Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 34 (2005) 29–54
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation

Pablo Vinuesa[a,b,*], Claudia Silva[a], Dietrich Werner[b], Esperanza Martínez-Romero[a]

[a] *Centro de Investigación sobre Fijación de Nitrógeno, Programa de Ecología Molecular y Microbiana, UNAM, Av. Universidad S/N, Col. Chamilpa, AP 565-A, Cuernavaca CP 62210, Morelos, México*

[b] *FB Biologie, FG für Zellbiologie und Angewandte Botanik, Philipps Universität Marburg, Karl von Frisch Str, D-35032 Marburg, Germany*

Received 17 November 2003; revised 19 June 2004

Abstract

A combination of population genetics and phylogenetic inference methods was used to delineate *Bradyrhizobium* species and to uncover the evolutionary forces acting at the population-species interface of this bacterial genus. Maximum-likelihood gene trees for *atpD*, *glnII*, *recA*, and *nifH* loci were estimated for diverse strains from all but one of the named *Bradyrhizobium* species, and three unnamed "genospecies," including photosynthetic isolates. Topological congruence and split decomposition analyses of the three housekeeping loci are consistent with a model of frequent homologous recombination within but not across lineages, whereas strong evidence was found for the consistent lateral gene transfer across lineages of the symbiotic (auxiliary) *nifH* locus, which grouped strains according to their hosts and not by their species assignation. A well resolved Bayesian species phylogeny was estimated from partially congruent *glnII* + *recA* sequences, which is highly consistent with the actual taxonomic scheme of the genus. Population-level analyses of isolates from endemic Canarian genistoid legumes based on REP-PCR genomic fingerprints, allozyme and DNA polymorphism analyses revealed a non-clonal and slightly epidemic population structure for *B. canariense* isolates of Canarian and Moroccan origin, uncovered recombination and migration as significant evolutionary forces providing the species with internal cohesiveness, and demonstrated its significant genetic differentiation from *B. japonicum*, its sister species, despite their sympatry and partially overlapped ecological niches. This finding provides strong evidence for the existence of well delineated species in the bacterial world. The results and approaches used herein are discussed in the context of bacterial species concepts and the evolutionary ecology of (brady)rhizobia.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Bayes factors; Canary islands; Coalescent simulations; Likelihood ratio test; Gene flow; Model selection; Neutrality tests; Partitioned models; Prokaryotes

1. Introduction

The working hypothesis of this study is that speciation in the bacterial world is the process that follows as gene flow diminishes and divergence increases among populations, as the result of a variety of genetic and ecological processes and historical contingencies, as broadly accepted for other asexual microbes and sexually reproducing macroscopic organisms (Avise, 2000; Barraclough and Nee, 2001; Carbone and Kohn, 2001; Cohan, 2001; Templeton, 1989). The "phylophenetic" approach adopted by bacterial taxonomists for species identification and delineation (Rosselló-Mora and Amann, 2001; Stackebrandt et al., 2002; Vandamme et al., 1996) does not capture this process adequately, as it relies on similarity criteria (i.e., DNA homology values $\geqslant$60–70%, $\leqslant$3% *rrs*

* Corresponding author. Fax: +52 777 3175581.
E-mail address: vinuesa@cifn.unam.mx (P. Vinuesa).

1055-7903/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2004.08.020

sequence divergence, and phenetic clustering), which cannot uncover the evolutionary forces that provide species with internal cohesiveness, and shape their population genetic structures. Based on the evidence accumulated during the last decade on the contrasting population genetic structures exhibited by different bacterial groups, which range from strictly clonal to essentially panmictic (Feil and Spratt, 2001; Istock et al., 1992; Maynard-Smith et al., 1993; Souza et al., 1992), the astronomic effective population sizes of many free-living prokaryotes, and the extraordinary diversity of prokaryotic lifestyles and ecological niches they occupy (DeLong and Pace, 2001; Torsvik et al., 2002), we postulate that an interplay of diverse cohesion forces (such as recombination, migration, and selective sweeps) maintain species as recognizable evolutionary lineages, and that these will differ among functional groups of prokaryotes and genomic regions of individual strains.

This study was designed to gain insights into the evolutionary processes operating at the population-species interface of the bacterial genus *Bradyrhizobium* by using a combination of population genetics and phylogenetic analyses (Lan and Reeves, 2001). This approach should more properly capture the ecological and evolutionary aspects of speciation in this cosmopolitan and diverse bacterial group, than that used in traditional rhizobial taxonomy, and systematics (Sawada et al., 2003; Willems et al., 2001a,b). Based on 16S rRNA gene (*rrs*) sequences, the genus *Bradyrhizobium* forms a clade in the α-2 subclass of the *Proteobacteria* along with oligotrophic soil or aquatic bacteria such as *Rhodopseudomonas palustris*, *Rhodoplanes roseus*, *Nitrobacter winogradskyi*, *Blastobacter denitrificans*, and the pathogen *Afipia* spp. (Saito et al., 1998; Sawada et al., 2003; van Berkum and Eardly, 2002; Willems et al., 2001b). Symbiotic *Bradyrhizobium* strains have been isolated from the nodules of highly divergent legume tribes, both from herbaceous and woody species, of tropical and temperate origin, including aquatic legumes such as *Aeschynomene* species, and the non-legume *Parasponia andersonii* (Sprent, 2001). Bradyrhizobia have also been found as endophytes of several rice species (Chaintreuil et al., 2000; Tan et al., 2001). Probably the most outstanding aspect of the evolutionary ecology of bradyrhizobia is that they cycle through highly contrasting lifestyles, as members of the communities of soil or aquatic oligotrophic bacteria, as rhizospheric or endophytic bacteria, or as N_2-fixing endosymbionts of diverse host legume species. Bradyrhizobia are metabolically diverse, some strains being capable of performing denitrification or derepressing nitrogenase activity under microaerobic free-living conditions, and yet others are photosynthetic or degrade xenobiotics such as 2,4-dichlorophenoxyacetate (2,4-D) or halobenzoates (Chaintreuil et al., 2000; Kitagawa et al., 2002; Kurz and LaRue, 1975; Mesa et al., 2002; Saito et al., 1998; So et al., 1994).

Despite this impressive diversity, only six species have been validly described so far: *Bradyrhizobium japonicum* (Jordan, 1982), *Bradyrhizobium elkanii* (Kuykendall et al., 1992), *Bradyrhizobium liaoningense* (Xu et al., 1995), *Bradyrhizobium yuanmingense* (Yao et al., 2002), *Bradryrhizobium betae* (Rivas et al., 2004), and *Bradyrhizobium canariense* (Vinuesa et al., 2004). The first three species were isolated from soybean (*Glycine max*) nodules. *B. yuanmingense* was isolated from *Lespedeza cuneata* in China, *B. betae* from tumor-like root deformations of sugar beet (*Beta vulgaris*) in Northern Spain, and *B. canariense* from diverse legume genera in the tribes Genisteae and Loteae growing naturally in the Canary Islands, Morocco, Spain, the Americas and probably Australia (Jarabo-Lorenzo et al., 2003; Vinuesa et al., 2004).

The phylogenetic relationships among these species are poorly understood, mainly because only *rrs* sequences were analyzed in the reports describing them, the exception being the description of *B. canariense*, which was largely supported by the population genetics and phylogenetic evidence presented herein. *B. japonicum* isolates have been proposed to form at least two subgroups (I and Ia) based on DNA–DNA hybridization values (Hollis et al., 1981). Recently, ITS sequences have been obtained for a diverse set of bradyrhizobia, including soybean isolates from the 17 described serotypes (van Berkum and Fuhrmann, 2000), isolates from diverse tropical and temperate legumes, including many photosynthetic strains, as well as sequences of the genera *Afipia*, *Blastobacter*, *Nitrobacter*, and *Rhodopseudomonas* (Tan et al., 2001; van Berkum and Eardly, 2002; Willems et al., 2001a; Willems et al., 2003). These studies, coupled with extensive DNA–DNA hybridizations led Willems et al. (2003) to propose seven new genospecies. However, ITS sequences do not properly resolve the evolutionary relationships among them (Vinuesa et al., 2004). Notable is the recent use of *dnaK* sequences for phylogenetic analyses of bradyrhizobia (Moulin et al., 2004), which corroborated rDNA-based groupings, but did not resolve the relationships among them.

In this work, we generated a high number of *atpD*, *glnII*, *recA*, and *nifH* sequences of all the major *Bradyrhizobium* lineages (see Table S1, provided as supplementary material), and used maximum likelihood (ML) and Bayesian frameworks to test diverse phylogenetic hypotheses, to infer a highly resolved species phylogeny for the genus, and to uncover the different evolutionary histories of housekeeping and symbiotic loci. These analyses revealed that at least four *Bradyrhizobium* lineages nodulate diverse endemic Canarian genistoid legumes (ECGLs) such as *Adenocarpus*, *Chamaecytisus*, *Lupinus*, *Spartocytisus*, and *Teline* species

(brooms) in the Canary Islands (Fig. 1). These plants belong to the Northern Hemisphere tribe Genisteae, which includes lupins (*Lupinus* spp.), and represent a highly supported clade within the Papilionoideae subfamily by morphological and molecular sequence analyses (Käss and Wink, 1997). ECGLs are dominant members of several plant communities found in Canarian ecosystems (Francisco-Ortega et al., 1994), and previous studies have shown that they are nodulated exclusively by *Bradyrhizobium* strains (Jarabo-Lorenzo et al., 2000; Vinuesa et al., 1998, 1999). We show that one of the lineages is *B. japonicum*, whereas the other three correspond to *B. canariense*, *Bradyrhizobium* genospecies α and genospecies β (Vinuesa et al., 2004). REP-PCR genomic fingerprinting was used to analyze the genetic diversity of a collection of 53 ECGL isolates recovered from different localities in Morocco and the Canary Islands at the clone or strain level of taxonomic resolution, whereas PCR-RFLP analysis of *rrs* amplicons was used to group clones at higher taxonomic levels (Vinuesa et al., 1998, 1999). Population genetics analyses of continental and insular populations of *B. canariense* based on multilocus enzyme electrophoresis and *atpD*, *glnII*, and *recA* sequence polymorphisms revealed that recombination and migration are key evolutionary forces providing cohesion to this species. The results and analytical approaches used herein are interpreted and discussed in the framework of prokaryotic evolutionary and population genetic theory, with particular reference to the contrasting conceptual views of bacterial species expressed by microbial taxonomists and evolutionary biologists (Cohan, 2002; Lan and Reeves, 2000, 2001; Lawrence, 2002; Rosselló-Mora and Amann, 2001; Stackebrandt et al., 2002; Ward, 1998).



Fig. 1. Geographic location of the Canary Islands and sampling sites of rhizospheric soils used to trap *Chamecytisus proliferus* microsymbionts. Sites I and II are located in Morocco, at the Maâmora forest and Mehdiya-Kénitra, respectively. Sites III and IV are located in Garafía and El Paso, in northern, and central La Palma, respectively. The location of other sampling sites is also highlighted by arrows. The geographic origin and host of each isolate is listed in Table 1.

2. Materials and methods

2.1. Strategies for sampling representative species and strain diversity for phylogenetic and population genetics analyses

A critical aspect of studies dealing with population-species interfaces is the use of an adequate taxon and population (strain, i.e., individual) sampling (Barraclough and Nee, 2001; Maynard-Smith et al., 1993; Rosenberg, 2002). For above species phylogenetic comparisons, strain selection was based on published work and the large dataset we have of combined PCR-RFLPs from *rrs*, ITS and *rrl* amplicons of a highly diverse collection of *Bradyrhizobium* strains from distant geographic regions and contrasting hosts (Vinuesa, unpublished). This ensured the selection of at least two or three highly diverse strains from potentially different demes, which could be considered as representative of the genetic diversity found within named species, and unnamed genospecies. The sources, hosts and geographic locations of the selected strains are listed in Table 1.

The population-level sampling and characterization was greatly aided by the use of REP-PCR genomic fingerprints (Cho and Tiedje, 2000; Judd et al., 1993; Rademaker et al., 2000; Vinuesa et al., 1998), which are known to have a high resolving power, allowing the rapid identification of clonemates, and distinct strains isolated from endemic Canarian genistoid legumes (ECGLs). This ability is of critical importance for population genetic analyses of host associated bacteria (mutualists or pathogens), as only particularly virulent or competitive strains, which generally represent a tiny fraction of the total population, are usually capable of invading the host niche, where they are overrepresented (Feil and Spratt, 2001; Maynard-Smith et al., 1993; Segovia et al., 1991; Sullivan et al., 1995, 1996). Sampling only such "epidemic" strains can lead to the erroneous conclusion that the species to which they belong is highly clonal, when it may be panmictic (Maynard-Smith et al., 1993; Silva et al., 1999).

2.2. Isolation of symbiotic Bradyrhizobium strains from endemic Canarian genistoid legumes

The sources, hosts and other characteristics of the *Bradyrhizobium* isolates from endemic Canarian genistoid legumes (ECGLs) used in this study for population genetic analyses are listed in Table 1. These isolates were obtained either from field sampled nodules reported in previous works (Jarabo-Lorenzo et al., 2000; Vinuesa

Table 1
Bradyrhizobium isolates and reference strains used in this study

Strain[a]	RFLP cluster-genotype[b]	Host[c]	Origin[d]	Source/reference
Isolates from genistoid legumes				
#*Bradyrhizobium* genosp. α BC-C1	III-AAB	*C. proliferus* (1)	G. Canaria	Vinuesa et al. (1998)
#*B. canariense* BC-C2	II-BBA	*C. proliferus* (1)	G. Canaria	Vinuesa et al. (1998)
#*B. canariense* BC-P1	II-BBA	*C. proliferus* (1)	La Palma	This study
#*B. canariense* BC-P5	II-BBA	*C. proliferus* (1)	La Palma	Vinuesa et al. (1998)
#*Bradyrhizobium* genosp. β BC-P6	I-AAA	*C. proliferus* (1)	La Palma	Vinuesa et al. (1998)
#*B. japonicum* BC-P7	I-AAA	*C. proliferus* (1)	La Palma	Vinuesa et al. (1998)
#*B. canariense* BC-P8	II-BBA	*C. proliferus* (1)	La Palma (IV)	This study
#*B. canariense* BC-P9	II-BBA	*C. proliferus* (1)	La Palma (IV)	This study
#*B. canariense* BC-P10	II-BBA	*C. proliferus* (1)	La Palma (IV)	This study
#*B. canariense* BC-P11	II-BBA	*C. proliferus* (1)	La Palma (IV)	This study
#*B. canariense* BC-P12	II-BBA	*C. proliferus* (1)	La Palma (IV)	This study
#*B. japonicum* BC-P14	I-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. japonicum* BC-P15	I-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. japonicum* BC-P16	I-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. japonicum* BC-P17	I-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. japonicum* BC-P18	I-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. japonicum* BC-P20	I-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. canariense* BC-P22	II-BBA	*C. proliferus* (1)	La Palma (III)	This study
#*B. canariense* BC-P23	II-BBA	*C. proliferus* (1)	La Palma (III)	This study
#*B. canariense* BC-P24	II-BBA	*C. proliferus* (1)	La Palma (III)	This study
#*B. japonicum* BC-P25	II-AAA	*C. proliferus* (1)	La Palma (III)	This study
#*B. canariense* BCO-1	II-BBA	*Ad. foliolosus* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BES-1	II-BBA	*C. proliferus* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BES-2	II-BBA	*C. proliferus* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. japonicum* BGA-1	I-AAA	*T. stenopetala* (1)	La Palma	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BGA-2	II-BBA	*T. stenopetala* (1)	La Palma	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BGA-3	II-BBA	*T. stenopetala* (1)	La Palma	Jarabo-Lorenzo et al. (2000)
#*Bradyrhizobium* genosp. β BRE-1	I-AAA	*T. canariensis* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BRE-4	I-BBA	*T. canariensis* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BTA-1^T	II-BBA	*C. proliferus* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BTA-2	II-BBA	*C. proliferus* (1)	La Palma	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BRT-2	II-BBA	*Sp. supranubius* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BRT-5	II-BBA	*Sp. supranubius* (1)	Tenerife	Jarabo-Lorenzo et al. (2000)
#*B. canariense* BC-MAM1	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM2	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM3	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM4	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM5	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM6	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM7	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM8	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM9	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM10	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM11	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*B. canariense* BC-MAM12	II-BBA	*C. proliferus* (1)	Maamora (I)	This study
#*Bradyrhizobium* genosp. β BC-MK1	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK2	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK3	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK4	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK5	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
#*Bradyrhizobium* genosp. β BC-MK6	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK7	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK8	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK9	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK10	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
BC-MK11	I-AAA	*C. proliferus* (1)	M.-Kénitra (II)	This study
#*B. canariense* ISLU16	II-BBA	*Lupinus sp.* (1)	Spain	Jarabo-Lorenzo et al. (2000)
#*B. japonicum* Blup-MR1	I-AAA	*L. polyphillus* (1)	Germany	This study
#*B. japonicum* FN13	I-AAA	*L. montanus* (1)	Mexico	Barrera et al. (1997)
Bradyrhizobium sp. CICS70	I-AAA	*L. montanus* (1)	Mexico	Barrera et al. (1997)

Table 1 (*continued*)

Strain[a]	RFLP cluster-genotype[b]	Host[c]	Origin[d]	Source/reference
Reference strains (named species)				
B. japonicum DSMZ30131^T	I-ABA	*G. max* (3)	Japan	DSMZ
B. japonicum USDA110	I-AAA	*G. max* (3)	USA	USDA
B. japonicum USDA122	I-AAA	*G. max* (3)	USA	USDA
#*B. japonicum* X1-3	I-AAA	*G. max* (3)	China	H. Blasum/this study
#*B. japonicum* X3-1	I-AAA	*G. max* (3)	China	H. Blasum/this study
#*B. japonicum* X6-9	I-AAA	*G. max* (3)	China	H. Blasum/this study
#*B. japonicum* Nep1		*G. max* (3)	Nepal	This study
B. elkanii USDA76^T	IV-BCC	*G. max* (3)	USA	P. van Berkum
B. elkanii USDA94	IV-BCC	*G. max* (3)	USA	P. van Berkum
B. elkanii USDA46	IV-BCC	*G. max* (3)	USA	P. van Berkum
B. liaoningense LMG18230^T	I-AAA	*G. max* (3)	China	Xu et al. (1995)
#*B. liaoningense* Spr3-7	I-AAA	*Ar. hypogaea* (2)	China	Zhang et al. (1999)
B. yuanmingense CCBAU10071^T	I-AAC	*Le. cuneata* (4)	China	Yao et al. (2002)
#*B. yuanmingense* LMTR28		*P. lunatus* (2)	Peru	E. Ormeño/this study
#*B. yuanmingense* TAL760		*I. hirsuta* (6)	Mexico	So et al. (1994)
#*Bradyrhizobium* genosp. α CIAT3101	III-AAB	*Ce. plumieri* (3)	Colombia	Vinuesa et al. (1998)
Photosynthetic strains				
Bradyrhizobium sp. BTAi1		*Ae. indica* (2)	USA	So et al. (1994)
Bradyrhizobium sp. IRBG127		*Ae. pratensis* (2)	Philippines	So et al. (1994)
Bradyrhizobium sp. IRBG231		*Ae. denticulata* (2)	Philippines	So et al. (1994)

[a] # Isolates classified in this study.

[b] Each letter refers to a restriction pattern obtained with enzymes *Cfo*I, *Dde*I, and *Msp*I, respectively. Restriction sites for each enzyme were mapped on the full-length *rrs* sequence of *B. japonicum* USDA 110 (Accession No. Z35330), to which an AAA pattern was assigned. The type B pattern for *Cfo*I contained an additional restriction site at nucleotide 351. The *Dde*I type B pattern contained an additional restriction site at nucleotide 589, and type C pattern resulted from an additional site at position 138. For *Msp*I digestions, the type B pattern resulted from the loss of site 951 and the gain of two new ones at positions 944 and 984, and the type C pattern resulted from two additional sites at positions 604 and 690. Roman numerals (I–IV) correspond to the clusters recovered in >60% of bootstrap pseudoreplicates of a Dice/UPGMA analysis of the combined RFLP patterns.

[c] Abbreviations for legume host genera are: *Aeschynomene* (*Ae*), *Adenocarpus* (*Ad*), *Arachis* (*Ar*), *Chamaecytisus* (*C*), *Centrosema* (*Ce*), *Glycine* (*G*), *Indigofera* (*I*), *Lespedeza* (*Le*), *Lupinus* (*L*), *Phaseolus* (*P*), *Sesbania* (*Se*), *Spartocytisus* (*Sp*), and *Teline* (*T*). The tribal classification of these genera (Polhill, 1994) is given in parenthesis after the species name: (1) Genisteae, (2) Aeschynomeneae, (3) Phaseoleae, (4) Desmodieae, (5) Robinieae, and (6) Indigofereae.

[d] The geographic origin (I–IV) of the isolates from endemic Canarian genistoid legumes obtained in this study is shown in Fig. 1. M.-Kénitra stands for Mehdiya-Kénitra.

et al., 1998), or from the nodules induced by bacteria present in four different rhizospheric soil samples on *C. proliferus* ssp. *proliferus* var. *palmensis* (tagasaste) trap plants, cultivated for 8 weeks in the Leonard jar setting described elsewhere (Vinuesa et al., 1998). The soil samples were collected at two sites in Morocco (the Maamora forest and Mehdiya-Kénitra, both close to Rabat) from *Spartium junceum* (Papilionoideae:Genisteae) and *Acacia cyanophylla* (Mimosoideae:Acacieae) stands, respectively, and from *C. proliferus* ssp. *proliferus* var. *palmensis* (tagasaste) rhizosperes collected in La Palma (Garafía and El Paso), Canary Islands. The localization of these sampling sites (I–IV, respectively) is indicated in Fig. 1. Three gram aliquots of air-dried soil were mixed with the sterile perlite–vermiculite substrate used to fill each cultivation unit. Three jars containing two axenically germinated plantlets per jar were used for each site, whereas jars without soil inoculum served as negative nodulation controls. Genomic DNA from purified isolates and reference strains was isolated as described (Vinuesa et al., 1998).

2.3. Repetitive sequence-based genomic fingerprinting with REP primers and pattern analysis

REP-PCR genomic fingerprints were generated with primers REP1R and REP2I, as previously reported (Vinuesa et al., 1998). Comparative analysis of electrophoretic REP patterns was performed with GelCompar*II* V.2 (Applied Maths BVBA, Belgium) using Pearson's product-moment correlation analysis to calculate pairwise similarity coefficients among pattern densitometric profiles (Rademaker et al., 1998). Similarity matrices were clustered using the unweighted pair group method with averages (UPGMA) algorithm (Sneath and Sokal, 1973). Gel normalization, background subtraction and zone definition were performed as previously described (Rademaker et al., 1998; Vinuesa et al., 1998).

2.4. PCR-RFLP analysis of rrs genes

The primers, cycling parameters, and reaction mixture used were those previously reported (Vinuesa et al., 1998). The PCR products were restricted with the endonucleases *Cfo*I, *Dde*I, and *Msp*I (Amersham and New England Biolabs, UK) and the RFLP patterns resolved on 2% (w/v) MetaPhor agarose gels (Biozym, Hess. Oldendorf, Germany), as described (Vinuesa et al., 1998).

2.5. Multilocus enzyme electrophoresis and population genetics analyses

Electrophoresis was performed on starch gels following standard procedures (Selander et al., 1986). Six enzymes were assessed: alcohol dehydrogenase (ADH, EC 1.1.1.1), esterase (EST, EC 3.1.1.1), isocitrate dehydrogenase (IDH, EC 1.1.1.42), indophenol oxydase (IPO, EC 1.15.1.1), malate dehydrogenase (MDH, EC 1.1.1.37) and malic enzyme (ME, EC 1.1.1.40). Distinctive mobility variants of each enzyme were considered alleles at the corresponding locus (Selander et al., 1986). The allele profiles for the six loci were equated with multilocus genotypes (electrophoretic types or ETs). The mean genetic diversity per locus (h), mean genetic diversity for the population (H) and the coefficient of genetic differentiation (G_{ST}) were estimated using the program ETDIV version 2.2, as previously described (Whittam, 1990; Silva et al., 2003). A χ^2 test of independence was performed to test if the G_{ST} values were significantly different from 0, as previously reported (Silva et al., 1999, 2003).

To determine the extent to which populations exhibit non-random associations of alleles between loci, we used a multilocus association index based on the distribution of allelic mismatches between pairs of isolates over all loci (Maynard-Smith et al., 1993; Souza et al., 1992). The ratio of the variance in mismatches observed in a population (V_0) to the expected variance of the corresponding population at linkage equilibrium (random association of alleles, V_e), provides a measure of linkage disequilibrium. If there is linkage equilibrium, then $V_0/V_e = 1$. The significance of the difference between V_0 and V_e was calculated using a Monte Carlo procedure with 1000 iterations, as implemented in the LDV program (Silva et al., 2003; Souza et al., 1992).

2.6. Amplification and sequencing of atpD, glnII, recA, and nifH loci

The mosaic nature of most bacterial genomes was taken into account to guide the choice of loci for sequencing. Bacterial genomes typically contain regions of so-called auxiliary or accessory genes, coding for ecotype-specific adaptive traits, which are intermixed with the broadly distributed core or housekeeping genes (Gogarten et al., 2002; Lan and Reeves, 2000; Ochman et al., 2000). Rhizobial genomes are good examples of such composite genomes (Galibert et al., 2001; Kaneko et al., 2002), in which accessory symbiotic (*sym*) loci containing the nodulation (*nod*, *nol*, and *noe*), and N_2-fixation (*nif*, *fix*) genes confer these bacteria the ability to nodulate and fix N_2 in symbiosis with particular legume hosts (Spaink et al., 1998; Sullivan et al., 1995). Three protein-coding core loci (*atpD*, *glnII*, and *recA*) and a *sym* locus (*nifH*) were selected to expand the corresponding database available for other rhizobial genera (Gaunt et al., 2001; Laguerre et al., 2001; Turner and Young, 2000; Wernegreen and Riley, 1999). Importantly, since these loci are unlinked, they provide independent genealogies from which to infer a species tree (Nichols, 2001; Rosenberg, 2002), and allow to distinguish the effects of different forces and historical contingencies (Daubin et al., 2003; Fu, 1997; Lan and Reeves, 2000; Ramos-Onsins and Rozas, 2002; Tajima, 1989; Wernegreen and Riley, 1999) that may have shaped the genome and population genetic structure of *Bradyrhizobium* lineages.

A 550 bp *atpD* fragment was amplified with primers *atpD* 273f and *atpD* 771r (Gaunt et al., 2001) or *atpD* 255F (GCTSGGCCGCATCMTSAACGTC) and *atpD* 782R (GCCGACACTTCMGAACCNGCCTG) using the same reaction mixture and cycling parameters used for ITS amplifications reported previously (Vinuesa et al., 1998). Partial *glnII* sequences were obtained with primers *glnII* 12F (YAAGCTCGAGTACATYTG GCT) and *glnII* 689R (TGCATGCCSGAGCCGTTC CA), using the same reaction mixture as for the *atpD* amplification experiments, with annealing at 58 °C and 1 min extension time. Partial *recA* fragments were amplified with primers *recA* 41F (TTCGGCAAGGG MTCGRTSATG) and *recA* 640R (ACATSACRCCGA TCTTCATGC), using the same amplification mix and protocol described for *glnII* sequences. Partial *nifH* gene fragments of 370 bp were amplified with the primers and cycling parameters described elsewhere (Ueda et al., 1995), or with primers *nifH* 40F (GGNATCGGCAA GTCSACSAC) and *nifH* 817R (TCRAMCAGCATG TCCTCSAGCTC), using the amplification mix and protocol described for *glnII* sequences. All amplifications were performed with *Taq* polymerase (USB-Amersham). All the new primer pairs designed in this study are notated in the 5′–3′ orientation and the coordinates correspond to the binding sites on the genome sequence of *B. japonicum* USDA110 (Kaneko et al., 2002). Amplification products were purified using the PCR product purification system of Roche, and subjected to cycle sequencing using the same primers as for PCR amplification, with ABI Prism Dye chemistry, and analyzed with an ABI377 automatic sequencer (ABI, Foster City, CA) at the sequencing facilities of the Institute of Bio-

technology of the UNAM in Cuernavaca, Mexico, and the DNA Laboratory at Tempe, AZ, USA. The GenBank accession numbers for all the sequences generated in this study are listed in Table S1, provided as supplementary online material.

2.7. Evolutionary analyses of nucleotide sequence alignments

Multiple nucleotide sequence alignments were generated with ClustalX (Thompson et al., 1997) and edited with BioEdit (Hall, 1999) and DAMBEv4.2.13 (Xia and Xie, 2001). The protein-coding nucleotide sequences were aligned based on the alignment of the encoded products using DAMBE and the BLOSUM62 matrix. The multiple alignments of individual sequence partitions were scanned for intragenic recombination using Bootscanning (Salminen et al., 1995), Chimaera (Posada and Crandall, 2001), GENECONV1.81 (Sawyer, 1998), MaxChi (Maynard-Smith, 1992) and RDP (Martin and Rybicki, 2000), as implemented in RDP2.

Model fitting was performed by likelihood ratio tests (LRTs (Huelsenbeck and Rannala, 1997)) using MODELTEST3.06 (Posada and Crandall, 1998), PAUP*10b (Swofford, 2002) and PhyMLv.2.1 (Guindon and Gascuel, 2003). The automatically calculated parameter estimates for the best-fit model selected by MODELTEST were manually optimized by consecutive cycles of parameter and maximum likelihood (ML) phylogeny estimation in PAUP*, until convergence. The gamma distribution of among-site rate variation was approximated by a discrete distribution with eight rate categories (Yang, 1996), each category being represented by its mean. To avoid model overparameterization by MODELTEST when it selected the general time reversible (GTR) model, simpler (more restricted) models were evaluated manually by LRTs (Swofford et al., 1996). The indices of substitution saturation (I_{ss} and I_{ssc}) for third codon positions were estimated using the procedure by Xia and Xie (Xia et al., 2002) under the best-fit model parameter estimates and ML topology, as implemented in DAMBE. The robustness of NJ and ML topologies was inferred by non-parametric bootstrap tests (Felsenstein, 1985) using 1000 pseudoreplicates for the NJ analyses in PAUP*, whereas for ML analyses, 100 bootstrapped alignments were generated with SEQBOOT (PHYLIP package V.3.6 Felsenstein, 2004) and the corresponding ML topologies were inferred with PhyML (Guindon and Gascuel, 2003) under best-fit models, calculating a consensus tree with CONSENSE (Felsenstein, 2004). Tree display and editing was performed with TreeView1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/rod.html/). Standard unweighted maximum parsimony (MP) analyses were performed in PAUP* using a heuristic search strategy with 100 random sequence additions and TBR branch swapping.

Phylogenetic congruence between different data partitions was assessed in a maximum parsimony (MP) framework by the incongruence length difference (ILD) test (Farris et al., 1994), using 1000 random partition replicates of informative sites only to ensure best accuracy of P values (Darlu and Lecointre, 2002), as implemented in PAUP*. Branch and bound searches were used when <20 sequences were included in the analyses, and heuristic searches were conducted for $\geqslant 20$ sequences, using 10 random additions of sequences in all cases. Competing tree topologies were also evaluated in an ML framework using the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999), as implemented in PAUP*4b10 under the RELL model (Goldman et al., 2000), with 1000 bootstrap pseudoreplicates. Topologies and parameter values were those obtained under the best-fit model.

Split decomposition analyses (Bandelt and Dress, 1992) were used to visualize the degree of conflicting phylogenetic signals present in concatenated datasets, and to identify the sequences involved in the conflict. Distance (Hamming) split decomposition analyses were performed on variable sites only, using SplitsTree v.2.4 (Huson, 1998).

Bayesian phylogenetic analyses of concatenated sequence alignments using mixed models were performed with MrBayes3b4 (Ronquist and Huelsenbeck, 2003). Characters within single or combined sequence sets were partitioned by gene, as well as by codon position (SSR). The best-fit model for each partition was selected either by LRTs or the Akaike information criterion (AIC), as implemented in MODELTEST, or by using Bayes factors (Kass and Raftery, 1995; Nylander et al., 2004). The gamma distribution of among-site rate variation was approximated as described above. The different data partitions were allowed to evolve at different rates, but branch lengths were assumed to be proportional across partitions using a rate multiplier (Ronquist and Huelsenbeck, 2003). All substitution model parameters were also allowed to be independent across partitions. Default MrBayes priors on model parameters were used, except for Ratepr = variable (for multiple partition analyses). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) was used to estimate the posterior probability distribution using three incrementally heated chains with the temperature parameter set to 0.15, all four chains starting from different random trees. Each analysis was replicated three times for 3×10^6 generations, sampling the posterior distribution every 100th or 200th generation. Evidence for convergence of the different Markov chains was obtained by examining the correlation between the posterior probabilities of individual clades in pairwise comparisons among runs, and by comparing their generation plots for overall model likelihood (marginal $-\ln L$ against generations). The latter were used to infer a proper burn-in value (i.e., the number of samples to discard before the distri-

bution reached apparent stationarity). We checked that all runs had similar mean and variance of model likelihood after burn-in; if so, the samples from the stationary phases of three independent runs were pooled to obtain a 50% majority rule consensus tree using the sumt command n MrBayes.

Analyses of sequence polymorphisms within and between species were performed with DnaSP4.0 (Rozas et al., 2003) and SITES1.1 (Hey and Wakeley, 1997), to test the neutral mutation and population equilibrium hypotheses, to infer the population mutation ($\theta = 2N_e\mu$) and recombination (R or $C = 2N_e\rho$) parameters (Hey and Wakeley, 1997; Hudson, 1987), and to obtain estimates of population differentiation (Hudson et al., 1992a) and gene flow (Hudson et al., 1992b), as detailed in the relevant sections. Coalescent simulations based on 10^4 genealogy replications were performed with DnaSP to estimate the 95% confidence interval of the *Rm* (minimal number of recombination events) and R_2 (population growth) test statistics (Hudson and Kaplan, 1985; Ramos-Onsins and Rozas, 2002). Permutation analyses with 10^4 replicates were run to test the significance of the population subdivision test statistics (Hudson et al., 1992a).

3. Results

3.1. Strain diversity of Canarian and Moroccan isolates from endemic Canarian genistoid legumes assessed by REP-PCR genomic fingerprinting

The first step in the hierarchical analysis of the genetic diversity found among bacteria associated with ECGLs in Canarian and Moroccan soils was to generate high-resolution REP-PCR genomic fingerprints of the nodule isolates obtained from the four sampling sites indicated in Fig. 1, which were compared with reference ECGL isolates characterized in previous works (Table 1). Rep-PCR genomic fingerprinting using BOX, ERIC, or REP primers have a much greater discriminatory power than serotyping, PCR-RFLP or MLEE analyses, as evidenced in this and other studies (Cho and Tiedje, 2000; Judd et al., 1993; Rademaker et al., 2000; Vinuesa et al., 1998). This allowed us to identify the distinct clonal lineages (i.e., the equivalents of individuals in vertebrates or spermatophytes) within our collection. Repeated REP-PCR fingerprinting of three isolates in three different amplification experiments, and fractionation of the resulting amplicons on three different gels, resulted in linkage values of $r = 0.9 \pm 0.06$, which is consistent with previous reports (Rademaker et al., 2000; Vinuesa et al., 1998). Therefore, a conservative threshold level of 80% similarity was used as a means of identifying putative clonemates ($r > 80\%$) and distinct strains or clonal lineages ($r \ll 80\%$).

The result of a product-moment/UPGMA analysis of 53 REP-patterns is shown in Fig. 2. Seven clusters (A–H) of nearly identical fingerprints ($r > 0.80$) were found. These always grouped isolates from the same sampling site and host, indicating that different "epidemic" clonal lineages are found at each sampling site, as reported for some human bacterial pathogens and *Rhizobium* populations (Maynard-Smith et al., 1993; Silva et al., 2003). However, the geographic distribution of such epidemic clones is apparently very restricted, as clonemates, or highly related clones were always associated to a particular sampling site, denoting strong levels of endemicity of individual clones, as found also for other soil *Proteobacteria* (Cho and Tiedje, 2000; Oda et al., 2003). Genetic relationships between strains grouped at linkage levels <65% are unreliable because the method rapidly approaches saturation (Cho and Tiedje, 2000; Rademaker et al., 2000; Vinuesa et al., 1998), precluding their classification or grouping at the species or higher taxonomic levels, as shown in Fig. 2 (compare with the RFLP patterns).

3.2. RFLP analysis of amplified rrs fragments

Our previous diversity surveys of bradyrhizobia associated with ECGLs had uncovered three PCR-RFLP genotypes when *rrs* amplicons are digested with *Cfo*I, *Dde*I, and *Msp*I (Jarabo-Lorenzo et al., 2000; Vinuesa et al., 1998, 1999). Two of these genotypes (AAA and BBA) were found among the new isolates obtained from the four sampling sites indicated in Fig. 1, as shown along with their REP-PCR genomic fingerprints in Fig. 2. We had previously shown that the AAA pattern is identical to that of *B. japonicum* USDA110 (Vinuesa et al., 1998), and here we show that the BBA genotype corresponds to *B. canariense* (Vinuesa et al., 2004). Twelve ECGL isolates, including all Moroccan isolates trapped with tagasaste from the Moroccan Mehdiya–Kénitra site (MK isolates), presented the AAA genotype, whereas the remaining 23 ECGL isolates, including all the Moroccan isolates from the Maâmora forest, displayed the BBA pattern. The restriction sites for each genotype were mapped on the *rrs* sequence of *B. japonicum* USDA110, as indicated in the footnote of Table 1.

3.3. Genetic diversity and population structure of ECGL isolates inferred from MLEE data

Twenty two isolates from ECGLs were analyzed by MLEE at six loci (ADH, EST, IDH, IPO, MDH, and ME), resulting 19 multilocus electrophoretic types (ETs). ET1 and ET2 contained 2 (BC-P11 and BC-P8) and 3 (BC-P22, BC-MAM3 and BC-MAM9) isolates, respectively, the others were unique (Fig. 2). Note that ET1 isolates were grouped in the REP cluster H, whereas ET2 isolates were recovered from insu-

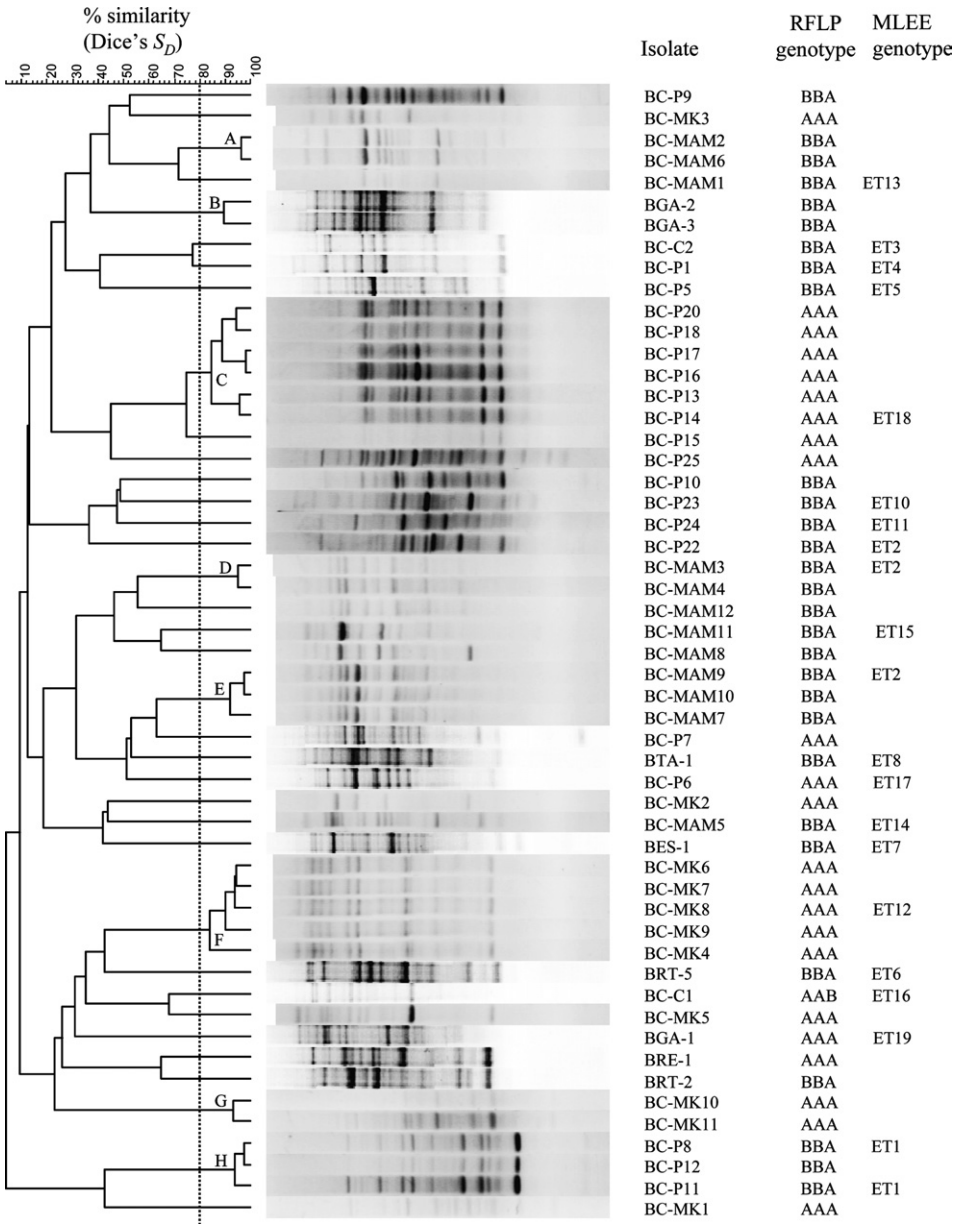

Fig. 2. Product-moment/UPGMA analysis of 53 REP-PCR genomic fingerprints of isolates from endemic shrub legumes of the Canary Islands, including Moroccan isolates trapped with *C. proliferus* plants using soils from Mehdiya-Kénitra and Maâmora as inoculum. Clusters A–H group clonemates at $r > 80\%$. The 16S rDNA PCR-RFLP and MLEE genotypes of selected strains are indicated to the right.

lar and continental populations and displayed distinct REP patterns (Fig. 2). Cluster analysis (UPGMA) of ETs resulted in two major groupings that were consistent with the AAA and BBA RFLP genotypes (data not shown). The 22 isolates were subdivided for subsequent hierarchical analyses into three populations as follows: (A) 12 *B. canariense* isolates from the Canaries, (B) six *B. canariense* isolates from the Moroccan locality Maamora, and (C) four Canarian *B. japonicum* isolates (Table 2). The 18 *B. canariense* isolates (populations A + B) displayed a lower genetic diversity ($H = 0.406$) than the four *B. japonicum* isolates ($H = 0.583$). The genetic differentiation between the *B. canariense* populations A and B was not significant ($G_{ST} = 0.063$, $\chi^2 = 14.93$, df = 10, $P = 0.135$), suggesting the existence of gene flow between the continental and insular populations, and supporting the notion that populations A and B belong to the same species (Table 2). The genetic differentiation between the *B. canariense* isolates (populations A + B) and the *B. japonicum* isolates was significantly different from 0 ($G_{ST} = 0.061$, $\chi^2 = 26.20$, df = 16, $P = 0.05$), suggesting that they belong to different species. These results were reinforced by the linkage disequilibrium analyses. When the 18 *B. canariense* isolates were analyzed, linkage equilibrium was found (V_0/V_e not significantly dif-

Table 2
Genetic diversity, genetic differentiation, and linkage disequilibrium estimates based on MLEE data for *Bradyrhizobium* populations associated with endemic Canarian genistoid legumes

Population[a]	No. of isolates	No. of ETs	H[b]	Mean No. alleles	G_{st}[c]	V_0/V_e[d]	P[e]
B. canariense (A + B)	18	15	0.406 (0.130)	2.7	0.063	0.96	NS
B. japonicum	4	4	0.583 (0.134)	2.3	ND	ND	ND
Total	22	19	0.476 (0.135)	3.7	0.061*	1.17	<0.001

NS, not significant. ND, not determined due to small sample size.
[a] A and B corresponds to samples from insular and continental populations.
[b] Mean genetic diversity, standard error is in parenthesis.
[c] Genetic differentiation, *G_{st} significantly different from 0.
[d] Observed variance/expected variance of the mismatch distribution.
[e] Probability of rejecting by chance the null hypothesis that $V_0 = V_e$.

ferent from 1), suggesting the existence of recombination within the population (Table 2). The inclusion of the four *B. japonicum* isolates in the analysis, however, revealed linkage disequilibrium, indicating the non-random association of alleles among members of the two species, which may be due to sexual isolation.

3.4. Scanning of multiple sequence alignments for intragenic mosaicism and other sources of phylogenetic distortions

Single DNA alignments were analyzed with several programs (see Section 2.7) to detect putative interspecies recombinant sequences. The identification of such sequences is important to minimize the distortions caused by mosaic DNA sequences in phylogeny estimation (Posada and Crandall, 2002; Schierup and Hein, 2000). Potential mosaic regions were considered as likely recombinants only if they were detected by at least two programs. Significant interspecies recombinant sequences were detected only in the *atpD* alignment, where the photosynthetic strain BTAi1 was found to contain a 3′-end segment nearly identical to that of *B. japonicum* USDA110, and strain BC-MK1 appeared to contain an internal segment related in sequence to that of *B. yuanmingense* TAL760 (see Figure S1 provided as supplementary online material). Evidence for gene mosaicism was also found for *B. canariense recA* sequences (data not shown). However, these were not removed as the parental and daughter sequences belonged exclusively to this species, reflecting within population recombination phenomena. This filtering did not guarantee, however, that all potential recombinants were identified, due to limited power of the programs to detect ancient recombinational events masked by superimposed point mutations (Posada and Crandall, 2001) and because the alignments are relatively short, implying that potential recombination breakpoints may lay outside of the sequenced segment of each locus. Therefore, we performed exploratory NJ analyses under the best-fit model of all individual alignments in order to identify sequences with unstable and unresolved (<60% bootstrap support) positions across phylogenies. Such sequences were not used in subsequent analyses if their removal did significantly improve the overall tree resolution, as indicated in the relevant sections.

3.5. Estimation of evolutionary sequence parameters from atpD, glnII, recA, and nifH multiple sequence alignments

No significant saturation was found at the third codon positions of *atpD*, *glnII*, *recA*, and *nifH* sequences ($I_{ss} < I_{ssc}$; $P < 0.001$ in all cases; data not shown). Therefore, all sites were considered in subsequent analyses.

Basic sequence alignment characteristics for each data partition and the corresponding ML estimates for the frequency and rate parameter values obtained under the best-fit models of sequence evolution are presented in Table 3. Gapped sites were excluded for model fitting or phylogeny estimation. All sequences presented biased nucleotide compositions, as expected for genes from organisms with high G + C content (typically >60%). The substitution patterns were complex, with three or more rate parameters as estimated by likelihood ratio tests (LRTs; Table 3). Accounting for among-site rate variation resulted in highly significant improvements of the corresponding models ($P < 0.0001$ in all cases). The automatic ηLRT performed by MODELTEST selected the GTR + G model for the *atpD*, *glnII*, and *recA* partitions. However, manual LRTs performed with the simpler models presented in Table 3 revealed that they explain the data equally well (data not shown), indicating that ηLRT performed by MODELTEST tends to yield significantly overparameterized models. This was evidenced in some instances by the AIC, also implemented in MODELTEST, which introduces a penalty term for models with more parameters (not shown).

A significant bias toward transitional substitutions was observed in all datasets (Table 3). The proportion of invariant sites was significant in the *recA* and *nifH* alignments. The G + C content across taxa for each data partition was homogeneous ($P \geqslant 0.92$ in all cases), as assessed with the χ^2 test implemented in PAUP*.

Table 3
Basic sequence alignment characteristics and maximum-likelihood estimates of frequency and rate parameter values under the best-fit models selected by likelihood ratio tests

Loci	No. seq[a]	No. haplo[b]	No. aln sites[c]	No. gaps	No. C sites[d]	No. V sites[e]	No. Pi sites[f]	π (A, C, G, T)[g]	Rate matrix[h]	I[i]	α[j]	Best fit model[k]
atpD	45	34	483	12	315 65.2%	168 34.8%	111 23.0%	0.193, 0.353 0.318, 0.135	Ra = 0.679, Rb = 2.315 Rc = 1.000, Rd = 2.315 Re = 8.242, Rf = 1.000	0.00	0.277	Custom 4 subst. types + G −ln L = 2495.1813
glnII	46	34	594	0	409 68.9%	185 31.1%	135 22.7%	0.201, 0.339 0.292, 0.168	Ra = 0.516, Rb = 1.550 Rc = 1.000, Rd = 1.550 Re = 5.662, Rf = 1.000	0.00	0.298	Custom 4 subst. types + G −ln L = 2701.678
recA	43	36	511	0	313 61.3%	198 38.7%	145 28.4%	0.188, 0.343 0.335, 0.134	Ra = 0.280, Rb = 1.343 Rc = 1.000, Rd = 1.000 Re = 5.015, Rf = 1.000	0.481	1.030	Custom 4 subst. types + I + G −ln L = 3163.6422
nifH	25	19	336	0	187 55.7%	149 44.3%	123 36.6%	0.197, 0.349 0.309, 0.145	Ra = 1.000, Rb = 3.827 Rc = 1.000, Rd = 1.000 Re = 5.642, Rf = 1.000	0.484	1.869	TrN+I+G −ln L = −2108.0913

[a] Number of sequences.
[b] Number of haplotypes, including outgroup sequences.
[c] Number of aligned sites.
[d] Number and percentage of constant sites.
[e] Number and percentage of variable sites.
[f] Number and percentage of parsimony informative sites.
[g] Nucleotide equilibrium frequencies.
[h] Substitution rates: Ra = [A–C], Rb = [A–G], Rc = [A–T], Rd = [C–G], Re = [C–T], Rf = [G–T].
[i] Proportion of invariant sites.
[j] Shape parameter of the gamma distributed across-sites rate variation (approximated by eight discrete rate categories).
[k] Best fit model and associated −ln likelihood value for alignment matrices without indels.

3.6. Maximum likelihood atpD, glnII, and recA gene phylogenies

A 5′-end fragment of the *atpD* gene, spanning positions 274–756 of the *B. japonicum* USDA110 (bll440) ortholog (Kaneko et al., 2002), was amplified from all bradyrhizobia tested. Values of key sequence parameters are shown in Table 3. Only the sequences from strains BC-MK6 and CICS70 exhibited significant disparity index (I_D) values (Kumar and Gadagkar, 2001). They were excluded from the analysis along with those from BTAi1, BC-MK1 and TAL760 because doing so significantly increased the resolution of several internal nodes. This effect is probably caused by mosaic gene structures detected for the latter three strains (Fig. S1, see online supplementary material). Twelve gaps were introduced in the alignment due to four extra codons found in all *Bradryrhizobium* sequences, which were absent in the three outgroup sequences used to root the *Bradyrhizobium* clade. Fig. 3A shows a ML topology for 31 *Bradyrhizobium* haplotypes. They form a well defined clade with 100% bootstrap support (BS), which has poor internal resolution. The two *B. yuanmingense* sequences represent a well resolved basal lineage. Other clades that correspond to named species are highlighted with brackets. The *B. japonicum* sequences are paraphyletic, being grouped in three independent clusters. The sequences from the two *B. japonicum* type strains (USDA6^T and DSMZ30131^T) differed at five synonymous substitution sites. Note that the *Bradyrhizobium* genospecies β sequence from strain BRE-1 grouped within the well resolved clade formed by *B. japonicum* bv. *genistearum* strains like BGA-1 (see below), highlighting a lateral transfer event across these lineages.

A 5′-end fragment spanning positions 274–756 of the *B. japonicum* USDA110 *glnII* (blr4196) sequence could be amplified from all *Bradyrhizobium* strains tested. The sequences from strains BC-C1, CIAT3101, and CICS70 were excluded from the analysis as their presence was found to decrease significantly the bootstrap support values of several clades. Fig. 3B shows a ML phylogram for 33 *Bradyrhizobium* haplotypes. Rooting *glnII* sequences has been shown to be unreliable when all rhizobial genera are considered, in part explained by lateral gene transfer across genera (Turner and Young, 2000). We found that *Bradyrhizobium glnII* sequences could be rooted with that from *R. palustris*. The topology is well resolved, revealing seven internal sequence clades (highlighted by brackets) that correspond to the current taxonomic scheme of the genus. Only the bipartition corresponding to the genospecies β clade is not significantly supported. Notice, however, that the photosynthetic strains are paraphyletic with respect to the other bradyrhizobia (also revealed by Bayesian analyses; not shown). This topology provides weak to moderate evidence for the monophyly of the *B. japonicum* sequences and moderate support for their grouping as sister to the monophyletic *B. canariense* sequence cluster. However, this relationship did not hold when the sequences of BC-C1 and CIAT3101 were included, which formed a basal split to the *B. canariense* clade, suggesting that lateral gene transfer may have been at play, although no evidence for gene mosaicism could be found in these sequences.

The *recA* primers designed in this study successfully amplified all *Bradyrhizobium* strains tested. Fig. 3C shows a ML phylogeny for 33 *Bradyrhizobium* haplotypes rooted with three outgroup sequences. The bradyrhizobia appear as a highly supported monophyletic group with six well resolved internal clades that are consistent with the taxonomic structure of the genus. This phylogeny supports the inclusion of the photosynthetic strains as a basal lineage of the genus, the monophyly of the *B. japonicum* I and Ia lineages, and their grouping as a sister lineage to the well resolved *B. canariense* clade. The greatest uncertainty in the topology is the bipartition connecting the *B. liaoningense*, *B. yuanmingense*, and *Bradyrhizobium* genospecies β lineages which, with the exception of *B. liaoningense*, appear to be monophyletic. The sequences of BC-C1 and CIAT3101 were again excluded, as their presence significantly weakened the overall tree resolution.

3.7. Maximum likelihood nifH phylogeny and the identification of new Bradyrhizobium biovarieties

Fig. 4 shows an ML phylogram for 16 *Bradyrhizobium* haplotypes, rooted with three outgroup rhizobial sequences. The tree shows that *Bradyrhizobium nifH* sequences constitute a monophyletic cluster (100% BS) with a well resolved internal structure. The most striking aspect of *nifH* evolution revealed by this tree is the correlation between the phylogenetic grouping of strains and their host ranges. All isolates from genistoid legumes, regardless of their original host, geographic origin or species designation, were recovered in a strongly supported clade, which indicates gene transfer events across species. It was previously found that none of the *B. canariense* and *B. japonicum* isolates from genistoid legumes recovered in this clade nodulates soybeans, whereas *B. japonicum* and *B. elkanii* isolates from soybeans do not nodulate genistoid legumes (Jarabo-Lorenzo et al., 2000, 2003; Vinuesa et al., 1998, 2004). To reflect the differential and genetically based adaptive phenotype (or "ecotype") displayed by *B. japonicum* isolates from soybeans and genistoid legumes, we propose the two new biovarieties *glycinearum* and *genistearum*, respectively. It is noteworthy that both *nifH* (Fig. 4) and *nodC* phylogenies (Jarabo-Lorenzo et al., 2003; Vinuesa et al., 2004) indicate that the alleles from biovar *genistearum* isolates split off at a basal position of the corresponding trees, which correlates with the early split

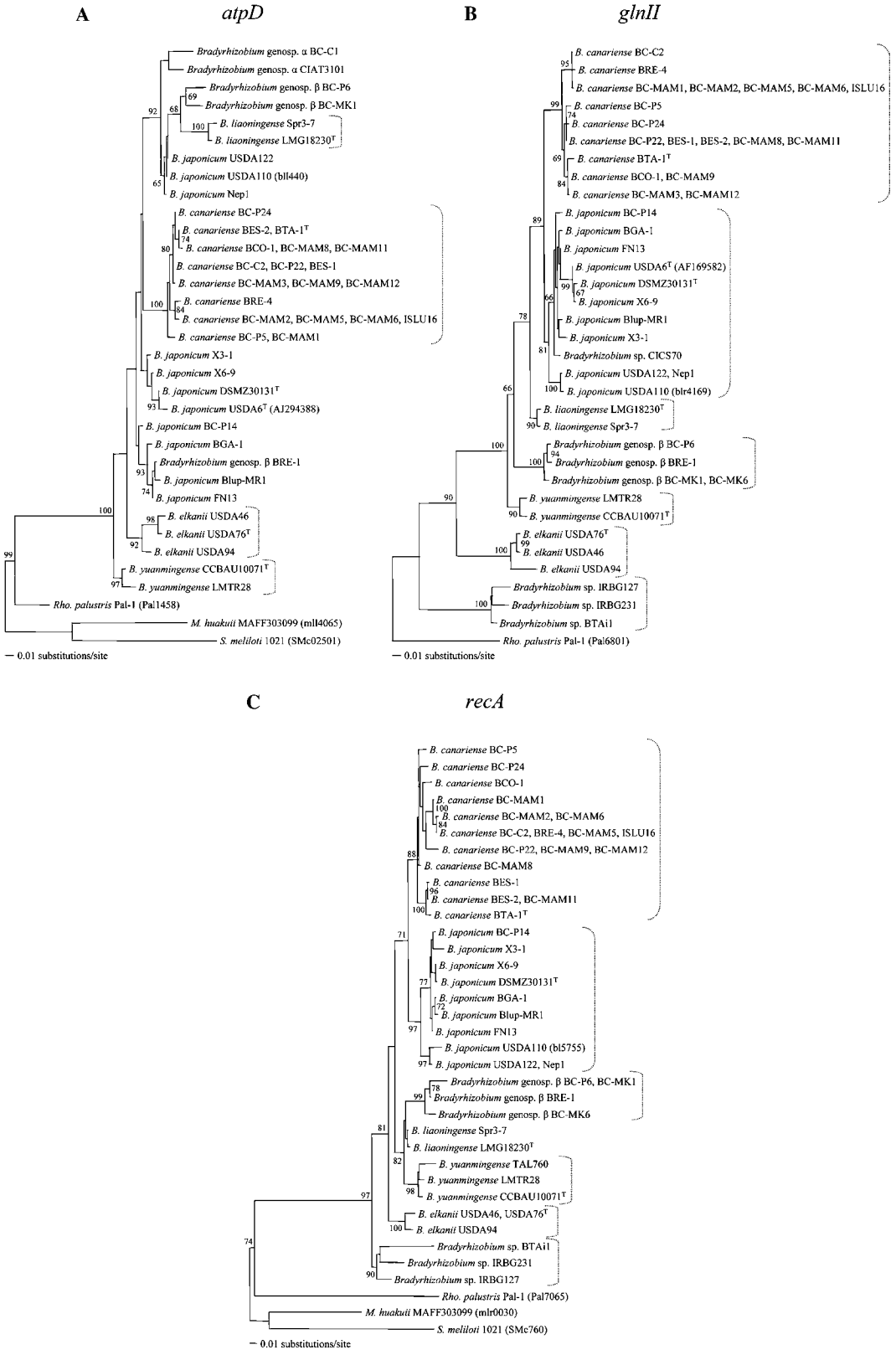

Fig. 3. Maximum likelihood gene phylograms under best-fit substitution models (see Table 3) inferred from the 5′-end *atpD*, *glnII* and *recA* sequences of *Bradyrhizobium* and outgroup strains. The gene ID numbers for outgroup sequences and the accession numbers of reference *Bradyrhizobium* sequences retrieved from sequence databases are indicated. The GenBank accession numbers for the sequences generated in this study are provided in Table S1 (see online supplementary material). The maximum likelihood bootstrap proportions out of 100 pseudoreplicates are indicated at the corresponding nodes when $\geqslant 65\%$. Brackets on the right side of each topology denote well resolved sequence clades (>80% bootstrap support) that correspond to named species or unnamed genospecies as defined by Vinuesa et al. (2004).

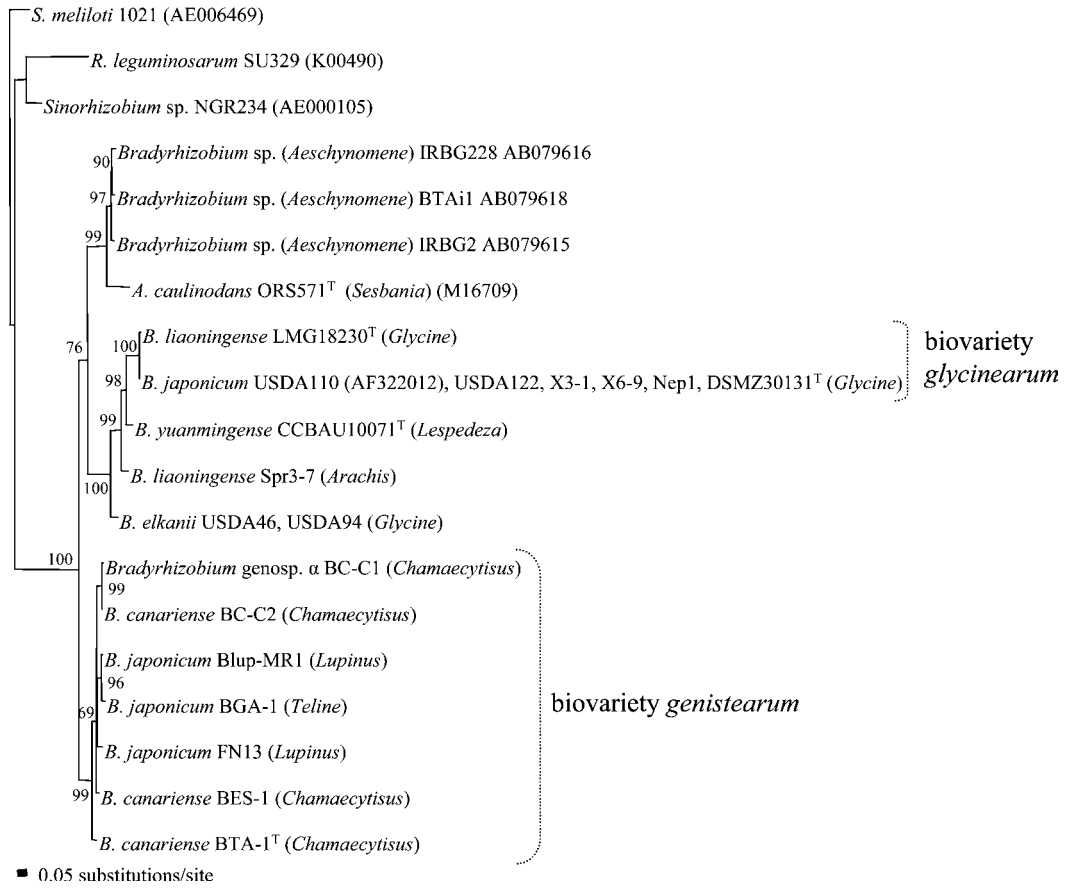

Fig. 4. Maximum likelihood *nifH* phylogeny estimated under the best-fit substitution model (see Table 3). The gene ID or accession numbers of outgroup sequences retrieved from sequence databases are indicated. The GenBank accession numbers for the sequences generated in this study are provided in Table S1 (online supplementary material). The maximum likelihood bootstrap proportions out of 100 pseudoreplicates are indicated at the corresponding nodes when $\geqslant 65\%$. Brackets on the right side of the tree denote the two new biovarieties discovered in this study (see text for details).

of genistoid legumes in relation to the those from the more derived tribe Phaseoleae:Glycinianae (Doyle and Luckow, 2003), which are nodulated by the biovar *glycinearum* strains. The sequences from photosynthetic bradyrhizobia formed another highly supported monophyletic lineage. The remaining *Bradyrhizobium* sequences from strains nodulating divergent hosts (*Arachis*, *Glycine*, *Lespedeza* and *Parasponia*) formed a third well supported clade. Interestingly, all *B. japonicum* isolates from *Glycine max* (soybean) shared a unique haplotype, closely related to, and significantly associated with that from *B. liaoningense* LMG18230^{T}, also a soybean isolate. However, the *B. liaoningense* isolate Spr3-7 from peanut (*Arachis hypogaea*) has a related but distinct *nifH* allele, characterized by a single codon deletion. Consequently, the soybean *B. liaoningense* isolates was also included in the biovar *glycinearum*.

3.8. Incongruence length difference and Shimodaira–Hasegawa tests of topological congruence

Since none of the gene phylogenies was totally resolved (Figs. 3A to C), and gene trees should not be confounded with species trees (Nichols, 2001; Rosenberg, 2002), an obvious step in the analysis of multiple sequence data sets obtained from the same individuals of different species is to concatenate the alignments. One can then test whether the different data partitions converge towards the same underlying species phylogeny or indicate discrepancies leading to conflicting solutions (Huelsenbeck et al., 1996).

The incongruence length difference test (ILD), also known as partition homogeneity test, has been shown to be useful for rapidly identifying the individuals contributing conflicting signals between different data partitions (Brown et al., 2002; Escobar-Páramo et al., 2004). Table S2 (supplementary online material) shows the results of pairwise ILD tests for the four loci sequenced herein, which indicate that the different data sets yield significantly incongruent topologies. The null hypothesis of congruent topologies was also significantly rejected by the SH test in all cases (data not shown). This was expected, as we had detected mosaic genes (Fig. S1), strains with unstable positions across the different gene phylogenies (i.e., BC-C1, BC-MK1, BRE-1, CIAT3101, or CICS70) and taxa, like *B. japonicum*, that appear as

paraphyletic in some gene trees and monophyletic in others. Further potential character incongruence between datasets was detected by visually comparing the topologies obtained for the different partitions (Figs. 3A to C), and confirmed by ILD tests performed in all possible pairwise comparisons for the *atpD*, *glnII*, and *recA* partitions. This analysis was started with sequences from the individual species to identify strains contributing conflicting phylogenetic signals within specific cladeo, which are likely to be the most abundant ones (Brown et al., 2002; Dykhuizen and Green, 1991; Escobar-Páramo et al., 2004). The next step was to detect conflicting sequences in across-species comparisons, using the within-species congruent partitions. The highest number of congruent *Bradyrhizobium* sequences was found for the *glnII* + *recA* partitions ($n = 36$ strains, 32 haplotypes; $P = 0.064$). The sequences yielding significant ILD values for particular comparisons are indicated in Table S1 (online supplementary material).

3.9. Estimation of a Bradyrhizobium species phylogeny based on combined and partially congruent glnII + recA sequences using different partitioning schemes and mixed models in a Bayesian framework

A Bayesian approach was chosen to infer a species phylogeny because its implementation in MrBayes3 allows the use of mixed models for partitioned data analyses (Nylander et al., 2004; Ronquist and Huelsenbeck, 2003), which is a desirable attribute for complex inferences such as those involving different sequence partitions. The evidence gained from the congruence tests presented above was used to select the partitions (*glnII* + *recA*) showing the highest number of sequences converging towards a common underlying topology. However, a few incongruent sequences (highlighted on the tree shown in Fig. 5) were included in the analysis to extend taxon sampling to all potential *Bradyrhizobium* lineages present in our collection. The inclusion of these sequences did not significantly affect the overall tree topology or resolution (data not shown). Proper model selection is a critical and complex issue in phylogeny estimation, especially when mixed models are implemented to analyze partitioned datasets (Lemmon and Moriarty, 2004; Nylander et al., 2004). We analyzed three partitioning schemes: (P1) partitioning by gene plus site specific (i.e., by codon position) rates (by gene + SSR); (P2) partitioning by gene (by gene); (P3) considering the combined gene sequences as a single partition (no partitioning). Model selection for each partitioning scheme was guided by ηLRTs and the AIC, as implemented in MODELTEST. If the best-fit model estimates obtained by these methods differed, the simpler model was selected, as likelihood-based model selection tends to favour overparameterized ones (Nylander et al., 2004). Alternative models were also tested using Bayes factors (Kass and Raftery, 1995), which take into account the entire parameter space (Holder and Lewis, 2003). An additional advantage of the Bayesian framework for model selection is that it marginalizes over nuisance parameters in the model (Holder and Lewis, 2003; Huelsenbeck et al., 2001; Nylander et al., 2004), accounting for model uncertainties and the interactions among phylogenetic signals in different partitions.

The relative merits of each of the three partitioning schemes under the corresponding best-fit models was assessed using Bayes factors (B_{10}), which were calculated as the ratios between the marginal model likelihoods of the more general model (M_1) over the more restricted one (M_0). Overall model likelihoods for each partitioning scheme were estimated as the harmonic means of marginal $-\ln L$ values from pooled post burn-in samples from 2 or 3 different convergent Markov chains (Nylander et al., 2004). Interpretation of the relative merits of the competing models (posterior odds of M_1 to M_0) was performed after Kass and Raftery (Kass and Raftery, 1995), using the critical values reported in Table 1 of Nylander et al. (2004). Table 4 shows the results of the Bayes factor analyses. They reveal that by far the SSR variation was the most important model component to take into account, dramatically increasing the model fit to the data (by over 457 log likelihood units when P1 is compared to P2 or P3), whereas partitioning the combined alignment by gene (P2) or not (P3), did not affect the overall model fit (Table 4).

Fig. 5 shows the 50% majority rule consensus tree obtained for the pooled set of post burn-in trees from three independent Markov chains using the P1 partitioning scheme. The posterior probabilities (PPs) of all bipartitions are highly significant, except for that corresponding to the *Bradyrhizobium* genospecies β clade. Moderate to high support for most of the bipartitions was also obtained from ML bootstrap analysis (Fig. 5). Only the bipartitions for *B. liaoningense* and *Bradyrhizobium* genospecies β clades were not consistently resolved by significant PPs ($\geqslant 0.95$) and BS ($\geqslant 75\%$), which probably reflects the incongruence detected for these sequences by the ILD test. A SH test indicated that only the *glnII* partition was significantly incongruent ($\Delta -\ln L = 139.680$, $P < 0.05$) with that inferred from the combined *glnII* + *recA* alignment, whereas the *recA* partition was congruent with it ($\Delta -\ln L = 33.317.680$, $P = 0.19$). The topology shown in Fig. 5 provides strong support for the monophyly of the genus, and provides the following evolutionary hypothesis for the splitting order of *Bradyrhizobium* species ((((((((*B. japonicum*, *B. canariense*), *Bradyrhizobium* genospecies α), *B. liaoningense**) *Bradyrhizobium* genospecies β*), *B. yuanmingense*), *B. elkanii*), photosynthetic bradyrhizobia), *R. palustris* = outgroup), where the bipartitions marked with an asterisk are not statistically supported. It should

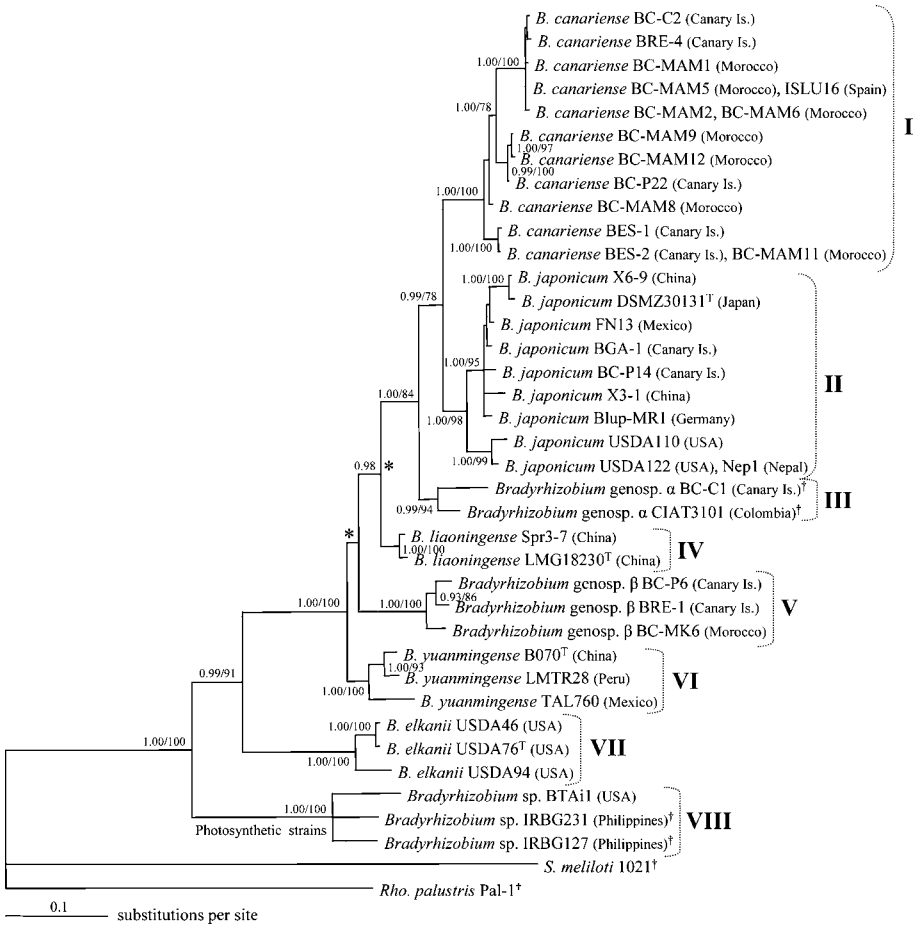

Fig. 5. Bayesian species phylogeny for the genus *Bradyrhizobium* inferred from partially congruent *glnII* + *recA* sequence partitions using a gene + site-specific rate partitioning scheme. Incongruent sequence partitions are denoted by † (see text for details). The topology represents the 50% majority rule consensus tree derived from 89997 trees resulting from the convergent and pooled post burn-in samples obtained from three independent MCMCMC samplers run for 3×10^6 generations and sampled every 100th. Notice the high resolution of this topology, as denoted by the significant and correlated posterior probabilities/ML bootstrap support values (100 pseudoreplicates) shown at the relevant nodes. Brackets to the right of the tree indicate the eight evolutionary lineages found among the collection of *Bradyrhizobium* strains analyzed. Notice also the broad geographic distribution and host-range (shown in Table 1) exhibited by most of them. Lineages I, II, III, and V contain strains isolated from genistoid legumes from the Canary Islands and other parts of the world. Lineage VIII groups the photosynthetic strains. Isolates marked with a^T are type strains. The nodes marked with an asterisk (*) were not fully resolved (i.e. $\geqslant$0.95 posterior probability and $\geqslant$75% bootstrap support).

Table 4
Effect of data partitioning on marginal model likelihood

Data partition[a]	Model likelihood[b] $\log_e f(X\|M)$	Bayes factor[c]	
		$\log_e B_{10}$ comparisons	2 $\log_e B_{10}$ comparisons
(P1) by gene + SSR	−5878.85802	(P1 vs. P2) 857.172*	(P1 vs. P2) 428.586*
(P2) by gene	−6307.41247	(P1vs. P3) 858.46*	(P1vs. P3) 429.230*
(P3) no partition	−6308.15693	(P2 vs. P3) 1.488	(P2 vs. P3) 0.744

[a] (P1) refers to the 38 combined *glnII* + *recA* sequences, partitioned by gene and by codon position (site specific rates, SSR), (P2) by gene alone, and (P3) without partitioning. The substitution model for each site for P1 is indicated in Figure S2 (provided as online supplementary material). Those for the P2 and P3 partitioning schemes were GTR + G in each case.

[b] Refers to the overall or marginal model likelihoods for each partitioning scheme, where the values represent the harmonic means of the pooled post burn-in −ln L values.

[c] Is the posterior odds of M_1 to M_0 (see text for details).

* Very strong evidence in favor of the most complex partitioning model was obtained, since $\log_e B_{10} \gg 150$ and 2 $\log_e B_{10} \gg 10$.

be noted that this topology is highly consistent with the current taxonomic structure of the genus (Sawada et al., 2003; Vinuesa et al., 2004; Willems et al., 2003).

Taken together, the marginal likelihood generation and clade correlation plots shown in Figures S2A and S2B (online supplementary material) provide good evi-

dence for the convergence and adequate mixing of the three independent Markov chains (Huelsenbeck et al., 2001; Huelsenbeck et al., 2002), whereas Figures S2C, S2D, S2E, S2F, and S2G, which show the mean and 95% credible intervals for all the parameter estimates of the model, reveal the significant levels of uncertainty of these estimates. When more complex models were used (e.g., using an invariant sites + gamma model to account for among-site rate variation, or six nucleotide substitution rates), extremely large standard deviations and significantly larger 95% credible intervals were obtained for several parameters (data not shown), suggesting that the selected model strikes a good compromise between model complexity and parameter variance. Importantly, the topology and the posterior clade PPs were nearly unaffected by the different models and partitioning schemes tested, highlighting their robustness and low sensitivity to model uncertainties.

3.10. Evolutionary inferences from DNA polymorphisms of atpD, glnII, and recA sequences from B. canariense and B. japonicum populations

Phylogenetic analyses of the *atpD*, *glnII*, and *recA* sequences provided compelling evidence for the monophyly of *B. canariense* strains (Figs. 3A to 3C and 5) and strong support for the hypothesis that it is the sister species of *B. japonicum* (Fig. 5). The genetic differentiation and linkage disequilibrium analyses based on MLEE polymorphisms suggested that both species are genetically isolated, and that significant recombination exists within but not between them (Table 2). These conclusions were confirmed, refined and extended by a population DNA polymorphism analysis of the three loci sequenced for both sister clades.

The results shown in Table 5 for the χ^2 and K^*_{ST} indices (Hudson et al., 1992a) indicate that no genetic differentiation (population subdivision) was detected between continental and insular isolates of *B. canariense*, which therefore were treated as a single population or deme in subsequent analyses. Significant gene flow was detected between them, as inferred from the high number of migrants (*Nm*) and low fixation index (F_{ST}) values. The *Nm* estimates are known to be skewed toward large values (Hudson et al., 1992b), which is why those shown in Table 5 may be somewhat overestimated. These data support the conclusion that migration is a significant evolutionary force providing cohesion to *B. canariense*. On the other hand, highly significant genetic differentiation and no significant gene flow were detected between the total *B. canariense* and *B. japonicum* populations, which was clearly reflected in the number of fixed differences between them (Table 5). Hence, a second major conclusion is that the bipartition separating *B. canariense* and *B. japonicum* in Fig. 5 corresponds to a speciation event, justifying their taxonomic recognition as *bona fide* evolutionary and cohesive species. Note that significant population subdivision was found within the *B. japonicum* sample for the three loci examined (data not shown), which indicates the *B. japonicum* bv. *glycinearum* strains grouped in the DNA-homology groups I and Ia (Hollis et al., 1981) form two differentiated gene pools within the species.

Table 6 shows basic descriptive statistics for the DNA polymorphisms found in the *B. canariense* and *B. japonicum* populations. The analyses are based on the segregating sites, excluding those that violate the infinite

Table 5
Genetic differentiation and gene flow estimates

Gene/populations	No. Fix Diff.[a]	Genetic differentiation				Gene flow	
		χ^2 (df)[b]	P[c]	K^*_{ST}[d]	P[e]	F_{ST}[f]	Nm[g]
atpD							
B. canariense insular vs. continental	0	13.333 (14)	0.5005 (NS)	0.00135	0.4151 (NS)	0.01458	33.78
B. japonicum vs. *B. canariense*	11	28.000 (17)	0.0449	0.20326	0.0000	0.57751	0.37
glnII							
B. canariense insular vs. continental	0	11.200 (8)	0.1906 (NS)	0.05014	0.1000 (NS)	0.09391	4.82
B. japonicum vs. *B. canariense*	9	28.000 (17)	0.0449	0.21687	0.0000	0.61930	0.31
recA							
B. canariense insular vs. continental	0	9.333 (10)	0.5008 (NS)	0.00107	0.3901 (NS)	0.04188	11.44
B. japonicum vs. *B. canariense*	10	28.000 (19)	0.0834 (NS)	0.16635	0.0000	0.58114	0.36

NS, not significant.

[a] Number of fixed differences between populations.

[b] Haplotype based statistic (Hudson et al., 1992a), degrees of freedom are indicated in parenthesis.

[c] Probability of rejecting the null hypothesis that the two populations are not genetically differentiated, based on the critical values from the χ^2 distribution.

[d] Sequence based statistic described in Hudson et al. (1992a).

[e] Probability obtained by the permutation test (Hudson et al., 1992a) with 1000 replicates.

[f] Sequence based estimate described in Hudson et al. (1992a).

[g] Effective number of migrants.

Table 6
Descriptive statistics of nucleotide polymorphisms neutrality and population growth tests

Species/gene	No. sites (bp)	S[a]	Pi[b]	ks/ka[c]	k[d]	h/Hd[e]	θ[f]	π[g]		Tajima's D[j]	Fu's D*[j]	Fu's F_S[j]	R_2[k]
								Total	Synon[h]	Nonsynon[i]				
B. canariense ($n = 18$)													
atpD	483	18	16	17/2	6.118	8/0.902	0.01025	0.01267	0.04731	0.00116	0.65277	0.99173	1.071	0.1657
glnII	594	23	12	22/2	6.974	9/0.863	0.01079	0.01174	0.04548	0.00135	0.16924	−1.11011	0.629	0.1333
recA	511	45	32	46/1	13.869	11/0.928	0.02456	0.02714	0.10034	0.00163	0.24644	−0.01796	1.160	0.1386
B. japonicum ($n = 10$)														
atpD	483	35	28	27/9	14.378	10/1.000	0.02493	0.02977	0.08727	0.01034	0.78334	0.78884	−2.528	0.1833
glnII	594	30	18	28/5	11.422	9/0.978	0.01615	0.01923	0.07180	0.00308	0.37023	−0.16482	−1.320	0.1499
recA	511	39	20	39/1	12.844	9/0.978	0.02708	0.02514	0.09326	0.00124	−0.33127	−0.49355	−1.061	0.1321

[a] Segregating sites.
[b] Parsimony informative sites.
[c] Total number of synonymous/nonsynonymous changes.
[d] Average number of nucleotide differences.
[e] Number of haplotypes/haplotype (gene) diversity.
[f] Theta per bp, after Watterson (1975).
[g] Nucleotide diversity.
[h] Nucleotide diversity of synonymous sites.
[i] Nucleotide diversity of nonsynonymous sites.
[j] Calculations using the total number of segregating sites; all the values are non-significant.
[k] Population growth test statistic of Ramos-Onsis and Rozas (2002); all the values were non-significant as determined by neutral coalescence simulations considering recombination if necessary.

sites model (i.e., those segregating two or more bases). The observed pattern of nucleotide substitution was compatible with that expected under the neutral equilibrium model, as revealed by Tajima's D (Tajima, 1989) Fu and Li's D^* and F^* (Fu and Li, 1993), and Fu's F_S (Fu, 1997) statistics, which are based on intraspecifc data of DNA polymorphisms to test the hypothesis that all mutations are selectively neutral (Kimura, 1983). The powerful R_2 test statistic (Ramos-Onsins and Rozas, 2002), which is particularly suited for small sample sizes with recombination, also failed to reject the equilibrium model, as shown by coalescent simulations. Therefore, all evidence indicates that the observed polymorphisms at the three housekeeping loci conform to the neutral equilibrium model (Fu, 1997; Ramos-Onsins and Rozas, 2002; Tajima, 1989).

We used two different approaches to estimate the population recombination parameter for 18 *B. canariense* isolates (C or $R = 2N_e\rho$ for haploid organisms, where N_e is the effective population size and ρ the recombination rate per site); estimations of R and Rm are based on the variance of the number of base pair differences between DNA sequences (Hudson and Kaplan, 1985; Hudson, 1987), whereas estimation of C is based on coalescence theory (Hey and Wakeley, 1997). The latter approach was shown in simulation studies to provide reliable estimates of C, with low to moderate bias, even for small data sets (Hey and Wakeley, 1997). The results presented in Table 7 provide significant evidence for intragenic recombination at all but the *glnII* locus. Estimates of R or C are high to intermediate for the other two loci, as confirmed by neutral coalescent simulations performed under the assumption of moderate levels of recombination (Rozas et al., 2003). The simulations were used to estimate the 95% confidence intervals of the conservative Rm parameter (Hudson and Kaplan, 1985). Table 7 shows that the estimated Rm values for the *atpD* and *recA* sequences lie within (*atpD*, $P \leqslant 0.56$) or outside of the upper interval (*recA*, $P \leqslant 0.99$), which is consistent with moderate and high recombination rates at these loci, respectively, although lower than that expected under free recombination (simulation data not shown). The diversity generated by point mutation is reflected in the estimates of the population mutation parameter ($\theta = 2N_e\mu$, where μ is the mutation rate per site), while the estimates of the parameter C ($2N_ec$) indicate the diversity generated by recombination. Dividing the latter by the former provides an estimate of the ration c/μ, which can be interpreted as the chance of recombination per site over point mutation per site. The c/μ values shown in Table 7 provide additional evidence for the conclusion that recombination is another major evolutionary force shaping the population genetic structure of *B. canariense*. The corresponding R and C estimates for *B. japonicum* point to the same conclusion, although they may not be conclusive due to the small

Table 7
Recombination estimates for the *Bradyrhizobium canariense* population ($n = 18$)

Species/gene	R[a]	Rm[b]	Coalescence simulations[c]			C[g]	c/μ[h]
			Conf. Int.[d]	$P \leqslant$ obs. Rm[e]	Rm avg.[f]		
atpD	10.0	2	(1.0, 5.0)	0.56	2.4	0.7	0.29
glnII	3.4	0	(0.0, 4.0)	0.11	1.5	0.0	0.00
recA	13.0	10	(2.0, 8.0)	0.99	5.1	14.8	2.37
Mean	8.8	4	(1.0, 5.6)	0.55	3.0	5.2	0.88

[a] Estimate of the population recombination parameter R (Hudson, 1987) corrected for haploid organisms.
[b] Minimum number of recombination events (Hudson and Kaplan, 1985).
[c] Neutral coalescence simulations (10^4) given the number of segregating sites, with an intermediate level of recombination.
[d] Confidence interval (lower limit, upper limit) for Rm under the neutral coalescent process.
[e] Probability that $Rm \leqslant$ the observed Rm under the neutral coalescent process.
[f] Average value of Rm under the neutral coalescent process.
[g] Estimate of population recombination parameter using the coalescent approach of Hey and Wakeley (1997).
[h] Recombination/mutation rate ratio.

sample size for these strains, especially given the evidence found for population subdivision within this group.

3.11. Visualization of recombination patterns using split-decomposition analysis of concatenated atpD, glnII, and recA sequences from B. canariense and B. japonicum strains

One reason for the topological incongruence revealed by the ILD and SH tests (Table S2, online supplementary material) could be recombination, as suggested by the linkage disequilibrium analyses of MLEE data (Table 2). Further evidence for this hypothesis was provided by an independent approach, split decomposition analysis (Bandelt and Dress, 1992; Huson, 1998), which was performed on concatenated pairs of data sets from the sister species *B. canariense* and *B. japonicum*. Figs. 6A and C show the corresponding splits graph solutions, which show a great number of parallel paths of sequence evolution, depicted as networks, as expected for incongruent data partitions resulting from recombination events (Brown et al., 2002; Dykhuizen and Green, 1991; Zhou et al., 1997). These graphs revealed that the highest number of conflicting phylogenetic signals is found in the comparisons involving the *atpD* partition, as predicted by the ILD tests (Table S2). Importantly, the split graphs indicate that the sympatric *B. japonicum* and *B. canariense* sequence lineages evolve along independent pathways, as reticulations are found only between sequences from conspecific strains, but not across species (notice that the genetic distances between them are not reflected in this graph, which were displayed with equal edges in order to illustrate the

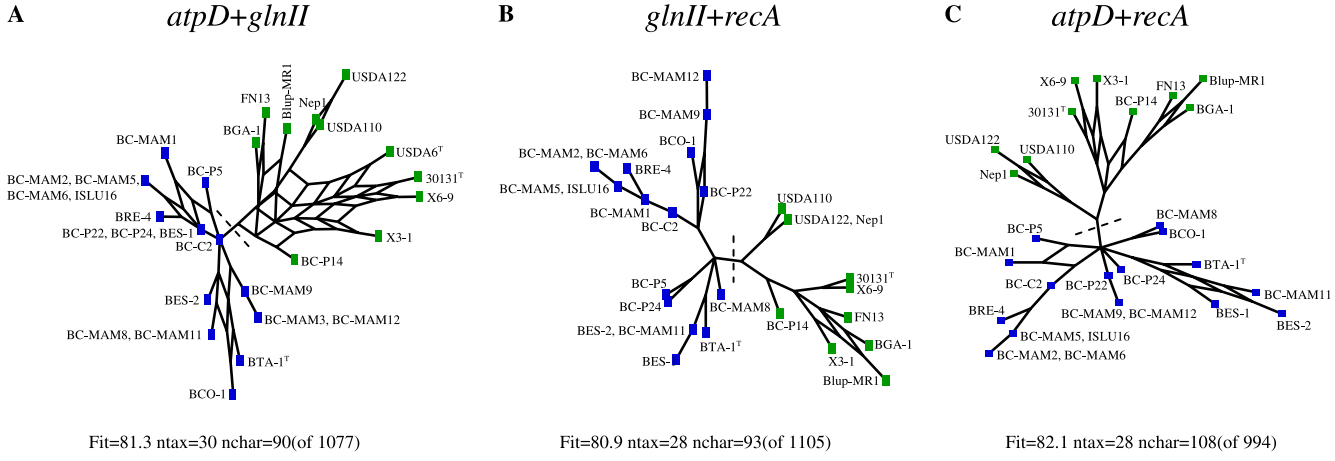

Fig. 6. Hamming distance split decomposition graphs depicting the parallel paths of evolution found among the indicated data partitions from the sister species *B. canariense* and *B. japonicum* (see Fig. 5). Notice that the graphs were displayed with equal edges for a clearer view of the reticulations, and therefore are not proportional to the genetic distance between sequences. A dashed line indicates the longest split found on graphs drawn proportionally to the genetic distance, which separated the sequences corresponding to both species with >90% bootstrap support in all cases (data not shown). The fit parameter values were >80% in all cases, indicating that a substantial fraction of the phylogenetic signals could be depicted on the graphs. The number of taxa and parsimony informative sites are also indicated.

parallel evolutionary paths more clearly; bootstrap analyses significantly resolved both groups and supported many of the reticulations). This finding provides further evidence for genetic differentiation and lack of gene flow between these sympatric species.

4. Discussion

The general objective of this study was to evaluate the adequacy of combined phylogenetic and population genetic inferences for bacterial species delineation at the population-species interface of the genus *Bradyrhizobium*, which has been classically considered a taxonomically difficult group (Sawada et al., 2003; Willems et al., 2001a,b). Our specific objectives were: (1) to obtain a well-resolved phylogeny for the genus; (2) to demonstrate the existence of biotic discontinuities within the genus, particularly between the sympatric sister species *B. canariense* and *B. japonicum*, which can be found in undisturbed ecosystems nodulating the same host plant and therefore have partially overlapped ecological niches (Jarabo-Lorenzo et al., 2003; Vinuesa et al., 1998, 2004); and (3) to identify the key evolutionary forces that provide cohesion to *B. canariense* bv. *genistearum* populations that nodulate endemic Canarian genistoid legumes (ECGLs) in the Canary Islands and nearby continental areas.

4.1. Species-level inferences from atpD, glnII recA, and nifH gene trees

This is the first molecular evolutionary and systematic study of bradyrhizobia that used a multilocus sequence analysis approach for species delineation. It is well established that, in general, phylogenetic methods perform significantly worse when the underlying model of evolution is incorrect, often leading researchers to derive wrong conclusions concerning clade support, or tree topologies (Bruno and Halpern, 1999; Buckley et al., 2001; Kelsey et al., 1999). Appropriate models are also absolutely critical for inferring tempo and mode of the evolutionary process analyzed (Buckley et al., 2001; Leitner et al., 1997), or for choosing among alternative tree topologies (Buckley, 2002). In addition to the use of best-fit models of nucleotide substitution, care was taken to identify and exclude from our analyses sequences causing phylogenetic distortions due to gene mosaicism (Posada and Crandall, 2002; Schierup and Hein, 2000) or other undefined reasons. This greatly improved the topological resolution of each individual gene tree, as well as the overall topological congruence between them, as shown in other studies (Escobar-Páramo et al., 2004).

Perhaps the most interesting finding from the gene tree comparisons was to discover contrasting evolutionary dynamics displayed by core and *sym* loci. Inheritance of the former seems to be dominated by vertical descent with frequent homologous recombinational exchanges within lineages, but not across them. In contrast, horizontal gene transfer was consistently detected for the symbiotic *nifH* locus across species nodulating ECGLs. The same pattern was detected for the *nodC* locus, as reported elsewhere (Vinuesa et al., 2004), which indicates that the evolutionary dynamics of *sym* loci from ECGL isolates is governed by a mixture of vertical and lateral inheritance. This is consistent with the conclusion of a recent report that compared the evolutionary dynamics of *Bradyrhizobium rrs* and *dnaK* with that of *nodA*, *nolL*, and *nodZ* (Moulin et al., 2004). The fact that the *nifH* and *nodC* phylogenies correlated well with the host range (Jarabo-Lorenzo et al., 2003; Vinuesa et al., 2004) was the basis for recognizing and proposing the two *B. japonicum* biovarieties *glycinearum* and *genistearum*, which nodulate soybeans and genistoid legumes, respectively. This finding highlights the independent evolutionary histories and dynamics of adaptive (accessory) and housekeeping (core) loci (Lan and Reeves, 2000; Wernegreen and Riley, 1999) and the importance of carefully choosing adequate core loci for the inference of the underlying species phylogeny (Daubin et al., 2003).

4.2. A well resolved Bayesian species phylogeny and its implications for the molecular systematics of Bradyrhizobium

This is the first molecular systematic study of rhizobia that attempted to infer a species phylogeny based on (partially) congruent sequence partitions, using a Bayesian framework. The model parameters of primary interest were the topology and associated branch lengths, as well as the bipartition posterior probabilities (PPs). Several studies have shown that PPs may be biased by type I errors, i.e., providing too high clade PPs (Cummings et al., 2003; Suzuki et al., 2002). However, other studies (Alfaro et al., 2003; Douady et al., 2003; Erixon et al., 2003) indicate that type I errors are systematically of concern only when oversimplified substitution models are used. Since our ML bootstrap analysis resolved all nodes having $\geqslant 0.95$ PP with $\geqslant 75\%$ BS, we can safely argue that our *glnII* + *recA* species phylogeny is actually well resolved, that the model chosen for Bayesian phylogeny estimation was reasonably complex, and that the PP and BS values for each node may be interpreted as the potential upper and lower bounds of clade reliability, as suggested by Douady et al. (2003).

This species phylogeny has several important implications for the molecular systematics of the genus *Bradyrhizobium*. First, the current *Bradyrhizobium* taxonomy is consistent with it. From the six named species only *B. betae* isolates remain to be positioned on the tree.

Second, it revealed that ECGLs are nodulated by at least four independent evolutionary lineages, namely the sister species *B. canariense* and *B. japonicum*, along with the genospecies α and β (Vinuesa et al., 2004). Third, the photosynthetic bradyrhizobia represent another unnamed lineage. Fourth, using a conservative classification criterion, *B. elkanii* and the photosynthetic strains may be considered distinct basal evolutionary lineages (species) of the genus *Bradyrhizobium* (*sensu lato*), which is in line with the conclusions reached by So et al. (1994) based on fatty acid and *rrs* sequence analyses. Previous suggestions that they may represent new genera (Eaglesham et al., 1990; van Berkum and Eardly, 1998; Willems et al., 2001b) should be considered cautiously at the moment, as these lineages are included in a perfectly supported clade along with the *bona fide* bradyrhizobia in both Bayesian and ML phylogenies. Interestingly, this finding provides compelling evidence that photosynthesis represents an ancestral physiologic trait in the genus *Bradyrhizobium* (*sensu lato*), which is consistent with the old evolutionary hypothesis that bradyrhizobia descend from a photosynthetic ancestor related to *Rhodopseudomonas* (Jarvis et al., 1986). Rooting of the *Bradyrhizobium* spp. clade with the *R. palustris* sequence is consistent with this hypothesis and with ML phylogenies of concatenated AtpD, GSI, RecA, RpoB, CysS, and SerS protein sequences from all α-proteobacterial genomes available to date (Vinuesa, unpublished). However, the relationships of *Afipia* spp., *Agromonas* spp., *Blastobacter* spp. and *Nitrobacter* spp., which are intermingled with bradyrhizobia in *rrs* and ITS phylogenies (van Berkum and Eardly, 2002; Vinuesa et al., 2004; Willems et al., 2001a,b; Willems et al., 2003), need to be critically revised based on the phylogenetic analysis of multiple protein-coding loci and an adequate strain sampling before a revision at a higher taxonomic level is made for these genera. Finally, the current classification of bradyrhizobia in a separate family *Bradyrhizobiaceae*, which is mainly based on *rrs* phylogenies (Sawada et al., 2003), is consistent with their monophyletic grouping in the Bayesian species phylogeny.

4.3. Migration and recombination are key evolutionary forces which provide B. canariense with internal cohesion and shape its population genetic structure

All methods and molecular markers used herein provided significant evidence for the impact of recombination on the population genetic structure and cohesiveness of *B. canariense*. Evidence for intergenic recombination was obtained from the linkage disequilibrium analyses of MLEE data and phylogenetic congruence analyses of the *atpD*, *glnII*, and *recA* sequence partitions. It is well known, however, that parsimony tests of incongruence like the ILD test (Farris et al., 1994) are more likely to confuse true topological incongruence (different evolutionary histories) with systematic error because they lack the inherent ability of model-based methods to accommodate biases in the data such as rate heterogeneity among sites, transitional bias, etc. (Darlu and Lecointre, 2002). In general, noisy data can result in significant test values (Dolphin et al., 2000). However, the same result was obtained by the conservative, ML-based SH test (Buckley, 2002; Shimodaira and Hasegawa, 1999). Therefore, type I errors (wrongly rejecting the true hypothesis of congruence) in the ILD tests due to high noise to signal ratios (Dolphin et al., 2000), or due to marked differences in the evolutionary characteristics of the different alignment matrices (Darlu and Lecointre, 2002); are unlikely in this case. Rather, the evolutionary force generating the phylogenetic incongruence is lateral transfer within populations, which is well known to produce incongruent topologies for bacterial intraspecies relationships based on multilocus sequence analyses (Brown et al., 2002; Dykhuizen and Green, 1991; Escobar-Páramo et al., 2004), as found herein. Split decomposition analysis (Bandelt and Dress, 1992) corroborated the notion that *B. canariense* and *B. japonicum* populations evolve along independent pathways, as reticulations were found only within but never between populations, suggesting that these recombinogenous sister species are well differentiated and therefore have completed a speciation event.

DNA polymorphism analyses of *atpD*, *glnII*, and *recA* sequences provided compelling evidence for intragenic recombination. Coalescent simulations provided quantitative evidence for intermediate to high levels of recombination. This was also revealed by the mean value of the estimated recombination to mutation rate ratios (c/μ), which suggests that the observed population DNA polymorphisms at the corresponding loci are as likely to originate by recombination as by point mutation. These c/μ ratios are about 4–15 times higher than those estimated for core genes of highly clonal *Pseudomonas syringae* strains (Sarkar and Guttman, 2004), but similar to those reported for the *porB* gene of *Neisseria gonorrhoeae*, a panmictic bacterium (Posada et al., 2000). All the neutrality tests (Tajima's D, Fu's D^* and Fu's F_S) failed to reject the neutral evolution hypothesis for these loci (Fu and Li, 1993; Tajima, 1989). As none of these tests nor the powerful R_2 test for population growth (Ramos-Onsins and Rozas, 2002) was significant, and these were performed on three unlinked loci, it is reasonable to conclude that there is no evidence for population growth, hitchhiking or background selection (Fu, 1997; Tajima, 1989), all of which might bias the estimates of C, R or Rm (Hey and Wakeley, 1997; Hudson, 1987; Hudson and Kaplan, 1985). On the other hand, the test statistics (particularly K_{ST}) used to detect genetic differentiation within and between species are more powerful in the presence of recombination (Hud-

son et al., 1992a). This was an important advantage, since the sample sizes were relatively small. All tests consistently indicated the lack of genetic differentiation between insular and continental populations of *B. canariense*, but revealed a highly significant differentiation between *B. canariense* and *B. japonicum* populations. In summary, all population genetic evidence supports the conclusion that both sister species are genetically differentiated and sexually isolated at housekeeping (core) loci, and therefore represent different species in an evolutionary sense.

The F_{ST} test statistic for gene flow, along with the *Nm* estimates revealed high levels of migration between continental and insular populations of *B. canariense*, which is the likely force that precludes their geographic differentiation. This contrasts markedly with the phylogeographic data available for many endemic animal and plant species of the Canaries, which indicate that these organisms speciated on the islands by adaptive radiation and vicariance mechanisms after their colonisation from the nearby continental areas (Francisco-Ortega et al., 1996; Juan et al., 2000). These contrasting phylogeographic patterns clearly reflect the difference in dispersion potential of microbes vs. plants or animals, which results from their radical difference in body sizes (Finlay, 2002; Vinuesa and Silva, 2004). We hypothesize that the most likely transportation vector for bradyrhizobial cells are the dust masses that regularly reach the Canarian Archipelago due to NW African dust storms, which can reach the European and American continents (Griffin et al., 2002; Vinuesa and Silva, 2004). This and other studies (Jarabo-Lorenzo et al., 2003) have revealed that *B. canariense*, *B. japonicum* or *B. yuanmingense* are distributed across different continents and hemispheres, which underlines the need for large strain samples from each lineage and geographic zone to delineate species boundaries precisely, uncover potential internal genetic subdivisions within them, and determine at which geographic scale migration acts as a significant cohesive evolutionary force (for a review on these topics see Vinuesa and Silva, 2004).

4.4. Implications for a bacterial species concept

We consider the finding of genetically differentiated sympatric evolutionary lineages with overlapping ecological niches, such as those represented by the recombinogenous sister species *B. canariense* and *B. japonicum*, as strong evidence for the existence of well delineated bacterial species in ecological time, at least in stable ecosystems. This finding challenges the view expressed by other authors who have questioned the existence of well delineated species in the bacterial world based on the high levels of lateral gene transfer uncovered by comparative genome analyses (Gogarten et al., 2002; Lawrence, 2002). Contradicting that view, however, are other recent and critical studies supporting our conclusion that bacterial species are delineable as "classical Darwinian" evolutionary lineages when neutrally evolving *bona fide* orthologous housekeeping (core) genes are chosen for phylogeny estimation (Daubin et al., 2003; Kurland et al., 2003).

Our REP-PCR genomic fingerprinting data uncovered a high *B. canariense* strain diversity across sampling sites and revealed that epidemic clones are geographically restricted to their sampling sites, where several of them coexist. This may indicate that periodic selection (Levin, 1981) is not as efficient in recurrently purging the complete genetic diversity within populations as recently proposed Cohan (2001, 2002) and therefore may not be the universal cohesiveness factor for bacterial speciation. The fact that the ECGL isolates are not significantly clustered by sampling site is not consistent with Cohan's views. Furthermore, considering the evidence shown herein and elsewhere that *Bradyrhizobium* and other rhizobial species are almost globally distributed (Jarabo-Lorenzo et al., 2003; Vinuesa and Silva, 2004), along with their astronomic effective population sizes and the well known fact that soils and plant rhizospheres are highly heterogeneous habitats, which permit the coexistence of multiple ecotypes in different microniches (Haubold and Rainey, 1996; Bent et al., 2003), there seems to be little chance for selective sweeps to be a strong diversity purging force in metapopulations, at least for neutrally evolving core loci. This argument seems to be particularly strong when high recombination and migration rates are also at play. However, selective sweeps may adequately explain the epidemic structure of symbiotic populations found at very small geographic and time scales, or the extremely low diversity found at the *nifH* locus of *B. japonicum* bv. *glycinearum*, which may be the result of a very recent spread and horizontal acquisition of the *sym* locus by soybean isolates in response to the world-wide distribution of soybean seeds and inoculants in the past century.

Our results also suggest that there is no theoretical or operational justification for the need of an ad hoc species concept for prokaryotes (Rosselló-Mora and Amann, 2001; Stackebrandt et al., 2002). In our view, its major limitation is that it is not well suited to identify the evolutionary forces that provide bacterial species with internal cohesion and shape their population genetic structures (Avise, 2000; Carbone and Kohn, 2001; Lan and Reeves, 2000, 2001; Templeton, 1989; Ward, 1998). These forces should be uncovered in order to define adequate sampling strategies of individuals and molecular markers suitable for the delineation of populations and species as the basis for a truly evolutionary systematics of bacteria.

In summary, our data and conclusions strongly support the proposal of Lan and Reeves (Lan and Reeves, 2001), who stated that a combination of population

genetics and phylogenetics is currently the best theoretical and practical approach to delineate species as natural and discrete lineages in the bacterial world.

Acknowledgments

The authors thank Heidemarie Thierfelder, Eva-Maria Kurz, Marco Antonio Rogel, Ernesto Ormeño, René Hernández, and Scott Bringham for excellent technical assistance. Holger Blasum, Jagdish Ladha, Milagros León-Barrios, Kristina Lindström, Ernesto Ormeño, Peter van Berkum, En Tao Wang, and Anne Willems are warmly thanked for providing reference strains. Paul V. Dunlap, Luis E. Eguiarte, Marcos Pérez Losada, J. Peter W. Young, and two anonymous reviewers are gratefully acknowledged for their critical and constructive comments on the manuscript. Julio Rozas and Rodolfo Salas are acknowledged for their advice on coalescent simulations. This work was partially supported by INCO-DEV ICA-4-CT-2001-10057 and DFG in the SFB395 project A6 from the EU.

Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2004.08.020.

References

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Avise, J.C., 2000. Phylogeography: the History and Formation of Species. Harvard University Press, Cambridge, MA, USA.

Bandelt, H.-J., Dress, A.W.M., 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. 1, 242–252.

Barraclough, T.G., Nee, S., 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16, 391–399.

Barrera, L.L., Trujillo, M.E., Goodfellow, M., Garcia, F.J., Hernandez-Lucas, I., Davila, G., van Berkum, P., Martinez-Romero, E., 1997. Biodiversity of bradyrhizobia nodulating *Lupinus* spp. Int. J. Syst. Bacteriol. 47, 1086–1091.

Bent, S.J., Gucker, C.L., Oda, Y., Forney, L.J., 2003. Spatial distribution of *Rhodopseudomonas palustris* ecotypes on a local scale. Appl. Environ. Microbiol. 69, 5192–5197.

Brown, E.W., Kotewicz, M.L., Cebula, T.A., 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. Mol. Phylogenet. Evol. 24, 102–120.

Bruno, W.J., Halpern, A.L., 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. 16, 564–566.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.

Buckley, T.R., Simon, C., Chambers, G.K., 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. Syst. Biol. 50, 67–86.

Carbone, I., Kohn, L.M., 2001. A microbial population-species interface: nested cladistic and coalescent inference with multilocus data. Mol. Ecol. 10, 947–964.

Chaintreuil, C., Giraud, E., Prin, Y., Lorquin, J., Ba, A., Gillis, M., de Lajudie, P., Dreyfus, B., 2000. Photosynthetic bradyrhizobia are natural endophytes of the African wild rice *Oryza breviligulata*. Appl. Environ. Microbiol. 66, 5437–5447.

Cho, J.C., Tiedje, J.M., 2000. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. Appl. Environ. Microbiol. 66, 5448–5456.

Cohan, F.M., 2001. Bacterial species and speciation. Syst. Biol. 50, 513–524.

Cohan, F.M., 2002. What are bacterial species?. Annu. Rev. Microbiol. 56, 457–487.

Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A., Winka, K., 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52, 477–487.

Darlu, P., Lecointre, G., 2002. When does the incongruence length difference test fail?. Mol. Biol. Evol. 19, 432–437.

Daubin, V., Moran, N.A., Ochman, H., 2003. Phylogenetics and the cohesion of bacterial genomes. Science 301, 829–832.

DeLong, E.F., Pace, N.R., 2001. Environmental diversity of bacteria and archaea. Syst. Biol., 470–478.

Dolphin, K., Belshaw, R., Orme, C.D., Quicke, D.L., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17, 401–406.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Doyle, J.J., Luckow, M.A., 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol. 131, 900–910.

Dykhuizen, D.E., Green, L., 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. 173, 7257–7268.

Eaglesham, A.R.J., Ellis, J.M., Evans, W.R., Fleischmann, D.E., Hungria, M., Hardy, W.F., 1990. The first photosynthetic N_2-fixing *Rhizobium*: characteristics. In: Gresshoff, P.M., Roth, L.E., Stacey, G., Newton, W.L. (Eds.), Nitrogen Fixation: Achievements and Objectives. Chapman & Hall, New York, pp. 805–811.

Erixon, P., Svennblad, B., Britton, T., Oxelman, B., 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52, 665–673.

Escobar-Páramo, P., Sabbagh, A., Darlu, P., Pradillon, O., Vaury, C., Denamur, E., Lecointre, G., 2004. Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. Mol. Phylogenet. Evol. 30, 243–250.

Farris, J.S., Källersjö, J., Kluge, A.G., Bult, C., 1994. Testing significance of congruence. Cladistics 10, 315–319.

Feil, E.J., Spratt, B.G., 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. 55, 561–590.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package). Distributed by the author. Department of Genetics, University of Washington, Seattle.

Finlay, B.J., 2002. Global dispersal of free-living microbial eukaryote species. Science 296, 1061–1063.

Francisco-Ortega, J., Jansen, R.K., Santos-Guerra, A., 1996. Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. Proc. Natl. Acad. Sci. USA 93, 4085–4090.

Francisco-Ortega, J., Jackson, M.T., Santos-Guerra, A., Fernández-Galván, M., Ford-Lloyd, B.V., 1994. The phytogeography of *Chamaecytisus proliferus* (L.fil.) Link complex (*Fabaceae: Genisteae*) in the Canary Islands: a multivariate analysis. Vegetatio 110, 1–17.

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

Fu, Y.-X., Li, W.-H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709.

Galibert, F., Finan, T.M., Long, S.R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R.W., Dreano, S., Federspiel, N.A., Fisher, R.F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R.W., Jones, T., Kahn, D., Kahn, M.L., Kalman, S., Keating, D.H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M.C., Pohl, T.M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F.J., Weidner, S., Wells, D.H., Wong, K., Yeh, K.C., Batut, J., 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293, 668–672.

Gaunt, M.W., Turner, S.L., Rigottier-Gois, L., Lloyd-Macgilp, S.A., Young, J.P., 2001. Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. Int. J. Syst. Evol. Microbiol. 51, 2037–2048.

Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. 19, 2226–2238.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Griffin, D.W., Kellogg, C.A., Garrison, V.H., Shinn, E.A., 2002. The global transport of dust—an intercontinental river of dust, microorganisms and toxic chemicals flows through the Earth's atmosphere. Am. Sci. 90, 228–235.

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

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser., pp. 95–98.

Haubold, B., Rainey, P.B., 1996. Genetic and ecotypic structure of a fluorescent *Pseudomonas* population. Mol. Ecol. 5, 747–761.

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

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275–284.

Hollis, A.B., Kloos, W.E., Elkan, G.E., 1981. DNA: DNA hybridization studies of *Rhizobium japonicum* and related *Rhizobiaceae*. J. Gen. Microbiol. 123, 215–222.

Hudson, R.R., 1987. Estimating the recombination parameter of a finite population model without selection. Genet. Res. 50, 245–250.

Hudson, R.R., Kaplan, N.L., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147–164.

Hudson, R.R., Boos, D.D., Kaplan, N.L., 1992a. A statistical test for detecting geographic subdivision. Mol. Biol. Evol. 9, 138–151.

Hudson, R.R., Slatkin, M., Maddison, W.P., 1992b. Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589.

Huelsenbeck, J.P., Rannala, B., 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science 276, 227–232.

Huelsenbeck, J.P., Bull, J.J., Cunningham, C.W., 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11, 152–157.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Syst. Biol. 51, 673–688.

Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.

Istock, C.A., Duncan, K.E., Ferguson, N., Zhou, X., 1992. Sexuality in a natural population of bacteria—*Bacillus subtilis* challenges the clonal paradigm. Mol. Ecol. 1, 95–103.

Jarabo-Lorenzo, A., Velazquez, E., Perez-Galdona, R., Vega-Hernandez, M.C., Martinez-Molina, E., Mateos, P.E., Vinuesa, P., Martinez-Romero, E., Leon-Barrios, M., 2000. Restriction fragment length polymorphism analysis of 16S rDNA and low molecular weight RNA profiling of rhizobial isolates from shrubby legumes endemic to the Canary islands. Syst. Appl. Microbiol. 23, 418–425.

Jarabo-Lorenzo, A., Pérez-Galdona, R., Donate-Correa, J., Rivas, R., Velázquez, E., Hernández, M., Temprano, F., Martínez-Molina, E., Ruiz-Argüeso, T., León-Barrios, M., 2003. Genetic diversity of bradyrhizobial populations from diverse geographic origins that nodulate *Lupinus* spp. and *Ornithopus* spp. Syst. Appl. Microbiol. 26, 611–623.

Jarvis, B.D.W., Gillis, M., De Ley, J., 1986. Intra- and intergeneric similarities between the ribosomal ribonucleic cistrons of *Rhizobium* and *Bradyrhizobium* species and some related bacteria. Int. J. Syst. Bacteriol. 36, 129–138.

Jordan, D.C., 1982. Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing root nodule bacteria from leguminous plants. Int. J. Syst. Bacteriol. 32, 136–139.

Juan, C., Brent, E., Oromí, P., Hewitt, G.M., 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends Ecol. Evol. 15, 104–109.

Judd, A.K., M., S., Sadowsky, M.J., de Bruijn, F.J., 1993. Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl. Environ. Microbiol. 59, 1702–1708.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., Tabata, S., 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. 9, 189–197.

Käss, E., Wink, M., 1997. Phylogenetic relationships in the Papilionoideae (family Leguminosqe) based on nucleotide sequences of cpDNA (*rbcL*) and ncDNA (ITS 1 and 2). Mol. Phylogenet. Evol. 8, 65–88.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kelsey, C.R., Crandall, K.A., Voevodin, A.F., 1999. Different models, different trees: the geographic origin of PTLV-I. Mol. Phylogenet. Evol. 13, 336–347.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, MA.

Kitagawa, W., Takami, S., Miyauchi, K., Masai, E., Kamagata, Y., Tiedje, J.M., Fukuda, M., 2002. Novel 2,4-dichlorophenoxyacetic acid degradation genes from oligotrophic *Bradyrhizobium* sp. strain HW13 isolated from a pristine environment. J. Bacteriol. 184, 509–518.

Kumar, S., Gadagkar, S.R., 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158, 1321–1327.

Kurland, C.G., Canback, B., Berg, O.G., 2003. Horizontal transfer: a critical view. Proc. Natl. Acad. Sci. USA 100, 9658–9662.

Kurz, W.F.W., LaRue, T.A., 1975. Nitrogenase activity in rhizobia in absence of plant host. Nature 256, 407–408.

Kuykendall, L.D., Saxena, B., Devine, T.E., Udell, S.E., 1992. Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov.. Can. J. Microbiol. 38, 501–505.

Laguerre, G., Nour, S.M., Macheret, V., Sanjuan, J., Drouin, P., Amarger, N., 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, R., Reeves, P.R., 2001. When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol. 9, 419–424.

Lan, R.T., Reeves, P.R., 2000. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol. 8, 396–401.

Lawrence, J.G., 2002. Gene transfer in bacteria: speciation without species?. Theor. Popul. Biol. 61, 449–460.

Leitner, T., Kumar, S., Albert, J., 1997. Tempo and mode of nucleotide substitutions in *gag* and *env* gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. J. Virol. 71, 4761–4770.

Lemmon, A.R., Moriarty, E.C., 2004. The importance of proper model assumptions in Bayesian phylogenetics. Syst. Biol. 53, 265–277.

Levin, B.R., 1981. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. Genetics 99, 1–23.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Maynard-Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Maynard-Smith, J., Smith, N.H., O'Rourke, M., Spratt, B.G., 1993. How clonal are bacteria?. Proc. Natl. Acad. Sci. USA 90, 4384–4388.

Mesa, S., Velasco, L., Manzanera, M.E., Delgado, M.J., Bedmar, E.J., 2002. Characterization of the *norCBQD* genes, encoding nitric oxide reductase, in the nitrogen fixing bacterium *Bradyrhizobium japonicum*. Microbiology 148, 3553–3560.

Moulin, L., Béna, G., Boivin-Masson, C., Stepkowski, T., 2004. Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. Mol. Phylogenet. Evol. 30, 720–732.

Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358–364.

Nylander, J.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.

Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299–304.

Oda, Y., Star, B., Huisman, L.A., Gottschal, J.C., Forney, L.J., 2003. Biogeography of the purple nonsulfur bacterium *Rhodopseudomonas palustris*. Appl. Environ. Microbiol. 69, 5186–5191.

Polhill, R.M., 1994. Complete synopsis of legume genera, Phytochemical Dictionary of the Leguminoseae. Chapman & Hall, London, pp. xlix-Ivii.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. USA 98, 13757–13762.

Posada, D., Crandall, K.A., 2002. The effect of recombination on the accuracy of phylogeny estimation. J. Mol. Evol. 54, 396–402.

Posada, D., Crandall, K.A., Nguyen, M., Demma, J.C., Viscidi, R.P., 2000. Population genetics of the *porB* gene of *Neisseria gonorrhoeae*: different dynamics in different homology groups. Mol. Biol. Evol. 17, 423–436.

Rademaker, J.L., Hoste, B., Louws, F.J., Kersters, K., Swings, J., Vauterin, L., Vauterin, P., de Bruijn, F.J., 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA–DNA homology studies: *Xanthomonas* as a model system. Int. J. Syst. Evol. Microbiol. 50 (Pt 2), 665–677.

Rademaker, J.L.W., Louws, F.J., Rossbach, U., Vinuesa, P., de Bruijn, F.J., 1998. Computer-assisted pattern analysis of molecular fingerprints and database construction. In: Akkermans, A.D.L., van Elsas, J.E., de Bruijn, F.J. (Eds.), Molecular Ecology Manual. Kluwer Academic Publishers, Dordrecht, pp. 1–33, Chapter 37.31.33.

Ramos-Onsins, S.E., Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. Mol. Biol. Evol. 19, 2092–2100.

Rivas, R., Willems, A., Palomo, J.L., García-Benavides, P., Mateos, P.F., Martinez-Molina, E., Gillis, M., Velázquez, E., 2004. *Bradryrhizobium betae* sp. nov. isolated from roots of *Beta vulgaris* affected by tumor-like deformations. Int. J. Syst. Evol. Microbiol. (in press).

Ronquist, F., Huelsenbeck, J.P., 2003. MrBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rosenberg, N.A., 2002. The probability of topological concordance of gene trees and species trees. Theor. Popul. Biol. 61, 225–247.

Rosselló-Mora, R., Amann, R., 2001. The species concept for prokaryotes. FEMS Microbiol. Rev. 25, 39–67.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Saito, A., Mitsui, H., Hattori, R., Minamisawa, K., Tsutomu Hattori, T., 1998. Slow-growing and oligotrophic soil bacteria phylogenetically close to *Bradyrhizobium japonicum*. FEMS Microbiol. Ecol. 25, 277–286.

Salminen, M.O., Carr, J.K., Burke, D.S., McCutcham, F.E., 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res. Hum. Retroviruses 11, 1423–1425.

Sarkar, S.F., Guttman, D.S., 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. Appl. Environ. Microbiol. 70, 1999–2012.

Sawada, H., Kuykendall, L.D., Young, J.M., 2003. Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. J. Gen. Appl. Microbiol. 49, 155–179.

Sawyer, S., 1998. Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–538.

Schierup, M.H., Hein, J., 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156, 879–891.

Segovia, L., Piñero, D., Palacios, R., Martínez-Romero, E., 1991. Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. Appl. Environ. Microbiol. 57, 426–433.

Selander, R.K., Caugant, D.A., Ochman, H., Musser, J.M., Gilmour, M.N., Whittam, T.S., 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. 51, 873–884.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Silva, C., Eguiarte, L.E., Souza, V., 1999. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar phaseoli in a traditionally managed locality in Mexico. Mol. Ecol. 8, 277–287.

Silva, C., Vinuesa, P., Eguiarte, L.E., Martinez-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. Appl. Environ. Microbiol. 69, 884–893.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy. W.H. Freeman, San Francisco.

So, R.B., Ladha, J.K., Young, J.P., 1994. Photosynthetic symbionts of *Aeschynomene* spp. form a cluster with bradyrhizobia on the basis of fatty acid and rRNA analyses. Int. J. Syst. Bacteriol. 44, 392–403.

Souza, V., Nguyen, T.T., Hudson, R.R., Piñero, D., Lenski, R.E., 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex?. Proc. Natl. Acad. Sci. USA 89, 8389–8393.

Spaink, H.P., Kondorosi, A., Hooykaas, P.J.J., 1998. The Rhizobiaceae. Kluwer Academic Publishers, Dordrecht.

Sprent, J.I., 2001. Nodulation in legumes. Royal Botanic Gardens, Kew, p. 146.

Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kampfer, P., Maiden, M.C., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B., 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 52, 1043–1047.

Sullivan, J.T., Eardly, B.D., van Berkum, P., Ronson, C.W., 1996. Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. Appl. Environ. Microbiol. 62, 2818–2825.

Sullivan, J.T., Patrick, H.N., Lowther, W.L., Scott, D.B., Ronson, C.W., 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal and symbiotic gene transfer in the environment. Proc. Natl. Acad. Sci. USA 92, 8995–8999.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16138–16143.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods (software). Sinuauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddel, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, pp. 407–514.

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

Tan, Z., Hurek, T., Vinuesa, P., Muller, P., Ladha, J.K., Reinhold-Hurek, B., 2001. Specific detection of *Bradyrhizobium* and *Rhizobium* srains colonizing rice (*Oryza sativa*) roots by 16S–23S ribosomal DNA intergenic spacer- targeted PCR. Appl. Environ. Microbiol. 67, 3655–3664.

Templeton, A.R., 1989. The meaning of species and speciation: a genetic perspective. In: Otte, D., Endler, J.A. (Eds.), Speciation and its Consecuences. Sinauer, Sunderland, Massachusetts, pp. 3–27.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Torsvik, V., Ovreas, L., Thingstad, T.F., 2002. Prokaryotic diversity—magnitude, dynamics, and controlling factors. Science 296, 1064–1066.

Turner, S.L., Young, J.P., 2000. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol. Biol. Evol. 17, 309–319.

Ueda, T., Suga, Y., Yahiro, N., Matsuguchi, T., 1995. Remarkable N_2-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of *nifH* gene sequences. J. Bacteriol. 177, 1414–1417.

van Berkum, P., Eardly, B.D., 1998. Molecular evolutionary systematics of the *Rhizobiaceae*. In: Spaink, H.P., Kondorosi, A., Hooykaas, P.J.J. (Eds.), The *Rhizobiaceae*: Molecular Biology of Plant-Associated Bacteria. Kluwer Academic Publishers, Dordrecht, pp. 1–24.

van Berkum, P., Fuhrmann, J.J., 2000. Evolutionary relationships among the soybean bradyrhizobia reconstructed from 16S rRNA gene and internally transcribed spacer region sequence divergence. Int. J. Syst. Evol. Microbiol. 50, 2165–2172.

van Berkum, P., Eardly, B.D., 2002. The aquatic budding bacterium *Blastobacter denitrificans* is a nitrogen-fixing symbiont of *Aeschynomene indica*. Appl. Environ. Microbiol. 68, 1132–1136.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., Swings, J., 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol. Rev. 60, 407–438.

Vinuesa, P., Silva, C., 2004. Species delineation and biogeography of symbiotic bacteria associated with cultivated and wild legumes. In: Werner, D. (Ed.), Biological Resources and Migration. Springer Verlag, Berlin, pp. 143–161.

Vinuesa, P., Rademaker, J.L.W., de Bruijn, F.J., Werner, D., 1998. Genotypic characterization of *Bradyrhizobium* strains nodulating endemic woody legumes of the Canary Islands by PCR-restriction fragment length polymorphism analysis of genes encoding 16S rRNA (16S rDNA) and 16S–23S rDNA intergenic spacers, repetitive extragenic palindromic PCR genomic fingerprinting and partial 16S rDNA sequencing. Appl. Environ. Microbiol. 64, 2096–2104.

Vinuesa, P., Rademaker, J.L.W., de Bruijn, F.J., Werner, D., 1999. Characterization of *Bradyrhizobium* spp. strains by RFLP analysis of amplified 16S rDNA and rDNA intergenic spacer regions. In: Martínez, E., Hernández, G. (Eds.), Highlights on Nitrogen Fixation. Plenum Publishing Corporation, New York, pp. 275–279.

Vinuesa, P., León-Barrios, M., Silva, C., Willems, A., Jarabo-Lorenzo, A., Pérez-Galdona, R., Werner, D., Martínez-Romero, E., 2004. *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont from the nodules of endemic woody legume species (Papilionoideae:Genisteae) growing in the Canary Islands along with *B. japonicum* bv. genistearum, *Bradyrhizobium* genospecies α and *Bradyrhizobium* genospecies β. Int. J. Syst. Evol. Microbiol. (in press. http://dx.doi.org/10.1099/ijs.0.63292-0).

Ward, D.M., 1998. A natural species concept for prokaryotes. Curr. Opin. Microbiol. 1, 271–277.

Wernegreen, J., Riley, M.A., 1999. Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. Mol. Biol. Evol. 16, 98–113.

Whittam, T.S., 1990. ETDIV and ETCLUS programs. Pennsylvania State University.

Willems, A., Coopman, R., Gillis, M., 2001a. Comparison of sequence analysis of 16S–23S rDNA spacer regions, AFLP analysis and DNA–DNA hybridizations in *Bradyrhizobium*. Int. J. Syst. Evol. Microbiol. 51, 623–632.

Willems, A., Coopman, R., Gillis, M., 2001b. Phylogenetic and DNA–DNA hybridization analyses of *Bradyrhizobium* species. Int. J. Syst. Evol. Microbiol. 51, 111–117.

Willems, A., Munive, A., de Lajudie, P., Gillis, M., 2003. In most *Bradyrhizobium* groups sequence comparison of 16S–23S rDNA internal transcribed spacer regions corroborates DNA–DNA hybridizations. Syst. Appl. Microbiol. 26, 203–210.

Xia, X., Xie, Z., 2001. DAMBE: Software package for data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y., 2002. An index of sustitution saturation and its application. Mol. Phylogenet. Evol. 26, 1–7.

Xu, L.M., Ge, C., Cui, Z., Li, J., Fan, H., 1995. *Bradyrhizobium liaoningense* sp. nov., isolated from the root nodules of soybeans. Int. J. Syst. Bacteriol. 45, 706–711.

Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–372.

Yao, Z.Y., Kan, F.L., Wang, E.T., Wei, G.H., Chen, W.X., 2002. Characterization of rhizobia that nodulate legume species of the genus *Lespedeza* and description of *Bradyrhizobium yuanmingense* sp. nov. Int. J. Syst. Evol. Microbiol. 52, 2219–2230.

Zhang, X., Nick, G., Kaijalainen, S., Terefework, Z., Paulin, L., Tighe, S.W., Graham, P.H., Lindström, K., 1999. Phylogeny and diversity of *Bradyrhizobium* strains isolated from the root nodules of peanut (*Arachis hypogaea*) in Sichuan, China. Syst. Appl. Microbiol. 22, 378–386.

Zhou, J., Bowler, L.D., Spratt, B.G., 1997. Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. Mol. Microbiol. 23, 799–812.

Supplementary Material of Vinuesa et al. 2004 (MPEV501)

Table S1. GenBank accession numbers for the sequences generated in this study, and results of the incongruence length difference test*.

Isolate or Strain[a]	*atpD*	*glnII*	*recA*	*nifH*
Isolates from genistoid legumes				
#*Bradyrhzobium* genosp. α BC-C1	AY3867351 [r]	AY386761 [r]	AY591540 [a,]	AY38678
#*B. canariense* BC-C2	AY386736 [r]	AY386762	AY591541 [a]	AY38678
#*B. canariense* BC-P5	AY386737 [r]	AY386763 [r]	AY591542 [a,]	
#*Bradyrhzobium* genosp. β BC-P6	AY653751 [g, r]		AY591543 [a]	
#*B. japonicum* BC-P14	AY386746	AY386771	AY591544	
#*B. canariense* BC-P22	AY386740	AY386766	AY591545	
#*B. canariense* BC-P24	AY653743 [r]			
#*B. canariense* BES-1	AY386738 [r]	AY386764	AY591548 [a]	
#*B. canariense* BES-2	AY653753	AY599109	AY591549	AY59909
#*B. canariense* BCO-1	AY653754 [r]	AY599110 [r]	AY591550 [a,]	
#*B. japonicum* BGA-1	AY386747	AY386772	AY591558	AY38678
#*Bradyrhzobium* genosp. β BRE-1	AY386748 [g, r]	AY599112 [a]	AY591551 [a]	
#*B. canariense* BRE-4	AY653755	AY599111	AY591552	
#*B. canariense* BTA-1^T	AY386739	AY386765 [r]	AY591553 [g]	AY38678
#*B. canariense* BC-MAM1	AY386741	AY386767	AY591546	
#*B. canariense* BC-MAM2	AY653744	AY599103		
#*B. canariense* BC-MAM3	AY386742	AY386768		
#*B. canariense* BC-MAM5	AY386743	AY386769	AY591547	
#*B. canariense* BC MAM6	AY653745			
#*B. canariense* BC-MAM7		AY599104		
#*B. canariense* BC-MAM8	AY653746 [r]	AY599105		
#*B. canariense* BC-MAM9	AY653747			
#*B. canariense* BC-MAM11	AY386744	AY386770		
#*B. canariense* BC-MAM12	AY653749			
#*Bradyrhzobium* genosp. β BC-	AY386756 [g]	AY386778 [a]		
#*Bradyrhzobium* genosp. β BC-	AY386757 [r]	AY386779	AY591554 [a]	
#*B. canariense* ISLU16	AY386745		AY591576	
#*B. japonicum* Blup-MR1	AY386751	AY386774	AY591559	AY38678
#*B. japonicum* FN13	AY386749	AY386773	AY591560	AY38678
Bradyrhizobium sp. CICS70	AY386750 [g]	AY599113 [a]	AY591561	
Reference strains				
B. japonicum DSMZ30131^T	AY386754	AY591555	AY591555	AY59908
B. japonicum USDA122	AY653767	AY386777	AY591562	AY59908
#*B. japonicum* X3-1	AY653764	AY599120	AY591556	AY59908
#*B. japonicum* X6-9	AY653765	AY599121	AY591557	AY59908
#*B. japonicum* Nep1	AY386753	AY386776	AY591563	AY59908
B. elkanii USDA76^T	AY386758 [r]	AY599117	AY591568 [a]	
B. elkanii USDA46	AY653766 [r]	AY599125	AY591576 [a]	AY59909
B. elkanii USDA94	AY386759	AY599118	AY591569	AY59909
B. liaoningense LMG18230^T	AY386752 [g, r]	AY386775 [a]	AY591564 [a]	AY59908
#*B. liaoningense* Spr3-7	AY653763 [g, r]	AY599116 [a]	AY591574 [a]	AY38678
B. yuanmingense CCBAU10071^T	AY386760	AY386780	AY591566	AY59909
#*B. yuanmingense* LTMR28	AY653768 [g]	AY599115 [a]	AY591573	
#*B. yuanmingense* TAL760	AY653761	AY599114	AY591565	
#*Bradyrhzobium* genosp. α	AY653762 [r]		AY591567 [a,]	
Photosynthetic strains				
Bradyrhizobium sp. BTAi1		AY599123	AY591570	
Bradyrhizobium sp. IRBG127		AY599122 [r]	AY591571 [g]	
Bradyrhizobium sp. IRBG231		AY599124 [r]	AY591572 [g]	

*The superindices [a, g, r] correspond to the first letter of the *atpD*, *glnII* and *recA* loci and indicate sequences pairs that presented significant values in the ILD test.

[a] # Isolates classified in this study.

Table S2. Pairwise incongruence length difference tests[a].

Locus	*glnII*	*recA*	*nifH*
atpD	*P* = 0.001 (n = 41)	*P* = 0.001 (n = 43)	*P* = 0.001 (n = 13)
glnII		*P* = 0.003 (n = 43)	*P* = 0.001 (n = 13)
recA			*P* = 0.001 (n = 13)

[a] *P* values for 1,000 random partitions for the number of sequences indicated in parenthesis. Branch & Bound searches were used when the number of sequences was ≤20, whereas heuristic searches were for 20 or more sequence comparisons. Only variable sites were included in the analyses.

Figure S1. Evidence for intragenic mosaicism in the *atpD* sequences of *Bradyrhizobium* strains BC-K1 (**panel I**) and BTAi1 (**panel II**) obtained by different recombination detection methods (see main text for details and references). Subpanels **A** shows the location of the mosaic segments predicted by CHIMAERA. Subpanels **B** show the corresponding confirmatory analyses performed with BootScanning. The dotted line indicates the 90% bootstrap support cutoff level. The HKY85 model was used for the corresponding phylogeny estimations. Subpanels **C** highlight the sequence blocks of polymorphic sites affected by recombination events. Subpanels **D** show the summary statistics of the confirmatory analyses performed on the initial CHIMAERA prediction, indicating the predicted daughter and parental sequences, breakpoint coordinates, and a table of the significant events detected by different programs, along with their average *P* values. Based on these results, the sequences from Strains BTAi1, BC-MK1 and TAL760 were excluded from subsequent analyses (see Fig. 3).

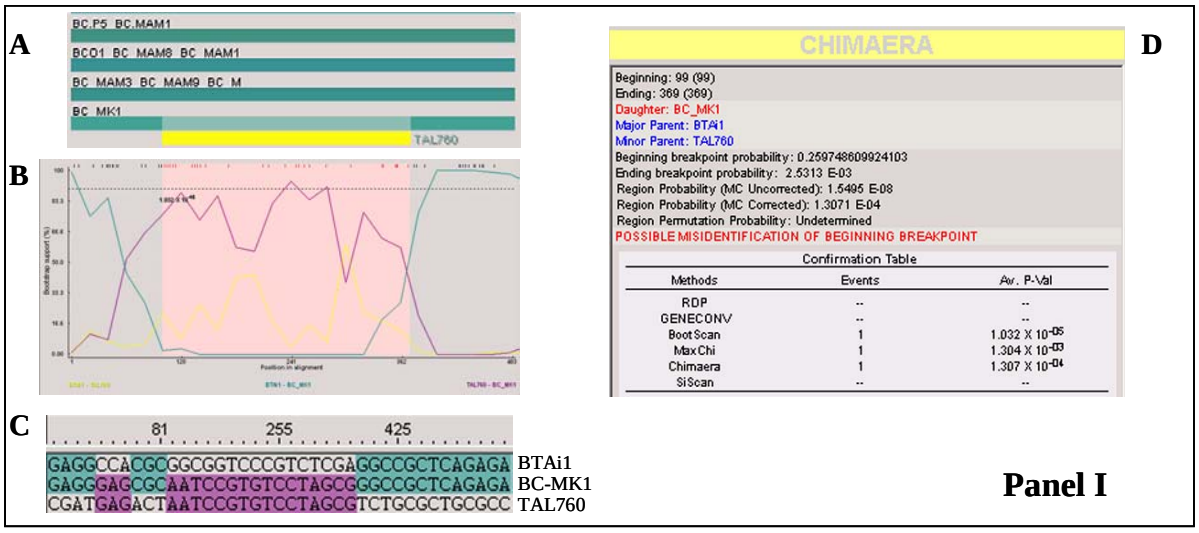

Figure S2. Evidence for the convergence and adequate mixing of the three MCMC runs and details of the substitution model used to infer the Bayesian species phylogeny based on partially congruent *glnII+recA* partitions shown in Figure 5 of the main text. Panel **A**) shows the marginal –lnL generation plots of for the three MCMC runs (over $3x10^6$ generations and sampled every 100th). Panel **B**) shows the correlation plot for clade posterior probabilities of the trees inferred from runs 2 and 3, and the correlation corresponding values for all possible pairwise comparisons among runs. Together, the analyses shown in the previous panels provide good evidence for the convergence and good mixing of the chains. Panels **C**, **D**, **E**, **F** and **G** show the mean and 95% credible intervals for the corresponding substitution model parameter estimates obtained for the concatenated *glnII+recA* sequences, partitioned by gene and by codon position (site-specific rate). The nucleotide frequencies are denoted by pi, the transition/transversion rate ratio by kappa; α is the shape parameter of the gamma distribution used to model among-site rate variation; pinv are the invariant sites; m is the rate multiplier for each codon position. The numbers 1, 2 and 3 correspond to the 1st, 2cnd and 3rd codon positions of the *glnII* partition, whereas numbers 4, 5 and 6 represent the corresponding codon positions of the *recA* sequence.

