Diversity Patterns of Protists are Highly Affected by Methods Disentangling Inter-specific Variants: A Case Study in Oligotrich (s.l.) Ciliates

Jiahui Xu
  Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, Guangdong Provincial Key Laboratory for Healthy and Safe Aquaculture

Jianlin Han
  Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, Guangdong Provincial Key Laboratory for Healthy and Safe Aquaculture

Hua Su
  Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, Guangdong Provincial Key Laboratory for Healthy and Safe Aquaculture

Changyu Zhu
  Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, Guangdong Provincial Key Laboratory for Healthy and Safe Aquaculture

Zijing Quan
  Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, Guangdong Provincial Key Laboratory for Healthy and Safe Aquaculture

Zhenzhen Yi (✉ zyi@scnu.edu.cn)
  South China Normal University College of Life Sciences   https://orcid.org/0000-0002-0693-9989

Lei Wu
  Guangzhou Key Laboratory of Subtropical Biodiversity and Biomonitoring, Guangdong Provincial Key Laboratory for Healthy and Safe Aquaculture

Research Article

Abstract

An enormous amount of environmental sequences produced by high-throughput sequencing (HTS) is popular for inferring diversity and distribution patterns of protists, which are widely distributed and playing important roles in energy flux and nutrient cycling. However, study testing whether methods disentangling inter-specific variants affect the diversity and distribution patterns of protists using field samples is insufficient. Using oligotrich (s.l.) ciliates, one group of abundant and dominate planktonic protists, in field samples as an example, the present study indicates that DADA2 performs better than UCLUST and UPARSE for inferring diversity patterns of oligotrich (s.l.) ciliates in Pearl River Estuary and surrounding regions. UPARSE, as an OTU-construction method might underestimate species richness and produce less reliable beta diversity pattern than DADA2. UCLUST with 97% and 99% clustering thresholds overestimate species richness, and the beta diversity pattern inferred by the former one is unreasonable. Salinity is shown to be one of the key factors responsible for variations in community distribution of ciliates, but infrequent marine-freshwater transitions occurred during evolutionary terms of this group.

Introduction

Protists, single-celled eukaryotes, are widely distributed in soil, marine and freshwater worldwide and play key roles in energy flux as well as trophic interactions and nutrient cycling [1−3]. However, protists with low abundance in a region are easily to be ignored in observation in vivo due to small size and difficulty in identification and culture [4, 5]. Foissner [6] assumed that probably more than a half of the diversity of many protist groups has not been revealed. Consequently, insufficient sampling dense of protists in studies using isolated cells might halt revealing their accurate patterns of diversity, phylogeny, transition and so on. Fortunately, a great number of sequences produced by high-throughput sequencing (HTS) without limitations of culturing and ignoring in observation that provide us a chance to investigate more accurate diversity and distribution patterns of protists [7−11].

Considering that both real biological variants and spurious sequences introduced by sequencing are included in the HTS datasets, method disentangling biological variation is one main factor affecting the accuracy of downstream analyses [12−15]. Nowadays, two sequence grouping approaches are popular for disentangling biological variations [12]. One is the construction of operational taxonomic units (OTUs) with quality filtering, e.g., UPARSE [14], UCLUST [13], MED [16], and different clustering thresholds of sequence have been reported to define OTUs of protists (e.g., 95% [17]; 97% [18]; 98% [19]; 99% [20]). Another is correcting amplicon errors by generating Amplicon Sequence Variants (ASVs) at single nucleotide resolution, e.g., DADA2 [21]. To our knowledge, a huge amount of references focusing on protist diversity based on HTS data are published every year, but only few studies comparing effects of different HTS data processing methods on analyses [15, 22, 23]. Some studies suggested that ASVs could more accurately reproduce a known alpha diversity than OTUs [23], USEARCH and QIIME strongly affected the number of predicted OTUs but not biogeography patterns [22]. Additionally, the numbers of OTUs varied with change of clustering thresholds [24]. Consequently, protists diversity inferred from HTS data are highly affected not only by sequence grouping approaches but also softwares and clustering

thresholds producing OTUs. However, some questions are still unclear. For instance, testing whether ASVs and OTUs affect beta diversity patterns of protists is lack in real case scenarios using field samples [23].

In the present study, we will explore whether methods disentangling biological variants will affect the diversity patterns of protists by using oligotrich (s.l.) ciliates, a major group of the abundant and dominate planktonic protist communities [1, 25–29], in field samples as an example. Our aim is to provide suggestions for data analyses in future studies. What's more, we investigate the community distribution and transition patterns of oligotrich (s.l.) ciliates using both released sequences of identified species in GenBank and environmental sequences from the Pearl River Estuary and surrounding regions, which includes various salinity gradients.

Materials And Methods

Environmental Sequences of Oligotrich (s.l.) Ciliates

Oligotrich (s.l.) ciliates comprise two subclasses, Oligotrichia and Choreotrichia [30]. Environmental V4 region of SSU rDNA sequences were downloaded from GenBank under the accession number PRJNA646537, the geographic locations of sampling sites and the salinity of each site are shown in our previous study (Figure 1, Table 1 in [11]). In summary, PRE1–PRE11 were sampling sites along the flow direction of Pearl River Estuary (PRE), sampling sites DY1–DY5, GZ1–GZ5, SZ1–SZ6, ZH1–ZH5 and ZJ1–ZJ6 were in sampling area Daya Bay (DY), Guangzhou (GZ), Shenzhen (SZ), Zhuhai (ZH) and Zhanjiang (ZJ), respectively. According to the Venice Salinity Classification System (Source: Limnology and Oceanography, 1958), ecological status was classified as freshwater (< 0.5 PSU), low-brackish (0.5–18 PSU), high-brackish (18–30 PSU) and marine (30–40 PSU).

Table 1
The number for OTUs/ASVs and reads of total sequences and oligotrichous sequences from four datasets

	UPARSE-97E	UCLUST-97E	UCLUST-99E	DADA2-100E
Number of total OTUs/ASVs	2991	38165	159501	3890
Reads number of total OTUs/ASVs	978977	1161520	1161520	874508
Number of oligotrichous OTUs/ASVs	63	521	2154	103
Reads number of oligotrichous OTUs/ASVs	18261	20348	20286	16063

To test whether the methods of defining species affect diversity patterns, all downloaded data were processed using three highly frequently used methods: UPARSE in USEARCH v9.2.64 [14], UCLUST in QIIME v1.9.1 [13], and DADA2 in QIIME2 v2019.10 [23], respectively. In UPARSE algorithm, OTUs could

only be clustered at 97% similarity. And 97% and 99% similarity thresholds were set for using UCLUST algorithm, considering that different similarity thresholds could be set in QIIME for outlining OTUs. DADA2 produced ASVs with one nucleotide difference, which was equivalent to constructing OTUs at 100% similarity threshold. Then the Silva 138 SSURef database (https://doi.org/10.5281/ZENODO.3891931) was utilized to annotate taxonomic information for each representative sequence. After that, only sequences annotated as oligotrich (s.l.) ciliates (similarity > 90%) were selected for downstream analyses. DY3, DY4, PRE11, SZ5, SZ6 and ZJ5 were deleted from subsequent analyses, because less than five oligotrich (s.l.) sequences detected in these sampling sites. Totally, four oligotrich (s.l.) datasets, viz., UPARSE-97E, UCLUST-97E, UCLUST-99E and DADA2-100E (the number of OTUs/ASVs and reads number of OTUs/ASVs are shown in Table 1), were produced from environmental sequences. The geographic locations of sampling sites and the salinity of each sites in this study are shown in Fig. 1.

In order to explore the diversity patterns of oligotrich (s.l.) ciliates among different datasets, Bray-Curtis distance matrix were used for quantifying the community dissimilarity. Hierarchical Clustering and Principal Co-ordinates Analysis (PCoA) were performed in the R platform [31] using the package 'ape' [32] and 'vegan' [33]. Hierarchical clustering were performed using the "ward.D2" algorithm of the hclust function, since the OTUs/ASVs numbers of different analyses methods were quite different, logarithmic calculations were performed on the data before performing the distance matrix calculation. In the PCoA, using "stat_ellipse (level = 0.95)" command to add ellipses by group except DY (too few points to calculate an ellipse), the differences between predefined groups based on sampling areas were statistically tested by permutational multivariate analysis of variance (ADONIS) [34] using 1,000 permutations.

Sequence Alignment and Phylogenetic Analyses

All available SSU rDNA sequences (>1,000 bp) of identified oligotrich (s.l.) species as of August 2021, as well as 6 sequences of the subclasses Hypotrichia used for outgroup, were downloaded from GenBank (accession numbers in Table S1). The phylogenetic dataset DATA2-100P contained identified sequences in Table S1 and representative sequences from dataset DATA2-100E were produced and aligned using MAFFT [35]. The resulting alignment was manually checked in SeaView v4 [36] for trimming two ends and ambiguous sites.

Maximum likelihood (ML) analyses were carried out using RAxML-HPC2 on XSEDE on CIPRES Science Gateway (http://www.phylo.org/sub_sections/portal) [37] using the GTRGAMMA model, and support for the best-scoring ML tree was assessed by 1,000 bootstrap replicates. Bayesian inference (BI) analysis was also performed on CIPRES Science Gateway using MrBayes on XSEDE v3.2.7a, using the GTR+I+G model, which was selected by jModeltest v2.1.10 [38]. Markov chain Monte Carlo simulations were run for 10,000,000 generations with four chains. Trees were sampled every 100 generations and the first 25% of trees were discarded as burn-in. The 50% majority rule consensus tree was used to calculate the posterior probabilities (PP) for each node. Trees were viewed and edited with FigTree v. 1.4.4 [39].

Finally, Mesquite v3.6 [40] was used to infer the most parsimonious pattern of marine-freshwater transitions of oligotrich (s.l.) ciliates using the ML tree above.

Results

Number and Proportion of Oligotrich (s.l.) OTUs/ASVs in Sampling Areas/Sites Varies among Methods

Since the total number of oligotrich (s.l.) OTUs/ASVs contained in each dataset varies heavily from 63 (UPARSE-97E) to 2154 (UCLUST-99E) (Table 1), the ratio of oligotrich (s.l.) OTUs/ASVs number to total ones within each dataset is also compared. The proportion of oligotrich (s.l.) OTUs/ASVs in each sampling area/site in UPARSE-97E is much higher than those in the other three datasets (Table 2 and Table S2). Among six sampling areas, the largest proportion of oligotrich (s.l.) OTUs/ASVs is detected in PRE among all four datasets, with about 70% in UPARSE-97E and 50% in the other three datasets. The lowest proportion of oligotrich (s.l.) OTUs/ASVs is detected in DY from UCLUST-97E (4.03%), UCLUST-99E (1.16%) and DADA2-100E (5.83%) (Table 2). By contrast, the lowest one is detected in SZ (14.29%) from UPARSE-97E, which is about twice that in UCLUST-99E (6.04%) and DADA2-100E (7.77%) (Table 2). Similarly, for each sampling site, the highest proportion of oligotrich (s.l.) OTUs/ASVs is detected in UPARSE-97E among four datasets (Table S2). And more than or nearly half oligotrich (s.l.) OTUs/ASVs are only present in one sampling site in UCLUST-97E (65.11%), UCLUST-99E (84.11%) and DADA2-100E (48.57%), while only 12.31% in UPARSE-97E (Table S3).

Table 2

Number and proportion of oligotrichous OTUs/ASVs in each sampling area from the four datasets

Sampling areas	UPARSE-97E	UCLUST-97E	UCLUST-99E	DADA2-100E
DY	14 (22.22%)	21 (4.03%)	25 (1.16%)	6 (5.83%)
GZ	32 (50.79%)	146 (28.02%)	421 (19.55%)	31 (30.10%)
PRE	45 (71.43%)	275 (52.78%)	834 (38.72%)	47 (45.63%)
SZ	9 (14.29%)	56 (10.75%)	130 (6.04%)	8 (7.77%)
ZH	28 (44.44%)	50 (9.60%)	123 (5.71%)	23 (22.33%)
ZJ	31 (49.21%)	154 (29.56%)	797 (37.00%)	32 (31.07%)

Beta Diversity Patterns of Oligotrich (s.l.) Ciliates

The hierarchical clustering analyses shows that oligotrich (s.l.) samples group into two major clades in all four datasets (Fig. 2). Generally, Clade A consists of most samples from sampling areas ZH and GZ, as well as samples of PRE1-PRE4. Clade B contains most samples from sampling areas ZJ, SZ and DY, as well as samples of PRE5-PRE10. Monophyletic groups are not revealed for samples collected from

sampling aeras ZH, DY and ZJ, respectively (Fig. 2). Differently from the general clustering pattern in other three datasets, samples of GZ1- GZ5 fall into Clade B instead of Clade A, and sample PRE5 falls into Clade A instead of Clade B in UPARSE-97E (Fig. 2b). Additionally, sample SZ1 falls into Clade A instead of Clade B in DADA2-100E, UPARSE-97E and UCLUST-99E (Fig. 2a, b and d), while it falls into Clade A in UCLUST-97E (Fig. 2c).

PCoA results show that oligotrich (s.l.) samples basically group depending on six sampling areas (DY, GZ, PRE, SZ, ZH, ZJ), and PRE is divided into two groups (PRE1-PRE4, PRE5-PRE10). ADONIS results support significant ($P < 0.001$) differentiations of community structures among these seven groups based on all four datasets (Fig. 3). Consistent with the hierarchical clustering results (Fig. 2), samples of PRE1-PRE4 are clearly separated from those of PRE5-PRE10, ZJ, SZ and DY in all four datasets (Fig. 3). However, for DADA2-100E, samples of PRE1-PRE4, ZH and GZ tend to group together and are separated from other sampling sites (Fig. 3a).

Phylogeny and Transition Patterns of Oligotrich (s.l.) Ciliates

To investigate the ecological transition patterns of oligotrich (s.l.) ciliates, the phylogenetic tree including 197 sequences from identified species (Table S1) and 103 ASVs from dataset DADA2-100E are constructed (Fig. 4). Monophyly of the subclass Choreotrichia is supported by high supports (98% ML, 1.00 BI), but that of subclass Oligotrichia is not. In the phylogenetic trees, crown clades usually contain identified oligotrich (s.l.) species and ASVs from various habits. Though most ASVs group with identified species, some ASVs form isolated clades (Clade 1, Clade 2 in Fig. 4) without identified species. This indicates that oligotrichous diversity in low-brackish habitats might be underestimated in previous morphological investigations. Transition pattern analyses show that there is a high-brackish ancestor for oligotrich (s.l.) ciliates (Fig. 4). The subclass Choreotrichia appears to be evolved from the high-brackish ancestor with high support values (98% ML, 1.00 BI), then some transitions to freshwater, low-brackish and marine, and even back to high-brackish for some species (e.g., *Tintinnopsis radix* and *Leprotintinnus nordqvisti*) (Fig. 4). By contrast, the subclasses Oligotrichia seems to be derived from an equivocal ancestor with poor support values (ML < 50%, BI < 0.50), and some transitions to freshwater, low-brackish, high-brackish and marine, and possible additional transitions back to low-brackish for some species (e.g., *Parallelostrombidium paraellipticum* and *P. obesum*) (Fig. 4).

Discussion

Methods Disentangling Biological Variants Highly Affect Diversity Patterns of Oligotrich (s.l.) Ciliates

Among four datasets generated by different algorithms and thresholds (viz. DADA2-100E, UPARSE-97E, UCLUST-97E, UCLUST-99E), the actual number of oligotrich (s.l.) OTU/ASV as well as ratio of the oligotrich (s.l.) OTUs/ASVs number to the total OTUs/ASVs number in each sample area/site are different (Table 1 and 2, Table S2 and S3). This is consistent with previous investigations that the alpha diversity

of protists is highly variable depending on sequence grouping approaches, as well as softwares and clustering thresholds producing OTUs [22–24, 41]. Recent investigations revealed that compared to OTUs, ASVs could more accurately reproduce a known alpha diversity from mock communities of various groups, e.g., bacteria [12, 42, 43], fungi [42], ciliated protists [23]. In this study, the numbers of oligotrich (s.l.) OTUs/ASVs inferred by UPARSE-97E (63) and DADA2-100E (103) are more reliable than those inferred by UCLUST-97E (521) and UCLUST-99E (2154) (Table 1). That is because only 592 morphological ciliate species were reported in South China Sea from 1991 to 2018 [44]. Surprisingly, the proportion of oligotrich (s.l.) OTUs/ASVs in each sampling area/site (Table 2 and Table S2) as well as shared OTUs/ASVs in two or more sampling sites (Table S3) is much higher in UPARSE-97E than in the other three datasets. One possible explanation is that UPARSE, as an OTU-construction method with the least false-positive results [14], might underestimate oligotrich (s.l.) species richness. Therefore, our study using field samples draws the same conclusion as studies using mock communities that the estimate of alpha diversity reproduced by ASVs in DADA2-100E might be better than OTUs.

In this study, a general beta diversity pattern of oligotrich (s.l.) ciliates is revealed in four datasets, but methods disentangling inter-specific variants also show some impact on inferring beta diversity patterns (Figs. 2 and 3). Clearly, community variations are observed between PRE1-PRE4 and PRE5-PRE10 in DADA2-100E, UCLUST-97E and UCLUST-100E, and PRE5 groups with PRE1-PRE4 instead of PRE6-PRE10 in UPARSE-97E. This might be explained by sharp increase of salinity between sampling sites PRE1-PRE4 (0.3-0.9 PSU) and PRE5-PRE10 (4.7-12.2 PSU). It is possible that the mixture of freshwater and seawater form a low-salinity front between sampling sites PRE4 and PRE5 [45]. Numerous studies have proved that salinity appears to be the factor that correlates best with distributions of phytoplankton and bacterioplankton in estuaries (e.g., [46–49]), and hence community compositions of oligotrich (s.l.) ciliates with phytoplankton and bacterioplankton as food also change heavily between PRE1-PRE4 and PRE5-PRE10. Sampling areas ZJ and DY are located in nearshore of South China Sea (surrounding regions of Pearl River Estuary), and most samples from these sampling areas cluster with PRE5-PRE10 in all four datasets (Figs. 2 and 3). Possibly, salinity also play an important role in this cluster pattern, since salinity value in sampling areas ZJ, DY and sampling sites PRE5-PRE10 are generally higher than other ones (Fig. 1). Similar to PRE1-PRE4, sampling sites GZ1-GZ5 are in upper estuary of the Pearl River Estuary, and are highly influenced by large freshwater discharge from Pearl River. Theoretically, community structures of oligotrich (s.l.) ciliates in upper Pearl River Estuary with lower salinity and higher nutrient content should be much different from those in lower Pearl River Estuary. In present study, GZ samples group with PRE1-PRE4 in DADA2-100E, UCLUST-97E and UCLUST-99E, but not in UPARSE-97E (Figs. 2 and 3). It seems that beta diversity pattern of GZ samples are more reasonable in the former three datasets than later ones. Hierarchical clustering analyses show that most ZH samples group with PRE1-PRE4 in Clade A (Fig. 2), and most SZ samples fall into the Clade B including PRE5-PRE10 (Fig. 2a, b and d), though sampling areas ZH and SZ are located at the similar latitude of PRE5. Different community structures of oligotrich (s.l.) ciliates in ZH and SZ samples might be due to two reasons. Firstly, the surface flow velocity on the west side (ZH) is usually greater than the east side (SZ) of Pearl River Estuary [50]. This indicates that community structures of oligotrich (s.l.) ciliates in ZH might be more highly

influenced by river runoff than in SZ. Thus, cluster pattern of SZ and ZH samples are less reliable in UCLUST-97E (Fig. 2c) than in other three datasets (Fig. 2a, b and d). Secondly, both sampling areas ZH and SZ are located in nearshore of Pearl River Estuary, where environmental factors and community structures of ciliates are highly influenced by human activities [44, 51, 52]. All these indicate that beta diversity patterns of DADA2-100E and UCLUST-99E are more reliable than those of UPARSE-97E and UCLUST-97E.

In conclusion, among four datasets compared in this study, DADA2-100E performs best for inferring diversity patterns of oligotrich (s.l.) ciliates in Pearl River Estuary and surrounding regions.

Community Distribution and Ecological Transitions of Oligotrich (s.l.) Ciliates in Environments with Various Salinity Gradients

As revealed in various groups of archaea, bacteria and protist (e.g., [46, 47, 49, 53, 54]), the present study also revealed that salinity gradients play a vital role in structuring patterns for community distribution of oligotrich (s.l.) ciliates in Pearl River Estuary and surrounding regions where encompass an entire freshwater-marine salinity gradient. Firstly, communities of oligotrich (s.l.) ciliates in sampling sites PRE1−PRE10 along the flow direction of Pearl River Estuary are divided into two distinct groups PRE1−PRE4 (Clade A) and PRE5-PRE10 (Clade B) (Fig. 2a), which is consistent with the sharp change of salinity gradients between sampling sites PRE1−PRE4 (0.3-0.9 PSU) and PRE5-PRE10 (4.7-12.2 PSU). Additionally, all samples from high-brackish and marine habits fall into Clade B, though these samples were collected from two sampling areas (ZJ and DY) with around 500 km distance. It is believed that both physicochemical barrier of salinity gradients and the presence of locally adapted taxa limit the colonization success of microbes in different habits with salinity gradients [55]. Within Clade B, several samples (DY2, ZJ1, PRE5-PRE10, SZ2-SZ4) are from freshwater and low-brackish habits, which indicates that except for salinity, other environmental factors and geographical distance also shape community distribution of oligotrich (s.l.) ciliates. This is also revealed in various prokaryotic and eukaryotic microbial groups in estuaries (e.g., [44, 54, 56]).

As above mentioned, salinity gradients have been proven to be important physicochemical factor in structuring community distribution and limiting transitions of microbes including protists. Ancestors of different ciliate groups seem originate in various habits. Oligotrich (s.l.) and hypotrich (present study, [57]) ciliates are ancestrally high-brackish, and peritrich and colpod ciliates have freshwater/terrestrial ancestors [58, 59]. However, species in various habitats are reported for each of these ciliate groups. That is because infrequent marine-freshwater transitions always occurred during their evolutionary terms (present study, [57−60]), due to the ability of microbes to rapidly adapt and highly colonize to new environments [55, 61]. In our phylogenetic trees (Fig. 4), some ASVs formed separated clades (Clade 1, Clade2) and might be new taxa. It is consistent with previous report that the oligotrichous morphospecies diversity is underestimated [28]. Additionally, these ASVs representing new taxa are from low-brackish

habits (Fig. 4). By contrast, among all SSU rDNA sequences of identified oligotrich (s.l.) species deposited into GenBank, only 4% (8 out of 191) and 8% (15 out of 191) ones are from freshwater and low-brackish habits (Table S1). In future, more detailed transition and evolutionary patterns of oligotrich (s.l.) ciliates are expected be outlined, with the broader sampling, especially from freshwater and low-brackish habits.

Declarations

Author Contributions

ZY conceived the study. JX and JH performed the data analyses. JX, JH, HS and CZ prepared the figures. ZY, JX, JH and HS primarily wrote this manuscript. CZ, ZQ and LW participated in discussion and revised the manuscript.

Consent for publication

The authors all contributed to this work and agree to publish the results.

Conflict of Interest

The authors declare they had no relevant financial or non-financial interests to disclose.

References

1. Pierce RW, Turner JT (1992) Ecology of planktonic ciliates in marine food webs. Rev Aquat Sci 6(2):139–181.
2. Fenchel T (2008) The microbial loop-25 years later. J Exp Mar Biol Ecol 366(1–2):99–103. https://doi.org/10.1016/j.jembe.2008.07.013
3. Worden AZ, Follows MJ, Giovannoni SJ et al (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science 347(6223):1257594. https://doi.org/10.1126/science.1257594
4. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59(1):143–169.

https://doi.org/10.1128/MMBR.59.1.143-169.1995

5. Eland LE, Davenport R, Mota CR (2012) Evaluation of DNA extraction methods for freshwater eukaryotic microalgae. Water Res 46(16):5355–5364. https://doi.org/10.1016/j.watres.2012.07.023

6. Foissner W (2006) Biogeography and dispersal of micro-organisms: a review emphasizing protists. Acta Protozool 45(2):111–136. https://doi.org/10.1021/jp805737a

7. Doherty M, Tamura M, Vriezen JA et al (2010) Diversity of oligotrichia and choreotrichia ciliates in coastal marine sediments and in overlying plankton. Appl Environ Microbiol 76(12):3924–3935. https://doi.org/10.1128/AEM.01604-09

8. Li G, Su L, Zhang Q et al (2018) Molecular diversity and biogeography of benthic ciliates in the Bohai Sea and Yellow Sea. Acta Oceanol Sin 38(2):78–86. https://doi.org/10.1007/s13131-018-1236-y

9. Massana R, Gobet A, Audic S et al (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. Environ Microbiol 17(10):4035–4049. https://doi.org/10.1111/1462-2920.12955

10. Zinger L, Boetius A, Ramette A (2014) Bacterial taxa-area and distance-decay relationships in marine environments. Mol Ecol 23(4):954–964. https://doi.org/10.1111/mec.12640

11. Zhu C, Bass D, Wang Y et al (2020) Environmental parameters and substrate type drive microeukaryotic community structure during short-term experimental colonization in subtropical eutrophic freshwaters. Front Microbiol 11:555795. https://doi.org/10.3389/fmicb.2020.555795

12. Callahan BJ, McMurdie PJ, Rosen MJ et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13(7):581–583. https://doi.org/10.1038/nmeth.3869

13. Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336. https://doi.org/10.1038/nmeth.f.303

14. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10(10):996–998. https://doi.org/10.1038/nmeth.2604

15. Forster D, Dunthorn M, Stoeck T et al (2016) Comparison of three clustering approaches for detecting novel environmental microbial diversity. PeerJ 4:e1692. https://doi.org/10.7717/peerj.1692

16. Eren AM, Morrison HG, Lescault PJ et al (2015) Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J 9(4):968–979. https://doi.org/10.1038/ismej.2014.195

17. Countway PD, Gast RJ, Dennett MR et al (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500m) of the western North Atlantic (Sargasso Sea and Gulf Stream). Environ Microbiol 9(5):1219–1232. https://doi.org/10.1111/j.1462-2920.2007.01243.x

18. Wu W, Liu H (2018) Disentangling protist communities identified from DNA and RNA surveys in the Pearl River–South China Sea Continuum during the wet and dry seasons. Mol Ecol 27(22):4627–4640. https://doi.org/10.1111/mec.14867

19. Nebel M, Pfabel C, Stock A et al (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. Environ Microbiol Rep 3(2):154–

158. https://doi.org/10.7717/peerj.5364

20. Doherty M, Costas BA, McManus GB et al (2007) Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. Aquat Microb Ecol 48(2):141–154. https://doi.org/10.3354/ame048141

21. Bolyen E, Rideout JR, Dillon MR et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37(8):852–857. https://doi.org/10.1038/s41587-019-0209-9

22. Boscaro V, Rossi A, Vannini C et al (2016) Strengths and biases of high-throughput sequencing data in the characterization of freshwater Ciliate microbiomes. Microb Ecol 73(4):865–875. https://doi.org/10.1007/s00248-016-0912-8

23. Forster D, Lentendu G, Filker S (2019) Improving eDNA-based protist diversity assessments using networks of amplicon sequence variants. Environ Microbiol 21(11):4109–4124. https://doi.org/10.1111/1462-2920.14764

24. Logares R, Mangot JF, Massana R (2015) Rarity in aquatic microbes: placing protists on the map. Res Microbiol 166(10):831–841. https://doi.org/10.1016/j.resmic.2015.09.009

25. Dale T, Einar D (1987) Mass occurrence of planktonic oligotrichous ciliates in a bay in southern Norway. J Plankton Res 9(5):871–879. https://doi.org/10.1093/plankt/9.5.871

26. Edwards ES, Burkill PH (1995) Abundance, biomass and distribution of microzooplankton in the Irish Sea. J Plankton Res 17(4):771–782. https://doi.org/10.1093/plankt/17.4.771

27. Huang H, Yang J, Huang S et al (2021) Spatial distribution of planktonic ciliates in the western Pacific Ocean: along the transect from Shenzhen (China) to Pohnpei (Micronesia). Mari Life Sci Technol 3(1):103–115. https://doi.org/10.1007/s42995-020-00075-7

28. Lu K, Liu W, Warren A et al (2019) Diversity of oligotrich ciliates (Ciliophora, Spirotrichea) in the northern coast of South China Sea as revealed in LSU rDNA sequences. J Oceanol Limnol 38(1):158–169. https://doi.org/10.1007/s00343-019-9021-0

29. Tanaka T, Rassoulzadegan F (2002) Full-depth profile (0–2000 m) of bacteria, heterotrophic nanoflagellates and ciliates in the NW Mediterranean Sea: Vertical partitioning of microbial trophic structure. Deep Sea Res Part II Top Stud Oceanogr 49(11):2093–2107. https://doi.org/10.1016/S0967-0645(02)00029-2

30. Gao F, Warren A, Zhang Q et al (2016) The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). Sci Rep 6: 24874. https://doi.org/10.1038/srep24874

31. R Core Team (2018) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. https://www.R-project.org/

32. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290. https://doi.org/10.1093/bioinformatics/btg412

33. Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14(6):927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x

34. Clarke KR, Somerfield PJ, Gorley RN (2008) Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. J Exp Mar Biol Ecol 366(1−2):56−69. https://doi.org/10.1016/j.jembe.2008.07.009

35. Penn O, Privman E, Ashkenazy H et al (2010) GUIDANCE: a web server for assessing alignment confidence scores. Nucleic Acids Res 38(2):W23−W28. https://doi.org/10.1093/nar/gkq443

36. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27(2):221−224. https://doi.org/10.1093/molbev/msp259

37. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES science gateway for inference of large phylogenetic trees. Gatew Comput Environ Work, New Orleans, LA,14 Nov 2010, pp 1−8. https://doi.org/10.1109/GCE.2010.5676129

38. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253−1256. https://doi.org/10.1093/molbev/msn083

39. Rambaut A (2012) FigTree v1.4.2, available from: http://tree.bio.ed.ac.uk/software/figtree/.

40. Maddison WP, Maddison DR (2018) Mesquite: a modular system for evolutionary analysis. Version 3.6. available from: http://www.mesquiteproject.org/

41. Kopylova E, Navas-Molina J, Mercier C et al (2016) Open-source sequence clustering methods improve the state of the art. mSystems 1(1):e00003−e00015. https://doi.org/10.1128/mSystems.00003-15

42. Nearing JT, Douglas GM, Comeau AM et al (2018) Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 6:e5364. https://doi.org/10.7717/peerj.5364

43. Xue Z, Kable ME, Marco ML (2018) Impact of DNA sequencing and analysis methods on 16S rRNA gene bacterial community analysis of dairy products. mSphere 3(5):e00410−e00418. https://doi.org/10.1128/mSphere.00410-18

44. Liu W, McManus G, Lin X et al (2021) Distribution patterns of ciliate diversity in the South China Sea. Front Microbiol 12:689688. https://doi.org/10.3389/fmicb.2021.689688

45. Huang S, Li K, Jiang G et al (2011) Influence of salinity fronts on nutrition substance and biology in Pearl River estuary (in Chinese). Water Resour Prot 27(2):18−25. https://doi.org/10.3969/j.issn.1004-6933.2011.02.005

46. Zhang X, Shi Z, Huang X et al (2013) Phytoplankton abundance and size-fractionated structure in three contrasting periods in the Pear River Estuary. J Mar Res 71(3):187−210. https://doi.org/10.1357/002224013807719464

47. Qiu D, Huang L, Zhang J et al (2010) Phytoplankton dynamics in and near the highly eutrophic Pearl River Estuary, South China Sea. Cont Shelf Res 30(2):177−186. https://doi.org/10.1016/j.csr.2009.10.015

48. Zhu J, Hong Y, Zada S et al (2018) Spatial variability and co-acclimation of phytoplankton and bacterioplankton communities in the Pearl River Estuary, China. Front Microbiol 9:2503.

https://doi.org/10.3389/fmicb.2018.02503

49. Liu J, Fu B, Yang H et al (2015) Phylogenetic shifts of bacterioplankton community composition along the Pearl Estuary: the potential impact of hypoxia and nutrients. Front Microbiol 6:64. https://doi.org/10.3389/fmicb.2015.00064

50. Hu J, Peng P, Chivas AR (2009) Molecular biomarker evidence of origins and transport of organic matter in sediments of the Pearl River estuary and adjacent South China Sea. Appl Geochem 24(9):1666–1676. https://doi.org/10.1016/j.apgeochem.2009.04.035

51. Xu Y, Soininen J, Zhang S et al (2021) Disentangling the relative roles of natural and anthropogenic-induced stressors in shaping benthic ciliate diversity in a heavily disturbed bay. Sci Total Environ 801:149683. https://doi.org/10.1016/j.scitotenv.2021.149683

52. Wang C, Li X, Lai Z et al (2009) Seasonal variations of *Aulacoseira granulata* population abundance in the Pearl River Estuary. Estuar Coast Shelf Sci 85(4):585–592. https://doi.org/10.1016/j.ecss.2009.09.031

53. Gordon W, O'Sullivan LA, Meng Y et al (2015) Archaeal community diversity and abundance changes along a natural salinity gradient in estuarine sediments. FEMS Microbiol Ecol 91(2):1–18. https://doi.org/10.1093/femsec/fiu025

54. Zhu C, Liu W, Li X et al (2021) High salinity gradients and intermediate spatial scales shaped similar biogeographical and co-ccurrence patterns of microeukaryotes in a tropical freshwater-saltwater ecosystem. Environ Microbiol 23(8):4778–4796. https://doi.org/10.1111/1462-2920.15668

55. Logares R, Brte J, Bertilsson S et al (2009) Infrequent marine-freshwater transitions in the microbial world. Trends Microbiol 17(9):414–422. https://doi.org/10.1016/j.tim.2009.05.010

56. Mai Y, Lai Z, Li X et al (2018) Structural and functional shifts of bacterioplanktonic communities associated with spatiotemporal gradients in river outlets of the subtropical Pearl River Estuary, South China. Mar Pollut Bull 136:309–321. https://doi.org/10.1016/j.marpolbul.2018.09.013

57. Xu J, Wilkinson M, Chen M et al (2021) Concatenated data and dense taxon sampling clarify phylogeny and ecological transitions within Hypotricha. Zool Scr 50(1):125–139. https://doi.org/10.1111/zsc.12459

58. Dunthorn M, Otto J, Berger SA et al (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. Mol Biol Evol 31(4):993–1009. https://doi.org/10.1093/molbev/msu055

59. Sun P, Clamp J, Xu D et al (2016) An integrative approach to phylogeny reveals patterns of environmental distribution and novel evolutionary relationships in a major group of ciliates. Sci Rep 6(1):21695. https://doi.org/10.1038/srep21695

60. Bachy C, Gomez F, Lopez-Garcia P et al (2012) Molecular phylogeny of tintinnid ciliates (Tintinnida, Ciliophora). Protist 163(6):873–887. https://doi.org/10.1016/j.protis.2012.01.001

61. Zhao Y, Yi Z, Warren A et al (2018) Species delimitation for the molecular taxonomy and ecology of the widely distributed microbial eukaryote genus *Euplotes* (Alveolata, Ciliophora). Proc Royal Soc B 285(1871):20172159. https://doi.org/10.1098/rspb.2017.2159
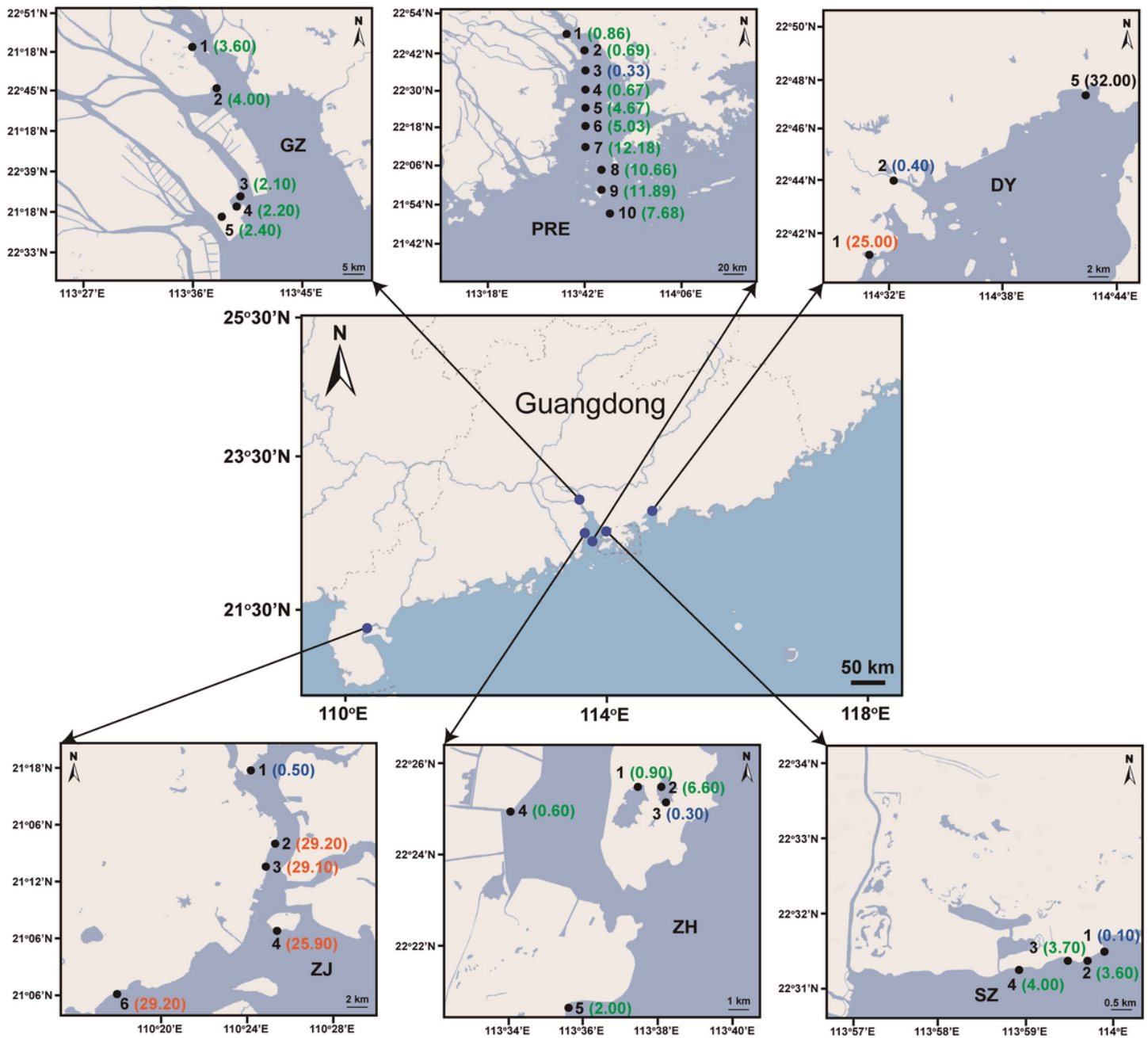
Figures

Figure 1

The location of sampling regions, modified from Zhu et al [11]. Sampling areas and their abbreviations are as follows: Daya Bay (DY1, DY2, DY5), the Pearl River Estuary (PRE1-PRE10), and the coast of Guangzhou (GZ1-GZ5), Shenzhen (SZ1-SZ4), Zhuhai (ZH1-ZH5) and Zhanjiang (ZJ1-ZJ4, ZJ6). The number shown in brackets represents the salinity of sampling sites (unit: PSU), and salinity values of freshwater, low-brackish, high-brackish and marine are colored in blue, green, orange and black.
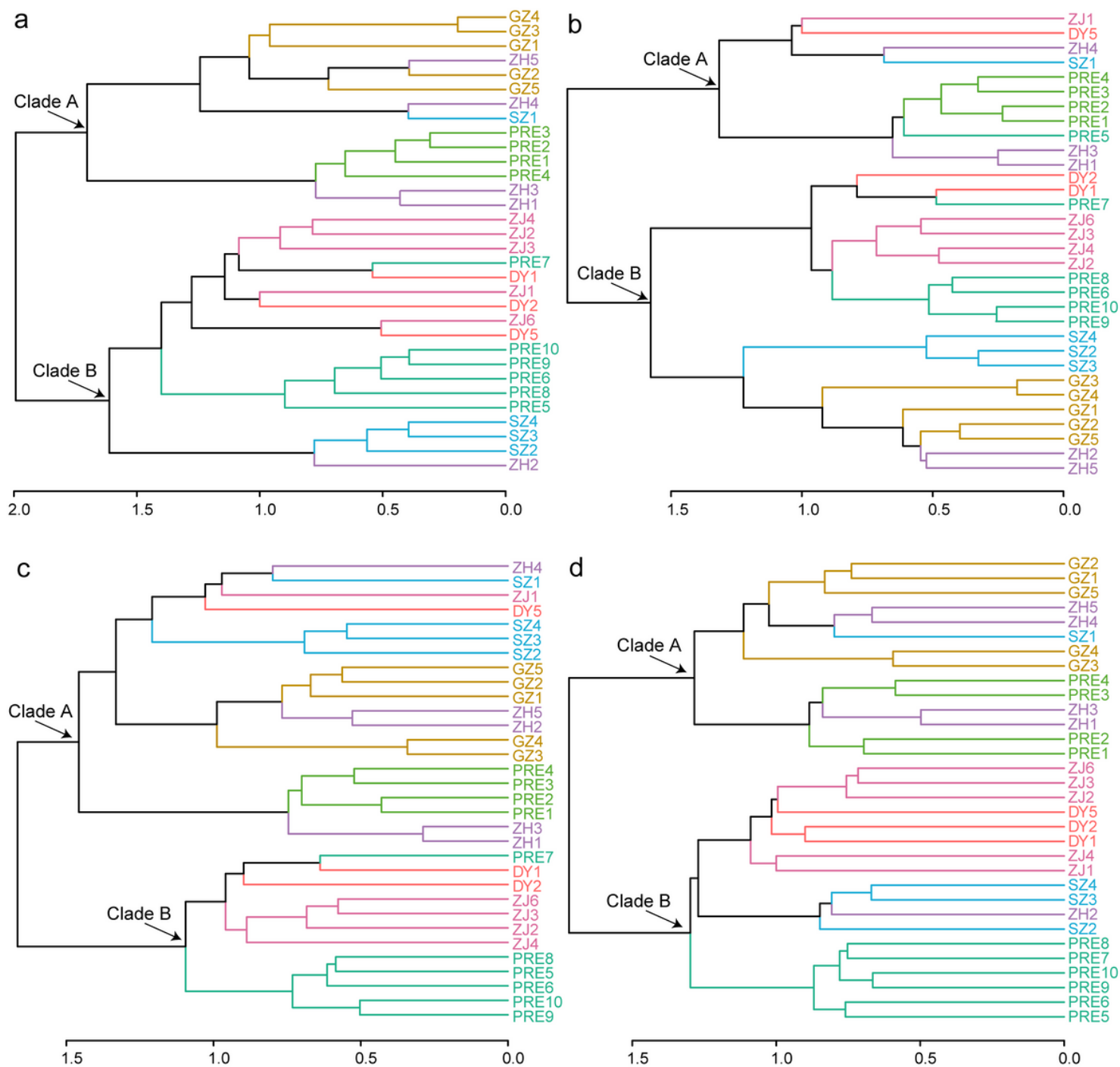
Figure 2

Hierarchical Clustering analysis for oligotrichous community of 32 samples from four datasets. a. DADA2-100E; b. UPARSE-97E; c. UCLUST-97E; d. UCLUST-99E. Samples from the same sampling area are indicated by the same color, except for samples form PRE. Samples named PRE1-4 and PRE5-10 are indicated by different colors, respectively.
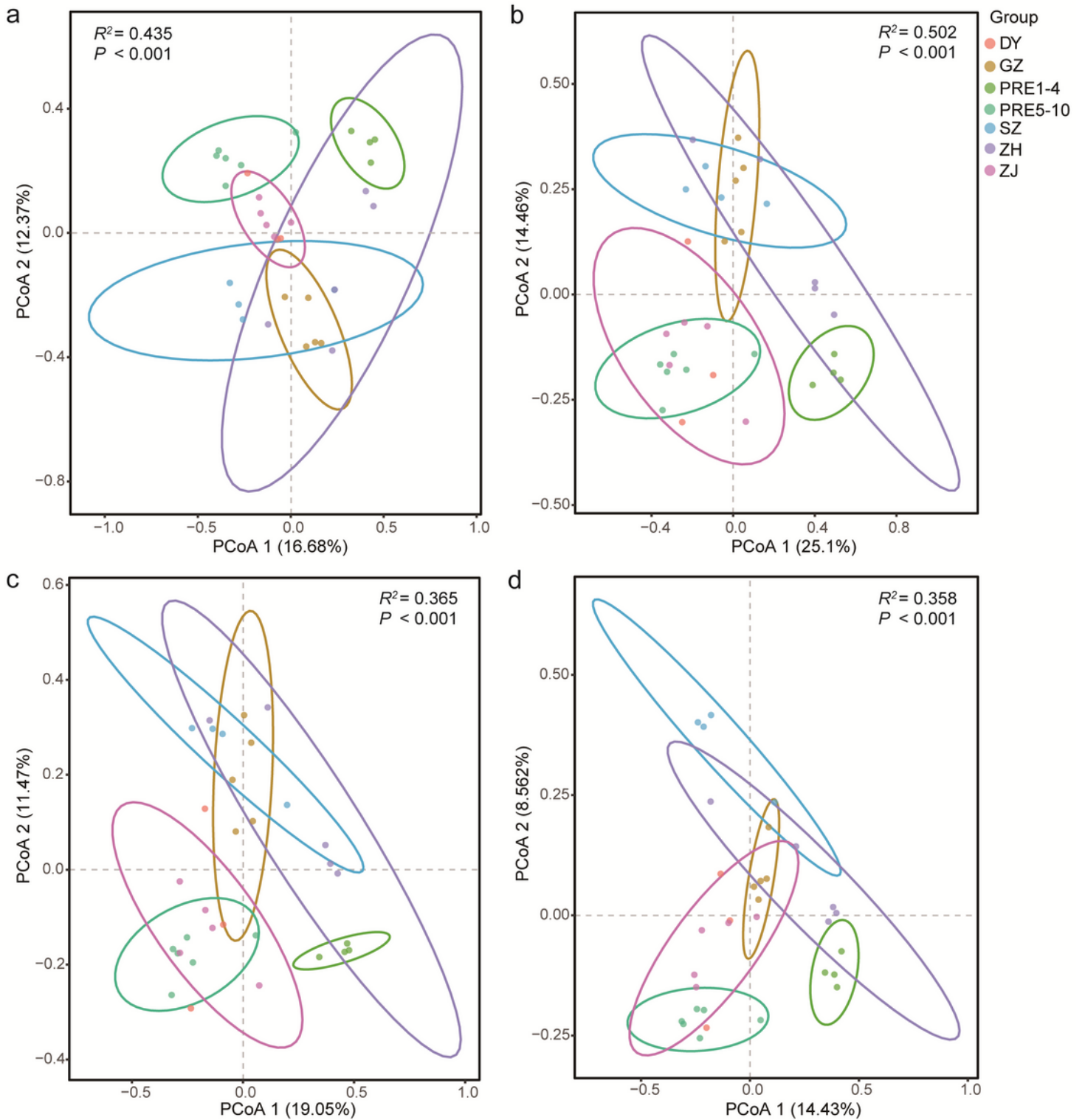
Figure 3

The distributions of oligotrichous communities from four datasets (a. DADA2-100E; b. UPARSE-97E; c. UCLUST-97E; d. UCLUST-99E) based on Principal Co-ordinates Analysis (PCoA). Samples from the same sampling area are indicated by the same color, except for samples from PRE. Samples named PRE1-4 and PRE5-10 are indicated by different colors, respectively. The ellipses represent 95% confidence intervals except for DY (too few samples to calculate an ellipse). P represents global significance among

oligotrichous communities from different sampling areas based on ADONIS analysis. P values < 0.05 are considered as significant differences among community structures.
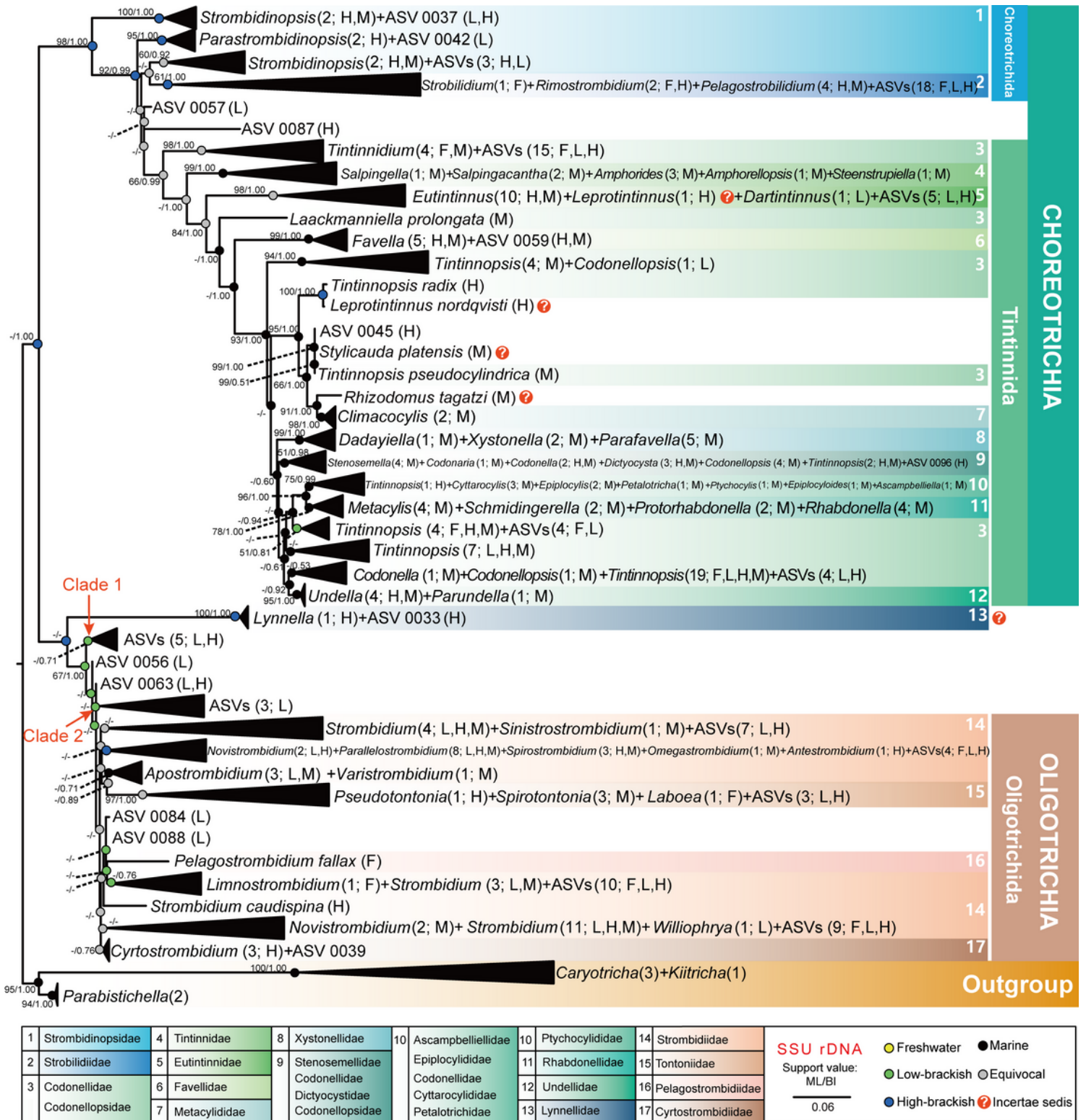

Figure 4

Maximum likelihood (ML) tree based on SSU rDNA of 103 environmental sequences (ASVs in DADA2-100E) and 197 sequences of identified species from GenBank (accession numbers in Table S1). The tree shows the topology and transition pattern, support values (ML/BI) > 50%/0.50 are labeled. The number of

collapsed sequences and ecological characteristics are shown in brackets. F, freshwater; L, low-brackish; H, high-brackish; M, marine.

Supplementary Files

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

- SupplementaryMaterials.docx