Updated 08/13/2024

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/v/JAXDNASeqFiles*

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

PREREQUISITES

Prior to implementing this lab, you 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-0JIw*

Additionally, you should have completed some 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

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 Playlist for Teaching the Genome Generation:

tinyurl.com/TtGG-Protocol-Videos

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

DNA sequence files

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 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 progress bar will appear and show when the file is uploaded to the site.

NOTES

☐ 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**.

NOTES:

a. Hovering over the grey bars will reveal the numerical quality score. A score of >40 is considered high quality.

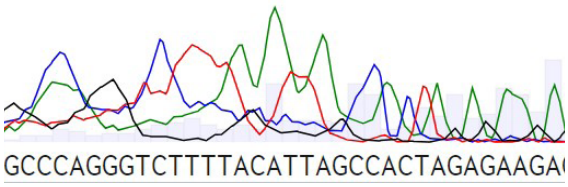
b. Heterozygous regions will have low quality scores, but that is expected.

c. The first 20-30 bp and last 10 bp of almost every sequencing reaction are typical of low quality, and should be trimmed from the data file.

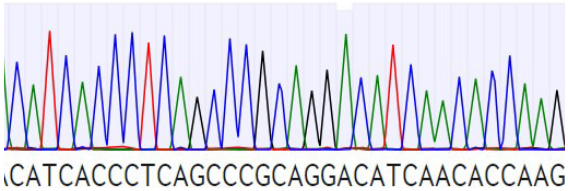
d. You should select the beginning of a large chunk of high-quality scores.

e. The trimmed sequence will become shaded gray

NOTES



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

To confirm the sequence matches the gene you expect, submit your high-quality sequence to NCBI BLAST. Select the high-quality sequence, then **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 reveals sequences from organisms that match your query sequence, as well as percent identity (how closely the sequences match).

NOTE: To select the high-quality sequence, press SHIFT and click on the beginning, scroll, then click on the end of the high-quality sequence.

NOTES

Record the top result listed from your BLAST query.

Description:

Organism:

Percent Identity:

BREAK POINT IF NEEDED

☐ STEP 12

Return to Benchling. Repeat Steps 3-10 with the "R" or "Reverse" sequence file to view it in Benchling and trim for quality.

☐ STEP 13

Return to the "F" or "Forward" sequence file. Align the full length forward and reverse reads to create one consensus sequence. Click the **ALIGNMENT** icon on the right-hand tool bar. Click **CREATE NEW ALIGNMENT**.

☐ STEP 14

A new window will appear. Under **CHOOSE INPUT**, search for the name of the reverse sequence file. Ensure **NUCLEOTIDE TYPE** is **DNA**, and you see the two **SEQUENCES** load on the page. Click **NEXT**.

☐ STEP 15

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 reverse sequence.

NOTES

☐ STEP 16

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.

BREAK POINT IF NEEDED

NOTES

☐ STEP 17

Find the SNP polymorphism(s) in your consensus sequence. 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)

For OXTR, search:

ATGCCCGAGG next SNP (G or A)

NOTES:

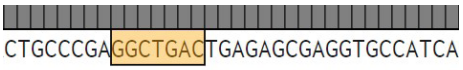
a. To locate the reference sequence of interest select all of your trace file by hitting Ctrl+A and then Ctrl+F to bring up the find toolbar. Type sequence appropriate for the gene of interest into window and program will automatically navigate to the region.

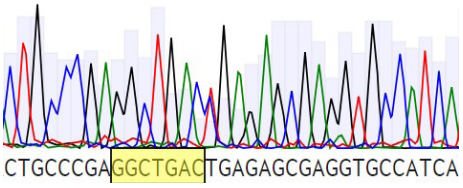
b. You can zoom in on the trace to assess the base call by using the vertical scroll bar on the left hand side of the trace window and shortening the viewing window on the gray bars at the bottom of the screen.

NOTES

ACTN3 EXAMPLE

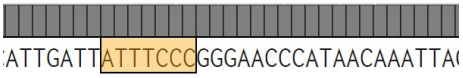
Search for the sequence GGCTGAC. The SNP is the next downstream base and will either be a C (the 577R variant) or a T (the R577X variant). The T creates a premature stop codon (TGA). This example is heterozygous CT, as it shows both T and C peaks.

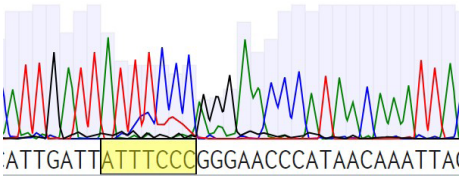




CYP2C19 EXAMPLE

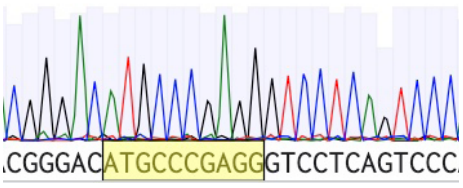
Search for the sequence ATTTCCC. The SNP is the next downstream base and will either be a G or an A. The A variant creates an aberrant splice site that causes a frameshift. This example is heterozygous





OXTR EXAMPLE

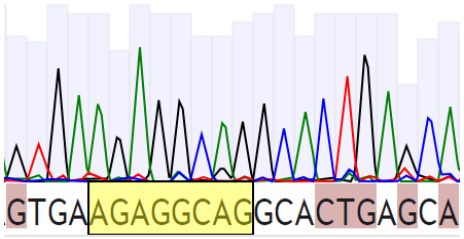
Search for the sequence ATGCCCGAGG. The SNP is the next downstream base and will either be a G or an A. This example is homozygous G.

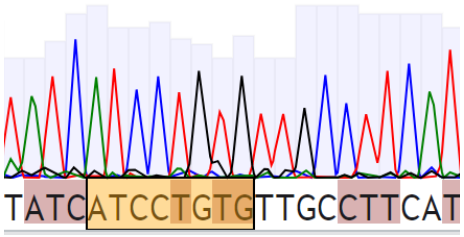


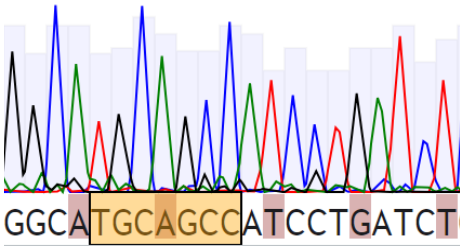
NOTES

TAS2R38 EXAMPLE

To find the first SNP, search for the sequence AGAGGCAG. To find the second SNP, search for the sequence ATCCTGTG. To find the third SNP, search for the sequence TGCAGCC. The SNP is the next downstream base. The first SNP will be either a G or a C, the second will be either a T or a C, and the third will be either an A or a G. This example is homozygous GTA.







☐ STEP 18

Determine the genotype of your sample.

Gene name:

Sequence name:

Genotype:

NOTES