Editorial

DATABASE: A new forum for biological databases and curation

David Landsman[1], Robert Gentleman[2], Janet Kelso[3] and B. F. Francis Ouellette[4]

[1]Potomac, MD, USA, [2]Seattle, WA, USA, [3]Leipzig, Germany and [4]Toronto, ON, Canada

Most computational tools for biologists preferably require data in large amounts. The larger the quantity of data, the more rigorous statistical analyses can support the discovery of new hypotheses for testing in a laboratory. A variety of technological developments during the past two decades have accelerated the rate of deposition of data into databases. Currently there are many public databases where data from, for example, DNA and protein sequences or 3D protein structures, and more complex information types, like ontologies, networks and pathways are deposited, maintained, annotated, curated and stored. Indeed, more recent efforts to store, for example, phenotype (in addition to genotypes) and clinical trials signify a new tendency to gather more complex data types. The data collected in these large public repositories represent valuable and significant resources for ongoing knowledge extraction. Mining of this data using computational tools is an increasingly indispensible part of modern research, and the organized storage of the data in databases is obligatory. Indeed such approaches are likely to have serious impact on the reproducibility of results. Resourceful tools for the establishment, interrogation, rearrangement, display and interpretation of new and large databases are frequently minor points in a publication and are relegated to brief statements in methods sections or in figure legends when the final work is published. However, there are often original and creative computational methods which resulted in these discoveries but which are not communicated in the scientific literature because the description of a database and the tools to interact with it are not deemed essential to the communication.

Accepting that the archiving, curation, analysis and understanding of all of this data is a challenge, *DATABASE: the Journal of Biological Databases and Curation* will publish articles which describe the construction of novel databases and the software tools designed to interact with these databases. All submissions should describe worthy resources for the scientific research endeavor. We also plan to invite reviews and tutorials that will make the databases described in these pages more user friendly and easier to match with the tasks that need to be accomplished. In addition, manuscripts that describe collections of data and associated tools where a biologically relevant discovery or example is presented will be reviewed more favorably. We would also be prepared to review opinions, discussions and/or demonstrations of how new technologies, new data models (or data exchange models) can be used to address complexities presented by the new large datasets and/or personal identification challenges the new initiatives are presenting. The journal will also accept update reports which describe new features and content of existing databases.

The maintenance and longevity (when appropriate) of databases is an ongoing point of discussion, and we welcome opinion pieces and the presentation of how such problems could best be addressed. Scalability and federation of a number of databases, the Web 2.0 and 3.0 integration and the semantic web are also pertinent discussions for the biological database community, and we hope that *DATABASE: the Journal of Biological Databases and Curation* becomes the place where some of these ideas are discussed and deliberated. We will provide online commenting and discussion tools on the journal's website to encourage this.

Extensive and ongoing curation of the biological data being stored in public databases ensures that these data can be discovered and used optimally, and facilitates the integration of information from multiple sources. Structured collection of metadata, using standard terminology, will foster more complex and relevant analyses. *DATABASE: the Journal of Biological Databases and*

Curation invites the submission of novel strategies for the efficient and accurate curation of biological data, including systems to support ongoing curation by both individual researchers and research communities in order to ensure long-term availability and reusability of these data.

In support of the new open access policies of many funding agencies as well as the open source software movement which started in the 1980s, *DATABASE: the Journal of Biological Databases and Curation* will be a fully open access journal from launch. In addition, it will be a condition of publication that all databases and software described in *DATABASE* articles are made publicly available. The journal will be online-only, providing fast access of its full content to scientists worldwide.

Submissions to *DATABASE: the Journal of Biological Databases and Curation* are welcomed via the journal's web site at www.database.oxfordjournals.org. We also welcome suggestions for how this new forum can best serve the needs of the increasingly important field it represents.