Title: Depth-dependent mycoplankton glycoside hydrolase gene activity in the open ocean – evidence from the Tara Ocean eukaryote metatranscriptomes

Running title: Fungal glycoside hydrolases in Tara Oceans metatranscriptomes

Nathan Alexis Mitchell Chrismas[1,2] & Michael Cunliffe[1,3]

[1]Marine Biological Association of the UK, The Laboratory, Citadel Hill, Plymouth, UK

[2]School of Geographical Sciences, University of Bristol, University Road, Bristol, UK

[3]School of Biological and Marine Sciences, University of Plymouth, Plymouth, UK

Corresponding authors:

Nathan Chrismas natchr@mba.ac.uk

Michael Cunliffe micnli@mba.ac.uk

Abstract

Mycoplankton are widespread components of marine ecosystems, yet the full extent of their functional roles remain poorly known. Marine mycoplankton are likely functionally analogous to their terrestrial counterparts, including performing saprotrophy and degrading high-molecular weight organic substrates using carbohydrate active enzymes (CAZymes). We investigated the diversity of transcribed oceanic fungal CAZyme genes using the Tara Oceans Metatranscriptomic Occurrences database. We revealed a high diversity of actively transcribed fungal glycoside hydrolases in the open ocean, including a particularly high diversity of enzymes that act upon cellulose in surface waters and the deep chlorophyll maximum (DCM). A variety of other glycoside hydrolases acting on a range of biogeochemically important polysaccharides including β-glucans and chitin were also found. This analysis demonstrates that mycoplankton are active saprotrophs in the open ocean and paves the way for future research into the depth-dependent roles of marine fungi in oceanic carbon cycling, including the biological carbon pump.

Even though our understanding of marine fungal diversity is increasing [1, 2], a comprehensive knowledge of their active functional ecology remains limited, especially in the open ocean [1]. Fungal activity has been detected in corals, deep sea and coastal sediments, and associated with phytoplankton blooms, including parasites [1, 2], but the full extent that fungi are functionally active throughout the open ocean water column is yet to be established.

Many terrestrial fungi occupy key roles as saprotrophs by decomposing and recycling biogenic matter, making them intrinsic components of healthy functioning ecosystems [3]. In coastal waters there is evidence that planktonic fungi (mycoplankton) degrade and utilise phytoplankton-derived carbohydrate-rich matter in broadly analogous functional modes [4]. However, the extent that carbohydrate-based fungal saprotrophy occurs in the open ocean remains largely speculated [1].

Glycoside hydrolases (GHs) are a ubiquitous group of carbohydrate-active enzymes (CAZymes) [5] that degrade complex polysaccharides and are categorised into substrate-specific families. In terrestrial fungi, secreted CAZymes are key to the functional potential of saprotrophs and are the primary mode of degradation of high-molecular weight (HMW) polysaccharides (e.g. wood). Coastal saprotrophic mycoplankton also employ secreted GHs to degrade phytoplankton-derived HMW carbohydrate-based substrates [4], but the prevalence and identity of the specific GHs families of active open ocean mycoplankton are unknown.

Metagenomes from the Tara Oceans project have been used to assess mycoplankton diversity [6, 7], but the associated metatranscriptomes are yet to be fully explored from a fungal perspective. We interrogated the Marine Atlas of Tara Ocean Unigenes (MATOU) metatranscriptomic occurrences database [8] for transcribed fungal GH genes to explore broad-scale depth-dependent structuring in the oceans. The MATOU database consists of all unique eukaryotic genes assembled from the Tara Oceans metatranscriptomes (Unigenes), their associated taxonomy and occurrence within samples (full methods described in [9]).

To identify GHs within the MATOU database, reference libraries were created for 61 fungal GH families and clustered using CD-Hit [9]. The MATOU Unigenes were searched against these libraries using Diamond v.0.9.22 [10], yielding a database where each positive Unigene match represents a unique GH (similar genes clustered when similarity <95% over 90% of the smallest sequence [8]). The database was screened for non-fungal Unigenes using the MATOU taxonomy (Figure 1a, Supplementary Figure 1). After removal of redundant matches (i.e. where multiple GHs matched to a single Unigene), 1,326 unique fungal GH Unigenes were found (~0.001% of the entire Unigene catalogue) that occurred 44,386 times in all Tara Oceans samples.

The top ten GH families containing the greatest number of unique genes were determined by ranking the sum of all Unigene occurrences from all samples for each family (Figure 1b). Overall, the greatest number of unique GHs were involved in cellulose degradation (GH7). Other substrates of the most diverse GH families included β-glucans (GH17 and GH72), β-glycans (GH5, GH16, GH3), α-glucans (GH13), chitin (GH18), N-/O-glycans (GH47) and xylan (GH43). GH diversity was dominated by genes originating from the Ascomycota, except for the abundant GH7s, many of which were unclassified fungal genes (Figure 1c).

Transcribed fungal GH genes were recovered from 66 stations (Figure 2). GH7 diversity was greatest in the surface and deep chlorophyll maximum (DCM), especially in high productivity areas such as the Mediterranean Sea and the Indian Ocean (Figure 2b). At stations where concomitant samples were available from the surface, DCM and mesopelagic, a distinct drop in GH7 diversity was seen in the mesopelagic. The diversity of other abundant GHs was more heterogeneous between sites (Figure 2c).

The high prevalence of unique GH7 transcribed genes in surface waters is likely a response to increased ‘fresh’ phytoplankton-derived matter in the photic zone. Cellulose is a key structural component of many phytoplankton cells [11, 12]. Polysaccharides, such as cellulose, also represent a primary source of particulate organic carbon (POC) [13]. Of the enzymes responsible for cellulose degradation, those within the GH7 family are typically the

most active, and are important in biomass degradation by terrestrial fungi [14]. The decline in the number of unique GH7 genes below the DCM in the mesopelagic zone in the six sites sampled suggests a shift in mycoplankton functionality due to depletion of readily available phytoplankton-derived carbon sources, and corresponds with the decrease in overall polysaccharide concentrations between surface waters and the mesopelagic zone [15].

Amongst the other diverse GH groups, the greatest number of unique genes was in the GH17 family, a group of CAZymes that degrade β-glucans. *Cladosporium* mycoplankton isolated from coastal waters have been shown to secrete GH17 β-1,3-glucosidase when utilising laminarin [4], which is an algal-derived β-glucan that is a major POC component [16]. Also prevalent were GH18 genes, responsible for degrading chitin, another major polysaccharide and important component of zooplankton and some phytoplankton, suggesting that chitin degradation is a functional role of open ocean mycoplankton, as with fungi in freshwater lake ecosystems [17].

While the diversity of transcribed GHs is a strong indicator of active carbohydrate metabolism by mycoplankton, the identity of these fungi remains uncertain. The extent of GH7 genes with unresolved fungal taxonomy opens interesting questions about the phylogeny of these taxa. The majority of the other GHs were affiliated to the Ascomycota, in line with phylogenetic studies that show the phylum dominates open ocean mycoplankton diversity [6]. However, there is a lack of early-diverging taxa (e.g. Chytridiomycota) within the MATOU database. Since Chytridiomycota parasitism of phytoplankton takes place in the open ocean [2] their absence highlights outstanding gaps in our understanding of the functional ecology of marine fungi.

The vertical flux of POC in the open ocean is an essential feature of the biological carbon pump (BCP), sustaining the oceans capability to sequester carbon [18]. Biogeochemical models of the BCP do not currently consider fungi (e.g. [19]). Given that marine snow is an apparent hotspot for fungi [20], and that here we show differences in fungal GH expression suggesting depth-dependent resource partitioning in relation to POC-

114 associated substrates, the recently proposed ‘mycoflux’ [2] should be considered within a
115 contemporary view of the BCP.
116
117 **Competing interests**
118 The authors declare no conflict of interest.
119
120 **Acknowledgments**
121 NC is supported by NERC grant NE/N006151/1. MC is supported by NERC grant
122 NE/N006151/1 and ERC grant 772584.

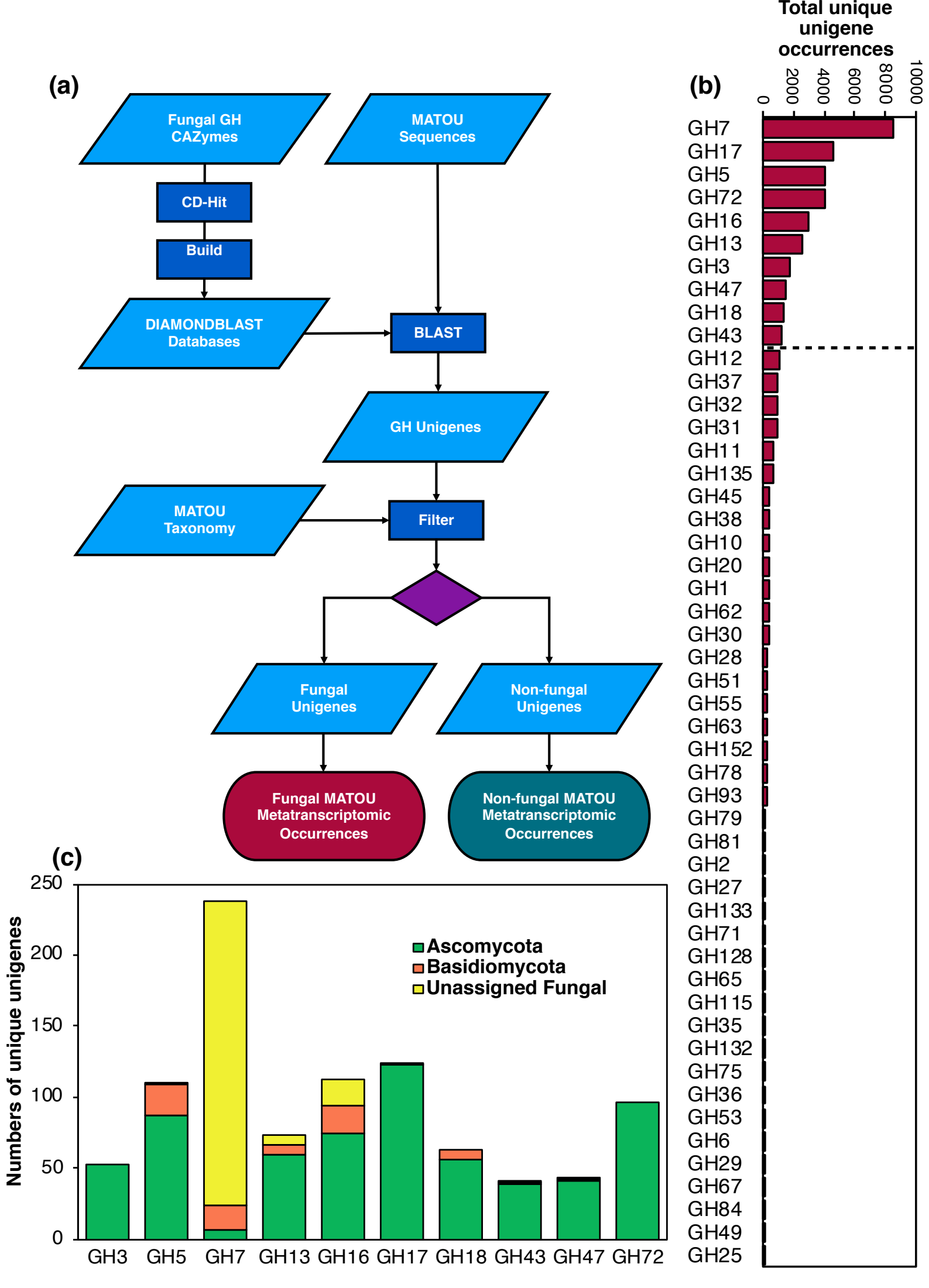
References

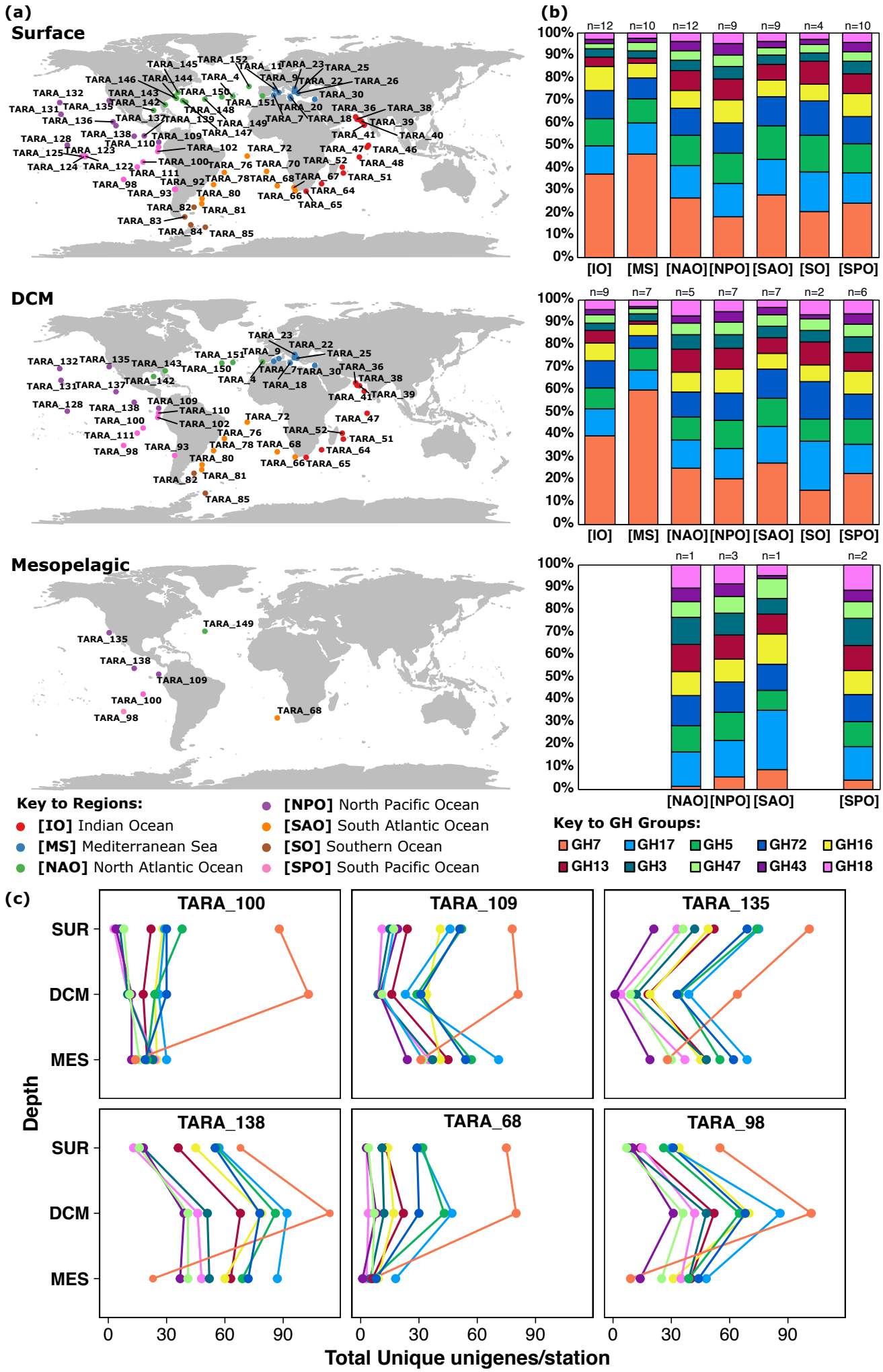
1. Amend A, Burgaud G, Cunliffe M, Edgcomb VP, Ettinger CL, Gutiérrez MH, et al. Fungi in the Marine Environment: Open Questions and Unsolved Problems. *mBio* 2019; **10**: e01189-18.
2. Grossart H-P, Wyngaert SV den, Kagami M, Wurzbacher C, Cunliffe M, Rojas-Jimenez K. Fungi in aquatic ecosystems. *Nat Rev Microbiol* 2019; 1.
3. Frąc M, Hannula SE, Bełka M, Jędryczka M. Fungal Biodiversity and Their Role in Soil Health. *Front Microbiol* 2018; **9**.
4. Cunliffe M, Hollingsworth A, Bain C, Sharma V, Taylor JD. Algal polysaccharide utilisation by saprotrophic planktonic marine fungi. *Fungal Ecol* 2017; **30**: 135–138.
5. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014; **42**: D490-495.
6. Hassett BT, Vonnahme TR, Peng X, Jones EBG, Heuzé C. Global diversity and geography of planktonic marine fungi. *Bot Mar* 2019; **0**.
7. Morales SE, Biswas A, Herndl GJ, Baltar F. Global Structuring of Phylogenetic and Functional Diversity of Pelagic Fungi by Depth and Temperature. *Front Mar Sci* 2019; **6**.
8. Carradec Q, Pelletier E, Silva CD, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. *Nat Commun* 2018; **9**: 373.
9. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinforma Oxf Engl* 2012; **28**: 3150–3152.
10. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015; **12**: 59–60.
11. Domozych D. Algal Cell Walls. *eLS*. 2019. American Cancer Society, pp 1–11.
12. Walker CE, Heath S, Salmon DL, Smirnoff N, Langer G, Taylor AR, et al. An Extracellular Polysaccharide-Rich Organic Layer Contributes to Organization of the Coccosphere in Coccolithophores. *Front Mar Sci* 2018; **5**.

13. Smetacek V, Hendrikson P. Composition of Particulate Organic-Matter in Kiel Bight in Relation to Phytoplankton Succession. *Oceanol Acta* 1979; **2**: 287–298.
14. Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, et al. Fungal Cellulases. *Chem Rev* 2015; **115**: 1308–1448.
15. Pakulski JD, Benner R. Abundance and distribution of carbohydrates in the ocean. *Limnol Oceanogr* 1994; **39**: 930–940.
16. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is a major molecule in the marine carbon cycle. *Proc Natl Acad Sci* 2020; **117**: 6599–6607.
17. Wurzbacher CM, Bärlocher F, Grossart H-P. Fungi in lake ecosystems. *Aquat Microb Ecol* 2010; **59**: 125–149.
18. Boyd PW, Claustre H, Levy M, Siegel DA, Weber T. Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature* 2019; **568**: 327–335.
19. Giering SLC, Sanders R, Lampitt RS, Anderson TR, Tamburini C, Boutrif M, et al. Reconciliation of the carbon budget in the ocean's twilight zone. *Nature* 2014; **507**: 480–483.
20. Bochdansky AB, Clouse MA, Herndl GJ. Eukaryotic microbes, principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine snow. *ISME J* 2017; **11**: 362–373.

Figure 1. a) Pipeline describing the steps involved in identifying fungal CAZymes within the Tara Ocean MATOU Unigenes database. A fungal GH protein sequence reference database was created from all the 61 characterised GH subfamilies found in fungi. This was consolidated by clustering sequences at 95 % identity using CD-Hit before Diamond BLAST databases were generated for each subfamily. The MATOU Unigenes were searched against each of these 61 databases using the following thresholds: evalue > 1e-30 , score > 1, Subject Cov > 75 %, keeping only the best alignments. Positive matches were then screened using the MATOU taxonomy to discriminate between fungal and non-fungal Unigenes. Occurrences of each Unigene within the Tara Oceans transcriptomes were returned. b) Fungal GH groups found in the MATOU Unigenes database Tara Oceans Metatranscriptomic occurrences ranked by abundance over all Tara Oceans samples. c) Total numbers and taxonomy of unique Fungal Unigenes from the 10 most abundant GH groups.

Figure 2. a) Global map indicating Tara Oceans stations searched for fungal GH Unigenes in the surface, deep chlorophyll maximum (DCM) and mesopelagic. b) Mean unique unigenes/station for each of the major oceanic regions sampled. c) Depth dependent partitioning of fungal GH Unigenes in the Surface (SUR), DCM and mesopelagic (MES).


(a)
Fungal GH CAZymes
MATOU Sequences
CD-Hit
Build
DIAMONDBLAST Databases
BLAST
GH Unigenes
MATOU Taxonomy
Filter
Fungal Unigenes
Non-fungal Unigenes
Fungal MATOU Metatranscriptomic Occurrences
Non-fungal MATOU Metatranscriptomic Occurrences
(b)
Total unique unigene occurrences
0
2000
4000
6000
8000
10000
GH7
GH17
GH5
GH72
GH16
GH13
GH3
GH47
GH18
GH43
GH12
GH37
GH32
GH31
GH11
GH135
GH45
GH38
GH10
GH20
GH1
GH62
GH30
GH28
GH51
GH55
GH63
GH152
GH78
GH93
GH79
GH81
GH2
GH27
GH133
GH71
GH128
GH65
GH115
GH35
GH132
GH75
GH36
GH53
GH6
GH29
GH67
GH84
GH49
GH25
(c)
Numbers of unique unigenes
250
200
150
100
50
0
Ascomycota
Basidiomycota
Unassigned Fungal
GH3
GH5
GH7
GH13
GH16
GH17
GH18
GH43
GH47
GH72


(a)
Surface
DCM
Mesopelagic
(b)
n=12 n=10 n=12 n=9 n=9 n=4 n=10
n=9 n=7 n=5 n=7 n=7 n=2 n=6
n=1 n=3 n=1 n=2
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
[IO] [MS] [NAO] [NPO] [SAO] [SO] [SPO]
Key to Regions:
[IO] Indian Ocean
[MS] Mediterranean Sea
[NAO] North Atlantic Ocean
[NPO] North Pacific Ocean
[SAO] South Atlantic Ocean
[SO] Southern Ocean
[SPO] South Pacific Ocean
Key to GH Groups:
GH7 GH17 GH5 GH72 GH16
GH13 GH3 GH47 GH43 GH18
(c)
TARA_100
TARA_109
TARA_135
TARA_138
TARA_68
TARA_98
SUR
DCM
MES
Depth
0 30 60 90
Total Unique unigenes/station