

# probeBase—an online resource for rRNA-targeted oligonucleotide probes: new features 2007

Alexander Loy, Frank Maixner, Michael Wagner and Matthias Horn\*

Department für Mikrobielle Ökologie, Universität Wien, Althanstrasse 14, A-1090 Wien, Austria

Received September 13, 2006; Revised and Accepted October 10, 2006

## ABSTRACT

**probeBase is a curated database of annotated rRNA-targeted oligonucleotide probes and supporting information. Rapid access to probe, microarray and reference data is achieved by powerful search tools and via different lists that are based on selected categories such as functional or taxonomic properties of the target organism(s) or the hybridization format (fluorescence *in situ* hybridization or microarray) in which the probes were applied. Additional information on probe coverage and specificity is available through direct submissions of probe sequences from probeBase to RDP-II and GreenGenes, two major rRNA sequence databases. A freely editable user comments field for each probe entry allows any user to add, modify or remove information or to report errors in real-time. probeBase entries increased from 700 to more than 1200 during the past three years. Several options for submission of single probes or entire probe sets, even prior to publication of newly developed probes, should further contribute to keeping probeBase an up-to-date and useful resource. probeBase is freely accessible at <http://www.microbial-ecology.net/probebase>. Email correspondence can be addressed to [probebase@microbial-ecology.net](mailto:probebase@microbial-ecology.net).**

## INTRODUCTION

Numerous ribosomal RNA (rRNA)-targeted oligonucleotide probes have been developed and are widely employed in a variety of hybridization formats for identification (1–3), quantification (4) and physiological characterization (5) of cultivated and un-cultivated organisms in environmental, bioengineered and clinical settings. In order to release researchers from the necessity to tediously survey the literature for already published probes, probeBase, an internet database for rRNA-targeted oligonucleotides [NAR Molecular Biology Database Collection entry number 390 (6)], was established in 2002 to provide a freely accessible,

rapid platform to search for published probes and annotated/pre-calculated information and to facilitate an up-to-date evaluation of probe coverage and specificity (7). Suitable probes in probeBase can be found via the ‘Search’ tool by indicating the name of a probe, a target organism, a certain probe target site or by searching for references. The ‘Match’ service can be used to easily retrieve all probes that are fully complementary to up to 150 query rRNA gene sequences without prior aligning and phylogenetic analysis.

Today, probeBase contains 1258 probes and 8 microarrays from 266 references (status October 2006) and is frequently used by researchers worldwide. For example, probeBase user statistics (accessible on the start page) have recorded an average of 723 hits (corresponding to 554 opened pages) per day in the year 2005.

Here we describe novel features and search services that were implemented into probeBase since 2003, accompanied by significant layout and interface changes, which should collectively enhance data interconnectivity and usability of the database. Furthermore, we present a meta-analysis of all probes deposited at probeBase and have evaluated the *in silico* specificity of frequently used probes targeting major bacterial groups.

## NEW DATA AND ANALYSIS TOOLS

### Additional probe features

A distinctive accession number that begins with the prefix ‘pB’ is now assigned to each probe, which is essentially defined by its unique sequence. For example, the probes EUB338, EUB338-II and EUB338-III, together targeting most Bacteria (8), have the accession numbers pB-00159, pB-00160 and pB-00161, respectively. This minor but very important new feature of probeBase prevents potential confusion in cases where the same name was incidentally used for more than one probe, especially if a new probe is just a modification of an already existing one.

probeBase now also includes data on rRNA-targeted microarrays [so-called PhyloChips e.g. (9)]. Entries for probes that are used on microarrays contain links to more detailed information on the respective microarray, including a brief description, the technology applied, bibliographic data and

\*To whom correspondence should be addressed: Tel: +43 1 4277 54393; Fax: +43 1 4277 54389; Email: [horn@microbial-ecology.net](mailto:horn@microbial-ecology.net)

a list of all probes used on this particular microarray. probeBase will generally host and welcome submission of all microarrays that consist of rRNA-targeted oligonucleotide probes with hierarchically nested and identical specificity for phylogenetically defined groups of organisms. Microarrays with random probes will not be accommodated.

Theoretical thermodynamic probe parameters according to Yilmaz *et al.* (10) have been pre-calculated and added to each probe entry. The standard free energy changes  $\Delta G_1$  and  $\Delta G_2$  for probe/rRNA duplex formation and intramolecular probe folding (hairpin formation), respectively, and the thermodynamic affinity of a probe to a hypothetical perfectly accessible target site,  $\Delta G_{12}$ , should aid in the optimization of existing probes, which have so far not been tested for fluorescence *in situ* hybridization (FISH) or which show only weak fluorescence signals [See (10,11) for details].

Furthermore, each probe entry contains a so-called Wiki user comments field, which allows users to instantly add, modify or remove information on the probe and its application without the need for prior registration. The opportunity to easily annotate probes in real-time should promote up-to-dateness of probe information, reporting of bugs and errors in the database or the original publication, and interactions among probe users. This combination of community driven and protected content as realized in the current version of probeBase should generate an optimal resource, less susceptible to abuse than pure Wiki systems.

Each probe entry contains links to directly match the probe sequence against RDP-II and Greengenes, major encompassing, regularly updated 16S rRNA sequence online databases (12,13). probeBase users thereby benefit from the more efficient probe match tool that was recently implemented in RDP-II to directly match probe sequences against *Bacteria* and for the first time are able to match probe sequences against *Bacteria* and *Archaea* at Greengenes. These tools enable researchers to quickly check the current *in silico* specificity and coverage of a probe and thus to validate its originally intended specificity—a step that becomes increasingly important with ever-accumulating rRNA sequence data in public databases.

### Search probeBase

In order to extend the search options of probeBase, it is now also possible to search for a certain probe sequence, the accession number of a probe and references. Regarding the latter option, users can survey all references (including titles, authors and abstracts) that are available in the database via a simple key word search. probeBase then returns a list of references and all probes originally described therein, allowing a quick overview of all probes that were published by a certain author. Finally, all search options can now also be restricted to only those probes that were already applied for FISH or microarray hybridization.

### Lists and probe categories

A new option to view and retrieve probes, references and microarray layouts from probeBase is the 'Lists' service. The following dynamic lists, (pre)compiled according to

different criteria, are currently available (suggestions of new probe categories by users are welcome).

- (i) Lists of probes that target a certain functional or taxonomic category of organisms (e.g. sulphate-reducing prokaryotes or organisms of medical/hygienical relevance)
- (ii) A list of all probes
- (iii) A list of all references
- (iv) A list of all probes that were successfully used for FISH
- (v) A list of all rRNA-targeted oligonucleotide microarrays
- (vi) A pre-calculated, but regularly updated list showing the coverage of selected probes targeting major taxonomic groups of bacteria (exemplified by Table 1)

Because probes targeting major taxonomic groups of bacteria are those most frequently applied in ecological, biotechnological and clinical studies, we have re-evaluated their *in silico* coverage against the RDP-II database [release 9.41, (12)]. A few selected probes with a good coverage of >70% for the respective target taxon are exemplarily shown in Table 1. Knowing the exact target range of a given probe is vital for meaningful interpretation of the hybridization results.

### Distribution of probes along the 16S/18S rRNA sequence

Using the roughly 1200 probes currently available at probeBase we have analyzed the distribution of their respective binding sites on the 16S/18S rRNA sequence (Figure 1). The number of all probes and FISH probes covering a certain target position on the 16S/18S rRNA displayed a weak but significant positive correlation with the variability of the respective sequence position (the non-parametric Spearman-Rho correlation coefficients were 0.404,  $P < 0.001$  and 0.410,  $P < 0.001$ , respectively). Most probes in the database target highly variable sequence regions, e.g. around *Escherichia coli* position 450 and 650, while conserved regions, e.g. around position 350 and 1400 are typically occupied by only few probes that are specific for higher taxa, such as the bacterial probe EUB338 or the universal probe UNIV1390. This analysis may provide guidance for the development of novel probes and the selection of suitable target sites.

### SUBMISSION OF PROBES

Several options are provided on the probeBase website for submission of single probes and whole probe sets, including microarray (PhyloChip) probes. It is also possible to deposit newly developed probes in probeBase already prior to their publication in a scientific journal, thereby apprising probeBase users of the new probes and the upcoming publication. In this case, the information on the actual probe sequence might be hidden until the day of publication, and only the short name, the specificity, and the target molecule of the probe are displayed. Additionally, the contact details of the person who has deposited the probe are shown.

### AVAILABILITY, ERROR REPORTING AND CONTACT

probeBase is maintained by the Department of Microbial Ecology, University of Vienna, Wien, Austria. Data and

**Table 1.** Selected *in silico* coverages of published group-specific probes in probeBase, as analyzed with the RDP II probe match tool

| pB accession number | Probe name | Specificity <sup>b</sup>  | Reference | Probe evaluation with RDP II probe match <sup>a</sup> | Coverage <sup>c</sup> (%) | Hits in target taxon <sup>d</sup> | Total non-target hits <sup>e</sup> | Major non-target taxa <sup>f</sup>   |
|---------------------|------------|---|-----------|---|---------------------------|-----------------------------------|------------------------------------|--|
|                     |            |   |           | RDP II target taxon                                   |                           |                                   |                                    |  |
| pB-00327            | UNIV1390   | all organisms   | (16)      | domain <i>Bacteria</i>                                | 92                        | 75934                             | NA <sup>g</sup>                    | —  |
| pB-00159            | EUB338     | <i>Bacteria</i>   | (17)      | domain <i>Bacteria</i>                                | 90                        | 169389                            | NA <sup>g</sup>                    | —  |
| pB-00021            | ALF968     | <i>Alphaproteobacteria</i> ,<br>except of <i>Rickettsiales</i>        | (18)      | class <i>Alphaproteobacteria</i>                      | 79                        | 9750                              | 1876                               | class <i>Deinococci</i> (38/364),<br>phylum <i>Proteobacteria</i> (11138/51072),<br>class <i>Fibrobacteres</i> (48/53)   |
| pB-00432            | DELTA495a  | <i>Deltaproteobacteria</i>  | (19)      | class <i>Deltaproteobacteria</i>                      | 73                        | 4236                              | 729                                | phylum <i>Gemmatimonadetes</i> (343/391),<br>phylum <i>Genera_incertae_sedis_W53</i> (23/68)   |
| pB-00050            | CIV/V/1342 | heterocystous strains of<br>cyanobacteria (sections IV and V)         | (20)      | <i>Cyanobacteria</i> subsection 4                     | 96                        | 491                               | 36                                 | —  |
| pB-00044            | CFB1082    | class <i>Bacteroidetes</i>  | (21)      | class <i>Bacteroidetes</i>                            | 85                        | 8485                              | 3                                  | —  |
| pB-00046            | CFB563     | <i>Flavobacteriales</i>   | (21)      | order <i>Flavobacteriales</i>                         | 84                        | 3301                              | 302                                | —  |
| pB-00718            | GNSB-941   | <i>Chloroflexi</i> (green nonsulfur bacteria)                         | (22)      | phylum <i>Chloroflexi</i>                             | 94                        | 1240                              | 816                                | phylum <i>Thermomicrobia</i> (8/11),<br>phylum <i>Genera_incertae_sedis_Dehalococcoides</i> (81/89),<br>unclassified <i>Bacteria</i> (706/6563)  |
| pB-00645            | GSB-532    | <i>Chlorobiaceae</i> (green sulfur bacteria)                          | (23)      | family <i>Chlorobiaceae</i>                           | 91                        | 221                               | 0                                  | —  |
| pB-00283            | PLA46      | <i>Planctomycetales</i>   | (24)      | family <i>Planctomycetales</i>                        | 75                        | 599                               | 310                                | phylum <i>Leptisphaerae</i> (38/54),<br>phylum <i>Genera_incertae_sedis_BRC1</i> (10/11),<br>phylum <i>Genera_incertae_sedis_W53</i> (21/25),<br>unclassified <i>Bacteria</i> (232/6286) |
| pB-00049            | ChIs-0523  | <i>Chlamydiales</i>   | (25)      | order <i>Chlamydiales</i>                             | 93                        | 189                               | 21                                 | —  |
| pB-00183            | HoAc1402   | phylum <i>Acidobacteria</i>   | (26)      | phylum <i>Acidobacteria</i>                           | 95                        | 693                               | 718                                | unclassified <i>Bacteria</i> (602/3249)  |
| pB-00056            | CYA361     | phylum <i>Cyanobacteria</i>   | (20)      | phylum <i>Cyanobacteria</i>                           | 91                        | 4043                              | 438                                | phylum <i>Firmicutes</i> (143/49033),<br>unclassified <i>Bacteria</i> (254/10078)  |
| pB-00435            | NTSPA714   | phylum <i>Nitrospira</i> , not<br><i>Thermodesulfobrio islandicus</i> | (19)      | phylum <i>Nitrospira</i>                              | 71                        | 373                               | 4                                  | —  |
| pB-00535            | TDSBM1282  | genera <i>Thermodesulfobacterium</i><br>and <i>Geothermobacterium</i> | (19)      | phylum<br><i>Thermodesulfobacteria</i>                | 84                        | 36                                | 0                                  | —  |
| pB-00601            | TM7522     | candidate phylum TM7  | (27)      | phylum <i>Genera_incertae_sedis_TM7</i>               | 92                        | 264                               | 12                                 | —  |

Note: The RDP II probe match only includes bacterial but no archaeal or eukaryal small subunit rRNA sequences in the analysis.

<sup>a</sup>RDP II probe match was performed with database release 9.41 (July 28, 2006) containing 253 813 bacterial 16S rRNA sequences. The search for each probe was restricted to sequences of good quality with data in the respective probe binding region.

<sup>b</sup>The intended specificity of the probe.

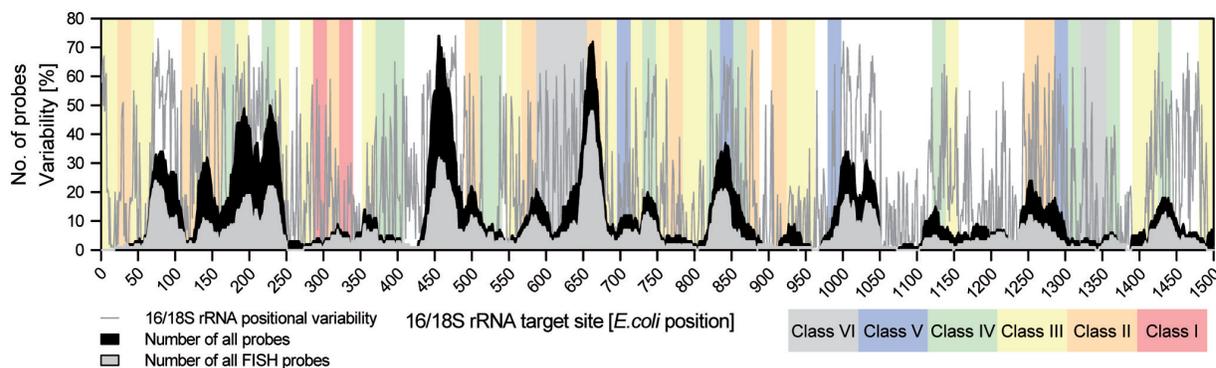
<sup>c</sup>The percentage of sequences within the RDP II target taxon that show a full match to the probe sequence.

<sup>d</sup>The number of sequences within the RDP II target taxon that show a full match to the probe sequence.

<sup>e</sup>The number of sequences outside the RDP II target taxon that show a full match to the probe sequence.

<sup>f</sup>If applicable, additional information on the major non-target taxa are provided. The number of sequences that show a full match to the probe versus the total number of sequences within a non-target taxon are shown in brackets.

<sup>g</sup>NA, not applicable because RDP II only contains bacterial 16S rRNA sequences.



**Figure 1.** Number of probes per target site and positional variability of the 16S/18S rRNA molecule (status July 2006). Positional variability profile of the 16S/18S rRNA in percentage was created using the ARB database “ssu\_jan04\_corr\_opt.arb” (14). Background colors indicate the different fluorescence brightness classes (I, bright to VI, dim) according to the consensus accessibility map of bacterial and archaeal 16S rRNA for FISH probes, established by Behrens *et al.* (15); uncolored regions were not analyzed.

analysis tools are available at <http://www.microbial-ecology.net/probebase/>. We welcome questions or comments concerning probeBase and highly appreciate reports of bugs, errors or missing probes. You may contact us by email to [probebase@microbial-ecology.net](mailto:probebase@microbial-ecology.net).

## ACKNOWLEDGEMENTS

The authors are indebted to Şafak Yilmaz and Daniel Noguera for calculation of the  $\Delta G$  values. We greatly thank Todd DeSantis for his efforts to facilitate the access to Greengenes and acknowledge help of Christian Baranyi for statistical analysis and of Carmen Krammer, Manuela Hartmann and Silvia Weber for database maintenance. We further thank Phil Hugenholz for valuable comments and all users who contributed to probeBase by submitting probes. probeBase was partly supported by grants of the bmb+f (01 LC 0021A-TP2 in the framework of the BIOLOG II program) to M.W. and the Austrian Science Fund to A.L. (P18836-B17) and M.H. (Y277-B03). Funding to pay the Open Access publication charges for this article was provided by the Austrian Science Fund.

*Conflict of interest statement.* None declared.

## REFERENCES

- Amann,R., Fuchs,B.M. and Behrens,S. (2001) The identification of microorganisms by fluorescence *in situ* hybridisation. *Curr. Opin. Biotechnol.*, **12**, 231–236.
- Loy,A. and Bodrossy,L. (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clin. Chim. Acta*, **363**, 106–119.
- Wagner,M., Horn,M. and Daims,H. (2003) Fluorescence *in situ* hybridisation for the identification and characterisation of prokaryotes. *Curr. Opin. Microbiol.*, **6**, 302–309.
- Daims,H., Lückner,S. and Wagner,M. (2006) daime, a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.*, **8**, 200–213.
- Wagner,M., Nielsen,P.H., Loy,A., Nielsen,J.L. and Daims,H. (2006) Linking microbial community structure with function: fluorescence *in situ* hybridization-microautoradiography and isotope arrays. *Curr. Opin. Biotechnol.*, **17**, 1–9.
- Galperin,M.Y. (2006) The Molecular Biology Database Collection: 2006 update. *Nucleic Acids Res.*, **34**, D3–D5.
- Loy,A., Horn,M. and Wagner,M. (2003) probeBase: an online resource for rRNA-targeted oligonucleotide probes. *Nucleic Acids Res.*, **31**, 514–516.
- Daims,H., Brühl,A., Amann,R., Schleifer,K.-H. and Wagner,M. (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.*, **22**, 434–444.
- Loy,A., Schulz,C., Lückner,S., Schöpfer-Wendels,A., Stoecker,K., Baranyi,C., Lehner,A. and Wagner,M. (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order “*Rhodocyclales*”. *Appl. Environ. Microbiol.*, **71**, 1373–1386.
- Yilmaz,L.S., Okten,H.E. and Noguera,D.R. (2006) Making all parts of the 16S rRNA of *Escherichia coli* accessible *in situ* to single DNA oligonucleotides. *Appl. Environ. Microbiol.*, **72**, 733–744.
- Yilmaz,L.S. and Noguera,D.R. (2004) Mechanistic approach to the problem of hybridization efficiency in fluorescent *in situ* hybridization. *Appl. Environ. Microbiol.*, **70**, 7126–7139.
- Cole,J.R., Chai,B., Farris,R.J., Wang,Q., Kulam,S.A., McGarrell,D.M., Garrity,G.M. and Tiedje,J.M. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.*, **33**, 294–296.
- DeSantis,T.Z., Hugenholz,P., Larsen,N., Rojas,M., Brodie,E.L., Keller,K., Huber,T., Dalevi,D., Hu,P. and Andersen,G.L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, **72**, 5069–5072.
- Ludwig,W., Strunk,O., Westram,R., Richter,L., Meier,H., Yadhukumar, Buchner,A., Lai,T., Steppi,S., Jobb,G. *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.*, **32**, 1363–1371.
- Behrens,S., Ruhland,C., Inacio,J., Huber,H., Fonseca,A., Spencer-Martins,I., Fuchs,B.M. and Amann,R. (2003) *In situ* accessibility of small-subunit rRNA of members of the domains Bacteria, Archaea, and Eucarya to Cy3-labeled oligonucleotide probes. *Appl. Environ. Microbiol.*, **69**, 1748–1758.
- Zheng,D., Alm,E.W., Stahl,D.A. and Raskin,L. (1996) Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Appl. Environ. Microbiol.*, **62**, 4504–4513.
- Amann,R.I., Binder,B.J., Olson,R.J., Chisholm,S.W., Devereux,R. and Stahl,D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.*, **56**, 1919–1925.
- Neef,A. (1997) Ph. D. thesis. Anwendung der *in situ* Einzelzell-Identifizierung von Bakterien zur Populationsanalyse in komplexen mikrobiellen Biozönosen. Technische Universität München, München.
- Loy,A., Lehner,A., Lee,N., Adamczyk,J., Meier,H., Ernst,J., Schleifer,K.-H. and Wagner,M. (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol.*, **68**, 5064–5081.

20. Schönhuber, W., Zarda, B., Eix, S., Rippka, R., Herdman, M., Ludwig, W. and Amann, R. (1999) *In situ* identification of cyanobacteria with horseradish peroxidase-labeled, rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.*, **65**, 1259–1267.
21. Weller, R., Glöckner, F.O. and Amann, R. (2000) 16S rRNA-targeted oligonucleotide probes for the *in situ* detection of members of the phylum *Cytophaga-Flavobacterium-Bacteroides*. *Syst. Appl. Microbiol.*, **23**, 107–114.
22. Gich, F., Garcia-Gil, J. and Overmann, J. (2001) Previously unknown and phylogenetically diverse members of the green nonsulfur bacteria are indigenous to freshwater lakes. *Arch. Microbiol.*, **177**, 1–10.
23. Tuschak, C., Glaeser, J. and Overmann, J. (1999) Specific detection of green sulfur bacteria by *in situ* hybridization with a fluorescently labeled oligonucleotide probe. *Arch. Microbiol.*, **171**, 265–272.
24. Neef, A., Amann, R., Schlesner, H. and Schleifer, K.-H. (1998) Monitoring a widespread bacterial group: *in situ* detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology*, **144**, 3257–3266.
25. Poppert, S., Essig, A., Marre, R., Wagner, M. and Horn, M. (2002) Detection and differentiation of chlamydiae by fluorescence *in situ* hybridization. *Appl. Environ. Microbiol.*, **68**, 4081–4089.
26. Juretschko, S., Loy, A., Lehner, A. and Wagner, M. (2002) The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. *Syst. Appl. Microbiol.*, **25**, 84–99.
27. Hugenholtz, P., Tyson, G.W., Webb, R.I., Wagner, A.M. and Blackall, L.L. (2001) Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. *Appl. Environ. Microbiol.*, **67**, 411–419.