



Module 5: Translation

Carina Endres Howell & Leocadia Paliulis

Objectives

- Determine the codons for specific amino acids and identify reading frames by looking at the Base Position track in the genome browser
- Assemble exons to maintain the open reading frame (ORF) for a given gene
- Define the phases of the splice donor and acceptor sites and describe how they impact the maintenance of the ORF
- Identify start and stop codons of an assembled ORF

Prerequisites

- “Understanding Eukaryotic Genes” Modules 1–4: <https://thegep.org/ueg/>
- Overview of the ribosome, tRNAs, and associated proteins involved in translation (Initiation Factors, Elongation Factors and Release Factors)
- Overview of the DNA codon table

Class Instruction

- Review the process of translation: Overview of the ribosome, tRNAs, and associated proteins involved in translation (Initiation Factors, Elongation Factors and Release Factors)
- Review the DNA codon table
- Work through the activities using the Genome Browser, with pauses to discuss the answers to the questions.
- Conclude with emphasis on main points:
 - mRNAs are translated into amino acids using triplet codons.
 - Identification of ORFs.
 - The ORF must be maintained across splice sites to generate a working mRNA.
 - The assembled ORF begins with a start codon and ends with a stop codon.

Associated Videos

- Splicing and Phase Video: <https://youtu.be/9VWjyu3PUJO>

Table of Contents

Investigation 1: Examining Open Reading Frames (ORFs) in the tra gene..... 2

Investigation 2 5

Investigation 1: Examining Open Reading Frames (ORFs) in the *tra* gene

Introduction: Review of reading frames.

In this exploration, we will continue to focus on the *transformer* gene (referred to as *tra*-RA or just *tra*), and will learn about how the *tra* mRNA is translated into a string of amino acids.

Given that DNA is double-stranded, and that the genetic code is based on triplets (3 consecutive bases), there are six possible reading frames. One can determine a reading frame by dividing the sequence of nucleotides in DNA or RNA into a set of consecutive, non-overlapping triplets. There are three possible reading frames (read 5'→3') in the forward direction on the top strand of DNA, and three (read 5'→3') in the reverse direction on the complementary bottom strand of the same DNA molecule. Hence, there are six possible reading frames for each gene (see illustration in Module 1).

Once it is determined in which direction a particular gene is transcribed (for review see Modules 2 and 3 on transcription), there remain three choices for the reading frame. To determine which of these reading frames is used during translation, evidence such as the presence of an initiation codon and the absence of stop codons is used. As you learned in Module 1, the initiation codon is ATG in the coding DNA strand (AUG in the mRNA) and specifies the amino acid methionine. Additional triplets code for the other 19 amino acids, and three triplets are stop codons, causing termination of translation. These stop codons are TAA, TAG and TGA in DNA, or UAA, UAG and UGA when found in mRNA. An Open Reading Frame is a string of consecutive codons that is uninterrupted by stop codons. Every mRNA contains one ORF that is translated by the ribosome from start codon to stop codon.

Let's investigate reading frames for the *tra* gene.

1. Go to the *GEP UCSC Genome Browser Mirror* site at <https://gander.wustl.edu/> and follow the instructions given in Module 1 to open contig1 of *Drosophila melanogaster* using the July 2014 (Gene) assembly.
2. The screen below will appear (Figure 1). As you will remember, this section of DNA is 11,000 base pairs long and is a small part of the left arm of chromosome 3, which is about 28,100,000 bp long.

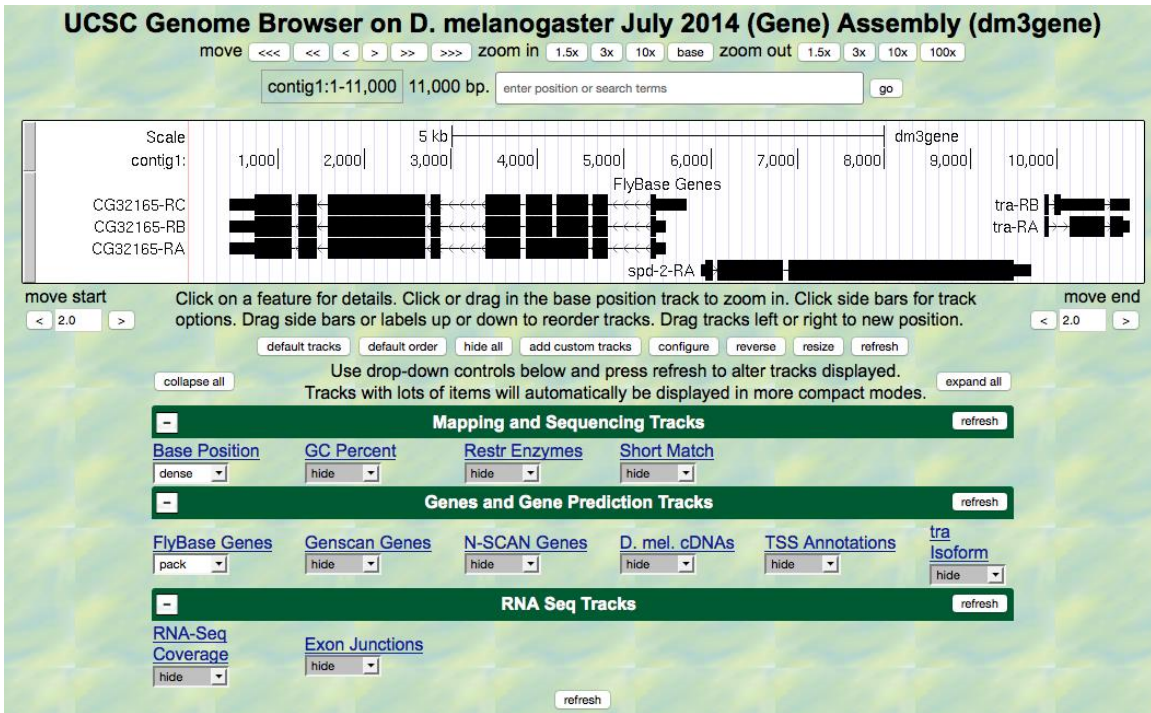


Figure 1 View of the region of the *tra* gene on chromosome 3.

3. Zoom in to view only the first exon of the *tra*-RA gene by entering “**contig1:9,840-9,900**” in the “enter position or search terms” text box and hitting the go button.
4. Open only the tracks that will provide information for this investigation. Set the Base Position track to “full,” and click “refresh.” Now we can see the amino acid tracks as well, giving the results from conceptual translation.
5. There are three possible reading frames for this transcript in the forward direction, here indicated by the numbers 1, 2 and 3 (red arrow) (Figure 2).

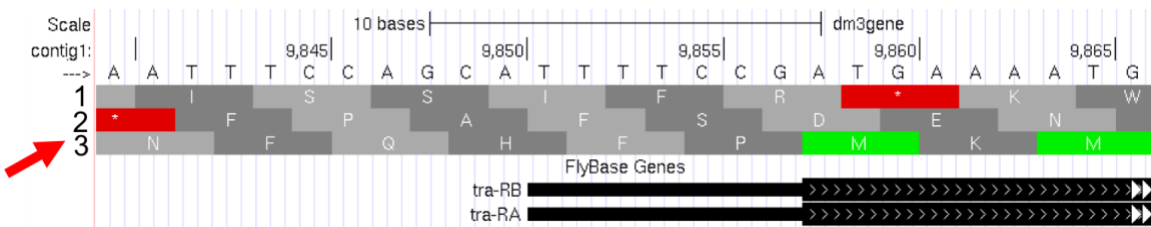


Figure 2 Three possible reading frames for the *tra* gene in the forward direction.

6. Three reading frames are possible in the forward direction, as one could start translating with the first base, the second base or the third base. In this case, starting translation with the first base shown in Figure 2 (i.e., position 9,840 in contig1) leads us to the reading frame 3. Starting with the second base leads us to the reading frame 1, and starting with the third base lead us to the reading frame 2. Starting with the fourth base is equivalent to starting with the first base, only missing the first codon. The DNA “top strand” is read from left to right (indicated by the arrow in the browser, right under the word “contig1” that looks like this --->). If you click on that arrow, the three reading frames in the reverse direction will appear, as the “bottom

strand” is read from right to left. As you learned in Module 1, genes have directionality. The *tra* gene is read from the “top strand” (left to right).

7. Notice that the third reading frame has a green M codon (methionine) at the location where the thick black rectangle indicates the first coding exon or CDS (Coding DNA Sequence) of *tra*-RA mRNA. Remember that the codon for methionine (ATG in DNA) is the start signal, the first codon used in translation. This gives us our first piece of evidence that reading frame 3 is the one used in translation of the first CDS of the *tra* gene. For simplicity let’s call this first CDS “CDS1” to distinguish it from other CDSs in the *tra* gene. Note that there is a stretch of RNA transcript upstream (to the left) of the ATG; this is the 5’UTR (5’ untranslated region), found at the 5’ end of all eukaryotic mRNAs.
8. Next carefully examine reading frame 2. Notice that in this reading frame there is no codon for methionine (no start codon) in the region that maps to the first exon. This gives us evidence that reading frame 2 is probably not being used during translation of CDS1 of the *tra* gene.
9. Finally, look at reading frame 1. Notice that there is a stop codon at the beginning of that reading frame (indicated by a red box with an asterisk in it). This evidence indicates that reading frame 1 probably is not being used during translation of CDS1 of the *tra* gene.
10. Let’s move on to looking at the reading frames for exon 2 of *tra*-RA. Zoom in to view only the second exon of the *tra* gene by jumping to “**contig1:10,120-10,570**” using the “enter position or search terms” text box. Remember that both the RNA-Seq data and the cDNA data have been used to map the positions of exons.

Q1. First examine reading frame 1. Are there any stop codons in the reported exon? _____ If there are early stop codons, do you think this is the reading frame used during translation?

Q2. Examine reading frame 2. Are there any stop codons in this reading frame within the exon?

Q3. Examine reading frame 3. Are there any stop codons in the reported exon? _____

Q4. Using the evidence above, which reading frame maintains an Open Reading Frame (ORF) across exon 2 of *tra*-RA? _____ Is this the same reading frame as that used for exon 1?

11. Finally, take a look at exon three (contig1:10,600-10,850). We anticipate that since this is the last CDS there will be one or more stop codons, the first of which will mark the site of translation termination. The 3’ UTR (3’ untranslated region) extends downstream from this point to the site of poly(A) addition where the last exon ends (see Module 3). Here all three reading frames have an ORF followed by one or more stop codons in the exon.

How can we figure out which reading frame is correct? We will investigate this in the next section by looking specifically at the splice junction.

Investigation 2

We can combine what we know about reading frames with what we know about splicing to learn exactly how tra-RA is put together. We'll note where the start codon, splice sites, and stop codon are so we can construct a gene model. Then, in Module 6, we'll use these same types of information to solve some mysteries about tra-RB.

1. Using the same Genome Browser page, reset the Browser by clicking on "hide all." Then open the tracks that will provide the information we want for Investigation 2.
 - **Base Position "full"**
 - Note that you will not be able to see the DNA sequence or amino acid tracks until you zoom in.
 - **FlyBase Genes "pack"**
 - **RNA-Seq Coverage "full"**
 - You will see blue and red histograms representing the RNA-Seq data (indicating the amount of mRNA synthesized) in females and males, respectively. We will focus on the blue histogram (Adult Females) again.
 - **Exon Junctions "full"**
 - These rectangular boxes joined by a thin black line will help us identify the exon-intron boundaries.
2. As we did in Module 3, let's customize the RNA-Seq track by setting the "Data view scaling" field to "use vertical viewing range setting" and the "max" "Vertical viewing range" to 37. (Remember that you gain access to these settings by clicking on the "RNA-Seq Coverage" label under the "RNA Seq Tracks" green bar found in the bottom section of the page.) Under the "List sub-tracks" section, unselect the "**Adult Males**" track.
3. Let's find the start codon for tra-RA. Zoom in on where the FlyBase Genes track shows that the translation starts (where the tracked black box gets thicker for the tra-RA isoform) as seen in Figure 3.

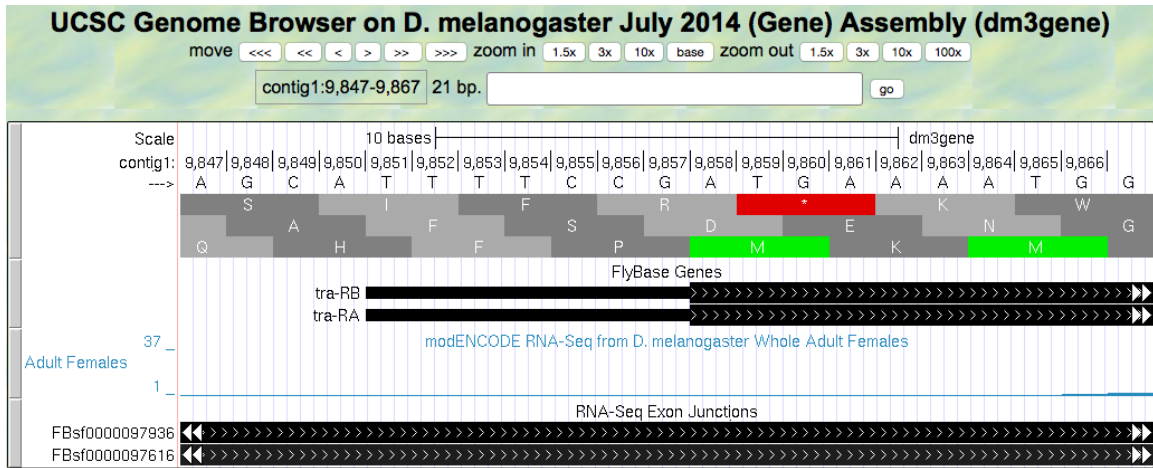


Figure 3 Translation start region of the *tra* gene.

Q5. Give the coordinates for the entire start codon for tra-RA (start codon coordinates should be three consecutive numbers, for example: nucleotides 212-214). _____

Q6. Which reading frame should we follow along to see the predicted amino acid sequence of tra-RA?

Q7. Zoom out to see the entire exon. Are there any stop codons in this reading frame in the first exon?

4. Now zoom in and find the last base of the first exon for tra-RA using your RNA-Seq data and Exon Junctions data (Figure 4).

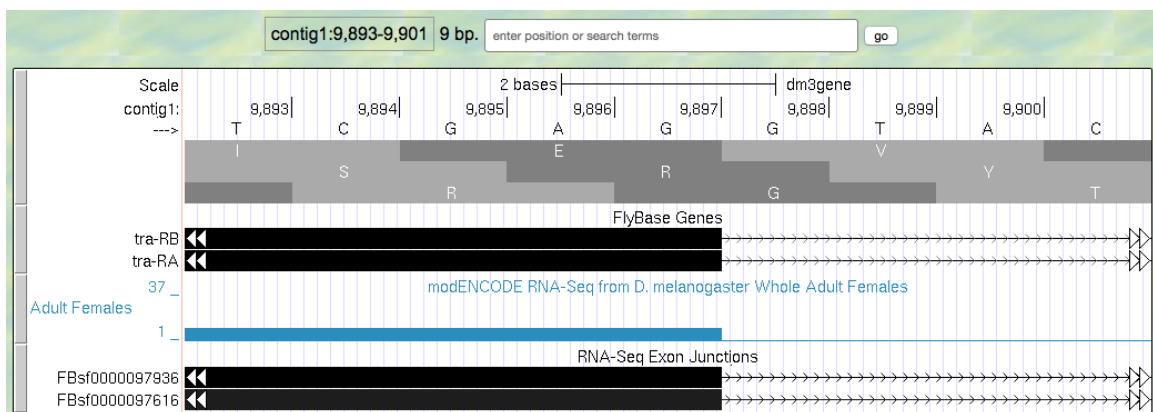


Figure 4 Region at the end of Exon 1 of the *tra* gene.

Q8. Give the coordinate for the very last base of the first exon. _____

So that we can follow the polypeptide through until we identify the stop codon, we need to figure out which reading frame we should follow in the second exon. This is not as easy as you might think, because in eukaryotes, different exons of the same gene may utilize different reading frames. You saw

an example of this previously in Module 1. Sometimes we can infer the correct reading frame given the pattern of start and stop codons within the region of the exon, identified by RNA-Seq data. But that sort of information does not always give a definitive answer — there may be more than one possible reading frame for a given exon. To figure out which reading frame is being translated at exon 2, we need to check the end of the first exon to see how many bases of the last codon are present before the 5' splice site consensus sequence. To do this, look closely at reading frame 3, just before the splice site (Figure 5).

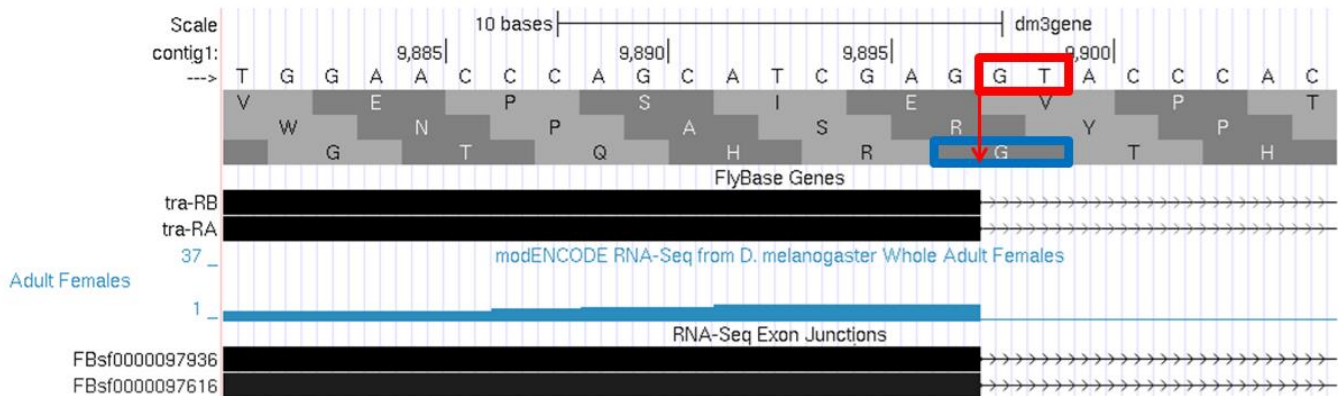


Figure 5 Region at the end of Exon 1 of the *tra* gene.

Note that the splice site (red box) cuts off the last codon of the first exon (blue box) after just one base. Therefore, we would say this exon has a **“phase 1” end because there is a partial codon at the end of the exon that is 1 base long.** (If there were a fully completed codon before the splice site, it would be in **phase 0**, and if there were two bases before the splice site, it would be in **phase 2**.)

For this exon with a phase 1 end, we will need two more bases from the next exon to complete the codon. Knowing this we can identify the reading frame that will be used in the second exon. Navigate to the 3' splice site of intron 1 (i.e., the location where the first intron ends and the second exon begins; Figure 6). To review splicing and the concept of phase, watch the [Splicing and Phase video](#).

Q9. Based on the evidence you see in the browser, give the coordinate for the first base of the second exon of *tra*-RA. _____

Q10. Do you observe an appropriate splice acceptor site just upstream within the intron? _____

Q11. Knowing that exon 1 ends with a partial codon of 1 base, what reading frame is being used in the second exon? _____

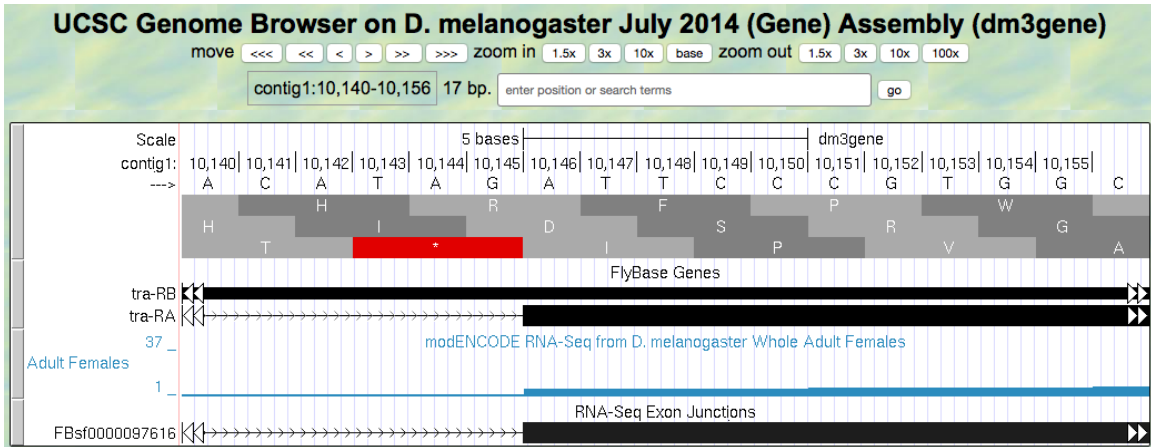


Figure 6 Region at the beginning of Exon 2 of the *tra* gene.

Now we will be using reading frame 2, because, after the splice site, there are two bases left in the codon. These two bases plus the one base left from the first exon make a complete codon.

5. Next, zoom out and look at reading frame 2 for all of exon 2 of *tra*-RA. You can see that there are no stop codons in this reading frame, which lends support to our conclusion that this is the proper reading frame.
6. Now, let's do the same for the 5' splice site of intron 2 for *tra*-RA. Zoom in on that splice site (Figure 7).

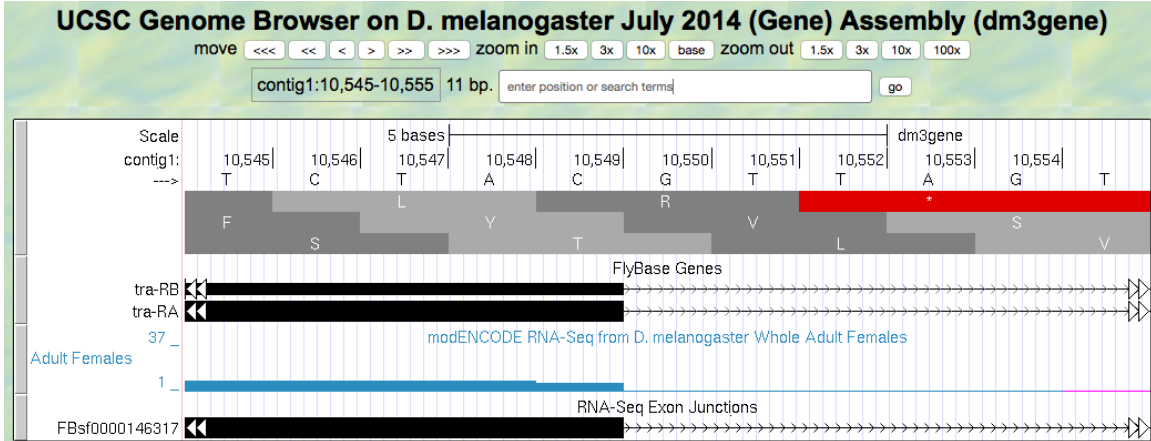


Figure 7 Region at the end of Exon 2 of the *tra* gene.

Q12. Give the coordinate of the base prior to the 5' splice site of intron 2. _____

Q13. How many bases are left in the codon before the splice site (i.e., is this phase 0, phase 1, or phase 2)? _____

7. Navigate to the start of the final exon (Figure 8).

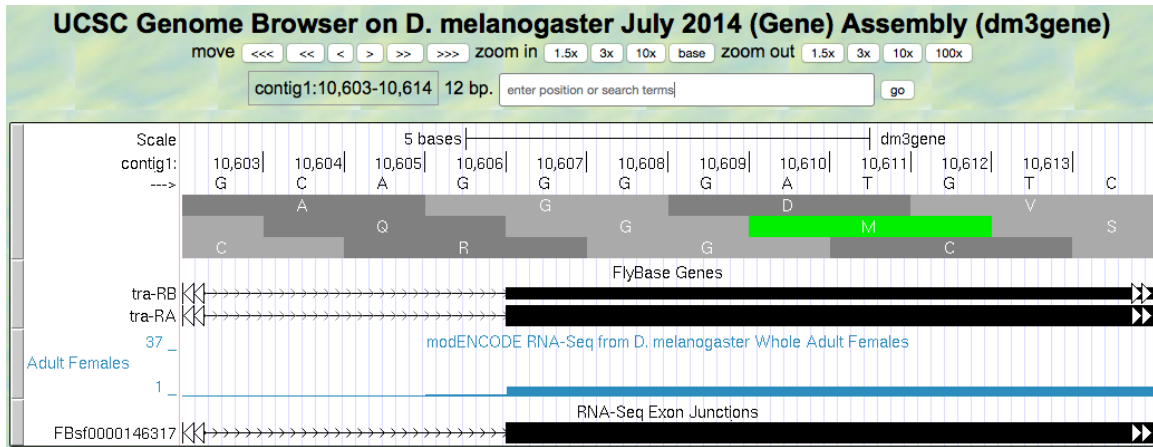


Figure 8 Region at the beginning of Exon 3 of the *tra* gene.

Q14. Locate the 3' splice site of Intron 2. Give the coordinate of the first base in exon 3 for tra-RA.

Q15. Which reading frame is being translated in the final exon? _____

- Now locate the first stop codon in the translated reading frame. Stop codons are shown as red boxes with asterisks (red arrows) as shown in Figure 9.

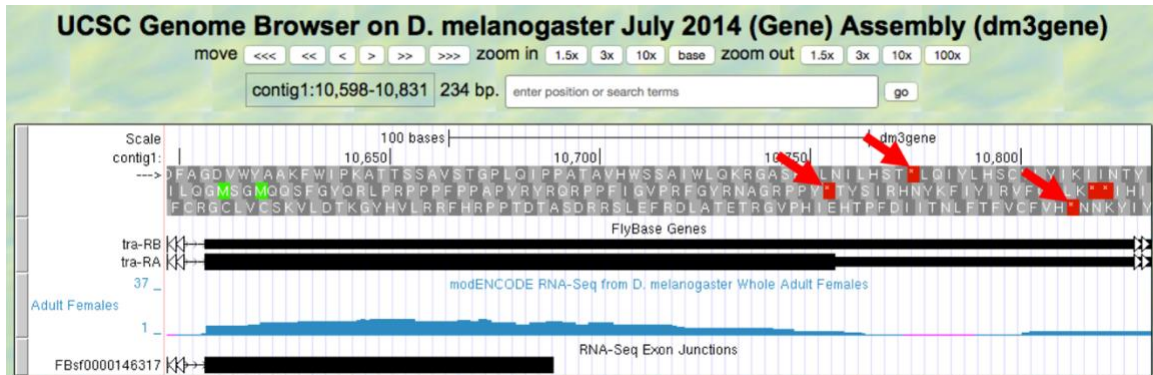


Figure 9 Region at the end of Exon 3 of the *tra* gene

Q16. Give the coordinates for the bases in the stop codon. _____

Q17. Let's consolidate all the data we found above in one place:

Gene model for tra-RA:

- Coordinates for start of translation: _____
- Coordinate for last base of exon 1: _____
- Coordinate for first base of exon 2: _____
- Coordinate for last base of exon 2: _____
- Coordinate for first base of exon 3: _____
- Stop codon coordinates: _____

Take the coordinate information above to draw a map of *tra*-RA using rectangles to represent exons and connecting lines to represent introns. Label the ends of the exons with the appropriate coordinates and indicate the transcription start site for the *tra*-RA initial transcript. Below this map, provide a map of the processed mRNA after intron removal. Below this map, indicate the regions that are translated into a protein. Give precise coordinates. Color coding may be helpful.

In Module 6, we will compare this model of *tra*-RA with a model of *tra*-RB.

Q18. To cement your knowledge of gene structure, you could construct a similar map of the *spd-2* gene. How many exons does this gene have? _____ How many introns? ____ How many isoforms? _____ Use the same approach to determine the coordinates for the exons, and the coordinates for the coding region (another name for the region that is translated).