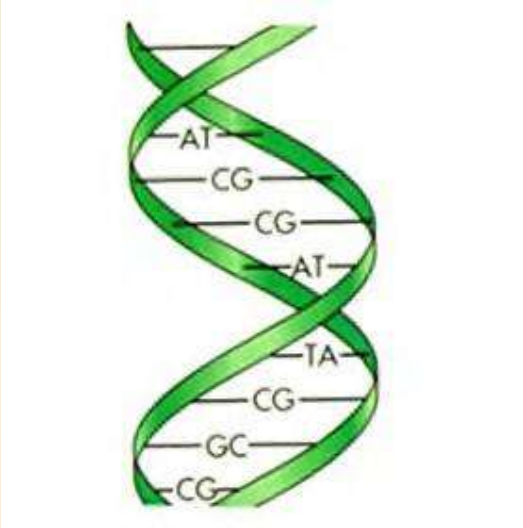
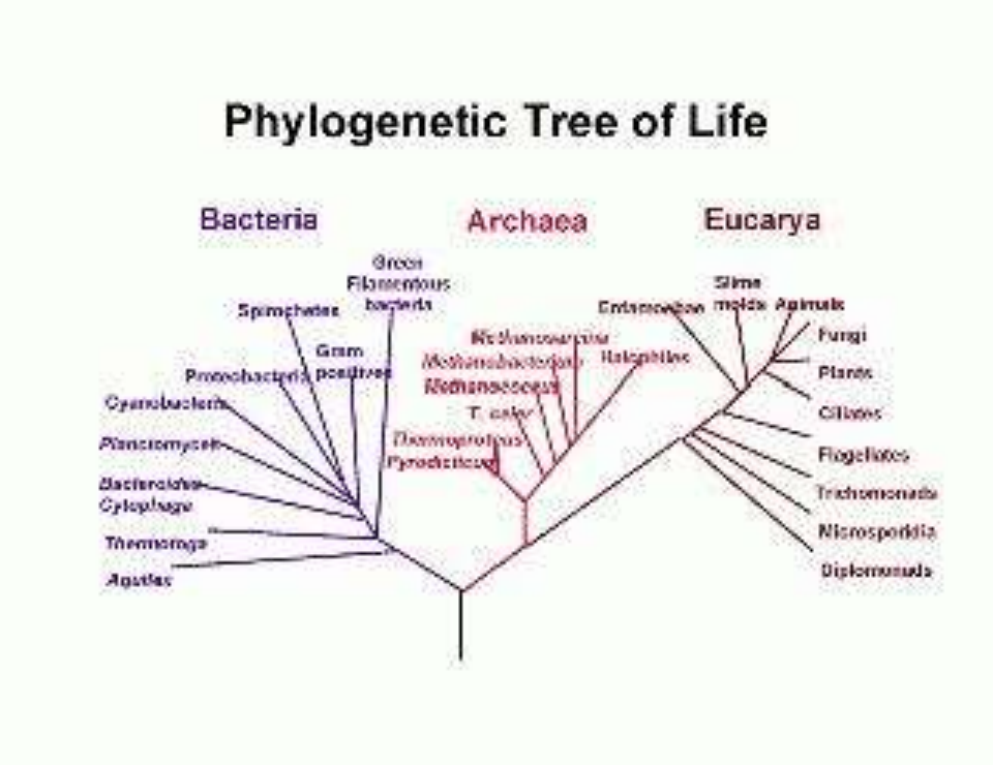
Molecular Phylogenies

```
Outgroup      AAGCTTCATAGGAGCAACCATTCTAATAATAAGCCTCATAAAGCC
Species A     AAGCTTCACCGGCGCAGTTATCCTCATAATATGCCTCATAATGCC
Species B     GTGCTTCACCGACGCAGTTGTCCTCATAATGTGCCTCACTATGCC
Species C     GTGCTTCACCGACGCAGTTGCCCTCATGATGAGCCTCACTATGCA
```





Phylogeny Basics

- Branching history of evolutionary lineages
- New branches arise via speciation
- Speciation occurs when gene flow is severed between populations
- Phylogenetic relationships depicted as a tree

Assumptions

Closer related organisms have more similar genomes.

Highly similar genes are homologous (have the same ancestor).

A universal ancestor exists for all life forms.

Molecular difference in homologous genes (or protein sequences) are positively correlated with evolution time.

Phylogenetic relation can be expressed by a dendrogram (a “tree”) .

Phylogenetic Characters

- Morphology
- Secondary chemistry
- Cytology
- Allele frequencies
- Protein sequences
- Restriction sites
- DNA sequences

} "Molecular" data

Types of molecular data

Antibodies (Serology, Immunology)

Proteins – sequences

DNA restriction sites - RFLPs

DNA sequences

Microsatellites - DNA regions w/tandem repeats

RAPDs - Random Amplification of Polymorphic DNA

AFLPs - Amplified Fragment Length Polymorphism

Allozymes - different forms of proteins

Why Use Molecular Data?

Many more molecular characters available for analysis than morphological ones.

Identity is easier to define: ATCG vs. whether a flower color is pink or white.

Fewer problems with analogous characters and convergence.

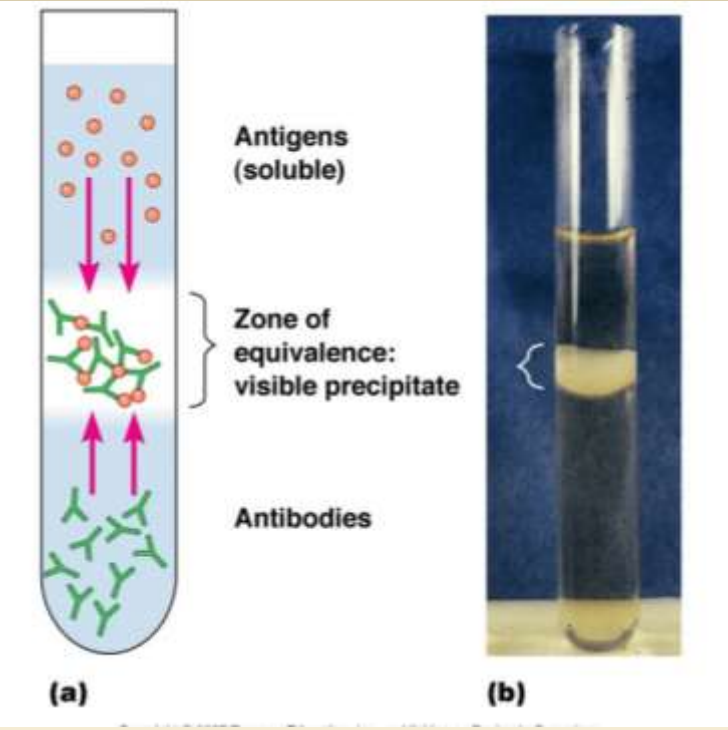
Still subject to homoplasy: alignment errors, reversals, and base saturation.

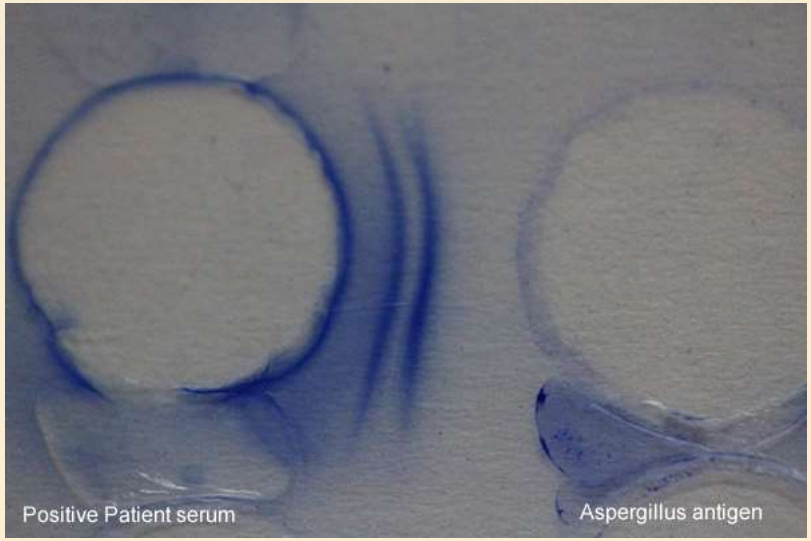
Molecular data vs. Morphology / Physiology

• Strictly heritable entities	• Can be influenced by environmental factors
• Data is unambiguous	• Ambiguous modifiers: “reduced”, “slightly elongated”, “somewhat flattened”
• Regular & predictable evolution	• Unpredictable evolution
• Quantitative analyses	• Qualitative argumentation
• Ease of homology assesment	• Homology difficult to assess
• Relationship of distantly related organisms can be inferred	• Only close relationships can be confidently inferred
• Abundant and easily generated with PCR and sequencing	• Problems when working with micro-organisms and where visible morphology is lacking

Serology - using antibodies in blood serum

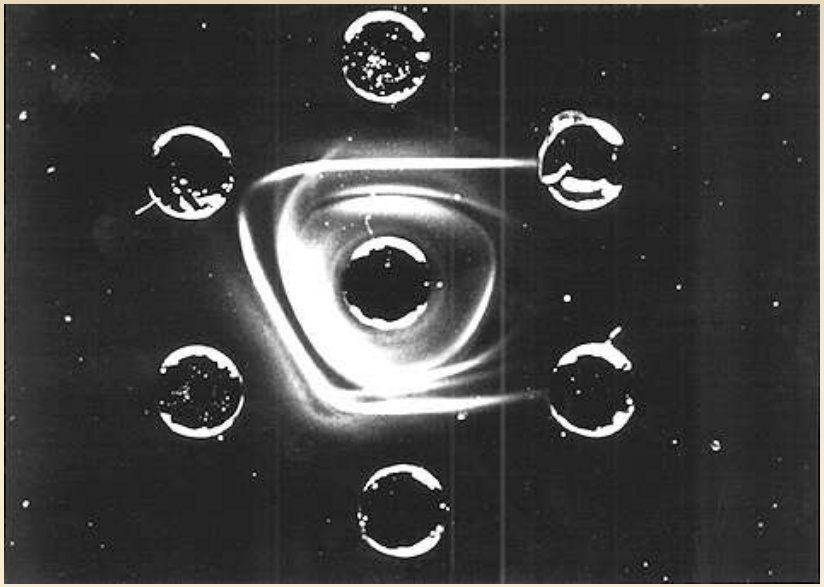
- Protein from Species A injected in rabbit.
- Rabbit makes antibodies to Species A
- Extract blood from rabbit, isolate serum
- Add protein antigen from different Species B
- Strength of antigen-antibody reaction used to calculate Immunological distances





Immunology was first molecular data used in phylogenetics

Blood Immunity and Blood Relationship (1904)



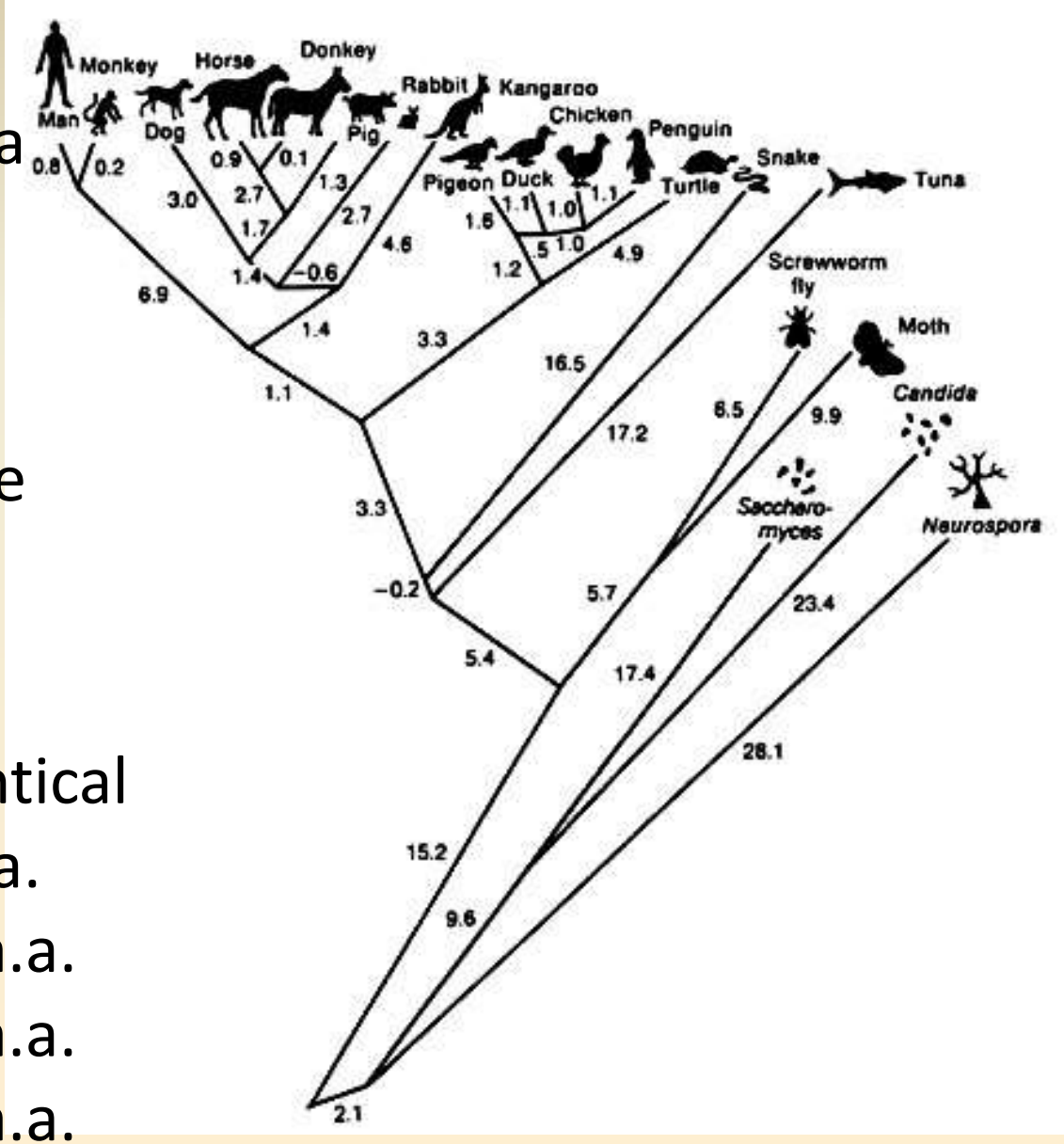
George Nuttall 1862-1937

serum antibodies used to identify different kinds of blood

"If we accept the degree of blood reaction as an index of the degree of blood-relationship within the Anthropoidea, then we find that the Old World apes are more closely allied to man than are the New World apes, and this is exactly in accordance with the opinion expressed by Darwin."

Protein Sequencing - Cytochrome C amino acids

1980s The a.a. sequence has been determined in a large number of taxa. Seems to change very slowly. The more closely related the organisms are the more a.a.s they have in common.



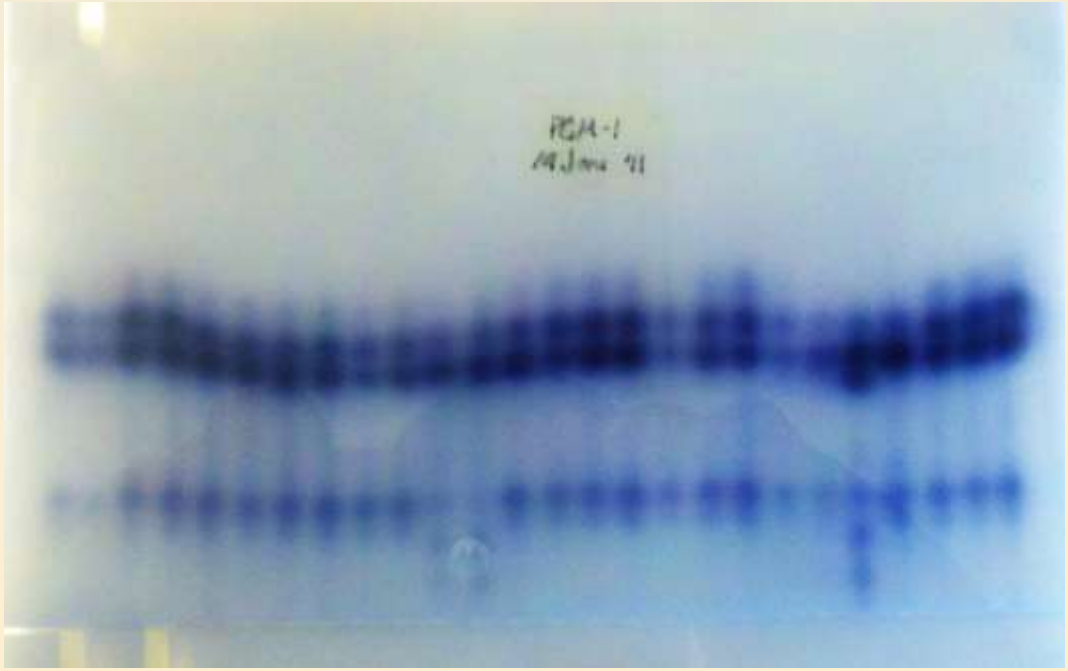
Man Chimp	identical
Man Rhesus	1 a.a.
Man Dog	13 a.a.
Man Rattlesnake	20 a.a.
Man Tuna	31 a.a.

Allozymes:

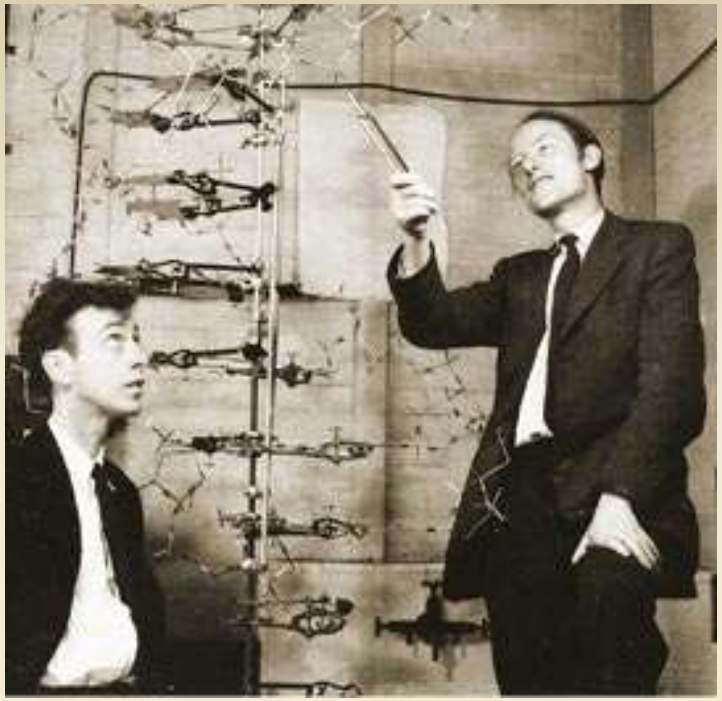
Different alleles produce slightly different proteins which migrate differently on an electrically charged starch gel.

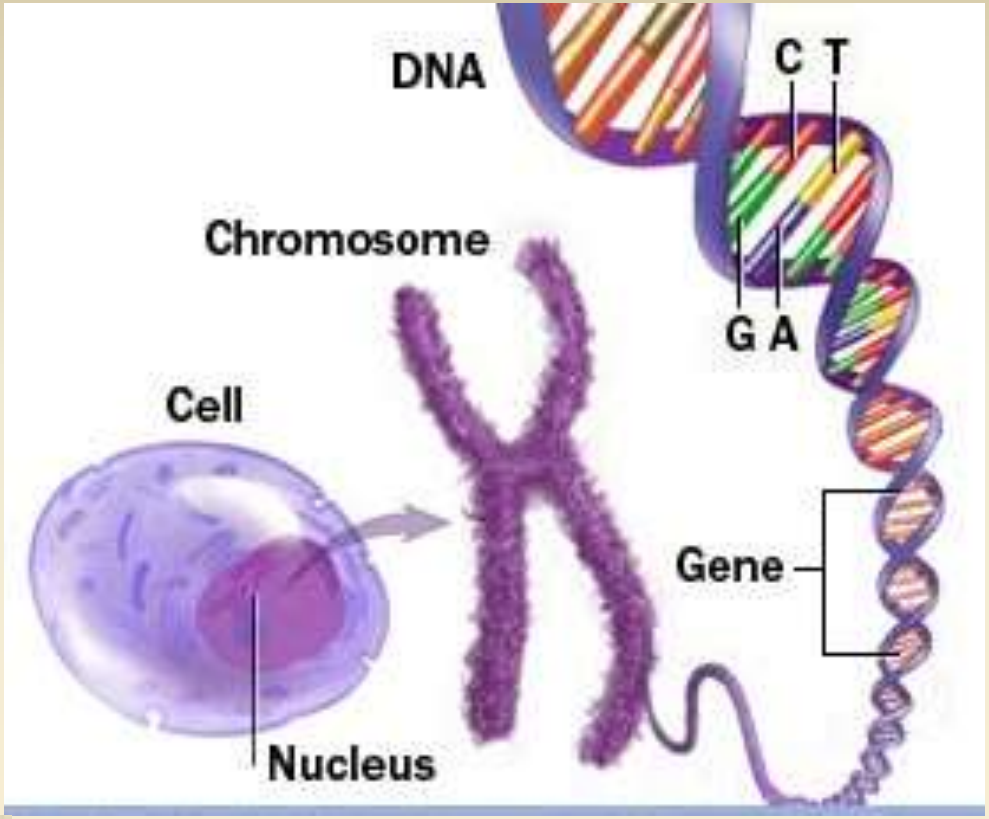
Data used to **measure genetic diversity, heterozygosity**, in populations.

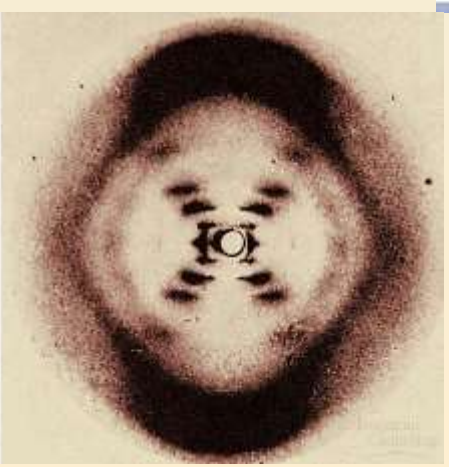
Used in the past frequently, now replaced by DNA methods.



Watson and Crick – 1953 - DNA Structure and Function







Rosalind Franklin,
X-ray diffraction of DNA

Landmarks of DNA Sequencing

1953 - Discovery of the structure of the DNA double helix.

1972 - Development of recombinant DNA technology, which permits isolation of defined fragments of DNA.

1977 - The **first complete DNA genome** to be sequenced is that of bacteriophage ϕX174 (F. Sanger).

1977 - Maxam and Gilbert "DNA sequencing by chemical degradation". Frederick Sanger, independently, publishes "DNA sequencing with chain-terminating inhibitors".

1983 - Kary Mullis invents **Polymerase Chain Reaction**

1987 - first automated sequencing machine, the model ABI 370.

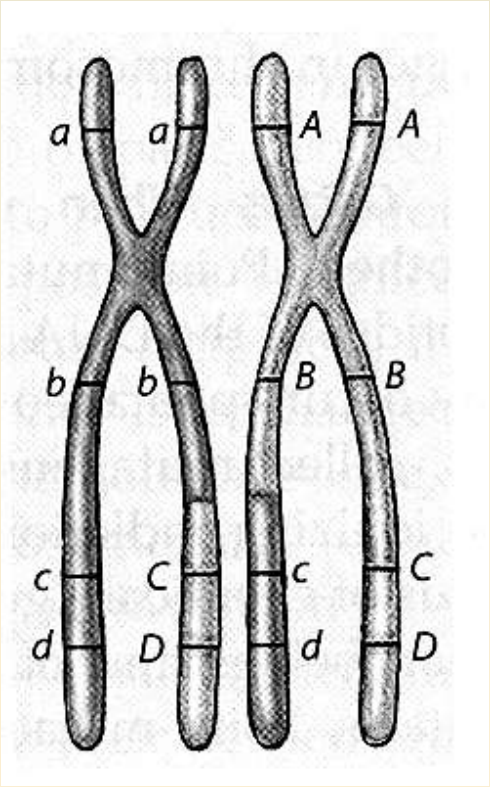
1995 - first bacterial genome sequenced, *Haemophilus influenzae*.

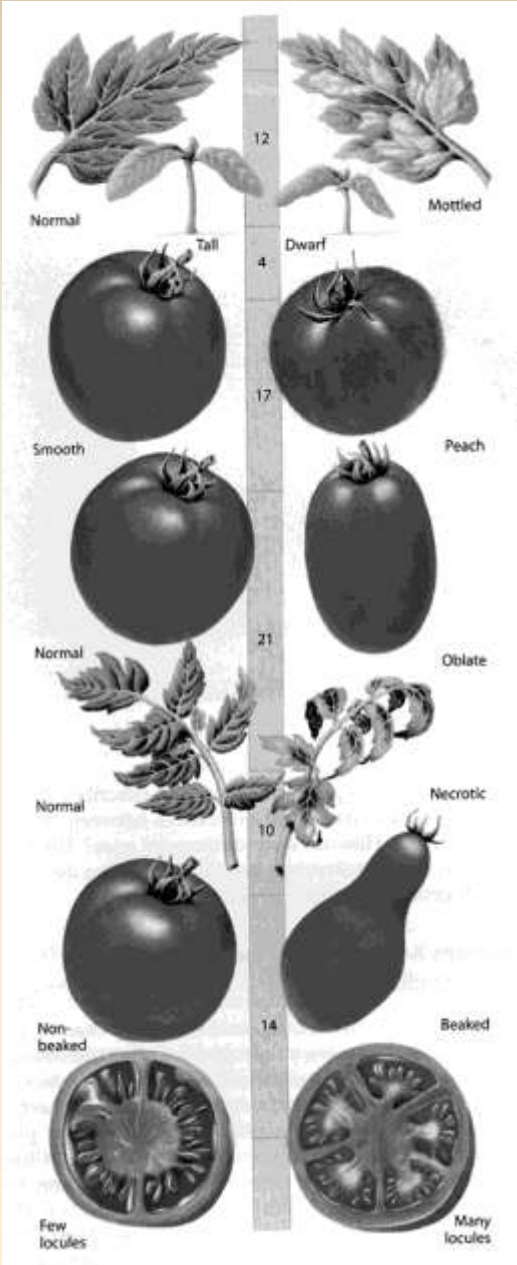
1999 - Complete sequence of a human chromosome (22) published.

2000 – first bead-based "next-generation" sequencing launched.

2004 - Complete sequencing of the human genome was finished

What are genes?

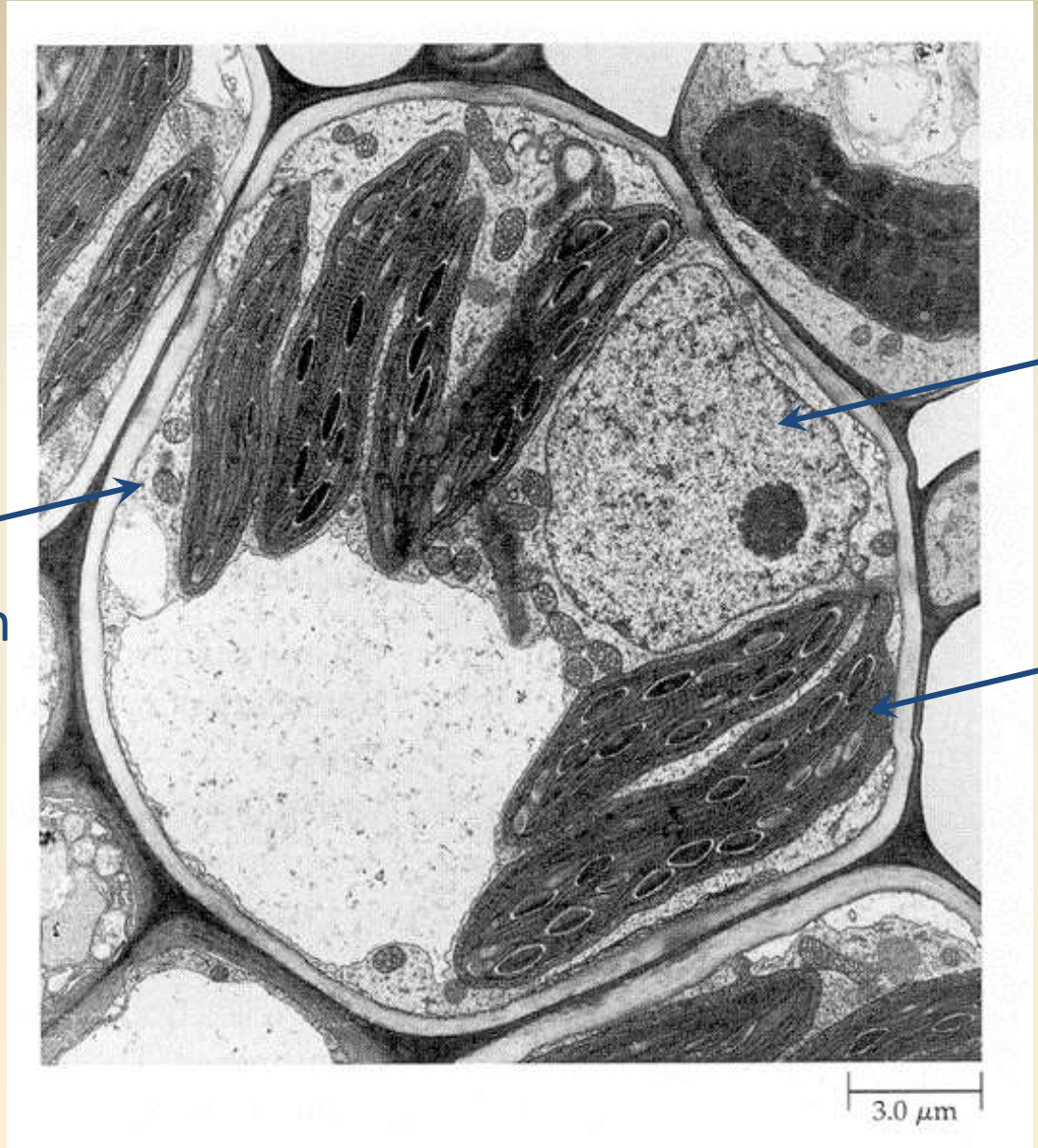




From Raven et al. (1999),
Biology of Plants

Genomes

- All of the genes within a cell are the genome
- Genes located in the nucleus are the nuclear genome
- Other genomes (organellar)
  - Mitochondrion: mitochondrial genome
  - Chloroplast: plastid genome



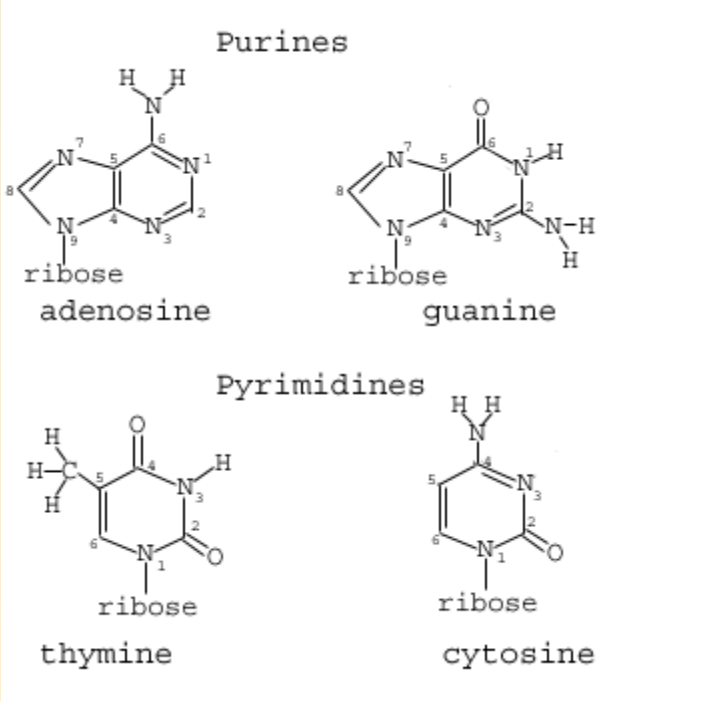
From Raven et al., 1999, *Biology of Plants*

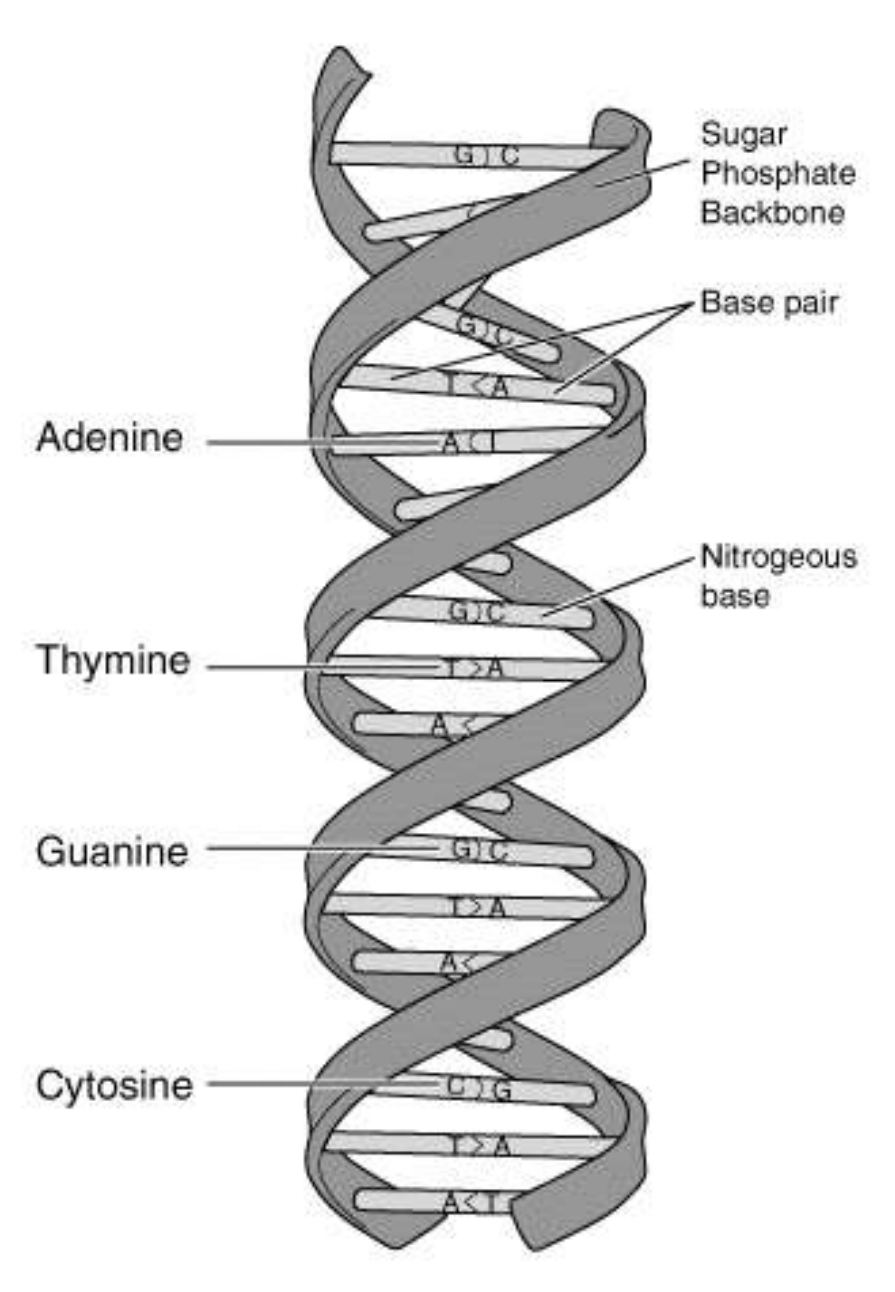
Comparison of Genomes

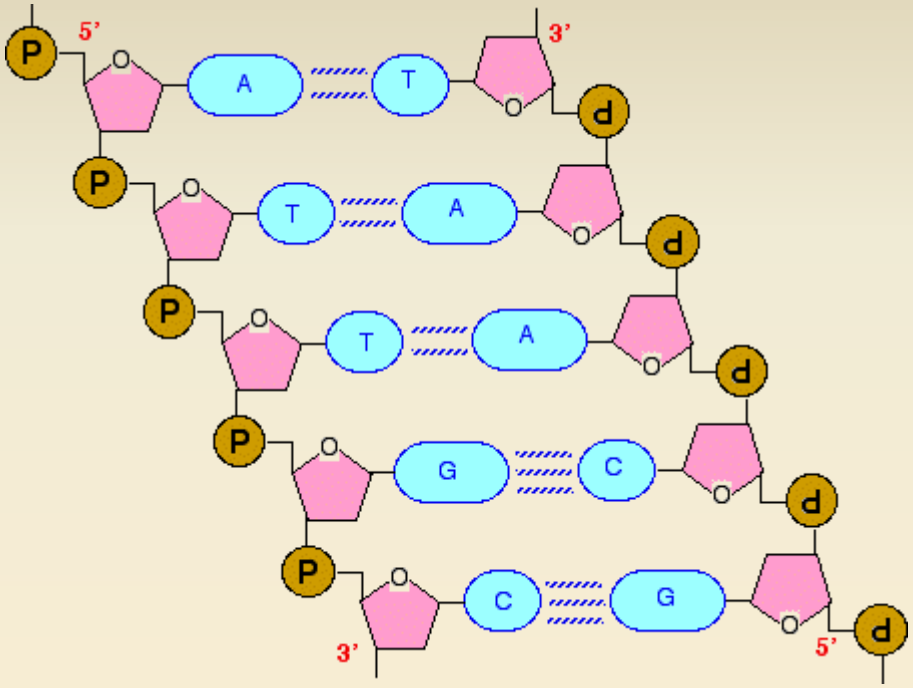
	Nuclear	Mitochondrial	Plastid
Size	Large	Small	Small
Number	Multiple	Single	Single
Shape of Chromosomes	Linear	Circular	Circular
Ploidy	Diploid	Haploid	Haploid
Inheritance	Biparental	Uniparental	Uniparental

Chemistry of Genes

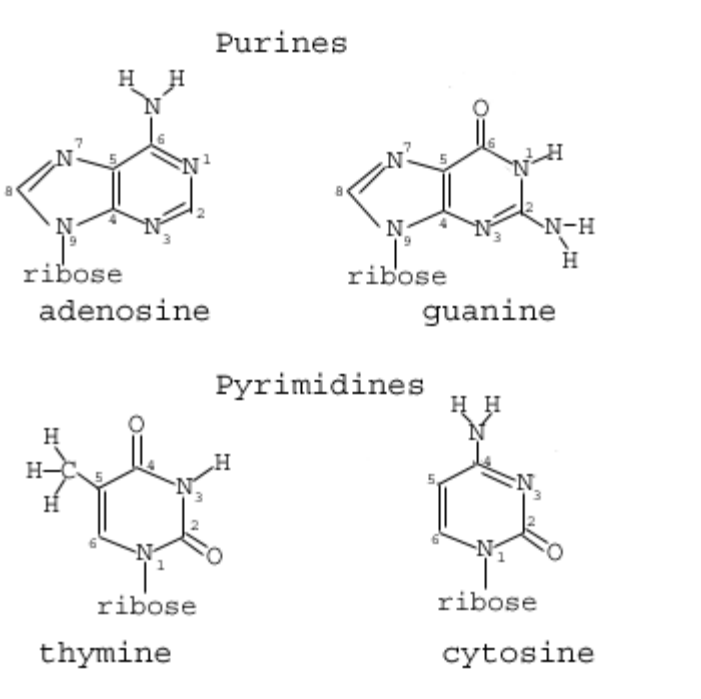
- DNA
- Parallel strands linked together
- Linear array of units called nucleotides
  - Phosphate
  - Sugar: deoxyribose
  - One of four bases
    - Adenine ("A")
    - Cytosine ("C")
    - Guanine ("G")
    - Thymine ("T")

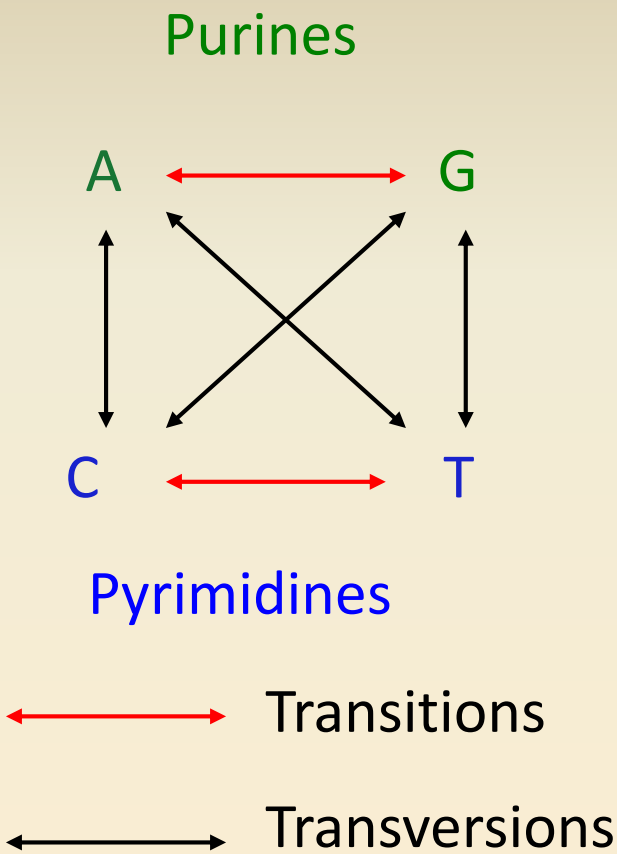



Sugar
Phosphate
Backbone
Base pair
Adenine
Nitrogeous
base
Thymine
Guanine
Cytosine


5'
3'
P
A
T
d
G
C

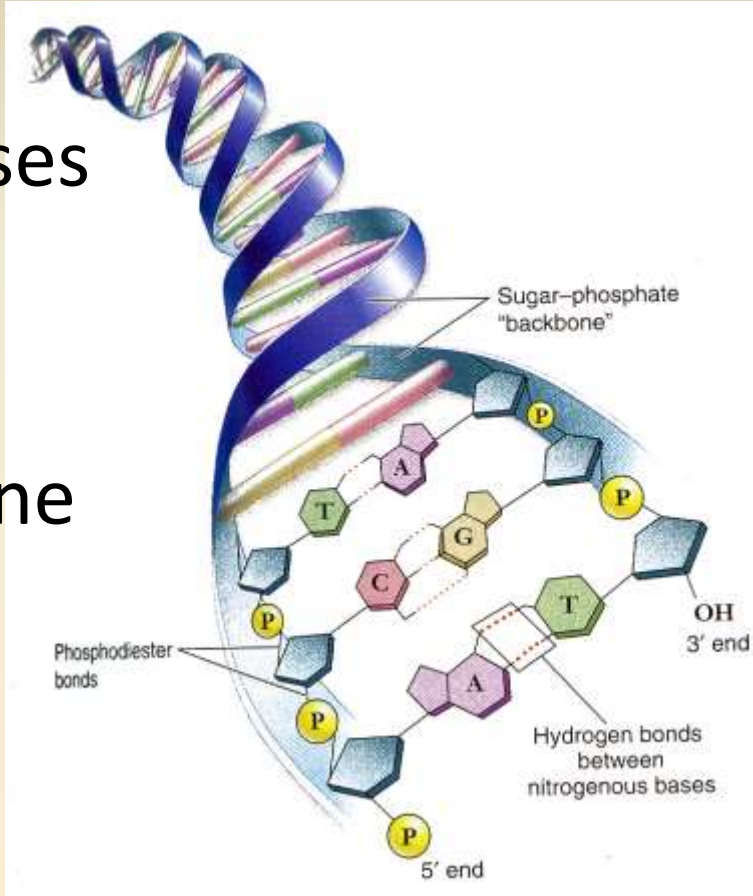
Transitions and Transversions





DNA Structure

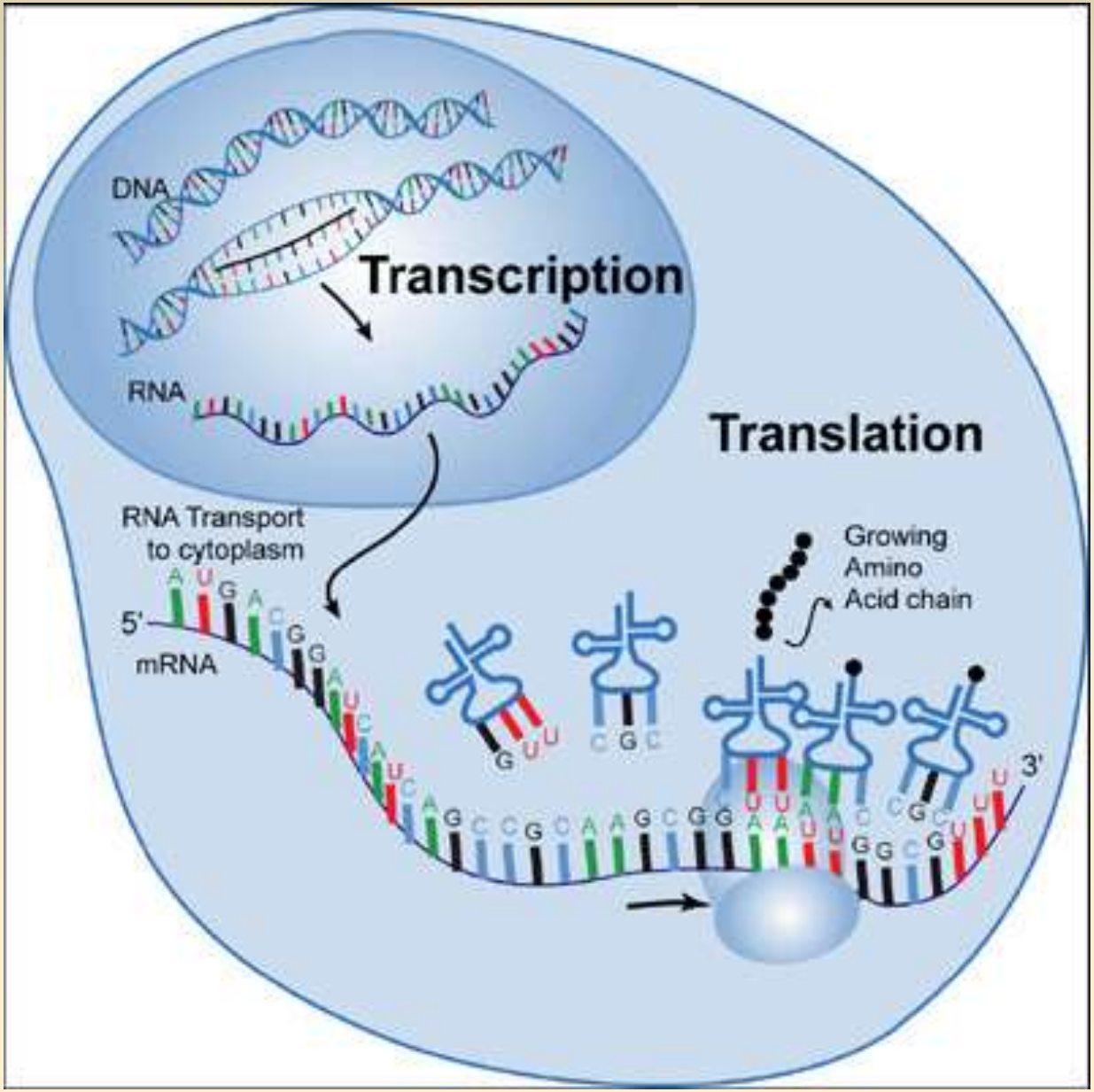
- Paired strands are linked by bases
  - A must bond with T
  - G must bond with C
- Each link is composed of a purine and a pyrimidine
  - A & G are purines
  - C & T are pyrimidines



DNA function

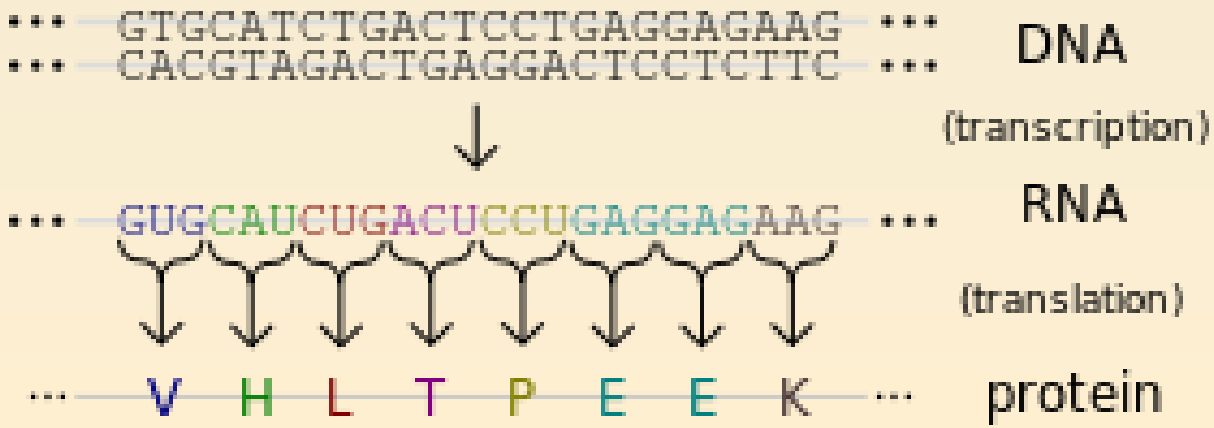
- DNA is the code for making proteins (and a few other molecules)
- Proteins are the structures and **enzymes that catalyze biochemical reactions** that are essential for the function of an organism
- DNA code is read and converted to protein in two steps
  - Transcription: DNA is copied to messenger RNA
  - Translation: messenger RNA is template for protein

The Central Dogma – DNA=>RNA=>Protein



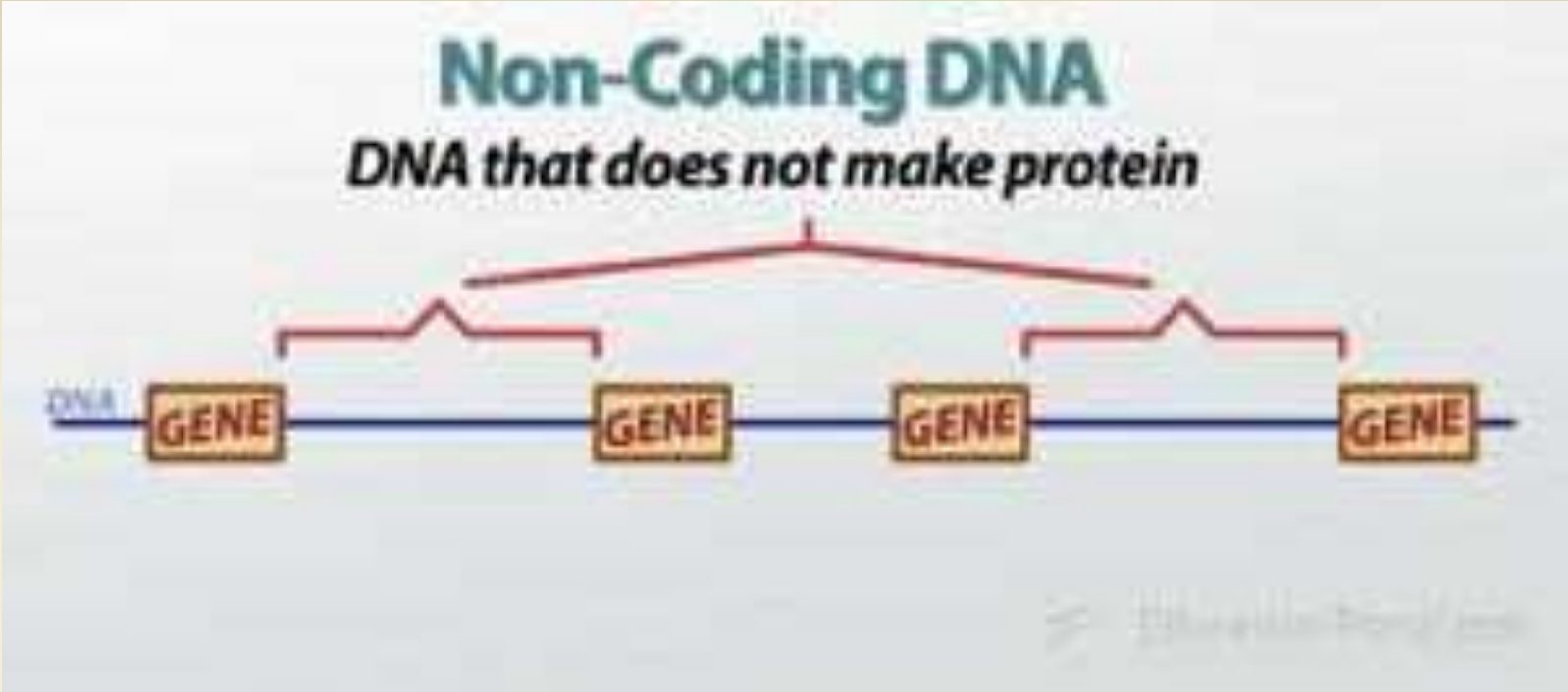
Codons and the Genetic Code

1st letter	Second Letter: U		Second Letter: C		Second Letter: A		Second Letter: G		3rd letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G



DNA functional classes

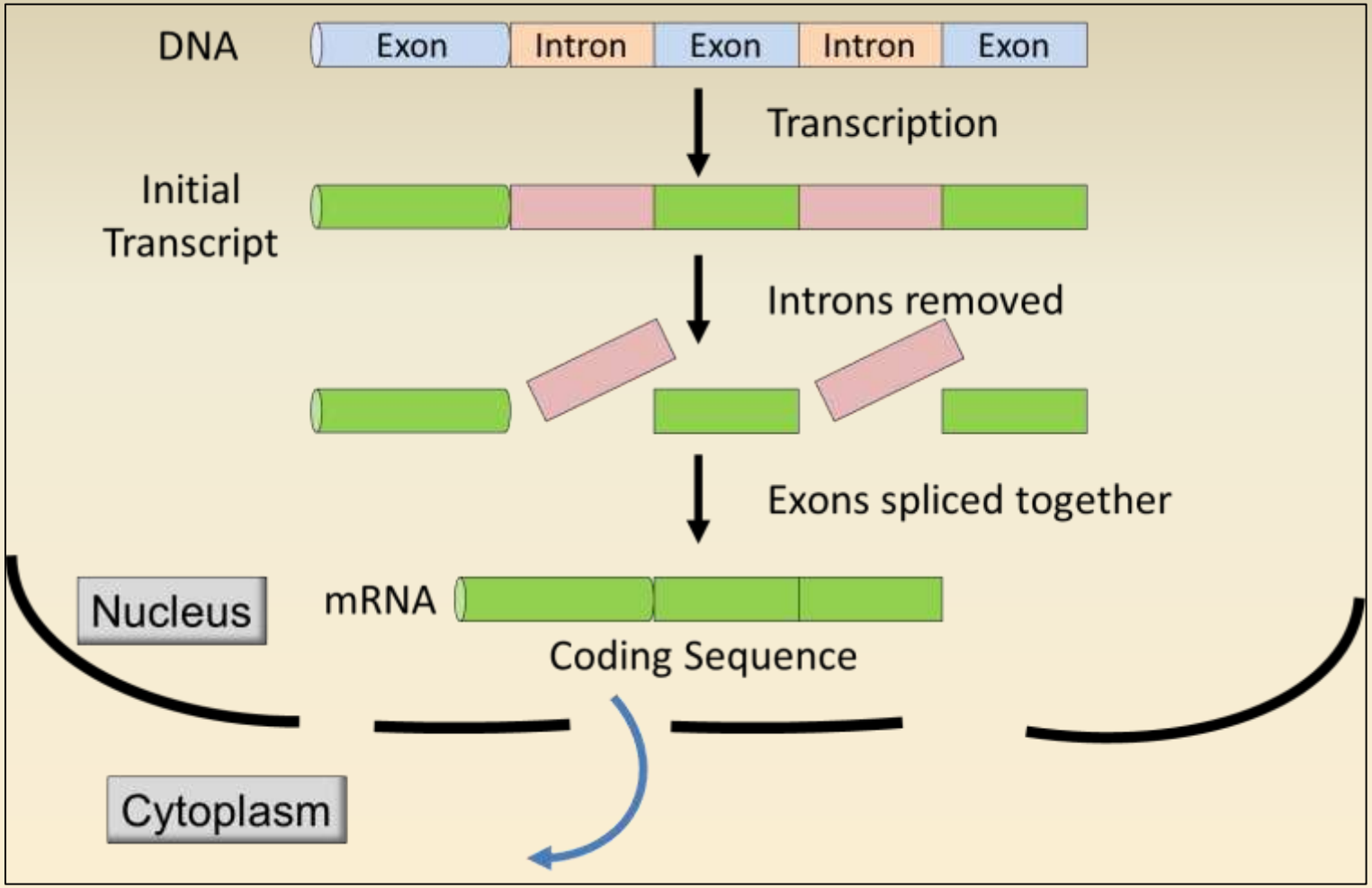
- Coding
  - Proteins (exons)
  - Ribosomes (RNA)
  - Transfer RNA
- “Non-coding”
  - Introns
  - Spacers



Intergenic Regions - Non-coding DNA regions Between Genes

“Spacer Regions”

Non-coding Regions within Genes - Introns



Exons – code for proteins

Introns - non-coding regions between exons, spliced out

How is DNA Used to Study Phylogeny?

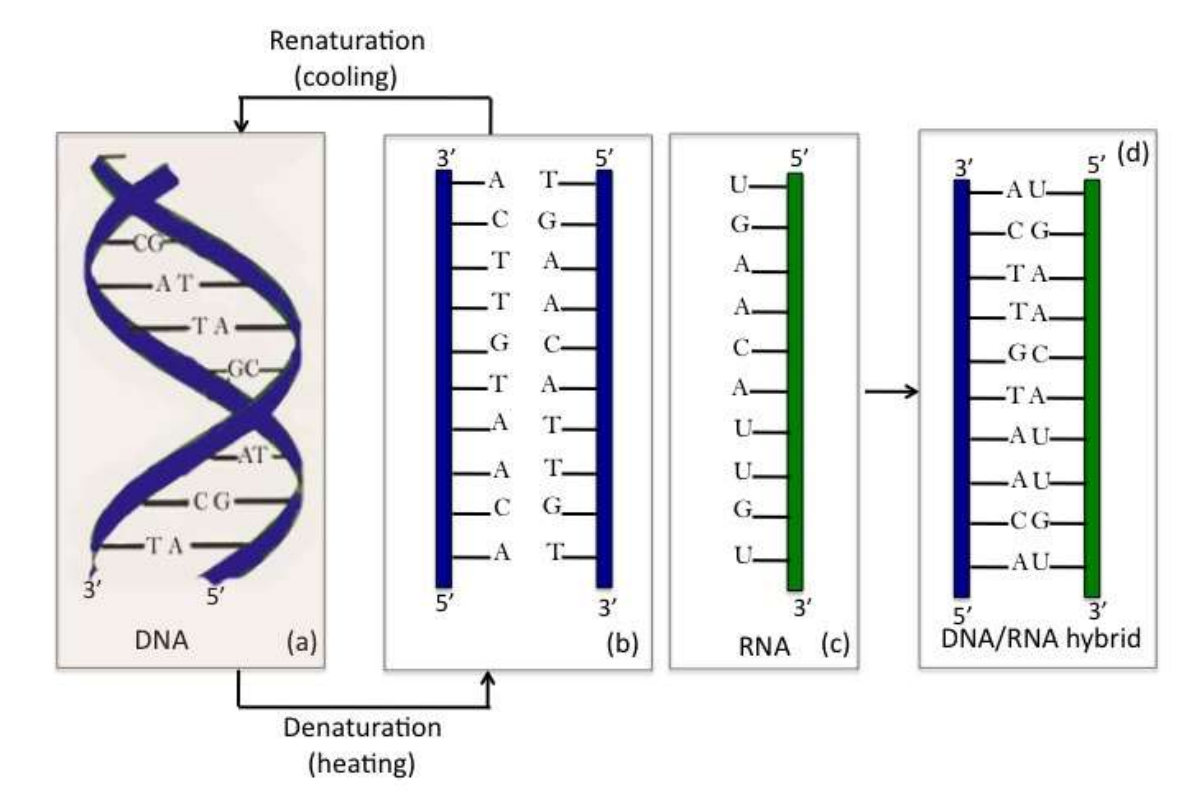
DNA-DNA Hybridization

DNA Restriction Site Analysis – RFLP

DNA Sequence Comparison

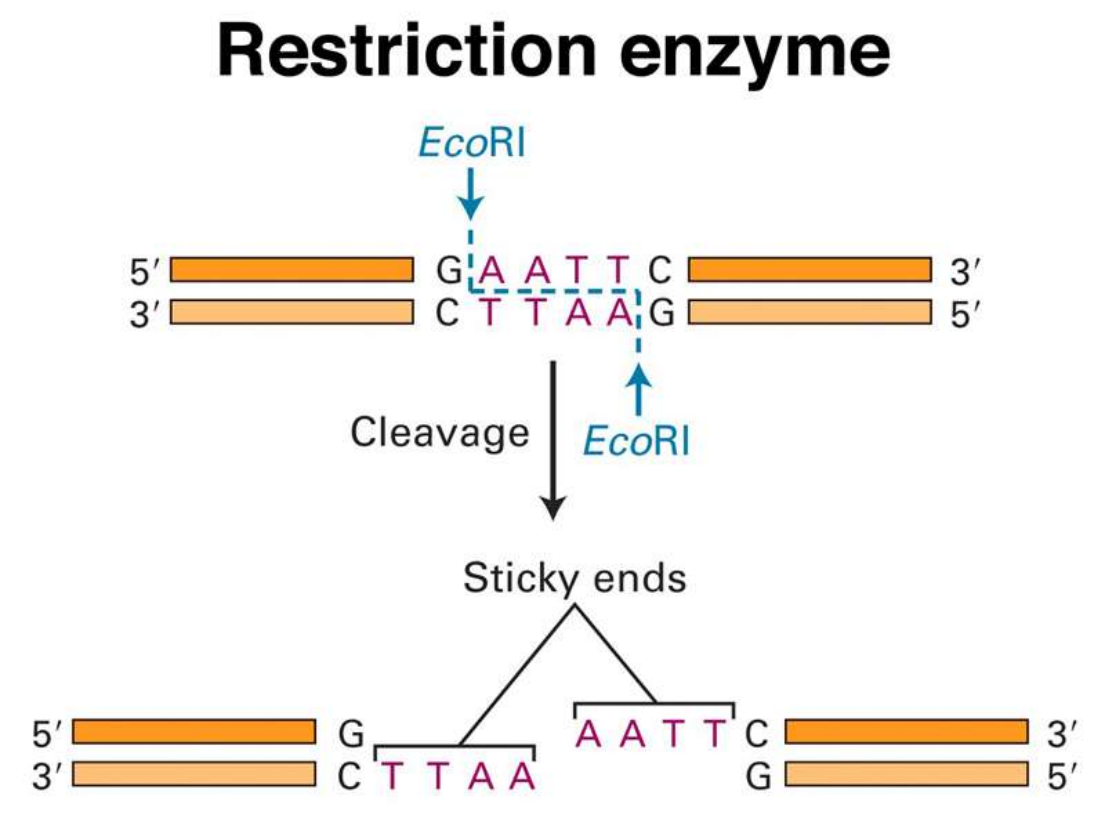
DNA-DNA Hybridization

- Early method used to compare species relationships
- Melt DNA from two species to separate strands
- Mix melted DNA single strands, allow to cool
- Measure how long it takes hybrid DNA to anneal back to double stranded DNA, gives measure of relationship



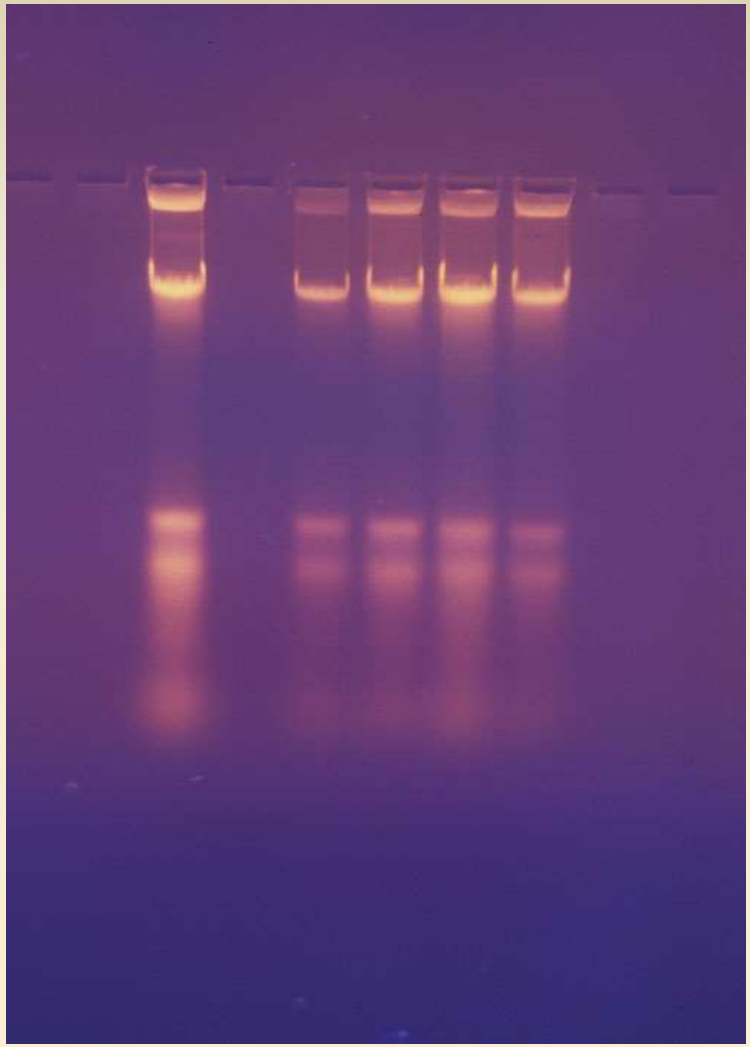
Restriction Fragment Length Polymorphism - RFLP

Cut DNA in pieces using restriction enzymes.
Look at length variation in the fragments.



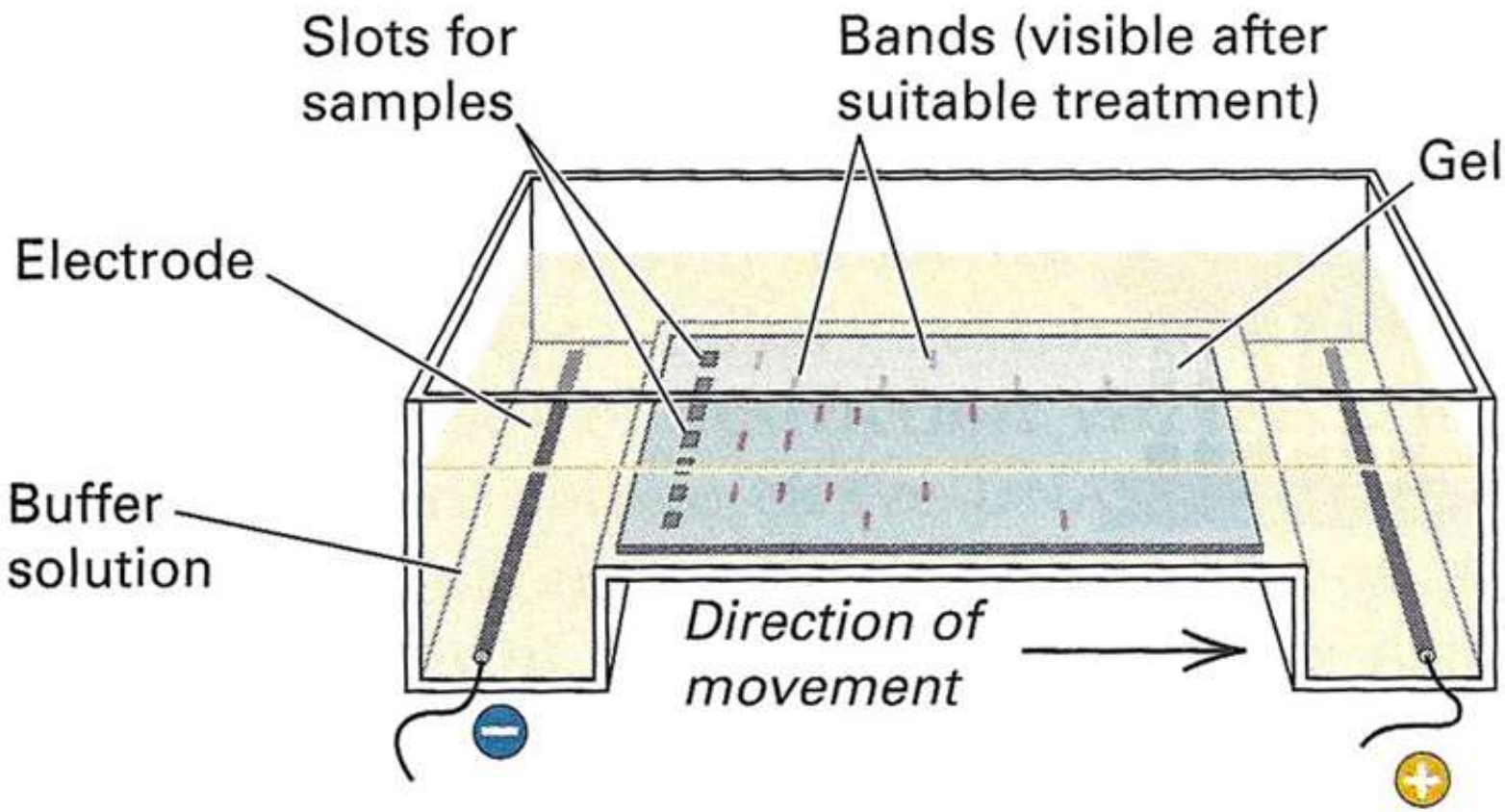
Working with Plant DNA – Agarose Gel Electrophoresis

Loading DNA on Agarose Gel



Visualizing DNA

Agarose gel electrophoresis of DNA



RFLP Analysis – Simple Explanation

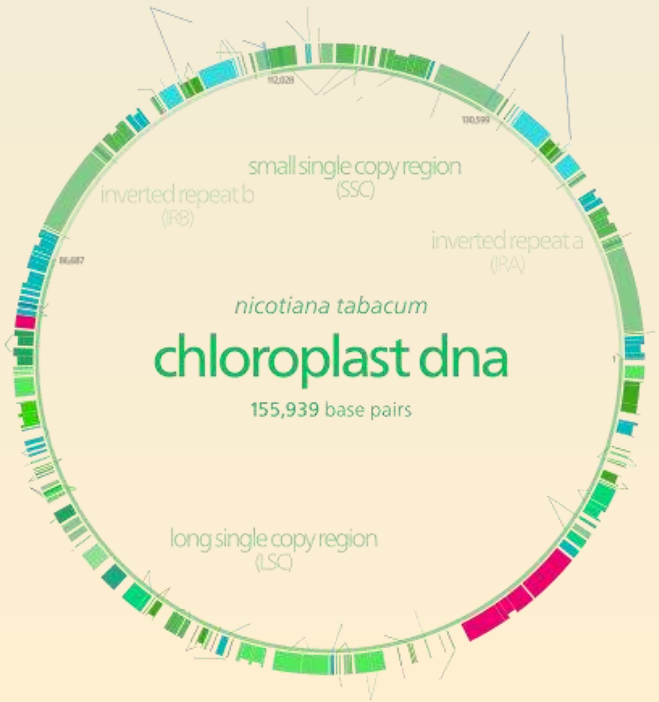
Extract DNA

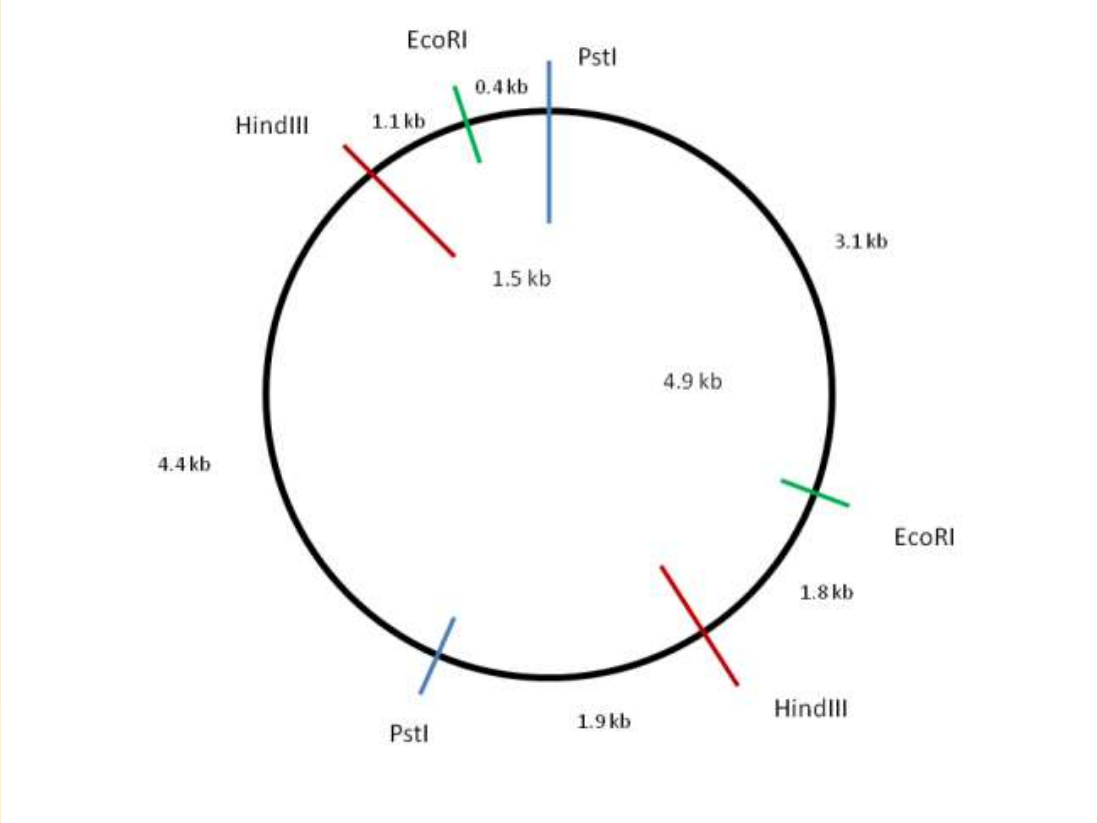
Cut with restriction enzymes

Analyze fragment patterns

If there is insertion or deletion fragment lengths will differ

If there is mutation in restriction site, enzyme will not cut, producing a larger fragment





Details: Restriction Fragment Length Polymorphism

Restriction fragment length polymorphism is the comparison of different lengths of DNA produced by restriction endonucleases to determine genetic differences. This process is usually used for larger sample size and compares the entire genome, rather than target sequences like in PCR.

The steps of RFLP are outlined below:

1. The entire genome is subjected to restriction enzymes, then run on an agarose gel.
2. The gel electrophoresis cannot distinguish differences in length because the amount of bands is so numerous; the gel is placed on a nylon membrane and a chemical is used to denature the double-stranded DNA into single strands.
3. **Southern Blotting:** An electric current is used to transfer the DNA onto a nylon membrane. The negatively charged DNA will be transferred from the gel onto the nylon and bind to it.
4. Radioactive probes are added to the membrane and they seek out regions that may have mutations in the DNA sequence. They will then pair to the bases in these areas, a process called hybridization.
5. The radioactive probes will be detected under an X-ray film and the differences in the DNA pattern can now be used for profiling of a suspect or to detect a mutation.

Restriction fragment analysis
Digoxigenin-labeled cpDNA probe

cp DNA Restriction Site Analysis

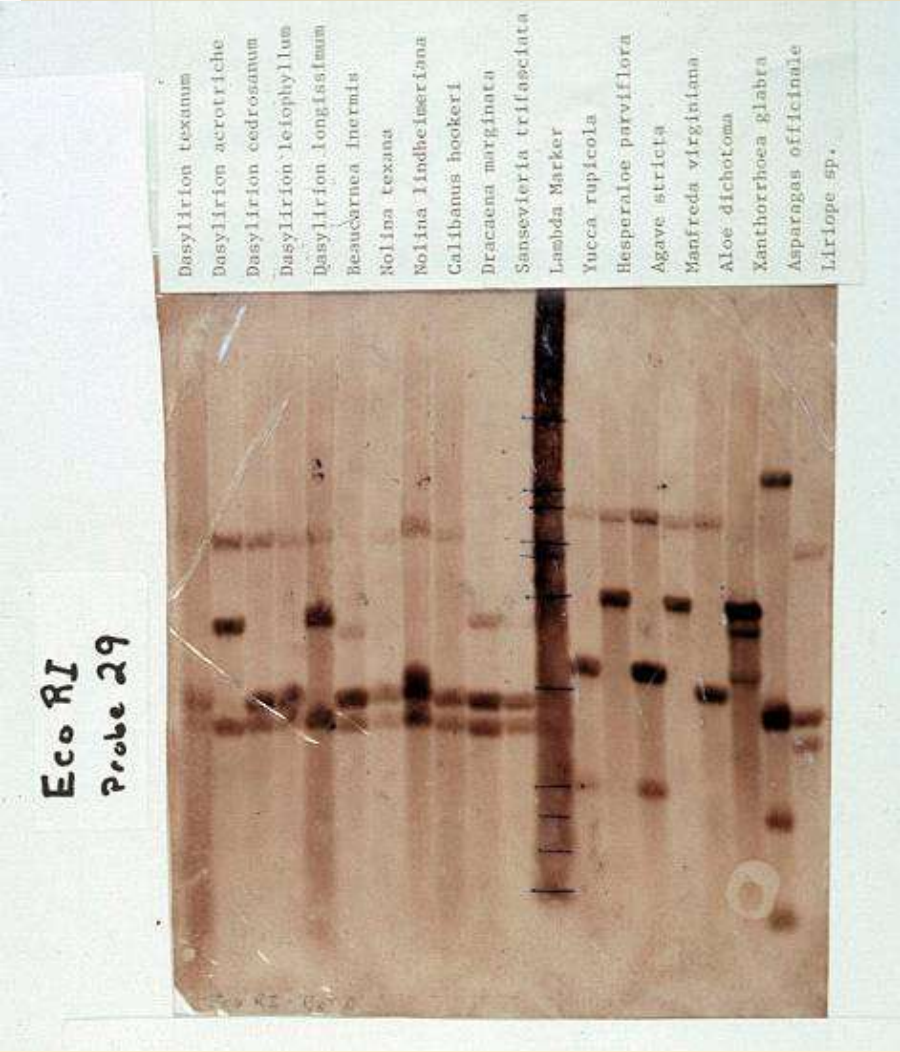
Restriction Enzymes Used:

- Bam HI
- Bcl I
- BstN I
- Dra I
- Eco RI
- Eco RV
- Hae II
- Hae III
- Hha I
- Hind III
- Msp I
- Xho I

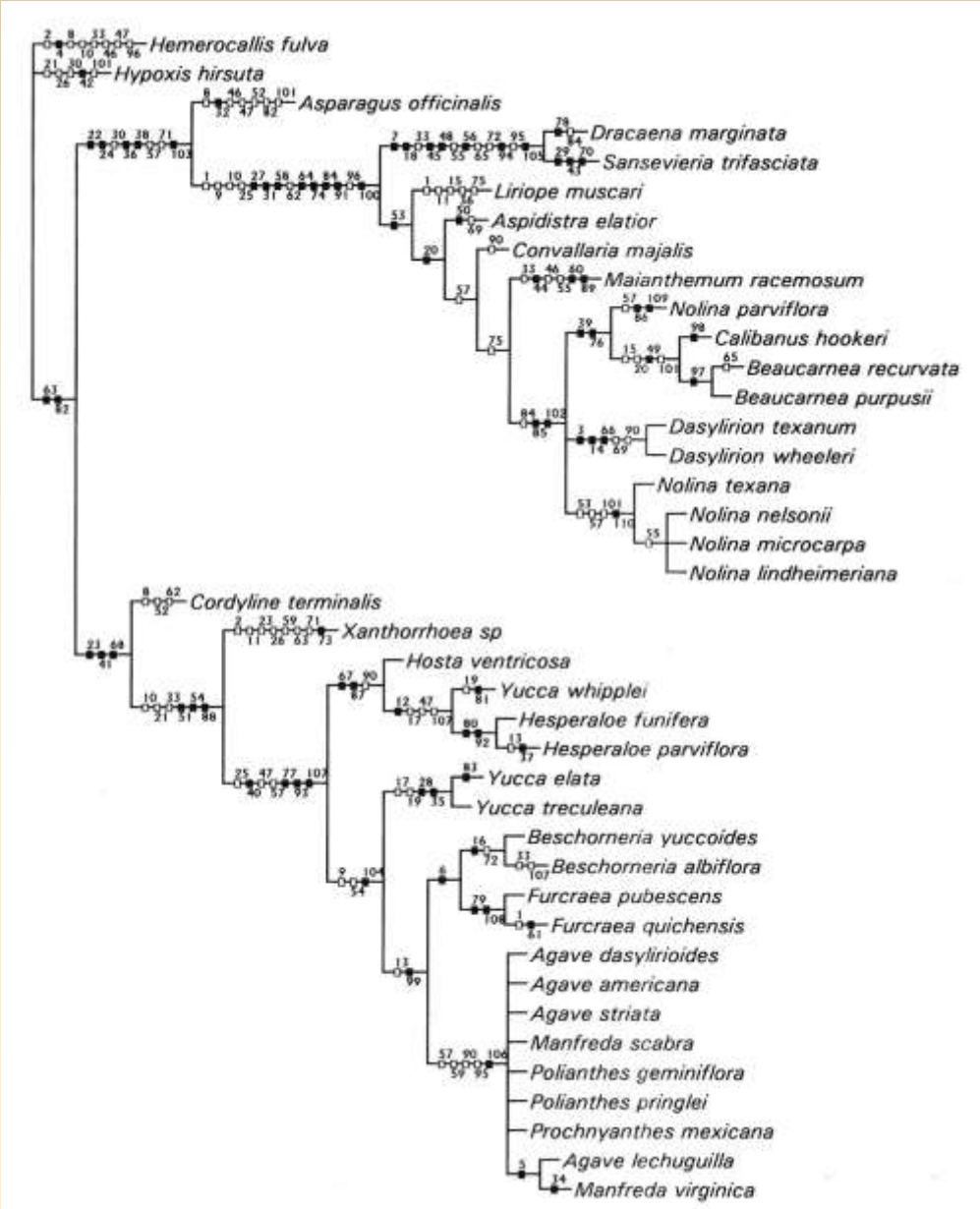
Total # Tobacco cpDNA Probes - 40
Subset of Probes Used for this Analysis - 20

100 Restriction Sites Surveyed

18 Autapomorphies
82 sites Shared by Two or More Taxa



cpDNA Restriction Site Analysis of Agavaceae



Dracaenaceae

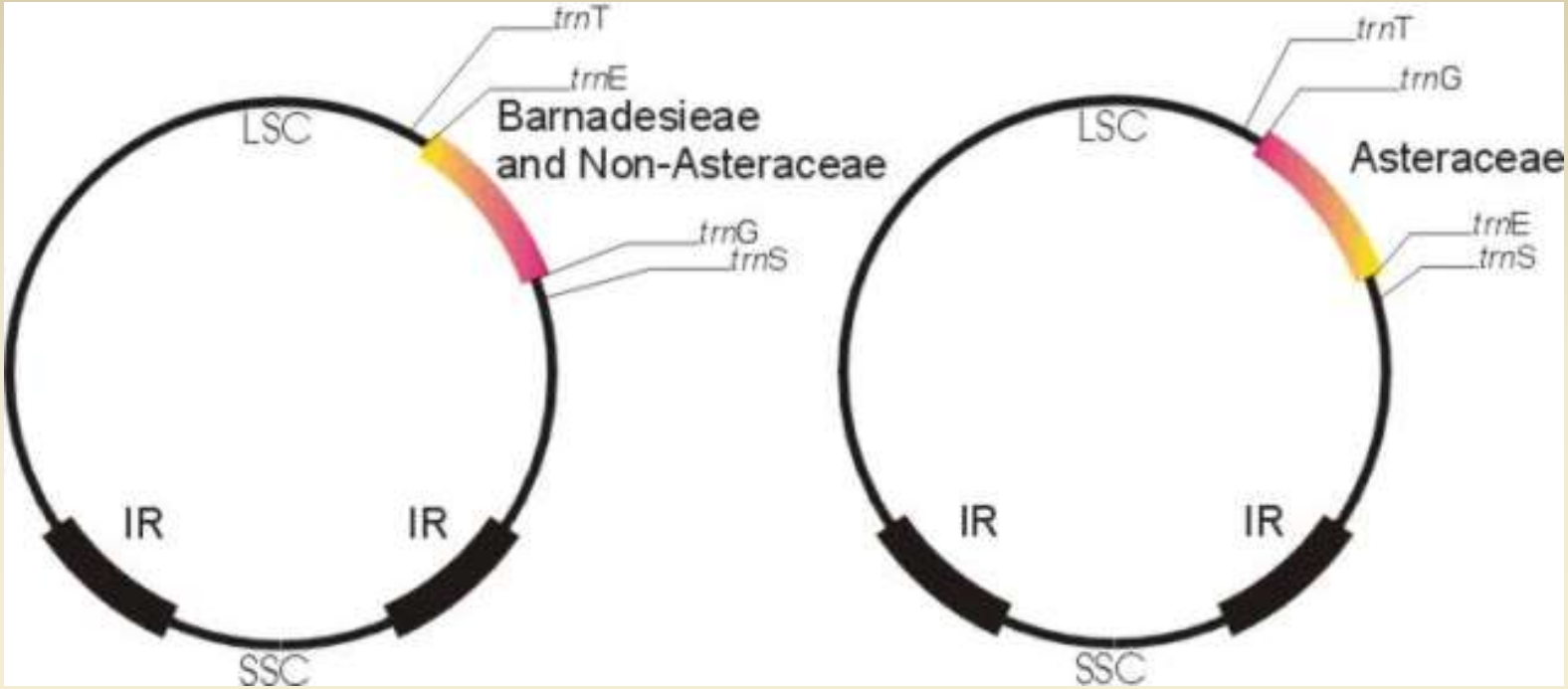
Convallariaceae

Nolinaceae

Agavaceae s.s.

Bogler and Simpson. 1995. Syst. Bot. 20: 191

A synapomorphy for the Asteraceae (excluding the Barnadesieae)



In 1987, Jansen and Palmer discovered that the whole family Asteraceae (some 20,000 species) shared a **22 kilobase inversion in the chloroplast DNA**, with the sole exception of the ***Barnadesia*** group (88 species), which had normal chloroplast DNA.

DNA Sequencing

Types of DNA sequence data:

- Chloroplast - cpDNA
- Nuclear - nDNA
- Mitochondrial - mtDNA (not used much with plants; used more with animals)

Cost – was expensive but has come down considerably. About $10 per sequence.

Much of the process has now been automated.

University core or outsource

Next Generation Sequencing coming within reach

DNA Sequencing Project – Basic Steps

1. Pick study group. Develop hypotheses to test.
2. Collect samples. Make voucher specimens for museum.
3. Store samples in freezer or silica gel. DNA is fragile.
4. Homogenize tissue – grinder or homogenizer.
5. Extract DNA – CTAB or micro kits.
6. Choose gene or region to study.
7. Design and/or order primers.
8. Amplify gene or region with PCR
9. Check amplification with gel
10. Sequence DNA in lab or outsource to commercial facility.
11. Download sequences. Check chromatograph.
12. Align sequenced single strands into contigs
13. Make data matrix
14. Analyze with phylogeny computer program

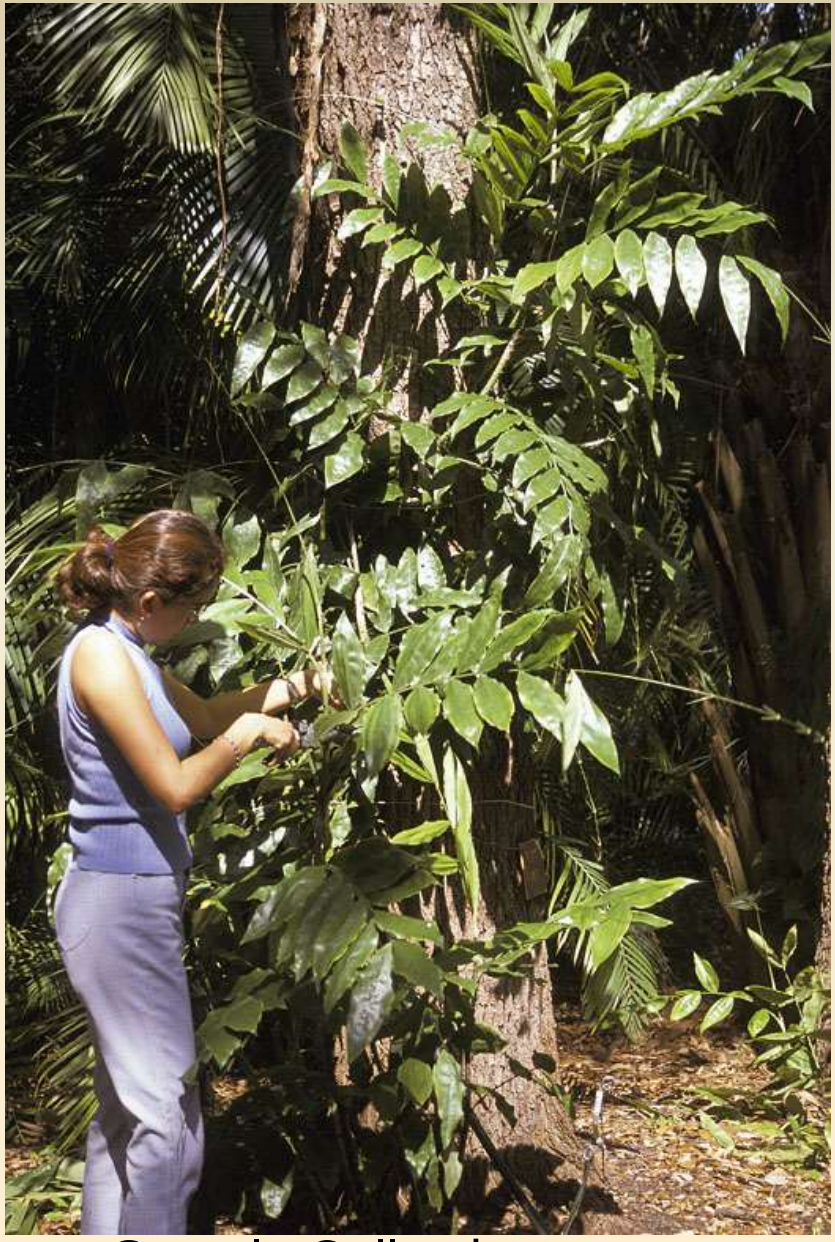
DNA Sequencing Project – Basic Steps

1. Pick study group. Develop hypotheses to test.
2. Collect samples. Make voucher specimens for museum.
3. Store samples in freezer or silica gel. DNA is fragile.
4. Homogenize tissue – grinder or homogenizer.
5. Extract DNA – CTAB or micro kits.
6. Choose gene or region to study.
7. Design and/or order primers.
8. Amplify gene or region with PCR
9. Check amplification with gel
10. Sequence DNA in lab or outsource to commercial facility.
11. Download sequences. Check chromatograph.
12. Align sequenced single strands into contigs
13. Make data matrix
14. Analyze with phylogeny computer program

Collecting Specimens
Pressing the Plant for Voucher

Livingstone Nganga, UMSL Undergraduate
2012 REU

Collecting plants and preparing vouchers….

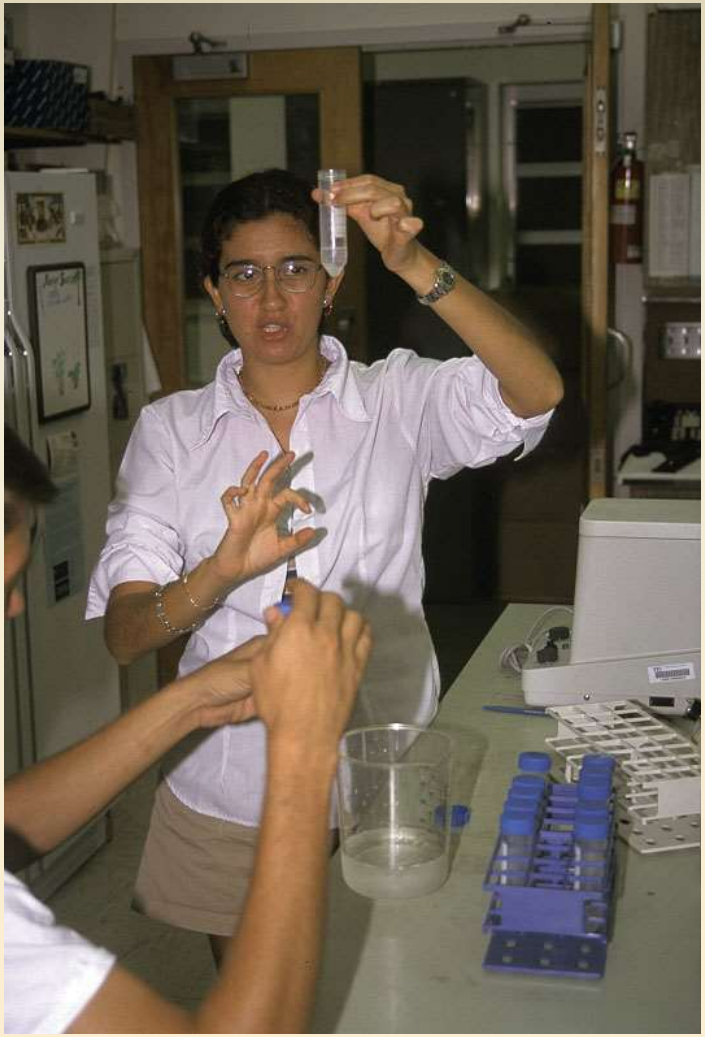


Sample Collection

Voucher Specimens

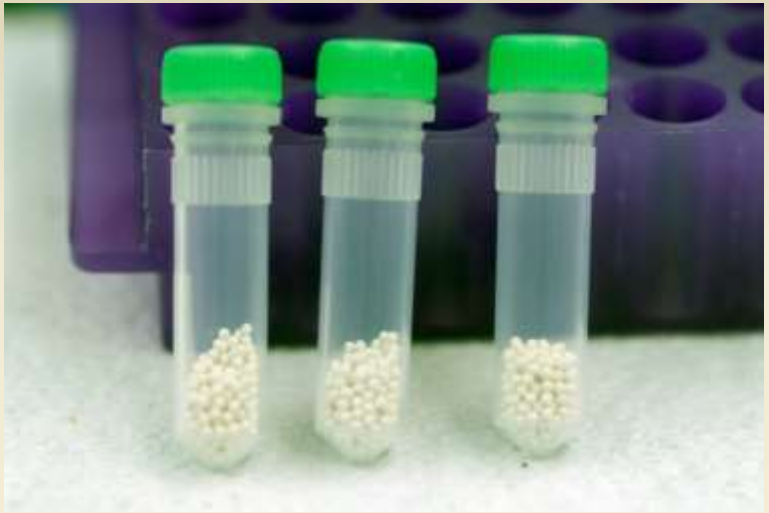
Extracting DNA

Grinding Tissue



Extracting DNA

Extracting DNA from Leaf Samples



FastPrep DNA Extraction Kit

Livingstone Nganga

DNA Sequencing Project – Basic Steps

1. Pick study group. Develop hypotheses to test.
2. Collect samples. Make voucher specimens for museum.
3. Store samples in freezer or silica gel. DNA is fragile.
4. Homogenize tissue – grinder or homogenizer.
5. Extract DNA – CTAB or micro kits.
6. Choose gene or region to study.
7. Design and/or order primers.
8. Amplify gene or region with PCR
9. Check amplification with gel
10. Sequence DNA in lab or outsource to commercial facility.
11. Download sequences. Check chromatograph.
12. Align sequenced single strands into contigs
13. Make data matrix
14. Analyze with phylogeny computer program

Each Gene Mutates at a Different Rate

- Genes coding for vital enzymes or structures tend to be more conserved.
- The frequency of a mutation of a gene determines its utility for addressing a specific question
- Slow rate of mutation– used for older groups
- Fast rate of mutation– used to asses relationships in closely related populations

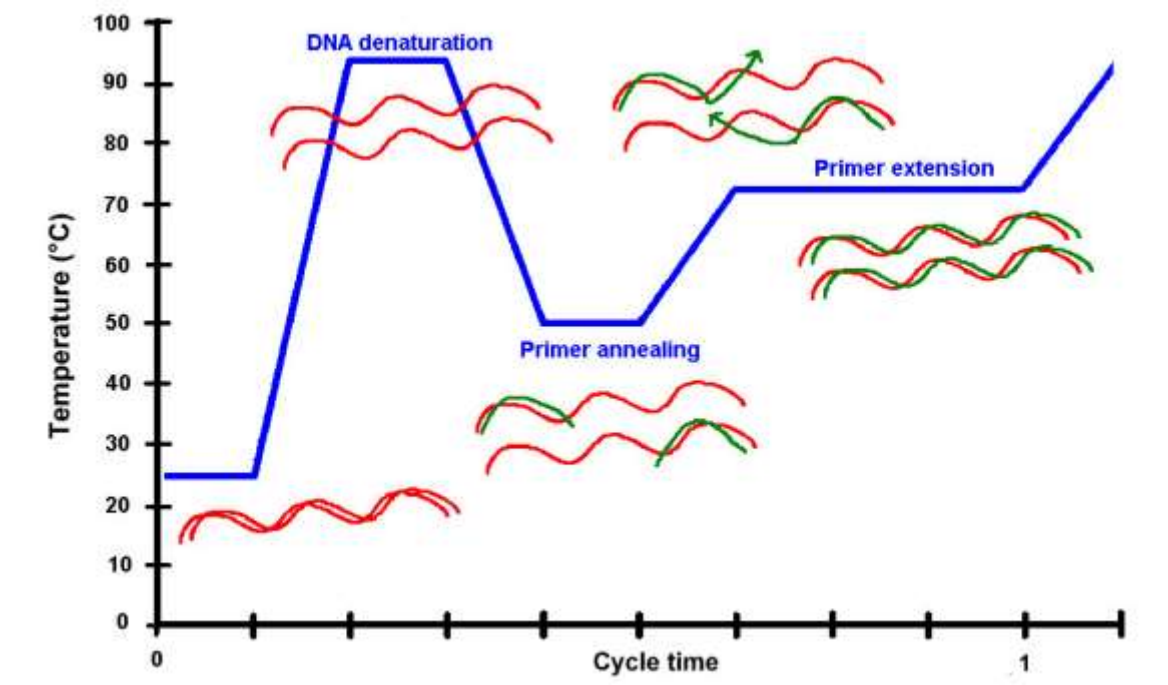
Gene Mutation Rate Problems

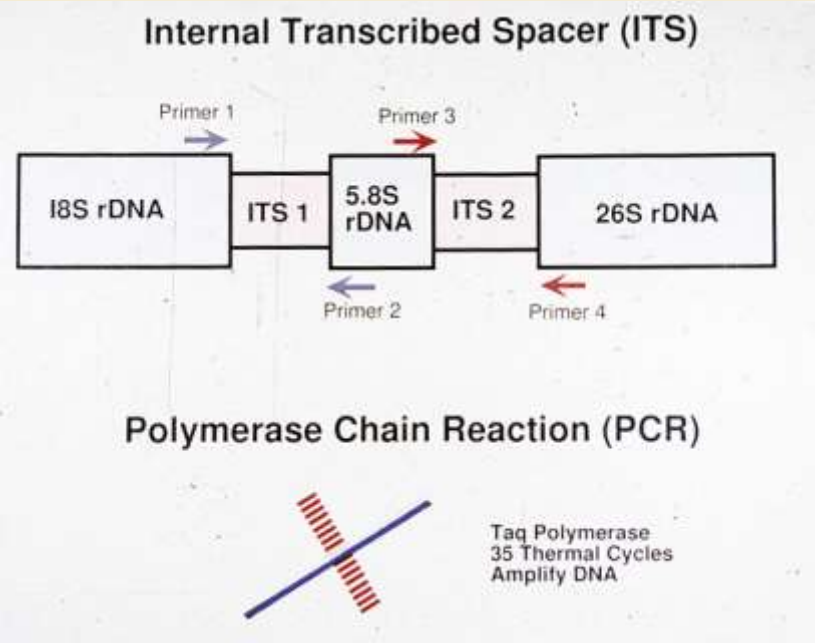
- If a gene is mutating very slowly, the level of variation approaches the sequencing error rate and inferences become unreliable
- If a gene is mutating very quickly, parallelisms and reversals accumulate so fast that all phylogenetic information is lost
- Genes have to be picked for a given study based on what information is desired and what rate of genetic mutation will be required for that goal.

DNA Sequencing Project – Basic Steps

1. Pick study group. Develop hypotheses to test.
2. Collect samples. Make voucher specimens for museum.
3. Store samples in freezer or silica gel. DNA is fragile.
4. Homogenize tissue – grinder or homogenizer.
5. Extract DNA – CTAB or micro kits.
6. Choose gene or region to study.
7. Design and/or order primers.
8. Amplify gene or region with PCR
9. Check amplification with gel
10. Sequence DNA in lab or outsource to commercial facility.
11. Download sequences. Check chromatograph.
12. Align sequenced single strands into contigs
13. Make data matrix
14. Analyze with phylogeny computer program

DNA Amplification





PCR: Polymerase Chain Reaction

What is it?

Process used to amplify DNA gene or spacer region, replication into thousands of copies.

How does it work?

DNA is isolated, purified.

PCR Tube: DNA + Taq polymerase, nucleotides, buffer/salts

Thermal cycler heats to denature, cools to anneal primer, warmed so taq polymerase makes copy.

Cycle repeated over and over, amplifying DNA

Primer:

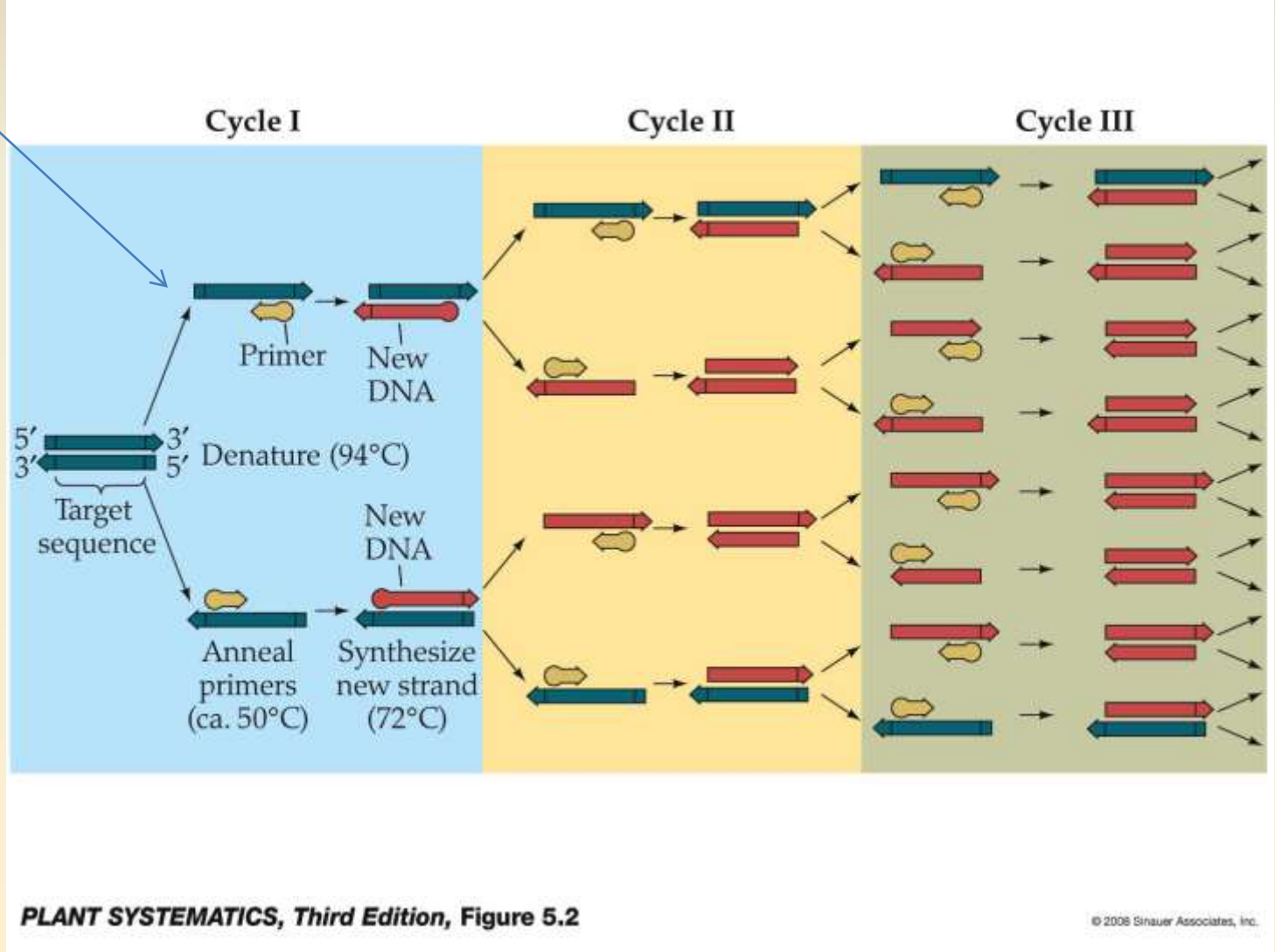
Primer = short, <u>conserved</u> DNA region

Complementary to ends of DNA region to be amplified

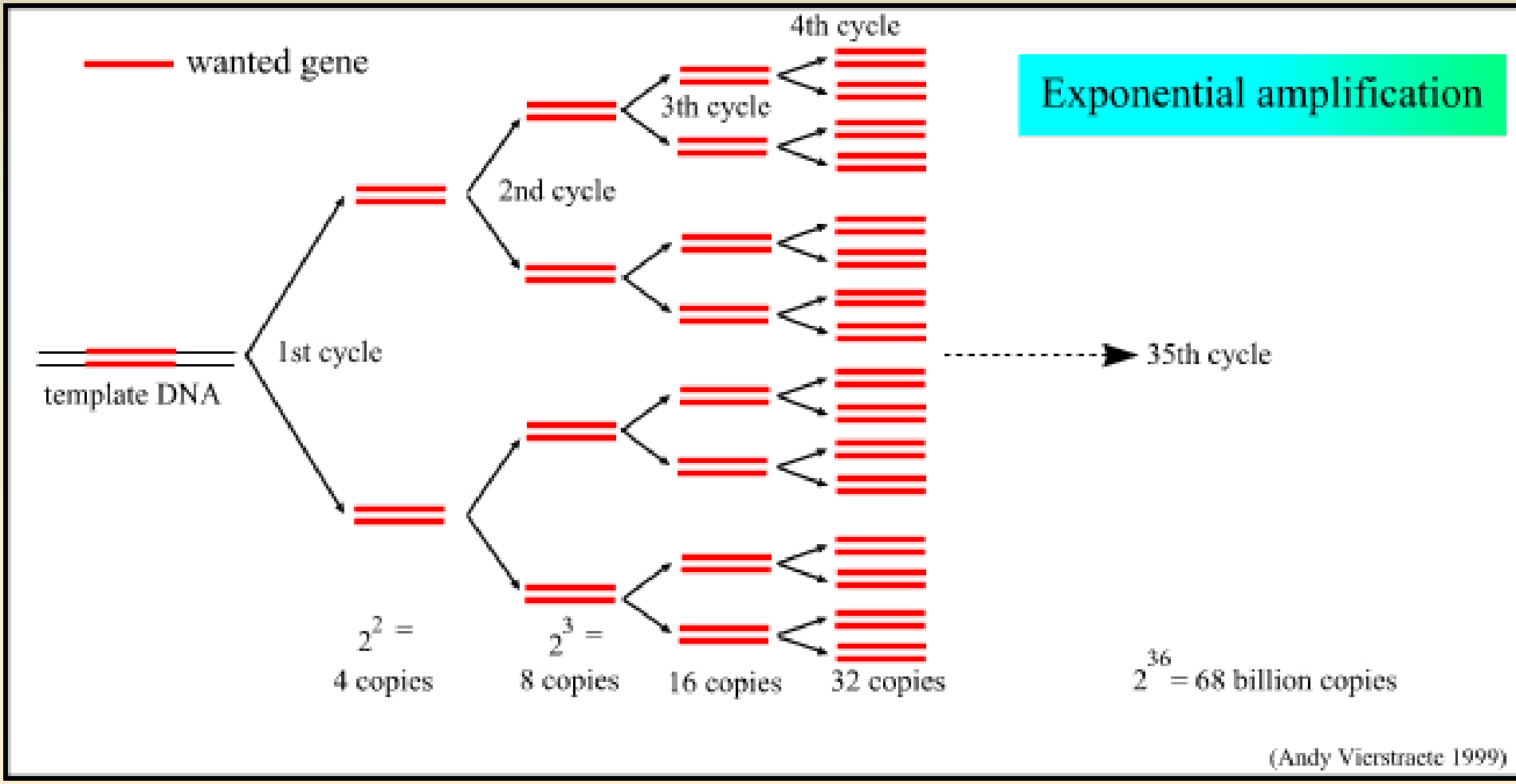
Taq Polymerase – isolated from thermal spring bacteria (Thermus aquaticus), stable in boiling water.

Polymerase Chain Reaction

Finding the primer is the hard part– you have to know something about the gene you want to sequence ahead of time.

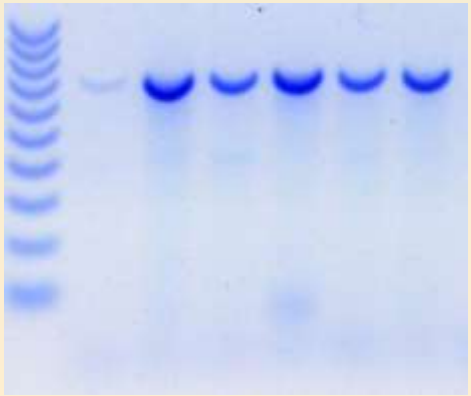


PCR


wanted gene
Exponential amplification
4th cycle
3th cycle
2nd cycle
1st cycle
template DNA
35th cycle
2^2 =
4 copies
2^3 =
8 copies
16 copies
32 copies
2^{36} = 68 billion copies
(Andy Vierstraete 1999)

DNA Amplification: PCR

Kelsey Huisman, 2013 REU



Sequencing Facility

PCR Product Gel Electrophoresis – check size

DNA Sequencing Project – Basic Steps

1. Pick study group. Develop hypotheses to test.
2. Collect samples. Make voucher specimens for museum.
3. Store samples in freezer or silica gel. DNA is fragile.
4. Homogenize tissue – grinder or homogenizer.
5. Extract DNA – CTAB or micro kits.
6. Choose gene or region to study.
7. Design and/or order primers.
8. Amplify gene or region with PCR
9. Check amplification with gel
10. Sequence DNA in lab or outsource to commercial facility.
11. Download sequences. Check chromatograph.
12. Align sequenced single strands into contigs
13. Make data matrix
14. Analyze with phylogeny computer program

Sanger Method DNA Sequencing

How does it work?

Similar to PCR amplification

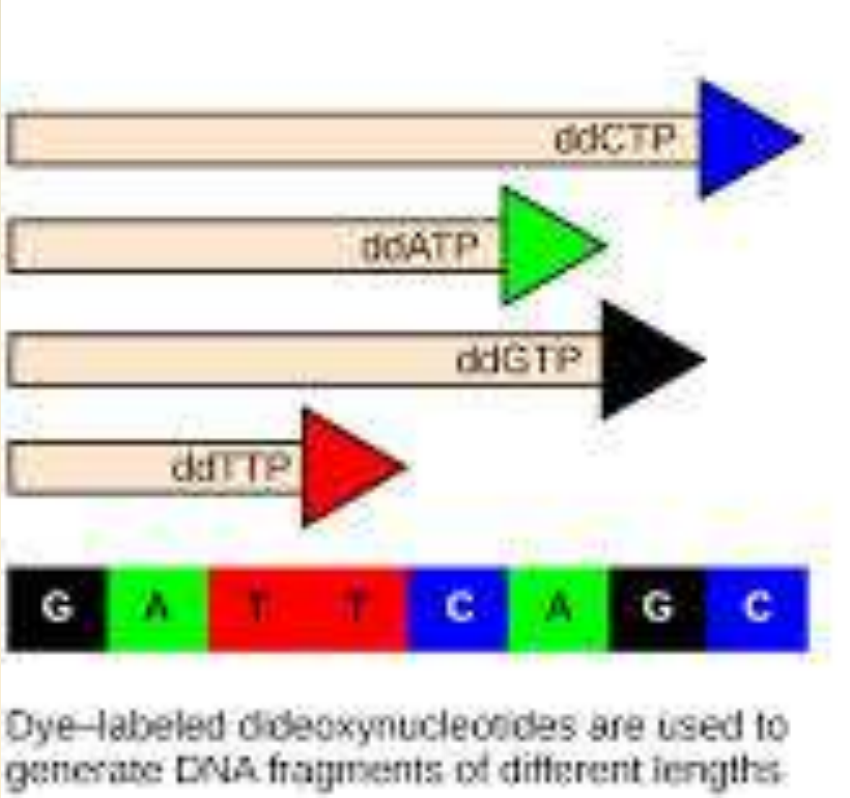
But, small amount of **Dideoxynucleotides** used (along with higher conc. of nucleotides)

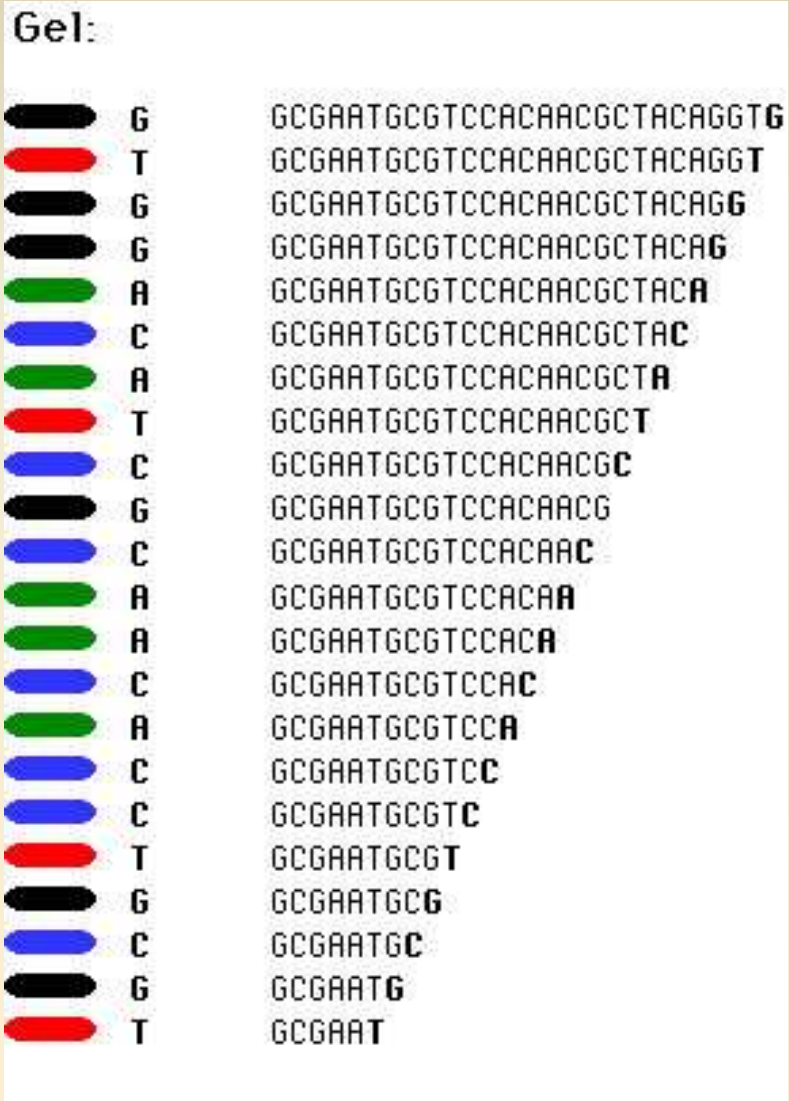
Dideoxynucleotides, once joined to new DNA strand, terminate polymerase reaction.

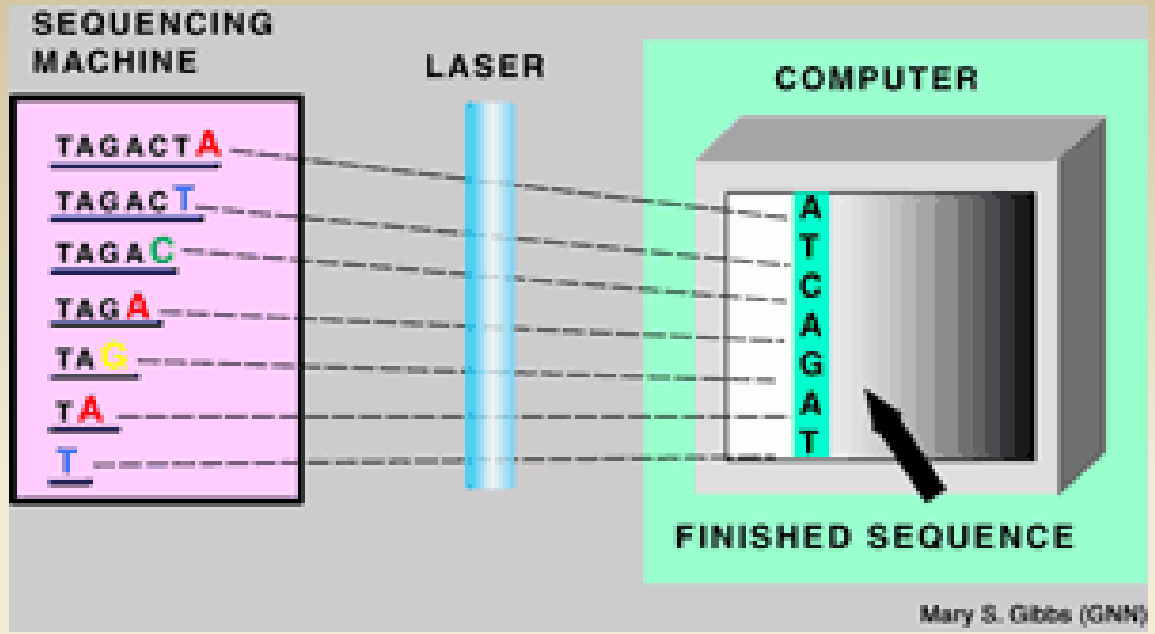
Dideoxynucleotides identified by fluorescence pattern.

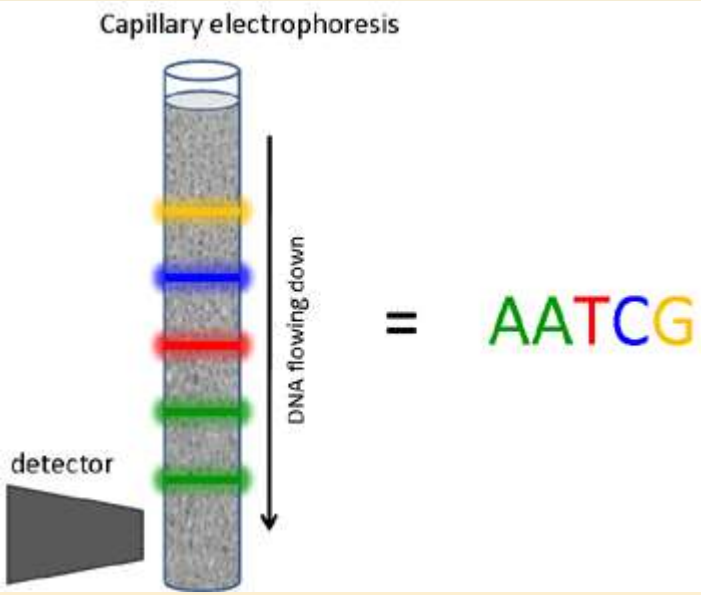
Length of DNA strands determined by electrophoresis.

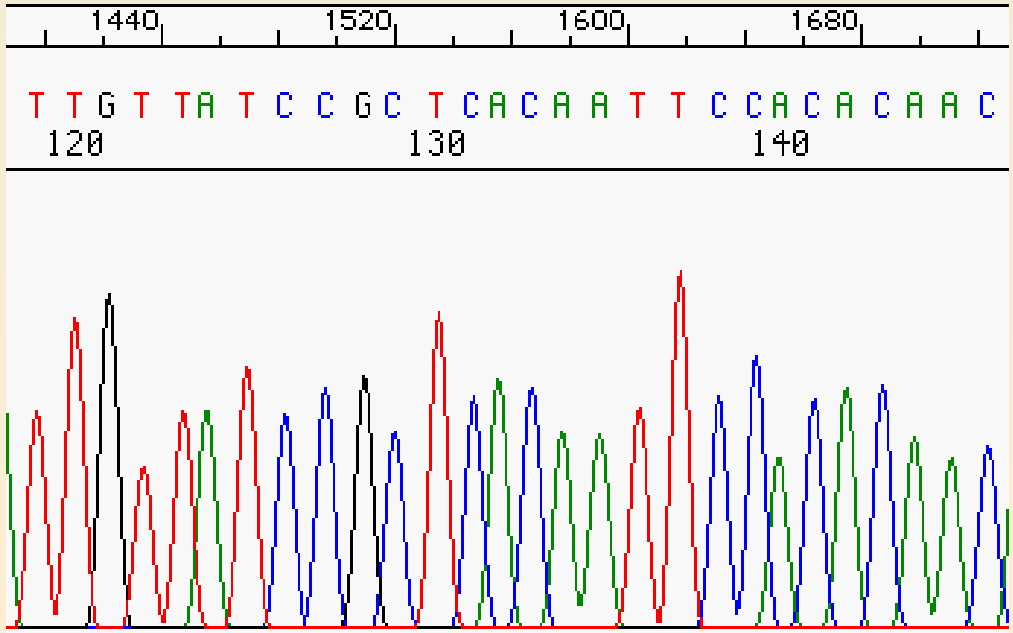
Sanger Dideoxy Sequencing Method



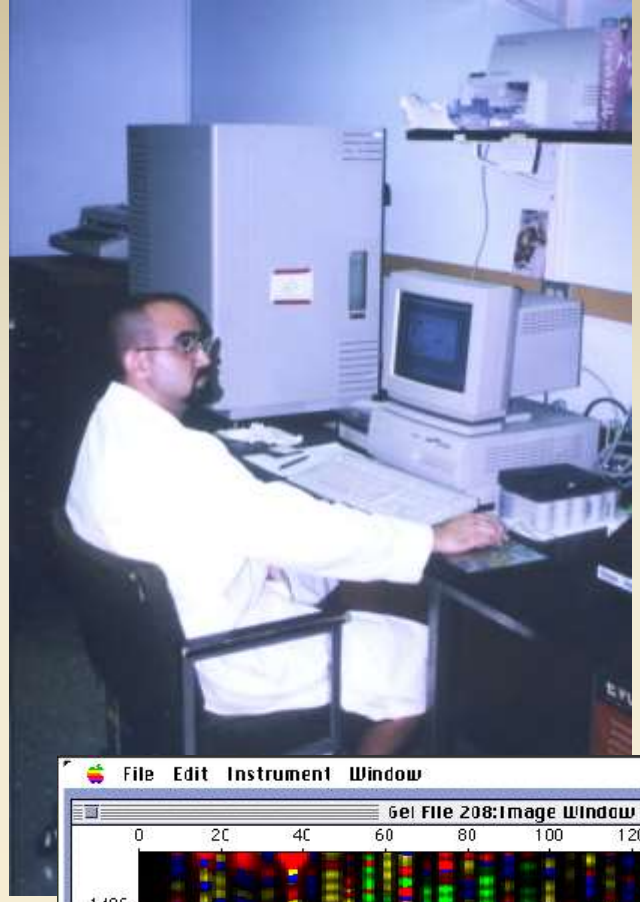


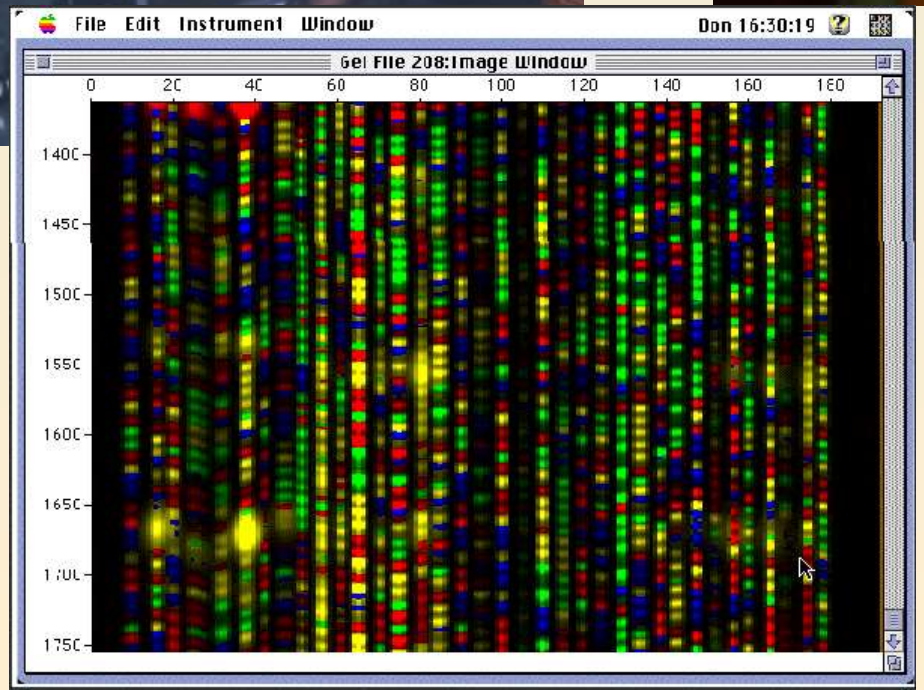

SEQUENCING MACHINE
LASER
COMPUTER
TAGACTA
TAGACT
TAGAC
TAGA
TAG
TA
T
ATCAGAT
FINISHED SEQUENCE
Mary S. Gibbs (GNN)

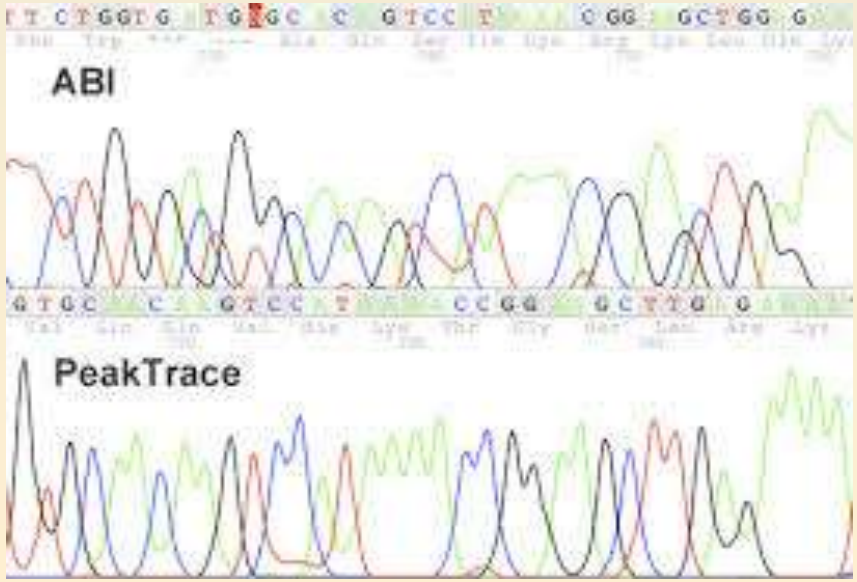

Capillary electrophoresis
DNA flowing down
detector
=
AATCG


1440
1520
1600
1680
TTGTTATCCGCTCACAATTCCACACAAC
120
130
140

Old DNA Sequencing – ABI 377



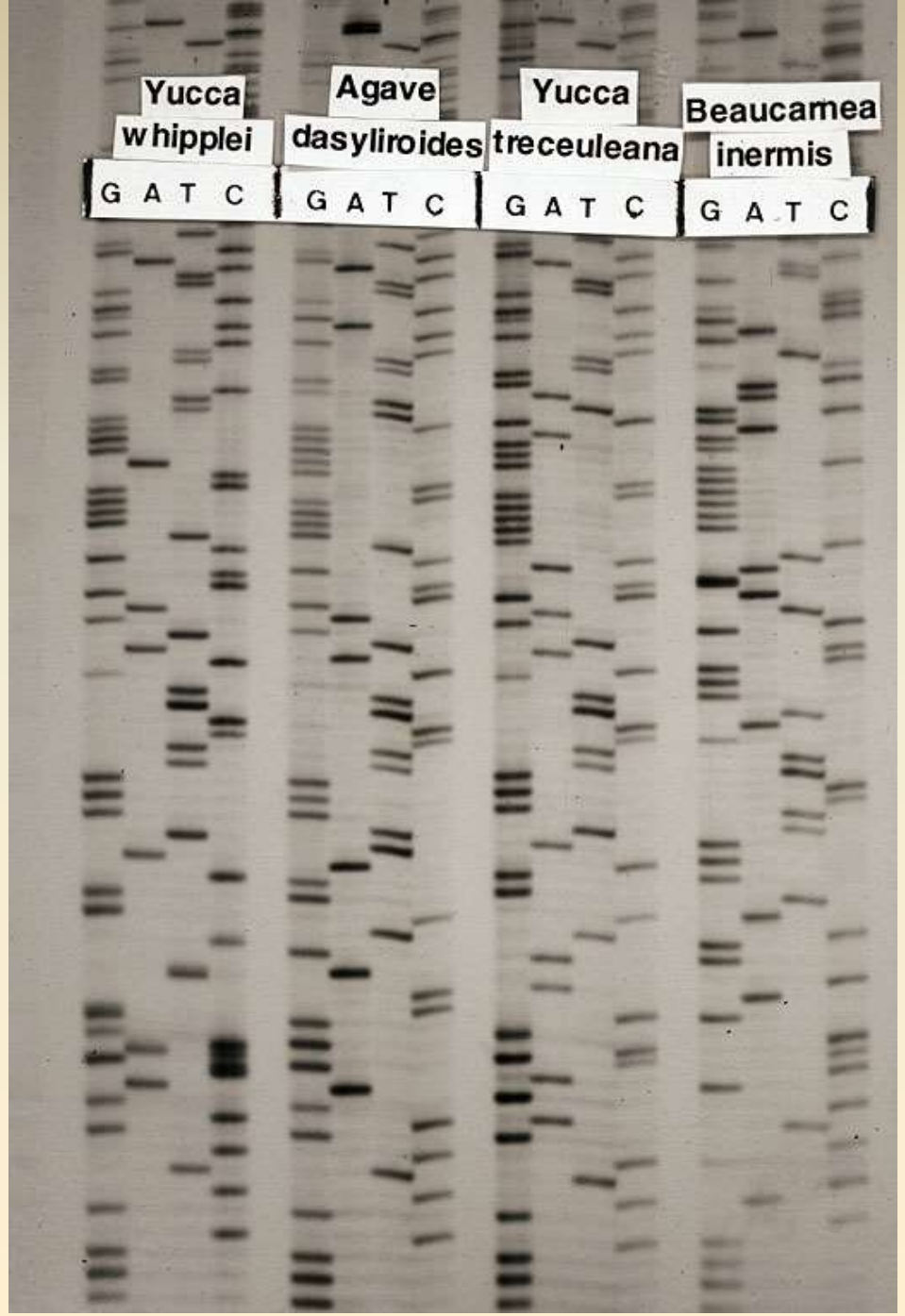




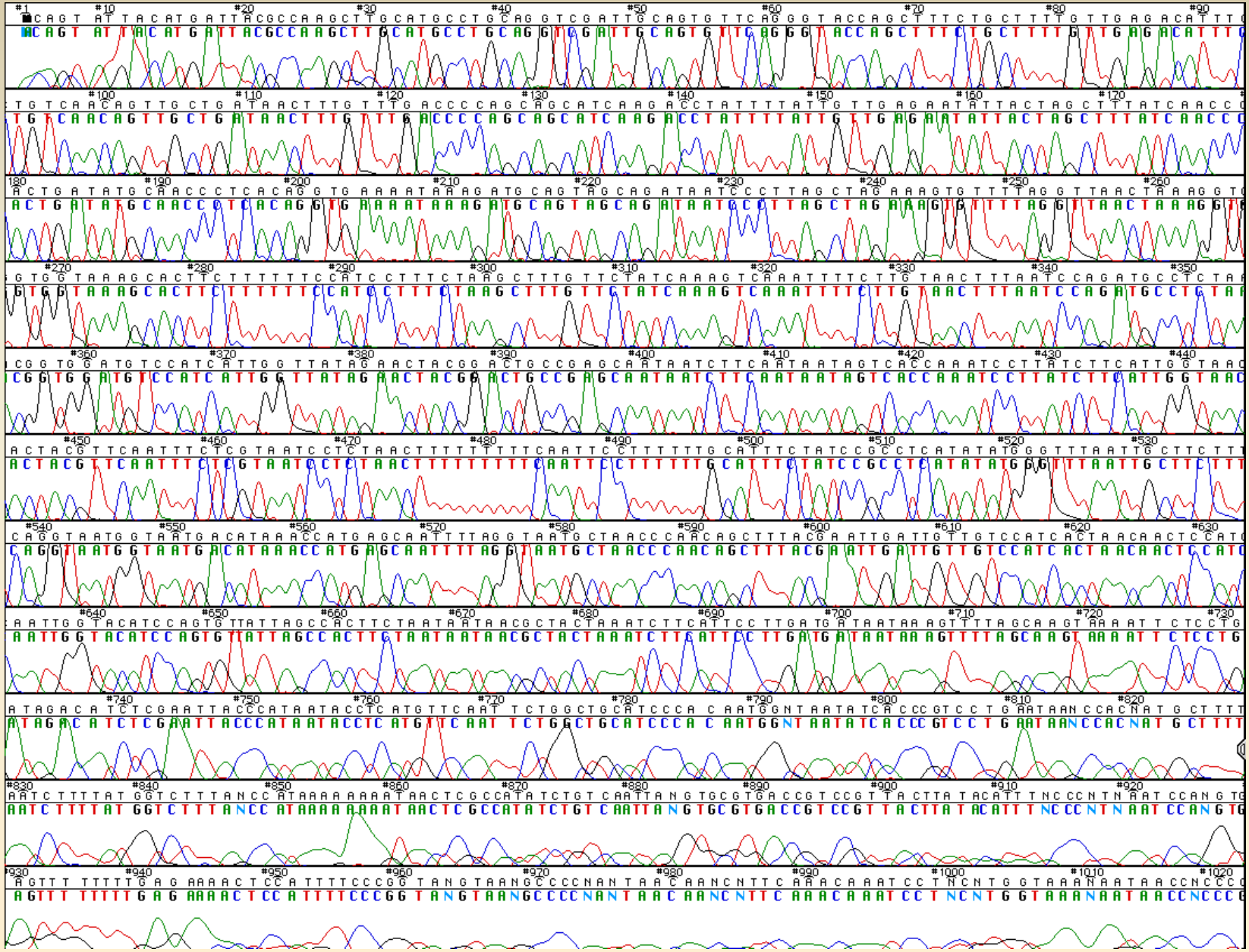
A modern capillary sequencer – ABI 3100

My ancient history…

Sanger Sequencing
P32-labeled dNTP
X-ray film

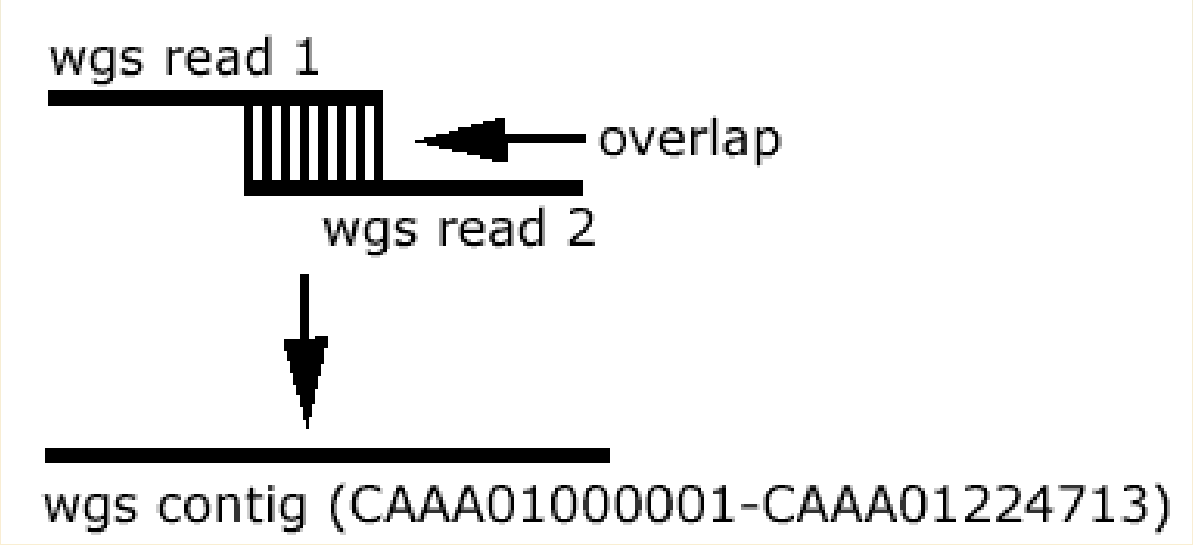


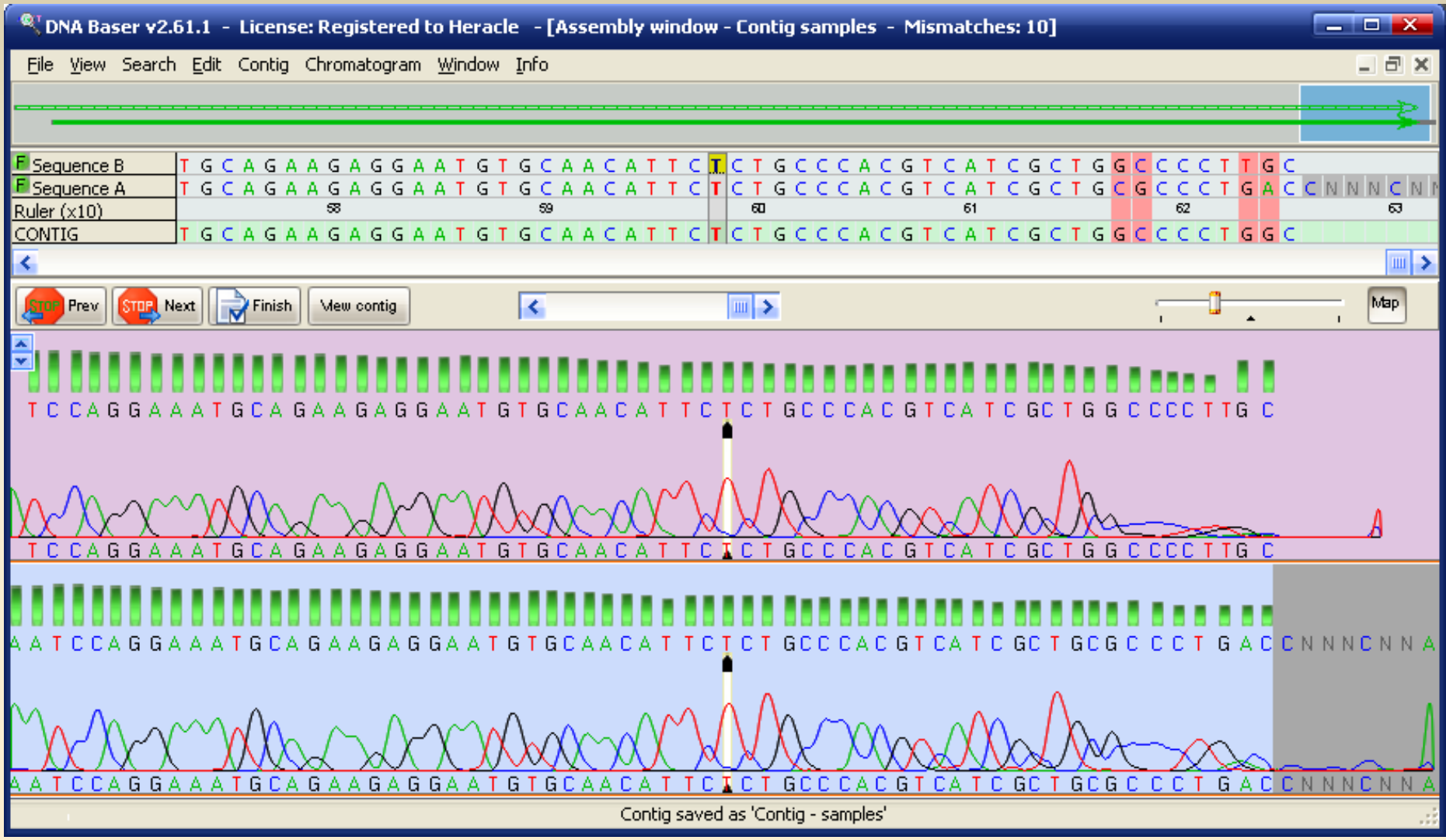
Chromatogram for One Sequence



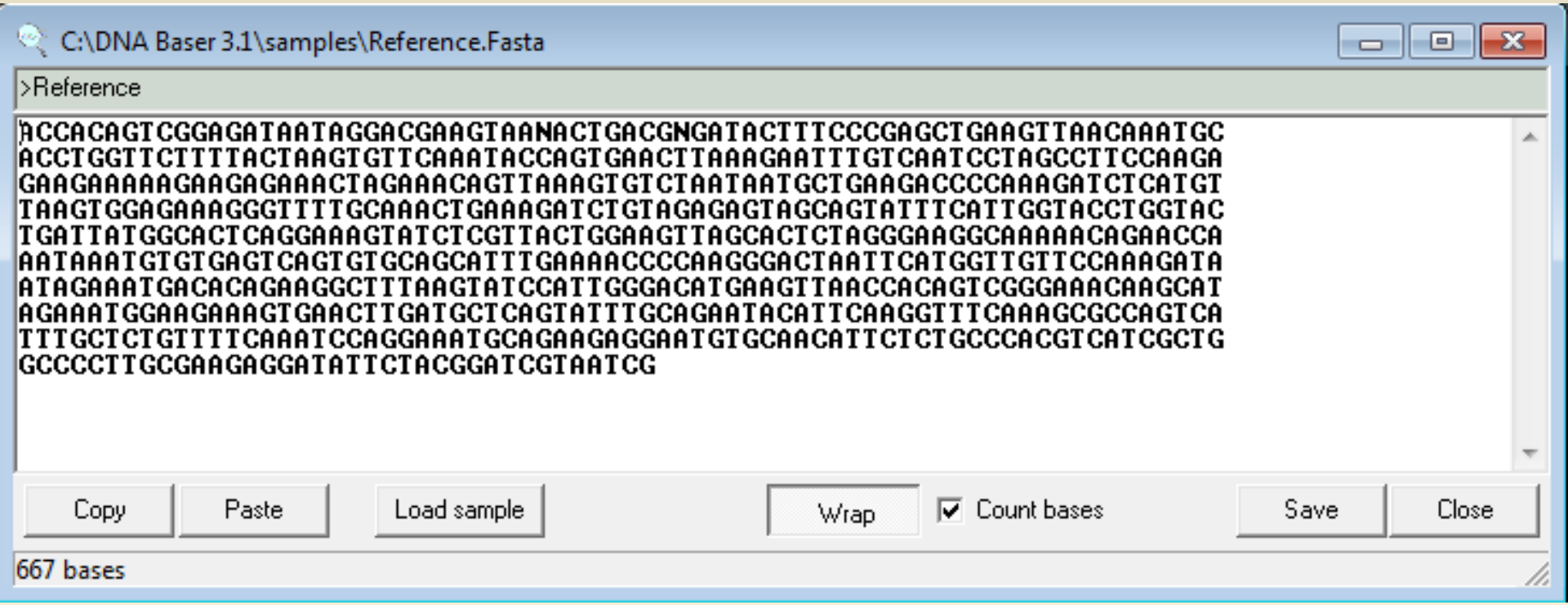
Contig Assembly

Match overlapping end regions of pieces to get full gene read consensus sequence




DNA Baser v2.61.1 - License: Registered to Heracle - [Assembly window - Contig samples - Mismatches: 10]
File View Search Edit Contig Chromatogram Window Info
Sequence B
Sequence A
Ruler (x10)
CONTIG
Prev
Next
Finish
View contig
Map
Contig saved as 'Contig - samples'

Finally, the Sequence Data for alignment and analysis…



Sequence Alignment Programs – Clustal, Muscle

FastA Format for Loading into ClustalW or Other Alignment Program

>Crataegus_castlegarensis_Exon1
gaagcagccgtaacgccagtagcggcagctgctgcggcggcggctggttatactttgcggccgccaagggagcttggacttggaggg
cttgaagacttgttccaggcttatggggttagatactacacgacggcgaagatagcggagcttggatttactgtgaacaccctcttgga
catgaaggatgatgagcttgatgacatgatgagcagcctctctcagatattccgctgggagttgcttgttggggagaggtatggtatca
aagctgccgtcagagccgagcgccgccgccttgaggaggaggactctcggcggcgcaaccttgtctctggtgataccaccaccaatg
ccctagatgctctctcccaagaag

>Cochearia_officinalis_Exon1
atggatcctgaaggtttcacgaatggcttattccgatggaacacaacaagagcaatgattcaacaacaacaacaattaccaccgcctc
aaatcactcctccgccgcaacaatcaccggcaacaccacaaacggcggcgtttgggatgagactaggtggtttagaaggtttgttcgg
tccttacgggatacgtttttacacggcggcgaagatagctgagctaggtttcacggcgagcacgcttgttggtatgaaagacgaagag
cttgaagatatgatgaatagtctctcacatatctttcgttgggagcttcttgtcggtgaacgttacggtatcaaagctgccgttagaactg
aacggaggagattgcaagaagaggaagaggaggaatcttctagacgccgtcattttatgctctccgccggtggtgattccggcactca
ccacgctcttgatgctctctctcaagaag

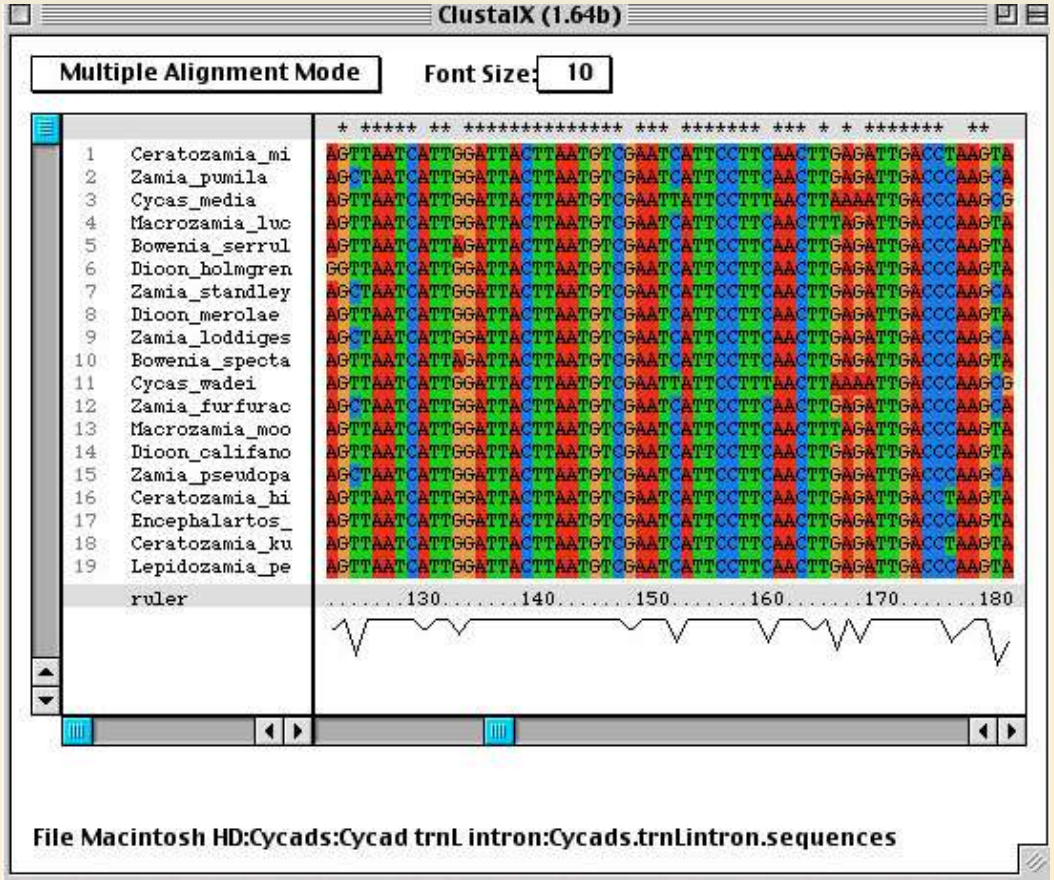
>Capsella_bursa-pastoris_Exon1
gtggcttattccgatggaacccaatgagagcaatggttcaagcaccacctccggttcctccttcgccgcagcagcaacagccggcaac
acctcagacggcggctttcgggatgcgacttggtggcttagagggactctttggtgcttacggtatccgtttctacacggcggcgaagat
agcggagttgggttttacggccagcacgctcgttggtatgaaggacgaggagcttgaggagatgatgaacagtctctctcacatcttta
ggtgggagcttctcgttggtgaacggtacggtatcaaagctgccgtaagagctgaacggagacgattgcaagaagaggaggaggaa
tcttctagacgccgtcatttgctgctctccgccgctggtgattccggtactcatcacgctcttgatgccctctcccaagaa

Pairwise alignments

```
43.2% identity;                    Global alignment score: 374

                10        20        30        40              50
alpha  V-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSA
       : :.: .:. : : ::::  .. : :.::: :... .: :. .:  : :::      :.
beta   VHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNP
               10          20        30        40        50

            60        70        80        90       100       110
alpha  QVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
       .::.::::: :.....::.:.. .....::.::  ::.::: ::.::.. :. .:: :.
beta   KVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHF
       60        70        80        90       100       110

           120       130       140
alpha  PAEFTPAVHASLDKFLASVSTVLTSKYR
         :::: :.:. .: .:.:...:. ::.
beta   GKEFTPPVQAAYQKVVAGVANALAHKYH
      120       130       140
```

Multiple Sequence Alignment

- Goal: create data matrix in which columns are homologous positions
- Problem: sequences vary in length
- Why?
  - Insertions
  - Deletions



Simple Sequence Alignment

Taxon 1	GTACGTTG
Taxon 2	GTACGTTG
Taxon 3	GTACGTTG
Taxon 4	GTACATTG
Taxon 5	GTACATTG
Taxon 6	GTACATTG

DNA Sequence Data Matrix

	C1	C2	C3	C4	C5	C6	C7	C8
T1	G	T	A	C	G	T	T	G
T2	G	T	A	C	G	T	T	G
T3	G	T	A	C	G	T	T	G
T4	G	T	A	C	A	T	T	G
T5	G	T	A	C	A	T	T	G
T6	G	T	A	C	A	T	T	G

Slightly Less Simple Sequence Alignment

Taxon 1	AGAGTGAC
Taxon 2	AGAGTGAC
Taxon 3	AGAGTGAC
Taxon 4	AGAGGAC
Taxon 5	AGAGGAC
Taxon 6	AGAGGAC

Slightly Less Simple Sequence Alignment

Taxon 1	AGAGTGAC
Taxon 2	AGAGTGAC
Taxon 3	AGAGTGAC
Taxon 4	AGAG–GAC
Taxon 5	AGAG–GAC
Taxon 6	AGAG–GAC

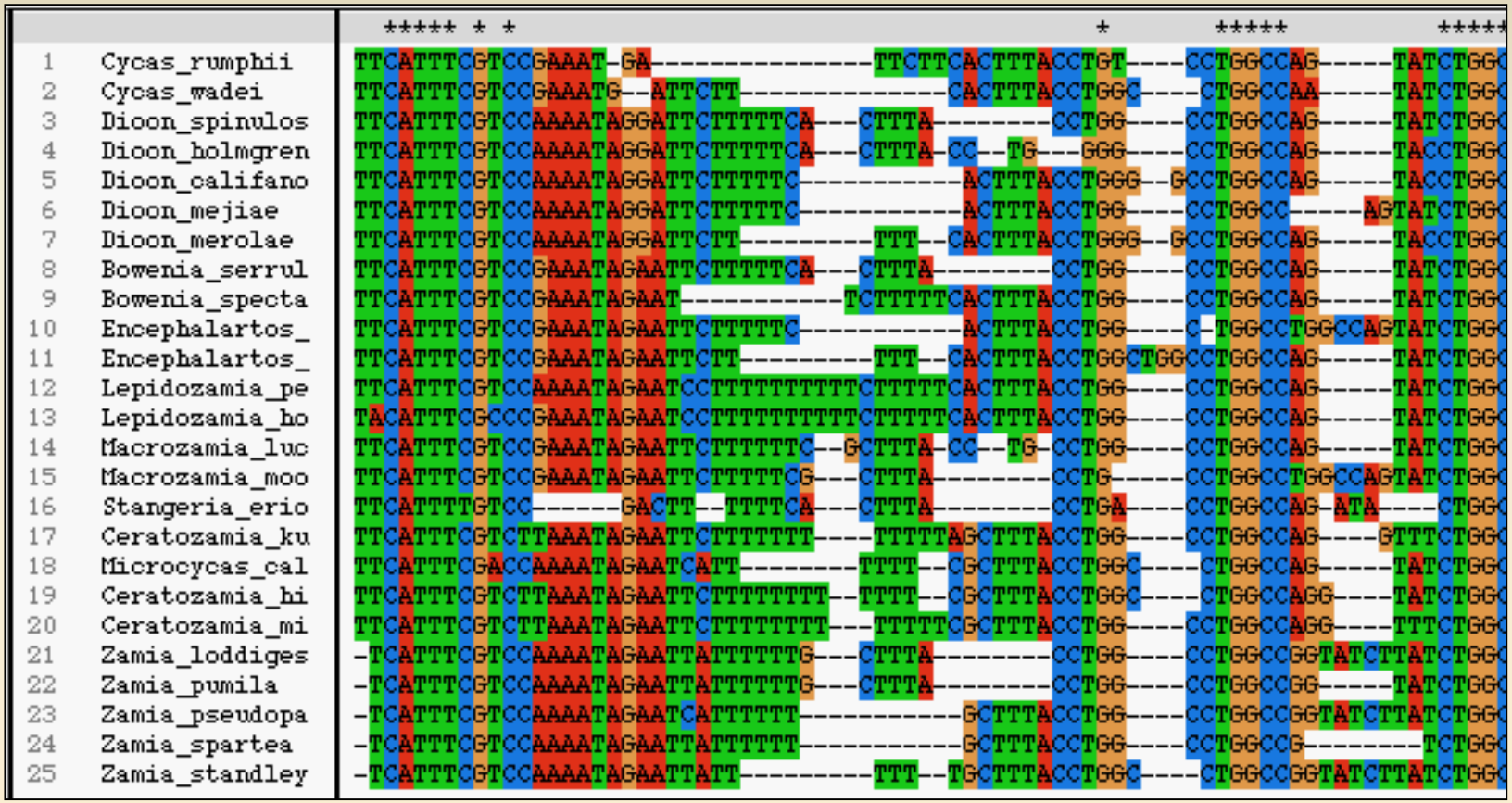
Alignment Gaps

- Gaps are inserted to maximize homology across nucleotide positions
- Gaps are hypothesized indels
- Inserting a gap assumes that an indel event is a better explanation of the differences among sequences than nucleotide substitution

Gap Number and Length

- All else being equal, is it better to assume fewer longer gaps, or more shorter gaps?
- In other words, what is more likely:
  - For a new indel to occur?
  - For an existing indel to lengthen?
- There is no general answer!
  - Alternate alignments are explored algorithmically

Hypervariable Region in *trn*S - *trn*G Spacer
an alignment nightmare



Clustal Alignment Algorithm

- Creates alignment based on penalties for gap opening (number of gaps) and gap extension (gap length)
- Multiple alignment built according to guide tree determined by pairwise alignments
- Order of adding sequences determined by a guide tree

Clustal Alignment Algorithm

Distance matrix calculated from pairwise comparisons

→

Dendrogram calculated from from distance matrix

↘

Alignment calculated for most similar pair of sequences, based on alignment parameters

←

Additional sequences are added according to dendrogram, until all sequences are added

DNA Sequencing Project – Basic Steps

1. Pick study group. Develop hypotheses to test.
2. Collect samples. Make voucher specimens for museum.
3. Store samples in freezer or silica gel. DNA is fragile.
4. Homogenize tissue – grinder or homogenizer.
5. Extract DNA – CTAB or micro kits.
6. Choose gene or region to study.
7. Design and/or order primers.
8. Amplify gene or region with PCR
9. Check amplification with gel
10. Sequence DNA in lab or outsource to commercial facility.
11. Download sequences. Check chromatograph.
12. Align sequenced single strands into contigs
13. Make data matrix
14. Analyze with phylogeny computer program

Mitochondrial partial NADH1 alignment for birds

```
#Nexus

Begin DATA;
  Dimensions ntax=29 nchar=10692;
  Format datatype=dna gap=-;
  Matrix
  Tinamou       AACTATCTATTCATATCCTTATCATACATCATTCCTATTCTTATTGCA..
  Emu           AACCATCTCACTATATCACTCTCCTATGCAATCCCCATTCTAATCGCA..
  Cassowary     AACCACCTCACCATATCCCTGTCCTATGCAATCCCAATTCTAATCGCA..
  Kiwi          AACTACCTCACTATATCACTATCATATGTCATCCCAATTCTGATTGCA..
  Rhea          AACTACCTAATTATGTCCCTGTCATATGCTATCCCAATTCTAATCGCA..
  Ostrich       ACACACCTGACTATAGCACTCTCATACGCTGTTCCAATCCTAATTGCA..
  Chicken       AACCTTCTAATCATAACCTTATCCTATATTCTCCCCATCCTAATCGCC..
  BrushTurkey   AAACACCTCATCATATCCCTATCCTATGTTCTCCCAATTTTAATCGCC..
  MagpieGoose   AATCACCTCATTATAACCCTATCGTATGCCATCCCAATCCTAATCGCC..
  Duck          AGCTACCTCATTATATCCCTCCTATACGCCATCCCCATTCTAATCGCC..
  Broadbill     ACTAACCTTACCATATCCCTATCCTACGCCATCCCCGTCCTAGTTGCC..
  Flycatcher    ACCCACCTCATTATATCACTATCCTATGCCGTACCCATCCTAATTGCT..
  ZebraFinch    ATTAACCTCATCATAGCCCTCTCCTATGCCCTCCCAATCCTGATCGCA..
  Rook          GTCAACCTCATTATAGCACTTTCTTATGCTATCCCTATTCTAATCGCC..
  Oystercatcher ACCTATCTCATTATATCCCTATCCTATGCCATCCCAATCCTGATCGCA..
  Turnstone     ACCTACTTCATCATATCCCTATCCTATGCAATCCCAATTCTAATTGCA..
  Penguin       GCTCACTTAGCCATATCCCTATCCTATGCCATCCCAATCCTCATTGCA..
  Albatross     ACCTATCTTGTCATGTCCCTATCATATGCCATCCCAATCCTAATCGCC..
  ;
End;
```

Next GenerationSequencing

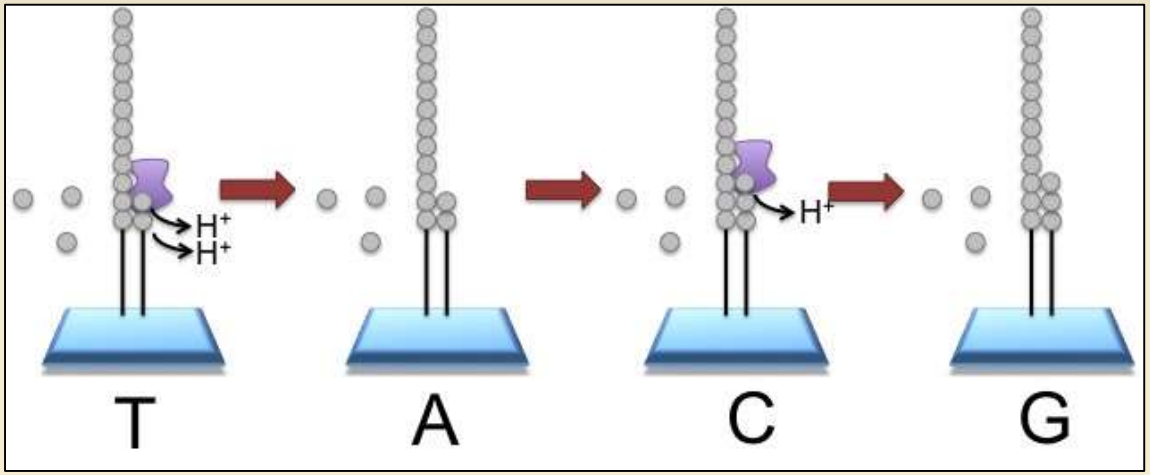
The Genome Institute, Washington Univ.

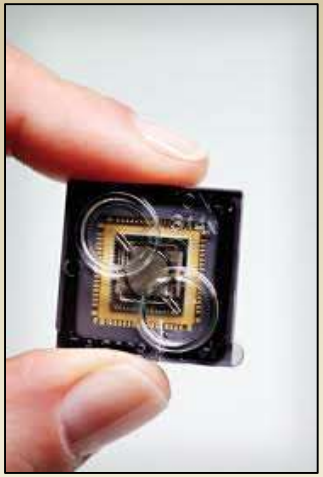
Illumina 454 NGS

Ion Torrent NGS

NGS in a Nutshell

DNA is fragmented.
Adapters are added.
DNA molecule is placed on a bead.
Each bead is placed in a single well on a slide





Semiconductor Chip

Molecules are amplified on the bead by emulsion PCR
Slide is flooded with a single species of dNTP, along with buffers and polymerase, one NTP at a time.
The pH is detected in each of the wells, as each H^+ ion released will decrease the pH.
The changes in pH allow us to determine if that base, and how many thereof, was added to the sequence read. Computer keeps track.
The dNTPs are washed away, and the process is repeated cycling through the different dNTP species.

Sequence fragments are assembled into fragments by software

Sequence fragments can also be screened for microsatellite regions
New NGS population techniques on horizon - RADSeq

Methods of tree estimation

- Distance based
  - Minimum distance
    - Shortest summed branch lengths
- Character based
  - Maximum parsimony (MP)
    - Fewest character changes
  - Maximum likelihood (ML)
    - Highest probability of observing data, given a model
  - Bayesian
    - Similar to ML, but incorporates prior knowledge

End