

pr2-primers: an 18S rRNA primer database for protists

Daniel Vaulot ^{1,2,*}, Stefan Geisen ^{3,4,5}, Frédéric Mahé ^{6,7}, David Bass ^{8,9}

¹ UMR 7144, ECOMAP, CNRS, Sorbonne Université, CNRS, Station Biologique de Roscoff, 29680 Roscoff, France

² Asian School of the Environment, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

³ Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), PO Box 50, 6700 AB Wageningen, The Netherlands

⁴ Laboratory of Nematology, Wageningen University, Wageningen, The Netherlands

⁵ Nanjing Agricultural University, Nanjing, 210095, P.R. China

⁶ CIRAD, UMR BGPI, F-34398, Montpellier, France

⁷ BGPI, Univ Montpellier, CIRAD, IRD, Montpellier SupAgro, Montpellier, France

⁸ Centre for Environment, Fisheries and Aquaculture Science, Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK

⁹ Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

ORCID

- Daniel Vaulot: 0000-0002-0717-5685
- Stefan Geisen: 0000-0003-0734-727X
- Frédéric Mahé: 0000-0002-2808-0984
- David Bass: 0000-0002-9883-7823

* Corresponding author: vaulot@gmail.com

Date: December 24, 2020

Keywords: 18S rRNA, primers, database, R, shiny, PCR, protists

1 Abstract

2 Metabarcoding of microbial eukaryotes (collectively known as *protists*) has developed tremendously in
3 the last decade, almost uniquely relying on the 18S rRNA gene. As microbial eukaryotes are extremely
4 diverse, many primers and primer pairs have been developed. To cover a relevant and representative
5 fraction of the protist community in a given study system, a wise primer choice is needed as no primer
6 pair can target all protists equally well. As such, a smart primer choice is very difficult even for
7 experts and there are very few on-line resources available to list existing primers. We built a database
8 listing 179 primers and 76 primer pairs that have been used for eukaryotic 18S rRNA metabarcoding.
9 *In silico* performance of primer pairs was tested against two sequence databases: PR² for eukaryotes
10 and a subset of Silva for prokaryotes. This allowed to determine the taxonomic specificity of primer
11 pairs, the location of mismatches as well as amplicon size. We developed a R-based web application
12 that allows to browse the database, visualize the taxonomic distribution of the amplified sequences
13 with the number of mismatches, and to test any user-defined primer set (<https://app.pr2-primers.org>).
14 This tool will provide the basis for guided primer choices that will help a wide range of ecologists to
15 implement protists as part of their investigations.

16 Introduction

17 Microbes are key players in all Earth ecosystems. Among them are protists that encompass all uni-
18 cellular or unicellular-colonial eukaryotes excluding some fungi. Protists perform a range of functions
19 from photosynthesis to organic matter degradation. Although some eukaryotic groups such as unicel-
20 lular algae (phytoplankton) have a long tradition of being studied as key players in marine primary
21 production, the importance of protists in other processes and other environments has only been re-
22 cently recognized, for example their role in nutrient cycling in soils or the complexity of symbiotic and
23 parasitic relationships in marine waters (Geisen et al. 2018a; Worden et al. 2015). This absence of
24 recognition stems in part from the inherent difficulties to identify them because of the lack of morpho-
25 logical features as well as of the difficulty to grow them in culture. In recent years the development of
26 metabarcoding has provided new tools to study protist roles.

27 Metabarcoding is defined as the use of a specific marker gene to analyse the composition of natural
28 communities in a specific environment (water, soil, animal gut, faeces, etc...). After DNA extrac-
29 tion, the gene is amplified using a pair of primers targeting one specific region, samples are labelled
30 with molecular tags and the resulting DNA is sequenced using a high throughput technology, mostly
31 Illumina currently. This approach was initially developed for prokaryotes (Sogin et al. 2006) and
32 expanded steadily in recent years for protists. For prokaryotes, the gene most commonly used is
33 the gene coding for the small sub-unit ribosomal RNA (SSU rRNA or 16S). SSU rRNA genes are
34 composed of conserved and variable regions. Conserved regions can be used to design primers while
35 variable region (V) can be used to assign taxonomy. In prokaryotes, the region targeted is very often
36 V3/V4, although other regions have been suggested as providing better resolution (e.g. Bukin et al.
37 2019). For eukaryotes, earlier metabarcoding work done on microbial communities used the 18S rRNA
38 gene (Amaral-Zettler et al. 2009; Stoeck et al. 2009), following up on what had been done using ear-
39 lier cloning and Sanger sequencing approaches (López-García et al. 2001; Moon-van der Staay et al.
40 2001). Other genes, and in particular the mitochondrial cytochrome oxidase 1 gene (COI or cox1),
41 have been used for Metazoa but their use is debated in particular because of the lack of universal
42 primers (Andújar et al. 2018; Deagle et al. 2014) and the absence of this gene in lineages that have
43 lost the mitochondrial genome (e.g. Yahalomi et al. 2020). For protists, the 18S rRNA gene appears
44 to be most appropriate as a general marker (Pawlowski et al. 2012), although other genes such as rbcL
45 (large unit of the RUBISCO) have been used for specific purposes such as targeting photosynthetic
46 organisms (e.g. Pujari et al. 2019). Two variable regions of the 18S rRNA gene have been mostly
47 targeted, V4 and V9: V4 is located in the second quarter of the 18S rRNA gene and V9 at the end

48 of the 18S rRNA gene, near the ITS (internally transcribed spacer) region. Initially, the V4 region
49 was favoured for 454 sequencing technology and V9 for Illumina which was then restricted to 2x75
50 bp. However with the development of the Illumina MiSeq (up to 2x300 bp), the V4 region is now
51 preferred, in particular because it is longer, more variable, and better covered in reference databases
52 (Pawlowski et al. 2012).

53 Primer selection is critical to obtain an accurate image of protist communities. Each primer (forward
54 and reverse) must amplify the target community with minimal biases and the region amplified must
55 be long enough to be taxonomically resolute but preferably short enough to be fully sequenced by
56 the chosen technology, although longer amplicons can also be partially sequenced. With Illumina
57 sequencing being now the most used technology, amplicon size must be about 50 bp smaller than
58 the sum of the forward and reverse sequences (called R1 and R2) to allow enough overlap to recon-
59 struct the complete amplicon: for example, the Illumina MiSeq 2x300 bp chemistry can sequence
60 amplicons of up to 550 bp. A large diversity of primer and primer sets targeting the 18S rRNA
61 gene have been developed over the years, although a few dominate in protist metabarcoding studies.
62 Few resources are available that list eukaryotic 18S primers and primer pairs, provide information on
63 their taxonomic specificity and allow to test new primer pairs. Most existing primer databases do
64 not focus on protists. For example the primer database linked the Barcode of Life Data System web
65 site Bold Systems (https://boldsystems.org/index.php/Public_Primer_PrimerSearch) is focusing on
66 metazoans and Probebase (<http://probebase.csb.univie.ac.at/node/8>, Greuter et al. 2016) focuses
67 on bacteria. A few programming tools have been developed to test primer set specificity, for exam-
68 ple EcoPCR (Ficetola et al. 2010), a Python program, or R libraries such as PrimerMiner (Elbrecht
69 and Leese 2017). Unfortunately, these tools need to be installed in a specific computing environment
70 and require some background programming skills. Many existing online tools such as Probematch
71 (<https://rdp.cme.msu.edu/probematch/search.jsp>) only allow testing primer sets against bacteria, ar-
72 chaea and fungi. Silva TestPrime (<https://www.arb-silva.de/search/testprime>) is the only tool that
73 covers protists. It provides very detailed feedback on the taxonomy of amplified sequences, and the
74 location of mismatches. Such detailed information comes at the expense of speed with a typical test
75 needing a few minutes to run. Moreover the taxonomic annotation of the Silva database for protists
76 is not optimal.

77 To fill this gap and to provide protist researchers with a usable tool, we constructed a database of
78 primer and primer sets used for eukaryotic 18S rRNA metabarcoding. These primer sets were tested
79 *in silico* against the PR² database (Guillou et al. 2013) that contains more than 180,000 18S rRNA
80 sequences with expert taxonomical annotation and a subset of the prokaryotic Silva database. We

81 developed a R-based web application that allows to explore the database,d to visualize pre-computed
82 *in silico* amplification results according to taxonomy (% of amplification, size of amplicons and location
83 of mismatches), and to test any user-defined primer set.

84 Material and Methods

85 18S rRNA gene primers (Table S1) and primer sets (Table S2) used in metabarcoding studies were
86 collected from the literature. Primer sequences and primer sets (knowing that several primer sets may
87 share at least one primer) were stored in a MySQL database. Primer sets were tested by performing
88 *in silico* amplification of eukaryotic sequences stored in the PR² reference database (Guillou et al.
89 2013) version 4.12.0 (<https://github.com/pr2database/pr2database/releases/tag/v4.12.0>). We also
90 used a small subset of the Silva database version 132 provided by the mothur web site (https://mothur.org/wiki/silva_reference_files) to test whether prokaryotes were amplified. Sequences with
91 ambiguities were discarded (any nucleotide that is not A, C, G or T). Sequences with length shorter
92 than 1350 bp were not considered except for the V4 region for which this threshold was lowered to
93 1200 bp, since most sequences from PR² contain the V4 region. In contrast, this limit was extended
94 to 1650 for the V9 region and since many 18S rRNA do not cover the full V9 region, we only kept
95 sequences that contained the canonical sequence GGATC[AT] which is located at the end of the V9
96 region, just before the start of the internally transcribed spacer 1 (ITS1). A R (R Development
97 Core Team 2013) script using the *Biostrings* package (Pagès et al. 2020) was used to compute the
98 number of mismatches to the forward and reverse primers allowing for a maximum of 2 mismatches
99 for each primer using the function *matchPattern* with the following parameters: max.mismatch=2,
100 min.mismatch=0, with.indels=FALSE, fixed=FALSE, algorithm="auto". We computed the position
101 of mismatches using the *mismatch* function with parameter fixed=FALSE. A faster version of the
102 script is also available that does not compute mismatch position using the vectorized form of the
103 *matchPattern* function (*vmatchPattern*). The latter function is used in the Shiny application (see
104 below) allowing users to test their own prime sets. The data were tabulated using the *dplyr* package
105 and plotted using the *ggplot2* package (Wickham 2016). A R shiny application to interact with the
106 database was developed using the following R packages: *shiny*, *shinyFeedback* and *shinyCSSloaders*
107 (Sali and Attali 2020).

109 All scripts including those for the Shiny application are available at <https://github.com/pr2database/pr2-primers>.
110

111 Results and Discussion

112 Database of primers and primer sets

113 We have been able to recover from the literature a total of 102 general eukaryotic primers and 77
114 specific to some taxonomic groups (Tables 1 and S1, <https://app.pr2-primers.org>). Some of these
115 primers were designed early on when researchers began to amplify and sequence the 18S rRNA gene
116 (e.g. primers EukA and EukB, Medlin et al. 1988). More recently, researchers have been designing
117 primer specific of some taxonomic groups mostly targeting the division-level (e.g. S19F and S15rF for
118 Foraminifera, Morard et al. 2011) or class-level (e.g. primer PRYM03+3 for Prymnesiophyceae, Egge
119 et al. 2013). Some primers are also designed to block specific taxa (e.g. 18SV1V2Block against the
120 coral *Pocillopora damicornis*, Clerissi et al. 2018) to be used in combination with more general primers
121 (18SV1V2F in this case) or to avoid amplification of some groups (e.g. EUK581-F and EUK1134-R
122 which do not amplify Metazoa, Carnegie et al. 2003). These primers are used when looking at the
123 eukaryotic microbiome of specific organisms (corals, oysters) to avoid amplification of host's genes
124 (Bass and del Campo 2020).

125 We identified a total of 76 primer sets that have been used in metabarcoding studies, mostly targeting
126 protists (Table S2). Of these, most are general, i.e. not targeting specific groups. The distribution
127 of these primer sets over the 18S rRNA gene is very heterogeneous, but the vast majority target the
128 V4 region (Table 2 and Fig. 1). In contrast the number of primer sets targeting the other favoured
129 metabarcoding region V9 is much lower. Most of the primer sets targeting a specific taxonomic group
130 are located in the V4 region, and none are in the V9 region (Table 2). In terms of usage, the V4 region
131 is much more popular (about 80% of published studies in marine systems, Lopes dos Santos et al.
132 2021), the three most commonly used primer sets being #8 (TAReuk454FWD1 and TAReukREV3,
133 Stoeck et al. 2010), #17 (E572F and E1009R, Comeau et al. 2011) and #16 (TAReuk454FWD1 and
134 V4 18S Next.Rev, Piredda et al. 2017), while for the V9 region the most popular sets are #27 (1391F
135 and EukB, Stoeck et al. 2010) and #28 (1380F and 1510R, Amaral-Zettler et al. 2009).

136 Testing primer sets by *in silico* matching

137 General primer sets

138 We used the PR² database (Guillou et al. 2013) which currently contains about 180,000 18S rRNA
139 sequences with a detailed taxonomic annotations to test all primer sets from the pr2-primers database.

140 We also verified on a small set of sequences representative of the different prokaryotic groups whether
141 these primers amplified bacteria or archaea. We only used long sequences (see Material and Methods)
142 and allowed for a maximum of 2 mismatches on both forward and reverse primers, i.e. a maximum of
143 4 mismatches. For general primers, amplification success varied from 32 to more than 97% (Table 3,
144 Fig. 2 and S1). In general, the reverse primer has a tendency to have more mismatches than the
145 forward primer (Table 3). Primer sets targeting regions other than V4 or V9 do not perform as well
146 in general (Fig. S1), although the best overall performance is for #76 targeting the V7 region (F-1183
147 and R-1443, 97.1% of sequences amplified, Lundgreen et al. 2019). If we focus on the V4 and V9
148 regions (Fig. 2), the best performing primer sets overall are #6 (616*f and 1132r, 96.5%, Hugerth
149 et al. 2014) and #29 (1389F and 1510R, 79.8%, Amaral-Zettler et al. 2009). The lower percentage
150 observed for the V9 primers have to be interpreted with caution: many 18S reference sequences do
151 not extend to the end of the V9 region and therefore will miss the signature of the reverse primer. To
152 minimize this problem we retained for the analysis of V9 primer sets only sequences that contain the
153 canonical signature GGATC[AT] located at the end of the V9 region. Despite performing well when
154 allowing for 4 mismatches, some of these primer sets have at least one mismatch to PR² sequences:
155 for example primer set #108 (545F and 1119R, Kataoka et al. 2017) amplifies only 7.9% of the
156 sequences with zero mismatch. Another important consideration is the size of the amplicon. Since
157 most metabarcoding studies are currently using Illumina sequencing technology, the maximum possible
158 size to allow some overlap between the two R1 and R2 reads is about 550 bp (assuming that one uses
159 the 2x300 bp sequencing kits), although smaller amplicons are preferable to allow more overlap. A
160 sizeable fraction of the primer sets produce amplicon close to or larger than 600 bp (Fig. 2). The post
161 sequencing analysis strategy in this case would be to only use one of the reads (R1 is in general less
162 noisy) without trying to assemble R1 and R2 (as done in Lambert et al. 2019).

163 Another important consideration is whether amplification is similar across the whole eukaryotic taxo-
164 nomic range. Taking as example the most used primer set targeting V4 (#8, Fig. 3A) and looking at
165 the amplification efficiency at the supergroup level, a significant fraction of Excavata and to a smaller
166 extent of Rhizaria present at least 5 mismatches to this primer set (Fig. 3A top-left). Amplifica-
167 tion is even more unlikely for sequences presenting mismatches with the forward primer because the
168 mismatches are located at the 3' end of the primer (Fig. 3A top-right) which is the most unfavorable
169 situation (mismatches at the 5' end are better tolerated). The average size of the amplicon is also
170 varying depending on the taxonomic group (Fig. 3B bottom). For example, Excavata have on average
171 longer amplicons because of the presence of introns: amplicon size is then beyond the current range of
172 Illumina sequencing and this may induce as well negative bias during PCR amplification (Geisen et al.

173 2015). For other groups such as Opisthokonta, although the average size is compatible with Illumina
174 sequencing, there is a large number of outlier sequences with long amplicons. This will mean that taxa
175 corresponding to these sequences (mostly Arthropoda) will be missed from surveys conducted with
176 this primer set, although of course this is less critical when protists are targeted. The situation with
177 V9 primer # 27 (Fig. 3) is somewhat similar although there is less dissimilarity between the different
178 supergroups. However for some supergroups, in particular Opisthokonta (Ascomycota) and Archaea-
179 plastida (Bangiophyceae), there is a number of outliers that will be missed by Illumina sequencing.
180 Again these groups are less relevant when focusing on protists. When looking at all the general primer
181 sets (Figure S2), some sets such as # 2, 25, and 110 appear to have more taxonomic biases than others.
182 Overall Excavata constitute the supergroup that is most often discriminated against.

183 Most primer sets will not amplify prokaryotes except primers such as set # 33 (515F and Univ 926R
184 Needham and Fuhrman 2016) that were designed to amplify both bacteria and eukaryotes (Figs. S3
185 and S4). However some primers specific of eukaryotes such as # 4 (563f and 1132r, Hugerth et al. 2014)
186 amplifies quite well prokaryotes. Interestingly, set # 12 (3NDF and 1132rmod, Geisen et al. 2018b)
187 amplify only archaea but not bacteria. In most cases, it is the reverse primer which was discriminating
188 against prokaryotes.

189 Specific primer sets

190 In order to access a deeper diversity within a given taxonomic group primers, primer sets have been
191 developed with specific targets (Tables S1 and S2). Target levels are most often at the division (e.g.
192 Haptophyta) and class levels (e.g. Chrysophyceae), although some sets are targeting supergroups
193 (e.g. SAR # 84). Some primer sets are extremely specific of their targets. One example is primer
194 # 65 targeting Cercozoa (S616F Cerco and S947R Cerco, Fiore-Donno et al. 2018) that presents at
195 least 5 mismatches to all other divisions (Fig. S5) and amplifies all Cercozoa groups. Primer # 38
196 targeting Chlorophyta (ChloroF and ChloroR, Moro et al. 2009) presents at least 5 mismatches to all
197 other divisions (Fig. S5). However, it does not amplify all Chlorophyta as it misses picoplanktonic
198 green algae such as Mamiellophyceae or Chloropicophyceae (Fig. S6). In contrast several primer sets
199 claimed to be specific of a given group are in fact quite general. For example set # 87 which targets
200 oxymonads (Oxy 18S-F and Oxy 18S-R Michaud et al. 2020) amplifies many other groups (Figs. S1
201 and S5). In this case, this is not critical since oxymonads only occur in termite guts and such primers
202 will be used in this specific context. Primer set # 21 (D512for and D978rev, Zimmermann et al. 2011)
203 which was designed to target diatoms would amplify actually most of the Ochrophyta (brown algae)

204 classes but also some green algae (Fig. S6).

205 **R Shiny application**

206 We have developed a web site based on a R Shiny application (<https://app.pr2-primers.org>) that
207 allows users to visualize and download the pr2-primers database, explore at different taxonomy levels
208 the results of *in silico* amplification against the PR² and Silva seed database for the primer sets from
209 the database and test their own primer sets. The application is composed of 6 panels. The first panel
210 (Fig. 4A) provides information on the database as well as link to report issues or new primers. The
211 second and third panels (Fig. 4B) provide an interface to the two tables for primers and primers sets
212 with the option of downloading it (item 1) and of revealing/hiding specific columns. The fourth and
213 fifth panels allow to explore the *in silico* amplification of the primer sets from the database. The
214 fourth panel present a synthesis of the results (similar to Fig. 2) while the fifth panel allows for a given
215 primer set (item 2) to look at amplification properties within a taxonomic level from the kingdom to
216 the class (item 3). The right panel shows general amplification characteristics (item 4), the location
217 of the mismatches, the number of mismatches for each group and the distribution of the amplicon
218 sizes (item 5). Finally the sixth panel allows users to provide a primer set and a maximum number of
219 mismatches (item 6) to run an *in silico* amplification against PR² and Silva seed databases. For the
220 sake of speed, only the number of mismatches is provided but not the position of the mismatches as
221 for the primer sets from the database. Global statistics on the amplification are provided which can
222 be explored at different taxonomic levels (item 7). The R shiny application has been incorporated
223 into a Docker container available at <https://hub.docker.com/repository/docker/vault/pr2-primers>.

224 **Conclusion**

225 The combination of the pr2-primers database with the PR² sequence database provides a very useful
226 resource for protist metabarcoding. It will help researchers to select the most suitable primer pairs
227 for both broadly-targeted surveys and studies focusing on target taxonomic groups, and to test and
228 validate *in silico* novel primers. We emphasize that primer pairs must also be tested on reference
229 culture material and natural samples as actual amplification may differ from *in silico* results. Hopefully
230 this database will grow with time as novel primer pairs are developed and tested on samples from a
231 range of environments. This will contribute to better design and comparability of microbiome analyses,
232 inventories of protist diversity across environments, and increase our understanding of this functionally
233 diverse and important group of organisms.

234 References cited

- 235 Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A Method for
236 Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions
237 of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, *4*, e6372. <http://dx.doi.org/10.1371/journal.pone.0006372>
- 238
- 239 Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode
240 should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, *27*, 3968–
241 3975. <https://doi.org/10.1111/mec.14844>
- 242 Bass, D., & del Campo, J. (2020). Microeukaryotes in animal and plant microbiomes: Ecologies of
243 disease? *European Journal of Protistology*, *76*, 125719. <https://doi.org/10.1016/j.ejop.2020.125719>
- 244
- 245 Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., & Zemskaya, T. I.
246 (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results.
247 *Scientific Data*, *6*, 190007. <https://doi.org/10.1038/sdata.2019.7>
- 248 Carnegie, R., Meyer, G., Blackbourn, J., Cochenne-Laureau, N., Berthe, F., & Bower, S. (2003).
249 Molecular detection of the oyster parasite *Mikrocytos mackini*, and a preliminary phylogenetic
250 analysis. *Diseases of Aquatic Organisms*, *54*, 219–227. <https://doi.org/10.3354/dao054219>
- 251 Clerissi, C., Brunet, S., Vidal-Dupiol, J., Adjeroud, M., Lepage, P., Guillou, L., Escoubas, J.-M., &
252 Toulza, E. (2018). Protists Within Corals: The Hidden Diversity. *Frontiers in Microbiology*, *9*,
253 2043. <https://doi.org/10.3389/fmicb.2018.02043>
- 254 Comeau, A. M., Li, W. K. W., Tremblay, J.-É., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean
255 microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE*,
256 *6*, e27492. <https://doi.org/10.1371/journal.pone.0027492>
- 257 Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding
258 and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, *10*,
259 20140562. <https://doi.org/10.1098/rsbl.2014.0562>
- 260 Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C., & Edvardsen, B. (2013). 454 Pyrose-
261 quencing to Describe Microbial Eukaryotic Community Composition, Diversity and Relative
262 Abundance: A Test for Marine Haptophytes. *PLoS ONE*, *8*, e74371. <https://doi.org/10.1371/journal.pone.0074371>
- 263

- 264 Elbrecht, V., & Leese, F. (2017). PrimerMiner: an R package for development and *in silico* validation
265 of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*, 622–626. <https://doi.org/10.1111/2041-210X.12687>
- 266
- 267 Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F.
268 (2010). An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*, 434.
269 <https://doi.org/10.1186/1471-2164-11-434>
- 270 Fiore-Donno, A. M., Rixen, C., Rippin, M., Glaser, K., Samolov, E., Karsten, U., Becker, B., &
271 Bonkowski, M. (2018). New barcoded primers for efficient retrieval of cercozoan sequences in
272 high-throughput environmental diversity surveys, with emphasis on worldwide biological soil
273 crusts. *Molecular Ecology Resources*, *18*, 229–239. <https://doi.org/10.1111/1755-0998.12729>
- 274 Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., & de Groot, G. A. (2015). Not all are free-living:
275 high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil
276 metazoa. *Molecular Ecology*, *24*, 4556–4569. <https://doi.org/10.1111/mec.13238>
- 277 Geisen, S., Mitchell, E. A., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L. D.,
278 Jousset, A., Krashevska, V., Singer, D., Spiegel, F. W., Walochnik, J., & Lara, E. (2018a). Soil
279 protists: A fertile frontier in soil biology research. *FEMS Microbiology Reviews*, *42*, 293–323.
280 <https://doi.org/10.1093/femsre/fuy006>
- 281 Geisen, S., Snoek, L. B., ten Hooven, F. C., Duyts, H., Kostenko, O., Bloem, J., Martens, H., Quist,
282 C. W., Helder, J. A., & der Putten, W. H. (2018b). Integrating quantitative morphological and
283 qualitative molecular methods to analyse soil nematode community responses to plant range
284 expansion. *Methods in Ecology and Evolution*, *9*, 1366–1378. <https://doi.org/10.1111/2041-210X.12999>
- 285
- 286 Greuter, D., Loy, A., Horn, M., & Rattei, T. (2016). ProbeBase—an online resource for rRNA-targeted
287 oligonucleotide probes and primers: New features 2016. *Nucleic Acids Research*, *44*, D586–
288 D589. <https://doi.org/10.1093/nar/gkv1232>
- 289
- 290 Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de
291 Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann,
292 M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., ... Christen, R. (2013).
293 The Protist Ribosomal Reference database (PR²): a catalog of unicellular eukaryote Small
294 Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*, D597–D604.
295 <https://doi.org/10.1093/nar/gks1160>
- 296
- Hugerth, L. W., Muller, E. E. L., Hu, Y. O. O., Lebrun, L. A. M., Roume, H., Lundin, D., Wilmes, P.,
& Andersson, A. F. (2014). Systematic Design of 18S rRNA Gene Primers for Determining

- 297 Eukaryotic Diversity in Microbial Consortia. *PLoS ONE*, 9, e95567. <http://dx.doi.org/10.1371/journal.pone.0095567>
- 298
- 299 Kataoka, T., Yamaguchi, H., Sato, M., Watanabe, T., Taniuchi, Y., Kuwata, A., & Kawachi, M.
300 (2017). Seasonal and geographical distribution of near-surface small photosynthetic eukaryotes
301 in the western North Pacific determined by pyrosequencing of 18S rDNA. *FEMS Microbiology
302 Ecology*, 93, fiw229. <https://doi.org/10.1093/femsec/fiw229>
- 303 Lambert, S., Tragin, M., Lozano, J.-C., Ghiglione, J.-F., Vaulot, D., Bouget, F.-Y., & Galand, P. E.
304 (2019). Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular
305 environmental perturbations. *The ISME Journal*, 13, 388–401. <https://doi.org/10.1038/s41396-018-0281-z>
- 306
- 307 Lopes dos Santos, A., Ong, D., Vaulot, D., Garczarek, L., Gérikas Ribero, C., Shi, X., & Gutiérrez-
308 Rodríguez, A. (2021). Phytoplankton diversity and ecology through the lens of High Through-
309 put sequencing technologies, In *Advances in phytoplankton ecology: Applications of emerging
310 technologies*. Elsevier.
- 311 López-García, P., Rodriguez-Valera, F., Pedrós-Alió, C., & Moreira, D. (2001). Unexpected diversity
312 of small eukaryotes in deep-sea Antarctic plankton. *Nature*, 409, 603–607.
- 313 Lundgreen, R. B. C., Jaspers, C., Traving, S. J., Ayala, D. J., Lombard, F., Grossart, H.-p., Nielsen,
314 T. G., Munk, P., & Riemann, L. (2019). Eukaryotic and cyanobacterial communities associated
315 with marine snow particles in the oligotrophic Sargasso Sea. *Scientific Reports*, 9, 8891. <https://doi.org/10.1038/s41598-019-45146-7>
- 316
- 317 Medlin, L., Elwood, H. J., Stickel, S., & Sogin, M. L. (1988). The characterization of enzymatically
318 amplified eukaryotic 16S-like rRNA-coding regions. *Gene*, 71, 491–499. [https://doi.org/10.1016/0378-1119\(88\)90066-2](https://doi.org/10.1016/0378-1119(88)90066-2)
- 319
- 320 Michaud, C., Hervé, V., Dupont, S., Dubreuil, G., Bézier, A. M., Meunier, J., Brune, A., & Dedeine,
321 F. (2020). Efficient but occasionally imperfect vertical transmission of gut mutualistic protists
322 in a wood-feeding termite. *Molecular Ecology*, 29, 308–324. <https://doi.org/10.1111/mec.15322>
- 323 Moon-van der Staay, S. Y., De Wachter, R., & Vaulot, D. (2001). Oceanic 18S rDNA sequences from
324 picoplankton reveal unsuspected eukaryotic diversity. *Nature*, 409, 607–610. <https://doi.org/10.1038/35054541>
- 325
- 326 Morard, R., Quillévéré, F., Douady, C. J., de Vargas, C., de Garidel-Thoron, T., & Escarguel, G.
327 (2011). Worldwide Genotyping in the Planktonic Foraminifer *Globoconella inflata*: Implications
328 for Life History and Paleoceanography. *PLoS ONE*, 6, e26665. <https://doi.org/10.1371/journal.pone.0026665>
- 329

- 330 Moro, C. V., Crouzet, O., Rasconi, S., Thouvenot, A., Coffe, G., Batisson, I., & Bohatier, J. (2009).
331 New design strategy for development of specific primer sets for PCR-based detection of Chloro-
332 phyceae and Bacillariophyceae in environmental samples. *Applied and Environmental Micro-*
333 *biology*, 75, 5729–5733. <https://doi.org/10.1128/AEM.00509-09>
- 334 Needham, D. M., & Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea
335 and bacteria following a spring bloom. *Nature Microbiology*, 1, 16005. <https://doi.org/10.1038/nmicrobiol.2016.5>
- 337 Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2020). *Biostrings: Efficient manipulation of*
338 *biological strings* [R package version 2.56.0]. R package version 2.56.0.
- 339 Pawłowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle,
340 J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling,
341 P. J., Kostka, M., Kudryavtsev, A., Lara, E., ... de Vargas, C. (2012). CBOL Protist Working
342 Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. *PLoS*
343 *Biology*, 10, e1001419. <http://dx.doi.org/10.1371/journal.pbio.1001419>
- 344 Piredda, R., Tomasino, M. P., D'Erchia, A. M., Manzari, C., Pesole, G., Montresor, M., Kooistra,
345 W. H. C. F., Sarno, D., & Zingone, A. (2017). Diversity and temporal patterns of planktonic
346 protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbi-*
347 *ology Ecology*, 93, fiw200. <https://doi.org/10.1093/femsec/fiw200>
- 348 Pujari, L., Wu, C., Kan, J., Li, N., Wang, X., Zhang, G., Shang, X., Wang, M., Zhou, C., & Sun, J.
349 (2019). Diversity and spatial distribution of chromophytic phytoplankton in the bay of bengal
350 revealed by RuBisCO Genes (rbcL). *Frontiers in Microbiology*, 10, 1–17. <https://doi.org/10.3389/fmicb.2019.01501>
- 352 R Development Core Team. (2013). R: A Language and Environment for Statistical Computing. <https://doi.org/10.1007/978-3-540-74686-7>
- 354 Sali, A., & Attali, D. (2020). *Shinycssloaders: Add loading animations to a 'shiny' output while it's*
355 *recalculating* [R package version 1.0.0]. R package version 1.0.0. <https://CRAN.R-project.org/package=shinycssloaders>
- 357 Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., &
358 Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere".
359 *Proceedings of the National Academy of Sciences of the United States of America*, 103, 12115–
360 12120.
- 361 Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W., & Richards, T. A.
362 (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex

- 363 eukaryotic community in marine anoxic water. *Molecular Ecology*, 19, 21–31. <https://doi.org/10.1111/j.1365-294X.2009.04480.x>
- 364
- 365 Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M. J., Chistoserdov, A.,
366 Orsi, W., & Edgcomb, V. P. (2009). Massively parallel tag sequencing reveals the complexity of
367 anaerobic marine protistan communities. *BMC Biology*, 7, 72. <https://doi.org/10.1186/1741-7007-7-72>
- 368
- 369 Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer International Publishing.
370 <https://books.google.com.sg/books?id=RTMFswEACAAJ>
- 371 Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J.
372 (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes.
373 *Science*, 347, 1257594. <https://doi.org/10.1126/science.1257594>
- 374 Yahalomi, D., Atkinson, S. D., Neuhof, M., Chang, E. S., Philippe, H., Cartwright, P., Bartholomew,
375 J. L., & Huchon, D. (2020). A cnidarian parasite of salmon (Myxozoa: *Henneguya*) lacks
376 a mitochondrial genome. *Proceedings of the National Academy of Sciences*, 117, 5358–5363.
377 <https://doi.org/10.1073/pnas.1909907117>
- 378 Zimmerman, J., Jahn, R., & Gemeinholzer, B. (2011). Barcoding diatoms: Evaluation of the V4
379 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity
380 and Evolution*, 11, 173–192. <https://doi.org/10.1007/s13127-011-0050-6>

381 **Acknowledgements**

382 We thank the ABIMS platform of the FR2424 (CNRS, Sorbonne Université) for bioinformatics re-
383 sources.

384 **Author contributions statement**

385 DV and SG conceived the study. DV, DB and FM scanned the literature for existing primers and
386 primer sets. DV developed the database, the analysis scripts and the R shiny application. DV wrote
387 the first draft of the paper and all co-authors edited and approved the final version.

388 **Additional information**

389 **Competing interests.** The authors declare no competing financial interests.

390 List of Tables

- 391 Table 1 Type of primers listed in the pr2-primers database. General primers target
392 all eukaryotes and specific primers only certain taxonomic groups.
393 Table 2 Regions of the 18S rRNA gene targeted by the primer sets from the pr2-
394 primers database.
395 Table 3 Overall characteristics of primer sets listed in the pr2-primers database.
396 Table S1 List of primers in the pr2-primers database ordered by start position rela-
397 tive to the sequence of the yeast extitSaccharomyces cerevisiae (FU970071).
398 Table S2 List of primer sets in the pr2-primers database.

399 List of Figures

- 400 Figure 1 Position of the primer sets listed in the pr2-primers database along the
401 18S RNA gene relative to the sequence of the yeast *Saccharomyces cere-*
402 *visiae* (FU970071). The label correspond to the primer set id, the 18S
403 region amplified, its identification name and the specific group it eventu-
404 ally targets. Bar shading indicates whether the primer is general (black)
405 or specific (grey) of a taxonomic group.
406 Figure 2 Evaluation of general primer sets (Table S2) targeting the V4 (top) and
407 V9 (bottom) regions of the 18S rRNA gene against the PR² reference
408 database (version 4.12.0). Left panel. Percentage of reference sequences
409 with at most 2 mismatches to either forward and reverse primer or to
410 both primers, corresponding to the percentage of sequences amplified by
411 the primer set. Central panel. Number of mismatches for each primer
412 set. Right panel. Amplicon sizes targeted by different primer pairs. The
413 vertical lines correspond to the lengths that can be covered by the most
414 commonly used Illumina sequencers (dashed line: 2x250 base pairs; dotted
415 line: 2x300 base pairs). Error bars represent the standard deviation.
416 Figure 3 Example of analysis for two primer sets amplifying two regions of the 18S
417 rRNA gene: V4 (primer set #8, A) and V9 (primer set #27, B). Top
418 left. Percentage of sequences with a given number of mismatches. Top
419 right. Position of the mismatches for different taxonomic supergroups on
420 the forward and reverse primer counted from the 5' end. Bottom left. Dis-
421 tribution of amplicon size for different supergroups. Bottom right. Box
422 plots of amplicon size. Colors correspond to taxonomy (division). Hacro-
423 bia represents the sum of Haptophytes, Cryptophytes and Centrohelids.
424 Figure 4 Shiny interface to the pr2-primers database. See text for details

- 425 **Figure S1** Evaluation of general (top) or specific (bottom) primer sets (Table S2) for
426 the 18S rRNA gene against the PR² reference database (version 4.12.0).
427 See Fig. 2 for legend.
- 428 **Figure S2** Number of mismatches (A) and amplicon size (mean \pm SD, B) for general
429 primer sets as a function of the supergroup.
- 430 **Figure S3** Evaluation of general primer sets (Table S2) targeting the V4 region of
431 the 18S rRNA gene against bacterial 16S rRNA sequences from the Silva
432 seed reference database (version 132). Legend as in Figure 2.
- 433 **Figure S4** Evaluation of general primer sets (Table S2) targeting the V4 region of
434 the 18S rRNA gene against archaeal 16S rRNA sequences from the Silva
435 seed reference database (version 132). Legend as in Figure 2.
- 436 **Figure S5** Number of mismatches (A) and amplicon size (mean \pm SD, B) for specific
437 primer sets as a function of the supergroup.
- 438 **Figure S6** Percentage of sequences amplified with specific primer sets for different
439 photosynthetic classes belonging to the Ochrophyta, Haptophyta, Dinoflag-
440 ellata and Chlorophyta divisions.

Table 1: Type of primers listed in the pr2-primers database. General primers target all eukaryotes and specific primers only certain taxonomic groups.

direction	general	specific
fwd	52	42
rev	49	36

Table 2: Regions of the 18S rRNA gene targeted by the primer sets from the pr2-primers database.

gene region	general	specific
37F-41F		1
V1-V2	1	1
V2-V3	1	3
V3		1
V3-V4		2
V4	32	13
V4-V5	1	
V5		2
V6		1
V6-V8	1	
V7	2	
V7-V8		1
V8-V9	2	
V9	4	

Table 3: Overall characteristics of primer sets listed in the pr2-primers database.

	general	specific
forward primer	% of sequences amplified	
min	36.4	0.0
mean	92.1	64.8
max	98.7	97.6
reverse primer		
min	43.2	0.0
mean	89.9	32.4
max	98.6	96.1
both primers		
min	30.0	0.0
mean	85.0	27.3
max	96.5	92.7
amplicon size	bp	
min	174.6	184.0
mean	460.1	441.5
max	737.6	785.9

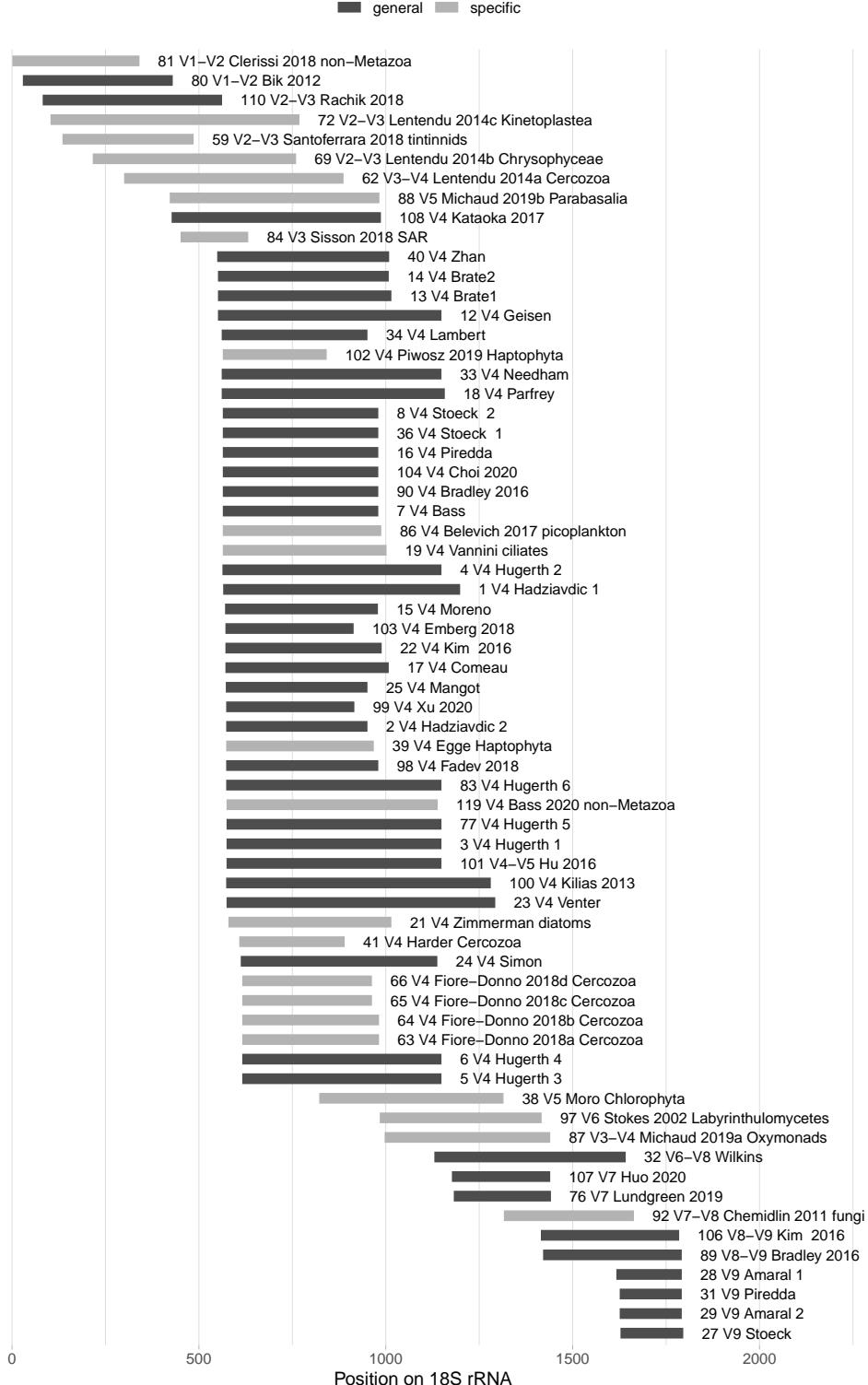


Figure 1: Position of the primer sets listed in the pr2-primers database along the 18S RNA gene relative to the sequence of the yeast *Saccharomyces cerevisiae* (FU970071). The label correspond to the primer set id, the 18S region amplified, its identification name and the specific group it eventually targets. Bar shading indicates whether the primer is general (black) or specific (grey) of a taxonomic group.

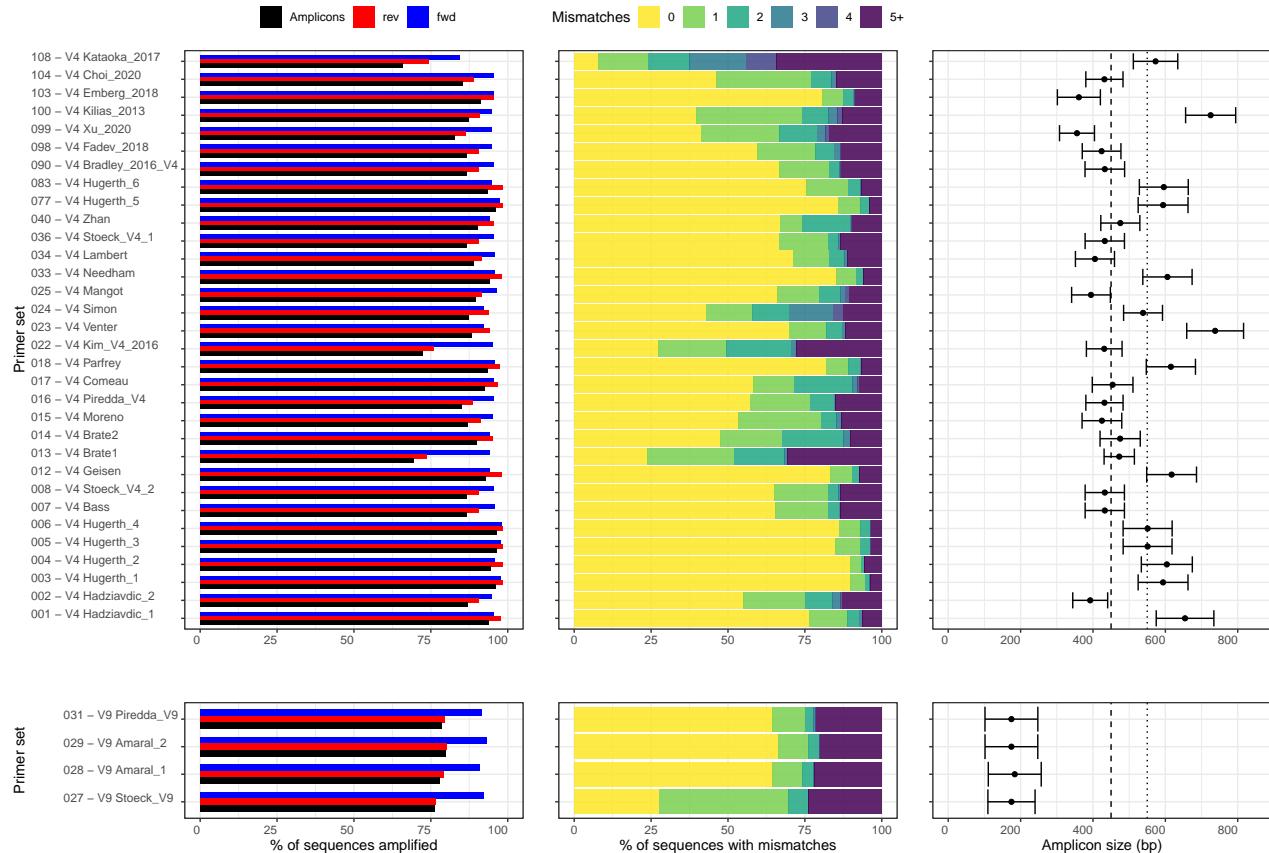


Figure 2: Evaluation of general primer sets (Table S2) targeting the V4 (top) and V9 (bottom) regions of the 18S rRNA gene against the PR² reference database (version 4.12.0). Left panel. Percentage of reference sequences with at most 2 mismatches to either forward and reverse primer or to both primers, corresponding to the percentage of sequences amplified by the primer set. Central panel. Number of mismatches for each primer set. Right panel. Amplicon sizes targeted by different primer pairs. The vertical lines correspond to the lengths that can be covered by the most commonly used Illumina sequencers (dashed line: 2x250 base pairs; dotted line: 2x300 base pairs). Error bars represent the standard deviation.

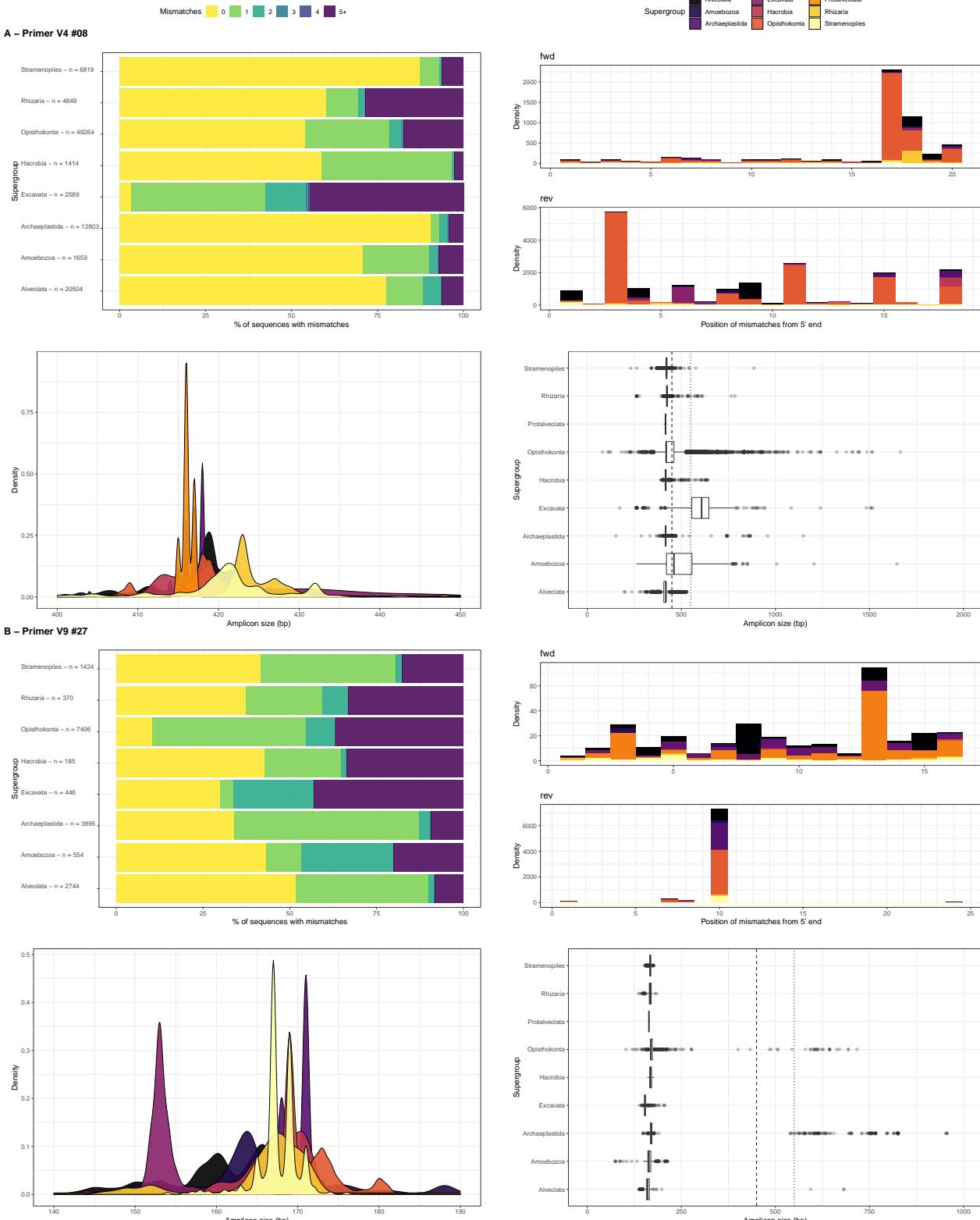


Figure 3: Example of analysis for two primer sets amplifying two regions of the 18S rRNA gene: V4 (primer set #8, A) and V9 (primer set #27, B). Top left. Percentage of sequences with a given number of mismatches. Top right. Position of the mismatches for different taxonomic supergroups on the forward and reverse primer counted from the 5' end. Bottom left. Distribution of amplicon size for different supergroups. Bottom right. Box plots of amplicon size. Colors correspond to taxonomy (division). Hacrobia represents the sum of Haptophytes, Cryptophytes and Centrohelids.

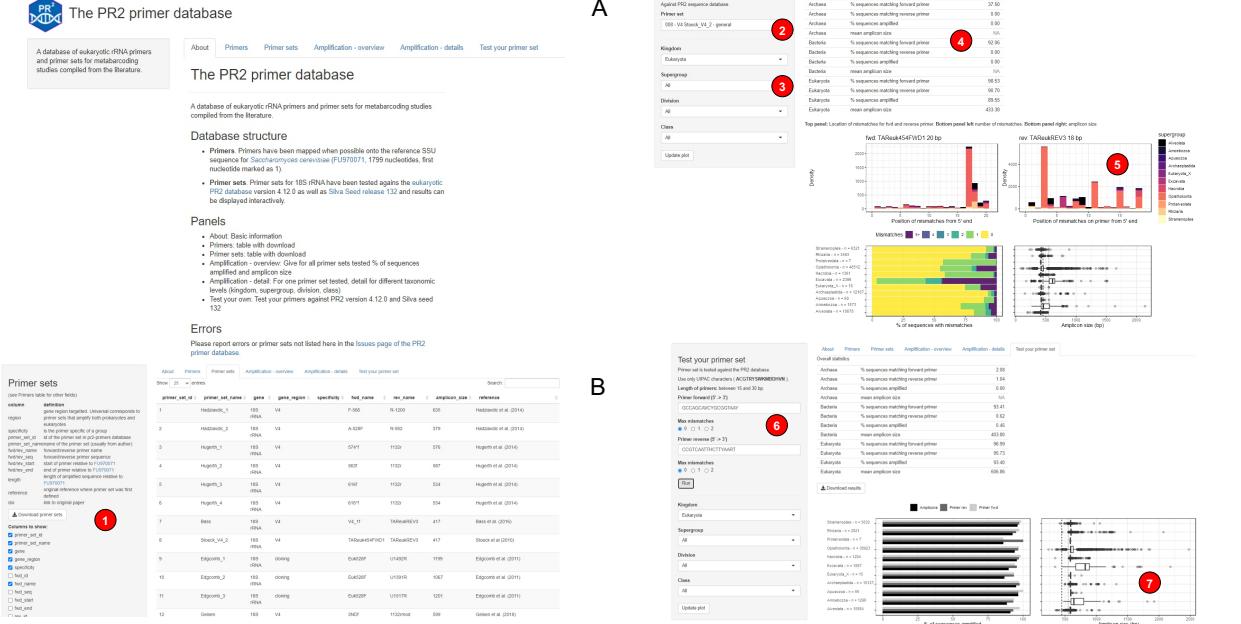


Figure 4: Shiny interface to the pr2-primers database. See text for details

441 Supplementary Material

Table S1: List of primers in the pr2-primers database ordered by start position relative to the sequence of the yeast *Saccharomyces cerevisiae* (FU970071).

id	Name	Sequence	Direction	Start (yeast)	Specificity	DOI
71	PF1	TGCGCTACCTGGTTGATCCTGCC	fwd	-5		10.1078/0932-4739-00874
78	EukA	AACCTGGTTGATCCTGCCAGT	fwd	0		10.1016/0378-1119(88)90066-2
81	Euk328F	ACCTGGTTGATCCTGCCAG	fwd	1		10.1038/35054541
138	18S1V2F	ACCTGGTTGATCCTGCCA	fwd	1	non-Metazoa	10.3389/fmicb.2018.02043
220	NSF4/18	CTGGTTGATYCTGCCAGT	fwd	3		10.1016/S0723-2020(89)80066-9
168	Pbr1	GGTTGATCCTGCCAGTAGTC	fwd	5	Plasmodiophora	10.1016/j.protis.2011.02.005
169	Pbr1r	GACTACTGGCAGGATCAACC	rev	5	Plasmodiophora	10.1016/j.protis.2011.02.005
109	SF2Dark	GTTGATCCTGCCAGTAGTGT	fwd	6	Myxomycetes	10.1038/srep19068
155	NS1	GTAGTCATATGCTTGCTC	fwd	19		10.1016/B978-0-12-372180-8.50042-1
142	25F	CATATGCTGTCTAAAGATTAAGCCA	fwd	24		10.1016/j.protis.2009.03.003
79	63f	ACGCTTGTCTCAAAGATTAAGGATTA	fwd	27		10.1111/j.1574-6941.2011.01072.x
136	F04	GCTTGTCTCAAAGATTAAGGCC	fwd	29		10.1111/j.1365-294X.2011.05297.x
234	18S-42F	CTCAARGAYTAAGGCATGCA	fwd	35		10.1073/pnas.0235779100
107	SFAca22	CGGYGAGACTGCCGATGG	fwd	78	Acanthamoeba	10.1038/srep19068
231	18S-82F	GAAACTGCGAATGGCTC	fwd	82		10.1073/pnas.0235779100
104	Kineto 80	CATCAGACGYAACCTGCCGC	fwd	103	Kinetoplastea	10.1111/mec.12819
82	152+	TTACATGGATAACCGTGGTAATT	fwd	135	Tintinnids	10.1093/plankt/fby011
98	Chryso 240	GGAAACCAATGCCGGGCAAC	fwd	216	Chrysophyceae	10.1111/mec.12819
88	Cer2F	ATTTCTGCCCTATCAGCT	fwd	300	Cercozoa	10.1111/mec.12819
140	18S1V2Block	CTACCTTACCATCGACAGTTGATAG	rev	309	blocking Pocillopora damicornis	10.3389/fmicb.2018.02043
139	18S1V2R	GTARKCCWMTAYMYTACC	rev	324	non-Metazoa	10.3389/fmicb.2018.02043
137	R22	GCCTGCTGCCTTCCTTGG	rev	412		10.1111/j.1365-294X.2011.05297.x
195	Par 18S-F	GCAGCAGCGYGAAC	fwd	422	Parabasalids	10.1111/mec.15322
227	545F	AGGCCGTAATTACCCAATC	fwd	427		10.1016/j.gene.2015.10.033
151	SAR V3 F	AYTCAGGGAGGTAGTGACAAG	fwd	451	SAR	10.1093/plankt/fby020
83	528-	CCCGGCCCGTTATTCTTGT	rev	467	Tintinnids	10.1093/plankt/fby011
232	Euk-516r	ACCAAGACTTGCCCTCC	rev	547		10.1128/AEM.56.6.1919-1925.1990
65	Uni18SF	AGGCCAAKYCTGGTGCAGC	fwd	549		10.1111/2041-210X.12037
9	3NDf	GGCAAGTCTGGTGCAG	fwd	551		10.1016/j.protis.2009.03.003
156	NS2	GGCTGCTGCCAACAGACTGC	rev	552		10.1016/B978-0-12-372180-8.50042-1
157	NS3	GCAAGTCTGGTGCAGCAGCC	fwd	552		10.1016/B978-0-12-372180-8.50042-1
13	515F	GTGCCAGCMGCCGCGTAA	fwd	561		10.3389/fmicb.2014.00298
19	515FY	GTGYCAGCMGCCGCGTAA	fwd	561		10.1111/1462-2920.13023
31	515F Univ	GTGYCAGCMGCCGCGTAA	fwd	561		10.1038/nm microbiol.2016.5
4	563f	GCCAGCAVCYGCCGTAAY	fwd	563		10.1371/journal.pone.0095567
7	V 4 f	CCAGCASCYGCCGTAATWCC	fwd	564		10.1111/1462-2920.13235
8	TAReuk454FWD1	CCAGCASCYGCCGTAATTCC	fwd	564		10.1111/j.1365-294X.2009.04480.x
67	Claudia Vannini (F)	CCAGCASCYGCCGTAATWCC	fwd	564	ciliates	10.1007/s00248-016-0912-8
187	Euf-V4	CCAGCASCYGCCGTAATWCC	fwd	564		10.1007/s00248-017-1076-x
218	TAReuk454FWD1 Choi	CCAGCAGCCGCCGTAATTCC	fwd	564		10.1038/s41598-020-63561-z
1	F-566	CAGCAGCCGCCGTAATTCC	fwd	565		10.1371/journal.pone.0087624
10	EUKAF	GCCGCCGTAATTCCAGCTC	fwd	570		10.1016/j.jjeh.2017.10.008
69	ParaV45F	GCYCGGTAATTCCAGCTCT	fwd	570	Parabasalids	10.1016/j.ejop.2017.09.001
12	E572F	CYCGGTAATTCCAGCTC	fwd	571		10.1371/journal.pone.0027492
14	528F	CCCGCGGTAATTCCAGCTC	fwd	571		10.1016/j.femsec.2004.10.006
17	NSF563	CGCGGTAATTCCAGCTCCA	fwd	572		10.1111/1462-2920.12065
75	SSU556F	CGCGGTAATTCCAGCTYC	fwd	572	dinoflagellates	10.1080/00288330.2017.1298632
2	A-528F	GCGGTAAATTCCAGCTCAA	fwd	573		10.1038/ismej.2010.26
73	DIV4for	GCGGTAAATTCCAGCTCAAATAG	fwd	573	diatoms	10.1021/es506158m
3	574*f	CGGTAAYTCCAGCTCYV	fwd	574		10.1371/journal.pone.0095567
15	590F	CGGTAATTCCAGCTCCAATAGC	fwd	574		10.1016/j.protis.2017.03.005
32	Euk528F	CGGTAATTCCAGCTCC	fwd	574		10.1038/ismej.2011.6
132	574f	CGGTAAYTCCAGCTCYAV	fwd	574		10.1371/journal.pone.0095567
20	D512for	ATTCAGCTCCAATAGCG	fwd	579	diatoms	10.1007/s13127-011-0050-6
199	FF1100	CCAGCTCCAATAGCGTATTA	fwd	582	Fungi	10.1017/S0953756200002471
207	S32 J	CCAGCTCCAATAGCGTATAC	fwd	582	Radiolaria	10.1016/j.protis.2011.10.002

id	Name	Sequence	Direction	Start (yeast)	Specificity	DOI
208	S32 TASN	CCAGCTCCAATAGCGTATRC	fwd	582	Radiolaria	10.1016/j.dsr2.2011.12.005
21	Cerc479F	TGTGCACTTAAAGCTCGT	fwd	608		10.1038/ismej.2016.31
16	EK-565F-NGS	GCAGTTAAAAGCTCGTAGT	fwd	612		10.1111/1462-2920.12591
152	SAR V3 R	RACTACGAGCTTTAACACTGC	rev	612	SAR	10.1093/plankt/fby020
5	616f	TTAAAAGVYTCGTAGTYG	fwd	616		10.1371/journal.pone.0095567
6	616'f	TTAAARVGVTGCGTAGTYG	fwd	616		10.1371/journal.pone.0095567
90	S616F Cerco	TTAAAAAGCTCGTAGTTG	fwd	616	Cercozoa	10.1111/1755-0998.12729
91	S616F Eocer	TTAAAAAGCGCGTAGTTG	fwd	616	Cercozoa	10.1111/1755-0998.12729
99	Chryso 651	CTATTTGCTCACAGTAAAGCAGAG	rev	735	Chrysophyceae	10.1111/mec.12819
105	Kineto 651	TTGGTCGCRCTTGTAGTCACAG	rev	746	Kinetoplastea	10.1111/mec.12819
23	ChloroF	TGGCTATCTTGTGGCTGT	fwd	822	Chlorophyceae	10.1128/AEM.00509-09
190	Haptor1	CGAAACCAACAAAATAGCAC	rev	823	Prymnesiophyceae	10.1371/journal.pone.0074371
206	S879	CCAACTGTCCCTATCAATCAT	rev	855	Radiolaria	10.1016/j.protis.2011.10.002
89	Cer1R	ATACTAGCACCCCCAAC	rev	870	Cercozoa	10.1111/mec.12819
52	Cerc750R	TGAATACTAGCACCCCCAAC	rev	871	Cercozoa	10.1038/ismej.2016.31
74	DIV4rev3	CTCTGACAATGGAAATACGAATA	rev	879	diatoms	10.1021/es506158m
22	DimA	RGGGACRGGTGAAATAGGATG	fwd	893	diplonemids	10.1186/s40168-018-0581-6
76	SSU911R	ATYCAAGAATTTCACCTCTGAC	rev	894	dinoflagellates	10.1080/00288330.2017.1298632
146	690R	ATCCAAGAATTTCACCTCTGAC	rev	894		10.5194/bg-8-2125-2011
77	B-706R	AATCCRAGAATTTCACCTCT	rev	897		10.1038/ismej.2010.26
127	897f	AGAGGTGRAATTCTHRGA	fwd	897		10.1371/journal.pone.0095567
128	897r	TCYDAGAATTYCACCTCT	rev	897		10.1371/journal.pone.0095567
35	R-952	TTGGCAAATGCTTCGC	rev	935		10.1371/journal.pone.0087624
49	NSR951	TTGGYRAATGCTTCGC	rev	935		10.1111/1462-2920.12065
93	S947R Cerco	AAGAAGACATCCTGGTG	rev	946	Cercozoa	10.1111/1755-0998.12729
141	18SV4Block	TCTGATTATGAAAAACATTCTTGGC	rev	947	blocking Pocillopora damicornis	10.3389/fmicb.2018.02043
55	PRYM01+7	GATCAGTAAAACATCCCTGG	rev	948	Haptophyta	10.1371/journal.pone.0074371
43	V4 18S Next.Rev	ACTTTCGTTCTTGATYRATGA	rev	960		10.1093/femssec/fiw200
92	S963R Cerco	CAACTTCGTTCTTGATAAA	rev	962	Cercozoa	10.1111/1755-0998.12729
37	TAReukREV3	ACTTCGTTCTTGATYRA	rev	963		10.1111/j.1365-294X.2009.04480.x
38	V4RB	ACTTCGTTCTTGATYRR	rev	963		10.3354/ame01740
42	EUKAR	CYTTCGYYCTTGATTRA	rev	963		10.1016/j.ijheh.2017.10.008
219	TAReukREV3 Choi	ACTTCGTTCTTGATAA	rev	963		10.1038/s41598-020-63561-z
196	Par 18S-R	CCTACTCTCGCYCTTGATCG	rev	964	Parabasalids	10.1111/mec.15322
188	picoR2	AKCCCCYAACTTCGTTCTTGAT	rev	966		10.1007/s00248-017-1076-x
202	V4r	ACTTCGTTCTTGAT	rev	966		10.1128/AEM.01630-16
46	Nex 18S 0964 R	GATCCCYYAACTTCGTTCTTGA	rev	967		10.1111/1462-2920.13523
228	1119R	TCCCCTAACTTCGTTCTTG	rev	968		10.1016/j.gene.2015.10.033
68	Claudia Vannini (R)	TCTGRTYGCTTTGATCCYTA	rev	981	ciliates	10.1007/s00248-016-0912-8
216	LABY-A	GGGATCGAAGATGATTAG	fwd	984	Labyrinthulomycetes	10.3354/dao052233
40	V4 euk R1	GACTACGACGGTATCTRATCRTCTTCG	rev	989		10.1038/ismej.2010.39
41	V4 euk R2	ACGGTATCTRATCRTCTTCG	rev	989		10.1038/ismej.2010.39
44	E1009R	AYGGTATCTRATCRTCTTYG	rev	989		10.1371/journal.pone.0027492
66	Uni18SR	GRCGGTATCTRATCGYCTT	rev	991		10.1111/2041-210X.12037
51	D978rev	GACTACGATGGTATCTAAAC	rev	996	diatoms	10.1007/s13127-011-0050-6
193	Oxy 18S-F	ATCAGAWACCGYCGTAGTC	fwd	997	Oxymonads	10.1111/mec.15322
198	FF700	GATACCGTNGTAGTCT	fwd	1001	Fungi	10.1017/S0953756200002471
48	EUK1134-R	TTTAAGTTTCAGCCTTGCG	rev	1120		10.3354/dao054219
265	UNonMet DB	CTTTAARTTCASYCTTGCG	rev	1120	non-Metazoan	10.1016/j.ejop.2020.125719
30	926wF	AAACTYAAAKGAATTGRCGG	fwd	1130		10.1038/ncomms3457
60	926R	CCGYCAATTYMTTRAGTT	rev	1130		10.1038/nmicrobiol.2016.5
159	NS5	AACTTAAAGGAATTGACGGAAG	fwd	1131		10.1016/B978-0-12-372180-8.50042-1
36	1132r	CCGTCAATTHTCTTYAART	rev	1132		10.1371/journal.pone.0095567
211	1132rmod	TCCGTCAATTYCTTAAGT	rev	1132		10.1371/journal.pone.0095567
158	NS4	CTTCCGTCAATTCTCTTAAAG	rev	1133		10.1016/B978-0-12-372180-8.50042-1
224	1132R modified	CCGTCAATTHTCTTYAAR	rev	1133		10.3389/fmicb.2016.00679
70	ParaV45R	AAGRAATTGACGGAAGNGCA	rev	1137	Parabasalids	10.1016/j.ejop.2017.09.001
45	1119r	GGTGCCTTCCGTCA	rev	1144		10.3389/fmicb.2014.00298
53	DimB	CAAATTGAGCCGAGACTCC	rev	1168		10.1186/s40168-018-0581-6
225	960F	GGCTTAATTGACTCAACRCG	fwd	1177		10.1128/AEM.70.4.2028-2037.2004
34	R-1200	CCCGTGTGAGTCAAATTAAGC	rev	1178		10.1371/journal.pone.0087624
129	F-1183	AATTGACTCAACACGGG	fwd	1182		10.1371/journal.pone.0087624
94	1301f	GATTGAAGCTTTCTGATCACTTC	fwd	1236	Plasmodiophorida	10.3389/fmicb.2018.00168

id	Name	Sequence	Direction	Start (yeast)	Specificity	DOI
148	1055R	ACGGCCATGCACCACCAACCAT	rev	1260		10.5194/bg-8-2125-2011
147	1055F	GGTGGTCATGGCGTTCTT	fwd	1266		10.5194/bg-8-2125-2011
47	1300R	CACCAACTAAGAACGGCCATGC	rev	1272		10.1016/j.protis.2017.03.005
54	ChloroR	GAATCAACCTGACAAGGCAAC	rev	1295	Chlorophyceae	10.1128/AEM.00509-09
197	FF390	CGATAACGAAACGAGACCT	fwd	1316	Fungi	10.1017/S0953756200002471
217	LABY-Y	CWCRAACTCCTCCGGT	rev	1400	Labyrinthulomycetes	10.3354/dao052233
160	NS6	GCATCACAGACCTGTTATTGCCCT	rev	1415		10.1016/B978-0-12-372180-8.50042-1
161	NS7	GAGGCATAAACAGGTCTGTGATGC	fwd	1415		10.1016/B978-0-12-372180-8.50042-1
221	Nex 18S 1434 F	GAGGCATAAACAGGTCTGTGATG	fwd	1415		10.1111/1462-2920.13523
201	V8f	ATAACAGGTCTGTGATGCCCT	fwd	1421		10.1128/AEM.01630-16
203	1422f	ATAACAGGTCTGTGATGC	fwd	1421		10.1371/journal.pone.0087624
226	NSR1438	GGGCATCACAGACCTGTTAT	rev	1421		10.1093/nar/28.1.175
194	Oxy 18S-R	GGGCATMACRGACCTGTTA	rev	1422	Oxymonads	10.1111/mec.15322
204	1424f	AACAGGTCHGWRATGCC	fwd	1423		10.1371/journal.pone.0095567
131	R-1443	AAGGGCATCACAGACCTG	rev	1425		10.1371/journal.pone.0087624
238	SL175pr5F	ACGAGGAATGCCTAGTAAAGCGCAA	fwd	1569	Mantoniella antarctica	10.1111/1574-6941.12334
27	1380F	CCCTGCCHTTGACACAC	fwd	1617		10.1371/journal.pone.0006372
62	U1391R	GGGCGGTGTGACARGR	rev	1623		10.1038/ismej.2011.6
28	1389F	TTGTACACACCGCC	fwd	1626		10.1371/journal.pone.0006372
29	1388F	TTGTACACACCGCCGTGCG	fwd	1626		10.1093/femsec/fiw200
26	1391F	GTACACACCGCCCGTC	fwd	1628		Lane, D.J. (1991)
59	1392-R	ACGGGCGGTGTGTRC	rev	1628		10.1038/ncomms3457
145	18r71	GCGACGGGCGGTGTGTC	rev	1628		10.5194/bg-8-2125-2011
166	ITS9MUNngs	TACACACGGCCCGTCG	fwd	1629		10.1111/1758-2229.12438
200	FR1	ANCCATTCAATCGGTANT	rev	1647	Fungi	10.1017/S0953756200002471
95	1801r	ACGGAAACCTTGTACGACTTC	rev	1753	Plasmodiophorida	10.3389/fmicb.2018.00168
61	U1492R	GGTTACCTTGTACGACTT	rev	1754		10.1038/ismej.2011.6
63	U1517R	ACGGCTACCTTGTACGACTT	rev	1754		10.1038/ismej.2011.6
80	1818r	ACGGAAACCTTGTACGA	rev	1757		10.1111/j.1574-6941.2011.01072.x
235	18S-1498R	CACCTACGGAAACCTTGTTA	rev	1760		10.1073/pnas.0235779100
222	Nex 18S 1757 R	CAGGTTCACCTACGGAAACCT	rev	1765		10.1111/1462-2920.13523
237	RS11pr4R	CTGCAGGTTCACCTACGGAAACC	rev	1766	Pyramimonas cf. tychotreta	10.1111/1574-6941.12334
162	NS8	TCCGCAGGTTCACCTACGGA	rev	1770		10.1016/B978-0-12-372180-8.50042-1
57	EukB	TGATCCTCTGCAGGTTCACCTAC	rev	1773		10.1016/0378-1119(88)90066-2
58	1510R	CCTTCYGCAGGTTCACCTAC	rev	1773		10.1073/pnas.0235779100
143	1801R	TGATCCTCTGCAGGTTCACCT	rev	1775		10.1016/j.protis.2009.03.003
25	EUK581-F	GTGCCAGCAGCCGCG	rev		non-Metazoan	10.3354/dao054219
96	s14f1	AAGGGCACCAACAAGAACGC	fwd		Foraminifera	10.1007/PL00006232
97	s15.3	CCTATCACATAATCATGAAAG	rev		Foraminifera	10.1111/1755-0998.12261
106	SRAc28	CCAATTACAAGACTTTRTCGAG	fwd		Acanthamoeba	10.1038/srep19068
108	SR19Dark	GTCCTCTAATTGTACTCGAD	fwd		Myxomycetes	10.1038/srep19068
112	Pdir1	GATTTCGGGGGGTTTACCCGGA	fwd		Pedinophyceae	10.1002/mbo3.892
113	Pdir2	GATCGGGCTCGGGTTCGAG	fwd		Pedinophyceae	10.1002/mbo3.892
114	Prev2	CTCGCGGAACTCGAACCGAAG	rev		Pedinophyceae	10.1002/mbo3.892
115	Pdir3	CCTCAGCCTGCTAAATAGCTAC	fwd		Pedinophyceae	10.1002/mbo3.892
116	Pdir4	GACTTCGGGGTTTACCCGGA	fwd		Pedinophyceae	10.1002/mbo3.892
134	S19F	GTGCATGGCCGTCTTAGTTC	rev		Foraminifera	10.1371/journal.pone.0026665
135	S15rF	CCCGTACRAGGCATTCTAG	fwd		Foraminifera	10.1371/journal.pone.0026665
144	329R	GTGAACCTGCRGAAGGATCA	rev			10.5194/bg-8-2125-2011
170	Pb121	GGATACAAAACCAAAACCTGGC	fwd		Plasmodiophora	10.1016/j.protis.2011.02.005
171	Pb121r	GCCAGGTTGGTTTTGTATCC	rev		Plasmodiophora	10.1016/j.protis.2011.02.005
186	SB	GTAGGTGAACCTGCAGAAGGATCA	rev			Sogin (1990)
192	PRYM03+3	GTAATTGCCCCGATCCTG	fwd		Prymniosiphycaceae	10.1371/journal.pone.0074371
205	17	CGGTACGTTCGTTG	rev		Foraminifera	10.1016/j.pocean.2019.102175
210	S51 TAS	YAAGAATTTCACCTCTCGCTT	rev		Radiolaria	10.1016/j.dsr2.2011.12.005
214	QPX-F	ATCCTCGGCTGCTTTAGTAG	fwd		Quahog parasite	10.3354/dao052233
215	QPX-R2	GAAGTCTCACCTTCTTGC	rev		Quahog parasite	10.3354/dao052233
230	s14F3	ACGCAMGTGTAAACTTG	fwd		Foraminifera	10.1111/j.1550-7408.2003.tb00248.x
236	RS11pr4F	ATGTCGGATCGGGCGAGAC	fwd		Pyramimonas cf. tychotreta	10.1111/1574-6941.12334
239	SL175pr5R	TAGAAAGCCACGGTCCGAACGC	rev		Mantoniella antarctica	10.1111/1574-6941.12334
240	Gempr2F	TCGGATTGCTGGGTAGAACCTCGT	fwd		Geminigera cryophila	10.1111/1574-6941.12334
241	Gempr2R	CACCTACGGAAACCTTGTACGAC	rev		Geminigera cryophila	10.1111/1574-6941.12334

Table S2: List of primer sets in the pr2-primers database.

id	Name	Primer fwd	Primer rev	Region	Specificity	DOI
1	Hadziavdic 1	F-566	R-1200	V4		10.1371/journal.pone.0087624
2	Hadziavdic 2	A-528F	R-952	V4		10.1371/journal.pone.0087624
3	Hugerth 1	574*f	1132r	V4		10.1371/journal.pone.0095567
4	Hugerth 2	563f	1132r	V4		10.1371/journal.pone.0095567
5	Hugerth 3	616f	1132r	V4		10.1371/journal.pone.0095567
6	Hugerth 4	616*f	1132r	V4		10.1371/journal.pone.0095567
7	Bass	V4 1f	TAReukREV3	V4		10.1111/1462-2920.13235
8	Stoeck V4 2	TAReuk454FWD1	TAReukREV3	V4		10.1111/j.1365-294X.2009.04480.x
12	Geisen	3NDf	1132rmod	V4		10.1111/2041-210X.12999
13	Brate1	3NDf	V4 euk R1	V4		10.1038/ismej.2010.39
14	Brate2	3NDf	V4 euk R2	V4		10.1038/ismej.2010.39
15	Moreno	EUKAF	EUKAR	V4		10.1016/j.jheh.2017.10.008
16	Piredda V4	TAReuk454FWD1	V4 18S Next.Rev	V4		10.1093/femsec/fiw200
17	Comeau	E572F	E1009R	V4		10.1371/journal.pone.0027492
18	Parfrey	515F	1119r	V4		10.3389/fmicb.2014.00298
19	Vannini	Claudia Vannini (F)	Claudia Vannini (R)	V4	ciliates	10.1007/s00248-016-0912-8
21	Zimmerman	D512for	D978rev	V4	diatoms	10.1007/s13127-011-0050-6
22	Kim V4 2016	528F	Nex 18S 0964 R	V4		10.1111/1462-2920.13523
23	Venter	590F	1300R	V4		10.1016/j.protis.2017.03.005
24	Simon	EK-565F-NGS	EUK1134-R	V4		10.1111/1462-2920.12591
25	Mangot	NSF563	NSR951	V4		10.1111/1462-2920.12065
27	Stoeck V9	1391F	EukB	V9		10.1111/j.1365-294X.2009.04480.x
28	Amaral 1	1380F	1510R	V9		10.1371/journal.pone.0006372
29	Amaral 2	1389F	1510R	V9		10.1371/journal.pone.0006372
31	Piredda V9	1388F	1510R	V9		10.1093/femsec/fiw200
32	Wilkins	926wF	1392-R	V6-V8		10.1038/incomms3457
33	Needham	515F Univ	926R	V4		10.1038/nmicrobiol.2016.5
34	Lambert	515FY	NSR951	V4		10.1038/s41396-018-0281-z
35	UNonMet	EUK581-F	EUK1134-R	V4	non-Metazoa	10.3354/dao054219
36	Stoeck V4 1	TAReuk454FWD1	V4RB	V4		10.3354/ame01740
38	Moro	ChloroF	ChloroR	V5	Chlorophyta	10.1128/AEM.00509-09
39	Egge	A-528F	PRYM01+7	V4	Haptophyta	10.1371/journal.pone.0074371
40	Zhan	Uni18SF	Uni18SR	V4		10.1111/2041-210X.12037
41	Harder	Cerc479F	Cerc750R	V4	Cercozoa	10.1038/ismej.2016.31
59	Santoferrara 2018	152+	528-	V2-V3	tintinnids	10.1093/plankt/fby011
62	Lentendu 2014a	Cer2F	Cer1R	V3-V4	Cercozoa	10.1111/mec.12819
63	Fiore-Donno 2018a	S616F Cерко	S963R Cерко	V4	Cercozoa	10.1111/1755-0998.12729
64	Fiore-Donno 2018b	S616F Eocer	S963R Cерко	V4	Cercozoa	10.1111/1755-0998.12729
65	Fiore-Donno 2018c	S616F Cерко	S947R Cерко	V4	Cercozoa	10.1111/1755-0998.12729
66	Fiore-Donno 2018d	S616F Eocer	S947R Cерко	V4	Cercozoa	10.1111/1755-0998.12729
69	Lentendu 2014b	Chryso 240	Chryso 651	V2-V3	Chrysophyceae	10.1111/mec.12819
72	Lentendu 2014c	Kineto 80	Kineto 651	V2-V3	Kinetoplastea	10.1111/mec.12819
76	Lundgreen 2019	F-1183	R-1443	V7		10.1038/s41598-019-45146-7
77	Hugerth 5	574f	1132r	V4		10.1371/journal.pone.0095567
80	Bik 2012	F04	R22	V1-V2		10.1111/j.1365-294X.2011.05297.x
81	Clerissi 2018	18SV1V2F	18SV1V2R	V1-V2	non-Metazoa	10.3389/fmicb.2018.02043
83	Hugerth 6	A-528F	1132r	V4		10.1371/journal.pone.0095567
84	Sisson 2018	SAR V3 F	SAR V3 R	V3	SAR	10.1093/plankt/fby020
86	Belevich 2017	EuF-V4	picoR2	V4	picoplankton	10.1007/s00248-017-1076-x
87	Michaud 2019a	Oxy 18S-F	Oxy 18S-R	V3-V4	Oxymonads	10.1111/mec.15322
88	Michaud 2019b	Par 18S-F	Par 18S-R	V5	Parabasalia	10.1111/mec.15322
89	Bradley 2016 V9	V8f	1510R	V8-V9		10.1128/AEM.01630-16
90	Bradley 2016 V4	TAReuk454FWD1	V4r	V4		10.1128/AEM.01630-16
92	Chemidlin 2011	FF390	FR1	V7-V8	fungi	10.1371/journal.pone.0024166
96	Stokes 2002	QPX-F	QPX-R2	V4	Quahog parasite	10.3354/dao052233
97	Stokes 2002	LABY-A	LABY-Y	V6	Labyrinthulomycetes	10.3354/dao052233
98	Fadev 2018	A-528F	V4RB	V4		10.3389/fmars.2018.00429
99	Xu 2020	A-528F	B-706R	V4		10.1016/j.pocean.2020.102309
100	Kilius 2013	A-528F	1055R	V4		10.1111/jpy.12109
101	Hu 2016	574*f	1132R modified	V4-V5		10.3389/fmicb.2016.00679
102	Piwoz 2019	TAReuk454FWD1	HaptoR1	V4	Haptophyta	10.1002/lno.11177

id	Name	Primer fwd	Primer rev	Region	Specificity	DOI
103	Emberg 2018	E572F	897r	V4		10.3354/meps12645
104	Choi 2020	TAReuk454FWD1	Choi TAReukREV3	Choi V4		10.1038/s41598-020-63561-z
106	Kim V9 2016	Nex 18S 1434 F	Nex 18S 1757 R	V8-V9		10.1111/1462-2920.13523
107	Huo 2020	960F	NSR1438	V7		10.1186/s12302-020-00321-w
108	Kataoka 2017	545F	1119R	V4		10.1093/femsec/fiw229
109	Li 2020	s14F3	17	37F-41F	foraminifera	10.1038/s41598-020-67221-0
110	Rachik 2018	18S-82F	Euk-516r	V2-V3		10.1371/journal.pone.0196987
119	Bass 2020	574*f	UNonMet DB	V4	non-Metazoa	10.1016/j.ejop.2020.125719

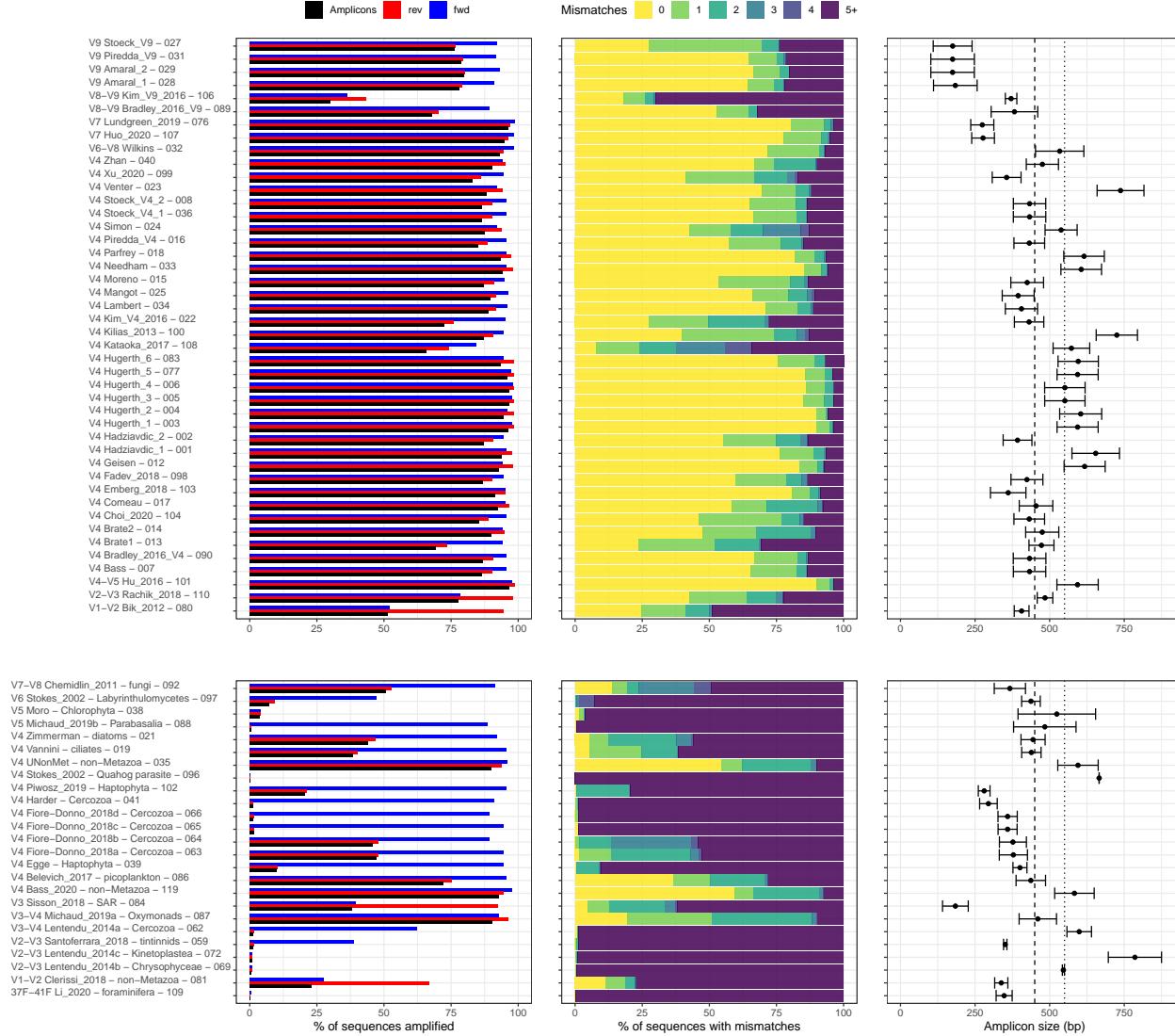


Figure S1: Evaluation of general (top) or specific (bottom) primer sets (Table S2) for the 18S rRNA gene against the PR² reference database (version 4.12.0). See Fig. 2 for legend.

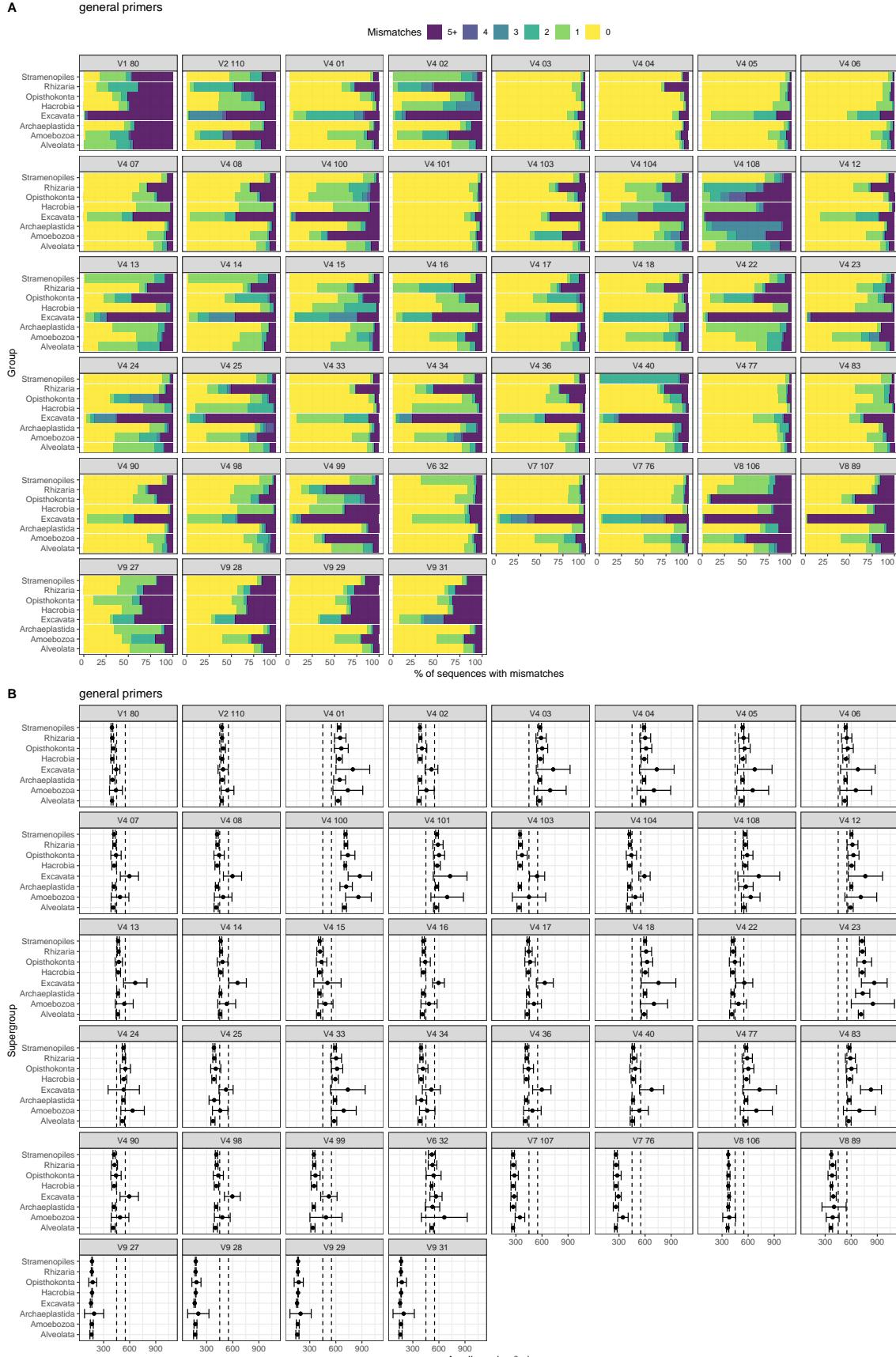


Figure S2: Number of mismatches (A) and amplicon size (mean \pm SD, B) for general primer sets as a function of the supergroup.

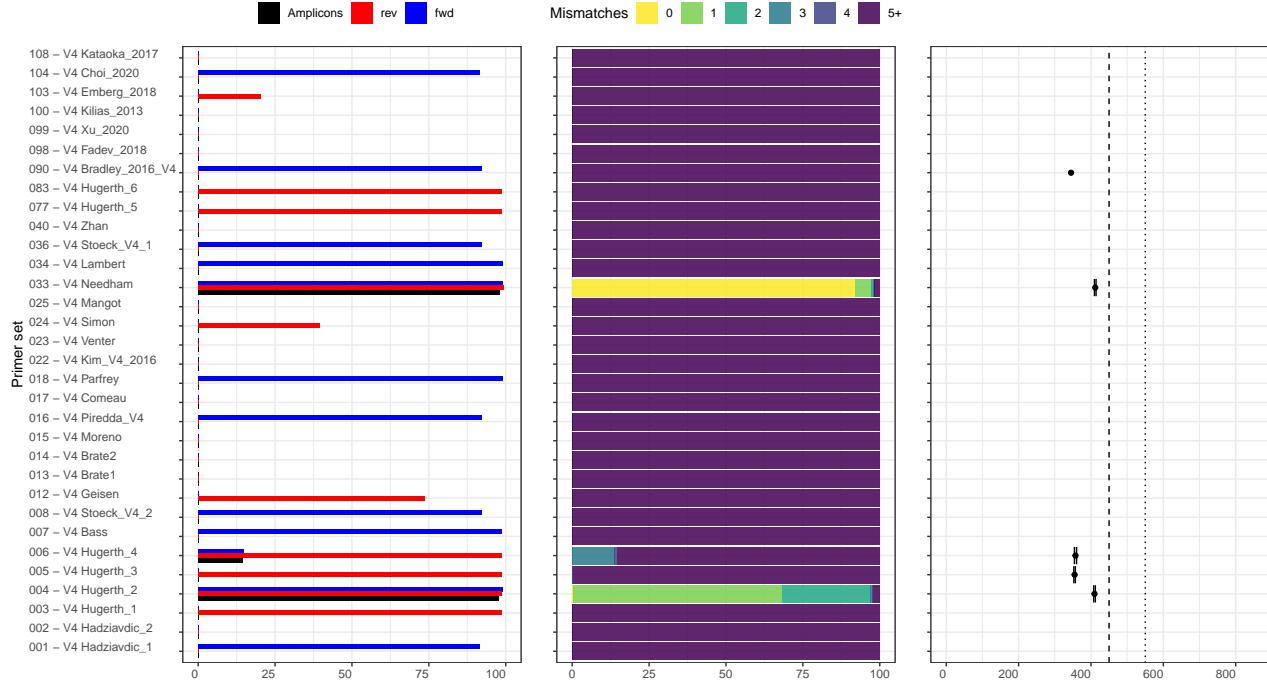


Figure S3: Evaluation of general primer sets (Table S2) targeting the V4 region of the 18S rRNA gene against bacterial 16S rRNA sequences from the Silva seed reference database (version 132). Legend as in Figure 2.

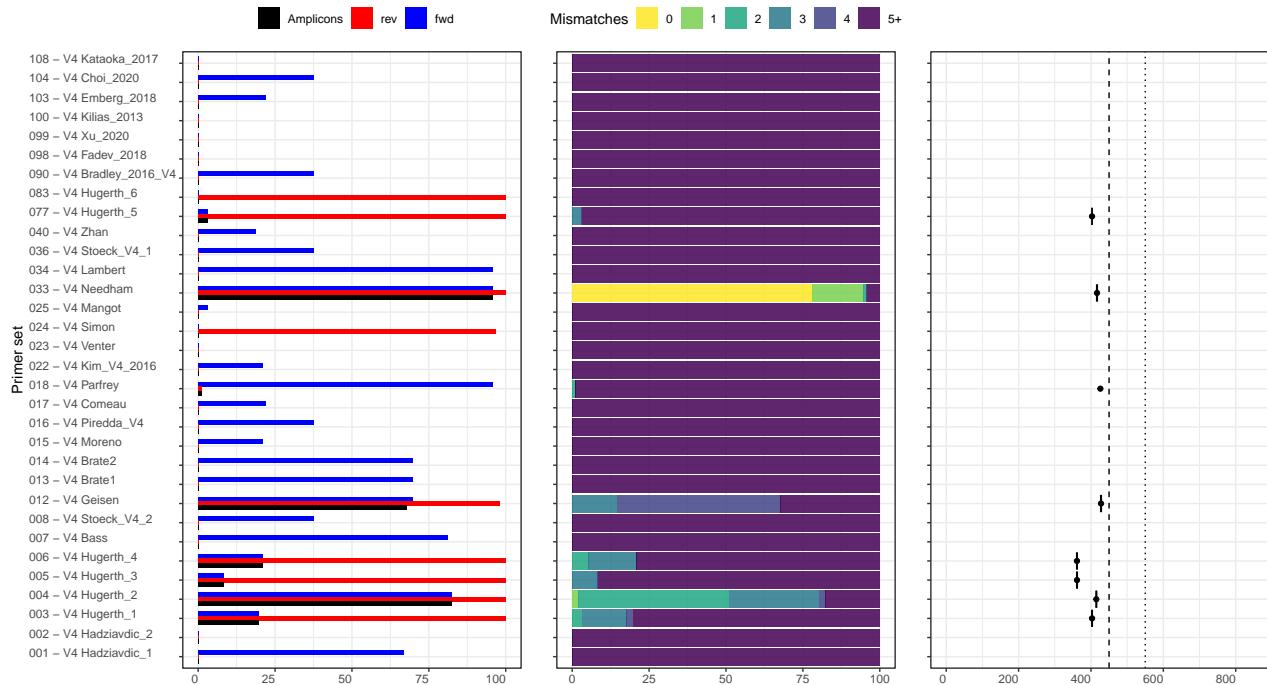
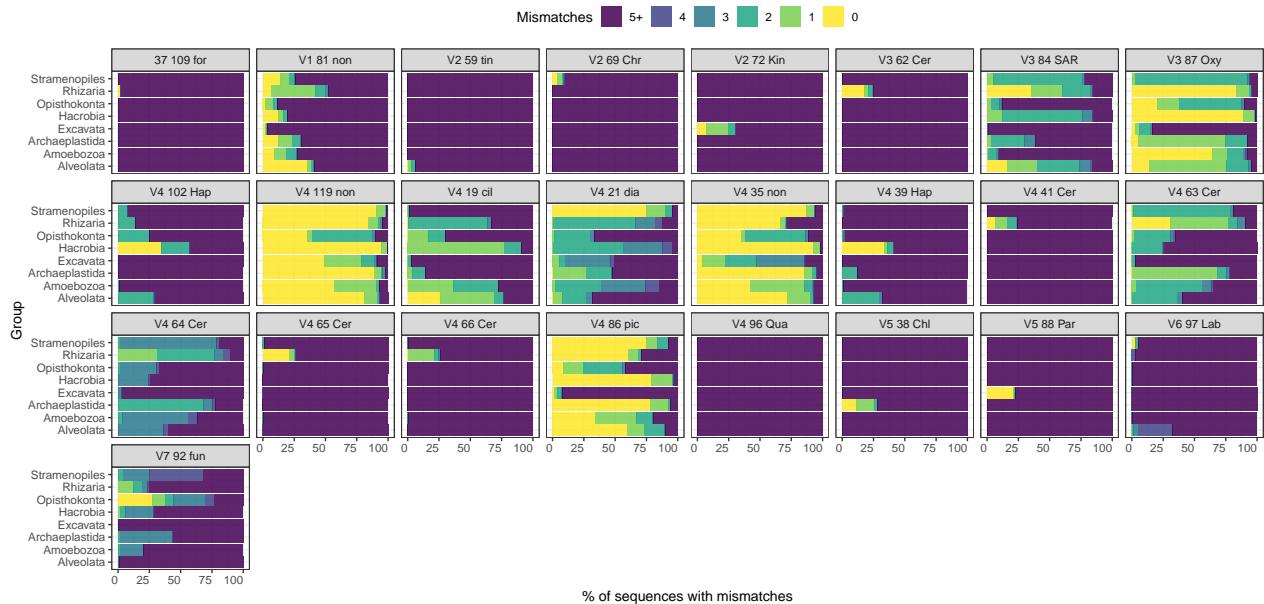


Figure S4: Evaluation of general primer sets (Table S2) targeting the V4 region of the 18S rRNA gene against archaeal 16S rRNA sequences from the Silva seed reference database (version 132). Legend as in Figure 2.

A specific primers



B specific primers

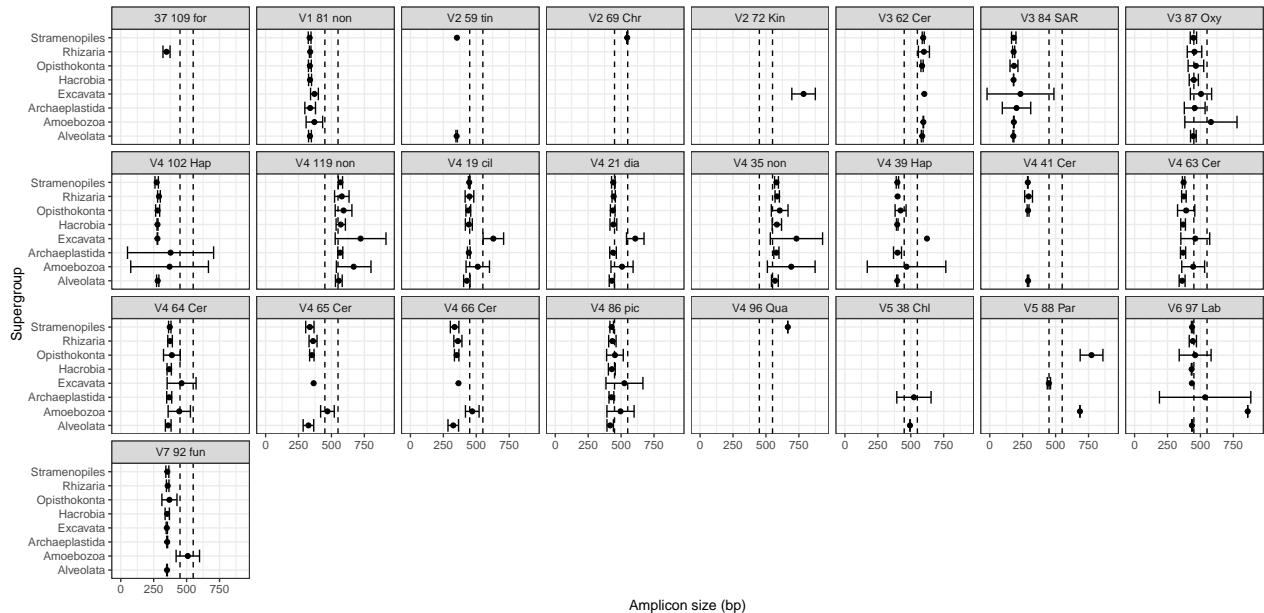


Figure S5: Number of mismatches (A) and amplicon size (mean \pm SD, B) for specific primer sets as a function of the supergroup.

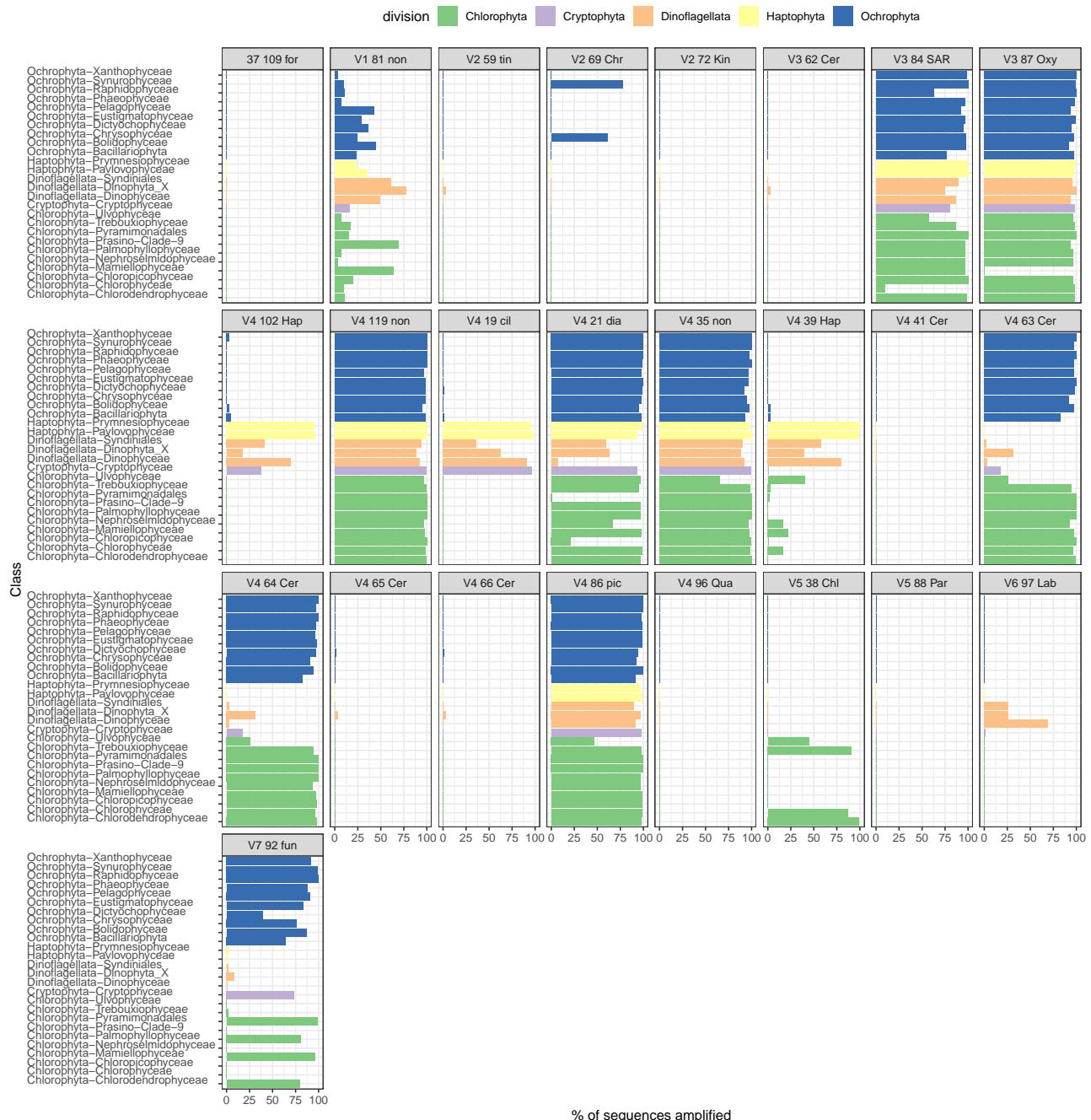


Figure S6: Percentage of sequences amplified with specific primer sets for different photosynthetic classes belonging to the Ochrophyta, Haptophyta, Dinoflagellata and Chlorophyta divisions.