

TEACHING THE GENOME GENERATION™

SEQUENCE ANALYSIS PROTOCOL

BEFORE YOU BEGIN

Download the sequence data files:

Download .ab1 sequence files from The Jackson Laboratory for your gene(s) of interest:

<https://thejacksonlaboratory.box.com/s/irrqmvpihxbwkqu5s2yrur8kucb1y68>

Prepare your computer:

Create a free account at Benchling www.benchling.com to analyze sequence files without downloading software

Optional software download:

For Macs — 4Peaks www.nucleobytes.com/4peaks/index.html

For PC — Chromas www.technelysium.com.au/wp/chromas

PREREQUISITES & GOALS

STUDENT PREREQUISITES

Prior to implementing this lab, students should understand:

- All previous prerequisites
- The connection between the genotypes and the DNA sequence
- The benefits of knowing a DNA sequence and its applications
- Watch the video “How does Sanger Sequencing Work? — Seq It Out #1” from Thermo Fisher Scientific:
www.youtube.com/watch?v=e2G5zx-OJlw

Additionally, students should have completed the Bioinformatics Exercises and understand:

- The DNA sequence for a gene contains both introns and exons and both can harbor sequence variants
- The effect of different types of mutations
- How DNA sequences code for proteins and how DNA mutations can affect amino acid sequence
- What the NCBI website and BLAST tool are and what information they can provide

STUDENT LEARNING GOALS

1. Interpret sequence quality and genotypes for ACTN3, CYP2C19 and/or TAS2R38 among sequenced individuals.
2. Locate the specific SNP variant(s) within DNA sequences.
3. Correlate restriction enzyme results for CYP2C19 data with sequence data, demonstrating that two techniques can be used to genotype.

NOTES

This protocol examines DNA Sanger sequence files provided by the The Jackson Laboratory. Several options are available for analysis, depending on computer setup and preferences. For the simplest method that does not require any software to download, this protocol follows steps using the Benchling web-based platform.

Watch our video tutorials, which take you through this sequencing analysis protocol step-by-step. Find these at The Jackson Laboratory’s YouTube Playlists for Teaching the Genome Generation Sequence Analysis:

Benchling, no download needed

www.youtube.com/playlist?list=PLWNP6Z5dXDZ6wvvyQdoiD1XaGwu5DFjkv

4Peaks for Mac, download needed

www.youtube.com/playlist?list=PLWNP6Z5dXDZ6pDkLBG7Su1-hldLsC0Yx_Z5dXDZ6pDkLBG7Su1-hldLsC0Yx_

Chromas for PC, download needed

www.youtube.com/playlist?list=PLWNP6Z5dXDZ4TWiDzqfg3SzhCk4X2v9Yx

CURRICULUM INTEGRATION

Use the planning notes space provided to reflect on how this protocol will be integrated into your classroom. You'll find every course is different, and you may need to make changes in your preparation or setup depending on which course you are teaching.

Course name:

1. What prior knowledge do the students need?

2. How much time will this lesson take?

3. What materials do I need to prepare in advance?

4. Will the students work independently, in pairs, or in small groups?

5. What might be challenge points for students during this lesson?

MATERIALS

REQUIRED LAB MATERIALS

Computers capable of connecting to the internet (for Benchling)
or able to download software (for 4Peaks and Chromas)

PROVIDED BY JAX

For these materials please contact ttgg@jax.org.

DNA sequence files

PROTOCOL STRUCTURE

STEPS 1-14 15 minutes

Break point if needed

STEPS 15-17 35 minutes

Break point if needed

STEP 18-23 40 minutes

PROCEDURE

□ STEP 1

Sign in or create a free Benchling account at www.benchling.com.

NOTES:

- a. This program is free to use, but an email address is required.
- b. At the time this protocol was written, a “Sign Up” button on the home page directs you to enter an email address then password to create your free account.
- c. This website may change over time and exact directions may differ as Benchling launches updates.

□ STEP 2

In the main navigation panel, click the **CREATE** icon to open a drop down menu.

□ STEP 3

From the drop down menu, hover or select **DNA / RNA SEQUENCE** then select **IMPORT DNA / RNA SEQUENCES**.

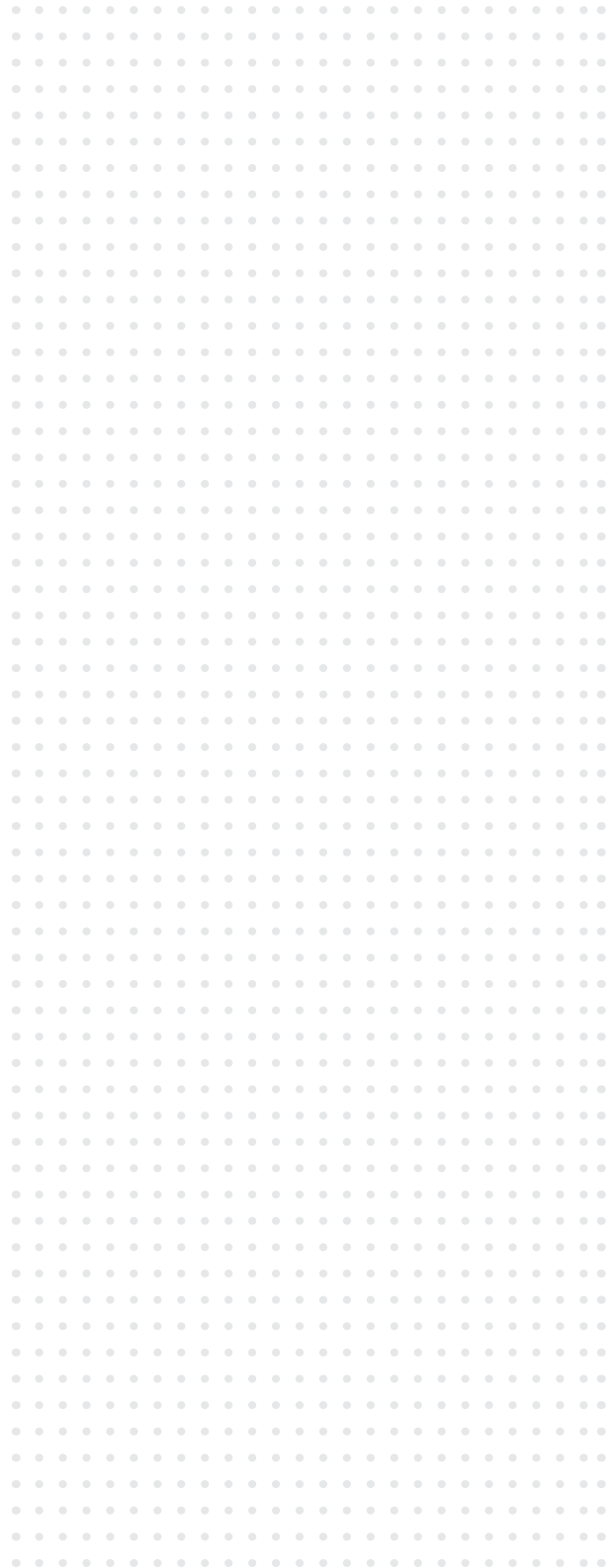
□ STEP 4

A pop-up menu called **CREATE DNA / RNA SEQUENCE** will appear. Ensure **UPLOAD FILES** is selected. Under **NUCLEOTIDE TYPE** select **DNA**. You may edit and create a folder within Benchling for this project by selecting a folder under **SET FOLDER**.

□ STEP 5

Ensure you have downloaded the sequence files provided by TtGG (see title page of protocol). Either

PLANNING NOTES

A large grid of small dots for planning notes, consisting of 20 columns and 30 rows of light blue dots.

drag and drop your “F” or “Forward” sequence into the upload box, or click **CHOOSE A FILE** and navigate to your sequence file.

NOTE:

- a. The file names are long. For example:
C09_7119_530413_1_075_2016-03-07.ab1
- b. A progress bar will appear and show when the file is uploaded to the site.

STEP 6

Select **OPEN SEQUENCE**. Your file will open in a tab within your Benchling notebook.

STEP 7

In order to view the trace file, select the **ALIGNMENTS** icon on the right-hand navigation window. Under **SAVED ALIGNMENTS**, select the file name. The window will reload showing single nucleotides and colorful traces from the Sanger sequence file.

STEP 8

You should now be able to see each individual nucleotide base call for this DNA sequence: red (T), green (A), blue (C) and black (G) peaks. The quality of those calls is indicated by light gray bars behind each of the peaks.

NOTE: You can scroll left and right to scan through your sequence file.

STEP 9

Trim your sequence to remove low-quality reads. Start at the beginning of the sequence read and find the nucleotide position where peaks are high quality (distinct, non-overlapping). Using your mouse, **RIGHT CLICK** on the base where the high-quality sequence begins. On the menu that pops up, select **TRIM...** and **TO START**.

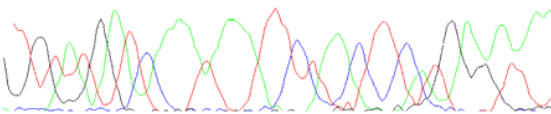
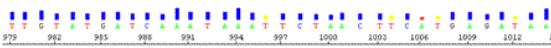
PLANNING NOTES



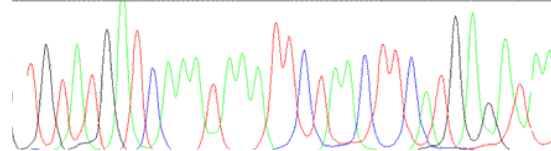
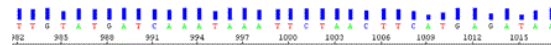
PLANNING NOTES

NOTES:

- Hovering over the grey bars will reveal the numerical quality score. A score of >40 is considered high quality.
- Heterozygous regions will have low quality scores, but that is expected.
- The first 20-30 bp and last 10 bp of almost every sequencing reaction are typically of low quality, and should be trimmed from the data file.
- You should select the beginning of a large chunk of high quality scores.
- The trimmed sequence will become shaded gray.



Example of low quality sequence trace data:
peaks are broad, jagged and overlapping



Example of high quality sequence trace data:
peaks are steep, smooth and non-overlapping

□ STEP 10

Trim your sequence to remove low-quality reads, now from the end of the sequence read. Find the nucleotide position where peaks are high quality (distinct, non-overlapping) and using your mouse, **RIGHT CLICK** on the base where the high-quality sequence ends. On the menu that pops up, select **TRIM...** and **TO END**.

NOTE: Your high-quality sequence should be unshaded.

□ STEP 11

Save this sequence in a text file for future use. Select your high quality sequence by selecting the entire high-quality region.

NOTE: Use may use SHIFT to click on the beginning, scoll, then click on the end of the high-quality sequence.

□ STEP 12

RIGHT CLICK on your highlighted sequence and select **COPY**.

□ STEP 13

Paste into a Word or Text document. Above the sequence, add a ">" symbol in the header followed by the sample identifier. For example:

```
>ACTN3_GROUP1 F
GGACTTAATTTGCGCAATTGNGGCCATATGTTTAAA
```

□ STEP 14

Save the file as **TEXT** only. This will enable you to have the sequence for future activities.

BREAK POINT IF NEEDED

□ STEP 15

Return to Benchling. To query the genomic origin of your sequence file, submit your high quality sequence to the NCBI database BLAST. Select the high-quality sequence as in STEP 11. **RIGHT CLICK** and select **NCBI BLAST**. A new window will appear. Scroll to the bottom of the NCBI submission page and click **BLAST**. Wait for the results to load and record your observations. The results table gives reveals sequences from organisms that match your query sequence, as well as percent identity (how much it matches).

PLANNING NOTES

Record the top result listed from your BLAST query.

Description:

Organism:

Percent Identity:

□ STEP 16

Return to Benchling and find the SNP polymorphism(s) in your sequence file, using sequence comparison. Find the search bar and enter the nucleotides immediately before the SNP of interest.

For ACTN3, search:

GGCTGAC, next SNP (C or T)

For CYP2C19, search:

ATTTCCC, next SNP (G or A)

For TAS2R38, search:

1st: AGAGGCAG, next SNP (G or C)

2nd: ATCCTCTC, next SNP (T or C)

3rd: TGCAGCC, next SNP (A or G)

NOTES:

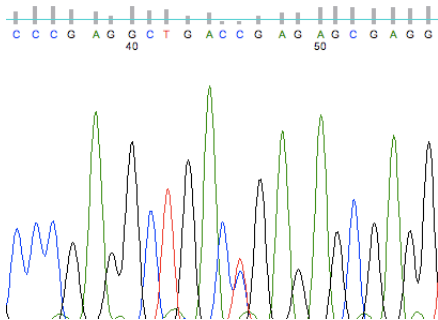
- To locate the reference sequence of interest select all of your trace file by hitting Ctrl+A and then Ctrl+F to bring up the find toolbar. Type sequence appropriate for the gene of interest into window and program will automatically navigate to the region.
- You can zoom in on the trace to assess the base call by using the vertical scroll bar on the left hand side of the trace window and shortening the viewing window on the gray bars at the bottom of the screen.

PLANNING NOTES



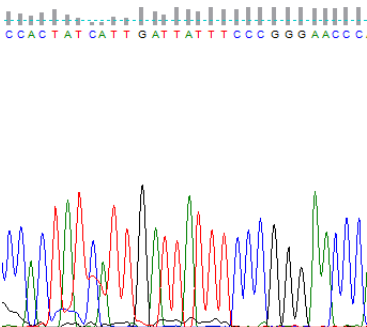
ACTN3 EXAMPLE

Look for the GGCTGAC sequence around the 40th base in the original sequence. The SNP is the next downstream base and will either be a C allele (the 577R variant) or a T allele (the R577X variant). The T creates a premature stop codon (TGA). This example is heterozygous (C/T) as it shows both T and C peaks corresponding to both bases.



CYP2C19 EXAMPLE

Look for the ATTTCCC sequence in the original sequence. The SNP is the next downstream base and will either be a G allele or an A allele. The A variant creates an aberrant splice site that causes a frameshift. This example is homozygous G/G.

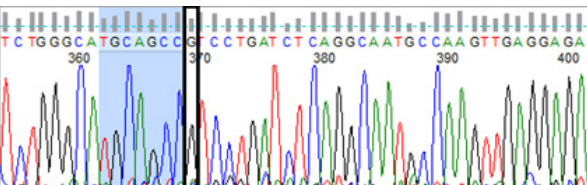
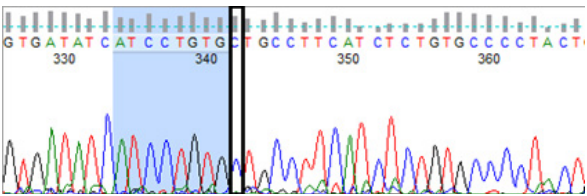
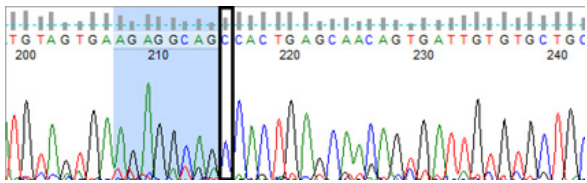


PLANNING NOTES

A large grid of small dots for planning notes.

TAS2R38 EXAMPLE

To find the first locus SNP, search for the AGAGGCAG signature. To find the second locus, search for the ATCCTGTG signature. To find the third locus, search for the TGCAGCC signature. The SNP is the next downstream base and will either be a G or a C for the first locus, either a T or a C for the second locus, and either an A or a G for the third locus. Example below is homozygous CCG allele.



□ STEP 17

Determine and record the genotype of your sample.

Gene name:

Sequence name:

Genotype:

BREAK POINT IF NEEDED

PLANNING NOTES

A large grid of dots for planning notes, consisting of 20 columns and 30 rows of small grey dots.

□ STEP 18

Repeat Steps 3-11 with the “R” or “Reverse” sequence file to view in Benchling, trim for quality, and select the high-quality region.

□ STEP 19

RIGHT CLICK on your highlighted high quality sequence. Select **COPY SPECIAL** and **REVERSE COMPLEMENT**. Select the **DNA REVERSE COMPLEMENT** option in the pop-up menu. Follow STEP 13 to paste this sequence into a text document.

NOTE: This takes the “Reverse” read and re-orientes it to match the “Forward” read. This process is done for two reasons:

1. As a double check of the (F)orward read
2. In the event that the (F)orward read is of such poor quality that it cannot be analyzed.

□ STEP 20

Align the full length forward and reverse reads to create one consensus sequence. Open one of the sequences and click the **ALIGNMENT** icon on the right-hand tool bar. Click **CREATE NEW ALIGNMENT**.

□ STEP 21

A new window will appear. Under **CHOOSE INPUT**, search for the name of the opposing sequence file using the name of the opposing sequence (i.e.: if you opened the forward file, you should select the reverse file). Ensure **NUCLEOTIDE TYPE** is **DNA**, and you see the two **SEQUENCES** load on the page. Click **NEXT**.

□ STEP 22

Under **DEFINE PARAMETERS**, select **CONSENSUS**. You may name your alignment under **GROUP(S)**, and choose a folder where you want this saved in Benchling. Choose the **AUTO (MAFFT)** alignment program for this alignment as it will automatically reverse-complement your R sequence.

PLANNING NOTES

A large grid of small dots for planning notes, consisting of 20 columns and 30 rows of light blue dots on a white background.

□ STEP 23

Click **CREATE ALIGNMENT**. This will create a new consensus alignment between the two sequences. A small window will appear on the bottom of the screen indicating the consensus sequence was created successfully. Select **OPEN** to view.

NOTE: You should see three horizontal bars at the bottom of your screen. The top is the consensus bar indicating the agreement between the two sequences. The middle is the first sequence file you chose, and the bottom is the second sequence file. Gray indicates agreement between the two files, whereas red indicates that only one file has that specific sequence. There should be a region of gray overlap between the files somewhere in the middle with red on the ends.

How much overlap is there between the (F)orward and (R)everse read?

Do the (F)orward and (R)everse reads indicate the same genotype? If not, provide a possible explanation.

Sources of Potential Error:

The most common errors in the SEQUENCE ANALYSIS PROTOCOL include:

- Copying the wrong sequence
- Using the low-quality section of the read (first 20-30 bases)
- Low read quality across the entire sequence, which could be due to:
 - » Pipetting errors in previous protocols
 - » Target gene not amplified
 - » Contamination of DNA samples or reagents

PLANNING NOTES

NEED HELP?

Email the experts: ttgg@jax.org