INTRODUCTION TO 16S RDNA SEQUENCING AND ANALYSIS, AND GEL DOCUMENTATION

Murugadas Vaiyapuri*, Sreejith V.N., and Vineetha Das

Microbiology, Fermentation and Biotechnology Division

Email: Murugadas.V@icar.gov.in

16s rDNA

rRNA gene is the most conserved and used to determine taxonomy, phylogeny (evolutionary relationships). It is also used to infer relationships between organisms that span the diversity of known life look. These genes are conserved through the billions of years of evolutionary divergence.

Microbial systematics

Novel species is recognized using the polyphasic approach - multidimensional aspects of organisms (phenotypic, genotypic and chemotaxonomic traits). Biochemical methods of identification of bacteria are called as phenotypic methods, 16s rRNA sequencing and DNA-DNA hybridization are called as genotypic methods and other FAME analysis are chemotaxonomic methods. Phylogenetic analysis based on 16S rRNA gene sequences and determination of similarity between sequences - first step in identifying novel organisms - most widely used methodology in the world. DNA−DNA hybridization (DDH), which measures indirectly the degree of genetic similarity between two genomes, has been the 'gold standard'

16s rRNA sequencing analysis

One or more copies of the operon dispersed in the genome (mostly 3, *E. coli* 7). Ribosomal RNAs in Prokaryotes: 5S 120 large subunit of ribosome; 16S 1500 Small subunit of ribosome; 23S 2900 large subunit of ribosome. The 16s rDNA sequence has hyper variable regions, where sequences have diverged over evolutionary time. Strongly conserved regions often flank these hyper variable regions. Primers are designed to bind to conserved regions and amplify variable regions.Numbered primers are named for the approximate position on the E. coli 16S rRNA molecule. More details can be sought from National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and the Ribosomal Database Project (http://rdp.cme.msu.edu/). Minimum: 500 to 525 bp sequenced; ideal: 1,300 to 1,500 bp sequenced 1% position ambiguities. Minimum: 99% sequence similarity; ideal: 99.5% sequence similarity. Sequence match is to type strain or reference strain of species that has undergone DNA-relatedness studies. For matches with distance scores 0.5% to the next

closest species, other properties, including phenotype, should be considered in final species identification.
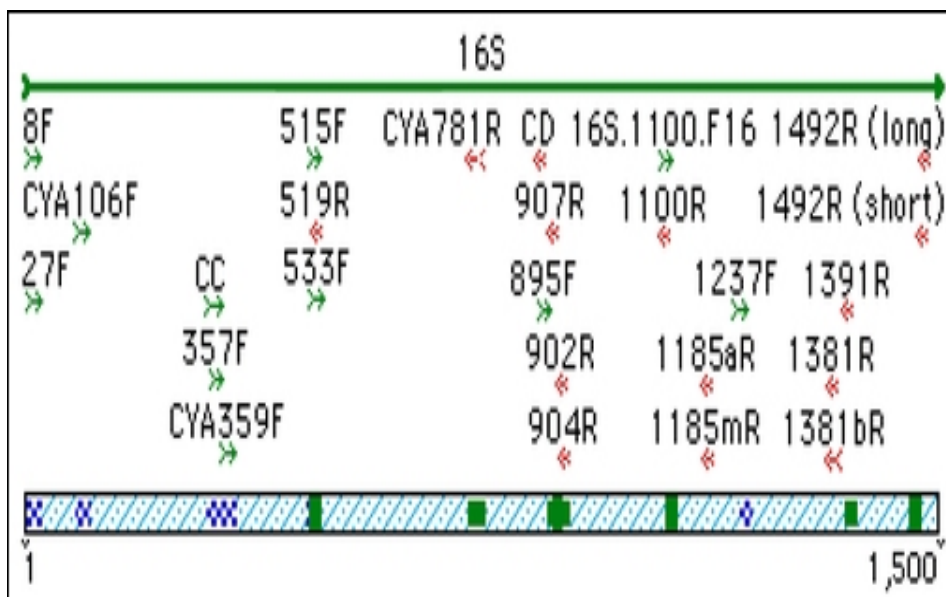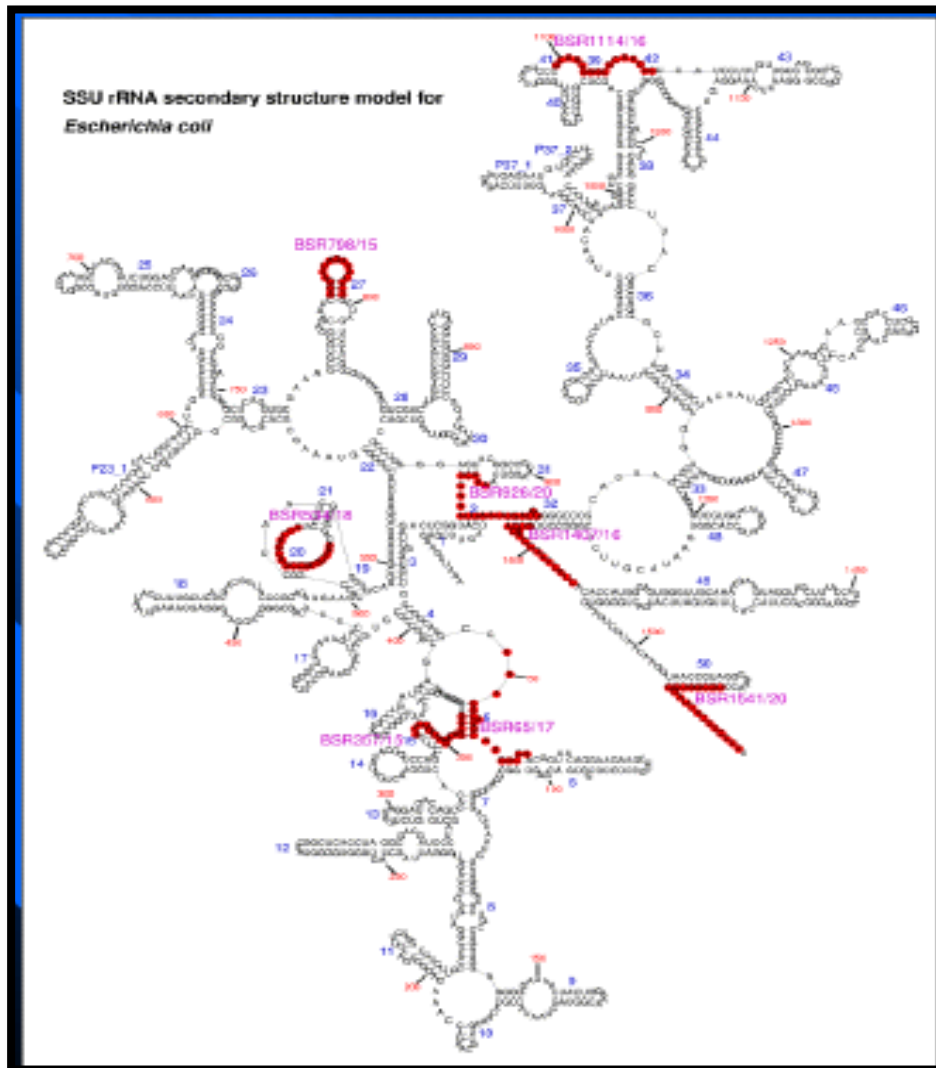
Primer		
Name	Sequence[a]	Amplified hypervariable region
V3F	5′ CCAgACTCCTACGGGAGGCAG 3′ (334–354)	V3 (334–537)[b]
V3R	5′ CGTATTACCGCGGCTGCTG 3′ (519–537)	
V6F	5′ TCGAtGCAACGCGAAGAA 3′ (961–78)	V6 (986–1043)
V6R	5′ ACATtTCACaACACGAGCTGACGA 3′ (1062–85)	
Molecular beacon probe[c]		
Name	Sequence	Target region
SEP-V6 probe	TxR-5′ tgcgcCTAGAGGGGTCAGAGGATgcgca 3′-BHQ2	1005–1022*

Other genes for identification or differentiation of bacteria

- 23s
- 16s-23s ITS
- *rpoB*
- *gyrB*
- *Hsp*
- *recB*

Bioinformatics

Bioinformatics is a new science that uses computational approaches to answer biological questions. Bioinformatics is a new scientific discipline created from the interaction of biology and computer. Biological questions raised from the researchers will be investigated with the large & complex data sets available in public as well as generated by the own laboratory in private to arrive at a valid biological conclusion.
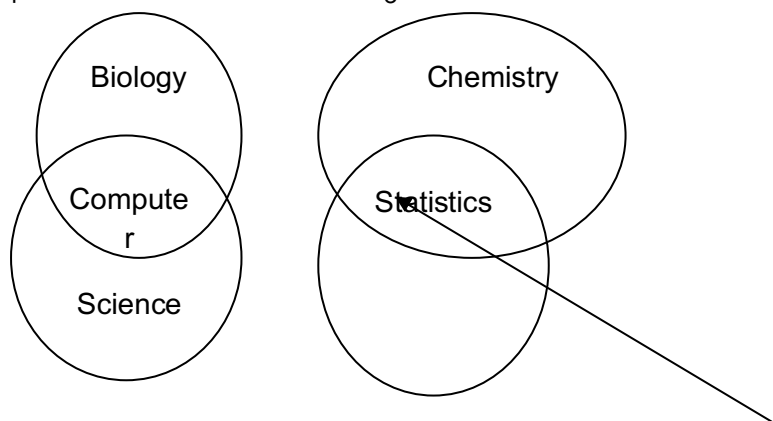
The National Center for Biotechnology Information (NCBI) defines bioinformatics as: "Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline"

Broad Areas in Bioinformatics

- Genomics
- Proteomics
- others

Some of the bioinformatics applicable are

✏Similarity search

✏Sequence comparison: Alignment, multiple alignments, retrieval

✏Sequence's analysis: Signal peptide, transmembrane domain,

✏Protein folding: secondary structure from sequence

✏Sequence evolution: phylogenetic trees

Important terms in Bioinformatics

Fasta sequences

The FASTA format is used in a variety of molecular biology software suites. In its simplest incarnation (as shown above) the "greater than" character (>) designates the beginning of a new file. An identifier (L04459 in the first of the preceding examples) is followed by the DNA sequence in lowercase or uppercase letters, usually with 60 characters per line. Users and databases can then, if they wish, add a certain degree of complexity to this format. For example, without breaking any of the rules just outlined, one could add more information to the FASTA definition line, making the simple format a little more informative, as follows:

>gi|171361|gb|L04459|YSCCYS3A Saccharomyces cerevisiae cystathionine gamma-lyase (CYS3) gene, complete cds.

GCAGCGCACGACAGCTGTGCTATCCCGGCGAGCCCGTGGCAGAGGACCTCGCTTG
CGAAAGCATCGAGTACCGCTACAGAGCCAACCCGGTGGACAAACTCGAAGTCATT
GTGGACCGAATGAGGCTCAATAACGAGATTAGCG

Similarly, the protein record infasta as follows

>P31373

MTLQESDKFATKAIHAGEHVDVHGSVIEPISLSTTFKQSSPANPIGTYEYSRSQNPNRENL
ERAVAALENAQYGLAFSSGSATTATILQSLPQGSHAVSIGDVYGGTHRYFTKVANAHGVE
TSFTNDLLNDLPQLIKENTKLVW

Majority of the procedure analysing either DNA or Protein sequences involves the use of fasta format

Practical on Blast analysis and identification of unknown bacteria from 16s rRNA gene sequence data

1. Check for the quality of sequence data with chromatogram file and pdf
2. Check the quality value of each sequence base call in the chromatogram file
3. Trim the sequence according to the sequence data quality value more than 20
4. Blast analysis of raw and trim sequence data in NCBI_Blast_nucleotide.

5. Perform merging with emboss merger after reverse complementing the reverse sequence data

6. Avoid the low-quality sequences in the analysis.

Agarose gel electrophoresis and documentation

Recommended agarose concentrations for resolution of linear dsDNA Agarose (%) are as per Size of fragments separated (kb) 0.5% is recommended to separate 1-40 kb sized fragments, and 0.7-0.8% for 0.8-12kb, 1.0% for 0.5-10kb, 1.2% for 0.4-7kb, 1.5% for 0.2-3kb, and 2.0% for 0.05-2kb. In general, for the PCR product electrophoresis purpose 1.5 to 2.0% gel is used. For genomic DNA electrophoresis 0.8% is used. Higher the molecular weight, slower the migration. Greater the concentration of agarose, smaller will be the pore size and consequently, slower the migration (usually concentration ranges from 0.7- 2%). Strength of electric field applied: Usually it ranges from 70-100 V. An increase above this range results in faster movement of the DNA but the process generates greater amount of heat. Composition and ionic strength of electrophoresis buffer: In the absence of ions, electrical conductivity is minimal and DNA migrates very slowly.

Materials required: 1. Agarose 2. Ethidium bromide (10 mg/ml) 3. Molecular weight marker 4. 50X TAE buffer 5. 6X Gel loading dye

50XTAE (1litre) 242 g of Tris base 57.1 ml of Glacial acetic acid 100 ml of 0.5 M EDTA (pH 8.0) D. H2O

6X Gel loading dye 0.25 % Bromophenol Blue 0.25 % Xylene cyanol 30 % Glycerol in water

Procedure (2.0% gel): 1. Prepare gel tray and place comb in gel tray about 1 inch from one end of the tray. 2. To prepare 100 ml of a 2.0% agarose solution, measure 2.0 g agarose into a glass flask and add 100 ml 1X TAE buffer. Microwave until agarose is dissolved and solution is clear. 3. Allow solution to cool to about 42-45°C before pouring. (Ethidium bromide can be added at this point to a concentration of 0.5 μg/ml). Add 5μl Ethidium bromide and mix well without foaming any bubbles. 4. Pour this warm gel solution into the tray to a depth of about 5 mm. Allow gel to solidify about 30-45 minutes at room temperature. (PS: UV will not penetrate efficiently in thicker gels, so make sure to make thin gels.) 5. To run, gently remove the comb, place tray in electrophoresis chamber, and cover (just until wells are submerged) with 1X TAE buffer. 6. To prepare samples for electrophoresis, add 1 μl of 6x gel loading dye for every 5 μl of DNA solution. Mix well and load 6μl of DNA per well. Load a suitable molecular weight marker in one lane. 7. Electrophorese at suitable volts until dye has migrated two third the length of the gel − *(Electrophoresis Power Pack* Connect the power, connect ‗BLACK (-)'& ‗RED (+)' pins with respective electrode. Set the ‗VOLT' using ‗▲,▼' keys. Press ‗START'). 8. Document the gel after electrophoresis by using Gel Documentation System. 9. Intact DNA appears as orange fluorescent bands. **Notes**: Ethidium bromide (EtBr) migrates in

392

the opposite direction of the DNA during electrophoresis. Hence if electrophoresis takes longer time, the visibility of the DNA reduces. Hence the gel will have to be stained again. EtBr is a powerful mutagen and is toxic and so wear gloves while handling the gel and should be decontaminated after the end of experiment. Ethidium bromide should be handled and disposed of as it is considered hazardous waste. This applies to gloves, pipette tips, test tubes, paper towels, etc., that are grossly contaminated with ethidium bromide as well.

Reference

1. Andreas D. Baxevanis and B. F. Francis Ouellette. BIOINFORMATICS-A Practical Guide to theAnalysis of Genes and Proteins. SECOND EDITION. WileysInerscience, A JOHN WILEY & SONS, INC., PUBLICATION. 2001.

2. Des Higgins and Willie Taylor. Bioinformatics-Sequences, structure, and databanks- A practical approach. Oxford University Press. 2000.

3. Andrzej Polanski and Marek Kimmel. Bioinformatics. Springer-Verlag Berlin Heidelberg 2007

****