Meeting report

Rhizobium goes genomic
Xavier Perret and William J Broughton

Address: Université de Genève, Laboratoire de Biologie Moléculaire des Plantes Supérieures, 1 chemin de l'Impératrice, 1292 Chambésy, Geneva, Switzerland.

Correspondence: William J Broughton. E-mail: broughtw@sc2a.unige.ch

A report from the Rhizobium Functional Genomics Workshop, Sevilla, Spain, 15-16 September 2000.

Although sequences of individual rhizobial genes involved in nitrogen fixation (*nif*) and nodulation (*nod, nol* and *noe*) began to be deposited in public databases in the early- to mid-1980s, the first report of their large-scale organization only appeared in 1997. This meeting was therefore the first in which the partial and complete genome sequences of various rhizobia could be compared. Although substantial progress has been made, it was obvious that the lack of public funding of large-scale sequencing of soil rhizobia severely restricts progress in the field. At the time of writing, only two rhizobial genomes have been completely sequenced - those of *Mesorhizobium loti* and *Sinorhizobium meliloti*. Highlights of these and other rhizobial genomes were presented at the meeting.

Bradyrhizobium japonicum
Michael Göttfert (Technical University, Dresden, Germany) reported the sequence of the 'symbiotic locus' (more than 400 kb) of *Bradyrhizobium japonicum* strain USDA110. This segment of the chromosome codes for 388 open reading frames (ORFs), 29% of which are 'pioneer sequences' (with no matching database entries) and of which are 19% insertion sequences (IS). Less than one third of *B. japonicum* proteins show similarities with those of the symbiotic plasmid of the broad host-range *Rhizobium* sp. NGR234 (*J Bacteriol* 2001, **183**:1405-1412). Gary Stacey (University of Tennessee, Knoxville, USA) reported the construction of a bacterial artificial chromosome (BAC) library that provides a 77-fold genome coverage based on an estimated genome size of 8.7 Mb. The library contains 4,608 clones with an average insert size of 146 kb. To generate a contig-based physical

map, the entire library was fingerprinted with *Hin*dIII and analyzed using IMAGE and Fingerprint Contig (FPC) software (Sanger Centre, Cambridge, UK), resulting in the assembly of six large contigs. High-density colony filters of the library were prepared and probed with 48 known genes. To develop a sequence-tagged connector (STC) framework, the ends of 2,256 BAC inserts were sequenced and searched against the public databases. In combination with the STC and hybridization data, the FPC contigs were aligned to create a physical map anchored to the known genetic map. This physical map represents an estimated 8.2 Mb of the genome. These results provide a framework to assist in the cloning of important genomic regions and the sequencing of the *B. japonicum* genome. (J.P. Tomkins, T.C. Wood, J. Loh, A. Judd, B. Stacey, C. Soderlund, D.A. Frisch, M. Sadowsky and R.A. Wing, personal communication).

Mesorhizobium loti
The genome of *M. loti* has recently been completely sequenced by a group led by Satoshi Tabata of the Kazusa DNA Research Institute, Kisarazu, Japan. John Sullivan (University of Otago, Dunedin, New Zealand) and Frans de Bruijn (INRA, Toulouse, France) reported the analysis of the 500 kb 'symbiotic island' of *M. loti*. Sullivan confirmed that the symbiotic island of *Mesorhizobium* sp. R7A is a 501.8 kb mobile element, which transfers to non-symbiotic mesorhizobia in the environment, converting them to *Lotus* symbionts. It integrates into a Phe-tRNA gene of the chromosome, regenerating a wild-type copy of the gene at one border of the island. Integration is mediated by a P4-like integrase encoded by *intS*. The island may belong to a class of elements termed 'fitness islands', which, when acquired, confer an advantage on the host under specific environmental conditions. The sequence analysis showed that the symbiotic island of *M. loti* has a mosaic structure, which suggests it evolved in a stepwise fashion via multiple recombination

events. As expected, it contains common *nod* and *nif* genes, including some that are spread across several replicons in other rhizobia. Other genes in the island include pioneer sequences, genes for transcriptional regulators, genes for cell-membrane-associated components including porins, as well as genes for unknown functions also found on symbiotic replicons in other species. Interestingly, the symbiotic element includes an unexpected array of metabolic genes, which may contribute to fine tuning of nodule metabolism. Several of these would normally be considered core housekeeping genes and are presumably present as duplicated copies in the genome. A number of genes similar to those required for transfer of the *Agrobacterium tumefaciens* Ti plasmid and RP4 replicon are located on the island, including *traG*, *traF* and a *trb* operon likely to be involved in the formation of type IV pili. No homologs of the genes *repA*, *repB* or *repC*, which are required for plasmid replication, were detected, however.

These findings suggest that the symbiosis island is a broad host-range site-specific conjugative transposon. The island contains a second type IV secretion system, consisting of a *virB* operon and *virD4*, with strong similarity to the *A. tumefaciens* type IV system that mediates T-DNA transfer into plant cells. Homologs of *virA* and *virG*, which in *A. tumefaciens* constitute a two-component regulatory system required for the expression of the *vir* operons, are also encoded by the island. As the *M. loti virA* homolog is preceded by a *nod* box, the expression of the island *vir* genes may be symbiotically regulated (J.R. Trzebiatowski and C.W. Ronson, personal communication).

Rhizobium sp. NGR234

The genome of the broad host-range *Rhizobium* sp. NGR234 consists of three replicons: a symbiotic plasmid of 536 kb (pNGR234*a*); a megaplasmid of more than 2 Mb (pNGR234*b*); and a chromosome of around 3.5 Mb. In addition to the complete sequence of pNGR234*a*, a collection of random sequences from the two remaining replicons has been collected. With the aim of comparing the genome of the narrow host-range *S. meliloti* with that of NGR234, a Swiss-German collaboration, described at the meeting by Wolfgang Streit (University of Göttingen, Germany), is now completing the sequence of pNGR234*b*. Eventually, comparative analysis of both genomes will help identify the molecular keys associated with symbiotic nitrogen fixation, host-range determination and various metabolic functions. It will also provide a better understanding of the mechanisms that shape the genomes of different rhizobial species and their closest pathogenic relatives. Initial sequencing data obtained from selected regions of the megaplasmid revealed the presence of many genes associated with the turnover of complex carbon sources, and a number of genes encoded by pNGR234*b* have orthologs on the chromosome of *S. meliloti*. In addition, comparison of the loci involved in the synthesis of exopolysaccharides (*exo* and *exs* genes) in NGR234 and *S. meliloti* showed a high degree of conservation at the amino acid level, suggesting that these loci have been acquired from a common ancestor. (C. Staehelin, R.A. Schmitz, G. Gottschalk, T. Hartsch, A. Johann, and W.R. Streit, personal communication).

Sinorhizobium meliloti

The genome of *S. meliloti* strain 1021 consists of one 3.65 Mb chromosome and two megaplasmids of 1.35 Mb (pSym*a*) and 1.68 Mb (pSym*b*). Francis Galibert (Rennes University, France) described how their group constructed a BAC library of the *S. meliloti* genome and selected a minimal set of BAC clones covering each of the three replicons. BACs representing each replicon were then parceled out to other groups. pSym*a* was sequenced in California (S. Long, Stanford University, USA), pSym*b* was sequenced by A. Pühler, A. Becker, J. Buhrmester and S. Weidner (Universität Bielefeld, Germany) and T. Finan and G.B. Golding (McMaster University, Hamilton, Canada). In parallel, the sequence of the chromosome was completed by a European consortium led by Galibert. About 86% of the genome (62.1% GC) appears to encode genes, 40% of which are pioneer sequences. Surprisingly, 54% of the known ORFs from pNGR234*a* were not found in *S. meliloti*, and those that were seem to be biased towards pSym*a,* which contains 1,298 ORFs. Homologs of *virB* are present, but their mutation does not affect the symbiotic properties of the bacterium. pSym*b* contains a high proportion of transporters of the ATP-binding-cassette superfamily: 70 were positively identified - about 34 seem to be responsible for sugar import, 12 for the transport of amino acids, 3 for the transport of iron and 3 are involved in sulphonate transport. In addition to the two well studied clusters of genes involved in the synthesis of extracellular polysaccharides (EPSs), several new loci were found, bringing to 12% the total proportion of EPS genes on pSym*b*. The chromosome carries 3,339 putative ORFs and encodes most housekeeping functions, with a few exceptions. The proportion of orphan genes is significantly lower on the chromosome (9%) than on the pSym megaplasmids (19% for pSym*a* and 16% for pSym*b*).

Genome architecture and genomic design

Many bacterial genomes contain large amounts of reiterated DNA sequences. Among those most commonly found are the ribosomal RNA genes and different types of insertion sequences. Reiterated DNA sequences are potential sites for homologous recombination that results in genomic rearrangements. Recombination between direct repeats leads to either amplification or deletion of DNA segments, whereas recombination between inverted sequences causes the inversion of the intervening region. Finally, recombination between repeated sequences present in different replicons may lead to the formation of cointegrates.

From the complete DNA sequence of an organism, the potential genomic structures generated by homologous recombination can be predicted, and a dynamic map of the genome can be proposed. To test this model, the symbiotic plasmid of *Rhizobium* sp. NGR234 was analyzed, as described at the meeting by Rafael Palacios (Universidad Nacional Autónoma de México, Cuernavaca, Mexico). This replicon of 536,165 base pairs (bp) carries most nodulation and nitrogen fixation genes, as well as five major families of repeated elements: the duplicated *nifHDK* operons and four types of IS elements. Using a PCR-based methodology, the different structures derived from predicted rearrangements were experimentally identified, and strain derivatives that are pure for each rearrangement were isolated. pNGR234*a* shares reiterated sequences with both the megaplasmid and the chromosome, however. Analysis of the plasmid profiles of numerous colonies derived from the wild-type strain NGR234 led to the identification of new genomic architectures. As predicted, symbiotic plasmid and megaplasmid were found to integrate into the chromosome. Together these results show that the complete DNA sequence of a genome offers the possibility of designing pathways of sequential rearrangements leading to alternative genome structures. Palacios also proposed an experimental strategy to isolate such structures.

Functional transcriptomics and proteomics were also discussed at the meeting, but it is clear that the use of microarrays to study the expression of symbiotic genes is still in its infancy. Nevertheless, Michael Djordjevic and Jeremy Weinman (The Australian National University, Canberra) presented convincing evidence that the separation of proteins using two-dimensional electrophoresis followed by their identification using mass-spectrometric tools is a powerful technique that, when coupled to sequence information, will shed light on some of the hitherto hazy details of rhizobial interactions with legumes.