Data-driven classification of individual cells by their non-Markovian motion

Anton Klimek,[1] Debasmita Mondal,[3,5] Stephan Block,[2] Prerna Sharma,[3,4] and Roland R. Netz[1,*]

[1]Fachbereich Physik, Freie Universität Berlin, Berlin, Germany; [2]Institut für Chemie und Biochemie, Freie Universität Berlin, Berlin, Germany; [3]Department of Physics, Indian Institute of Science, Bangalore, India; [4]Department of Bioengineering, Indian Institute of Science, Bangalore, India; and [5]James Franck Institute, University of Chicago, Chicago, Illinois

ABSTRACT   We present a method to differentiate organisms solely by their motion based on the generalized Langevin equation (GLE) and use it to distinguish two different swimming modes of strongly confined unicellular microalgae *Chlamydomonas reinhardtii*. The GLE is a general model for active or passive motion of organisms and particles that can be derived from a time-dependent general many-body Hamiltonian and in particular includes non-Markovian effects (i.e., the trajectory memory of its past). We extract all GLE parameters from individual cell trajectories and perform an unbiased cluster analysis to group them into different classes. For the specific cell population employed in the experiments, the GLE-based assignment into the two different swimming modes works perfectly, as checked by control experiments. The classification and sorting of single cells and organisms is important in different areas; our method, which is based on motion trajectories, offers wide-ranging applications in biology and medicine.

SIGNIFICANCE   Classification of cells is a common task in biology and medicine. We introduce the framework to accomplish such classification based on cell-center trajectories. Our method is based on the systematic theory for the dynamics of coarse-grained variables and extracts the underlying parameters describing active and passive cell motion. We apply our methodology to confined unicellular microalgae that exhibit two different swimming modes and show that we can accurately distinguish the two populations solely based on their motion pattern. Our method can be applied to time-series data of general observables from unicellular and multicellular organisms. We anticipate numerous applications in biology and medicine that require the label-free distinction and analysis of individual cells and organisms.

INTRODUCTION

Classifying individual cells or organisms is a challenging task that has been approached in many different ways and has ample applications. Distinguishing different types of cancer cells (1,2), foodborne pathogens (3), sperm cells (4), or types of neurons (5) are just a few examples. Different techniques have been introduced to distinguish and classify organisms on the multi-cell down to the single-cell level. One approach involves markers that bind cell specifically (2,6,7). Since individual cells contain unique genetic and epigenetic information, it is also possible to distinguish cells by their specific DNA or RNA content. Indeed, biotechnological advances enable single-cell RNA sequencing (8,9), which can be used in combination with machine-learning approaches (10,11), to efficiently distinguish single cells. However, RNA sequencing, as well as usage of markers, requires cell perturbation or even destruction for data acquisition. In many cases, it is desirable to classify cells without perturbing them, which requires label-free techniques such as spectroscopic approaches (3,12) or microscopy (13). In this way, cell information can be extracted almost instantaneously (14) from living organisms (1). Spectroscopic and microscopic images can be processed using machine learning to classify cells; however, the outcomes can be hard to interpret and require massive training data. One way to simplify the processing of cell-image data is to reduce the parameter space. This can be achieved by feature selection (15) or by projection onto important parameters, such as by principal-component analysis (11). A prime feature of mobile cells is their positional trajectory, which is relatively easy to obtain in experiments and contains hidden information on the motion-generating processes within the cell (16–20). Machine-learning algorithms have been proposed to classify
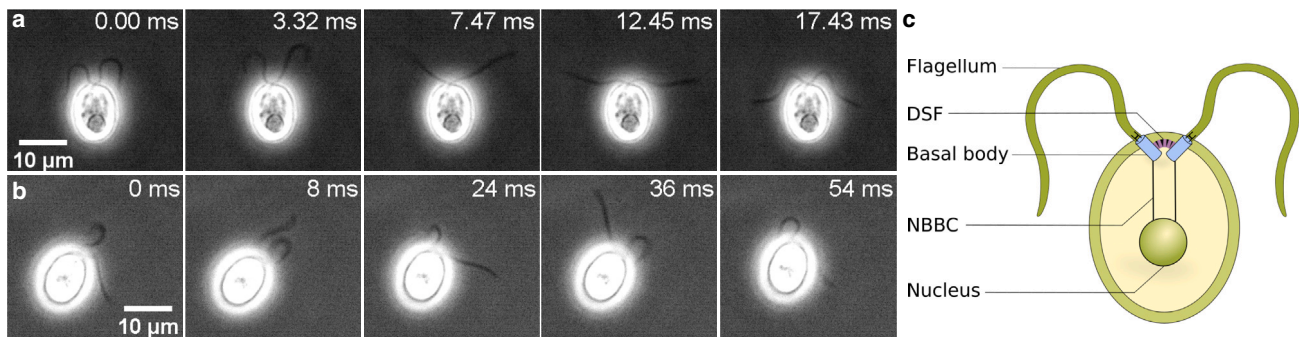
FIGURE 1   Unicellular microalgae microscopy. Sequences of phase-contrast microscopy images of CR algae exhibiting (*a*) synchro and (*b*) wobbler-type flagellar motion. The white halo around the cells is typical for phase-contrast microscopy (46). (*c*) Sketch of a CR cell: the distal striated fiber (DSF) connects the two basal bodies (47), which anchor the flagella and are connected to the nucleus by nuclear basal-body connectors (NBBCs) (45,48). To see this figure in color, go online.

trajectories (21) and some have been applied to single-cell trajectories (15). These approaches mostly focus on classification and not on the interpretation of the motion patterns. To yield a mechanistic interpretation, some specific model is usually assumed (22–27), which makes the interpretation model dependent. A general model that describes the motion of a particle in a complex environment, captures the stochasticity of its motion and can be derived from first principles is the GLE, which has been shown to accurately describe the motion of different cell types (28–32). In fact, the GLE is not an *ad hoc* model but can be derived from the underlying general many-body Hamiltonian (33–35). Living cells are intrinsically out of equilibrium (36), a fact that can be properly accounted for by the GLE used to describe the cell motion (37). In fact, there are many other models besides the GLE that have been successfully used to describe active and passive stochastic motion (38–42). The advantage of the GLE is that it makes minimal assumptions on the type of motion and encompasses many previously introduced models, such as the run-and-tumble model used to describe bacterial motion (43), as has been shown recently (44).

Here, we present a method to classify individual organisms based on the GLE parameters extracted from their motion trajectories. We apply our methods to experimental trajectories of individual unicellular biflagellate algae *Chlamydomonas reinhardtii* (CR) (45) and find two distinct groups of swimmers, which are illustrated in Fig. 1 *a* and *b*. In contrast to other existing methods for cell sorting and classification, our method requires only trajectories as input, does not need any training of a network, and avoids human bias in the selection of relevant features. Additionally, our approach allows us to interpret motion characteristics in terms of simple mechanistic models derived from the GLE parameters. In the case of CR cells, the data suggest some type of elastic coupling that presumably involves the anchoring of the flagella (45,47,48), as schematically depicted in Fig. 1 *c*, or a chemical feedback loop. Our approach is applicable to any kind of cell or organism motility data with sufficiently long trajectories and suffi-

ciently fine temporal discretization if the coordinate describing the motion corresponds to a Gaussian process, as will be explained in detail further below.

MATERIALS AND METHODS

Cell growth and sample preparation

Wild-type CR cell cultures (strain, CC-1690) are grown in tris-acetate-phosphate (TAP + P) medium by alternating light:dark (12:12 h) cycles for 3 days. We collect the cell suspension in their actively growing phase (between third and sixth day of culture) 2–3 h after the beginning of the light cycle and re-suspend it in fresh TAP + P medium. After 30 to 40 min of equilibration to recover from the mechanosensitive shock during re-suspension (49), we inject the cells inside a rectangular quasi-2D microfluidic chamber of height 10 μm and area 18 × 6 mm. This chamber is assembled by using a glass slide and coverslip sandwiched with a 10 − μm double-sided tape (Nitto Denko corporation) as spacer. The glass surfaces are pre-cleaned and coated with a polyacrylamide brush to suppress nonspecific adhesion of cell body and flagella (50). The chamber height is determined as 10.88 ± 0.68 μm across different samples (51).

Recording of trajectories

The cells in the chamber are placed under red light illumination ($>610\ nm$) to prevent phototaxis (52) and flagellar adhesion (53) of CR (51). We use high-speed video microscopy (Olympus IX83/IX73) at 500 frames per second with a 40× phase-contrast objective (Olympus, 0.65 NA, Plan N, PH2) connected to a metal oxide semiconductor (CMOS) camera (Phantom Miro C110, Vision Research, pixel size = 5.6 μm) for imaging the mid-plane between the confining glass plates. This setup enables us to simultaneously image cell position and flagellar shape. To capture very long trajectories to probe the long-time diffusive behavior in the supporting material, we use a 10× bright-field objective (Olympus, 0.25 NA, PlanC N) connected to a high-speed CMOS camera of higher pixel length (pco.1200hs, pixel size = 12 μm) at 50 frames per second. We determine cell trajectories by binarizing the image sequences with appropriate threshold parameters and tracking their centers using standard MATLAB routines (54).

Velocity autocorrelation function

Fourier transformation of Eqs. 23 and 24 leads to

$$\tilde{v}(\omega) = \frac{\tilde{F}_R(\omega)}{\tilde{\Gamma}_v^+(\omega) + i\omega} \qquad \text{(Equation 1)}$$

with the single-sided Fourier transform defined as $\tilde{\Gamma}_v^+(\omega) = \int_0^\infty e^{-i\omega t}\Gamma_v(t)dt$ and

$$\langle \tilde{F}_R(\omega)\tilde{F}_R(\omega')\rangle = 2\pi B\delta(\omega+\omega')\tilde{\Gamma}_R(\omega') . \qquad \text{(Equation 2)}$$

From the Fourier transform of the velocity autocorrelation function (VACF)

$$\begin{aligned}
\tilde{C}_{vv}(\omega) &= \int_{-\infty}^{\infty} dt e^{-i\omega t}\langle v(0)v(t)\rangle \\
&= \int_{-\infty}^{\infty} e^{-i\omega t}dt \int_{\infty}^{\infty} e^{i\omega t}\frac{d\omega}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\langle \tilde{v}(\omega)\tilde{v}(\omega')\rangle
\end{aligned}$$
$$\text{(Equation 3)}$$

we obtain by inserting Eqs. 1 and 2

$$\tilde{C}_{vv}(\omega) = \frac{B\tilde{\Gamma}_R(\omega)}{\left(\tilde{\Gamma}_v^+(\omega) + i\omega\right)\left(\tilde{\Gamma}_v^+(-\omega) - i\omega\right)} . \qquad \text{(Equation 4)}$$

Equating the nonequilibrium and the surrogate VACF in Eqs. 26 and 27 leads to

$$\frac{\tilde{\Gamma}_R(\omega)}{\left(\tilde{\Gamma}_v^+(\omega) + i\omega\right)\left(\tilde{\Gamma}_v^+(-\omega) - i\omega\right)} = \frac{\tilde{\Gamma}(\omega)}{|\tilde{\Gamma}^+(\omega) + i\omega|^2} . \qquad \text{(Equation 5)}$$

In the supporting material, we show that, for every correlation function $C_{vv}(t)$, we can determine a unique $\Gamma(t)$.

Memory kernel extraction

Multiplying the GLE Eq. 23 by $\dot{x}(t_0)$, averaging over the random force and integrating from t_0 to t leads to

$$(C_{vv}(t) - C_{vv}(0)) = -\int_0^t C_{vv}(s)G(t-s)ds, \qquad \text{(Equation 6)}$$

where we used that $\langle \dot{x}(t_0)F_R(t)\rangle = 0$ (33–35), set $t_0 = 0$, and introduced the integral kernel

$$G(t) = \int_0^t \Gamma(s)ds . \qquad \text{(Equation 7)}$$

To invert Eq. 6, we discretize it. Since $C_{vv}(t)$ is even but $G(t)$ is odd, we discretize $G(t)$ on half steps and $C_{vv}(t)$ on full steps and obtain (30)

$$G_{i+1/2} = \frac{2\left(C_{vv}^0 - C_{vv}^{i+1}\right)}{\Delta\left(C_{vv}^1 + C_{vv}^0\right)} - \sum_{j=1}^{i} G_{i-j+1/2}\frac{C_{vv}^{j+1} + C_{vv}^j}{C_{vv}^1 + C_{vv}^0}. \qquad \text{(Equation 8)}$$

The kernel Γ_i is obtained by the discrete derivative $\Gamma_i = \frac{G_{i+1/2}+G_{i-1/2}}{\Delta}$ with the initial value $\Gamma_0 = 2G_{1/2}/\Delta$.

Two-point velocity distribution

A stationary Gaussian process is completely described by its two-point probability distribution as shown in the supporting material. Here we show that the joint and conditional velocity distributions only depend on the VACF. The joint probability to observe v_2 at time t_2 and v_1 at time t_1 can be written in terms of the velocity vector $\vec{v} = (v_1, v_2)^T$ as

$$p(v_2, t_2; v_1, t_1) = \frac{\exp\left(-\vec{v}^T\Sigma^{-1}(t_2-t_1)\vec{v}/2\right)}{\sqrt{2\pi|\Sigma(t_2-t_1)|}} \qquad \text{(Equation 9)}$$

with $|\Sigma(t_2 - t_1)|$ denoting the determinant of the covariance matrix

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \langle v(t_1)v(t_1)\rangle & \langle v(t_1)v(t_2)\rangle \\ \langle v(t_1)v(t_2)\rangle & \langle v(t_2)v(t_2)\rangle \end{pmatrix} \\
&= \begin{pmatrix} C_{vv}(0) & C_{vv}(t_2 - t_1) \\ C_{vv}(t_2 - t_1) & C_{vv}(0) \end{pmatrix}.
\end{aligned}$$
$$\text{(Equation 10)}$$

Using the normal velocity distribution

$$p(v) = \frac{\exp\left(\frac{-v^2}{2C_{vv}(0)}\right)}{\sqrt{2\pi C_{vv}(0)}}, \qquad \text{(Equation 11)}$$

we obtain the conditional probability that $v(t_2) = v_2$ given $v(t_1) = v_1$, as

$$\begin{aligned}
p(v_2, t_2|v_1, t_1) &= \frac{p(v_2, t_2; v_1, t_1)}{p(v_1, t_1)} \\
&= \frac{\exp\left(-\frac{[v_2 - v_1(C_{vv}(t_2 - t_1))/C_{vv}(0)]^2}{\left(C_{vv}(0) - C_{vv}^2(t_2 - t_1)/C_{vv}(0)\right)}\right)}{\sqrt{2\pi\left(C_{vv}(0) - C_{vv}^2(t_2 - t_1)/C_{vv}(0)\right)}} .
\end{aligned}$$
$$\text{(Equation 12)}$$

Thus, the joint and conditional distributions only depend on the VACF $C_{vv}(t)$.

Fitting of friction kernel

The extracted friction kernel of each individual cell is fitted to Eq. 29 by least-square minimization (using the curve fit function of python's *scipy*) of the first 0.2 s of the data, which allows to estimate the parameter standard deviation by the diagonal of the parameter covariance. The δ-peak in Eq. 29 leads to the initial kernel value $\Gamma_0 = 2a/\Delta + b$ where Δ denotes the discretization time (see section "memory kernel extraction"). Since single cells exhibit large variations of the friction kernels, we constrain a and b in Eq. 29 to be between 0.1 % and 99.9 % of Γ_0. The decay time τ is constrained to be between 0.05 and 3 s and the frequency Ω is constrained between 20 and 250 s^{-1}.

Discretized VACF including localization noise

To test whether the cell motion is actually described by the friction kernel (Eq. 29) and does not originate from the experimental finite time step or noise, we fit the experimental VACF of individual cells with an analytical model that accounts for finite time discretization and noise (30). Here, we explain the fitting procedure; the analytical expression for the

mean-squared displacement (MSD) using a friction kernel in the form of Eq. 29 is derived in the supporting material.

We denote discrete values of a function $f(t)$ as $f(i\Delta) = f_i = f^i$ and the discretization time step as Δ. After smoothing the data by averaging over consecutive positions to reduce the localization noise, as discussed in detail in the supporting material, the velocities at half time steps follow as

$$v_{i+\frac{1}{2}} = \frac{x_{i+1} - x_i}{\Delta} . \qquad \text{(Equation 13)}$$

From the velocities, the VACF defined by Eq. 25 is calculated according to

$$C_{vv}^i = \frac{1}{N + 1 - i} \sum_{j=0}^{N-i} v_{j+\frac{1}{2}} v_{j+i+\frac{1}{2}}, \qquad \text{(Equation 14)}$$

with N being the number of trajectory steps. To account for localization noise, we assume Gaussian uncorrelated noise of width σ_{loc} at every time step, which gives the noisy MSD as (30)

$$C_{\text{MSD}}^{\text{noise}}(t) = C_{\text{MSD}}^{\text{theo}}(t) + 2(1 - \delta_{t0})\sigma_{\text{loc}}^2, \qquad \text{(Equation 15)}$$

where $C_{\text{MSD}}^{\text{theo}}(t)$ is the theoretical expression for the model MSD given in the supporting material, Eq. S51, and δ_{t0} is the Kronecker delta reflecting the uncorrelated nature of the localization noise. Since the observed trajectories are sampled with a finite time step Δ, we discretize the relation

$$C_{vv}(t) = \frac{1}{2}\frac{d^2}{dt^2}C_{\text{MSD}}(t), \qquad \text{(Equation 16)}$$

which leads to

$$C_{vv}^{\text{fit}}(i\Delta) = \frac{C_{\text{MSD}}^{\text{noise}}((i+1)\Delta) - 2C_{\text{MSD}}^{\text{noise}}(i\Delta) + C_{\text{MSD}}^{\text{noise}}((i-1)\Delta)}{2\Delta^2} . \qquad \text{(Equation 17)}$$

Finally, fits are performed by minimizing the cost function

$$E_{\text{cost}} = \sum_{i=0}^{n} \left(C_{vv}^{\text{exp}}(i\Delta) - C_{vv}^{\text{fit}}(i\Delta) \right)^2 \qquad \text{(Equation 18)}$$

with SciPy's least squares function in python and using Eq. 17 to determine $C_{vv}^{\text{fit}}(t)$ at discrete time points. As the MSD and VACF follow from the GLE Eq. 23 and the friction kernel (Eq. 29), the parameters to optimize are the kernel parameters a, b, τ, Ω, the mean-squared velocity B, and the localization noise width σ_{loc}. The data is fitted up to 0.2 s to disregard the noisy part of the VACF (see the supporting material for details). The cluster analysis is performed on the reduced parameter sets a, b, τ, Ω, B, which are obtained from the direct fit of Eq. 29 to the extracted kernel from the data.

Cluster analysis

The friction kernel in Eq. 29 contains four parameters; together with the mean-squared velocity B, each individual cell is characterized by five parameters. We perform an X-means cluster analysis (55), which is a generalization of the k-means algorithm (56). The k-means algorithm assigns unlabeled data to a predetermined number of k clusters by minimizing distances to the cluster centers. In the X-means algorithm, the number of clusters is not predetermined; we allow cluster numbers from 2 to 20. The algorithm starts with the minimal number of clusters and finds the cluster centers using k-means. It then splits every cluster into two subclusters whose centers are again determined by k-means. New subclusters are accepted if they improve the clustering quality accounting for the increased number of parameters. For this, we use the minimal noiseless description length criterion (57,58). We use an implementation of the X-means algorithm in Python (59) and use individual cell parameters as initial cluster centers (60). The X-means algorithm can converge to different final results depending on the initial cluster centers. Thus, we use all $59 \cdot 58/2$ possible combinations of initial cluster centers and use the result that occurs most often. We rescale each parameter by the median of its distribution.

Markovian embedding

A similar kernel to Eq. 29, namely

$$\Gamma(t) = 2a\delta(t) + be^{-t/\tau}\left(\cos(\Omega t) + \frac{1}{\tau\Omega}\sin(\Omega t)\right), \qquad \text{(Equation 19)}$$

can be derived from a system of harmonically coupled degrees of freedom. In fact, Eq. 19 becomes equivalent to Eq. 29 if the oscillation period $1/\Omega$ is much smaller than the decay time τ, which is the case for the extracted algae kernels. The Hamiltonian describing the coupled degrees of freedom takes the form

$$H = \frac{m}{2}v^2 + \frac{m_y}{2}v_y^2 + \frac{K}{2}(x - y)^2, \qquad \text{(Equation 20)}$$

where m_i are the effective masses of the two degrees of freedom and $K = bm$ is the harmonic coupling strength. In the presence of friction, quantified by friction coefficients γ_i, and coupling the degrees of freedom to a heat bath at temperature T, the coupled equations of motion are given by

$$\begin{aligned}
\dot{x}(t) &= v(t) \\
m\dot{v}(t) &= -\gamma_x v + bm(y(t) - x(t)) + F_{Rx}(t) \\
\dot{y}(t) &= v_y(t) \\
m_y\dot{v}_y(t) &= -\gamma_y v_y + bm(x(t) - y(t)) + F_{Ry}(t),
\end{aligned} \qquad \text{(Equation 21)}$$

where $F_{Rx}(t)$ and $F_{Ry}(t)$ are random forces with zero mean and second moment $\langle F_{Ri}(0)F_{Rj}(t)\rangle = \delta_{ij}2\gamma_i k_B T\delta(t)$. In the supporting material, it is shown that the coupled equations of motion Eq. 21 are equivalent to a GLE in the form of Eq. 23 for the coordinate $x(t)$ with a memory kernel $\Gamma(t)$ given by Eq. 19 (61). The friction of the first degree of freedom leads to the δ-contribution of $\Gamma(t)$ and the harmonic coupling to the second degree of freedom leads to the oscillating exponentially decaying contribution. The parameters of Eq. 21 translate into the parameters of Eq. 19 as

$$\begin{aligned}
a &= \frac{\gamma_x}{m} \\
\tau &= 2\frac{m_y}{\gamma_y} \\
\Omega &= \sqrt{\frac{bm}{m_y} - \frac{1}{\tau^2}} .
\end{aligned} \qquad \text{(Equation 22)}$$

RESULTS

Experimental trajectories

Our analysis is based on videos of 59 CR cells that are strongly confined between two glass plates separated by a distance similar to the cell diameter $\sim 10~\mu\text{m}$, which resembles the natural habitat of CR in soil (45) and simplifies the recording of long two-dimensional trajectories, as the cells cannot move out of the image plane of the microscope objective (see sections "cell growth and sample preparation" and "recording of trajectories" for experimental

details). Videos that resolve the flagella motion are shown in the supporting material and reveal two different types of flagellar motion (51). In one type, the flagella move synchronously as in a breaststroke called "synchro" (Fig. 1 *a* and Video S1); in the other type, the flagella move asynchronously, which results in a wobbling cell motion called "wobbler" (Fig. 1 *b* and Video S2). The emergence of two different swimming modes reflects cell-size variation and constitutes the tactile cell response to the confining surfaces (51), where the synchros tend to exhibit slightly larger cell bodies and therefore are more strongly confined. The confining surfaces are coated with an anti-adhesive polymer brush to prevent sticking of cells to the surfaces. The strong confinement leads to cell-surface friction, which is fully accounted for by the GLE analysis. In fact, unconfined CR cells do not exhibit distinct synchro and wobbler swimming modes (51,52). Switching events between synchronous and asynchronous flagella motion are never observed and thus are negligible. We use the classification into synchros and wobblers based on the flagella motion in the high-resolution video data as a test of our classification method that is based on the cell-center trajectories.

Theoretical trajectory model

The experiments yield two-dimensional trajectories $x(t), y(t)$ for the cell center position, which we describe by the GLE

$$\ddot{x}(t) = -\int_{t_0}^{t} \Gamma_v(t - t')\dot{x}(t')dt' + F_R(t) \quad \text{(Equation 23)}$$

with an identical equation for $y(t)$. Here, $\ddot{x}(t) = \dot{v}(t)$ denotes the acceleration of the cell position, $\Gamma_v(t)$ is a memory kernel that describes how the acceleration at time t depends on the cell velocity $\dot{x}(t') = v(t')$ at previous times and therefore accounts for non-Markovian friction effects, and $F_R(t)$ is a random force that describes interactions with the surrounding and within the interior of the cell. Since the experimental system is isotropic and homogeneous in space, no deterministic force term appears in the GLE. In fact, the GLE in Eq. 23 can be derived by projection at time t_0 from the underlying many-body Hamiltonian even in the presence of nonequilibrium effects, which obviously are present for living organisms (33–35,37).

If the cell motion can be described as a Gaussian process, which for CR cells is suggested by the fact that the single-cell velocity distributions are perfectly Gaussian, as will be demonstrated further below, the random force is a Gaussian process with correlations given by

$$\langle F_R(t)F_R(0)\rangle = B\Gamma_R(t), \quad \text{(Equation 24)}$$

where $B = \langle v^2 \rangle$ denotes the mean-squared cell velocity and the symmetric random-force kernel is denoted as $\Gamma_R(t)$. In

this case, the equation of motion is linear and there is no coupling between the motion in x and y direction, and we thus average all cell-trajectory data over the two directions.

For an equilibrium system, the fluctuation dissipation theorem (FDT) predicts $\Gamma_R(t) = \Gamma_v(|t|)$ with the mean-squared velocity given by $B = k_B T/m$ according to the equipartition theorem, where m is the mass of the moving object and $k_B T$ denotes the thermal energy (33,34). For living cells, both FDT and equipartition theorem do not hold in general and thus there is no a priori reason why $\Gamma_v(|t|)$ and $\Gamma_R(t)$ should be equal (36,62). Nevertheless, one can construct a surrogate model with an effective kernel $\Gamma(|t|) = \Gamma_R(t) = \Gamma_v(|t|)$ that exactly reproduces the dynamics described by the nonequilibrium GLE with $\Gamma_R(t) \neq \Gamma_v(|t|)$. This can be most easily seen by considering the VACF defined by

$$C_{vv}(t) = \langle v(0)v(t)\rangle, \quad \text{(Equation 25)}$$

whose Fourier transform $\tilde{C}_{vv}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t}C_{vv}(t)dt$ follows from Eqs. 23 and 24 as (30)

$$\tilde{C}_{vv}(\omega) = \frac{B\tilde{\Gamma}_R(\omega)}{\left(\tilde{\Gamma}_v^+(\omega) + i\omega\right)\left(\tilde{\Gamma}_v^+(-\omega) - i\omega\right)},$$
$$\text{(Equation 26)}$$

where $\tilde{\Gamma}_v^+(\omega)$ denotes the single-sided Fourier transform of $\Gamma_v(t)$ (see section Velocity autocorrelation function for the derivation). The VACF of the surrogate model with $\Gamma(|t|) = \Gamma_R(t) = \Gamma_v(|t|)$ follows from Eq. 26 as

$$\tilde{C}_{vv}^{\text{sur}}(\omega) = \frac{B\tilde{\Gamma}(\omega)}{(\tilde{\Gamma}^+(\omega) + i\omega)(\tilde{\Gamma}^+(-\omega) - i\omega)}.$$
$$\text{(Equation 27)}$$

For each combination of $\Gamma_R(t)$ and $\Gamma_v(t)$, there is a unique $\Gamma(t)$ that produces the same VACF; i.e., for which $\tilde{C}_{vv}^{\text{sur}}(\omega) = \tilde{C}_{vv}(\omega)$ holds (see the supporting material for the derivation) and which can be uniquely extracted from trajectories via the VACF (as shown in section "memory kernel extraction"). Since the VACF completely determines the dynamics of a Gaussian system, as shown in section "two-point velocity distribution" and in more detail in the supporting material, this implies that the extracted effective kernel $\Gamma(t)$ not only describes the VACF exactly but also characterizes the system completely (63). In fact, recent work, where the nonequilibrium GLE is derived from a suitably chosen time-dependent Hamiltonian, shows that, for Gaussian nonequilibrium observables, the condition $\Gamma_R(t) = \Gamma_v(|t|)$ is actually satisfied (37), in line with our method that is based on extracting an effective kernel $\Gamma(t)$.

Trajectories and velocity distributions

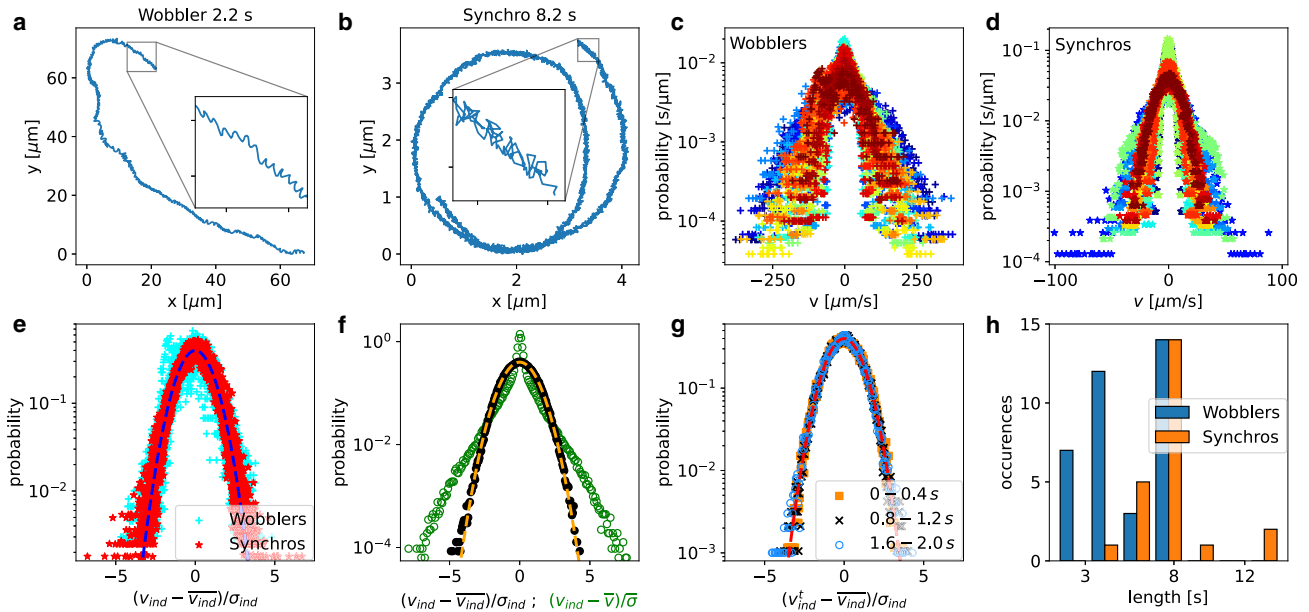Due to the asynchronous flagella motion of the wobblers, the cells turn in the flagella-beating rhythm and exhibit

FIGURE 2 Cell-center trajectories. Exemplary cell-center trajectory (54) (*a*) of a wobbler of duration 2.2 s and (*b*) of a synchro of duration 8.2 s. The insets show trajectory fragments of duration 0.2 s each. Velocity distributions of (*c*) wobblers and (*d*) synchros, individual cells are distinguished by color. (*e*) Velocity distributions of individual cells rescaled by subtracting the mean velocity of individual cells $\overline{v_{\rm ind}}$ and dividing by their standard deviation $\sigma_{\rm ind}$ for wobblers (*cyan crosses*) and synchros (*red stars*). The dashed line is the normal distribution. (*f*) Mean velocity distribution averaged over all cells (wobblers and synchros). For the green circles, the cell velocities are rescaled by subtracting the ensemble mean velocity $\overline{v}$ and dividing by the ensemble standard deviation $\overline{\sigma}$; for the black circles, the cell velocities are rescaled by subtracting the mean velocity of individual cells $\overline{v_{\rm ind}}$ and dividing by their standard deviation $\sigma_{\rm ind}$ as in (*e*). The normal distribution is indicated by a dashed line. (*g*) Individually rescaled velocity distributions of all cells for three different time windows. (*h*) Distribution of recorded trajectory lengths for wobblers and synchros. To see this figure in color, go online.

monotonically forward-moving wiggly trajectories, as shown in Fig. 2 *a*. In contrast, the synchronous flagella beating of synchros leads to fast switching between forward and backward motion, as shown in Fig. 2 *b*. As a consequence, synchros exhibit much slower net-forward motion than wobblers, as seen in Fig. 2 *a* and *b*, where trajectories with total duration 2.2 s (wobbler) and 8.2 s (synchro) are compared. Synchros also exhibit a somewhat narrower instantaneous velocity distribution, as seen in Fig. 2 *c* and *d*. However, a differentiation of the two cell types solely based on their speed does not work, as we will show later.

Even though individual cells exhibit pronounced variations in their velocity distributions, as seen from the large spread in Fig. 2 *c* and *d*, their velocity distributions are Gaussian, as demonstrated in Fig. 2 *e*: When subtracting from the cell velocities the mean velocity of each individual cell and dividing by the corresponding velocity standard deviation, $(v_{\rm ind} - \overline{v_{\rm ind}})/\sigma_{\rm ind}$, all individual velocity distributions collapse onto the Gaussian (normal) distribution (dashed line in Fig. 2 *e*). In contrast, when subtracting from the cell velocities the cell-ensemble mean velocity and dividing by the cell-ensemble standard deviation, $(v_{\rm ind} - \overline{v})/\sigma$, the velocity distribution averaged over all cells deviates strongly from a Gaussian (green circles in Fig. 2 *f*). In contrast, the individually rescaled velocity distribution averaged over all cells (black dots) perfectly agrees with the Gaussian normal distribution (dashed line in Fig. 2

f). Thus, single cells exhibit perfectly Gaussian velocity distributions, which suggests that the GLE with Gaussian noise is appropriate to analyze experimental single-cell trajectories and that the condition $\Gamma_R(t) = \Gamma_v(|t|)$ holds (37).

The GLE in Eq. 23 features time-independent parameters and thus describes a stationary process. That the cell velocity distribution does not change over the observational time is demonstrated in Fig. 2 *g*, where velocity distributions in three consecutive time intervals are compared (again subtracting the individual cell mean velocities and dividing by the corresponding velocity deviations of the entire trajectory). This suggests that the motion of individual CR algae can indeed be modeled by the GLE in Eq. 23. In this context, it is to be noted that wobblers are relatively fast and tend to move out of the camera window more quickly than synchros, leading to slightly shorter wobbler trajectories, as shown in Fig. 2 *h*.

Trajectory analysis and friction-kernel extraction

Trajectories are standardly characterized by the VACF or by the MSD

$$C_{\rm MSD}(t) = \langle (x(0) - x(t))^2 \rangle . \qquad \text{(Equation 28)}$$

Although the VACF is simply the curvature of the MSD, $C_{vv}(t) = \frac{1}{2}\frac{d^2}{dt^2}C_{\rm MSD}(t)$, the MSD and the VACF highlight
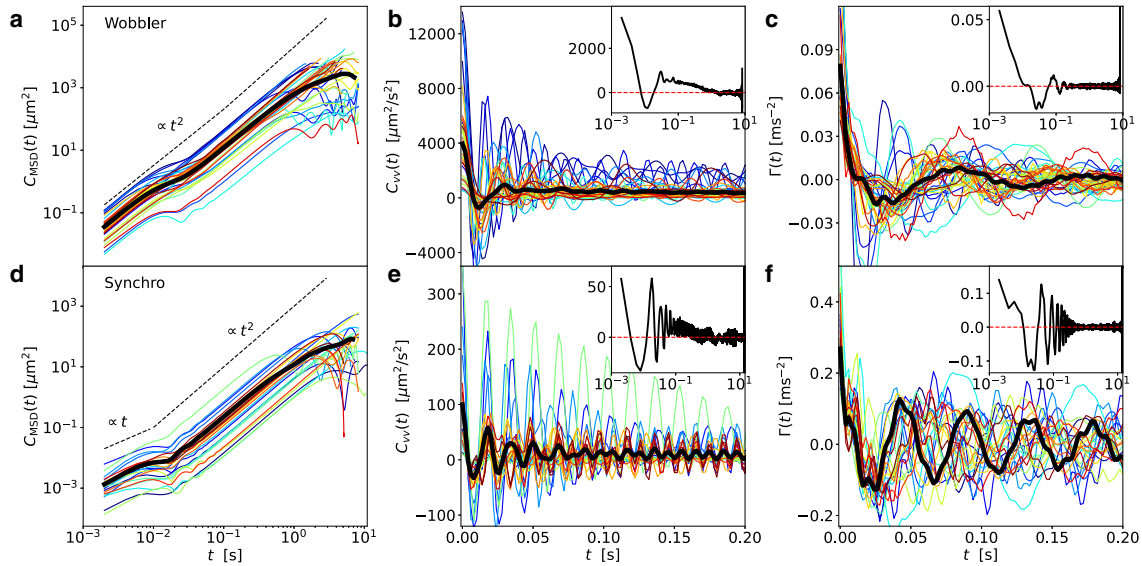
FIGURE 3 Results for the MSD, the VACF, and the friction kernel for wobblers and synchros. Results for the MSD, $C_{MSD}(t)$ defined in Eq. 28; the VACF, $C_{vv}(t)$ defined in Eq. 24; and the friction kernel $\Gamma(t)$, extracted according to Eq. 8, for wobblers in (*a*)–(*c*) and synchros in (*d*)–(*f*). Different colors represent results for individual cells; the black lines in (*a*), (*b*), (*d*), and (*e*) denote the average over all cells. For the friction kernels, the black line is computed from the average VACF. The dashed lines in (*a*) and (*d*) indicate ballistic and diffusive scaling. Insets show the long-time behavior of the average quantities on a lin-log scale. To see this figure in color, go online.

different aspects of the trajectories. In fact, the different propulsion modes of wobblers and synchros lead to drastically different MSDs: the wobblers exhibit ballistic behavior $C_{MSD}(t) \propto t^2$ on both short and long timescales (Fig. 3 *a*) with an intermediate crossover at $t \sim 0.02$ s. In contrast, the forward-backward motion of the synchros leads to short-time diffusive behavior $C_{MSD}(t) \propto t$ up to $t \sim 0.01$ s followed by a long-time ballistic regime for $t > 0.02$ s (Fig. 3 *d*). The transition to the intermediate ballistic regime occurs for both synchros and wobblers around the flagella oscillation period of the order of $t \sim 0.02$ s; the MSD for shorter times is dominated by the flagella motion and for longer times by the ballistic net-forward motion. The transition to asymptotic diffusive behavior, expected for long times, is for the synchros observed for $t > 2s$ in extended low-resolution microscopy data, whereas wobblers stay in the ballistic regime for the entire observation time of tens of seconds (see the supporting material). As wobblers move faster than synchros, their absolute VACF values are higher compared to the synchros, as seen in Fig. 3 *b* and *e*. The synchros exhibit less variation of the VACF among individual cells, which leads to slowly decaying oscillations in the VACF averaged over all cells (black line in Fig. 3 *e* compared to Fig. 3 *b*). These oscillations reflect the flagella-beating cycle.

We extract effective friction kernels $\Gamma(t)$ from the VACF of individual cells, as described in section "memory kernel extraction." The results, shown as colored lines in Fig. 3 *c* and *f*, demonstrate that the algal motion deviates strongly from the simple persistent random walk model, which is widely used to describe the motion of cells (28,29,32) and

which in the GLE formulation would correspond to $\Gamma(t)$ exhibiting a delta peak at $t = 0$ and otherwise being zero. The extracted memory kernels also reveal a substantially higher friction for synchros compared to wobblers, in line with the fact that synchros are larger and thus interact more strongly with the confining surfaces.

Comparing Fig. 3 *b* with Fig. 3 *c* or Fig. 3 *e* with Fig. 3 *f*, one notes that the oscillation period of the friction kernel is substantially longer than that of the VACF. The complex relation between the kernel and the VACF is discussed in the supporting material, where it is shown that the extracted values of the kernel decay time and oscillation amplitude achieve high directionality and speed of CR cells.

For both wobblers and synchros, the friction kernels exhibit an initial sharp peak followed by a decaying oscillation and are well described by

$$\Gamma(t) = 2a\delta(t) + be^{-t/\tau}\cos(\Omega t), \qquad \text{(Equation 29)}$$

with δ-peak amplitude a, oscillation amplitude b, exponential decay time τ, and oscillation frequency Ω. This is demonstrated in Fig. 4 *c* and *f* for one exemplary synchro and wobbler, where we compare the extracted memory kernels with fits according to Eq. 29, see section "fitting of friction kernel" for details. From these fits, we thus obtain four memory kernel parameters for each cell.

Before further analysis of the obtained individual cell parameters, we test whether the GLE Eq. 23 actually describes the cell motion. We thus compare the experimental MSD of a single wobbler and synchro in Fig. 4 *a* and *d* (orange dots) with the analytical prediction based on the GLE using the fit result for the friction kernel and the mean-squared velocity

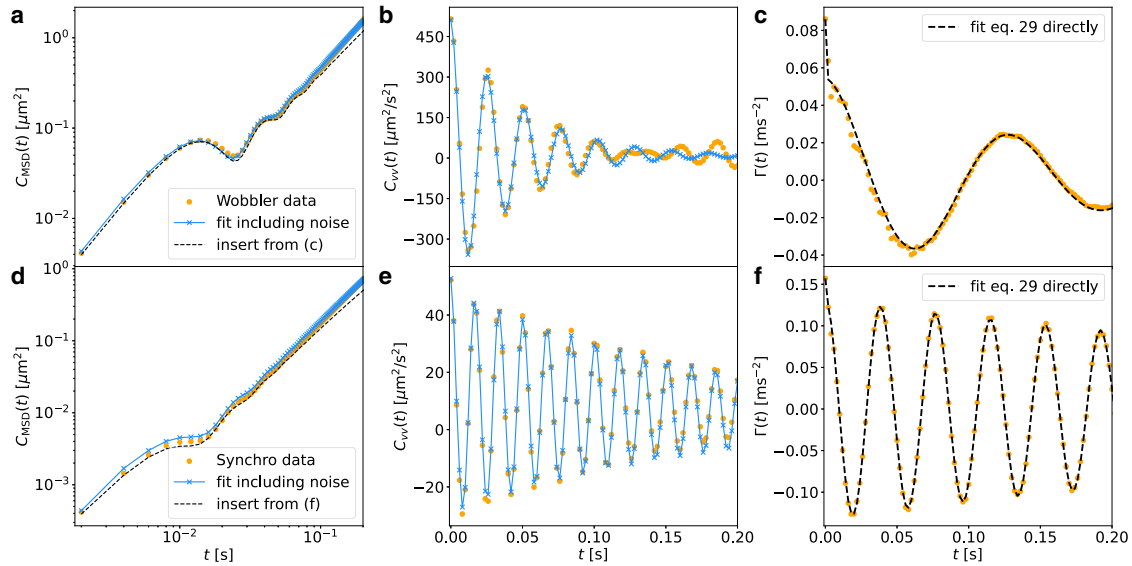FIGURE 4 Check the accuracy of the GLE for a single wobbler and synchro. Results for the MSD, $C_{MSD}(t)$; VACF, $C_{vv}(t)$; and friction kernel $\Gamma(t)$, of a single (*a–c*) wobbler and (*d–f*) synchro (*orange dots*). The black dashed lines in (*c*) and (*f*) denote fits of Eq. 29 to the extracted friction kernel determining the individual cell parameters shown in Fig. 5. The black dashed lines in (*a*) and (*d*) denote the analytical result Eq. S51 obtained from the friction-kernel fit in (*c*) and (*f*) and the mean-squared velocity $B = C_{vv}(0)$. Blue crosses in (*b*) and (*e*) denote a fit of the discretized expression for the VACF, including localization noise, (Eq. 17); blue crosses in (*a*) and (*d*) denote the corresponding prediction for the MSD, Eq. 15, using the same parameters as in (*b*) and (*e*); blue crosses are connected by blue straight lines. To see this figure in color, go online.

$B = C_{vv}(0)$ (black broken lines, the derivation of the analytical MSD expression is given in the supporting material). The agreement is very good, meaning that the memory extraction works well, which demonstrates that the GLE is an accurate model for the motion of organisms.

The MSD and VACF calculated from the GLE neglect the finite experimental recording time step of 0.002 s, and they also neglect the localization noise of the cell position, due to the finite spatial resolution of the microscopy images and the projection of a three dimensional object onto a two-dimensional point position (30) (see section "discretized VACF including localization noise" for details). The blue crosses in Fig. 4 *b* and *e* represent fits of the GLE-based analytical expression for the VACF, which includes localization noise and discretization effects, given in Eq. 17, to the experimental data. The blue crosses in Fig. 4 *a* and *d* show the corresponding MSD results with the same parameters according to Eq. 15. The agreement between experimental data and the discretized model is perfect, and the fitted localization noise strength, defined in Eq. 15, is of the order of $\sigma_{loc} \sim 0.02$ μm, similar to the pixel size, as expected (see discussion in the supporting material). This means that temporal and spatial discretization effects in the experimental data can be straightforwardly incorporated in the GLE model.

Clustering of single-cell parameters

The GLE Eq. 23 in conjunction with the random-force strength B defined in Eq. 24 and the effective friction kernel Eq. 29 has five parameters. This gives rise to 10 distinct two-

dimensional projections, which are shown in Fig. 5. Each data point corresponds to a single cell. The parameters exhibit substantial spread among individual cells, but an unambiguous separation between wobblers and synchros, here colored in blue and red, is not obvious. As can already be seen in Fig. 3 *c* and *f*, the friction amplitudes a and b are larger for the synchros, whereas the mean-squared velocity B is larger for wobblers, which leads to a separation of the two populations in Fig. 5 *b* and *c*. Each flagellar beating cycle leads to a net forward cell motion, which is reflected by the positive correlation between memory oscillation frequency Ω and mean-squared velocity B in Fig. 5 *g* for each cell type. The uncertainty of the parameters in Fig. 5, estimated from the diagonal fit covariances, is rather low except for the decay time τ, as seen in Fig. S5 in the supporting material. Since the parameter τ contributes only marginally to the clustering confidence, as demonstrated by the results in Figs. 5 and S4, we conclude that our cluster analysis is not affected by parameter uncertainties.

We perform an unbiased cluster analysis using the X-means algorithm (55), which is a general version of the k-means algorithm (56) that self-consistently determines the optimal number of clusters (details are given in section "cluster analysis"). Applying this unbiased cluster analysis to the single-cell parameters in five-dimensional space, we obtain two distinct groups, which perfectly coincide with the assignment into wobblers and synchros from visual analysis of the flagellar motion in the video data (51). This means that we can classify the cells by just knowing their center of mass trajectories with an accuracy of 100 %. In comparison,
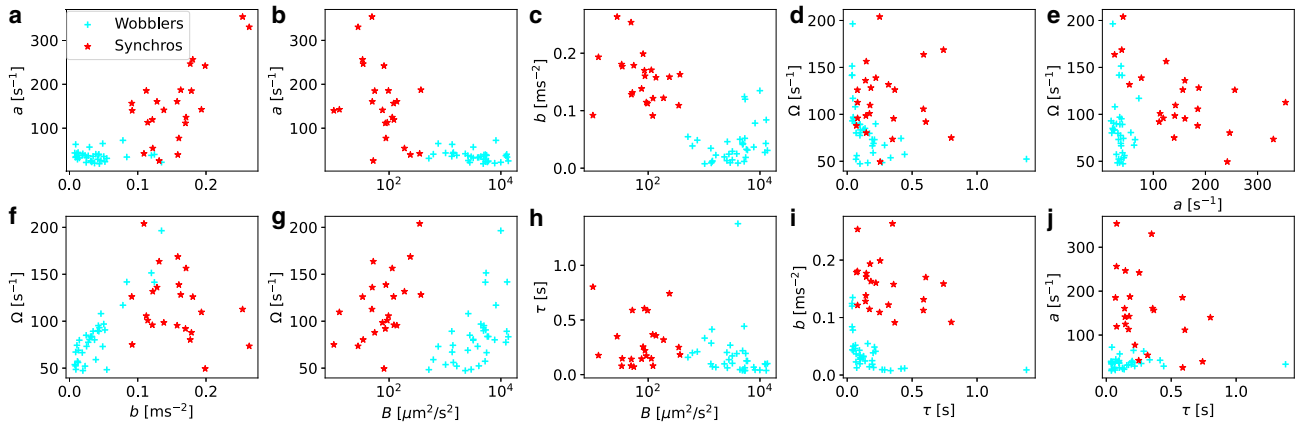
FIGURE 5 Scatter correlation plots of individual CR cell parameters. These consist of the friction-kernel parameters (a, b, τ, Ω) defined in Eq. 29 and the mean-squared velocity B. All parameters except B are presented on a linear scale. Synchros are shown in red and wobblers in cyan according to our cluster analysis, which perfectly matches a categorization based on the visual analysis of flagella motion. Error estimates of the extracted parameters are omitted here for clarity and shown in Fig. S5. To see this figure in color, go online.

a cluster analysis solely based on the mean-squared velocity B leads to an accuracy of only 69 %, whereas a cluster analysis using the first two principal-component analysis components leads to an accuracy of 90 % (see the supporting material). Using only the four friction-kernel parameters for a cluster analysis without the mean-squared velocity B still reaches an accuracy of 92 %. This clearly demonstrates that the GLE model, which parameterizes cell motion based on friction-kernel parameters in Eq. 29 together with the mean-squared velocity B, allows for accurate classification of cells based only on their motion.

DISCUSSION AND CONCLUSIONS

We demonstrate that the rather complex motion of individual CR algae can be accurately described by the GLE Eq. 23 and extract all GLE model parameters for individual algae from their cell-center trajectories in a data-driven manner. Based on the extracted GLE parameters, we detect two distinct algae classes by an unbiased cluster analysis; this unsupervised clustering result is confirmed by comparison with a categorization based on visual inspection of the flagella beating patterns. Our method is applicable to any kind of cells and even higher organisms if the motion is a Gaussian process. Cell and animal motion often exhibits aging effects (64), which in the GLE framework would show up as very slowly decaying contributions to the memory function. To accurately extract slowly decaying memory functions from data, very long trajectories would be needed. Since the GLE Eq. 23 is not restricted to positional degrees of freedom, our approach can be applied to any observable (for instance, cell extension or deformation). As the only inputs needed are trajectories, our method requires minimal interaction with the organisms and can be easily used as a stand-alone tool or to improve existing machine-learning algorithms for cell classification.

Additionally, our approach allows for a mechanistic interpretation of cell-motion characteristics. In fact, a friction-kernel model that is very similar to Eq. 29 and describes the data equally well can be derived from the equation of motion of two elastically coupled objects (see section "markovian embedding" and the supporting material for details). Without further experimental input, our approach does not reveal what these objects are, so we can only speculate that the elastic coupling between the cell body and the flagella, which presumably involves the connection between the flagellar basal bodies by the distal striated fiber (schematically shown in Fig. 1 c), causes the slowly decaying oscillations in the memory kernel. This seems in line with previous models for CR algae motion and flagella synchronization (65,66). Alternatively, the kernel oscillations could also be caused by some chemical or hydrodynamic feedback loop. Clearly, more experiments that resolve the relative motion of the cell center and the flagella are needed to resolve these issues.

Our extraction of GLE parameters from cell trajectories yields a single effective memory kernel and does not allow detection of the nonequilibrium character of cell motion, in agreement with recent general arguments (63). Conversely, based on an explicit nonequilibrium model for cell motion, it is rather straightforward to derive the GLE Eq. 23 and the functional form of the extracted kernel (Eq. 29), as shown in the supporting material. A similar GLE model can also be derived for motion in a confining potential to describe confined neurons (5) or bacteria moving in mucus (67) (see the supporting material for details).

As a final note, we mention that the angular orientation of CR algae can be accurately extracted from their cell-center trajectory, as shown in the supporting material. Thus, the orientational cell dynamics is included in our GLE model.

In summary, our approach allows for cell classification by positional cell-center trajectories or any other kind of

time-series data and at the same time for interpretation of the motion pattern in terms of intracellular interactions. We anticipate numerous applications in biology and medicine that require the label-free distinction of individual cells and organisms.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2024.03.023.

AUTHOR CONTRIBUTIONS

D.M. and P.S. designed and performed the experiments. S.B. contributed analytic tools and analyzed data. A.K. and R.R.N. designed the models, analyzed the experimental data, and wrote the manuscript.

ACKNOWLEDGMENTS

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Gessert, N., M. Bengs, …, D. B. Ellebrecht. 2019. Deep transfer learning methods for colon cancer classification in confocal laser microscopy images. *Int. J. Comput. Assist. Radiol. Surg.* 14:1837–1845.

2. Teng, M. W. L., S. F. Ngiow, …, M. J. Smyth. 2015. Classifying cancers based on T-cell infiltration and PD-L1. *Cancer Res.* 75:2139–2145.

3. Kang, R., B. Park, …, K. Chen. 2020. Single-cell classification of food-borne pathogens using hyperspectral microscope imaging coupled with deep learning frameworks. *Sensor. Actuator. B Chem.* 309, 127789.

4. Davis, R. O., and R. J. Siemers. 1995. Derivation and reliability of kinematic measures of sperm motion. *Reprod. Fertil. Dev.* 7:857–869.

5. Zeng, H., and J. R. Sanes. 2017. Neuronal cell-type classification: challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* 18:530–546.

6. Zhang, L., F. Wang, …, Y. Zhu. 2020. Classifying Cell Types with DNA-Encoded Ligand–Receptor Interactions on the Cell Membrane. *Nano Lett.* 20:3521–3527.

7. Sanz, I., C. Wei, …, F. E.-H. Lee. 2019. Challenges and opportunities for consistent classification of human B cell and plasma cell populations. *Front. Immunol.* 10:2458.

8. Saliba, A.-E., A. J. Westermann, …, J. Vogel. 2014. Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* 42:8845–8860.

9. Kolodziejczyk, A. A., J. K. Kim, …, S. A. Teichmann. 2015. The technology and biology of single-cell RNA sequencing. *Mol. Cell.* 58:610–620.

10. Papalexi, E., and R. Satija. 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 18:35–45.

11. Qi, R., A. Ma, …, Q. Zou. 2020. Clustering and classification methods for single-cell RNA-sequencing data. *Briefings Bioinf.* 21:1196–1208.

12. Chen, C. L., A. Mahjoubfar, …, B. Jalali. 2016. Deep learning in label-free cell classification. *Sci. Rep.* 6:21471.

13. Shen, S., K. Syal, …, S. Wang. 2015. Note: An automated image analysis method for high-throughput classification of surface-bound bacterial cell motions. *Rev. Sci. Instrum.* 86, 126104.

14. Liu, X., and Y.-T. Chang. 2022. Fluorescent probe strategy for live cell distinction. *Chem. Soc. Rev.* 51:1573–1591.

15. Sbalzarini, I. F., J. Theriot, and P. Koumoutsakos. 2002. Machine learning for biological trajectory classification applications. *In* Studying Turbulence Using Numerical Simulation Databases-IX: Proceedings of the 2002 Summer Program.

16. Boedeker, H. U., C. Beta, …, E. Bodenschatz. 2010. Quantitative analysis of random ameboid motion. *Europhys. Lett.* 90, 28005.

17. Pohl, O., M. Hintsche, …, H. Stark. 2017. Inferring the chemotactic strategy of P. putida and E. coli using modified Kramers-Moyal coefficients. *PLoS Comput. Biol.* 13, e1005329.

18. Maiuri, P., J.-F. Rupprecht, …, R. Voituriez. 2015. Actin flows mediate a universal coupling between cell speed and cell persistence. *Cell.* 161:374–386.

19. Heuzé, M. L., P. Vargas, …, A. M. non-Lennon-Duménil. 2013. Migration of dendritic cells: physical principles, molecular mechanisms, and functional implications. *Immunol. Rev.* 256:240–254.

20. Liu, Y.-J., M. Le Berre, …, M. Piel. 2015. Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells. *Cell.* 160:659–672.

21. da Silva, C. L., L. M. Petry, and V. Bogorny. 2019. A survey and comparison of trajectory classification methods. *In* 2019 8th Brazilian Conference on Intelligent Systems (BRACIS) IEEE, pp. 788–793.

22. Amselem, G., M. Theves, …, C. Beta. 2012. A stochastic description of Dictyostelium chemotaxis. *PLoS One.* 7, e37213.

23. Selmeczi, D., L. Li, …, H. Flyvbjerg. 2008. Cell motility as random motion: A review: Cell motility as random motion. *Eur. Phys. J. Spec. Top.* 157:1–15.

24. Selmeczi, D., S. Mosler, …, H. Flyvbjerg. 2005. Cell motility as persistent random motion: theories from experiments. *Biophys. J.* 89:912–931.

25. Codling, E. A., M. J. Plank, and S. Benhamou. 2008. Random walk models in biology. *J. R. Soc. Interface.* 5:813–834.

26. Pedersen, J. N., L. Li, …, H. Flyvbjerg. 2016. How to connect time-lapse recorded trajectories of motile microorganisms with dynamical models in continuous time. *Phys. Rev. E.* 94, 062401.

27. Dieterich, P., R. Klages, …, A. Schwab. 2008. Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. USA.* 105:459–463.

28. Gail, M. H., and C. W. Boone. 1970. The locomotion of mouse fibroblasts in tissue culture. *Biophys. J.* 10:980–993.

29. Wright, A., Y.-H. Li, and C. Zhu. 2008. The differential effect of endothelial cell factors on in vitro motility of malignant and non-malignant cells. *Ann. Biomed. Eng.* 36:958–969.

30. Mitterwallner, B. G., C. Schreiber, …, R. R. Netz. 2020. Non-Markovian data-driven modeling of single-cell motility. *Phys. Rev. E.* 101, 032408. https://doi.org/10.1103/PhysRevE.101.032408.

31. Li, L., E. C. Cox, and H. Flyvbjerg. 2011. 'Dicty dynamics': Dictyostelium motility as persistent random motion. *Phys. Biol.* 8, 046006.

32. Li, L., S. F. Nørrelykke, and E. C. Cox. 2008. Persistent cell motion in the absence of external signals: a search strategy for eukaryotic cells. *PLoS One.* 3, e2093.

33. Mori, H. 1965. Transport, Collective Motion, and Brownian Motion. *Prog. Theor. Phys.* 33:423–455. https://doi.org/10.1143/PTP.33.423.

34. Zwanzig, R. 1961. Memory effects in irreversible thermodynamics. *Phys. Rev.* 124:983–992.

35. Ayaz, C., L. Scalfi, …, R. R. Netz. 2022. Generalized Langevin equation with a nonlinear potential of mean force and nonlinear memory

friction from a hybrid projection scheme. *Phys. Rev. E.* 105, 054138. https://doi.org/10.1103/PhysRevE.105.054138.

36. Mizuno, D., C. Tardin, …, F. C. MacKintosh. 2007. Nonequilibrium mechanics of active cytoskeletal networks. *Science.* 315:370–373.

37. Netz, R. R. 2023. Derivation of the non-equilibrium generalized Langevin equation from a generic time-dependent Hamiltonian. Preprint at arXiv. https://doi.org/10.48550/arXiv.2310.00748.

38. Bechinger, C., R. Di Leonardo, …, G. Volpe. 2016. Active particles in complex and crowded environments. *Rev. Mod. Phys.* 88, 045006.

39. Viswanathan, G. M., M. G. Da Luz, …, H. E. Stanley. 2011. The Physics of Foraging: An Introduction to Random Searches and Biological Encounters. Cambridge University Press.

40. Romanczuk, P., M. Bär, …, L. Schimansky-Geier. 2012. Active Brownian particles: From individual to collective stochastic dynamics. *Eur. Phys. J. Spec. Top.* 202:1–162.

41. Ramaswamy, S. 2010. The mechanics and statistics of active matter. *Annu. Rev. Condens. Matter Phys.* 1:323–345.

42. Brockmann, D., L. Hufnagel, and T. Geisel. 2006. The scaling laws of human travel. *Nature.* 439:462–465.

43. Tailleur, J., and M. E. Cates. 2008. Statistical mechanics of interacting run-and-tumble bacteria. *Phys. Rev. Lett.* 100, 218103.

44. Mitterwallner, B. G., L. Lavacchi, and R. R. Netz. 2020. Negative friction memory induces persistent motion. *Eur. Phys. J. E Soft Matter.* 43:67.

45. Jeanneret, R., M. Contino, and M. Polin. 2016. A brief introduction to the model microswimmer Chlamydomonas reinhardtii. *Eur. Phys. J. Spec. Top.* 225:2141–2156. https://doi.org/10.1140/epjst/e2016-60065-3.

46. Nguyen, T. H., M. Kandel, …, G. Popescu. 2017. Halo-free phase contrast microscopy. *Sci. Rep.* 7:44034.

47. Dutcher, S. K., and E. T. O'Toole. 2016. The basal bodies of Chlamydomonas reinhardtii. *Cilia.* 5:18.

48. Wan, K. Y., and R. E. Goldstein. 2016. Coordinated beating of algal flagella is mediated by basal coupling. *Proc. Natl. Acad. Sci. USA.* 113:E2784–E2793.

49. Wakabayashi, K.-i., T. Ide, and R. Kamiya. 2009. Calcium-dependent flagellar motility activation in Chlamydomonas reinhardtii in response to mechanical agitation. *Cell Motil Cytoskeleton.* 66:736–742. https://doi.org/10.1002/cm.20402.

50. Mondal, D., R. Adhikari, and P. Sharma. 2020. Internal friction controls active ciliary oscillations near the instability threshold. *Sci. Adv.* 6, eabb0503. https://doi.org/10.1126/sciadv.abb0503.

51. Mondal, D., A. G. Prabhune, …, P. Sharma. 2021. Strong confinement of active microalgae leads to inversion of vortex flow and enhanced mixing. *Elife.* 10, e67663. https://doi.org/10.7554/eLife.67663.

52. Goldstein, R. E. 2015. Green Algae as Model Organisms for Biological Fluid Dynamics. *Annu. Rev. Fluid Mech.* 47:343–375. https://doi.org/10.1146/annurev-fluid-010313-141426.

53. Kreis, C. T., M. Le Blay, …, O. Bäumchen. 2018. Adhesion of Chlamydomonas microalgae to surfaces is switchable by light. *Nat. Phys.* 14:45–49. https://doi.org/10.1038/nphys4258.

54. Blair, D., and E. Dufresne. 2008. The Matlab Particle Tracking Code Repository. Particle-tracking code. http://site.physics.georgetown.edu/matlab/code.html.

55. Pelleg, D., and A. W. Moore. 2000. X-means: Extending k-means with efficient estimation of the number of clusters. *Icml.* 1:727–734.

56. Bock, H.-H. 2007. Clustering methods: a history of k-means algorithms. *In* Selected Contributions in Data Analysis and Classification, pp. 161–172.

57. Beheshti, S., and M. A. Dahleh. 2005. A new information-theoretic approach to signal denoising and best basis selection. *IEEE Trans. Signal Process.* 53:3613–3624.

58. Shahbaba, M., and S. Beheshti. 2012. Improving x-means clustering with mndl. *In* 2012 11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA) IEEE, pp. 1298–1302.

59. Novikov, A. 2019. PyClustering: Data Mining Library. *J. Open Source Softw.* 4:1230. https://doi.org/10.21105/joss.01230.

60. Celebi, M. E., H. A. Kingravi, and P. A. Vela. 2013. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Syst. Appl.* 40:200–210.

61. Brünig, F. N., O. Geburtig, …, R. R. Netz. 2022. Time-dependent friction effects on vibrational infrared frequencies and line shapes of liquid water. *J. Phys. Chem. B.* 126:1579–1589.

62. Netz, R. R. 2018. Fluctuation-dissipation relation and stationary distribution of an exactly solvable many-particle model for active biomatter far from equilibrium. *J. Chem. Phys.* 148, 185101.

63. Netz, R. R. 2023. Multi-point distribution for Gaussian non-equilibrium non-Markovian observables. https://arxiv.org/abs/2310.08886.

64. Metzler, R., J.-H. Jeon, …, E. Barkai. 2014. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* 16:24128–24164.

65. Friedrich, B. M., and F. Jülicher. 2012. Flagellar synchronization independent of hydrodynamic interactions. *Phys. Rev. Lett.* 109, 138102.

66. Quaranta, G., M.-E. Aubin-Tam, and D. Tam. 2015. Hydrodynamics versus intracellular coupling in the synchronization of eukaryotic flagella. *Phys. Rev. Lett.* 115, 238101.

67. Wang, B. X., C. M. Wu, and K. Ribbeck. 2021. Home, sweet home: how mucus accommodates our microbiota. *FEBS J.* 288:1789–1799.