

DNA Sequencing Overview

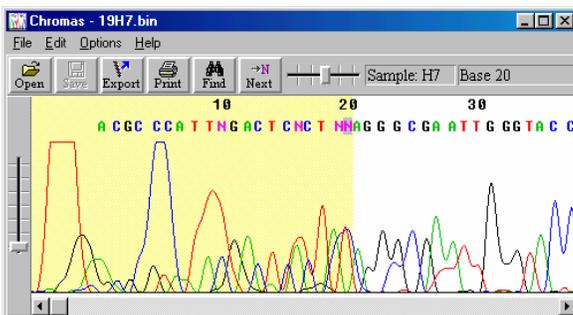
DNA sequencing involves the determination of the sequence of nucleotides in a sample of DNA. It is presently conducted using a modified PCR reaction where both normal and labeled dideoxy-nucleotides are included in the reaction mix. The dideoxy-nucleotides lack the OH at their 3' carbon atom and are labeled with fluorescent dyes (each nucleotide has a different color). During PCR, mostly normal nucleotides are added to the DNA template. However, when a labeled nucleotide happens to add to the growing DNA strand, chain elongation stops because there is no 3' OH to which the next nucleotide can be attached. Hence, the dideoxy method is also called the chain termination method. At the end of the sequencing reaction, the fragments with fluorescent termini are separated by length from longest to shortest using a polyacrylamide gel (either a big thin slab gel or a narrow capillary tube filled with gel solution) that is scanned with a laser detection device. As each band moves past a viewer, the laser excites the dye, and the color of fluorescence is read by a photocell and recorded on a computer. A computer program evaluates the resulting histogram of fluorescence peaks and “calls” the nucleotides in the order they passed through the viewer.

Interpreting Sequencing Results

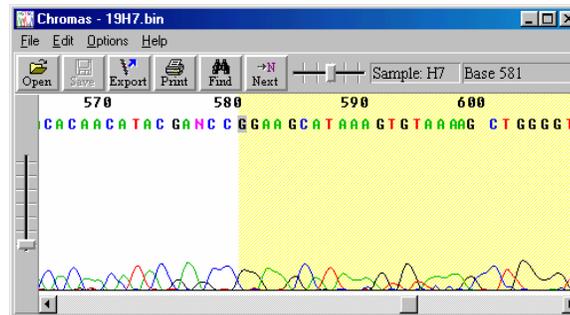
The computer output for a sequencing run consists of chromatogram (*typically a sequencing facility will email you a file*) which can be opened in one of many free DNA sequence viewers that are available on the web. We will be using a program called FinchTV. When viewing chromatograms, keep in mind that there is some ambiguity at various sites along the DNA sequence as the sequencing machine that is used to “call” the bases can't always precisely determine what nucleotide is represented by either a broad peak or a set of overlaying peaks. In such cases, a letter other than A, C, G, or T is recorded, most commonly “N”. When you obtain a sequence you should proofread it to ensure that all ambiguous sites are correctly called and determine the overall quality of your data. At the very minimum you should proofread and evaluate the quality of your sequence prior to uploading it to NCBI. If your sequence is suboptimal, you want to redo the sequencing or have both a forward and reverse sequencing reaction done and compare your two results to alleviate any possible “miscalls”.

Assignment: Viewing your sequence and Analyses

Open the FinchTV program. (Finch TV folder is on desktop, open it and hit the icon that resembles a TV). You will be directed to “drag” a sequencing file into the “screen”. Drag in your sequencing file (Sequencing files are present in a file on the desktop). Examine your sequence from beginning to end by using the horizontal and vertical scroll bars. Good sequence generally begins around base 20, and is represented by tall distinct peaks that have little overlap. Poor reactions will result in low or multiple peaks as illustrated below (very typical of beginning or end of a reaction).



Beginning of Sequence

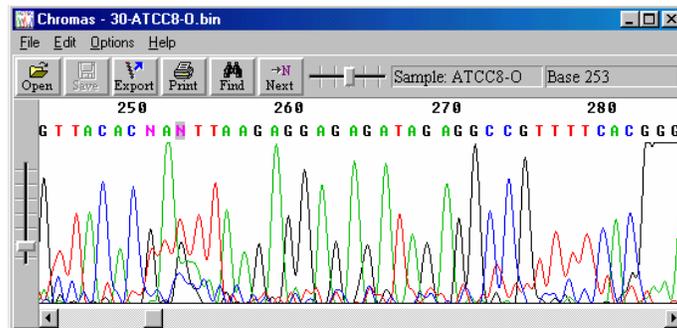


End of sequence

At what base (indicated by the numbers at the top of the sequence) does your sequence become of high quality? (Note: some will not be of high quality anywhere, if this is the case compare with that of a friend)

How long is your sequence? Does the quality deteriorate towards the end? Do you find a stretch of AAA's at the end? (FYI these AAA's are added by the Taq polymerase and are non-template derived)

Next, check your sequence for ambiguity by searching for places where "N" is the base call instead of an A, C, G, or T. For example, in the case shown below, the correct nucleotide was likely a T. Do you find any instances where the N designation might not be appropriate?



Look at the overall quality of the sequence, do you find that you have nice peaks that are easy to score? Does it appear that more than one product was sequenced?

Determining homology: in other words, is your sequence similar to any other published sequences and if so, to what degree? This can be accomplished using BLAST, a program supported by the National Center for Biotechnology Information (NCBI). In FinchTV, select from the File dropdown menu the command "export FASTA sequence" and save the export onto the desktop. Open the exported file (you may need to open this up in the notepad program). You should see your sequence as a string of letters, copy the sequence and do a BLAST search at: <http://www.ncbi.nlm.nih.gov/BLAST/>

In the Nucleotide (top left) section of this page, select "Nucleotide-nucleotide BLAST [blastn]." Paste your sequence in the "Search" box. Click the "Blast!" button to submit your sequence data.

NCBI **nucleotide-nucleotide BLAST**
Nucleotide Protein Translations Retrieve results for an RID

Search

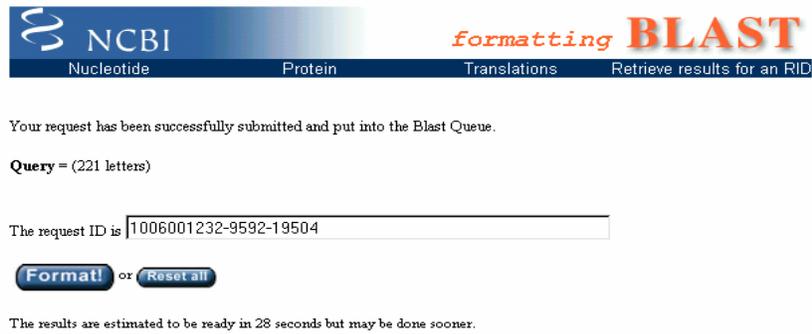
ATTCAGTGTCTTTTATTCAGAGTCAGATTTATATTATTCCTATCAGATTAACATAATA
CCGCCTTTTGTTCCTTGTAGATGTCACCAACAGTTATACAGGGTATGCAAAGTTCT
GTCAGGTAAATTGAATTTGACGTAATTAATAATTTAAACGTAAGTTAAATAATCAGCAATC
TTACAGTTAAATCCACTTTATTAACGCTTAAAGATTTATTAATGCTATAATAAAAAGAATT
GATTGGTTTAAATTACAGCTGCCATTCAAATGTTACACAGCATGTGCTCTGTGTTATC

Set subsequence From: To:

Choose database nr

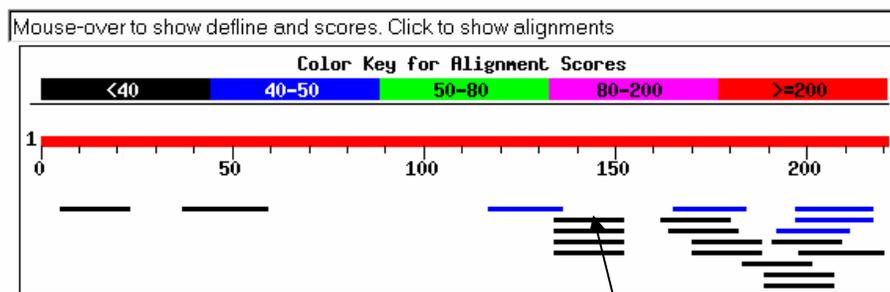
Now: **BLAST!** or Reset query Reset all

Next, click the “Format!” button so that your search results will be displayed in an html window.



Finally (sometimes after a lengthy wait), a new window will appear showing any “hits” your sequence made. The results will be color coded and annotated so that you can either click on the horizontal homology bar(s) or you can simply scroll down the page.

Distribution of 20 Blast Hits on the Query Sequence



Sequences producing significant alignments:

| | Score (bits) | E Value |
|---|--------------------|---------|
| gi 15383789 gb AC092719.3 Homo sapiens chromosome 16 clone... | 42 | 0.20 |
| gi 12053852 emb AJ278581.1 HSA278581 Homo sapiens DREV gene... | 42 | 0.20 |
| gi 14572121 gb AC008632.6 AC008632 Homo sapiens chromosome ... | 40 | 0.77 |
| gi 18745063 emb AL137006.5 AL137006 Human DNA sequence from ... | 40 | 0.77 |
| gi 2956679 emb AL008720.1 HS343C1 Human DNA sequence from c... | 40 | 0.77 |
| gi 15538860 gb AC092757.2 Homo sapiens chromosome 15 clone... | 38 | 3.1 |
| gi 15187230 gb AC016574.7 Homo sapiens chromosome 5 clone ... | 38 | 3.1 |

Links to other GenBank pages showing matching accessions.

Links to subsequent sections on the current results page.

The bars show what places along your sequence are similar to other published sequences; the colors indicate how many bases were involved in homology determination. The scores indicate how much homology occurred for particular hits. Clicking on a score or bar in the graphic representation will take you to the section of the results page that describes a sequence that is similar to yours. Clicking on a “gi” link at the beginning of any line will take you to the GenBank accession page for a sequence showing similarity to yours. There you can find a wealth of information about the published sequence to which yours showed some homology.

BLAST RESULTS

| | NCBI Best match (Genus species) | % Identity | Organization where matching sequence was uploaded | Ribosomal Database (Genus species of match) |
|---------------|------------------------------------|------------|---|--|
| Your Sequence | | | | |
| Alternate #1 | | | | |
| Alternate #2 | | | | |

To access the 16S rRNA database: <http://rdp.cme.msu.edu/>

Skip Nav

Ribosomal Database Project - II

[About](#) | [Announcements](#) | [Citation](#) | [Contacts](#) | [Resources](#) | [Related Sites](#)

Ribosomal Database Project-II Release 9

The Ribosomal Database Project (RDP) provides ribosome related data and services to the scientific community, including online data analysis and aligned and annotated 16S rRNA sequences.

Release 9 of the RDP is designed to deliver up-to-date data to our users in an era of cheap sequencing. Release 9 contains 184,990 aligned and annotated Bacterial small-subunit rRNA sequences, and is updated frequently. More on Release 9 is available in the [release notes](#).

RDP Analysis Tools

Hierarchy Browser Browse a phylogenetic hierarchy and compile a list of 16S rRNA sequences for download or use.

Classifier Assign 16S rRNA sequences to our taxonomical hierarchy.

Library Compare Compare two sequence libraries using the RDP Classifier.

Sequence Match Upload your sequence and search for its nearest neighbors.

Release 9.32
October 07, 2005
184,990 16S rRNAs

News [RSS](#)

10/10/2005
Important Annotation Fixes to Update 32. [\[more\]](#)

10/07/2005
RDP 9, Update 32 Released. [\[more\]](#)

08/23/2005
SIMO uses RDP's Sequence Match in their new RDPQuery application [\[more\]](#)

08/11/2005
The RDP webservice has been upgraded. Please [contact](#) us if you experience any problems. [\[more\]](#)

03/14/2005
Survey concerning

Choose "Sequence Match"

Paste in sequence

Click "Submit"

Click "Show Printer Friendly Results"

Did results from NCBI and Ribosomal Database at the Genus species level come out the same?

If there was a discrepancy between NCBI and the Ribosomal Database project which would you suspect is generally more accurate? Why?