Policy proposal for publication of papers with data sets from genome-wide studies

The text below summarizes the position that the Editors of *Microbiology* and the *International Journal of Systematic and Evolutionary Microbiology* (IJSEM) plan to adopt on this important issue. Publication of this draft proposal has been prompted by the accompanying letter from the Microarray Gene Expression Data (MGED) Society, which specifically addresses the issue of microarray data. The Editors of *Microbiology* and IJSEM feel that the same principles apply to other types of genome-wide studies. Both journals will continue to provide a platform for publication of any type of data that significantly increases our understanding of microbes. But this must be done in a way to preserve the quality of the scientific process and allow open movement of information between researchers. The journals, therefore, are endeavouring to define a workable process for deposition of and access to the data on which conclusions are based when it is impossible to publish complete data sets in each paper. If you have any comments please contact either of the corresponding authors at the email addresses given below.

Premise

Genome-based studies of microbial metabolic and regulatory processes are producing and will continue to produce papers that include large data sets. The large data sets obtained from genome-wide transcript, proteomics or metabolite profiles are raising new issues for journals, editors, reviewers and authors to consider as this work is reviewed and published.

Microbiology and IJSEM have accepted the so-called MIAME guidelines for expectations in publication of such studies. However, the MIAME guidelines were formulated to consider publication of results from microarray studies. As a

consequence, MIAME compliance does not apply to publications reporting on the results of genome-wide proteomics or metabolomics studies.

When considering the establishment of policies for publication of papers that use the results of genome-wide studies to analyse biological problems, issues to consider include:

- Expectations regarding description of strains, media and other growth conditions
- Deposition of wild-type reference strains in stock collections
- Primary and supplementary data needed during review or upon publication
- Access to supplementary data that benefits the scientific community after publication
- Replicates required for publication and error in biological or experimental replicates

As a consequence of these deliberations, the following proposal is made for establishment of guidelines for publication of the results of such studies in *Microbiology* and IJSEM.

- The journals encourage the publication of papers using genome-wide approaches to answer important questions in microbes. Papers which report on the results of such studies without using them to increase our understanding of a system, process or organism are unlikely to be well received by reviewers and members of the Editorial Boards.
- Authors who submit papers describing large data sets from genome-wide studies of a biological question should be prepared to provide reviewers with supplementary material at a password-protected online depository (GEO, etc.).

- Where authors wish to publish a paper whose conclusions explicitly or implicitly depend on genome-wide profiling, they should be prepared to associate the relevant complete set of genome-wide data with the published paper as supplementary data so that members of the general scientific community can access this information after publication. Links to the precise set of supplementary data that was evaluated prior to publication should be made available at public domain data repositories (GEO, etc.). Files containing supplementary data should also include details on organisms, growth conditions, etc. (so-called metadata) that might not be included in the paper.

- To enforce compliance, the standard review form will be modified to include a way for reviewers to indicate whether such supplementary material has been posted and will be accessible as soon as the paper is published.

- There should be sufficient replicates to make the data biologically meaningful, but setting rigid rules on numbers of replicates seems prescriptive since the numbers will depend on the question being addressed and the size of the differences observed.

Timothy J. Donohue¹
and Christopher M. Thomas²

¹Bacteriology Department, University of Wisconsin-Madison, Room 390, 420 Henry Mall, Madison, WI 53706, USA

²School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Correspondence: Timothy J. Donohue (tdonohue@bact.wisc.edu) or Christopher M. Thomas (c.m.thomas@bham.ac.uk)