

Student Handout

Fruit Fly Ethomics & Genomics

Summary of Laboratory Exercise

In this laboratory unit, students will connect behavioral phenotypes to their underlying genes and molecules in the model genetic system, *Drosophila melanogaster*.

Students will:

1. Observe and characterize how male and female fruit flies interact with each other in a mating arena by constructing an ethogram.
2. Query a particular mating behavior in the model organism database, FlyBase, to identify candidate genes that are known to code for a particular trait.
3. Describe detailed attributes of two chosen candidate genes, including their length, gene structure (e.g., exon/introns), number of transcripts, and expression location (e.g., tissues, embryonic stage, etc.)
4. Develop new functional hypotheses on how the two genes contribute to the phenotypic expression of their selected behavior.

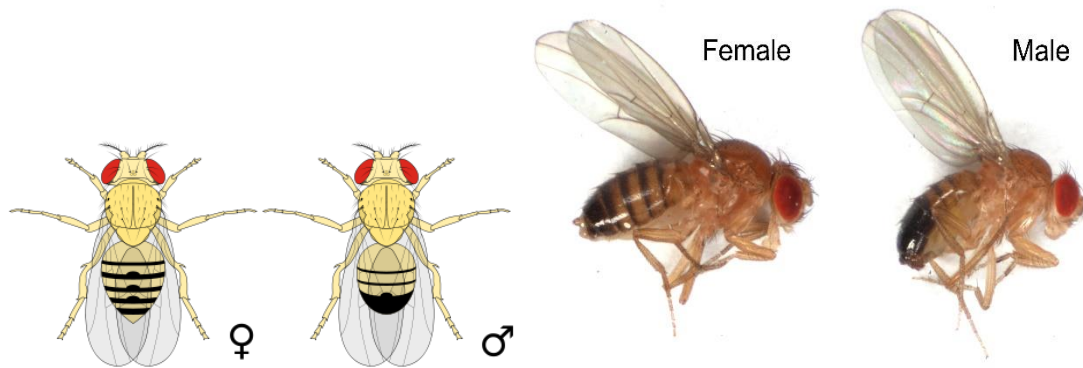


Figure 1. Pairs of *Drosophila melanogaster* (female on the left and male on the right).

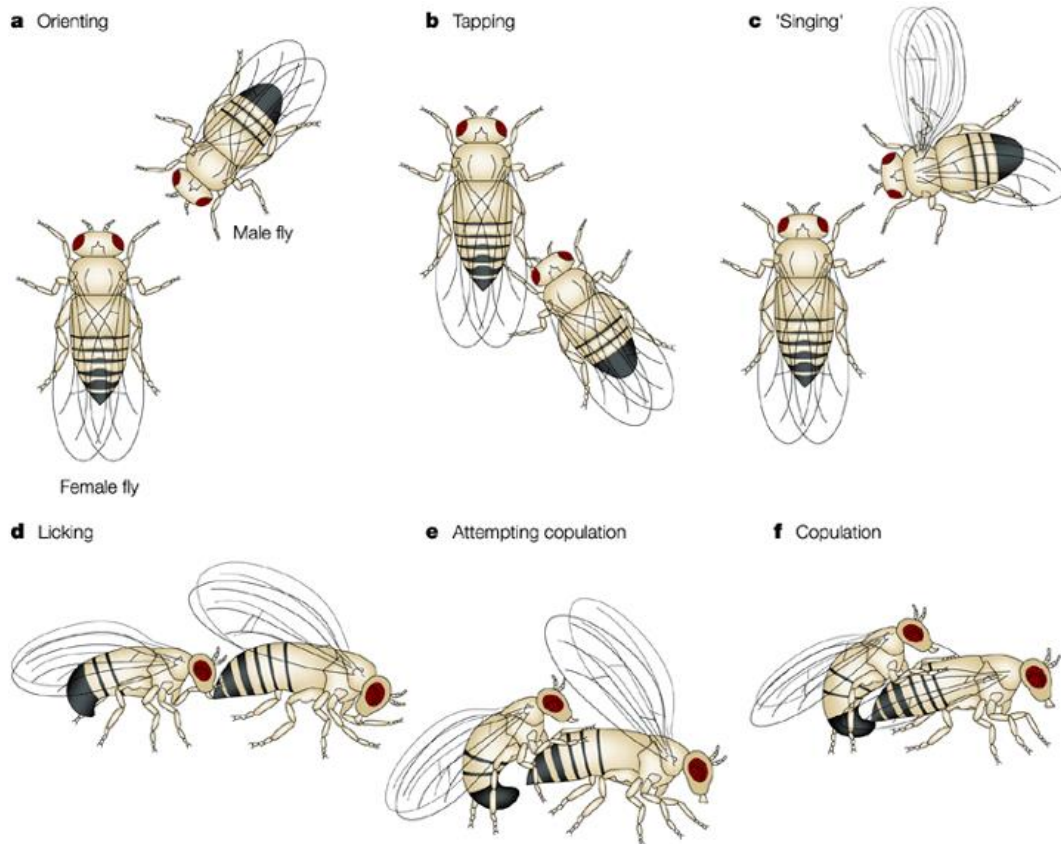


Figure 2. Fruit fly courtship behaviors (taken from Sokolowski 2001).

Behavioral Observation Worksheet (Part 1)

With your lab partner, open a browser on your computer and go to (or paste in) the YouTube URL that was assigned by your instructor. The 15 minute video will comprise of five males and five females. Be sure that you are able to distinguish between the sexes. You will be constructing a basic “ethogram” that quantifies behavioral activity in the arena by recording the frequency of successive events. Record any mating behaviors (refer to Figure 2) on the observation worksheet, below, including the frequency of each behavior (use ticks or check marks). Briefly define each of the behaviors after the observation period in your own words. If you observe a mating behavior that is not pre-defined in Figure 2, describe it and the number of observations in an unlabeled row. In the exercise, each behavior should be treated as a continuous event separated from other events or periods of inactivity. Gene Ontology (GO) IDs of each behavior are provided for Part 2.

Ethogram

	Behaviors	Frequency of behavior	Description of behavior
A	Orienting (GO:0016543)		
B	Tapping (GO:0016544)		
C	Male song (GO:0045433)		
D	Male licking (GO:0008049)		
E	Following/chasing (GO:0016546)		
F	Copulation (GO:0007620)		

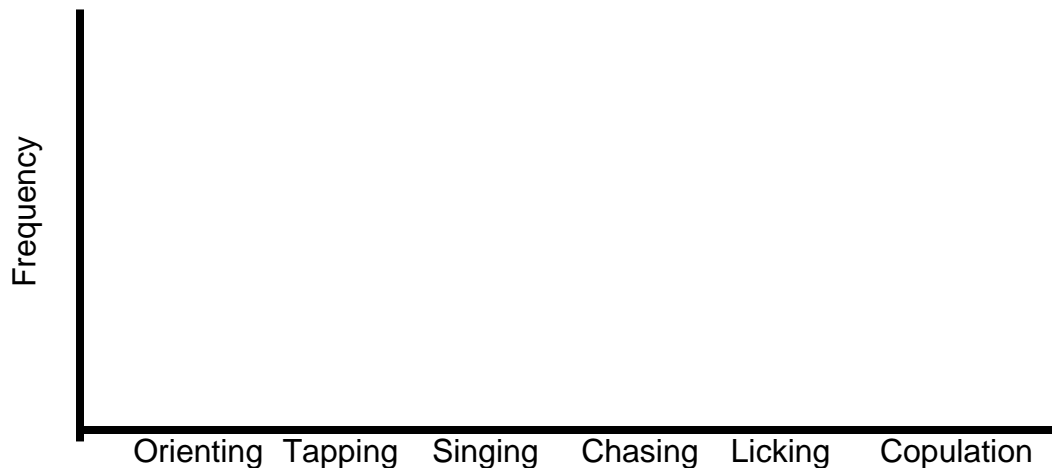
Share the final tally of your observations per behavior with your class on the blackboard.

Questions on courtship behaviors

i. What was the most frequent courtship behavior(s) observed in the flies?

ii. In this exercise, you recorded the frequency of events and not their duration time. Can you think of another method to quantify behavioral observations?

iii. In the space below hand-draw a bar graph that compares the frequency of mating behaviors that you observed in your 15 minute video.



iv. How reproducible were your observations? For example, do other groups have similar observations, in terms of the relative frequencies of behaviors? From the class results on the blackboard, calculate the average number of observations in each behavior per group (total observations per behavior divided by the number of lab groups) and add this to the above bar graph using another color.

v. Select a specific mating behavior from Figure 2 that you would like to further examine, in terms of its genetic basis. In the next part of this lab, you will identify genes involved in this behavioral phenotype and characterize their function from a molecular perspective.

Mating behavior: _____

Exploring FlyBase (Part 2)

FlyBase is an online biological database designed primarily for researchers studying the model genetic organism, *Drosophila melanogaster*, and related fruit fly species. FlyBase contains an enormous number of tools and data, and it is very easy to get disoriented. Using a morphological phenotypic mutation, we will first walk you through FlyBase to demonstrate how to quickly navigate through its remarkable array of resources. At the end of this exercise, you will select one of the courtship behaviors you observed in the previous exercise and use FlyBase to characterize two of its “candidate” genes, i.e., genes that may contribute to a particular behavioral phenotype. Once you complete your Candidate Gene Worksheet (last page of this handout), we encourage you to further explore FlyBase and its many links.

To identify genes that encode a particular phenotype, we will search FlyBase’s library of functional ontologies. These gene ontologies have been defined by experts in the community and provide geneticists with a common dictionary of phenotypes that can be connected to individual genes via richly curated data from *Drosophila*’s extensive genetic and phenotypic literature base. In today’s lab, we will search ontological classes based on known genetic mutations.

1. Go to FlyBase

On your computer, open a web browser and go to “www.flybase.org”. The FlyBase homepage should look like the image on the right.

In Part 1, you identified a behavioral phenotype that you wanted to study further. However, in this demonstration, we will focus on a known complex morphological trait: **the lack of female genitalia**. Once you are done, your instructor will provide an answer key based on the latest version of FlyBase (FB2015_04): Of course, due to regular FlyBase updates, values will change.

The process for identifying the genes responsible for this particular phenotypic trait will be similar for addressing the genetic basis of any fruit fly trait you wish to study via FlyBase.

The screenshot displays the FlyBase homepage. At the top, it features the FlyBase logo and the text "A Database of *Drosophila* Genes & Genomes". Below this is a navigation bar with links for Home, Tools, Downloads, Links, Community, Species, About, Help, Archives, and a "Jump to Gene" search box. A secondary row of tools includes BLAST, GBrowse, QueryBuilder, RNA-Seq Profile, GO PHENOTYPE ANATOMY DISEASE MORE, ImageBrowse, and Batch Download. The main content area is divided into several sections: "Fast-Track Your Paper", "FlyBase Newsletter", "Fly Board & White Papers", "News" (with links to NAR articles, 2016 Fly Board election, and 2015 releases), "Upcoming Meetings" (2016 Creole Fly Meeting, 57th ADRC), and "Courses" (GEP Workshop, W.T Dros Genetics/omics). A "QuickSearch" box is prominently displayed, allowing users to search by Species, Protein Domains, Gene Groups, Human Disease, Data Class, Expression, Phenotype, or References. A "Commentary" section highlights a "Fast-Track Your Paper tool upgrade" from September 3, 2015. The footer contains support information from the National Human Genome Research Institute, the British Medical Research Council, the Indiana Genomics Initiative, and the National Science Foundation, along with the version number (FB2015_04, released September 3, 2015) and social media icons.

2. Choose the correct ontological category and associated controlled vocabulary

On the **FlyBase** homepage, click the “**Vocabularies**” toolbox at the top of the page (it is the box labeled with the text “GO, Phenotype, Anatomy, More”). This will bring you to the “**Vocabularies Search Page**”.

Enter a search term

CV Hierarchy: All

Enter text:

Note: Consider using wild cards (*) and/or singular search terms.

Or browse the following hierarchy structures:

- Anatomy (FBdv)
- Biological Process (GO)**
- Cellular Component (CC)
- Molecular Function (GO)
- Human Disease (DO) NEW
- Molecular Interaction (MI)
- Development (FBdv)
- Allele Class (FBcv)
- Phenotypic Class (FBcv)
- Origin of mutation (FBcv)
- Stock descriptor (FBsv)
- Publication descriptor (FBcv)
- Imaging method (FBbi)
- Sequence ontology (SO)
- Gene Class (SO)
- Chromosome Structure Variation (SO)
- Sequence Variant (SO)

Biological Process (GO)

- | [biological adhesion](#)
- | [biological regulation](#)
- | [cell killing](#)
- | [cellular process](#)
- | [death](#)
- | [developmental process](#)
- | [establishment of localization](#)
- | [growth](#)
- | [immune system process](#)
- | [localization](#)
- | [locomotion](#)
- | [metabolic process](#)
- | [multi-organism process](#)
- | [multicellular organismal process](#)
- | [pigmentation](#)
- | [reproduction](#)
- | [reproductive process](#)
- | [response to stimulus](#)
- | [rhythmic process](#)

On the left hand side, select the “**Biological Process (GO)**” radio button (indicated, in the above figure, in the green oval). This particular ontological class accesses Drosophila Gene Ontologies (GO), a well-curated dictionary of fruit fly phenotypes. Once you select a particular ontological class, there are many ontological subcategories, also known as Controlled Vocabularies (CVs), that you could choose from and they are organized in a hierarchical “**Spanning Tree**”. Each CV will be associated with genes that have been curated from the Drosophila community’s extensive literature. Navigating through the submenus will reduce the number of gene hits for each term until you find a particular gene of interest.

On the right hand side of the page (in the grey box indicated by the red arrow in the above figure), you will need to choose the term that best describes the phenotype that we wish to study its genetic basis, in this case, a female fly lacking genitals.

Scanning the list of available ontological terms under “**Biological Process (GO)**”, what do you think is/are the most appropriate term(s) to click with respect to this particular morphological phenotype of interest?

(a)

The screenshot shows the FlyBase website interface. At the top, there is a navigation bar with links for Home, Tools, Downloads, Links, Community, Species, and About. Below this is a section for 'General information' with fields for Term (developmental process), Definition, and Comment. The 'Annotations' section shows a table with 'Data Class' (Genes (FBgn)) and 'Field' (GO_BIOLOGICAL_PROCESS). Below this is a section for 'Records which annotation includes this term OR any of its CHILDREN TERMS' with a grid of colored boxes: Insertions (147), Genes (5086, circled in red), Constructs (40), Images (216), and Alleles (6). At the bottom, there is a 'Spanning Tree (Parents/Children)' section with a dropdown menu set to 'is_a' and a 'Search All' button.

The absence of female genitalia phenotype would likely be under “**developmental process**”. Click this term and the “**Term Report**” page appears (see figure on left). The teal blue boxes” provide the number of records available for this parent CV and any of its child ontologies. Clicking on the “**Genes**” box (circled in red on the lefthand figure) will bring you to an alphabetized list of all the genes that are associated with this term. Scroll down past the numbered genes.

```

biological_process
|_ developmental_process 5495 rec.
|   |_ abscission
|   |   |_ cell_wall_modification_involved_in_abscission(+)
|   |   |_ cell_wall_polysaccharide_catabolic_process_involved_in_abscission
|   |   |_ floral_organ_abscission(+)
|   |   |_ fruit_abscission
|   |   |_ leaf_abscission
|   |   |_ plant-type_cell_wall_loosening_involved_in_abscission

```

How many genes are associated with “development process”?

(b)

List three gene names with their associated symbols.

(c)

Obviously, here these are too many gene records to individually sort through. So, let’s further narrow our search. Navigate back to the previous page, using the browser’s back button. Below the records box, you will see a hierarchical “**Spanning Tree**”, a very long list of CVs that are classified as “development process” (see above figure). You will have to use some logic and thought to narrow things down, but in the interest of time, do the following: under the child term “**anatomical structure development**” (since the genitals appeared to never have formed correctly this was probably a developmental process) you should see the term, “**analia development**”. Click on this term. Now, click the term, “**female analia development**”. There should be ~2 genes: certainly a much more manageable list of genes to explore.

3. Choose a gene(s) that codes for this particular phenotype

Click on the “**Genes**” box. The genes, “*doublesex*” and “*ken and barbie*” should be listed. From their gene names, guess which gene has a phenotype that does not develop external genitalia.

FB2015_04, released September 3, 2015

Gene Dmelldsx

Home Tools Downloads Links Community Species About Help Archives Jump to Gene Go

FlyGene Wiki Help Open All Close All

General Information	
Symbol	Dmelldsx (1)
Species	<i>D. melanogaster</i>
Name	doublesex
Annotation symbol	CG11094
Feature type	protein_coding_gene
FlyBase ID	FBgn0000504
Gene Model Status	Current
Stock availability	29 publicly available
Also Known As	doublesex, Dmdsx

Genomic Location	
Cytogenetic map	84E5-84E6
Sequence location	3R:7,924,323..7,967,408 [-] (2)

Genomic Maps

GBrowse

Decorated FASTA
Get genome region

Gene region
Get FastA

Families, Domains and Molecular Function	
Gene Group Membership (FlyBase)	
Protein Family (UniProt, Sequence Similarities)	-
Protein Domains/Motifs	<p>UniProt (Sequence Similarities)</p> <p>Contains 1 DM DNA-binding domain. {ECO:0000255 PROSITE-ProRule:PRU00070}. (P23023)</p> <p>InterPro</p> <p>DM DNA-binding domain; Doublesex dimerisation; DMRT/protein doublesex/protein male abnormal 3</p>
Molecular Function (see GO section for details)	<p>Experimental Evidence</p> <p>protein binding; protein homodimerization activity; RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription; RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription</p> <p>Predictions/Assertions</p> <p>DNA binding; mRNA binding; sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity; zinc ion binding</p>

Click the *ken* gene. *ken* is the gene symbol for *ken and barbie*. Above is their “**Gene Page**”. There is an enormous amount of information about that gene here including (labeled in red in the figure above): the species and symbol (1), sequence location (2), and a map of the immediate genomic region (3), evidence-based molecular functions (4) and a summary of the gene’s known functionality are also included if you open the boxes below the Genomic Map. Refer back to your observational data to help understand the molecular basis of the chosen behavior.

regulation of transcription

Predictions/Assertions
DNA binding; mRNA binding; sequence-specific DNA binding; sequence-specific DNA binding transcription factor activity; zinc ion binding

Gene Ontology (GO): (42 terms)

- Molecular Function (9 terms)
- Biological Process (32 terms)
- Cellular Component (1 term)

Summaries

Automatically generated summary

See sections below for more information

The gene *doublesex* is referred to in FlyBase by the symbol *Dme/ldsx* (CG11094, FBgn0000504). It is a protein_coding_gene from *Drosophila melanogaster*. It has **6 annotated transcripts** and **6 polypeptides** (3 unique). Gene sequence location is **3R:7924323..7967408**. It has the cytological map location **84E5-84E6**.

- Protein features** are: DM DNA-binding domain; DMRT/protein doublesex/protein male abnormal 3; Doublesex dimerisation.
- Its **molecular function** is described by: sequence-specific DNA binding; protein homodimerization activity; RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription; protein binding; mRNA binding; zinc ion binding; DNA binding; RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription; sequence-specific DNA binding transcription factor activity.
- It is involved in the **biological process** described with 32 unique terms, many of which group under: gamete generation; taxis; male courtship behavior; veined wing vibration; male sex determination; organic cyclic compound biosynthetic process; RNA processing; developmental pigmentation; male courtship behavior; sex-specific pigmentation; sensory organ development; post-embryonic morphogenesis; regulation of multicellular organismal process; negative regulation of developmental pigmentation.
- 140 alleles are reported.**
- The **phenotypes** of these alleles manifest in: genital disc; embryonic/first instar larval cuticle; trunk mesoderm derivative; compound sense organ; gonadal sheath proper primordium; segmental subdivision of integument; reproductive system; germline cell; prothoracic metatarsus; male accessory gland; developing embryonic structure.
- The **phenotypic classes** of alleles include: phenotype: increased mortality during development; increased mortality; fertile; sterile.
- Summary of **modENCODE Temporal Expression Profile**: Temporal profile ranges from a peak of moderate expression to a trough of extremely low expression. Peak expression observed during early pupal stages, in adult male stages.

Phenotypic Descriptions of Classical Alleles

Click open the “**Summaries**” tab (circled in red, above) and you will be presented with curated information about this particular gene. For our purposes, the most informative section is the “**Phenotypic Description of Classical Alleles**” button at the bottom of the first summaries box (also circled in red, above) presenting a compilation of the extensive and curated *Drosophila* literature. Click it.

From the enormous literature generated by the *Drosophila* community over the last century, observed mutant phenotypes have been collected and curated by FlyBase. A separate description for each observed mutant phenotype of *ken* is listed in each row. In *Drosophila*, an allele is usually denoted in superscript after the gene symbol but in databases with character type restrictions, they are typically denoted in square brackets (e.g., *ken*[00628]). Alleles are specific versions of a gene. Geneticists typically generate non-functional gene versions to study their effect on the organism.

Summarize one of the *ken* phenotypes **in your own words** (pick any one):

(d)

How many different alleles of *ken* have been studied in the *Drosophila* literature? [Hint: count the rows or number of “Submitted IDs”] Remember, *ken* is only one of ~14,000 genes in *D. melanogaster*: lots of research has been performed on this insect.

(e)

4. Explore the molecular and functional features of this gene

Going back to the **Summaries** section, you should see a bullet that says “*The phenotypes of these alleles manifest in*”. This section lists the specific tissues where this phenotype is detected.

List the tissues in which this phenotype is expressed that supports your hypothesis of this gene’s function.

_____ (f)

Under the “**General Information**” section found at the top, we find basic data about the gene including the Gene **Symbol** which is represented as “species\symbol”. “D” stands for *Drosophila* and “mel” for *melanogaster*, and *ken* is the official gene symbol denoting the “*ken and barbie*” gene, i.e., Dmel\ken.

Why do you think this name was chosen for this gene?

_____ (g)

You should also see the **FlyBase ID** in this section, “FBgn0011236”. This represents the unique identifier assigned to each gene on FlyBase. You can use this FlyBase ID or the symbol to look up the gene if you need to refer to it later.

What type of gene is this? [Hint: look in “**Feature type**”]

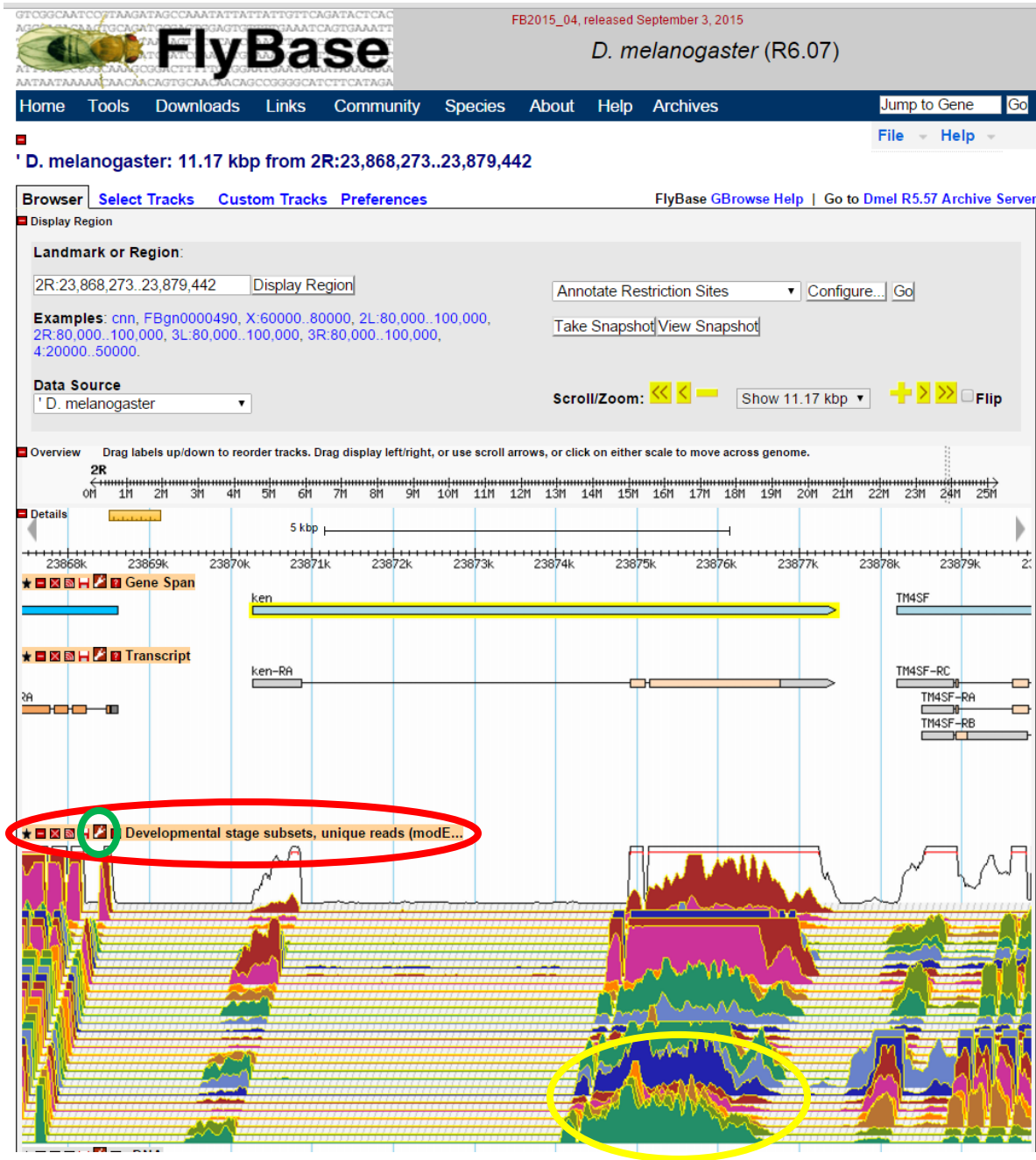
_____ (h)

Under the “**General Information**” section is the “**Genomic Location**” section. The first thing you should notice is the “**Sequence location**”. The *ken* gene is on the right arm of the second chromosome (2R) on the positive strand (+).

How large is the *ken* gene? [Hint: subtract the two numbers in “Sequence location”]

_____ (i)

Also displayed is a snapshot of a genomic map that details genes, their content, and their position in the genome. Next to the genomic map are links to download the actual gene sequence in FASTA format for other analyses (like a BLAST search). Perhaps, the most powerful tool is the “**GBrowse**” link on the left hand side of the gene map. This takes you to a genome browser to visualize the Dmel/ken locus. Click on it. Identify the **Landmark** section and the **Gene Span** and **Transcript** tracks.



The “**Transcript**” track provides gene models including UTRs (grey rectangle), exons (orange rectangle), and introns (black line). Draw *ken*’s gene model below.

(j)

Scroll down to the “**Developmental stages subsets...**” (in red oval, in figure above) tracks and you will see a multi-colored series of graphs that span this genomic region. Each of the color-coded graphs represents the gene expression level for a specific developmental stage. If you mouse-over the graphs (or click the little wrench in the track

label, circled in green in figure above), you will see the color-coded legend denoting transcriptomic levels from stages 0-2hrs in the embryo (em0-2hr) to the third larval instar (L3_12hr) to adult females 30 days after hatching (AdF_Ecl_30days). These graphs refer to the “**Temporal Expression**” of each gene/transcript and represent the unprecedented “big data” that is currently being generating in genomics. It is easy to see that *ken* is expressed in only certain developmental stages. Note that only exonic regions are expressed—compare the location of these graphs to the gene models found in the **Transcript** tracks.

For a better view, again, click on the little wrench in the track label located just above the transcriptomic tracks. When the settings menu pops up under the “**Signal scaling method**”, set it to “**linear**”. Hitting “**Apply Changes**” will close this box and redraw the graphs of gene expression as linear values. With these settings you should be able to see more clearly how expression levels change at different points during development. From this new view, which specific developmental stages have the highest and/or most consistent expression levels across exons? [Hint: three of them are circled in yellow, in the above figure.]

_____ (k)

Since the phenotypic trait we are studying (lack of female genitalia) is associated with development, it is most likely expressed during certain points in the life cycle of the fly. Using these data, we can see when in the life cycle a specific gene is expressed and get a better idea of its function. This information is available for all genes and, in fact, the entire genome of *D. melanogaster*. Using these expression data, what can you hypothesize about the function of the gene, *ken*, and its observed phenotype (lack of proper female genitalia)?

_____ (l)

Which gene is *ken*'s immediate 3' downstream neighbor, and how many alternative transcripts does this gene appear to have?

_____ (m)

Explore some of the other tracks. As you can see, we have so much fine-grained information concerning the genome of *fruit flies*. These little critters are considered pests to most non-geneticists yet, for over a century, they have been the go-to organism for genetic studies. Why do we spend so much time and energy to learn so much about this tiny insect? Use a few minutes to think about this question and then write down your thoughts in a few sentences.

_____ (n)

Take some time to explore the genomic landscape you see here by playing with the zoom size and then exploring the wide range of tracks. If at any point you get lost all you have to do is hit the back button on your web browser and it will take you back to the Gene Page. Then, simply navigate to the **GBrowse** page again. The tools and resources you see on FlyBase are typical of most model organism databases (MODs) as well as major human databanks such as NCBI and the UCSC Genome Browser, so you have now learned how to access the amazing world of genomic data!

Behavioral Candidate Gene Worksheet (Part 2)

Which behavioral phenotype (from Part 1) did you choose to study further?

Using FlyBase's extensive database of ontological "Vocabularies", identify a number of genes identified in "Biological Process (GO)" for your selected behavioral phenotype. Exploring phenotypic descriptions will help to evaluate whether a gene is a candidate for your behavioral phenotype of interest. If the description does not match your observations, it is not a strong candidate (at least, without further experimentation). However, if you see a description that matches your observations, then you have found an ideal candidate gene for your behavioral phenotype.

After you have identified **two** potential candidate genes, complete the worksheet. Include a brief hypothesis on how this gene may have molecularly contributed to your chosen courtship behavioral phenotype.

- Gene name: _____
 Gene symbol: _____ Genomic location: _____
 Gene size (bps): _____ Protein size (amino acids): _____
 Number of exons: _____ Number of transcripts: _____

Hypothesis on the molecular basis of behavioral phenotype:

- Gene name: _____
 Gene symbol: _____ Genomic location: _____
 Gene size (bps): _____ Protein size (amino acids): _____
 Number of exons: _____ Number of transcripts: _____

Hypothesis on the molecular basis of behavioral phenotype:

3. By looking at when and where these genes are expressed, is it possible that these two genes (or their expressed proteins) interact to generate the mating behavior?
