Acta Biochimica Polonica

Vol. 48 No. 2/2001

367–381

QUARTERLY

Review

Reduction of bacterial genome size and expansion resulting from obligate intracellular lifestyle and adaptation to soil habitat★

Tomasz Stêpkowski[½] and Andrzej B. Legocki

Institute of Bioorganic Chemistry Polish Academy of Sciences, Poznañ, Poland

Received: 14 February, 2001; accepted: 9 May, 2001

Key words: genome, symbiosis, rhizobium, *Buchnera*, *Rickettsia*

Prokaryotic organisms are exposed in the course of evolution to various impacts, resulting often in drastic changes of their genome size. Depending on circumstances, the same lineage may diverge into species having substantially reduced genomes, or such whose genomes have undergone considerable enlargement. Genome reduction is a consequence of obligate intracellular lifestyle rendering numerous genes expendable. Another consequence of intracellular lifestyle is reduction of effective population size and limited possibility of gene acquirement *via* lateral transfer. This causes a state of relaxed selection resulting in accumulation of mildly deleterious mutations that cannot be corrected by recombination with the wild type copy. Thus, gene loss is usually irreversible. Additionally, constant environment of the eukaryotic cell renders that some bacterial genes involved in DNA repair are expandable. The loss of these genes is a probable cause of mutational bias resulting in a high A+T content.

While causes of genome reduction are rather indisputable, those resulting in genome expansion seem to be less obvious. Presumably, the genome enlargement is an indirect consequence of adaptation to changing environmental conditions and requires the acquisition and integration of numerous genes. It seems that the need for a great number of capabilities is common among soil bacteria irrespective of their phylogenetic relationship. However, this would not be possible if soil bacteria lacked

---

★Presented at the International Conference on "Molecular Architecture of Evolution, Primary and Secondary Determinants" Poznañ, Poland, October 29–31, 2000.

[½]Address for correspondence: Institute of Bioorganic Chemistry Polish Academy of Sciences, 61-704 Poznañ, Z. Noskowskiego 12/14, Poland; tel.: (48 61) 852 8919; fax: (48 61) 852 0532; e-mail: sttommic@ibch.poznan.pl

Abbreviations: LCOs, lipochitooligosaccharides; Nfs, Nod factrors; ORF, open reading frame; SAM, *S*-adenosylmethionine; SFG, spotted fever group; T-DNA, transferred DNA; TG, typhus group.

indigenous abilities to exchange and accumulate genetic information. The latter are considerably facilitated when housekeeping genes are physically separated from adaptive loci which are useful only in certain circumstances.

GENOME REDUCTION OF INTRACELLULAR OBLIGATE SYMBIOTIC AND PATHOGENIC *BACTERIA*

A great number of *Bacteria* live solely in eukaryotic cells or tissues as chronic pathogens or mutualistic bacteriocyte associates. Some of these species, usually human (or animal) pathogens are the objects of extensive research including sequencing of their complete genomes. Two such complete genome sequences have become available in recent years in public databases. The first relates to *Rickettsia prowazekii* Madrid E strain pathogenic to humans, whereas the other is *Buchnera* sp. ASP, a mutualistic bacterium living in symbiosis with Aphids. There are some other obligate intracellular bacteria whose genomes should be completed in near future. They include one *Buchnera aphidicola* strain, at least two *Rickettsia* species and three *Wolbachia* strains. Despite distinct phylogenetic origins, all these bacteria bear certain common characteristics, directly resulting from their intracellular life style. This mode of life limits the possibility of acquisition of foreign genes *via* lateral gene transfer. Moreover, all obligate intracellular bacteria have a much lower effective population size than free-living ones. The small population size causes a state of relaxed selection, thus allowing accumulation of moderately deleterious mutations (Wernegreen & Moran, 1999). This phenomenon is known as near-neutral evolution or Muller's ratchet (Moran, 1996). One of the consequences of Muller's ratchet is accelerated evolution of all gene sequences (Brynnel *et al.*, 1998). On the other hand, the availability of compounds in the host cell, and relative safety in constant cellular environment renders many genes expendable. The consequence of these two factors is inactivation and subsequent loss of genes, which finally causes genome reduction often approaching the lowest size limits. Additionally, the loss of certain functions involved in DNA repair and recombination results in a strong mutational bias towards a high A+T content, a feature usually associated with intracellular mode of life. All these features define the so-called "resident" genome (Andersson & Kurland, 1998). However, obligate intracellular bacteria share some other features as well. They include a small number of regulatory genes as well as a reduced amount of genes linked to uptake or transport of the compounds from the outside environment.

Phylogenetic analysis reveals that intracellular obligate bacteria belong to distinct and usually deeply branching lineages. This is interpreted as the factor which precludes frequent shift from symbiotic to pathogenic lifestyle and *vice versa* (Moran & Wernegreen, 2000). Irreversible loss of genes which could contribute to either pathogenic or mutualistic association appears to be the main cause of lineage stability that is observed in most obligate intracellular bacteria. Nevertheless, there are examples of closely related lineages comprising both mutualistic and pathogenic bacteria, such as Flavobacteria, one lineage of which contains an obligate mutualist of cockroaches, whereas the other comprises male-killing parasites in ladybird beetles. Similarly, *Wolbachia* spp. include reproductive (male-killing) parasites of arthropods and mutualists of nematodes (Bandi *et al.*, 1992; Hurst *et al.*, 1996). Perhaps, for that reason genomes of *Wolbachia* spp. are still rather large (1.4–1.7 Mb), suggesting that genome reduction in these bacteria is in its initial phase.

In this paper we focus on *Buchnera* which is a symbiont of Aphids and whose genome has probably reached one of the lowest size limits, as well as on *Rickettsia* spp. in which the genome reduction is still an ongoing process.

GENOMES OF *BUCHNERA APHIDICOLA* AND ITS APHID HOST ENCODE COMPLEMENTARY METABOLIC FUNCTIONS

Buchnera spp. are endosymbionts of Aphids, in which they spend their entire life inhabiting a specialized cell-line, the so-called bacteriocytes. A phylogenetic analysis has revealed that this symbiotic association between *Buchnera* and Aphids was established some 200–250 million years ago (Baumann *et al.*, 1995; Brynnel *et al.*, 1998; Ochman *et al.*, 1999). Such a long time has resulted in a close integration of their metabolisms, and complete mutual dependence of the partners on each other. It has been noticed that *Buchnera* shares features of both pathogenic bacteria and eukaryotic organelles, being probably intermediate between the two (Andersson, 2000). *Buchnera* provides its host with a variety of nutrients, including essential amino acids, vitamins, and probably some nucleotides (Baumann *et al.*, 1995). Recently, a complete, 0.64 Mb genomic sequence of *Buchnera* sp. APS strain has been published (Shigenobu *et al.*, 2000). *Buchnera* sp. APS strain is an endosymbiont of the pea aphid, *Acyrthiosiphon pisum*. This second, smallest genome published to date is composed of a circular chromosome and two small plasmids, harboring 583 open reading frames a total. One of the plasmids carries *leuABCD* operon (pLeu plasmid), while the other *trpEG* operon (pTrp plasmid) (Rouhbakhsh *et al.,* 1996; Silva *et al.*, 1998). Thus, some genes that essentially contribute to this unique association occur in multiple copies that may positively influence the amount of the amino acids synthesized. Actually, all genes irrespective of their location are multiple copy ones, since each cell of these bacteria contains an average of 120 genomic copies (Komaki & Ishikawa, 1999). The average G+C content of *Buchnera* genome is 26.3%. Similarly to other prokaryotes of comparable genome size, including all intracellular *Bacteria*, *Buchnera* genome harbors single copies of 16S, 5S and 23S rRNA genes and only 32 tRNA genes. The chromosome harbors 564 ORFs, with average size of 988 bp, which cover 88% of chromosome length. Both the ORF size and the percentage of coding regions are similar to those found in the majority of sequenced prokaryotic genomes. Interestingly, unlike free-living prokaryotes, the *Buchnera* genome misses insertion or phage-related sequences. This implies that lateral gene transfer played a very limited role in the evolution of these bacteria, as well as that there was a strong pressure to eliminate redundant or expendable sequences. *Buchnera* sequences have been the first published ones, and majority of them have their counterparts in the database: For 500 out of 583 ORFs a function based on similarity searches in the database could be assigned. For other 79 ORFs, similar genes albeit of unknown functions were found, while only four ORFs appear to be unique. As expected, the majority of most similar ORFs originate from *Escherichia coli*, which is phylogenetically most related among all fully sequenced *Bacteria* (Shigenobu *et al.*, 2000).

Genome analysis has revealed that *Buchnera* harbors genes for biosynthesis of essential amino acids, while those which are responsible for non-essential amino acids are almost completely missing. Thus, *Buchnera* possesses only those genes which are lacking in the host genome. Similar mutual dependence can be found for pantothenate-coenzyme A (CoA) biosynthesis. The genes for pantothenate are present in *Buchnera*, while the host cells lack such functions. On the other hand, no genes for the pathway from pantothenate to CoA were found in *Buchnera*, while the eukaryotic cell expresses this part of the pathway. For that reason, finding of only a few genes involved in transport was rather unexpected. Besides, *Buchnera* genome carries only a few genes for cell-surface components, since the host provides some components necessary for lipopolysaccharide synthesis. There are also only a few genes encoding

outer membrane proteins and lipoproteins. Scarcity of genes coding for cell surface components renders *Buchnera* cells vulnerable to environmental challenges and fully dependent on its host's cells (Shigenobu *et al.*, 2000).

Another peculiarity of *Buchnera* genome is the lack of *recA*, *lexA*, *umuCD*, and *uvrABC* genes, which are responsible for homologous recombination and DNA repair, as well as the lack of genes involved in DNA methylation and restriction. Presumably, the lack of these genes is responsible for mutational bias towards a high AT content, which is usually observed among intracelullar species (Moran & Wernegreen, 2000).

GENOME DEGRADATION IN *RICKETTSIA*

Rickettsia spp. are obligate intracellular parasites that belong to the alpha proteobacteria. These bacteria are usually associated with arthropods, from which they are transfer to humans (Raoult & Roux, 1997). This genus can be divided into two groups: one (the typhus group or TG) comprises *R. prowazekii* and *R. typhi* spp., which are pathogenic to humans and mice, while the other so-called spotted fever group (SFG) includes *R. rickettsii*, a species known as an etiological agent of Rocky Mountain spotted fever. *Rickettsia* genomes are larger than those of *Buchnera*, and range from 1.1 to 1.4 Mb. Presumably, this genus originates from a free-living ancestor whose genome was much larger (Andersson *et al.*, 1998). The genome of *R. prowazekii* strain Madrid E contains 834 complete ORFs of average length 1005 bp. A biological role has been assigned to 62.7% of ORFs, while 12.5% have similar counterparts although of unknown function. Interestingly, this genome carries a much higher proportion of non-coding sequences (24%) than most prokaryotic chromosomes characterized to date, the average percentage for which is about 10%. Only small fractions of *R. prowazekii* non-coding sequences, i.e., 0.9% and 0.2% are represented by pseudogenes or non-coding repeating sequences, respectively. The remaining 22.9% do not code for proteins composed of more than 100 amino acids. A small number of reiterated sequences is common among obligate intracellular bacteria, moreover, all these sequences are relatively short (< 500 bp), and occur in intergenic regions. A low G+C content (mean 23.7%), slightly lower than the average for the whole genome (29.1%) is a characteristic feature of non-coding sequences.

Consistently with other findings concerning obligate intracellular bacteria, the number of genes involved in biosynthetic pathways in *Rickettsia* is highly reduced. This concerns the genes responsible for amino acid synthesis, as well as those genes involved in *de novo* synthesis of nucleosides. The latter are most likely taken up from the host cell cytoplasm in the form of monophosphates, which later are converted into di- and triphosphates by enzymatic machinery of the pathogen. Unlike in *Buchnera*, *Rickettsia* genome harbors a full complement of genes coding for tricarboxylic acid cycle–respiratory chain complexes. It includes also ATP/ADP translocases that enable the uptake of ATP directly from the host in initial stages of infection.

Like *Buchnera*, *R. prowazeki* carries genes encoding α, β, and β' subunites of RNA polymerase, and σ^{70} and σ^{32} factors. The latter is absent in the majority of small genomes such as those of *Borrelia burdorferi*, *Helicobacter pylori*, *Chlamydia trachomatis* although these bacteria have heat shock encoding genes (Alm *et al.*, 1999; Fraser *et al.*, 1997; Stephens *et al.*, 1998; Tomb *et al.*, 1997). Likewise, *Rickettsia* has fewer genes involved in DNA repair and recombination, for instance *mutH*, *mutY* genes and *recBCD* operon are missing.

The genome of *Rickettsia* has 21 genes coding for 18 out of 20 aminoacyl-tRNA synthetases which are necessary for protein synthesis. The genes encoding glutaminyl-tRNA (*glnS*) and asparaginyl-tRNA (*asnS*) synthetases are missing. This suggests that, like in

the majority of *Bacteria*, Gln-tRNAs and Asn-tRNAs are formed following transamidation reactions of glutamic and apartic acids, respectively (Handy & Doolittle, 1999).

Like in other intracellular obligate bacteria, the number of regulatory genes in *Rickettsia* seems to be significantly reduced. Among these genes are a few members of two-component regulatory systems, such as *barA*, *envZ*, *ntrY*, *ompR* and *phoR*, all of which are also missing in *Buchnera* genome.

Unlike *Buchnera*, *Rickettsia* genome has most genes involved in lipopolysaccharide synthesis, including *lpxA*, *lpxB*, *lpxC*, *lpxD* genes. This genome contains most genes implicated in ketodeoxyoctonate synthesis (including *kdsA*, *kdsB* and *kdtA*), and several genes coding for outer membrane proteins. It carries genes involved in protein excretion such as *secA secB*, *secD*, *secE*, *secF*, *secG* and *ffH* genes. In comparison, *Buchnera* genome has *secA*, *secB*, *secE*, *secG*, and *ffH*, but not *secC*, *secD*, and *secF* genes (Shigenobu *et al.*, 2000).

The molecular basis of pathogenicity is still unclear. Nevertheless, the genome analysis has revealed two types of genes whose homologs may belong to principal factors rendering the bacterium pathogenic. The first relates to *virB* homologs of *Agrobacterium tumefaciens*, in which genes of this type are associated with transfer of T-DNA (Kado, 2000). However, the lack of *virD2* and *virE2*, which in *A. tumefaciens* encode proteins conferring DNA transfer by binding to single-stranded T-DNA, may suggest yet another role. It seems likely, especially taking into account that homologues of *virB* genes in *Bordetella pertussis* and *H. pylori* are related to protein secretion. Thus, in *R. prowazekii* these *virB* homologs may be involved in both conjugal DNA transfer and protein export. Two other putative determinants of pathogenicity are homologous to *capD* and *capM* genes of *Staphylococcus aureus*, in which these two genes participate in synthesis of capsular polysaccharide, which is one of the principal determinants of pathogenicity in this species (Lin *et al.*, 1994).

Unlike other obligate intracellular bacteria whose genome sequences have been determined in recent years, *R. prowazekii* carries a much higher proportion of non-coding sequences. This indicates that in this group of bacteria the genome reduction is still an ongoing process. This may imply that *Rickettsia* ancestors for a much longer time have remained in at least partially free-living state than other obligate intracellular species. The most striking example of initial evolutionary processes that lead to gene inactivation comes from a comparison of gene sequences coding for *S*-adenosylmethionine synthetase (*metK*) in several *Rickettsia* spp. This gene has housekeeping function and the encoded enzyme is responsible for biosynthesis of *S*-adenosylmethionine (SAM), a substrate necessary for methylation processes (Newman *et al.*, 1998). In most of the *Rickettsia* species analysed, this gene is inactive although in each case the mutation has distinct nature (Andersson & Andersson, 1999).

SYMBIOTIC NITROGEN FIXATION WITH LEGUMINOUS PLANTS IS A FEATURE THAT IS LARGELY CONFINED TO THE ALPHA PROTEOBACTERIA

Like *Rickettsia*, all rhizobia belong to alpha proteobacteria. Currently, this group of symbiotic bacteria is classified into the genera: *Allorhizobium*, *Azorhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* (Broughton & Perret, 1999; van Berkum & Eardly, 1998). However, the identification of *Methylobacterium nodulans* extends the scope of this symbiosis onto another group of the alpha proteobacteria (Sy *et al.*, 2001). A unique property of symbiotic nitrogen fixation by legumes and rhizobia is the formation of nodules on the roots, or, in some cases, also on the stems. The nodules are novel plant organs

whose main, and presumably the only function, is nitrogenfixation. Within these structures the rhizobia undergo transformation into bacteroids that are able to fix atmospheric nitrogen. Although the legume determines a nodule type and a site of nodule formation, a signal molecule that triggers nodule formation is produced by a rhizobium. Bacterial signals are lipochitooligosaccharides known as Nod factors (NFs or LCOs) (Lerouge *et al.*, 1990). There are two main features which are associated with Nod factors. One concerns the recognition process that allows a legume plant to select its proper microsymbiont, while another is induction of divisions of meristematic cells in the root tissues, which in consequence gives rise to a nodule.

Noteworthy is the fact that the majority, if not all genera of rhizobium have diverged prior to the emergence of leguminous plants, which occurred not earlier than 140 million years ago. Taking into account that the most phylogenetically distant *Bradyrhizobium* genus diverged from the last common ancestor of all rhizobia some 500 million years ago, i.e., well before the emergence of land plants (Turner & Young, 2000), it is apparent that during most of the time these rhizobia were non-symbiotic. Presumably, the same might be said about the remaining genera. One can only speculate about rhizobium life style prior to the emergence of legumes. Nonetheless, there are suggestions that they may have been soil saprophytes, possibly living in the rhizosphere or in an endophytic association with plants (Chaintreuil *et al.*, 2000). In contrast to obligate intracellular bacteria, the intracellular stage in this case is limited to a fraction of all cells that multiply in the rhizosphere, or within the plant root. This is logical, since transformation into bacteroids is presumed to be irreversible. For this reason, like other soil bacteria, rhizobia must carry numerous genes necessary for living in soil, being moreover equipped with functions allowing for invasion into and survival in an eukaryotic cell. The latter ability might be a very ancient one, possibly carried even since the time preceding the appearance of mitochondria (Andersson *et al.*, 1998).

ORGANIZATION OF RHIZOBIUM GENOME FACILITATES ACQUISITION OF FOREIGN GENES NECESSARY TO COPE WITH ADVERSE ENVIRONMENTAL CONDITIONS

Apparently, living in soil is something very different from life in a rather constant eukaryotic cell environment. Thus, a rhizobium cell is usually well adapted to various adverse conditions such as suboptimal temperatures, drought (or excess of water), salinity, alkaline or acid conditions. Moreover, living in an environment which is poor in nutrients, they must be able to compete with other microorganisms, some of which produce toxic or inhibitory compounds. As could be expected, rhizobium genome harbors all these functions that are necessary to survive in such a rigid environment. It includes the ability to synthesize cell components from simple substrates and to use numerous compounds as energy, carbon and nitrogen sources. Additionally, the rhizobium must be equipped with a number of genes involved in quorum-sensing, intercellular communication and signaling, as well as with a number of regulatory, uptake and secretion genes. To accommodate all these functions, rhizobia are expected to have larger genomes than the species sequenced to date. Indeed, the genome of *Bradyrhizobium japonicum* USDA110 has been assessed to be of 8.7 Mb (Kundig *et al.*, 1993), the size that is close to the largest prokaryotic genomes of *Myxococcus xanthus* and *Stigmatella aurantiaca*, whose sizes have been estimated for 9.2 MB and 9.2–9.9 Mb, respectively (Casjens, 1998). Actually, the 7.6 Mb-large genome of *Mesorhizobium loti* MAFF303099 strain is the largest prokaryotic one completed to date (Kaneko *et al.*, 2000).

Genome expansion would not be possible without lateral gene transfer. Actually, lateral gene transfer seems to be the major force responsible for shaping gene content and organization of prokaryotic genomes. It is also the most effective mechanism responsible for acquisition of foreign genes that are necessary for occupation of a novel niche (Ochman *et al.*, 2000). Some bacterial species show the linkage equilibrium, which implies a high frequency of recombination caused by lateral gene transfer (Smith *et al.*, 1993). However, "novel" genes, if not carried by a broad range plasmid, rarely find homologous sequences in the recipient genome. In such cases, their successful integration usually depends on recombination (either legitimate or illegitimate) mediated by other mobile elements, such as insertion sequences, transposons or phages (Ochman *et al.*, 2000). Genome analysis shows that mobile elements, as well as reiterated and other accessory sequences are not randomly distributed, but are located mainly in discrete regions known as recombination hot spots (Romero *et al.*, 1991). Such a state could be caused by a pure chance although it might be also an indirect consequence of selection. It is conceivable that the presence of housekeeping genes on a plasmid may indirectly limit the number of those DNA elements, which enhance genomic instability, rendering such plasmid more stable. This in turn may favor the acquisition of other housekeeping genes, resulting in the final change of its status from "accessory" to "chromosomal". In this way some plasmids may have evolved into a second chromosome. On the contrary, the lack (or loss) of housekeeping genes makes more likely an accumulation of accessory elements, since a higher level of genomic instability has a limited effect on cell fitness, if deletions are confined only to expendable sequences. Probably, the same logic could be applied to explain the mosaic character of chromosomes, where adaptive genes are often not only separated from housekeeping ones, but are also located in regions rich in insertion elements, transposons and reiterated sequences. Thus, deletions that are often formed during recombination events encompass mostly non-coding sequences or genes that are dispensable. This allows the integration of numerous genes without loss of the essential ones (Moxon *et al.*, 1994). Another mechanism often reported in the regions which bear genes related to pathogenicity, exploits tRNA genes as target sequences for integration through homologous recombination. Such discrete regions usually flanked by two direct tRNA gene repeats are termed pathogenicity islands. Their distinct G+C content, as well as the presence of integrase (and other mobility loci), and genetic instability argue for the generation of pathogenicity islands by lateral gene transfer, a process that is well known to contribute to microbial evolution (Hacker & Kaper, 2000). High level of conservation of tRNA gene sequences makes them ideal targets for recombination of DNA fragments, even when a sequence originates from a phylogenetically distant species. Although most newly acquired sequences are neutral or deleterious, and therefore are lost (or the bacteria harboring them become outcompeted), some may ultimately develop into a function that allows occupation of a novel niche. Thus, the mosaic structure of chromosomes and plasmids, as well as a distinct selective status of particular regions facilitate genome plasticity necessary for adaptation to changing environment and reduce the costs related to this process. Finally, this is one of the mechanisms responsible for the increase of genome size.

SYMBIOSIS PLASMIDS, ISLANDS AND REGIONS

Symbiotic nitrogen fixation is a "composite" function. Conventionally, symbiotic genes are divided into two groups: genes involved in nodulation (*nod*, *nol* and *noe*), and those related to nitrogen fixation (*nif* and *fix*). All these genes belong to "adaptive" loci, i.e., they

seem to be dispensable for cell functioning (at least in laboratory conditions) but enable occupation of a discrete ecological niche (Perret *et al.*, 2000; Preston *et al.*, 1998). In some rhizobia, the majority of symbiotic loci are located on an indigenous plasmid, the so-called symbiosis, or Sym plasmid. It seems that such location of symbiotic functions occurs in all species belonging to the genera *Rhizobium* and *Sinorhizobium*, as well as in many *Mesorhizobium* spp., while in *Azorhizobium*, *Bradyrhizobium* and *Mesorhizobium loti* symbiotic genes are located on the chromosome (Schlaman *et al.*, 1998).

Although more than 60 genes directly involved in nodulation have been identified to date, a given strain carries only 15–20 *nod* genes (Schlaman *et al.*, 1998). There is no single gene arrangement of nodulation loci. The most frequently found is composed of three common *nodA*, *nodB* and *nodC* genes (occur in all rhizobia) which are followed by *nodI* and *nodJ*, both involved in Nod factor transport, as well as by a few *hsn* (host specificity nodulation) genes responsible for chemical modifications of the Nod factor (Mergaert *et al.*, 1997). Such *nod* gene operon is under control of *nodD* gene of the LysR family of prokaryotic transcriptional regulators (Downie, 1998). Interestingly, the nodulation clusters have not been reported in organisms other than the rhizobia, albeit somewhat lower G+C content of nodulation genes in comparison to G+C content of non-symbiotic loci could suggest the opposite. While the origin of nodulation functions remains unknown, some Nod proteins show a distant similarity to proteins found in unrelated organisms.

The presence of non-symbiotic rhizobia clones in natural populations has suggested that the loss of symbiotic loci has rather a moderate effect on strain survival. Although this issue deserves additional studies, those carried out so far indicate that non-symbiotic rhizobia may constitute the majority of clones in soil populations (Segovia *et al.*, 1991). The recurrent loss and acquisition (*via* lateral transfer) of symbiotic loci could have some evolutionary significance provided that such functions are organized as discrete gene clusters. Actually, lateral gene transfer seems to be a major factor responsible for clustering genes into functional operons (Preston *et al.*, 1998). While curing of symbiosis plasmid is generally easy, a derivative of *S. meliloti* missing pSym megaplasmid has been reported very recently, suggesting the presence of genes which influence the growth rate even in a rich medium (Oresnik *et al.*, 2000). Symbiotic genes can be harbored by a 1200 kb (or larger) replicon, i.e., some of them are larger than the whole prokaryotic chromosomes. The symbiosis plasmids differ significantly even among closely related strains, however it seems rather unlikely that the main cause of differences is the number of symbiotic genes (Baldani *et al.*, 1992; Hynes & McGregor, 1990). Actually, earlier studies have suggested that a small number of genes is necessary for development of effective symbiosis. For instance, the *nod-nif-fix* region on 180 kb pSym (plasmid **a**) of *R. leguminosarum* bv. *trifolii* ANU843 is confined to a 32 kb DNA fragment (Innes *et al.*, 1988). However, pSym of ANU843 lacks some essential genes, e.g., it misses the *fixNOQP* operon, therefore such a conclusion on the limited number of symbiotic genes may not be justified. Actually, taking into account only recent sequencing data, there could be as many as several hundred genes. For such conclusion seems to indicate the studies concerning the symbiosis regions (both plasmids and islands) in *Sinorhizobium* sp. NGR234, *M. loti* strains and in *B. japonicum* USDA110, all of which comprise DNA fragments of >400 kb (Freiberg *et al.*, 1997; Göttfert *et al.*, 2001).

In *M. loti* the chromosomal symbiotic genes form the so-called symbiosis island. Unlike other symbiotic regions, a symbiosis island carries genes responsible for excision and integration within the target phenylalanine tRNA gene sequence (Sullivan & Ronson,

1998). Both excision and integration are carried out by an integrase of the phage P4 family. Importantly, neither of these two processes disrupts the continuity of tRNA gene. Actually, the island integrates into phe-tRNA gene, reconstructing the gene at the integrase end and forming a 17 bp repeat of the 3′ end of phe-tRNA at the other end of this island. Additionally, the island harbors genes involved in biosynthesis of biotin, thiamine and nicotinate for which non-symbiotic clones are auxotrophic. This gives a selective advantage over non-symbiotic clones even prior to the onset of symbiosis, explaining the dissemination of the symbiotic genes among cognate *Mesorhizobium* strains. This >500 kb DNA fragment carries all genes which are known to be associated with symbiosis including *fixNOQP*, *fixGHIS*, *exsBCD* and *dctABD* operons, which usually are not carried by symbiosis plasmids, as well as many other genes of largely unclear function. It is noteworthy that, the genome of *M. loti* MAFF303099 carries a certain number of genes, the finding of which was rather unexpected. To such genes belong *nodE*, and *nodF* as well as *nodG*, *nodP* and *nodQ* genes. The presence of well conserved *nodE, nodF* together with *nodZ*, *nolL* genes is surprising as the $\alpha-\beta$ unsaturation of Nod factor fatty acid chain (for which *nodE* and *nodF* are responsible) has not been so far reported in rhizobia whose Nod factor reducing end is glycosylated (in this case it carries acetylfucose) (Downie, 1998; Kaneko *et al.*, 2000). The presence of acetylfucose moiety, which is conferred by *nodZ* and *nolL* genes, appears to be a crucial modification responsible for recognition of *Lotus* (or lupine) plants by *M. loti* (Lopez-Lara *et al.*, 1995; Stacey *et al.*, 1994). Although, in this strain, the genes related to specific modifications of Nod factor fatty acyl chains presumably are inactive (or silenced), their presence potentially gives to the strain a possibility to infect (or adapt to) those legumes species which recognize distinct unsaturation levels. This may happen provided that gene(s) for instance *nodZ*, is inactivated. That this could be the case, is shown by rather recent inactivation of *noeE* gene, whose still well-preserved sequence remains in the 410 kb-long symbiotic region of *B. japonicum* USDA110. The *noeE* gene encodes a sulfotransferase specific for fucosylated Nod factors. Interestingly, sulfation of fucose molecule which is present at Nod factor reducing end has never been reported in soybean rhizobia (Carlson *et al.*, 1993; Hanin *et al.*, 1997; Quesada-Vincens *et al.*, 1998). Thus, the loss of *noeE* gene may be regarded as a specific adaptation towards the soybeans. This symbiotic region carries some other genes whose inactivation took place rather recently. They include for instance, a few genes involved in hydrogen uptake (*hupD*, *hupH*, *hupK*, *hypA* and *hypB*), several genes encoding type III protein secretion system, and a gene involved in transport of branched amino acids (*braC*). It is not clear whether the loss of these genes had any impact on symbiosis, or if genes of similar function compensate for their loss. Nevertheless, this also suggests that certain genes may be linked to symbiosis loci rather accidentally, probably the linkage resulting from co-transfer with the symbiotic genes.

It seems that symbiosis regions can significantly differ even among closely related rhizobia. In lupine-nodulating *Bradyrhizobium* sp. WM9, a DNA fragment carrying most nodulation and a few nitrogen fixation genes has the same gene content and gene arrangement as that of *B. japonicum* USDA110. However, apart from the genes present in both strains, the symbiosis clusters of *Bradyrhizobium* sp. WM9 carry genes which are not present on the 400 kb symbiosisregion of *B. japonicum* USDA110. Moreover, nucleotide sequences of nodulation genes of *Bradyrhizobium* sp. WM9 are much less similar with respect to *nod* genes of *B. japonicum* USDA110 than the latter are to *B. elkanii*, and *nod* gene phylogeny contradicts the phylogeny deduced upon analysis of nonsymbiotic genes. In this case, *Bradyrhizobium* sp. WM9 and *B. japo-*

nicum USDA110 are in the same branch on 16S rRNA and *dnaK* phylogenetic trees (Legocki *et al.*, 1997; Sikorski *et al.*, 1999; Stêpkowski *et al.*, 2001).

To some degree, various proportions of non-coding sequences, insertion and other accessory elements may be responsible for differences in size among symbiosis plasmids. This seems to be the case, since (as it has been discussed above) symbiosis clusters, like other adaptive loci carry more insertion and mosaic elements than those mainly composed of the housekeeping genes. The sequencing of symbiosis regions in *Sinorhizobium* sp. NGR234, *M. loti* and *B. japonicum* fully confirms this assertion. In all these species, insertion and mosaic elements make up approximately one-fifth of the total symbiotic sequence. Most insertion sequences or mosaic elements are clustered, and some flank the functionally important genes, implying that these genes have been acquired by recent lateral gene transfer. Interestingly, some repeated sequences are sufficiently preserved to be potential targets for homologous recombination; moreover, some of them have counterparts in the genomes of other rhizobium species (Göttfert *et al.*, 2001). In the symbiotic region of USDA110, several copies of well-preserved insertion element (all in the same orientation), referred to as RSα flank the genes related to hydrogen uptake, *nif-fix* cluster, type III protein excretion gene cluster, as well as *nod-nol-noe* nodulation cluster, respectively, implying that these distinct categories of symbiotic genes may have been acquired independently. Such "modular" arrangement facilitates accumulation of various genes, further emphasizing the "composite" character of symbiosis loci.

SYMBIOSIS PLASMID OF *SINORHIZOBIUM* SP. NGR234

The sequencing of the symbiosis plasmid of *Sinorhizobium* sp. NGR234 was a milestone in the studies on symbiotic nitrogen fixation (Freiberg *et al.*, 1997). This strain characterizes the broadest host range among known rhizobium spp., comprising more than 300 species of 112 genera (Pueppke & Broughton, 1999). The molecular basis for such extremely broad nodulation potential is still an unresolved issue. At least in part, it results from NGR234 unique ability to produce a much higher number of various Nod factors (mostly differing at their reducing end) than any other rhizobium sp. (Berck *et al.*, 1999; Jabbouri *et al.*, 1998; Perret *et al.*, 2000; Price *et al.*, 1992; Quesada-Vincens *et al.*, 1998). Nevertheless, this broad host range must be determined by some unrecognized factors as well, since the closely related *S. fredii* (shares > 95% sequence identity with nodulation genes of NGR234) nodulates many species of Leguminosae, even though it produces only one or two types of Nod factors. However, all legumes infected by *S. fredii* are nodulated by NGR234 (Pueppke & Broughton, 1999). Probably, most legumes nodulated by these two rhizobium species are promiscuous plants that tolerate various NFs. Nevertheless, nodulation of certain legumes requires the presence of intrinsic modifications conferred by host specificity genes, which are present exclusively in NGR234 (Berck *et al.*, 1999; Hanin *et al.*, 1997).

The symbiosis plasmid of NGR234 is 536 kb large, i.e., its size is close to the smallest-known genome of *Mycoplasma genitalium* (Fraser *et al.*, 1995). Out of its 416 open reading frames, 136 lack similarity to any known protein in the database. For the majority of the remaining 280, the role is still rather theoretical, based upon predictions of biochemical functions of their most similar counterparts in the database. However, what seems to be important, neither of the genes found in this symbiosis plasmid is related to transcription, translation or primary metabolic functions. This fact explains why symbiosis plasmid can be eliminated from the cell. The nodulation genes are uniquely arranged into three dis-

tinct clusters (*hsnI*, *hsnII* and *hsnIII*), all dispersed around the whole plasmid (Freiberg *et al.*, 1997). This plasmid carries all genes that are implicated in modifications of the Nod factor, including those that encode fucose transferase (*nodZ*), and fucose-specific; acetyl (*nolL*), methyl (*noeI*), and sulfate (*noeE*) transferases. However, the plasmid misses some symbiotic genes, e.g., *nodEG*, and *nodPQ* nodulation genes, as well as *fixNOQP* and *fixGHIS* operons. The latter two are necessary for respiration under microaerobic conditions that occur in nodule during nitrogen fixation (Preisig *et al.*, 1993).

Perhaps, the most essential finding was the identification of many genes never before implicated in symbiosis. Transcriptional analysis has revealed expression of 247 ORFs, while the remaining 169 ORFs, i.e., nearly 40% were either inactive genes, or their expression was undetectable or uninduced under conditions tested. Out of the expressed ones, only 22 (mostly insertion-related genes) were constitutive. Intriguingly, daidzein (a flavonoid) induced expression of as many as 147 ORFs, among which nodulation genes constituted only 20. While nodulation genes were expressed during the first hours of induction, the majority of the remaining daidzein-induced ORFs were maximally expressed after 24 h. Almost all daidzein-inducible genes were under control of nod box elements. However, only 5 nod boxes precede nodulation genes, two are not functional, while the remaining 12 regulate the expression of genes whose roles have yet not been elucidated. Among genes whose expression was not under control of nod box elements were ORFs involved in rhamnose synthesis (Hurst *et al.*, 1996).

The study of Perret *et al.* (1999), has revealed a number of genes expressed in the nodule under control of NifA-σ^{54} promoters, as well as those regulated in a NifA-σ^{54} independent manner. In addition, certain differences were found in gene expression patterns in determinate and indeterminate nodules. For instance, 20 ORFs including *nodD1* and genes coding for components of ABC transporters and trehalose synthesis, respectively, were induced only in determinate nodules. In contrast, much fewer genes were found exclusively in indeterminate nodules. Most nodule-expressed genes comprise a 55 kb cluster that harbors 10 NifA-σ^{54}-dependent promoters. Among the remaining six NifA-σ^{54} promoters, one regulates the expression of a cluster carrying cytochrome P450 operon, while two others control two opposing operons (y4nGHIJ and y4nMN), both involved in sugar metabolism. Subtractive DNA hybridization has shown which genes are missing in *S. fredii*. Among them are not only *nolL* or *noeE* (both linked to specific fucose modifications), but also genes involved in sugar transport, as well as sugar epimerase (y4nG) and aminotransferase (y4uB) genes.

More detailed studies conducted for a few genes have confirmed their symbiotic significance. Among them, the genes involved in type III protein excretion system (TTSS) attracted the greatest attention. The TTSS genes were previously described in various (unrelated) pathogenic bacteria, implying lateral transfer as a way of their dissemination among distant species. In NGR234, type III excretion may be one of the key determinants responsible for the broad host range (Viprey *et al.*, 1998). Mutations in TTSS genes abolish secretion of at least two proteins (y4xL and NolX) and strongly affect nodulation of a variety of tropical legumes including *Pachyrhizus tuberosus* and *Tephrosia vogelii*. The presence of these genes on symbiosis plasmid suggests that similar mechanisms function in both symbiotic and pathogenic associations. It can be assumed that also some other genes harbored by symbiotic plasmid may have primarily evolved in a pathogenic association.

REFERENCES

Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., de Jonge, B.L, Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F. & Trust, T.J. (1999) Genomic-sequence comparison of two unrelated isolates of the humangastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180.

Andersson, J.O. & Andersson, S.G. (1999) Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* **16**, 1178–1191.

Andersson, J.O. (2000) Evolutionary genomics: is *Buchnera* a bacterium or an organelle? *Curr. Biol.* **10**, 866–868.

Andersson, S.G. & Kurland, C.G. (1998) Reductive evolution of resident genomes. *Trends Microbiol.* **6**, 263–268.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. & Kurland, C.G. (1998) The genome sequence *of Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140.

Baldani, J.I., Weaver, R.W., Hynes, M.F. & Eardly, B.D. (1992) Utilization of carbon substrates, electrophoretic enzyme patterns, and symbiotic performance of plasmid-cured clover rhizobia. *Appl. Envir. Microbiol.* **58**, 2308–2314.

Bandi, C., Anderson, T.J., Genchi, C. & Blaxter, M.L. (1998) Phylogeny of *Wolbachia* in filarial nematodes. *Proc. R. Soc. Lond. B Biol. Sci.* **265**, 2407–2413.

Baumann, P., Baumann, L., Lai, C.Y., Rouhbakhsh, D., Moran, N.A. & Clark, M.A. (1995) Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: Intracellular symbionts of aphids. *Annu. Rev. Microbiol.* **49**, 55–94.

Berck, S., Perret, X., Quesada-Vincens, D., Prome, J.-C., Broughton, W.J. & Jabbouri, S. (1999) NolL of *Rhizobium* sp. strain NGR234 is required for *O*-acetyltransferase activity. *J. Bacteriol.* **181**, 957–964.

Broughton, W.J. & Perret, X. (1999) Genealogy of legume-rhizobium symbioses. *Curr. Opin. Plant Biol.* **2**, 305–311.

Brynnel, E.U., Kurland, C.G., Moran, N.A. & Andersson, S.G. (1998) Evolutionary rates for *tuf* genes in endosymbionts of aphids. *Mol. Biol. Evol.* **15**, 574–582.

Carlson, R.W., Sanjuan, J., Bhat, U.R., Glushka, J., Spaink, H.P., Wijfjes, A.H., van Brussel, A.A., Stokkermans, T.J., Peters, N.P. & Stacey, G. (1993) The structures and biological activities of the lipo-oligosaccharidenodulation signals produced by type I and II strains of *Bradyrhizobium japonicum*. *J. Biol. Chem.* **268**, 18372–18381.

Casjens, S. (1998) The diverse and dynamic structure of bacterial genomes. *Annu. Rev. Genet.* **32**, 339–377.

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

Downie, A.J. (1998) Functions of rhizobial nodulation genes; in *The Rhizobiaceae. Molecular Biology of Model Plant-Associated Bacteria* (Spaink, H.P., Kondorosi, A., Hooykaas, P.J.J., eds.) pp. 387–402, Kluwer Academic Publishers.

Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Venter, J.C., *et al.* (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., *et al.* (1995). The minimal gene complement of

Mycoplasma genitalium. Science **270**, 397–403.

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

Göttfert, M., Röthlisberger, S., Kündig, C., Beck, C., Marty, R. & Hennecke, H. (2001) Potential symbiosis-specific genes uncovered by sequencing a 410-kb DNA region of the *Bradyrhizobium japonicum* chromosome. *J. Bacteriol.* **183**, 1405–1412.

Hacker, J. & Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679.

Handy, J. & Doolittle, R.F. (1999) An attempt to pinpoint the phylogenetic introduction of glutaminyl-tRNA synthetase among bacteria. *J. Mol. Evol.* **49**, 709–715.

Hanin, M., Jabbouri, S., Quesada-Vincens, D., Freiberg, C., Perret, X., Promé, J.-C., Broughton, W.J. & Fellay, R. (1997) Sulphation of *Rhizobium* sp. NGR234 Nod factors is dependent on *noeE*, a new host-specificity gene. *Mol. Microbiol.* **24**, 1119–1129.

Hurst, G.D.D., Walker, L.E. & Majerus, M.E.N. (1996) Bacterial infections of hemocytes associated with the maternally inherited male-killing trait in britishpopulations of the two spot ladybird, *Adalia bipunctata. J. Invertebr. Pathol.* **68**, 286–292.

Hynes, M.F. & McGregor, N.F. (1990) Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum. Mol. Microbiol.* **4**, 567–574.

Innes, R.W., Hirose, M.A. & Kuempel, P.L. (1988) Induction of nitrogen-fixing nodules on clover requires only 32 kilobase pairs of DNA from the *Rhizobium trifolii* symbiosis plasmid. *J. Bacteriol.* **170**, 3793–3802.

Jabbouri, S.B., Reliæ, B., Hanin, M., Kamalaprija, P., Burger, U., Promé, D., Promé, J.-C. & Broughton, W.J. (1998) *nolO* and *noeI* (HsnIII) of *Rhizobium* sp. NGR234 are involved in 3-*O*-carbamoylation and 2-*O*-methylation of Nod factors. *J. Biol. Chem.* **273**, 12047–12055.

Kado, C.I. (2000) The role of the T-pilus in horizontal gene transfer and tumorigenesis. *Curr. Opin. Microbiol.* **3**, 643–648.

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.* **31**, 331–338.

Komaki, K. & Ishikawa, H. (1999) Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. *J. Mol. Evol.* **48**, 717–722.

Kundig, C., Hennecke, H. & Gottfert, M. (1993) Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome. *J. Bacteriol.* **175**, 613–622.

Legocki, A.B., Kar³owski, W.M., Podkowiñski, J., Sikorski, M.M. & Stêpkowski, T. (1997) Advances in molecular characterization of the yellow lupin-Bradyrhizobium sp. (Lupinus) symbiotic model; in *NATO ASI Series, vol. G39: Biological Nitrogen Fixation for Ecology and Sustainable Agriculture* (Legocki, A.B., Bothe, H. & Puhler, A., eds.) pp. 263–266, Springer-Verlag, Berlin, Heidelberg.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Prome, J.-C. & Denarie, J. (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781–784.

Lin, W.S., Cunneen, T. & Lee, C.Y. (1994) Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus. J. Bacteriol.* **176**, 7005–7016.

Lopez-Lara, I.M., van den Berg, J.D., Thomas-Oates, J.E., Glushka, J., Lugtenberg, B.J. & Spaink, H.P. (1995) Structural identification of the lipo-chitin oligosaccharide nodulation. signals of *Rhizobium loti. Mol. Microbiol.* **15**, 627–638.

Mergaert, P., Van Montagu, M. & Holsters, M. (1997) Molecular mechanisms of Nod factor diversity. *Mol. Microbiol.* **25**, 811–817.

Moran, N.A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 2873–2878.

Moran, N.A. & Wernegreen, J.J. (2000) Lifestyle evolution in symbiotic bacteria: Insights from genomics. *Trends Ecol. Evol.* **15**, 321–326.

Moxon, E.R., Rainey, P.B., Nowak, M.A. & Lenski, R.E. (1994) Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33.

Newman, E.B., Budman, L.I., Chan, E.C., Greene, R.C., Lin, R.T., Woldringh, C.L. & D'Ari, R. (1998) Lack of *S*-adenosylmethionine results in a cell division defect in *Escherichia coli*. *J. Bacteriol.* **180**, 3614–3619.

Ochman, H., Elwyn, S. & Moran, N.A. (1999) Calibrating bacterial evolution. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12638–12643.

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

Oresnik, I.J., Liu, S.L., Yost, C.K. & Hynes, M.F. (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. *J. Bacteriol.* **182**, 3582–3586.

Perret, X., Freiberg, C., Rosenthal, A., Broughton, W.J. & Fellay, R. (1999) High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol. Microbiol.* **32**, 415–425.

Perret, X., Staehelin, C. & Broughton, W.J. (2000) Molecular basis of symbiotic promiscuity. *Microbiol. Mol. Biol. Rev.* **64**, 180–201.

Preisig, O., Anthamatten, D. & Hennecke, H. (1993) Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3309–3313.

Preston, G., Haubold, B. & Rainey, P.B. (1998) Bacterial genomics and adaptation to life on plants: Implications for the evolution of pathogenicity and symbiosis. *Curr. Opin. Microb.* **1**, 589–597.

Price, N.P.J., Relic, B., Talmont, F., Lewin, A., Prome, D., Pueppke, S.G., Maillet, F., Denarie, J., Prome, J.-C. & Broughton, W.J. (1992) Broad-host range *Rhizobium* species strain NGR234 secretes a family of carbamoylated and fucosylated, nodulation signals that are *O*-acetylated and sulphated. *Mol. Microbiol.* **6**, 3575–3584.

Pueppke, S.G. & Broughton, W.J. (1999) *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Mol. Plant Microbe Interact.* **12**, 293–318.

Quesada-Vincens, D., Hanin, M., Broughton, W.J. & Jabbouri, S. (1998) *In vitro* sulfotransferase activity of NoeE, a nodulation protein of *Rhizobium* sp. NGR234. *Mol. Plant Microbe Interact.* **11**, 592–600.

Raoult, D. & Roux, V. (1997) Rickettsioses as paradigms of new or emerging infectious diseases. *Clin. Microbiol. Rev.* **10**, 694–719.

Romero, D., Brom, S., Martinez-Salazar, J., Girard, M.L., Palacios, R. & Davila, G. (1991) Amplification and deletion of a *nod-nif* region in the symbiotic plasmid of *Rhizobium phaseoli*. *J. Bacteriol.* **173**, 2435–2441.

Rouhbakhsh, D., Lai, C.Y., von Dohlen, C.D., Clark, M.A., Baumann, L., Baumann, P., Moran, N.A. & Voegtlin, D.J. (1996) The tryptophan biosynthetic pathway of aphid endosymbionts (*Buchnera*): Genetics and evolution of plasmid-associated anthranilate synthase (*trpEG*) within the aphididae. *J. Mol. Evol.* **42**, 414–421.

Schlaman, H.R.M., Philips, D.A. & Kondorosi, E. (1998) Genetic organization and transcriptional regulation of rhizobial nodulation genes; in *The Rhizobiaceae. Molecular Biology of Model Plant-Associated Bacteria* (Spaink, H.P., Kondorosi, A. & Hooykaas, P.J.J., eds.) pp. 361–386, Kluwer Academic Publishers.

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

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86.

Sikorski, M.M., Stêpkowski, T, Œwiderska, A., Biesiadka, J., £otocka, B., Kopcińska, J., Golinowski, W. & Legocki, A.B. (1999) Differential expression of symbiosis-related genes in yellow lupine; *in Highlights of Nitrogen Fixation Research* (Martinez, E. & Hernandez, G., eds.) pp. 125–129, Kluwer Academic/Plenum Publishers, New York.

Silva, F.J., van Ham, R.C., Sabater, B. & Latorre, A. (1998) Structure and evolution of the leucine plasmids carried by the endosymbiont (*Buchnera aphidicola*) from aphids of the family Aphididae. *FEMS Microbiol. Lett.* **168**, 43–49.

Smith, J.M., Smith, N.H., O'Rourke, M. & Spratt, B.G. (1993) How clonal are bacteria? *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4384–4388.

Stacey, G., Luka, S., Sanjuan, J., Banfalvi, Z., Nieuwkoop, A.J., Chun, J., Forsberg, L.S. & Carlson, R.W. (1994) *nodZ*, a unique host-specific nodulation gene, is involved in the fucosylation of the lipooligosaccharide nodulation signal of *Bradyrhizobium japonicum*. *J. Bacteriol.* **176**, 620–633.

Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V. & Davis, R.W. (1998) Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**, 754–759.

Stêpkowski., T., Œwiderska, A., Miedzinska, K., Czaplińska, M., Œwiderski, M., Biesiadka, J. & Legocki, A.B. (2001) Molecular and phylogenetic analysis of nodulation genes in *Bradyrhizobium* sp. WM9 (*Lupinus*) suggests early divergence of "lupine" lineage within the *Bradyrhizobium* genus. Submitted.

Sullivan, J.T. & Ronson, C.W. (1998) 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.* **95**, 5145–5149.

Sy, A., Giraud, E., Jourand, P., Garcia, N., Willems, A., de Lajudie, P., Prin, Y., Neyra, M., Gillis, M., Boivin-Masson, C. & Dreyfus, B. (2001) Methylotrophic *Methylobacterium* nodulate and fix nitrogen in symbiosis with legumes. *J. Bacteriol.* **183**, 214–220.

Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzegerald, L.M., Lee, N., Adams, M.D., Venter, J.C., *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547.

Turner, S.L. & Young, J.P.W. (2000) The glutamine synthetases of rhizobia: Phylogenetics and evolutionary implications. *Mol. Biol. Evol.* **17**, 309–319.

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

Viprey, V., Del Greco, A., Golinowski, W., Broughton, W.J. & Perret, X. (1998) Symbiotic implications of type III protein secretion machinery in *Rhizobium*. *Mol. Microbiol.* **28**, 1381–1389.

Wernegreen, J.J. & Moran, N.A. (1999) Evidence for genetic drift in endosymbionts (*Buchnera*): Analyses of protein-coding genes. *Mol. Biol. Evol.* **16**, 83–97.