

Encapsulated Bacteria Session 2: Bacterial identification, Linux, and BLAST

Genomics and Clinical Microbiology 2024

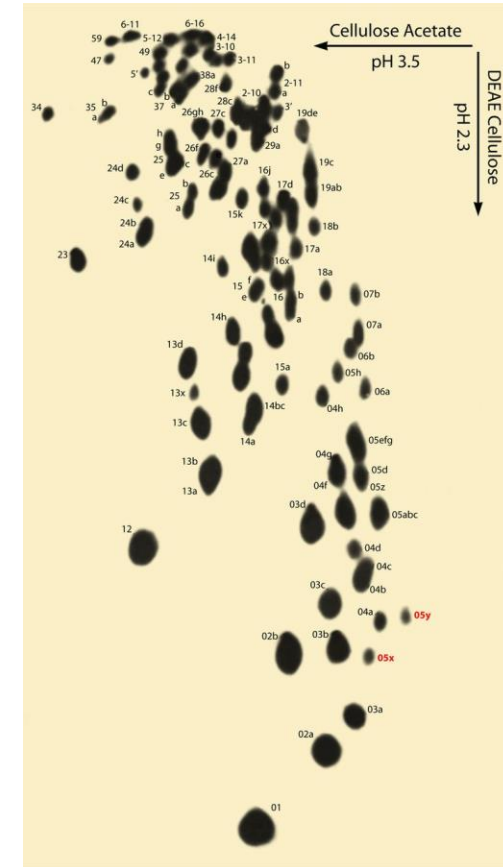
Keith Jolley, Made Krisna, Kasia Parfitt, Martin Maiden,
Department of Biology



UNIVERSITY OF
OXFORD

16S Ribosomal RNA sequencing

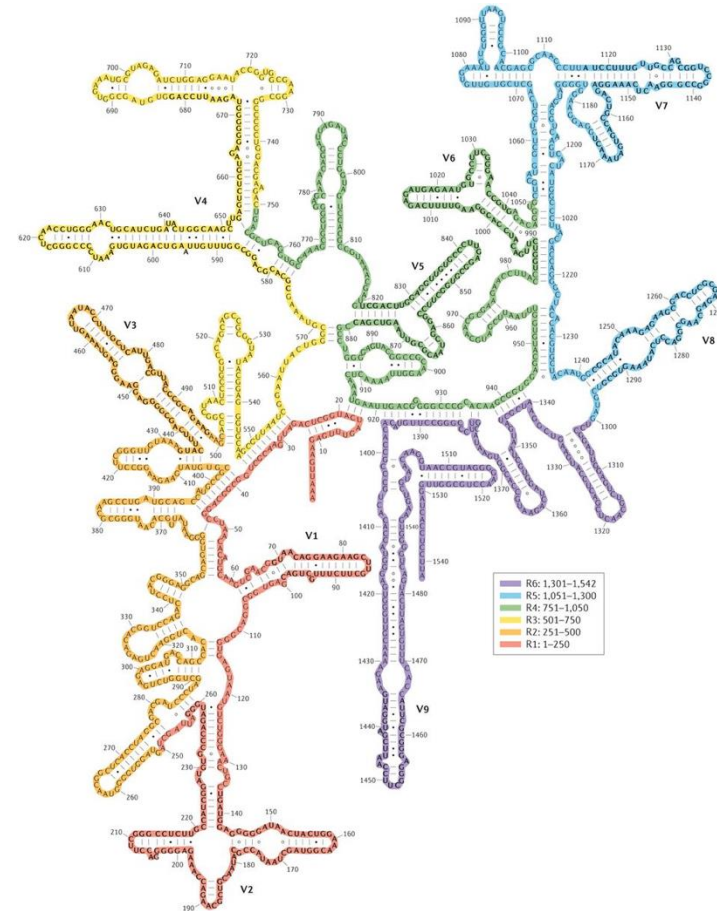
- Carl Woese initiated a unique research programme to obtain a tree of life.
- Ribosomal RNA was an accessible and tractable nucleic acid, which he characterised with thin layer chromatography sequencing.
- To his surprise, these analyses resulted in the discovery of a new domain of life first called the Archaeobacteria (1977) and subsequently the Archaea (1990).



Sapp, J. & Fox, G. E. (2013). The singular quest for a universal tree of life. *Microbiol Mol Biol Rev.* **77**, 541-550.

16S rRNA gene analysis

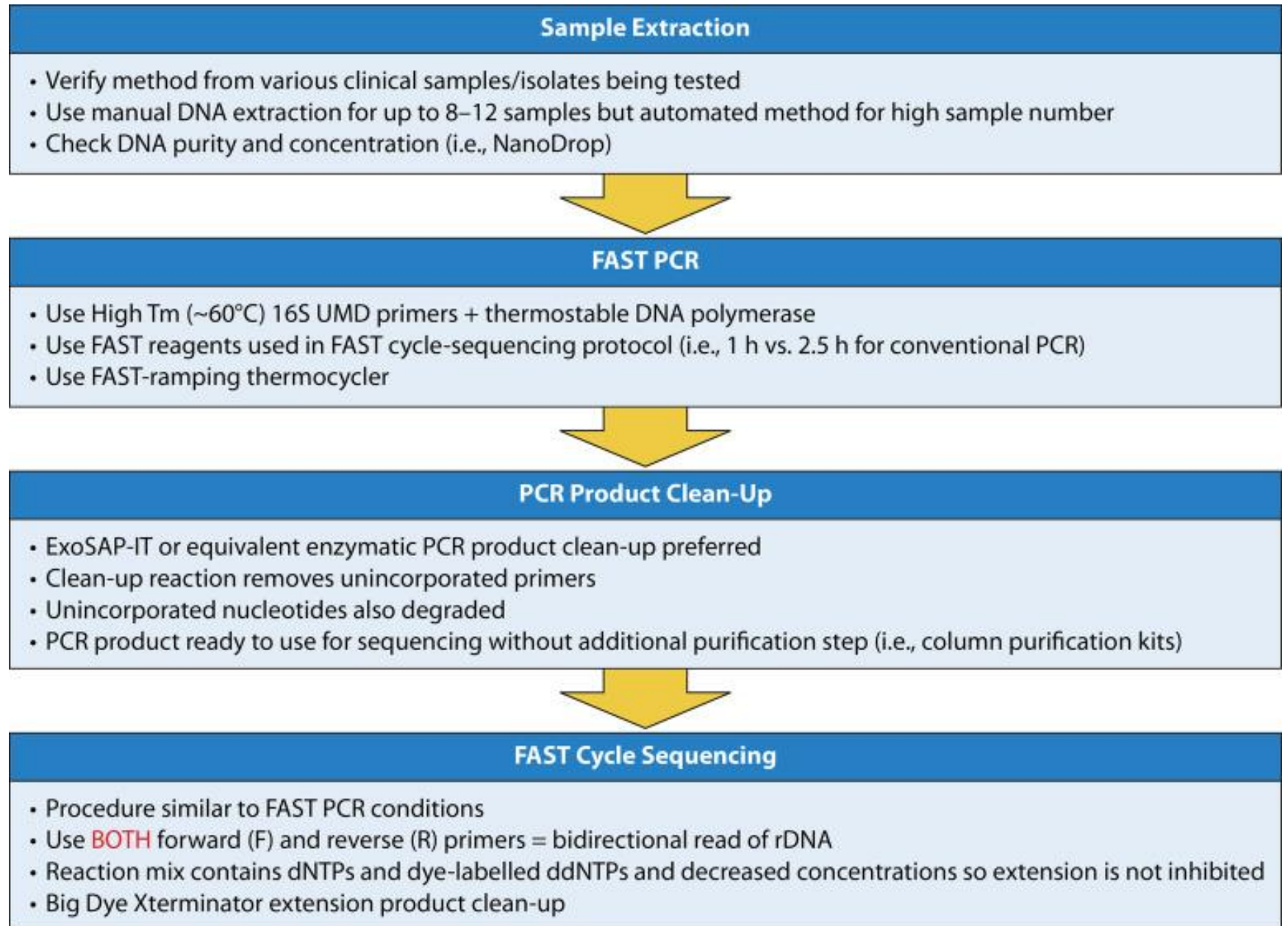
- 16S rRNA an essential structural component of the ribosome,
 - highly conserved.
- Varies among organisms,
 - most variation is at the genus level.
- 16S rRNA gene, ~1.5Kbp, with:
 - universal, conserved regions, which can be used to generate primers;
 - nine variable regions which can be used as signatures to determine genus/species.



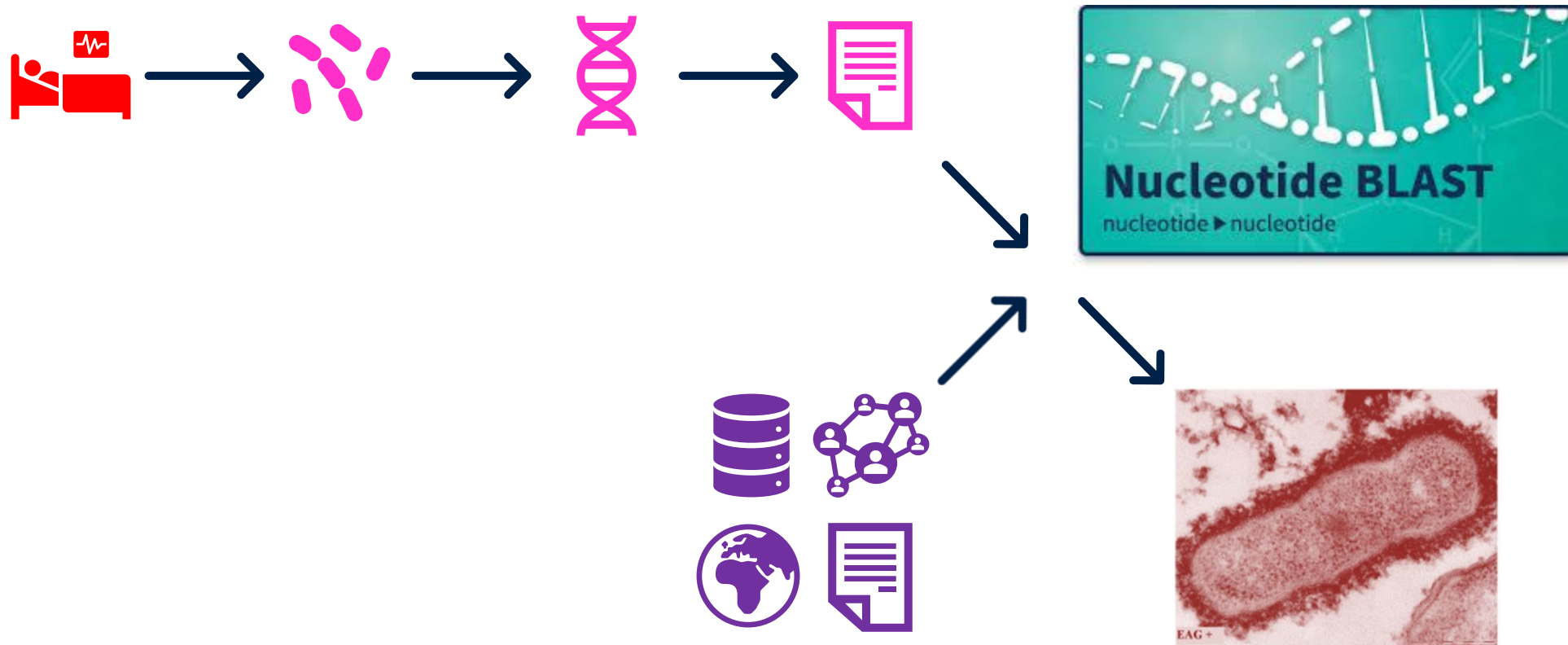
Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F. O., Ludwig, W., Schleifer, K. H., Whitman, W. B., Euzéby, J., Amann, R. & Rossello-Mora, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol.* **12**, 635-645.

16S rRNA sequencing in the clinic

Church, D. L., Cerutti, L., Gurtler, A., Griener, T., Zelazny, A. & Emler, S. (2020). Performance and Application of 16S rRNA Gene Cycle Sequencing for Routine Identification of Bacteria in the Clinical Microbiology Laboratory. *Clin Microbiol Rev.* **33**, e00053-1



16S rRNA 'species' id: process overview




Exercise: searching an NCBI database with 16SrRNA
sequences

16s rRNA BLAST search online

- Navigate to NCBI BLAST home page:
 - <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

- Click on:



- Under 'Enter Query sequence' paste in a test sequence (Test 1-6)
 - https://github.com/WCSCourses/Molecular_Approaches_Clinical_Microbiology_2024/blob/main/course_data/bioinformatics/16S/16S_samples.fas
 - <https://tinyurl.com/228hpuev>
- Under 'Choose Search Set' click on 'rRNA/ITS databases'
- Check box 'Show results in a new window'. Click on 




Important update
The core nucleotide database (**core_nt**) is now the default nucleotide BLAST database. [Learn more about core_nt.](#)

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query, more...

[Reset page](#) [Bookmark](#)

Enter Query Sequence


Enter accession number(s), gi(s), or FASTA sequence(s) 

```
>Test_1
AATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACA
CATGCAAG
TCGAACGGTAACAGGAAGAAGCTTGCTCTTTGCTGACGAGTGGCGGACGG
```


Query subrange 


From

To

Or, upload file No file selected. 


Job Title

Enter a descriptive title for your BLAST search 

☐ Align two or more sequences 


Choose Search Set

Database ☐ Standard databases (nr etc.): ☒ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus ☐ Experimental databases

16S ribosomal RNA sequences (Bacteria and Archaea) 

[Targeted Loci Project Information](#)


Organism Optional ☐ exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown 

Exclude Optional ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional ☐ Sequences from type material

Entrez Query Optional [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search 

Program Selection

Optimize for ☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm 

BLAST Search database 16S ribosomal RNA sequences (Bacteria and Archaea) using Megablast (Optimize for highly similar sequences)

☒ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with  sign

+ Algorithm parameters

What bacterium is it?

- Look at the results of the search (scroll down the page as necessary).
- What is the likely organism that the 16S rRNA sequence came from?

Your Results

Sample	Candidate Bacterium
Test 1	<i>Escherichia/Shigella???</i>
Test 2	<i>Mycobacterium tuberculosis</i>
Test 3	<i>Vibrio cholerae</i>
Test 4	<i>Neisseria mucosa</i>
Test 5	<i>Neissera gonorrhoeae</i>
Test 6	<i>Neisseria meningitidis</i>

Source sequences

Sample	Candidate Bacterium
Test 1	<i>Escherichia coli</i> (ENA MN900682)
Test 2	<i>Mycobacterium tuberculosis</i> (ENA AJ536031)
Test 3	<i>Vibrio cholerae</i> (ENA U10955)
Test 4	<i>Neisseria lactamica</i> (ENA FN995097.1)
Test 5	<i>Neisseria gonorrhoeae</i> (ENA X07714)
Test 6	<i>Neisseria meningitidis</i> (ENA AJ239309)

Reflections

- How easy is it to get results?
- How easy is it to interpret them?
- How would you report them?
- How confident are you about the results?
- What information do you need to improve your confidence?

Practical: Introduction to Linux & BLAST