Revealing the Composition of the Eukaryotic Microbiome of Oyster Spat by CRISPR-Cas Selective Amplicon Sequencing (CCSAS)

Kevin Xu Zhong
University of British Columbia https://orcid.org/0000-0002-1386-8816
Anna Cho
UBC: The University of British Columbia
Christophe M. Deeg
UBC: The University of British Columbia
Amy M. Chan
UBC: The University of British Columbia
Curtis A. Suttle (suttle@science.ubc.ca)
UBC: The University of British Columbia https://orcid.org/0000-0002-0372-0033

Methodology

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

Posted Date: February 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-240715/v1

Version of Record: A version of this preprint was published at Microbiome on November 26th, 2021. See the published version at https://doi.org/10.1186/s40168-021-01180-0.

Revealing the composition of the eukaryotic microbiome of oyster spat by CRISPR-Cas Selective Amplicon Sequencing (CCSAS)

Kevin Xu Zhong[1]*, Anna Cho[2], Christophe M. Deeg[2], Amy M. Chan[1] and 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: Microbiome – Methodology

(https://microbiomejournal.biomedcentral.com/submission-guidelines/preparing-your-manuscript/methodology)

Abstract

Background: The microbiome affects the health of plants and animals, including humans, and has many biological, ecological and evolutionary consequences. Microbiome studies typically rely on sequencing ribosomal 16S RNA gene fragments, which serve as taxonomic markers for prokaryotic communities; however, for eukaryotic microbes this approach is compromised, because 18S rRNA gene sequences from microbial eukaryotes are swamped by contaminating host rRNA gene sequences.

Results: 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 rRNA gene sequences of the host, while leaving protistan and fungal sequences intact. We validated the specificity of the sgRNA on ten model organisms and an artificially constructed (mock) community of nine protistan and fungal pathogens. The results showed that >96.5% of host rRNA gene amplicons were cleaved, while 18S rRNA gene sequences from protists and fungi were unaffected. When used to assess the eukaryotic microbiome of oyster spat from a hatchery, CCSAS revealed a diverse community of eukaryotic microbes, typically with much less contamination from oyster 18S rRNA gene sequences than other methods using non-metazoan or blocking primers. However, each method revealed taxonomic groups that were not detected using the other methods, showing that a single approach is unlikely to uncover the entire eukaryotic microbiome in complex communities. To facilitate the application of CCSAS, 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.

Conclusion: CCSAS provides a high-through-put and cost-effective approach for resolving the eukaryotic microbiome of metazoa and plants with minimal contamination from host 18S rRNA gene sequences.

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

Background

There is a growing interest in understanding how the composition of the microbiome affects the health of plants [1-2] and animals [3-7], including humans [8-9]. For example, in humans the gut microbiome is associated with both positive and adverse health effects, and changes in the microbiome have been linked to a number of diseases [10-13], such as obesity [14-15], diabetes [16-17], inflammatory bowel disease [18-21], cancer [22-25], cadiovascular disease [26-27], and even mental illness [28-30]. As well, a wide span of biological, ecological and evolutionary questions have been addressed through microbiome studies [3, 6-9, 31-32]. Microbes have been shown to affect host metabolism [33], host immunity [34-35], and human development [8, 36] incuding the brain [37-38], and may even influence the evolution of animals and plants through microbe-host interactions [3, 7, 32, 39-45].

Microbiome studies have largely been facilitated through deep sequencing of ribosomal RNA gene fragments [46-49]; yet, our knowledge of the eukaryotic component of the microbiome, particularly protists, is relatively limited compared to that of prokaryotes [6, 49-53]. This is largely due to the challenge of profiling host-associated eukaryotic microbes, as the standard "universal" primers [54] used to amplify 18S rRNA gene sequences from eukaryotic microbes also amplify host 18S rRNA gene sequences, which will dominate the sequencing library [46, 52, 55].

A number of approaches have been use to minimize contamination by host 18S rRNA gene sequences. For example, primers can be designed that will not amplify host 18S rRNA sequences, but will amplify sequences from microeukaryotes (e.g. reference 56-58);

alternatively, other marker genes can be targeted such as the ITS region of fungi [59]. However, designing primers to amplify ribosomal RNA gene sequences from a broad range of microeukaryotes, but not the host, can be challenging.

Another approach is to use primers to block amplification of host 18S rRNA sequences to study the eukaryotic microbiome [60]. Such "blocking primers" typically use a short blocking-oligonucleotide with a modified 3' end that binds to the 18S rRNA gene of the host, and prevents "universal" 18S primers from amplifying host sequences [60]. Such an approach has been successfully applied to krill [60], fish [61-62], coral [63], primates [64], shrimp [65-66], flying squid [67], mosquitos [68-69] and Pacific oysters [57], although a large proportion of the sequences can still be host-derived (e.g. up to 92% in coral, 42% in krill, and 45% in fish) [57, 72, 63]. This approach also requires designing and optimizing the blocking-primers for each animal host, which remains a challenge [71-72].

Recently, a method involving the usage of non-metazoan (UNonMet) primers [58] was developed [71-72] and was shown to be effective in coral and humans [71, 73]. This "non-metazoan primers" method employs a nested-PCR approach which involves a two-step PCR procedure. The first-PCR step uses UNonMet primers [58] to generate ~600-bp fragments of 18S rRNA gene that are specific to microeukaryotes but not to metazoans; the products from the first PCR are reamplified using the "universal" 18S primers to produce a shorter 18S rRNA gene fragment [71]. This method has the advantage of not requiring host-specific primer design, but based on *in silico* analysis cannot be used for sponges and ctenophores [71].

Here we describe CRISPR-Cas Selective Amplicon Sequencing (CCSAS), an alternative approach to resolve the eukaryotic microbiome of metazoa and plants. Clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9) system provides bacteria and archaea adaptive immunity against viruses and plasmids by cleaving invading double-stranded (ds) DNA [74]. 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) crispr RNA (crRNA) guide sequence that is complementary to the DNA target site [75-78]. Cas9 can be programmed to target any DNA sequence by modifying the 20-nt guide sequence [78-79]. 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) [78], CRISPR-Cas9 has emerged as a powerful tool in a wide variety of applications [79-80]. CCSAS leverages this tool by using a custom sgRNA to direct Cas9 to specifically cut host 18S rRNA gene sequences in the region flanked 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 removed during the preparation of the sequencing library. This results in a library highly enriched in 18S amplicons from microeukaryotes, allowing for high-resolution surveys of the taxonomic composition of eukaryotic microbes associated with any eukaryotic host.

Results

Design of the taxon-specific sgRNA

The key to CCSAS is the 20-nt guide sequence of gRNA that directs Cas9 to selectively cut the 18S rRNA gene 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 rRNA gene region spanned by "universal" primers that can serve as target-sites for gRNA, and which are complementary to the sgRNA's guide sequence. The selected gRNA and sgRNA duplex thus dictates the specificity of the sgRNA-CRISPR-Cas complex, allowing a user to easily synthesize a taxon-specific sgRNA.

To validate the target-specificity of sgRNA, taxon-specific sgRNAs were designed and tested for 18S rRNA sequences from 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 an artificially constructed (mock) community composed of nine protists and fungi (Table S1 and S2). 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. S1). Thus, the sgRNAs effectively targeted host sequences, while leaving sequences from microeukaryotes intact.

Using CCSAS to reveal host-associated microeukaryotic populations

The next step was to evaluate the effectiveness of Cas9 when complexed with the host-specific sgRNA. After CRISPR-Cas9 treatment, about 0.6% to 3.5% of the remaining 18S amplicons were still host-derived, but in most cases still dominated the sequencing library (data not shown). Hence, to further reduce the host-derived 18S rRNA gene sequences, we introduced a two-step CRISPR-Cas9 procedure (Fig. 2). 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 rRNA gene, 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 clean-up step using SPRI magnetic beads. Second, following the first size selection, another Cas9 cut, PCR amplification and size selection is conducted (Fig. 2), resulting in almost the complete removal of host 18S amplicons, while leaving the protistan and fungal

amplicons intact. This allows for high-resolution characterization of the composition of the microeukaryotic community with a fraction of the sequencing effort typically used.

We applied two-step CCSAS 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 that using CCSAS in conjunction with "universal" 18S primers resulted in almost the complete removal of oyster 18S amplicons, while leaving the protistan and fungal amplicons intact and highly enriched for sequencing (Fig. S2; Fig. 3). With CCSAS, the percentage of sequences from metazoa (mostly assigned to oysters, although some were from nematodes in the order Monhysterida; Fig. S2) was at most 7.4%, while in three out of eight samples, sequences from metazoa were undetectable (Fig. 3). In contrast, with non-metazoan and blocking primers, most sequences were still from metazoa (Fig. 3), primarily oysters (Fig. S2). When compared to non-metazoan and blocking primers, CCSAS revealed all the major eukaryotic microbial groups including members of the Ochrophyta, Labyrinthulomycetes and Ciliophora (Fig. 3; Fig. 4). Nevertheless, given the three methods employ different “universal” 18S primers, there were differences among the taxa detected (Fig. 4). For example, CCSAS detected the genus *Telonema*, peronosporomycetes in the Stramenopiles and Picomonadida in the Picozoa; while non-metazoan and blocking primers did not. Yet, CCSAS did not detect the MAST4-group of stramenophiles or hyphochytriomyctes, while, non-metazoan primers did; and prymnesiophytes, cryptophytes and fungi in the phylum Cryptomycota and the division Chytridiomycota were only revealed by the blocking primers. As well, members of the genus *Mantamonas* and the family Acanthocystidae were detected by the non-metazoan and blocking primers, but not by CCSAS (Fig. 4). Additionally, using CCSAS the relative abundances of cercozoans and dinoflagellates were less than with the other methods. Thus, amplification with "universal" 18S primers combined with CCSAS had less contamination by host sequences and revealed some additional taxa compared to non-metazoan

and blocking primers; however, there were also some taxa that were absent using CCSAS. Nonetheless, the composition of the eukaryotic microbiome detected by the three methods was quite similar, and taxa that were not detected by one or more methods were always a minor component of the overall community.

Database of gRNA-target-sites for metazoa and plants

To enable CCSAS to be easily applied for characterizing eukaryotic microbiomes in a wide range of metazoa and plants, we used CasOligo to identify gRNA-target-sites for 99.6% of the 15907 metazoa and plant taxa (metaphyta of Embryophyta group) in the SILVA SSU database [81] (*version 119, released on 24 July 2014*) (Fig. 5). For each taxon we identified between 3 and 217 (average 33) gRNA-target-sites that are compatible with the CRISPR-Cas9 system (Fig. S3); of these, between 1 and 214 targeted the putative host 18S sequence, but not protistan or fungal sequences. Thus, the database provides a wide selection of gRNA-target-site sequences from which to design taxon-specific sgRNAs.

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 (Fig. S4). 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 (Fig. S4). CasOligo, can also be used to retrieve gRNA-target-site sequences for specific taxa by entering the species name of the host using the function, search.db.byname(). 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 sequence-specific and the gRNA-target-site database does not cover all sequence variants for a specific taxon.

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, CCSAS requires much less sequencing than other methods 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. We identified taxon-specific gRNA-target-sites for 99.6% of the taxa in the SILVA database, with an average of 33 taxon-specific gRNA-target-sites per taxon, showing that CCSAS can be applied to nearly all metazoa and plants. Additionally, the CasOligo package provides an oligonucleotide design function, Cas9.oligo.search2(), that can be used to design custom sgRNAs 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, gRNA can recognize the wrong target [82-84], which might lead Cas9 to cut some protistan and fungal sequences, or incompletely cleave host sequences. This problem can be minimized by careful design of the gRNA, and in silico analysis against the most comprehensive databases of 18S gene sequences. Second, efficient sequencing requires effective removal of the cut host amplicons. This can be accomplished by optimizing the size selection of SPRI magnetic beads or may reduce sequencing efficiency, or adapting other methods for size selecting DNA

fragments. Third, there are inherent amplification biases associated with PCR [85]; thus the accuracy of differences in the relative abundances of specific sequences using CCSAS, or any other PCR-based approach is unknown. Fourth, the design of host-specific gRNA target sites is only as good as the available 18S rRNA gene references for microeukaryotes. However, the design of gRNA target sites will continue to improve as SSU sequence databases continue to expand. Despite these caveats, CCSAS can be used to obtain high-resolution data on the composition of eukaryotic microbiomes with relatively low sequencing effort. Moreover, it has broad application because gRNA-target-sites can be identified for thousands of host species.

Conclusions

CCSAS is a powerful tool with which to investigate the composition of the eukaryotic microbiome for a vast array of host organisms. Relative to approaches using non-metazoan or blocking primers, CCSAS provides similar resolution of the eukaryotic microbial community, but with much less contamination by sequences from host 18S rRNA genes. Moreover, the ease with which specific sgRNA can be designed allows CCSAS to be used to explore the eukaryotic microbiome of almost any host organism. Thus, CCSAS can facilitate significant advances for investigations of the eukaryotic microbiome across a wide diversity of hosts.

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 nematode (*Caenorhabditis elegans*), as well as nine species of protists and fungi were obtained from either commercial markets or laboratories at The University of British Columbia (Table S1). As well, eight samples of seven- to 28-day old oyster spat, with sizes ranging

between 0.4 and 1.0 mm, were obtained from a hatchery that was experiencing mortality events. The oyster spat were immediately frozen using liquid nitrogen following collection, and stored at -80ºC 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).

An artificial community of microeukaryotes was made by pooling equal amounts (~50 ng) of genomic DNA from each protist and fungus (Table S1).

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 rRNA gene 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 using a EnGen™ sgRNA Synthesis Kit from New England Biolabs (NEB).

Obtaining the host 18S rRNA gene sequences. Prior to the design of the sgRNA, we obtained the 18S rRNA gene sequences of the host organisms for identifying gRNA-target-sites, and employed the following cloning and sequencing approaches:

For each host, 18S rRNA gene fragments were PCR amplified using the "universal" primers TAReuk454FWD1 and TAReukREV3 [54] to produce 380-450 bp amplicons that were 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 (NEB), 4 mM $MgCl_2$, 20 µg of Bovine Serum Albumin (NEB), 200 nM 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 described [54], 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 (Invitrogen) 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. These DNA sequences were then used to design the taxon-specific 20-nt gRNA-target-site sequences, which were used to synthesize the taxon-specific sgRNAs that guide 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) to design taxon-specific 20-nt gRNA-target-site oligonucleotide sequences, which allows 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 TAReukREV3 [54]. 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 rRNA gene. The gRNA-target-site oligonucleotide sequences designed in this study are shown in Table S2.

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* (NEB) 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 a 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.

Preparation of the host 18S amplicons. For each of the ten host organisms and the mock community of protists and fungi, 18S rRNA gene fragments were obtained using PCR with the "universal" primers TAReuk454FWD1 and TAReukREV3 [54] following the conditions detailed above. The 18S amplicons were purified using Agencourt SPRI magnetic beads (Beckman Coulter) at a 1:1 (vol:vol) ratio of beads:DNA.

DNA cleavage using CRISPR-Cas9. For each of the ten host organisms and the mock community of protists and fungi, the purified 18S amplicons were cut using Cas9 Nuclease, *S. pyogenes* (NEB) in the presence of a sgRNA, following the manufacturer's directions. Briefly, the 10 µL reaction contained approximately 0.1 pmol of dsDNA, 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. The reaction was incubated at 37°C for 4 h in a thermocycler, followed

342 by 70°C for 10 min to deactivate the CRISPR-Cas9. For each sample, in parallel with the CRISPR-Cas9
343 treatment we also prepared the reaction without CRISPR-Cas9 treatment, in which Cas9 nuclease and sgRNA
344 were replaced with molecular grade ultrapure water (Invitrogen). Thus, each reaction of both treatments
345 contained the same amount of template dsDNA (18S amplicons at 0.1 pmol) and was subjected to the same
346 incubation conditions.
347
348 *Visualization using gel electrophoresis.* The size of the 18S rRNA gene fragments with and without CRISPR-
349 Cas9 treatment was visualized by gel electrophoresis using a Bioanalyser (Agilent). Prior to loading into the gel,
350 the Cas9-cut products (5 µL out of 10 µL) were treated with 1 mg/mL (final) Proteinase K (Invitrogen) at room
351 temperature for 15 min to digest the Cas9 nuclease. Then, 1 to 2 µL of this proteinase-K-treated product was
352 added into a well of an Agilent High Sensitive DNA Chip in a Bioanalyzer (Agilent) to visualize and verify
353 cutting by CRISPR-Cas9.
354
355 *Quantitative PCR.* To determine the efficiency of CRISPR-Cas9 for eliminating host-derived 18S sequences, we
356 used quantitative PCR (qPCR) and the primers TAReuk454FWD1 and TAReukREV3 (Table S3) that targets a
357 380-450 bp fragment of the V4 region of the 18S rRNA gene, to assess the proportion of 18S amplicons cut by
358 Cas9. The 10 µL qPCR reactions contained 1 X SsoFast™ EvaGreen® Supermix (Bio-Rad), 0.5 µM of each
359 primer, and a 1 µL 1/10000 dilution of DNA template consisting of amplified products, either with or without
360 the addition of Cas9. Thermal cycling was done in a CFX96 real-time PCR detection system (Bio-Rad) with the
361 following program: 3 min denaturation at 95°C, followed by 40 cycles of denaturation at 95°C for 30 s, and
362 annealing and extension at 49°C for 30 s. Nine, 10-fold serially diluted standards (ranging from 5 x 10^0 to 5 x
363 10^9 molecules per mL) were run in duplicate along with two no-template control reactions containing 1 µL of
364 nuclease-free water. The amplicon standards were made from a cloned 18S rRNA gene fragment amplified from
365 a culture of the prasinophyte microalga, *Micromonas pusilla*, using the primer set
366 TAReuk454FWD1/TAReukREV3 [54]. The amplicons were purified using a MiniElute® PCR Purification Kit
367 (Qiagen), and quantified using a Qubit® dsDNA High Sensitivity Assay Kit (Invitrogen). The size of the
368 amplicon was checked using gel-electrophoresis, and the qPCR melting curves were used to confirm that the
369 fluorescence signal corresponded to a single-sized DNA fragment. The qPCR amplification efficiency was
370 between 0.95 and 1.05 for the cloned amplicons (with $r>0.98$, $n=9$).
371

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 rRNA gene, followed by a second cut of any host-derived 18S amplicons. Details of the method are provided below.

Cas9 cutting of host genomic DNA. Genomic DNA of the host was 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 was incubated at 37°C for 4 h in a thermocycler.

The first PCR and size selection. The Cas9-cleaved genomic DNA was used as a template in the first PCR to generate 380-450 bp amplicons from the V4 region of the 18S rRNA gene that are depleted in host sequences. To ensure representative amplification of 18S sequences from microeukaryotes, four parallel PCR reactions were run at different annealing temperatures (45, 47, 48 or 49°C), using the "universal" 18S primers TAReuk454FWD1-Nxt and TAReukREV3-Nxt (Table S3). Compared to TAReuk454FWD1 and TAReukREV3 [54], this modified primer set contained overhang adapter sequences (Table S3), 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 Illumina [86].

Details on the first PCR reactions are as follows. Briefly, each 25 µL reaction mix contained 1X PCR buffer (NEB), 4 mM $MgCl_2$, 20 µg of Bovine Serum Albumin (NEB), 200 nM 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 2 bp shorter than the forward primer and has a lower annealing temperature, we used the two-step PCR approach of Stoeck et al. [54], 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. At the end, the PCR product of the four reactions were pooled together. Then amplicons were 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 cutting of the 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 (NEB), 4 mM $MgCl_2$, 200 nM of each dNTP (Invitrogen), 5 µL of each index primer (N7XX and S5XX of Nextera® XT Index Kit, Illumina), 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 a 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 universal 18S primers

Sequencing libraries for 18S amplicons generated using the "universal" 18S primers, and not cut using CRISPR-Cas9, were prepared using protocols adapted from Illumina [86]. 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 (Table S3) 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 index PCR were the same as above the second CCSAS PCR, except here the PCR amplification cycle was reduced to be 16 cycles. 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 blocking primers

Preparation of the sequencing library for the 18S amplicons obtained using blocking primers was similar to that described above, except that the first PCR used the primer set 18SV4-F-Nxt / 18SV4-R-Nxt and the oyster-blocking primer 18SV4-Block-oyster (Table S3), which was adapted from Clerissi et al. [57] , to amplify a ~377 bp fragment of 18S rRNA gene that is specific to microeukaryotes but not Pacific oysters. This 30-nt oyster-blocking primer was modified at the 3' end with theSpacer C3 CPG (3 hydrocarbons) and contained a 10-bp overlap with the reverse primer 18SV4-R-Nxt, which prevents the amplification of the 18S rRNA gene from Pacific oysters, and thus enriches the proportion of amplicons from microeukaryotes [57] . In the first PCR, the 25-µL reaction mix comprised 1X PCR buffer (NEB), 4 mM $MgCl_2$, 20 µg of Bovine Serum Albumin (NEB), 200 nM of each dNTP (Invitrogen), 0.4 µM of primer 18SV4-F-Nxt, 0.4 µM of primer 18SV4-R-Nxt, 1.2 µM of the oyster-blocking primer 18SV4-Block-oyster, 0.5 U of Q5® high-fidelity polymerase (NEB) and approximately 5 ng of genomic DNA. The PCR cycling was as per Clerissi et al. [57], which included an initial incubation of 15 min at 96°C followed by 35 cycles of denaturation at 96°C for 30 s, annealing at 52°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 16-cycles index PCR to add a Nextera® XT index (Illumina) to each 3' and 5' end of the amplicons.

Sequencing library preparation for amplicons generated using non-metazoan primers

The non-metazoan primers, UNonMet primers [58], and a two-step nested-PCR were used to generate 18S amplicons from non-metazoan eukaryotes following del Campo et al. [71]. The first step of the nested-PCR uses the primers 18s-EUK581-F and 18s-EUK1134-R [58] (Table S3) to generate ~600-bp 18S rRNA gene fragments from microeukaryotes. Then, these fragments are used in a second PCR with the universal V4 primer set E572F-Nxt / E1009R-Nxt [87] (Table S3) to amplify a ~440-bp 18S rRNA gene fragment to which overhanging adapter sequences (Table S3) are added that are compatible with the Illumina indexes and sequencing adapters. Finally, a third PCR is used to add a Nextera® XT index (Illumina) to each 3' and 5' end of the amplicons.

In the first PCR, the 25 µL reaction mix comprised 1X PCR buffer (NEB), 4 mM $MgCl_2$, 20 µg of Bovine Serum Albumin (NEB), 200 nM 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. The initial denaturation of 2 min at 98°C was followed by 25 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).

In the second PCR, adapted from Comeau et al. [87], the 25 µL reaction mix comprised 1X PCR buffer (NEB), 4 mM $MgCl_2$, 20 µg of Bovine Serum Albumin (NEB), 200 nM of each dNTP (Invitrogen), 0.4 µM of each primer, 0.5 U of Q5® high-fidelity polymerase (NEB) and approximately 5 ng of the purified 1st PCR amplicons. There initial denaturation of 2 min at 98°C was followed by 20 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s and elongation at 72°C for 30 s, and a final elongation for 10 min at 72°C. The 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).

Last, the amplicon libraries were completed as described above, with a 16-cycles of PCR to add the Nextera® XT index (Illumina) to the 3' and 5' ends 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, blocking 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 sequenced 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 1.9 [88]. Briefly, sequences were de-multiplexed by their forward and reverse indexes, and the paired-end reads merged using PEAR version 1.10.4 [89]. Then, sequences from different samples were pooled, and Uclust [90] 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 database [81] (*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 within a sample. The downstream analysis was conducted in R v3.5.3 [91] using packages such as Phyloseq version 1.26.1 [92] and the figures were generated using ggplot2 version 3.3.0 [93] and metacoder [94].

Abbreviations

CRISPR: Clustered regularly interspaced short palindromic repeats; Cas: CRISPR-associated protein; CCSAS: CRISPR-Cas Selective Amplicon Sequencing; gRNA: guide RNA; sgRNA: single-guide RNA; dsDNA: double-stranded DNA; ssDNA: single-stranded DNA; nt: nucleotide; crRNA: crispr RNA; tracrRNA: trans-activating RNA; V4 region of the 18S rRNA gene: Variable region 4 of the 18S rRNA gene; PCR: Polymerase chain reaction; OTU: Operational taxonomic unit; SSU: Small subunit; dNTP: deoxyribonucleotide triphosphate; bp: base pairs.

Additional files

Additional file 1: **Table S1.** List of organisms used in this study.

Additional file 2: **Table S2.** List of the 20-nt sgRNA-target-site oligonucleotide sequences designed for cutting V4 region of 18S rRNA genes of ten model organisms using CRISPR-Cas9.

Additional file 3: **Table S3.** List of primers used in this study.

Additional file 4: **Figure S1.** Percentage of intact 18S amplicons remaining from the model organisms after cutting with one-step CRISPR-Cas9. The concentration of intact 18S amplicons was measured using Quantitative PCR for samples both with and without CRISPR-Cas9 treatment. The portion of remaining intact 18S amplicons was determined by dividing the concentration 18S amplicons in sample with CRISPR-Cas9 treatment by that of without CRISPR-Cas9 treatment. The labels on the X-axes indicate the ID of the taxon-specific sgRNA and its corresponding host.

Additional file 5: **Figure S2.** Eukaryotic taxa representing >1% of the sequences revealed by deep-sequencing of the 18S rRNA amplicons for oyster spat samples, using "universal" 18S primers (Table S3), non-metazoan primers (Table S3), blocking primers (Table S3), or CRISPR-Cas Selective Amplicon Sequencing (**CCSAS**) combining "universal" 18S primers and CRISPR-Cas9 with pacific-oyster-specific sgRNA m258 (Table S2).

Additional file 6: **Figure S3.** Distribution of the number of gRNA-target-sites of each metazoan and plant species from the SILVA SSU database v119 [81]. These gRNA-target-site 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 set TAReuk454FWD1 / TAReukREV3 [54], and are used for designing and synthesizing the CRISPR-Cas9-compatible sgRNA. The taxon-specific gRNA-target-sites allows the 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.

Additional file 7: **Figure S4.** 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 / TAReukREV3 [54]. These nine sgRNAs, which are among 205242 unique taxon-specific sgRNA designed from the SILVA SSU database (*version 119*) [81] using CasOligo, are selected to show that some

549 sgRNAs can target more than 1000 species and broad taxonomic groups based on an *in-silico* analysis (*i.e.* 100%
550 match to the 18S rRNA gene sequences of the metazoan host at the gRNA-target-site, but no match for protists
551 and fungi). Taxon names on the left side of the panel are shown as SILVA taxonomic hierarchy with levels ranging
552 from D0 (kingdom) to D7. The D7 taxonomic level comprises eukaryotic classes and families.
553
554

555 Ethics approval and consent to participate

556 Not applicable.
557

558 Consent for publication

559 Not applicable.
560

561 Availability of data and materials

562 All next-generation sequencing data generated in this study have been deposited in the NCBI Sequence Read
563 Archive (SRA) under the accession numbers SRR13658714 to SRR13658745. The Sanger cloning-sequencing
564 data were deposited in GenBank under the accession numbers MT328571 to MT328580 for 18S rRNA gene
565 sequences of ten model organisms. All related scripts, functions and algorithms for designing gRNA-target-site
566 oligonucleotide sequences are included in the custom R package: CasOligo
567 (https://github.com/kevinzhongxu/CasOligo). The gRNA-target-sites database was included in the CasOligo
568 package, as well. The authors declare that all other data supporting the findings of this study are available within
569 the paper and/or the associated supplementary files.
570

571 Competing Interests

572 The authors declare no competing interests.
573

574 Funding

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).

Authors' 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.

Acknowledgments

We gratefully acknowledge 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, for providing biological materials for testing (Table S1).

Author details

[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.

Authors' information

Current address of A.C. is Department of Botany, University of British Columbia, Vancouver, BC, Canada. C.M.D. is currently at Pacific Biological Station, Nanaimo, BC, Canada.

References

1. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. Going back to the roots: the microbial ecology of the rhizosphere. Nature Reviews Microbiology. Nature. 2013;11:789–99.
2. Haney CH, Ausubel FM. Plant microbiome blueprints. Science. 2015;349:788–9.
3. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. PNAS. 2013;110:3229–36.
4. Bourne DG, Morrow KM, Webster NS. Insights into the Coral Microbiome: Underpinning the Health and Resilience of Reef Ecosystems. Annu Rev Microbiol. 2016;70:317–40.
5. Apprill A. Marine Animal Microbiomes: Toward Understanding Host–Microbiome Interactions in a Changing Ocean. Front Mar Sci. 2017;4 10.3389/fmars.2017.00222
6. Parfrey LW, Moreau CS, Russell JA. Introduction: The host-associated microbiome: Pattern, process and function. Mol Ecol. 2018;27:1749–65.
7. Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I. The role of microorganisms in coral health, disease and evolution. Nat Rev Microbiol. 2007;5:355–62.
8. Knight R, Callewaert C, Marotz C, Hyde ER, Debelius JW, McDonald D, et al. The Microbiome and Human Biology. Annu Rev Genomics Hum Genet. 2017;18:65–86.
9. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med. 2018;24:392–400.
10. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13:260–70.
11. Pflughoeft KJ, Versalovic J. Human Microbiome in Health and Disease. Annu Rev Pathol. 2012;7:99–122.
12. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Curr Opin Gastroenterol. 2015;31:69–75.
13. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. Nat Rev Microbiol. 2021;19:55–71.
14. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Human gut microbes associated with obesity. Nature. 2006;444:1022–3.

15. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science. 2013;341:1241214.
16. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490:55–60.
17. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen A-M, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe. 2015;17:260–73.
18. Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U.S.A. 2007;104:13780–5.
19. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The Treatment-Naive Microbiome in New-Onset Crohn's Disease. Cell Host Microbe. 2014;15:382–92.
20. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019;4:293–305.
21. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569:655–62.
22. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012;338:120–3.
23. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. Cell Host Microbe. 2013;14:207–15.
24. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013;13:800–12.
25. Xavier JB, Young VB, Skufca J, Ginty F, Testerman T, Pearson AT, et al. The Cancer Microbiome: Distinguishing Direct and Indirect Effects Requires a Systemic View. Trends Cancer. Elsevier; 2020;6:192–204.
26. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature. 2011;472:57–63.
27. Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun. 2017;8:845.

28. Luna RA, Foster JA. Gut brain axis: diet microbiota interactions and implications for modulation of anxiety and depression. Curr Opin Biotechnol. 2015;32:35–41.
29. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol. 2019;4:623–32.
30. Capuco A, Urits I, Hasoon J, Chun R, Gerald B, Wang JK, et al. Current Perspectives on Gut Microbiome Dysbiosis and Depression. Adv Ther. 2020;37:1328–46.
31. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats Across Space and Time. Science. 2009;326:1694–7.
32. Cordovez V, Dini-Andreote F, Carrión VJ, Raaijmakers JM. Ecology and Evolution of Plant Microbiomes. Annu Rev Microbioly. 2019;73:69–88.
33. Sonnenburg JL, Bäckhed F. Diet–microbiota interactions as moderators of human metabolism. Nature. 2016;535:56–64.
34. Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell. 2016;165:842–53.
35. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science. 2015;350:1084–9.
36. Sommer F, Bäckhed F. The gut microbiota — masters of host development and physiology. Nat Rev Microbiol. 2013;11:227–38.
37. Mayer EA, Knight R, Mazmanian SK, Cryan JF, Tillisch K. Gut microbes and the brain: paradigm shift in neuroscience. J Neurosci. 2014;34:15490–6.
38. Sharon G, Sampson TR, Geschwind DH, Mazmanian SK. The Central Nervous System and the Gut Microbiome. Cell. 2016;167:915–32.
39. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol. 2008;6:776–88.
40. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of Mammals and Their Gut Microbes. Science. 2008;320:1647–51.
41. Brucker RM, Bordenstein SR. The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus Nasonia. Science. 2013;341:667–9.

690 42. Rosenberg E, Zilber-Rosenberg I. Microbes Drive Evolution of Animals and Plants: the Hologenome
691 Concept. MBio. 2016;7:e01395.
692 43. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, et al. Cospeciation
693 of gut microbiota with hominids. Science. 2016;353:380–2.
694 44. Davenport ER, Sanders JG, Song SJ, Amato KR, Clark AG, Knight R. The human microbiome in
695 evolution. BMC Biol. 2017;15:127.
696 45. Sharpton TJ. Role of the Gut Microbiome in Vertebrate Evolution. mSystems 2018;3
697 10.1128/mSystems.00174-17
698 46. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, et al. Experimental and
699 analytical tools for studying the human microbiome. Nat Rev Genet. 2011;13:47–58.
700 47. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput
701 microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–4.
702 48. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for
703 analysing microbiomes. Nat Rev Microbiol. 2018;16:410–22.
704 49. Popovic A, Parkinson J. Characterization of eukaryotic microbiome using 18S amplicon sequencing. in
705 Microbiome Analysis: Methods and Protocols (eds. Beiko, R. G., Hsiao, W. & Parkinson, J.) 29–48
706 (Springer New York, 2018).
707 50. Parfrey LW, Walters WA, Knight R. Microbial eukaryotes in the human microbiome: ecology,
708 evolution, and future directions. Front Microbiol. 2011;2:153.
709 51. Andersen LO, Vedel Nielsen H, Stensvold CR. Waiting for the human intestinal Eukaryotome. ISME J.
710 2013;7:1253–5.
711 52. Laforest-Lapointe I, Arrieta M-C. Microbial Eukaryotes: a Missing Link in Gut Microbiome Studies.
712 mSystems. 2018;3.
713 53. Hooks KB, O'Malley MA. Contrasting Strategies: Human Eukaryotic Versus Bacterial Microbiome
714 Research. J Eukaryot Microbiol. 2020;67:279–95.
715 54. Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, et al. Multiple marker parallel tag
716 environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic
717 water. Mol Ecol. 2010;19:21–31.
718 55. Campo J del, Bass D, Keeling PJ. The eukaryome: Diversity and role of microeukaryotic organisms
719 associated with animal hosts. Funct Ecol. 2020;34:2045–54.

720 56. Waidele L, Korb J, Voolstra CR, Künzel S, Dedeine F, Staubach F. Differential Ecological Specificity
721 of Protist and Bacterial Microbiomes across a Set of Termite Species. Front Microbiol. 2017;8
722 10.3389/fmicb.2017.02518
723 57. Clerissi C, Guillou L, Escoubas J-M, Toulza E. Unveiling protist diversity associated with the Pacific
724 oyster Crassostrea gigas using blocking and excluding primers. BMC Microbiol. 2020;20:193.
725 58. Bower SM, Carnegie RB, Goh B, Jones SR, Lowe GJ, Mak MW. Preferential PCR amplification of
726 parasitic protistan small subunit rDNA from metazoan tissues. J Eukaryot Microbiol. 2004;51:325–32.
727 59. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal
728 internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. PNAS. 2012;
729 16:6241-6.
730 60. Vestheim H, Jarman SN. Blocking primers to enhance PCR amplification of rare sequences in mixed
731 samples - a case study on prey DNA in Antarctic krill stomachs. Front Zool. 2008;5:12.
732 61. Leray M, Agudelo N, Mills SC, Meyer CP. Effectiveness of annealing blocking primers versus
733 restriction enzymes for characterization of generalist diets: unexpected prey revealed in the gut contents
734 of two coral reef fish species. PLoS ONE. 2013;8:e58076.
735 62. Takahashi M, DiBattista JD, Jarman S, Newman SJ, Wakefield CB, Harvey ES, et al. Partitioning of
736 diet between species and life history stages of sympatric and cryptic snappers (Lutjanidae) based on
737 DNA metabarcoding. Sci Rep. 2020;10:4319.
738 63. Clerissi C, Brunet S, Vidal-Dupiol J, Adjeroud M, Lepage P, Guillou L, et al. Protists Within Corals:
739 The Hidden Diversity. Front Microbiol. 2018;9 10.3389/fmicb.2018.02043
740 64. Mann AE, Mazel F, Lemay MA, Morien E, Billy V, Kowalewski M, et al. Biodiversity of protists and
741 nematodes in the wild nonhuman primate gut. ISME J. 2020;14:609–22.
742 65. van der Reis AL, Laroche O, Jeffs AG, Lavery SD. Preliminary analysis of New Zealand scampi
743 (Metanephrops challengeri) diet using metabarcoding. PeerJ. 2018;6:e5641.
744 66. Liu C, Qi R-J, Jiang J-Z, Zhang M-Q, Wang J-Y. Development of a Blocking Primer to Inhibit the PCR
745 Amplification of the 18S rDNA Sequences of Litopenaeus vannamei and Its Efficacy in Crassostrea
746 hongkongensis. Front Microbiol. 2019;10:830.
747 67. Fernández-Álvarez FÁ, Machordom A, García-Jiménez R, Salinas-Zavala CA, Villanueva R. Predatory
748 flying squids are detritivores during their early planktonic life. Sci Rep. 2018;8:3440.

68. Hino A, Maruyama H, Kikuchi T. A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing; parasitome analysis method. Parasitol Int. 2016;65:572–5.
69. Belda E, Coulibaly B, Fofana A, Beavogui AH, Traore SF, Gohl DM, et al. Preferential suppression of Anopheles gambiae host sequences allows detection of the mosquito eukaryotic microbiome. Sci Rep. 2017;7:3241
71. Campo J del, Pons MJ, Herranz M, Wakeman KC, Valle J del, Vermeij MJA, et al. Validation of a universal set of primers to study animal-associated microeukaryotic communities. Environ Microbiol. 2019;21:3855–61.
72. Bass D, del Campo J. Microeukaryotes in animal and plant microbiomes: Ecologies of disease? Eur J Protistol. 2020;76:125719.
73. Kwong WK, del Campo J, Mathur V, Vermeij MJA, Keeling PJ. A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. Nature. 2019;568:103–7.
74. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315:1709–12.
75. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science. 2008;321:960–4.
76. Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010;468:67–71.
77. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471:602–7.
78. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21.
79. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346:1258096.
80. Barrangou R, Doudna JA. Applications of CRISPR technologies in research and beyond. Nat Biotechnol. 2016;34:933–41.
81. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl Acids Res. 2013;41:D590–6.
82. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotech. 2013;31:839–43.

83. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotech. 2013;31:822–6.

84. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotech. 2013;31:827–32.

85. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. Appl Environ Microbiol. 2005;71:8966–9.

86. Anonymous. 16S Metagenomic Sequencing Library Preparation. Illumina Inc. https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.

87. Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLoS ONE. 2011;6:e27492.

88. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Meth. 2010;7:335–6.

89. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614–20.

90. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.

91. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

92. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE. 2013;8:e61217.

93. Wickham H. ggplot2 [Internet]. New York, NY: Springer New York; 2009 [cited 2019 Aug 20]. Available from: http://link.springer.com/10.1007/978-0-387-98141-3

94. Foster ZSL, Sharpton TJ, Grünwald NJ. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. PLoS Comput Biol. 2017;13:e1005404.

FIGURE AND TABLE LEGEND

Figure 1 Agilent Bioanalyzer gel images of 18S amplicons from ten model organisms (**a** and **b**) and an artificial community of protists and fungi (**c**) to which Cas9 with the taxon-specific sgRNA (as shown in Table S2) was either added (+) or not (-). Gel bands show the amplicon length in base pairs (bp) relative to a DNA ladder. The labels on the X-axes of panel **c** indicate the ID of the taxon-specific sgRNAs and its corresponding host to target.

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

Figure 3 Relative abundances of eukaryotic groups in eight oyster samples revealed using deep-sequencing of 18S amplicons of rRNA genes using non-metazoan primers (**NM**), blocking primers (**BP**) and CCSAS using 18S "universal" primers and cleavage using CRISPR-Cas9 (**CCSAS**).

Figure 4 Comparison of the number of 18S rRNA gene OTUs summed for the eight oyster samples for each microeukaryotic group. The methods are based on using non-metazoan primers (**NM**), blocking primers (**BP**) and CCSAS combining 18S "universal" primers and cleavage using CRISPR-Cas9 (**CCSAS**).

Figure 5 Illustration showing the number and taxonomic distribution of gRNA-target-sites for metazoa and plants that are available for 18S rRNA sequences in the SILVA SSU database v119 [81] These gRNA-target-site oligonucleotide sequences are used for designing and synthesizing the taxon-specific and the CRISPR-Cas9-compatible sgRNAs that are used to guide CRISPR-Cas9 to cut the 18S rRNA genes of metazoan or plant hosts, but not those of microeukaryotes (protists and fungi). The node size indicates the number of species at each corresponding taxonomic level, while the size of the edge presents the number of gRNA-target-sites. Nodes and edges with the highest values are purple, while the smallest ones are grey. Only taxa with more than 50 gRNA-target-sites per taxon are shown.

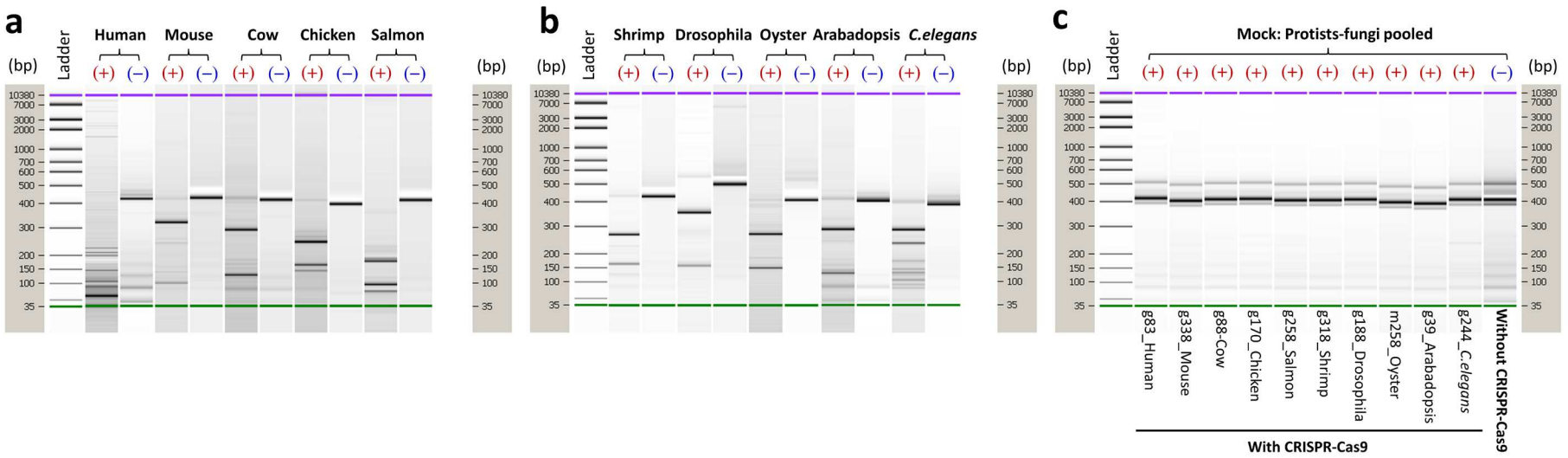


Fig. 1 Agilent Bioanalyzer gel images of 18S amplicons from ten model organisms (**a** and **b**) and an artificial community of protists and fungi (**c**) to which Cas9 with the taxon-specific sgRNA (as shown in Table S2) was either added (+) or not (-). Gel bands show the amplicon length in base pairs (bp) relative to a DNA ladder. The labels on the X-axes of panel **c** indicate the ID of the taxon-specific sgRNAs and its corresponding host to target.

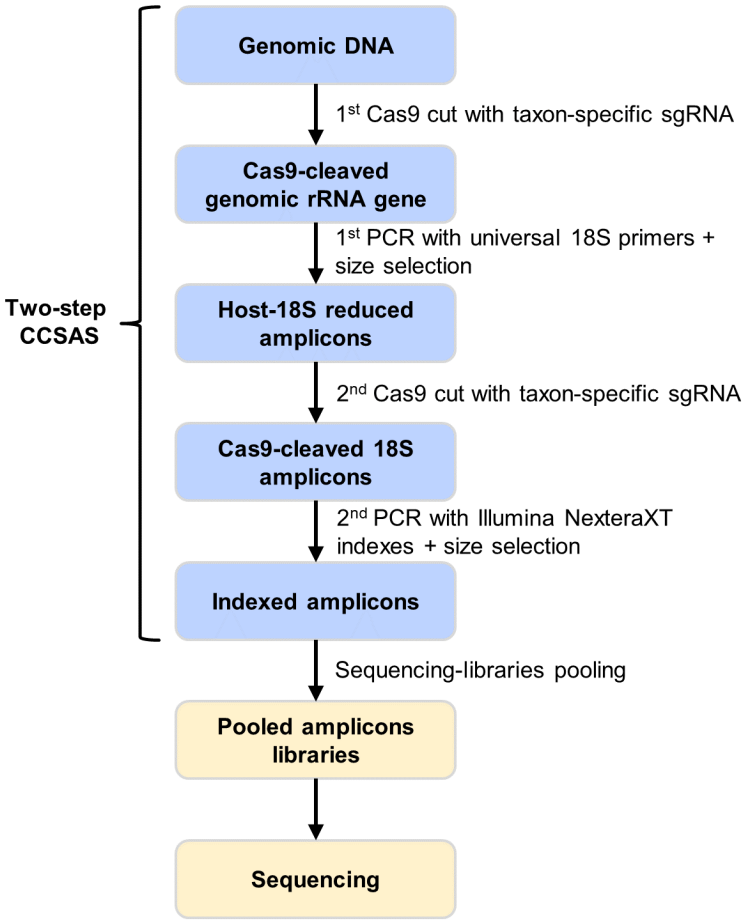


Fig. 2 Workflow for two-step CRISPR-Cas Selective Amplicon Sequencing (CCSAS) to study the composition of the host-associated eukaryotic microbiome.

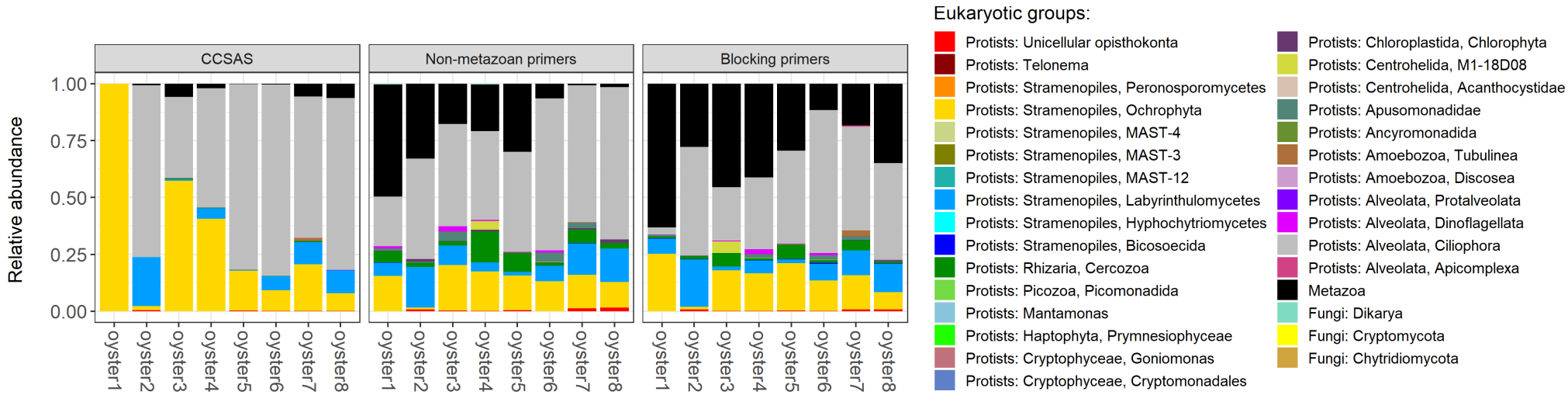
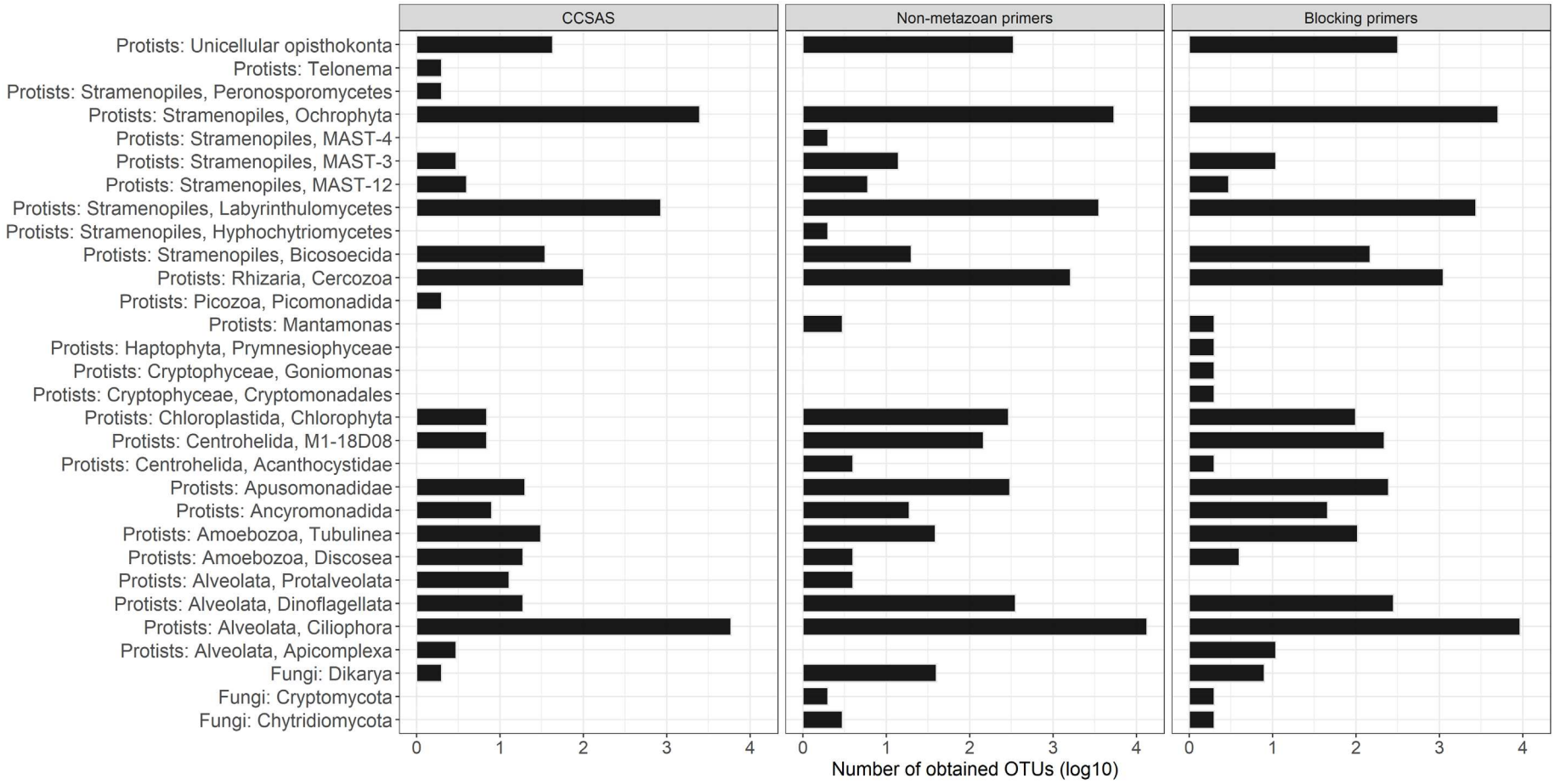


Fig. 3 Relative abundances of eukaryotic groups in eight oyster samples revealed using deep-sequencing of 18S amplicons of rRNA genes using non-metazoan primers (**NM**), blocking primers (**BP**) and CCSAS using 18S "universal" primers and cleavage using CRISPR-Cas9 (**CCSAS**).

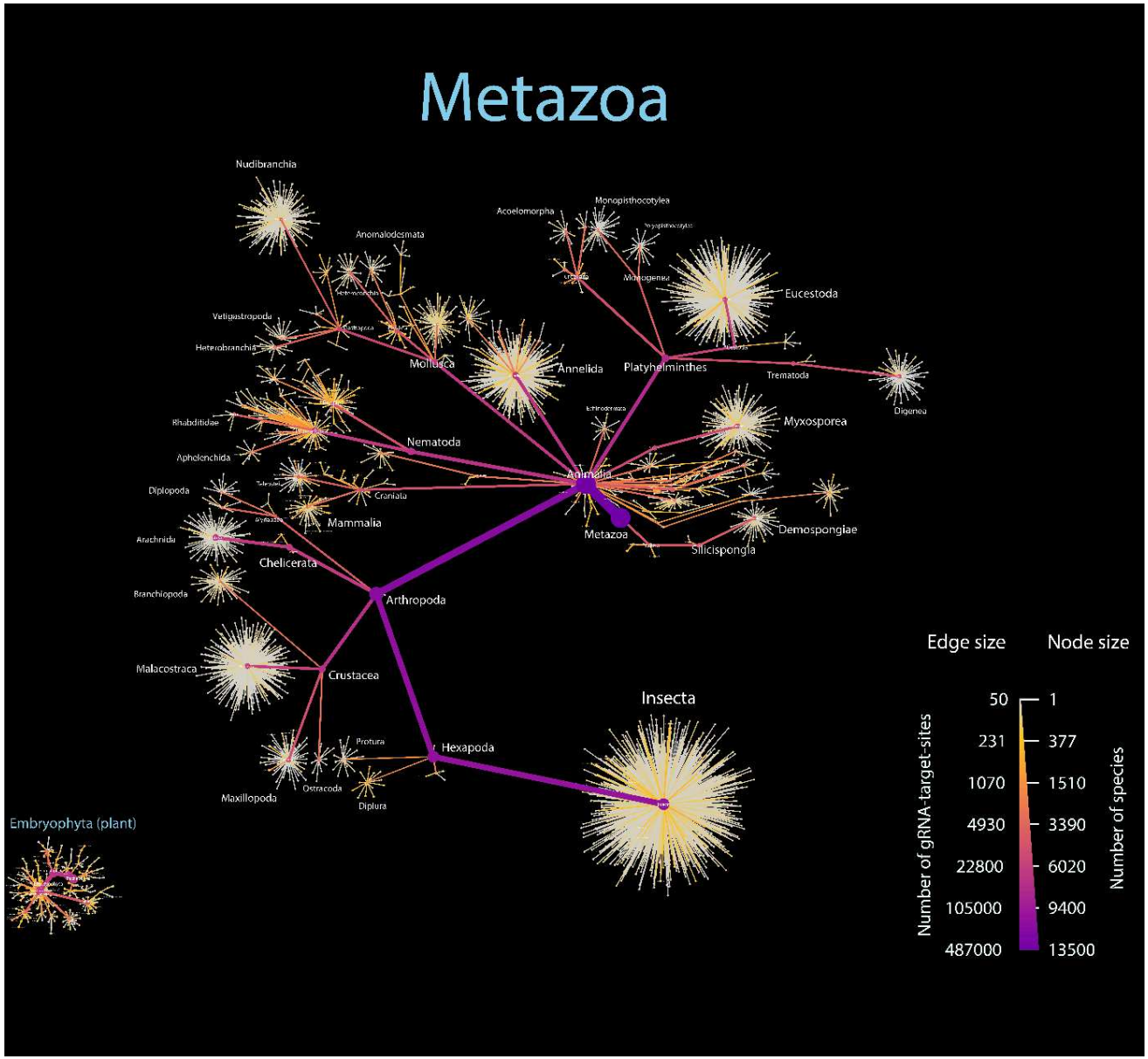


855

856 **Fig. 4** Comparison of the number of 18S rRNA gene OTUs summed for the eight oyster samples for each microeukaryotic group. The methods are based on using non-metazoan

857 primers (**NM**), blocking primers (**BP**) and CCSAS combining 18S "universal" primers and cleavage using CRISPR-Cas9 (**CCSAS**).

858


Metazoa
Nudibranchia
Anoinalodesmata
Monopisthocotylea
Acoelomorpha
Eucestoda
Vetigastropoda
Heterobranchia
Mollusca
Annelida
Platyhelminthes
Trematoda
Digenea
Rhabditidae
Nematoda
Myxosporea
Aphelenchida
Diplopoda
Craniata
Mammalia
Arachnida
Chelicerata
Metazoa
Demospongiae
Silicispongia
Branchiopoda
Arthropoda
Malacostraca
Crustacea
Insecta
Protura
Hexapoda
Maxillopoda
Ostracoda
Diplura
Embryophyta (plant)
Edge size
Node size
Number of gRNA-target-sites
50
231
1070
4930
22800
105000
487000
Number of species
1
377
1510
3390
6020
9400
13500

859

860 **Fig. 5** Illustration showing the number and taxonomic distribution of gRNA-target-sites for metazoa and plants that are available for 18S rRNA sequences in the SILVA SSU
861 database v119 [81] These gRNA-target-site oligonucleotide sequences are used for designing and synthesizing the taxon-specific and the CRISPR-Cas9-compatible sgRNAs
862 that are used to guide CRISPR-Cas9 to cut the 18S rRNA genes of metazoan or plant hosts, but not those of microeukaryotes (protists and fungi). The node size indicates the
863 number of species at each corresponding taxonomic level, while the size of the edge presents the number of gRNA-target-sites. Nodes and edges with the highest values are
864 purple, while the smallest ones are grey. Only taxa with more than 50 gRNA-target-sites per taxon are shown.
865

866

Figures

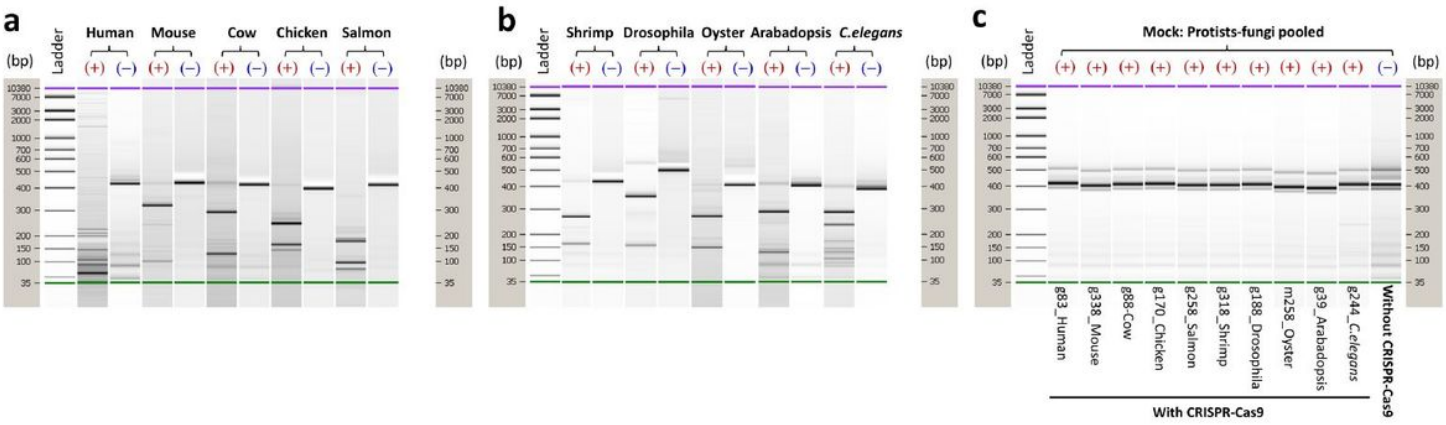


Figure 1

Agilent Bioanalyzer gel images of 18S amplicons from ten model organisms (a and b) and an artificial community of protists and fungi (c) to which Cas9 with the taxon-specific sgRNA (as shown in Table S2) was either added (+) or not (-). Gel bands show the amplicon length in base pairs (bp) relative to a DNA ladder. The labels on the X-axes of panel c indicate the ID of the taxon-specific sgRNAs and its corresponding host to target.

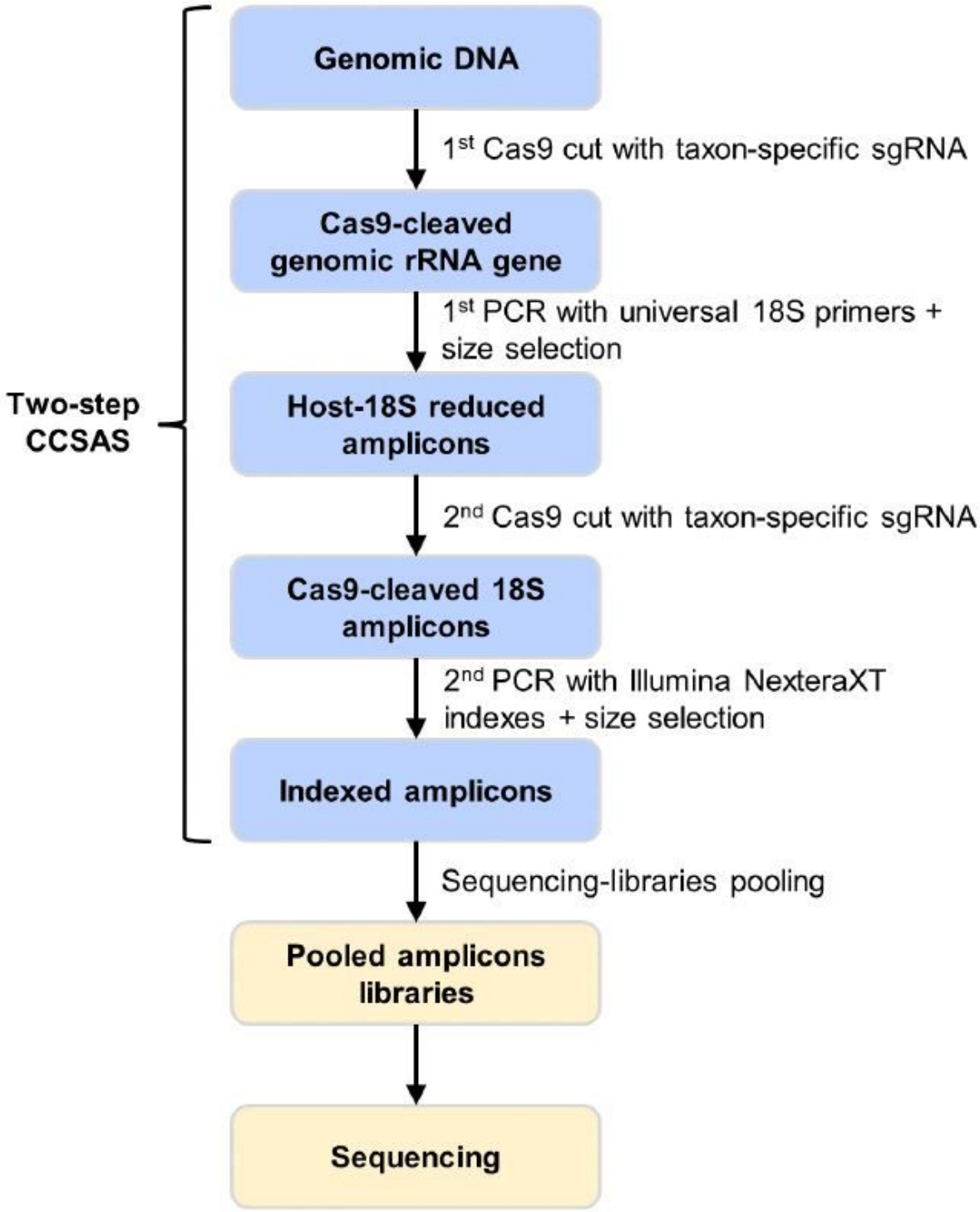


Figure 2

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

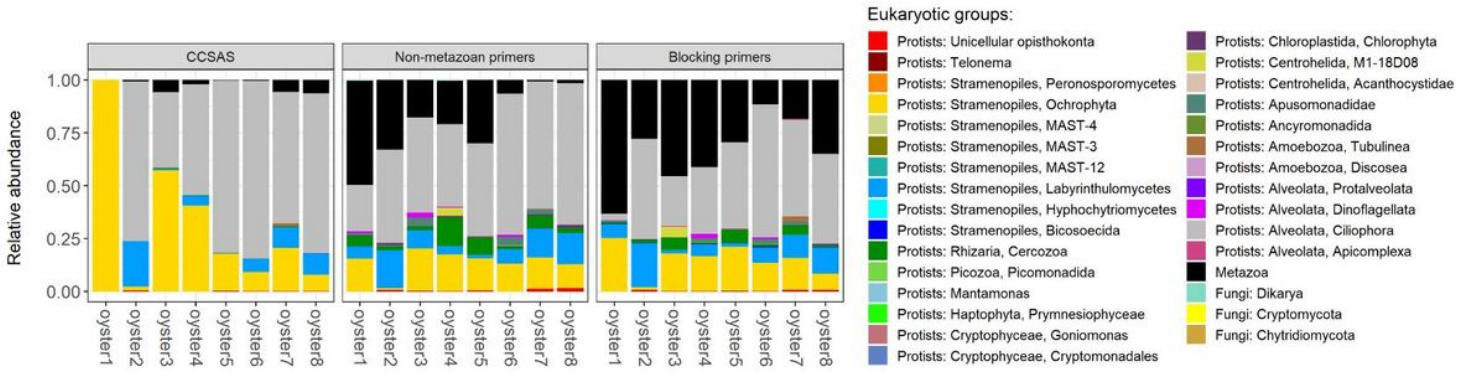


Figure 3

Relative abundances of eukaryotic groups in eight oyster samples revealed using deep-sequencing of 18S amplicons of rRNA genes using non-metazoan primers (NM), blocking primers (BP) and CCSAS using 18S "universal" primers and cleavage using CRISPR-Cas9 (CCSAS).

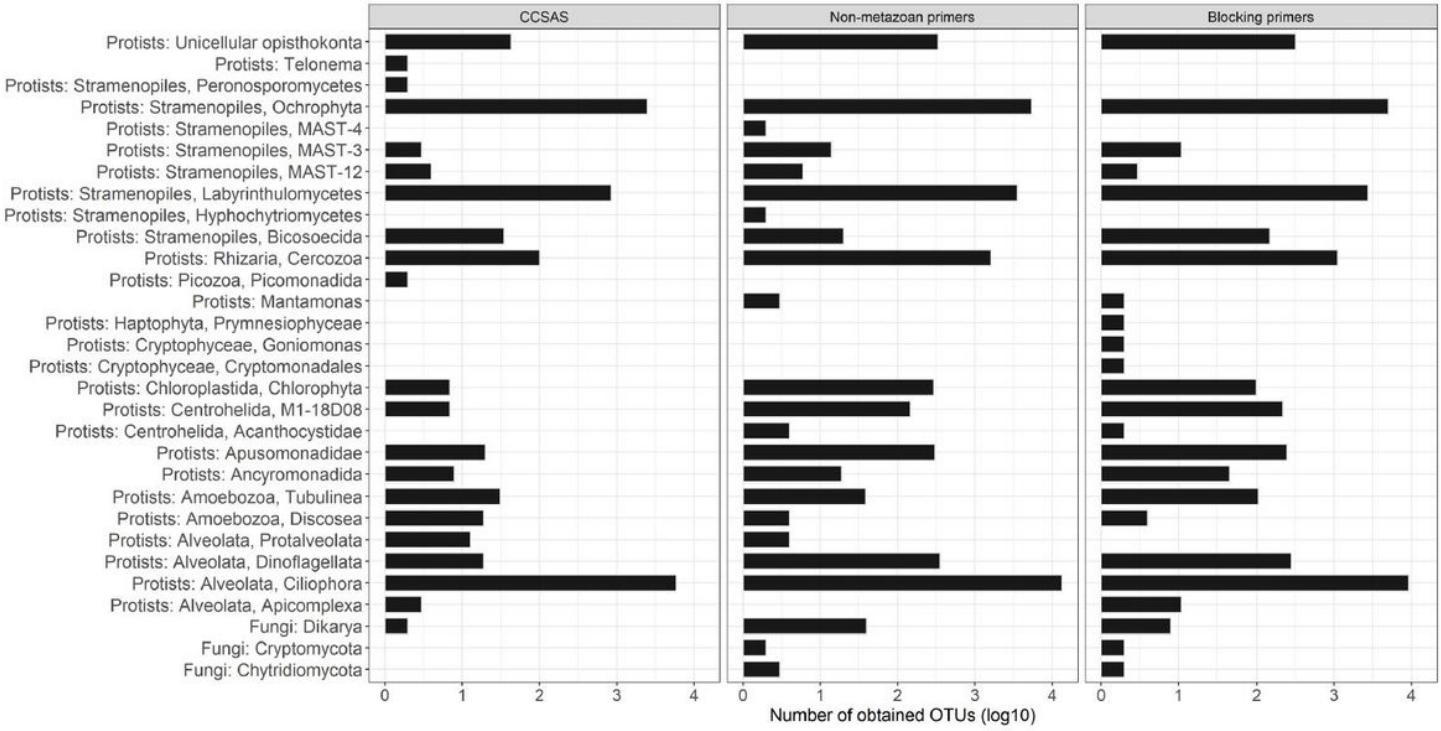


Figure 4

Comparison of the number of 18S rRNA gene OTUs summed for the eight oyster samples for each microeukaryotic group. The methods are based on using non-metazoan primers (NM), blocking primers (BP) and CCSAS combining 18S "universal" primers and cleavage using CRISPR-Cas9 (CCSAS).

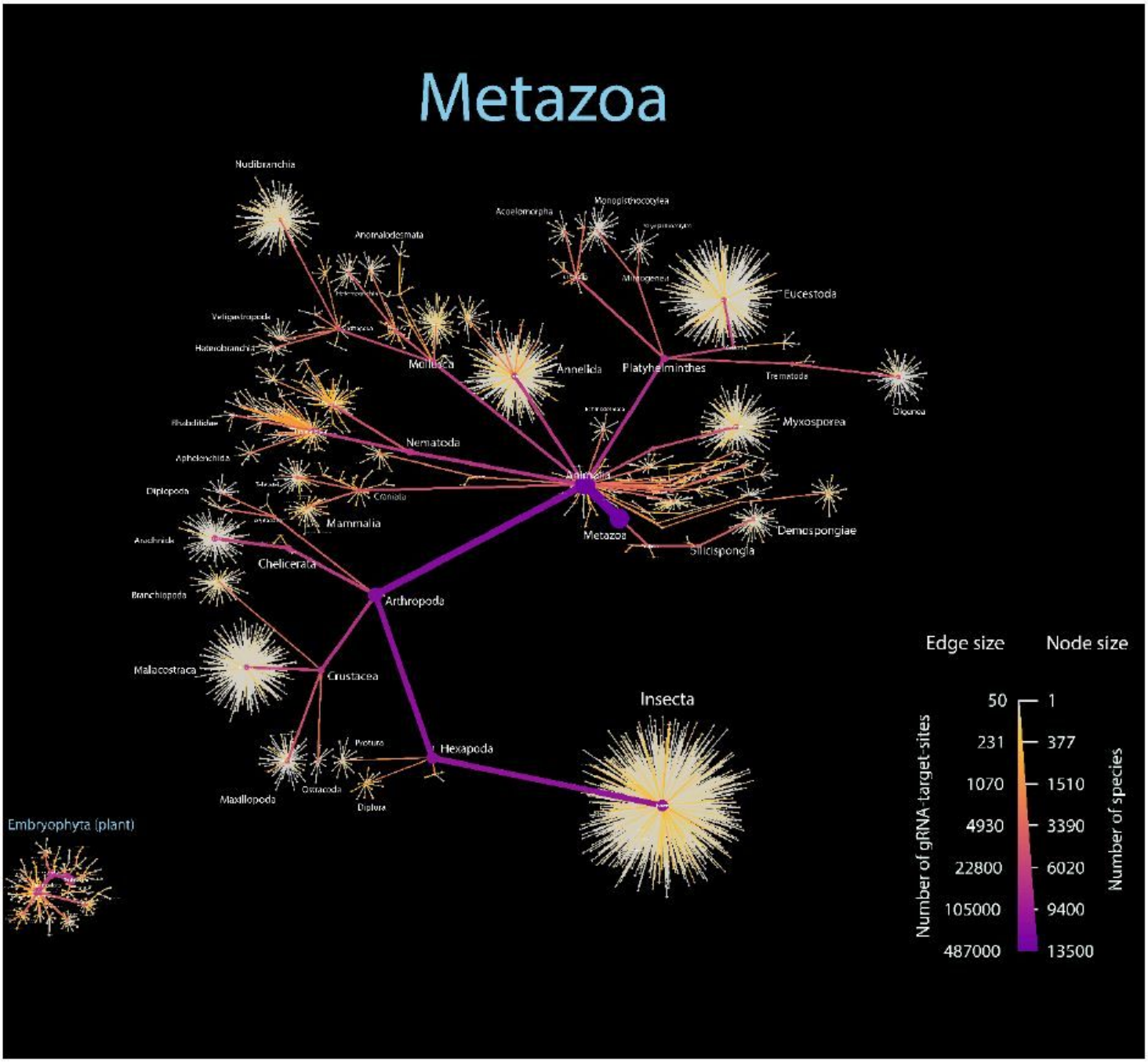


Figure 5

Illustration showing the number and taxonomic distribution of gRNA-target-sites for metazoa and plants that are available for 18S rRNA sequences in the SILVA SSU database v119 [81] These gRNA-target-site oligonucleotide sequences are used for designing and synthesizing the taxon-specific and the CRISPR-Cas9- compatible sgRNAs that are used to guide CRISPR-Cas9 to cut the 18S rRNA genes of metazoan or plant hosts, but not those of microeukaryotes (protists and fungi). The node size indicates the number of species at each corresponding taxonomic level, while the size of the edge presents the number of gRNA-target-sites. Nodes and edges with the highest values are purple, while the smallest ones are grey. Only taxa with more than 50 gRNA- target-sites per taxon are shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CCSASZhongMicrobiomeSupplementarysubmitted.pdf