FlyMine
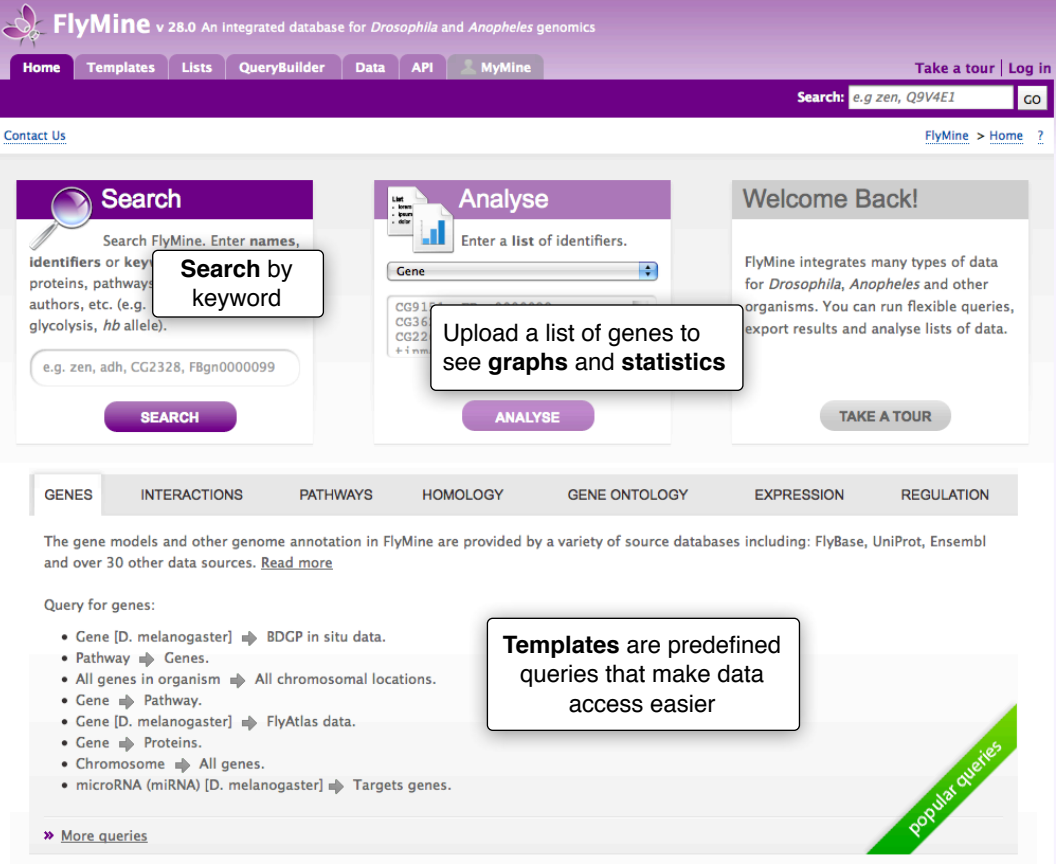
www.flymine.org

An open source integrated database of genomic, expression and protein data for *Drosophila*, *Anopheles* and other organisms.

How to use FlyMine?

As an integrated database FlyMine makes it possible to run **data mining** queries that span domains of biological knowledge. Inclusion of orthologue data allows **comparative analysis** between organisms.

Search and Browse
Enter **names**, **identifiers** or **keywords** for genes, proteins, pathways, ontology terms, publications, etc to search the entire FlyMine database and access report pages with integrated tools.
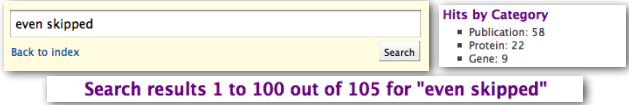
Run Template Queries
Dozens of useful queries have been predefined. These range from simple: *find homologues of a gene or list of genes in selected organisms* to more advanced data mining queries: *find all genes that have transposon insertions overlapping exons*. Search for available templates on the templates tab.

Query and Analyze Lists
Upload a list of gene identifiers to view graphical and statistical analysis including **expression graphs** and **enrichment** of Gene Ontology terms, protein domains, publications, etc. You can create lists from the results of queries, use saved lists in any query and merge, subtract and intersect multiple lists.

Use the FlyMine QueryBuilder
Advanced users can use a flexible query interface to construct their own data mining queries across the multiple integrated data sources, to modify existing template queries or to create new template queries.

Export
You can export results as tab or comma separated values to import into **Excel**. Genome features can be exported as GFF3 and FASTA. You can also send the results of any query to **Galaxy** for further analysis.

Personal Account
If you log in, you can save all your lists and queries permanently in **MyMine**. To create an account you just need to supply an email adress. All of your data are kept private.

FlyMine data sources include:

- Genome annotation for 12 *Drosophila* species-FlyBase
- *A. gambiae* genome annotation - Ensembl
- Orthologues & Paralogues between multiple organisms including 12 *Drosophila* species - InParanoid, FlyBase, TreeFam, KEGG
- Protein data for multiple organisms - UniProt
- Protein-protein interactions for *D. melanogaster*, *S. cerevisiae* and *C. elegans* - IntAct, BioGRID
- Genetic interactions for *D. melanogaster*, *S. cerevisiae* and *C. elegans* - BioGRID
- Protein Family and Domain data for *D. melanogaster*, *A. gambiae* and *C. elegans* - InterPro
- *D. melanogaster* and *A. gambiae* protein domain 3D structure predictions - Kenji Mizuguchi, PDB
- RNAi phenotypes for *C. elegans* genes - WormBase
- RNAi screens for *D. melanogaster* - DRSC
- GO annotations for *D. melanogaster*, *A. gambiae*, *C. elegans* and other organisms
- CRMs and TF binding sites for *D. melanogaster* - REDfly
- mRNA in situ hybridization data - FlyFISH & BDGP
- Microarray gene expression data, including FlyAtlas and Arbeitman et al (2002) *D. melanogaster* time course
- *D. melanogaster* P-element insertions - FlyBase, DrosDel
- *D. melanogaster* deletions - DrosDel
- Human disease to *Drosophila* gene data set - Homophila
- Microarray oligo set for *D. melanogaster* - INDAC
- *D. melanogaster* whole genome tiling amplimer set
- Gene versus publications - PubMed
- Pathway data for *D. melanogaster* - KEGG

Template Queries

Templates are useful pre-defined queries, they can all be run on a single identifier or a whole list at once. Some examples:

- Find all genes that have binding sites for a particular transcription factor
- Find which tissue(s) particular genes are up- or down-regulated in
- Find which *D. melanogaster* genes are involved in the Wnt signaling pathway and show the orthologues of these genes in *A. gambiae*?
- Find all interologs between *D. melanogaster* and *S. cerevisiae*
- When and where are particular genes expressed during embryogenesis?
- Find P-elements inserted in the exons of all the genes that are associated with a particular GO term
- Find the protein domains that are coded by my genes of interest.
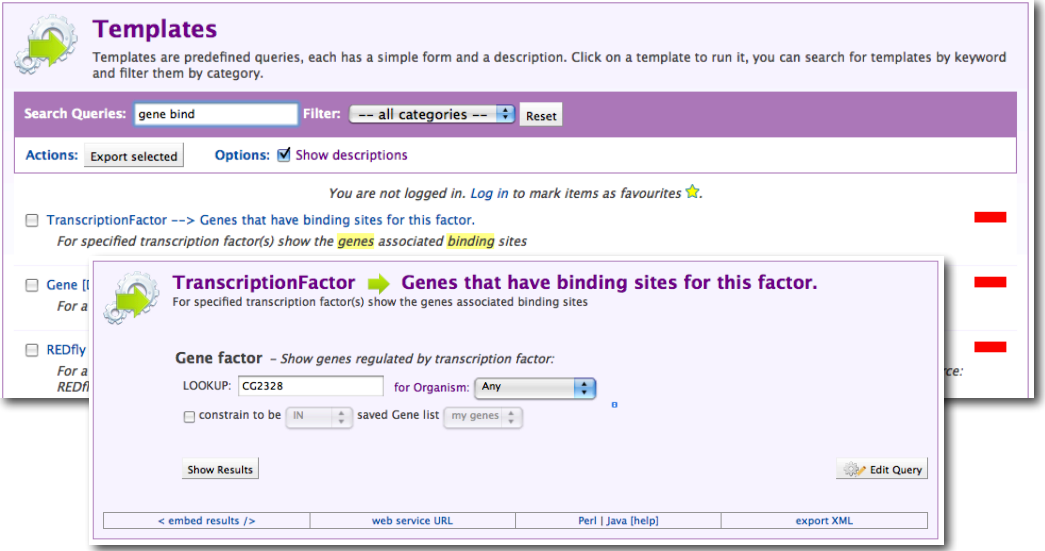

Lists

You can create lists of genes, proteins or any other type by uploading identifiers or from the results of a FlyMine query. Working with lists:

- Upload a mixed list of identifiers and convert it to a list of gene symbols.
- For any list of genome annotation features export their sequences or locations
- Find out which genes are unique between different lists
- Find all genes that particular species have in common
- Export a list of gene identifiers to FlyBase
- Convert a list to a different type, for example genes to mRNAs, exons or proteins

Widgets

For any list of genes a **list analysis page** displays graphs and summary statistics (widgets) to help interpret the biological properties of the list. All widgets are interactive - click to find the genes in a category for export or to view another list analysis page.

Results

When viewing results from a Template or a query:
- Use the column summary icon Σ to display statistical information about each column (number of time each term appears, averages).
- Use the menu at the top to create lists, export data, send results to Galaxy
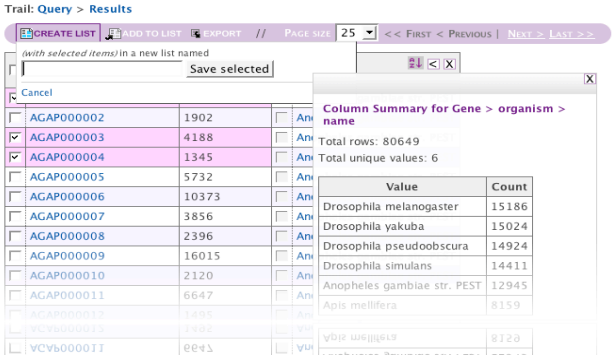
Web services

It's easy to run FlyMine queries directly from Perl or Java programs using the RESTful web service. From the web interface you can automatically generate code to run any Template or custom query, or you can create your own query using a simple API.

You can also embed results of FlyMine queries directly in other web pages using our new JavaScript library. Queries are executed at FlyMine and the results appear in the page you create.