Genome database & information systems for Daphnia

Don Gilbert, gilbertd@bio.indiana.edu  October 2002
Talk doc at http://iubio.bio.indiana.edu/daphnia/docs/ genome-dbs-talk.doc, .ppt

Anatomy of Genome database & info. system

Structure, Biology information Content, Data exchange, Informatics / software
Genome database examples: Start at euGenes http://eugen.org/ (Indiana Univ.)
GMOD - Generic Model Organism Database Construction Set at http://www.gmod.org/

FlyBase.net

Distributed project (4 sites, ~6 PI’s, ~15 curators, ~15 informaticians) ; 10 years old
Multiple databases; project data flow and exchange critical
Curated and computed data, from expt. literature, genome sequence

euGenes.org

Automated genome summaries for Human, Fruitfly, Mouse, Mosquito, Arabidopsis, C. elegans, Saccharomyces, Zebrafish
3 year, computational db project, 1 part-time bioinformatician

Preliminary Daphnia database & info system

http://iubio.bio.indiana.edu/daphnia/
Sample data include microsatellite DNA, GenBank Daphnia seqs, Medline abstracts, Blast searches, reports, data searches

Requirements for genome system

Data components??: biosequence types, literature, external data (insects, others), expression info, pathways, maps, anatomy, populations, species, ecology, organismal, stocks, people
Standard data structure and exchange schema (sequences, tables, XML)
Internet-shared, standards-based, open-source preferred
Relational database for data management
Search and retrieval software for flat file data
Flexible – data schema changes common
Performance constraints
Project uses: sequence analyses, external database comparisons
One-time analyses, publishing results
Pipeline for automated analyses, rerun as needed
Public uses (e.g. BLAST search)
Detail biological object views (sequences, genes, etc.)
Queries: simple-common, ad-hoc/general
Graphic viewers
Interactive – document editing
Batch data updates

Compute parts of info system

Web server (Apache) and modules
FTP server for bulk data exchange
Relational DBMS: PostgreSQL.org, MySQL.com, Oracle, etc.
Analysis programs: BLAST, other bioinformatics tools
Perl, Java middleware for data access & analysis, search and report
Limited, secure access for project data management (web,ftp,other)
Public access for released data (web, ftp)