

568A - Annotation of Drosophila genomes by NCBI's RefSeq project

Terence Murphy, Françoise Thibaud-Nissen, Valerie Schneider, and the NCBI Genome Annotation Team

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

TAGC 2020 April 27, 2020



U.S. National Library of Medicine
National Center for Biotechnology Information

Overview

NCBI RefSeq is your best source for annotated *Drosophila* genomes

- Genome annotations for [30 *Drosophila* species](#) are available now in NCBI RefSeq, with more to come
 - All produced with NCBI's annotation pipeline using the latest public assemblies and RNA-seq data
 - Species other than *D. melanogaster* are being [retired at FlyBase](#) to optimize resources
- Many tools available to use the data at NCBI
 - [NCBI Gene](#)
 - [BLAST](#)
 - [Genome Data Viewer](#)
- Everything available by FTP
 - Try the new [NCBI Datasets!](#)
 - [Orthology data](#)

Abstract

Advances in sequencing technology over the last decade have led to a cornucopia of *Drosophila* genome assemblies, with over 140 assemblies from 70 *Drosophila* species now available in public sequence archives. Many species have new, high-quality assemblies based on PacBio, Oxford Nanopore (ONT), or other technologies along with abundant RNA-seq datasets, generated by many researchers from around the world. To help maximize the utility of these genomes for the *Drosophila* research community, NCBI's Reference Sequence (RefSeq) project is now the primary annotation source for nearly 30 *Drosophila* species, allowing FlyBase to focus their resources on *D. melanogaster*. NCBI's automated annotation pipeline provides rapid, high-quality gene annotations across many taxa, with consistent processing that benefits comparative genomic studies. Annotations are available in NCBI's Gene resource, BLAST databases, and Genome Data Viewer (GDV). Gene and GDV also provide access to other genomic information including orthologs, RNA-seq expression data, and whole genome alignments to previous assembly versions or assemblies from different strains. This presentation will give an overview of these resources, including the scope and quality of annotation data currently available. Further information about NCBI's annotation resources and GDV is available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/ and <https://www.ncbi.nlm.nih.gov/genome/gdv/>.

Genomes annotated by NCBI

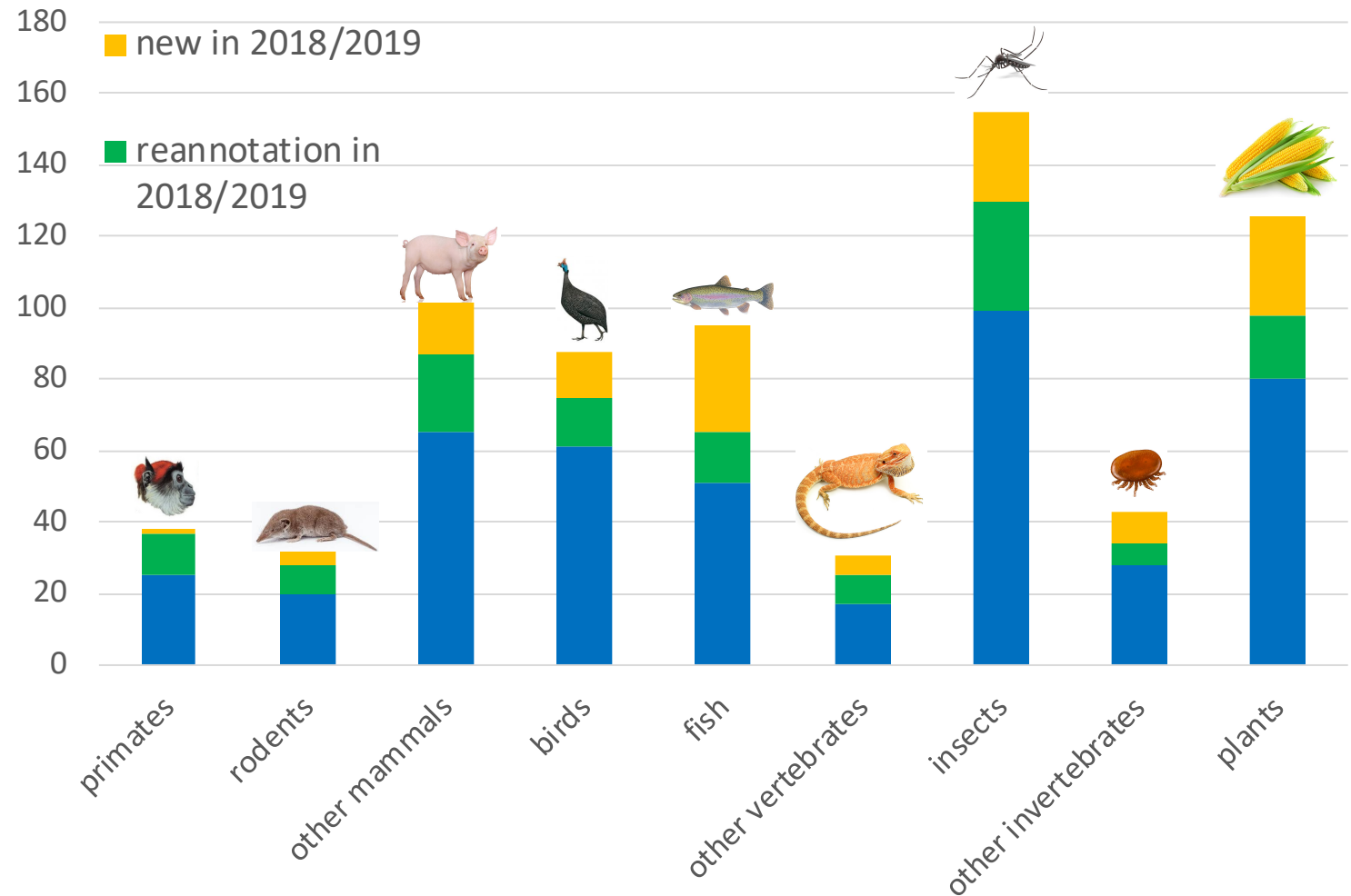
RefSeq project started in 1999

Over 600 eukaryotic species annotated to date

RefSeq also includes:

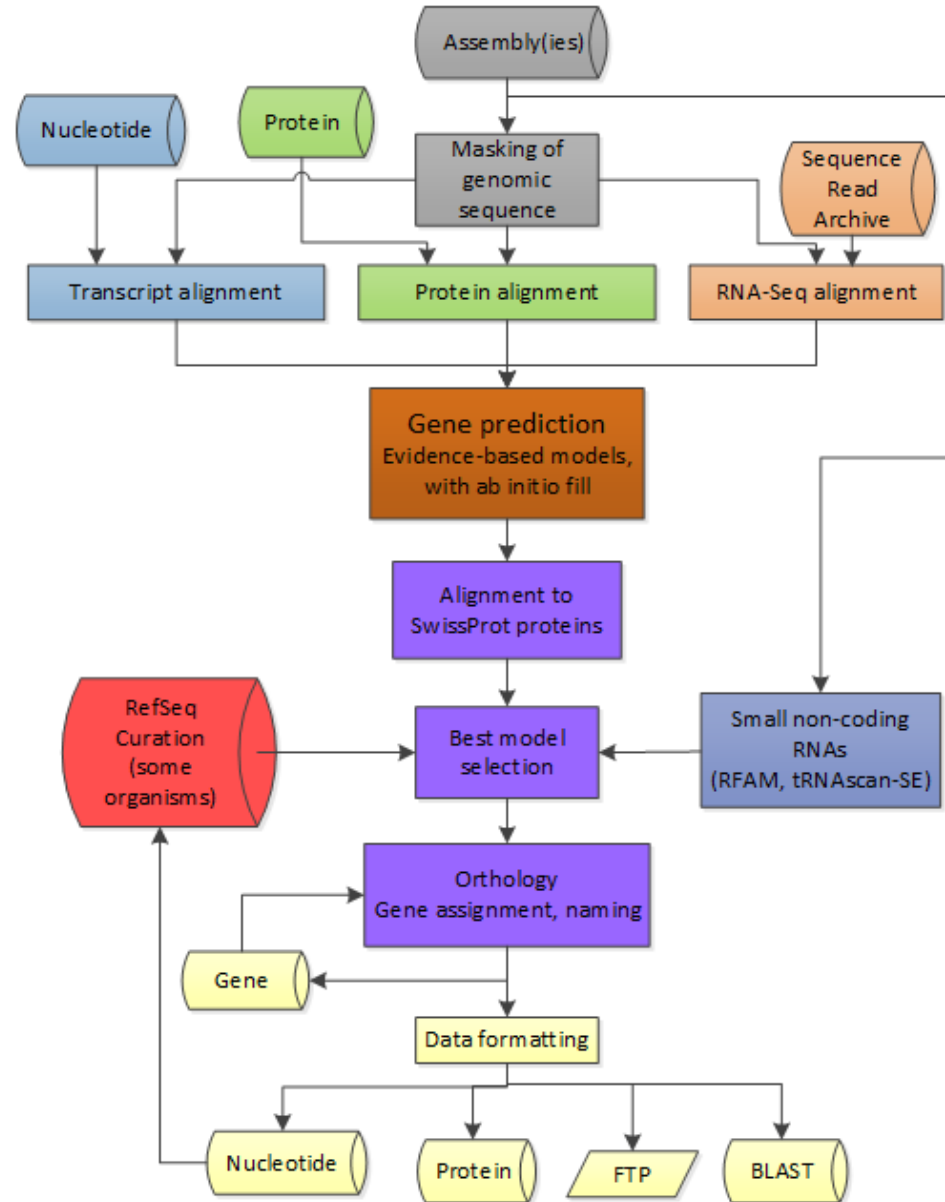
- >400 other eukaryotes
- >180,000 bacteria genomes annotated by NCBI (PGAP)
- >9,000 viruses
- >14,000 organelles

More info: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/



NCBI's Eukaryotic Genome Annotation Pipeline (EGAP)

- Automated
- Evidence-based
 - RNA-seq
 - IsoSeq/ONT long reads
 - Proteins
- Data in:
 - NCBI Gene
 - nucleotide/protein dbs
 - BLAST
 - FTP / Datasets
 - Genome Data Viewer



NCBI RefSeq has annotations for 30 *Drosophila* species

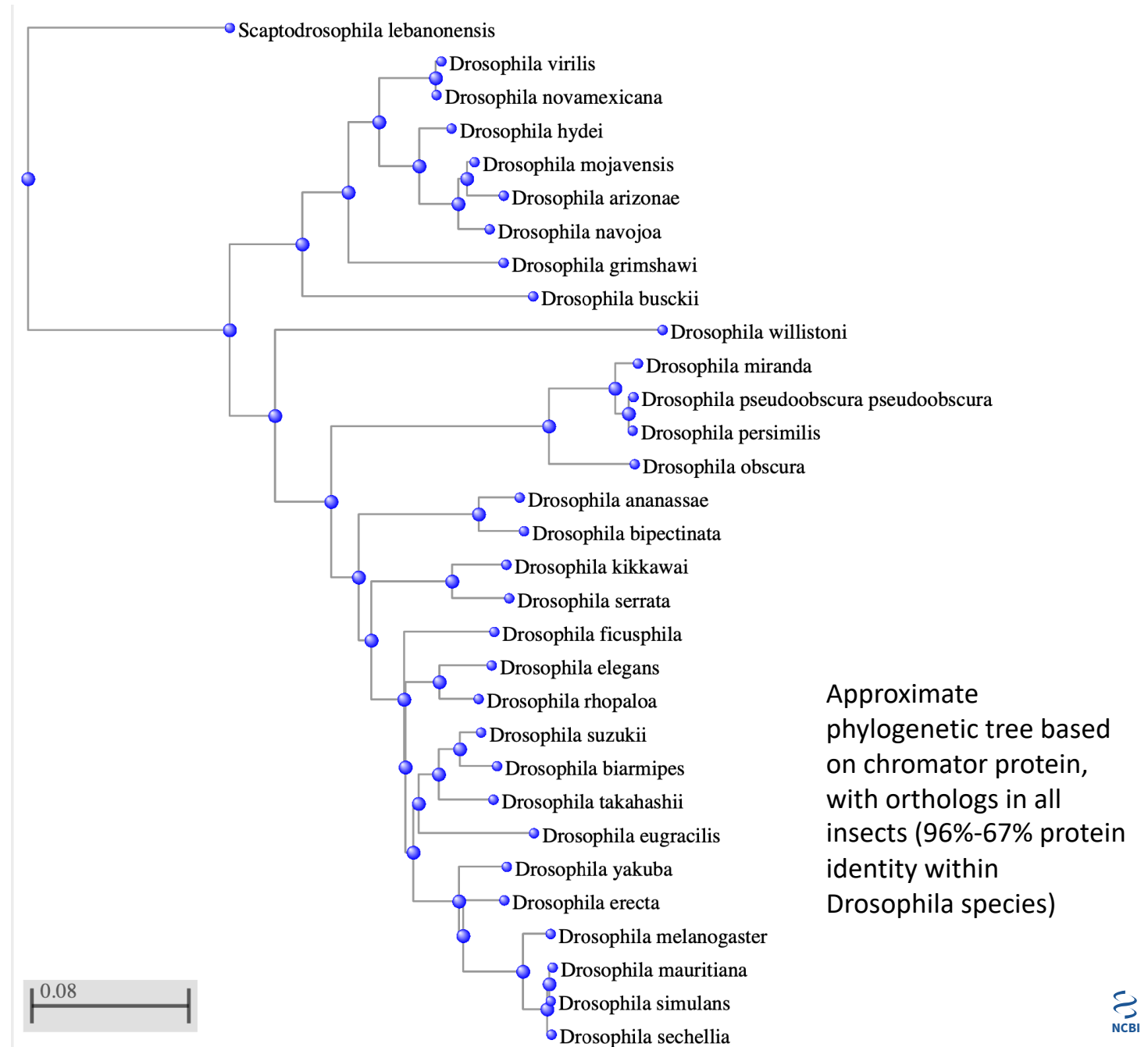
We select public assemblies for
annotation

Annotated with NCBI's EGAP
pipeline using RNA-seq from SRA

- *D. melanogaster* using
FlyBase (Release 6.26)

Non-melanogaster species are
transitioning out of FlyBase and
only in NCBI RefSeq

- More consistency across
species
- Faster updates to new
assemblies
- More support in NCBI tools
- Adding more in the future



High-quality annotations

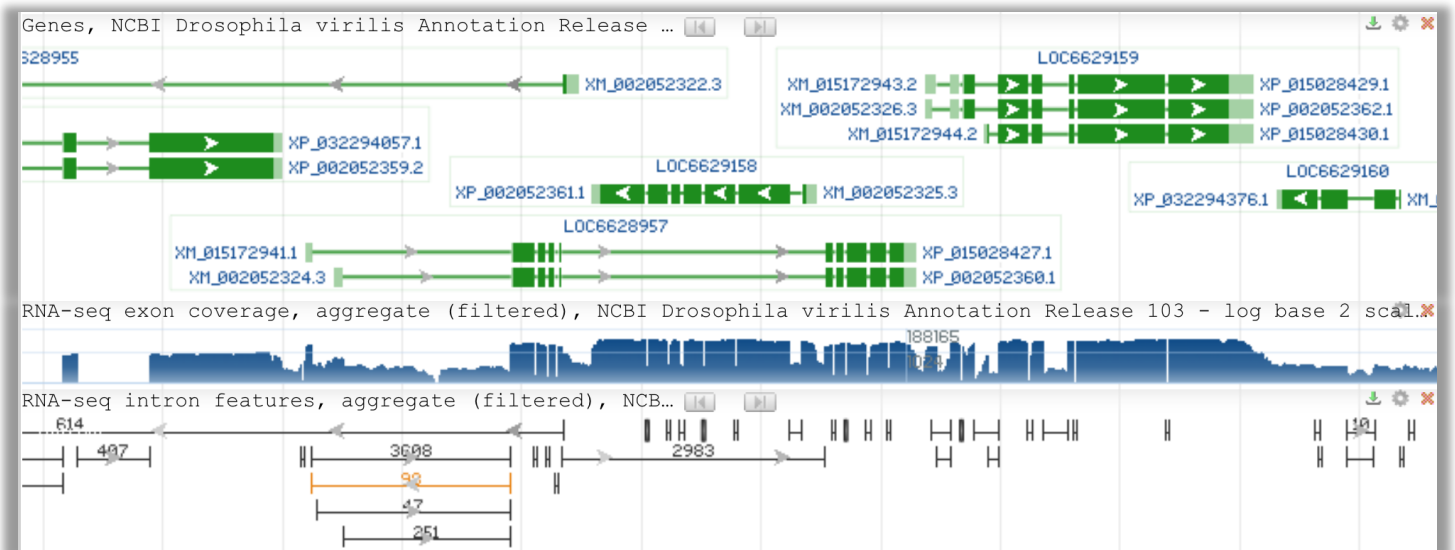
NCBI EGAP produces accurate gene annotations, even in regions of densely spaced and overlapping genes

Annotations are similar to the curated FlyBase annotation for *D. melanogaster* orthologs

D. melanogaster ([live view](#) in GDV)

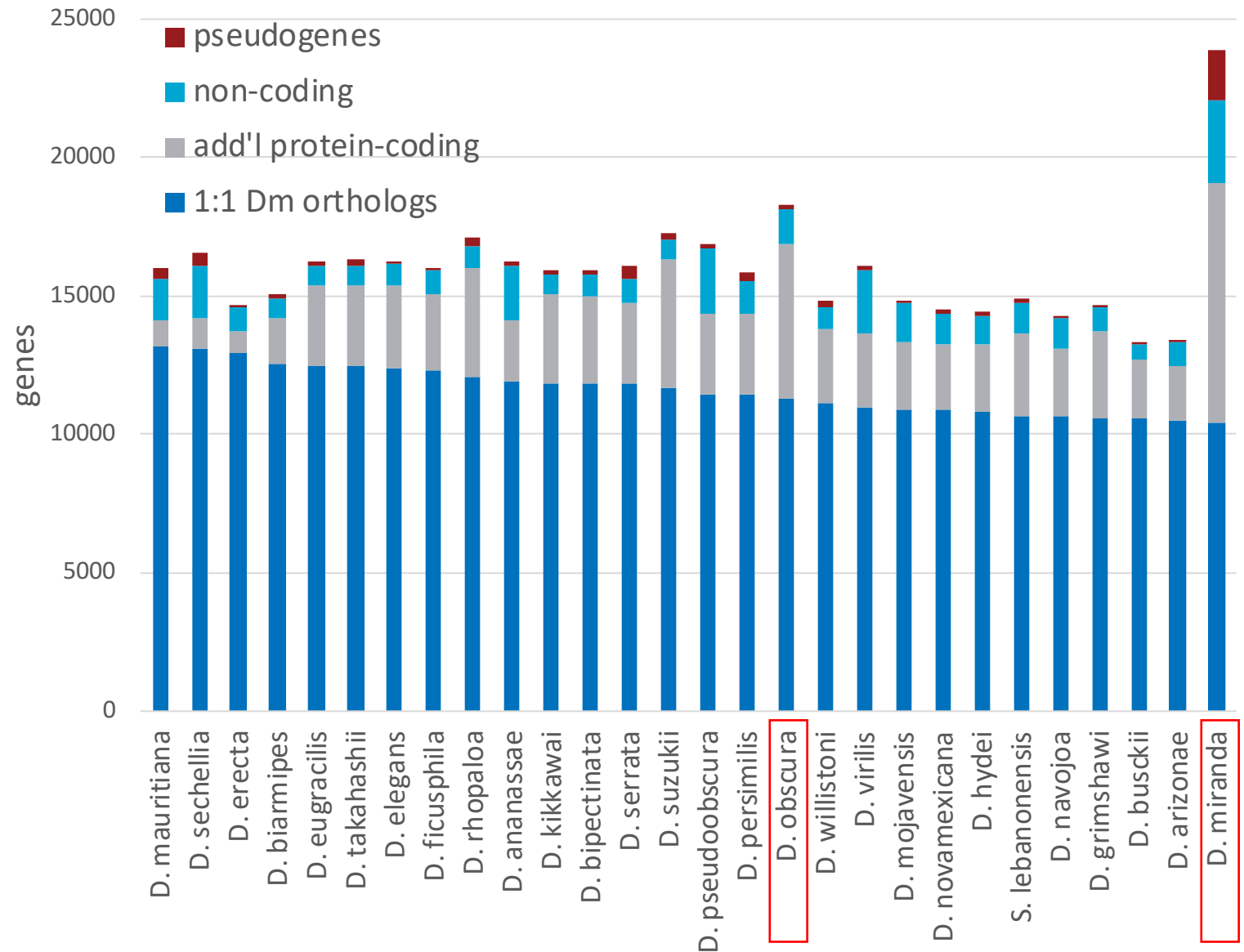


D. virilis ([live view](#) in GDV)



Drosophila Gene Annotation Statistics

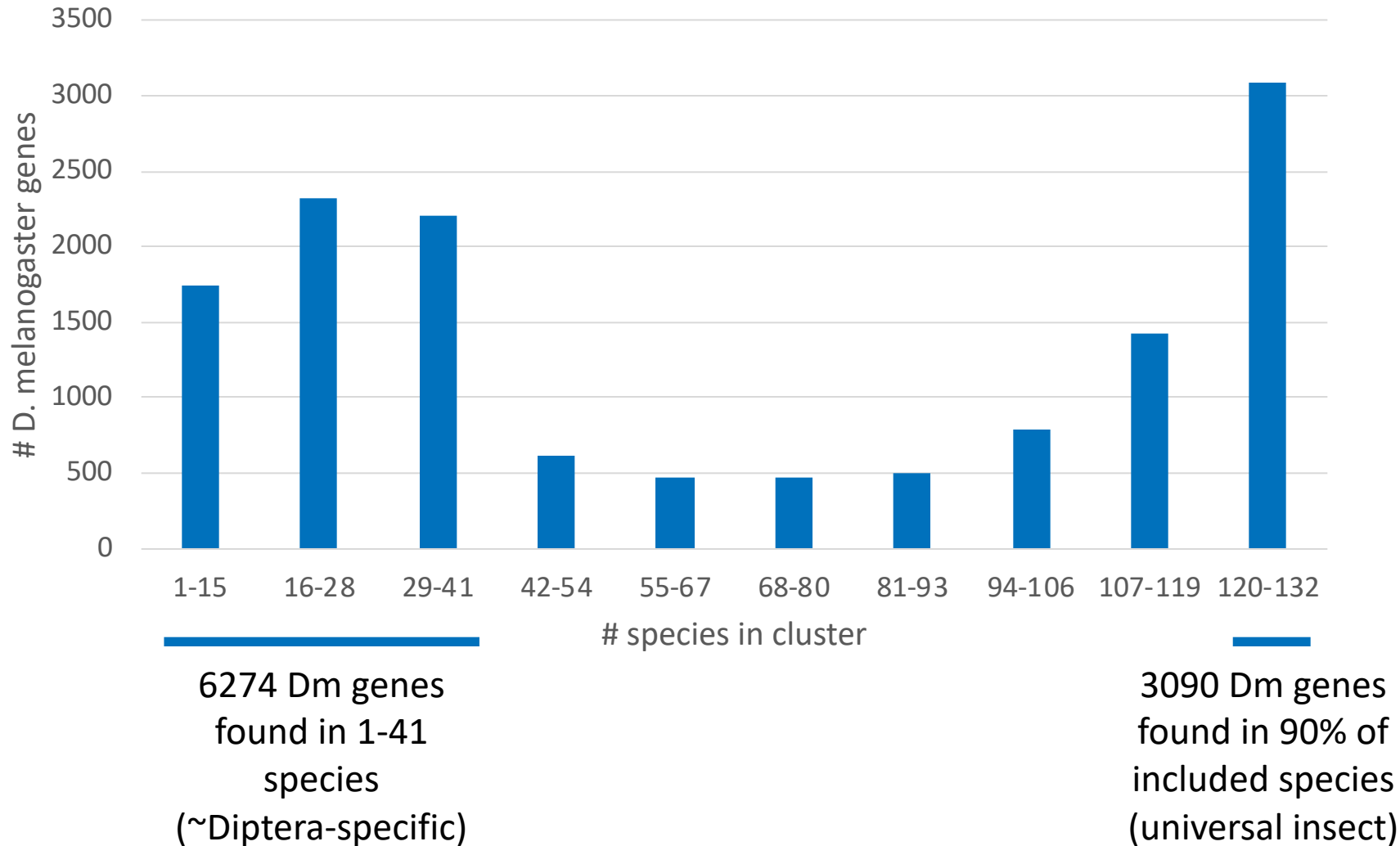
- Median 14206 protein-coding genes (compared to 13933 in *D. melanogaster*)
 - ~75-95% have *D. melanogaster* orthologs
- Lower-quality genomes (e.g. *D. obscura*) have an excess of protein-coding genes due to fragmentation
- *D. miranda* has an evolving Y chromosome derived from chr3 (Mahajan et al 2018), resulting in more coding genes and affecting identification of 1:1 orthologs



Orthology

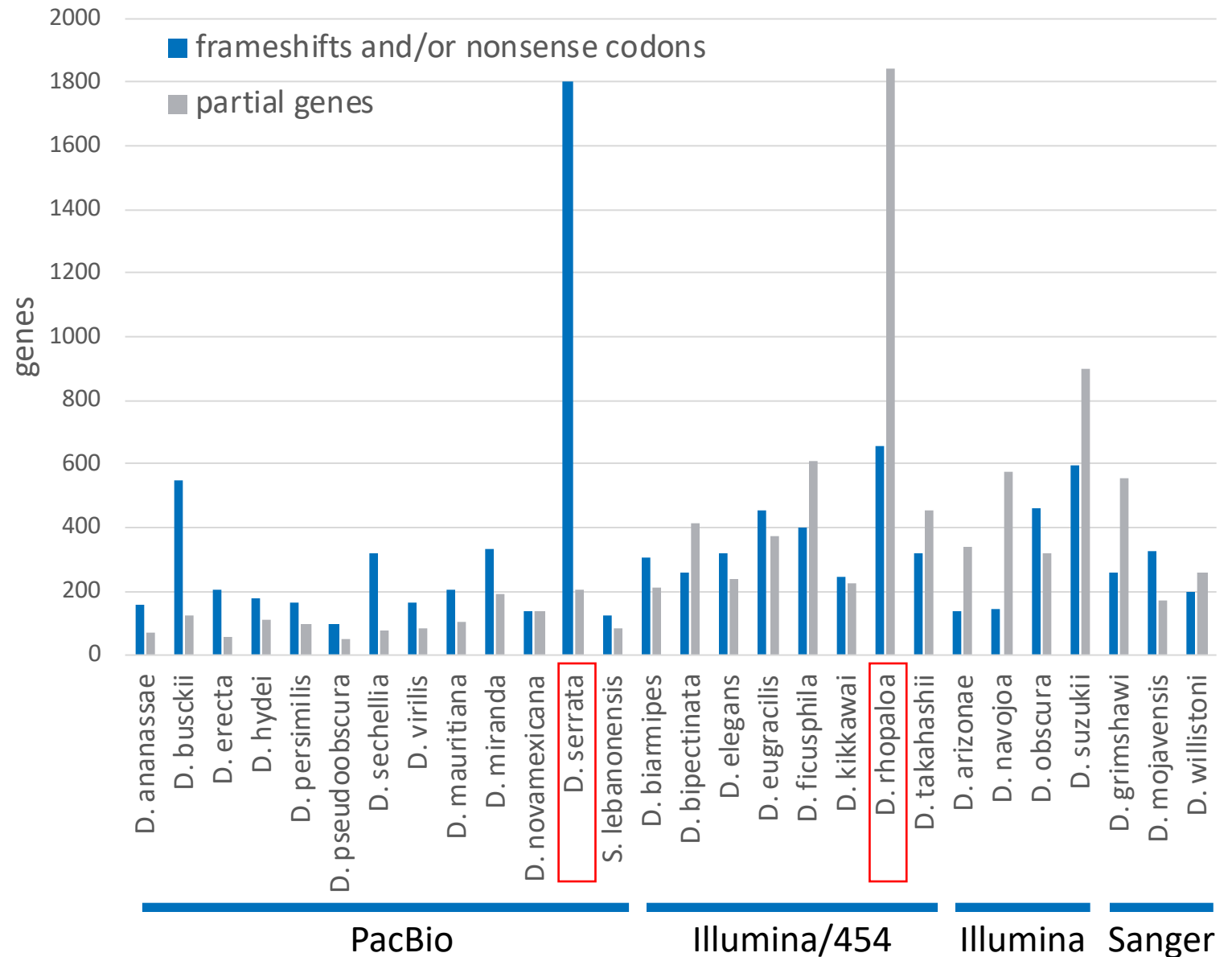
- Computed vs *D. melanogaster* using protein alignments and local synteny
- 10,551 *Dm* genes identified as 1:1 orthologs in >80% of other *Drosophila* species
- Data available for >130 insects
 - Identifies core of ~3090 nearly universal insect genes
- Can use to connect to functional information from *D. melanogaster*
- Data on FTP. Will be integrated into web pages in the future

Data at: https://ftp.ncbi.nlm.nih.gov/gene/DATA/special_requests/



Quality Analysis

- Most of the new PacBio assemblies are excellent, with few protein-coding genes identified as partial or with suspected assembly errors (frameshifting indels or nonsense codons)
- Some genomes are outliers with many partials (e.g. *D. rhopaloa*) or frameshifts (*D. serrata*)
 - Polishing or other measures are critical to remove indels from PacBio assemblies



Everything you might want to know about the assemblies...

species	assembly accession	assembly name	level	scaffold N50	contig N50	technology	assembly submitter	assembly publ.
* RefSeq uses annotation from FlyBase								
Drosophila melanogaster	GCF_000001215.4	Release 6	Chromosome	25.3 Mbp	21.5 Mbp	Sanger	FlyBase/BDGP/Celera	
* RefSeq annotation recently updated and no longer maintained at FlyBase								
Drosophila ananassae	GCF_003285975.2	DanaRS2.1	Contig	6.21 Mbp	6.21 Mbp	PacBio, Illumina	University of Arizona	
Drosophila erecta	GCF_003286155.1	DereRS2	Contig	22.1 Mbp	22.1 Mbp	PacBio, Illumina	University of Arizona	
Drosophila grimshawi	GCF_000005155.2	dgri_caf1	Scaffold	8.40 Mbp	91 kb	Sanger	Agencourt	PMID:17994087
Drosophila mojavensis	GCF_000005175.2	dmoj_caf1	Scaffold	24.8 Mbp	122 kb	Sanger	Agencourt	PMID:17994087
Drosophila persimilis	GCF_003286085.1	DperRS2	Contig	5.21 Mbp	5.21 Mbp	PacBio, Illumina	University of Arizona	
Drosophila pseudoobscura	GCF_009870125.1	UCI_Dpse_MV25	Chromosome	32.4 Mbp	30.7 Mbp	PacBio	University of California, Irvine	
Drosophila sechellia	GCF_004382195.1	ASM438219v1	Chromosome	30.0 Mbp	19.9 Mbp	PacBio	University of California, Irvine	
Drosophila virilis	GCF_003285735.1	DvirRS2	Contig	8.70 Mbp	8.70 Mbp	PacBio, Illumina	University of Arizona	
Drosophila willistoni	GCF_000005925.1	dwil_caf1	Scaffold	4.51 Mbp	180 kb	Sanger	JCVI	PMID:17994087
* updates planned (last annotated by NCBI for FlyBase in 2014)								
Drosophila simulans	GCF_000754195.2	ASM75419v3	Chromosome	23.5 Mbp	452 kb	Illumina, Sanger	Princeton University	PMID:22936249
Drosophila yakuba	GCF_000005975.2	dyak_caf1	Chromosome	21.8 Mbp	116 kb	Sanger	WUGSC	PMID:17994087
* other Drosophila species with NCBI-produced annotation in RefSeq								
Drosophila arizonae	GCF_001654025.1	ASM165402v1	Scaffold	26.5 Mbp	15 kb	Illumina	Biotechnology Institute of UNAM	PMID:27489210
Drosophila biarmipes	GCF_000233415.1	Dbia_2.0	Scaffold	3.39 Mbp	475 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila bipectinata	GCF_000236285.1	Dbip_2.0	Scaffold	0.66 Mbp	149 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila busckii	GCF_011750605.1	ASM1175060v1	Chromosome	22.7 Mbp	1.00 Mbp	Illumina, PacBio	Max Planck Institute	PMID:31601616
Drosophila elegans	GCF_000224195.1	Dele_2.0	Scaffold	1.71 Mbp	213 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila eugracilis	GCF_000236325.1	Deug_2.0	Scaffold	0.98 Mbp	224 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila ficusphila	GCF_000220665.1	Dfic_2.0	Scaffold	1.05 Mbp	276 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila hydei	GCF_003285905.1	DhydRS2	Contig	3.37 Mbp	3.37 Mbp	PacBio, Illumina	University of Arizona	
Drosophila kikkawai	GCF_000224215.1	Dkik_2.0	Scaffold	0.90 Mbp	209 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila mauritiana	GCF_004382145.1	ASM438214v1	Chromosome	24.2 Mbp	22.1 Mbp	PacBio	University of California, Irvine	
Drosophila miranda	GCF_003369915.1	D.miranda_PacBio2.1	Chromosome	35.2 Mbp	12.0 Mbp	PacBio	University of California, Berkeley	PMID:30059545
Drosophila navojoa	GCF_001654015.2	UFRJ_Dnav_4.2	Scaffold	0.39 Mbp	27 kb	Illumina	Biotechnology Institute of UNAM	PMID:27489210
Drosophila novamexicana	GCF_003285875.2	DnovRS2.1	Contig	3.16 Mbp	3.16 Mbp	PacBio, Illumina	University of Arizona	
Drosophila obscura	GCF_002217835.1	Dobs_1.0	Scaffold	0.47 Mbp	115 kb	Illumina	Tokyo Metropolitan University	PMID:27897175
Drosophila rhopaloa	GCF_000236305.1	Drho_2.0	Scaffold	0.05 Mbp	19 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Drosophila serrata	GCF_002093755.1	Dser1.0	Contig	0.94 Mbp	0.94 Mbp	PacBio	University of Queensland	PMID:28143951
Drosophila suzukii	GCF_000472105.1	Dsuzukii.v01	Scaffold	0.39 Mbp	25 kb	Illumina	BGI-Shenzhen	PMID:24142924
Drosophila takahashii	GCF_000224235.1	Dtak_2.0	Scaffold	0.39 Mbp	126 kb	Illumina, 454	Baylor College of Medicine	PMID:24985915
Scaptodrosophila lebanonensis	GCF_003285725.1	SlebRS2	Contig	7.85 Mbp	7.85 Mbp	PacBio, Illumina	University of Arizona	

...and the annotations.

the annotations.

species	annotation release	annotation date	RNA-seq			genes					
			reads (M)	runs	samples	protein-coding	non-coding	pseudo	orthologs	partial	low-quality
<i>* RefSeq uses annotation from FlyBase</i>											
Drosophila melanogaster	R6.26	25-Apr-2019				13933	3531	308	n/a		
<i>* RefSeq annotation recently updated and no longer maintained at FlyBase</i>											
Drosophila ananassae	101	18-Feb-2020	5041	299	216	14128	1924	223	11942	69	158
Drosophila erecta	101	6-Nov-2018	618	32	32	13718	867	96	12922	57	204
Drosophila grimshawi	102	10-Mar-2020	1982	208	105	13754	814	106	10611	556	262
Drosophila mojavensis	101	9-Mar-2020	2825	205	99	13329	1385	68	10891	173	330
Drosophila persimilis	101	7-Nov-2018	624	112	56	14397	1104	353	11422	99	164
Drosophila pseudoobscura	104	8-Apr-2020	8872	290	228	14343	2410	127	11480	50	100
Drosophila sechellia	101	3-Mar-2020	2701	88	81	14182	1939	432	13136	77	320
Drosophila virilis	103	11-Feb-2020	4491	208	125	13685	2215	202	10947	83	165
Drosophila willistoni	101	4-Dec-2017	378	56	56	13769	806	276	11095	257	196
<i>* updates planned (last annotated by NCBI for FlyBase in 2014)</i>											
Drosophila simulans	R2.01	8-Oct-2014	3211	n/a	41	14179	900	187	nd	nd	nd
Drosophila yakuba	R1.04	9-Jan-2014	995	n/a	20	14837	1100	338	nd	nd	nd
<i>* other Drosophila species with NCBI-produced annotation in RefSeq</i>											
Drosophila arizonae	100	29-Aug-2016	3025	74	29	12476	827	26	10473	341	141
Drosophila biarmipes	101	17-Jun-2016	400	3	3	14230	704	132	12557	211	307
Drosophila bipectinata	101	17-Jun-2016	505	3	3	14981	800	135	11806	413	256
Drosophila busckii	101	8-Apr-2020	876	3	3	12712	569	30	10567	127	547
Drosophila elegans	101	17-Jun-2016	499	3	3	15407	750	115	12369	240	323
Drosophila eugracilis	101	17-Jun-2016	351	3	3	15375	689	187	12500	377	452
Drosophila ficusphila	101	17-Jun-2016	433	5	3	15062	860	98	12343	612	401
Drosophila hydei	101	12-Aug-2019	464	4	4	13282	983	135	10811	113	177
Drosophila kikkawai	101	17-Jun-2016	311	3	3	15096	675	139	11807	225	244
Drosophila mauritiana	100	9-Apr-2020	1130	24	23	14112	1535	359	13163	106	208
Drosophila miranda	102	8-Apr-2020	5415	115	95	19112	2964	1788	10422	195	331
Drosophila navojoa	101	16-Aug-2019	5943	248	118	13082	1118	45	10644	577	142
Drosophila novamexicana	100	1-Sep-2019	386	11	4	13260	1117	103	10856	135	136
Drosophila obscura	100	24-Aug-2017	1806	34	17	16865	1269	168	11263	320	464
Drosophila rhopaloa	101	17-Jun-2016	367	3	3	16017	770	363	12096	1842	656
Drosophila serrata	100	2-May-2017	336	8	3	14775	809	497	11802	205	1800
Drosophila suzukii	101	10-Jun-2016	445	14	7	16335	727	226	11688	901	594
Drosophila takahashii	101	17-Jun-2016	352	3	3	15410	719	172	12495	456	318
Scaptodrosophila lebanonensis	100	26-Aug-2019	67	6	6	13655	1077	171	10654	83	126

Data Access

NCBI Gene

Your portal into Gene related content:

- Search
- Nomenclature
- Primary identifier: GeneID
 - Old FlyBase identifiers work for search
- Transcript and protein sequences
- Link to Genome Data Viewer
- Some publications
 - Add your own via GeneRIF!

<https://www.ncbi.nlm.nih.gov/gene/>

NCBI Resources How To

murphyte2 My NCBI Sign Out

Gene Gene drosophila virilis HP1 Search

Create RSS Save search Advanced Help

Full Report Send to: Hide sidebar >>

Showing Current items.

LOC6627672 heterochromatin protein 1 [*Drosophila virilis*]

Gene ID: 6627672, updated on 15-Apr-2020

Summary

Gene symbol LOC6627672

Gene description heterochromatin protein 1

Locus tag Dvir_GJ17281

Gene type protein coding

RefSeq status MODEL

Organism [Drosophila virilis](#)

Lineage Eukaryota; Metazoa; Ecdysozoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Holometabola; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila

Also known as HP1; HP-1; HP1A; DvHP1; GJ17281; HP1A-PA; Su(var); DvirHP1A; GJ17281-PA; Su(var)205

Genomic context

Location: chromosome: Un

Exon count: 5

Annotation release	Status	Assembly	Chr	Location
103	current	DvirRS2 (GCF_003285735.1)	Unplaced Scaffold	NW_022587572.1 (3698920..3700255)

See LOC6627672 in [Genome Data Viewer](#)

Table of contents

- Summary
- Genomic context
- Genomic regions, transcripts, and products
- Bibliography
- General gene information
 - Homology
- General protein information
- NCBI Reference Sequences (RefSeq)
- Related sequences
- Additional links

Genome Browsers

- Genome Data Viewer

Related information

- BioProjects
- Conserved Domains
- Full text in PMC

NCBI Genome Data Viewer

- Over 600 species supported
- RNA-seq expression tracks
- Assembly alignments to see previous annotation
- View BLAST alignments
- Track Hub support
- Upload your own tracks
- Download gene tables
- Make publication-quality SVG or PDF images

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information murphyte2 My NCBI Log out

Genome Data Viewer

Share this page Reset All More Info

Drosophila virilis
(NA)
Assembly: DvirRS2

Search assembly
HP1

Name	Location
LOC6627672	NW_022587572.1: 4M - 4M

Examples

Pick Assembly

Ideogram View

Unplaced/unlocalized scaffolds: 222

User Data and Track Hubs

BLAST

Add Tracks

History

NW_022587572.1 (unplaced): 3,698,786 - 3,700,389

Region LOC6627672 Transcript XM_002051912.3 Exons: click an exon above to zoom in, mouse over to see details

Tools Tracks Download

Genes, NCBI Drosophila virilis Annotation Release 103

LOC6627672

RNA-seq exon coverage, aggregate (filtered), NCBI Drosophila virilis Annotation Release 103

RNA-seq intron features, aggregate (filtered), NCBI Drosophila virilis Annotation Release 103

RNA-seq exon coverage, Blastoderm, Whole embryo (Drosophila virilis, male, SAMN03492546, filtered), NCBI Drosophila virilis Annotation Release 103

RNA-seq exon coverage, Ovary (Drosophila virilis, SAMN07367980, filtered), NCBI Drosophila virilis Annotation Release 103

RNA-seq exon coverage, Testis (Drosophila virilis, SAMN07367979, filtered), NCBI Drosophila virilis Annotation Release 103

Assembly-assembly alignments, DvirRS2 (GCF_00328...)

View previous assembly and annotation

Feedback

NCBI Orthologs

- Accessed from NCBI Gene
- Currently supports vertebrate orthologs
- Planning to add insects in the future
- Functions:
 - Browse and filter sets
 - Generate alignments
 - Download sets
 - Explore literature


 National Library of Medicine
 National Center for Biotechnology Information


 murphyte2

BTRC – beta-transducin repeat containing E3 ubiquitin protein ligase

[Genes similar to BTRC](#)

NCBI Orthologs [How was this calculated?](#)

0 selected.

SEARCH THE TAXONOMY TREE

- tetrapods
 - birds
 - alligators and others
 - turtles
 - lizards
 - mammals
 - amphibians

Species	Gene	Architecture	aa
☐ <i>Homo sapiens</i> human	BTRC beta-transducin repeat containing E3 ubiquitin protein ligase		605
☐ <i>Mus musculus</i> house mouse	Btrc beta-transducin repeat containing protein		605

RefSeq transcripts (29)	RefSeq proteins (29)	Architecture	aa
NM_001037758.2	NP_001032847.2		605
NM_009771.3	NP_033901.1		569
NM_001286465.1	NP_001273394.1		605
NM_001286466.1	NP_001273395.1		383

NCBI Assembly

- Search for genome assemblies
- Filter by:
 - Just RefSeq
 - Assembly level
 - Annotation
- View assembly stats and info
- Download via FTP

<https://www.ncbi.nlm.nih.gov/assembly/>

NCBI Resources ▾ How To ▾

murphyte2 My NCBI Sign Out

Assembly Assembly ▾ drosophila[orgn] Search

Create alert Advanced Browse by organism Help

Organism group
Animals (33)
Customize ...

Summary ▾ 20 per page ▾ Sort by Significance ▾

Download Assemblies

Send to: ▾

Find related data
Database: Select ▾
Find items

Search details
("Drosophila"[Organism] OR "Drosophila"[Organism]) AND ("latest refseq"[filter] AND (all[filter] NOT "derived from surveillance project"[filter])

Recent activity
Turn Off Clear
drosophila[orgn] AND ("latest refseq"[filter] AND (all[filter] NO... (33) Assembly
LOC6627672 [Drosophila virilis] Gene
drosophila virilis HP1 AND (alive[prop]) (1) Gene
6627672[uid] AND (alive[prop]) (1) Gene
LOC6629158 [Drosophila virilis] Gene

Status clear
Latest (32)
Latest GenBank (32)
✓ Latest RefSeq (33)
Replaced (1)

Assembly level
Complete genome (0)
Chromosome (11)
Scaffold (15)
Contig (7)

RefSeq category
Reference (1)
Representative (28)

Exclude clear
Exclude partial (0)
✓ Exclude derived from surveillance project (0)
✓ Exclude anomalous (0)
Customize ...

Annotation status
Has annotation (33)
GenBank has annotation (10)
RefSeq has annotation (33)

Relation to type

Search results
Items: 1 to 20 of 33

Filters activated: Latest RefSeq, Exclude derived from surveillance project, Exclude anomalous.
Clear all to show 176 items.

1. ☐ Organism: **Drosophila melanogaster** (fruit fly)
Submitter: The FlyBase Consortium/Berkeley **Drosophila** Genome Project/Celera Genomics
Date: 2014/08/01
Assembly level: Chromosome
Genome representation: full
RefSeq category: reference genome
Synonyms: dm6
GenBank assembly accession: GCA_000001215.4 (latest)
RefSeq assembly accession: GCF_000001215.4 (latest)
IDs: 202931 [UID] 1186808 [GenBank] 1186828 [RefSeq]

2. ☐ Organism: **Drosophila mauritiana** (flies)
Intraspecific name: Strain: mau12
Sex: male
Submitter: University of California, Irvine
Date: 2019/04/16
Assembly level: Chromosome
Genome representation: full
RefSeq category: representative genome
GenBank assembly accession: GCA_004382145.1 (latest)
RefSeq assembly accession: GCF_004382145.1 (latest)

NCBI Datasets - Beta

- New way to access genome data
 - FASTA
 - GFF3
 - Reports
- Download a single zip file with multiple genomes and file types
- Access to individual gene data coming soon
- Web interface, API, and command-line tool
- Try it and provide [feedback!](#)

<https://www.ncbi.nlm.nih.gov/datasets/>

ncbi.nlm.nih.gov/datasets/genomes/?txid=32281&source_db=RefSeq

Search NCBI

Genomes - NCBI Datasets BETA

NCBI Datasets is an experimental project for finding and building datasets

[NCBI Datasets](#) [Command-line tool](#) [API documentation](#)

3 Drosophila X RefSeq X

	Species	Assembly	Annotation	Level	Contig N50	Size	Year
	<i>Drosophila arizonae</i>						1 assembly
☒	Isolate: ariz_Son04	ASM165402v1 reference RefSeq: GCF_001654025.1	NCBI Release 100	Scaffold	15 kb	141.387 Mbp	2016
	<i>Drosophila hydei</i>						1 assembly
☒	Strain: 15085-1641.00,03,60	DhydRS2 reference RefSeq: GCF_003285905.1	NCBI Release 101	Contig	3,367 kb	153.742 Mbp	2019
	<i>Drosophila mojavensis</i>						1 assembly
☒	Strain: TSC#15081-1352.22	dmoj_caf1 reference RefSeq: GCF_0000005175.2	NCBI Release 101	Scaffold	122 kb	193.826 Mbp	2006

NCBI BLAST

- RefSeq annotations included in nr (protein) and nt (nucleotide) databases
- Or use RefSeq databases
 - Smaller, faster, better signal
 - Can further restrict by taxonomy
- Organism specific BLAST databases also available

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

The screenshot shows the NCBI BLAST homepage. At the top, the URL <https://blast.ncbi.nlm.nih.gov/Blast.cgi> is displayed in the browser's address bar. The main heading is "Basic Local Alignment Search Tool". Below it, a description states: "BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance." A link "Learn more" is provided. To the right, a "NEWS" sidebar contains a "Search Betacoronavirus Database" announcement, mentioning the creation of a new BLAST database for SARS-CoV-2 sequences, dated Mon, 03 Feb 2020 10:00:00 EST, with a link to "More BLAST news...". The "Web BLAST" section features three main options: "Nucleotide BLAST" (nucleotide to nucleotide), "blastx" (translated nucleotide to protein), and "tblastn" (protein to translated nucleotide). To the right of these is "Protein BLAST" (protein to protein). Below this, the "BLAST Genomes" section has a search bar with "Drosophila a" entered, a "Search" button, and a dropdown list of suggestions including "Drosophila ananassae (taxid:7217)", "Drosophila ananassae Doleschall, 1858 (taxid:7217)", "Drosophila annanassae (taxid:7217)", "Drosophila americana (taxid:40366)", "Drosophila arizona (taxid:7263)", and "Drosophila arizonae (taxid:7263)". At the bottom left, a link for "Standalone and API BLAST" is visible.

We want to hear from you!

- This work was done as part of the Authors' official duties as NIH employees and is a Work of the United States Government. Therefore, copyright may not be established in the United States. 17 U.S.C. § 105
- This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.
- The annotation produced for RefSeq is within the public domain. Please cite the assembly providers and NLM/NCBI (PMID:[26553804](#)) if using these data.
- Send questions and feedback to:
 - RefSeq: <https://www.ncbi.nlm.nih.gov/projects/RefSeq/update.cgi>
 - Annotation requests: <https://support.nlm.nih.gov/support/create-case/>
- Watch NCBI News for updates!
 - <http://www.ncbi.nlm.nih.gov/news/>
 - <https://www.youtube.com/user/NCBINLM>

Thank you!

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

RefSeq/Gene

Eric Cox	Shashi Pujar
Catherine Farrell	Bhanu Rajput
Tamara Goldfarb	Lillian Riddick
Diana Haddad	Barbara Robbertse
John Jackson	Brian Smith-White
Vinita Joardar	Pooja Strobe
Kelly McGarvey	Anjana Vatsan
Michael Murphy	David Webb

RefSeq Developers

Alex Astashyn
Olga Ermolaeva
Vamsi Kodali
Craig Wallin

Annotation Pipeline

Francoise Thibaud-Nissen
Wratko Hlavina
Paul Kitts
Avi Kimchi

Jinna Hoffman
Patrick Masterson
Eyal Mozes
Robert Smith
Alexandre Souvorov

GDV/Remap/Datasets/Orthologs

Valerie Schneider
Peter Meric
Nathan Bouk
Hsiu-Chuan Chen
Cliff Clausen
Nuala O'Leary
Sanjida Rangwala
Anatoliy Kuznetsov

A cast of thousands

Lukas Wagner
Andrei Shkeda
Donna Maglott
Kim Pruitt
Jim Ostell
David Lipman