Environmental Microbiology (2008) **10**(10), 2894–2898 doi:10.1111/j.1462-2920.2008.01706.x

Brief report

probeCheck – a central resource for evaluating oligonucleotide probe coverage and specificity

Alexander Loy,[1] Roland Arnold,[2] Patrick Tischler,[2] Thomas Rattei,[2] Michael Wagner[1] and Matthias Horn[1]*

[1]*Department of Microbial Ecology, Universität Wien, Althanstraße 14, A-1090 Wien, Austria.*
[2]*Department of Genome Oriented Bioinformatics, Technische Universität München, Am Forum 1, D-85354 Freising, Germany.*

Summary

The web server probeCheck, freely accessible at http://www.microbial-ecology.net/probecheck, provides a pivotal forum for rapid specificity and coverage evaluations of probes and primers against selected databases of phylogenetic and functional marker genes. Currently, 24 widely used sequence collections including the Ribosomal Database Project (RDP) II, Greengenes, SILVA and the Functional Gene Pipeline/Repository can be queried. For this purpose, probeCheck integrates a new online version of the popular ARB probe match tool with free energy (ΔG) calculations for each perfectly matched and mismatched probe-target hybrid, allowing assessment of the theoretical binding stabilities of oligo-target and non-target hybrids. For each output sequence, the accession number, the GenBank taxonomy and a link to the respective entry at GenBank, EMBL and, if applicable, the query database are displayed. Filtering options allow customizing results on the output page. In addition, probeCheck is linked with probe match tools of RDP II and Greengenes, NCBI BLAST, the Oligonucleotide Properties Calculator, the two-state folding tool of the DINAMelt server and the rRNA-targeted probe database probeBase. Taken together, these features provide a multifunctional platform with maximal flexibility for the user in the choice of databases and options for the evaluation of published and newly developed probes and primers.

Received 5 May, 2008; accepted 5 June, 2008. *For correspondence. E-mail horn@microbial-ecology.net; Tel. (+43) 1427 754 393; Fax (+43) 1427 754 389.

Introduction

Diagnostic hybridization and PCR assays employing oligonucleotides as probes/primers (subsequently referred to as diagnostic oligos) are routinely applied for identifying microbes of interest and for studying the composition of polymicrobial communities in clinical, biotechnological and environmental specimens. The accuracy and performance of these molecular assays are intimately connected with the physicochemical characteristics of the diagnostic oligos and their discriminatory capacity against non-target sequences. Various tools are available that assist researchers in the *in silico* development of diagnostic oligos for (groups of) genes and microorganisms of interest (e.g. Ashelford *et al*., 2002; Kumar *et al*., 2006). Such oligos ideally have a high coverage for the target group (i.e. a high percentage of sequences in the target group possess a perfectly matching probe binding site) and a high specificity (i.e. no or only a low number of perfectly matching sequences not belonging to the target group are known and all other non-target sequences have many strongly discriminating mismatches). Furthermore, theoretical thermodynamic criteria, such as Gibbs-free energy; for efficient formation of the oligo-target hybrid and for avoiding non-specific binding to non-target genes should also be carefully considered during the design of a new diagnostic oligo (Yilmaz and Noguera, 2004; 2007). However, even the best *in silico* evaluation can currently only provide an estimate of the actual binding and discriminatory performance of a probe and thus empirical evaluation of the optimal experimental conditions by using suitable target and non-target reference sequences/organisms remains the final requirement during the development of effective new probes (Daims *et al*., 2005; Loy and Bodrossy, 2006).

Most experiments do not involve the *de novo* design and empirical testing of new diagnostic oligos, but employ already-published probes and primers. The frequent interest in such diagnostic oligos is, for example, mirrored in the user statistics of the rRNA-targeted oligonucleotide probe database probeBase (Loy *et al*., 2007), which show 265 705 page views in the year 2007. A considerable problem of naïve application of already-published probes is, however, that their originally intended coverage and specificity might no longer hold true, taken in consideration the rapid accumulation of sequences in public repositories. Periodic evaluation of a diagnostic oligo is thus of utmost importance for its reliable application in molecular assays (Lücker *et al*., 2007).

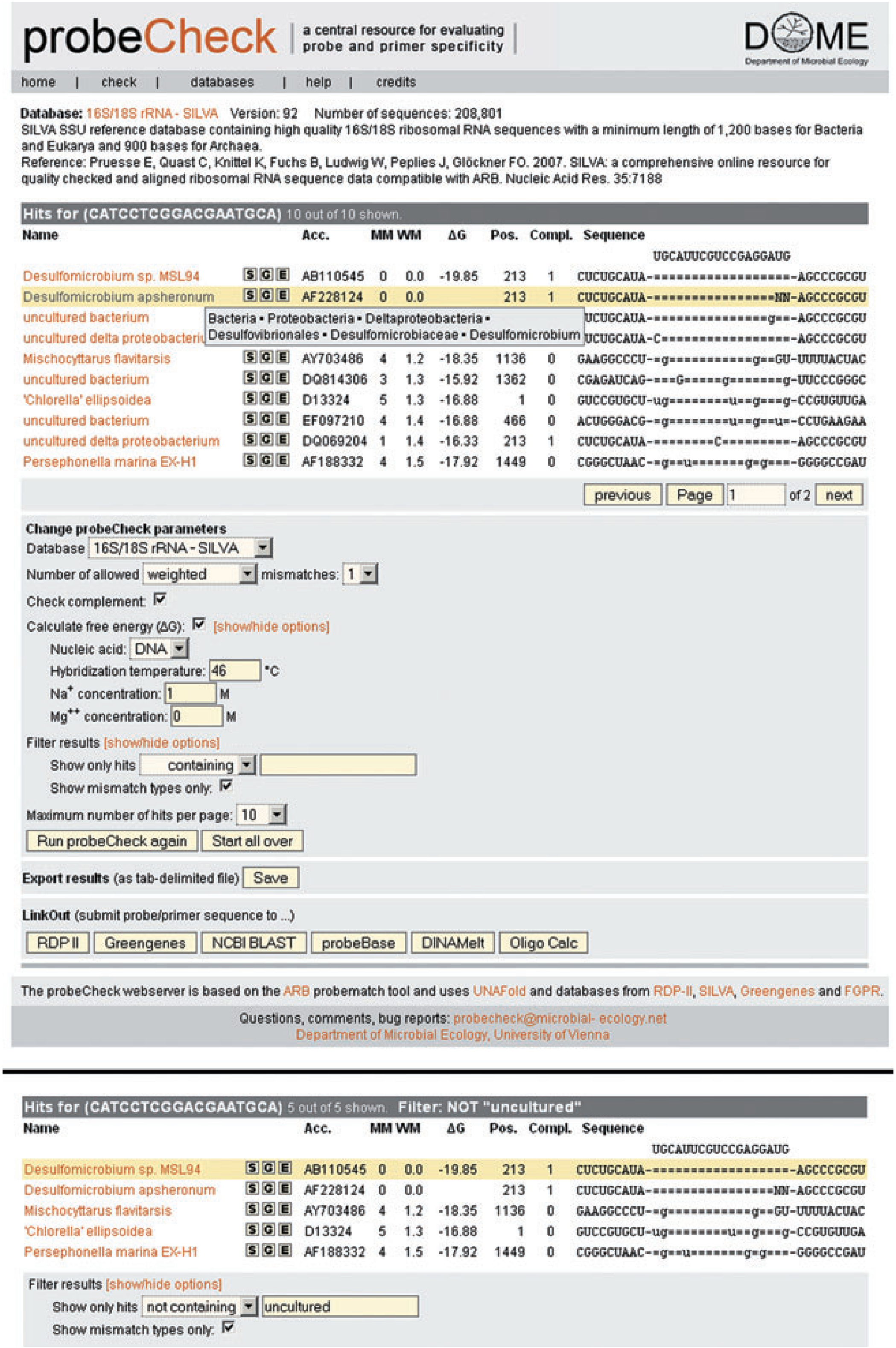
probeCheck provides a freely accessible, central platform for rapid *in silico* specificity and coverage evaluations of diagnostic oligos against the latest sequence collections of selected phylogenetic and functional marker genes. By integrating various existing (online-) tools and databases, probeCheck offers a number of unique features (e.g. an online version of the ARB probe match tool (Ludwig *et al*., 2004), its combination with ΔG calculations and new data filtering options, and the possibility to query the currently largest rRNA sequence database SILVA (Pruesse *et al*., 2007) and should thus be a useful web resource for all microbiologists interested in the detection of genes or microorganisms with oligonucleotide-based assays.

Features of probeCheck

Using a common user interface up to 10 oligonucleotide sequences (8–100 mer, in FASTA format and IUPAC coding, which is resolved automatically) can be queried against a number (currently 24) of sequence databases retrieved from public repositories such as the Ribosomal Database Project (RDP) II (Cole *et al*., 2007), Greengenes (DeSantis *et al*., 2006), SILVA (Pruesse *et al*., 2007) and the Functional Gene Pipeline/Repository (http://fungene.cme.msu.edu/). The probeCheck server employs the established ARB probe match tool, which creates difference alignments of the tested oligonucleotide and complementary sequences as output. The user can adjust the following search parameters. The *check complement* option causes probeCheck to not only search for sequences that are identical to the query sequence (i.e. have the same orientation), but also to target sequences that are in reverse complementary orientation; a feature required for, e.g. checking rRNA-targeted probes used for fluorescence *in situ* hybridization of microorganisms. The *number of allowed weighted* or *unweighted mismatches* can range between 0 and 4. The weighted mismatch value, calculated by the ARB probe match method using default settings, considers the relative strength of base pairings and the position of the mismatch to estimate the stability of the probe-target hybrid. This estimation is best suited for fluorescence-labelled probes applied for whole-cell hybridization (Strunk, 2001), but might also be useful for other hybridization formats such as DNA microarrays (Sanguin *et al*., 2006). In addition, the *free energy (ΔG)* can be determined for each perfectly matched and mismatched oligo-target hybrid by using the two-state hybridization algorithm of the UNAfold software (Markham and Zuker, 2005), allowing rough assessment of the differential theoretical melting properties of oligo-(non-)target hybrids (see Yilmaz and Noguera, 2007 for further information; Loy *et al*., 2005). probeCheck further offers new possibilities for filtering of the results on the output page. A keyword can be entered next to the option *show only hits (not) containing* in order to filter the list of hits for the presence/absence of this keyword in the sequence/species name (see Fig. 1 for an example). Multiple keywords can be entered, separated by 'OR'. The option *show mismatch types only* restricts the output to only one example sequence per each perfectly matching target and mismatching non-target type, thus presenting a quick overview of the different types of mismatches to the query sequence and facilitating the selection of appropriate non-target references for empirical oligo performance tests.

In the output table (Fig. 1), difference alignments showing the probe binding site and its flanking regions in the target sequence (5′→3′ orientation) are ordered according to the number, position and type of mismatching bases. A short description, the accession number and a link to the respective entry at GenBank (Benson *et al*., 2007), the European Molecular Biology Laboratory (EMBL) database and, if applicable, the query database are given for each output sequence. For rRNA-targeted oligos, the position of the 5′-terminal nucleotide in the oligo binding site relative to the rRNA sequence of *Escherichia coli* is indicated (Brosius *et al*., 1981). The unified NCBI/EMBL taxonomy is displayed on mouse-over the name of each target sequence. The output table can be exported and saved as tab-delimited text file for further processing. In addition, on the results page (Fig. 1) the query sequence can be directly submitted to a number of other web servers for further analysis, including the probe match tools of RDP II and Greengenes, BLAST (search for short nearly exact matches) at NCBI (Benson *et al*., 2007), the Oligonucleotide Properties Calculator OligoCalc (Kibbe, 2007), the two-state folding tool of the DINAMelt server (enabling evaluation of the oligo's thermodynamic tendency to form self-structures, i.e. hairpins) (Markham and Zuker, 2005) and the rRNA-targeted probe database probeBase (Loy *et al*., 2007).

probeCheck also includes a help page with a detailed description of the input and output features. A separate page contains an overview over the sequence databases available in probeCheck, including information on the release version, a web link to each database homepage, and references.


probeCheck
a central resource for evaluating probe and primer specificity
DOME
Department of Microbial Ecology
home | check | databases | help | credits

Database: 16S/18S rRNA - SILVA Version: 92 Number of sequences: 208,801
SILVA SSU reference database containing high quality 16S/18S ribosomal RNA sequences with a minimum length of 1,200 bases for Bacteria and Eukarya and 900 bases for Archaea.
Reference: Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acid Res. 35:7188

Hits for (CATCCTCGGACGAATGCA) 10 out of 10 shown.

Name		Acc.	MM	WM	ΔG	Pos.	Compl.	Sequence
								UGCAUUCGUCCGAGGAUG
Desulfomicrobium sp. MSL94	S G E	AB110545	0	0.0	-19.85	213	1	CUCUGCAUA-==================-AGCCCGCGU
Desulfomicrobium apsheronum	S G E	AF228124	0	0.0		213	1	CUCUGCAUA-================NN-AGCCCGCGU
uncultured bacterium								UCUGCAUA-===============g==-AGCCCGCGU
uncultured delta proteobacteri								UCUGCAUA-C=================-AGCCCGCGU
Mischocyttarus flavitarsis	S G E	AY703486	4	1.2	-18.35	1136	0	GAAGGCCCU-=g===========g==GU-UUUUACUAC
uncultured bacterium	S G E	DQ814306	3	1.3	-15.92	1362	0	CGAGAUCAG-===G=====g=======g-UUCCCGGGC
'Chlorella' ellipsoidea	S G E	D13324	5	1.3	-16.88	1	0	GUCCGUGCU-ug========u==g===g-CCGUGUUGA
uncultured bacterium	S G E	EF097210	4	1.4	-16.88	466	0	ACUGGGACG-=g========u==g==u=-CCUGAAGAA
uncultured delta proteobacterium	S G E	DQ069204	1	1.4	-16.33	213	1	CUCUGCAUA-========C=========-AGCCCGCGU
Persephonella marina EX-H1	S G E	AF188332	4	1.5	-17.92	1449	0	CGGGCUAAC-=g==u=======g=g===-GGGGCCGAU

Bacteria • Proteobacteria • Deltaproteobacteria •
Desulfovibrionales • Desulfomicrobiaceae • Desulfomicrobium

previous Page 1 of 2 next

Change probeCheck parameters
Database 16S/18S rRNA - SILVA
Number of allowed weighted mismatches: 1
Check complement: ☑
Calculate free energy (ΔG): ☑ [show/hide options]
Nucleic acid: DNA
Hybridization temperature: 46 °C
Na$^+$ concentration: 1 M
Mg^{++} concentration: 0 M
Filter results [show/hide options]
Show only hits containing
Show mismatch types only: ☑
Maximum number of hits per page: 10

Export results (as tab-delimited file) Save

LinkOut (submit probe/primer sequence to ...)

The probeCheck webserver is based on the ARB probematch tool and uses UNAFold and databases from RDP-II, SILVA, Greengenes and FGPR.

Questions, comments, bug reports: probecheck@microbial- ecology.net
Department of Microbial Ecology, University of Vienna

Hits for (CATCCTCGGACGAATGCA) 5 out of 5 shown. **Filter: NOT "uncultured"**

Name		Acc.	MM	WM	ΔG	Pos.	Compl.	Sequence
								UGCAUUCGUCCGAGGAUG
Desulfomicrobium sp. MSL94	S G E	AB110545	0	0.0	-19.85	213	1	CUCUGCAUA-==================-AGCCCGCGU
Desulfomicrobium apsheronum	S G E	AF228124	0	0.0		213	1	CUCUGCAUA-================NN-AGCCCGCGU
Mischocyttarus flavitarsis	S G E	AY703486	4	1.2	-18.35	1136	0	GAAGGCCCU-=g===========g==GU-UUUUACUAC
'Chlorella' ellipsoidea	S G E	D13324	5	1.3	-16.88	1	0	GUCCGUGCU-ug========u==g===g-CCGUGUUGA
Persephonella marina EX-H1	S G E	AF188332	4	1.5	-17.92	1449	0	CGGGCUAAC-=g==u=======g=g===-GGGGCCGAU

Filter results [show/hide options]
Show only hits not containing uncultured
Show mismatch types only: ☑

Fig. 1. probeCheck screen shot. (Upper panel) An example of the probeCheck output using the *Desulfomicrobium*-specific 16S rRNA-targeted probe DSM213 (probeBase Accession No. pB-00507) (Lücker *et al.*, 2007) as query sequence and the 16S/18S rRNA sequence database from SILVA. (Lower panel) Same as above, except that the results were filtered to exclude environmental sequences by excluding hits containing the term ‘uncultured’ in the species/sequence name.

Database updates and call for submission of ARB sequence databases

probeCheck is maintained by the Department of Microbial Ecology at the University of Vienna. Databases behind probeCheck are retrieved on a regular basis from public database projects such as RDP II, greengenes, SILVA and FUNGENE, and, if required, are adapted to the ARB database format. The dates of last and upcoming updates are indicated.

Curators of own nucleic acid sequence ARB databases are strongly encouraged to make their databases available for probe/primer evaluations on the probeCheck server. Note that probeCheck only enables matches against the database. The actual database remains hidden in the background and is not available for download. The probeCheck staff can be contacted by email (probecheck@microbial-ecology.net) for questions and bug reports.

Technology behind probeCheck

The probeCheck website is hosted on a Linux server (openSuSE 10.2) with 4GB RAM and two Intel Xeon processors (2.4 GHz). Perl (including BioPerl) scripts are used to parse user input, ARB probematch and UNAfold output, and to create web pages on the fly.

Acknowledgements

We would like to thank Harald Meier, Safak Yilmaz and Holger Daims for their help in the initial phase of the project. We would like to express our sincere gratitude to Wolfgang Ludwig and the entire ARB team for developing and maintaining the ARB program package. probeCheck is dependent on the quality and up-to-dateness of the underlying sequence databases of phylogenetic and functional marker genes. We thus greatly acknowledge the efforts and endurance of the individual researchers and larger consortia in maintaining and curating the various database projects. A. Loy is supported by the Austrian Science Fund (FWF) grants P18836-B17 and P20185-B17 and by the bmb+f grant 01LC0621D. M. Horn is supported by FWF grant Y277-B03. M. Wagner is supported by a grant from the University of Vienna in the framework of the University Research Focus 'Symbiosis research and molecular principles of recognition'.

References

Ashelford, K.E., Weightman, A.J., and Fry, J.C. (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res* **30:** 3481–3489.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2007) GenBank. *Nucleic Acids Res* **35:** D21–D25.

Brosius, J., Dull, T.L., Sleeter, D.D., and Noller, H.F. (1981) Gene organization and primary structure of a ribosomal operon from *Escherichia coli*. *J Mol Biol* **148:** 107–127.

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., *et al.* (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35:** D169–D172.

Daims, H., Stoecker, K., and Wagner, M. (2005) Fluorescence *in situ* hybridization for the detection of prokaryotes. In *Advanced Methods in Molecular Microbial Ecology*. Osborn, A.M., and Smith, C.J. (eds). Abingdon, UK: BIOS Scientific Publishers, pp. 213–239.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35:** W43–W46.

Kumar, Y., Westram, R., Kipfer, P., Meier, H., and Ludwig, W. (2006) Evaluation of sequence alignments and oligonucleotide probes with respect to three-dimensional structure of ribosomal RNA using ARB software package. *BMC Bioinformatics* **7:** 240.

Loy, A., and Bodrossy, L. (2006) Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clin Chim Acta* **363:** 106–119.

Loy, A., Schulz, C., Lücker, S., Schöpfer-Wendels, A., Stoecker, K., Baranyi, C., *et al.* (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order '*Rhodocyclales*'. *Appl Environ Microbiol* **71:** 1373–1386.

Loy, A., Maixner, F., Wagner, M., and Horn, M. (2007) probeBase – an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Res* **35:** D800–D804.

Lücker, S., Steger, D., Kjeldsen, K.U., MacGregor, B.J., Wagner, M., and Loy, A. (2007) Improved 16S rRNA-targeted probe set for analysis of sulfate-reducing bacteria by fluorescence *in situ* hybridization. *J Microbiol Methods* **69:** 523–528.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.

Markham, N.R., and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33:** W577–W581.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Sanguin, H., Herrera, A., Oger-Desfeux, C., Dechesne, A., Simonet, P., Navarro, E., *et al.* (2006) Development and validation of a prototype 16S rRNA-based taxonomic microarray for *Alphaproteobacteria*. *Environ Microbiol* **8:** 289–307.

Strunk, O. (2001) ARB: Entwicklung eines Programmsystems zur Erfassung, Verwaltung und Auswertung von Nuklein- und Aminosäuresequenzen. PhD Thesis. München, Germany: Department of Microbiology, Technische Universität München.

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.

Yilmaz, L.S., and Noguera, D.R. (2007) Development of thermodynamic models for simulating probe dissociation profiles in fluorescence *in situ* hybridization. *Biotechnol Bioeng* **96:** 349–363.