MARCO-BOLO

STRENGTHENING BIODIVERSITY OBSERVATION IN SUPPORT OF DECISION MAKING

Deliverable 2.3

Final report on the congruence between traditional and eDNA-based biodiversity observations, and their robustness across bioinformatics issues and diversity metrics. Version 1.0

28.11.2025

Joanna Warwick-Dugdale[1], Emilie Boulanger[2], Saara Suominen[2], Hanneloor Heynderickx[3], Mads Reinholdt Jensen[4], Pedro Ciarlini Junger[5], Chris Bowler[5], Haesung Jee[5], Fabrice Not[6], Jennifer Beatty[6], Aubrie Onoufriou[7], Iveta Matejusová[7], Domenico D'Alelio[8], Daniel Kumazawa Morais[4], Michael Cunliffe[1]

[1]**MBA:** Marine Biological Association
[2]**UNESCO-OBIS:** United Nations Educational, Scientific and Cultural Organization - Ocean Biodiversity Information System
[3]**VLIZ:** Flanders Marine Institute
[4]**UiT:** The Arctic University of Norway
[5]**CNRS-IBENS**
[6]**SU:** Sorbonne Université
[7]**MS:** Marine Directorate, Scottish Government
[8]**SZN:** Stazione Zoologica Anton Dohrn

MARK IF PUBLIC / SENSITIVE / CLASSIFIED

Funded by the European Union

Funded by the European Union under the Horizon Europe Programme, Grant Agreement No. 101082021 (MARCO-BOLO). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK Research and Innovation

UK participants in MARCO-BOLO are supported by the UKRI's Horizon Europe Guarantee under the Grant No. 10068180 (MS); No. 10063994 (MBA); No. 10048178 (NOC).

authority can be held responsible for them.

UK participants in Empowerlife are supported by UKRI Grant No. 10040189 (QUB).

MARCO-BOLO
STRENGTHENING BIODIVERSITY OBSERVATION IN SUPPORT OF DECISION MAKING

Document Information

Grant Agreement	101082021
Project Acronym	MARCO-BOLO
Project Title	MARCO-BOLO will strengthen European marine, coastal and freshwater biodiversity observation to understand and restore ocean health.

Deliverable Number	D2.3
Work Package Number	WP2
Deliverable Title	Report on the congruence between traditional and eDNA-based biodiversity observations, and their robustness across bioinformatics issues and diversity metrics.
Lead Beneficiary	10. MBA
Author(s)	Joanna Warwick-Dugdale (MBA), Emilie Boulanger (UNESCO), Saara Suominen (UNESCO), Hanneloor Heynderickx (VLIZ), Mads Reinholdt Jensen (UiT), Pedro Ciarlini Junger (CNRS), Chris Bowler (CNRS), Haesung Jee (CRNS), Fabrice Not (SU), Aubrie Onoufriou (MS), Iveta Matejusová (MS), Domenico D'Alelio, Jennifer Beatty (SU), Daniel Kumazawa Morais (UiT), Michael Cunliffe (MBA)
Due Date	30.11.2025
Submission Date	28.11.2025
Dissemination Level	PU - Public[1]
Type of Deliverable	R - Document, report[2]

Version 1.0	28.11.2025, Joanna Warwick-Dugdale

[1] Dissemination level: **PU:** Public, information as referred to in European Commission Decision 2015/844

[2] Type of deliverable: **R: Document,** Report

Executive summary

Here, we present the outcomes of Task 2.3: "Comparison of spatial and temporal eDNA-based vs traditional observations". The report summarises the level of agreement between traditional and eDNA-based biodiversity observations and evaluates how stable these comparisons remain when different analytical decisions are applied. In this context, performance or congruence refers to the extent to which eDNA observations reproduce the ecological patterns detected by traditional methods, including seasonal cycles, interannual trends and diversity gradients. Robustness refers to the consistency of these eDNA-derived patterns when analytical choices vary, for example the choice of bioinformatic workflow or the use of different diversity metrics.

Performance was quantified through direct comparisons of alpha and beta diversity metrics between eDNA and microscopy datasets, including statistical tests that evaluated the strength and significance of these relationships. Robustness was assessed by examining how much these diversity comparisons changed when different analytical approaches were applied, particularly across independent bioinformatic pipelines and alternative diversity metrics. Together, these assessments provide an integrated view of how reliably eDNA reflects established biodiversity records across multiple datasets and methodological approaches.

Background: This work was conducted in conjunction with tasks 2.1 and 2.2. The former produced a "Meta-analysis and review of the performance of eDNA-based approaches for quantitative and species richness metrics". This report considered the main methods employed for generation of eDNA-based biodiversity observations, and the depth and breadth of published research. The results of the meta-analysis indicated that in total-species data, the relative abundance of sequencing reads (obtained with metabarcoding), were significantly and positively correlated with the relative abundance of taxa obtained with traditional monitoring methods. In addition, consideration of eDNA compatible diversity metrics and statistical tools presented potential pitfalls in using richness-only diversity estimates (i.e. presence/absence), and promoted indices influenced by the taxa 'evenness' (i.e. relative abundance). Task 2.2, "Set of databases and software/pipelines facilitating the implementation of eDNA-based monitoring in terms of study design, data analysis and sharing", included previously unreleased long-term (20-year) time-series data from the Western English Channel, generated for task 2.3, and analysed here. Currently, WP2 is in the process of generating a similarly long-term time series from the Western English Channel of complementary eDNA data for traditional fish count data, facilitating an ongoing investigation into the performance of such data in comparison to traditional larval fish and fish egg net tow data.

Approach: In task 2.3 we built on the above by comparing eDNA-based observations against traditional counts data for retrieving different biodiversity facets (e.g., communities taxonomic diversity, taxa abundance, etc.) in space and time, across taxa and ecosystems. To facilitate this task, we undertook two primary routes of investigation. To compare eDNA with traditional observations over the course of time, we generated, curated and analysed a long-term 'total eukaryote'

metabarcoding time series (18S rRNA gene V9 region amplicons) from the Western English Channel, Western Channel Observatory (WCO) coastal station: 'L4', covering a 20-year period at a resolution of ~weekly samples (surface waters). This eDNA dataset was then compared to a complementary WCO dataset of plankton counts obtained by microscopy methods (i.e. 'traditional' data). Both datasets were filtered to retain the major top-level plankton groups observed in the microscopy data: Diatoms, Dinoflagellates, Coccolithophores, Flagellates and Ciliates. Similarities and discordances between the intra- and interannual patterns in biodiversity were investigated, and implications/possible mechanisms behind these considered.

To compare eDNA derived observations with traditional data from spatially resolved data, we leveraged metabarcoding (18S rRNA gene V4 and V9 regions) and microscopy datasets generated by the *Tara Oceans* expedition. Here, *Tara* samples from across the Global Ocean that were taken from the surface (3-7 m depth) and from the 20–180 µm size fraction were used to ensure congruence of eDNA samples with data available from light microscopy counts (N = 46 stations). Direct comparisons of phytoplankton eDNA and traditional microscopy count diversity values per-sample (i.e. alpha diversity: that of a given area/timepoint) and across samples (i.e. beta diversity: derived from differences between areas/timepoints) were performed. Congruency in diversity values were compared across the photosynthetic community as a whole, and in size major phytoplankton groups: Diatoms; Dinophyceae; Haptophyta; Chrysophyceae; Dictyochophyceae and 'Other Chlorophyta'. The methodological effects of different metabarcoding regions and analytical approaches were also assessed. For completeness, a direct comparison of phytoplankton diversity values was also applied to the Western English Channel ('L4') time-series.

In addition, we further assessed the variation of eDNA performance to bioinformatics issues (i.e. pipeline and parameter choices) and diversity metrics employed by inviting the global eDNA community to partake in a 'Data analysis Challenge'. Participants were provided with eDNA datasets: 1. The Plankton Time series collected from SOMLIT-Astan station, Western English Channel (bimonthly sampling at 60 m depth across eight years; metabarcoding of 18S rRNA gene V4 region). 2. The 12S/16S/COI aquarium samples from the Lisbon Aquarium 'Oceanário de Lisboa' (multiple targeting of fish species using 12S, 16S and COI genes). Contributors were also supplied with standardised reference libraries (for identity of taxa), and were asked to process the data with their favoured pipeline, before sharing the results. Here, performance variation was assessed by quantifying the species detection capabilities of various pipelines, considering true positives, false positives and false negatives, between known species present and eDNA detections. Alpha diversity metrics were also compared descriptively. The plankton time series dataset with complementary microscopy data is most pertinent to the other analyses performed here, and congruency of the different eDNA datasets produced to the traditional microscopy data is reported below.

Summary of the results: A. eDNA-Traditional data comparison for L4 plankton intra- and interannual patterns in biodiversity

- The degree of congruence in intra-annual (seasonal) patterns of alpha (i.e. per sample) diversity between data types was higher from diversity metrics that accounted for the relative abundance of taxa (rather than presence-absence only).
- Diatoms, dinoflagellates and pico-algae were the most abundant phytoplankton groups in both dataset, however the different data types showed different seasonal patterns in relative abundance of high-level taxa groups.
- Across the 20 years covered by the timeseries, diversity estimation which accounted for evenness/abundance of taxa resulted in reasonable congruence between the data types, which agreed on long-term seasonal patterns of maximal community similarity, and across-season changes in community similarity (although differences were noted in which seasons had minimal community similarity). When only presence/absence of taxa was considered, there was little discernible congruency in seasonal and monthly patterns of across-year similarity between datatypes.
- Interannual patterns of diversity in the eDNA data showed evidence of a biodiversity shift, which was robust across diversity metrics and potential sequencing-related confounders. Evidence of such a shift was limited in the microscopy data, and was dependent on the type of diversity metric used.

Summary of the results B: Direct comparison of phytoplankton diversity values between different metabarcoding (eDNA) methods and microscopy data in the *Tara* dataset:

- Total phytoplankton community richness was significantly lower from microscopy data than from eDNA; the latter did not differ significantly across the different eDNA methods applied (i.e. types of marker and pipelines).
- For (alpha) diversity metrics that include abundance/evenness, differences in total phytoplankton diversity were both method- and metric dependent. Shannon and Inverse Simpson were congruent between microscopy and eDNA methods, whereas Pielou's Evenness and Hill (q = 1.5) diversity values were greater in the eDNA data generated by some eDNA approaches than in the microscopy data.
- Phytoplankton community (beta) diversity comparisons indicated significant differences in dissimilarity values obtained from microscopy data compared to those from eDNA data, across all eDNA methods and diversity metrics.
- Method comparisons across major phytoplankton groups (Diatoms, Dinophyceae, Haptophyta, Chrysophyceae, Dictyochophyceae, and 'Other Chlorophyta') indicated that for the majority of phytoplankton groups and alpha diversity metrics, diversity from microscopy was lower than from either one or both of the eDNA markers employed. Although diversity values were dependent on eDNA-approach, group-specific consistency was observed.

Summary of the results C: Comparison in diversity values between different metabarcoding (eDNA) methods and microscopy data in the coastal station L4 dataset:

- When considered as a whole, the congruence in 'L4' phytoplankton diversity values derived from eDNA and microscopy data was different for alpha and beta diversity: alpha diversity values were higher in the eDNA data, and the difference to microscopy diversity was large, whereas differences in beta diversity between data types were negligible.
- When major phytoplankton groups were tested separately, differences in alpha diversity values from eDNA and microscopy data were group dependant; although eDNA data always produced higher diversity values, the difference between datatypes was smaller for diatoms and dinoflagellates, than for coccolithophores.

Summary of 'Data Analysis Challenge' results: congruency between microscopy and eDNA diversity from ASTAN plankton times series and the effect of pipeline choice:

- Different eDNA pipelines produced different magnitudes of alpha diversity, and affected the strength and shape of seasonal patterns in alpha diversity.
- Community (Beta) diversity in microscopy data was strongly dissimilar to that observed in eDNA data when values were compared simultaneously.
- Pipeline choice was more important than season or year to the community structure of the eDNA samples; those which employed V-SEARCH pipeline were most unlike the results from alternative pipelines.
- When beta diversity patterns were calculated per data type, most eDNA pipelines were highly congruent in long-term, seasonal patterns (with winter samples being more alike across years than spring samples in microscopy data and in seven of the eight eDNA pipelines), but agreement in long-term differences in community diversity between microscopy and eDNA data sets were not evident.

Summary of conclusions/patterns noted across the results

- The degree of congruence between eDNA (metabarcoding) datasets and traditional (microscopy) data is most dependent on the diversity metric applied, however both the methods employed in data production and processing, and the nature of the community being observed, further impacts the degree of similarity in the diversity patterns between data types.
- Time series data suggests that long-term patterns in seasonal diversity are more congruent between microscopy and eDNA data than changes in plankton diversity between years.
- Largely, estimations of diversity that account for differences in abundance between taxa (i.e. the 'evenness' of a community) produced the most congruent patterns of diversity between microscopy and metabarcoding data, as observed in both intra-annual (alpha- and beta) diversity of the plankton community in the coastal station 'L4' timeseries, and the total

phytoplankton alpha diversity values of the *Tara Oceans* dataset; Diversity estimators that considered only presence-absence (i.e. 'richness'; Jaccard dissimilarity) produced largely divergent pictures of diversity from the traditional microscopy and eDNA datasets.

- Designation of taxa into major plankton/phytoplankton groups revealed group-dependent differences in diversity patterns between the data types, both in intra-annual relative abundances and in total diversity values.

Contents

1. Importance

Biodiversity is essential to the future health and welfare of humanity, but is in widespread, human-driven decline (Diaz *et al*, 2019). Accurate monitoring of biodiversity is vital for a comprehensive understanding of ecosystem functioning and effective management (Strange *et al*, 2024), and escalation of biodiversity loss further necessitates urgent expansion of current monitoring efforts (Leadly *et al*, 2022). Changes to plankton biodiversity are of especial concern, having the potential to jeopardise major ecosystem services including food security (as the basis of aquatic food webs) and climate regulation (via carbon cycling) (Grigoratou *et al*, 2005). Standard plankton monitoring approaches (i.e. 'Traditional' approaches) are based on light microscopy, are supported by international standards and have been used to generate long time series of data (Holland *et al*, 2025). However, this method requires highly skilled taxonomists, and is time consuming/labour intensive, presenting scalability issues. In addition, light microscopy cannot resolve smaller taxa (i.e. pico- and nano-eukaryotes).

The application of eDNA-based methods to plankton biodiversity monitoring has great potential to overcome such issues. The most established eDNA survey method is 'metabarcoding', where 'universal genes' (often encoding ribosomal DNA) of the target kingdom (e.g. Eukaryotes) are amplified via Polymerase Chain Reaction (PCR). Reference databases of DNA sequences are used to identify the members of the plankton community, and the numbers of sequences (or 'reads') can be used as a proxy for the 'relative abundance' of those taxa. Compared to traditional microscopy, this approach is comparatively cost effective, high throughput, and requires only generic molecular lab methods/skills. However, viewing plankton diversity through the 'lens of eDNA' also has inherent constraints, e.g. lack of biomass data; incomplete databases limiting taxonomic classification. Comparison of effectiveness and congruence of eDNA approaches with traditional techniques is made challenging by innate differences in data types, e.g. different taxonomic classification systems and methodological biases. However, identifying the similarities and differences in plankton biodiversity patterns observed from complementary eDNA and microscopy datasets is key to expanding diversity monitoring efforts (Holland *et al*, 2025).

By comparing spatial and temporal patterns across different diversity estimators (i.e. diversity metrics), this work will elucidate areas of compatibility and consistency, and those of discordance, between data types, playing a vital role in the integration of eDNA to traditional plankton monitoring for comprehensive assessment of plankton biodiversity.

2. Objectives

Work Package 2 aims "to enable eDNA-based approaches (i.e. single species detection, DNA metabarcoding and metagenomics) for biodiversity monitoring across trophic and functional groups (i.e. from microbes to vertebrates) and in marine, terrestrial and freshwater systems." This Report presents the outcome of Task 2.3, "Comparison of spatial and temporal eDNA-based vs traditional observations." (lead: MBA; Contributors: CNRS, UiT, VLIZ, SU, MS, SZN, UNESCO). This document, deliverable 2.3, is a "Report on the congruence between traditional and eDNA-based biodiversity observations, and their robustness across bioinformatics issues and diversity metrics." This task was approached via three main analyses, each of which included comparisons between a traditional data set obtained from light microscopy, and a complementary eDNA dataset:

A. Comparison of complementary plankton datasets spanning 20 years of sampling (2001-2021) at ~weekly resolution from the WCO Coastal Station 'L4': 'eDNA' (18S-V9 rRNA gene metabarcode sequences) and 'microscopy' (count data obtained via inverted microscopy and Utermohl counting technique; Widdicombe and Harbour (2021)). The full diversity of plankton groups identified by microscopy were included, and the temporal nature of the data was leveraged by investigation of differences in seasonal (intra-annual) and long-term (interannual) diversity patterns in each dataset.

B. Assessment of the congruency of phytoplankton diversity values generated by different metabarcoding (eDNA) methods and microscopy data in the Global Ocean *Tara Oceans* dataset. Data was filtered to include photosynthetic plankton, which are key indicators of "good environmental status" in the EU's Water Framework and Marine Strategy Framework Directives (Official Journal of the European Union, 2010). The effect of different metabarcoding approaches (i.e. targeting of different regions of the 18S rRNA gene; different bioinformatic pipelines) and of different estimators of diversity were investigated. Diversity values generated from complementary microscopy data were employed in total phytoplankton diversity values comparisons.

C. Investigation of the effect of pipeline- and parameter choices on intra- and interannual plankton diversity from an eDNA dataset, and of the resulting differences in congruency to traditional microscopy data. This assessment was undertaken via a 'Data Analysis Challenge', where investigators of eDNA data were invited to run a supplied metabarcoding dataset (18S-V4 rRNA gene sequences) from the SOMLIT-Astan time series through their own preferred pipeline. The results facilitated comparisons between the magnitude and seasonality of plankton diversity derived from traditional microscopy data and the different methods employed.

3. Approach

Below is an overview of the methods applied for creation and curation of the datasets, and comparisons of spatial and temporal eDNA-based vs traditional observations (details in Appendix IA and B):

Collection of samples and environmental variables:

A. Time series of Western English Channel Coastal Station 'L4': (https://www.westernchannelobservatory.org.uk/index.php)
B. Global Ocean samples from *Tara Oceans* sampling campaign: (https://fondationtaraocean.org/en/expedition/tara-oceans/)
C. SOMLIT-Astan time series (for the 'Data analysis Challenge'): (https://www.seanoe.org/data/00854/96634/)

Figure 1. Locations of the sampling sites for: A. Time series of Western English Channel Coastal Station 'L4' (WCO); B. Global Ocean samples from *Tara Oceans* sampling campaign (adapted from Tara Oceans Consortium, 2015); Here, samples from 46 stations with complementary eDNA- and light microscopy counts data from surface waters were included in the analysis; C. SOMLIT-Astan time series (adapted from Caracciolo *et al*, 2022).

For eDNA datasets:

1. DNA extraction, metabarcode amplification and sequencing
2. Sequence processing, dereplication, taxonomic classification (via reference databases), quality control (QC: removal of samples with few sequences and very rare sequences) (Figure 2)
3. Filtering of individual biological sequences (i.e. Amplicon Sequence Variants: ASVs) to retain plankton/phytoplankton; summing of abundance at species level (i.e. for 'operational taxonomic units' - OTUs)

For microscopy datasets:

1. Concentration and 'fixing' of cells (post sample collection)
2. Utermöhl counting technique and morphology based classification

For the comparisons:

1. Alpha diversity metrics calculated from eDNA and microscopy data
2. Transformation to relative abundance: eDNA abundance data consisting of sequence counts of ASVs/clustered ASVs (called Operational Taxonomic Units: OTUs), and microscopy abundance data (i.e. counts of individual cells).
4. Beta diversity metrics calculated
5. Relative abundance/total diversity values of major plankton/phytoplankton groups calculated
6. Intra- and interannual patterns and total diversity values plotted.
7. Significance tests

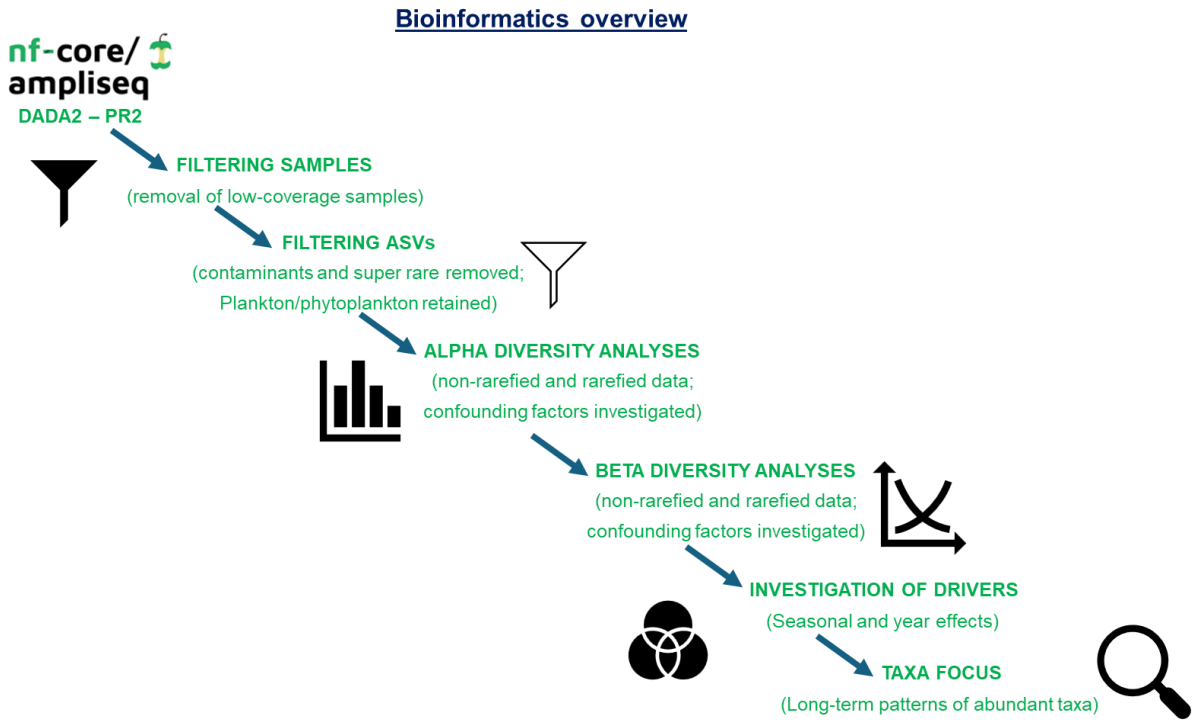


Figure 2: eDNA bioinformatic pipeline overview.

More details on the application and methods of the 'Data Analysis Challenge' are available at the dedicated public github: https://github.com/marco-bolo/data-analysis-challenge/

4. Intra-annual and Interannual biodiversity Patterns of Plankton at coastal station 'L4': Comparison of eDNA to traditional microscopy data

4.1. Comparison of long-term Intra-annual biodiversity patterns at coastal station L4.

Intra-annual (seasonal) patterns of alpha diversity varied by both data type and the diversity indices employed, however some universal patterns were observed. With metrics that accounted for relative abundance (i.e. Shannon; Hill numbers: q = 1.5), spring samples had consistently lower diversity than winter samples in both eDNA and microscopy datasets (Figure 3; Appendix II). When accounting for presence-absence only (i.e. taxa 'Richness'), both data types had higher diversity in the autumn and summer than in the spring, however spring samples had lowest diversity in the eDNA data only; microscopy samples had lowest diversity in the winter. Rarefaction of the eDNA data (i.e. subsampling all samples to 1000 reads to test effects of sequencing depth) had little effect on seasonal patterns of alpha diversity calculated from metrics that accounted for relative abundance, but eDNA richness patterns from rarefied data were similar to those from Shannon and Hill numbers (q = 1.5), and thus totally divergent from those in microscopy data (Appendix II). Accumulation curves of species richness per sample number (Figure 4A) suggest the seasonal patterns in alpha diversity observed in the eDNA dataset (i.e. winter higher diversity than spring) were not due to undersampling. However, the lower richness of winter samples recorded in the microscopy data could potentially have been influenced by sampling effort, as the accumulation curve for winter samples in the microscopy data appears to be further from reaching an asymptote than those for the other seasons (Figure 4B).

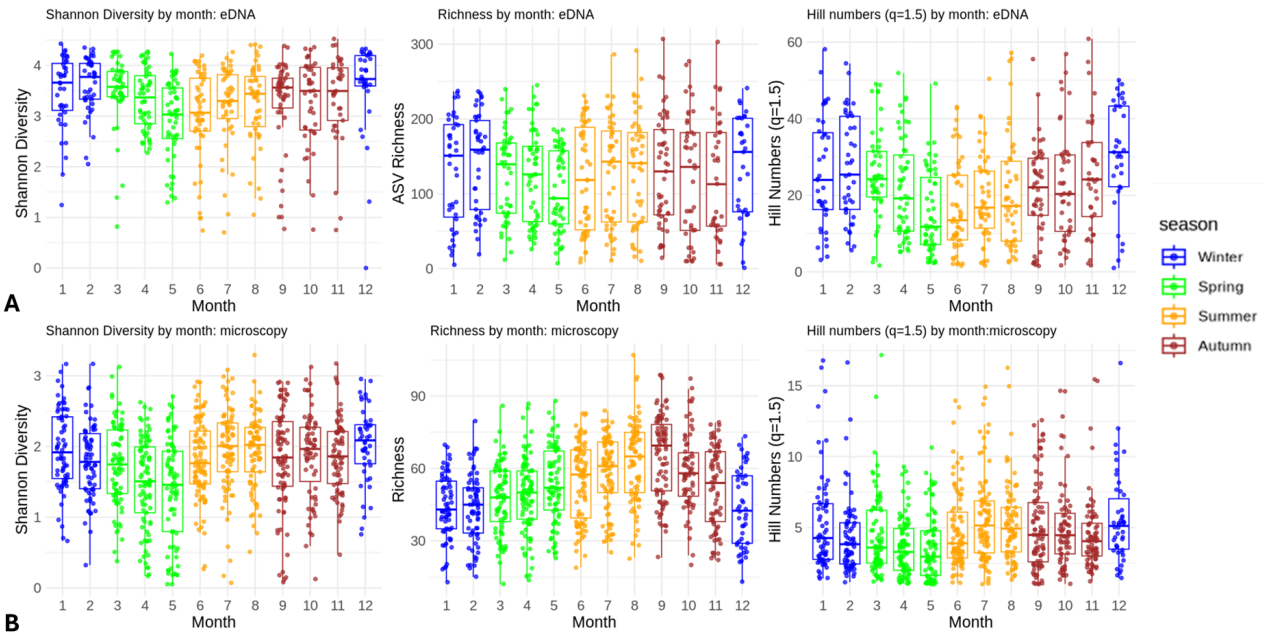


Figure 3: Seasonal and monthly plankton alpha diversity patterns in: A. eDNA data; B. Microscopy data, as calculated by the Shannon Index (left), Richness (middle) and Hill numbers (q = 1.5)

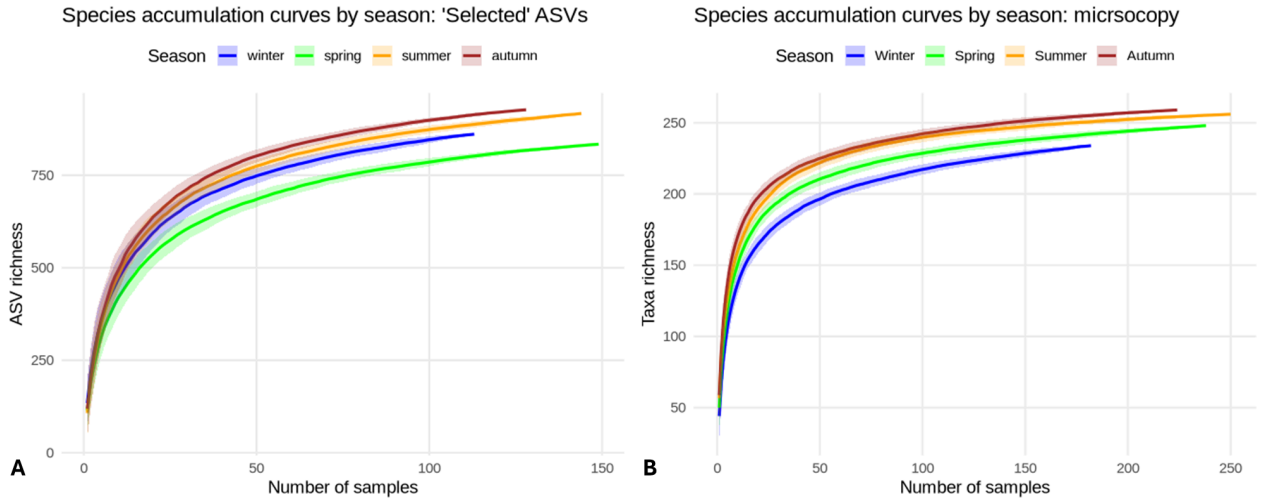


Figure 4: Accumulation curves for: A eDNA (ASVs), and B. microscopy (Taxa).

The different data types indicated different seasonal patterns of relative abundance in top-level plankton groups (Figure 5). Dinoflagellates had highest relative abundance in the eDNA data across seasons, peaking in mid summer, and although diatoms were the second most abundant group, no seasonal peak was observed. Diatoms were most abundant in the microscopy data and peaked over midsummer, whereas dinoflagellates peaked in September in this dataset. Flagellates were the second most abundant group in the microscopy data, whereas ciliates were more prevalent in the eDNA dataset than in the microscopy data.

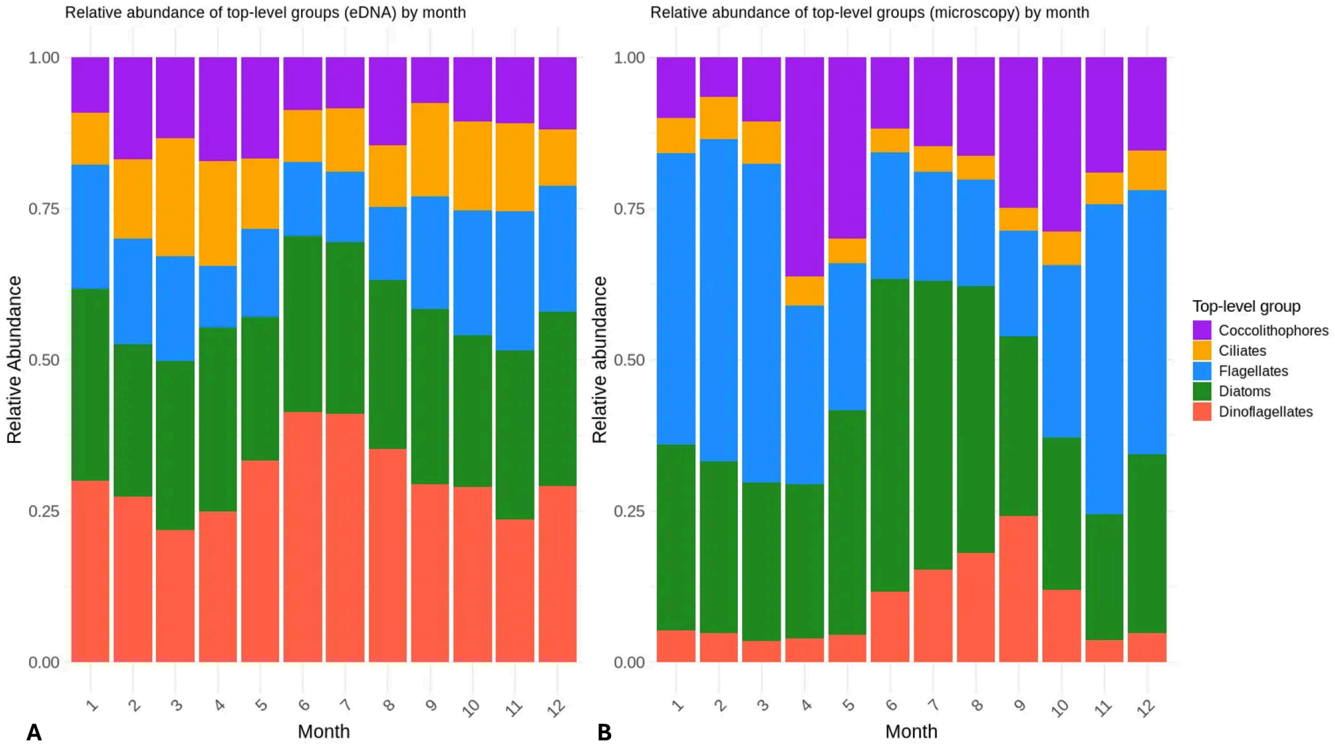


Figure 5: Interannual relative abundance of Western English Channel coastal station 'L4' plankton groups in: A. eDNA; and B. microscopy datasets (in order of most abundant groups in eDNA data; bottom up).

Diversity patterns observed as common to both data types revealed that across the 20 years, plankton community composition varied significantly according to season, and to a lesser extent, month. When community evenness was accounted for (i.e. in Bray-curtis dissimilarities), the communities from winter samples were most alike, being more similar across years than those from the summer and autumn months in both the eDNA and microscopy data (figure 6 and 7; Appendix III; dispersion plots show the distance/dissimilarity of samples to the 'centroid' (i.e. the average of all samples per season/month), thus seasons/months with lower values are more alike). Both datasets also indicated that with the progression of spring, across-year similarity of plankton communities decreased; autumn progress was characterised by an increased similarity of plankton communities across time (Figure 8). However, autumn plankton communities were less similar between years than spring communities in the eDNA dataset, whereas the microscopy data indicated that spring communities were less similar across the time series than autumn communities (Figure 7). The similarities and differences in patterns of across-year seasonal and monthly diversity observed between data types were robust to tests for the effect of eDNA sequencing depth, although, autumn samples had somewhat higher across-year similarity in the rarefied eDNA data, and variation in dissimilarity values were reduced (Appendix III).

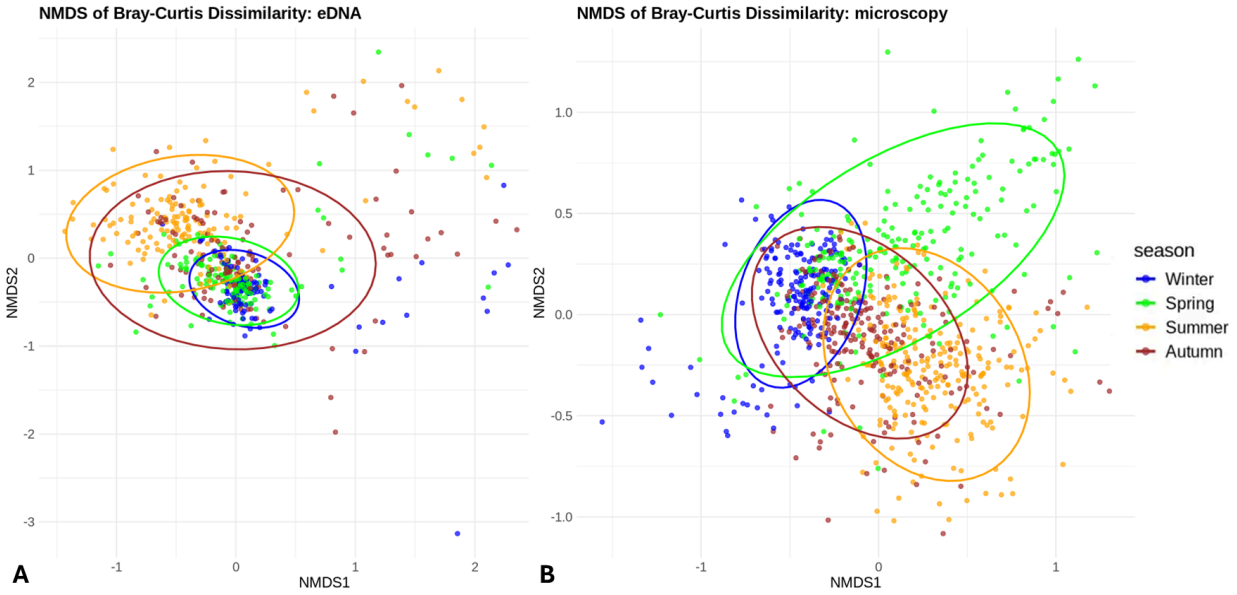


Figure 6: NMDs plots (K = 3) of Hellinger transformed Bray-Curtis Dissimilarity in Coastal Station L4 time series for: A. eDNA data (stress = 0.14), and B. microscopy data (stress = 0.17); points are coloured by season (ADONIS test).

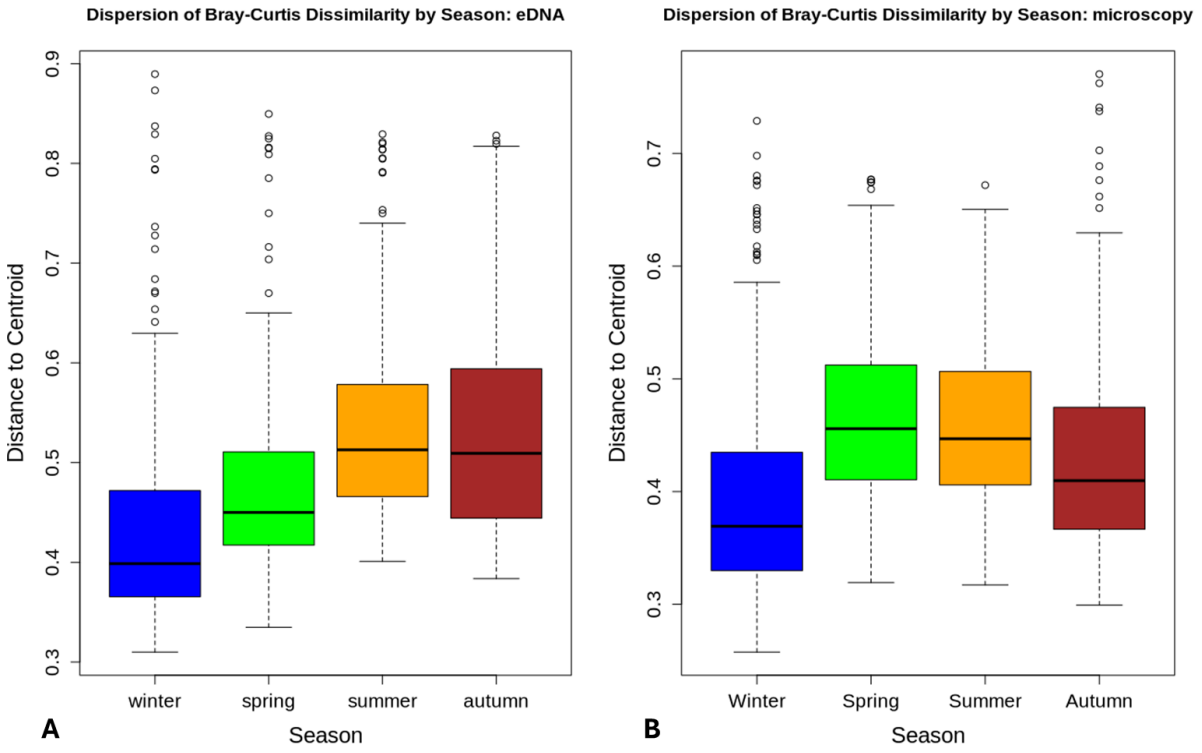


Figure 7: Dispersion test for Hellinger transformed Bray-Curtis dissimilarity in Coastal Station L4 time series: dispersion analysis (betadisper) with permutation test (permutest) to check for significant differences among seasonal groups in A: eDNA data; B. Microscopy data.

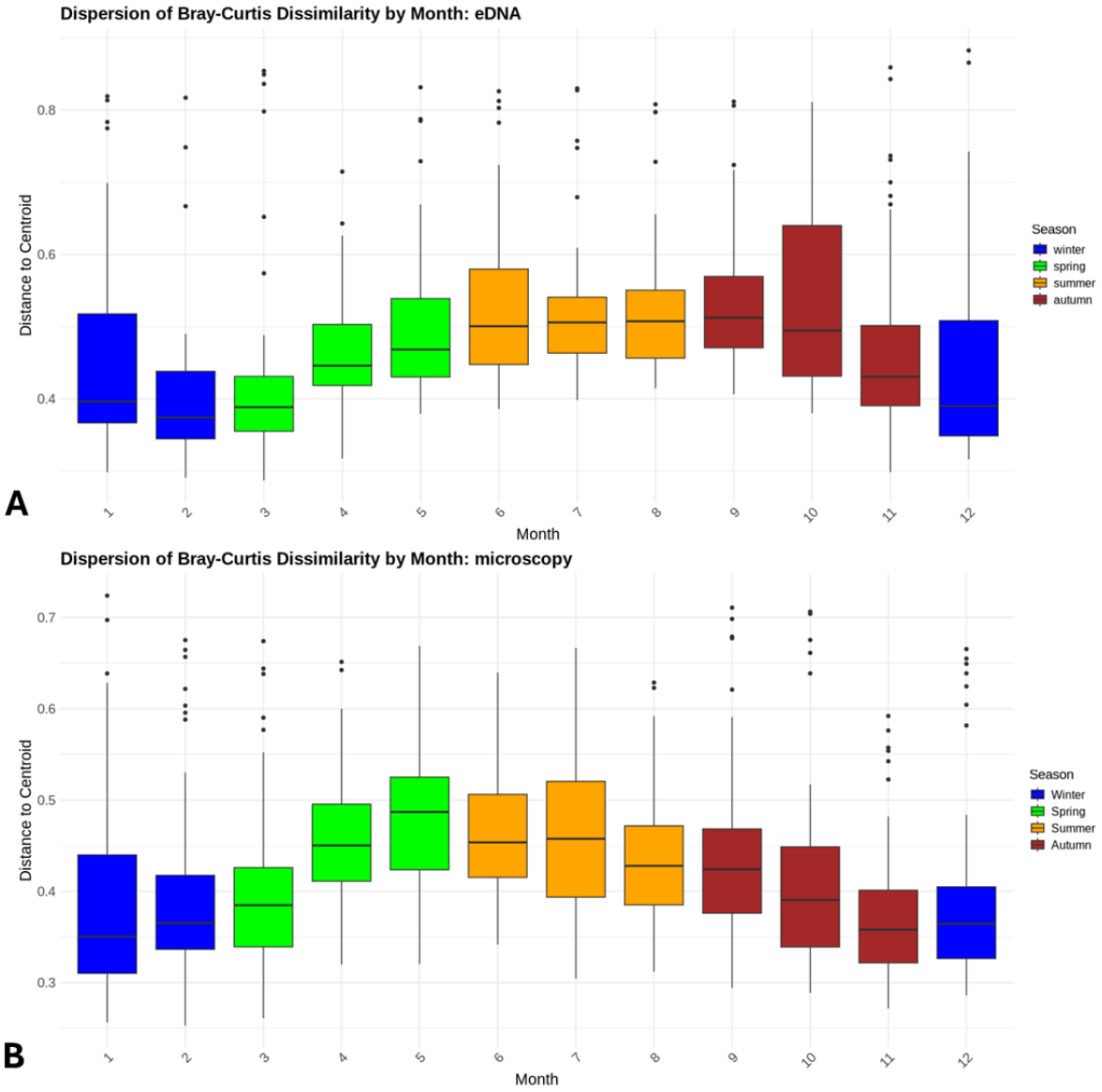


Figure 8: Dispersion test for Hellinger transformed Bray-Curtis dissimilarity in Coastal Station L4 time series: dispersion analysis (betadisper) with permutation test (permutest) to check for significant differences among months in A: eDNA data; B. Microscopy data.

When only presence-absence of taxa was taken into account (i.e. in Jaccard dissimilarities), the differences in across-year seasonal and monthly patterns between eDNA and microscopy data were pronounced (figure 9, 10 and 11). The eDNA data retained the major seasonal and monthly similarity patterns in Jaccard dissimilarity as observed for Bray-Curtis dissimilarity, with winter being most similar across years, and autumn samples being more less alike than those from spring (Figure 9A and 10A). Rarefaction (subsampling) of the data did not change these patterns, although as observed above for Bray-Curtis dissimilarity, autumn samples appeared more similar with rarefaction

(Appendix III). However, the seasonal and monthly patterns of Jaccard dissimilarity were very different in the microscopy samples: the samples from summer months are most similar across the dataset, with little difference between the samples from the other seasons (figure 9B and 10B).

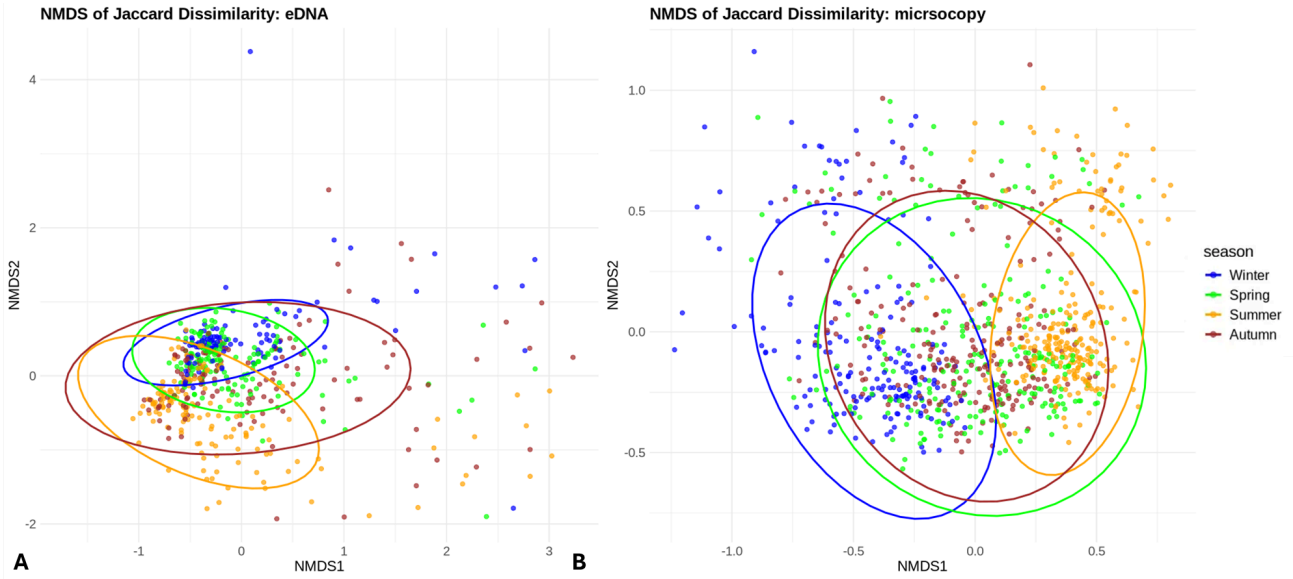


Figure 9: NMDs plots (K = 3) of Jaccard Dissimilarity in Coastal Station L4 time series for A:. eDNA data (stress = 0.14), and B. microscopy data (stress = 0.18); points are coloured by season (ADONIS test).

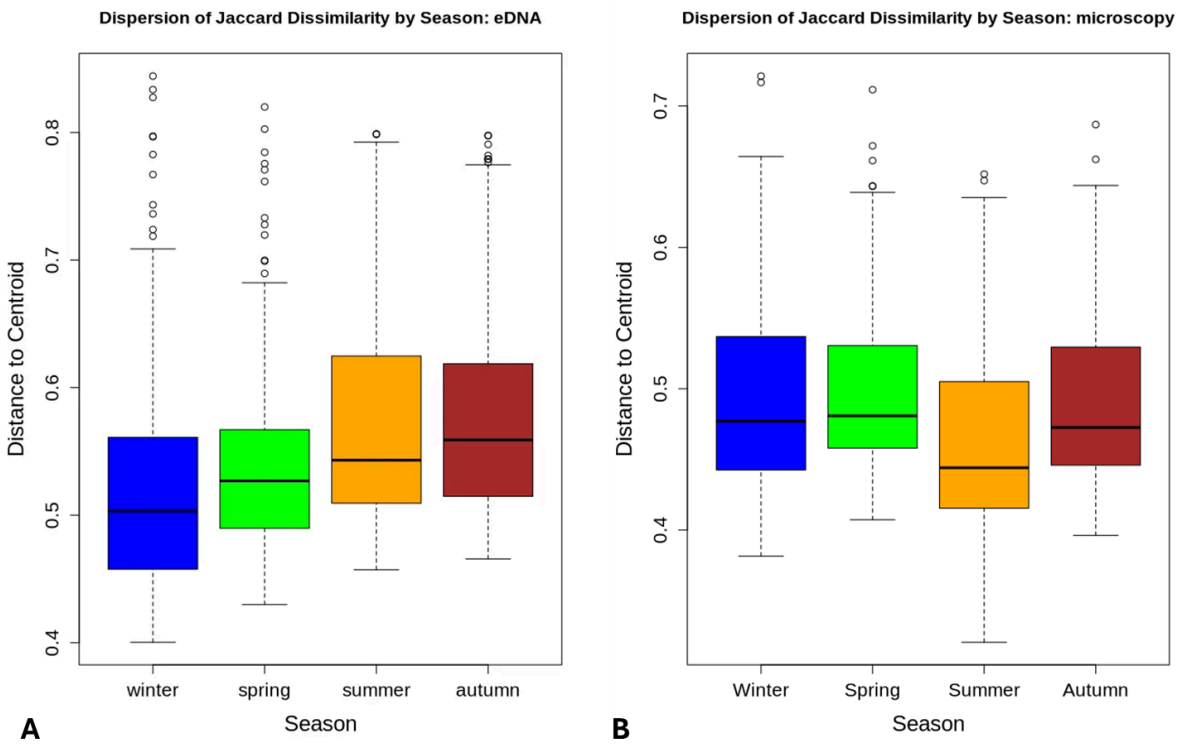


Figure 10: Dispersion test for Jaccard dissimilarity in Coastal Station L4 time series: dispersion analysis (betadisper) with permutation test (permutest) to test for significant differences among seasonal groups in A: eDNA data; B. Microscopy data.

.

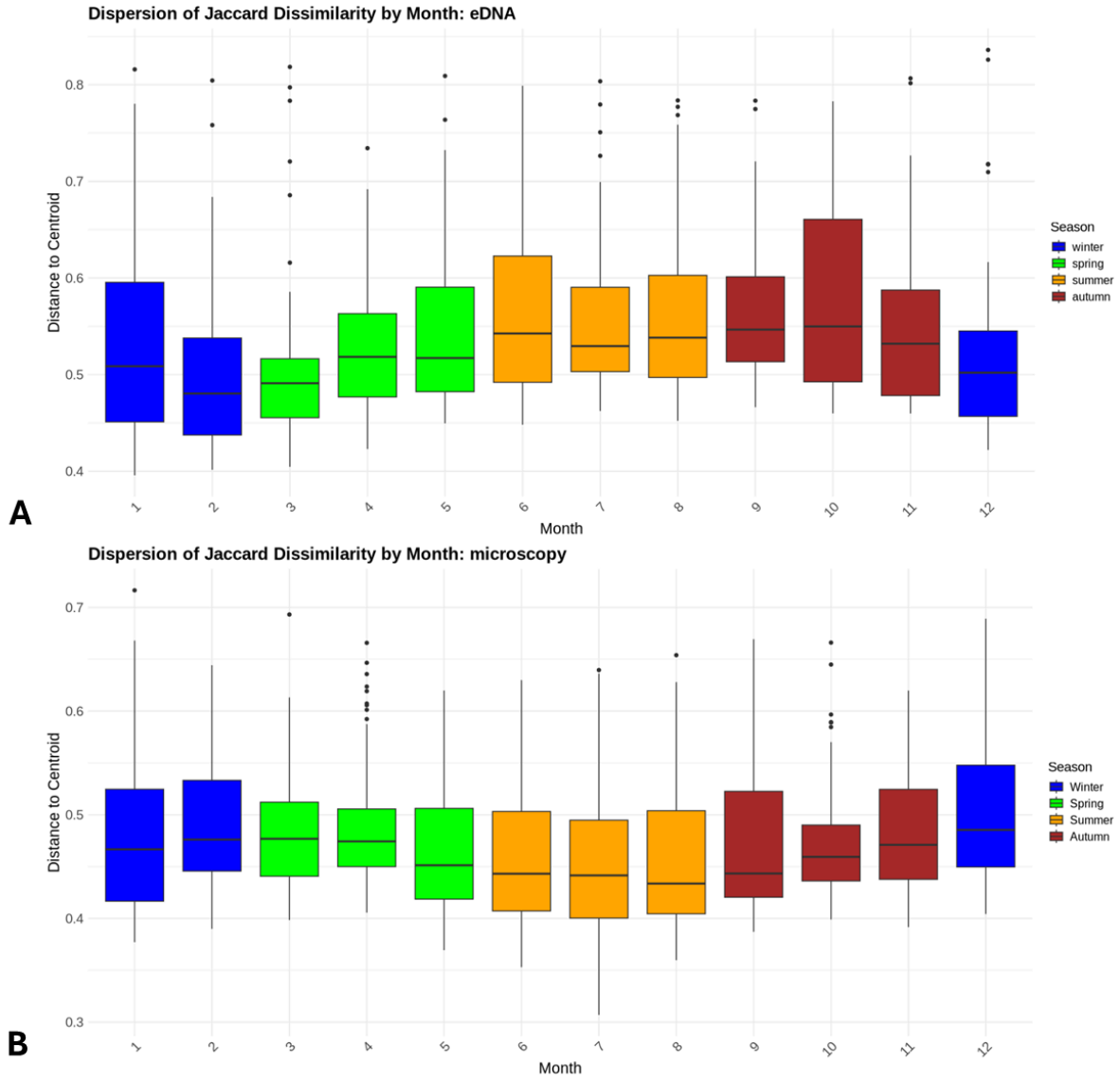


Figure 11: Dispersion test for Jaccard dissimilarity in Coastal Station L4 time series: dispersion analysis (betadisper) with permutation test (permutest) to check for significant differences among months in A: eDNA data; B. Microscopy data.

4.2. Comparison of interannual biodiversity patterns across 20 years at coastal station L4.

Interannual (long-term) signals observed in the eDNA dataset showed evidence of a biodiversity shift in the years around 2010 (figure 12A and 13).

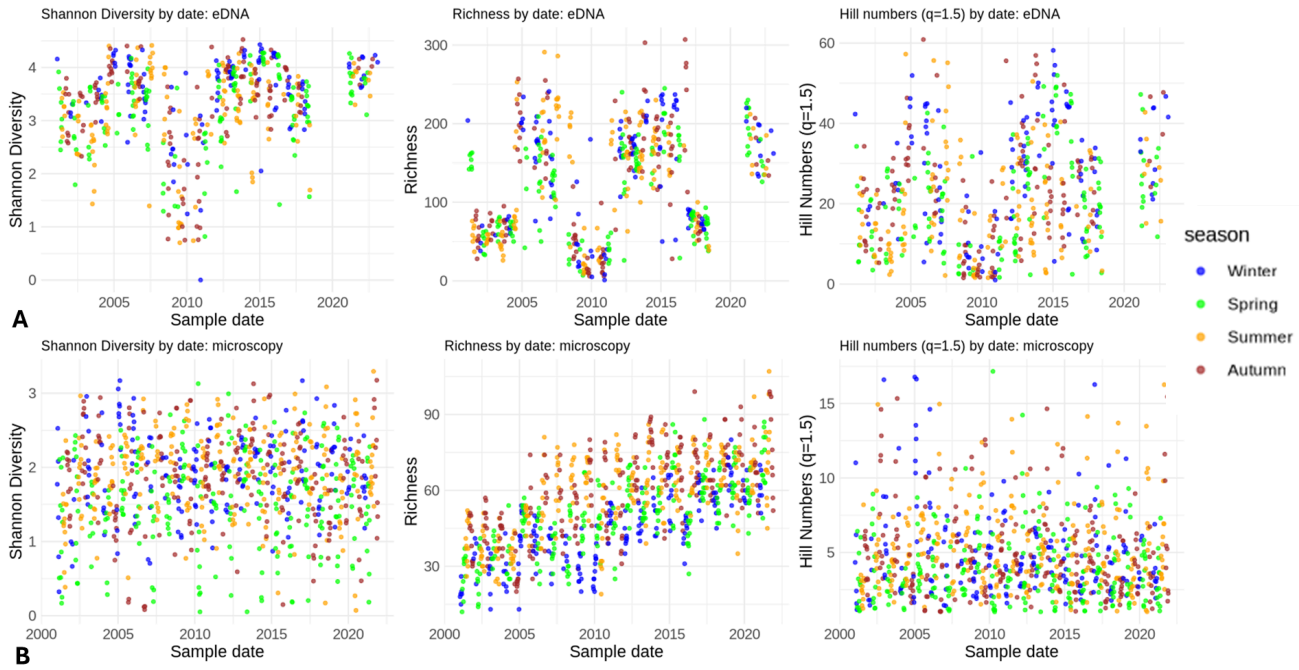


Figure 12: Long-term plankton alpha diversity patterns in: A. eDNA samples; B. Microscopy samples, as calculated by the Shannon Index (left), Richness (middle) and Hill numbers (Q = 1.5; right).

Regime shifts were characterised by a sharp decline in alpha diversity after 2007, prior to a rise and return to previous diversity levels in 2011. The shift signal was robust throughout diversity metrics and after accounting for sequencing depth and run effects (Figure 13; Appendix IV). This shift was less clear in the microscopy data (Figure 12B): the exact years identified as 'changepoints' in annual mean alpha diversity differed depending on the diversity metric employed (Figure 13; Appendix IV). No changepoints were detected in the Shannon diversity index; in addition to a general increase in microscopy taxa richness over time, the years 2004 and 2011 were identified as changepoints (the former coincides with a change in taxonomist),;and 2013 was identified as a changepoint via Hill Numbers (q =1.5) (Figure 14).

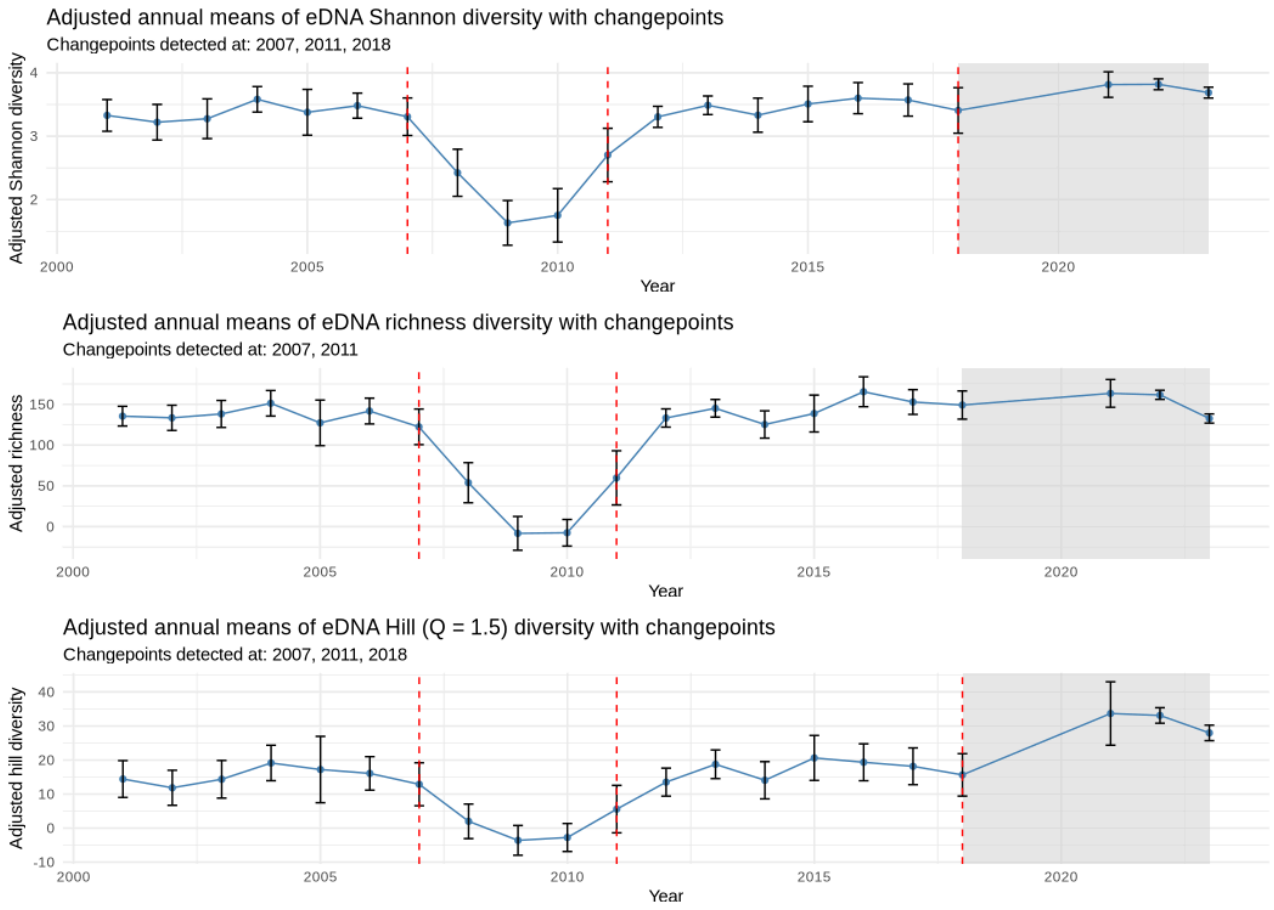


Figure 13: Adjusted annual means of plankton eDNA alpha diversity across the time series. Annual means were estimated using a linear model with year included as a fixed factor and adjusted for season and sequencing run. Points show estimated marginal means with 95% confidence intervals. Grey areas show years where data was not available and proceeding years. Dashed lines indicate regime 'changepoints' in the adjusted mean series (mean/variance shifts) using PELT. Changepoints in years 2007 and 2011 were retained in models that excluded post 2018 years..

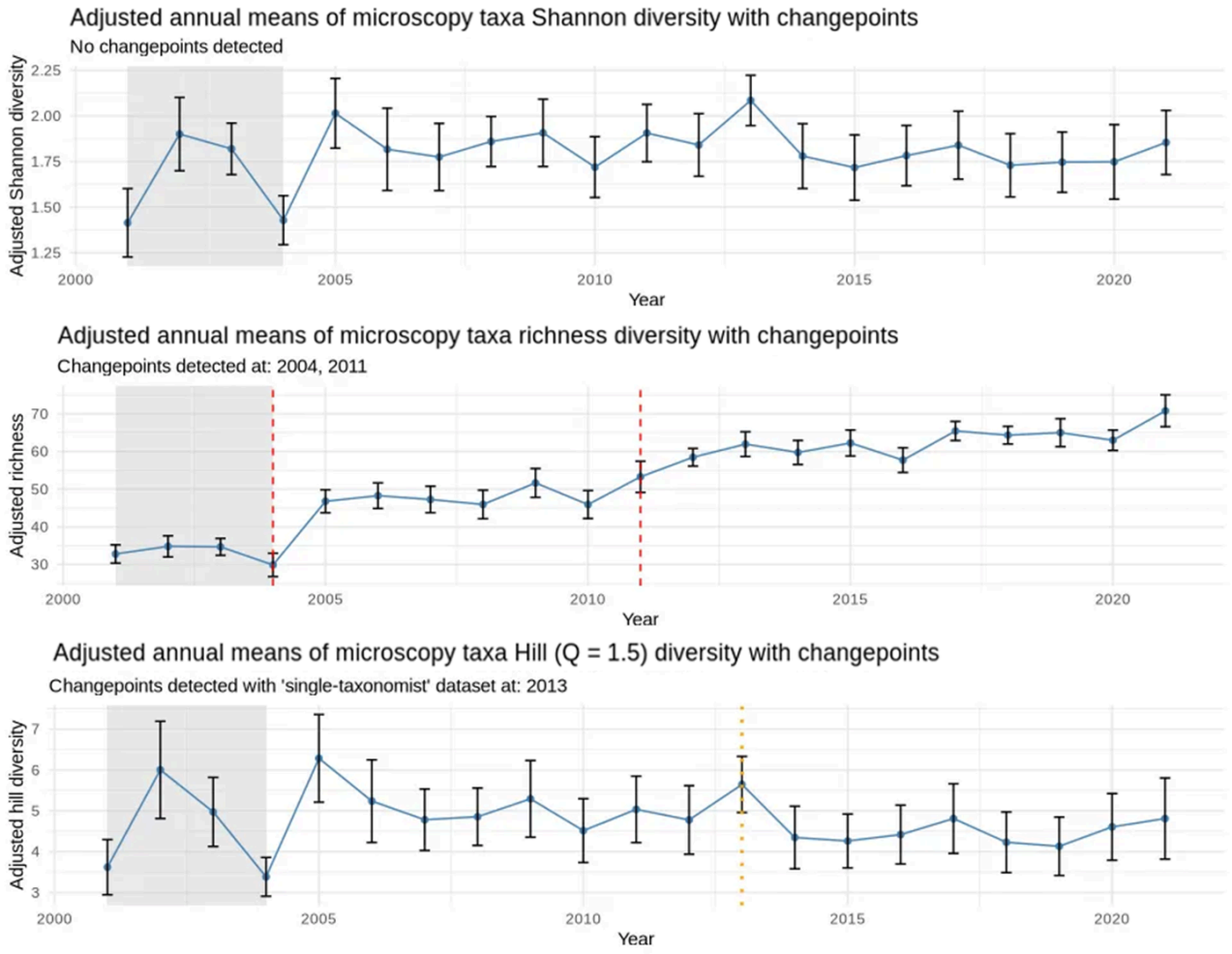


Figure 14: Adjusted annual means of plankton microscopy alpha diversity across the timeseries. Annual means were estimated using a linear model with year included as a fixed factor and adjusted for season and sequencing run. Points show estimated marginal means with 95% confidence intervals. Grey areas show years where microscopy was conducted by a separate (initial) taxonomist. Dashed red lines indicate regime 'changepoints' in the adjusted mean series (mean/variance shifts) using PELT. Identification of the year 2011 as a richness changepoint was observed in both models that included and excluded 'initial taxonomist' years; dotted orange lines mark additional changepoint when initial taxonomist years were excluded (i.e. in the 'single taxonomist' dataset).

- Summary of congruence in intra- and interannual diversity between eDNA (metabarcoding) and microscopy data in the L4 time series:
    - The degree of congruence in intra-annual (seasonal) patterns of alpha (i.e. per sample) diversity between data types was higher from diversity metrics that accounted for the relative abundance of taxa (rather than presence-absence only)

- Diatoms, dinoflagellates and pico-algae were the most abundant phytoplankton groups in both dataset, however the different data types showed different seasonal patterns in relative abundance of high-level taxa groups.
- Across the 20 years covered by the time series, diversity estimation which accounted for evenness/abundance of taxa resulted in reasonable congruence between the data types, which agreed on long-term seasonal patterns of maximal community similarity, and across-season changes in community similarity (although differences were noted in which seasons had minimal community similarity). When only presence/absence of taxa was considered, there was little discernible congruency seasonal and monthly patterns of across-year similarity between datatypes.
- Interannual patterns of diversity in the eDNA data showed evidence of a biodiversity shift, which was robust across diversity metrics and potential sequencing-related confounders. Evidence of such a shift was limited in the microscopy data, and was dependent on the type of diversity metric used.

5. Direct comparison of phytoplankton diversity values from eDNA and traditional microscopy data

5.1. Phytoplankton diversity eDNA-microscopy comparison from spatial data on a global scale: the *Tara Oceans* dataset

Differences in total '*Tara*' phytoplankton ('All Species') diversity derived from different methods were dependent on the diversity metric applied. Shannon diversity and Inverse Simpson (Hill Q:2) were the most congruent, with no significant differences to microscopy across sequencing methods, and only one metabarcoding approach was significantly different to microscopy in Pielous's Evenness (18S-V9; Swarm) (Figure 15). However, richness was significantly higher across all eDNA methods applied than captured in the microscopy data, and Hill (Q: 1.5) values were lower from three of the four metabarcoding methods than recorded in the microscopy data (no significant difference to 18S-V4; Swarm).

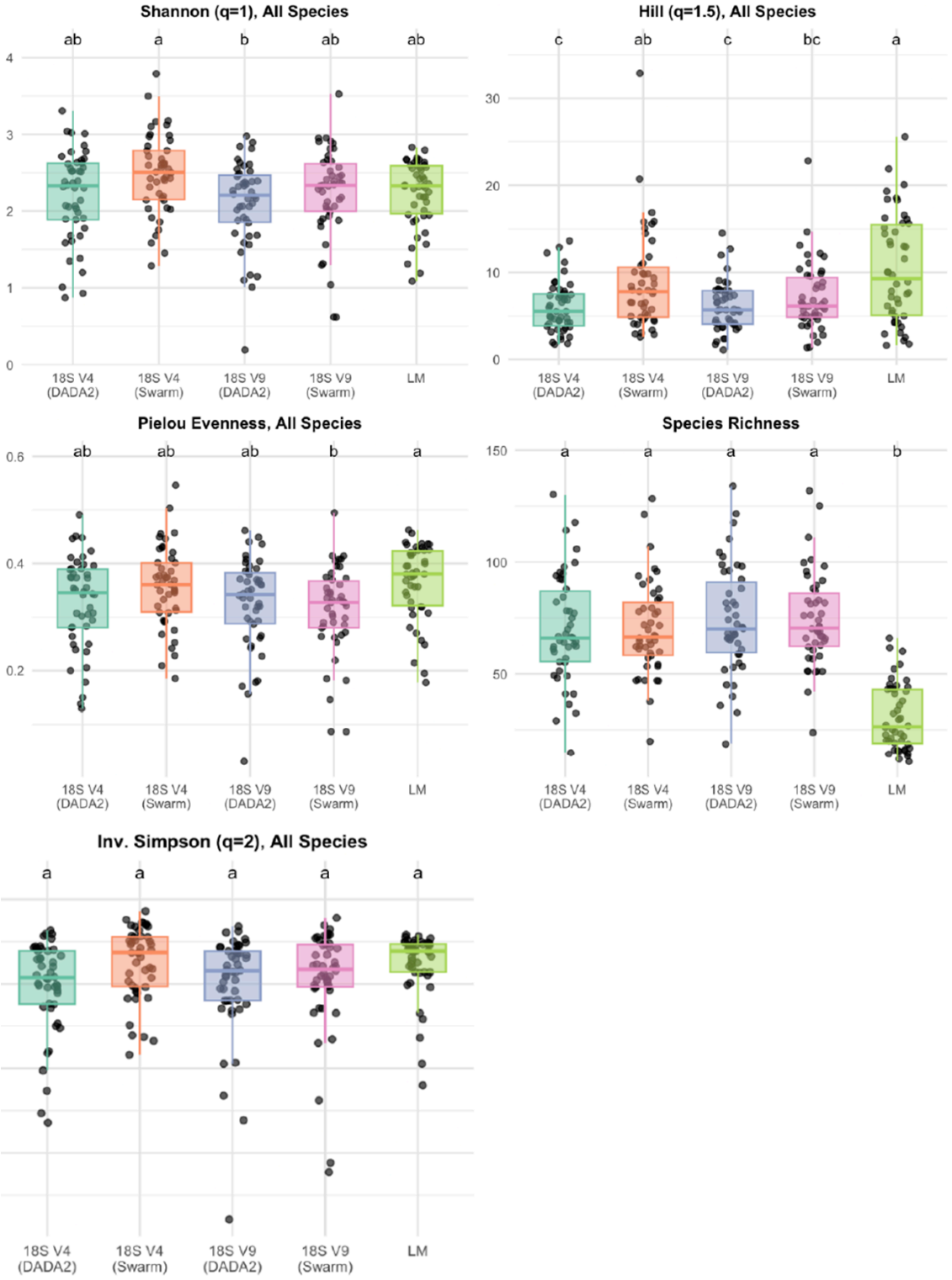


Figure 15. Boxplot comparison of *Tara* total phytoplankton alpha diversity metrics across methods. Letters above boxes indicate significant differences between sequencing methods (18S V4/V9,

DADA2 and Swarm) and light microscopy (LM) from post-hoc Tukey HSD tests ($p < 0.05$); shared letters denote nonsignificant differences.

Beta diversity comparisons of phytoplankton showed significantly higher dissimilarity in microscopy (LM) than eDNA samples across all methods and diversity metrics (Figure 16). However some differences in dissimilarity were observed between different metabarcoding methods and the three diversity metrics. For two of the three diversity indices (Bray-Curtis and Jaccard) the same patterns were observed: Processing which included internal pipeline clustering of similar sequences (i.e. 'Swarm') resulted in the highest dissimilarity across methods, with no significant difference between 18S-V4 and 18S-V9 markers. In addition, the cluster-free pipeline (i.e. DADA2) produced higher dissimilarity in 18S-V9 than in 18S-V4 data. Interestingly, Sørensen diversity was quite different, with highest dissimilarity in '18S-V4 DADA2' data and lowest from the '18S-V9 Swarm' method.

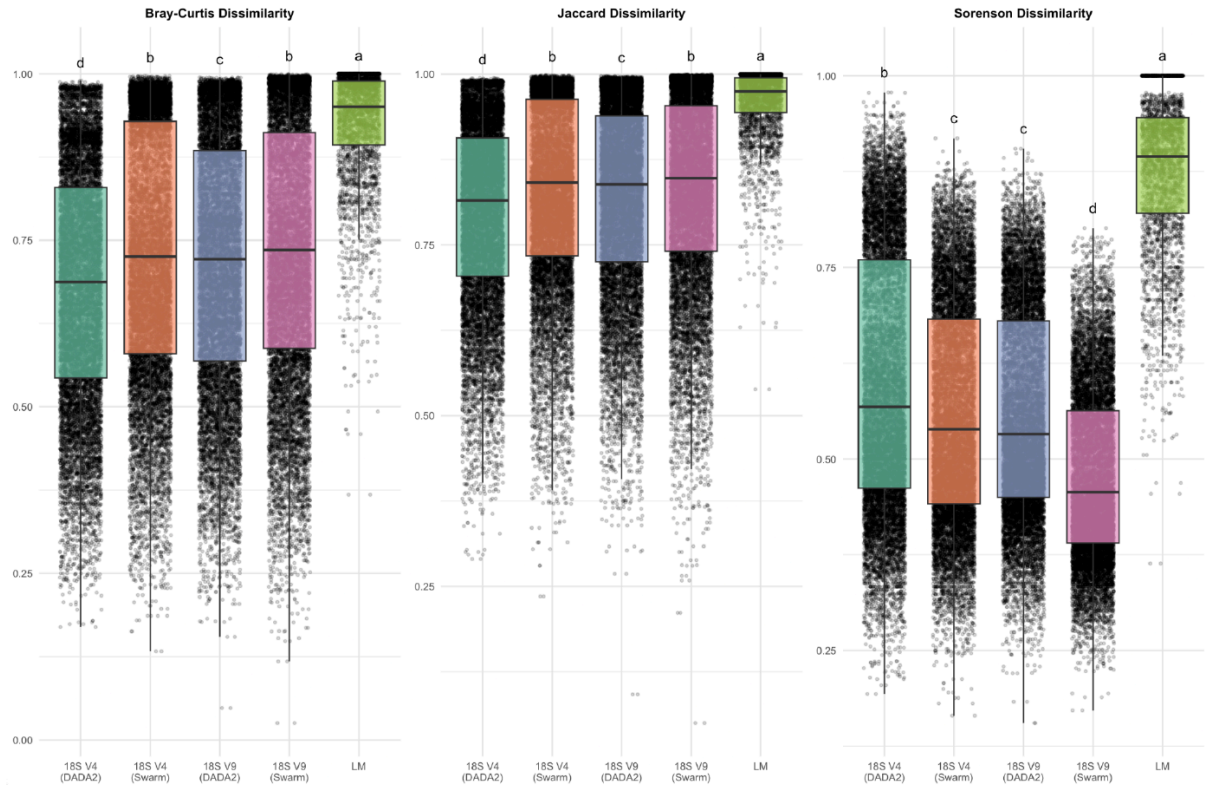


Figure 16. Comparison of beta diversity (Bray–Curtis, Jaccard, and Sørensen dissimilarities) across all methods. Letters above boxes indicate significant differences determined via ANOVA followed by post-hoc Tukey's HSD tests ($p < 0.05$).

Division of phytoplankton into major groups (Diatoms, Dinophyceae, Haptophyta, Chrysophyceae, Dictyochophyceae, and 'Other Chlorophyta') revealed both method-dependent differences and group-specific consistency and in alpha diversity values, dependent on both metric and group (figure 17, 18 and 19; Appendix V). Alpha diversity from microscopy was observed as lower than diversity values from either one or both of the eDNA markers for the majority of phytoplankton groups across most diversity metrics.

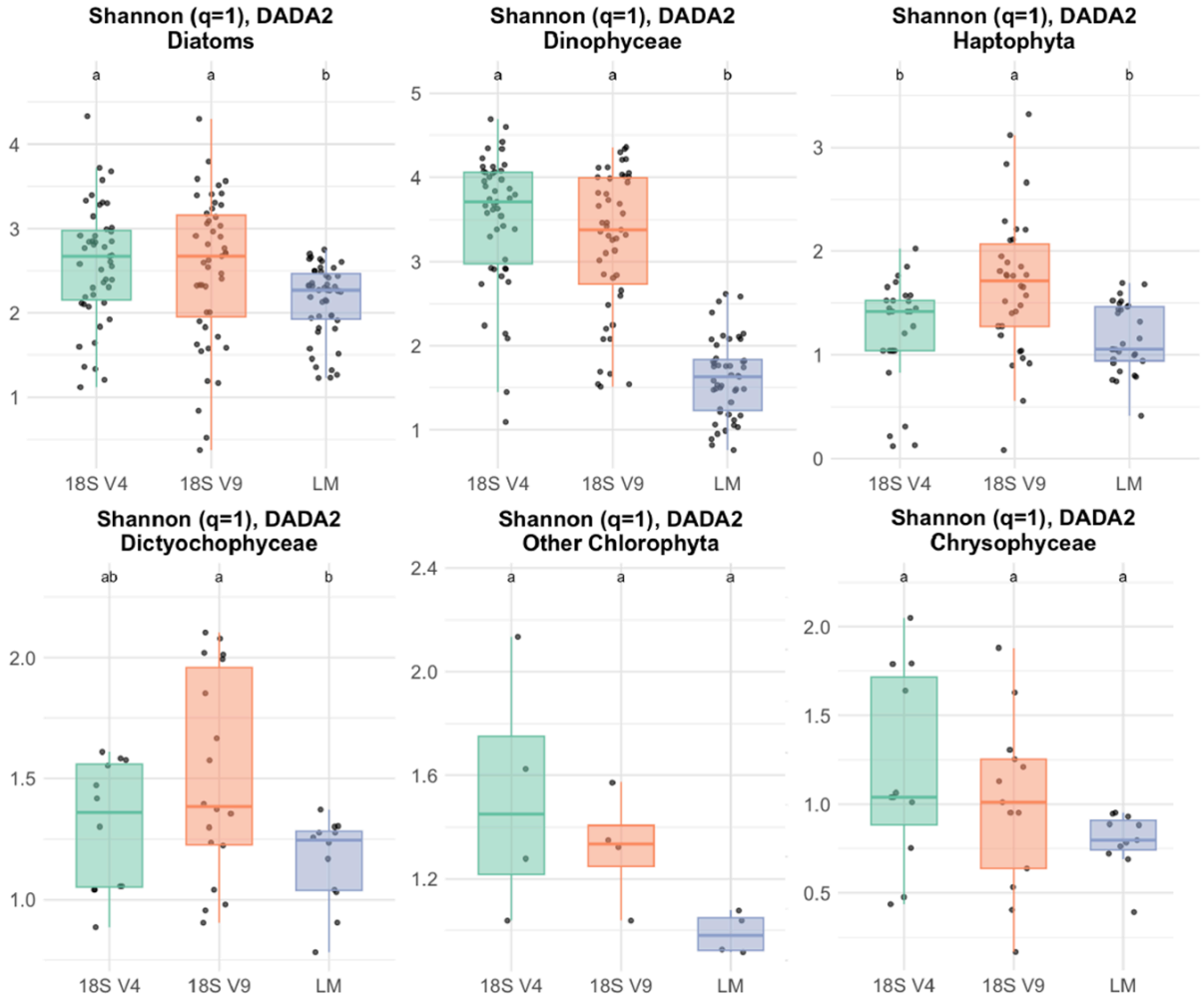


Figure 17. Shannon diversity by phytoplankton group, processed by DADA2. Letters indicate groups from ANOVA and $p < 0.05$ post-hoc Tukey HSD tests.

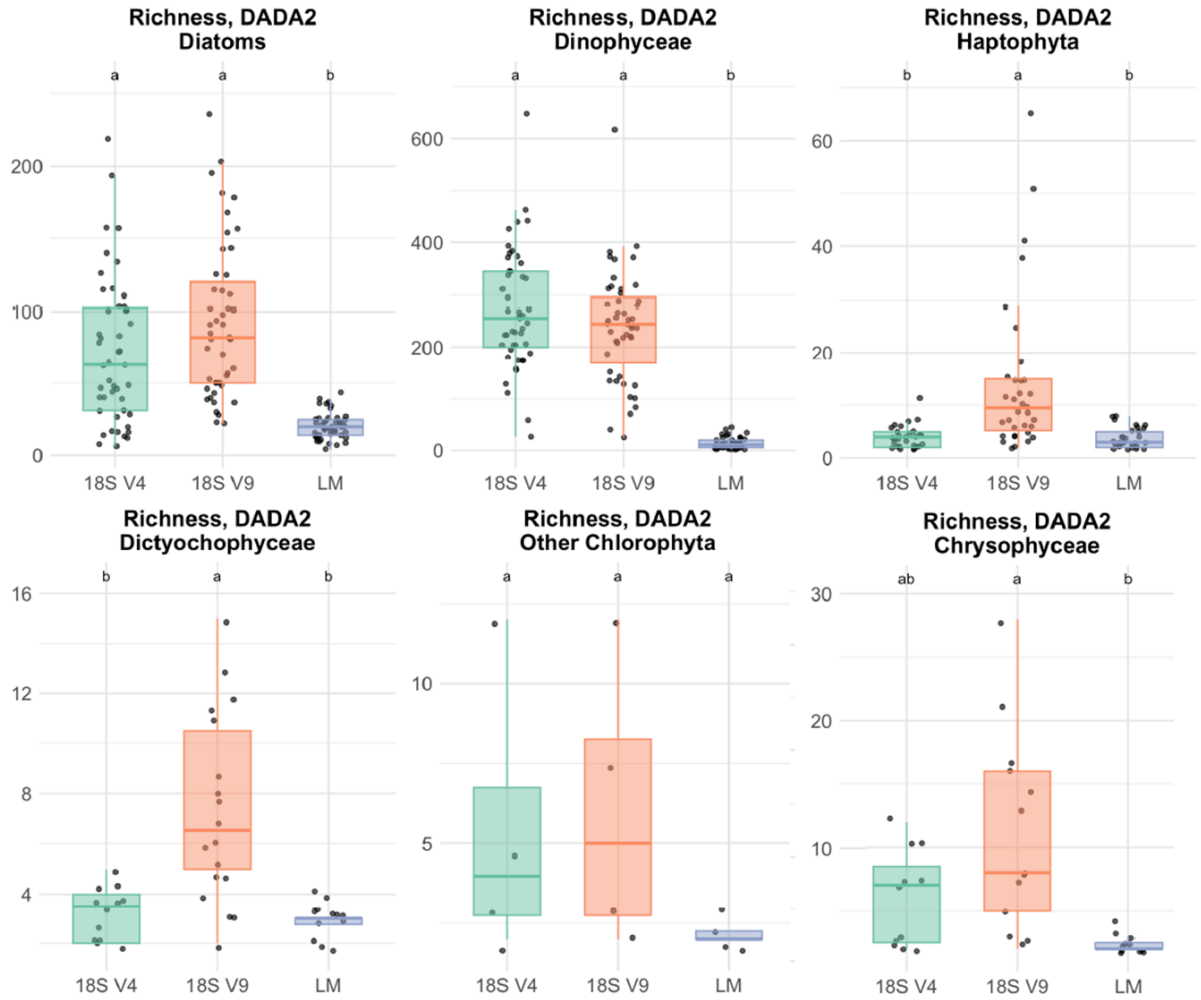


Figure 18. Richness by phytoplankton group, processed by DADA2. Letters indicate groups from ANOVA and $p < 0.05$ post-hoc Tukey HSD tests.

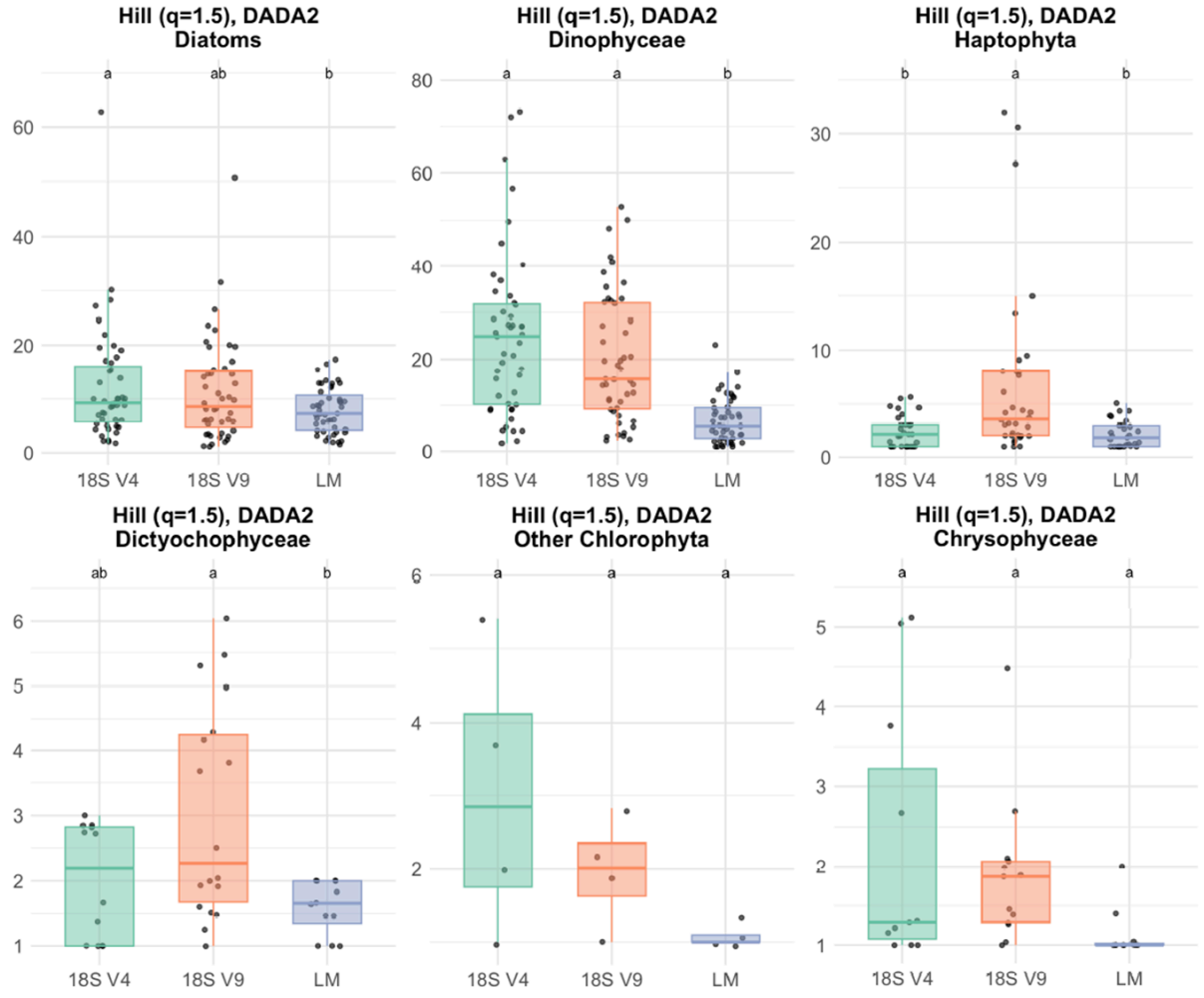


Figure 19. Hill (Q = 1.5) diversity by phytoplankton group, processed by DADA2. Letters indicate groups from ANOVA and $p < 0.05$ post-hoc Tukey HSD tests.

Some exceptions were observed: no significant difference in diversity was apparent across most methods for the 'Other Chlorophyta' and Chrysophyceae. Differences in diversity from marker choice (V4 or V9 region of 18S) were evident but dependent on the phytoplankton group and pipeline method employed. Diversity values across eDNA methods were most consistent for the Diatoms and Dinophyceae, with the only significant difference being higher diversity in Richness of V9 data compared to V4 data resulting from the 'Swarm' method. In comparison, Haptophyta and Dictyochophyceae diversity was significantly higher in V9 data compared to V4 across Shannon-, Richness, and Hill (Q=1.5) with DADA2 processing (less evident in the Swarm pipeline).

- Summary of congruence in diversity values between different metabarcoding (eDNA) methods and microscopy data in the *Tara* dataset:

- Total phytoplankton community richness was significantly lower from microscopy data than from eDNA; the latter did not differ significantly across the different eDNA methods (i.e. types of marker and pipelines) applied.
- For (alpha) diversity metrics that include abundance/evenness, differences in total phytoplankton diversity were both method- and metric dependent. Shannon and Inverse Simpson were congruent between microscopy and eDNA methods, whereas Pielou's Evenness and Hill (q = 1.5) diversity values were greater in the eDNA data generated by some eDNA approaches than in the microscopy data.
- Phytoplankton community (beta) diversity comparisons indicated significant differences in dissimilarity values obtained from microscopy data compared to those from eDNA data, across all eDNA methods and diversity metrics.
- Method comparisons across major phytoplankton groups (Diatoms, Dinophyceae, Haptophyta, Chrysophyceae, Dictyochophyceae, and 'Other Chlorophyta') indicated that for the majority of phytoplankton groups across most diversity metrics, alpha diversity from microscopy was lower than from either one or both of the eDNA markers employed. However both method dependent differences and group-specific consistency in alpha diversity values were observed.

5.1. Phytoplankton diversity eDNA-microscopy comparison from a long-term temporal dataset: coastal station 'L4'

For combined phytoplankton taxa at L4, all alpha diversity metrics (Shannon Index; richness; Hill Numbers (q = 1.5) had significantly higher values for eDNA than microscopy, and the difference was of a 'large magnitude' (Wilcoxon test with Cliff's delta; Figure 20; Appendix VI). Rarefaction of the eDNA decreased variance in eDNA richness, but did not change the direction or magnitude of these differences observed (Appendix VI).

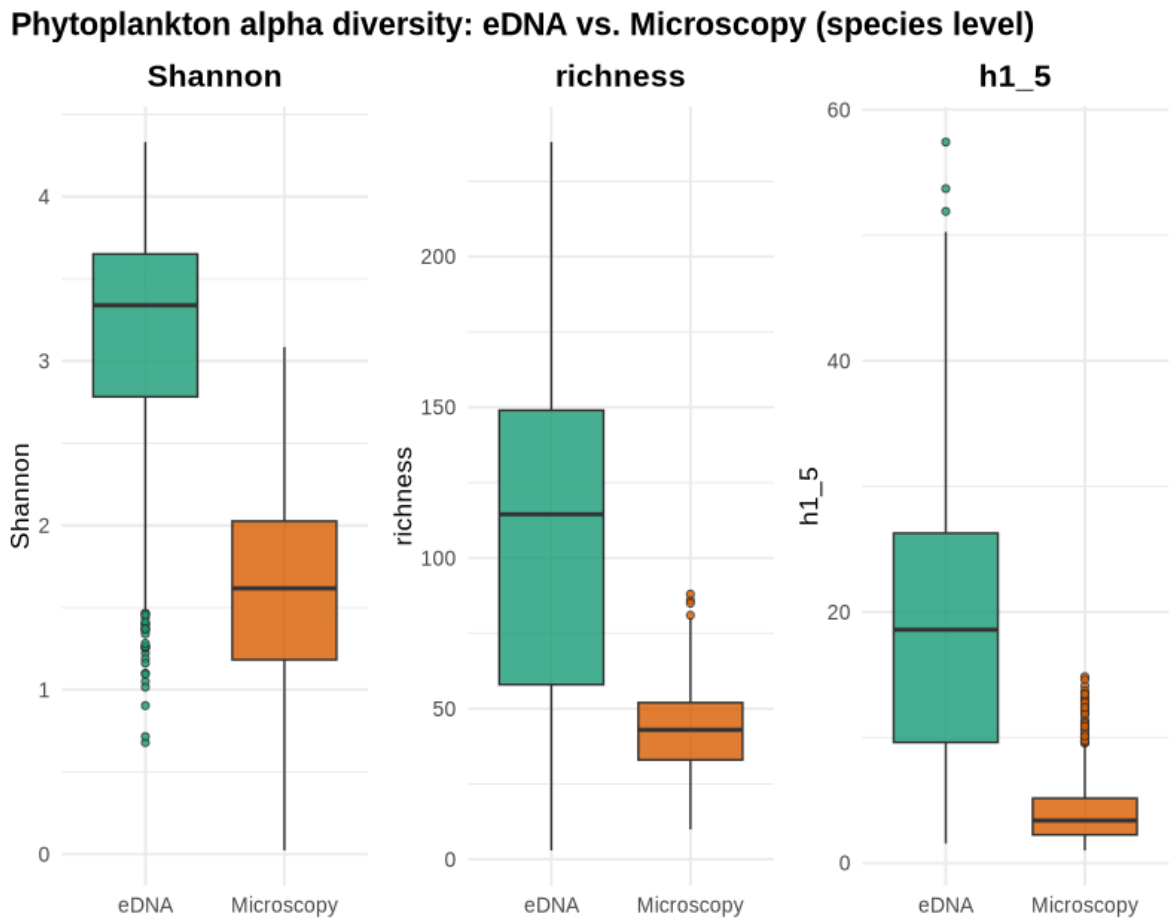


Figure 20. Boxplot comparisons of coastal station 'L4' total phytoplankton alpha diversity metrics from eDNA (metabarcoding) and microscopy data.

Difference in community (beta: dissimilarity) diversity values from phytoplankton eDNA (metabarcoding) and microscopy data at coastal station 'L4' were 'negligible' (Wilcoxon test with Cliff's delta) for both Bray-Curtis (from Hellinger transformed abundance) and Jaccard (presence/absence) dissimilarity metrics (Figure 21; Appendix VI). Rarefaction of the eDNA data caused a minor alteration in Jaccard dissimilarity only, eDNA data having higher dissimilarity than microscopy data of a 'small' magnitude (Wilcoxon test with Cliff's delta; Appendix VI)

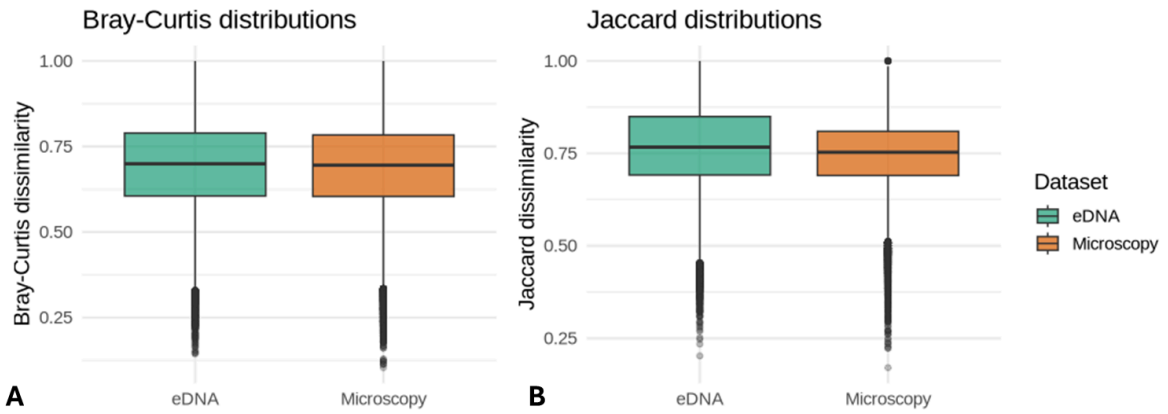


Figure 21. Comparisons of coastal station 'L4' total phytoplankton beta diversity metric values from eDNA (metabarcoding) and microscopy data: A. Bray-Curtis dissimilarity of Hellinger transformed abundance; B. Jaccard dissimilarity.

For three major groups of phytoplankton - the diatoms, Dinophyceae (dinoflagellates) and Haptophyta (coccolithophores) - all diversity metric values generated were higher in eDNA data than in microscopy data at coastal station L4 (Wilcoxon test) (Figure 22). However, the magnitude of difference was dependent on the group: Diatoms: small magnitude across metrics; Dinoflagellates: small magnitude of difference for Shannon and Hill (Q = 1.5), medium difference in richness; coccolithophores: large magnitude of difference across metrics.

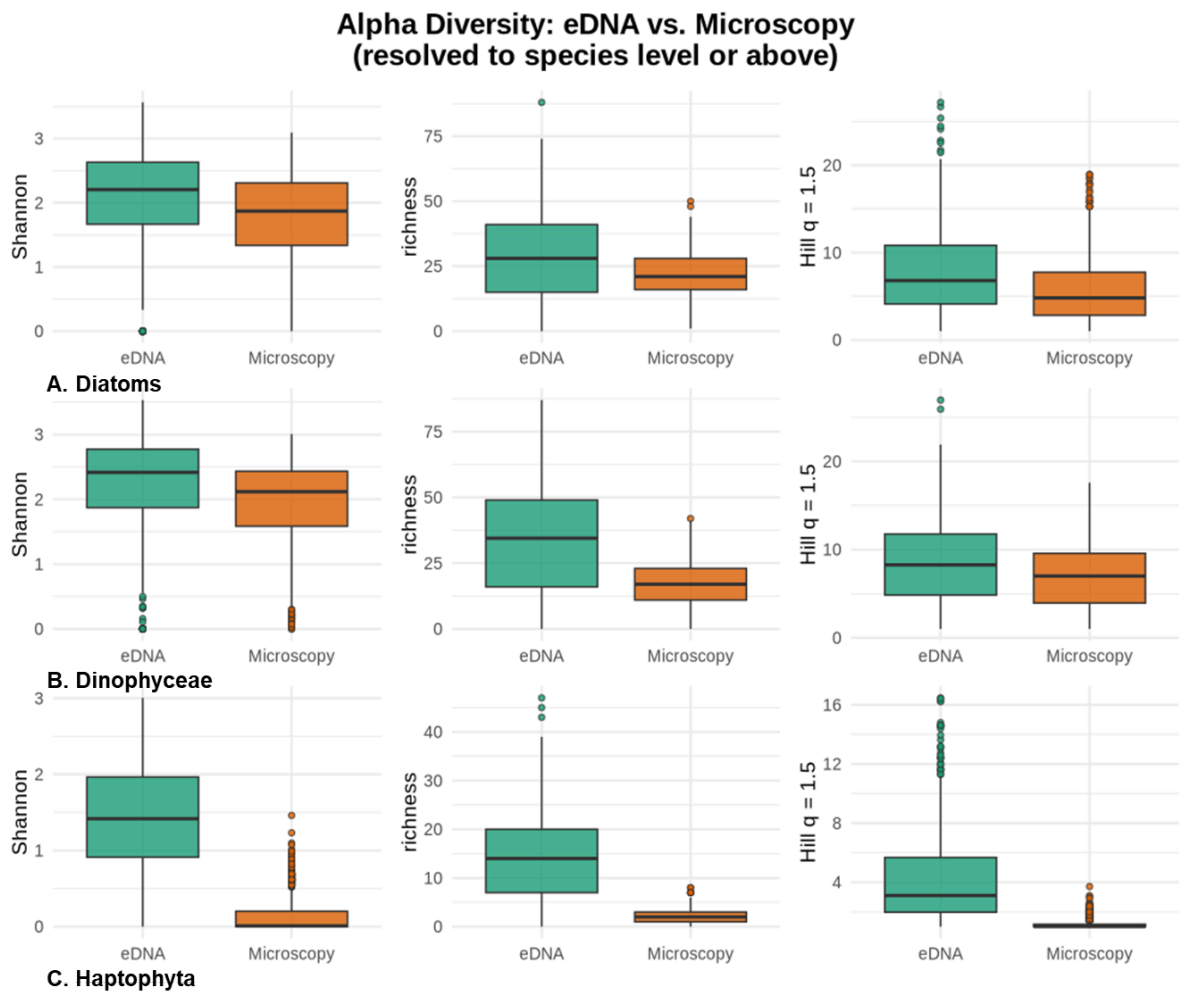


Figure 22. Comparisons of alpha diversity values from eDNA (metabarcoding) and microscopy datasets for the major phytoplankton groups: A. Diatoms; B. Dinophyceae (dinoflagellates); C. Haptophyta (coccolithophores).

- Summary of congruence in diversity values between different metabarcoding (eDNA) methods and microscopy data in the coastal station L4 dataset:
    - When considered as a whole, the congruence in 'L4' phytoplankton diversity values derived from eDNA and microscopy data was dependent on diversity type: alpha diversity was higher in eDNA data, and the difference was large, whereas differences in beta diversity were negligible.
    - When major phytoplankton groups were tested separately, differences in alpha diversity values from eDNA and microscopy data was group dependent; although eDNA data always produced higher diversity values, the difference between datatypes was smaller for diatoms and dinoflagellates, than for coccolithophores.

6. Results of the Bioinformatic Data Challenge: How different eDNA pipelines have affected eDNA-microscopy comparisons

The pipelines have produced seasonal patterns in alpha diversity that vary in degree of similarity to microscopy data (Figure 23). For Shannon diversity, three pipelines which utilised the 'VSEARCH' pipeline (Rognes *et al.*, 2016) produced seasonal signals that were more reduced compared to the microscopy dataset: the two pipelines based on APSCALE, which are most similar to the microscopy data in terms of Shannon magnitude, and the SLIM pipeline. Although no single pipeline appears fully congruent with the microscopy data for which months have minimum- and maximum Shannon diversity (in microscopy data, May and September, respectively), several pipelines are close, including 'Casabel' (min: May; max: October), SAMBA (min: May-June; max: November) and QIIME2 (min: May; max: October), all three of which use elements of the DADA2- (Callahan *et al.* 2016) and QIIME (Caporaso *et al*. 2010) pipelines. The majority of pipelines have produced higher magnitudes of Shannon diversity than captured in the microscopy data

As observed in the analyses presented above, calculation of richness produced starkly different seasonal patterns in the microscopy data than those calculated via Shannon diversity (Figure 24): seasonal signals in the richness of the microscopy data from the ASTAN timeseries were negligible, and the magnitude of richness in microscopy data was less than any of the eDNA datasets. Both APSCALE pipelines also produced negligible differences in richness across seasons. The FROGS 'Swarm' pipeline, which clusters taxa into OTUs, also showed little in the way of season-related variability in richness. However the majority of the eDNA pipelines retained a similar seasonal signal to that observed in Shannon diversity, with richness minimums in June, and maximums in October-November. Interestingly, seasonal differences in the SLIM pipeline were stronger for richness than for Shannon. Lastly, seasonal differences in Hill Numbers (q=1.5) diversity have a mixture of the characteristics noted above (Figure 25). Seasonal signals were modest in the microscopy data, with minimum and maximum values in the same months as observed for Shannon. The APSCALE eDNA pipelines were similar in magnitude to the microscopy data, but with negligible differences in richness across seasons. Meanwhile, the'Casabel', SAMBA and QIIME2 pipelines Hill numbers (q=1.5) diversity were more different across months than observed for Shannon, with steep increases after the May minimum to the October-November maximum. Here, the FROGS Swarm (clustering) and FROGS DADA2 (ASVs) pipelines 'bottom out' with little change in diversity through the April-August period.

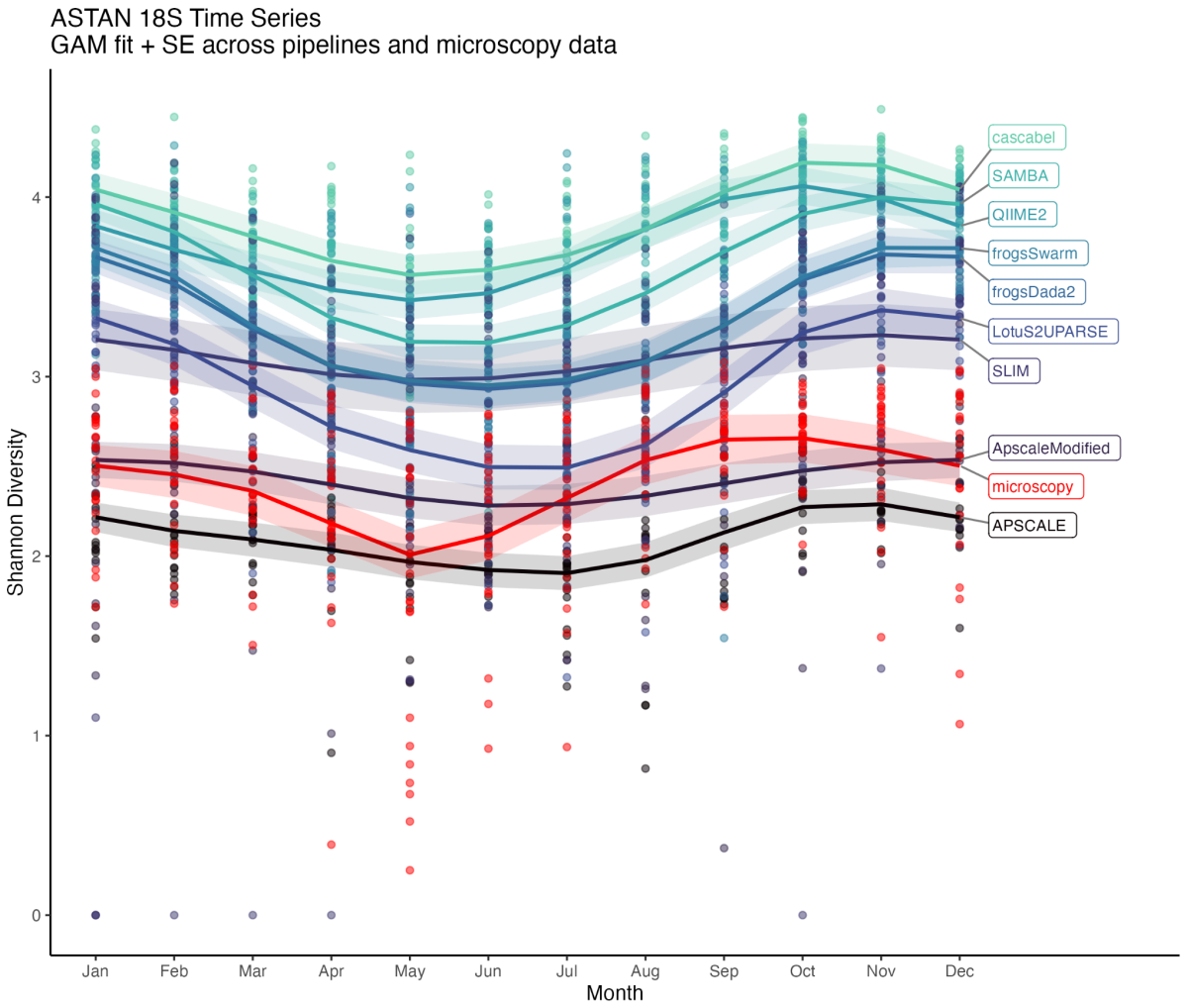


Figure 23. A Generalized Additive Model (GAM) of intra-annual Shannon diversity of ASTAN plankton over the course of 8 years (2009-2016; bimonthly sampling), showing microcopy data and eDNA (18S-V4 metabarcoding) data processed through a variety of bioinformatic pipelines.

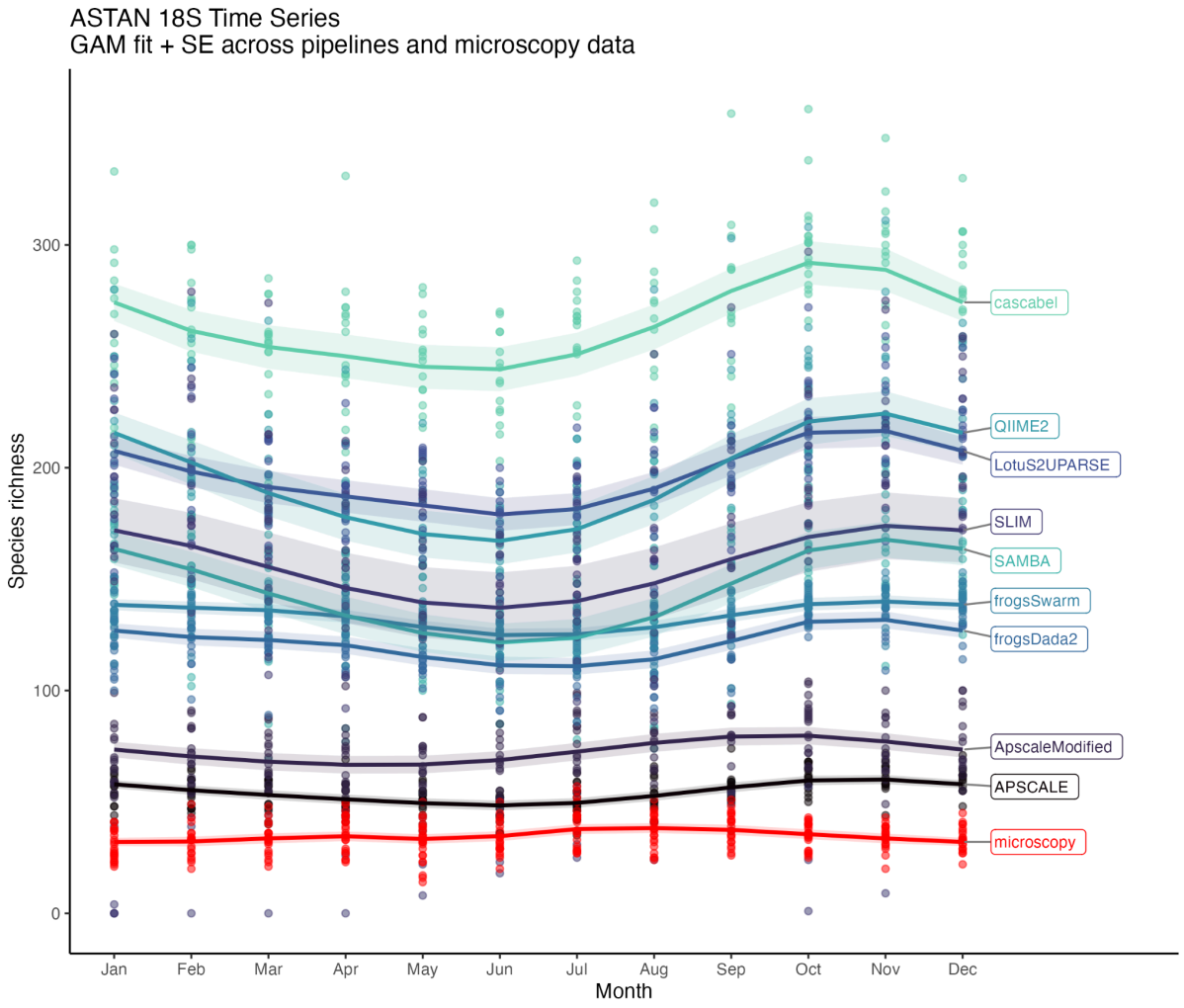


Figure 24. A Generalized Additive Model (GAM) of intra-annual richness of ASTAN plankton over the course of 8 years (2009-2016; bimonthly sampling), showing microcopy data and eDNA (18S-V4 metabarcoding) data processed through a variety of bioinformatic pipelines.

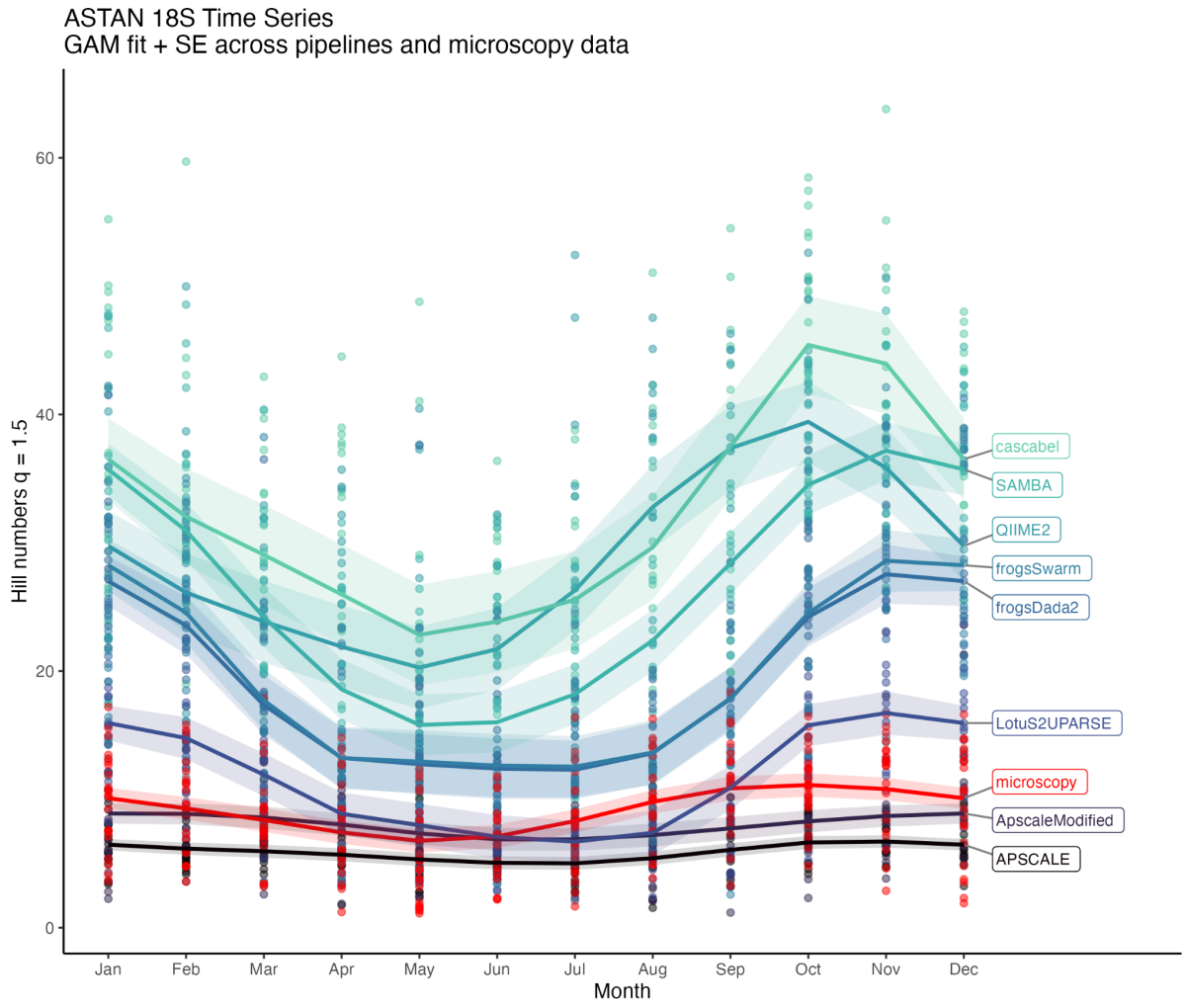


Figure 25. A Generalized Additive Model (GAM) of intra-annual Hill Numbers (q=1.5) diversity of ASTAN plankton over the course of 8 years (2009-2016; bimonthly sampling), showing microcopy data and eDNA (18S-V4 metabarcoding) data processed through a variety of bioinformatic pipelines (SLIM pipeline not shown: SLIM Hill numbers (q=1.5) could not be modelled via GAM).

The plankton community structure (i.e. beta diversity) was heavily influenced by both data type (i.e. microscopy; eDNA) and choice of pipeline (Figure 26 and 27). For both beta diversity metrics (Bray-Curtis and Jaccard), the samples clustered into 3 main groups: microscopy samples, eDNA Samples processed by both of the APSCALE-based pipelines, and those processed by alternative approaches (Figures 26 and 27).

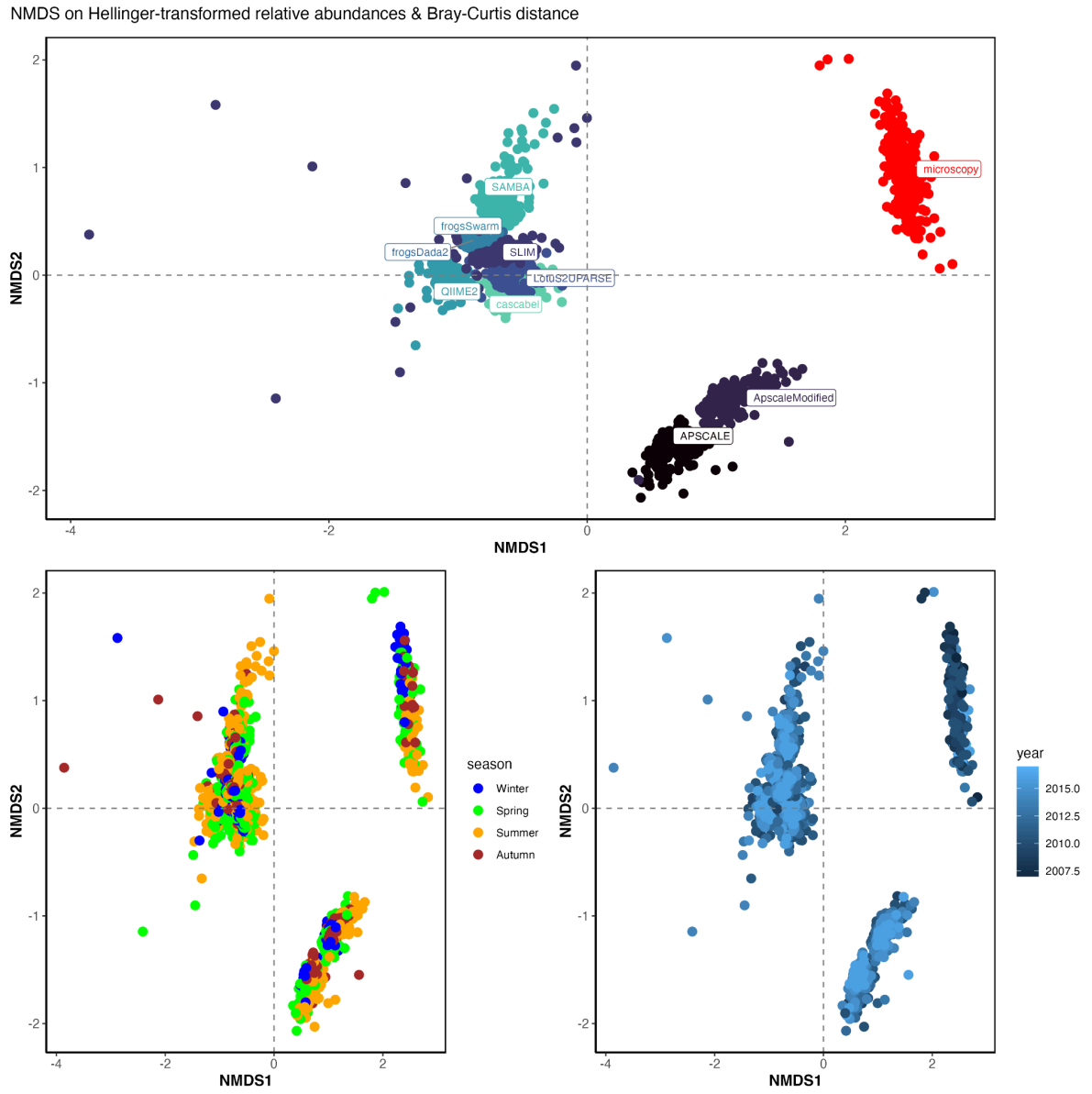


Figure 26. Non-metric multidimensional scaling (NMDS) of the relative abundances of protist taxa detected through the different participating pipeline entries, analysed together with the taxa identified through morphology/microscopy. The NMDS stress = 0.12. Relative abundances of taxa first underwent a Hellinger-transformation before computing the NMDS with Bray-Curtis distance.

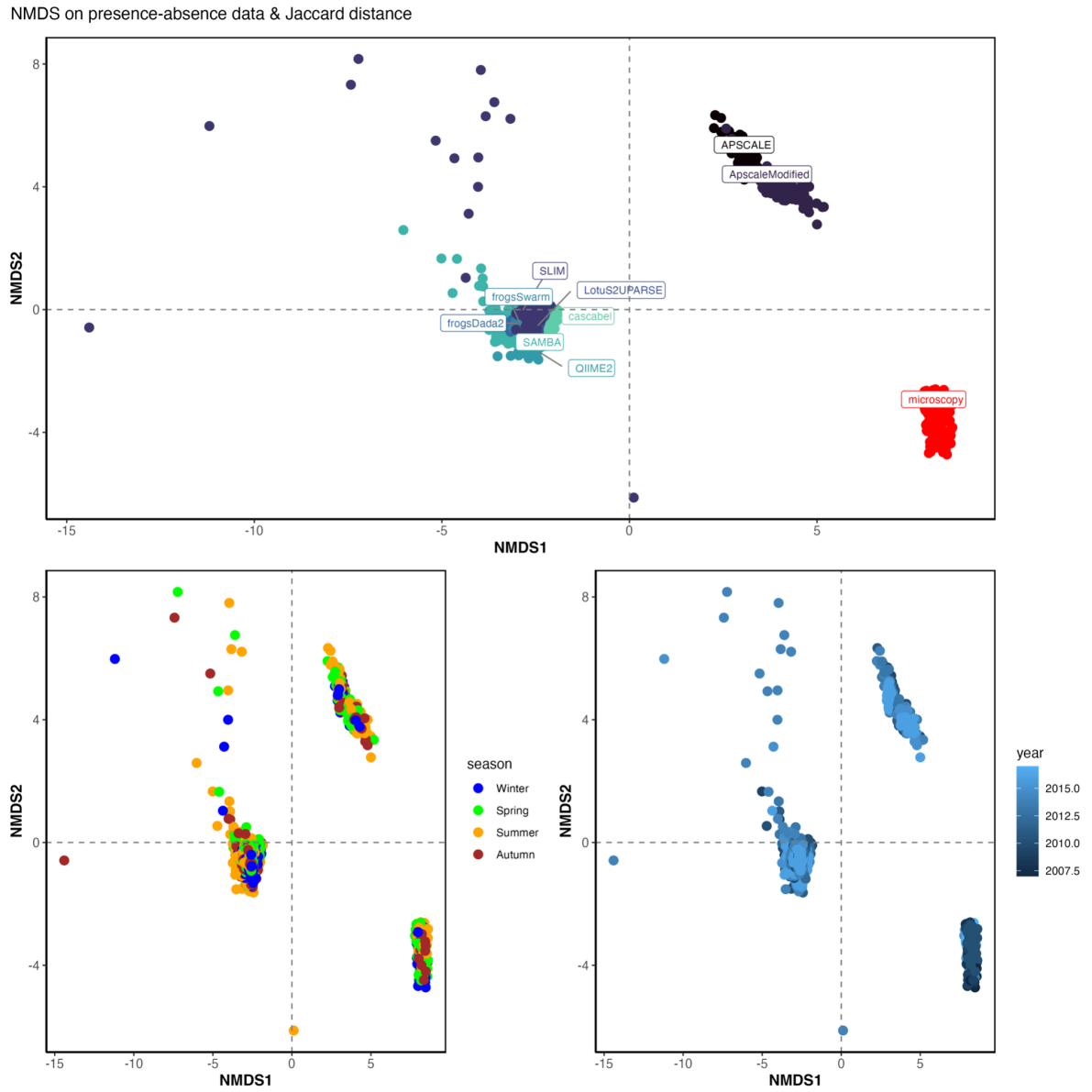


Figure 27. Non-metric multidimensional scaling (NMDS) of the presence-absence of protist taxa detected through the different participating pipeline entries, analysed together with the taxa identified through morphology/microscopy. The NMDS stress = 0.07 and is computed with the Jaccard distance.

Examination of the seasonal patterns in community abundance (Bray-Curtis) dissimilarity produced by the microscopy data and each of the pipelines separately reveals high levels of eDNA-microscopy congruence in the across-year similarity of samples per season (Figure 28). Winter samples were more similar across the time series than spring samples in both the microscopy data and all eDNA datasets except one: the signal is less clear in the SLIM pipeline. This same pattern and congruence was also observed in the L4 time series. When calculated by presence-absense (i.e.

Jaccard dissimilarity), the congruence in seasonal patterns of community diversity between the microscopy data and different eDNA pipeline datasets was less clear, with different pipelines pointing to different seasons being the least- and most similar across the time series (figure 29).

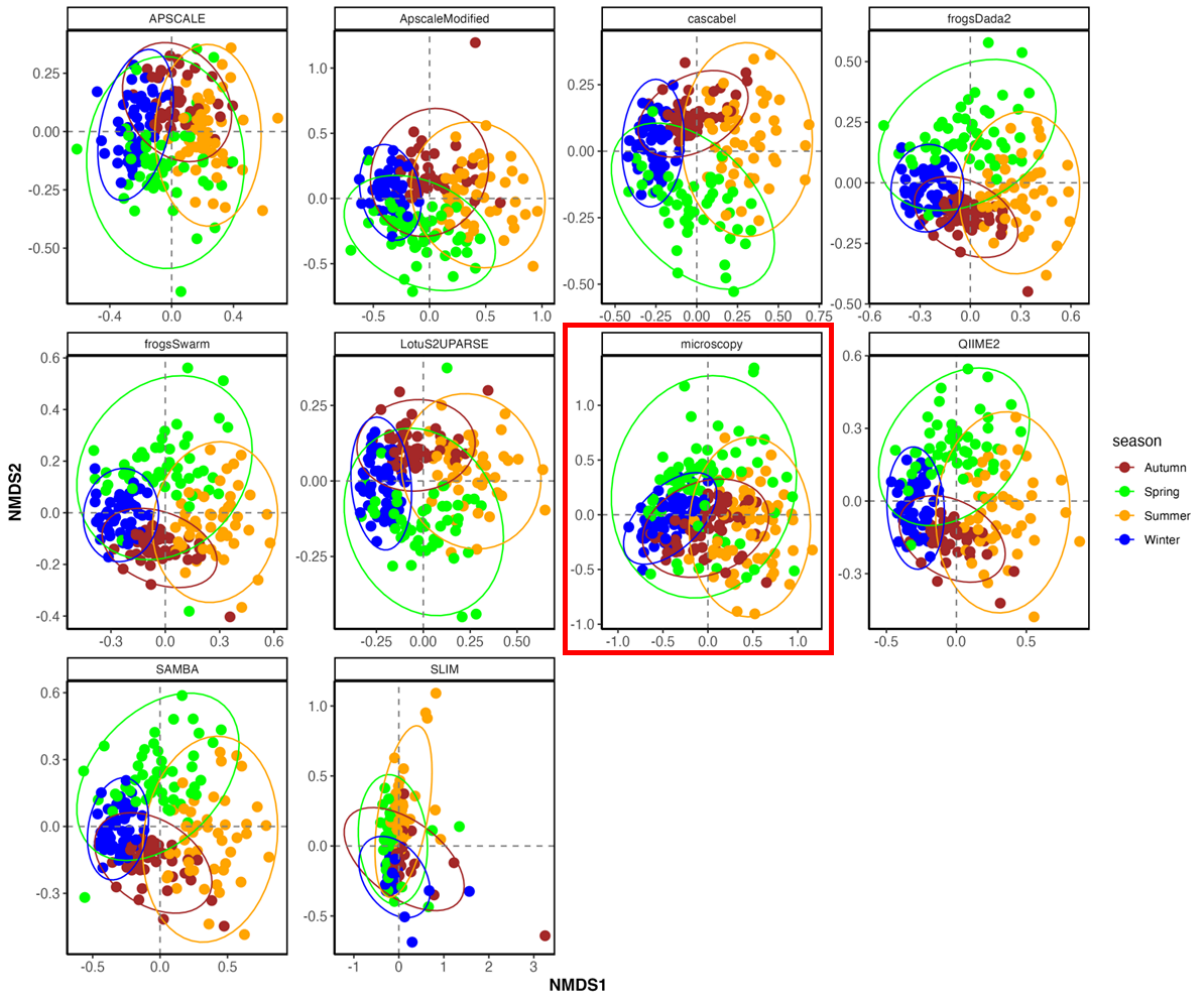


Figure 28. Non-metric multidimensional scaling (NMDS) of the relative abundances of protist taxa detected through the different participating pipeline entries, analysed together with the taxa identified through morphological/microscopy. Relative abundances of taxa first underwent a Hellinger-transformation before computing the NMDS with Bray-Curtis distance. Points are coloured by season, and the microscopy dataset is highlighted with a red square.

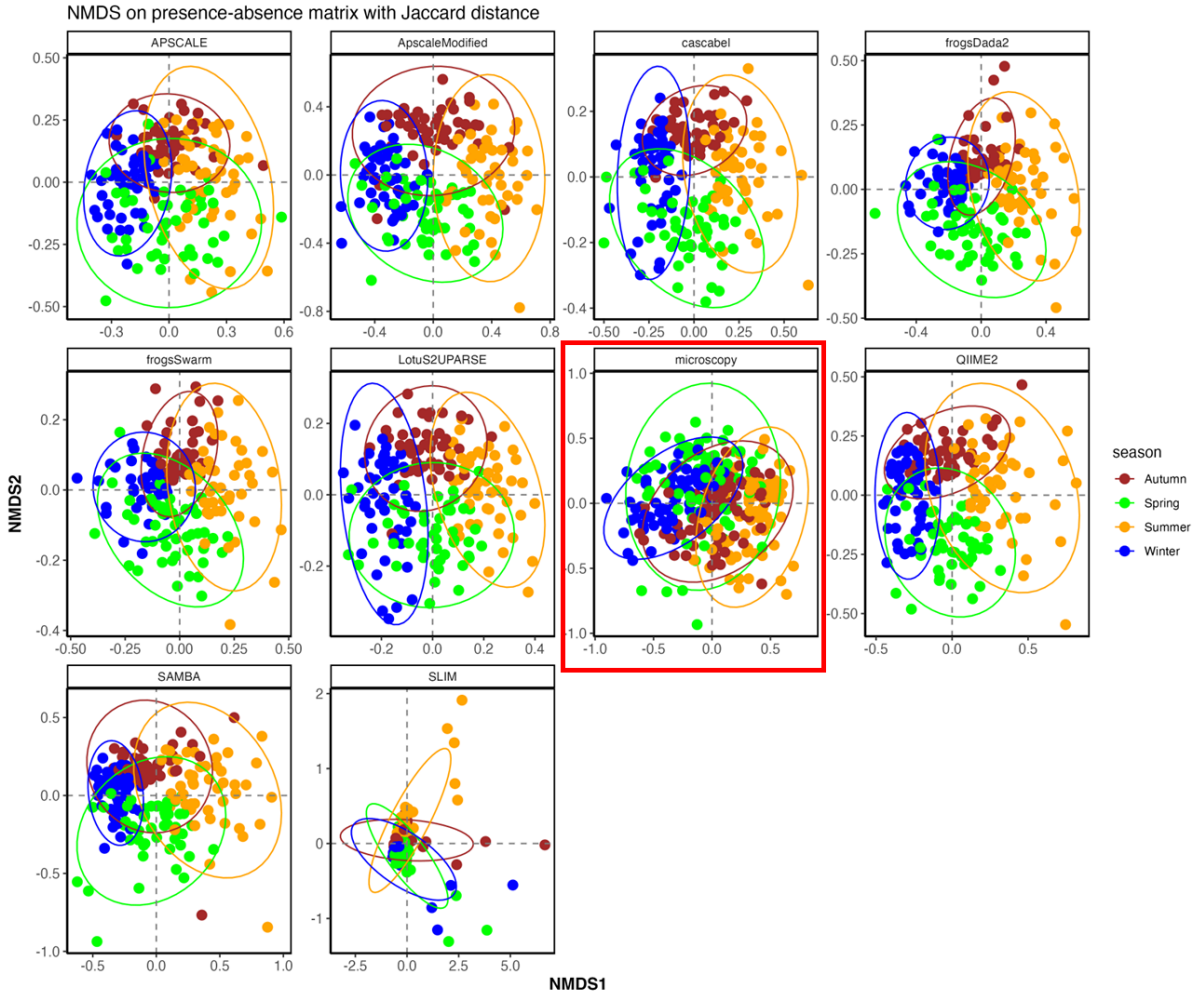


Figure 29. Non-metric multidimensional scaling (NMDS) of the presence-absense of protist taxa detected through the different participating pipeline entries, analysed together with the taxa identified through morphological/microscopy. NMDS was computed with Jaccard distance. Points are coloured by season, and the microscopy dataset is highlighted with a red square.

A similar examination of the community dissimilarity produced by the microscopy data and each of the pipelines separately and with consideration of sample year does not suggest agreement between data types (Figure 30, 31). In both abundance and presence-absence analyses, the microscopy data shows that samples taken early in the time series are somewhat dissimilar to those taken later in the time series, although there is a high degree of overlap. However, this pattern does not appear in any of the datasets from the eDNA pipelines. In addition, the SLIM pipeline has a number of outlier samples, mostly from the year 2014, which are not apparent in the microscopy data.

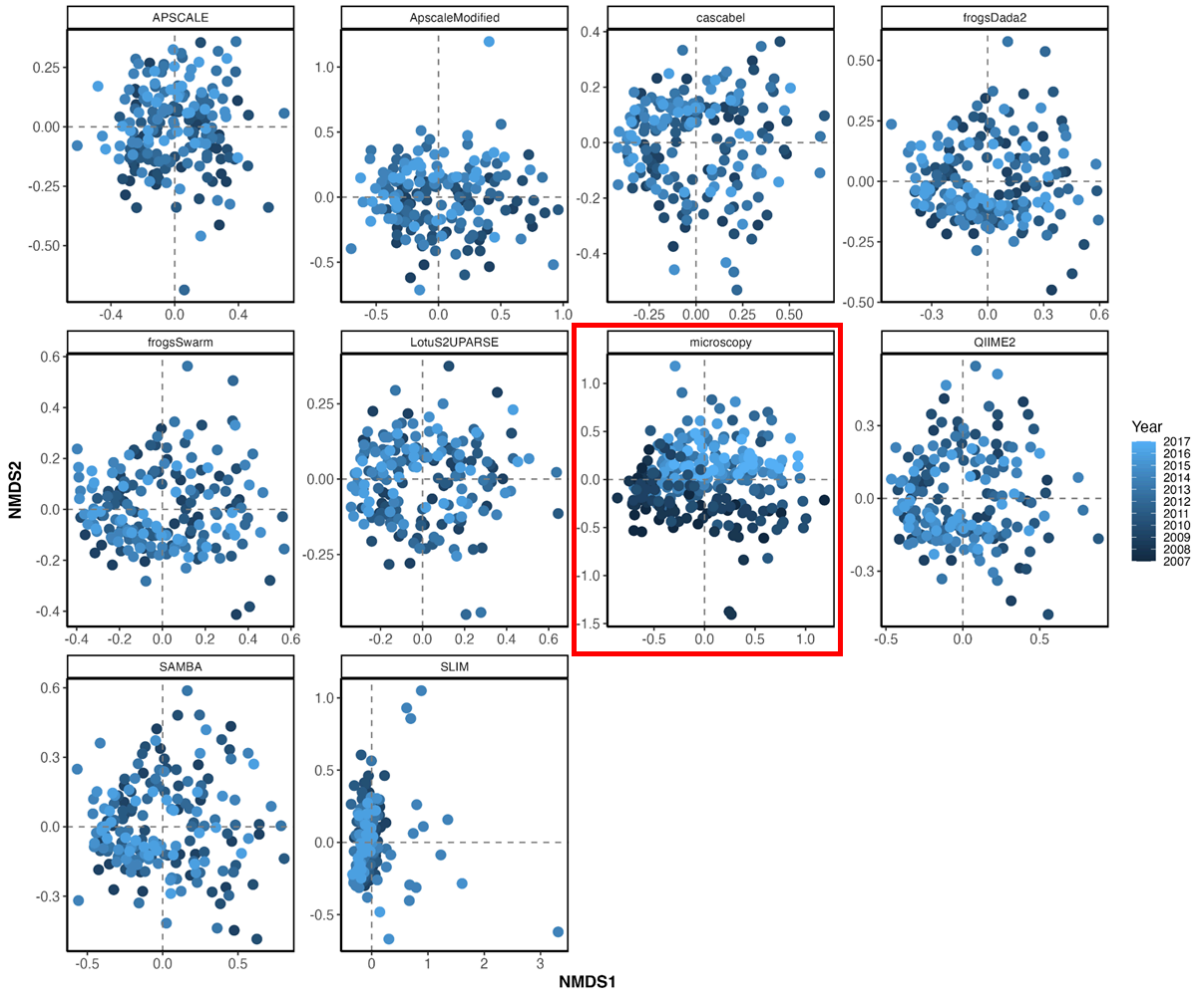


Figure 30. Non-metric multidimensional scaling (NMDS) of the presence-absence of protist taxa detected through the different participating pipeline entries, analysed together with the taxa identified through morphological/microscopy. NMDS was computed with Jaccard distance. Points are coloured by year,and the microscopy dataset is highlighted with a red square.

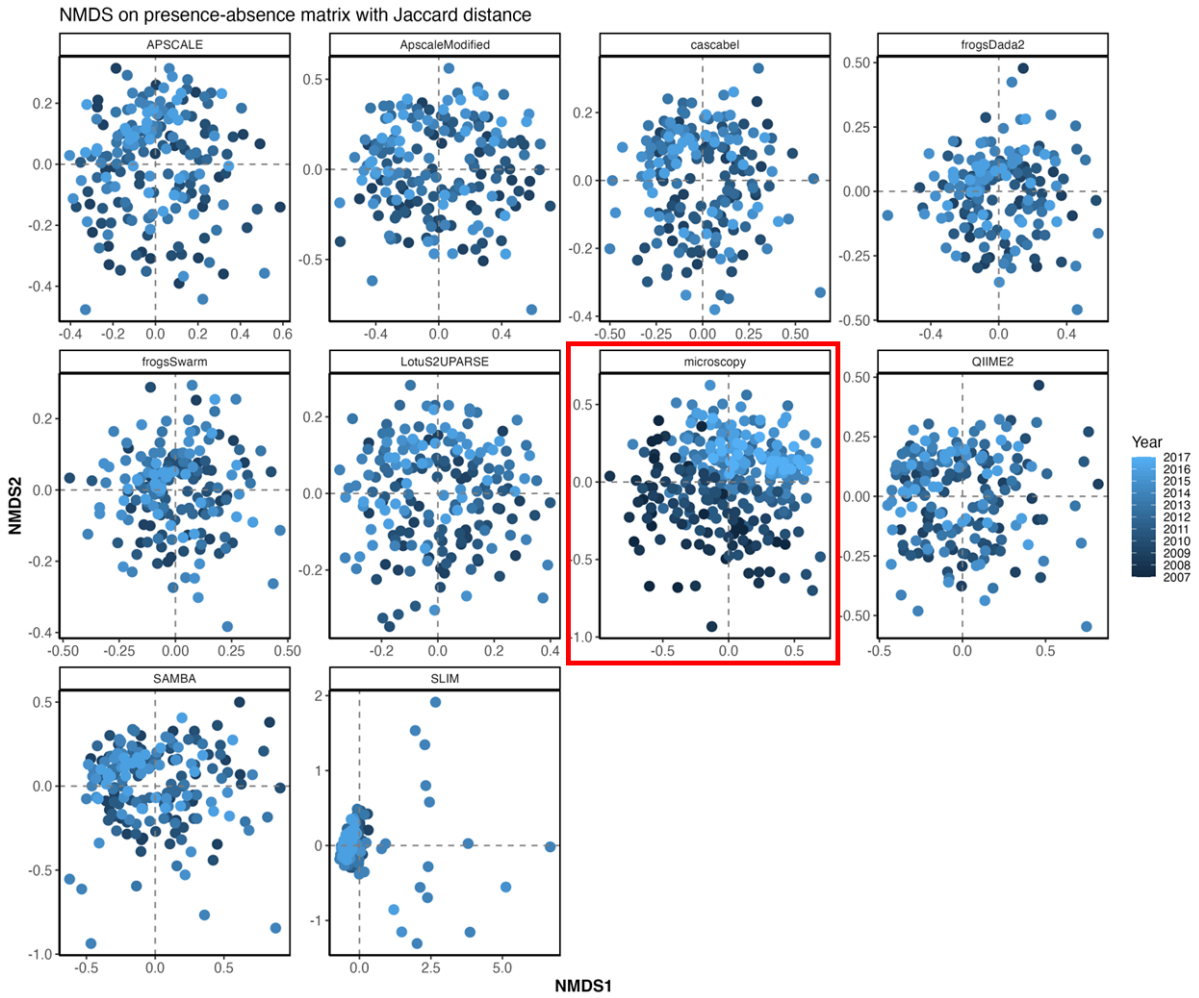


Figure 30. Non-metric multidimensional scaling (NMDS) of the presence-absense of protist taxa detected through the different participating pipeline entries, analysed together with the taxa identified through morphology/microscopy. NMDS was computed with Jaccard distance. Points are coloured by year, and the microscopy dataset is highlighted with a red square.

- Summary of 'Data Analysis Challenge' results: congruency between microscopy and eDNA diversity from ASTAN plankton time series and the effect of pipeline choice:
    - Different eDNA pipelines produced different magnitudes of alpha diversity, and were important in the strength and shape of seasonal patterns in alpha diversity
    - Community (Beta) diversity in microscopy data was strongly dissimilar to that observed in eDNA data when values were compared simultaneously.
    - Pipeline choice was more important than season or year to the community structure of the eDNA samples; those which employed V-SEARCH pipeline were most unlike the results from alternative pipelines

- When beta diversity patterns were calculated per data type, most eDNA pipelines were highly congruent in long-term, seasonal patterns (with winter samples being more alike across years than spring samples in microscopy data and in seven of the eight eDNA pipelines), but agreement in long-term differences in community diversity between microscopy and eDNA data sets were not evident

7. Possible mechanisms and Implications of eDNA-traditional data differences

The results of the analyses presented here indicate that the degree of congruence between eDNA (metabarcoding) datasets and traditional (microscopy) data is most dependent on the diversity metric applied. However, both the eDNA methods employed in data production and processing, and the nature of the community being observed (e.g. plankton/phytoplankton; different phytoplankton groups), appeared to further impact the degree of similarity in the diversity patterns observed between data types. Largely, estimations of diversity that account for differences in abundance between taxa (i.e. the 'evenness' of a community) produced the most congruent patterns of diversity between microscopy and metabarcoding data, as observed in both intra-annual (alpha- and beta) diversity of the plankton community in the coastal station 'L4' time series (figure 3A, 6, 7, and 8), the total phytoplankton alpha diversity values of *Tara Oceans* dataset (Figure 15 and 16), and the eDNA-microscopy congruence in the across-year similarity of samples per season from the ASTAN times series (Figure 28). However, the magnitude of the diversity values generated by such estimators (i.e. Shannon; Hill numbers: q = 1.5) was different between the data types for two of the three datasets investigated (Figure 3, 20, 21, 23, 25), and only the total phytoplankton diversity values in the *Tara* datasets were consistent between microscopy and the different eDNA approaches applied (Figure 15).

There were other differences in the diversity patterns of traditional and eDNA data, as produced by estimators that accounted for evenness. Although some seasonal signals were congruent between data types in the 'L4' time series for Shannon and Hill numbers (q = 1.5) (e.g. samples from winter being most diverse and most similar across years; spring samples being the least diverse; Figure 6 and 7), others were divergent between data types (e.g. which season was least similar across years; the presence/timing of a diversity 'shift' in long term data; Figure 6, 7, 12, 13 and 14). In addition, designation of taxa into major plankton/phytoplankton groups revealed group-dependent differences in diversity patterns between the data types, both in intra-annual relative abundances (Figure 5), and in total diversity values (Figure 20), the latter also evident in the *Tara* data (figure 17-19). Diversity estimators that considered only presence-absence (i.e. 'richness'; Jaccard dissimilarity) produced largely divergent pictures of diversity from the traditional microscopy

and eDNA dataset. Exceptions were limited to identification of a matching 'changepoint' year in 'richness' (Figure 14), negligible difference in total phytoplankton Jaccard dissimilarity values in the L4 time series (Figure 12), and congruent patterns of dissimilarity in microscopy data compared to eDNA data for Jaccard- and Bray-Curtis values across the 'Data Analysis Challenge' (Figure 26 and 27).

The majority of the differences observed in diversity patterns between data types are likely due to limitations/strengths of each data type and each approach undertaken. Here, the 'L4' eDNA time series was processed in two different ways: 1). Inclusion of the maximum number of taxa possible, comprising all the major plankton groups identified by the complementary microscopy dataset, facilitating identification of seasonal and long-term patterns of diversity; 2) Filtration of phytoplankton groups only, summing of abundance at species level, and calculation of comparative diversity values, as performed for the *Tara* datasets. For the former, the analysis of diversity patterns, the maximum diversity capturing potential for each data type was preserved: intraspecific variation in the form of ASVs in the eDNA data, and size-based differentiation in the microscopy taxa, were retained. For the latter, the groups whose diversity was most likely to align between data types were preserved in order to maximise potential congruence. Yet the similarities in phytoplankton diversity between the data types was limited, and most congruency was observed in the seasonal diversity patterns of plankton.

The congruencies between diversity patterns from the data types identified are interesting because eDNA and microscopy data sets have innate differences. One major limitation of light microscopy being that it cannot be used to identify or resolve the taxonomy of pico- and nano eukaryotes. For example, coccolithophores without liths are unlikely to be preserved (and thus observed) by standard microscopy approaches, and differentiation of flagellates is severely limited. These factors are likely to have influenced the magnitude of diversity observed in the different datatypes, differences in the relative abundance of major plankton groups, and the differences in group-specific phytoplankton diversity values (e.g. diatoms had the highest level similarity between data types; coccolithophores had low levels of diversity in the microscopy data compared to eDNA data). However, identification and taxonomic resolution via metabarcoding is also limited and influenced by 'marker' choice and by incomplete reference data bases (Holland *et al*, 2025). The different diversity values for *Tara* phytoplankton obtained for the different groups by the different markers illustrate this issue (Figure 17-19; Appendix V). In addition, different eDNA processing pipelines deal with intraspecific variation (i.e. ASV-level variation) and very rare taxa in a variety of ways, some of which may work to inflate or flatten diversity, likely accounting for some of the differences in seasonal patterns of alpha diversity and community diversity between eDNA pipelines observed in the Data Analysis Challenge results (Figure 23-25).

Another potential confounder for traditional-eDNA congruence is how relative abundance values for metabarcoding data are produced. Metabarcoding based relative abundance values are known to be biased by a number of factors, including the number of copies of the marker genes encoded by different taxa, which can inflate relative abundance in taxa which encode high numbers

of the marker gene (Yates *et al*, 2019). Dinoflagellates have high 18S rRNA gene copy numbers, and significantly higher dinoflagellate diversity values were observed in both the 'L4' and *Tara* eDNA dataset compared to the respective microscopy data (although the difference was of smaller magnitude for dinoflagellates than for coccolithophores in the 'L4' data). Sampling effort may also have produced some of the differences in diversity from the different data types observed. Standard microscopy methods count cells in 50 mL of water, whereas 1-2 L of water were filtered for eDNA collection. There are also known issues with both data types in capturing the presence of rare taxa, especially during plankton 'blooms'. The sampling effort for both methods is affected by the increased cell density/biomass of such events. When cell density reaches a threshold, standard microscopy methods include taking transects of the slides rather than counting whole samples, meaning that rare cells are more likely to be missed. Similarly, when eDNA samples are dominated by particular taxa, these will represent the majority of the sequences produced; very rare taxa may fall below level of detection, or may be removed as part of the quality control stages of the pipeline (to account for possible sequencing errors).

In conclusion, despite the innate differences in between data types and myriad sources of potential bias against consistency, there was some - though not full - congruence in the across-year seasonal diversity patterns and total diversity values observed in eDNA and traditional 'microscopy' data from temporal and spatial investigations, respectively. Due to the limitations of eDNA methods, there have been calls for eDNA-based approaches to be applied to presence-absense surveys only (Holland *et al*, 2025). However, the analyses presented here indicate higher congruence between data types in diversity estimates that account for the evenness of the communities in question (e.g. Shannon diversity), and show that the application of different metabarcoding markers and data processing methods are likely to affect the degree of congruence to traditional data.

8. Acknowledgements

Dr Claire Widdicombe (Plymouth Marine Laboratory) for Western Channel Observatory coastal station 'L4' microscopy data and associated metadata. The Western Channel Observatory is funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Atlantic Climate and Environment Strategic Science (AtlantiS), grant number NE/Y005589/1.

Dr Nicolas Henry (Station Biologique de Roscoff; CNRS; Sorbonne Université) for assistance with retrieval and use of the Roscoff-ASTAN plankton time series data sets.

Crews of the research vessels ‘Quest’ (operated by Plymouth Marine Laboratory) and ‘Sepia’ (operated by the Marine Biological Association), for collection of the coastal station ‘L4’ samples and metadata.

9. References

Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., Albini, G., Aury, J.-M., Belser, C., Bertrand, A., Cruaud, C., Da Silva, C., Dossat, C., Gavory, F., Gas, S., Guy, J., Haquelle, M., Jacoby, E. 'krame, Jaillon, O., … Wincker, P. (2020). 18S and 16S rRNA genes amplicon generation for eukaryotic and prokaryotic metabarcoding. https://www.protocols.io/view/18s-and-16s-rrna-genes-amplicon-generation-for-euk-qwhdxb6

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. http://doi.org/10.1371/journal.pone.0006372

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

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

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6, 1621–1624. http://doi.org/10.1038/ismej.2012.8

Caracciolo M, Rigaut-Jalabert F, Romac S, Mahé F, Forsans S, Gac JP, Arsenieff L, Manno M, Chaffron S, Cariou T, Hoebeke M, Bozec Y, Goberville E, Le Gall F, Guilloux L, Baudoux AC, de Vargas C, Not F, Thiébaut E, Henry N, Simon N. (2022). Seasonal dynamics of marine protist communities in tidally mixed coastal waters. Mol Ecol. 2022 Jul;31(14):3761-3783. doi: 10.1111/mec.16539. Epub 2022 Jun 16. PMID: 35593305; PMCID: PMC9543310.

da Veiga Leprevost, F., Grüning, B. A., Alves Aflitos, S., Röst, H. L., Uszkoreit, J., Barsnes, H., Vaudel, M., Moreno, P., Gatto, L., Weber, J., Bai, M., Jimenez, R. C., Sachsenberg, T., Pfeuffer, J., Vera Alvarez, R., Griss, J., Nesvizhskii, A. I., & Perez-Riverol, Y. (2017). BioContainers: An open-source and community-driven framework for software standardization. Bioinformatics, 33(16), 2580–2582. https://doi.org/10.1093/bioinformatics/btx192

Delage, E., Henry, N., Salazar, G., Ruscheweyh, H.-J., Poulain, J., Sunagawa, S., de Vargas, C., & Chaffron, S. (2024a). Tara Oceans (2009-2013) rDNA 18S V4 ASV table (dada2) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.13881376

Delage, E., Henry, N., Salazar, G., Ruscheweyh, H.-J., Poulain, J., Sunagawa, S., de Vargas, C., & Chaffron, S. (2024b). Tara Oceans (2009-2013) rDNA 18S V9 ASV table (dada2) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.13881418

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., … Zayas, C. N. (n.d.). *REVIEW SUMMARY Pervasive human-driven decline of life on Earth points to the need for transformative change*. https://doi.org/10.1126/science.aaw3100

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ewels, P.A., Peltzer, A., Fillinger, S. et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol 38, 276–278 (2020). https://doi.org/10.1038/s41587-020-0439-x

Grigoratou, M., Menden-Deuer, S., McQuatters-Gollop, A., Arhonditsis, G., Artigas, L. F., Ayata, S. D., Bedikoğlu, D., Beisner, B. E., Chen, B., Davies, C., Diarra, L., Elegbeleye, O. W., Everett, J. D., Garcia, T. M., Gentleman, W. C., Gonçalves, R. J., Guy-Haim, T., Halfter, S., Hinners, J., … Richardson, A. (2025). The immeasurable value of plankton to humanity. BioScience, 75(9), 706–721. https://doi.org/10.1093/biosci/biaf049

Grüning, B., Dale, R., Sjödin, A., Rowe, J., Chapman, B. A., Tomkins-Tinch, C. H., Valieris, R., Batut, B., Caprez, A., Cokelaer, T., Yusuf, D., Beauchamp, K. A., Brinda, K., Wollmann, T., Corguillé, G. le, Ryan, D., Bretaudeau, A., Hoogstrate, Y., Pedersen, B. S., … Köster, J. (2018). Bioconda: Sustainable and comprehensive software distribution for the life sciences. Nature Methods, 15(7), 475–476. https://doi.org/10.1038/s41592-018-0046-7

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, Del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WH, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 2013 Jan;41(Database issue):D597-604. doi: 10.1093/nar/gks1160. Epub 2012 Nov 27. PMID: 23193267; PMCID: PMC3531120.)

Henry, N., & Salazar, G. (2022). Summarized contextual data about metabarcoding Tara Oceans samples (2009-2013) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.7229815.

Holland, M. M., Artigas, L. F., Atkinson, A., Best, M., Bresnan, E., Devlin, M., Eerkes Medrano, D., Johansen, M., Johns, D. G., Machairopoulou, M., Pitois, S., Scott, J., Schilder, J., Stern, R., Tait, K., Whyte, C., Widdicombe, C., & McQuatters-Gollop, A. (2025). Mind the gap - The need to integrate novel plankton methods alongside ongoing long-term monitoring. Ocean & Coastal Management, 262(107542), 107542. https:// doi.org/10.1016/j.ocecoaman.2025.107542

Killick, R. & Eckley, I. (2014). Changepoint: An R package for changepoint analysis. Journal of Statistical Software, 58(3), 1–19.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, 82(13), 1–26. doi:10.18637/jss.v082.i13.

Leadley, P., Gonzalez, A., Obura, D., Krug, C. B., Londoño-Murcia, M. C., Millette, K. L., Radulovici, A., Rankovic, A., Shannon, L. J., Archer, E., Armah, F. A., Bax, N., Chaudhari, K., Costello, M. J., Dávalos, L. M., Roque, F. de O., DeClerck, F., Dee, L. E., Essl, F., … Xu, J. (2022). Achieving global biodiversity goals by 2050 requires urgent and integrated actions. In One Earth (Vol. 5, Issue 6, pp. 597–603). Cell Press. https://doi.org/10.1016/j.oneear.2022.05.009

Lenth, R. V., & Piaskowski, J. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 2.0.0. https://CRAN.R-project.org/package=emmeans.

Li, D. (2018). "hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers." Journal of Open Source Software, 3(31), 1041. https://doi.org/10.21105/joss.01041.

Mahé, F., Henry, N., de Vargas, C., Tara Oceans Consortium, Coordinators, & Tara Oceans Expedition, Participants. (2022a). rDNA 18S V4 metabarcoding tables (Swarm) for Tara Oceans Expedition (2009-2013), including Tara Polar Circle Expedition (2013) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.7235995

Mahé, F., Henry, N., de Vargas, C., Tara Oceans Consortium, Coordinators, & Tara Oceans Expedition, Participants. (2022b). rDNA 18S V9 metabarcoding tables (Swarm) for Tara Oceans Expedition (2009-2013), including Tara Polar Circle Expedition (2013) [Dataset]. Zenodo. https://doi.org/10.5281/ZENODO.7236051

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. PeerJ, 2, e593. https://doi.org/ 10.7717/peerj.593

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17(1), 10. https://doi.org/10.14806/ej.17.1.200

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE, 8(4). https://doi.org/10.1371/journal.pone.0061217

Niels Strange, Sophus zu Ermgassen, Erica Marshall, Joseph W. Bull, Jette Bredahl Jacobsen, Why it matters how biodiversity is measured in environmental valuation studies compared to conservation science, Biological Conservation, Volume 292, 2024,110546, ISSN 0006-3207, https://doi.org/10.1016/j.biocon.2024.110546.

Official Journal of the European Union, (2010). Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (Text with EEA relevance)

Oksanen, (2001). vegan: Community Ecology Package. R package version 2.7-2, https://cran.r-project.org/web/packages/vegan. Accessed 26 Nov. 2025.

Pesant, Stephane; Not, Fabrice; Picheral, Marc; Kandels-Lewis, Stefanie; Le Bescot, Noan; Gorsky, G; Iudicone, Daniele; Karsenti, Eric; Speich, Sabrina; Troublé, Romain; Dimier, Céline; Searson, Sarah; Acinas, Silvia G; Bork, Peer; Boss, Emmanuel; Bowler, Chris; De Vargas, Colomban; Follows, Michael J; Grimsley, Nigel; Hingamp, Pascal; Jaillon, Olivier; Karp-Boss, Lee; Krzic, Uros; Ogata, Hiroyuki; Raes, Jeroen; Reynaud G., Emmanuel; Sardet, Christian; Sieracki, Michael E; Stemmann, Lars; Sullivan, Matthew B; Sunagawa, Shinichi; Velayoudon, Didier; Weissenbach, Jean; Wincker, Patrick (2015): Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data, 2, 150023, https://doi.org/10.1038/sdata.2015.23

Pierella Karlusich, J. J., Lombard, F., Irisson, J.-O., Bowler, C., & Foster, R. A. (2022). Coupling imaging and omics in plankton surveys: State-of-the-art, challenges, and future directions. Frontiers in Marine Science, 9, 878803..https://doi.org/10.3389/fmars.2022.878803

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016 Oct 18;4:e2584. doi: 10.7717/peerj.2584. PMID: 27781170; PMCID: PMC5075697.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology, 19 Suppl 1, 21–31. http://doi.org/10.1111/j.1365-294X.2009.04480.x

Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., & Kleindienst, S. (2020). Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA

(Gene) Amplicon Sequencing Pipeline. Frontiers in Microbiology, 11. https://doi.org/10.3389/fmicb.2020.550420

Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2015): Registry of all stations from the Tara Oceans Expedition (2009-2013) [dataset]. PANGAEA, https://doi.org/10.1594/PANGAEA.842237; Accessed 25/11/2025.

Torchiano, M. (2020). effsize: Efficient Effect Size Computation. R package version 0.8.1.

Widdicombe C.E.; Harbour D.(2021). Phytoplankton taxonomic abundance and biomass time-series at Plymouth Station L4 in the Western English Channel, 1992-2020. NERC EDS British Oceanographic Data Centre NOC. doi:10/grks.

Yates, M. C., Fraser, D. J., & Derry, A. M. (2019). Meta-analysis supports further refinement of eDNA for monitoring aquatic species-specific abundance in nature. In *Environmental DNA* (Vol. 1, Issue 1, pp. 5–13). Blackwell Publishing Inc. https://doi.org/10.1002/edn3.7

Zeileis, A., Köll, S., & Graham, N. (2020). "Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R." Journal of Statistical Software, 95(1), 1–36.

Appendices

Appendix I. Details of methods

A. Methods for coastal station 'L4' Timeseries – 2001-2023 - Metabarcoding Total Eukaryotes (18S V9 region rRNA gene)

Collection of samples and environmental variables: Samples were collected approximately once per week from the surface (≤5 m depth) of the Western English Channel (WEC) coastal Western Channel Observatory (WCO) site 'L4' (50◦15.00′ N, 4◦13.02′ W; site depth ~55 m). From the years 2001-2019, seawater was collected via CTD Niskin-rosette from the Plymouth Marine Laboratory RV 'Plymouth Quest'; 1 L per sample was filtered onto 0.45 µm cellulose-nitrate or polycarbonate filters, and stored at −80°C. Samples obtained in the years 2022-2023 were collected from the Marine Biological Association RV 'Sepia', either via CTD Niskin rosette (2 L volume filtered per sample), or by the under-way sampling system (1 L volume filtered per sample). These samples were collected onto 0.22 µm cellulose-nitrate filters, and stored as above.

Corresponding environmental parameters (metadata) were collected by WCO operations (https://www.westernchannelobservatory.org.uk), including temperature, salinity and nutrients (ammonia; nitrate; nitrite-nitrate; phosphate; silicate). Riverine input at time of sampling was calculated as the Total Gauged Daily Flow (GDF; m3/s), measured from six sites where water courses run into the Plymouth Sound (as reported by the National River Flow Archive; NRFA).

DNA extraction, metabarcode amplification and sequencing: All samples were extracted from filters within a 12-month extraction campaign. The majority of samples were extracted with the ZymoBIOMICS DNA Miniprep Kit (Zymo Research); 12 samples were extracted with the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research). Filter sections were added to 700 µl of kit lysis solution in kit lysis tubes, and were bead beaten three times for one minute duration at 10 m s-1, with five-minute intervals on ice. From here, DNA extraction followed manufacturer's protocols with 100 µl final elution volume. Possible kit/extraction associated contaminants were determined via the inclusion of one blank (i.e. no sample) sample per extraction batch of 50 samples. The 'Total-eukaryote' 18S-V9 region rRNA gene was amplified using PCR with the forward primer 1391f (5'-GTACACACCGCCCGTC -3') and the reverse primer EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC -3') (based on Amaral-Zettler et al, 2009, and Stoek et al, 2010), according to the Earth Microbiome Project (EMP) protocol (including reference to Caporaso et al., 2012), excepting the use of KAPA2G Robust Master Mix (Roche Diagnostics). Amplicons were sequenced on the Illumina MiSeq platform by NU-OMICS at Northumbria University.

Sequence processing, dereplication and taxonomic classification: In total, 20,941,365 18S-V9 rRNA reads were generated from 575 samples (plus 32 extraction kit blanks and 20 sequencing negative controls). Sequencing adapters were removed with CASAVA (Illumina), and primers were trimmed with CutAdapt v1.5 (Martin, 2011). Amplicon sequences were further processed using the using

nf-core/ampliseq version 2.8.0 (doi: 10.5281/zenodo.1493841) (Straub et al., 2020) of the nf-core collection of workflows (Ewels et al., 2020), utilising reproducible software environments from the Bioconda (Grüning et al., 2018) and Biocontainers (da Veiga Leprevost et al., 2017) projects. Data quality was evaluated with FastQC (Andrews, 2010: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and summarized with MultiQC (Ewels et al., 2016). Sequences were processed independently, then re-evaluated as one pool (pseudo-pooled) with DADA2 (Callahan et al., 2016) for: additional trimming (including elimination of PhiX contamination; trim of forward reads at 80 bp and reverse reads at 80 bp, reads shorter than this were discarded); removal of low-quality reads (discarding reads with > 3 expected errors; error correction); merging of read pairs, removal of PCR chimeras, dereplication and determination of individual biological sequences (i.e. Amplicon Sequence Variants: ASVs; total 8860). An average of 19,206 reads were retained per sample.

Taxonomic classification was performed by DADA2 against the Protist Ribosomal Reference (PR2) database ('PR2 - Protist Reference Ribosomal Database - Version 5.0.0') (Guillou et al., 2013). Eleven samples were sequenced twice due to low success on the initial run; read counts were summed per sample. Taxonomy and abundance of ASVs were primarily interrogated using the R (R Core Team, 2004) package Phyloseq (McMurdie and Holmes, 2013). Although the primers chosen favour amplification of eukaryotic DNA, some binding and amplification of prokaryotic sequences has previously been observed (Amaral-Zettler et al., 2009). ASVs that were classified as bacteria or archaea (n = 1584) were removed from downstream analysis. DNA extraction and sequencing negative controls were examined to identify ASVs likely resulting from contamination using the R package Decontam (Davis et al, 2018). Of the 61 ASVs flagged, 52 were removed by subsequent pruning steps (i.e. removal of negative controls/extraction kit blanks and very rare ASVs as described below). The remaining nine ASVs were further investigated via: NCBI BLAST database; prevalence across samples; presence in low-biomass samples only. Three of the nine ASVs identified by Decontam were judged as not being contamination and were retained. After removal of negative controls (i.e. blank kit- and sequencing samples), vary rare ASVs identified in <2 samples and <3 reads (n = 3443) were also removed (as Caracciolo et al., 2022).

Further investigation of some abundant ASVs classified only as 'Domain: Eukaryota' was undertaken via NCBI BLAST, revealing that ASVs with high similarity (95%-98%) to archaeal sequences had been identified as eukaryotic by PR2. For conservatism, ASVs not characterised beyond domain level (n = 816) were removed. Per-sample read abundance was assessed and samples with <1000 reads (n = 41) were discarded (lowest Goods Coverage: 97.7%; remaining samples: n = 534). Due to inherent biases of PCR amplification and unavoidable differences in sequencing depth between samples, the abundance data (sequence counts) of the remaining 2980 ASVs was transformed to proportional values (i.e. relative abundance). Prior to Nonmetric Multidimensional Scaling (NMDS), eDNA abundance underwent Hellinger transformation. This was required for consistency in comparisons to microscopy data, which required Hellinger transformation for the production of informative visualisations.

Calculation of diversity indices and significance tests: Shannon diversity and ASV richness were calculated per sample using the R package Vegan (Oksanen, 2001), and Hill numbers (effective species; q=1.5) were calculated using R package hillR (Li, 2018). Bray-Curtis and Jaccard dissimilarity were calculated using the R package Vegan as were the seasonal and monthly beta diversity dispersion analyses (functions: 'betadisper' and 'permutest'). Vegan was also used for Nonmetric Multidimensional Scaling (NMDS; function 'metaMDS), rarefaction (i.e. subsampling) of the eDNA samples to 1000 reads (function 'rrarefy') and generation of species accumulation curves by sample, separated per season (function:'specaccum'). Linear mixed-effects models (LMM) for the effect of season on alpha diversity metrics were calculated using the R package lme4 (Bates et al., 2015) (function: 'lmer'), and significance of seasonal variance was tested using Type III ANOVA with (Satterthwaite's method) from lmerTest (Kuznetsova et al, 2017) (function: anover(m)). Seasonal and yearly estimated marginal means and post-hoc pairwise comparisons of the former were conducted with R package emmeans (functions 'emeans()' and 'pairs()', respectively) (Lenth and Piaskowski, 2025). Robust standard errors for yearly estimated marginal means were calculated using the R package Sandwich (Zeileis et al., 2020), and changepoint years were calculated using the R package changepoint (Killick and Eckley, 2014). Differences in eDNA and microscopy total alpha- and beta diversity values were tested for significance using Wilcoxon rank–sum tests (base R) and effect sizes were calculated from Cliff's delta with R package effsize (Torchiano, 2020).

B. Overview of method for *Tara Oceans* total phytoplankton eDNA-traditional direct diversity values analyses.

Analyses were performed on publicly available metabarcoding sequences and cellular imaging datasets generated from *Tara Oceans* (Pesant et al., 2015; Table A1). Details of the methods used for production and curation of the data sets are presented in: Pesant *et al*. (2015) and Pierella Karlusich *et al.* (2024) (relating to light microscopy); Alberti *et al.* (2020) (amplification of 18S rRNA gene barcodes and Illumina sequencing protocol); Delage *et al*. (2024a; 2024b) and Mahé *et al*. (2022a; 2022b) (18S-V4 and -V9 metabarcoding table generation).

Table A1. Publicly available metabarcoding and imaging datasets used for the *Tara Oceans* total phytoplankton eDNA-traditional direct diversity values analyses.

Dataset	Method	Source
Tara Oceans (2009-2013) rDNA 18S V4 ASV table (dada2)	18S rRNA gene (V4 region) metabarcoding, processed by DADA2	(Delage et al., 2024a)
Tara Oceans (2009-2013) rDNA 18S V9 ASV table (dada2)	18S rRNA gene (V9 region) metabarcoding, processed by DADA2	(Delage et al., 2024b)
rDNA 18S V4 metabarcoding tables (Swarm) for Tara Oceans Expedition (2009-2013), including Tara Polar Circle Expedition (2013)	18S rRNA gene (V4 region) metabarcoding, processed by Swarm	(Mahé et al., 2022a)
rDNA 18S V9 metabarcoding tables (Swarm) for Tara Oceans Expedition (2009-2013), including Tara Polar Circle Expedition (2013)	18S rRNA gene (V9 region) metabarcoding, processed by Swarm	(Mahé et al., 2022b)
Light microscopy (LM) counts	Inverted light microscopy images from surface net samples, size fraction 20-180 µm	(Pierella Karlusich et al. 2023)
Summarized contextual data about metabarcoding Tara Oceans samples (2009-2013)	Information such as depth, time, geographic position, size fraction collected from the PANGAEA Data Publisher.	(Henry & Salazar, 2022)

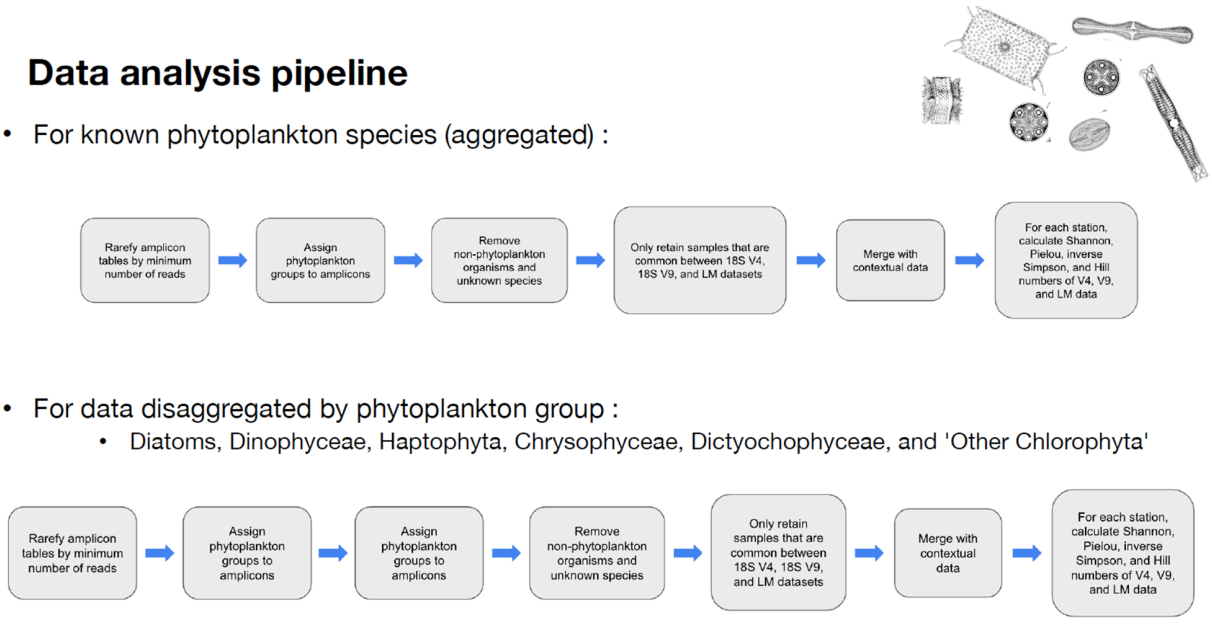


Figure A1. Overview of bioinformatic pipeline methods for comparison of eDNA-microscopy diversity values of *Tara* phytoplankton

Appendix II. Results of Analysis of Variance tests on differences in alpha diversity between seasons

Table A2. Results of Type III Analysis of Variance (ANOVA; Satterthwaite's method) and Post-hoc pairwise comparisons (Holm-adjusted) of Linear mixed-effect models (LMMs) on effect of season (fixed effect) on three diversity indices (Shannon Index; Richness; Hill numbers (q = 1.5)), after accounting for year (as a random effect). Coastal station 'L4' time series Data types: eDNA (metabarcode sequences); Microscopy (cell counts); eDNA-sub (subsampled/rarefied eDNA sequences). Global test model F and P values displayed; Different letters indicate significant differences in estimated marginal means.

Data type	Metric	F	P	Seasonal ranking (estimated marginal means with significance groups)
eDNA	Shannon	9.46	<0.001	Winter (2.26) a > Autumn (2.18) ab > Summer (2.04) bc > Spring (1.88) c
Microscopy	Shannon	21.11	<0.001	Winter (1.90) a > Summer (1.90) a > Autumn (1.85) a > Spring (1.52) b
eDNA-sub	Shannon	10.13	<0.001	Winter (3.46) a > Autumn (3.34) ab > Summer (3.11) bc > Spring (3.08) c
eDNA	Richness	5.62	<0.001	Autumn (129.31) a > Summer (127.037) a > Winter (125.33) a > Spring (109.14) b
Microscopy	Richness	96.79	<0.001	Autumn (59.04) a > Summer (58.11) a > Spring (48.87) b > Winter (43.57) c
eDNA-sub	Richness	9.91	<0.001	Winter (101.63) a > Autumn (99.67) ab > Summer (91.15) bc > Spring (84.14) c
eDNA	Hill (q1.5)	11.44	<0.001	Winter (26.52) a > Autumn (23.62) a > Spring (19.64) b > Summer (19.25) b
Microscopy	Hill (q1.5)	11.30	<0.001	Summer (5.27) a > Winter (4.99) a > Autumn (4.91) a > Spring (3.88) b
eDNA-sub	Hill (q1.5)	11.11	<0.001	Winter (27.19) a > Autumn (23.75) ab > Spring (19.84) bc > Summer (18.58) c

Appendix III. Across-year patterns in seasonal and monthly dissimilarity of beta diversity in coastal station 'L4' eDNA and microscopy data.

A. Results of permutation-dispersion tests of seasonal and monthly dispersion differences in beta diversity in the coastal station 'L4' time series.

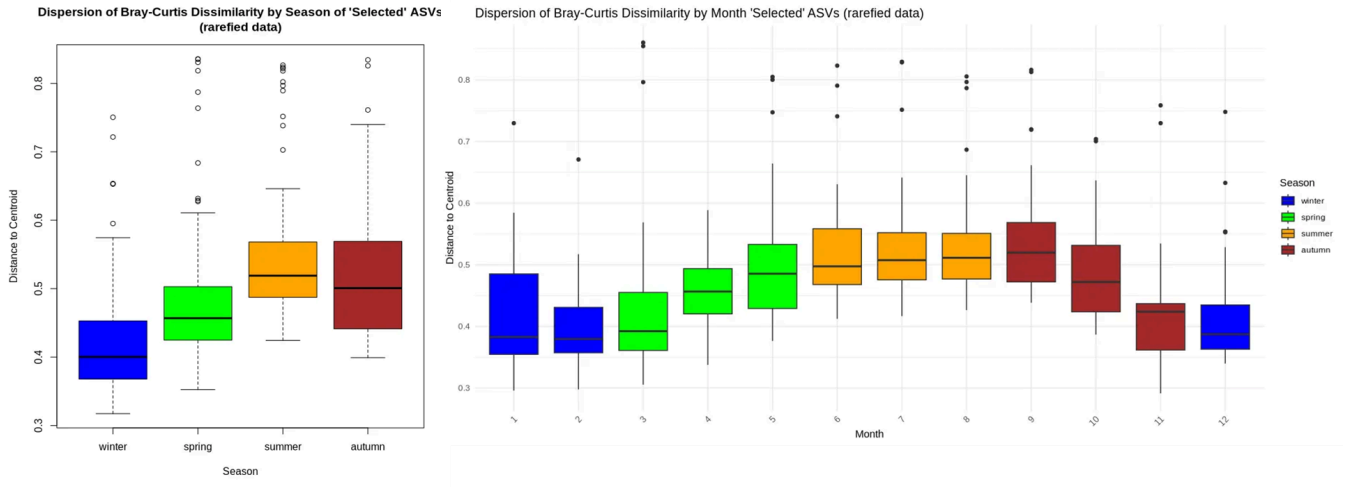
Table A3. PERMDISP results for dispersion of Bray–Curtis and Jaccard distances by season for each for each coastal station 'L4' time series data type: : eDNA (metabarcode sequences); Microscopy (cell counts); eDNA-sub (subsampled/rarefied eDNA sequences). Tests were run with 999 permutations; group dispersion means (i.e. distance to centroids) are shown for each season; degrees of freedom (Df) for groups (G) and residuals (R) are listed.

Data type	Metric	Df (G, R)	F	P	Winter	Spring	Summer	Autumn
eDNA	Bray-Curtis	3, 530	16.94	0.001	0.45	0.48	0.54	0.54
Microscopy	Bray-Curtis	3, 897	28.63	0.001	0.40	0.47	0.46	0.43
eDNA-sub	Bray-Curtis	3, 428	30.70	0.001	0.42	0.48	0.54	0.52
eDNA	Jaccard	3, 530	9.53	0.001	0.53	0.54	0.57	0.58
Microscopy	Jaccard	3, 897	11.89	0.001	0.50	0.50	0.47	0.49
eDNA-sub	Jaccard	3, 428	23.71	0.001	0.51	0.54	0.58	0.57

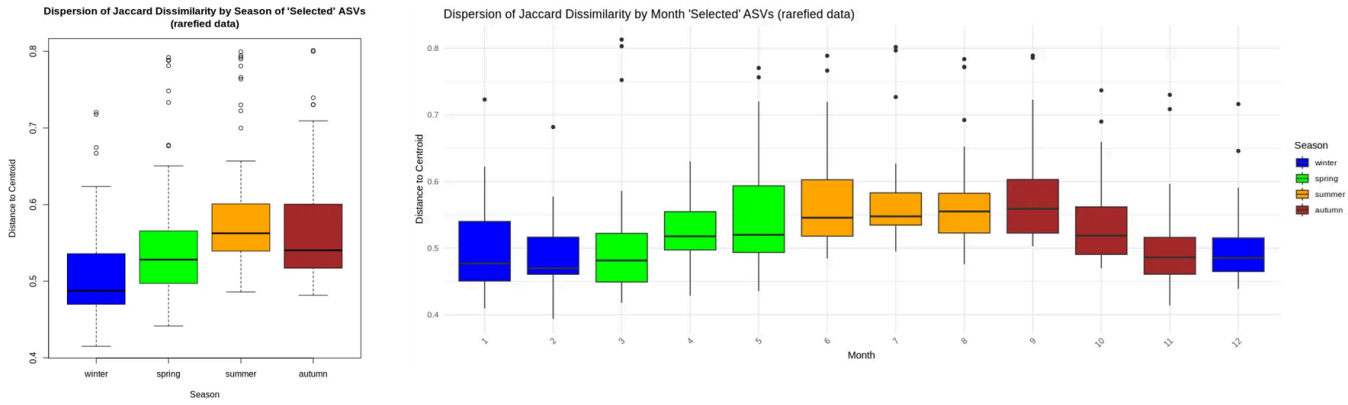
Table A4. PERMDISP results for dispersion of Bray–Curtis and Jaccard distances by month for each for each coastal station 'L4' time series data type: : eDNA (metabarcode sequences); Microscopy (cell counts); eDNA-sub (subsampled/rarefied eDNA sequences). Tests were run with 999 permutations; group dispersion means (i.e. distance to centroids) are shown for each month; degrees of freedom (Df) for groups (G) and residuals (R) are listed.

Data type	Metric	Df (G, R)	F	P	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
eDNA	Bray-Curtis	11, 522	5.97	0.001	0.47	0.41	0.43	0.46	0.50	0.53	0.52	0.53	0.53	0.54	0.48	0.45
Microscopy	Bray-Curtis	11, 889	12.26	0.001	0.39	0.39	0.40	0.45	0.48	0.46	0.46	0.44	0.44	0.41	0.37	0.40
eDNA-sub	Bray-Curtis	11, 420	10.54	0.001	0.42	0.40	0.43	0.46	0.50	0.53	0.53	0.53	0.54	0.49	0.43	0.43
eDNA	Jaccard	11,522	3.15	0.001	0.54	0.50	0.51	0.53	0.55	0.56	0.56	0.56	0.57	0.58	0.56	0.53
Microscopy	Jaccard	11, 889	3.51	0.001	0.48	0.49	0.49	0.49	0.47	0.46	0.46	0.46	0.47	0.48	0.49	0.50
eDNA-sub	Jaccard	11, 420	7.63	0.001	0.50	0.49	0.51	0.52	0.55	0.57	0.57	0.57	0.58	0.54	0.51	0.51

B. Across-year seasonal and monthly patterns of Bray-Curtis and Jaccard dissimilarity in rarefied coastal station L4 eDNA data.



A.



B.

Figure A2: Dispersion tests for A: Hellinger transformed Bray-Curtis dissimilarity; B. Jaccard dissimilarity, in rarefied (subsampled) eDNA data from the Coastal Station L4 time series: dispersion analysis (betadisper) with permutation tests (permutest) performed to check for significant differences among (left) seasonal- and (right) monthly. groups.

Appendix IV. Changepoints in alpha diversity across 20 years in the coastal station 'L4' time series.

Table A5. Years identified as 'changepoints' in the adjusted annual means of alpha diversity metrics (Shannon Index; Richness; Hill numbers (q = 1.5)), across the coastal station 'L4' time series for each for each data type: eDNA (metabarcode sequences); Microscopy (cell counts); eDNA-sub (subsampled/rarefied eDNA sequences). Adjusted annual mean values were estimated using a linear model (LM) with year as a fixed factor and controlled for season; LMs on eDNA data also controlled for the potential confounding factors of sequencing run. Changepoints were identified with the "Pruned Exact Linear Time" algorithm (PELT).

Data type	Diversity metric	Changepoint years
eDNA	Shannon	2007; 2011; 2018
Microscopy	Shannon	Null
eDNA-sub	Shannon	2007; 2011; 2018
eDNA	Richness	2007; 2011
Microscopy	Richness	2004; 2011
eDNA-sub	Richness	2007; 2011
eDNA	Hill (q1.5)	2007; 2011; 2018
Microscopy	Hill (q1.5)	2013*
eDNA-sub	Hill (q1.5)	2007; 2011; 2018

*Changepoint detected in pruned 'single taxonomist' dataset only

Appendix V. Further comparisons of *Tara* Alpha diversity values by phytoplankton group

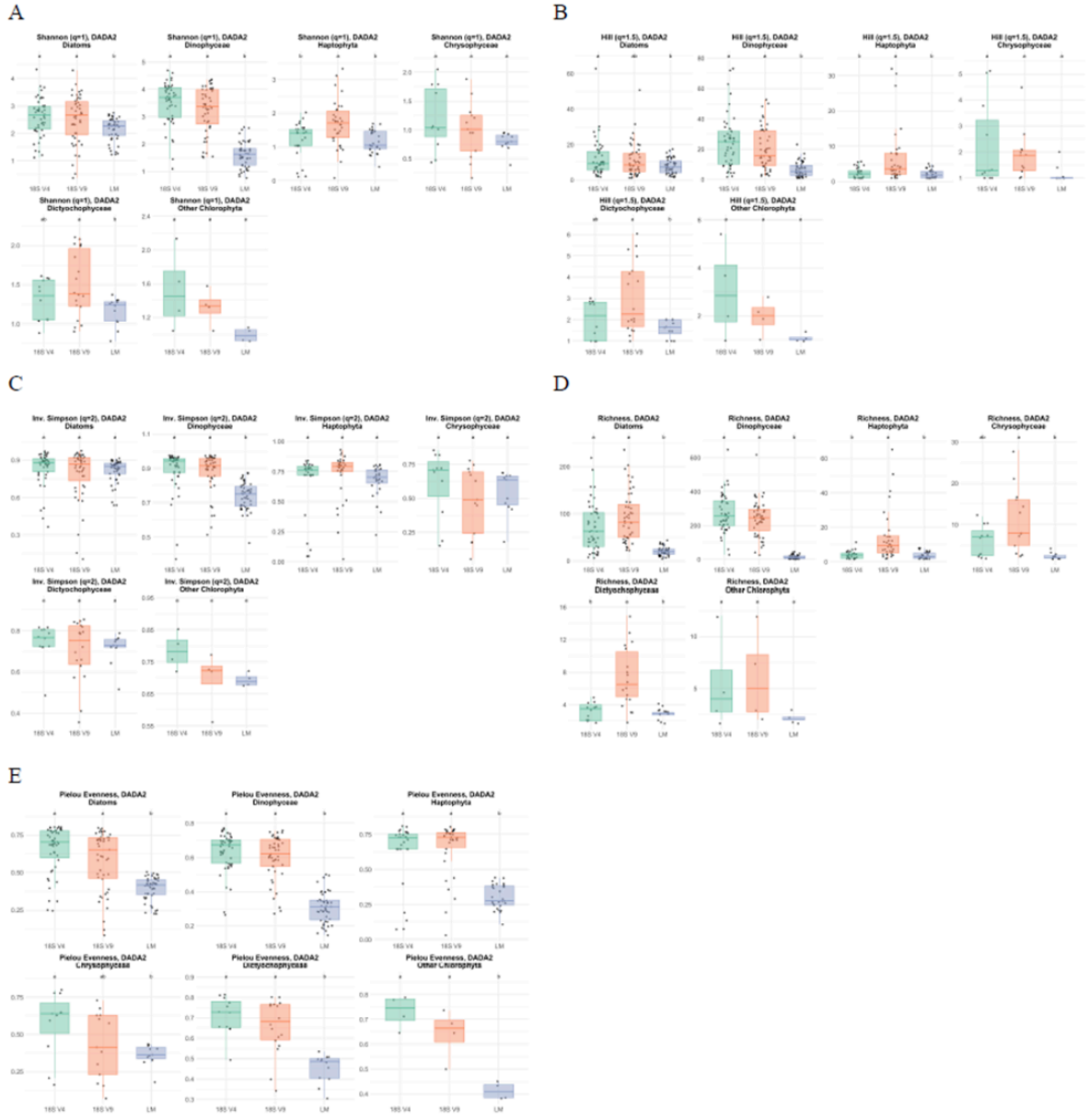


Figure A3. Alpha diversity indices by phytoplankton group, processed by DADA2. Letters indicate groups from ANOVA and $p < 0.05$ post-hoc Tukey HSD tests. (A) Shannon ($q = 1$); (B) Hill numbers ($q = 1.5$); (C) Inverse Simpson ($q = 2$); (D) Richness; (E) Pielou's evenness.

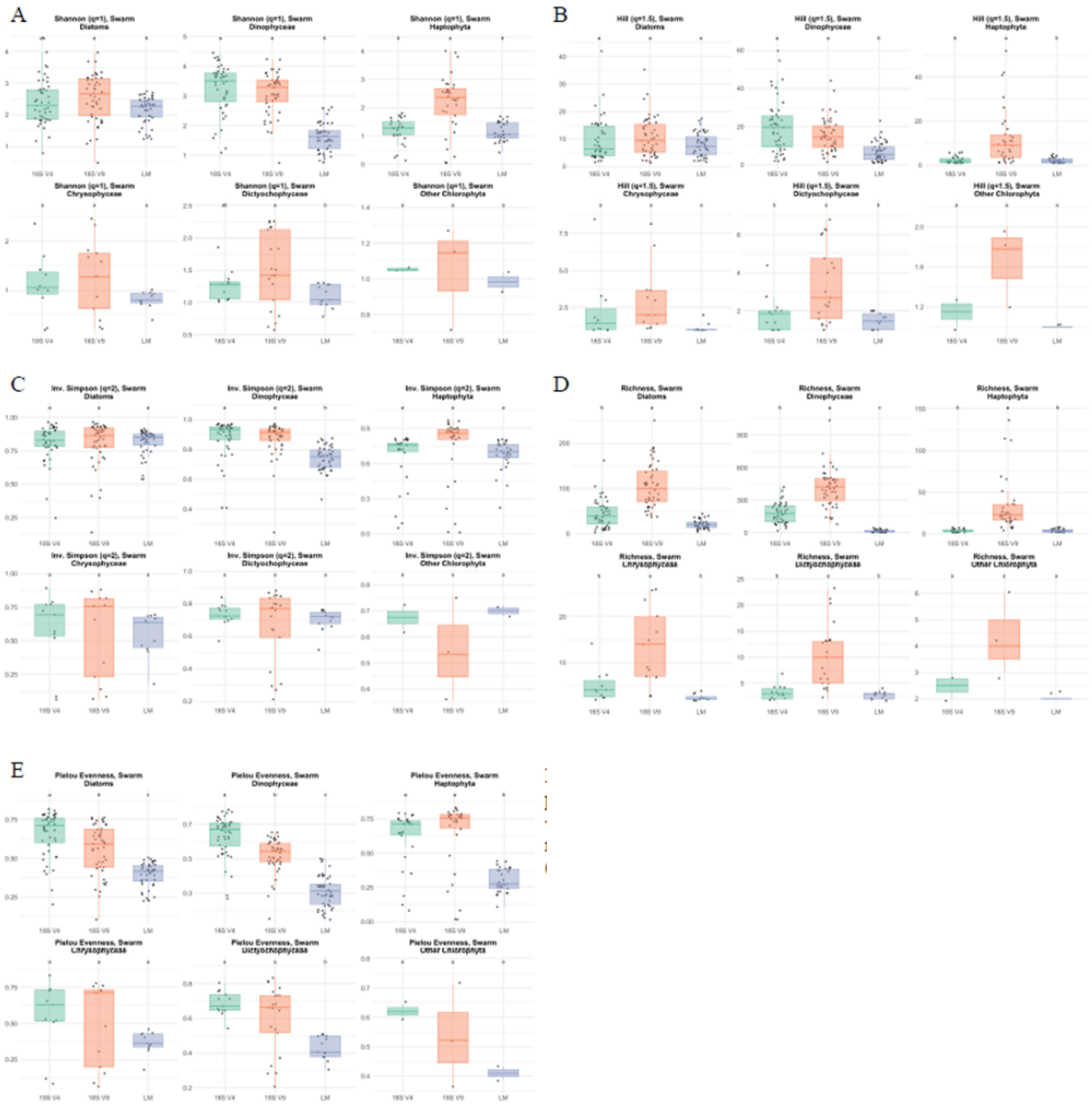


Figure A4. Alpha diversity indices by phytoplankton group, processed by Swarm. Letters indicate groups from ANOVA and $p < 0.05$ post-hoc Tukey HSD tests. (A) Shannon ($q = 1$); (B) Hill numbers ($q = 1.5$); (C) Inverse Simpson ($q = 2$); (D) Richness; (E) Pielou's evenness.

Appendix VI: Results of Wilcoxon tests with Cliff's delta for direct comparison of eDNA and microscopy diversity values at coastal station 'L4'.

Table A6. Wilcoxon rank–sum tests comparing coastal station 'L4' phytoplankton alpha- (Shannon Index; Richness; Hill numbers (q =1)) and beta diversity (Bray-Curtis and Jaccard dissimilarity) in: A. eDNA (metabarcoding) and microscopy; B. rarefied eDNA ('eDNA-sub') and microscopy. Effect sizes based on Cliff's delta with 95% confidence intervals. Due to the volume of distance values per dataset, beta diversity comparisons were run on a subsampled set of distance values (n = 100,000) to facilitate calculations.

A.

Metric	eDNA Median	Microscopy Median	W	P-value	Cliffs Δ	95% CI	Magnitude
Shannon	3.34	1.62	38053	<0.001	-0.84	-0.90, -0.76	Large
Richness	114	43	74989	<0.001	-0.69	-0.75, -0.62	Large
Hill (q1.5)	18.6	3.38	39638	<0.001	-0.83	-0.89 - 0.76	Large
Bray-Curtis	0.70	0.70	2.9268e+10	<0.001	0.02	0.01 -0.03	Negligible
Jaccard	0.77	0.75	3.2087e+10	<0.001	0.11	0.11 - 0.12	Negligible

B.

Metric	eDNA-sub. Median	Microscopy Median	W	P-value	Cliffs Δ	95% CI	Magnitude
Shannon	3.36	1.62	31385	<0.001	-0.85	-0.91 -0.77	Large
Richness	85	43	60053	<0.001	-0.72	-0.78 -0.65	Large
Hill (q1.5)	19.6	3.38	32931	<0.001	-0.85	-0.90 -0.76	Large
Bray-Curtis	0.71	0.70	2.3708e+10	<0.001	0.04	0.03 - 0.05	Negligible
Jaccard	0.77	0.75	2.6312e+10	<0.001	0.15	0.14 - 0.16	Small

Appendix VII: Overview of taxa shared between metabarcoding (9 pipelines) and microscopy data for the ASTAN time series.

Table A7. Overview of the taxa shared between the outputs of the 9 participating metabarcoding pipelines and the morphology-based identifications from Rigaut-Jalabert et al. (2021). The original metabarcoding sequence data comes from Caracciolo et al. (2022). Common taxa are extracted at the species level (N=33) (A), genus level (N=57) (B) and family level (N=38) (c).

A.

domain	phylum	kingdom	class	order	family	genus	species
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Hemiaulales	Hemiaulaceae	Cerataulina	Cerataulina pelagica
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae	Chaetoceros	Chaetoceros danicus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae	Chaetoceros	Chaetoceros diadema
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae	Chaetoceros	Chaetoceros eibenii
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae	Chaetoceros	Chaetoceros socialis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Coscinodiscales	Coscinodiscaceae	Coscinodiscus	Coscinodiscus radiatus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Skeletonemaceae	Detonula	Detonula pumila
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Lithodesmiales	Lithodesmiaceae	Ditylum	Ditylum brightwellii
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Hemiaulales	Hemiaulaceae	Eucampia	Eucampia zodiacus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae	Guinardia	Guinardia flaccida
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae	Guinardia	Guinardia striata
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Lithodesmiales	Lithodesmiaceae	Helicotheca	Helicotheca tamesis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Lauderiaceae	Lauderia	Lauderia annulata
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Leptocylindrales	Leptocylindraceae	Leptocylindrus	Leptocylindrus danicus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Leptocylindrales	Leptocylindraceae	Leptocylindrus	Leptocylindrus minimus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Naviculaceae	Meuniera	Meuniera membranacea
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Triceratiales	Triceratiaceae	Odontella	Odontella mobiliensis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Triceratiales	Triceratiaceae	Odontella	Odontella regia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Paraliales	Paraliaceae	Paralia	Paralia sulcata
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Bacillariales	Bacillariaceae	Psammodictyon	Psammodictyon panduriforme
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Striatellales	Striatellaceae	Striatella	Striatella unipunctata
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassionematales	Thalassionemataceae	Thalassionema	Thalassionema nitzschioides
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira	Thalassiosira anguste-lineata
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira	Thalassiosira hendeyi
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira	Thalassiosira nordenskioeldii
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira	Thalassiosira punctigera
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira	Thalassiosira rotula
Eukaryota	Ochrophyta	Chromista	Dictyochophyceae	Dictyochales	Dictyochaceae	Dictyocha	Dictyocha speculum
Eukaryota	Myzozoa	Chromista	Dinophyceae	Dinophysiales	Dinophysiaceae	Phalacroma	Phalacroma rotundatum
Eukaryota	Myzozoa	Chromista	Dinophyceae	Noctilucales	Noctilucaceae	Noctiluca	Noctiluca scintillans
Eukaryota	Myzozoa	Chromista	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium	Protoperidinium bipes
Eukaryota	Chlorophyta	Plantae	Prasinophyceae	Halosphaerales	Pterospermataceae	Pterosperma	Pterosperma cristatum
Eukaryota	Haptophyta	Chromista	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	Phaeocystis globosa

B.

domain	phylum	kingdom	class	order	family	genus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Achnanthales	Achnanthaceae	Achnanthes
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Coscinodiscales	Heliopeltaceae	Actinoptychus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Fragilariales	Fragilariaceae	Asterionellopsis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Asterolamprales	Asterolampraceae	Asteromphalus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Bacillariales	Bacillariaceae	Bacillaria
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae	Bacteriastrum
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Hemiaulales	Hemiaulaceae	Cerataulina
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae	Chaetoceros
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Achnanthales	Cocconeidaceae	Cocconeis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Corethrales	Corethraceae	Corethron
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Coscinodiscales	Coscinodiscaceae	Coscinodiscus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Bacillariales	Bacillariaceae	Nitzschia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae	Dactyliosolen
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhaponeidales	Rhaphoneidaceae	Delphineis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Skeletonemaceae	Detonula
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Lithodesmiales	Lithodesmiaceae	Ditylum
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Hemiaulales	Hemiaulaceae	Eucampia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Striatellales	Striatellaceae	Grammatophora
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae	Guinardia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Naviculaceae	Haslea
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Lithodesmiales	Lithodesmiaceae	Helicotheca
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Hemiaulales	Hemiaulaceae	Hemiaulus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Lauderiaceae	Lauderia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Leptocylindrales	Leptocylindraceae	Leptocylindrus
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Licmophorales	Licmophoraceae	Licmophora
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Naviculaceae	Meuniera
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Naviculaceae	Navicula
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Triceratiales	Triceratiaceae	Odontella
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Paraliales	Paraliaceae	Paralia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Cymatosirales	Cymatosiraceae	Plagiogrammopsis
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Pleurosigmataceae	Pleurosigma
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae	Proboscia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Bacillariales	Bacillariaceae	Psammodictyon
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Bacillariales	Bacillariaceae	Pseudo-nitzschia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae	Rhizosolenia
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Skeletonemaceae	Skeletonema
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Striatellales	Striatellaceae	Striatella
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassionematales	Thalassionemataceae	Thalassionema
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira
Eukaryota	Ochrophyta	Chromista	Dictyochophyceae	Dictyochales	Dictyochaceae	Dictyocha
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gonyaulacales	Gonyaulacaceae	Alexandrium
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Amphidinium
Eukaryota	Myzozoa	Chromista	Dinophyceae	Dinophysiales	Dinophysiaceae	Dinophysis
Eukaryota	Myzozoa	Chromista	Dinophyceae	Peridiniales	Protoperidiniaceae	Diplopsalis
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gonyaulacales	Gonyaulacaceae	Gonyaulax
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gymnodinium
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Gyrodinium
Eukaryota	Myzozoa	Chromista	Dinophyceae	Peridiniales	Peridiniida incertae sedis	Heterocapsa
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Kareniaceae	Karenia
Eukaryota	Myzozoa	Chromista	Dinophyceae	Noctilucales	Noctilucaceae	Noctiluca
Eukaryota	Myzozoa	Chromista	Dinophyceae	Prorocentrales	Prorocentraceae	Prorocentrum
Eukaryota	Myzozoa	Chromista	Dinophyceae	Peridiniales	Protoperidiniaceae	Protoperidinium
Eukaryota	Myzozoa	Chromista	Dinophyceae	Peridiniales	Peridiniaceae	Scrippsiella
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Torodinium
Eukaryota	Chlorophyta	Plantae	Prasinophyceae	Halosphaerales	Pterospermataceae	Pterosperma
Eukaryota	Haptophyta	Chromista	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis
Eukaryota	Chlorophyta	Plantae	Pyramimonadophyceae	Pyramimonadales	Pyramimonadaceae	Halosphaera

C.

domain	phylum	kingdom	class	order	family
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Achnanthales	Achnanthaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Coscinodiscales	Hemidiscaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Coscinodiscales	Heliopeltaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Fragilariales	Fragilariaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Asterolamprales	Asterolampraceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Bacillariales	Bacillariaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Chaetocerotanae incertae sedis	Chaetocerotaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Hemiaulales	Hemiaulaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Achnanthales	Cocconeidaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Corethrales	Corethraceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Coscinodiscales	Coscinodiscaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhizosoleniales	Rhizosoleniaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Rhaponeidales	Rhaphoneidaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Skeletonemaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Lithodesmiales	Lithodesmiaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Striatellales	Striatellaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Naviculaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Lauderiaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Leptocylindrales	Leptocylindraceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Licmophorales	Licmophoraceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Triceratiales	Triceratiaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Paraliales	Paraliaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Cymatosirales	Cymatosiraceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Naviculales	Pleurosigmataceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Melosirales	Hyalodiscaceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassionematales	Thalassionemataceae
Eukaryota	Ochrophyta	Chromista	Bacillariophyceae	Thalassiosirales	Thalassiosiraceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gonyaulacales	Gonyaulacaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Gymnodiniaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gonyaulacales	Ceratiaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Dinophysiales	Dinophysiaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Peridiniales	Protoperidiniaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Pyrocystales	Pyrocystaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Gymnodiniales	Kareniaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Noctilucales	Noctilucaceae
Eukaryota	Myzozoa	Chromista	Dinophyceae	Prorocentrales	Prorocentraceae
Eukaryota	Haptophyta	Chromista	Prymnesiophyceae	Phaeocystales	Phaeocystaceae
Eukaryota	Chlorophyta	Plantae	Pyramimonadophyceae	Pyramimonadales	Pyramimonadaceae

Project Coordinator
Nicolas Pade | nicolas.pade@embrc.eu

Project Manager
Giulia Vecchi | giulia.vecchi@embrc.eu

Press and Communications
Mathilde Vidal | mathilde@erinn.eu
Website: MarcoBolo-Project.eu
Twitter: @MARCOBOLO_EU
LinkedIn: MARCO-BOLO

Funded by the European Union

Funded by the European Union under the Horizon Europe Programme, Grant Agreement No. 101082021 (MARCO-BOLO). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK participants in MARCO-BOLO are supported by the UKRI's Horizon Europe Guarantee under the Grant No. 10068180 (MS); No. 10063994 (MBA); No. 10048178 (NOC).

UK Research and Innovation

authority can be held responsible for them.

UK participants in Empowering are supported by UKRI Grant No. 10040188 (QUB).