Chapter 4
Genomics and Evolution of Rhizobia

4.1 The General Organisation of Rhizobial Genomes

In the first decade of the twenty-first century, people working on rhizobia had been very excited about the release of the complete genomes of model strains including *Mesorhizobium loti* MAFF303099 (reclassified as *M. japonicum*) (Kaneko et al. 2000), *Sinorhizobium meliloti* 1021 (Galibert et al. 2001), *Bradyrhizobium japonicum* USDA110 (reclassified as *B. diazoefficiens*) (Kaneko et al. 2002), *Rhizobium etli* CFN42 (González et al. 2006), *Rhizobium leguminosarum* bv. *viciae* 3841 (Young et al. 2006), *Bradyrhizobium* sp. BTAi1 and ORS278 (Giraud et al. 2007), *Azorhizobium caulinodans* ORS571 (Lee et al. 2008), *Cupriavidus taiwanensis* LMG19424 (Amadou et al. 2008), *Sinorhizobium* sp. NGR234 (Schmeisser et al. 2009) and *Sinorhizobium medicae* WSM419 (Reeve et al. 2010). Notably, all of these genomes were sequenced using the Sanger platform, and these valuable earlier efforts have provided us essential information regarding general features of rhizobial genomes. For example, symbiosis genes are intensively clustered in a symbiosis island or a symbiosis plasmid, and genome organisation and gene content can vary drastically between different species. These features have been further validated by more than 100 complete rhizobial genomes obtained later on using next-generation sequencing platforms such as Illumina, Roche 454, Ion Torrent and PacBio. As shown in Fig. 4.1, there is a great variation in genome size of different strains/species within each genus, indicating diverse metabolic abilities of rhizobial germplasms.

E. T. Wang et al., *Ecology and Evolution of Rhizobia*,
https://doi.org/10.1007/978-981-32-9555-1_4

Strain	Genus	Genome Size (Mb)	BioProject
B.diazoefficiens NK6	*Bradyrhizobium*	10.48	PRJDB3027
B. japonicum J5	*Bradyrhizobium*	10.14	PRJNA347501
Mic. ossetica V5/3M	*Microvirga*	9.63	PRJNA329489
B. japonicum E109	*Bradyrhizobium*	9.22	PRJNA270102
B. japonicum USDA 6	*Bradyrhizobium*	9.21	PRJDA67463
B. diazoefficiens USDA 122	*Bradyrhizobium*	9.14	PRJNA298974
B. diazoefficiens USDA 110	*Bradyrhizobium*	9.11	PRJNA17
Methy. nodulans ORS 2060	*Methylobacterium*	8.84	PRJNA20477
Para. phymatum STM815	*Paraburkholderia*	8.68	PRJNA17409
R. leguminosarum Vaf10	*Rhizobium*	8.57	PRJNA316801
B. sp. BTAi1	*Bradyrhizobium*	8.49	PRJNA16137
R. leguminosarum Vaf-108	*Rhizobium*	8.45	PRJNA316801
B. icense LMTR 13	*Bradyrhizobium*	8.32	PRJNA325367
B. sp. SK17	*Bradyrhizobium*	8.29	PRJNA420598
B. oligotrophicum S58	*Bradyrhizobium*	8.26	PRJDB684
R. leguminosarum BIHB	*Rhizobium*	7.95	PRJNA395792
Burk. sp. *CCGE1002*	*Burkholderia*	7.88	PRJNA37719
R. leguminosarum UPM791	*Rhizobium*	7.84	PRJNA417467
B. sp. CCGE-LA001	*Bradyrhizobium*	7.83	PRJNA172908
Para. sprentiae WSM5005	*Paraburkholderia*	7.83	PRJNA344837
B. sp. ORS 285	*Bradyrhizobium*	7.80	PRJEB20226
R. leguminosarum Norway	*Rhizobium*	7.79	PRJNA417364
R. leguminosarum 3841	*Rhizobium*	7.75	PRJNA344
Methy. sp. 4-46	*Methylobacterium*	7.74	PRJNA18809
S. meliloti RU11/001	*Sinorhizobium*	7.69	PRJEB4559
Para. phenoliruptrix BR3459a	*Paraburkholderia*	7.65	PRJNA174166
M. japonicum MAFF 303099	*Mesorhizobium*	7.60	PRJNA18
B. sp. BF49	*Bradyrhizobium*	7.55	PRJEB10689
M. loti NZP2037	*Mesorhizobium*	7.48	PRJNA325064
R. gallicum IE4872	*Rhizobium*	7.47	PRJNA293856
B. sp. ORS 278	*Bradyrhizobium*	7.46	PRJNA19575
R. leguminosarum	*Rhizobium*	7.42	PRJNA20097
M. amorphae CCNWGS0123	*Mesorhizobium*	7.34	PRJNA318467
R. gallicum R602	*Rhizobium*	7.31	PRJNA227036
R. etli 8C-3	*Rhizobium*	7.31	PRJNA293876
R. sp. NXC24	*Rhizobium*	7.30	PRJNA415482
S. meliloti M270	*Sinorhizobium*	7.28	PRJNA388336
B. sp. S23321	*Bradyrhizobium*	7.23	PRJDA72425
S. meliloti USDA1157	*Sinorhizobium*	7.21	PRJNA388336
M. sp. WSM1497	*Mesorhizobium*	7.20	PRJNA323416
R. etli Mim1	*Rhizobium*	7.20	PRJNA200310
S. meliloti SM11	*Sinorhizobium*	7.17	PRJNA41117
S. meliloti Rm41	*Sinorhizobium*	7.15	PRJEB436
S. meliloti USDA1021	*Sinorhizobium*	7.15	PRJNA388336
S. meliloti AK83	*Sinorhizobium*	7.14	PRJNA41993
S. meliloti GR4	*Sinorhizobium*	7.14	PRJNA175860
S. meliloti Rm41	*Sinorhizobium*	7.14	PRJNA388336
S. fredii CCBAU 83666	*Sinorhizobium*	7.08	PRJNA353922
R. sp. 10195	*Rhizobium*	7.07	PRJNA399130
R. sp. IE4771	*Rhizobium*	7.06	PRJNA230502
S. meliloti T073	*Sinorhizobium*	7.04	PRJNA388336
R. etli bv. *phaseoli* IE4803	*Rhizobium*	7.00	PRJNA231318
S. meliloti HM006	*Sinorhizobium*	7.00	PRJNA388336
S. meliloti BL225C	*Sinorhizobium*	6.98	PRJNA42477
S. meliloti KH46	*Sinorhizobium*	6.98	PRJNA388336
R. sp. N324	*Rhizobium*	6.97	PRJNA293118
R. leguminosarum BIHB	*Rhizobium*	6.96	PRJNA395793
M. ciceri CC1192	*Mesorhizobium*	6.94	PRJNA317272
R. phaseoli N671	*Rhizobium*	6.91	PRJNA293118
R. phaseoli N771	*Rhizobium*	6.91	PRJNA293118
R. leguminosarum	*Rhizobium*	6.90	PRJNA62289

Strain	Genus	Genome Size (Mb)	BioProject
S. fredii NGR234	*Sinorhizobium*	6.89	PRJNA21101
M. opportunistum WSM2075	*Mesorhizobium*	6.88	PRJNA33861
M. ciceri WSM1284	*Mesorhizobium*	6.88	PRJNA317273
R. leguminosarum	*Rhizobium*	6.87	PRJNA20179
R. phaseoli N841	*Rhizobium*	6.86	PRJNA293118
S. americanum CCGM7	*Sinorhizobium*	6.85	PRJNA222537
R. sp. Kim5	*Rhizobium*	6.82	PRJNA29555
S. medicae WSM419	*Sinorhizobium*	6.82	PRJNA16304
S. meliloti M162	*Sinorhizobium*	6.81	PRJNA388336
R. etli NXC12	*Rhizobium*	6.76	PRJNA383588
R. phaseoli N831	*Rhizobium*	6.75	PRJNA293118
R. phaseoli R630	*Rhizobium*	6.75	PRJNA293118
R. phaseoli N931	*Rhizobium*	6.75	PRJNA293118
S. americanum CFNEI 73	*Sinorhizobium*	6.75	PRJNA298565
S. meliloti KH35c	*Sinorhizobium*	6.75	PRJNA388336
R. phaseoli N261	*Rhizobium*	6.74	PRJNA293118
R. phaseoli R723	*Rhizobium*	6.74	PRJNA293118
R. phaseoli R650	*Rhizobium*	6.73	PRJNA293118
S. meliloti RMO17	*Sinorhizobium*	6.73	PRJNA244712
R. phaseoli N161	*Rhizobium*	6.72	PRJNA293118
S. meliloti USDA1106	*Sinorhizobium*	6.72	PRJNA388336
R. leguminosarum CB782	*Rhizobium*	6.70	PRJNA67103
R. phaseoli R611	*Rhizobium*	6.70	PRJNA293118
S. meliloti CCMM B554	*Sinorhizobium*	6.70	PRJNA369312
M. ciceri WSM1271	*Mesorhizobium*	6.69	PRJNA48991
R. tropici CIAT 899	*Rhizobium*	6.69	PRJNA42391
R. sp. NXC14	*Rhizobium*	6.69	PRJNA383589
S. meliloti 1021	*Sinorhizobium*	6.69	PRJNA19
R. phaseoli Brasil 5	*Rhizobium*	6.67	PRJNA29557
R. sp. CIAT894	*Rhizobium*	6.66	PRJNA29565
R. phaseoli R620	*Rhizobium*	6.63	PRJNA293118
R. sp. N1341	*Rhizobium*	6.56	PRJNA293118
R. sp. N741	*Rhizobium*	6.56	PRJNA293118
S. fredii NXT3	*Sinorhizobium*	6.56	PRJNA415486
R. etli CFN 42	*Rhizobium*	6.53	PRJNA13932
R. sp. N113	*Rhizobium*	6.52	PRJNA293118
C. taiwanensis LMG 19424	*Cupriavidus*	6.48	PRJNA15733
R. etli N561	*Rhizobium*	6.48	PRJNA293118
R. sp. N871	*Rhizobium*	6.48	PRJNA293118
S. meliloti B399	*Sinorhizobium*	6.47	PRJNA361265
N. galegae HAMBI 540	*Neorhizobium*	6.46	PRJEB1950
R. etli CIAT 652	*Rhizobium*	6.45	PRJNA28021
R. sp. N621	*Rhizobium*	6.43	PRJNA293118
R. sp. N6212	*Rhizobium*	6.42	PRJNA293118
N. galegae HAMBI 1141	*Neorhizobium*	6.41	PRJEB1951
R. sp. TAL182	*Rhizobium*	6.40	PRJNA294178
S. sp. CCBAU 05631	*Sinorhizobium*	6.39	PRJNA353922
R. sp. ACO-34A	*Rhizobium*	6.28	PRJNA384228
R. sp. N731	*Rhizobium*	6.27	PRJNA293118
R. sp. N1314	*Rhizobium*	6.27	PRJNA293118
P. sp. Tri-48	*Phyllobacterium*	6.21	PRJNA329489
M. australicum WSM2073	*Mesorhizobium*	6.20	PRJNA47287
S. meliloti B401	*Sinorhizobium*	6.14	PRJNA361251
S. sojae CCBAU 05684	*Sinorhizobium*	6.09	PRJNA353922
R. sp. S41	*Rhizobium*	5.52	PRJNA327265
R. sp. IRBG74	*Rhizobium*	5.46	PRJEB4411
A. caulinodans ORS 571	*Azorhizobium*	5.37	PRJDA19267
M. sp. B7	*Mesorhizobium*	5.32	PRJNA354894
R. sp. Y9	*Rhizobium*	5.32	PRJNA351913
O. anthropi ATCC 49188	*Ochrobactrum*	5.21	PRJNA19485
R. sp. NT-26	*Rhizobium*	4.58	PRJEB4806

Fig. 4.1 122 complete rhizobial genomes. Genomes retrieved on March 30, 2018, from BioProject as indicated are ordered according to the genome sizes. *Bradyrhizobium*, *Rhizobium*, *Mesorhizobium* and *Sinorhizobium* are coloured

4.1.1 Replicons: Chromosome, Chromid and Plasmid

4.1.1.1 Replicons

Although some archaea harbour a chromosome with the multiple-origin mode of replication (Lindås and Bernander 2013), all DNA molecules of bacterial genomes studied to date have a single replication origin. Consequently, replicon has been widely used to refer such a DNA molecule in bacterial genomes (Harrison et al. 2010; diCenzo and Finan 2017). Among the 122 complete rhizobial genomes from 12 genera accessible till March of 2018 (Fig. 4.1), 107 genomes from 11 genera have two or more replicons (Fig. 4.2a), a genome organisation feature described as a multipartite genome. The multipartite genome organisation is apparently overrepresented in rhizobia compared to an estimated 10% for bacterial genomes (diCenzo and Finan 2017).

The multipartite genome architecture is scarce in *Bradyrhizobium* and not found in the genus *Azorhizobium*, from which only one complete genome is available (Fig. 4.2a). More replicons per genome do not necessarily lead to a larger genome size (Fig. 4.2b). Genome sizes of *Bradyrhizobium* strains are among the largest in rhizobia (Figs. 4.1 and 4.3), even though most *Bradyrhizobium* strains have only one replicon. In a recent global exploration of the soil microbiome, *Bradyrhizobium* was identified as one of the most ubiquitous phylotypes of bacteria (Delgado-Baquerizo et al. 2018). It would be interesting to investigate the relationship between the genome size and the adaptation ability of different *Bradyrhizobium* strains with a great variation in their genome sizes (such as from 7.23 to 10.48 Mb; Fig. 4.1). The size of individual replicons in a multipartite genome can be smaller than the single replicon found in *Bradyrhizobium* and *Azorhizobium*, and the total genome size of a rhizobial strain is generally above 4.5 Mb (Fig. 4.3). This number is several times larger than

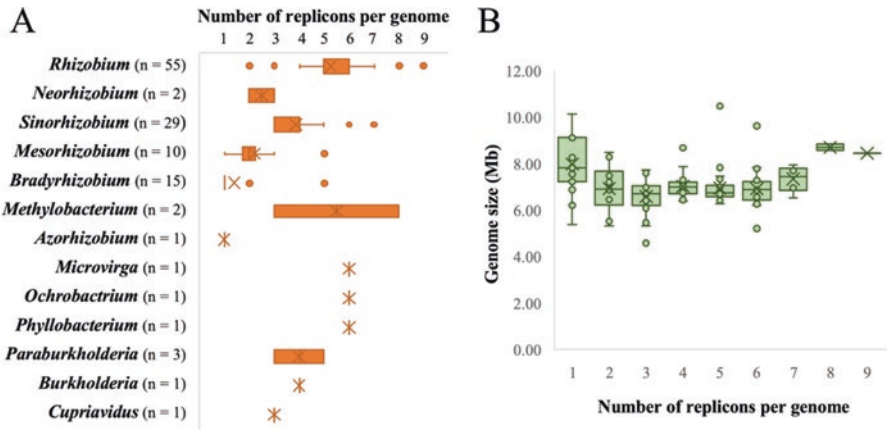

Fig. 4.2 The number of replicons in rhizobial genomes. (**a**) Box plot of the number of replicons within individual complete genomes available in the corresponding genus. The number of analysed genomes is indicated in brackets. (**b**) Box plot of the genome size in rhizobia with different numbers of replicons. One hundred twenty-two rhizobial genomes were analysed

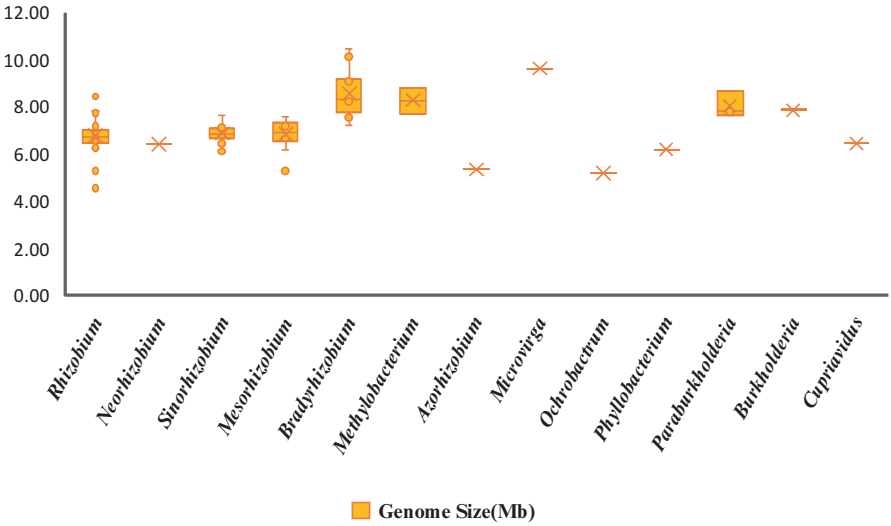

Fig. 4.3 Box plot showing the size of rhizobial genomes. The numbers of analysed genomes are the same as those indicated in Fig. 4.2a for each genus

those of animal-associated obligate endosymbionts (Toft and Andersson 2010), which cannot be cultivated in the laboratory. It is also larger than the average and median bacterial genome sizes, 3.87 Mb and 3.65 Mb, respectively (diCenzo and Finan 2017).

4.1.1.2 Chromosome, Chromid and Plasmid

The single replicon or the largest replicon in a multipartite genome is called a chromosome. In most cases for a multipartite genome, information genes such as rRNA genes and highly conserved housekeeping genes and most essential genes are located on the chromosome. But the presence of a sole rRNA operon in a nonchromosomal replicon has been reported for a plant-associated α-proteobacterium *Aureimonas* sp. AU20 (Anda et al. 2015). In addition to the chromosome, secondary chromosome, chromid, megaplasmids (above 350 kb in size) and plasmids are terms that have been proposed to classify different replicons present in a multipartite genome (diCenzo and Finan 2017). Among them, secondary chromosome and chromid refer to a secondary replicon harbouring some essential genes (either under all conditions or environmentally) (Harrison et al. 2010; diCenzo and Finan 2017). The name "secondary chromosome" was used to indicate that it resulted from a split of the ancestor chromosome into two (Fig. 4.4). With the increasing number of genomes in the database, convincing evidence for this split event is however rare (diCenzo and Finan 2017) and not easily obtained for researchers not fully involved in this specific field. On the other hand, nearly all secondary replicons with essential genes are considered to be chromids evolved from plasmids (Harrison et al. 2010; diCenzo

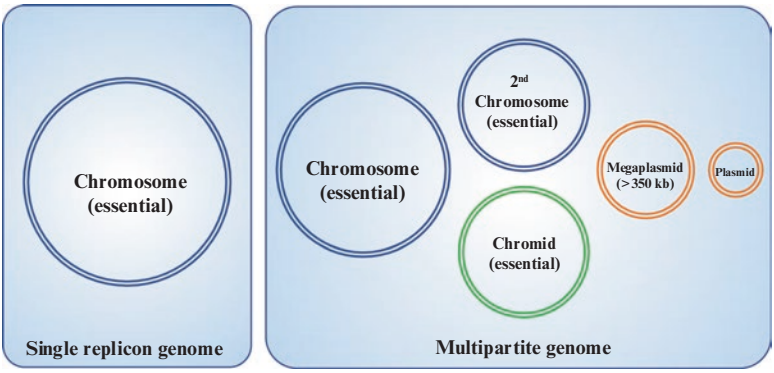

Fig. 4.4 Classification of bacterial replicons. A single replicon genome (left) and a multipartite genome (right) are shown. Different replicons in a multipartite genome are ordered according to their sizes from left to right, but the diameter of each replicon is not scaled to actual size (e.g. >350 kb for megaplasmid). The same colour of second chromosome and the primary chromosome indicates that two replicons result from a split of the ancestor chromosome into two. Essential genes are harboured by the corresponding replicon as indicated. Notably, a chromid smaller than megaplasmids has been reported in a non-rhizobial strain *Deinococcus deserti* VCD115. (Harrison et al. 2010)

Table 4.1 General features of different replicons

Characteristic	Chromosome	Chromid	(Mega)plasmid
Size	The largest	Usually secondary replicons	Smallest replicons
G + C	Varies	Usually within 1% of chromosome	Often >1% lower than chromosome
Maintenance and replication systems	Chromosome-type	Plasmid-type	Plasmid-type
Core genes	Most essential genes	Some essential genes	Few genes shared at any phylogenetic level
Phylogenetic distribution of genes	Wide conservation of genes and synteny between genera	Gene conservation and shared synteny only within genus	Genes specific to strain or species

Adapted from Harrison et al. (2010)

and Finan 2017). Megaplasmids (>350 kb) and plasmids are used to refer replicons lacking essential genes but enriched with dispensable genes and characterised with biased sequence features such as lower GC content and biased codon usage compared to the chromosome. By contrast, the sequence features including GC content and codon usage of chromids are similar to those of the chromosome (Harrison et al. 2010). Distinct characteristics of different replicons have been excellently reviewed by Harrison et al. (2010) and listed herein in Table 4.1.

4.1.2 Symbiosis Plasmid and Symbiosis Island

Despite the great diversity of alpha- and beta-rhizobia (Peix et al. 2015), most rhizobia have a cluster of key symbiosis genes, including *nod*, *nif* and *fix*, localised within a symbiosis island in the chromosome or in the symbiosis plasmid. These genes are specifically involved in nodulation and nitrogen fixation processes during the symbiosis with compatible legumes. Deletion of the key nodulation genes usually leads to a complete loss of symbiotic ability of rhizobia associated with either specific (such as *Medicago sativa*) or promiscuous (such as *Sophora flavescens*) legume hosts (Marvel 1985, 1987; Horvath et al. 1986; Liu et al. 2018). Some *Bradyrhizobium* strains do not require canonical *nod* genes and typical lipochitooligosaccharidic Nod factors for symbiosis with certain *Aeschynomene* species (Giraud et al. 2007; Miche et al. 2010); nevertheless the *nif* and *fix* genes are clustered in a 45-kb island in their genomes (Giraud et al. 2007).

As expected from Table 4.1, symbiosis plasmids have a lower GC content compared to the chromosome in a multipartite rhizobial genome. Here we take *Sinorhizobium* strains associated with soybeans as examples (Fig. 4.5a, b). *S. fredii* CCBAU45436 is an epidemic and efficient soybean microsymbiont in alkaline soils (Zhang et al. 2011; Tian et al. 2012). Five replicons were identified in its multipartite genome (Jiao et al. 2018): chromosome (cSF45436), chromid (pSF45436b), symbiosis plasmid (pSF45436a) and two accessory plasmids (pSF45436d and pSF45436e) (Fig. 4.5a). The symbiosis plasmid pSF45436a is a megaplasmid of 0.42 Mb, which is around 10% and 20% of the size of the chromosome (cSF45436) and chromid (pSF45436b), respectively (Fig. 4.5a). Its GC% (59.9%) is at least 3% lower than those of chromid and chromosome (Fig. 4.5a). Although the replicon size of the symbiosis plasmid varies in different *S. fredii* strains nodulating soybeans, such as 0.40–0.74 Mb in CCBAU45436, CCBAU25509 and CCBAU83666, the average GC% varies little among symbiosis plasmids of *Sinorhizobium* spp. nodulating soybeans (Fig. 4.5b). By contrast, the GC% of chromid is only slightly (0.5%) but also significantly lower than that of chromosome (Fig. 4.5b). Another notable feature of the symbiosis plasmid in these *Sinorhizobium* strains is the enrichment of insertion sequences (ISs), particularly those high-copy ones, compared to chromosome and chromid (Zhao et al. 2018). Although transposable elements had been considered as junk and selfish components in genomes, accumulative evidence supports their critical roles in the evolution of both eukaryotes and prokaryotes (Biémont 2010). A recent experimental evolution study demonstrated that insertion mutation of type three secretion system (T3SS) genes by parallel transpositions of ISs, enriched on the same symbiosis plasmid, is the major mutagenesis mechanism during adaptive evolution of symbiotic compatibility of *Sinorhizobium* associated with soybeans (Zhao et al. 2018). It should be noted, however, that the symbiosis plasmid is not essential for the free-living stage of its rhizobial host, as experimentally demonstrated in *S. meliloti* (diCenzo et al. 2014, 2018). Transcriptomics analyses recurrently show that most genes on the symbiosis plasmid of diverse rhizobia are specifically induced during nodulation and nitrogen fixation, but not under free-living conditions lacking a compatible host or its

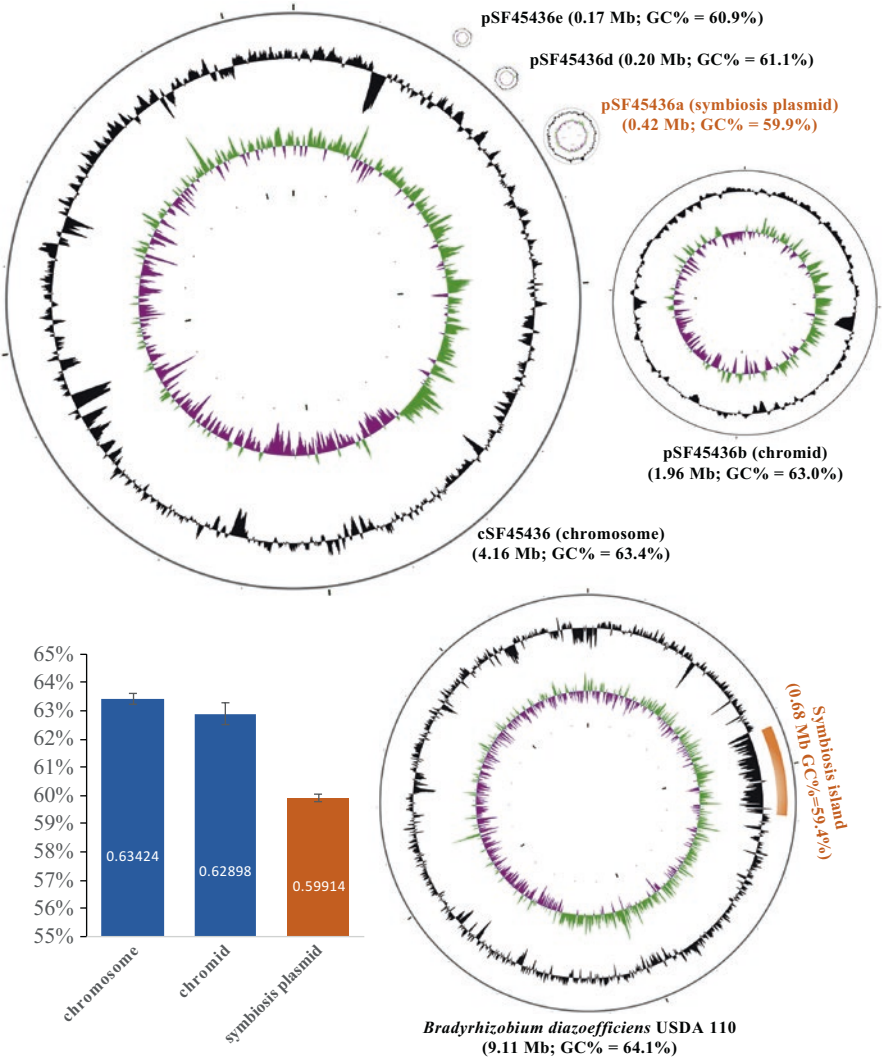

Fig. 4.5 Representative symbiosis plasmid and symbiosis island. (**a**) Five replicons including the symbiosis plasmid pSF45436a in the genome of *Sinorhizobium fredii* CCBAU45436 nodulating soybeans. (**b**) Average GC% of three major replicons in soybean microsymbionts belonging to *Sinorhizobium* (*S. fredii* CCBAU45436, *S. fredii* CCBAU25509, *S. fredii* CCBAU83666, *S. sojae* CCBAU05684, *Sinorhizobium* sp. CCBAU05631). Significant GC% difference of chromid or symbiosis plasmid compared to that of chromosome is shown (T-test; ∗, $p < 0.05$; ∗∗∗, $p < 0.001$). (**c**) The genome of *Bradyrhizobium diazoefficiens* USDA 110 nodulating soybeans. The size and GC% of the symbiosis island are indicated. (**a** and **c**) GC content (black ring) and GC skew (the ring in green and purple) are shown. The genome size of USDA 110 in (**c**) is at a scale of one third of the CCBAU45436 genome in (**a**). A window size of 10,000 and a step of 100 were used in GC content and GC skew analyses for USDA 110, cSF45436 and pSF45436b, while a size of 1000 and a step of 10 were used for pSF45436a, pSF45436d and pSF45436e

symbiotic signal molecules (Ampe et al. 2003; Capela et al. 2006; Vercruysse et al. 2011; Li et al. 2013; Jiao et al. 2016).

In rhizobia with a single replicon and some rhizobia (such as certain *Mesorhizobium* strains) with multiple replicons, key symbiosis genes are found on the chromosome. As shown in the genome of *Bradyrhizobium diazoefficiens* USDA 110 (Fig. 4.5c), a genomic island of six hundred eighty-one kilobyte in length is characterized by its lower GC% (59.4%) than the genomic average (64.4%).

Six hundred eighty-one kilobyte in length is characterised by its lower GC% (59.4%) than the genomic average (64.4%). This island contains key symbiosis genes *nod*, *nif* and *fix* (Göttfert et al. 2001; Kaneko et al. 2002) and many uncharacterised genes, which are highly transcribed in soybean nodules (Pessi et al. 2007). Similarly, a genomic island of 611 kb containing *nod/nif* genes was identified on the chromosome of *Mesorhizobium japonicum* MAFF303099, which harbours two more replicons (plasmids) (Kaneko et al. 2000). Consequently, "symbiosis island" has been used to refer this kind of genomic island (Sullivan and Ronson 1998). As in symbiosis plasmids, there is an overrepresentation of ISs in symbiosis islands, such that 60% of the ISs of *B. diazoefficiens* USDA 110 were localised in this island (Kaneko et al. 2002). The symbiosis island of *M. japonicum* is also characterised by its enrichment of transposable elements compared to the chromosome background and the two plasmids (Kaneko et al. 2000).

More than 20 years ago, it was demonstrated that the symbiosis island of *Mesorhizobium loti* can be transferred into non-symbiotic mesorhizobia under field and lab conditions and integrated into a phe-tRNA gene (Sullivan et al. 1995; Sullivan and Ronson 1998). Recently, Ling et al. provided evidence that the symbiosis island of *Azorhizobium caulinodans* is an integrative and conjugative element that can be transferred to a specific site in a gly-tRNA gene of other rhizobial genera (Ling et al. 2016). Moreover, the horizontal transfer frequency of this symbiosis island increased in the legume rhizosphere or in the presence of plant flavonoids (Ling et al. 2016), highlighting an intriguing host-dependent evolutionary scenario of rhizobia. As shown in Table 4.2, one or two conserved met-tRNA gene(s) can be identified in symbiosis plasmids but not other extrachromosomal replicons of *Sinorhizobium* strains nodulating soybeans. These data imply that integration into a tRNA gene may have played an important role in the horizontal transfer of symbiosis genes in many rhizobial genera in the long run. It is noteworthy that the symbiosis plasmid itself can be subject to conjugative transfer, as demonstrated in *Rhizobium* and *Sinorhizobium* (Danino et al. 2003; Perez-Mendoza et al. 2004, 2005). This is in line with the finding that extremely similar symbiosis plasmids were found in different *Rhizobium* species associated with common bean (Perez Carrascal et al. 2016). If we look at the alignment of symbiosis plasmids from *Sinorhizobium* strains associated with soybeans (Fig. 4.6), a similar conclusion can be drawn for certain *S. fredii* and *S. sojae* strains (CCBAU45436, CCBAU25509 and CCBAU05684). Although highly conserved locally collinear blocks can also be found in *S. fredii* CCBAU83666 and *Sinorhizobium* sp. CCBAU05631, extensive rearrangement and the presence of other accessory sequences can be found in symbiosis plasmids of these two strains (Fig. 4.6).

Table 4.2 Distribution of tRNA genes in multipartite genomes of *Sinorhizobium* nodulating soybeans

	S. sojae	*S. fredii*			*Sinorhizobium* sp.
Replicon	CCBAU05684	CCBAU45436	CCBAU25509	CCBAU83666	CCBAU05631
Chromosome	51	53	53	53	50
Chromid	0	0	0	0	0
Symbiosis plasmid	1 met-tRNA	1 met-tRNA	1 met-tRNA	2 met-tRNA	2 met-tRNA
Other plasmids	0	0	0	0	0

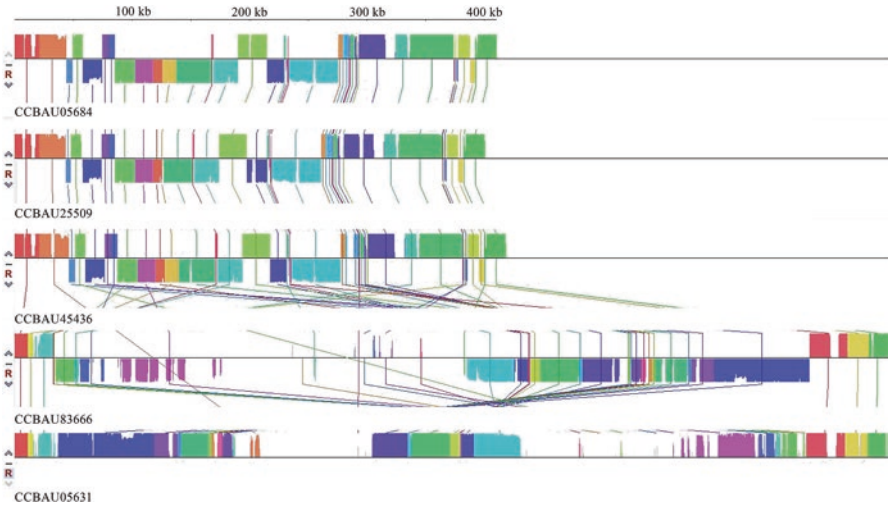

Fig. 4.6 Progressive Mauve alignment of symbiosis plasmids from *Sinorhizobium* microsymbionts of soybean. From the first to fifth row: *S. sojae* CCBAU05684, *S. fredii* CCBAU25509, *S. fredii* CCBAU45436, *S. fredii* CCBAU83666 and *Sinorhizobium* sp. CCBAU05631. Locally collinear blocks conserved between different strains are indicated in the same colour and connected

4.2 Evolution of Core and Accessory Genes

4.2.1 Characteristics of Core and Accessory Genes

In the previous Sect. (4.1), intraspecies and intra-genus variation in rhizobial genome size at the scale of Mb can be observed. If we simply take 1 kb as the average length of a gene, the difference in gene number can be up to several thousand between strains. This phenomenon is widespread in prokaryotes. In 2005, a term "pan-genome" ("pan" – "**παν**" in Greek – means "whole") was introduced to refer the gene repertoire accessible to any given species (Medini et al. 2005; Tettelin et al. 2005). The pan-genome is composed of a "core genome" containing genes present

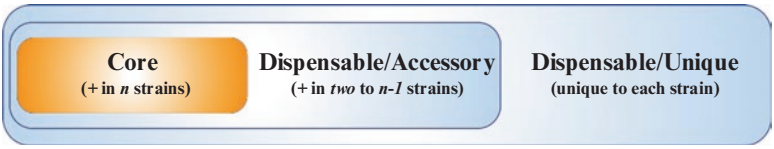

Fig. 4.7 A schematic diagram illustrating the partition of a pan-genome for n strains of a given taxonomic unit. +, present

in all strains and a "dispensable genome" (also called accessory, flexible or adaptive) with genes present in a subset of strains (Medini et al. 2005) (Fig. 4.7). The dispensable genome can be further divided into two elements: genes shared by some but not all strains (named "accessory" genes in some publications to distinguish it from "core" and "unique" elements) and genes unique to each strain (Medini et al. 2005, 2008; Rouli et al. 2015) (Fig. 4.7). Although the species is usually considered to be an evolutionary unit, the pan-genome concept has been extended to higher taxonomic units (Lapierre and Gogarten 2009). This is biologically meaningful, since accessory gene functions may provide adaptive advantages for their host cells in a specific niche and the pan-genome analysis of different species inhabiting the same niche can provide novel insight into the evolutionary mechanisms underlying their adaptation and competition. For example, *S. sojae* CCBAU05684, *Sinorhizobium* sp. CCBAU05631 and *S. fredii* CCBAU45436 share certain wild soybean hosts (Li et al. 2011; Zhang et al. 2011; Liu et al. 2017; Zhao et al. 2018). A pan-genome analysis followed by reverse genetics has revealed that an accessory gene cluster present in CCBAU45436 and CCBAU05631 but absent in CCBAU05684 is essential for effective symbiosis of its host strains (Liu et al. 2017).

It has been estimated that the pan-genome of the bacterial domain is of infinite size, likely due to numerous niches on earth (Lapierre and Gogarten 2009; McInerney et al. 2017), i.e. the number of new genes grows indefinitely with the number of sequenced strains. An "open" pan-genome is used to refer this pattern (Medini et al. 2005). By contrast, if the size of a pan-genome quickly saturates to a limiting value, a "closed" pan-genome can be proposed (Medini et al. 2005). A closed pan-genome has been reported for species living in isolated niches with limited access to the global microbial gene pool, such as *Bacillus anthracis*, *Mycobacterium tuberculosis* and *Chlamydia trachomatis* (Medini et al. 2005). As facultative microsymbionts, rhizobia are expected to have a large pan-genome to cope with fluctuating biotic and abiotic stimuli in soils and during symbiosis with legumes. Indeed, rhizobia such as the model species *Sinorhizobium meliloti* associated with *Medicago* and *S. fredii* nodulating soybeans have a typical open pan-genome (Tian et al. 2012; Galardini et al. 2013). The same conclusion can be drawn for species belonging to *Rhizobium*, *Mesorhizobium* and *Bradyrhizobium* (Tian et al. 2012; Kumar et al. 2015; Perez Carrascal et al. 2016; Porter et al. 2017).

A genome-wide average nucleotide identity (ANI) value of 95% has been widely used to determine if two prokaryotic strains can be considered to be the same spe-

cies (Richter and Rossello-Mora 2009), and a discontinuity in ANI space is observed around this boundary (Konstantinidis and Tiedje 2005; Richter and Rossello-Mora 2009). This gap in sequence space has also been reported in several independent analyses of rhizobia using either a fixed number of shared core genes or a genome scale alignment (Tian et al. 2012; Zhang et al. 2012; Kumar et al. 2015). Therefore, it is established that core genome determines the taxonomy of rhizobia, as for other prokaryotes (Ormeno-Orrillo et al. 2015). By contrast, representative features used in polyphasic taxonomy in pre-genomics studies only capture a tiny fraction of the inter-species variation, and it is not uncommon that these features can also vary at the intraspecies level (Ormeño-Orrillo and Martínez-Romero 2013; Kumar et al. 2015; Vernikos et al. 2015; Young 2016), thus blurring the species boundary. Comparative genomics of rhizobia from 8 genera suggested that the phyletic distribution of 887 functional genes with experimental evidence can reflect the species phylogeny of test strains, while the distribution of the whole pan-genome could not (Tian et al. 2012). This highlights that accessory genes in the open pan-genome of rhizobia are differentially integrated with the genome backgrounds of individual species. As typical accessory genes, key nodulation and nitrogen fixation genes within symbiosis islands or symbiosis plasmids of rhizobia determine the symbiovar and hence the corresponding legume host, rather than the bacterial species assignments (Rogel et al. 2011). These key symbiosis genes provide adaptive advantage for rhizobia in the presence of compatible legumes, while many other accessory genes can be adaptive in diverse niches in soils. For example, in contrast to *Sinorhizobium*, *Bradyrhizobium* strains are enriched with accessory genes involved in secondary metabolism, which may explain the high global abundance of *Bradyrhizobium* in soils (Tian et al. 2012; Delgado-Baquerizo et al. 2018).

4.2.2 Main Evolutionary Forces Shaping the Diversity of Core and Accessory Genes

It is estimated that the divergence of rhizobial genera predates the origin of legumes (Turner and Young 2000), and transferable accessory symbiosis genes can be considered "microsymbionts" that have spread across diverse bacteria (Remigi et al. 2016). That is to say, these symbiosis genes succeed, regarding their wide phyletic distribution in at least two bacterial orders, by improving the adaptation of their host strains. This regime has largely dominated the evolutionary study of rhizobia in past decades.

With the burst of new rhizobial species being documented in the literature and the development of sequencing technology, our knowledge of rhizobial core genes has been extended from information on the 16S rRNA gene and few housekeeping genes (such as *atpD*, *glnII*, *recA*, *rpoB*, etc.) to hundreds and thousands of core genes. It is notable that both intragenic and intergenic recombination, in addition to

point mutation, have played a substantial role in creating the observed diversity of chromosomal housekeeping genes in rhizobial species such as *Bradyrhizobium canariense*, *B. japonicum*, *B. elkanii*, *B. liaoningense*, *B. yuanmingense*, *B. diazoefficiens*, *Rhizobium gallicum* sensu *lato*, *Rhizobium leguminosarum* bv. *viciae* and *Sinorhizobium fredii* (Vinuesa et al. 2005, 2008; Silva et al. 2005; Tian et al. 2010; Zhang et al. 2014; Guo et al. 2014). This view has been further verified in a comparison of individual core gene trees to the species tree based on 295 core genes in alpha- and beta-rhizobia (Tian et al. 2012). Around 90% of these core genes have undergone horizontal gene transfer or intergenic recombination, and only 20 out of 295 genes in test strains were free of either inter- or intragenic recombination (Tian et al. 2012). Therefore, strict vertical evolution is rare in rhizobial chromosomal core genes.

The multipartite architecture of many rhizobial genomes (Fig. 4.2 and Table 4.1) provides a unique opportunity to investigate the evolution of core and accessory genes. Extrachromosomal replicons thought to be essential for the saprophytic lifestyle in soils and rhizospheres usually show higher rates of recombination than the chromosomes, as demonstrated in *Rhizobium* and *Sinorhizobium* (Bailly et al. 2011; Guo et al. 2014; Perez Carrascal et al. 2016). The chromid of *Sinorhizobium* species such as *S. meliloti* and *S. fredii* is characterised by its distinct role in intraspecies differentiation and enrichment with accessory genes (Galardini et al. 2013; Jiao et al. 2018). Moreover, the chromid is a hot spot for positively selected genes such as those involved in the synthesis of polysaccharides (Bailly et al. 2011; Galardini et al. 2013), which can influence diverse aspects including host range and phage tolerance (Campbell et al. 2003; Parada et al. 2006; Staehelin et al. 2006; Müller et al. 2009; López-Baena et al. 2016). Horizontal gene transfer has a greater effect on gene content of symbiosis plasmids/islands than of chromids or chromosomes (Bailly et al. 2011; Zhang et al. 2014; Guo et al. 2014; Kumar et al. 2015; Perez Carrascal et al. 2016). Symbiosis plasmids are more prone to share a gene pool with accessory plasmids, as reported in *S. meliloti* strains (Nelson et al. 2018). A low frequency of horizontal gene transfer on chromosomes does not equal none. Although accessory genes can be interspersed throughout the chromosome, most are concentrated in flexible genomic islands (fGIs) (Rodriguez-valera and Lo 2016). This phenomenon can be clearly identified in the example of *S. fredii* strains (Fig. 4.8). Several fGIs are present in locally collinear blocks. These fGIs may contribute to intraspecies variation and increase the adaptation potential of populations. For example, an accessory operon encoding a multidrug efflux system in *S. fredii* CCBAU45436 is located within a fGI on the chromosome (indicated in Fig. 4.8) and is essential for efficient symbiosis of CCBAU45436 with soybeans (Jiao et al. 2018).

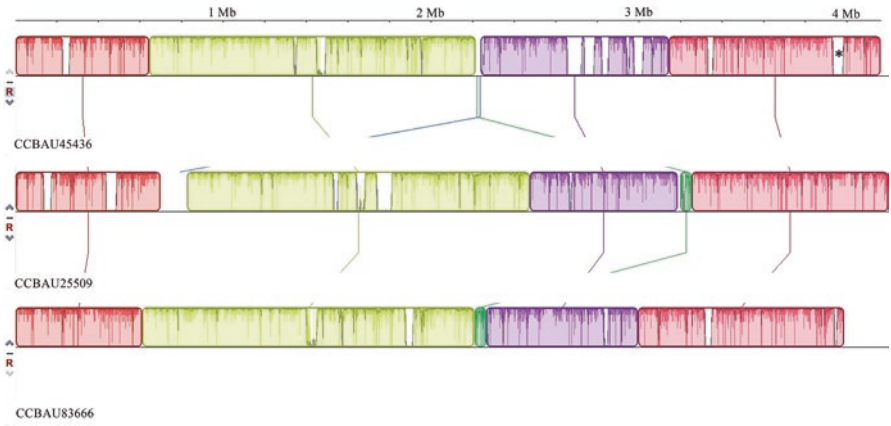

Fig. 4.8 Progressive Mauve alignment of chromosomes from soybean microsymbionts of *Sinorhizobium fredii*. From the first to third row: *S. fredii* CCBAU45436, CCBAU25509 and CCBAU83666. Locally collinear blocks conserved between different strains are indicated in the same colour and connected. ∗ indicates a flexible genome island containing genes encoding a multidrug efflux system that is essential for effective symbiosis of CCBAU45436 with soybeans. (Jiao et al. 2018)

References

Amadou C, Pascal G, Mangenot S, et al. Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. Genome Res. 2008;18:1472–83. https://doi.org/10.1101/gr.076448.108 [pii]

Ampe F, Kiss E, Sabourdy F, Batut J. Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. Genome Biol. 2003;4:R15.

Anda M, Ohtsubo Y, Okubo T, et al. Bacterial clade with the ribosomal RNA operon on a small plasmid rather than the chromosome. Proc Natl Acad Sci U S A. 2015;112:14343-7. https://doi.org/10.1073/pnas.1514326112.

Bailly X, Giuntini E, Sexton MC, et al. Population genomics of *Sinorhizobium medicae* based on low-coverage sequencing of sympatric isolates. ISME J. 2011;5:1722–34. https://doi.org/10.1038/ismej.2011.55 [pii].

Biémont C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics. 2010;186:1085–93.

Campbell GR, Sharypova LA, Scheidle H, et al. Striking complexity of lipopolysaccharide defects in a collection of *Sinorhizobium meliloti* mutants. J Bacteriol. 2003;185:3853–62.

Capela D, Filipe C, Bobilk C, et al. *Sinorhizobium meliloti* differentiation during symbiosis with alfalfa: a transcriptomic dissection. Mol Plant-Microbe Interact. 2006;19:363–72. https://doi.org/10.1094/MPMI-19-0363.

Danino VE, Wilkinson A, Edwards A, Downie JA. Recipient-induced transfer of the symbiotic plasmid pRL1JI in *Rhizobium leguminosarum* bv. *viciae* is regulated by a quorum-sensing relay. Mol Microbiol. 2003;50:511–25. doi: 3699 [pii]

Delgado-Baquerizo M, Oliverio AM, Brewer TE, et al. A global atlas of the dominant bacteria found in soil. Science. 2018;359:320–5.

diCenzo GC, Finan TM. The divided bacterial genome: structure, function, and evolution. Microbiol Mol Biol Rev. 2017;81:1–37.

diCenzo GC, MacLean AM, Milunovic B, et al. Examination of prokaryotic multipartite genome evolution through experimental genome reduction. PLoS Genet. 2014;10:e1004742. https://doi.org/10.1371/journal.pgen.1004742.

diCenzo GC, Benedict AB, Fondi M, et al. Robustness encoded across essential and accessory replicons in an ecologically versatile bacterium. 2018.

Galardini M, Pini F, Bazzicalupo M, et al. Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. Genome Biol Evol. 2013;5:542–58. https://doi.org/10.1093/gbe/evt027.

Galibert F, Finan TM, Long SR, et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science. 2001;(80)293:668–72. https://doi.org/10.1126/science.1060966293/5530/668. [pii]

Giraud E, Moulin L, Vallenet D, et al. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science. 2007;(80)316:1307–12. doi: 316/5829/1307 [pii] 10.1126/science.1139548

González V, Santamaría RI, Bustos P, et al. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci U S A. 2006;103:3834–9. doi: 0508502103 [pii] 10.1073/pnas.0508502103

Gottfert M, Rothlisberger S, Kundig C, et al. Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. J Bacteriol. 2001;183:1405–12. https://doi.org/10.1128/JB.183.4.1405-1412.2001.

Guo HJ, Wang ET, Zhang XX, et al. Replicon-dependent differentiation of symbiosis-related genes in *Sinorhizobium* nodulating *Glycine max*. Appl Environ Microbiol. 2014;80:1245–55. https://doi.org/10.1128/AEM.03037-13.

Harrison PW, Lower RPJ, Kim NKD, Young JPW. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol. 2010;18:141–8. https://doi.org/10.1016/j.tim.2009.12.010.

Horvath B, Kondorosi E, John M, et al. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. Cell. 1986;46:335–43. doi:0092-8674(86)90654-9 [pii]

Jiao J, Wu LJ, Zhang B, et al. MucR is required for transcriptional activation of conserved ion transporters to support nitrogen fixation of *Sinorhizobium fredii* in soybean nodules. Mol Plant-Microbe Interact. 2016;29:352–61.

Jiao J, Ni M, Zhang B, et al. Coordinated regulation of core and accessory genes in the multipartite genome of *Sinorhizobium fredii*. PLoS Genet. 2018;14:e1007428. https://doi.org/10.1371/journal.pgen.1007428.

Kaneko T, Nakamura Y, Sato S, et al. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. 2000;7:331–8.

Kaneko T, Nakamura Y, Sato S, et al. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. 2002;9:189–97.

Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102:2567–72. https://doi.org/10.1073/pnas.0409727102.

Kumar N, Lad G, Giuntini E, et al. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. Open Biol. 2015;5:140133. https://doi.org/10.1098/rsob.140133.

Lapierre P, Gogarten JP. Estimating the size of the bacterial pan-genome. Trends Genet. 2009;25:107–10. https://doi.org/10.1016/j.tig.2008.12.004.

Lee KB, De Backer P, Aono T, et al. The genome of the versatile nitrogen fixer *Azorhizobium caulinodans* ORS571. BMC Genomics. 2008;9:271. doi:1471-2164-9-271 [pii] 10.1186/1471-2164-9-271

Li QQ, Wang ET, Zhang YZ, et al. Diversity and biogeography of rhizobia isolated from root nodules of *Glycine max* grown in Hebei Province, China. Microb Ecol. 2011;61:917–31. https://doi.org/10.1007/s00248-011-9820-0.

Li Y, Tian CF, Chen WF, et al. High-resolution transcriptomic analyses of *Sinorhizobium* sp. NGR234 bacteroids in determinate nodules of *Vigna unguiculata* and indeterminate nodules of *Leucaena leucocephala*. PLoS One. 2013;8:e70531. https://doi.org/10.1371/journal.pone.0070531.

Lindås A-C, Bernander R. The cell cycle of archaea. Nat Rev Microbiol. 2013;11:627–38. https://doi.org/10.1038/nrmicro3077.

Ling J, Wang H, Wu P, et al. Plant nodulation inducers enhance horizontal gene transfer of *Azorhizobium caulinodans* symbiosis island. Proc Natl Acad Sci U S A. 2016;113:13875–80. https://doi.org/10.1073/pnas.1615121113.

Liu LX, Li QQ, Zhang YZ, et al. The nitrate-reduction gene cluster components exert lineage-dependent contributions to optimization of *Sinorhizobium* symbiosis with soybeans. Environ Microbiol. 2017;19:4926–38. https://doi.org/10.1111/1462-2920.13948.

Liu YH, Jiao YS, Liu LX, et al. Nonspecific symbiosis between sophora flavescens and different rhizobia. Mol Plant-Microbe Interact. 2018;31:224–32. https://doi.org/10.1094/MPMI-05-17-0117-R.

López-Baena F, Ruiz-Sainz J, Rodríguez-Carvajal M, Vinardell J. Bacterial molecular signals in the *Sinorhizobium fredii*-soybean symbiosis. Int J Mol Sci. 2016;17:755. https://doi.org/10.3390/ijms17050755.

Marvel DJ. Conservation of nodulation genes between *Rhizobium meliloti* and a slow-growing rhizobium strain that nodulates a nonlegume host. Proc Natl Acad Sci U S A. 1985;82:5841–5. https://doi.org/10.1073/pnas.82.17.5841.

Marvel DJ. *Rhizobium* symbiotic genes required for nodulation of legume and nonlegume hosts. Proc Natl Acad Sci U S A. 1987;84:1319–23. https://doi.org/10.1073/pnas.84.5.1319.

McInerney JO, McNally A, O'Connell MJ. Why prokaryotes have pangenomes. Nat Microbiol. 2017;2:17040. https://doi.org/10.1038/nmicrobiol.2017.40.

Medini D, Donati C, Tettelin H, et al. The microbial pan-genome. Curr Opin Genet Dev. 2005;15:589–94. https://doi.org/10.1016/j.gde.2005.09.006.

Medini D, Serruto D, Parkhill J, et al. Microbiology in the post-genomic era. Nat Rev Microbiol. 2008;6:419–30. doi: nrmicro1901 [pii]10.1038/nrmicro1901

Miche L, Moulin L, Chaintreuil C, et al. Diversity analyses of *Aeschynomene* symbionts in Tropical Africa and Central America reveal that nod-independent stem nodulation is not restricted to photosynthetic bradyrhizobia. Environ Microbiol. 2010;12:2152–64. https://doi.org/10.1111/j.1462-2920.2009.02090.x.

Müller MG, Forsberg LS, Keating DH. The *rkp-1* cluster is required for secretion of Kdo homopolymeric capsular polysaccharide in *Sinorhizobium meliloti* strain Rm1021. J Bacteriol. 2009;191:6988–7000. https://doi.org/10.1128/JB.00466-09.

Nelson M, Guhlin J, Epstein B, et al. The complete replicons of 16 *Ensifer meliloti* strains offer insights into intra- and inter-replicon gene transfer, transposon-associated loci, and repeat elements. Microb Genomics. 2018:1–11. https://doi.org/10.1099/mgen.0.000174.

Ormeño-Orrillo E, Martínez-Romero E. Phenotypic tests in rhizobium species description: an opinion and (a sympatric speciation) hypothesis. Syst Appl Microbiol. 2013; https://doi.org/10.1016/j.syapm.2012.11.009.

Ormeno-Orrillo E, Servin-Garciduenas LE, Rogel MA, et al. Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. Syst Appl Microbiol. 2015;38:287–91. https://doi.org/10.1016/j.syapm.2014.12.002.

Parada M, Vinardell JM, Ollero FJ, et al. *Sinorhizobium fredii* HH103 mutants affected in capsular polysaccharide (KPS) are impaired for nodulation with soybean and *Cajanus cajan*. Mol Plant-Microbe Interact. 2006;19:43–52. https://doi.org/10.1094/Mpmi-19-0043.

Peix A, Ramírez-Bahena MH, Velázquez E, Bedmar EJ. Bacterial associations with legumes. Crit Rev Plant Sci. 2015;34:17–42. https://doi.org/10.1080/07352689.2014.897899.

Perez Carrascal OM, VanInsberghe D, Juarez S, et al. Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing rhizobium species associated with *Phaseolus vulgaris*. Environ Microbiol. 2016;18:2660–76. https://doi.org/10.1111/1462-2920.13415.

Perez-Mendoza D, Dominguez-Ferreras A, Munoz S, et al. Identification of functional mob regions in *Rhizobium etli*: evidence for self-transmissibility of the symbiotic plasmid pRetCFN42d. J Bacteriol. 2004;186:5753–61. https://doi.org/10.1128/JB.186.17.5753-5761.2004186/17/5753. [pii]

Pérez-Mendoza D, Sepúlveda E, Pando V, et al. Identification of the *rctA* gene, which is required for repression of conjugative transfer of rhizobial symbiotic megaplasmids. J Bacteriol. 2005;187:7341–50. https://doi.org/10.1128/JB.187.21.7341-7350.2005.

Pessi G, Ahrens CH, Rehrauer H, et al. Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. Mol Plant-Microbe Interact. 2007;20:1353–63. https://doi.org/10.1094/Mpmi-20-11-1353.

Porter SS, Chang PL, Conow CA, et al. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. ISME J. 2017;11:1–15. https://doi.org/10.1038/ismej.2016.88.

Reeve W, Chain P, O'Hara G, et al. Complete genome sequence of the *Medicago* microsymbiont *Ensifer* (*Sinorhizobium*) *medicae* strain WSM419. Stand Genomic Sci. 2010;2:77–86. https://doi.org/10.4056/sigs.43526.

Remigi P, Zhu J, Young JPW, Masson-Boivin C. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. Trends Microbiol. 2016;24:63–75. https://doi.org/10.1016/j.tim.2015.10.007.

Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106:19126–31. https://doi.org/10.1073/pnas.0906412106.

Rodriguez-valera F, Lo M. Flexible genomic islands as drivers of genome evolution. Curr Opin Microbiol. 2016;31:154–60.

Rogel MA, Ormeño-Orrillo E, Martinez Romero E. Symbiovars in rhizobia reflect bacterial adaptation to legumes. Syst Appl Microbiol. 2011;34:96–104. https://doi.org/10.1016/j.syapm.2010.11.015.

Rouli L, Merhej V, Fournier P-E, Raoult D. The bacterial pangenome as a new tool for analyzing pathogenic bacteria. New Microbes New Infect. 2015;7:72–85. https://doi.org/10.1016/j.nmni.2015.06.005.

Schmeisser C, Liesegang H, Krysciak D, et al. *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. Appl Environ Microbiol. 2009;75:4035–45. doi: AEM.00515-09 [pii]10.1128/AEM.00515-09

Silva C, Vinuesa P, Eguiarte LE, et al. Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. Mol Ecol. 2005;14:4033–50. https://doi.org/10.1111/j.1365-294X.2005.02721.x.

Staehelin C, Forsberg LS, D'Haeze W, et al. Exo-oligosaccharides of *Rhizobium* sp. strain NGR234 are required for symbiosis with various legumes. J Bacteriol. 2006;188:6168–78. doi: 188/17/6168 [pii]10.1128/JB.00365-06

Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci U S A. 1998;95:5145–9.

Sullivan JT, Patrick HN, Lowther WL, et al. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci U S A. 1995;92:8985–9. https://doi.org/10.1073/pnas.92.19.8985.

Tettelin H, Masignani V, Cieslewicz MJ, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. 2005;102:13950–5. https://doi.org/10.1073/pnas.0506758102.

Tian CF, Young JPW, Wang ET, et al. Population mixing of *Rhizobium leguminosarum* bv. *viciae* nodulating *Vicia faba*: the role of recombination and lateral gene transfer. FEMS Microbiol Ecol. 2010;73:563–76. https://doi.org/10.1111/j.1574-6941.2010.00909.x.

Tian CF, Zhou YJ, Zhang YM, et al. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. Proc Natl Acad Sci U S A. 2012;109:8629–34. https://doi.org/10.1073/pnas.1120436109.

Toft C, Andersson SGE. Evolutionary microbial genomics: insights into bacterial host adaptation. Nat Rev Genet. 2010;11:465–75. https://doi.org/10.1038/Nrg2798.

Turner SL, Young JPW. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol Biol Evol. 2000;17:309–19.

Vercruysse M, Fauvart M, Beullens S, et al. A comparative transcriptome analysis of *Rhizobium etli* bacteroids: specific gene expression during symbiotic nongrowth. Mol Plant-Microbe Interact. 2011;24:1553–61. https://doi.org/10.1094/MPMI.

Vernikos G, Medini D, Riley DR. Ten years of pan-genome analyses. Curr Opin Microbiol. 2015;23:148–54.

Vinuesa P, Silva C, Werner D, Martinez-Romero E. Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. Mol Phylogenet Evol. 2005;34:29–54. https://doi.org/10.1016/j.ympev.2004.08.020.

Vinuesa P, Rojas-Jimenez K, Contreras-Moreira B, et al. Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent. Appl Environ Microbiol. 2008;74:6987–96. https://doi.org/10.1128/aem.00875-08.

Young JPW. Bacteria are smartphones and mobile genes are apps. Trends Microbiol. 2016;24:931–2. https://doi.org/10.1016/j.tim.2016.09.002.

Young JPW, Crossman LC, Johnston AW, et al. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. 2006;7:R34. doi:gb-2006-7-4-r34 [pii] 10.1186/gb-2006-7-4-r34

Zhang YM, Li Y Jr, Chen WF, et al. Biodiversity and biogeography of rhizobia associated with soybean plants grown in the North China plain. Appl Environ Microbiol. 2011;77:6331–42. https://doi.org/10.1128/AEM.00542-11.

Zhang YM, Tian CF, Sui XH, et al. Robust markers reflecting phylogeny and taxonomy of rhizobia. PLoS One. 2012;7:e44936. https://doi.org/10.1371/journal.pone.0044936.

Zhang XX, Guo HJ, Wang R, et al. Genetic divergence of *Bradyrhizobium* nodulating soybeans as revealed by multilocus sequence analysis of genes inside and outside the symbiosis island. Appl Environ Microbiol. 2014;80:3181–90. https://doi.org/10.1128/AEM.00044-14.

Zhao R, Liu LX, Zhang YZ, et al. Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences. ISME J. 2018;12:101–11. https://doi.org/10.1038/ismej.2017.136.