

An integrated database for *Drosophila* and *Anopheles* genomics

Introduction

With the completion of increasing numbers of genomic sequences has come an explosion in the development of both computational and experimental high-throughput techniques for deciphering molecular functions and their interactions. However, while for certain objectives each of these techniques is useful in its own right, the strength of any inference can be increased when predicted by a combination of measures. Currently, biological data is stored in a wide variety of formats in numerous different places, making complex queries across these data sources very difficult, if not impossible. One of the current challenges for bioinformatics is therefore to develop resources for the integration and combined analysis of such data.

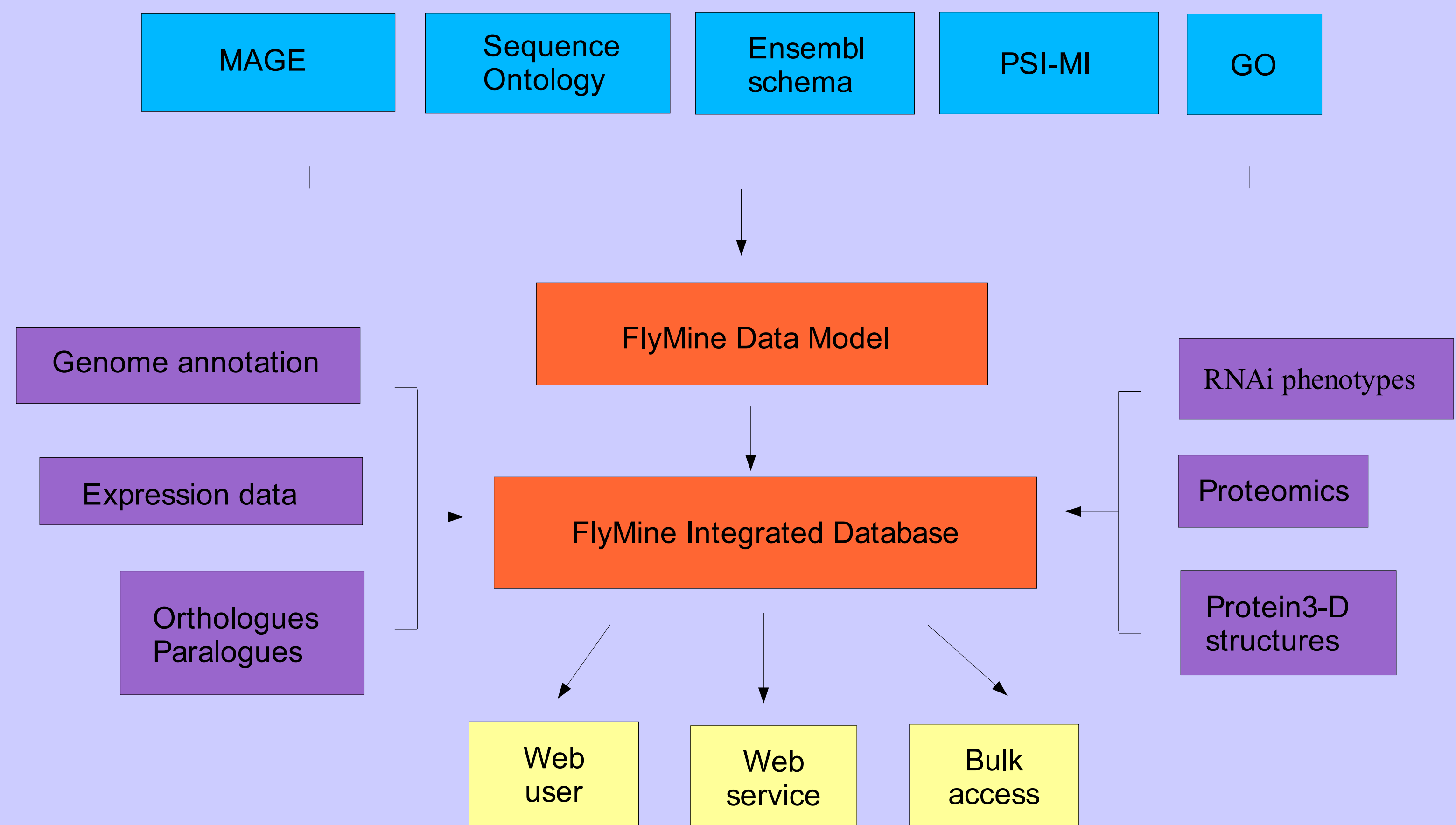
FlyMine is an open-source data warehouse system for integrating the data from many sources in to one object-based database, allowing complex queries to be performed on the integrated data. Initially, the focus is on genomic, proteomic and microarray data for *Drosophila* and *Anopheles*, however the system is designed to be completely generic and is not tied to any particular data model.

The Data Model

- The data model is generated from existing and emerging ontologies and standards (MAGE, SO, PSI etc).
- Enables data to be loaded from a variety of sources.
- Provides the flexibility to incorporate new data as they emerge.
- All model-specific parts of the system are automatically generated from a single UML diagram, or from DAG and FlyMine OWL ontologies.

Queries

- Ability to perform arbitrary and complex queries on the data.
- Multiple query Interfaces: Java, OQL, graphical etc.
- Query flexibility: not constrained to 'fill-in-the-blanks' query templates or predefined queries.



Example: select *Drosophila* genes which are annotated with GO:0007560 which have an *Anopheles* orthologue.

Project
About
Team
Funding
Presentations

Documentation
Software
Database
API Javadoc

Download software
Latest release
Nightly build

CVS
Browse
Anonymous

Mailing lists
Internal

Add class to query:

FlankingRegion
FoldbackElement
Fosmid
GOTerm
Gene
GeneGroup
GoldenPath
GoldenPathFragment
GroupIntron
GroupIntron

Select Browse...

Current query:

Gene **Gene_0** [Edit] [Remove]
organism **EQUALS** D. melanogaster
properties **CONTAINS** GOTerm_0

GOTerm **GOTerm_0** [Edit] [Remove]
termId = GO:0007560

Orthologue **Orthologue_0** [Edit] [Remove]
subject **CONTAINS** Gene_0
object **CONTAINS** Gene_1

Gene **Gene_1** [Edit] [Remove]
organism **EQUALS** A. gambiae

Reset query Run query

Graphical Query Interface

- Queries are constructed by selecting and combining classes.
- Constraints are applied to the attributes of the selected classes.
- The query is displayed as it is built up.
- Result display will include links to external databases, display and analysis tools and class and attribute browsing capability.
- Queries and results can be saved and fed back into new queries.
- Help will be available through query examples, on-line help pages and an analysis support help-desk.

FlyMine
An integrated database for *Drosophila* and *Anopheles* genomics

Project
About
Team
Funding
Presentations

Documentation
Software
Database
API Javadoc

Download software
Latest release
Nightly build

CVS
Browse
Anonymous

Mailing lists
Internal

Gene_0 [sort] [hide] [c]	Gene_1 [sort] [hide] [c]
<input type="checkbox"/> CG2671	<input type="checkbox"/> ENSANGG00000014570
<input type="checkbox"/> CG9885	<input type="checkbox"/> ENSANGG00000009878
<input type="checkbox"/> CG4889	<input type="checkbox"/> ENSANGG00000007666
<input type="checkbox"/> CG4316	<input type="checkbox"/> ENSANGG00000019529
<input type="checkbox"/> CG5462	<input type="checkbox"/> ENSANGG00000003691
<input type="checkbox"/> CG1725	<input type="checkbox"/> ENSANGG00000009998

Displaying 1 to 6 of 6 rows

Page size 10 Change

Save selections in new collection

Save query

Return to query

Initial Data Sources: first release (expected by mid May)

Data Type	Organism	Source	Web site
Genome annotation	<i>D. melanogaster</i>	Chado	http://flybase.net/annot/
2-hybrid protein interactions	<i>D. melanogaster</i>	intAct	www.ebi.ac.uk/intact/index.html
Gene expression data	<i>D. melanogaster</i>	ArrayExpress	www.ebi.ac.uk/arrayexpress/
Genome annotation	<i>A. gambiae</i>	Ensembl	www.ensembl.org
RNAi phenotypes	<i>C. elegans</i>	WormBase	www.wormbase.org
Orthologues	<i>D.melanogaster</i> <-> <i>A. gambiae</i>	Ensembl	www.ensembl.org
Orthologues and Paralogues	<i>D.melanogaster</i> <-> <i>C. elegans</i>	InParanoid	http://inparanoid.cgb.ki.se/index.html

Further Information and Download

Further information and documentation can be found on the FlyMine website (www.flymine.org), by joining one of the FlyMine electronic mailing lists (details on the website), or by email to info@flymine.org. The code is available for download under the open-source LGP Licence. We welcome co-developers.

The FlyMine team: Andrew Varley, Richard Smith, Matthew Wakeling, Mark Woodbridge, François Guillier, Rachel Lyne, Kim Rutherford and Gos Micklem.



Department of Genetics, University of Cambridge,
Downing Street, Cambridge CB2 3EH, UK
Tel: +44 1223 333377 Email: info@flymine.org



FlyMine is funded by the Wellcome Trust (grant no. 067205), awarded to M. Ashburner, G. Micklem, S. Russell, K. Lilley, and K. Mizuguchi.