



Parasitoid Wasps Project: Annotation Workbook Directions

Modified by Marisol Santisteban from Version 2.2 created by Marisol Santisteban & Charles Hauser from Version 2.1 created by Nate Mortimer, Melanie Van Stry, Rebecca Burgess, Michael Rubin, Rivka Glaser, & Chris Jones

Resources & Tools

- All links for this lesson can be found on the Parasitoid Wasps Project page of the GEP website (<https://thegep.org/wasps/>).

Table of Contents

Introduction	1
Workbook Instructions and Transcript Sheets	2
Feature:.....	3
First Base, Last Base: the first and last bases of the feature.....	4
Reading Frame:.....	4
Acceptor Phase and Donor Phase:.....	4
Checking Accuracy of the Gene Model (Alternative to Gene Model Checker)	6
Getting the Sequence Data for the pep and fasta files	9

Introduction

For the wasp venom gene annotation project, your project submission will consist of a completed annotation workbook, a screenshot of the user track on the genome browser, and sequence files for the entire transcript and encoded protein.

The annotation workbook is an Excel file with space to fill out all the necessary information for the gene annotation. If it is done correctly, filling out the workbook will also auto-populate a GFF file that can be used for testing the accuracy of gene models (described in the “An alternative to GMC” section below). Please note that everything should be filled out exactly as described: the vocabulary and formatting need to be very precise to make a functional GFF file.

You can download the annotation workbook, including an example venom gene annotation from the “Project Curriculum” section of the Parasitoid Wasps Project page (<https://thegep.org/wasps/>) on the GEP website.

Workbook Instructions and Transcript Sheets

The first sheet of the Excel document will have the instructions (Figure 1) and the third sheet has the template for inputting your data. Use a separate sheet within the workbook for each transcript of the gene that you annotate.

Gene Name	SP_Name	Found on the claim form			
Scaffold Name	scaffold_XXX	Name format will be scaffold_ followed by a number			
Your Name		How you want your name to appear in the gene annotation record (use _ to separate (no spaces))			
Strand	+ or -				
Transcript ID		IDs of additional transcripts from the wasp transcriptome track that are incorporated into the final gene model (one per column) or N/A			
Other IDs		Do any venom peptides align with the CDS of your model?			
Venom Peptides	Yes or no (Ganaspis or Leptopilina heterotoma or Leptopilina boulardi)	Input the species name here, follow the examples given			
Species					
Faculty		Enter the last name of your professor			
Feature	Left Base	Right Base	Reading Frame	Acceptor Phase	Donor Phase
Options are:			Top line: 0 Middle: 1 Bottom: 2	Phases for splicing . for UTRs	Phases for splicing . for UTRs
SUTR					
CDS					
3UTR					
Genome Browser tab:	You can use the information automatically generated in this tab to look at your gene model in the Genome Browser. Only use the lines that contain annotation information.				
Concat tab and GFF tab:	These tabs are for data extraction. You won't need to use them.				
Concat	Genome Browser	GFF			

Figure 1. The first sheet contains the instructions. The fields that will need to be filled out are highlighted in the above figure for illustrative purposes only.

Start by filling in the basic information on the third tab (Transcript). Be very careful to follow this guide for formatting.

- **Gene Name:** The name of the gene; this is given to you by your instructor with your assignment.
- **Scaffold Name:** This is found in the top left corner of the Genome Browser (Figure 2). Be careful to fill this out exactly (lower cases in scaffold, underscore rather than a space).
- **Your Name:** Write your name how you want it to appear in the gene annotation record with underscores to separate (no spaces) (e.g., John_Doe or John_C_Doe).
- **Strand:** Use + or – to indicate the DNA strand the gene is on.
 - Note: Do NOT use the transcriptome track to determine + or – strand. Use the other data (e.g., StringTie, gene predictors) instead.
 - Hint: type ‘+ (apostrophe +) or ‘- (apostrophe -) to prevent Excel from thinking you are typing a formula.
- **Transcript ID:** ID corresponding to the Species Transcriptome track. This is what you use to search for the gene and is given to you with your assignment (Figure 2).

- **Other IDs:** Any other transcripts from the Species Transcriptome track that form part of your final gene model. Many of the Transcriptome track gene models are incomplete so you may have to combine multiple transcripts to make a complete model.
- **Venom Peptides:** Indicate whether there are any peptides in the “Species Venom Peptides” track that align with the CDS of your gene.
- **Species:** The species of parasitoid that the transcript comes from; this is given to you by your instructor with your assignment.
- **Faculty:** Input the last name of your instructor.

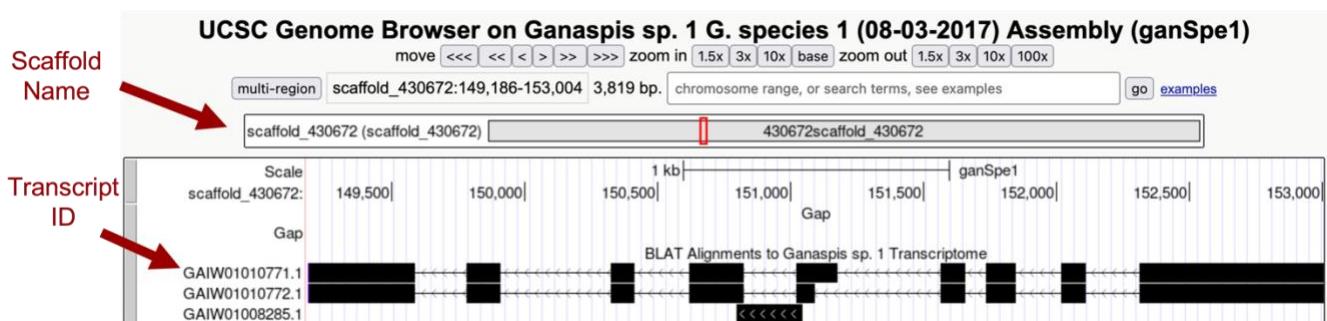


Figure 2. Parasitoid Wasps Genome Browser showing where to find the “Scaffold Name” and “Transcript ID” to input in your workbook. NOTE: You may have to set the “G1 Transcriptome” track to “full” in order to see the Transcripts IDs.

Then start annotating the gene as you learned previously. Add more rows to the annotation region using the insert command in Excel if you need more lines (and sorry that you have so many exons!).

Feature:

You will want to annotate the 5' UTR, CDS exons, and 3' UTR. Be sure to specify each feature by using:

Feature
Options are:
5UTR
CDS
3UTR

The above three features are the only options that will work for later analysis so use these names *exactly*.

Note: The transcriptome track does not attempt to differentiate between UTR and CDS. You will have to use what you've learned about these features to determine the UTR and CDS regions of your gene.

First Base, Last Base: the first and last bases of the feature

For 5UTR, the last base should be the final base before the ATG of the start codon. The first CDS will start with the A of the ATG and the final CDS will end with the final base of the STOP codon. The 3UTR then starts with the first base following the STOP.

Note: Regardless of which strand your gene is on, the first base should be a lower number than the last base. So, if you are annotating a gene on the minus strand the numbers will have to go in “backwards” so that the “first base” is the smaller number. You’ll use the strand field as described above to designate + or – strand.

Reading Frame:

Your options are 0, 1, 2. What we have referred to as Frame 1 (either +1 or -1) now becomes 0, Frame 2 becomes 1, and Frame 3 becomes 2. Because the 5UTR and 3UTR don’t have a reading frame, use a period(.) in these boxes.

Acceptor Phase and Donor Phase:

These fields are for you to keep track of your splicing phases as determined in your annotation. Because the 5UTR and 3UTR don’t have splicing phases, use a period(.) in these boxes. Also, the first CDS exon won’t have an acceptor phase and the last CDS exon won’t have a donor phase so use a period(.) in these boxes as well.

Once you have filled all the data for the gene in this example, the Transcript tab would look like this (Figure 3):

Gene Name	G1_PHGPx.right				
Scaffold Name	scaffold_430672				
Your Name	John_Doe				
Strand	+				
Transcript ID	GAIW01010771.1				
Other IDs	N/A				
Venom Peptides	Yes				
Species	Ganaspis				
Faculty	Mortimer				
Feature	First Base	Last Base	Reading Frame	Acceptor Phase	Donor Phase
5UTR	150642	150822	.	.	.
5UTR	151023	151100	.	.	.
CDS	151101	151176	2	.	1
CDS	151565	151658	0	2	2
CDS	151736	151845	2	1	1
CDS	152019	152110	1	2	0
CDS	152311	152526	0	0	.
3UTR	152527	153004	.	.	.

Figure 3. Workbook “Transcript” sheet filled in.

As you are filling out the data in the third tab of the Excel file (Transcript), the other tabs—Concat, Genome Browser, and GFF—will automatically populate (Figures 4, 5, and 6).

	A	B	C	D	E	F
1	gene_id "G1_PHGPx.right"					
2	transcript_id "GAIW01010771.1"					
3						
4	Feature	First Base	Last Base	Reading Frame	Acceptor Phase	Donor Phase
5	SUTR	150642	150822	.	.	.
6	SUTR	151023	151100	.	.	.
7	CDS	151101	151176		2	1
8	CDS	151565	151658		0	2
9	CDS	151736	151845		2	1
10	CDS	152019	152110		1	2
11	CDS	152311	152526		0	0
12	3UTR	152527	153004	.	.	.
13		0	0		0	0

Figure 4. The “Concat” sheet will automatically populate based on the information you enter in Figure 3.

	A	B	C	D	E	F	G	H	I
1	GFF info:								
2	scaffold_430672	John_Doe	SUTR	150642	150822	.	+	.	G1_PHGPx.right
3	scaffold_430672	John_Doe	SUTR	151023	151100	.	+	.	G1_PHGPx.right
4	scaffold_430672	John_Doe	CDS	151101	151176	2	+		2 G1_PHGPx.right
5	scaffold_430672	John_Doe	CDS	151565	151658	0	+		0 G1_PHGPx.right
6	scaffold_430672	John_Doe	CDS	151736	151845	2	+		2 G1_PHGPx.right
7	scaffold_430672	John_Doe	CDS	152019	152110	1	+		1 G1_PHGPx.right
8	scaffold_430672	John_Doe	CDS	152311	152526	0	+		0 G1_PHGPx.right
9	scaffold_430672	John_Doe	3UTR	152527	153004	.	+	.	G1_PHGPx.right
10	scaffold_430672	John_Doe		0	0	0	0	+	0 G1_PHGPx.right
11	scaffold_430672	John_Doe		0	0	0	0	+	0 G1_PHGPx.right

Figure 5. The “Genome Browser” sheet will automatically populate based on the information you enter in Figure 3.

	A	B	C	D	E	F	G	H	I
1	scaffold_430672	John_Doe	SUTR	150642	150822	.	+	.	gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
2	scaffold_430672	John_Doe	SUTR	151023	151100	.	+	.	gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
3	scaffold_430672	John_Doe	CDS	151101	151176	2	+	.	gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
4	scaffold_430672	John_Doe	CDS	151565	151658	0	+		2 gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
5	scaffold_430672	John_Doe	CDS	151736	151845	2	+		1 gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
6	scaffold_430672	John_Doe	CDS	152019	152110	1	+		2 gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
7	scaffold_430672	John_Doe	CDS	152311	152526	0	+		0 gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
8	scaffold_430672	John_Doe	3UTR	152527	153004	.	+	.	gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";
9	scaffold_430672	John_Doe		0	0	0	0	+	0 gene_id "G1_PHGPx.right"; transcript_id "GAIW01010771.1";

Figure 6. The “GFF” sheet will automatically populate based on the information you enter in Figure 3.

The “Genome Browser” sheet will allow you to add your own annotation into the Genome Browser. This can be used to check the accuracy of your annotation and to get the predicted sequence of your gene as detailed below.

Checking Accuracy of the Gene Model (Alternative to Gene Model Checker)

To see your gene model in the context of the other evidence on the browser, you will need to upload the data from your workbook. Go to the browser window and click on the “add custom tracks” button found under the viewer window (Figure 7).

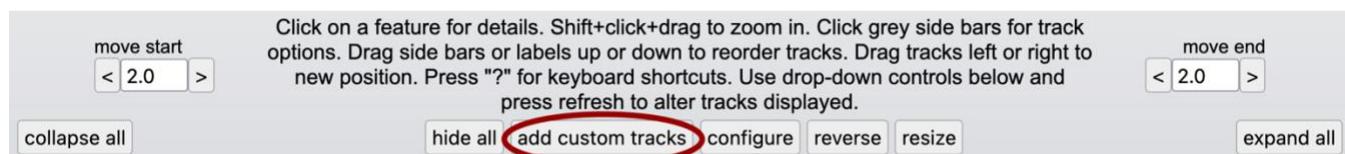


Figure 7. Click on the “add custom tracks” button to add your gene model to the browser window.

In your Annotation Workbook, click on the “Genome Browser” tab and copy the lines that contain the full GFF information (shaded) (Figure 8).

	A	B	C	D	E	F	G	H	I
1	GFF info:								
2	scaffold_430672	John_Doe	5UTR	150642	150822	.	+	.	G1_PHGPx.right
3	scaffold_430672	John_Doe	5UTR	151023	151100	.	+	.	G1_PHGPx.right
4	scaffold_430672	John_Doe	CDS	151101	151176	2	+		2 G1_PHGPx.right
5	scaffold_430672	John_Doe	CDS	151565	151658	0	+		0 G1_PHGPx.right
6	scaffold_430672	John_Doe	CDS	151736	151845	2	+		2 G1_PHGPx.right
7	scaffold_430672	John_Doe	CDS	152019	152110	1	+		1 G1_PHGPx.right
8	scaffold_430672	John_Doe	CDS	152311	152526	0	+		0 G1_PHGPx.right
9	scaffold_430672	John_Doe	3UTR	152527	153004	.	+		G1_PHGPx.right
10	scaffold_430672	John_Doe	0	0	0	0	+		0 G1_PHGPx.right
11	scaffold_430672	John_Doe	0	0	0	0	+		0 G1_PHGPx.right
12	scaffold_430672	John_Doe	0	0	0	0	+		0 G1_PHGPx.right
13	scaffold_430672	John_Doe	0	0	0	0	+		0 G1_PHGPx.right
14	scaffold_430672	John_Doe	0	0	0	0	+		0 G1_PHGPx.right

Below the table is a navigation bar with buttons for back, forward, instructions, venom gene example, transcript, concat, genome browser (which is selected and highlighted in green), GFF, and a plus sign.

Figure 8. The “Genome Browser” sheet of the Annotation Workbook.

Paste the data into the “Paste URLs or data” field of the Add Custom Tracks screen and click on “Submit” (Figure 9).

The screenshot shows the 'Add Custom Tracks' interface. At the top, there are dropdown menus for 'clade' (Wasps Genomes), 'genome' (Ganaspis sp. 1), and 'assembly' (G. species 1 (08-03-2017)). A button labeled '[ganSpe1]' is also present. Below these, a note specifies supported file formats: bigBed, bigBarChart, bigChain, bigGenePred, bigInteract, bigLolly, bigMaf, bigPsl, bigWig, BAM, barChart, VCF, BED, BED detail, bedGraph, broadPeak, CRAM, GFF, GTF, hic, interact, MAF, narrowPeak, Personal Genome SNP, PSL, or WIG. A list of instructions follows:

- You can paste just the URL to the file, without a "track" line, for bigBed, bigWig, bigGenePred, BAM and VCF.
- To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#).

Examples are [here](#). If you do not have web-accessible data storage available, please see the [Hosting](#) section of the Track Hub Help documentation.

Please note a much more efficient way to load data is to use [Track Hubs](#), which are loaded from the [Track Hubs Portal](#) found in the menu under My Data.

Paste URLs or data: (highlighted with a red oval)

scaffold_430672	John_Doe	SUTR	150642	150822	.	+	.	G1_PHGPx.right
scaffold_430672	John_Doe	SUTR	151023	151100	.	+	.	G1_PHGPx.right
scaffold_430672	John_Doe	CDS	151101	151176	2	+	2	G1_PHGPx.right
scaffold_430672	John_Doe	CDS	151565	151658	0	+	0	G1_PHGPx.right
scaffold_430672	John_Doe	CDS	151736	151845	2	+	2	G1_PHGPx.right
scaffold_430672	John_Doe	CDS	152019	152110	1	+	1	G1_PHGPx.right
scaffold_430672	John_Doe	CDS	152311	152526	0	+	0	G1_PHGPx.right
scaffold_430672	John_Doe	3UTR	152527	153004	.	+	.	G1_PHGPx.right

Or upload: [Browse...](#) No file selected. [Submit](#) (with a red arrow pointing to it) [Clear](#)

Figure 9. Pasting the coordinates of the annotated gene in the “Add Custom Tracks” window.

This will bring you to the Manage Custom Tracks screen. You can rename your gene model if you wish by clicking on the “User Track” link (Figure 10, arrow) then replacing the highlighted text following “track name=” in the “Edit configuration” field (Figure 10, shaded).

The screenshot shows the 'Manage Custom Tracks' interface. It lists a single track named 'User Track' with a description of 'User Supplied Track'. An arrow points to the 'User Track' link. Below is the 'Update Custom Track' interface for 'User Supplied Track [hub_175539_ganSpe1]'. It contains the same note about supported file formats and examples. A list of instructions follows:

- You can paste just the URL to the file, without a "track" line, for bigBed, bigWig, bigGenePred, BAM and VCF.
- To configure the display, set [track](#) and [browser](#) line attributes as described in the [User's Guide](#).

Examples are [here](#). If you do not have web-accessible data storage available, please see the [Hosting](#) section of the Track Hub Help documentation.

Please note a much more efficient way to load data is to use [Track Hubs](#), which are loaded from the [Track Hubs Portal](#) found in the menu under My Data.

Edit configuration: (shaded area containing 'track name='User Track' description='User Supplied Track')

Figure 10. Renaming your gene model.

There is no need to change the track name every time you make a track, but this does let you view two or more of your own gene models at the same time which is often useful. Clicking on “Submit” will take you back to the Manage Custom Tracks screen.

Click on the link with the project name, scaffold_430672 in our example, to view your annotation in the browser (Figure 11).

Name	Description	Type	Doc Items	Pos	delete
User Track	User Supplied Track	gff		1 scaffold_430672:	<input type="checkbox"/>

view in [Genome Browser](#) [go to first annotation](#) [return to current position](#)
[add custom tracks](#)

Figure 11. Add a custom track containing your gene model to the Genome Browser window.

The next screen will have a new track with your gene model centered around the coordinates you annotated. In this example, because we did not change the Name, the track appears as “User Supplied Track” (Figure 12, red box). You can customize the number of tracks shown on the viewing window.

At a minimum, you would want to have the following tracks turned on:

- SPALN Alignments
- At least one Gene Predictor track (in this example Augustus, N-SCAN, Geneid)
- At least one RNA-Seq track (in this example Unpaired and Paired-end RNA-Seq Coverage, StringTie)
- Ganaspis Venom Peptides

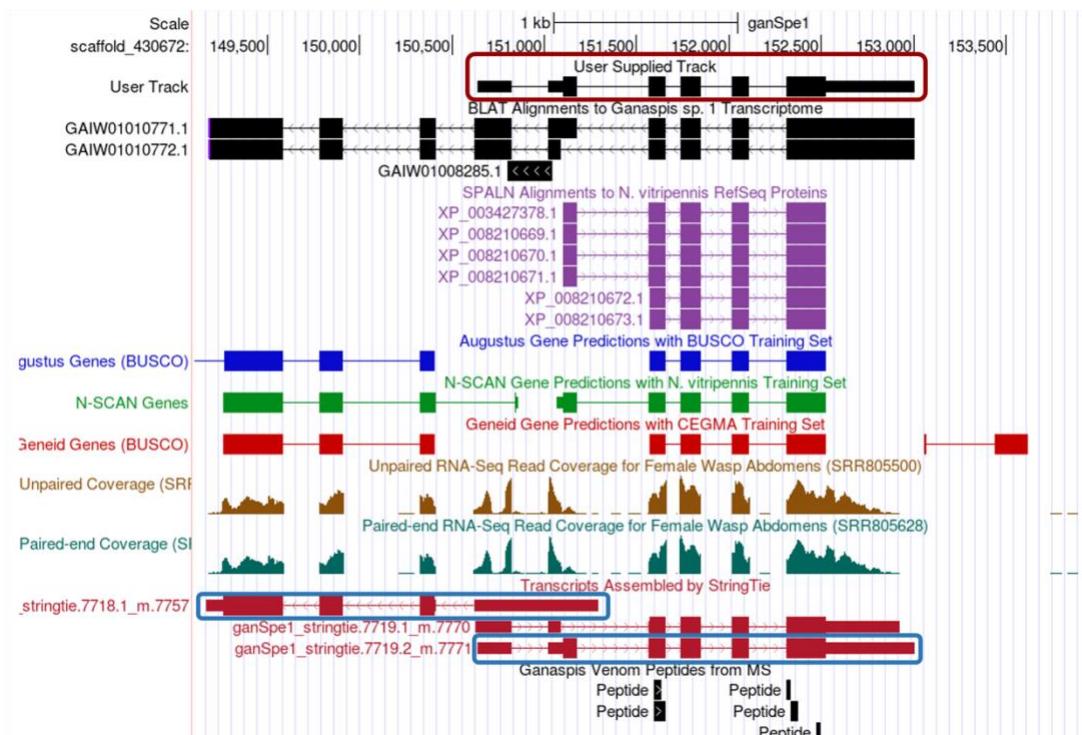


Figure 12. Your custom gene model visualized in the Genome Browser window.

Under this zoomed-out view, it can be appreciated that the *de novo* transcriptome GAIW01010771.1 prediction is probably bringing together two separate genes. This commonly happens when the genes have overlapping UTRs (one of the big issues with unstranded RNA-Seq).

Both Augustus and N-SCAN predict a gene on the left part of the region and StringTie predictions fall into two separate sections (Figure 12, blue boxes).

For this example, we have annotated the gene on the right side of the transcript ID—corresponds to the StringTie prediction on the right. However, the gene to the left side would also need to be annotated using a separate workbook.

For the submission of your project, you will need to take a screenshot of the browser with your gene model in it.

Getting the Sequence Data for the pep and fasta files

In the main navigation menu at the top of the Genome Browser, click on “Tools” and then “Table Browser” (Figure 13).

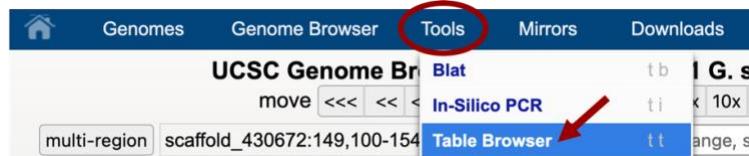


Figure 13. Navigate to the “Table Browser.”

Within Table Browser, set the “group” field to “All Tracks,” the “track” field to “User Track,” the “region” field to “position,” and “output format” to “sequence.” Then click on the “get output” button (Figure 14).

Table Browser

Use this tool to retrieve and export data from the Genome Browser annotation track database. You can limit retrieval based on data attributes and intersect or merge with data from another track, or retrieve DNA sequence covered by a track. [More...](#)

Select dataset

clade: Wasps Genomes genome: Ganaspis sp. 1 assembly: G. species 1 (08-03-2017)
 group: All Tracks track: User Track
 table: ct_UserTrack_3545 data format description

Define region of interest

region: genome position scaffold_430672:149,100-154,000 lookup define regions
 identifiers (names/accessions): paste list upload list

Optional: Subset, combine, compare with another track

filter: create
 intersection: create
 correlation: create

Retrieve and display data

output format: sequence Send output to Galaxy GREAT
 output filename: (leave blank to keep output in browser)
 file type returned: plain text gzip compressed

get output summary/statistics

Figure 14. Configure the Table Browser.

To retrieve the protein sequence, select “CDS Exons”, “One FASTA record per gene” and “All upper case”, then click on “get sequence” (Figure 15).

User Track Genomic Sequence

Sequence Retrieval Region Options:

- Promoter/Upstream by bases
- 5' UTR Exons
- CDS Exons
- 3' UTR Exons
- Introns
- Downstream by bases
- One FASTA record per gene.
- One FASTA record per region (exon, intron, etc.) with extra bases upstream (5') and extra downstream (3')
 - Split UTR and CDS parts of an exon into separate FASTA records

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

- Exons in upper case, everything else in lower case.
- CDS in upper case, UTR in lower case.
- All upper case.
- All lower case.
- Mask repeats: to lower case to N

get sequence **cancel**

Figure 15. Retrieve the protein sequence.

Copy and paste the nucleotide sequence from the next screen into [NCBI's ORFfinder](#). Then set “Minimal ORF length (nt)” to “300” and check the box next to “Ignore nested ORFs.” Click on the “Submit” button (Figure 16).

Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

```
>hub_175539_ganSpe1_ct_UserTrack_3545_G1_PHGpx.right
ATGAATATTAACACTTATTTCGAATACCCCTGATTGGTAAGCTAAA
CTTGGGAGCAAGAAGTTTGCACAGCAACAATGAGTGGAAATCAGGATT
ATAAAAGTCTTCATCAATTTCAGATTTCACAGCAAATTCTATTAAGGA
GAAGAAGTACCAATTATCAAATACAAGGGTCACGTATGTTGATTGAAA
CGTTGCTTCAAATGTGGACTAACAAAAACTAATTACAAGAATTAAACG
AGCTTCAATGATAAAATACGCAGAACATCTAAAGGTTGGAAATACTAGCTTT
CCCTGCAATCAATTATGGACAAAGAACAGGTGATGCCAGGACATTG
CAGTTTGCGGATCGACAGAAGGTTGAAATTGATCTATTGAGAAAATTG
ATGTGAATGGAGATTCACTCATCCACTGTGGAAATATTGAGAAAGGAG
CAAGGAGGAACCTTAGGAGATTTCATCAAATGGAAATTTCACAAAATTCT
TATAAATAAAGAGGAAAGTCGTGGAAAGACATGGACCTTATGTGGAAAC
CAAAGAAATTGGAACCTCATCTGGAGAAATATTCTAA
```

Choose Search Parameters

- Minimal ORF length (nt):
- Genetic code: 1. Standard
- ORF start codon to use:
 - "ATG" only
 - "ATG" and alternative initiation codons
 - Any sense codon
- Ignore nested ORFs:

Start Search / Clear

Submit **Clear**

Figure 16. To search for Open Reading Frames in your DNA sequence, paste the nucleotide sequence in the ORFfinder.

If your annotation is correct, your sequence will be translated in the correct frames and you will get a single ORF that spans the entire sequence (Figure 17). Take a screenshot of the ORF map and copy/paste the amino acid sequence into a new text file. Name this file “xxxx_protein.” If you get an ORF that is smaller or incomplete, something is annotated wrong. Use the user track on the Genome Browser to verify your exon coordinates.

The screenshot shows the Open Reading Frame Viewer interface. At the top, it says "Open Reading Frame Viewer". Below that is a "Sequence" section with the following details: "ORFs found: 1", "Genetic code: 1", "Start codon: 'ATG' only", and "Nested ORFs removed". The main view shows a blue track labeled "(U) ORFFinder_12.26.123832675" with a red arrow indicating the direction of the ORF. A red arrow points to the text "Single ORF" above the track. Below the track, a scale from 20 to 588 is shown, with "1: 1.588 (588 nt)" at the bottom. To the right, there are "Tools", "Tracks", and a question mark icon. At the bottom right, it says "Tracks shown: 2/3".

This part of the screenshot shows the protein sequence details. It includes a table with the following columns: Label, Strand, Frame, Start, Stop, and Length (nt | aa). The data in the table is:

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF1	+	1	<1	588	588 195

A red arrow points to the "Mark" button in the top bar of this panel. Another red arrow points to the text "Copy/Paste in a new text file" above the sequence box.

Figure 17. Open Reading Frame Viewer and protein sequence (box).

To get the sequence of the whole transcript, go back to the tab with the Genome Browser, currently showing the nucleotide sequence of the translated portion of the gene (Figure 16, left), and then click on the back button to navigate back to the Figure 15 screen. Now select “5' UTR Exons,” “CDS Exons,” “3' UTR Exons,” “One FASTA record per gene,” and “All upper case.” Then click on “get sequence” (Figure 18).

This screenshot shows the "User Track Genomic Sequence" retrieval options. Under "Sequence Retrieval Region Options:", the following checkboxes are checked and highlighted with a red box: "5' UTR Exons", "CDS Exons", and "3' UTR Exons". Other options like "Promoter/Upstream by 1000 bases" and "Downstream by 1000 bases" are also present but not checked. A red arrow points to the "One FASTA record per gene" option, which is also checked. Below this, a note states: "Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome." Under "Sequence Formatting Options:", the "All upper case." option is selected and highlighted with a red box. Other options include "Exons in upper case, everything else in lower case.", "CDS in upper case, UTR in lower case.", "All lower case.", and "Mask repeats: to lower case or to N". At the bottom, there are "get sequence" and "cancel" buttons.

Figure 18. Retrieve the transcript sequence.

Copy/paste the sequence (Figure 19) into a new text file. Name the file “xxxxx_mRNA.”

```
>hub_175539_ganSpe1_ct_UserTrack_3545_G1_PHGPx.right range=scaffold_430672:150642-153004 5'pad=0 3'pad=0 strand=+ repeatMasking=none
ACTTTTTCTAGATTCAAACGCTATATAAATGAAAGAATAAACCTTC
GTCCAATTTCAAATTTCAGGTTATGGCAATGTTACTGCGGTCAT
GGCAAGCGACGTAATAAAATATATAATTGTACAACTGAGTTTCTG
TTTTAACTGACTACAGTCGTTATTCGAACCTTAACTATATAGACA
CTCCAAACATTCTACGGATTCAAATACTCAAAGGATCAAAGGAGTC
GGTGAGAATATGAAATTAATTAACCTTATTCGAATACCTGATTTTGT
AAGCTCAAACTTGGGAGCAAGAAGTTTGCACACAGCACAACTGAGTGGAA
ATCAGGATTATAAAAGTGCCTCATCAATTACGATTTACAGCAAATTCT
ATTAAGGGAGAAGAAGTACCAATTACAAATACAAGGGTCACGTATGTTT
GATTGAAACCGTTGCTTAAATGTGACTAACAAAACTAATTACAAG
AATTAACGAGCTTCATGATAATACGAGAACCTAAAGGTTGCGAATA
CTAGCTTTCCCTGCAATCAATTAAATGGACAAGAACCCAGGTGATGCCA
GGACATTGCAGTTGCAGTCGACAGAAGGTGAAATTGATCTATTG
AGAAAATTGATGTGAATGGAGATTCTAGCTCATCCACTGTGAAATTTG
AAGAAGGAGCAAGGAGGAACCTTACGGAGATTCTCATCAAATGGAATTTCAC
AAAATTCAATTAAATAAAAGAGGGAAAAGTCGTGAAAGACATGGACCTT
ATGTGAAACCAAAGAAATTGAACTCATCTGGAGAAATTCTAAATT
ACAAATGATCAAGATCTATTTGTATTGCACAAAGTCACAGTGA
TTATGAAATTGAGGTGATAAAATTAAATTACATGAAATATGCAAATTCTTGTCAA
CAAATTAAATTACTTTTACATGAAATATGCAAATTCTTGTCAA
TAGAGTTAAGAATCGAAGTAGCGAAGGTTCTTATTGCAACGTT
AATCGATAATCACTTACTTTTGTATTCTAAATTACATTATA
CAACTCTGAAACTTGTGAATAAAACTTCATGAATTAAATTGAAAC
GAACACCCACCTTTACAATTTCAGATATAACGTGTATTAAATTCTGA
TAAATCAATTCTTATTCATTAAATTGAGTATTCTGTAT
ATTTAAATGAAACTTCTAAATATTCAAAATTGTATTAAATAAA
GAAAATAAATACGTTTATCAAAAT
```

Figure 19. Save the transcript sequence as a txt file.