Housekeeping genes essential for pantothenate biosynthesis are plasmid-encoded in *Rhizobium etli* and *Rhizobium leguminosarum*

Tomás Villaseñor¹, Susana Brom¹, Araceli Dávalos¹, Luis Lozano², David Romero¹ and Alejandro García-de los Santos¹*

**Abstract**

**Background:** A traditional concept in bacterial genetics states that housekeeping genes, those involved in basic metabolic functions needed for maintenance of the cell, are encoded in the chromosome, whereas genes required for dealing with challenging environmental conditions are located in plasmids. Exceptions to this rule have emerged from genomic sequence data of bacteria with multipartite genomes. The genome sequence of *R. etli* CFN42 predicts the presence of *panC* and *panB* genes clustered together on the 642 kb plasmid p42f and a second copy of *panB* on plasmid p42e. They encode putative pantothenate biosynthesis enzymes (pantoate-β-alanine ligase and 3-methyl-2-oxobutanoate hydroxymethyltransferase, respectively). Due to their ubiquitous distribution and relevance in the central metabolism of the cell, these genes are considered part of the core genome; thus, their occurrence in a plasmid is noteworthy. In this study we investigate the contribution of these genes to pantothenate biosynthesis, examine whether their presence in plasmids is a prevalent characteristic of the *Rhizobiales* with multipartite genomes, and assess the possibility that the *panCB* genes may have reached plasmids by horizontal gene transfer.

**Results:** Analysis of mutants confirmed that the *panC* and *panB* genes located on plasmid p42f are indispensable for the synthesis of pantothenate. A screening of the location of *panCB* genes among members of the *Rhizobiales* showed that only *R. etli* and *R. leguminosarum* strains carry *panCB* genes in plasmids. The *panCB* phylogeny attested a common origin for chromosomal and plasmid-borne *panCB* sequences, suggesting that the *R. etli* and *R. leguminosarum* *panCB* genes are orthologs rather than xenologs. The *panCB* genes could not totally restore the ability of a strain cured of plasmid p42f to grow in minimal medium.

**Conclusions:** This study shows experimental evidence that core *panCB* genes located in plasmids of *R. etli* and *R. leguminosarum* are indispensable for the synthesis of pantothenate. The unusual presence of *panCB* genes in plasmids of *Rhizobiales* may be due to an intragenomic transfer from chromosome to plasmid. Plasmid p42f encodes other functions required for growth in minimal medium. Our results support the hypothesis of cooperation among different replicons for basic cellular functions in multipartite rhizobia genomes.

**Background**

Multipartite genomes are common among members of the α-proteobacteria [1]. Most symbiotic nitrogen-fixing bacteria belonging to the genera *Rhizobium*, *Sinorhizobium*, *Mesorhizobium* and *Bradyrhizobium* possess multipartite genomes organized as a single circular chromosome and a variable number of large plasmids [2]. In some species plasmids can represent, in terms of size, up to 40% of the total genome. In *Rhizobium* and *Sinorhizobium* species one plasmid (pSym) concentrates most of the genes required for nodulation and nitrogen fixation [3]. The complete genome sequences of different rhizobia have revealed that plasmids harbor mainly accessory genes and that most encode predicted transport systems and a variety of catabolic pathways that may contribute to the adaptation of rhizobia to the heterogeneous soil and nodule environments [2,4]. These genes are absent from closely related genomes, lack synteny and their G+C composition differs from that of the core genes. The core genes are mainly...

* Correspondence: alex@ccg.unam.mx
Full list of author information is available at the end of the article
located on chromosomes, have essential functions in cell maintenance and have orthologs in related species [5,6]. In spite of this evidently biased distribution of core genes in the chromosome and accessory genes in plasmids, it is important to highlight the fact that there are interesting exceptions to this genomic rule: several typical core genes have been found encoded on rhizobia plasmids. Some are copies of genes located on chromosomes, with redundant functions that are totally dispensable for normal growth. Examples of these genes are the multiple copies of chaperonin-encoding genes groEL/groES [7,8], two tpiA genes encoding putative triose phosphate isomerase, a key enzyme of central carbon metabolism [4,6,9], and two putative S. meliloti asparagine synthetases (asnB and asnO), which may have a role in asparagine synthesis from aspartate by ATP-dependent amidation [10]. In contrast to these reiterated genes, a few single copy core genes have also been localized in plasmids. The tRNA specific for the second most frequently used arginine codon, CCG, is located on pSymB in S. meliloti [10]. Since this gene lies within a region of pSymB that could not be deleted [11], it is assumed to be essential for cell viability. The single copy of the minCDE genes, conceivably involved in proper cell division, have also been found in plasmids of S. meliloti, R. leguminosarum and R. etli [4,6,10]. Studies in S. meliloti have demonstrated that even though these genes are expressed in free-living cells and within nodules they are nonessential for cell division, since their deletion did not produce the small chromosomless minicells observed in E. coli and Bacillus subtilis [12].

A recent bioinformatic study revealed that approximately ten percent of the 897 complete bacterial genomes available in 2009 carry some core genes on extrachromosomal replicons [13]. However, very few of these genes have been functionally characterized and so their real contribution to bacterial metabolism is still an open question.

The complete genome sequence of R. etli CFN42 predicts that two putative “housekeeping” genes, panC and panB, which may be involved in pantothenate biosynthesis, are clustered together on plasmid p42f. Pantothenate is an essential precursor of coenzyme A (CoA), a key molecule in many metabolic reactions including the synthesis of phospholipids, synthesis and degradation of fatty acids, and the operation of the tricarboxylic acid cycle [14]. The R. etli panC gene is predicted to encode the sole pantoate-β-alanine ligase (PBAL), also known as pantothenate synthetase (PS) (EC 6.3.2.1), present in the R. etli genome. The function of this enzyme is the ATP-dependent condensation of D-pantoate with β-alanine to form pantothenate, the last step of the pantothenate biosynthesis pathway. The panB gene encodes the putative 3-methyl-2-oxobutanoate hydroxymethyltransferase (MOHMT) (EC 2.1.2.11), also known as ketopantoate hydroxymethyltransferase (KPHMT), the first enzyme of the pathway, responsible for the formation of α-ketopantoate by the transfer of a methyl group from 5,10-methylenetetrahydrofolate to alpha-ketoisovalerate. The complete genome sequence of R. etli CFN42 predicts that a second putative MOHMT enzyme (RHE_PE00443), similar to the product of panB, is encoded on plasmid p42e.

In this work we describe the isolation and use of panC and panB mutants to analyze the involvement of these plasmid-encoded genes in pantothenate biosynthesis. A survey of the localization of panCB genes among members of the Rhizobiales with multipartite genomes allowed us to infer a panCB phylogeny and to establish the probable chromosomal origin of these plasmid-borne genes. We also report that the panCB genes could not totally restore the growth in minimal medium (MM) of a strain cured of plasmid p42f, suggesting that other functions essential for growth in MM are encoded in this plasmid.

**Results**

**Functional characterization of plasmid p42f encoded panCB genes**

The predicted function of the product of panC (RHE_PF00001) annotated as PBAL, is the catalysis of the last step of pantothenate synthesis. This PBAL (298 amino acids) showed 43% identity and 62% similarity over 279 amino acids with the functionally characterized PBAL of E. coli K12 (284 amino acids). A search for conserved domains (CD-search) at NCBI-CDD revealed the presence of a typical pantoate-binding site. The panB gene (RHE_PF00002) is located immediately downstream of panC. The four nucleotide overlap between the panC TGA codon and panB ATG codon suggest that these genes might be transcribed as an operon. The panB gene encodes a putative MOHMT, the first enzyme of the pantothenate pathway. A BlastP comparison between the functionally characterized MOHMT of E. coli K12 (264 amino acids) and the putative MOHMT encoded on plasmid p42f of R. etli CFN42 (273 amino acids) showed 37% identity and 56% similarity over a length of 240 amino acids. A CD-search indicated that in the putative MOHMT of R. etli CFN42 the magnesium binding and active site domains are conserved. Additionally, Paralog Search (KEGG SSDB) and pathway tools predicted a second probable MOHMT, encoded on plasmid p42e (locus tag RHE_PE00443). Both proteins are similar in length (273 and 270 aa for the products encoded by panB and RHE_PE00443, respectively). However, a BlastP comparison of these sequences showed only 36% identity and 56% similarity over a tract of 140 amino acids. A CD-search revealed that only 5 of 12 of the invariable residues present in the active site domain are conserved in RHE_PE00443. The metal binding domain could not be detected by the
CD-search. To determine whether the \textit{panC} and \textit{panB} genes located on plasmid p42f are required for pantothenate synthesis, mutations in these genes were generated by site-directed vector integration mutagenesis via a single cross-over recombination (see details in Material and Methods and Table 1). Mutants ReTV1 (\textit{panC} \textsuperscript{−}) and ReTV2 (\textit{panB} \textsuperscript{−}) were unable to grow in minimal medium (MM) lacking calcium pantothenate (Figure 1a). Supplementation of MM with 1 \(\mu\)M calcium pantothenate allowed the \textit{panC} and \textit{panB} mutants to recover their wild-type growth rate (Figure 1b). The pantothenate auxotrophy displayed by the \textit{panB} mutant ReTV2 allowed us to discard a functional role of the putative MOHMT encoded by RHE\_PE00443 in pantothenate biosynthesis. Moreover, a pBBRMC3 clone constitutively expressing RHE\_PE00443 (pTV7) was unable to complement the pantothenate auxotrophy of the \textit{panB} mutant (data not shown).

Plasmid pTV4, harboring the \textit{panC} and \textit{panB} genes, as well as plasmids pTV5 and pTV6, carrying only \textit{panC} or \textit{panB} respectively, were introduced into mutant strains ReTV1 and ReTV2 and the growth phenotype of each construction was evaluated in MM. The \textit{panC} mutant ReTV1 complemented with the \textit{panCB} genes

| Table 1 Bacterial strains and plasmid. |
|--------------------------------------|
| **Strain or plasmid** | **Relevant genotype** | **Reference or source** |
|-----------------------|----------------------|-------------------------|
| \textbf{Rhizobium etli} |
| CFN42 | Wild type; Na\(^{1}\) | [6] |
| ReTV1 | CFN42 \textit{panC}\textsuperscript{−}pTV1; Km\(^{1}\) | This study |
| ReTV1-4 | CFN42 \textit{panC}\textsuperscript{−}pTV1 complemented with pTV4; Tcr Km\(^{1}\) | This study |
| ReTV1-5 | CFN42 \textit{panC}\textsuperscript{−}pTV1 complemented with pTV5; Tcr Km\(^{1}\) | This study |
| ReTV2 | CFN42 \textit{panB}\textsuperscript{−}pTV2; Km\(^{1}\) | This study |
| ReTV2-4 | CFN42 \textit{panB}\textsuperscript{−}pTV2 complemented with pTV4; Tc\(^{1}\) Km\(^{1}\) | This study |
| ReTV2-6 | CFN42 \textit{panB}\textsuperscript{−}pTV2 complemented with pTV6; Tc\(^{1}\) Km\(^{1}\) | This study |
| ReTV2-7 | CFN42 \textit{panB}\textsuperscript{−}pTV2 complemented with pTV7; Tc\(^{1}\) Km\(^{1}\) | This study |
| ReTV3 | CFN42 \textit{argE}\textsuperscript{−}pTV3; Km\(^{1}\) | This study |
| CFNX186 | CFN42 cured of plasmid p42f; Na\(^{1}\) | [18] |
| CFNX186-4 | CFNX186 cured of p42f; pTV4; Km\(^{1}\) | This study |
| CFNX186-24 | CFNX186 cured of p42f; pCos24; Tc\(^{1}\) | [30] |
| CIAT 652 | Wild type; Na\(^{1}\) | [38] |
| CIAT 894 | Wild type; Na\(^{1}\) | [38] |
| Kim5 | Wild type; Na\(^{1}\) | J. Handelsman, University of Wisconsin, MD |
| IE4771 | Wild type; Na\(^{1}\) | [15] |
| \textbf{Escherichia coli} |
| DH\textit{5}a | Host for recombinant plasmids; Na\(^{1}\) | Stratagene |
| S17-1 | C600-RP4-2(Tc::Mu) (Km::Tn7) | [39] |
| | Donor for conjugation |
| \textbf{Plasmids} |
| pBC | pBluescript II SK(+) phagemid vector; Cm\(^{1}\) | Stratagene. |
| pk18mob | pk18, derivative mob; Km\(^{1}\) | [29] |
| pRK7813 | Broad-host-range cosmid vector, Mob, IncP, Tc\(^{1}\) | [40] |
| pBRRMCS3 | Broad-host-range cloning vector, Mob; Tc\(^{1}\) | [41] |
| pBC1 | pBC harboring a 400-bp \textit{BamHI}-XbaI PCR fragment of \textit{panC}; Cm\(^{1}\) | This study |
| pBC2 | pBC harboring a 400-bp \textit{BamHI}-XbaI PCR fragment of \textit{panB}; Cm\(^{1}\) | This study |
| pTV1 | pK18mob harboring a 400-bp \textit{KpnI}-XbaI PCR fragment of \textit{panC}; Km\(^{1}\) | This study |
| pTV2 | pK18mob harboring a 400-bp \textit{KpnI}-XbaI PCR fragment of \textit{panB}; Km\(^{1}\) | This study |
| pTV3 | pK18mob harboring a 400-bp \textit{KpnI}-XbaI PCR fragment of \textit{argE}; Km\(^{1}\) | This study |
| pTV4 | pRK7813 harboring a 3.1 kb \textit{EcoRI} fragment of pCos24 containing \textit{panC} and \textit{panB}; Tc\(^{1}\) | This study |
| pTV5 | pBRRMCS3 harboring a 1.2 kb \textit{KpnI}-XbaI PCR fragment containing \textit{panC}; Tc\(^{1}\) | This study |
| pTV6 | pBRRMCS3 harboring a 1 kb \textit{KpnI}-XbaI PCR fragment containing \textit{panB}; Tc\(^{1}\) | This study |
| pTV7 | pBRRMCS53 harboring a 1 kb \textit{KpnI}-XbaI PCR fragment containing RHE\_PE00443; Tc\(^{1}\) | This study |
| pcos24 | 20 Kb EcoRI fragment of plasmid p42f cloned in pLAFR1 containing \textit{panC}, \textit{panB}, oxyR and \textit{katG}; Tc\(^{1}\) | [30] |
ReTV1-4) recovered wild type growth in MM. In contrast, when complemented only with panC (strain ReTV1-5) no growth occurred in the absence of pantothenate. These results strongly suggest that the panCB genes form a single transcriptional unit. As expected, wild type growth of panB mutant ReTV2 was recovered by complementation with the panCB genes or with the panB gene (strains ReTV2-4 and ReTV2-6 respectively).

The occurrence of panCB genes in plasmids is highly conserved among R. etli and R. leguminosarum strains but not in other members of the Rhizobiales with multipartite genomes

To investigate whether the presence of the panCB genes in plasmids is a common characteristic of the Rhizobiales we examined the location of panCB genes in 22 members of the Rhizobiales having fully sequenced multipartite genomes (Table 2). To date, the genomes of seven R. etli strains, in addition to CFN42, have been totally sequenced [15]. However, with the exception of strain CIAT 652, the genomes were released as draft assemblies, precluding panCB localization. We experimentally determined the localization of panCB genes in the genome of four of these R. etli strains (CIAT 894, Kim5, 8C-3, and IE4771) by hybridization of their plasmid profiles with [32P]dCTP-labelled panC and panB genes from CFN42 under high stringency conditions. Both probes produced intense hybridization signals on the same plasmid of each strain, indicating that the panCB genes are also plasmid-borne in these R. etli strains (Table 2). Coincidentally, in the three R. leguminosarum strains with fully sequenced genomes reported in the NCBI database, the panCB genes are assigned to plasmids. In contrast, in other species of Rhizobiales with multipartite genomes, the panCB genes are always confined to the chromosome, or to chromosome I in those species harboring two chromosomes, with exception of Agrobacterium tumefaciens C58 which carries panCB on the linear chromosome II and Methylobacterium nodulans ORS2060 that carries panC on their single chromosome and panB on plasmid pMNOD02 (Table 2).

Phylogenetic analysis of rhizobial panCB genes indicates a common origin of chromosomal and plasmid-borne sequences

Two possible hypotheses were considered to explain the presence of panCB genes in plasmids of R. etli and R. leguminosarum strains: (1) an intragenomic rearrangement of panCB genes from chromosome to plasmid, which must have occurred in the last common ancestor of both species; (2) by xenologous gene displacement, that is, a horizontal transfer event in which a gene is displaced by a horizontally transferred ortholog acquired from another lineage [16]. In the latter hypothesis we assume that the presence of these xenolog genes in plasmids conferred a selective advantage that may have eventually led to the loss of the chromosome-located panCB genes. To test these hypotheses the phylogeny of 16 rhizobial species inferred from ten orthologous single copy housekeeping genes (fusA, guaA, ileS, infB, recA, rplB, rpoB, rpoC, secY and valS) located on primary chromosomes, was compared with the phylogeny of the same rhizobial species inferred from the panCB genes located on plasmids and chromosomes. The rationale for this comparison was that if the plasmid-borne
panCB phylogeny agrees with the current phylogeny of the Rhizobiales, inferred from the housekeeping genes, it would support the hypothesis of intragenomic transfer of the panCB genes. On the other hand, if both phylogenies are incongruent, it would favor the hypothesis of horizontal transfer of the panCB genes. Concatenated nucleic acids multiple alignments were used to infer both phylogenies with the maximum likelihood method described in materials and methods. The resulting phylogenetic trees are shown in Figure 2. The housekeeping genes inferred tree (Figure 2a) was consistent with the recently reported phylogeny of 19 Rhizobiales performed on a data set of 507 homologous proteins from the primary chromosome [17]. Both trees are in close agreement with the phylogeny inferred from the panCB genes (Figure 2b). Thus the phylogeny of R. etli and R. leguminosarum inferred from plasmid-encoded panCB genes is consistent with the phylogeny deduced from their housekeeping genes supporting the hypothesis of a chromosomal origin for the plasmid-encoded panCB genes.

The panCB genes do not fully complement the growth deficiency of a R. etli CFN42 p42f cured derivative in MM

It was reported previously that R. etli CFNX186, a p42f-cured derivative of R. etli CFN42, is unable to grow in MM [18]. To assess if the growth deficiency of strain CFNX186 in MM was due to the absence of the panC and panB genes, plasmid pTV4 (panCB) was introduced into strain CFNX186. The growth of the transconjugant (CFNX186-4) after 15 hours of culture in MM was only 50% that of the WT strain grown under the same conditions (Figure 3a). The growth of CFNX186-4 did not improve even after 72 h in culture (data not shown).

Table 2 Localization of the panCB genes in representative members of the Rhizobiales with multipartite genomes

| Strain                        | Genome number | Structure of Plasmids | Localization of panC | Localization of panB |
|-------------------------------|---------------|-----------------------|----------------------|----------------------|
| Brucella abortus bv. 1 str. 9-941 | 2             | 0                     | ChrI                 | ChrI                 |
| B. melitensis 16M             | 2             | 0                     | ChrII                | ChrII                |
| B. ovis ATCC 25840            | 2             | 0                     | Chr                  | Chr                  |
| Sinorhizobium meliloti 1021   | 1             | 2                     | Chr                  | Chr                  |
| S. medicae WSM419             | 1             | 3                     | Chr                  | Chr                  |
| Ochrobactrum anthropi ATCC 49188 | 2             | 4                     | ChrI                 | ChrI                 |
| Agrobacterium radiobacter K84 | 2             | 3                     | Chr                  | Chr                  |
| A. vitis S4                   | 2             | 5                     | ChrI                 | ChrI                 |
| A. tumefaciens CS8             | 2             | 2                     | ChrII                | ChrII                |
| Rhizobium etli CFN42          | 1             | 6                     | p42f                 | p42f                 |
| R. etli C1652                 | 1             | 3                     | pci                  | pdi                  |
| R. etli C1654                 | 1             | 4                     | pci                  | pdi                  |
| R. etli BE4771                | 1             | 4                     | pd                   | pd                   |
| R. etli BE3-3                 | 1             | 3                     | pci                  | pdi                  |
| R. leguminosarum bv. viciae 3841 | 1             | 6                     | pRL12                | pRL12                |
| R. leguminosarum WSM1325      | 1             | 5                     | pRL132501            | pRL132501            |
| R. leguminosarum WSM2304      | 1             | 4                     | pRLG201              | pRLG201              |
| Rhizobium sp. NGR234          | 1             | 2                     | Chr                  | Chr                  |
| Mesorhizobium loti MAFF303099 | 1             | 2                     | Chr                  | Chr                  |
| M. sp. BNC1                   | 1             | 3                     | Chr                  | Chr                  |
| Methylbacterium extorquens AM1 | 1             | 4                     | Chr                  | Chr                  |
| M. radiotolerans JCM2831      | 1             | 8                     | Chr                  | Chr                  |
| M. nodulans ORS2060           | 1             | 7                     | Chr                  | pMNOD2               |
| Bradyrhizobium sp. BTAI1      | 1             | 1                     | Chr                  | Chr                  |
| Nitrobacter hamburgensis X14  | 1             | 3                     | Chr                  | Chr                  |
| Xanthobacter autotrophicus Hy2 | 1             | 1                     | Chr                  | Chr                  |

Abbreviations are as follows: Chr, chromosome of those Rhizobiales with one chromosome; Chr I and Chr II, chromosome I and chromosome II respectively in those Rhizobiales harboring two chromosomes; p, plasmid. *Rhizobium species in which localization of panCB genes was done by Southern blot hybridization of plasmid profiles. †Plasmids with very similar electrophoretic mobility gave as result ambiguous plasmid localization of panC and panB homologous sequences.
with 1 μM calcium pantothenate (Figure 3b). This shows that the growth deficiency of CFNX186 is only partly due to the absence of the panCB genes and indicates that other functions encoded in plasmid p42f are required for growth in MM.

Previous studies have demonstrated that the katG gene, which encodes the sole catalase-peroxidase expressed in free-living growth conditions, is located on plasmid p42f of R. etli CFN42. These studies also revealed that the growth rate of a katG mutant in MM was significantly reduced in comparison with that of the wild-type parental strain [19]. On plasmid p42f, katG, as well as its putative transcriptional regulator protein encoded by oxyR, are located 80 bp downstream of the panCB genes. We speculated that introduction of the panCB genes together with the katG and oxyR genes might improve the growth of CFNX186 in MM. To test this hypothesis, we used pCos24, which contains a 20 kb EcoRI fragment of p42f carrying panC, panB, oxyR and katG (see Material and Methods). pCos24 was introduced into CFNX186 and the resulting transconjugant (CFNX186-24) grown in MM. Figure 3 shows that after 15 hours of culture there was no significant difference between the growth rate of CFNX186 complemented only with panCB (CFNX186-4), and CFNX186 complemented with cosmid pCos24 (CFNX186-24). Furthermore, the

![Figure 2 Comparison of phylogenetic trees constructed from core and panCB genes](image)

**Figure 2** Comparison of phylogenetic trees constructed from core and panCB genes. Maximum-likelihood phylogenetic trees of 16 Rhizobiales constructed using the concatenated nucleic acid sequences of 10 housekeeping genes (a) or panC and panB concatenated genes (b). Bootstrap values are shown over each branch (based on 100 pseudo-replicates).

![Figure 3 panCB genes do not fully restore the growth deficiency of CFNX186](image)

**Figure 3** panCB genes do not fully restore the growth deficiency of CFNX186. Growth of R. etli CFN42 wild-type strain, its p42f-cured derivative CFNX186, CFNX186 complemented with the panCB genes (CFNX186-4) and CFNX186 complemented with a 20 kb EcoRI fragment of plasmid p42f containing the panC, panB, oxyR and katG genes (CFNX186-24) in: (a) minimal medium, (b) minimal medium supplemented with 1 μM pantothenate. Growth curves are the mean of at least three independent experiments; error bars represent standard deviations.
growth of CFNX186-24 did not increase even after 72 h of culture (data not shown) indicating that katG and oxyR did not improve the growth rate of \textit{panCB} complemented CFNX186 in MM. We also tested the possibility that arginine might improve the growth of strain CFNX186-24 due to the presence of a putative \textit{N}-acetylornithinase (EC 3.5.1.16) encoded in the plasmid p42f. In the \textit{Enterobactericeae} this enzyme catalyzes the conversion of \textit{N}-acetylornithine to ornithine, a key step in the arginine biosynthesis pathway [20]. However, the growth deficiency of strain CFNX186-24 in MM was not corrected by the addition of 1, 5, 10 or 15 mM arginine (data not shown). Furthemore, we constructed an \textit{argE} mutant strain (ReTV3, Table 1) that was able to grow in MM without exogenous arginine at the same rate as parental strain CFN42 (data not shown), confirming that this gene is not essential for arginine synthesis.

**Discussion**

Seminal studies on the phenotypic characterisation of plasmid-cured strains of \textit{R. leguminosarum} and \textit{R. etli} revealed that the absence of several plasmids cause a growth deficiency in rich and minimal medium [18,21]. These findings suggested that undefined metabolic traits are present on rhizobial plasmids. The bioinformatic analysis of 897 bacterial genomes performed by Harrison et al [13] revealed the presence of extrachromosomal core genes in 82 genomes mainly belonging to the Proteobacteria. In contrast with these \textit{in silico} data, there is little experimental information on the contribution of these core genes to bacterial metabolism or cellular process. The few genes that have been functionally characterized encode redundant functions and are totally dispensable for the cell [7-9,12]. Our study provides experimental evidence that the enzymes MOHMT (EC 2.1.2.11) and PBAL (EC 6.3.2.1) encoded on plasmid p42f are indispensable for the synthesis of pantothenate. Moreover, our results showed that the cluster of \textit{panCB}, \textit{katG} and \textit{oxyR} genes was insufficient to restore full growth capacity to the p42f cured derivative CFNX186, implying that in addition to pantothenate synthesis, there are more functions encoded on plasmid p42f required for growth in MM. Obvious candidates for these functions could not be identified \textit{a priori} among the 567 proteins encoded in p42f even though their predicted functions were recently updated with KAAS (KEGG Automatic Annotation Server and Pathway Reconstruction Server). We discarded arginine limitation as the cause for the growth deficiency of strain CFNX186-24. The arginine prototrophy displayed by a mutation in the p42f encoded \textit{argE} suggests that in \textit{R. etli} the conversion of \textit{N}-acetylornithine to ornithine is catalyzed by the chromosome-encoded ArgJ, an ornithine acetyltransferase (OATase, EC 2.3.1.35), which transfers the acetyl group of \textit{N}-acetylornithine to glutamate to produce ornithine and \textit{N}-acetylglycinate. Functional OATases have been found in the majority of bacteria [20].

Also, we have demonstrated that plasmid-localization of \textit{panCB} in \textit{R. etli} CFN42 is not unique to this strain. A screening of the location of \textit{panCB} genes among members of the \textit{Rhizobiales}, showed that the occurrence of these genes in plasmids is a highly conserved trait among \textit{R. etli} and \textit{R. leguminosarum} strains. Furthermore, the synteny of the \textit{panCB}, \textit{oxyR}, \textit{katG} genes in \textit{R. etli} CFN42 is conserved in \textit{R. etli} CIAT652 and in \textit{R. leguminosarum} strains 3841, WSM1325 and WSM2304. In contrast, genomes of \textit{Rhizobium sp.}, \textit{Sinorhizobium}, \textit{Bradyrhizobium} and \textit{Mesorhizobium} species carried chromosomal \textit{panCB} genes. Only in \textit{A. tumefaciens} C58 the \textit{panCB} genes are localized in the linear chromosome, whereas in all other \textit{Rhizobiales} harboring secondary chromosomes the \textit{panCB} genes were located in chromosome I. A bioinformatic analysis with MicrobesOnline operon predictions [22] indicates that \textit{panCB} genes are organized as possible operons in most of the \textit{Rhizobiales} examined in this work: all these predicted operons conserve the four nucleotide overlap between the \textit{panC} TGA codon and the \textit{panB} ATG codon observed in \textit{R. etli} CFN42 (data not shown). In the genomes of \textit{Bradyrhizobium sp.} BTAi1, \textit{Nitrobacter hamburgensis} X14, \textit{Methylobacterium extorquens} AM1, \textit{Methylobacterium radiotolerans} JCM2831 and \textit{Xanthobacter autotrophicus} Ry2, \textit{panC} and \textit{panB} are encoded in separate chromosomal loci, whereas in \textit{Methylobacterium nodulans} ORS2060 \textit{panC} is located in the chromosome and \textit{panB} in plasmid pMNOD02.

The \textit{Rhizobiales} phylogeny inferred from concatenated \textit{panC} and \textit{panB} genes was consistent with the phylogeny deduced from 10 concatenated housekeeping genes. The low bootstrap values obtained for some nodes of the \textit{panCB} phylogeny might be due to the small number of informative characters in the alignments of only two genes (1 977 nucleotides). This is consistent with previous reports that state that trees from longer alignments obtained by the concatenation of genes encoding multiple-protein families have higher bootstrap support than trees inferred from genes encoding single proteins [23]. The phylogenetic relationships among \textit{Rhizobium} species carrying \textit{panCB} genes in plasmids with their closest relatives, \textit{Agrobacterium} and \textit{Sinorhizobium} species, harboring \textit{panCB} genes in the chromosome was also observed in neighbor-joining trees inferred from single \textit{panC} and \textit{panB} genes (data not shown). These data agree with the hypothesis that plasmid-encoded \textit{panCB} genes are orthologs of the \textit{panCB} genes located in chromosome. From these results, we propose that the presence of the \textit{panCB} genes in plasmids in \textit{R. etli} and \textit{R. leguminosarum} species may be due to an intragenomic transfer event from chromosome to plasmid. The mechanism leading to the
transfer of core genes from chromosome to plasmids could involve cointegration and excision events between the replicons, similar to rearrangements that have been visualized in *S. meliloti* [24]. The translocation of genes from chromosome to plasmids may be part of the complex evolution of multipartite genomes. A study based on the analysis of clusters of syntenic genes shared among plasmids and secondary chromosomes of bacteria with multipartite genomes suggested that secondary chromosomes may have originated from an ancestral plasmid to which genes had been transferred from a primary chromosome [17].

Our pioneering work on plasmid-encoded functions in *R. etli* CFN42 established that a functional relationship among different replicons is required for symbiotic and free-living functions [18,25]. More recently, a functional connectivity among most of the proteins encoded in the replicons of *R. etli* CFN42 was predicted in silico [6]. Our results demonstrated that the putative MOHMT encoded by RHE_PE00443 is not functional under the conditions studied and provides evidence of functional cooperation between p42f and chromosomally encoded proteins for pantothenate biosynthesis.

**Conclusions**

Our study shows that the presence of the core *panCB* genes in a plasmid is a characteristic conserved in *R. etli* and *R. leguminosarum* strains but not in other *Rhizobiales*. The phylogenetic approach used in this study suggests that the unusual presence of *panCB* in plasmids may be due to an intragenomic transfer event from chromosome to plasmid rather than a xenologous gene displacement. Using *R. etli* CFN42 as a model, we showed that the plasmid-encoded core *panCB* genes were indispensable for the synthesis of pantothenate. The *panCB* genes could not totally restore growth of a strain cured of plasmid p42f in minimal medium, suggesting that other functions essential for growth in this medium are encoded in this plasmid. Our results support the hypothesis of functional cooperation among different replicons for basic cellular functions in multipartite rhizobial genomes.

**DNA manipulations**

Standard techniques described by Sambrook *et al.* [28] were used for plasmid and total DNA isolation, restriction, cloning, transformations, and agarose gel electrophoresis. Plasmid mobilization from *E. coli* to *Rhizobium* was done by conjugation performed on PY plates at 30°C by using overnight cultures grown to stationary phase. Donors (*E. coli* strain S17-1) and recipients (*R. etli* CFN42 wild type and mutant strains) were mixed at a 1:2 ratio, and suitable markers were used for transconjugant selection.

**Mutagenesis of the *panC* and *panB* genes and genetic complementation of mutant strains**

Mutants were generated by site-directed vector integration mutagenesis. Internal 400 bp DNA fragments of *panC* and *panB* were amplified by PCR with primers A and B; C and D, respectively (Table 3). PCR fragments of *panC* and *panB* were cloned in vector pBC as 400 bp BamHI-XbaI fragments, generating pBC1 and pBC2 respectively, and then subcloned as KpnI-XbaI fragments into suicide vector pK18mob [29] to form plasmids pTV1 and pTV2, respectively. These plasmids were mobilized into *R. etli* CFN42 by conjugation and single crossover recombinants selected on PY plates containing Km and Nal. The disruption of the *panC* and *panB* genes was confirmed by Southern blot analysis using a 400-bp PCR internal fragment of each gene as a probe (data not shown). The resultant mutants were named ReTV1 and ReTV2 respectively. To complement the phenotype of...
Table 3 Primers used in this work.

| Primer | Sequence (5'-3') |
|--------|-----------------|
| A      | GGGGATCCGGAGAGCGCTCCAGAAATACCCGC |
| B      | CGGAGATCCCGCGCAAGGACGACCAGGGCCGCGCC |
| C      | CGGGTCTGAGATCCGGAGATGGCTTCTGCTGGAAGGCG |
| D      | GACGTCTAGAACTGGAAGTGGAGAGGCCC |
| E      | TGTAGGTTACCTTGGAGGCGGTTGGAATACGGG |
| F      | GGGGCTCTGAGATTCCGGCTGCTCTTTCG |
| G      | GGGGATCCGGAGAGCGCTCCAGAAATACCCGC |
| H      | AAGGTCTAGAGCGGCTGATCCGGGGGTAAGGAGTG |
| I      | CGTCAAGGCCATCCCTGCTGGCAGCG |
| J      | CCCCCCTCTAGACGCTGGGGGAGAAGGAGT |
| K      | GCTGTGGTACCAGCGGCTTCCGGCACCCTCGG |
| L      | ACCCTCTAGATGCGGACCTGGAGAGG |

The restriction sites are indicated in bold.

the panC and panB mutants, plasmids pTV4, pTV5, pTV6 and pTV7 were constructed as follows: a 3.1 kb EcoRI fragment from cosmid vector pCos24, isolated from a genomic library of R. etli CFN42 [30] and containing the panC and panB genes, was subcloned in broad-host-range vector pTV7813, generating plasmid pTV4. To construct plasmid pTV5, a 1.2 kb fragment containing only panC (894 bp) was amplified by PCR with primers E and F and cloned in the KpnI-XbaI sites in the broad-host-range vector pBBRMSC3 so that the gene would be constitutively expressed from the vector’s lacZ promoter. Primers G and H (Table 3) were used to amplify a 1 kb PCR fragment containing only the panB gene (822 bp). This DNA fragment was cloned in plasmid pBBRMSC3 in the KpnI-XbaI restriction sites, generating plasmid pTV6. Plasmid pTV7 contains the second panB gene (RHE_PEO00443), encoded on R. etli plasmid p42e, amplified with primers I and J. The resultant 1 kb PCR fragment was cloned in the KpnI-XbaI sites of plasmid pBBRMSC3. To complement the growth deficiency of strain CFNX186, a derivative of R. etli CFN42 cured of plasmid p42f, plasmid pTV5 and cosmid vector pCos24 were introduced by conjugation. The complemented strains obtained were named CFNX186-4 and CFNX186-24 respectively. The argE gene was disrupted as described above. Briefly, an internal 400 bp PCR fragment of argE amplified with primers K and L was cloned directly in pK18mob using the KpnI and XbaI sites to give pTV3 (Table 1). This recombinant suicide plasmid was mobilized into R. etli CFN42 and the resultant mutant named RetV3 (Table 1).

Filter blots hybridization and plasmid visualization

For Southern-type hybridizations, genomic DNA was digested with appropriate restriction enzymes, electrophoresed in 1% (w/v) agarose gels, blotted onto nylon membranes, and hybridized under stringent conditions, as previously reported by [31], using Rapid-hyb buffer. To use the panC and panB genes as probes, both genes were amplified by PCR, separated on a 1% agarose and purified by a PCR purification kit (QIAquick). They were labeled with [α-32P]dCTP using a Rediprime DNA labeling system. Plasmid profiles were visualized by the Eckhardt technique as modified by [21], and hybridized in a similar manner.

Identification of orthologous proteins, multiple sequence alignments and phylogenetic analysis

All genomic sequences analyzed in this study were obtained from the Integrated Microbial Genomes System of the DOE Joint Genome Institute http://img.jgi.doe.gov/). We obtained protein and gene sequences of panB, panC and 10 chromosomal housekeeping genes (fusA, guaA, ileS, infB, recA, rplB, rpoB, rpoC, secY and valS) from 16 rhizobial species. Accession numbers for these sequences and the species list are shown in Table S1 (see Additional file 1). An orthologous data set for each gene was constructed using Blast [32] and the bidirectional best hit method applying the criteria reported by Poggio et al [33]. Multiple alignments of putative orthologous proteins were performed using the MUSCLE program [34] with default settings. After removing poorly conserved regions two concatenated protein alignments were obtained, one for the 10 chromosomal housekeeping genes (8469 amino acids) and the other for panB and panC (659 amino acids). Both concatenated protein multiple alignments were used to generate nucleic acids multiple alignments of their respective genes with the TranAlign program of the EMBOSS suit http://emboss.sourceforge.net/. Nucleic acids multiple alignments were used to obtain two phylogenies with the maximum likelihood method implemented in PHYML [35] with HKY as substitution model [36]. The phylogenetic reconstruction was carried out with a nonparametric bootstrap analysis of 100 replicates for each alignment. TreeDyn program [37] was used to visualize and edit both phylogenies.

Additional material

Additional file 1: Table S1 Rhizobial species list and accession numbers of housekeeping and panC8 genes used for phylogenetic analysis.

List of abbreviations

PBAL: pantoate-β-alanine ligase; MOHMT: 3-methyl-2-oxobutanoate hydroxymethyltransferase; MM: minimal medium.

Acknowledgements

We are grateful to Laura Cervantes and Javier Rivera for their excellent technical assistance. We acknowledge Michael F. Dunn for critically reviewing the manuscript. This work was supported by DGAPA-PAPIIT-UNAM grant IN200309-2. Tomás Villaseñor was supported by a Ph. D. scholarship.
characterization of genes involved in symbiosis. Microbiology 2007, 153:375-387.

13. Harrison PW, Lower RJ, Kim NDO, Young JPW. Introducing the bacterial "chromid": not a chromosome, not a plasmid. Trends Microbiol 2010, 18:141-148.

14. Jackowski S. Biosynthesis of pantothetic acid and coenzyme A. In Escherichia coli and Salmonella: cellular and molecular biology. Edited by: Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE. Washington, DC: ASM Press; 1996:1310-1324.

15. González V, Acosta JL, Santamaría RL, Bustos P, Fernández JL, Hernández IL, Díaz R, Flores M, Palacios R, Mora J, Dávila G. Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of Rhizobium etli. Appl Environ Microbiol 2010, 76:1604-1614.

16. Koozin EV, Makarova KS, Aravind L. Horizontal gene transfer in prokaryotes: quantification and classification. Ann Rev Microbiol 2003, 57:709-742.

17. Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW. Characterization of two plasmid-borne nitrofixing plasmids of R. etli strains and assisted in DNA cloning. LL carried out the phylogenetic analysis and the discussion of results. AGS conceived the study, supervised the experimental work and wrote the manuscript. All authors read and approved the final manuscript.

Received: 17 November 2010 Accepted: 5 April 2011 Published: 5 April 2011

References

1. Jumas-Bilak E, Michaux-Charachon S, Bourg G, Ramuz M, Allard-Sevent A. Unconventional genomic organization in the alpha subgroup of the Proteobacteria. J Bacteriol 1998, 180:2749-2753.

2. MacLean AM, Finan TM, Sadowsky MJ. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. Plant Physiol 2007, 144:515-522.

3. Romero D, Brom S. The symbiotic plasmids of the Rhizobiaceae. In Plasmid biology. Edited by: Phillips G, Funnell BE, Washington, D.C: American Society for Microbiology; 2004:271-290.

4. Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazouei ZF, Hull KH, Wexler M, Cunnon AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Parkhill J, González V: Characterization of two plasmid-borne nitrofixing plasmids of R. etli strains and assisted in DNA cloning. LL carried out the phylogenetic analysis and the discussion of results. AGS conceived the study, supervised the experimental work and wrote the manuscript. All authors read and approved the final manuscript.

Published: 5 April 2011

http://www.biomedcentral.com/1471-2180/11/66
32. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.

33. Poggio S, Abreu-Goodger C, Fabela S, Osorio A, Dreyfus G, Vinuesa P, Camarena L: A complete set of flagellar genes acquired by horizontal transfer coexists with the endogenous flagellar system in Rhodobacter sphaeroides. J Bacteriol 2007, 189:3208-3216.

34. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004, 5:113.

35. Guindon S, Gascuel O: Efficient biased estimation of evolutionary distances when substitution rates vary across sites. Mol Biol Evol 2002, 19:534-543.

36. Hasegawa M, Kishino H, Yano T: Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Biol 1985, 22:160-174.

37. Chevenet F, Brun C, Barulis AL, Jacq B, Chisten R: TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 2006, 7:439.

38. Piñero D, Martínez E, Selander RK: Genetic diversity and relationships among isolates of Rhizobium leguminosarum biovar phaseoli. Appl Environ Microbiol 1988, 54:2825-2832.

39. Simon R: High frequency mobilization of gram-negative bacterial replicons by the in vitro constructed Tn5-mob transposon. Mol Gen Genet 1984, 196:413-420.

40. Jones JDS, Gutteron N: An efficient mobilizable cosmid vector, pIK7813, and its use in a rapid method for marker exchange in Pseudomonas fluorescens strain HV37a. Gene 1987, 61:299-306.

41. Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM II, Peterson RM: Four new derivatives of the broad-host-range cloning vector pBR1MCS, carrying different antibiotic-resistance cassettes. Gene 1995, 160:175-176.

doi:10.1186/1471-2180-11-66
Cite this article as: Villaseñor et al.: Housekeeping genes essential for pantothenate biosynthesis are plasmid-encoded in Rhizobium etli and Rhizobium leguminosarum. BMC Microbiology 2011 11:66.