Genetic snapshots of the Rhizobium species NGR234 genome

ABSTRACT

'Skimming' the genome of Rhizobium sp. NGR234 sheds new light on the fine structure and evolution of its replicons, as well as on the integration of symbiotic functions in the genome of a soil bacterium. Although most putative coding sequences could be distributed into functional classes similar to those in Bacillus subtilis, functions related to transposable elements were more abundant in NGR234. In contrast to ISs that accumulated in pNGR234a and pNGR234b, the hundreds of RIME elements seem mostly attributes of the chromosome.

INTRODUCTION

Background Many different Gram-negative bacteria colonize the nutrient-rich rhizospheres of plant roots. Some bacteria are pathogenic, whereas others form beneficial associations. In nitrate-poor soils, strains of Azorhizobium, Bradyrhizobium, Mesorhizobium and Rhizobium (collectively known as rhizobia), form nitrogen-fixing symbioses with leguminous plants. In compatible interactions, invading rhizobia penetrate their hosts through infection threads, which develop centripetally. At the same time, new structures called nodules develop from meristems induced in the cortex of infected roots. When infection threads reach nodule cells, rhizobia are released as symbiosomes into the cytoplasm of infected cells where they eventually enlarge and differentiate into nitrogen-fixing bacteroids. Continuous exchange of chemical signals between the two symbionts coordinates expression of bacterial and plant genes required for a symbiotic development. Flavonoids released by legume roots are amongst the first signals exchanged in this molecular dialog. By interacting with rhizobial regulators of the NodD family, flavonoids trigger the expression of nodulation genes (nod, noe and nol). In turn, most nodulation genes participate in the synthesis and secretion of a family of lipochito-oligosaccharide molecules, the Nod factors that are required for bacterial entry into root hairs. Little is known about how rhizobia migrate inside the infection threads, although it seems likely that genetic determinants of both partners are again involved (see). Once within the cortex, the rhizobia differentiate into bacteroids where low free-oxygen tensions help coordinate the expression of genes involved in nitrogen fixation (nif and fix). Taxonomic proposals based on DNA sequences of highly conserved genes indicate that rhizobia are a group of genetically diverse soil bacteria. Other data suggest that in populations of soil bacteria, natural genetic mechanisms exist which can transform isolates with widely different chromosomal backgrounds into nodulating bacteria (that is, rhizobia) (for review see). Comparisons of genomes of soil bacteria will help define the pools of symbiotic genes. Unfortunately, genomic studies of this kind have been hindered by the relatively large size of rhizobial genomes (6.5 to 8.7 Mb for R. meliloti and B. japonicum, respectively). Instead, as many symbiotic loci are often clustered on large plasmids in Rhizobium strains, or in chromosomal 'symbiotic islands' as in B. japonicum and M. loti, physical and genetic analyzes of symbiotic plasmids or 'islands' prevailed. Rhizobium sp. NGR234 was selected for its exceptionally broad host range, which includes more than 112 genera of legumes in addition to the non-legume Parasponia andersonii. As in R. meliloti, the genome of NGR234 is partitioned into three replicons, a chromosome of about 3.5 Mb, a megaplasmid of more than 2 Mb (pNGR234b) and pNGR234a, a 536 kb symbiotic plasmid. Although various experiments have shown that most symbiotic genes are amongst

the 416 open reading frames (ORFs) identified in the complete sequence of pNGR234a, others are carried by the chromosome and/or the mega-plasmid. Many ways of finding genes exist, but with the rapid advances in genomics, among the most effective are those that involve sequencing parts of or entire genomes. Although contiguous sequences of several symbiotic islands/plasmids will be released in the near future, R. meliloti strain 1021 as well as the phytopathogens Ralstonia solanacearum and Xanthomonas citri are the only plant-interacting microbes currently being sequenced. The cost of sequencing a complete genome is still well beyond the capability of most laboratories, however. Nevertheless, extensive information on the structure and content of genomes can be gained by randomly sequencing libraries made from total DNA. Here, we have used this approach to analyze the megaplasmid and chromosome of NGR234. A total of 2,275 individual shotgun sequences of ANU265 (a derivative strain of NGR234 cured of its symbiotic plasmid) were searched for protein and/or DNA homologies, and putative coding sequences were grouped into 28 classes according to their putative function. In addition, clones carrying various Rhizobium-specific repeated elements such as RIME1 and RIME2 were also analyzed.

CONCLUSION

Conclusions Random sequencing of ANU265 followed by homology searches of public databases resulted in the identification of 1,130 putative protein-coding sequences, of which 922 (41%) could be classified into functional groups. Comparison of these data with those derived from the complete sequence of the B. subtilis genome showed a similar distribution of putative coding sequences, except perhaps for functions related to transposable elements (Table 2). In fact, the genome of ANU265 carries more putative transposases and other IS-related functions (5.5% of all identified genes, and 2.2% of all shotgun sequences) than that of B. subtilis. Nevertheless, in proportion to their size, the chromosome and megaplasmid of NGR234 carry fewer IS sequences than pNGR234a. Furthermore, hybridization data indicate that the density of known transposable elements is higher in pNGR234b than on the chromosome (order of IS accretion is: pNGR234a > pNGR234b > chromosome). This suggests that IS elements preferentially accumulate on plasmids, possibly because they are less likely to disrupt essential functions. In contrast, the many RIME elements present in NGR234 are clearly more abundant on the chromosome and megaplasmid than on pNGR234a. Together, the distinct G+C contents and structural features of the symbiotic plasmid, megaplasmid and chromosome suggest that different evolutionary constraints and histories contributed to shape these three replicons. 'Skimming' the genome of Rhizobium sp. NGR234 has given new insights into the evolution of its replicons and the integration of symbiotic functions in the genome of a soil bacterium. It also reinforced the assumption, which originated from host-range extension experiments, that pNGR234a carries most of the symbiotic genes. Although few nod, nif and fix homologs were found amongst the random clones, it is likely that additional chromosome- and megaplasmid-encoded functions contribute to successful symbioses between NGR234 and its many host plants. In this respect, transcriptional analyses using shotgun sequences as hybridization templates will help identify such new symbiotic loci.