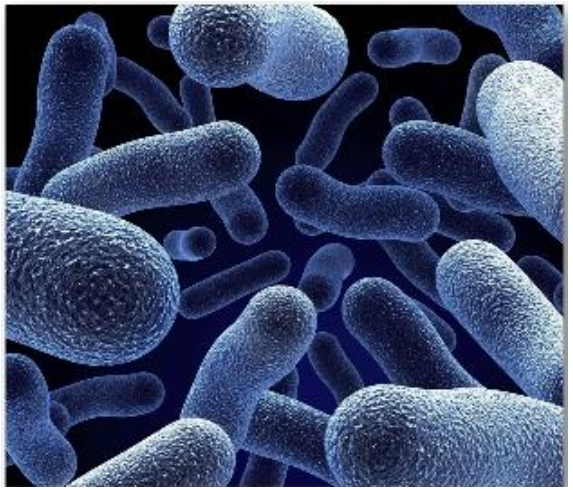
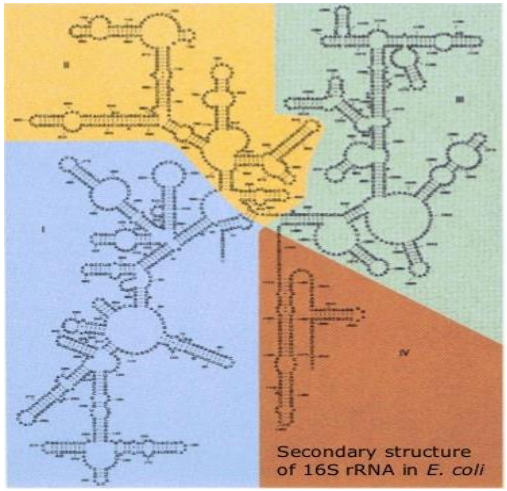
Technological
Educational
Institute of Athens

Molecular typing in Microbiology





Lecture 1b

Department of Medical laboratories

Theodoros Rampias

08/05/ 2017

Molecular typing

- Stable target
- Objective
- No specific growth requirements
- Easier to standartize

Criteria for evaluating typing systems

Typeability	Capacity to produce clearly interpretable results with most strains of the bacterial species
Reproducibility	Capacity to repeatedly obtain the same typing profile result with the same bacterial strain
Discriminatory power	Ability to produce results that clearly allow differentiation between unrelated strains of the same bacterial species
Practicality (ease of performance & interpretation)	Method should be versatile, relatively rapid, inexpensive, technically simple and provide readily interpretable results

GENETIC MARKER

- A genetic marker is gene or DNA sequence with a known location on a chromosome that cam be used to identify individual or species

- It may be a **short DNA sequence**

Taking advantage of evolution

The affinities of all the beings of the same class have sometimes been represented by a great tree... The green and budding twigs may represent existing species; and those produced during former years may represent the long succession of extinct species.

C. Darwin, 1872

Nothing in biology makes sense, except in the light of evolution.

T. Dobzhansky, 1973

http://tolweb.org

FEATURES

- Should be easy, fast and cheap to detect
- Should be reproducible
- Should be polymorphic
- Should have co-dominant inheritance to allow discrimination between homo and heterozygote in diploids

Powerful Genetic Markers：

Locus-specific
Co-dominant
PCR-based
Highly Polymorphic

Molecular Typing Techniques

- **Restriction analysis**
  - Plasmid profiling
  - Restriction fragment length polymorphism (RFLP)
  - Ribotyping
  - Pulse Field Gel Electrophoresis (PFGE)

- **PCR amplification of particular genetic targets**
  - Amplified fragment length polymorphism (AFLP)
  - Random Amplified Polymorphic DNA (RAPD)
  - Repetitive element PCR (Rep-PCR)
  - Variable number of tandem repeat (VNTR) analysis and
  - Multiple locus VNTR analysis (MLVA)

- **Sequencing-based methods**
  - Multilocus sequence typing (MLST)
  - Single nucleotide polymorphism (SNPs)

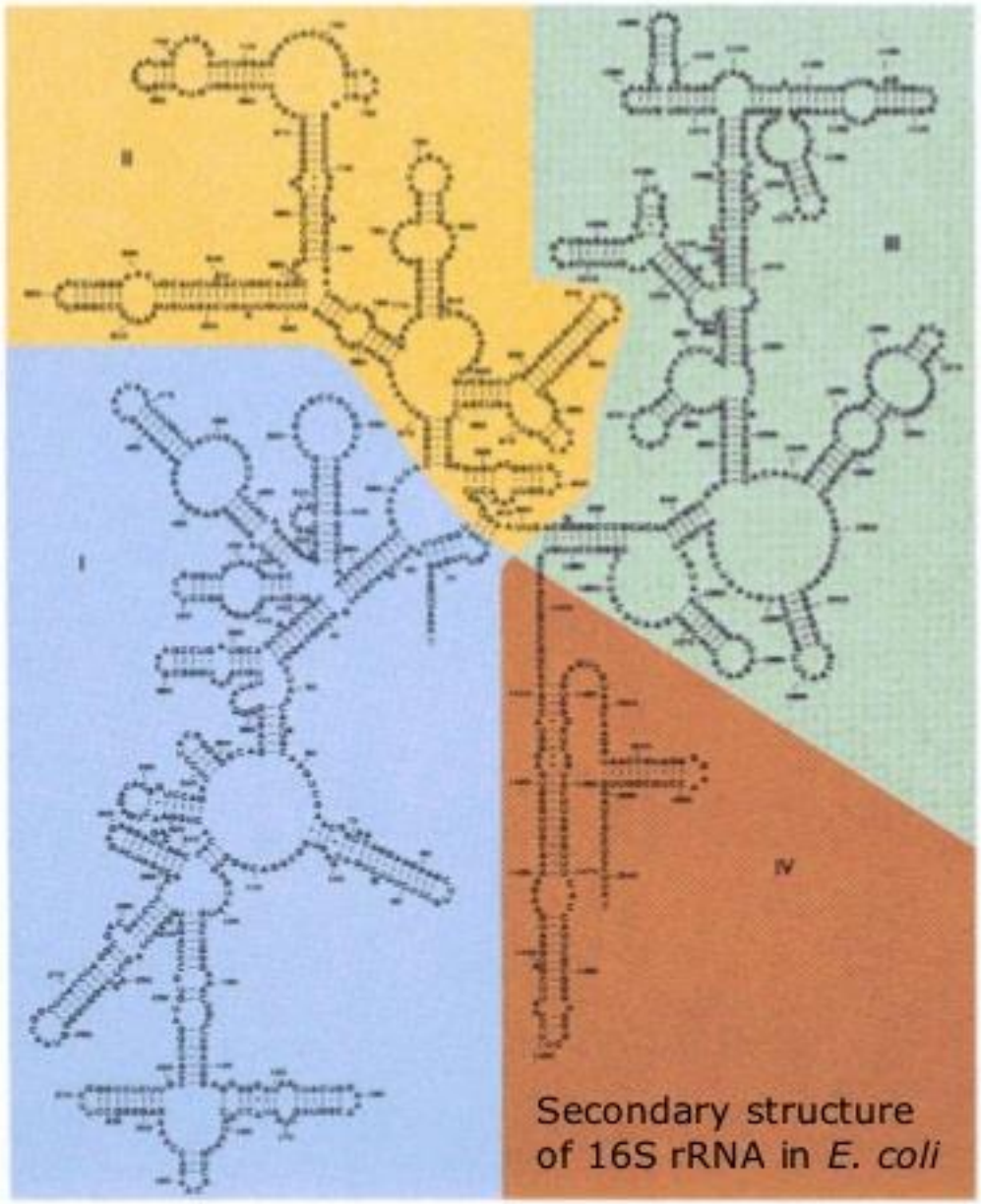
Molecular Phylogenetics

Step 1. Select a DNA region that is *homologous*, or similar across species due to common ancestry.

Ribosomal RNA (rRNA)

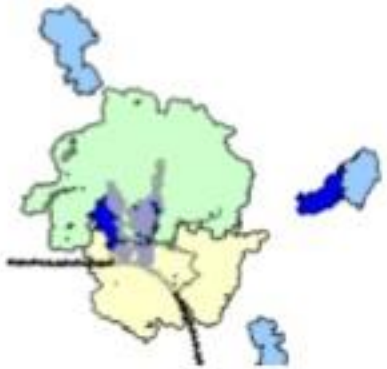
Ideal gene for phylogenetic studies because it :

- is an essential gene that is present in all organisms.
- is a common target for sequencing studies; large database for comparisons.
- contains sites that are relatively conserved (stems) and sites that are more free to vary (loops).



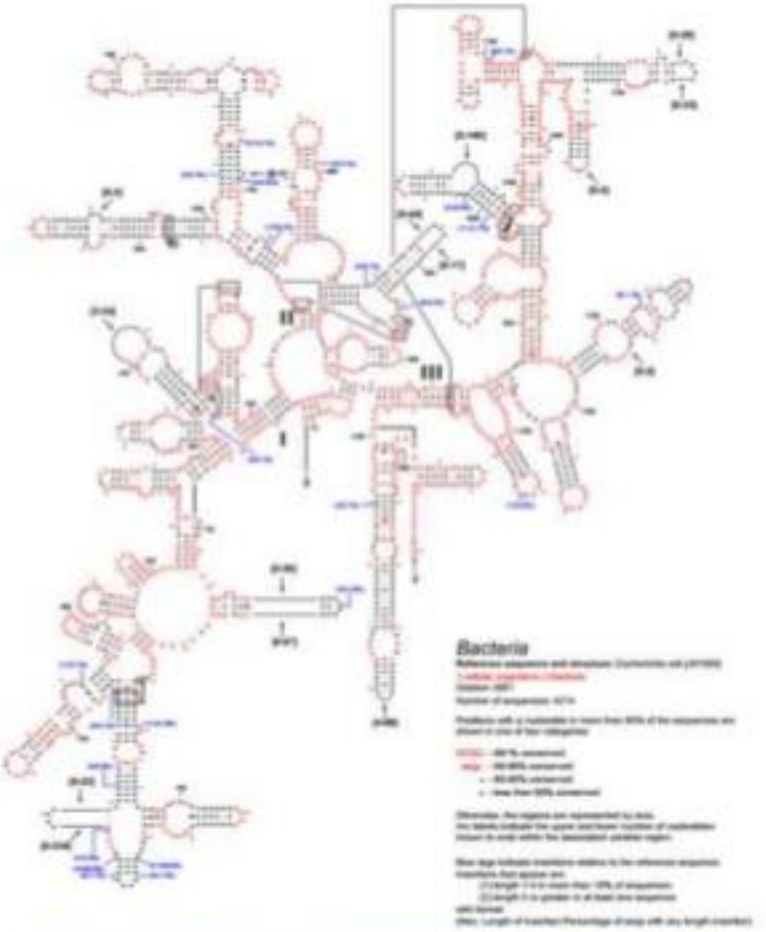
Secondary structure of 16S rRNA in *E. coli*

Why do we use the 16S gene?



Ribosomes are universal

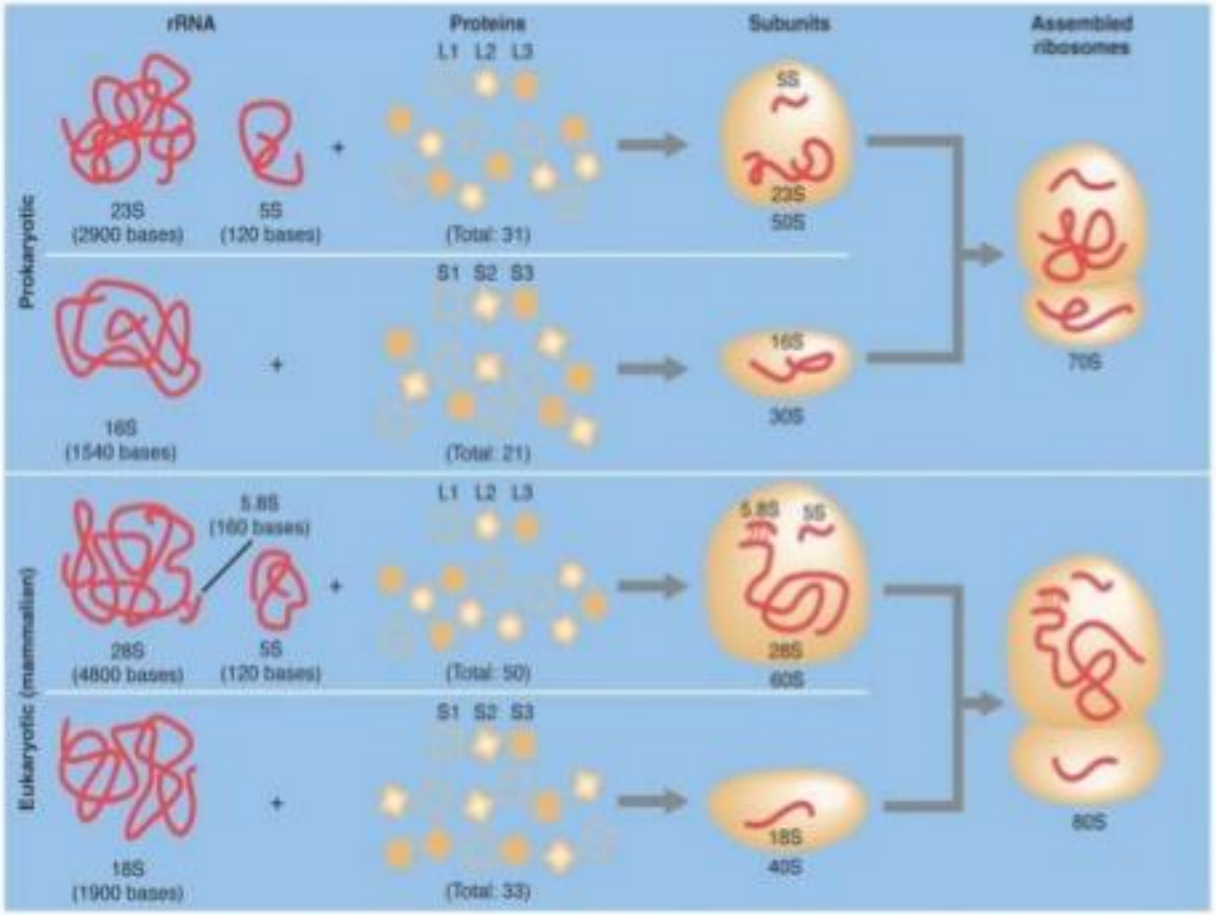
rRNA = Structural RNA



http://www.rna.icmb.utexas.edu/SAE/2B/ConsStruc/Diagrams/cons.16.b.Bacteria.pdf

Types

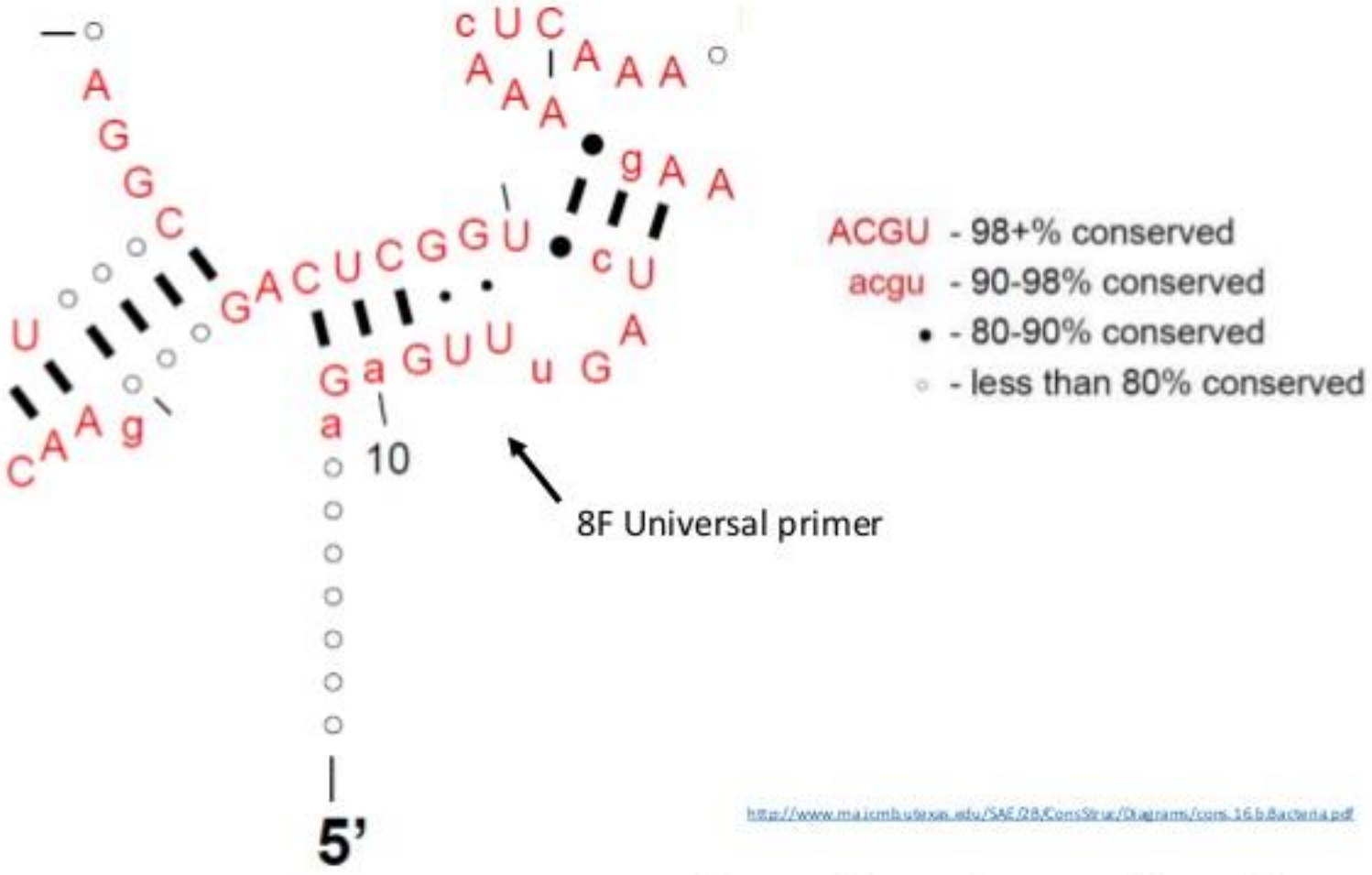
- In prokaryotes: 23S, 5S,16S
- In eukaryotes: 28S, 5.8S, 5S, 18S



Ribosomal RNAs in Prokaryotes:

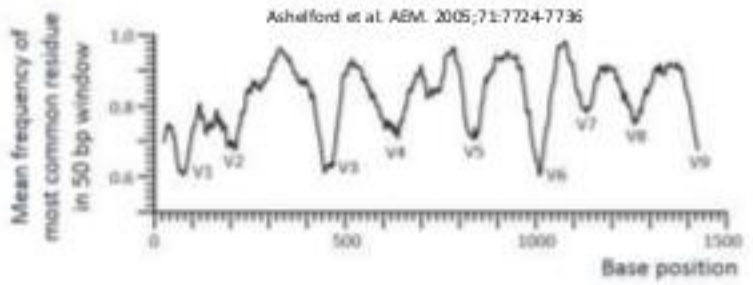
NAME	SIZE (NUCLEOTIDES)	LOCATION
5S	120	Large subunit of ribosome
16S	1500	Small subunit of ribosome
23S	2900	Large subunit of ribosome

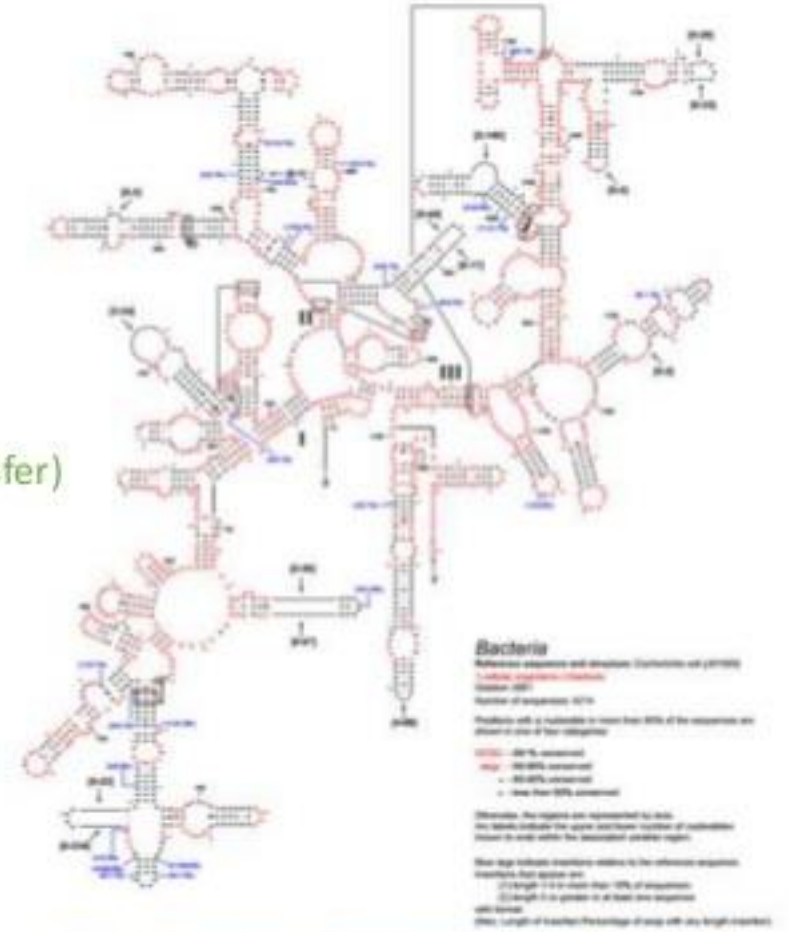
Why do we use the 16S gene?



http://www.rna.icmb.utexas.edu/SAE/2B/ConsStruc/Diagrams/cons.16.bBacteria.pdf

Why do we use the 16S gene?





http://www.rna.icmb.utexas.edu/SAE/2B/ConsStruc/Diagrams/cons.16.b.Bacteria.pdf

- Advantages:
  - Universal gene (No horizontal gene transfer)
  - Conserved regions
  - Variable regions
  - Great databases and alignments
- Problems:
  - Variable copy number
  - No universal (unbiased) primers
  - (Not directly correlated with activity)
  - (Lack of functional information)

- The 16s rDNA sequence has hypervariable regions, where sequences have diverged over evolutionary time.
- Strongly conserved regions often flank these hypervariable regions.
- Primers are designed to bind to conserved regions and amplify variable regions.
- The DNA sequence of the16S rDNA gene has been determined for an extremely large number of species. In fact, there is no other gene that has been as well characterized in as many species.
- Sequences from tens of thousands of clinical and environmental isolates are available over the Internet through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and the Ribosomal Database Project (http://rdp.cme.msu.edu/).
- These sites also provide search algorithms to compare new sequences to their database.

Genome

- Genes and regulatory sequences make up a small proportion of the genome
- The majority of DNA sequences in all higher eukaryotic genomes are repetitive sequences (50-90%)
- FUNCTION?

- Different sequence classes evolve at different rates

Major Genomic Components

	Typical Fraction
• Tandem Repeats	10%
• Simple Sequence Repeats	5%
• Dispersed Repeats	10%
• Functional Repeats	15%
• Retroelements	50%
• Genes	10%

Tandem repeats

- Polymorphism in chromosomal DNA can arise from the presence of a variable number of tandem repeats .
- These are short sequences of DNA at scattered locations in the genome, repeated in tandem (one after another).

Simple sequence repeats

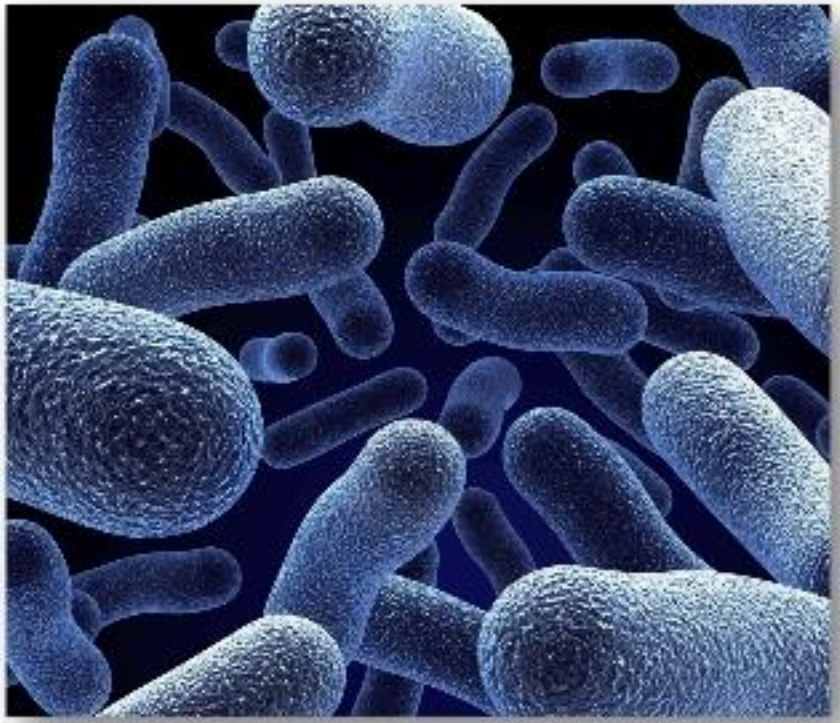
- GGCTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGGTCGTAATG
- Flanked by unique sequences (SSR/microsatellite markers) or
- Part of other repetitive elements

- Dispersed OR clustered in genome
- SSR markers are dispersed!

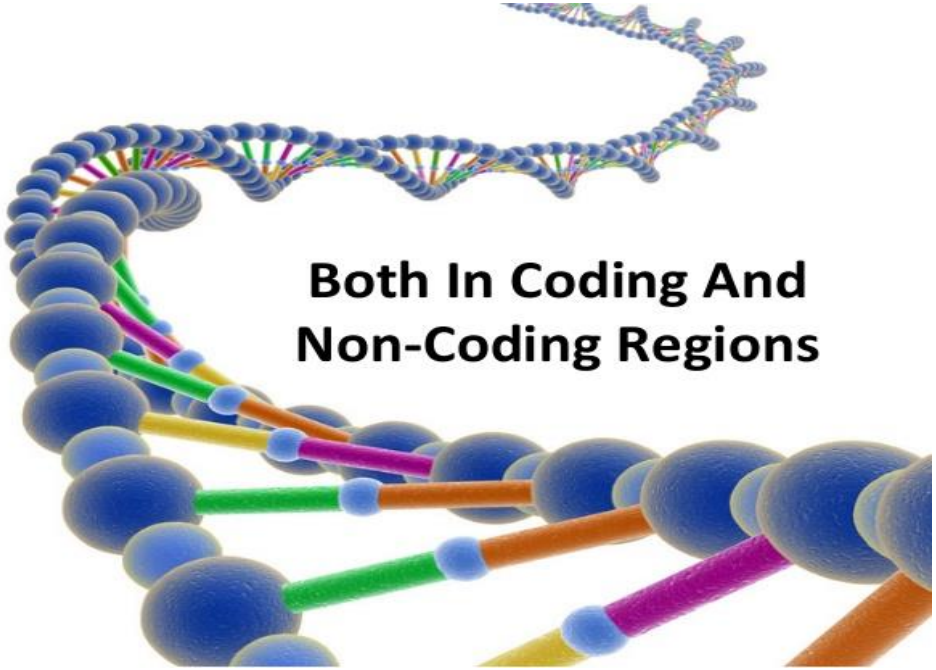
Microsatellites…

What is a microsatellite?

- Tandemly repeated DNA (**may see in the literature as STRs -** Short tandem repeats)
  - Poly A/T most common
  - 1-10 bp tandemly repeated = 'micro' satellite
  - >10 = 'mini' satellite
- Types of microsats
  - Di, tetra and tri nucleotide (**used in that order**)
  - Perfect
  - Imperfect/interrupted
  - Compound
    - Varying levels of variation associated with each type
    - Difficulty in scoring



Highly Abundant In Prokaryotes And Eukaryotes



Neutral?

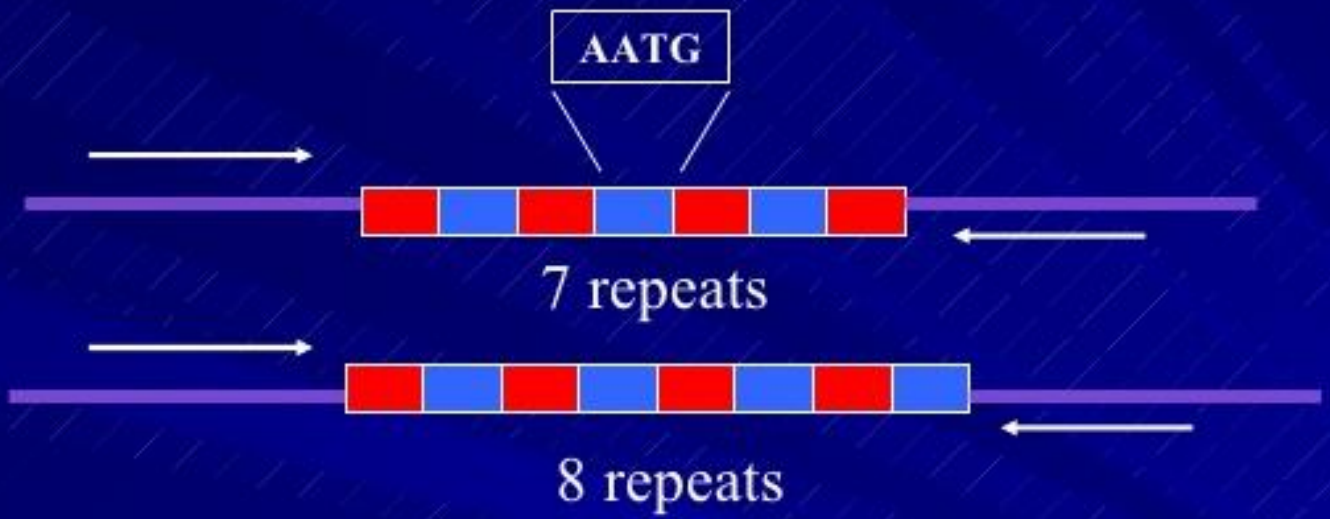
The Major Proportion Of SSRs

Intergenic Regions

Genic And Intragenic Regions


Highly
Variable

Short Tandem Repeats (STRs)



the repeat region is variable between samples while the flanking regions where PCR primers bind are constant

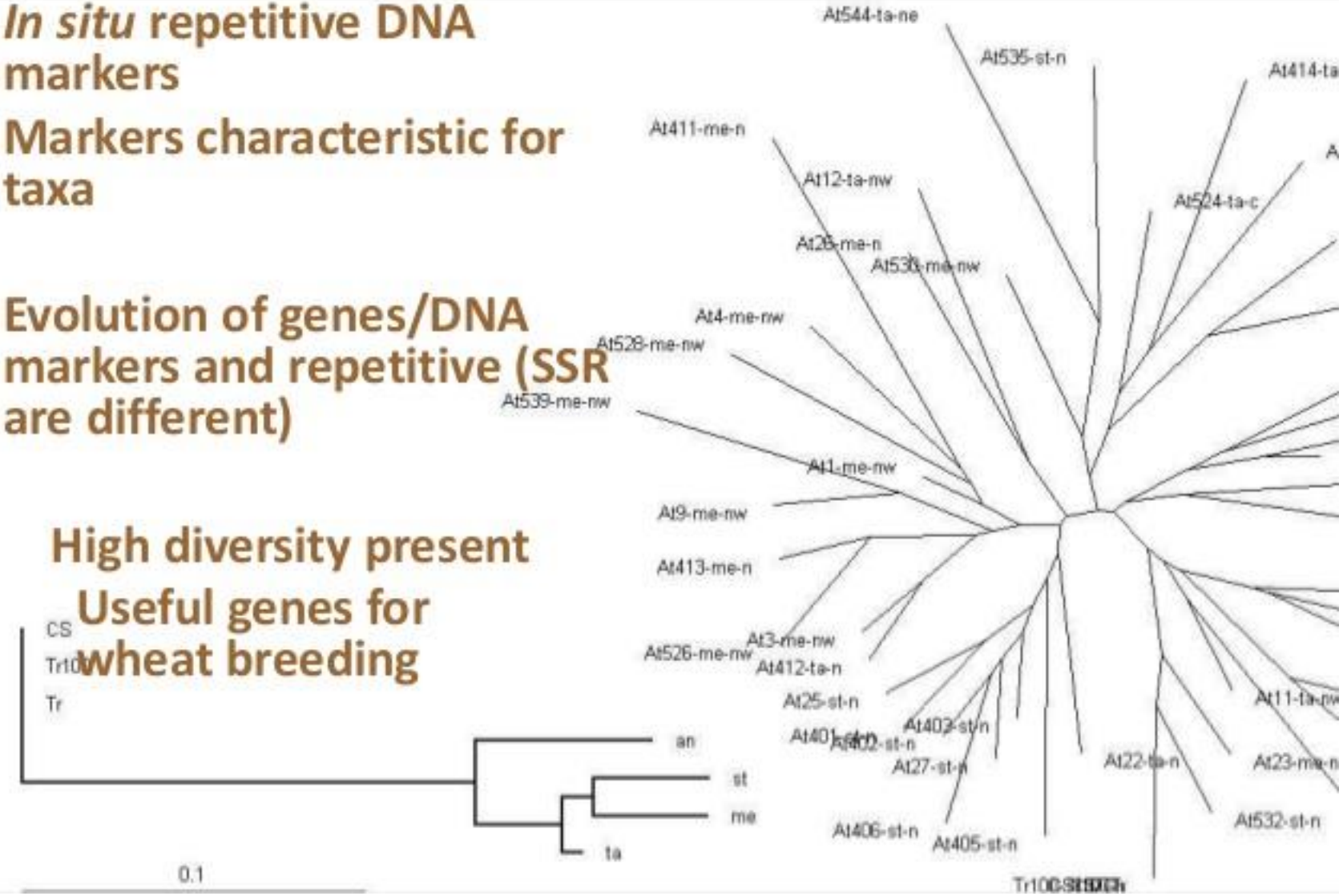
In situ repetitive DNA markers

Markers characteristic for taxa

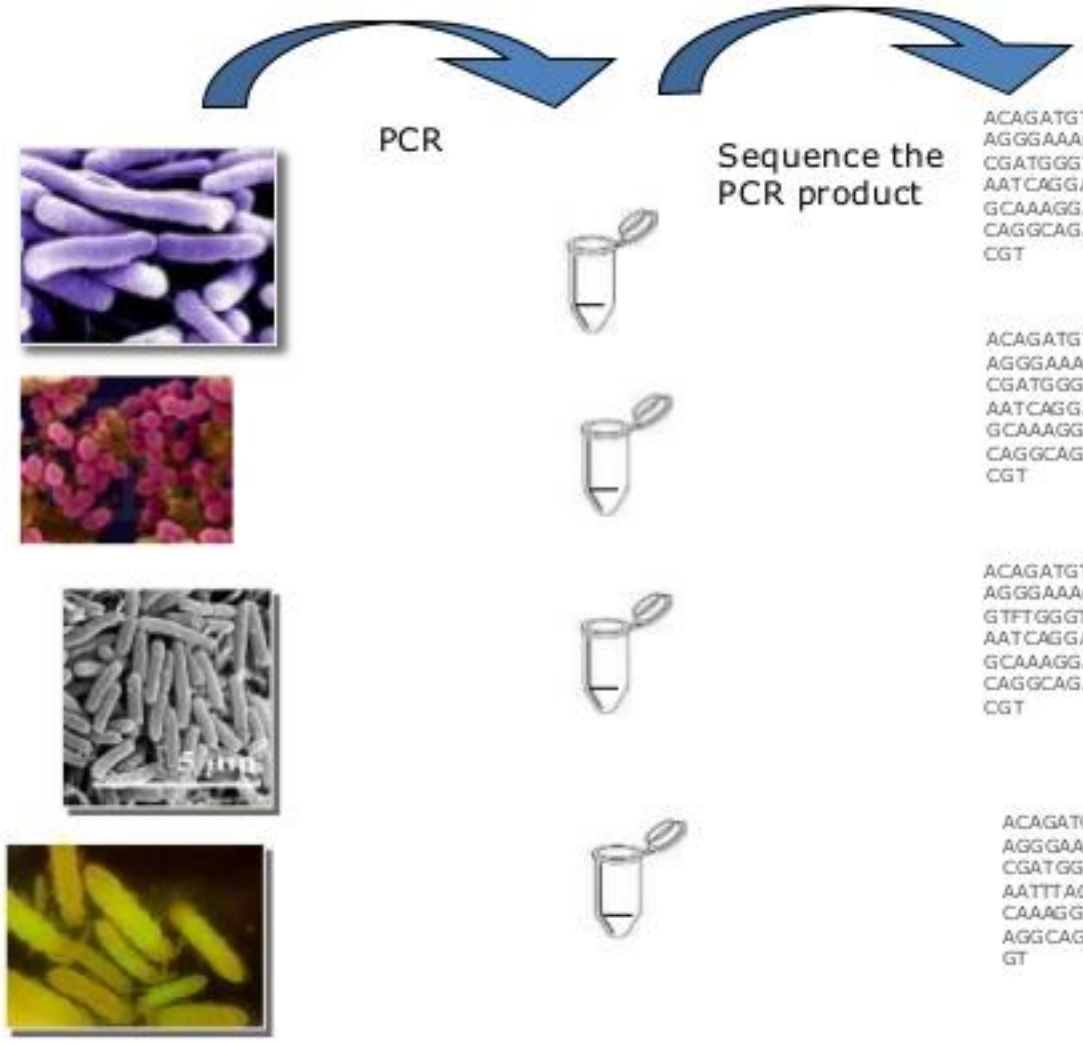
Evolution of genes/DNA markers and repetitive (SSR are different)

High diversity present

Useful genes for wheat breeding



2. **Amplify and Sequence** this region across isolates….



Restriction analysis-based typing

RFLP: Restriction Fragment length polymorphism

TFLP: Terminal Restriction Fragment length polymorphism

PFGE: Pulse Field Gel Electrophoresis

What are Restriction Enzymes?

- Also referred to as **Restriction Endonucleases**
- Enters and recognizes a certain sequence on a double helix strand of DNA, usually 4-6 base-pairs long, and cuts it
- Precise by cutting both strands in same location though strands move in reverse directions; REs are able to depict the precise spot to cut
- Able to restrict and destroy foreign DNA, such as viruses, preventing them from entering the cell
- Used in biotechnology for cutting DNA into smaller strands for research in gene cloning or fragment lengths among different individuals

Few Restriction Enzymes

Enzyme	Organism from which derived	Target sequence (cut at *) 5' -->3'
Bam HI	*Bacillus amyloliquefaciens*	G* G A T C C
Eco RI	*Escherichia coli RY 13*	G* A A T T C
Hind III	*Haemophilus inflenzae Rd*	A* A G C T T
Mbo I	*Moraxella bovis*	*G A T C
Pst I	*Providencia stuartii*	C T G C A * G
Sma I	*Serratia marcescens*	C C C * G G G
Taq I	*Thermophilus aquaticus*	T * C G A
Xma I	*Xanthamonas malvacearum*	C * C C G G G

Recognition Sequences

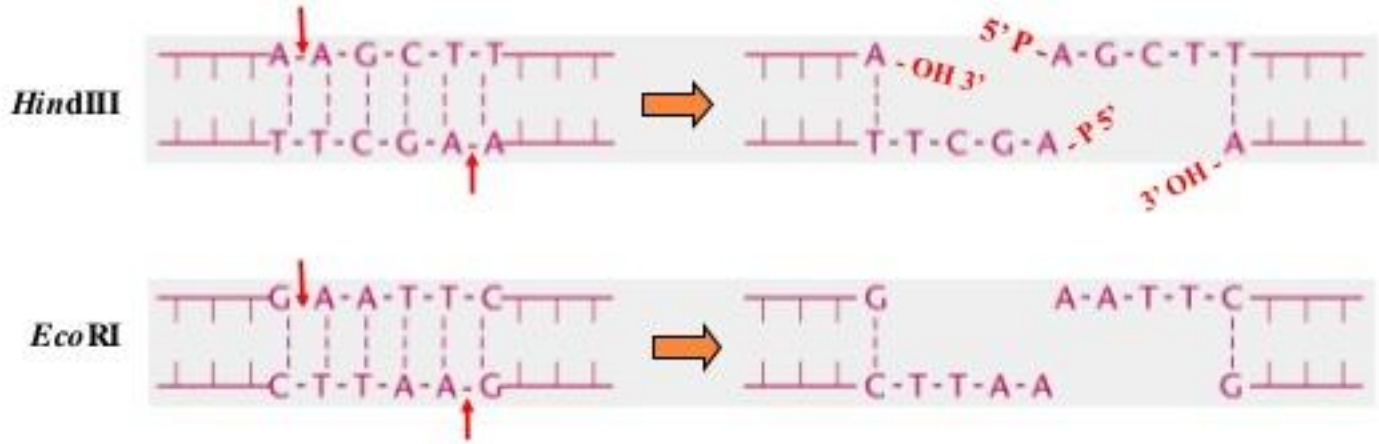
- Each restriction enzyme always cuts at the same recognition sequence.
- Produce the same gel banding pattern (fingerprint).
- Many restriction sequences are palindromic. For example.

5' GAATTC 3'
3' CTTAAG 5'

(Read the same in the opposite direction (eg. madam, race car…)

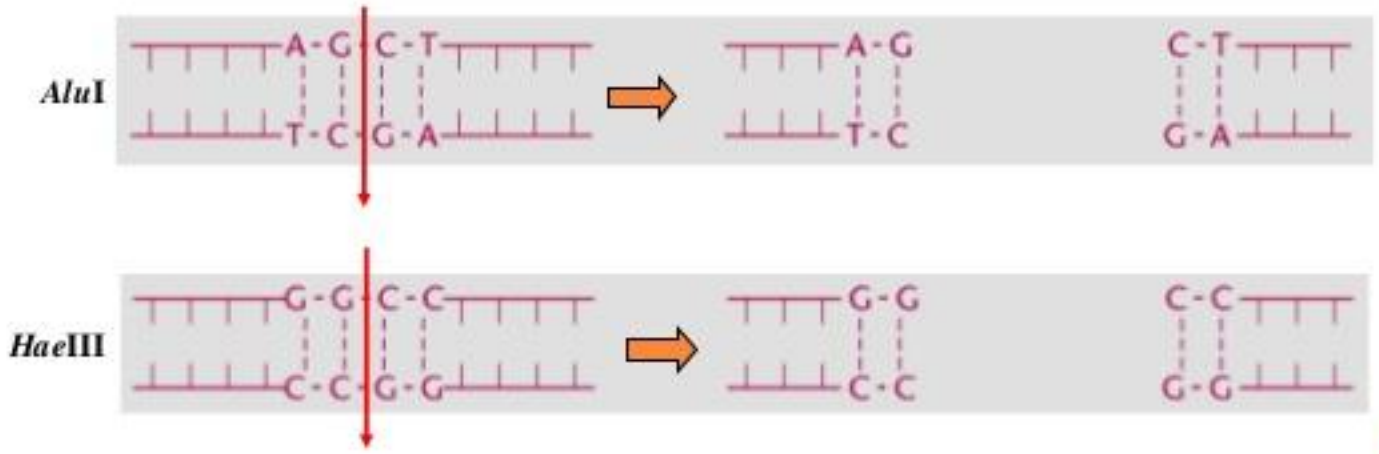
Sticky End Cutters

- Most restriction enzymes make staggered cuts.
- Staggered cuts produce single stranded "sticky-ends".
- DNA from different sources can be spliced easily because of sticky-end overhangs.



Blunt End Cutters

- Some restriction enzymes cut DNA at opposite base
- They leave blunt ended DNA fragments
- These are called blunt end cutters

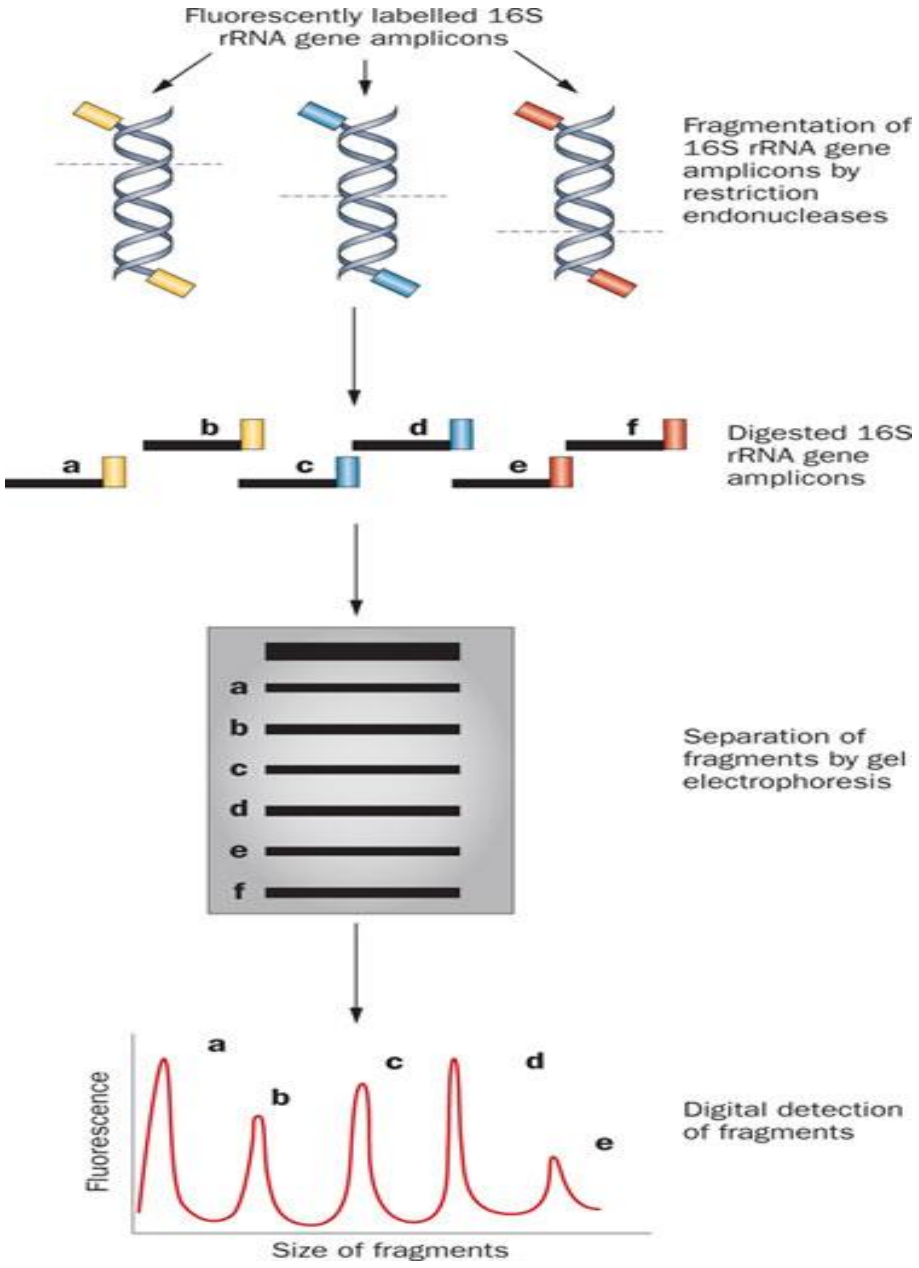


Restriction Fragment Length Polymorphism (RFLP)

- RFLP is a technique in which organisms may be differentiated by analysis of patterns derived from cleavage of their DNA.
- If two organisms differ in the distance between sites of cleavage of particular *Restriction Endonucleases,* the length of the fragments produced will differ when the DNA is digested.

- The similarity of the patterns generated can be used to differentiate species (and even strains) from one another.
- This technique is mainly based on the special class of enzyme i.e. Restriction Endonucleases.
- The variability of restriction sites have their origin in the DNA rearrangements, point mutations within the restriction enzyme recognition site sequences, insertions or deletions within the fragments, and unequal crossing over

RFLP


Fluorescently labelled 16S rRNA gene amplicons
Fragmentation of 16S rRNA gene amplicons by restriction endonucleases
a
b
c
d
e
f
Digested 16S rRNA gene amplicons
Separation of fragments by gel electrophoresis
Fluorescence
Size of fragments
Digital detection of fragments

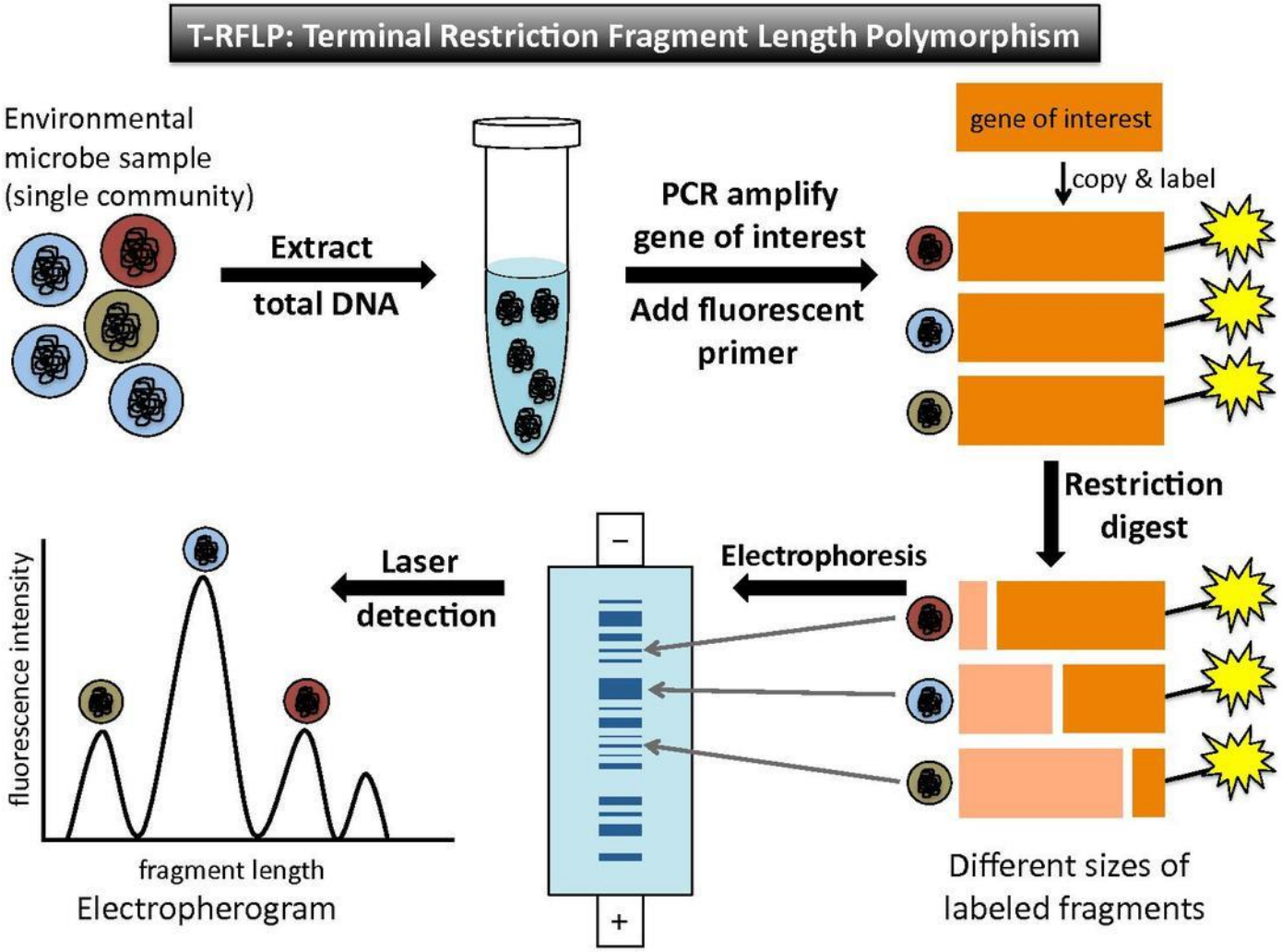
TRFLP Analysis

- TRFLP = (terminal restriction fragment length polymorphism analysis)
- A way to separate multiple PCR products of the same size. These products can be generated by a 16S-rRNA PCR of community DNA
- The PCR is performed as usual with two primers, but one is fluorescently labeled
- The PCR products are then cut up using a restriction enzyme
- The fluorescently labeled PCR pieces are detected
- TRFLP steps:

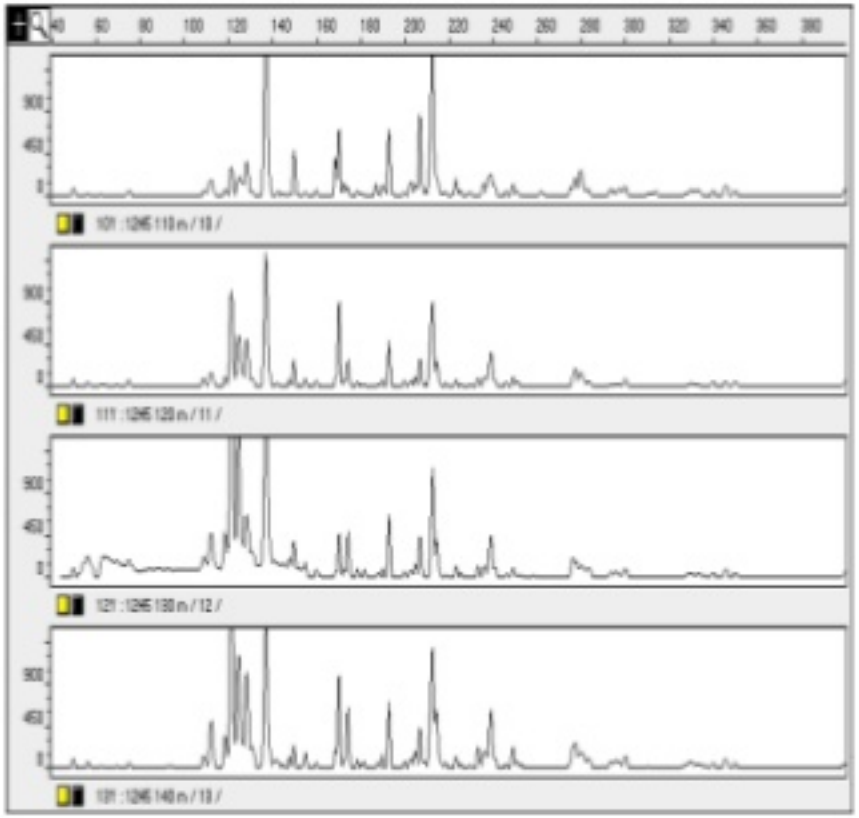
1. Extract DNA

2. Perform 16S rRNA PCR using fluorescently-labeled primer

3. Choose a restriction enzyme for TRFLP that will give the greatest diversity in restriction product size


T-RFLP: Terminal Restriction Fragment Length Polymorphism
Environmental microbe sample (single community)
Extract total DNA
PCR amplify gene of interest
Add fluorescent primer
gene of interest
copy & label
Restriction digest
Electrophoresis
Laser detection
fluorescence intensity
fragment length
Electropherogram
Different sizes of labeled fragments
–
+

TRFLP (cont.)



- Advantages
  - Very sensitive
  - Fast, easy and cheap
- Disadvantages
  - Can NOT cut bands to get sequence data
  - Requires capillary sequencer
  - Hard to distinguish noise from little peaks sometimes

Pulsed field gel electrophoresis (PFGE)

Conventional gel electrophoresis techniques:

- separates DNA fragments from 100 to 200 bp to 50 kilobase pairs (kb) only
- DNA(>50kb) cant be separated by this method.

- In 1982, Schwartz introduced the concept that DNA molecules larger than 50 kb can be separated by using two alternating electric fields.

- In conventional gels, the current is applied in a single direction (from top to bottom).

- But in PFGE, the direction of the current is altered at a regular interval.

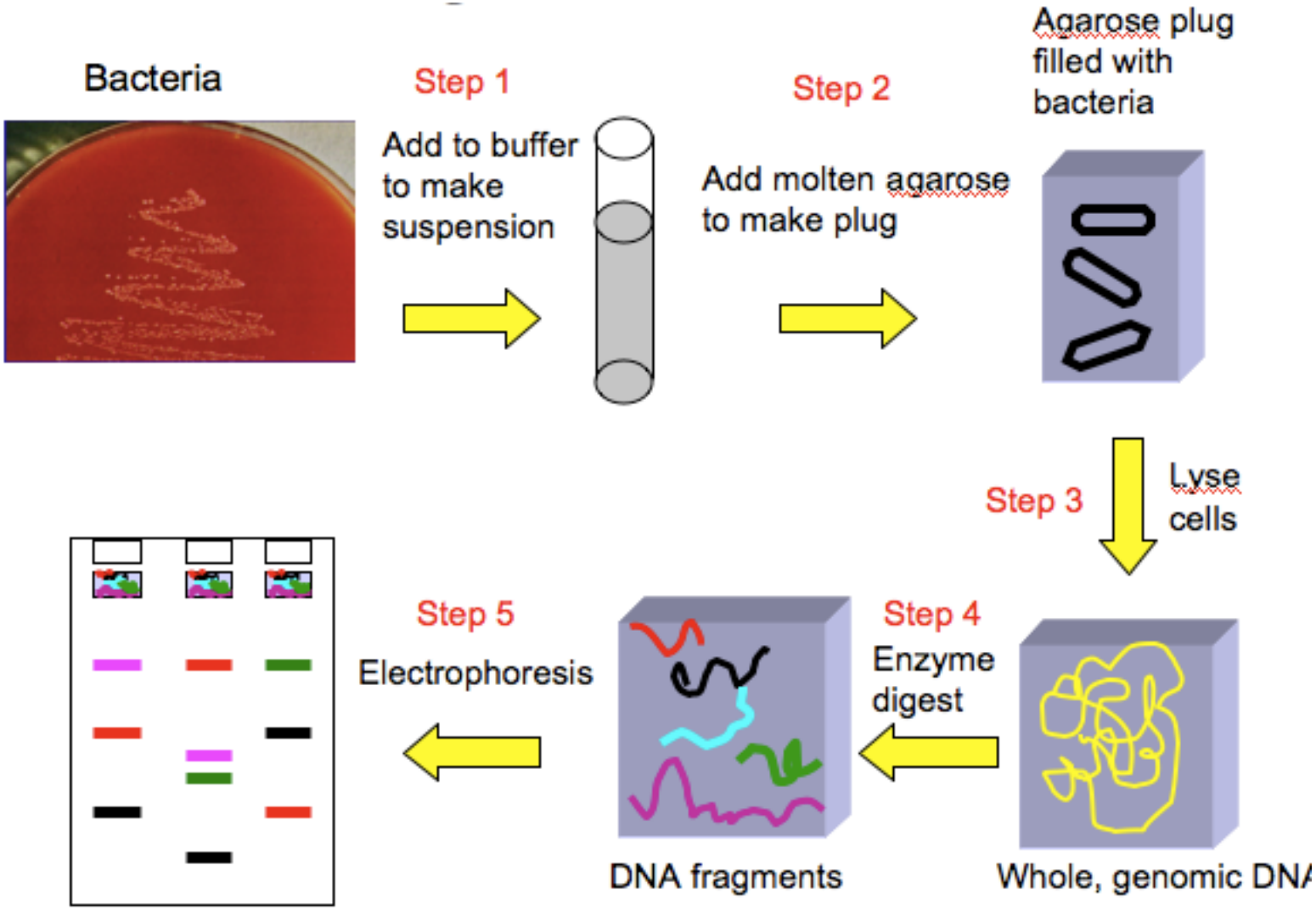
- Pulsed-field gel electrophoresis is based on the digestion of bacterial DNA with restriction endonucleases that recognize few sites along the chromosome, generating large DNA fragments (30-800 Kb)

- The basis for PFGE separation is the size-dependent time-associated reorientation of DNA migration achieved by periodic switching of the electric field in different directions.

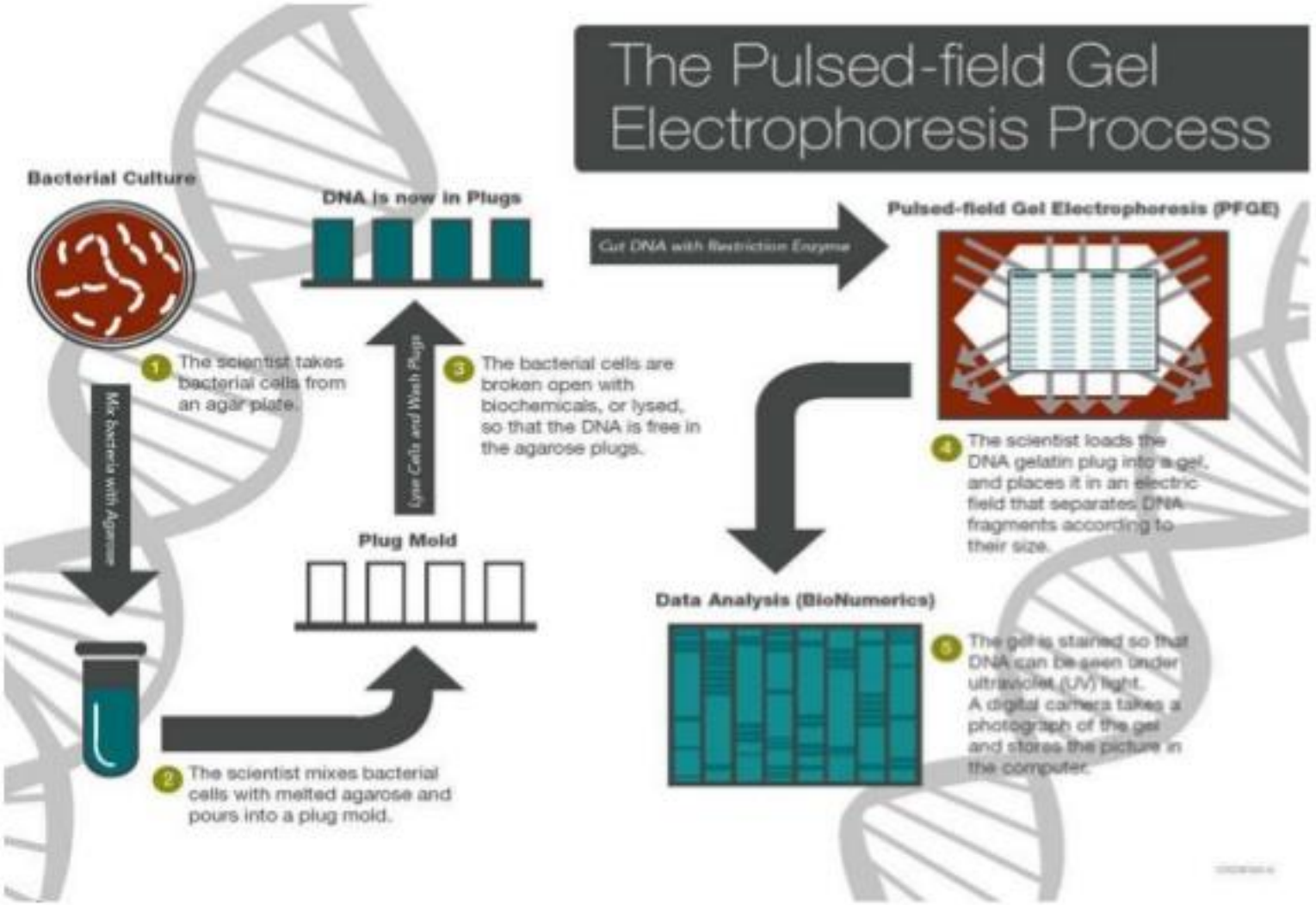
- The DNA fragments will form a distinctive pattern of bands in the gel, which can be analyzed visually and electronically.

- Bacterial isolates with identical or very similar band patterns are more likely to be related genetically than bacterial isolates with more divergent band patterns.

PFGE


Bacteria
Step 1
Add to buffer to make suspension
Step 2
Add molten agarose to make plug
Agarose plug filled with bacteria
Step 3
Lyse cells
Step 4
Enzyme digest
Step 5
Electrophoresis
DNA fragments
Whole, genomic DNA

The Pulsed-field Gel Electrophoresis Process



Advantages of PFGE

- PFGE has proved to be an efficient method for genome size estimation
- In PFGE DNA fragments obtained by using endonucleases produce a discrete pattern of bands useful for the fingerprinting and physical mapping of the chromosome.
- The PFGE technique is useful to establish the degree of relatedness among different strains of the same species.

Applications of PFGE

- PFGE is used for epidemiological studies of pathogenic organisms.
- PFGE is often employed to track pathogens, such as *Salmonella, Shigella, Escherichia coli (including O157), Campylobacter, and Listeria species*
- PFGE has remarkable discriminatory power and reproducibility. It is currently considered the strain typing method of choice for many commonly encountered pathogens.
- PFGE has proven extremely powerful in the analysis of large DNA molecules from a variety of sources including intact chromosomal DNAs from fungi, parasitic protozoa and specifically fragmented genomes of bacteria and mammal.

LIMITATIONS OF PFGE

- Time consuming (2-4 days)
- Requires a trained and skilled technician.
- Pattern results vary slightly between technicians.
- Don't really know if bands of same size are same pieces of DNA.
- Not applicable for all organisms.
- The choice of restriction enzyme may be important to optimize the results

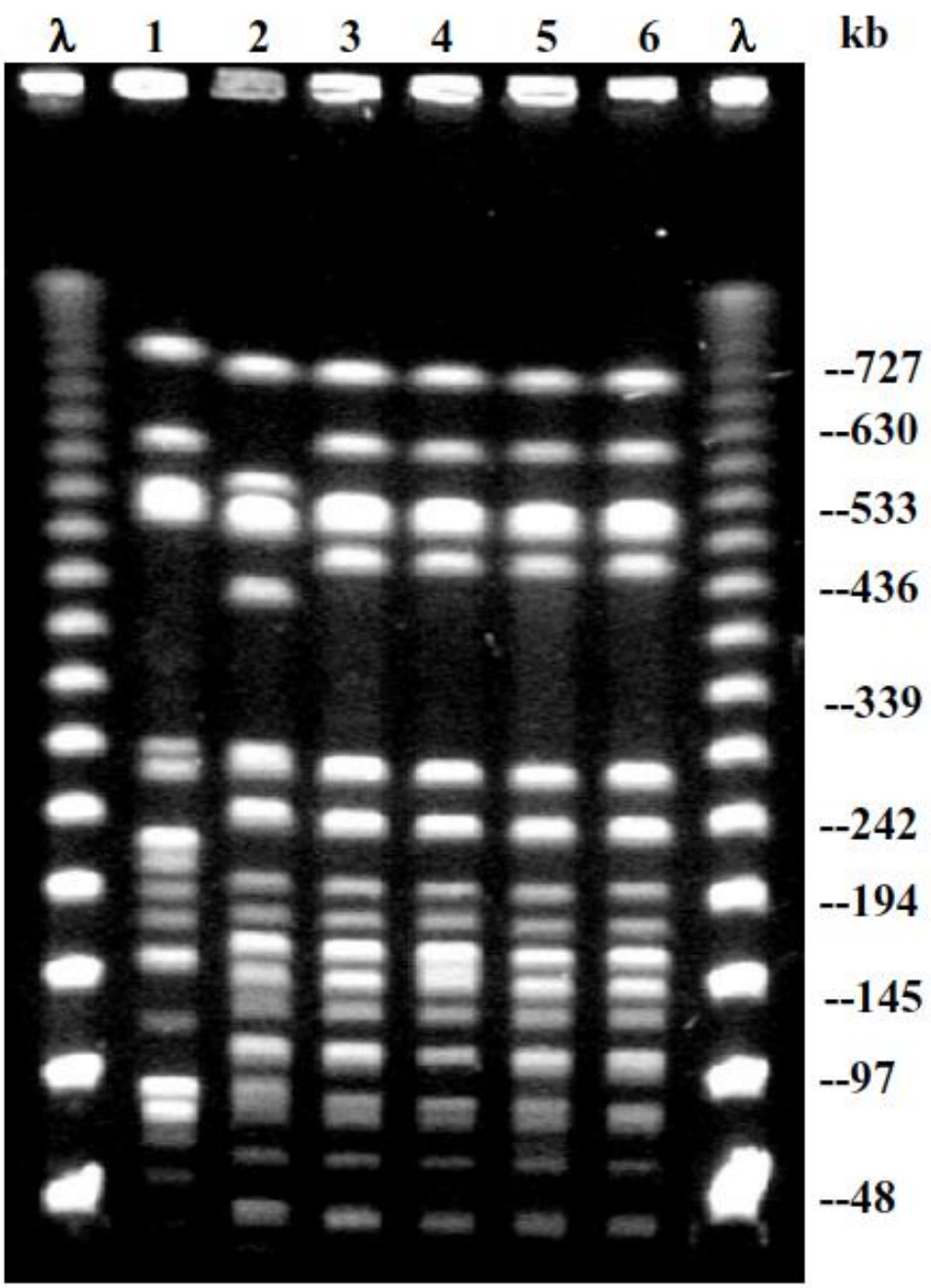
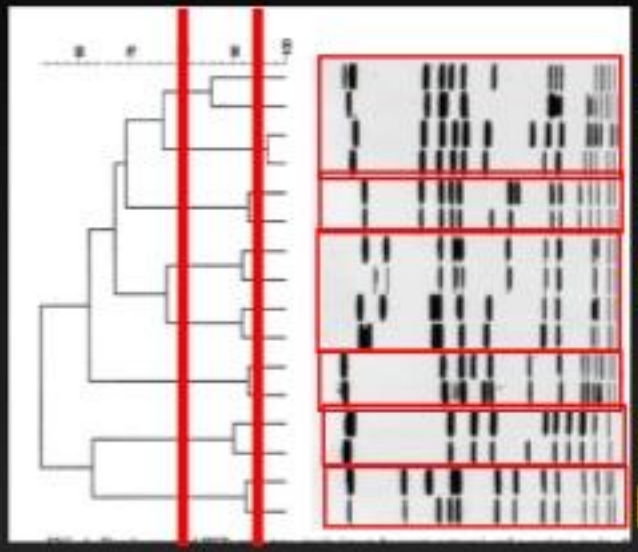


FIGURE 9.1. PFGE of *K. pneumoniae* isolates. Genetic profiles were obtained by digestion of chromosomal DNA with *Xba*I restriction endonuclease. Lanes 2 to 6 represent chromosomal DNA patterns of *K. pneumoniae* isolates recovered from five patients during an outbreak. Lane 1 shows chromosomal DNA from unrelated isolates obtained from other hospital units during the course of the outbreak. Lane λ shows the molecular weight standard.

Typing methods: types / subtypes

PFGE :
PFGE Type (cut-off 80% DICE/UPGMA)
PFGE Subtype (cut-off 80% DICE/UPGMA)

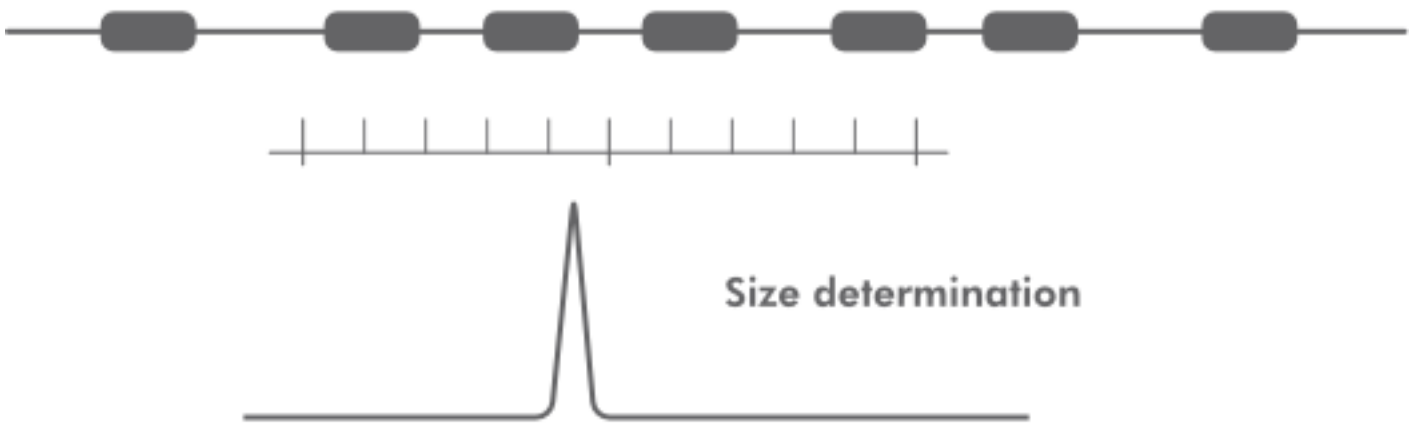


Dis advantage

- Require large quantities of high molecular weight DNA.
- Expensive process
- Time consuming
- Labor intensive

PCR- based typing

- Variable number of tandem repeat (VNTR) analysis and
- Multiple locus VNTR analysis (MLVA)



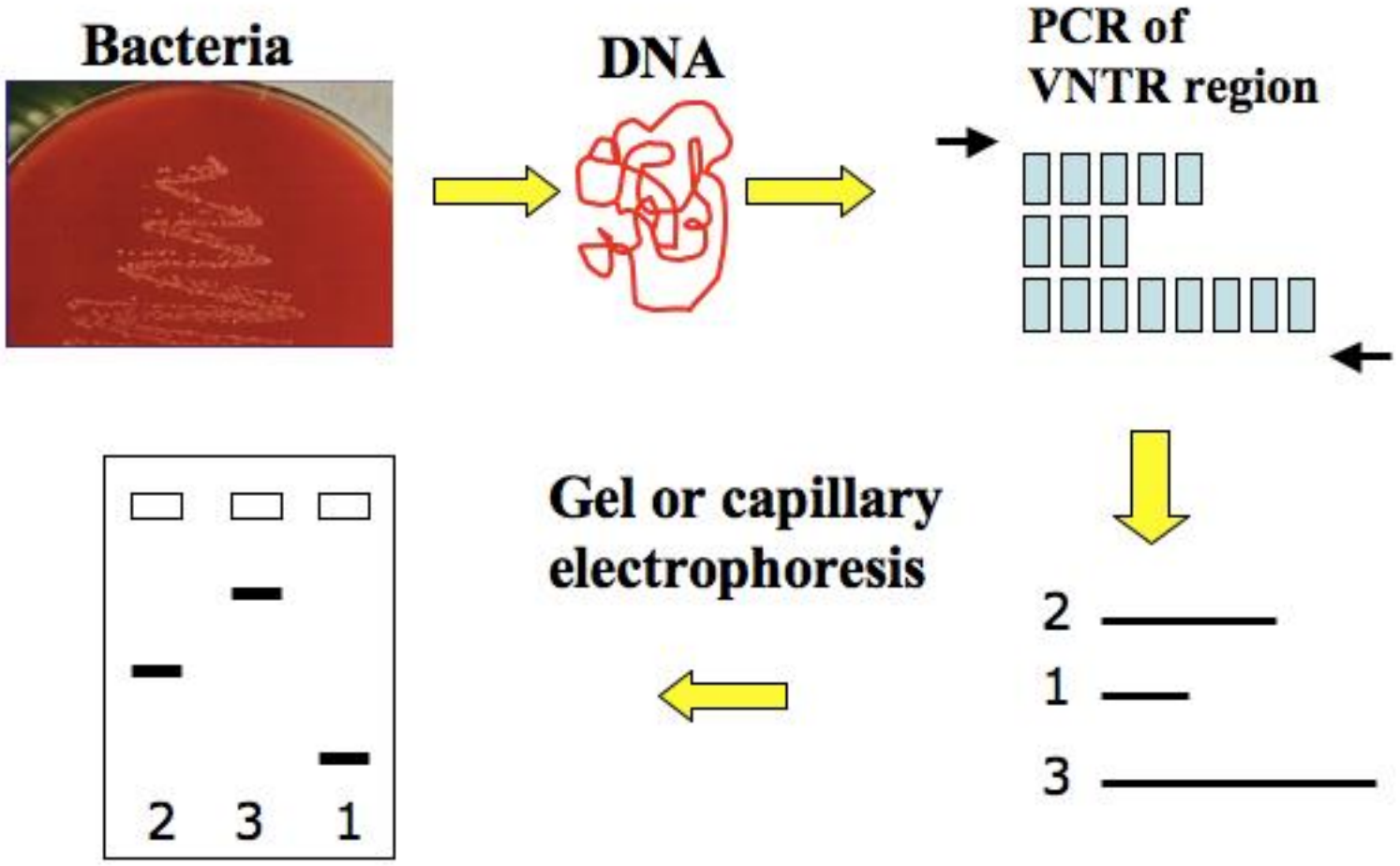
Strain A

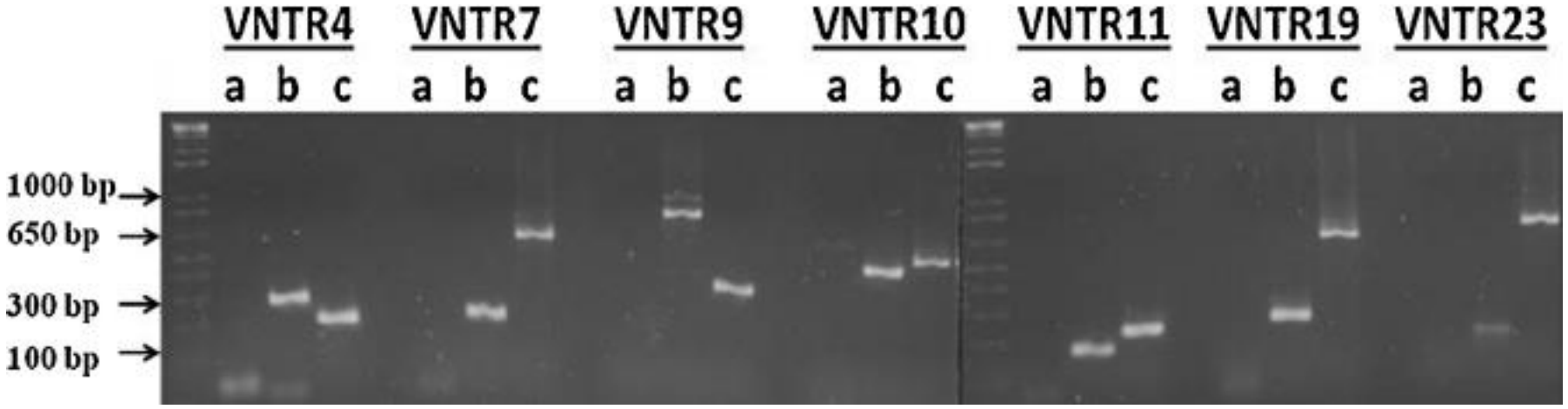
VNTR code: 5-3-7-4

Strain B

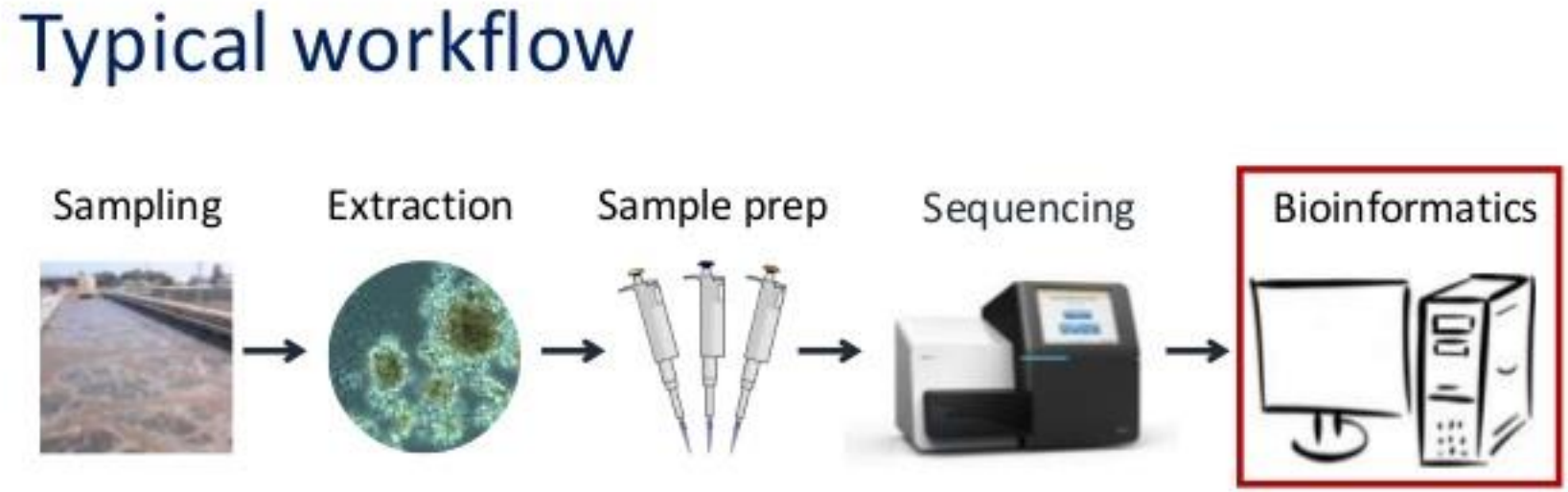
VNTR code: 3-4-9-2

MLVA : MULTI LOCUS VNTR ANALYSIS

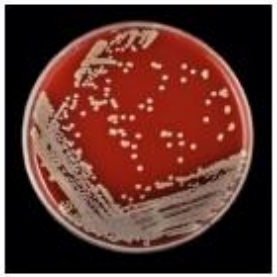



VNTR4
VNTR7
VNTR9
VNTR10
VNTR11
VNTR19
VNTR23
a b c a b c a b c a b c a b c a b c a b c
1000 bp
650 bp
300 bp
100 bp

Sequencing approaches



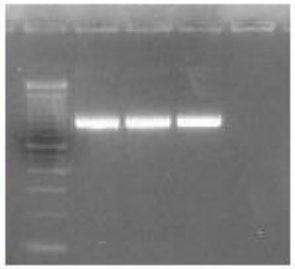
Protocol:



1. Bacteria in pure culture

2. DNA isolation

3. PCR amplification
(standard program, universal primers)



4. Agarose gel electrophoresis

5. Purify PCR products

6. Sequence

How does this organism fit into the world of available sequence data?

PCR

Sequence the PCR product

ACAGATGTCTTGTAATCCGGCCGTTGGTGGCAT
AGGGAAAGGACATTTAGTGAAAGAAATTGATG
CGATGGGTGGATCGATGGCTTATGCTATCGATC
AATCAGGAATTCAATTTAGAGTACTTAATAGTA
GCAAAGGAGCTGCTGTTAGAGCAACACGTGCT
CAGGCAGATAAAATATTATATCGTCAAGCAATA
CGT

"Blast" sequence to Genbank

Blast output:
Lists sequences that are most similar to yours

GENBANK = NIH genetic database with all publicly available DNA sequences. As of 2004: > 44 billion bp, and > 40 million sequences

BLAST ANALYSIS

Sequence alignment

This is a hard example.

That is another easy example.

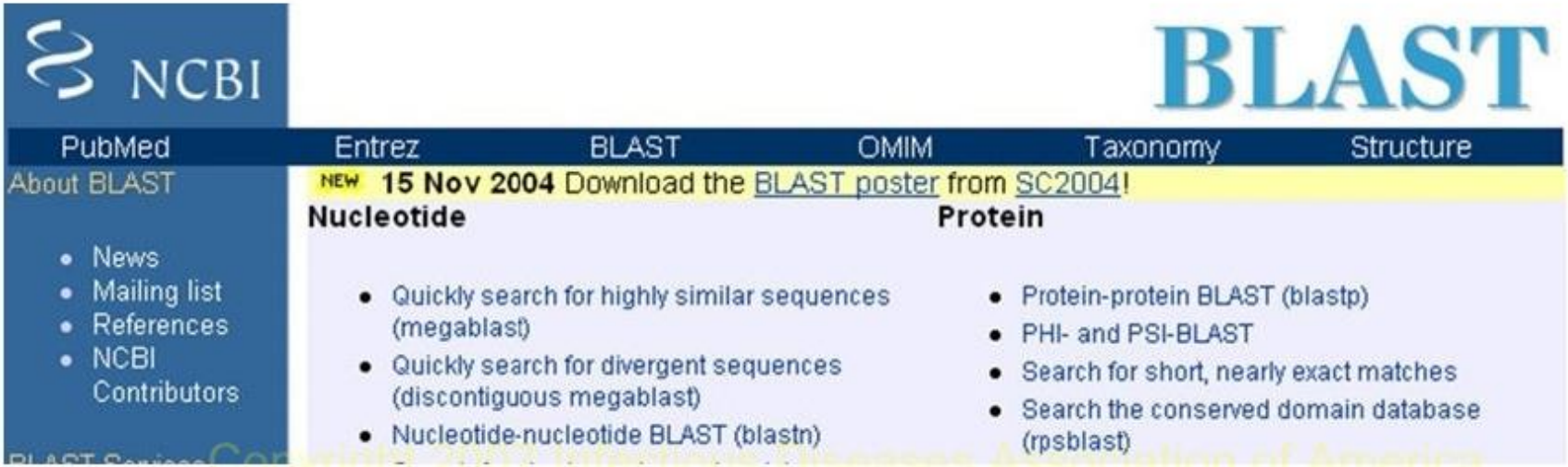
```
This is a --hard---- example.
||  |||||   | |     |||||||||
That is another easy example.
```

```
This is a-- h-ard---- example.
||  |||||   |  |     |||||||||
That is anothe-r easy example.
```

```
This is a hard example.------
||  |||||         |   |
That is another easy example.
```

Protocol:

- NCBI GenBank webpage: http://www.ncbi.nlm.nih.gov/BLAST/
  - an annotated collection of nucleotide sequences
  - short sequences to whole genomes
  - open access
- Nucleotide-nucleotide BLAST
  - paste in the linear sequence data, submit
  - search is performed
  - list of matches is provided



Comparisment of assembled 16S rRNA sequences with database library

- Genbank database

Description	Max score	Total score	Query coverage	E value	Max ident
Uncultured bacterium clone SJTU F 04 61 16S ribosomal RNA gene,	2575	2575	100%	0.0	99%
Streptococcus salivarius strain ATCC 7073 16S ribosomal RNA gene,	2575	2575	100%	0.0	99%
Uncultured bacterium clone RL181 aah39e03 16S ribosomal RNA gen	2571	2571	99%	0.0	99%
Uncultured bacterium clone SJTU C 06 81 16S ribosomal RNA gene,	2569	2569	100%	0.0	99%
Uncultured bacterium clone SJTU D 10 61 16S ribosomal RNA gene,	2569	2569	100%	0.0	99%
Uncultured bacterium clone SJTU D 09 61 16S ribosomal RNA gene,	2569	2569	100%	0.0	99%
Uncultured bacterium clone SJTU D 06 49 16S ribosomal RNA gene,	2569	2569	100%	0.0	99%
Uncultured bacterium clone 014B-B9 16S ribosomal RNA gene, partial	2569	2569	100%	0.0	99%
Uncultured bacterium clone RL182 aah31c06 16S ribosomal RNA gen	2566	2566	99%	0.0	99%
Uncultured bacterium clone SJTU C 03 56 16S ribosomal RNA gene,	2564	2564	100%	0.0	99%
Uncultured bacterium clone SJTU D 06 44 16S ribosomal RNA gene,	2564	2564	100%	0.0	99%
Streptococcus salivarius 16S ribosomal RNA gene, partial sequence	2564	2564	100%	0.0	99%
Streptococcus salivarius 16S rRNA gene, clone 1V5	2564	2564	100%	0.0	99%
Uncultured bacterium partial 16S rRNA gene, isolate BF0002A102	2560	2560	99%	0.0	99%
Uncultured bacterium clone RL182 aah34f08 16S ribosomal RNA gene	2560	2560	99%	0.0	99%
Uncultured bacterium clone 002C-h1 16S ribosomal RNA gene, partial	2560	2560	99%	0.0	99%
Uncultured bacterium clone SJTU C 09 54 16S ribosomal RNA gene,	2558	2558	100%	0.0	99%
Uncultured bacterium clone SJTU C 02 44 16S ribosomal RNA gene,	2558	2558	100%	0.0	99%
Uncultured bacterium clone SJTU F 03 24 16S ribosomal RNA gene,	2558	2558	100%	0.0	99%

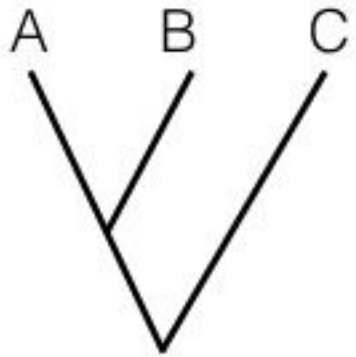
Identification criteria

≥ 99% sequence similarity	≥97% and < 99% of sequence similarity	≥95% and <97% of sequence similarity
Species level	Genus level	Family level

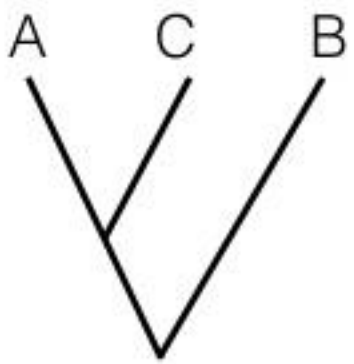
INPUT

1) AGT
2) AT
3) ATC

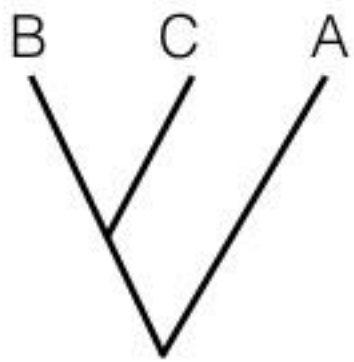
1) AGT
2) A–T
3) ATC



1) AGT
2) AT–
3) ATC

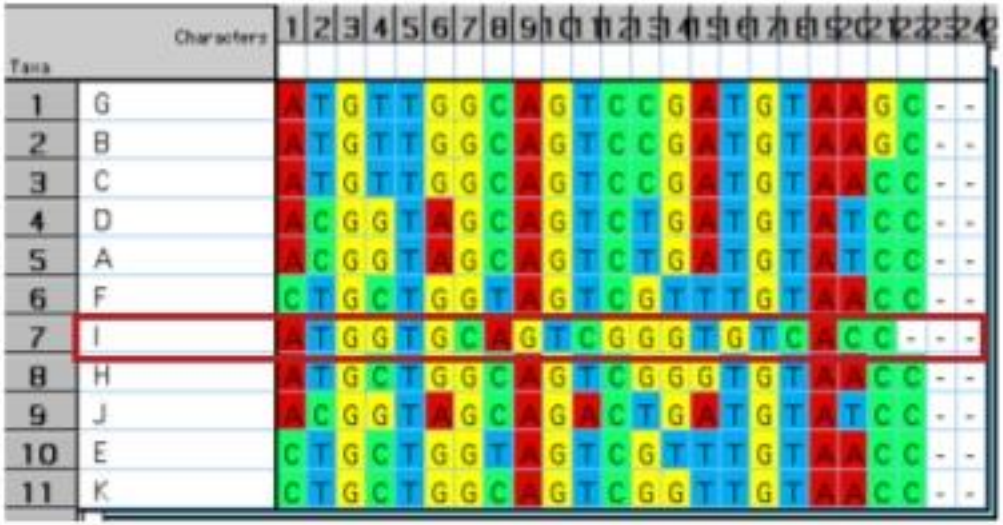


1) AGT -
2) A - T -
3) A - TC

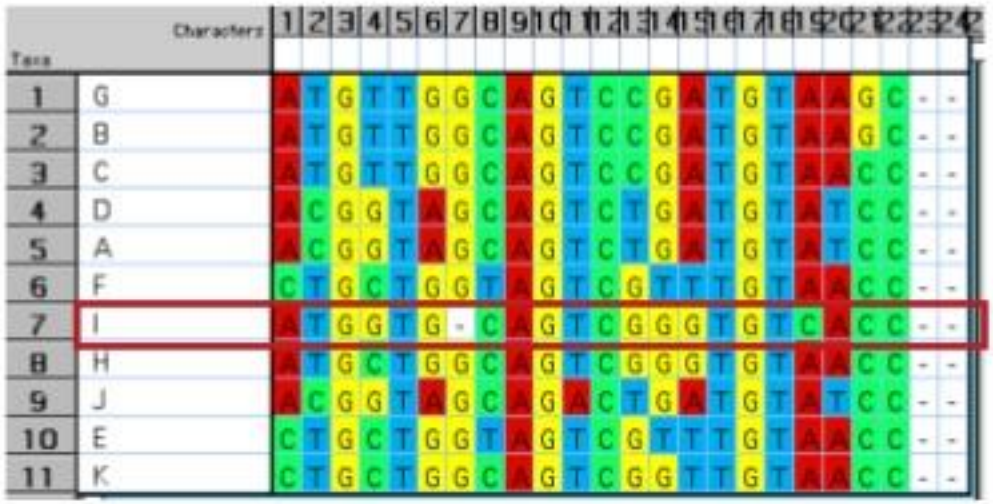


3. Sequence alignment is crucial for inferring how DNA sites have changed.

Poor alignment
Implies that species "I" is divergent from the others, but this is not the case.



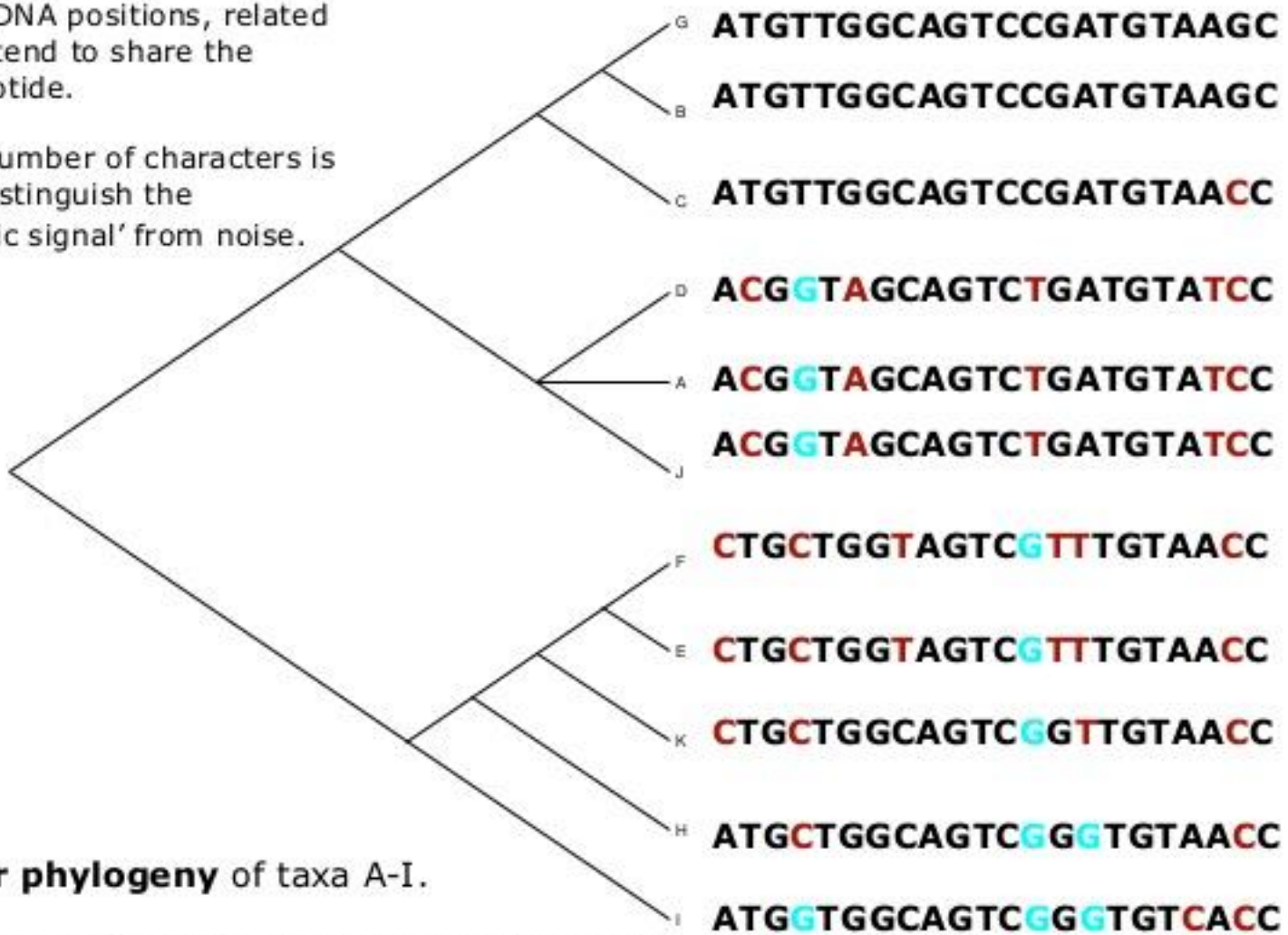
Good alignment.
Species "I" has probably experienced a deletion event at position #6 or #7.



4. Estimate relationships based on extent of DNA similarity.

At variable DNA positions, related groups will tend to share the same nucleotide.

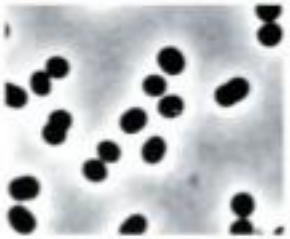
The sheer number of characters is helpful to distinguish the 'phylogenetic signal' from noise.



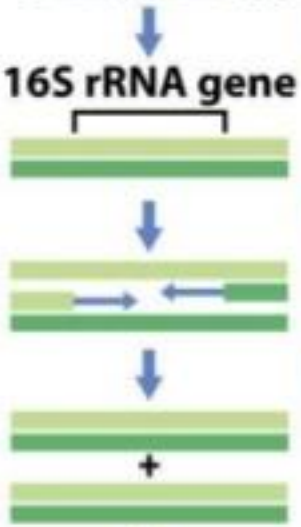
Molecular phylogeny of taxa A-I.

Colored letters = different from top sequence (taxon G)

rRNA gene sequencing



16S rRNA gene



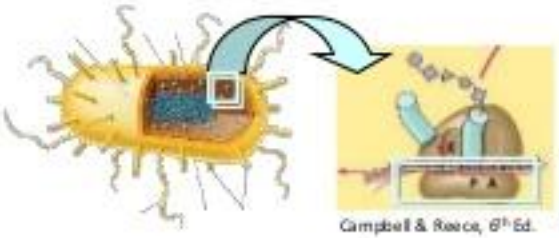
1. Isolate DNA
2. Heat to separate strands; add specific primers
3. Primer extension with DNA polymerase
4. Repeat above steps to obtain many copies of 16S rRNA gene
5. Run agarose gel and check for correct sized product
6. Purify and sequence PCR product

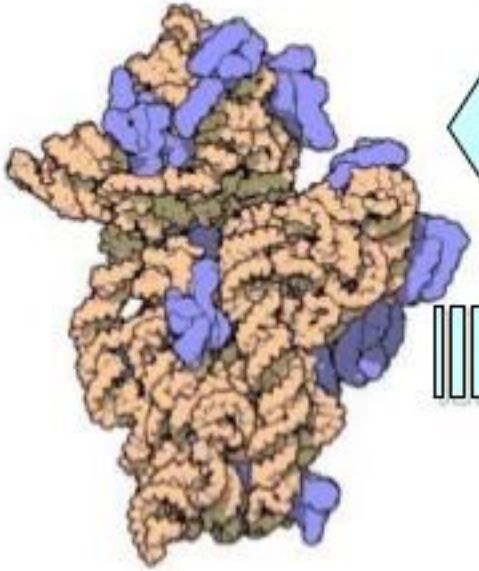
Figure 11-12 Brock Biology of Microorganisms 11/e

Use of primers to copy the 16S rDNA gene in bacteria

Bacterium with ribosomes

Ribosome synthesizing a protein

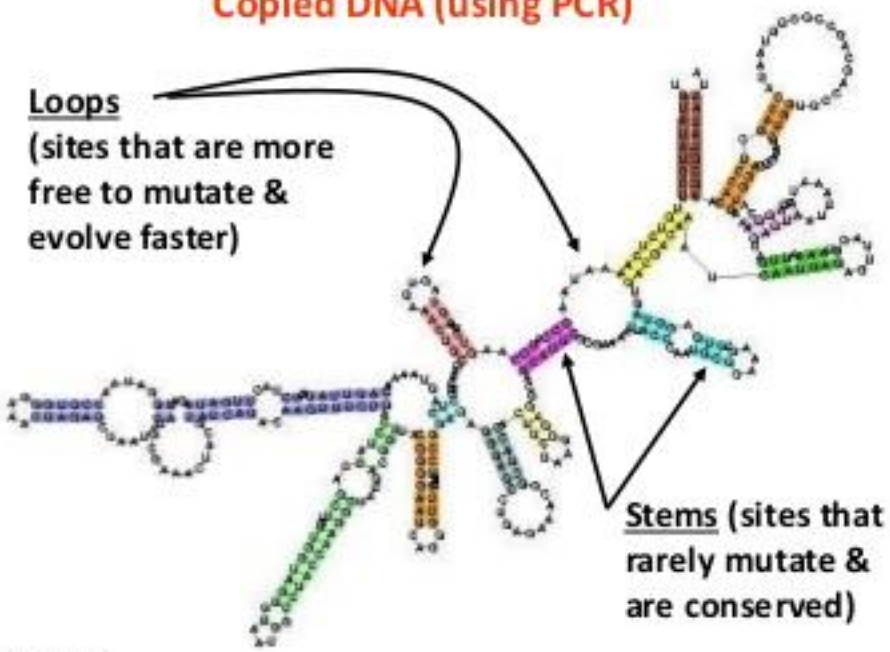




Atomic structure of the small subunit a ribosome. The rRNA, shown in orange, helps match the mRNA (codon) to the tRNA (anticodon).

16S rDNA gene – codes for making SSU rRNA

Copied DNA (using PCR)

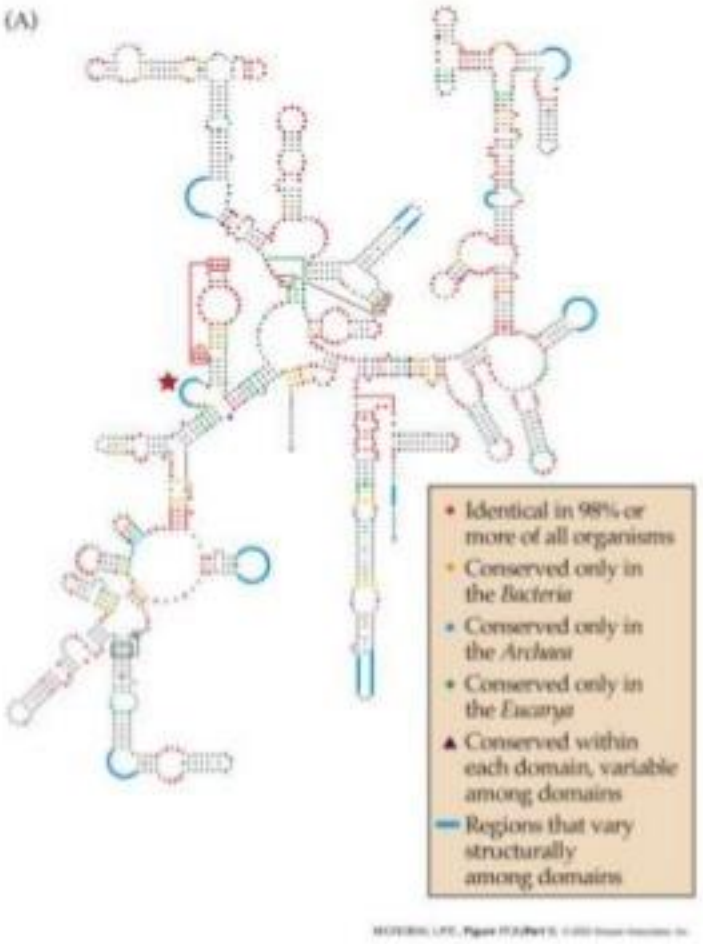


Small subunit ribosomal RNA

- The 16s rDNA sequence has hypervariable regions, where sequences have diverged over evolutionary time.
- Strongly conserved regions often flank these hypervariable regions.
- Primers are designed to bind to conserved regions and amplify variable regions.
- The DNA sequence of the16S rDNA gene has been determined for an extremely large number of species. In fact, there is no other gene that has been as well characterized in as many species.
- Sequences from tens of thousands of clinical and environmental isolates are available over the Internet through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and the Ribosomal Database Project (http://rdp.cme.msu.edu/).
- These sites also provide search algorithms to compare new sequences to their database.

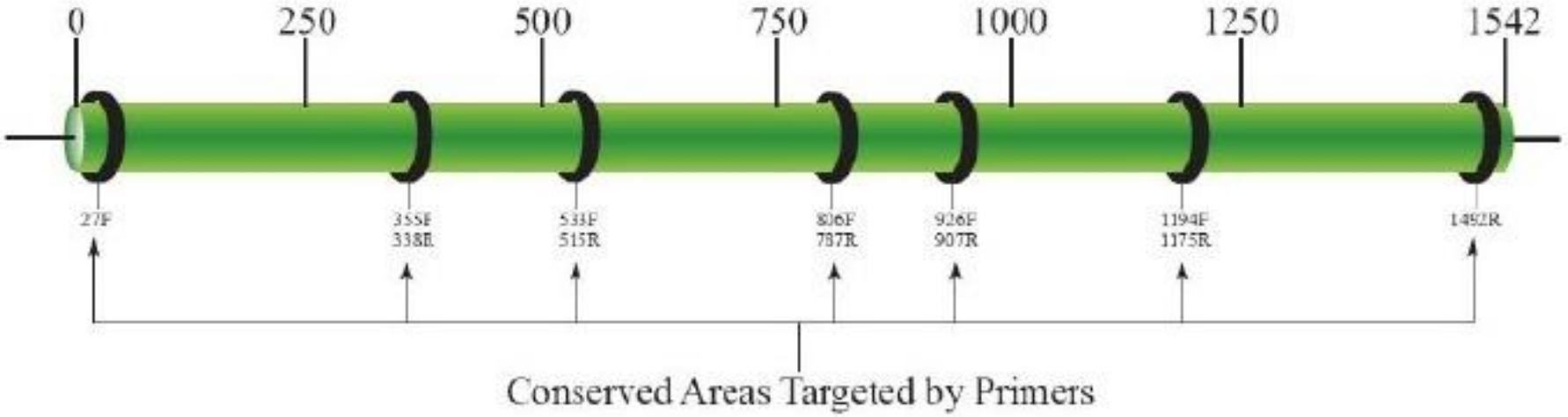
Why is the small subunit rRNA gene so useful ?

- Conserved in parts – highly variable in other parts. Thus it a very good phylogenetic marker
- VERY large database of sequences
- Cell have many ribosomes which can be targeted with probes (e.g. FISH, &TRFLP) for community analysis
- 16S rRNA gene sequencing is now the gold standard for community analysis



16S rRNA gene (cont.)

- It consists of conserved and variable regions



Which hyper-variable regions to sequence?

Region	Position	# b.p.
V1	69-99	30
V2	137-242	105
V3	338-533	195
V4	576-682	106
V5	822-879	57
V6	967-1046	79
V7	1117-1173	56
V8	1243-1294	51
V9	1435-1465	30

E.coli 16S SSU rRNA hyper-variable regions

CAVEATS TO 16S SEQUENCING

- ✓ Requires pure culture
- ✓ Different spp. can have an identical 16S
  - *B. bronchoseptica* + *B. parapertussis*, *M. gastri* + *M. kansasii*
- ✓ Different spp. can have minimally variable 16S
  - *S. pneumoniae* + *S. mitis*, *M. abscessus* + *M. chelonae*
- ✓ Genomovars of a single "species" may have relatively different 16S sequences (*P. vulgaris*, *E. cloacae*, *B. fragilis*)
- ✓ Multiple 16S alleles within a strain

Caveats to Using GenBank

- ✓ Not a quality-controlled database
  - Many sequencing errors (N, misreads, gaps)
  - Many incorrect IDs
- ✓ Lots of junk sequences (anaerobes)
- ✓ Paucity of entries (Coaggulase-negative Staphylococci)
- ✓ Dated entries (*Legionella micdadei* vs. *Tatlockia micdadei*)
- ✓ Submitters assign names to new species that are not validly published
- ✓ Highest score is not necessarily the correct species

- ✓ Alternatives (RIDOM) lack the same breadth

16S Amplicon

PROS

- Ubiquitous gene
- Contains both conserved and variable regions

CONS

- Copy number variations
- It's only a (single) gene

Typical workflow

Sampling → Extraction → Sample prep → Sequencing → Bioinformatics





- Standardisation, standardization, standardizasion..!
- Use biological replicates and evaluate your variation…!

Driving idea

- Well established method
  that can be used to **compare different** samples

- At any step we introduce **bias**, that have to be taken into account
  - **Sampling** (replicate or lie)
  - **Cell breakage** (are you strong enough?)
  - **Amplification** (where do your primers come from?)
  - **Sequencing** (how good is your machine?)
  - **Analysis** (database annotation?)

16S rRNA gene sequencing process

Bacterial colonies

↓

DNA extraction

↓

PCR

↓

PCR product cleanup

↓

Cycle sequencing

↓

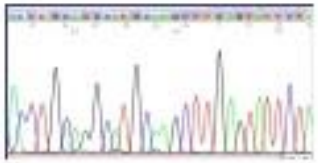
Sequencing product cleanup

↓

Sequence detection

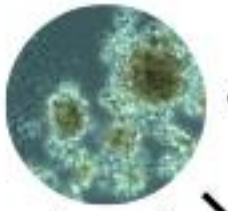
↓

Data assembly and analysis



Typical workflow


Sampling


Extraction


Sample prep


Sequencing


Bioinformatics


Storage

- Fresh
- 24 h @ 4°C
- 24 h @ 20 °C

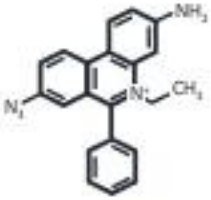

Input (mg)

1 2 4 9 22

eDNA removal

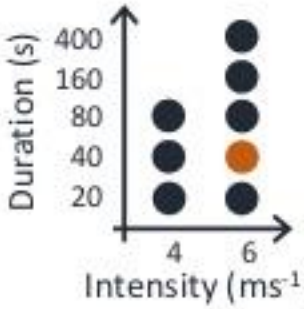

650 W 10 min

+



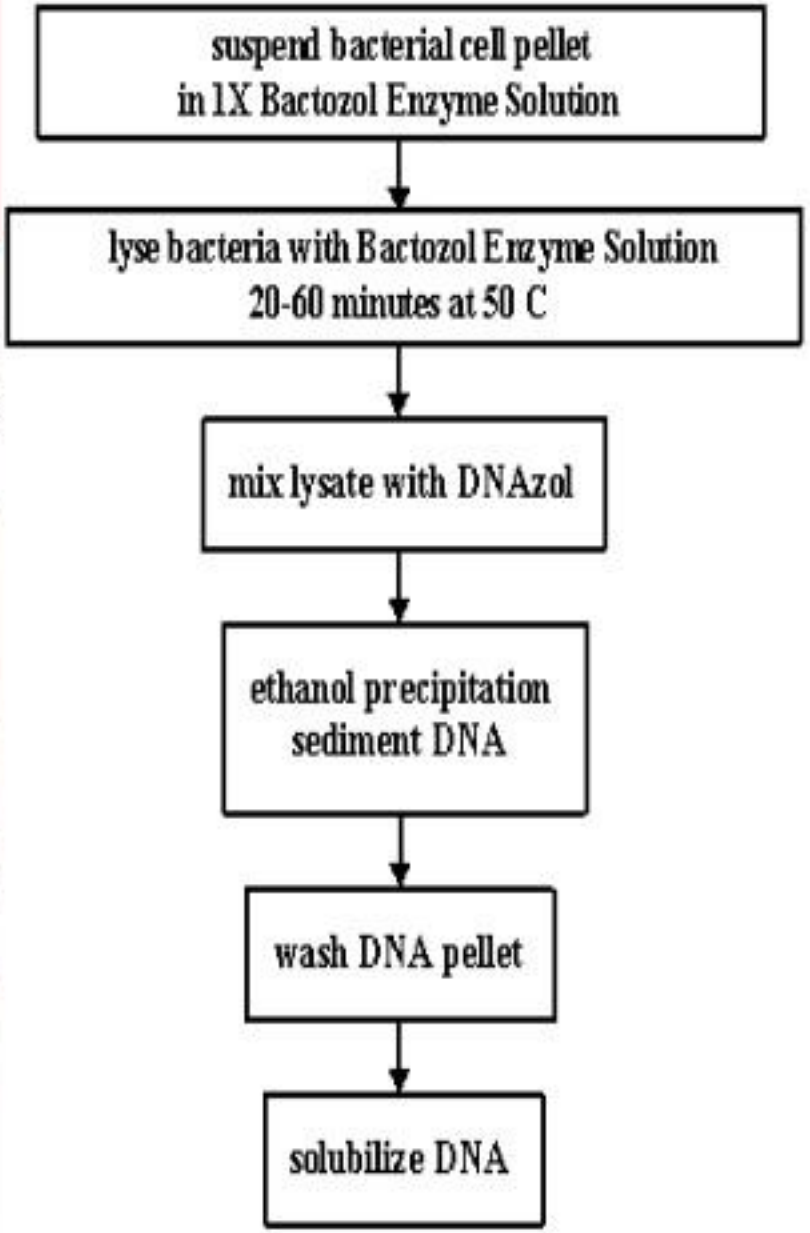
PMA

Bead beating



Extraction of DNA from the sample

- A variety of extraction methods can be used for the extraction of DNA for use in 16S rRNA sequencing. The choice of extraction methods rest with the source of the DNA sample and the amount of purity desired.
- For soil bacteria, the most common methods used are bead beating, sonication, enzymatic lysis, etc.
- For bacteria derived from other sources, the mainstream DNA isolation methods like Phenol-Chloroform method, CTAB method are also used.
- Modern laboratories depend on readily available kits to achieve quick, efficient and highly pure DNA Extraction.



Bacterial DNA Extraction

0.5 McFarland solution

Heating for 10 min at 95-100°C

No DNA present

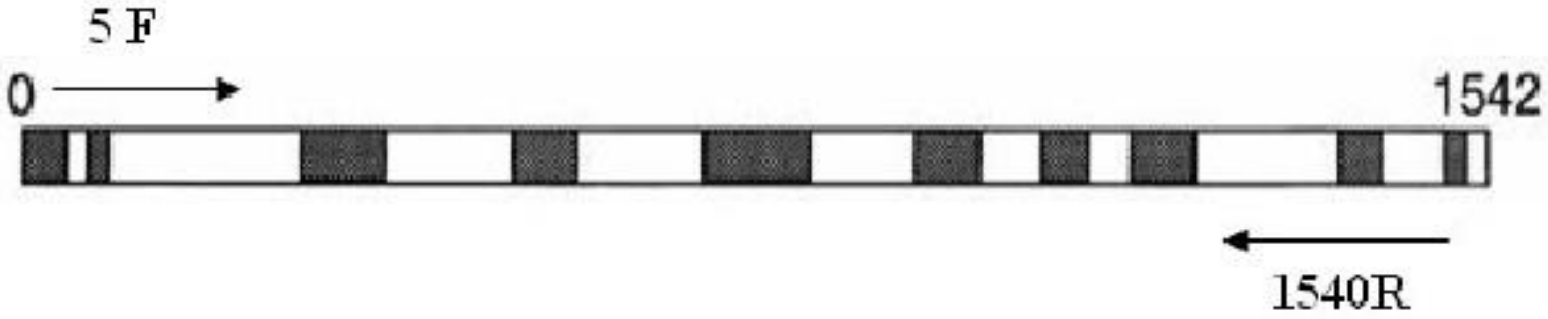
Bead beating extraction method

Amplification of 16S rRNA Sequence

- The DNA extracted is used as the template for PCR to amplify a segment of about 500 or 1,500 bp of the 16S rRNA gene sequence.
- Broad-based or universal primers complementary to conserved regions are used so that the region can be amplified from any bacteria. The PCR products are purified to remove excess primers and nucleotides.
- The PCR Amplification results in multiple copies of the target DNA Sequence being produced.
- This resulting sequence is then used as the template for the next step of the process known as Cycle Sequencing.

PCR of 16S rRNA gene sequence

- Universal primers were selected to amplify 1,500bp 16S rRNA gene sequence.



Black: conserved regions

White: variable regions

PCR optimization

- Data showing results when PCR conditions were not optimized:

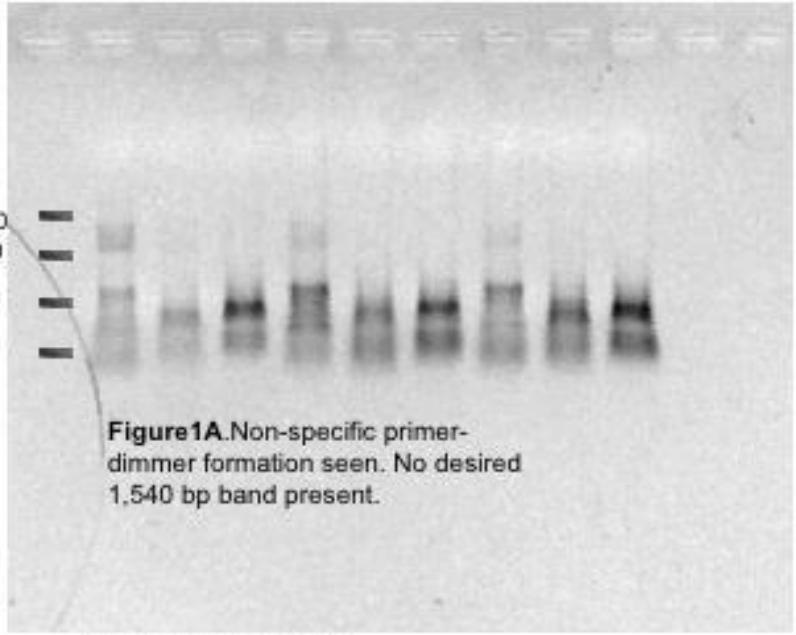


Figure1A.Non-specific primer-dimmer formation seen. No desired 1,540 bp band present.

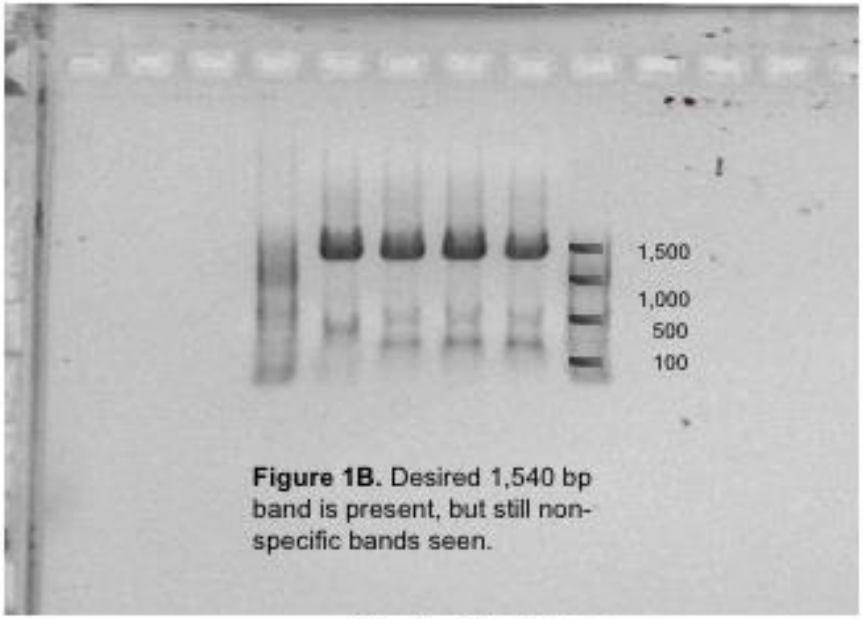


Figure 1B. Desired 1,540 bp band is present, but still non-specific bands seen.

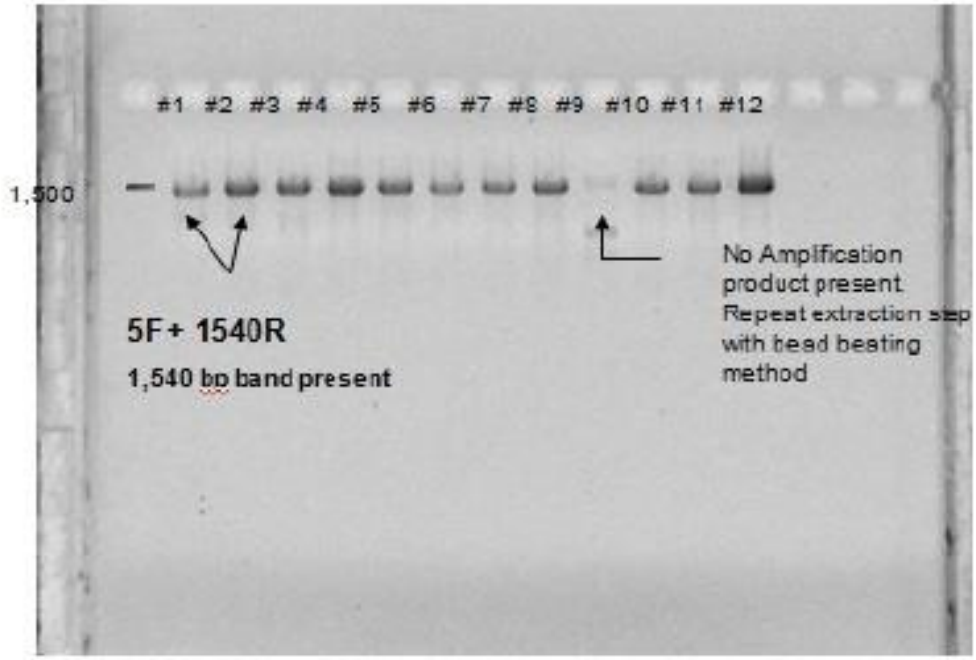
Cycling Conditions:

Cycles	Temp.	Time
1	95°C	5min
35	94°C	30sec
	52°C	20sec
	72°C	2min
1	72°C	10min

Cycling Conditions:

Cycles	Temp.	Time
1	95°C	5min
35	94°C	**40sec**
	56°C	20sec
	72°C	2min
1	72°C	10min

PCR optimization



Cycles	Temp.	Time
1	95°C	5min
35	94°C	40sec
	60°C	20sec
	72°C	2min
1	72°C	10min

16S rRNA gene sequencing process

Bacterial colonies

↓

DNA extraction

↓

PCR

↓

PCR product cleanup

↓

Cycle sequencing

↓

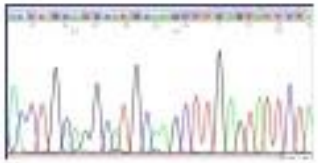
Sequencing product cleanup

↓

Sequence detection

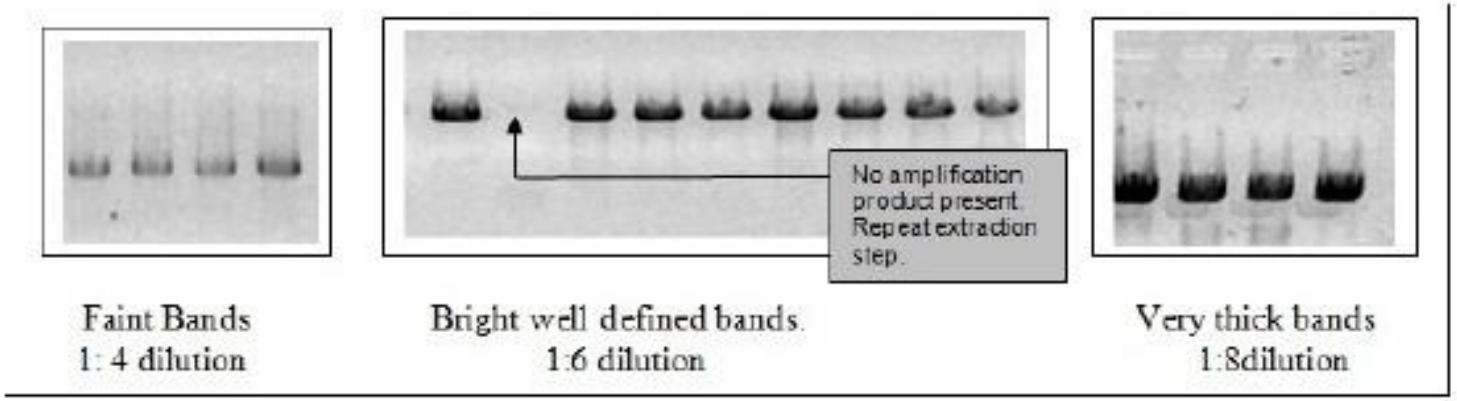
↓

Data assembly and analysis



Purification of PCR product

- Amplification product was diluted 1:6 (or more depending on band size) with PCR grade water.



16S rRNA gene sequencing process

Bacterial colonies

↓

DNA extraction

↓

PCR

↓

PCR product cleanup

↓

Cycle sequencing

↓

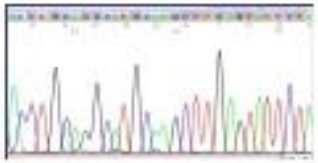
Sequencing product cleanup

↓

Sequence detection

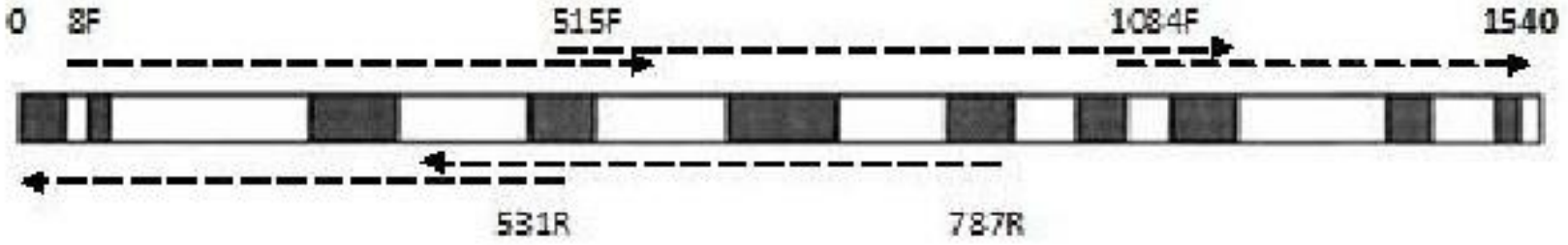
↓

Data assembly and analysis



Sequencing of PCR product

- To sequence 16S rRNA gene several universal primers were selected



Black: conserved regions

White: variable regions

Sanger Sequencing

16S rRNA gene sequencing process

Bacterial colonies

↓

DNA extraction

↓

PCR

↓

PCR product cleanup

↓

Cycle sequencing

↓

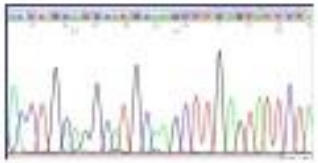
Sequencing product cleanup

↓

Sequence detection

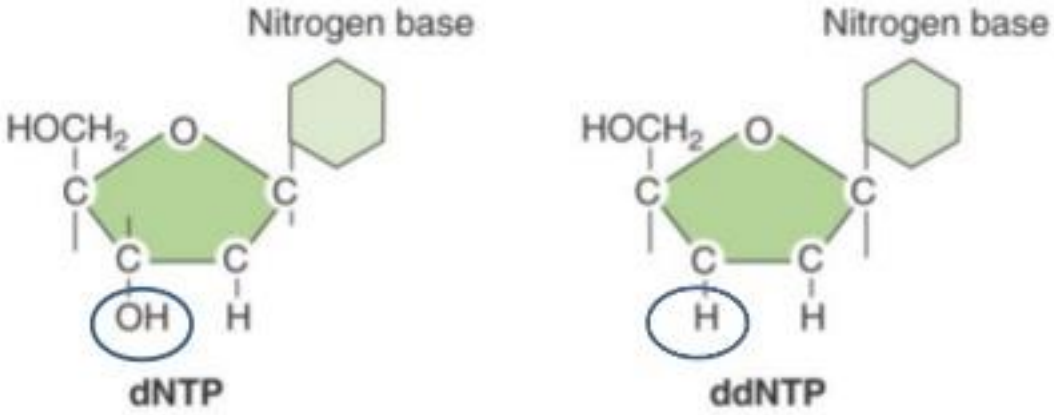
↓

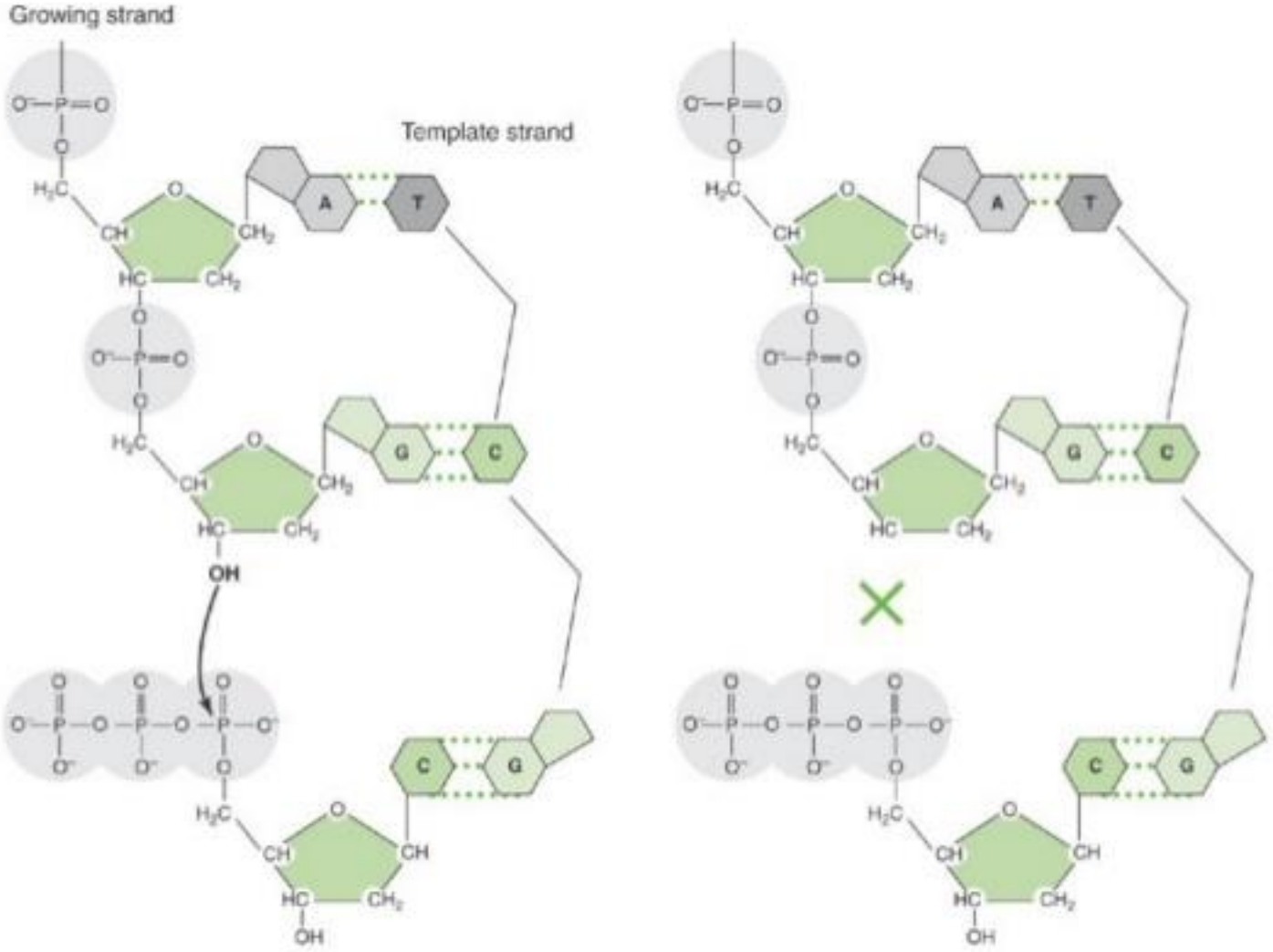
Data assembly and analysis



Sanger; Chain Termination Sequencing

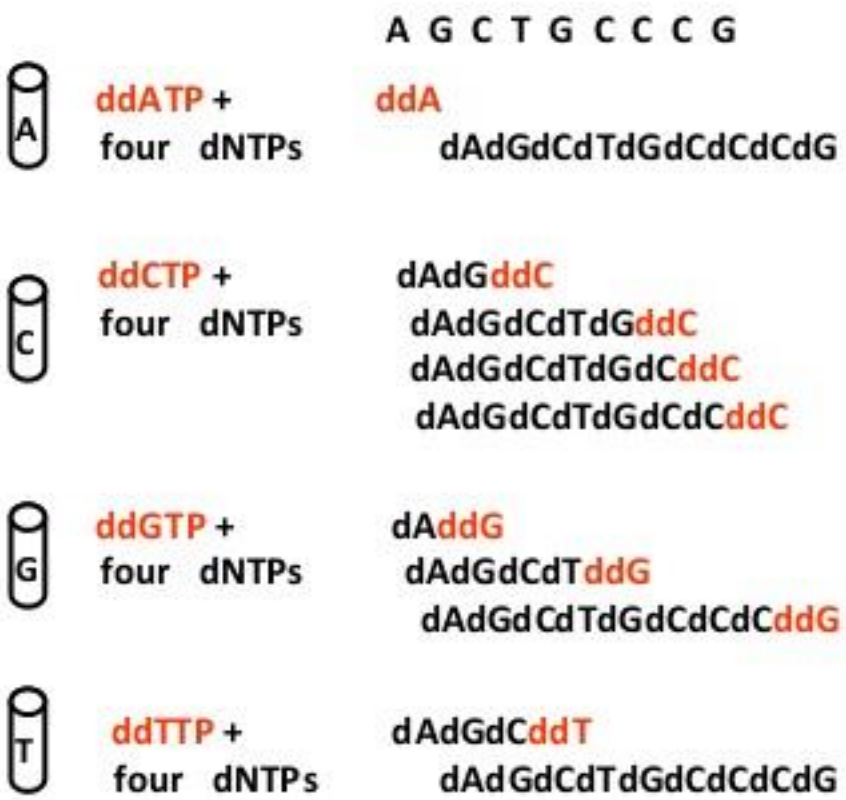
- It is PCR based method
- A modified DNA replication reaction
- Growing chains are terminated by dideoxynucleotides





The 3′-OH group necessary for formation of the phosphodiester bond is missing in ddNTPs

Sanger; Chain Termination Sequencing

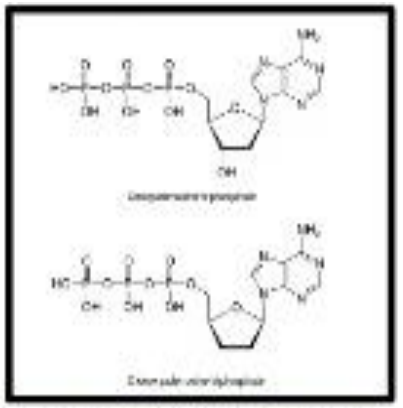


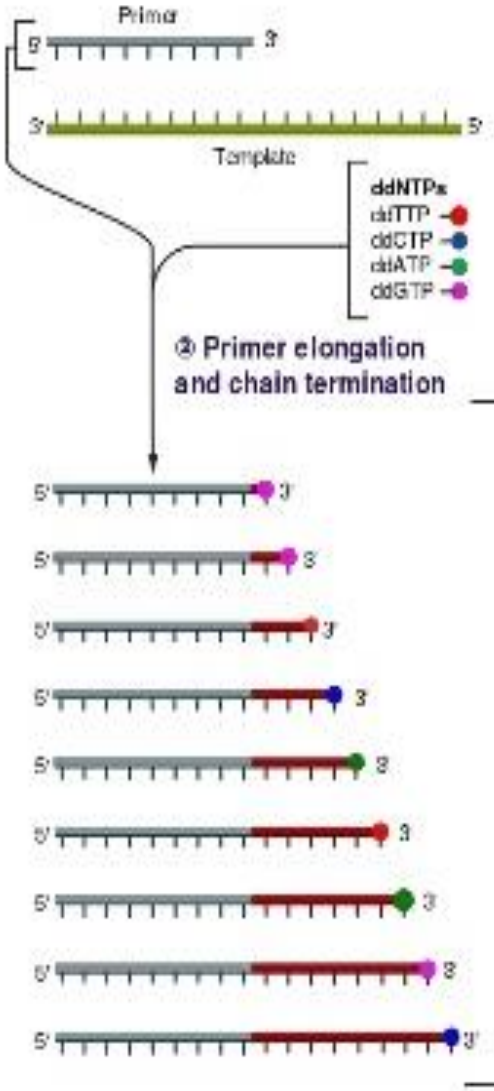
Purification and detection of sequencing product

- Sequenced product was purified with Big Dye Xterminator purification kit.
- By capillary electrophoresis (ABI PRISM Genetic analyzer)

① Reaction mixture

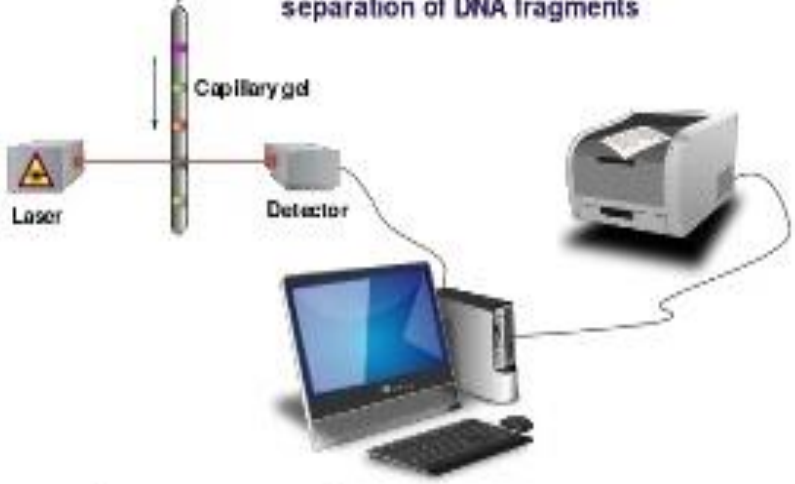
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)

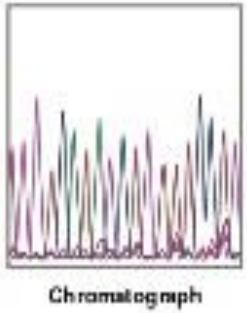




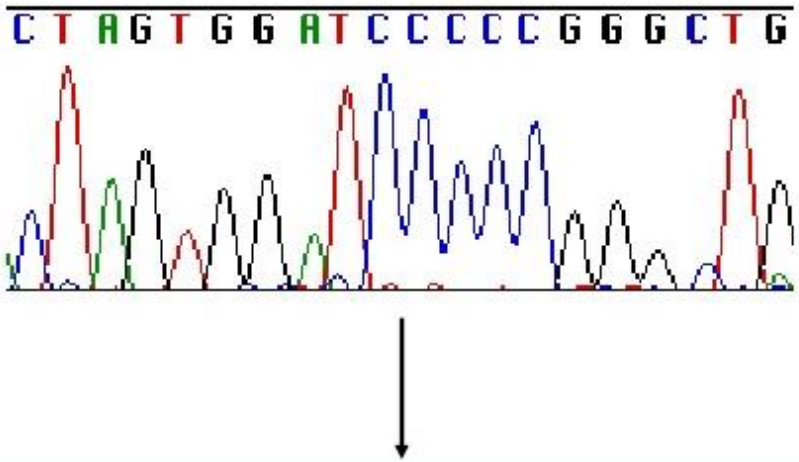
② Primer elongation and chain termination

③ Capillary gel electrophoresis separation of DNA fragments



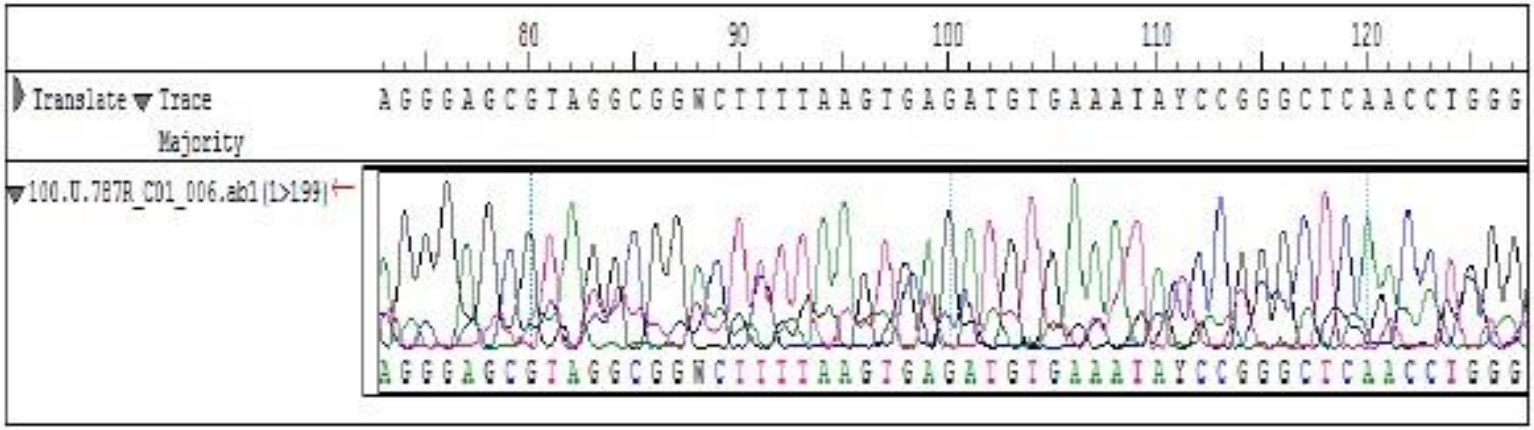


④ Laser detection of flourochromes and computational sequence analysis


C T A G T G G A T C C C C C G G G C T G
EditSeq
File Edit Search Speech Features Goodies Windows Net Search Help
Unknown S55691 Bact16S fR rc.seq : SEQUENCE
Position: 525
786 bp
10 20 30 40 50 60
AGAGTTTGATCATGGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAAC 60
GATGAAGCCTAGCTTGCTAGGTGGATTAGTGGCGAACGGGTGAGTAATACGTGAGTAACC 120
TACCTTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATACGACCAATCTC 180
CGCATGGGGTGTTGGTGGAAAGCGTTATGTAGTGGTTATAGATGGGCTCACGGCCTATCA 240
GCTcGTTGGTGAGGTAACGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGA 300
CCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATA 360
TTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGT 420
TGTAAACCTCTGTTAGCAGGGAAGAAGCAGAGATTGACGGTACCTGCAGAGAAAGCGCCGG 480
CTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCGAGCGTTGTCCGGAATTATTG 540
GGCGTAAAGAGCTTGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGGCCGGAGCTTAACTCC 600
GTGTATTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACTGGAATTCCTGGTG 660
TAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCT 720
GTAACTGACGCTGAGAAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTC 780
CATGCC 786
Unspecified Search

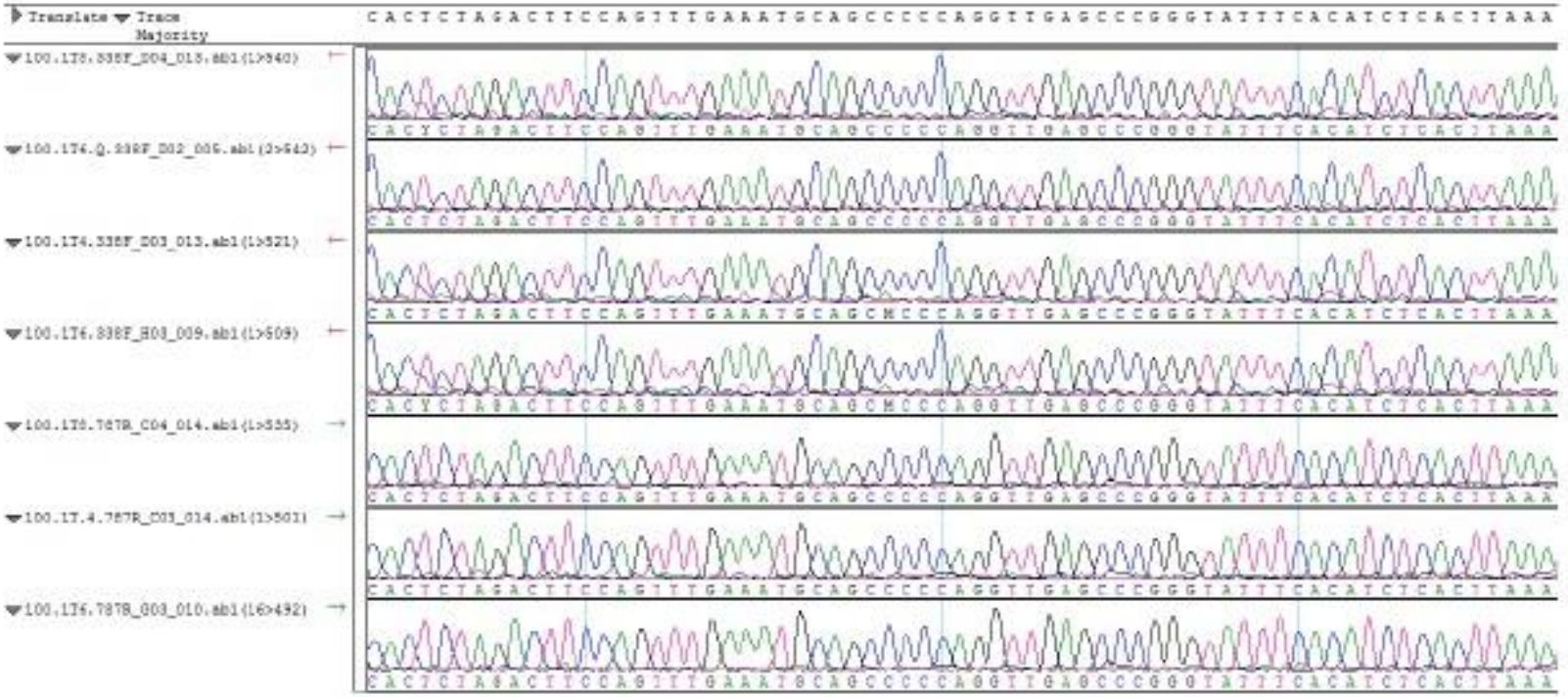
Optimization of PCR product purification step

- Sequencing with undiluted PCR product generated very poor data with high background noise and sequences not being able to asseble into one contiq.



Optimization of PCR product purification step

- Dilution of PCR product to 1:2 generated good quality but shorter sequences (~300bp). Dilution of PCR product to1:4, 1:6 and 1:8 resulted in very good quality long (>500bp) sequences .



Labor intensiveness and Turn-around time assessment of sequencing

Steps	Procedure	Labor Time (hands-on)	Waiting time (machine time)
1	Extraction of bacterial DNA	5 min	10 min
2	Master mix preparation, PCR amplification	10 min	3 hours, 30 min
3	Analysis of the PCR product. Loading, running, and examining gel.	10 min	20 min
4	Dilution of PCR product and Sequencing step	15 min	2 hours, 30 min
5	Purification of PCR products	5 min	32 min
6	Assembling capillary tray for sequencing, loading tray to the Genetic analyzer.	5 min	1 hour
7	Sequence assembly, editing, database search	10 min	
9	Reporting of results.	5 min	
	Total labor time	**1 hours 5 min/** per 1 isolate (add 10 min to each additional isolate)	8 hours and 35 min

Validation of 16S rRNA sequencing

- 129 isolates consisting of ATCC strains and known bacterial strains were tested with 16S rRNA sequencing

No. of known strains tested	Identified to genus level	Identified to species level
129	129(100%)	127 (98.4%)

Next Generation Sequencing

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 2011, p. 3846–3852 Vol. 77, No. 11
0099-2240/11/$12.00 doi:10.1128/AEM.02772-10

Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads▽†

Andrea K. Bartram,[1] Michael D. J. Lynch,[2] Jennifer C. Stearns,[1] Gabriel Moreno-Hagelsieb,[2] and Josh D. Neufeld[1]*

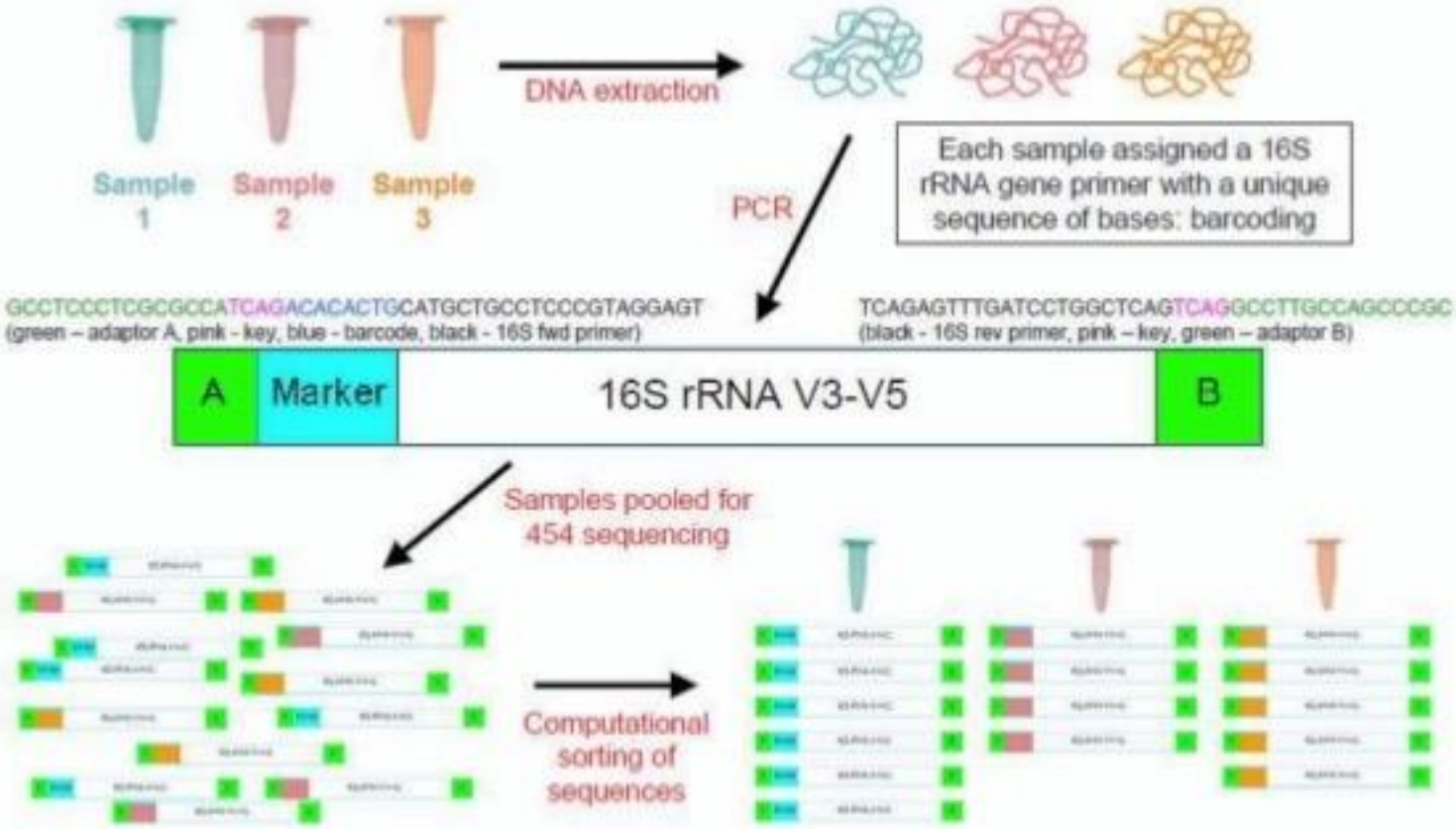
Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1,[1] and Department of Biology, Wilfrid Laurier University, Waterloo, Ontario, Canada N2L 3C5[2]

Received 25 November 2010/Accepted 24 March 2011

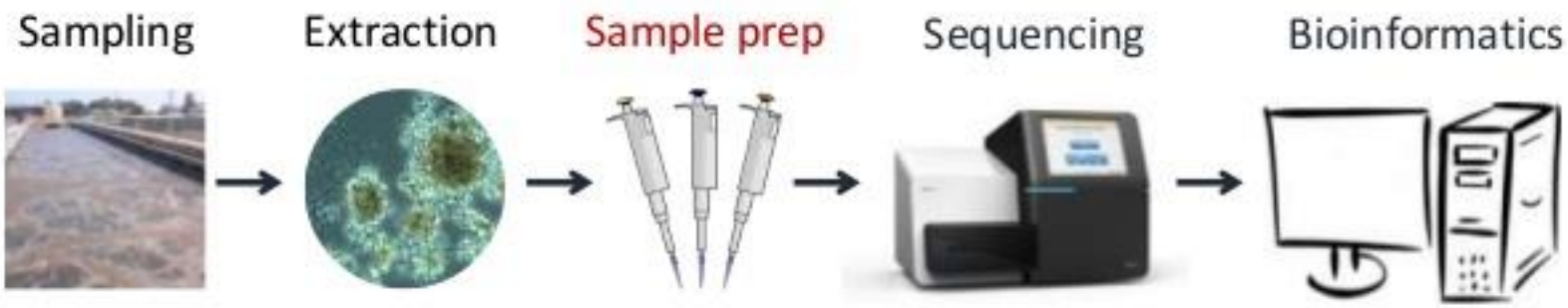
Microbial communities host unparalleled taxonomic diversity. Adequate characterization of environmental and host-associated samples remains a challenge for microbiologists, despite the advent of 16S rRNA gene sequencing. In order to increase the depth of sampling for diverse bacterial communities, we developed a method for sequencing and assembling millions of paired-end reads from the 16S rRNA gene (spanning the V3 region; ~200 nucleotides) by using an Illumina genome analyzer. To confirm reproducibility and to identify a suitable computational pipeline for data analysis, sequence libraries were prepared in duplicate for both a defined mixture of DNAs from known cultured bacterial isolates (>1 million postassembly sequences) and an Arctic tundra soil sample (>6 million postassembly sequences). The Illumina 16S rRNA gene libraries represent a substantial increase in number of sequences over all extant next-generation sequencing approaches (e.g., 454 pyrosequencing), while the assembly of paired-end 125-base reads offers a methodological advantage by incorporating an initial quality control step for each 16S rRNA gene sequence. This method incorporates indexed primers to enable the characterization of multiple microbial communities in a single flow cell lane, may be modified readily to target other variable regions or genes, and demonstrates unprecedented and economical access to DNAs from organisms that exist at low relative abundances.

Now generate V3-V4 bacterial amplicons (~450 bases)
Usually PE 300

454-based 16S amplicon sequencing



Typical workflow



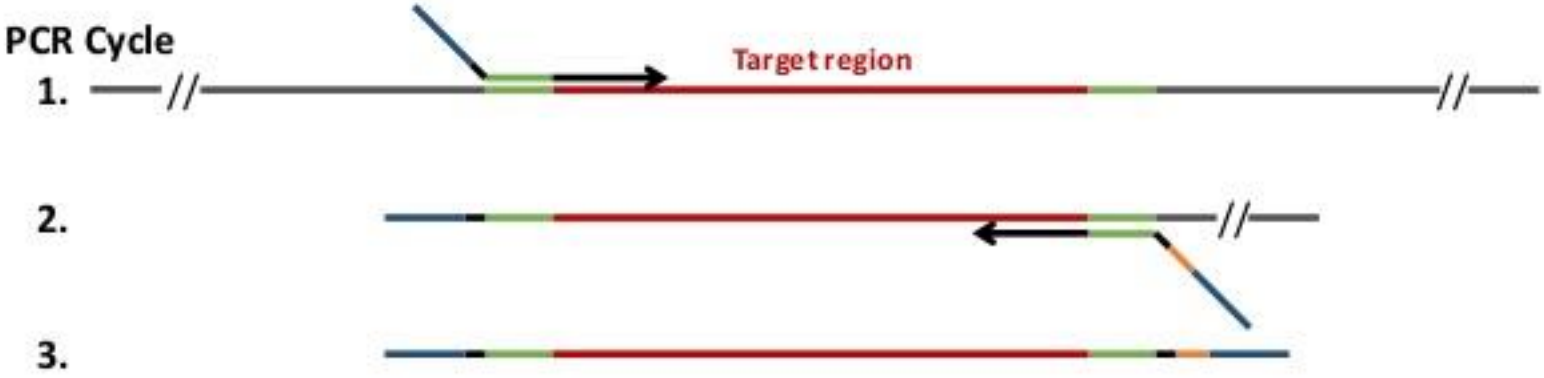
PCR with modified 16S primers

Illumina adapter Pad linker 27F

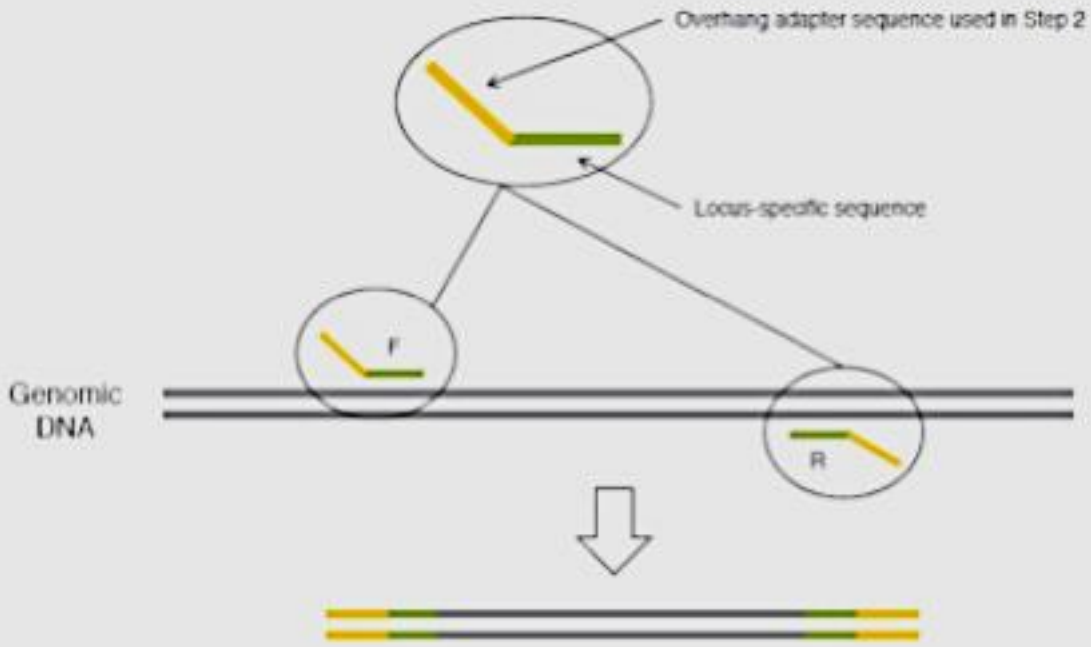
5′ -AATGATACGGCGACCACCGAGATCTACAC GTACGTACG GT AGAGTTTGATCCTGGCTCAG-3′

Illumina adapter Barcode Pad linker 534R

5′ -CAAGCAGAAGACGGCATACGAGAT TCCCTTGTCTCC ACGTACGTAC CCG ATTACCGCGGCTGCTGG-3′

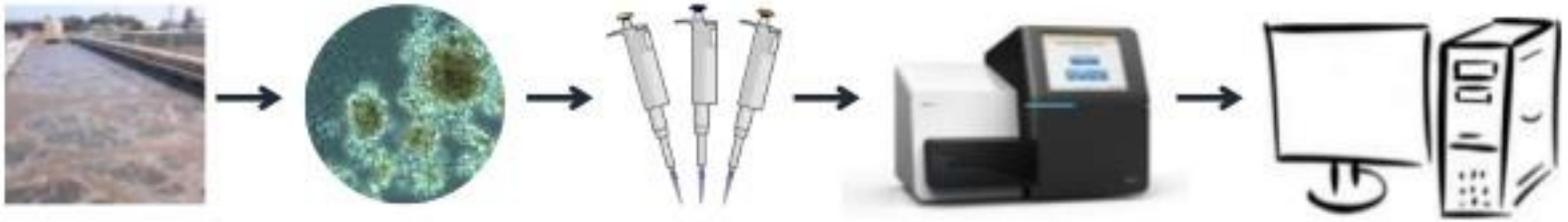


16S Amplicon

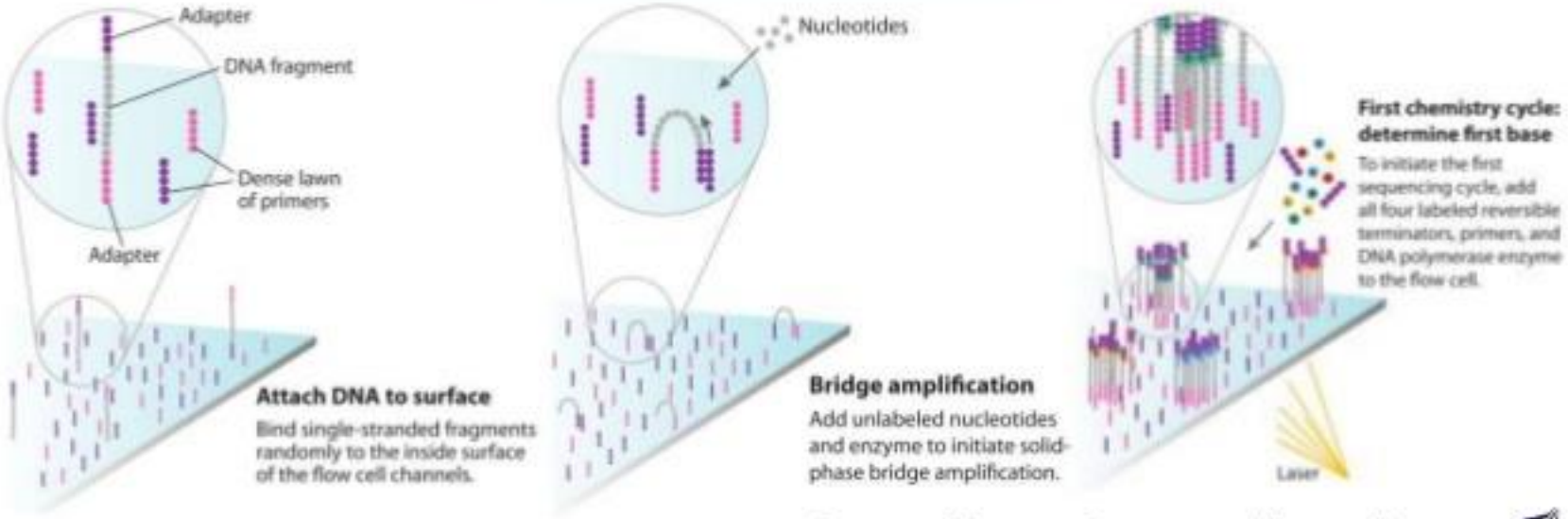


Typical workflow

Sampling → Extraction → Sample prep → Sequencing → Bioinformatics

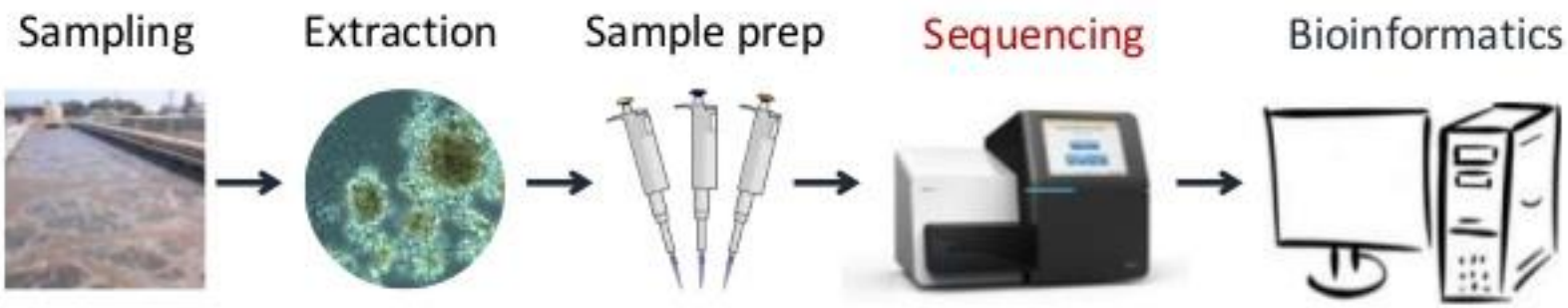


≈ 500 bp target amplicon

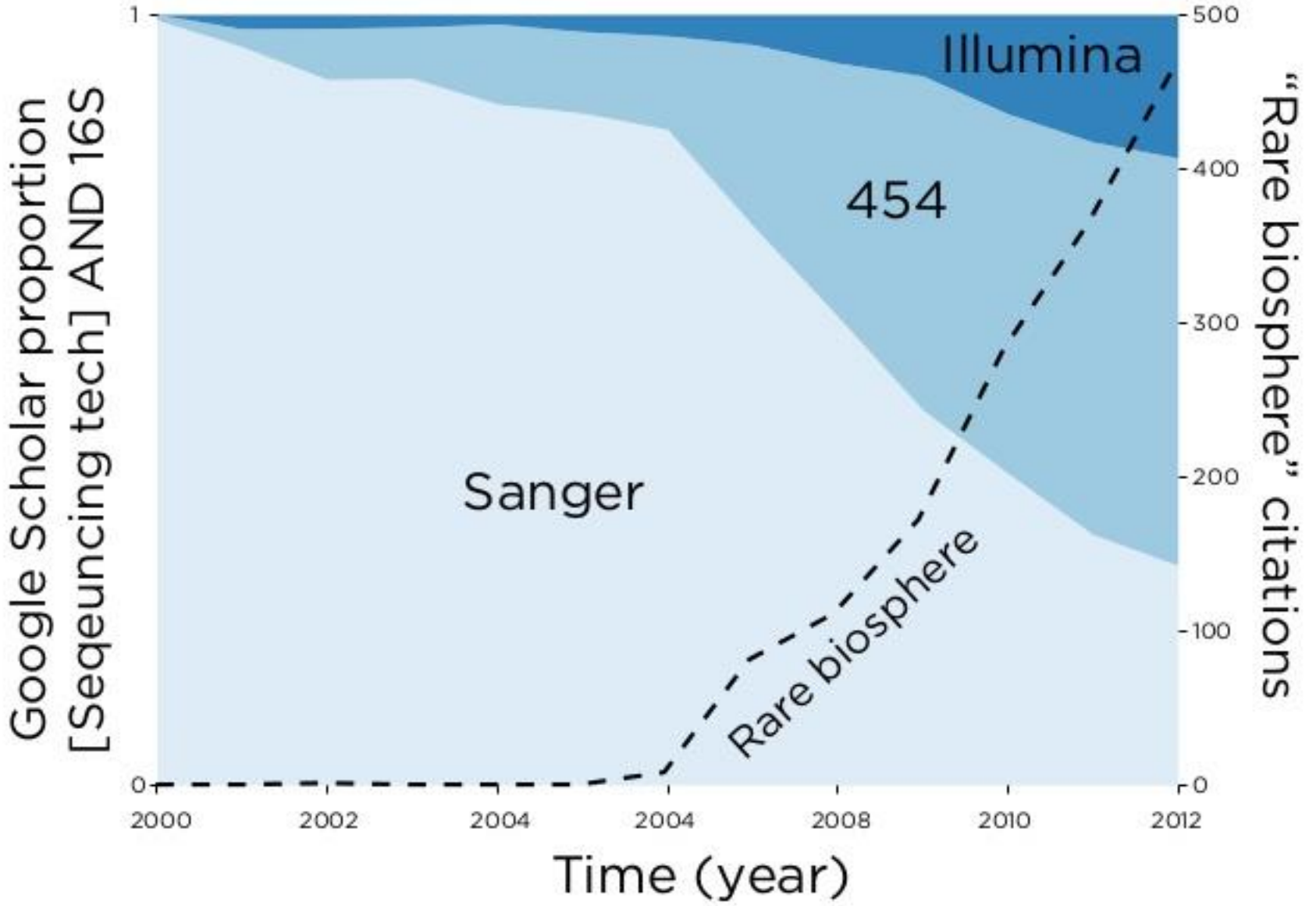


Mardis, 2008 (PMID 18576944)

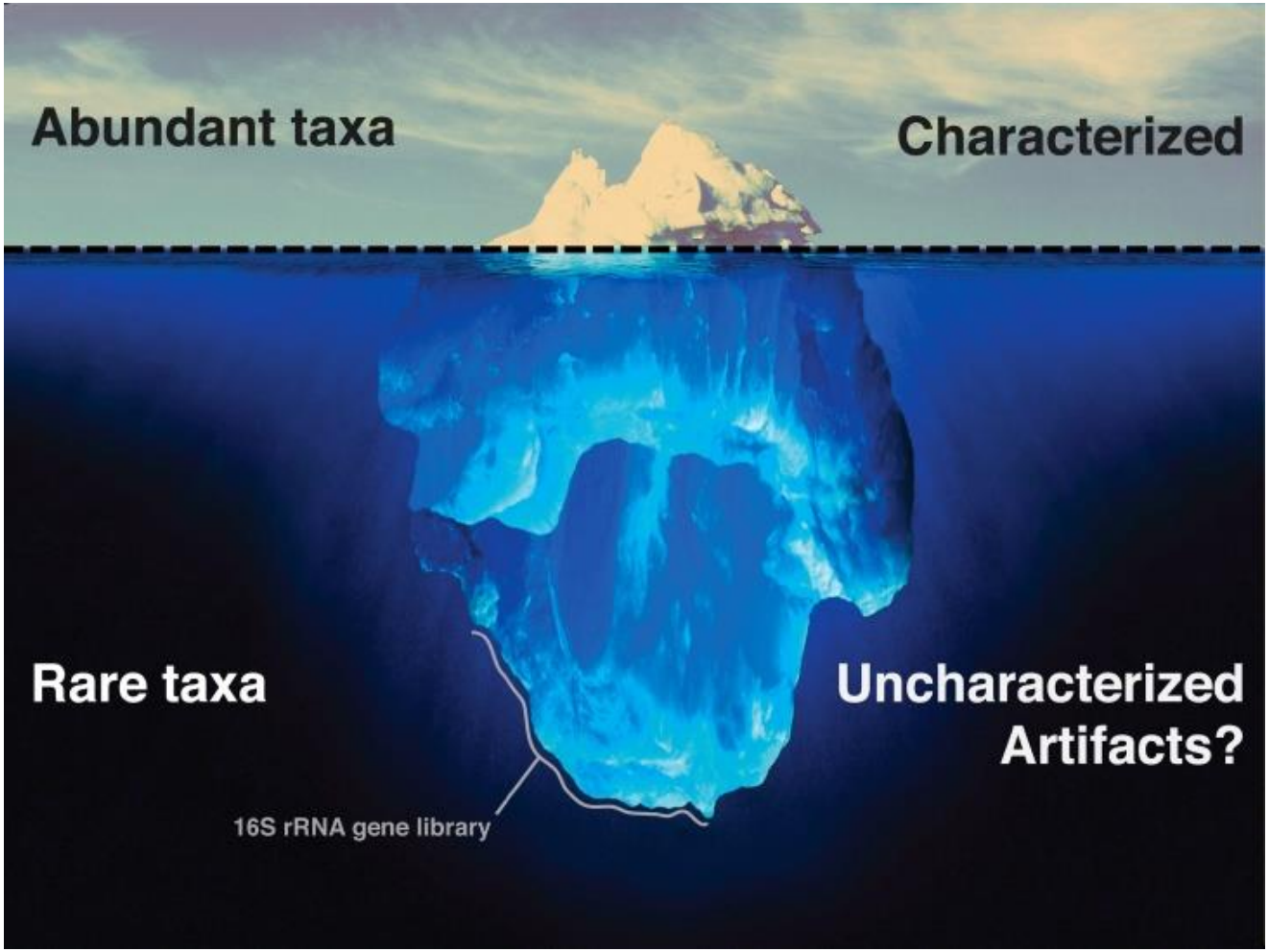
Typical workflow



After Sequencing:



Lynch and Neufeld. 2013. Nat. Rev. Microbiol. In preparation.


Abundant taxa
Characterized
Rare taxa
Uncharacterized
Artifacts?
16S rRNA gene library

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Microbiology

High-throughput sequencing and clinical microbiology: progress, opportunities and challenges

Mark J Pallen, Nicholas J Loman and Charles W Penn

High-throughput sequencing is sweeping through clinical microbiology, transforming our discipline in its wake. It is already providing an enhanced view of pathogen biology through rapid and inexpensive whole-genome sequencing and more sophisticated applications such as RNA-seq. It also promises to deliver high-resolution genomic epidemiology as the ultimate typing method for bacteria. However, the most revolutionary effect of this 'disruptive technology' is likely to be creation of a novel sequence-based, culture-independent diagnostic microbiology that incorporates microbial community profiling, metagenomics and single-cell genomics. We should prepare for the coming 'technological singularity' in sequencing, when this technology becomes so fast and so cheap that it threatens to out-compete existing diagnostic and typing methods in microbiology.

Address
School of Biosciences, University of Birmingham, Birmingham, B15 2TT, United Kingdom

http://pathogenomics.bham.ac.uk/blog/2011/08/are-diagnostic-and-public-health-bacteriology-ready-to-become-branches-of-genomic-medicine/