

FlyBase bulk data files

Survey Questions Feb 2019

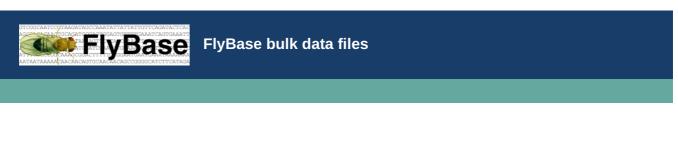
Introduction

FlyBase provides many (>100) downloadable bulk files, each containing defined subsets of data in different formats. These files are accessed via the 'Downloads' menu of the main navigation bar on any FlyBase page and can be obtained via a <u>FlyBase-styled webpage interface</u> (left panel in figure below) or standard <u>ftp site interface</u> (right panel). Details of the contents and format of each file are provided in <u>this document</u>, which is available via the 'Overview' link under the 'Downloads' menu on the navigation bar or by clicking the 'Help' button at the top of the webpage interface. Please take a few minutes to browse these pages before completing the survey.

See FlyBase	Current Release	Index of ftp://ftp.flybase.net/releases/c	current/
fome Tools Downloads Links Commun	ity Species About Help Archives	20 - June to Game 00 Im Up to higher level directory	
		Help Open All Close All	
Main Data Set		Name	Size Last Modified
Synonyms		File: FB2018.06	21/12/2018 18:07:00 GMT
Genes		README	21/12/2018 20:32:00 GMT
	ML Other		
ern Chado. Senes data downle		🔛 chado-xml	21/12/2018 20:51:00 GMT
Senes data downer	gene genetic interactions fb 2018 06.tsv.gz	Collaborators	21/12/2018 18:10:00 GMT
NA-Seg RPKM values	gene_rpkm_report_fb_2018_06.tsv.gz		
hysical Interaction table TSV	physical interactions fb 2018 06 tsv gz,	dana_r1.06	21/12/2018 18:24:00 GMT
MITAB TSV	physical_interactions_mitab_fb_2018_06.tsv.gz	dere_r1.05	21/12/2018 18:25:00 GMT
unctional complementation table	gene_functional_complementation_fb_2018_06.tsv.gz	dari_r1.05	21/12/2018 18:25:00 GMT
Bgn ++ DB Accession IDs	fbgn_NAseq_Uniprot_fb_2018_06.tsv.gz	dmel_r6.25	21/12/2018 18:25:00 GMT
Bgn ++ Annotation ID	fbgn_annotation_ID_fb_2018_06.tsv.gz		
Bgn ++ GLEANR IDs	fbgn_gleanr_fb_2018_06.tsv.gz	dmoj_r1.04	21/12/2018 18:26:00 GMT
Bgn ↔ FBtr ↔ FBpp IDs	fbgn_fbtr_fbpp_fb_2018_08.tsv.gz	dper_r1.3	21/12/2018 18:26:00 GMT
Bgn Exons ++ Atty1	fbgn_exons2affy1_overlaps.tsv.gz	dpse_r3.04	21/12/2018 18:24:00 GMT
Bgn Exons ++ Affy2	fbgn_exons2affy2_overlaps.tsv.gz		21/12/2018 18:27:00 GMT
lenes GO data	gene_association.fb.gz	dsec_r1.3	
lenes map table	gene_map_table_fb_2018_06.tsv.gz	dsim_r2.02	21/12/2018 18:24:00 GMT
utomated gene summaries	automated_gene_summaries.tsv.gz	dvir_r1.07	21/12/2018 18:24:00 GMT
Jene snapshots Drosophila Orthologs	gene_snapshots_fb_2018_06.tsv.gz	dwiLr1.05	21/12/2018 18:28:00 GMT
Inique Protein Isoforms	dmeLorthologs_in_drosophila_species_fb_2018_06.tsv.gz dmeLunique_protein_isoforms_fb_2018_06.tsv.gz		
Ion-coding RNA genes (TSV)	ncRNA_genes_fb_2018_06.tsv.gz	dyak_r1.05	21/12/2018 18:28:00 GMT
Ion-coding RNAs (JSON)	ncRNA genes fb 2018.06.json.gz	precomputed_files	21/12/2018 18:10:00 GMT
Gene Groups	Harris Garris 10, 10 10, 00 plongs		
Alleles and Stocks		🖾 psqt	21/12/2018 20:53:00 GMT
Orthologs			
Human Disease			
Nomenclature			
Ontology Terms			
Genomes: Annotation and Sequence			
Genomes: Annotation and Sequence Transcripts and Polypeptides			
Transcripts and Polypeptides Transposons, Transgenic Constructs, and			
	I Insertouris		
Aberrations			
Large Dataset Metadata			
Clones			
References			
Drosophila Researchers			
Map Conversion Tables			

This survey is designed to discover how researchers use FlyBase bulk data files and how we can improve their accessibility and utility.

Note: The Safari browser doesn't support browsing of FTP site directories - please use an alternative browser when completing this survey.



* 1. Which of the following best describes your role/expertise?

- Lab researcher
- Bioinformatician/developer
- Bit of both
- Other

* 2. Do you use the bulk data files provided by FlyBase?

- Yes frequently (every FlyBase release)
- Yes occasionally (more than once/year)
- Yes infrequently (once/year or less)
- No I don't need to access FlyBase data in bulk
- > No I've wanted to, but wasn't aware of/couldn't find these files
- No I use alternative methods/sites to access FlyBase data in bulk. Please specify:

GLOCE FlyBase FlyBase bulk data files

* 3. How do you usually access the file(s)? (Check as many as applicable)

Manual download via the <u>FlyBase-styled webpage interface</u>

Manual download via the <u>ftp site interface</u>

Command line/scripted access via the ftp site

GAPERING SET STATE FlyBase bulk data files

* 4. We are considering retiring the <u>FlyBase-styled webpage interface</u>, leaving the <u>ftp site interface</u> as the sole method to browse and access bulk files via a web browser. How would this affect you?

- No impact/opinion
- Minor inconvenience I can use the ftp interface instead
- Major inconvenience I rely on the webpage interface

Reasons/additional comments:

* 5. Which of the following bulk file types do you use, or would use now that you are aware of them? (Check as many as applicable)

Precomputed data files (e.g. synonyms, genetic interactions, gene summaries, orthologs)

Genomic sequence files (FASTA, GFF, GTF)

Ontology files (e.g. Gene Ontology, fly anatomy)

Map conversion tables (e.g. cytogenetic map <-> genomic coordinates)

Database dumps for individual data types (Chado XML)

- Entire FlyBase database (PSQL)
 - None

* 6. Are there additional bulk files, or changes to the content/file format of the current files, that would benefit your work?

🔵 No

Yes (please specify)

* 7. Do you have any suggestions to improve the <u>documentation</u> about the bulk files?

- 🔵 No
- Yes (please specify)
- * 8. Have you used 'Precomputed files' as a data source in our <u>Batch Download</u> tool, allowing bulk retrieval of data from these files for a specified set of genes/transcripts (etc)?
 - 🔵 Yes
 - No, this functionality isn't useful to me
 - No, I wasn't aware of this functionality but it looks useful
 - I don't use Batch Download

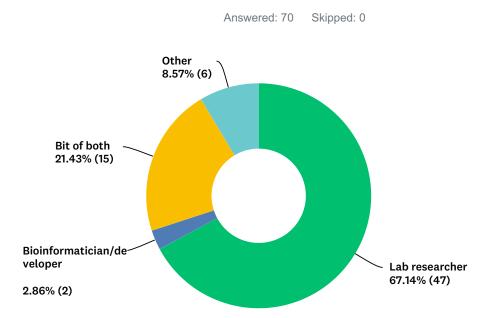
9. Please give any additional comments about the provision of or access to bulk data files at FlyBase below.



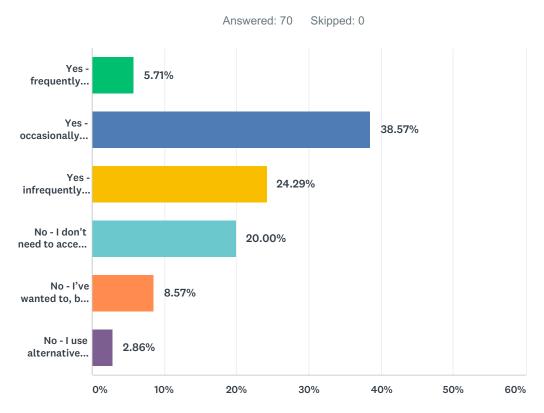
FlyBase bulk data files

Survey Answers Feb 2019

Q1 Which of the following best describes your role/expertise?



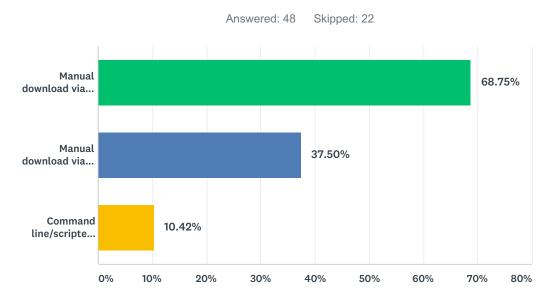
ANSWER CHOICES	RESPONSES	
Lab researcher	67.14%	47
Bioinformatician/developer	2.86%	2
Bit of both	21.43%	15
Other	8.57%	6
TOTAL		70



Q2 Do you use the bulk data files provided by FlyBase?

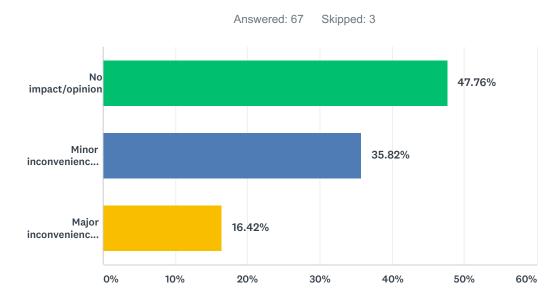
ANSWER CHOICES	RESPONSES	
Yes - frequently (every FlyBase release)	5.71%	4
Yes - occasionally (more than once/year)	38.57%	27
Yes - infrequently (once/year or less)	24.29%	17
No - I don't need to access FlyBase data in bulk	20.00%	14
No - I've wanted to, but wasn't aware of/couldn't find these files	8.57%	6
No - I use alternative methods/sites to access FlyBase data in bulk. Please specify:	2.86%	2
TOTAL		70

Q3 How do you usually access the file(s)? (Check as many as applicable)



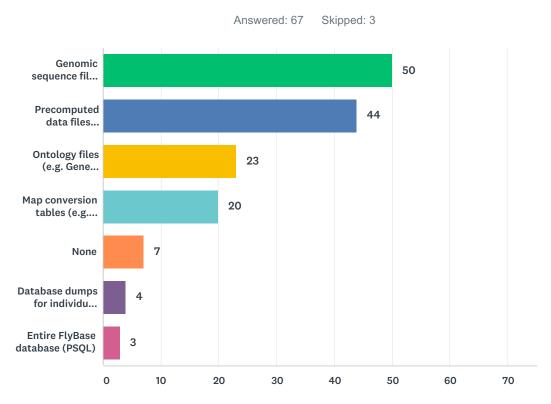
ANSWER CHOICES	RESPONSES	
Manual download via the FlyBase-styled webpage interface	68.75%	33
Manual download via the ftp site interface	37.50%	18
Command line/scripted access via the ftp site	10.42%	5
Total Respondents: 48		

Q4 We are considering retiring the FlyBase-styled webpage interface, leaving the ftp site interface as the sole method to browse and access bulk files via a web browser. How would this affect you?



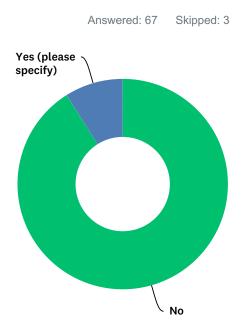
ANSWER CHOICES	RESPONSES	
No impact/opinion	47.76%	32
Minor inconvenience - I can use the ftp interface instead	35.82%	24
Major inconvenience - I rely on the webpage interface	16.42%	11
TOTAL		67

Q5 Which of the following bulk file types do you use, or would use now that you are aware of them? (Check as many as applicable)



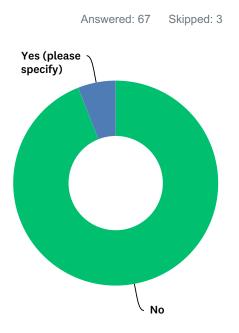
ANSWER CHOICES	RESPONSES	
Genomic sequence files (FASTA, GFF, GTF)	74.63%	50
Precomputed data files (e.g. synonyms, genetic interactions, gene summaries, orthologs)	65.67%	44
Ontology files (e.g. Gene Ontology, fly anatomy)	34.33%	23
Map conversion tables (e.g. cytogenetic map <-> genomic coordinates)	29.85%	20
None	10.45%	7
Database dumps for individual data types (Chado XML)	5.97%	4
Entire FlyBase database (PSQL)	4.48%	3
Total Respondents: 67		

Q6 Are there additional bulk files, or changes to the content/file format of the current files, that would benefit your work?



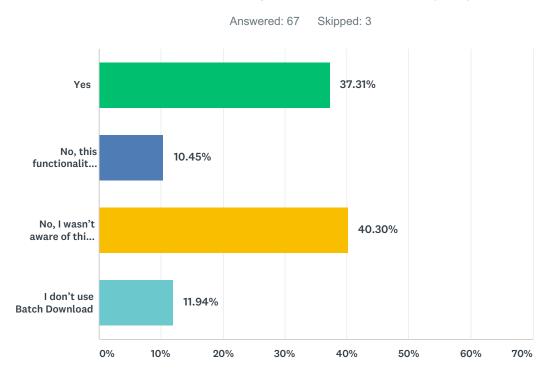
ANSWER CHOICES	RESPONSES	
No	91.04%	61
Yes (please specify)	8.96%	6
TOTAL		67

Q7 Do you have any suggestions to improve the documentation about the bulk files?



ANSWER CHOICES	RESPONSES	
No	94.03%	63
Yes (please specify)	5.97%	4
TOTAL		67

Q8 Have you used 'Precomputed files' as a data source in our Batch Download tool, allowing bulk retrieval of data from these files for a specified set of genes/transcripts (etc)?



ANSWER CHOICES	RESPONSES	
Yes	37.31%	25
No, this functionality isn't useful to me	10.45%	7
No, I wasn't aware of this functionality but it looks useful	40.30%	27
I don't use Batch Download	11.94%	8
TOTAL		67

FlyBase bulk data files

Q9 Please give any additional comments about the provision of or access to bulk data files at FlyBase below.

Answered: 8 Skipped: 62