

## Genome database & information systems for Daphnia

Don Gilbert, [gilbertd@bio.indiana.edu](mailto:gilbertd@bio.indiana.edu) October 2002

Talk doc at <http://iubio.bio.indiana.edu/daphnia/docs/> genome-dbs-talk.doc, .ppt

### Genome database examples

Drosophila: FlyBase, <http://flybase.net/> (Indiana Univ.)

C. elegans: WormBase, <http://www.wormbase.org/>

Mouse: MGD, <http://www.informatics.jax.org/>

Saccaromyces: SGD, <http://genome-www.stanford.edu/Saccharomyces/>

Human: LocusLink, <http://www.ncbi.nlm.nih.gov/LocusLink/>

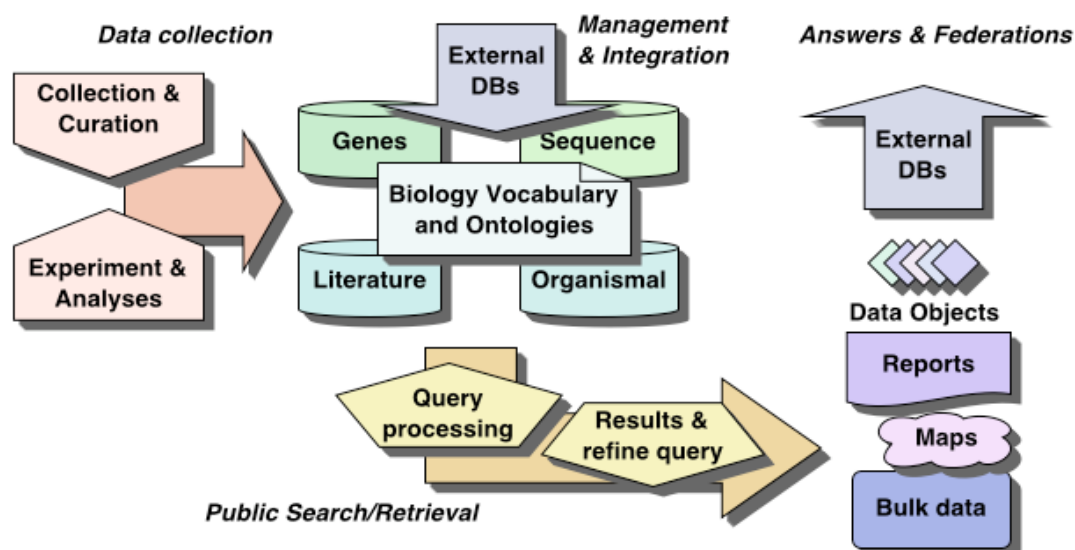
Human: GeneCards <http://bioinfo.weizmann.ac.il/cards/>

Various eukaryotes: Ensembl <http://www.ensembl.org/>

Various eukaryotes: euGenes <http://eugen.es.org/> (Indiana Univ.)

Many newly developing organism genome systems for Daphnia, insects, vertebrates, new full-genome organisms

### Anatomy of genome database & info system



## **Anatomy of Genome DB/IS**

### **Structure**

Complex document structure; tabular data; etc.

Organize: Table of contents, Reports, Indexing

Browse contents; Search / retrieve from biological questions

Bulk data search / retrieve for bioinformatics

### **Content**

Literature (abstracted and curated), Sequence and feature analyses, maps, controlled vocabulary/ontologies, people, biologics, contacts, etc.

Metadata describing primary data, along with protocols, notes, sources

## **Anatomy of Genome DB/IS, cont.**

### **Data exchange**

XML data definitions & schema

Controlled vocabularies of science terms, ontologies

Minimal information for collaboration, sharing

### **Informatics / software**

Backend database, data collection, management, analyses

Front-end services (hypertext web, search/retrieval); ease of understanding and usage (HCI)

Middleware software, interfaces

Genome specialized: maps, BLAST searches, ontologies

## **GMOD - Generic genome database tools**

Generic Model Organism Database Construction Set, <http://www.gmod.org/>

Database schemas

Literature curation tools

Genome database & information systems for Daphnia

Gene ontology management tools

Visualization tools

Data processing pipelines

## FlyBase and euGenes

The screenshot shows a web browser window with the URL <http://flybase.bio.indiana.edu/>. The page title is "FlyBase A Database of the Drosophila Genome". The main content is organized into two columns. The left column contains navigation links such as "Getting Started", "Documents", "News, meetings & announcements", "Drosophila links", "FlyBase mirrors", "Alternative views", and "Set preferences". The right column is titled "Data Classes" and "Selected Searches & Tools", listing various search categories like "Maps", "Genes", "Sequences", "Stocks", "Transgenes & Transposons", "Aberrations", "Anatomy & Images", "References", and "People". At the bottom of the right column, there is a search interface for BLAST sequence search, with a dropdown menu showing "All sections", "Genes", "References", "Stocks", and "People".

## 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

Integrated database modules

Genetics, Sequences, Maps, Expression

Genome database & information systems for Daphnia

Controlled vocabularies & Ontologies

Computational analyses

Organism, taxonomy, phylogenetic/comparative

Publications, General

Will be developed for generic use with GMOD

## euGenes.org

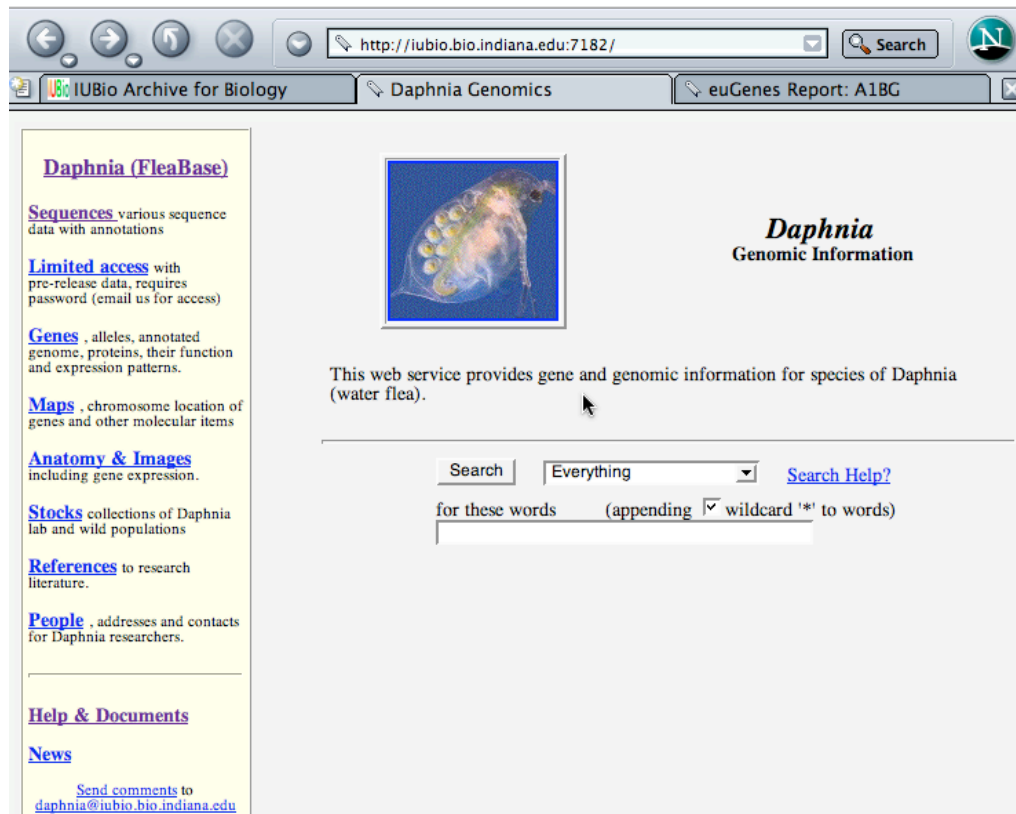
Automated genome summaries for Human, Fruitfly, Mouse, Mosquito, Arabidopsis, C. elegans, Saccharomyces, Zebrafish

3 year, computational db project, 1 part-time informatician (dgg ☺)

genome maps, sequences, gene reports, external database links

cross-species comparisons: similar genes, genome features, gene function

## A database & info system for Daphnia



The screenshot shows a web browser window with the URL <http://iubio.bio.indiana.edu:7182/>. The browser tabs include "IUBio Archive for Biology", "Daphnia Genomics", and "euGenes Report: A1BC". The main content area features a navigation menu on the left with links for "Daphnia (FleaBase)", "Sequences", "Limited access", "Genes", "Maps", "Anatomy & Images", "Stocks", "References", "People", "Help & Documents", and "News". The central part of the page displays a photograph of a Daphnia water flea with a green fluorescent signal. To the right of the image is the heading "Daphnia Genomic Information". Below the image, a text box states: "This web service provides gene and genomic information for species of Daphnia (water flea)." A search bar is located below the text, with a "Search" button, a dropdown menu set to "Everything", and a "Search Help?" link. Below the search bar, there is a checkbox for "wildcard '\*' to words" and a text input field.

## **Preliminary example**

<http://iubio.bio.indiana.edu/daphnia/>

Sample data include microsatellite DNA of J. Colbourne, GenBank Daphnia seqs, Medline abstracts

Blast searches, reports

Text data searches

## **Requirements for a genome db/ info 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, XML)

Architecture

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

## **Requirements for genome system, cont.**

Analysis software

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)

Publication interface

Genome database & information systems for Daphnia

Detail biological object views (sequences, genes, etc.)

Queries: simple-common, ad-hoc/general

Graphic viewers

Editing / data management interface

Interactive – document editing

Batch data updates

## **Components of web system**

Web server (Apache) and modules

FTP server for bulk data exchange

Core programs: relational database, BLAST and bioinformatics tools

Perl, Java middleware for data access & analysis, search and report

Limited, secure access for project data management

Public access for released data