

I. Hands on exercise #1:

Analysis of a primate VNTR (D1S80 locus)

Several types of repetitive DNA are often used for DNA fingerprinting: Variable Number of Tandem Repeats (VNTRs, or minisatellites), and Short Tandem Repeats (STRs, or microsatellites) VNTR core sequences can vary from about 7-100 bps in length, while STR core sequences are less than 5 bps long. The number of repeats vary considerably (i.e., are highly polymorphic).

The **D1S80** locus on human chromosome #1 contains a 16 bp VNTR with consensus sequence AGGACCACCAGGAAGG. The smallest allele found in humans (so far) contains 14 repeats, while the largest alleles contain up to 72 repeats. A study in 1992 of close to 3,000 randomly chosen individuals revealed the most common alleles contain 18 and 24 repeats, while the rarest contained 14 and 38.

D1S80 primer sequences: Forward primer: 5' -GAAACTGGCCTCCAAACACTGCCC GCCG- 3'
Reverse primer: 5' -GTCTTGTTGGAGATGCACGTGCCCTTGC- 3'

☐ For this exercise, follow the steps below:

1. Go to the NCBI homepage (Google search "NCBI").
2. On left side of page (Resource List) open "DNA & RNA" link
3. Scroll down to the "GenBank" link, and open
4. In search window, type in the accession number for the human D1S80 locus ("**D28507.1**").
5. When the Genbank file opens, click on the "Run BLAST" link (right side). When the BLAST screen opens, go to the bottom and click on "BLAST". *You will have to wait a few seconds for the results.*
6. Scroll down the resulting list of entries and click on links for "Pan troglodytes chromosome 1 clone". The accession number is on the right side of the entry underneath (yup!) the "Accession" heading. Write this accession number down. Now scroll a little further down to the "Gorilla gorilla VNTR" link, and write down this accession number.
7. Now go back to the BLAST link, open it, and then open the "Nucleotide BLAST" link.
8. Top left, type in the human D1S80 accession # (D28507.1) in window saying "Enter Query Sequence". Just below this window find box saying "Align two or more sequences", and check it.
9. Below this, in box saying "Enter Subject Sequence", enter the Pan troglodytes and Gorilla gorilla accession numbers (one below the other), one below the other.
10. Before hitting the BLAST button at bottom of page, check the box saying "show results in a new window". Then click the BLAST button.
11. When you get the BLAST results back, click on the "Alignment" tab, and save the results.

II. Hands-on Exercise #2:

Analysis of calmodulin - a highly conserved calcium signaling protein.

Part 1 (single sequence search):

1. Go to NCBI website → on left side open "Sequence analysis" link → scroll down main "database" column, and under "Tools" heading open "Basic Local Alignment Search Tool (BLAST)"
2. Select "human" under "BLAST genomes" search box, and then open "Protein BLAST" window.
 - *Note tab options at top indicating "blastn", "blastp", "blastx", etc.*
3. Cut/paste the entire protein sequence shown below (human calmodulin) in the "Enter Query Sequence" window (top left). In "Job Title" box, type in "Human CaM-1" → Go to bottom of screen and click on "BLAST" link to activate search. *Search may take 15-30 seconds to generation results.*
 - Human Calmodulin protein sequence (single letter amino acid code):
**MADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGT
IDFPEFLTMMARKMKD TDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMI
READIDGDGQVNYEEFVQMMAK**
4. Click on "Alignments" tab to view results. Scroll down page and open/view several resulting sequence alignments (see links on left side under "description" column). *Note virtual identity among human calmodulin isoform sequences.* If desired, you can save blast results in a text file (optional).
5. Open accession number (right side) of first entry: **NM_001329922.1**
 - Scroll down data sheet and see "Name" and "Source".
 - Under "Features" (near bottom) find "CCDS" → open link: **CCDS 86782.1** and scroll down to see "chromosome 19", "Nucleotide sequence" (mRNA), and translation product.

Part 2 (multiple sequence alignment comparison):

1. Go back to main BLAST page (repeat step 1, part 1 above). However, do not make any selections under "BLAST genomes" search box (leave at default setting), then open "Protein BLAST" window (just as in part 1 above).
2. Just under "Job Title" heading, check box "Align two or more sequences", and then cut/paste entire human calmodulin sequence from part 1 (above) in "Enter Query Sequence" box (top). In "Job Title" box. type "Human-plant CaM align".
3. Cut/paste pear entire calmodulin sequence below in second box ("Enter subject sequence")
 - Pear Calmodulin protein sequence:
**MADQLTDDQIAEFKEAFSLFDKDGDCITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGT
IDFPEFLNLMARKMKD TDSEEEIKEAFRVFDKDGNGFISAAELRHVMTNLGEKLTDEEVDEMI
READVDGDGQINYEEFVKVMAK**
5. Activate BLAST, wait, and review results when complete. *Note 97% sequence identity.*

Part 3 (DNA sequence alignment):

In this final exercise you will perform a nucleotide sequence comparison between a human and chimp calmodulin sequences, and then .

1. Go to main "BLAST" page → Open the "Nucleotide BLAST" link.
2. Type in any name in the "Job Title" box, and then check the box "Align two or more sequences".
3. Cut/paste accession **NM_001329922.1** in the "Enter Query Sequence" box, and cut/paste accession **XM_003817488.3** (chimp) in the "Subject Query box". Activate BLAST.
4. Repeat step 3, with **NM_001329922.1** as "Query" sequence, but cut/paste accession **XM_014666665.1** (Mung bean) in "Subject Query box". Activate BLAST.
5. Review alignment results.

□ If time allows: Brief description of Alu insertions in human genome

(I) Genome database entry for human Alu repeat (PV92) sequence:

LOCUS HUMALURP7 1002 bp DNA linear PRI 27-APR-1993
DEFINITION Human Alu repeat.
ACCESSION **M57427** M33776
VERSION M57427.1 GI:178509
KEYWORDS Alu repeat; repeat region.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1002)
AUTHORS Matera,A.G., Hellmann,U. and Schmid,C.W.
TITLE A transpositionally and transcriptionally competent Alu subfamily
JOURNAL Mol. Cell. Biol. 10 (10), 5424-5432 (1990)
PUBMED [2169023](#)
COMMENT Original source text: **Human DNA, clone PV 92.**
FEATURES
Location/Qualifiers
source 1..1002
/organism="Homo sapiens"
/mol_type="genomic DNA"
/tissue_type="lymphocyte"

[repeat unit](#) **70..76**
/note = "5' insertion target sequence"

[repeat unit](#) **77..384**
/rpt_family="Alu"

[repeat unit](#) **385..391**
/note = "3' insertion target sequence"

```
1 aactgggaaa atttgaagag aaagtacaca gatacatttg taaggttgtc tgttacttag
61 gctacagagg aaagaaggcc gggcgcggtg gctcacgoot gtaatcccag cactttggga
121 ggccgaggcg ggcggatcac gaggtcagga gatcgagacc atcccggcta aaacgctgaa
181 acctcgtctc tactaaaaat acaaaaaatt agccgggcgt agtggcgggc gctgtagtcc
241 cagctacttg ggaggctgag gcaggagaat ggcgtgaacc cgggaggcgg agcttgcagt
301 gagccgagat cctgccactg cactccagcc tgggcgacag agcgagactc cgtctcaaaa
361 aaaaaaaaaa aaaaaaaaaa aaaagaaaga attccctctc tctaaacaca ctctaacaca
421 caggagttga gaactcatcg gatttattcc agaccctcct nnnnnnngga cttctggtag
481 ctgcctttct tggtaagatt tctcagagaa ctgggtagat tcgatggagc ctggggggcat
541 ggttgttggt ccaatgggcg ggggttgggg agacacaagc agttggcagg gaacttacaa
601 atctccagga ggttattcag aataaaaaatt cagatattgc catcttttcc atttcaaagc
661 accaagagtc tgtcatcagc aattgtgcct ttctaggtgt cacctgatat ctaacattga
721 ggcattgaaa gatagatcag gactgatttt actgagcatt ttcaaaggca gcaggcaggc
781 ttttcaatgc tgcatgagat cctcaccttc tctgaggtga cagggccttct gtttcttgag
841 tctccctggt tggctgatgt catcccctcc taccccatga ccatgacacc aactacatcc
901 atgatgtata tttccaatgt gtttcaactt atggtcagct cccaccagct ccatattctg
961 gtgggataag caagccttgc cttgagcttt gcaggctggg cc //
```

Notes: (1) target site duplications (direct repeats) are shown in yellow;
(2) *Alu* sequence in green;
(3) poly A "tail" in purple.

(II) Genome database entry for Alu PV92 insertion sequence into existing Alu sequence:

LOCUS AF302689 788 bp DNA linear PRI 22-JAN-2001
DEFINITION Homo sapiens isolate BAS101 AluPV92 repeat sequence.
ACCESSION **AF302689**
VERSION AF302689.1 GI:11095297
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
AUTHORS Comas,D., Plaza,S., Calafell,F., Sajantila,A. and Bertranpetit,J.
TITLE Recent insertion of an Alu element within a polymorphic human-specific
Alu insertion
JOURNAL Mol. Biol. Evol. 18 (1), 85-88 (2001)
PUBMED [11141195](#)
JOURNAL Submitted (04-SEP-2000) Facultat de Ciències de la Salut i de la
vida, Universitat Pompeu Fabra, Dr Aiguader,80, Barcelona 08003, Spain
FEATURES Location/Qualifiers
source 1..788
/organism="Homo sapiens"
/mol_type="genomic DNA"

[repeat region](#) 1..788
/rpt_family="AluPV92"
/rpt_type=dispersed

[variation](#) 33
/replace="t"

[repeat region](#) 86..744
/rpt_family="AluYa5"
/rpt_type=dispersed

[variation](#) 185
/replace="g"

[repeat region](#) 220..565
/rpt_family="AluYb8"
/rpt_type=dispersed

```
1 aactgggaaa atttgaagag aaagtcacac agatacattt cagtaagggtt gtctctgtta
61 cttgaggctt acaagaagga aagaaggccg ggcgcggttg ctcacgcctg taatcccagc
121 actttgggag gccgaggcgg gcggatcaag aggtcaggag atcgagacca tcccggtctaa
181 aacgctgaaa cctcgtctct actaaaaa ta caaaaaattg ccgggcgcgg tggctcacgc
241 ctgtaatccc agcacttttg gaggccgagg cgggtggatc atgaggtcag gagatcgaga
301 ccatacctggc taacaagggtg aaaccccgtc tctactaaaa atacaaaaaa ttagccgggc
361 gcggtggcgg ggcgcctgta gtcccagcta ctcgggaggc tgaggcagga gaatggcgtg
421 aacccgggaa gcggagcttg cagtgaaccg agattgcgcc actgcagtcc gcagtccggc
481 ctgggcgcaca gagcgagact ccgtctcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
541 aaaaaaaaaa aaaaatacaaa aaattagccg ggcgtagtgg cgggcgcctg tagtcccagc
601 tacttgggag gctgaggcag gagaatggcg tgaacccggg aggcggagct tgcagtgagc
661 cgagatcctg ccactgcact ccagcgtggg cgacagagcg agactccgtc tcaaaaaaaaa
721 aaaaaaaaaa aaaaaaagaa agaattccct ctctaacac actctaacac acaggagttg
781 agaactca //
```

- Notes: (1) flanking target site duplications are shown in yellow;
(2) the original Alu sequence is shown in green (split in two by "new" Alu insertion);
(3) the recently inserted, intact (i.e., "young") Alu element is shown in blue;
(4) poly-A "tails" are shown in purple.