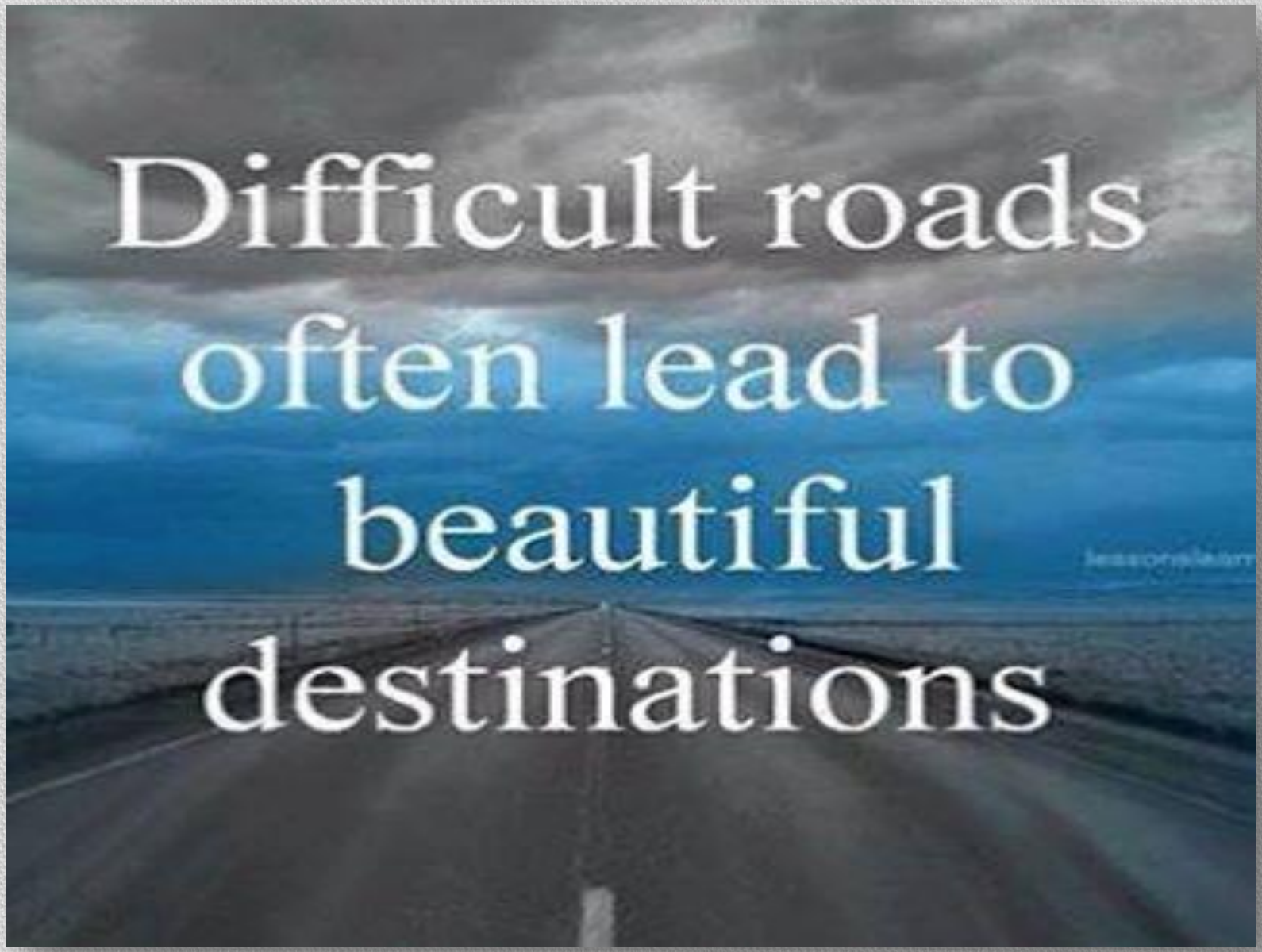

Difficult roads
often lead to
beautiful
destinations

Amira A. AL-Hosary
PhD of infectious diseases
Department of Animal Medicine
(Infectious Diseases)
Faculty of Veterinary Medicine
Assiut University
Egypt

INTERPRETATION OF SEQUENCE RESULTS

An overview on DNA sequencing:

DNA sequencing involves the determination of the sequence of nucleotides in a sample of DNA.

It use a modified PCR reaction where both normal and labeled dideoxy-nucleotides are included in the reaction mix.

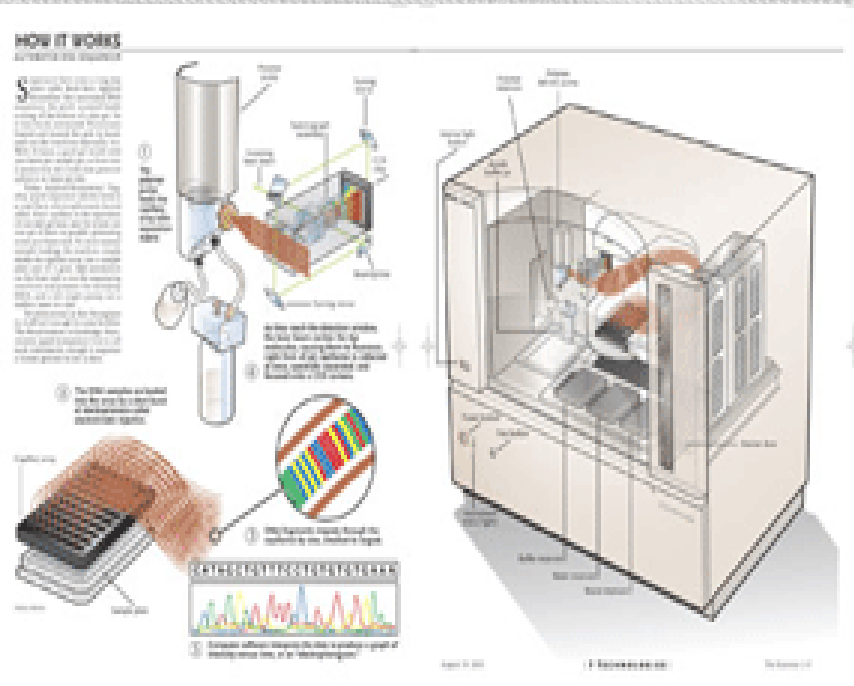
Each dideoxy-nucleotides were labeled with **fluorescent dyes** (**Each nucleotide has a different color**).

At the end of the sequencing reaction,

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.

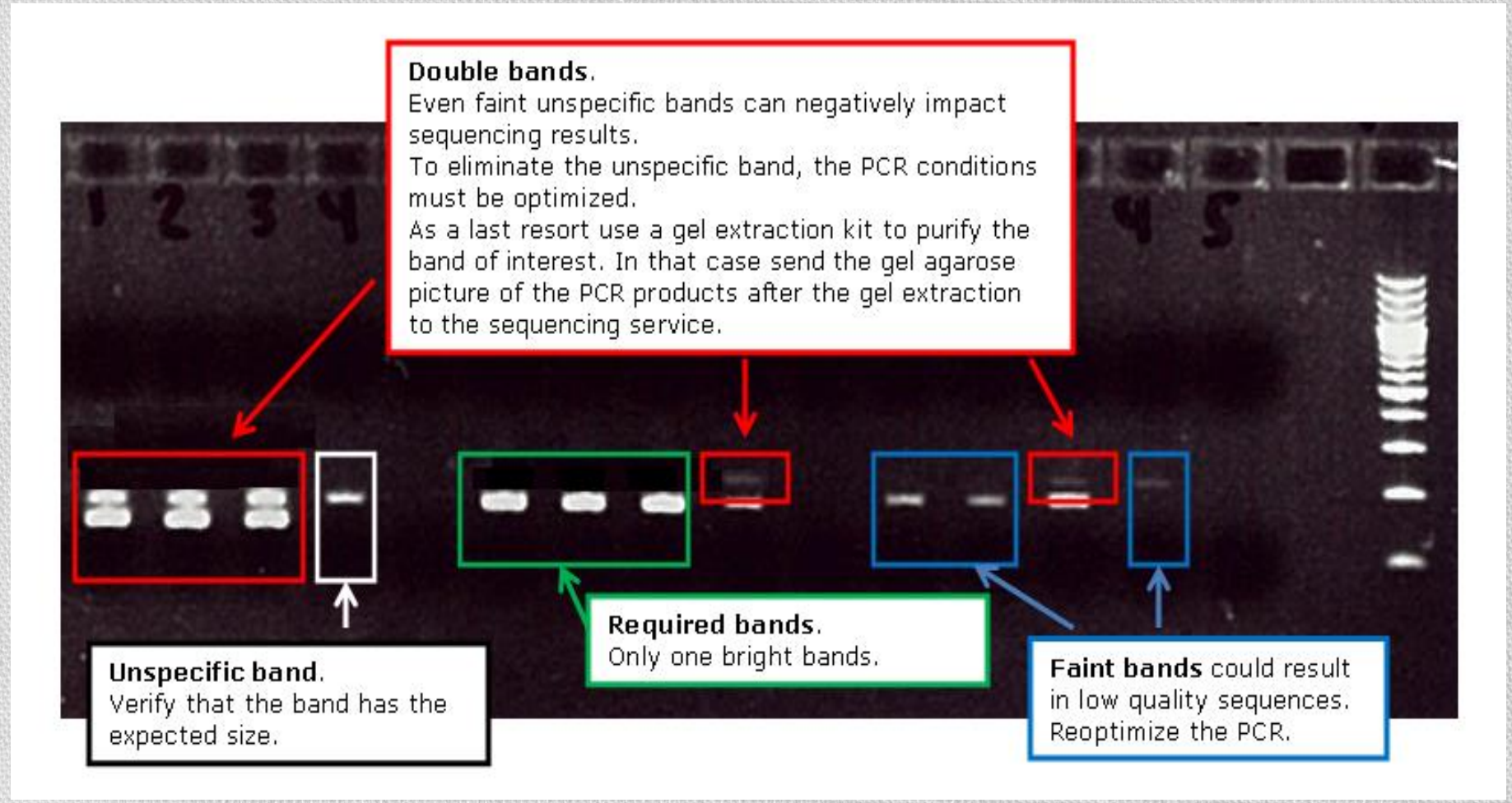
As each band moves past a viewer, the laser excites the dye, and the color of fluorescence is read by a photocell and recorded on a computer.

An overview on DNA sequencing:



Sequencer	Ion Torrent PGM	454 GS FLX	HiSeq 2000	SOLiDv4	PacBio	Sanger 3730xl
Manufacturer	Ion Torrent (Life Technologies)	454 Life Sciences (Roche)	Illumina	Applied Biosystems (Life Technologies)	Pacific Biosciences	Applied Biosystems (Life Technologies)
Amplification approach	Emulsion PCR	Emulsion PCR	Bridge amplification	Emulsion PCR	Single-molecule; no amplification	PCR
Data output per run	100-200 Mb	0.7 Gb	600 Gb	120 Gb	100-700 Mb	1.9~84 Kb
Accuracy	**99%**	**99.9%**	**99.9%**	**99.94%**	**88.0% (>99.9% CCS)**	**99.999%**
Time per run	2 hours	24 hours	3–10 days	7–14 days	2-3 hours	20 minutes - 3 hours
Read length	**200-400 bp**	**700 bp**	**100x100 bp paired end**	**50x50 bp paired end**	**5,500-10,000 bp**	**400-900 bp**
Cost per run	$350 USD	$7,000 USD	$6,000 USD (30x human genome)	$4,000 USD	$125-300 USD	$4 USD (single read/reaction)
Cost per Mb	$1.00 USD	$10 USD	$0.07 USD	$0.13 USD	$0.20 - $3.00 USD	$2400 USD
Cost per instrument	**$80,000 USD**	**$500,000 USD**	**$690,000 USD**	**$495,000 USD**	**$695,000 USD**	**$95,000 USD**

1- The Band:

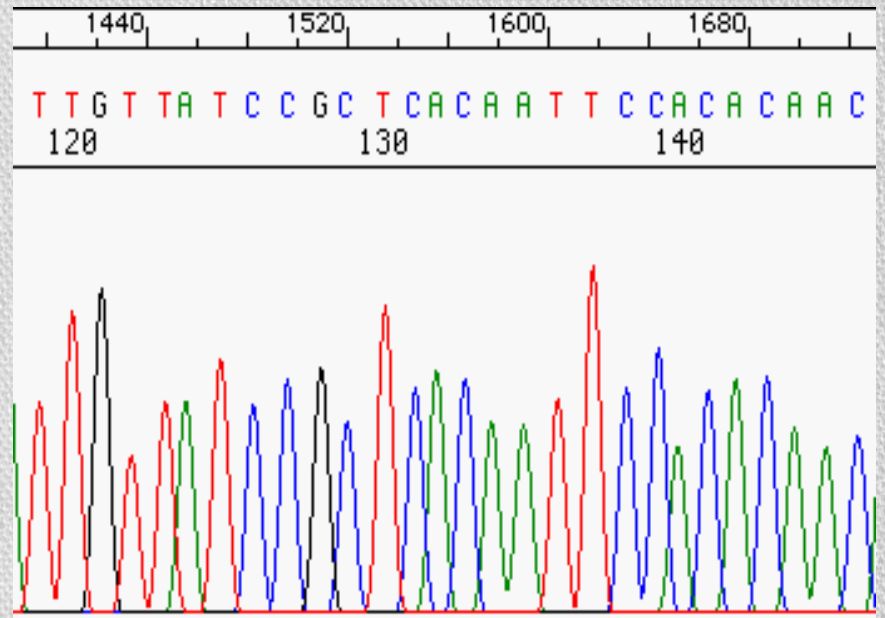

Double bands.
Even faint unspecific bands can negatively impact sequencing results.
To eliminate the unspecific band, the PCR conditions must be optimized.
As a last resort use a gel extraction kit to purify the band of interest. In that case send the gel agarose picture of the PCR products after the gel extraction to the sequencing service.
1 2 3 4
4 5
Unspecific band.
Verify that the band has the expected size.
Required bands.
Only one bright bands.
Faint bands could result in low quality sequences.
Reoptimize the PCR.

Interpreting Sequencing Results

Automated DNA Sequencers generate

1- A four-color chromatogram showing the results of the sequencing run.

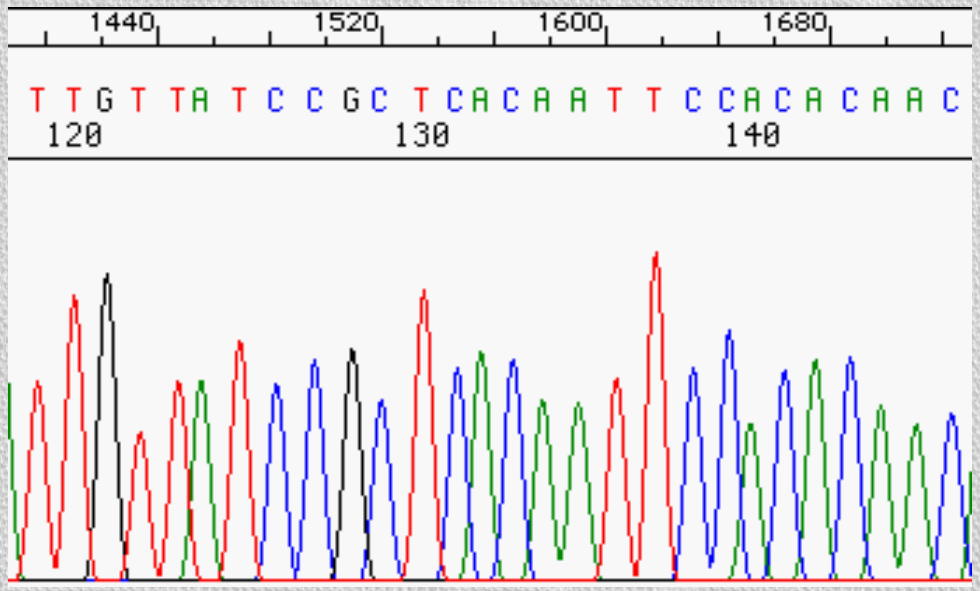
2- In addition to a text file of sequence data.




```
>GXP_210035 loc=GXL_175098|sym=FAM149A|taxid=9606|spec=Homo
sapiens|chr=4|ctg=NC_000004|str=(+)|start=187065495|end=187066181|len=687|comm=Promoter
Region
GGACGGGCGTGGAAGGGTCCACGTCTTTAGTATGCATGCTTAGATCTAGCGTTCCTGTTGATGGAGTAATGGTTCTCGCA
TTGACCAGATCCGGGGCTTCATTTTTTAAACCTCATTCGTCCACTCCCCACCCCAGCCTGGTGTGCGCACCCTTTGATGG
GGCGGGGATAGGCGAGATGGTCCTGTGGTTCTCTGCCTTCTTCTGGTGAATTAAAATCCGATTTGGAAGAGAGAAGGGCA
GCCAGCACCAGTATGCACAGCCCCCGGCCCCAGAGACCCGGGAAGGAGTAGGGAGGCCGGGCCGTGCGCGGAGGAGTGGC
CGCTGGGTTGGAAACCCGGCCCGGCAGGGAGCGGGGAAGGCGCGCTTTCCCGGAGGTCGGCGCGGGGCCGGGGCCGGGGC
CGGGGCCCGGAGCGGGGATGGGCGGGCGCAGCCGGGATTAGCTGGCGGGCGAGGGCGCAGCGCAGGGAGGAGGAGGGGAG
GCGGCGCCGGCGCGGGCGGGGCGGAGGATCTGGAGAGGGAAGGGGCGTGCGAGCCCCGCGGACCCCGGGCGCGCCCGGGC
CGCCTGAGCTGGGCCAGCCGCGCGGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGGGTGGGGAGCCCCAGCCCC
GGGGCCGCGGGGGCGCGTGACCGGCTGTCTGCGTGGGGCCCGCGCGC
```

Interpreting Sequencing Results

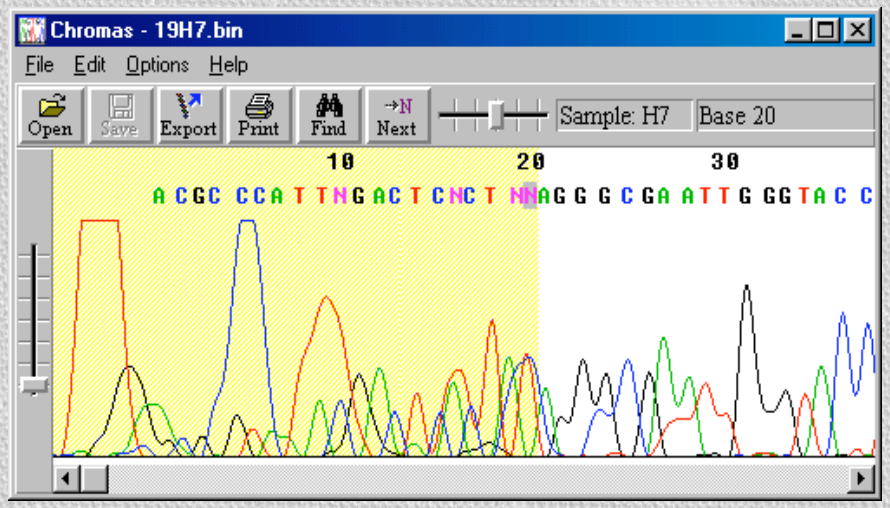
- When you obtain a sequence you should proofread it to ensure that all ambiguous sites are correctly called and determine the overall quality of your data.



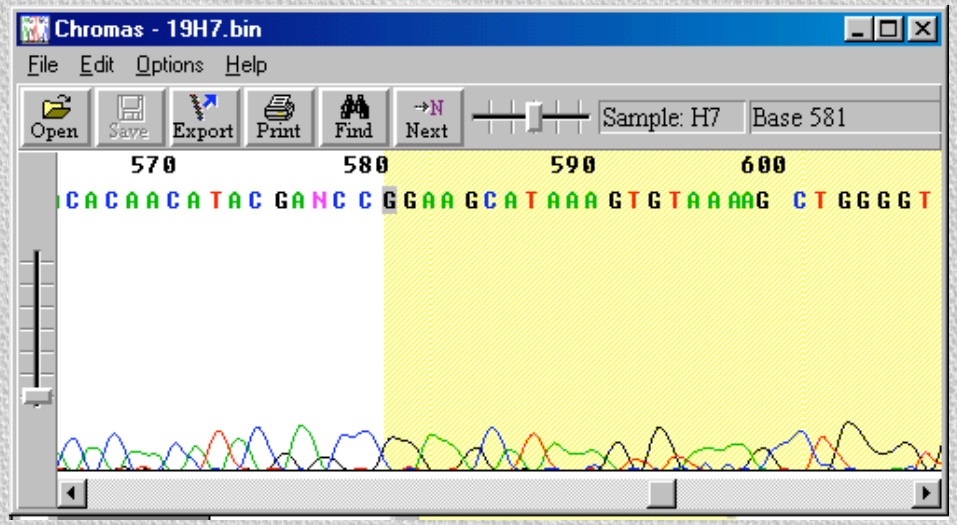
- Base Designations
- *"A" designation*—green peaks
- *"G" designation*—black peaks
- *"T" designation*—red peaks
- *"C" designation*—blue peaks
- *"N" designation*—peaks that, for whatever reason, are not clear enough to designate as A, G, T, or C.

Interpreting Sequencing Chromatograms

Good sequence generally begins roughly around base 20.



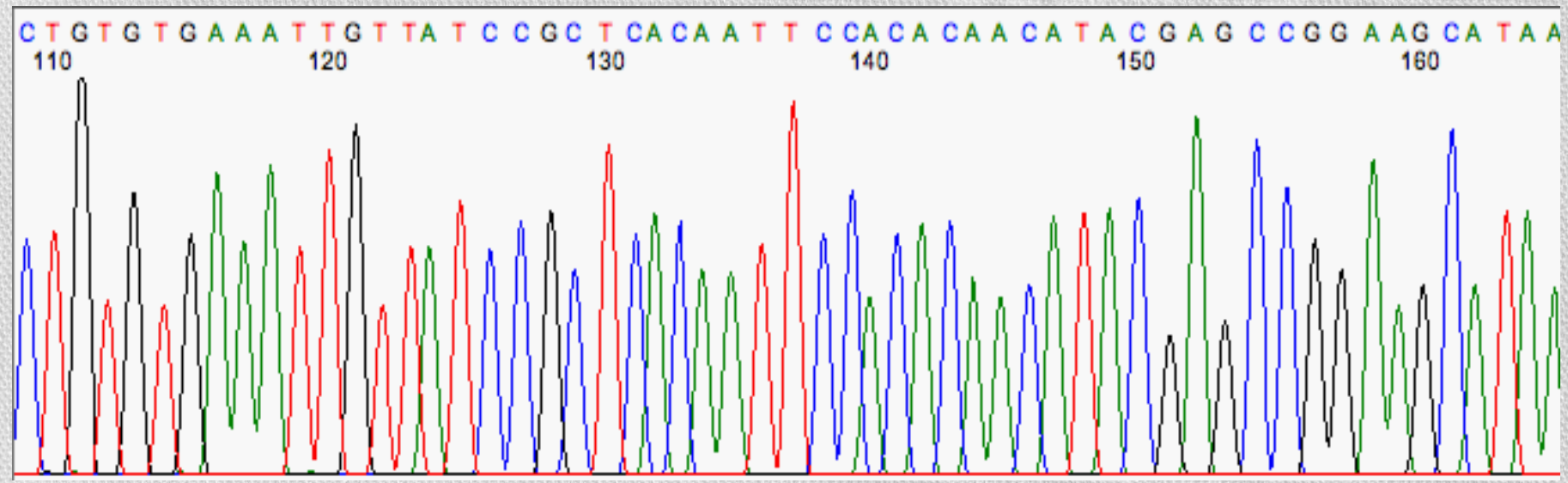
Beginning of Sequence



End of sequence

Interpreting Sequencing Chromatograms

With a little practice, you can scan a chromatogram in less than a minute and spot problems. It is not necessary to read each and every base.

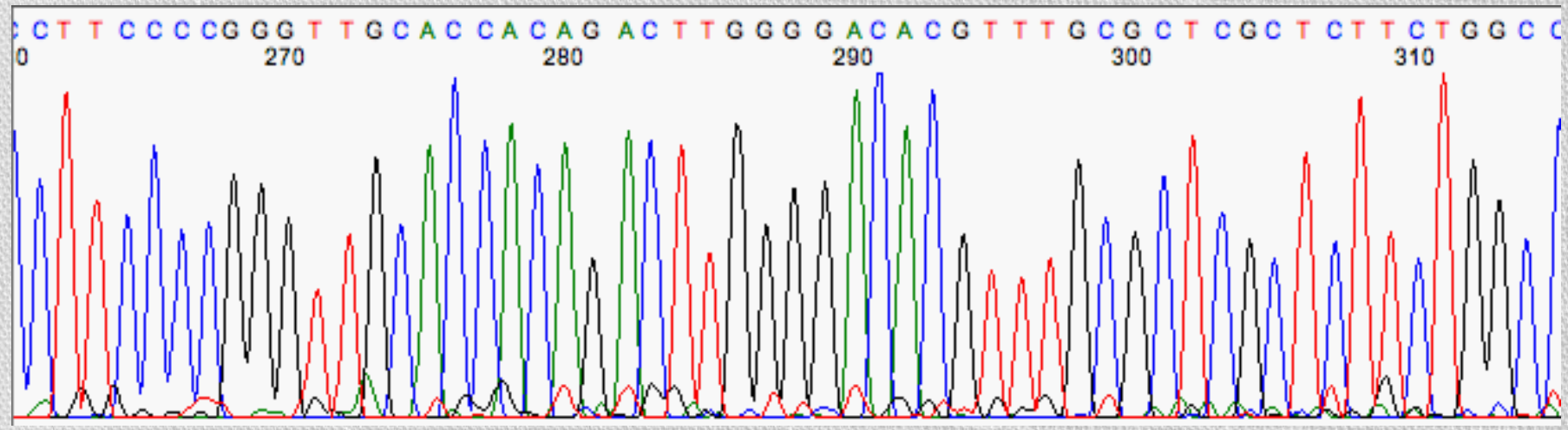


An example of excellent sequence. Note the evenly-spaced peaks and the lack of baseline 'noise'

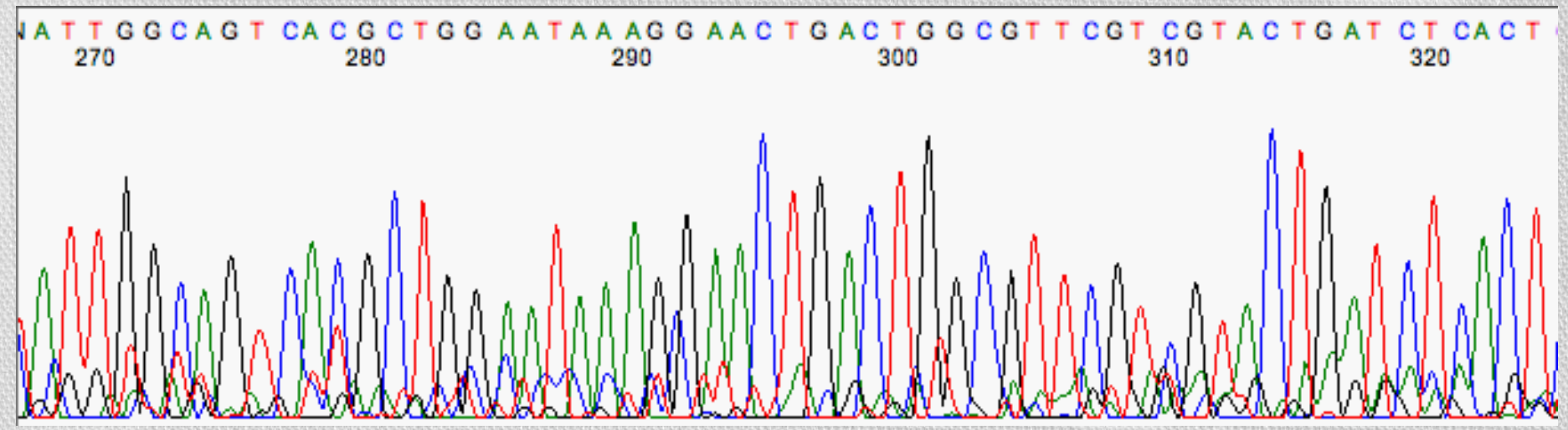
Interpreting Sequencing Chromatograms

Background noise

This example has a little baseline noise, but the 'real' peaks are still easy to call, so there's no problem with this sample



Interpreting Sequencing Chromatograms



Noise like the above most commonly arises when the sample itself is too dim.

Types of Polymorphisms

1- Transitions: A ←→ G or C ←→ T

(purines to purines OR pyrimidines to pyrimidines)

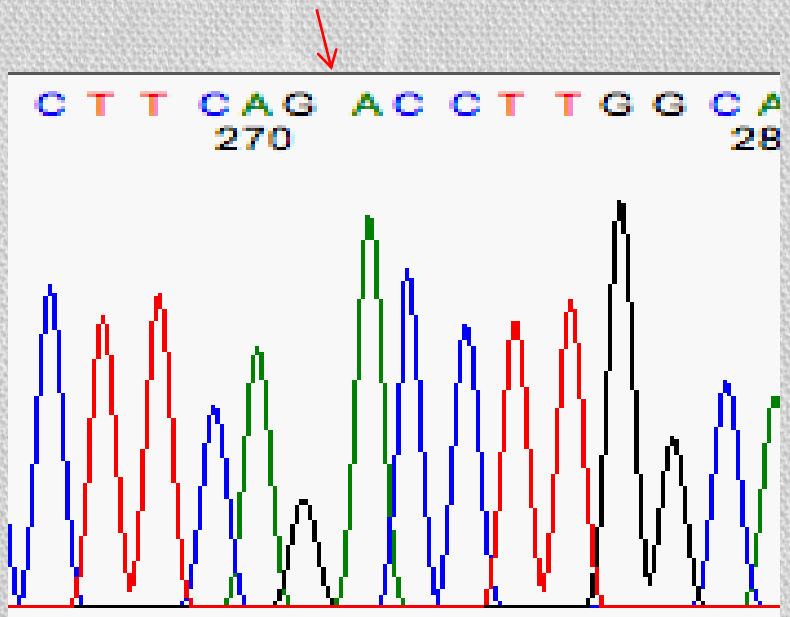
2-Insertions: an extra base is present when compared to the Anderson reference sequence.

3- Deletions: a base is missing when compared to the Anderson reference sequence.

4- Mis-Called

(a) Irregular spacing:

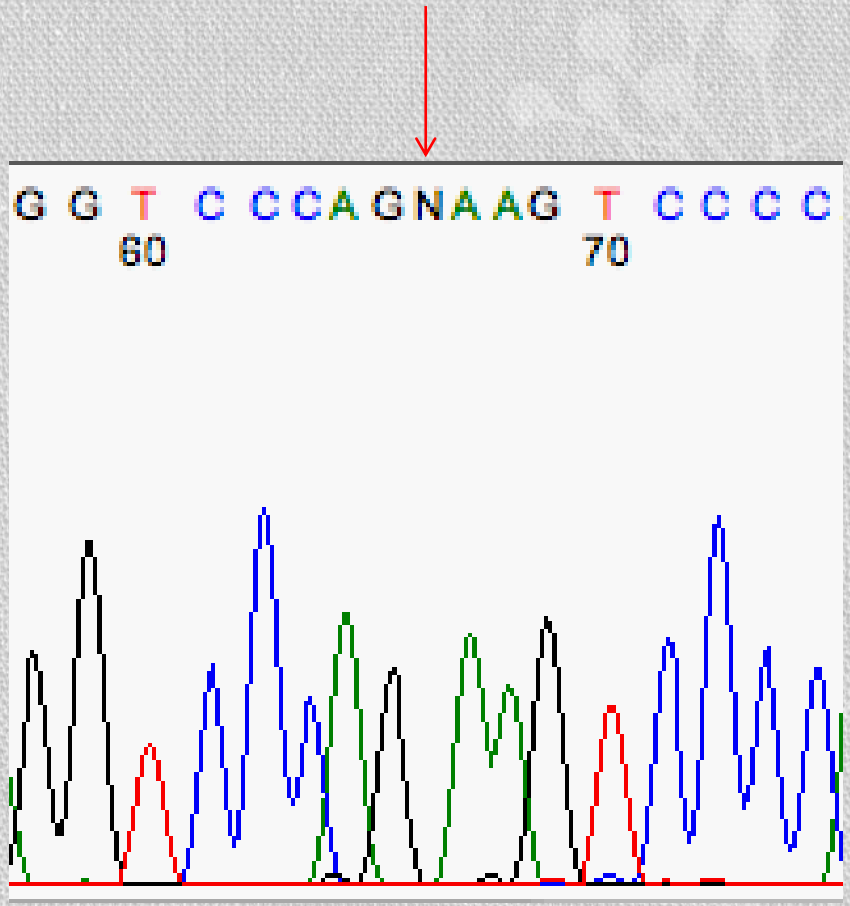
Common one for us is a G-A dinucleotide, which leaves a little extra space between them.



4- Mis-Called

(b) Mis-call a nucleotide:

Sometimes the computer will mis-call a nucleotide when a human could do better. Most often, this occurs when the basecaller calls a specific nucleotide, when the peak really was ambiguous and should have been **called as 'N'.**

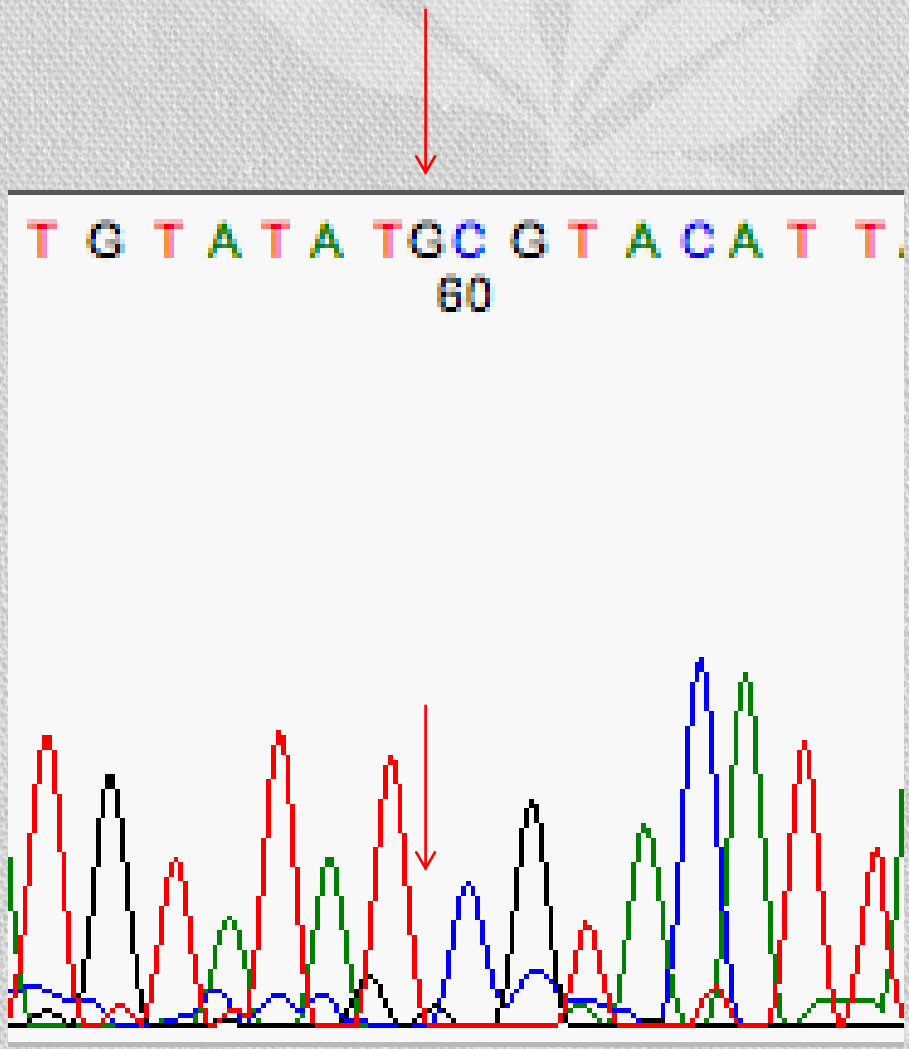


4- Mis-Called

(b) The real problem comes when the base caller attempts to interpret a gap as a real nucleotide.

Note the real T peak (nt 58) and the real C peak (nt 60), with the G barely visible between them. Despite it size, the baseline-noise G peak was picked as if it were real. The clues to spot are (i) the oddly-spaced letters, with the G squeezed in, and (ii) the gap in the 'real' peaks, containing a low noise peak.

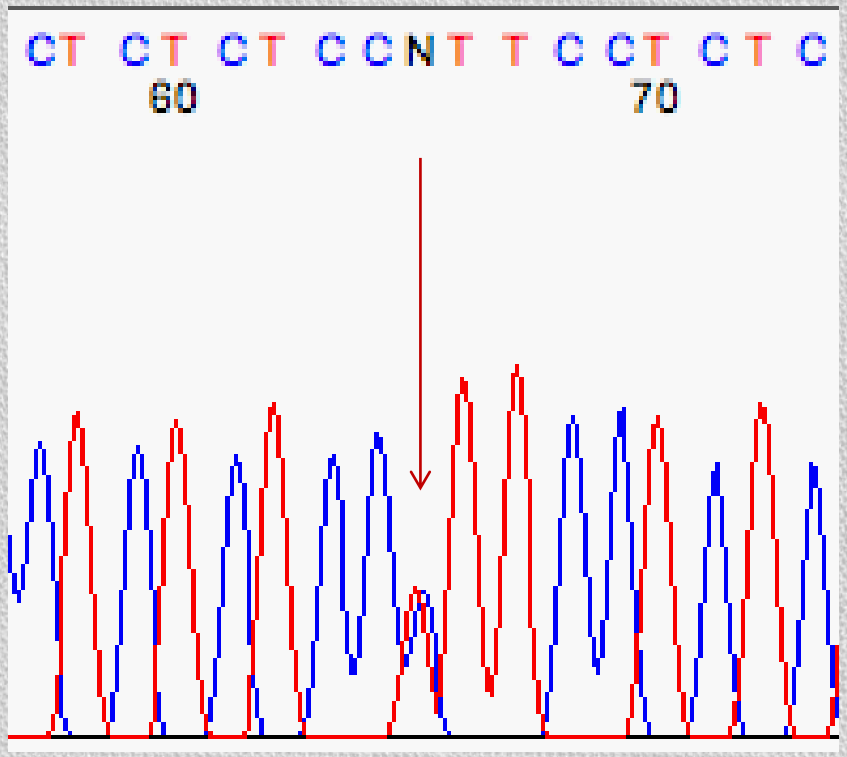
This is a great example of why a weak sample, with its consequent noisy chromatogram, is untrustworthy.



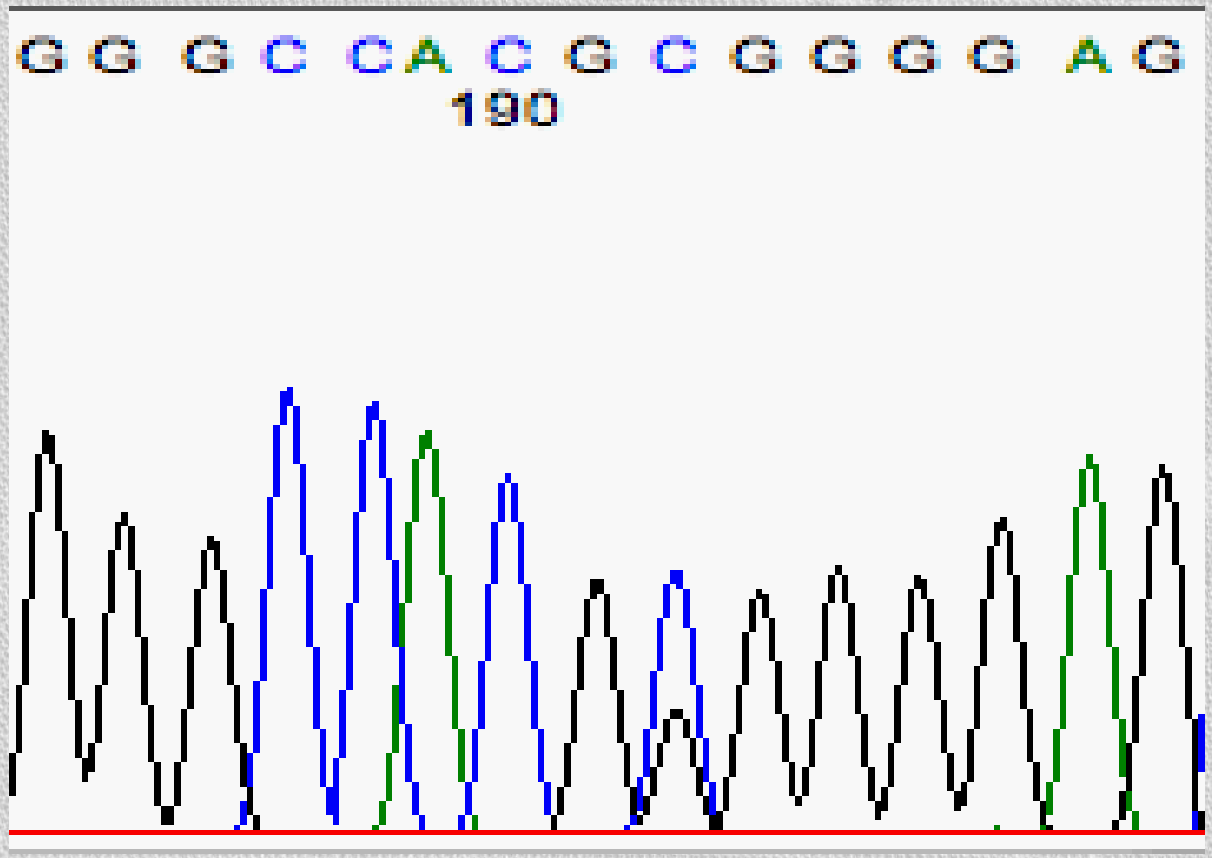
5- Heterozygous (double) peaks:

A single peak position within a trace may have but two peaks of different colors instead of just one. This is common when sequencing a PCR product derived from diploid genomic DNA, where polymorphic positions will show both nucleotides simultaneously. Note that the base caller may list that base position as an 'N', or it may simply call the larger of the two peaks.

Here's a great example of a PCR amplicon from genomic DNA, with a clear heterozygous **single-nucleotide polymorphism (SNP).**

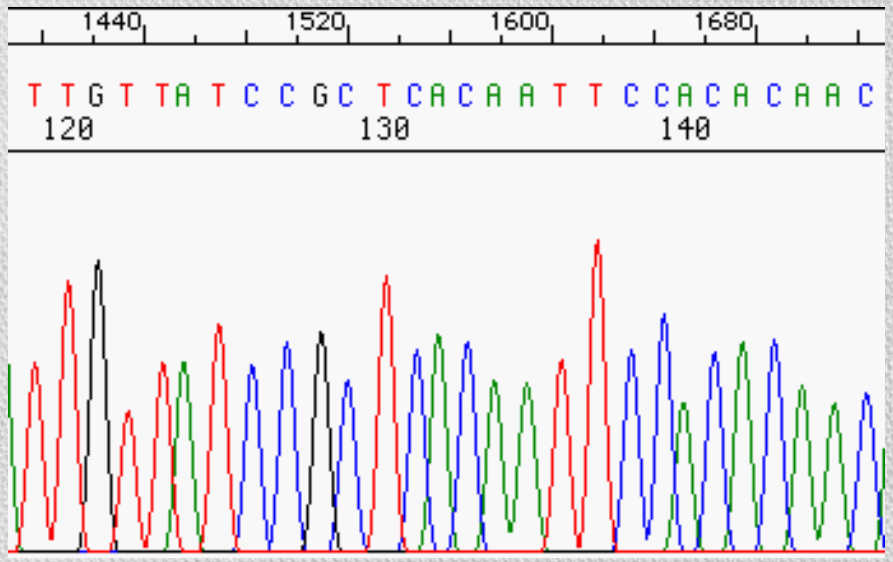


6- *Negative samples / No DNA*—chromatograms displaying peaks from which no useable sequence can be obtained may be due to an absence of DNA. These chromatograms generally have one or two predominant colors.

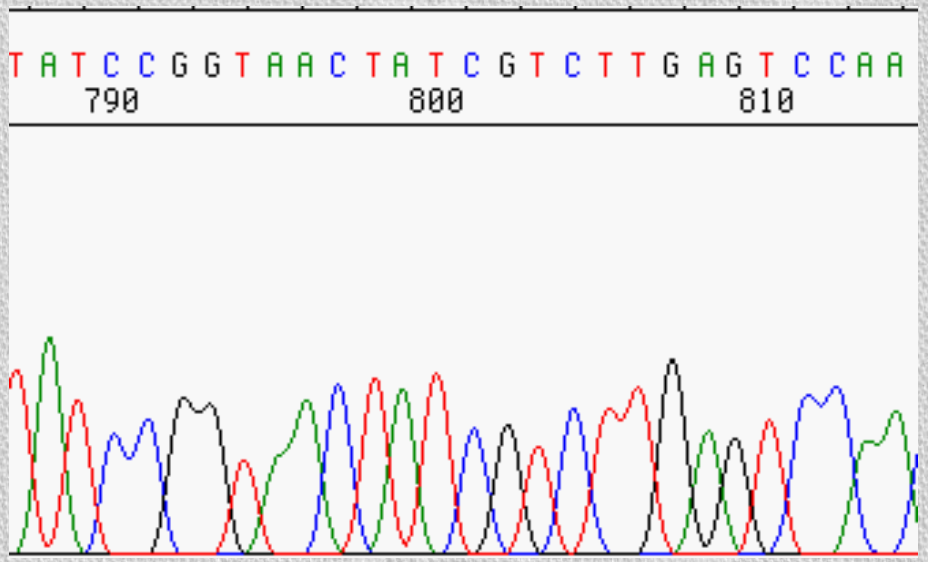


7- Loss of resolution later in the gel:

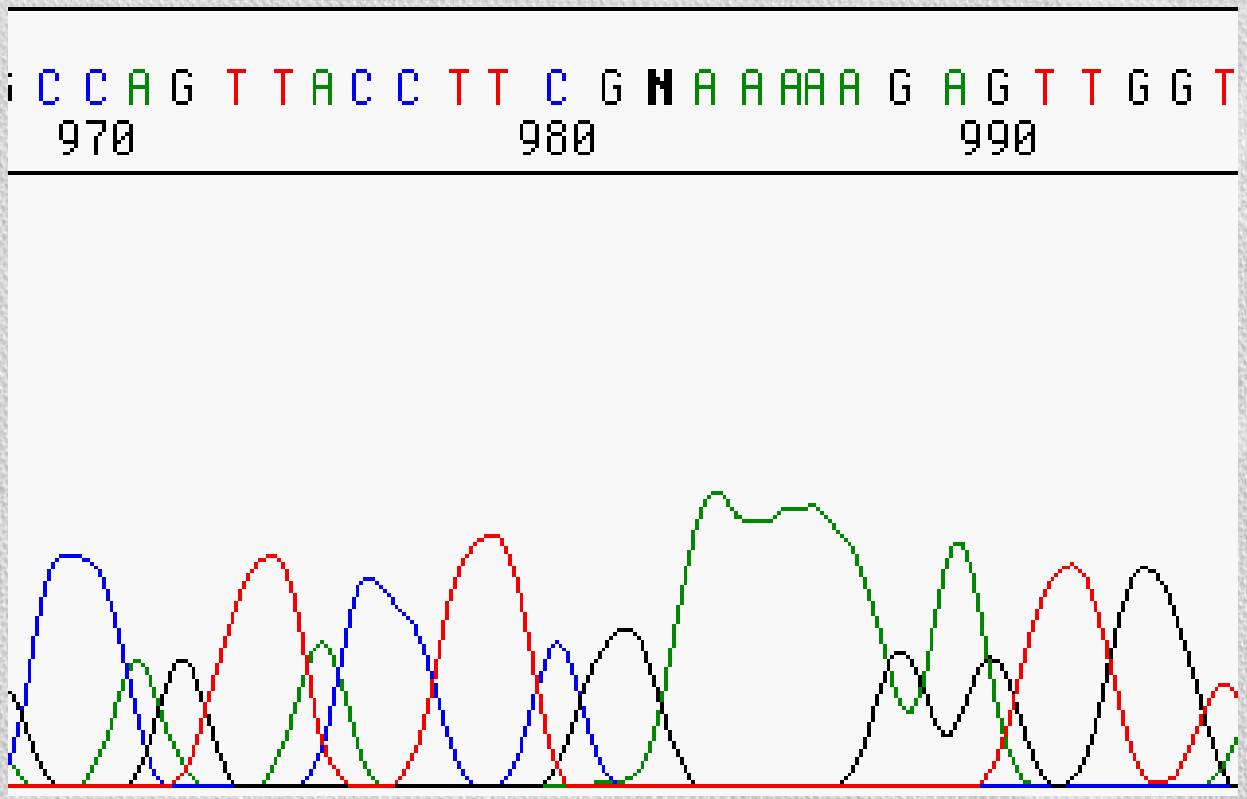
As the gel progresses, it loses resolution. This is normal; peaks broaden and shift, making it harder to make them out and call the bases accurately. The sequencer will continue attempting to "read" this data, but errors become more and more frequent.



This is a typical example of data from a very good sample

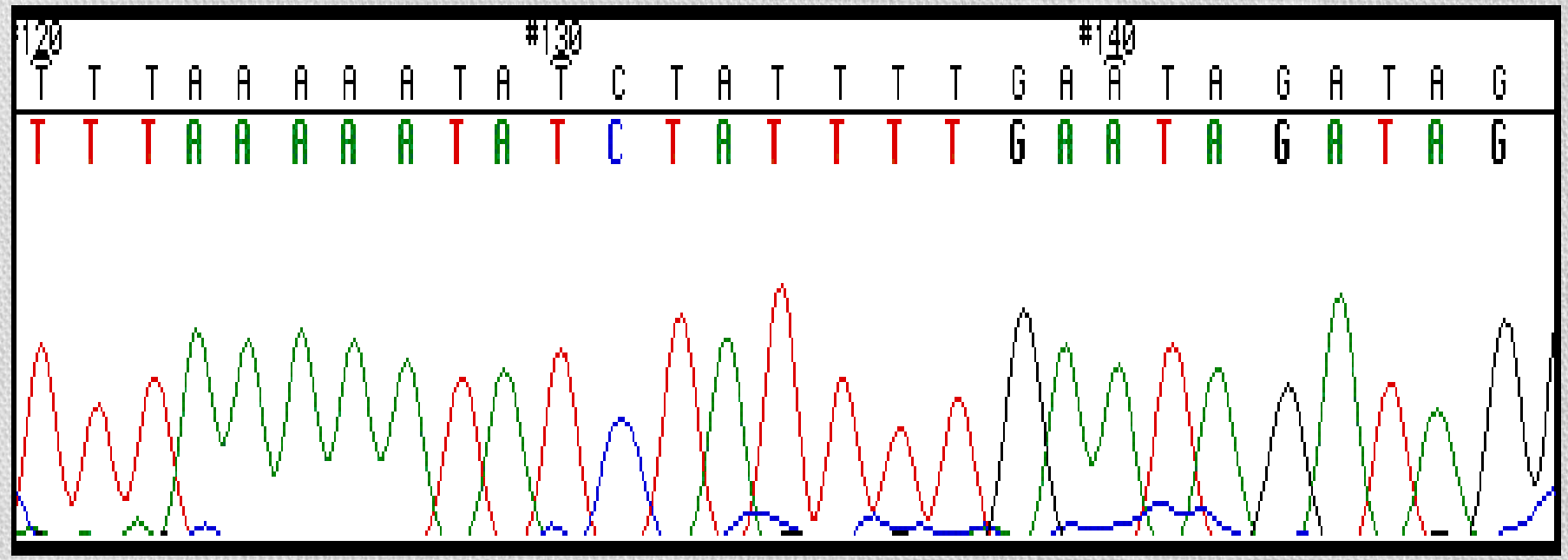


the spacing between the basecall letters at top is regular, which is often a good indication of the reliability of the data.



There are only a few base calls that can be considered reliable. The G at 981 may in fact be two G's, the N could be a G or an A, and who knows how many A's there are afterwards.

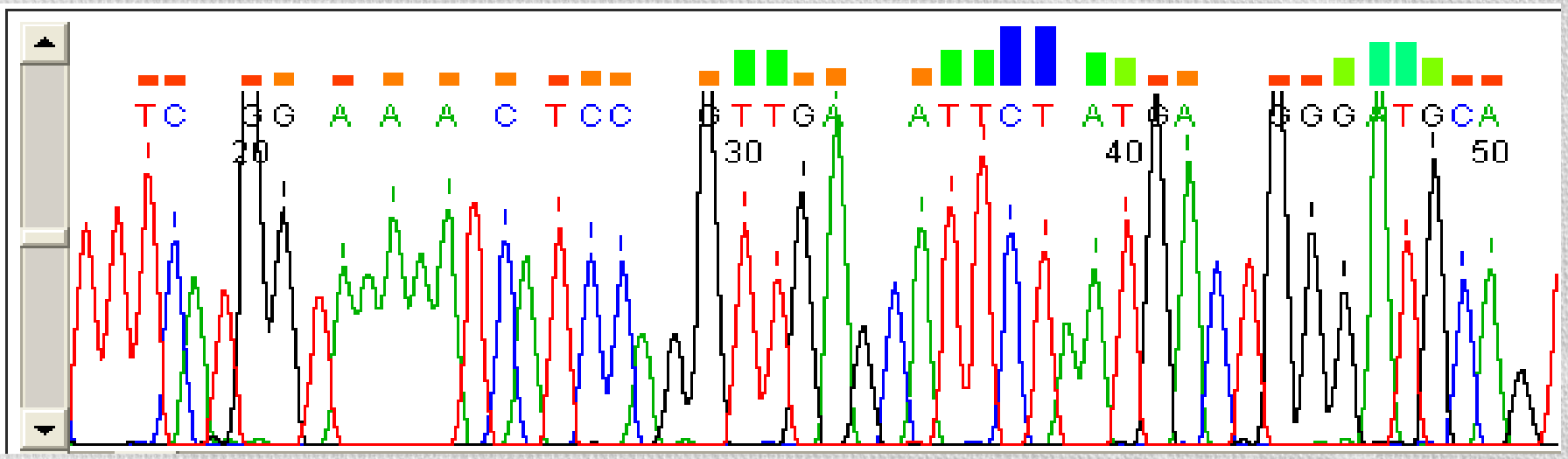
8- Non-discrete peaks—these may occur when several of the same nucleotide appears in a row. For example, if the sequence includes the region TAAAAAT, it may be represented by one wavy peak as opposed to 5 distinct peaks.



9- Good sequence with bad base calling:

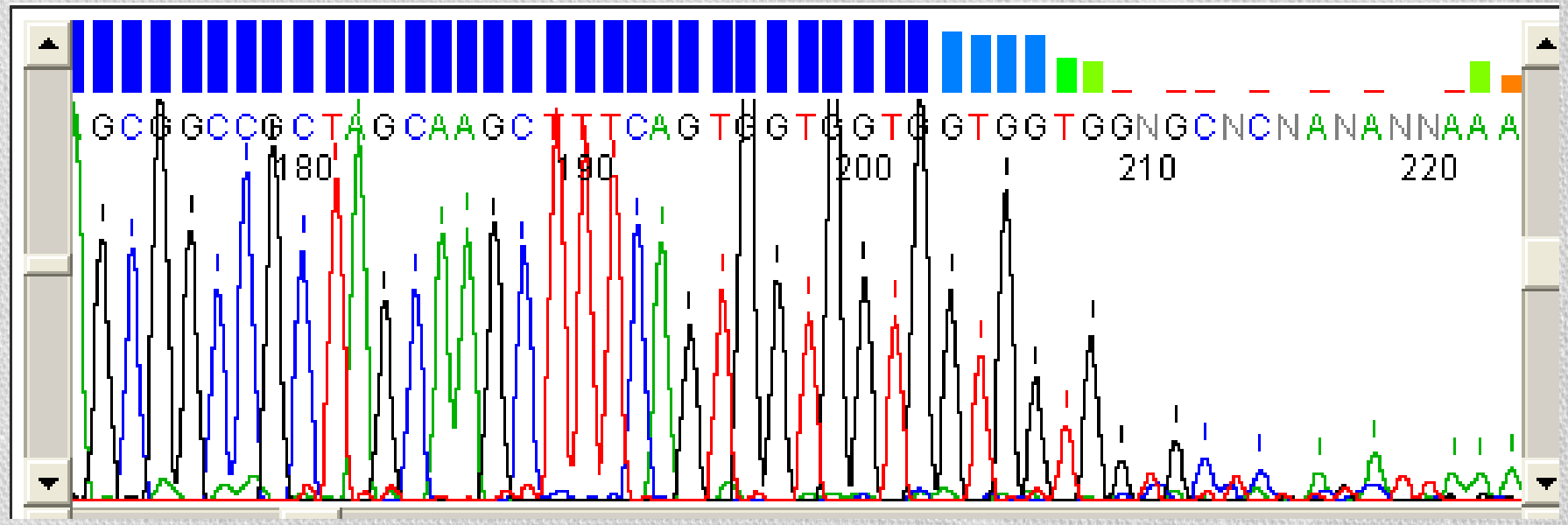
Failed analysis

Ask the Sequencing Service to reanalyze the sequence.



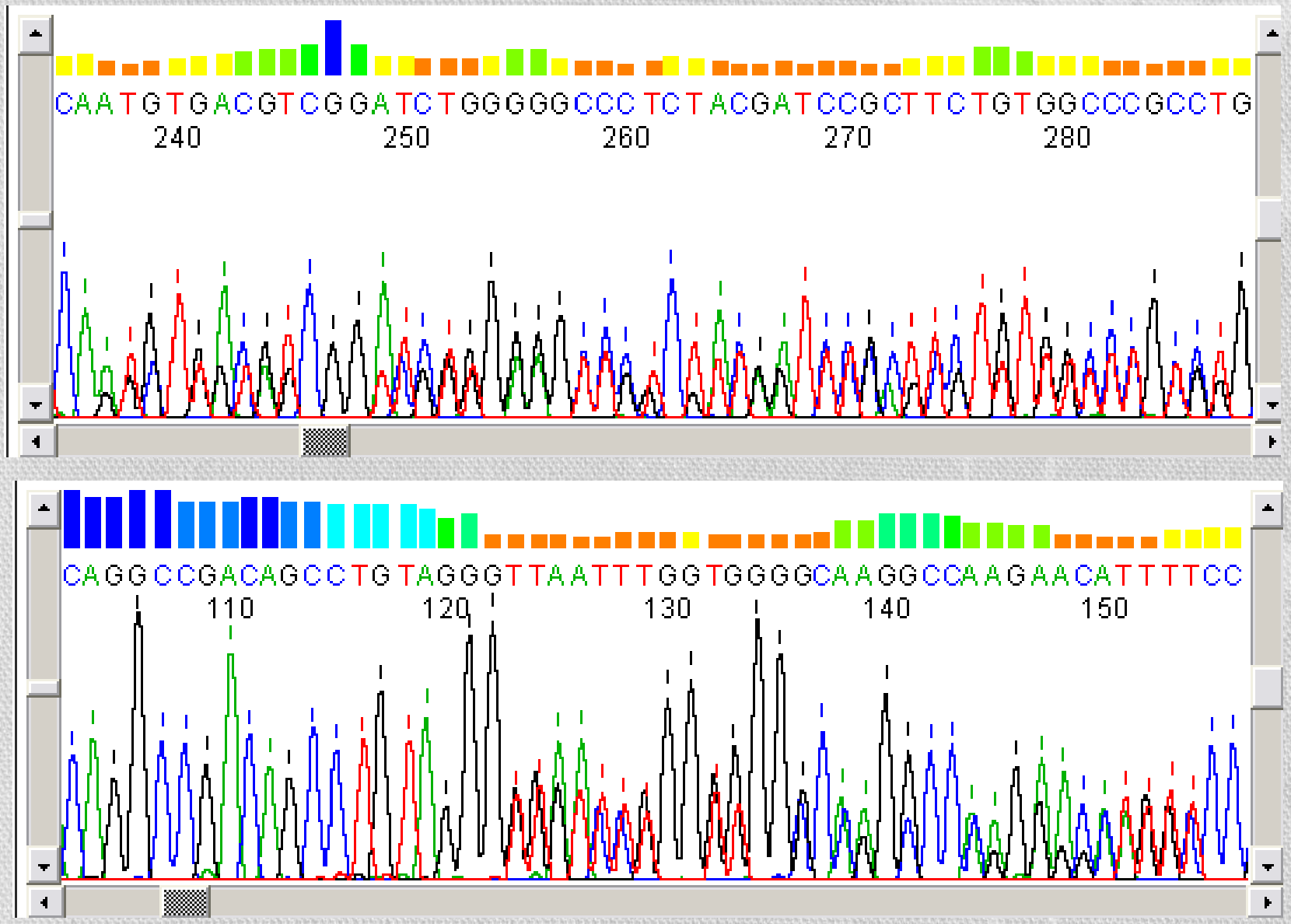
10- DNA template has a secondary structure:

Secondary structures create a distortion that makes it impossible for elongation to continue and so the sequence ends abruptly.



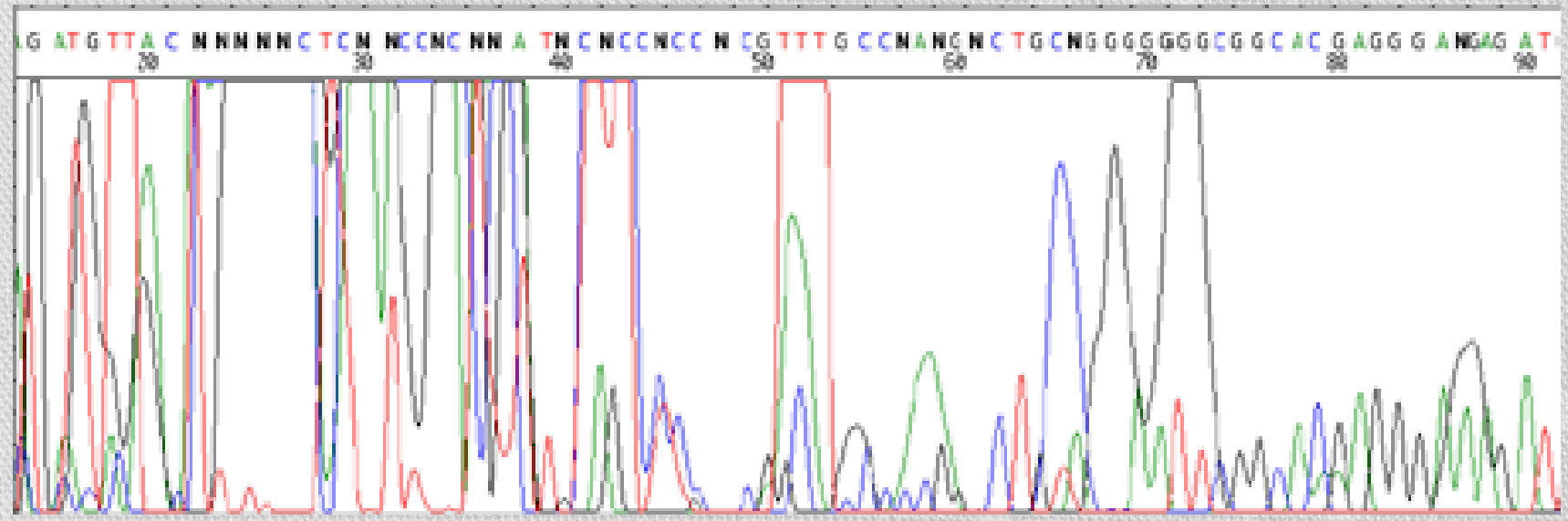
The sequence ends after approximately 200 bp

11- DNA contamination:

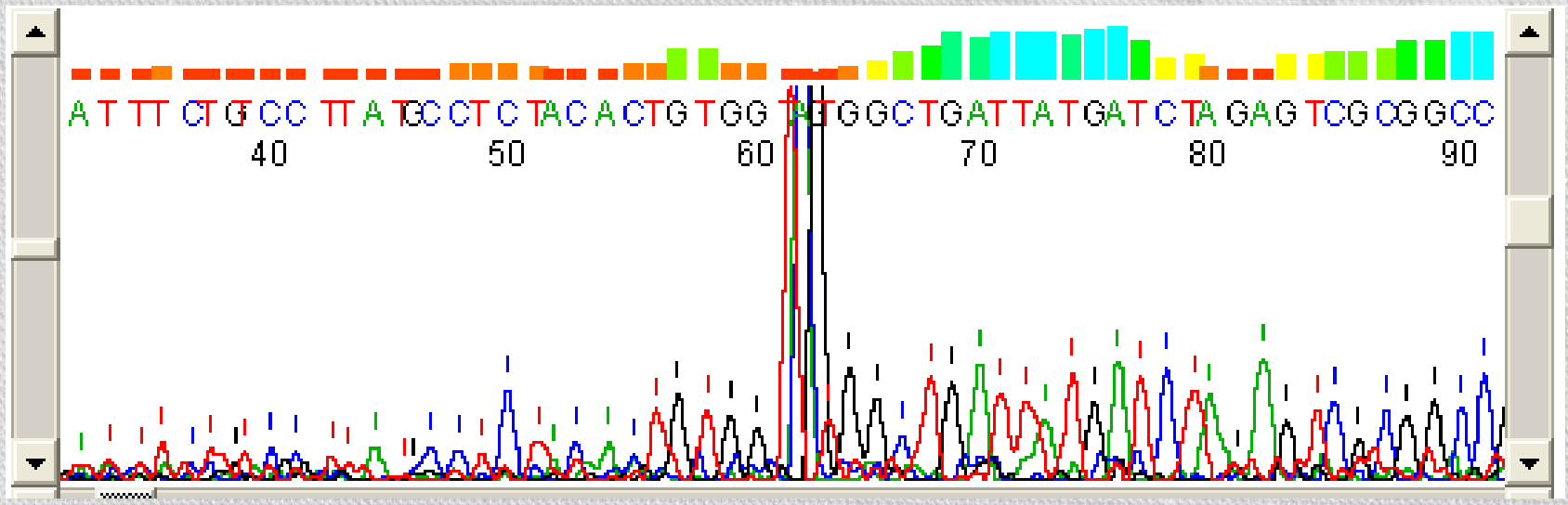


12- Excess dye peaks at the beginning of the sequence

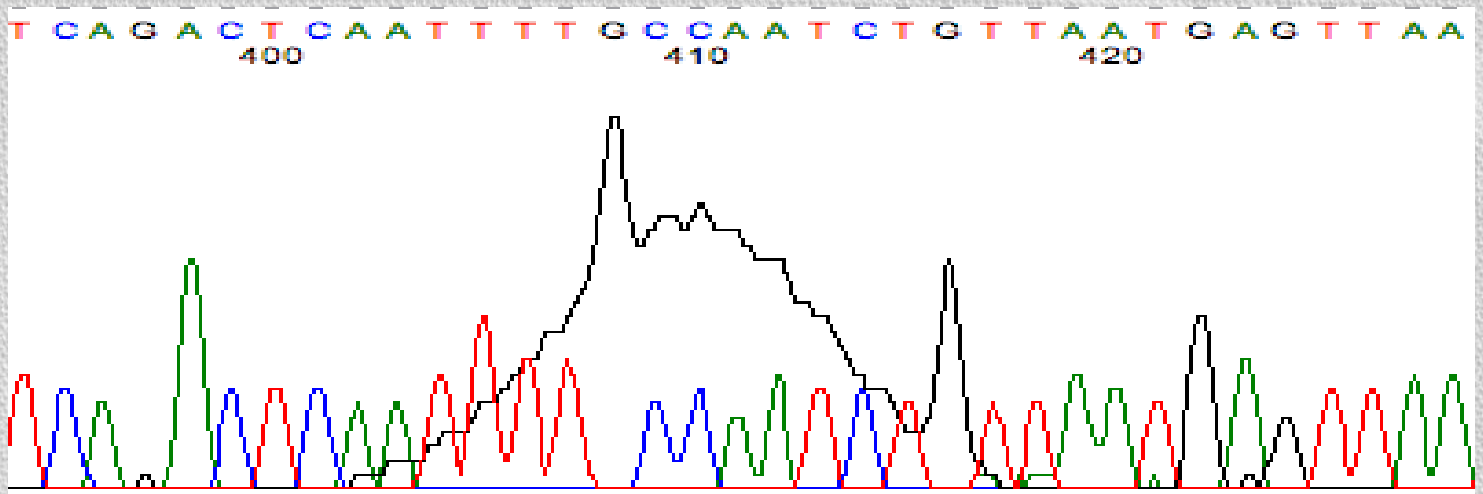
Cause related to sequencing: Poor removal of unincorporated dye terminators during the post-sequencing clean up



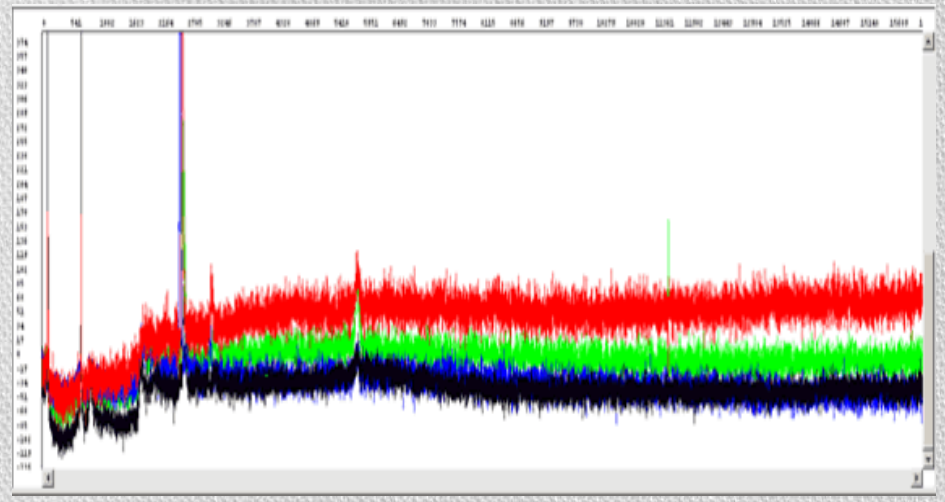
13- Sharp peaks in the sequence:

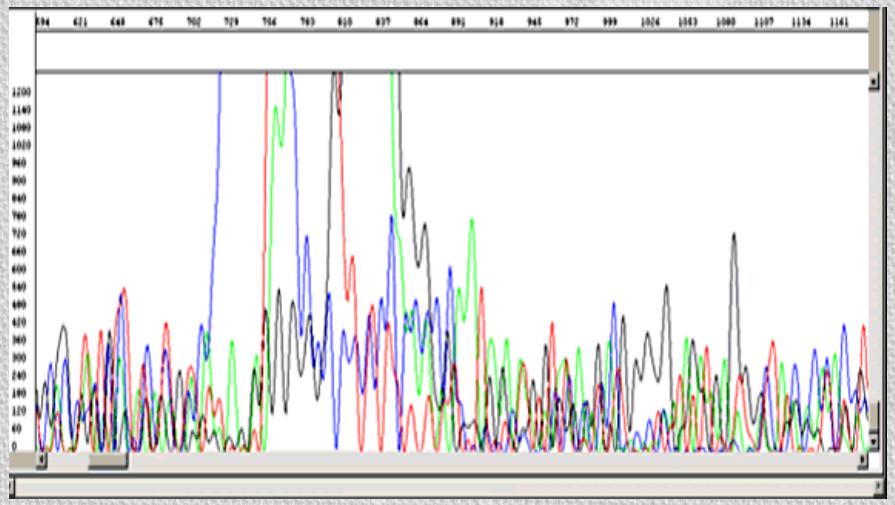


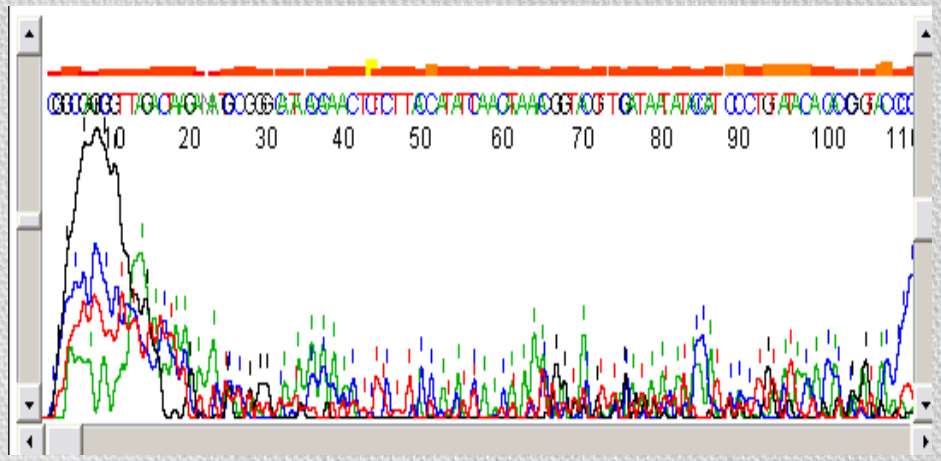
14- Sequence with "G" dye blob

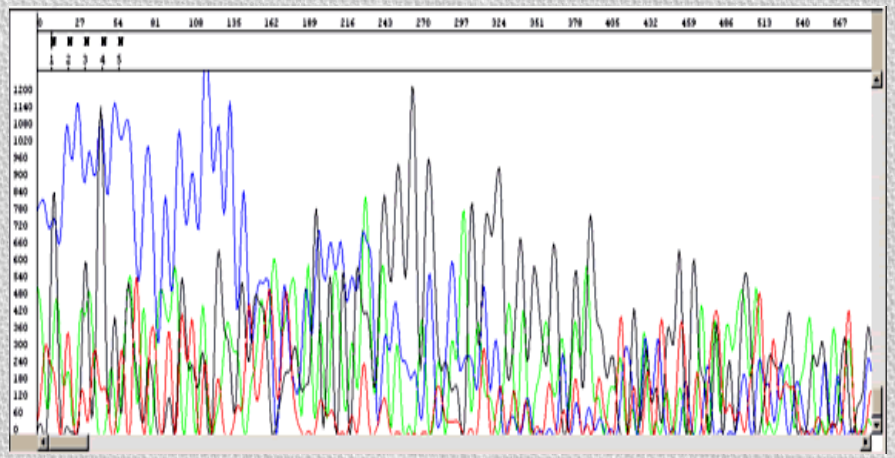


15- Reaction failed, No sequencing data









Realize, too, that it's easy for a human to miss these. If you want to be sure you've detected all of the polymorphic positions, you should be using a computer program to scan your chromatograms!

Interpreting of Sequencing Results

>GXP_210035 loc=GXL_175098|sym=FAM149A|taxid=9606|spec=Homo sapiens|chr=4|ctg=NC_000004|str=(+)|start=187065495|end=187066181|len=687|comm=Promoter Region

GGACGGGCGTGGAAGGGTCCACGTCTTTAGTATGCATGCTTAGATCTAGCGTTCCTGTTGATGGAGTAATGGTTCTCGCA
TTGACCAGATCCGGGGCTTCATTTTTTAAACCTCATTCGTCCACTCCCCACCCCAGCCTGGTGTGCGCACCCTTTGATGG
GGCGGGGATAGGCGAGATGGTCCTGTGGTTCTCTGCCTTCTTCTGGTGAATTAAAATCCGATTTGGAAGAGAGAAGGGCA
GCCAGCACCAGTATGCACAGCCCCCGGCCCCAGAGACCCGGGAAGGAGTAGGGAGGCCGGGCCGTGCGCGGAGGAGTGGC
CGCTGGGTTGGAAACCCGGCCCGGCAGGGAGCGGGGAAGGCGCGCTTTCCCGGAGGTCGGCGCGGGGCCGGGGCCGGGGC
CGGGGCCCGGAGCGGGGATGGGCGGGCGCAGCCGGGATTAGCTGGCGGGCGAGGGCGCAGCGCAGGGAGGAGGAGGGGAG
GCGGCGCCGGCGCGGGCGGGGCGGAGGATCTGGAGAGGGAAGGGGCGTGCGAGCCCCGCGGACCCCGGGCGCGCCCGGGC
CGCCTGAGCTGGGCCAGCCGCGCGGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGCGGGCGGGTGGGGAGCCCCAGCCCC
GGGGCCGCGGGGGCGCGTGACCGGCTGTCTGCGTGGGGCCCGCGCGC

Interpreting of Sequencing Results

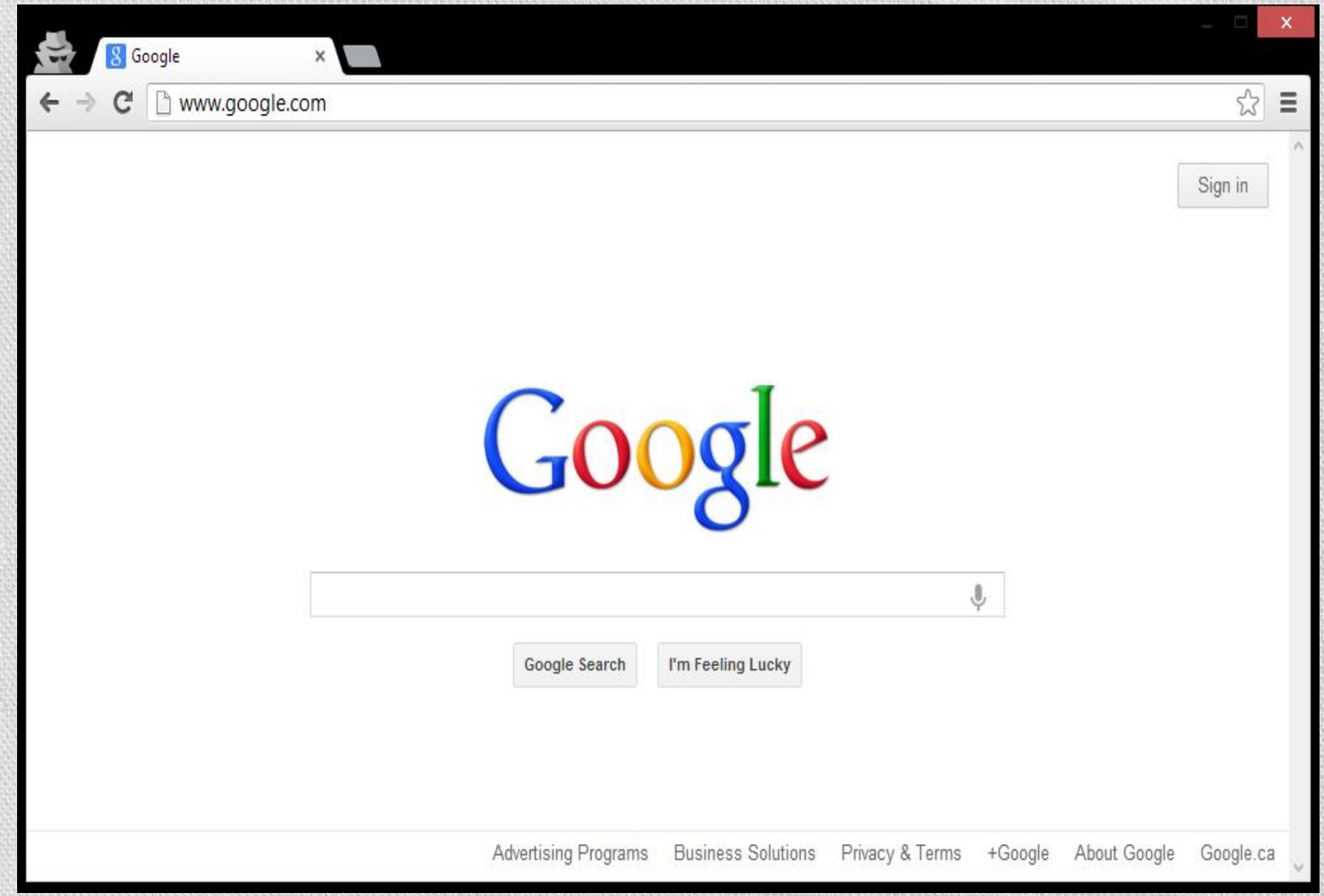
Determining homology:

In other words, is your sequence similar to any other published sequences and if so, to what degree?

This can be accomplished using **BLAST**, (**B**asic **L**ocal **A**lignment **S**earch **T**ool): This program supported by the National Center for Biotechnology Information (NCBI).

The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

This program is accessible at: http://www.ncbi.nlm.nih.gov/BLAST/ (GenBank database; National Center for Biotechnology Information, National Institutes of health).


Google
www.google.com
Sign in
Google
Google Search
I'm Feeling Lucky
Advertising Programs
Business Solutions
Privacy & Terms
+Google
About Google
Google.ca

BLAST: **Basic Local Alignment Search Tool**

blast.ncbi.nlm.nih.gov/

The **Basic Local Alignment Search Tool** (**BLAST**) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to **...**

Align two or more - Protein BLAST: ***search ... - Nucleotide BLAST Rat - sequences

Nucleotide **BLAST**: Search nucleotide databases using a nucleotide **...**

blast.ncbi.nlm.nih.gov/Blast.cgi?...blastn...BlastSearch...

No BLAST database contains all the sequences at NCBI. BLAST databases **...**

BLAST - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/BLAST

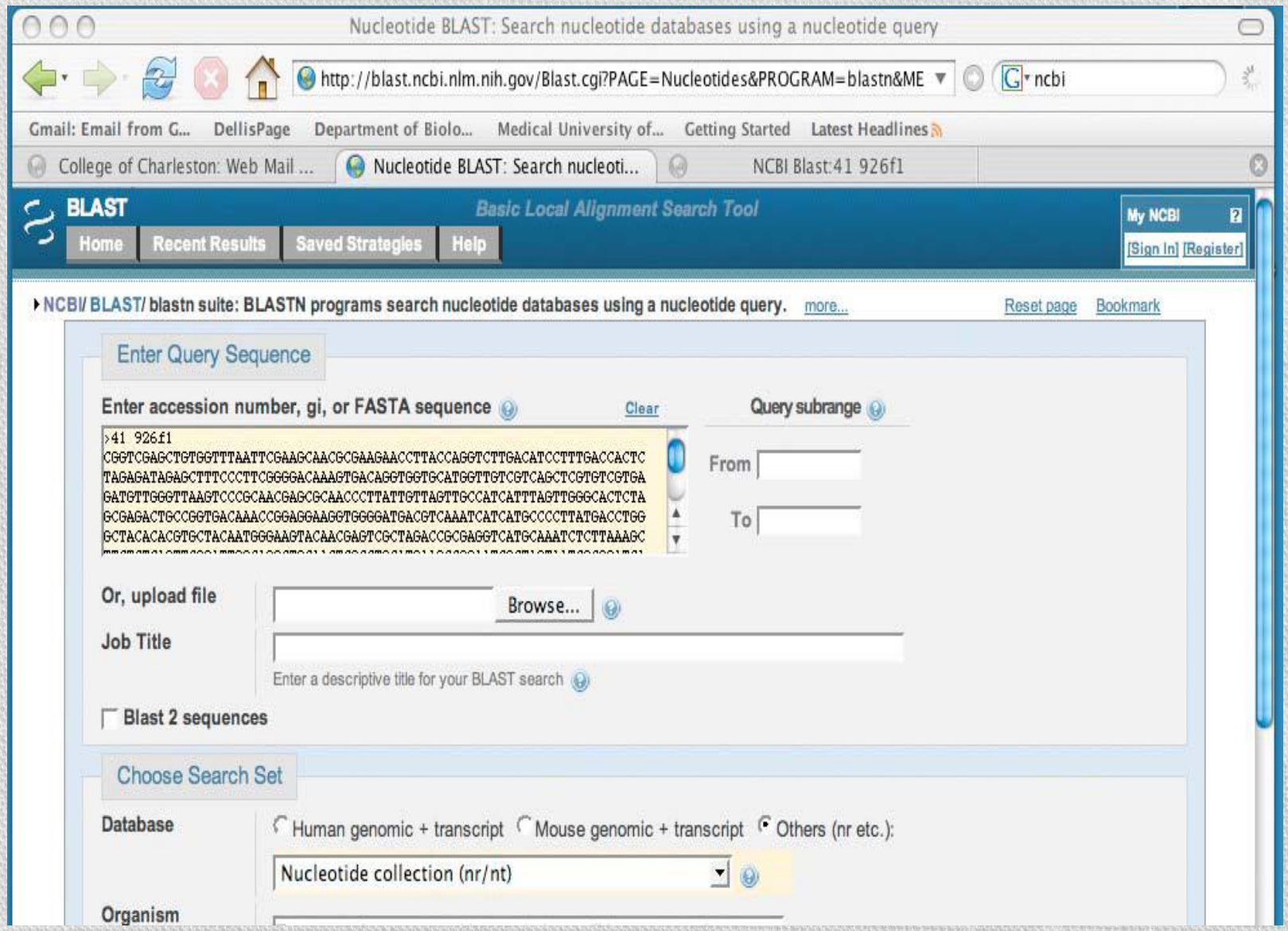
In bioinformatics, **Basic Local Alignment Search Tool**, or **BLAST**, is an algorithm for comparing primary biological sequence information, such as the amino-acid **...**

Process - Output - Input - Background

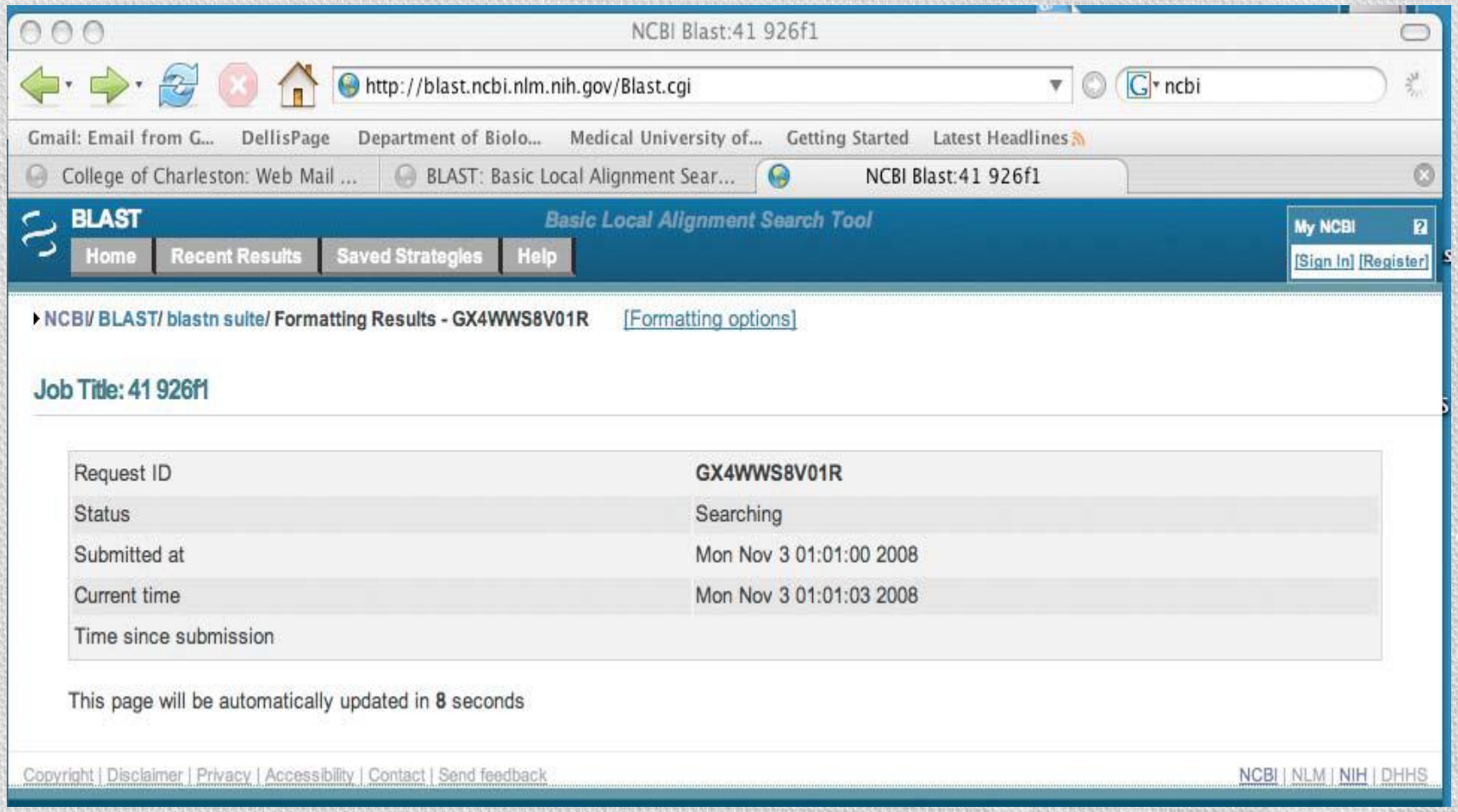
Basic BLAST

Choose a BLAST program to run.

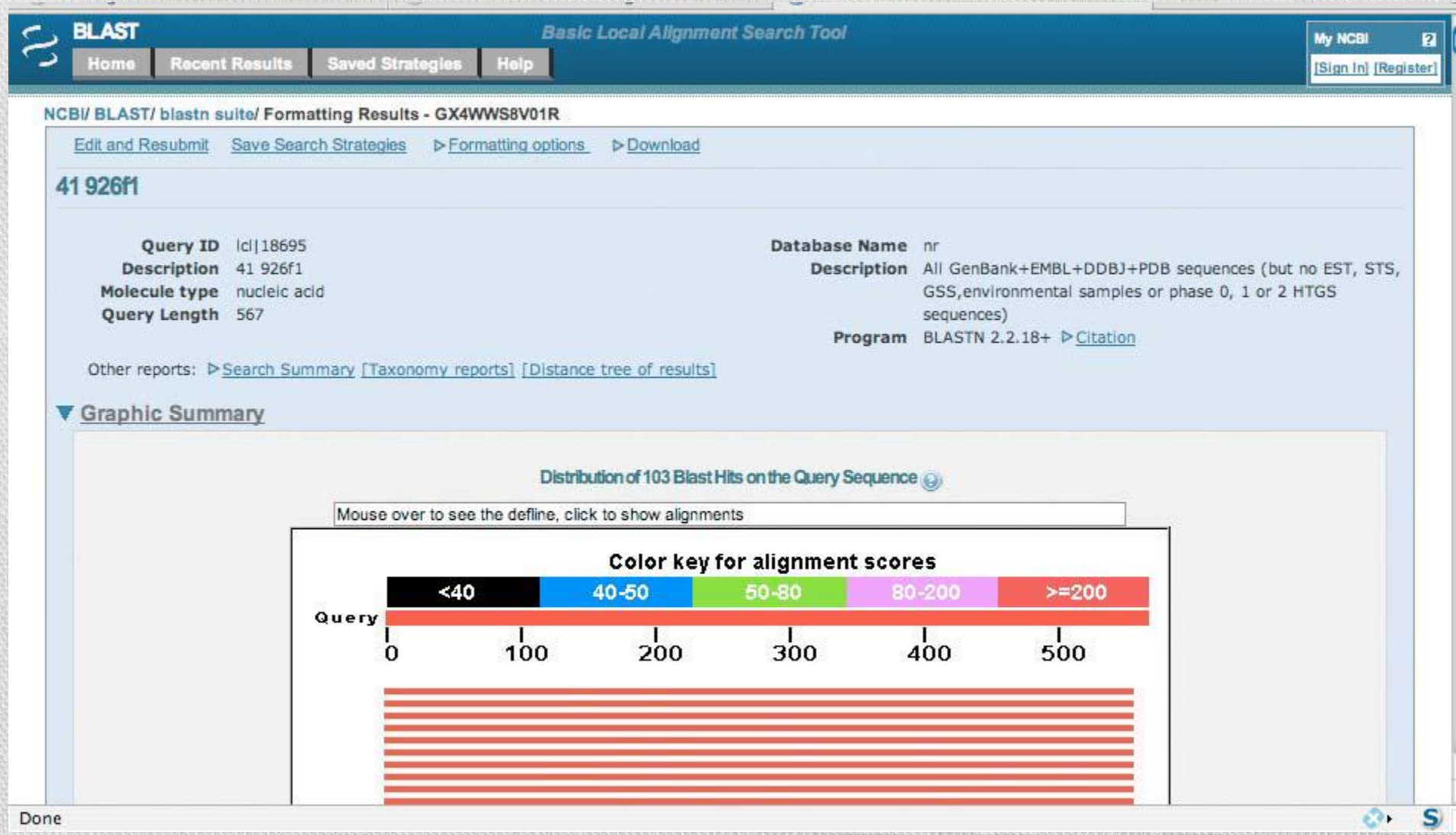
nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query



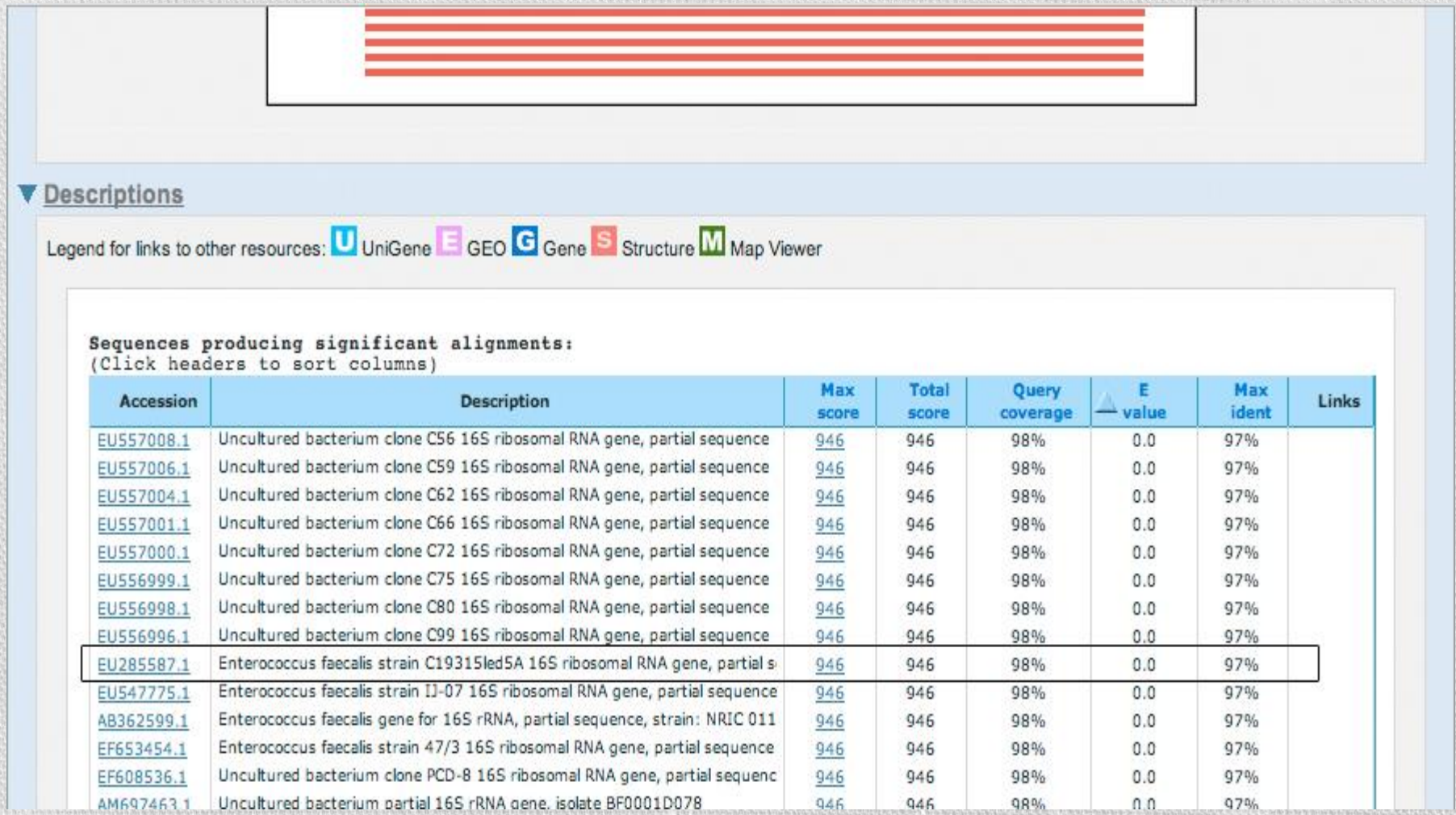
Click the "Blast!" button at the bottom to submit your sequence data.



This screen will come up next. 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



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.



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU557008.1	Uncultured bacterium clone C56 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557006.1	Uncultured bacterium clone C59 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557004.1	Uncultured bacterium clone C62 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557001.1	Uncultured bacterium clone C66 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557000.1	Uncultured bacterium clone C72 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556999.1	Uncultured bacterium clone C75 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556998.1	Uncultured bacterium clone C80 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556996.1	Uncultured bacterium clone C99 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU285587.1	Enterococcus faecalis strain C19315led5A 16S ribosomal RNA gene, partial s	946	946	98%	0.0	97%	
EU547775.1	Enterococcus faecalis strain IJ-07 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
AB362599.1	Enterococcus faecalis gene for 16S rRNA, partial sequence, strain: NRIC 011	946	946	98%	0.0	97%	
EF653454.1	Enterococcus faecalis strain 47/3 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EF608536.1	Uncultured bacterium clone PCD-8 16S ribosomal RNA gene, partial sequenc	946	946	98%	0.0	97%	
AM697463.1	Uncultured bacterium partial 16S rRNA gene, isolate BF0001D078	946	946	98%	0.0	97%	

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.

>gb|EU285587.1| Enterococcus faecalis strain C19315led5A 16S ribosomal RNA gene, partial sequence
Length=1456

Score = 946 bits (512), Expect = 0.0
Identities = 550/566 (97%), Gaps = 12/566 (2%)
Strand=Plus/Plus

```
Query  1     CGGTCGAGC-TGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCC  59
             |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  893   CGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCC  952

Query  60    TTTGACCACTCTAGAGATAGAGCTTTCCCTTCGGGGACAAAGTGACAGGTGGTGCATGGT  119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  953   TTTGACCACTCTAGAGATAGAGCTTTCCCTTCGGGGACAAAGTGACAGGTGGTGCATGGT  1012

Query  120   TGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATT  179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1013  TGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATT  1072

Query  180   GTTAGTTGCCATCATTTAGTTGGGCACTCTAGCGAGACTGCCGGTGACAAACCGGAGGAA  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1073  GTTAGTTGCCATCATTTAGTTGGGCACTCTAGCGAGACTGCCGGTGACAAACCGGAGGAA  1132

Query  240   GGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAAT  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1133  GGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAAT  1192

Query  300   GGGAAGTACAACGAGTCGCTAGACCGCGAGGTCATGCAAATCTCTTAAAGCTTCTCTCAG  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1193  GGGAAGTACAACGAGTCGCTAGACCGCGAGGTCATGCAAATCTCTTAAAGCTTCTCTCAG  1252

Query  360   TTCGGATTGGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCGGATC  419
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1253  TTCGGATTG-CAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCGGATC  1311

Query  420   AGCACGCCGCGGTGAATACGTTGCCGGGGCCTTGTACACACCGCCCGTCACACCACGAGA  479
             |||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||||
Sbjct  1312  AGCACGCCGCGGTGAATACGTTCCCGGG-CCTTGTACACACCGCCCGTCACACCACGAGA  1370

Query  480   GTTTGTAACACCCGAAGTCGG-GAGGTACCCTTTT-GGAGC-A-CCGCCTTAGGTGG-AT  534
             |||||||||||||||||||| |||||| |||||| ||||| | |||||| |||||| ||
Sbjct  1371  GTTTGTAACACCCGAAGTCGGTGAGGTAACCTTTTTGGAGCCAGCCGCCTAAGGTGGGAT  1430

Query  535   AGATGAT-GGGGTGA-GTTC-TAACA  557
             ||||||| ||||||| || | |||||
Sbjct  1431  AGATGATTGGGGTGAAGT-CGTAACA  1455
```

Interpretation of sequences which coding for protein

Translation and Open Reading Frame Search

Regions of DNA that encode proteins are first transcribed into messenger RNA and then translated into protein.

By examining the DNA sequence alone we can determine the sequence of amino acids that will appear in the final protein.

In translation codons of three nucleotides determine which amino acid will be added next in the growing protein chain.

It is important then to decide which nucleotide to start translation, and when to stop, this is called an **open reading frame**.

Once a gene has been sequenced it is important to determine the correct **open reading frame (ORF).**

Every region of DNA has six possible **reading frames,** three in each direction.

The reading frame that is used determines which amino acids will be encoded by a gene.

Typically only one reading frame is used in translating a gene and this is often the longest open reading frame.

Once the open reading frame is known the DNA sequence can be translated into its corresponding amino acid sequence. An open reading frame starts with an **ATG (Met)** in most species and ends with a **stop codon (TAA, TAG or TGA).**

For example,

the following sequence of DNA can be read in six reading frames.

Three in the forward and three in the reverse direction.

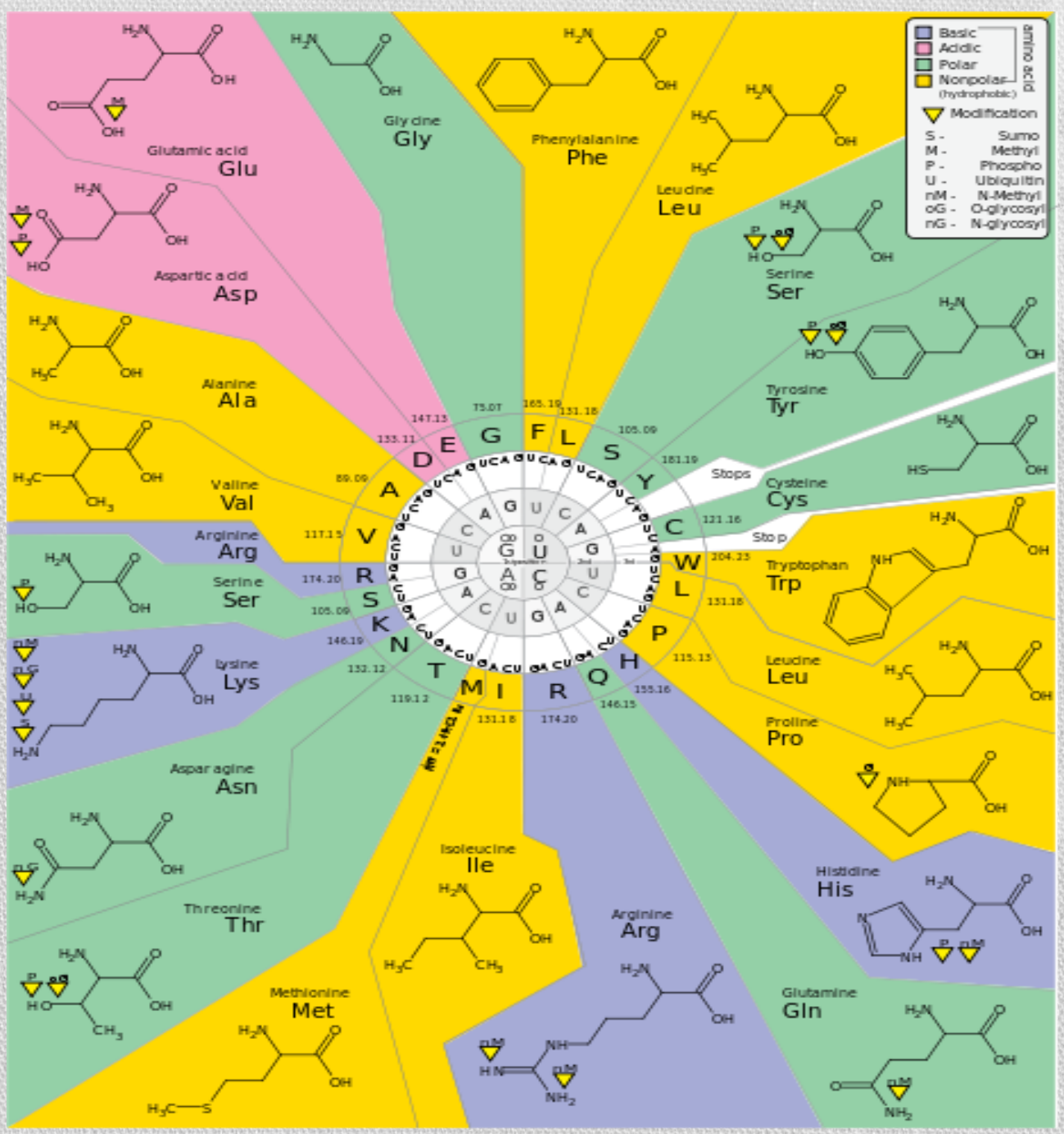
The three reading frames in the forward direction are shown with the translated amino acids below each DNA seqeunce.

Frame **1** starts with the "a", Frame **2** with the "t" and Frame **3** with the "g". Stop codons are indicated by an "*" in the protein sequence.

5' 3'

atgcccaagctgaatagcgtagaggggttttcatcatttgaggacgatgtataa

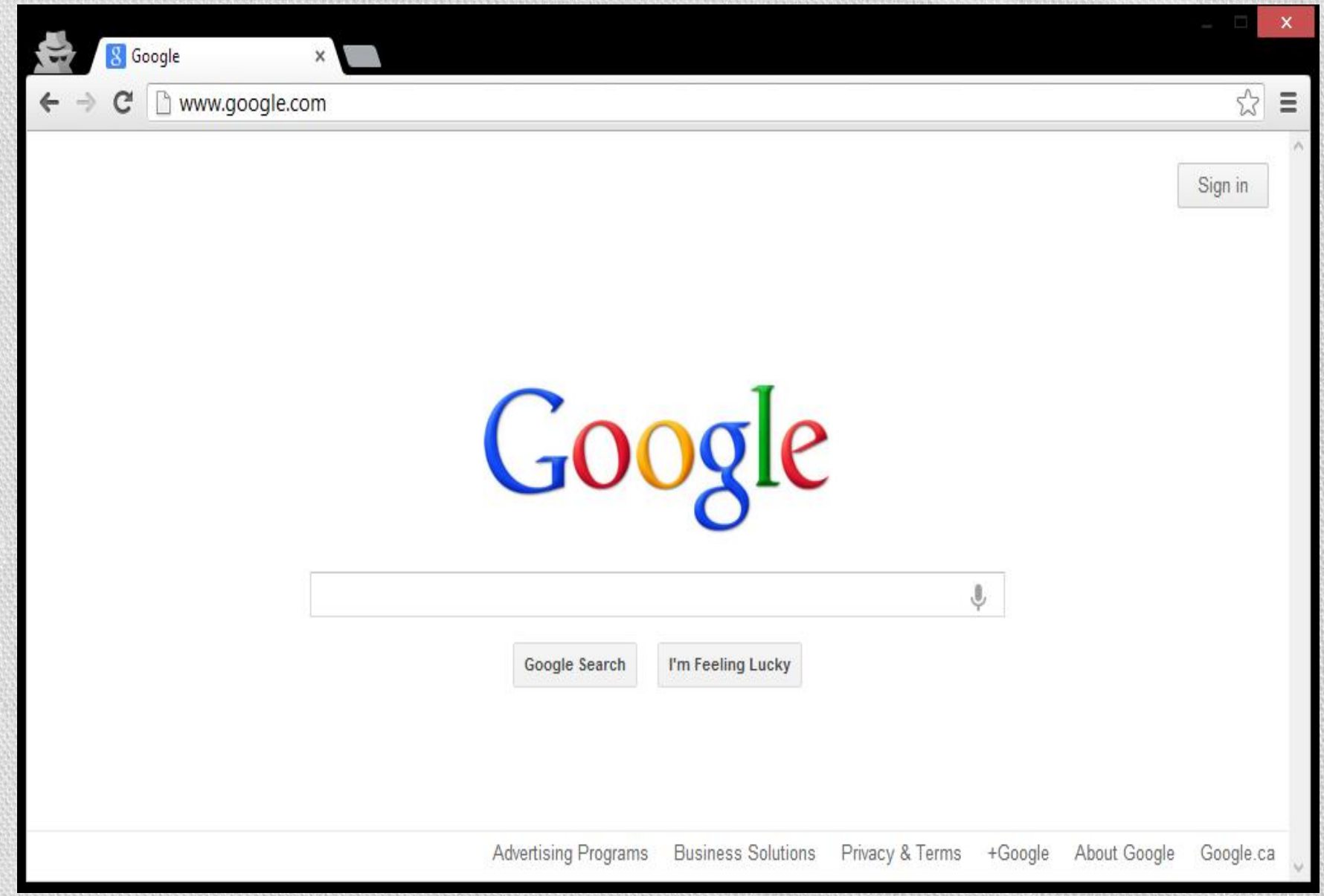
1	atg	ccc	aag	ctg	aat	agc	gta	gag	ggg	ttt	tca	tca	ttt	gag	gac	gat	gta	taa
	M	P	K	L	N	S	V	E	G	F	S	S	F	E	D	D	V	*
2	tgc	cca	agc	tga	ata	gcg	tag	agg	ggt	ttt	cat	cat	ttg	agg	acg	atg	tat	
	C	P	S	*	I	A	*	R	G	F	H	H	L	R	T	M	Y	
3	gcc	caa	gct	gaa	tag	cgt	aga	ggg	gtt	ttc	atc	att	tga	gga	cga	tgt	ata	
	A	Q	A	E	*	R	R	G	V	F	I	I	*	G	R	C	I	

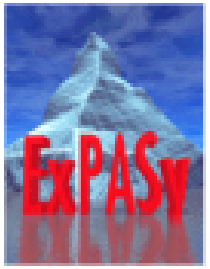

Basic
Acidic
Polar
Nonpolar
(hydrophobic)
amino acid
Modification
S - Sumo
M - Methyl
P - Phospho
U - Ubiquitin
nM - N-Methyl
oG - O-glycosyl
nG - N-glycosyl
Glutamic acid
Glu
Glycine
Gly
Phenylalanine
Phe
Leucine
Leu
Aspartic acid
Asp
Serine
Ser
Alanine
Ala
Tyrosine
Tyr
Valine
Val
Stops
Cysteine
Cys
Stop
Arginine
Arg
Tryptophan
Trp
Serine
Ser
Lysine
Lys
Leucine
Leu
Proline
Pro
Asparagine
Asn
Isoleucine
Ile
Histidine
His
Threonine
Thr
Arginine
Arg
Glutamine
Gln
Methionine
Met

Translation:

Each sequence must be translate to its amino acids (aa) by using

Expasy.translatesoftware


Google
www.google.com
Sign in
Google
Google Search
I'm Feeling Lucky
Advertising Programs
Business Solutions
Privacy & Terms
+Google
About Google
Google.ca



Translate tool

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

Please enter a DNA or RNA sequence in the box below (numbers and blanks are ignored).

```
3601 AAGATACTAG TTTTGCTGAA AATGACATTA AGGAAAGTTC TGCTGTTTTT AGCAAAAGCG
3661 TCCAGAAAGG AGAGCTTAGC AGGAGTCCTA GCCCTTTCAC CCATACACAT TTGGCTCAGG
3721 GTTACCGAAG AGGGGCCAAG AAATTAGAGT CCTCAGAAGA GAACTTATCT AGTGAGGATG
3781 AAGAGCTTCC CTGCTTCCAA CACTTGTTAT TTGGTAAAGT AAACAATATA CCTTCTCAGT
3841 CTACTAGGCA TAGCACCGTT GCTACCGAGT GTCTGTCTAA GAACACAGAG GAGAATTTAT
3901 TATCATTGAA GAATAGCTTA AATGACTGCA GTAACCAGGT AATATTGGCA AAGGCATCTC
3961 AGGAACATCA CCTTAGTGAG GAAACAAAAT GTTCTGCTAG CTTGTTTTCT TCACAGTGCA
4021 GTGAATTGGA AGACTTGACT GCAAATACAA ACACCCAGGA TCCTTTCTTG ATTGGTTCTT
4081 CCAAACAAAT GAGGCATCAG TCTGAAAGCC AGGGAGTTGG TCTGAGTGAC AAGGAATTGG
4141 TTTCAGATGA TGAAGAAAGA GGAACGGGCT TGGAAGAAAA TAATCAAGAA GAGCAAAGCA
4201 TGGATTCAAA CTTAGGTGAA GCAGCATCTG GGTGTGAGAG TGAAACAAGC GTCTCTGAAG
4261 ACTGCTCAGG GCTATCCTCT CAGAGTGACA TTTTAACCAC TCAGCAGAGG GATACCATGC
4321 AACATAACCT GATAAAGCTC CAGCAGGAAA TGGCTGAACT AGAAGCTGTG TTAGAACAGC
4381 ATGGGAGCCA GCCTTCTAAC AGCTACCCTT CCATCATAAG TGACTCTTCT GCCCTTGAGG
4441 ACCTGCGAAA TCCAGAACAA AGCACATCAG AAAAAGCAGT ATTAACTTCA CAGAAAAGTA
```

Output format: Verbose ("Met", "Stop", spaces between residues)

Reset or TRANSLATE SEQUENCE

Strand 1:

1st ORF: 2 stop codons

CGA-GAT-GCC-TAA-ATG-AGT-TGG-CCA-GCA-GAG-CGA-GCA-TGG-ATG-TAA-TCA-G

R D A * M S W P A E R A W M * S

2nd ORF: 1 stop codons

GAG-ATG-CCT-AAA-TGA-GTT-GGC-CAG-CAG-AGC-GAG-CAT-GGA-TGT-AAT-CAG

E M P K * V G Q Q S E H G C N Q

3rd ORF: 0 stop codons

AGA-TGC-CTA-AAT-GAG-TTG-GCC-AGC-AGA-GCG-AGC-ATG-GAT-GTA-ATC-AG

R C L N E L A S R A S M D V I

Reverse complementary strand:

4th ORF: 0 stop codons

CTG-ATT-ACA-TCC-ATG-CTC-GCT-CTG-CTG-GCC-AAC-TCA-TTT-AGG-CAT-CTC-G

L I T S M L A L L A N S F R H L

5th ORF: 1 stop codons

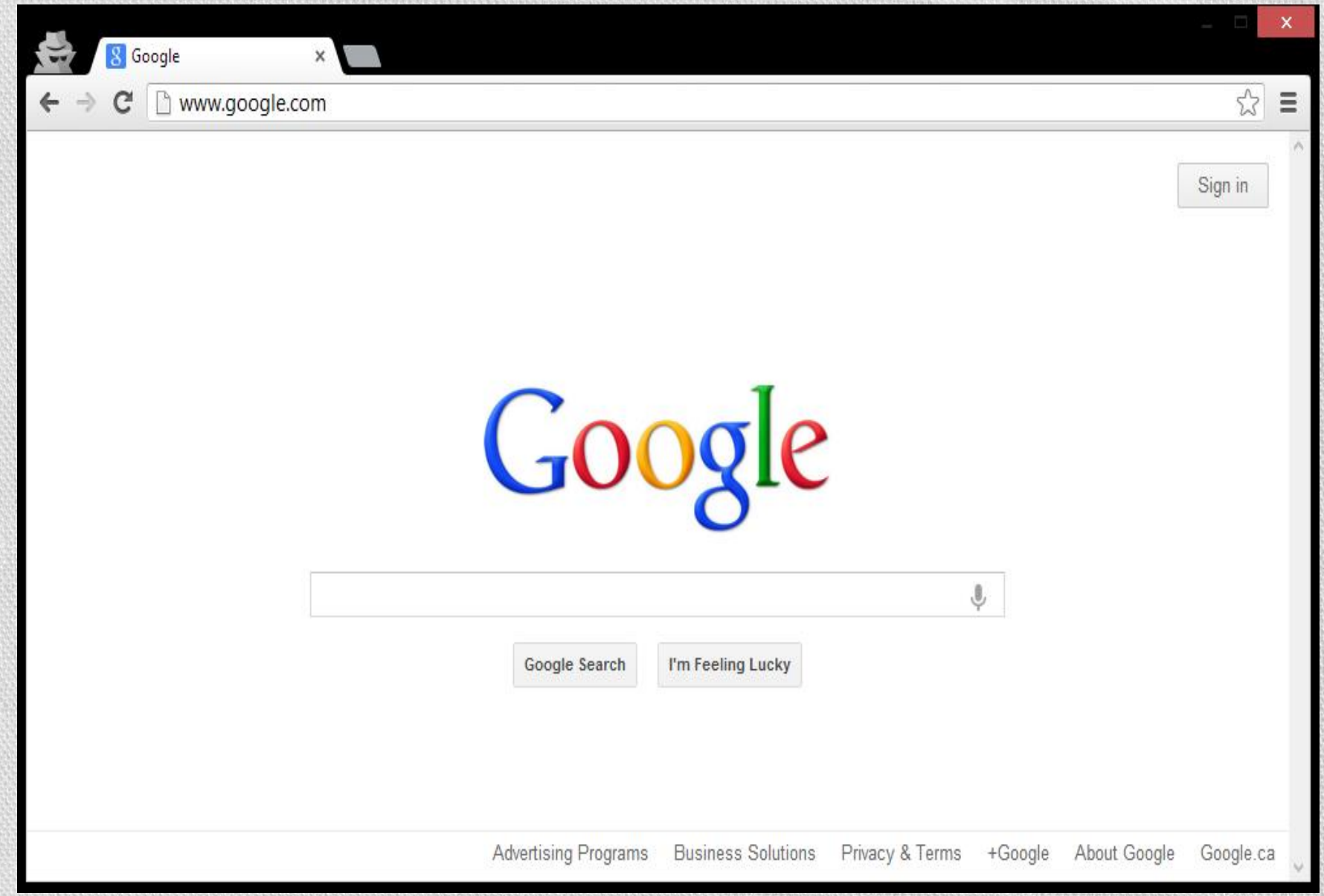
TGA-TTA-CAT-CCA-TGC-TCG-CTC-TGC-TGG-CCA-ACT-CAT-TTA-GGC-ATC-TCG

* L H P C S L C W P T H L G I S

6th ORF: 1 stop codons

GAT-TAC-ATC-CAT-GCT-CGC-TCT-GCT-GGC-CAA-CTC-ATT-TAG-GCA-TCT-CG

D Y I H A R S A G Q L I * A S


Google
www.google.com
Sign in
Google
Google Search
I'm Feeling Lucky
Advertising Programs
Business Solutions
Privacy & Terms
+Google
About Google
Google.ca

BLAST: **Basic Local Alignment Search Tool**

blast.ncbi.nlm.nih.gov/

The **Basic Local Alignment Search Tool** (**BLAST**) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to **...**

Align two or more - Protein BLAST: ***search ... - Nucleotide BLAST Rat - sequences

Nucleotide **BLAST**: Search nucleotide databases using a nucleotide **...**

blast.ncbi.nlm.nih.gov/Blast.cgi?...blastn...BlastSearch...

No BLAST database contains all the sequences at NCBI. BLAST databases **...**

BLAST - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/BLAST

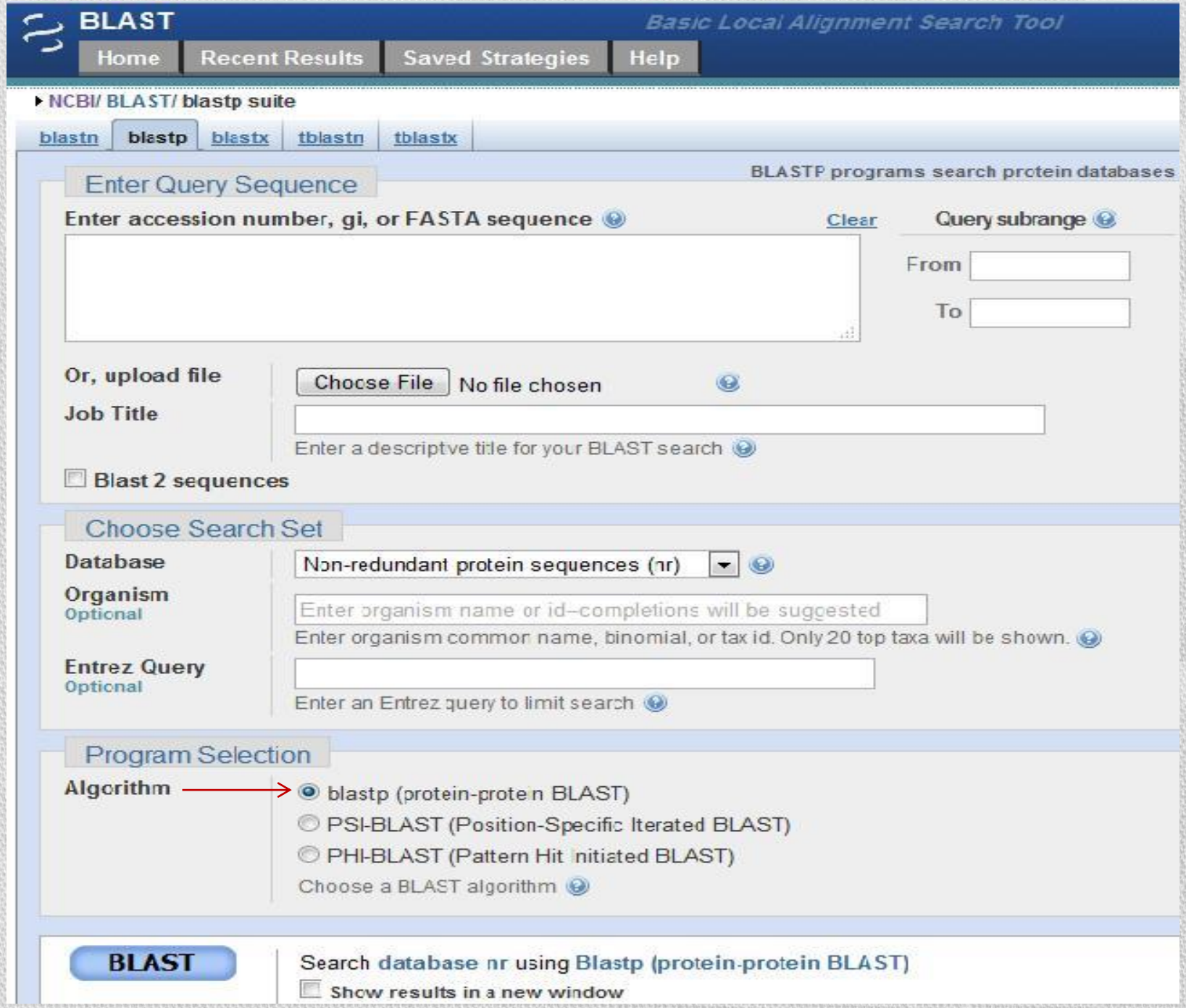
In bioinformatics, **Basic Local Alignment Search Tool**, or **BLAST**, is an algorithm for comparing primary biological sequence information, such as the amino-acid **...**

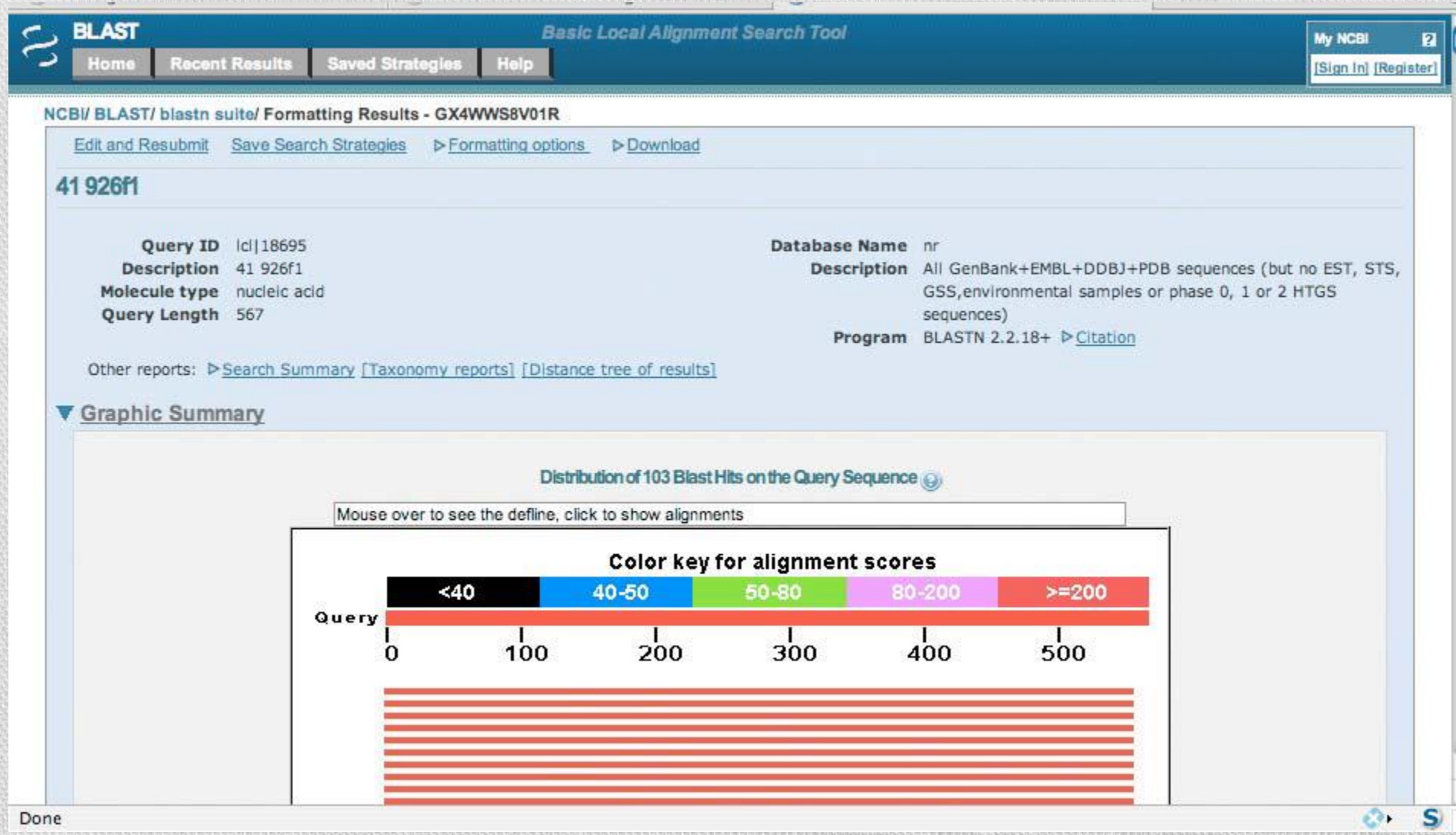
Process - Output - Input - Background

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query


BLAST
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
NCBI/ BLAST/ blastp suite
blastn
blastp
blastx
tblastn
tblastx
BLASTP programs search protein databases
Enter Query Sequence
Enter accession number, gi, or FASTA sequence
Clear
Query subrange
From
To
Or, upload file
Chocse File
No file chosen
Job Title
Enter a descriptve title for your BLAST search
Blast 2 sequences
Choose Search Set
Database
Non-redundant protein sequences (nr)
Organism
Optional
Enter organism name or id--completions will be suggested
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.
Entrez Query
Optional
Enter an Entrez query to limit search
Program Selection
Algorithm
blastp (protein-protein BLAST)
PSI-BLAST (Position-Specific Iterated BLAST)
PHI-BLAST (Pattern Hit Initiated BLAST)
Choose a BLAST algorithm
BLAST
Search database nr using Blastp (protein-protein BLAST)
Show results in a new window



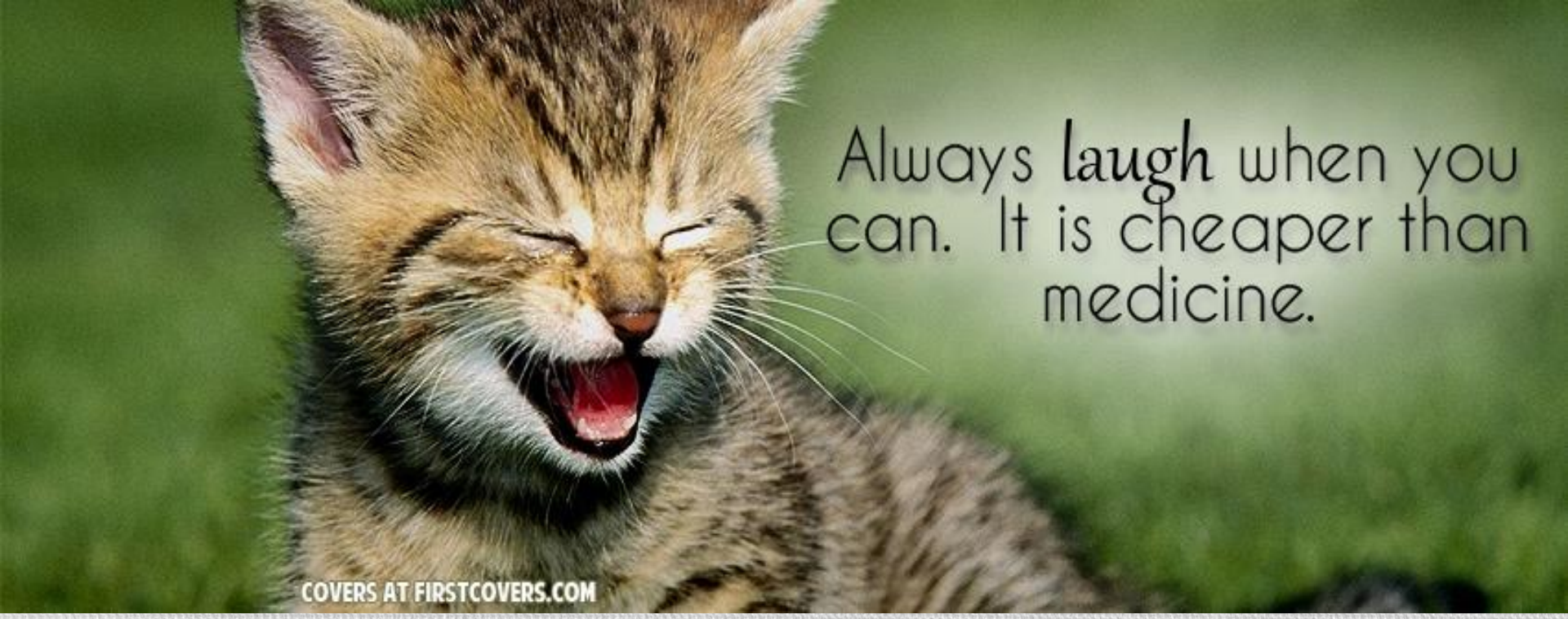
The bars show what places along your aa are similar to other published; the colors indicate how many bases were involved in homology determination.

Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU557008.1	Uncultured bacterium clone C56 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557006.1	Uncultured bacterium clone C59 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557004.1	Uncultured bacterium clone C62 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557001.1	Uncultured bacterium clone C66 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557000.1	Uncultured bacterium clone C72 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556999.1	Uncultured bacterium clone C75 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556998.1	Uncultured bacterium clone C80 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556996.1	Uncultured bacterium clone C99 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU285587.1	Enterococcus faecalis strain C19315led5A 16S ribosomal RNA gene, partial s	946	946	98%	0.0	97%	
EU547775.1	Enterococcus faecalis strain IJ-07 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
AB362599.1	Enterococcus faecalis gene for 16S rRNA, partial sequence, strain: NRIC 011	946	946	98%	0.0	97%	
EF653454.1	Enterococcus faecalis strain 47/3 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EF608536.1	Uncultured bacterium clone PCD-8 16S ribosomal RNA gene, partial sequenc	946	946	98%	0.0	97%	
AM697463.1	Uncultured bacterium partial 16S rRNA gene, isolate BF0001D078	946	946	98%	0.0	97%	



Thanks a lot

with my Best Regards and My Best wishes

Amira A. AL-Hosary
E-mail: Amiraelhosary @yahoo.com
Mob. (002) 01004477501