

A Replicable Model Organism Information System

FlyBase next-generation

Don Gilbert, gilbertd@indiana.edu

May 2003





Portable FlyBase 1996-2002

FlyBase last-generation portable system structure (1996-2002):

- common Unices supported; source code included
- installs and runs in any Unix path
- all via Apache web server & CGI (no separate standalone servers beyond FTP)
- update nightly from main server via FTP mirror job; 'live-file-system' mirroring (not tarballs)
- `ftp://fbserver:FlyBase@ FlyBase.net/`





FlyBase old structure

flybase-server/

cgi-bin/ -- web CGI programs

bin-local -> OS specific folder

data -> ~ftp/flybase/data/ -- public data files (with FTP access)

indices/ -- data search index files

logs/ -- web logs

mirror/ -- FTP mirroring software

server/ -- public web pages and accessory data

.etc/ -- accessory data, miscellany

.srs/ -- SRS search engine

aberrations/ -- many subject class folders

anatomy/ genes/ maps/ refs/ seqs/ and others

source/ -- program source code

sun-sparc-solaris-bin/, sgi-irix-bin/ , linux-i86-elf-bin/ -- OS specific binaries

Install*, Readme -- installation documents





FlyBase next-generation

/bio/biodb/

rsync://FlyBase.net/biodb

- common/
 - java/ ; perl/ -- language packages
 - servers/ -- major programs (blast, dbms, internet servers)
 - systems/ -- OS binaries of programs, packages
- docs/
- logs/
- myorg/ -- template information system structure
- flybase/ -- implemented project structures
- eugenes/
- daphnia/





FlyBase next-generation

- segregate common infrastructure from project-specific parts
- want customer-choice per-package installations and updates
 - need to find/make package distribution management utility
 - include logic to update infrastructure from source sites
 - focus on 'rsync' now as main distribution tool
 - evaluating RPM, pacman, cluster-backup tools, grid packaging tools
 - CVS management of biodb structure, package info, configs but not main programs, data, binaries
- per-project packages should be flexible in structure, content
 - project needs to specify infrastructure packages
 - need security/authentication options; private and public sections
- retain daily mirror-ability of current server
 - retain 'live file system' mirror/replication mechanism
 - for distribution & update of active servers
 - for local clusters to manage high-volume traffic
 - issues with rdbms and other stand-alone server updates
 - need install/update script to allow path choices, auto-restart servers



FlyBase NG structure details

/bio/biodb common:

- *java*: axis, lsid, lucene, ogsa, xindice (others to move in)
 - *perl*: lsid, (others to move in)
 - *servers*:
 - apache, tomcat
 - berkeleydb, mysql, postgresql, srs
 - blast
 - ldap, mirror, rsync, wuftp
- hope to use 'plain vanilla' copies of these so updates are easy and customers can replace as desired
- customize per project and via configurations

/bio/biodb common:

- *source*: fbapache_1.3.26.tar.gz, mod_backhand-1.2.2.tar.gz, mod_layout-3.2.tar.gz, mod_throttle-312.tgz; postgresql-7.3.2.tar.gz, berkeleydb, mysql, rsync, blast, ...
- *system-local* -- common reference link to active system binaries
- *systems*: -- compiled binaries for common servers
 - apple-powerpc-darwin, intel-linux, sgi-irix, sun-sparc-solaris





MyOrganism project template

cgi-bin: -- web CGI programs

common -- symlink to common infrastructure

conf:

 apache.conf -- virtual host include for main httpd.conf

 apache.conf.local -- local host config (not mirrored)

 apache.conf.in -- path-independent ; other project configs here

data: -- public data (symlink to FTP folder)

dfs: -- project databases (configs, scripts, common symlinks)

etc: -- miscellany

indices:-- database indices (update often)

secure:-- secure, authenticated access data, web

tmp: -- temporaries

web: -- public web structure

webapps: -- web Servlet programs

rsync.exclude.local -- project mirroring configurations





FlyBase NG implementation

cgi-bin: -- web CGI programs

common -- symlink to common programs

conf: apache.conf, apache.conf.in, apache.conf.local, cvsweb.conf

data: aberrations: allied-data: docs: extdb: genes: images: maps:
news: nomenclature: refs: work:

dbs: blast: srs:

etc: cytodb: expdb: gmod-fb: gnomap: icons: insitus: jdata: jlib: kevin:
other: people: perlbio: pix: plib: prefs: sean: stockxgene: templates:
tomcat transmolmaps: transseq:

indices: blast: srs: postgres:

tmp:

web: aberrations: allied-data: alt-views: anatomy: annotfb: clones:
docs: fbservlet: gbrowse_fb genes: genome-projects: images:
index.html maps: people: pep: refs: robots.txt search: sequences:
stocks: transposons:

webapps: cvservlet: fbchado:

rsync.exclude.local





FlyBase NG issues

- Need soon for production FlyBase (summer 2003)
- Make work with gmod-web, other gmod tools
- What packaging distribution management?
- Mechanics for public/private mixtures

In progress at <rsync://FlyBase.net/biodb/>

Web docs <http://FlyBase.net/flybase-ng/>

