Exercise 13 — Phylogeny of prokaryotes based on 16S rRNA genes

Advisor

Kurt Hanselmann, hanselma@botinst.unizh.ch

Reading

Chapters in BBOM 10th: 11.4 - 11.8, 18.3 – 18.5
BBOM: Madigan M.T., J.M. Martinko and J. Parker: "Brock - Biology of Microorganisms", 10th edition, 2003. Prentice Hall.

Objectives

- Familiarize yourself with a number of useful sequence analysis databases which are available via the internet
- Learn to work with a computer-based resource for sequence analysis
- Understand the molecular background of 16S rRNA gene based phyolgeny
- Learn how to generate phylgenetic trees based on 16S rDNA sequence comparison
- Understand how the universal tree of life was derived
- Experience how to identify signature sequences for the design of oligonucleotide hybridization probes

Background

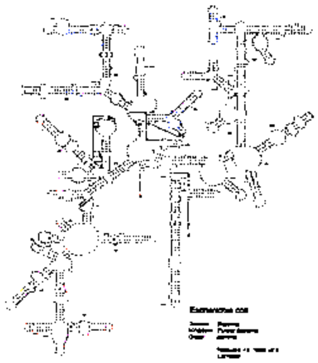
Originally, the taxonomic **classification of prokaryotes** depended exclusively on **phenotypic characteristics** such as

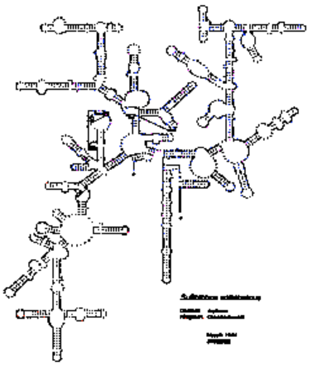
- **shape**: coccus, rod, spirillum, vibrio, etc. (BBOM 10th, 4.4),
- **motility**: movement by gliding, with flagellum or flagella (monotrich, polytrich, peritrich), etc. (BBOM 10th, 4.10-4.11)
- **behavior**: chemotactic, phototactic, magnetotactic (BBOM 10th, 4.12)
- **membrane structure**: e.g. ester-lipids in bacteria vs. ether-lipids in archaea, chlorophylls and other pigments (BBOM 10th, 4.5),
- **cell inclusions and surface structures**: slime layers, capsules, glycogen, sulfur, magnetosomes, chromatophores, carboxysomes, spores etc. (BBOM 10th, 4.13-4.15)
- **metabolism**: phototrophic vs. chemotrophic, lithotrophic vs. organotrophic, autotrophic vs. heterotrophic (BBOM 10th, 5.14)
- **sensitivity and resistance** to antibiotics (BBOM 10th, 20.12)
- **cell wall**: Gram-positive or Gram-negative, LPS (BBOM 10th, 4.8-4.9)
- **pathogenicity, virulence** etc. (BBOM 10th, 1.4, 1.5)
- **and ecological characters:** temperature-, salt-, pH-tolerance, ecotype, etc.

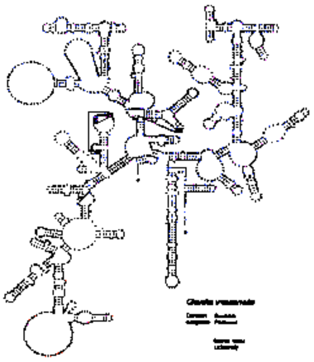
Some of these properties turned out to be good distinguishing characteristics (e.g., the Gram stain), while others are not (e.g., cell shape).

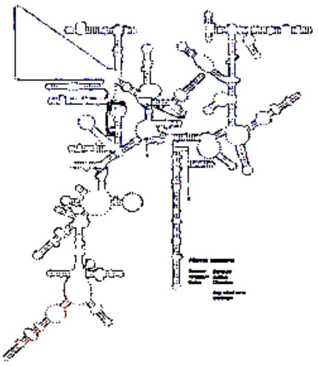
Genotypic classification based on nucleotide sequence comparison of **16S ribosomal RiboNucleicAcid** (**16S rRNA**) genes is now available as an additional taxonomic tool (BBOM 10th, 11.4, 11.5). 16S rRNA, along with the 23S rRNA, has properties which predestine it as a **universal phylogenetic marker**.

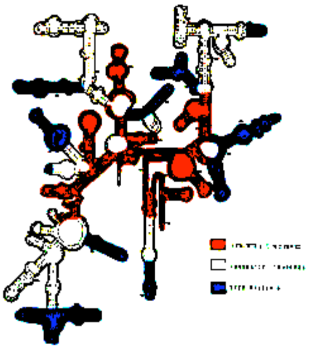
Ribosomes of all living organisms, prokaryotes as well as eukaryotes, consist of a **small subunit** (30S in prokaryotes, 40 S in eukaryotes) and a **large subunit** (50S in prokaryotes, 60 S in eukaryotes), which contain functionally homologous **ribosomal RNA** (16S/18S in the small subunit and 23S/28S and 5S/5-5.8S in the large subunit of prokaryotes/eukaryotes, respectively). S stands for **"Svedberg"** which is a unit for the sedimentation rate during centrifugation) (BBOM 10th, Fig. 11.8, Table 7.4).

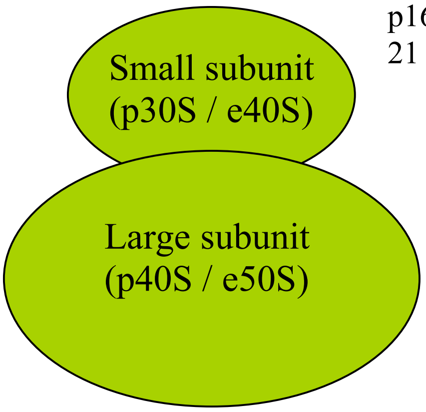












rRNA:
p16S / e18S; p1500 / e1900 nucleotides
21 proteins

p prokaryote
e eukaryote

rRNA:
p23S / e28S; p2900 / e4700 nucleotides
p5S; 120 nucleotides
e5.8S; 160 nucleotides
34 proteins

Ribosomal RNA must have been present since very early in the development of life, because it is essential for protein synthesis. A mutation in the 16S rRNA gene can directly affect the proper functioning of the ribosome and lead to the elimination of less efficient organisms through selection. One may assume, therefore, that the 16S rRNA genes contain a large number of **highly conserved sequence patterns**. There are a number of sequence differences which do not impair the functioning of the ribosome, however, and which were maintained over evolutionary times. These can be used to distinguish **phylogenetically different organisms**.

There are regions on the 16S rRNA which are quite conserved and others which are variable. Comparing the differences in the **base sequence of 16S rRNA genes** is therefore a good tool to detect **evolutionary changes** and **phyolgenetic relatedness** of organisms. The question remains, however, how many point mutations might have occurred at one base position up to now, from A to G to U and back to A for instance. This uncertainty and the fact that we do not know when the particular mutations took place make it difficult to use the 16S rRNA as an exact **evolutionary clock**.

One branch of **bio-informatics** is studying the relatedness between organisms based on sequence comparison. The molecular sequence data obtained worldwide from sequencing projects are collected in public **databases**. With this information, one can test novelty and possibly function of new sequences, search for homologous patterns and regulatory domains and create hypothetical **phylogenetic relationships** called **evolutionary trees**. The information which is encoded in sequences can be analyzed by comparing against existing sequences whose functions are already known. The computer uses algorithms to find similarities between the query sequence and every sequence in the database and scores them according to the degree of relative similarity. Different databases might lead to differences in scoring depending on the algorithm and the dataset used. It is reasonable, therefore, to conduct searches on at least two different database libraries in order to obtain the best possible homologous sequences.

As an exercise, we will use the database provided by the National Center for Biotechnology Information (NCBI) and apply the **BLAST algorithm** to the nucleotide sequences given below (BLAST = Basic Local Alignment Search Tool). The BLAST algorithm searches for patches of similarity. It is based on ungapped sequence alignments, i.e. the alignments created by BLAST do not allow for gaps, but BLAST does allow multiple hits on the same sequence. With this type of statistical model one increases search speed but one reduces sensitivity, i.e. BLAST might miss certain matches.

The basic concept underlying all **comparing and tree building methods** is: Take two sequences, make them the same length and divide the number of identical nucleotides by the number of all nucleotides in one sequence (and gaps if there are any). What you get is the basic similarity percentage between two sequences (BBOM 10th, 11.10). It is presently agreed, that similarities of less than 95% define different genera, less than 97% different species.

Literature

- Stackebrandt, Erko and Michael Goodfellow (eds.) 1991. "Nucleic Acid Techniques in Bacterial Systematics", John Wiley & Sons; Chichester, New York, Brisbane, Toronto, Singapore. ISBN: 0-471-92906-9
- Peruski Jr., Leonard F. and Anne Harwood Peruski. 1997. “The Internet and the New Biology - Tools for Genomic and Molecular Research”. American Society for Microbiology, Washington DC. ISBN 1-55581-119-1

www. Links

- U.S. National Library of Medicine and National Institutes of Health; **National Center for Biotechnology Information (NCBI)**: http://www.ncbi.nlm.nih.gov/

- The **Ribosomal Database Project (RDP):** http://rdp.cme.msu.edu/

- The **European Molecular Biology Laboratory (EMBL)** Bioinformatics Services: http://www-db.embl.de/jss/EmblGroupsHD/serv_0?t=1&p=1

- EMBL’s **European Bioinformatics Institute (EBI)**: http://www.ebi.ac.uk/index.html

- The **Oligonucleotide Probe Base:** http://www.microbial-ecology.net/probebase/default.asp

- The **ARB Project**: Phylogenetic tree building http://www.arb-home.de/

Exercises

Below you will find a number 16S rDNA nucleotide sequences relating to the rumen organisms listed in table 1 of Experiment 1. Try to find out the following, applying the NCBI internet resources:

1. What are the sequences available in BLAST that share the highest relative level of similarity with the sequences given?
2. To which organisms do the sequences belong to?
3. What is the minimal length of the nucleotide sequence to be typed into the BLAST window to still get the correct nearest organism?
4. Is this length always the same or does it depend on the fragment location which you choose from the entire sequence?
5. Take a small fragment of your favorite sequence and exchange one or more nucleotides in it. Does this new sequence still have the same nearest relatives?
6. Does it matter where along the 16S rRNA gene you introduce mutations?
7. On the data bank of the Ribosomal Database Project (RDP) it is possible to create sequence-based trees. Can we fit the inhabitants of the rumen into adequate groups?
8. Are there rumen organisms of which the entire genome has been sequenced?
9. What can we learn from whole genome sequences?

Organisms and genomic sequences to be analyzed:

the sequences are also available as text files in the course folder

Exercise 1

With this sequence you should find *Prevotella ruminicola* ATCC 19189
Is it a partial or a full sequence for this organism’s 16S rRNA gene?

TNCAATGAAGAGTTTGATCCTGGCTCAGGATNAACGCTAGCTNCAGGCTTAACACA
TGCAAGTCGAGGGGCAGCATAATCGAAGCTTGCTTTGATTGATGGCGACCGGCGCA
CGGGTGAGTAACGCGTATCCAACCTTCCCTATAGTAGAGAATAGCCCGGCGAAAGT
CGGATTAATGCTCTATGTTGTATTTAGAGGACATCTNAAGAATACCAAAGGTTTACC
GCTATAGGATGGGGATGCGTCTGATTAGGTAGTAGGCGGGGTAACGGCCCACCTAG
CCGACGATCAGTAGGGGTTCTGAGAGGAAGGTCCCCCACATTGGAACTGAGACACG
GTCCAAACTCCTACGGGAGGCAGCAGTGAGGAATATTGGTCAATGGACGGAAGTCT
GAACCAGCCAAGTAGCGTGCAGGATGACGGCCCTNTGGGTTGTAAACTNCTTTTAT
ATAGGGATAAAGTCGGGGACGTGTCCCNGTTTGTAGGTACTATATGAATAAGGACC
GGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGGTCNNGGCGTTATCCGGAT
TTATTGGGTTTAAAGGGAGCGCAGGCTGATGATTAAGCGTGACGTGAAATGTAGCC
GCTNAACGGCNNAACTGCGTCGCGAACTGGTTATCTTGAGTGAGTTCGATGTTGGC
GGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCACGAAGAACTCCGATTGCGA
AGGCAGCCAACAAGGCCTNTACTGACGCTAAAGCTCGAAGGTGCGGGTATCGAAC
AGGATTNGATACCCTGGTAGTCCGCACGGTNAACGATGGATGCCCGCTNTTTGCGA
TATACTGTGAGCGGCCAAGAGAAATCGTTAAGCATCCCACCTGGGGAGTACGCCGG
CAACGGTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGAGGAACATGTG
GTTTAATTCGATGATACGCGAGGAACCTTACCCGGGCTTGAACTGCCAGCGAACGA
TTCAGAGATGATGAGGTCCTTCGGGACGCTGGTGGAGGTGCTGCATGGTTGTCGTC
AGCTCGTGCCGTGAGGTGTCGGCTTNAGTGCCATAACGAGCGCAACCCTNTTCTTTA
GTTGCCATCAGGTAATGCTGGGCACTCTGGAGATACTGCCACCGTAAGGTGTGAGG
AAGGTGGGGATGACGTCAAATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGT
TACAATGGGGGGTACAGAGAGTCGGAGCAATCTAATCCTTAAAGCCTTCCTCAGTT
CGGATTGGGGTCTGCAACCCGACCCCATGAAGCTGGATTCGCTAGTAATCGCGCAT
CAGCCATGGCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAAGCCA
TGAAAGCCGGGGGCGCTTGAAGTCCGTGACCGCAAGGATCGGCCTAGAGCGAAAC
TGGAAT

Exercise 2

Is the following a full or a partial 16S rRNA gene sequence? From which organism?

CAGATTGAACGCTGGCGGCAGGCTTAATACATGCAAGTCGAACGGTAGCAGCAGG
AAAGCTTGCTTTTCTGGCTGACGAGTGGCGGACGGGTGAGTAATACCTGGGAAACT
GCCTATAAGAGGGGGATAACACCTGGAAACGGGTGCTAATACCGCATAAACCCTG
AGGGGAAAAGGATGGGAAACCATGCGCTTATAGATGTGCCCAGGAGGGATTAGCT
AGTTGGGGAGGTAAAGGCTCACCAAGGCGACGATCTCTAGCTGGTCTGAGAGGATG
ATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTAG
GGAATATTGCACAATGGGGGAAACCCTGATGCAGCCATGCCGCGTGTGTGAAGAA
GGCCTTCGGGTTGTAAAGCACTTTCAGCGGGGAGGAAGCTGAAGTTCGAACAGGGC
AGAAAGCAGACGTTACCCGCAGAAGAAGCGCCGGCTAACTCCGTGCCAGCAGCCG
CGGTAATACGGAGGGCGCAAGCGTTAATCGGAATAACTGGGCGTAAAGCGCACGC
AGGCGGTTATGCAAGTCAGATGTGAAATCCCAGCGCTTAACGTTGGGCGGTCATTT
GAAACTGCAAGACTAGAGTACTATAAGGGGAGCTGGAATTCCAGGTGTAGCGGTG
AAATGCGTAGATATCTGGAAGAACACCGGTGGCGAAGGCGAGCTTCTGGATAGTA
ACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAG
TCCACGCTGTAAACGATGTCAATTAGGAGTCTAGTCCGAAGAGGCTGGGTTCCGCA
GCTAACGCTTTAAATTGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAAT
GAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGC
GAAGAACCTTACCTGAACTTGACATATACGATGATCTGTCTGAGAGGATGGAGTAC
CCGCAAGGGAATCGTAAACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTCTTTGTTGCCAGCGCGTAAA
GGCGGGAACTCAAAGAAGACTGCCGGTGATAAACCGGAGGAAGGCAGGGATGACG
TCAAGTCATCATGGCCCTTACGTTCAGGGCTACACACGTGCTACAATGGTTTGTACA
GAGGGAAGCGAAGTCGCGAGACGGGGCAGATCTCAGAAAGCAAATCGTAGTCCGG
ATTGGAGTCTGCAATTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGAATCAG
AATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGG
AGTGAATTGCACCAGAAGTAGTTAGCGTAACCGCAAGGGGCGCGATTACCACGGTG
TGGTTTATGACTGGGG

Exercise 3

The accession number of the organism is >gi|75766461|gb|DQ186901.1|
Which organism is represented by this accession number and what is its 16S rRNA gene sequence ?
Does the following sequence line up with the one you found yourself ?

```
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTC
GAACGAGGTGGATTTTAACACTGAATGCTTGAGAACGTAACTTCGAGCGTTGTGGC
GGCGTAGCCGACACTACACTTTATATAAAGTGCTTAACTTGAACTTACGGTCAAGA
AATAATCGAGCATTGAGTGTTAAAAGACACCGAGTGGCAAACGGGTGAGTAACGC
GTAGACAACCTGCCGCAAAGATGGGGACAACAGTCCGAAAGGACTGCTAATACCG
AATGTTGTCAGTTTCTCGCATGGGAGACTGATTAAAGATGGCCTCTACTTGTAAGCT
ATCGCTTTGCGATGGGTCTGCGTCTGATTAGCTGGTTGGTGGGGTAACGGCCTACCA
AGGCGACGATCAGTAGCCGGTCTGAGAGGATGAACGGCCACATTGGAACTGAGAC
ACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTCCGCAATGGGCGAAA
GCCTGACGGAGCAACGCCGCGTGAGTGAAGAAGGGTTTCGGCTCGTAAAGCTCTGT
TGACGGGGACGAACGTGCGGTTTGCAAATAGTGAACAGCAATGACGGTACCCGTCG
AGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGC
GTTGTCCGGAATTATTGGGCGTAAAGGGAGCGCAGGCGGGAAGGCAAGTCAGTCTT
AAAAGTGCGGGGCTCAACCCCGTGATGGGATTGAAACTGTCTTTCTTGAGTGCAGG
AGAGGAAAGCGGAATTCNTAGTGTAGCNGTGAAATGCGTAGATATTAGGAGGAAC
ACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACGCTGANGCTCGAAAGCCAG
GGGAGCGAACGGGATTAGATACCCCGGTAGTCCTGGCCGTAAACGATGAATACTAG
GTGTGGGAGGTATCGACCCCTTCCGTGCCGGAGTTAACGCAATAAGTATTCCGCCT
GGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAG
CGGTGGAGTATGTGGTTTAATTCGACGCAACGCGAAGAACCTTACCAGGGCTTGAC
ATTGAGTGAAATATCTAGAGATAGATACCTCCCTTCGGGGACACGAAAACAGGTGG
TGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGC
GCAACCCCTATCATTTGTTGCCAGCACGTAAAGGTGGGAACTCAAATGAGACTGCC
GCGGACAACGCGGAGGAAGGCGGGGATGACGTCAAGTCATCATGCCCCTTATGTCC
TGGGCTACACACGTACTACAATGGGATGGACAGAGAGCAGCGACCCCGCGAGGGC
AAGCGAACCCCATAAACCATCTCCCAGTTCGGATTGCAGGCTGCAACCCGCCTGCA
TGAAGTCGGAATCGCTAGTAATCGCTGGTCAGCATACAGCGGTGAATACGTTCCCG
GGCCTTGTACACACCGCCCGTCACACCACGGAAGTCATTCACACCCGAAGCCGGTG
GAGGAACCGCAAGGACCTAGCCGTCTAAGGTGGGGGCGATGACTGGGGTGAAGTC
GTAACAAGGTAGCCGT
```

Exercise 4

Introduce mutations into the 16S rRNA gene of *Streptococcus bovis*. How will they affect the outcome of the analysis ?

```
TGCCTAATACATGCAAGTAGAACGCTGAAGACTTTAGCTTGCTAAAGTTGGAAGAG
TTGCGAACGGGTGAGTAACGCGTAGGTAACCTGCCTACTAGCGGGGGATAACTATT
GGAAACGATAGCTAATACCGCATAACAGCATTTAACTCATGTTAGATGCTTGAAAG
GAGCAATTGCTTCACTAGTAGATGGACCTGCGTTGTATTAGCTAGTTGGTGAGGTA
ACGGCTCACCAAGGCGACGATACATAGCCGACCTGAGAGGGTGATCGGCCACACT
GGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCGGC
AATGGGGGCAACCCTGACCGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATC
GTAAAGCTCTGTTGTAAGAGAAGAACGTGTGTGAGAGTGGAAAGTTCACACAGTGA
CGGTAACTTACCAGAAAGGGACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGT
AGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTTAAT
AAGTCTGAAGTTAAAGGCAGTGGCTTAACCATTGTTCGCTTTGGAAACTGTTAGACT
TGAGTGCAGAAGGGGAGAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGATAT
ATGGAGGAACACCGGTGGCGAAAGCGGCTCTCTGGTCTGTAACTGACGCTGAGGCT
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACG
ATGAGTGCTAGGTGTTAGGCCCTTTCCGGGGCTTAGTGCCGCAGCTAACGCATTAA
GCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGG
GCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTAC
CAGGTCTTGACATCCCGATGCTATTCCTAGAGATAGGAAGTTTCTTCGGAACATCGG
TGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCC
GCAACGAGCGCAACCCCTATTGTTAGTTGCCATCATTAAGTTGGGCACTCTAGCGA
GACTGCCGGTAATAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCT
TATGACCTGGGCTACACACGTGCTACAATGGTTGGTACAACGAGTCGCGAGTCGGT
GACGGCAAGCAAATCTCTTAAAGCCAATCTCAGTTCGGATTGTAGGCTGCAACTCG
CCTACATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACG
TTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAG
TCGGTGAGGTAACCTTTTAGGAGCCAGCCGCCTAAGGTGGGATAGATGATTGGGGT
GAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCAC
```

Exercise 5

The following two sequences stem from 16S rRNA genes *of Ruminococcus flavefaciens* strain Y1 and of *Ruminococcus albus*, respectively. Is it justified to classify these two organisms under the same genus?

Ruminococcus flavefaciens strain Y1
TGGCTCAGGACGAACGCTGGCGGCACGCTTAACACATGCAAGTCGAACGGAGATAT
TTTAAGTTTACTTAGGATATCTTAGTGGCGGACGGGTGAGTAACACGTGAGCAACC
TACCTTAGAGAGAGGGATAGCTTCTGGAAACGRGTGGTAATACCTCATAACATACA
GAGACCGCATGATTTCTGTATCAAAGATTTATCACTCTGAGATGGGCTCGCGTCTGA
TTAGATAGTTGGTGAGGTAACGGCTCACCAAGTCGACGATCAGTAGCCGGACTGAG
AGGTTGAACSGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGC
AGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGATGCCGCGTGGAGG
AAGAAGGTTTTCGGATTGTAAACTCCTGTCTTAAAGGACGATAATGACGGTACTTT
AGGAGGAAGCTCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGAGCG
AGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCGGGAGTGCAAGTCAG
ATGTGAAATACATGGGCTCAACCCATGGGCTGCATTTGAAACTGCATTTCTTGAGTG
AAGTAGAGGTAAGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCAGGAG
GAACAYCGGTGGCGAAGGCGGCTTACTGGGCTTTTACTGACGCTGAGGCTCGAAAG
CGTGGGGAGCAAACAGGATTAGATAACCCTGGTAGTCCACGCTGTAAACGATGATT
ACTAGGTGTGGGGGGACTGACCCCTTCCGTGCCGCAGTTAACACAATAAGTAATCC
ACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCA
CAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCT
TGACATCGTATGCATAGTCTAGAGATAGATGAAATCCCTTCGGGGACATATAGACA
GGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAAC
GAGCGCAACCCTTACCTTTAGTTGCTACGCAAGAGCACTCTAAAGGGACTGCCGTT
GACAAAACGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGG
GCTACACACGTACTACAATGGCAATTAACAAAGAGAAGCAAGACGGTGACGTGGA
GCGAATCTCAAAAAATTGTCCCAGTTCAGATTGCAGGCTGCAACTCGCCTGCATGA
AGTCGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTCCCCGGGC
CTTGTACACACCGCCCGTCACACCATGGGAGTCGGTAACACCCGAAGTCGGTAGTC
TAACAGCAATGAGGACGCCGCCGAAGGTGGGATTGATGACTGGGGTGAAGTCTAC
AAGGT

Ruminococcus albus
AGAGTTTGATCCATGGCTCAGGACGAACGCTGGCGGCACGCTTAACACATGCAAGT
CGAACGAGCGAAAGGTGCTTGCACCTCTAGCTAGTGGCGGACGGGTGAGTAACAC
GTGAGCAATCTGCCTTTCGGTGGGGGATACCAATTGGAAACGATTGTTAATACCCC
ATAACATAACGAAGCCGCATGACTTTGTTATCAAAGATTTATTGCCGGAAGATGAG
CTCGCGTCTGATTAGGTAGTTGGTGAGGTAACGGCTCACCAAGCCGACGATCAGTA
GCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGAGACACGGCCCAGAYTCCT
ACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGAT
GCCGCGTGAGGGAAGAAGGTTTTAGGATTGTAAACCTCTGTCTTTGGGGACGATAA
TGACGGTACCCAAGGAGGAAGCTCCGGCTAACTACGTGCCGGCAGCCGCGGTAATA
CGTAGGGAGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGAGCGTAGGCGGGA
TGCAAGTCAGGTGTGAAATTTAGGGGCTTAACCCCTGAACTGCACTTGAAACTGTA
GTTCTTGAGTGAAGTAGAGGTAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAG
ATATTAGGAGGAACATCAGTGGCGAAGGCGGCTTACTGGGCTTTAACTGACGCTGA
GGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGATTACTAGGTGTGGGGGGACTGACCCCTTCCGTGCCGCAGTTAACACAA
TAAGTAATCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACG
GGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCT
TACCAGGTCTTGACATCGTACGCATAGCATAGAGATATGTGAAATCCCTTCGGGGA
CGTATAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTA
AGTCCCGCAACGAGCGCAACCCTTACTGTTAGTTGCTACGCAAAAGCGCTCTAGCA
GGACTGCCGTTGACAAAACGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCC
CTTATGACCTGGGCTACACACGTACTACAATGGCTGTTAACAGAGGGTAGCAAAGT
GGTGACACGGAGCAAAACCCTAAAAGCAGTCTTAGTTCGGATTGTAGGCTGCAACC
CGCCTACATGAAGTCGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATA
CGTTCCCGGGCCTTGTACACACCGCCCGTCACGCCATGGGAGTCGGTAACACCCGA
AGCCTGTGTTCTAACCGCAAGGAGGAAGCAGTCGAAGGTGGGATTGATGACTGGG
GTGAAGTCGTAACAAGGT

Exercise 6

>gi|42627617|emb|AJ505938.1|FSU505938 *Fibrobacter succinogenes* strain R, 16S rRNA gene, partial sequence

Which is the most distantly related organism to this *Fibrobacter succinogenes* strain of the ones which you analyzed so far?

```
GAACGCTGGTGGCGTGTCTTATACATGCAAGTCGAGCGAGGCAGCAATGCCGAGCG
GCGAACGGGTGAGTAACGCGTAAGCAATCTGCCCCATATCAGGAAATACCCGTGCC
AACGCGCGGTTAATGTCCAGGAGAGTGGCCCTCTGCATGGAGGGTTGACTAGAGAT
TTATCGGTATGGGATGAGCTTGCGTCCGATTAGCTAGTTGGCGGGGCAACGGCCCA
CCAAGGCGACGATCGGTAGCCGGCCTGAGAGGGTGATCGGCCACATTGGGACTGA
GATACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATATTGCACAATGGGGG
AAACCCTGATGCAGCAACGCCACGTGTGGGAAGAAGCATTTCGGTGTGTAAACCAC
TGTCATGAGGGAATAAGGCCCGCCTTCGGACGGGATTGAATGTACCTTGAGAGGAA
GCACCGGCAAACTTCGTGCCAGCAGCCGCGGTAATACGAGGGGTGCAAGCGTTGTT
CGGAATTACTGGGCGTAAAGGGAGCGTAGGCGGAGATTCAAGCGGATTGTACAAT
CCCGGGGCCCAACCCCGGCTCTGCAGTCCGAACTGGATCTCTTGGATAGTTCAGGG
GCAGGCGGAATTCCTGGTGTAGCGGTGGAATGCGTAGAGATCAGGAAGAACACCG
ATGGCGAAGGCAGCCTGCTGGGGACTTATCGACGCTGAGGCTCGAAAGTGCGGGTA
GCAAACAGGATTAGATACCCTGGTAGTCCGCACCGTAAACAATGCATACTGGGTGT
CCGGGGGTTCCCCCGGGTACCGTAGCCAACGCGTTAAGTATGCCGCCTGGGGAGTA
CGTACGCAAGTATGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG
CATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGGGTTTGACATGGGAAC
GCCGCGGCGAGAGATCGCCGTTTTGCAGCAATGCAACGTTCCGCACAGGTGCTGCA
TGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAA
CCCACGTTTCCAGTTGCCACCCGCAAGGGGGCCCTCTGGAGAGACTGCCGGGGACA
ACCCGGAGGAAGGTGTGGATGACGTCAAGTCCTCATGGCCCTTACATCCTGGGCTA
CACACGTGCTACAATGGTCGGTACAATGGGTCGCAACGCCGCGAGGCGGAGCCAAT
CCTCAAAGCCGTCCTCAGTTCGGATCGGAGTCTGCAACTCGACTCCGTGAAGCTGG
AATCGCTAGTAATCGTGGGTCAGCACACCACGGTGAATACGTTCCCGGGCCTTGTA
CACACCGCCCGTCAAGCCATGGGAGAAGGGAGTGCTCTAAGTCGTGCAAGCGCCTA
AAGCAAGACCTTTGACTGGGGCTAAGTCGTAACAAGGTA
```

Exercise 7

>gi|51512124|gb|AY699275.1| *Butyrivibrio fibrisolvens* isolate L10 16S ribosomal RNA gene, partial sequence

What can you find out about *Butyrivibrio fibrisolvens* from the information which is accessible via the NCBI / all database search?
http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi
Has the organism’s genome been fully sequenced ?

```
AGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAACACATGCAAGTC
GAACGGATTTTGTTCGCTGACAAAGCTTCGGCAGAGTCTTGAACAAAGTTAGTGGC
GGACGGGTGAGTAACGCGTGGGCAACCTGCCTCATACTGGGGGATAACAGCTGGA
AACGACTGTTAATACCGCATAAGCGCACGGTGTCGCATGACACAGTGTGAAAAACT
CCGGTGGTATGAGATGGGCCCGCGTCAGATTAGCTAGTTGGTAAGGTAACGGCTTA
CCAAGGCGACGATCTGTAGCCGGACTGAGAGGTCGGACGGCCACATTGGGACTGA
GACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGGATATTGCACAATGGAGG
AAACTCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATCTCGGTATGTAAAGCTC
TATCAGCAGGGAAGAAAGGCTCGCAAGAGAGATGACGGTACCTGACTAAGAAGCC
CCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGG
ATTTACTGGGTGTAAAGGGAGCGCAGACGGTTATGCAAGTCTGAAGTGAAACCCCA
CGGCTCAACCGTGGGCTTGCTTTGGAAACTGTATAACTAGAGTACTGGAGAGGTAA
GCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACATCGGTGGC
GAAGGCGGCTTACTGGACAGCAACTGACGTTGAGGCTCGAAGGCGTGGGGAGCAA
ACAGGATTAGATACCCTGGTAGTCCACGCGGTAAACGATGAATACTAGGTGTTGGG
GACCAAAGGTCTTCAGTGCCGTCGCAAACGCAGTAAGTATTCCACCTGGGGAGTAC
GTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGC
ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGATCTTGAGATCCTGCTG
AACATCAGGTAACGCTGATGGGCCTTCGGGACAGTAGAGACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTT
GTCCATAGTAGCCAGCAGTAAGATGGGCACTCTATGGAGACTGCCAGGGATAACCT
GGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGATCTGGGCCACACA
CGTGCTACAATGTCGTAACAAAGGGACGCGAAGGAGCGATCCGGAGCAAATCTCA
```

AAAATAACGACCCAGTTCGGACTGTAGGCTGCAACCCGCCTGCACGAAGCTGGAAT
CGCTAGTAATCGCAGATCAGCATGCTGCGGTGAATACGTTCCCGGGTCTTGTACAC
ACCGCCCGTCACACCATGGGAGTCGGAAATGCCCGAAGCCGGTGACTTAACCGTAA
GGAGAGAGCCGTCGAAGGCAGGTCGGATAACTGGGGTGAAGTCGTAACAAGGTAG
CCGTAGGAGAACCTGCGGCTGGAACACCTCCTT

Exercise 8

>gi|293025|gb|AH000856.1|SEG_LAHRRDA0 partial 16S ribosomal RNA

For *Lachnospira multiparus* you have a sequence with 1119 bases. What is the least number of bases needed to still blast for *Lachnospira multiparus* ?

GCGTGGGTAACCTGCCTTATACAGGGGGATAGCAGCTGGAAACGGCTGATAATACC
GCATAAGCGCACGGCATCGCATGATGCAGTGTGAAAAACTCCGGTGGTATAAGATG
GACCCGCGTCTGATTAGCTAGTTGGTGAGGTAACGGCCYACCAAGGCAACGATCAG
TAGCCGACCTGAGAGGGTGACCGGCCACATTGGGACTGAGACACGGCCCAGACTCC
TRCGGGAGGCNGCMGTGGGGAATATTGCACMATGGAGGAAACTCTGATGCAGCGA
CGCCGCGTGAGCGAAGAAGTATTTCGGTATGTAAAGCTCTATCAGCAGGGAAGATA
ATGACGGTACCTGACTAAGAAGCTCCGGCTAAATACGTGCCAGCAGCCGCGGTAAT
ACGTATGGAGCNAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGTGTAGGTGGCA
AAGCAAGTCAGTAGTGAAAATCCGGGGCTCAACCTCGGAACTGCTATTGAAACTGT
TTAGCTAGAGTGCAGGAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTA
GATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTG
AGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGT
AAACGATGAATACTAGGTGTTGGGTCTCATAAGAGATTCGGTGCCGCACGTAACGC
AATAAGTATTCCNNCTGGGGAGCAAGTATGTAATGTACTCTTTCTTCGGAACATCG
GTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCC
CGCAACGAGCGCAACCCCTATTTCTAGTAGCCAGCAGTTCGGCTGGGCACTCTAGA
GAGACTGCCAGGGATAACCTGGAGGAAGGTGGGGATGACGTCNAATCATCATGCC
CCTTATGACTAGGGCTACRCACGTGCTACNATGGCGTAAACNAAGTGAAGCGAGVG
TGTGAGCTTAAGCAAATCACAAAAATAACGTCTCNGTTCGGATTGTAGTCTGCAAC
TCGACTACATGAAGCTGGAATCGCTAGTAATCGCAGATCNGAATGCTGCGGTGAAT
A

Exercise 9

>gi|15408522|dbj|AB070610.1| *Lactobacillus sp.* B5407 gene for 16S rRNA

This is the partial 16S rRNA gene sequence (1481 bp) of a *Lactobacillus sp*. Would it be allowed to name the species ?

CTGGCGGCGTGCCTAATACATGCAAGTCGAACGAACTCTGGTATTGATTGGTGCTT
GCATCATGATTTACATTTGAGTGAGTGGCGAACTGGTGAGTAACACGTGGGAAACC
TGCCCAGAAGCGGGGGATAACACCTGGAAACAGATGCTAATACCGCATAACAACTT
GGACCGCATGGTCCGAGTTTGAAAGATGGCTTCGGCTATCACTTTTGGATGGTCCCG
CGGCGTATTAGCTAGATGGTGGGGTAACGGCTCACCATGGCAATGATACGTAGCCG
ACCTGAGAGGGTAATCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGG
AGGCAGCAGTAGGGAATCTTCCACAATGGACGAAAGTCTGATGGAGCAACGCCGC
GTGAGTGAAGAAGGGTTTCGGCTCGTAAAACTCTGTTGTTAAAGAAGAACATATCT
GAGAGTAACTGTTCAGGTATTGACGGTATTTAACCAGAAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGT
AAAGCGAGCGCAGGCGGTTTTTTAAGTCTGATGTGAAAGCCTTCGGCTCAACCGAA
GAAGTGCATCGGAAACTGGGAAACTTGAGTGCAGAAGAGGACAGTGGAACTCCAT
GTGTAGCGGTGAAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTG
TCTGGTCTGTAACTGACGCTGAGGCTCGAAAGTATGGGTAGCAAACAGGATTAGAT
ACCCTGGTAGTCCATACCGTAAACGATGAATGCTAAGTGTTGGAGGGTTTCCGCCC
TTCAGTGCTGCAGCTAACGCATTAAGCATTCCGCCTGGGGAGTACGGCCGCAAGGC
TGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAAT
TCGAAGCTACGCGAAGAACCTTACCAGGTCTTGACATACTATGCAAATCTAAGAGA
TTAGACGTTCCCTTCGGGGACATGGATACAGGTGGTGCATGGTTGTCGTCAGCTCGT
GTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTATCAGTTGCC
AGCATTAAGTTGGGCACTCTGGTGAGACTGCCGGTGACAAACCGGAGGAAGGTGG
GGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATG
GATGGTACAACGAGTTGCGAACTCGCGAGAGTAAGCTAATCTCTTAAAGCCATTCT
CAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGTCGGAATCGCTAGTAATCG
CGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAC
ACCATGAGAGTTTGTAACACCCAAAGTCGGTGGGGTAACCTTTTAGGAACCAGCCG
CCTAAGGTGGGACAGATGATTAGGGTG

Exercise 10

Blast the two sequences of *Schwartzia succinivorans* and *Veillonella parvula* using http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi and interprete the results.

>gi|1890230|emb|Y09434.1|SS16SRIBO *Schwartzia succinivorans* 16S rRNA gene, partial sequence

```
AGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGGCGATTTGAAA
AGCTTGCTTTTGATGAGCCTAGTGGCAAACGGGTGAGTAACGCGTAGGCAACCTGC
CCTCAGGATGGGGACAACGGCCCGAAAGGACCGCTAATACCGAATGACGTATCGG
GATCGCATGATCCTGATACCAAAGGCGGAGCAATCCGTCACCTGAGGATGGGCCTG
CGTCTGATTAGGTAGTTGGTGAGGTAACGGCCCACCAAGCCGACGATCAGTAGCCG
GTCTGAGAGGATGAACGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGG
GAGGCAGCAGTGGGGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCAACGCCG
CGTGAGTGAAGAAGGTCTTCGGATCGTAAAGCTCTGTTGTCGGGGACGAAGGGGCA
GATCAGGAAATGGGTCTGCTTAGACGGTACCTGACGAGGAAGCCACGGCTAACTAC
GTGCCAGCAGCCGCGGTAATACGTAGGCGGCAAGCGTTGTCCGGAATCATTGGGCG
TAAAGGGAGCGCAGGTGGGACGGTAAGTCTTTCTTAAAAGTGCGGGGCTCACCCCC
GTGAGGGGAAAGAAACTATCGTTCTTGAGTGCAGGAGAGGAAAGCGGAATTCCTA
GTGTAGCGGTGAAATGCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGGCTTT
CTGGACTGTAACTGACACTGAGGCTCGAAAGCCAGGGGAGCGAACGGGATTAGAT
ACCCCGGTAGTCCTGGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCC
TTCTGTGCCGGAGTTAACGCAATAAGTATCCCGCCTGGGGAGTACGGTCGCAAGAT
TGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAAT
TCGACGCAACGCGAAGAACCTTACCAGGGCTTGACATTGATCGAAAGCGCCAGAG
ATGGTGCCCTCTCTTCGGAGACGAGAAAACAGGTGGTGCATGGCTGTCGTCAGCTC
GTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATGCTTTGTTGC
CAGCACGTAAAGGTGGGCACTCAAAGGAGACTGCCGCGGAGAACGCGGAGGAAGG
CGGGGATGACGTCAAGTCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACA
ATGGGATGGACAGAGAGCAGCGACCCCGCGAGGGTAAGCGAACCTCATAAACCAT
CTCCCAGTTCGGATTGCAGGCTGAAACCCGCCTGCATGAAGTTGGAATCGCTAGTA
ATCGCAGGTCAGCATACTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCG
TCACACCACGGAAGTCATTCACACCCGAAGCCGGCGCAGCCGTCTAAGGTGGGGGC
GATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAA
```

>gi|62766472|gb|AY995769.1| *Veillonella parvula* strain ATCC 17745 16S ribosomal RNA gene, partial sequence

```
GCGAAGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGAAGAGCGATGGAAGCTTGCTTCTA
TCAATCTTAGTGGCGAACGGGTGAGTAACGCGTAATCAACCTGCCCTTCAGAGGGGGACAACA
GTTGGAAACGACTGCTAATACCGCATACGATCTAATCTCGGCATCGAGGATAGATGAAAGGTG
GCCTCTACATGTAAGCTATCACTGAAGGAGGGGATTGCGTCTGATTAGCTAGTTGGAGGGGTA
ACGGCCCACCAAGGCGATGATCAGTAGCCGGTCTGAGAGGATGAACGGCCACATTGGGACTG
AGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTTCCGCAATGGACGAAAGTC
TGACGGAGCAACGCCGCGTGAGTGATGACGGCCTTCGGGTTGTAAAGCTCTGTTAATCGGGAC
GAAAGGCCTTCTTGCGAATAGTGAGAAGGATTGACGGTACCGGAATAGAAAGCCACGGCTAA
CTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAA
AGCGCGCGCAGGCGGATMGGTCAGTCTGTCTTAAAAGTTCGGGGCTTAACCCCGTGATGGGAT
GGAAACTGCCAATCTAGAGTATCGGAGAGGAAAGTGGAATTCCTAGTGTAGCGGTGAAATGC
GTAGATATTAGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGACGAAAACTGACGCTGAGG
CGCGAAAGCCAGGGGAGCGAACGGGATTAGATACCCCGGTAGTCCTGGCCGTAAACGATGGG
TACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACGCAATAAGTACCCCGCCTG
GGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAG
TATGTGGTTTAATTCGACGCAACGCGAAGAACCTTACCAGGTCTTGACATTGATGGACAGAAC
CAGAGATGGTTCCTCTTCTTCGGAAGCCAGAAAACAGGTGGTGCACGGTTGTCGTCAGCTCGT
GTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCTTATGTTGCCAGCACTTT
GGGTGGGGACTCATGAGAGACTGCCGCAGACAATGCGGAGGAAGGCGGGGATGACGTCAAAT
CATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGGAGTTAATAGACGGAAGCGA
GATCGCGAGATGGAGCAAACCCGAGAAACACTCTCTCAGTTCGGATCGTAGGCTGCAACTCGC
CTACGTGAAGTCGGAATCGCTAGTAATCGCAGGTCAGCATACTGCGGTGAATACGTTCCCGGG
CCTTGTACACACCGCCCGTCACACCACGAAAGTCGGAAGTGCCCAAAGCCGGTGGGGTAACCT
TCGGGAGCCAGCCGTCTAAGGTAAAGTCGATGATTGGGGTGAAGTCGTAACAAGGTAG
```

Exercise 11

>gi|73487026|gb|DQ146765.1| *Megasphaera elsdenii* strain 5T 16S ribosomal RNA gene, partial sequence

Find out whether the following sequence of *Megasphaera elsdenii* strain 5T contains a foreign vector or a sequence contamination using VecScreen at http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html

TCACCAGCCTTAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAAC
ACATGCAAGTCGAACGAGAAGAGATGAGAAGCTTGCTTCTTATCGATTCGAGTGGC
AAACGGGTGAGTAACGCGTAAGCAACCTGCCCTTCAGATGGGGACAACAGCTGGA
AACGGCTGCTAATACCGAATACGTTCTTTTTGTCGCATGGCAGAGGGAAGAAAGGG
AGGCTCTTCGGAGCTTTCGCTGAAGGAGGGGCTTGCGTCTGATTAGCTAGTTGGAG
GGGTAACGGCCCACCAAGGCGACGATCAGTAGCCGGTCTGAGAGGATGAACGGCC
ACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATCTT
CCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAACGATGACGGCCTTCG
GGTTGTAAAGTTCTGTTATACGGGACGAATGGCGTAGCGGTCAATACCCGTTACGA
GTGACGGTACCGTAAGAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAA
TACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCGCGCAGGCGGC
GTCGTAAGTCGGTCTTAAAAGTGCGGGGCTTAACCCCGTGAGGGGACCGAAACTGC
GATGCTAGAGTATCGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGT
AGATATTAGGAGGAACACCAGTGGCGAAAGCGGCTTTCTGGACGACAACTGACGCT
GAGGCGCGAAAGCCAGGGGAAGCAAACGGGATTAGATACCCCGGTAGTCCTGGCC
GTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAAC
GCAATAAGTATCCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATT
GACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAATTCGACGCAACGCGAAG
AACCTTACCAAGCCTTGACATTGATTGCTATGGGTAGAGATACCCAGTTCCTCTTCG
GAGGACAAGAAAACAGGTGGTGCACGGCTGTCGTCAGCTCGTGTCGTGAGATGTTG
GGTTAAGTCCCGCAACGAGCGCAACCCCTATCTTCTGTTACCAGCGAGTTAAGTCG
GGGACTCAGGAGAGACTGCCGCAGACAATGCGGAGGAAGGCGGGGATGACGTCAA
GTCATCATGCCCCTTATGGCTTGGGCTACACACGTACTACAATGGCTCTTAATAGAG
GGAAGCGAAGGAGCGATCCGGAGCAAGCCCCAAAAACAGAGTCCCAGTTCGGATT
GCAGGCTGCAACTCGCCTGCATGAAGCAGGAATCGCTAGTAATCGCAGGTCAGCAT
ACTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAAAGT
CATTCACACCCGAAGCCGGTGAGGTAACCTTTTGGAGCCAGCCGTCGAAGGTGGGG
GCGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGT

Exercise 12

What is Ruminobacter amylophilum ? Why is this organism also called Bacteriodes amylophilus

>gi|2739312|emb|Y15992.1|RAY15992 Ruminobacter amylophilus 16S rRNA gene
AAAGTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCTTAATACAT
GCAAGTCGAACGGTAACAGCAGGAAGCTTGCTTCCTGGCTGACGAGTGGCGGACG
GGTGAGTAATACCTGGGGAGCTGCCTGAATGAGGGGGACAACACCTGGAAACGGG
TGCTAATACCGCGTAAGCCTGAGGGGGAAAGGCTGGGCAACCAGTCGCATTCAGAT
GCGCCCAGGTGGGATTAGCTAGTTGGTGGGGTAACGGCCTACCAAGGCGACGATCT
CTAGCTGGTCTGAGAGGATGATCAGCCACACTGGAACTGAGACACGGTCCAGACTC
CTACGGGAGGCAGCAGTAGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCC
ATGCCGCGTGTGTGAAGAAGGCCTTTGGGTTGTAAAGCACTTTCAGTATGGAGGAA
GTGTAGTATGTTAACAGCATGCTGCATTGACGTTACATACAGAAGAAGCACCGGCT
AACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCGAGCGTTAATCGGAATAAC
TGGGCGTAAAGAGCTCGTAGGCGGTTTGTCAAGTCAGATGTGAAAGCCCCGGGCTT
AACCTGGGAACCGCATTTGAAACTGACAGACTAGAGTACTGTAGAGGGAGGTAGA
ATTCCAGGTGTAGCGGTGAAATGCGTAGATATCTGGAGGAATACCGGTGGCGAAGG
CGGCCTCCTGGACAGAGACTGACGCTGAGGAGCGAAAGCGTGGGGAGCAAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCAATTAGAAGCATGTTGCCA
TGAGTAGTGTGTTTCTAAGCTAACGCGATAAATTGACCGCCTGGGGAGTACGGCCG
GCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGT
GGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGACTTGACATATTGAGAAGTA
TTTAGAGATAGATACGTGCCGCAAGGAGCTCAAATACAGGTGCTGCATGGCTGTCG
TCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTCT
TTGTTGCCAGCACGTAAAGGTGGGAACTCAAAGAAGACTGCCGGTGACAAACCGG
AGGAAGGCAGGGATGACGTCAAGTCATCATGGCCCTTACGTCCAGGGCTACACACG

TGCTACAATGGGTTGTACAGAGGGAAGCGAAGTCGCGAGGTAGAGCGGAACCCAG
AAAGCAACTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATC
GCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACA
CCGCCCGTCACACCATGGGAGTGAATTGCACCAGAAGTAGTTAGCTTAACCCGCAA
GGGAGGGCGATTACCACGGTGTGGTTTATGACTGGGGTGAAGTCGTAACAAGGTAA
CCATAGGGGAACCTGTGGTTGGATCACCTCCTTA

Exercise 13

Where will you find out more about the following two microorganisms?

>gi|37719324|gb|AY196666.1| *Methanobrevibacter ruminantium* 16S ribosomal RNA gene, partial sequence

GCTCAGTAACACGTGGATAACCTACCCTTAGGACTGGGATAACCCTGGGAAACTGG
GGCTAATACCGGATAGATAATTTTTCCTGGAATGGTATTTTGTTTAAATGTTTTTTCG
CCTAAGGATGGGTCTGCGGCAGATTAGGTAGTTGGTTAGGTAATGGCTTACCAAGC
CTATGATCTGTACGGGTTGTGAGAGCAAGAGCCCGGAGATGGAACCTGAGACAAG
GTTCCAGGCCCTACGGGGCGCAGCAGGCGCGAAACCTCCGCAATGTGAGAAATCGC
GACGGGGGGATCCCAAGTGCCATTCTTAACGGGATGGCTTTTCTTAAGTGTAAAAA
GCTTTTGGAATAAGAGCTGGGCAAGACCGGTGCCAGCCGCCGCGGTAACACCGGCA
GCTCTAGTGGTAGCTACTTTTATTGGGCCTAAAGCGTTCGTAGCCGGTTTGATAAGT
CACTGGTGAAATCCTATAGCTTAACTGTGGGAATTGCTGGTGATACTGTTGAACTTG
AGGTCGGGAGAGGTTAGCGGTACTCCCAGGGTAGAGGTGAAATTCTGTAATCCTGG
GAGGACCACCTGTGGCGAAGGCGGCTAACTGGAACGAACCTGACGGTGAGGGACG
AAAGCCAGGGGCGCGAACCGGATTAGATACCCGGGTAGTCCTGGCCGTAAACGAT
GCGGACTTGGTGTTGGGATGGCTTTGAGCCGCTCCGGTGCCGAAGGGAAGCTGTTA
AGTCCGCCGCCTGGGAAGTACGGTCGCAAGACTGAAACTTAAAGGAATTGGCGGG
GGAGCACCACAACGCGTGGAGCCTGCGGTTTAATTGGATTCAACGCCGGACATCTC
ACCAGGAGCGACAGCTGTATGATTACCAGGCTGATGACCTTGTTTGACTAGCTGAG
AGGAGGTGCATGGCCGCCGTCAGCTCGTACCGTGAGGCGTCCTGTTAAGTCAGGCA
ACGAGCGAGACCCACGCCCTTAGTTACCATCAGATTCTCCGGAATGCTGGGCACAC
TAAGGGGACCGCCAGTGATAAACTGGAGGAAGGAGTGGACGACGGTAGGTCCGTA
TGCCCCGAATCCCCTGGGCTACACGCGGGCTACAATGGCTGGGACAATGGGTTCCG
ACGCCGAAAGGCGAAGGCAATCTTTTAAACCTAGTCGTAGTTCGGATTGAGGGCTG
TAACTCGCCCTCATGAAGCTGGAATGCGTAGTAATCGCGTGTCACAATCGCGCGGT
GAATACGTCCCTGCTCCTTGCACACACCG

>gi|175257|gb|M59142.1|MHARR16S *Methanomicrobium mobile* 16S ribosomal RNA

ATTCTGGTTGATCCTGCCAGAGGTCACTGCTATCGGGGTTCGATTAAGCCATGCGAG
TCGAGAGGGTTCAGACCCTCGGCGGACTGCTCAGTAACACGTGGATAATCTGCCCT
ATGGTGGAGGATAACCCCGGGAAACTGGGGATAATACTCCATAAGTCAGGAGTACT
GGAATGTCTCCTAACTGAAAGTTGCGGCGCCATAGGATGAATCTGCGGCCGATTAG
GTAGTTGTTGGGGTAACGGCCCAACAAGCCCGTAATCGGTACGGGTTGTGGGAGCA
AGAGCCCGGAGATGGAATCTGAGACACGATTCCAGGCCCTACGGNNNGCAGCAGG
CGCGAAAACTTTACAATGCAGGAAACTGTGATAAGGGAACCCCGAGTGCCCGTACA
CGCGGGCTGTCCCGGTGTTTAAAACACATCGGAAGAAAGGGCCGGGCAAGACCGG
TGCCAGCCGCCGCGGTAATACCGGCNNCTCGAGTGGTGACCACTTTTATTGGGCTT
AAAGCGTTCGTAGCTTGATTTTTAAGTCTCTTGGGAAATCTCACGGCTTAACTGTGA
GGCGTCTAAGAGATACTGGGAATCTAGGGACCGGGAGAGGTAAGAGGTACTTCAG
GGGTAGAAGTGAAATTCTGTAATCCTTGAGGGACCACCGATGGCGAAGGCATCTTA
CCAGAACGGCTTCGACAGTGAGGAACGAAAGCTGGGGGAGCGAACGGGATTAGAT
ACCCCGGTAGTCCCAGCCGTAAACTATGCGCGTTAGGTGTGCCTGTAACTACGAGT
TACCGGGGTGCCGAAGTGAAAACGTGAAACGTGCCGCCTGGGAAGTACGGTCGCA
AGGCTGAAACTTAAAGGAATTGGCGGGGGAGCACCACAACGGGTGGAGCCTGCGG
TTTAATTGGACTCAACGCCGGGCAGCTCACCGGATAGGACAGCGGAATGATAGCCG
GGCTGAAGACCTTGCTTGACCAGCTGAGAGGAGGTGCATGGCCGTCGTCAGTTCGT
ACTGTGAAGCATCCTGTTTAGTCAGGCAACGAGCGAGACCCACGCCAACAGTTGCT
AGCATGTTCTCCGGAATGATGAGGACACTGTTGGGACCGCCTCTGCTAAAGAGGAG
GAAGGAATGGGCAACGGTAGGTCAGCATGCCCCGAATTATCCGGGCTACACGCGG
GCTACAATGGTCAGGACAATGGGAAACGGCACTGAAAAGTGTAGTCAATCTCCTAA
ACCTGTCCCAAGTTCGGATTGTGGGCTGCAACTCGCCCACATGAAGCTGGAATCCG
TAGTAATCGCGTTTCAAAATAGCGCGGTGAATAAGTCCCTGCTCCTTGCACACACC
GCCCGTCAAACCATCTTAGTGAGGTTTGGATGAGGCTGTGGTTAACTGCCACAGTC
GAATCTAGGTTTCGCAAGGAGGGTTAAGNNGTAACAAGNNNNNNNNNNNNGAATC
TGNNNNNNGATCACCTCCT

Equipment	Computer on the internet and browser
Rules & Precautions	No microbial risks
Experiences gained	• Familiarize yourself with a few theoretical and pratical aspects of computer-aided molecular sequence analysis techniques • Experience how the internet can be used to get phylogenetic information from a nucleotide sequence • Perform sequence database searches • Practice to download and understand sequence records from international data bases • Learn about the usefulness and limitations of some computer algorithms for sequence analysis
Timing	90 minutes
Reporting	Take notes on the exercise and present the results in your report.
Questions to be answered	1. What is the RDP ? 2. What is an evolutionary distance (ED) ? 3. Why are ribosomal RNA genes good evolutionary chronometers? Why are they not? 4. Describe why the results of phylogenetic community analyses of microbial habitats have been rather surprising. 5. Why are we using rRNA genes for studying phylogenetic relationships? 6. What effect do slightly different sequences have on the positioning in a phylogenetic group? 7. What is a Svedberg Unit (S) ? The Svedberg Unit (S) equals 10^{-13} seconds; it is used to express sedimentation velocities of colloidal macromolecules. The symbol S is named after Theodor Svedberg, a Swedish chemist, (1884-1971), who pioneered the development of the ultracentrifuge. Svedberg won the Nobel prize for chemistry in 1926 for his work on disperse systems. Since ribosomes and ribosomal particles were first isolated from cell lysates by ultracentrifugation, they were named according to their sedimentation characteristics during centrifugation. The two prokaryotic ribosomal subunits have sedimentation coefficients of 30 x10^{-13} and 50 x 10^{-13}. As one Svedberg (S) unit is 10^{-13}, the two ribosomal subunits are referred to as the 30S and the 50S ribosomal subunits.