The use of CRISPR-Cas Selective Amplicon Sequencing (CCSAS) to reveal the eukaryotic microbiome of metazoans

Kevin Xu Zhong^{1*}, Anna Cho^{2,3}, Christophe M. Deeg², Amy M. Chan¹, Curtis
A. Suttle^{1,2,3,4*}

1. Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada
2. Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada
3. Department of Botany, University of British Columbia, Vancouver, BC, Canada
4. Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, BC, Canada

*For correspondence: suttle@science.ubc.ca; xzhong@eoas.ubc.ca

Submitted to: Nature Methods - Article

This manuscript contains ~2000 words in the main text, 5 display items and 41 references.

Keywords: Microeukaryote, Microbiome, CRISPR-Cas, Taxon-specific single-guide RNA, Host-associated eukaryotic microbiome, 18S rRNA gene, gRNA-target-site, CasOligo, CCSAS

Running title: Revealing the eukaryotic microbiome of metazoans

Abstract

Characterization of the eukaryotic microbiome is required to understand the role of microbial communities in health and disease. Such investigation relies on sequencing 18S ribosomal RNA genes (rDNA), which serve as taxonomic markers; however, this is compromised by contaminating host rDNA sequences. To overcome this problem, we developed CRISPR-Cas Selective Amplicon Sequencing (CCSAS), a high-resolution and efficient approach for characterizing eukaryotic microbiomes. CCSAS uses taxon-specific single-guide RNA (sgRNA) to direct Cas9 to cut 18S rDNA sequences of the host. Validation shows that >96.5% of rDNA amplicons from ten model organisms were cleaved, while rDNA from protists and fungi were unaffected. In oyster spat, CCSAS resolved ~8.5-fold more taxa, and several additional major phylogenetic groups when compared to the best available alternative approach. We designed taxon-specific sgRNA for ~16,000 metazoan and plant taxa, making CCSAS widely available for characterizing eukaryotic microbiomes that have largely been neglected because of methodological challenges.

Introduction

There is an explosion in interest and understanding of how the composition of the microbiome affects the health of plants¹⁻² and animals³⁻⁶, including humans⁷⁻⁹. For example, the composition of the microbiome is associated with both positive and adverse health effects in humans¹⁰⁻¹⁴. As well, a wide span of biological, ecological and evolutionary questions have been addressed through microbiome studies^{6-8, 15-17}.

Deep sequencing of ribosomal RNA gene (rDNA) fragments is widely used to characterize the microbiome associated with a wide range of organisms¹⁸⁻²¹. Yet, to date, our knowledge on the eukaryotic component of the microbiome, particularly protists, is relatively limited compared to that of prokaryotes^{6, 21-25}. This is due to the challenge in profiling host-associated eukaryotic microbes, as the standard "universal" primers²⁶ used to amplify 18S rDNA sequences from eukaryotic microbes also amplify the host 18S rDNA, which will predominate the sequencing library¹⁸. Sampling strategies to minimize host-tissue contamination, including using fecal samples for the gut microbiome, or using gentle brushing to sample the skin microbiome are not possible when the host organism is too small or the microbes are too closely associated with host tissue. Other strategies such as using taxon-specific primers to target microbial 18S sequences (e.g. non-metazoan primers²⁷), using other marker genes such as the ITS region of fungi²⁸, or using blocking primers²⁹ to prevent the amplification of host 18S rDNA sequences have significant challenges when trying to target the entire eukaryotic microbial community. For example, blocking primers²⁹ focus on a short fragment within the hypervariable region that has a reduced taxonomic resolution compared to "universal" 18S primers. Second, designing a primer that blocks amplification of host DNA, but discriminates a wide variety of taxa is often not possible and hinders the broad application of this approach. Third, even with blocking primers, a large proportion of sequences can be host-derived (e.g. up to 92% in coral, 42% in

krill, and 45% in fish) as briefly reviewed³⁰. Consequently, studies on the eukaryotic microbiome are languishing in the absence of an approach that allows effective amplification of microeukaryotic 18S sequences with minimal contamination from host sequences.

Clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9) system offers bacteria and archaea adaptive immunity against viruses and plasmids by cleaving invading double-stranded (ds) DNA³¹. The sequence-specific cleavage is performed by Cas9 endonuclease in the presence of guide RNA (gRNA). This gRNA is a duplex comprising a trans-activating RNA (tracrRNA) that is a scaffold for binding the Cas9 protein, and an approximately 20 nucleotide (nt) crRNA guide sequence that is complementary to the DNA target site³²⁻³⁵. Cas9 can be programmed to target any DNA sequence by modifying the 20-nt guide sequence³⁵⁻³⁶. Due to its precision in DNA cutting, the simplicity in programming and the ability to artificially fuse the gRNA duplex (tracrRNA-crRNA) into a single-guide RNA (sgRNA)³⁵, CRISPR-Cas9 has emerged as a powerful biotechnological tool in a wide variety of applications³⁶⁻³⁷. We used this approach and designed a custom sgRNA to direct Cas9 to specifically cut host 18S rDNA sequences in the region targeted by "universal" primers. The cleaved host 18S fragments contain only a 3' or 5' primer binding region, resulting in short single-stranded (ss) DNA products produced by PCR, which are then removed by size-selection using SPRI magnetic beads during the preparation of the sequencing library. This results in a library highly enriched in 18S amplicons from protists and fungi, allowing for high-resolution surveys of the taxonomic composition of the eukaryotic microbes associated with any eukaryotic host. This sequencing library preparation method is termed as CRISPR-Cas Selective Amplicon Sequencing (CCSAS).

Results

Design of the taxon-specific sgRNA. The key to CCSAS is the 20-nt guide sequence of gRNA that directs Cas9 to cut the 18S rDNA sequences of the host, but not those of the associated microeukaryotes. We developed CasOligo (<https://github.com/kevinzhongxu/CasOligo>), an R package that contains the algorithm `Cas9.gRNA.oligo1()`, which identifies 20-nt sequences in the 18S rDNA region spanned by "universal" primers that can serve as target-sites for gRNA, and which are compatible with the CRISPR-Cas9 system and specific to the 18S rRNA gene of the host but not to microeukaryotes ([Supplementary Materials and Methods](#)). This 20-nt gRNA-target-site oligonucleotide sequence, complementary to the sgRNA's guide sequence, determines the specificity of the sgRNA and thereby the CRISPR-Cas action, and can be used to design and synthesize a taxon-specific sgRNA in the laboratory ([Supplementary Materials and Methods](#)).

To validate the design of sgRNA in taxon-specific cleavage, taxon-specific sgRNAs were designed and tested for each of the following ten model organisms: human (*Homo sapiens*), salmon (*Salmo salar*), shrimp (*Solenocera crassicornis*), chicken (*Gallus gallus domesticus*), cow (*Bos taurus*), mouse (*Mus musculus*), fruit fly (*Drosophila melanogaster*), rock cress (*Arabidopsis thaliana*), oyster (*Crassostrea gigas*) and the nematode (*Caenorhabditis elegans*), as well as being tested against a mock community composed of nine pathogenic protists and fungi ([Supplementary Table 1 and 2](#)). The results showed that the CRISPR-Cas9 treatment effectively cleaved the host 18S amplicons, while amplicons from the mock community of protists and fungi remained intact ([Fig. 1](#)). Comparisons using qPCR with and without CRISPR-Cas9 treatment showed that only 0.6% to 3.5% of the intact 18S amplicons remained after CRISPR-Cas9 cutting ([Fig. 1D](#)). Thus, the sgRNAs effectively targeted host sequences, while leaving sequences from microeukaryotes intact.

Using CCSAS to reveal host-associated microeukaryotic populations. Cas9 in association with the host-specific sgRNA will specifically cut the amplified host 18S sequences, while leaving intact the DNA amplified from microeukaryotes. This allows for high-resolution characterization of the taxonomic composition of the microeukaryotic community with a fraction of the sequencing effort typically used. However, after Cas9 treatment 0.6% to 3.5% of the 18S amplicons remaining were still primarily host-derived, and these could still dominate the sequencing library (data not shown). Hence, to further reduce host-derived 18S rRNA gene sequences we introduced a two-step CCSAS approach (Fig. 2) which employs a two-step CRISPR-Cas9 procedure. First, Cas9 with a taxon-specific sgRNA that is complementary to the host 18S rRNA gene sequence at the 20-nt target-site is used to cut the host genomic 18S rDNA, and then the remaining uncut 18S sequences are amplified using PCR. Any amplification of the cut fragments yields short pieces of ssDNA that are removed during size-selection using SPRI magnetic beads (Fig. 2). Second, following the first size selection, another Cas9 cut, PCR amplification and size selection were conducted, resulting in almost the complete removal of host 18S amplicons, while leaving the protistan and fungal amplicons intact and enriched for sequencing (Fig. 2). This two-step CCSAS was used to examine the eukaryotic microbiome from eight different samples of oyster spat (*C. gigas*) collected from a hatchery that was experiencing mortality events.

The results showed almost the complete removal of oyster 18S amplicons by CCSAS, while leaving the protistan and fungal amplicons intact and highly enriched for sequencing (Fig. 3). Furthermore, the taxonomic resolution of the eukaryotic microbial community associated with the oysters was much higher when using CCSAS in conjunction with "universal" 18S primers than with the best available alternative approach using primers that target non-metazoan

eukaryotes (Fig. 3, Fig. 4). Relative to using non-metazoan primers, CCSAS revealed ~8.5-fold more unique taxa of microeukaryotes (Fig. 4B). Of these, ~5.7-fold more were assigned to established taxa (Fig. 4C), such as Picozoa, Apusomonadidae, Chloroplastida, Centrohelida and Amoebozoa, as well as several groups within the Alveolata including Protalveolata, Dinoflagellata and Apicomplexa (Fig. 4D). Thus, using amplification with "universal" 18S primers combined with CCSAS provided much better taxonomic resolution of the eukaryotic microbiome, and more efficient sequencing compared to using non-metazoan primers. Moreover, with non-metazoan primers, most sequences were from metazoa (Fig. 4A), although the taxonomic resolution was not adequate to assign them to be from oysters (Fig. 3). In contrast, for CCSAS, the highest percentage of sequences from metazoa (mostly assigned to oysters and nematodes of order Monhysterida; Fig. 3) was only 7.4%, while in three out of eight samples, sequences from metazoa were undetectable (Fig. 4A).

Database of gRNA-target-sites for metazoa and plants. To ease the application of CCSAS to the study of eukaryotic microbiomes in a wide range of metazoa and plants, we used CasOligo to identify the gRNA-target-sites for almost all 15907 metazoa and plant taxa (metaphyta of Embryophyta group) in the SILVA SSU database³⁸ (version v119, released on 24 July 2014) (Fig. 5). This created a database of gRNA-target-sites, comprising, on average, 33 taxon-specific gRNA-target-sites per taxon (Supplementary Fig. 1), which provides a wide choice for designing taxon-specific sgRNA for various host organisms. For each metazoan and plant species in SILVA, we revealed between 3 and 217 different gRNA-target-sites that are compatible with the CRISPR-Cas9 system (Supplementary Fig. 1). For 99.6% of the 15907 metazoan and plant species in SILVA, we found between 1 and 214 gRNA-target-sites that specifically target the 18S sequence for each taxon, but not protistan or fungal 18S sequences. Although it is not possible to design a "universal" sgRNA that targets all metazoa and plants

but not microeukaryotes, some sgRNAs target broad taxonomic groups ([Supplementary Fig. 2](#)). For example, based on in-silico analysis, sgRNA_058534 targets 3099 species from 22 classes and families of Animalia, primarily 72.7% of the 4014 Insecta species in SILVA ([Supplementary Fig. 2](#)). CasOligo, also allows retrieval of the gRNA-target-site oligonucleotide sequence for specific taxa by entering the scientific name of the host species using the `search.db.byname()` function. Nonetheless, it is best to identify the taxon-specific gRNA-target-site based on the 18S rRNA gene sequence of the host, because the action of CRISPR-Cas9 is highly sequences-specific and the database does not cover all sequence variants for specific taxa.

Discussion

CCSAS provides a new way to obtain high-resolution taxonomic data for the eukaryotic microbiomes of plants, animals and other metazoa. By employing CRISPR-Cas9 with taxon-specific gRNAs, the background of host 18S sequences is greatly reduced or eliminated; thus, only a fraction of the sequencing effort is required to obtain high-resolution taxonomic data for the eukaryotic microbiome. Moreover, the creation of a database of gRNA-target-sites, and the primary gRNA-target-site oligonucleotide design functions of the CasOligo package, makes it easy to profile the eukaryotic microbiome of metazoa and plants. Additionally, the CasOligo package provides an oligonucleotide design function, `Cas9.oligo.search2()`, that can be used to design custom sgRNA for any gene for which the sequence is known, and for which there is a reference database for comparison, so that the specificity of the sgRNA can be ascertained. This includes genes encoding other regions of rRNA, such as the 16S and 23S rRNA genes, or metabolic genes (e.g. COX1). Thus, CCSAS makes it possible to study the genetic diversity of any gene in complex systems, including genes that are rare, by removing any sequence that

would otherwise dominate the data. The sequence-specific removal of any amplicon has a wide range of applications, including pathogen diagnosis, and studies of symbiosis and microbiome therapy.

There are a few considerations in applying CCSAS to microbiome studies. First, there are examples of gRNA recognizing the wrong target³⁹⁻⁴⁰, which might lead Cas9 to cut some protistan and fungal sequences, or incompletely cleave host sequences. As well, incomplete removal of cut host amplicons by SPRI magnetic beads might reduce sequencing efficiency. Other more efficient size-selection methods may overcome this potential difficulty. Despite these caveats, in our hands CCSAS was far superior to other currently used approaches for obtaining high-resolution eukaryotic microbiome data.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at the online [Supplementary Information](#).

References

1. Philippot, L., Raaijmakers, J. M., Lemanceau, P. & van der Putten, W. H. Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* **11**, 789–799 (2013).
2. Haney, C. H. & Ausubel, F. M. Plant microbiome blueprints. *Science* **349**, 788–789 (2015).
3. McFall-Ngai, M. et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci. USA* **110**, 3229–3236 (2013).
4. Bourne, D. G., Morrow, K. M. & Webster, N. S. Insights into the coral microbiome: underpinning the health and resilience of reef ecosystems. *Annu. Rev. Microbiol.* **70**, 317–340 (2016).

- 230 5. Apprill, A. Marine animal microbiomes: toward understanding host–microbiome interactions in a
231 changing ocean. *Front. Mar. Sci.* **4**, (2017).
- 232 6. Parfrey, L. W., Moreau, C. S. & Russell, J. A. Introduction: The host-associated microbiome: pattern,
233 process and function. *Mol. Ecol.* **27**, 1749–1765 (2018).
- 234 7. Knight, R. et al. The microbiome and human biology. *Annu. Rev. Genomics Hum. Genet.* **18**, 65–86
235 (2017).
- 236 8. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
- 237 9. Proctor, L. M. et al. The integrative human microbiome Project. *Nature* **569**, 641–648 (2019).
- 238 10. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity.
239 *Nature* **444**, 1022–1023 (2006).
- 240 11. Frank, D. N. et al. Molecular-phylogenetic characterization of microbial community imbalances in
241 human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. USA* **104**, 13780–13785 (2007).
- 242 12. Arthur, J. C. et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science*
243 **338**, 120–123 (2012).
- 244 13. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.*
245 **13**, 260–270 (2012).
- 246 14. Valles-Colomer, M. et al. The neuroactive potential of the human gut microbiota in quality of life and
247 depression. *Nat. Microbiol.* **4**, 623–632 (2019).
- 248 15. Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution
249 of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6**, 776–788 (2008).
- 250 16. Sommer, F. & Bäckhed, F. The gut microbiota — masters of host development and physiology. *Nat.*
251 *Rev. Microbiol.* **11**, 227–238 (2013).
- 252 17. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism.
253 *Nature* **535**, 56–64 (2016).
- 254 18. Kuczynski, J. et al. Experimental and analytical tools for studying the human microbiome. *Nat. Rev.*
255 *Genet.* **13**, 47–58 (2011).
- 256 19. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and
257 MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
- 258 20. Knight, R. et al. Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **16**, 410–422 (2018).

21. Popovic, A. & Parkinson, J. Characterization of eukaryotic microbiome using 18S amplicon sequencing. in *Microbiome Analysis: Methods and Protocols* (eds. Beiko, R. G., Hsiao, W. & Parkinson, J.) 29–48 (Springer New York, 2018).
22. Parfrey, L. W., Walters, W. A. & Knight, R. Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. *Front. Microbiol.* **2**, 153 (2011).
23. Andersen, L. O., Vedel Nielsen, H. & Stensvold, C. R. Waiting for the human intestinal Eukaryotome. *ISME J.* **7**, 1253–1255 (2013).
24. Laforest-Lapointe, I. & Arrieta, M.-C. Microbial eukaryotes: a missing link in gut microbiome studies. *mSystems* **3**, (2018).
25. Hooks, K. B. & O'Malley, M. A. Contrasting strategies: human eukaryotic versus bacterial microbiome research. *J. Eukaryot. Microbiol.* **67**, 279–295 (2020).
26. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
27. Bower, S. M. et al. Preferential PCR amplification of parasitic protistan small subunit rDNA from metazoan tissues. *J. Eukaryot. Microbiol.* **51**, 325–332 (2004).
28. Schoch, C. L. et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. USA* **109**, 6241–6246 (2012).
29. Vestheim, H., Deagle, B. E. & Jarman, S. N. Application of blocking oligonucleotides to improve signal-to-noise ratio in a PCR. *Methods Mol. Biol.* **687**, 265–274 (2011).
30. Clerissi, C. et al. Protists within corals: the hidden diversity. *Front. Microbiol.* **9**, (2018).
31. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
32. Brouns, S. J. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
33. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
34. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
35. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

36. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
37. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **34**, 933–941 (2016).
38. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
39. Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
40. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).

Acknowledgments

V. Au, M.L. Berbee, J. Caleta, M. Edgley, G. Hu, J. Kronstad, D.G. Moerman, G. Xiang, C. Yang and Y. Zhang of the University of British Columbia, and K.M. Miller from Fisheries and Oceans, Canada, are gratefully acknowledged for providing organisms for testing (Supplementary Table 1). This work was supported by the Gordon and Betty Moore Foundation (grant: GBMF#5600) and an infrastructure award from the Canada Foundation for Innovation and British Columbia Knowledge Development Fund to C.A.S. A.C. was supported by an award from the Pacific Research Board (grant: NPRB GSRA#1335).

Author Contributions

During a stimulating discussion to profile the eukaryotic microbiota associated with pacific oysters, C.M.D. proposed the idea of using CRISPR-Cas9 technology to reduce the background of 18S rRNA genes and conducted a pilot trial. K.X.Z. conceived, developed and implemented the CCSAS method, the CasOligo package, and the database of gRNA-target-sites. A.C.

contributed to the optimization of the CCSAS method. C.A.S and A.M.C. were involved in discussions of experimental design and interpretation of the data. K.X.Z and C.A.S. wrote the manuscript with input from all authors.

Competing Interests

The authors declare no competing financial interests.

Additional information

Supplementary Information is available in the online version of the paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.A.S. (suttle@science.ubc.ca) or K.X.Z (xzhong@eoas.ubc.ca).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Fig. 1 | Gel images of 18S amplicons from ten model organisms (A and B) and a mock community of protists and fungi (C) to which Cas9 with the taxon-specific sgRNA (as shown in Supplementary Table 2) was either added (+) or not (-). Gel bands show the amplicon length in base pairs (bp) relative to a DNA ladder. Quantitative PCR was used to assess the remaining intact 18S amplicons from the model organisms after cutting with CRISPR-Cas9 (D). The labels on the X-axes of C and D indicate the ID of the taxon-specific gRNAs and its corresponding host.

Fig. 2 | Workflow of the two-step CRISPR-Cas Selective Amplicon Sequencing (CCSAS) to study the host-associated eukaryotic microbiome.

Fig. 3 | Eukaryotic taxa representing >1% of the sequences revealed by amplicon deep-sequencing of the 18S rRNA genes for oyster spat samples, using "universal" 18S primers²⁶ (Supplementary Table 3), non-metazoan primers²⁷ (Supplementary Table 3), as well as CRISPR-Cas Selective Amplicons Sequencing (CCSAS), which combined "universal" 18S primers and CRISPR-Cas9 with pacific-oyster-specific sgRNA m258 (Supplementary Table 2).

Fig. 4 | Distribution of eukaryotic groups (**panel A**) in eight oyster samples revealed using deep-sequencing of 18S amplicons of rRNA genes using non-metazoan primers (NM), and CCSAS that combines 18S "universal" primers and cleavage using CRISPR-Cas9 (CCSAS). The number of unique operational taxonomic units (OTUs) of microeukaryotes (**panel B**) and their assigned taxa (**panel C**) are summarized for the NM and CCSAS methods. **Panel D** compares the number of OTUs detected for different microeukaryotic groups using the CCSAS and NM methods.

384
385
386

Supplementary Information for

The use of CRISPR-Cas Selective Amplicon Sequencing (CCSAS) to reveal the eukaryotic microbiome of metazoans

Kevin Xu Zhong^{1*}, Anna Cho^{2, 3}, Christophe M. Deeg², Amy M. Chan¹, Curtis A. Suttle^{1,2,3,4*}

1. Department of Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada
2. Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada
3. Department of Botany, University of British Columbia, Vancouver, BC, Canada
4. Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, BC, Canada

✉Correspondence to:

Curtis Suttle
Department of Earth, Ocean & Atmospheric Sciences
University of British Columbia
#2020-2207 Main Mall
Vancouver, BC
Canada V6T 1Z4
E-mail: suttle@science.ubc.ca

Kevin Xu Zhong
Department of Earth, Ocean & Atmospheric Sciences
University of British Columbia
#2020-2207 Main Mall
Vancouver, BC
Canada V6T 1Z4
E-mail: xzhong@eoas.ubc.ca

This PDF file includes Supplementary Information:

Supplementary Materials and Methods

Supplementary References (41 - 46)

Data availability

Supplementary Table 1-3

Supplementary Figure 1-2

Supplementary Materials and Methods

Organisms and samples

Ten model organisms, human (*Homo sapiens*), salmon (*Salmo salar*), shrimp (*Solenocera crassicornis*), chicken (*Gallus gallus domesticus*), cow (*Bos taurus*), mouse (*Mus musculus*), fruit fly (*Drosophila melanogaster*), rock cress (*Arabidopsis thaliana*), oyster (*Crassostrea gigas*) and the nematode (*Caenorhabditis elegans*), as well as nine different species of protists and fungi were obtained from either commercial markets or various laboratories at University of British Columbia ([Supplementary Table 1](#)). As well, eight samples of seven- to 28-day old oyster spat, with sizes ranging between 400 and 1000 μM , were obtained from a hatchery that was experiencing mortality events. The oyster spat were immediately frozen using liquid nitrogen following collection, and stored at -80°C until analysis.

Genomic DNA extraction

DNA from the model organisms, protists, fungi and oyster spat were extracted using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's directions, and quantified using the Qubit™ DNA HS Assay Kit (Invitrogen).

To make a mock community of microeukaryotes, equal amounts of DNA from each protist and fungus ([Supplementary Table 1](#)) were pooled.

Design and synthesis of taxon-specific sgRNA

The specificity of CRISPR-Cas9 is determined by a 20-nt guide sequence within the sgRNA, which directs Cas9 to cut a target DNA at the 20-nt target site that is complementary to this guide sequence. Thus, the design of a taxon-specific sgRNA requires identifying a 20-nt gRNA-target-site oligonucleotide sequence in the host 18S rRNA gene, which is, absent in microeukaryotes. This taxon-specific 20-nt gRNA-target-site oligonucleotide sequence, reverse-complement to the sgRNA's guide sequence, determines the specificity of the sgRNA and thereby the CRISPR-Cas action that is to cut 18S rDNA from the host but not from microeukaryotes. This taxon-specific 20-nt gRNA-target-site oligonucleotide sequence is used to synthesize the taxon-specific sgRNA in the laboratory using the NEB EnGen™ sgRNA Synthesis Kit.

Obtaining the host 18S rRNA gene sequences

Prior to the design of the sgRNA, we need to obtain the 18S rRNA gene sequence of the host organism for the identification of the gRNA-target-sites. In this study, we employed the cloning-sequencing approach as follows.

For each host, 18S rRNA gene fragments were PCR amplified using the "universal" primers TAREuk454FWD1 and TAREukREV3²⁶, and sequenced to facilitate the design of gRNA-target-site oligos targeting each host. Briefly, PCR was conducted in four separate reactions run at annealing temperatures of 45, 47, 48 or 49°C, to ensure amplification of a 380-450 bp fragment from the V4 region of the 18S rRNA gene. Each 25 µL reaction mix was made with 1X PCR buffer, 4 mM MgCl₂, 50 µg of Bovine Serum Albumin (Invitrogen), 200 mM of each dNTP (Invitrogen), 0.4 µM of each primer, 0.5 U of Q5® high fidelity polymerase (NEB) and 10 ng of genomic DNA template. As previously described²⁶, the initial denaturation and activation was at 95°C for 5 min, followed by 10 cycles consisting of 95°C for 30 s, 57°C for 45 s, and 72°C for 1 min, followed by 25 cycles of denaturation at 95°C for 30 s, annealing at 45, 47, 48 or 49°C for 45 s, elongation at 72°C for 60 s, and a final elongation for 10 min at 72°C. The PCR products from the four reactions were then pooled, and the 18S amplicons purified using Agencourt SPRI magnetic beads (Beckman Coulter) at a 1:1 (vol:vol) ratio of beads:DNA to remove fragments <200bp.

These purified amplicons were then cloned into pCR2-TOPO vectors using the TOPO TA Cloning Kit (Invitrogen). Four 18S rRNA gene clones from each model organism were sent for Sanger sequencing at the NAPS Unit sequencing facility at The University of British Columbia. The obtained DNA sequences were then used to design the taxon-specific 20-nt gRNA-target-

site sequence, which was be used for synthesizing the taxon-specific sgRNA that guides Cas9 to cleave the host 18S sequences, as outlined below.

Design of the taxon-specific gRNA-target-site oligonucleotide sequences

We developed the R package CasOligo (<https://github.com/kevinzhongxu/CasOligo>) for designing taxon-specific 20-nt gRNA-target-site oligonucleotide sequences, which allow sgRNA to recognize 18S sequences from specific taxa. Taxon-specific gRNA-target-site oligonucleotide sequences were designed for each model organism using the Cas9.gRNA.oligo1() function in CasOligo by providing a fasta file of the V4 region of the 18S rRNA gene from each organism that is amplified by the "universal" 18S primers, TAREuk454FWD1 and TAREukREV3²⁶. The same approach can be used to design taxon-specific gRNA-target-site oligonucleotide sequences for any host organism. First, Cas9.gRNA.oligo1() searches the forward and reverse strands of the 18S rRNA gene for 20-nt gRNA-target-site oligonucleotide sequences that are compatible with Cas9 nuclease; compatibility requires that the protospacer-adjacent-motif (PAM), NGG, is immediately adjacent to the 3' downstream region of the 20-nt target-site sequence. Each of these 20-nt gRNA-target-site sequences is potentially a target for the combined actions of sgRNA and Cas9. Next, each potential gRNA-target-site sequence is searched against the SILVA SSU database for the V4 region of 18S rRNA genes, in order to determine if the sequence is absent in protistan and fungal microeukaryotes. If so, this gRNA-target-site sequence can be used to synthesize a sgRNA that will guide Cas9 to specifically cut the host 18S rDNA. The gRNA-target-site oligonucleotide sequences designed in this study are shown in [Supplementary Table 2](#).

Synthesis of sgRNA-template oligonucleotides

Once suitable taxon-specific 20-nt gRNA-target-site oligonucleotide sequences were identified, the sgRNA-template oligonucleotide sequences were obtained using the EnGen™ sgRNA Template Oligo Designer (<https://nebiocalculator.neb.com/#!/sgrna>), which adds a T7 promoter sequence at the 5' end, and a 14-nt overlap sequence at the 3' end of the 20-nt gRNA-target-site sequence. For our studies, this sgRNA-template oligonucleotide was synthesized by Integrated DNA Technologies (IDT), and diluted to 1 µM with molecular grade ultrapure water (Invitrogen).

Synthesis of sgRNA

The 1 µM sgRNA-template oligonucleotide was used as a DNA template to synthesize the sgRNA using the EnGen™ sgRNA Synthesis Kit, *S. pyogenes* (New England Biolabs) by following the manufacturer's instructions. The resulting sgRNA was treated with amplification grade DNase I (Invitrogen) at room temperature for 15 min to remove any remaining DNA, and then purified using RNA Clean & Concentrator-25 Kit (Zymo Research) by following the manufacturer's instructions. Finally, the fragment size of the sgRNA was assessed using an Agilent RNA 6000 Pico Kit (Agilent) and its concentration measured using a Qubit™ RNA HS Assay Kit (Invitrogen).

Validation of the design of taxon-specific sgRNA

To validate the design of gRNA for taxon-specific cleavage, we first generated 18S amplicons for each model organism and the mock community of protists and fungi. Then, these 18S amplicons were used to ascertain the effect of CRISPR-Cas9, in conjunction with taxon-specific sgRNA, on cleavage of the amplicons. The results were visualized on a gel using a Bioanalyzer (Agilent) and assessed using quantitative PCR (qPCR) as detailed below.

542

543 *Preparation of the host 18S amplicons*

544 For each of the ten host organisms and the mock community of protists and fungi, 18S rRNA
545 gene fragments were obtained using PCR with the "universal" primers TAREuk454FWD1 and
546 TAREukREV3²⁶ following the conditions as detailed above. The 18S amplicons were purified
547 using Agencourt SPRI magnetic beads (Beckman Coulter) at a 1:1 (vol:vol) ratio of
548 beads:DNA.

549

550 *DNA cleavage using CRISPR-Cas9*

551 For each of the ten host organisms and the mock community of protists and fungi, the purified
552 18S amplicons were cut using Cas9 Nuclease, *S. pyogenes* (New England Biolabs) in the
553 presence of a sgRNA, following the manufacturer's directions. Briefly, the 10 µL reaction
554 contained approximately 0.1 pmol of dsDNA, 1 pmol of sgRNA, and 1 pmol of Cas9, as well
555 as 1x Cas9 reaction buffer to keep the molar ratio of Cas9:sgRNA:template DNA at 10:10:1.
556 The reaction was incubated at 37°C for 4 h in a thermocycler, and then followed by 70°C for
557 10 min to deactivate the CRISPR-Cas9. For each sample, in parallel with the CRISPR-Cas9
558 treatment we also prepared the reaction without CRISPR-Cas9 treatment, in which Cas9
559 nuclease and sgRNA were replaced with molecular grade ultrapure water (Invitrogen). Thus,
560 each reaction of both treatments contained the same amount of template dsDNA (18S amplicons
561 at 0.1 pmol) and was subjected to the same incubation conditions.

562

563 *Visualization using gel electrophoresis*

564 The size of the 18S rDNA fragments with and without CRISPR-Cas9 treatment was visualized
565 by gel electrophoresis using a Bioanalyser (Agilent). Prior to loading into the gel, the Cas9-cut
566 product (5 µL out of 10 µL) were treated with 1 mg/mL (final) Proteinase K (Invitrogen) at

room temperature for 15 min to digest the Cas9 nuclease. Then, 1 to 2 μ L of this proteinase-K-treated product was added into a well of an Agilent High Sensitive DNA Chip in a Bioanalyzer (Agilent) to visualize and verify cutting by CRISPR-Cas9.

Quantitative PCR

To determine the efficiency of CRISPR-Cas9 for eliminating host-derived 18S sequences, we used quantitative PCR (qPCR) and the primer set TAREuk454FWD1/TAREukREV3²⁶ (Supplementary Table 3) that targets a 380-450 bp fragment of the V4 region of the 18S rRNA gene, to assess the proportion of 18S amplicons cut by Cas9. The 10 μ L qPCR reactions contained 1 X SsoFast™ EvaGreen® Supermix (Bio-Rad), 0.5 μ M of each primer, and a 1 μ L 1/10000 dilution of DNA template consisting of amplified products, either with or without the addition of Cas9. Thermal cycling was conducted in a CFX96 real-time PCR detection system (Bio-Rad) with the following program: 3 min denaturation at 95°C, followed by 40 cycles of denaturation at 95°C for 30 s, and annealing and extension at 49°C for 30 s. Nine, 10-fold serially diluted standards (ranging from 5 x 10⁰ to 5 x 10⁹ molecules per mL) were run in duplicate along with two no-template control reactions containing 1 μ L of nuclease-free water. The amplicon standards were made from a cloned 18S rRNA gene fragment amplified from a culture of the prasinophyte microalga, *Micromonas pusilla*, using the primer set TAREuk454FWD1/TAREukREV3²⁶. The amplicons were purified using a MiniElute® PCR Purification Kit (Qiagen), and quantified using a Qubit® dsDNA High Sensitivity Assay Kit (Invitrogen). The size of the amplicon was checked using gel-electrophoresis, and the qPCR melting curves were used to confirm that the fluorescence signal corresponded to a single-sized DNA fragment. The qPCR amplification efficiency was between 0.95 and 1.05 for the cloned amplicons (with $r > 0.98$, $n = 9$).

Sequencing library preparation using CRISPR-Cas Selective Amplicon Sequencing (CCSAS)

To profile host-associated eukaryotic microbiomes, we developed CRISPR-Cas Selective Amplicon Sequencing (CCSAS), which combines the use of CRISPR-Cas9 and universal 18S primers to prepare a sequencing library that is compatible with Illumina sequencing platforms. The method uses a taxon-specific sgRNA to guide Cas9 nuclease to selectively cleave 18S rRNA gene sequences from metazoa and plants, which then can be removed by size selection with SPRI beads; sequences from microeukaryotes are left intact, and can be amplified by PCR. Therefore, CCSAS allows high-resolution profiling of host-associated eukaryotic microbiomes with relatively low sequencing effort. In this study, we present CCSAS (Fig. 2); the two-step CRISPR-Cas procedure first uses Cas9 to cut the host gene encoding 18S rDNA, followed by a second cut of any host-derived 18S amplicons. Details of the method are provided below.

Cas9 cut on host genomic DNA

Genomic DNA of the host is cut using Cas9 Nuclease, *S. pyogenes* (NEB) following the manufacturer's directions. Briefly, a 10-μL reaction mix containing approximately 0.1 pmol of genomic DNA, 1 pmol of sgRNA and 1 pmol of Cas9, as well as 1x Cas9 reaction buffer to keep the molar ratio of Cas9:sgRNA:template DNA at 10:10:1 is incubated at 37°C for 4 h in a thermocycler.

The first PCR and size selection

The Cas9-cleaved genomic DNA is used as a template in the first PCR to generate 380-450 bp amplicons from the V4 region of the 18S rDNA gene that are depleted in host sequences. To ensure representative amplification of 18S sequences from microeukaryotes, four parallel PCR

reactions are run at different annealing temperatures (45, 47, 48 or 49°C), using the “universal” 18S primers TAREuk454FWD1-Nxt and TAREukREV3-Nxt ([Supplementary Table 3](#)). Compared to TAREuk454FWD1 and TAREukREV3²⁶, this modified primer set contained overhang adapter sequences ([Supplementary Table 3](#)), which are compatible with Illumina indexes and sequencing adapters. These adapters allowed for a second PCR to append Illumina Nextera XT indexes to each side of the amplicons as forward and reverse primers, thus creating a dual-indexed library. This dual-indexed library preparation approach is adapted from the ref [41](#).

Details on the first PCR reactions are as follows. Briefly, each 25 µL reaction mix contained 1X PCR buffer, 4 mM MgCl₂, 50 µg of Bovine Serum Albumin (Invitrogen), 200 mM of each dNTP (Invitrogen), 0.4 µM of each primer, 0.5 U of Q5® high fidelity polymerase (NEB) and 5 µL of the Cas9-cleaved genomic DNA. Because the reverse primer is 2bp shorter than the forward primer and has a lower annealing temperature, we used the two-step PCR approach of Stoeck et al.²⁶, in which there is an initial ten PCR cycles at an annealing temperature where only the forward primer will bind and amplify, followed by 25 cycles at one of four lower annealing temperatures (45, 47, 48 or 49°C) where both forward and reverse primers amplify. The program has an initial denaturation and activation at 95°C for 5 min, followed by ten three-step cycles consisting of 94°C for 30 s, 57°C for 45 s, and 72°C for 1 min, followed by 25 cycles of denaturation at 94°C for 30 s, annealing at either 45, 47, 48 or 49°C for 45 s and elongation at 72°C for 60 s, with a final elongation for 10 min at 72°C. In the end, the PCR product of four reactions are pooled together. Then amplicons are size-selected and purified using magnetic Agencourt SPRI beads (Beckman Coulter) at an 0.8:1 (vol:vol) ratio of beads:DNA to remove fragment < 300bp.

Cas9 cut on 18S amplicons

To further remove 18S host amplicons, the size-selected amplicons described above were cut again using Cas9 Nuclease, *S. pyogenes* (NEB). Briefly, the 10 µL reaction contained approximately 0.1 pmol of DNA amplicons, 1 pmol of sgRNA, 1 pmol of Cas9, 1x Cas9 reaction buffer to keep the molar ratio of Cas9:sgRNA:template DNA at 10:10:1. The reaction was incubated at 37°C for 4 h in a thermocycler.

The 2nd PCR and size selection

The product of the second Cas9-cut was used as the template for a second PCR (index PCR) to generate the indexed amplicons libraries. The 50-µL reaction mix of the second PCR comprised 1X PCR buffer, 4 mM MgCl₂, 200 mM of each dNTP (Invitrogen), 5 µL of each index primer (N7XX and S5XX of Nextera® XT Index Kit), 1 U of Q5® high fidelity polymerase (NEB) and 5 µL of the product of the second Cas9 cut. The second PCR (index PCR) consisted of an initial denaturation and activation at 95°C for 3 min, followed by 29 three-step cycles consisting of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a final elongation for 10 min at 72°C. The indexed amplicons generated by the second PCR were size-selected and purified using magnetic Agencourt SPRI beads (Beckman Coulter) at a ratio of 0.8:1 (vol:vol) for beads:DNA to remove fragments < 300bp.

During size selection with SPRI magnetic beads, the bead:DNA ratio depends on the size of the fragments that need to be separated. As the size of the fragments generated by cutting the ~424-bp metazoan 18S rRNA gene sequences will vary depending on the cut site, the beads:DNA ratio of specific sgRNA may need to be optimized to remove all of the cleaved fragments. It is important to remove sequence fragments generated by amplification of the cleaved host 18S rRNA genes, as these can reduce sequencing efficiency.

Sequencing library preparation for amplicons generated using the universal 18S primers

Sequencing libraries for 18S amplicons generated using the “universal” 18S primers and without cutting using CRISPR-Cas9 were prepared using protocols adapted from Illumina⁴¹. Briefly, two successive runs of PCR were performed as follows: For the first PCR, 29 cycles of amplification using the modified primers TAREuk454FWD1-Nxt and TAREukREV3-Nxt (Supplementary Table 3) were used to generate 380 to 450 bp amplicons of the V4 region of the 18S rRNA genes. The reaction conditions for the first PCR were as detailed above for the first CCSAS PCR, except that there was about 5 ng of genomic DNA in the sample. The amplicons were purified using magnetic Agencourt SPRI beads (Beckman Coulter) at a ratio of 1:1 (vol:vol) for beads:DNA to remove fragments < 200bp.

Five µL of the purified amplicons from the first PCR were used as templates for the second PCR (index PCR). The PCR reactions and conditions for the second PCR were as detailed above for the second CCSAS PCR. The amplicon libraries generated were purified using magnetic Agencourt SPRI beads (Beckman Coulter) at a ratio of 1:1 (vol:vol) for beads:DNA to remove fragments < 200bp.

Sequencing library preparation for amplicons generated using the non-metazoan primers

Preparation of the sequencing library for the 18S amplicons obtained using non-metazoan primers²⁷ was similar to that described above, except that the first PCR used the primers 18s-EUK581-F-Nxt and 18s-EUK1134-R-Nxt (Supplementary Table 3) to amplify a 300-600 bp region of 18S rRNA genes that is specific to microeukaryotes but absent in metazoa²⁷. The 25 µL reaction mix comprised 1X PCR buffer, 4 mM MgCl₂, 50 µg of Bovine Serum Albumin (Invitrogen), 200 mM of each dNTP (Invitrogen), 0.4 µM of each primer, 0.5 U of Q5® high-

fidelity polymerase (NEB) and approximately 5 ng of genomic DNA. There initial denaturation of 2 min at 98°C was followed by 29 cycles of denaturation at 98°C for 30 s, annealing at 51.5°C for 30 s and elongation at 72°C for 60 s, and a final elongation for 10 min at 72°C. The first PCR product was purified using magnetic Agencourt SPRI beads (Beckman Coulter) at a ratio of 1:1 (vol:vol) for beads:DNA to remove fragments less < 200bp (e.g. dimers). The amplicon libraries were completed as described above, with a second PCR to add a Nextera® XT index to each 3' and 5' end of the amplicons.

Next-generation sequencing and data analysis

The DNA concentrations of the 18S amplicon sequencing libraries that were prepared using "universal" 18S primers, non-metazoan primers, or the CCSAS method were measured using the Qubit® dsDNA High Sensibility Assay Kit (Invitrogen). The fragment size for each type of library was determined using an Agilent bioanalyzer with the High Sensitive DNA Chip (Agilent). Equimolar amounts of these barcoded and purified amplicon sequencing libraries were pooled and sent for sequencing at the BRC Sequencing Core at the University of British Columbia using MiSeq Illumina 2 x 300bp chemistry.

Sequences were processed and analyzed using QIIME version 1.9⁴². Briefly, sequences were de-multiplexed by their forward and reverse indexes, and the paired-end reads merged using PEAR version 1.10.4⁴³. Then, sequences from different samples were pooled, and Uclust⁴⁴ was used for OTU picking with 99% nucleotide sequence similarity. Taxonomy was assigned for representative OTU sequences using the Uclust consensus taxonomy assigner and the SILVA SSU database³⁸ (*version v132, released on 13 December 2017*) at a 90% confidence cutoff. The samples were normalized by analyzing the relative abundance for each OTU or taxon as the

proportion of all sequences (tags) within a sample. The data is visualized using ggplot2⁴⁵ and metacoder⁴⁶.

Data Availability

All next-generation sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject (PRJNA625176). The Sanger cloning-sequencing data were deposited in GenBank under the accession numbers MT328571 to MT328580 for 18S rRNA gene sequences of ten model organisms. The authors declare that all other data supporting the findings of this study are available within the paper and/or the associated supplementary files. All related scripts, functions and algorithms for designing gRNA-target-site oligonucleotide sequences were included in the custom R package: CasOligo (<https://github.com/kevinzhongxu/CasOligo>). The gRNA-target-sites database was included in the CasOligo package as well.

Supplementary References

41. Anonymous. 16S Metagenomic Sequencing Library Preparation. https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html. (27 November 2013).
42. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Meth.* **7**, 335–336 (2010).
43. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End read merger. *Bioinformatics* **30**, 614–620 (2014).
44. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

45. Wickham, H. *ggplot2*. (Springer New York, 2009). doi:10.1007/978-0-387-98141-3.
46. Foster, Z. S. L., Sharpton, T. J. & Grünwald, N. J. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput. Biol.* **13**, e1005404 (2017).

757

758

759 **Supplementary Figure 1.** Distribution of the number of gRNA-target-sites of each metazoan and plant species from the SILVA SSU v119 database³⁸. These gRNA-target-site
 760 oligonucleotide sequences were identified, using the Cas9.gRNA.oligo1() algorithm, from the V4 region of the 18S rRNA gene that is flanked by the 18S "universal" primer
 761 set TAREuk454FWD1 / TAREukREV3²⁶, and are used for designing and synthesizing the CRISPR-Cas9-compatible sgRNA. The taxon-specific gRNA-target-sites allows the
 762 design of the sgRNA to taxon-specifically cut the 18S rRNA gene sequence of a metazoan or plant host but not microeukaryotes (protists and fungi) using CRISPR-Cas9.

763

764

Supplementary Figure 2. Summary of the number of eukaryotic species at each D7 taxonomic level that the sgRNA can cut at the V4 region of the 18S rRNA genes that are flanked by the 18S "universal" primer set TAREuk454FWD1 / TAREukREV3²⁶. These nine sgRNAs, which are among 205242 unique taxon-specific sgRNA designed from the SILVA SSU database (*version 119*)³⁸ using CasOligo, are selected to show that some sgRNAs can target more than 1000 species and broad taxonomic groups based on an *in-silico* analysis (*i.e.* 100% match to the 18S rRNA gene sequences of the metazoan host at the gRNA-target-site, but no match for protists and fungi). Taxon names on the left side of the panel are shown as SILVA taxonomic hierarchy with levels range from D0 (kingdom) to D7. The D7 taxonomic level comprises eukaryotic classes and families.

774 **Supplementary Table 1** List of organisms used in this study.

Organisms	Classification	Representative groups	Applications	Source	NCBI accession no.
<i>Cryptococcus neoformans</i> variety grubii	Fungi, Basidiomycota	Fungi	Mock community of protists-fungi	J. Kronstad, UBC	-
<i>Ustilago maydis</i> strain 002 WT	Fungi, Basidiomycota	Fungi	Mock community of protists-fungi	J. Kronstad, UBC	-
<i>Aspergillus citrisporus</i> strain E1	Fungi, Ascomycete	Fungi	Mock community of protists-fungi	M.L. Berbee, UBC	-
<i>Rhizopus</i> sp. strain AFTDL	Fungi, Zygomycete	Fungi	Mock community of protists-fungi	M.L. Berbee, UBC	-
<i>Batrachochytrium dendioatriodio</i> strain JEL 423	Fungi, Chytride	Fungi	Mock community of protists-fungi	M.L. Berbee, UBC	-
<i>Coemunsia</i> sp.	Fungi, Zygomycete	Fungi	Mock community of protists-fungi	M.L. Berbee, UBC	-
<i>Creolimax fragrantissima</i> strain Pdd4	Protist, Ichthyosporaea	Protists	Mock community of protists-fungi	M.L. Berbee, UBC	-
<i>Micromonas pusilla</i>	Protist, Mamiellaceae	Protists	Mock community of protists-fungi	This lab	-
<i>Tetrahymena</i> sp.	Protist, Tetrahymenidae	Protists	Mock community of protists-fungi	This lab	-
<i>Caenorhabditis elegans</i>	Animalia, Nematoda, Caenorhabditis	Worm	Model organisms	D.G. Moerman, UBC	MT328580
<i>Arabidopsis thaliana</i>	Plantae, Angiosperms, Arabidopsis	Plant	Model organisms	Y. Zhang, UBC	MT328579
<i>Homo sapiens</i> (Human)	Animalia, Chordata, Mammals	Mammals	Model organisms	The skin tissue of my lip	MT328572
<i>Mus musculus</i> (Mouse)	Animalia, Chordata, Mammals	Mammals	Model organisms	J. Caleta, UBC	MT328573
<i>Bos taurus</i> (Cow)	Animalia, Chordata, Mammals	Mammals	Model organisms	Commercial market	MT328574
<i>Gallus gallus</i> (Chicken)	Animalia, Chordata, Aves	Aves	Model organisms	Commercial market	MT328575
<i>Salmo salar</i> (Salmon)	Animalia, Chordata, Actinopterygii	Fish	Model organisms	Commercial market	MT328576
<i>Solenocera crassicornis</i> (Shrimp)	Animalia, Arthropoda, Malacostraca	Malacostraca	Model organisms	Commercial market	MT328577
<i>Crassostrea gigas</i> (Oyster)	Animalia, Mollusca, Bivalvia	Bivalvia	Model organisms	K.M. Miller, Fisheries & Oceans, Canada	MT328571
<i>Drosophila melanogaster</i>	Animalia, Arthropoda, Insecta	Insect	Model organisms	Wild nature	MT328578

775

776

777

778

779

780

781

782

783 **Supplementary Table 2** List of the 20-nt sgRNA-target-site oligonucleotide sequences designed for cutting V4 region of 18S rRNA genes of ten model organisms using
784 CRISPR-Cas9.

Name of sgRNA	Name of the 20-nt gRNA-target-site oligo	The 20-nt gRNA-target-site oligo sequence (5'-3')	Target strand	Oligo position (5' - 3' of forward strand)	Number of protists & fungi of SILVA v119 SSU database that the oligo target	Percentage of protists & fungi of SILVA v119 SSU database that the oligo target (%)	Targeted model organisms
g83	g83	ACACTCAGCTAAGAGCATCG	Reverse	278 to 259	1	0.003%	<i>Homo sapiens</i> (Human)
g338	g338	ACGGGCGGTGACTCGCCTCG	Reverse	326 to 307	0	0.000%	<i>Mus musculus</i> (Mouse)
g88	g88	AGGGGGCGCCGAGAGGCAAG	Reverse	298 to 279	1	0.003%	<i>Bos taurus</i> (Cow)
g170	g170	GCTCTTAAGTGTGAGTGTCCCG	Forward	151 to 170	0	0.000%	<i>Gallus gallus domesticus</i> (Chicken)
g258	g258	AGTAAACGCTTCGGACCCCG	Reverse	257 to 238	0	0.000%	<i>Salmo salar</i> (Salmon)
g318	g318	GGAATGATGGAACAGGACCT	Forward	253 to 272	0	0.000%	<i>Solenocera crassicornis</i> (Shrimp)
g188	g188	AACCAATGAAAGCAGAACAG	Reverse	163 to 144	0	0.000%	<i>Drosophila melanogaster</i>
m258	m258	CAACTCAGTCAAGAGCACCA	Reverse	277 to 258	0	0.000%	<i>Crassostrea gigas</i> (Oyster)
g39	g39	AGGAGCGTATCGCCGACCGA	Reverse	291 to 272	0	0.000%	<i>Arabidopsis thaliana</i>
g244	g244	TGGTTAAGAGGGACAAACCG	Forward	281 to 300	0	0.000%	<i>Caenorhabditis elegans</i>

785

786

787

788

789

790

791

792

793

794

795 **Supplementary Table 3** List of primers used in this study.

Gene markers	Primer names	Primer (5'-3')	Target range	Reference
18S rRNA gene	TAReuk454FWD1	CCAGCASCYGCGGTAATTCC	Universal 18S	Ref. 26
	TAReukREV3	ACTTTCGTTCTTGATYRA		
	TAReuk454FWD1-Nxt	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCAGCASCYGCGGTAATTCC		This study; Ref. 41
	TAReukREV3-Nxt	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG ACTTTCGTTCTTGATYRA		
18S rRNA gene	18s-EUK581-F	GTGCCAGCAGCCGCG	Non-metazoan eukaryotes	Ref. 27
	18s-EUK1134-R	TTTAAGTTTCAGCCTTGCG		
	18s-EUK581-F-Nxt	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GTGCCAGCAGCCGCG		This study; Ref. 41
	18s-EUK1134-R-Nxt	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG TTTAAGTTTCAGCCTTGCG		

796

797