Research Note

An Open Reading Frame Downstream of *Rhizobium meliloti nodQ1* Shows Nucleotide Sequence Similarity to an *Agrobacterium tumefaciens* Insertion Sequence

Julie Schwedock and Sharon R. Long

Department of Biological Sciences, Stanford University Stanford CA 94305-5020 U.S.A.
Received 10 May 1993. Accepted 14 September 1993.

We sequenced a small uncharacterized region in the *Rhizobium meliloti nod* gene cluster downstream of *nodQ1*. We found the beginning of a large open reading frame (260 amino acids) in this fragment. The sequence reported here has striking similarity to IS66 (Y. Machida, M. Sakurai, S. Kiyokawa, A. Ubasawa, and S. Yasushiro, 1984, Proc. Natl. Acad. Sci. USA 81:7495-7499), an insertion element found in an *Agrobacterium tumefaciens* mutant.

The *Rhizobium meliloti* pSym-a megaplasmid bears a number of nodulation (*nod*) genes, many clustered together in a 25-kb segment near the nitrogen fixation (*nif*) genes. The *R. meliloti nod* genes have been identified by DNA sequencing, mutation, and phenotypic analysis, and protein studies (Dénarié *et al.* 1992; Fisher and Long 1992; Kondorosi *et al.* 1991). Two genes, *nodP* and *nodQ*, were identified as open reading frames (Cervantes *et al.* 1989; Schwedock and Long 1989, 1990) that encode ATP sulfurylase and APS kinase functions (Schwedock and Long 1990; Swanson *et al.* 1987). We sequenced the fragment downstream of *nodQ1* to determine whether there might be open reading frames for other known sulfate metabolism functions. We instead found a 260-amino acid open reading frame (ORF) encoded by a sequence that shows striking similarity to IS66, an insertion sequence identified in *Agrobacterium tumefaciens*.

The nucleotide sequencing was done using a modified form of the dideoxy chain termination method (Sanger *et al.* 1988; Vieira and Messing 1987), and computer analyses were performed using the software of Devereux *et al.* on a VAX (1984). We used the TFASTA program (Pearson and Lipman 1988) to compare the new *R. meliloti* ORF to translations of the DNA sequences in GenBank and EMBL.

The nucleotide sequence of the 1.2-kb fragment downstream of *nodQ1* and the 260-amino acid ORF it encodes are shown in Figure 1. The ORF continues past the second *Eco*RI site, as no stop codon was found in this sequence. Cervantes *et al.* (1989) reported part of this sequence previously (nucleotides 1–847); our sequence differs at nucleotide 625, which in the Cervantes *et al.* sequence creates a frameshift such that the ORF would be truncated.

This sequence has striking similarity to IS66 (Machida *et al.* 1984), an insertion element found in an *Agrobacterium tumefaciens* mutant (Fig. 2A). The sequences show 63% identity over 797 nucleotides. In addition, an open reading frame in the insertion element, ORF2, has 68% identity with the ORF reported here, over a 125-amino acid overlap (Fig. 2B).

The *R. meliloti* ORF is continuous over at least 779 nucleotides (260 amino acids). While the *A. tumefaciens* IS66 open reading frame is similar over 125 amino acids, there is also significant similarity between the *R. meliloti* ORF and other reading frames of the IS66 sequence lying upstream and downstream of the amino acid correspondence shown. It might be profitable to resequence portions of IS66, to determine whether the ORF may in fact be completely co-linear with that of *R. meliloti*. The sequence downstream of *nodQ1* appears to be singular in the *R. meliloti* genome, as we did not detect any evidence for additional copies of this sequence in the *R. meliloti* genome by Southern blots (Schwedock and Long 1989; J. Ogawa, J. Schwedock, unpublished data).

The discovery of the similarity between this *R. meliloti* sequence and IS66 poses some interesting questions. It is possible that the ORFs may be homologs, each representing a conserved domain for an as-yet unknown function encoded both in *Agrobacterium tumefaciens* IS66 and in the *R. meliloti nod* region. This function might be important in the biology of these related soil bacteria, a possibility that can be tested genetically. In previous work (Tabor and Richardson 1987), we obtained five transposon insertions in this segment and assayed homogenotized versions of the mutants: JO916, JT512, JO902, JT215, and JT711. All of these were scored as phenotypically Nod$^+$ in our standard assay on agar slopes. A slight decrease in performance of some of these strains compared to wild type at 3–4 wk can be seen in the data, but this difference is not significant. Tests on other plant species will be needed to assay for possible host range effects.

Corresponding author: Sharon R. Long, Department of Biological Sciences, Stanford CA 94305-5020 U.S.A.
Current address of J. Schwedock: The Biological Laboratories, Harvard University, Cambridge MA 02138 U.S.A.

MPMI Vol. 7, No. 1, 1994, pp. 151-153

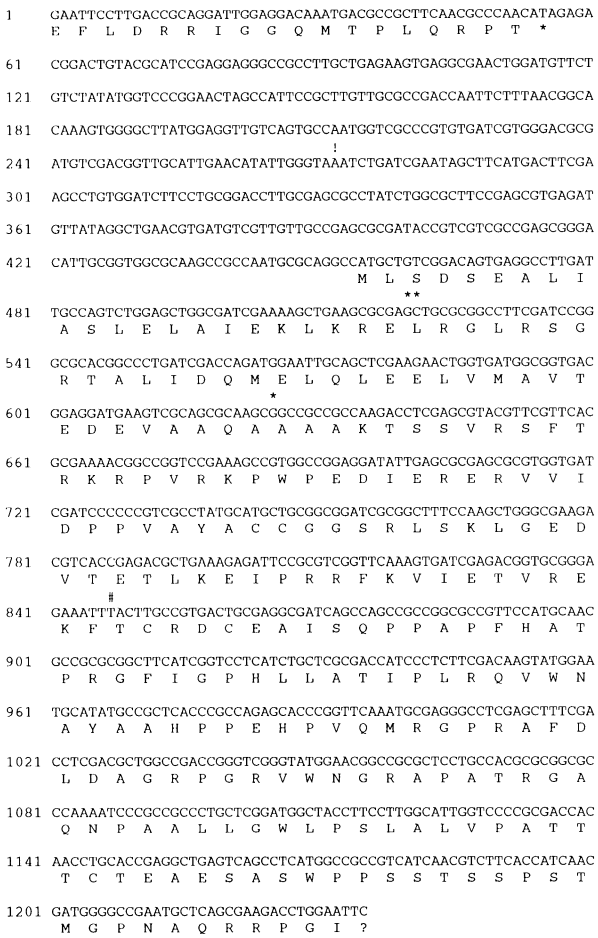

Fig. 1. Nucleotide sequence of the 1.2-kb *Eco*RI fragment downstream of *nodQ*, with the translation of a large open reading frame. The amino acid sequence under the first line represents the end of NodQ. The other amino acid sequence represents the largest open reading frame in this fragment, and continues past the sequenced region as indicated by the question mark (?). An exclamation point (!) above the sequence indicates the beginning of similarity with the nucleotide sequence of IS66. Asterisks above the sequence indicate differences between this sequence and the published sequence of Cervantes *et al.* (1989), where the first pair (**) represents a change from GC to CG at their nucleotides #3711 and 3712, and the lone asterisk (*) represents the absence of the G between their nucleotides #3815 and #3816. A number sign (#) above the sequence indicates the end of the overlap between this sequence and that of Cervantes *et al.* This sequence has been entered into the GenBank database under accession number L08667.

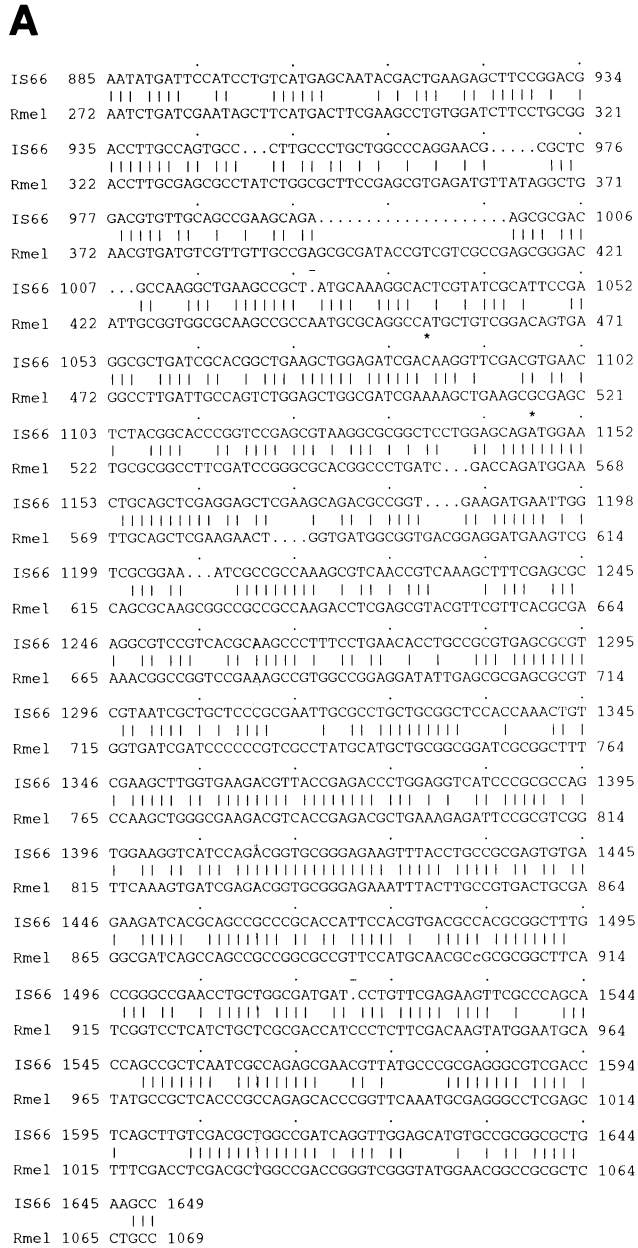

Fig. 2. The comparison of nucleotide (A) and amino acid (B) sequences from the fragment downstream of *nodQ* and IS66 from *Agrobacterium tumefaciens* (Devereux 1984). **A,** IS66 sequences are the top lines and *Rhizobium meliloti* sequences are the bottom lines of each pair. Identical bases are indicated by vertical lines (|). An asterisk (*) below the nucleotide comparison indicates the start of the ORF from *R. meliloti*, and one above indicates the start of IS66 ORF2. Minus signs (–) above the comparison indicate positions where frameshifts occur. **B,** ORFdnQ indicates the sequence of the large open reading frame downstream of *nodQ* and ORF2 is indicates the sequence of ORF2 from IS66 of *A. tumefaciens*. A "1" marks the position of the putative start codon of ORF2, though amino acids upstream of this start codon which are potentially translated in the event of a frameshift are included. Identical amino acids are indicated by vertical bars (|) and similar amino acids are indicated by colons (:). A minus sign (–) below the amino acid comparison indicates the position of the frameshift.

Alternatively, the sequence similarity reported here may represent the remnants of an insertion sequence related to IS66 in *R. meliloti*. If so, we note that the similarity breaks down 800 bp away from the first inverted repeat of IS66, and similarity to other ORFs in the insertion element is not found. Thus, it is possible that the fragment we have characterized is the result of a rearrangement or imprecise excision of such a putative insertion sequence. We also note that the position of the IS66 similarity occurs adjacent to the apparent end of a *nod* regulon unit. The *nod* genes are arranged differently in various *Rhizobium* species and how they arrived in their present arrangements is an interesting question. Insertion sequence-mediated recombination is one possible mechanism for gene rearrangement and/or for capture from exogenous genetic sources. If IS66 has played such a role, then remnants may possibly be found at the borders of *nod* gene units in other *Rhizobium* species.

ACKNOWLEDGMENTS

We thank all the members of our group for useful discussions, and in particular acknowledge J. Ogawa for unpublished genome hybridization data. This research was supported by Department of Energy contract DE-FG03-90ER20010 to S.R.L.

LITERATURE CITED

Cervantes, E., Sharma, S. B., Maillet, F., Vasse, J., Truchet, G., and C. Rosenberg, C. 1989. The *Rhizobium meliloti* host-range *nodQ* gene encodes a protein which shares homology with translation elongation and initiation factors. Mol. Microbiol. 3:745-755.

Dénarié, J., Debellé, F., Rosenberg, C. 1992. Signalling and host range variation in nodulation. Annu. Rev. Microbiol. 46:497-531.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

Fisher, R. F., and Long, S. R. 1992. *Rhizobium*-plant signal exchange. Nature 357: 655-660.

Kondorosi, A., Kondorosi, E., John, M., Schmidt, J., and Schell, J. 1991. The role of nodulation genes in bacterium-plant communication. Gen. Engin. 13:115-136.

Machida, Y., Sakurai, M., Kiyokawa, S., Ubasawa, A., and Yasushiro, S. 1984. Nucleotide sequence of the insertion sequence found in the T-DNA region of mutant Ti plasmid pTiA66 and distribution of its homologues in octopine Ti plasmid. Proc. Natl. Acad. Sci. USA 81:7495-7499.

Pearson, W. R., and Lipman, D. J. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85:2444-2448.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schwedock, J. S., and Long, S. R. 1989. Nucleotide sequence and protein products of two new nodulation genes of *Rhizobium meliloti*, *nodP*, and *nodQ*. Mol. Plant-Microbe Interact. 2:181-194.

Schwedock, J. S., and Long, S. R. 1990. ATP sulfurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. Nature 348: 644-647.

Schwedock, J. S., Liu, C.-X., Leyh, T., and Long, 1994. S. R. *Rhizobium meliloti* NodP and NodQ form a multifunctional sulfate activating complex that requires GTP for activity. J. Biol. Chem. In press.

Swanson, J., Tu, J. K., Ogawa, J. M., Sanga, R., Fisher, R. F., and Long, S. R. 1987. Extended nodulation gene region in *Rhizobium meliloti* 1021. Physical and genetic map of loci in clone pRmJT5. Genetics 117:181-189.

Tabor, S., and Richardson, C. C. 1987. Sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767-4771.

Vieira, J., and Messing, J. 1987. Production of single stranded plasmid DNA. Methods Enzymol. 153:3-11.