TITLE: Integration of Gene Family and Protein Family Information in the Mouse Genome Database

AUTHORS: Judith A. Blake, Joel E. Richardson, Lois J. Maltais, Richard M. Baldarelli, Ken S. Frazer, Martin Ringwald, and Janan T. Eppig

AFFILIATION: The Jackson Laboratory, Bar Harbor, Maine, U.S.A.

jblake@informatics.jax.org

The Mouse Genome Database (MGD) was developed and is maintained at The Jackson Laboratory. MGD is a highly integrated database, inclusive for genomics and biology of the laboratory mouse. Gene family characterization in MGDs is an important organizing concept and, with the growth in the number of characterized genes, places ever growing demands on our resources. Here are presented the various types of gene and protein family classifications curated in MGD. The relationships to classic and emerging concepts of gene and protein family groupings are detailed.

Collaboration with domain experts on gene family classification and nomenclature significantly improves database representation, and is a long standing practice for MGD. Gene family groupings are implicit both in the coordination of gene symbol determination between mouse and other mammalian gene nomenclature committees and in the naming of related genes within the mouse. Recently, self-organized groups of researchers with interest in a particular group of genes have emerged and have become active participants in the determination of gene family relationships. The cooperative efforts of these researchers and the genome nomenclature groups can and often does result in a revised and structured assemblage of gene symbols, names and family designations that has the pre-approval of most interested members of the scientific community.

Similarly, close coordination of the relationship between gene objects in MGD and Swiss-Prot results in the linking of MGD gene family information with Swiss-Prot protein family information. While gene/protein family relationships determined by phylogenetic analysis are understood, other groupings of proteins based on structural or functional classifications may not overlap with specific gene family groupings. Multiple classification systems and their various presentations in MGD are presented and the advantages and disadvantages discussed.

Supported by NHGRI grant HG0033O
http://www.informatics.jax.org