

**1 Examine genomic neighborhood surrounding target gene in *D. melanogaster***  
Use [Genome Browser](#) to view target gene in *D. melanogaster* (Assembly: Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6))

Annotation Form  
Take screenshot, and make a sketch, of the genomic neighborhood

1. Identify orientation (+/- DNA strand) of target gene and its nearest genomic neighbors in *D. melanogaster*
2. Once you identify the two neighboring genes on each side of your target gene, the arrows on the target gene will determine upstream (before, 5') or downstream (after, 3')

**2 Identify genomic location of ortholog in target species**  
Use the [Genome Browser](#) to retrieve protein sequence of target gene from *D. melanogaster* (click on an isoform in 'FlyBase Protein-Coding Genes' track; then click on 'Translated Protein from predicted mRNA')

Annotation Form  
1. Number of isoforms your target gene has in *D. melanogaster*  
2. Take screenshot of the "Descriptions" panel for the results of your *tblastn* search and summarize your findings  
3. Complete TABLE 1 with search results for the best scaffold match

1. Use 'Genome BLAST' link of target species genome to run *tblastn* search of *D. melanogaster* protein
  - **Query:** *D. melanogaster* protein
  - **Database:** target species' whole genome assembly
2. Identify best collinear set of protein alignments in target species-coordinates, scaffold name & accession

**3 Examine genomic neighborhood of putative ortholog in target species**

Annotation Form  
Complete TABLE 2 with the search results for the best *blastp* hits, take screenshot, and make a sketch, of the genomic neighborhood

1. Use target species' [Genome Browser](#) to navigate to the collinear set of alignments identified in Part 2
2. Use synteny to verify ortholog assignment
  - Run *blastp* search for target gene AND its two upstream and two downstream genes on either side
    - **Query:** accession of each protein in target species (click on isoform in 'NCBI RefSeq Genes' track; then, in the GenBank Record window, scroll down to the 'CDS' section and copy accession number for the translated protein sequence labeled 'protein\_id')
    - **Database:** refseq\_protein | Organism: *Drosophila melanogaster* (taxid: 7227)

**4 Determine structure of target gene in *D. melanogaster***  
Use the [Gene Record Finder](#) to identify the gene structure of the target gene in *D. melanogaster*

Annotation Form  
Take screenshot of Gene Record Finder showing mRNA details and CDS usage map. How many isoforms and unique coding exons does the target gene have in *D. melanogaster*? If there's more than one isoform, how many are unique?

**5 Determine approximate location of coding exons in target species**

Annotation Form  
Summarize *tblastn* CDS-by-CDS search results in TABLE 4. How many isoforms and unique coding exons does the target gene have in your target species? If there's more than one isoform, how many are unique? Explain any differences between *D. melanogaster* and your target species.

1. Obtain the protein sequence of each CDS of the target gene in *D. melanogaster* using the [Gene Record Finder](#) (in Polypeptide Details tab, click on each CDS and then copy/paste the entire sequence from the pop-up window)
2. Run *tblastn* search (check the box to align two or more sequences) for each CDS of the target gene
  - **Query:** each individual CDS of target gene in *D. melanogaster*
  - **Database:** accession number of scaffold containing ortholog in target species (identified in Part 2) and narrow region of scaffold to search via "Subject subrange"
  - **Algorithm parameters:** "Compositional adjustments"- No adjustment and uncheck box for filtering "Low complexity regions"

**6 Refine coordinates of coding exons**  
Use the [Genome Browser](#) to:

Annotation Form  
Complete TABLE 5 with the refined coordinates of your gene model for each unique isoform

1. Verify start and stop codon coordinates identified in Part 5
2. Determine phases of donor and acceptor splice sites using Frame identified in *tblastn* CDS-by-CDS search in Part 5; donor is typically a "GT" and acceptor is typically a "AG" (Georgia Tech is in Atlanta Georgia)
3. Use splice junction predictions to verify coordinates of each intron

**7 Verify and submit gene model**  
Use the [Gene Model Checker](#) to verify that your proposed gene model satisfies the basic biological constraints (e.g., begins with a start codon, has compatible splice sites, and ends with a stop codon)

Annotation Form

1. Take screenshot of "Configure Gene Model" section
2. Take screenshots and describe anomalies of the dot plot and protein alignment
3. Take screenshot of your gene model as shown on the Genome Browser, including the following tracks (in addition to "default" tracks): RNA-Seq tracks for at least one transcript prediction track (e.g., TransDecoder Transcripts) and at least one splice-site prediction track (e.g., Combined Splice Junctions), and a comparative genomics track for *Drosophila melanogaster* (e.g., Drosophila Chain/Net)
4. Download the GFF, fasta, and pep files of your proposed gene model

If you have more than one unique isoform, repeat verification process and Step 4 for each, and then merge the 3 file types using the [Annotation Files Merger](#).

Final three files you submit should have the following names: Use your assigned species ("d" followed by the first three letters of your species) + "\_gene.filetype" as the format. For example, if you're annotating Ilp8 in *D. grimshawi*, the merged GFF would be titled "dgril\_lip8.gff"