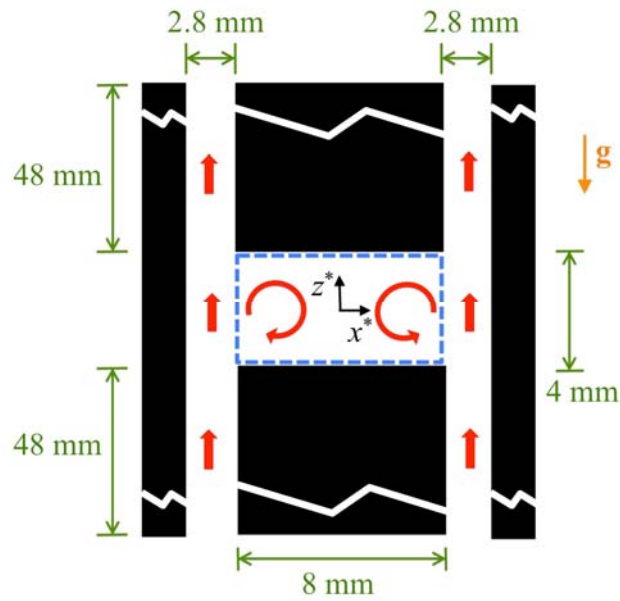
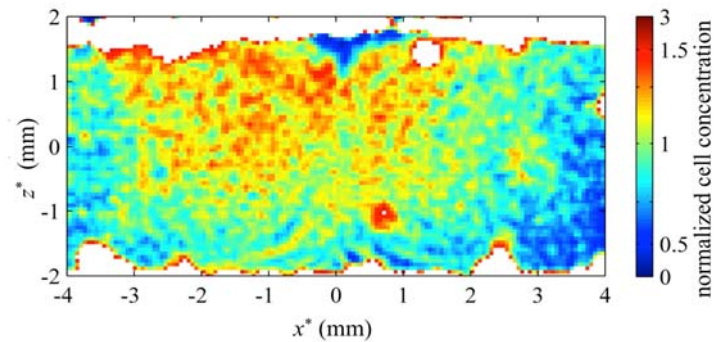


SUPPLEMENTARY INFORMATION

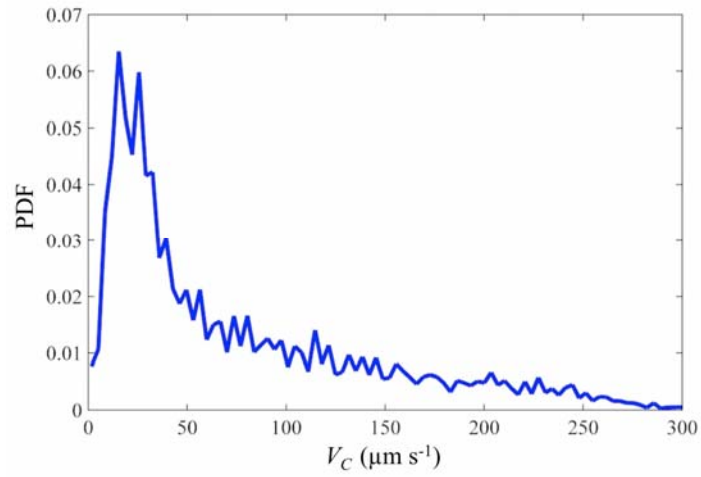
Supplementary Figures



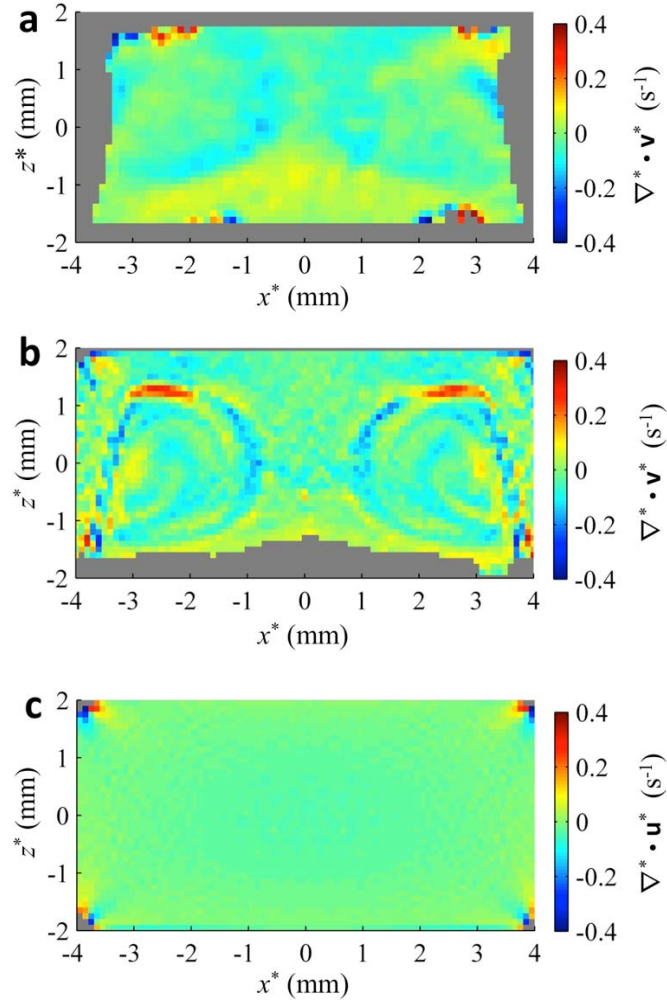
Supplementary Figure S1 | Schematic of the experimental device. The device extends 18 mm in the y^* direction (*i.e.*, into the plane). The central plane is denoted with the dashed blue box. Not drawn to scale.



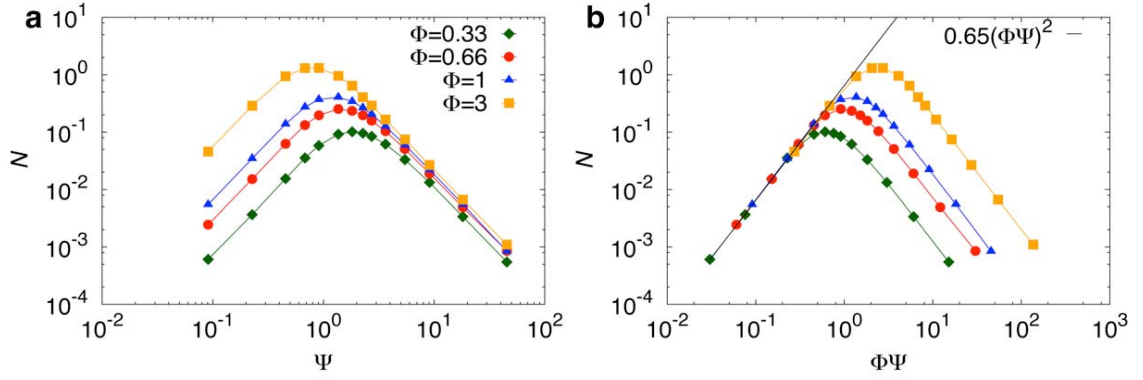
Supplementary Figure S2 | A histogram of cell concentration from experiments using dead cells reveals no aggregations. Cell concentrations were normalized using the same method as in Fig. 1c.



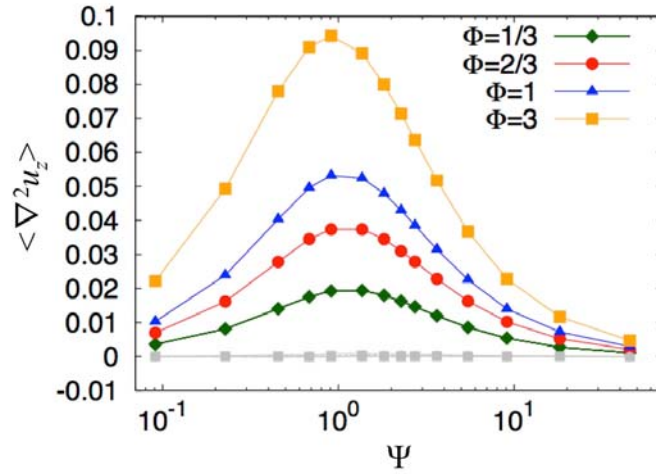
Supplementary Figure S3 | Probability density function of the three-dimensional swimming speed, V_C , of *Heterosigma akashiwo*. The three-dimensional swimming speed was estimated assuming $V_C = (2v_x^{*2} + v_z^{*2})^{1/2}$, where v_x^* and v_z^* are the instantaneous swimming speeds in the x and z direction, respectively.



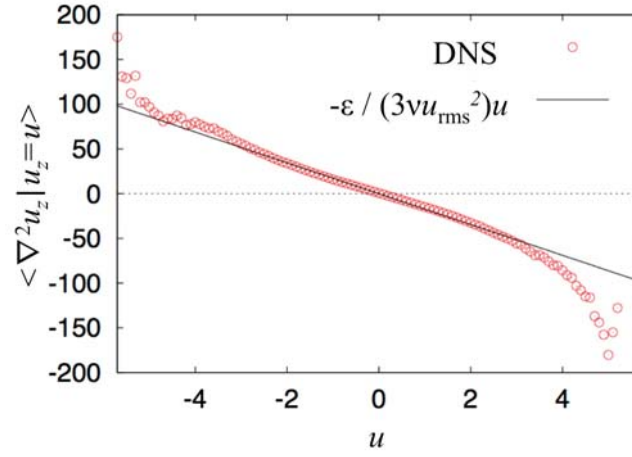
Supplementary Figure S4 | Measurements of divergence. The two-dimensional divergence of the y^* -averaged cell velocity ($\nabla^* \cdot \mathbf{v}^*$) in experiments (a) and from a model that simulates gyrotactic motility within the experimental flow (b) bear strong resemblance to one another, suggesting that the similar accumulation patterns in both (Fig. 1c,d) are borne by the same processes. In contrast, the divergence of the y^* -averaged fluid flow velocity ($\nabla^* \cdot \mathbf{u}^*$) exhibits a divergence of approximately zero within the central plane (c), except within the corner regions where the spatial sampling of the simulated velocity field is not sufficient for accurate differentiation.



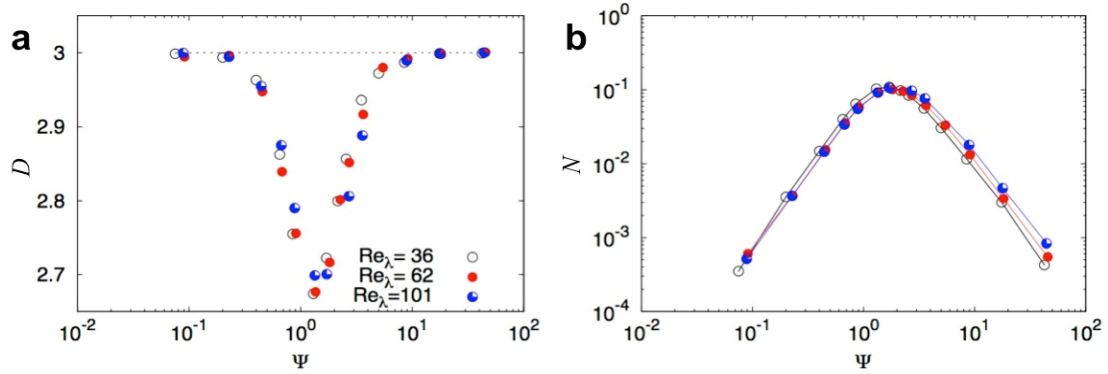
Supplementary Figure S5 | The normalized box probability index, N , as a function of (a) the cell stability number, Ψ , and (b) the predicted scaling (see Section 4 of Supplementary Methods). Results obtained using $\text{Re}_\lambda = 62$.



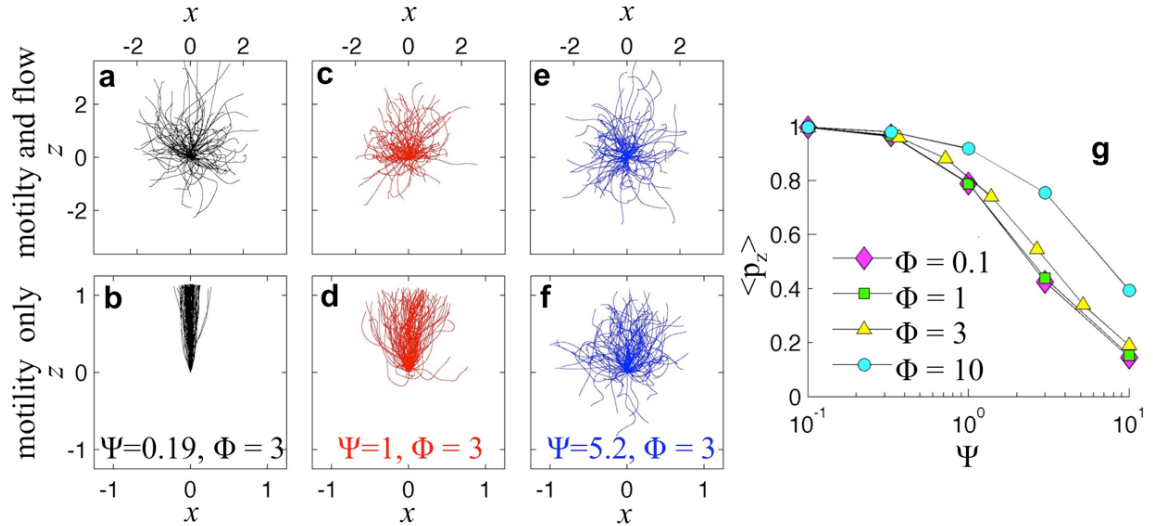
Supplementary Figure S6 | The non-dimensional Laplacian of the vertical component of the velocity of the fluid, $\nabla^2 u_z$, averaged over the position of 300,000 cells, as a function of the stability number Ψ for different swimming numbers Φ . Grey symbols show $\langle \nabla^2 u_x \rangle$ and $\langle \nabla^2 u_y \rangle$ for comparison. Results obtained using $\text{Re}_\lambda = 62$.



Supplementary Figure S7 | The average of the dimensional Laplacian of the vertical component of the fluid velocity conditioned on the vertical velocity value, $\langle \nabla^2 u_z | u_z = u \rangle$, as a function of the dimensional local vertical velocity, u (asterisks are omitted for brevity). The straight line shows the theoretical prediction (Eq. 10). The DNS data (red circles) were obtained by sampling the flow ($\text{Re}_\lambda = 62$) at random locations in the domain.



Supplementary Figure S8 | The aggregation intensity is only weakly dependent on the Taylor Reynolds number, Re_λ . The effect of Re_λ on (a) D and (b) N . Regardless of the Re_λ , the maximum level of aggregation occurs at a $\Psi \sim 1$, confirming that cell aggregations originate at the Kolmogorov scale. All results shown are for $\Phi = 1/3$.



Supplementary Figure S9 | Turbulence can inhibit the vertical migration of phytoplankton. (a-f), The trajectories of 100 cells swimming in a turbulent flow at three different stability numbers, Ψ . Although vertical migration is difficult to discern in the raw trajectories (a,c,e) due to advection by the flow, when the latter is subtracted (b,d,f) it is clear that turbulence progressively hampers vertical migration as Ψ increases. (g) The erosion of vertical migration was quantified by the average vertical projection of the swimming direction, $\langle p_z \rangle$, of 10^5 cells. Very stable cells ($\Psi \ll 1$) always swim vertically ($\langle p_z \rangle = 1$), whereas progressively more unstable cells (increasing Ψ) lose the ability to vertically migrate. In addition, vertical migration is favored by large swimming speeds (Φ), as fast swimming cells quickly traverse regions of locally enhanced shear, reducing the time over which they can be reoriented by flow. Panels a-f each show the two-dimensional projection of trajectories randomly distributed in the flow over 20 Kolmogorov timescales. Trajectories in (a,c,e) show movement due both advection and motility ($d\mathbf{X}/dt = \mathbf{u}(\mathbf{X}) + \Phi\mathbf{p}$), while trajectories in (b,d,f) have been postprocessed to show movement due to motility only ($d\mathbf{X}/dt = \Phi\mathbf{p}$). For visualization purposes the starting point of each trajectory has been translated to the origin. $Re_\lambda = 62$ for all panels.

Supplementary Table

Re_λ	M	Number of cells	λ/η_K
36	64^3	300,000	11.7
62	128^3	300,000	15.7
101	256^3	2,400,000	19.5
123	512^3	3,200,000	21.7

Supplementary Table S1 | Parameters of the direct numerical simulations of turbulence. Re_λ is the Taylor Reynolds number, M is the number of nodes in the computational mesh, and λ/η_K is the Taylor scale normalized by the Kolmogorov scale.

Supplementary Methods

1. Analysis of experimental measurements

Histograms of cell concentration within the region illuminated by the laser sheet (Fig. 1c and Supplementary Fig. S2) were obtained by identifying the (x^*, z^*) positions of the centroids of individual cells over a period of 4.4 min (5,300 images) and 3.8 min (4,540 images) for the swimming and killed cell treatments, respectively. To obtain the normalized cell concentration, the cumulative number of cells residing within $114 \mu\text{m} \times 114 \mu\text{m}$ bins over the entire experimental time was divided by the mean number of cells per bin within the region defined by $-4 \text{ mm} < x^* < 4 \text{ mm}$ and $-1.6 \text{ mm} < z^* < 1.6 \text{ mm}$, away from the device's solid boundaries.

To calculate the two-dimensional cell velocity, $\mathbf{v}^* = (v_x^*, v_z^*)$, within the central plane and its in-plane divergence, $\nabla^* \cdot \mathbf{v}^* = \partial v_x^* / \partial x^* + \partial v_z^* / \partial z^*$ (Supplementary Fig. S4a), we reconstructed 3.6×10^5 trajectories of individual cells swimming within the central plane ($-4 \text{ mm} < x^* < 4 \text{ mm}$, $-0.8 \text{ mm} < y^* < 0.8 \text{ mm}$, $-2 \text{ mm} < z^* < 2 \text{ mm}$) over time and space, using automated software (PredictiveTracker, ref. 42). The trajectories had an average length of 0.49 s (9.8 time points). Cell trajectories were used to obtain the Eulerian velocity field by calculating the mean cell velocity within $114 \mu\text{m} \times 114 \mu\text{m}$ bins in (x^*, z^*) space. An average of 1,515 velocity measurements were obtained for each bin. The velocity profile in Fig. 1b was obtained by tracking dead phytoplankton cells.

While the flow velocity, $\mathbf{u}^* = (u_x^*, u_y^*, u_z^*)$, within the device is three-dimensional, by symmetry we predict that $u_y^* = 0$ at every point within the plane, $y^* = 0$, and similarly that u_y^* averaged over a slice that symmetrically spans $y^* = 0$, $-0.8 \text{ mm} < y^* < 0.8 \text{ mm}$ (*i.e.* the laser illuminated ‘central plane’), equals zero. Therefore, the divergence of the y -averaged flow within the central plane is expected to be zero ($\nabla^* \cdot \mathbf{u}^* = \partial u_x^* / \partial x^* + \partial u_z^* / \partial z^* = 0$, due to the flow's incompressibility. In contrast, the divergence of the cell velocity field in the central plane is non-zero ($\nabla^* \cdot \mathbf{v}^* = \partial v_x^* / \partial x^* + \partial v_z^* / \partial z^* \neq 0$) and its sign can be used to diagnose regions of the flow where cells converge ($\nabla^* \cdot \mathbf{v}^* < 0$),

increasing the local cell concentration, from regions where cells diverge ($\nabla^* \cdot \mathbf{v}^* > 0$), decreasing the local cell concentration. The divergence field was calculated via numerical differentiation of the velocity field and was smoothed with a two-dimensional Gaussian filter (with standard deviation of 0.1 mm) to reduce noise (Supplementary Fig. S4a).

2. Analysis of simulations of gyrotactic motility within experimental flow field

The histogram of cell concentration for the simulations (Fig. 1d) was computed by sampling the x^* and z^* positions of the cells in $100 \mu\text{m} \times 100 \mu\text{m}$ bins within the region $-1 \text{ mm} < y^* < 1 \text{ mm}$, at intervals of 1 s. The cumulative counts over a simulation period of 2 min were divided by the mean cell concentration in the center region ($-4 \text{ mm} < x^* < 4 \text{ mm}$ and $-1.6 \text{ mm} < z^* < 1.6 \text{ mm}$) to obtain normalized cell concentrations.

The divergence field for the simulations (Supplementary Fig. S4b) was computed in a manner analogous to what done for the experiments: for trajectories residing in the region of the device defined by $-1 \text{ mm} < y^* < 1 \text{ mm}$, the x^* and z^* velocities were sampled in $100 \mu\text{m} \times 100 \mu\text{m}$ bins. The cell velocity field was constructed using the mean velocity from each bin; the two-dimensional divergence, $\nabla^* \cdot \mathbf{v}^* = \partial v_x^* / \partial x^* + \partial v_z^* / \partial z^*$, was computed from this velocity field using two-dimensional Gaussian filtering (with standard deviation of 0.05 mm) to reduce noise. An average of 386 velocity measurements was recorded in each bin.

The divergence field of the flow within the center plane ($\nabla^* \cdot \mathbf{u}^* = \partial u_x^* / \partial x^* + \partial u_z^* / \partial z^*$, Supplementary Fig. S4c) was obtained by averaging (u_x^* , u_z^*) over $-1 \text{ mm} < y^* < 1 \text{ mm}$ and calculating the derivatives numerically without smoothing.

3. Quantification of cell aggregation intensity in DNS simulations

To quantify the patchiness generated the interaction of gyrotactic motility with isotropic turbulence, we used three different aggregation metrics: the patch concentration enhancement factor, Q (Fig. 2e,f), the correlation fractal dimension, D (Fig. 3b,c; ref. 32), and the normalized box probability index, N (Supplementary Figs. S5 and S8b; ref. 53). Three different metrics were used to ensure that conclusions were not sensitive to the particular choice of metric. All three metrics indicate that maximal aggregation intensity occurred at intermediate stability (*i.e.*, $\Psi \sim 1$) and increased monotonically with increasing swimming speed (*i.e.*, increasing Φ).

Calculation of Q . Three-dimensional Voronoi tessellations of the particle distributions were obtained using the software Voropp²⁶, which accounted for the triply periodic boundaries of the computational volume (the Voronoi tessellation shown in Fig. 2d uses non-periodic boundaries, as only a small segment of the full computational domain is shown). The Q value presented for each $[\Psi, \Phi]$ pair was computed as the average value from eighteen snapshots of the instantaneous positions of 100,000 cells, each separated in time by at least 10 Kolmogorov timescales.

Calculation of D . The probability $p(r)$ (Fig. 3a) was obtained by computing the pairwise Euclidean distance between $K = 300,000$ cells in each of ≈ 100 snapshots of the particle distribution, each separated in time by more than 20 Kolmogorov timescales. In each snapshot, $p(r)$ was calculated as $p(r) = g(r)/K^2$, where $g(r)$ is the total number of pairwise distances less than r . Applying the definition $p(r) \sim r^D$, we

calculated D as the slope of $d(\log[p(r)])/d(\log[r])$ for $r/\eta_K \lesssim 1$ using the average $p(r)$ from the snapshots.

Calculation of N . The computational domain was partitioned into boxes of length $\Lambda \approx 3\eta_K$. The number of cells residing in each box, n , was used to obtain the box occupancy function, $f(n)$. While the mean of $f(n)$, $b = \langle n \rangle$, is a constant (equal to the mean cell number density), the standard deviation of $f(n)$, $\sigma = (\langle n^2 \rangle - \langle n \rangle^2)^{1/2}$, depends on the manner in which cells are distributed, with more patchy distributions having larger σ . If cells are randomly distributed throughout the domain, the standard deviation is that of a Poisson distribution, $\sigma_P = b^{1/2}$. To measure the particle distribution's deviation from Poisson, we calculated $N = (\sigma - \sigma_P)/b$ (ref. 53): for $N > 0$, the distribution of cells is more patchy than a random distribution (Supplementary Figs. S5 and S8b). To control for statistical fluctuations, this metric was calculated every 2 Kolmogorov time scales and N was taken as the average over more than 10^3 Kolmogorov time scales.

4. Theoretical prediction of N

The theoretical prediction of the fractal dimension D (section 8.2) can be related to the normalized box probability index N (ref. 54). Briefly, with the assumption that cells are randomly (Poisson) distributed over the fractal set of dimension $D < 3$ (and hence not uniformly distributed over the volume), one finds that the box probability $f(n)$ (the probability to have n cells in a box of size Λ) is related to the box probability on the fractal set $f_F(n)$, which is given by a Poisson distribution with mean b_F rescaled as $b_F = b (L_B/\Lambda)^{3-D}$, where L_B is the size of the domain.

Similarly, the variance of n , σ^2 , can be related to b as⁵⁴.

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = b \left[1 + b \left(\left(\frac{\Lambda}{L_B} \right)^{D-3} - 1 \right) \right]. \quad (\text{S1})$$

Applying the definition of N and via a Taylor expansion of $D-3$, one finds

$$N \approx \frac{(3-D)b^{1/2}}{2} \ln \left(\frac{L_B}{\Lambda} \right). \quad (\text{S2})$$

Substitution of Eq. (11) into Eq. (S2) yields the prediction $N \propto (\Phi\Psi)^2$ (for $\Psi \ll 1$), which is in agreement with simulations (Supplementary Fig. S5b).

5. Cell aggregation in flows with variable Taylor Reynolds number, Re_λ

The range of length scales within a turbulent flow is quantified by the Taylor Reynolds number, $\text{Re}_\lambda = u_{\text{rms}}\lambda/\nu$, where $\lambda = u_{\text{rms}}(15\nu/\varepsilon)^{1/2}$ is the Taylor length scale and u_{rms} is the root mean square fluid velocity. Larger Re_λ correspond to a greater separation between the size of integral-scale fluctuations, L , and the size of Kolmogorov-scale fluctuations, η_K . We calculated turbulent flows for four different Taylor Reynolds numbers, ranging from 36 to 123, by varying the number of mesh points M from 64^3 to 512^3 to ensure that the Kolmogorov scale η_K was fully resolved (i.e. $k_{\text{max}}\eta_K \geq 1$) in each case (Supplementary Table S1). In all simulations, the separation between forcing and dissipation scales was sufficient to produce the universal characteristics of a homogeneous isotropic turbulent flow. Data shown in Figs. 2 and 3 were obtained at $\text{Re}_\lambda = 62$.

We found that the aggregation intensity (as measured by N and D) was only weakly dependent on Re_λ for the parameters we tested (Supplementary Fig. S8), suggesting that motility-driven unmixing can overcome turbulent dispersion in a broad range of natural phytoplankton habitats, from pycnoclines ($Re_\lambda \approx 20$) to deep chlorophyll maxima ($Re_\lambda \approx 150$) (refs. 55,56). Furthermore, the finding that maximal aggregation occurs at $\Psi \sim 1$ is robust to changes in Re_λ within the range tested (Supplementary Fig. S8), provides further substantiation that motility-driven unmixing is controlled by the velocity gradients at the Kolmogorov scale.

Supplementary References

- 53 Fessler, J. R., Kulick, J. D. & Eaton, J. K. Preferential concentration of heavy particles in a turbulent channel flow. *Phys. Fluids* **6**, 3742-3749 (1994).
- 54 Dubrulle, B. & Lachi  ze-Rey, M. The poisson-distribution on a fractal. *Astron. Astrophys.* **287**, 361-367 (1994).
- 55 Maar, M., Nielsen, T. G., Stips, A. & Visser, A. W. Microscale distribution of zooplankton in relation to turbulent diffusion. *Limnol. Oceanogr.* **48**, 1312-1325 (2003).
- 56 Tennekes, H. & Lumley, J. L. *A first course in turbulence*. (MIT Press, Cambridge, USA, 1972).