Molecular Ecology (2005) **14**, 4033–4050 doi: 10.1111/j.1365-294X.2005.02721.x

Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes

CLAUDIA SILVA,* PABLO VINUESA,* LUIS E. EGUIARTE,† VALERIA SOUZA† and ESPERANZA MARTÍNEZ-ROMERO*

**Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, AP 565A, Cuernavaca, Morelos, México, †Instituto de Ecología, Universidad Nacional Autónoma de México, AP 70-275, México D. F., México*

Abstract

We used phylogenetic and population genetics approaches to evaluate the importance of the evolutionary forces on shaping the genetic structure of *Rhizobium gallicum* and related species. We analysed 54 strains from several populations distributed in the Northern Hemisphere, using nucleotide sequences of three 'core' chromosomal genes (*rrs*, *glnII* and *atpD*) and two 'auxiliary' symbiotic genes (*nifH* and *nodB*) to elucidate the biogeographic history of the species and symbiotic ecotypes (biovarieties) within species. The analyses revealed that strains classified as *Rhizobium mongolense* and *Rhizobium yanglingense* belong to the chromosomal evolutionary lineage of *R. gallicum* and harbour symbiotic genes corresponding to a new biovar; we propose their reclassification as *R. gallicum* bv. *orientale*. The comparison of the chromosomal and symbiotic genes revealed evidence of lateral transfer of symbiotic information within and across species. Genetic differentiation analyses based on the chromosomal protein-coding genes revealed a biogeographic pattern with three main populations, whereas the 16S rDNA sequences did not resolve that biogeographic pattern. Both the phylogenetic and population genetic analyses showed evidence of recombination at the *rrs* locus. We discuss our results in the light of the contrasting views of bacterial species expressed by microbial taxonomist and evolutionary biologists.

Keywords: bacterial systematics, biogeography, ecotypes, migration, recombination, species concept

Received 23 April 2005; revision received 8 July 2005; accepted 25 July 2005

Introduction

Rhizobia are free-living heterotrophic bacteria that have the ability to establish symbiotic associations with leguminous plants. They therefore represent a good model to investigate diverse aspects of bacterial ecology and evolution (Young & Johnston 1989; Wernegreen & Riley 1999). In *Rhizobium* and *Sinorhizobium* species, most of the symbiosis-specific genes required for nodulation and nitrogen fixation are borne on large plasmids, called symbiotic plasmids (García de los Santos *et al*. 1996; Romero & Brom 2004). The suite of nodulation genes carried by a strain largely determines its host range and specificity (Spaink 2000). Within a rhizobial species, different symbiotic ecotypes (biovarieties) can exist due to the presence of alternative symbiotic plasmids encoding for different host specificities. The subspecific taxonomic rank of biovariety was created to handle the different symbiotic ecotypes found within species (Graham *et al*. 1991). In *Rhizobium gallicum* two biovars have been described: *gallicum* and *phaseoli*. The biovar *gallicum* has been reported only for *R. gallicum* strains isolated in Europe, America and North Africa, and it is characterized by a broad host range, nodulating legumes in the genera *Phaseolus*, *Leucaena*, *Macroptilium*, *Onobrychis*, *Sesbania*, *Caliandra*, *Gliricidia* and *Piptadenia* (Amarger *et al*. 1997; Laguerre *et al*. 1997; Sessitsch *et al*. 1997b; Herrera-Cervera *et al*. 1999; Mhamdi *et al*. 1999; Silva *et al*. 1999, 2003; Rodriguez-Navarro *et al*. 2000; Zurdo-Piñeiro *et al*. 2004). The biovar *phaseoli* strains have been exclusively isolated from *Phaseolus vulgaris* nodules and they are not able to establish nitrogen-fixing

Correspondence: Claudia Silva, Fax: 52-777-3175581; E-mail: csilva@ccg.unam.mx

© 2005 Blackwell Publishing Ltd

symbioses with *Leucaena*, *Macroptilium* and *Onobrychis* (Amarger *et al*. 1997; Silva *et al*. 2003). Other *Phaseolus* nodulating species of *Rhizobium* have strains ascribed to biovar *phaseoli*, namely *Rhizobium etli*, *Rhizobium leguminosarum* and *Rhizobium gardinii* (Amarger *et al*. 1997; Herrera-Cervera *et al*. 1999). The accepted biogeographic hypothesis for the origin of biovar *phaseoli* symbiotic plasmids is that they evolved in American *R. etli* strains associated with native *P. vulgaris* plants (Martínez-Romero & Caballero-Mellado 1996; Martínez-Romero 2003; Aguilar *et al*. 2004). It is postulated that *R. etli* bv. *phaseoli* was introduced to Europe and other parts of the world in historic times along with common bean seeds, providing the ecological opportunity for the biovar *phaseoli* symbiotic plasmids to be laterally transferred from *R. etli* to autochthonous *Rhizobium* species (Segovia *et al*. 1993; Martínez-Romero & Caballero-Mellado 1996; Amarger *et al*. 1997; Sessitsch *et al*. 1997a; Pérez-Ramírez *et al*. 1998; Mhamdi *et al*. 1999; Laguerre *et al*. 2001; Brom *et al*. 2002; Martínez-Romero 2003).

In addition to the lateral transfer of symbiotic information among rhizobia, population genetic analyses using multilocus enzyme electrophoresis (MLEE) data have demonstrated that chromosomal homologous recombination is frequent within rhizobial populations (Maynard Smith *et al*. 1993; Souza *et al*. 1994; Strain *et al*. 1995; Hagen & Hamrick 1996; Silva *et al*. 1999; Vinuesa *et al*. 2005a), but rare between sympatric species (Silva *et al*. 2003; Vinuesa *et al*. 2005c). Furthermore, like recombination, migration can have profound effects on the genetic structure of rhizobial populations by homogenizing gene pools that are distant, thus increasing the effective population size (Strain *et al*. 1995; Hagen & Hamrick 1996; Vinuesa & Silva 2004; Vinuesa *et al*. 2005c). Although reports of the same rhizobial species isolated from very distant places of the world are widespread in the literature, few studies have addressed the question of genetic differentiation among populations. Some studies have detected significant differences among geographically separated populations (Souza *et al*. 1994; Hagen & Hamrick 1996) while some others have found little geographic variation (Strain *et al*. 1995; Vinuesa *et al*. 2005c).

In previous studies we analysed the spatial and temporal genetic structure of local *R. etli* and *R. gallicum* populations associated with common bean plants (*Phaseolus vulgaris*) cultivated in San Miguel, Mexico (Silva *et al*. 1999, 2003). Population genetic analyses showed that recombination and migration are important evolutionary forces shaping the genetic structure of these local *Rhizobium* populations. In the present study, we analyse a representative sample of the San Miguel strains together with a large collection of *R. gallicum* strains isolated in different continents from beans and other hosts. In recent phylogenetic studies of rhizobia, strains described as *R. gallicum* (Amarger *et al*. 1997), *Rhizobium mongolense* (van Berkum *et al*. 1998) and *Rhizobium yanglingense* (Tan *et al*. 2001) form a consistent and highly supported clade based on the use of different markers (Sawada *et al*. 2003; Young *et al*. 2004; Vinuesa *et al*. 2005b), suggesting that they may be one species. For this reason some *R. mongolense* and *R. yanglingense* strains were also included in the analyses here. Phylogenetic and population genetics approaches based on multilocus sequence analyses of three 'core' genes (*rrs*, *atpD* and *glnII*) and two 'auxiliary' symbiotic loci (*nodB* and *nifH*) were used to address several issues of taxonomic, evolutionary and ecological interest. The particular objectives were (i) to determine by multilocus sequence analyses if the strains of *R. mongolense* and *R. yanglingense* belong to the *R. gallicum* evolutionary lineage, (ii) to address the importance of different evolutionary forces in shaping the genetic structure of several *R. gallicum* populations distributed in the Northern Hemisphere, and (iii) to compare the evolutionary histories of core chromosomal housekeeping genes with those of plasmid-borne symbiotic genes. We discuss our results in the context of current concepts for bacterial biogeography and in the light of the different species concepts used in bacterial systematics.

Materials and methods

Bacterial strains and population sampling

The strains analysed are listed in Table 1 according to their classification, host and country of origin. Most *Rhizobium gallicum* strains were isolated from *Phaseolus vulgaris* (Piñero *et al*. 1988; Eardly *et al*. 1995; Amarger *et al*. 1997; Sessitsch *et al*. 1997b; Herrera-Cervera *et al*. 1999; Mhamdi *et al*. 2002); only three *Onobrychis* isolates were included (Laguerre *et al*. 1997). Nine *Medicago ruthenica* isolates classified as *Rhizobium mongolense* (van Berkum *et al*. 1998) and the only available *Rhizobium yanglingense* isolate (the type strain isolated from *Gueldenstaedtia multiflora*, Tan *et al*. 2001) were included in the analyses. Additional relevant reference strains of related taxa were included in the phylogenetic analyses, yielding a total of 91 strains (Table 1).

PCR amplification and DNA sequencing

Five loci were selected for polymerase chain reaction (PCR) amplification and sequencing with the primers and conditions listed in Table S1, based on a survey of publicly available sequences generated in previous studies. We chose three chromosomally encoded housekeeping genes: (i) the widely used *rrs* (Amarger *et al*. 1997; Wernegreen *et al*. 1997; Gaunt *et al*. 2001; Laguerre *et al*. 2001; Sawada *et al*. 2003; van Berkum *et al*. 2003; Young *et al*. 2004; Vinuesa *et al*. 2005a, b), coding for 16S rRNA [516 bp, spanning from nucleotides 626–1142 of the *R. gallicum*

Table 1 Bacterial strains used in this study

Strain	Classification (reclassification)*	Host	Locality	Reference
IE988	*Rhizobium gallicum* bv. *gallicum*	*Phaseolus vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE992	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2703	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2729	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2735	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2751	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4770	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4845	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4868	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4872	*R. gallicum* bv. *gallicum*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
FL27	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Tehuacán, Mexico	Piñero *et al*. (1988)/Eardly *et al*. (1995)/ Sessitsch *et al*. (1997b)
FL34	*R. leguminosarum* bv. *phaseoli* (*R. gallicum* bv. *gallicum*)	*P. vulgaris*	Tehuacán, Mexico	Piñero *et al*. (1988)/ Eardly *et al*. (1995)
FL44	*R. leguminosarum* bv. *phaseoli* (*R. gallicum* bv. *gallicum*)	*P. vulgaris*	Tehuacán, Mexico	Piñero *et al*. (1988)/ Eardly *et al*. (1995)
FL48	*R. leguminosarum* bv. *phaseoli* (*R. gallicum* bv. *gallicum*)	*P. vulgaris*	Tehuacán, Mexico	Piñero *et al*. (1988)/ Eardly *et al*. (1995)
COC-11	*R. leguminosarum* bv. *phaseoli* (*R. gallicum* bv. *gallicum*)	*P. coccineus*	Tepoztlán, Mexico	Piñero *et al*. (1988)/ Eardly *et al*. (1995)
116-A15	*R. gallicum* bv. *gallicum*	*Onobrychis viciifolia*	Alberta, Canada	Laguerre *et al*. (1997)
SM2	*R. gallicum* bv. *gallicum*	*O. viciifolia*	Alberta, Canada	Laguerre *et al*. (1997)
USDA 3736	*R. gallicum* bv. *gallicum*	*O. transcaucasia*	Unknown	Laguerre *et al*. (1997)
R602sp^T	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Maine et Loire, France	Amarger *et al*. (1997)
DC-22	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Cote d'Or, France	Amarger *et al*. (1997)
PhF29	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Cote d'Or, France	Amarger *et al*. (1997)
PhF222	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Tarn et Garonne, France	Amarger *et al*. (1997)
PhD12	*R. gallicum* bv. *phaseoli*	*P. vulgaris*	Gers, France	Amarger *et al*. (1997)
PhI21	*R. gallicum* bv. *phaseoli*	*P. vulgaris*	Gers, France	Amarger *et al*. (1997)
GR-45	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
GR-89	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
GR-46	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
GR-55	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
GR-18	*R. gallicum* bv. *phaseoli*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
GR-42	*R. gallicum* bv. *phaseoli*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
GR-60	*R. gallicum* bv. *phaseoli*	*P. vulgaris*	Granada, Spain	Herrera-Cervera *et al*. (1999)
CbS-1	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Seibersdorf, Austria	Sessitsch *et al*. (1997)
CbS-3	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Seibersdorf, Austria	Sessitsch *et al*. (1997)
CbS-17	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Seibersdorf, Austria	Sessitsch *et al*. (1997)
LAAP 23C2	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Bizerte, Tunisia	Mhamdi *et al*. (2002)
LAAP 28B3	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Bizerte, Tunisia	Mhamdi *et al*. (2002)
LAAP 32A5	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Bizerte, Tunisia	Mhamdi *et al*. (2002)
LAAP 28B7	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Bizerte, Tunisia	Mhamdi *et al*. (2002)
LAAP 8A3	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Cap bon, Tunisia	Mhamdi *et al*. (2002)
LAAP 21B7	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Bizerte, Tunisia	Mhamdi *et al*. (2002)
LAAP 7B4	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Cap bon, Tunisia	Mhamdi *et al*. (2002)
LAAP 10B2	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Cap bon, Tunisia	Mhamdi *et al*. (2002)
LAAP 7B6	*R. gallicum* bv. *gallicum*	*P. vulgaris*	Cap bon, Tunisia	Mhamdi *et al*. (2002)
USDA 1844^T	*R. mongolense* (*R. gallicum* bv. *orientale*)	*Medicago ruthenica*	Tongliao, China	van Berkum *et al*. (1998)
USDA 1832	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Huhehot, China	van Berkum *et al*. (1998)
USDA 1834	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Xilinhot, China	van Berkum *et al*. (1998)
USDA 1836	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Xilinhot, China	van Berkum *et al*. (1998)
USDA 1849	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Tongliao, China	van Berkum *et al*. (1998)
USDA 1877	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Xilinhot, China	van Berkum *et al*. (1998)

Table 1 *Continued*

Strain	Classification (reclassification)*	Host	Locality	Reference
USDA 1890	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Huhehot, China	van Berkum *et al*. (1998)
USDA 1904	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Xilinhot, China	van Berkum *et al*. (1998)
USDA 1929	*R. mongolense* (*R. gallicum* bv. *orientale*)	*M. ruthenica*	Xilinhot, China	van Berkum *et al*. (1998)
SH22632T	*R. yanglingense* (*R. gallicum* bv. *orientale*)	*Gueldenstaedtia multiflora*	Gansu, China	Tan *et al*. (2001)
SH2462	*R. yanglingense* (*R. gallicum* bv. *orientale*)	*Coronilla varia*	Gansu, China	Tan *et al*. (2001)
SH445	*R. yanglingense* (*R. gallicum* bv. *orientale*)	*Coronilla varia*	Shaanxi, China	Tan *et al*. (2001)
SH1718	*R. yanglingense* (*R. gallicum* bv. *orientale*)	*Coronilla varia*	Shaanxi, China	Tan *et al*. (2001)
Sh19351	*Rhizobium* sp. (*Rhizobium* sp. bv. *orientale*)	*Glycyrrhiza mulenses*	Shaanxi, China	Wernegreen & Riley (1999)
N95	*Rhizobium* sp. (*Rhizobium* sp. bv. *orientale*)	*Hedysarum mongolicum*	China	Wernegreen & Riley (1999)
IE950	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE954	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE1004	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE1009	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2704	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2730	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2737	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE2755	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4794	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4795	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4771	*R. etli* bv. *phaseoli* (*R. etli* bv. *mimosae*)	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4803	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4804	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4810	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4837	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4874	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
IE4876	*R. etli* bv. *phaseoli*	*P. vulgaris*	San Miguel, Mexico	Silva *et al*. (2003)
CFN42T	*R. etli* bv. *phaseoli*	*P. vulgaris*	Mexico	Segovia *et al*. (1993)
KIM5S	*R. etli* bv. *phaseoli*	*P. vulgaris*	Idaho, USA	Josephson *et al*. (1984)
GR-12	*R. etli* bv. *phaseoli*	*P. vulgaris*	Granda, Spain	Herrera-Cervera *et al*. (1999)
4PR-2	*R. etli* bv. *phaseoli*	*P. vulgaris*	Alcolea del Rio, Spain	Rodriguez-Navarro *et al*. (2000)
Mim2	*R. etli* bv. *mimosae*	*Mimosa affinis*	Huautla, Mexico	Wang *et al*. (1999)
GR-84	*R. leguminosarum* bv. *phaseoli*	*P. vulgaris*	Granda, Spain	Herrera-Cervera *et al*. (1999)
VF39	*R. leguminosarum* bv. *viciae*	*Vicia faba*	Bielefeld, Germany	Priefer (1989)
3841	*R. leguminosarum* bv. *viciae*	*Pisum sativa*	England	http://www.sanger.ac.uk/Projects/R_leguminosarum/
CIAT 899T	*R. tropici*	*P. vulgaris*	Colombia	Martínez-Romero *et al*. (1991)
CFN 299	*R. tropici*	*P. vulgaris*	Brazil	Martínez-Romero *et al*. (1991)
USDA 1920	*Rhizobium* sp. (*Rhizobium* sp. bv. *orientale*)	*M. ruthenica*	Xilinhot, China	van Berkum *et al*. (1998)
TJ167	*Rhizobium* sp. (*R. etli* bv. *mimosae*)	*Mimosa diplotricha*	Taoyuan, Taiwan	Chen *et al*. (2003)
TJ173	*Rhizobium* sp. (*R. etli* bv. *mimosae*)	*Mimosa diplotricha*	Nantou, Taiwan	Chen *et al*. (2003)
A321T	*Sinorhizobium medicae*	*Medicago truncatula*	Aude, France	Rome *et al*. (1996)
1021	*S. meliloti*	*M. sativa*	Australia	http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/
USDA 110	*Bradyrhizobium japonicum*	*Glycine max*	USA	www.kazusa.or.jp/rhizobase/Bradyrhizobium

*Strain reclassifications proposed in this work are presented in parentheses.

R602sp^T sequence (U86343), which includes one of the hypervariable regions of the gene]; (ii) *glnII* coding for glutamine synthetase II (540 bp), which has been selected as a molecular marker in several studies (Wernegreen & Riley 1999; Turner & Young 2000; Vinuesa *et al*. 2005b, c) and (iii) *atpD* coding for the beta subunit of ATPase (441 bp) (Gaunt *et al*. 2001; Vinuesa *et al*. 2005b, c). Two plasmid-encoded symbiotic genes were also selected: *nifH* coding for the dinitrogenase reductase (387 bp), which is the most widely studied nitrogen fixation gene (Laguerre *et al*. 2001; Vinuesa *et al*. 2005b, c), and *nodB* coding for N-acetylglucosmine deacetylase (516 bp), an essential gene for Nod factor biosynthesis, which was previously analysed by Wernegreen & Riley (1999).

Genomic DNA was prepared by using the Amersham Tissue and Cell Kit and amplifications were performed with *Taq* polymerase (Roche) using the cycling parameters and reaction mixture described in Silva *et al*. (2003). Amplification products were purified using the PCR product purification system of Roche and sequenced using BigDye chemistry and an ABI377 automatic sequencer (ABI). The accession numbers for the sequences generated in this study are AY907379–AY907521 and AY929391–AY929593. All the accession numbers used in this study are listed in Table S2.

Phylogenetic analyses

Multiple nucleotide sequence alignments were generated with CLUSTAL_X (Thompson *et al*. 1997) and edited with BIOEDIT (Hall 1999) and DAMBE version 4.2.13 (Xia & Xie 2001). The protein-coding nucleotide sequences were aligned based on the alignment of the encoded products using DAMBE. Neighbour-joining (NJ), maximum-likelihood (ML) and Bayesian phylogenetic inferences were carried out for each data set and for the concatenated *glnII* and *atpD* partitions according to the procedures detailed in Vinuesa *et al*. (2005). Model fitting was performed by likelihood-ratio tests (LRT, Huelsenbeck & Rannala 1997) using MODELTEST 3.06 (Posada & Crandall 1998), PAUP*4.0v10b (Swofford 2002) and PHYML version 2.1 (Guindon & Gascuel 2003), and manually optimized by consecutive cycles of parameter and ML phylogeny estimation in PAUP*, until convergence. The robustness of NJ and ML topologies were inferred by nonparametric bootstrap test using MEGA2 (Kumar *et al*. 2001) and PHYML with 1000 and 100 pseudoreplicates, respectively. Bayesian phylogenetic analyses were performed with MRBAYES 3b4 (Ronquist & Huelsenbeck 2003). The best-fit model for each alignment was selected either by LRT or the Akaike information criterion as implemented in MODELTEST (Posada & Crandall 1998). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) with four chains was used to sample trees. Each analysis was replicated two times for 3×10^6 generations, sampling the posterior distribution every 100th generation. The post burn-in samples from the stationary phase were pooled to obtain a 50% majority rule consensus tree using MRBAYES. For the analysis of concatenated sequences, characters were partitioned by gene as well as by codon position, estimating best-fit models for each partition as previously described (Vinuesa *et al*. 2005c).

Population genetics analyses of DNA polymorphisms

The K_{ST}* statistic was chosen to detect geographic subdivision directly from sequence polymorphisms, as it is more powerful than the haplotype-based statistics (Hudson *et al*. 1992a). The permutation-based method for assessing statistical significance was used with 10^4 replicates (Hudson *et al*. 1992a). We calculated F_{ST} statistics to estimate the average level of gene flow, as measured by the effective number of migrants per generation *Nm* (Hudson *et al*. 1992b). To determine departures from the neutral model of evolution, we calculated three indexes: Tajima's *D*, Fu's F_S and R_2 (Tajima 1989; Fu 1997; Ramos-Onsins & Rozas 2002). Departures from the neutral model expectation can be due to natural selection acting on the locus, genetic hitch-hiking or demographic processes (Tajima 1989; Ramos-Onsins & Rozas 2002). The statistical significance of the neutrality and population growth tests was determined by coalescent simulations based on 10^4 replicates. The descriptive statistic of nucleotide polymorphisms, genetic diversity, population differentiation and gene flow estimates, as well as neutrality and population growth tests were computed with DNASP 4.0 (Rozas & Rozas 1999). The level of intragenic recombination was estimated using the *R* parameter (Hudson 1987) calculated with DNASP, whereas the coalescent estimator of the population recombination rate (*C*) and the recombination/mutation rate ratio (c/μ) were determined using SITES 1.1 (Hey & Wakeley 1997). The proportion of total genetic variance attributable to geographic population subdivision was estimated by analyses of molecular variance (AMOVA, Excoffier *et al*. 1992) using ARLEQUIN 2.0 (Schneider *et al*. 2000). The AMOVAs were carried out with Tamura–Nei + Gamma distances and the best-fit alpha value determined for each sequence set under the ML criterion. To determine the association between pairwise geographic and genetic distances (F_{ST}) of populations, Mantel tests (Smouse *et al*. 1986) were performed with 10^4 permutations using ARLEQUIN.

Results

*Phylogenetic inferences based on the 16S rRNA gene (*rrs*)*

All *rrs* phylogenies (NJ, ML and Bayesian) were poorly resolved. Different topologies and node support values were obtained depending on the strain sampling, reconstruction

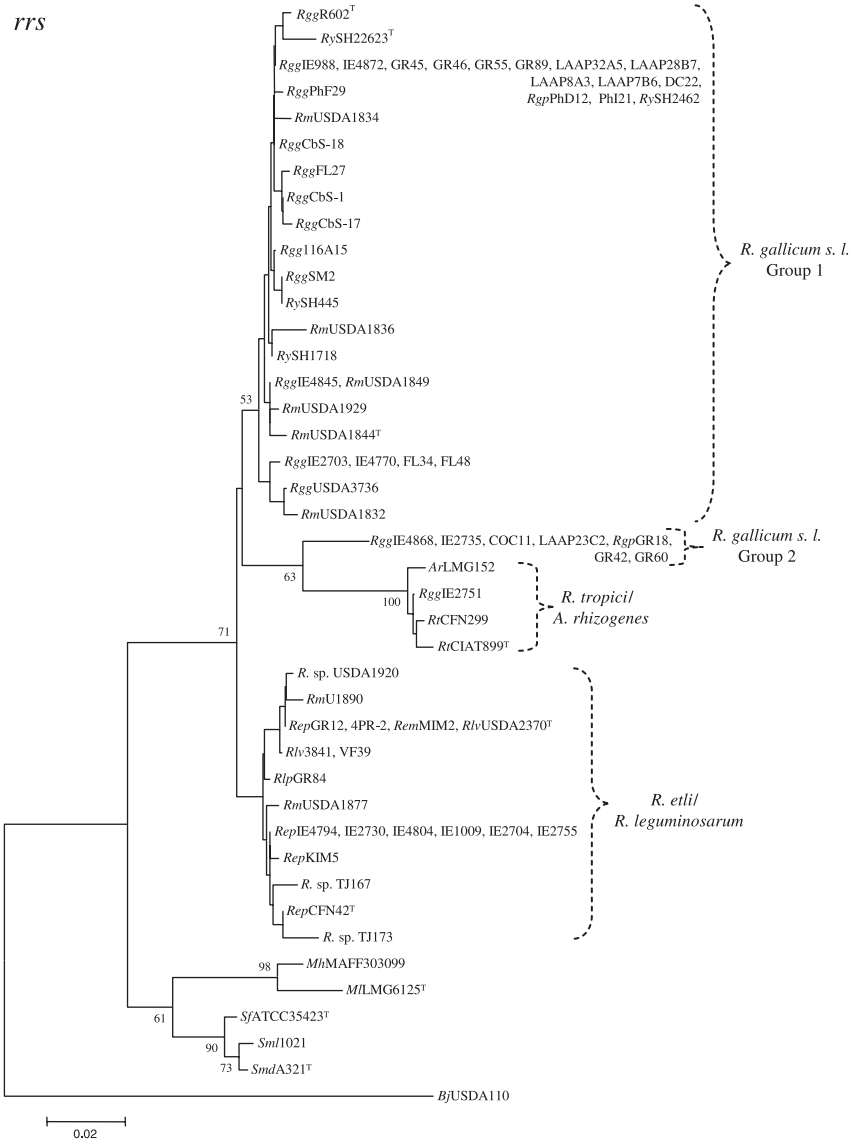

Fig. 1 Neighbour-joining *rrs* gene tree estimated using the TrN + G substitution model ($\alpha = 0.15$) for 75 sequences and 541 sites. Bootstrap support values > 50% are indicated at the relevant nodes. Strains with identical haplotypes are included in the same terminal node (note that sequences can belong to different species or biovars). Type strains are indicated by a superscript T. The genus, species and biovar of the strains are abbreviated as follows: *Rgg*, *Rhizobium gallicum* bv. *gallicum*; *Rgp*, *R. gallicum* bv. *phaseoli*; *Rm*, *R. mongolense*; *Ry*, *R. yanglingense*; *Rt*, *R. tropici*; *Ar*, *Agrobacterium rhizogenes*; *Rep*, *R. etli* bv. *phaseoli*; *Rem*, *R. etli* bv. *mimosae*; *Rlv*, *R. leguminosarum* bv. *viciae*; *Mh*, *Mesorhizobium huakuii*; *Ml*, *M. loti*; *Sf*, *Sinorhizobium fredii*; *Sml*, *S. meliloti*; *Smd*, *S. medicae*; *Bj*, *Bradyrhizobium japonicum*.

method and substitution model used. In Fig. 1 we present an NJ tree using Tamura–Nei + Gamma distances ($\alpha = 0.15$). Most of the *Rhizobium gallicum* isolates had closely related haplotypes (1% differences in average) and grouped together with several *Rhizobium mongolense* and *Rhizobium yanglingense* strains in a tight but weakly supported clade, that hereafter will be referred to as Group 1. The *R. mongolense* and *R. yanglingense* strains were intermingled with *R. gallicum* strains, providing evidence that they are part of the same evolutionary lineage. Moreover, identical haplotypes were shared between these species. The most abundant *rrs* haplotype was detected in *R. gallicum* strains isolated in Mexico, France, Spain and Tunisia, along with a *R. yanglingense* strain isolated in China. The haplotype of the Mexican *R. gallicum* strain IE4845 was identical to that of *R. mongolense* strain USDA 1849.

A subset of *R. gallicum* strains displayed a divergent *rrs* haplotype for which we did not find any closely related (> 97% identity) sequence in GenBank. This haplotype was found in strains from Mexico, Spain and Tunisia, and hereafter is referred to as *R. gallicum* Group 2. This lineage was associated either with the *Rhizobium tropici*/*A. rhizogenes* group (as shown in Fig. 1) or basal to *R. gallicum* Group 1, depending on the choice of strains and reconstruction method used. In the *rrs* phylogeny three other phylogenetic/taxonomic incongruences were detected. Two strains classified as *R. mongolense* (USDA 1877 and USDA 1890) displayed haplotypes related to those of the *Rhizobium etli*/*Rhizobium leguminosarum* group. In this phylogeny the latter group is not statistically supported (bootstrap support < 50%), and it is not possible to establish a clear differentiation between *R. etli* and *R. leguminosarum*; this difficulty

may be due to the small divergence between the *rrs* haplotypes of these species, or to genetic exchange between them, as previously proposed (Eardly *et al*. 1995). The third phylogenetic/taxonomic incongruence detected for the *rrs* gene involved the Mexican *R. gallicum* strain IE2751, which displayed a haplotype closely related to those of *R. tropici*/*A. rhizogenes*. These incongruences could be due to genetic exchange at the *rrs* locus between several *Rhizobium* species, as documented in other studies (Eardly *et al*. 1995, 1996; van Berkum *et al*. 2003; Vinuesa *et al*. 2005b).

Phylogenetic inferences based on the protein-coding genes glnII *and* atpD

The phylogenetic inferences based on the protein-coding chromosomal housekeeping genes *glnII* and *atpD* were very similar to each other (data not shown) and produced much better resolved topologies than the *rrs* sequences. Due to their similar results, we decided to concatenate the *glnII* and *atpD* sequences to generate a combined phylogeny (Fig. 2) employing a Bayesian analysis since this approach

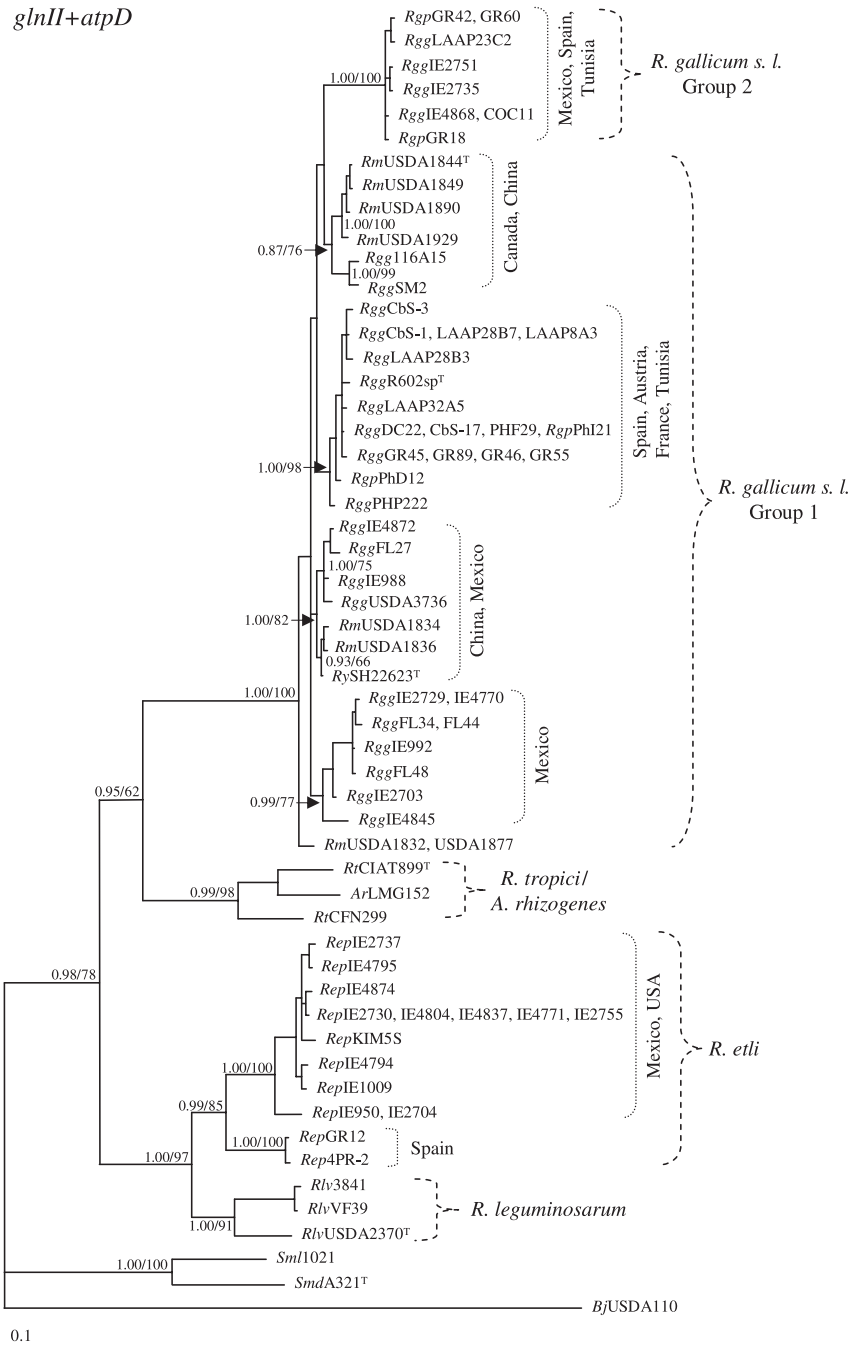

Fig. 2 Bayesian phylogeny inferred from *glnII+atpD* sequence partitions for 72 sequences and 993 sites. The topology represents the 50% majority rule consensus tree derived from 40 000 trees resulting from the convergent and pooled post burn-in samples obtained from two independent MCMCMC samplers. The substitution models were partitioned by gene as well as by codon position using the HKY + G for the first and third positions, and $F_{81} + I$ for the second positions. Posterior probabilities > 0.85 and the maximum-likelihood bootstrap support values > 60% are indicated at the relevant nodes. Strains with identical haplotypes are included in the same terminal node (note that sequences can belong to different species or biovars). Type strains are indicated by a superscript T. Abbreviations are detailed in Fig. 1.

allows the use of mixed models for partitioned data analyses (Ronquist & Huelsenbeck 2003; Nylander *et al*. 2004; Vinuesa *et al*. 2005c). In the *glnII+atpD* Bayesian phylogeny all the *R. mongolense*, *R. yanglingense* and *R. gallicum* strains are grouped in a strongly supported clade, reinforcing the conclusion that these three named species constitute a single evolutionary lineage. Since *R. gallicum* was the first species from this lineage to be described, we will refer to this clade as the *R. gallicum sensu lato* (*s. l.*) lineage. This clade exhibits significant internal structure, displaying five internal subclades. Group 2 strains formed the most differentiated and supported subclade in the *R. gallicum s. l.* lineage, suggesting that they constitute a divergent branch containing strains from various geographic origins. The other four subclades within this lineage grouped strains largely, although not strictly, according to their geographic origin (Fig. 2). The most basal split in the *R. gallicum s. l.* clade is formed by a haplotype shared by two strains from China. In contrast with the *rrs* phylogeny, the *R. etli* and *R. leguminosarum* species clades are well resolved as sister lineages. Within *R. etli* two subclades were recovered, grouping the New and Old World strains.

Phylogenetic inferences based on the symbiotic genes nifH *and* nodB

The phylogenetic reconstructions based on the sequences of each of the two plasmid-encoded symbiotic loci (*nifH* and *nodB*) produced similar topologies under the different reconstruction methods. The ML phylogenies for these genes are presented in Figs 3 and 4, respectively. For both genes three different clades were delineated within *R. gallicum s. l.* strains, corresponding to the biovars *gallicum*, *phaseoli* and *orientale* (see below). In the *nifH* phylogeny the biovars *gallicum* and *phaseoli* form a well-supported clade, in which the alleles of *R. etli* biovar *mimosae* are placed as the sister lineage of the biovar *phaseoli*. Although the alleles of the biovars *gallicum* and *phaseoli* are only slightly differentiated, of the 10 average differences, three of them are fixed differences that clearly differentiate them. In the *nodB* phylogeny the clades representing the *gallicum* and *phaseoli* biovars are well supported as sister groups, but in contrast with the *nifH* phylogeny they are much more divergent. The biovar *mimosae* alleles could not be amplified with the *nodB* primers and conditions used. The third clade revealed by the *nifH* and *nodB* genes grouped the *R. mongolense* and *R. yanglingense* strains. In the *nifH* phylogeny this clade was related to *R. leguminosarum* bv. *viciae* haplotype, whereas in the *nodB* phylogeny it was closely related to *Sinorhizobium meliloti* and *Sinorhizobium medicae* sequences. This newly discovered lineage of symbiotic loci is found among strains isolated from very diverse legume tribes (Table 1), and the only obvious ecological feature we could find shared by these strains is that they were isolated in China; hereafter, we will refer to this lineage as biovar *orientale*.

The topologies recovered for the *nifH* and *nodB* genes were clearly different. In the *nifH* phylogeny the polymorphisms among the biovars *gallicum*, *phaseoli* and *mimosae* were due to neutral mutations involving synonymous substitutions, whereas for the *nodB* gene the protein products were highly divergent, providing evidence for distinct evolutionary histories of the nitrogen fixation and nodulation genes, as previously noted in other studies (Young & Johnston 1989; Laguerre *et al*. 2001).

Several phylogenetic incongruences between the groupings established by chromosomal and symbiotic loci provided evidence of lateral transfer of symbiotic information. In the *R. gallicum* bv. *gallicum* clades of the *nifH* and *nodB* phylogenies (Figs 3 and 4), *R. gallicum* Group 1 and Group 2 strains appear intermingled. This indicates that Group1 and Group 2 strains are not genetically isolated, and that they share part of the symbiotic pool of the species. The lateral transfer of symbiotic information between *R. gallicum* and *R. etli* is evident from the monophyletic grouping of *R. etli* bv. *phaseoli* and *R. gallicum* bv. *phaseoli* Group 1 and Group 2 strains in the *nifH* and *nodB* phylogenies (Figs 3 and 4). A similar conclusion can be drawn for the biovar *orientale*, since it seems to share common ancestors with symbiotic alleles from *R. leguminosarum* bv. *viciae* (*nifH*) or *S. meliloti* and *S. medicae* (*nodB*).

Genetic diversity and DNA polymorphisms of R. gallicum sensu lato

Analyses of DNA polymorphisms were performed for the total *R. gallicum s. l.* population and for the monophyletic groupings; therefore, the chromosomal genes were analysed separately for the *R. gallicum s. l.* Groups 1 and 2, whereas the polymorphisms of the symbiotic genes were analysed by grouping the strains according to biovariety. Descriptive statistics of nucleotide polymorphisms and neutrality tests for the five loci analysed are presented in Table 2. For the *rrs* locus all the Group 2 strains displayed the same haplotype, and thus population genetic analyses could not be performed. Group 1 strains presented several haplotypes that on average were 1% divergent, and if the complete sample of *R. gallicum s. l.* strains (including strains IE2751, U1877 and U1890) was taken into account, a relatively high genetic diversity was found ($\pi = 0.0138$). The levels of genetic diversity of the *glnII* and *atpD* loci were similar ($\pi = 0.0230$ and $\pi = 0.0209$, respectively). In both cases Group 1 exhibited higher genetic and haplotype diversity than Group 2; however, this difference could result from a smaller sample size for Group 2.

The genetic diversity of the symbiotic genes was higher than that of the chromosomal genes, due to the presence of three different biovars within the *R. gallicum s. l.* symbiotic

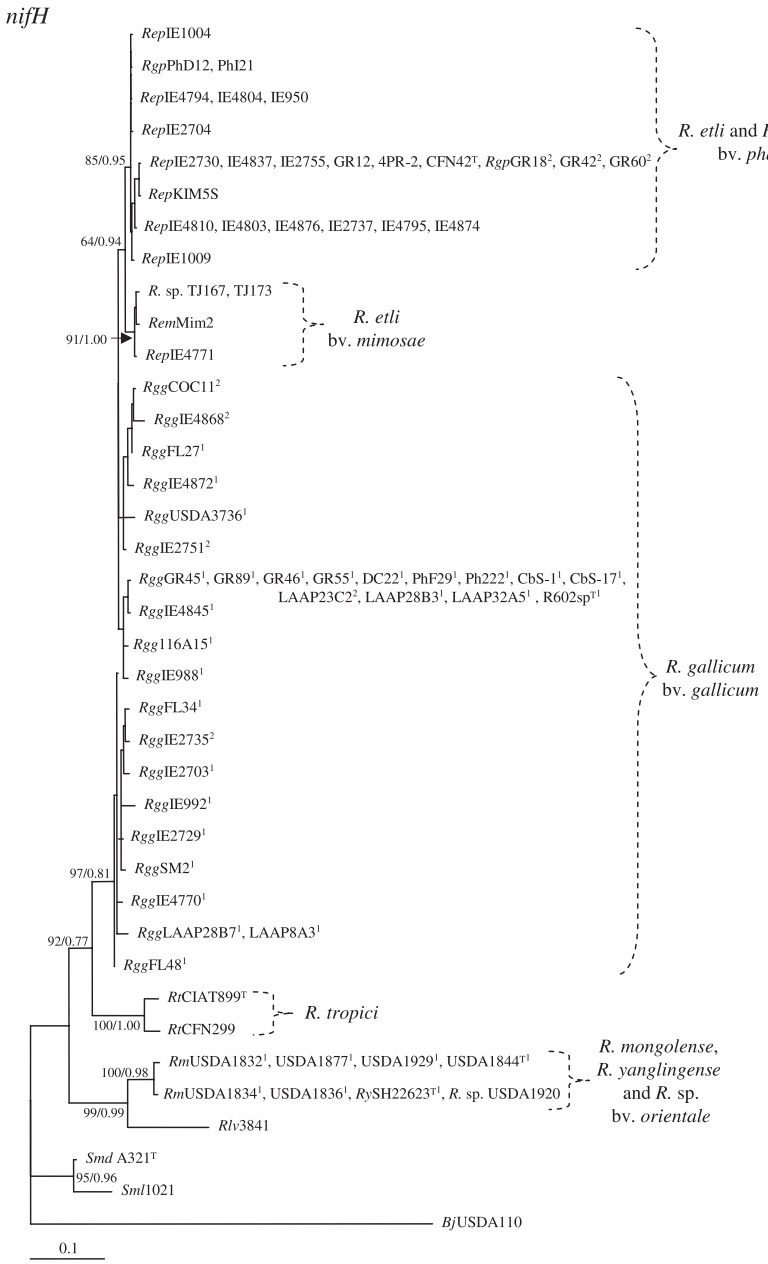

Fig. 3 Maximum-likelihood *nifH* gene tree estimated using the HKY + G substitution model ($\alpha = 0.26$) for 75 sequences and 387 sites. Bootstrap support values > 60% and posterior probabilities > 0.80 are indicated at the relevant nodes. The brackets to the right of the tree indicate the clades constituting different biovars. Strains with identical haplotypes are included in the same terminal node (note that sequences can belong to different species). Type strains are indicated by a superscript T. *Rhizobium gallicum s. l.* Group1 and Group 2 strains are denoted by superscripts 1 and 2, respectively. Abbreviations are detailed in Fig. 1.

gene pool. The *nifH* locus displayed a characteristically lower genetic diversity ($\pi = 0.0458$) than the *nodB* locus ($\pi = 0.1243$). The *nodB* gene was the most polymorphic of the five loci analysed; it exhibited a large number of non-synonymous substitutions within and among biovarieties (Table 2), suggesting that diversifying selection may have contributed to the divergence of these proteins, which are involved in the synthesis of the nodulation factors, and their co-evolution with host plants. For both symbiotic genes biovar *orientale* was the least polymorphic of the three biovars, despite its larger sample size compared with that of biovar *phaseoli*, and despite the fact that it contains isolates from diverse leguminous host (Table 1). We can speculate that this difference could be due to a founder effect caused by the recent acquisition of these symbiotic genes by the *R. gallicum s. l.* evolutionary lineage, or to the fact that the hosts of the biovar *orientale* are wild legumes, in contrast with *Phaseolus vulgaris*, which is cultivated.

Most of the neutrality and population growth tests were nonsignificant for the *R. gallicum s. l.* population (Table 2), indicating that the overall observed polymorphisms conform to the neutral equilibrium model (Tajima 1989). However, the Group 1 strains showed significant values for the chromosomal loci *rrs* and *atpD*, and biovar *gallicum* strains showed significant values for the *nifH* locus. These results

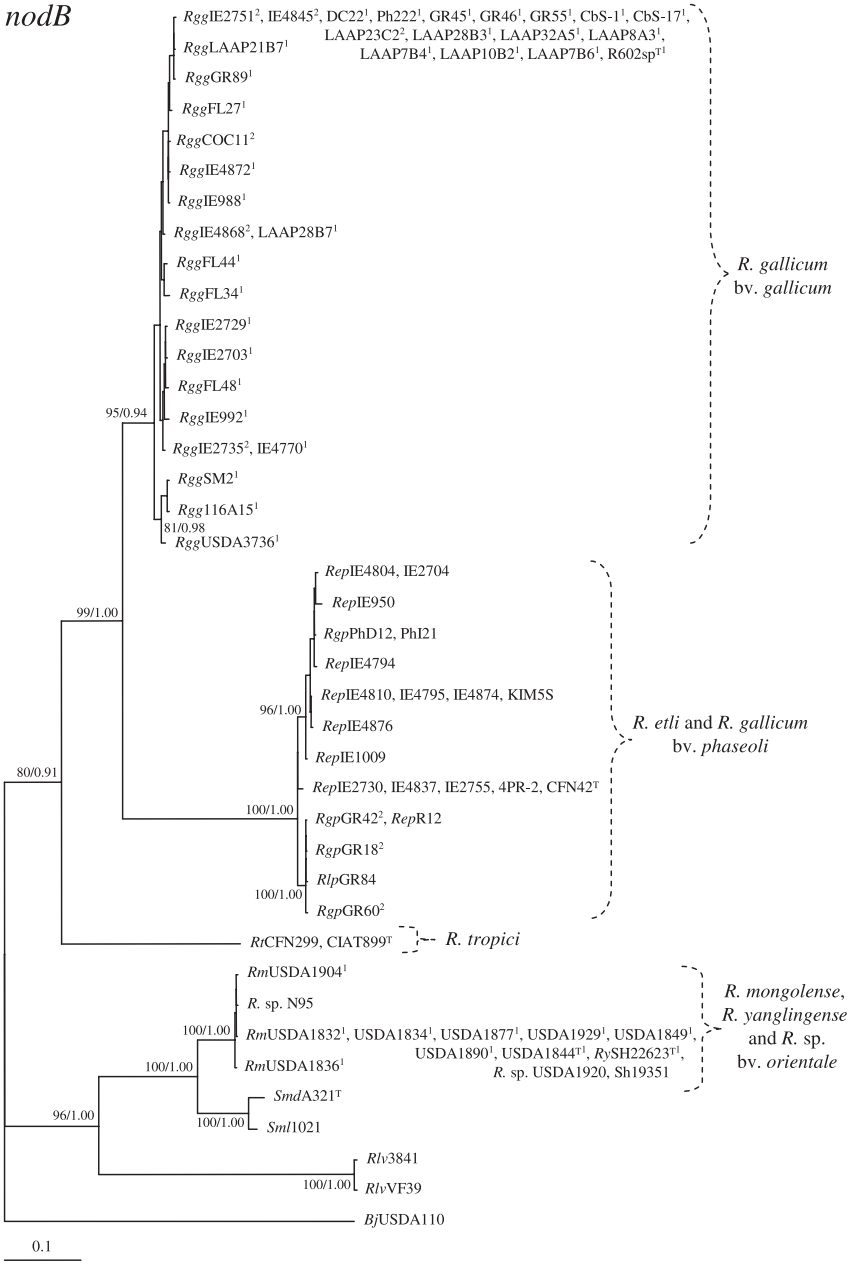

Fig. 4 Maximum-likelihood *nodB* gene tree estimated using the HKY + G substitution model ($\alpha = 0.66$) for 78 sequences and 516 sites. Bootstrap support values > 60% and posterior probabilities > 0.85 are indicated at the relevant nodes. The brackets to the right of the tree indicate the clades constituting different biovars. Strains with identical haplotypes are included in the same terminal node (note that sequences can belong to different species). Type strains are indicated by a superscript T. *Rhizobium gallicum s. l.* Group1 and Group 2 strains are indicated by superscripts 1 and 2, respectively. Abbreviations are detailed in Fig. 1.

could be due to purifying selection acting on these loci or to population expansion of the *R. gallicum* bv. *gallicum* strains (Fu 1997; Ramos-Onsins & Rozas 2002).

Recombination in the R. gallicum sensu lato *lineage*

The recombination estimates were performed both for the total population and for the different haplotypes. Due to the small sample size and the low levels of polymorphism found in the Group2, biovar *phaseoli* and *orientale* strains, the recombination estimates were performed only for the total sample, Group 1 and biovar *gallicum* strains. Some general patterns can be observed in the recombination estimates shown in Table 3. First, the recombination estimates are much higher if only the different haplotypes are analysed, as expected for a population with an epidemic genetic structure, in which the frequency of a few dominant genotypes increases to produce epidemic clones, but frequent genetic exchange occurs among the members of the population (Maynard Smith *et al*. 1993; Silva *et al*. 1999, 2003). Second, in most of the cases the estimates of recombination increased for the chromosomal genes if only the Group 1 strains were analysed, providing evidence of genetic compartments within the species. Third, despite

Table 2 Descriptive statistics of nucleotide polymorphism, genetic diversity estimates, neutrality and growth tests for the *Rhizobium gallicum sensu lato* population

Gene	Group	No. of sites	No. of seqs†	*S*‡	Pi§	*ks*/*ka*¶	*k*††	*h*/*Hd*‡‡	π§§	θ¶¶	Tajima's *D*†††	Fu's F_S†††	R_2†††
rrs	Total	539	47	44	26		7.02	22/0.868	0.0138	0.0194	−1.005	−4.340	0.075
	Group 1 + 2		44	31	20		5.65	19/0.849	0.0111	0.0139	−0.688	−3.832	0.085
	Group 1		37	23	11		3.12	18/0.818	0.0061	0.0107	−1.459*	−8.879*	0.064*
	Group 2		7	0	0	0/0	0	0/0.000					
glnII	Total	540	54	61	48	61/4	12.41	27/0.939	0.0230	0.0248	−0.252	−2.853	0.098
	Group 1		46	50	37	48/4	9.36	25/0.927	0.0173	0.0211	−0.619	−4.955	0.088
	Group 2		8	1	1	1/0	0.57	2/0.571	0.0011	0.0007	1.444	0.966	0.286
atpD	Total	441	48	50	39	50/3	9.22	19/0.923	0.0209	0.0256	−0.630	−0.685	0.089
	Group 1		40	38	24	37/3	6.28	16/0.897	0.0143	0.0203	−1.034*	−1.566	0.080*
	Group 2		8	3	3	3/0	1.50	3/0.750	0.0034	0.0026	1.220	0.971	0.250
nifH	Total	387	44	68	56	70/5	16.22	23/0.903	0.0458	0.0442	0.135	−0.453	0.114
	gallicum		32	33	16	36/1	5.83	19/0.841	0.0165	0.0232	−1.036*	−6.126	0.076*
	phaseoli		5	6	6	6/0	3.60	2/0.600	0.0093	0.0074	1.718	3.967	0.300
	orientale		7	3	3	3/0	1.71	2/0.571	0.0044	0.0032	1.811	2.920	0.286
nodB	Total	516	51	203	198	118/66	63.99	23/0.855	0.1243	0.0876	1.507	14.016	0.157
	gallicum		36	38	24	30/10	7.91	17/0.754	0.0153	0.0178	−0.488	−1.593	0.095
	phaseoli		5	13	12	8/5	7.60	3/0.800	0.0148	0.0121	1.583	3.183	0.2786
	orientale		10	2	0	1/1	0.40	3/0.378	0.0008	0.0014	−1.401	−1.164	0.2000

†Number of sequences; ‡segregating sites; §parsimony informative sites; ¶total number of synonymous/nonsynonymous changes; ††average number of nucleotide differences; ‡‡number of haplotypes/haplotype (gene) diversity; §§pi, nucleotide diversity per site; ¶¶theta per site; †††calculations using the total number of segregating sites.
*Significant values, as estimated by coalescent simulations.

Table 3 Recombination estimates for the *Rhizobium gallicum sensu lato* population

		Total				Haplotypes			
Gene	Group	*n*	*R**	*C*†	*c*/μ‡	*n*	*R*	*C*	*c*/μ
rrs	Total	48	0.7	0.0043	0.411	24	2.0	0.0189	1.545
	Group 1 + 2	45	0.8	0.0037	0.506	21	6.1	0.0203	2.279
	Group 1	38	3.9	0.0082	1.470	20	31.6	0.0216	3.283
glnII	Total	54	5.5	0.0089	0.719	27	19.2	0.0166	1.134
	Group 1	46	6.9	0.0050	0.518	25	32.0	0.0128	1.042
atpD	Total	48	4.9	0.0006	0.045	19	14.9	0.0040	0.246
	Group 1	40	12.8	0.0003	0.030	16	92.0	0.0046	0.355
nifH	Total	44	0.001	0.0100	0.443	23	0.001	0.0237	0.884
	gallicum	32	8.1	0.0084	0.680	19	174.0	0.0227	1.601
nodB	Total	51	0.001	0.0212	0.485	24	0.001	0.0375	0.711
	gallicum	36	1.2	0.0039	0.437	17	18.4	0.0135	1.239

*Estimate of the population recombination parameter *R* (Hudson 1987) corrected for haploid organisms.
†Estimate of population recombination parameter using the coalescent approach of Hey & Wakeley (1997) corrected for haploid organisms.
‡Recombination/mutation rate ratio.

the evidence that lateral transfer events have a significant impact on the evolution of the symbiotic genes, when the three biovars were analysed together the recombination parameters (*R*) were close to 0, indicating that homologous recombination is not taking place among alleles of the different biovars, possibly due to sequence divergence or selective constrains imposed on these adaptive loci. Fourth, the highest levels of recombination were detected

for the *rrs* locus, in agreement with the evidence gathered from the phylogenetic analyses.

Evidence for gene flow and geographic structuring among R. gallicum sensu lato *populations*

Genetic differentiation analyses were performed on populations with reference to their geographic origins. For the two sampling sites in Mexico (San Miguel and Tehuacán), nonsignificant genetic differentiation and *Nm* values higher than 4 were obtained for the three chromosomal genes (data not shown), indicating that these strains could be pooled to constitute the Mexican population. For *R. yanglingense* we could obtain only the type strain; however, three additional *rrs* sequences could be retrieved from GenBank for this species. Genetic differentiation analyses with the *rrs* sequences showed that the *R. yanglingense* strains were not significantly differentiated from those of *R. mongolense* (K_{ST} = 0.044 $P > 0.05$, F_{ST} = 0.123, *Nm* = 3.59). Therefore, in further analyses the *R. yanglingense* type strain was pooled with the *R. mongolense* strains to form the Chinese population.

The pairwise genetic differentiation and gene flow estimates based on the *rrs* sequences are presented in Table 4. Only a few of the comparisons showed significant differentiation (K_{ST}*) between populations at a $P < 0.05$ level, but none was significant at the 0.01 level. Likewise, most of the pairwise *Nm* estimates are higher than 1, suggesting that within *R. gallicum s. l.* enough gene flow occurs to consider it a single evolutionary unit. In contrast to the *rrs* gene, the *glnII* and *atpD* sequences revealed significant genetic differentiation for some comparisons, uncovering patterns of biogeographic differentiation. All the pairwise comparisons between the European populations (Spain, Austria and France) and the North African population (Tunisia) showed nonsignificant K_{ST}* values, and the gene flow estimates (*Nm*) were higher than 1 (Table 5), suggesting that they can be considered as a single population, accordingly designated as the Mediterranean population. On the other hand, all the comparisons involving the populations from Mexico or China showed significant differentiation (Table 5), although some of the *Nm* values involving the Mexican population were higher than 1, which may reflect historical dispersion pathways of bean seeds. The general pattern of genetic differentiation uncovered three geographically structured populations within our sample, which will be referred to as the Mexican, Mediterranean and Chinese regions. The results of the pairwise comparison between

Table 4 Genetic differentiation and gene flow estimates between *Rhizobium gallicum sensu lato* populations (the sample size is indicated in parentheses) based on *rrs* sequences. Above the diagonal are the K_{ST}* values†. Below the diagonal are the F_{ST} values and their associated *Nm* values

	Mexico (11)	Spain (7)	Austria (3)	France (5)	Tunisia (5)	China (10)
Mexico	—	0.024 NS	0.037 NS	0.094*	0.039 NS	0.054*
Spain	0.015, 33.02	—	0.047 NS	0.074 NS	0.000 NS	0.075*
Austria	0.339, 0.97	0.333, 1.00	—	0.087 NS	0.017 NS	0.000 NS
France	0.320, 1.06	0.288, 1.23	0.208, 1.90	—	0.000 NS	0.015 NS
Tunisia	0.045, 10.50	0.000‡	0.098, 4.60	0.000	—	0.018 NS
China	0.144, 2.98	0.215, 1.82	0.223, 1.74	0.135, 3.22	0.043, 11.24	—

†NS, non significant; *$0.01 < P < 0.05$.
‡with F_{ST} values of 0, *Nm* tends to infinite.

Table 5 Genetic differentiation and gene flow estimates between *Rhizobium gallicum sensu lato* populations (the sample size is indicated in parentheses) based on *glnII+atpD* concatenated sequences. Above the diagonal are the K_{ST}* values†. Below the diagonal are the F_{ST} values and their associated *Nm* values

	Mexico (15)	Spain (7)	Austria (3)	France (6)	Tunisia (5)	China (9)
Mexico	—	0.092**	0.080**	0.146***	0.089**	0.095***
Spain	0.153, 2.77	—	0.059 NS	0.124 NS	0.040 NS	0.183***
Austria	0.523, 0.46	0.333, 1.00	—	0.000 NS	0.000 NS	0.149**
France	0.463, 0.58	0.277, 1.31	0.165, 2.54	—	0.047 NS	0.211***
Tunisia	0.242, 1.57	0.000‡	0.000	0.003, 156.00	—	0.162**
China	0.276, 1.31	0.352, 0.92	0.638, 0.28	0.562, 0.39	0.357, 0.90	—

†NS, non significant; **$0.001 < P < 0.01$; ***$P < 0.001$.
‡with F_{ST} values of 0, *Nm* tends to infinite.

Table 6 Genetic differentiation and gene flow estimates between the three main biogeographic regions found for *Rhizobium gallicum sensu lato* (the sample size is indicated in parenthesis) based on *rrs* and *glnII+atpD* sequences. Above the diagonal are the K_{ST}* values†. Below the diagonal are the F_{ST} values and their associated *Nm* values

rrs	Mexico (11)	Mediterranean (20)	China (10)
Mexico	—	0.088*	0.054*
Mediterranean	0.100, 4.50	—	0.063*
China	0.144, 2.98	0.112, 3.95	—
glnII+atpD	Mexico (15)	Mediterranean (26)	China (10)
Mexico	—	0.154***	0.095***
Mediterranean	0.279, 1.29	—	0.171***
China	0.276, 1.31	0.399, 0.75	—

†*$0.01 < P < 0.05$, ***$P < 0.001$.

the Mexican, Mediterranean (Spain, Austria, France and Tunisia) and Chinese regions for the *rrs* and *glnII+atpD* sequences are shown in Table 6. A sharp genetic differentiation was found between these three biogeographic regions for the protein-coding genes, in contrast with the *rrs* gene, for which low genetic differentiation values and high migration rates were recorded. This differentiation is reinforced by the results of the analyses of molecular variance (AMOVA) performed to test the structure of these three biogeographic regions. Both the rRNA and the protein-coding genes coincide in showing that the major part of the genetic variance is found within populations (86 and 68%, respectively), probably reflecting the enormous genetic diversity of the populations present at each sampling site. However, the *glnII+atpD* sequences revealed that one-third (28%) of the genetic variation is explained by differences among the three biogeographic regions, whereas the *rrs* sequences assigned less than 8% of the genetic variation to differences among regions. Both analyses coincide in showing that the variance among populations within regions accounts for the smallest part of the variance (7% and 4%, respectively), providing evidence that these biogeographic regions correspond with real biological discontinuities. The Mantel tests performed for the pairwise comparisons of the six populations based on the *rrs* sequences did not detect a significant correlation between genetic and geographic distances ($r = 0.223$, $P = 0.286$), whereas the analysis based on the *glnII+atpD* sequences showed a strong isolation-by-distance effect ($r = 0.733$, $P = 0.007$). It is noteworthy that the Mantel test for the pairwise comparisons of the four Mediterranean populations (Spain, Austria, France and Tunisia) revealed a nonsignificant geographic isolation ($r = 0.527$, $P = 0.123$), supporting our previous conclusion that enough gene flow exists among these populations to consider them as a single population.

Finally, we performed genetic analyses based on the chromosomal genes to ask if there was significant differentiation among the individuals of the different biovars and to establish if they can be considered compartments within *R. gallicum s. l.* lineage. When the three biovars where analysed together, significant differentiation was found for the *rrs*, *glnII*, *atpD* and *glnII+atpD* sequences (data not shown). This was expected since the biovar *orientale* was represented only in the Chinese population, which was significantly differentiated from the other populations (see Tables 5 and 6). However, when only the strains from the biovars *gallicum* and *phaseoli* were compared, the detected differentiation was not significant (data not shown), suggesting that in the places where these two biovars are sympatric they do not represent genetic compartments or subdivided populations of the chromosomal genetic pool of *R. gallicum*. This lack of separation is not surprising since the symbiotic haplotypes of these two biovars were found in strains with the same chromosomal background, for example, the Group 2 strains (compare Figs 1– 4).

Discussion

In this study we have shown the power of combining population genetics and phylogenetic approaches to address several interrelated issues in bacterial evolutionary ecology and systematics, such as bacterial species delineation, the uncovering of the evolutionary forces that shape their genetic structures, and the discovery of population subdivisions. The robustness and resolving power of such inferences are maximized when large strain collections are analysed using sequences from multiple loci (Palys *et al*. 1997; Whitaker *et al*. 2003; Vinuesa & Silva 2004; Vinuesa *et al*. 2005c). In this context, we have targeted several housekeeping and symbiotic loci in order to disentangle the different evolutionary histories of so-called core and auxiliary loci that make up the mosaic genomes of many bacterial groups such as enterics and rhizobia (Wernegreen & Riley 1999; Lan & Reeves 2000; Welch *et al*. 2002; Romero & Brom 2004).

Proper recognition of the different evolutionary patterns exhibited by core and auxiliary loci is fundamental to an adequate molecular systematic interpretation of the phenotypes exhibited by different ecotypes of a particular species (Lan & Reeves 2001; Vinuesa *et al*. 2005a). The importance of this is clearly illustrated in our study, where we have provided compelling evidence that demonstrates: (i) that *Rhizobium gallicum*, *Rhizobium mongolense* and *Rhizobium yanglingense* are members of the same evolutionary lineage; (ii) that the *R. mongolense* and *R. yanglingense* strains constitute a new biovar, and (iii) that the *R. gallicum s. l.* lineage contains two genetic groups (Group 1 and Group 2) that coexist sympatrically in distinct geographic regions

and share the ecological niche and symbiotic pool of the species.

The synonymy of *R. gallicum*, *R. mongolense* and *R. yanglingense* was suggested previously based on phylogenetic analyses of *rrs* (Young *et al*. 2004) and *atpD+recA* sequences (Vinuesa *et al*. 2005b). The synonymy is confirmed in this study through phylogenetic analyses involving several isolates of most of these named species. The existence of two divergent chromosomal groups in *R. gallicum* also had been suggested previously on the basis of *rrs* PCR-RFLPs (PCR-restriction fragment length polymorphism) and MLEE profiles (Herrera-Cervera *et al*. 1999; Mhamdi *et al*. 1999), and those observations are amply corroborated by our sequence-based analyses. Similarly, Laguerre *et al*. (2001) previously noted on the basis of PCR-RFLPs that the type strain of *R. mongolense* contains *nifH* and *nodC* alleles closely related to those of other *Medicago* symbionts. Our phylogenetic analyses of *nifH* and *nodB* sequences for multiple isolates confirmed this point, and allowed us to recognize that *R. mongolense* and *R. yanglingense* strains represent a new symbiotic ecotype of *R. gallicum*.

Sequencing three unlinked housekeeping loci provided sufficient DNA polymorphisms to make inferences about the genetic structure of populations, and to uncover key evolutionary processes acting on them at different geographic scales. At a local scale, nodulation is frequently dominated by a few genotypes (Silva *et al*. 1999; McInnes *et al*. 2004; Vinuesa *et al*. 2005c). Therefore, in order to be able to uncover the effects of recombination in an epidemic population, it is important to perform population analyses using only the distinct genotypes (Maynard Smith *et al*. 1993; Silva *et al*. 1999). The recombination estimates obtained using all sequences, or only distinct haplotypes, are consistent with an epidemic genetic structure for the *R. gallicum s. l.* population. This finding is in agreement with the results from linkage disequilibrium analyses based on MLEE data obtained for the Mexican (San Miguel) population (Silva *et al*. 1999, 2003). Both population genetic and phylogenetic analyses revealed that the *rrs* locus exhibited the highest levels of recombination. This present report augments the cumulative evidence of homologous recombination at the *rrs* locus in rhizobia and other bacterial species (e.g. Eardly *et al*. 1995, 1996; Yap *et al*. 1999; Schouls *et al*. 2003; van Berkum *et al*. 2003; Vinuesa *et al*. 2005b). Recombination at this locus explains at least in part the topological instability of *rrs* phylogenies with respect to the sequenced region, which is also influenced by taxon and model choice. In addition, phylogenetic incongruences found between the groupings established by the chromosomal and symbiotic loci uncovered likely events of the lateral transfer of symbiotic information across chromosomal backgrounds within and among species.

Only the chromosomal protein-coding genes showed a pattern of biogeographic differentiation; a biogeographic pattern was not revealed by the *rrs* gene. The difference in these results could be explained by the combined effects of higher evolutionary rates exhibited by the protein-coding genes, by stronger regional selective and/or random drift effects acting on the protein-coding genes, and lower recombination levels compared with those found for the rRNA genes. These results demonstrate that the *atpD* and *glnII* genes analysed here contain valuable information about biogeographic processes. Therefore, protein-coding genes should be chosen as the preferred molecular markers for evolutionary, biogeographic and systematic studies, as shown in other studies (Palys *et al*. 1997; Lan & Reeves 2000, 2001; Whitaker *et al*. 2003; Zeigler 2003; Vinuesa & Silva 2004; Vinuesa *et al*. 2005a, c).

Gene flow estimates showed that migration has a significant impact on the *R. gallicum s. l.* population. We determined that most of the observed genetic variation is found within sampling sites. This pattern accords with that of other studies (Souza *et al*. 1994; Hagen & Hamrick 1996; Silva *et al*. 1999), and demonstrates that enormous genetic diversity exists at a local scale. At higher geographic scales, the general trend is to record low levels of genetic differentiation (Strain *et al*. 1995; Hagen & Hamrick 1996; Vinuesa *et al*. 2005c), which led Hagen & Hamrick (1996) to propose that gene flow and founder effects act differentially at geographic and local scales. We found that only very distantly separated populations (> 10 000 km) showed significant genetic differentiation, indicating that very large geographic distances are required to function as effective barriers to gene flow in rhizobial populations. Populations sampled in the same region, such as the four Mediterranean populations (separated by ~1000–2000 km), did not show evidence of geographic subdivision. The general trend of low genetic differentiation at local scales and high differentiation at larger geographic scales is in concordance with patterns found for populations of hyperthermophilic archaea *Sulfolobus* (Whitaker *et al*. 2003). Despite the significant genetic differentiation estimates among the Mexican and Mediterranean populations, some haplotypes were shared by both populations, suggesting that they were either derived from a very recent common ancestor, or that migration may sometimes overcome large distances. Little is known about the dispersal mechanisms of rhizobia; both anthropogenic and natural mechanisms may act with varying intensities. Humans can disperse rhizobia either intentionally by deliberated inoculation of crop legumes or unintentionally, by the transportation of legume seeds along with their associated rhizobia to very distant regions (Pérez-Ramírez *et al*. 1998). This is certainly true in the case of the common bean (Martínez-Romero 2003). Running water and wind-driven dust could account for the natural dispersal of soil bacteria (Griffin *et al*. 2002; Vinuesa & Silva 2004; Vinuesa *et al*. 2005c). We think that both anthropogenic and natural

mechanisms of dispersal might have a role in the wide distribution of *R. gallicum s. l.*

The accepted notion for rhizobial biogeography is that the symbiotic bacteria co-evolved with their hosts in the centres of origin and diversification of the legumes (Martínez-Romero & Caballero-Mellado 1996). However, molecular clock analyses have shown that the divergence times of the rhizobial genera pre-date the existence of legumes (Turner & Young 2000). Several lines of evidence support the hypothesis that nodulation is a relatively recent acquisition of rhizobia that occurred later than legume emergence (Young & Johnston 1989; Wernegreen & Riley 1999; Laguerre *et al*. 2001). Here we expand the co-evolution hypothesis by pointing out that it might be true for symbiotic genes and ecotypes (biovars), but not necessarily for the housekeeping genes and species as a whole. This suggestion is supported by the results of the study of *Bradyrhizobium* sp. from North America, Central America, Asia and Austria, which revealed that the 16S rDNA haplotypes were randomly distributed whereas the *nifD* haplotypes were distributed according to the geographic region (Parker *et al*. 2002). It has been speculated that *R. gallicum* could be of American origin (Sessitsch *et al*. 1997b; Zurdo-Piñeiro *et al*. 2004) or, alternatively, that it could have a wide geographic distribution and a long evolutionary history of adaptation to different environments and leguminous hosts (Amarger *et al*. 1997; Silva *et al*. 2003). Taken together, the evidence presented here based on phylogenetic and population genetics inferences along with the wide spectrum of legume hosts nodulated by the *R. gallicum s. l.* lineage, suggests that the cosmopolitan distribution hypothesis is more likely. The contrasting biogeographic patterns revealed by the slowly evolving *rrs* and faster-evolving protein-coding genes may be taken as evidence for an ancient panglobal lineage, in which migration and recombination occurred extensively until the continental masses drifted apart, giving rise to the actual pattern of significant levels of gene flow among proximal (i.e. regional) populations, but not across the global population.

The analyses of two plasmid-borne symbiotic genes demonstrated the existence of three symbiotic ecotypes within *R. gallicum s. l.* the previously described biovars *gallicum* and *phaseoli* (Amarger *et al*. 1997), and the newly discovered biovar *orientale*. The American origin of the biovar *phaseoli* is well supported by several studies (Laguerre *et al*. 2001; Brom *et al*. 2002; Martínez-Romero 2003; Aguilar *et al*. 2004). The symbiotic alleles found in the biovar *gallicum* strains seem to be the ones that co-evolved with the *R. gallicum* lineage, as they have a wide geographic distribution and confer a broad host range. We propose the *R. mongolense* and *R. yanglingense* strains be reclassified as *R. gallicum* bv. *orientale*. Since this biovar was found only among Chinese strains, and due to the fact that no chromosomal haplotypes of these isolates were shared with those sampled from other populations, it was not possible to disentangle the evolutionary effects of geographic isolation from those related to host selection or specificity.

The description of rhizobial species based on the delineation of chromosomal evolutionary lineages and on the assignation of biovarieties based on symbiotic features provides a sound biological framework for rhizobial systematics (Graham *et al*. 1991; Amarger *et al*. 1997; Wang *et al*. 1999; Laguerre *et al*. 2001; Vinuesa *et al*. 2005a, c). This approach helps avoid the joining of different chromosomal evolutionary lineages that share a particular symbiotic genetic pool (probably by lateral transfer) and host range, or the incorrect separation of individuals of the same species into different taxa based on distinct host ranges, which result from having different symbiotic alleles (probably also as the result of lateral transfer). An example for the latter situation is provided in the present work, in which members of the *R. gallicum* evolutionary lineage had been named as two new species (*R. mongolense* and *R. yanglingense*) based on the standard taxonomic practice (Stackebrandt *et al*. 2002), mainly DNA : DNA hybridization values lower than 70% between type strains, phenotypic features and different hosts of isolation (van Berkum *et al*. 1998; Tan *et al*. 2001). A similar situation is found for several pathogenic bacteria, for which new species or even genera were described based on biochemical traits or pathological symptoms, and which were later shown to represent clones or ecotypes within species by phylogenetic analyses of multiple loci (Lan & Reeves 2001).

It is important to recognize ecotypes as components of species' diversity. Ecotypes structure populations at the ecological scale, and in evolutionary time they bear the diversity of the species. Nonetheless, we believe that species (which may contain multiple ecotypes) are the fundamental evolutionary units. In this study, we show that strains with monophyletic chromosomal backgrounds can harbour divergent symbiotic genotypes and conversely, the same symbiotic genotype can be found in divergent chromosomal backgrounds. This result points out the transient nature of the symbiotic ecotypes, and it exemplifies how a species can have different ecotypes without loosing its cohesiveness. For this reason, we do not share the view that bacterial ecotypes are the biological units that exhibit the quintessential features of species, as proposed by Cohan (2002). Ecotypes should be recognized taxonomically at a subspecific rank (subspecies, biovars, pathovars, clones, etc.) instead of creating new species for them (Lan & Reeves 2001; Vinuesa *et al*. 2005a).

Bacterial species delineation ideally should be based on synapomorphies encoded by single copy, neutrally evolving housekeeping genes (Lan & Reeves 2000; Vinuesa *et al*. 2005c); however, housekeeping genes are not immune to the effects of evolutionary forces, such as selection and recombination (Lawrence & Hendrickson 2003). The use of multiple loci and individuals from different demes is

currently the best strategy for detecting those forces and determining their impact on structuring the genetic diversity of species (Istock *et al*. 1996; Hanage *et al*. 2005; Vinuesa *et al*. 2005c). The fact that recombination and lateral transfer events may affect some loci of certain individuals in a large population sample does not obliterate the existence of species in the bacterial world. We think that it is very useful to adopt an evolutionary species concept in bacteriology that incorporates the growing body of knowledge about bacterial ecology and evolution (Istock *et al*. 1996; Ward 1998; Vinuesa *et al*. 2005c). We stress the importance of incorporating 'population thinking' (Mayr 1970) into bacterial systematics and of avoiding the establishment of arbitrary similarity cut-off values for species delineation, since for each species the outcomes of historical contingencies, genetic and ecological processes, and the relative strengths of the evolutionary forces can differ (Mayr 1970; Istock *et al*. 1996; Ward 1998; Vinuesa *et al*. 2005c). We agree with the proposal of Mayden (1997) that the evolutionary species concept should be used as the primary concept and should be supported by other concepts (e.g. biological, phylogenetic, ecological, etc.) to make it operational (Mayden 1997). The application of this approach would place bacterial taxonomy more fully within the mainstream of current biological molecular systematics and would avoid the need for creating ad hoc species definitions for prokaryotes (Stackebrandt *et al*. 2002).

Acknowledgements

We are grateful to Noelle Amarger, Bertrand Eardly, Danielle Prevoist, Ridha Mhamdi, Juan San Juan, Angela Sessitsch, Peter van Berkum and En Tao Wang for kindly providing strains. We thank Marco A. Rogel and Aldo Valera for technical assistance, Brandon and Rebeca Gaunt for access to the sequencing facilities at University of California-Irvine, and Scott Bringham for sequencing services at Arizona State University. Paul V. Dunlap and three anonymous reviewers are gratefully acknowledged for the critical reading of the manuscript. This research was partially supported by SEP-CONACyT grant 44673-Q to V. S. and by the budget from Universidad Nacional Autónoma de México to E. M.-R.

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2721/MEC2721sm.htm

Table S1 Primers and annealing temperatures used in this study

Table S2 Accession numbers for the sequences used in this study. The sequences generated in this study are highlighted in boldface

References

Aguilar OM, Riva O, Peltzer E (2004) Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. *Proceedings of the National Academy of Sciences, USA*, **101**, 13548–13553.

Amarger N, Macheret V, Laguerre G (1997) *Rhizobium gallicum* sp. nov. and *Rhizobium giardinii* sp. nov., from *Phaseolus vulgaris* nodules. *International Journal of Systematic Bacteriology*, **47**, 996–1006.

Brom S, Girard L, García-de los Santos A *et al*. (2002) Conservation of plasmid-encoded traits among bean-nodulating *Rhizobium* species. *Applied and Environmental Microbiology*, **68**, 2555–2561.

Chen WM, Moulin L, Bontemps C, *et al*. (2003) Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. *Journal of Bacteriology*, **185**, 7266–7272.

Cohan FM (2002) What are bacterial species? *Annual Review of Microbiology*, **56**, 457–487.

Eardly BD, Wang FS, Whittam TS, Selander RK (1995) Species limits in *Rhizobium* populations that nodulate the common bean (*Phaseolus vulgaris*). *Applied and Environmental Microbiology*, **61**, 507–512.

Eardly BD, Wang F-S, van Berkum P (1996) Corresponding 16S rRNA gene segments in Rhizobiaceae and *Aeromonas* yield discordant phylogenies. *Plant Soil*, **186**, 69–74.

Excoffier L, Somouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

García de los Santos A, Brom S, Romero D (1996) *Rhizobium* plasmids in bacteria–legume interactions. *World Journal of Microbiological Biotechnology*, **12**, 119–125.

Gaunt MW, Turner SL, Rigottier-Gois L, Lloyd-Macgilp SA, Young JPW (2001) Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *International Journal of Systematic and Evolutionary Microbiology*, **51**, 2037–2048.

Graham PH, Sadowsky MJ, Keyser HH *et al*. (1991) Proposed minimal standards for the description of new genera and species of root- and stem-nodulating bacteria. *International Journal of Systematic and Evolutionary Microbiology*, **41**, 582–587.

Griffin DW, Kellogg CA, Garrison VH, Shinn EA (2002) The global transport of dust—an intercontinental river of dust, microorganisms and toxic chemicals flows through the Earth's atmosphere. *American Scientist*, **90**, 228–235.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hagen MJ, Hamrick JL (1996) A hierarchical analysis of population genetic structure in *Rhizobium leguminosarum* bv. *trifolii*. *Molecular Ecology*, **5**, 177–186.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hanage WP, Fraser C, Spratt BG (2005) Fuzzy species among recombinogenic bacteria. *BMC Biology*, **3**, 6.

Herrera-Cervera JA, Caballero-Mellado J, Laguerre G *et al*. (1999) At least five rhizobial species nodulate *Phaseolus vulgaris* in a Spanish soil. *FEMS Microbiology Ecology*, **30**, 87–97.

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. *Genetics*, **145**, 833–846.

Hudson RR (1987) Estimating the recombination parameter of a finite population model without selection. *Genetical Research*, **50**, 245–250.

Hudson RR, Boos DD, Kaplan NL (1992a) A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution*, **9**, 138–151.

Hudson RR, Slatkin M, Maddison WP (1992b) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.

Huelsenbeck JP, Rannala B (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, **276**, 227–232.

Istock CA, Bell JA, Ferguson N, Istock NL (1996) Bacterial species and evolution: theoretical and practical perspectives. *Journal of Industrial Microbiology*, **17**, 137–150.

Josephson KL, Paau AS (1984) Competitiveness and effectiveness of strains of *Rhizobium phaseoli* isolated from the Sonora desert. *Soil Biology and Biochemistry*, **16**, 651–655.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244–1245.

Laguerre G, van Berkum P, Amarger N, Prevost D (1997) Genetic diversity of rhizobial symbionts isolated from legume species within the genera *Astragalus*, *Oxytropis*, and *Onobrychis*. *Applied and Environmental Microbiology*, **63**, 4748–4758.

Laguerre G, Nour SM, Macheret V *et al.* (2001) Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology*, **147**, 981–993.

Lan RT, Reeves PR (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology*, **8**, 396–401.

Lan R, Reeves P (2001) When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiology*, **9**, 419–424.

Lawrence JG, Hendrickson H (2003) Lateral gene transfer: when will adolescence end? *Molecular Microbiology*, **50**, 739–749.

Martínez-Romero E (2003) Diversity of *Rhizobium–Phaseolus vulgaris* symbiosis: overview and perspectives. *Plant Soil*, **252**, 11–23.

Martínez-Romero E, Caballero-Mellado J (1996) *Rhizobium* phylogenies and bacterial genetic diversity. *Critical Reviews in Plant Science*, **15**, 113–140.

Martínez-Romero E, Segovia L, Mercante FM *et al.* (1991) *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. beans and *Leucaena* sp. trees. *International Journal of Systematic Bacteriology*, **41**, 417–426.

Mayden RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: the Units of Biodiversity* (eds Claridge MF, Dawah HA, Wilson MR). Chapman & Hall, London.

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences, USA*, **90**, 4384–4388.

Mayr E (1970) *Populations, Species, and Evolution*. Belknap Press of Harvard University, Cambridge, Massachusetts.

McInnes A, Thies JE, Abbot JG, Howeieson JG (2004) Structure and diversity among rhizobial strains, populations and communities — a review. *Soil Biology and Biochemistry*, **36**, 1295–1308.

Mhamdi R, Jebara M, Aouani ME, Ghrir R, Mars M (1999) Genotypic diversity and symbiotic effectiveness of rhizobia isolated from root nodules of *Phaseolus vulgaris* L. grown in Tunisian soils. *Biology and Fertility of Soils*, **28**, 313–320.

Mhamdi R, Laguerre G, Aouani ME, Mars M, Amarger N (2002) Different species and symbiotic genotypes of field rhizobia can nodulate *Phaseolus vulgaris* in Tunisian soils. *FEMS Microbiology Ecology*, **41**, 77–84.

Nylander JA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.

Palys T, Nakamura LK, Cohan FM (1997) Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *International Journal of Systematic Bacteriology*, **47**, 1145–1156.

Parker MA, Lafay B, Burdon JJ, van Berkum P (2002) Conflicting phylogoegraphic patterns in rRNA and *nifD* indicate regionally restricted gene transfer in *Bradyrhizobium*. *Microbiology*, **148**, 2557–2565.

Pérez-Ramírez NO, Rogel MA, Wang E, Castellanos JZ, Martínez-Romero E (1998) Seeds of *Phaseolus vulgaris* bean carry *Rhizobium etli*. *FEMS Microbiology Ecology*, **26**, 289–296.

Piñero D, Martínez E, Selander RK (1988) Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. *Applied and Environmental Microbiology*, **54**, 2825–2832.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Priefer UB (1989) Genes involved in lipopolysaccharide production and symbiosis are clustered on the chromosome of *Rhizobium leguminosarm* biovar *viciae* VF39. *J. Bacteriol*, **171**, 6161–6168.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rodriguez-Navarro DN, Buendia AM, Camacho M, Lucas MM, Santamaria C (2000) Characterization of *Rhizobium* spp. bean isolates from South-West Spain. *Soil Biology and Biochemistry*, **32**, 1601–1613.

Rome S, Brunel B, Normand P, Fernandez M, Cleyet Marel JC (1996) Evidence that two genomic species of *Rhizobium* are associated with *Medicago truncatula*. *Archives Microbiology*, **165**, 285–288.

Romero D, Brom S (2004) The symbiotic plasmids of the *Rhizobiaceae*. In: *Plasmid Biology* (eds Funnell BE, Philips GJ), pp. 271–290. ASM Press, Washington.

Ronquist F, Huelsenbeck JP (2003) MRBAYES3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rozas J, Rozas R (1999) DNASP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Sawada H, Kuykendall LD, Young JM (2003) Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. *Journal of General and Applied Microbiology*, **49**, 155–179.

Schneider S, Roessli D, Excoffier L (2000) *ARLEQUIN: A software for population genetic data analysis*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland (http://anthro.unige.ch/arlequin).

Schouls LM, Schot CS, Jacobs JA (2003) Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *Journal of Bacteriology*, **185**, 7241–7246.

Segovia L, Young JP, Martínez-Romero E (1993) Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *International Journal of Systematic Bacteriology*, **43**, 374–377.

Sessitsch A, Hardarson G, Akkermans ADL, de Vos WM (1997a) Characterization of *Rhizobium etli* and other *Rhizobium* spp. that nodulate *Phaseolus vulgaris* L. in an Austrian soil. *Molecular Ecology*, **6**, 601–608.

Sessitsch A, Ramírez-Saad H, Hardarson G, Akkermans AD, de Vos WM (1997b) Classification of Austrian rhizobia and the

Mexican isolate FL27 obtained from *Phaseolus vulgaris* L. as *Rhizobium gallicum*. *International Journal of Systematic Bacteriology*, **47**, 1097–1101.

Silva C, Eguiarte LE, Souza V (1999) Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Molecular Ecology*, **8**, 277–287.

Silva C, Vinuesa P, Eguiarte LE, Martínez-Romero E, Souza V (2003) *Rhizobium etli* and *Rhizobium gallicum* nodulate common bean (*Phaseolus vulgaris*) in a traditionally managed milpa plot in Mexico: population genetics and biogeographic implications. *Applied and Environmental Microbiology*, **69**, 884–893.

Smouse PE, Long JC, Sokal RR (1986) Mutiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627–632.

Souza V, Eguiarte LE, Avila G *et al.* (1994) Genetic structure of *Rhizobium etli* biovar *phaseoli* associated with wild and cultivated been plants (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. *Applied and Environmental Microbiology*, **60**, 1260–1268.

Spaink HP (2000) Root nodulation and infection factors produced by rhizobial bacteria. *Annual Review of Microbiology*, **54**, 257–288.

Stackebrandt E, Frederiksen W, Garrity GM *et al.* (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, **52**, 1043–1047.

Strain SR, Whittam TS, Bottomley PJ (1995) Analysis of genetic structure in soil populations of *Rhizobium leguminosarum* recovered from the USA and the UK. *Molecular Ecology*, **4**, 105–114.

Swofford DL (2002) PAUP*: *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tan ZY, Kan FL, Peng GX *et al.* (2001) *Rhizobium yanglingense* sp. nov., isolated from arid and semi-arid regions in China. *International Journal of Systematic and Evolutionary Microbiology*, **51**, 909–914.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Turner SL, Young JP (2000) The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Molecular Biology and Evolution*, **17**, 309–319.

van Berkum P, Beyene D, Bao G, Campbell TA, Eardly BD (1998) *Rhizobium mongolense* sp. nov. is one of three rhizobial genotypes identified which nodulate and form nitrogen-fixing symbioses with *Medicago ruthenica* [(L.) Ledebour]. *International Journal of Systematic Bacteriology*, **48**, 13–22.

van Berkum P, Terefework Z, Paulin L *et al.* (2003) Discordant phylogenies within the *rrn* loci of rhizobia. *Journal of Bacteriology*, **185**, 2988–2998.

Vinuesa P, León-Barrios M, Silva C *et al.* (2005a) *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (Papilionoideae: Genisteae) from the Canary Islands, along with *Bradyrhizobium japonicum* bv. *genistearum*, *Bradyrhizobium* genospecies α and *Bradyrhizobium* genospecies β. *International Journal of Systematic and Evolutionary Microbiology*, **55**, 569–575.

Vinuesa P, Silva C (2004) Species delineation and biogeography of symbiotic bacteria associated with cultivated and wild legumes. In: *Biological Resources and Migration* (ed. Werner D), pp. 143–155. Springer Verlag, Berlin.

Vinuesa P, Silva C, Lorite MJ *et al.* (2005b) Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from *rrs*, *atpD*, *recA* and *nifH* sequences, and their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands. *Systematic Applied and Microbiology*, doi: 10.1016/j.syapm.2005.05.007.

Vinuesa P, Silva C, Werner D, Martínez-Romero E (2005c) Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Molecular Phylogenetics and Evolution*, **34**, 29–54.

Wang ET, Rogel MA, García-de los Santos A *et al.* (1999) *Rhizobium etli* bv. *mimosae*, a novel biovar isolated from *Mimosa affinis*. *International Journal of Systematic Bacteriology*, **49**, 1479–1491.

Ward DM (1998) A natural species concept for prokaryotes. *Current Opinion in Microbiology*, **1**, 271–277.

Welch RA, Burland V, Plunkett G 3rd *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA*, **99**, 17020–17024.

Wernegreen J, Riley MA (1999) Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Molecular Biology and Evolution*, **16**, 98–113.

Wernegreen JJ, Harding EE, Riley MA (1997) *Rhizobium* gone native: unexpected plasmid stability of indigenous *Rhizobium leguminosarum*. *Proceedings of the National Academy of Sciences, USA*, **94**, 5483–5488.

Whitaker RJ, Grogan DW, Taylor JW (2003) Geographic barriers isolate endemic populations of hyperthermophilic Archaea. *Science*, **301**, 976–978.

Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, **92**, 371–373.

Yap WH, Zhang Z, Wang Y (1999) Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *Journal of Bacteriology*, **181**, 5201–5209.

Young JPW, Johnston AWB (1989) The evolution of specificity in the legume-*Rhizobium* symbiosis. *Trends in Ecology & Evolution*, **4**, 341–349.

Young JM, Park DC, Weir BS (2004) Diversity of 16S rDNA sequences of *Rhizobium* spp. implications for species determinations. *FEMS Microbiology Letters*, **238**, 125–131.

Zeigler DR (2003) Gene sequences useful for predicting relatedness of whole genomes in bacteria. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 1893–1900.

Zurdo-Piñeiro JL, Velázquez E, Lorite MJ *et al.* (2004) Identification of fast-growing rhizobia nodulating tropical legumes in Puerto Rico as *Rhizobium gallicum* and *Rhizobium tropici*. *Systematic and Applied Microbiology*, **27**, 469–477.

The authors are interested in bacterial diversity, evolutionary genetics and phylogeography. Our recent work has focused on studying the relative importance of different evolutionary forces in structuring bacterial populations and their implications in bacterial systematics and biogeography.
