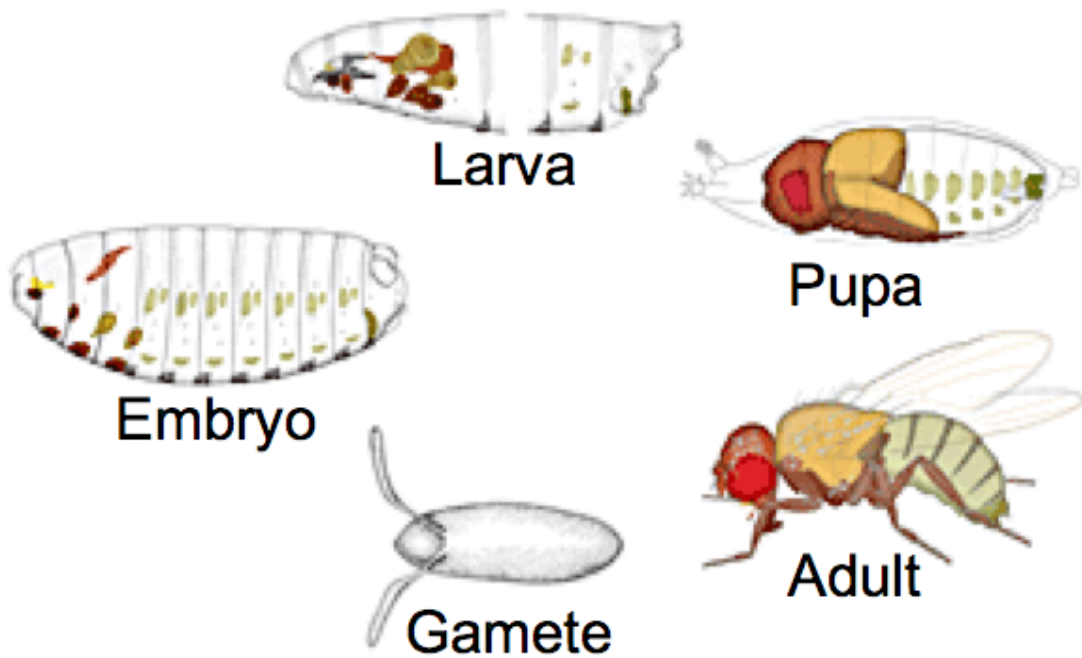


# Section I

## FlyBase Workshop



FlyBase (Cambridge)



What can I do with FlyBase? .....	4
What can I get from FlyBase? .....	5
Navigating FlyBase .....	6-10
QuickSearch .....	11
Linking of reports in FlyBase .....	12
Help menu & documentation .....	13
Search Tools introduction .....	14
Genomic Search Tools .....	15-16
FlyBase BLAST .....	17
GBrowse on FlyBase .....	18
Searching 12 species and aberrations .....	19-20
Search Strategies: Expression data, stocks .....	21
Search Strategies: Mutant phenotype, references .....	22
Ontologies in FlyBase .....	23
TermLink & ImageBrowse .....	24
Term Report & Image Report .....	25
Query Results Analysis Tools .....	26
Comparison tables of search tools .....	27
QueryBuilder tutorial .....	28-36
Examples of QueryBuilder searches .....	37-38

## What can I *do* with FlyBase:

- **QuickSearch/Jump to Gene:**
  - Enter any identifier or symbol for 12 different data classes (e.g. genes, alleles) to access different report pages and data
- **BLAST:**
  - The integrated FlyBase BLAST server allows searching of 21 different genomes from Drosophilid and insect species using standard BLAST programs
  - These BLAST hits can then be traced onto Drosophilid genomes in GBrowse, our integrated genome viewer
  - Access sequences of hit sequences, link to FlyBase and external websites for BLAST hits
- **Search Genomes:**
  - **GBrowse** Genome Browser - allows viewing of 12 Drosophila genomes
    - Fully customizable
    - Can add own/3rd party annotations
  - Access sequences
  - Compare different Drosophila genomes using Orthoview
  - Convert co-ordinates between different genome releases
  - Find insertions, high-throughput data, affymetrix probes on the genome
- Find **curated** information from papers:
  - Browse between reports for different types of object, reference to gene, to allele, to construct, to anatomy term to stocks, to whatever you like!
- Find **stocks:**
  - Stock report pages, linked to relevant papers, genes, alleles etc., with links to external stock centers where possible
- **Search all our data:**
  - Perform combinatorial searches on the database using **QueryBuilder** to access particular parts of the database, or particular genes, alleles, transcripts, references
  - Search using controlled terms for expression or phenotype in specific body parts using **TermLink**
  - Search for specific molecular functions, e.g. kinase activity.

## What can I *get* from FlyBase:

- FlyBase includes information about the structure and function of genes and gene products of the *Drosophila* genome
- Contains information on over **400 *Drosophila* species**, and houses genomic information for 12 *Drosophila* species
- **Phenotypic** and **genetic interaction** information about mutants, and wild-type gene and enhancer-trap **expression patterns** are linked to strains in the *Drosophila* stock centers
- Mutant phenotypes and gene expression patterns are described using **controlled vocabularies**, including anatomical terms linked to illustrations in the Anatomy section of FlyBase, allowing complex searching
- Chromosome aberrations, natural transposons, genetically engineered constructs and transgene insertions are presented with hyperlinks to affected genes and resulting mutant alleles
- Batch download of large datasets such as all fasta sequences for a particular genome, all reference data, the entire database in chadoXML format
- Clone information
- References
- Links to other *Drosophila* and related resources
- Curated 'value-added' integration of information from a variety of sources.



## Navigating FlyBase

Home Tools Files Species Documents Resources News Help Archives Jump to Gene Go

The top of each page on FlyBase comprises the FlyBase banner and navigation bar, which has been used as a header in this document.

The text in the grey box serves as a title for each page and therefore changes as one navigates from page to page, but the rest of the banner remains consistent.

Users can return to the homepage from any page on FlyBase either by clicking on 'Home' or by clicking on the fly image, as shown by the arrows.



- ID Converter
- ImageBrowse
- Interactions Browser
- QueryBuilder
- QuickSearch
- TermLink

- Find A Person
- Add A New Person
- Update An Address

- Tools** Files Species
- Tools Overview
- General Tools (A-G) ▶
- General Tools (H-Z) ▶
- Data Submission ▶
- People ▶

- Aberration Maps
- Batch Download
- BLAST
- Chromosome Maps
- Coordinate Converter
- CytoSearch
- GBrowse
- Google™ FlyBase
- Add A Paper

The 'Tools' menu on the navigation bar provides quick access to all of the search tools available on FlyBase.

At the top of the tools menu there is a link to documentation providing an overview of the search tools and providing examples of some search strategies.

Please see each individual tool section for more information on the specific usage of each tool.

Files	Species	Doc
<b>Downloads:</b>		
Precomputed files		
Archived data		
Releases (FTP)		
Genomes (FTP)		
AAA data (FTP)		
<b>Lists:</b>		
Transposons (Dmel)		
Transposons (other)		
Vectors & Constructs		
Anatomical Terms		
Map Conversion		

The links available under the 'Files' menu are split into two sections: Downloads and Lists, reflecting the diversity of information that can be reached from this part of the navigation bar.

### Downloads:

The 'Precomputed files' link goes to a page from which many types of FlyBase data can be downloaded. Examples include data on stocks, genes, controlled vocabulary terms, genome annotation and sequences.

'Archived data' links to a page that contains downloadable data from previous FlyBase releases.

'Releases (FTP)' and 'Genomes (FTP)' link to the FlyBase FTP site, where various archives of the database or previous releases of the genome, are available for download.

'AAA data (FTP)' links to the fasta and GFF files for each genome sequenced by the AAA consortium.

### Lists:

'Transposons (Dmel)' links to a list of natural *Drosophila melanogaster* transposable elements and genes which provide further links to transposon reports. 'Transposons (other)' is a similar table for all other species. 'Vectors and Constructs' links to a list that is subdivided according to experimental use.

'Anatomical terms' links to a glossary of controlled vocabulary terms relating to anatomy.

The 'Map conversion' link goes to a page that contains a table showing the estimated values used by FlyBase to convert between cytogenetic map positions, genetic map positions and sequence coordinates.

Species	Document
Phylogeny	
Synteny Table	
Drosophilidae	
Abbreviations	

The '**Species**' menu provides links to pages about Drosophilidae species.

'Phylogeny' links to a page with an alphabetical index of Drosophilidae.

'Synteny table' goes to a table that shows the syntenic relationships between the 12 *Drosophila* species whose genomes have been sequenced.

'Drosophilidae' links to images of the different species.

'Abbreviations' goes to a table showing all 4-letter species abbreviations used in FlyBase.

Documents	Resou
About FlyBase	
<b>Reference Manual:</b>	
Sections	
Contents	
Nomenclature	
Release Notes	

The '**Documents**' menu contains links to documentation on FlyBase.

The 'Reference Manual' is a comprehensive document on various aspects of FlyBase.

'Nomenclature' includes the rules used by FlyBase on naming genetic objects.

'Release notes' contains information on the most recent FlyBase data update, which includes both genomic annotation and literature curation.



Resources	News
All Resources	
Model Organisms	
Stock Collections	
<b>Other links:</b>	
BDGP	
DGRC	
DHGP	
FlyExpress	
Interactive Fly	
modENCODE	
Textpresso for Fly	

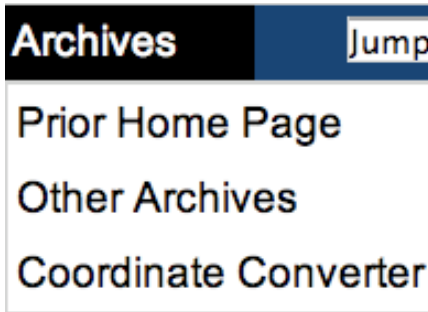
The '**Resources**' menu contains links to stock centers and other *Drosophila*-related project sites.

The exception is 'Model Organisms' which provides links to the sites of other model organisms, such as yeast and mouse.

For a comprehensive list of links, click on 'All resources'.

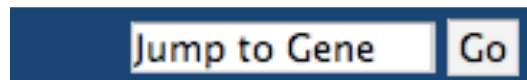
The '**News**' menu contains links to current developments in the *Drosophila* world, jobs available at FlyBase, funding contacts for researchers and a searchable archive of bionet.dros postings.

News	Help	Archive
News		
FlyBase Forum		
FlyBase Positions		
Meetings		
Courses		
Fly Board		
White papers		
Funding		
bionet.dros		



The '**Archives**' menu provides links to previous FlyBase releases.

'Prior Home Page' will take users back to the previous release of FlyBase (fb2008\_08).



The 'Jump to Gene' (J2G) box allows users to move directly to specific Gene Report pages from any FlyBase page. This navigation aid accepts valid gene symbols or synonyms, full gene names, or FBids.

J2G searches *D.melanogaster* genes by default. If you would like to search for a non-Dmel gene, you need to use the unique, 4-letter species abbreviation, followed by a backslash, and then the gene symbol (e.g. *Dpse\dpp*). This is the same for all searches on FlyBase, not just J2G.

Example identifiers:

Gene symbol : amn

Annotation symbol (i.e. the gene model) : CG11937

FlyBase identifier : FBgn0000076

Allele symbol : amn [ 1 ]

FlyBase identifier : FBa10000500

**QuickSearch**

Species:  Dmel only  All species [Find A Fly Person](#)

Search:  ID/Symbol/Name  All text [QuickSearch help](#)

Data Class:

Enter text:

Note: Wild cards (\*) can be added to your search term

The QuickSearch tool on the home page provides access to the FlyBase report pages. Searching can be performed in either *D.melanogaster* only, or in "All species". Data other than genes can be queried by selecting one of the options from the "Data Class" drop-down menu.

A gene search using the ID/Symbol/Name option (the default selection) will start by looking for an FBgn or CG number, and then if nothing is found the search engine checks for a gene symbol, synonym, or gene name or gene name synonym. To search for greek symbols, type out the full word, e.g. alpha for  $\alpha$ .

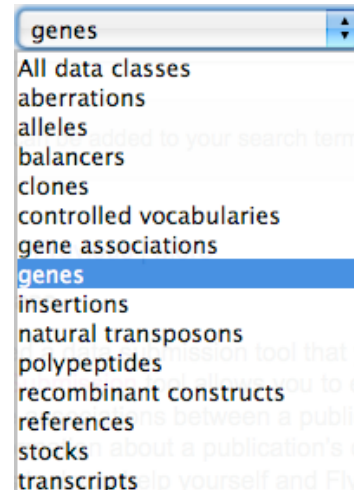
If anything in these fields matches the search string, then the gene concerned is reported. If more than one gene is found, then a 'hit list' is reported. An error only occurs if nothing matches the input string. QuickSearch supports the AND, OR, or NOT boolean operators in the "All text" search of any data class and the

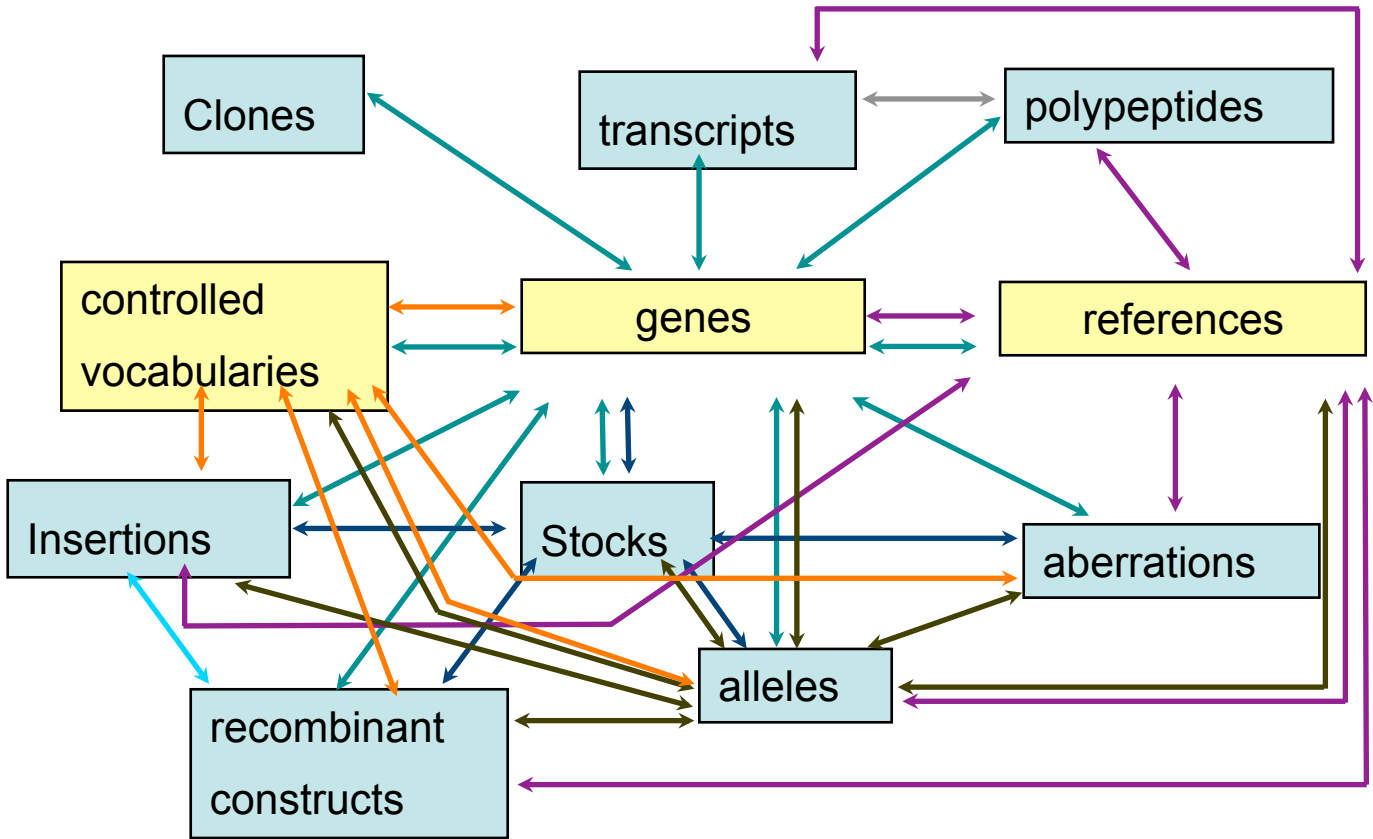
"Author(s)" and "Year(s)" fields of a "references" data class search. These operators are not supported for any ID/Symbol/Name search and the "gene association" or "controlled vocabularies" data class options.

To search for GAL4 (a *Saccharomyces cerevisiae* protein, for which our species abbreviation is "Scer") or GFP (an *Aequorea victoria* protein - "Avic") lines, select 'All species'. If one knows that the GFP is only used as a tag, fused to coding region of a Dmel gene, then this is classified as a Dmel gene.

The "controlled vocabularies" data class option allows you to search for a term in the controlled vocabularies (CVs) used by FlyBase. When this option is selected you can choose to search All CVs or select a particular CV from the drop down button.

The "gene associations" option allows you to enter a single gene identifier (ID, symbol, synonym, annotation symbol, name, etc...) and see a list of FlyBase objects that are directly related to that gene. From there you can click the "HitList" and "Refine" buttons to further evaluate the items returned.





Many of the reports in FlyBase are linked. Where this occurs, it is possible to go directly from one report to another. For example, genes are expressed as transcripts and it is possible to click on transcript names within gene reports to be taken to the transcript report of interest. The above diagram demonstrates some of the links between reports.

Help	Archives
Google™ FlyBase	
Site Map	
FAQs	
Report help	
New to Flies	
Pers. comm.	
Author guidelines	
Known Problems	
Contact FlyBase	

## Help Documentation - under help menu

The FlyBase Site Map gives a comprehensive listing of the searches and resources available on FlyBase

Author Guidelines - provides advice on how to make your publications FlyBase-friendly

Known Problems

Contact FlyBase

Tool-specific help is available on each tool page

Report help: Field-by-field guide to the reports:



Beta Release FB2006\_00.2

## FlyBase

Report A Bug

Home Tools Files Species Documents Resources News Help Archives

---

Last Updated: 6 November 2006

The **Aberration Report** provides information on individual chromosomal aberrations. Aberrations are chromosomal rearrangements that differ from the wild-type arrangement of chromosomes.

This is a field-by-field guide to the information that is provided in the **Aberration Report**.

FlyBase attributes data to the publication that reported it, so that users can easily refer back to the original publication if they wish. Thus, where possible in the fields below, the publication(s) that are the source of the information are listed, typically in parentheses to the right of the data. The exception in the **Aberration Report** is the **General Information** section which contains a summary of the identity of the aberration and a summary of cytological data.

GENERAL INFORMATION	
<b>Symbol</b>	The <b>valid symbol</b> that is used in FlyBase for the aberration.  The first part of the symbol (before the '1') is the standard prefix for the species (from the <a href="#">Species Abbreviations</a> list). For species other than <i>D.melanogaster</i> , the species prefix is displayed wherever the aberration symbol is used throughout FlyBase. For <i>D.melanogaster</i> aberrations, the species prefix is only displayed in the <b>GENERAL INFORMATION</b> section at the top of a Report.  See <a href="#">General principles for naming aberrations</a> for a detailed description of how aberration symbols are constructed.
<b>Name</b>	The valid full name that is used in FlyBase for the aberration.
<b>Feature type</b>	A list of <a href="#">controlled vocabulary</a> terms from the <a href="#">Sequence Ontology</a> that describe the nature of the chromosomal aberration relative to the wild-type chromosomal order.
<b>Species</b>	The organism that the aberration originates from, with the initial letter of the genus and the full species name listed.
<b>FlyBase ID</b>	The <b>Primary FlyBase identifier number</b> of the aberration, used to uniquely identify the aberration in the database.  An aberration may also have any number of <b>Secondary FlyBase identifier numbers</b> , which are listed in the <b>SECONDARY FLYBASE IDs</b> section of the <b>Aberration Report</b> .
<b>Created/Updated</b>	The date that the aberration was entered into the chado database underlying FlyBase (we migrated to the chado database on 2006-08-22) and the date that the aberration report was last updated. The dates are in the format: year-mm-dd.

## Search tools in FlyBase:

There are many portals into FlyBase data and tools to analyze those data. FlyBase contains a wealth of data (see our release notes), and knowing how to search this data is the key to using FlyBase to its full potential. For an overview of the tools available, take a look at our new tools overview page, under the tools menu in the navbar.

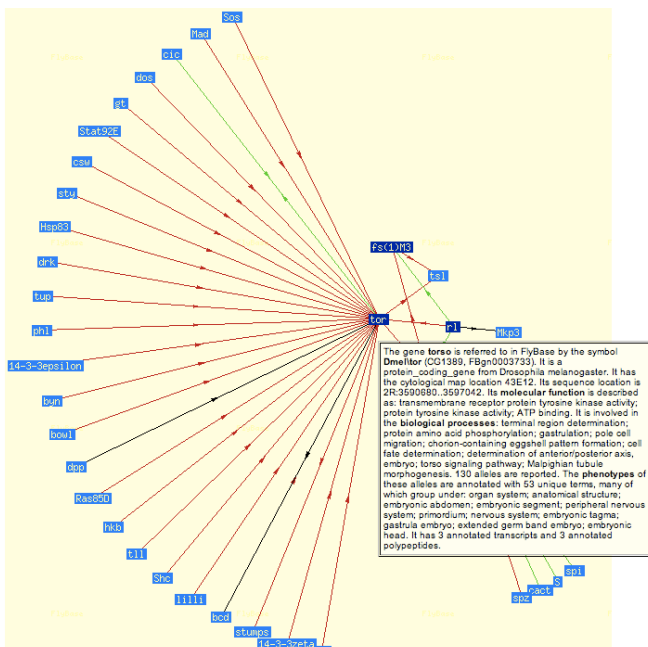
Our search tools can be grouped into 3 broad groups.

We have five genomic search tools/browsers - BLAST, GBrowse, CytoSearch, and our new tools, Chromosome Maps and Coordinates Converter.

Our main query tools, for searching phenotype, expression, genes, alleles, insertions, stocks, references etc., are QuickSearch, QueryBuilder and TermLink (for searching controlled terms).

Our other tools include Interactions Browser (providing a graphical representation of the genetic interactions displayed in the allele reports), ImageBrowse, Find a Person, Update an Address, and Google FlyBase. In addition to these tools, Jump to Gene is available on every page on the right hand side of the navigation bar.

Tools	Files	Spec
<b>Tools Overview</b>		
<b>General Tools:</b>		
Batch Download		
BLAST		
Chromosome Maps		
Coordinate Converter		
CytoSearch		
GBrowse		
Google™ FlyBase		
ImageBrowse		
Interactions Browser		
QueryBuilder		
QuickSearch		
TermLink		
<b>People:</b>		
Find A Person		
Add A New Person		
Update An Address		



Our tools overview page also provides examples of search strategies, illustrating how and when to use each tool.

Through a combination of our tools overview, the Site Map, and the tool and report-specific help documentation, you should be able to use our search tools to find what YOU are looking for on FlyBase. Our hit list refinement and batch download tools, available for every hit list generated from our tools, should enable you to manipulate the search output to provide the results format you were looking for.

We have five genomic search tools/browsers - BLAST, GBrowse, CytoSearch, and our new tools, Chromosome Maps and Coordinates Converter. BLAST and GBrowse are covered in the following pages, here, we quickly outline some of the other tools and strategies available in FlyBase. For more details, please see our Tools Overview section, in the Tools menu.

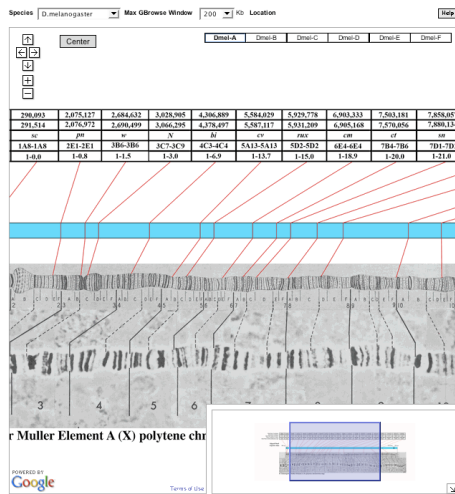
Over the next few pages, we will outline some of the uses of each of these tools. The two made genomic tools are GBrowse and CytoSearch. GBrowse is especially useful when looking for molecularly mapped sequences, insertions, or Affymetrix probes. CytoSearch comes into its own when searching for cytologically defined features, such as cytologically-mapped genes or deficiencies, that haven't been molecularly mapped to the sequence.

**CytoSearch** lists are regional maps of the *Drosophila melanogaster* genome incorporating both sequence-based and cytology-based map data. Sequence-based data trumps cytology when both are available, cytology trumps meiotic data when both are available, and estimated cytology is used when only meiotic data are available. The FlyBase correspondence tables for cytological and sequence level maps are used to estimate cytology from sequence range and sequence range from cytology, for both the underlying data and the query input.

## Finding orthologs using GBrowse

By adding 'Similarity' tracks to the *D.melanogaster* genome view you can use the resulting ortholog links to navigate to orthologs in the other species. You can also find an ortholog by selecting the species from the 'Data Source' menu and entering the *D.melanogaster* gene symbol or FBgn ID in the 'Landmark or Region' box. For genomes other than *D.melanogaster* GBrowse is configured by default to show two windows that indicate how the region displayed is related to *D.melanogaster*. The top window of this 'Orthoview' displays a representation of the genome of the selected species showing the predicted gene. If this window contains a putative ortholog, a second window will appear with the *D.melanogaster* genome aligned to the ortholog closest to the center of the upper window. The relationship between the genomes is shown by sets of green lines that connect the orthologs in the region displayed. Furthermore, we have added pop-ups to GBrowse to provide a gene summary when you mouse-over a gene.

Tools	Files	Spec
Tools Overview		
<b>General Tools:</b>		
		Batch Download
→		BLAST
→		Chromosome Maps
→		Coordinate Converter
→		CytoSearch
→		GBrowse
		Google™ FlyBase
		ImageBrowse
		Interactions Browser
		QueryBuilder
		QuickSearch
		TermLink
<b>People:</b>		
		Find A Person
		Add A New Person
		Update An Address



The **chromosome maps** show sequence scaffolds aligned to polytene chromosome maps for the Muller elements of the sequenced *Drosophila* species. For more information on the syntenic relationships among the 12 sequenced genomes, their standard chromosomal numbering and corresponding Muller element please see the Muller Element Arm Synteny Table.

- Tools
- Files
- Spec
- Tools Overview
- General Tools:**
  - Batch Download
  - BLAST
  - Chromosome Maps
  - Coordinate Converter
  - CytoSearch
  - GBrowse
  - Google™ FlyBase
  - ImageBrowse
  - Interactions Browser
  - QueryBuilder
  - QuickSearch
  - TermLink
- People:**
  - Find A Person
  - Add A New Person
  - Update An Address

The aligned sequence scaffolds, shown in blue on the maps, provide access to the sequence data and gene models. When you move your cursor over one of the blue scaffolds a yellow box appears that corresponds to a GBrowse window, and clicking on the box will take you to the corresponding location in GBrowse.

The **Coordinates Converter** tool is simple, yet incredibly useful and allows you to convert genomic coordinates between different genome releases. Just select the input and output assembly, enter your list of coordinates (or load them from a file), and away you go! It's that simple. Coordinates converter is also found under the Archives menu.

D.melanogaster Sequence Coordinates Converter

Input Assembly:   
  Output Assembly:   
  Send results to:

Enter D.melanogaster Coordinates:  or Upload File of Coordinates:

Examples: 3L:18386078..18396077 or X:2684632





# Genomic Search Tools: FlyBase BLAST

**Tools** Files Species

Tools Overview

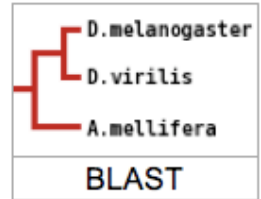
**General Tools:**

- Batch Download
- BLAST**
- Chromosome Maps
- Coordinate Converter
- CytoSearch
- GBrowse
- Google™ FlyBase
- ImageBrowse
- Interactions Browser
- QueryBuilder
- QuickSearch
- TermLink

**People:**

- Find A Person
- Add A New Person
- Update An Address

**FlyBase BLAST** gives the user the opportunity to BLAST against the 12 completed *Drosophila* genomes, along with related insect species. BLAST provides access to the FASTA sequences of all sequenced *Drosophila* sequences, as well as providing links to GenBank. In addition, you can BLAST an unknown sequence and identify its position on GBrowse.



The BLAST homepage is split into three sections; the first allows the user input the query sequence and set-up the standard BLAST parameters (e.g. Expectation value, database to be searched); the second section allows the species to be selected; while the third allows the user to specify advanced BLAST options.

To perform a BLAST, copy and paste your sequence into the BLAST window and start the query. From section 2, pick as many species as you wish. When your query is returned you will see the standard list of hits. Pick one of the top hits and go to its specific record/alignment.

You will note that there is an html link on the Score line that reads "BLAST HIT on Genome Map". Click on this link and you will be taken to the position on the appropriate species genome wherein the hit resides in the GBrowse genome viewer tool. The position of the hit will be highlighted in grey and this highlight will extend over as many of the evidence tracks as are open. The best hits should take you to the ortholog of your gene if there is one.

**BLAST**

Database: Genome Assembly (NT)

Program: blastn: NT -> NT

Sequence file: [Browse...]

Sequence: [Text area]

[BLAST] [Clear sequence]

**Species (optional)**

Quick Tip: Click a node in the tree below to select all species under that node.

- Drosophila melanogaster<sup>1,2,3,4</sup>
- Drosophila simulans<sup>5,7</sup>
- Drosophila sechellia<sup>5,7</sup>
- Drosophila yakuba<sup>5,7</sup>
- Drosophila erecta<sup>5,7</sup>
- Drosophila ananassae<sup>6,7</sup>
- Drosophila pseudoobscura<sup>5,6,7</sup>
- Drosophila persimilis<sup>5,7</sup>
- Drosophila willistonii<sup>5,7</sup>
- Drosophila mojavensis<sup>5,7</sup>
- Drosophila virilis<sup>5,7</sup>
- Drosophila grimshawi<sup>6,7</sup>
- Culex pipiens (mosquito)
- Aedes aegypti (mosquito)<sup>13</sup>
- Anopheles gambiae (mosquito)<sup>11,12</sup>
- Bombyx mori (silkworm)<sup>8</sup>
- Bombyx mori (silkworm)<sup>10</sup>
- Tribolium castaneum (red flour beetle)<sup>11</sup>
- Apis mellifera (honey bee)<sup>9</sup>
- Nasonia vitripennis (wasp)
- P. gulosus humanus corporis (human body louse)

**Advanced Settings (optional)**

Expect value: 1

Output format: Graphical Alignment View

MegaBLAST (blastn only):

Use legacy BLAST engine:

Number of descriptions: 25

Number of alignments: 25

Low complexity filter:

Word size: [ ]

Matrix: BLOSUM62

Genetic Codes: Standard (1)

Other options: (e.g. -G 7 -E 3)

Search

Advanced Search: Cytolocalization

Landmark or Region: [ ]

Data Source: D. melanogaster (R5.12)

Overview of 3L

Gene Span: 033220

HSP13: 0332242-0h, 0332242-8h, 0332244-8B, 0332244-8C, 0332244-9B, 0332244-9C

Orthologs: [ ]

TRN: [ ]

Non coding RNA: [ ]

Natural transposon: [ ]

Transgene insertion site: [ ]

cDNA and aligned genomic sequences: [ ]

BLAST: HSP13

3



# Genomic Search Tools: GBrowse on FlyBase

Home Tools Files Species Documents Resources News Help Archives Jump to Gene Go

**Tools** Files Species

Tools Overview

**General Tools:**

- Batch Download
- BLAST
- Chromosome Maps
- Coordinate Converter
- CytoSearch

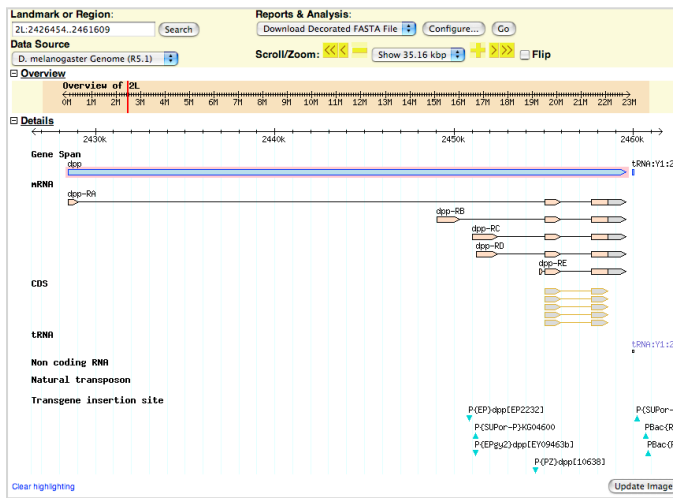
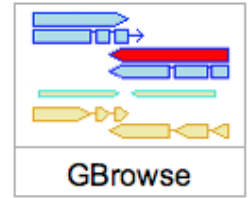
**GBrowse**

- Google™ FlyBase
- ImageBrowse
- Interactions Browser
- QueryBuilder
- QuickSearch
- TermLink

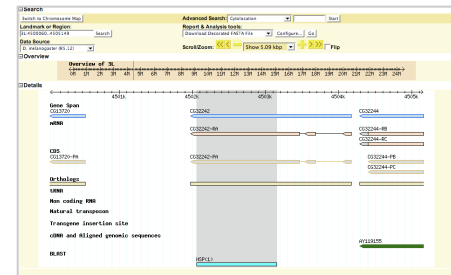
**People:**

- Find A Person
- Add A New Person
- Update An Address

**FlyBase GBrowse** provides a graphical or tabular representation of the 12 sequenced *Drosophila* genomes. Genes, insertions, and deficiencies, along with other mapped objects are illustrated along with orthologous regions in other *Drosophila* genomes, and affymetrix probes. There is a separate help manual for GBrowse that can be accessed from the GBrowse pages, along with information about the different evidence tiers available.



By default FlyBase present a view of *D.melanogaster* that displays gene models, transcript and polypeptide data, natural transposon insertion sites, and cDNAs. These and many additional tracks are easily configured to create a customised view of the data. You can navigate to a specific location by entering a precise sequence range, or any valid FlyBase identifier for a gene, gene product, or insertion in the 'Landmark or Region' box. 'Advanced Search' enables you to move to a particular cytological location. Additionally, FlyBase BLAST output includes GBrowse links that display each BLAST alignment as a highlighted feature in the context of neighbouring gene models and other features of the region. This is an extremely useful entry path into the sequence data of species other than *D.melanogaster*, which in some cases is comprised of a large number of relatively short unlinked scaffolds.



Individual gene reports for genes from all 12 sequenced *Drosophila* genomes are now available in FlyBase. There are four main ways in which this data can be browsed and queried in FlyBase:

- \* Precomputed files
- \* BLAST
- \* GBrowse
- \* Gene Report Pages

For those interested in genome-wide analyses, bioinformatics and comparative genomics, there are a selection of pre-computed files available for download from our precomputed files page (in the Genomes:Annotation and Sequence section, for example), found in the 'Files' menu.

For those with an interest in a specific gene/protein/region across the different species, there are a number of ways to query the data. Our BLAST server allows querying of all 12 genomes, either individually, as a subset, or all together. Each BLAST hit can then be localised and shown on the genome through GBrowse. Orthoview in GBrowse allows movement through the different genomes, illustrating the same region (where possible) in different genomes. In the near future, multiple alignments will be available, enabling direct sequence comparison between the different genomes.



# Genomic Search Tools: Searching 12 species and aberrations

**Select species to search against**

Clicking a node in the tree below selects all species under that node. More information about the CAF1 assemblies can be obtained from the [AAA](#) site.

Species

- Drosophila melanogaster* [A,G,P,I,T,D,U] - v5.1
- Drosophila simulans* [A,D,GR] - CAF1
- Drosophila sechellia* [A,D,GR] - CAF1
- Drosophila yakuba* [A,D,GR] - CAF1
- Drosophila erecta* [A,D,GR] - CAF1
- Drosophila ananassae* [A,D,GR] - CAF1
- Drosophila pseudoobscura* [A,G,P,I,B,T,D,GR] - Z1
- Drosophila persimilis* [A,D,GR] - CAF1
- Drosophila willistoni* [A,D,GR] - CAF1
- Drosophila mojavensis* [A,D,GR] - CAF1
- Drosophila vitta* [A,D,GR] - CAF1
- Drosophila grimshawi* [A,D,GR] - CAF1
- Anopheles gambiae* (mosquito) [A,D]
- Aedes aegypti* (mosquito) [A,D]
- Bombyx mori* (silkworm)<sup>1</sup> [A,D]
- Drosophila obscura* [A,D,GR] - CAF1

Sequences producing significant alignments:

gi	Accession	Species	Score	E Value
gi110313979	gb CH954177 CH954177	<i>Drosophila erecta</i> strain ...	2054	0.0
gi199109335	gb CM000157 CM000157	<i>Drosophila yakuba</i> strain T...	1592	0.0
gi110313975	gb CH954181 CH954181	<i>Drosophila erecta</i> strain ...	44	0.049
gi199109333	gb CM000159 CM000159	<i>Drosophila yakuba</i> strain T...	44	0.049
gi110313978	gb CH954178 CH954178	<i>Drosophila erecta</i> strain ...	42	0.19
gi110313977	gb CH954179 CH954179	<i>Drosophila erecta</i> strain ...	42	0.19
gi110313976	gb CH954180 CH954180	<i>Drosophila erecta</i> strain ...	42	0.19
gi199109334	gb CM000158 CM000158	<i>Drosophila yakuba</i> strain T...	42	0.19
gi199109332	gb CM000160 CM000160	<i>Drosophila yakuba</i> strain T...	42	0.19
gi199095574	gb CH93835 CH93835	<i>Drosophila yakuba</i> strain T...	40	0.76
gi110313974	gb CH954182 CH954182	<i>Drosophila erecta</i> strain ...	38	3.0

The 12 *Drosophila* genomes, along with other insect genomes, can be searched using BLAST

**Instructions**  
Search using a sequence name or range, gene symbol (but not full or other landmark. Searches are case-sensitive. Wildcards are not supported to change magnification and position.

**Examples:** cnn, FBgn0000490, scaffold\_4820:290000..330000, scaf scaffold\_4845:80000..100000.

[Hide banner] [Bookmark this] [Link to Image] [Help] [Reset]

**Search**  
Landmark or Region:  Search

Data Source:  Search

**Overview** Overview of scaffold\_4929

**Details**

Putative ortholog *Dere1398*

GLENNR / high confidence set *dere\_GLENNR\_10390*

GLENNR / medium confidence set

GLENNR / low confidence set

GLENNR / non-orthologous

Score = 2054 bits (1036), Expect = 0.0 **BLAST HIT on Genome Map**  
Identities = 1306/1382 (94%), Gaps = 18/1306 (1%)  
Strand = Plus / Plus

Each *Drosophilid* hit on the BLAST list has a link to GBrowse, allowing you to see the BLAST hit on the respective genome

Predicted Dere gene

GBrowse can be used to navigate between the homologous regions of the different genomes, allowing sequence and genomic comparisons to be made

**Gene Span**

nRNA

CDS

Ortholog (FlyBase)

Ortholog (GLENNR)

Clear highlighting [Help for FlyBase evidence tiers]

**Tracks**

Gene Model features  All on  All off

CDS  Gene Span  mRNA

General  All on  All off

3-frame translation (forward)  Blast hit

3-frame translation (reverse)  DNA/GC Content

## Browsing Expression Data

Information about transcript and protein expression patterns can be found on **gene reports** (e.g. the *cnn* gene), data for reporter constructs can be found on recombinant construct reports (e.g. P{dpp-lacZ.B}), and data for enhancer traps can be found on insertion reports (e.g. P{GawB}elav[C155]). In all cases, expression data will be found in the "Expression Data" section of the report. Please note that, in the case of transcript and protein expression, text descriptions of expression patterns are available on the **Gene Expression Report** (e.g. *cnn*), which is linked to at the top of the Expression Data section.

We also cooperate with several other databases of expression data and either display a portion of their data within FlyBase (e.g. FlyExpress) and/or link to their database (e.g. FlyAtlas). These types of data can be found in the "External Data & Images" subsection of the "Expression Data" section.

High throughput expression data can also be viewed using the GBrowse tool. By selecting the tracks found under the "High Throughput data (Arrays)" section, the available data will be shown on GBrowse.

## Searching for Expression Patterns

Expression data can be searched most easily and accurately by using **QueryBuilder** or **TermLink**, depending on how you'd like to search. You'll want to use QueryBuilder if you're interested in a multipart query (e.g. generate a list of genes which have the GO term "transcription factor activity" and whose protein products are expressed in the central nervous system). However, if you're interested in all genes expressed in a bodypart, tissue, or developmental stage, you can find that using TermLink.

## Stocks

One of the easiest ways to search for a stock in FlyBase is to use **QuickSearch**. Simply change the data class to 'stocks', type in the feature of interest (e.g. a gene symbol, allele symbol), and search. A further way to identify stocks is through the hit list produced after a search. At the top of the hit list there is a toggle allowing you to 'Show related' stocks. Stocks can also be found for individual alleles by clicking on the Stocks section on the allele report page. Stocks can also be searched using **QueryBuilder**.

## Mutant phenotype data

Mutant phenotype data is associated with alleles in FlyBase, so you need to search allele data if you are interested in mutant phenotype. In addition to free text describing the phenotype, the alleles are indexed with controlled vocabulary (CV) terms, which makes it easier for you to search for a particular phenotype, e.g. searching for mutants that affect the wing. You can search with these CV terms using either **TermLink** or **QueryBuilder**.

You can find mutant alleles affecting the wing from all species using TermLink. If you enter the term "wing" into TermLink search page and then click on the "Alleles" button in the report page, you will obtain a list of mutant alleles that affect the wing. However, to search in a specific species, or to search for mutant phenotypes as part of a multipart query, QueryBuilder must be used. In this case, you should pick the "CV Hierarchy (GO/etc.)" dataset and then use the term picker to choose the body part, e.g. wing. In both cases, the default is to search both for alleles specifically labelled with the CV term, e.g. wing and also with child CV terms that are a subset of the term chosen, e.g. wing vein. If you want to restrict your search to just the precise term chosen, use QueryBuilder and select 'Retrieve records annotated with "This CV term only"' before you run the query.

## References

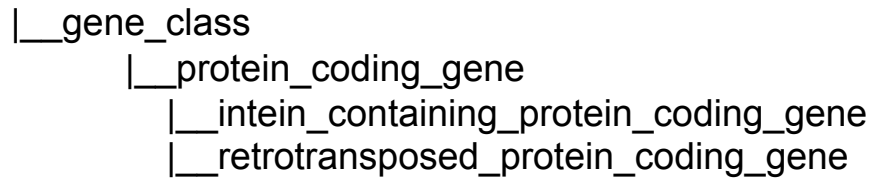
FlyBase is an excellent source of Drosophila references. References can be searched in a number of ways. The easiest way is through **QuickSearch**, on our homepage. Changing the 'Data Class' to search to references alters the layout of the QuickSearch box, providing a text box for Author, Year(s), and All text.

More refined reference searches can be performed using **QueryBuilder** (QB). Click on the box titled 'Query is empty.. Click here to start building' on the QB start page to begin the search. At this stage the window will be displaying all the fields available to search for the 'Genes' dataset. Change the dataset to 'References'. Now the fields found in the reference reports are displayed. From here, you can search all the data found in the reference report, including pubmed ID, author, and type (e.g. review).

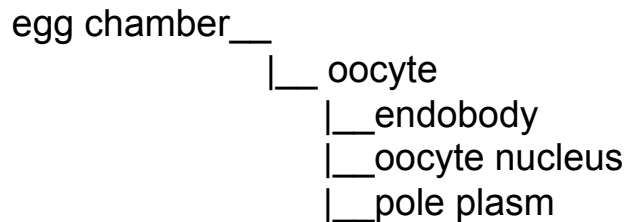
A popular way to search for references is to search for a (list of) objects (e.g. genes, GO terms) and then to use the 'Show related' toggle on the hits page to change the hit list to the related references. The **'Results Analysis/Refinement'** button, found on the hit list page, can be used to analyse the distribution of the references over year, journal, author, and type of publication (e.g. review, paper, abstract).

CV terms are organized into hierarchies which encode the logical relationships between terms:

- protein coding gene is a type of gene



- oocyte is part of the egg chamber



- CV hierarchies allow things carrying related labels to be grouped and therefore aids accurate searching

Controlled vocabulary	Example term	Associated object
GO process	cell death	genes
GO function	nuclease	genes
GO cell component	nucleus	genes/alleles
sequence ontology	tRNA_gene	genes, aberrations...
anatomy	head	alleles/transcripts/polypeptides
stage	larval stage	alleles/transcripts/polypeptides
phenotypic class	lethal	alleles
mutagen	x ray	alleles
allele class	hypomorph	alleles

**Tools** Files Species

Tools Overview

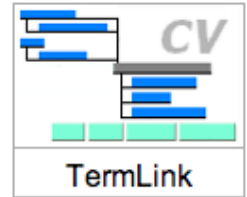
**General Tools:**

- Batch Download
- BLAST
- Chromosome Maps
- Coordinate Converter
- CytoSearch
- GBrowse
- Google™ FlyBase
- ImageBrowse
- Interactions Browser
- QueryBuilder
- QuickSearch
- TermLink**

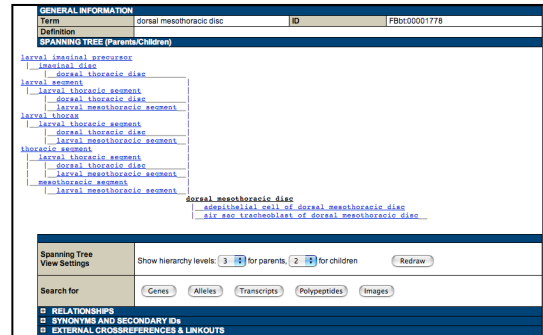
**People:**

- Find A Person
- Add A New Person
- Update An Address

**TermLink** is the only search tool in FlyBase that allows users to search directly for a controlled vocabulary (CV) term from any of the controlled vocabularies used by FlyBase.



TermLink provides hierarchy structures (called ontologies) of the Anatomy, GO Cellular Component and GO Biological process CVs, plus other ontologies. Users can browse through these controlled vocabulary terms by clicking on the links provided.



TermLink also has a search box that allows users to search through ALL of the CVs used by FlyBase. These controlled terms are detailed in the Term Report.

**Tools** Files Species

Tools Overview

**General Tools:**

- Batch Download
- BLAST
- Chromosome Maps
- Coordinate Converter
- CytoSearch
- GBrowse
- Google™ FlyBase
- ImageBrowse**
- Interactions Browser
- QueryBuilder
- QuickSearch
- TermLink

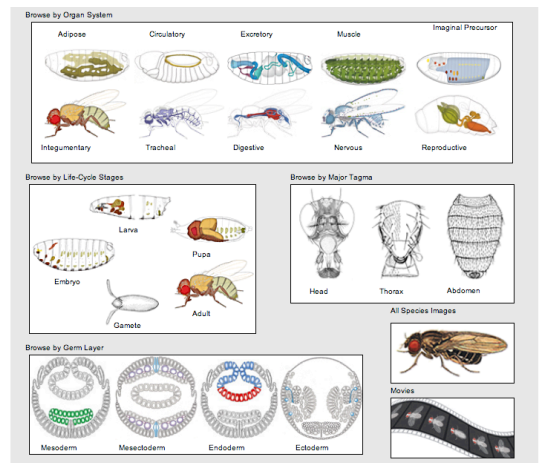
**People:**

- Find A Person
- Add A New Person
- Update An Address

**ImageBrowse** allows the user to browse through image reports by organ system, life-cycle, tagma, or germ layer, as well as browsing images of different Drosophilids.



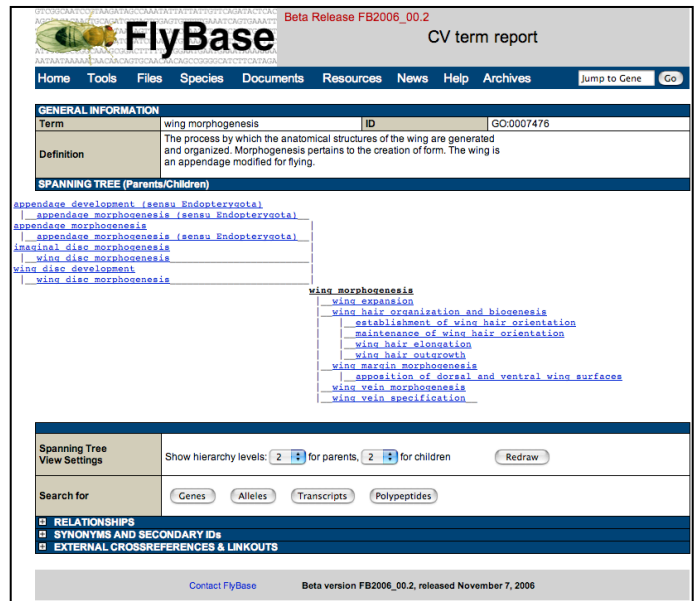
Miscellaneous images and quick-time films are also accessible from this section. V terms are used to annotate and label the images. To search images, and to link relevant gene, allele, transcript and protein records to stages of development, a region of the body or to a specific body part, Termlink can also be used.





The **term report page** provides information regarding a controlled vocabulary term used in FlyBase. Where available a definition is provided, along with a spanning tree, detailing the relationship of the term to other terms in the structured controlled vocabulary (ontology). The spanning tree can be expanded or collapsed both up and down the tree from the term of interest. Synonyms and external linkouts are also provided where possible.

Clicking on any term in the spanning tree takes you to the report for that term. A list of the different types of relationships between the term of interest and other terms in the hierarchy is provided, along with links to search for related objects in FlyBase, such as genes, alleles, transcripts, polypeptides and images, associated with the term.



**CV term report**

Home Tools Files Species Documents Resources News Help Archives Jump to Gene  Go

<b>Term</b>	wing morphogenesis	<b>ID</b>	GO:0007476
<b>Definition</b>	The process by which the anatomical structures of the wing are generated and organized. Morphogenesis pertains to the creation of form. The wing is an appendage modified for flying.		

**SPANNING TREE (Parents/Children)**

```

appendage_development (sensu Endopterygota)
├── appendage_morphogenesis (sensu Endopterygota)
├── appendage_morphogenesis
│   ├── appendage_morphogenesis (sensu Endopterygota)
│   ├── imaginal_disc_morphogenesis
│   ├── wing_disc_morphogenesis
│   └── wing_disc_development
│       └── wing_disc_morphogenesis
└── wing_morphogenesis
    ├── wing_morphogenesis
    ├── wing_hair_organization_and_biosynthesis
    │   ├── establishment_of_wing_hair_orientation
    │   ├── maintenance_of_wing_hair_orientation
    │   └── wing_hair_elongation
    ├── wing_hair_outgrowth
    ├── wing_margin_morphogenesis
    │   ├── apposition_of_dorsal_and_ventral_wing_surfaces
    │   ├── wing_vein_morphogenesis
    │   └── wing_vein_specification
    └── wing_vein_specification
    
```

Spanning Tree View Settings: Show hierarchy levels: 2 for parents, 2 for children

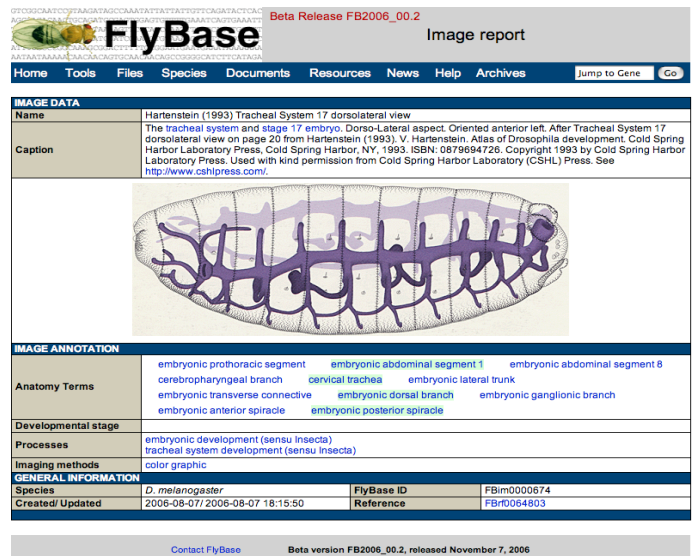
Search for:

RELATIONSHIPS  
 SYNONYMS AND SECONDARY IDs  
 EXTERNAL CROSSREFERENCES & LINKOUTS

Contact FlyBase Beta version FB2006\_00.2, released November 7, 2006

The **Image report page** includes an annotated image and a listing of associated vocabulary terms. Where possible, a link to the reference from which the image is taken is provided, along with the caption associated with the original image.


A list of the controlled vocabulary terms annotated in the image is shown underneath the image. Clicking on a term will take you to the Term report. From here, you can search for images, alleles, genes, and transcripts relating to that term.



**Image report**

Home Tools Files Species Documents Resources News Help Archives Jump to Gene  Go

<b>Name</b>	Hartenstein (1993) Tracheal System 17 dorsolateral view		
<b>Caption</b>	The tracheal system and stage 17 embryo, Dorsolateral aspect. Oriented anterior left. After Tracheal System 17 dorsolateral view on page 20 from Hartenstein (1993). V. Hartenstein. Atlas of Drosophila development. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993. ISBN: 0879694726. Copyright 1993 by Cold Spring Harbor Laboratory Press. Used with kind permission from Cold Spring Harbor Laboratory (CSHL) Press. See <a href="http://www.cshlpress.com/">http://www.cshlpress.com/</a> .		



<b>Anatomy Terms</b>	embryonic prothoracic segment	embryonic abdominal segment 1	embryonic abdominal segment 8
	cerebropharyngeal branch	cervical trachea	embryonic lateral trunk
	embryonic transverse connective	embryonic dorsal branch	embryonic ganglionic branch
	embryonic anterior spiracle	embryonic posterior spiracle	

<b>Developmental stage</b>	embryonic development (sensu Insecta)		
<b>Processes</b>	tracheal system development (sensu Insecta)		
<b>Imaging methods</b>	color graphic		

<b>Species</b>	<i>D. melanogaster</i>	<b>FlyBase ID</b>	FBim0000674
<b>Created/Updated</b>	2006-08-07 2006-08-07 18:15:50	<b>Reference</b>	FBID064803

Contact FlyBase Beta version FB2006\_00.2, released November 7, 2006

Where known and appropriate, the developmental stage of the fly shown in the image is given. The species which is shown in the image, and the date that the image was created is also given.

When you perform any search that returns multiple hits, you are presented with a hit-list, that can be modified or refined. By default all records are selected for inclusion in subsequent manipulations, but the checkboxes allow user-defined subsets to be created. The first data column links directly to the report for each record that matched your search. Other columns link to GBrowse or to searches that return hits directly related to that record. In addition to these links, the hit list provides a set of powerful tools for query refinement or batch processing.

The 'Show related' drop down menu enables you to see all objects of a particular class that are related to the hits selected in your list. For example, selecting 'clones' from the 'Show related' menu of a gene search will return a list of clones that are related to the selected genes.

5977 matches Show related

#	Symbol	Annotation	Cytology	Alleles	Stocks	Clones
1	7SLRNA	CR32864	8A45-8A45	-	-	-
2	abdA	CG10325	89E2-89E2	87	22	32
3	AbdB	CG11648	89E4-89E5	203	60	27
4	abo	CG6093	32C1-32C1	8	7	11
5	ac	CG3796	1A8-1A6	44	56	9
6	Ace	CG17907	87E2-87E3	65	7	40
7	nAcrB-64B	CG11348	64B6-64B6	8	5	39
8	Act5C	CG4027	5C7-5C7	47	29	177
9	Act42A	CG12051	42A7-42A7	7	7	71
10	Act67E	CG18290	87E11-87E11	2	2	61

**EXPORT selected HITS TO:**

- Batch Download
- QueryBuilder (new session)
- FlyBase (Isa file to store on your computer)
- Cross-referencing FlyBase Records Table
- Got FlyBase-curated external data links:
  - Nucleotide Sequences
  - Protein Data
  - Protein Domains
- Got LinkOut external data links:
  - InParanoid computed orthology calls
  - FlyExpress - Expression Pattern Search
  - Fly GRID Interaction Data
  - FlyAtlas - Drosophila adult expression atlas
  - FLIGHT - Integrating Genomic and High-Throughput data
  - DRSC - Drosophila RNAi Screening Center
  - DEDB - Drosophila melanogaster Exon Database
  - FlyMine - integrated genomics and proteomics
  - NCBI Gene Expression Omnibus (GEO)
  - Heidelberg Database for RNAi Phenotypes
  - Hybridomics Drosophila PIMRider
  - Interactive Fly
  - PANTHER Classification System
  - BDGP in situ Gene Expression Database
  - REDfly - Regulatory Element Database for Drosophila

For the selected records, analyze frequency of values of:

- Molecular Function
- Biological Process
- Cellular Component
- Polypeptide Subcellular Localization
- Transcript Subcellular Localization
- Chromosome arm

For the selected records, browse Interactions::

- Protein-Protein
- Genetic Interactions for the effect of:
  - Enhancement
  - Enhancement + Suppression
  - Suppression

#	Most frequent field values (out of 3143)	Related records	%
1	[ empty field - no data available ]	2887	48%
2	non-traceable author statement	625	10%
3	inferred from sequence or structural similarity	508	8%
4	protein binding ; GO:000515	350	5%
5	DNA binding ; GO:0003677	311	5%
6	zinc ion binding ; GO:0008270	282	4%
7	ATP binding ; GO:0005524	253	4%
8	inferred from direct assay	250	4%
9	no biological data available	205	3%
10	molecular_function ; GO:0003674	205	3%
11	transcription factor activity ; GO:0003700	181	3%
12	binding ; GO:0005488	149	2%
13	nucleic acid binding ; GO:0003676	129	2%
14	traceable author statement	124	2%
15	structural constituent of ribosome ; GO:0003735	122	2%
16	sequence-specific DNA binding ; GO:0043565	89	1%
17	serine-type endopeptidase activity ; GO:0004252	78	1%
18	RNA binding ; GO:0003723	77	1%
19	calcium ion binding ; GO:0005509	61	1%
20	transporter activity ; GO:0005215	60	1%

The 'Results Analysis/Refinement' button allows you to see the frequency of values within your selected hits for a predefined list of fields. Selecting 'Molecular Function', for example, from the Results Analysis/Refinement tool, will result in a page listing the distribution of the different molecular function controlled vocabulary terms associated with the list. Clicking on the number in the 'Related records' column will return the genes from your hit-list that are annotated to be involved in that GO term.

Lastly, the 'Hit-List Conversion Tools' button allows you to send the selected hits to our Batch Download tool for use offline, to a new QueryBuilder session for further querying, or to link-out HTML tables of various third party data sources with data linked to the hits in your result list.

## Query Tools:

	Jump to Gene	QuickSearch	QueryBuilder
<b>Data class</b>	Genes	Any one data class	Any combination of data classes
<b>Input</b>	Symbol, full name, synonym, ANY valid FBid	...plus "all text" option	Field specific
<b>Booleans</b>	None		AND, OR, BUT NOT

## Controlled terms:

	TermLink	ImageBrowse	QueryBuilder
<b>Data class</b>	Genes, alleles, transcripts, polypeptides, insertions, images	Images	Any combination of data classes except images
<b>Input</b>	Controlled term from any of the the controlled vocabularies	Browse visually through the anatomy and developmental stage controlled vocabularies	Controlled term from any of the the controlled vocabularies
<b>Booleans</b>	None		AND, OR, BUT NOT

**QueryBuilder** (QB) provides the most powerful way to search FlyBase on a field-by-field level. QB presents a simple user interface that supports powerful searches by offering access to every DataSet|Field pair (for example, Genes|CV:GO:Molecular Function) in FlyBase, along with the ability to include any combination of datasets in the same search.

QB automatically creates sets of records that are cross-referenced to the records that match your query, providing links to all related records in FlyBase from a single page. Both simple and complex queries can be built in a few steps. A search can be focused to a particular piece of data within a report page, such as the 'mapped features and mutations' associated with a gene, and Boolean operators can be used to combine two or more searches.

QB allows a user to perform much more sophisticated searches compared to QuickSearch or other search tools on FlyBase, that take full advantage of how the data is stored in FlyBase. A useful feature of QB is that a list of FlyBase identifiers or valid symbols can be imported from an external file to use as a query segment. In addition, a set of results can be exported to QB, as described in the 'Hit list refinement' section, and then modified to refine the search by adding additional query segments. Thus, QB is a very powerful tool that can be used in many different ways to explore the data in FlyBase. However, QB is also the most complicated tool on FlyBase and therefore the rest of this booklet is devoted to outlining the use of QB.

### Introducing the data that is searchable using QB

To use QB effectively, it helps to know a little about how FlyBase data are structured and how QB interacts with that structure.

QB allows the user to search through 11 different Datasets:

- Annotations
- Genes
- Alleles
- Aberrations
- Balancers
- Transposons
- Insertions
- Transcripts
- Polypeptides
- Stocks
- References

Specific data within each DataSet lives within database fields. Examples of database fields are 'cytogenetic map', 'class of gene', and 'Anatomy'. You can search through these using the QuerySegments in QB. DataBase fields are linked directly to a reference and to the most relevant type of genetic object. For example, fields containing expression data are linked to transcripts or peptides, while fields containing phenotypic data are linked to alleles.

These fields fall into five major categories and are marked as such:

CV = controlled vocabulary or ontology. FlyBase uses set terms to describe certain biological aspects of *Drosophila* in order to make these things easy to search. These CVs include Gene Ontology (GO) terms to describe the wild-type gene function, an anatomy CV, a developmental stage CV and a phenotype CV. These can be found by selecting GO/Anatomy CV DB from the Query Segment of QB.

Flag = In those fields marked with 'Flag', you can only search for the presence or absence of information. QB treats any text entered for Flag fields as if the text reads "IS NOT NULL", meaning that any records included in the query results contain relevant data for that particular database field.

Map = any database fields containing data relevant to *Drosophila* genetic maps.

Symbol = database fields that are searchable by entering FlyBase symbols for genes, alleles, aberrations etc.

Text = database fields that are searchable by entering free text.

### **A note on the cross-referencing section of QB**

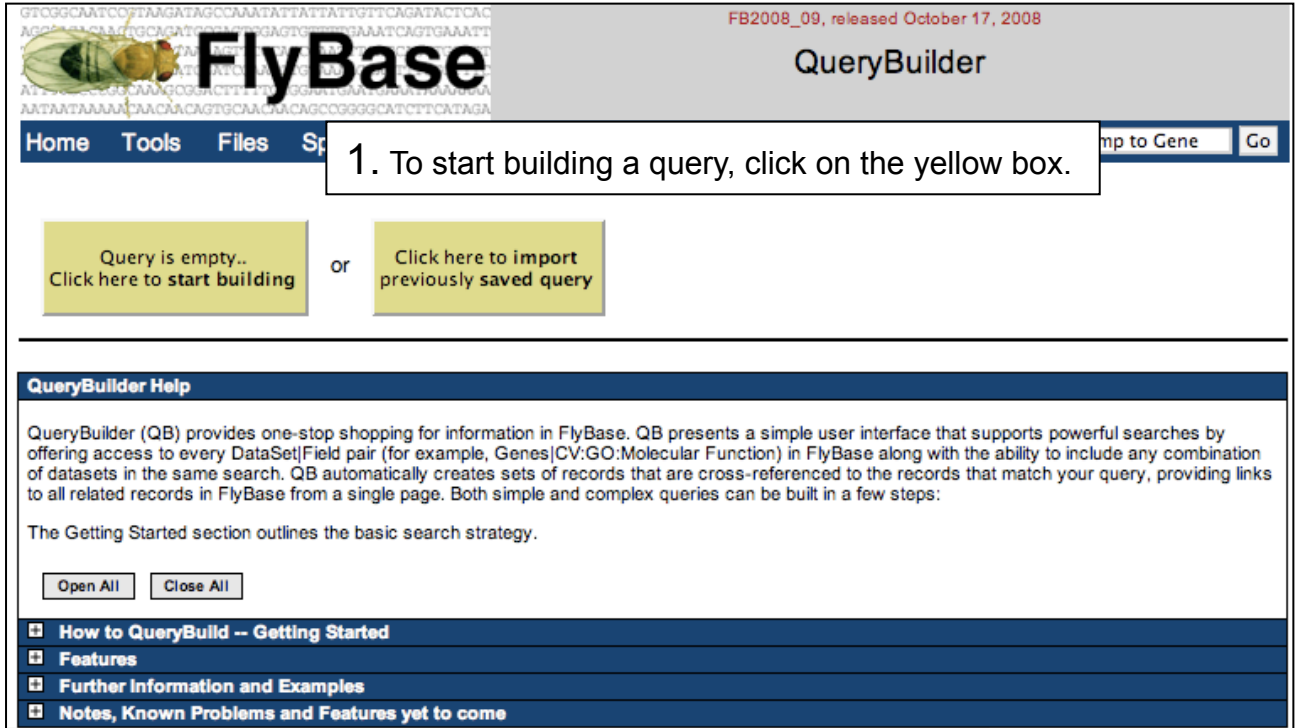
QB generates cross-referencing lists according to the links between the various genetic objects; for example, genes can have alleles, transcripts and peptides. The output of this can be switched on or off as this tutorial will show. This cross-referencing gives QB the capacity to combine multiple searches even when those searches target data linked to different types of genetic object.

Data linked to genetic objects can take a number of forms. In some cases the same CV may be used for more than one purpose. For example, an anatomy term might have a direct link to a transcript (indicating expression) and a direct link to an allele (indicating a phenotype). As a result a list of genes generated by searching with an anatomy term can combine hits due to expression with those due to phenotype.

All associations with ontology terms live in one big table. You can search this directly by choosing the option 'GO/Anatomy CV DB' from the menu headed 'Search DataSet'. Choosing any data field populated by CV terms will cause QB to default to searching the whole table. As a result, any search using an anatomy term can find alleles, transcripts and peptides. These can be displayed separately, but all contribute to the list of indirectly linked genes.

The following pages illustrate how to perform a simple search, and are followed by various examples of the types of searches possible with QB.

## How to Perform a basic search



1. To start building a query, click on the yellow box.

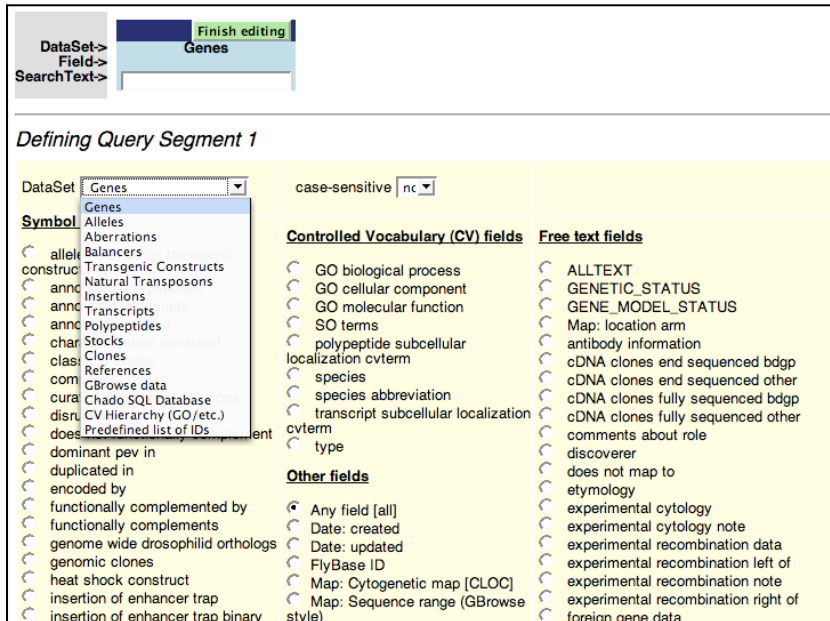
Query is empty..  or

**QueryBuilder Help**

QueryBuilder (QB) provides one-stop shopping for information in FlyBase. QB presents a simple user interface that supports powerful searches by offering access to every DataSet|Field pair (for example, Genes|CV:GO:Molecular Function) in FlyBase along with the ability to include any combination of datasets in the same search. QB automatically creates sets of records that are cross-referenced to the records that match your query, providing links to all related records in FlyBase from a single page. Both simple and complex queries can be built in a few steps:

The Getting Started section outlines the basic search strategy.

- How to QueryBuild -- Getting Started**
- Features**
- Further Information and Examples**
- Notes, Known Problems and Features yet to come**



DataSet-> Genes

Field->

SearchText->

**Defining Query Segment 1**

DataSet: Genes case-sensitive:  nc

Symbol	Controlled Vocabulary (CV) fields	Free text fields
<input type="checkbox"/> Alleles	<input type="checkbox"/> GO biological process	<input type="checkbox"/> ALLTEXT
<input type="checkbox"/> Aberrations	<input type="checkbox"/> GO cellular component	<input type="checkbox"/> GENETIC_STATUS
<input type="checkbox"/> Balancers	<input type="checkbox"/> GO molecular function	<input type="checkbox"/> GENE_MODEL_STATUS
<input type="checkbox"/> Transgenic Constructs	<input type="checkbox"/> SO terms	<input type="checkbox"/> Map: location arm
<input type="checkbox"/> Natural Transposons	<input type="checkbox"/> polypeptide subcellular localization cvterm	<input type="checkbox"/> antibody information
<input type="checkbox"/> Insertions	<input type="checkbox"/> species	<input type="checkbox"/> cDNA clones end sequenced bdgp
<input type="checkbox"/> Transcripts	<input type="checkbox"/> species abbreviation	<input type="checkbox"/> cDNA clones end sequenced other
<input type="checkbox"/> Polypeptides	<input type="checkbox"/> transcript subcellular localization cvterm	<input type="checkbox"/> cDNA clones fully sequenced bdgp
<input type="checkbox"/> Stocks	<input type="checkbox"/> type	<input type="checkbox"/> cDNA clones fully sequenced other
<input type="checkbox"/> Clones	<b>Other fields</b>	<input type="checkbox"/> comments about role
<input type="checkbox"/> References	<input checked="" type="radio"/> Any field [all]	<input type="checkbox"/> discoverer
<input type="checkbox"/> GBrowse data	<input type="checkbox"/> Date: created	<input type="checkbox"/> does not map to etymology
<input type="checkbox"/> Chado SQL Database	<input type="checkbox"/> Date: updated	<input type="checkbox"/> experimental cytology
<input type="checkbox"/> CV Hierarchy (GO/etc.)	<input type="checkbox"/> FlyBase ID	<input type="checkbox"/> experimental cytology note
<input type="checkbox"/> Predefined list of IDs	<input type="checkbox"/> Map: Cyto genetic map [CLOC]	<input type="checkbox"/> experimental recombination data
<input type="checkbox"/> element	<input type="checkbox"/> Map: Sequence range [GBrowse style]	<input type="checkbox"/> experimental recombination left of
<input type="checkbox"/> dominant pev in		<input type="checkbox"/> experimental recombination note
<input type="checkbox"/> duplicated in		<input type="checkbox"/> experimental recombination right of
<input type="checkbox"/> encoded by		<input type="checkbox"/> foreign gene data
<input type="checkbox"/> functionally complemented by		
<input type="checkbox"/> functionally complements		
<input type="checkbox"/> genome wide drosophilid orthologs		
<input type="checkbox"/> genomic clones		
<input type="checkbox"/> heat shock construct		
<input type="checkbox"/> insertion of enhancer trap		
<input type="checkbox"/> insertion of enhancer trap binary		

A query segment will automatically open. The box at the top of the segment shows you what choices you have made as you build the segment.

2. The first step is to select the DataSet in which you would like to search.

DataSet->  
Field->  
SearchText->

Finish editing  
References  
Any field

---

### Defining Query Segment 1

DataSet

case-sensitive

**Controlled Vocabulary (CV) fields**

- abstract languages
- publication languages
- published in abbreviation
- published in language
- published in title
- type

**Other fields**

- Any field [all]
- FlyBase ID
- biosis
- isbn

- ALLTEXT
- URL
- also published as
- associated files
- authors
- comments
- conference abstract
- fulltext url
- issue
- miniref
- pages
- place of publication

For most of the datasets, e.g. genes, alleles, or references, database fields are provided to focus your search. Many different types of data, including symbols, controlled terms, sequence ranges and free text can be searched within certain database fields.

DataSet->  
Field->  
SearchText->

Finish editing  
CV Hierarchy (GO/etc.)  
Any field  
-

---

### Defining Query Segment 1 - selecting CV term

DataSet

Retrieve records annotated with

[search hierarchies](#) or select root CV term to browse hierarchy from top:

<b>GO: Biological Process</b>	<a href="#">reproduction</a> <a href="#">development</a> <a href="#">physiological process</a> <a href="#">cellular process</a> <a href="#">growth</a> <a href="#">pigmentation</a> <a href="#">regulation</a> <a href="#">response to stimulus</a> <a href="#">interaction between organisms</a>
<b>GO: Molecular Function</b>	<a href="#">binding</a> <a href="#">catalytic</a> <a href="#">enzyme re</a> <a href="#">translation regulator</a> <a href="#">transpo</a> <a href="#">energy transducer</a> <a href="#">chemoattr</a>
<b>GO: Cellular Component</b>	<a href="#">cell</a> <a href="#">extracellular region</a> <a href="#">ext</a>
<b>Anatomy CV</b>	<a href="#">developing embryonic structure</a>
<b>Developmental Stage</b>	<a href="#">oogenesis</a> <a href="#">egg</a> <a href="#">embryonic</a> <a href="#">larval</a> <a href="#">prepupal</a> <a href="#">pupal</a> <a href="#">adult</a> <a href="#">pnarate adult</a>
<b>Phenotype Ontology</b>	<a href="#">allele morphy class</a> <a href="#">phenotypic class</a> <a href="#">origin of mutation</a>

Clicking on certain DataSets in the table above changes the appearance of the QuerySegment. For example, clicking on the CV Hierarchy (GO/ etc) dataset, will open the QuerySegment shown here.

## Defining Query Segment 1 - selecting CV term

DataSet:  Retrieve records annotated with:

search hierarchies or select root CV term to browse hierarchy from top:

<b>GO: Biological Process</b>	reproduction development physiological process cellular process growth pigmentation regulation response to stimulus interaction between organisms
<b>GO: Molecular Function</b>	binding catalytic enzyme regulator signal transduction transcription regulator translation regulator transport chaperone regulator protein tag energy transduction nutrient reservoir
<b>GO: Cellular Component</b>	cell extracellular region extracellular matrix organelle
<b>Anatomy CV</b>	developing embryonic structure extraembryonic structure
<b>Developmental Stage</b>	oogenesis egg embryonic larval prepupal pupal
<b>Phenotype Ontology</b>	allele morphy class phenotypic class origin of mutation

4. To find a specific CV term using a text search, type the test in the 'search hierarchies' box. You can also browse through the CV terms by clicking the parent terms in each ontology.

**CV terms matching your search context (total of 8):**

Use this term ID	dorsal mesothoracic disc ; FBbt:00001778 (synonym: wing disc)
Use this term ID	determination of wing disc primordium ; GO:0035294
Use this term ID	wing disc morphogenesis ; GO:0007472 (synonym: wing disc metamorphosis)
Use this term ID	wing disc anterior/posterior pattern formation ; GO:0048100
Use this term ID	wing disc development ; GO:0035220
Use this term ID	wing disc dorsal/ventral pattern formation ; GO:0048190
Use this term ID	wing disc pattern formation ; GO:0035222
Use this term ID	wing disc proximal/distal pattern formation ; GO:0007473

5. To select the term, click on the green 'Use this term ID' button

CV spanning tree for "adepithelial cell" (FBbt:00003219)

```

material_anatomical_entity
├── anatomical_structure
│   ├── cell
│   │   ├── organ_system
│   │   │   ├── muscle_system
│   │   │   │   ├── adult_muscle_system
│   │   │   │   └── organiam
│   │   │   │       ├── adult
│   │   │   │       │   ├── adult_muscle_system
│   │   │   │       │   └── organ_system
│   │   │   │       │       ├── muscle_system
│   │   │   │       │       └── adult_muscle_system
│   │   └── adepithelial_cell
│   │       ├── adepithelial_cell_of_clypeo-labral_disc
│   │       ├── adepithelial_cell_of_dorsal_mesothoracic_disc
│   │       ├── adepithelial_cell_of_dorsal_mesothoracic_disc
│   │       ├── adepithelial_cell_of_dorsal_mesothoracic_disc
│   │       ├── adepithelial_cell_of_eye-antennae
│   │       ├── adepithelial_cell_of_genital_disc
│   │       ├── adepithelial_cell_of_labial_disc
│   │       ├── adepithelial_cell_of_ventral_mesothoracic_disc
│   │       └── adepithelial_cell_of_ventral_mesothoracic_disc
    
```

To find out more about a term, click the yellow term name button, which will take you to a hierarchical view of the term illustrating its relationship with other terms.

6. To finish a query segment, click the green 'Finish editing' button.

DataSet->

Field-> CV Hierarchy (GO/etc.)

SearchText-> wing disc morphogenesis



DataSet-> CV Hierarchy (GO/etc.) x

Field-> CV term +

SearchText-> wing disc morphogenesis

7. To start the search, click the green 'Run query' button.

Show me related 
Species: 
Run query
Interface style:

Your results will appear in a table like the one below. Additionally, QB creates a set of cross-references for the records that match your search criteria. These separate into direct and indirect links, and provide access to related records in other data classes.

[ # of leads ] 227 x

DataSet-> CV Hierarchy (GO/etc.) +

Field-> CV term

SearchText-> wing disc morphogenesis

**Direct links: Records matching your query *and* annotated with relevant CV term(s) :**

Genes  
223

**Indirect links: Records matching your search *or* crossreferencing its results (Species Filter == Dmel) :**

Aberrations Alleles Clones Genes Polypeptides Stocks Insertions Transgenic Constructs Transcripts

223 matches
Store This Query
Results Analysis/Refinement
HitList Conversion Tools

	#	^Symbol^	^Name^	^Annotation ID^	^Cytology^	^Alleles #^	^Stocks #^	^Clones #^
<input checked="" type="checkbox"/>	1	<a href="#">2D5</a>	2D5	-	-	1	-	-
<input checked="" type="checkbox"/>	2	<a href="#">2R-F</a>	2R-F	-	-	1	1	-
<input checked="" type="checkbox"/>	3	<a href="#">2R-L</a>	2R-L	-	-	1	1	-

**For the selected records, analyze frequencies of values of:**

- [Molecular Function](#)
- [Biological Process](#)
- [Cellular Component](#)
- [Polypeptide Subcellular Localization](#)
- [Transcript Subcellular Localization](#)
- [Chromosome arm](#)

**For the selected records, browse Interactions::**

- **Genetic Interactions for the effect of:**
- [Enhancement](#)
- [Enhancement + Suppression](#)
- [Suppression](#)

**EXPORT selected HITS TO:**

- [Batch Download](#)
- [QueryBuilder \(new session\)](#)
- [FlyBase IDs file \(to store on your computer\)](#)
- [Cross-referencing FlyBase Records Table](#)

**Get FlyBase-curated external data links:**

- [Nucleotide Sequences](#)
- [Protein Data](#)
- [Protein Domains](#)

**Get LinkOut external data links:**

- [InParanoid computed orthology calls](#)
- [FlyExpress - Expression Pattern Search](#)
- [Fly GRID Interaction Data](#)
- [FlyAtlas - Drosophila adult expression atlas](#)
- [FLIGHT - Integrating Genomic and High-Throughput data](#)
- [DRSC - Drosophila RNAi Screening Center](#)
- [DEDB - Drosophila melanogaster Exon Database](#)
- [FlyMine - integrated genomics and proteomics](#)
- [NCBI Gene Expression Omnibus \(GEO\)](#)
- [Heidelberg Database for RNAi Phenotypes](#)
- [Hybrigenics Drosophila PIMRider](#)
- [Interactive Fly](#)
- [PANTHER Classification System](#)
- [BDGP in situ Gene Expression Database](#)
- [REDfly - Regulatory Element Database for Drosophila](#)

If the records identified by your search include links to external databases, these links are available by clicking on the HitList Conversion Tools box. You can then choose the database of interest from the list in the box that appears.

33

Note: your hitlist may differ slightly from the one shown here as this is from an older release.

## How to download your results

[ # of leads ] 227 x

DataSet-> CV Hierarchy (GO/etc.) +

Field-> CV term

SearchText-> wing disc morphogenesis

**Direct links: Records matching your query *and* annotated with relevant CV term(s) :**

Genes  
223

**Indirect links: Records matching your search *or* crossreferencing its results (Species):**

Aberrations  
465
Alleles  
10644
Clones  
6810
Genes  
CURRENT VIEW
Polypeptides  
494
Stocks  
7970
Insertions  
2618
Transcripts  
2963
Proteins  
494

223 matches Store This Query Results Analysis/Refinement HitList Conversion Tools

✓	#	Symbol	Name	Genes	Stocks	Clones
✓	1	2D5	2D5			
✓	2	2R-F	2R-F			
✓	3	2R-L	2R-L			
✓	4	3F43	3F43			
✓	5	3L-B	3L-B			
✓	6	3R-C	3R-C			
✓	7	Adv	Additional veins			
✓	8	alph	alphabet			2
✓	9	ap	apterous		78	57
✓	10	apb	apterblister			

**EXPORT selected HITS TO:**

- [Batch Download](#)
- [QueryBuilder \(new session\)](#)
- [FlyBase IDs file \(to store on your computer\)](#)

**Get FlyBase-curated external data links:**

- [Nucleotide Sequences](#)
- [Protein Data](#)
- [Protein Domains](#)

**Get LinkOut external data links:**

- [InParanoid computed orthology calls](#)
- [FlyExpress - Expression Pattern Search](#)
- [Fly GRID Interaction Data](#)
- [FlyAtlas - Drosophila adult expression atlas](#)
- [FLIGHT - Integrating Genomic and High-Throughput data](#)
- [DRSC - Drosophila RNAi Screening Center](#)
- [DEDB - Drosophila melanogaster Exon Database](#)
- [FlyMine - integrated genomics and proteomics](#)
- [NCBI Gene Expression Omnibus \(GEO\)](#)
- [Heidelberg Database for RNAi Phenotypes](#)
- [Hybrigenics Drosophila PIMRider](#)
- [Interactive Fly](#)
- [PANTHER Classification System](#)
- [BDGP in situ Gene Expression Database](#)
- [REDfly - Regulatory Element Database for Drosophila](#)

1. Click on the HitList Conversion Tools box

2. Click on the BatchDownload Tool option.

### Batch Download

#### Output Format

**FASTA Sequence**

**Database Format**  
Full Data Only

**Field Data**  
Selected Fields Only

#### Output Options

Gene region

Chado XML

As HTML table

**Send results to:** Browser

You may enter FlyBase IDs or Symbols, including Annotation Symbols and Clone Names.

**Enter IDs, Symbols or Sequence Coordinates:**

FBgn0067331

FBgn0021755

FBgn0021754

FBgn0067329

FBgn0021753

FBgn0021751

FBgn0004614

**or Upload File of IDs:** Browse...

Allow synonyms and multi-variant conversions of IDs or Symbols

Get FastA Reset

3. Select the Field Data radio button.

A list of IDs that correspond to the genes in the QueryBuilder Hit-List will be automatically entered into this box.


**Batch Download Help**

Last Updated: 30 Septem

The Batch Download tool provides access to a variety of data and data formats for a specified list of IDs. The specified list of IDs can be large (e.g. all genes) or small (e.g. one gene), but each ID should be provided on a separate line. The types of data available through Batch Download include FASTA sequences, XML files, and data from specified fields on reports. Results can be downloaded or viewed online.

Open All Close All

- Getting Started**
- How to Download FASTA Sequence**
- How to Download Database XML**
- How to Download Field Data**



FB2008\_09, released October 17, 2008

### Batch Download

Home Tools Files Species Documents Resources News Help Archives

Jump to Gene

#### Output Format

FASTA Sequence

#### Output Options

Gene region

Chado XML

As HTML table

**Send results to:**

You may enter FlyBase IDs or Symbols, including Annotation Symbols and Clone Names.

**Enter IDs, Symbols or Sequence Coordinates:**

FBgn0067331

FBgn0021755

FBgn0021754

FBgn0067329

FBgn0021753

FBgn0021751

FBgn0004614

or Up

Allow synonyms and multi-variant conversions of IDs or Symbols

Tip: if you want to save your results, choose 'file' instead of 'browser' here.

4. Click on Select Fields to choose which aspects of your results to download.

Date: created

Date: updated

FlyBase ID

location max

location min

location strand

CV: GO biological process

CV: GO cellular component

CV: GO molecular function

CV: SO terms

CV: polypeptide subcellular localization cvterm

CV: species abbreviation

CV: transcript subcellular localization cvterm

CV: type

Symbol: alleles carried on transgenic constructs

Symbol: annotated polypeptides

Symbol: annotated transcripts

Symbol: annotation symbol

Symbol: characterization construct

Symbol: classical alleles

5. From here, you can choose which fields you would like to download for your results. Only a subset of the complete list of field is shown here.

Get field data

SUBMITTED ID	FBID KEY	SYMBOL
<a href="#">FBgn0067331</a>	<a href="#">FBgn0067331</a>	2D5
<a href="#">FBgn0021755</a>	<a href="#">FBgn0021755</a>	2R-F
<a href="#">FBgn0021754</a>	<a href="#">FBgn0021754</a>	2R-L
<a href="#">FBgn0067329</a>	<a href="#">FBgn0067329</a>	3F43
<a href="#">FBgn0021753</a>	<a href="#">FBgn0021753</a>	3L-B
<a href="#">FBgn0021751</a>	<a href="#">FBgn0021751</a>	3R-C
<a href="#">FBgn0004614</a>	<a href="#">FBgn0004614</a>	Adv
<a href="#">FBgn0086361</a>	<a href="#">FBgn0086361</a>	alph
<a href="#">FBgn0000099</a>	<a href="#">FBgn0000099</a>	ap
<a href="#">FBgn0000101</a>	<a href="#">FBgn0000101</a>	apb
<a href="#">FBgn0015904</a>	<a href="#">FBgn0015904</a>	ara
<a href="#">FBgn0004569</a>	<a href="#">FBgn0004569</a>	argos
<a href="#">FBgn0000139</a>	<a href="#">FBgn0000139</a>	ash2
<a href="#">FBgn0015905</a>	<a href="#">FBgn0015905</a>	ast
<a href="#">FBgn0029907</a>	<a href="#">FBgn0029907</a>	Atx-1

A table of your results will then appear in your browser.

### Examples of QueryBuilder searches.

1. How many publications exist for Ashburner in the year 2001?

Answer: start query and select **DataSet**: References, **Free text fields**: authors, enter text in the querysegment: Ashburner, and click 'Finish editing'. Click on the plus to open a second query segment, select **DataSet**: References, **Other Fields**: year, enter the text to find: 2001, click 'Finish editing'. Now select *References* from Show me related menu and click 'Run query'.

Result = 11.

2. Find all of the enhancer trap GAL4 lines that are expressed in the nervous system.

Start query and select **DataSet**: Insertions, **Free text fields**: expression marker for, Enter text: nervous system, click 'Finish editing'. Click plus to open a new segment, select **DataSet**: Insertions, **Free text field**: expression data, enter text: GAL4. Select *Insertions* from the Show me related menu and click 'Run query'.

Result = 23 matches. (2 query segments are needed here as the 'expression marker for' field contains data for both lacZ and GAL4. Therefore, the second segment refines the search to GAL4 only.)

3. Find all the lethal alleles of hep that result from mutation at the locus.

Answer: start query and select **DataSet:** Genes, **Symbol field:** symbol, Enter text: **hep**, and click Done. Click on the plus to open a new segment, select **DataSet:** CV Hierarchy, click to browse *phenotypic class*. Choose *lethal* from the menu of options that opens and click the green 'Use marked CV' box to add to the query segment. The term CV will be entered automatically in the box so now click Done. Click on the next plus to open a third query segment, select **DataSet:** Alleles, **Controlled vocabulary (CV) field:** mutagen, and type 'in vitro construct', click Done. Now click twice on the AND box between the lethal and the in vitro construct query segments so that it becomes BUT NOT. Select Alleles from the Show me related menu and click 'Run query'.

Result = 16 hits.

4. Uploading lists into QueryBuilder.

This can be done using the **DataSet:** Predefined list of IDs. You can either attach a file containing FBIDs or you can upload results from previous searches straight into QB. Here is an example:

Use QuickSearch, select the 'all text' radio button and type kinase. You will get a hitlist of 1912 genes. Let's find out which of these genes are actually annotated with the GO term 'kinase activity'. Click on the HitList Conversion Tool and select '*QueryBuilder (new session)*'. You will be taken to the QB page and the exported Ids will already be present in a query segment. Now add another query segment, select the **Dataset:** CV Hierarchy, then type '*kinase activity*' in the search hierarchies box. The matching terms are in alphabetical order so scroll down to find the matching term that has the ID GO:0016301. Click the green 'Use this term ID' box, then click 'Finish editing'. Make sure that the Show me related menu is set to genes. You will see that 397 of the initial hit list from QuickSearch are annotated with this term.

**This example illustrates the advantages of using QueryBuilder to find a list of relevant genes. QuickSearch found more genes as it found the term 'kinase' somewhere in the gene report page for each hit, whereas QueryBuilder identified those genes that are actually annotated with 'kinase activity'.**