Visualization of Probe Accessibility of ribosomal RNA using ARB Software

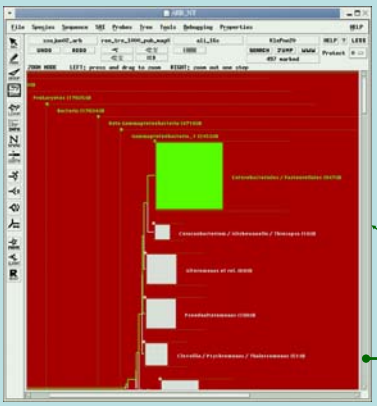
Yadhukumar, R. Westram, L. Richter, H. Meier†, A. Bode†, W. Ludwig

Lehrstuhl für Mikrobiologie, Am Hochanger 4, 85350 Freising, † Lehrstuhl für Rechnertechnik und Rechnerorganisation, Bolzmannstr. 3, 85747 Garching, Technische Universität München.

Designing, Evaluation and Visualization of Probes

Increasingly, classification of prokaryotes has involved molecular phylogenetic methods. This identification invariably relies on a genotypic approach, typically involving an analysis of 16s rRNA gene. More frequently, 16S rRNA information is used to identify oligonucleotide sequences unique to specific bacteria, for use as hybridization probes or as PCR primers. Such oligonucleotides can be specific to phylogenetic groupings as diverse as bacterial species or an entire Phylum. Furthermore, they are regularly used to enumerate specific groups of bacteria in natural environments by fluorescence *in situ* hybridization (FISH) technique and also in the gene expression studies of several genome projects. In this regard, a good probe design plays a crucial role in determining a potential probe that matches as many of its intended target group as possible, while ignoring the non-members of the target group.

ARB software package provides a workbench, where probes can be designed, evaluated and visualized in more intuitive way, using interactive graphical user interface (GUI) of ARB. Algorithms of the ARB programs 'Probe Design' and 'Probe Match' are searching the PT-server to identify short (10 - 100 monomers) diagnostic sequence stretches which are evaluated against the background of all full and partial sequences in the respective database the PT-server has been built from. In principle, no alignment of the sequence data is needed for specific probe design. However, in case of taxon specific probes, alignment and phylogenetic analyses are necessary to allow defining groups of phylogentically (taxonomically) related organisms as the targets of specific probes. The design of taxon specific oligonucleotide probes with ARB is performed in three steps. Firstly, the user selects the organism or a group of organisms for which he wants to design a diagnostic probe. Secondly, the software 'Probe Design' searches the PT-server for potential target sites. The results are shown in a ranked list of proposed targets, probes and additional information. The ranking is according to several compositional and thermodynamical criteria. Thirdly, the proposed oligonucleotide probes are evaluated against the whole database by using the program "Probe Match". Local alignments are determined between the probe target sequences and the most similar reference sequences (optionally from 0 to 5 mismatches) in the respective database. Furthermore, these sequence strings can automatically be visualised in the primary and secondary structure editors. Especially the latter information is of high importance when designing probes for in situ cell hybridisation. A special advancement is the ARB multi probe software component. It determines sets of up to five probes optimally identifying the target group. This probe set can be used for multiple fluorescence in situ hybridisation experiments.



ARB Tree Window

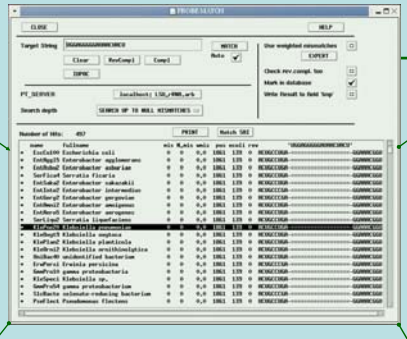
ARB Tree Window with the Phylogenetic tree constructed using ARB Parsimony Method from 16S ribosomal RNA of 25,000 species. Probes are designed by selecting the desired group / species in the tree and then, using Probe Design Tool, parameters like %GC content, melting temp, % group hits, % non-group hits are set, and the probes are designed against all the sequences in the ARB database.

As an example, probes were designed for the Enterobacteria group (colored in fluorescent green) in the phylogenetic tree (IRS Dendrogram tree view) of small subunit (16S) ribosomal RNA.

Probe Match Result Window

Probe Match Result Window displaying potential Probe Targets that were designed by Probe Design Tool. Both Probe Design and Probe Match Tools use PT SERVER, a special server built for faster designing and retrieval of probes from the ARB Databases.

In this window, results of the potential probe target (5` UGGAGGGGGAUAACUACU 3`), designed for Enterobacteria group, hits 497 members of the group containing 947 species. Additional information such as no. of mismatches, ecoli position, region on either side of the probe in the actual nucleotide sequence (around 10 nucleotide bases) are also shown.



Probe Visualization Window

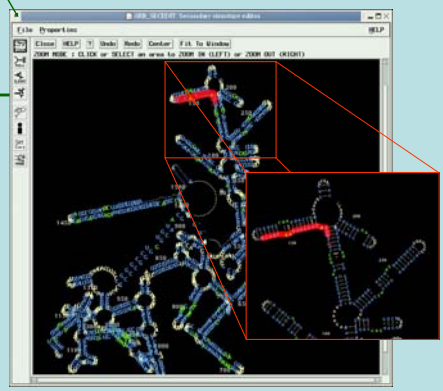
Probe Visualization window displaying the probe match results and respective target species in different colors. User can display any information that is associated with the sequences such as evolutionary models (secondary structure of ribosomal RNA, consensus models), column statistics (maximum column base frequency, Positional variability model) and any other user defined models, filters or statistics, in different colors. Additionally, user has an option to visualize the sequence associated information below each target species.

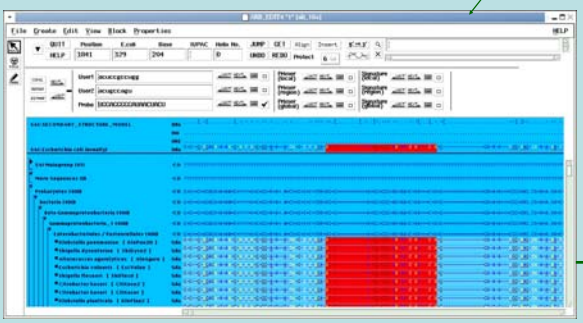
In this screenshot, well accepted secondary structure model of 16S rRNA is visualized in red color (helix region), where the starting and ending positions of helices are colored in black, and bases without background represents the loop regions of the structure.

Secondary Structure Editor Window

Secondary Structure Editor Window displaying the well accepted "16S rRNA secondary structure model" and the sequence, selected in primary structure editor, fitted to the structure model. Probe targets and any other sequence associated information can be visualized in different colors, as a background of the nucleotide bases in the structure, which is of more importance for fluorescence *in situ* hybridization (FISH) experiments.

In this screenshot, the potential probe target (5` UGGAGGGGGAUAACUACU 3`) region found in the structure of Enterobacteria group member, ***Klebsiella pneumoniae***, is colored in red background.





Primary Structure Editor Window displaying the sequences along with the alignment information with respect to *Escherichia coli* positions and well accepted secondary structure model of 16S rRNA. Any other sequence associated information can as well be displayed in the same window in different colors.

In this screenshot, all the members that are targeted by the potential probe (5` UGGAGGGGGAUAACUACU 3`), designed for Enterobacteria group, are displayed. The probe target region is colored in red. In addition, other sequence associated information can be displayed in various colors as a sequence background.

Primary Structure Editor Window

The ARB project

The ARB (derived from the Latin word, *arbor*, meaning tree) Project, was started as an interdisciplinary initiative of the Lehrstuhl für Mikrobiologie and the Lehrstuhl für Rechnertechnik und Rechnerorganisation, at the Technical University of Munich, more than a decade ago.

The ARB software package provides with an extensive and greedy software environment for the researchers in the area of Molecular Phylogeny, Molecular Ecology and Molecular Taxonomy. The ARB Project is based on the concept of integrated databases of raw and/or processed sequence and any other type of additional data assigned to the individual sequence entries. The work bench facilitates curation, alignment, visualization, and analysis of databases of any genes or genomes. This is provided by a selection of directly cooperating software tools which comprise primary and secondary structure editors, automated or user guided aligners, programs for establishing profiles and filters, different treeing approaches for homogenous or concatenated heterogeneous gene sets, taxon or gene specific probe/primer design and evaluation as well as other facilities.

Integrated ARB databases are continuously maintained for full genomes as well as for evolutionary conserved genes (e.g. ribosomal RNAs, elongation-initiation factors, RNA polymerases, recA, heat shock proteins, DNA gyrases/topoisomerases, aminoacyl tRNA synthetases, ATP sythetases). The ARB software as well as the ARB databases, are made available to the public via world wide web. Please use our homepage , www.arb-home.de , for further information regarding the ARB software.