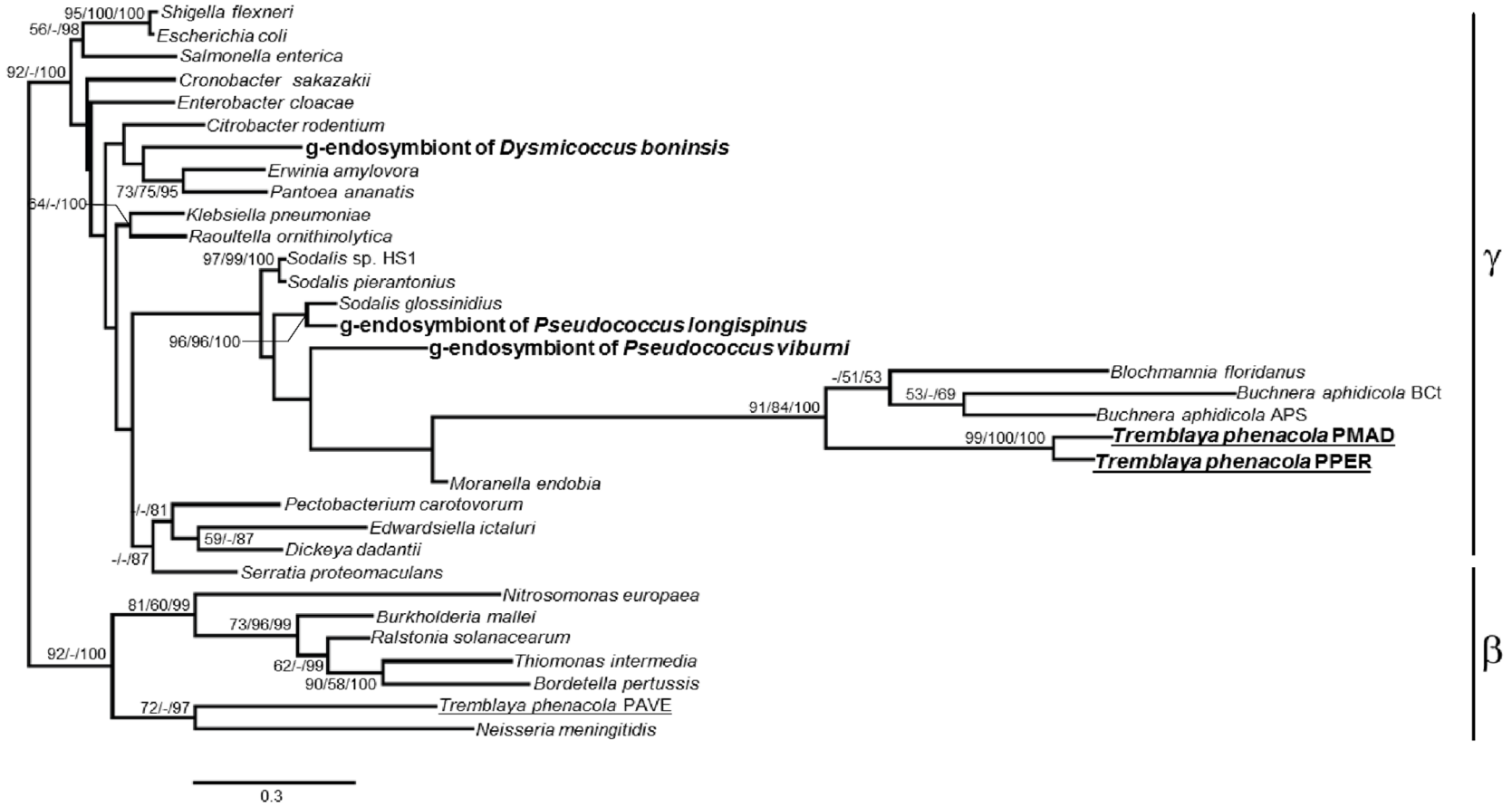


Figure 2. ML phylogenetic analysis of the *trp*B partial nucleotide sequences obtained from *P. peruvianus* and *P. madeirensis* samples. Sequences obtained in this work are in bold; those from the Phenacoccinae are underlined. Bayesian and MP analysis gave essentially the same results. ML and MP bootstrap values, and Bayesian posterior probabilities over 50% are represented. Scale bar represents substitutions per site.

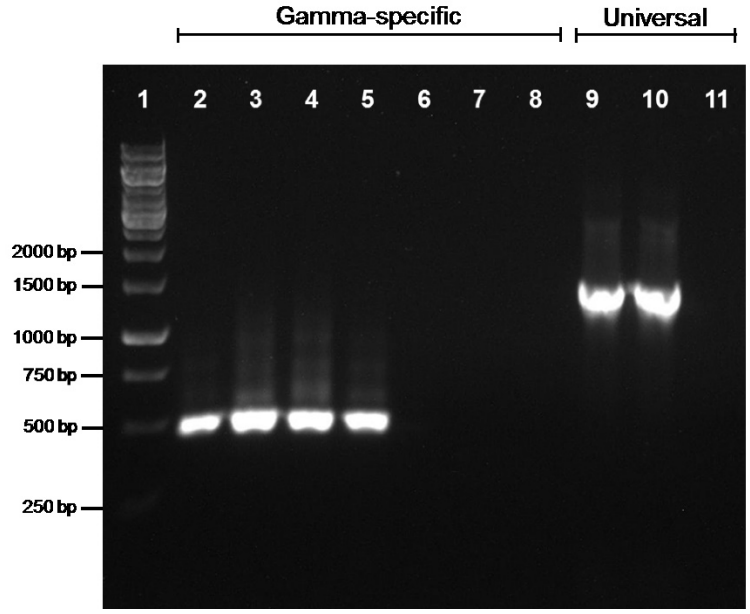


Figure 3. PCR screening for gammaproteobacterial endosymbionts. Gammaproteobacterial-specific primers were used on ᴛDNA from *P. citri* (lane 2), *D. boninsis* (lane 3), *P. viburni* (lane 4), *P. longispinus* (lane 5), *P. peruvianus* (lane 6), and *P. madeirensis* (lane 7). The quality of samples from *P. peruvianus* and *P. madeirensis* was tested using 16S rDNA universal primers (lanes 9 and 10, respectively). Lanes 8 and 11 are the results of negative controls for each pair of primers.

The absence of a second bacterium was further confirmed by FISH analysis of *P. peruvianus* adult females. As it can be seen on Figure 4, the bacteriome appears as a well-defined organ inside the insect body cavity. It is visible under DAPI staining because DNA is found both in the nucleus (insect genome) and the cytoplasm (bacterial genomes). The only bacteria detected are exclusively located in the bacteriome, and the same fluorescent pattern is observed using both a universal probe and a "*Ca*. Tremblaya phenacola" specific probe.

To ascertain the location of the amplified *trp*B gene, either in the nuclear or the bacterial genome, we also followed a PCR approach (see Materials and Methods; Figure 5). The 18S rRNA gene, used as a positive control, was amplified in both head and body samples, while no amplification of the 16S rRNA gene was obtained from the head sample, thus confirming that it was not contaminated with bacteriocytes. As for the 16S rRNA gene, *trp*B was only amplified in the body samples, an indication that it is not present in the nuclear genome. Since no other bacterium was detected, *trp*B is likely to be located in the "*Ca*. Tremblaya phenacola" PPER genome.

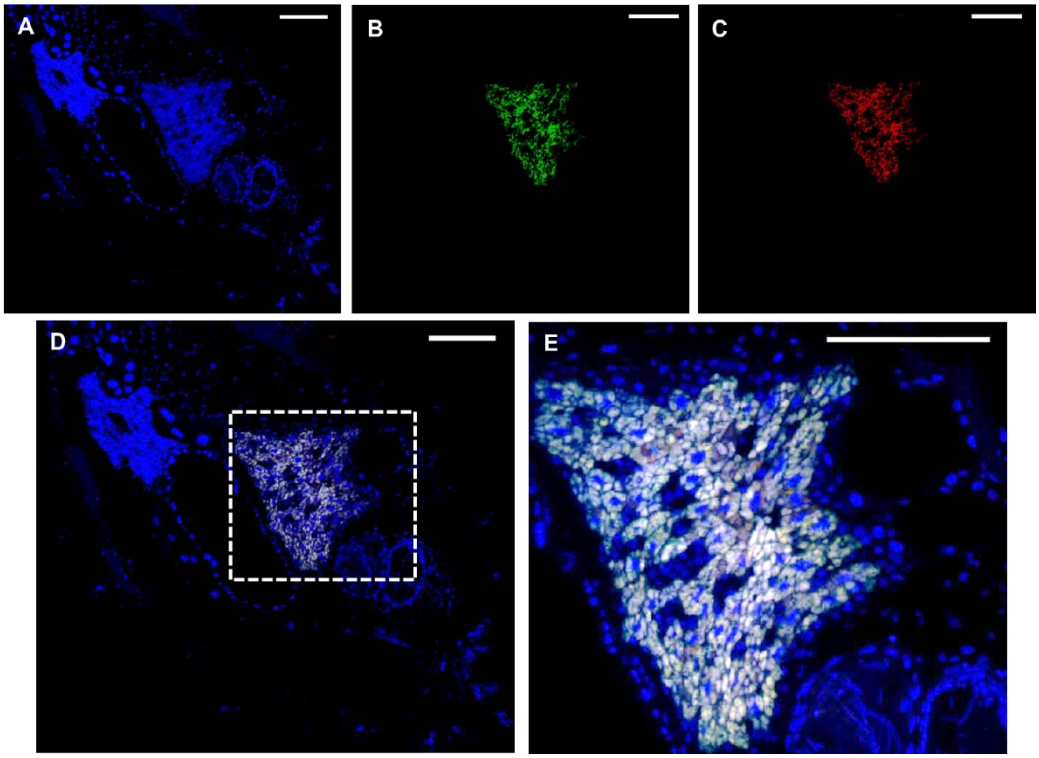


Figure 4. FISH analysis of *P. peruvianus* bacteriome. Adult female mealybugs stained with DAPI (blue: A, D, E) and probed with Cy5-EUB338 (green: B, D, E) and Cy3-TphPPER1290 (red: C–E). (A–D) Complete insect section showing a compact bacteriome. (E) Amplification of the region indicated in the dashed square in (D) to show the endosymbiotic system in more detail. Scale: 100 μm.

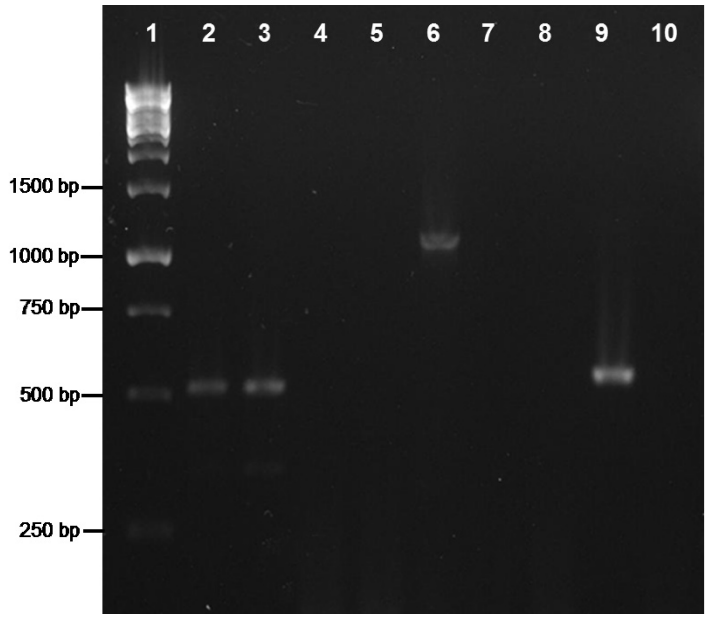


Figure 5. PCR analysis on ₜDNA extracted from *P. peruvianus*. Heads (lines 2, 5, 8) and bodies (lines 3, 6, 9) were analyzed. Lane 1, Molecular Weight Marker; lines 2–4, amplification of 18S rDNA; lanes 5–7, amplification of the 16S rDNA; lanes 8–10, amplification of *trp*B. Negative controls: lanes 4, 7, 10.

Molecular characterization and phylogenetic analysis of the endosymbionts from the analyzed mealybugs

TDNA from *D. boninsis*, *P. longispinus*, *P. peruvianus*, and *P. madeirensis* was used for the amplification of 16S rDNA with universal primers, and the corresponding amplicons were cloned and sequenced. BLAST analysis of the obtained sequences revealed the presence of two different haplotypes, corresponding to a beta and a gammaproteobacterium, in the two Pseudoccocinae species (*D. boninsis* and *P. longispinus*), whereas the Phenacoccinae (*P. peruvianus* and *P. madeirensis*) yielded a single haplotype from a betaproteobacterium. We have obtained the almost-complete sequence of the 16S rRNA gene (1467 bp) of the betaproteobacterium from *P. madeirensis*, less than 50% of which was previously available (Gruwell et al., 2010). The sequence of betaproteobacterial origin amplified from *P. longispinus* is identical to the one that had been previously identified (Acc. no. JN182336). However, the gammaproteobacterial sequence obtained in this study (deposited in GenBank under Acc. no. KF742536) is not the same as the partial 16S rDNA sequences previously reported by Duron et al. (2008) and Rosenblueth et al. (2012). In fact, there are 13 polymorphic sites among the three haplotypes in the compared region (525 bp). A diagnostic screening by restriction enzyme analysis performed on the common region of the 16S rRNA genes among "*Ca*. Tremblaya princeps" and the three gamma-haplotypes from *P. longispinus* confirmed that the sequences previously published were not present in our sample (Figure 6).

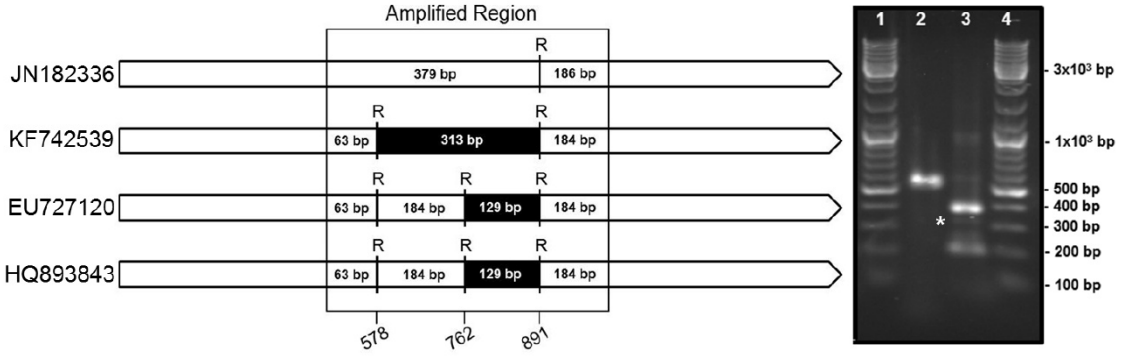


Figure 6. Endosymbionts survey in *P. longispinus*. Left panel: Restriction maps for the four putatively present 16S rDNA sequences, identified by their accession numbers in GenBank, i.e., "*Ca*. Tremblaya princeps" (JN182336) and the sequences of gamma-proteobacterial origin identified by us (KF742539), Duron et al. (2008) (EU727120), and Rosenblueth et al. (2012) (HQ893843). RsaI restriction sites and expected fragments length are shown. Position numbers refer to the *E. coli* homolog. Right panel: RFLP results. Lanes 1 and 4, Molecular Weight Marker; lane 2, undigested amplicon; line 3, amplicon digested with RsaI. A band around 313 bp (labeled with an asterisk) is indicative of the presence of the KF742539 sequence (this work). Absence of a band around 129 bp rules out the presence of EU727120 and HQ893843.

The phylogenetic relationship of the endosymbionts under study was determined. The analysis was performed on 1221 unambiguously aligned sites of the 16S rRNA gene from 50 different prokaryote lineages, including gamma and betaproteobacteria, both free-living and endosymbionts (Figure 7). Two separate clades for beta and gammaproteobacteria are well defined. Among the beta-endosymbionts, the one from *D. boninsis* forms a monophyletic clade with the other "*Ca*. Tremblaya princeps," while the endosymbionts of both Phenacoccus belong to the "*Ca*. Tremblaya phenacola" clade. They form a monophyletic cluster with the endosymbiont from *Phenacoccus solani* (Ferris), and appear separated from the subclade of strain PAVE. The obtained tree topology is congruent with that of the hosts (Hardy et al., 2008). Regarding the new gamma-endosymbionts characterized in this work, those from *D. boninsis* and *P. longispinus* do not cluster together. Nevertheless, these clustering have little support. To better define the phylogenetic position of the newly described gamma-endosymbionts, we took advantage of the *arg*H and *trp*B genes of gammaproteobacterial origin that we detected in the three Pseudococcinae mealybugs analyzed in this work (Figure 1). We performed a phylogenetic reconstruction using a concatenate of the available sequences from the two coding-genes of gamma-proteobacteria used in Figure 7. The analysis includes 1308 unambiguously aligned sites (489 from *arg*H and 819 from *trp*B) (Figure 8). This new tree confirms that the gamma-endosymbionts of *D. boninsis* and *P. longispinus* do not belong to the same clade. Furthermore, the gamma-endosymbionts of the two Pseudococcus species analyzed (*P. viburni* and *P. longispinus*) cluster with the Sodalis-like endosymbionts, but they do not form a monophyletic group either with "*Ca*. Moranella endobia" or between them.

Finally, we performed a molecular characterization of the 16S rDNA sequences from the analyzed beta-endosymbionts from the two Phenacoccus species under study. Strain PPER (from *P. peruvianus*) and PMAD (from *P. madeirensis*) have a GC-content of 48.5 and 48.1%, respectively. This GC-content fits well into the range described for the other characterized "*Ca*. Tremblaya phenacola" strains (from 45.8% in *Peliococcus turanicus* Kiritshenko to 50.6% in *Mirococcus* sp.), and it is clearly lower than the GC-content of the 16S rRNA gene from "*Ca*. Tremblaya princeps" (whose known minimum is 55.4% in *Pseudococcus comstocki* Kuwana; Koga et al., 2013). Multiple sequence alignment of the 16S rDNA sequences from "*Ca*. Tremblaya phenacola" strains PPER and PMAD revealed that both strains present four out of five motifs used by Gruwell et al. (2010) to define this species. However, in both cases the motif AGTT is modified to AGCT (positions 1240–1243 in the sequence from strain PPER). Additionally, the motifs AATGTC and TTTTA (sites 160–165 and 1121–1125, respectively, in PPER), also present in "*Ca*. Tremblaya phenacola" from *P. solani*, appear to be unique for this subclade members.

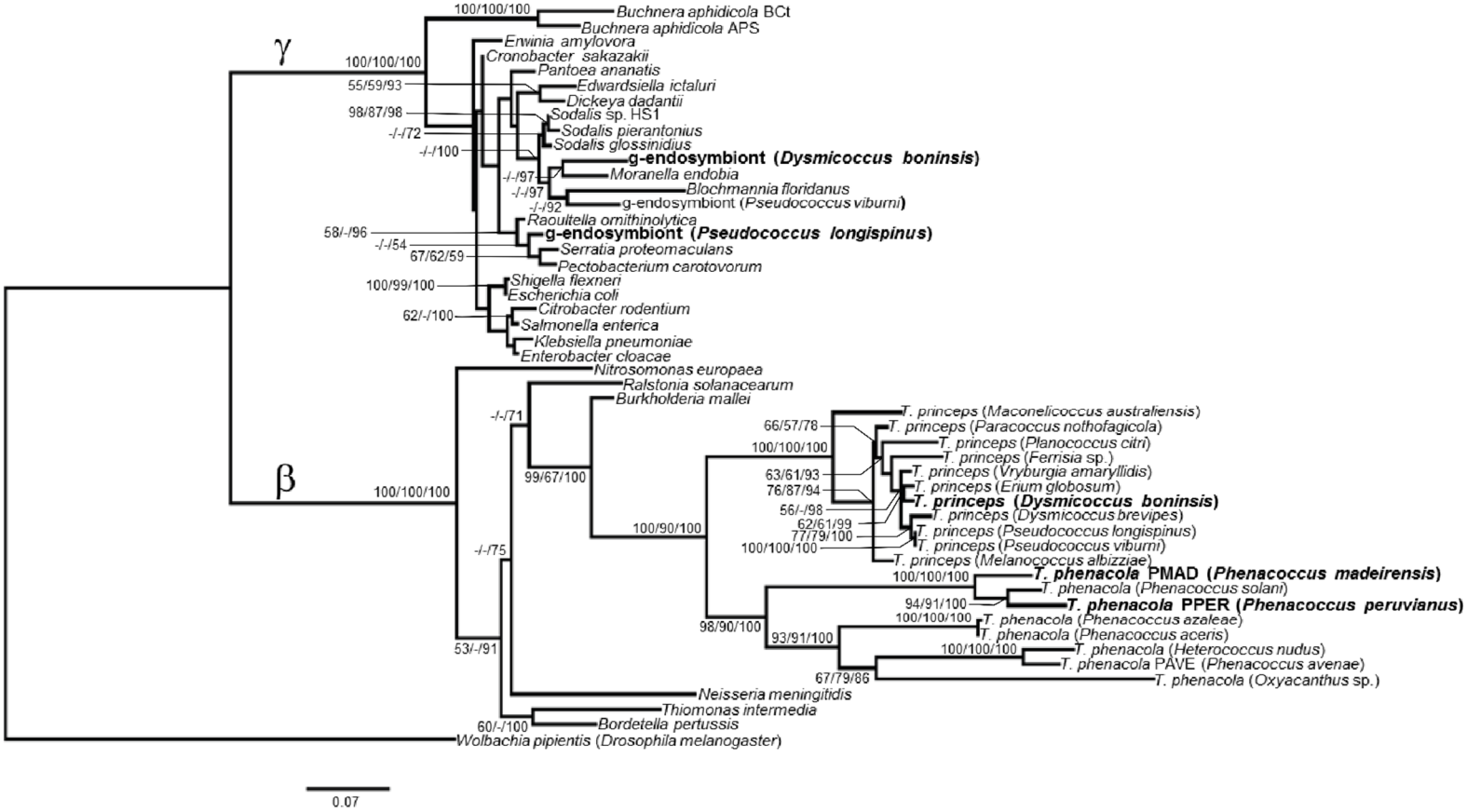


Figure 7. ML phylogenetic analysis of gamma-endosymbionts and *Tremblaya* lineages based on their 16S rDNA sequences. *Wolbachia pipientis* Dmel (alphaproteobacterial endosymbiont of *Drosophila melanogaster*) was used as outgroup. Bayesian and MP analysis gave essentially the same results. ML and MP bootstrap values, and Bayesian posterior probabilities over 50% are represented. Scale bar represents substitutions per site.

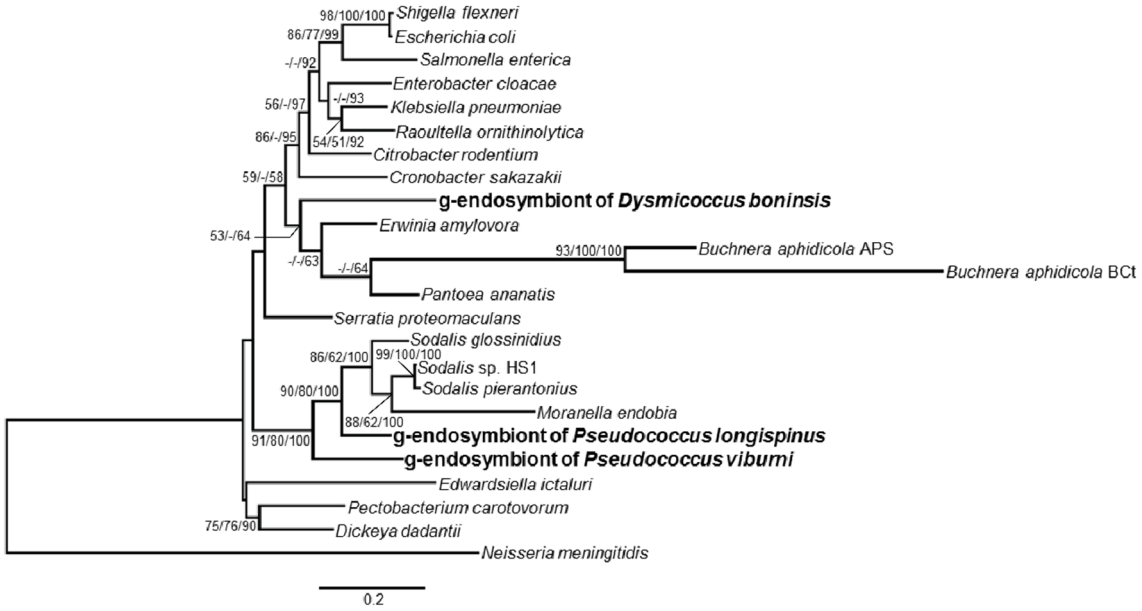


Figure 8. Phylogenetic relationships among gamma-endosimbionts of mealybugs from subfamily Pseudoccinae, based on concatenated sequences of genes *arg*H and *trp*B. Sequences obtained in this work are in bold. The betaproteobacterium *Neisseria meningitidis* MC58 was used as outgroup. Bayesian and MP analysis gave essentially the same results. ML and MP bootstrap values, and Bayesian posterior probabilities over 50% are represented. Scale bar represents substitutions per site.

DISCUSSION

The availability of the complete "*Ca*. Tremblaya" genomes from two mealybug species revealed that "*Ca*. Tremblaya phenacola" PAVE alone is able to provide its host with the same essential amino acid biosynthetic capabilities as the consortium composed by "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" in *P. citri* (Husnik et al., 2013). However, no other mealybug species have been thoroughly analyzed for their essential amino acids biosynthetic capabilities. To address this issue, we performed a genetic screening in five endosymbiotic systems from different subclades of the mealybug subfamilies Pseudococcinae and Phenacoccinae. We screened for selected genes involved in the last step usually performed by the endosymbionts in the biosynthetic pathways of arginine (*arg*H), branched amino acids (*ilv*D and *leu*B), methionine (*met*E), threonine (*thr*C), and tryptophan (*trp*B) (Figure 1). Many targeted genes were detected, which is consistent with the critical relevance of these bacteria in essential amino acid supply. Even though we were unable to obtain some amplicons, primer-pair amplification problems seem an improbable cause, because different beta and gammaproteobacteria were successfully amplified in the lineages under study, and no single DNA template or primer pair led to complete negative results. In any case, apparent absence of some genes in certain lineages should be interpreted carefully. Negative results do not necessarily imply the absence of a certain *locus* in the endosymbiotic system, although they might indicate the absence of a functional one. Since the degenerate primers were designed on gene regions

encoding highly conserved residues, changes affecting the primer target sequences could affect both gene functionality and PCR results, potentially preventing the detection of pseudogenes. Nevertheless, we were able to amplify some likely recent pseudogenes that maintain a high level of identity with predicted functional homologs in closely related species and still retain most known critical residues for protein functioning (i.e., *arg*H and *met*E). Functional redundancy tends to be lost following a stochastic process in endosymbiotic consortia, since only one copy is necessary to fulfill host needs. Therefore, the detected recent gene inactivations, as well as the great variety in functional redundancies and gene retention patterns among endosymbionts of the Pseudococcinae lineages, suggest an ongoing specification of the role of each endosymbiotic partner in metabolic complementation. Moreover, our phylogenetic analyses (Figures 7, 8) indicate that each lineage was infected with different gammaproteobacteria, so that the detected gene losses are independent events. Our results indicate that all screened genes must have been present in the "*Ca*. Tremblaya" ancestor. Loss of *arg*H apparently occurred after acquisition of the gammaproteobacterial partner in Pseudococcinae. At present, all analyzed species have retained an ortholog of gammaproteobacterial origin, but this gene appears to be functional in "*Ca*. Tremblaya princeps" from *D. brevipes* (Baumann et al., 2002), and it is pseudogenized in "*Ca*. Tremblaya princeps" from *P. citri*. *ilv*D has been retained by all "*Ca*. Tremblaya phenacola" analyzed, while Pseudococcinae shows various alternatives: in *P. citri* and *P. viburni*, it is only present in "*Ca*. Tremblaya princeps", while in *D. boninsis* the gamma-endosymbiont performs this function. *leu*B has been retained in all "*Ca*. Tremblaya," whereas it is redundant in *D. boninsis* and *P. longispinus*. *thr*C of gamma-endosymbiont origin was not detected in any analyzed lineage. *met*E is redundant in *D. boninsis* and *P. longispinus*, although it is pseudogenized in the gammaproteobacterium of the latter. The strict conservation of MetE key residues suggests that "*Ca*. Tremblaya princeps" performs the last step in methionine biosynthesis from cysteine in all surveyed Pseudococcinae, as well as in "*Ca*. Tremblaya phenacola" PAVE (Husnik et al., 2013). However, this gene was not detected in the two Phenococcinae analyzed here. The methionine synthase MetH (EC 2.1.1.13) may perform this last step in these endosymbionts, as in "*Candidatus* Hodgkinia cicadicola", endosymbiont of cicadas (McCutcheon and Moran, 2010).

The most intriguing case relates to *trp*B. Even though Phenacoccinae harbor "*Ca*. Tremblaya phenacola" as a single endosymbiont, and in contrast to findings of the genome project of strain PAVE (Husnik et al., 2013), both Phenacoccus species analyzed here present only a gammaproteobacterial homolog (Figure 2). The two sequences we obtained are highly similar but not identical, and they present the common AT-content bias of P-endosymbionts. Both facts appear to discard DNA contamination. However, a PCR screening for gammaproteobacterial endosymbionts gave negative results for both Phenacoccus species (Figure 3). FISH and PCR analyses also showed that "*Ca*. Tremblaya phenacola" is the only bacteria found in *P. peruvianus*, where it is confined in the bacteriome (Figures 4, 5). Husnik et al. (2013) had recently reported several horizontally transferred genes of bacterial origin in the nuclear genome of *P. citri*, some of which are involved in the biosynthesis of several nutrients including the amino acid lysine. The source of such genes was not any of the members of the mealybug endosymbiotic consortium

(i.e., "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia"), even though many of them seem to complement gene loses in the consortium genomes. The authors suggest that several facultative symbionts, which are not essential for host survival, and can be free in the environment and infect the host sporadically, have been involved in HGT to the insect genome. Altogether, our findings suggest that HGT events have also occurred in mealybugs of the subfamily Phenacoccinae, affecting the evolution of, at least, one of the metabolic pathways for essential amino acids biosynthesis. In this case, however, the *trp*B gene appears to have been transferred to the "*Ca*. Tremblaya phenacola" genome. If confirmed by sequencing the whole endosymbiont genome, this would be (to our knowledge) the second case of HGT described in endosymbiotic bacteria. The other described case corresponds to "*Candidatus* Profftella armature," a defensive symbiont from the psyllid *Diaphorina citri*, in which the genes involved in the biosynthesis of a cytotoxic metabolite appear to have been horizontally acquired (Nakabachi et al., 2013).

The analysis of the obtained 16S rDNA sequences (Figure 7) showed the presence of a beta-endosymbiont in all mealybug species under study, whereas in the Pseudococcinae there is also a gammaproteobacterium, as expected (Thao et al., 2002; Hardy et al., 2008; Gruwell et al., 2010). The phylogenetic reconstruction using a concatenate of *arg*H and *trp*B allowed us a better characterization of the position of the newly identified gamma-endosymbionts (Figure 8). Our analyses show that both Pseudococcus gamma-endosymbionts analyzed in this work belong to the *Sodalis*-like clade. However, they do not form a monophyletic group either with "*Ca*. Moranella endobia" or between them. On the other hand, the gamma-endosymbiont of *D. boninsis* is not a *Sodalis*-like bacterium, consistently with the gamma-endosymbionts described for other species of the genus *Dysmicoccus*, based on 16S rDNA sequences (Thao et al., 2002).

The identification of the gamma-endosymbiont of *P. longispinus* had been controversial. Some authors were unable to detect it (Thao et al., 2002; Gatehouse et al., 2012), while two different haplotypes have been identified in other studies (Duron et al., 2008; Rosenblueth et al., 2012). In all these cases the genetic screening was not exhaustive (due to the analysis of a limited number of clones), and the amplified sequences were shorter than the one obtained in this work. Our results, based on the analysis of 36 clones, indicate the presence of a single gamma-endosymbiont in this species, which is consistent with what had been found in other Pseudococcinae. However, our sequence does not correspond to any of the previously described (Figure 6). The existence of different haplotypes might indicate high levels of intraspecific polymorphisms in the gamma-endosymbionts of *P. longispinus*. Alternatively, due to the high morphological similarity among mealybug species, problems in the identification of insect host species cannot be ruled out.

We have also performed phylogenetic and molecular characterization of "*Ca*. Tremblaya phenacola" strains PPER and PMAD. Only a few strains have been reported in this species previously, and two subclades have been described (Gruwell et al., 2010). As revealed by the phylogenetic analysis based on 16S rDNA sequences (Figure 7, Table S3 in Supplementary Material), strains PPER and PMAD are members of the most unexplored subclade. While they

present most of the characteristic sequences used by Gruwell et al. (2010) to define this species, we also identified some unique molecular signatures for this subclade. Considering the high morphological similarity exhibited by mealybug species, these sequences could be useful as potential targets on strategies for both bacteria and insect molecular identification (Cox, 1983; Charles et al., 2000). Specifically, the mealybug species *P. solani*, *P. peruvianus*, and *P. madeirensis*, which belong to this subclade and are invasive pests of horticultural and ornamental plants, represent a relevant threat in several European countries (Pellizzari and Germain, 2010). Therefore, the ability to differentiate them, at the molecular level, from other widespread polyphagous species such as *P. citri*, *P. viburni*, and *P. longispinus*, could provide a rapid pests detection tool for import/export controls in Europe.

In summary, our molecular and phylogenetic analyses provide a more complete picture of the complex evolutionary history of the two "*Ca*. Tremblaya" lineages. The genetic screening of selected genes confirmed the importance of mealybug endosymbionts in providing essential amino acids to their hosts. In several Pseudococcinae analyzed, the complementation of "*Ca*. Tremblaya princeps" and the gamma-endosymbionts is ongoing, given the gene redundancies found. We have also identified a putative case of HGT in "*Ca*. Tremblaya phenacola" for the biosynthesis of tryptophan. Finally, from an applied point of view, two diagnostic motifs in the 16S rDNA sequence have been identified, which could be potentially used to implement a rapid detection method to differentiate mealybug pests in horticultural and ornamental plants.

ACKNOWLEDGMENTS

This work was supported by grants BFU2012-39816-C02-01 (co-financed by FEDER funds and Ministerio de Economía y Competitividad, Spain) to Amparo Latorre and Prometeo/2009/092 (Conselleria d'Educació, Generalitat Valenciana, Spain) to Andrés Moya. Sergio López-Madrigal and Aleixandre Beltrà are recipients of fellowships from the Ministerio de Educación (Spain).

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25, 3389-3402. doi:10.1093/nar/25.17.3389

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol*. 56, 1919-1925.

Baumann, L., Thao, M.L., Hess, J.M., Johnson, M.W., and Baumann, P. (2002). The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts

of plant sap-sucking insects. *Appl. Environ. Microbiol*. 68, 3198–3205. doi:10.1128/AEM.68.7.3198-3205.2002

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol*. 59, 155-189. doi:10.1146/annurev.micro.59.030804.121041

Charles, J.G., Froud, K.J., and Henderson, R.C. (2000). Morphological variation and mating compatibility within the mealybugs *Pseudococcus calceolariae* and *P. similans* (Hemiptera: Pseudococcidae), and a new synonymy. *Syst. Entomol*. 25, 285-294. doi:10.1046/j.1365-3113.2000.00104.x

Cox, J.M. (1983). An experimental study of morphological variation in mealybugs (Homoptera: Coccoidea: Pseudococcidae). *Syst. Entomol*. 8, 361–382. doi:10.1111/j.1365-3113.1983.tb00489.x

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods*. 9, 772. doi:10.1038/nmeth.2109

Duron, O., Bouchon, D., Boutin, S., Bellamy, L., Zhou, L., Engelstadter, J., and Hurst, G.D. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol*. 6, 27. doi:10.1186/1741-7007-6-27

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. 360, 1427-1434. doi:10.1098/rstb.2005.1669

Fukatsu, T., Watanabe, K., and Sekiguchi, Y. (1998). Specific detection of intracellular symbiotic bacteria of aphids by oligonucleotide-probed in situ hybridization. *Appl. Entomol. Zool*. 33, 461-472.

Gatehouse, L.N., Sutherland, P., Forgie, S.A., Kaji, R., and Christeller, J.T. (2012). Molecular and histological characterization of primary (betaproteobacteria) and secondary (gammaproteobacteria) endosymbionts of three mealybug species. *Appl. Environ. Microbiol*. 78, 1187-1197. doi:10.1128/AEM.06340-11

Gil, R., Latorre, A., and Moya, A. (2010). “Evolution of prokaryote-animal symbiosis from a genomics perspective,” in *(Endo)symbiotic methanogenic archaea (Microbiology Monographs)*, ed. J. H. P. Hackstein (Berlin Heidelberg, Springer-Verlag), 207-233. doi:10.1007/978-3-642-13615-3_11

Gruwell, M.E., Hardy, N.B., Gullan, P.J., and Dittmar, K. (2010). Evolutionary relationships among primary endosymbionts of the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). *Appl. Environ. Microbiol*. 76, 7521–7525. doi:10.1128/AEM.01354-10

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*. 52, 696-704. doi:10.1080/10635150390235520

Hardy, N.B., Gullan, P.J., and Hodgson, C.J. (2008). A subfamily-level classification of mealybugs (Hemiptera: Pseudococcidae) based on integrated molecular and morphological data. *Syst. Entomol*. 33, 51-71. doi:10.1111/j.1365-3113.2007.00408.x

Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., von Dohlen, C.D., Fukatsu, T., McCutcheon, J.P. (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*. 153, 1567–1578. doi:10.1016/j.cell.2013.05.040

Koga, R., Nikoh, N., Matsuura, Y., Meng, X.Y., and Fukatsu, T. (2013). Mealybugs with distinct endosymbiotic systems living on the same host plant. *FEMS Microbiol*. Ecol. 83, 93–100. doi:10.1111/j.1574-6941.2012.01450.x

Kono, M., Koga, R., Shimada, M., Fukatsu, T. (2008). Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. *Appl. Environ. Microbiol*. 74, 4175–4184. doi:10.1128/AEM.00250-08

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. 23, 2947–2948. doi:10.1093/bioinformatics/btm404

Littlewood, D.T.J., and Olson, P.D. (2001). "SSU rDNA and the Platyhelminthes: signal, noise, conflict and compromise," in *Interrelationships of the Platyhelminthes*, ed. D. T. J. Littlewood & R. A. Bray (London, Taylor & Francis), 262–278.

Lopez-Madrigal, S., Latorre, A., Porcar, M., Moya, A., and Gil, R. (2011). Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. *J. Bacteriol*. 193, 5587–5588. doi:10.1128/JB.05749-11

Lopez-Madrigal, S., Latorre, A., Porcar, M., Moya, A., and Gil, R. (2013). Mealybugs nested endosymbiosis: going into the 'matryoshka' system in *Planococcus citri* in depth. *BMC Microbiol*. 13:74. doi:10.1186/1471-2180-13-74

McCutcheon, J.P., and Moran, N.A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol. Evol*. 2, 708–718. doi:10.1093/gbe/evq055

McCutcheon, J.P., and von Dohlen, C.D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol*. 21, 1366–1372. doi:10.1016/j.cub.2011.06.051

Moran, N.A., McCutcheon, J.P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet*. 42, 165–190.

Moya, A., Gil, R., and Latorre, A. (2009). The evolutionary history of symbiotic associations among bacteria and their animal hosts: a model. *Clin. Microbiol. Infect*. 15 Suppl 1, 11–13.

Mühling, M., Woolven-Allen, J., Murrell, J.C., and Joint, I. (2008). Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J*. 2, 379–392. doi: 10.1038/ismej.2007.97

Nakabachi, A., Ueoka, R., Oshima, K., Teta, R., Mangoni, A., Gurgui, M., Oldham, N.J., van Echten-Deckert, G., Okamura, K., Yamamoto, K., Inoue, H., Ohkuma, M., Hongoh, Y., Miyagishima, S., Hattori, M., Piel, J., and Fukatsu, T. (2013). Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol*. 23, 1478–1484. doi: 10.1016/j.cub.2013.06.027

Pellizzari, G., and Germain, J.F. (2010). "Scales (Hemiptera, Superfamily Coccoidea)," in *Alien terrestrial arthropods of Europe (Biorisk)*, ed. A. Roques, M. Kenis, D. Lees, C. Lopez-Vaamonde, W. Rabitsch, J. Y Rasplus, Roy DB (Sofia, Pensoft Publishers), 475-510.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol*. 61, 539–542. doi:10.1093/sysbio/sys029

Rosenblueth, M., Sayavedra, L., Sámano-Sánchez, H., Roth, A., and Martínez-Romero, E. (2012). Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). *J. Evol. Biol*. 25, 2357–2368. doi:10.1111/j.1420-9101.2012.02611.x

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics*. 16, 944–945. doi:10.1093/bioinformatics/16.10.944

Staden R., Beal K.F., and Bonfield J.K. (2000). The Staden package, 1998. *Methods Mol. Biol.* 132, 115-130.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22, 2688-2690. doi:10.1093/bioinformatics/btl446

Thao, M.L., Gullan, P.J., and Baumann, P. (2002). Secondary (γ-proteobacteria) endosymbionts infect the primary (β-proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Appl. Environ. Microbiol*. 68, 3190–3197. doi:10.1128/AEM.68.7.3190

van Ham, R.C., Moya, A., and Latorre, A. (1997). Putative evolutionary origin of plasmids carrying the genes involved in leucine biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids). *J. Bacteriol*. 179, 4768–77.

von Dohlen, C.D., Kohler, S., Alsop, S.T., and McManus, W.R. (2001). Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature*. 412, 433–436. doi:10.1038/35086563

SUPPLEMENTARY MATERIAL

Table S1. Bacterial species included in the multiple alignment leading to degenerate primers design.

Bacterial Species	**Class***	**Accession Numbers**
Burkholderia glumae BGR1	Beta	CP001503, CP001504
Burkholderia multivorans ATCC 17616	Beta	AP009385, AP009386, AP009387
Burkholderia pseudomallei 1106a	Beta	CP000572, CP000573
Burkholderia thailandensis E264	Beta	CP000085, CP000086
"Candidatus Tremblaya princeps" PCVAL	Beta	CP002918
Escherichia coli K-12 substr. MG1655	Gamma	U00096
Salmonella enterica Typhimurium str. LT2	Gamma	AE006468
Serratia proteomaculans 568	Gamma	CP000826
Sodalis glossinidius	Gamma	AP008232
Dickeya dadantii 3937	Gamma	CP002038
Yersinia pestis Angola	Gamma	CP000901
"Candidatus Moranella endobia" PCVAL	Gamma	CP003881

*Beta: betaproteobacteria; Gamma: gammaproteobacteria

Table S2. List of additional primers used in this work.

Gene	Primer pairs	Sequence (5'→3')	Specificity	Insect sample	Application (Results section)	Reference
16S	16S-up 16S-down	AGAGTTTGATCATGGCTCAGATTG TACCTTGTTACGACTTCACCCCAG	Bacteria	*D. boninsis* *P. longispinus* *P. peruvianus* *P. madeirensis*	Endosymbionts molecular characterization (3.3)	van Ham *et al*., 1997
16S	Gamma395f Gamma871r	CMATGCCGCGTGTGTGAA ACTCCCCAGGCGGTCDACTTA	γ-proteobacteria	*P. peruvianus* *P. madeirensis*	Endosymbionts molecular characterization (3.2)	Mühling *et al*., 2008
18S	930F 1270R	GCATGGAATAATGGAATAGG CCGTCAATTCCTTTAAGT	Eukarya	*P. peruvianus*	*trp*B localization (3.2)	Littlewood and Olson, 2001
16S	HGT-16-F HGT-16-R	GGTGGCTAATACCGCATAATGTC GGCTAGTTCTTACGAAGTTGCAG	"*Ca*. Tremblaya phenacola" from *P. peruvianus*	*P. peruvianus*	*trp*B localization (3.2)	This work
trpB	HGT-TRPB-F HGT-TRPB-R	CATGGTGTGGCTTCTTCAATAG GGAAGTCAAGACCAGCAGATA	"*Ca*. Tremblaya phenacola" from *P. peruvianus*	*P. peruvianus*	*trp*B localization (3.2)	This work
16S	RFLP-up RFLP-down	TGCCAGCAGCCGCGGTAATAC ACACGAGCTGACGACAGCCATG	Endosymbionts of *P. longispinus*	*P. longispinus*	RFLP analysis (3.3)	This work

Table S3. Bacterial species included in the phylogenetic analyses, by alphabetical order in each class. Accession numbers refer to complete genomes or available sequences for 16S rRNA[1], *argH*[2] and *trpB*[3] genes. Sequences obtained in this work appear in bold.

Species	Class	Insect host	Accession number
Wolbachia pipientis	Alpha	*Drosophila melanogaster*	AE017196
Bordetella pertussis CS	Beta	-	CP002695
Burkholderia mallei ATCC 23344	Beta	-	CP000011
"Ca. Tremblaya phenacola*"*	Beta	*Phenacoccus solani*	HM449980[1]
"Ca. Tremblaya phenacola*"*	Beta	*Phenacoccus aceris*	HM449982[1]
"Ca. Tremblaya phenacola*"*	Beta	*Phenacoccus azaleae*	AB627026[1]
"Ca. Tremblaya phenacola*"*	Beta	*Heterococcus nudus*	HM449976[1]
"Ca. Tremblaya phenacola*"*	Beta	*Oxyacanthus sp.*	HM449972[1]
"Ca. Tremblaya phenacola*"* PAVE	Beta	*Phenacoccus avenae*	CP003982
"Ca. Tremblaya phenacola*"* PMAD	Beta	*Phenacoccus madeirensis*	**KF444180**[1], **KF444184**[3]
"Ca. Tremblaya phenacola*"* PPER	Beta	*Phenacoccus peruvianus*	**KF444174**[1], **KF444179**[3]
"Ca. Tremblaya princeps*"*	Beta	*Maconellicoccus australiensis*	AF476088[1]
"Ca. Tremblaya princeps*"*	Beta	*Melanococcus albizziae*	AF476087[1]
"Ca. Tremblaya princeps*"*	Beta	*Paracoccus nothofagicola*	AF476094[1]
"Ca. Tremblaya princeps*"* PCVAL	Beta	*Planococcus citri*	CP002918
"Ca. Tremblaya princeps*"*	Beta	*Ferrisia* sp.	AF476086[1]
"Ca. Tremblaya princeps*"*	Beta	*Vryburgia amaryllidis*	AF476097[1]
"Ca. Tremblaya princeps*"*	Beta	*Erium globosum*	AF476084[1]
"Ca. Tremblaya princeps*"*	Beta	*Pseudococcus longispinus*	AF476093[1]
"Ca. Tremblaya princeps*"*	Beta	*Dysmicoccus brevipes*	AF476082[1]
"Ca. Tremblaya princeps*"*	Beta	*Pseudococcus viburni*	JN182337[1]
"Ca. Tremblaya princeps*"*	Beta	*Dysmicoccus boninsis*	**KF911099**[1]
Neisseria meningitidis MC58	Beta	-	AE002098
Nitrosomonas europaea ATCC 19718	Beta	-	AL954747
Ralstonia solanacearum GMI1000	Beta	-	NC003295
Thiomonas intermedia K-12	Beta	-	CP002021
Blochmannia floridanus	Gamma	*Camponotus floridanus*	NC_005061
Buchnera aphidicola APS	Gamma	*Acyrthosiphon pisum*	NC_002528
Buchnera aphidicola BCt	Gamma	*Cinara tujafilina*	NC_015662
"Ca. Moranella endobia*"* PCVAL	Gamma	*Planococcus citri*	CP003881
Citrobacter rodentium ICC168	Gamma	-	FN543502
Cronobacter sakazakii ATCC BAA-894	Gamma	-	NC_009778
Dickeya dadantii 3937	Gamma	-	NC_014500
Edwardsiella ictaluri 93-146	Gamma	-	NC_012779
Enterobacter cloacae ATCC 13047	Gamma	-	NC_014121
Erwinia amylovora ATCC 49946	Gamma	-	FN666575
Escherichia coli K-12 MG1655	Gamma	-	U00096
Klebsiella pneumoniae 342	Gamma	-	CP000964
Pantoea ananatis AJ13355	Gamma	-	NC_017531
Pectobacterium carotovorum subsp. carotovorum PC1	Gamma	-	NC_012917
Raoultella ornithinolytica B6	Gamma	-	NC_021066
Salmonella enterica DT104	Gamma	-	HF937208
Serratia proteamaculans 568	Gamma	-	CP000826
Shigella flexneri 2a str. 301	Gamma	-	NC_004337
Sodalis sp. HS1	Gamma	-	CP006569
Sodalis glossinidius	Gamma	*Glossina morsitans*	AP008232
Sodalis pierantonius SOPE	Gamma	*Sitophilus oryzae*	CP006568
Unnamed endosymbiont	Gamma	*Dysmicoccus boninsis*	**KF911098**[1], **KF444196**[2], **KF444200**[3]
Unnamed endosymbiont	Gamma	*Pseudococcus longispinus*	**KF742539**[1], **KF444192**[2], **KF444195**[3]
Unnamed endosymbiont	Gamma	*Pseudococcus viburni*	JN182341[1], **KF444190**[2], **KF444191**[3]

Table S4. Summary of residues involved in MetE enzymatic function.

		P. avenae	***P. citri***	***D. boninsis***		***P. viburni***	***P. longispinus***	
MetE functional motifs	*E. coli* key amino acids	β	β	β	γ	β	β	γ
Zn^{2+} Site	H641	+	+	+	+	+	+	+
	C643	+	+	+	+	+	+	+
	E665	+	+	+	+	+	+	+
	C726	+	+	n. a.	n. a.	n. a.	n. a.	n. a.
MetE-Hcy binary complex	I431	V	V	V	+	V	V	+
	S433	+	+	+	+	+	+	+
	E484	+	+	+	+	+	+	+
	M490	+	+	+	+	+	+	+
	D599	+	+	+	+	+	+	+
MetE-folate binary complex	R17	+	+	n. a.	n. a.	n. a.	n. a.	n. a.
	K20	+	+	n. a.	n. a.	n. a.	n. a.	n. a.
	K117	+	+	n. a.	n. a.	n. a.	n. a.	n. a.
	R515	+	+	+	+	+	+	+
	K518	+	+	+	+	+	+	+
	W561	+	+	+	+	+	+	+
	E605	+	+	+	+	+	+	+
DMV sequence	D489	+	+	+	+	+	+	+
	M490	+	+	+	+	+	+	+
	V491	+	+	+	+	+	+	+

Amino acid numbers refer to the *E. coli* K12 substr. MG1655 homolog. Since *met*E sequences are only partially available for *D. boninsis*, *P. viburni* and *P. longispinus* endosymbionts, information about the conservation of R17, K20, K117 and C726 homolog residues is not available (n. a.). Full identity was detected for all analyzed amino acids except for I431 (involved in MetE-Hcy binary complex formation), turned into the also non-polar amino acid valine in all analyzed *Tremblaya* lineages

.

Supplementary Results

Analysis of pseudogenes involved in the biosynthesis of essential amino acids. The multiple alignment of homolog *loci* from a set of selected betaproteobacteria (*Burkholderia mallei* NCTC10229, "*Ca*. Tremblaya princeps" from *D. brevipes*, *Ralstonia solanacearum* GMI1000, *Nitrosomonas europea*, *Neisseria meningitidis* MC58, *Thiomonas intermedia* K-12 and *Bordetella pertusis*), and *E. coli* allowed us to identify the inactivating mutations in the *arg*H *locus* in "*Ca*. Tremblaya princeps" PCVAL (GenBank acc. no. CP002918) and the *met*E *locus* in the gamma-endosymbiont from *P. longispinus* (GenBank acc. no. KF444194).

*arg*H. We have identified two deletions, involving 6 and 57 nucleotides (sites 37548-49 and 37890-91, respectively), which cause the loss of the fully conserved amino acids A56, Y172, M175, R178, D179 and R182 (amino acid numbers refer to the homolog protein in *E. coli*). We also detected an inactivating frameshift caused by a single cytosine deletion between sites 38038 and 38039, which is part of the GCG codon (A236) in the homolog gene of "*Ca*. Tremblaya princeps" from *D. brevipes* (Baumann et al., 2002). However, two protein functional domains (PF00206 and PF14698 at sites 37402-38222 and 38409-38594, respectively) and their predicted active residues (H156, S277 and E290) are still recognizable in the "*Ca*. Tremblaya princeps" PCVAL pseudogene, according to Pfam information.

*met*E. The *met*E homolog in the *P. longispinus* gamma-endosymbiont appears to be inactivated due to the presence of a nonsense mutation (TGG→TAG) affecting the highly conserved W140 in the *E. coli* homolog protein. However, a careful analysis of all known residues that are important for protein function (Pejchal et al., 2005; Koutmos et al., 2008) showed that they are still preserved, as in all other mealybug endosymbionts analyzed, which indicates a recent pseudogenization event. The results of the essential motif analysis are presented in Table S4.

The methionine synthase MetE (EC 2.1.1.13) catalyzes the transfer of a methyl group from N5-methyl-5,6,7,8-tetrahydrofolate to L-homocysteine, activating L-homocysteine by binding the thiolate form of the substrate to Zn^{2+}. The motifs analyzed in this work include the four residues at the zinc-binding site, the residues that participate at the zinc-replete MetE-L-homocysteine complex formation, those that participate at the generation of the MetE-N5-methyl-5,6,7,8-tetrahydrofolate complex, and the DMV region involved in the substrate binding sites communication, so that the binding of any of the substrates increases MetE affinity for the acquisition of the other substrate.

Supplementary References

Koutmos, M., Pejchal, R., Bomer, T.M., Matthews, R.G., Smith, J.L., and Ludwig, M.L. (2008). Metal active site elasticity linked to activation of homocysteine in methionine synthases. Proc. Natl. Acad. Sci. 105, 3286–3291. doi: 10.1073/pnas.0709960105

Pejchal, R., and Ludwig, M.L. (2005). Cobalamin-independent methionine synthase (MetE): a face-to-face double barrel that evolved by gene duplication. PLoS Biol. 3:2. doi: 10.1371/journal.pbio.0030031

CHAPTER 5

Status: Submitted.

Tremblaya princeps acquisition of intracellular bacteria: bulls in a China shop

Sergio López-Madrigal[1], Amparo Latorre[1,2], Andrés Moya[1,2], Rosario Gil[1]*.

1 Institut Cavanilles de Biodiversitat y Biologia Evolutiva (ICBiBE), Universitat de València, Apartado Postal 22085, València 46071, Spain. 2 Área de Genómica y Salud, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana FISABIO–Salud Pública, Avenida Cataluña 21, València 46020, Spain.

*Corresponding author: rosario.gil@uv.es

ABSTRACT

Many insect species establish mutualistic symbiosis with intracellular bacteria that complement their unbalanced diets. The betaproteobacteria "*Candidatus* Tremblaya" maintain an ancient symbiosis with mealybugs (Hemiptera: Coccoidea: Pseudococcidae), which are classified in two subfamilies. Most Phenacoccinae mealybugs have "*Candidatus* Tremblaya phenacola" as their unique intracellular endosymbiont, while most Pseudococcinae mealybugs show a nested dual symbiosis where every "*Candidatus* Tremblaya princeps" cell harbors several cells of its corresponding gammaproteobacterial partner. Genomic characterization of the endosymbiotic consortium from *Planococcus citri,* composed by "*Ca*. Tremblaya princeps" and the gammaproteobacteria "*Candidatus* Moranella endobia", unveiled that the genome of the former presents several atypical features regarding gene repertoire, coding density, genomic instability and *loci* redundancy. Its comparison with the recently sequenced genome of "*Ca*. Tremblaya phenacola", single endosymbiont of *Phenacoccus avenae*, suggested that the atypical reductive evolution of the former could be linked to the acquisition of "*Ca.* Moranella endobia", which possess an almost complete set of genes involved in homologous recombination. We have checked for "*Ca*. Tremblaya princeps" susceptibility to homologous recombination in *P. citri,* and analyzed concerted evolution and presence of homologous recombination genes in four additional mealybug species, *Dysmicoccus boninsis*, *Planococcus ficus*, *Pseudococcus longispinus* and *Pseudococcus viburni*. Our results support the link between nested symbiosis and the atypical "*Ca.* Tremblaya princeps" structural genome features.

INTRODUCTION

The advances in genome sequencing and the development of metagenomic methods have been critical for our knowledge of the bacterial world. Hundreds of bacterial genomes have been deposited in open databases since 1996, allowing comparative analysis of bacterial genome

content and organization, and revealing the complex evolutionary dynamics of bacterial chromosomes. Now that complete genomes from closely related species or even from different strains of the same species are available, numerous studies have focused on the diversity of gene repertoire and genome rearrangements [1]. Horizontal gene transfer (HGT), transposition and intragenomic recombination lead to extensive structural diversity and fluidity of prokaryotic genomes, which are remarkable among free-living bacteria [2,3]. These phenomena are known to be important sources of evolutionary novelties, being responsible for bacterial huge metabolic diversity and adaptive potential. However, the analysis of bacteria that have acquired an intracellular host-dependent life-style revealed important constrains to these evolutionary mechanisms.

During the last 15 years, the complete genomes of many endosymbionts (i.e., obligate symbiotic bacteria that live inside eukaryotic cells) have become available. Comparative genomics has allowed the identification of several commonalities among them, which are related with the stage of integration of the bacteria with their respective hosts [4,5]. The best studied cases of endosymbiosis involve mutualistic associations with insects, which have been considered a driving force in the diversification of this animal group. Generally, intracellular bacteria have smaller genomes than their free-living relatives. The reduction in genome size is not only due to the reduction in gene content, but also in both gene and intergenic region (IGR) size, which lead to highly compact genomes [5]. Gene losses affect *loci* performing functions that are unnecessary in an intracellular environment or that can be provided by the host. Moreover, the obligatory intracellular life-style prevents the acquisition of genetic material by HGT. Thus, highly reduced genomes (i.e., those from endosymbionts that have maintained a long relationship with their hosts) have typically lost most genes involved in DNA recombination and repair, present almost no gene duplications, and lack transposable elements and prophages. There are some exceptions to these rules. Several strains of the betaproteobacteria *"Ca.* Tremblaya princeps*"*, the obligate endosymbiont of mealybugs from subfamily Pseudococcinae [6], present a duplicated region containing the complete ribosomal operon evolving under concerted evolution. Surprisingly, the complete sequencing of the genome of two strains belonging to *Planococcus citri* revealed that they possess a tiny 139-kb genome [7,8]. In addition, gene density of these genomes is one of the lowest among endosymbiotic bacteria (about 70%). *P. citri* presents an unusual nested endosymbiotic system in which each *"Ca.* Tremblaya princeps*"* cell harbors several cells of the gammaproteobacterium *"Candidatus* Moranella endobia*"* [7,9]. It is worth noticing that, even though the latter possess a typically reduced genome for an endosymbiont regarding gene content and density, it has a quite complete set of genes for DNA repair and recombination, including genes involved in both RecF and RecBCD recombination pathways, the two redundant mechanisms for this function that are nearly ubiquitous in free-living bacterial species [10,11,12].

Contrary to what has been observed in free-living bacteria, the high degree of stability in genome structure is also a common trait of most endosymbiont genomes. An extreme case of

genome stability has been observed after sequencing the genomes of *Buchnera aphidicola* from eight different aphid species [13,14,15,16,17,18,19]. This bacterial strains have maintained a nearly perfect fossil gene-order over 180 million years of evolution [17]. However, there are also notable exceptions to this high genomic stability. Lack of synteny conservation and genomic instability have been recently described in "*Candidatus* Portiera aleyrodidarum", endosymbiont of the whitefly *Bemisia tabaci* [20]. Synteny has not been either preserved in the genus "*Ca*. Tremblaya*"*. *"Ca.* Tremblaya princeps*"* PCVAL and PCIT genomes differ from that of *"Ca.* Tremblaya phenacola*"* PAVE, primary endosymbiont of the mealybug *Phenacoccus avenae*, in one large inversion and one plasmid containing two ribosomal-protein genes in the latter [21]. Additionally, during the genome sequencing of *"Ca.* Tremblaya princeps*"* PCIT, a 7-kb region flanked by two 71-bp inverted repeats was found in both orientations [7]. Strikingly, both "*Ca*. Portiera aleyrodidarum" and *"Ca.* Tremblaya princeps*"* share some other unusual structural features among obligate insect endosymbionts (i.e., low gene density and paralogous *loci* evolving under concerted evolution).

Mealybugs (Hemiptera: Coccoidea: Pseudococcidae) are phloem-feeding insects that have been classified in subfamilies Pseudococcinae and Phenacoccinae [22], and present an intricate variety of endosymbiotic relationships. Based on phylogenetic analysis, it has been suggested that a betaproteobacterial ancestor of *"Ca.* Tremblaya*"* infected a mealybug ancestor before the split of the two subfamilies. In subfamily Phenacoccinae, *"Ca.* Tremblaya phenacola*"* is the obligate endosymbiont in most tested mealybug species, excluding the tribe *Rhizoecini* and the genus *Rastrococcus*, where it has been replaced by different *Bacteroidetes* [21,23]. In subfamily Pseudococcinae, the obligate endosymbiont *"Ca.* Tremblaya princeps*"* has been classified in up to six different clusters (A to F) [24]. Except for those of *Ferrisia* and *Maconellicoccus* (B and F, respectively), where no additional endosymbiont has been reported, "*Ca*. Tremblaya princeps" has been infected multiple times by different gammaproteobacteria, establishing nested endosymbiotic consortia in which each *"Ca.* Tremblaya princeps*"* cell contains several cells of the corresponding gammaproteobacterium [9,24,25,26].

The nested endosymbiotic system identified in *P. citri* (cluster E), where *"Ca.* Tremblaya princeps*"* harbors *"Ca.* Moranella endobia*"*, has been extensively studied [7,8,9,12,21,27]. Both endosymbiont genomes have been sequenced for two different strains, PCVAL and PCIT [7,8,12]. Since the two strains present almost identical genomes, and strain PCVAL was sequenced in our laboratory, we will refer to this strain from now on throughout the text, except when indicated. The analysis of the functional capabilities of the nested consortium found in *P. citri* suggests the existence of an intricate case of complementation, involving both metabolic and informational functions (i.e., DNA replication, transcription and translation). Later on, the genome of "*Ca*. Tremblaya phenacola" PAVE was sequenced and compared with that of *"Ca.* Tremblaya princeps*"*. Both "*Ca*. Tremblaya" species possess tiny genomes (138.9 and 171.5 kb, respectively), an indication that a severe gene loss must have occurred at the beginning of the obligate intracellular symbiosis in the common ancestor of both species. However, they present

remarkable differences in several genome features, including gene repertoire, coding density and genome GC-content [21]. The genome of *"Ca.* Tremblaya princeps*"* PCVAL is an almost perfect subset of that from *"Ca.* Tremblaya phenacola*"* PAVE [21], which has retained many genes involved in metabolic and informational functions that are absent in *"Ca.* Tremblaya princeps*"* and must be provided by its nested endosymbiont "*Ca*. Moranella endobia" for both proper cell functioning and fulfilling the host needs.

The maintenance of homologous recombination (HR) ability in *"Ca.* Moranella endobia*"* could be on the root of concerted evolution and structural instability of the "*Ca*. Tremblaya princeps" genome. If this hypothesis is correct, we expect to find genes involved in HR and/or signals of recent concerted evolution in other endosymbionts of mealybugs belonging to subfamily Pseudococcinae. In this study, we have analyzed concerted evolution and presence of HR-related genes in the endosymbiotic systems from four unexplored mealybug species. The gray sugarcane mealybug *Dysmicoccus boninsis*, the long tailed mealybug *Pseudococcus longispinus*, and the obscure mealybug *Pseudococcus viburni* are phylogenetically distant members of the tribe Pseudococcinni [22]. The gamma-endosimbionts described for these species have been independently acquired [28]. The vine mealybug *Planococcus ficus* is a close relative of *P. citri*, whose endosymbiotic system has been widely characterized. Our results suggest a complex role of the nested gamma-endosymbionts on the reductive evolution of the *"Ca.* Tremblaya princeps*"* genome, not only at the functional but also at the structural level. Additionally, we have checked for "*Ca*. Tremblaya princeps" susceptibility to HR in *P. citri*. Based on our findings, we suggest an explanation for the low gene density of the highly reduced genome of "*Ca*. Tremblaya princeps" PCVAL.

MATERIALS AND METHODS

Insect sample collection and DNA extraction

Insects belonging to the species *P. longispinus, P. viburni* and *D. boninsis* were field collected in the Botanical Garden of the Universitat de València (València, Spain. 39º 28' 11.667'' N, 0° 22' 34.637 W), with permission from the curator of the garden, Dr. Jaime Güemes. *P. ficus* was sampled from a population reared on *Vitis vinifera* at the Mediterranean Agroforestal Institute, Universitat Politècnica de València (València, Spain. 39º 29' 1.699 N, 0º 20' 28.978'' W). This study did not involve endangered or protected species. Insects were stored in absolute ethanol at -20ºC. Total insect DNA ($_T$DNA) was extracted from adult female insects, where endosymbiont populations are expected to reach a peak [29], using JETFLEX Genomic DNA Purification Kit (GENOMED).

DNA amplification and sequencing

PCR amplifications were performed on insect TDNA with appropriate primer pairs (see below), using 50-60 µmoles of each primer per 50 µl reaction, and the KAPATaq DNA Polymerase Kit (Kapa Biosystems). *P. citri* TDNA was used as positive control. The thermal cycling protocol was as follows: an initial denaturation at 95ºC for 5 min, followed by 35 cycles of 50 s at 95ºC, 40 s at 55ºC (or 52ºC when indicated), and 2 min at 72ºC, plus a final extension step of 7 min at 72ºC. Amplicons were ABI sequenced at the sequencing facility of the Universitat de València.

Sequencing reads were quality surveyed and assembled with Staden Package (http://staden.sourceforge.net) [30]. Artemis software was used for sequence data management (http://www.sanger.ac.uk/resources/software/artemis/) [31].

Molecular and evolutionary analysis

"Ca. Tremblaya princeps*"* genomic fragments *leuA-rrs*1 and *prs-rrs*2 were PCR amplified with proper combinations of the already described primers leuA, prs5/6 and U16S [6]. The same primers, as well as OR-leuAR2 (5´-TCAGTMATTAHGGCWACCTGCAC-3´), OR-prsR2 (5´-AATAGCYAAGCGGGTCAAGGC-3´) and OR-UF2 (5´-TGGCGCATGCTGTATGAGTTC-3´), were used to sequence the PCR products. tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) [32] and ARAGORN (http://mbio-serv2.mbioekol.lu.se/ARAGORN/) [33] were used for the prediction of tRNA genes. All other genes were annotated by BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi/) [34]. The newly obtained sequences have been deposited in the GenBank database (*D. boninsis*, KF591104 and KF591105; *P. longispinus*, KF591108 and KF591109; *P. viburni*, KF591110 and KF591111; *P. ficus*, KF591106 and KF591107).

Multiple alignments were performed with ClustalW [35]. The ancient state of sites under concerted evolution was inferred for the last common ancestor (LCA) of "*Ca*. Tremblaya princeps" strains from clusters C and E. Analysis was performed by Maximum Likelihood (ML) with the DNAML program of the PHYLIP v3.69 package [36], predefining the tree topology as already determined [22].

The 16S rRNA gene sequences from 19 gammaproteobacterial endosymbionts of mealybugs were retrieved from GenBank and aligned with ClustalW [35]. Later edition with Gblocks 0.91b [37] yielded a total of 1389 unambiguously aligned sites. Phylogenetic analyses were performed by ML, Maximum Parsimony (MP) and Bayesian methods using RAxML [38], DNAPARS from PHYLIP v3.69 package [36], and MrBayes 3.2 [39], respectively. A separate general time-reversible evolutionary model with gamma-distributed rates and a proportion of invariant sites (GTR+I+G) was applied for ML and Bayesian phylogenetic reconstructions, according to inferences by JModelTest [40,41]. ML and MP reconstructions included a 1000-replications bootstrap analysis. On the other hand, Bayesian reconstruction was generated from two runs of 150,000 generations. Likelihood settings were set to nst=6, rates=invgamma and ngammacat=4. Sampling was performed every 100 generations. First 3,400 generations were discarded as "burn

in". Figure on phylogenetic analysis was prepared using FigTree v1.4.0 software (http://tree.bio.ed.ac.uk/software/figtree/) and Inkscape (https://inkscape.org/es/).

The complete genomes of *"Ca.* Tremblaya princeps*"* PCVAL and *"Ca.* Moranella endobia*"* PCVAL were scanned with plugin "Find Repeats" from Unipro UGENE v1.12.2 [42] to identify DNA repeats. Direct and inverted repeats (DR and IR, respectively) of at least 20 nucleotides in length were analyzed. GC content of independent repeats was calculated with MEGA5 [43].

Perfect tandem repeat sequences with a total length of at least 18 bp and a unit size of up to 40 bp were identified using Phobos v3.3.12 [44] in the genomes of "*Ca.* Portiera aleyrodidarum" BT-B, *"Ca.* Tremblaya princeps*"* PCVAL, *"Ca.* Moranella endobia*"* PCVAL, and *"Ca.* Tremblaya phenacola*"* PAVE.

Gene screening

Complete sequences of HR-related genes *recA* (encoding the recombination protein RecA), *recG* (encoding the ATP-dependent DNA helicase RecG, EC:3.6.4.12), *ruvA* (encoding the Holliday junction DNA helicase RuvA, EC:3.6.4.12), *ruvB* (encoding the Holliday junction DNA helicase RuvB, EC:3.6.4.12), *ruvC* (encoding the crossover junction endodeoxyribonuclease RuvC, EC:3.1.22.4), and *priA* (encoding the primosomal protein N', EC:3.6.4.-) were retrieved from GenBank for a set of selected betaproteobacteria (*Burkholderia glumae* BGR1, *B. multivorans* ATCC 17616, *B. pseudomallei* 1106a, and *B. thailandensis* E264), and gammaproteobacteria (*Escherichia coli* K-12 MG1655, *Salmonella enterica* Typhimurium LT2, *Serratia proteomaculans* 568, *Sodalis glossinidius*, *Dickeya dadantii* 3937, *Yersinia pestis* Angola, and *"Ca.* Moranella endobia" PCVAL). Multiple alignments were performed in ClustalW [35] in order to identify conserved motifs where degenerate primers for PCR amplification and sequencing could be designed (Table S1). The annealing temperature in the PCR amplifications was 52ºC. Most of the primer pairs (named as BG) are expected to amplify both beta- and gammaproteobacterial homologs of the targeted *locus*. Primers G-ruvAF/R were designed on a multiple alignment including gamma-endosymbiont homologs only, in order to obtain a *ruvA* sequence larger than the one amplified by BG-ruvAF/R. BLAST searches against the non-redundant protein database (http://blast.ncbi.nlm.nih.gov/Blast.cgi/) [34] were performed to identify the putative taxonomic origin of the obtained sequences.

RESULTS AND DISCUSSION

The ancestral duplicated ribosomal genomic region in "*Ca*. Tremblaya*"*

One of the unusual genomic features of *"Ca.* Tremblaya princeps*"* from *P. citri* (cluster E) is the presence of an identical 5702-bp partial genome duplication. It includes the complete ribosomal operon (16S, 23S and 5S rRNA genes), plus the 3' region of *leuA* and *rpsO* upstream

the ribosomal operon, and the 5' region of *rsmH* downstream (Figure 1). The same duplicated region had been identified in strains from several clusters belonging to different mealybug species from subfamily Pseudococcinae: *D. brevipes* (cluster A), *M. albizziae* (cluster C), and *M. australiensis* and *M. hirsutus* (cluster F) [6,24]. In order to determine if the duplication of the ribosomal operon took place before the divergence of the two *"Ca.* Tremblaya*"* lineages, we performed a comparative genomic analysis of the complete genomes of *"Ca.* Tremblaya princeps*"* PCVAL [8] and *"Ca.* Tremblaya phenacola*"* PAVE [21]. The analysis revealed the presence of an identical 385-bp inverted duplication in the latter. It contains the 3' end of a pseudogenized 23S rRNA gene (not annotated originally in the genome), the 5S rRNA gene and the TPPAVE_188 pseudogene, which corresponds to a truncated paralog of *rsmH* (Figure 1). This result suggests that the duplication of the ribosomal operon took place before the split of the two *"Ca.* Tremblaya*"* lineages. Based on the genomic context of the ribosomal genes in both genomes, it can be deduced that the original copy of the ribosomal operon has undergone massive decay in *"Ca.* Tremblaya phenacola*"* PAVE, while the two identical copies found in *"Ca.* Tremblaya princeps*"* PCVAL have been evolving under concerted evolution.

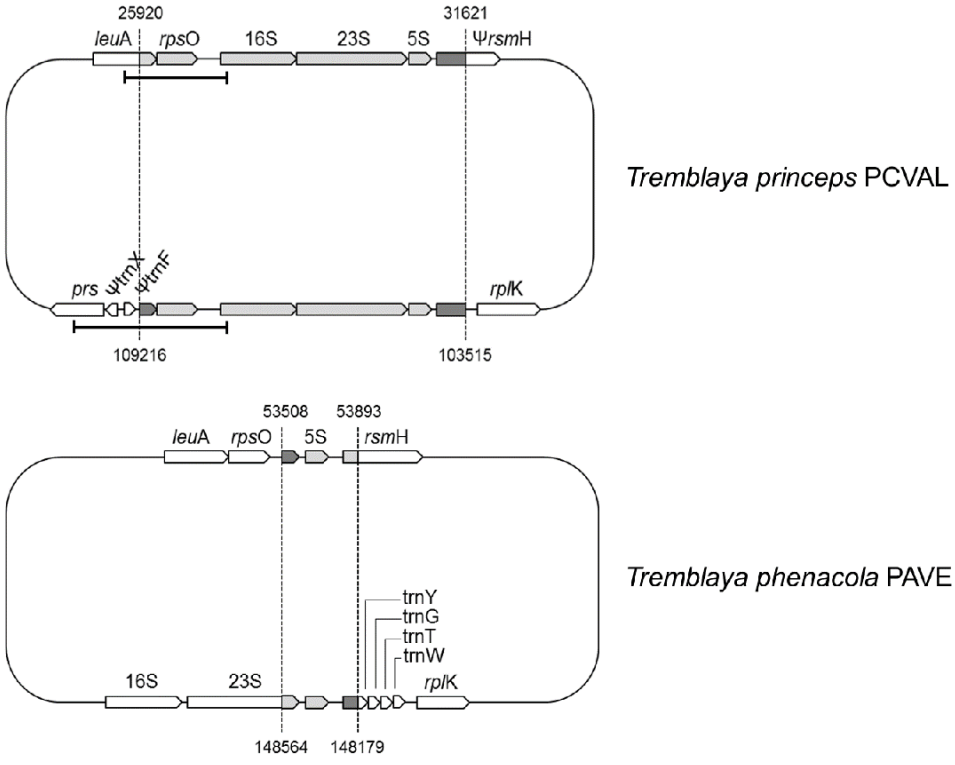


Figure 1. Duplicated genomic regions in *"Ca.* Tremblaya princeps*"* PCVAL and *"Ca.* Tremblaya phenacola*"* PAVE. The identical regions coordinates within each genome are indicated. Genes in the duplicated region appear in grey. Dark grey represent pseudogenes. The bars in the PCVAL genome indicate the regions that have been amplified in several *"Ca.* Tremblaya princeps*"* strains in this work.

Searching for signals of concerted evolution in Pseudococcinae endosymbiotic systems

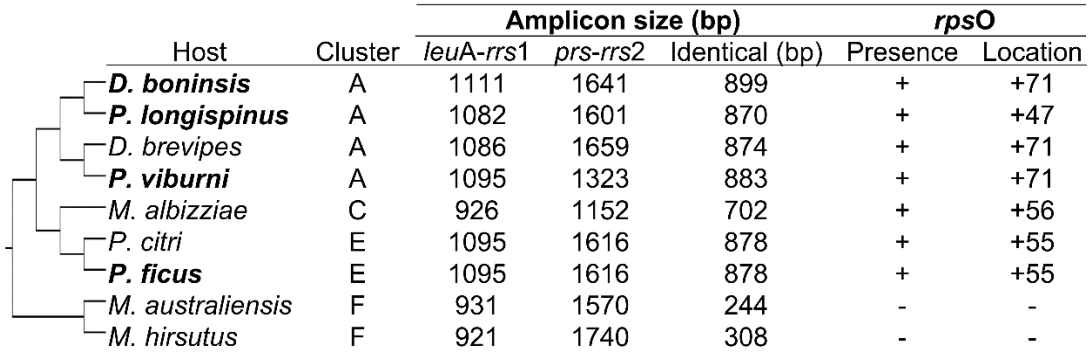
Concerted evolution is a molecular process that leads to homogenization of duplicated DNA sequences within one species. Consequently, paralogous *loci* are more closely related to each other than to the corresponding orthologous region in another species, even though the duplication event preceded the speciation event. DNA recombination mechanisms are responsible for this universal biological phenomenon first discovered in gene families.

As it is typical for most endosymbionts with reduced genomes, the two available genomes for *"Ca.* Tremblaya*"* species lack genes involved in DNA recombination and repair. In contrast, an almost complete set of genes involved in HR were annotated in the genome of *"Ca.* Moranella endobia*"* from *P. citri*. Therefore, these genes could be on the root of the link between the nested organization of the endosymbiotic system and the atypical characteristics of the *"Ca.* Tremblaya princeps*"* genome [21]. In order to test this hypothesis, we searched for signs of concerted evolution and the presence of HR-related genes in four mealybug species from subfamily Pseudoccinae (*D. boninsis, P. longispinus, P. viburni* and *P. ficus*).

Concerted evolution in "Ca. Tremblaya princeps". To search for signals of concerted evolution, we have focused on the molecular analysis of the 5'-flanking regions of the duplicated ribosomal operons (*leuA-rrs1* and *prs-rrs2*). In addition to the duplications already described in *"Ca.* Tremblaya princeps*"* from different clusters [6], we have included in our analysis the PCR-amplified sequences of four additional strains belonging to clusters A (from *D. boninsis*, *P. viburni* and *P. longispinus*) and E (from *P. ficus*). The amplicons include the homologous regions of the last 251 nucleotides of "*Ca*. Tremblaya princeps" PCVAL *leuA* or an almost complete *prs*, respectively, as well as the complete sequence of *rpsO,* several tRNA genes and the 5'-end of the 16S rRNA gene (Figures 1 and 2). The results obtained fit with our predictions. The alignment of the newly amplified sequences revealed the existence of identical paralogous fragments ranging from 870 bp in *"Ca.* Tremblaya princeps*"* strain PLON (beta-endosymbiont of *P. longispinus*) to 899 bp in strain DBON (beta-endosymbiont of *D. boninsis*). The length of regions putatively affected by concerted evolution remains relatively homogeneous (702-899 bp) among *"Ca.* Tremblaya princeps*"* lineages from clusters A, E and C. In coherence with their evolutionary relationships, identical duplicated *loci* start at orthologous positions for all available members of cluster A (nucleotide 25915/109218 in *"Ca.* Tremblaya princeps*"* PCVAL) and cluster E (nucleotides 25920/109216 in strain PCVAL; Figure 1), respectively. This is in contrast with the dramatic reduction previously observed in *"Ca.* Tremblaya princeps*"* from *M. australiensis* and *M. hirsutus* [6], whose initial nucleotides are orthologous of sites 26557/108579 and 26387/108749 in strain PCVAL, respectively. No nested intracellular bacteria has been described in cluster F [24], where the perfectly conserved duplicated regions are significantly shorter.

P. citri and *P. ficus* are so closely related that they have been considered as cryptic species [45]. The comparison among the identical paralogous regions in the genomes of their "*Ca*.

Tremblaya princeps" strains revealed homogenization of polymorphisms between them. Four indels and (at least) 15 nucleotide substitutions have been detected (Table S2). In order to characterize the mutations leading to homogenized polymorphic sites, their ancestral state in the LCA of "*Ca*. Tremblaya princeps" of clusters C and E were inferred with over 95% probability. These results suggest that concerted evolution has acted very recently, at least in *"Ca.* Tremblaya princeps*"* from cluster E.



Host	Cluster	Amplicon size (bp)			*rpsO*	
		*leu*A-*rrs*1	*prs*-*rrs*2	Identical (bp)	Presence	Location
D. boninsis	A	1111	1641	899	+	+71
P. longispinus	A	1082	1601	870	+	+47
D. brevipes	A	1086	1659	874	+	+71
P. viburni	A	1095	1323	883	+	+71
M. albizziae	C	926	1152	702	+	+56
P. citri	E	1095	1616	878	+	+55
P. ficus	E	1095	1616	878	+	+55
M. australiensis	F	931	1570	244	-	-
M. hirsutus	F	921	1740	308	-	-

Figure 2. Characteristics of the *leu*A-*rrs*1 and *prs*-*rrs*2 regions. If present, *rpsO* location with respect to the beginning of the identical region is indicated. Host species from which sequences have been obtained in this work are in bold. The phylogenetic relationship among the insect hosts [22] is indicated.

Genetic screening of genes involved in homologous recombination. In order to evaluate the putative HR capability of the endosymbiotic systems from different Pseudococcinae species, we investigated the presence of a diverse set of genes in the four mealybug species under study. Screened *loci* include *recA*, *recG*, *ruvA*, *ruvB*, *ruvC*, and *priA*. All of them have been identified in the genome of "*Ca*. Moranella endobia" from *P. citri*. Most of them (*recA*, *recG, ruvA*, *ruvB*, *ruvC*) are common elements of both RecF and RecBCD pathways [10,11]. RecG may functionally replace RuvABC [46]. On the other hand, PriA is exclusively involved in the RecBCD pathway [47]. It has been proposed to catalyze the assembly of the "*Ca*. Moranella endobia" incomplete primosome [12].

The results are presented in Table 1, including the GenBank accession numbers for all newly amplified sequences. BLAST searches against the non-redundant protein database showed all detected genes to be of gammaproteobacterial origin, suggesting that the internalization of a nested endosymbiont have likely represented the acquisition of HR potential by the corresponding *"Ca.* Tremblaya princeps*"* lineages. Since gamma-endosymbionts from *P. ficus, P. longispinus* and *P. viburni* were acquired independently [25,28], these results indicate three putative events of acquisition of an HR machinery.

Although all primer combinations successfully amplify their target when applied to *P. citri* as a positive control, none of the analyzed consortia gave positive results for all the screened genes. Negative results should be interpreted with caution, since they do not necessarily imply the absence of undetected *loci*. Degenerate primers were designed on gene regions encoding highly conserved motifs among beta and gammaproteobacterial homologs of the analyzed genes (Table S1). However, although highly conserved between distantly related bacteria, motifs acting as primer templates are not directly involved in protein functionality. Therefore, it is possible that non-synonymous substitutions affecting the target sequence lead to false negative results. Nevertheless, in concordance with the close evolutionary relationship between *P. citri* and *P. ficus*, five of the six screened *loci* were detected in the latter. Only *priA* could not be detected. PriA is needed for the assembly of the primosome, which is already incomplete in "*Ca*. Moranella endobia" PCVAL, due to the loss of *dnaT* and *priC* [12]. Thus, its absence suggests a relatively recent inactivation of the RecBCD pathway in the nested endosymbiont of *"Ca.* Tremblaya princeps*"* strains from cluster E. In contrast, as revealed by the very recent homogenization of polymorphisms (Table S2), the RecF pathway appears to be still acting on this cluster. However, none of the components of the RecFOR complex, which enhances RecA loading onto SSB-coated single stranded DNA, are present in *"Ca.* Moranella endobia*"* [7,12]. Furthermore, none of the RecA mutations known to bypass the RecFOR complex deficiency (i.e., recA441, recA730, recA803) [55] were detected in this genome. Therefore, in the absence of the RecFOR complex, the RecF pathway is expected to be attenuated [56,57].

Table 1. Genetic screening of selected *loci* involved in homologous recombination. (+) detected gene; (-) non-detected gene.

	P. citri	***P. ficus***	***P. longispinus***	***P. viburni***	***D. boninsis***
recA	+	+ (KJ140122)	-	-	-
recG	+	+ (KJ140123)	-	-	-
ruvA	+	+ (KJ140120)	+ (KJ140121)	-	-
ruvB	+	+ (KJ140117)	+ (KJ140118)	+ (KJ140119)	-
ruvC	+	+ (KJ140124)	-	-	-
priA	+	-	+ (KJ140115)	+ (KJ140116)	-

As for the endosymbiotic consortia involving the three strains from cluster A under study, our results suggest that the RecF and RecBCD pathways are currently inactive. Different patterns of HR-related genes conservation were observed, which is coherent with the independent evolutionary origin of the gamma-endosymbionts (Table 1). None of the screened genes was found in *D. boninsis*. Cluster A is very wide, including endosymbionts from mealybugs of the tribe Pseudococcinni and the southern Africa group [22]. Additionally, both genera *Dysmicoccus* and *Pseudococcus* are polyphyletic, and the two species of each genus analyzed in this work are phylogenetically distant, belonging to different clades of the tribe Pseudococcinni. In order to place these three gamma-endosymbionts in the phylogenetic tree of those already described for mealybugs, we performed a phylogenetic analysis based on 16S rDNA sequences (Figure S1).

Our results show that the gamma-endosymbiont from *D. boninsis* belongs to cluster A, as expected. However, this is not the case for the gamma-endosymbionts from *P. longispinus* and *P. viburni.* Previous studies [25,28] indicated that these three gamma-endosymbionts have been independently acquired. The present analysis suggests that two independent replacement events have occurred in *P. longispinus* and *P. viburni.* The more ancient gamma-endosymbiont of *D. boninsis* is expected to be more affected by the genome reduction process typically observed in intracellular symbionts. Therefore, it is possible that the HR machinery is more degraded in this symbiont or that nucleotide changes have accumulated in the sequences targeted by the PCR primers, thus hampering amplification.

Susceptibility to homologous recombination in "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" from *P. citri*

Repeat sequences are the substrate for the HR machinery and are responsible of genomic instability. It has been shown that repeat density correlates positively with rates of chromosome rearrangement [58]. Recombination between DR causes DNA duplication or DNA deletion, while recombination between IR generates DNA inversion, which are usually less abundant than deletions. Repeats with very variable lengths are able to promote HR events [59,60]. In the endosymbionts of *P. citri*, it has been demonstrated that a 71-bp inverted repeat is able to promote recombination [7]. Smaller repeats, ranging from 18 to 24 bp have been suggested to be long enough to do so [20,59,61].

In order to analyze the susceptibility to HR of both *"Ca.* Tremblaya princeps*"* and *"Ca.* Moranella endobia*"* from *P. citri*, we performed a comprehensive search of direct and inverted repeats with a minimal length of 20 bp in the complete genomes of both *"Ca.* Tremblaya princeps*"* and "*Ca*. Moranella endobia" PCVAL (Table S3). We found 16 DR (TDR01 to 16) and 12 IR (TIR01 to 12) in *"Ca.* Tremblaya princeps*"*. Except for TIR12 (i.e., the duplicated region containing the ribosomal operon), all other repeats seem to have been randomly generated.

In "*Ca*. Moranella endobia", twenty-four DR (MDR01 to 24) and sixteen IR (MIR01 to 16) were found. Several of them are also the consequence of ancestral duplications. Thus, MDR01, MDR02 and MDR11 map on a functional *pdx*J (*locus* MPC_094 in the genome) and its pseudogenized copy (MPC_306), while MDR07, MDR14-16, MDR19, MDR22 and MDR23 are linked to a duplication including genes *sec*E (MPC_278) and *tuf* (MPC_279). Additionally, seven DR and five IR appear to derivate from tRNA genes proliferation events [42]. These results suggest that there is a close relationship among *loci* MPC_120 (*trn*M), MPC_418 (*trn*V), and MPC_317 (*trn*I); MPC_018 (*trn*P), MPC_350 (*trn*R), and MPC_453 (*trn*R); MPC_262 (*trn*F), MPC_175 (*trn*A), MPC_316 (*trn*A) and MPC_178 (*trn*K); MPC_041 (*trn*L) and MPC_118 (*trn*L); MPC_119 (*trn*Q) and MPC_121 (*trn*Q); and MPC_443 (*trn*S) and MPC_454 (*trn*S). Such repeats conservation might be due to mutational constraints, since 36-71% of their sequences correspond to tRNA stem regions (Table S4).

According to our results, repeats are larger in "*Ca.* Tremblaya princeps" (mean length =235.9 bp) than in *Ca*. Moranella endobia" (mean length =127.6 bp). Moreover, some of them (TDR8 and TDR12; TIR05 and TIR07) appear to derive from larger ancestral repeats. They are also more abundant, with a repeats density (abundance/kb) 2.85 times larger in the former (Table S3). This is likely due to a higher bias affecting the nucleotide composition of the *"Ca.* Tremblaya princeps*"* genome. The molecular characterization of the independently generated repeats identified in these genomes revealed that those of *"Ca.* Tremblaya princeps*"* are GC-enriched compared with the whole genome ($GC_{repeats}$=67.7% versus GC_{genome}=58.8%, SD_{G+C}=6.5), while no bias is observed in the case of "*Ca*. Moranella endobia" ($GC_{repeats}$=42.4% versus GC_{genome}=43.5%, SD_{G+C}=11.0) (Figure 3).

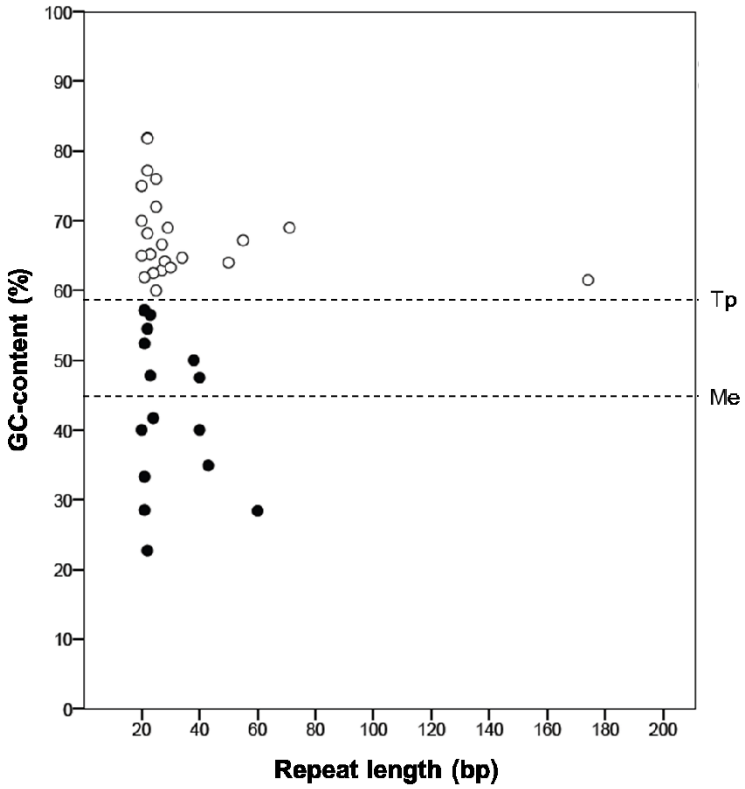


Figure 3. Molecular characterization of independent repeats identified in the genomes of *"Ca.* Tremblaya princeps*"* (white circles) and *"Ca.* Moranella endobia*"* (black circles). The horizontal lines indicate the mean GC-content of each genome.

In any case, HR events are not expected to be highly frequent at least in the *"Ca.* Tremblaya princeps*"* genome, where the mean distance between DR is about 50 kb (36% of the chromosome), because further genome reduction derived from such events would be strongly deleterious and, therefore, under negative selection. As for rearrangements mediated by IR, half of such repeats map on functional *loci*, including genes involved in translation (*rplS*, *rpsF*, *rpmA*)

and essential amino acids biosynthesis (*pheA*, *ilvI*, *leuA*, *aroB*) (Table S3). Thus, apparent inactivation of the HR pathways in endosymbionts from *D. boninsis*, *P. longispinus* and *P. viburni* and its attenuation in cluster E may lead to an increased stability of the corresponding bacterial consortia.

"*Ca*. Tremblaya princeps" gene density: low and heterogeneous

Another peculiarity of the *"Ca.* Tremblaya princeps*"* genome is its low gene density. In order to get a more comprehensive picture of this feature, we performed a comparative analysis of IGR length among endosymbionts with highly reduced genomes (Figure 4). We used in our comparison the strain PCVAL of both "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" plus the representative species used by McCutcheon and Moran (2012) [5] in their review on extreme genome reduction in symbiotic bacteria. Additionally, we have included in our analysis the recently sequenced genomes of *"Ca.* Tremblaya phenacola*"* PAVE [21], the closest relative to *"Ca.* Tremblaya princeps*"* whose genome has been sequenced; "*Candidatus* Nasuia deltocephalinicola" NAS-ALF (primary endosymbiont of the deltocephaline leafhopper *Macrosteles quadrilineatu*) [49], the smallest sequenced bacterial genome; and "*Ca.* Portiera aleyrodidarum" BT-B [50], which displays atypical genomic features similar to those found in *"Ca.* Tremblaya princeps*"*.

Almost all analyzed endosymbionts present highly compact genomes. Although the amount of detected pseudogenes and gene imbrication frequency depend on annotation criteria, there are clear differences between reduced and tiny genomes, likely due to minimization of IGRs (Figure 4). Tiny genomes (smaller than 300 kb) [5], with none or very few pseudogenes and extremely reduced IGRs, represent the most extreme cases. No pseudogenes have been annotated in the genomes of "*Ca*. Nasuia deltocephalinicola", "*Ca.* Sulcia muelleri" [51] and "*Ca.* Carsonella rudii" [52], while only one has been identified in "*Ca.* Hodkinia cidadicola" [53] and "*Ca.* Zinderia insecticola" [54]. IGR mean size ranges from 33 bp ("*Ca*. Nasuia deltocephalinicola") to 83 bp ("*Ca.* Zinderia insecticola"). The coding capacity of these tiny bacterial genomes increases by the presence of overlapping events, which in "*Ca.* Carsonella ruddii" (97% gene density) represents the 65%. Similarly, *"Ca.* Tremblaya phenacola*"* PAVE presents typical characteristics of a tiny genome, both regarding coding density (93.5%), and IGR length homogeneity (79 bp, 122.2 SD) [21]. On the contrary, while 34% overlapping events have been annotated in *"Ca.* Tremblaya princeps*"* PCVAL, placing it in the range of other analyzed tiny genomes, its IGR mean size is 467 bp, and they present a huge length heterogeneity, from 2304 pb (*dna*N-*tal*B) to -257pb (*rps*L-*rpo*C), with more than 30% of IGRs (29/86) larger than 450 bp. Since only 62% of them (18/29) include inactivated *loci*, low gene density cannot be explained exclusively by pseudogenes abundance [7,8].

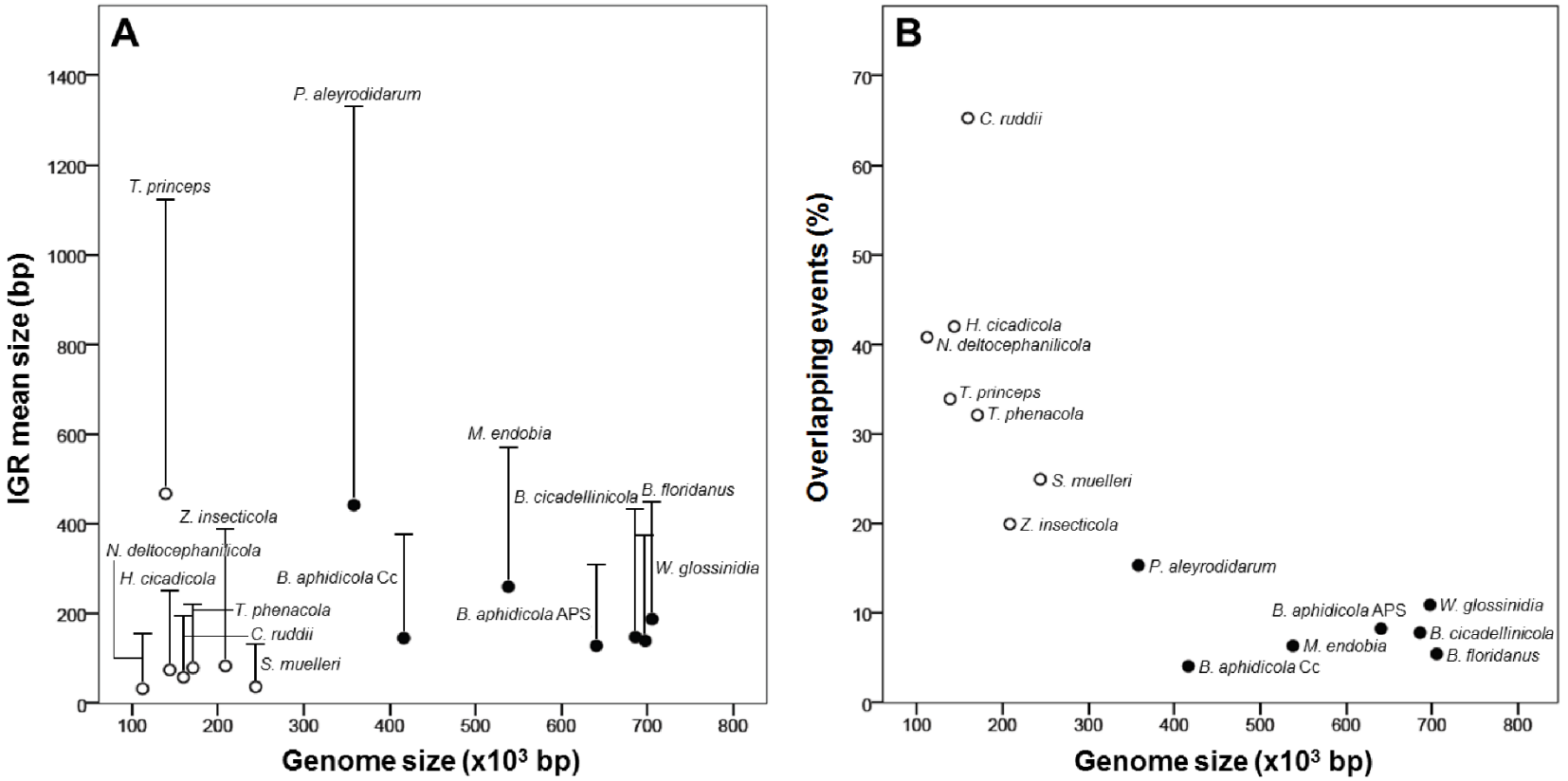


Figure 4. Trends in structural genome evolution of obligate endosymbionts. IGR characteristics (A) and abundance of overlapping genes (B) have been studied for a set of reduced (black) and tiny (white) bacterial genomes. Error bars represent standard deviation (SD) for IGR lengths at the corresponding genomes.

"*Ca.* Portiera aleyrodidarum*"* BT-B resembles *"Ca.* Tremblaya princeps*"* PCVAL regarding these two genome features. Low gene density is known to be caused by the expansion of short tandem repeats in "*Ca.* Portiera aleyrodidarum*"* BT-B [20]. The authors suggest that it is the consequence of the loss of *dna*Q (subunit epsilon of the DNA polymerase III), which is nearly universal among bacteria and is responsible for 3' to 5' exonuclease proofreading during DNA replication. Mutations on this gene lead to severe replication errors, including the increase of the instability of microsatellite regions [62]. This is not the case in *"Ca.* Tremblaya princeps*"* PCVAL, where a functional *dna*Q gene has been annotated. In coherence, the analysis of this type of repeats in the two *P. citri* endosymbionts (*"Ca.* Tremblaya princeps*"* and "*Ca*. Moranella endobia"), as well as in *"Ca.* Tremblaya phenacola*"* PAVE (Figure 5) revealed the scarcity of tandem repeats in these mealybug endosymbionts.

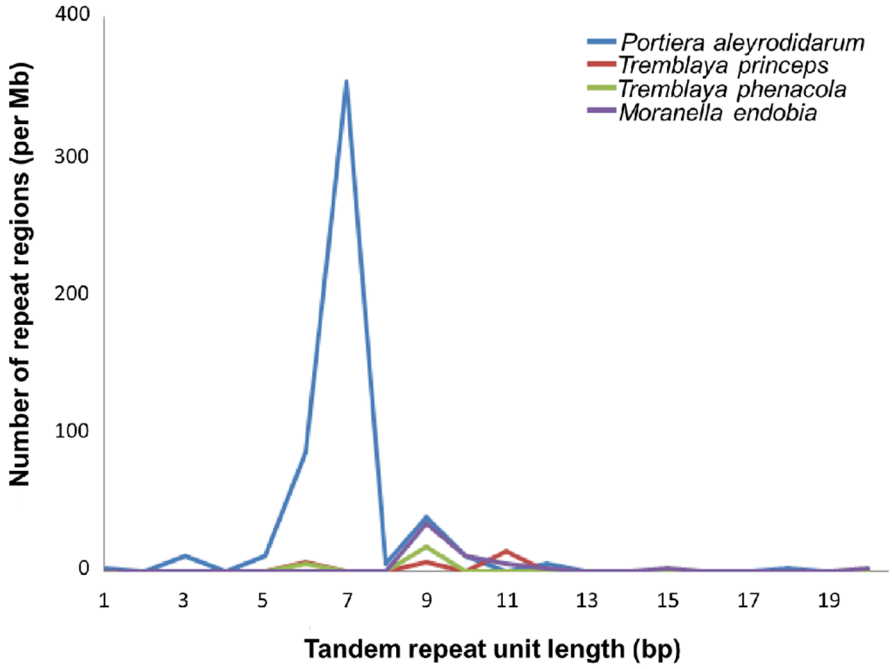


Figure 5. Short tandem repeats analysis. Genomic tandem repeat sequences with a total length of at least 18 bp and unit lengths of 1-20 bp are represented.

In summary, taking into account our results regarding the apparently ongoing inactivation of the HR mechanisms, and the big size of the regions flanked by DR in the *"Ca.* Tremblaya princeps*"* PCVAL genome, we propose that the acquisition of functional HR mechanisms throughout gamma-endosymbionts internalization could have led, not only to genomic instability and concerted evolution of paralogous *loci*, but also to the atypical low gene density showed by *"Ca.* Tremblaya princeps*"*. Genome shuffling mediated by IRs inside non-essential genes would have accelerated pseudogenization. The negative selection putatively affecting further HR events mediated by DR would impede the deletion of the regions containing these pseudogenes, which

would accumulate mutations until no signal of the ancient *loci* would be detected. The future metagenomics characterization of nested consortia from unexplored Pseudococcinae species would allow testing our hypothesis of a link between the HR potential, mediated by independently acquired gamma-endosymbionts, and the atypical "*Ca.* Tremblaya princeps" reductive genome evolution.

ACKNOWLEDGMENTS

We thank Dr. Antonia Soto and Dr. Aleixandre Beltrà, from the Mediterranean Agroforestal Institute (Universitat Politècnica de València) for kindly supplying the biological samples. This work was supported by grants BFU2012-39816-C02-01 (co-financed by FEDER funds and Ministerio de Economía y Competitividad, Spain) to A. Latorre and PrometeoII/2014/065 (Conselleria d'Educació, Generalitat Valenciana, Spain) to A. Moya. S. López-Madrigal is a recipient of a fellowship from the Ministerio de Educación (Spain).

REFERENCES

1. Abby S, Daubin V (2007) Comparative genomics and the evolution of prokaryotes. Trends Microbiol 15: 135–141. doi:10.1016/j.tim.2007.01.007

2. Casjens S (1998) The diverse and dynamic structure of bacterial genomes. Annu Rev Genet 32: 339–377. doi:10.1146/annurev.genet.32.1.339

3. Rocha EPC (2008) The organization of the bacterial genome. Annu Rev Genet 42: 211–233. doi:10.1146/annurev.genet.42.110807.091653

4. Moya A, Peretó J, Gil R, Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9: 218–229. doi:10.1038/nrg2319

5. McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10: 13–26. doi:10.1038/nrmicro2670

6. Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P (2002) The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. Appl Environ Microbiol 68: 3198–3205. doi:10.1128/AEM.68.7.3198

7. McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr Biol 21: 1366–1372. doi:10.1016/j.cub.2011.06.051.

8. López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R (2011) Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. J Bacteriol 193: 5587–5588. doi:10.1128/JB.05749-11

9. von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. Nature 412: 433–436. doi:10.1038/35086563

10. Spies M, Kowalczykowski SC (2005) Homologous recombination by the RecBCD and the RecF pathways. In: Higgins NP, editor. The bacterial chromosome: ASM Press. pp. 389–403.

11. Rocha EPC, Cornet E, Michel B (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet 1: e15. doi:10.1371/journal.pgen.0010015

12. López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R (2013) Mealybugs nested endosymbiosis: going into the "matryoshka" system in *Planococcus citri* in depth. BMC Microbiol 13: 74. doi:10.1186/1471-2180-13-74

13. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407: 81–86. doi:10.1038/35024074

14. Tamas I, Klasson L, Canbäck B, Näslund a K, Eriksson A-S, et al. (2002) 50 million years of genomic stasis in endosymbiotic bacteria. Science 296: 2376–2379. doi:10.1126/science.1071278

15. van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, et al. (2003) Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci U S A 100: 581–586. doi:10.1073/pnas.0235981100

16. Pérez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, et al. (2006) A small microbial genome: the end of a long symbiotic relationship? Science 314: 312–313. doi:10.1126/science.1130441

17. Lamelas A, Gosalbes MJ, Moya A, Latorre A (2011) New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the aphid *Cinara tujafilina*. Appl Environ Microbiol 77: 4446–4454. doi: 10.1128/AEM.00141-11

18. Degnan PH, Ochman H, Moran NA (2011) Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont Buchnera. PLoS Genet 7: e1002252. doi:10.1371/journal.pgen.1002252

19. Jiang Z, Jones DH, Khuri S, Tsinoremas NF, Wyss T, et al. (2013) Comparative analysis of genome sequences from four strains of the *Buchnera aphidicola* Mp endosymbion of the green peach aphid, Myzus persicae. BMC Genomics 14: 917. doi:10.1186/1471-2164-14-917

20. Sloan DB, Moran NA (2013) The evolution of genomic instability in the obligate endosymbionts of whiteflies. Genome Biol Evol 5: 783–793. doi:10.1093/gbe/evt044

21. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, et al. (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153: 1567–1578. doi:10.1016/j.cell.2013.05.040

22. Hardy NB, Gullan PJ, Hodgson CJ (2008) A subfamily-level classification of mealybugs (Hemiptera: Pseudococcidae) based on integrated molecular and morphological data. Syst Entomol 33: 51–71. doi:10.1111/j.1365-3113.2007.00408.x

23. Gruwell ME, Hardy NB, Gullan PJ, Dittmar K (2010) Evolutionary relationships among primary endosymbionts of the mealybug subfamily phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). Appl Environ Microbiol 76: 7521–7525. doi:10.1128/AEM.01354-10

24. Thao M, Gullan P, Baumann P (2002) Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl Environ Microbiol 68: 3190–3197. doi:10.1128/AEM.68.7.3190

25. Gatehouse LN, Sutherland P, Forgie SA, Kaji R, Christeller JT (2012) Molecular and histological characterization of primary (betaproteobacteria) and secondary (gammaproteobacteria) endosymbionts of three mealybug species. Appl Environ Microbiol 78: 1187–1197. doi:10.1128/AEM.06340-11

26. Koga R, Nikoh N, Matsuura Y, Meng XY, Fukatsu T (2013) Mealybugs with distinct endosymbiotic systems living on the same host plant. FEMS Microbiol Ecol 83: 93–100. doi:10.1111/j.1574-6941.2012.01450.x

27. López-Madrigal S, Balmand S, Latorre A, Heddi A, Moya A, et al. (2013) How does *Tremblaya princeps* get essential proteins from its nested partner *Moranella endobia* in the mealybug *Planoccocus citri*? PLoS One 8: e77307. doi:10.1371/journal.pone.0077307

28. López-Madrigal S, Beltrà A, Resurrección S, Soto A, Latorre A, et al. (2014) Molecular evidence for ongoing complementarity and horizontal gene transfer in endosymbiotic systems of mealybugs. Front Microbiol 5: 449. doi:10.3389/fmicb.2014.00449

29. Kono M, Koga R, Shimada M, Fukatsu T (2008) Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. Appl Environ Microbiol 74: 4175–4184. doi:10.1128/AEM.00250-08

30. Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Methods Mol Biol 132: 115–130. doi:10.1385/1-59259-192-2:115

31. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. Bioinformatics 16: 944–945. doi:10.1093/bioinformatics/16.10.944

32. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964. doi:10.1093/nar/25.5.0955

33. Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32: 11–16. doi:10.1093/nar/gkh152

34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402. doi:10.1093/nar/25.17.3389

35. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948. doi:10.1093/bioinformatics/btm404

36. Felsenstein J (2005) Using the quantitative genetic threshold model for inferences between and within species. Philos Trans R Soc Lond B Biol Sci 360: 1427–1434. doi:10.1098/rstb.2005.1669

37. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552.

38. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690. doi:10.1093/bioinformatics/btl44637.

39. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539–542. doi:10.1093/sysbio/sys029

40. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696–704. doi:10.1080/10635150390235520

41. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9: 772. doi:10.1038/nmeth.2109

42. Okonechnikov K, Golosova O, Fursov M (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28: 1166–1167. doi:10.1093/bioinformatics/bts091

43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739. doi:10.1093/molbev/msr121

44. Kraemer L, Beszteri B, Gäbler-Schwarz S, Held C, Leese F, et al. (2009) STAMP: Extensions to the STADEN sequence analysis package for high throughput interactive microsatellite marker design. BMC Bioinformatics 10: 41. doi:10.1186/1471-2105-10-41

45. Kol-Maimon H, Ghanim M, Franco JC, Mendel Z (2014) Evidence for gene flow between two sympatric mealybug species (Insecta; Coccoidea; Pseudococcidae). PLoS One 9: e88433. doi:10.1371/journal.pone.0088433

46. Meddows TR, Savory AP, Lloyd RG (2004) RecG helicase promotes DNA double-strand break repair. Mol Microbiol 52: 119–132. doi:10.1111/j.1365-2958.2003.03970.x

47. Ng JY, Marians KJ (1996) The ordered assembly of the FX174-type primosome I. Isolation and identification of intermediate protein-DNA complexes. J Biol Chem 271: 15642–15648. doi:10.1074/jbc.271.26.15642

48. Withers M, Wernisch L, dos Reis M (2006) Archaeology and evolution of transfer RNA genes in the *Escherichia coli* genome. RNA 12: 933–942. doi:10.1261/rna.2272306

49. Bennett GM, Moran N a (2013) Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. Genome Biol Evol 5: 1675–1688. doi:10.1093/gbe/evt118

50. Sloan DB, Moran NA (2012) Endosymbiotic bacteria as a source of carotenoids in whiteflies. Biol Lett 8: 986–989. doi:10.1098/rsbl.2012.0664

51. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, et al. (2010) One bacterial cell, one complete genome. PLoS One 5: e10314. doi:10.1371/journal.pone.0010314

52. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, et al. (2006) The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314: 267. doi:10.1126/science.1134196

53. McCutcheon JP, McDonald BR, Moran NA (2009) Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. PLoS Genet 5: e1000565. doi:10.1371/journal.pgen.1000565

54. McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol Evol 2: 708–718. doi:10.1093/gbe/evq055

55. Lavery PE, Kowalczykowski SC (1992) Biochemical basis of the constitutive repressor cleavage activity of recA730 protein. A comparison to recA441 and recA803 proteins. J Biol Chem 267: 20648–20658.

56. Morimatsu K, Kowalczykowski SC (2003) RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell 11: 1337–1347. doi:10.1016/S1097-2765(03)00188-6

57. Handa N, Morimatsu K, Lovett ST, Kowalczykowski SC (2009) Reconstitution of initial steps of dsDNA break repair by the RecF pathway of E. coli. Genes Dev 23: 1234–1245. doi:10.1101/gad.1780709

58. Rocha EPC (2003) DNA repeats lead to the accelerated loss of gene order in bacteria. Trends Genet 19: 600–603. doi:10.1016/j.tig.2003.09.011

59. Aras RA, Kang J, Tschumi AI, Harasaki Y, Blaser MJ (2003) Extensive repetitive DNA facilitates prokaryotic genome plasticity. Proc Natl Acad Sci U S A 100: 13579–13584. doi:10.1073/pnas.1735481100

60. Achaz G, Coissac E, Netter P, Rocha EPC (2003) Associations between inverted repeats and the structural evolution of bacterial genomes. Genetics 164: 1279–1289.

61. Shen P, Huang H V (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 112: 441–457.

62. Zahra R, Blackwood JK, Sales J, Leach DRF (2007) Proofreading and secondary structure processing determine the orientation dependence of CAG x CTG trinucleotide repeat instability in *Escherichia coli*. Genetics 176: 27–41. doi:10.1534/genetics.106.069724

SUPPLEMENTARY MATERIAL

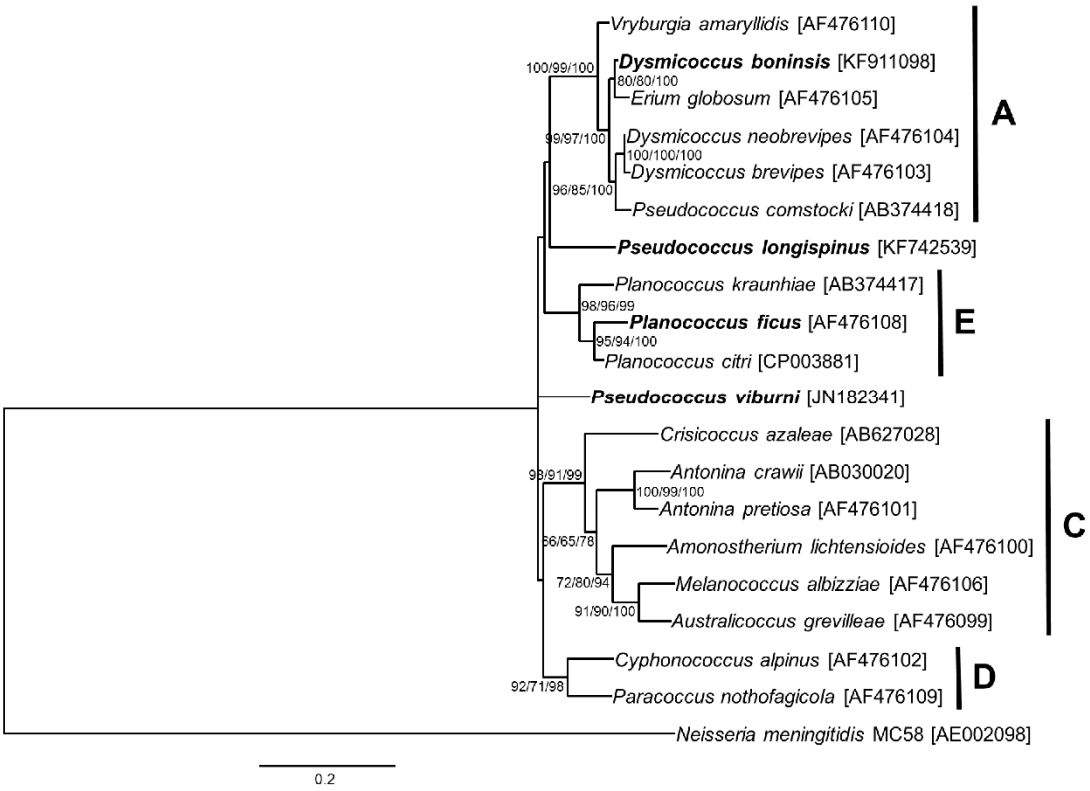
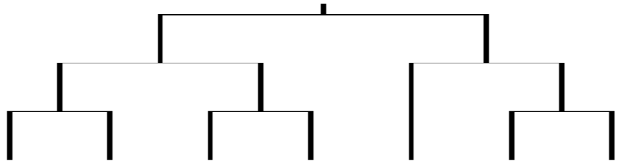


Figure S1. Phylogenetic relationships among gamma-endosimbionts of Pseudococcinae mealybugs. Already defined clusters A, C, D and E are represented. Host species used in this work are in bold. The betaproteobacterium *Neisseria meningitidis* MC58 was used as outgroup. ML, MP and Bayesian analysis gave essentially the same results. ML and MP bootstrap values, and Bayesian posterior probabilities over 50% are indicated. Scale bar represents substitutions per site.

Table S1. Degenerate primers designed for the gene screening. Primers degeneration (DG) was calculated as the proportion of ambiguous sites. The conserved motifs position refers to the *Escherichia coli* homologous proteins.

Gene	Primer	Sequence (5´→3´)	Length (nt)	DG	Conserved motifs	Position
recA	BG-recAF	TTYGGNAARGGNWSNATHATG	21	38%	FGKGSIM	22-28
	BG-recAR	RTANSWRTACCANGCNCCNGC	21	38%	AGAWYSY	288-294
recG	BG-recGF	MWNGAYYTNBTNYTNCAYYTNCC	23	56%	Q/IDLL/VLHLP	34-41
	BG-recGR	GGNCCNCKDATYTCNARRTC	20	40%	DLEIRGP	626-632
ruvA	BG-ruvAF	GARGAYGCNCANYTNYTNTWYGG	23	43%	EDAQ/HLLY/FG	55-62
	BG-ruvAR	ARNCKYTCNGCNGTYTTYTTNCC	23	39%	GKKTAERL	117-124
	G-ruvAF	ACNTGYTTYTAYGMNYTNCC	20	40%	TCFYE/ALP	33-39
	G-ruvAR	ACNADNCKNSHNGCYTCYTG	20	50%	QEAS/GRM/LV	175-181
ruvB	BG-ruvBF	CCNCCNGGNYTNGGNAARACNAC	23	34%	PPGLGKTT	63-70
	BG-ruvBR	KNGGNGTNCKYTGNADRAANCC	22	45%	GFL/IQRTPR	309-316
ruvC	BG-ruvCF	GAYCCNGGNHBNMGNRTNACNGG	23	47%	DPGS/LRV/ITG	8-15
	BG-ruvCR	TGNSWDATNGCNRTNSCNARNGC	23	52%	ALA/GI/T/MAIT/CH	143-150
priA	BG-priAF	GGNRTNACNGGNWSNGGNAARAC	23	43%	GV/ITGSGKT	224-231

Table S2. Homogenized polymorphisms in "*Ca*. Tremblaya princeps" PCVAL and PFIC. The genome position indicates the corresponding nucleotide along the "*Ca*. Tremblaya princeps" PCVAL genome. The state of orthologous sites in the analyzed strains from clusters A, C and E as well as the most likely nucleotide in the LCA of clusters C and E is represented. A cladogram depicting the evolutionary relationships of the studied bacteria is also included.



Clade A				Clade C	Clade E		LCA (C, E)	Genome position	
DBON	PLON	DBRE	PVIB	MALB	PFIC	PCVAL		*leuA*-rrs1	*prs*-rrs2
G	-	G	G	G	C	A	**G**	25955	109181
C	-	C	C	C	C	G	**C**	25957	109179
T	-	T	T	T	T	G	**T**	25961	109175
C	-	C	C	C	C	A	**C**	25962	109176
T	T	A	T	T	T	G	**T**	26057	109079
T	T	T	T	T	C	G	**T**	26102	109034
G	G	G	G	G	C	A	**G**	26106	109030
G	G	G	G	G	T	A	**G**	26107	109029
A	A	-	A	-	A	T	**A**	26417	108719
T	T	T	G	-	T	A	**T**	26437	108699
G	G	G	C	-	G	C	**G**	26438	108698
T	T	T	A	-	T	A	**T**	26439	108697
C	C	C	C	-	-	C	**C**	26479	108657
A	A	A	A	-	A	-	**A**	26479-80	108656-7
G	G	G	G	-	G	-	**G**	26479-80	108656-7
C	C	C	C	-	-	C	**C**	26480	108656
C	C	C	C	-	A	C	**C**	26482	108654
T	T	T	T	-	T	G	**T**	26494	108642
T	T	T	T	T	G	T	**T**	26526	108610

Table S3. Repetitive sequences detected in the genomes of the *P. citri* endosymbionts. Sequences have been sorted by size. Information about their position and distribution (distance), as well as if they map on coding (gene name) or non-coding (-) regions, is provided.

Organism	Type	ID	Position 1	Position 2	Size (bp)	Distance (bp)	Context
***"Ca.* Tremblaya princeps"**	DR	TDR01	776-795	55251-55270	20	54455	*-/lysC*
		TDR02	3058-3077	53837-53856	20	50759	*-/-*
		TDR03	16248-16267	97630-97649	20	81362	*-/rpoC*
		TDR04	44279-44300	79739-79760	22	35438	*-/rpoA*
		TDR05	58673-58694	59973-59994	22	1278	*-/-*
		TDR06	6373-6395	21488-21510	23	15092	*-/dnaE*
		TDR07	63473-63495	126023-126045	23	62527	*-/clpX*
		TDR08	58059-58082	121811-121834	24	63728	*trnK1/-*
		TDR09	101936-101959	121556-121579	24	19596	*rpoB/-*
		TDR10	69981-70005	111177-111201	25	41171	*engA/-*
		TDR11	15006-15032	121818-121844	27	106785	*trnW/rpoD*
		TDR12	58103-58129	121833-121859	27	63703	*trnK1/-*
		TDR13	51226-51253	56851-56878	28	5597	*aroC/-*
		TDR14	73171-73220	130598-130647	50	57377	*trnK2/-*
		TDR15	47369-47423	69420-69474	55	21996	*infB/engA*
		TDR16	8141-8314	128322-128495	174	120007	*-/argG*
	IR	TIR01	22754-22773	43644-43663	20	20870	*ilvI/rpsF*
		TIR02	45099-45119	52235-52255	21	7115	*-/-*
		TIR03	52182-52202	110655-110675	21	58452	*-/-*
		TIR04	13832-13853	14058-14079	22	204	*rplS/rplS*
		TIR05	18479-18500	60398-60419	22	41897	*-/-*
		TIR06	10513-10537	66135-66159	25	55597	*pheA/-*
		TIR07	18448-18472	60427-60451	25	41954	*-/-*
		TIR08	78375-78403	121525-121553	29	43121	*aroB/-*
		TIR09	34369-34398	50394-50423	30	15995	*-/-*
		TIR10	32831-32864	60265-60298	34	27400	*rpmA/-*
		TIR11	56533-56603	63636-63706	71	7032	*-/-*
		TIR12	25920-31621	103515-109216	5702	71893	*leuA/-*
***"Ca.* Moranella endobia"**	DR	MDR01	113273-113292	338282-338301	20	224989	*pdxJ/-*
		MDR02	113332-113351	338341-338360	20	224989	*pdxJ/-*
		MDR03	138725-138744	138745-138764	20	-	-
		MDR04	141407-141426	141625-141644	20	198	*trnQ1/trnQ2*
		MDR05	141448-141467	141666-141685	20	198	*trnQ1/trnQ2*
		MDR06	289025-289044	347776-347795	20	58731	*trnF/trnA2*
		MDR07	308929-308948	329232-329251	20	20283	*-/-*
		MDR08	191271-191291	387527-387547	21	196235	*tsf/-*
		MDR09	52452-52473	141300-141321	22	88826	*trnL2/trnL3*
		MDR10	171181-171202	231113-231134	22	59910	*pnp/-*
		MDR11	113369-113391	338378-338400	23	224986	*pdxJ/-*
		MDR12	141540-141562	465203-465225	23	323640	*trnM2/trnV2*
		MDR13	242541-242563	326616-326638	23	84052	*-/cydD*
		MDR14	308808-308830	329209-329231	23	20378	*tuf/-*
		MDR15	307460-307483	329023-329046	24	21539	*secE/-*
		MDR16	307485-307508	329048-329071	24	21539	*secE/-*
		MDR17	206190-206215	347777-347802	26	141561	*trnA1/trnA2*
		MDR18	495524-495551	505578-505605	28	10026	*trnS3/trnS4*
		MDR19	307517-307546	329071-329100	30	21524	*secE/-*
		MDR20	130684-130721	132186-132223	38	1464	*-/cysE*
		MDR21	513474-513513	513540-513579	40	26	*acpP/-*
		MDR22	307551-307595	329105-329149	45	21509	*secE/-*
		MDR23	308751-308806	329152-329207	56	20345	*tuf/-*
		MDR24	435418-435492	435493-435567	75	-	-
	IR	MIR01	23640-23659	505488-505507	20	481828	*trnP1/trnR4*
		MIR02	141541-141560	347884-347903	20	206323	*trnM2/trnI*
		MIR03	208052-208071	289024-289043	20	80952	*trnK/trnF*
		MIR04	238440-238459	308844-308863	20	70384	*typA/tuf*
		MIR05	347884-347903	465204-465223	20	117300	*trnI/trnV2*
		MIR06	133087-133107	133112-133132	21	4	*-/-*
		MIR07	170354-170374	170535-170555	21	160	*rpsO/-*
		MIR08	389208-389228	505528-505548	21	116299	*trnR3/trnR4*
		MIR09	435164-435184	435186-435206	21	1	*-/-*
		MIR10	459924-459945	516044-516065	22	56098	*-/-*
		MIR11	81288-81310	81314-81336	23	3	*-/-*
		MIR12	237863-237886	237897-237920	24	10	*ydjM/ydjM*
		MIR13	40549-40588	45537-45576	40	4948	*-/-*
		MIR14	52338-52380	54429-54471	43	2048	*-/-*
		MIR15	52376-52435	54375-54434	60	1939	*-/-*
		MIR16	109083-113105	343701-347723	4023	230595	*trnE/-*

Table S4. DNA repeats apparently linked to tRNA genes proliferation. Percentages of each repeat mapping on tRNA stem regions (STEM) or self-paired at the tRNA predicted secondary structure (SELF-PAIRED) are indicated.

Repeat	Positions	Sequence	tRNA	Locus	STEM	SELF-PAIRED
MDR04	141407-141426 141625-141644	TATAGCCAAGCGGTAAGGCA	Gln(TTG) Gln(CTG)	MPC_119 MPC_121	40% (8/20)	30% (6/20)
MDR05	141448-141467 141666-141685	CCCAGGTTCGAATCCTGGTA	Gln(TTG) Gln(CTG)	MPC_119 MPC_121	60% (12/20)	50% (10/20)
MDR06	289025-289044 347776-347795	TGCTCTACCAACTGAGCTAT	Phe(GAA) Ala(TGC)	MPC_262 MPC_316	45% (9/20)	40% (8/20)
MDR09	52452-52473 141300-141321	AGTGGCGAAATTGGTAGACGCA	Leu (CAA) Leu (TAG)	MPC_041 MPC_118	36,4% (8/22)	27,3% (6/22)
MDR12	141540-141562 465203-465225	CGTAGCTCAGTTGGTTAGAGCAC	Met(CAT) Val (GAC)	MPC_120 MPC_418	48% (11/23)	34,8% (8/23)
MDR17	206190-206215 347777-347802	GCTCTACCAACTGAGCTATAGCCCCA	Ala(GGC) Ala(TGC)	MPC_175 MPC_316	57.7% (15/26)	30,8% (8/26)
MDR18	495524-495551 505578-505605	TGAGGGAGGGATTCGAACCCTCGATACA	Ser(GGA) Ser(GCT)	MPC_443 MPC_454	71.4% (20/28)	35,7% (10/28)
MIR01	23640-23659 505488-505507	GGTCGGAGGTTCGAATCCTC	Pro(GGG) Arg(ACG)	MPC_018 MPC_453	45% (9/20)	40% (8/20)
MIR02	141541-141560 347884-347903	GTAGCTCAGTTGGTTAGAGC	Met(CAT) Ile(GAT)	MPC_120 MPC_317	45% (9/20)	40% (8/20)
MIR03	208052-208071 289024-289043	TAGCTCAGTTGGTAGAGCAG	Lys(TTT) Phe(GAA)	MPC_178 MPC_262	50% (10/20)	40% (8/20)
MIR05	347884-347903 465204-465223	GCTCTAACCAACTGAGCTAC	Ile (GAT) Val(GAC)	MPC_317 MPC_418	45% (9/20)	40% (8/20)
MIR08	389208-389228 505528-505548	CCGTAGCTCAGCTGGATAGAG	Arg(CCG) Arg(ACG)	MPC_350 MPC_453	47.6% (10/21)	28,5% (6/21)

4. RESULTS AND DISCUSSION

An atypical cellular organization was early reported for the endosymbiotic systems of many mealybug species (Buchner, 1965). Later molecular and histological characterization showed that two bacterial endosymbionts inhabit the bacteriocytes of most Pseudococcinae mealybugs (Fukatsu and Nikoh, 2000; von Dohlen et al., 2001; Thao et al., 2002; Kono et al., 2008; Gatehouse et al., 2011; Koga et al., 2012). This dual symbiosis is organized in an unprecedented nested-manner, in which every cell of the betaproteobacterium "*Ca*. Tremblaya princeps" contains several cells of a gammaproteobacterium (von Dohlen et al., 2001; Kono et al., 2008; Gatehouse et al., 2011; Koga et al., 2012). Phylogenetic analyses confirmed the monophyly of "*Ca*. Tremblaya princeps", suggesting it to be the P-endosymbiont of Pseudococcinae mealybugs. In contrast, gammaproteobacterial endosymbionts are polyphyletic, which means that they have infected "*Ca*. Tremblaya princeps" through multiple independent events (Thao et al., 2002). Nevertheless, both endosymbionts are transmitted together from mothers to offspring, while neither empty "*Ca*. Tremblaya princeps" cells nor free gammaproteobacteria cells have been reported (von Dohlen et al., 2001). In order to infer the molecular bases of the dual symbiosis, we undertook the metagenomic characterization of the nested endosymbiotic system of *P. citri* and performed structural and functional analysis of the sequenced bacterial genomes. Our results provided important information regarding (1) reductive evolution, (2) metabolic complementation and (3) molecular communication in the nested endosymbiotic consortium of *P. citri*. Finally, the study of some relevant functional and structural features was extended to additional Pseudococcinae and/or Phenacoccinae mealybug species.

4.1. Genomic Characterization

Bacterial chromosomes display complex evolutionary dynamics. The study of genome rearrangements and gene repertoire diversity (Abby and Daubin, 2007) reveals that HGT, transposition and intragenomic recombination mostly explain the structural fluidity typically found among free-living bacteria genomes (Casjens, 1998; Rocha, 2008). These phenomena are important sources of evolutionary novelties, being responsible for the bacterial huge metabolic diversity and adaptive potential. In contrast, bacteria that have acquired an intracellular host-dependent lifestyle, including obligate endosymbionts of insects, show important constrains to these molecular mechanisms. Comparative genomics has allowed the identification of several commonalities among them (McCutcheon and Moran, 2012). Endosymbionts genomes tend to be smaller than those of their free-living relatives. This is mostly due to the drastic reduction of the gene content (affecting functions that are unnecessary in the intracellular environment or that can be provided by the host), but also to the depletion of redundant or nonfunctional genomic features (e.g., duplicated genes, transposable elements, prophages, intergenic regions). Thus, endosymbiont genomes usually display very high gene densities. Moreover, bacterial isolation due to the obligatory intracellular lifestyle, together with the early loss of genes involved in DNA acquisition, recombination and repair, are thought to enhance such a reductive process by

preventing the occurrence of HGT events. Additionally, endosymbiont genomes usually show high levels of structural stability, mostly due to the loss of repetitive DNA and recombination-related *loci* (Silva et al., 2003). Finally, the genomes of bacterial endosymbionts typically undergo a significant increase in AT-content. This has been explained by the higher energetic cost of the biosynthesis of GTP and CTP nucleotides (Rocha and Danchin, 2002), and/or by the combination of both the universal mutational bias towards AT and the early loss of DNA repair systems (Lind and Andersson, 2008; Hershberg and Petrov, 2010).

4.1.1. General features of "*Ca*. Tremblaya princeps" genome

The partial characterization of the genome of "*Ca*. Tremblaya princeps" strain from *D. brevipes* showed that it displays unexpected molecular characteristics for an endosymbiont after, at least, 100-200 million years of intracellular lifestyle (Baumann et al., 2002). The analysis of two genomic fragments with a total length of 64 kb indicated that this genome has a high GC-content and a low gene density. Moreover, the analysis revealed the presence of a partial genome duplication affecting the ribosomal operon and its closest genomic context. It included a total of 5.7-kb identical paralogous sequences. The same trait was found in four additional "*Ca*. Tremblaya princeps" strains, belonging to *Maconelicoccus australiensis*, *M. hirsutus*, *M. albizziae* and *P. citri*, thus suggesting that CE would be hampering the independent evolution of these paralogous *loci*.

The metagenomic characterization of the *P. citri* endosymbionts showed "*Ca*. Tremblaya princeps" to possess a tiny genome (Chapter 1; McCutcheon and von Dohlen, 2011) composed by a single 139-kb chromosome, one of the most drastically reduced prokaryotic genomes described to date. It also confirmed all previously reported unconventional traits, suggesting that reductive evolution of "*Ca*. Tremblaya princeps" occurred in an atypical manner. The genome of "*Ca*. Tremblaya princeps" from *P. citri* displays a high overall GC-content (59%), which has an strong effect on both the codon usage and the amino acidic content of its proteome (Figure 2 and Table S2 in Chapter 2, pages 56 and 73), although no alterations of the genetic code was noticed. No clear evidences for AT accumulation were observed, since GC-content differences among the coding fraction, pseudogenes and IGRs were not found (Table 1 in Chapter 2, page 54). The 130 annotated genes include 116 CDS and 14 RNA genes, which cover only 71% of the genome (Table 1 in Chapter 2, page 54). Such a low gene density is partially due to pseudogenes abundance, which is rare among tiny genomes (McCutcheon and von Dohlen, 2012). However, the "*Ca*. Tremblaya princeps" genome also presents very large IGRs were neither genes nor pseudogenes could be annotated (Figure 4A in Chapter 5, page 136). A very similar situation has been described in the genome of "*Ca*. Portiera aleyrodidarum", the P-endosymbiont of whiteflies. In the latter, this is mostly due to the expansion of DNA tandem repeats, and has been linked to the pseudogenization of *dna*Q (encoding the subunit epsilon of the DNA polymerase III, EC 3.1.11.-) (Sloan and Moran, 2013). This is not the case for "*Ca*. Tremblaya princeps", where a

functional copy of *dna*Q has been annotated and tandem repeats are scarce (Figure 5 in Chapter 5, page 137).

As already noticed in *D. brevipes*, the "*Ca*. Tremblaya princeps" strains from *P. citri* presents two 5.7-kb identical fragments containing the 3'-end of *leu*A (encoding alpha-isopropylmalate synthase, EC 2.3.3.13), *rps*O (encoding 30S ribosomal protein S15), the ribosomal operon, and the 5'-end of *rsm*H (encoding 16S rRNA m^4C1402 methyltransferase, EC 2.1.1.199) (Figure 1 in Chapter 2, page 55). Additional DNA repeats with at least 20 bp in length are present in this genome, although none of them seems to be related to either gene or partial genome duplication events (Table S3 in Chapter 5, page 146). In contrast, their biased GC-content suggests that they are highly influenced by the nucleotide composition of the genome (Figure 3 in Chapter 5, page 134). The full identity of the paralogous *loci* including the complete ribosomal operon indicates that CE is also acting on "*Ca*. Tremblaya princeps" from *P. citri*. This non-reciprocal recombination process is a universal biological phenomenon leading to the homogenization of the polymorphisms independently accumulated by repetitive DNA sequences (Liao 1999; Innan and Kondrashov, 2010; Hastings, 2010). Many genes are involved in the two pathways known to mediate HR events in bacteria, i.e., RecF and RecBCD (Spies and Kowalczykowski, 2005; Rocha et al., 2005). Although no genes specifically involved in HR were detected in the "*Ca*. Tremblaya princeps" genome, we showed this molecular phenomenon to have occurred very recently in "*Ca*. Tremblaya princeps" of cluster E, according to the comparative analysis between the closely related strains from *P. citri* and *P. ficus* (Table S2 in Chapter 5, page 145). Moreover, an inversion polymorphism has been noticed in the genome of "*Ca*. Tremblaya princeps" PCIT, characterized by McCutcheon and von Dohlen (2011). It affects a 7-kb genomic fragment, which was found in both orientations in the endosymbiont population of individual insects. As well as CE, structural instability is only conceivable due to the action of HR-related genes. Interestingly, structural instability was also noticed in the genome of "*Ca*. Portiera aleyrodidarum", suggesting that HR-related genes might have been recently lost in this bacterium (Sloan and Moran, 2013).

4.1.2. General features of "*Ca*. Moranella endobia" genome

During the metagenomic study of the endosymbiotic system of *P. citri*, the genome of its gamma-endosymbiont, "*Ca*. Moranella endobia", was also characterized. It is composed by a circular 538-kb chromosome with a general GC-content of 44%, and it includes 411 CDS and 47 RNA genes (Chapter 2). Thus, "*Ca*. Moranella endobia" genome is 3.87 fold larger than that of "*Ca*. Tremblaya princeps", and contains 3.52 fold more functional genes (Table 1 in Chapter 2, page 54). In contrast to "*Ca*. Tremblaya princeps", there are evident differences regarding GC-content of functional and nonfunctional DNA (Table 1 in Chapter 2, page 54), suggesting AT to be accumulating in "*Ca*. Moranella endobia", as usually observed among endosymbionts.

"*Ca*. Moranella endobia" gene density is not as high as expected when compared with similarly reduced genomes (McCutcheon and Moran, 2012), in part due to de presence of pseudogenes.

In this case, IGR lengths are much more homogeneous than in "*Ca*. Tremblaya princeps" (Figure 4A in Chapter 5, page 136).

Similarly to "*Ca*. Tremblaya princeps", there is a 4-kb identical duplication in "*Ca*. Moranella endobia", affecting the 23S and 5S rRNA genes and their closest genomic context (Figure 1 in Chapter 2, page 55), comprising *trn*D, *acp*S (encoding holo-[acyl-carrier-protein] synthase, EC 2.7.8.7) and the 3'-end of *pdx*J (encoding pyridoxine 5'-phosphate synthase, EC 2.6.99.2). This suggests that CE would be also acting on these paralogous *loci*. Interestingly, "*Ca*. Moranella endobia" encodes an almost complete set of HR-related proteins (Table 2 in Chapter 3, page 83), which have been pointed out as responsible for CE both in "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" (Chapter 2; Chapter 5). This observation is coherent with the exportation of, at least, part of the "*Ca*. Moranella endobia" cytoplasm content towards "*Ca*. Tremblaya princeps", including HR-related proteins (Chapter 3).

Many additional DNA repeats were found in the "*Ca*. Moranella endobia" genome when direct and inverted repeats with at least 20 bp in length were screened. In contrast to "*Ca*. Tremblaya princeps", several of them do not appear to have a random molecular origin, but mapped inside putatively paralogous *loci*, including both coding and tRNA genes (Table S3 and Table S4 in Chapter 5, pages 146 and 147).

4.1.3. Unconventional reductive evolution in a nested endosymbiosis

Co-occurrence of atypical genomic features with the unprecedented nested organization of the endosymbiotic consortium led authors to suggest that the nested gamma-endosymbionts would be responsible for the unconventional reductive evolution of the "*Ca*. Tremblaya princeps" genome (McCutcheon and von Dohlen, 2011; Husnik et al., 2013). Since "*Ca*. Tremblaya phenacola" appears alone in the bacteriocytes of Phenacoccinae mealybugs (Gruwell et al., 2010; Koga et al., 2012), and no major differences exist in the biology of these mealybugs with respect to those belonging to subfamily Pseudococcinae (Koga et al., 2012), it could be considered as a quasi-perfect negative control in order to test this hypothesis. Recent characterization of the complete genome of "*Ca*. Tremblaya phenacola" PAVE strongly supports it, since it presents lower GC-content and higher gene density than the "*Ca*. Tremblaya princeps" genome (Husnik et al., 2013). Moreover, comparative genomics between "*Ca*. Tremblaya princeps" and "*Ca*. Tremblaya phenacola" revealed that the partial genome duplication involving the ribosomal operon, which was thought to be ancestral in the lineage of Pseudococcinae, occurred before the split of the two subfamilies (Figure 1 in Chapter 5, page 129). The original ribosomal operon has not been conserved in "*Ca*. Tremblaya phenacola", where only a 385-bp fragment remains identical. This fact perfectly fits into the above mentioned hypothesis, since no CE is expected without the contribution of an additional nested endosymbiont. Although in a lesser extent, identity between paralogous *loci* in the explored "*Ca*. Tremblaya princeps" strains from cluster F was also reduced, somehow representing a middle point between "*Ca*. Tremblaya princeps" from *P. citri*

and "*Ca*. Tremblaya phenacola" PAVE. According to the phylogenetic relationships among "*Ca*. Tremblaya", these results can be explained in two diferente ways: (1) although originally present, nested endosymbionts got finally extinct in "*Ca*. Tremblaya princeps" strains from cluster F; (2) even if present, any unreported, nested endosymbiont inhabiting the cytoplasm of these "*Ca*. Tremblaya princeps" strains has lost the ability to drive CE of paralogous *loci* in the genome of its betaproteobacterial partner.

Since both genome instability and CE of paralogous *loci* must be caused by HR events, and a broad set of HR-related genes has been annotated in the genome of "*Ca*. Moranella endobia" from *P. citri* (Table 2 in Chapter 3, page 83), we explored if the function of these genes could explain, at least to some extent, the link between nested endosymbiosis and the "*Ca*. Tremblaya princeps" atypical reductive evolution (Chapter 5). In coherence with this hypothesis, we found several HR-related genes of gammaproteobacterial origin in three different mealybug species (Table 1 in Chapter 5, page 132), which likely represents the acquisition of HR potential by the corresponding "*Ca*. Tremblaya princeps" strains through the independent internalization of three different gammaproteobacterial lineages (Figure 8 in Chapter 4, page 109; Figure S1 in Chapter 5, page 144). Additionally, we detected that CE events occurred very recently in "*Ca*. Tremblaya princeps" strains from cluster E (Table S2 in Chapter 5, page 145). The genomic characterization of "*Ca*. Moranella endobia" from *P. citri* suggested the existence of simplified but still functional RecF and RecBCD pathways (Chapter 2), while later analysis on the nested consortium from *P. ficus* suggested a recent inactivation of the RecBCD pathway in the gamma-endosymbionts inhabiting "*Ca*. Tremblaya princeps" from clade E (Chapter 5).

DNA repeats are the substrate for HR machinery. Recombination between direct repeats causes DNA duplication or DNA deletion, while recombination between inverted repeats generates DNA inversion. Inverted repeats of, at least, 71 bp are known to promote recombination events in the endosymbiotic system of *P. citri* (McCutcheon and von Dohlen, 2011). However, smaller repeats (ranging from 18 to 24 bp) have been suggested to be long enough to do so in prokaryotic genomes (Shen and Huang, 1986; Aras et al., 2003; Sloan and Moran, 2013). In order to study the relative susceptibility to HR of the nested endosymbionts of *P. citri*, we searched for DNA repeats with a minimal length of 20 bp in both partners' genomes (Table S3 in Chapter 5, page 146). Evident differences between "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" were found in terms of DNA repeats density (abundance/genome length) and mean length. Our results suggest "*Ca*. Tremblaya princeps" to be more sensitive than "*Ca*. Moranella endobia" to the action of the HR machinery encoded by the latter (Rocha, 2003; Chapter 5). The putatively great impact of HR-related genes on the "*Ca*. Tremblaya princeps" genome could be explained assuming the exportation of, at least, part of the "*Ca*. Moranella endobia" cytoplasmic content (Chapter 1; McCutcheon and von Dohlen, 2011).

In addition to CE and genome structural instability, HR between genome repeats could also explain the low gene density of "*Ca*. Tremblaya princeps" (Chapter 5). Genome shuffling mediated by inverted repeats inside non-essential genes would have accelerated pseudogenization, while

negative selection putatively affecting further HR events (both DNA deletions and inversions might seriously risk the stability of the endosymbiotic system) would impede the final elimination of these regions in "*Ca*. Tremblaya princeps". The deleterious effects of additional HR events could explain the tendency to loose recombinogenic pathways. As we have seen, the RecBCD pathway has recently been inactivated in "*Ca*. Moranella endobia" (Chapter 5), while the absence of both the RecFOR complex and *rec*A mutations known to bypass such a deficiency are expected to hinder the action of the remaining RecF pathway (Lavery and Kowalczykowski, 1992).

4.2. Metabolic Complementation

Symbiosis between endocellular bacteria and insects is considered critical for the evolutionary success of the latter. Most insects depend on restricted and nutritionally unbalanced diets, so that their bacterial endosymbionts supply them with essential nutrients otherwise unavailable. Obligatory intracellular lifestyle lead endosymbiont genomes to undergo drastic functional degradation in a progressive manner, which is thought to jeopardize their fitness (McCutcheon and Moran, 2012; Koga et al., 2013). Only the eventual recruitment of additional endosymbiotic bacteria could compensate for the loss of essential functions, thus temporarily avoiding the lineage extinction (Moya et al., 2009). In coherence with the predictions from the Black Queen Hypothesis (Morris et al., 2012; Sachs and Hollowell, 2012), co-existence of two or more intracellular symbionts in the same bacteriome, or even in the same bacteriocyte, frequently leads to an scenario of metabolic complementation involving the members of the endosymbiotic consortium.

4.2.1. Essential amino acids biosynthesis, on the root of the symbiosis

The metagenomic characterization of the nested endosymbionts from *P. citri* suggests them to complement their host diet with essential amino acids, as usually found in endosymbiotic systems from sap-feeding insects (Takiya et al., 2006; McCutcheon et al., 2009a; Lamelas et al., 2011; Urban and Cryan, 2012). Twenty-nine functional genes involved in essential amino acids biosynthesis were annotated in the genome of "*Ca*. Tremblaya princep" PCVAL (Chapter 1), despite its reduced length and low gene density. This data are coincident with those of "*Ca*. Tremblaya princeps" PCIT, except for the pseudogene *trp*G (previously encoding the glutamine amidotransferase component of anthranilate synthase, EC 4.1.3.27) (Table S1 in Chapter 2, page 72), which was annotated as a functional gene by McCutcheon and von Dohlen (2011), in spite of the dramatic reduction of the amidotransferase functional domain (Chapter 2). These genes are mostly involved in the biosynthesis of the branched-chain amino acids (7), histidine (7), tryptophan and phenylalanine (7), while some of them participate in the biosynthesis of arginine (1), methionine (1), threonine and lysine (6). Only four (*arg*F, *arg*H, *car*B and *trp*G) out of the 25 pseudogenes annotated in the genome of "*Ca*. Tremblaya princeps" PCVAL correspond to this

category, and all of them have a functional homolog in the genome of "*Ca*. Moranella endobia". Comparative genomics between strains PCVAL and PCIT also revealed the occurrence of innocuous indel mutations in *aro*C (encoding chorismate synthase, EC 4.2.3.5, involved in the biosynthesis of aromatic amino acids) and *ilv*D (encoding dihydroxy acid dehydratase, EC 4.2.1.9, involved in the biosynthesis of branched-chain amino acids), for which no additional copies exist in the bacterial consortium (Chapter 2).

Only 15 out of the 458 functional genes of "*Ca*. Moranella endobia" PCVAL are directly involved in the biosynthesis of essential amino acids. They mostly participate in the pathways leading to tryptophan and phenylalanine (8), although some of them are involved in the biosynthesis of arginine (4), threonine and lysine (3). It also includes *cys*E (encoding serine acetyltransferase, EC 2.3.1.30), which might be involved in the biosynthesis of isoleucine and methionine by participating in the biosynthesis of the nonessential cysteine. Thus, "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" cooperate in the production of most essential amino acids. Moreover, they do it in an unprecedented tight manner, possibly because the nested organization of the bacterial consortium facilitates the bidirectional traffic of molecular precursors (McCutcheon and von Dohlen, 2011; Husnik et al., 2013). In contrast to mealybugs and with very limited exceptions (Lamelas et al., 2011), bacterial members of symbiotic consortia are usually specialized in the full production of complementary sets of essential amino acids (Takiya et al., 2006; McCutcheon and Moran, 2007; Gosalbes et al., 2008; McCutcheon et al., 2009a; McCutcheon et al., 2009b; McCutcheon and Moran, 2010; Urban and Cryan, 2012). Very few functional redundancies are observed between "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia", which suggests that they have very specialized roles regarding essential amino acids biosynthesis. In coherence, "*Ca*. Tremblaya phenacola" PAVE has retained exactly the complete set of metabolic capabilities collectively retained by "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" in the dual symbiosis of *P. citri* (Husnik et al., 2013).

Finally, a metatranscriptomic analysis of *P. citri* has recently confirmed the implication of the insect host in the biosynthesis of essential amino acids. It includes the participation of several nuclear genes of both eukaryotic (AAT, BCAT, CBL, CGL and OAT) and prokaryotic (*cys*K, *dap*F and *lys*A) origin. Thus, multiple HGT events from several transient alpha- and gammaproteobacteria S-symbionts have drastically influenced the evolution of the symbiosis in *P. citri* (Husnik et al., 2013). Collaboration of the insect host is especially evident for methionine biosynthesis, for which only *met*E has been conserved in the bacterial consortium of *P. citri*. At least one apparently functional *met*E has been detected in the endosymbiotic consortia from several Pseudococcinae mealybug species analyzed (Figure 1 and Table S4 in Chapter 4, pages 101 and 119).

4.2.2. Additional ways of complementation between *"Ca.* Tremblaya princeps*"* and *"Ca.* Moranella endobia*"*

Due to its extreme genomic degradation, "*Ca*. Tremblaya princeps" displays almost null metabolic capacities except for those involved in the production of essential amino acids (Chapter 1). However, the analysis of the metabolic abilities predicted for the two members of the bacterial consortium pointed out additional ways of metabolic complementation putatively functioning in this consortium (Chapter 2). Metabolic complementation could be taking place in the assembly of [Fe-S] clusters, one of the most ubiquitous and versatile prosthetic groups in nature (Johnson et al., 2005). Two sets of proteins, the Suf (*suf*ABCDSE) and the Isc (*isc*SUA) systems, are known to catalyze the assembly in *E. coli*. While "*Ca*. Moranella endobia" encodes the complete *suf* operon, the IscS system is almost lost, having *isc*S as its sole representative. In contrast, both *isc*S and *isc*U are present in the genome of "*Ca*. Tremblaya princeps", which lacks *isc*A instead. In spite of this, "*Ca*. Tremblaya princeps" encodes *hes*B, whose protein product might putatively compensate for the absence of IscA, since both of them belong to the same protein family. An additional case of metabolic complementation between both endosymbionts might occur though the participation of the transketolase Tkt and the transaldolase TalB, whose coding genes were annotated in the genomes of "*Ca*. Moranella endobia" and "*Ca*. Tremblaya princeps", respectively. Both proteins could create a reversible link between the pentose phosphate pathway and glycolysis. The pentose phosphate pathway appears incomplete in "*Ca*. Moranella endobia", where only a few of the corresponding genes were found, including an inactivated transaldolase. On the other hand, "*Ca*. Moranella endobia" is the only partner predicted to perform glycolysis, which is very important since the whole consortium relies on substrate-level phosphorylation as a source of ATP. This is the first case in which all energy sources are apparently provided by only one of the members of the bacterial consortium (Chapter 2).

Complementation between nested endosymbionts is expected to occur also at the informational level, including "*Ca*. Moranella endobia" supply of tRNAs and aminoacyl-tRNA synthetases. Only 7 functional tRNA genes were annotated in the genome of "*Ca*. Tremblaya princeps", where no functional aminoacyl-tRNA synthetase was found. In contrast, a complete set of aminoacyl-tRNA synthetases is encoded by the genome of "*Ca*. Moranella endobia", which strickingly include more than 80% of the tRNA genes annotated in the endosymbiotic consortium (Figure 3 and Table S3 in Chapter 2, pages 61 and 74).

4.2.3. Exploring amino acids supply in mealybugs beyond *P. citri* and *P. avenae*

In order to investigate the evolution of essential amino acids supply in mealybugs beyond the particular cases of *P. citri* and *P. avenae*, we explored the presence of genes belonging to this functional category in additional endosymbiotic lineages from Pseudococcinae (*D. boninsis*, *P.*

longispinus and *P. vinurni*) and Phenacoccinae mealybugs (*P. peruvianus* and *P. madeirensis*). We took into account those genes performing the last conserved step of the corresponding biosynthetic pathways both in "*Ca*. Tremblaya phenacola" PAVE and in the bacterial consortium from *P. citri*. Genes whose length and conservation level allowed for the design of reliable degenerate primers were screened. Our approach revealed the presence of a broad set of biosynthetic genes in all tested insects, confirming the key importance of essential amino acids supply in the symbiosis between intracellular bacteria and mealybugs (Figure 1 in Chapter 4, page 101). Moreover, all tested Pseudococcinae species showed only one single haplotype (of gamma- or betaproteobacterial origin) for most of the screened genes, suggesting that endosymbionts specialization is generalized among mealybug dual endosimbioses. Only consortia from *P. longispinus* and *D. boninsis* showed functional redundancy regarding *leu*B and both *leu*B and *met*E, respectively (Figure 1 in Chapter 4, page 101).

As expected for Phenacoccinae mealybugs from the genus *Phenacoccus*, where "*Ca*. Tremblaya phenacola" have been reported as the sole endosymbiont (Gruwell et al., 2010; Koga et al., 2012), most of the genes amplified from *P. peruvianus* and *P. madeirensis* total DNA samples had a betaproteobacterial origin (Figure 1 in Chapter 4, page 101). In contrast, phylogenetic analyses confirmed the gammaproteobacterial origin of the two amplified *trp*B haplotypes (Figure 2 in Chapter 4, page 103). Both PCR and FISH approaches allowed us to discard the presence of any additional gammaproteobacterial endosymbiont (Figure 3 and 4 in Chapter 4, pages 104 and 105), while the strict co-occurrence of *trp*B and the corresponding "*Ca*. Tremblaya phenacola" from *P. peruvianus* discarded such gene to be present in the insect host genome (Figure 5 in Chapter 4, page 105). Thus, our results suggest that HGT has affected the genomes of "*Ca*. Tremblaya phenacola" from *P. peruvinus* and *P. madeirensis* influencing, at least, the evolution of the tryptophan supply strategy. The existence of HGT in "*Ca*. Tremblaya phenacola" represents a novel evolutionary strategy in the symbiosis between bacteria and mealybugs. In contrast to free-living bacteria, HGT events are not common among endosymbionts, mostly due to the isolation and genome streamlining associated to their intracellular lifestyle (Silva et al., 2003). Actually, this is the second reported case after that of "*Candidatus* Profftella armature", defensive symbiont of the Asian citrus psyllid *Diaphorina citri* (Hemiptera: Psyllidae), which has horizontally acquired the genes involved in the biosynthesis of a cytotoxic metabolite (Nakabachi et al., 2013). The complete genome sequencing of "*Ca*. Tremblaya phenacola" from *P. peruvianus* (currently in progress) will reveal the impact of HGT in the reductive evolution of this endosymbiont.

4.3. Molecular communication between nested endosymbionts

As already stated, a very limited set of functional genes have been annotated in the genome of "*Ca*. Tremblaya princeps", which displays almost null metabolism. Moreover, it seems even

unable to deal with basic informational functions since several members of the corresponding minimal molecular machineries had not been found (Gil et al., 2004). In contrast, most of these *loci* are present in the genome of "*Ca*. Moranella endobia", which has been postulated as their single source for the whole consortium (McCutcheon and von Dohlen, 2011). According to *in silico* predictions, many "*Ca*. Moranella endobia" proteins are expected to go into the "*Ca*. Tremblaya princeps" cytoplasm. However, the way such exportation process takes place remained unclear (McCutcheon and von Dohlen, 2011).

4.3.1. Proteins in the "*Ca*. Moranella endobia" cell envelope

The analysis of the inferred proteome of "*Ca*. Moranella endobia" allows making predictions about the presence of proteins in its inner and outer membranes (Chapters 2; Chapter 3).

Regarding protein translocation, genes coding for a putatively functional Sec translocation complex (Mori and Ito, 2001; de Keyzer et al., 2003) were annotated in the genome of "*Ca*. Moranella endobia" (Chapter 2). These include the inner membrane proteins SecY, SecE and SecG (members of the SecYEG complex), FtsY (signal recognition particle receptor, involved in the targeting and integration of inner membrane proteins), and both LepB and LspA (responsible for the cleavage of the signal peptides typically found in the N-terminal of immature secretory proteins). Additionally, SecA (which recognizes the signal peptide of secretory proteins and energize the translocation process with its ATPase activity), and Ffh (which form, together with 4.5S RNA, the signal recognition particle, co-translationally directing the secreted proteins to the SecYEG complex) are also present. Alternative post-translational secretion, mediated by the chaperone SecB, can also take place in other bacteria. However, this route is not working in "*Ca*. Moranella endobia", whose *sec*B homolog is inactivated (McCutcheon and von Dohlen, 2011; Chapter 2). On the other hand, BamA, BamB and BamE are thought to associate into a simplified BamABCDE complex, involved in the assembly of beta-barrel outer membrane proteins (OMPs), which would be fed by the periplasmic chaperones SurA and Skp. Such chaperones prevent folding and aggregation of transported proteins during their passage through the periplasm, assisting in the proper folding of OMPs once they reach their final location. Although *deg*P (encoding the protease in charge of the degradation of misfolded OMPs) was not found in the "*Ca*. Moranella endobia" genome, it possess *deg*Q, which encodes a periplasmic protease showing functional overlap with DegP (Singh et al., 2011; Sawa et al., 2011). Finally, "*Ca*. Moranella endobia" genome includes *yid*C, which would be participating in inner membrane proteins insertion in a Sec-independent manner, since neither *sec*D, *sec*F nor *yaj*C have been annotated (Nouwen and Driessen, 2002; Facey et al., 2007).

A very limited set of transporters has been predicted in "*Ca*. Moranella endobia" (Chapter 2). These include several ion transporters, such as the CydDC complex, which participates in the transport of glutathione and cysteine from the cytoplasm to the periplasm (Pittman et al., 2005); PitA, a low-affinity transporter of inorganic phosphate and Zn (II) (Beard et al., 2000); the ATP

Binding Cassette (ABC) transporter ZnuABC, which functions as a high-affinity Zn (II) uptake system (Patzer and Hantke, 1998); YbaL, an uncharacterized member of the CPA2 family of monovalent cation:proton antiporters; and YggT, a putative but still uncharacterized ion transporter.

Being the only member of the consortium able to perform glycolysis, "*Ca*. Moranella endobia" also possess a complete phosphotransferase system (PTS) for the uptake (and simultaneous phosphorylation) of hexoses to be used as carbon source, composed by the integral membrane proteins ManY and ManZ, which form the transmembrane channel, and the cytoplasmic component ManX. Additionally, it encodes both PtsI and PtsH, which mediate the tranfer of a phosphoryl group from phosphoenol pyruvate (PEP) to the imported hexoses.

Lipopolysaccharides (LPS) are major constituents of the outer membrane of Gram-negative bacteria, being involved in many aspects of bacterial ecology. They are considered essential for proper function of the outer membrane, which mostly acts as a protective permeability barrier against environmental large molecules and hydrophobic compounds. LPS usually consist of a hydrophobic domain known as lipid A (which anchor it in the outer leaflet of the outer membrane), a nonrepeating core oligosaccharide, and a distal polysaccharide known as O-antigen (which is displayed in the cell surface) (Wang and Quinn, 2010). Gram-negative endosymbionts usually lack most of the genes involved in LPS biosynthesis and neither "*Ca*. Tremblaya princeps" nor "*Ca*. Moranella endobia" are exceptional. However, instead of the LPS typically observed in gram-negative bacteria, the latter still retains the ability to synthesize the lipid IV_A, lacking any secondary acyl chains and Kdo (2-keto 3-deoxy-D-manno-octulosonate). The "*Ca*. Moranella endobia" genome also encodes all essential components of the Lpt transport system, which is responsible for transporting LPS from the inner to the outer membrane and their proper assembly. It is composed by an ABC transporter, including the three components located in the inner membrane (LptCFG) plus the cytoplasmic protein LptB; the periplasmic protein LptA, involved in the movement of LPS across the periplasm (Ma et al., 2008), and the LptED complex, devoted to LPS assembly in the outer membrane (Wu et al., 2006). It also encodes MsbA, mostly involved in LPS translocation across the inner membrane, but also in the transportation of the LPS precursor N-acetyl-glucosamine, lipid A-core and glycerophospholipid molecules (Polissi and Georgopoulos, 1996; Doerrler et al., 2004).

Finally, two channels are expected to exist in "*Ca*. Moranella endobia" membranes: the outer membrane OmpF, which might allow the passage of small solutes under 0.6 kDa (e.g., sugars, ions, amino acids) (Cowan et al., 1992), and the inner membrane mechanosensitive channel MscL, which is involved in cell homeostasis and has not been found in any other reduced genome of Sternorrhyncha endosymbionts.

4.3.2. Exploring the "*Ca*. Moranella endobia" potential for protein exportation

4.3.2.1. The Sec translocon

As above mentioned, an apparently functional Sec translocation system has been retained by "*Ca*. Moranella endobia". This is the most generally employed mechanism for protein translocation across the inner membrane in Gram-negative bacteria, endosymbionts included (McCutcheon and von Dohlen, 2011; Chapter 2). We have explored if "*Ca*. Moranella endobia" proteins apparently essential for "*Ca*. Tremblaya princeps" could be exported through this mechanism.

Signal peptide sequences prediction along "*Ca*. Moranella endobia" proteome. Sec-dependent secretory proteins, which include periplasmic and outer membrane proteins, are synthesized as precursor macromolecules in the cell cytoplasm, typically carrying amino-terminal signal peptide (SP) sequences that would be first recognized by SecA and finally cleaved by the signal peptidases I (LepB) or II (LspA). So, if protein communication between the nested endosymbionts takes place in a Sec-dependent manner, one would expect SP sequences to be broadly distributed in the "*Ca*. Moranella endobia" proteome.

A comprehensive scanning of the 411 coding sequences from "*Ca*. Moranella endobia" PCVAL was performed using both the PRED-TAT (Bagos et al., 2010) and SignalPv4.1 (Petersen et al., 2011) softwares. Putative SP sequences were predicted for 33 proteins at least by one of the two softwares (Table 1 in Chapter 3, page 81). Results are coherent with those roughly reported by McCutcheon and von Dohlen on "*Ca*. Moranella endobia" PCIT, which indicated that only 27 proteins harbor N-terminal secretory SP sequences (McCutcheon and von Dohlen, 2011). Since some of these proteins display strict cytoplasmic distribution in other bacteria, the corresponding predictions could represent false positive results or, alternatively, the gain of SP mimicking sequences by traditionally non-secretory proteins. The latter situation may reflect possible adaptations of the corresponding "*Ca*. Moranella endobia" proteins to the nested endosymbiosis. In any case, most of the predicted SP sequences are harbored by typical cell envelope proteins, while only a small amount of essential cytoplasmic proteins have been identified in the screening. Furthermore, the two ribosomal proteins detected (RplK and RplL) are also encoded by the "*Ca*. Tremblaya princeps" genome. Therefore, it is not plausible that a canonical Sec translocon could be used by "*Ca*. Moranella endobia" in order to supply its beta-proteobacterial partner with the essential proteins it lacks (Chapter 3).

Permeability of the Sec translocon machinery in "*Ca*. Moranella endobia". Several dominant mutations allowing the exportation of SP-defective precursors or even SP-less proteins led to the molecular characterization of the Sec translocon machinery in *E. coli* (Khatib and Belin, 2002; Smith et al., 2005). Collectively known as *prl* alleles, these mutations are able to expand the repertoire of secretory proteins. Mutations affecting a total of 37 codons at genes *sec*A (*prlD* mutants) (Fikes and Bassford, 1989; Huie and Silhavy, 1995), *sec*E (*prlG* mutants), *sec*G (*prlH*

mutants) (Bost and Belin, 1995) and *sec*Y (*prlA* mutants) (Smith et al., 2005) have been linked to well characterized *prl* phenotypes in *E. coli*. Since the putative occurrence of these mutations could allow for protein exportation in an unconventional SP-independent manner, we performed an *in silico* evaluation of the Sec translocon permeability in "*Ca*. Moranella endobia". No changes were observed in 33 out of the 37 amino acids known to be affected in *E. coli prl* mutants, with 19 cases displaying synonymous substitutions (Table S1 in Chapter 3, page 92). Non-synonimous changes with respect to *E. coli* homologous proteins affect SecA (E148D), SecG (N50T, I65V) and SecY (A277 of *E. coli* corresponds to "*Ca*. Moranella endobia" T279, according to the alignment). However, the impact of these mutational changes might be negligible, since amino acid polarity and/or molecular weight is almost unaffected. Moreover, the effect of two of these mutations (A277E in SecY and E148K in SecA) is not clear, since they were found in *E. coli* double mutant strains whose additional changes (L407R and Y134S, respectively) were individually linked to *prl* phenotypes.

According to our results, no evidences for abnormal permeability of the Sec translocon are observable in "*Ca*. Moranella endobia", suggesting that SP sequences would be still necessary for proteins exportation in a Sec-dependent manner. However, analyzed residues are probably just a fraction of those potentially relevant for the proper functioning of the Sec machinery. Additionaly, overexpression of some of its components, such as SecA, had also been linked to *prl* phenotypes (Fikes and Bassford, 1989). Thus, we cannot rule out that protein communication in the nested system would be facilitated by a putatively relaxed permeability of "*Ca*. Moranella endobia" Sec translocon.

4.3.2.2. The mechanosensitive channel MscL

No active molecular machinery in "*Ca*. Moranella endobia" seems able to deal with the protein traffic predicted to occur in the nested endosymbiosis. Since only a short set of "*Ca*. Moranella endobia" predicted proteins harbor secretory SP sequences, the potential relevance of passive and nonspecific transport mechanisms such as pores formed by MscL, which is maintained in "*Ca*. Moranella endobia", should be taken into account (Chapter 2). Neither this one nor the rest of bacterial mechanosensitive ion channels are frequently found in obligatory endosymbionts of insects, including "*Ca*. Tremblaya princeps".

MscL forms, mainly as a homopentamer (Sukharev et al., 1999), a mechanosensitive channel of large conductance involved in the transduction of membrane osmotic/mechanical stress into an electrochemical response (Oakley et al., 1999). It includes two transmembrane segments (TM1, TM2) connected by a periplasmic loop and a cytoplasmatic helix connected to TM2 through a linker (Blount et al., 1996). It reaches the bacterial inner membrane in a YidC-dependent manner (Pop et al., 2009). MscL pores are known to release low molecular weight molecules (e.g., ions, metabolites, osmoprotectants) but also small proteins under 6.5 kDa (van den Bogaart et al., 2007) in response to osmotic downshock, preventing cell lysis (Sukharev, 1999; Berrier et al., 1992). There is some controversy regarding the passage of thioredoxin (11.5 kDa) through MscL pores. Additionally, the 41-kDa chaperone DnaK (Berrier et al., 2000; el Yaagoubi et al., 1994),

the 52-kDa elongation factor Tu (Berrier et al., 2000; Jacobson and Rosenbusch, 1976), and the 142-kDa enterobactin synthase EntF (Hantash et al., 1997) have been suggested to go through it in osmotically stressed *E. coli*, although no direct demonstration has been performed (van den Bogaart et al., 2007). The higher metabolic activity of "*Ca*. Moranella endobia" (in comparison with that predicted for "*Ca*. Tremblaya princeps") is expected to increase osmotic pressure on its membranes. Thus, conservation of *msc*L might allow "*Ca*. Moranella endobia" to deal with recurrent osmotic stress conditions derived from the extremely biased distribution of metabolic capabilities between the nested endosymbionts. Moreover, alleviation of osmotic pressure by the action of MscL pores could lead to the release of (at least) some cytoplasmic proteins in an unspecific manner so that the participation of MscL in their molecular communication cannot be ruled out (Chapter 3).

4.3.2.3. Additional ways for unspecific protein exportation

In spite of MscL expected limitations regarding protein extrusion (van den Bogaart et al., 2007), osmotic shock itself can lead to the release of up to 10% of *E. coli* protein content, including periplasmic and cytoplasmic components under 100 kDa (Vázquez-Laslop et al., 2004). In contrast to other bacteria, almost half (44%) of "*Ca*. Tremblaya princeps" predicted ORFs are under 0.5 kb, likely because of the great abundance of ribosomal proteins. As for "*Ca*. Moranella endobia", the 74% of its predicted ORFs are larger than 0.5 kb. However, more than 95% of "*Ca*. Moranella endobia" predicted proteins are under 100 kDa, including most of the HR-related proteins apparently responsible for the CE noticed in the "*Ca*. Tremblaya princeps" ribosomal operons (Table 2 in Chapter 3, page 83) (Baumann et al., 2002; McCutcheon and von Dohlen, 2011).

Vázquez-Laslop and coworkers (2001) have suggested that osmotic shock causes transient perforation of the plasma membrane, leading proteins to exit the damaged cells through a molecular sieve, whose limiting size is determined by the surrounding peptidoglycan mesh. The recent discovery of several genes involved in peptidoglycan biosynthesis (*mur*ABCDE) and recycling (*mlt*D and *ami*D) in the *P. citri* nuclear genome suggested that the strength of the murein sacculus might be controlled by the host (Husnik et al., 2013). If this is true, it is tempting to speculate that the peptidoglycan mesh could allow the passage of even larger proteins (i.e., proteins equal or larger than 100 kDa) in a controlled manner. Furthermore, "*Ca*. Moranella endobia" outer membrane might be more fluid and permeable. Instead of the LPS typically observed in gram-negative bacteria, it putatively contains lipid IV_A only. This particular composition might also facilitate the efflux of large proteins (Meredith et al., 2006).

4.3.3. Hypothesis on the mechanisms allowing protein exportation

The determination of the nested organization of the bacterial consortium early suggested a putative facilitation of gene products exchange between both endosymbionts (von Dohlen et al., 2001). In coherence, studies performed on the Pseudococcinae mealybugs *P. kraunhiae* and *P. comstocki*, revealed the parallel dynamics of both the beta- and the gamma-endosymbionts along the insect life cycle, an indication of the intimate association between these bacteria (Kono et al., 2008). Based on the polyphyletic evolutionary origins of the gamma-endosymbionts, their apparent absence in some Pseudococcinae mealybugs (i.e., those belonging to *Ferrisia* and *Maconellicoccus* groups, at least), and their preferential elimination during the endosymbiotic system degradation in male insects, a dominant role for "*Ca*. Tremblaya princeps" was assumed (Thao et al., 2002; Kono et al., 2008). However, the dramatic functional erosion revealed by the genomic characterization of "*Ca*. Tremblaya princeps" from *P. citri*, ruled out this possibility (Chapter 1; McCutcheon and von Dohlen, 2011). In the absence of any evident explanation for the molecular communication between endosymbionts, simple lysis of "*Ca*. Moranella endobia" cells was suggested as the mechanism for "*Ca*. Tremblaya princeps" acquisition of the required essential gene products (McCutcheon and von Dohlen, 2011). This hypothesis was supported by the observation of a subset of amorphous, apparently degenerating, gammaproteobacterial cells in the bacteriocytes of the Pseudoccocinae mealybug *C. azalea* (Koga et al. 2013). Based on the detection of genes of prokaryotic origin involved in peptidoglycan biosynthesis in the *P. citri* genome, it has been suggested that the lysis of "*Ca*. Moranella endobia" cells would be controlled by the host. This is coherent with early observations by Paul Buchner (1965), who reported at least two morphological forms of "*Ca*. Moranella endobia": a "reproductive form" in which cells were in the process of dividing and small sized; and a "degenerative phase" in which cells were unevenly shaped and elongated. Apparently synchronized within a bacteriocyte, these forms were dependent on the insect developmental stage. Husnik and coworkers (2013) stated that insect control may affect "*Ca*. Moranella endobia" only, assuming that "*Ca*. Tremblaya princeps" likely uses host-derived membranes to define its cytoplasm. However, as in many other P-endosymbionts, "*Ca*. Tremblaya princeps" envelope is formed by three membrane bilayers (von Dohlen et al., 2001). Moreover, they do not explain the way controlled cell-lysis could affect only a set of "*Ca*. Moranella endobia" cells inside each "*Ca*. Tremblaya princeps" cell, which must be assumed in order to avoid the extinction of the nested consortium. In coherence with this idea, no empty "*Ca*. Tremblaya princeps" cells have been ever reported (von Dolhen et al., 2001).

Taking into account the unprecedented conservation of the mechanosensitive channel MscL by "*Ca*. Moranella endobia", as well as the expected osmotic imbalance within the nested consortium, an alternative scenario can be conceived in which horizontally-acquired genes would not lead directly to cell lysis but, instead, would regulate the strength of the bacterial peptidoglycan mesh. If this was correct, only small quantities of "*Ca*. Moranella endobia" proteins would be present in "*Ca*. Tremblaya princeps" cytoplasm at a given time, contrary to what would be expected if they were released by cell lysis. To test this prediction, we conducted experimental

immunolocation studies in order to determine the spatial distribution of two "*Ca*. Moranella endobia" proteins, the chaperone GroEL and the mechanosensitive channel MscL, across the nested endosymbiotic system (Chapter 3). Empirical assays have been rarely applied to the study of the molecular communication between endosymbionts, traditionally approached through *in silico* predictions only. Effective synthesis of interesting gene products had been eventually checked at the transcriptional and/or translational level (Hara et al., 1990; Charles et al., 1997; Tamas et al., 2008; McCutcheon and von Dohlen, 2011; Poliakov et al., 2011; Fan et al., 2013). However, subcellular proteins location had been rarely assayed (Hara et al., 1990; Tamas et al., 2008; Nakabachi et al., 2014), even though it could be relevant in order to evaluate protein functionality and tight molecular complementation is expected to occur among the members of most endosymbiotic consortia.

GroEL is an essential protein required for the proper folding of a wide range of cytosolic proteins (Kerner et al., 2005) It was used as a target in our study because it is highly expressed in endosymbiotic systems, where it apparently alleviates the destabilizing effects of non-synonymous mutations during protein folding (Baumann, 2005; Fares et al., 2002; Tokuriki and Tawfik, 2009). Detection of "*Ca*. Moranella endobia" GroEL was performed using polyclonal antibodies originally raised against GroEL from *B*. *aphidicola* APS, endosymbiont of the pea aphid, which also belongs to the Gammaproteobacteria (Hara et al., 1990). We showed the protein to be mostly confined inside "*Ca*. Moranella endobia" cells, since fluorescent signal aggregated into bacilliform bodies, following a pattern that accurately mimics that obtained when specific "*Ca*. Moranella endobia" detection is performed by FISH experiments (Figure 1 in Chapter 3, page 84). GroEL is a highly conserved protein. Nevertheless, exclusive detection of the "*Ca*. Moranella endobia" protein was expectable, since "*Ca*. Moranella endobia" and "*Ca*. Tremblaya princeps" homologs display important differences in terms of inmuno-histochemical detection. Most of the identical residues are located in short motives with null immunogenic potential (Figure S1 in Chapter 3, page 91), so that only half of the conserved residues map on putative minimal epitopes (5-8 residues), and only 38% of them map on amino acidic motives over 10 residues in length.

The detection of MscL was approached using polyclonal antibodies against cytoplasmic/periplasmic segments of the "*Ca*. Moranella endobia" protein. Once again, only "*Ca*. Moranella endobia" showed detectable levels of MscL (Figure 1 in Chapter 3, page 84). Moreover, in coherence with its expected subcellular location, the shiner signal was recurrently detected in cell membranes. Occasionally, a few degenerating cells might show some MscL staining in the "*Ca*. Tremblaya princeps" cytoplasm, suggesting that the protein is not performing its functional role at the membrane in this bacterium.

In summary, we did not find evidences for the involvement of the Sec translocon machinery in the broad molecular communication predicted to exist between both endosymbionts. Neither putatively exported-protein adaptations, nor mutational changes allowing abnormal permeability of the Sec translocon machinery were noticed. Thus, *in silico* analyses suggest that exportation

of “*Ca*. Moranella endobia” proteins should take place in a nonspecific-manner. This is coherent with cell lysis as the mechanism allowing for the share of the complete "*Ca*. Moranella endobia" cytoplasmic content in the nested system, as early suggested by McCutcheon and von Dohlen. However, even if controlled and/or sporadic cell lysis can occur, empirical location of "*Ca*. Moranella endobia" GroEL and MscL appears to rule out a massive supply of proteins to "*Ca*. Tremblaya princeps", as expected if molecular communication was accomplished mostly by cell lysis. In contrast, the role of passive and non-specific communication gates, such as transient perforation of the plasma membrane and membrane MscL channels, both controlled by osmotic stress, can provide alternative explanations for the endosymbionts molecular communication. Putatively recurrent osmotic stress events could result from the extremely biased distribution of metabolic capabilities in the nested system. Moreover, the unusual composition of the outer membrane lipopolysaccharides, together with the putative modulation of the peptidoglycan mesh strength by the insect host, might help on the provision of essential proteins from “*Ca*. Moranella endobia” to the “*Ca*. Tremblaya princeps” cytoplasm.

5. CONCLUSIONS

1. Sequencing and characterization of the complete genome of "*Ca*. Tremblaya princeps" PCVAL confirmed the presence of atypical traits for such a long-term P-endosymbiont (i.e., high GC-content, low gene density, concerted evolution of paralogous *loci*). In spite of this, "*Ca*. Tremblaya princeps" shows one of the most reduced genomes known so far.
2. Neither the abundance of pseudogenes nor the expansion of tandem repeats can fully explain the low gene density noticed in the drastically reduced "*Ca*. Tremblaya princeps" genome.
3. Comparative genomics between "*Ca*. Tremblaya princeps" PCVAL and "*Ca*. Tremblaya phenacola" PAVE suggests that the partial genome duplication event involving the ribosomal operon and its closest genomic context occurred before the divergence of subfamilies Pseudococcinae and Phenacoccinae.
4. A partial genome duplication event and the concerted evolution of the corresponding paralogous *loci* was also noticed in the genome of "*Ca*. Moranella endobia".
5. Sequencing and characterization of the genomic regions *leu*A-*rrs*1 y *prs*-*rrs*2 of "*Ca*. Tremblaya princeps" from *P. ficus* confirmed the very recent occurrence of HR events in "*Ca*. Tremblaya princeps" strains belonging to the clade E.
6. The presence of HR-related genes in "*Ca*. Moranella endobia" from *P. citri* and *P. ficus* but also in the independently-acquired gammaproteobacterial endosymbionts from *P. longispinus* and *P. viburni* support the link between nested endosymbiosis and the atypical evolution of "*Ca*. Tremblaya princeps" genome.
7. The characterization of repetitive sequences in the genomes of "*Ca*. Tremblaya princeps" and "*Ca*. Moranella endobia" suggests the first to be more sensitive to the action of the HR-related genes encoded by the latter.
8. The functional analysis of both endosymbionts genomes suggest that (1) "*Ca*. Moranella endobia" is the consortium single source of energy and that (2) "*Ca*. Tremblaya princeps" needs to import a diverse set of proteins exclusively encoded by "*Ca*. Moranella endobia" genes. Complementation in the bacterial consortium is both nutritional and informational, so that both endosymbionts could be considered as part of a novel, unprecedented biological entity.
9. Both members of the microbial consortium collaborate in the biosynthesis of essential amino acids, as usually observed in endosymbionts from plant-sucking insects. Endosymbionts specialization regarding essential amino acids supply is recurrently found among Pseudococcinae mealybugs.
10. At least one HGT event has influenced the evolution of essential amino acids biosynthesis in "*Ca*. Tremblaya phenacola" from *P. peruvianus* and *P. madeirensis*.
11. "*Ca*. Moranella endobia" conservation of the mechanosensitive channel MscL and protein immunostaining assays suggest that molecular communication between nested endosymbionts might be mediated by osmotic stimuli, pointing out a sporadic exportation pattern.

6. ABBREVIATIONS & ACRONYMS

Abbreviations & Acronyms	Definition
ABC	ATP binding cassette
ABI	Applied Biosystems, Inc
AAT	Aspartate aminotransferase
BCAT	Branched-chain amino acid aminotransferase
BI	Bayesian inference
BSA	Bovine serum albumin
C	Cytoplasm
CBL	Cystathionine beta-lyase
CDS	Coding DNA sequence
CE	Concerted evolution
CGL	Cystathionine gamma-lyase
CI	Cytoplasmic incompatibility
CP	Co-primary
CS	Cleavage site
CTAB	Cetyltrimethylammonium bromide
DAPI	4',6-diamidino-2-phenylindole
DG	Primers degeneration
DR	Direct repeats
dN	Nonsynonymous substitution per nonsynonymous site
dS	Synonymous substitution per synonymous site
EC	Enzyme commission number
EF-1a	Elongation factor 1 alpha
FISH	Fluorescence *in situ* hybridization
GLRaV	Grapevine leafroll-associated viruses
HGT	Horizontal gene transfer
HP1	Heterochromatic protein 1
HR	Homologous recombination
IGR/ITS	Intergenic region
IR	Inverted repeats
LCA	Last common ancestor
LPS	Lipopolysaccharydes
LSCM	Laser-scanning confocal microscopy
ML	Maximun likelihood

MP	Maximum parsimony
MW	Molecular weight
OAT	Ornithine aminotransferase
OMP	Outer membrane proteins
ORF	Open reading frame
P	Primary/Periplasm
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEP	Phosphoenol pyruvate
PL	Protein length
PTS	Phosphotransferase system
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
S	Secondary
SD	Standard deviation
SET	Serial endosymbiotic theory
SP	Signal peptide
SRD	Sex ratio distortion
SRP	Signal recognition particle
TCA	Tricarboxylic acid cycle
TDH	Threonine dehydratase
$_{T}$DNA	Total DNA
TEM	Transmission electron microscopy
TM	Transmembrane segment

7. REFERENCES

Abby S, Daubin V. 2007. Comparative genomics and the evolution of prokaryotes. Trends Microbiol. 15:135–41.

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. Nat. Genet. 32:402–7.

Allen JM, Reed DL, Perotti MA, Braig HR. 2007. Evolutionary relationships of "*Candidatus* Riesia spp.," endosymbiotic enterobacteriaceae living within hematophagous primate lice. Appl. Environ. Microbiol. 73:1659–64.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–402.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl D a. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56:1919–25.

Aras RA, Kang J, Tschumi AI, Harasaki Y, Blaser MJ. 2003. Extensive repetitive DNA facilitates prokaryotic genome plasticity. Proc. Natl. Acad. Sci. U. S. A. 100:13579–84.

Ashbolt NJ, Inkerman P A. 1990. Acetic acid bacterial biota of the pink sugar cane mealybug, *Saccharococcus sacchari*, and its Environs. Appl. Environ. Microbiol. 56:707–12.

Ausubel F. 1999. Short protocols in molecular biology: a compendium of methods from current protocols in molecular biology. Wiley, New York.

Awmack CS, Leather SR. 2002. Host plant quality and fecundity in herbivorous insects. Annu. Rev. Entomol. 47:817–44.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD. 2010. Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 26:2811–7.

Baumann L, Baumann P. 2005. Cospeciation between the primary endosymbionts of mealybugs and their hosts. Curr. Microbiol. 50:84–7.

Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P. 2002. The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. Appl. Environ. Microbiol. 68:3198–3205.

Baumann P. 2005. Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. 59:155–89.

Beard SJ, Hashim R, Wu G, Binet MR, Hughes MN, Poole RK. 2000. Evidence for the transport of zinc(II) ions via the pit inorganic phosphate transport system in *Escherichia coli*. FEMS Microbiol. Lett. 184:231–5.

Ben-Dov Y. 1994. A systematic catalogue of the mealybugs of the world (Insecta: Homoptera: Coccoidea: Pseudococcidae and Putoidae) with data and geographical distribution, host plants, biology and economic importance. Intercept Limited, Andover, UK.

Bennett GM, Moran N a. 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. Genome Biol. Evol. 5:1675–88.

Berrier C, Coulombe A, Szabo I, Zoratti M, Ghazi A. 1992. Gadolinium ion inhibits loss of metabolites induced by osmotic shock and large stretch-activated channels in bacteria. Eur. J. Biochem. 206:559–65.

Berrier C, Garrigues A, Richarme G, Ghazi A. 2000. Elongation factor Tu and DnaK are transferred from the cytoplasm to the periplasm of *Escherichia coli* during osmotic downshock presumably via the mechanosensitive channel *msc*L. J. Bacteriol. 182:248–51.

Beuning LL, Murphy P, Wu E, Batchelor TA, Morris BA. 1999. Molecular-based approach to the differentiation of mealybug (Hemiptera: Pseudococcidae) species. J. Econ. Entomol. 92:463–72.

Blount P, Sukharev SI, Moe PC, Schroeder MJ, Guy HR, Kung C. 1996. Membrane topology and multimeric structure of a mechanosensitive channel protein of *Escherichia coli*. EMBO J. 15:4798–805.

Bost S, Belin D. 1995. A new genetic selection identifies essential residues in SecG, a component of the *Escherichia coli* protein export machinery. EMBO J. 14:4412–21.

Brumin M, Kontsedalov S, Ghanim M. 2011. Rickettsia influences thermotolerance in the whitefly *Bemisia tabaci* B biotype. Insect Sci. 18:57–66.

Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. Interscience, New York.

Buglia GL, Dionisi D, Ferraro M. 2009. The amount of heterochromatic proteins in the egg is correlated with sex determination in *Planococcus citri* (Homoptera, Coccoidea). Chromosoma 118:737–46.

Buglia GL, Ferraro M. 2004. Germline cyst development and imprinting in male mealybug *Planococcus citri*. Chromosoma 113:284–94.

Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream M-A. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 24:2672–6.

Casjens S. 1998. The diverse and dynamic structure of bacterial genomes. Annu. Rev. Genet. 32:339–77.

Cavalieri V, Mazzeo G, Garzia GT, Buonocore E, Russo A. 2008. Identification of *Planococcus ficus* and *Planococcus citri* (Hemiptera: Pseudococcidae) by PCR-RFLP of COI gene. Zootaxa. 68: 65-68.

Charles H, Heddi a, Guillaud J, Nardon C, Nardon P. 1997. A molecular aspect of symbiotic interactions between the weevil *Sitophilus oryzae* and its endosymbiotic bacteria: over-expression of a chaperonin. Biochem. Biophys. Res. Commun. 239:769–74.

Chen X, Li S, Aksoy S. 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J. Mol. Evol. 48:49–58.

Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G, Heddi A. 2008. Long-term evolutionary stability of bacterial endosymbiosis in curculionoidea: additional evidence of symbiont replacement in the Dryophthoridae family. Mol. Biol. Evol. 25:859–68.

Copland MJ, Tingle CC, Saynor M, Panis A. 1985. Biology of glasshouse mealybugs and their predators and parasitoids, p. 82-86. In Hussey NW, Scopes N (ed), Biological pest control. Cornell University Press, Ithaca, NY.

Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. 1992. Crystal structures explain functional properties of two *E. coli* porins. Nature 358:727–33.

Cox JM. 1983. An experimental study of morphological variation in mealybugs (Homoptera: Coccoidea: Pseudococcidae). Syst. Entomol. 8:361–382.

Daane KM, Middleton MC, Sforza R, Cooper ML, Walton VM, Walsh DB, Zaviezo T, Almeida RPP. 2011. Development of a multiplex PCR for identification of vineyard mealybugs. Environ. Entomol. 40:1595–603.

Dadd RH. 1985. Nutrition: organisms, p. 313–390. In Kerkut GA, Gilbert LI (ed) Comprehensive insect physiology, biochemistry and pharmacology, vol 4. Pergamon Press: Oxford.

Dale C, Welburn SC. 2001. The endosymbionts of tsetse flies: manipulating host-parasite interactions. Int. J. Parasitol. 31:628–31.

Dasch GA, Weiss E, Chang K. 1984. Endosymbionts of insects, p. 881–883. In Krieg NR (ed), Bergeys Manual of Systematic Bacteriology, vol. 1. Williams & Williams, Baltimore.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9:772.

Davidson DW, Cook SC, Snelling RR, Chua TH. 2003. Explaining the abundance of ants in lowland tropical rainforest canopies. Science 300:969–72.

de Duve C. 2007. The origin of eukaryotes: a reapparisal. Nat Rev Genet. 8:395-405.

de Keyzer J, van der Does C, Driessen AJM. 2003. The bacterial translocase: a dynamic protein channel complex. Cell. Mol. Life Sci. 60:2034–52.

Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. 2009. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. Proc. Natl. Acad. Sci. U. S. A. 106:9063–8.

Demontis MA, Ortu S, Cocco A, Lentini A, Migheli Q. 2007. Diagnostic markers for *Planococcus ficus* (Signoret) and *Planococcus citri* (Risso) by random amplification of polymorphic DNA-polymerase chain reaction and species-specific mitochondrial DNA primers. J. Appl. Entomol. 131:59–64.

Dinant S, Bonnemain J-L, Girousse C, Kehr J. Phloem sap intricacy and interplay with aphid feeding. C. R. Biol. 333:504–15.

Doerrler WT, Gibbons HS, Raetz CRH. 2004. MsbA-dependent translocation of lipids across the inner membrane of *Escherichia coli*. J. Biol. Chem. 279:45102–9.

Douglas AE. 1989. Mycetocyte symbiosis in insects. Biol. Rev. Camb. Philos. Soc. 64:409–34.

Douglas AE. 1993. The nutritional quality of phloem sap utilized by natural aphid populations. Ecol. Entomol. 18:31–38.

Douglas AE. 1994. Symbiotic interactions. Oxford University Press, Oxford.

Downie DA, Gullan PJ. 2004. Phylogenetic analysis of mealybugs (Hemiptera: Coccoidea: Pseudococcidae) based on DNA sequences from three nuclear genes, and a review of the higher. Syst. Entomol. 238–259.

Downie DA, Gullan PJ. 2005. Phylogenetic congruence of mealybugs and their primary endosymbionts. J. Evol. Biol. 18:315–24.

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Hurst GD. 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol. 6:27.

el Yaagoubi A, Kohiyama M, Richarme G. 1994. Localization of DnaK (chaperone 70) from *Escherichia coli* in an osmotic-shock-sensitive compartment of the cytoplasm. J. Bacteriol. 176:7074–8.

Facey SJ, Neugebauer SA, Krauss S, Kuhn A. 2007. The mechanosensitive channel protein MscL is targeted by the SRP to the novel YidC membrane insertion pathway of *Escherichia coli*. J. Mol. Biol. 365:995–1004.

Fan Y, Thompson JW, Dubois LG, Moseley MA, Wernegreen JJ. 2013. Proteomic analysis of an unculturable bacterial endosymbiont (*Blochmannia*) reveals high abundance of chaperonins and biosynthetic enzymes. J. Proteome Res. 12:704–18.Fares MA, Moya A, Barrio E. 2004. GroEL and the maintenance of bacterial endosymbiosis. Trends Genet. 20:413-6.

Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E. 2002. Endosymbiotic bacteria: *gro*EL buffers against deleterious mutations. Nature 417:398.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3:121.

Feldhaar H, Straka J, Krischke M, Berthold K, Stoll S, Mueller MJ, Gross R. 2007. Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. BMC Biol. 5:48.

Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360:1427–34.

Ferrari J, Darby AC, Daniell TJ, Godfray HCJ, Douglas AE. 2004. Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. Ecol. Entomol. 29:60–65.

Fikes JD, Bassford PJ. 1989. Novel *sec*A alleles improve export of maltose-binding protein synthesized with a defective signal peptide. J. Bacteriol. 171:402–9.

Franke IH, Fegan M, Hayward C, Leonard G, Sly LI. 2000. Molecular detection of Gluconacetobacter sacchari associated with the pink sugarcane mealybug *Saccharicoccus sacchari* (Cockerell) and the sugarcane leaf sheath microenvironment by FISH and PCR. FEMS Microbiol. Ecol. 31:61–71.

Franke-Whittle IH, O'shea MG, Leonard GJ, Sly LI. 2004. Molecular investigation of the microbial populations of the pink sugarcane mealybug, *Saccharicoccus sacchari*. Ann. Microbiol 54:455–470.

Fukatsu T, Ishikawa H. 1996. Phylogenetic position of yeast-like symbiont of *Hamiltonella styraci* (Homoptera, Aphididae) based on 18S rDNA sequence. Insect Biochem. Mol. Biol. 26:383-388.

Fukatsu T, Nikoh N. 2000. Endosymbiotic microbiota of the bamboo pseudococcid *Antonina crawii* (Insecta, Homoptera). Appl. Environ. Microbiol. 66:643–50.

Garcia JR, Gerardo NM. 2014. The symbiont side of symbiosis: do microbes really benefit? Front. Microbiol. 5.

Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res. 37:D136–40.

Gatehouse LN, Sutherland P, Forgie SA, Kaji R, Christeller JT. 2012. Molecular and histological characterization of primary (betaproteobacteria) and secondary (gammaproteobacteria) endosymbionts of three mealybug species. Appl. Environ. Microbiol. 78:1187–97.

Gherna RL, Werren JH, William W, Cote R, Woese CR, Mandelco L, Brenner DJ. 1991. *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. International Journal of Systematic Bacteriology. 41:563-565.

Gil R, Belda E, Gosalbes MJ, Delaye L, Vallier A, Vincent-Monégat C, Heddi A, Silva FJ, Moya A, Latorre A. 2008. Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil *Sitophilus oryzae*. Int. Microbiol. 11:41-8.

Gil R, Silva FJ, Peretó J, Moya A. 2004. Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. Rev. 68:518–37.

Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Hölldobler B, van Ham RCHJ, Gross R, Moya A. 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc. Natl. Acad. Sci. U. S. A. 100:9388–93.

Goldasteh S, Talebi AA, Fathipour Y, Ostovan H, Zamani A, Shoushtari RV. 2009. Effect of temperature on life history and population growth parameters of *Planococcus citri* (Homoptera, Pseudococcidae) on coleus [*Solenostemon scutellarioides* (L.) Codd.]. Arch. Biol. Sci., Belgrade 61:329–336.

González-Domenech CM, Belda E, Patiño-Navarrete R, Moya A, Peretó J, Latorre A. 2012. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two *Blattabacterium cuenoti* strains, primary endosymbionts of cockroaches. BMC Microbiol. 12 Suppl 1:S5.

Gosalbes MJ, Lamelas A, Moya A, Latorre A. 2008. The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. J Bacteriol. 190:6026–9.

Gotoh T, Noda H, Ito S. 2007. *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. Heredity (Edinb). 98:13–20.

Gray M, Spencer D. 1996. Organellar evolution, p. 109–126. In Roberts D, Sharp P, Altderson G, Collins M (ed). Evolution of Microbial Life. Cambridge University Press, Cambrige.

Gruwell ME, Hardy NB, Gullan PJ, Dittmar K. 2010. Evolutionary relationships among primary endosymbionts of the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae). Appl. Environ. Microbiol. 76:7521–5.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Gullan PJ, Martin JH. 2009. Sternorrhyncha (jumping plant-lice, whiteflies, aphids, and scale insects), p. 957–967. In Resh VH, Cardé RT (ed), Encyclopedia of Insects. Elsevier, San Diego.

Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol. 2:95–108.

Hantash FM, Ammerlaan M, Earhart CF. 1997. Enterobactin synthase polypeptides of *Escherichia coli* are present in an osmotic-shock-sensitive cytoplasmic locality. Microbiology 143 (Pt 1):147–56.

Hara E, Fukatsu T, Ishikawa H. 1990. Characterization of symbionin with anti-symbionin antiserum. Insect Biochem. 20:429–436.

Hardy NB, Gullan PJ, Hodgson CJ. 2008. A subfamily-level classification of mealybugs (Hemiptera: Pseudococcidae) based on integrated molecular and morphological data. Syst. Entomol. 33:51–71.

Hardy NB. 2007. Phylogenetic utility of dynamin and triose phosphate isomerase. Syst. Entomol. 32:396–403.

Hastings PJ. 2010. Mechanisms of ectopic gene conversion. Genes. 1:427–39.

Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6:e1001115.

Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, Chiel E, Duckworth VE, Dennehy TJ, Zchori-Fein E, Hunter MS. 2011. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. Science 332:254–6.

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol. 4:e337.

Houk E, Griffiths GW. 1980. Intracellular symbiotes of Homoptera. A. Rev. Entomol. 25:161-187.

Huie JL, Silhavy TJ. 1995. Suppression of signal sequence defects and azide resistance in *Escherichia coli* commonly result from the same mutations in *sec*A. J. Bacteriol. 177:3518–26.

Hunter MS, Perlman SJ, Kelly SE. 2003. A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. Proc. Biol. Sci. 270:2185–90.

Hurst GDD, Bandi C, Sacchi L, Cochrane AG, Bertrand D, Karaca I, Majerus MEN. 1999. *Adonia variegata* (Coleoptera: Coccinellidae) bears maternally inherited Flavobacteria that kill males only. Parasitology 118:125–134.

Hurst GDD, Hammarton TC, Bandi C, Majerus TMO, Bertrand D, Majerus MEN. 1997. The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria. Genet. Res. 70:1–6.

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson ACC, von Dohlen CD, Fukatsu T, McCutcheon JP. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153:1567–78.

Iaccarino FM, Tremblay E. 1970. Osservazioni ultrastrutturali sulla endosimbiosi del *Planococcus citri* (Risso) (Homoptera, Coccoidea). Boll. Lab. Entomol. Agrar. Portici 28:35–48.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11:97–108.

Jacobson GR, Rosenbusch JP. 1976. Abundance and membrane association of elongation factor Tu in *E. coli*. Nature 261:23–6.

Johnson DC, Dean DR, Smith AD, Johnson MK. 2005. Structure, function, and formation of biological iron-sulfur clusters. Annu. Rev. Biochem. 74:247–81.

Kantheti P, Jayarama KS, Chandra HS. 1996. Developmental analysis of a female-specific 16S rRNA gene from mycetome-associated endosymbionts of a mealybug, *Planococcus lilacinus*. Insect Biochem. Mol. Biol. 26:997–1009.

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang H-C, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 122:209–20.

Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muñiz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD. 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res. 39:D583–90.

Khatib K, Belin D. 2002. A novel class of *sec*A alleles that exert a signal-sequence-dependent effect on protein export in *Escherichia coli*. Genetics 162:1031–43.

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva E V, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Krause E, Mittapalli O, Margam VM, Li H-M, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner R V, Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. U. S. A. 107:12168–73.

Kneip C, Lockhart P, Voss C, Maier UG. 2007. Nitrogen fixation in eukaryotes - New models for symbiosis. BMC Evol. Biol. 7:55.

Koga R, Bennett GM, Cryan JR, Moran NA. 2013. Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. Environ. Microbiol. 15:2073–81.

Koga R, Nikoh N, Matsuura Y, Meng X-Y, Fukatsu T. 2012. Mealybugs with distinct endosymbiotic systems living on the same host plant. FEMS Microbiol. Ecol. 83:93–100.

Koga R, Tsuchida T, Fukatsu T. 2003. Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. Proc. Biol. Sci. 270:2543–50.

Köhler M and Schwartz W. 1962. Über die beziehungen zwischen symbionten und wirtsorganismus bei *Pseudococcus citri*, *Ps. maritimus* und *Orthezia insignis*. Z. Allg. Mikrobiol. 2: 10-31.

Kol-Maimon H, Ghanim M, Franco JC, Mendel Z. 2014. Evidence for gene flow between two sympatric mealybug species (Insecta; Coccoidea; Pseudococcidae). PLoS One 9:e88433.

Kono M, Koga R, Shimada M, Fukatsu T. 2008. Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. Appl. Environ. Microbiol. 74:4175–84.

Kraemer L, Beszteri B, Gäbler-Schwarz S, Held C, Leese F, Mayer C, Pöhlmann K, Frickenhaus S. 2009. STAMP: Extensions to the STADEN sequence analysis package for high throughput interactive microsatellite marker design. BMC Bioinformatics 10:41.

Laflin HM, Parrella MP. 2004. Developmental biology of citrus mealybug under conditions typical of California rose production. Ann. Entomol. Soc. Am. 97:982–988.

Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. 2011. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. PLoS Genet. 7:e1002357.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–8.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11–6.

Latorre A, Durbán A, Moya A, Peretó J. The role of symbiosis in eukaryotic evolution, p. 326–343. In Gargaud M, López-García P, Martin H (ed), Origins and Evolution of Life: an Astrobiology Perspective. Cambridge University Press, New York.

Lavery PE, Kowalczykowski SC. 1992. Biochemical basis of the constitutive repressor cleavage activity of recA730 protein. A comparison to recA441 and recA803 proteins. J. Biol. Chem. 267:20648–58.

Lawson ET, Mousseau TA, Klaper R, Hunter MD, Werren JH. 2001. *Rickettsia* associated with male-killing in a buprestid beetle. Heredity (Edinb). 86:497.

Lefèvre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A. 2004. Endosymbiont phylogenesis in the dryophthoridae weevils: evidence for bacterial replacement. Mol. Biol. Evol. 21:965–73.

Liao D. 1999. Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. 64:24–30.

Lind PA, Andersson DI. 2008. Whole-genome mutational biases in bacteria. Proc. Natl. Acad. Sci. USA. 105:17878–83.

Lo N, Bandi C, Watanabe H, Nalepa C, Beninati T. 2003. Evidence for cocladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. Mol. Biol. Evol. 20:907–13.

López-Madrigal S, Balmand S, Latorre A, Heddi A, Moya A, Gil R. 2013. How does *Tremblaya princeps* get essential proteins from its nested partner *Moranella endobia* in the mealybug *Planoccocus citri*? PLoS One 8:e77307.

López-Madrigal S, Beltrà A, Resurrección S, Soto A, Latorre A, Moya A, Gil R. 2014. Molecular evidence for ongoing complementarity and horizontal gene transfer in endosymbiotic systems of mealybugs. Front. Microbiol. 5:449.

López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status. J. Bacteriol. 193:5587–5588.

López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. 2013. Mealybugs nested endosymbiosis: going into the "matryoshka" system in *Planococcus citri* in depth. BMC Microbiol. 13:74.

López-Sánchez MJ, Neef A, Peretó J, Patiño-Navarrete R, Pignatelli M, Latorre A, Moya A. 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLoS Genet. 5:e1000721.

Louis C, Giannotti J. 1974. Formation lysosomales et membranaires associees a des symbiotes et a des mycoplasmes endocellulaires d'insectes. Bull. Obs. Mer. 43-55.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–64.

Ma B, Reynolds CM, Raetz CRH. 2008. Periplasmic orientation of nascent lipid A in the inner membrane of an *Escherichia coli* LptA mutant. Proc. Natl. Acad. Sci. U. S. A. 105:13823–8.

Majerus TMO, Majerus MEN. 2010. Discovery and identification of a male-killing agent in the Japanese ladybird *Propylea japonica* (Coleoptera: Coccinellidae). BMC Evol. Biol. 10:37.

Malausa T, Fenis A, Warot S, Germain J-F, Ris N, Prado E, Botton M, Vanlerberghe-Masutti F, Sforza R, Cruaud C, Couloux A, Kreiter P. 2011. DNA markers to disentangle complexes of cryptic taxa in mealybugs (Hemiptera: Pseudococcidae). J. Appl. Entomol. 135:142–155.

Malleshaiah BK, Rajagopal K, Gowda KNM. 2000. Biology of the citrus mealybug *Planococcus citri* (Risso) (Hemiptera: Pseudococcidae). Crop Research. 20:130-133.

Margulis L. 1993. Symbiosis in Cell Evolution. W. H. Freeman & Co, New York.

Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature. 393:162–5.

McCutcheon JP, McDonald BR, Moran NA. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. U. S. A. 106:15394–9.

McCutcheon JP, McDonald BR, Moran NA. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. PLoS Genet. 5:e1000565.

McCutcheon JP, Moran NA. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. U. S. A. 104:19392–7.

McCutcheon JP, Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol. Evol. 2:708–18.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13–26.

McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr. Biol. 21:1366–72.

McFall-Ngai M. 2008. Are biologist in “future shock?” Symbiosis integrates biology across domains. Nat Rev Microbiol. 6:789–92.

Meredith TC, Aggarwal P, Mamat U, Lindner B, Woodard RW. 2006. Redefining the requisite lipopolysaccharide structure in *Escherichia coli*. ACS Chem. Biol. 1:33–42.

Millar IM. 2002. Mealybugs genera (Hemiptera: Pseudococcidae) of South Africa: identification and review. African Entomology. 10: 185-233.

Minic Z, Hervé G. 2004. Biochemical and enzymological aspects of the symbiosis between the deep-sea tubeworm *Riftia pachyptila* and its bacterial endosymbiont. Eur J Biochem. 271:3093–102.

Moe PC, Blount P, Kung C. 1998. Functional and structural conservation in the mechanosensitive channel MscL implicates elements crucial for mechanosensation. Mol. Microbiol. 28:583–92.

Montllor CB, Maxmen A, Purcell AH. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. Ecol. Entomol. 27:189–195.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on earth and in the ocean? PLoS Biol. 9: e1001127

Moran NA, Dale C, Dunbar H, Smith WA, Ochman H. 2003. Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) form a distinct clade with a small genome. Environ. Microbiol. 5:116–26.

Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr. Opin. Genet. Dev. 14:627–33.

Moran NA, Russell JA, Koga R, Fukatsu T. 2005. Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects. Appl. Environ. Microbiol. 71:3302–10.

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. U.S.A. 93:2873–8.

Mori H, Ito K. 2001. The Sec protein-translocation pathway. Trends Microbiol. 9:494–500.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–5.

Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio 3:e00036–12.

Moya A, Gil R, Latorre a. 2009. The evolutionary history of symbiotic associations among bacteria and their animal hosts: a model. Clin. Microbiol. Infect. 15 Suppl 1:11–3.

Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet. 9:218–29.

Moya A, Peretó J. 2011. Simbiosis, seres que evolucionan juntos. Síntesis, Madrid.

Munson MA, Baumann P, Clark MA, Baumann L, Moran NA, Voegtlin DJ, Campbell BC. 1991. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. J. Bacteriol. 173:6321–4.

Munson MA, Baumann P, Moran NA. 1992. Phylogenetic relationships of the endosymbionts of mealybugs (Homoptera: Pseudococcidae) based on 16S rDNA sequences. Mol. Phylogenet. Evol. 1:26–30.

Murray RG, Schleifer KH. 1994. Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes. Int. J. Syst. Bacteriol. 44:174–6.

Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. Curr. Biol. 24:R640–1.

Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, Oldham NJ, van Echten-Deckert G, Okamura K, Yamamoto K, Inoue H, Ohkuma M, Hongoh Y, Miyagishima S, Hattori M, Piel J, Fukatsu T. 2013. Defensive bacteriome symbiont with a drastically reduced genome. Curr. Biol. 23:1478–84.

Narai Y, Murai T. 2002. Individual rearing of the Japanese mealybug, *Planococcus kraunhiae* (Kuwana) (Homoptera: Pseudococcidae) on germinated broad bean seeds. Appl. Entomol. Zool. 37: 295-298.

Nogge G. 1981. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. Parasitology. 82:101-104.

Nouwen N, Driessen AJM. 2002. SecDFyajC forms a heterotetrameric complex with YidC. Mol. Microbiol. 44:1397–405.

Nováková E, Hypsa V, Moran NA. 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol. 9:143.

Oakeson KF, Gil R, Clayton AL, Dunn DM, von Niederhausern AC, Hamil C, Aoyagi A, Duval B, Baca A, Silva FJ, Vallier A, Jackson DG, Latorre A, Weiss RB, Heddi A, Moya A, Dale C. 2014. Genome degeneration and adaptation in a nascent stage of symbiosis. Genome Biol. Evol. 6:76–93.

Oakley AJ, Martinac B, Wilce MC. 1999. Structure and function of the bacterial mechanosensitive channel of large conductance. Protein Sci. 8:1915–21.

Okonechnikov K, Golosova O, Fursov M. 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166–7.

Oliver KM, Russell JA, Moran NA, Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. U. S. A. 100:1803–7.

Pan H, Li X, Ge D, Wang S, Wu Q, Xie W, Jiao X, Chu D, Liu B, Xu B, Zhang Y. 2012. Factors affecting population dynamics of maternally transmitted endosymbionts in *Bemisia tabaci*. PLoS One 7:e30760.

Paracer S, Ahmadjian V. 2000. Symbiosis. An introduction to biological associations. Oxford University Press, NY.

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL, Cerdeño-Tárraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, Whitehead S, Barrell BG, Maskell DJ. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. 35:32–40.

Patino-Navarrete R, Piulachs M-D, Belles X, Moya A, Latorre A, Pereto J. 2014. The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. Biol. Lett. 10:20140407–20140407.

Patzer SI, Hantke K. 1998. The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. Mol. Microbiol. 28:1199–210.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8:785–6.

Pittman MS, Robinson HC, Poole RK. 2005. A bacterial glutathione transporter (*Escherichia coli* CydDC) exports reductant to the periplasm. J. Biol. Chem. 280:32254–61.

Polat F, Ulgenturk S, Kaydan M.B. 2007. Developmental biology of citrus mealybug, *Planococcus citri* (Risso) (Hemiptera: Pseudococcidae), on ornamental plants, p. 177-184. In Branco M, Franco

JC, and Hodgson C (ed), Proceedings, Symposium: XI International Symposium on Scale Insect Studies, 24-27 September 2007, Oeiras, Portugal. ISA Press, Lisbon, Portugal.

Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ. 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol. Cell. Proteomics 10:M110.007039.

Polissi A, Georgopoulos C. 1996. Mutational analysis and properties of the *msb*A gene of *Escherichia coli*, coding for an essential ABC family transporter. Mol. Microbiol. 20:1221–33.

Pop OI, Soprova Z, Koningstein G, Scheffers D-J, van Ulsen P, Wickström D, de Gier J-W, Luirink J. 2009. YidC is required for the assembly of the MscL homopentameric pore. FEBS J. 276:4891–9.

Prosser WA, Douglas AE. 1992. A test of the hypotheses that nitrogen is upgraded and recycled in an aphid (*Acyrthosiphon pisum*) symbiosis. J. Insect Physiol. 38:93–99.

Rispe C, Delmotte F, Van Ham RC, Moya A. 2004. Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. Genome Res. 14:44–53.

Rocha EP, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. Trends Genet. 18:291–4.

Rocha EPC, Cornet E, Michel B. 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet. 1:e15.

Rocha EPC. 2003. DNA repeats lead to the accelerated loss of gene order in bacteria. Trends Genet. 19:600–3.

Rocha EPC. 2008. The organization of the bacterial genome. Annu. Rev. Genet. 42:211–33.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–42.

Roques A, Rabitsch W, Rasplus J, Lopez-Vaamonde C, Nentwig W, Kenis M. 2009. Handbook of alien species in Europe, p. 63-79. In DAISIE (ed), Alien terrestrial invertebrates of Europe. Springer, Dordrecht, Netherlands.

Rosenblueth M, Sayavedra L, Sámano-Sánchez H, Roth A, Martínez-Romero E. 2012. Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). J. Evol. Biol. 25:2357–68.

Rung A, Scheffer SJ, Evans G, Miller D. 2008. Molecular identification of two closely related species of mealybugs of the genus *Planococcus* (Homoptera: Pseudococcidae). Ann. Entomol. Soc. Am. 101:525–532.

Russell JA, Latorre A, Sabater-Muñoz B, Moya A, Moran NA. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol. Ecol. 12:1061–75.

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. 2012. Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. Appl. Environ. Microbiol. 78:204–10.

Saccaggi DL, Krüger K, Pietersen G. 2008. A multiplex PCR assay for the simultaneous identification of three mealybug species (Hemiptera: Pseudococcidae). Bull. Entomol. Res. 98:27–33.

Sachs JL, Hollowell AC. 2012. The origins of cooperative bacterial communities. MBio 3:e00099–12.

Saffo M. 1992. Coming to terms with a field: words and concepts in symbiosis. Symbiosis. 14:17–31.

Sandström J, Pettersson J. 1994. Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. J. Insect Physiol. 40:947–955.

Sauer C, Stackebrandt E, Gadau J, Hölldobler B, Gross R. 2000. Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus* Blochmannia gen. nov. Int. J. Syst. Evol. Microbiol. 50:1877–86.

Sawa J, Malet H, Krojer T, Canellas F, Ehrmann M, Clausen T. 2011. Molecular adaptation of the DegQ protease to exert protein quality control in the bacterial cell envelope. J. Biol. Chem. 286:30680–90.

Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. Mol. Biol. Rev. 61:262–80.

Serbus LR, Casper-Lindley C, Landmann F, Sullivan W. 2008. The genetics and cell biology of *Wolbachia*-host interactions. Annu. Rev. Genet. 42:683–707.

Shen P, Huang H V. 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 112:441–57.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–6.

Silva FJ, Latorre A, Moya A. 2003. Why are the genomes of endosymbiotic bacteria so stable? Trends Genet. 19:176–80.

Singh N, Kuppili RR, Bose K. 2011. The structural basis of mode of activation and functional diversity: a case study with HtrA family of serine proteases. Arch. Biochem. Biophys. 516:85–96.

Singh ST, Kumar J, Thomas A, Ramamurthy V V, Rajagopal R. 2013. Detection and localization of *Rickettsia* sp in mealybug. Environ. Entomol. 42:711–6.

Slansky F, Rodriguez JG. 1987. Nutritional ecology of insects, mites, spiders, and related invertebrates. John Wiley & Sons: New York.

Sloan DB, Moran NA. 2013. The evolution of genomic instability in the obligate endosymbionts of whiteflies. Genome Biol. Evol. 5:783–93.

Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Mol. Biol. Evol. 31:857–71.

Smith MA, Clemons WM, DeMars CJ, Flower AM. 2005. Modeling the effects of prl mutations on the *Escherichia coli* SecY complex. J. Bacteriol. 187:6454–65.

Spies M, Kowalczykowski SC. 2005. Homologous recombination by the RecBCD and the RecF pathways, p. 389–403. In Patrick Higgins, N (ed), The Bacterial Chromosome. ASM Press, Washington D.C.

Staden R, Beal KF, Bonfield JK. 2000. The Staden package, 1998. Methods Mol. Biol. 132:115–30.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–90.

Stewart FJ, Newton IL, Cavanaugh CM. 2005. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. Trends Microbiol. 13:439–48.

Sukharev SI, Schroeder MJ, McCaslin DR. 1999. Stoichiometry of the large conductance bacterial mechanosensitive channel of *E. coli*. A biochemical study. J. Membr. Biol. 171:183–93.

Takiya DM, Tran PL, Dietrich CH, Moran NA. 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Mol. Ecol. 15:4175–91.

Tamas I, Klasson L, Canbäck B, Näslund AK, Eriksson A-S, Wernegreen JJ, Sandström JP, Moran NA, Andersson SGE. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 296:2376–9.

Tamas I, Wernegreen JJ, Nystedt B, Kauppinen SN, Darby AC, Gomez-Valero L, Lundin D, Poole AM, Andersson SGE. 2008. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. Proc. Natl. Acad. Sci. U. S. A. 105:14934–9.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–9.

Tanada Y, Kaya HK. 1993. Insect Pathology. In Kaya HK, Vega FE (ed). Academic Press, London, UK.

Thao ML, Baumann P. 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. Appl. Environ. Microbiol. 70:3401–6.

Thao ML, Gullan P, Baumann P. 2002. Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Appl. Environ. Microbiol. 68:3190–3197.

Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. 2000. Cospeciation of psyllids and their primary prokaryotic endosymbionts. Appl. Environ. Microbiol. 66:2898–2905.

Thomas GH, Zucker J, Macdonald SJ, Sorokin A, Goryanin I, Douglas AE. 2009. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. BMC Syst. Biol. 3:24.

Tokuriki N, Tawfik DS. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459:668–73.

Tremblay E. 1989. Coccoidea endocytobiosis, p. 145-173. In Schwemmler W, Gassner G (ed), Insect Endocytobiosis: Morphology, Physiology, Genetics, Evolution. CRC Press Inc., Boca Raton, Fla.

Tremblay E. 1990. Endosymbionts, p. 275-283. In Rosen D (ed), Armored Scale Insects: Their Biology, Natural Enemies and Control, vol. 4A. Elsevier, Amsterdam, The Netherlands.

Tremblay E. 1997. Endosymbionts, p. 261-267. In Ben-Dov Y and Hodgson CJ (ed), Soft Scale Insects: Their Biology, Natural Enemies and Control, vol. 7A. Elsevier, Amsterdam, The Netherlands.

Tsai CW, Rowhani A, Golino DA, Daane KM, Almeida RPP. 2010. Mealybug transmission of Grapevine leafroll viruses: an analysis of virus-vector specificity. Phytopathology 100:830–4.

Tsuchida T, Koga R, Fukatsu T. 2004. Host plant specialization governed by facultative symbiont. Science 303:1989.

Urban JM, Cryan JR. 2012. Two ancient bacterial endosymbionts have coevolved with the planthoppers (Insecta: Hemiptera: Fulgoroidea). BMC Evol. Biol. 12:87.

van den Bogaart G, Krasnikov V, Poolman B. 2007. Dual-color fluorescence-burst analysis to probe protein efflux through the mechanosensitive channel MscL. Biophys. J. 92:1233–40.

Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 33:W455–9.

van Ham RC, Moya A, Latorre A. 1997. Putative evolutionary origin of plasmids carrying the genes involved in leucine biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids). J. Bacteriol. 179:4768–77.

van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM, Jiménez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Morán F, Moya A. 2003. Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. U. S. A. 100:581–6.

Vázquez-Laslop N, Lee H, Hu R, Neyfakh AA. 2001. Molecular sieve mechanism of selective release of cytoplasmic proteins by osmotically shocked *Escherichia coli*. J. Bacteriol. 183:2399–404.

von Dohlen CD, Kohler S, Alsop ST, McManus WR. 2001. Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. Nature 412:433–6.

Walker DE. 2000. Culturing mealybug parasites. Assoc. Educ. Res. Greenhouse Curators News. 13:1-3.

Wang X, Quinn PJ. 2010. Lipopolysaccharide: Biosynthetic pathway and structure modification. Prog. Lipid Res. 49:97–107.

Watson GW and Kubiriba J. 2005. Identification of mealybugs (Hemiptera: Pseudococcidae) on banana and plantain in Africa. African Entomology. 13: 35-47.

Weinstock O, Sella C, Chipman DM, Barak Z. 1992. Properties of subcloned subunits of bacterial acetohydroxy acid synthases. J. Bacteriol. 174:5560–5566.

Wicker C, Nardon P. 1982. Development responses of symbiotic and aposymbiotic weevils *Sitophilus oryzae* L. (Coleoptera, curculionidae) to a diet supplemented with aromatic amino acids. J. Insect Physiol. 28:1021–1024.

Wilson ACC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, Thomas GH, Douglas AE. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. Insect Mol. Biol. 19 Suppl 2:249–58.

Wolschin F, Hölldobler B, Gross R, Zientz E. 2004. Replication of the endosymbiotic bacterium *Blochmannia floridanus* is correlated with the developmental and reproductive stages of its ant host. Appl. Environ. Microbiol. 70:4096–102.

Worning P, Jensen LJ, Hallin PF, Staerfeldt H-H, Ussery DW. 2006. Origin of replication in circular prokaryotic chromosomes. Environ. Microbiol. 8:353–61.

Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. 4:e188.

Wu T, McCandlish AC, Gronenberg LS, Chng S-S, Silhavy TJ, Kahne D. 2006. Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 103:11754–9.

Yule KM, Miller TEX, Rudgers JA. 2013. Costs, benefits, and loss of vertically transmitted symbionts affect host population dynamics. Oikos. 122:1512–1520.

Zientz E, Dandekar T, Gross R. 2004. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. Microbiol. Mol. Biol. Rev. 68:745–70.

Zug R, Hammerstein P. 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS One 7:e38544.