Unlocking The Secrets of Soil Biodiversity: The Molecular Barcoding for Soil Protists

Thesis report submitted by

Fajri Maulana Anugrah

1049449

In

March 2023

In partial fulfilment of the requirements for the Degree of Master of Science in Plant Sciences at Wageningen University and Research, The Netherlands

Abstract

Soil protists are a group of all eukaryotes except plants, animals, and fungi that provide essential functions, such as nutrient cycling and energy transfer. One of the most common approaches to studying soil protists is via molecular tools. Sanger sequencing was applied to perform DNA-barcoding for species identification of individual protist cultures using 18S rDNA gene sequences. The results showed a 100% sequence identity with the references published on the NCBI database for twenty individual protist cultures, while nine samples showed 91-99% sequence identity. The sequences obtained were recorded for a partial region of the 18S rDNA gene sequence (between 219 and 893 bp). The quality of the sequences varied due to the presence of ambiguous signals during chromatogram analysis. On the other hand, the results demonstrate that a portion of the 18S rDNA sequence can be used to identify different individual protist cultures and ultimately create a reference database. Together, our study extends the molecular information of soil protists by including three reported references, counting ambiguous signals, and providing taxonomical information of different individual protist cultures.

Keywords: soil protist, DNA barcoding, Sanger sequencing, 18S rDNA gene, species identification

Table of Contents

1 Introduction

Protists are single-celled eukaryotic microbes that cover almost the entire eukaryotic tree of life (Geisen et al., 2018). In the tree of eukaryotic life, protists are classified into different supergroups. They are SAR (which includes Stramenopiles, Alveolata, Rhizaria), Obazoa (including Ophisthokonta), Archaeplastida, Excavata, and Amoebozoa (Geisen et al., 2018). As they are highly diverse, protists can be found in a diverse conditions, such as in environments with a wide range of temperatures and pH values, as well as in high salinity conditions (Amaral Zettler et al., 2002; De Jonckheere, 2006; Rodriguez-Zaragoza et al., 2005; Shmakova et al., 2016). Among these diverse ecosystems, the soil ecosystem is of particular interest, because more than tens of thousands of individual protist can be found in one gram of soil (Geisen et al., 2018). They live and provide significant contributions to below-ground ecosystems.

Protists can regulate the structure of microbial communities as well as the functioning of the remainder of the soil community via predation (Gao et al., 2019). About 85% of total soil protist community are predators of many microbes, such as bacteria, fungi, nematodes, and even other protists (Gao et al., 2019; Oliverio et al., 2020). Their predation is essential for the plants, as nutrients that reside inside the cells of rhizosphere microbes can be released into the soil and subsequently be taken up by plants (Gao et al., 2019). Protist predation is highly specific, as protists can selectively choose their prey, for example pathogenic microbes. Via this mechanism, protists enhance plant protection (Amacker et al., 2022). Therefore, a sustainable plant protection strategy can potentially be achieved by integrating soil protists into belowground ecosystems (Gao et al., 2019).

There are many different methodologies available to study soil protists. Up until now, a classical identification method based on morphological characteristics has been widely applied for species identification (Page, 1967). However, species identification by means of morphology is often problematic in protists (Gao et al., 2022). The problem arises particularly due to presence of cryptic species in protistan lineages (Geisen et al., 2015). Cryptic species are defined as protist species that possess similar morphological characteristics, but are genetically different (Geisen et al., 2015). As a result, morphological characteristics cannot fully identify protist species and have low taxonomic resolution (Geisen et al., 2014).

Initially, molecular tools were introduced to solve low taxonomic resolution derived from morphological characterization (Gao et al., 2022; Borg Dahl et al., 2018; Geisen et al., 2016, 2014; Kosakyan et al., 2013). In one of these tools, DNA barcoding is used to provide genetic information for species identification in individual protist samples (Gao et al., 2022; Kosakyan et al., 2013). One of the most commonly used genetic barcodes is located in the small subunit ribosomal DNA, called 18S rDNA (Borrelli et al., 2018; Gao et al., 2022; Lara et al., 2022). This gene has been widely applied for identification of protist species, as the 18S rDNA gene can be found universally among eukaryotes (Hadziavdic et al., 2014). Furthermore, the 18S rDNA gene contains highly conserved regions that can facilitate primer design targeting soil protists (Hadziavdic et al., 2014). The 18S rDNA gene also contains nine variable regions, ranging from V1 to V9 (Hadziavdic et al., 2014). The full length of the 18S rDNA gene can vary between 1500 and 4500 bp (Ishaq & Wright, 2014). In Figure 1, a visual representation of the 18S rDNA gene is shown.

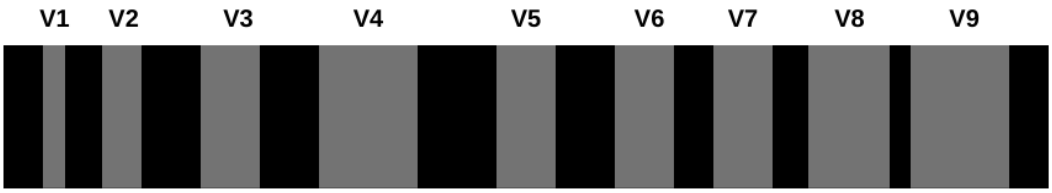

Figure 1. Schematic structure of the18S rDNA gene, including nine variable regions (V1-V9) from left to right.

The aim of the current study is to provide molecular information on protist cultures in order to obtain taxonomic information and to create a reference database of the protist cultures in the Laboratory of Nematology, WUR. Protist cultures used in this study has been purified to remove any excess bacteria in laboratory setting and contain one individual taxa. The reference database was created based on Sanger-sequencing the full-length of 18S rDNA gene sequences.

2 Materials and Methods

2.1 Preparation of Protist Cultures

This study was conducted at the Laboratory of Nematology, WUR. Sixty-seven protist individual cultures (Appendix 1) previously obtained and grown at the Laboratory of Nematology were subcultured in Petri dishes (50 x 15 mm) containing 10% of Nutrient Broth (NB)- Neff's Modified Amoebae Saline (NMAS) and 90% NMAS media. Protist growth was established by incubating protist cultures at 16°C in the dark for 3 to 5 days (Amacker et al., 2020). Afterward, using the inverted IXplore Standard microscope was used to check the morphological characteristics using 10x and 20x magnifications. Three Petri dishes containing similar protist cultures were collected in 50 ml centrifuge tubes and filled with NMAS up to 50 ml. Following this process, tubes were centrifuged (Heraeus, US) for 5 minutes at 1200rpm. Supernatant was discarded until 15 ml solution was left in the tubes to remove the excessive amount of bacteria. The remaining solution was filled up to 40 ml and centrifugation was repeated twice with the same configuration, as mentioned above. After the last centrifugation, the supernatant was removed until an end volume of approximately 25 ml.

2.2 DNA Extraction

After the washing steps were completed, the DNA extraction was performed using the protocol used by Holterman et al. (2006). 25 µl of the culture was placed in a 0.2 ml polymerase chain reaction (PCR) tube and mixed with a lysis buffer. Afterward, lysis was carried out in a Thermomixer for 30 minutes at 65 °C and 750 rpm, followed by a 5-minute incubation at 100°C. The lysate was used immediately or stored at –20 °C for further analysis.

2.3 Primer Testing and PCR Assay

The extracted DNA was prepared for primer testing and PCR assay. In the first batch, we tested different primer combinations targeting the universal barcode, 18S rDNA (Appendix 2). Several common regions, such as V4, V5, and V9 were targeted. The second batch of primer testing focused on amplifying different markers, such as 28S rDNA and COI (Appendix 2). Each forward (5 µl) and reverse primer (5 µl) was prepared individually in a tube containing 190 µl of Milli-Q water (MQ). In a total reaction volume of 20 µl, 2.5 µl DNA template was mixed with 2.5 µl primer pairs, 5 µl MQ water, and 10 µl Phire Hot Start II DNA Polymerase.

The PCR cycles for 18S rDNA and 28S rDNA genes started with DNA denaturation at 95°C for 3 minutes, followed by an annealing step at 63°C for 30 seconds and extension at 72°C for 1 minute. In total, the PCR consisted of 40 amplification cycles. The PCR cycles for COI had 35 cycles with the following parameters: denaturation for 1 minute at 95°C, annealing step for 1.5 minutes at 72°C, and extension at 72°C for 7 minutes.

All PCR products (5 ul) were subjected to gel electrophoresis in 1.5% agarose dissolved in TAE buffer. 1 Kb Plus DNA ladder (Invitrogen) was added to verify the presence of PCR products. Only PCR products with single and clear bands were sent for Sanger sequencing to ensure amplification success.

2.4 Sequence analysis and Interpretation of Sequencing Chromatograms

Prior species identification, sequences were assessed to check any low-quality reads or background noise. The low quality-reads are typically found at the beginning and end of the sequences (Fig. 1A). Using 4 Peak (version 1.8) (https://nucleobytes.com/4peaks/), those low-quality signals were removed. Additionally, double peaks and ambiguous signals were marked as ‘N’ for further analysis (Fig. 1B).

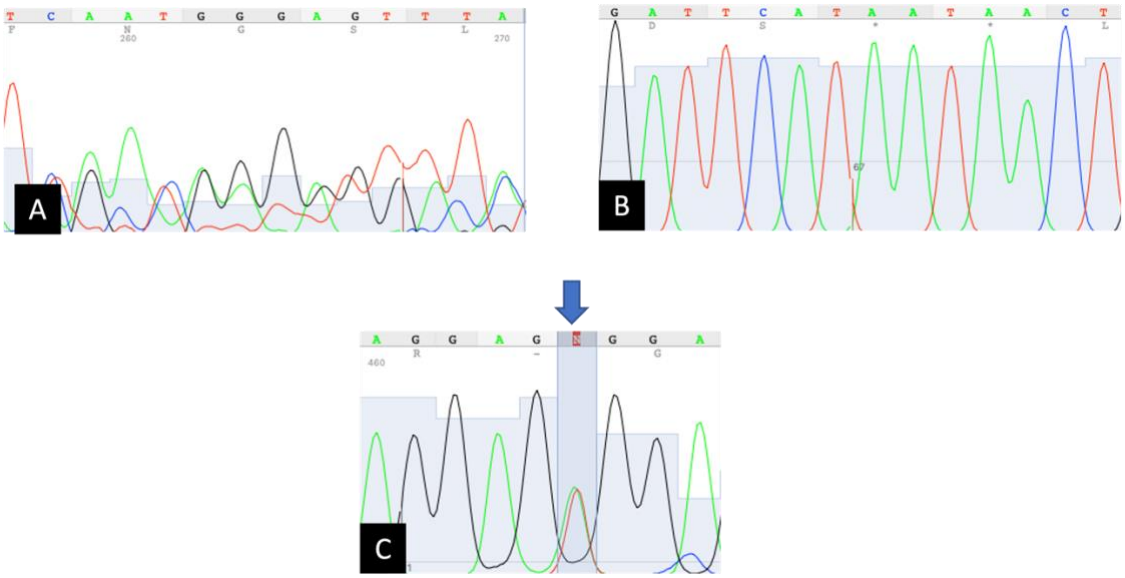

Figure 2. Visual examples of chromatogram analysis using 4 Peak. Low-quality reads show ambiguous signals (A), clear signals (B), and a blue arrow indicates two peaks were detected in the chromatogram and annotated as 'N'.

The taxonomic information of each protist culture was compared to three reference sequences found in the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The criteria used to select the references was based on percentage identity and query coverage. All datasets were aligned using Muscle method within Mega 11 (Tamura et al., 2021) and further manually adjusted in 4 Peak (version 1.8) to double check aligned regions base pairs.

3 Results

3.1 Molecular information of individual protist cultures

As quality differed per culture, sequence quality of twenty-nine individual protist cultures was divided into two categories. The first category consists of sequences that have clear signals and separated peaks. These sequences were 34M, 139, 75, 45, 119, 32M, S9-F4, S8-H7, RE2, P5, P10, S9-G12, S24-D2, SG2-H2, 19M, 37, and P8. Results shown in Table 1 represent the remaining sequences that had some ambiguous signals (N). The least ambiguous signal of 1 was detected in five protist cultures, and the highest ambiguous signal of 6 was detected in two protist cultures. The percentage of ambiguous signals varies between 0.13% and 2.23%, depending on the lengths of the sequences.

Table 1. The ambiguous signals observed (N) in the chromatogram analysis. Ambiguous signals, their percentage, and position are depicted.

Culture	Total ambiguous signals	Percentage ambiguous signals (%)	Position
P2881	1	0.20	248
C5D3	1	0.13	529
24M	1	0.17	465
25M	1	0.13	566
33	1	0.20	294
2M	2	0.47	125, 126

Culture	Total ambiguous signals	Percentage ambiguous signals (%)	Position
114-W21	3	0.38	206, 207, 638
S28-D2	3	0.38	285, 288, 302
43M	3	0.57	418, 506, and 514
16	5	0.58	139, 140, 424, 634, 638
133	6	2.23	150, 182, 187, 196, 200, 229
P-147	6	0.88	403, 409, 436, 486, 592, 600

Taxonomical information is shown in the following examples. The partial 18S rDNA gene sequence of culture 139 was 100% identical to the first reference (Table 2): *Cercomonas* sp. culture C5D3 (Amacker et al., 2022) and the second reference: *Cercomonas lenta* (Bass et al., 2009). The third reference shows that culture 139 more closely resembles *Cercomonas* sp. strain SF52_P_Pla (Flues et al., 2018) with both 99% identity and query coverage.

The BLAST search of the partial 18S rDNA region of culture 45 suggests that this culture can be classified within the genus *Vannella.* The first and the second references show 100% sequence identity to culture 45: *Vannella* sp. culture Amoebae3 and *Vanella* sp. culture 33 (Amacker et al., 2022), respectively. *Vannella planctonica* strain A2FBB (Van Wichelen et al., 2016), as a third reference, had 93% identity and 100% query coverage.

The taxonomical identification suggests that culture 19M is closely affiliated to *Dimastigella trypaniformis* at 100% (Callahan et al., 2002). In contrast, the BLAST search suggests that culture 19M more closely resembles *Dimastigella mimosa strain HFCC20* with 97% identity and 100% query coverage (Scheckenbach et al., 2006) *and Dimastigella trypaniformis* strain SCCAP DIM74 with 97% identity and 100% query coverage (Koch & Ekelund, 2005).

The taxonomical information of some of our cultures resembles environmental samples. For example, culture S9-G12 was 96% identical to uncultured chrysophyte clone PR4_4E_14 (Lara et al., 2011) and resembles uncultured freshwater eukaryote with a sequence identity of 92%. Percentage identity is less than 100%, because some nucleotide differences were recorded at several positions, namely 115, 151, and 164. In the sequence, these positions were detected as C, C, and A with clear signals, while the reference sequences were T, T, and G.

By incorporating the taxonomic information from three reference sequences obtained in the NCBI database, several prominent supergroups and genera were identified (Table 2). SAR, Amoebozoa, and Exacavata are the three most prevalent protist supergroups. Regarding SAR, *Ochromonas, Cercomonas, Apoikia, Spumella, Colpoda,* and *Heteromita* are the most frequently observed genera. There are five genera for Excavata, including *Allovahlkampfia, Bodo, Naegleria, Dimastigella,* and *Vahlkampfia.* The Amoebozoa supergroup contains *Vermamoeba, Cochliopodium, Didymium, Cryptodifflugia,* and *Vannella.*

Table 2. Taxonomical information of different individual protist cultures used in this study, excluding uncultured environmental sample sequences.

Super group	Taxonomical information	Group	Genus	Individual Culture
SAR	Eukaryota; Sar; Stramenopiles; Ochrophyta; Chrysophyceae	Chrysophyceae	*Ochromonas*	34M and 43M
SAR	Eukaryota; Sar; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae	Cercomonads	*Cercomonas*	139, P5, S24-D2, P8, 16, and C5D3
SAR	Eukaryota; Sar; Stramenopiles; Ochrophyta; Chrysophyceae; Apoikiaceae	Chrysophyceae	*Apoikia*	32M, S9-F4, RE2, S9-G12, and 119
SAR	Eukaryota; Sar; Stramenopiles; Ochrophyta; Chrysophyceae;Chromulinales; Chromulinaceae	Chrysophyceae	*Spumella*	S8-H7
SAR	Eukaryota; Sar; Alveolata; Ciliophora; Intramacronucleata; Colpodea; Colpodida; Colpodidae	Colpodea	*Colpoda*	P10

Super group	Taxonomical information	Group	Genus	Individual Culture
SAR	Eukaryota; Sar; Rhizaria; Cercozoa; Cercomonadida; Heteromitidae	Cercomonads	*Heteromita*	SG2-H2 and 37
Excavata	Eukaryota; Discoba; Heterolobosea; Tetramitia; Eutetramitia; Acrasidae; Allovahlkampfia	Heteroloboseans	*Allovahlkampfia*	S28D2
Ecavata	Eukaryota; Discoba; Euglenozoa; Kinetoplastea; Metakinetoplastina; Eubodonida; Bodonidae	Kinetoplastids	*Bodo*	133
Excavata	Eukaryota; Discoba; Heterolobosea; Tetramitia; Eutetramitia; Vahlkampfiidae	Heteroloboseans	*Naegleria*	P2881
Excavata	Eukaryota; Discoba; Euglenozoa; Kinetoplastea; Metakinetoplastina; Neobodonida; Rhynchomonadidae	Kinetoplastids	*Dimastigella*	19M
Excavata	Eukaryota; Discoba; Heterolobosea; Tetramitia; Eutetramitia; Vahlkampfiidae	Heteroloboseans	*Vahlkampfia*	24M
Amoebozoa	Eukaryota; Amoebozoa; Tubulinea; Echinamoebida	Echinamoebids	*Vermamoeba*	25M
Amoebozoa	Eukaryota; Amoebozoa; Discosea; Himatismenida; Cochliopodiidae	Himastimenids	*Cochliopodium*	33

Super group	Taxonomical information	Group	Genus	Individual Culture
Amoebozoa	Eukaryota; Amoebozoa; Evosea; Eumycetozoa; Myxogastria; Myxogastromycetidae; Physariida; Didymiaceae	Myxogastrids	*Didymium*	2M and 114-W21
Amoebozoa	Eukaryota; Amoebozoa; Tubulinea; Elardia; Arcellinida; Phryganellina; Cryptodifflugiidae	Arcellinids	*Cryptodifflugia*	75
Amoebozoa	Eukaryota; Amoebozoa; Discosea; Flabellinia; Vannellida	Vannelids	*Vannella*	P-147

This study also demonstrates the existence of sequences that are specific to uncultured environmental samples (Table 3). Uncultured freshwater eukaryotes, uncultured chrysophytes, and uncultured cercozoan cultures are examples of this type of sequences. The categorization is done using groups, because some of these sequences could not be identified up to the genus level. The uncultivated chrysophyte, for instance, is a member of the protist group Chrysophytes (golden algae) (Lara et al., 2011). The Cercomonads group includes uncultured cercozoan cultures. No identified groups could be obtained for uncultured eukaryotes. Therefore, these taxa were grouped individually as environmental samples.

Table 3. Taxonomical information of uncultured environmental sample sequences. N/A = no data available.

Super group	Taxonomical information	Group	Individual Culture
SAR	Eukaryota; Sar; Stramenopiles; Ochrophyta; Chrysophyceae; environmental samples	Chrysophyceae	43M, 34M, 119, 32M, S9-F4, RE2, S9-G12, and S8-H7
SAR	Eukaryota; Sar; Rhizaria; Cercozoa; environmental samples	Cercomonads	37
N/A	Eukaryota; environmental samples	N/A	75, 32M, S9-F4, RE2, P10, S9-G12, SG2-H2, 37, P-147, and 43M

3.2 Reference database

Table 4 provides an overview of a reference database of the individual protist cultures used in this study. Overall, the length of the sequences obtained varied from 219 to 893 bp.

Table 4. A reference database of twenty-nine individual protist cultures. Shown are three reference sequences (Ref) along with the percentage identity (% Ident.); query coverage (Query cov.); GenBank accession number (Acc).

Culture	Length (bp)	Ref 1	% Ident.	Query Cov.	Acc	Ref 2	% Ident.	Query cov.	Acc	Ref 3	% Ident.	Query cov.	Acc
34M	219	Uncultured chrysophyte clone PR4_4E_14	97	100	GQ330571	Uncultured chrysophyte clone PR2_3E_121	96	100	GQ330576	*Ochromonas tuberculata*	94.5	100	AF123293
139	463	*Cercomonas* sp. culture C5D3	100	100	OL989987	*Cercomonas lenta*	100	100	FJ790691	*Cercomonas* sp. strain SF52_P_Pla	99	99	MG77558
75	621	*Cryptodifflugia operculata*	100	100	JF694280	Uncultured eukaryote	97	100	MT533860	*Cryptodifflugia oviformis*	98	93	JQ366062
45	559	*Vannella* sp. culture Amoebae3	100	100	MN907843	*Vannella* sp. culture 33	100	100	OL989989	*Vannella planctonica* strain A2FBB	93	100	KP719193
119	599	Uncultured	96	98	GQ330571	Uncultured chrysophyte	98	82	GQ330576	*Apoikia lindahlii*	96	82	FJ971855

Culture	Length (bp)	Ref 1	% Ident.	Query Cov.	Acc	Ref 2	% Ident.	Query cov.	Acc	Ref 3	% Ident.	Query cov.	Acc
		chrysophyte clone PR4_4E_14				clone PR2_3E_121							
32M	827	Uncultured chrysophyte clone PR4_4E_14	96	100	GQ330571	Uncultured eukaryote clone 270C05	92	100	KJ925240	*Apoikia lindahlii*	91	100	FJ971855
S9-F4	728	Uncultured chrysophyte clone PR4_4E_14	96	100	GQ330571	Uncultured freshwater eukaryote	91	100	AB721020	*Apoikia lindahlii*	90	100	FJ971855
S8-H7	657	*Spumella* sp. OF-40	100	100	KF651119	Uncultured Chrysophyceae	100	100	AB749100	Uncultured chrysophyte	100	100	AB520724
RE2	715	Uncultured chrysophyte clone PR4_4E_14	96	100	GQ330571	Uncultured freshwater eukaryote	92	100	AB721020	*Apoikia lindahlii*	91	100	FJ971855
P5	696	*Cercomonas sp. strain SF42_R_Pla*	100	100	MG775567	*Cercomonas* sp. strain SF34_R_Tri	100	100	*MG775566*	*Cercomonas* sp. culture S24D2	100	100	OL989988

Culture	Length (bp)	Ref 1	% Ident.	Query Cov.	Acc	Ref 2	% Ident.	Query cov.	Acc	Ref 3	% Ident.	Query cov.	Acc
P10	733	*Colpoda cucullus*	99	100	LC441007	Uncultured eukaryote	99.7	100	LC222948	Uncultured eukaryote clone HRT3hrs_A1_euk_14	99	100	LC222943
P2881	340	*Naegleria clarki* strain MSED4	100	100	JQ271704	*Naegleria clarki* strain ALM1A	100	100	*JQ271698*	*Naegleria clarki* strain 6HFKL	100	100	JQ271696
S9-G12	893	Uncultured chrysophyte clone PR4_4E_14	96	93	GQ330571	Uncultured freshwater eukaryote	92	100	AB622338	*Apoikia lindahlii*	91	100	FJ971855
S24-D2	714	*Cercomonas* sp. strain SF34_R_Tri	100	100	MG775566	*Cercomonas* sp. culture S24D2	100	100	OL989988	*Cercomonas* sp. strain SF42_R_Pla	100	100	MG775567
SG2-H2	867	*Heteromita globosa*	100	100	U42447	Uncultured eukaryote clone NYS000592	99	99	JX394796	Uncultured cercozoan	99	100	EU709218
19M	654	*Dimastigella trypaniformis*	100	100	AY028447	*Dimastigella mimosa*	97	100	DQ207576	*Dimastigella trypaniformis* strain SCCAP DIM74	97	100	AY998641

Culture	Length (bp)	Ref 1	% Ident.	Query Cov.	Acc	Ref 2	% Ident.	Query cov.	Acc	Ref 3	% Ident.	Query cov.	Acc
24M	572	*Vahlkampfia inornata*	100	100	AJ224887	*Vahlkampfia* sp. culture CN7	97	100	MT739329	*Vahlkampfia* sp. culture - 4ES E1	97	94	MZ338491
25M	735	*Vermamoeba vermiformis* strain MG1	100	100	KU519742	*Vermamoeba vermiformis* culture 1	99	100	KT185625	*Hartmannella vermiformis* strain PM11	99	100	JQ271689
33	521	*Cochliopodium minus* clone 9804	100	100	KJ569708	*Cochliopodium minus* clone 1105	99	100	KJ569717	*Cochliopodium minus* clone 1104	99	100	KJ569716
37	726	*Heteromita* sp. strain DH1	100	98	KR047867	Uncultured cercozoan culture HetChi22	99	100	EU709218	Uncultured eukaryote clone NYS000592	99	100	JX394796
P8	610	*Cercomonas* sp. culture C5D3	100	100	OL989987	*Cercomonas lenta*	100	100	FJ790691	*Cercomonas nebulosa*	100	100	FJ790690
2M	420	*Didymium dachnayum*	100	100	AY062881	*Didymium* sp. ECH49/I	98	100	EF118761	*Pseudodidymium cryptomastigophorum*	96	100	EF118760

Culture	Length (bp)	Ref 1	% Ident.	Query Cov.	Acc	Ref 2	% Ident.	Query cov.	Acc	Ref 3	% Ident.	Query cov.	Acc
16	853	*Cercomonas* sp. culture C5D3	100	100	OL989987	Soil flagellate AND25	96	100	AY965868	*Neocercomonas* sp. strain New Zealand 1-7E	96	100	AY884322
C5D3	729	*Cercomonas* sp. culture C5D3	100	100	OL989987	*Cercomonas lenta* strain WA84	100	100	FJ790691	*Cercomonas nebulosa* strain xt164	99	100	FJ790690
P-147	680	*Vannella sp.*	100	100	OL989990	*Vannella* sp. ACN1	98	100	JQ271724	Uncultured eukaryote	98	68	HQ999648
S28D2	776	*Allovahlkampfia* sp.	100	100	MT739328	*Allovahlkampfia spelaea*	95	100	EU696948	*Heterolobosea* sp. OSA	94	100	DQ388520
43M	523	Uncultured chrysophyte clone PR4_4E_14	96	93	GQ330571	Uncultured eukaryote	98	82	KJ925240	*Ochromonas tuberculata*	97	82	AF123293
114-W21	784	*Didymium* sp. ECH49/I	98	100	EF118761	*Didymium* sp. ECH1/I	98	100	EF118758	*Didymium dachnayum*	96	100	AY062881
133	268	*Bodo* sp. TGS2	100	100	AB585965	*Bodo saltans* strain HFCC12	100	100	DQ207569	*Bodo saltans*	100	100	AY028452

4 Discussion

This study provides molecular information, taxonomic information, and a reference database of the individual protist cultures provided by the Laboratory of Nematology, WUR. The 18S rDNA gene as a genetic marker used in the present study produced sequences that varied between 219 and 893 bp. This suggests that the sequences obtained are only part of the18S rDNA gene marker. The full length of the 18S rDNA gene sequences ranges from 1500 bp to 4500 bp (Ishaq & Wright, 2014).

The limited genetic information presented in this study may in part be caused by the application of a sequencing platform that affects the length of sequences obtained (Geisen et al., 2014; Kudryavtsev et al., 2005). Previous research has shown that Sanger sequencing can generate sequences between 800 and 1000 bp (Gupta & Verma, 2019). Additionally, sequences generated by Sanger sequencing may contain low-quality signals. Low-quality signals in Sanger sequencing are commonly found within the first 15-40 bp, as this is where primer annealing occurs for sequencing initiation (Crossley et al., 2020).

It is important to note that identification based on 18S rDNA gene can be potentially biased because of low taxonomic resolution (Lara et al., 2022). This low taxonomic resolution can be explained by the fact that the 18S rDNA gene is not variable enough, as was reported in previous studies (Nassonova et al., 2010; Borrelli et al., 2018; Lara et al., 2022). For example, Nassonova et al., (2010) in their study demonstrated that the 18S rDNA gene failed to differentiate three different *Vanella* species: *V.arabica*, *V. bursella*, and *V. calycinucleolus.* This was caused by low inter-specific sequence divergences in the 18S rDNA gene sequence of these three species. The low-interspecific sequence divergences could also have been found in this study. For example, sequences obtained from culture 139 showed the same identity (100%) for *Cercomonas* sp. culture C5D3 (Amacker et al., 2022) and *Cercomonas lenta* (Bass et al., 2009).

The 18S rDNA gene, however, also exhibits considerable variability, particularly in a number of protist groups that are frequently found in soil, including Amoebozoa and Heterolobosea (Lara et al., 2022). A study by Kudryavtsev & Gladkikh (2017) showed that there is genetic variation in the genus *Ripella* (Amoebozoa, Vannellida) based on the 18S rDNA gene sequence. This genetic variability is possibly associated with the high copy number and the presence of extrachromosomal structures of the 18S rDNA gene, which is common in many types of unicellular eukaryotes (Gong et al., 2013; Long & Dawid, 1980).

It is challenging to obtain sequences that have precisely 100% identity with the reference sequence. This could be caused by several factors. Firstly, the 18S rDNA gene that was used in this study was the only genetic barcode applied for protist culture identification. This gene has several limitations, as has been discussed in previous paragraphs. Another possible reason could be that the NCBI database is not fully completed and curated (Gao et al., 2022). As such, information lacking from the database could hinder the acquisition of accurate molecular and taxonomical information of individual protist cultures. Subsequently, this could have impaired the formation of an accurate reference database in the current study.

All in all, the final product, which is the reference database provided here, broadens the molecular and taxonomical information of individual protist cultures for the Laboratory of Nematology, WUR.

5 Conclusion and future recommendations

This study demonstrates that a portion of the 18S rDNA sequence can be used to identify different individual protist cultures. This gene provides molecular information for acquisition of taxonomic information and formation of a reference database for the Laboratory of Nematology, WUR. However, using only the 18S rDNA gene is not sufficient to obtain a high taxonomic resolution, as this gene can only provide partial taxonomic profiles.

Future studies should include long-read sequencing platforms, such as (MinION) and Pacific Biosciences (PacBio). Using these platforms, reads up to 300 kb can be sequenced (Gupta & Verma, 2019). Furthermore, both long-read sequencing platforms can provide real-time data and have low sequencing errors (Gupta & Verma, 2019; Jain et al., 2016; Maloney et al., 2020; Orr et al., 2018).

In addition to the use of long-sequencing platforms, also the use of multiple genetic markers is required for obtaining reliable sequencing results. A promising genetic barcode, such as cytochrome c oxidase I (COI) is important to be tested in the future on the individual protist cultures. This gene has been shown to be informative for DNA barcoding, for instance in amoebae (Kosakyan et al., 2012; Nassonova et al., 2010) and ciliates (Chantangsi et al., 2007). Additionally, the 28S rDNA gene, particularly the D1-D2, D2-D3 has been positively tested for sequencing in some protistan lineages (Gentekaki & Lynn, 2009; Trobajo et al., 2010).

In conclusion, DNA barcoding focusing on genes other than the 18S rDNA gene together with long-read sequencing could be able to provide a high resolution for acquisition of taxonomic information, which in turn could be used for creating a more in-depth reference database for individual protist cultures.

Acknowledgement

I would like to express gratitude and appreciation to the following people for their guidance, support, inspiration, and assistance during the entire period of research works and writing up of the thesis: my thesis supervisors Alejandro Berlinches de Gea and Assistant Professor Stefan Geisen, Laboratory of Nematology, Department of Plant Sciences. Lab technicians: Sven van den Elsen, for his help with the laboratory protocols and guidance in the lab. To Sara Cazzaniga, who introduced me to DNA extraction. To the entire Molecular Soil Ecology group, students, and supervisors of my thesis ring for providing fruitful feedbacks and discussion.

I wish to thank the Indonesia Endowment Fund for Education (LPDP) for providing me with scholarship endorsement throughout my master's study. Also, I would like to thank my study advisor, Cindy ten Brooke for her assistance throughout my study at Wageningen University and Research, The Netherlands.

Both my late mother and father, I have all your guidance that I will always remember. Last but not least, my sister and my brother for their endless support while I am being away from home.

References

Amacker, N., Gao, Z., Agaras, B. C., Latz, E., Kowalchuk, G. A., Valverde, C. F., Jousset, A., & Weidner, S. (2020). Biocontrol Traits Correlate With Resistance to Predation by Protists in Soil Pseudomonads. *Frontiers in Microbiology*, *11*, 614194. https://doi.org/10.3389/fmicb.2020.614194

Amacker, N., Gao, Z., Hu, J., Jousset, A. L. C., Kowalchuk, G. A., & Geisen, S. (2022). Protist feeding patterns and growth rate are related to their predatory impacts on soil bacterial communities. *FEMS Microbiology Ecology*, *98*(6), fiac057. https://doi.org/10.1093/femsec/fiac057

Amaral Zettler, L. A., Gómez, F., Zettler, E., Keenan, B. G., Amils, R., & Sogin, M. L. (2002). Eukaryotic diversity in Spain's River of Fire. *Nature*, *417*(6885), 137–137. https://doi.org/10.1038/417137a

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, *4*(7), e6372. https://doi.org/10.1371/journal.pone.0006372

Bass, D., Howe, A. T., Mylnikov, A. P., Vickerman, K., Chao, E. E., Edwards Smallbone, J., Snell, J., Cabral Jr, C., & Cavalier-Smith, T. (2009). Phylogeny and Classification of Cercomonadida (Protozoa, Cercozoa): Cercomonas, Eocercomonas, Paracercomonas, and Cavernomonas gen. nov. *Protist*, *160*(4), 483–521. https://doi.org/10.1016/j.protis.2009.01.004

Borg Dahl, M., Brejnrod, A. D., Unterseher, M., Hoppe, T., Feng, Y., Novozhilov, Y., Sørensen, S. J., & Schnittler, M. (2018). Genetic barcoding of dark-spored myxomycetes (Amoebozoa)-Identification, evaluation and application of a sequence similarity threshold for species differentiation in NGS studies. *Molecular Ecology Resources*, *18*(2), 306–318. https://doi.org/10.1111/1755-0998.12725

Borrelli, C., Hou, Y., Pawlowski, J. W., Holzmann, M., Katz, M. E., Chandler, G. T., & Bowser, S. S. (2018). Assessing SSU rDNA Barcodes in Foraminifera: A Case Study using *Bolivina quadrata*. *Journal of Eukaryotic Microbiology*, *65*(2), 220–235. https://doi.org/10.1111/jeu.12471

Callahan, H. A., Litaker, R. W., & Noga, E. J. (2002). Molecular Taxonomy of the Suborder Bodonina (Order Kinetoplastida), Including the Important Fish

Parasite, Ichthyobodo necator. *The Journal of Eukaryotic Microbiology*, *49*(2), 119–128. https://doi.org/10.1111/j.1550-7408.2002.tb00354.x

Chantangsi, C., Lynn, D. H., Brandl, M. T., Cole, J. C., Hetrick, N., & Ikonomi, P. (2007). Barcoding ciliates: A comprehensive study of 75 isolates of the genus Tetrahymena. *International Journal of Systematic and Evolutionary Microbiology*, *57*(Pt 10), 2412–2423. https://doi.org/10.1099/ijs.0.64865-0

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., Clement, T., & Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, *32*(6), 767–775. https://doi.org/10.1177/1040638720905833

De Jonckheere, J. F. (2006). Isolation and molecular identification of free-living amoebae of the genus Naegleria from Arctic and sub-Antarctic regions. *European Journal of Protistology*, *42*(2), 115–123. https://doi.org/10.1016/j.ejop.2006.02.001

Flues, S., Blokker, M., Dumack, K., & Bonkowski, M. (2018). Diversity of Cercomonad Species in the Phyllosphere and Rhizosphere of Different Plant Species with a Description of *Neocercomonas epiphylla* (Cercozoa, Rhizaria) a Leaf-Associated Protist. *Journal of Eukaryotic Microbiology*, *65*(5), 587–599. https://doi.org/10.1111/jeu.12503

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Gao, Z., Jousset, A., Kowalchuk, G. A., & Geisen, S. (2022). Five Groups in the Genus Allovahlkampfia and the Description of the New Species Vahlkampfia bulbosis n.sp. *Protist*, *173*(3), 125870. https://doi.org/10.1016/j.protis.2022.125870

Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., & Jousset, A. (2019). Protists: Puppet Masters of the Rhizosphere Microbiome. *Trends in Plant Science*, *24*(2), 165–176. https://doi.org/10.1016/j.tplants.2018.10.011

Geisen, S., Bonkowski, M., Zhang, J., & De Jonckheere, J. F. (2015). Heterogeneity in the genus Allovahlkampfia and the description of the new genus Parafumarolamoeba (Vahlkampfiidae; Heterolobosea). *European Journal of Protistology*, *51*(4), 335–349. https://doi.org/10.1016/j.ejop.2015.05.003

Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T., & Bonkowski, M. (2016). The soil food web revisited: Diverse and widespread mycophagous soil protists. *Soil Biology and Biochemistry*, *94*, 10–18. https://doi.org/10.1016/j.soilbio.2015.11.010

Geisen, S., Kudryavtsev, A., Bonkowski, M., & Smirnov, A. (2014). Discrepancy between Species Borders at Morphological and Molecular Levels in the Genus Cochliopodium (Amoebozoa, Himatismenida), with the Description of Cochliopodium plurinucleolum n. Sp. *Protist*, *165*(3), 364–383. https://doi.org/10.1016/j.protis.2014.04.002

Geisen, S., Mitchell, E. A. D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L. D., Jousset, A., Krashevska, V., Singer, D., Spiegel, F. W., Walochnik, J., & Lara, E. (2018). Soil protists: A fertile frontier in soil biology research. *FEMS Microbiology Reviews*, *42*(3), 293–323. https://doi.org/10.1093/femsre/fuy006

Gentekaki, E., & Lynn, D. H. (2009). High-level genetic diversity but no population structure inferred from nuclear and mitochondrial markers of the peritrichous ciliate Carchesium polypinum in the Grand River basin (North America). *Applied and Environmental Microbiology*, *75*(10), 3187–3195. https://doi.org/10.1128/AEM.00178-09

Gong, J., Dong, J., Liu, X., & Massana, R. (2013). Extremely High Copy Numbers and Polymorphisms of the rDNA Operon Estimated from Single Cell Analysis of Oligotrich and Peritrich Ciliates. *Protist*, *164*(3), 369–379. https://doi.org/10.1016/j.protis.2012.11.006

Guerra, C. A., Bardgett, R. D., Caon, L., Crowther, T. W., Delgado-Baquerizo, M., Montanarella, L., Navarro, L. M., Orgiazzi, A., Singh, B. K., Tedersoo, L., Vargas-Rojas, R., Briones, M. J. I., Buscot, F., Cameron, E. K., Cesarz, S., Chatzinotas, A., Cowan, D. A., Djukic, I., van den Hoogen, J., … Eisenhauer, N. (2021). Tracking, targeting, and conserving soil biodiversity. *Science*, *371*(6526), 239–241. https://doi.org/10.1126/science.abd7926

Gupta, N., & Verma, V. K. (2019). Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. In P. K. Arora (Ed.), *Microbial Technology for the Welfare of Society* (Vol. 17, pp. 313–341). Springer Singapore. https://doi.org/10.1007/978-981-13-8844-6_15

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLOS ONE*, *9*(2), e87624. https://doi.org/10.1371/journal.pone.0087624

Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., Bakker, J., & Helder, J. (2006). Phylum-Wide Analysis of SSU rDNA Reveals Deep Phylogenetic Relationships among Nematodes and Accelerated Evolution toward Crown Clades. *Molecular Biology and Evolution*, *23*(9), 1792–1800. https://doi.org/10.1093/molbev/msl044

Hugerth, L. W., Wefer, H. A., Lundin, S., Jakobsson, H. E., Lindberg, M., Rodin, S., Engstrand, L., & Andersson, A. F. (2014). DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Applied and Environmental Microbiology*, *80*(16), 5116–5123. https://doi.org/10.1128/AEM.01403-14

Ishaq, S. L., & Wright, A.-D. G. (2014). Design and Validation of Four New Primers for Next-Generation Sequencing To Target the 18S rRNA Genes of Gastrointestinal Ciliate Protozoa. *Applied and Environmental Microbiology*, *80*(17), 5515–5521. https://doi.org/10.1128/AEM.01644-14

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 239. https://doi.org/10.1186/s13059-016-1103-0

Koch, T. A., & Ekelund, F. (2005). Strains of the Heterotrophic Flagellate Bodo designis from Different Environments Vary Considerably with Respect to Salinity Preference and SSU rRNA Gene Composition. *Protist*, *156*(1), 97–112. https://doi.org/10.1016/j.protis.2004.12.001

Kosakyan, A., Gomaa, F., Mitchell, E. A. D., Heger, T. J., & Lara, E. (2013). Using DNA-barcoding for sorting out protist species complexes: A case study of the Nebela tincta–collaris–bohemica group (Amoebozoa; Arcellinida, Hyalospheniidae). *European Journal of Protistology*, *49*(2), 222–237. https://doi.org/10.1016/j.ejop.2012.08.006

Kosakyan, A., Heger, T. J., Leander, B. S., Todorov, M., Mitchell, E. A. D., & Lara, E. (2012). COI barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the Hyalospheniidae Schultze. *Protist*, *163*(3), 415–434. https://doi.org/10.1016/j.protis.2011.10.003

Kudryavtsev, A., Bernhard, D., Schlegel, M., E-Y Chao, E., & Cavalier-Smith, T. (2005). 18S Ribosomal RNA Gene Sequences of Cochliopodium (Himatismenida) and the Phylogeny of Amoebozoa. *Protist*, *156*(2), 215–224. https://doi.org/10.1016/j.protis.2005.03.003

Kudryavtsev, A., & Gladkikh, A. (2017). Two new species of Ripella (Amoebozoa, Vannellida) and unusual intragenomic variability in the SSU rRNA gene of this genus. *European Journal of Protistology*, *61*, 92–106. https://doi.org/10.1016/j.ejop.2017.09.003

Lara, E., Mitchell, E. A. D., Moreira, D., & López García, P. (2011). Highly Diverse and Seasonally Dynamic Protist Community in a Pristine Peat Bog. *Protist*, *162*(1), 14–32. https://doi.org/10.1016/j.protis.2010.05.003

Lara, E., Singer, D., & Geisen, S. (2022). Discrepancies between prokaryotes and eukaryotes need to be considered in soil DNA -based studies. *Environmental Microbiology*, *24*(9), 3829–3839. https://doi.org/10.1111/1462-2920.16019

Long, E. O., & Dawid, I. B. (1980). Repeated Genes in Eukaryotes. *Annual Review of Biochemistry*, *49*(1), 727–764. https://doi.org/10.1146/annurev.bi.49.070180.003455

Machida, R. J., & Knowlton, N. (2012). PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences. *PLoS ONE*, *7*(9), e46180. https://doi.org/10.1371/journal.pone.0046180

Maloney, J. G., Molokin, A., & Santin, M. (2020). Use of Oxford Nanopore MinION to generate full-length sequences of the Blastocystis small subunit (SSU) rRNA gene. *Parasites & Vectors*, *13*(1), 595. https://doi.org/10.1186/s13071-020-04484-6

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

Nassonova, E., Smirnov, A., Fahrni, J., & Pawlowski, J. (2010). Barcoding amoebae: Comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist*, *161*(1), 102–115. https://doi.org/10.1016/j.protis.2009.07.003

Needham, D. M., & Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature Microbiology*, *1*(4), 16005. https://doi.org/10.1038/nmicrobiol.2016.5

Oliverio, A. M., Geisen, S., Delgado-Baquerizo, M., Maestre, F. T., Turner, B. L., & Fierer, N. (2020). The global-scale distributions of soil protists and their contributions to belowground systems. *Science Advances*, *6*(4). Scopus. https://doi.org/10.1126/sciadv.aax8787

Orr, R. J. S., Zhao, S., Klaveness, D., Yabuki, A., Ikeda, K., Watanabe, M. M., & Shalchian-Tabrizi, K. (2018). Enigmatic Diphyllatea eukaryotes: Culturing and targeted PacBio RS amplicon sequencing reveals a higher order taxonomic diversity and global distribution. *BMC Evolutionary Biology*, *18*(1), 115. https://doi.org/10.1186/s12862-018-1224-z

Page, F. C. (1967). Taxonomic Criteria for Limax Amoebae, with Descriptions of 3 New Species of *Hartmannella* and 3 of *Vahlkampfia* *. *The Journal of Protozoology*, *14*(3), 499–521. https://doi.org/10.1111/j.1550-7408.1967.tb02036.x

Rodriguez-Zaragoza, S., Mayzlish, E., & Steinberger, Y. (2005). Vertical Distribution of the Free-Living Amoeba Population in Soil under Desert Shrubs in the Negev Desert, Israel. *Applied and Environmental Microbiology*, *71*(4), 2053–2060. https://doi.org/10.1128/AEM.71.4.2053-2060.2005

Scheckenbach, F., Wylezich, C., Mylnikov, A. P., Weitere, M., & Arndt, H. (2006). Molecular Comparisons of Freshwater and Marine Isolates of the Same Morphospecies of Heterotrophic Flagellates. *Applied and Environmental Microbiology*, *72*(10), 6638–6643. https://doi.org/10.1128/AEM.02547-05

Shmakova, L., Bondarenko, N., & Smirnov, A. (2016). Viable Species of Flamella (Amoebozoa: Variosea) Isolated from Ancient Arctic Permafrost Sediments. *Protist*, *167*(1), 13–30. https://doi.org/10.1016/j.protis.2015.11.001

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Trobajo, R., Mann, D. G., Clavero, E., Evans, K. M., Vanormelingen, P., & McGregor, R. C. (2010). The use of partial *cox* 1, *rbc* L and LSU rDNA sequences for phylogenetics and species identification within the *Nitzschia palea* species complex (Bacillariophyceae). *European Journal of Phycology*, *45*(4), 413–425. https://doi.org/10.1080/09670262.2010.498586

Van Wichelen, J., D'Hondt, S., Claeys, M., Vyverman, W., Berney, C., Bass, D., & Vanormelingen, P. (2016). A Hotspot of Amoebae Diversity: 8 New Naked Amoebae Associated with the Planktonic Bloom-forming Cyanobacterium Microcystis. *Acta Protozoologica*, *55*, 61–87. https://doi.org/10.4467/16890027AP.16.007.4942

Appendices

1. List of primer pairs used in this study.

Region	Fwd name	Fwd sequence	Fwd start	Fwd end	Rev name	Rev sequence	Rev start	Rev end	Length (bp)	Ref
V4	563F	GCCAGCAVCYGCGGTAAY	563	580	Euk895R	TCHNHGNATTTCACCNCT	x	x	x	
V4-V5	563F	GCCAGCAVCYGCGGTAAY	563	580	1132R	CCGTCAATTHCTTYAART	1132	1149	587	(Hugerth et al., 2014)
V4	563F	GCCAGCAVCYGCGGTAAY	563	580	926R	CCGYCAATTYMTTTRAGTTT	x	x	x	
V4	616*F	TTAAARVGYTCGTAGTYG	616	633	Euk895R	TCHNHGNATTTCACCNCT	x	x	x	(Guerra et al., 2021; Hugerth et al., 2014)
V4-V5	616*F	TTAAARVGYTCGTAGTYG	616	633	1132R	CCGTCAATTHCTTYAART	1132	1149	534	(Hugerth et al., 2014)
V4	616*F	TTAAARVGYTCGTAGTYG	616	633	926R	CCGYCAATTYMTTTRAGTTT	x	x	x	(Hugerth et al., 2014)
V4	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	Euk895R	TCHNHGNATTTCACCNCT	x	x	x	(Hugerth et al., 2014; Needham & Fuhrman, 2016)
V4-V5	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	1132R	CCGTCAATTHCTTYAART	1132	1149	589	(Hugerth et al., 2014; Needham & Fuhrman, 2016)

V4	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	926R	CCGYCAATTYMTTTRAGTTT	x	x	x	(Hugerth et al., 2014; Needham & Fuhrman, 2016)
V4	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	Euk895R	TCHNHGNATTTCACCNCT	x	x	310-330	(Guerra et al., 2021)
V4-V5	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	1132R	CCGTCAATTHCTTYAART	1132	1149	x	(Guerra et al., 2021; Hugerth et al., 2014)
V4	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	926R	CCGYCAATTYMTTTRAGTTT	x	x	x	(Guerra et al., 2021; Hugerth et al., 2014)
V4	574F	x	x	x	Euk895R	TCHNHGNATTTCACCNCT	x	x	x	(Guerra et al., 2021)
V4-V5	574F	x	x	x	1132R	CCGTCAATTHCTTYAART	1132	1149	x	(Hugerth et al., 2014)
V4	574F	x	x	x	926R	CCGYCAATTYMTTTRAGTTT	x	x	x	(Needham & Fuhrman, 2016)
V4	574*F	x	x	x	Euk895R	TCHNHGNATTTCACCNCT	x	x	x	(Guerra et al., 2021)
V4-V5	574*F	x	x	x	1132R	CCGTCAATTHCTTYAART	1132	1149	x	(Hugerth et al., 2014)
V4	574*F	x	x	x	926R	CCGYCAATTYMTTTRAGTTT	x	x	x	(Needham & Fuhrman, 2016)
V9	1389F	TTGTACACACCGCCC	1626	1640	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	(Amaral-Zettler et al., 2009; Medlin et al., 1988)
V9	1389F	TTGTACACACCGCCC	1626	1640	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	(Amaral-Zettler et al., 2009)
V4-V9	563F	GCCAGCAVCYGCGGTAAY	563	580	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	(Amaral-Zettler et al., 2009; Hugerth et al., 2014)

V4-V9	563F	GCCAGCAVCYGCGGTAAY	563	580	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	(Hugerth et al., 2014; Medlin et al., 1988)
V9-28S	563F	GCCAGCAVCYGCGGTAAY	563	580	RM3R 1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	(Hugerth et al., 2014; Machida & Knowlton, 2012)
V9-28S	563F	GCCAGCAVCYGCGGTAAY	563	580	RM3R 925	TTCGATTRGTCTTTCGCCCCT	x	x	x	(Hugerth et al., 2014; Machida & Knowlton, 2012)
V9-D2(LSU)	563F	GCCAGCAVCYGCGGTAAY	563	580	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	(Hugerth et al., 2014)
V4-V5	616*F	TTAAARVGYTCGTAGTYG	616	633	1132R	CCGTCAATTHCTTYAART	1132	1149	534	(Hugerth et al., 2014)
V4-V9	616*F	TTAAARVGYTCGTAGTYG	616	633	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	(Hugerth et al., 2014)
V4-V9	616*F	TTAAARVGYTCGTAGTYG	616	633	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	(Hugerth et al., 2014; Medlin et al., 1988)
V4-28S	616*F	TTAAARVGYTCGTAGTYG	616	633	RM3R 1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	(Hugerth et al., 2014; Machida & Knowlton, 2012)
V4-28S	616*F	TTAAARVGYTCGTAGTYG	616	633	RM3R 925	TTCGATTRGTCTTTCGCCCCT	x	x	x	(Hugerth et al., 2014; Machida & Knowlton, 2012)
V4-D2(LSU)	616*F	TTAAARVGYTCGTAGTYG	616	633	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	(Hugerth et al., 2014)

V4-V9	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	(Amaral-Zettler et al., 2009; Needham & Fuhrman, 2016)
V4-V9	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	(Medlin et al., 1988; Needham & Fuhrman, 2016)
V4-28S	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	RM3R1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	(Machida & Knowlton, 2012; Needham & Fuhrman, 2016)
V4-28S	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	RM3R925	TTCGATTRGTCTTTCGCCCCT	x	x	x	(Machida & Knowlton, 2012; Needham & Fuhrman, 2016)
V4-D2(LSU)	515Funi	GTGCCAGCMGCCGCGGTAA	561	579	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	(Needham & Fuhrman, 2016)
V4-V9	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	(Amaral-Zettler et al., 2009; Guerra et al., 2021)
V4-V9	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	(Guerra et al., 2021; Medlin et al., 1988)
V4-28S	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	RM3R1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	(Guerra et al., 2021; Machida & Knowlton, 2012)
V4-28S	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	RM3R925	TTCGATTRGTCTTTCGCCCCT	x	x	x	(Guerra et al., 2021; Machida & Knowlton, 2012)
V4-D2(LSU)	Euk575F	ASCYGYGGTAAYWCCAGC	x	x	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	(Guerra et al., 2021)
V4-V9	574F	x	x	x	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	(Amaral-Zettler et al., 2009)

V4-V9	574F	x	x	x	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	(Guerra et al., 2021)
V4-28S	574F	x	x	x	RM3R1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	(Machida & Knowlton, 2012)
V4-28S	574F	x	x	x	RM3R925	TTCGATTRGTCTTTCGCCCCT	x	x	x	(Machida & Knowlton, 2012)
V4-D2(LSU)	574F	x	x	x	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	x
V4-V9	574*F	x	x	x	1510R	CCTTCYGCAGGTTCACCTAC	x	x	x	x
V4-V9	574*F	x	x	x	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	x	x
V4-28S	574*F	x	x	x	RM3R1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	x
V4-28S	574*F	x	x	x	RM3R925	TTCGATTRGTCTTTCGCCCCT	x	x	x	x
V4-D2(LSU)	574*F	x	x	x	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	x
V9-28S	1389F	TTGTACACACCGCCC	1626	1640	RM3R1141	CRCCAGTTCTGCTTACCAAAA	x	x	x	(Amaral-Zettler et al., 2009; Machida & Knowlton, 2012)
V9-28S	1389F	TTGTACACACCGCCC	1626	1640	RM3R925	TTCGATTRGTCTTTCGCCCCT	x	x	x	(Amaral-Zettler et al., 2009; Machida & Knowlton, 2012)

V9-D2(LSU)	1389F	TTGTACACACCGCCC	1626	1640	TWI3R	GGTCCGTGTTTCAAGACG	x	x	x	(Amaral-Zettler et al., 2009)
COI	LCO	GGTCAACAAATCATAAAGATATTGG	x	x	HCO	TAAACTTCAGGGTGACCAAAAAATCA	x	x	710	(Folmer et al., 1994)
18S rDNA	EukA	AACCTGGTTGATCCTGCCAGT	x	x	EukB	TGATCCTTCTGCAGGTTCACCTAC	x	x	1700	(Medlin et al., 1988)

2. Preparation and Composition

To create NMAS, combine 5 ml of each stock solution 1 and 2 and fill up with 1 L of demineralized water. Furthermore, solution is autoclaved for 90 minutes at 121°C. While, NB-NMAS media is done by adding nutrient broth (NB) mix to 1L of NMAS.

Table 5. Composition of stock solution 1

Component	Quantity
NaCl	12 g
$MgSO_4$	0.4 g
$CaCl_2$	0.6 g
Distilled water (MQ)	500 ml

Table 6. Composition of stock solution 2

Component	Quantity
Na_2HPO_4	14.2 g
KH_2PO_4	13.6 g
Distilled water (MQ)	500 ml

Table 7. Composition of nutrient broth

Component	Quantity
Bacterial peptone	0.32 g
Beef extract	0.32 g
NaCl	0.16 g