Module 1: Basic Information

Objectives

The objectives of this module are:

1. To navigate to the gene information page for a gene assigned to be annotated.
2. To record information about the DNA sequence of the gene to be annotated and about the amino acid sequence of the protein encoded by the gene that will be used in subsequent modules. Some of this information may be modified at a later point based on the results of your annotation.

Materials

To perform this activity you will need:

- Access to the internet on a computer equipped with the most recent version of Firefox (preferred), Chrome or Safari (Firefox should be used to have the best functionality).
- To have completed the signup for GENI-ACT described in the Signing Up for GENI-ACT section of the manual.
- To have an assignment visible on your assignments page.

Procedures

1. Log In to GENI-ACT
   - Open the GENI-ACT website http://GENI-ACT.org using one of the browsers mentioned in the Materials section above.
   - Click on the Login button and type in your email address and password.
   - Press Login.
   - Click the hyperlink for the course name to take you to your GENI-ACT homepage (Figure 2.1) and then click on the hyperlink for your assignment to access the genes that you are working on. This is your GENI-ACT working page (Figure 1.2).
2. Opening the Gene Notebook
   - In Figure 2.2, you will see how a page with a single assignment appears.
   - Click on the link under the Assignment heading (called Test in the example) to open your assignment (Figure 1.2).
   - You will next open your online lab notebook.

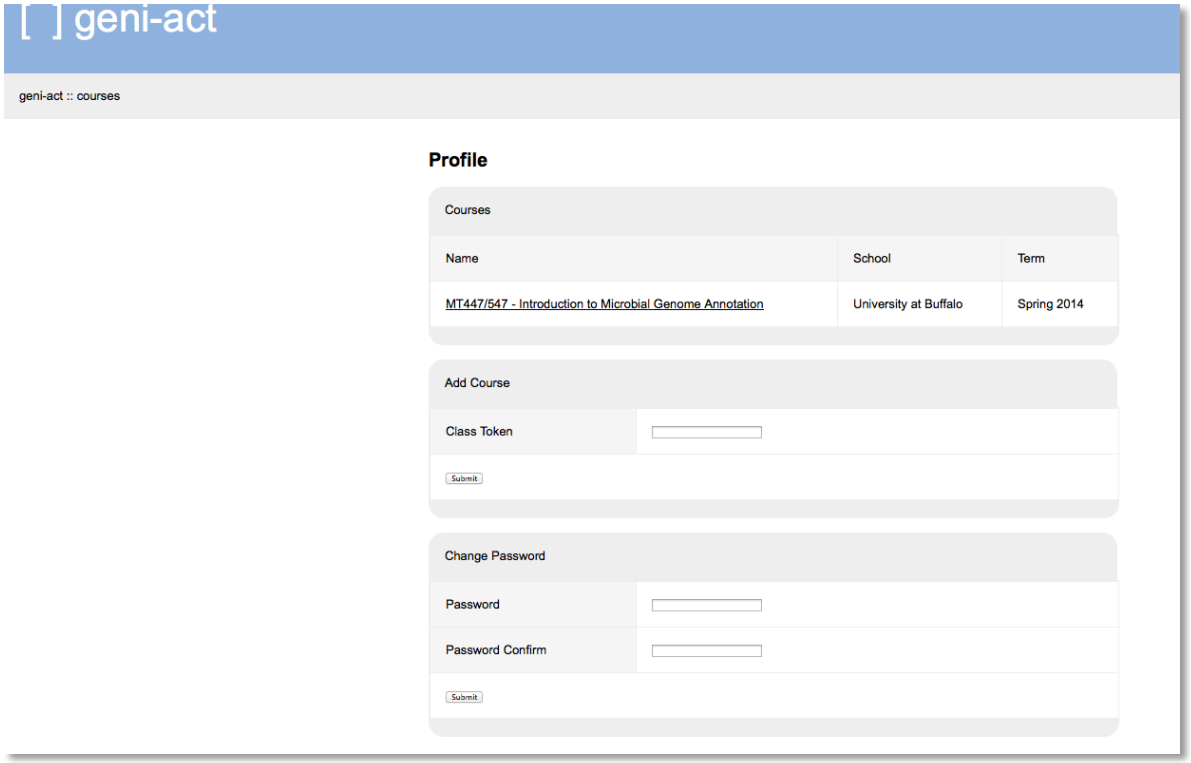


Figure 1.1. The student homepage of GENI-ACT

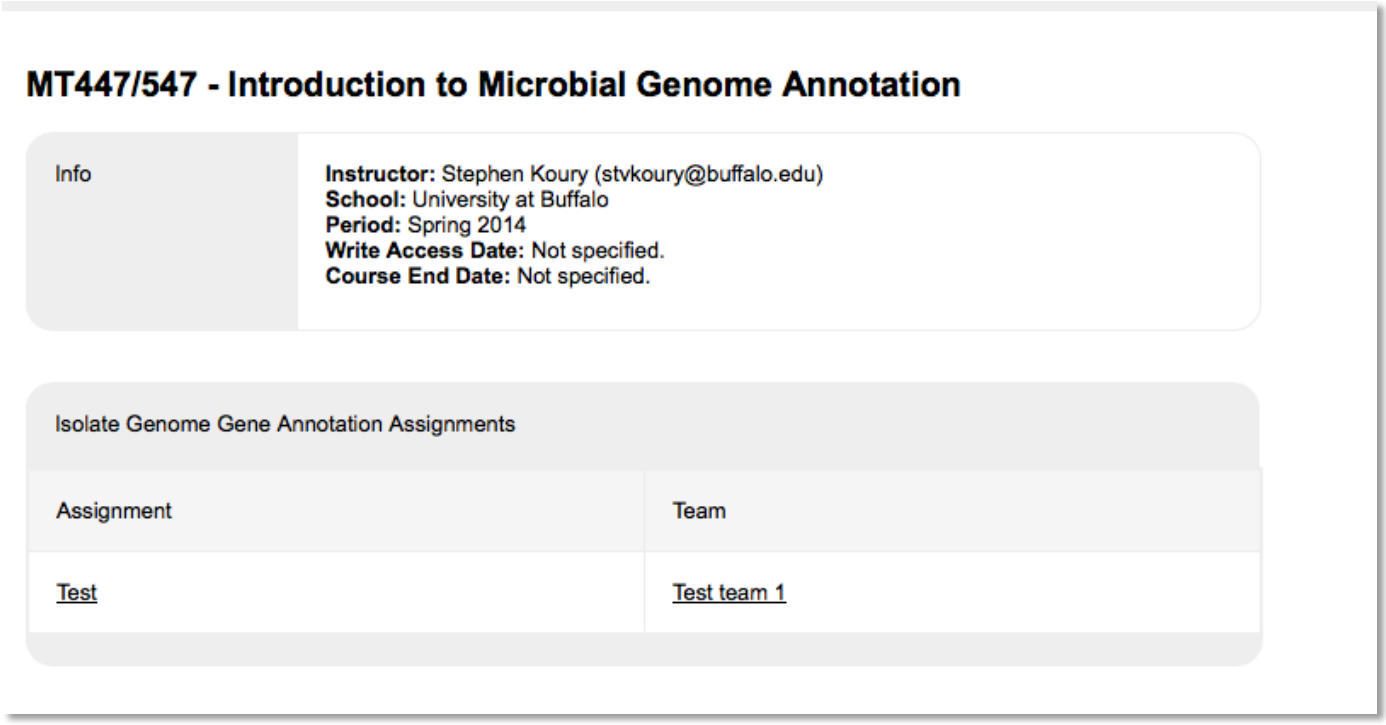


Figure 1.2. A Gene Annotation Assignment

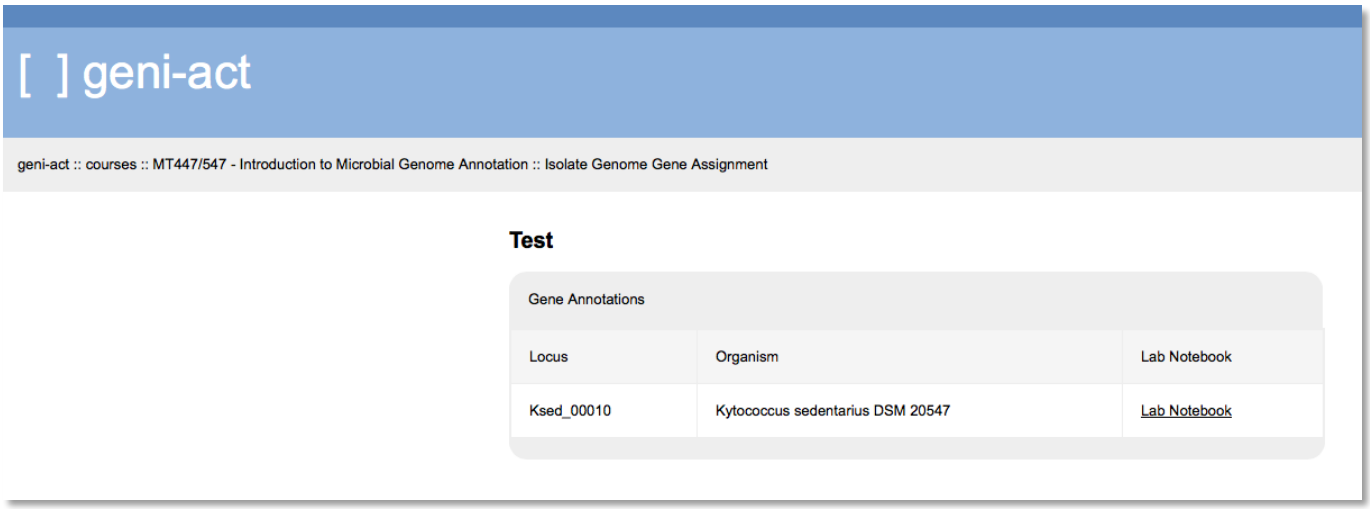


Figure 1.3. A Gene Annotation Assignment Notebook Page

3. Appearance of the Notebook Page (Figure 2.4)
   - Click on the lab notebook link to open your GENI-ACT notebook page
   - At the top of the page you will see the genome of the organism from which the gene is taken (*Kytococcus sedentarius* 541 DSM 20547 in the example shown in figure 1.4) and the Genbank locus tag of your gene. The locus tag is unique for each gene in the Genbank database and it is an active hyperlink that will be used many times during your annotation assignment.
   - A box labeled Instructions on the notebook page give a quick link to online GENI-ACT instructions for each module. **They are similar, but not identical or as detailed as the one you will have in this manual.**
   - The bottom of the page has all of the 9 modules of GENI-ACT collapsed for easy navigation between modules.

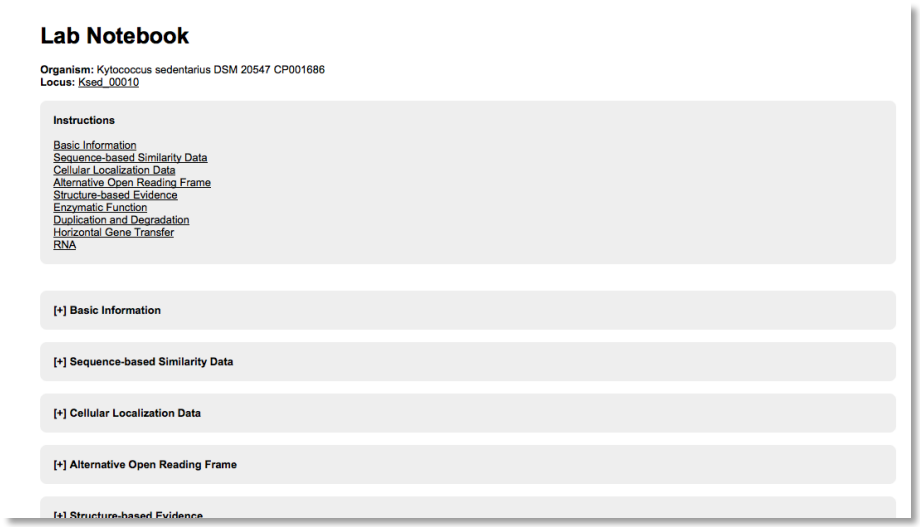


Figure 1.4. The GENI-ACT notebook

4. DNA Coordinates

- Log onto your GENI-ACT working page.
- Open the gene notebook (in a new tab or window) corresponding to the gene that you are working on as described in Opening Gene Notebook.
- Hold the command key down while clicking on the locus tag (Ksed_00010 in the example in figure 1.4) to open the gene details page in a new tab (Figure 1.5 below). Tabbed browsing is the best way to keep the notebook, the gene details page and any other links you work on open and organized.

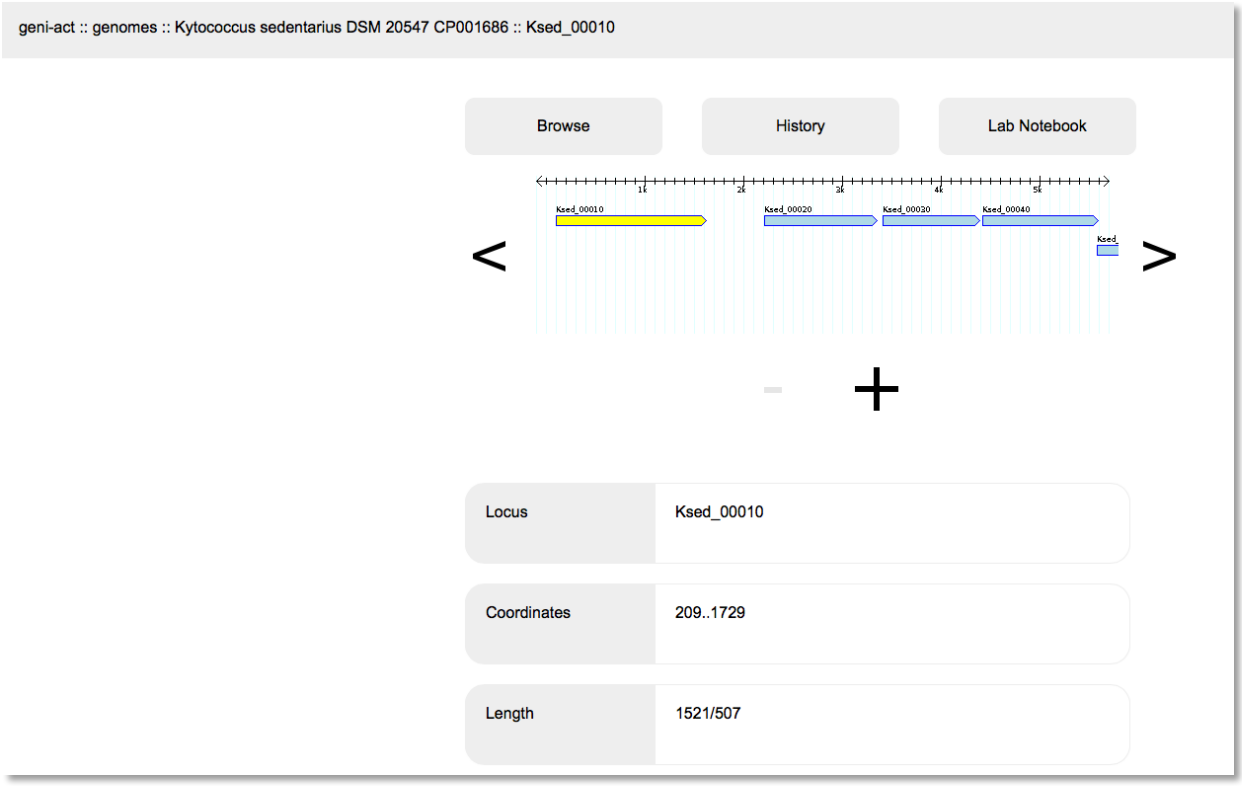


Figure 1.5. A portion of the gene details page. Your gene will be indicated in yellow along with other genes in the same chromosomal neighborhood as your gene shown in blue. The arrowhead pointing to the right for all genes in this example indicates that genes are on the top strand of DNA and oriented with their 5' end to the left and their 3' end to the right. **If your gene or other genes in the neighborhood have the arrowhead oriented to the left, then the genes of question are on the bottom strand of the DNA molecule and oriented in the opposite direction.**

- Find the subheading labeled Coordinates. To the right are the DNA coordinates for your gene. This refers to the first and last nucleotide positions in the genome of *K. sedentarius.* Coordinates could range between 1 and 2785024. In the example shown in figure 1.5 above, the coordinates are 209..1729.
- Copy and paste the DNA coordinates sequence into your open lab notebook (Figure 1.6). You can edit the notebook by clicking on the notepad icon indicated by the arrow in Figure 1.6. Be sure to note if the

gene is located on either the top (forward) or bottom (reverse) strand of the double stranded DNA genome (Figure 1.6).

- Click Save (in the lower left corner of the notebook editor as seen in Figure 1.7) to save changes to your notebook. Be sure to save your work frequently to avoid a loss of data. There is also a scroll bar at the right of the notebook that allows you to move to different data recording sections of each module.

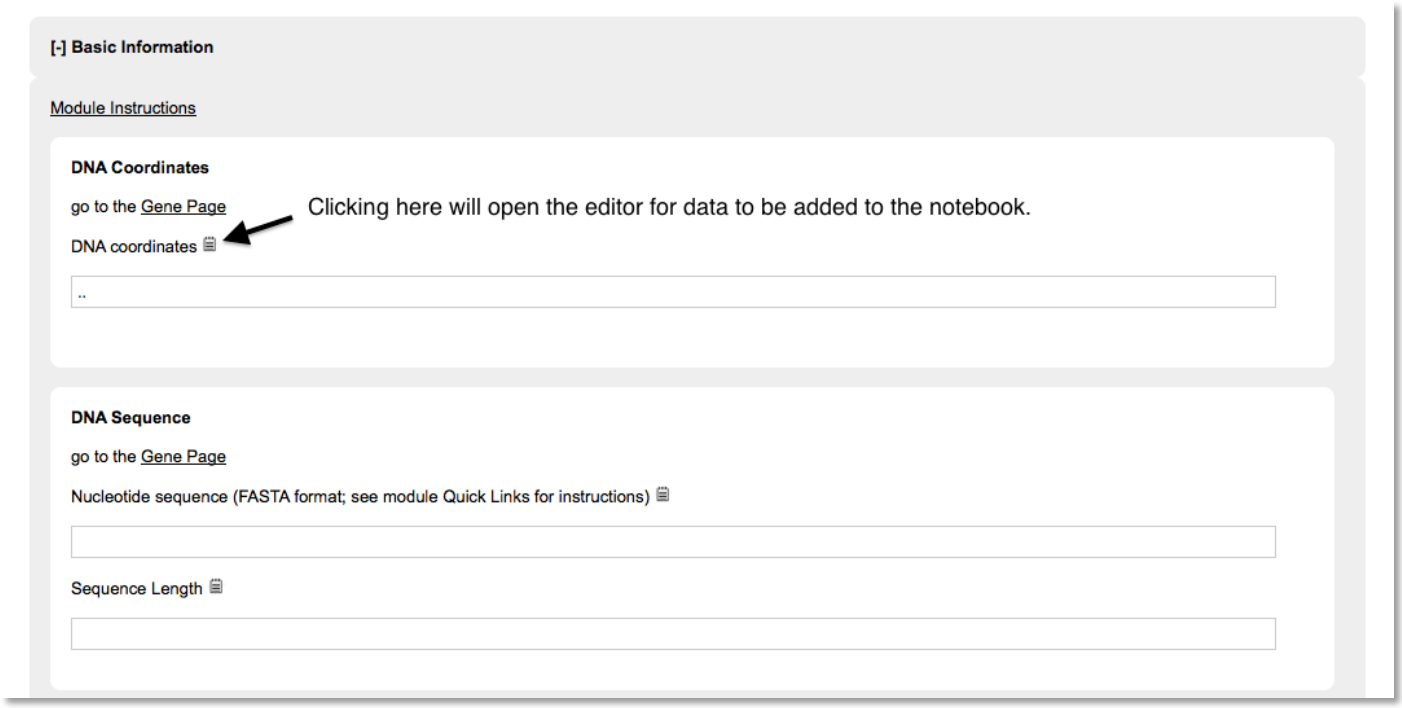
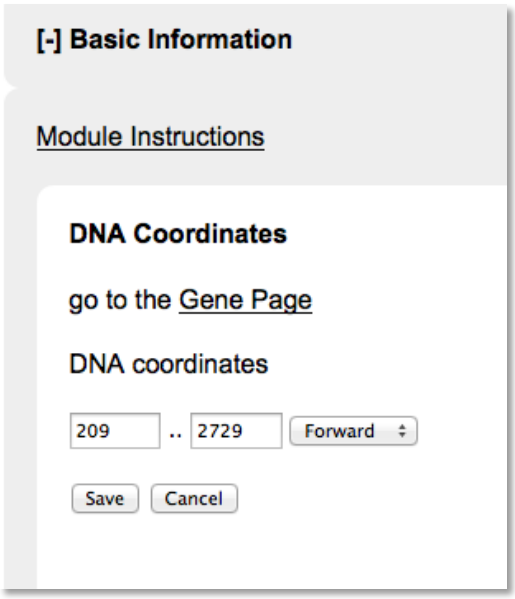


Figure 1.6. A portion of the Module 1 Notebook Page. The arrow indicates the icon to click to add data to the notebook.

Figure 1.7. The DNA coordinates section that opens after clicking the icon shown by the arrow in Figure 2.6.



5. DNA Sequence

- On your gene information scroll down until you see the nucleotide sequence section as show in figure 1.8. The dark letters represent the actual sequence of the gene you are annotating and the lighter letters represent nucleotides upstream or downstream from your gene.
- Select and copy the DNA sequence in the dark letters (beginning with CTG and ending with TGA in the example DNA sequence in figure 1.8)

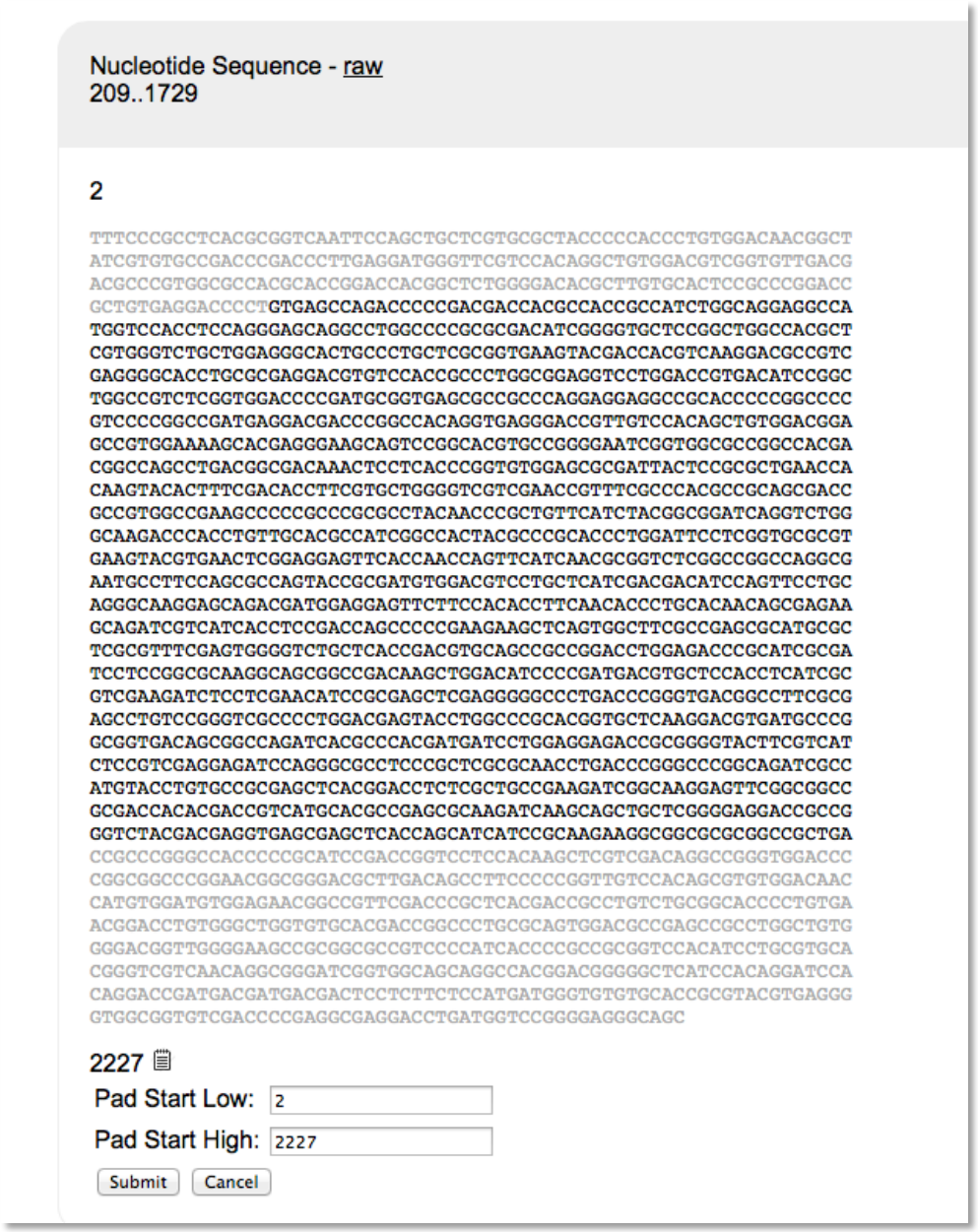


Figure 1.8. The nucleotide sequence section from the gene information page.

- Use the scroll bar to navigate down to the box entitled DNA sequence in your notebook. Click on the editor link as you did above to open the DNA sequence editor box and paste the raw sequence into the box (Figure 1.9) and click on the save button.

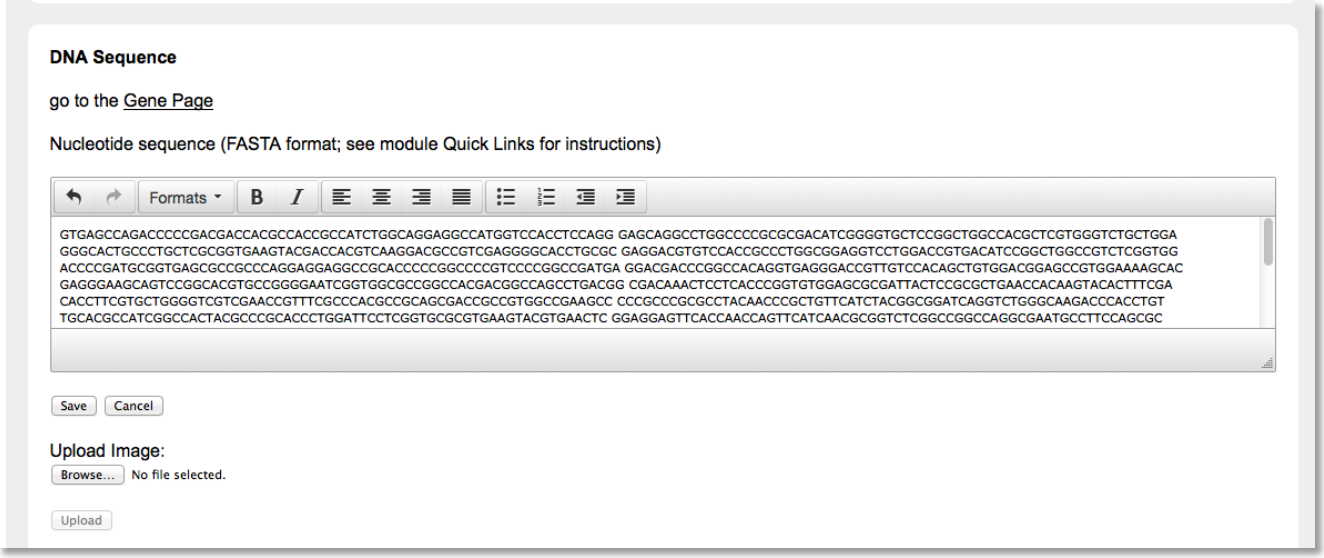


Figure 1.9. The DNA sequence editor box in the notebook. The sequence from figure 2.9 is shown pasted in the box

- **We will use a FASTA header to allow you to keep track of your sequences as you plug them into the modules that follow:**
  - FASTA format uses a first line to give information about the sequence that follows. The line must begin with a ">". Any information that follows on the line will not be used when the sequence information is submitted to a database for a search.
  - FASTA format allows you to keep track of the sequences you are working with and will be used routinely during your annotations.
  - Click on editor link again to open the nucleotide sequence editor box. Insert the cursor before the first letter of the sequence (be sure the indicator on the scroll bar is all the way at the top so that you know you are at the start of your sequence) and then hit return to put a blank space at the top of your sequence.
  - In the example the locus tag for the gene is Ksed_00010, so the FASTA header that would be used for this gene is >Ksed_00010 nucleotide sequence, as shown in Figure 1.10.

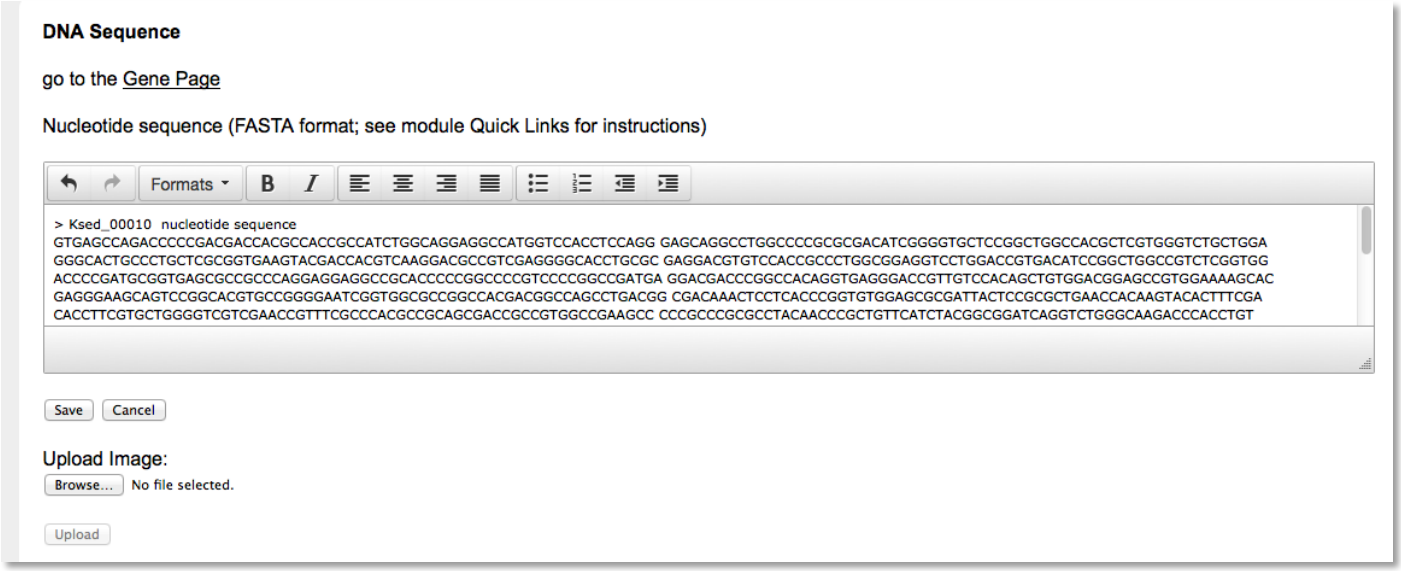


Figure 1.10. The nucleotide sequence for Ksed_00010 in FASTA format. The FASTA header provides information about the source of the sequence.

- Click Save after adding the FASTA header specific for your gene in the lower left hand corner of your lab notebook to save changes. **Be sure to save your work periodically while annotating to make sure you do not lose any data in your notebook.**
- Go back to the gene information page and locate the box labeled Length (see figure 2.5 above). The first number in the box is the length of the nucleotide sequence of your gene then a / and then a second number which corresponds to the amino acid length the protein encoded by your gene. In the example from figure 2.5 the length box reads: 1521/507. Write the length of the DNA sequence of your gene in the sequence length box of your notebook (first click on the editor icon, fill in the length and hit save).
- When completed the DNA sequence information of your notebook should look something like that shown in figure 1.11

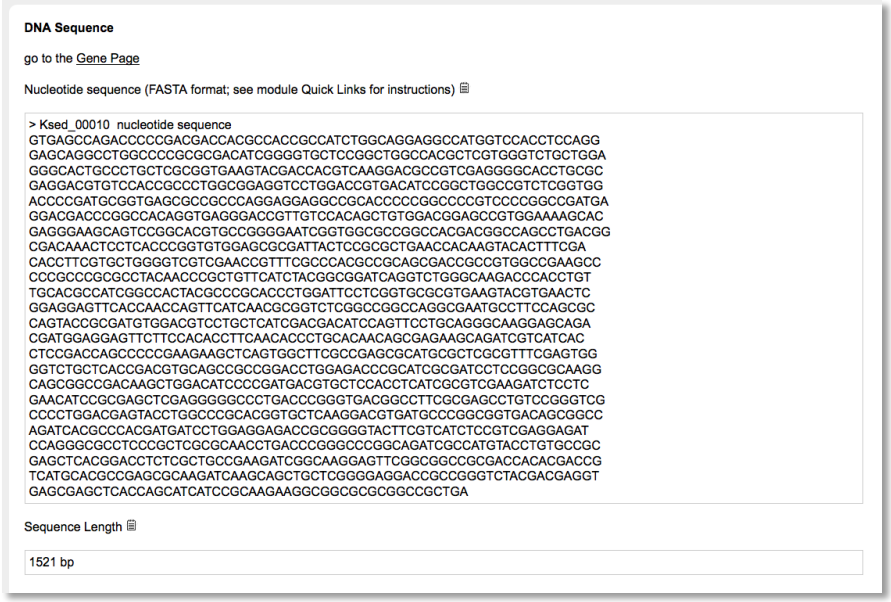


Figure 1.11. The complete example entry for DNA sequence information.

6. Amino Acid Sequence

- Return to the gene information page and scroll down until you see the Amino Acid Sequence section (Figure 1.12). As was the case for the DNA sequence information, the amino acids encoded by your gene are indicated in bold font, while translated amino acid sequence upstream and downstream from that proposed for your gene are indicated in a lighter font.

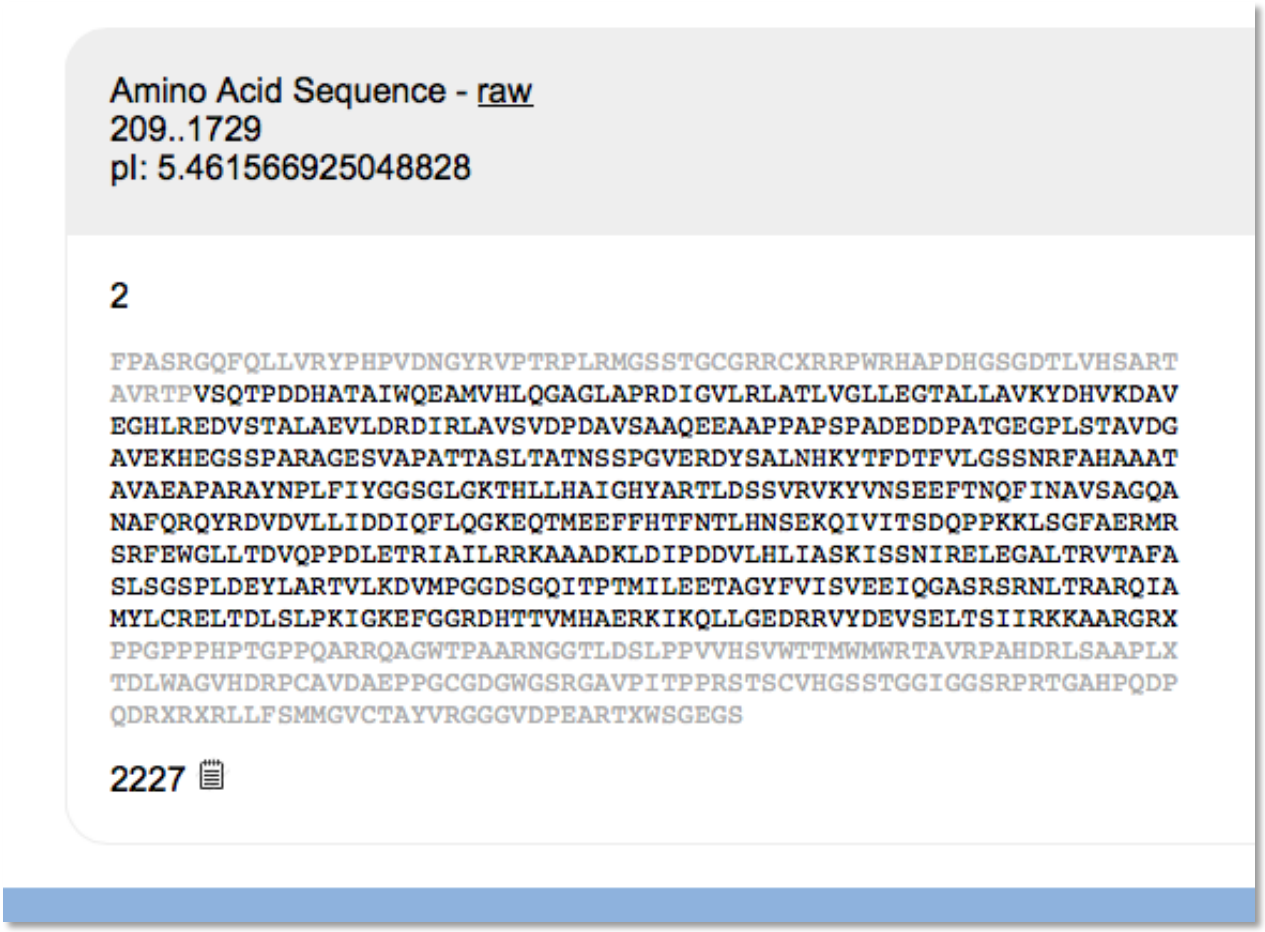


Figure 1.12. The amino acid sequence information section for Ksed_00010. See the text for a full explanation

- Copy the bold font sequence and paste it into the Protein Sequence box in your notebook (remember to click on the editor icon first and to save after pasting). Add a FASTA header to the amino acid sequence in the same way you did for the nucleotide sequence. In the example in figure 1.13 the FASTA header would read: >Ksed_00010 Amino Acid Sequence.
- Go back to the gene information page and locate the box labeled Length (see figure 1.5 above). The first number in the box is the length of the nucleotide sequence of your gene then a / and then a second number which corresponds to the amino acid length the protein encoded by your gene. In the example from Figure 1.5 the length box reads: 1521/507.
- Write the length of the amino acid sequence of your protein in the sequence length box of your Protein Sequence notebook (first click on the editor icon, fill in the length and hit save)

- When completed the Protein sequence section of your notebook should look something like that shown in figure 1.13.

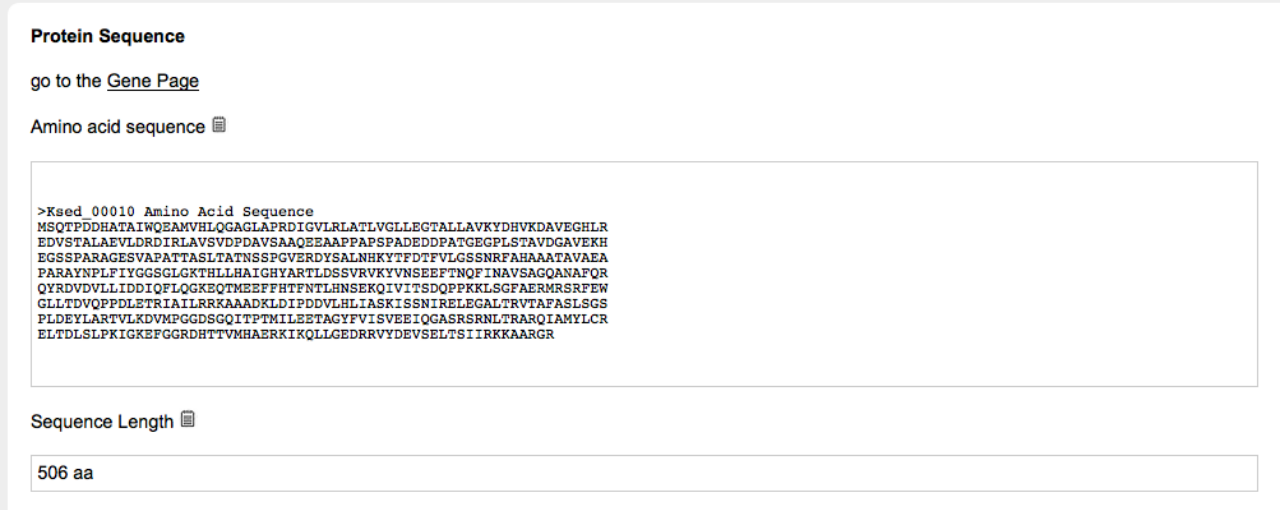


Figure 2.13. The amino acid sequence information section for Ksed_00010

- Click Save to save your work.
- Repeat the process for the other genes in your assignment
- You have completed the required information for the Basic Information Module