DNA SEQUENCE DATA

-From template DNA to Sequence Alignment...

Case Study:

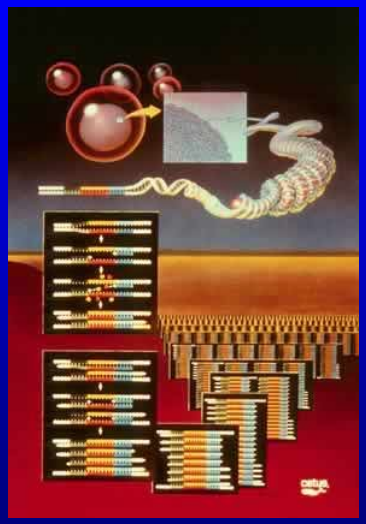
Western Diamondback Rattlesnake (*Crotalus atrox*)

Protocol

1. Collect tissue samples from *C. atrox* individuals and extract tDNA
2. Amplify specific gene using PCR (Polymerase Chain Reaction)
3. Sequence PCR products
4. Align our sequence with published sequences
5. Analyze with phylogenetic software

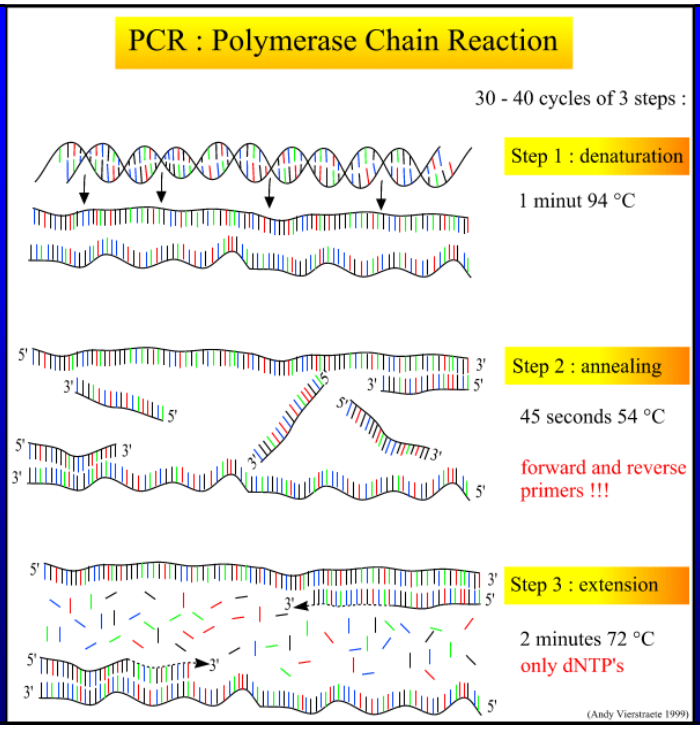
PCR – Purpose

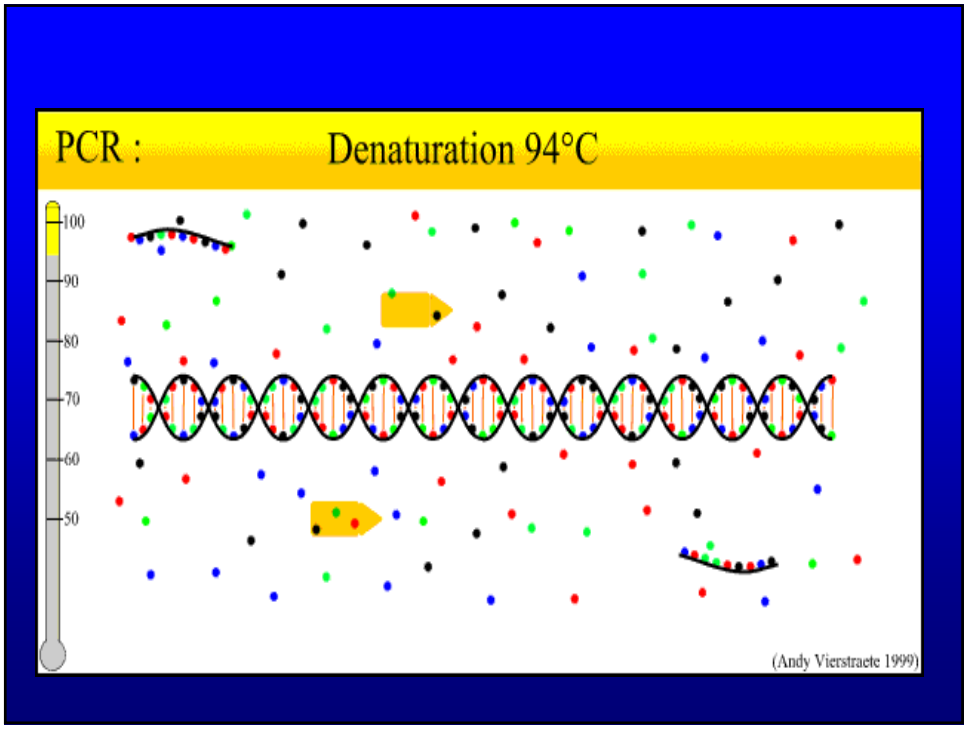
- Need multiple copies of the gene in order to sequence it
- Primer extension reaction for amplification of specific nucleic acids in vitro

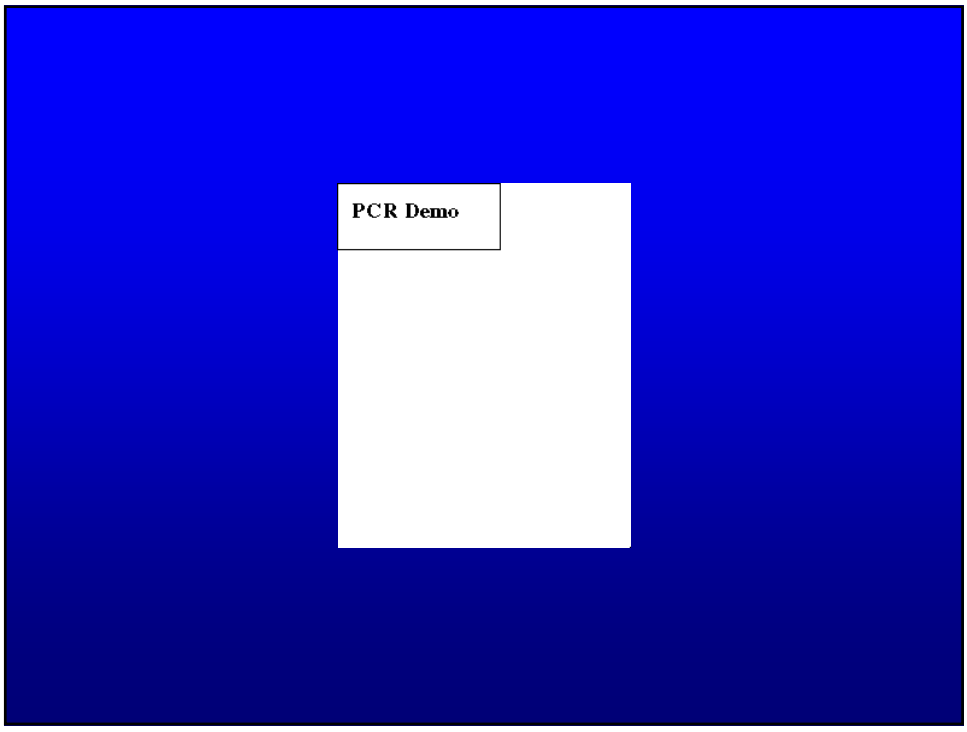


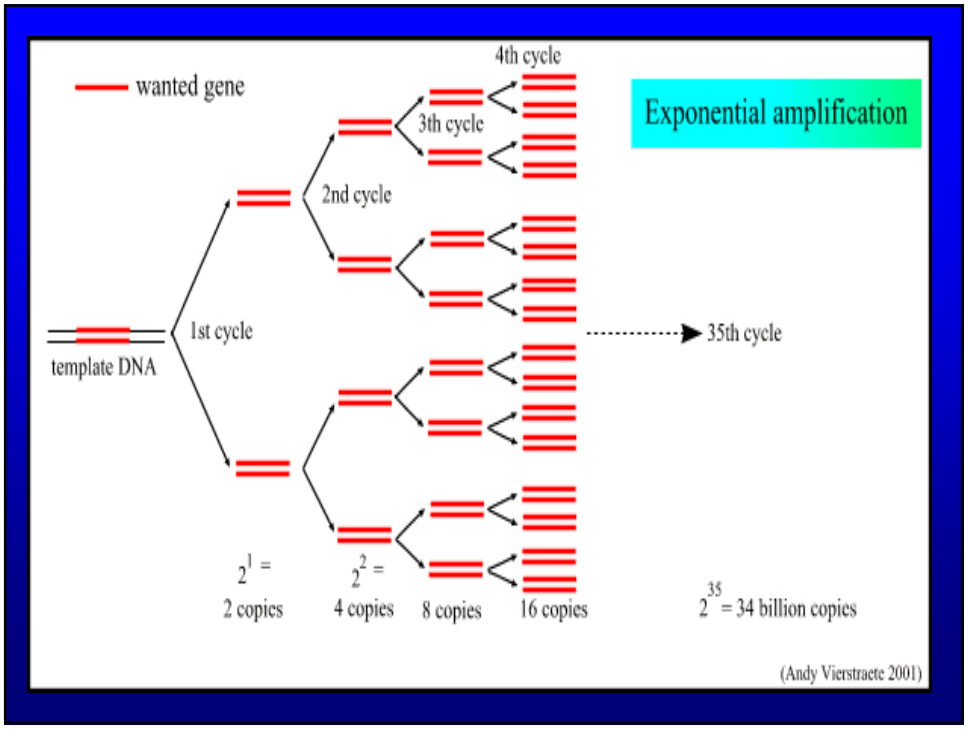
PCR – Reaction Composition

- tDNA
- Sequence specific primers
- dNTP's
- Taq polymerase
- Buffer
- Thermocycler

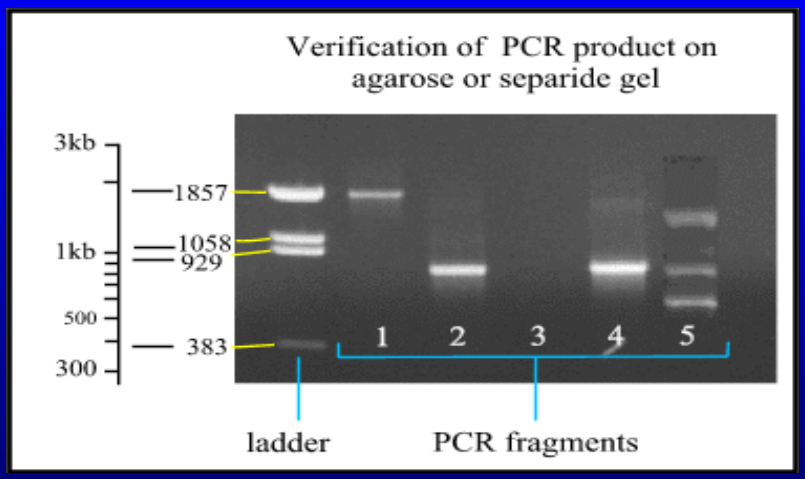



PCR : Denaturation 94°C
100
90
80
70
60
50
(Andy Vierstraete 1999)

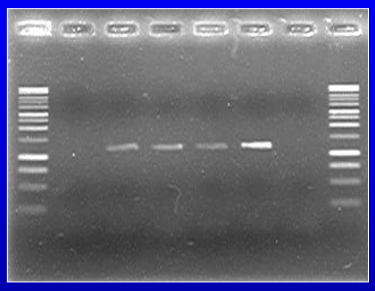

PCR Demo



PCR – How do we know it worked?



DNA Sequencing



How to get from this…

To this!

TATCCGCATAATACAGATCCTCCCCACAACAAAAACCGACCTATTCCTTCCATTCATCAT
TCTAGCCCTCTGAGGGGCAATTCTAGCCAATCTCACATGCCTACAACAGACAGACCTAAA
ATCCCTAATCGCCTACTCCTCCATCAGCCACATAGGCCTAGTAGTAGCCGCAATTATTAT
CCAAACCCCATGAGGCCTATCCGGAGCCATAGCTCTAATAATCGCACACGGATTTACCTC
CTCAGGCACTCTTCTGCCTAGCTAACACAACCTATGAACGAACACACACCCGAGTCCTAAT
TCTTACACGAGGATTCCACAATATCCTACCCATAGCTACAACCTGATGACTAGTAACAAA
CCTCATAAACATCGCCATCCCCCCCTCCATAAACTTCACCGGAGAGCTCCTAATTATATC
CGCCCTATTTAACTGATGCCCAACAACAATCATCATACTAGGAATATCAATACTTATCAC
CGCCTCTTACTCCCTACATATATTTCTGTCAACACAAATAGGGCCAACTCTACTAAACAA
CCAAACAGAACCCACACACTCCCGAGAACACCTACTAATAACCCTCCACCTTGCCCCCCT
ACTTATGATCTCCCTCAAACCAGAATTAGTCATCAGGAGTGTGCGTAATTTAAAGAAAAT
ATCAAGCTGTGACCTTGAAAATAGATTAACCTCGCACACCGAGAGGTCCAGAAGACCTG
C TAACTCTTCAATCTGGCGAA--CACACCAGCCCTCTCTTCTATCAAAGGAGAATAGTTA-
CCCGCTGGTCTTAGGCACCACAACTCTTGGTGCAAAT

Automated DTCS

(Dye Terminator Cycle Sequencing)

- Typically provides accurate reads of 600-800 b.p.
- For long fragments, two or more sequencing reactions are run
- Up to 96 run at once in a plate

- Reaction is similar to a PCR reaction, but there is no logarithmic replication, so technically a primer extension reaction

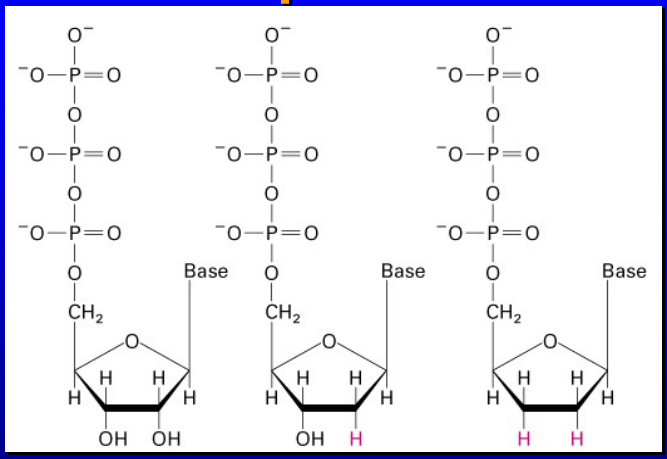
Components

- Purified PCR product (template)
- Primer (1 per sequencing reaction)

Components

- Thermostable DNA polymerase
- Buffer, $MgCl_2$
- Deoxynucleoside triphosphates (dNTPs)
- Dideoxynucleoside triphosphates (ddNTPs)
  - Each with a different fluorescent label
  - Much smaller molar concentration than dNTPs

Components



Ribonucleoside triphosphate
RNA

Deoxynucleoside triphosphate
DNA

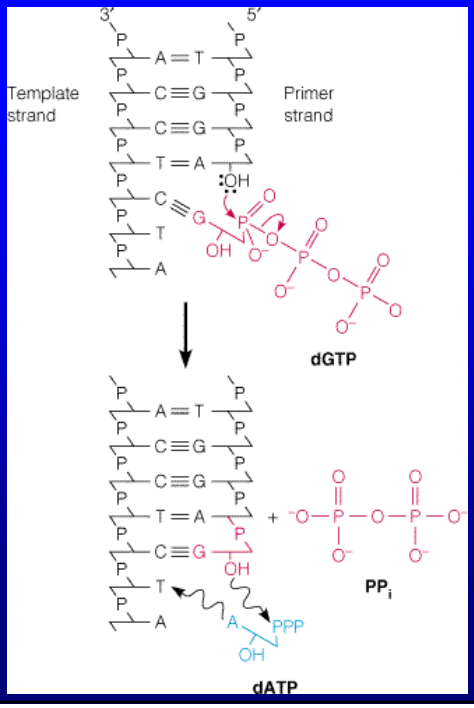
Dideoxynucleoside triphosphate
DDNA

Reaction

- Similar to a PCR reaction:
  - Denature at ~96°C
  - Anneal primer at ~50°C
  - Extend primer at ~60°C
- Primer extension occurs normally as long as dNTPs are incorporated
- When a ddNTP is incorporated, extension stops

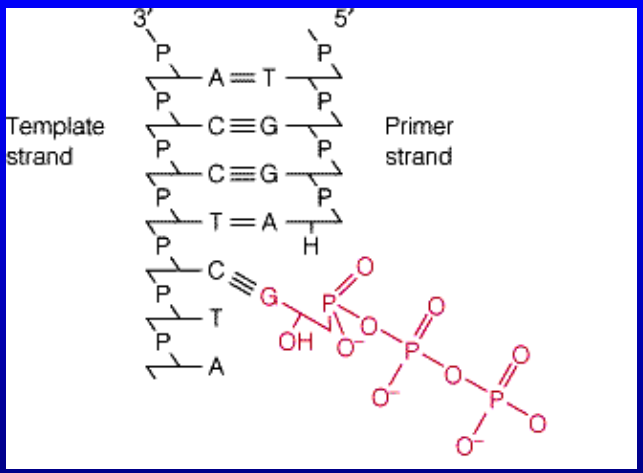
Reaction

- Extension occurs via nucleophilic attack
  - 3'-hydroxyl group at the 3' end of the growing strand
  - attacks the 5'-α-phosphate of the incoming dNTP,
  - releasing pyrophosphate (PP_i).
  - $(dNMP)_n + dNTP \rightarrow (dNMP)_{n+1} + PP_i$
  - Catalyzed by DNA polymerase
  - Synthesis occurs 3' → 5'



Reaction

- ddNTPs lack a 3'-OH group
- Once a ddNTP is incorporated, nucleophilic attack cannot occur, so primer extension is terminated



http://www.lsic.ucla.edu/ls3/tutorials/gene_cloning.html

Reaction

- Produces a mixture of single-stranded DNA products of varying lengths
  - Each ends with a dye-labelled ddNTP
  - Hopefully, everything from P + 1 to P + n

Reading the sequence

- DNA from the sequencing reaction is purified via ethanol precipitation
- DNA is resuspended in deionized formamide
- Plate is loaded into the automated sequencer

Automated sequencing

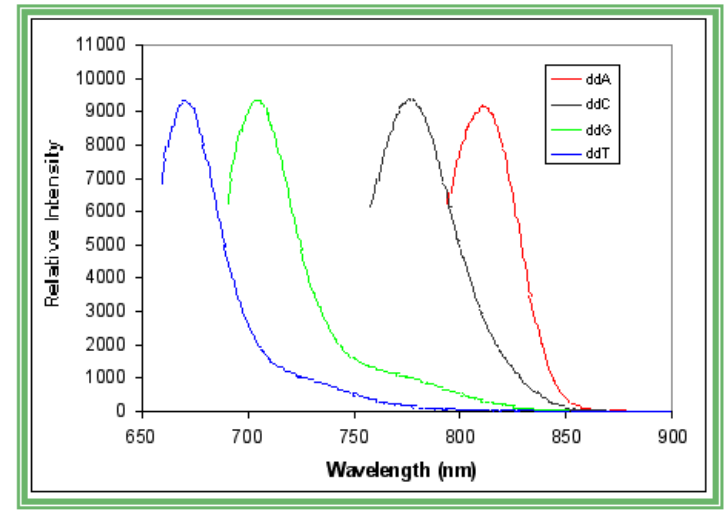
- Capillary array contains polyacrylamide gel

- DNA fragments migrate through gel by electrophoresis
- Separate by size

Automated sequencing

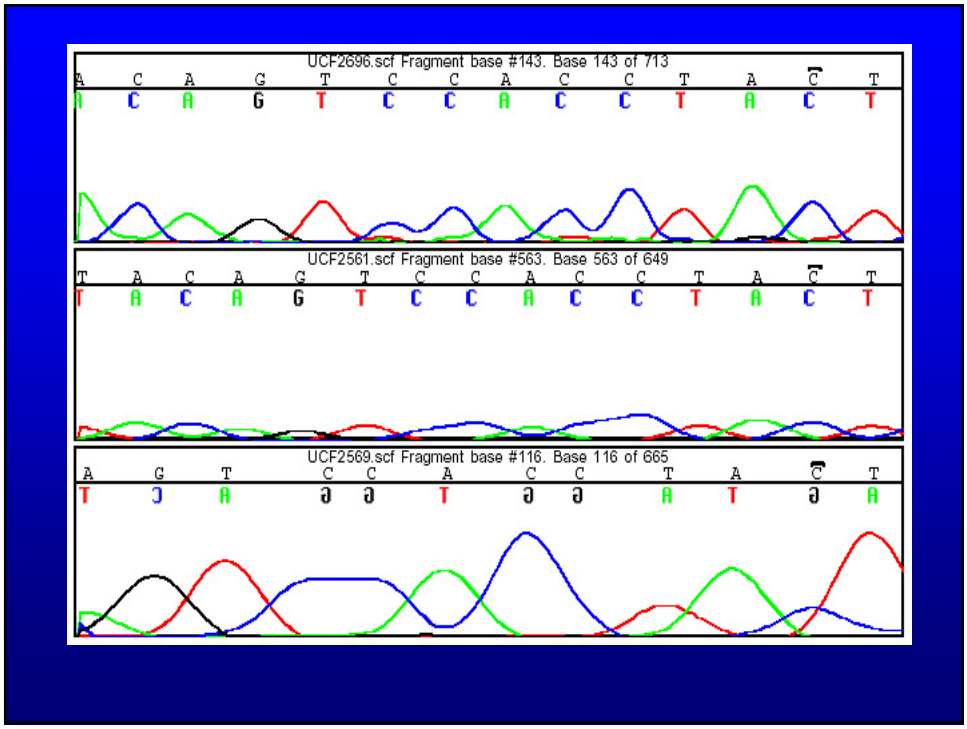
- Capillary passes through a laser
- Each dye fluoresces a different wavelength when excited by the laser
- Fluorescence is detected by a CCD

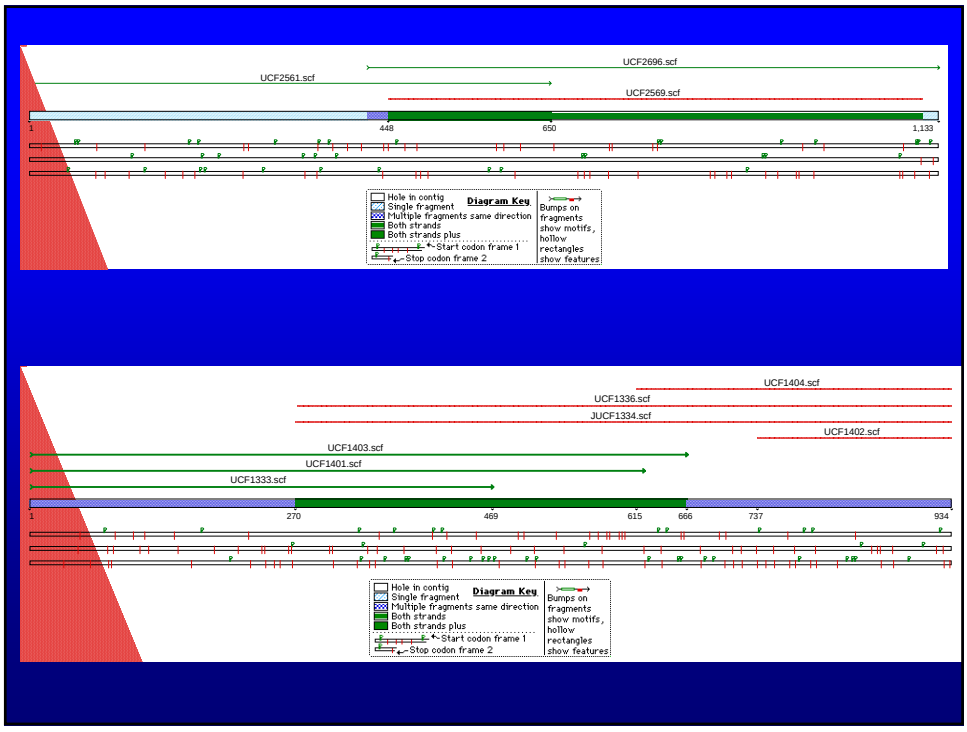
Fluor Emission Spectra



Automated sequencing

- Fluorescences are processed into an electropherogram
- Base "calls" made by sequencing software, but can be analyzed manually


UCF2696.scf Fragment base #143. Base 143 of 713
A C A G T C C A C C T A C T
UCF2561.scf Fragment base #563. Base 563 of 649
T A C A G T C C A C C T A C T
UCF2569.scf Fragment base #116. Base 116 of 665
A G T C C A C C T A C T


UCF2696.scf
UCF2561.scf
UCF2569.scf
1
448
650
1,133
Hole in contig
Single fragment
Multiple fragments same direction
Both strands
Both strands plus
Start codon frame 1
Stop codon frame 2
Diagram Key
Bumps on fragments show motifs, hollow rectangles show features
UCF1404.scf
UCF1336.scf
JUCF1334.scf
UCF1402.scf
UCF1403.scf
UCF1401.scf
UCF1333.scf
1
270
469
615
666
737
934
Hole in contig
Single fragment
Multiple fragments same direction
Both strands
Both strands plus
Start codon frame 1
Stop codon frame 2
Diagram Key
Bumps on fragments show motifs, hollow rectangles show features

NCBI – National Center for Biotechnology Information

- http://www.ncbi.nlm.nih.gov/
- Literature databases
- Entrez databases
- Nucleotide databases
- Genome resources
- Analytical tools

Literature databases

- PubMed – searchable citation database of life science literature
- PubMed Central – digital versions of life science journals
- Bookshelf – online versions of textbooks
- OMIM – catalog of human genes and genetic disorders
- PROW – Protein Reviews On the Web – reviews of proteins and protein families

Entrez databases

- System for searching several linked databases:
  - PubMed
  - Protein sequence databases
  - Nucleotide sequence databases
  - Genome databases
  - Pop sets
  - Books

Nucleotide databases

- GenBank - annotated collection of all publicly available nucleotide and amino acid sequences
- SNPs - Single base Nucleotide Polymorphisms - substitutions and short deletion and insertion polymorphisms
- ESTs - Expressed Sequence Tags - short, single-pass sequence reads from mRNA

Genome resources

- Whole genomes of over 800 organisms
- Others in progress
- Viroids, viruses
- Plasmids
- Bacteria
- Eukaryotic organelles

Genome resources

- Eukaryotes
  - Yeast
  - Fruit fly
  - Zebrafish
  - Human
  - *C. elegans*
  - *Rattus, Mus*
  - *Plasmodium*
  - Plants

Analytical tools

- Sequence analysis tools
- Macromolecular and 3-dimensional structure analysis
- Software downloads
- Citation searching
- Taxonomy searching
- Sequence similarity searching – BLAST

Where are we now??

- Kelly has shown you PCR….
- Matt has explained sequencing…
- Now we must use BLAST with our sequence to determine if we have the correct:
  - Gene
  - Animal

BLAST

- Basic Local Alignment Search Tool
- Similarity Program
  - Compares input sequences with all sequences (protein or DNA) in database
  - Each comparison given a score
    - Degree of similarity between query (input sequence) and sequence that it is being compared to
    - Higher the score, the greater the degree of similarity

BLAST, cont'd

- Significance of each alignment composed as an E-value
  - The number of different alignments with scores equal to or greater than the given score that are expected to occur in a database search by chance
  - The lower the E-value, the more significant the score