Nucleic Acid Sequence Citations: Need for More-Specific Guidelines

It is a policy of the Journal of Clinical Microbiology (JCM) and other ASM journals to encourage prompt submission of nucleic acid sequences to GenBank/EMBL. As stated in the Instructions to Authors, "It is expected that GenBank/EMBL accession numbers for primary nucleotide and/or amino acid sequence data will be included in the original manuscript or be inserted when the manuscript is modified."

After discussions with Dr. Richard Tilton, editor in chief of JCM, and with other colleagues, we concluded that although the policy regarding inclusion of sequence data appears to be clearly stated, it is not always followed. We also question the procedure for citing original references when publishing results of comparative nucleic acid sequence analysis. Since both of us inadvertently neglected to supply adequate information regarding previously published nucleotide sequences in recent publications, we believe that specific guidelines for references to such sequences are necessary. As both of us are strongly in favor of maintaining the highest quality of scientific publication in ASM journals and as we believe that nucleic acid sequence data are useful to many researchers in many ways, we would like to bring this to the attention of the readers of this journal.

The following are two examples of DNA or RNA sequence results that were not accompanied by either the sequences or GenBank accession numbers.

(i) In the July 1991 issue, in an article by Brenner et al. (1), the sequencing of rickettsial 16S rRNA is described in the Materials and Methods section and extensively discussed in the Results and Discussion sections. The actual sequence data could be found in reference 16 (O'Connor et al., submitted for publication). The O'Connor paper was published in the October 1991 issue of JCM.

Another feature of the References of this paper, germane to the subject of nucleic acid sequence data, is the reference to Bilofsky and Burks (GenBank) as the primary source for other 16S rRNA sequences. On the other hand, Weisburg et al. (3) cited GenBank as the source of a number of sequences previously published by these and other authors. No policy or guideline addressing the appropriateness of referral to GenBank, the use of GenBank accession numbers, or the inclusion of original references for gene sequences cited can be found in the instructions to authors. It is a question that deserves attention.

(ii) A second example of nondisclosure of sequence occurs in a paper by Gaydos et al. (2), in which the method for and results from 16S ribosomal DNA sequencing are extensively discussed. The sequence is not presented in a figure, and no GenBank accession number is given.

These examples illustrate that the instructions given to authors do not ensure the inclusion of sequence data and, further, that lack of guidelines for citation of previously published sequence data is also a problem. We would like to propose some suggestions for consideration. (i) Manuscripts in which the determination of a nucleic acid or amino acid sequence is described should require a GenBank accession number prior to final acceptance. (ii) Reviewers should be provided with a hard copy (or floppy diskette version) of such sequence data. (iii) When it is practical or desirable to split publication of a primary sequence determination from publication of the use of that sequence (for example, sequencing a gene as opposed to targeting the gene for polymerase chain reaction diagnostic applications), the publication of the sequence should precede or be concurrent with the other publication. Alternatively, the sequence can be deposited with GenBank in an accessible (unpublished) form, and the reference can be updated in the future. (iv) When references to previously published sequences used for comparative analysis are being made, every attempt should be made to cite the original publication and to refer to the GenBank accession number, if that is from where the sequence was retrieved. (v) Finally, in instances in which comparative analysis, such as phylogenetic determination, is used, other sequences used in analysis should be accessible. If unique masking or other alignment tools are used, they should be explicitly described. For example, an alignment mask should be shown in a figure.

We believe that these considerations, and possibly suggestions from other investigators, will remove the present ambiguity in the Instructions to Authors. Reviewers and editors would need to be fully apprised of the instructions. An updated policy regarding nucleic acid sequence data would be beneficial to the microbiology research community.

REFERENCES


William G. Weisburg
GENE-TRAK Systems
Framingham, Massachusetts 01701

Don J. Brenner
Division of Bacterial and Mycotic Diseases
National Center for Infectious Diseases
Centers for Disease Control and Prevention
Atlanta, Georgia 30333

K. H. Wong
Division of Bacterial and Mycotic Diseases
National Center for Infectious Diseases
Centers for Disease Control
Atlanta, Georgia 30333