Shape matters: the relationship between cell geometry and diversity in phytoplankton.

Alexey Ryabov[1,2*], Onur Kerimoglu[1,3], Elena Litchman[4], Irina Olenina[5,6], Leonilde Roselli[7], Alberto Basset[8,9], Elena Stanca[8], and Bernd Blasius[1,2]

Affiliations:

[1]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany

[2]Helmholtz Institute for Functional Marine Biodiversity (HIFMB), Carl von Ossietzky University Oldenburg, Oldenburg, Germany

[3]Insistute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

[4]W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI 49060, USA

[5]Environmental Protection Agency, Klaipėda, Lithuania

[6]Marine Research Institute of the Klaipeda University, Klaipėda, Lithuania

[7]Agency for the Environmental Prevention and Protection (ARPA Puglia), Lecce, Italy

[8]Department of Biological and Environmental Science and Technologies, University of Salento, Lecce, Italy

[9]Institute of Research on Terrestrial Ecosystems, National Research Council, Lecce Italy

*Correspondence to: alexey.ryabov@uni-oldenburg.de.

Summary

Organisms' size and shape profoundly influence their ecophysiological performance and evolutionary fitness, suggesting a link between morphology and diversity. We analyse global datasets of unicellular phytoplankton, major group of photosynthetic microbes with an astounding diversity of cell sizes and shapes, and explore the distribution of taxonomic diversity across different cell shapes and sizes. We find that cells of intermediate volume have the greatest shape variation, from oblate to extremely elongated forms, while small and large cells are mostly compact (e.g., spherical or cubic). Taxonomic diversity varies across cell elongation and cell volume, with both traits explaining up to 92% of its variance. It decays exponentially with cell elongation and displays a log-normal dependence on cell volume, peaking for compact, intermediate-volume cells. Our findings point to the presence of different selective pressures and constraints on the geometry of phytoplankton cells and, thus, improve our understanding of the evolutionary rules of life.

Phytoplankton are major aquatic primary producers that form the base of most marine food webs and are vital to the functioning of marine ecosystems. Marine unicellular phytoplankton exhibit an enormous diversity (Hutchinson, 1961), with cell volumes spanning many orders of magnitude and dozens of different shape types, from simple spherical to extremely complex cells (Reynolds, 2006). This huge variation in phytoplankton cell volumes and shapes presents a unique opportunity for investigating evolutionary constraints on morphological traits and their connection to taxonomic richness, because the geometry of a phytoplankton cell plays an important role in its adaptation to the environment. Cell size and shape affect most aspects of phytoplankton survival, from grazing by zooplankton (Pančić and Kiørboe, 2018; Sunda and Hardison, 2010) to sinking (Durante et al., 2019) and diffusion (Padisák et al., 2003), diffusive transport limitation (Kiørboe, 2008) and nutrient uptake (Edwards et al., 2012; Grover, 1989; Karp-Boss and Boss, 2016; Tambi et al., 2009). While the role of cell size in determining phytoplankton fitness and diversity has been documented previously (Cermeño and Figueiras, 2008; Ignatiades, 2017), not much is known about the role of cell shapes.

Here, we characterize broad patterns in cell shapes and their relationship with cell volume and taxonomic richness across main phyla of unicellular marine phytoplankton and heterotrophic dinoflagellates (together called below, for brevity, phytoplankton). We compiled one of the most comprehensive data sets of phytoplankton in terms of sizes, shapes and taxonomic diversity from seven globally distributed marine areas: North Atlantic (Scotland), Mediterranean Sea (Greece and Turkey), Indo-Pacific (the Maldives), South-western Atlantic (Australia), Southern Atlantic (Brazil) and Baltic Sea (see Methods). The data comprises 5,743 cells of unicellular phytoplankton from 402 genera belonging to 16 phyla. We classified each cell as one of 38 fundamental geometric shapes, such as spheres, cylinders, prisms, etc., measured cell linear dimensions and calculated the surface area and volume for each cell (Hillebrand et al., 1999; Olenina et al., 2006; Vadrucci et al., 2007) (see Methods). Cell volumes span almost 10 orders of magnitude, from 0.065 $\mu\mathrm{m}^3$ for the cyanobacterium *Merismopedia* to $5 \cdot 10^8\ \mu\mathrm{m}^3$ for *Dinophyceae's Noctiluca.*

The degree of shape elongation can be expressed as the aspect ratio and surface relative extension (see Methods). The aspect ratio, r, characterizes the linear dimension of cell elongation, and is less than one for oblate (flattened at the poles) shapes, and is greater than 1 for prolate (stretched) shapes. We also define a shape as compact if $2/3 < r < 3/2$. The surface extension, ϵ, shows the relative gain in surface area of a cell compared to a sphere with the same volume. The minimum level of surface extension is shape-specific and equals 1 for spheres, 1.14 for cylinders, 1.24 for

cubes, and 1.09 for double cones (see Methods). The two measures of shape elongation are related, and the logarithm of the aspect ratio changes approximately with the square root of surface extension (Extended Data Fig. 2).

Variation in cell shape

We found that the taxonomic diversity across different phyla varies with cell shape type and elongation (Fig. 1A). Most Bacillariophyta (diatoms) are cylindrical or prismatic, while other phyla are mostly ellipsoidal, with additional shapes, e.g., conic or of a more complex geometry, being relatively rare. In our database, 46% of genera are prolate, 38% compact and only 16% oblate (Fig. 1B). These proportions vary across phyla and shapes (Extended Data Fig. 1). For instance, more than half of genera classified as elliptical cells have a compact shape, while for other shapes more than half of genera have prolate cells. Oblate shapes comprise up to 20% of genera in diatoms, dinoflagellates (Miozoa), Haptophyta, Charophyta, Cryptophyta, and Euglenozoa, but are rarer (< 10%) in other phyla. Half-shapes such as half-spheres or half-cones are more dominated by oblate forms.

Shape elongation is hypothesized to influence phytoplankton fitness. Several studies argued that elongation is beneficial for the volume-specific nutrient uptake and, therefore, large cells should be elongated to increase the surface to volume ratio (Lewis, 1976; Niklas, 2000). However, our analysis, based on the order of magnitude more cell measurements than in previous studies (Lewis, 1976; Niklas, 2000), shows that cell surface area increases with volume approximately to the power of 2/3 (Fig. 2A), indicating that cell dimensions scale on average isometrically with volume, and there is no evidence for more shape elongation with increasing volume.

By contrast, the variation in cell elongation strongly depends on cell volume (Fig. 2C, Extended Data Fig. 3). The distribution of the surface extension as a function of cell volume is approximately hump-shaped, with a peak of cell elongation at intermediate volumes (between $10^3 - 10^4\ \mu m^3$), where the cell surface area can exceed the surface area of a sphere with an equivalent volume up to 5-fold. In contrast, for cells of very small or large volume, surface extension approaches its minimum values, implying that these cells have a compact shape minimizing their surface area. The hump-shaped pattern is also seen in the 75% and 90% quantiles (Fig. 2C), confirming that this is not a sample artifact. The same pattern emerges for the aspect ratio, which reaches 100 for prolate cells and drops to 0.025 for oblate cells (Extended Data Fig. 3). This pattern also holds across different trophic guilds (autotrophic, mixotrophic or heterotrophic); however, the maximum cell elongation is reached only by the autotrophs, while in heterotrophs and mixotrophs the maximum aspect ratio equals 10 and the maximum surface extension equals 2 (Extended Data Fig. 4), likely because these two groups need to swim actively.

Phytoplankton diversity distribution

Taxonomic diversity, D, measured here as richness of genera, depends on both cell volume and surface extension. It follows a lognormal function of volume with a peak of diversity at $V_0 = 1100 \pm 90\ \mu m^3$ (Fig. 2D, $R^2_{adj} = 0.98$) and decreases exponentially with shape surface extension ϵ as $D \sim e^{-1.43\epsilon}$ (Fig. 2E, $R^2_{adj} = 0.97$). Both relationships vary across cell shapes (Extended Data Fig. 5, 6). The ellipsoidal cells have the diversity distribution peaking at the smallest volume, compared to other shapes ($V_0 = 330 \pm 40\ \mu m^3$, $R^2_{adj} = 0.96$) and the fastest rate of diversity decrease with surface extension $D \sim e^{-2.4\epsilon}$, $R^2_{adj} = 0.8$), with 54% of the genera exceeding the surface area of a sphere by less than 10%. By contrast, for cylindrical cells (mainly diatoms), diversity peaks at the largest volume compared to other shapes ($V_0 = 8{,}700 \pm 800\ \mu m^3$, $R^2_{adj} = 0.98$) and declines more slowly with surface extension ($D \sim e^{-1.4\epsilon}$, $R^2_{adj} = 0.92$). There is a comparable effect of surface

extension on diversity for conic shapes ($D \sim e^{-1.2\epsilon}$, $R^2_{adj} = 0.77$). The effect is weaker for prismatic ($D \sim e^{-0.95\epsilon}$, $R^2_{adj} = 0.71$) and complex shapes ($D \sim e^{-0.75\epsilon}$, $R^2_{adj} = 0.62$) which can be attributed to the fact that both prismatic and complex shapes occur mainly in diatoms. The secondary peaks of diversity at ϵ between 1.5 and 3 for these shapes suggest that for specific shapes cell elongation might have a nonmonotonic effect on cell fitness, such that both compact and elongated cells can have high diversity (Grover, 1989). The weaker correlation of diversity with cell elongation for complex shapes could also be caused by the fact that representing complex shapes requires more parameters than just simple composites such as aspect ratio or surface extension.

Both cell volume and surface extension are important drivers of taxonomic diversity. Assuming that volume and surface extension are independent, we can approximate the diversity distribution as product of a lognormal function of volume and a decreasing exponential function of surface extension

$$D \sim \exp\left[-\frac{(\log V - \log V_0)^2}{2\sigma^2} - k\epsilon\right]$$

As shown in Fig. 3, this function describes the dependence of diversity on both cell volume and extension remarkably well, with $V_0 = 1{,}000 \pm 200\ \mu m^3$ (mean volume), $\sigma = 1.74 \pm 0.08$ (variance logarithm of volume) and $k = 1.47 \pm 0.06$ (the rate of exponential diversity decrease with surface extension) explaining 92% of the variation of phytoplankton diversity for the entire dataset. Across shape types the fit parameters have the same variance as above: the best match is obtained for ellipsoidal, cylindrical and conic shapes (Fig. 3B-D), and a poorer fit for prismatic and other shape types (Fig. 3E-F). A comparison of the predicted and the observed diversity shows that we get an unbiased fit across all shapes, and also in the group of ellipsoidal, cylindrical and conic shapes (Extended Data Fig. 7A-D). However, it overestimates taxonomic diversity of prismatic and other shapes for the ranges of volume and surface extension where the observed diversity is low (Extended Data Fig. 7E-F). Note that correlations in Fig. 3 obtained across all shapes ($R^2 = 0.92$) are higher that those obtained for some specific shapes (except cylindrical). The reason for this is a niche separation between shape classes in a gradient of surface extension. This niche separation reduces the quality of fit for specific shape types but does not play a role when we consider all shapes together (see Extended Data Fig. 6 for detail).

Similarly, diversity can be correlated to cell volume and aspect ratio (Methods). However, the aspect ratio has a more complicated functional relationship with taxonomic diversity, which is likely due to the non-linear relationship between these two parameters (Extended Data Fig. 2). Although, on average, the diversity predictions obtained using aspect ratio are poorer than those based on surface extension, aspect ratio is easier to measure with automated plankton monitoring(Pomati et al., 2011).

Discussion

Our study shows that cell surface area increases approximately isometrically with cell volume, but the variation in cell elongation exhibits a hump-shaped dependence on cell volume. Interestingly, the shapes of cells of intermediate volume are very diverse and range from oblate and to extremely prolate forms, while cells of both large and small volumes are compact (mostly spherical). To what extent can this pattern be explained by the constraints on cell dimensions? Linear cell dimensions range from $0.5\ \mu\mathrm{m}$ to $1{,}000\ \mu\mathrm{m}$ (Fig. 2B). The minimum cell size is likely constrained by the size of organelles; for instance, for autotrophs the minimum chloroplast size equals 1 μm(Li et al., 2013; Raven, 1998). The maximal feasible cell size can be limited by the scale of diffusive displacement of

proteins in cytoplasm during the cell cycle (see Methods). Thus, the minimal (or maximal) cell volume can only be realized in a compact geometry where all three linear dimensions are approximately equal. A model based on these constraints correctly predicts that the smallest and largest cells should be compact, while cells of intermediate volumes can have a diverse geometry (Extended Data Fig. 3). The model, however, overestimates the measured range in surface extension, yielding values of $\epsilon > 10$ for prolate cells and $\epsilon > 30$ for oblate cells. This discrepancy indicates the existence of further physiological constraints on cell geometry. In an improved model we assume that cell aspect ratios can vary from 0.025 to 100 only (Fig. 2B). As the longest linear cell dimension $L_{max} < 1000\ \mu m$, the allowed range of r reduces with increasing the shortest cell dimension L_{min}, so that r approaches 1 when L_{min} approaches $1{,}000\ \mu\mathrm{m}$ (Fig. 2B, solid line). This constraint may reflect limitations due to mechanical instability, material transport needs within a cell, or reduced predator defence experienced by extremely prolate or oblate cells. With this constraint, the model and the data agree well for prolate cells, but the theoretical model still overestimates the potential surface extension for oblate cells of large volumes (Extended Data Fig. 3). This suggests that there may be unknown additional constraints that prevent the evolution of extremely wide oblate cells with large volume.

Our study shows that cell shape elongation, along with cell size, is an important driver of taxonomic diversity distribution with both traits explaining up to 92% of its variance. Diversity distribution is a lognormal function of volume and decreases exponentially with cell surface extension. As diversity typically increases with abundance (Siemann et al., 1996), we hypothesize that species with compact cells of intermediate volume have the highest fitness among unicellular plankton. Thus, a reduction of cell surface area is likely advantageous as it leads to greater diversification rates resulting in higher diversity of compact cells compared to elongated cells.

For all phyla, except for prismatic and complex shapes (mainly diatoms), the minimization of cell surface area is a beneficial strategy independent of cell volume. Reducing cell surface area likely reduces the cost of cell wall, which may be expensive, and makes a cell less vulnerable to predators. In contrast, having a non-spherical shape is easy only for species with a rigid cell wall, such as diatoms (Martin-Jézéquel et al., 2000; Monteiro et al., 2016). This can explain why for prismatic and complex shapes we observe secondary peaks of richness for elongated shapes, resulting in significant diversity of diatom shapes across a wide range of cell elongation. This suggests that the appearance of silica cell walls in diatoms is a major evolutionary innovation that allows diatoms to achieve an unusually large shape diversity, which may have contributed to the ecological success of this group (Malviya et al., 2016; Nelson et al., 1995).

The surprisingly good prediction of global taxonomic richness of marine plankton by cell volume and surface relative extension implies either a fundamental metabolic relationship between these parameters and speciation rates or a specific global distribution of niches favouring oblate and prolate shapes in competition with compact shapes, as the environment can select certain cell morphology (Charalampous et al., 2018; Kruk and Segura, 2012). In particular, very elongated shapes occur in deep waters (Reynolds, 1988). Our study suggests that this phenomenon can have another explanation, as elongated shapes might dominate at depths because building complex cell wall is cheaper under high nutrient conditions characteristic of deeper layers, compared to low nutrients of the upper layer.

A link between phytoplankton diversity and morphology has not been explored much and previous studies on the topic did not show a consistent pattern. In particular, local species richness showed either a hump-shaped function or was independent of cell volume (Cermeño and Figueiras, 2008), or decreased as a power function of volume (Ignatiades, 2017). There may be several explanations for

the discrepancy between our and previous results. First, unlike previous studies, we consider cell surface extension as an important driver and separate its effects from the effects of cell volume. Second, our study includes a wider range of cell volumes and, third, it includes samples from world's ocean ecosystems of various typology and in different times of the year, so this global pattern may be different from the local patterns influenced by specific environmental conditions, such as nutrient or light levels, grazing, species sorting or mass effects.

Our findings show that taxonomic richness correlates not only with cell size but also with cell shape and open new avenues of biodiversity research. As different environmental factors affect both cell shape and size, they can change shape-size distributions of phytoplankton communities, and therefore, may indirectly affect biodiversity. In particular, temperature and nutrients often change cell volume and, thus, may alter diversity (Acevedo-Trejos et al., 2013; Agawin et al., 2000), which would be important to investigate in the context of rapid environmental change. Similarly, indirect changes in diversity and community composition can be caused by grazing, through its differential effect on cells of various shapes and sizes or by environmental factors through a potential link between cell elongation and generalist or specialist strategies. Finally, many phytoplankton genera are present in the natural environment as colonies or chains, thus, the colony shape and length and the geometry of chains formation might also become important evolutionary factors leading to species dominance or high speciation rates. Answering these questions would help us further understand the ecological and evolutionary constraints on phytoplankton diversity in the ocean.

Methods

Databases

We combined two databases on biovolumes and size-classes of marine unicellular phytoplankton (see Data Availability statement).

Baltic Sea

The first database includes information on phytoplankton species and heterotrophic dinoflagellates, covering a total of 308 genera found in the different parts of the Baltic Sea since the 80s of the 20th century to 2018 (PEG_BVOL, http://www.ices.dk/marine-data/Documents/ENV/PEG_BVOL.zip). The measurements were prepared by the HELCOM Phytoplankton Expert Group (PEG) and originally in more detail described by Olenina et al.(Olenina et al., 2006). The phytoplankton samples were taken in accordance with the guidelines of HELCOM (1988) as integrated samples from surface 0-10, or 0-20 m water layer using either a rosette sampler (pooling equal water volumes from discrete 1; 2,5; 5; 7,5 and 10 m depth) or with a sampling hose. The samples were preserved with acid Lugol's solution(Willén, 1962). For the phytoplankton species identification and determination of their abundance and biomass, the inverted microscope technique(Utermöhl, 1958) was used. After concentration in a sedimentation 10-, 25-, or 50-ml chamber, phytoplankton cells were measured for the further determination of species-specific shape and linear dimensions. All measurements were performed under high microscope magnification (400–945 times) using an ocular scale.

Different ecoregions around the globe

The second database includes a biogeographical snapshot survey of natural phytoplankton and heterotrophic dinoflagellates communities obtained by Ecology Unit of Salento University (https://www.lifewatch.eu/web/guest/catalogue-of-data)(Roselli et al., 2017). The data cover a total of 193 genera and were sampled in five different coastal ecoregions: North Atlantic Sea (Scotland), Mediterranean Sea (Greece and Turkey), Indo-Pacific Ocean (the Maldives), South-

Western Atlantic Ocean (Australia) and Southern Atlantic Ocean (Brazil). The data covers 23 ecosystems belonging to different typology (coastal lagoons, estuaries, coral reefs, mangroves and inlets or silled basins) that were sampled during the summer period in the years 2011 – 2012. Three to nine ecosystems per ecoregion and three locations for each system, yielding a total of 116 local sites replicated three times, were sampled. Phytoplankton were collected with a 6 µm mesh plankton net equipped with a flow meter for determining filtered volume. Water samples for phytoplankton quantitative analysis were preserved with Lugol (15mL/L of sample). Phytoplankton were examined following Utermöhl's method (Utermöhl, 1958). Phytoplankton were analysed by inverted microscope (Nikon T300E, Nikon Eclipse Ti) connected to a video-interactive image analysis system (L.U.C.I.A Version 4.8, Laboratory Imaging). Taxonomic identification, counting and linear dimensions measurements were performed at individual level on 400 phytoplankton cells for each sample. Overall, an amount of 142 800 cells constitutes the present data set. The data on the dimensions of the same species were averaged for each replica. For the present analysis, to reduce the effect of intraspecific variability, the data were averaged again for each genus and local site. Phytoplankton were identified to species or genus level, each cell was associated with a species-specific geometric model and their relative linear dimensions were measured. Detailed information about sampling design, sampled environments and taxonomic list of phytoplankton can be found on the website of the project (http://phytobioimaging.unisalento.it/) (Roselli et al., 2017).

Combined data set

Combining both data set, we obtained a data base that contains information on phytoplankton cell shape type and linear dimensions of a total of 402 genera of unicellular marine phytoplankton (phytoplankton and heterotrophic dinoflagellates) from 7 locations: Baltic Sea, North Atlantic Ocean (Scotland), Mediterranean Sea (Greece and Turkey), Indo-Pacific Ocean (the Maldives), South-Western Atlantic Ocean (Australia), South Atlantic Ocean (Brazil). Phyla were identified according to www.algaebase.org (Guiry and Guiry, 2018).

The datasets were obtained in different regions and by different research groups. "Baltic sea" dataset compared to "Different ecoregions" dataset was obtained during a longer period of time and different techniques. The regular screening of plankton in Baltic sea is performed over the past 25 years and include cells from 1 μm length, while the second dataset includes single screenings in various regions around the globe with mesh grid of 6 μm. The first dataset includes a wider range of cell volumes from 0.065 $\mu\mathrm{m}^3$ to $5 \cdot 10^8\ \mu\mathrm{m}^3$ and more species, and the second dataset represents only a part of the entire distribution in the range of volumes from 5.9 $\mu\mathrm{m}^3$ to $3.9 \cdot 10^6\ \mu\mathrm{m}^3$. Despite these differences in the techniques and origin of the data, we find similar distributions of diversity for both datasets in the range of volumes where the datasets overlap, and these distributions are also close to the distributions presented here for the combined dataset.

Cell volume and surface area

We calculated cell volume and surface area based on formulae published earlier(Hillebrand et al., 1999; Sun and Liu, 2003; Vadrucci et al., 2007) and http://phytobioimaging.unisalento.it/AtlasofShapes. To standardize the calculations for both databases and automate the process, we have rederived all formulae using Maple software and corrected some formulae, yielding a list of analytic expressions for cell volume and cell surface area for each of the 38 shape types (see Supplementary material for the entire list of rederived formulae and a Maple script, which can be used as a tool for further derivations).

Cell dimensions

To characterize cell linear dimensions in 3D space additionally to cell microscopic characteristics, which can include up to 10 measurements of different cell parts, we use 3 orthogonal dimensions of each cell, charactering the minimal, middle and maximal cell linear dimensions, which are denoted as L_{min}, L_{mid} and L_{max}. For most of shapes such as sphere, ellipsoid, cube or cone the meaning of these dimensions is clear. For some asymmetrical cells with, for instance, different horizontal extents at the top and bottom, we used the largest of these two extends, because the smallest one (or average) does not properly describe the geometric limitations. For instance, a truncated cone is characterized by the height and the radius at the top and bottom. However, the top radius is typically extremely small, and is not related to the geometric constrains. Thus, for such shapes we used height as one dimension and the doubled bottom radius as the other two dimensions. For more complex shapes, consisting of few parts measured separately (e.g., half ellipsoid with a cone), we used the sum of linear dimensions of these parts as projected to each orthogonal axis (see Supplementary material for the details for each shape type).

Measures of cell elongation

To characterize cell elongation, we used aspect ratio and relative surface extension (calculated as the inverse shape sphericity). For cells with axial symmetry the *aspect ratio* is defined as the ratio between the principal axis of revolution and the maximal diameter perpendicular to this axis. It indicates the linear cell elongation and is greater than one for prolate shapes, equal to one for shapes with equal linear dimensions (cubes, spheres, cones with equal height and bottom diameter, etc.), and less than one for oblate shapes. To generalize the definition of aspect ratio for cells without axial symmetry, we classify a cell as prolate, if $L_{mid} < \sqrt{L_{max}L_{min}}$, so L_{mid} is closer to the minimal dimension in terms of geometric averaging, and as oblate, if $L_{mid} > \sqrt{L_{max}L_{min}}$. For prolate cells the aspect ratio equals L_{max}/L_{min}, for oblate cells we use the inverse value. Note that due to intraspecific and intragenus variability cells of the same genera can be attributed to various elongation types.

The *relative surface area extension,* ϵ, shows the gain in surface area due to the deviation from a spherical shape and is calculated as the ratio of the surface area S of a cell with a given morphology to the surface area of a sphere with the same volume, $\epsilon = \sqrt[3]{36\pi}\, S/V^{2/3}$. Mathematically it can also be termed the inverse shape sphericity.

Prolate, oblate and compact cells

Prolate or oblate cells can have an extremely large values of cell surface extension, but the minimal value of cell surface extension, ϵ_{min}, is shape specific. To find ϵ_{min} for a given shape type (e.g. ellipses or cylinders), we need to find a specific shape with minimal surface area for given volume. Assume that $L_{max} = \alpha L_{min}$ and $L_{mid} = \beta L_{min}$ where α and β are some positive numbers. Then for basic geometric shapes, the surface area can be expressed as $S = s(\alpha,\beta)L_{min}^2$ and volume as $V = v(\alpha,\beta)L_{min}^3$, where $s(\alpha,\beta)$ and $v(\alpha,\beta)$ are shape specific functions which do not depend on L_{min}. Then surface extension becomes a function of only α and β: $\epsilon(\alpha,\beta) = \sqrt[3]{36\pi}\, s(\alpha,\beta)/v(\alpha,\beta)^{2/3}$. The minimal surface extension can be found as $\epsilon_{min} = \min_{\alpha,\beta} \epsilon(\alpha,\beta)$ and the values $(\alpha^*,\beta^*) = \arg\min_{\alpha,\beta} \epsilon(\alpha,\beta)$ are the ratios between the linear dimensions of the specific shape with the minimal surface area. If a shape has rotational symmetry, then $\alpha = \beta$ and the problem becomes even simpler. Solving this minimisation problem for different shape type, we find that for ellipses the minimal surface extension $\epsilon_{min} = 1$ is achieved when all semi-axes are equal, that is, if the ellipse is a sphere. For a cylinder $\epsilon_{min} = (3/2)^{1/3} = 1.14$, when its height equals diameter; for a

parallelogram or prism on a rectangular base $\epsilon_{min} = (6/\pi)^{1/3} \approx 1.24$ (when it is a cube). In all these cases $\alpha^* = \beta^* = 1$.

Strictly speaking, only cells with aspect ratio of 1 are neither prolate nor oblate and can therefore be identified as compact. However, the aspect ratio changes over four orders of magnitude, and cells with a small difference in linear dimensions are closer to the compact shapes than to extremely oblate or prolate cells. To separate these groups, we define a cell to be *compact* if $L_{max}/L_{min} < 3/2$, so that the maximal cell dimensions is less than 150% of the minimal dimension. Such a choice of the border between compact, prolate and oblate cells is due to the specific dependence between the aspect ratio and surface extension (Extended Data Fig. 2). As shown in the Extended Data figure, for cells with small ϵ, the aspect ratio changes much faster than surface extension. As the border of the aspect ratio can be approximated as $\log r = \pm 1.3\sqrt{\epsilon - 1}$, the aspect ratio of 3/2 (or 2/3) can correspond to only a 2% increase in the surface area with respect to a ball.

Using aspect ratio for predicting biodiversity

Like the surface extension, the aspect ratio can be used as predictor of taxonomic diversity. The regression analysis based on volume and aspect ratio gives $R^2_{adj} = 0.89$ across all data and R^2_{adj} ranging from 0.23 to 0.86 for specific shapes (Extended Data Fig. 8). The reduced R^2_{adj} values compared to the fitting based on surface extension probably occur because of a more complicated functional dependence of diversity on aspect ratio (Extended Data Fig. 9). For instance, for ellipsoidal prolate shapes diversity monotonically decreases with aspect ratio but shows a peak for oblate shapes at $r \approx 1/2$. For cylinders the picture is even more complicated with two peaks of diversity at $r \approx 3$ and $1/3$.

The discrepancy between the dependence of diversity on the surface extension and aspect ratio occurs likely from the nonlinear relationships between these parameters (Extended Data Fig. 2). The logarithm of aspect ratio changes approximately as $\sqrt{\epsilon - 1}$, implying an extremely high rate of change of aspect ratio with ϵ for compact shapes, and a much smaller rate for elongated shapes. Consequently, projecting diversity onto the surface extension axis results in an exponential decrease, while projecting it on the aspect ratio axes results in a bimodal distribution with a local minimum of diversity shapes for $r = 1$. However, the projections show only a part of the entire picture. As, shown in the bivariate plot (Extended Data Fig. 2A) the diversity peaks for spherical cells (both surface extension and aspect ratio of around 1) and then decreases with deviation from this shape towards prolate or oblate forms. This decrease is asymmetric and occurs faster for oblate shapes.

Diffusion constraints on the cell's longest linear dimension

The mean diffusive displacement in a 3D space equals $\sqrt{< x^2 >} = \sqrt{6Dt}$, where D is the diffusion coefficient and t time interval. The maximal cell size can be limited by the mean diffusive displacement of molecules in cell cytoplasm during one life cycle. For instance, the diffusion of proteins in cytoplasm of bacteria, *Escherichia coli,* ranges from 0.4 to 7 $\mu m^2/s$ (Ref(Kumar et al., 2010)). Diffusion rates in cytoplasm presented in the *Cell Biology by then Numbers* database (http://book.bionumbers.org/what-are-the-time-scales-for-diffusion-in-cells/) lay also in this range. According with this data, the mean diffusive displacement in the cell cytoplasm during one day (a typical reproduction time scale for phytoplankton) should range from 455 to 1900 μm. These values are close to the maximal cell size of 1000 μm, found in our study.

References

Acevedo-Trejos, E., Brandt, G., Merico, A., and Smith, S.L. (2013). Biogeographical patterns of phytoplankton community size structure in the oceans. Glob. Ecol. Biogeogr. *22*, 1060–1070.

Agawin, N.S.R., Duarte, C.M., and Agustí, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. Limnol. Oceanogr. *45*, 591–600.

Cermeño, P., and Figueiras, F.G. (2008). Species richness and cell-size distribution: size structure of phytoplankton communities. Mar. Ecol. Prog. Ser. *357*, 79–85.

Charalampous, E., Matthiessen, B., and Sommer, U. (2018). Light effects on phytoplankton morphometric traits influence nutrient utilization ability. J. Plankton Res. *40*, 568–579.

Durante, G., Basset, A., Stanca, E., and Roselli, L. (2019). Allometric scaling and morphological variation in sinking rate of phytoplankton. J. Phycol.

Edwards, K.F., Thomas, M.K., Klausmeier, C.A., and Litchman, E. (2012). Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. Limnol Ocean. *57*, 554–566.

Grover, J.P. (1989). Influence of Cell Shape and Size on Algal Competitive Ability. J. Phycol. *25*, 402–405.

Guiry, M.D., and Guiry, G.M. (2018). AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org. AlgaeBase.

Hillebrand, H., Dürselen, C.-D., Kirschtel, D., Pollingher, U., and Zohary, T. (1999). Biovolume calculation for pelagic and benthic microalgae. J. Phycol. *35*, 403–424.

Hutchinson, G.E. (1961). The paradox of the plankton. Am Nat *95*, 137–145.

Ignatiades, L. (2017). Size scaling patterns of species richness and carbon biomass for marine phytoplankton functional groups. Mar. Ecol. *38*, e12454.

Karp-Boss, L., and Boss, E. (2016). The Elongated, the Squat and the Spherical: Selective Pressures for Phytoplankton Shape. In Aquatic Microbial Ecology and Biogeochemistry: A Dual Perspective, (Springer), pp. 25–34.

Kiørboe, T. (2008). A mechanistic approach to plankton ecology (Princeton University Press).

Kruk, C., and Segura, A.M. (2012). The habitat template of phytoplankton morphology-based functional groups. Hydrobiologia *698*, 191–202.

Kumar, M., Mommer, M.S., and Sourjik, V. (2010). Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli. Biophys. J. *98*, 552–559.

Lewis, W.M. (1976). Surface/volume ratio: implications for phytoplankton morphology. Science *192*, 885–887.

Li, Y., Ren, B., Ding, L., Shen, Q., Peng, S., and Guo, S. (2013). Does chloroplast size influence photosynthetic nitrogen use efficiency? PloS One *8*, e62036.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., and Bittner, L. (2016). Insights into global diatom distribution and diversity in the world's ocean. Proc. Natl. Acad. Sci. *113*, E1516–E1525.

Martin-Jézéquel, V., Hildebrand, M., and Brzezinski, M.A. (2000). Silicon metabolism in diatoms: implications for growth. J. Phycol. *36*, 821–840.

Monteiro, F.M., Bach, L.T., Brownlee, C., Bown, P., Rickaby, R.E.M., Poulton, A.J., Tyrrell, T., Beaufort, L., Dutkiewicz, S., Gibbs, S., et al. (2016). Why marine phytoplankton calcify. Sci. Adv. *2*, e1501822.

Nelson, D.M., Tréguer, P., Brzezinski, M.A., Leynaert, A., and Quéguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Glob. Biogeochem. Cycles *9*, 359–372.

Niklas, K.J. (2000). The evolution of plant body plans—a biomechanical perspective. Ann. Bot. *85*, 411–438.

Olenina, I., Hajdu, S., Edler, L., Andersson, A., Wasmund, N., Busch, S., Göbel, J., Gromisz, S., Huseby, S., Huttunen, M., et al. (2006). Biovolumes and size-classes of phytoplankton in the Baltic Sea. HELCOM Balt.Sea Environ. Proc. *106*.

Padisák, J., Soróczki-Pintér, É., and Rezner, Z. (2003). Sinking properties of some phytoplankton shapes and the relation of form resistance to morphological diversity of plankton—an experimental study. In Aquatic Biodiversity, (Springer), pp. 243–257.

Pančić, M., and Kiørboe, T. (2018). Phytoplankton defence mechanisms: traits and trade-offs: Defensive traits and trade-offs. Biol. Rev. *93*, 1269–1303.

Pomati, F., Jokela, J., Simona, M., Veronesi, M., and Ibelings, B.W. (2011). An automated platform for phytoplankton ecology and aquatic ecosystem monitoring. Environ. Sci. Technol. *45*, 9658–9665.

Raven, J.A. (1998). The twelfth Tansley Lecture. Small is beautiful: the picophytoplankton. Funct. Ecol. *12*, 503–513.

Reynolds, C.S. (2006). Ecology of phytoplankton (Cambridge UniversityPress).

Reynolds, C.S. (Colin) (1988). Functional Morphology and the Adaptive Strategies of Freshwater Phytoplankton.

Roselli, L., Litchman, E., Stanca, Elena, Cozzoli, Francesco, and Basset, Alberto (2017). Individual trait variation in phytoplankton communities across multiple spatial scales. J. Plankton Res. *39*, 577–588.

Siemann, E., Tilman, D., and Haarstad, J. (1996). Insect species diversity, abundance and body size relationships. Nature *380*, 704–706.

Sun, J., and Liu, D. (2003). Geometric models for calculating cell biovolume and surface area for phytoplankton. J. Plankton Res. *25*, 1331–1346.

Sunda, W.G., and Hardison, D.R. (2010). Evolutionary tradeoffs among nutrient acquisition, cell size, and grazing defense in marine phytoplankton promote ecosystem stability. Mar. Ecol. Prog. Ser. *401*, 63–76.

Tambi, H., Flaten, G.A.F., Egge, J.K., Bødtker, G., Jacobsen, A., and Thingstad, T.F. (2009). Relationship between phosphate affinities and cell size and shape in various bacteria and phytoplankton. Aquat. Microb. Ecol. *57*, 311–320.

Utermöhl, H. (1958). Zur vervollkommnung der quantitativen phytoplankton-methodik: Mit 1 Tabelle und 15 abbildungen im Text und auf 1 Tafel. Int. Ver. Für Theor. Angew. Limnol. Mitteilungen *9*, 1–38.

Vadrucci, M.R., Cabrini, M., and Basset, A. (2007). Biovolume determination of phytoplankton guilds in transitional water ecosystems of Mediterranean Ecoregion. Transitional Waters Bull. *1*, 83–102.

Willén, T. (1962). Studies on the phytoplankton of some lakes connected with or recently isolated from the Baltic. Oikos *13*, 169–199.

Acknowledgements

We thank H. Hillebrand for useful comments on the manuscript; Michael Guiry for providing data on phylogenetic classification. A.R. acknowledges funding by the Ministry of Science and Culture, State of Lower Saxony, through the projects project POSER. E.L. was in part supported by the sabbatical fellowship from the German Centre of Integrative Biodiversity (iDiv). A.B. acknowledges funding by the Puglia Region, through the strategic project PS126 'PhytoBioImaging'. O.K. was supported by the German Research Foundation, DFG [KE 1970/1-1 and KE 1970/2-1]. B.B. acknowledges funding project BL 772/6-1 within the DFG Priority Programme 1704 DynaTrait.

Author contributions

A.R. designed the research and performed the analysis. A.R. and O.K. calcualted cell surface and volume. A.R. wrote the manuscript with contribution from B.B., E.L., I.O., L.R. and O.K.; I.O. and L.R. described methods. I.O., L.R. A.B., E.S. provided data.

Correspondence and requests for materials should be addressed to A.R.

Competing interests

The authors declare no competing financial interests.

Data availability

Data on Baltic sea are publicly available under http://www.ices.dk/marine-data/Documents/ENV/PEG_BVOL.zip. Data on the global ecosystems are available under https://www.lifewatch.eu/web/guest/catalogue-of-data.

Figure	Model	R^2_{adj}	$b_1 \pm \delta_1$ (p)	$b_2 \pm \delta_2$ (p)	$b_3 \pm \delta_3$ (p)	$b_4 \pm \delta_4$ (p)
Fig. 2A	$\log S = b_1 + b_2 \log V$	0.98	0.767±0.005	0.678±0.001		
Fig. 2B	$r = \pm \frac{2}{\exp\left[\frac{\log L_{min} - b_1}{b_2}\right] + 1}$	0.24	1.79±0.2	0.24±0.2 (0.13)		
Fig. 2D	$D = b_1 \exp\left(-\frac{(\log V - \log b_2)^2}{2(b_3)^2}\right)$	0.98	140±3	1100±90	1.34±0.04	
Fig. 2E	$\ln D = b_1 - b_4 \epsilon$	0.97	6.2±0.1	1.43±0.06		
Fig. 3A	$\ln D = b_1 - \frac{(\log V - \log b_2)^2}{2(b_3)^2} - b_4 \epsilon$	0.92	7.0±0.1	1000±200	1.47±0.06	1.74±0.08
Fig. 3B		0.85	8.7±0.4	380±100 (0.0091)	1.54±0.1	3.6±0.3
Fig. 3C		0.93	5.6±0.1	5900±900	1.38±0.05	1.58±0.08
Fig. 3D		0.79	4.8±0.3	430±100 (0.0048)	1.36±0.1	1.5±0.1
Fig. 3E		0.65	3.7±0.3	1800±400 (4e-05)	1.02±0.08	0.7±0.1
Fig. 3F		0.55	2.2±0.2	800±300 (0.014)	1.59±0.2	0.5±0.1 (6e-05)

Table 1. Fitting parameters for figures in the main text. Parameter values b_i are specified with standard error δ_i and p-value in brackets (only when $p > 10^{-5}$). Fitting in Fig. 2B is done to the outer hull of the data points.

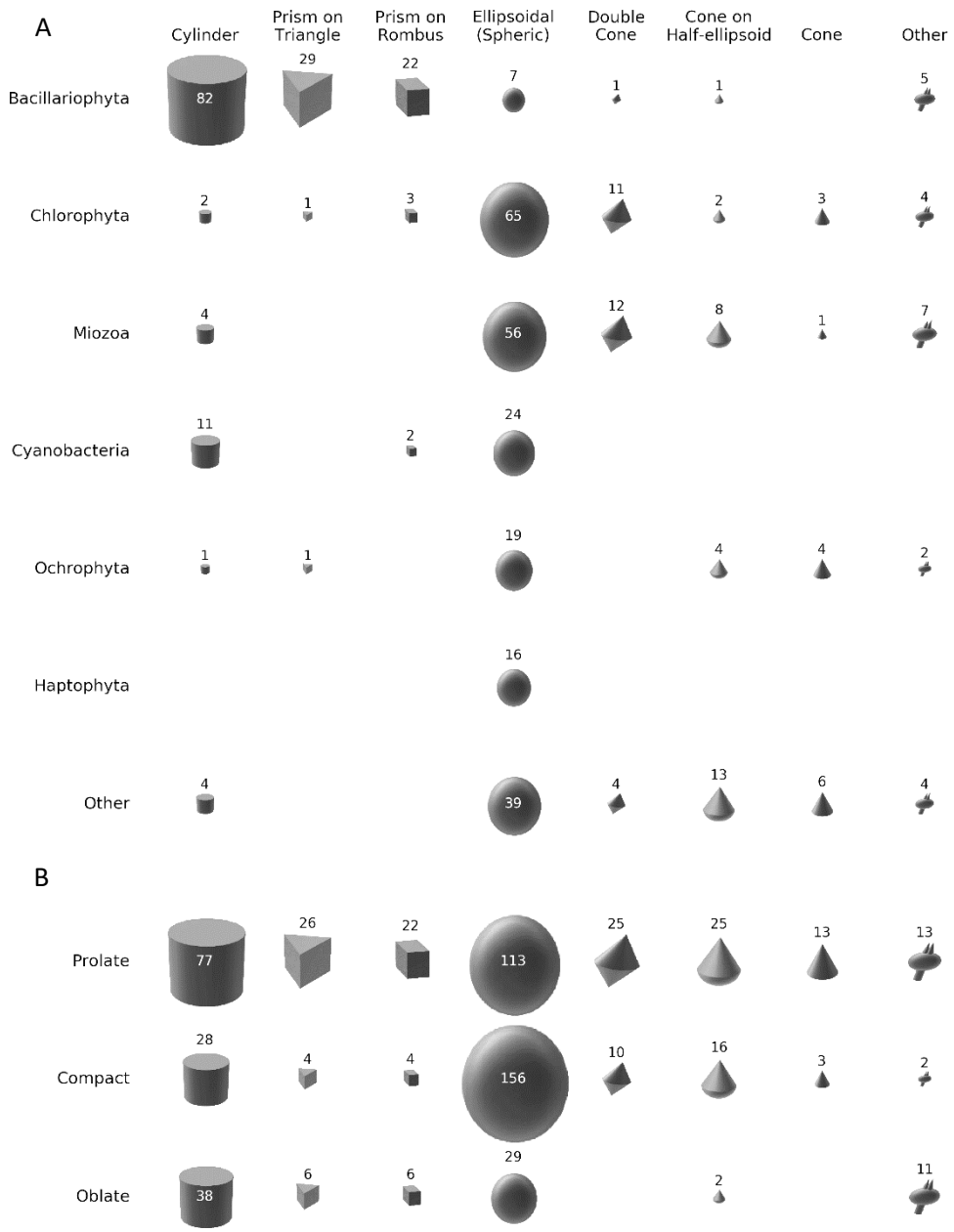


Fig. 1. Diversity distribution of various shape types (columns) across phyla (A, rows) and across cell shape elongation (B, rows). The area of each figure is proportional to the number of genera (shown next to or within it). See Extended Data Fig. 1 for detailed analysis.

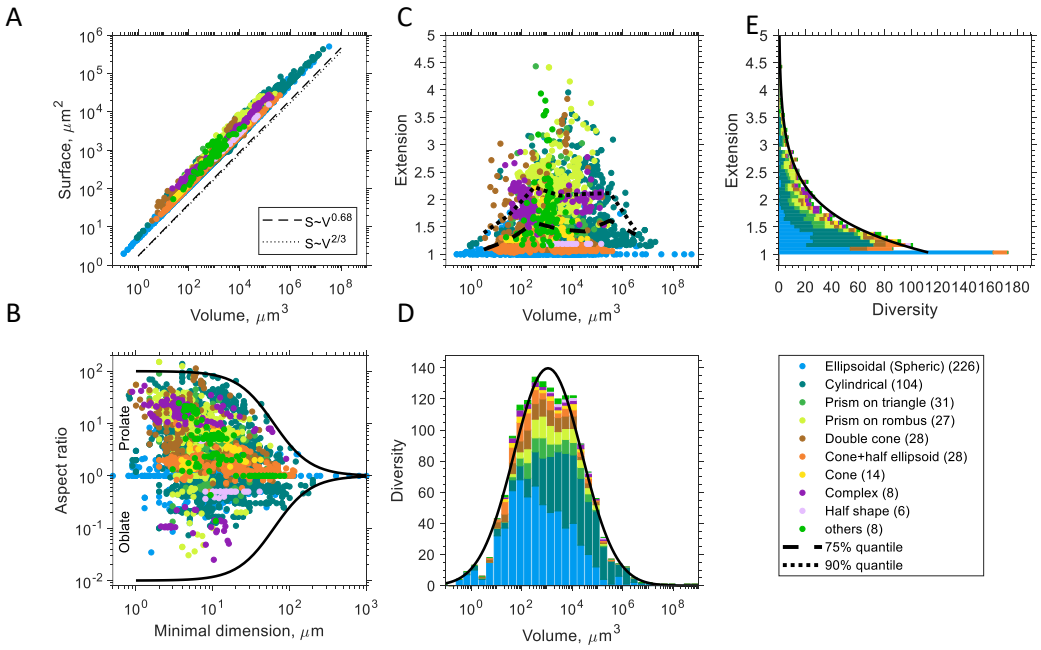


Fig. 2. Geometry of unicellular plankton for various cell shape types (A) Surface area as a function of cell volume. The dashed, and dotted, lines show the slope of a power law fit, and a scaling with the power of 2/3, respectively. (B) Aspect ratio, *r*, as a function of minimal cell dimension. The solid line shows a fitted sigmoidal function to the upper boundary of $|\log r|$ (black solid line). (C) Surface relative extension as a function of cell volume. The dotted and dashed black lines show 75% and 90% quantiles. (D) Distribution of taxonomic diversity as a function of cell volume. The black line shows a fitted Gaussian function. (E) Distribution of taxonomic diversity over cell surface extension (note the interchanged axes). The black line shows a fitted exponential function. The legend depicts the colour coding for different shape types, with the number of genera for each shape type given in parenthesis. See Table 1 for fitting parameters.

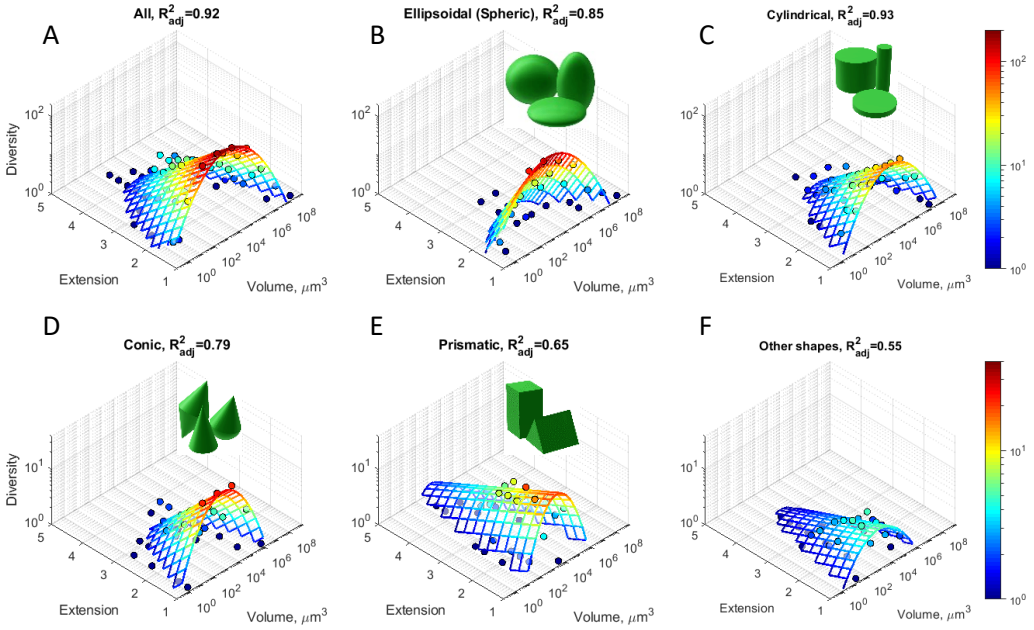
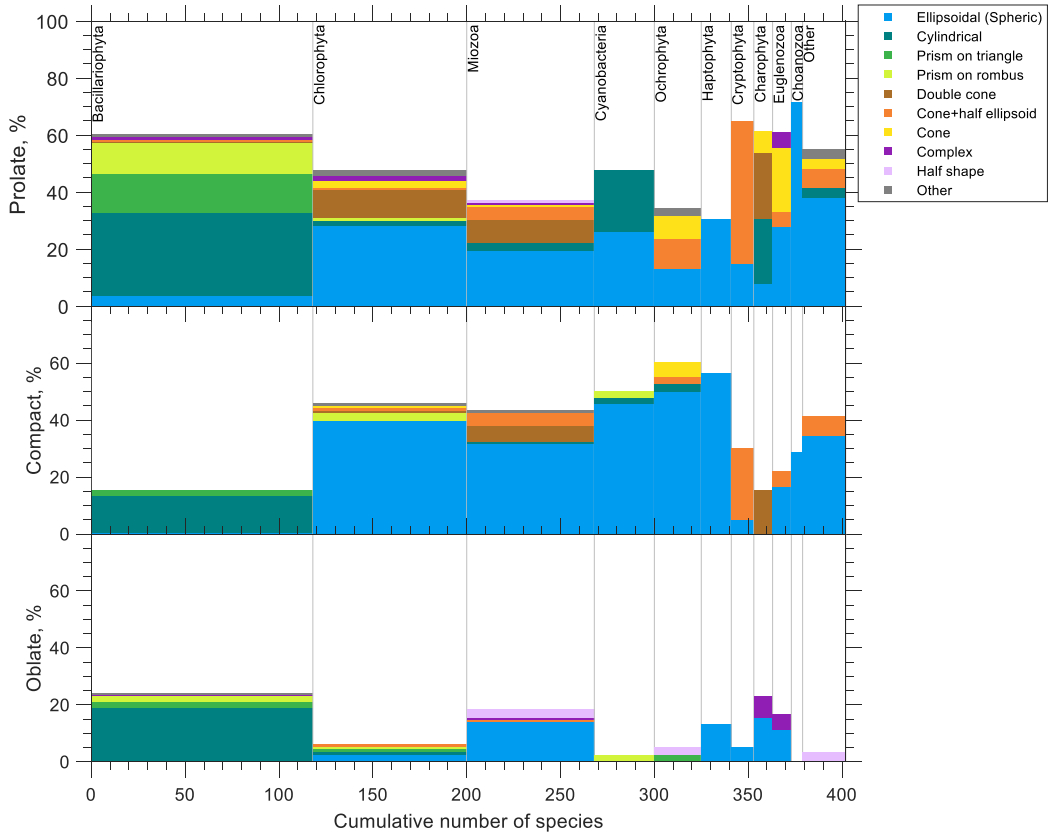
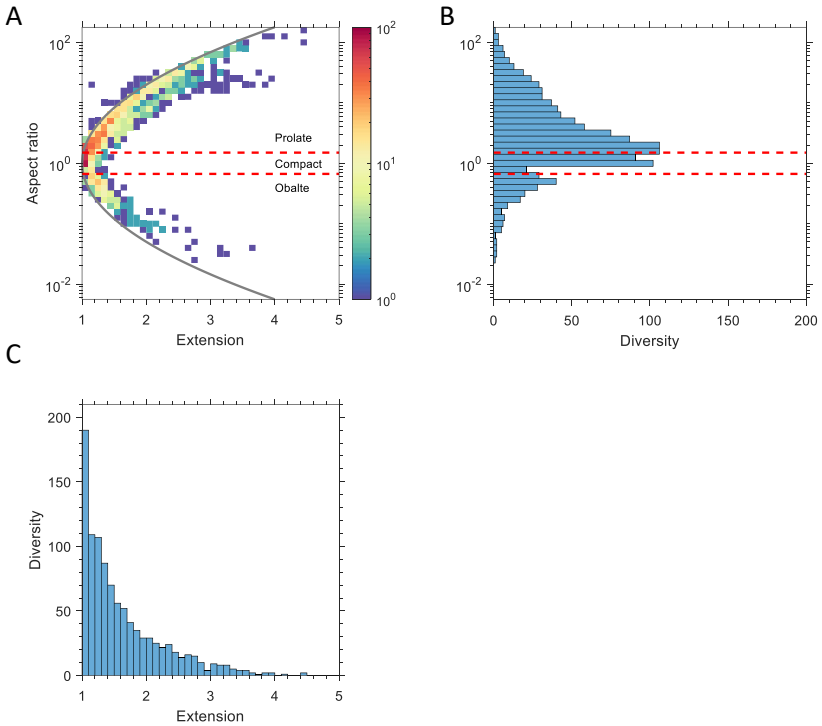


Fig. 3. Diversity distribution of unicellular phytoplankton. (A-F) Bivariate histograms of taxonomic diversity, D, as a function of surface extension, ϵ, and logarithm of cell volume, V, (dots), aggregated over all shape types (A) and for different shape types (B-F). Note that due to intraspecific and intragenus variability cells of the same genera can contribute to diversity in different bins. The mesh (solid lines) shows a fit by the function $\ln D = a - (\log V - \log V_0)^2/(2\ \sigma^2) - \alpha\epsilon$, weighted with diversity. The colours indicate taxonomic diversity from $D = 1$ (blue) to $D = 200$ (red) in A-C and to $D = 40$ in D-F. See Table 1 for regression results, and Extended Data Fig. 7 for comparison between predicted and observed diversity.

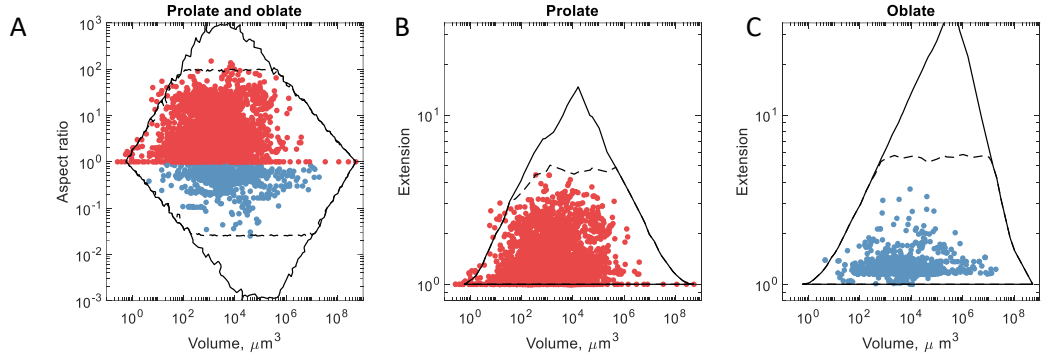
Extended Data



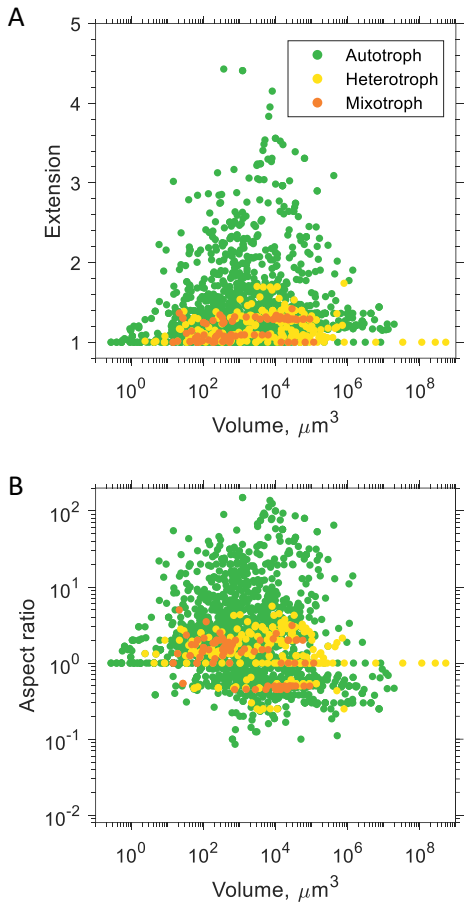
Extended Data Fig. 1. Diversity of phytoplankton across cell shapes (colour coded) and shape elongation (top, middle and bottom panel) for different phyla (columns). See Methods for classification of prolate, oblate and compact cells. Most of compact and prolate cells have cylindrical or prismatic shape in Bacillariophyta, conic shapes in Cryptophyta and Charophyta, and ellipsoidal shapes in the other phyla. Oblate cells are present in Bacillariophyta, Miozoa and Haptophyta, while for the other phyla their frequency is less than 10%, in particular oblate cells absent in cyanobacteria, Ochrophyta and Cryptophyta. Most of cylindrical and prismatic species belong to Bacillariophyta. Bacillariophyta almost do not contain ellipsoids which have a large fraction in the other phyla. Half-shapes (e.g. half-spheres or half-cones) are more dominated by oblate forms.



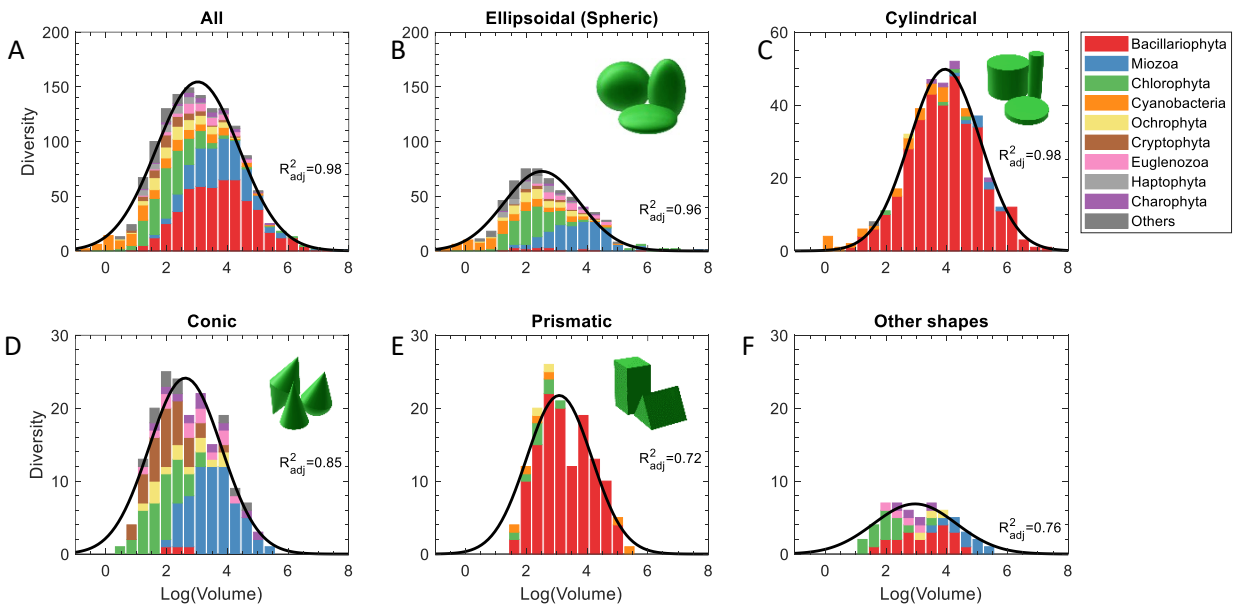
Extended Data Fig. 2. Bivariate effect of cell surface extension and aspect ratio on diversity. (A) Distribution of taxonomic diversity (shown by colour) over aspect ratio (logarithmic binning) and surface extension. The grey line shows a fitting parabola $\log r = \pm 1.3\sqrt{\epsilon - 1}$ to the upper boundary of the aspect ratio for a given surface extension. Horizontal red lines at $r = 3/2$ and $r = 2/3$ show the borders between compact, oblate and prolate cells, as defined in Methods. Diversity peaks for compact cells with smallest sphericity ($r = 1, \epsilon = 1$) and decreases both with increasing surface extension and absolute value of logarithm of aspect ratio. (B) Distribution of taxonomic diversity over aspect ratio. When projected on this axis the distribution of diversity shows peaks for cells with $r = 2$ and $r = 1/2$. We suppose that these peaks occur due to the specific shape of the distribution in Fig. A, where aspect ratios of compact shapes can change very fast with a small increase in surface extension, so the distribution is strongly stretched in the vertical direction resulting in a local minima at $r = 1$. (C) Distribution of taxonomic diversity over surface extension. In this projection the diversity distribution decays exponentially.



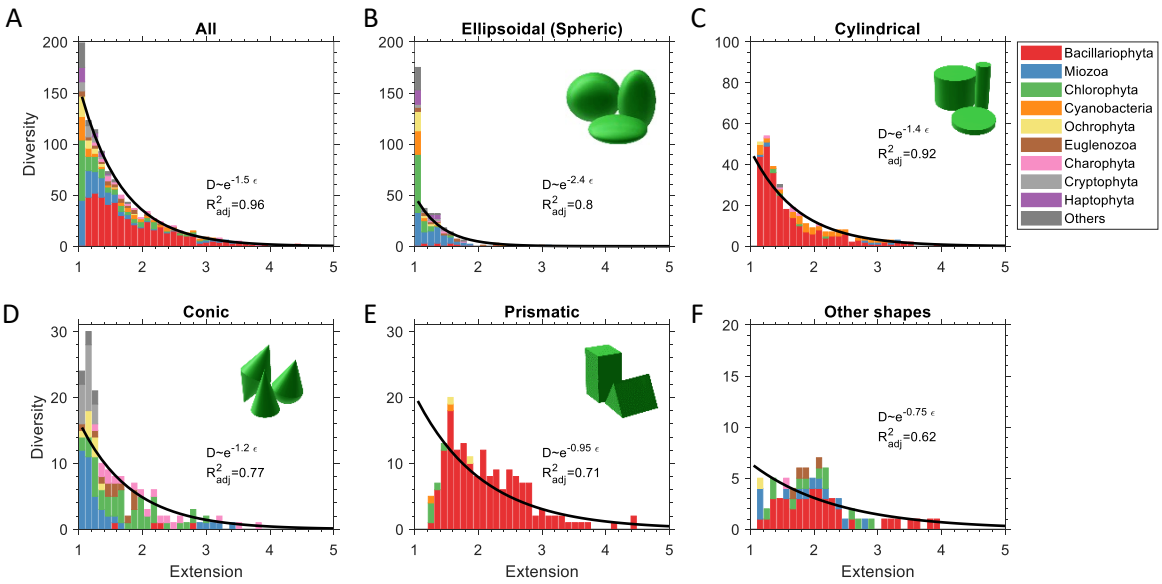
Extended Data Fig. 3. Prediction of surface extension and aspect ratio using various constraints on linear cell sizes. To make a theoretical prediction of a potential variation in aspect ratio and surface extension across cells we calculate the surface area and volume of ellipsoidal cells based on two models: (i) assuming all three linear dimensions are log-uniformly distributed in the range from 1 to 1000 μm and (ii) additionally assuming that the aspect ratio is constrained, $L_{max}/L_{min} < 100$ for prolate cells and $L_{max}/L_{min} < 40$ for oblate cells. (A) Comparison of the aspect ratio of prolate (red circulars) and oblate (blue circulars) cells with outer hulls for volume and aspect ratio in the first model (black line) and in the additionally constrained model (black dashed line). (B, C) the same for combinations of volume and surface extension for prolate (B) and oblate (C) cells. The first model, assuming only that cell dimensions can vary from 1 to 1000 μm, reproduces the hump-shaped dependence of maximal aspect ratio and elongation on volume (black solid lines show the outer hulls across 50,000 ellipsoids with randomly chosen linear dimensions), but this model strongly overestimates the maximal possible aspect ratio (ranges from 10^{-3} to 10^{3}) and surface extension (achieves 20 for prolate ellipsoids and more than 30 for oblate ellipsoids). The second model, with an additional constraint on cell aspect ratio, makes a relatively good prediction of the variation of aspect ratio and surface extension as a function of volume for prolate cells, but it overestimates the aspect ratio and surface extension for oblate cells. In particular, the model predicts that surface extension for oblate species can reach 5, while the observed maximal surface extension for oblate cells equals 2 for cells with intermediate.



Extended Data Fig. 4. Dependence of the geometry of unicellular plankton on cell volume for different nutritional modes. (A) Surface extension and (B) aspect ratio for different heterotrophic groups of plankton cells (colour coded are autotrophs, mixotrophs and heterotrophs) in dependence of cell volume. Based on data from Baltic sea, because only this data contained information on trophic levels of organisms.

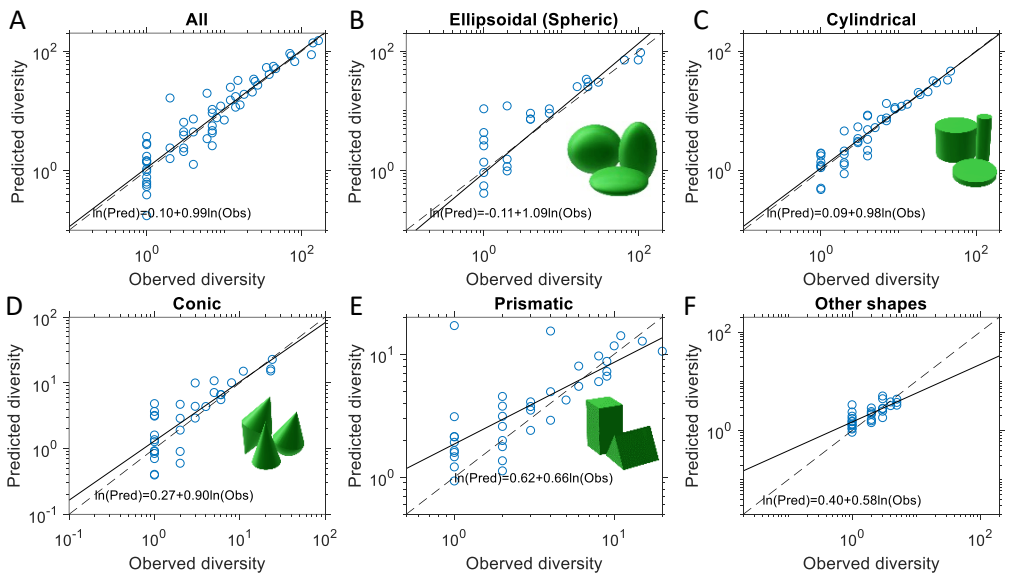


Extended Data Fig. 5. Distribution of taxonomic diversity as a function of volume for the most common shapes partitioned by phyla groups. Black lines show a least square fit of a Gaussian function $D = a \exp\left(-\frac{(\log V - \log V_0)^2}{2\sigma^2}\right)$ to the histogram (see Extended Data Table 1 for fitting parameters).

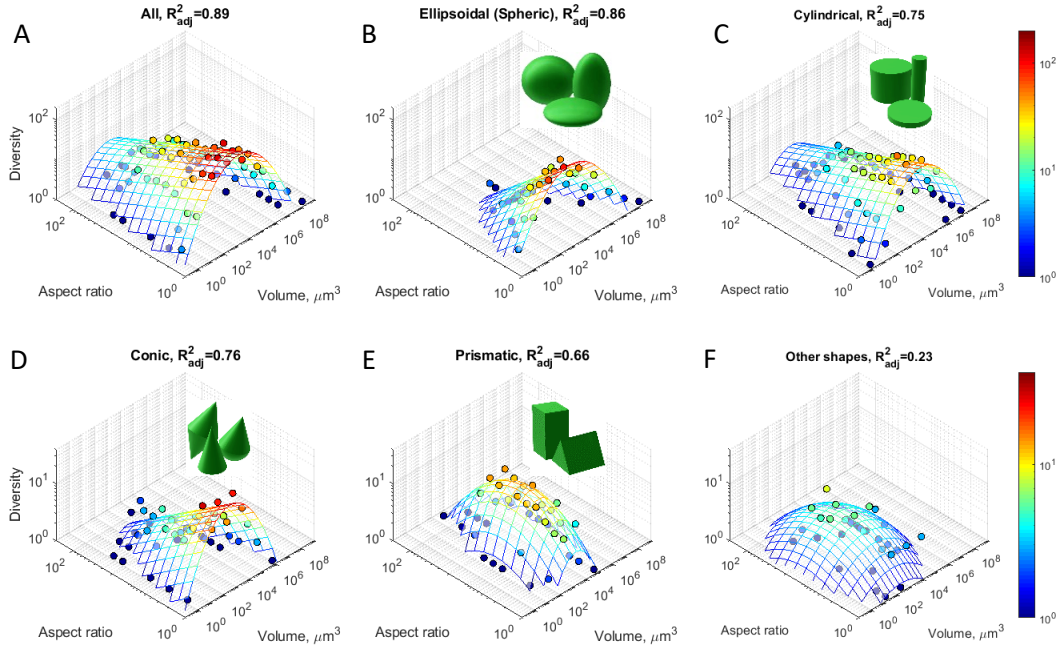


Extended Data Fig. 6. Distribution of taxonomic diversity as a function of surface extension for the most common shapes types partitioned by phyla. Solid lines show a least square fit of a linear function $\ln D = a - k\epsilon$ to the log-transformed histogram (see Extended Data Table 1 for fitting parameters).

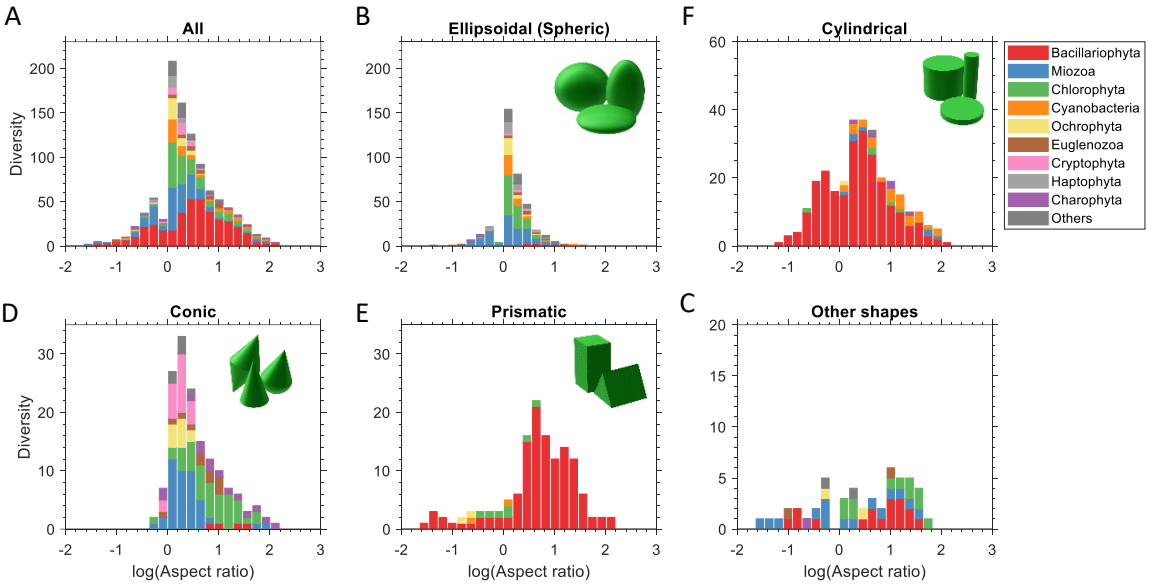
How correlations obtained across all shapes ($R^2 = 0.96$) can be larger that those obtained for specific shapes? The diversity distribution of elliptical genera (Fig. B) abruptly decreases with ϵ and elliptical genera have a strong tendency to be compact ($\epsilon \approx 1$). The maximum of diversity distribution for cylindrical and conic genera occurs at $\epsilon \approx 1.2$ (Fig. C,D). Finally, the diversity distributions of prismatic and other genera (Fig. E,F) decreases much slower with ϵ and exhibit some secondary maxima. This can be interpreted as a kind of niche separation between shapes classes along the gradient of surface extension. This niche separation diminish the quality of fit for each specific shape class, but it is not visible any more when we consider the entire distribution (Fig. A). The same explanation is applied for Fig. 3 (main text).



Extended Data Fig. 7. Comparison of observed diversity and diversity predicted based on nonlinear regression models in Fig. 3 (blue dots). Black dashed lines shows 1:1 diagonals and solid lines are linear regressions through the data points. The closer the solid line is to the dashed line, and the smaller the variability of datapoints around this line, the better is the prediction of diversity by the model function $D = a\exp(-(\log V - v_0)^2/(2\ \sigma^2) - k\epsilon)$ in Fig. 3 (main text). An increase in the variation of the predicted diversity in the range of small D can partly be explained by the fact that observed D is constrained by 1, while predicted values can be less than 1. The regression analysis shows that the predictions for ellipsoidal (**B**), cylindrical (**C**) and conic (**D**) shapes are unbiased, because the solid and dashed lines are almost parallel. By contrast, for prismatic (**E**) and other shapes (**F**) the regression lines deviate from the diagonals, and the model is biased as predictions of diversity in the range of small observed D are overestimated. However, as prismatic and other shapes are relatively rare, the model provides also a good and unbiased prediction of diversity across all shapes (**A**).



Extended Data Fig. 8. Diversity distribution of unicellular phytoplankton. Bivariate histogram of taxonomic diversity as a function of aspect ratio and volume. To reduce the number of fitting parameters the aspect ratio here is measured as L_{max}/L_{min}, so that no distinction between prolate and oblate cells has been made. Note that due to intraspecific and intragenus variability cells of the same genera can contribute to diversity in different bins. To provide a better fit for prismatic and other shapes (**E, F**), where diversity peaks at intermediate values of the aspect ratio, we also assumed a Gaussian dependence on the aspect ratio. See Extended Data Table 1 for the results of regression analysis.



Extended Data Fig. 9. The same as in Extended Data Fig. 6 but plotted as a function of the logarithm of the cell aspect ratio.

Figure	Model	R^2_{adj}	$b_1 \pm \delta_1$ (p)	$b_2 \pm \delta_2$ (p)	$b_3 \pm \delta_3$ (p)	$b_4 \pm \delta_4$ (p)	$b_5 \pm \delta_5$ (p)
Ext. Fig. 5A	$D = b_1 \exp\left(-\frac{(\log V - \log b_2)^2}{2(b_3)^2}\right)$	0.98	155±4	1100±100	1.36±0.04		
Ext. Fig. 5B		0.96	73±3	330±40	1.25±0.06		
Ext. Fig. 5C		0.98	50±1	8700±800	1.17±0.04		
Ext. Fig. 5D		0.85	24±2	400±10 (10^{-3})	1.19±0.1		
Ext. Fig. 5E		0.72	22±2 (1.5e-05)	1200±40 (10^{-2})	1.07±0.2 (10^{-4})		
Ext. Fig. 5F		0.76	6.9±0.5	900±30 (10^{-2})	1.36±0.2 (10^{-5})		
Ext. Fig. 6A	$\ln D = b_1 - b_4 \epsilon$	0.96	6.6±0.2	1.50±0.06			
Ext. Fig. 6B		0.8	6.3±0.6	2.4±0.3			
Ext. Fig. 6C		0.92	5.2±0.2	1.36±0.08			
Ext. Fig. 6D		0.77	4.3±0.3	1.2±0.1			
Ext. Fig. 6E		0.71	4.0±0.3	0.95±0.1			
Ext. Fig. 6F		0.62	2.6±0.3	0.75±0.1			
Ext. Fig. 7A	$\ln D_{pred} = b_1 + b_2 \ln D_{obs}$	0.85	0.1±0.1 (0.44)	0.99±0.05			
Ext. Fig. 7B		0.66	-0.1±0.3 (0.74)	1.2±0.2			
Ext. Fig. 7C		0.87	0.1±0.1 (0.41)	0.98±0.06			
Ext. Fig. 7D		0.65	0.3±0.2 (0.11)	0.9±0.1			
Ext. Fig. 7E		0.54	0.6±0.1 (1e-04)	0.6±0.1			
Ext. Fig. 7F		0.54	0.4±0.08 (5e-05)	0.6±0.1			
Ext. Fig. 8A	$\ln D = -\frac{(\log V - b_1)^2}{2\,(b_2)^2} + b_3 - b_4 r$	0.89	3.09±0.06	1.37±0.04	5.2±0.07	1.58±0.09	
Ext. Fig. 8B		0.86	2.7±0.1	1.47±0.07	4.8±0.1	3.18±0.3	
Ext. Fig. 8C		0.75	3.76±0.08	1.26±0.06	4.2±0.1	1.39±0.1	
Ext. Fig. 8D		0.76	2.5±0.1	1.4±0.1	3.5±0.1	1.78±0.2	
Ext. Fig. 8E	$\ln D = -\frac{(\log V - b_1)^2}{2\,(b_2)^2} - \frac{(\log V - b_1)^2}{2\,(b_2)^2}$	0.66	3.18±0.09	1.03±0.08	2.7±0.1	0.91±0.05	0.56±0.05
Ext. Fig. 8F		0.23	2.3±0.7 (0.0037)	2.1±0.7 (10^{-2})	1.7±0.2	1.0±0.1	0.6±0.2 (5 10^{-4}))

Extended Data Table 1. Fitting parameters for figures in the Extended Data. Parameter values are specified with standard error, p-value in brackets is shown only when $p > 10^{-5}$.