Exploring FlyBase Data Using QuickSearch

Steven J. Marygold[1,*], **Giulia Antonazzo**[1], **Helen Attrill**[1], **Marta Costa**[1], **Madeline A. Crosby**[2], **Gilberto dos Santos**[2], **Joshua L. Goodman**[3], **L. Sian Gramates**[2], **Beverley B. Matthews**[2], **Alix J. Rey**[1], **Jim Thurmond**[3], and **The FlyBase Consortium**[**]

[1]Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, CB2 3DY, UK

[2]The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

[3]Department of Biology, Indiana University, Bloomington, IN 47405, USA

Abstract

FlyBase (flybase.org) is the primary online database of genetic, genomic and functional information about *Drosophila* species, with a major focus on the model organism *Drosophila melanogaster*. The long and rich history of *Drosophila* research, combined with recent surges in genomic-scale and high-throughput technologies, mean that FlyBase now houses a huge quantity of data. Researchers need to be able to rapidly and intuitively query these data, and the QuickSearch tool has been designed to meet these needs. This tool is conveniently located on the FlyBase homepage and is organized into a series of simple tabbed interfaces that cover the major data and annotation classes within the database. This unit describes the functionality of all aspects of the QuickSearch tool. With this knowledge, FlyBase users will be equipped to take full advantage of all QuickSearch features and thereby gain improved access to data relevant to their research.

Keywords

INTRODUCTION

FlyBase is the primary online database of genetic, genomic and functional information about *Drosophila* species, with a major focus on the model organism *Drosophila melanogaster*. Data are incorporated through curation of published research papers and via computational uploads of third-party datasets. Much of these data are partitioned across ~20 'data classes' within FlyBase (Marygold et al., 2016; Table 1). These include classical divisions, such as 'genes', 'alleles' or 'references', as well as more recent additions such as 'sequence features', 'human disease models' or 'gene groups'. Certain data classes are associated with functional, phenotypic or expression annotations, as appropriate, with links established between all related objects. For example, a phenotype will be annotated to a particular allele

*Corresponding author. sjm41@cam.ac.uk Phone: +44 (0)1223 333170.
**The members of the FlyBase Consortium are listed in the Acknowledgements.

of a specified gene, all of which will be associated with the source reference. This organization underlies the presentation of related data within distinct report pages within FlyBase, and enables users to browse, navigate and search effectively across the website.

The QuickSearch tool provides rapid querying of almost all data in FlyBase through a series of concise tabbed interfaces that are accessible directly from the homepage (Fig. 1A; Table 2). It is designed to provide quick, intuitive access to searches that users commonly need to perform. The 'Simple' tab of QuickSearch allows for Google-like querying across all data classes, while the other tabs permit more targeted queries of particular data classes (References, Gene Groups, Human Disease, Orthologs, Data Class), controlled vocabularies used during annotation (Gene Ontology (GO), Disease Ontology (DO)) or direct annotations associated with one or more these data classes (Expression, Phenotype).

The protocols in this unit cover the functionality and usage of each QuickSearch tab. The concluding section describes general features that are applicable to all tabs, together with advice on how to interpret the results of a search and how to trouble-shoot any problems. The protocols and screenshots are based on FlyBase release FB2016_05 (September 2016).

BASIC PROTOCOL 1: USING THE Simple TAB OF QuickSearch

This tab (Fig. 1) performs a comprehensive case-insensitive search across most text fields associated with almost all FlyBase data classes. It provides a 'Google'-like functionality, in that no prior knowledge of the content or organization of FlyBase data is required, the search box accepts any text, and all possible matches are retrieved and reported rapidly. The 'Simple' tab is often the best place to start a search of data within FlyBase.

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1. Choose whether to limit the search to *D. melanogaster* data (the default) or to include reports from additional *Drosophila* species by checking the 'include non-Dmel species' box (Fig. 1A).

   Note that FlyBase also includes reports (e.g. recombinant constructs, stocks) pertaining to genes from species outside of the Drosophila genus. This allows accommodation of data on heterologous expression of, for example, GAL4, GFP, lacZ, and a growing number of human genes. All these cases are classed as 'honorary D. melanogaster genes' for the purposes of the Simple tab, and so will be found regardless of whether the 'include non-Dmel species' box is checked.

2. Enter one or more search terms.

   The search term box of the Simple tab supports a number of additional features that can be used to narrow or broaden the query. A wild card

character (*) can be appended, prepended or added within a search term to broaden the query. When specifying multiple terms, a Boolean 'AND' is used for searches by default and does not require any special notation (e.g. a search for 'neurogenesis microtubule polymerization' will return only hits that have all of those three terms somewhere in the record). A Boolean 'OR' can be added to find records that have one or another of a list of specified terms (e.g. 'cnn OR cbs'). To exclude certain terms from the results, prefix the term(s) to be excluded with a '-' character (e.g. 'Parkinson -CG5680'). Finally, results can be specified to contain an exact phrase by surrounding the search term with double quotes (e.g. "SH3 domain").

3. Click the green 'Search' button or press 'enter'.

4. View matching hits in the results overview page (Fig. 1B).

This displays the list of FlyBase data classes that are searched via the Simple tab and the number of matching records from each. The original query can be modified in the box at the bottom of this page.

5. Click on a data class heading to view the hit-list of results specific to that data class.

Results are listed in descending order of their relevance score.

6. Click on an individual entry in a hit-list to view its report page.

FlyBase report pages display all of the data associated with that entity, organized into separate sections. Reports contain many links to other related reports within and external to FlyBase to aid further navigation and exploration.

BASIC PROTOCOL 2: USING THE Data Class TAB OF QuickSearch

If a query is well defined, it is often useful to perform a more targeted or nuanced search than that offered by the Simple tab. This can be achieved using the 'Data Class' tab (Fig. 2). Queries can be restricted to a single class of data (Table 1) and the search may be limited to just the symbols, names or identifiers (IDs) of entities (rather than searching the text of all report fields). In this way, the resulting hit-list contains only the data type of interest with far fewer false positive hits. Note that 'References', 'Human Disease Models', 'Gene Groups' or 'Orthologs' are not searchable via this tab as these data classes have their own dedicated tabs (Table 2).

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1.   Choose whether to limit the search to *D. melanogaster* reports (the default) or to include data from additional species by checking the 'include non-Dmel species' box (Fig. 2).

In addition to data on additional Drosophila species, FlyBase includes reports pertaining to genes from species outside of the Drosophila genus. This allows accommodation of data on heterologous expression of, for example, GAL4, GFP, lacZ, and a growing number of human genes. Leaving the 'include non-Dmel species' box unchecked will only retrieve D. melanogaster-specific reports in most cases – that is, the function of this checkbox within the 'Data Class' tab is more restrictive than in the 'Simple' tab (see PROTOCOL 1).

2.   Choose whether to restrict the search to 'ID/Symbol/Name' (the default) or to search for matches in the text of all report fields by clicking the 'All text' button.

The 'ID/Symbol/Name' option includes synonyms. Searching 'all text' results in a slower query and many more hits returned, but this option may be useful if searching for a text string within a particular data class that is not an ID/symbol/name.

3.   Select the desired data class from the drop-down menu.

Options are shown in Table 1.

4.   Enter a search term appropriate to that class.

Valid entries are ID/Symbol/Name (or their synonyms), if that is option has been selected. If the 'QuickSearch autocomplete' option is checked, valid entry symbols will start to appear when typing and can be clicked to populate the field. The search is case-insensitive and a wild card (*) can be added to match partial terms.

To restrict a symbol-based search to a particular species, prepend the symbol with the FlyBase 4-letter species prefix followed by a backslash (\). (Species abbreviations can be found by selecting the 'Abbreviations' link from the 'Species' menu on the navigation bar found on every FlyBase page.) E.g., enter 'Dvir\dpp' in the Data Class 'genes' to search specifically for the dpp gene of D. virilis. Note this is an override feature that works regardless of whether the 'include non-Dmel species' box is checked.

5.   Click the green 'Search' button or press 'enter'.

6.   View the hit-list of results matching the search criteria within the selected data class.

If the search criteria match a single hit, then its report page is displayed directly (see step 7).

7.   Click on an individual entry in a hit-list to view its report page.

FlyBase report pages display all of the data associated with that entity organized into separate sections. Reports contain many links to other related reports within and external to FlyBase to aid further navigation and exploration.

BASIC PROTOCOL 3: USING THE References TAB OF QuickSearch

FlyBase maintains a large and varied bibliography of *Drosophila*-specific publications (Marygold et al., 2013). It currently comprises >215,000 references, covering ~40 distinct publication types, with dates of publication extending from the 17th century to the present day. Approximately half of these references are primary research papers, identified through regular searches of the NCBI PubMed database. The remaining 50% include reviews, commentaries and errata (also identified in PubMed), as well as otherwise unpublished 'personal communications' from the *Drosophila* community and internal 'FlyBase analyses'. The entire FlyBase bibliography is searchable using the 'References' tab (Fig. 3).

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1.   Select a search field, either using one of the four defaults or selecting another (Fig. 3A).

       Each of the four search boxes can be toggled to search any of seven options, namely: Author(s), Year(s), Title/Abstract, Journal, Publication type, unique ID (PubMed ID, PubMed Central ID, Digital Object Identifier (DOI), FlyBase reference ID (FBrf)) or 'all fields'. The default set comprises the first four of these.

2.   Enter one or more terms that are appropriate to the selected field.

       It is recommended that 'QuickSearch autocomplete' is selected – valid entry terms will then start to appear in the 'Author(s)', 'Journal' and 'Publication type' fields when typing and can be clicked to populate the field. The search is case-insensitive and a wild card (*) can be added match partial terms. Boolean operators (AND, OR, BUT NOT) can be used within any field to search using specific combinations or exclusions. The 'Year(s)' field also accepts ranges and greater/less than notations (e.g. '2004–2008' or '>2001').

3.   If desired, repeat steps 1–2.

       The specificity of the query is proportional to the number of search terms and fields used.

4.   Click the green 'Search' button or press 'enter'.

5.   View the hit-list of matching references.

If the search criteria match a single hit, then its report page is displayed directly (step 6).

6. Click on an individual reference to view its Reference Report (Fig 3B).

These reports display the full citation data of the reference, together with links to the journal article, PubMed record and (where available) the full text at PubMed Central. In addition, Reference Reports include a summary of the data associated with the publication in FlyBase - all associated entities are linked to their own dedicated report.

BASIC PROTOCOL 4: USING THE Gene Groups TAB OF QuickSearch

The Gene Groups resource in FlyBase comprises sets of *D. melanogaster* genes or gene products that have been described in the published literature to share a set of characteristics (Attrill et al., 2016). Examples include evolutionarily related gene families (e.g. actins, odorant receptors), subunits of macromolecular complexes (e.g. spliceosome, ribosome), and gene products with a common molecular function (e.g. deubiquitinases, GTPase activating proteins). These collections are useful starting points for further analyses and explorations of associated data, both within FlyBase and in other species databases. The Gene Groups resource is an ongoing project with new groups being added each release. All current groups are searchable and browsable via the 'Gene Groups' tab (Fig. 4).

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1. Enter a name, symbol or other identifier corresponding to a group or a gene member (Fig 4A).

Valid entries are symbols or full names (or their synonyms) or FlyBase identifiers for the Gene Groups themselves or any member genes (e.g. entering sirtuin, SIRT, Sirt1, CG5216, FBgn0024291 or FBgg0000278 will all find the SIRTUIN LYSINE DEACETYLASES group). Valid names of Gene Groups will start to appear when typing and can be clicked to populate the field. The search is case-insensitive and a wild card (*) can be added to match partial terms. Alternatively, click on the 'browse' link to see a page listing all the Gene Groups compiled to date. Here, the top 'parent' groups are arranged in alphabetical order, with any subgroups ('children') displayed as nested subsets. These may either be browsed or searched for in-page. Clicking on a Gene Group opens the corresponding report page (step 4).

2. Click the green 'Search' button or press 'enter'.

3. View the hit-list of matching gene groups.

If the search criteria match a single hit, then its report page is displayed directly (step 4).

4. Click on an individual group to view its Gene Group Report (Fig 4B).

These reports include a brief description of the group, a summary of the references used to compile the group, the key Gene Ontology terms that typify the members, and links to related Gene Groups within FlyBase and at external sites (e.g. human gene families at the HUGO Gene Nomenclature Committee (HGNC) website (Gray et al., 2016)). The 'Members' table lists all members of the group with links to the corresponding Gene Report pages, together with convenient buttons to export members to other tools, such as the 'Orthologs' tool (see PROTOCOL 10), in order to facilitate further analyses.

BASIC PROTOCOL 5: USING THE Human Disease TAB OF QuickSearch

The Human Disease tab incorporates multiple options to access data relevant to human disease models in *D. melanogaster* (Millburn et al., 2016) (Fig. 5). First, it provides an entry point into the Disease Ontology (DO) (Kibbe et al., 2015), which is a rich controlled vocabulary (CV) of human diseases used to annotate mutant alleles and transgenic constructs that generate a model of human disease, or which ameliorate or exacerbate such a model. Second, the tab allows a direct search of Human Disease Model Reports, which have been produced by FlyBase for those specific diseases for which *Drosophila* serves as a model. These reports include a summary of the human disease and implicated gene(s), information about the *D. melanogaster* model, and links to related data in FlyBase and at external sites, such as Online Mendelian Inheritance in Man (OMIM) (Amberger et al., 2015). A specific disease may be modeled by expressing a human transgene in flies and/or by manipulating the fly ortholog(s), and this format integrates these data into a single report. Third, the tab also provides access to a browsable list of existing Human Disease Model Reports. (Note that the process of creating Human Disease Model Reports is ongoing and new reports are currently being added in each release.)

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

Search Human Disease Model Reports

1 In the upper section of the tab, choose to search by a disease term, a human gene or a *D. melanogaster* gene by selecting the appropriate option (Fig. 5A).

2 Enter a term or identifier corresponding to a disease or gene, as appropriate.

Valid 'disease' entries are disease names, synonyms or identifiers, including those used in the FlyBase Disease Model Reports, the DO, and OMIM (e.g.

entering 'ALS6', 'FUS-ALS', 'FBhh0000018', 'DOID:0060198', '608030' will all find the 'amyotrophic lateral sclerosis 6' Human Disease Model Report). Valid 'human gene' entries are the gene symbols and IDs used by the HGNC (Gray et al., 2015) and OMIM (e.g. 'FUS', 'HGNC:4010' or '137070'). Valid fly gene symbols are FlyBase gene symbols or identifiers (e.g. 'caz', 'CG3606' or 'FBgn0011571'). Valid entry terms will start to appear when typing and can be clicked to populate the field.

Alternatively, click on the 'browse' link at the foot of the tab to see a page listing all the Human Disease Model Reports produced to date. The list is alphabetical by major disease or mechanistic category; disease sub-types are organized as subsections under a major disease listing. A specific disease may be listed more than once, if it falls into more than one major category.

3   Click the green 'Search' button or press 'enter'.

4   View the hit-list of human disease models that match the search criteria.

Note that only those diseases and genes that are associated directly with a Human Disease Model Report are retrieved. That is, a search with a disease or gene that is not associated with a Drosophila model, or has not yet been curated as such, will not generate any hits.

5   Click on an individual disease to view its Human Disease Model Report.

These reports include an overview of the human disease, information on human-fly orthology relationships, the human/fly alleles annotated with DO terms, potentially useful stocks, and lists of relevant references and external websites.

Search the Disease Ontology

6   In the lower section of the tab, enter a disease term or identifier in the search box (Fig 5A).

Valid entries are DO terms/synonyms or DO identifiers (e.g. entering 'amyotrophic lateral sclerosis 6', 'ALS6' or 'DOID:0060198' will all find the 'amyotrophic lateral sclerosis type 6' DO term). Valid entry terms will start to appear when typing and can be clicked to populate the field. The search is case-insensitive and a wild card (*) can be added to match partial terms.

7   Click the green 'Search' button or press 'enter'.

8   View the hit-list of matching DO terms.

9   Click on a DO term to open its associated Term Report page (Fig. 5C), which contains additional details about the term and links to annotations.

These reports include a definition, synonyms and a view of the tree displaying the hierarchical relationships between any parent and child terms. Within the tree view, the total number of records annotated with each DO term (including its children) is indicated in green text.

10 If present, click the green 'Alleles' button to show a hit-list of all alleles (either classical or transgenic) annotated with the given DO term or any of its child terms, or click on one of the specific annotation statements in the table below to show a more refined hit-list.

The 'model of' prefix denotes experimental models of the given disease, whereas the 'ameliorates' and 'exacerbates' prefixes indicate alleles that modify the disease model phenotype. The 'DOES NOT' prefix is only used where an expected/predicted effect was not observed.

11 Alternatively, and if present, click the green 'Genes' button to show a hit-list of unique genes associated with those alleles annotated with the given DO term (or any of its child terms), or click on the 'Human Diseases' button to transfer to the Human Disease Model Report associated with that DO term.

12 When viewing a hit-list of alleles or genes associated with a DO term, click on any individual entry to view its report page and all associated DO annotations.

On both reports, the section titled 'Human Disease Model Data' displays the full set of DO annotations, as well a link to any relevant Human Disease Model Reports.

BASIC PROTOCOL 6: USING THE Gene Ontology (GO) TAB OF QuickSearch

The Gene Ontology (GO) is a structured CV used in FlyBase and many other biological databases to describe the attributes of gene products in terms of their molecular functions, associated biological processes, and cellular components (The Gene Ontology Consortium, 2015; Tweedie et al., 2009). The GO tab (Fig. 6) can be used to find terms of interest within the GO and identify *Drosophila* genes annotated with those terms. GO annotations may be based on experimental data in published research (e.g. mutational analysis or direct assays) or predicted from computational analysis (e.g. sequence similarity to a different protein of known function).

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1. Choose to search for terms across the entire GO (which is the default) or to restrict the search to the individual molecular function, biological process or cellular component ontologies by using the 'Data Field' drop-down menu (Fig. 6A).

2. Enter the search text or identifier.

Valid entries are GO terms/synonyms or GO identifiers (e.g. entering 'Wingless signaling pathway', 'Wg pathway' or 'GO:0016055' will all

find the 'Wnt signaling pathway' GO term). Valid GO terms that match the entered text, confined to the ontology aspect selected in step 1, will start to appear when typing and can be clicked to populate the field. The search is case-insensitive and a wild card (*) can be added to match partial terms.

3. Click the green 'Search' button or press 'enter'.

4. View the hit-list of matching GO terms.

5. Click on the desired GO term to open its associated Term Report page (Fig. 6B), which contains additional details about the term and links to annotations.

   These reports include a definition, synonyms and a view of the tree displaying the hierarchical relationships between any parent and child terms. Within the tree view, the total number of records annotated with each DO term (including its children) is indicated in green text. Link-outs to the corresponding term page at the QuickGO (Binns et al., 2009) *and AmiGO* (Carbon et al., 2009) *websites are also provided in the final section of this page.*

6. If present, click the green 'Genes' button to show a FlyBase hit-list of genes annotated with the given term or any of its child terms.

7. Click on an individual gene to view its Gene Report and all associated GO annotations.

   Molecular function annotations are summarized in the 'Families, Domains and Molecular Function' section of the report, and the full set of annotations is listed in the 'Gene Ontology (GO) section' below. In both cases, annotations are separated into those based on experimental evidence and those based only on predictions/assertions.

BASIC PROTOCOL 7: USING THE Expression TAB OF QuickSearch

This tab (Fig. 7) is used to search for *D. melanogaster* genes with a specified spatial and/or temporal expression pattern, as curated from low-throughput studies of the transcript or polypeptide described in individual research papers. Expression data may be searched by developmental stage, gross anatomical structure or subcellular location, or any combination of these. The allowed input terms are from the fly developmental stage and anatomy CVs (Costa et al., 2013) and the cellular component aspect of the GO. The lower panel of the Expression tab allows access to different options for querying or browsing high-throughput RNA-Seq expression data – click the 'RNA-Seq overview page' link in that section for further details.

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1.   Enter a term into the 'Stage', 'Tissue' or 'Cell Loc.' (subcellular location) field (Fig. 7A).

        Valid entries are CV terms or identifiers from the fly developmental stage, fly anatomy or the GO cellular component CVs, respectively. Valid CV terms that match the entered text will start to appear when typing and can be clicked to populate the field. Note that the autocomplete function only lists terms for which annotations exist. The search is case-insensitive; wild cards (*) are not accepted.

2.   If desired, click the 'refine search by adding qualifier terms' text and enter a field-appropriate 'qualifier' term in the box that appears below the main search box.

        For example, the developmental stage can be qualified by specifying gender, while the anatomy and subcellular fields can be qualified by a term adding detail about spatial distribution (e.g., 'dorsal') or signal intensity (e.g. 'faint'). Valid qualifier terms that match the entered text will start to appear when typing and can be clicked to populate the field; only those qualifiers associated with the CV term chosen in step 1 will appear. The search is case-insensitive; wild cards (*) are not accepted.

3.   If desired, repeat steps 1–2 for the other search field(s).

        The autocomplete function is coordinated between the terms entered in the primary search boxes; only terms that have been curated in a temporospatial statement together with those already in a filled search box will be offered.

        It is not necessary to fill all three primary search boxes; two or three boxes may be completed in any combination in order to add additional constraints to a query.

4.   Click the green 'Search' button or press 'enter'.

5.   View the hit-list of genes that exhibit a transcript or protein expression pattern consistent with the search criteria.

        Instead of viewing matching genes, the results corresponding to enhancer-trap insertions or transgenic constructs for the same query can be accessed by clicking on the 'Insertions', 'Recombinant Constructs' or 'Alleles' buttons at the top of the hit-list page. The 'Alleles' button brings both insertions and recombinant constructs into a common hit-list and is the most straightforward option when looking for GAL4 drivers associated with a specific expression pattern.

6. Click on an individual gene to view its Gene Report. Detailed expression information associated with the transcript, polypeptide or deduced from a reporter are included within the 'Expression Data' section (Fig. 7B).

Expression data derived from high-throughput studies and link-outs to relevant external sites are also displayed in this section of the Gene Report.

BASIC PROTOCOL 8: USING THE Phenotype TAB OF QuickSearch

This tab (Fig. 8) can be used to search for alleles (both classical and transgenic) associated with a specific class of phenotype, such as 'lethal', 'sterile' or 'behavior defective', or for those that affect a certain anatomical structure (Drysdale, 2001). These searches may be conducted in combination in order to retrieve alleles that match both search criteria. There is also the option of restricting each search to a particular developmental stage. The allowed input terms are from the CVs used by FlyBase to describe phenotypic class, anatomy and developmental stage (Costa et al., 2013; Osumi-Sutherland et al., 2013).

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1. Input a term in the 'Phenotypic class' or the 'Tissue/cell affected' fields (Fig. 8A).

Valid entries are CV terms or identifiers from the phenotypic class or fly anatomy CVs.Valid CV terms that match the entered text will start to appear when typing and can be clicked to populate the field. Note that the autocomplete function only lists terms for which annotations exist. The search is case-insensitive; wild cards (*) are not accepted.

2. If desired, fill in one or both 'refinement' boxes.

These can be used to specify a stage of development or to specify qualifiers such as 'somatic clone', 'recessive' or 'heat sensitive. Valid qualifier terms that match the entered text will start to appear when typing and can be clicked to populate the field; only those qualifiers associated with the CV term chosen in step 1 will appear. The search is case-insensitive; wild cards (*) are not accepted.

3. If desired, repeat steps 1–2 for the other search field.

It is not necessary to fill in both the 'Phenotypic class' and the 'Tissue/cell affected' but doing so will add an additional constraint to a query.

Note the autocomplete function is not coordinated between the 'Phenotypic class' and the 'Tissue/cell affected' fields.

4. Click the green 'Search' button or press 'enter'.

5. View the hit-list of alleles associated with phenotype(s) matching the search criteria.

6. Click on an individual allele to view its Allele Report. Detailed phenotypic information is included within the 'Phenotypic Data' section (Fig. 8B).

CV terms associated with the given allele are shown in the 'Phenotypic Class' and 'Phenotype Manifest In' subsections, together with any qualifiers and accessory or trans-acting alleles. The 'Detailed Description' subsection includes a 'free text' description of the phenotype with additional clarifications or qualifications that are not captured adequately by CV terms alone.

BASIC PROTOCOL 9: USING THE Protein Domains TAB OF QuickSearch

This tab (Fig. 9) is used to generate lists of *D. melanogaster* genes whose protein products share a specified domain or functional site, as determined by InterPro (Mitchell et al., 2015). InterPro integrates protein signatures from several individual sources (including Pfam, PROSITE, and SMART) and then associates InterPro entries with protein products in the UniProtKB collection (The UniProt Consortium, 2015). UniProtKB IDs are linked to gene records in FlyBase, thereby providing the connection between InterPro protein features and *D. melanogaster* genes.

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1. Input a search term or an identifier (Fig. 9A).

Valid entries are InterPro identifiers or signatures, which include protein domains, families, repeats, or sites (e.g. 'SH3 domain' or 'IPR001452'; 'WD40 repeat' or 'IPR001680'). (A link is provided on the tab to the InterPro documentation describing the different entry types.) Valid entry terms/IDs will start to appear when typing and can be clicked to populate the field. The search is case-insensitive and a wild card (*) can be added to match partial terms.

2. Click the green 'Search' button or press 'enter'.

3. View the hit-list of genes whose product(s) are associated with the specified InterPro term.

4. Click on an individual gene to view its Gene Report. Protein domain information from InterPro is displayed within the 'Families, Domains and Molecular Function' section (Fig. 9B) and within the 'Polypeptide Data' subsection under the 'Gene Models and Products' section (not shown).

Click on an InterPro term to transfer to its report on the InterPro site, which contains additional information on the domain/family etc.

Click on the GBrowse link in the Gene Report and choose to display the 'Protein domains (PFAM)' track to view isoform-specific Pfam domains, which indicates how these domains align with the intron/exon structure of the gene.

BASIC PROTOCOL 10: USING THE Orthologs TAB OF QuickSearch

This tab (Fig. 10) can be used to quickly search for orthologs of *D. melanogaster*, human or other model organism genes, as provided by the DRSC Integrative Ortholog Prediction Tool (DIOPT) (Hu et al., 2011) or OrthoDB (Kriventseva et al., 2015). The DIOPT dataset integrates ortholog predictions for several model organisms originating from many different algorithms based on sequence homology, phylogenetic trees, and functional similarity. The OrthoDB dataset (as implemented in FlyBase) comprises ~60 species, biased towards those that are closely related to *D. melanogaster*, such as other *Drosophila* species and other diptera.

Necessary Resources

Hardware—Computer or other device with access to the internet

Software—An up-to-date web browser, such as Firefox, Chrome or Safari, with JavaScript and cookies enabled.

1.   Select the input species from the 'Species' drop-down menu (Fig. 10A).

The current input species are: H. sapiens (human), R. norvegicus (Norway rat), M. musculus (house mouse), X. tropicalis (Western clawed frog), D. rerio (Zebrafish), D. melanogaster (Fruit fly), C. elegans (Nematode, roundworm), S. cerevisiae (Brewer's yeast) and S. pombe (Fission yeast).

2.   Enter one or more gene symbols/identifiers in the adjacent 'Gene(s)' box.

Valid entries are gene symbols/IDs approved by the relevant species database, or NCBI gene IDs. OMIM and Ensembl gene IDs are also accepted for human genes. Valid entry terms/IDs will start to appear when typing and can be clicked to populate the field. Symbol-based searches are case-sensitive; wild cards (*) are not accepted. Multiple entries must be separated a space character.

3.   Select one or more output species using the check-boxes.

Where the input species is D. melanogaster, there is a choice between searching the DIOPT or OrthoDB datasets. For DIOPT-based searches, the choice of output species is identical to the input species (see step 1); for OrthoDB-based searches, the output species are arranged into four distinct groups: 'Drosophila species', 'non-Drosophila Dipterans', 'non-Dipteran Insects', 'non-Insect Arthropods' and 'non-Arthropod Metazoa'.

4. Click the green 'Search' button or press 'enter'.

5. View the results table (Fig. 10B).

For DIOPT-based searches, the 'score' is expressed as the number of algorithms supporting a given ortholog call compared to the total number of relevant algorithms for that call; the explicit list of supporting algorithms is also shown, The 'Best score' and 'Best Rev Score' columns indicate whether the given ortholog has the highest score for the query gene, and whether the reciprocal relationship is also true. The results page for OrthoDB-based searches is similar, except that the DIOPT-specific columns are absent and the OrthoDB 'orthology group' to which the species belongs is shown on the right of each species line.

6. If desired, click on '+' symbol in the 'Best Rev Score' column to perform a new orthology search with the given gene as the input.

Conducting a full de novo reciprocal search gives a more complete view of complicated orthology relationships.

7. If desired, click on the '+' symbol in the 'Align' column to view a pairwise alignment of the amino acid sequences.

This links to an alignment page on the DIOPT website, which also tabulates the domains present in the two proteins and presents additional information on the individual orthology algorithm scores.

8. If desired, click on a link in the 'Ortholog Gene Reports' or 'Transgene in Fly' columns to view a corresponding gene report.

Links are provided to report pages at model organism databases and NCBI; for human genes, additional links are provided to Ensembl and OMIM. Links are also given to FlyBase Gene Reports for non-Drosophila genes that have been expressed transgenically in flies.

9. If desired, download the visible results table as a TSV file.

Click on the 'Save results as tsv file' text at the top of the table to download all the visible results to a file in tab-separated value format.

GUIDELINES FOR UNDERSTANDING RESULTS

Results obtained using QuickSearch are an excellent starting place for further exploration. Nonetheless, there are some factors to bear in mind when interpreting these results, as outlined below.

Broad searches, such as those using the 'Simple' tab or an 'all text' search within the 'Data Class' tab, can often return undesired or false positive hits, while very restricted searches may result in relevant data being omitted from a hit-list. Often the most effective strategy is to begin a search with inclusive criteria (e.g. using the 'Simple' tab), and then to narrow

these criteria (e.g. searching within a specific Data Class or using a more specialized tab) once the query is better defined and the underlying data are better understood.

The contextual reason for an item being returned in a search may not be obvious from the initial hit-list. Therefore, before drawing any important conclusions, additional data associated with any hit of interest should be examined by clicking through to its report page in FlyBase. Additional context/clarification may be obtained by consulting the original source, which is facilitated by the provision of clear data attributions and links to references in FlyBase reports.

The FlyBase website is updated regularly (see the 'Release Schedule' link in the 'About' menu of the navigation bar), which means that the results of any QuickSearch query may change between releases. It is therefore important to note the FlyBase release number being used for the query, which is present in the header and footer of every FlyBase page. For example, the screenshots herein were taken from release FB2016_05.

COMMENTARY

Background Information

The QuickSearch tool allows rapid querying across most FlyBase data classes and major data annotations through a concise interface. It is both easily found (in the center of the FlyBase homepage) and intuitive to use, factors which likely explain it accounting for 40% of all tool usage within FlyBase. However, the input parameters and output display options of QuickSearch are relatively limited and certain types of data, principally genomic data, are not accessible via this tool. FlyBase provides an extensive set of additional tools to accommodate these use cases (Marygold et al., 2016; McQuilton et al., 2012; St Pierre et al., 2014). For example, QueryBuilder can be used to construct multi-leg queries that search on a field-by-field basis, while three different RNA-seq tools offer the ability to search modENCODE RNA-seq data based on expression patterns or region-specific expression levels. Genomic data may be queried via BLAST, CytoSearch or Feature Mapper. These tools, together with documentation on their use, are accessible via the 'Tools' menu of the navigation bar that is present on every FlyBase webpage.

Critical Parameters

First, it is important to use a QuickSearch tab that is suitable for the query. Most tabs search only a specific class of data or annotation and so will result in an unproductive search if a term outside that class is entered (Table 1). If in doubt, or if a broad search is intended, then use the 'Simple' tab, though note that orthology, image and genomic data are outside the scope of this tab. (In these cases, instead use the Orthologs QuickSearch tab, the ImageBrowse tool, or a suitable genomic query tool (see Background Information), respectively.)

Most QuickSearch tabs require entry of a 'valid' term or ID, or doing so will at least result in a more accurate search. 'Valid' in this sense means the 'official' and current symbol, name or ID of the entity (gene symbol, protein domain, GO term etc.) in the source database. The source for *Drosophila* genetic symbols is FlyBase, while different external

databases are the ultimate sources of other symbols or terms (e.g. InterPro for protein domains, the GO Consortium for GO terms, or the HGNC for human gene symbols). To aid selection of a valid term, an auto-completion feature is active/available wherever a search term needs to be a CV term (the GO, Human Disease, Expression and Phenotype tabs), a genetic symbol (Data Class, Gene Groups, Human Disease, Orthologs) or other defined term/ID (Protein Domains and certain fields within the References tab). The auto-complete feature presents a list of matching terms that becomes more restricted as more characters are added to the search string. Note that 'coordinated auto-completion' occurs for the Expression and Phenotype tabs in which several CV terms may be used simultaneously - when a term is entered in one field, the auto-completion options in the associated field(s) are constrained to suggest only terms that actually occur in combination with the first term.

An asterisk (*) can be used as a wild card in most QuickSearch tabs – exceptions are Expression, Phenotype and Orthologs tabs. Wild cards can be added to the beginning, end or the middle of a search term to broaden the query. Flanking a search term with wild cards will return all phrases containing the search term.

While the majority of FlyBase data pertains to *D. melanogaster*, reports also exist for genetic entities from other *Drosophila* species as well as non-*Drosophila* species – the latter are required to accommodate data on heterologous expression, where a gene from a 'foreign' species has been expressed transgenically in *D. melanogaster*. Two QuickSearch tabs, Simple and Data Class, provide an option to restrict or expand the breadth of species data being searched.

The Data Class tab offers the choice of searching only the ID/Symbol/Name fields (which includes synonyms) of reports versus the entire text of all fields. The default behavior is the former option as this is the most common usage. Selecting the 'all text' option will result in a slower query with many more hits returned, but is useful if searching for a text string within a particular data class that is not an ID/symbol/name.

Additional help documentation can be accessed online by clicking on the '?' icon above the QuickSearch interface or within any individual tab.

Troubleshooting

Common problems and possible solutions are shown in Table 3. Users are encouraged to use the 'Contact FlyBase' link present at the foot of every FlyBase webpage to request assistance, report a bug or request a new feature.

Acknowledgments

Gillian Millburn, Alix Rey, Nicole Staudt, Jose-Maria Urbano (FlyBase-Cambridge); Thomas Kaufman, Joshua Goodman, Gary Grumbling, Victor Strelets, Jim Thurmond (FlyBase-Indiana); Richard Cripps, Maggie Werner-Washburne, Phillip Baker (FlyBase-New Mexico).

Literature Cited

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015; 43:D789–798. [PubMed: 25428349]

Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ. the FlyBase Consortium. FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. Nucleic Acids Res. 2016; 44:D786–792. [PubMed: 26467478]

Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics. 2009; 25:3045–3046. [PubMed: 19744993]

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S. the AmiGO Hub and the Web Presence Working Group. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009; 25:288–289. [PubMed: 19033274]

Costa M, Reeve S, Grumbling G, Osumi-Sutherland D. The *Drosophila* anatomy ontology. J Biomed Semantics. 2013; 4:32. [PubMed: 24139062]

Drysdale R. Phenotypic data in FlyBase. Brief Bioinform. 2001; 2:68–80. [PubMed: 11465064]

Gray KA, Seal RL, Tweedie S, Wright MW, Bruford EA. A review of the new HGNC gene family resource. Hum Genomics. 2016; 10:6. [PubMed: 26842383]

Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. 2015; 43:D1079–1085. [PubMed: 25361968]

Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, Mohr SE. An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics. 2011; 12:357. [PubMed: 21880147]

Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, Parkinson H, Schriml LM. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. Nucleic Acids Res. 2015; 43:D1071–1078. [PubMed: 25348409]

Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 2015; 43:D250–256. [PubMed: 25428351]

Marygold, SJ.; Crosby, MA.; Goodman, JL. and the FlyBase Consortium. Using FlyBase, a Database of *Drosophila* Genes and Genomes. In: Dahmann, C., editor. Drosophila: Methods and Protocols. 2. Vol. 1478. Springer; New York: 2016. p. 372

Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ. the FlyBase Consortium. FlyBase: improvements to the bibliography. Nucleic Acids Res. 2013; 41:D751–757. [PubMed: 23125371]

McQuilton P, St Pierre SE, Thurmond J. the FlyBase Consortium. FlyBase 101--the basics of navigating FlyBase. Nucleic Acids Res. 2012; 40:D706–714. [PubMed: 22127867]

Millburn GH, Crosby MA, Gramates LS, Tweedie S. the FlyBase Consortium. FlyBase portals to human disease research using *Drosophila* models. Dis Model Mech. 2016; 9:245–252. [PubMed: 26935103]

Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJ, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2015; 43:D213–221. [PubMed: 25428371]

Osumi-Sutherland D, Marygold SJ, Millburn GH, McQuilton PA, Ponting L, Stefancsik R, Falls K, Brown NH, Gkoutos GV. The *Drosophila* phenotype ontology. J Biomed Semantics. 2013; 4:30. [PubMed: 24138933]

St Pierre SE, Ponting L, Stefancsik R, McQuilton P. the FlyBase Consortium. FlyBase 102--advanced approaches to interrogating FlyBase. Nucleic Acids Res. 2014; 42:D780–788. [PubMed: 24234449]

The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43:D1049–1056. [PubMed: 25428369]

The UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015; 43:D204–212. [PubMed: 25348405]

Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-Sutherland D, Schroeder A, Seal R, Zhang H. the FlyBase Consortium. FlyBase: enhancing *Drosophila* Gene Ontology annotations. Nucleic Acids Res. 2009; 37:D555–559. [PubMed: 18948289]

A

QuickSearch

Human Disease	Expression	GO	Phenotype	References
Simple	Orthologs	Protein Domains	Gene Groups	Data Class

Species: ☐ include non-Dmel species   Search

Enter text: cadherin   ⑦

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

B

Your query phrase "cadherin" returned 1108 matches

Hits	Data Class	Hits	Data Class
76	Genes	138	Alleles
24	Stocks	29	Insertions
702	References	0	Clones
18	Transcripts	38	Polypeptides
0	Aberrations	0	Balancers
0	Natural Transposons	59	Transgenic Constructs
0	Sequence Features	0	Large Dataset Metadata
0	Cell Lines	14	Physical Interactions
0	Strains	5	Gene Groups
4	Human Disease		

Edit your search query phrase

Query phrase: cadherin

Resubmit

Figure 1.

(A) The 'Simple' tab of QuickSearch, with the text 'cadherin' entered as the search term.

(B) The initial results page resulting from running that query.
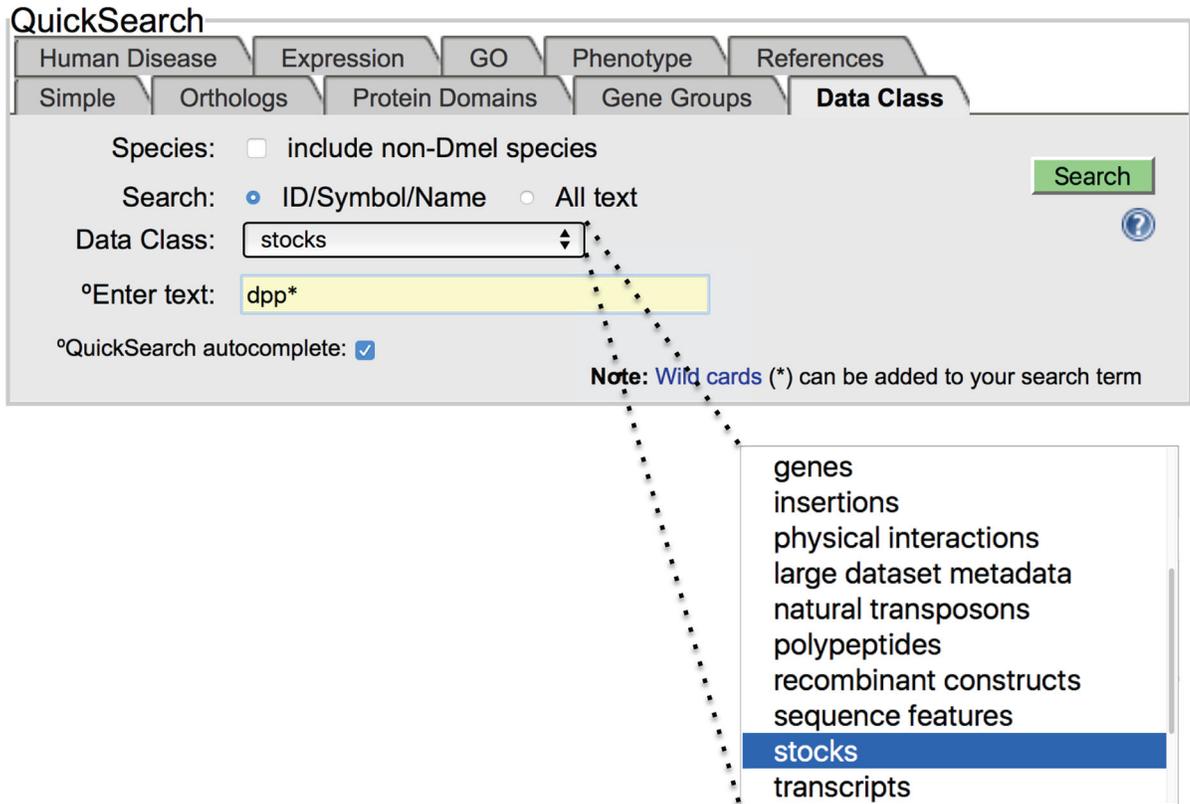
Figure 2.
The 'Data Class' tab of QuickSearch, with 'stocks' selected from the drop-down menu, 'ID/Symbol/Name' selected as the search mode, and 'dpp*' entered as the search term. The expanded box shows a subset of the options available from the Data Class menu.

A

QuickSearch

| Human Disease | Expression | GO | Phenotype | **References** |
| Simple | Orthologs | Protein Domains | Gene Groups | Data Class |

°Author(s) ⬍ : Perrimon AND Kwon

Year(s) ⬍ : >2001

Title/Abstract ⬍ : stem cells

°Journal ⬍ : Dev. Cell

°QuickSearch autocomplete: ☑

Search ❓

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

B

Reference	
Citation	Kwon, Y., Song, W., Droujinine, I.A., Hu, Y., Asara, J.M., Perrimon, N. (2015). Systemic Organ Wasting Induced by Localized Expression of the Secreted Insulin/IGF Antagonist ImpL2. Dev. Cell **33(1)**: 36--46. (Export to RIS)
FlyBase ID	FBrf0228030
Publication Type	Research paper
Abstract	Organ wasting, related to changes in nutrition and metabolic activity of cells and tissues, is observed under conditions of starvation and in the context of diseases, including cancers. We have developed a model for organ wasting in adult Drosophila, whereby overproliferation induced by activation of Yorkie, the Yap1 oncogene ortholog, in intestinal stem cells leads to wasting of the ovary, fat body, and muscle. These organ-wasting phenotypes are associated with a reduction in systemic insulin/IGF signaling due to increased expression of the secreted insulin/IGF antagonist ImpL2 from the overproliferating gut. Strikingly, expression of rate-limiting glycolytic enzymes and central components of the insulin/IGF pathway is upregulated with activation of Yorkie in the gut, which may provide a mechanism for this overproliferating tissue to evade the effect of ImpL2. Altogether, our study provides insights into the mechanisms underlying organ-wasting phenotypes in Drosophila and how overproliferating tissues adapt to global changes in metabolism.
PubMed ID	25850671
PubMed Central ID	PMC4437243 (PMC) (EuropePMC)
DOI	10.1016/j.devcel.2015.02.012
Related Publication(s)	
Note	Cancer metabolism: A waste of insulin interference. Wagner and Petruzzelli, 2015, Nature 521(7553): 430--431 [FBrf0228802]

⊞ Recent Updates
⊞ Associated Information
⊞ Other Information
⊞ Parent Publication
⊟ Data from Reference
⊞ Alleles (12)
⊞ Constructs (10)
⊞ Genes (27)
⊞ Insertions (1)

Figure 3.
(A) The 'References' tab of QuickSearch, with several criteria entered as search terms. (B) The Reference Report for one of the resulting hits.

Figure 4.
(A) The 'Gene Groups' tab of QuickSearch, with the text 'Sirtuin' entered as the search term. Note the link to 'browse all Gene Groups'. (B) The Gene Group Report for SIRTUIN LYSINE DEACETYLASES.

Figure 5.
(A) The 'Human Disease' tab of QuickSearch, with the text 'Alexander' entered as the search term of the lower 'Search Disease Ontology' box. (B) The Term Report for 'Alexander disease' showing the link to the associated Human Disease Model Report and to hit-lists of alleles/genes annotated with the term.

Figure 6.

(A) The 'GO' tab of QuickSearch, with the text 'axonogenesis' entered as the search term.

(B) The Term Report for 'axonogenesis' showing the link to a hit-list of genes annotated with the term.

Figure 7.
(A) The 'Expression' tab of QuickSearch, with several criteria entered as search terms. (B) The 'Expression Data' section of the Gene Report for the *AstA* gene (one of the resulting hits). The 'Polypeptide Expression' subsection is open.

Figure 8.
(A) The 'Phenotype' tab of QuickSearch, with several criteria entered as search terms within the upper 'Phenotypic class' section. (B) The 'Phenotypic Data' section of the Allele Report for the $polo^{11}$ allele (one of the resulting hits).

A

QuickSearch

| Human Disease | Expression | GO | Phenotype | References |
| Simple | Orthologs | **Protein Domains** | Gene Groups | Data Class |

Search using InterPro IDs or signatures, including protein domains, families, repeats, and sites.

[Search]

❓

Enter text: SH3 domain

See the InterPro FAQ page.

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

B

Families, Domains and Molecular Function	
Gene Group Membership (FlyBase)	NON-RECEPTOR TYROSINE KINASES
Protein Family (UniProt, Sequence Similarities)	Belongs to the protein kinase superfamily. Tyr protein kinase family. ABL subfamily. (P00522)
Protein Domains/Motifs	**UniProt (Sequence Similarities)** Contains 1 protein kinase domain. (P00522) ; Contains 1 SH2 domain. (P00522) ; Contains 1 SH3 domain. (P00522) **InterPro** Protein kinase domain; SH2 domain; Serine-threonine/tyrosine-protein kinase catalytic domain; SH3 domain; Tyrosine-protein kinase, active site; Protein kinase-like domain; F-actin binding; Protein kinase, ATP binding site; Tyrosine-protein kinase, catalytic domain; Tyrosine-protein kinase ABL1/transforming protein Abl
Molecular Function (see GO section for details)	**Experimental Evidence** non-membrane spanning protein tyrosine kinase activity; protein binding; protein tyrosine kinase activity **Predictions/Assertions** ATP binding; protein tyrosine kinase activity

Figure 9.
(A) The 'Protein Domains' tab of QuickSearch, with the text 'SH3 domain' entered as the search term. (B) The 'Families, Domains and Molecular Function' section of the Gene Report for the *Abl* gene (one of the resulting hits).

Figure 10.
(A) The 'Orthologs' tab of QuickSearch, searching for orthologs of the *Psn* gene of *D. melanogaster* in all organisms in the DIOPT set. (B) The results table from running that query.

Table 1

Data classes in FlyBase and how they can be queried using QuickSearch.

Data Class	Searchable via these QuickSearch tabs
Genetic/Genomic Data	
Genes	Simple, Data Class: genes
Transcripts	Simple, Data Class: transcripts
Polypeptides	Simple, Data Class: polypeptides
Alleles	Simple, Data Class: alleles
Aberrations	Simple, Data Class: aberrations
Balancers	Simple, Data Class: balancers
Recombinant Constructs	Simple, Data Class: recombinant constructs
Insertions	Simple, Data Class: insertions
Natural Transposons	Simple, Data Class: natural transposons
Sequence Features	Simple, Data Class: sequence features
Physical Interactions	Simple, Data Class: physical interactions
Orthologs	Orthologs
Integrated Data	
Large Dataset Metadata	Simple, Data Class: large dataset metadata
Gene Groups	Simple, Gene Groups
Human Disease Models	Simple, Human Disease
Reagents	
Stocks	Simple, Data Class: stocks
Strains	Simple
Cell Lines	Simple, Data Class: cell lines
Clones	Simple, Data Class: clones
Other	
References	Simple, References
Images	*Not searchable in this interface - use ImageBrowse*

Table 2

Summary of QuickSearch tabs and their functionality.

QuickSearch tab	Scope/Type of search	Primary entity searched/retrieved
Simple	Most data classes	Most data classes
Data Class	Single data class (user-specified)	*Same data class as searched*
References	Single data class	References
Gene Groups	Single data class	Gene Groups
Human Disease	Single data class/CV	Human Disease Models/DO terms
GO (Gene Ontology)	CV	GO terms
Expression	Multiple annotation types	Genes
Phenotype	Multiple annotation types	Alleles
Protein Domains	Single annotation type	Genes
Orthologs	Single data class	Genes

Table 3

Troubleshooting common QuickSearch problems

Problem	Possible Cause	Solution
No results	Search term is invalid; inappropriate tab has been used	Use auto-complete and select a search term from drop-drop list; use 'Simple' tab
Too few results	Search criteria are too restrictive	Expand criteria by adding a wild card, including non-Dmel species, or searching 'all text' (options differ by tab). Try using the 'Simple' tab.
Too many results	Search criteria are too broad	Restrict criteria by removing a wild card, excluding non-Dmel species, or searching 'ID/Symbol/Name' (options differ by tab). Try using the 'Data Class' tab.
Missing data	Data have not yet been curated; data have been curated are not yet shown on website	Use the 'Fast-Track Your Paper' tool to check/request data curation; use the 'Contact FlyBase' link to check status.