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 Green-genes, 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.

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

© 2006 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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 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>
<thead>
<tr>
<th>pB accession number</th>
<th>Probe name</th>
<th>Specificity</th>
<th>Reference</th>
<th>Probe evaluation with RDP II probe match</th>
<th>RDP II target taxon</th>
<th>Coverage (%)</th>
<th>Hits in target taxon</th>
<th>Total non-target hits</th>
<th>Major non-target taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td>pB-00327</td>
<td>UNIV1390</td>
<td>all organisms</td>
<td>(16)</td>
<td>domain Bacteria</td>
<td>92</td>
<td>75934</td>
<td>NA</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>pB-00159</td>
<td>EUB338</td>
<td>Bacteria</td>
<td>(17)</td>
<td>domain Bacteria</td>
<td>90</td>
<td>169389</td>
<td>NA</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>pB-00021</td>
<td>ALF968</td>
<td>Alphaproteobacteria, except of Rickettsiales</td>
<td>(18)</td>
<td>class Alphaproteobacteria</td>
<td>79</td>
<td>9750</td>
<td>1876</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00432</td>
<td>DELTA495a</td>
<td>Deltaproteobacteria</td>
<td>(19)</td>
<td>class Deltaproteobacteria</td>
<td>73</td>
<td>4236</td>
<td>729</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00050</td>
<td>CIV/V1342</td>
<td>heterocystous strains of cyanobacteria (sections IV and V)</td>
<td>(20)</td>
<td>Cyanobacteria subsection 4</td>
<td>96</td>
<td>491</td>
<td>36</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00044</td>
<td>CFB1082</td>
<td>class Bacteroidetes</td>
<td>(21)</td>
<td>class Bacteroidetes</td>
<td>85</td>
<td>8485</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00046</td>
<td>CFB563</td>
<td>Flavobacteriales</td>
<td>(21)</td>
<td>order Flavobacteriales</td>
<td>84</td>
<td>3301</td>
<td>302</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00718</td>
<td>GNSB-941</td>
<td>Chloroflexi (green nonsulfur bacteria)</td>
<td>(22)</td>
<td>phylum Chloroflexi</td>
<td>94</td>
<td>1240</td>
<td>816</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00645</td>
<td>GSB-532</td>
<td>Chlorobiaceae (green sulfur bacteria)</td>
<td>(23)</td>
<td>family Chlorobiaceae</td>
<td>91</td>
<td>221</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00283</td>
<td>PL-A46</td>
<td>Planctomycetales</td>
<td>(24)</td>
<td>family Planctomycetales</td>
<td>75</td>
<td>599</td>
<td>310</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00049</td>
<td>Chls-0523</td>
<td>Chloridiales</td>
<td>(25)</td>
<td>order Chloridiales</td>
<td>93</td>
<td>189</td>
<td>21</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00183</td>
<td>HoAc1402</td>
<td>Acidobacteria</td>
<td>(26)</td>
<td>phylum Acidobacteria</td>
<td>95</td>
<td>693</td>
<td>718</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00056</td>
<td>CYA361</td>
<td>Cyanobacteria</td>
<td>(20)</td>
<td>phylum Cyanobacteria</td>
<td>91</td>
<td>4043</td>
<td>438</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00435</td>
<td>NTSPA714</td>
<td>phylum Nitrospira, not Thermodesulfobacterium and Geothermobacterium</td>
<td>(19)</td>
<td>phylum Nitrospira</td>
<td>71</td>
<td>373</td>
<td>4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00535</td>
<td>TDSBM1282</td>
<td>genera Thermodesulfobacterium and Geothermobacterium</td>
<td>(19)</td>
<td>phylum Thermodesulfobacterium</td>
<td>84</td>
<td>36</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>pB-00601</td>
<td>TM7522</td>
<td>candidate phylum TM7</td>
<td>(27)</td>
<td>phylum Genera incertae_sedis_TM7</td>
<td>92</td>
<td>264</td>
<td>12</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

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

aRDP 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.

The intended specificity of the probe.

cThe percentage of sequences within the RDP II target taxon that show a full match to the probe sequence.

dThe number of sequences within the RDP II target taxon that show a full match to the probe sequence.

eThe number of sequences outside the RDP II target taxon that show a full match to the probe sequence.

fIf applicable, additional information on the major non-target taxon is 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.

gNA, not applicable because RDP II only contains bacterial 16S rRNA sequences.
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.

ACKNOWLEDGEMENTS
The authors are indebted to Şafak Yilmaz and Daniel Noguera for calculation of the ΔG 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 Krummer, Manuela Hartmann and Silvia Weber for database maintenance. We further thank Phil Hugenholtz 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

Figure 1. Number of probes per target site and positional variability of the 16/18S rRNA molecule (status July 2006). Positional variability profile of the 16S/18S rRNA molecule (status July 2006). Positional variability profile of the 16S/18S rRNA molecule (status July 2006).


