Bio 211 Genetics Laboratory
Experiment 5: Bioinformatics

Bioinformatics is the field and study of biological information in DNA using computer-based approaches. Through program algorithms, coding sequences, promoters, and other functional DNA sequences can be identified from databases of genomic information, and interspecific comparisons can be made to address evolutionary relationships. One of the most widely-used in silico (in silicon, or on the computer) tools is the Basic Local Alignment Search Tool (BLAST).

A. PCR Amplification of TAS2R38 Gene

The two primers that were used for your PCR experiment are shown below to amplify a 172-bp DNA sequence of the TAS2R38 gene.

forward primer:

5' CCTTCGTTTTCTTGGTGAATTTTTGGGATGTAGTGAAGAGGCGG 3'

reverse primer:

5' CTGTGGTTCAGTGGTTCACTCAACTTCTGG 3'

These primers produce the following PCR products, also referred to as amplicons.

Amplicon for non-taster allele:

CCTTCGTTTTCTTGGTGAATTTTTGGGATGTAGTGAAGAGGCGGGCACTGAGCAACAGTGAT
TGTGTGCTGCTGTGTCTCAGCATCAGCCGGCTTTTCCTGCATGGACTGCTGTTCCTGAGTGCT
ATCCAGCTTACCCACTTCCAGAAGTTGAGTGAACCACTGAACCACAG

Amplicon for taster allele:

CCTTCGTTTTCTTGGTGAATTTTTGGGATGTAGTGAAGAGGCGGCCACTGAGCAACAGTGAT
TGTGTGCTGCTGTGTCTCAGCATCAGCCGGCTTTTCCTGCATGGACTGCTGTTCCTGAGTGCT
ATCCAGCTTACCCACTTCCAGAAGTTGAGTGAACCACTGAACCACAG

B. Using BLAST to Search for DNA Sequences

1. Visit the National Center for Biotechnology Information (NCBI) website at

   http://www.ncbi.nlm.nih.gov

   a. Click on the BLAST link under Popular Resources in the right column.

   b. Click on nucleotide blast under the Basic BLAST section.

c.   Enter the following primer sequences in the Enter Query Sequence box. Do a carriage return after the first sequence (enter each primer sequence as a separate line, one directly below the other). The reverse primer, in this example, is a commercially available DNA sequence to amplify the same region of the TAS2R38 gene.

CCTTCGTTTTCTTGGTGAATTTTTGGGATGTAGTGAAGAGGCGG
AGGTTGGCTTGGTTTGCAATCATC

d.   Under Choose Search Set, use the drop down menu to Nucleotide Collection (nr/nt).

e.   Under Program Selection, click on Somewhat similar sequences (blastn).

f.   Next to the BLAST button, click on Show results in a new window.

g.   Click on the BLAST button. The BLAST algorithm will query a match between the primer sequences and the millions of DNA sequences in the NCBI database.

2.   The results of your BLAST search should appear in a separate browser window, and the results page contains 3 different sections.

a.   The first section, titled Graphic Summary, shows the number of matches or hits. Color differences of the bars indicate matches of different lengths. Clicking on a bar will link you to the third section.

b.   The second section, titled Descriptions, lists sequences with significant alignments and information links about each. Clicking on a Max. score number for an Accession item (such as AY258957.1) will link you to the third section. Clicking on the Accession link will take you to a page of information about the matched DNA sequence. E-value (or Expection value) for a matched DNA sequence is the expected number of occurrences of alignments with the query sequence. A low E-value (such as 6e-12) indicates a high probability that the match is uniquely related to the query sequence.

c.   The third section, titled Alignments, provides detailed alignment between each primer sequence (Query) and a match to a DNA sequence (Sbjct) in the database. Alignments that have E-values less than 0.1 tend to be related to each other. In this analysis, all such alignments are related to bitter taste receptor genes in humans and primates.

Exact match between the primer base and the corresponding base in the matched sequence is indicated by a vertical line, while a mismatch does not have this vertical line. For example, find gb|AY258597.1| and you will note that base position 43 of the forward primer does not match the DNA sequence in the database. The forward primer that you used for your analysis, by design, produces amplicons that contain G-C at that position instead of A-T. The

significance of this change is the creation of an *Hae*III site (GGCC) in the amplicon from the taster allele sequence, but not in the amplicon from the non-taster allele (GGGC). The diagnostic utility of this change is that, the two different amplicons can be distinguished from each other by the presence or absence of this *Hae*III site.

The borders of the amplicon are defined by the lowest and the highest nucleotide base positions in the subject (sbjct) sequence. The length of the amplicon is the difference between these two numbers plus one (since both ends are included in the actual length of the amplicon).

3. Complete DNA sequences that correspond to hits (such as PTC taster and non-taster alleles in Homo sapiens) can be obtained from the Description (second) section of this BLAST search.

    a. From the list of sequences producing significant alignments with the lowest E-values, find Homo sapiens PTC bitter taste receptor (PTC) gene, PTC-taster allele and click on the Accession link to view information on this match.

    b. The data page provides basic information about the DNA sequence, such as the total nucleotide length, biological source of the sequence, references, initial and terminal base position numbers, translated amino acid sequence, and the complete nucleotide sequence

    c. Click on the FASTA link at the top of the page to obtain the DNA sequence. Make a copy of the complete nucleotide sequence into a text or Word document. Be sure that the base sequence is continuous.

    d. Follow steps a. – c. to make a copy of the complete nucleotide sequence for:

        -   Homo sapiens PTC bitter taste receptor (PTC) gene, PTC-taster allele
        -   Homo sapiens PTC bitter taste receptor (PTC) gene, PTC-non-taster allele
        -   Pan troglodytes (chimpanzee) TAS2R38 gene (AY736062.1)
        -   Pan paniscus (bonobo) TAS2R38 gene (AY677143.1)
        -   gorilla TAS2R38 gene (AY566403)

        These complete sequences will be used for analysis in part 5.

4. A useful tool is the Map Viewer, which can be accessed from the NCBI home page through the Maps & Markers link in the left column. Click through the Map View link under Quick Links in the right column to access chromosome location of a DNA sequence in a species and information about genes that neighbor the DNA sequence on that chromosome.

    a. Click on the Map Viewer link, then click on the BLAST (B) button located on the same row for Homo sapiens.

    b. Enter the two primer sequences from B.1.c. in the Enter an accession, gi, or a sequence in FASTA format: box. Choose genome (all assemblies) from the

Database drop down menu and BLASTN: Compare nucleotide sequences from the Program drop down menu. Finally, click on Begin Search button at the bottom left of the page.

c. On the results page, click on the View report button. This will link you to a BLAST search results page, titled NCBI/ BLAST/ Formatting Results.

d. Under the Descriptions (second) section, click on a choice from the list of Sequences producing significant alignments. You will be directed to a results page. Near the top of the page, chromosome location of the alignment is shown.

e. Change the zoom bar in the left column to 1/100 (corresponding to the middle of the zoom range). In the Genes_seq mode (second column image from the right, between RefSeq RNA and Contig modes), you will see genes that neighbor the TAS2R38 gene. To obtain more extensive information about any of these genes, click on the name of the gene to open an information box, then click on the link after SYMBOL: at the top of the box.

5. Evolutionary comparisons can be made from multiple sequence alignments.

a. Visit the Dolan Learning Center website at

http://www.bioservers.org   which automatically redirects to
http://www.bioservers.org/bioservers/

b. Under the SEQUENCE SERVER column on the left, click on the ENTER button. (Registration is recommended if you wish to save your work for future reference.)

c. To conduct sequence comparisons, you need to first create individual files for each DNA sequence that you wish to use in multiple sequence alignments. For each sequence, create a file as follows.

1) Click on the CREATE SEQUENCE button at the top of the page. This opens a new browser window.

2) In the Name: box, type in a name for the file, such as Taster Allele.

3) In the Sequence: box, cut an paste the complete sequence of the PTC Taster Allele nucleotide sequence that you saved in a text file from B.3.

4) Click on the OK button to save the file and sequence. This should now appear on the SEQUENCE SERVER page.

5) Create individual files for all of the other sequences you saved from B.3.d. In addition, create a file called "220" for the 220 nucleotides of the PTC taster allele shown below that corresponds to the 220 basepair amplicon sequence (based on primers from B.1.c.):

ccttcgttttcttggtgaatttttgggatgtagtgaagaggcagccactgagcaacagtgattgtgtgctgctgtgtc
tcagcatcagccggcttttcctgcatggactgctgttcctgagtgctatccagcttacccacttccagaagttgagtg
aaccactgaaccacagctaccaagccatcatcatgctatggatgattgcaaaccaagccaacct

d. Multiple sequence alignments are conducted by the following steps. The
    CLUSTAL W algorithm is the default sequence alignment program that is
    executed on a server at Cold Spring Harbor Laboratory, when sequences are
    submitted for analysis from the Dolan Learning Center's SEQUENCE SERVER
    site.

    1) Click on the boxes located to the left of the names of the files you wish to
        compare.

    2) Click on the COMPARE button. This opens a new browser window (titled
        CLUSTAL W ALIGNMENT) for the results of sequence alignments.

    3) Enter 1100 in show (number) per page box to display complete DNA
        sequences on a single page. Sequence alignments are displayed in rows of
        25 nucleotides each. Highlighted areas in yellow indicate mismatches
        between the sequences.

Bioinformatics Assignment

1.   Use the Map Viewer at the NCBI website as outlined in section **B.4.**

     a.   To which chromosome is the *TAS2R38* gene mapped?

     b.   List the name (symbol) of genes located immediately on each side of the *TAS2R38* gene (one gene per side).

          Click on the names of the genes you listed above and follow links for more information about them. The *TAS2R38* gene contains no introns. Is this also the case for these genes?

     c.   Set the zoom bar to 1/100 and click on up to 15 neighboring genes beyond the most adjacent genes on each side of the *TAS2R38* gene. *TAS2R38* is a gene for a taste receptor. In general, is there a functional similarity about the gene products of genes surrounding the *TAS2R38* gene (what type of proteins do these genes encode and what do they do)? List some of the general functions of these gene products (proteins).

2. Refer to secttion **B.5** to address the following questions.

    a. Compare the DNA sequences of human PTC taster allele, human PTC non-taster allele, and the 220 basepair amplicon.

       Does the amplicon cover the entire PTC taste receptor gene sequence? How did you base your answer?

       From what beginning base position number to what ending base position number does the 220 basepair amplicon match the two human *TAS2R38* alleles?

       At which base position number is the SNP between the PTC taster allele and the PTC non-taster allele, and what is the difference in nitrogenous base between them?

    b. Compare the DNA sequences of the human PTC taster allele and the human PTC non-taster allele. Consider the complete nucleotide sequence for each.

       List all base position(s) and nucleotide differences of SNP.

       Count triplets of nucleotides from the **initial ATG** start codon to determine codon(s) affected by the SNP(s) you have identified. Determine if an amino acid is changed by each SNP, using a standard genetic code dictionary.

c. Compare the DNA sequences of human PTC taster allele, human PTC non-taster allele, chimpanzee, bonobo, and gorilla.

What is the ancestral state (most common nucleotide base) at base positions 145, 785, and 886?

Position: 145   785   886

Ancestral state: _____   _____   _____

Are chimpanzee, bonobo, and gorilla tasters or non-tasters?

Does this suggest a selective advantage to either phenotype? If so, how might being a taster or non-taster confer a selective advantage for these species?

d. Compare again the DNA sequences of human PTC taster allele, human PTC non-taster allele, chimpanzee, bonobo, and gorilla, except choose Phylogenetic Tree as the option in the drop down menu next to the COMPARE button. **Print a copy of the resulting page.**

Based on the resulting phylogenetic tree (Tree Style: Cladogram; Use length? No), which two sequences are most evolutionarily related?

Based on the resulting phylogenetic tree (Tree Style: Phenogram; Use length? Yes), which two sequences are most evolutionarily separated?