

RIBOSOMAL RNA AS A TOOL IN MICROBIAL ECOLOGY

INTRODUCTION

Traditional techniques. Knowledge of bacterial diversity obtained after more than 100 years of pure culture study is incomplete, and very few of the total number of microbial species are in culture (Torsvik *et al.*, 1995). Until the middle of the last decade enumeration and identification of soil microorganisms had to rely on phenotypic methods. However, phenotypic methods are restricted to only those bacteria that can be isolated and cultured. Other, maybe completely unsuspected groups, which may be abundant or very active, will not be considered, rendering the emerging picture of the soil microbial community false. Only a small portion of all prokaryotes has so far been cultivated, and the majority of soil bacteria observed microscopically cannot be cultivated. In addition, the selective enrichment culture has severe limitations as an approach to study the community composition of naturally occurring microorganisms (Ard *et al.*, 1992; Amann *et al.*, 1995). Results with pure culture isolates are not always reproducible due to the variability of phenotypic properties in relation to culture conditions. Furthermore, laboratory cultivation introduces serious bias to community analysis (Boivin-Jahns *et al.*, 1995; Ferris *et al.*, 1996), since the bacterial populations obtained through planting are mainly dependent on the isolation media used (Sørheim *et al.*, 1989), as well as on purification and maintenance procedures. Since nutrient-rich media are used the selection might be biased towards copiotrophic bacteria rather than dominant community members. In addition, properties of microbes *ex situ* may differ, because they lack the interaction among populations or with the natural environment, or change their physiological state as a result of adaptation to the changed environment. Consequently, *in situ* diversity and community structure are unlikely to be represented by collections of natural isolates.

Molecular approaches. Alternatively, molecular approaches based on phylogenetic analyses of biological markers amongst cellular components are capable of better encompassing the diversity of all well represented members of the entire community. Since the mid-1980s, the use of small-subunit ribosomal ribonucleic acid (SSU rRNA) based techniques has facilitated a culture-independent approach of investigating microorganisms as they occur in nature (Olsen *et al.*, 1986; Ward *et al.*, 1992; Amman *et al.*, 1995). The comparison of these molecular "signature" sequences transformed microbial taxonomy from a pure identification system to an evolutionarily-based framework (Gray *et al.*,

1984; Woese, 1987; Olsen *et al.*, 1994). Based on these studies all forms of life are separated into three major evolutionary lines, the three so-called domains: *Bacteria*, *Archaea*, and *Eucarya* (Fox *et al.*, 1980; Woese *et al.*, 1990).

Ribosomal RNA. Ribosomal RNA and the corresponding genes (*rrn*) are now widely used as powerful evolutionary and investigative biomarkers for the following reasons (Olsen *et al.*, 1986): (i) Ribosomal RNAs are essential to protein synthesis, and therefore are ubiquitous to all organisms, and structurally and functionally conserved; (ii) ribosomal RNAs are readily isolated and identified, (iii) they contain variable and highly conserved regions in both primary and secondary structure, (iv) and they appear to change in sequence very slowly, and they do not exhibit horizontal gene transfer found with many other prokaryotic genes; therefore relationships between rRNAs reflect evolutionary relationships. These traits make rRNAs not only the most widely used biomarker, but also a powerful tool for microbial ecology studies, particularly for complex terrestrial environments such as the soil with enormous and undiscovered diversity.

Amongst a variety of cellular biomarkers studied, the SSU rRNA gene provides certain aspects of information that makes it an extremely versatile tool and the best culture-independent biomarker to study microorganisms. Each SSU rRNA gene contains highly conserved regions found among all living organisms as well as unique variable regions and hence diagnostic to certain organisms or related groups. Furthermore, the primary structure of the approximately 1,500 base SSU rRNA gene allows the inference of phylogenies based on comparative sequence analysis. By estimating the phylogenetic relatedness to known microorganisms based on the homology of the gene sequence, the closest affiliation of a newly isolated or molecularly detected microorganism can be established. In combination with a large and growing SSU rRNA database, microorganisms can be sorted according to their phylogenetic affiliation, and conversely, gene probes can be constructed at different levels of specificity. To date, over 6,000 SSU rRNA sequences have been made available for comparison (Maidak *et al.*, 1996). As these databases rapidly expand, they constantly improve the process of matching new sequences to known microorganisms.

Ribosomal RNA Databases. Since the development and widespread application of the polymerase chain reaction (PCR; Saiki *et al.*, 1988), rRNA sequences can directly be obtained from lysed cells, which have contributed to the exponential increase in known prokaryotic SSU rRNA sequences in recent years. Results from molecular ecological studies within the last seven years from marine (Giovannoni *et al.*, 1990; Schmidt *et al.*, 1991), thermophilic (Weller *et al.*, 1991; Ferris *et al.*, 1996), terrestrial (Liesack and Stackebrandt, 1992; Borneman *et al.*, 1996) environments and on symbionts (Amann *et al.*, 1991) documents the success of this strategy.

Potential biases of SSU rDNA-based studies. As with every method, bias might occur during its construction. In community studies the most critical step is the extraction of nucleic acids from the ecosystem. Isolated DNA should reflect the existing genetic diversity. But microbial cells do not lyse equally well.

Especially Gram-positive cells might be underrepresented in the clone library due to their resistance to lysis. Some DNA might get lost during purification treatment.

Another potential source of bias is the PCR reaction, where certain SSU rRNA genes might preferentially be amplified (Amann *et al.*, 1995). The primer sequences used in the PCR-mediated amplification of SSU rDNA should cover the phylogenetic diversity present in the sample, but primers can only be as good as the databases on which they are based. Our databases, however, are focused on economically important strains, *i.e.* medical isolates, food and industry strains, and do not reflect the high diversity of terrestrial and other environmental samples.

A third potential source of bias is the PCR coamplification of mixed genomes, which leads to the formation of chimeric SSU rDNA molecules (Liesack *et al.*, 1991; Wang *et al.*, 1997). PCR-induced chimeras are formed by annealing of partial-length fragments for different SSU rDNA genes via highly conserved regions. The following primer independent elongation phase then creates full fragments, thereby leading to reports of nonexistent organisms.

A fourth factor that could influence selective recovery is copy number variation of the *rm* gene between bacterial species, which could result in a quantitative misrepresentation of the community profile. This render clone numbers of low abundance per phylotype unreliable for quantitative evaluation. Other, less important sources of bias in DNA sequence-based analyses of natural communities have been discussed extensively elsewhere (Ward *et al.*, 1992; Stackebrandt *et al.*, 1993; Liesack *et al.*, 1997).

The error rate of sequencing is difficult to determine. The materials supplied by one of the biggest manufacturers for automated sequencing. ABI, state a 98-99% fidelity rate. Standards in the sequencing facility at MSU typically adhere to that rule. But behind that are questions of positional fidelity (errors in the first hundred bases are less frequent than in the second one hundred bases, etc.), as well as template and primer specific questions. For good sequencing results make more than one run in each direction with fresh primer and purified template.

Despite the care taken in the design of your study, there are still method limitations and biases inherent to each step in an analysis that might influence the results and conclusions. Nevertheless, any study should use currently available methods as critically as possible with the present knowledge to minimize the introduction of potential biases.

SEQUENCE THEORY

DNA Sequencing

Dideoxynucleotide chain termination, as a primary mode of DNA sequencing, is based upon the synthesis of sets of radioactively labeled or fluorescence-labeled fragments of progressively increasing lengths of each nucleotide, from the 5' to 3' ends, of the entire fragment. Synthesis is followed by polyacrylamide gel electrophoretic separation of fragments sets that are either radio-or fluorescence-labeled.

Specific steps include: preparation of a single-stranded DNA fragment, preparation of a primer complementary to the 3' end of the fragment, DNA polymerase I synthesis of progressively increasing DNA length fragments by the introduction of a trace amount of either radioactive (^{32}P labeled dideoxynucleoside triphosphate) or fluorescence-labeled dideoxynucleoside triphosphate (different wavelengths for each nucleotide) into each of four standard synthesis reactions containing complete sets of the four nucleotides. In each of the four reactions, specific to each DNA base, the ratio of the trace labeled nucleotide to the unlabeled nucleotide determines the probability of triphosphate incorporation versus deoxynucleoside triphosphate incorporation. The absence of the 2'-hydroxyl in the ribose sugar of the triphosphate prevents chain elongation that normally occurs through phosphate to deoxyribose 3'-hydroxyl sugar additions. The result, in the presence of the dideoxy substitute, is chain termination. Thus, in the complete absence of dideoxynucleoside triphosphates, chain elongation occurs over the length of the DNA segment; in the exclusive presence of dideoxynucleoside triphosphates, no chain elongation would occur. In the presence of both, probability of incorporation of either allows for synthesis, for each DNA base, of sets of labeled fragments for every base occurrence in the DNA segment. Given the same process for each of the four DNA bases, the generation of sets of labeled fragments for every nucleotide site on the DNA segment provides for a autoradiographic-based or fluorescence-based "read" of the sequence.

It should be noted that DNA sequencing can be accomplished using either DNA and DNA Polymerase I or RNA and Reverse Transcriptase. Sources of DNA include: double-stranded (denature) or single-stranded DNA, examples including intact plasmids or prepared fragments. In identification of microbial isolates, PCR-synthesized fragments are particularly useful due to the specific interest, for all isolates, in 16S RNA. Likewise, DNA primers are derived from various sources, including universal primers and specific oligonucleotide primers. "Primer 3" is a recommended program useful in primer design. Given an appropriate primer for a given DNA sequence, it is generally possible to analyze 500 DNA bps using the dideoxy method of enzymatic chain termination for DNA sequence determination.

TIPS FOR COMMUNITY ANALYSIS

- | | |
|-------|---|
| (i) | Culturing biases can be omitted by using a molecular approach. |
| (ii) | By using the direct lysis method (lysis cells in the sample rather than extracting them first) one can extract high yields of non-sheared DNA. The high molecular weight of the extracted DNA reduces the probability of chimera formation. |
| (iii) | Use updated primers for the PCR-based amplification of SSU rDNA. |
| (iv) | Only 22 to 25 PCR cycles should be used to stay within the quantifiable linear range of the amplification reaction series. |
| (v) | Soil DNA templates are denature immediately before adding them into the PCR reaction assay to free the target DNA from potential contamination, and to optimally expose all target sites. |

STEPS OF SEQUENCE ANALYSIS

GOAL

Learn and understand the steps necessary to make a phylogenetic assignment for a nucleotide sequence (here the bacterial 16S rDNA).

BACKGROUND

The nucleotide sequence of extracted nucleic acid can be compared with a database of already assigned sequences.

PRINCIPLE

1. Extract nucleic acid from isolates or amplify sequence in the polymerase chain reaction.
2. Purify pure culture or amplified DNA using a kit.
3. Send to sequencing facility for PCR based sequencing with an appropriate primer.
4. Edit sequence.
5. Using the RDP Website find closest relative to the primary structure of this sequence with Sequence_Match.
6. Align sequences with aligned sequences of the closest relatives from the RDP database using alignment software. Proof read the alignment: base the alignment of conserved and variable areas on secondary structure of a closely related organism (available from your instructor or specified web pages).
7. Create a distance matrix.
8. Test sequence for potential chimeric structures.
9. Export the aligned product and the aligned sequences of the closest relatives into software able to create a tree of phylogenetic relationship based on specific algorithms.

Now enough is known to design a probe for the analyzed sequence.

WEB BASED SEQUENCE ANALYSIS

GOAL

Search for DNA and protein deposited in the Genebank. Search for papers, information, and documents about your sequences.

BLAST search:

National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/
--

BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all the available sequence databases regardless of whether the query is protein or DNA.

1. Open a window in your WEB browser and type:
<http://www.ncbi.nlm.nih.gov/>
2. You are now in the National Center for Biotechnology Information WEB page.
Go to *Database services* and click on *<BLAST Sequence Similarity Searching>*.
3. Go to BLAST 2.0. Choose *<Basic Blast search>*.
The difference between choosing Basic Blast or Advanced Blast is your degree of familiarity with sequence search programs. If you are a beginner start with “Basic” and experiment the “Advanced Blast” later.
4. The next page will show *<Program>*. Do not change the default choice.

The BLAST family of programs allows all combinations of DNA or protein query sequences with searches against DNA or protein databases:	
blastp	compares an amino acid query sequence against a protein sequence data base
blastn	compares a nucleotide query sequence against a nucleotide sequence database
blastx	compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database
tblastn	compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands)
tblastx	compares the six-frame translations of a nucleotide query sequence against the six-frame translations of s nucleotide sequence database

<Database> will allow you to choose a specific database for your search. Do not change the default. The *<nr>* will allow you to search for all non-redundant nucleotide or protein sequences present in different databases (Genebank, Swiss Protein database, EMBL).

5. Go to the open windows and write:
>Consensus sequence # (29 or 80), hit *<Enter>*
Then copy and paste your consensus sequence into the next line of the open window.
Hit *<Submit Query>* and wait.
Your results will look like that:

Score	E		(bits)
Sequences producing significant alignments:			
Value			
dbjID84585ID84585	Arthrobacter sp. DNA for 16S ribosomal RNA, p.571	e-	
161			
dbjID84573ID84573	Arthrobacter sp. DNA for 16S ribosomal RNA, p.555	e-	
156			
dbjID84561ID84561	Arthrobacter sp. DNA for 16S ribosomal RNA, p.555	e-	
156			

The number before the match takes you to a new WEB page where you can see the sequence and get information about the source of the organism, it's authors, addresses, publications, etc. You can also retrieve the sequence.

The score number informs how many times you query hits the same sequence in the database.

The Expect value (E) is a parameter that describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. For example, an E value of 1 assigned to a hit can be interpreted as in database of the current size one might expect to see 1 match with a similar score simply by chance.

SEQUENCE_MATCH

GOAL

Make sequences comparable at both levels, the primary and the secondary structure.

PRINCIPLE

This is a crude but fast way of measuring similarity between your sequence and unaligned sequences in the RDP database. Neither rRNA nor rDNA similarity values can be used to detect fine differences between highly related species.

This program counts the number of unique oligomers (7-mers in the case of SSU-RNA) in your sequence that are common to other sequences in the database. The obtained result is the number of shared oligomers divided by either the number of unique oligomers in the submitted sequence or the database sequence, whichever is lower. This ensures that a good match between partial sequences; between a partial and a complete sequence it is scored higher than a less good match between complete sequences.

The output shows the sequences that are most similar to yours with the better ones in the top list.

There is no direct relationship between Sequence_Match and the sequence similarity, because Sequence_Match does not use mismatches for its calculation.

For more information see Sequence Match description in the RDP-II website's Info box.

The Sequence Match can also be done with the NCBI website program BLAST. However, BLAST outputs are more complicated, and are not useful for partial sequences.

SEQUENCE ALIGNMENT

GOAL

Make sequences comparable at both levels, the primary and the secondary structure.

PRINCIPLE

Comparing sequence one with sequence two seems simple. Just take the number of matching bases divided by the number of bases you compare which gives you the fraction of sequence similarity. However, it is more complicated than that. The major problem is determining when a base mismatch is the result of a base change and not the result of a base insertion or of a base deletion. In order to minimize the insertion and deletion effect the conserved regions of the sequences have to be aligned, i.e. you align a new sequence derived from an environmental isolate with a group of related sequences. Test sequences are usually aligned with respect to the sequence of *Escherichia coli* 16S rRNA. The Ribosomal Database Project contains a list of many thousand 16S rRNA sequences aligned to each other. We can use this database to find where our sequence would fit best.

THE DISTANCE MATRIX

GOAL

Create a matrix of pair wise aligned sequences to show their similarity in percent.

PRINCIPLE

Distance Matrix will find sequences in the RDP database that are most similar to your sequence. The program works by comparing two sequences (pair wise alignment) that are placed on the X and Y axes of a matrix.

Attention: If you are using NETSCAPE WEB BROWSER, you will need to change all back slashes '\ ' to forward slashes '/' in your file path.

For example, C:\mydir\myfile would become C:/mydir/myfile

The program does not work for INTERNET EXPLORER WEB BROWSER

Distance Matrix considers only ATGC bases during search and analysis. Any other unknown characters (e.g. like an N) are removed. The results are presented as a table. The upper right triangle shows percentage similarity values, the lower left dissimilarity percentages. The similarity values indicate the similarity of your sequence compared to others. The most similar sequences are shown in red colors, the most dissimilar in blue colors. The ID's in boldface are from your data, the hyper linked ones are from RDP.

The areas highlighted in green/yellow are the positions that could be aligned with confidence by the alignment program throughout all your sequences, and are those from which the matrix was calculated. Although the program searches for the most complete sequences in the database that align with your sequence, the DISTANCE MATRIX does not give you an actual phylogenetic analysis. It gives you a good idea of that. The number of displayed organisms is limited, so there may be many almost equally similar sequences not shown. The power of the matrix similarity approach lies in the fact that the real work of finding similarities is done not by the computer, but by the user's brain. It is the user's ability to recognize those numbers which make this approach highly valuable. For more details see http://www.cme.msu.edu/RDP/docs/distance_doc.html.

THE IDEA OF PHYLOGENETIC TREES

GOAL

Make graphical displays of phylogenetic assignments.

PRINCIPLE

To place your sequence on a tree you must **first create an alignment** that includes your sequence and others, preferably closely related, SSU rRNA sequences. This is the most critical step. The alignment is effectively your hypothesis about the evolution of the sequences being aligned. You will have to choose some sequences with which to align yours – download them from the RDP-II or GenBank. You might choose sequences that appear similar to yours on the basis of a BLAST search or you could find the most similar sequences using the Sequence Match tool at the RDP-II. Once the alignment is completed, you will need to edit it, since no algorithm is perfect.

To edit the alignment you need a sequence editor; Seq Pup is free from Don Gilbert and others are available including commercial ones like GCG. (Addresses for *Clustal SeqPup* and a host of other free programs for desktops are available at <http://iubio.bio.indiana.edu/soft/molbio/docs/biosoft-free.html>). Be sure to save your alignment in a format compatible with the phylogeny package you will be using. Before the alignment is used, a mask should also be made so that the most variable positions in the alignment are not used in the phylogenetic analysis. The more distantly related two organisms are (i.e., the longer is has been since they diverged), the more positions must be excluded since in the course of evolution the more variable positions will have change so many times that any information they may have relevant to

evolutionary history on a given timescale is lost. Most phylogeny programs support the inclusion of a mask line in the input.

Next, **decide on the approach to tree making** you want to take. Approaches to tree building may be crudely divided into two categories:

- character-based methods
- distance-based methods

In the later approach distances between taxa are calculated as sequence dissimilarity. The distances are corrected for multiple changes at the same site using a model that specifies the rate of evolutionary change at each site in the molecule. A tree is then constructed based on the table of distances. Methods like *Neighbor Joining* and *UPGMA* use no optimality criteria and construct only one tree. Methods like the *DeSoete tree fit* or *FITCH* evaluate many alternative tree topologies and try to find the topology with the best fit to the data, using various least-squares fit measurements as optimality criteria. Character-based approaches to tree building include maximum parsimony and maximum likelihood methods. In maximum parsimony, a tree is sought that minimizes the number of evolutionary steps (i.e., the number of required base changes to get from one base to another at each position) between all the taxa. Maximum likelihood methods seek the tree that represents the most probable evolutionary hypothesis for the observed data, based on a specific model of evolution. After inferring a tree, you may wish to assess the probability that the tree is the correct one.

Note that if your sequences were generated using the PCR, it is important that you screen them for possible chimeras; over 30% of the clones from a PCR reaction can be chimeric.

The two most popular **packages that incorporate the tree building methods** mentioned above (plus many refinements of them) are **PHYLIP** <http://evolution.genetics.washington.edu/phylip.html> and **PAUP*** <http://www.lms.si.edu/PAUP/>.

A good **introduction to phylogenetics** is Hills et al., 1993, Analysis of DNA Sequence Data: Phylogenetic Inference. Meth.Enzymol.224:456-89. The books "Molecular Evolution" by Wen-Hsiung Li (1997, Sinauer Associates, Sunderland, MA) and "Molecular Systematics" edited by Hillis et al. (1996, Sinauer Associates, Sunderland, MA) are also helpful.

If you plan to do phylogeny on a larger scale the ARB software package from the Technical University of Munich <http://www.mikro.biologie.tu-muenchen.de/> can be very useful. This package comes with a collection of pre-aligned SSU RNA sequences, incorporates Clustal (among other programs) and allows you to edit your alignments taking into account secondary structure of the SSU rRNA. It is possible to export your alignment from ARB in a variety of formats or generate phylogenetic trees within the package.

BOOTSTRAPPING

GOAL

Bootstrapping to rate phylogenetic trees by confidence.

PRINCIPLE

Evolutionary trees are often estimated from DNS or RNA sequence data. But, how much confidence should we have in the estimated trees? In 1985, Felsenstein [Felsenstein, J. (1985) *Evolution* 39, 783-791] suggested the use of the bootstrap to answer this question. If you create a tree it is the best guess of phylogenetic relationships of its members based on the comparison of your aligned sequences. You have looked at 1500 or so bases comprising the 16S rRNA gene as a sample of evolutionary change. Now you want to know whether this sample is large enough to give us the same evolutionary tree we would get if we had the sequences of a lot more genes. To find out if our data are sufficient to create the correct tree you can recalculate it over and over again with independent calculations. If you calculate this tree one hundred times, the most likely tree will appear. Each branch then has a percentage number attached to it showing how often it fell in that position.

Let us assume that our data sample (e.g. 1500 bases from 16S rRNA) is large enough to be representative of the underlying distribution (i.e. the set of all sequences that evolve like 16S rRNA). This implies that our answer is 'correct'. Now, if this assumption is true: how often would we expect to get the same tree if we sequenced 1500 bases from a gene that behaves 'just like 16S'? If the probability of getting the same answer is low, then either we were very lucky to get a representative sample (and the 'correct' answer) or our assumption is wrong and we don't have a representative sample (nor the 'correct' answer).

When you see a bootstrap number like 70%, it means that if this is the 'correct' answer, we only had a 70% chance of finding it. It is both common and reasonable to turn this around and say that the answer has an approximately 70% chance of being 'correct'. You will often see confidence numbers like 20%, the trees are still the best, (although probably wrong) guess. There are not enough data to give an answer.

So, the bootstrap doesn't tell us if our method of calculating the tree is right, or if our model of evolution is right, or even if genes like 16S are representative of the 'true' evolutionary history. It only tells us if we have enough sequence data to apply our models.

MORE INTERNET RESOURCES OF INTEREST

1. Old Ribosomal Database project:
<http://rdpwww.life.uiuc.edu/index2.html>
2. Molecular Sequence Analysis Tools:
<http://www.wi.mit.edu/bio/biopage2.html> Links to major available sequence analysis software in the WWW. Submit your sequences,

- nucleic acid analysis, protein analysis, search for protein domains and conserved regions.
3. T-RFLP: <http://rdp.cse.msu.edu:8000/~saxmanpa/>
Release of the Beta version of the T-RFLP analysis program. A WEB browser is required for the following platforms: Windows95/98/NT, Solaris Sparc/Intel.
 4. Cold Spring Harbor Laboratory WWW home page: <http://www.cshl.org/>
Found under CSHL Press.
Chapter 7 of Cold Spring Harbor Genome Analysis Laboratory Manual provides helpful introductory information for users of molecular biology database and software. This chapter is available over the WWW at: http://www.cshl.org/books/g_a/bk1ch7/
 5. European Bioinformatics Institute (EBI) BioCatalog: <http://www.ebi.ac.uk/biocat>
List of software's for sequence analysis. Link to other homepages, where you can download some free software.
 6. Indiana University IUBio Archive: <http://iubio.bio.indiana.edu>
Archive of biology and software's.
 7. Pedro's BioMolecular Research Tools: http://www.public.iastate.edu/~pedro/research_tools.html
One of the best sites for biologists and molecular biologists. Everything you would like to know about biology and never had chance to ask.
 8. The tree of life: <http://phylogeny.arizona.edu/tree/phylogeny.html>
The Tree of Life is a project containing information about the diversity of organisms on Earth, their history, and characteristics.
 9. ExPASy Molecular Biology Server: <http://www.expasy.ch/>
This is the ExPASy molecular biology WWW server of the Swiss Institute of Bioinformatics (SIB). This server is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE. Links to other sequence analysis servers.
 10. Ribosomal Secondary Structure: <http://pundit.icmb.utexas.edu>
Robin Gutell's WEB site. Download secondary structures of different organisms. You have to know how to handle Postscript files.
 11. Algorithms, thermodynamics and databases for RNA secondary structure: <http://www.ibc.wustl.edu/~zucker/rna/>
Information about how RNA molecules fold. Just for the really curious mind.

LIST OF REFERENCES

1. **Amann, R., N. Springer, W. Ludwig, H. -D. Görtz, and K. H. Schleifer.** 1991. Identification *in situ* and phylogeny of uncultured bacterial endosymbionts. *Nature* **351**:161-164.
2. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143-169.

3. **Boivin-Jahns, V., A. Bianchi, R. Ruimy, J. Garcin, S. Daumas, and R. Christen.** 1995. Comparison of phenotypical and molecular methods for the identification of bacterial strains isolated from a deep subsurface environment. *Appl. Environ. Microbiol.* **61**:3400-3406.
4. **Borneman, J., P. W. Skroch, K. M. O'Sullivan, J. A. Paulus, N. G. Rumjanek, J. L. Jansen, J. Nienhus, and E. W. Triplett.** 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl. Environ. Ecol.* **62**:1935-1943.
5. **Ferris, M. J., A. L. Ruff-Roberts, E. D. Kopczynski, M. M. Bateson, and D. M. Ward.** 1996. Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat habitat. *Appl. Environ. Microbiol.* **62**:1045-1050.
6. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. *Science* **209**:457-463.
7. **Giovannoni, S., T. Britschgi, C. L. Moyer, and K. G. Field.** 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**:60-63.
8. **Gray, M. W., D. Sankoff, and R. J. Cedergren.** 1984. On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in a small subunit ribosomal RNA. *Nucleic Acids Res.* **12**:5837-5852.
9. **Liesack, W. and E. Stackebrandt.** 1992. Occurrence of a novel group of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* **174**:5072-5087.
10. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The RDP (Ribosomal Database Project). *Nucleic Acids Res.* **24**:82-85.
11. **Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl.** 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Ann. Rev. Microbiol.* **40**:337-365.
12. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. FastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* **10**:41-48.
13. **Rheims, H., C. Sproer, F. A. Rainey, and E. Stackebrandt.** 1996. Molecular biological evidence for the occurrence of uncultured members of the actinomycete line of descent in different environments and geographical locations. *Microbiology* **142**:2863-2870.
14. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. G. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Ehrlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487-491.
15. **Schmidt, T. M., B. Pace, and N. R. Pace.** 1991. Detection of DNA contamination in Taq polymerase. *BioTechniques* **11**:176-177.
16. **Sørheim, R., V. L. Torsvik, and J. Goksøyr.** 1989. Phenotypical divergences between populations of soil bacteria isolated on different media. *Microb. Ecol.* **17**:181-192.

17. **Torsvik, V., J. Goksøyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**:782-787.
18. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature, p. 219-286. *In* K. C. Marshall (ed.), *Advances in Microbial Ecology*, vol. 12. Plenum Press, New York, U.S.A.
19. **Weller, R., J. W. Weller, and D. M. Ward.** 1991. 16S rRNA sequences of uncultivated hot spring cyanobacterial mat inhabitants retrieved as randomly primed cDNA. *Appl. Environ. Microbiol.* **57**:1146-1151.
20. **Woese, C. R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221-271.
21. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87**:4576-4579.