Related sets of genes or gene products, such as gene families (e.g. actins), subunits of protein complexes (e.g. proteasome subunits) or other functional groupings (e.g. Ubiquitin E3 ligases), are frequently described in the literature. Moreover, researchers often wish to obtain a list of related genes from a database and be able to easily download/query data associated with them. We wish to improve the representation of such 'gene groups' in FlyBase through the provision of new ‘Gene Group Reports’. These Reports will tabulate the membership of each group, clearly attributed to source references, together with buttons to export the member genes to FlyBase tools to facilitate further analyses. Gene Group Reports will also include a brief description of the group, links to parent/child groups within FlyBase, and links to orthologous groups available at external sites such as HGNC, WormBase and TAIR. Through doing this work and adopting a ‘group’ approach to curation, we have also been able to improve the consistency and accuracy of Gene Ontology annotations and nomenclature of related genes.

The General Information panel displays basic data on the Gene Group, including the number of members in the group and the date the group was last reviewed by FlyBase.

Any ‘edge cases’ where the inclusion/exclusion of a specific gene is unclear or debated are described in the Notes on Membership section.

Key Gene Ontology (GO) terms are those terms that largely define the group and which are used to annotate most/all member genes.

Data downloads and further analyses of the Gene Group are available via the 'Export to Hitlist' and 'Export to Batch Download' buttons.

Hyperlinks to equivalent Gene Groups at other databases, such as the HGNC, WormBase or TAIR, are given in the Orthologous Group(s) section.

Curation strategy

1. Identify a gene group of interest
2. Systematic search for relevant literature
3. Compile list of member genes
4. Assign ‘meta’ data to group
   - Assign group name and symbol
   - Add any synonyms
   - Write description
5. Review gene-level data of individual genes

Flag that reference describes a gene group

Add fast track your paper

Review gene-level data of individual genes - see next panel

Added benefits of ‘group curation’

Compiling sets of related genes provides the ideal opportunity to review the consistency and accuracy of gene-level data in FlyBase. The following areas are routinely examined:

1. Gene Ontology annotations
   - Add missing terms (~4,500 added to date)
   - Update existing terms (e.g. if the available ontology terms or underlying data have changed)
   - Remove any redundant or incorrect terms

2. Gene nomenclature
   - Check compliance with guidelines
   - Check accuracy and that it reflects usage
   - Add full-name if missing
   - Improve consistency within group where feasible (e.g. related genes that are named for their molecular function/orthology should have the same prefix and format)

3. Un-annotated genes
   - Check if there are any genes not mapped to the sequence that have GO annotations and/or gene symbols that suggest they are members of the current group
   - Merge FlyBase Gene Reports as appropriate

(These improvements mean that GO- or symbol-based searches to identify groups of related genes are more successful after review, but such searches will always suffer from false positives/negatives – thus the need for the Gene Group Reports.)

The Source Material section serves to: (i) clearly indicate that the group has been curated manually by FlyBase curators; and (ii) summarizes the primary references used to compile the group.

When relevant, parent super-groups or child sub-groups within FlyBase are shown as hyperlinks in the Related Gene Groups panel.

The Members table lists the members of the group and includes:
- Gene Symbol (linked to Gene Report)
- Gene Name
- Annotation ID (linked to GBrowse)
- Source Material (i.e. the reference(s) that state that this specific gene is a member of this particular group)

All citations are given in full in the Reference section and each hyperlinked to its Reference Report

Hyperlinks to all relevant and important external websites provided in the Other resources panel.