An article submitted to Environmental Microbiology

Constant abundances of ubiquitous uncultured protists in the open sea assessed
by automated microscopy

Jean-François Mangot^{1*}, Irene Forn¹, Aleix Obiol¹ and Ramon Massana^{1**}

Author affiliation

¹ Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), Barcelona,
Catalonia, Spain.

For correspondence:

*E-mail: jean-francois.mangot@wanadoo.fr. Tel: (+34) 932309571; Fax: (+34) 93 2309555.

**E-mail: ramonm@icm.csic.es. Tel: (+34) 93 2309500; Fax: (+34) 93 2309555.

Running head: Automatic microscopic counts of small protists

17 **Summary**

18 Protists have fundamental ecological roles in marine environments and their diversity is being
19 increasingly explored, yet little is known about the quantitative importance of specific taxa in
20 these ecosystems. Here we optimised a newly developed automated system of image
21 acquisition and image analysis to enumerate minute uncultured cells of different sizes targeted
22 by fluorescence in situ hybridization. The automated counting routine was highly
23 reproducible, well correlated with manual counts, and was then applied on surface and DCM
24 samples from the Malaspina 2010 circumnavigation. The three targeted uncultured taxa
25 (MAST-4, MAST-7 and MAST-1C) were found in virtually all samples from several ocean
26 basins (Atlantic, Indian and Pacific) in fairly constant cell abundances, following typical
27 lognormal distributions. Their global abundances averaged 49, 23 and 7 cells ml⁻¹,
28 respectively, and altogether the three groups accounted for about 10-20% of heterotrophic
29 picoeukaryotes. Our innovative high-throughput cell counting routine allows for the first time
30 a direct assessment of the biogeographic distribution of small protists (< 5 µm) and shows the
31 ubiquity in sunlit oceans of three bacterivorous taxa, suggesting their key roles in marine
32 ecosystems.

33

34 **Introduction**

35 Protists are critical components of marine systems, largely contributing to biogeochemical
36 processes and ecosystem functioning as major players in primary production, nutrient cycling,
37 and food-web dynamics (Sherr and Sherr, 2008; Worden *et al.*, 2015). In particular, pico- and
38 small nanosized heterotrophic protists (1-5 μm) are important mortality agents of planktonic
39 prokaryotes and thus link prokaryotic biomass with upper trophic levels (Jürgens and
40 Massana, 2008). A key component of this functional group are the MARine STRamenopiles
41 (MASTs), initially detected in molecular diversity surveys and formed by at least 18
42 independent lineages of essentially uncultured protists (Massana *et al.*, 2014). Some MAST
43 lineages have shown active bacterivory (Massana *et al.*, 2009; Piwosz *et al.*, 2013) and a
44 widespread distribution in sequencing datasets (Lin *et al.*, 2012; Logares, *et al.*, 2012;
45 Seeleuthner *et al.*, 2018). Furthermore, MAST cells may be of quantitative importance and
46 one group in particular, MAST-4, averaged 9% of the picosized heterotrophic protists in a
47 sparse set of marine surface samples, being proposed as the most abundant heterotrophic
48 protist in the oceans (Massana *et al.*, 2006). Unfortunately, despite the apparent importance of
49 MAST cells, our knowledge of their biogeographic distribution is still very incomplete.

50 Since the beginning of the 21st century, technological progresses in molecular ecology have
51 substantially boosted our understanding of the biogeography of eukaryotic microbes. High-
52 Throughput Sequencing (HTS) surveys, allowing the parallel analysis of multiple samples,
53 have been very successful in unveiling the patterns of the small protists community structure
54 at various spatial and temporal scales (Amaral-Zettler *et al.*, 2009; Lepère *et al.*, 2013;
55 Mangot *et al.*, 2013; de Vargas *et al.*, 2015; Massana *et al.*, 2015). These approaches,
56 however, are intrinsically limited in revealing protist abundances. Even though HTS surveys
57 may provide reasonable relative quantitative estimates (Mangot *et al.*, 2013; Giner *et al.*,

2016), it is known that these may deviate significantly from true cell abundances (Medinger *et al.*, 2010). Perhaps their most critical drawback is the large variation in the ribosomal (r)DNA operon copy number among protist taxa, which in microbial eukaryotes can vary by several orders of magnitude and is roughly correlated with cell size (Zhu *et al.*, 2005) and genome size (Prokopowich *et al.*, 2003). These large differences, from a few copies per cell in some green algae to several thousands in some dinoflagellates or ciliates (Zhu *et al.*, 2005; Vd'ačný *et al.*, 2012), will indeed influence the interpretation of the relative abundance of protist taxa in DNA surveys (Medinger *et al.*, 2010). Consequently, manual cell counting under the microscope is still the method of choice to quantify microbial eukaryotes.

Since microscopic counting is time consuming and operator-dependent, tools have been developed to automate some steps, with significant advances in image analysis (Bloem *et al.*, 1995; Singleton *et al.*, 2001; Abramoff *et al.*, 2004; Selinummi *et al.*, 2005; Thiel and Blaut, 2005; Zhou *et al.*, 2007) and image acquisition with a motorized microscope (Pernthaler *et al.*, 2003). The latter automated system has been applied to quantify specific-labelled prokaryotic taxa with fluorescence *in situ* hybridization (FISH) in thousands of samples from laboratory experiments (Pernthaler *et al.*, 2003) and oceanographic surveys (Schattenhofer *et al.*, 2009). Further developments of the system have been made to minimize human intervention (Zeder and Pernthaler, 2009; Zeder *et al.*, 2011) and to facilitate and speed the whole process (Bennke *et al.*, 2016). So far, this automated counting system has only been used for prokaryotes, and its application to count protist cells requires some adjustments, as protist are orders of magnitude less abundant than their prokaryotic counterparts.

In the present work, we optimized the above automated cell counting system on three uncultured protist taxa of different cell size, MAST-1C (~5 μm) and MAST-4 and MAST-7 (~2 μm). These groups were selected because they are widely found in sequencing surveys

82 and FISH probes are already publicly available for their detection (Massana *et al.*, 2002,
83 2006; Giner *et al.*, 2016). We first optimized the counting routine for pico- and
84 nanoeukaryotes using coastal and enrichment samples from past projects. Then, we applied
85 this method on the Malaspina 2010 circumglobal expedition where 129 stations were sampled
86 at two depths in the photic zone of the Atlantic, Indian and Pacific Oceans (Supporting
87 Information Fig. S1). We show here the spatial distribution of the three MAST groups in
88 parallel with that of small pigmented and colourless protists ($< 5 \mu\text{m}$) and report their ubiquity
89 and constancy in cell numbers. Our approach constitutes the first attempt to describe the
90 abundance of several small protist components in epipelagic oceanic waters using state-of-
91 the-art automated microscopy.

92

93 **Results**

94 *Optimizing the automated image acquisition and analysis for FISH counts*

95 To calibrate the automated enumeration approach for small protists, we selected samples from
96 past coastal surveys (Giner *et al.*, 2016) and enriched natural communities (del Campo *et al.*,
97 2013). The manually counted abundances of MAST-1C varied from 0 to 304 cells ml^{-1}
98 (median = 6) while MAST-4 densities ranged from 4 to 1145 cells ml^{-1} (median = 261;
99 Supporting Information Table S1). We first identified the settings for the best distinction
100 between true FISH-positive cells and undesired signal by combining DAPI and FISH images.
101 FISH-positive cells were those fulfilling the following criteria: 1) an object area between 100
102 and 2000 pixels and a MGVP90 (mean grey value of the darker pixels) > 200 in the DAPI
103 image, 2) an object area between 300 and 3000 pixels, an SBR (signal background ratio) > 3

and a $MGVp90 > 250$ in the FISH image, and 3) an object overlap between DAPI and FISH images $>75\%$.

Then we evaluated the number of fields of view (FOVs) needed to reach an accurate estimate by plotting the mean cell abundance and its standard error obtained by increasing the number of FOVs considered. We illustrated this with MAST-4 counts obtained in three coastal samples with contrasting cell abundances (Supporting Information Fig. S2). The sample with the highest MAST-4 density required less images to have an accurate cell count. In the sample with the lowest abundance, a fairly constant estimate was obtained after considering 200 FOVs. Similarly, about 180 FOVs were required to obtain a reliable estimate of MAST-1C abundance (Supporting Information Fig. S2).

Validation of the automated enumeration of small protists

We compared the automated counts of MAST-1C and MAST-4 cells with published manual counts (Supporting Information Table S1). Manual and automated MAST-4 counts for coastal and enrichment samples analysed separately (data not shown) and together were strongly correlated ($r = 0.966$, $P < 0.001$; Fig. 1A). MAST-1C counts displayed similar results ($r = 0.996$, $P < 0.001$; Fig. 1B). Despite these good correlations, the slopes of the linear regression were significantly greater than 1 (about 1.3 in both cases), indicating a slight underestimation of automated counting (Figs. 1A and 1B). In order to understand this, we counted MAST cells by visually inspecting the acquired images, avoiding then the image analysis step. In doing so, we observed a strong and significant correlation between automated counts and those derived from manually inspecting the images in both MAST-4 ($r = 0.991$, $P < 0.001$; Fig. 1A) and MAST-1C counts ($r = 0.999$, $P < 0.001$; Fig. 1B). Although still significantly greater than 1, the slopes were a bit smaller than before (~ 1.1 in both cases), indicating that the visual inspection of the images allowed to detect few cells missed by the image analysis (Figs. 1A

and 1B). These “missed” cells were explained by two main factors: i) the overlap of two or more FISH-positive cells that create an object area beyond the limits defined for a single cell and ii) FISH-positive cells placed at the image edge and escaping automated detection.

Next, we investigated the replicability of the automated counting by performing independent image capture sets (different FOVs) on the same filter section. Triplicate counts on MAST-4 cells in coastal samples covering a broad range of cell densities performed very well, as shown by the small standard errors in each sample (Supporting Information Fig. S3). Coefficients of variation (CV) were typically below 10% for samples above 35 cells ml⁻¹ and were larger and variable for low-abundance samples, ranging from 47% (Varna’10 DCM, 28 cells ml⁻¹) to 111% (Varna’10 Anoxic, 4 cells ml⁻¹).

The previous analysis showed that a given filter section yielded a consistent value, but another section from the same filter could provide a different estimate if cell distribution was uneven. Thus, we investigated the distribution of MAST-4 and MAST-7 cells on entire filters by processing about 3000 FOVs per filter (~15% of the total area). The abundance of both groups differed in the test sample (380 cells ml⁻¹ for MAST-4 and 60 cells ml⁻¹ for MAST-7) but their respective cells were evenly distributed in the entire filters (Fig. 2A). This was illustrated by calculating cell abundance in 10 virtually cut filter sections (Fig. 2A). In each virtual section, the abundance of MAST-4 varied from 315 to 427 cells ml⁻¹, while that of MAST-7 ranged from 49 to 68 cells ml⁻¹. This represented CV of 9.5% for MAST-4 and 9.2% for MAST-7, variation which became smaller when only five or two virtual filter sections were considered (Fig. 2B).

Epipelagic abundance of picosized and small nanosized protists over the oceans

We aimed to count phototrophic and heterotrophic eukaryotes smaller than 5 μm by standard epifluorescence microscopy in all epipelagic stations visited during the Malaspina circumglobal expedition, but some preparations had lost the chlorophyll red autofluorescence during storage and were not processed. In the end, we obtained counts for 94 surface samples and 114 DCM (Deep Chlorophyll Maximum) samples (Table 1). Overall, small protists were ubiquitous and exhibited only minor variations in cell abundance, with averaged values of 1942 (± 166) cells ml^{-1} at surface and 5227 (± 507) cells ml^{-1} at DCM depths (Table 1). Pigmented and colourless cells also showed this constancy in abundance in both depths (Fig. 3A). In general, heterotrophs were less abundant than phototrophs both at surface (mean of 385 vs 1557 cells ml^{-1}) and DCM depths (mean of 799 vs 4440 cells ml^{-1}) (Table 1). Consequently, the heterotrophic community accounted on average for 20-25% of total small eukaryotes in the sunlit water (Fig. 3B). When comparing the two depths, about 2-3 times more cells were detected at the DCM than at surface (Fig. 3A, Table 1). As expected, phototrophic cells represented a larger proportion of small protists ($\sim 85\%$) at the DCM (Fig. 3B). The abundances of small protists also exhibited slight variations when considering the different oceans and geographical regions visited during the cruise (Supporting Information Fig. S1). Globally, phototrophs and heterotrophs were more abundant in the Pacific Ocean with, respectively, about 1.4 and 1.9 times more cells in these waters (Table 1). They notably reached highest abundances in the North Pacific, where they averaged 9332 phototrophs ml^{-1} and 1290 heterotrophs ml^{-1} (Table 1, Supporting Information Table S2). On the other hand, lowest abundances of both phototrophs and heterotrophs were found in the Indian and South Atlantic Oceans (Table 1, Supporting Information Table S2).

Automated counts of three MAST lineages at a wide geographical scale

173 We obtained automated counts of three MAST lineages in surface (n=128) and DCM (n=122)
174 samples taken during the Malaspina cruise. The three taxa were found in almost all samples as
175 we were unable to detect MAST-1C, MAST-4 and MAST-7 cells in only 4, 7 and 20
176 epipelagic samples, respectively (Supporting Information Table S2). Globally, the two
177 epipelagic layers displayed very similar averaged abundances: $7 (\pm 1)$ and $6 (\pm 1)$ cells ml^{-1}
178 for MAST-1C, $51 (\pm 5)$ and $47 (\pm 6)$ cells ml^{-1} for MAST-4 and $24 (\pm 3)$ and $22 (\pm 2)$ cells ml^{-1}
179 for MAST-7 in surface and DCM, respectively (Table 1). Despite similar cell numbers of
180 the three groups in the two depths (Fig. 3C), their contribution to total heterotrophic protists
181 was larger at the surface than at the DCM (Fig. 3D). The averaged proportions of MAST-1C,
182 -4 and -7 cells with respect to total heterotrophs were 2.4, 15.8 and 6.6% at the surface, and
183 1.0, 6.9 and 3.1% at the DCM (Fig. 3D). Altogether, the three MAST groups can locally reach
184 very high contributions, with a maximal value of 85.3% of heterotrophs in a surface sample of
185 the Equatorial Pacific (Fig. 4A).

186 With respect to their geographic distribution, the three MAST groups displayed slight
187 differences (Table 1, Figs. 4C and 4D). MAST-1C, the least abundant of the three groups,
188 maintained relative similar abundances at the three visited oceans, while MAST-4 and
189 MAST-7 were locally variable (Table 1, Figs. 4C and 4D). MAST-4 exhibited highest
190 abundances in the Indian Ocean (near the African coast) both at surface ($264 \text{ cells ml}^{-1}$) and
191 DCM waters ($503 \text{ cells ml}^{-1}$), whereas MAST-7 was more abundant in the North Pacific, with
192 maximal values of $171 \text{ cells ml}^{-1}$ recorded at surface and $152 \text{ cells ml}^{-1}$ at the DCM (Figs. 4C
193 and 4D). Nevertheless, these peaks were uncommon and MAST-4 and MAST-7 abundances
194 remained relatively stable and within a relatively small range over the cruise (Table 1, Figs.
195 4C and 4D).

Abundance spectra and factors explaining the global distribution of small eukaryotes and MASTs

We explored the abundance spectra of small protists and the three MASTs groups in the epipelagic zone (surface and DCM combined) by comparing their logarithmic abundances with a set of statistical distribution models. Globally, we observe that all groups better fitted to normal-like models (bell-shaped curve), with yet slight differences in the width of the bell expressed by the standard deviation (σ) estimate (Fig. 5). The logarithmic abundance spectra of both phototrophic and heterotrophic protists followed a unimodal gamma distribution, with a wider variation in phototrophs ($\sigma = 0.39$) than in heterotrophs ($\sigma = 0.29$) (Fig. 5A).

Regarding the MASTs, MAST-4 followed a normal distribution while MAST-1C and MAST-7 better fitted to a Weibull distribution (Fig. 5B). The width of the bell also varied among groups, with MAST-4 and -7 showing a wider dispersion ($\sigma = 0.51$ - 0.53) than MAST-1C ($\sigma = 0.33$). In normal distributions most values (~68% for a standard normal model) are within one standard deviation of the mean. Applying this to our data, it shows that the abundance of small protists and MASTs in the sunlit ocean generally varied within one order of magnitude (Fig. 5).

We further performed a constrained ordination analysis (redundancy analysis, RDA) to explore whether the abiotic and biotic variables collected during the cruise could explain the variations in abundance of small phototrophic and heterotrophic eukaryotes and of the three MAST groups (Supporting Information Fig. S4). With respect to the small protists, a minor part of the variability (24.8%) was explained by the first two axes (adjusted $R^2 = 0.246$, $P < 0.001$; Supporting Information Fig. S4A). The factors that better explained the variations in phototrophic abundances were the concentration of chlorophyll a and phosphate, while changes in heterotrophic abundances were related to water temperature (Supporting

Information Fig. S4A). On the other hand, the RDA analysis conducted on the three MAST groups revealed a non-significant correlation with environmental parameters (adjusted $R^2 = 0.014$, $P = 0.07$; Supporting Information Fig. S4B).

Finally, we explored the relationship between the abundances of MAST-4 and MAST-7 in each sample (Supporting Information Fig. S5) to see whether they compete for the same ecological niche or if they follow the same environmental drivers, as these taxa have a similar cell size. These two taxa showed a weak but significant positive correlation when combining all samples ($r = 0.41$, $P < 0.001$, $n=250$; Supporting Information Fig. S5B). Considering each ocean separately (Supporting Information Fig. S5A), the relationship was very strong in the Indian Ocean ($r = 0.77$, $P < 0.001$, $n=65$), less clear in the Atlantic Ocean ($r = 0.51$, $P < 0.001$, $n=109$), and absent in the Pacific Ocean ($r = 0.11$, $P = 0.33$, $n=76$). The slope of the linear regression was ~ 1 in the Atlantic and ~ 2 in the Indian Ocean (Supporting Information Fig. S5A), meaning that MAST-4 was equal or twice more abundant than MAST-7 in these oceans.

Discussion

A high-throughput automated cell counting mastered for pico- and nanoeukaryotes

In this study, we tested and optimized an automated image acquisition and image analysis system for the enumeration of planktonic pico- and nanoeukaryotes. Test samples covered the broad range of cell densities typically found in marine systems (from 1 to 10^3 cells mL^{-1}) and included enrichment samples with different microbial life stages that might have variable rRNA content. Our routine provided accurate counts with both natural and enrichment samples, likely because we used an enzymatic signal amplification step during TSA-FISH

that increased 20-40 times the fluorescence intensity of target cells (Schönhuber *et al.*, 1997; Biegala *et al.*, 2003). Overall, we propose a consensus of settings for enumerating FISH-positive cells, both picosized (MAST-4 and MAST-7) and nanosized (MAST-1C), which appear of similar size (~700 pixels per cell) when acquired at different magnifications. We believe that this approach can be applied to other small eukaryotes.

In contrast to prokaryotic taxa that require about 10 high-quality images to obtain statistically relevant counts (Bennke *et al.*, 2016), many more images are needed to estimate protist abundance. We showed that acquiring about 200 high-quality images was sufficient for an accurate estimation of the abundance of nanosized or picosized eukaryotes. This is consistent with the area inspected during manual epifluorescence microscopic counting of small eukaryotes (Pernice *et al.*, 2015), often based in a transect of about 20 mm at 1000x (i.e., 200 FOVs). We also analysed the distribution of cells on an entire membrane filter. The relative homogeneous cell distribution on the filter supports the practice of using a filter section for FISH. Commonly, 25 mm membranes are cut in 8-12 sections to perform separate counts of microbial taxa by FISH (Not *et al.*, 2002; Karl, 2007), and we show here that counts performed with separate filter sections are reasonably similar.

Besides its replicability and accuracy, the automated counting is faster and saves time for the operator. Once filter sections are hybridized, up to 100 samples per week can be processed by the automated image acquisition and analysis, while a single expert operator dedicating 6 hours a day on manual counting would process up to 30 samples per week. This system opens up new possibilities of addressing the global abundance of specific microbial taxa, as exemplified here with the samples from the Malaspina global expedition (Duarte, 2015). Our automated routine is well suited to count FISH-labelled microbial cells that have regular object properties and allows assessing cell biovolumes as well, as reports of parameters useful

for this calculation (object area, perimeter and circularity) are generated for each recognized cell. Our automated method can also be adapted to dual FISH assays, which couples the quantification of predator/prey and parasite/host populations, as long as homogenous cell characteristics can be defined in the groups of interest. Therefore, it has a high potential to assess specific functional diversity and responses of specific grazers or parasites in natural environments. In the future, we envision the application of our automated method to count the whole small protist community, a more difficult task as this includes a mix of cells with variable size, shape and pigmentation, making harder the generalization of object features for image analysis. Meanwhile, the method of choice to quantify phototrophic and heterotrophic cells is still manual counting.

Worldwide distribution of epipelagic small protists and MASTs

The unprecedented scale and geographical coverage of our sampling effort allowed a comprehensive investigation of the abundance of small protists and three of their constituting taxa in the epipelagic zone of the three main oceans. The abundance and relative contribution of phototrophic and heterotrophic protists in the two layers was within the range previously observed: $0.3 - 8.0 \times 10^3$ at surface and $0.6 - 46.0 \times 10^3$ cells ml^{-1} at DCM for phototrophs, and $0.1 - 2.0 \times 10^3$ at surface and $0.2 - 5.0 \times 10^3$ cells ml^{-1} at DCM for heterotrophs, the latter representing 10-30% of protists in upper marine waters (Sanders *et al.*, 2000; Jürgens and Massana, 2008; Lepère *et al.*, 2009; Li, 2009; Cabello *et al.*, 2016). Despite some spots of exceptionally high abundances, particularly in the North Pacific, a relative constancy of these communities along the cruise track was evident, and this could be explained by the relative uniformity of hydrological conditions. Thus, temperature at both surface and DCM depths averaged 22.7°C (CV = 17.3%) and salinity averaged 35.7 psu (CV = 2.4%) (Supporting Information Table S2). One reason for such uniformity is that about 75% of the stations were

in the Trades Biome (Supporting Information Fig. S1), known as having a strong permanent thermocline and a weak seasonality (Longhurst, 2007).

Within heterotrophic protists, we investigated the abundance in the epipelagic global ocean of three uncultured MAST groups with the automated counting approach. The three groups showed a ubiquitous distribution (present in >90% of the samples) and only slight variations in cell abundances. With ~ 50 cells ml^{-1} on average, MAST-4 was the most abundant of the three taxa and accounted for 15.8% of heterotrophic cells at surface and 6.9% at DCM. Previous studies reported the absence of MAST-4 cells in waters colder than 5°C (Massana *et al.*, 2006; Rodríguez-Martínez *et al.*, 2009) and the impact of temperature in the distribution patterns of MAST-4 subclades (Rodríguez-Martínez *et al.*, 2013; Seeleuthner *et al.*, 2018). We did not find an effect of temperature on MAST-4 spatial dynamics, probably due to the narrow range of the water temperature sampled here (Supporting Information Table S2). Our data supports, with a much larger sampling effort, previous observations of MAST-4 being widespread and abundant in surface waters (Massana *et al.*, 2006). Thus, MAST-4 is confirmed as one of the most abundant heterotrophic protist taxa in sunlit oceans, where it likely plays an important role in food-web dynamics and nutrient re-mineralization as a bacterial grazer (Massana *et al.*, 2006, 2009).

The second most abundant group was MAST-7, which averaged 22-24 cells ml^{-1} and represented 6.6 % of heterotrophic cells at surface and 3.1% at DCM. A recent study (Giner *et al.*, 2016) showed higher cell densities in coastal samples (160 cells ml^{-1} on average) and our study is the first investigating its abundance in the open sea. Little is known about the ecological role of MAST-7 and, to the best of our knowledge, only one study has caught a glimpse of its activity as a putative *Prochlorococcus* grazer (Frias-Lopez *et al.*, 2009). Given the similar small size and phylogenetic proximity with MAST-4, known to be bacterivorous

(Massana *et al.*, 2002, 2009) and occasionally picoalgalivorous (Massana *et al.*, 2009), it is likely that MAST-7 is a bacterial grazer as well. MAST-7 and -4 cells were positively correlated, especially in the Indian and Atlantic Oceans, confirming that they follow the same environmental drivers and suggesting that competition among them was not a major driving force. Finally, MAST-1C was also ubiquitous but at lower densities than the other two taxa (<10 cells ml^{-1}). This group represented 2.4% (surface) to 1.0% (DCM) of heterotrophic protists, values similar to those previously reported in temperate waters (Massana *et al.*, 2006). Much higher cell abundances (up to 200 cells ml^{-1}) were reported in Antarctic waters (Massana *et al.*, 2006), showing a putative preference of MAST-1C for cold waters. Although it is assumed that most MAST clades are free-living bacterivorous flagellates (Piwosz *et al.*, 2013; Massana *et al.*, 2014), we haven't yet observed any significant correlation between their spatial dynamics and the total prokaryotic abundance, most likely because the abundance spectra of marine bacteria in our samples is relatively similar to that observed for the MAST groups. Seasonal variation of bacterioplankton community structure is well known (Ghiglione *et al.*, 2005; Mary *et al.*, 2006; Gilbert *et al.*, 2012; Fuhrman *et al.*, 2015) and correlations between specific bacterial and MAST clades, indicative of putative feeding preferences among specific MASTs, could be envisioned.

The abundances of phototrophic and heterotrophic protists and of the MAST taxa were relatively constant along the main oceans investigated, showing a typical unimodal lognormal distribution (i.e., log-transformed abundances normally distributed). This distribution is commonly used to describe taxa/species abundance data, and emerges from the multiplicative interactions of stochastic processes that shape biodiversity (Preston, 1948; McGill *et al.*, 2007; Shoemaker *et al.*, 2017). Specifically, it has been associated to microbial species abundance curves, where many taxa are rare and a few very abundant, and has allowed to provide theoretical diversity estimates (Curtis *et al.*, 2002; Prosser *et al.*, 2007). More

recently, lognormal distributions have been fitted to the abundance profiles of widely distributed prokaryotic Operational Taxonomic Units (Niño-García *et al.*, 2016). Our findings show a broader application of the model and suggest that lognormal dynamics underpin the fundamental nature of microbial communities (Prosser *et al.*, 2007; Shoemaker *et al.*, 2017). Our data highlights the ubiquity and uniformity in the abundance pattern of several protist components in sunlit oceans, indicative of high tolerance to putative environmental changes (Niño-García *et al.*, 2016). Slight differences in the width of the lognormal distribution were observed among the groups, with abundance values being more dispersed in MAST-4 and MAST-7 than in MAST-1C, phototrophs and heterotrophs. Indeed, photosynthetic and heterotrophic protists include many different species, each one responding differently to variation in environmental conditions (such as the three MASTs groups), but collectively reacting in unison to these changes. Overall, the fact that small protists and MAST cells have near-constant abundances over large oceanic areas indicate that they are part of the microbial ocean's veil as defined by Smetacek (2002). Under this scenario, the small size of these marine components would allow efficient dispersion and fast growth controlled by equally fast-growing predators (Smetacek, 2002; Durham *et al.*, 2009; Massana and Logares, 2013). Along the cruise track, the abundance of small protists and MASTs generally varied by a factor of 10, similar to the variation observed among bacterial populations. This contrasts with larger protists like diatoms, dinoflagellates (Brussaard *et al.*, 1996; Leblanc *et al.*, 2012; Paterson *et al.*, 2017) or ciliates (Vaqué *et al.*, 1997; Santoferrara and Alder, 2009), which have orders of magnitude variation in their local abundances, exhibiting sporadic blooms in which growth and mortality are decoupled (Smetacek, 2002). In conclusion, the ubiquity and near-constant abundances of small protists and of some of their taxonomic components in sunlit oceans exemplifies their ecological success and their pivotal role in microbial food webs.

365

366 **Experimental procedures**367 *Sampling*

368 For the calibration of the automated microscopic cell counting, we used data from past
369 projects where we did manual FISH counts (Supporting Information Table S1). This includes
370 natural samples from two depth layers in six coastal European sites from the BioMarKs
371 consortium (Giner *et al.*, 2016) and modified assemblages of heterotrophic protists from the
372 Blanes Bay Microbial Observatory (BBMO) (del Campo *et al.*, 2013). In the latter, surface
373 seawater was prefiltered by 3 µm and incubated in 8L containers for 8 days in the dark,
374 resulting in the growth and enrichment of heterotrophic pico-sized protist cells. An extra
375 surface sample from BBMO was collected on June 6th, 2017 to assess the distribution of cells
376 on filters.

377 Samples from the three major oceans (Supporting Information Fig. S1) were collected in 129
378 stations visited from December 2010 to July 2011 during the Malaspina 2010 expedition
379 (Duarte, 2015). Two depths (surface and DCM) were sampled with Niskin bottles attached to
380 a rosette equipped with a Seabird 911Plus CTD probe that measured temperature,
381 fluorescence and salinity (Supporting Information Fig. S2). Nutrients (nitrate, phosphate and
382 silicate) were determined as explained in Catalá *et al.* (2016). Chlorophyll *a* concentration
383 was measured with a Turner Designs fluorometer (Estrada *et al.*, 2016) and bacterial
384 abundance was determined by flow cytometry as described in Lara *et al.* (2017). Bacterial
385 biomass and bacterial heterotrophic production (³H-leucine incorporation method) were
386 determined as explained in Morán *et al.* (2017).

387 *Epifluorescence microscopy counts by DAPI staining*

Samples for picoeukaryotic cell counts were prefiltered by gravity through a 200 μm nylon mesh, fixed with ice-cold glutaraldehyde (1% final concentration), and left for 1 to 24 h in the dark at 4°C. Then, aliquots of 30 ml were filtered through 0.6 μm pore-size black polycarbonate filters (25 mm diameter; DHI; Hørsholm, Denmark) and stained with 0.5 $\mu\text{g ml}^{-1}$ DAPI (4',6-diamidino-2-phenylindole). Preparations were mounted on a slide with low-autofluorescence oil and kept at -80°C. Filters were observed with an epifluorescence microscope (Olympus BX61, Olympus America Inc., Center Valley, PA, USA) at 1000x magnification under ultraviolet excitation (360 ± 25 nm) for DAPI-stained DNA blue fluorescence and blue excitation (465 ± 15 nm) for chlorophyll red autofluorescence to discriminate between pigmented and colourless protists. Counted cells were classified into three size classes: <3 μm , 3-5 μm and >5 μm . Data shown here refer to the two smaller classes (≤ 5 μm), which accounted for the large majority of the cells (97% on average).

Quantification by TSA-FISH (FISH coupled with Tyramide Signal Amplification)

Aliquots of 100 ml of 200 μm -prefiltered seawater were fixed with formaldehyde (3.7% final concentration), left for 1 to 24 h in the dark at 4°C, filtered through 0.6 μm pore-size polycarbonate filters (25 mm diameter; DHI), and kept at -80°C until processed. TSA-FISH was carried out on filter sections (1/10 of the filter) as described in Pernice *et al.* (2015) and Giner *et al.* (2016). Three probes were used (Supporting Information Table S3): NS1C targeting MAST-1C (Massana *et al.*, 2006), NS4 targeting MAST-4 (Massana *et al.*, 2002), and NS7 targeting MAST-7 (Giner *et al.*, 2016), the latter used with oligonucleotide helpers. Manual counts of hybridized cells were done by exciting with blue light and observing about 40 randomly chosen fields (40-200 cells in total) or, when cell densities were lower than 1 cell per field, from a transect across the filter of about 13 mm (equivalent to 130 fields, about 30 to 300 cells in total).

412 *Automated microscopy: image acquisition and image analysis*

413 The automated analysis is adapted from Bennke *et al.* (2016) and uses the motorized
414 microscope ZEISS Axio Imager.Z2 (Carl Zeiss MicroImaging GmbH, Göttingen, Germany)
415 equipped with a Colibri.2 LED light source (Carl Zeiss) and a HE-62 multifilter module (Carl
416 Zeiss). Four filter sections were mounted with an antifading mix in a single slide and up to 8
417 slides were placed at the microscopic stage. A first overview picture of the microscope stage
418 was taken in bright-field illumination with the 1x objective and the SamLoc 1.4 software
419 (Zeder *et al.*, 2011). This step allows the definition of coordinates (i.e., fields of view, FOVs)
420 in each filter piece for image acquisition.

421 Image acquisition was performed with the MPISYS program (modified after Zeder *et al.*
422 (2011)) implemented in the AxioVision software 4.8.2 (Carl Zeiss) using the channels defined
423 for DAPI and Alexa Fluor 488 dyes. For each channel, parameters defined by the user were
424 the objective, excitation light, exposure time, focusing procedure, and number of images per
425 z-stack to compensate for filter unevenness. Image acquisition was performed with the 40x
426 objective for MAST-1C cells (~5 μm) and the 63x objective for MAST-4 and MAST-7 cells
427 (~2 μm), resulting in similar cell sizes in the images in the three cases. In each FOV, focusing
428 was first done on the DAPI channel (UV excitation: 385 ± 4.5 nm) and set as a fixed focal
429 position for the FISH channel (blue light excitation: 470 ± 14 nm). Images were taken at 10-
430 25 ms in the DAPI channel and at 50-100 ms in the FISH channel. A single extended depth of
431 field image (EDF) resulting from a z-stack of 7 layers was created and saved per DAPI and
432 FISH signals.

433 Final EDF images were first quality-controlled using the Automated image quality control
434 (AIQC) software (Zeder *et al.*, 2010), and then by a fast visual inspection. Low-quality
435 images with over- or underexposed parts or areas out of focus were removed (typically <10%

of the set). Cell detection was done with the Automated Cell Measuring and Enumeration tool 2.0 (ACMEtool2.0; www.technobiology.ch) software. Parameters of all detected objects in the image (area, length, width, mean grey value [MGV], mean grey value of the darker pixels [MGVp90], signal to background ratio [SBR]) were evaluated by manual cross-checks to determine the values defining FISH-positive cells. Then, cells were automatically counted by ACMEtool2.0, producing a report of the number of cells per FOV in each filter section. A final manual count from the pictures was done in samples with very low (0-5) or very high (100-200) number of cells per filter section.

Data analyses

An iterative procedure adapted from Niño-García *et al.* (2016) was done to find out the model that best fits the abundance distribution of the targeted groups. Since abundance data was log-transformed and some fitting models require non-zero and positive values, samples with less than 1 cell ml⁻¹ were removed from the analysis. The Hartigan's dip test statistic (Maechler, 2015) implemented in the dip test package in R 3.3.2. (R Core Team, 2015) was first computed to assess the non-unimodality of the abundance distribution. According to this test, all groups had a unimodal distribution ($P > 0.01$). Fits to unimodal models (normal, Weibull, logistic, gamma, lognormal, Cauchy and exponential) were computed by maximum likelihood with the fitdist function of the fitdistrplus package in R (Delignette-Muller and Dutang, 2014). The best fitting model was the one with the lowest Akaike Information Criterion (AIC) value.

To investigate the relationship between the distribution of each group and biotic or abiotic factors, redundancy analysis (RDA) (van den Wollenberg, 1977) was performed with the vegan v2.4-0 package in R (Oksanen *et al.*, 2016). Prior to analysis, the abundance of each group was standardized by Hellinger transformation (Legendre and Gallagher, 2001). Abiotic

factors considered were nitrate (NO₃), phosphate (PO₄), silicate (SiO₂) and temperature, whereas biotic factors were chlorophyll *a*, bacterial biomass, production and abundance. For each RDA, forward selection (Blanchet *et al.*, 2008) with 999 Monte Carlo permutation tests was performed to identify the set of parameters that best explain group variation.

Acknowledgements

This work was supported by the Spanish projects MEFISTO (CTM2013-43767-P, MINECO) and ALLFLAGS (CTM2016-75083-R, MINECO) and by a call that funded the motorized microscope (CSIC15-EE-5379). JFM was beneficiary of a Marie Curie Fellowship (PIEF-GA-2012-331190, EU). Samples were collected during the Malaspina 2010 circumnavigation expedition (grant CSD2008-00077, MINECO). We thank the crew and chief scientists on board of R/V Hesperides and our fellows R. MacKenzie, L. Alonso-Sáez, M.M. Varela, M. Mestre, E. Teira, V. Hernando-Morales and J. Valencia for collecting the FISH filters, and J.M. Gasol and X.A.G. Morán for bacterial data. We are grateful to M. Zeder (Roche Diagnostics, Rotkreuz, Switzerland) for his help in the installation and use of the automated microscope. We finally thank C. Ruiz-González and J.P. Niño-García for their advices on computing model distributions. The authors declare no conflict of interest.

References

- Abràmoff, M.D., Magalhães, P.J., and Ram, S.J. (2004) Image processing with Image J. *Biophotonics Int.* **11**: 36–42.
- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009) A method

- 482 for studying protistan diversity using massively parallel sequencing of V9 hypervariable
483 regions of small-subunit ribosomal RNA genes. *PLoS One* **4**: e6372.
- 484 Bennke, C.M., Reintjes, G., Schattenhofer, M., Ellrott, A., Wulf, J., Zeder, M., *et al.* (2016)
485 Modification of a high-throughput automatic microbial cell enumeration system for
486 shipboard analyses. *Appl. Environ. Microbiol.* **82**: 3289–3296.
- 487 Biegala, I.C., Not, F., Vaultot, D., and Simon, N. (2003) Quantitative assessment of
488 picoeukaryotes in the natural environment by using taxon-specific oligonucleotide
489 probes in association with tyramide signal amplification-fluorescence *in situ*
490 hybridization and flow cytometry. *Appl. Environ. Microbiol.* **69**: 5519–5529.
- 491 Blanchet, G., Legendre, P., and Borcard, D. (2008) Forward selection of spatial explanatory
492 variables. *Ecology* **89**: 2623–2632.
- 493 Bloem, J., Veninga, M., and Shepherd, J. (1995) Fully automatic determination of soil
494 bacterium numbers, cell volumes, and frequencies of dividing cells by confocal laser
495 scanning microscopy and image analysis. *Appl. Environ. Microbiol.* **61**: 926–936.
- 496 Brussaard, C.P.D., Gast, G.J., van Duyl, F.C., and Riegman, R. (1996) Impact of
497 phytoplankton bloom magnitude on a pelagic microbial food web. *Mar. Ecol. Prog. Ser.*
498 **144**: 211–221.
- 499 Cabello, A.M., Latasa, M., Forn, I., Morán, X.A.G., and Massana, R. (2016) Vertical
500 distribution of major photosynthetic picoeukaryotic groups in stratified marine waters.
501 *Environ. Microbiol.* **18**: 1578–1590.
- 502 del Campo, J., Balagué, V., Forn, I., Lekunberri, I., and Massana, R. (2013) Culturing bias in
503 marine heterotrophic flagellates analyzed through seawater enrichment incubations.
504 *Microb. Ecol.* **66**: 489–499.
- 505 Catalá, T.S., Álvarez-Salgado, X.A., Otero, J., Iuculano, F., Companys, B., Horstkotte, B., *et*

- 506 *al.* (2016) Drivers of fluorescent dissolved organic matter in the global epipelagic ocean.
507 *Limnol. Oceanogr.* **61**: 1101–1119.
- 508 Curtis, T.P., Sloan, W.T., and Scannell, J.W. (2002) Estimating prokaryotic diversity and its
509 limits. *Proc. Natl. Acad. Sci.* **99**: 10494–10499.
- 510 Delignette-Muller, M.L. and Dutang, C. (2014) fitdistrplus : an R package for fitting
511 distributions. *J. Stat. Softw.* **64**: 1–34.
- 512 Duarte, C.M. (2015) Seafaring in the 21th century: the Malaspina 2010 circumnavigation
513 expedition. *Limnol. Oceanogr. Bull.* **24**: 11–14.
- 514 Durham, W.M., Kessler, J.O., and Stocker, R. (2009) Disruption of vertical motility by shear
515 triggers formation of thin phytoplankton layers. *Science*. **323**: 1067–1070.
- 516 Estrada, M., Delgado, M., Blasco, D., Latasa, M., Cabello, A.M., Benítez-Barrios, V., *et al.*
517 (2016) Phytoplankton across tropical and subtropical regions of the Atlantic, Indian and
518 Pacific Oceans. *PLoS One* **11**: e0151699.
- 519 Frias-Lopez, J., Thompson, A., Waldbauer, J., and Chisholm, S.W. (2009) Use of stable
520 isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface
521 waters. *Environ. Microbiol.* **11**: 512–525.
- 522 Fuhrman, J.A., Cram, J.A., and Needham, D.M. (2015) Marine microbial community
523 dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**: 133–146.
- 524 Ghiglione, J., Larcher, M., and Lebaron, P. (2005) Spatial and temporal scales of variation in
525 bacterioplankton community structure in the NW Mediterranean Sea. *Aquat. Microb.*
526 *Ecol.* **40**: 229–240.
- 527 Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., *et al.*
528 (2012) Defining seasonal marine microbial community dynamics. *ISME J.* **6**: 298–308.
- 529 Giner, C.R., Forn, I., Romac, S., Logares, R., de Vargas, C., and Massana, R. (2016)

- 530 Environmental sequencing provides reasonable estimates of the relative abundance of
531 specific picoeukaryotes. *Appl. Environ. Microbiol.* **82**: 4757–4766.
- 532 Jürgens, K. and Massana, R. (2008) Protist grazing on marine bacterioplankton. In *Microbial*
533 *Ecology of the Oceans*. Kirchman, D.L. (ed). Hoboken, USA: John Wiley & Sons, Inc.,
534 pp. 383–424.
- 535 Karl, D.M. (2007) Microbial nucleotide fingerprints in Nature. In *Methods for General and*
536 *Molecular Microbiology*. Reddy, C.A., Beveridge, T.J., Breznak, J.A., Marzluf, G.A.,
537 Schmidt, T.M., and Snyder, L.R. (eds). Washington DC, USA: ASM Press, pp. 869–878.
- 538 Lara, E., Vaqué, D., Sà, E.L., Boras, J.A., Gomes, A., Borrull, E., *et al.* (2017) Unveiling the
539 role and life strategies of viruses from the surface to the dark ocean. *Sci. Adv.* **3**:
540 e1602565.
- 541 Leblanc, K., Aristegui, J., Armand, L., Assmy, P., Beker, B., Bode, A., *et al.* (2012) A global
542 diatom database – abundance, biovolume and biomass in the world ocean. *Earth Syst.*
543 *Sci. Data* **4**: 149–165.
- 544 Legendre, P. and Gallagher, E.D. (2001) Ecologically meaningful transformations for
545 ordination of species data. *Oecologia* **129**: 271–280.
- 546 Lepère, C., Domaizon, I., Taïb, N., Mangot, J.-F., Bronner, G., Boucher, D., *et al.* (2013)
547 Geographic distance and ecosystem size determine the distribution of smallest protists in
548 lacustrine ecosystems. *FEMS Microbiol. Ecol.* **85**: 85–94.
- 549 Lepère, C., Vaultot, D., and Scanlan, D.J. (2009) Photosynthetic picoeukaryote community
550 structure in the South East Pacific Ocean encompassing the most oligotrophic waters on
551 Earth. *Environ. Microbiol.* **11**: 3105–3117.
- 552 Li, W. (2009) From cytometry to macroecology: a quarter century quest in microbial
553 oceanography. *Aquat. Microb. Ecol.* **57**: 239–251.

- 554 Lin, Y.-C.C., Campbell, T., Chung, C.-C.C., Gong, G.-C.C., Chiang, K.-P.P., and Worden,
555 A.Z. (2012) Distribution patterns and phylogeny of marine stramenopiles in the North
556 Pacific Ocean. *Appl. Environ. Microbiol.* **78**: 3387–3399.
- 557 Logares, R., Audic, S., Santini, S., Pernice, M.C., de Vargas, C., and Massana, R. (2012)
558 Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled
559 with pyrosequencing. *ISME J.* **6**: 1823–1833.
- 560 Longhurst, A.R. (2007) Biomes: The primary partition. In *Ecological Geography of the Sea*.
561 Longhurst, A.R. (ed). London, UK: Academic Press, pp. 89–99.
- 562 Maechler, M. (2015) diptest: Hartigan's dip test statistic for unimodality - Corrected . R
563 package version 0.75-7. Available at: <http://CRAN.R-project.org/package=diptest>.
- 564 Mangot, J.-F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G., *et al.* (2013)
565 Short-term dynamics of diversity patterns: evidence of continual reassembly within
566 lacustrine small eukaryotes. *Environ. Microbiol.* **15**: 1745–1758.
- 567 Mary, I., Cummings, D., Biegala, I., Burkill, P., Archer, S., and Zubkov, M. (2006) Seasonal
568 dynamics of bacterioplankton community structure at a coastal station in the western
569 English Channel. *Aquat. Microb. Ecol.* **42**: 119–126.
- 570 Massana, R., del Campo, J., Sieracki, M.E., Audic, S., and Logares, R. (2014) Exploring the
571 uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within
572 stramenopiles. *ISME J.* **8**: 854–866.
- 573 Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., *et al.* (2015) Marine
574 protist diversity in European coastal waters and sediments as revealed by high-
575 throughput sequencing. *Environ. Microbiol.* **17**: 4035–4049.
- 576 Massana, R., Guillou, L., Díez, B., and Pedrós-Alió, C. (2002) Unveiling the organisms
577 behind novel eukaryotic ribosomal DNA sequences from the ocean. *Appl. Environ.*

- 578 *Microbiol.* **68**: 4554–4558.
- 579 Massana, R. and Logares, R. (2013) Eukaryotic versus prokaryotic marine picoplankton
580 ecology. *Environ. Microbiol.* **15**: 1254–1261.
- 581 Massana, R., Terrado, R., Forn, I., Lovejoy, C., and Pedrós-Alió, C. (2006) Distribution and
582 abundance of uncultured heterotrophic flagellates in the world oceans. *Environ.*
583 *Microbiol.* **8**: 1515–1522.
- 584 Massana, R., Unrein, F., Rodríguez-Martínez, R., Forn, I., Lefort, T., Pinhassi, J., *et al.* (2009)
585 Grazing rates and functional diversity of uncultured heterotrophic flagellates. *ISME J.* **3**:
586 588–596.
- 587 McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., *et al.*
588 (2007) Species abundance distributions: moving beyond single prediction theories to
589 integration within an ecological framework. *Ecol. Lett.* **10**: 995–1015.
- 590 Medinger, R., Nolte, V., Pandey, R.V., Jost, S., Ottenwälder, B., Schlötterer, C., *et al.* (2010)
591 Diversity in a hidden world: Potential and limitation of next-generation sequencing for
592 surveys of molecular diversity of eukaryotic microorganisms. *Mol. Ecol.* **19**: 32–40.
- 593 Morán, X.A.G., Gasol, J.M., Pernice, M.C., Mangot, J.-F., Massana, R., Lara, E., *et al.* (2017)
594 Temperature regulation of marine heterotrophic prokaryotes increases latitudinally as a
595 breach between bottom-up and top-down controls. *Glob. Chang. Biol.* **23**: 3956–3964.
- 596 Niño-García, J.P., Ruiz-González, C., and del Giorgio, P.A. (2016) Landscape-scale spatial
597 abundance distributions discriminate core from random components of boreal lake
598 bacterioplankton. *Ecol. Lett.* **19**: 1506–1515.
- 599 Not, F., Simon, N., Biegala, I., and Vaultot, D. (2002) Application of fluorescent *in situ*
600 hybridization coupled with tyramide signal amplification (FISH-TSA) to assess
601 eukaryotic picoplankton composition. *Aquat. Microb. Ecol.* **28**: 157–166.

- 602 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al.* (2016)
603 *vegan*: Community Ecology Package. R package version 2.4-0. Available at:
604 <https://CRAN.R-project.org/package=vegan>.
- 605 Paterson, R.F., McNeill, S., Mitchell, E., Adams, T., Swan, S.C., Clarke, D., *et al.* (2017)
606 Environmental control of harmful dinoflagellates and diatoms in a fjordic system.
607 *Harmful Algae* **69**: 1–17.
- 608 Pernice, M.C., Forn, I., Gomes, A., Lara, E., Alonso-Sáez, L., Arrieta, J.M., *et al.* (2015)
609 Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME J.* **9**:
610 782–792.
- 611 Pernthaler, J., Pernthaler, A., and Amann, R. (2003) Automated enumeration of groups of
612 marine picoplankton after fluorescence *in situ* hybridization. *Appl. Environ. Microbiol.*
613 **69**: 2631–2637.
- 614 Piwosz, K., Wiktor, J.M., Niemi, A., Tatarek, A., and Michel, C. (2013) Mesoscale
615 distribution and functional diversity of picoeukaryotes in the first-year sea ice of the
616 Canadian Arctic. *ISME J.* **7**: 1461–1471.
- 617 Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology* **29**: 254–283.
- 618 Prokopowich, C.D., Gregory, T.R., and Crease, T.J. (2003) The correlation between rDNA
619 copy number and genome size in eukaryotes. *Genome* **46**: 48–50.
- 620 Prosser, J.I., Bohannon, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et*
621 *al.* (2007) The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* **5**:
622 384–392.
- 623 R Core Team (2015) R: A Language and Environment for Statistical Computing. Vienna,
624 Austria. <http://www.r-project.org/>.
- 625 Rodríguez-Martínez, R., Labrenz, M., del Campo, J., Forn, I., Jürgens, K., and Massana, R.

- 626 (2009) Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake
627 Passage and Mediterranean Sea assessed by real-time quantitative PCR. *Environ.*
628 *Microbiol.* **11**: 397–408.
- 629 Rodríguez-Martínez, R., Rocap, G., Salazar, G., and Massana, R. (2013) Biogeography of the
630 uncultured marine picoeukaryote MAST-4: temperature-driven distribution patterns.
631 *ISME J.* **7**: 1531–1543.
- 632 Sanders, R.W., Berninger, U.-G., Lim, E.L., Kemp, P.F., and Caron, D.A. (2000)
633 Heterotrophic and mixotrophic nanoplankton predation on picoplankton in the Sargasso
634 Sea and on Georges Bank. *Mar. Ecol. Prog. Ser.* **192**: 103–118.
- 635 Santoferrara, L. and Alder, V. (2009) Abundance trends and ecology of planktonic ciliates of
636 the south-western Atlantic (35–63°S): a comparison between neritic and oceanic
637 environments. *J. Plankton Res.* **31**: 837–851.
- 638 Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M. V, Tarran, G.A., and Pernthaler, J.
639 (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic
640 Ocean. *Environ. Microbiol.* **11**: 2078–2093.
- 641 Schönhuber, W., Fuchs, B., Juretschko, S., and Amann, R. (1997) Improved sensitivity of
642 whole-cell hybridization by the combination of horseradish peroxidase-labeled
643 oligonucleotides and tyramide signal amplification. *Appl. Environ. Microbiol.* **63**: 3268–
644 3273.
- 645 Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., *et al.*
646 (2018) Single-cell genomics of multiple uncultured stramenopiles reveals underestimated
647 functional diversity across oceans. *Nat. Commun.* **9**: 310.
- 648 Selinummi, J., Seppälä, J., Yli-Harja, O., and Puhakka, J.A. (2005) Software for
649 quantification of labeled bacteria from digital microscope images by automated image

- 650 analysis. *Biotechniques* **39**: 859–862.
- 651 Sherr, E. and Sherr, B. (2008) Understanding roles of microbes in marine pelagic food webs:
652 A brief history. In *Microbial Ecology of the Oceans*. Kirchman, D.L. (ed). Hoboken,
653 USA: John Wiley & Sons, Inc., pp. 27–44.
- 654 Shoemaker, W.R., Locey, K.J., and Lennon, J.T. (2017) A macroecological theory of
655 microbial biodiversity. *Nat. Ecol. Evol.* **1**: 0107.
- 656 Singleton, S., Cahill, J.G., Keith Watson, G., Allison, C., Cummins, D., Thurnheer, T., *et al.*
657 (2001) A fully automated microscope bacterial enumeration system for studies of oral
658 microbial ecology. *J. Immunoass. Immunochem.* **22**: 253–274.
- 659 Smetacek, V. (2002) Microbial food webs: The ocean's veil. *Nature* **419**: 565–565.
- 660 Thiel, R. and Blaut, M. (2005) An improved method for the automated enumeration of
661 fluorescently labelled bacteria in human faeces. *J. Microbiol. Methods* **61**: 369–379.
- 662 Vaqué, D., Blough, H.A., and Duarte, C.M. (1997) Dynamics of ciliate abundance, biomass
663 and community composition in an oligotrophic coastal environment (NW
664 Mediterranean). *Aquat. Microb. Ecol.* **12**: 71–83.
- 665 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., *et al.* (2015)
666 Eukaryotic plankton diversity in the sunlit ocean. *Science*. **348**: 1261605–1261605.
- 667 Vd'ačný, P., Bourland, W.A., Orsi, W., Epstein, S.S., Foissner, W., Vd'ačný, P., *et al.* (2012)
668 Genealogical analyses of multiple loci of litostomatean ciliates (Protista, Ciliophora,
669 Litostomatea). *Mol. Phylogenet. Evol.* **65**: 397–411.
- 670 van den Wollenberg, A.L. (1977) Redundancy analysis an alternative for canonical
671 correlation analysis. *Psychometrika* **42**: 207–219.
- 672 Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling,
673 P.J. (2015) Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles

- 674 of microbes. *Science*. **347**: 1257594–1257594.
- 675 Zeder, M., Ellrott, A., and Amann, R. (2011) Automated sample area definition for high-
676 throughput microscopy. *Cytom. Part A* **79**: 306–310.
- 677 Zeder, M., Kohler, E., and Pernthaler, J. (2010) Automated quality assessment of
678 autonomously acquired microscopic images of fluorescently stained bacteria. *Cytom.*
679 *Part A* **77**: 76–85.
- 680 Zeder, M. and Pernthaler, J. (2009) Multispot live-image autofocusing for high-throughput
681 microscopy of fluorescently stained bacteria. *Cytom. Part A* **75**: 781–788.
- 682 Zhou, Z., Pons, M.N., Raskin, L., and Zilles, J.L. (2007) Automated image analysis for
683 quantitative fluorescence *in situ* hybridization with environmental samples. *Appl.*
684 *Environ. Microbiol.* **73**: 2956–2962.
- 685 Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of picoeucaryotes in
686 marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol.*
687 *Ecol.* **52**: 79–92.

688

689 **Table and Figure legends**

690 **Table 1.** Mean abundance ($\pm$ standard error; cells ml⁻¹) of total small eukaryotes (cells $\leq 5 \mu\text{m}$)
 691 and the three MAST groups across the different biogeographic areas sampled.

692 **Fig. 1.** Comparison between manual and automated cell counts of MAST-4 (A) and MAST-
 693 1C (B) in test samples. Automated cell counts were compared with manual counts done at the
 694 microscope (black triangles, solid lines) and from visual inspection of the pictures (white
 695 circles, dashed lines). Regression lines and their statistics (slope $\pm$ standard error, Pearson
 696 correlation coefficient [r], and *P* value) are shown.

697 **Fig. 2.** Distribution of MAST-4 and MAST-7 cells on entire membrane filters. (A) Number of
 698 cells detected in the about 3000 FOVs processed on whole filters. Inset images show the
 699 estimated cell abundances in 10 virtually cut filter sections. (B) MAST-4 and MAST-7 cell
 700 abundance and their coefficients of variation (in italics) obtained by virtually splitting the
 701 whole filter in a different number of sections.

702 **Fig. 3.** Global abundance of total phototrophic and heterotrophic protists ($< 5 \mu\text{m}$) and of three
 703 MAST groups. Box plots capture the variation in abundance of the phototrophs and
 704 heterotrophs (A) and of the MAST groups (C) at surface (orange) and DCM (grey). The
 705 variation of the proportion of phototrophs and heterotrophs to the total counts of small protists
 706 (B) and of the proportion of the MAST groups to total counts of heterotrophic protists (D) are
 707 shown for both depths. Each data point represents an individual sample. Significant
 708 differences between surface and DCM are shown with Mann–Whitney–Wilcoxon tests (NS: *P*
 709 > 0.05 , *: *P* < 0.01 , **: *P* < 0.001).

710 **Fig. 4.** Biogeographical distribution of heterotrophic protists and MAST groups in the
 711 epipelagic zone. The abundances of heterotrophs at surface (A) and DCM depths (B) are

represented by the circle size, while the summed contributions of the MASTs groups to heterotrophs are shown by the blue-scale. The longitudinal distribution of MAST-1C, MAST-4 and MAST-7 cells are displayed for surface (**C**) and DCM (**D**) samples. Two DCM stations in the Indian Ocean with MAST-4 abundances out of the scale are marked by arrows (*, station 45, 316 cells ml⁻¹; ‡, station 47, 502 cells ml⁻¹). Loess regressions were computed to illustrate the relative constancy of the three taxa in both depths. Pale colour area displays a 0.95 confidence interval around the trend.

Fig. 5. Density plots of the logarithmic abundance distribution of small phototrophic and heterotrophic protists (**A**) and of the three MAST groups (**B**). Mean (μ) and standard deviation (σ) of each abundance spectra are shown to emphasise the width of the distribution. The curve of the best fitting model is also displayed for each group (gamma in small protists; Weibull in MAST-1C and MAST-7; normal in MAST-4).

Table 1. Mean abundance ($\pm$ standard error; cells ml^{-1}) of total small eukaryotes (cells $\leq 5 \mu\text{m}$) and the three MAST groups across the different biogeographic areas sampled.

A. Small eukaryotes ($< 5\mu\text{m}$)											
Ocean		Atlantic				Indian			Pacific		
Biogeographic subdivisions		Whole ocean	Leg			Whole ocean	Leg		Whole ocean	Leg	
			North Atlantic	Equat. Atlantic	South Atlantic		Indian	Great Austral. Bight		Equat. Pacific	North Pacific
Number of stations	<i>Surf.</i>	45	14	20	11	18	11	7	31	15	16
	<i>DCM</i>	51	17	20	14	30	20	10	33	15	18
Phototrophs	<i>Surf.</i>	1049 (± 81)	967 (± 95)	1394 (± 119)	527 (± 62)	1459 (± 286)	684 (± 100)	2677 (± 405)	2351 (± 364)	2283 (± 434)	2416 (± 591)
	<i>DCM</i>	3519 (± 375)	2823 (± 427)	4209 (± 678)	3381 (± 797)	3687 (± 292)	3699 (± 351)	3664 (± 549)	6498 (± 1409)	3098 (± 302)	9332 (± 2398)
Heterotrophs	<i>Surf.</i>	367 (± 31)	469 (± 58)	392 (± 43)	192 (± 26)	337 (± 41)	236 (± 24)	497 (± 63)	440 (± 60)	331 (± 34)	542 (± 108)
	<i>DCM</i>	755 (± 84)	594 (± 62)	887 (± 156)	763 (± 197)	607 (± 66)	575 (± 83)	671 (± 110)	1050 (± 174)	764 (± 129)	1290 (± 292)
Total	<i>Surf.</i>	1416 (± 97)	1436 (± 124)	1786 (± 134)	719 (± 86)	1796 (± 320)	920 (± 116)	3174 (± 440)	2791 (± 408)	2614 (± 452)	2958 (± 681)
	<i>DCM</i>	4275 (± 447)	3417 (± 474)	5096 (± 803)	4143 (± 987)	4294 (± 338)	4274 (± 409)	4335 (± 632)	7549 (± 1525)	3861 (± 344)	10621 (± 2593)
B. MAST groups											
Ocean		Atlantic				Indian			Pacific		
Biogeographic subdivisions		Whole ocean	Leg			Whole ocean	Leg		Whole ocean	Leg	
			North Atlantic	Equat. Atlantic	South Atlantic		Indian	Great Austral. Bight		Equat. Pacific	North Pacific
Number of stations	<i>Surf.</i>	56	18	23	15	33	23	10	39	18	21
	<i>DCM</i>	53	18	20	15	32	22	10	37	16	20
MAST-1C	<i>Surf.</i>	8 (± 1)	12 (± 3)	5 (± 1)	9 (± 2)	7 (± 1)	5 (± 1)	12 (± 2)	7 (± 1)	6 (± 1)	7 (± 1)
	<i>DCM</i>	5 (± 1)	6 (± 1)	4 (± 1)	6 (± 1)	7 (± 1)	7 (± 2)	8 (± 2)	7 (± 1)	6 (± 1)	8 (± 2)
MAST-4	<i>Surf.</i>	48 (± 6)	60 (± 9)	48 (± 10)	33 (± 9)	49 (± 11)	35 (± 12)	83 (± 20)	56 (± 9)	64 (± 11)	49 (± 14)
	<i>DCM</i>	43 (± 5)	46 (± 11)	43 (± 6)	40 (± 8)	67 (± 18)	75 (± 23)	47 (± 14)	34 (± 6)	42 (± 10)	28 (± 7)
MAST-7	<i>Surf.</i>	15 (± 3)	16 (± 4)	19 (± 5)	4 (± 1)	16 (± 6)	11 (± 7)	27 (± 8)	44 (± 7)	39 (± 7)	48 (± 11)
	<i>DCM</i>	17 (± 2)	28 (± 5)	13 (± 3)	8 (± 3)	22 (± 5)	27 (± 7)	11 (± 3)	31 (± 6)	21 (± 4)	38 (± 9)

N/A: not applicable, N/D: not determined.
Abbreviations: *Surf.*: Surface, *DCM*: Deep Chlorophyll Maximum depth, *Equat. Atlantic*: Equatorial Atlantic, *Great Austral. Bight*: Great Australian Bight and *Equat. Pacific*: Equatorial Pacific.
In bold and underlined, Mean abundance of picoeukaryotes or MASTs significantly different (Welch's t-test, $P < 0.05$) in the different Legs and/or Longhurst Biomes within their respective ocean.

Fig. 2. Distribution of MAST-4 and MAST-7 cells on entire membrane filters. **(A)** Number of cells detected in the about 3000 FOVs processed on whole filters. Inset images show the estimated cell abundances in 10 virtually cut filter sections. **(B)** MAST-4 and MAST-7 cell abundance and their coefficients of variation (in italics) obtained by virtually splitting the whole filter in a different number of sections.

158x159mm (300 x 300 DPI)

Fig. 3. Global abundance of total phototrophic and heterotrophic protists ($<5 \mu\text{m}$) and of three MAST groups. Box plots capture the variation in abundance of the phototrophs and heterotrophs (**A**) and of the MAST groups (**C**) at surface (orange) and DCM (grey). The variation of the proportion of phototrophs and heterotrophs to the total counts of small protists (**B**) and of the proportion of the MAST groups to total counts of heterotrophic protists (**D**) are shown for both depths. Each data point represents an individual sample. Significant differences between surface and DCM are shown with Mann-Whitney-Wilcoxon tests (NS: $P > 0.05$, *: $P < 0.01$, **: $P < 0.001$).

156x157mm (300 x 300 DPI)

Fig. 4. Biogeographical distribution of heterotrophic protists and MAST groups in the epipelagic zone. The abundances of heterotrophs at surface (**A**) and DCM depths (**B**) are represented by the circle size, while the summed contributions of the MASTs groups to heterotrophs are shown by the blue-scale. The longitudinal distribution of MAST-1C, MAST-4 and MAST-7 cells are displayed for surface (**C**) and DCM (**D**) samples. Two DCM stations in the Indian Ocean with MAST-4 abundances out of the scale are marked by arrows (*, station 45, 316 cells ml⁻¹; †, station 47, 502 cells ml⁻¹). Loess regressions were computed to illustrate the relative constancy of the three taxa in both depths. Pale colour area displays a 0.95 confidence interval around the trend.

153x122mm (300 x 300 DPI)

Fig. 5. Density plots of the logarithmic abundance distribution of small phototrophic and heterotrophic protists (**A**) and of the three MAST groups (**B**). Mean (μ) and standard deviation (σ) of each abundance spectra are shown to emphasise the width of the distribution. The curve of the best fitting model is also displayed for each group (gamma in small protists; Weibull in MAST-1C and MAST-7; normal in MAST-4).

90x45mm (300 x 300 DPI)