

January 1995 D.G. Gilbert

*New avenues of access to Drosophila data for biologists. Since 1992, I have been working toward goals of making Drosophila data available more easily and with greater sophistication and higher quality to biologists. Aspects of these goals have been met through the development and improvements to FlyBase Internet services at IUBio archive. Further aspects of this have been met through development of Internet client software with new features of data access and presentation that current network clients lack. Some of the key points in this development include: * searching by keywords through the various parts of Drosophila data for quick selection of portions of interest to the user. Additional complex queries and interactions between user/client and data server, made possible using gopher, and recently html, dialogs. Browsing and searching genetic data through graphical maps. Searches based on selected fields within records to allow better focussing on topics. * access to several different views of data which provide biologists a richer choice, enabling them to obtain data in publication-quality formats for best reading, or in computer-readable formats for easiest use with software; in formats with network links to other relevant data for easier selection of common threads of interest; in formats suitable for simple as well as sophisticated client software; in graphical as well as textual formats. * access to these services both over Internet and from non-networked computers, using local disks and eventually CD-ROM. This allows for quick and full access to data services to those lacking good Internet connections, as well as direct access to data on central servers with more timely updates. Many of the details and full implementation of these points remain to be fleshed out. However the underlying foundation has been built and verified as workable. Here are details of the above points for examination. Many of these new services from the FlyBase @ IUBio server are under development still, and are located at a non-standard Internet port 71. If you prefer, you can find these by connecting to the standard gopher server/port here (flybase.bio.indiana.edu, port 70), and selecting items called Gopher door to FlyBase@IUBio (new features; in progress) and HTML door to FlyBase@IUBio (Mosaic; new features; in progress) GopherPup network client software. Several of these new features are only available with a new network client, GopherPup. Where possible, I've made equivalent or near-equivalent services available to existing gopher and html network clients. A preliminary release of GopherPup is available for Macintosh, MSWindows, Sun and SGI computer systems (other XWindows variants will follow). This release lacks needed documentation, and has several bugs. But it basically works. The corrections and documentation will be made available in coming months. Most of the effort for development of GopherPup has been done in the author's weekend and vacation time. It is hoped that biologists will gain new and interesting methods of access to Drosophila data using this new client. It is also flexible enough to*

work well with other genome information services. GopherPup has several unique strengths that make it potentially very useful for *Drosophila* and other genome information delivery for use by biologists. - Display documents in Microsoft's Rich Text Format (RTF) produced by popular wordprocessors. RTF not only is a widely-available standard, it has full support for fonts, styles, embedded images, and the formatting needed to display scientific, publication-quality documents. - Display graphics in a standard, widely-used vector & bitmap format, Apple Macintosh's PICT. This format provides high-resolution output, smaller/faster image transfer, and editable graphics. Biologists will find these features quite useful for putting genomic maps to use in their work. - Display documents in the network HyperText Markup Language (HTML), and in the popular graphic bitmap format, CompuServe's GIF, as well as plain text. - Easily link network resources into any of these documents. Documents created in a wordprocessor can be directly displayed and network links can be placed by simple drag-and-drop methods. - The hypertext method used is easily extensible to any new format which can be displayed. - Browse and search through information on your local computer system using the same easy methods used for remote systems. The search methods use the same query dialogs, and relies on a variant of FreeWAIS software which provides for complex and field-specified queries. - Supports all Gopher+ protocol features, including user dialogs for client-server interactions; simple and sophisticated multiple view handling; display of document dates, sizes, and abstracts. - Supports many HTTP/HTML protocol features (more later), and supports SMTP Mail protocol. - "One source fits all" -- this program is freely available in one source version that compiles on Macintosh, Microsoft Windows and X- Windows computer systems. This preliminary GopherPup software is available through IUBio Archive.

Anonymous FTP is the best way to obtain it: ftp to iubio.bio.indiana.edu ftp> cd util/dclap/apps/gopherpup/ ftp> mget GopherPup.* get all documents with this name
ftp> binary Get one or more of these programs compiled for different systems get
gopherpup-mac-68k.hqx Macintosh, 68000 processor get gopherpup-mac-ppc.hqx Macintosh,
PowerPC processor get gopherpup-sun-sparc-stat.Z Sun Sparcstation, static-linked Motif
libraries get gopherpup-sun-sparc-dyn.Z Sun Sparcstation, dynamic-linked libraries get
gopherpup-sgi-mips.Z SGI Irix, MIPS processor get gopherpup-mswin-32.zip Microsoft
Windows, 32-bit, Intel processor get gopherpup-mswin-16.zip Microsoft Windows, 16-bit, Intel
processor The Macintosh version is recommended as the most stable and bug free. The Sun version is
compiled in both static and dynamic library link forms. The static version is for those whose Sun
system lacks Motif XWindow libraries. I've found this program will run on both SunOS4 and Solaris 2
Sun systems. The 32-bit MSWindows version is more stable than the 16-bit version, but may require
32-bit Windows extensions not available on a standard version 3.1 of MS Windows. For the following
examples, use the server flybase.bio.indiana.edu, at Internet port 71, or port 80.

- Searching data for quick selections Example: *Drosophila* gene data Use any gopher client (including Mosaic and similar browsers) Open this folder at the server: FlyBase> Genes> Choose the "Genes-Search" item. See also the document "Genes-search-help" in that folder for query help. - Complex queries for such searches Example: *Drosophila* clone data For query dialogs that provide easy selection of both general searches, and searches of selected fields in the data, see the following items: Using a Gopher+ client, access this folder at server: FlyBase> Clones > Choose item "All clones search (gopher form)". Enter a relevant query for clone data, such as cytological minimum: 55C cytological maximum: 55D With the Class selector, you can restrict the search to just cosmids or just P1 or just YAC. Or using an HTML client, access this folder <<http://flybase.bio.indiana.edu/1/Flybase/clones/>> Choose the item "All clones search (html form)", and use equivalent complex searches. For a dialog that shows how to build a very complex query using multiple fields, joined with Boolean selectors, with nested levels (similar to NCBI Entrez query forms), try the following item. Using a Gopher+ client, open the folder: FlyBase> Genes> Choose item "Genes-search (complex query)" See also the document "Genes-search-help-complex" in that folder for complex query help. - Variety in output formats for various needs of the biologists and their client software Example: *Drosophila* gene data Use GopherPup preferably, or alternatively other Gopher+ clients or HTML clients. Open this folder at the server: FlyBase> Genes> Choose the "Genes-Search" item, search on the word "tubulin" (without quote marks) and select some of the resulting items, such as alphaTub84 and betaTub60D. From GopherPup and other Gopher+ clients, you will get a choice of view formats, including application/rtf, text/plain, text/html and text/x-ashburner-star-code. RTF, plain text and HTML versions include network links to relevant data, including GenBank, Swiss-Protein and Prosite databases, and FlyBase references. The references section includes network links to NCBI's Entrez Medline abstract databank. If you use other Gopher+ clients, you will likely have to configure a helper application to view the RTF and x-ashburner-star-code formats. If you use an HTML/HTTP client, such as Mosaic or Lynx, to <<http://flybase.bio.indiana.edu/1/Flybase/genes/>>, choose the item "Genes-search (HTML/Mosaic output)" and perform this same query. You will obtain results in HTML format, including the above mentioned links to other data. An important comparison is the display of say the betaTub60D gene in RTF and HTML formats. The former includes all the Greek symbols and allele superscripting normally found in *Drosophila*

scientific publications, and includes fuller formatting, than does the HTML format. Some other useful output formats that are easily deliverable with this means include postscript, computer-readable comma-separated-values, and refer format for references, various biosequence formats for sequence data.- Browsing and searching genetic data through graphical maps To browse through the genetic, clone and other data which can be located by genetic and cytological map positions, it is helpful to see a graphical display and pick genes or clones that are located in given positions. Using the GopherPup client, open this folder: FlyBase> Maps> Graphical maps (html & net-pict) Open one of the subset folders such as Gene data or Clone data. This example is still incompletely developed. Later versions will have a graphic main map to avoid the long list of items in these folders. Currently there are also no aberrations data maps. Open any of the maps in the subset folder. You will find in the top portion, names of genes or clones located above their map location. These names can be clicked on to bring up full descriptions. Below the map lines, you find other dialog buttons. The only active ones now are the button to select "Clones" from Genes, or "Genes" from Clones maps, and the Forward and Reverse selectors for browsing up and down the full map. Of these selectors the single left/right arrows move 1 unit, and double arrows, move by 10 units. Using a graphical HTML client such as Mosaic, find the same information at <<http://flybase.bio.indiana.edu/0/Flybase/maps/grmaps/main.html>> This comparison of GopherPup and HTML browsing of graphic map data shows one of the strengths of the GopherPup innovations. The netdoc/pict format that GopherPup uses provides for about 5 times faster access to the maps than the HTML image/gif format, it also allows researchers to print the maps at high resolution (without the GIF bitmap "jaggies" found with HTML browsers), and it provides map pictures that can be edited, a feature that researchers will find useful in preparing reports.- Access the same services from non-networked computers GopherPup has the ability to browse through data on local disks or CDROM disks. With helper applications built from FreeWAIS searching software, it can also query and search the same indexed data as is done on a remote server. This ability isn't quite ready for demonstration, but the basic features have been tested by the author.