DATA ARCHIVING STRATEGY

The proposed work will generate laboratory data and will use metagenomic data from MG-RAST web servers. All sequencing and microarray data and source code will be released to the public on the PI's web site (peteranoble.com) and results archived in the Dryad Digital repository 90 days before the end of the project period.

Laboratory data
The data collected daily will be kept as handwritten document in laboratory note books and saved on MS Excel spreadsheets. Roche/454 sequencing output (fna and sff files) and files containing information on subject data will be stored on two independent computers. As in past projects, we will produce C++ code to analyze the data. Metagenomic data, figures, tables, posters, and peer-reviewed papers will be used in the development of curriculum materials for the PI's courses. All of the data will be captured in MS Excel files, txt files, or SAS JMP files for statistical analyses.

Metagenomic and genomic data from web servers
Our project will use data downloaded from servers at the Metagenomics (MG)-RAST at Argonne National Laboratory. We use these data as references to compare with data we obtain from the study. We cite these sites in our publications.

Data and Metadata Standards
We have previously followed the metagenomic and metadata standards outlined by Ivanova et al. (2010) and Yilmaz et al. (2011) in our ongoing study. Moreover, to analyze metagenomic samples at MG-RAST, one has to follow their stringent metagenomic and metadata policies. We have followed those policies in our ongoing project and we intend to follow with them in the future because we use other scientists' data in our research, and we want other scientists to use our data in their research.

Microarray data. The Noble laboratory is compliant in terms of MIAME standards (Brazma et al. 2001) for DNA microarray data.

Policies for access and sharing and provisions for appropriate protection/privacy
We will make all the data generated in this project available to any scientist, as has been our policy in past years. For example, the data generated in Pozhitkov et al. (2006) paper has been re-analyzed by Hooyberghs et al. (2009). They found interesting relationships we did not observe in our original study. Another example of data sharing in nucleic acids is represented by Mueckstein et al. (2010) that used Pozhitkov et al. (2006) data in their study. We support and encourage scientists to critically evaluate our work and find interesting results.

We will only release the data after we have published our research in a peer-reviewed journal. The Roche/454 and microarray data will be available to public and permission will not be required. Other data will be released upon request at no charge.

Policies and provisions for re-use, re-distribution

We will endeavor to adhere to the data management policies of the NIJ, and to practice data sharing consistent with Alabama State University policies governing intellectual property, copyright and the dissemination of research products.

Plans for archiving and Preservation of access
Noble's laboratory computers are backed up twice a year and are permanent. We believe that all metagenomic data and metadata will also be permanently retained by MG-RAST.

Literature Cited

Brazma, A, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenission, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. Nature Genetics. 29: 365-371.

Hooyberghs J, Van Hummelen P, Carlon E. The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters. Nucleic Acids Res. 2009 Apr;37(7):e53. Epub 2009 Mar 6.

Ivanova et al., A call for standardized classification of metagenome projects. Environ Microbiol. 2010 Jul;12(7):1803-5.

Mueckstein U, Leparc GG, Posekany A, Hofacker I, Kreil DP. 2010. Hybridization thermodynamics of NimbleGen microarrays. BMC Bioinformatics. 11:35.

Pozhitkov A., Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M, Tautz D. 2006. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Res. 34:e66.

Yilmaz P, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol. 2011 May;29(5):415-20.