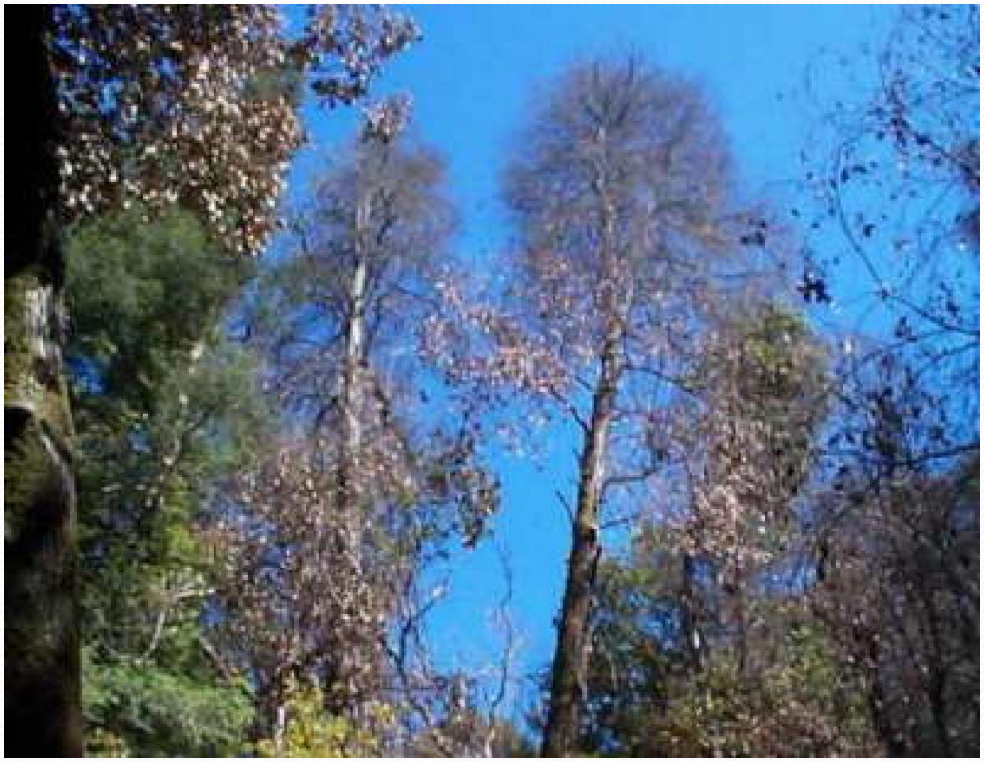
Invasive Plant Diseases Molecular Diagnostics

27 October 2013

Catherine Eyre, PhD.
UC Berkeley
Forest Pathology and Mycology Laboratory

Outline

- What can molecular biology tell us about a pathogen
- Tools and techniques used for diagnostics
  - ELISA
  - PCR
  - Sequencing
  - Sequence alignment
  - Phylogenetic trees
- Investigation scenario



www.suddenoakdeath.org

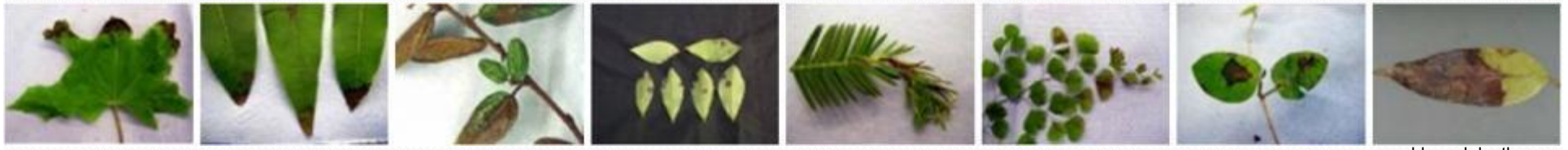
What do we want to know about a pathogen?

- **Detect and identify pathogen**
- **Determine strain/lineage/subgroup**
  - They can behave differently – more/less pathogenic, infect different species
- **Quantify size and range of pathogen population**
  - Where is it, and how many trees/plants are infected
- **Epidemiology: following the movement and migration of pathogens**
  - How quickly is it spreading and to where
- **Determine origins of disease outbreaks**
  - Where did it come from? Should we quarantine? Can we prevent another outbreak and/or further spread?
- **Predict how it will behave in the future**

Pathogen detection and identification

- Symptoms, host species, environment
- Culturing - pathogen morphology
- **Molecular diagnostics**
  - help confirm ID
  - Can also detect cryptic pathogens that you can't see or culture

- 2 approaches
  - Antibodies i.e. ELISA or DNA



www.suddenoakdeath.org

ELISA assay

- **E**nzyme **L**inked **I**mmuno**S**orbent **A**ssay
- Detects pathogen proteins using antibodies

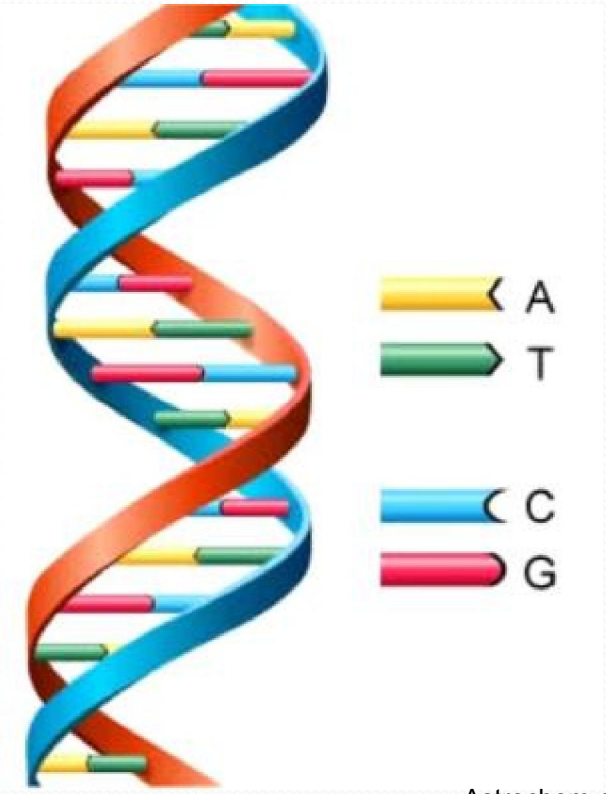
- Agdia field test Immunostrip kit
  - 1 sample at a time

- Lab test kit
  - test 96 samples at a time

Agdia Inc.

DNA

- Molecule found in the cells of all organisms
- Double helix
- Four nucleotides building blocks
- Complementary pairs
  - Adenine, Thymine
  - Cytosine, Guanine
- Unique sequence in all organisms



Astrochem.org

Organism – **Cell nucleus** – **Chromosome** – **Gene** – **Codons** - **Bases**

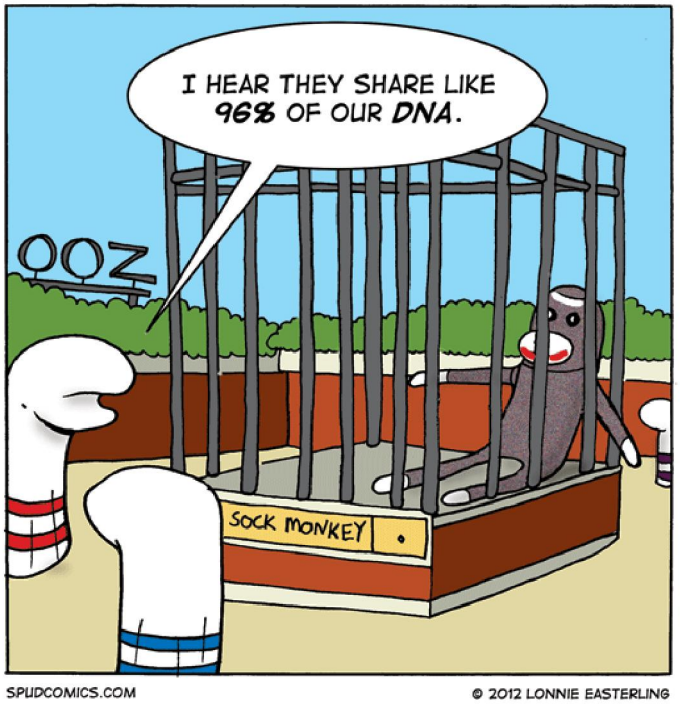
Library - **Book** - **Chapter** - **Sentence** – **Words** - **Letters**

The Language of Genes by Steve Jones

DNA Barcoding

- Every individual has a unique DNA sequence, but…
- *We are also very similar*

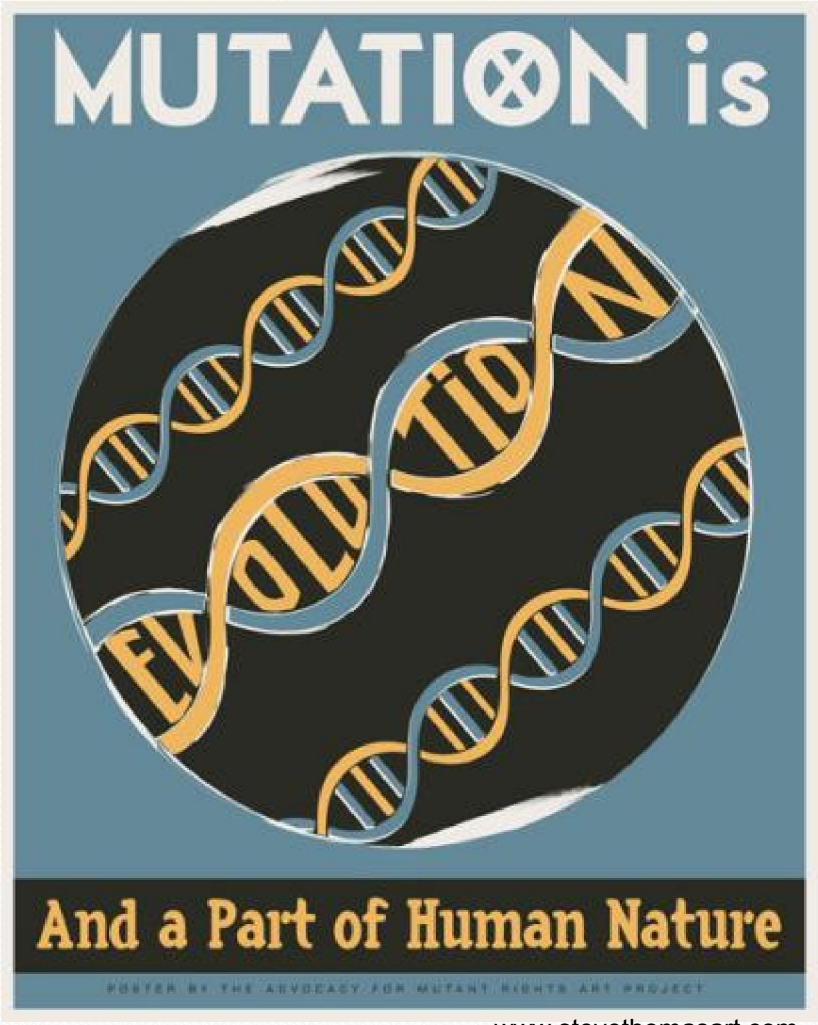
- 99% of our DNA is identical to other humans
- 98% similar to Chimpanzees
- 92% with Mice
- 26% with Yeast (Ref: koshland-science-museum.org)



- To differentiate between individuals, species or organisms, you need to find where the differences are

DNA Polymorphisms

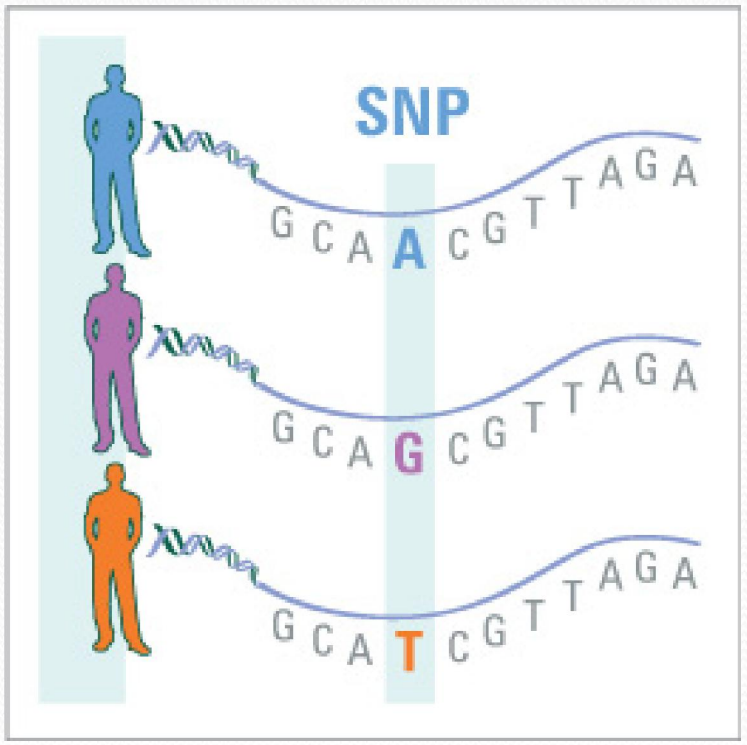
- Differences between genes in different organisms
- Mutations in the DNA sequence
  - Mistakes during DNA replication
  - DNA damage from environment – UV, chemicals



www.stevethomasart.com

SNP mutation

- **S**ingle **N**ucleotide **P**olymorphism
- Change of a single base pair to another



broadinstitute.org

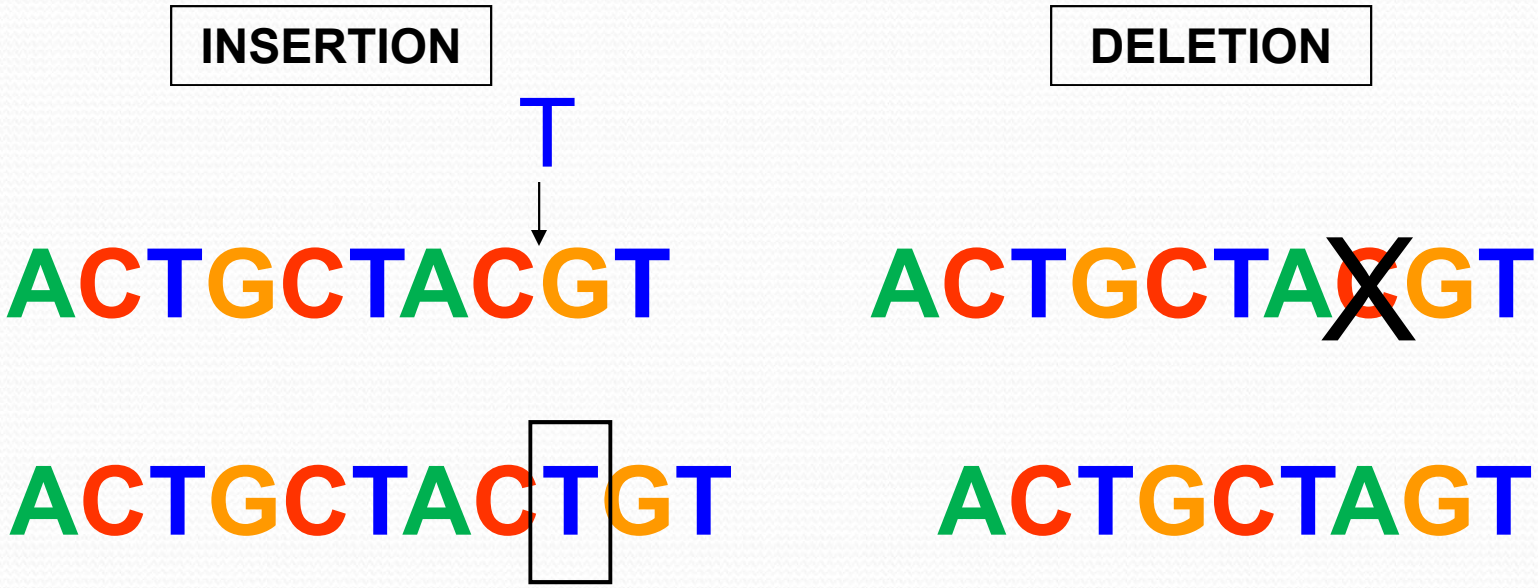
ACTGCTACGT

ACTTCTACGT

DNA replicates, and makes a mistake, or mutation

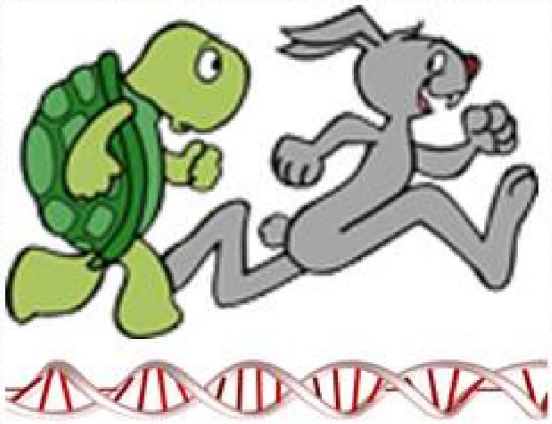
Indel mutation

- **In**sertions or **del**etions
- One or more basepairs are lost or gained



Mutation rates vary

- **Different regions** of DNA gain mutations and evolve at **different rates**
  - How fast depends on the function of the affected DNA
  - **Vital functional gene** – fewer mutations
    - Damaged DNA may not survive
  - **Non-functional DNA** – a change may have no effect on the host, so mutations can persist

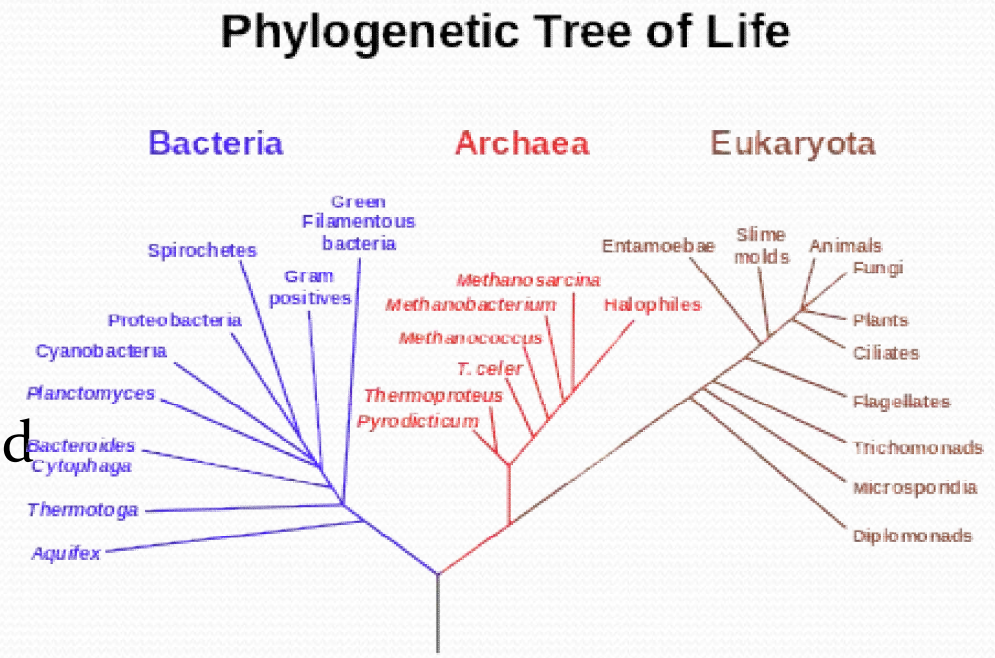


Which gene to study

Depends on what taxonomic level you want to separate

Conserved

- Slow mutation rate
- Takes many generations to accumulate differences
- Usually have vital functional roles
- Use to separate domains, kingdoms, and distantly related organisms
- Example:
  - Ribosomal genes – 16s

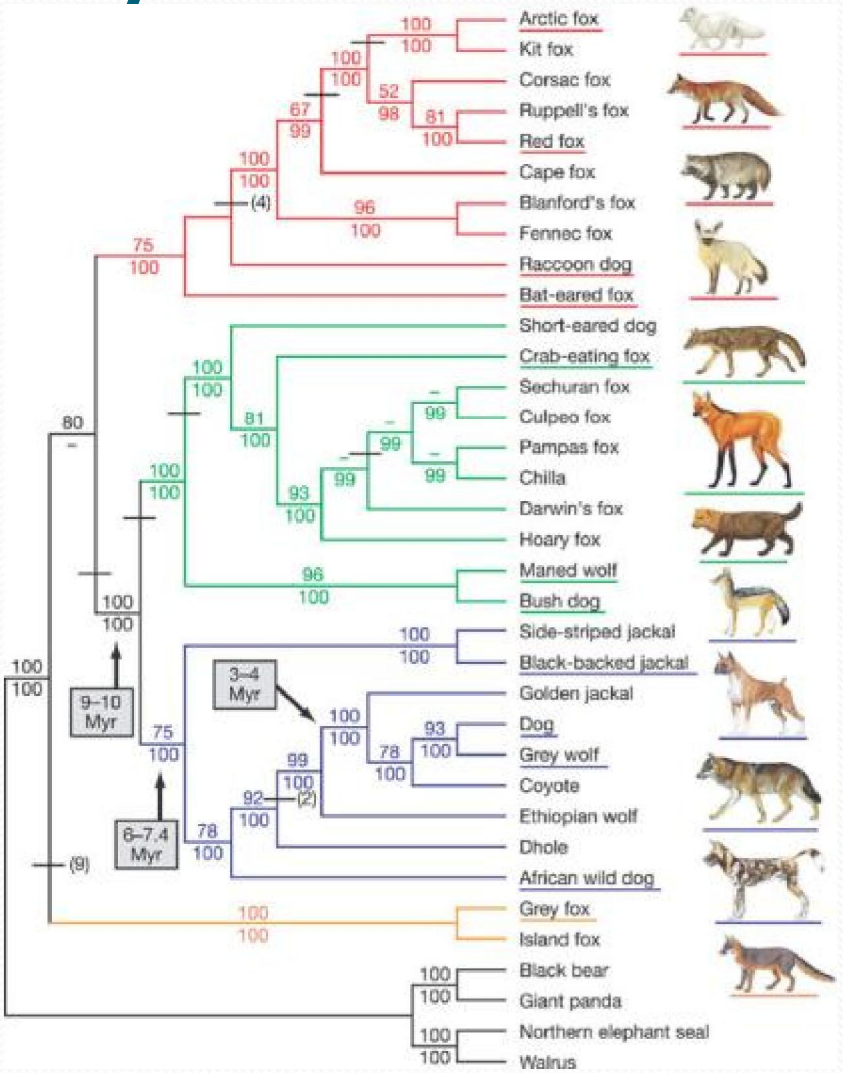


wikipedia

Which gene to study

Variable

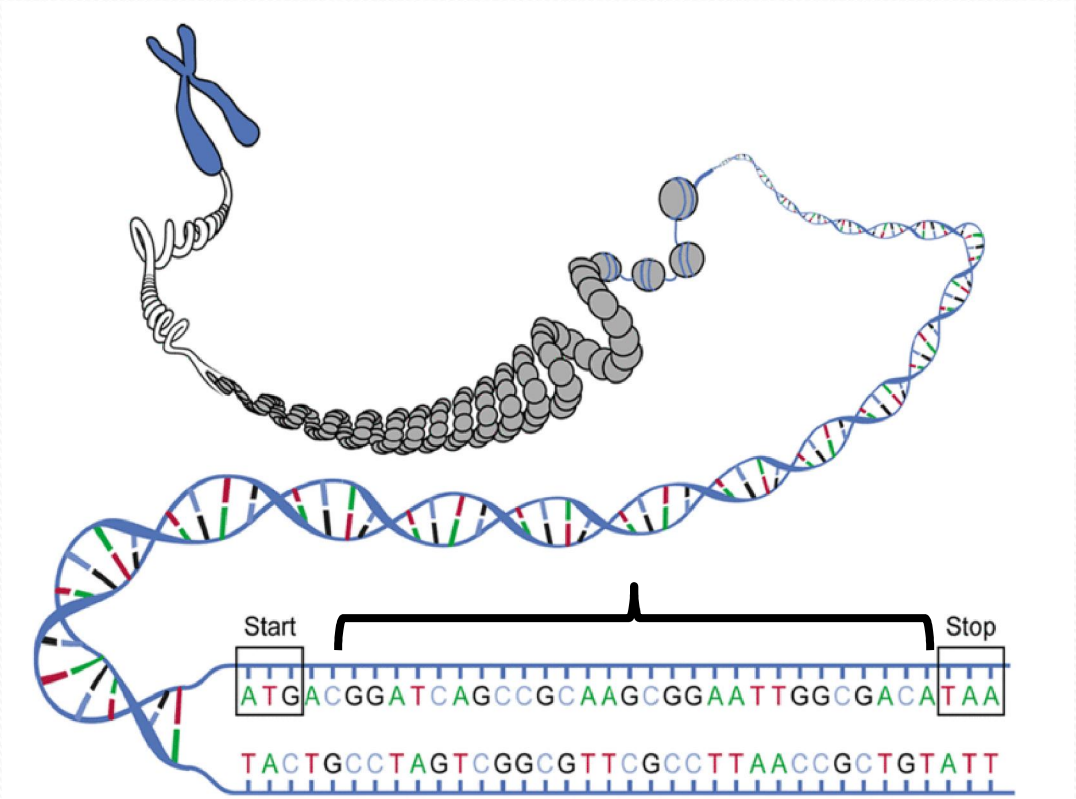
- Fast mutation rate
- Accumulate differences relatively quickly
- Usually non-functional areas of DNA
- Use to compare closely related species, or strains and subgroups within the same species
- Examples:
  - Mitochondrial genes - Cytochrome oxidase
  - Chloroplast genes - RUBISCO
  - Internal transcribed spacer regiong - ITS



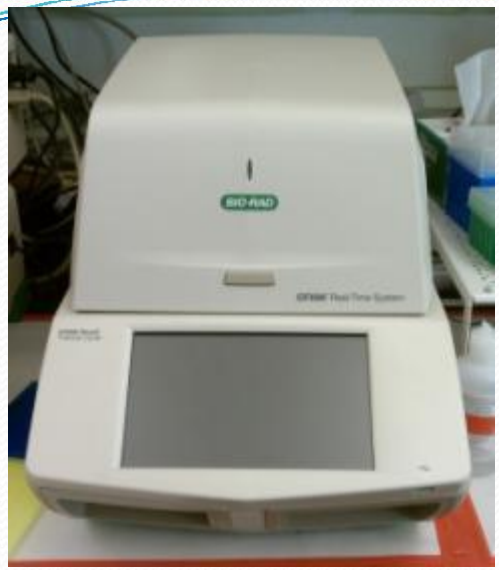
www.bio.miami.edu

PCR

- **P**olymerase **C**hain **R**eaction = **Amplifying DNA**
- Choose gene to study
- Make more copies of it so you can **detect** it



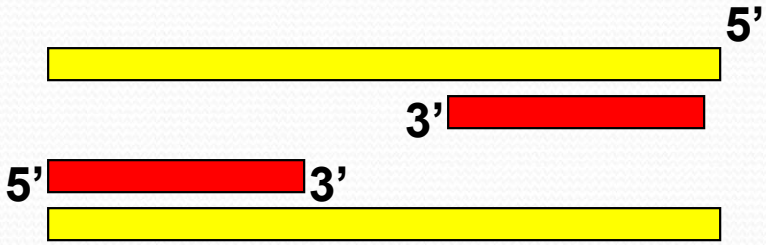
- Very sensitive
- Only small amount of starting DNA needed
- Can use to detect the DNA of a pathogen on leaves
- Pathogen doesn't have to be alive!



PCR reaction

1. Double strand denaturation **95°C**

2. Annealing of the primers **50-65°C**



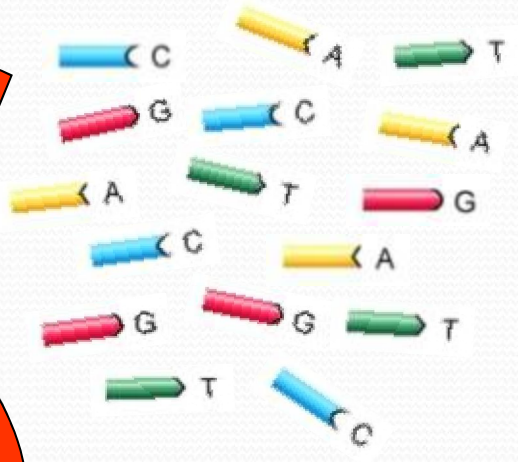
3. Elongation **72°C**

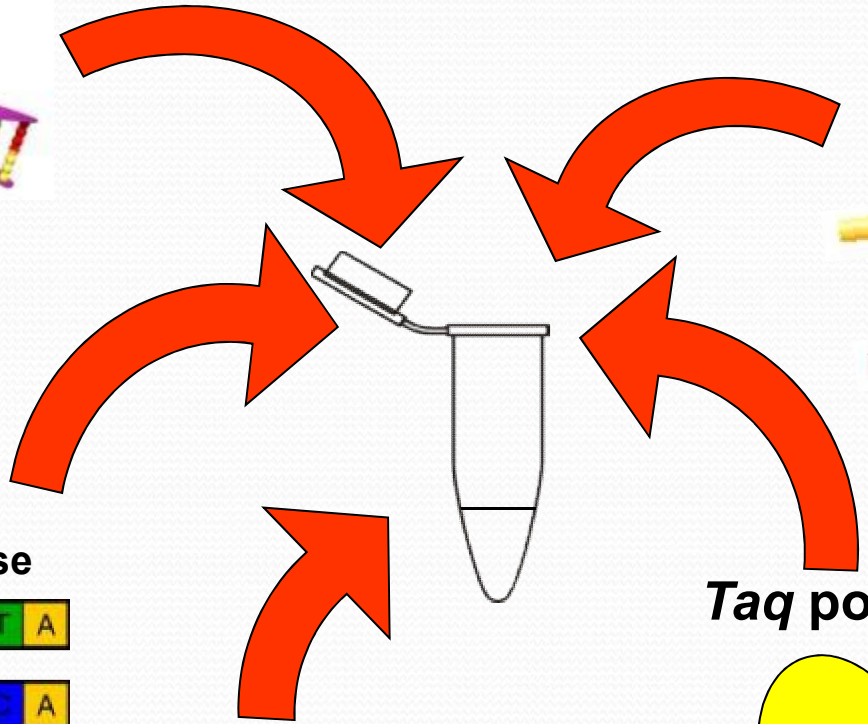
PCR amplification

DNA template

Nucleotides

dATP, dCTP, dGTP, dTTP





Primers

Forward and Reverse

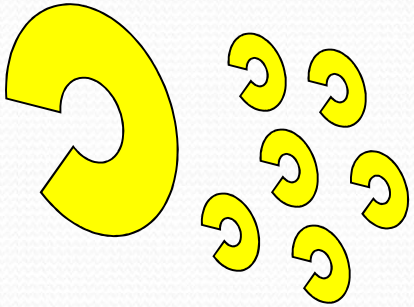
Designed to specific areas of DNA

Water

Buffer

$MgCl_2$

***Taq* polymerase**



PCR videos

- **Conventional PCR**
- http://www.youtube.com/watch?v=HMC7c2T8fVk
- http://www.youtube.com/watch?v=DkT6XHWne6E

- **Real time PCR**
- http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/real-time-pcr.html

- The PCR song!
- http://www.youtube.com/watch?v=x5yPkxCLads

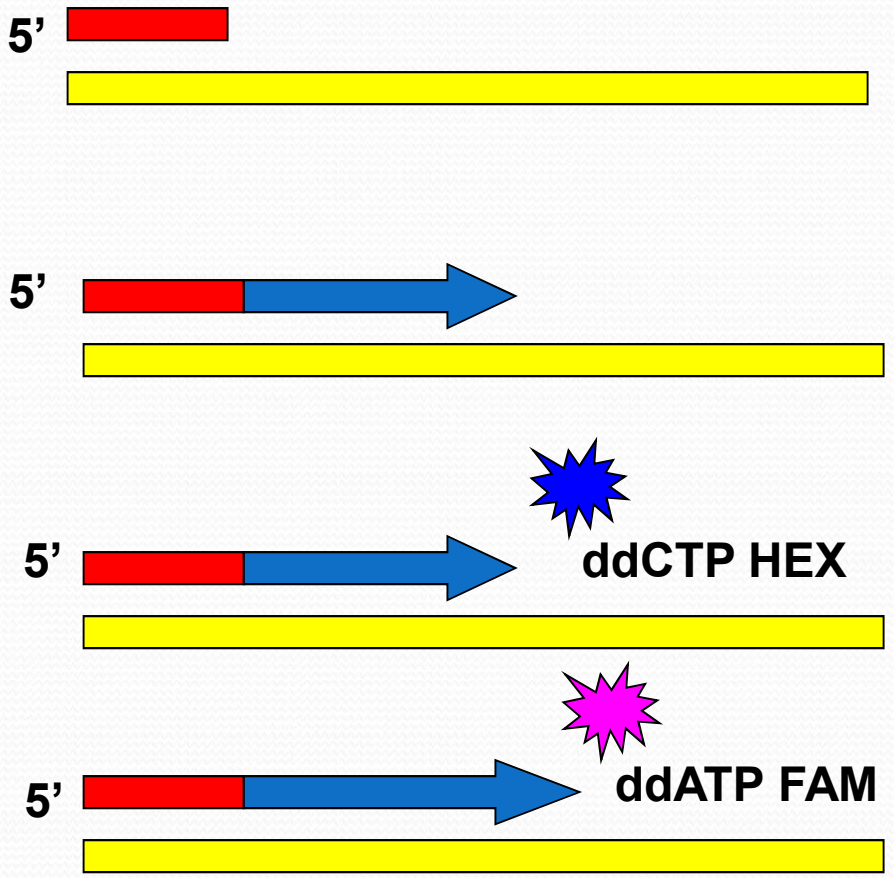
Sequencing

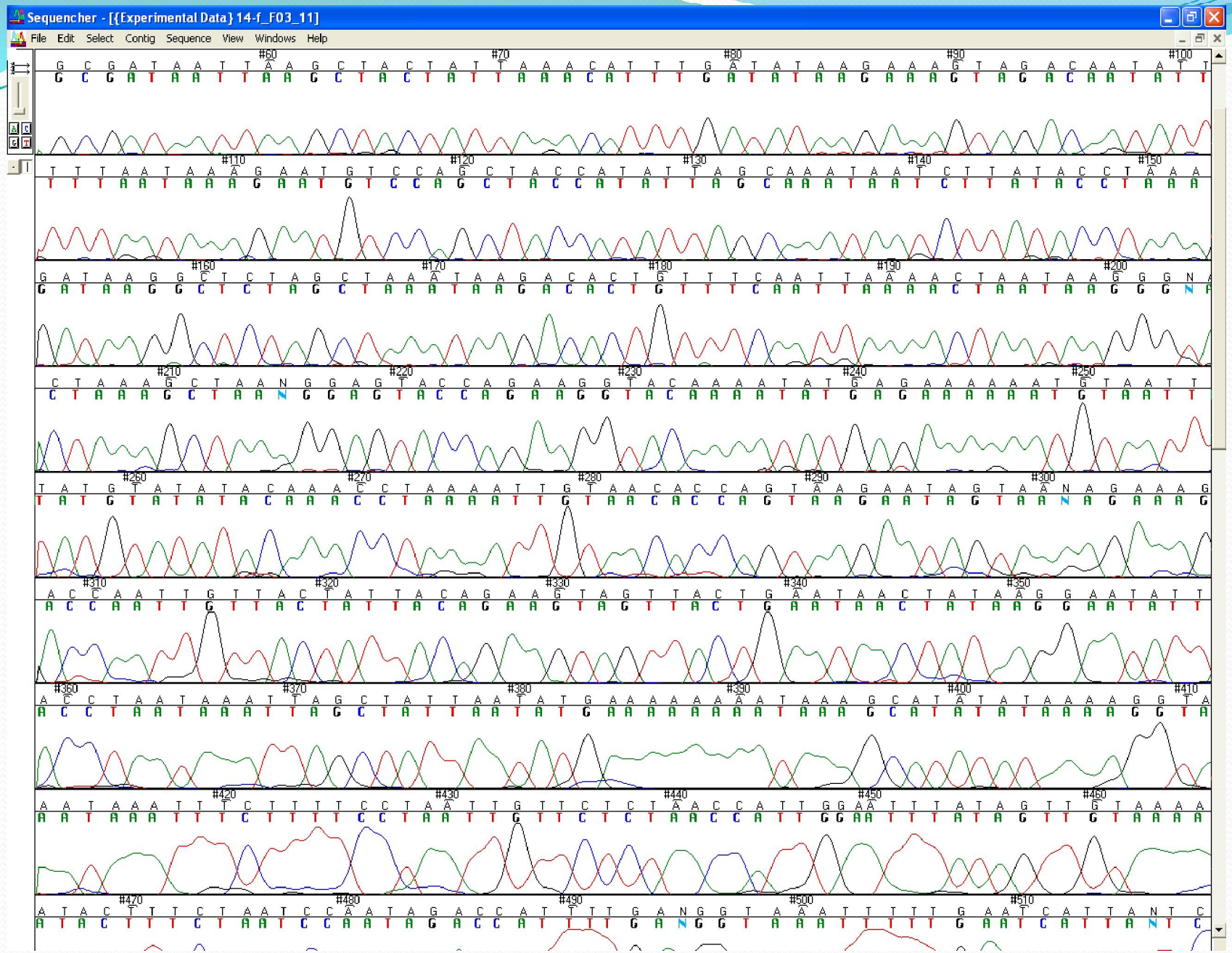
- Determine the exact sequence of As, Ts, Gs and Cs in a particular region of DNA
- Often use the ITS region – good for distinguishing species – especially in fungi
- Use the sequence to:
  - **Identify** an unknown organisms
    - Using the BLAST database
  - **Compare** with other organisms to infer relatedness
    - Alignment and Phylogenetic tree building

>gi|331702969|emb|FR850495.1| **Heterobasidion annosum** genomic DNA containing 18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2, 28S rRNA gene, isolate H4_10
TGGCCTCTCGGGGCATGTGCTCGCCTTGTTCATATCCATCTCACACCTGTGCACACTCGCG
TGGGTCGGTCGGGTTCTTTTGACCCCTTCCGAGCCGCGTCTTCTCACAAACTCTTCGTATGT
CTTTAGAATGGTATCAATGCTATAAAACGCATCTTATACAACTTTCAACAATGGATCTCTCGG
TTCTCGCATCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTG
AATCATCGAATCTTTGAACGCACCTTGCGCCCTTTGGTATTCCGAAGGGCACGCCTGTTTGA
GTGTCGTGAAATTCTCAACCCTGTGCTTTTCTTGTGAAAGCGCGTGGGCTTGGACTTGGAG
GCTTTGCTGGTCCTCGCGGATCGGCTCCTCTCAAATGCATTAGCGAGACCCTTGTGGTGCC
GCCCCCGGTGTGATAATT

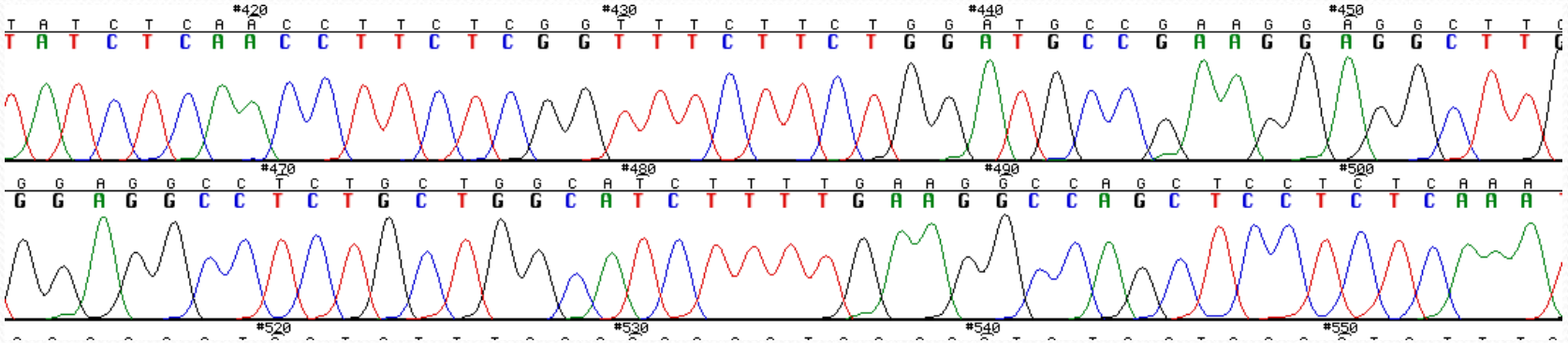
Sequencing reaction

1. Annealing
2. Elongation
3. Incorporation of ddNTP and stop of the elongation

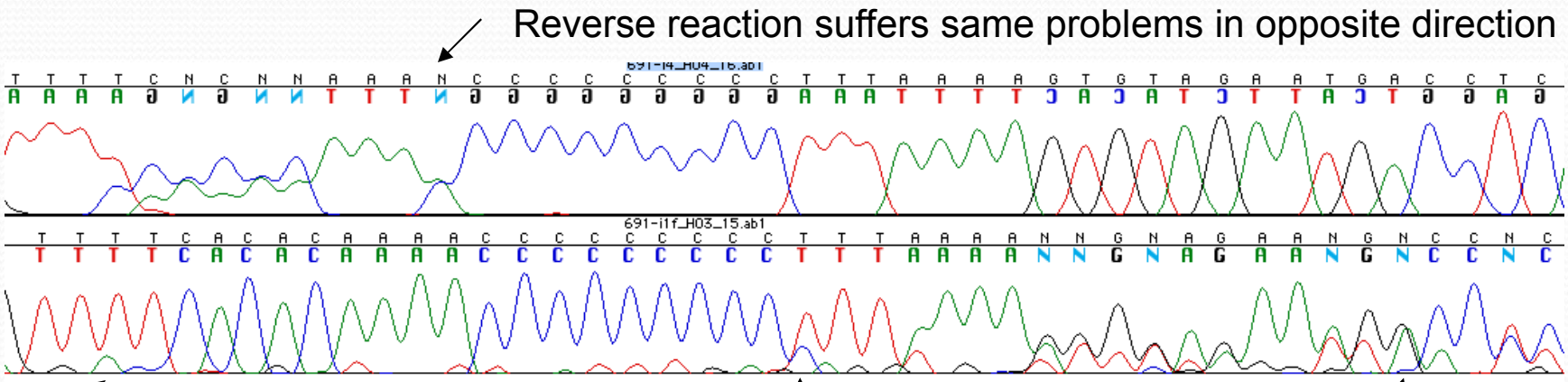



Sequencher - [{Experimental Data} 14-f_F03_11]
File Edit Select Contig Sequence View Windows Help

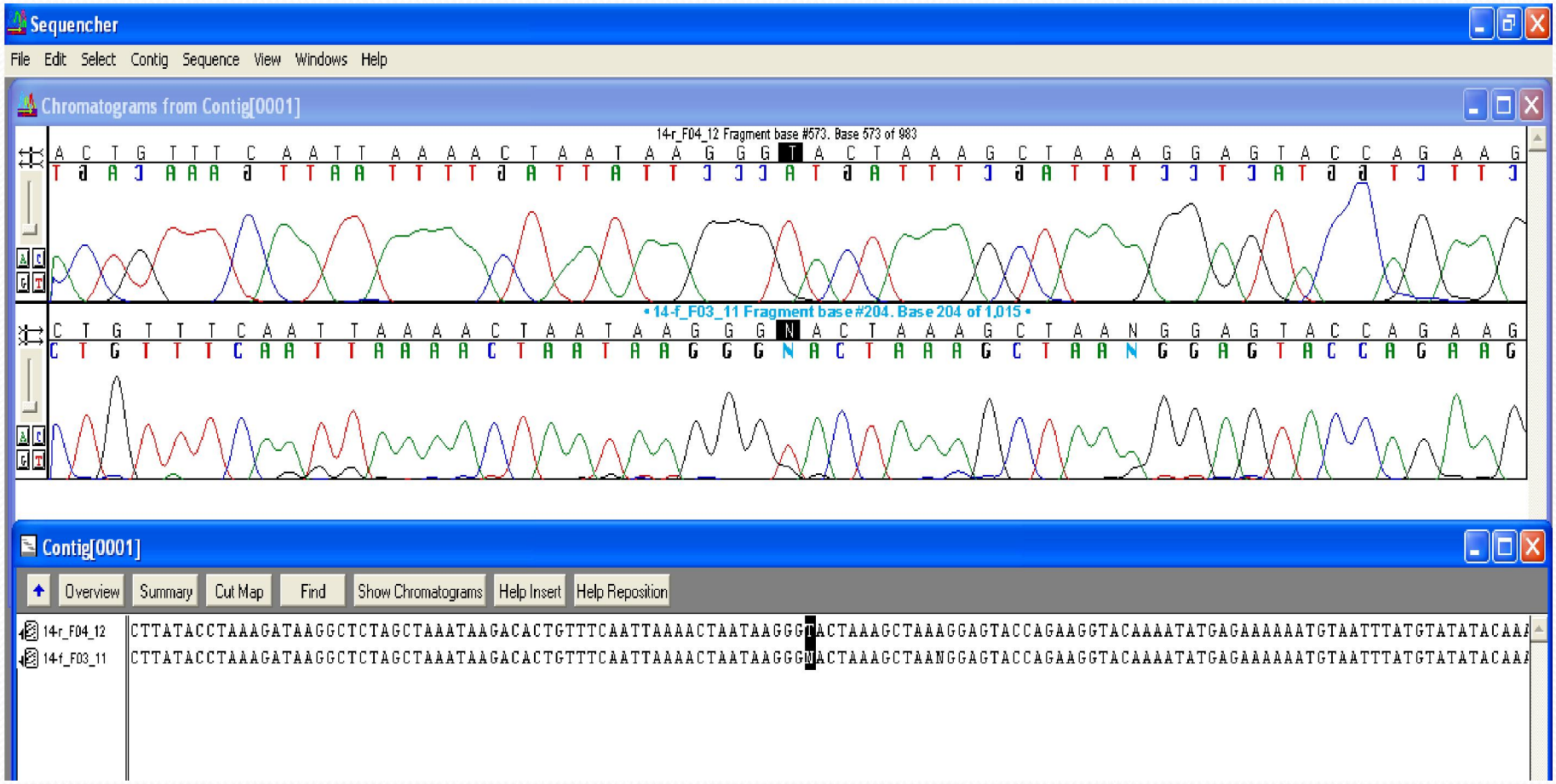
Good chromatogram!



Bad chromatogram…



Alignment of the 2 sequences obtained using the Forward and the Reverse primers on the same PCR amplification product



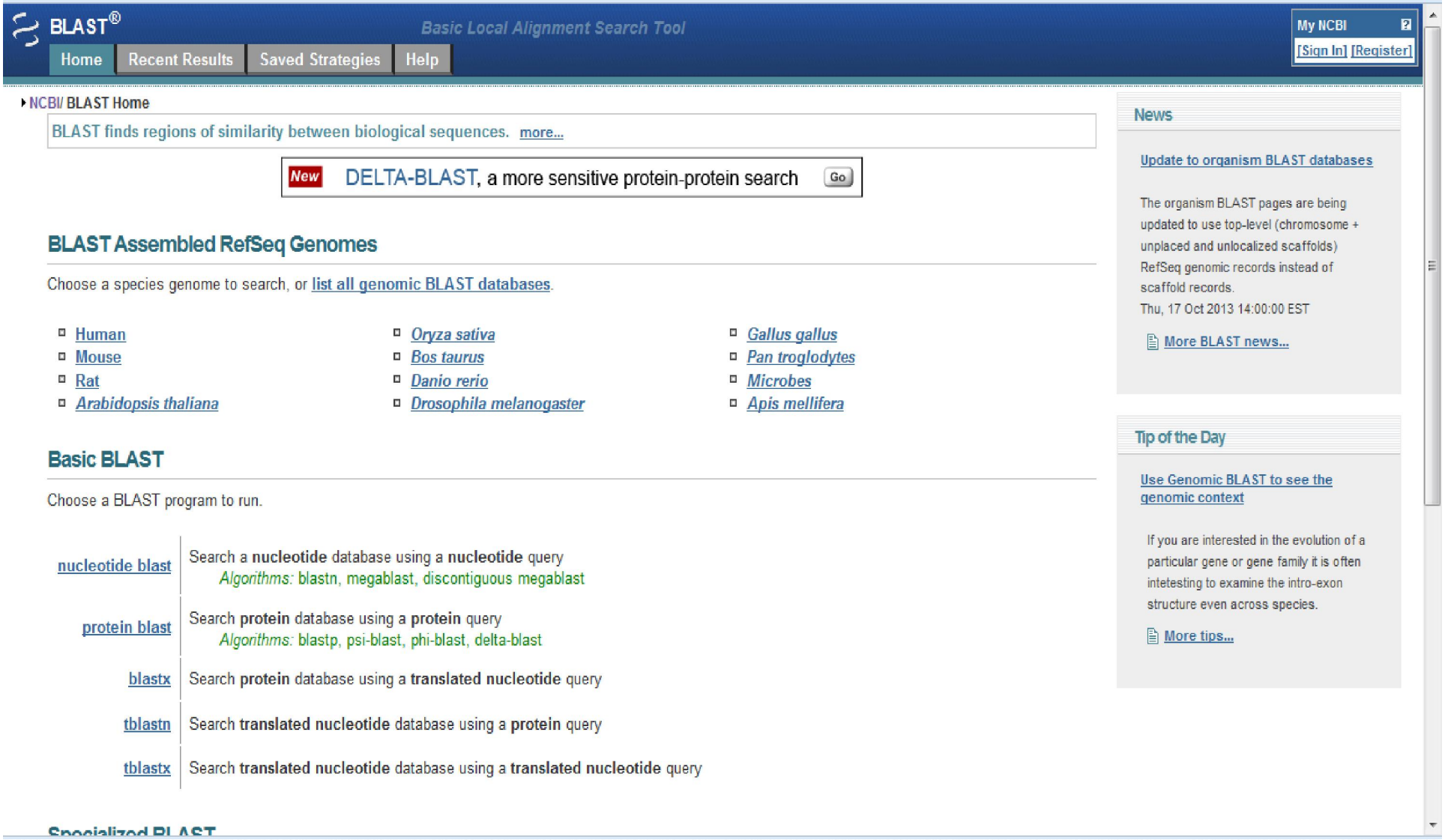
FASTA DNA sequence file

>gi|183013890|gb|EU427473.1|
ATGGAAGGTATTATTAACTTTCACCATGATTTAATGTTTTTTTTTAATTATGATTACTGTTTTGTTTTTGGA
TGTTATTAGAGTTATTACTCTTTTTGATGAAAAAAAAAATAAAATTCCTTCAACGGTAGTACATGGCGCT
ACTATTGAAATTATTGGACATCTATTCCAGCTTTAATTTTATTAGTTGTTGCAGTACCATCTTTTGCTTTA
TTATATTCAATGGATGAGGTAATTGATCCAATTATTACATTAAAAGTAATTGGTAGTCAATGGTATTGGAG
TTATGAATATTCTGATAATTTAGAATTTTCTGATGAACCTTTAATTTTTGATAGTTACATGATACAAGAAG
ATGATTTAGCAATAGGTCAATTTAGAGTTTTAGAAGTAGATAATCGTGTAGTTGTACCAACAAATAGTCAT
ATTAGAGTATTAATTACCGCATCAGATGTTTTACATTCATGGGCTATTCCTTCATTAGGTATTAAATTAGA
TGCATGTCCTGGACGTTTAAATCAAACATCAATGTTTATTAAAAGAGAAGGTGTTTTTTATGGACAATGT
AGTGAAATTTGTGGAGTAAATCATGGATTTATGCCTATTGTTGTAGAAGCTGTTTCATTAGAAGATTATTT
AACTTGGTTAAAAAATAAAATCAATTTTGATTTTAATGTATAATGATTTTAATGTATAATTAAATTAATTT
TATGGTATTTAAAATCATGGGTGTAATTTGTTTAATATTATTATTATTTACAGATATTAAACAAATTATATA
TAAAATCAAACAATTTTTTAATAAATAAAAAAATATTATCAATGATAATATAAATTAAAAAAACCAACGC
TTTTTTTAATTAAAAAAATATATAATTTTGCATTTAAATTAAATTTTAAAAATATTCAAAA

NCBI BLAST

- **B**asic **L**ocal **A**lignment **S**earch **T**ool
- **http://blast.ncbi.nlm.nih.gov/Blast.cgi**
- Tool to compare your sequence to all other sequences held in the National Center for Biotechnology Information (NCBI) Genbank Database
- **http://www.ncbi.nlm.nih.gov/genbank/**
- Finds the sequence that is most similar
- If you are lucky you get a direct hit and your sequence matches perfectly allowing you to ID your specimen

BLAST


BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]
NCBI/ BLAST Home
BLAST finds regions of similarity between biological sequences. more...
New DELTA-BLAST, a more sensitive protein-protein search Go
BLAST Assembled RefSeq Genomes
Choose a species genome to search, or list all genomic BLAST databases.
Human
Mouse
Rat
Arabidopsis thaliana
Oryza sativa
Bos taurus
Danio rerio
Drosophila melanogaster
Gallus gallus
Pan troglodytes
Microbes
Apis mellifera
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, delta-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
News
Update to organism BLAST databases
The organism BLAST pages are being updated to use top-level (chromosome + unplaced and unlocalized scaffolds) RefSeq genomic records instead of scaffold records.
Thu, 17 Oct 2013 14:00:00 EST
More BLAST news...
Tip of the Day
Use Genomic BLAST to see the genomic context
If you are interested in the evolution of a particular gene or gene family it is often intetesting to examine the intro-exon structure even across species.
More tips...

Genbank

GenBank Overview

What is GenBank?

GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*. 2013 Jan;41(D1):D36-42). GenBank is part of the International Nucleotide Sequence Database Collaboration , which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

The complete release notes for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release.

An example of a GenBank record may be viewed for a *Saccharomyces cerevisiae* gene.

Access to GenBank

There are several ways to search and retrieve data from GenBank.

- Search GenBank for sequence identifiers and annotations with Entrez Nucleotide , which is divided into three divisions: CoreNucleotide (the main collection), dbEST (Expressed Sequence Tags), and dbGSS (Genome Survey Sequences).
- Search and align GenBank sequences to a query sequence using BLAST (Basic Local Alignment Search Tool). BLAST searches CoreNucleotide, dbEST, and dbGSS independently; see BLAST info for more information about the numerous BLAST databases.
- Search, link, and download sequences programatically using NCBI e-utilities .

GenBank Data Usage

The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted. NCBI is not in a position to assess the validity of such claims, and therefore cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in GenBank.

Confidentiality

GenBank Resources

GenBank Home

Submission Types

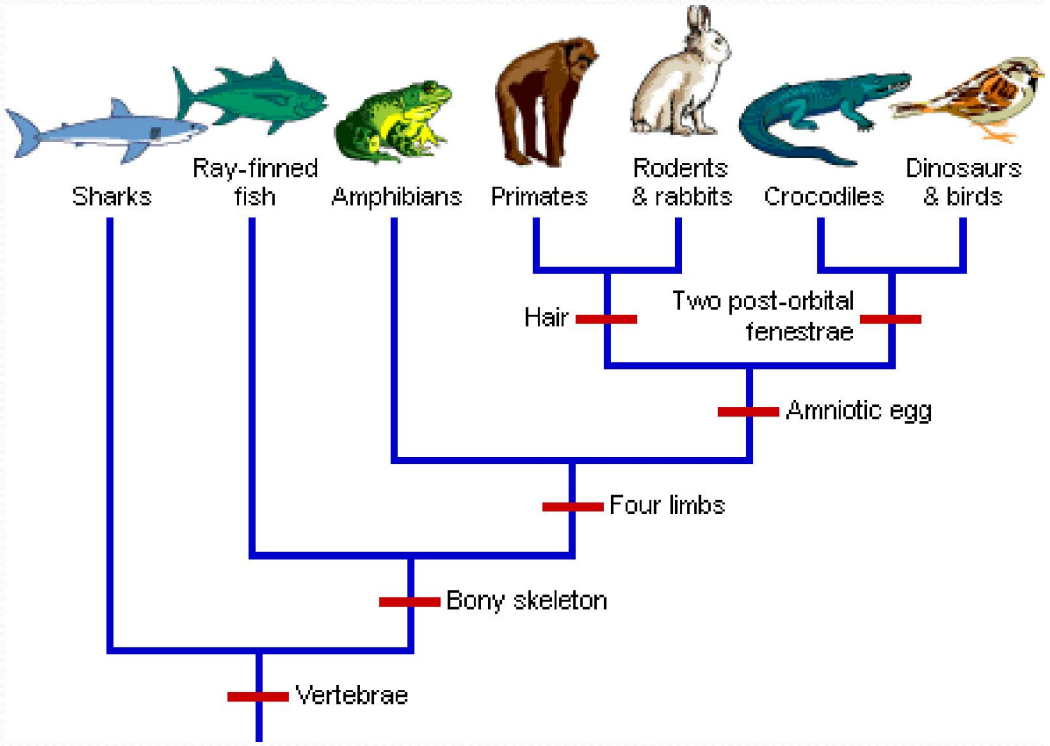
Submission Tools

Search GenBank

Update GenBank Records

Tree Building

- Can build trees based on shared features
- The more features shared, the more closely related things are likely to be
- E.g. morphology - number of toes, gill patterns etc.
- **Can also build trees based on DNA sequence similarity**



Evolution.berkeley.edu

Common ancestors

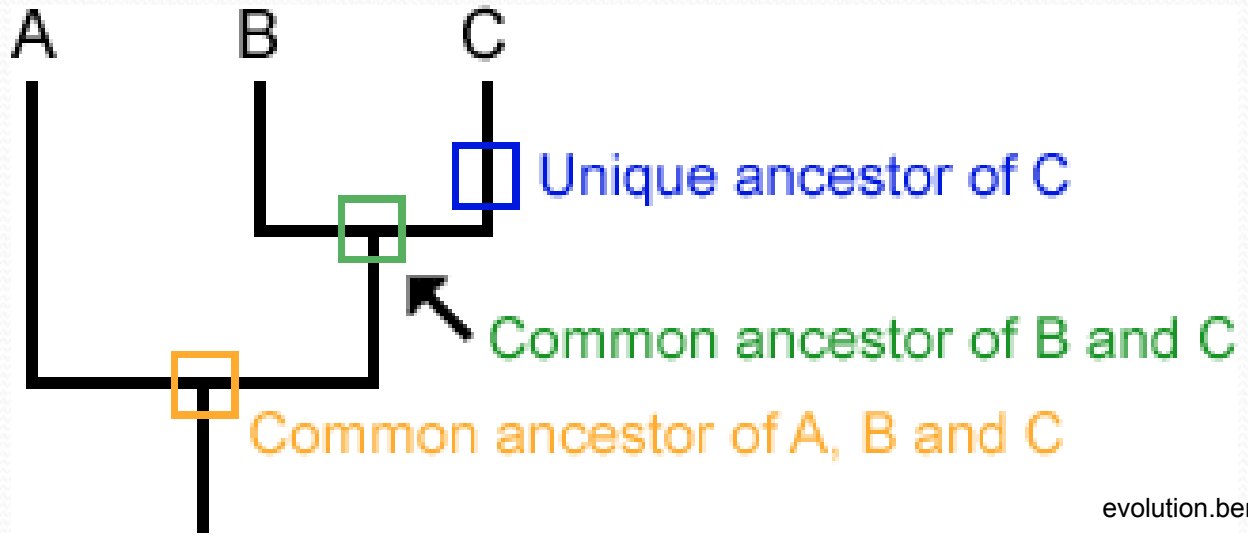
Molecular phylogenetic trees are constructed based on the differences between aligned sequences, and the relative time since individuals shared a common ancestor

Distantly related species

- long time since shared a common ancestor
- more time to accumulate mutations

Closely related species

- less time since common ancestor
- less mutations



evolution.berkeley.edu

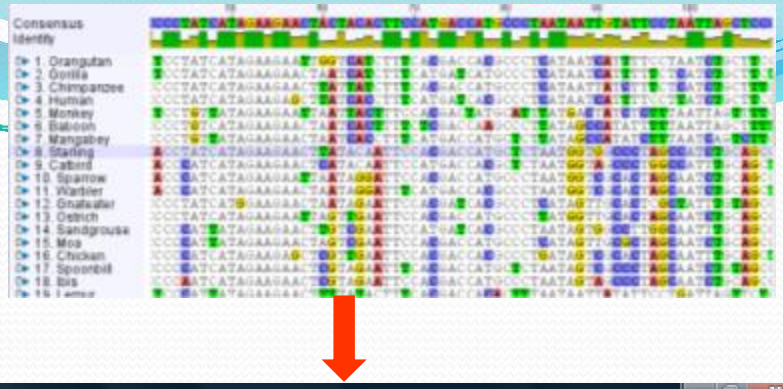
Sequence alignment

- Compare the sequences of the **SAME** area of DNA for **DIFFERENT** individuals
- Maximise the number of matches between sequences
- Easy for 2!

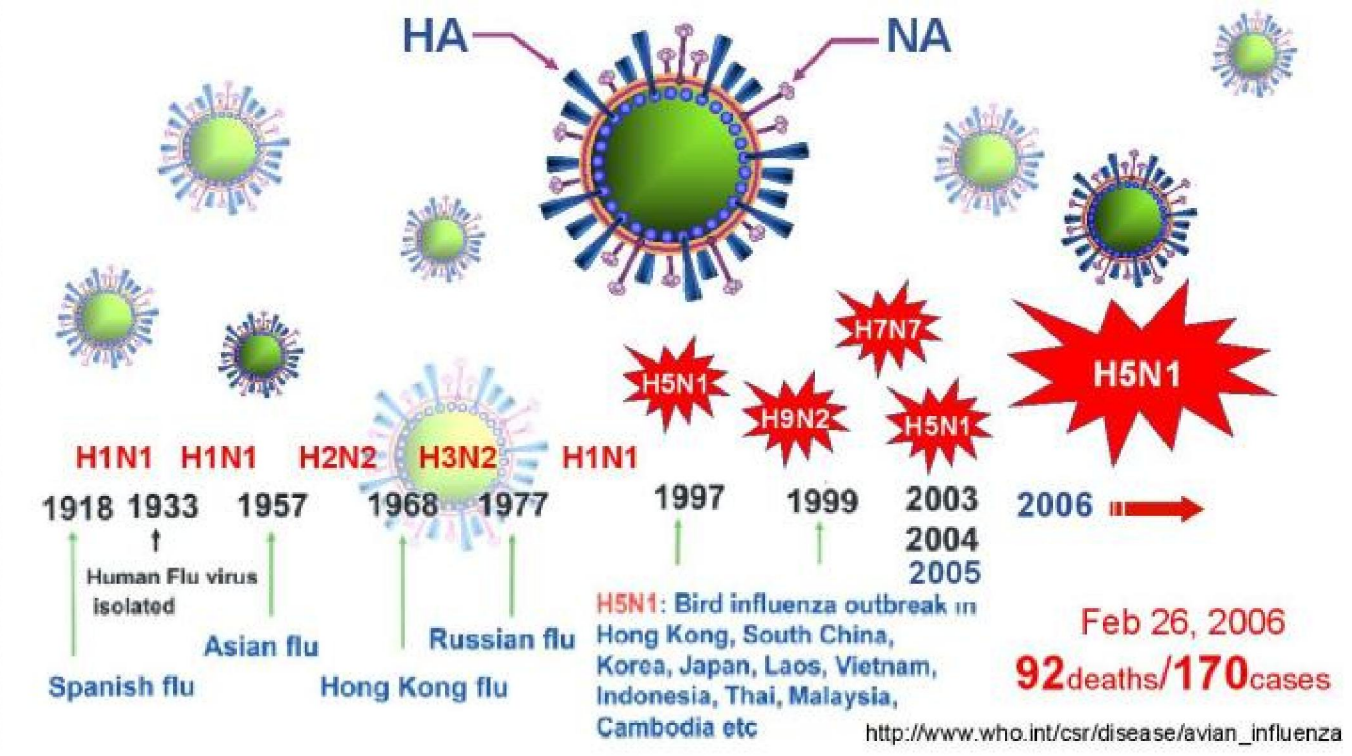
```
ACAGAC-GA
ACATACAGA
* * *  * *  * *
```

- Need specialist programs for more sequences

Tree building



Influenza



- Many different strains – H1N1
- Strain determined by two proteins on the surface
- Want to know which in order to make vaccines and track spread and where came from

Investigation

- Scenario: A California bay laurel tree has been identified in the Sierra Nevada that has the symptoms of *Phytophthora ramorum*. It is located in an area that was not previously known to be infected, and close to some commercial nurseries.

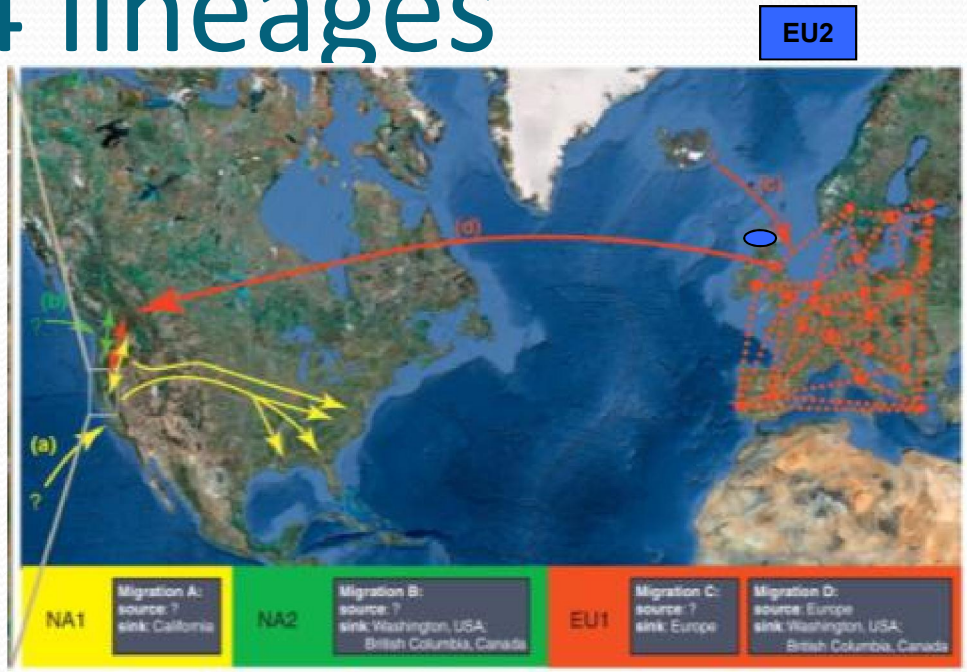
Supertopo.com

Questions

1. **Can we confirm that the tree is infected with *P. ramorum*?**
   - If a culture will grow then morphology can tell give us a very good idea that it is *P. ramorum*
   - If can't culture it, then can use PCR to detect any pathogen DNA, if present .
2. **If it is *P. ramorum*, how did it get here, and what lineage does it belong to?**
   - 4clonal lineages of *P. ramorum* exist, which are usually found in different areas. Detemining which lineage it is, provides us with information about where this infection originated from

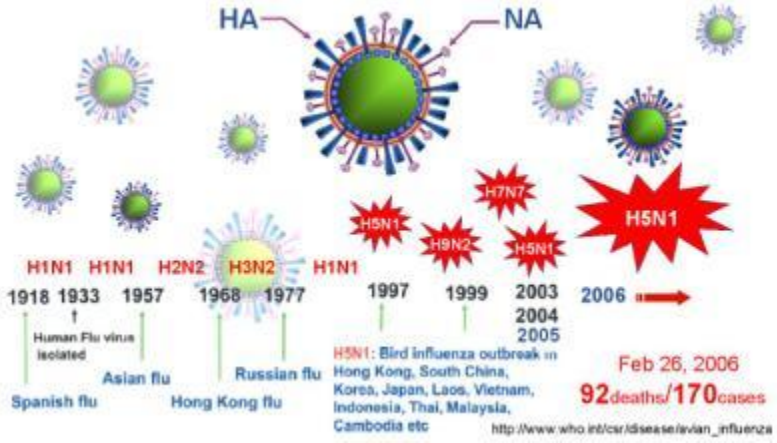
P. ramorum – 4 lineages

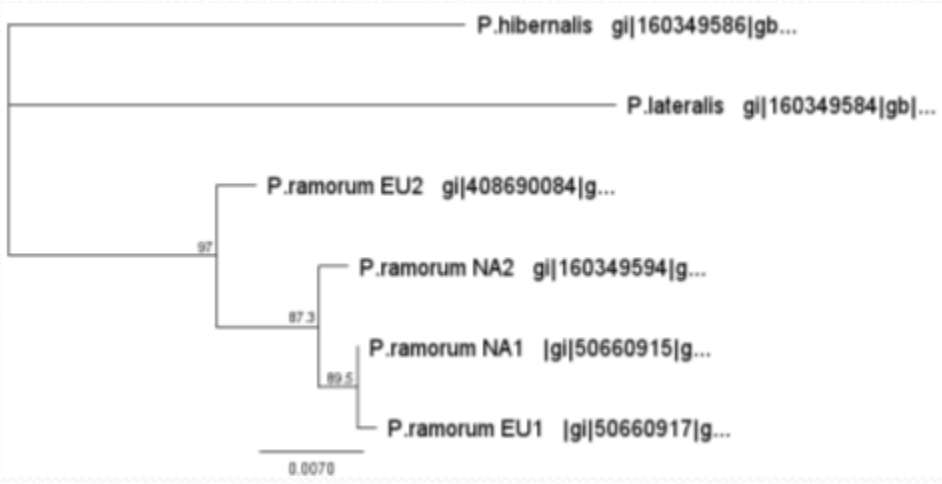
- **NA1** – mostly North American forests and nurseries
- **NA2** – mostly North American nurseries
- **EU1** – mostly in Europe
- **EU2** – Northern Ireland and Scotland



Van Poucke *et al* 2012 TRENDS in Microbiology

- Lineages analagous to strains of flu

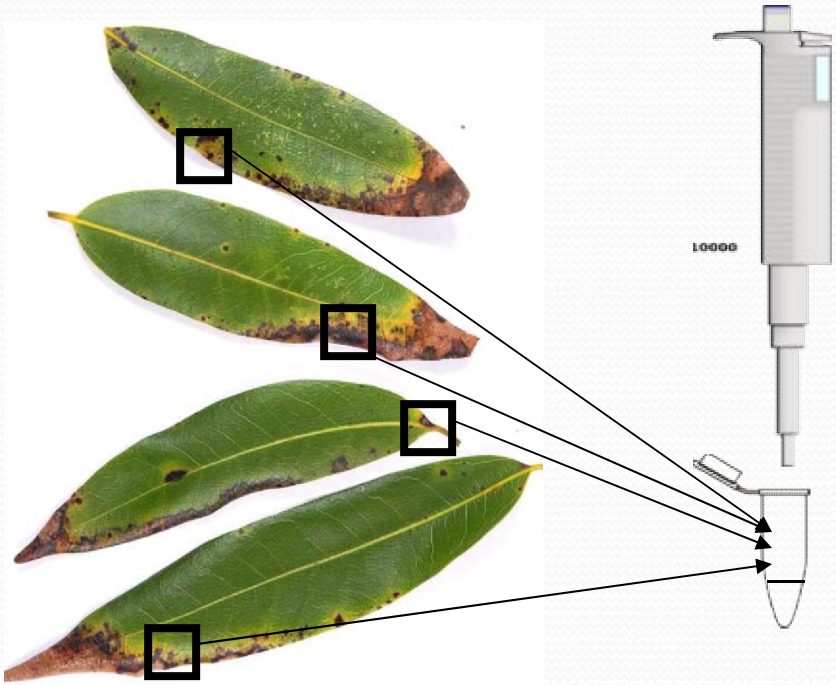


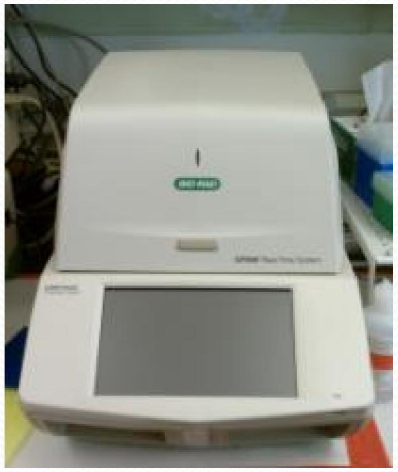


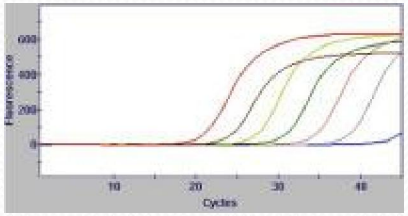
Plan

- **Q1: Is *P. ramorum* present on the symptomatic leaves?**
  - Use species specific DNA primers to detected any *P. ramorum* DNA on the symptomatic leaves
- **Q2: What lineage does the isolate belong to?**
  - Sequence a variable gene and compare it with other isolates of *P. ramorum*, with known lineages
  - Build a tree and see which of the other sequences the unknown groups with – i.e. which is it most similar to

Q1. Method: use species specific primers to identify P. ramorum DNA



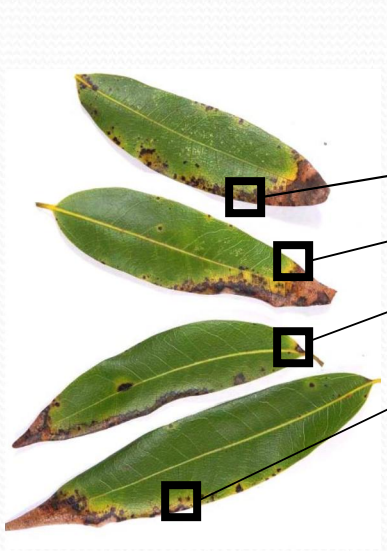


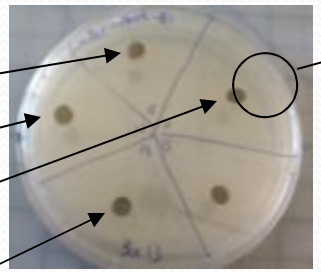


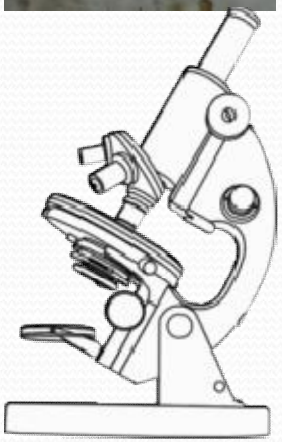
Sample leaf symptoms → Extract DNA → Amplify DNA with *P. ramorum* specific primers → Analyse the results!

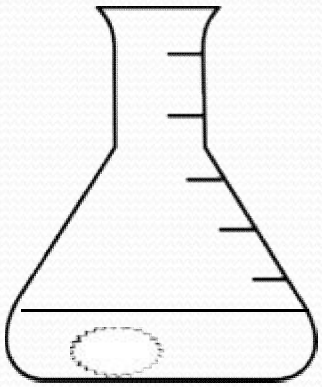
YES/NO

Q2: Culturing and sequencing



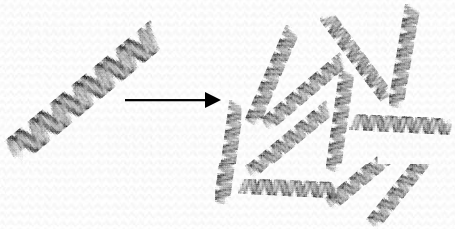


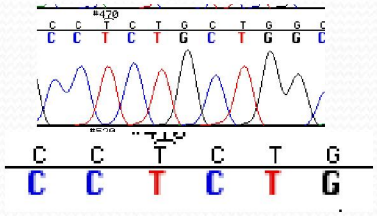


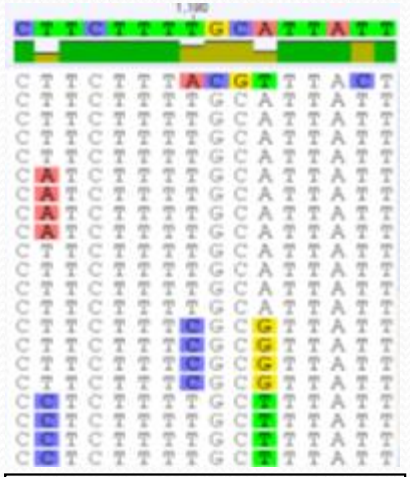


Sample leaf symptoms → Plate onto selective growth medium → Identify *P. ramorum* from morphology → Grow mycelium in liquid culture →

Leaf punch – plate – identify with morphology – grow in liq. Culture – DNA extract – PCR amplify – Sequence – BLAST for ID - align with other P. r. sequence of known clade





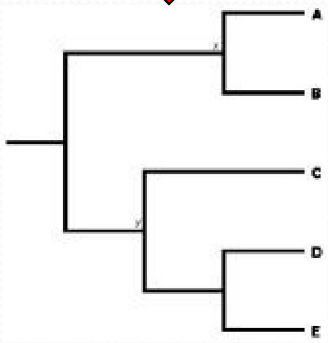


Extract DNA

Amplify specific region of DNA

Sequence DNA

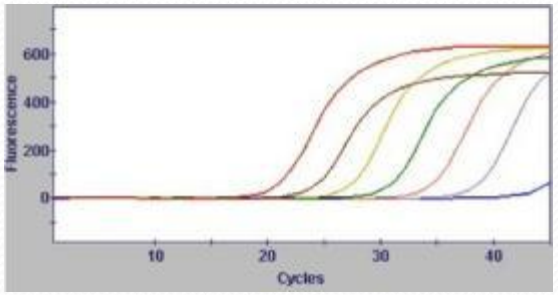
Align with other known DNA sequences



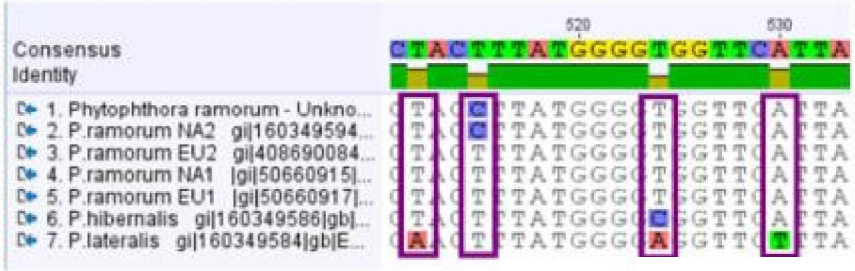
Build phylogenetic tree

Results

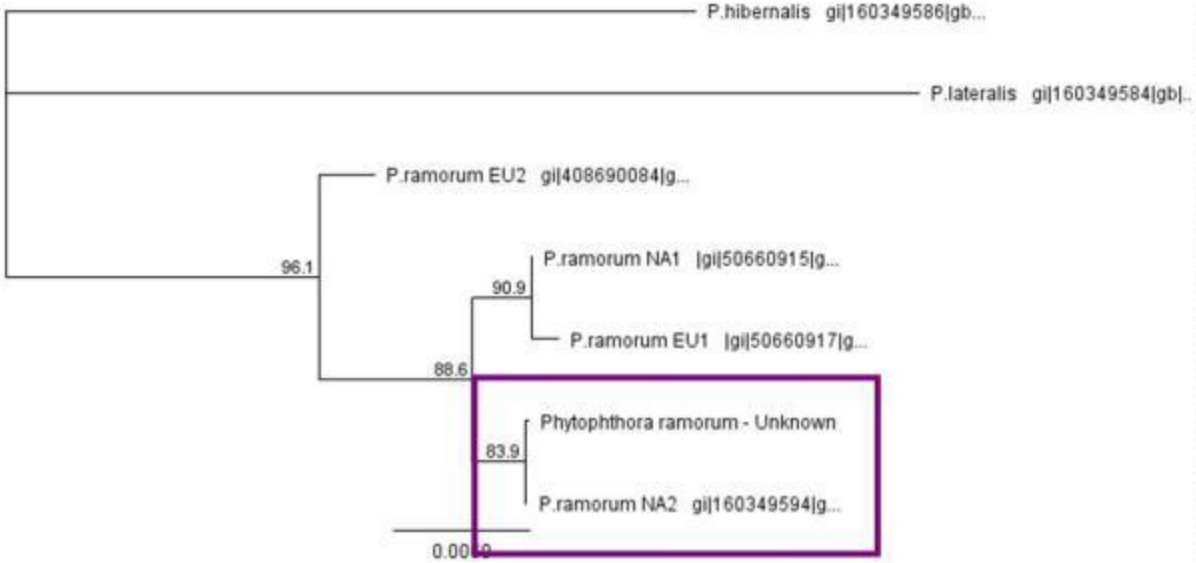
- Is *P. ramorum* present on the leaf?
  - Realtime PCR – PRESENCE/ABSENCE



- What lineage does the unknown sample group with?



New isolate is mostly likely to be NA2



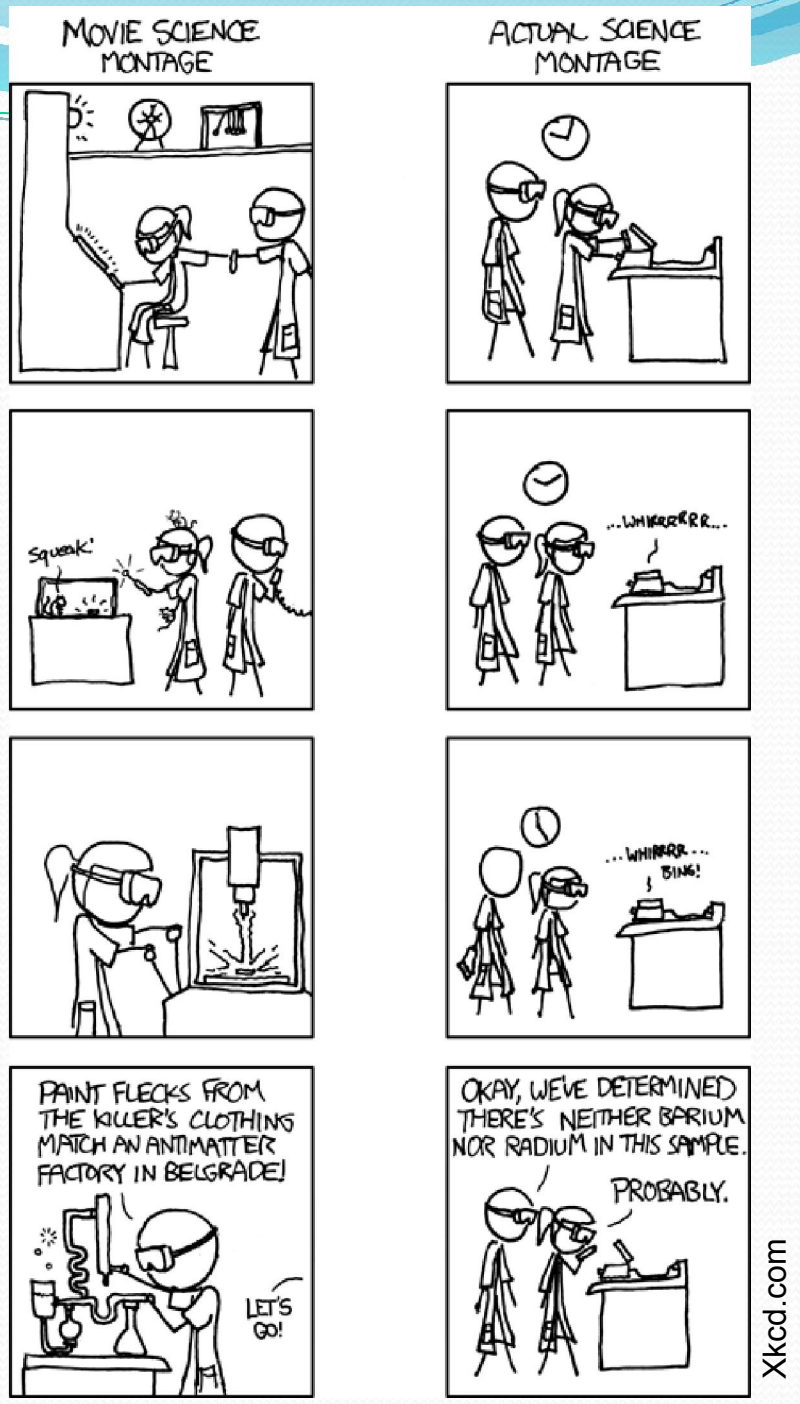
Other considerations

- Environmental samples add complexity to analysis
  - Different species present – 'dirty samples'
    - not necessarily the ones you are looking for
  - Pathogen can only survive on plants for a certain amount of time
    - Need to culture quickly to get a 'live' pathogen
  - DNA can degrade
    - Need to keep samples cool and process quickly
- How reliable are the results?
  - Need to include controls
  - Consistent sample handling – field collection to lab work
  - Must be reproducible

Diagnostics summary

- Need to be rigorous
- Use different techniques depending on the questions asked
- Time consuming
- Can be complicated by the nature of the sample

- 50% designing and setting up a good assay
- 50% controls and good handling of samples to ensure sample integrity



Thanks!