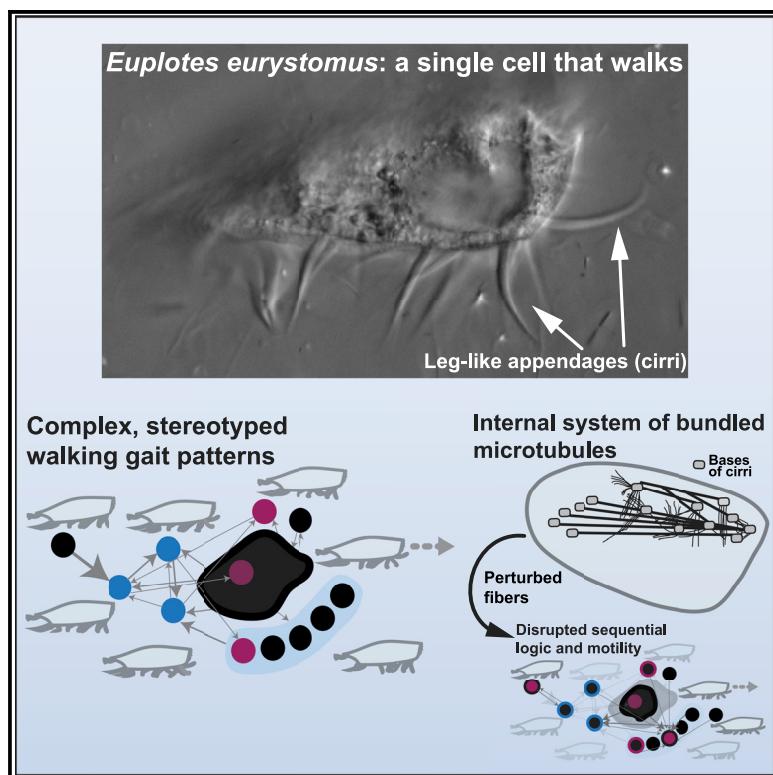


### A unicellular walker controlled by a microtubule-based finite-state machine

#### Graphical abstract



#### Authors

Ben T. Larson, Jack Garbus,  
Jordan B. Pollack, Wallace F. Marshall

#### Correspondence

blarson@berkeley.edu (B.T.L.),  
wallace.ucsf@gmail.com (W.F.M.)

#### In brief

Larson et al. analyze the motility of the unicellular organism *Euplates eurystomus*, which uses leg-like appendages called cirri to walk across surfaces. Walking involves an unusual gait with elements of variability and stereotypy in patterns of cirral activity. Gait coordination is mediated by an internal system of bundled microtubules.

#### Highlights

- *Euplates eurystomus* is a cell that walks with 14 leg-like appendages called cirri
- The walking gait involves complex, partially stereotyped patterns of cirral activity
- Sequential patterns defining gait coordination are necessary for proper motility
- Gait coordination is mediated by an internal system of bundled microtubules



## Article

# A unicellular walker controlled by a microtubule-based finite-state machine

Ben T. Larson,<sup>1,3,\*</sup> Jack Garbus,<sup>2</sup> Jordan B. Pollack,<sup>2</sup> and Wallace F. Marshall<sup>1,4,5,\*</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>2</sup>Computer Science Department, Brandeis University, Waltham, MA 02453, USA

<sup>3</sup>Twitter: @BEuplates

<sup>4</sup>Twitter: @WallaceUCSF

<sup>5</sup>Lead contact

\*Correspondence: blarson@berkeley.edu (B.T.L.), wallace.ucsf@gmail.com (W.F.M.)

<https://doi.org/10.1016/j.cub.2022.07.034>

## SUMMARY

Cells are complex biochemical systems whose behaviors emerge from interactions among myriad molecular components. Computation is often invoked as a general framework for navigating this cellular complexity. However, it is unclear how cells might embody computational processes such that the theories of computation, including finite-state machine models, could be productively applied. Here, we demonstrate finite-state-machine-like processing embodied in cells using the walking behavior of *Euplotes eurystromus*, a ciliate that walks across surfaces using fourteen motile appendages (cirri). We found that cellular walking entails regulated transitions among a discrete set of gait states. The set of observed transitions decomposes into a small group of high-probability, temporally irreversible transitions and a large group of low-probability, time-symmetric transitions, thus revealing stereotypy in the sequential patterns of state transitions. Simulations and experiments suggest that the sequential logic of the gait is functionally important. Taken together, these findings implicate a finite-state-machine-like process. Cirri are connected by microtubule bundles (fibers), and we found that the dynamics of cirri involved in different state transitions are associated with the structure of the fiber system. Perturbative experiments revealed that the fibers mediate gait coordination, suggesting a mechanical basis of gait control.

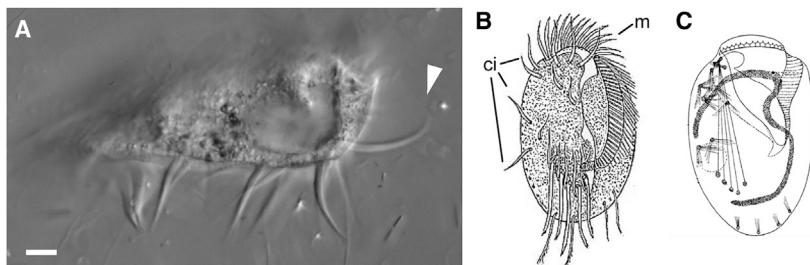
## INTRODUCTION

Cells are complex physical systems controlled by networks of signaling molecules. Single cells can display sophisticated, animal-like behaviors,<sup>1–3</sup> orchestrating active processes far from thermodynamic equilibrium in order to properly carry out biological functions.<sup>4,5</sup> Indeed, single cells can make decisions,<sup>6</sup> execute coordinated, directed movements,<sup>7</sup> solve mazes,<sup>8,9</sup> and learn.<sup>10,11</sup> Such behaviors in animals arise from neural activity, but we know comparatively little about the mechanisms of cellular behavior, which emerge from a combination of chemical reactions,<sup>12</sup> cellular architecture,<sup>3</sup> physical constraints,<sup>13,14</sup> and interactions with the environment.<sup>15</sup> The involvement of information processing in cell state transitions suggest that cellular behavior can be understood as an embodied computation.<sup>16,17</sup> The theory of computation has often been invoked as a general framework for understanding cellular dynamics<sup>16,18,19</sup>—environmental sensing by bacteria being a deeply studied example<sup>18,19</sup>—and has been used to engineer programmable cell states.<sup>20</sup> Ciliates display some of the most striking examples of unicellular behavior, including hunting,<sup>3</sup> sensorimotor navigation,<sup>21</sup> and predator avoidance.<sup>22</sup> Spirotrichous ciliates of the genus *Euplotes* are notable for their complex locomotion,<sup>23–25</sup> using bundles of specialized cilia (cirri) to walk across surfaces<sup>23,24</sup> (Figure 1A; Videos S1 and S2). Depending on the species, these cells

generally have 14 to 15 ventral cirri arranged in a consistent pattern used for walking locomotion.<sup>26</sup> *Euplotes* live in aquatic environments and, in addition to walking, use their cirri for swimming and rapid escape responses<sup>27</sup> (Video S2). Oral membranelles (Figure 1B) are also used for swimming and to generate feeding currents for capturing bacterial and protistan prey. Early 20<sup>th</sup> century protistologists were so impressed by the apparent coordination of cirri that they proposed the existence of a rudimentary nervous system, the neuromotor apparatus, to account for their observations.<sup>25</sup> This theory was motivated by the presence of tubulin-based<sup>28</sup> intracellular fibers emanating from the bases of cirri (Figure 1C).

How can a single cell coordinate a walking gait without a nervous system? Coordination, to the extent that it exists in the gait of *Euplotes*, requires dynamical coupling among cirri or between cirri and some shared external influence. Although the walking movements of *Euplotes* appear superficially similar to those of animals such as insects, the existence of stereotyped sequences of appendage movements that define a gait is unclear. Recently, analytical techniques from statistical physics have been used to understand mesoscale dynamics in biological systems, including cellular behavior,<sup>4,5,31,32</sup> by coarse graining the complexity of biological dynamics into states and analyzing the transitions among states. State representation allows us to ask whether forward and reverse transitions among states are equal,





**Figure 1. *Euplotes* cell architecture**

(A) *Euplotes eurystomus* cell in profile showing ventral cirri, used for walking locomotion (arrowhead indicates a single cirrus). Scale bar, 10  $\mu$ m.

(B) *Euplotes* cell viewed from the ventral surface, highlighting the complex, asymmetric structure. Notable features include the cirri (ci) and the membranellar band (m), wrapping from the top of the cell to the center, which is used to generate a feeding current to draw in prey items. Drawing adapted from Kahl<sup>29</sup> via Wikimedia Commons.

(C) A drawing of a *Euplotes* cell, highlighting the fiber system associated with the cirri. Adapted from Yocom.<sup>30</sup>

See also [Video S1](#).

a condition known as detailed balance.<sup>4,33</sup> Systems that violate detailed balance operate in a non-equilibrium mode, display net probability flows, and can produce directed cycles in state space.<sup>4</sup> Broken detailed balance has been observed in the motility dynamics of cultured mammalian cells as well as the motility dynamics of a freely behaving flagellate protist<sup>5,31</sup> and implies that non-equilibrium models are most applicable to such systems.<sup>32</sup> Identification of broken detailed balance, therefore, highlights dynamical stereotypy in terms of temporal irreversibility and can indicate active control of biological dynamics.

When information processing drives patterns of state transitions, such a system can be analyzed using automata theory, a fundamental level in the theory of computation.<sup>34,35</sup> We hypothesized that walking cells might be governed by finite-state automata with directed, processive movement arising from reproducible, stereotyped patterns of state transitions. We chose to focus on the relatively simple case of spontaneous linear walking, which might require some form of information processing to coordinate the movements of cirri.

Here, we use time-lapse microscopy and quantitative analysis to show that *Euplotes eurystomus* walks with a cyclic stochastic gait displaying broken detailed balance and exhibiting elements of stereotypy and variability, in accord with a finite-state automaton representation.

## RESULTS

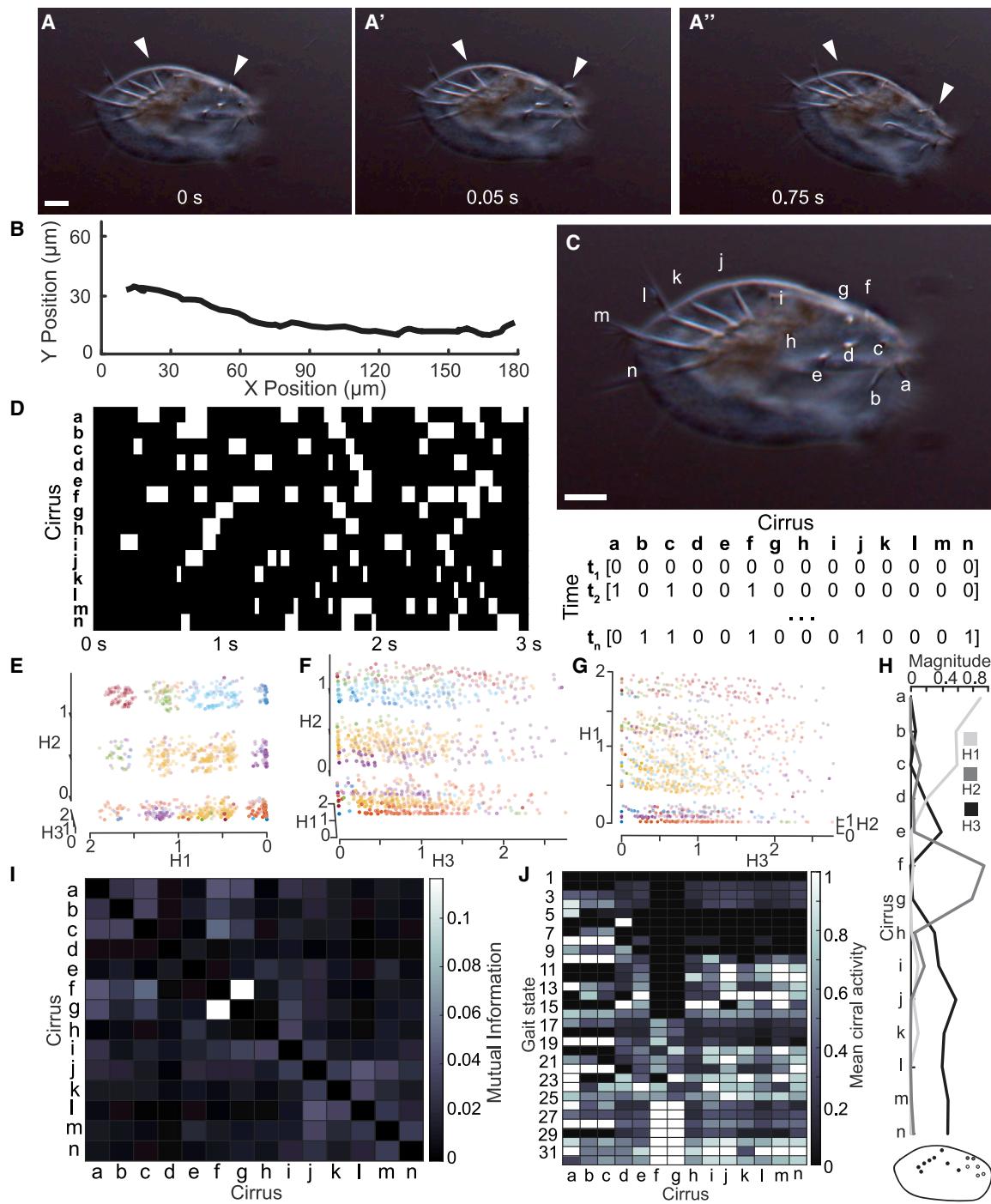
### A reduced state space is sufficient to describe walking dynamics

We analyzed the walking behavior of *Euplotes eurystomus* cells,<sup>27</sup> during uninterrupted, linear walking trajectories (Figures 2A and 2B; [Video S1](#)). Cells were observed by video microscopy at 33 frames/s in a focal plane at the cirrus-coverslip interface to clearly observe cirral dynamics (Figure 2A). The consistency of cirral position across cells allowed us to give each of the 14 cirri an alphabetic label from a–n (Figure 2C). During walking, cirri move in a manner analogous to the recovery stroke-power stroke cycle executed by many eukaryotic flagella, first lifting off the substrate and sweeping close to the cell body before extending in roughly the direction of cell orientation before sweeping downward to reestablish contact with the substrate<sup>24,36</sup> (Figure 2B; [Video S1](#)). In each video frame, walking state was encoded as a 14-element binary vector, with each element corresponding to a cirrus and receiving a value of “0” if the cirrus was in contact with the coverslip and stationary

and a “1” if the cirrus was in motion or had moved in the preceding interval between frames. (Instances of stationary cirri held above the coverslip for a sustained period of time were not observed.) The trajectories of 13 cells were manually tracked and annotated for a total of 2,343 time points. This quantitative analysis revealed stepping-like cirral dynamics: cirri tend to undergo rapid movements followed by longer periods of quiescence (Figure 2D). Cirral dynamics seemed to lack any obvious patterns such as periodicity or repeating sequences (e.g., Figure 2D), implying that the movements are generated either by stochastic processes or complex deterministic mechanisms. This lack of periodicity (confirmed by autocorrelation analysis; [Figure S1](#)) or fixed-phase relationships between appendage movements differs markedly from those of various unicellular organisms and the gaits of most animals.<sup>37,38</sup>

Despite the apparent complexity of cirral dynamics, we wondered whether there might be some underlying pattern. We first sought to obtain a reduced state space that could accurately describe the dynamics, as has proven successful in behavioral analysis of diverse living systems.<sup>31,32,40,41</sup> Because our ultimate goal was to identify motifs among the patterns of cirral activity, which entail strictly non-negative values, we performed dimensionality reduction by non-negative matrix factorization (NMF), a technique used to identify patterns in a wide range of other contexts.<sup>42</sup> NMF revealed the cirral states to be well-described by a 3D space (Figures 2E–2G and S1; see [method details](#) for more details). Factors of the NMF (H1, H2, and H3) constitute the basis of the NMF space and correspond to nonoverlapping, spatially distinct groups of cirri (Figure 2H). These factors indicate features of cirral activity that in various weighted combinations can be used to compactly represent all of the cirral activity measurements. This dimensionality reduction of the gait state space arises in part from shared pairwise mutual information among cirri (Figure 2I). Here, mutual information quantifies the amount of information about the activity of one cirrus gained by measuring the activity of another and acts like a generalized measurement of correlation. Therefore, dimensionality reduction by NMF reflects correlations in cirral activity.

Noting the apparent structure in the NMF space in the form of clusters of points (Figures 2E–2G), we identified gait states by applying the density-based spatial clustering of applications with noise (DBSCAN) algorithm<sup>39</sup> to group the output of NMF into clusters with members of a given cluster sharing similar patterns of cirral activity (see [method details](#) and [Figure S1](#)). Although NMF can itself be used to define clusters, by clustering



**Figure 2. Euplates walking described in a discrete, reduced state space with gait states corresponding to the patterns of cirral activity**

(A) (A-A') Movements of cirri during walking locomotion visualized by brightfield microscopy. Three snapshots depict time points during a walking trajectory; white arrowheads indicate cirri. In the panels from left to right, the cirrus indicated by the left arrowhead is stationary, stationary, and then moving, and the cirrus indicated by the right arrowhead is stationary, moving, and then stationary. Scale bar, 15 μm.

(B) Trajectory of a cell walking across a coverslip.

(C) Scheme for encoding cirral dynamics during walking as a 14-bit binary vector. Each entry is either 0 if the cirrus is not moving and in contact with the coverslip or 1 if the cirrus is moving. Scale bar, 15 μm.

(D) Representative visualization of encoded cirral dynamics for a single trajectory, corresponding to trajectory in (E). White denotes 1 and black 0.

(E-G) Three roughly orthogonal views of cirral dynamics, encoded as in (E) using NMF, from 13 cells over 2,343 time points. Axes correspond to the components of the NMF (H1, H2, and H3), each point is one timepoint. Randomized colors highlight the 32 clusters identified using DBSCAN.<sup>39</sup> These clusters constitute gait states corresponding to unique configurations of cirral activity.

in two steps, i.e., NMF followed by DBSCAN, we could use the visual observation of clustering in the NMF output to confirm performance of the subsequent clustering step. Visual inspection in conjunction with silhouette coefficient analysis (a metric of cluster cohesion and separation) revealed that 32 clusters accurately captured the visible structure in the reduced state space without overfitting (Figures 2E–2G and S1; [method details](#)). These reduced gait states correspond to distinct patterns of cirral activity (Figure 2J).

### Euplotes walks with a cyclic stochastic gait

In order to relate the gait states identified in our cluster analysis to cell motility, we asked how changes in the number of active cirri may relate to cell movement. Naively, one might expect that the force associated with locomotion is roughly proportional to the number of moving appendages. Alternatively, velocity might inversely correlate with the net change in cirral activity, which would be expected if stationary cirri were generating a pushing traction force as in crawling animals. Our data supported neither expectation: cell velocity was only weakly correlated with number of active cirri ( $R^2 = 0.03$ ). Instead, the largest cell velocities corresponded to small-to-moderate changes in the number of active cirri (Figure 3A). We hypothesized that transitions between gait states must be important to driving the forward progression of walking cells and thus looked for active coordination in the observed gait dynamics.

To detect active coordination dictating gait state transitions, we calculated the forward and reverse transition rates between states from the 1,423 pairwise transitions in our dataset as  $N_{ij}/T$ , where  $N_{ij}$  is the total number of transitions observed between states  $i$  and  $j$ , and  $T$  is the total observation time (Figure 3B; see [method details](#)). The transition rate defined here is proportional to the probability current  $J_{ij} = P_i k_{ij}$  where  $P_i$  is the probability of state  $i$  and  $k_{ij}$  is the conditional transition probability analogous to a chemical reaction rate. The presence of strongly unbalanced transitions such as from gait state 3 to 17 versus 17 to 3 suggested broken detailed balance, and indeed, a number of forward and reverse transitions were found to be significantly unbalanced by the binomial test (see [method details](#)). Unbalanced transitions can also arise in equilibrium systems that have not yet reached steady state, in which case transition rates may change in time. To test whether the gait operates at steady state, we checked whether the total number of transitions into each state were balanced by total transitions out of that state:  $\sum_i N_{ij} = \sum_j N_{ji}$ . Consistent with steady-state dynamics, we found that this condition held to within a difference of at most a single transition. Although there are cases where non-equilibrium state transitions can nevertheless be balanced, or appear that way due to limited sampling, the presence of unbalanced transitions existing as parts of loops in state space unambiguously indicated non-equilibrium dynamics (see [method details](#)). To further understand the degree to which detailed balance was broken, or, similarly, the distance from equilibrium, we calculated the entropy production

rate.<sup>5</sup> Following the procedure from Wan and Goldstein,<sup>5</sup> we obtained a lower bound estimate for an entropy production rate of 0.4 nats, similar to the value reported for strongly non-equilibrium gait transitions observed in a flagellate.<sup>5</sup> Walking *Euplotes* cells, therefore, have a non-equilibrium gait, displaying temporal order in sequences of appendage movement, despite a lack of standard gait periodicity.

We next sought to better understand the organization and sequential logic of this unusual gait. First, we noted that gait state transitions appear restrained: only 322 of the 1,024 possible types of transitions were observed to occur at least once, and within this restricted set, only 173 occurred more than once (Figure 3C). We then grouped transitions into two categories: balanced transitions, which satisfy detailed balance, and unbalanced transitions, which do not (see [method details](#)). This partitioning allowed us to separately investigate unbalanced, non-equilibrium-like and balanced, equilibrium-like transitions ([method details](#); Figures 3D and 3E). We found that unbalanced transitions occur at relatively high frequency but involve a small number of states (Figure 3D). Only six of the 32 gait states are associated with unbalanced transitions (Figure 3F), and among these states, three form a directed cycle following  $2 \rightarrow 3 \rightarrow 17 \rightarrow 2$ . We had expected that unbalanced transitions might be associated with a “power stroke” (in the sense of occurring simultaneously with cell movement) but found instead that high cellular velocities tend to be associated with balanced transitions (Figures 3D and 3E) and that relatively few transitions corresponded to substantial cell movement (Figure 3C).

Despite the presence of high frequency unbalanced transitions, the gait of *Euplotes* involves highly variable trajectories through gait state space. The picture of walking trajectories that emerges is one of stochastic excursions from gait state 1, a unique “rest state,” which involves no cirral movement (Figure 2J), into nondeterminate paths through state space involving a mix of balanced and unbalanced transitions. The majority of cell movement occurs during infrequent, equilibrium-like (balanced) transitions leading both toward and away from unbalanced transitions. We note here that although our coarse-graining procedure to identify gait states does not constitute a unique representation of gait structure, we can be confident that our analysis captures structure in gait dynamics and the presence of broken detailed balance in particular. In general, the coarse graining of the state space of a system can obscure broken detailed balance, but the net flux of transitions in the state space of a system should not arise in an illusory fashion based on a coarse-graining procedure or partial observation of a system.<sup>43</sup> Additionally, we have chosen a coarse-graining procedure based on the properties of our particular data and demonstrate its performance on simulated cirral dynamics with varying noise and for the case of unpatterned, random cirral fluctuations (Figure S1; [method details](#)).

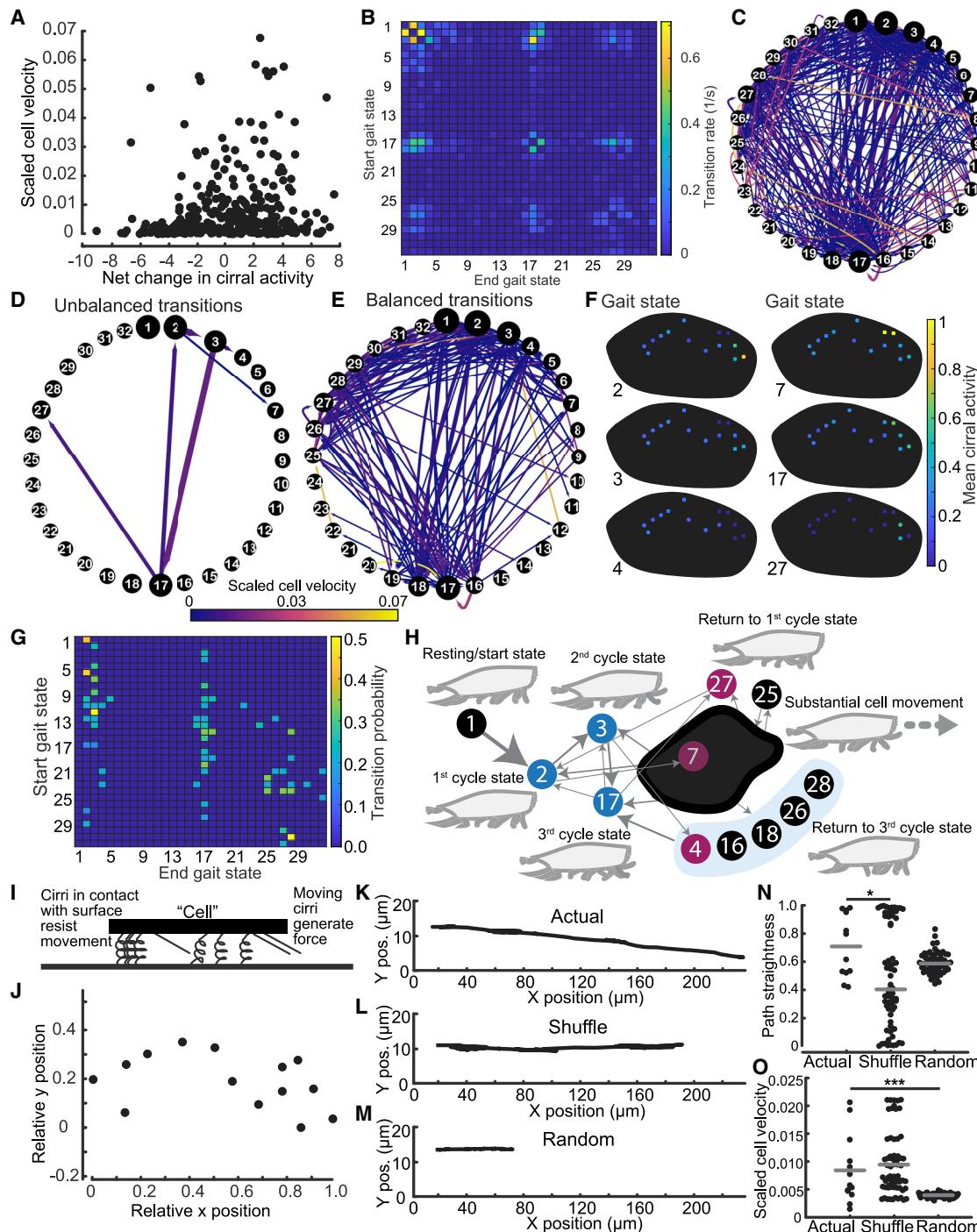
A natural question is whether gait state transitions are a “memoryless” or first-order Markov process, such that transition

(H) Magnitudes associated with each cirrus for distinct components of the NMF. The depiction of a cell, including the positions of cirri, has the same color map as the plot above and shows the grouping of the cirri corresponding to each NMF component.

(I) Heatmap of mutual information between all pairs of cirri showing correlations in cirral activity corresponding to the NMF components displayed in (H).

(J) Heatmap representation of cirral activity associated with each gait state. Values for each cirrus are the mean over all instances of the gait state.

See also Figure S1 and Videos S1 and S2.



**Figure 3. Cyclic stochastic gait exhibiting broken detailed balance, stereotypy, and state-machine-like dynamics**

(A) Mean net change in cirral activity versus the net scaled cell velocity associated with all transitions between the 32 gait states identified in Figure 2 shows that the change in number of active cirri is not strongly correlated with cell velocity ( $R^2 = 0.03$ ).

(B) Transition matrix of gait state transitions, with rows representing the starting state and columns indicating the ending state, exhibits broken detailed balance.

(C) Directed graph representation of gait state transitions. Nodes correspond to the 32 gait states, with node sizes scaled by proportion of total time cells spent in each state. Arrows between nodes (directed edges) signify state transitions. Arrow size is scaled by transition rates as in (B). Edge color represents scaled cell velocity, as in (A), according to the indicated color scale.

(D) A subset of transitions visualized as in (C) shows restricted and relatively high frequency of unbalanced, non-equilibrium-like transitions. Only transitions that were observed to happen more than one time and exhibiting a significant difference between forward and reverse transitions ( $p < 0.05$  by binomial test, see method details) are displayed.

probabilities are determined completely by the present state with previous history contributing no additional predictive information.<sup>44,45</sup> Several analyses (see [method details](#) and [Figure S1](#)) suggested that *Euplates* retains some memory of the prior sequence of cirral movements during locomotion such that the gait may not conform to a continuous or discrete-time first-order Markov process.

Taken together, our analysis revealed a mixture of unbalanced transitions driving cycles and balanced transitions arranged as networks, for which we propose to apply the term “cyclic stochastic gait.” It has been argued that physical systems exhibiting such a mix of stereotypy and variability can be viewed as performing computations<sup>34,46,47</sup> in the sense that the evolution of the system over time is most compactly described as the result of a computational process involving state transitions, memory, and decision rules.

To better understand the sequential logic of the gait, we focused on the dominant structure of gait transitions in terms of transition probabilities ([Figure 3G](#); [method details](#)) allowing us to derive a simplified representation of stereotypy in gait dynamics ([Figure 3H](#)). We found that few states were the recipients of the majority of the highest probability transitions and that many received no high-probability transitions ([Figure 3G](#)). Additionally, we found a “cloud” of states linked by low-probability balanced, equilibrium-like fluctuations. Nearly all of the states receiving high-probability transitions were either the three “cycle states” or else fed cycle states with their highest probability transitions, with the majority feeding gait state 17. Although gait state 1 is not the recipient of any individual highest probability transitions, we identified it as the unique “start” state from which cells initiate walking. Beginning with this start state, cells transition with high probability to gait state 2, also one of the highest frequency transitions and the first state in the  $2 \rightarrow 3 \rightarrow 17 \rightarrow 2$  cycle of unbalanced transitions. The sequence from the start state through the cycle states corresponds to increasing amounts of cirral activity. Although the highest probability transitions from

the third cycle state to any single gait state tend to return to the first or second cycle state with equal probability, cells in fact transition to the equilibrium “cloud” of motility-associated states with overall higher probability. Return to the cycle states tend to occur through various moderately high-probability transitions from the motility state cloud or through a restricted set of intermediate states. In conjunction with this set of transitions, we also noted unbalanced transitions stemming from the cycle states to the motility state as well as the presence of intermediate states from a given cycle state that subsequently feed the next cycle state.

Altogether, the picture of stereotypical gait dynamics that emerges is one of biased transitions involving cycle states preceding relatively low-probability, unbiased transitions associated with substantial cell movement before returning to the start or cycle states and beginning the sequence again. We hypothesized that sequences involving the cycle states serve to establish configurations of cirri necessary for cells to later transition between states from which substantial forward progress of the cell is generated.

To better understand the functional significance of the sequential logic of the gait in driving processive cell movement, we developed a simple model based on a coarse-grained physical picture of cellular walking ([Figure 3I](#)). Briefly, we consider a 2D system where a cell walks using its 14 cirri, which can exist in two states: moving or not moving. We modeled cirri as producing constant force in the direction of cell orientation while moving and resisting displacement by acting as linear springs when not moving. Cell position and orientation is defined in terms of the equilibrium positions of the cirri. For simulations, we used relative positions of cirri taken from stationary cell measurements ([Figure 3J](#)). In each time step of the simulation, we calculated a cell displacement based on the sum of forces due to the cirri and changes in cell orientation from the sum of torques (see [method details](#)). This simple model was sufficient to reproduce the linear trajectories of walking cells when using the actual patterns of cirral

- (E) A subset of transitions, similar to (D), except that only the balanced transitions, lacking a significant difference between forward and reverse transitions ( $p < 0.05$  by binomial test), are displayed, also show a complex and widespread structure, this time of balanced, equilibrium transitions.
  - (F) Examples illustrating the spatial organization of cirral activity corresponding to gait states. These gait states are involved in unbalanced transitions.
  - (G) Heatmap of transition probabilities between states, showing only the most probable transitions from a given state with all others set to zero. In cases where multiple state transitions from a state were tied for the highest probability, all of these transitions are displayed.
  - (H) A representation of functional states and transitions between them highlights the state-machine-like nature of the gait of *Euplates*. Gait states are represented as colored circles with numerical labels. Blue represents states that are both recipients and sources of unbalanced transitions as identified in (D) and constitute the three cycle states. Red represents states that are recipients but not sources of unbalanced transitions as identified in (D). Black corresponds to gait states that are associated only with balanced transitions as in (E). Arrows represent the highest probability transitions between the states, including compound functional states composed of multiple gait states (dark gray blob and blue background) as well as unbalanced transitions with size scaled by their proportional probability compared with all other transitions emanating from the source functional state. Diagrams depict walking cell in profile with cirri in a configuration corresponding to the functional state.
  - (I) Diagram summarizing our model of a walking cell.
  - (J) A plot of the relative average resting cirral-surface contact positions measured from the 13 cells used for gait analysis, which were used for all simulations and define simulated cell position and orientation.
  - (K) Simulated trajectory of walking cell using experimentally recorded cirral activity.
  - (L) Simulated trajectory using experimentally recorded cirral activity but in shuffled order.
  - (M) Simulated trajectory using random patterns of cirral activity with the same average level of cirral activity and the same number of timesteps as in (K) and (L).
  - (N) Path straightness measured from simulations using experimentally measured cirral activity (actual), this activity in a randomly shuffled order (shuffle), and randomly generated cirral activity with the same average cirral activity as actual patterns (random). Asterisk indicates  $p = 0.04$  by Wilcoxon rank-sum test.
  - (O) Scaled cell velocity for the simulations described for (K) predicts that the gait states are responsible for generating cell velocities as indicated by decreased scaled cell velocity for random patterns of cirral activity. Scaled velocities were averaged over each cell trajectory. Three asterisks indicate  $p < 0.001$  by Wilcoxon rank-sum test.
- See also [Figures S1](#) and [S2](#).

activity from walking cells (Figures 3K and S2). When we ran simulations with either the same gait states as those from actual cells, but in a shuffled order or random cirral activity with the same average cirral activity as actual cells, we found that path straightness significantly decreased in the case of shuffled transitions ( $p = 0.04$  by Wilcoxon rank-sum test) and scaled cell velocity significantly decreased in the case of random activity ( $p = 0.003$  by Wilcoxon rank-sum; Figures 3K–3O). Note that in both the cases of shuffled transitions and random cirral activity, gait state transitions satisfy detailed balance (Figure S2). These results suggest that the sequential logic of the gait does indeed matter for producing linear walking.

### The microtubule-based fiber system mediates gait coordination

The complex yet sequentially structured gait patterns in conjunction with our simulation results are consistent with the existence of some form of nontrivial gait coordination. What physical machinery could embody the information processing required to generate the stochastic, cyclic state transitions seen during *Euplotes*' walking? Since the early 1900s, the role of the system of cytoskeletal fibers associated with cirri as conduits of information between cirri during cellular locomotion, supported by microsurgical experiments, has been a dominant yet contentious hypothesized mechanism of gait coordination.<sup>48,49</sup>

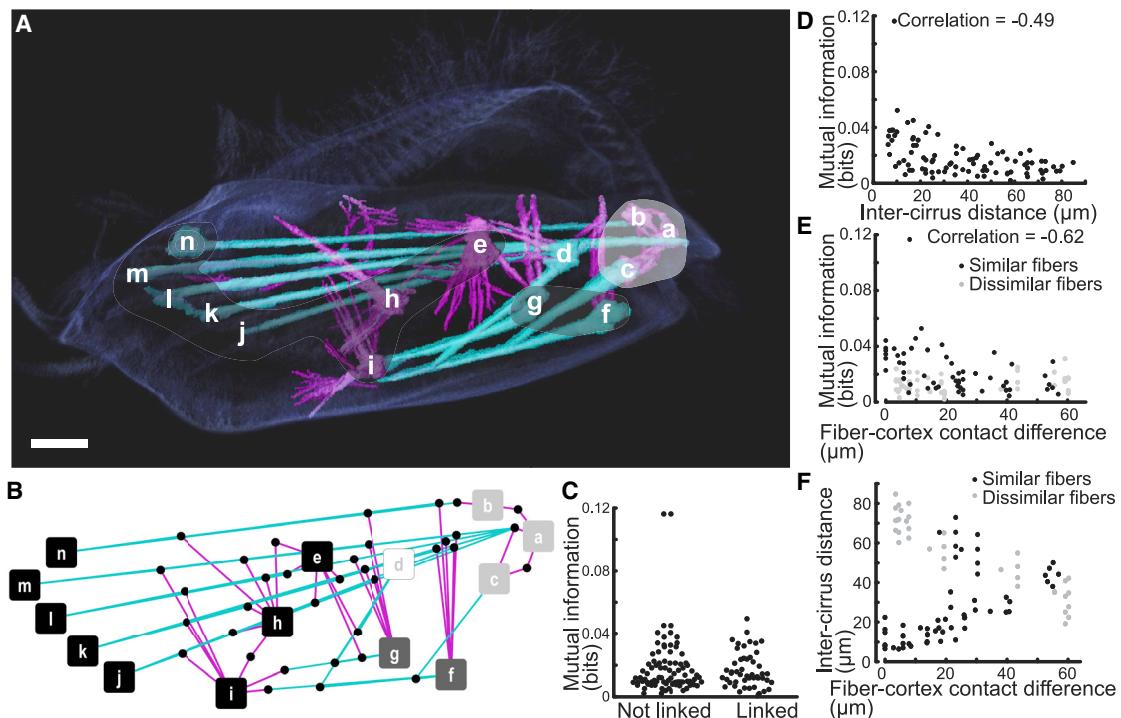
We reconstructed in 3D the microtubule-based fiber system of *Euplotes* associated with cirri and lying just beneath the cell cortex.<sup>25,28</sup> Upon inspection of our confocal reconstructions of SiR-tubulin-labeled cells (Figures 4A and S3), we noted the presence of two morphologically distinct classes of fibers, one thicker and linear and the other thinner, splayed, and less linear, consistent with previous observations<sup>25,28</sup> (Figure 1C). Fibers emanate from the bases of all cirri, appear to intersect one another, and also intersect the cell cortex at various points. Some cirri were found to be associated with only thick fibers, while others have both or only thin fibers. Based on apparent fiber-fiber intersections and convergences, we found that the fiber system forms a continuous network between all cirri (Figures 4A and 4B).

Contrary to the long-standing standing hypothesis from the literature,<sup>48</sup> the functional modules (groups of co-varying cirri) identified in our dynamical analysis were not exclusively linked by dense fiber intersections (Figures 4A and 4B).<sup>25,28,48</sup> In fact, connections between cirri are not generally associated with any statistically significant difference in mutual information compared with unlinked pairs of cirri ( $p = 0.14$  by Wilcoxon rank-sum test; Figure 4C). However, information flow became apparent when fiber-fiber links were grouped by type (i.e., thick to thick fiber, thick to thin fiber, or thin to thin fiber). Under this grouping, we found that pairs of cirri associated with only thick to thick fiber and only thin to thin fiber links have greater mutual information compared with those without links (Figure S3). Interestingly, we found that cirri near one another and connected by fibers to similar regions of the cell cortex shared the most mutual information (Figures 2C, 2G, 4D, and 4E), suggesting that if the fibers play a role in cirral coordination, coupling may also be mediated by mechanisms involving the cirrus and the fiber-cortex interface. Indeed, distances between pairs of cirri and cross nearest neighbor distance between paired sets of cirrus-cortex contact points both show significant Spearman correlations

( $-0.49$ ,  $p < 0.001$  and  $-0.62$ ,  $p < 0.001$ , respectively) to mutual information (Figures 4D and 4E). These correlations indicate that mutual information between pairs of cirri tends to increase with proximity and also tends to increase with similarity between fiber-cortex contact locations. Thus, the cirri with the highest mutual information are those that are close together with similar fiber-cortex connections (Figures 4D–4F).

Together, these observations suggest a mechanism of mechanical coordination in which microtubule bundles allow groups of cirri to influence successive behavior of other groups of cirri. When we perturbed the fiber system using nocodazole, a drug that affects the polymerization of microtubules, we found that treated cells walked along curved rather than straight trajectories (Figures 5A and 5B; Video S3). In contrast to the effect of nocodazole, we found that the microtubule stabilizer paclitaxel caused cells to walk along less convoluted trajectories compared with controls (Figure 5C). Quantifying these effects in terms of a scaled path length (integrated path length divided by the maximum radial distance traversed), we found that nocodazole significantly decreased and that paclitaxel significantly increased the scaled path length of cells compared with controls (Figure 5D). Nocodazole acted in a dose-dependent and reversible manner (Figures 5D and S4). Actin inhibitors had no effect on motility (Figure S4). Next, we checked whether nocodazole treatment had an observable effect on the fiber system. We found that fiber length significantly decreased compared with controls in cells where microtubule polymerization was disrupted by nocodazole (Figures 5E–5G). Further, we were unable to detect the presence of any thin fibers in four out of seven nocodazole-treated cells used for fiber analysis. Of the three cells with detectable thin fibers, we never observed connections between fibers associated with cirri a, b, and c and any other fibers. Additionally, we did not observe any thin fibers making distal cortical contacts. When thin fibers were visible, connections appeared to be the same as those in Figures 4A and 4B.

We then investigated how cirral dynamics were affected when microtubules were perturbed. Following the gait annotation procedure detailed previously, we characterized the walking dynamics of 6 nocodazole-treated cells for a total of 1,133 timepoints. Of those timepoints, 681 corresponded to cirral configurations never observed in untreated cells with a total of 391 new, unique configurations. Projecting these new configurations onto the NMF space we obtained previously, however, revealed that most of the cirral configurations in nocodazole-treated cells were near or within the clusters corresponding to the gait states we obtained from untreated cells (Figure S4). This allowed us to map new cirral configurations onto the gait states (see *method details*). New cirral configurations tended to skew toward more active cirri, and we also noted the presence of a new cluster involving movement in all or nearly all cirri, to which we assigned a new gait state (Figure S4). Mutual information between cirri was higher in general, with many pairs of cirri sharing higher mutual information than the highest values obtained in untreated cells (Figures 5H and S4). This increase in cirral activity and correlations is consistent with the fibers playing a role in conveying inhibitory information during unperturbed walking. In further support of this inhibitory role, we found that paclitaxel treatment also caused an overall increase in mutual information of pairs of cirri (Figure 5H).



**Figure 4. Fiber system structure correlates with dynamical associations between cirri**

(A) SiR-tubulin-labeled cell (faint, dark blue) imaged by confocal microscopy. Two morphologically distinct classes of fibers, one thick and linear and the other thin and curving are indicated by cyan and magenta, respectively (see Figure S3 for more image data). The base of each cirrus is indicated by corresponding letters (as in Figure 2C). Gray shading indicates dynamical groups identified by dimensionality reduction and follows the same color scheme as in Figure 2H. Scale bar, 10  $\mu$ m.

(B) A graph representation of fiber-fiber connections. Nodes correspond to the cirri to which each fiber system is associated, and edges indicate connections between fiber systems. Colors of nodes and colors of edges indicate the same groupings as in (A).

(C) Mutual information between all pairs of cirri grouped by the absence (not linked) or presence (linked) of associated fiber-fiber connections. Statistical significance was evaluated by the Wilcoxon rank-sum test. Note that when pairs of cirri were grouped by fiber-fiber connection type, we did observe a significant increase in mutual information (see Figure S3).

(D) Mutual information as a function of intercirrus distance displays negative correlation, with a Spearman correlation coefficient of  $-0.49$  ( $p < 0.001$ ). Plotted values are defined with respect to pairs of cirri.

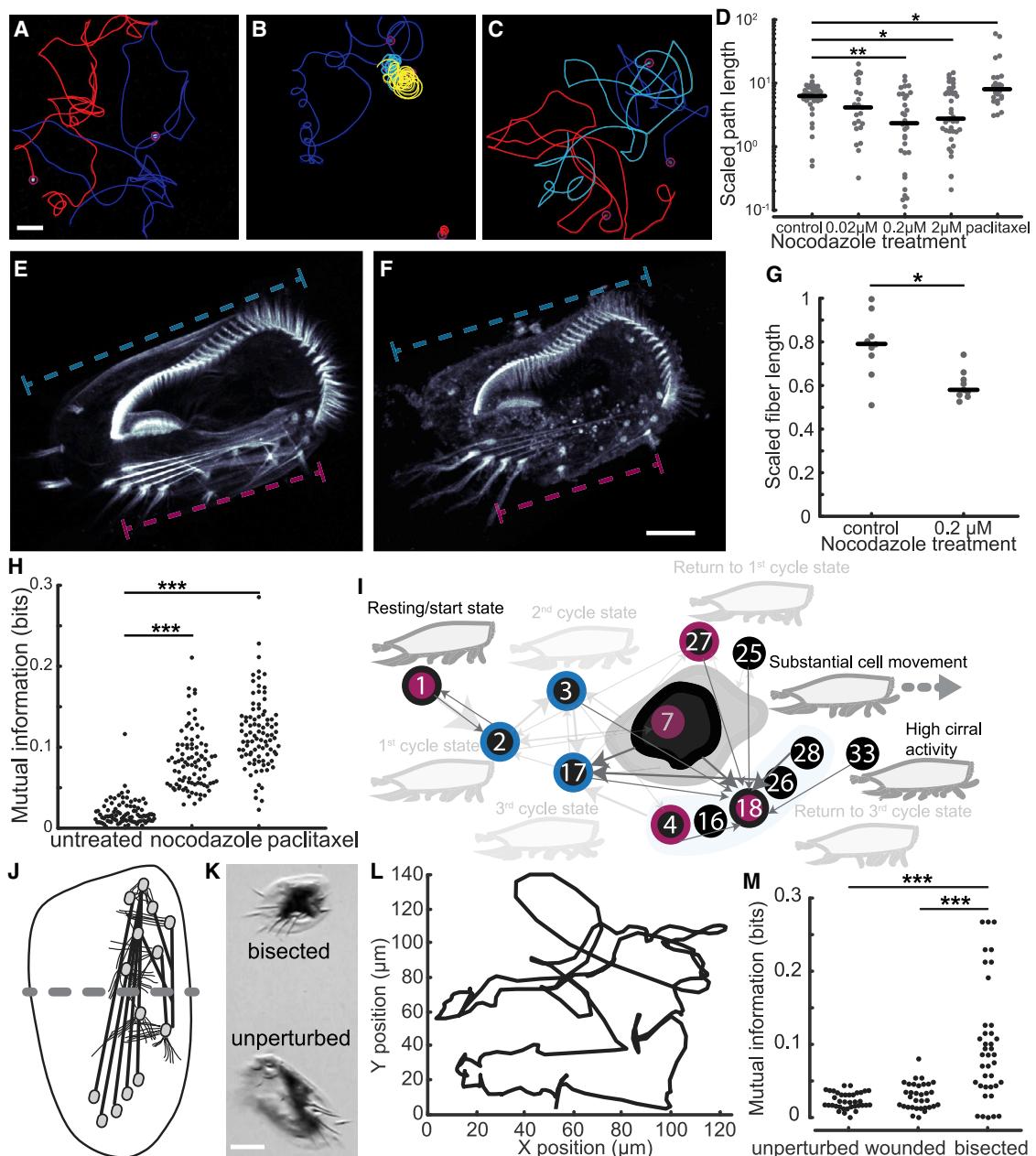
(E) Mutual information as a function of fiber-cortex contact distance grouped by fiber type similarity and lack thereof displays negative correlation, with a Spearman correlation coefficient of  $-0.62$  ( $p < 0.001$ ) for pairs of cirri with similar fiber types and no significant correlation for those with dissimilar fiber types. Fiber-cortex contact difference was measured by the mean cross nearest neighbor distance (see method details) for all fiber-cortex contact points associated with each cirrus.

(F) Fiber-cortex contact difference versus intercirrus difference (as in D and E) illustrating that nearby cirri tend to have similar associated fiber-cortex contacts, highlighting that nearby cirri with similar fiber-cortex contacts share the most mutual information.

Next, we investigated how the dynamics of gait state transitions were affected. Following our previous analysis, we found that the structure of state transitions differed greatly from that of unperturbed cells. Gait state transitions of cells with perturbed fiber systems exhibited less broken detailed balance and closer to equilibrium-like dynamics as indicated by an entropy production rate of 0.1 nats (compared with 0.4 nats in the unperturbed case), including the loss of the unbalanced, cyclic transitions (Figure S4). Figure 5I summarizes the change in the structure of gait state transitions including changes in broken detailed balance, reduction in transitions toward the states previously involved in cyclic transitions, and the reduction in the occupancy of states associated with the cloud of states involved in substantial cell movement, with only 10 of the original 21 cloud states sampled. Many of the new highest probability transitions feed gait state 18 (Figures 5I and S4), which involves nearly exclusive

activity in cirri f and g located close to one another at the edge of the cell. A persistent bias toward activation of these cirri, farthest from the central axis of the cell, may in part explain the increase in turning trajectories. These results, together with the fact that cell velocities were indistinguishable from control conditions (Figure S4) are consistent with predictions from simulations in which shuffling the order of gait state transitions reduced procressivity without affecting speed (Figures 3N and 3O).

The increase in mutual information between cirri under disrupted gait coordination is due to more frequent joint activation or synchronous movements among cirri and thus stems from the reduced complexity of gait dynamics. For a complementary perspective linking these changes in gait dynamics to an underlying computational process, we applied the Causal State Splitting Reconstruction (CSSR) algorithm to construct  $\epsilon$ -machines corresponding to walking cells<sup>34,50</sup>;  $\epsilon$ -machines are automaton



**Figure 5.** The microtubule fiber system of *Euplotes* mediates gait coordination

(A–C) Representative cell motility trajectories in control (A), nocodazole-treated (B), and paclitaxel-treated (C) cells highlight the curved and curled trajectories of cells treated with the microtubule polymerization inhibitor nocodazole and smoother, less convoluted trajectories in cells treated with the microtubule stabilizer paclitaxel. Different colors represent the trajectories of different cells. Scale bar, 500  $\mu$ m.

(D) Nocodazole affects motility in a dose-dependent and specific manner. Scaled path length decreased with increased nocodazole concentration. Paclitaxel, which stabilizes microtubules, increased scaled path length compared with the control. Trajectories of at least 20 cells were analyzed for each condition. The black bars are median values. A single asterisk indicates  $p < 0.05$ , and a double asterisk indicates  $p < 0.005$ , as computed by Wilcoxon rank-sum test.

(E and F) Representative images illustrating effect of nocodazole on the fiber system. Images are maximum intensity projections of confocal Z stack images of cells labeled by SiR-tubulin. Scale bar, 20  $\mu$ m.

(H) A plot of mutual information of all pairs of cirri shows that nocodazole and paclitaxel change the distribution of mutual information compared with untreated cells. The data for untreated cells are from [Figures 2](#) and [3](#). Nocodazole treatment was  $0.2 \mu\text{M}$ , and 6 cells over 1,133 time points were analyzed. Paclitaxel treatment was  $0.02 \mu\text{M}$ , and 6 cells over 441 time points were analyzed. Three asterisks indicate  $p < 0.001$  as computed by two-sample Kolmogorov-Smirnov test.

(legend continued on next page)

models consisting of a set of causal states with transitions between them and represent the minimal model consistent with accurate prediction of a stochastic process.<sup>34</sup> The causal states of an  $\varepsilon$ -machine indicate how the process from which it is constructed stores information, and state transitions indicate how the process transforms information.<sup>34</sup> We found that  $\varepsilon$ -machines constructed from the cirral activity of untreated cells tended to be similar to one another and were more complex in terms of having more causal states and transitions than those constructed from nocodazole-treated cells, which also tended to be similar to one another (Figure S5; method details). This reduction in complexity may reflect reduced computational capacity of the nocodazole-treated cells.

As an additional test of the role of the fibers in mediating gait coordination, we revisited historical microdissection experiments. Although Taylor's 1921 microdissection experiments indicated a role of the fiber system gait coordination,<sup>48</sup> a subsequent study failed to observe disrupted coordination among cirri upon bisection of cells.<sup>49</sup> Importantly, neither of these reports involved analysis of walking behavior. We performed microdissections on cells using pulled quartz microneedles, severing cells transversely just in front of cirrus h, ensuring that we had severed all fibers associated with cirri j–n (Figures 5J and 5K). Similar to previous reports, we found that cell fragments regained spontaneous cirral activity after a brief recovery period. After 24 h, we found that anterior portions of cells began to exhibit spontaneous walking activity (Figures 5K and 5L; Video S4), which persisted for up to 72 h. Although these fragments could walk, they followed circular trajectories reminiscent of nocodazole-treated cells but in the reverse direction (Figure 5L; Video S4). We found significantly higher mutual information shared between cirri indicating more synchronous cirral activity compared with merely wounded cells, which displayed apparently normal motility (Video S5) or unperturbed cells ( $p < 0.001$  in both cases by Wilcoxon rank-sum test), which displayed values consistent with one another (Figure 5M). As with the microtubule inhibitor experiments, these results are consistent with the fibers playing an inhibitory role in gait coordination.

## DISCUSSION

Traditionally, studies of computational processes in cells have focused on combinatorial logic, in which a molecular network generates an output that depends only on the current input.<sup>16,18,19</sup> We have focused on sequential logic, in which outputs depend on the system state as well.<sup>51,52</sup> Automata theory, using finite-state

machine models based on sequential logic, provides tools for understanding structure and stereotypy in transitions between dynamical states, which are increasingly appreciated as features of the behavior of eukaryotic cells. Related approaches for the coarse graining of complex dynamics have revealed simplicity and stereotypy in the behavior patterns of various organisms.<sup>3,5,40,41,53</sup> Although there are examples of locomotor coordination reminiscent of the stochastic, non-equilibrium gait dynamics of *Euplotes* in other cells and animals,<sup>5,31,54</sup> most appendage-based locomotor systems employ stereotyped, determinate patterns of activity.<sup>7,37,38</sup> In the run-and-tumble motility in *Escherichia coli*<sup>55</sup> or analogous behaviors observed in protists,<sup>56,57</sup> motility can be described by equilibrium processes,<sup>5</sup> in contrast to the non-equilibrium character of the gait of *Euplotes*.

We propose that in *Euplotes*, biased, actively controlled cyclic transitions store stress in certain cirri, and the spontaneous release of these cirri from the substrate, during a series of unbiased gait state transitions, allows the cell to move forward. Return to the cycle states reset this process by winding up the system for continued, proper cell movement. Disruption in this resetting may lead to defects in walking as shown by simulations (Figures 3K–3O and 5) and experiments (Figures 5B, 5D, and 5I; Video S3). Our proposed mechanism is consistent with previous observations of cyclic velocity fluctuations in the trajectories of walking *Euplotes*.<sup>23</sup>

The results of experiments perturbing the tubulin-based cytoskeletal fiber system are consistent with its role in mechanically mediating communication both among cirri and between cirri and the cell cortex (Figures 4 and 5). We conjecture that movement of cirri relative to one another can establish tension in the fiber system and that the tension state of fibers associated with each cirrus may then modulate cirral activity in a manner reminiscent of basal coupling in flagellates.<sup>13</sup> Microtubules can respond directly to mechanical forces inside cells<sup>58</sup> and may be involved in more complex signal-transduction pathways.<sup>59,60</sup> Our results show that perturbation of the microtubule fiber system shift the gait of *Euplotes* from a regime of asynchronous yet coordinated movements to a dysregulated regime with synchronous yet improperly coordinated movements. Our work lays a foundation for studying sensorimotor behavior in *Euplotes*, which will shed light on principles of cellular behavior. Because biological function often emerges from the productive management of stochastic fluctuations, we expect that our conceptual and analytical approaches may apply to the behavior of other living systems that produce directed or sequential behavior from random molecular processes.

(I) A representation of the changes in the nature and organization of functional states as well as transitions between them highlights the effects of nocodazole treatment. This panel is partially adapted from Figure 3H and was produced by following the same analysis procedure. Any functional state and transition depicted in Figure 3H no longer observed under nocodazole treatment appear faded. Dark arrows represent new highest probability transitions associated with the states. Outer colors of nodes denote the type or absence of broken detailed balance associated with the gait state for untreated cells, while the core color represents that for the treated cells. Node colors follow those in Figure 3H. The reduction in size of the black blob indicates a reduction in gait states associated with this functional state.

(J) Diagram illustrating the locations of microsurgical cuts (dashed gray line).

(K) Representative image showing a recovered bisected cell and a fully intact, unperturbed cell for reference. Scale bar, 50  $\mu\text{m}$ .

(L) Manually tracked trajectory of a walking, bisected cell illustrating unsteady, curving cell movement.

(M) Mutual information of all pairs of cirri h–n for unperturbed, wounded, and bisected cells. Three asterisks indicate  $p < 0.001$  from two-sample Kolmogorov–Smirnov tests. Data for unperturbed cells are from Figure 2; data for wounded and bisected conditions involved 10 cells over 1,308 time points in the case of the wounded cells and 1,815 time points for bisected cells.

See also Figures S4 and S5 and Videos S3, S4, and S5.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell lines
  - Cell husbandry
- **METHOD DETAILS**
  - Live cell brightfield microscopy
  - Quantification of walking dynamics
  - Dimensionality reduction
  - Clustering
  - State transition analysis
  - Biophysical model and simulations
  - Confocal microscopy
  - Fiber reconstruction and analysis
  - Drug treatment experiments and analysis
  - $\epsilon$ -machine construction
  - Microsurgery experiments
- **QUANTIFICATIONS AND STATISTICAL ANALYSIS**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.07.034>.

### ACKNOWLEDGMENTS

We would like to thank Nicole King for the use of the confocal microscope and Adair Oesterle for assistance with microsurgery. We would like to thank the students and faculty of the 2016 Marine Biological Laboratory Physiology Course and current and former members of the Marshall Lab and King Lab for comments, critiques, and encouragement during the development of this project. We would also like to thank David Booth, Greysen Lewis, Dennis Bray, and members of the Fourmentin-Guilbert Scientific Foundation for comments on the manuscript. This work was funded by the I2CELL Seed Award of the Fourmentin-Guilbert Scientific Foundation (W.F.M. and J.B.P.). Additional funding was provided by NIH grant R35 GM130327 (W.F.M.), NSF grant MCB-2012647 (W.F.M.), and a Merck Fellowship of the Jane Coffin Childs Memorial Fund for Medical Research (B.T.L.).

### AUTHOR CONTRIBUTIONS

B.T.L., J.D.P., and W.F.M. conceived the study. B.T.L., J.G., and W.F.M. developed methodology. B.T.L. performed the experiments, and all authors analyzed and interpreted the data. B.T.L. and W.F.M. originally drafted the manuscript, and all authors reviewed and edited the work.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 25, 2022

Revised: May 20, 2022

Accepted: July 14, 2022

Published: August 12, 2022

### REFERENCES

1. Jennings, H.S. (1906). *Behavior of the Lower Organisms* (Columbia University Press).
2. Reid, C.R., MacDonald, H., Mann, R.P., Marshall, J.A.R., Latty, T., and Garnier, S. (2016). Decision-making without a brain: how an amoeboid organism solves the two-armed bandit. *J. R. Soc. Interface* 13, 20160030.
3. Coyle, S.M., Flbaum, E.M., Li, H., Krishnamurthy, D., and Prakash, M. (2019). Coupled active systems encode an emergent hunting behavior in the unicellular predator *Lacrymaria olor*. *Curr. Biol.* 29, 3838–3850.e3.
4. Battle, C., Broedersz, C.P., Fakhri, N., Geyer, V.F., Howard, J., Schmidt, C.F., and MacKintosh, F.C. (2016). Broken detailed balance at mesoscopic scales in active biological systems. *Science* 352, 604–607.
5. Wan, K.Y., and Goldstein, R.E. (2018). Time irreversibility and criticality in the motility of a flagellate microorganism. *Phys. Rev. Lett.* 121, 058103.
6. Woznicka, A., Cantley, A.M., Beemelmanns, C., Freinkman, E., Clardy, J., and King, N. (2016). Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. USA* 113, 7894–7899.
7. Wan, K.Y. (2018). Coordination of eukaryotic cilia and flagella. *Essays Biochem.* 62, 829–838.
8. Nakagaki, T., Yamada, H., and Tóth, A. (2000). Maze-solving by an amoeboid organism. *Nature* 407, 470.
9. Tweedy, L., Thomason, P.A., Paschke, P.I., Martin, K., Machesky, L.M., Zagnoni, M., and Insall, R.H. (2020). Seeing around corners: cells solve mazes and respond at a distance using attractant breakdown. *Science* 369,
10. Applewhite, P.B. (1979). Learning in Protozoa. In *Biochemistry and Physiology of Protozoa* (Elsevier), pp. 341–355.
11. Jennings, H.S. (1902). Studies on reactions to stimuli in unicellular organisms. IX—on the behavior of fixed infusoria (*Stentor* and *Vorticella*), with special reference to the modifiability of protozoan reactions. *Am. J. Physiol. Content* 8, 23–60.
12. Greenfield, D., McEvoy, A.L., Shroff, H., Crooks, G.E., Wingreen, N.S., Betzig, E., and Liphardt, J. (2009). Self-organization of the *Escherichia coli* chemotaxis network imaged with super-resolution light microscopy. *PLoS Biol.* 7, e1000137.
13. Wan, K.Y., and Goldstein, R.E. (2016). Coordinated beating of algal flagella is mediated by basal coupling. *Proc. Natl. Acad. Sci. USA* 113, E2784–E2793.
14. Höfer, T., Sherratt, J.A., and Maini, P.K. (1995). *Dictyostelium discoideum*: cellular self-organization in an excitable biological medium. *Proc. Biol. Sci.* 259, 249–257.
15. Kantsler, V., Dunkel, J., Polin, M., and Goldstein, R.E. (2013). Ciliary contact interactions dominate surface scattering of swimming eukaryotes. *Proc. Natl. Acad. Sci. USA* 110, 1187–1192.
16. Bray, D. (2009). *Wetware* (Yale University Press).
17. Pfeifer, R., Lungarella, M., and Iida, F. (2007). Self-organization, embodiment, and biologically inspired robotics. *Science* 318, 1088–1093.
18. Mehta, P., and Schwab, D.J. (2012). Energetic costs of cellular computation. *Proc. Natl. Acad. Sci. USA* 109, 17978–17982.
19. Berg, H.C., and Purcell, E.M. (1977). Physics of chemoreception. *Biophys. J.* 20, 193–219.
20. Andrews, L.B., Nielsen, A.A.K., and Voigt, C.A. (2018). Cellular checkpoint control using programmable sequential logic. *Science* 361, eaap8987.
21. Machemer, H. (2001). The swimming cell and its world: structures and mechanisms of orientation in protists. *Eur. J. Protistol.* 37, 3–14.
22. Gilbert, J.J. (1994). Jumping behavior in the oligotrich ciliates *Strobilidium velox* and *Halteria grandinella*, and its significance as a defense against rotifer predators. *Microb. Ecol.* 27, 189–200.
23. Lueken, W., Ricci, N., and Krüppel, T. (1996). Rhythmic spontaneous depolarizations determine a slow-and-fast rhythm in walking of the marine hypotrich *Euplates vannus*. *Eur. J. Protistol.* 32, 47–54.

24. Erra, F., Iervasi, A., Ricci, N., and Banchetti, R. (2001). Movement of the cirri during the creeping of *Euplotes crassus* (Ciliata, Hypotrichida). *Can. J. Zool.* **79**, 1353–1362.
25. Taylor, C.V. (1919). The neuromotor system of *Euplotes*. *Exp. Biol. Med.* **16**, 101–103.
26. Syberg-Olsen, M.J., Irwin, N.A.T., Vannini, C., Erra, F., Di Giuseppe, G., Boscaro, V., and Keeling, P.J. (2016). Biogeography and character evolution of the ciliate genus *Euplotes* (Spirotrichea, Euplotia), with description of *Euplotes curdsi* sp. nov. *PLoS One* **11**, e0165442.
27. Ricci, N., Giannetti, R., and Miceli, C. (1988). The ethogram of *Euplotes crassus* (ciliata, hypotrichida): I. The wild type. *Eur. J. Protistol.* **23**, 129–140.
28. Fleury, A. (1991). Dynamics of the cytoskeleton during morphogenesis in the ciliate *Euplotes* I. Basal bodies related microtubular system. *Eur. J. Protistol.* **27**, 99–114.
29. Kahl, A. (1935). Urtiere oder Protozoa (Fischer).
30. Yocom, H.B. (1917). The Neuromotor Apparatus of *Euplotes Patella* (University of California).
31. Kimmel, J.C., Chang, A.Y., Brack, A.S., and Marshall, W.F. (2018). Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance. *PLoS Comput. Biol.* **14**, e1005927.
32. Chang, A.Y., and Marshall, W.F. (2019). Dynamics of living cells in a cytomorphological state space. *Proc. Natl. Acad. Sci. USA* **116**, 21556–21562.
33. Boltzmann, L. (1872). Weitere Studien über das warmegleich-gewich unter gasmolekülen. *Sitzber. Akad. Wiss. Wein* **II** *66*, 275–370.
34. Shalizi, C.R., and Crutchfield, J.P. (2001). Computational mechanics: pattern and prediction, structure and simplicity. *J. Stat. Phys.* **104**, 817–879.
35. Crutchfield, J.P. (1994). The calculi of emergence: computation, dynamics and induction. *Phys. D Nonlinear Phenom.* **75**, 11–54.
36. Mogami, Y., Pernberg, J., and Machemer, H. (1992). Ciliary beating in three dimensions: steps of a quantitative description. *J. Math. Biol.* **30**, 215–249.
37. Collins, J.J., and Stewart, I.N. (1993). Coupled nonlinear oscillators and the symmetries of animal gaits. *J. Nonlinear Sci.* **3**, 349–392.
38. McNeill Alexander, R. (2003). Principles of Animal Locomotion (Princeton University Press).
39. Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (AAAI Press).
40. Stephens, G.J., Johnson-Kerner, B., Bialek, W., and Ryu, W.S. (2008). Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput. Biol.* **4**, e1000028.
41. Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672.
42. Devarajan, K. (2008). Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS Comput. Biol.* **4**, e1000029.
43. Esposito, M. (2012). Stochastic thermodynamics under coarse graining. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **85**, 041125.
44. Markov, A.A. (1907). Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain. Notes of the Imperial Academy of Sciences of St. Petersburg VIII Series, Physio-Mathematical College, 22 (John Wiley and Sons).
45. Howard, R.A. (1971). Dynamic and Probabilistic Systems, Volume 1 (Dover Publications), Markov models.
46. Crutchfield, J.P., and Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Lett.* **63**, 105–108.
47. Moore, C. (1998). Dynamical recognizers: real-time language recognition by analog computers. *Theor. Comput. Sci.* **201**, 99–136.
48. Taylor, C.V. (1921). Demonstration of the function of the neuromotor apparatus in euplates by the method of microdissection. *Sci. Nat.* **9**, 910.
49. Okajima, A., and Kinoshita, H. (1966). Ciliary activity and coordination in *Euplotes eurystomus*-I. Effect of microdissection of neuromotor fibres. *Comp. Biochem. Physiol.* **19**, 115–131.
50. Shalizi, C.R., Shalizi, K.L., and Crutchfield, J.P. (2002). An algorithm for pattern discovery in time series. ArXiv. <https://doi.org/10.48550/arXiv.cs/0210025>.
51. Bar-Ziv, R., Tlusty, T., and Libchaber, A. (2002). Protein-DNA computation by stochastic assembly cascade. *Proc. Natl. Acad. Sci. USA* **99**, 11589–11592.
52. Dexter, J.P., Prabakaran, S., and Gunawardena, J. (2019). A complex hierarchy of avoidance behaviors in a single-cell eukaryote. *Curr. Biol.* **29**, 4323–4329.e2.
53. Berman, G.J., Bialek, W., and Shaevitz, J.W. (2016). Predictability and hierarchy in *Drosophila* behavior. *Proc. Natl. Acad. Sci. USA* **113**, 11943–11948.
54. Moore, T.Y., Cooper, K.L., Biewener, A.A., and Vasudevan, R. (2017). Unpredictability of escape trajectory explains predator evasion ability and microhabitat preference of desert rodents. *Nat. Commun.* **8**, 440.
55. Berg, H.C., and Brown, D.A. (1972). Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* **239**, 500–504.
56. Polin, M., Tuval, I., Drescher, K., Golub, J.P., and Goldstein, R.E. (2009). *Chlamydomonas* swims with two “gears” in a eukaryotic version of run-and-tumble locomotion. *Science* **325**, 487–490.
57. Almagor, M., Ron, A., and Bar-Tana, J. (1981). Chemotaxis in *Tetrahymena thermophila*. *Cell Motil.* **1**, 261–268.
58. Li, Y., Kučera, O., Cuvelier, D., Rutkowski, D.M., Deygas, M., Rai, D., Pavlović, T., Vicente, F.N., Piel, M., Giannone, G., et al. (2022). Compressive forces stabilise microtubules in living cells. bioRxiv. 2022. 02.07.479347. <https://doi.org/10.1101/2022.02.07.479347>.
59. Maniotis, A.J., Chen, C.S., and Ingber, D.E. (1997). Demonstration of mechanical connections between integrins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure. *Proc. Natl. Acad. Sci. USA* **94**, 849–854.
60. Craddock, T.J.A., and Tuszyński, J.A. (2010). A critical assessment of the information processing capabilities of neuronal microtubules using coherent excitations. *J. Biol. Phys.* **36**, 53–70.
61. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682.
62. Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830.
63. Darmon, D. (2020). transCSSR (GitHub Repos).
64. King, N., Young, S.L., Abedin, M., Carr, M., and Leadbeater, B.S.C.C. (2009). Starting and maintaining *Monosiga brevicollis* cultures. *Cold Spring Harb. Protoc.* **2009**. pdb.prot5148.
65. Tinevez, J.Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Elceiri, K.W. (2017). TrackMate: an open and extensible platform for single-particle tracking. *Methods* **115**, 80–90.
66. Kanagal, B., and Sindhwan, V. (2010). Rank selection in low-rank matrix approximations: a study of cross-validation for NMFs. In Advances in Neural Information Processing Systems (MIT Press).
67. Kawamoto, T., and Kabashima, Y. (2017). Cross-validation estimate of the number of clusters in a network. *Sci. Rep.* **7**, 1.
68. Kolmogoroff, A. (1936). Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112**, 155–160.
69. Sánchez, R., Newman, D.E., and Carreras, B.A. (2002). Waiting-time statistics of self-organized-criticality systems. *Phys. Rev. Lett.* **88**, 068302.
70. Billingsley, P. (1961). Statistical Methods in Markov Chains (Institute of Mathematical Statistics).

71. Hiscott, R.N. (1981). Chi-square tests for Markov chain analysis. *Mathematical Geology* 13, 69–80.
72. Purcell, E.M. (1977). Life at low Reynolds number. *Am. J. Phys.* 45, 3–11.
73. Happel, J., and Brenner, H. (1981). Low Reynolds Number Hydrodynamics, Volume 1 (Springer).
74. Cardona, A., Saalfeld, S., Schindelin, J., Arganda-Carreras, I., Preibisch, S., Longair, M., Tomancak, P., Hartenstein, V., and Douglas, R.J. (2012). TrakEM2 software for neural circuit reconstruction. *PLoS One* 7, e38011.
75. Thévenaz, P., Ruttimann, U.E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7, 27–41.
76. Kingma, D.P., and Welling, M. (2013). Auto-encoding variational Bayes. arXiv. <https://doi.org/10.48550/arXiv.1312.6114>.
77. Kingma, D.P., and Ba, L.J. (2015). Adam: a method for stochastic optimization. arXiv. <https://doi.org/10.48550/arXiv.1412.6980>.
78. Bach, F.R., and Jordan, M.I. (2003). Learning spectral clustering. *Adv. Neural Inf. Process. Syst.* 16, 305–312.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Nocodazole	Sigma-Aldrich	Cat#M1404
Paclitaxel	Sigma-Aldrich	Cat#T7191
Latrunculin B	Thomas Scientific	Cat#C834E37
Cytochalasin B	Fisher Scientific	Cat#1493-96-2
Jasplakinolide	Fisher Scientific	Cat#42-012-750UG
SiR-Tubulin	Cytoskeleton, Inc.	Cat#CY-SC002
Cereal Grass Medium	Thermo Fisher Scientific	Cat#S25242
Experimental models: Cell lines		
<i>Euplotes eurystromus</i>	Carolina Biological Supply	Cat#131480
Software and algorithms		
FIJI	Schindelin et al. <sup>61</sup>	<a href="https://imagej.net/software/fiji/downloads">https://imagej.net/software/fiji/downloads</a>
MATLAB	Mathworks	N/A
Code for simulating the biophysical model of a walking cell	This paper	<a href="https://github.com/btlarson/EuplotesWalkSim">https://github.com/btlarson/EuplotesWalkSim</a>
scikit-learn	Pedregosa et al. <sup>62</sup>	<a href="https://scikit-learn.org/stable/">https://scikit-learn.org/stable/</a>
CSSR algorithm	Darmon <sup>63</sup>	<a href="https://github.com/ddarmon/transCSSR">https://github.com/ddarmon/transCSSR</a>

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests for resources, data, and code should be directed to and will be fulfilled by the lead contact, Wallace F. Marshall ([wallace.ucsf@gmail.com](mailto:wallace.ucsf@gmail.com)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

All data reported in this paper will be shared by the [lead contact](#) upon request.

All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

## Cell lines

Cultures of *Euplotes eurystromus* were obtained from Carolina Biological Supply Company (Item #131480) and were kept at room temperature under ambient light conditions.

## Cell husbandry

Individual cells were isolated from cultures, which contained other protists and meiofauna, by pipetting and placed in non-treated 6-well plates (Thermo Fischer Scientific 08-772-49) containing spring water taken from cultures. Cells were kept in wells for no longer than five days before imaging, and if cells were to be kept for longer than 48 hours, wells containing cells were supplemented with 1% Cereal Grass Medium<sup>64</sup> (Thermo Fisher Scientific S25242) to prevent depletion of prey bacteria and otherwise maintain *Euplotes* under constant growth conditions.

## METHOD DETAILS

## Live cell brightfield microscopy

Cells were concentrated by centrifugation (500 × g for 5 min) and resuspended either in 0.5 mL of spring water in coverglass bottomed FluoroDishes (World Precision Instruments FD35-100) or in 0.2 mL spring water on a coverslip (FisherScientific, 12-545-D) for

imaging. No more than three cells were kept in 0.5 mL imaging samples and only one cell was ever kept in 0.2 mL imaging samples in order to minimize cell-cell interactions. Cells were observed to exhibit spontaneous walking activity on coverglass. Walking cells in FluoroDishes were imaged under brightfield illumination using a Zeiss Z.1 Observer and Hamamatsu Orca Flash 4.0 V2 CMOS camera (C11440-22CU) with a 20x, 0.8 NA Plan-Apochromat (Zeiss) objective. Cells on coverslips were imaged under brightfield illumination with coverslips inverted over a well containing a small amount of distilled water to reduce evaporation using a Zeiss Axio Zoom.V16 and a PCO pco.dimax S1 camera. Importantly, in both imaging systems, the focal plane was set at the interface between cirri of walking cells and the glass surface upon which they were walking. Images were acquired at 0.033 seconds per frame with a 0.005 second exposure in order to capture all cirral dynamics during walking with minimal blur.

### Quantification of walking dynamics

Movies of walking cells were viewed using FIJI.<sup>61</sup> Movement of cirri, or lack thereof was clearly visible in each movie frame (see [Figure 2A](#); [Video S1](#)). The dynamical state of each cirrus in each movie frame was manually annotated. For each frame, each cirrus received a label of “1” if the cirrus was in motion and “0” if the cirrus was not moving and in contact with the coverslip. Motion of cirri was evident in terms of a change in cirrus shape or tip position often in addition to blur due to motion during image acquisition or position out of the focal plane (see [Figure 2A](#); [Video S1](#)). While only slowly walking cells were recorded, sometimes cells nevertheless exhibit brief, spontaneous departures from slow walking during the course of movie acquisition. Any frame in which the movement of the cell and/or cirri were too fast to be resolved, such as during spontaneous escape responses<sup>27</sup> ([Video S2](#)), was excluded from analysis such that some videos were split into a number of separate continuous sequences. Thus, each movie frame associated with a particular time point in the walking trajectory, with the exception of those excluded from analysis as described, yielded a corresponding 14-element binary vector encoding the motility state of the cell in terms of the movement of cirri. Cell movement was tracked using the manual tracking feature of the TrackMate plugin in FIJI.<sup>65</sup> The center of each cell was used as the reference feature for tracking. We analyzed the walking dynamics of 13 different cells.

### Dimensionality reduction

Dimensionality reduction was performed by non-negative matrix factorization (NMF) implemented in MATLAB release 2019b (Mathworks, Natick). NMF was chosen as a dimensionality reduction technique to allow us to obtain a reduced, sparse, and interpretable representation of walking dynamics. Because NMF derives non-negative factors, the basis vectors in NMF space correspond straightforwardly to patterns of cirral activity. NMF involves factoring data,  $A$ , an  $n$  by  $m$  matrix, into non-negative factors  $W$ , an  $n$  by  $k$  matrix, and  $H$ , a  $k$  by  $m$  matrix where the product  $W^*H$  approximates  $A$ . To determine the appropriate number of dimensions or rank,  $k$ , that are necessary to accurately represent the data without overfitting, we performed cross-validation by imputation with random holdouts,<sup>66</sup> also implemented in MATLAB. We randomly held out 15% of our walking dynamics data, performed NMF for a given  $k$ , and then used the NMF reconstruction  $W^*H$ , to update the missing data entries. This process of updating is known as imputation, and we repeated the imputation process 50 times, by which point the imputed values had stabilized, to obtain a final NMF reconstruction. We then computed the root mean squared residual (RMSR) between the final NMF reconstruction,  $W^*H$ , and our dataset,  $A$ . We performed this entire process 100 times for each value of  $k$ . As is generally the case for NMF, we observed a monotonic decrease in reconstruction error with increasing  $k$  without performing the imputation procedure ([Figure S1](#)). In contrast to this trend, we observed an increase in RMSR of imputed values with increasing  $k$  indicating overfitting<sup>66</sup> ([Figure S1](#)). We chose  $k=3$  because this value was the highest value before a notable increase in imputation error ([Figure S1](#)), which would indicate overfitting.<sup>66</sup> Thus, our choice of rank 3 selects the lowest rank approximation that captures structure of the dataset without overfitting that structure. Further, our choice facilitated the visual inspection of the structure of data in the reduced dimensional reconstruction.

Finally, we noted that for our chosen value of  $k$ , due to the stochastic nature of the NMF algorithm, which involves a random initialization step, we obtained slightly different solutions for different iterations.<sup>66</sup> In order to choose the best reduced dimensional approximation, therefore, we performed NMF 500 times and chose the particular solution corresponding to the lowest RMSR compared to our dataset.

### Clustering

Clustering on the dataset obtained using NMF was performed by density-based spatial clustering of applications with noise (DBSCAN) algorithm<sup>39</sup> implemented in MATLAB release 2019b (Mathworks, Natick). Structure in NMF space was clearly visible ([Figures 2E–2G](#)), and DBSCAN using a Euclidean distance metric, was initially chosen as a clustering method because it yielded qualitatively good partitioning of the data. The DBSCAN algorithm involves stochastic search within neighborhoods of a given radius  $\epsilon$  around datapoints, and points with a minimum number of neighbors,  $n$ , within their neighborhood are grouped as belonging to the same cluster, leaving two free parameters to determine. We set  $\epsilon$  by first using the `clusterDBSCAN.estimateEpsilon` function in MATLAB (release 2020b, Phased Array System Toolbox), which yielded a value of 0.15. We next set about determining the minimum neighbor number,  $n$ . To do so, we computed the average Silhouette coefficient, a commonly used measure of clustering quality that indicates how well-separated clusters are, for various values of  $n$ . The results of this analysis are plotted in [Figure S1](#). Higher Silhouette coefficients indicate better clustering, and we found that a value of  $n=8$  maximized the mean Silhouette coefficient ([Figure S1](#)). We also noted, however, that for this value, many datapoints were found to be outliers, not belonging to any cluster due to having too few points within a distance of  $\epsilon$ . [Figure S1](#) displays percentage of datapoints found to be outliers as a function of  $n$ . In order to avoid categorizing more than 5% of datapoints as outliers, we chose to settle on  $n=4$ , which does not have a significantly

different mean Silhouette coefficient compared to any of the others in the range  $n=2\text{--}7$ . This choice was further supported by the fact that major clusters involving more than 5 datapoints identified with  $n=8$  were also identified with  $n=4$ .

Although this set of parameters gave qualitatively and quantitatively reasonable clustering results, we sought to further refine our clusters and to further reduce the outlier datapoints. We noted the obvious partitioning of the NMF dataset into three groups along the H2 axis (Figure 2E). We found the previously determined parameter values to yield good clustering for the top and middle partitions ( $H2 \geq 1.1$  and  $0.2 < H2 < 1.1$ ), with no outliers. For the lower partition ( $H2 \leq 0.2$ ), however, we found that we were able to improve clustering by using  $\varepsilon=0.1182$ . With this updated value, we found no statistically significant change in Silhouette coefficient and reduced outliers to 0%. The clusters obtained by this process constituted the identification of the 32 gait states. We note here that the problem of determining the true or optimal number of clusters is an unresolved problem,<sup>67</sup> and we note that we have followed standard methods to determine cluster number, and we found that our key results do not depend sensitively on the precise number of clusters identified (see following section and Figure S1 for more details).

### State transition analysis

Following dimensionality reduction and clustering to identify gait states, we proceeded to characterize state transition dynamics. For each cell trajectory, we identified all unique gait state transitions for a total of 1423 unique pairwise transitions over the cumulative 2343 video frames for 77.14 s of recording. We computed empirical transition rates between states as the total number of observed transitions divided by the total time of observation. In order to determine which transitions were balanced and which were unbalanced, we followed Chang and Marshall,<sup>32</sup> and performed binomial tests of statistical significance. Assuming a system at equilibrium, with all transitions obeying detailed balance, we expect to observe some deviation from exactly reciprocal transitions and can calculate the probability of observing a given set of ratios given underlying probabilities of forward and reverse transitions. The binomial probability of observing a set of transitions with known forward and reverse probabilities is given by

$$P(X = f) = \binom{n}{f} p_{\text{forward}}^f p_{\text{reverse}}^{n-f}$$

where  $\binom{n}{f} = \frac{n!}{f!(n-f)!}$  is the number of combinations,  $f$  is the number of forward transitions,  $n$  is the total number of transitions (such that  $n-f$  is the number of reverse transitions), and the probabilities  $p_{\text{forward}}$  and  $p_{\text{reverse}}$  are the forward and reverse probabilities. Considering only the set of transitions involving a specific pair of states, and calculating the probability that a transition between those states is either in the forward or reverse direction, the values of forward and reverse probabilities in the balanced case must be equal such that  $p_{\text{forward}} = p_{\text{reverse}} = 0.5$ . With an  $\alpha$  level of 0.05, we then considered reciprocal transition pairs with binomial probabilities less than 0.05 to be significantly unbalanced. Figure S1 displays the binomial probabilities associated with all transitions.

To further illustrate the nonequilibrium or temporally irreversible character of gait state transitions, we apply the Kolmogorov criterion, which specifies the necessary and sufficient condition for reversibility that the product of transition probabilities traversing any closed loop in state space must equal the product of the transition probabilities in the reverse direction.<sup>5,68</sup> Due to the presence of unbalanced transitions, the gait of *Euplates* clearly violates this condition; for example,  $k_{1,2}k_{2,3}k_{3,17}k_{17,1} = 0.003 \neq k_{1,17}k_{17,3}k_{3,2}k_{2,1} = 0.0005$ , where each  $k_{i,j}$  is the conditional probability of transitioning from state  $i$  to state  $j$  estimated as  $N_{ij}/N$ , with  $N$  the total number of transitions from state  $i$ .

In order to calculate the estimated entropy production rate, we followed Wan and Goldstein,<sup>5</sup> where the entropy production rate is defined as

$$\dot{S} = \frac{1}{2} \sum_{i \neq j} J_{ij} A_{ij}$$

with conjugate fluxes  $J_{ij} = p_i k_{ij} - p_j k_{ji}$  and forces  $A_{ij} = \ln\left(\frac{p_i k_{ij}}{p_j k_{ji}}\right)$  where the  $p_i$  are the probabilities of being in state  $i$  at steady state and the  $k_{ij}$  are the transition probabilities from states  $i$  to  $j$ . We estimate the state occupancy probabilities  $p_i$  as  $\frac{T_i}{T_{\text{Total}}}$ , where  $T_i$  is the amount of time spent in state  $i$  over all trajectories and  $T_{\text{Total}}$  is the total recorded time, and the transition probabilities  $k_{ij}$  as  $\frac{N_{ij}}{N_i}$ , where  $N_{ij}$  is the total number of observed transitions from state  $i$  to state  $j$  and  $N_i$  is the total number of transitions emanating from state  $i$ . To avoid  $k_{ji} = 0$  for pairs of states for which we did not observe any transitions during our experiments, we let  $k_{ji} = \frac{1}{p_j T_{\text{max}}}$  where  $T_{\text{max}} = 238$  is the maximum number of observed transitions for any single recorded walking trajectory.

In the course of our state transition analysis, we also checked whether the waiting times between instances of each state might be non-exponentially distributed, with exponential distributions indicative of an embedded Markov process or possibly self-organized criticality.<sup>69</sup> Using the Lilliefors test implemented in MATLAB, we found that in general, waiting times were not exponentially distributed, although states 2, 3, 6, 16, 17, 18, 25, 27, 28, 32 were found to have waiting times consistent with exponential distributions with Benjamini-Hochberg corrected p-values of 0.046, 0.046, 0.022, 0.008, 0.046, 0.017, 0.046, 0.0081, 0.0046, 0.0046 respectively. Interestingly, none of the waiting times between the movements of individual cirri were found to be consistent with exponential distributions. These results are consistent with mechanisms constraining the temporal dynamics of cirri and state transitions.

In order to begin evaluating whether state transitions obeyed the Markov property for a discrete-time, first-order Markov process, where the transition probabilities from one state to the next are completely determined by current state,<sup>44</sup> we estimated the transition

matrix for walking dynamics, consisting of the transition probabilities between all states. We estimate the transition probability from state  $i$  to state  $j$  as  $k_{ij} = \frac{N_{ij}}{\sum_k N_{ik}}$  such that  $\sum_i k_{ij} = 1$ . The entries of the transition matrix,  $P$ , are these transition probabilities with indices  $i$  for rows and  $j$  for columns. If gait state transitions obeyed the Markov property, we expect that the product of the transition matrix with itself,  $P^2$ , would be equivalent to the two-step transition matrix where transition probabilities are computed as before except that state  $j$  is the state to which  $i$  has transitioned after an intervening transition. [Figure S1](#) displays the results of this analysis showing that the two matrices show some quantitative and qualitative differences. Although these results strongly suggest violation of the Markov property, we applied the Billingsley test for a more statistically rigorous evaluation.[70,71](#) This test was implemented and performed in MATLAB. The Billingsley test gives a  $\chi^2$  metric with  $M^2 - 2M$  degrees of freedom given by

$$\sum_{i=1}^M \sum_{j=1}^M \frac{(N_{ij} - R_{ij} \sum_{l=1}^M N_{il})^2}{R_{ij} \sum_{l=1}^M N_{il}},$$

where  $R_{ij}$ , the independent trials probability matrix, is given by  $R_{ij} = \sum_{k=1}^M N_{kj} / (\sum_{h=1}^M \sum_{l=1}^M N_{hl} - \sum_{k=1}^M N_{ik})$ . Applying this test to our gait state transition data, we found that the null hypothesis that the gait conforms to a first-order discrete time Markov process was rejected ( $p=0.005$ ).

Importantly, we also noted that the key qualitative results of our state transition analysis are robust to the details of clustering results. In particular, we find that strongly unbalanced transitions and violation of the Markov property exist for a range of clustering parameters. [Figure S1](#) displays the transition matrices for different clustering results.

To arrive at the simplified, state machine representation of the gait, we focused on the highest probability transitions emanating from each state. Transition probabilities were estimated as  $k_{ij}$  (as defined above). This allowed us to prune away rare transitions in order to reveal the dominant structure of gait state transitions. [Figure 3G](#) displays the pruned transition matrix as a heatmap. We found that relatively few states were the recipients of the majority of high probability transitions, and many states received none. To more clearly visualize the structure of transitions, we grouped together all gait states receiving no more than one unique high probability transition based on the idea that state transitions into this group show little bias in terms of source state, and within the group, transitions between states exhibit low probability, time unbiased, equilibrium-like fluctuations.

### Biophysical model and simulations

For our simple biophysical model, we consider a 2D system in which a *Euplates* cell is walking across a surface in a low Reynolds number environment.[72](#) The cell has 14 cirri, which exist in one of two states: actively moving or not actively moving, following our quantitative gait characterization. Cell position and orientation is defined in terms of equilibrium position of the cirri. Cirri can generate a motive force to drive cell motility when moving and resist displacement when not moving and in contact with the surface. For our model, we remain agnostic to the details by which cirri produce generate force noting only that in our experiments, no cell displacement was observed when cirri were not moving and in contact with the coverslip. We therefore let cirri generate a constant force in the direction dictated by cell orientation when moving. We conceptualize the resistance to displacement of unmoving cirri as stemming from the adhesive interaction between the cirrus and the substrate on which the cell is walking and the energy required to bend or deflect a cirrus. Consistent with experimental observations, we do not allow for translation of a cirrus-substrate contact point while a cirrus is not actively moving.

For a particle moving through a fluid at low Reynolds number, such as our cell, velocity  $\vec{v}$  will be given by

$$\vec{v} = \frac{\sum \vec{F}}{\gamma},$$

where  $\sum \vec{F}$  is the sum of the external forces acting on the particle, and  $\gamma$  is constant related to the geometry of the particle and the viscosity of the fluid[73](#) accounting for drag. In our model, cirri are responsible for the forces involved in motility, so  $\sum \vec{F} = \vec{F}_a + \vec{F}_i$  with the motive force  $\vec{F}_a = n \vec{F}_a \vec{o}$  where  $n$  is the number of active cirri,  $\vec{F}_a$  is the magnitude of the constant force generated by active cirri, and  $\vec{o}$  is the unit vector in the direction of cell orientation, and the resistive force  $\vec{F}_i = \sum_i k \vec{d}_i$  is a sum over the inactive cirri where  $k$  is a constant controlling the resistance of a cirrus to deformation and  $\vec{d}_i$  is the displacement vector of inactive cirrus  $i$ . We note that the forces driving cell motility in *Euplates* stem from complex mechanical interactions, but for our model, we have chosen simple, first order expression to capture very basic features.

Similar to the expression for velocity above, angular velocity of a walking cell in our model is given by

$$\vec{\omega} = \frac{\sum \vec{F} \times \vec{r}}{\beta},$$

where  $\beta$  is a constant related to the geometry of the cell and viscosity of the fluid,  $\vec{F}_i$  is the force due to cirrus  $i$  ( $\vec{F}_a \vec{o}$  for active cirri and  $k \vec{d}_i$  for inactive cirri), and  $\vec{r}$  is the vector pointing from the center of the cell to cirrus  $i$ .

In addition to the relative positions of the cirri and patterns of cirral activity, the four parameters  $\gamma$ ,  $\beta$ ,  $F_a$  and  $k$  govern cell motility in our model. From these four parameters, we obtain three related dimensionless parameters:  $\alpha = \beta / \gamma l$  where  $l$  is the maximum distance between cirri, which can be thought of as characterizing the unsteadiness of the cell the intrinsic susceptibility of the cell to turning

due to cell geometry;  $M = F_a/kl$ , which can be thought of as characterizing degree to which cirral activity will tend to induce cell movement in opposition to inactive cirri; and  $\mathcal{F} = F_a t / \gamma l$  where  $t$  is the duration of a timestep in simulations, which can be thought of as the strength of the cirral motive force relative to the viscous drag experienced by the cell due to its fluid environment.

For all simulations, relative equilibrium cirral positions calculated from the average cirral positions in a video frame with no cirral activity over the 13 cells used for gait analysis (Figure 2) were used. In each simulation timestep, defined by the timestep used for recording videos for gait analysis, velocity and angular velocity are calculated based on the positions and activity of cirri, and the positions of all cirri are updated accordingly before proceeding. To calibrate the parameters of the model, we used the cirral patterns recorded from the cells used for gait analysis. We swept parameter space and found that simulations qualitatively and semi-quantitatively recapitulated experimentally measured cell motility with  $\alpha = 0.001$ ,  $M = 0.26$ , and  $\mathcal{F} = 0.008$  (Figures 3K and S1). These parameter values were subsequently used for all simulations.

For simulations with shuffled gait state transitions, we used MATLAB's shuffle command on the cirral dynamics of actual cells to obtain sequences of gait state transitions of the same length as those that were experimentally obtained except in a random order. To obtain random patterns of cirral activity similar to those measured experimentally, we generated cirral activity according to a process defined by two probabilities:  $p_a$ , the probability of transitioning from inactive to active at each timestep and  $p_i$ , the probability of transitioning from active to inactive at each timestep. We initialized sequences with no cirral activity and then updated cirral activity according to these probabilities for each timestep in the sequence. We found that setting  $p_a=0.3$  and  $p_i=0.1$  yielded the same average cirral activity as that recorded experimentally,  $0.23 \pm 0.42$  per frame and  $0.24 \pm 0.43$  per frame respectively where the values are mean  $\pm$  standard deviation. All simulations were performed in MATLAB.

### Confocal microscopy

Cells were prepared for imaging and placed into a FluoroDish as described in the [live cell brightfield microscopy](#) section. Cells were then labeled with SiR-tubulin (Spirochrome provided by Cytoskeleton, Inc, CY-SC002) at  $1 \mu\text{M}$  concentration. Cells were imaged using a Zeiss LSM 880 AxioExaminer and a 40x, 1.2 NA C-Apochromat water immersion objective (Zeiss) and excitation provided by a 633 nm laser (Zeiss). Only one full confocal z-stack of a complete cell was obtained during imaging to avoid effects of photodamage.

### Fiber reconstruction and analysis

The image stack resulting from confocal imaging was first aligned in FIJI using the StackReg plugin.<sup>61</sup> Next, fibers were manually segmented in each of the aligned z-stack images using the TrakEM2 plugin in FIJI.<sup>74,75</sup> Thick and thin fibers (Figure 4A) were morphologically distinguished, with thick fibers having a diameter of no less than  $5 \mu\text{m}$  at the thinnest point. Fibers were traced from their distal termini to their convergences at the base of the cirri with which they were associated. Following segmentation, 3D surfaces were reconstructed in TrakEM2. Inter-fiber contacts were then found by inspection of 3D reconstructions and verified by examining individual z-stack frames to confirm intersections between fibers.

### Drug treatment experiments and analysis

For all cytoskeleton inhibitor treatment experiments, 1 mL of cells in culture were placed in wells of 12-well plates (Thermo-Scientific, 12-565-321). Nocodazole (Sigma-Aldrich, M1404) as a stock solution of 6.64 mM in DMSO diluted further in distilled water, was added to achieve appropriate concentrations, with no more than  $1 \mu\text{L}$  of additional volume added, and  $1 \mu\text{L}$  of distilled water with equivalent DMSO concentration to nocodazole treatments added to controls. Cells were incubated for 1 hr before initiating experiments. No cell death was observed at any concentration of nocodazole in the 6 hours following nocodazole treatment. Paclitaxel (Sigma-Aldrich, T7191) as a stock solution of 2.23 mM in DMSO diluted further to  $20 \mu\text{M}$  in distilled water, was added to achieve a final concentration of 20 nM to cells in solution. Latrunculin B (Thomas Scientific, C834E37) as a stock solution of 1.1 mM in ethanol was further diluted in distilled water and added to achieve a final concentration of  $10 \mu\text{M}$  with cells in solution. Cytochalasin B (Fisher Scientific 1493-96-2) as a stock solution of 2.1 mM in ethanol was further diluted in distilled water was added to achieve a final concentration of  $50 \mu\text{M}$  with cells in solution. Jasplakinolide (Fisher Scientific 42-012-750UG) as a stock solution of 1 mM in DMSO was further diluted in distilled water to achieve a final concentration of  $10 \mu\text{M}$  with cells in solution. No cell death was observed in the 6 hours following treatment with any of the actin inhibitors. For the control condition for actin inhibitor experiments, both DMSO and ethanol was added to match the concentrations added in the cytochalasin B and jasplakinolide conditions.

### Washout experiments

For nocodazole washout experiments, motility assays (described below) were also performed after placement of cells into well plates and before nocodazole treatment. Following  $0.2 \mu\text{M}$  nocodazole treatment and another motility analysis, cells were picked in  $5 \mu\text{L}$  of media and placed into wells of 6 well plates (Corning, CLS3736) each containing 2mL of fresh media. No more than  $20 \mu\text{L}$  of nocodazole treatment media was added to any well so that the resultant nocodazole concentration in the washed condition was no more than 2 nM. Cells were allowed to recover in this condition for 4 hr, and then a final motility assay was performed.

### Motility analysis

For motility analysis, cells were picked from well-plates and placed onto well-slides created by using a paper hole punch tool to punch a hole in 0.25 mm thick silicone spacer material (CultureWell, 664475), and adhering the spacer material to a glass slide (Corning, 2947-75X25). A total volume of  $20 \mu\text{L}$  including up to 4 cells was added to well slides. A glass coverslip (FisherScientific, 12-545-D) was placed atop the well slide, sealing the well and creating an imaging chamber. After creating the imaging chamber, cells were allowed to acclimate to their new environment for 10 min. Cells were then imaged on a Zeiss Axio Zoom.V16 microscope under darkfield

illumination with a Canon EOS T5i DSLR camera recording at 30 fps for 2 min. Videos were then processed using FIJI.<sup>61</sup> First, images were cropped to remove extraneous parts of the field of view that did not contain the imaging chamber, and then background subtraction was performed by creating an image composed of mean pixel intensity values over all frames of the video and subtracting this mean image from all frames of the video. A mean filter with a four-pixel radius was then applied to each frame of the video for the purpose of smoothing. After processing, tracking was performed using the TrackMate FIJI plugin.<sup>65</sup> For detection of objects (cells), a Laplacian of Gaussian filter was applied with an estimated blob diameter of 25 pixels and threshold of 0.1. A quality threshold was set manually when necessary to filter out any detected objects that were not cells. The Linear Assignment Tracker with a linking max distance of 15 pixels, gap-closing max distance of 150 pixels, and a gap-closing max frame gap of 200 frames was then used to generate linked tracks (trajectories) of detected cells. Trajectories were analyzed in MATLAB. Scaled path length for each tracked cell was calculated by summing the length of all segments of the track and dividing by the maximum distance the cell traveled from its starting point.

#### Fiber length analysis

Just prior to confocal imaging, cells were washed by picking up to five cells in 10  $\mu\text{L}$  volume each and placing into 1 mL fresh media in a 12-well plate (Thermo-Scientific, 12-565-321). Cells were then prepared for imaging as described in the [confocal imaging](#) section of [method details](#). Confocal z-stacks were then loaded into FIJI and aligned using the StackReg plugin.<sup>61</sup> Cell lengths were determined by finding the maximal distance between two points on the front and rear ends of the cell. Because of variability in the detectable fibers in nocodazole treated cells, only fibers associated with the rear cirri (j-n), which were visible in all cells, were used for analysis. All of these rear fibers were measured, and the reported scaled fiber length was obtained by dividing the length of the longest fiber by the corresponding cell length. In all cases, the fiber associated with cirrus m was the longest fiber.

#### Analysis of cirral dynamics in nocodazole and paclitaxel treated cells

Cells were prepared for imaging and imaged as described in the [live cell brightfield microscopy](#) section of the [method details](#) with the exception that a Canon EOS T5i DSLR mounted on a Zeiss Axio Zoom.V16 microscope was used to record movements. Additionally, video was recorded at 0.066 seconds per frame to avoid blurring and then videos were downsampled to 0.033 seconds per frame for analysis. Cirral dynamics were quantified as described in the [quantification of walking dynamics](#) section of the [method details](#).

To assign cirral configurations of nocodazole treated cells to previously identified gait states, we first matched any cirral configurations with known gait state identity. Next, due to proximity of new cirral configurations not observed in untreated cells ([Figure S1](#)), we were able to map the new cirral configurations onto the clusters defining the gait states by determining the nearest cluster to the new cirral configuration. Distance between new cirral configurations and clusters were determined by finding the shortest distances to points defining convex hulls of each cluster. The shortest of all these distances then indicated the nearest cluster and corresponding gait state to which the new cirral configuration was assigned. We noted an obvious dense cluster of points corresponding to activation of nearly all cirri, and we identified a new cluster and corresponding gait state by applying the DBSCAN algorithm as described in the [clustering](#) section of the [method details](#). Evaluation of transition dynamics was performed as described in the [state transition analysis](#) section of the [method details](#). This analysis was all performed in MATLAB.

#### $\varepsilon$ -machine construction

Our representation of the Euplotes at timestep t takes the form of a length 14 binary string  $x^t = [x_1^t, x_2^t, \dots, x_{14}^t]$  where  $x_i^t \in \{0, 1\}$ . For reducing dimensions, we found that bigrams of the 14-dimensional strings, yielded more consistent, interpretable results than unigrams, so  $x^t = [x_1^t, x_2^t, \dots, x_{14}^t, x_1^{t+1}, x_2^{t+1}, \dots, x_{14}^{t+1}] \in \{0, 1\}^{28}$ .

In order to learn latent states, we used Variational AutoEncoders (VAEs)<sup>76</sup> to reduce each fourteen dimensional timestep to three dimensions. VAEs in particular are used for their ability to learn an interpolatable latent space where high-dimensional training data points are mapped to points in low-dimensional space that mimic a normal distribution centered around the origin. This processes yields a smooth latent space with dimensions that represent core aspects of the data. We used a very minimal VAE with an Adam optimizer,<sup>77</sup> consisting of one dense layer for the generating encoder means and one dense layer for generating encoder variance. This creates a 3-dimensional normal-like distribution, which we sample from and decode with a one layer decoder.

#### Variational Encoder ( $x^t \rightarrow z^t \in \mathbb{R}^3$ )

After reducing the dimensionality of our data, we perform spectral clustering<sup>78</sup> using scikit-learn<sup>62</sup> with 100 k-means runs and 4 clusters, where the cluster count was decided by inspection of the resulting classes.

#### Spectral Clustering ( $z^t \rightarrow c^t \in \{1, 2, 3, 4\}$ )

Applying this procedure to the data from the cells used for gait analysis (the same as used for [Figure 2](#)) with bigram training parameters lr=0.0011, weight\_decay=0.000727, k=3, hidden\_dim=5, beta1=0.5, and beta2=0.900, we found the following four clusters corresponding to particular patterns of cirral activity: 1) Front cirral activity 2) High overall cirral activity 3) Rear cirral activity and 4) Little cirral activity ([Figure S5](#)). When we applied this procedure to the data from cells treated with 0.2  $\mu\text{M}$  nocodazole (the same data used for [Figure 5](#)) with bigram training parameters lr=0.0026, weight\_decay=0.000106, k=3, hidden\_dim=3, beta1=0.5, and beta2=0.999, we obtained only three distinct clusters with similar cirral activity patterns to those of untreated cells except that the rear cirral activity cluster (3 above) disappeared ([Figure S5](#)). These clusters represent an independent, alternative coarse graining of patterns of cirral activity into gait states to that presented in [Figure 2](#). Although independent, this procedure identified the same underlying structure in the data. In particular, the alternative gait states here involve similar patterns of cirral activity to those defining the NMF components depicted in [Figure 2H](#). Whereas the spatially distinct groups of cirri sharing the most mutual information define the space of gait states under the NMF-based approach, under this alternative, activity or complete lack thereof in those same groups of cirri defines the gait

states. Furthermore, the loss of cluster 4 under nocodazole treatment is consistent with the loss of gait states as depicted in [Figure 5I](#).

To construct  $\varepsilon$ -machines,<sup>46</sup> characterized by a set of causal-states and transitions between them, we feed a stream of behavioral symbols  $c_0, c_1, \dots, c_T$  into a Causal-State Splitting Reconstruction (CSSR)<sup>50,63</sup> algorithm. Due to the nature of the CSSR algorithm, we can only construct  $\varepsilon$ -machines from unbroken data streams, so we constructed  $\varepsilon$ -machines for each cell from a given dataset. Performing this process over gait state data streams derived from cirral activity recordings as described above for each cell from the untreated and nocodazole datasets yielded a set of  $\varepsilon$  machines. Representative examples of  $\varepsilon$ -machines obtained from each dataset are depicted in [Figure S5](#).

### Microsurgery experiments

All microsurgeries were performed by hand under observation with a Zeiss Axio Zoom.V16 using quartz microneedles pulled from quartz rods (Sutter Instrument, QR100-10) using a Sutter Instrument P-2000 Laser-based micropipette puller. Individual cells were picked and placed in a 5  $\mu\text{L}$  droplet at the center of an imaging chamber as described in the [motility analysis](#) section of the [method details](#). For bisections, microneedles were quickly and firmly pressed onto cells just anterior to cirrus h, cleanly severing the cell including the fibers associated with cirri j-n and sealing the cell membrane. Cells were allowed to recover for at least 5 minutes until some motility resumed, and then 15  $\mu\text{L}$  of fresh culture media was added to the imaging chamber, which was subsequently sealed with a glass coverslip (FisherScientific, 12-545-D) to minimize evaporation. After 24 hr, once cell fragments resumed walking motility, gait analysis was performed as described in the [quantification of walking dynamics](#) section of the [method details](#), except that cell fragments only had 7 cirri (j-n) instead of the full 14. Experiments with wounded cells were conducted in the same fashion as described for bisections except that instead of cutting cells in half, wounding was performed by stabbing with a microneedle a portion of the cell that does not have any fibers and was confirmed visually by the loss of some amount of cytoplasm.

### QUANTIFICATIONS AND STATISTICAL ANALYSIS

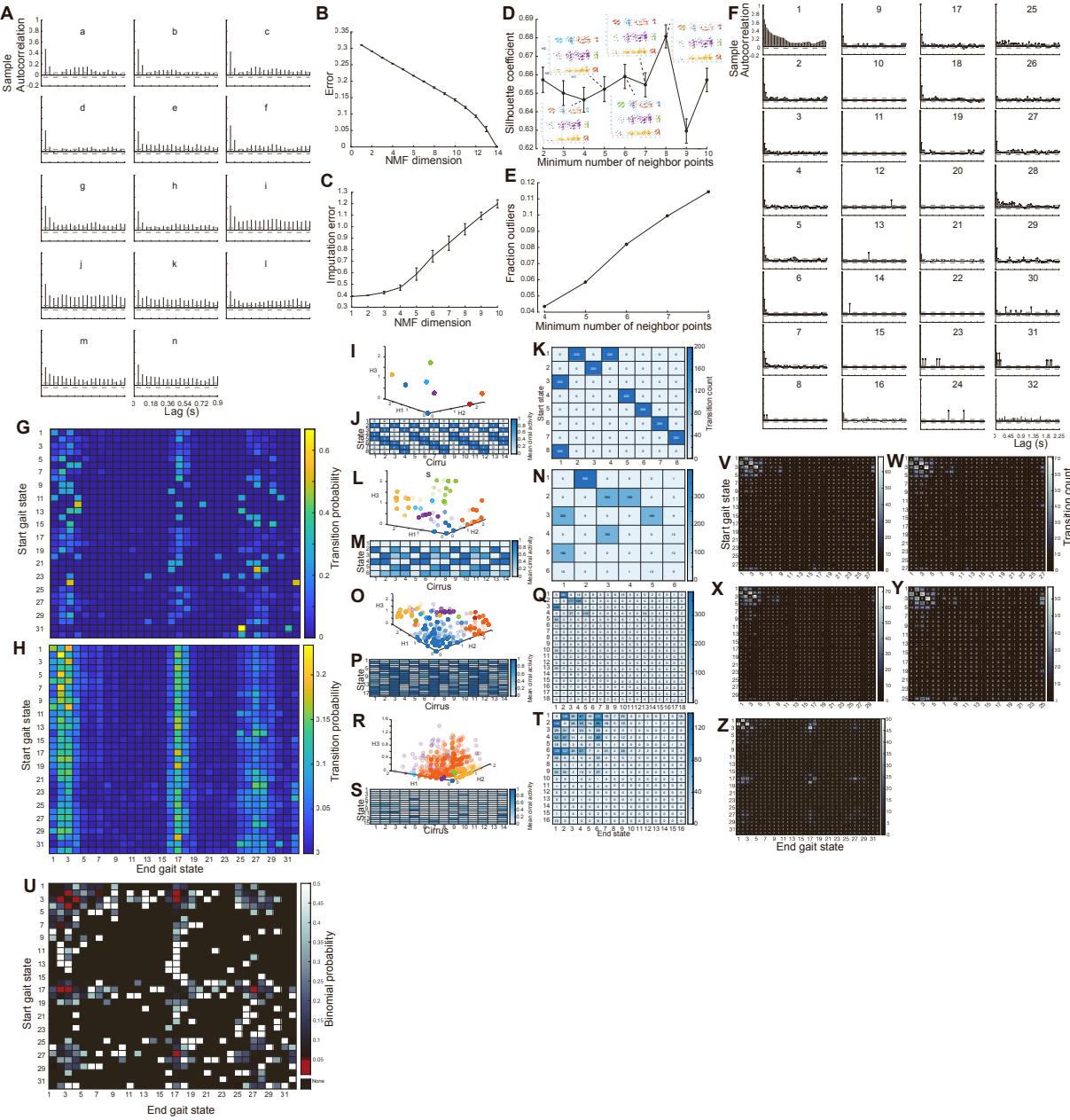
Statistical details of the experiments can be found in the figure legends, the main text, or the [method details](#) section. Statistical details include exact value of n, what n represents (generally the number of cells measured), definitions of center, and dispersion and precision measures. Statistical tests and fits were performed as described in the [method details](#) section using MATLAB release 2019b or 2020b (Mathworks, Natick).

**Current Biology, Volume 32**

**Supplemental Information**

**A unicellular walker controlled  
by a microtubule-based finite-state machine**

**Ben T. Larson, Jack Garbus, Jordan B. Pollack, and Wallace F. Marshall**



**Figure S1. Additional data and information concerning analysis of walking dynamics and our approach to coarse-graining. Related to Figure 2 and Figure 3.** (A) Autocorrelation analysis of the dynamics of cirri shows no clear periodicity in cirral activity. Plots represent the results for each cirrus *a-n* individually. Although we observe statistically significant autocorrelation at various lag times for all cirri, we do not observe clear peaks indicative of

periodic activity at a particular frequency. Some cirri, for example  $i$ ,  $j$ , and  $m$  display significant autocorrelation signal consistent with non-stationary dynamics and possibly weak periodic behavior. Although the signal is not strong, cirrus  $a$  has an autocorrelation profile with a single peak and so is most consistent with periodic activity at a particular frequency, in this case, at around 0.45 s. Dashed lines are the upper and lower confidence bounds consisting of two standard errors. (B, C) Cross-validation of non-negative matrix factorization (NMF) by imputations indicated that a rank of 3 yielded an acceptable reduced dimension representation of the 14-dimensional dataset of patterns of cirral activity for all walking trajectories. (B) A plot of reconstruction error, as measured by root mean square residual (RMSR) versus NMF dimension (or rank) shows, as expected that error decreases monotonically with increasing dimensionality. Plot is of mean RMSR and error bars are the standard error of the mean from 100 NMF reconstructions (see Method Details). (C) A plot of RMSR of imputed values versus NMF dimension shows that imputation error increases with increasing NMF reconstruction rank. Increasing error in imputed values is indicative of overfitting<sup>1</sup>. Plot is of mean RMSR and error bars are the standard error of the mean from 100 instances of the final result of 50 imputation iterations for each NMF dimension. See also Method Details for more information. Selecting a rank of 3 allowed us to select the lowest rank approximation that best described structure in the data without overfitting while also allowing for visual inspection of dimensionality reduction results, which also facilitated the evaluation of clustering. (D, E) Evaluation of parameters for clustering was performed by examining Silhouette coefficients and by minimizing outlier points. (D) Plot displaying Silhouette coefficient, a measure of cluster separation, as a function of the minimum number of neighbor points parameter,  $n$ , from the density-based spatial clustering of applications with noise (DBSCAN) algorithm along with visual representations of clustering

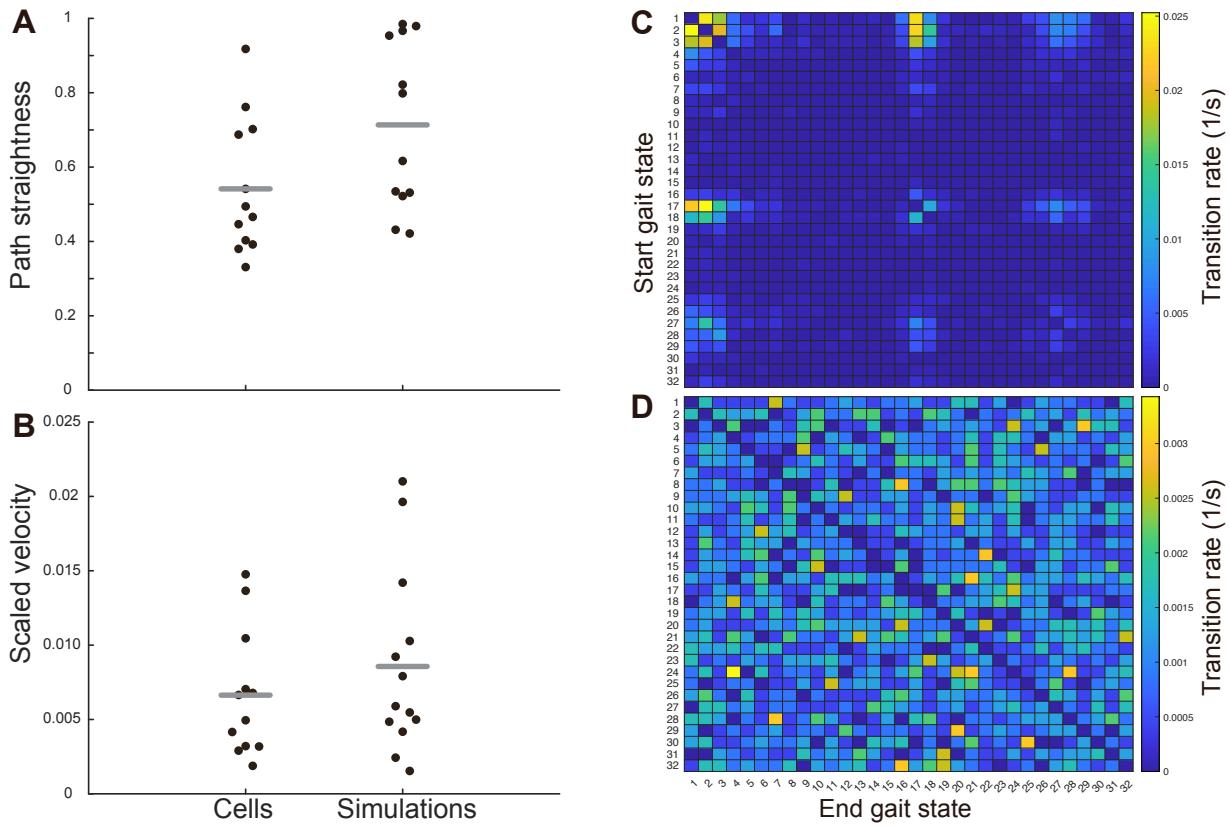
results. Silhouette coefficient was found to be maximized at  $n=8$  and was found to have values that were not significantly different from one another for  $n=2-7$ . In visual representations of clustering results, points belonging to clusters were given the same color, according to a random colormap. Note that for all values of  $n$  less than 8, the major clusters containing more than 5 points picked out for  $n=8$  are present. Points of the main plot are the mean Silhouette coefficient for clusters obtained by DBSCAN (see Method Details), and error bars are standard error of the mean. (E) A plot of the fraction of points identified as outliers by DBSCAN as a function of  $n$  shows that outliers increase with  $n$ . Although we noted that  $n=8$  maximized the mean Silhouette coefficient, over 10% of points were identified as outliers. The existence of outliers was problematic for our identification of gait states by clustering in that it left us unable to classify a significant fraction of cirral configurations as belonging to particular gait states. We therefore chose  $n=4$ , with fewer than 5% outliers, as a starting point, and further refined our clustering methods to eliminate all outliers (see Method Details). (F) Autocorrelation analysis of gait state dynamics shows no clear periodicity, which would manifest as distinct peaks in autocorrelation, in state transitions and that gait state 1 has the most predictable dynamics. Autocorrelation was calculated for binarized sequences corresponding to cells occupying the particular gait state denoted by the number associated with each plot. The substantial and slowly decaying autocorrelation of gait state 1 indicates predictability in the dynamics of gait state 1. For other gait states, we see little to no significant autocorrelation. For those gait states with a single or few very sharp increases in autocorrelation lasting for single time lags, we note that cells occupy those states very infrequently, and the elevated autocorrelation signal corresponds to single instances of state occupancy recurrences. Dashed lines are the upper and lower confidence bounds consisting of two standard errors. (G, H) Comparison between the 2-step transition

matrix computed directly from data and that computed by taking the product of the single step transition matrix does not rule out the possibility that the gait of *Euplates* satisfies the Markov property. (G) The 2-step transition matrix computed from walking cell trajectories directly from gait state transitions displayed as a heatmap. (H) The 2-step transition matrix computed by taking the product of the single step transition matrix with itself displayed as a heatmap. The structure of the 2-step transition matrixes displayed in panel G and H show some differences but are not inconsistent with one another, which leaves open the question of whether the gait of *Euplates* satisfies the Markov, “memoryless,” property. (I-T) Our procedure for coarse graining gait states from cirral dynamics robustly preserves underlying gait structure. (I-K) The results of the coarse graining procedure for gait state identification and transition analysis on a sequence with known, determinate patterns of cirral activity. The patterns of cirral activity are the same as those depicted in panel J repeated according to the transitions depicted in panel K for a total of 1800 timepoints. (I) The result of NMF and DBSCAN clustering on the patterns of cirral activity with points representing particular configurations of cirral activity, and the randomly selected colors representing cluster membership is depicted in NMF-space. (J) The gait states corresponding to the clusters in panel I represented as a heatmap with values obtained by averaging over all points within a given cluster show that we recover the distinct patterns of cirral activity. (K) The transition matrix obtained by analysis of gait state dynamics recovers the determinate structure of gait state transitions. (L-N) The results of the coarse graining procedure for gait state identification and transition analysis on a sequence with the same known, determinate patterns of cirral activity as in panels I-K but with noise applied by changing cirral activity with a probability  $p=0.01$  for each cirrus at each timepoint. Panels follow the same structure as I-K. Although we added random noise to the patterns of cirral activity, resulting in the spread of

points in panel I, we nevertheless largely recover the underlying determinate character of gait state transitions. (O-Q) The results of the coarse graining procedure for gait state identification and transition analysis on a sequence with the same known, determinate patterns of cirral activity as in panels I-K but with noise applied by changing cirral activity with a probability  $p=0.10$  for each cirrus at each timepoint. In this case, we observe an expected increase in the spread of points due to the substantial added noise. Note that with this dimensionality reduction we obtain 68% explained variance, which is comparable to the 64% explained variance for the data from Figure 2. Note that we again obtained a different number of gait states and even more variability within the gait states compared to panel M. Although the structure of the gait states and the cirral dynamics from which they are derived were more complicated in this case, the underlying determinate character of gait state transitions is still reflected in the transition matrix, although the noise level and coarse graining has shifted the transition counts substantially. (R-T) The results of the coarse graining procedure for gait state identification and transition analysis on a sequence of random cirral activity generated according to the process described in the Biophysical Model and Simulations section of the Method Details. In this case, we observed very little structure as expected due to the lack of underlying pattern in cirral configurations. This lack of structure is also reflected in the gait states corresponding to the clusters in panel I. As expected, the gait state transitions display symmetric structure depicted in the transition matrix. None of the gait state transitions were found to be significantly unbalanced by the binomial test. (U) Binomial probabilities corresponding to gait state transitions show the presence of significantly unbalanced transitions indicating broken detailed balance. The heatmap displays the values of binomial probabilities of obtaining the observed transition frequencies under detailed balance in which reciprocal transitions are balanced. Significantly unbalanced transitions are

those with binomial probability less than 0.05, and these cells appear red in the colormap. Note that many possible gait state transitions have no values, denoted by gray. These state transitions were not observed, in either the forward or reverse direction, during recorded walking trajectories analyzed in this study. (V-Z) Transition frequencies for gait states identified by partitioning of data using different clustering parameters illustrate robustness of key qualitative results to details of clustering. Heatmaps of raw transition counts, with numerical counts also displayed within the heatmap, measured according to gait states identified by clustering using DBSCAN with different parameters showing the presence of unbalanced transitions is robust to details of clustering. For example, in all cases, gait state transitions from state 2 to 3 and 3 to 4 are significantly unbalanced when compared to their associated reciprocal transitions ( $P < 0.05$  by binomial test, see Method Details). (V) Parameters for clustering were search neighborhood  $\epsilon = 0.15$  and minimum number of neighbors  $n = 5$ , yielding 28 gait states. (B) Parameters for clustering were  $\epsilon = 0.15$  and  $n = 6$ , yielding 27 gait states. (W) Parameters for clustering were  $\epsilon = 0.15$  and  $n = 7$ , yielding 30 gait states. (X) Parameters for clustering were  $\epsilon = 0.15$  and  $n = 8$ , yielding 25 gait states. (Y) This transition count matrix represents the gait transitions and gait states analyzed in the main text of the paper. To obtain these gait states, the dataset was partitioned first partitioned into three groups. Parameters for clustering for two of the partitions were  $\epsilon = 0.15$  and  $n = 4$ , and Parameters for clustering for the final partition were  $\epsilon = 0.1182$  and  $n = 4$ , yielding a total of 32 gait states. Note that for clustering displayed in panels V-Y outliers were present and were grouped into the gait state with the highest numerical index in each case. As such, they do not represent a distinct gait state associated with similar patterns of cirrial activity in the same way as the other gait states. No outliers were present in clustering yielding the gait states in panel Z. The apparent structural differences between the transitions in

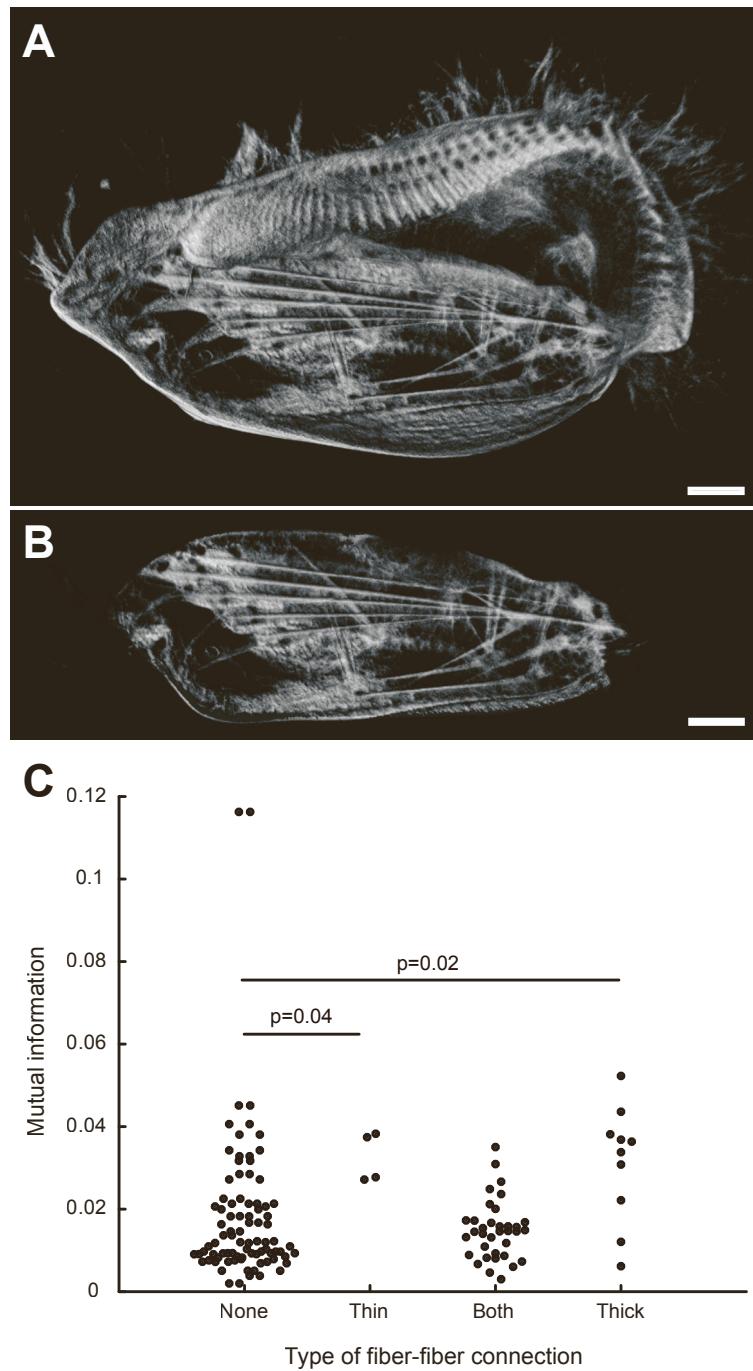
panel Z versus the others (e.g. the unbalanced transitions associated with gait state 17) can be explained by the presence of outlier points, which were all grouped together in panels V-Y in addition to the fact that gait states 3 and 4 from panels V-Y were partitioned differently and contain points that were grouped as gait states 17 and 18 in panel Z.



**Figure S2. Additional information concerning simulations of the biophysical model of a walking cell, including model calibration and simulations of cirral dynamics. Related to Figure 3.** (A, B) Simulations can recapitulate experimentally measured features of cell motility.

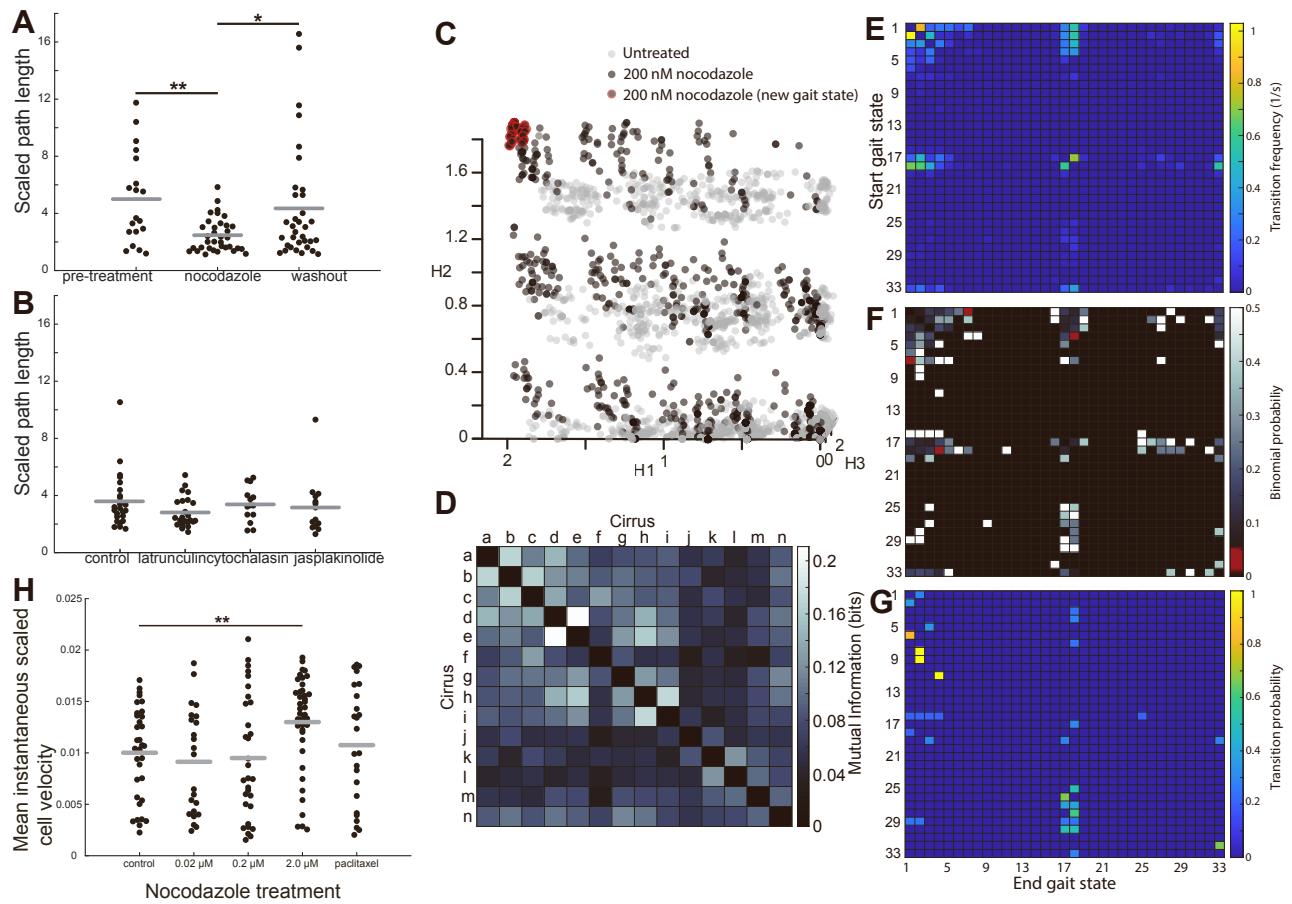
(A) A plot of path straightness from experimentally measured cell trajectories (data is the same as that from Figure 3) and from simulations shows that our model is able to semi-quantitatively reproduce motility dynamics in terms of the processivity of cell movement. Simulations were run using the same patterns of cirral activity as those recorded from cell experimentally. We note that while path straightness does not match exactly on a cell-by-cell basis, average path straightness is not significantly different ( $p>0.05$  by the Wilcoxon rank sum test). (B) A plot of path scaled velocity from experimentally measured cell trajectories (data is the same as that from Figure 3) and from simulations shows that our model is able to semi-quantitatively reproduce motility

dynamics in terms of cell speed. As for path straightness, cell speeds do not match exactly on a cell-by-cell basis, but average scaled velocity is not significantly different between the two ( $p>0.05$  by the Wilcoxon rank sum test). (C, D) Transition matrices display the structure of gait state transitions for shuffled and randomly generated patterns of cirral activity, both of which, in contrast to gait state transitions of walking cells, satisfy detailed balance. (C) The transition matrix obtained by randomly shuffling the order of experimentally measured patterns of cirral activity. These transition rates represent the average over five sets of separate shuffles of the gait state transition from each of the 13 experimentally measured cells. Note the symmetric structure of this matrix indicating that transitions obey detailed balance. We found no cases of significantly unbalanced transitions by binomial test for any shuffled set of gait state transitions. (D) The transition matrix obtained from random patterns of cirral activity. As in panel C, the values represent averages over five sets of gait state transitions with the total number of timepoints equal to that from experiments. Although the structure may look asymmetric, note the difference in heatmap values. As for the shuffled sets, we did not observe any significantly unbalanced transitions by the binomial test.



**Figure S3. Additional visualization of data and information concerning the microtubule cytoskeleton and fiber system as well as the relationship between the fiber system and cirral dynamics. Related to Figure 4.** (A, B) Confocal reconstruction of a SiR-tubulin labeled cell shows the tubulin-based fiber system associated with cirri. (A) An entire cell shows the

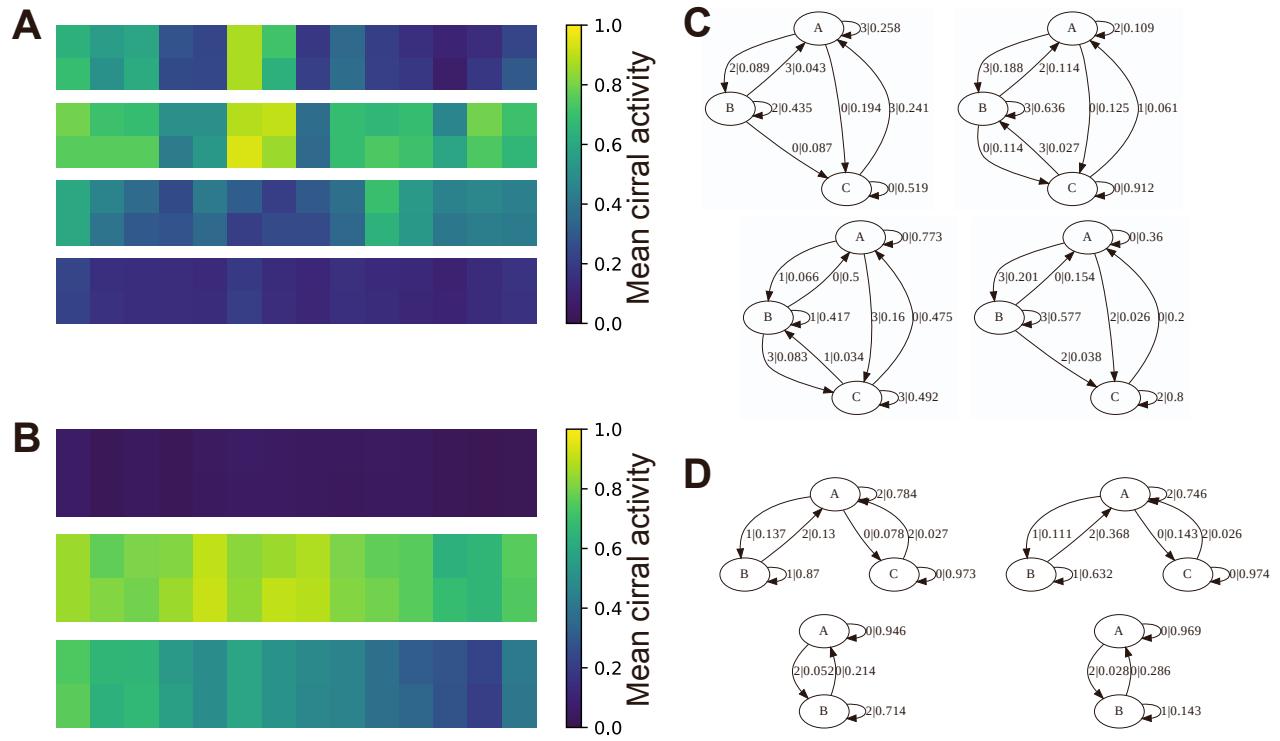
complex structure of the tubulin-based cytoskeleton of *Euplotes*. (B) The portion of the cell containing only fibers highlights the fiber system reconstructed in Figure 5 of the main text. Images were rendered using Imaris Viewer 9.6.0. Scale bars are 10  $\mu\text{m}$ . (C, D) Cirri associated with only thin or only thick fiber-fiber connections show increased mutual information compared to those lacking fiber-fiber connections. Bee-swarm plots are of mutual information (y-axis) with points representing mutual information shared by all pairs of cirri grouped by type of fiber-fiber connection or lack thereof. “None” indicates lack of fiber-fiber contacts, “Thin” indicates thin fiber to thin fiber connections, “Both” indicates thin fiber to thick fiber connections, and “Thick” indicates thick fiber to thick fiber connections. Both thin and thick fiber-fiber connections were found to be associated with increased mutual information (medians of 0.032 and 0.034 respectively) compared to cirri associated with no fiber-fiber connections (median of 0.012). Cirri associated with both thick and thin fiber-fiber connections (median of 0.015) show no significant difference in mutual information compared to those with none. Statistical significance was evaluated by the Kruskal-Wallis test, and p-values are displayed on the plot.



**Figure S4. Additional data, information, and results concerning drug treatments. Related to Figure 5.** (A, B) The effects of nocodazole treatment are reversible and specific to the perturbation of the microtubule cytoskeleton. (A) Nocodazole washout experiments show that the motility effect of nocodazole treatment measured by scaled path length are reversible. Scaled path length is significantly different from nocodazole treatment both before treatment and following washout ( $p=0.004$  and  $p=0.017$  by the Wilcoxon rank sum test respectively), and not significantly different when comparing pre-treatment with the washout condition ( $p=0.379$  by the Wilcoxon rank sum test). Data for each condition is aggregated from experiments conducted over three separate days and involves measurements of at least 20 cells. (B) Actin inhibitors have no discernable effect on cell motility as quantified by scaled path length ( $p>0.09$  by Wilcoxon

rank sum test in all cases). Measurements are from at least 14 cells and are aggregated from experiments conducted on three different days. (C-G) Cells with fiber systems disrupted by the microtubule inhibitor nocodazole exhibit altered cirral dynamics. All data are pooled from the analysis of 6 cells under 0.2  $\mu$ M nocodazole treatment over a total of 1133 timepoints. (C) Disruption of the fiber system biases cells toward increased cirral activity, although most new configurations are near the original gait states. The axes of this plot of the cirral configurations in NMF space are those obtained by the analysis of untreated cells (Figure 2E-G). Gray points are the cirral configurations of untreated cells as in Figure 2E-G, and black points are the cirral configurations obtained from 6 cells under 0.2  $\mu$ M nocodazole treatment over a total of 1133 timepoints. Red outlines of points denote those forming a new cluster, identified as gait state 33, which corresponds to activity in nearly all of the cirri. (D) Mutual information increases and shows an altered pattern under nocodazole treatment. Instead of the small clusters of cirri sharing mutual information observed in untreated cells, cirri a-i, those from the front to the center of the cell share increased mutual information. Cirri d and e, located near the center of the cell, share the highest mutual information instead of cirri f and g, located near the edge of the cell, in untreated cells. (E) Transitions exhibit altered frequency patterns less broken detailed balance compared to untreated cells as illustrated by a heatmap of transition frequencies. Note that transitions from gaits state 2 to 3 are reduced in frequency and that gait state 18 in particular exhibits an increase in high frequency transitions. (F) A heatmap of binomial probabilities illustrates the reduction and change in state transitions exhibiting broken detailed balance compared to untreated cells. (G) A heatmap of transition probabilities between states, showing only the most probable transitions from a given state with all others set to zero, as in Figure 3G, highlights the structure of transitions. In cases where multiple state transitions from a state were

tied for the highest probability, all of these transitions are displayed. The structure is similar to that of untreated cells in that few states are the recipients of the majority of high probability transitions. Note that a major difference compared to untreated cells is that gait state 18 is the recipient of many more high probability transitions. (H) A plot of the mean instantaneous scaled velocity of walking cells measured over entire trajectories shows that microtubule inhibitors do not affect the instantaneous velocity of walking cells. We did observe a significant increase in instantaneous velocity of cells in the highest nocodazole concentration ( $p=0.003$  by the Wilcoxon rank sum test), but we note that this is due to many cells swimming instead of walking, something that was not observed in any of the other conditions. Measurements are made from the same dataset as Figure 5D.



**Figure S5.  $\epsilon$ -machines constructed from the cirral dynamics of untreated cells and nocodazole treated cells highlight the reduction in complexity upon perturbation of the fibers and are consistent with the results summarized in Figures 3H and 5I. Related to Figure 5.** (A) A heatmap representation of the four gait states identified by an alternative coarse-graining process applied to the same dataset as from Figure 2 involving VAEs and spectral clustering highlights distinct patterns of cirral activity. Heatmap values are averages over all cirral configurations within a particular cluster as in Figure 2J. (B) A heatmap representation of the three gait states identified by an alternative coarse-graining process applied to  $0.2 \mu\text{M}$  nocodazole treated cells (the same dataset as from Figure 5) involving VAEs and spectral clustering highlights distinct patterns of cirral activity. Heatmap values are averages over all cirral configurations within a particular cluster as in Figure 2J. Note the loss of the gait state involving low cirral activity biased toward the rear of the cell. (C) Representative  $\epsilon$ -machines constructed from gait state dynamics for individual cells (the same dataset as in panel A) using

the CSSR algorithm. These  $\epsilon$ -machines tend to show similar structure in terms of causal states and transitions between them. While the majority have similar structures to those depicted, some display more states and transitions and a small minority display fewer states. (D) Representative  $\epsilon$ -machines constructed from gait state dynamics for individual cells (the same dataset as in panel B) using the CSSR algorithm. These  $\epsilon$ -machines notably tend to be less complex than those constructed from the dynamics of untreated cells (e.g. panel C) in terms of causal states and transitions.

### **Supplemental References**

- S1. Kanagal, B., and Sindhwan, V. (2010). Rank Selection in Low-rank Matrix Approximations: A Study of Cross-Validation for NMFs. In Advances in Neural Information Processing Systems.