

REACHING
OTHERS



Assignment 1: Basic Information Module

Navigate to geni-act.org



Select current classes after logging in

The screenshot shows the GENI-ACT website interface. At the top, there is a navigation bar with the GENI-ACT logo and menu items for Communities, Projects, and Classes. The 'Classes' menu is open, showing options: Join a Class, Current Classes, New Class, and Past Classes. The 'Join a Class' button is highlighted in green. Below the navigation, there is a 'Join a Class' section with a 'Class Pin' input field and a 'Join' button. The main content area features a 'Welcome Developer' heading and a list of roles for developers: Project lead and Partner. The 'Recent Activity' sidebar on the right shows two entries: 'Course Activated' for Mount Morris Genome Annotation and 'School Added' for Pacific University and Xavier University.

GENI-ACT Communities Projects Classes

Join a Class [Join](#)

Class Pin

Current Classes

New Class

Past Classes

Welcome Developer

As a developer you have two key roles:

- Project lead:** The lead initiates a project, builds courses and can invite partners to help build the project. There is only one lead per project, though you can transfer the lead role to any of your partners should you desire to do so. As project lead you will approve all student data submitted by instructors if you feel it is of sufficient quality to be included in the project or community dataset.
- Partner:** Partners can join any project and create and edit course materials that will be added to the project. There is no limit on the number of partners in a project. Please contact the project lead to help build a project.

Additional roles:

- Community curator:** Developers may also create and lead communities as a curator. Community curators have a similar role to project leads and will provide the final approval for student data to be included in the community data set. Please contact a GENI-

Recent Activity

- + Course Activated**
[Mount Morris Genome Annotation](#)
S. Koury - 3 days
- School Added**
Pacific University
13 days
- School Added**
Xavier University
14 days



Classes

Classes I'm Enrolled In

Title	Status	Instructor
Bioinformatics 2018 - Julie McLaughlin - CSCS	Active	Julie McLaughlin
2018 2019 Gene annotation club - Gene Club	Active	Pamela Patterson
MT447/547 Spring 2019 - Introduction to Microbial Genome Annotation	Active	Stephen Koury
Wolbachia 2018-2019 - Liffner - Sherman High School	Active	Stephen Koury
2018-2019 School Year - Mioducki - Research Laboratory for Bioinformatics & Life Sciences	Active	Stephen Koury
2018-2019 School Year - East Aurora High School - Mr. Larry Grisanti	Active	Stephen Koury
2018-2019 School Year - Rich-Frontier	Active	Stephen Koury

When you have logged into GENI-ACT, and selected “Current Classes”, you will see a list of classes in which you are currently enrolled (your list will look different than the example above). Click on the course in which you want to work, and you will see something similar to what is shown in the next slide.



Introduction to Microbial Genome Annotation

MT447/547 Spring 2019

This is a course for undergraduate and graduate students at the University at Buffalo that is part of the project to reannotate the genome of the bacterium *Kytococcus sedentarius*. The following are demonstration genes that will be used to illustrate the application of different modules in the course. Old locus tags are from prior to the automated pipeline reannotation of the *Kytococcus* genome in GenBank and are needed to access gene information at the site IMG/M that will be used later in the course.

Old Locus Tag	New Locus tag
Ksed_00010	KSED_RS00005
Ksed_00020	KSED_RS00010
Ksed_03810	KSED_RS01840
Ksed_04030	KSED_RS01950
Ksed_04070	KSED_RS01970
Ksed_04340	KSED_RS02105
Ksed_04870	KSED_RS02350
Ksed_05700	KSED_RS02745
Ksed_18780	KSED_RS09210
Ksed_25760	KSED_RS12585
AWG71_20800	same
RCFBP_mp10153	same

No class specific overview provided...

Select Team

- [Hrithik Bagga](#)
- [Dhruv Prasad](#)
- [Yvette Rodriguez](#)
- [Emily Brooks](#)
- [Jillian O'Shaughnessy](#)
- [Syed Adnan Uddin](#)
- [Erica Wicher](#)
- [Matthew Spink](#)
- [Instructor Demonstration Genes - Rama](#)
- [SEPA Teacher Overview](#)
- [Steve Genes](#)

Your team
selection will
look different
than shown here.

REACHING OTHERS



Select Team

Selected Team: Steve Genes [Switch to Hrithik Bagga](#) [Switch to Dhruv Prasad](#) [Switch to Yvette Rodriguez](#) [Switch to Emily Brooks](#)
[Switch to Jillian O'Shaughnessy](#) [Switch to Syed Adnan Uddin](#) [Switch to Erica Wicher](#) [Switch to Matthew Spink](#)
[Switch to Instructor Demonstration Genes - Rama](#) [Switch to SEPA Teacher Overview](#)

Select Team Assignment

[Make Selection](#)

The team “Steve Genes” has been selected by clicking on it from the list. After selecting the team and clicking “make selection”, you will see the list of genes in the assignment as shown in the next slide.



Select the gene on which you wish to work, and then click on “Back to Class”

[MT447/547 Spring 2019](#) / [Manage](#) / [Team](#) / Assignments

Steve Genes - Assignments

[Back to Class](#)

Team Assignments

My Selection

[KSED_RS00010](#)

Select

[KSED_RS00005](#)

Currently Selected

[KSED_RS01840](#)

Select

[KSED_RS01950](#)

Select

[KSED_RS02350](#)

Select

[KSED_RS01970](#)

Select

[KSED_RS02745](#)

Select

[KSED_RS09210](#)

Select

[KSED_RS12585](#)

Select

REACHING OTHERS



Select Team Assignment

Selected Locus : [KSED_RS00005](#) [Change Selection](#)

Modules

#	Name
1	Pre Course Survey
2	Basic Information
3	Sequence Based Similarity
4	Structure Based Similarity
5	Cellular Localization
6	Alternative Open Reading Frame
7	Enzymatic Function
8	Duplication and Degradation
9	Horizontal Gene Transfer
10	RNA Family
11	Final Annotation
12	Post Course Survey

Screenshot

You will see the list of modules. The locus tag of the gene for which you are collecting data is shown at the top, along with the option to switch to a different gene if desired.

The Basic Information notebooks should be completed for the gene assigned specifically to you.



After selecting the Basic Information link, you will be taken to a Google Drive where all notebook templates are found. Select the Basic Information Module.docx

My Drive > SEPA GENI-ACT Word Notebook Files 



Name ↑

Owner

Last modified

File size

 1. Basic Information Module.docx 	me	Jan 6, 2019 me	14 KB
 2. Sequence Similarity Module.docx 	me	Jan 6, 2019 me	16 KB
 3. Structure Based Evidence.docx 	me	Jan 6, 2019 me	17 KB
 4. Cellular Localization Data Module.docx 	me	Jan 6, 2019 me	16 KB
 5. Alternative Open Reading Frame Module.docx 	me	Jan 6, 2019 me	16 KB
 6. Enzymatic Function Module.docx 	me	Jan 6, 2019 me	15 KB
 7. Duplication and Degradation Module.docx 	me	Jan 6, 2019 me	13 KB
 8. Horizontal Gene Transfer.docx 	me	Jan 6, 2019 me	15 KB

REACHING OTHERS



Be sure to download the template and then save in the format “locus tag_basic information” to your local drive.

The screenshot shows a Google Drive interface with a document titled "1. Basic Information Module.docx" open. The document content is as follows:

- Basic Information Module**
- [Module Instructions](#)
- Locus Tag of Your Assigned Gene**
Enter the locus tag of the gene assigned to you for annotation:
- DNA Coordinates**
DNA coordinates
- DNA Sequence**
go to the IMG Gene Details page for the proposed gene
Nucleotide sequence (FASTA format)

Sequence Length
- Protein Sequence**
go to the IMG Gene Details page for the proposed gene
Amino acid sequence (FASTA format)

The document is displayed in a dark-themed Google Drive interface. The left sidebar shows navigation options like "My Drive", "Computers", "Shared with me", "Recent", "Starred", "Trash", "Backups", and "Storage". The right sidebar shows a list of files with their sizes. The bottom of the screen shows a macOS dock with various application icons.

REACHING OTHERS



Select Team Assignment

Selected Locus : [KSED_RS00005](#) [Change Selection](#)

Modules

#	Name
1	Pre Course Survey
2	Basic Information
3	Sequence Based Similarity
4	Structure Based Similarity
5	Cellular Localization
6	Alternative Open Reading Frame
7	Enzymatic Function
8	Duplication and Degradation
9	Horizontal Gene Transfer
10	RNA Family
11	Final Annotation
12	Post Course Survey

Screenshot

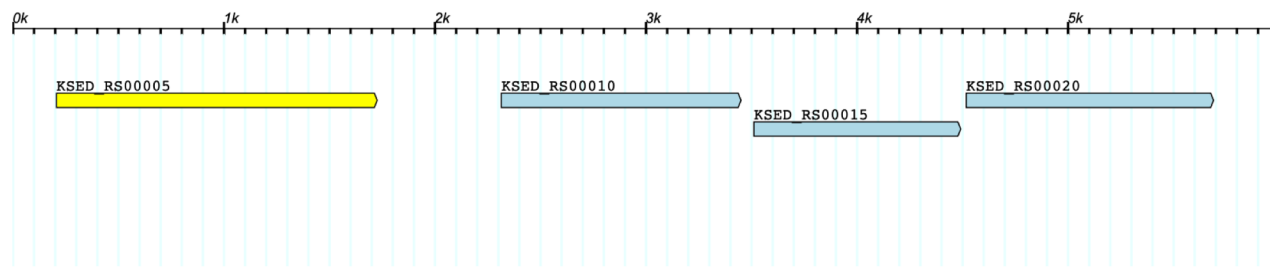
Right click on the locus tag of the gene to open the gene information page for that gene.

REACHING OTHERS



[Microbial Genomics](#) / [Kytococcus sedentarius DSM 20547](#) / [NC_013169](#) / [KSED_RS00005](#) ★

◀ Gene Browser (reset) ▶▶



Genome

[Kytococcus sedentarius DSM 20547](#)

Replicon

[NC_013169](#)

Locus

KSED_RS00005

Coordinates

209..1729

This is the top of the gene information page where you will see a graphical representation of the gene under investigation in yellow. The arrowhead pointing right indicates the gene is on the top strand of DNA. If it points to the left the gene is on the bottom strand.

The start and ending nucleotide coordinates are shown as well.

REACHING OTHERS



Old Locus Tag

Ksed_00010

Products

chromosomal replication initiation protein DnaA

Length

1521 / 506

Literature Links

Functional Support

Amino Acid Sequence

```
VSQTPDDHATAIWQEAMVHLQGAGLAPRDIGVLRRLATLVGLEGTALLAVKYDHVKDAVEGHLREDVSTALAEVLRDRDIRLAVSVDPPDAVSAAQEEAAPPAPSPAEDDDP  
ATGEGPLSTAVDGAWEKHEGSSPARAGESVAPATTASLTATNSSPGVERDYSALNHKYTFDTFVLGSSNRFAAAAATAVAEAPARAYNPLFIYGGSGLGKTHLLHAIGHY  
ARTLDSSVRVKYVNSEEFNQFINAVSAGQANAFQRYRDVLLIDDIQFLQKQETMEEFFHTFNTLHNSEKQIVITSDQPPKLSGFAERMRSRFEWGLLTDVQPPD  
LETRIAILRRKAAADKLDIPDDVLHLIASKISSNIRELEGALTRVTAFASLSGSPLEYLARTVLKDVMPGGDSGQITPTMILEETAGYFVISVEEIQGASRSRNLTRAR  
QIAMYLCRELTDLSLPKIGKEFGGRDHTTVMHAERKIKQLLGEDRRVYDEVSELTSIIRKKAAGR*
```

This is the middle section of the gene information page where you will see the “old” locus tag for the gene (if applicable).

The products sections gives the computer pipeline prediction of the gene product.

The length section gives the nucleotide length in base pairs / amino acid length in amino acids.

The amino acid sequence gives the single letter code amino acid sequence predicted for the protein, with an * to show the stop codon.

REACHING OTHERS



Nucleotide Sequence

```
gtgagccagacccccgacgaccacgccaccgccatctggcaggaggccatgggccacctccagggagcaggcctggccccgcgcgacatcgggggtgc  
tccggctggccacgctcgtgggtctgctggagggcactgccctgctcgcggtgaagtacgaccacgtcaaggacgccgtcgaggggacacctgcgcga  
ggacgtgtccaccgccctggcggaggtcctggaccgtgacatccggctggccgtctcgggtggaccccgatgcgggtgagcgcgcgccagggaggagcc  
gcacccccggccccgtccccggccgatgaggacgaccggccacaggtgagggaccgttgtccacagctgtggacggagccgtggaaaagcacgagg  
gaagcagtccggcacgtgccggggaatcgggtggcgccggccacgacggccagcctgacggcgacaaaactcctcaccgggtgtggagcgcgat tactc  
cgcgctgaaccacaagtacactttcgacacctcgtgctggggtcgtcgaccgtttcgcccacgccgacgcgaccgctggccgaagccccgcc  
cgcgctacaacccgctgttcacatcagggcgatcaggtctgggcaagaccacactgttgacgccaatcgccactacgccgcacctggattcct  
cgggtgcgcgtgaagtacgtgaactcggaggagttaccaaccagttcatcaacgcggtctcggccggccaggcgaatgccttcagcgcagttaccg  
cgatgtggacgtcctgctcatcgacgacatccagttcctgcagggaaggagcagacgatggaggagtcttccacacctcaacacctgcacaac  
agcgagaagcagatcgtcatcacctccgaccagccccgaagaagctcagtggtctcgccgagcgcgatgcgctcgcgtttcgagtggtgctgctca  
ccgacgtgcagcccggaacctggagaccgcatcgcgatcctccggcgaaagcagcggccgacaagctggacatccccgatgacgtgctccacct  
catcgcgtgaaagatctcctogaacatccgcgagctcagggggccctgaccgggtgacggccttcgcgagcctgtccgggtcgccccggacgag  
tacctggccccgacgggtgctcaaggacgtgatgccccggcgtgacagcggccagatcacgccacgatgatcctggaggagaccgccccggtacttcg
```

Note

Derived by automated computational analysis using gene prediction method: Protein Homology.

Further down the page
you will see the
nucleotide sequence

REACHING OTHERS



Member Contributions

[S. Koury - University at Buffalo](#) ←
[B. Lewis](#)
[J. McLaughlin - Campbell-Savona Jr/Sr High School](#)
[P. Hentschke](#)
[K. Flint](#)
[K. Carretto](#)
[K. Mioducki](#)
[D. Hildreth](#)
[R. Schwartz](#)
[L. Knappenberger - University at Buffalo](#)
[J. Hahn](#)
[T. Simchick](#)
[L. Grisanti - University at Buffalo](#)
[V. Rich](#)
[S. Kolbert](#)
[S. VanOstrand - University at Buffalo](#)
[L. Liffner - University at Buffalo](#)
[M. Chiulli - LeRoy Jr-Sr High School](#)
[J. Clancy - University at Buffalo](#)
[S. Proctor-Bates](#)
[B. Vinton](#)

Microbial Genomics / Kytococcus sedentarius DSM 20547 / NC_013169 / KSED_RS00005 / S. Koury
Classes with contributions by Stephen Koury
SEPA Training 2018
MT447/547 Spring 2019 ←

Dataset History

At the bottom of the page you will see contributions made by others who may have worked on the gene. KSED_RS00005 is used often as a demonstration gene and so there a large number of notebooks that have been uploaded. You will likely not see any notebooks for the genes on which you are working.

A good example of a completed Basic Information Word notebook can be seen by selecting KS. Koury – University at Buffalo (arrow) from the list and selecting the MT447/547 Spring 2019 option (arrow in insert image).

REACHING
OTHERS



Completing the Basic Information Notebook

REACHING OTHERS



An example completed notebook section for the DNA coordinates section. In this case the gene neighborhood image was “snipped” and added along with a text explanation of the meaning of the coordinates.

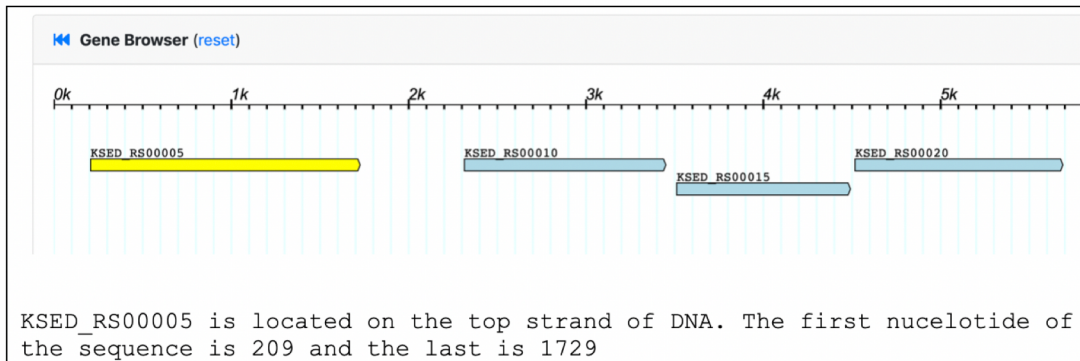
[Module Instructions](#)

Locus Tag of Your Assigned Gene

Enter the locus tag of the gene assigned to you for annotation:

DNA Coordinates

DNA coordinates



Had the gene been on the bottom strand, the coordinates would have been presented in the same way, but the word complement should be inserted in front of the coordinates (indicating the 5' end of the gene is to the right on the bottom strand of the DNA)

REACHING OTHERS



DNA Sequence

go to the IMG Gene Details page for the proposed gene

Nucleotide sequence (FASTA format)

```
KSED_RS00005 nucleotide sequence
gtgagccagacccccgacgaccacgcccacgcccacatctggcaggaggccatgggtccacctccaggggagcaggcc
tggcccccgcgcgacatcggggtgctccggctggccacgctcgtgggtctgctggagggaactgcctgctcgc
ggtgaagtacgaccacgtcaaggacgcccgtcgaggggacactgcccgcgaggacgtgtccaccgcccctggcggag
gtcctggaccgtgacatccggctggccgtctcgggtggaacccgatgcccgtgagcgcgcccaggaggaggccg
cacccccggcccctgcccggccgatgaggacgaccggccacaggtgagggaccgttgtccacagctgtgga
cggagccctggaaaagcagaggggaagcagtcgggcacgtgcccggggaatcgggtggcgcggccacgacggcc
agcctgacggcgcaaaactcctcaccgggtgaggagcgcgattactccgcgctgaaccacaagtaactttcg
acaccttgcgtgctggggtcgtcgaaccgtttcggcccacgcccagcgcgaccgcccgtggccgaagccccgcgg
cgectacaaccgcgtgttcatctacggcggatcaggtctgggcaagaccacctgttgacgcccacatcgggccac
tacgcccgcaccctggattcctcgggtgcccgtgaagtaactcggaggagttcaccacaccgattcatca
acgcccgtctcggccggccaggcgaatgccttccagcgcagtagtggacgtcctgctcatcgacga
catccagttcctgcagggcaaggagcagacgatggaggagttcttccacacctcaacacctgcacaacagc
gagaagcagatcgtcatcacctccgaccagccccgaagaagctcagtggtctcggcagcgcgatgcccgtcgc
gtttcagtgagggtcgtcaccgacgtgacgcccggacacctggagaccgcatcgcgatcctcgggcgcaa
ggcagcggccgacaagctggacatccccgatgacgtgctccacctcatcgctcgaagatctcctogaacatc
cgcgagctcgagggggccctgacccgggtgacggccttgcgagcctgtccgggtcgcccctggacgagtagtacc
tggcccgcacgggtgctcaaggacgtgatgccggcgggtgacagcggccagatcacgcccacgatgatcctgga
ggagaccgcccgggtacttctcgtcatctcggctcagaggatccagggcgcctcccgcctcgcgcaacctgacccgg
gcccggcagatcgccatgtacctgtgcccgcgagctcacggacctctcgtgcccgaagatcggcaaggagttcg
gcggccgcgaccacacgaccgtcatgacgcccagcgaagatcaagcagctgctcggggaggaccgcccgggt
ctacgacgaggtgagcgaagctcaccagcatcatccgcaagaaggcggcgcgcccggctga
```

Sequence Length

1521 nucleotides

An example completed notebook section for the nucleotide sequence and nucleotide sequence length.

A FASTA header is added 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 “>” and end with a “return” to start the sequence on a new line. Any information that follows the > on the line will not be used when the sequence information is submitted to a database for a search.

FASTA formatting allows you to keep track of the sequences you are working with and will be used routinely during your annotations.



Protein Sequence

go to the IMG Gene Details page for the proposed gene

Amino acid sequence (FASTA format)

```
>KSED_RS00005 amino acid sequence
VSQTPDDHATAIWQEAMVHLQGAGLAPRDIGVLRRLATLVGLLEG TALLAVKYDHVKDAVEGHLREDVSTALAE
VLDRDIRLAVSVDPDAVSAAQEEAAPPAPSPAEDDDPATGEGPLSTAVDGAVEKHEGSSPARAGESVAPATTA
SLTATNSSPGVERDYSALNHKYTFDTFVLGSSNRFAHAATAVAEAPARAYNPLFIYGGSGLGKTHLLHAIGH
YARTLDSSVRVKYVNSEEF TNQFINAVSAGQANAFQRQYRDVDVLLIDDIQFLQGKEQTMEEFFHTFNTLHNS
EKQIVITSDQPPKKLSGFAERMRSRFEWGLLTDVQPPDLETRIAILRRKAAADKLDIPDDVLHLIASKISSNI
RELEGALTRVTAFAASLSGSP LDEYLARTVLKDVMPGGDSGQITPTMILEETAGYFVISVEEIQGASRSRNLTR
ARQIAMYLCRELTDLSLPKIGKEFGGRDHTVMHAERKIKQLLGEDRRVYDEVSELTSIIRKKAARGR
```

Sequence Length

506 Amino acids

Be sure to save this document after completing the basic information module!

An example completed notebook section for the amino acid sequence and amino acid sequence length.

A FASTA header has also been added to the amino acid sequence and the * has been omitted from the amino acid sequence found on the gene information page.



Have your instructor review your notebook after completion

Once your instructor feels the notebook has been completed correctly to the best of their knowledge, upload the notebook to your geni-act assignment as follows.

REACHING OTHERS



Basic Information

Basic Information Module

Download the file entitled 1. Basic Information.docx [here](#) (be sure to open in a new tab) and save it to your computer using the format "locustag_basicinformation.doc" (or save it as a Google Doc with the same name in a folder on your Google drive).

Instructions for completion of the Basic Information Module can be found in the hard copy of the manual provided to you by your instructor, or the most recent update to the Basic Information module instructions can be found [here](#) (be sure to open in a new tab).

Fill in the required information and then have your instructor review it for accuracy.

Once approved by your instructor, you should attach the word file below. Be sure to clearly identify any information that you have found that differs from that supplied by GENI-ACT.

Upload your completed and reviewed Basic Information notebook here.

🔗 Basic Information Notebook - Collecting for: [KSED_RS00005](#)

course

Upload your completed and reviewed Basic Information notebook here.

Choose File KSED_RS00005_Basic Information Module.docx

Submit

After your notebook has been saved and reviewed for accuracy, return to the Basic Information assignment for your gene in GENI-ACT.

Upload the reviewed notebook by clicking the "Choose File" link (arrow) and selecting the notebook you wish to upload (note the file naming nomenclature in the file to be uploaded in this image).

Click "Submit" to upload the file

REACHING OTHERS



🔗 Basic Information Notebook - Collecting for: [KSED_RS00005](#)

course

Upload your completed and reviewed Basic Information notebook here.

Choose File no file selected

Submit

Value

Uploader

[KSED_RS00005_Basic_Information_Module.docx](#)

S. Koury

You will see the file you uploaded below the file submission window. Check back at regular intervals to see if any feedback has been provided (i.e., suggestions for correction of data submitted). If feedback has been provided you will see a comment appear in your Basic Information assignment. You can then make the corrections, delete your original file and then upload the edited version of the notebook.

If the notebook is reviewed and not found to have any errors your GENI-ACT mentor will recommend it for review and inclusion as a contribution credited to you on the gene information page of the gene.