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

Daniel Vaulot, Stefan Geisen, Frédéric Mahé, David Bass

▶ To cite this version:

Daniel Vaulot, Stefan Geisen, Frédéric Mahé, David Bass. pr2-primers: an 18S rRNA primer database for protists. Molecular Ecology Resources, 2022, 22 (1), pp.168-179. 10.1111/1755-0998.13465 . hal-03287553

HAL Id: hal-03287553

https://hal.sorbonne-universite.fr/hal-03287553

Submitted on 15 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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 PHIM, F-34398 Montpellier, France

[7] PHIM, Univ Montpellier, CIRAD, INRAE, Institut Agro, 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

Submitted date: July 8, 2021

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

Short title: pr2-primers database

Abstract

Metabarcoding of microbial eukaryotes (collectively known as protists) has developed tremendously in the last decade, almost solely relying on the 18S rRNA gene. As microbial eukaryotes are extremely diverse, many primers and primer pairs have been developed. To cover a relevant and representative fraction of the protist community in a given study system, an informed primer choice is necessary, as no primer pair can target all protists equally well. As such, a smart primer choice is very difficult even for experts and there are very few on-line resources available to list existing primers. We built a database listing 285 primers and 83 unique primer pairs that have been used for eukaryotic 18S rRNA gene metabarcoding. *In silico* performance of primer pairs was tested against two sequence databases: PR^2 version 4.12.0 for eukaryotes and a subset of SILVA version 132 for bacteria and archaea. We developed an R-based web application enabling browsing of the database, visualization of the taxonomic distribution of the amplified sequences with the number of mismatches, and testing any user-defined primer or primer set (https://app.pr2-primers.org). Taxonomic specificity of primer pairs, amplicon size and location of mismatches can also be determined. We identified universal primer sets that matched the largest number of sequences and analysed the specificity of some primer sets designed to target certain groups. This tool enables guided primer choices that will help a wide range of researchers to include protists as part of their investigations.

Introduction

Microbes are key players in all Earth ecosystems. Among them are protists that encompass all unicellular or unicellular-colonial eukaryotes, excluding some fungi. Protists perform a range of functions from photosynthesis to organic matter degradation. Although some eukaryotic groups such as unicellular algae (e.g. phytoplankton) have a long tradition of being studied as key players in marine primary production, the importance of protists in other processes and other environments has only been recently recognized, for example their role in nutrient cycling in soils or as symbionts and phagotrophs in marine waters (Geisen et al. 2018a; Worden et al. 2015). This late recognition stems in part from the inherent difficulties of visually identifying them and growing them in culture. In recent years, the development of metabarcoding has provided new tools to study protist diversity and ecology.

Metabarcoding is defined (Taberlet et al. 2012) as the use of a specific marker gene to analyse the composition of natural communities in a specific environment (water, soil, animal gut, faeces, etc...). After DNA extraction, the gene is amplified using a pair of primers targeting one specific region, samples are labelled with molecular tags and the resulting DNA is sequenced using a high throughput technology, mostly Illumina currently. This approach was initially developed for bacteria (Sogin et al. 2006) and expanded later for protists (Amaral-Zettler et al. 2009; Stoeck et al. 2009). The gene most commonly used is the small sub-unit ribosomal RNA gene (SSU rRNA: 16S rRNA for archaea and bacteria, 18S rRNA for eukaryotes). The SSU rRNA gene is composed of conserved regions that can be used to design general primers and variable regions (V) that can be used to assign taxonomy and design specific probes. In bacteria, the regions targeted are very often V3/V4 or V4/V5 (Parada et al. 2016), although other regions have been suggested as providing better resolution (e.g. Bukin et al. 2019). For eukaryotes, two variable regions of the 18S rRNA gene have mostly been targeted, the V4 and V9 regions: the V4 region is located in the second quarter of the 18S rRNA gene and the V9 region at the end of the 18S rRNA gene, near the internally transcribed spacer (ITS) region. Initially, the V9 region was favoured because of the limitation in sequence size (Amaral-Zettler et al. 2009; Stoeck et al. 2009): for example, initially Illumina sequences were restricted to 2x75 bp. However, with the development of the Illumina MiSeq (up to 2x300 bp), the V4 region is now preferred, in particular because it is longer, more variable, and better covered in reference databases (Pawlowski et al. 2012). Other eukaryotic genes, and in particular the mitochondrial cytochrome oxidase 1 gene (COI or cox1), have been used for Metazoa (Valentini et al. 2009) but their use is debated in particular because of the lack of universal primers (Andújar et al. 2018; Deagle et al. 2014) and the absence of this gene in lineages that have lost the mitochondrial genome (e.g. Yahalomi et al. 2020). For protists, the 18S rRNA gene appears to be most appropriate as a general marker (Pawlowski et al. 2012), although other genes such as rbcL (large unit of the RUBISCO) have been used for targeting photosynthetic organisms (e.g. Pujari et al. 2019).

Primer selection is critical to obtain an accurate taxonomic profiling of protist communities. Each primer (forward and reverse) must amplify the target community with minimal biases. The region amplified must be long enough to differentiate between closely related taxa by including enough variable positions. It should also be preferably short enough to be fully sequenced by the chosen technology, although longer amplicons can be also be partially sequenced. With Illumina sequencing being now the preferred technology, amplicon size must be ideally (although this is not absolutely necessary, see Lambert et al. 2019; Needham and Fuhrman 2016) about 50 bp smaller than the sum of the forward and reverse sequences (called R1 and R2) to allow enough overlap to reconstruct the complete amplicon: for example, the Illumina MiSeq 2x300 bp chemistry can sequence amplicons of up to 550 bp. A large diversity of primer and primer sets targeting the 18S rRNA gene have been developed over the years, although a few dominate in protist metabarcoding studies. Few resources are available that list eukaryotic 18S primers and primer pairs, provide information on their taxonomic specificity and allow testing of new primer pairs. Most existing primer databases do not focus on protists. For example, the primer database linked to the Barcode of Life Data System website Bold Systems (https://boldsystems.org/index.php/Public_Primer_PrimerSearch) focuses on metazoans, and Probebase (http://probebase.csb.univie.ac.at/node/8, Greuter et al. 2016) focuses on bacteria. A few programming tools have been developed to test primer set specificity, for example EcoPCR (Ficetola et al. 2010), a Python program, or R libraries such as PrimerMiner (Elbrecht and Leese 2017). The phylogenic program ARB offers a function to design and test probes and primers (Ludwig 2004). Unfortunately, these tools need to be installed in a specific computing environment and require some background programming skills. Many existing online tools such as Probematch (https://rdp.cme.msu.edu/probematch/search.jsp) only allow testing primer sets against bacteria, archaea and fungi. Silva TestPrime (https://www.arb-silva.de/search/testprime) is the only tool that covers protists. It provides very detailed feedback on the taxonomy of amplified sequences, and the location of mismatches. Such detailed information comes at the expense of speed, with a typical test needing a few minutes to run. Moreover, the taxonomic annotation of the Silva database for protists is not optimal at this time, particularly for environmental sequences which are often only assigned at the class level or above (typically "Chrysophyceae;uncultured;eukaryotic picoplankton environmental sample").

To fill this gap and to provide protist researchers with a usable tool, we constructed a database of primers and primer sets used for eukaryotic 18S rRNA metabarcoding. These primer sets were tested *in silico* against the PR^2 database (Guillou et al. 2013) that contains more than 180,000 18S rRNA sequences with expert taxonomical annotation and a subset of the Silva database for archaea and bacteria. We developed an R-based web application that allows exploration of the database, to visualize pre-computed *in silico* amplification results according to taxonomy (% of amplification, size of amplicons and location of mismatches), and to test any user-defined primer set.

Material and Methods

18S rRNA gene primers (Table S1) and primer sets (Table S2) used in metabarcoding studies were collected from the literature. Primer sequences and primer sets (knowing that several primer sets may share at least one primer) were stored in a MySQL database. Primer sets were tested by performing *in silico* amplification of eukaryotic sequences stored in the PR^2 reference database (Guillou et al. 2013) version 4.12.0 (https://github.com/pr2database/pr2database/releases/tag/v4.12.0). We also used a small subset of the Silva database version 132 provided by the mothur website (https://mothur.org/wiki/silva_reference_files) containing 8517 bacteria and 147 archaea sequences to test whether these two groups were amplified. Database sequences with ambiguities were discarded (any nucleotide that is not A, C, G or T). Sequences with length shorter than 1350 bp were not considered except for the V4 region, for which this threshold was lowered to 1200 bp, since most sequences in PR^2 contain the V4 region. In contrast, this limit was extended to 1650 for the V9 region and since many 18S rRNA do not cover the full V9 region, we only kept sequences that contained the canonical sequence GGATC[AT] which is located at the end of the V9 region, just before the start of the internally transcribed spacer 1 (ITS1). An R (R Development Core Team 2013) script using the *Biostrings* package (Pagès et al. 2020) was used to compute the number of mismatches to the forward and reverse primers, allowing for a maximum of 2 mismatches for each primer using the function *matchPattern* with the following parameters: max.mismatch=2, min.mismatch=0, with.indels=FALSE, fixed=FALSE, algorithm="auto". We computed the position of mismatches using the *mismatch* function with parameter fixed=FALSE. A faster version of the script is also available that does not compute mismatch position using the vectorized form of the *matchPattern* function (*vmatchPattern*). The latter function is used in the Shiny application (see below) allowing users to test their own primer or primer sets. The data were tabulated using the *dplyr* package and plotted using the *ggplot2* package (Wickham 2016). An R shiny application to interact with the database was developed using the following R packages: *shiny*, *shinyFeedback* and *shinycssloaders* (Sali and Attali

2020).

Results and Discussion

Database of primers and primer sets

We were able to recover a total of 108 general eukaryotic primers and 177 primers specific to some taxonomic groups from the literature (Tables 1 and S1, https://app.pr2-primers.org). Some of these primers were designed early on when researchers began to amplify and sequence the 18S rRNA gene (e.g. Medlin et al. 1988). More recently, researchers have been designing primers specific to some taxonomic groups, mostly targeting phylum level (e.g. S19F and S15rF for Foraminifera, Morard et al. 2011) or class level (e.g. primer PRYM03+3 for Prymnesiophyceae, Egge et al. 2013). Some primers were also designed to block specific taxa (e.g. 18SV1V2Block against the coral *Pocillopora damicornis*, Clerissi et al. 2018) to be used in combination with more general primers (18SV1V2F in this case) or to avoid amplification of some groups (e.g. EUK581-F and EUK1134-R which do not amplify Metazoa, Carnegie et al. 2003). "Anti-metazoan" primers are used when looking at the eukaryotic microbiome of eukaryotic organisms (e.g. corals, oysters) to avoid amplification of host's genes (Bass and del Campo 2020).

We identified a total of 83 unique primer sets (pairs) that have been used in metabarcoding studies (Table S2). Not all primers have been used for metabarcoding, in particular those that amplify the whole 18S rRNA gene, such as EukA and EukB (Medlin et al. 1988). Most metabarcoding primer sets do not target specific groups. The localization of these primer sets over the 18S rRNA gene is quite diverse, but the vast majority target the V4 region (Table 2 and Fig. 1). In contrast, the number of primer sets targeting the other favoured metabarcoding region V9 is much lower. Most of the primer sets targeting a specific taxonomic group are located in the V4 region, and none are in the V9 region (Table 2). In terms of usage, the V4 region is much more popular (about 80% of published studies in marine systems, Lopes Dos Santos et al. 2022), the three most commonly used primer sets being # 8 (TAReuk454FWD1 and TAReukREV3, Stoeck et al. 2010), # 17 (E572F and E1009R, Comeau et al. 2011) and # 16 (TAReuk454FWD1 and V4 18S Next.Rev, Piredda et al. 2017), while for the V9 region the most popular sets are # 27 (1391F and EukB, Stoeck et al. 2010) and # 28 (1380F and 1510R, Amaral-Zettler et al. 2009).

Testing primer sets by *in silico* matching

General primer sets

We used the PR^2 database (Guillou et al. 2013) which currently contains about 180,000 18S rRNA sequences with detailed taxonomic annotations to test all primer sets from the pr2-primers database. We also determined, using a set of more than 8,500 sequences representative of diverse archaeal and bacterial groups, whether these primers amplified bacteria or archaea. We only used long sequences (see Material and Methods) and allowed for a maximum of 2 mismatches on both forward and reverse primers, i.e. a maximum of 4 mismatches. For general primers, amplification success varied from 32 to more than 97% (Table S3, Fig. 2 and S1). In general, the reverse primer had a tendency to have more mismatches than the forward primer (Table S3). Primer sets targeting regions other than V4 or V9 did not perform as well in general (Fig. S1), although the best overall performance was for # 76 targeting the V7 region (F-1183 and R-1443, 97.1% of sequences amplified, Lundgreen et al. 2019). If we focus on the V4 and V9 regions (Fig. 2), the best performing primer sets overall were # 6 (616*f and 1132r, 96.5%, Hugerth et al. 2014) and # 29 (1389F and 1510R, 79.8%, Amaral-Zettler et al. 2009). Interestingly, the original paper describing this primer set also used another forward primer (1380F, primer set # 28) on the same samples and recommended using both forward primers together, an advice which was not followed in subsequent studies (but see Lie et al. 2014). The lower percentage observed for the V9 primers should be interpreted with caution: many 18S reference sequences do not extend to the end of the V9 region and therefore will miss the signature of the reverse primer. To minimize this problem, we retained for the analysis of V9 primer sets only sequences that contain the canonical signature GGATC[AT] located at the 3' end of the V9 region. Despite performing well when allowing for 4 mismatches, some of these primer sets have at least one mismatch to PR^2 sequences: for example, primer set # 108 (545F and 1119R, Kataoka et al. 2017) amplifies only 7.9% of the sequences with zero mismatch. Another important consideration is the size of the amplicon. Since most metabarcoding studies currently use Illumina sequencing technology, the maximum possible size to allow some overlap between the two R1 and R2 reads is about 550 bp (assuming that one uses the 2x300 bp sequencing kits), although smaller amplicons are preferable to allow more overlap. A sizeable fraction of the primer sets produce amplicons close to or larger than 600 bp (Fig. 2). The post sequencing analysis strategy in this case would be to only use one of the reads (R1 is in general less noisy) without trying to assemble R1 and R2 (Lambert et al. 2019) or to assemble the non-overlapping R1 and R2 reads with an intercalated N base (Needham and Fuhrman 2016).

Another important consideration is whether amplification is similar across the whole eukaryotic taxonomic range. Taking as an example the most frequently used primer set targeting V4 (# 8, Fig. 3A) and looking at the amplification efficiency at the supergroup level, a significant fraction of Excavata and to a smaller extent of Rhizaria present at least 5 mismatches to this primer set (Fig. 3A top-left). Amplification is even more unlikely for sequences presenting mismatches with the forward primer because the mismatches are located at the 3' end of the primer (Fig. 3A top-right) which is the most unfavourable situation (mismatches at the 5' end are better tolerated). The average size of the amplicon also varies depending on the taxonomic group (Fig. 3B bottom). For example, Excavata have on average longer amplicons, in particular because of the presence of introns (Torres-Machorro et al. 2010). Amplicon size is then beyond the current range of Illumina sequencing. This may also induce negative bias during PCR amplification (Geisen et al. 2015). For other groups such as Opisthokonta, although the average size is compatible with Illumina sequencing, there is a large number of outlier sequences with long amplicons. This will mean that taxa corresponding to these sequences (mostly Arthropoda) will be missed from surveys conducted with this primer set, although of course this is less critical when protists are targeted. The situation with the V9 primer set # 27 (Fig. 3) is somewhat similar, although there is less length variation between the different supergroups. However, for some groups, in particular Ascomycota and Bangiophyceae, there is a number of outliers that will be missed by Illumina sequencing. Again, these groups are less relevant when focusing on protists. When looking at all the general primer sets (Figure S2), some sets such as # 2, 25, and 110 appear to have more taxonomic biases than others. Overall, Excavata constitute the supergroup that is most often discriminated against.

Most primer sets will not amplify archaea and bacteria, except primer sets such as # 33 (515F and Univ 926R Needham and Fuhrman 2016) that were specifically designed to amplify both bacteria and eukaryotes (Figs. S3 and S4). However, some primer sets assumed to be specific to eukaryotes such as # 4 (563f and 1132r, Hugerth et al. 2014) amplifies quite well archaea and bacteria. Interestingly, set # 12 (3NDf and 1132rmod, Geisen et al. 2018b) amplify only eukaryotes and archaea, but not bacteria. In most cases we tested, the reverse primer was most discriminating against archaea and bacteria.

Specific primer sets

In order to access a deeper diversity within a given taxonomic group primer sets have been developed with specific targets (Tables S1 and S2). Target levels are most often at the division (e.g. Haptophyta) and class levels (e.g. Chrysophyceae), although some sets are targeting supergroups (e.g. SAR

84). Some primer sets have even more specific targets. One example is primer # 65 targeting Cercozoa (S616F Cerco and S947R Cerco, Fiore-Donno et al. 2018) that contains at least 5 mismatches to all other divisions (Fig. S5) and amplifies all cercozoan groups. Primer # 38 targeting Chlorophyta (ChloroF and ChloroR, Moro et al. 2009) contains at least 5 mismatches to all other divisions (Fig. S5). However, it is does not amplify all Chlorophyta as it misses picoplanktonic green algae such as Mamiellophyceae or Chloropicophyceae (Fig. S6). In contrast, several primer sets claimed to be specific of a given group are in fact quite general. For example set # 87 which targets oxymonads (Oxy 18S-F and Oxy 18S-R Michaud et al. 2020) amplifies many other groups (Figs. S1 and S5). In this case, this is not critical since oxymonads only occur in termite guts and such primers will only be used in this specific context. Primer set # 21 (D512for and D978rev, Zimmermann et al. 2011) which was designed to target diatoms would amplify actually most of the Ochrophyta classes but also some green algae (Fig. S6).

R Shiny application

We have developed a website based on an R Shiny application (https://app.pr2-primers.org) that allows users to visualize and download the pr2-primers database, explore at different taxonomic levels the results of *in silico* amplification against the PR^2 and Silva databases for the primer sets from the pr2-primers database and test their own primer sets. The application is composed of 7 panels. The first panel (Fig. 4A) provides information on the database as well as a link to report issues or new primers. The second and third panels (Fig. 4B) provide an interface to the primer and primer set tables, respectively, with the options of downloading the tables and revealing/hiding specific columns. The fourth and fifth panels are used to display the results of pre-computed *in silico* amplification of primer sets from the database. The fourth panel (Fig. 4C) shows a synthesis of the results (similar to Fig. 2) for all primer sets. The fifth panel (Fig. 5A) is a tool to explore amplification properties of a given primer set within a taxonomic level from kingdom to class levels. The right-hand section of this panel shows general amplification characteristics, the location of the mismatches, the number of mismatches for each group and the distribution of the amplicon sizes. Finally, the sixth and seventh panels (Fig. 5B) allow users to run an *in silico* amplification with their own primers/probes (panel 6) and primer sets (panel 7) against PR^2 and Silva seed databases. Users can fix the maximum number of mismatches (up to 2 for each primer). For the sake of speed, only the number of mismatches is provided, not their position. Global statistics on the amplification are provided, which can be explored at different taxonomic levels. The R shiny application has been incorporated into a Docker container available at https://hub.docker.com/repository/docker/vaulot/pr2-primers.

Conclusion

The combination of the pr2-primers database with the PR^2 sequence database provides a very useful resource for protist metabarcoding. It will help researchers to select the most suitable primer pairs for both broadly-targeted surveys and studies focusing on target taxonomic groups, and to test and validate *in silico* novel primers. We emphasize that primer pairs must also be tested on reference culture material and natural samples, as actual amplification may differ from *in silico* results. Hopefully this database will grow with time as novel primer pairs are developed and tested on samples from a range of environments. This will contribute to better design and comparability of microbiome analyses, inventories of protist diversity across environments, and increase our understanding of this functionally diverse and important group of organisms.

References cited

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE*, *4*, e6372. https://doi.org/10.1371/journal.pone.0006372

Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the Metazoa. *Molecular Ecology*, *27*, 3968–3975. https://doi.org/10.1111/mec.14844

Bass, D., & del Campo, J. (2020). Microeukaryotes in animal and plant microbiomes: Ecologies of disease? *European Journal of Protistology*, *76*, 125719. https://doi.org/10.1016/j.ejop.2020.125719

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., & Zemskaya, T. I. (2019). The effect of 16s rRNA region choice on bacterial community metabarcoding results. *Scientific Data*, *6*, 190007. https://doi.org/10.1038/sdata.2019.7

Carnegie, R. B., Meyer, G. R., Blackbourn, J., Cochennec-Laureau, N., Berthe, F. C., & Bower, S. M. (2003). Molecular detection of the oyster parasite *Mikrocytos mackini*, and a preliminary phylogenetic analysis. *Diseases of Aquatic Organisms*, *54*, 219–227. https://doi.org/10.3354/dao054219

Clerissi, C., Brunet, S., Vidal-Dupiol, J., Adjeroud, M., Lepage, P., Guillou, L., Escoubas, J. M., & Toulza, E. (2018). Protists within corals: The hidden diversity. *Frontiers in Microbiology*, *9*, 2043. https://doi.org/10.3389/fmicb.2018.02043

Comeau, A. M., Li, W. K., Tremblay, J. É., Carmack, E. C., & Lovejoy, C. (2011). Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE*, *6*, e27492. https://doi.org/10.1371/journal.pone.0027492

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, *10*, 20140562. https://doi.org/10.1098/rsbl.2014.0562

Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C., & Edvardsen, B. (2013). 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: A test for marine haptophytes. *PLoS ONE*, *8*, e74371. https://doi.org/10.1371/journal.pone.0074371

Elbrecht, V., & Leese, F. (2017). PrimerMiner: An r package for development and *in silico* validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*, 622–626. https://doi.org/10.1111/2041-210X.12687

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., & Pompanon, F. (2010). An *In silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*, 434. https://doi.org/10.1186/1471-2164-11-434

Fiore-Donno, A. M., Rixen, C., Rippin, M., Glaser, K., Samolov, E., Karsten, U., Becker, B., & Bonkowski, M. (2018). New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Molecular Ecology Resources*, *18*, 229–239. https://doi.org/10.1111/1755-0998.12729

Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., & De Groot, G. A. (2015). Not all are free-living: High-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Molecular Ecology*, *24*, 4556–4569. https://doi.org/10.1111/mec.13238

Geisen, S., Mitchell, E. A., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L. D., Jousset, A., Krashevska, V., Singer, D., Spiegel, F. W., Walochnik, J., & Lara, E. (2018a). Soil protists: A fertile frontier in soil biology research. *FEMS Microbiology Reviews*, *42*, 293–323. https://doi.org/10.1093/femsre/fuy006

Geisen, S., Snoek, L. B., ten Hooven, F. C., Duyts, H., Kostenko, O., Bloem, J., Martens, H., Quist, C. W., Helder, J. A., & van der Putten, W. H. (2018b). Integrating quantitative morphological and qualitative molecular methods to analyse soil nematode community responses to plant range expansion. *Methods in Ecology and Evolution*, *9*, 1366–1378. https://doi.org/10.1111/2041-210X.12999

Greuter, D., Loy, A., Horn, M., & Rattei, T. (2016). ProbeBase-an online resource for rRNA-targeted oligonucleotide probes and primers: New features 2016. *Nucleic Acids Research*, *44*, D586–D589. https://doi.org/10.1093/nar/gkv1232

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The Protist Ribosomal Reference database (PR^2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*, D597–D604. https://doi.org/10.1093/nar/gks1160

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 eukaryotic diversity in microbial consortia. *PLoS ONE*, *9*, e95567. https://doi.org/10.1371/journal.pone.0095567

Kataoka, T., Yamaguchi, H., Sato, M., Watanabe, T., Taniuchi, Y., Kuwata, A., & Kawachi, M. (2017). Seasonal and geographical distribution of near-surface small photosynthetic eukaryotes in the western North Pacific determined by pyrosequencing of 18S rDNA. *FEMS microbiology ecology*, *93*, fiw229. https://doi.org/10.1093/femsec/fiw229

Lambert, S., Tragin, M., Lozano, J. C., Ghiglione, J. F., Vaulot, D., Bouget, F. Y., & Galand, P. E. (2019). Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. *ISME Journal*, *13*, 388–401. https://doi.org/10.1038/s41396-018-0281-z

Lie, A. A., Liu, Z., Hu, S. K., Jones, A. C., Kim, D. Y., Countway, P. D., Amaral-Zettler, L. A., Cary, S. C., Sherr, E. B., Sherr, B. F., Gast, R. J., & Caron, D. A. (2014). Investigating microbial eukaryotic diversity from a global census: Insights from a comparison of pyrotag and full-length sequences of 18S rRNA genes. *Applied and Environmental Microbiology*, *80*, 4363–4373. https://doi.org/10.1128/AEM.00057-14

Lopes Dos Santos, A., Ribeiro Gérikas, C., Ong, D., Garczarek, L., Shi, X. L., Nodder, S., Vaulot, D., & Gutierrez-Rodriguez, A. (2022). Phytoplankton diversity and ecology through the lens of high throughput sequencing technologies, In *Advances in Phytoplankton Ecology. Applications of emerging technologies*. Elsevier.

Ludwig, W. (2004). ARB: A software environment for sequence data. *Nucleic Acids Research*, *32*, 1363–1371. https://doi.org/10.1093/nar/gkh293

Lundgreen, R. B., Jaspers, C., Traving, S. J., Ayala, D. J., Lombard, F., Grossart, H. P., Nielsen, T. G., Munk, P., & Riemann, L. (2019). Eukaryotic and cyanobacterial communities associated with marine snow particles in the oligotrophic Sargasso Sea. *Scientific Reports*, *9*, 8891. https://doi.org/10.1038/s41598-019-45146-7

Medlin, L., Elwood, H. J., Stickel, S., & Sogin, M. L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene*, *71*, 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Michaud, C., Hervé, V., Dupont, S., Dubreuil, G., Bézier, A. M., Meunier, J., Brune, A., & Dedeine, F. (2020). Efficient but occasionally imperfect vertical transmission of gut mutualistic protists in a wood-feeding termite. *Molecular Ecology*, *29*, 308–324. https://doi.org/10.1111/mec.15322

Morard, R., Quillévéré, F., Douady, C. J., de Vargas, C., de Garidel-Thoron, T., & Escarguel, G. (2011). Worldwide genotyping in the planktonic foraminifer *Globoconella inflata*: Implications for life history and paleoceanography. *PLoS ONE*, *6*, e26665. https://doi.org/10.1371/journal.pone.0026665

Moro, C. V., Crouzet, O., Rasconi, S., Thouvenot, A., Coffe, G., Batisson, I., & Bohatier, J. (2009). New design strategy for development of specific primer sets for PCR-based detection of Chlorophyceae and Bacillariophyceae in environmental samples. *Applied and Environmental Microbiology*, *75*, 5729–5733. https://doi.org/10.1128/AEM.00509-09

Needham, D. M., & Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature microbiology*, *1*, 16005. https://doi.org/10.1038/nmicrobiol.2016.5

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2020). *Biostrings: Efficient manipulation of biological strings* (https://bioconductor.org/packages/release/bioc/html/Biostrings.html).

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., … de Vargas, C. (2012). CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, *10*, e1001419. https://doi.org/10.1371/journal.pbio.1001419

Piredda, R., Tomasino, M. P., D'Erchia, A. M., Manzari, C., Pesole, G., Montresor, M., Kooistra, W. H., Sarno, D., & Zingone, A. (2017). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiology Ecology*, *93*, fiw200. https://doi.org/10.1093/femsec/fiw200

Pujari, L., Wu, C., Kan, J., Li, N., Wang, X., Zhang, G., Shang, X., Wang, M., Zhou, C., & Sun, J. (2019). Diversity and spatial distribution of chromophytic phytoplankton in the Bay of Bengal revealed by RuBisCO Genes (rbcL). *Frontiers in Microbiology*, *10*, 1–17. https://doi.org/10.3389/fmicb.2019.01501

R Development Core Team. (2013). R: A language and environment for statistical computing. *R foundation for statistical computing*, *1*, 409. https://doi.org/10.1007/978-3-540-74686-7

Sali, A., & Attali, D. (2020). *Shinycssloaders: Add loading animations to a 'shiny' output while it's recalculating* (https://CRAN.R-project.org/package=shinycssloaders).

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 12115–12120. https://doi.org/10.1073/pnas.0605127103

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, *19*, 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M. J., Chistoserdov, A., Orsi, W., & Edgcomb, V. P. (2009). Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biology*, *7*, 72. https://doi.org/10.1186/1741-7007-7-72

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Torres-Machorro, A. L., Hernández, R., Cevallos, A. M., & López-Villaseñor, I. (2010). Ribosomal RNA genes in eukaryotic microorganisms: Witnesses of phylogeny? *FEMS Microbiology Reviews*, *34*, 59–86. https://doi.org/10.1111/j.1574-6976.2009.00196.x

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, *24*, 110–117. https://doi.org/10.1016/j.tree.2008.09.011

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer International Publishing.

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, *347*, 1257594. https://doi.org/10.1126/science.1257594

Yahalomi, D., Atkinson, S. D., Neuhof, M., Sally Chang, E., Philippe, H., Cartwright, P., Bartholomew, J. L., & Huchon, D. (2020). A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 5358–5363. https://doi.org/10.1073/pnas.1909907117

Zimmermann, J., Jahn, R., & Gemeinholzer, B. (2011). Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity and Evolution*, *11*, 173–192. https://doi.org/10.1007/s13127-011-0050-6

Data availability

No new data were created or analysed in this study. All scripts, including those for the Shiny application, are available at https://github.com/pr2database/pr2-primers (doi:10.5281/zenodo.4849528). The database is available at https://app.pr2-primers.org.

Acknowledgements

We thank the ABIMS platform of the FR2424 (CNRS, Sorbonne Université) for bioinformatics resources. The authors declare no conflicts of interest.

Author contributions statement

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

Additional information

Competing interests. The authors declare no competing financial interests.

List of Tables

List of Figures

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. Colours correspond to taxonomy (division). Hacrobia represents the sum of haptophytes, cryptophytes and centrohelids.

Figure 4 Shiny interface to the pr2-primers database. A. First panel introducing the database. Numbers in red correspond to the different panels. B. Second panel displaying the list of primers. The third panel is analogous, but for primer sets. C. Fourth panel showing *in silico* amplification results for all pre-computed primer sets.

Figure 5 Shiny interface to the pr2-primers database. A. Fifth panel introducing showing *in silico* detailed amplification results for a given pre-computed primer set. Taxonomy can be explored in detail. B. Seventh panel displaying *in silico* amplification results for user-provided primer sets. The sixth panel is identical but for a single primer or probe.

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

Figure S2 Number of mismatches 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.

Figure S5 Number of mismatches for specific primer sets as a function of the supergroup. Target group is indicated inside the corresponding supergroup bar (e.g., Foraminifera are inside Cercozoa).

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

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

direction	general	specific
fwd	55	89
rev	53	88
total	108	177

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

gene region	general	specific
37F		1
37F-41F		2
V1-V2	1	1
V1-V3		1
V2		3
V2-V3	1	3
V3		1
V3-V4		2
V4	32	15
V4-V5	1	
V5		3
V5-V7	1	
V5-V9		2
V6		1
V6-V8	1	
V7	2	
V7-V8		1
V7-V9	1	1
V8-V9	2	
V9	4	

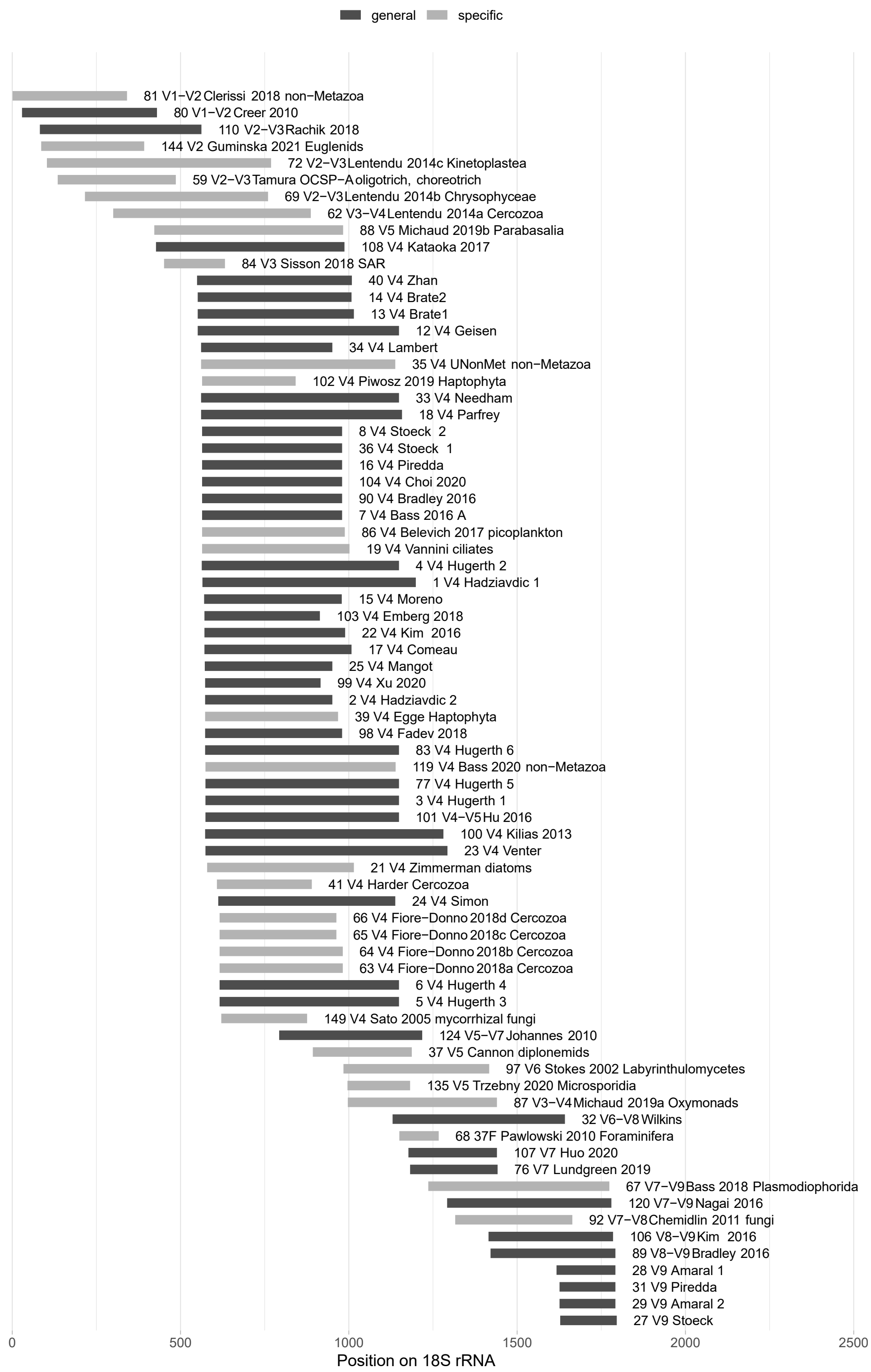

Figure 1: Position of the amplified region when using different primer sets listed in the pr2-primers database along the 18S RNA gene relative to the sequence of the yeast *Saccharomyces cerevisiae* (FU970071). The labels correspond to the primer set id, the 18S region amplified, its identification name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

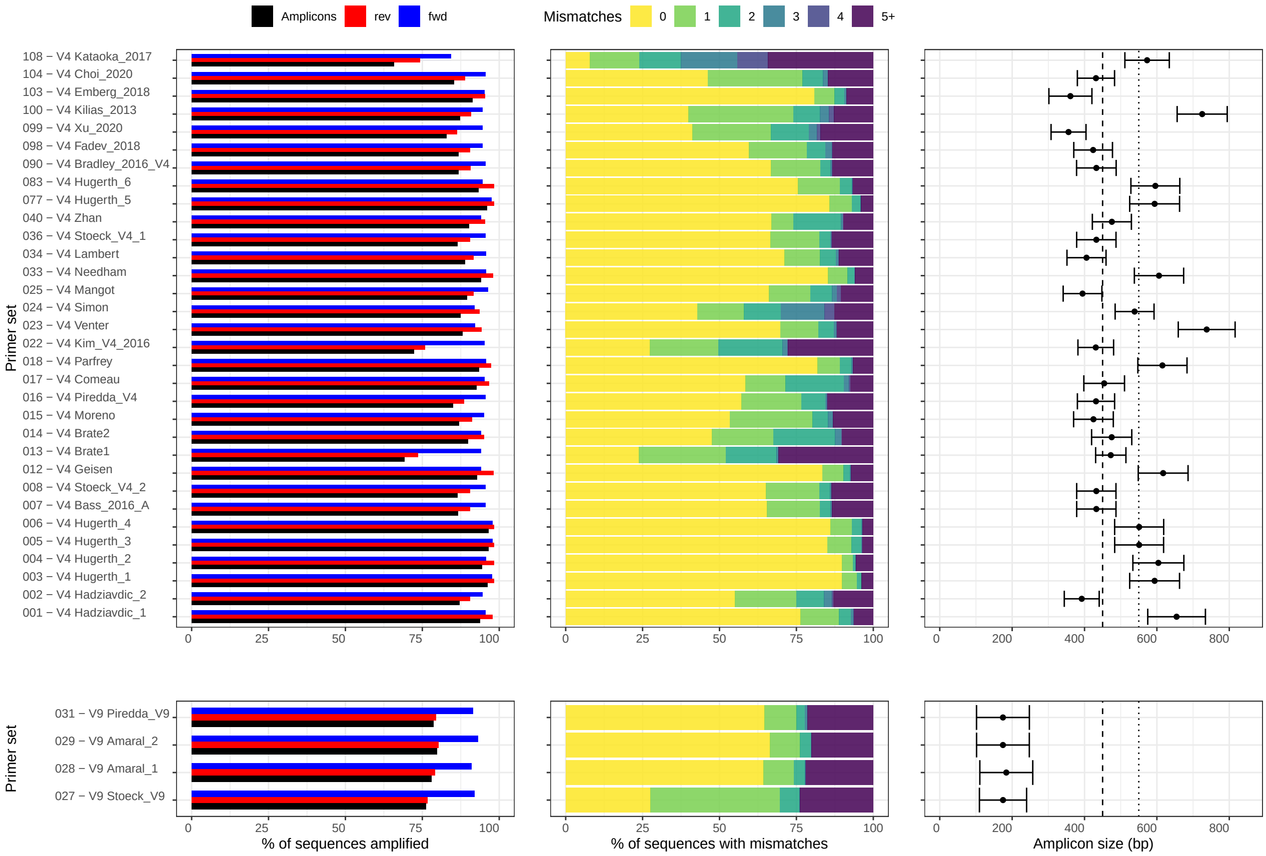

Figure 2: Evaluation of general primer sets (Table S2) targeting the V4 (top) and V9 (bottom) regions of the 18S rRNA gene against the PR^2 reference database (version 4.12.0). Left panel. Percentage of reference sequences with at most 2 mismatches to either forward and reverse primer or to both primers, 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. See Fig. S1 for the complete set of primer sets.

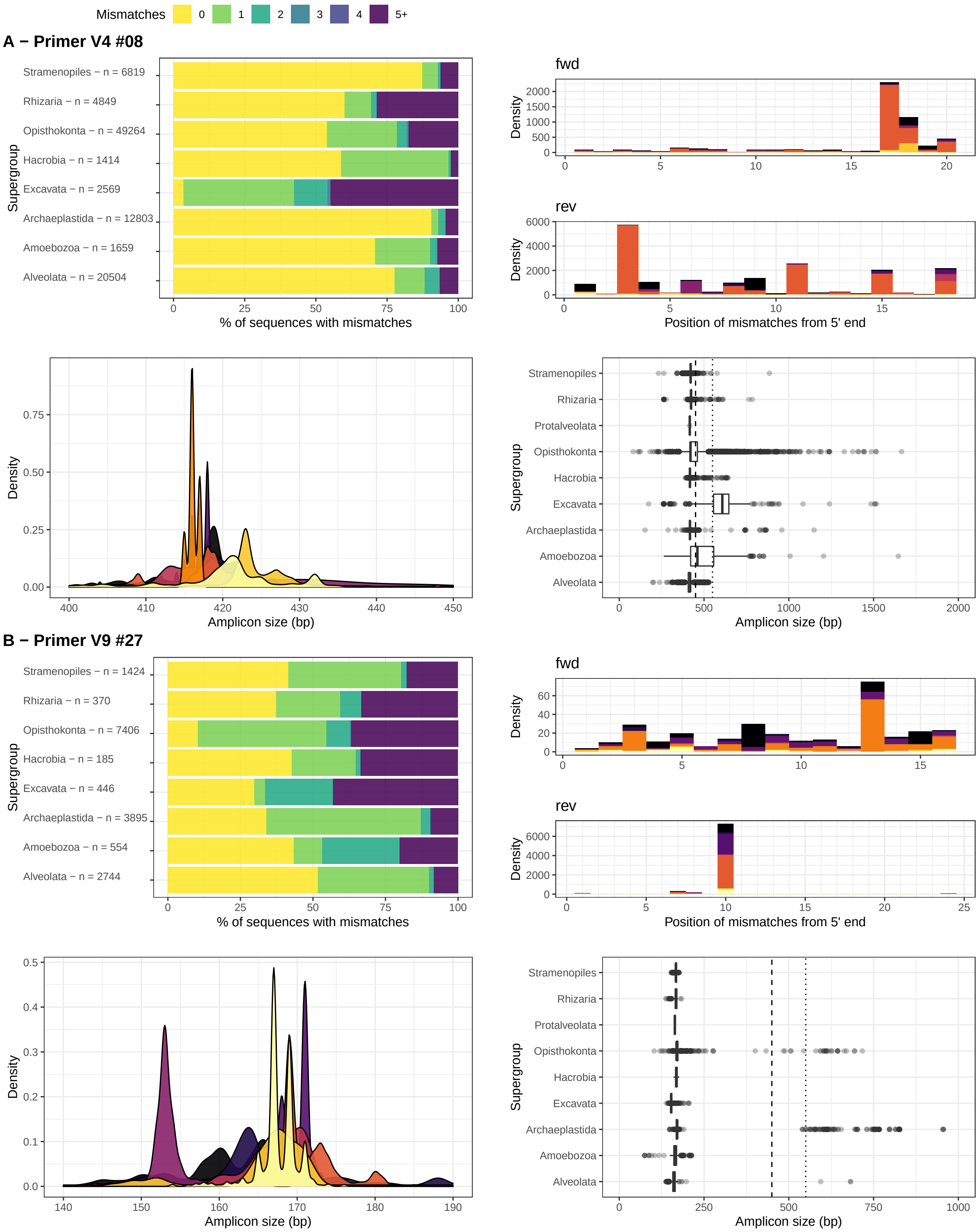

Figure 3: Example of analysis for two primer sets amplifying two regions of the 18S rRNA gene: V4 (primer set # 8, A) and V9 (primer set # 27, B). Top left. Percentage of sequences with a given number of mismatches. Top right. Position of the mismatches for different taxonomic supergroups on the forward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

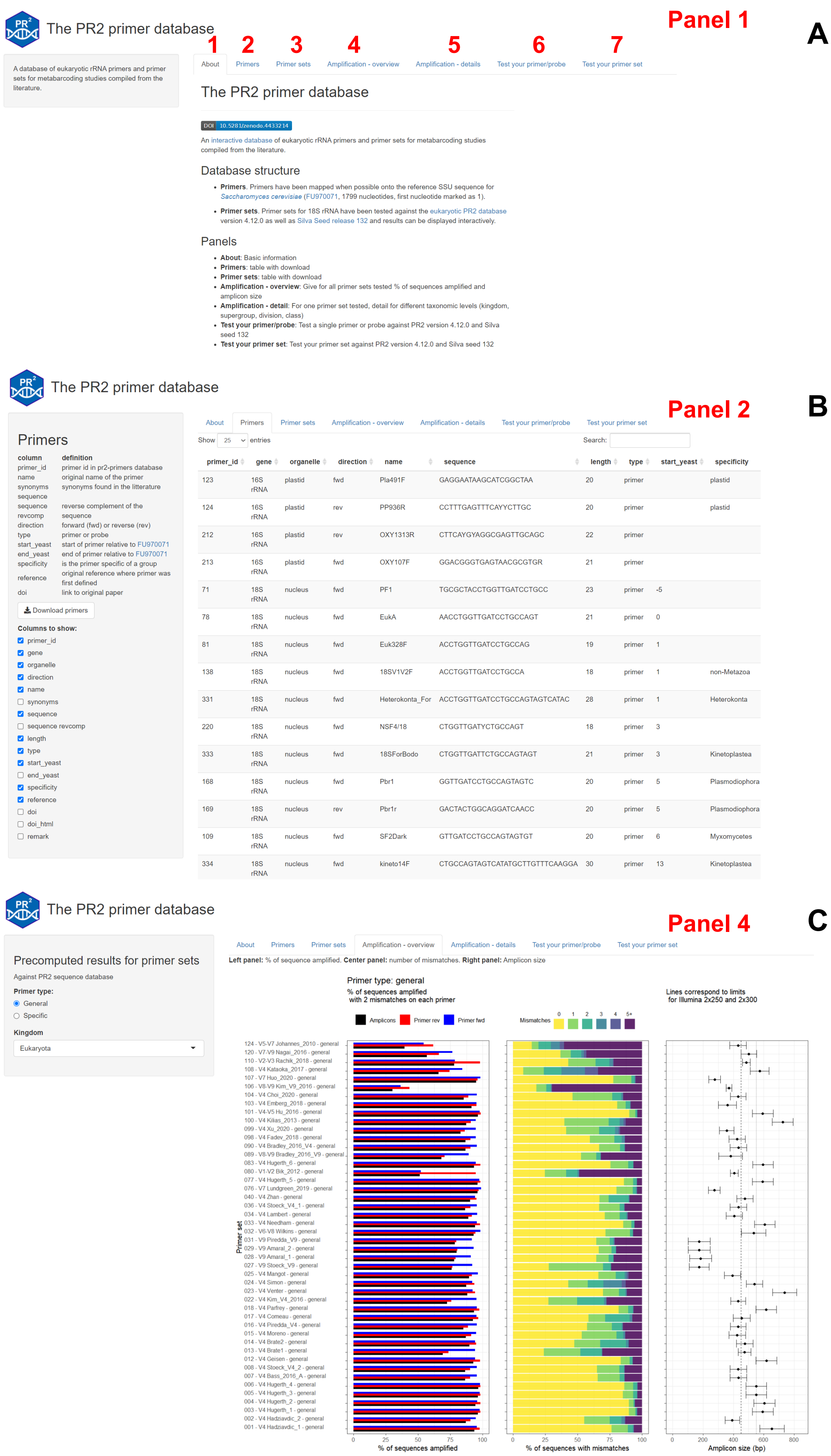

Figure 4: Shiny interface to the pr2-primers database. A. First panel introducing the database. Numbers in red correspond to the different panels. B. Second panel displaying the list of primers. The third panel is analogous, but for primer sets. C. Fourth panel showing *in silico* amplification results for all pre-computed primer sets.

Figure 5: Shiny interface to the pr2-primers database. A. Fifth panel introducing showing *in silico* detailed amplification results for a given pre-computed primer set. Taxonomy can be explored in detail. B. Seventh panel displaying *in silico* amplification results for user-provided primer sets. The sixth panel is identical but for a single primer or probe.

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 PHIM, F-34398 Montpellier, France

[7] PHIM, Univ Montpellier, CIRAD, INRAE, Institut Agro, 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

Supplementary Material

Table S1: List of 18S rRNA primers in the pr2-primers database ordered by start position relative to the sequence of the yeast *Saccharomyces cerevisiae* (FU970071). Some specific primers do not match the yeast sequence and are found at the bottom of the table. DOI for reference can be found in the on-line web application.

id	Name	Sequence	Direction	Start	Specificity	Reference
71	PF1	TGCGCTACCTGGTTGATCCTGCC	fwd	-5		Keeling, (2002)
78	EukA	AACCTGGTTGATCCTGCCAGT	fwd	0		Medlin et al. (1988)
81	Euk328F	ACCTGGTTGATCCTGCCAG	fwd	1		Moon et al. (2001
138	18SV1V2F	ACCTGGTTGATCCTGCCA	fwd	1	non-Metazoa	Clerissi et al. (2018)
331	Heterokonta For	ACCTGGTTGATCCTGCCAGTAGTCATAC	fwd	1	Heterokonta	Scheckenbach et al. (2010)
220	NSF4/18	CTGGTTGATYCTGCCAGT	fwd	3		Hendriks et al. (1989)
333	18SForBodo	CTGGTTGATTCTGCCAGTAGT	fwd	3	Kinetoplastea	Scheckenbach et al. (2010)
168	Pbr1	GGTTGATCCTGCCAGTAGTC	fwd	5	Plasmodiophora	Niwa et al. (2011)
169	Pbr1r	GACTACTGGCAGGATCAACC	rev	5	Plasmodiophora	Niwa et al. (2011)
109	SF2Dark	GTTGATCCTGCCAGTAGTGT	fwd	6	Myxomycetes	Fiore-Donno (2016)
334	kineto14F	CTGCCAGTAGTCATATGCTTGTTTCAAGGA	fwd	13	Kinetoplastea	von der Heyden and Cavalier-Smith (2005)
155	NS1	GTAGTCATATGCTTGTCTC	fwd	19		White et al. (1990)
142	25F	CATATGCTTGTCTCAAAGATTAAGCCA	fwd	24		Cavalier-Smith et al. (2009)
346	Thx25F	CATATGCTTGTCTCAAAGATTAAGCCA	fwd	24		Cavalier-Smith and von der Heyden (2007)
79	63f	ACGCTTGTCTCAAAGATTA	fwd	27		Lepere et al. (2011)
136	F04	GCTTGTCTCAAAGATTAAGCC	fwd	29		Blaxter et al. (1998)
255	HapGenFor33	TTGYCTYAAAGATTAAGCCATGCA	fwd	31	Haplosporidia	Ward et al. (2019)
234	18S-42F	CTCAARGAYTAAGCCATGCA	fwd	35		López-García et al. (2003)
294	73F	GCCATGCATGTCTAAGTATAAACWTC	fwd	47	Glissomonads	Howe et al. (2009)
329	Tel103-126F	TACACGGTGAGACTGCGAAT	fwd	74	Telonemia	Bråte et al. (2010)
246	Plas1f	TCAGTGAATCTGCGGATGGC	fwd	77	Plasmodiophorids	Neuhauser et al. (2014)
107	SFAca22	CGGYGAGACTGCGGATGG	fwd	78	Acanthamoeba	Fiore-Donno (2016)
256	HapGenFor84	CTGTGAAACTGCAKATGGCTC	fwd	78	Haplosporidia	Ward et al. (2019)
300	sA4-gra	CNGTGAAACWGCAGATGG	fwd	78	Granofilosea incl. Reticulamoeba spp.	Bass et al. (2012)
231	18S-82F	GAAACTGCGAATGGCTC	fwd	82		López-García et al. (2003)
313	NSF83	GAAACTGCGAATGGCTCATT	fwd	82		Hendriks et al. (1989)
366	18S V2i F	CTGTGAATGGCTCCTTACATCAG	fwd	86	Euglenids	Guminska et al. (2021)
104	Kineto 80	CATCAGACGYAATCTGCCGC	fwd	103	Kinetoplastea	Lentendu, G. et al. (2014)
82	152+	TTACATGGATAACCGTGGTAATTC	fwd	135	oligotrich, choreotrich	Tamura et al. (2011)
280	LGD-135	TAAGACGACGATTGCTGATTT	rev	135	Cyclotrichiid Ciliates	Bass et al. (2009)
301	V2f-d5	GGATAGCCGTACTAATTGTGG	fwd	142	Reticulamoeba spp.	Bass et al. (2012)
368	150+	AHTTACATGGATAACCGTGG	fwd	150	ciliates	Doherty et al. (2007)
98	Chryso 240	GGAAACCAATGCGGGGCAAC	fwd	216	Chrysophyceae	Lentendu, G. et al. (2014)
289	243F	CCAATGCACCCTCTGGGTGGTT	fwd	243	Cercomonas-clade A	Karpov et al. (2006)
310	Diphy257F	AAGWGGARTCATAATAACTTTTGCG	fwd	257	Diphyllatea	Orr et al. (2018)
88	Cer2F	ATTTCTGCCCTATCAGCT	fwd	300	Cercozoa	Lentendu, G. et al. (2014)
315	Oom278F	CTATCAGCTTTGGATGGTAGGA	fwd	309	Oomycetes	Holt et al. (2018)
252	Hap-E312f	CATAGCAGATGGAAGTTTGAGG	fwd	312	Haplosporidium sp.	Ward et al. (2019)
308	AU2	TTTCGATGGTAGGATAGDGG	fwd	317	Fungi	Vandenkoornhuyse et al. (2002)
139	18SV1V2R	GTARKCCWMTAYMYTACC	rev	324	non-Metazoa	Clerissi et al. (2018)
302	C3f-d5	GACATCTGAGGTGATAACGAA	fwd	344	Reticulamoeba spp.	Bass et al. (2012)
292	369R	TCGCATTACGTATCGCATTTCGCTG	rev	369		Karpov et al. (2006)
367	18S V2i R	GCTSCCTCTCCGGAATCRAACC	rev	371	Euglenids	Guminska et al. (2021)
137	R22	GCCTGCTGCCTTCCTTGGA	rev	412		Blaxter et al. (1998)
251	Hap-M412r	CGAGGTTGCCAAGTTCTTTCG	rev	412	Minchinia mytili	Ward et al. (2019)
195	Par 18S-F	GCAGCAGGCGYGAAAC	fwd	422	Parabasalids	Michaud et al. (2020)
227	545F	AGGCGCGTAAATTACCCAATC	fwd	427		Kawachi et al. (2016)
311	Diphy453F	CGCAAATTACCCAATCCTG	fwd	432	Diphyllatea	Orr et al. (2018)
254	Hap-E449r	TTGGATGCACTTTCAAGATTACC	rev	449	Haplosporidium sp.	Ward et al. (2019)
151	SAR V3 F	AYTCAGGGAGGTAGTGACAAG	fwd	451	SAR	Sisson et al. (2018)
242	mik451F	GCCGAGAYGGTTAAWGAGCCTCCT	fwd	451	Mikrocytid	Hartikainen et al. (2014)
83	528-	CCCGGCCCGTTATTTCTTGT	rev	467	oligotrich, choreotrich	Tamura et al. (2011)
344	MARa 502F	CAGAGATTTTCAATGGGGGATATTTAAYG	fwd	502	Neobodo designis marine clade	von der Heyden and Cavalier-Smith (2005)
250	Hap-M258f	AACTTTTAGCGTCCAGCCCA	fwd	528	Minchinia mytili	Ward et al. (2019)
232	Euk-516r	ACCAGACTTGCCCTCC	rev	547		Amann et al. (1990)
65	Uni18SF	AGGGCAAKYCTGGTGCCAGC	fwd	549		Zhan et al. (2013)
9	3NDf	GGCAAGTCTGGTGCCAG	fwd	551		Cavalier-Smith et al. (2009)
156	NS2	GGCTGCTGGCACCAGACTTGC	rev	552		White et al. (1990)
157	NS3	GCAAGTCTGGTGCCAGCAGCC	fwd	552		White et al. (1990)

id	Name	Sequence	Direction	Start	Specificity	Reference
13	515F	GTGCCAGCMGCCGCGGTAA	fwd	561		Parfrey et al. (2014)
19	515FY	GTGYCAGCMGCCGCGGTA	fwd	561		Parada et al. (2015)
25	EUK581-F	GTGCCAGCAGCCGCG	rev	561	non-Metazoan	Carnegie et al. (2003)
31	515F Univ	GTGYCAGCMGCCGCGGTAA	fwd	561		Needham and Fuhrman (2016)
320	530R	CCGCGGCKGCTGGCAC	rev	561	Microsporidia	Williams et al. (2018)
4	563f	GCCAGCAVCYGCGGTAAY	fwd	563		Hugerth et al. (2014)
7	V4 1f	CCAGCASCYGCGGTAATWCC	fwd	564		Bass et al. (2016)
8	TAReuk454FWD1	CCAGCASCYGCGGTAATTCC	fwd	564		Stoeck et al (2010)
67	Claudia Vannini (F)	CCAGCASCCGCGGTAATWCC	fwd	564	ciliates	Boscaro et al. (2017)
187	EuF-V4	CCAGCASCCGCGGTAATWCC	fwd	564		Belevich et al. (2017)
218	TAReuk454FWD1 Choi	CCAGCAGCCGCGGTAATTCC	fwd	564		Choi and Park (2020)
1	F-566	CAGCAGCCGCGGTAATTCC	fwd	565		Hadziavdic et al. (2014)
369	568	GGTSTAAATTCRKYTCATTKC	rev	568	ciliates	Doherty et al. (2007)
10	EUKAF	GCCGCGGTAATTCCAGCTC	fwd	570		Moreno et al. (2018)
69	ParaV45F	GCYGCGGTAATWCCAGCTCT	fwd	570	Parabasalids	Jasso-Selles et al. (2017)
12	E572F	CYGCGGTAATTCCAGCTC	fwd	571		Comeau et al. (2011)
14	528F	CCGCGGTAATTCCAGCTC	fwd	571		Zhu et al. (2005)
17	NSF563	CGCGGTAATTCCAGCTCCA	fwd	572		Mangot et al. (2013)
75	SSU556F	CGCGGTAATTCCAGCTYC	fwd	572	dinoflagellates	Smith et al (2017)
2	A-528F	GCGGTAATTCCAGCTCCAA	fwd	573		Cheung et al. (2010)
73	DIV4for	GCGGTAATTCCAGCTCCAATAG	fwd	573	diatoms	Visco et al. (2015)
3	574*f	CGGTAAYTCCAGCTCYV	fwd	574		Hugerth et al. (2014)
15	590F	CGGTAATTCCAGCTCCAATAGC	fwd	574		Venter et al. (2017)
32	Euk528F	CGGTAATTCCAGCTCC	fwd	574		Edgcomb et al. (2011)
132	574f	CGGTAAYTCCAGCTCYAV	fwd	574		Hugerth et al. (2014)
20	D512for	ATTCCAGCTCCAATAGCG	fwd	579	diatoms	Zimmermann et al. (2011
199	FF1100	CCAGCTCCAATAGCGTATATTA	fwd	582	Fungi	Vainio and Hantula (2000)
207	S32 J	CCAGCTCCAATAGCGTATAC	fwd	582	Radiolaria	Decelle et al. (2012)
208	S32 TASN	CCAGCTCCAATAGCGTATRC	fwd	582	Radiolaria	Ishitani et al. (2012)
21	Cerc479F	TGTTGCAGTTAAAAAGCTCGT	fwd	608		Harder et al. (2016)
16	EK-565F-NGS	GCAGTTAAAAAGCTCGTAGT	fwd	612		Simon et al. (2015)
152	SAR V3 R	RACTACGAGCTTTTTAACTGC	rev	612	SAR	Sisson et al. (2018)
5	616f	TTAAAAVGYTCGTAGTYG	fwd	616		Hugerth et al. (2014)
6	616*f	TTAAARVGYTCGTAGTYG	fwd	616		Hugerth et al. (2014)
90	S616F Cerco	TTAAAAAGCTCGTAGTTG	fwd	616	Cercozoa	Fiore-Donno et al. (2018)
91	S616F Eocer	TTAAAAAGCGCGTAGTTG	fwd	616	Cercozoa	Fiore-Donno et al. (2018)
253	Hap-E620r	GGAGCCAAATCCGAGGACTT	rev	620	Haplosporidium sp.	Ward et al. (2019)
376	AMV4.5NF	AAGCTCGTAGTTGAATTTCG	fwd	621	mycorrhizal fungi	Sato et al. (2005)
342	FWb 681F	GGAGTCGGTTACGTCCCRTCCTCCGRRYCG	fwd	681	Neobodo designis freshwater clade	von der Heyden and Cavalier-Smith (2005)
279	LGD-698	GCTTAGGTTTCTCGTCTTAGGA	fwd	698	Cyclotrichiid Ciliates	Bass et al. (2009)
99	Chryso 651	CTATTTTGCTCACAGTAAATGACGAG	rev	735	Chrysophyceae	Lentendu, G. et al. (2014)
303	V4r-d5b	GGATGACAATGTTTGCGGTGA	rev	740	Reticulamoeba gemmipara	Bass et al. (2012)
105	Kineto 651	TTGGTCGCRCTTYTTTAGTCACAG	rev	746	Kinetoplastea	Lentendu, G. et al. (2014)
274	817F	TTAGCATGGAATAATRRAATAGGA	fwd	793		Yang et al. (2020)
23	ChloroF	TGGCCTATCTTGTTGGTCTGT	fwd	822	Chlorophyceae	Valiente Moro et al. (2009)
190	HaptoR1	CGAAACCAACAAAATAGCAC	rev	823	Prymnesiophyceae	Egge et al. (2013)
296	Gv847F	ATCATTYAGCATGGAATAAACAYAAC	fwd	847	Sainouroids	Bass et al. (2016)
304	V4r-d5a	CTCGGATTCCTGAAACCAATG	rev	850	Reticulamoeba spp.	Bass et al. (2012)
206	S879	CCAACTGTCCCTATCAATCAT	rev	855	Radiolaria	Decelle et al. (2012)
377	AMDGR	CCCAACTATCCCTATTAATCAT	rev	855	mycorrhizal fungi	Sato et al. (2005)
244	mik868F	GGACTACCAGWGGCGAAAGCGCCT	fwd	868	Mikrocytid	Hartikainen et al. (2014)
89	Cer1R	ATACTAGCACCCCCAACT	rev	870	Cercozoa	Lentendu, G. et al. (2014)
52	Cerc750R	TGAATACTAGCACCCCCAAC	rev	871	Cercozoa	Harder et al. (2016)
74	DIV4rev3	CTCTGACAATGGAATACGAATA	rev	879	diatoms	Visco et al. (2015)
22	DimA	RGGGACRGGTGAAATAGGATG	fwd	893	diplonemids	Cannon et al. (2018)
76	SSU911R	ATYCAAGAATTTCACCTCTGAC	rev	894	dinoflagellates	Smith et al. (2017)
146	690R	ATCCAAGAATTTCACCTCTGAC	rev	894		Alves-de-Souza et al (2011)
77	B-706R	AATCCRAGAATTTCACCTCT	rev	897		Cheung et al. (2010
127	897f	AGAGGTGRAATTCTHRGA	fwd	897		Hugerth et al. (2014)
128	897r	TCYDAGAATTYCACCTCT	rev	897		Hugerth et al. (2014)
337	Kin1240rev	GCCTTCGCTGTAGTTCGTC	rev	924	Kinetoplastea	Scheckenbach et al. (2010)
35	R-952	TTGGCAAATGCTTTCGC	rev	935		Hadziavdic et al. (2014)
49	NSR951	TTGGYRAATGCTTTCGC	rev	935		Mangot et al. (2013)

id	Name	Sequence	Direction	Start	Specificity	Reference
93	S947R Cerco	AAGAAGACATCCTTGGTG	rev	946	Cercozoa	Fiore-Donno et al. (2018)
55	PRYM01+7	GATCAGTGAAAACATCCCTGG	rev	948	Haptophyta	Egge et al. (2013)
43	V4 18S Next.Rev	ACTTTCGTTCTTGATYRATGA	rev	960		Piredda et al. (2017)
92	S963R Cerco	CAACTTTCGTTCTTGATTAAA	rev	962	Cercozoa	Fiore-Donno et al. (2018)
37	TAReukREV3	ACTTTCGTTCTTGATYRA	rev	963		Stoeck et al (2010)
38	V4RB	ACTTTCGTTCTTGATYRR	rev	963		Balzano and Leterme (2015)
42	EUKAR	CYTTCGYYCTTGATTRA	rev	963		Moreno et al. (2018)
219	TAReukREV3 Choi	ACTTTCGTTCTTGATTAA	rev	963		Choi and Park (2020)
196	Par 18S-R	CCTACTCTCGCYCTTGATCG	rev	964	Parabasalids	Michaud et al. (2020)
188	picoR2	AKCCCCYAACTTTCGTTCTTGAT	rev	966		Belevich et al. (2017)
202	V4r	ACTTTCGTTCTTGAT	rev	966		Bradley et al. (2016)
46	Nex 18S 0964 R	GATCCCYYAACTTTCGTTCTTGA	rev	967		Kim et al. (2016)
228	1119R	TCCCCTAACTTTCGTTCTTG	rev	968		Kawachi et al. (2016)
68	Claudia Vannini (R)	TCTGRTYGTCTTTGATCCCYTA	rev	981	ciliates	Boscaro et al. (2017)
216	LABY-A	GGGATCGAAGATGATTAG	fwd	984	Labyrinthulomycetes	Stokes et al. (2002)
40	V4 euk R1	GACTACGACGGTATCTRATCRTCTTCG	rev	989		Bråte et al. (2010)
41	V4 euk R2	ACGGTATCTRATCRTCTTCG	rev	989		Bråte et al. (2010)
44	E1009R	AYGGTATCTRATCRTCTTYG	rev	989		Comeau et al. (2011)
66	Uni18SR	GRCGGTATCTRATCGYCTT	rev	991		Zhan et al. (2013)
51	D978rev	GACTACGATGGTATCTAATC	rev	996	diatoms	Zimmermann et al. (2011)
321	CM-V5F	GATTAGANACCNNNGTAGTTC	fwd	996	Microsporidia	Trzebny et al. (2020)
324	18S Naslb R	GAGACTACGACGGTATCTGATC	rev	996	Nassellaria	Sandin et al. (2019)
193	Oxy 18S-F	ATCAGAWACCGYCGTAGTC	fwd	997	Oxymonads	Michaud et al. (2020)
198	FF700	GATACCGTNGTAGTCT	fwd	1001	Fungi	Vainio and Hantula (2000)
257	C5f-Hapl	GTAGTCCCARCYATAAACBATGTC	fwd	1010	Haplosporidia	Hartikainen et al. (2014b)
316	Oom1024R	CTCATACGGTGCTGACAAGG	rev	1024	Oomycetes	Holt et al. (2018)
370	1199+	GCCGACTCGGGATCGGGGGC	fwd	1031	ciliates	Doherty et al. (2007)
380	1199+	GCCGACTCGGGATCGGGGGC	fwd	1031	oligotrich, choreotrich	Tamura et al. (2011)
305	V5r-d5b	GTCAACGCTCGCTGATCCCTG	rev	1055	Reticulamoeba spp.	Bass et al. (2012)
298	Gv1063F	AGCRAAAGCATTCATCAAT	fwd	1063	Sainouroids	Bass et al. (2016)
306	V5r-d5a	GGTGCCAACGAGGTCGTTTCA	rev	1075	Reticulamoeba spp.	Bass et al. (2012)
230	s14F3	ACGCAMGTGTGAAACTTG	fwd	1119	Foraminifera	Holzmann et al. (2003)
48	EUK1134-R	TTTAAGTTTCAGCCTTGCG	rev	1120		Carnegie et al. (2003)
265	UNonMet DB	CTTTAARTTTCASYCTTGCG	rev	1120	non-Metazoan	Bass and del Campo (2020)
30	926wF	AAACTYAAAKGAATTGRCGG	fwd	1130		Wilkins et al. (2013)
60	926R	CCGYCAATTYMTTTRAGTTT	rev	1130		Needham and Fuhrman (2016)
159	NS5	AACTTAAAGGAATTGACGGAAG	fwd	1131		White et al. (1990)
36	1132r	CCGTCAATTHCTTYAART	rev	1132		Hugerth et al. (2014)
211	1132rmod	TCCGTCAATTYCTTTAAGT	rev	1132		Geisen et al. (2018)
158	NS4	CTTCCGTCAATTCCTTTAAG	rev	1133		White et al. (1990)
224	1132R modified	CCGTCAATTHCTTYAAR	rev	1133		Hu et al. (2016)
70	ParaV45R	AAGRAATTGACGGAAGNGCA	rev	1137	Parabasalids	Jasso-Selles et al. (2018)
45	1119r	GGTGCCCTTCCGTCA	rev	1144		Parfrey et al. (2014)
96	s14f1	AAGGGCACCACAAGAACGC	fwd	1150	Foraminifera	de Vargas et al. (1997)
322	CM-V5R	TAANCAGCACAMTCCACTC	rev	1164	Microsporidia	Trzebny et al. (2020)
53	DimB	CAAATTGAGCCGCAGACTCC	rev	1168		Cannon et al. (2018)
225	960F	GGCTTAATTTGACTCAACRCG	fwd	1177		Gast et al. (2004)
34	R-1200	CCCGTGTTGAGTCAAATTAAGC	rev	1178		Hadziavdic et al. (2014)
129	F-1183	AATTTGACTCAACACGGG	fwd	1182		Hadziavdic et al. (2014)
275	1196R	TCTGGACCTGGTGAGTTTCC	rev	1199		Yang et al. (2020)
291	1259F	GGTCCRGACAYAGTRAGGATTGACAGATTGAAG	fwd	1211	Cercozoa	Karpov et al. (2006)
94	1301f	GATTGAAGCTCTTTCTTGATCACTTC	fwd	1236	Plasmodiophorida	Bass et al. (2018)
288	1256R	GCACCACCACCCAYAGAATCAAGAAAGAWCTTC	rev	1242	Cercozoa	Bass and Cavalier-Smith (2004)
343	FWb 1244R	TATTCTCTTTTGGCGGGMTCAGCAAGCGAG	rev	1244	Neobodo designis freshwater clade	von der Heyden and Cavalier-Smith (2005)
97	s15.3	CCTATCACATAATCATGAAAG	rev	1247	Foraminifera	Pawlowski, J., et al., (2014)
148	1055R	ACGGCCATGCACCACCACCCAT	rev	1260		Alves-de-Souza et al (2011)
147	1055F	GGTGGTGCATGGCCGTTCTT	fwd	1266		Alves-de-Souza et al (2011)
47	1300R	CACCAACTAAGAACGGCCATGC	rev	1272		Venter et al (2017)
266	SSR-F1 289	TGGAGYGATTTGTCTGGTTDATTCCG	fwd	1292		Nagai et al. (2016)
54	ChloroR	GAATCAACCTGACAAGGCAAC	rev	1295	Chlorophyceae	Valiente Moro et al. (2009)
299	hkx1295R	TCAATCCACTCACTTCCCAAAGGC	rev	1295	Sainouroids	Bass et al. (2016)
197	FF390	CGATAACGAACGAGACCT	fwd	1316	Fungi	Vainio and Hantula (2000)
345	MARa 1321R	GGACGTGCTGAGGATATTCCCGWTA	rev	1321	Neobodo designis marine clade	von der Heyden and Cavalier-Smith (2005)

id	Name	Sequence	Direction	Start	Specificity	Reference
245	mik1340	TGCATCACGGACCTACCTTWGACC	rev	1340	Mikrocytid	Hartikainen et al. (2014)
217	LABY-Y	CWCRAACTTCCTTCCGGT	rev	1400	Labyrinthulomycetes	Stokes et al. (2002)
160	NS6	GCATCACAGACCTGTTATTGCCTC	rev	1415		White et al. (1990)
161	NS7	GAGGCAATAACAGGTCTGTGATGC	fwd	1415		White et al. (1990)
221	Nex 18S 1434 F	GAGGCAATAACAGGTCTGTGATG	fwd	1415		Kim et al. (2016)
201	V8f	ATAACAGGTCTGTGATGCCCT	fwd	1421		Bradley et al. (2016)
203	1422f	ATAACAGGTCTGTGATGC	fwd	1421		Hadziavdic et al. (2016)
226	NSR1438	GGGCATCACAGACCTGTTAT	rev	1421		Van De Peer et al. (2000)
194	Oxy 18S-R	GGGCATMACRGACCTGTTA	rev	1422	Oxymonads	Michaud et al. (2020)
204	1424f	AACAGGTCHGWRATGCCC	fwd	1423		Hugerth et al. (2014)
131	R-1443	AAGGGCATCACAGACCTG	rev	1425		Hadziavdic et al. (2014)
297	hkx1442R	ATCTAAGAGCATCACGGACCTTTTATC	rev	1442	Sainouroids	Bass et al. (2016)
243	mik1511	CCTATTCAGCGCGCTCTGTTGAGA	rev	1511	Mikrocytid	Hartikainen et al. (2014)
238	SL175pr5F	ACGAGGAATGCCTAGTAAGCGCAA	fwd	1569	Mantoniella antarctica	Gast et al. (2014)
330	Tel3250-3230R	GACGTAATCAGGGCGGTCT	rev	1590	Telonemia	Bråte et al. (2010)
277	ParaGENrGW	GTGTACAAAGGRCAGGGACT	rev	1614	Paramyxids	Ward et al. (2016)
27	1380F	CCCTGCCHTTTGTACACAC	fwd	1617		Amaral Zettler et al (2009)
318	Microsp1342r	ACGGGCGGTGTGTACAAAGAACAG	rev	1619	Microsporidia	Stentiford et al. (2017)
62	U1391R	GGGCGGTGTGTACAARGR	rev	1623		Edgcomb et al. (2011)
28	1389F	TTGTACACACCGCCC	fwd	1626		Amaral Zettler et al (2009)
29	1388F	TTGTACACACCGCCCGTCGC	fwd	1626		Piredda et al. (2017)
59	1392-R	ACGGGCGGTGTGTRC	rev	1628		Wilkins et al. (2013)
145	18r71	GCGACGGGCGGTGTGTAC	rev	1628		Alves-de-Souza et al (2011)
166	ITS9MUNngs	TACACACCGCCCGTCG	fwd	1629		Tedersoo and Lindahl (2016)
200	FR1	ANCCATTCAATCGGTANT	rev	1647	Fungi	Vainio and Hantula (2000)
309	AU4	RTCTCACTAAGCCATTC	rev	1657	Fungi	Vandenkoornhuyse et al. (2002)
295	1682R	ATCCGTGAAGCTCACTAATC	rev	1682	Glissomonads	Howe et al. (2009)
290	1733R	TGATCAAGTTTGATTCAGTTCTCGGAT	rev	1733	Cercomonas-clade A	Karpov et al. (2006)
260	sB2hap	CCTTGTTACGACTTBTYCTTCCTC	rev	1744	Haplosporidia	Hartikainen et al. (2014b)
307	sB2-d5	CCTTGTTACGACTTTTGC	rev	1750	Reticulamoeba spp. + some eukaryotes	Bass et al. (2012)
332	Heterokonta Rev	GGTTCACCTACGGAAACCTTGTTACGACTTCA	rev	1752	Heterokonta	Scheckenbach et al. (2010)
95	1801r	ACGGAAACCTTGTTACGACTTC	rev	1753	Plasmodiophorida	Bass et al. (2018)
61	U1492R	GGTTACCTTGTTACGACTT	rev	1754		Edgcomb et al. (2011)
63	U1517R	ACGGCTACCTTGTTACGACTT	rev	1754		Edgcomb et al. (2011)
80	1818r	ACGGAAACCTTGTTACGA	rev	1757		Lepere et al. (2011
267	SSR-R1 772	TCACCTACGGAAACCTTGTTACG	rev	1758		Nagai et al. (2016)
235	18S-1498R	CACCTACGGAAACCTTGTTA	rev	1760		López-García et al. (2003)
222	Nex 18S 1757 R	CAGGTTCACCTACGGAAACCT	rev	1765		Kim et al. (2016)
371	1765-	CCCCAKCACGACDCMTATTGCTG	rev	1765	ciliates	Doherty et al. (2007)
237	RS11pr4R	CTGCAGGTTCACCTACGGAAACC	rev	1766	Pyramimonas cf. tychotreta	Gast et al. (2014)
162	NS8	TCCGCAGGTTCACCTACGGA	rev	1770		White et al. (1990)
259	sB1N	GATCCHTCYGCAGGTTCACCTACG	rev	1772		Hartikainen et al. (2014b)
262	Sb1n	GATCCHTCYGGAGGTTCACCTACG	rev	1772	Paradinids	Ward et al. (2018)
57	EukB	TGATCCTTCTGCAGGTTCACCTAC	rev	1773		Medlin et al. (1988)
58	1510R	CCTTCYGCAGGTTCACCTAC	rev	1773		López-García et al. (2003)
143	1801R	TGATCCTTCTGCAGGTTCACCT	rev	1775		Cavalier-Smith et al. (2009)
336	18SRevBodo	TGATCCAGCTGCAGGTTCACC	rev	1776	Kinetoplastea	Scheckenbach et al. (2010)
312	Diphy1881R	CGACCAAAACTCCAAAGATTTC	rev	1860	Diphyllatea	Orr et al. (2018)
347	Helio1979R	CACACTTACWAGGAYTTCCTCGTTSAAGACG	rev	1979	Centrohelid heliozoa	Cavalier-Smith and von der Heyden (2007)
335	kineto2026R	GATCCTTCTGCAGGTTCACCTACAGCT	rev	2026	Kinetoplastea	von der Heyden and Cavalier-Smith (2005)
106	SRAca28	CCAATTACAAGACTCTTRTCGAG	fwd		Acanthamoeba	Fiore-Donno (2016)
108	SR19Dark	GTCCTCTAATTGTTACTCGAD	fwd		Myxomycetes	Fiore-Donno (2016)
112	Pdir1	GATTTCGGGCGGGTTCCA	fwd		Pedinophyceae	Milyutina et al. (2019)
113	Pdir2	GATCGGGCTTCGGTTCGAG	fwd		Pedinophyceae	Milyutina et al. (2019)
114	Prev2	CTCGCGGAACTCGAACCGAAG	rev		Pedinophyceae	Milyutina et al. (2019)
115	Pdir3	CCTCAGCCTGCTAAATAGCTAC	fwd		Pedinophyceae	Milyutina et al. (2019)
116	Pdir4	GACTTTCGGGGTTTTACCCGGA	fwd		Pedinophyceae	Milyutina et al. (2019)
134	S19F	GTGCATGGCCGTTCTTAGTTC	rev		Foraminifera	Morard et al. (2011)
135	S15rF	CCCGTACRAGGCATTCCTAG	fwd		Foraminifera	Morard et al. (2011)
144	329R	GTGAACCTGCRGAAGGATCA	rev			Alves-de-Souza et al (2011)
170	Pb121	GGATACAAAAACCAAACCTGGC	fwd		Plasmodiophora	Niwa et al. (2011)
171	Pb121r	GCCAGGTTTGGTTTTTGTATCC	rev		Plasmodiophora	Niwa et al. (2011)
186	SB	GTAGGTGAACCTGCAGAAGGATCA	rev			Sogin (1990)

id	Name	Sequence	Direction	Start	Specificity	Reference
192	PRYM03+3	GTAAATTGCCCGAATCCTG	fwd		Prymnesiophyceae	Egge et al. (2013)
205	17	CGGTCACGTTCGTTGC	rev		Foraminifera	Cordier et al. (2019)
210	S51 TAS	YAAGAATTTCACCTCTCGCTT	rev		Radiolaria	Ishitani et al. (2012)
214	QPX-F	ATCCTCGGCCTGCTTTTAGTAG	fwd		Quahog parasite	Stokes et al. (2002)
215	QPX-R2	GAAGTCTCTACCTTTCTTGCGA	rev		Quahog parasite	Stokes et al. (2002)
236	RS11pr4F	ATGTTCGGATCGCGGCGAGAC	fwd		Pyramimonas cf. tychotreta	Gast et al. (2014)
239	SL175pr5R	TAGAAAGCCACGGTCCGAACGC	rev		Mantoniella antarctica	Gast et al. (2014)
240	Gempr2F	TCGGATTGCTGGGTAGAACTTCGT	fwd		Geminigera cryophila	Gast et al. (2014)
241	Gempr2R	CACCTACGGGAAACCTTGTTACGAC	rev		Geminigera cryophila	Gast et al. (2014)
247	Plas1r	GGTGCSKCKAGRTVCAAGAGGC	rev		Plasmodiophorids	Neuhauser et al. (2014)
248	Plas2f	TGGATGTACGAGAGTACTACATGG	fwd		Plasmodiophorids	Neuhauser et al. (2014)
249	Plas2r	CGTTGAACCTAGCATTGTAGCG	rev		Plasmodiophorids	Neuhauser et al. (2014)
258	V5f-Hapl	GGACTCRGGGGGAAGTATGCT	fwd		Haplosporidia	Hartikainen et al. (2014b)
261	V4fAsce	GGAATAATAWGATAGGACTTCRGCA	fwd		Paradinids	Ward et al. (2018)
263	V5fAsce	GYTCRGCACCKTATTYGAGAAATCA	fwd		Paradinids	Ward et al. (2018)
264	EndoR1	CGACTTCTCCTTCCTCTAARYRDTAWG	rev		Paradinids	Ward et al. (2018)
276	Para1fGW	GGGCGAGGGGTAAAATCT	fwd		Paramyxids	Ward et al. (2016)
278	Para3fGW	GGCTTYTGGGAGAKTACGG	fwd		Paramyxids	Ward et al. (2016)
281	myxo 617F all	CGCGCAAATTACCCAMTCCA	fwd		Myxozoa	Hartikainen et al. (2016)
282	myxo 764F all	CCGCGGTAATTCCAGCTCCAG	fwd		Myxozoa	Hartikainen et al. (2016)
283	myxo 2313R all	CGTTACCGGAATRRCCTGACAG	rev		Myxozoa	Hartikainen et al. (2016)
284	myxo 1817 v1	ATTTCACCTCTCGCCATCGA	rev		Myxozoa	Hartikainen et al. (2016)
285	myxo 1817 v2	ATTTCACCTCTCGCGGCMAA	rev		Myxozoa	Hartikainen et al. (2016)
286	myxo 1817 v3	ATTTCACCTCTCGCTGCCAA	rev		Myxozoa	Hartikainen et al. (2016)
317	CTMicrosp	CACCAGGTTGATTCTGCCTGACG	fwd		Microsporidia	Stentiford et al. (2017)
319	V1F	CACCAGGTTGATTCTGCCTGAC	fwd		Microsporidia	Williams et al. (2018)
323	APU-1R	CTTCCTTTGGTTAAAACAC	rev		Apusomonads	Torruella et al. (2017)
325	18S NassII F	AGCATGGAATAATAACTGATGA	fwd		Nassellaria	Sandin et al. (2019)
326	18S NassII R	CACCARTTCATCCAATCGGTAG	rev		Nassellaria	Sandin et al. (2019)
340	DiploF	GATATCTAAACCTGTC	fwd		diplonemids	Lara et al. (2009)
341	DiploR	GCATTCCTCATTCAAGGA	rev		diplonemids	Lara et al. (2009)
378	BaciF	AGATTGCCCAGGCCTCTCG	fwd		Bacillariophyceae	Valiente Moro et al. (2009)
379	BaciR	CCATCGTAGTCTTAACCATAAAC	rev		Bacillariophyceae	Valiente Moro et al. (2009)
381	1765-	CCCCAKCACGACDCMTATTGCTG	rev		oligotrich, choreotrich	Tamura et al. (2011)
172	Pbr2r	CTCTATGCCCGAATCGCTTC	rev		Plasmodiophora	Niwa et al. (2011)
173	Pbr4	GTGTCGCTTAAGATATAGTC	fwd		Plasmodiophora	Niwa et al. (2011)
174	Pbr4r	GACTATATCTTAAGCGACAC	rev		Plasmodiophora	Niwa et al. (2011)

Table S2: List of 18S rRNA primer sets used for metabarcoding in the pr2-primers database. Size corresponds to the average amplicon size (bp) for sequences from the PR2 database. DOI for reference can be found in the on-line web application.

id	Name	Primer fwd	Primer rev	Region	Size	Specificity	Reference
1	Hadziavdic 1	F-566	R-1200	V4	654		Hadziavdic et al. (2014)
2	Hadziavdic 2	A-528F	R-952	V4	392		Hadziavdic et al. (2014)
3	Hugerth 1	574*f	1132r	V4	594		Hugerth et al. (2014)
4	Hugerth 2	563f	1132r	V4	604		Hugerth et al. (2014)
5	Hugerth 3	616f	1132r	V4	551		Hugerth et al. (2014)
6	Hugerth 4	616*f	1132r	V4	551		Hugerth et al. (2014)
7	Bass 2016 A	V4 1f	TAReukREV3	V4	433		Bass et al. (2016)
8	Stoeck V4 2	TAReuk454FWD1	TAReukREV3	V4	433		Stoeck et al (2010)
12	Geisen	3NDf	1132rmod	V4	617		Geisen et al. (2018)
13	Brate1	3NDf	V4 euk R1	V4	473		Bråte et al. (2010)
14	Brate2	3NDf	V4 euk R2	V4	475		Bråte et al. (2010)
15	Moreno	EUKAF	EUKAR	V4	425		Moreno et al. (2018)
16	Piredda V4	TAReuk454FWD1	V4 18S Next.Rev	V4	432		Piredda et al. (2017)
17	Comeau	E572F	E1009R	V4	454		Comeau et al. (2011)
18	Parfrey	515F	1119r	V4	615		Parfrey et al. (2014)
19	Vannini	Claudia Vannini (F)	Claudia Vannini (R)	V4	439	ciliates	Boscaro et al. (2017)
21	Zimmerman	D512for	D978rev	V4	444	diatoms	Zimmermann et al. (2011)
22	Kim V4 2016	528F	Nex 18S 0964 R	V4	431		Kim et al. (2016)
23	Venter	590F	1300R	V4	738		Venter et al (2017)
24	Simon	EK-565F-NGS	EUK1134-R	V4	538		Simon et al. (2015)
25	Mangot	NSF563	NSR951	V4	394		Mangot at al. (2013)
27	Stoeck V9	1391F	EukB	V9	175		Stoeck et al (2010)
28	Amaral 1	1380F	1510R	V9	184		Amaral Zettler et al (2009)
29	Amaral 2	1389F	1510R	V9	175		Amaral Zettler et al (2009)
31	Piredda V9	1388F	1510R	V9	175		Piredda et al. (2017)
32	Wilkins	926wF	1392-R	V6-V8	534		Wilkins et al. (2013)
33	Needham	515F Univ	926R	V4	606		Needham and Fuhrman (2016)
34	Lambert	515FY	NSR951	V4	405		Lambert et al. (2019)
35	UNonMet	EUK581-F	EUK1134-R	V4	595	non-Metazoa	Carnegie et al. (2003)
36	Stoeck V4 1	TAReuk454FWD1	V4RB	V4	433		Balzano and Leterme (2015)
37	Cannon	DimA	DimB	V5	284	diplonemids	Cannon et al. (2018)
39	Egge	A-528F	PRYM01+7	V4	400	Haptophyta	Egge et al. (2013)
40	Zhan	Uni18SF	Uni18SR	V4	476		Zhan et al. (2013)
41	Harder	Cerc479F	Cerc750R	V4	295	Cercozoa	Harder et al. (2016)
59	Tamura OCSP-A	152+	528-	V2-V3	351	oligotrich, choreotrich	Tamura et al. (2011)
62	Lentendu 2014a	Cer2F	Cer1R	V3-V4	599	Cercozoa	Lentendu et al. (2014)
63	Fiore-Donno 2018a	S616F Cerco	S963R Cerco	V4	378	Cercozoa	Fiore-Donno et al. (2018)
64	Fiore-Donno 2018b	S616F Eocer	S963R Cerco	V4	377	Cercozoa	Fiore-Donno et al. (2018)
65	Fiore-Donno 2018c	S616F Cerco	S947R Cerco	V4	359	Cercozoa	Fiore-Donno et al. (2018)
66	Fiore-Donno 2018d	S616F Eocer	S947R Cerco	V4	359	Cercozoa	Fiore-Donno et al. (2018)
67	Bass 2018	1301f	1801r	V7-V9	544	Plasmodiophorida	Bass et al. (2018)
68	Pawlowski 2010	s14f1	s15.3	37F	160	Foraminifera	Pawlowski and Lecroq (2010)
69	Lentendu 2014b	Chryso 240	Chryso 651	V2-V3	546	Chrysophyceae	Lentendu et al. (2014)
72	Lentendu 2014c	Kineto 80	Kineto 651	V2-V3	786	Kinetoplastea	Lentendu et al. (2014)
73	Fiore-Donno 2016a	SRAca28	SFAca22	V2		Acanthamoeba	Fiore-Donno et al. (2016)
74	Fiore-Donno 2016b	SR19Dark	SF2Dark	V2		Myxomycetes	Fiore-Donno et al. (2016)
76	Lundgreen 2019	F-1183	R-1443	V7	274		Lundgreen et al. (2019)
77	Hugerth 5	574f	1132r	V4	594		Hugerth et al. (2014)
80	Creer 2010	F04	R22	V1-V2	406		Creer et al. (2010)
81	Clerissi 2018	18SV1V2F	18SV1V2R	V1-V2	338	non-Metazoa	Clerissi et al. (2018)
83	Hugerth 6	A-528F	1132r	V4	596		Hugerth et al. (2014)
84	Sisson 2018	SAR V3 F	SAR V3 R	V3	184	SAR	Sisson et al. (2018)
86	Belevich 2017	EuF-V4	picoR2	V4	437	picoplankton	Belevich et al. (2017)
87	Michaud 2019a	Oxy 18S-F	Oxy 18S-R	V3-V4	460	Oxymonads	Michaud et al. (2020)
88	Michaud 2019b	Par 18S-F	Par 18S-R	V5	484	Parabasalia	Michaud et al. (2020)
89	Bradley 2016 V9	V8f	1510R	V8-V9	382		Bradley et al. (2016)
90	Bradley 2016 V4	TAReuk454FWD1	V4r	V4	433		Bradley et al. (2016)
91	Cordier 2019	s14f1	17	37F-41F	323	Foraminifera	Cordier et al. (2019)
92	Chemidlin 2011	FF390	FR1	V7-V8	367	fungi	Chemidlin Prevost-Boure et al. (2011)

id	Name	Primer fwd	Primer rev	Region	Size	Specificity	Reference
97	Stokes 2002	LABY-A	LABY-Y	V6	437	Labyrinthulomycetes	Stokes et al. 2002
98	Fadev 2018	A-528F	V4RB	V4	424		Fadeev et al. (2018)
99	Xu 2020	A-528F	B-706R	V4	356		Xu et al. (2020)
100	Kilias 2013	A-528F	1055R	V4	725		Kilias et al. (2013).
101	Hu 2016	574*f	1132R modified	V4-V5	594		Hu et al. (2016)
102	Piwosz 2019	TAReuk454FWD1	HaptoR1	V4	280	Haptophyta	Piwosz (2019)
103	Emberg 2018	E572F	897r	V4	361		Enberg et al. (2018)
104	Choi 2020	TAReuk454FWD1 Choi	TAReukREV3 Choi	V4	432		Choi and Park (2020)
106	Kim V9 2016	Nex 18S 1434 F	Nex 18S 1757 R	V8-V9	371		Kim et al. (2016)
107	Huo 2020	960F	NSR1438	V7	277		Huo et al. (2020)
108	Kataoka 2017	545F	1119R	V4	573		Kataoka et al. (2017)
109	Li 2020	s14F3	17	37F-41F	347	Foraminifera	Li et al. (2020)
110	Rachik 2018	18S-82F	Euk-516r	V2-V3	484		Rachik et al. (2018)
117	Ward 2018 round 1	V4fAsce	Sb1n	V5-V9	995	Paradinids	Ward et al. (2018)
118	Ward 2018 round 2	V5fAsce	EndoR1	V5-V9	797	Paradinids	Ward et al. (2018)
119	Bass 2020	574*f	UNonMet DB	V4	583	non-Metazoa	Bass and del Campo (2020)
120	Nagai 2016	SSR-F1 289	SSR-R1 772	V7-V9	502		Nagai et al. (2016)
124	Johannes 2010	817F	1196R	V5-V7	431		Yang et al. (2020)
128	Hartikainen 2016 round 1	myxo 617F all	myxo 2313R all	V4	996	Myxozoa	Hartikainen et al. (2016)
129	Hartikainen 2016 round 2	myxo 764F all	myxo 1817 v1	V4	351	Myxozoa	Hartikainen et al. (2016)
134	Williams 2018	V1F	530R	V1-V3	441	Microsporidia	Williams et al. (2018)
135	Trzebny 2020	CM-V5F	CM-V5R	V5	195	Microsporidia	Trzebny et al. (2020)
144	Guminska 2021	18S V2i F	18S V2i R	V2	356	Euglenids	Guminska et al. (2021)
149	Sato 2005	AMV4.5NF	AMDGR	V4	260	mycorrhizal fungi	Sato et al. (2005)

Table S3: Overall statistics for *in silico* % amplification of PR^2 sequences for primer sets listed in the pr2-primers database.

	general	specific
forward primers		
min	36.4	0.0
mean	91.0	49.7
max	98.7	97.6
reverse primers		
min	43.2	0.0
mean	88.7	32.4
max	98.6	98.9
primer sets		
min	30.0	0.0
mean	83.4	18.6
max	96.5	92.7

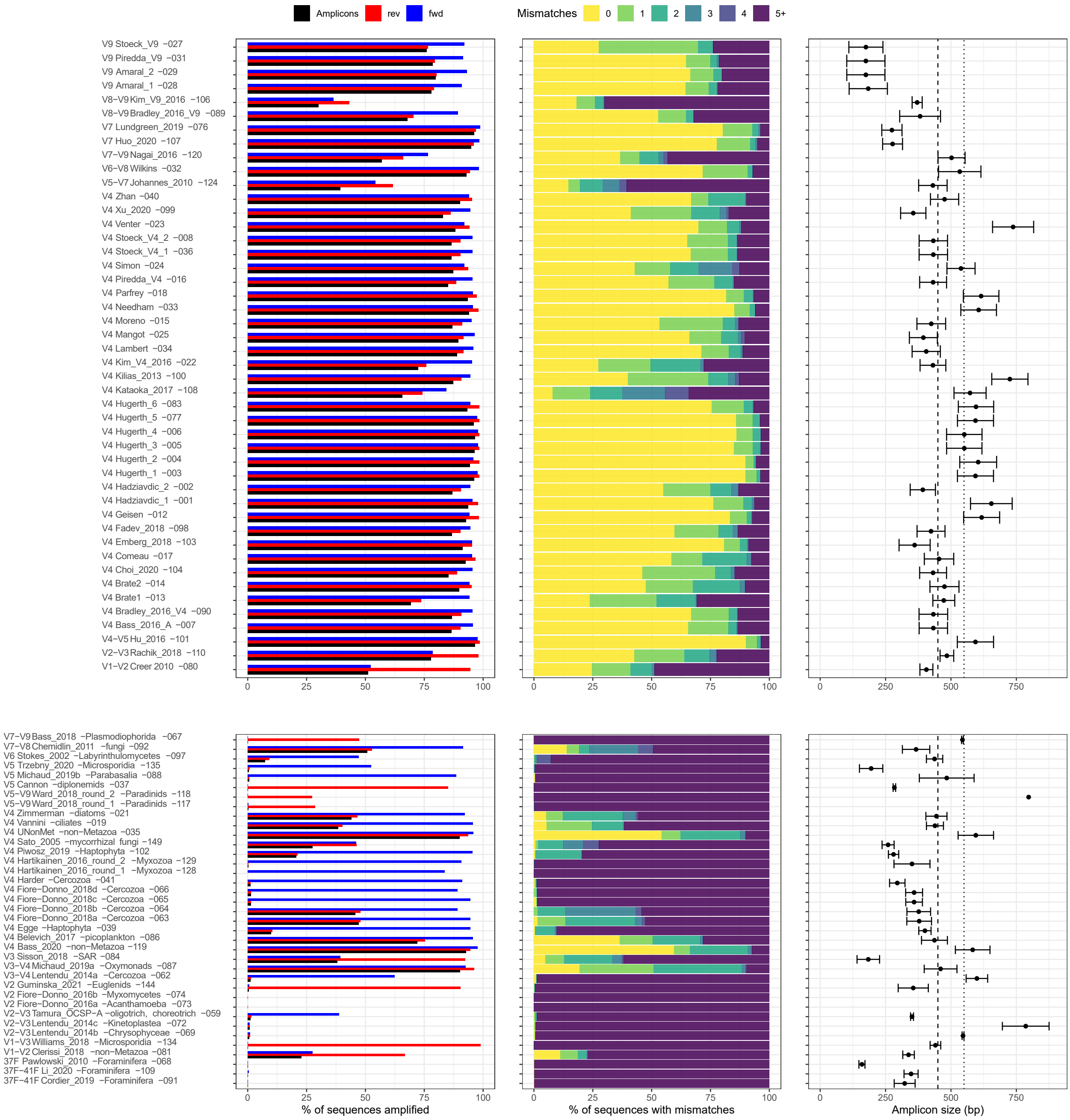

Figure S1: Evaluation of all general (top) or specific (bottom) primer sets (Table S2) for the 18S rRNA gene against the PR^2 reference database (version 4.12.0). See Fig. 2 for legend.

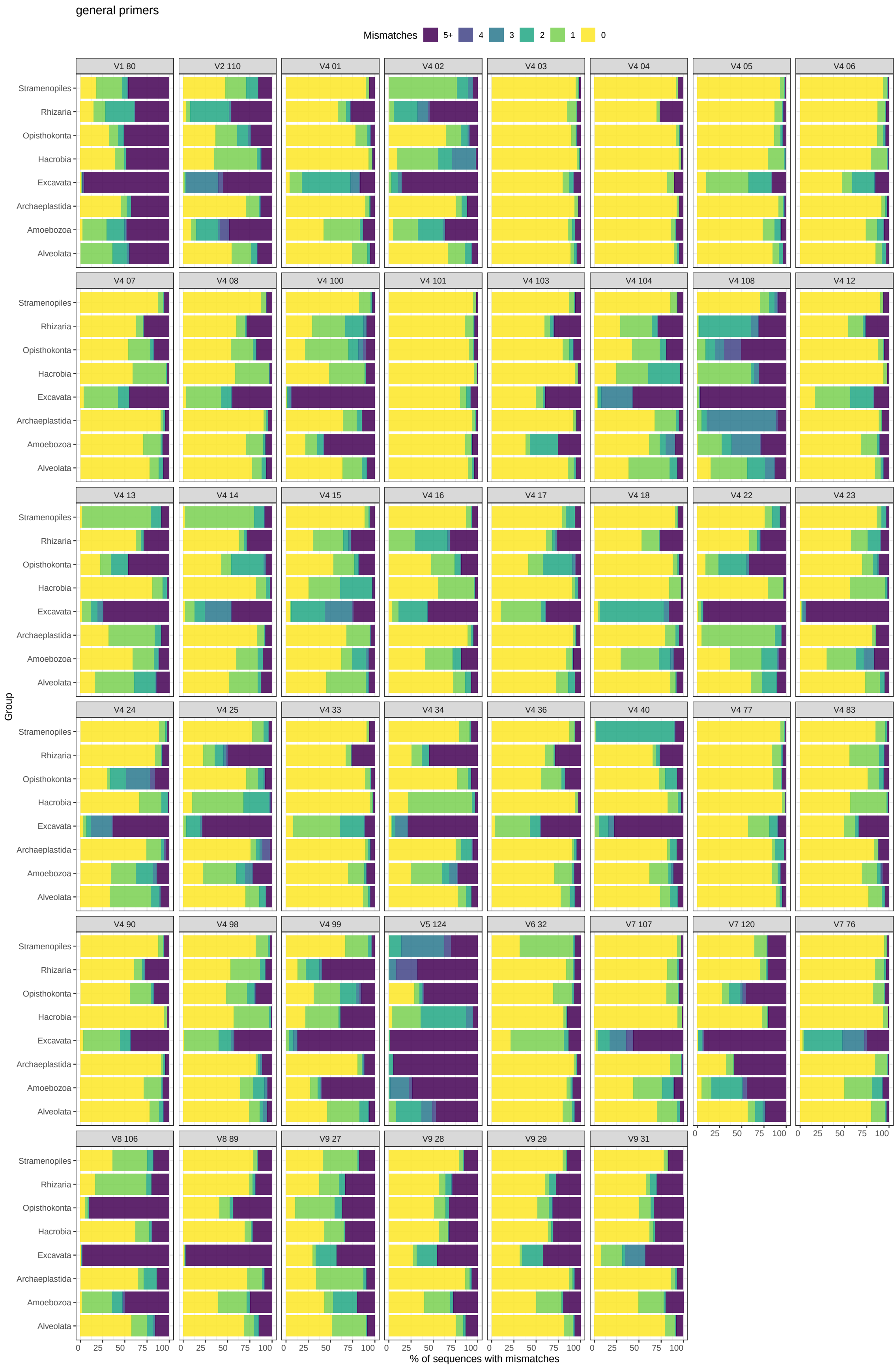

Figure S2: Number of mismatches for general primer sets as a function of the supergroup.

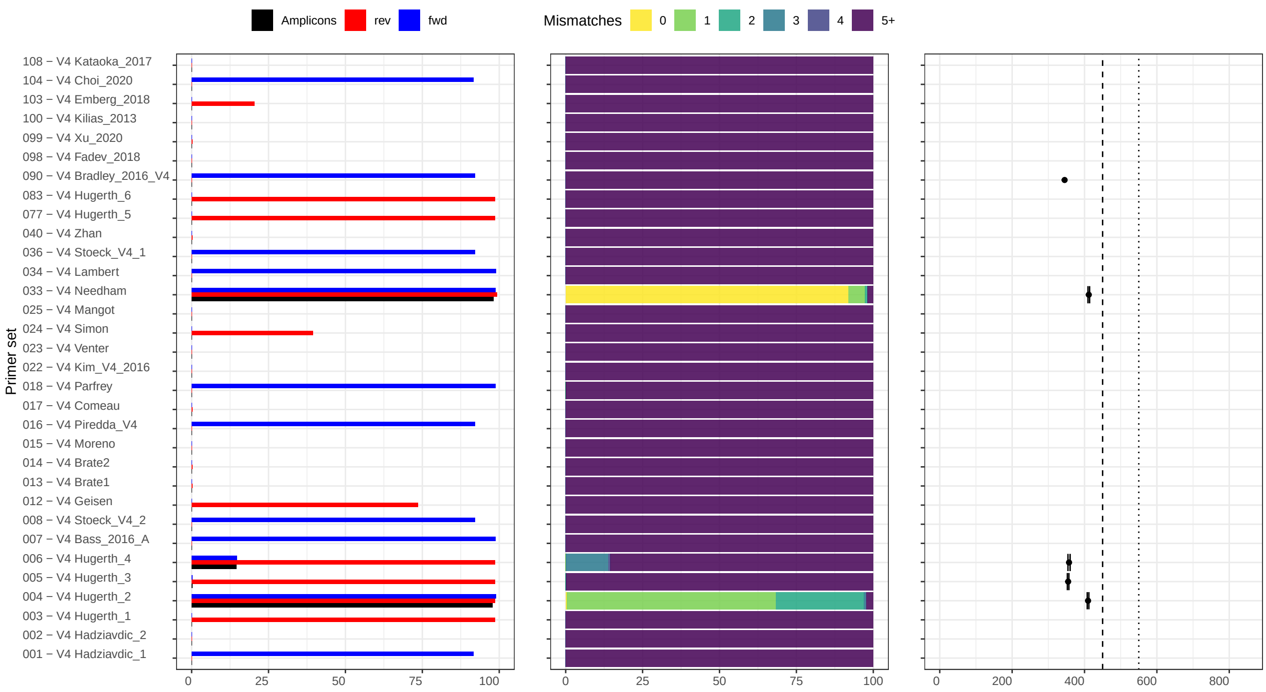

Figure S3: Evaluation of general primer sets (Table S2) targeting the V4 region of the 18S rRNA gene against bacterial 16S rRNA sequences from the Silva seed reference database (version 132). Legend as in Figure 2.

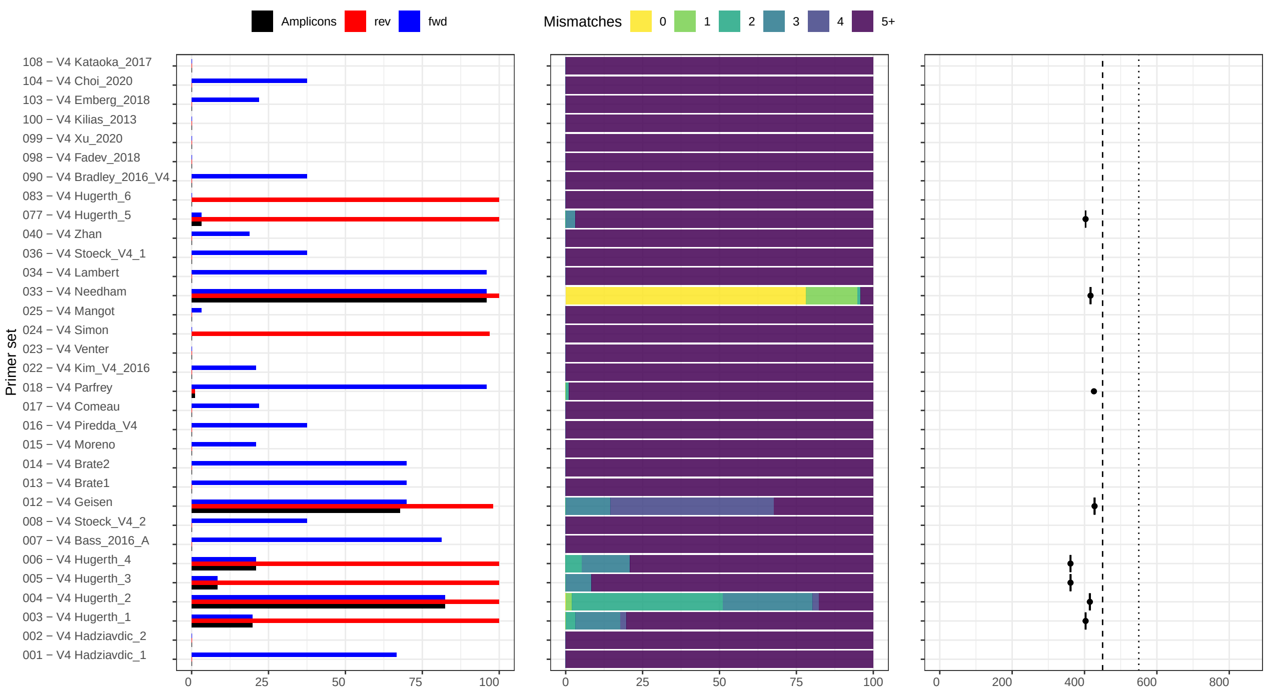

Figure S4: Evaluation of general primer sets (Table S2) targeting the V4 region of the 18S rRNA gene against archaeal 16S rRNA sequences from the Silva seed reference database (version 132). Legend as in Figure 2.

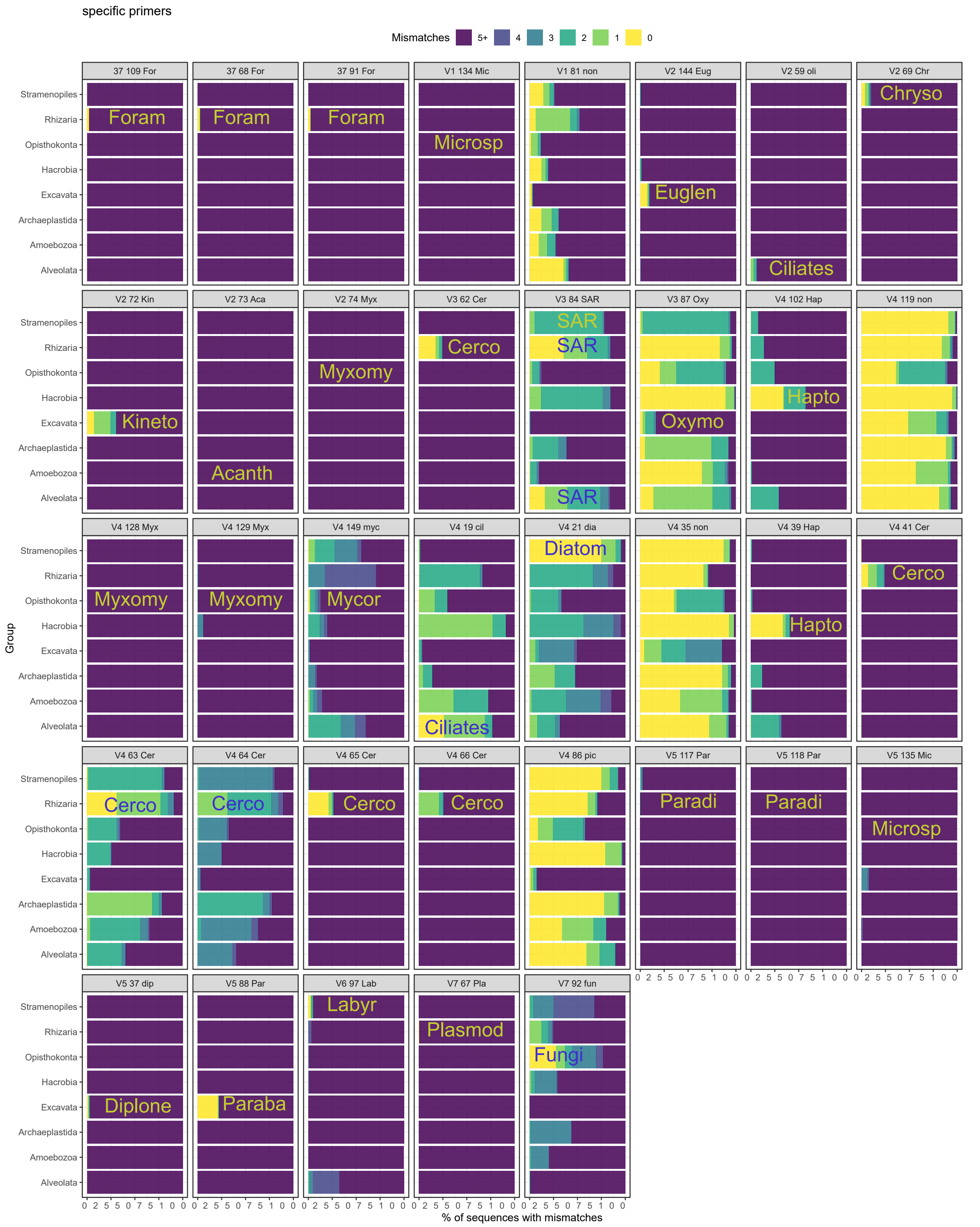

Figure S5: Number of mismatches for specific primer sets as a function of the supergroup. Target group is indicated inside the corresponding supergroup bar (e.g., Foraminifera are inside Cercozoa).

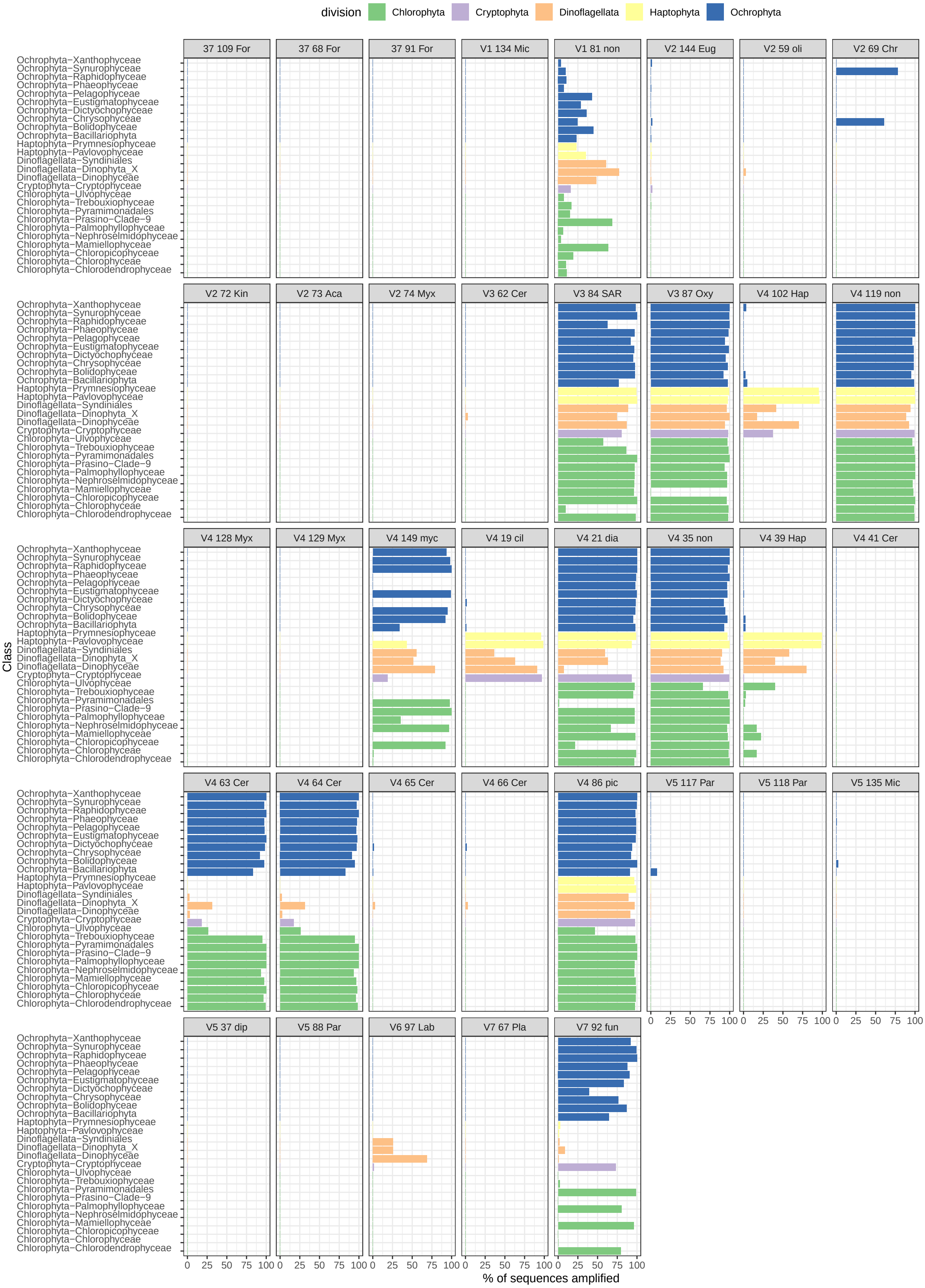

Figure S6: Percentage of sequences amplified with specific primer sets for different photosynthetic classes belonging to the Ochrophyta, Haptophyta, Dinoflagellata and Chlorophyta divisions.