Chapter 7
Evolutionary Genomics of the Nitrogen-Fixing Symbiotic Bacteria

Víctor González(✉), Luis Lozano, Santiago Castillo-Ramírez, Ismael Hernández González, Patricia Bustos, Rosa I. Santamaría, José L. Fernández, José L. Acosta, and Guillermo Dávila

7.1 Introduction

Soil contains the most complex communities of microorganisms (Tringe et al. 2005). The earth's global ecology depends largely on the metabolic activities of different soil bacteria. One of these processes is biological nitrogen fixation. Nitrogen from the atmosphere is made available to plants either by free-living bacteria or symbionts associated with leguminous plants. There are ample bacterial species that are able to establish nitrogen fixing symbiosis. Ordinarily known as rhizobia, they are grouped in different taxonomic families of the α-proteobacteria classified as Rhizobiaceae, Phylobacteriaceae and Bradyrhizobiaceae (Table 7.1) (Garrity et al. 2002). Recently, it was reported that some strains of *Burkholderia*, a member of the β-proteobacteria is also able to nodulate tropical legumes like *Aspalathus carnosa* (Moulin et al. 2001). The fact that the common nodulation genes *nodA* and *nodB* of *Burkholderia* spp STM678 are phylogenetically close to *nod* genes of rhizobia suggests a horizontal gene transfer mechanism for acquisition of the nodulating ability (Moulin et al. 2001).

Many years of research have been dedicated to different aspects of nitrogen fixation in symbiosis. This has given a good general view of the process, but it is far from complete (Palacios and Newton 2005). Nowadays, the ability to sequence the complete genome of almost any organism has produced an integral view of the physiology and evolution of the bacterial cell (Dávila and Palacios 2005). Several genomes of rhizobia and related bacteria have been

V. González
Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México Av. Universidad N/C, Col. Chamilpa CP 62210. Apdo. Postal 565-A, Cuernavaca, Morelos, México
e-mail: vgonzal@ccg.unam.mx

C.S. Nautiyal, P. Dion (eds.) *Molecular Mechanisms of Plant and Microbe Coexistence.* Soil Biology 15, DOI: 10.1007/978-3-540-75575-3

Table 7.1 General features of the complete genomes of nitrogen fixing symbiotic bacteria

Family	Rhizobiaceae			Phylo-bacteriaceae	Brady-rhizobiaceae
Features	*R. etli*	*R. leguminosarum*	*S. meliloti*	*M. loti*	*B. japonicum*
Size, bp	6,535,229	7,751,309	6,691,694	7,036,071	9,105,828
Number of replicons	1cc*, 6p**	1cc, 6p	1cc, 2p	1cc, 2p	1cc
GC average %	60.54%	60.86%	62.10%	60%	64.10%
Ribosomal RNA operons	3	3	3	2	1
tRNAs	50	52	54	50	50
Total CDS	6,034	7,263	6,204	6,752	8,317
CDS in functional classes	(71%)	–	3,703 (60%)	3,675 (54%)	4,348 (52%)
Hypothetical CDS	1,389 (23%)	–	1,993 (32%)	1,423 (21%)	2,506 (30%)
Orphan CDS	358 (6%)	–	508 (8%)	1,654 (25%)	1,463 (18%)
Transcriptional regulators	536 (8%)	–	539 (7.4%)	539 (7.4%)	567 (6.8%)
Transporters	837 (13.7%)	816	744 (12%)	764 (12%)	752 (9%)
External elements	157 (2.6%)		136 (2.2%)	171 (2.5%)	167 (2%)
Sigma subunits	23	16	13	23	23

*Circular chromosome

**Plasmids

sequenced, and others are in process. To date, complete sequences are available for *Mesorhizobium loti* MAFF303099 (Kaneko et al. 2000), *Sinorhizobium meliloti* 1021 (Galibert et al. 2001), *Bradyrhizobium japonicum* USDA 110 (Kaneko et al. 2002), *Rhizobium etli* CFN42 (González et al. 2003, 2006), and *Rhizobium leguminosarum* 3841 (Young et al. 2006). There is also information about the symbiotic plasmid of *Rhizobium* spp NGR234 (Freiberg et al. 1997) and the chromosomal symbiotic island of *M. loti* R7A (Sullivan et al. 2002). All of these rhizobia are of great economic importance because their hosts are domestic legumes like alfalfa, pea and beans. In addition, very recently it was reported the complete genome sequence of two strains of *Bradyrhizobium* spp that make stem nodules in the aquatic leguminous plant *Aeschynomene*. Remarkably, no canonical *nod*ABC genes neither typical lipochitoligosaccarides Nod factors are present in these strains (Giraud et al. 2007). This chapter is aimed at highlighting the main features of the complete genomes of nitrogen fixing symbiotic rhizobia available so far, concurrently with genome comparisons and discussion on the current views on the evolution of these complex genomes.

7.2 Origin, Taxonomy and Phylogeny

Several important reviews concerning the diversity and phylogenetic relationships of rhizobia have been published during the recent years (Lloret and Martínez-Romero 2005; Martínez-Romero 2003; Sessitsh et al. 2002). Analyses and simulations of the earth's early atmosphere indicate that nitrogen fixation is an ancient function estimated to have originated in the archean period about 3000 million years ago (MYA) (Kasting and Siefert 2001; Navarro-González et al. 2001). The scattered distribution of *nif* genes among different species of bacteria and archea suggests two possible scenarios about the origin. As an ancient function, nitrogen fixation may have originated from the last common ancestor, and was then inherited by all species but not all of these retained the trait. Alternatively, lateral transfer of *nif* genes among lineages may have resulted in the same phylogenetic pattern. Nevertheless, this is still a controversial issue.

Estimates of the divergence dates of rhizobia based on paralogous glutamine synthetase genes (GSI and GSII) suggest that fast-growing rhizobia lineages were established 203–324 MYA (Turner and Young 2000). Bradyrhizobia, the slow-growing rhizobia, is the more ancient branch calculated to have diverged about 507–553 MYA (Turner and Young 2000). In contrast, nodulation may have emerged when terrestrial plants appeared, 400 MYA and flourished as symbiosis with the expansion of the Leguminosae family 100 MYA (Lloret and Martínez-Romero 2005).

Rhizobia diversity has been assessed by a variety of methods. All indicate that rhizobia are highly heterogeneous since they differ in growth rates, biosynthetic pathways, catabolic activities, habitats, plasmid content, and the lipo and exopolysacharides structures. Phylogenetic reconstructions fail to place rhizobia as a coherent clade. Instead, phylogenies locate rhizobia species intermingled with non-symbiotic species. Actually, seven different symbiotic genera are recognized among the α-proteobacteria, *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Azorhizobium*, *Devosia,* and *Methylobacterium*. The whole genome sequence of model species of four genera has been achieved (Table 7.1). These are included in the order Rhizobiales for which a phylogeny based on the complete genomes available is shown here (Fig. 7.1). The tree constructed by using 636 orthologous proteins located in the chromosome matches very well phylogenetic reconstructions based on 16 S ribosomal RNA genes. As seen in Fig. 7.1, symbiotic species are in the same evolutionary branches of species without symbiotic traits. Moreover, these non-symbiotic species have very different lifestyles that in some cases contrast with symbiosis. More specifically, *B. japonicum* is related to *Rhodopseudomonas palustris*, a photoautothrophic and chemoautothrophic organism; *S. meliloti*, *R. etli* and *R. leguminosarum* are near relatives of the plant pathogen *A. tumefaciens*; and *M. loti* seems to be in the same branch that leads to mammal pathogens *Bartonella* and *Brucella*. Such distribution indicates that

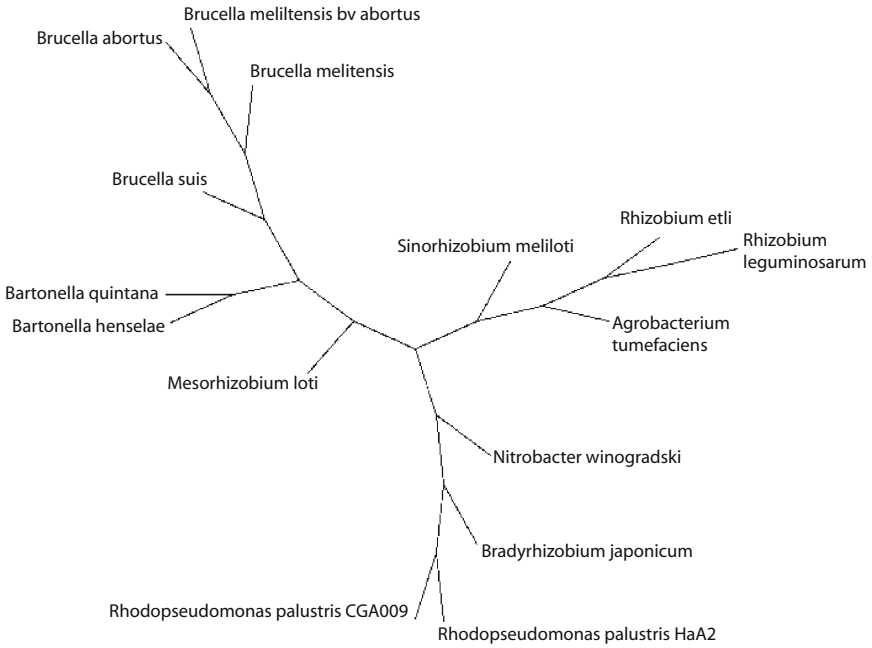

Fig. 7.1 Phylogenetic relationships among 15 species of the order Rhizobiales for which complete genomes are available. The unrooted tree was constructed by using 636 concatenated orthologous proteins common to all the species and parsimony methods (Protpars program of the Phylip package). Orthologous proteins were defined as reciprocal best-hits between pairs of species. Branch length is not equivalent to evolutionary distance

symbiosis, once acquired, has evolved in different and divergent genomic backgrounds.

7.3 Genome Structure

Rhizobia genomes are partitioned into several replicons, variable in size and number (Jumas-Bilak et al. 1998). A circular chromosome contains the majority of the essential genes, whereas plasmids encode diverse functions that are considered auxiliary. The exception to this arrangement is *B. japonicum* which has only one replicon. *A. tumefaciens,* closely related to *Rhizobium* genus, has an unusual genome organization including a linear chromosome (Allardet-Servent et al. 1993). Differences in genome architectures could be explained by rearrangements, horizontal transfer and cointegration of replicons. Some recent experiments have shown that *S. meliloti* and *Rhizobium* spp NGR234 could fuse their replicons into a single molecule without altering the symbiotic or growth properties of the strains (Guo

et al. 2003; Mavingui et al. 2002). However, this arrangement is unstable and reverts to its original state after several replications (Guo et al. 2003). It has been proposed that a subdivided genome might confer adaptive advantages to rhizobia allowing the redistribution of dispensable genes to cope with challenging environments (González et al. 2006).

Plasmids have long been considered unessential molecules, or, as proposed by Campbell, needed only occasionally under certain conditions during the life of the bacterium (Campbell 1981). Genome sequences have revealed that some genes that are essential in different bacteria, in rhizobia are present in plasmids. For instance, the *minCDE* genes required for cell division in *E. coli* have been found on the pSymB of *S. meliloti*, the p42e of *R. etli*, the pRL11 of *R. leguminosarum,* and on the linear chromosome of *A. tumefaciens* and chromosome II of *Brucella*. Recent genetic analysis of the *min*CDE genes of *S. meliloti* has shown that they are dispensable for cell viability. Mutants in *min*CDE do not alter the growth properties and only mutations in *min*E produce branched cells and diminished symbiotic performance (Cheng et al. 2007). Undoubtedly, plasmids represent an important source of variation in natural rhizobia population, and we know only a fraction of the functions encoded by them.

Almost all plasmids found in rhizobia replicate using RepABC proteins, a system only found in α-proteobacteria (Ramírez-Romero et al. 1997). The linear chromosome of *A. tumefaciens* and chromosome II of *Brucella* species also have this type of replicator. The proteins RepA and RepB are involved in partition and RepC is the initiator protein for replication (Ramírez-Romero et al. 2000). Because of the high number of plasmids present in every rhizobia isolated from the field, efficient incompatibility systems are expected to be operating (Soberón et al. 2004). Necessarily, plasmids that coexist in the same genomic background belong to distinct incompatibility groups. They could have originated by duplication and divergence of the RepABC system or, as proposed by Cevallos and colleagues, they have moved among different species by horizontal transfer evolving independently (Cevallos et al. 2002).

The complexity of the rhizobial genomes raises important questions about advantages related to such large genomes. How have the replicons present in the same strain coevolved and coadapted? What is the origin of multiple replicons? Some insights on these matters are emerging from the analysis of *R. etli* and *R. leguminosarum*, which possess the most complex system of replicons of any sequenced genome so far (González et al. 2006; Young et al. 2006). *R. etli* contains six plasmids, four of them showing similar features such as GC content, codon usage, and comparable distribution of groups of orthologous genes (COGs) that resemble the chromosome. The other two plasmids, called the p42a and p42d (pSym) differ completely in these features and also harbor most of the ISs found in the genome (González et al. 2006). Furthermore, bioinformatic analyses of the predicted protein associations among replicons show that p42a and p42d are the replicons that are connected least with the rest of the genome (González et al. 2006). These data are consistent with the notion that plasmids p42a and p42d were acquired recently, while the rest of the plasmids share long-term coevolution

with the chromosome. Like that of *R. etli*, the genome of *R. leguminosarum* is subdivided into seven replicons. By means of analysis of the nucleotide composition and quartop phylogenies, Young et al. (2006) have shown that the *R. leguminosarum* genome can be formed by two components: a "core", which is higher in GC and mainly placed in the chromosome, and an "accessory" component, lower in GC and located on the plasmids and chromosomal islands. This observation leads to the suggestion that horizontal gene transfer by phages and plasmids may have been occurred frequently in the history of *R. leguminosarum* and related bacteria.

7.4 Symbiotic Genome Compartments

The majority of the genes needed for establishing symbiosis are in specific genome compartments (Symbiotic Genome Compartments, SGCs), either in plasmids or in islands integrated in the chromosome (González et al. 2003). They are very heterogeneous in size and gene content and represent a mosaic of genes possibly assembled from different sources. Comparative analysis of the symbiotic plasmids pNGR234a of *Rhizobium* NGR234 spp, pSymA of *S. meliloti*, p42d of *Rhizobium etli* CFN42, and the symbiotic islands of *Mesorhizobium loti* and *Bradyrhizobium japonicum* has evidenced the lack of synteny. As few as 20 homologous genes are common among the SGCs already compared. These include the *nodABC*, *nodIJ*, *nodD*, *nifHDKENXAB*, *fixABCX, fdxN,* and *fdxB*. A few other genes like *nodN*, *fixK*, *fixNOQP*, *fixGHIS* and *fixL* are also common symbiotic genes, but they are not always in SGCs. In contrast, there are more than 30 genes involved in symbiosis that are particular to certain rhizobial species (*nodP, nodM, nodU, nod0, noeI, noeK, noeL, nolR, nolG, nfeD*).

Current ideas about the origin of nodulation point to a probable recruitment of genes of the biosynthetic pathways of lipids and exopolysacharides. In fact, some nodulation genes like *nodM* and *nodG* are recent paralogs of the housekeeping *glsM* and *fabG* genes which participate in the biosynthesis of glucosamine and fatty acids respectively (López-Lara and Geiger 2001; Marie et al. 1992). Other nodulation proteins encoded by *nodL* and *nodJ* have been shown to be homologous to proteins for the transportation of capsular polysaccharides in gram-negative bacteria (Vázquez et al. 1993). In *R. etli*, nodulation genes are included into large clusters of genes coding for components of the external surface of the bacteria (González et al. 2006). The high number of paralogous genes found in rhizobia (see below) may have contributed to the origin and diversification of symbiosis.

Rhizobium symbiotic plasmids are related to *Agrobacterium* plasmids pTi and pRi that induce crown-galls or hairy-roots in dicotyledonous plants. Several authors have reported on the close evolutionary relationship between these classes of plasmids, suggesting that they may have chimeric origin (Moriguchi

et al. 2001). Supporting this idea, it has been found, in different SGCs, complete or partial genetic systems for conjugation (*tra*/*trb*) or the type IV transport system (*vir*) that in *Agrobacterium* plasmids are responsible for transporting the T-DNA to plant cells. For instance, the pSymA of S. *meliloti* has retained partially the *vir*, the pNGR234a has complete *tra* and *trb* systems but lacks the *vir* genes (Barnett et al. 2001). Moreover, the p42d (the pSym of *R. etli*) has an entire set of *virB* genes and partially the *tra* region, but lacks *trb* and other *vir* genes (González et al. 2003).

The genetic heterogeneity of the SGCs suggests that they have been shaped during evolution by rearrangements, recombination, horizontal transfer, and transposition (González et al. 2003). Genetic rearrangements have been shown to occur in *R. etli* involving both plasmids and chromosome (Brom et al. 1991; Flores et al. 1988). In particular, the pSym of *R. etli* is prone to undergo deletions and amplifications by recombination of the reiterated *nifH* genes (Romero et al. 1991). The high number of repeated sequences present in the SGCs suggests that the occurrence of rearrangements is very common, as was shown for the pSym of *Rhizobium* spp NGR234 (Flores et al. 2000). Furthermore, rearrangements involving the symbiotic region of *R. tropici* lead to increased symbiotic capabilities, such as competition for nodulation (Mavingui et al. 1997).

With the SGCs available so far, only general comparisons can be made. Nevertheless, to understand the evolution of SGCs, comparisons of SGCs from strains of the same specie are needed. Such studies have been done for the symbiotic islands of two strains of *M. loti* and for sequences of the pSym of various strains of *R. etli* (Flores et al. 2005; Sullivan et al. 2002). The symbiotic islands of the *M. loti* strains R7A and MAFF303099 have probably derived from an ancestral island. A conserved backbone of 248 kb shows clear colinearity, and about 98% nucleotide identity. In both islands the backbone is interrupted by insertions and deletions of diverse DNA segments. These specific regions harbor ISs (that are strain specific) and hypothetical genes. Interestingly, some segments of the islands are homologs of regions of the plasmids pMLa and pMLb of *M. loti*. Like in other SGCs, in the symbiotic island R7A there is a complete set of *vir* genes for the type IV secretion system, whereas the symbiotic island MAFF303099 lacks *vir* genes but has the genes for the type III secretion system.

Further insights on the mechanisms that contribute to the diversification of SGCs were published recently (Flores et al. 2005). A comparison of the variations in nucleotide sequences among pSym of different *R. etli* strains shows an asymmetric distribution of single nucleotide polymorphisms (Flores et al. 2005). There are regions with few changes and regions with a high number of nucleotide substitutions. Moreover, some highly polymorphic sites share exactly the same changes in several strains. The authors propose that the majority of the nucleotide substitutions are produced in the population by recombination, and that the contribution of mutations to polymorphism is relatively low (Flores et al. 2005).

7.5 Horizontal Gene Transfer and Mobile Elements

Horizontal gene transfer is one of the major evolutionary forces in most bacteria (Gogarten et al. 2002; Lawrence and Hendrickson 2003). The rhizobial species are no exception. Different methods have been employed to infer cases of horizontal transfer, like GC composition, codon usage, GC composition in the third base of the codon, dinucleotide frequency, and phylogenetic testing for congruency (Ragan et al. 2006). In the rhizobial chromosome there are regions of lower GC content or dinucleotide composition that differ from the average. These islands contain genes related to integrases, transposons, DNA transfer and plasmid stabilization genes (Capela et al. 2001; González et al. 2006; Young et al. 2006). Insertion sequences belonging to distinct families like IS66, IS630, IS110, IS3, IS4, and others are particularly abundant in the SGCs. The prevalence of an IS family in a particular *Rhizobium* contrasts with its poor representation in others. For instance, the IS66 is the most abundant IS in *R. etli* but it is poorly represented in *B. japonicum* (González et al. 2006). Even though the role of horizontal transfer in rhizobial evolution is widely recognized, precise estimates of the degree and rates of gene acquisitions are still lacking.

There is good evidence suggesting that symbiotic plasmids and islands might have been acquired during the evolution of rhizobia. In general, SGCs have a lower GC content and different codon usage compared with the rest of the genome. As mentioned before, they also have numerous insertion sequences and, in some cases, phage sequences on their borders, which may indicate the acquisition of these regions by lateral transfer. Because of these characteristics, it has been proposed that the symbiotic regions correspond to entire mobile genetic elements (Kaneko et al. 2002).

It is well known that rhizobia have conjugative plasmids and phages that could be the means of gene dissemination in the bacterial population. The role of phages in gene transfer has been poorly studied, whereas self-transmissible symbiotic plasmids have been described in *Rhizobium leguminosarum*, *Rhizobium* spp NGR234, *R. etli,* and *S. meliloti* (Jhonston 1978; Pérez-Mendoza et al. 2004; Tun-Garrido et al. 2003; Herrera-Cervera et al. 1998). The plasmids pRL1J1 of *R. leguminosarum* and pNGR234a of *Rhizobium* spp are able to conjugate at relatively high frequencies in laboratory conditions by a mechanism that involves quorum sensing (Danino et al. 2003; He et al. 2003). Indeed, genomic sequences of the symbiotic plasmids and islands have revealed the presence of conjugative systems encoded by *tra* and *trb* genes (Freiberg et al. 1997; González et al. 2003; Sullivan et al. 2002). In *R. etli* there are two well-characterized conjugative plasmids. One is the 194-kb p42a plasmid that is transmissible at high frequencies depending on cell-density (Tun-Garrido et al. 2003). Moreover, p42a can mobilize the symbiotic plasmid p42d of *R. etli* by a mechanism that involves cointegration of both replicons (Brom et al. 2004). Recently, it has been shown that the symbiotic plasmid p42d of *R. etli* transfers itself, this process requiring a protein encoded by the gene *yp028* (Pérez-Mendoza et al. 2004). Furthermore, it was also shown that the conjugal transfer of the pSym of *R. etli* and *S. meliloti* are repressed in laboratory conditions by the

product of *rctA*, a DNA winged-helix DNA binding transcriptional regulator (Pérez-Mendoza et al. 2005). These data indicate that conjugative transfer of the symbiotic plasmids could be activated in natural conditions once an environmental signal, as yet unknown, is present.

The most persuasive evidence of horizontal transfer of symbiotic plasmids comes from experiments carried out in New Zealand (Sullivan et al. 1995). After the introduction of an inoculant strain of *M. loti* into a field devoid of native *Mesorhizobium*, a set of genetically diverse strains that contain symbiotic regions identical to that of the introduced strain was recovered (Sullivan et al. 1995). Furthermore, it was shown that the symbiotic region that was transferred is about 500 kb-long and inserted into the chromosomal loci for tRNA-phe (Sullivan and Ronson 1998).

7.6 Genetic and Metabolic Redundancy

Rhizobia face many challenging situations in the soil and in the nodule. They compete for energy sources with other microorganisms and cope with variable environmental conditions like humidity, drought, salinity, pH, temperature and others. We may think that some strategies exist for successful survival in such an environment. Rhizobia are heterotrophic obligate microaerophiles that can assimilate a wide range of rhizosphere carbon and nitrogen sources (Galibert et al. 2001; González et al. 2003; Goodner et al. 2001; Kaneko et al. 2000, 2002; Wood et al. 2001). They can exploit many sugars (mannose and rhamnose) present in plant root exudates as well as other rhizosphere compounds such as rhizopines (Prell and Poole 2006). They can also uptake many nutrients that exist in low concentrations in the soil by means of a large number of ABC transporters (Prell and Poole 2006). Such metabolic plasticity can be accounted for by the extensive genetic redundancy already evidenced by the high number of gene duplications in rhizobia (González et al. 2006). For instance, global genomic analysis shows that there are more isozymes in the genomes of *R. etli* and *S. meliloti* than in those of *E. coli* (González et al. 2006). In fact, in *R. etli* there are many isozymes for aminoacid biosynthesis distributed in both plasmids and chromosome. There are also an important number of pathways for fermentation, degradation and assimilation of aminoacids, aromatic compounds, carboxylates, sugars and polysaccharides. Such metabolic plasticity might be related to different degrees of physiological responses and the alternative regulation needed to be successful in the soil. In effect, large families of transcriptional regulators exist in rhizobia accounting for 9% of the total gene content, including members of the families LysR, TetR, AraC, LacI, and GntR. In addition, substantial amounts of two-component regulators indicate that signal transduction is a frequent mean for coordinating environmental responses. Moreover, Rhizobia harbor a high number of sigma factors of unknown function (Table 7.1). There are three basic sigma factors, RpoD, RpoN, and RpoH and several sigma factors belonging to the class of extracytoplasmic factors (ECFs). Various ECFs have been

studied in species other than rhizobia, having diverse roles like iron uptake, toxin secretion, alginate biosynthesis, and others. With the exception of ECF RpoI, that in *R. leguminosarum* controls the siderophore synthesis, the physiological role of the other ECFs is unknown. RpoS, the sigma factor that controls the transition to the stationary phase of growth in enterobacteria is notably absent from rhizobia.

7.7 An Ancestral Chromosome?

It has been suggested that an ancestral chromosome was present in the origin of the α-proteobacteria group and consequently in species of the order Rhizobiales (Boussau et al. 2004; Galibert et al. 2001). Such an ancestral chromosome might have had 3000–5000 genes and could have been well adapted to saprophytic life. Thus, the ability to interact with animal or vegetal cells is a derived character specific for some species. Interestingly, two opposite phenomena have taken place in the α-proteobacteria: massive genome expansions in Rhizobiales associated to plants, and genome contractions in several branches of Rhizobiales and Ricketssiales, pathogens of mammals (*Ricketssia*, *Brucella*, *and Bartonella*) or endosymbionts of insects (*Wolbacchia*) (Boussau et al. 2004). In fact, rhizobia have a high amount of paralogous genes. Most of them are ancient duplications. The number of paralogous genes in the five complete genomes of rhizobia represents 30–40% of the total gene content (Galibert et al. 2001; González et al. 2006). As was shown for *R. etli*, paralogous families with the largest number of members belong to diverse ABC transporters, transcription regulators, and a variety of genes related to the metabolism of amino acids, nucleotides, carbohydrates, coenzymes, lipids, inorganic ions, and secondary metabolites (González et al. 2006). It has been proposed that bacteria with large genomes are more ecologically successful in environments where resources are scarce but diverse, and there is no disadvantage for slow growth (Konstantinidis and Tiedje 2004). This is the expected situation that rhizobia confront in the soil.

It has been shown that gene order is well correlated with the phylogenetic distance that separates the different clades (Tamames 2001). Synteny maps of Rhizobiales chromosomes reveal a very high degree of colinearity between pairs of related species (Fig. 7.2) (Boussau et al. 2004; González et al. 2006; Goodner et al. 2001). Indeed, they reveal the close relationship of symbiotic species with plant or mammal pathogens. For instance, *S. meliloti* and *R. etli* chromosomes are very conserved among themselves and with *A. tumefaciens* (plant pathogen), but they also are very syntenic with the chromosome I of *Brucella* species. In contrast, plasmids or secondary chromosomes do not display any synteny pattern (González et al. 2006).

As has been discussed, rhizobia are not a homogeneous group, and the proportion of orthologous genes shared between closer species like *R. etli* and *S. meliloti*, is nearly 60% of the total gene content, whereas *B. japonicum*, the most divergent lineage, shares 42% of the orthologs with other rhizobia (González et al. 2006).

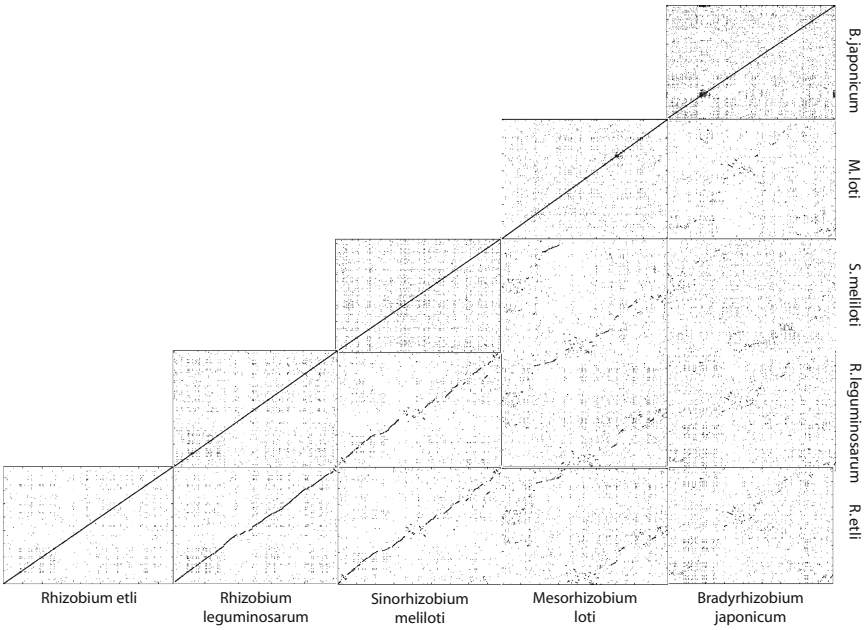

Fig. 7.2 Syntenic relationships between the chromosomes of pairs of rhizobia species. Complete chromosomes were aligned using the PROMER application of the MUMMER package (Delcher et al. 1999). All matching and alignments were performed on the six frame aminoacid translation of the DNA input sequences. These are used to perform pairwise alignments. Diagonals represent maximal matches between pairs of sequences. Points or lines outside the diagonals are matches due to repeated aminoacid stretches of paralogous genes or domains. Recently, two additional complete genomes of *Bradyrhizobium* spp have been reported and comparisons with *B. japonicum* USDA110 show several syntenic regions but also extensive rearrangements of other segments. Authors suggest that *Bradyrhizobium* genomes are highly plastic by the presence of numerous mobile elements (Giraud et al. 2007)

These observations suggest that not only gene duplications or gene acquisitions by horizontal transfer have had a role in rhizobia evolution but also severe gene losses have occurred. Conversely, it may be expected that a set of common genes would be present in rhizobia but absent from close relatives like *A. tumefaciens* or *Brucella*. In fact, by discounting the common orthologs between the plant pathogen *A. tumefaciens* and rhizobia, a set of about 500 genes constitute a core shared exclusively by *R. etli* and *S. meliloti* or by *R. leguminosarum*, *B. japonicum* and *S. meliloti*, and are not found in close relatives. These genes include the well-known *nif*, *fix* and *nod* genes, large families of adenylate cyclases and glutathione reductases, and several sigma factors that are not found in *A. tumefaciens* (González et al. 2006). Moreover, the core includes genes of all functional classes but not essential for cell survival, as well as hypothetical genes, perhaps related to symbiosis (González et al. 2006; Young et al. 2006).

7.8 Conclusions

The advent of genomics has changed our point of view about biology. Integrative approaches based on the knowledge of all genetic elements of the cell are currently used to answer fundamental questions. For many years, researchers in the field of nitrogen fixation have had questions about the origin and evolution of symbiosis. How do *nod* and *nif* genes unite to perform a complex function? Why is the distribution of the symbiotic trait so scattered? What is the functional connection between the symbiotic metabolism and the rest of the cellular metabolism? What are the commonalities and differences within rhizobia? Surely, field researchers have many other questions, but we can conclude on some suggestive ideas derived from genomic analysis. The ancestor of the present day rhizobia species likely had a large genome with broad biosynthetic capabilities to survive in variable environments where energy sources were scarce (Boussau et al. 2004). The high number of paralogous families presently found in the rhizobia genomes is the reflection of ancient mechanisms (Galibert et al. 2001; González et al. 2006). Thus, the ancestor of rhizobia was well-adapted to the saprophytic life-style, and the ability to perform symbiosis might have been a function acquired by rhizobia along the course of the evolution. Gene recruitment, through gene duplication and horizontal transfer, not only enriched the evolution of symbiosis but contributed to expanding the metabolic repertoire of rhizobia.

Acknowledgements The authors wish to thank to José Espíritu and Víctor del Moral for technical assistance during the writing of this chapter. This work is supported by grants from PAPIIT IN223005-3 and CONACyT U46333-Q.

References

Allardet-Servent A, Michaux-Charachon S, Jumas-Bilak E, Karayan L, Ramuz M (1993) Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. J Bacteriol 175:7869–7874

Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, Gurjal M, Hong A, Huizar L, Hyman RW, Kahn D, Kahn ML, Kalman S, Keating DH, Palm C, Peck MC, Surzycki R, Wells DH, Yeh K-C, Davis RW, Ederspiel NA, Long SR (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc Natl Acad Sci USA 98:9883–9888

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG (2004) Computational inference of scenarios for alpha-proteobacterial genome evolution. Proc Natl Acad Sci USA 101:9722–9727

Brom S, García de los Santos A, de Lourdes Girard M, Dávila G, Palacios R, Romero D (1991) High-frequency rearrangements in *Rhizobium leguminosarum* bv. *phaseoli* plasmids. J Bacteriol 173:1344–1346

Brom S, Girard L, Tun-Garrido C, Santos AG, Bustos P, González V, Romero D (2004) Transfer of the symbiotic plasmid of *Rhizobium etli* CFN42 requires cointegration with p42a, which may be mediated by site-specific recombination. J Bacteriol 186:7538–7548

Campbell A (1981) Evolutionary significance of accessory DNA elements in bacteria. Annu Rev Microbiol 35:55–83

Capela D, Barloy-Hubler F, Gouzy J, Bothe G, Ampe F, Batut J, Boistard P, Becker A, Boutry M, Cadieu E, Dréano S, Gloux S, Godrie T, Goffeau A, Kahn D, Kiss E, Lelaure V, Masuy D, Pohl T, Portetelle D, Pühler A, Purnelle B, Ramsperger U, Renard C, Thébault P, Vandenbol M, Weidner S, Galibert F (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. Proc Natl Acad Sci USA 98:9877–9882

Cevallos MA, Porta H, Izquierdo J, Tun-Garrido C, García-de-los-Santos A, Dávila G, Brom S (2002) *Rhizobium etli* CFN42 contains at least three plasmids of the repABC family: a structural and evolutionary analysis. Plasmid 48:104–116

Cheng J, Sibley CD, Zaheer R, Finan TM (2007) A *Sinorhizobium meliloti minE* mutant has an altered morphology and exibits defects in legume symbiosis. Microbiology 153:375–387

Danino VE, Wilkinson A, Edwards A, Downie JA (2003) Recipient-induced transfer of the symbiotic plasmid pRL1JI in *Rhizobium leguminosarum bv. viciae* is regulated by a quorum-sensing relay. Mol Microbiol 50:511–525

Dávila G, Palacios R (2005) Origins of genomics in nitrogen fixation research. In: Palacios R, Newton WE (eds) Genomes and genomics of nitrogen-fixing organisms. Springer, Berlin Heidelberg New York, pp 1–6

Delcher AL, Kasif S, Fleishman RD, Peterson J, White O, Salzberg SL (1999) Alignment of whole genomes. Nucleic Acids Res 27:2369–2376

Flores M, González V, Pardo MA, Leija A, Martínez E, Romero D, Piñero D, Dávila G, Palacios R (1988) Genomic instability in *Rhizobium phaseoli*. J Bacteriol 170:1191–1196

Flores M, Mavingui P, Perret X, Broughton WJ, Romero D, Hernández G, Dávila G, Palacios R (2000) Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: toward a natural genomic design. Proc Natl Acad Sci USA 97:9138–9143

Flores M, Morales L, Avila A, González V, Bustos P, García D, Mora Y, Guo X, Collado-Vides J, Piñero D, Dávila G, Mora J, Palacios R (2005) Diversification of DNA sequences in the symbiotic genome of *Rhizobium etli*. J Bacteriol 187:7185–7192

Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. Nature 387:394–401

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

Garrity GM, Jhonson KL, Bell JA, Searles DB (2002) Taxonomic outline of the Procaryotes. Release 3.0. In: Bergey's manual of systematic bacteriology. Springer, Berlin Heidelberg New York, p 365

Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Vermeglio A, Medigue C, Sadowsky M (2007) Legumes symbiosis: abscence of *nod* genes in photosynthetic Bradyrhizobia. Science 316:1307–1311

Gogarten JP, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. Mol Biol Evol 19:2226–2238

Gonzalez V, Bustos P, Ramirez-Romero MA, Medrano-Soto A, Salgado H, Hernandez-Gonzalez I, Hernandez-Celis JC, Quintero V, Moreno-Hagelsieb G, Girard L, Rodriguez O, Flores M, Cevallos MA, Collado-Vides J, Romero D, Davila G (2003) The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. Genome Biol 4:R36

Gonzalez V, Santamaria RI, Bustos P, Hernandez-Gonzalez I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramirez MA, Jimenez-Jacinto V, Collado-Vides J, Davila G (2006) The partitioned

Rhizobium etli genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci USA 103:3834–3839

Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo B, Goldman BS, Cao Y, Askenazi M, Halling C, Mullin L, Houmiel K, Gordon J, Vaudin M, Iartchouk O, Epp A, Liu F, Wollam C, Allinger M, Doughty D, Scott C, Lappas C, Markelz B, Flanagan C, Crowell C, Gurson J, Lomo C, Sear C, Strub G, Cielo C, Slater S (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294:2323–2328

Guo X, Flores M, Mavingui P, Fuentes SI, Hernández G, Dávila G, Palacios R (2003) Natural genomic design in *Sinorhizobium meliloti*: novel genomic architectures. Genome Res 13:1810–1817

He X, Chang W, Pierce DL, Seib LO, Wagner J, Fuqua C (2003) Quorum sensing in *Rhizobium* sp. strain NGR234 regulates conjugal transfer (*tra*) gene expression and influences growth rate. J Bacteriol 185:809–822

Herrera-Cervera JA, Sanjuan-Pinilla JM, Olivares J, Sanjuan J (1998) Cloning and identification of conjugative transfer origins in the *Rhizobium meliloti* genome. J Bacteriol 180:4583–4590

Jhonston AWB, Beynon JL, Buchanan-Wollaston AV, Setchell SM, Hirsh PR, Beringer JE (1978) High frequency transfer of nodulating ability between strains and species of *Rhizobium*. Nature 276:635–636

Jumas-Bilak E, Michaux-Charachon S, Bourg G, Ramuz M, Allardet-Servent A (1998) Unconventional genomic organization in the alpha subgroup of the Proteobacteria. J Bacteriol 180:2749–2755

Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S, Watanabe A, Idesawa K, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Mochizuki Y, Nakayama S, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res 7:331–338

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 (supplement). DNA Res 9:225–256

Kasting JF, Siefert JL (2001) Biogeochemistry: the nitrogen fix. Nature 412:26–27

Konstantinidis KT, Tiedje JM (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci USA 101:3160–3165

Lawrence JG, Hendrickson H (2003) Lateral gene transfer: when will adolescence end? Mol Microbiol 50:739–749

Lloret L, Martínez-Romero E (2005) Evolución y Filogenia de *Rhizobium*. Rev Latinoam Microbiol 47:43–60

López-Lara IM, Geiger O (2001) The nodulation protein NodG shows the enzymatic activity of an 3-oxoacyl-acyl carrier protein reductase. Mol Plant Microbe Interact 14:349–357

Marie C, Barny MA, Downie JA (1992) *Rhizobium leguminosarum* has two glucosamine synthases, GlmS and NodM, required for nodulation and development of nitrogen-fixing nodules. Mol Microbiol 6:843–851

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

Mavingui P, Flores M, Romero D, Martínez-Romero E, Palacios R (1997) Generation of *Rhizobium* strains with improved symbiotic properties by random DNA amplification (RDA). Nat Biotechnol 15:564–569

Mavingui P, Flores M, Guo X, Dávila G, Perret X, Broughton WJ, Palacios R (2002) Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. J Bacteriol 184:171–176

Moriguchi K, Maeda Y, Satou M, Hardayani NS, Kataoka M, Tanaka N, Yoshida K (2001) The complete nucleotide sequence of a plant root-inducing (Ri) plasmid indicates its chimeric structure and evolutionary relationship between tumor-inducing (Ti) and symbiotic (Sym) plasmids in Rhizobiaceae. J Mol Biol 307:771–784

Moulin L, Munive A, Dreyfus B, Boivin-Masson C (2001) Nodulation of legumes by members of the beta-subclass of Proteobacteria. Nature 411:948–950

Navarro-González R, McKay CP, Mvondo DN (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. Nature 412:61–64

Palacios R, Newton WE (2005) Nitrogen fixation: origins, applications, and research progress. In: Palacios R, Newton WE (eds) Genomes and genomics of nitrogen-fixing organisms Springer, Berlin Heidelberg New York, pp ix–xiv

Pérez-Mendoza D, Domínguez-Ferreras A, Muñoz S, Soto MJ, Olivares J, Brom S, Girard L, Herrera-Cervera JA, Sanjuán J (2004) Identification of functional *mob* regions in *Rhizobium etli*: evidence for self-transmissibility of the symbiotic plasmid pRetCFN42d. J Bacteriol 186:5753–5761

Pérez-Mendoza D, Sepúlveda E, Pando V, Muñoz S, Nogales J, Olivares J, Soto MJ, Herrera-Cervera JA, Romero D, Brom S, Sanjuán J (2005) Identification of the *rctA* gene, which is required for repression of conjugative transfer of rhizobial symbiotic megaplasmids. J Bacteriol 187:7341–7350

Prell J, Poole P (2006) Metabolic changes of rhizobia in legume nodules. Trends Microbiol 14:161–168

Ragan MA, Harlow TJ, Beiko RG (2006) Do different surrogate methods detect lateral genetic transfer events of different relative ages? Trends Microbiol 14:4–8

Ramírez-Romero MA, Bustos P, Girard L, Rodríguez O, Cevallos MA, Dávila G (1997) Sequence, localization and characteristics of the replicator region of the symbiotic plasmid of *Rhizobium etli*. Microbiology 143(8):2825–2831

Ramírez-Romero MA, Soberón N, Pérez-Oseguera A, Téllez-Sosa J, Cevallos MA (2000) Structural elements required for replication and incompatibility of the *Rhizobium etli* symbiotic plasmid. J Bacteriol 182:3117–3124

Romero D, Brom S, Martínez-Salazar J, Girard ML, Palacios R, Dávila G (1991) Amplification and deletion of a *nod-nif r*egion in the symbiotic plasmid of *Rhizobium phaseoli*. J Bacteriol 173:2435–2441

Sessitsh A, Howieson JG, Perret X, Antoun H, Martínez E (2002) Advances in *Rhizobium* research. Crit Rev Plant Sci 21:323–378

Soberón N, Venkova-Canova T, Ramírez-Romero MA, Téllez-Sosa J, Cevallos MA (2004) Incompatibility and the partitioning site of the repABC basic replicon of the symbiotic plasmid from *Rhizobium etli*. Plasmid 51:203–216

Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci USA 95:5145–5149

Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci USA 92:8985–8989

Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, Elliot RM, Fleetwood DJ, McCallum NG, Rossbach U, Stuart GS, Weaver JE, Webby RJ, De Bruijn FJ, Ronson CW (2002) Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. J Bacteriol 184:3086–3095

Tamames J (2001) Evolution of gene order conservation in prokaryotes. Genome Biol 2(6): research0020.1-0020.11

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005) Comparative metagenomics of microbial communities. Science 308:554–557

Tun-Garrido C, Bustos P, González V, Brom S (2003) Conjugative transfer of p42a from *Rhizobium etli* CFN42, which is required for mobilization of the symbiotic plasmid, is regulated by quorum sensing. J Bacteriol 185:1681–1692

Turner SL, Young JP (2000) The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol Biol Evol 17:309–319

Vázquez M, Santana O, Quinto C (1993) The NodL and NodJ proteins from *Rhizobium* and *Bradyrhizobium* strains are similar to capsular polysaccharide secretion proteins from Gram-negative bacteria. Mol Microbiol 8:369–377

Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF Jr, Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D Sr, Chapman P,

Clendenning J, Deatherage G, Gillet W, Grant C, Kutyavin T, Levy R, Li MJ, McClelland E, Palmieri A, Raymond C, Rouse G, Saenphimmachak C, Wu Z, Romero P, Gordon D, Zhang S, Yoo H, Tao Y, Biddle P, Jung M, Krespan W, Perry M, Gordon-Kamm B, Liao L, Kim S, Hendrick C, Zhao ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW (2001) The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294:2317–2323

Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol 7:R34