

GENOMICS

INTRODUCTION

In 1992, Craig Venter proposed an innovative approach to genome sequencing called whole-genome “shotgun” sequencing. Although this idea was met with ridicule, Venter pursued his method and was able to quickly sequence the bacterial genome of *Haemophilus influenzae*. Meanwhile, the Human Genome Project was underway headed by the US Government using slower traditional methods. In 1998, Venter formed the company Celera Genomics and jumped into the fray by declaring his intentions to finish the human genome before the government consortium. Both groups completed the genome at about the same time and jointly announced its completion in 2003.

The technique of shotgun sequencing involves breaking chromosomal DNA into small pieces to create a series of overlapping DNA fragments called contiguous sequences or “contigs.” The breakage can be done by physical shearing of chromosomal DNA or digestion with different restriction enzymes. Contigs are subjected to DNA sequencing and then bioinformatics-based programs are used to arrange contigs in their correct order based on overlapping sequences of nucleotides. For a very nice animated tutorial on contigs and whole genome sequencing methods click the link below.

<http://smcg.cifn.unam.mx/enp-unam/03-EstructuraDelGenoma/animaciones/humanShot.swf>.

In this exercise, you will carry out a simulation of contig alignment to help you understand how contigs can be arranged to create sequence maps of a chromosome. You will be given one of four unknown gene fragments and you will go to the National Center for Biotechnology Information BLAST site to use a DNA alignment program called “bl2seq.” After piecing together the contigs, you will enter the sequence into Genbank to search for a match. You will then learn about the gene fragment you have just pieced together.

ARRANGING CONTIGS TO CREATE A MAP

1. Click on this link <http://www.cabrillo.edu/~ytan/Bio1AF04/Genomics%20sequences.htm> to access the four gene fragments. You will be assigned one of the gene fragments for your lab work and report. Each gene fragment has been taken from actual human chromosome sequences deposited in GenBank and are involved in important biological processes. Each fragment has been broken into six contig sequences, rearranged, and labeled A through F. For this exercise we have used short fragments but in reality contigs would usually be several thousand base pairs long.
2. In a separate tab or window, access BLAST from the NCBI website at <http://www.ncbi.nlm.nih.gov/BLAST/>. Locate the “Align two sequences using BLAST (bl2seq)” category at the bottom of the BLAST homepage. Click the link for “Align.”

The bl2seq feature allows you to compare two DNA sequences to check for sequence similarity alignments.

3. To complete this exercise, copy and paste any two sequences from your gene fragment into the “Align” feature of BLAST then run an alignment. Your goal is to determine which sequences overlap and to create a sequence map showing overlapping contigs. Here are a few important things to consider:
 - a. Develop a strategy to be sure that you have analyzed alignments for all pairs of contigs.
 - b. Keep a running list of overlapping pairs as you go. You will hand this in with your report.
 - c. Only consider alignment overlaps that show 100% sequence similarity.
4. Piece together your contigs and draw them out to give you a complete overlapping sequence map. You will hand in a map with your report.

SEARCH GENBANK TO ANALYZE YOUR GENE

1. Access BLAST from the NCBI website at <http://www.ncbi.nlm.nih.gov/BLAST/>. Under “Basic BLAST” find and click on “Nucleotide BLAST.”
2. In the box under “Enter Query Sequence” paste in the sequence from the first contig of your map.
3. Click the button “BLAST” at the bottom of the page to run the search.
4. You will now get a screen that is extremely information rich. Take your time to explore this page. Some tips to exploring this page are below:
 - a. The “Graphic Summary” area shows you how well your sequence aligned with Genbank data. Because I took the sequence directly from Genbank, you should have 100% alignment.
 - b. The “Descriptions” area gives you click-able information on your gene. You will notice that you will have one or more transcripts listed. Find the one with the maximum identity and click the links on the right. “G” will give you gene information and “M” will give you chromosomal mapping information. When looking at the chromosome, be sure to zoom in and out using the links on the left.
 - c. The “Alignments” area shows you how well your output sequences lineup with Genbank sequences.
 - d. As you are exploring information on this site, you will want to look at the lab report questions below for what information to look for. Be sure to bookmark or write down references so that you can access the information later and so that you can cite it if needed.

LAB REPORT QUESTIONS

1. Present an abstract of your work. Your abstract should include a brief description (1/2 page) of 1) what you set out to learn, 2) what procedures you performed, and 3) a summary of your data or discoveries.
2. Present a table showing matched pairs of contigs from your gene fragment.
3. Present a map of your gene fragment showing the overlapping contigs. Be sure each contig is labeled (A-F) and that they are in the correct order.
4. Give a brief summary of your gene. Give its abbreviation, full name, function in the cell, and chromosomal location.
5. (2 pts) Your gene is important in a human disease. Give a detailed description of its role in the disease. Cite at least one reference for this section.
6. (2 pts) Describe the impact of the disease on society. If you have any personal connection with the disease, you may discuss it too.
7. (2 pts) Describe a way in which your gene can be manipulated to treat the disease. Assume you can make any changes to the protein product and describe specifically how it will affect its interaction with other molecules.