1   **A unicellular walker embodies a finite state machine**

2   Authors: Ben T. Larson[1], Jack Garbus[2], Jordan B. Pollack[2], and Wallace F. Marshall[1]

3   [1] Department of Biochemistry and Biophysics, University of California, San Francisco,

4   California, USA

5   [2] Computer Science Department, Brandeis University, Waltham, Massachusetts, USA

6

7   **Cells are complex biochemical systems whose behavior emerges from interactions**

8   **among myriad molecular components. The idea that cells execute computational processes**

9   **is often invoked as a general framework for understanding cellular complexity. However,**

10   **the manner in which cells might embody computational processes in a way that the**

11   **powerful theories of computation, such as finite state machine models, could be**

12   **productively applied, remains to be seen. Here we demonstrate finite state machine-like**

13   **processing embodied in cells, using the walking behavior of *Euplotes eurystomus*, a ciliate**

14   **that walks across surfaces using fourteen motile appendages called cirri. We found that**

15   **cellular walking entails a discrete set of gait states. Transitions between these states are**

16   **highly regulated, with distinct breaking of detailed balance and only a small subset of**

17   **possible transitions actually observed. The set of observed transitions decomposes into a**

18   **small group of high-probability unbalanced transitions forming a cycle and a large group**

19   **of low-probability balanced transitions, thus revealing stereotypy in sequential patterns of**

20   **state transitions. Taken together these findings implicate a machine-like process. Cirri are**

21   **connected by microtubule bundles, and we find an association between the involvement of**

22   **cirri in different state transitions and the pattern of attachment to the microtubule bundle**

23   **system, suggesting a mechanical basis for the regularity of state transitions. We propose a**

24   **model where the actively controlled, unbalanced transitions establish strain in certain cirri,**

25   **the release of which from the substrate causes the cell to advance forward along a linear**

26   **trajectory. This demonstration of a finite state machine embodied in a living cell opens up**

27   **new links between theoretical computer science and cell biology and may provide a general**

28   **framework for understanding and predicting cell behavior at a super-molecular level.**

29

30   **Introduction**

31   Cells are complex physical systems controlled by networks of signaling molecules.

32   Single cells can display remarkably sophisticated, seemingly animal-like behaviors [1–3],

33   orchestrating active processes far from thermodynamic equilibrium in order to carry out proper

34   biological functions [4,5]. Indeed, single cells can make decisions by sensing and responding to

35   diverse cues and signals [6], execute coordinated movements [7,8] and directed motility [9–12], and even

36   solve mazes [13,14] and possibly learn [15–18]. Such behaviors in animals arise from neural activity

37   and have been studied extensively, but we know comparatively little about the mechanisms of

38   cellular behavior [19,20]. In individual cells, behaviors emerge directly through the joint action of

39   chemical reactions [21], cellular architecture [3], physical mechanisms and constraints within the cell

40   [22,23] and interactions of the cell with its local environment [24]. The links between information

41   processing, decision-making, and their physical manifestation as cell state transitions suggest

42   that cellular behavior might be understood as an embodied computation [25,26]. The theory of

43   computation has often been invoked as a general framework for understanding cellular dynamics

44   [25,27–32], with environmental sensing by bacteria being a particularly deeply studied example [32–34],

45   and has been used to engineer programmable cell states [35], but the manner and extent to which

46   cells might embody functional, computational processes as well as the extent to which a

47   computational perspective on cellular behavior might prove productive remains to be seen.

48   Among the microbial eukaryotes (protists), ciliates display striking examples of

49   unicellular behavior including hunting [3], sensorimotor navigation [10], and predator avoidance [36].

50   Spirotrichous ciliates of the genus *Euplotes* are notable for their complex locomotion [37–39], using

51   bundles of specialized cilia called cirri to walk across surfaces [37,38] (Fig. 1a, Movies S1 and S2).

52   Depending on the species, cells generally have 14 to 15 ventral cirri arranged in a highly

53   stereotyped pattern used for walking locomotion [40]. *Euplotes* live in aquatic environments, and in

54   addition to walking, their cirri can be used for swimming and rapid escape responses [41] (Movie

55   S2). Oral membranelles (Fig. 1b) generate feeding currents to capture bacteria and small

56   protistan prey and are also used for swimming. Early 20th century protistologists were so

57   impressed by the apparent coordination of cirri that they proposed the existence of a rudimentary

58   nervous system, the neuromotor apparatus, to account for their observations [39]. This theory was

59   motivated in part by the presence of intracellular fibers connecting various cirri (Fig. 1C), now

60   known to be to be tubulin-based structures [42,43]. Although the walking movements of *Euplotes*

61   are superficially similar to those of animals such as insects, the low Reynolds environment of

62   aquatic microorganisms, where viscous forces dominate over inertial forces, imposes significant

63   physical constraints on all movements that do not impinge on the movements of larger terrestrial

64   animals [44].

65   How can a single cell coordinate a gait without a nervous system? Coordination, to the

66   extent that it exists in the gait of *Euplotes*, would require some kind of dynamical coupling

67   among cirri or between cirri and some shared external influence. Recently, analytical techniques

68   from statistical physics have been used to characterize, understand, and predict mesoscale

69   dynamics in biological systems, including cellular behavior [4,5,45,46]. These approaches rely on

70   coarse-graining the complexity of biological dynamics into states and analyzing the nature of

71   transitions between states. For example, a state representation allows us to ask whether forward

72   and reverse transitions between pairs of states are equal, a condition known as detailed balance

73   [4,47]. Systems that violate detailed balance operate in a non-equilibrium mode and can produce

74   directed cycles in state space [4,48]. Broken detailed balance has been observed in the motility

75   dynamics of cultured mammalian cells as well as the motility dynamics of a freely behaving

76   flagellate protist [5,45] and implies that non-equilibrium models are most applicable to such

77   systems [46].

78   When information processing drives patterns of state transitions, such a system can be

79   viewed in terms of automata theory, a fundamental level in the theory of computation [49–

80   51].Automata theory can be used to address problems of decision-making and control in complex

81   systems by providing predictive understanding independently of the underlying details of how a

82   given process is implemented [50]. Inspired by work considering cellular behavior in the context of

83   the theory of computation [25], we hypothesized that walking cells might be governed by finite

84   state automata with directed, processive movement arising from reproducible patterns of state

85   transitions.

86   The consistent structure of *Euplotes*, its mode of motility, and its ease of observation

87   makes these cells an ideal biological test-bed in which to apply theories of non-equilibrium

88   statistical mechanics and embodied computation, both of which rely on describing a system in

89   terms of discrete state transitions. Here, we use time-lapse microscopy and quantitative analyses

90   to show that *Euplotes eurystomus* walks with a cyclic stochastic gait displaying broken detailed

91   balance and exhibiting elements of stereotypy and variability, consistent with a finite state

92   automaton representation. The observed dynamics are reminiscent of behavioral regulation in

93   some cells and animals [5,52] but contrast with many well-characterized examples of cellular

4

94   motility [7,9,10,12,53–56]. Our results provide a clear demonstration of machine-like processes

95   governing cellular state transitions and serve as a framework for investigating the principles of

96   behavioral control and non-equilibrium dynamics in single cells.

97

98   **Results**

99   ***Walking dynamics can be described in a reduced state space***

100   In order to ask whether cell behavior is governed by a finite state machine, we analyzed

101   the walking behavior of *Euplotes eurystomus* cells, [41], focusing on the simplest case of

102   uninterrupted, linear walking trajectories (Fig. 2a, b, Movie S1). Cells were placed onto

103   coverslips on which free, spontaneous walking behavior was observed by microscopy. A focal

104   plane at the cirrus-coverslip interface was chosen in order to clearly observe cirral dynamics

105   (Fig. 2a). The relative spatial positioning of cirri is highly stereotyped from cell to cell, allowing

106   us to give each of the 14 cirri an alphabetic label from a-n (Figure 2c). In each video frame (33

107   frames/s), the walking state of the cell was encoded as a 14-bit binary vector, with each bit

108   corresponding to a cirrus and receiving a value of "0" if the cirrus was in contact with the

109   coverslip and stationary and a "1" if the cirrus was in motion (instances of stationary cirri held

110   above the coverslip for a sustained period of time were not observed). The trajectories of 13 cells

111   were manually tracked and annotated for a total of 2343 time points. This quantitative analysis

112   revealed stepping-like cirral dynamics in that cirri tend to undergo rapid movements followed by

113   longer periods of quiescence (Fig. 2d). Cirral dynamics appeared to lack any obvious patterns

114   such as periodicity or repeating sequences of states (e.g. Fig. 2d), implying that the state

115   sequences are either stochastic, or generated by complex deterministic processes. This lack of

116   periodicity or fixed phase relationships between appendage movements is different from the gaits

117   of most animals or those reported for various flagellates [57–59]. Autocorrelation analysis confirmed

118   the observed lack of clear periodicity (Fig. S1).

119   Despite the apparent complexity of cirral dynamics, we suspected that discernable

120   structure might exist, which would allow us to obtain a reduced state space that accurately

121   described the dynamics as has proved successful in behavioral analysis of diverse living systems

122   [45,46,60–63]. We performed dimensionality reduction using non-negative matrix factorization

123   (NMF), and cross-validation by imputation [64,65] (see Methods and Fig. S2 for more details)

124   revealed the cirral states to be well-described in three dimensions (Fig. 2e-g). The components of

125   the NMF analysis correspond to distinct groups of cirri, and these groups constitute spatially

126   distinct partitions of cirri with respect to their positions on the cell body (Fig. 3h). The

127   dimensionality reduction of the gait state space arises at least in part from shared pairwise mutual

128   information between groups of cirri (Fig. 3i).

129   We next used the density-based spatial clustering of applications with noise (DBSCAN)

130   algorithm [66] to group our data into clusters in an unbiased fashion, with members of a given

131   cluster sharing similar patterns of cirral activity. Visual inspection in conjunction with silhouette

132   coefficient [67] (a metric of cluster cohesion and separation) analysis revealed that 32 clusters

133   accurately captured the visible structure in the reduced state space without overfitting (Fig. 2e-g,

134   see Methods and Fig. S3 for more details)[*]. These reduced gait states correspond to distinct

135   patterns of cirral activity (Fig. 2j). Taken together, our results reveal stereotypy in the

136   spatiotemporal patterns of cirral activity. The discrete, reduced state space, or set of gait states,

---

[*] The problem of determining the true number of clusters is an unresolved problem [89]. We have followed standard methods to determine cluster number but have found that our key results do not depend sensitively on the precise number of clusters identified (see SI and Fig. S7).

137   demonstrates that cells make use of a subset of the possible patterns of appendage movement

138   during walking locomotion.

Euplotes walks with a cyclic stochastic gait

140   In order to relate the gait states identified in our cluster analysis, we asked how changes

141   in the number of active cirri may relate to cell movement. Naively, one might expect that the

142   force associated with locomotion is roughly proportional to the number of moving appendages [68].

143   Alternatively, we velocity might inversely correlate with the net change in cirral activity, which

144   would be expected if stationary cirri were generating a pushing traction force as in crawling or

145   climbing animals [58,69] or if cirri execute a power stroke just before coming to rest as has been

146   suggested by Erra et al. [38]. At low Reynolds number, velocity should be proportional to the

147   difference between the net force generated by the cell and the opposing drag [44]. Examining cell

148   velocity versus the net change in cirral activity, however, showed that neither of these

149   expectations were in fact the case (Fig. 3a). Cell velocity was only weakly correlated with

150   number of active cirri (R^2=0.03), and instead, small to moderate positive and negative changes in

151   the number of active cirri corresponded to the largest cell velocities (Fig. 3a). We reasoned that

152   transitions between gait states must be important to driving the forward progression of walking

153   cells, and so sought to determine whether this active coordination might manifest in the observed

154   gait dynamics.

155   Analysis of the 1423 pairwise transitions in our dataset yielded the transition matrix

156   displayed in Fig. 3b (see Methods for more details). The presence of strongly unbalanced

157   transitions such as from gait state 3 to 17 versus 17 to 3 suggested broken detailed balance, and

158   indeed, a number of forward and reverse transitions were found to be significantly unbalanced by

159   the binomial test (see Methods). Entropy production rate has been used to quantify the degree of

160   broken detailed balance, or, similarly, the distance from equilibrium where the entropy

161   production rate will be zero [5]. Following the procedure detailed in [5], we obtain a lower bound

162   estimate for an entropy production rate of 0.4, similar to the value reported for strongly non-

163   equilibrium gait transitions observed in a flagellate [5]. Walking *Euplotes* cells, therefore, have a

164   strongly non-equilibrium gait despite lack of periodicity.

165   Only 322 of the 1024 possible types of transitions were observed to occur at least once,

166   and within this restricted set, only 173 occurred more than once (Fig. 3c). Also, we found that

167   relatively few transitions corresponded to substantial cellular movement (Fig. 3c). Crucially, the

168   presence of broken detailed balance revealed the existence of directed cycles of cirral activity

169   during locomotion. To get a better understanding of the nature of these cycles, we grouped

170   transitions into two categories: balanced transitions, which satisfy detailed balance, and

171   unbalanced transitions, which do not (see Methods for details). This partitioning allowed us to

172   separately investigate unbalanced, non-equilibrium-like and balanced, equilibrium-like

173   transitions (Fig. 3d, e). Significantly unbalanced transitions ($p < 0.05$ by binomial test) are among

174   the most frequent transitions, but only involve a small number of states (Fig. 3d). Of the 32 gait

175   states, we found that only states 2, 3, 4, 7, 17, and 27 were associated with unbalanced

176   transitions. We noted the presence of one complete cycle with unbalanced transitions following

177   2→3→17→2. We had expected that unbalanced transitions might be associated with a "power

178   stroke" in the sense of occurring simultaneously with cell movement, but in fact high cellular

179   velocities tended to be associated with balanced transitions (Fig. 3d, e). Additionally, we found

180   that, with the exception of transitions between states 1 and 2, transitions occurring at the highest

181   frequencies were unbalanced (Fig. 3d).

182   Notably, the most frequent balanced transitions were associated with transitions into and

183   out of gait state 1, a unique "rest state" which involves no cirral movement (Fig. 3c, d).

184   Furthermore, we found by computing the autocorrelation function for a binarized sequence of

185   each state that gait state 1 has the most predictable dynamics in terms of significant positive

186   autocorrelation in contrast to the lack of significant autocorrelation seen for other states (see Fig.

187   S4). Although transitions between states 1 and 2 are balanced, the most frequent transitions out

188   of state 2 are strongly biased toward transitions into state 3, from which other strongly biased

189   transitions also frequently occur, including the cycle of biased transitions mentioned in the

190   preceding paragraph. The presence of high frequency unbalanced transitions does not preclude

191   the existence of highly variable trajectories through state space. The picture of walking

192   trajectories that emerges is of stochastic excursions from gait state 1 into a stochastic cycle

193   involving a mix of balanced and unbalanced transitions, with the majority of cell movement

194   occurring during infrequent, equilibrium-like transitions. Biased transitions, occurring at

195   relatively high frequency from a subset of states, introduce temporal irreversibility to the gait of

196   *Euplotes* due to their strongly non-equilibrium character.

197   Finally, we checked whether gait state transition dynamics had the Markov property,

198   which entails that transition probabilities are determined completely by the present state, and that

199   previous dynamics contribute no additional predictive information [70,71]. Lack of past dependence

200   has led to Markov processes often being referred to as "memoryless" [72]. When we compared the

201   gait transition matrix (Fig. 3b) with a computed transition matrix over two timesteps, we

202   observed these matrices to be different from one another, which is inconsistent with the Markov

203   property (see Fig. S5). Finally, we applied a Billingsley test, a chi-squared measure for

204   Markovness [73], which revealed that the null hypothesis that the process was Markov could be

9

205 rejected (p=0.005). These analyses showed that *Euplotes* retains some "memory" of the prior

206 sequence of cirral movements during locomotion.

207 Taken together, our analysis revealed a mixture of unbalanced transitions arranged in

208 cycles and balanced transitions arranged as networks, for which we propose to apply the term

209 "cyclic stochastic gait". The cyclic stochastic gait of *Euplotes eurystomus* incorporates elements

210 of both stereotypy and variability in gait dynamics. Forward progress of the cell is not produced

211 merely by a physical ratchetting process driven by unpatterned fluctuations in cirral activity, nor

212 is it produced by a highly regular, deterministic process like a clock. It has been argued that

213 significant computation arises in physical systems exhibiting such a mix of stereotypy and

214 variability [49,74,75] in the sense that the time-evolution of the system is most compactly described

215 by the result of a computation involving state transitions, memory, and decision rules, rather than

216 by a periodic oscillation or a random coin flip.

217 While our analysis revealed a computational underpinning of gait, we sought to better

218 understand the functional organization of the dynamical patterns driving processive motion of

219 the cell. To do so, we first focused on the highest transition probabilities emanating from each

220 state. Transition probabilities were estimated as N_{ij}/N_i where N_{ij} is the number of transitions from

221 state i to state j, and N_i is the total number of transitions from state i. This allowed us to prune

222 away rare transitions in order to reveal the dominant structure of gait state transitions. Figure 3g

223 displays the pruned transition matrix as a heatmap. We found that relatively few states were the

224 recipients of the majority of high probability transitions, and many states received none. To more

225 clearly visualize the structure of transitions, we grouped together all gait states receiving no more

226 than one unique high probability transition with the idea being that state transitions into this

227 group show little bias in terms of source state, and within the group, transitions between states

10

228   exhibit low probability, time unbiased, equilibrium-like fluctuations. the majority of cell

229   movement was associated with transitions between states within this group. In contrast to the

230   "cloud" of states linked by low-probability, balanced transitions, nearly all of the states receiving

231   high probability transitions were either the three "cycle" states, or else fed cycle states with their

232   highest probability transitions, with the majority feeding gait state 17.

233   Focusing on the structure of transitions in this way allowed us to derive a simplified,

234   functional representation of stereotypy in gait dynamics as depicted in Fig. 3h. Although gait

235   state 1 is not the recipient of any individual high probability transitions, we identified it as the

236   unique "start" state from which cells initiate walking. Beginning with this start state, cells

237   transition with high probability to gait state 2, also one of the highest frequency transitions and

238   the first state in the 2→3→17→2 cycle of unbalanced transitions. From this first cycle state,

239   cells transition to gait state 3, the second cycle state, with highest probability and frequency and

240   then similarly on to gait state 17, the third cycle state. This sequence from the start state through

241   the cycle states corresponds to increasing amounts of cirral activity. Although the highest

242   probability transitions from the third cycle state to any single gait state return to the first and

243   second cycle state with equal probability and return to the first cycle state also being unbalanced,

244   cells in fact transition to the equilibrium "cloud" of motility-associated states with overall higher

245   probability. Return to the cycle states tend to occur through various moderately high probability

246   transitions from the motility state cloud or through intermediate states. In conjunction with this

247   set of transitions, we also noted unbalanced transitions stemming from the cycle states to the

248   motility state as well as intermediate state subsequently feeding the next cycle state.

249   Altogether, the picture that emerges of stereotypical gait dynamics is of biased transitions

250   through cycle states before relatively low probability, unbiased transitions associated with

11

251   substantial cell movement before returning to the start or cycle states and beginning the sequence

252   again. While this general sequence is repeated during walking, there is variability or apparent

253   stochasticity in the details of gait state transitions with increasingly variable transitions as any

254   given sequence progresses. We propose that the cycle states serve to establish configurations of

255   cirri necessary for cells to later transition to between states from which forward progress of the

256   cell is generated. Many state transitions along any instance of the stereotyped sequence are

257   unbiased, but biased, high probability transitions, presumably resulting from active cellular

258   control, give temporal irreversibility to the sequence.

259

260   ***The fiber system of Euplotes constrains models of gait coordination***

261   What physical machinery could embody this information processing required to generate

262   the stochastic cyclic state transitions seen during *Euplotes*' walking? We reasoned that there

263   must be some form of coupling or communication between cirri or feedback between gait state

264   and cirral dynamics. The role of the system of cytoskeletal fibers associated with cirri as conduits

265   of information between cirri during cellular locomotion, supported by microsurgical

266   experiments, has been a key hypothesized mechanism of gait coordination since the early 1900s

267   [76,77]. We wondered whether the structure of the cytoskeletal fiber system associated with cirri

268   (Fig. 4a) could give some insight into how cirri might be coordinated.

269   We sought to characterize and reconstruct in 3D the tubulin-based fiber system of

270   *Euplotes* associated with cirri and lying just beneath the cell cortex [39,42,43]. Upon inspection of

271   our confocal reconstructions of SiR-tubulin labeled cells (Fig. 4a, Fig. S6), we noted the

272   presence of two morphologically distinct classes of fibers, one thicker, linear class and the other

273   more filamentous and less linear, consistent with previous observations (Fig. 1c, [39,42,43]).

274   Additionally, we observed a group of thick linear fibers associated with some of the frontal cirri,

275   which to our knowledge has not been previously reported. Fibers emanate from the base of all

276   cirri, appear to intersect one another, and also connect to the cortex of the cell at various points

277   (Fig. 3g). Some cirri were found to be associated with only thick fibers while others have both or

278   only thin fibers. Based on apparent fiber intersections and convergences, we found the fiber

279   system to constitute a continuous network between all cirri, with the fibers associated with the

280   base of each cirrus intersecting the fiber system associated with at least one other cirrus (Fig. 4a,

281   b). Contrary to the long-standing standing hypothesis from the literature [76], the functional

282   modules (groups of co-varying cirri) identified in our dynamical analysis were not exclusively

283   linked by dense fiber intersections [39,43,76] (Fig. 4a, b). In fact, connections between cirri are not

284   generally associated with any statistically significant difference in mutual information (defined

285   in terms of the information that the activation state of one cirrus has concerning the other)

286   compared to unlinked pairs of cirri (p=0.14 by Wilcoxon rank sum test, Fig. 4C). However,

287   information flow became apparent when fiber-fiber links are grouped by type (i.e. thick to thick

288   fiber, thick to thin fiber, or thin to thin fiber). We found that pairs of cirri associated with only

289   thick fiber to thick fiber and only thin fiber to thin fiber links have increased mutual information

290   compared to those without links (Fig. S6). Interestingly, we found that cirri nearby one another

291   and connected by fibers to similar regions of the cell cortex shared the most mutual information

292   (Fig. 2i, 3c, 4d, e), suggesting that if the fibers play a role in cirral coordination, coupling may

293   also be mediated by mechanisms at the cirrus and fiber-cortex interface. Cirri d, e, h, i, for

294   example, share very little mutual information with any of the other cirri, and fibers emanating

295   from the base of these cirri contact the cell cortex and other fibers at various unique points. Cirri

296   g and f, on the other hand, which share more mutual information than any other pair, are

297   associated with both thick and thin fibers terminating at similar regions of the cell cortex. Indeed,

298   both distances between pairs of cirri and cross nearest neighbor distance (a measure of similarity

299   between discrete spatial distributions) between paired sets of cirrus-cortex contact points show

300   significant Spearman correlations to mutual information (-0.49, $p < 0.001$ and -0.62, $p < 0.001$

301   respectively) (Fig. 4d, e). These correlations indicate that mutual information between pairs of

302   cirri tends to increase with proximity and also tends to increase with similarity between fiber-

303   cortex contact locations, so the cirri with the highest mutual information are those that are close

304   together with similar fiber-cortex connections (Fig. 4d-f). Together, these observations strongly

305   support a mechanical coordination mechanism in which microtubule bundles allow groups of

306   cirri to influence successive behavior of other groups of cilia.

307

308   **Discussion**

309   In order to meet the challenge of accounting for the emergence of apparently

310   sophisticated cellular behavior, we conceptualized the cell as a finite state machine.

311   Traditionally, studies of computational processes performed by cells have tended to focused on

312   combinatorial logic, where the output of a computational process depends only on the current

313   input, performed by networks of molecules in bacterial cells [25,32–34]. We have focused on

314   sequential logic, where outputs depend on the system state as well, an equally important aspect

315   of the theory of computation with notable yet less developed representation in studies of cellular

316   and sub-cellular dynamics [16,20,78]. Behavior of eukaryotes has frequently been observed to

317   involve stereotyped transitions between dynamical states, and our results suggest that automata

318   theory, which includes finite state machine models and necessarily involves sequential logic,

319   may be particularly well-suited to studying the behavior of eukaryotic cells. Our approach,

14

320 relying on dimensionality reduction to identify dynamical states, revealed modularity in cellular

321 dynamics associated with structural modularity of the cell (Fig. 2, 4) in addition to cyclic

322 patterns of sequential dynamical activity (Fig. 3).

323   Walking locomotion in *Euplotes* represents a departure from many of the best studied

324 appendage-based locomotor systems. For example, limbed locomotion in animals tends to

325 proceed by highly stereotyped, determinate patterns of activity [57,58], and many small, aquatic

326 animals exhibit periodic movements of appendages, often cilia, during locomotion [7,59,79]. Many

327 forms of unicellular locomotion involve such dynamics as well including in sperm cells [80],

328 diverse flagellates with various numbers of flagella [59], and ciliates [59,81,82]. Even in cases where

329 cellular locomotion involves fundamentally stochastic dynamics such as in run-and-tumble

330 motility in *E. coli* [12] or analogous behaviors observed in protists [11,83–85], motility can be described

331 by equilibrium processes [5], in contrast to the non-equilibrium character of the gait of *Euplotes*.

332 There are examples, however, of locomotor dynamics in both animals and unicellular organisms

333 that are reminiscent of those we have observed in *Euplotes*. Gait switching in kangaroo rats has

334 been shown to have a stochastic, non-equilibrium character, perhaps to facilitate predator

335 avoidance by being difficult to predict [52]. Most saliently, gait switching in an octoflagellate [5] and

336 motility dynamics in cultured mammalian cells [45] have been shown to exhibit broken detailed

337 balance. We propose that broken detailed balance in the gait of *Euplotes* indicates active

338 coordination of motility processes. Here, broken detailed balance in gait state transitions revealed

339 cyclic activity, characterized by transitions into and out of a resting state with a mixture of

340 stereotypy and variability in the intervening steps, in the gait of a single cell (Fig. 3, 4). To

341 explain how these dynamics give rise to directed walking, we propose a mechanism in which

342 biased, actively controlled cyclic transitions serve to establish strain, effectively storing stress, in

343   certain cirri, and the spontaneous release of these cirri from the substrate, during a series of

344   unbiased gait state transitions, allows the cell to move forward. The cloud of unbiased transitions

345   associated with substantial cellular movement is consistent with the fact that the details of the

346   precise order in which the strained cirri are released does not matter for generating motility.

347   Consistent with this mechanism, inspection of videos revealed that substantial cell movement

348   appears to be correlated with the movement of notably bent inactive cirri. Return to the cycle

349   states then are necessary to establish this process anew by winding up the system for continued

350   cell movement. This mechanism and these patterns of cirral activity are consistent with previous

351   observations of cyclic velocity fluctuations in the trajectories of walking *Euplotes* [37]. We argue

352   that subcellular processes must be involved in actively coordinating cirri in order to accomplish

353   the observed stereotypy in biased sequential activity. Our analysis of the tubulin-based

354   cytoskeletal fiber system is consistent with its role in mechanically mediating communication

355   among cirri and between cirri and cell cortex (Fig. 4). Thus, by combining information

356   processing to properly dictate patterns of cirral activity and the mechanical actions of cirral

357   movement, walking *Euplotes* embodies the sequential computation of a finite state machine.

358   Among the domains of life, eukaryotes uniquely display remarkable complexity and

359   diversity in cellular behavior [86]. Our approach, grounded in finite state-machine analysis, has

360   revealed modularity and stereotypy underlying a complex cellular behavior, implicating a

361   machine-like process. Our results suggest that integrating approaches from theoretical computer

362   science, non-equilibrium statistical physics, and cell biology stands to shed light on the

363   regulation of cellular behavior in eukaryotes more broadly. By revealing principles of cellular

364   behavior, the line of research established here stands to advance our ability to predict and even

365   one day engineer cellular behavior across diverse eukaryotic systems.

16

366

References

1. Jennings, H. S. *Behavior of the lower organisms. Behavior of the lower organisms.* (Columbia University Press, 1906). doi:10.1037/10817-000

2. Reid, C. R. *et al.* Decision-making without a brain: how an amoeboid organism solves the two-armed bandit. *J. R. Soc. Interface* **13**, 20160030 (2016).

3. Coyle, S. M., Flaum, E. M., Li, H., Krishnamurthy, D. & Prakash, M. Coupled Active Systems Encode an Emergent Hunting Behavior in the Unicellular Predator Lacrymaria olor. *Curr. Biol.* **29**, 3838-3850.e3 (2019).

4. Battle, C. *et al.* Broken detailed balance at mesoscopic scales in active biological systems. *Science* **352**, 604–7 (2016).

5. Wan, K. Y. & Goldstein, R. E. Time Irreversibility and Criticality in the Motility of a Flagellate Microorganism. *Phys. Rev. Lett.* **121**, 58103 (2018).

6. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7894–7899 (2016).

7. Wan, K. Y. Coordination of eukaryotic cilia and flagella. *Essays Biochem.* **62**, 829–838 (2018).

8. Nomura, M. & Ishida, K. I. Fine-structural Observations on Siliceous Scale Production and Shell Assembly in the Testate Amoeba Paulinella chromatophora. *Protist* **167**, 303–318 (2016).

9. Bennett, R. R. & Golestanian, R. A steering mechanism for phototaxis in Chlamydomonas. *J. R. Soc. Interface* **12**, 20141164–20141164 (2015).

10. Machemer, H. The swimming cell and its world: Structures and mechanisms of orientation

389   in protists. *Eur. J. Protistol.* **37**, 3–14 (2001).

390   11.   Almagor, M., Ron, A. & Bar-Tana, J. Chemotaxis in Tetrahymena thermophila. *Cell*

391   *Motil.* **1**, 261–268 (1981).

392   12.   Berg, H. C. & Brown, D. A. Chemotaxis in Escherichia coli analysed by three-

393   dimensional tracking. *Nature* **239**, 500–504 (1972).

394   13.   Nakagaki, T., Yamada, H. & Tóth, Á. Maze-solving by an amoeboid organism. *Nature*

395   **407**, 470–470 (2000).

396   14.   Tweedy, L. *et al.* Seeing around corners: Cells solve mazes and respond at a distance

397   using attractant breakdown. *Science (80-.).* **369**, (2020).

398   15.   Applewhite, P. B. Learning in Protozoa. in *Biochemistry and Physiology of Protozoa* 341–

399   355 (Elsevier, 1979). doi:10.1016/b978-0-12-444601-4.50018-7

400   16.   Dexter, J. P., Prabakaran, S. & Gunawardena, J. A Complex Hierarchy of Avoidance

401   Behaviors in a Single-Cell Eukaryote. *Curr. Biol.* **29**, 4323-4329.e2 (2019).

402   17.   Jennings, H. S. Studies on reactions to stimuli in unicellular organisms. IX—On the

403   behavior of fixed infusoria (Stentor and Vorticella), with special reference to the

404   modifiability of protozoan reactions. *Am. J. Physiol. Content* **8**, 23–60 (1902).

405   18.   Tang, S. K. Y. & Marshall, W. F. Cell learning. *Current Biology* **28**, R1180–R1184

406   (2018).

407   19.   Lyon, P. The cognitive cell: Bacterial behavior reconsidered. *Frontiers in Microbiology* **6**,

408   (2015).

409   20.   Marshall, W. F. Cellular Cognition: Sequential Logic in a Giant Protist. *Current Biology*

410   **29**, R1303–R1305 (2019).

411   21.   Greenfield, D. *et al.* Self-Organization of the Escherichia coli Chemotaxis Network

412   Imaged with Super-Resolution Light Microscopy. *PLoS Biol.* **7**, e1000137 (2009).

413   22.   Wan, K. Y. & Goldstein, R. E. Coordinated beating of algal flagella is mediated by basal

414   coupling. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E2784-93 (2016).

415   23.   Höfer, T., Sherratt, J. A. & Maini, P. K. Dictyostelium discoideum : cellular self-

416   organization in an excitable biological medium. *Proc. R. Soc. London. Ser. B Biol. Sci.*

417   **259**, 249–257 (1995).

418   24.   Kantsler, V., Dunkel, J., Polin, M. & Goldstein, R. E. Ciliary contact interactions

419   dominate surface scattering of swimming eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **110**,

420   1187–1192 (2013).

421   25.   Bray, D. *Wetware*. (Yale University Press, 2009).

422   26.   Pfeifer, R., Lungarella, M. & Iida, F. Self-organization, embodiment, and biologically

423   inspired robotics. *Science* **318**, 1088–1093 (2007).

424   27.   Regev, A. & Shapiro, E. Cellular abstractions: Cells as computation. *Nature* **419**, 343–343

425   (2002).

426   28.   Mitchell, M. Biological Computation. *Comput. J.* **55**, 852–855 (2012).

427   29.   Navlakha, S. & Bar-Joseph, Z. Algorithms in nature: the convergence of systems biology

428   and computational thinking. *Mol. Syst. Biol.* **7**, 546 (2011).

429   30.   Savin, T. *et al.* On the growth and form of the gut. *Nature* **476**, 57–62 (2011).

430   31.   Mehta, P. & Schwab, D. J. Energetic costs of cellular computation. *Proc. Natl. Acad. Sci.*

431   *U. S. A.* **109**, (2012).

432   32.   Berg, H. C. & Purcell, E. M. Physics of chemoreception. *Biophys. J.* **20**, 193–219 (1977).

433   33.   Bialek, W. & Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl. Acad.*

434   *Sci. U. S. A.* **102**, 10040–5 (2005).

435   34.   Mehta, P. & Schwab, D. J. Energetic costs of cellular computation. *Proc. Natl. Acad. Sci.*

436   *U. S. A.* **109**, 17978–17982 (2012).

437   35.   Andrews, L. B., Nielsen, A. A. K. & Voigt, C. A. Cellular checkpoint control using

438   programmable sequential logic. *Science (80-.).* **361**, (2018).

439   36.   Gilbert, J. J. Jumping behavior in the oligotrich ciliates Strobilidium velox and Halteria

440   grandinella, and its significance as a defense against rotifer predators. *Microb. Ecol.* **27**,

441   189–200 (1994).

442   37.   Lueken, W., Ricci, N. & Krüppel, T. Rhythmic spontaneous depolarizations determine a

443   slow-and-fast rhythm in walking of the marine hypotrich Euplotes vannus. *Eur. J.*

444   *Protistol.* **32**, 47–54 (1996).

445   38.   Erra, F., Iervasi, A., Ricci, N. & Banchetti, R. Movement of the cirri during the creeping

446   of Euplotes crassus (Ciliata, Hypotrichida). *Can. J. Zool.* **79**, 1353–1362 (2001).

447   39.   Taylor, C. V. The neuromotor system of Euplotes. *Exp. Biol. Med.* **16**, 101–103 (1919).

448   40.   Syberg-Olsen, M. J. *et al.* Biogeography and character evolution of the ciliate genus

449   Euplotes (Spirotrichea, Euplotia), with description of Euplotes curdsi sp. nov. *PLoS One*

450   **11**, (2016).

451   41.   Ricci, N., Giannetti, R. & Miceli, C. The ethogram of Euplotes crassus (ciliata,

452   hypotrichida): I. The wild type. *Eur. J. Protistol.* **23**, 129–140 (1988).

453   42.   Li, Y., Shi, L. & Gu, F. Comparison of ciliature microtubule organelles in three

454   hypotrichous ciliate species*. *Chinese J. Oceanol. Limnol.* **28**, 443–450 (2010).

455   43.   Fleury, A. Dynamics of the cytoskeleton during morphogenesis in the ciliate Euplotes I.

456   Basal bodies related microtubular system. *Eur. J. Protistol.* **27**, 99–114 (1991).

457   44.   Purcell, E. M. Life at low Reynolds number. *Am. J. Phys.* **45**, 3–11 (1977).

20

458   45.   Kimmel, J. C., Chang, A. Y., Brack, A. S. & Marshall, W. F. Inferring cell state by

459   quantitative motility analysis reveals a dynamic state system and broken detailed balance.

460   *PLoS Comput. Biol.* **14**, e1005927 (2018).

461   46.   Chang, A. Y. & Marshall, W. F. Dynamics of living cells in a cytomorphological state

462   space. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21556–21562 (2019).

463   47.   Boltzmann, L. Weirere Studien uber das warmegleich-gewich unter gasmolekulen.

464   *Sitzungsberichte der Akad. der Wissenschaften, Vienna, part II* **66**, 275–370 (1872).

465   48.   Alberts, B. *et al. Molecular biology of the cell*. (Garland Science, 2002).

466   49.   Shalizi, C. R. & Crutchfield, J. P. Computational Mechanics: Pattern and Prediction,

467   Structure and Simplicity. *J. Stat. Phys.* **104**, 817–879 (2001).

468   50.   Crutchfield, J. P. The calculi of emergence: computation, dynamics and induction. *Phys.*

469   *D Nonlinear Phenom.* **75**, 11–54 (1994).

470   51.   Hopcroft, J. E., Motwani, R. & Ullman, J. D. *Introduction to Automata Theory,*

471   *Languages, and Computation*. (Pearson Education, 2008).

472   52.   Moore, T. Y., Cooper, K. L., Biewener, A. A. & Vasudevan, R. Unpredictability of escape

473   trajectory explains predator evasion ability and microhabitat preference of desert rodents.

474   doi:10.1038/s41467-017-00373-2

475   53.   Goldstein, R. E. Green Algae as Model Organisms for Biological Fluid Dynamics. *Annu.*

476   *Rev. Fluid Mech.* **47**, 343–375 (2015).

477   54.   Simon, M. & Plattner, H. Unicellular Eukaryotes as Models in Cell and Molecular

478   Biology: Critical Appraisal of Their Past and Future Value. in *International Review of*

479   *Cell and Molecular Biology* **309**, 141–198 (Elsevier Inc., 2014).

480   55.   Durham, W. M. *et al.* Turbulence drives microscale patches of motile phytoplankton. *Nat.*

481   *Commun.* **4**, (2013).

482   56.   Mogilner, A. & Oster, G. Cell motility driven by actin polymerization. *Biophys. J.* **71**,

483   3030–3045 (1996).

484   57.   Collins, J. J. & Stewart, I. N. Coupled nonlinear oscillators and the symmetries of animal

485   gaits. *J. Nonlinear Sci.* **3**, 349–392 (1993).

486   58.   McNeill Alexander, R. *Principles of Animal Locomotion*. (Princeton University Press,

487   2003).

488   59.   Wan, K. Y. & Jékely, G. On the unity and diversity of cilia. *Philos. Trans. R. Soc. Lond.*

489   *B. Biol. Sci.* **375**, 20190148 (2020).

490   60.   Ahamed, T., Costa, A. C. & Stephens, G. J. Capturing the Continuous Complexity of

491   Behavior in C. elegans. doi:10.1101/827535

492   61.   Stephens, G. J., Johnson-Kerner, B., Bialek, W. & Ryu, W. S. Dimensionality and

493   Dynamics in the Behavior of C. elegans. *PLoS Comput. Biol.* **4**, e1000028 (2008).

494   62.   Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped

495   behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672 (2014).

496   63.   Stephens, G. J., Osborne, L. C. & Bialek, W. Searching for simplicity: Approaches to the

497   analysis of neurons and behavior. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15565–15571

498   (2010).

499   64.   Kanagal, B. & Sindhwani, V. Rank Selection in Low-rank Matrix Approximations: A

500   Study of Cross-Validation for NMFs. in *Advances in Neural Information Processing*

501   *Systems* **23**, (2010).

502   65.   Owen, A. B. & Perry, P. O. Bi-Cross-Validation of the SVD and the Nonnegative Matrix

503   Factorization. *Source Ann. Appl. Stat.* **3**, 564–594 (2009).

22

504 66. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. *A Density-Based Algorithm for Discovering*
505 *Clusters in Large Spatial Databases with Noise*. (1996).

506 67. Kauffman, L. & Rousseeuw, P. J. *Finding Groups in Data: and introduction to cluster*
507 *analysis*. (John Wiley & Sons, Inc., 1990). doi:10.1002/9780470316801

508 68. Kirkegaard, J. B., Marron, A. O. & Goldstein, R. E. Motility of Colonial
509 Choanoflagellates and the Statistics of Aggregate Random Walkers. *Phys. Rev. Lett.* **116**,
510 038102 (2016).

511 69. M Cartmill. Climbing. in *Functional Vertebrate Morphology* 73–88 (Harvard University
512 Press, 1985).

513 70. Markov, A. A. Extension of the Limit Theorems of Probability Theory to a Sum of
514 Variables Connected in a Chain. *Notes Imp. Acad. Sci. St. Petersbg. VIII Ser. Physio-*
515 *Mathematical Coll.* **22**, (1907).

516 71. Howard, R. A. *Dynamic and Probabilistic Systems, Volume 1: Markov models*. (Dover
517 Publications, 1971).

518 72. Gagniuc, P. *Markov Chains: From Theory to Implementation and Experimentation*.
519 (Wiley, 2017).

520 73. Billingsley, P. *Statistical Methods in Markov Chains*. *The Annals of Mathematical*
521 *Statistics* **32**, 12–40 (Institute of Mathematical Statistics, 1961).

522 74. Crutchfield, J. P. & Young, K. Computation at the Onset of Chaos. in *Complexity,*
523 *Entropy, and Physics of Information* (ed. Zurek, W.) (Addison-Wesley, 1989).

524 75. Crutchfield, J. P. & Young, K. Inferring statistical complexity. *Phys. Rev. Lett.* **63**, 105–
525 108 (1989).

526 76. Taylor, C. V. Demonstration of the function of the neuromotor apparatus in Euplotes by

527   the method of microdissection. *Naturwissenschaften* **9**, 910 (1921).

528   77.   Okajima, A. & Kinosita, H. Ciliary activity and coordination in Euplotes eurystomus-I.

529   Effect of microdissection of neuromotor fibres. *Comp. Biochem. Physiol.* **19**, 115–131

530   (1966).

531   78.   Bar-Ziv, R., Tlusty, T. & Libchaber, A. Protein-DNA computation by stochastic assembly

532   cascade. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11589–11592 (2002).

533   79.   Guasto, J. S., Rusconi, R. & Stocker, R. Fluid Mechanics of Planktonic Microorganisms.

534   *Annu. Rev. Fluid Mech.* **44**, 373–400 (2012).

535   80.   Goldstein, R. E., Polin, M. & Tuval, I. Noise and Synchronization in Pairs of Beating

536   Eukaryotic Flagella. *Phys. Rev. Lett.* **103**, 168103 (2009).

537   81.   Párducz, B. Ciliary Movement and Coordination in Ciliates. *Int. Rev. Cytol.* **21**, 91–128

538   (1967).

539   82.   Funfak, A. *et al.* Paramecium swimming and ciliary beating patterns: A study on four

540   RNA interference mutations. *Integr. Biol. (United Kingdom)* **7**, 90–100 (2015).

541   83.   Kirkegaard, J. B., Bouillant, A., Marron, A. O., Leptos, K. C. & Goldstein, R. E.

542   Aerotaxis in the closest relatives of animals. *Elife* **5**, e18109 (2016).

543   84.   Miño, G. L., Koehl, M. A. R., King, N. & Stocker, R. Finding patches in a heterogeneous

544   aquatic environment: pH-taxis by the dispersal stage of choanoflagellates. *Limnol.*

545   *Oceanogr. Lett.* **2**, 37–46 (2017).

546   85.   Polin, M., Tuval, I., Drescher, K., Gollub, J. P. & Goldstein, R. E. Chlamydomonas

547   Swims with Two 'Gears' in a Eukaryotic Version of Run-and-Tumble Locomotion.

548   *Science (80-.).* **325**, 487–490 (2009).

549   86.   Keeling, P. J. Combining morphology, behaviour and genomics to understand the

550   evolution and ecology of microbial eukaryotes. *Philos. Trans. R. Soc. B Biol. Sci.* **374**,

551   (2019).

552   87.   Kahl, A. *Urtiere oder Protozoa.* (Fischer, 1935).

553   88.   Yocom, H. B. The neuromotor apparatus of Euplotes patella. (University of California,

554   Berkeley, 1917).

555   89.   Kawamoto, T. & Kabashima, Y. Cross-validation estimate of the number of clusters in a

556   network. *Sci. Rep.* **7**, 1–17 (2017).

557

558

559 **Figures**
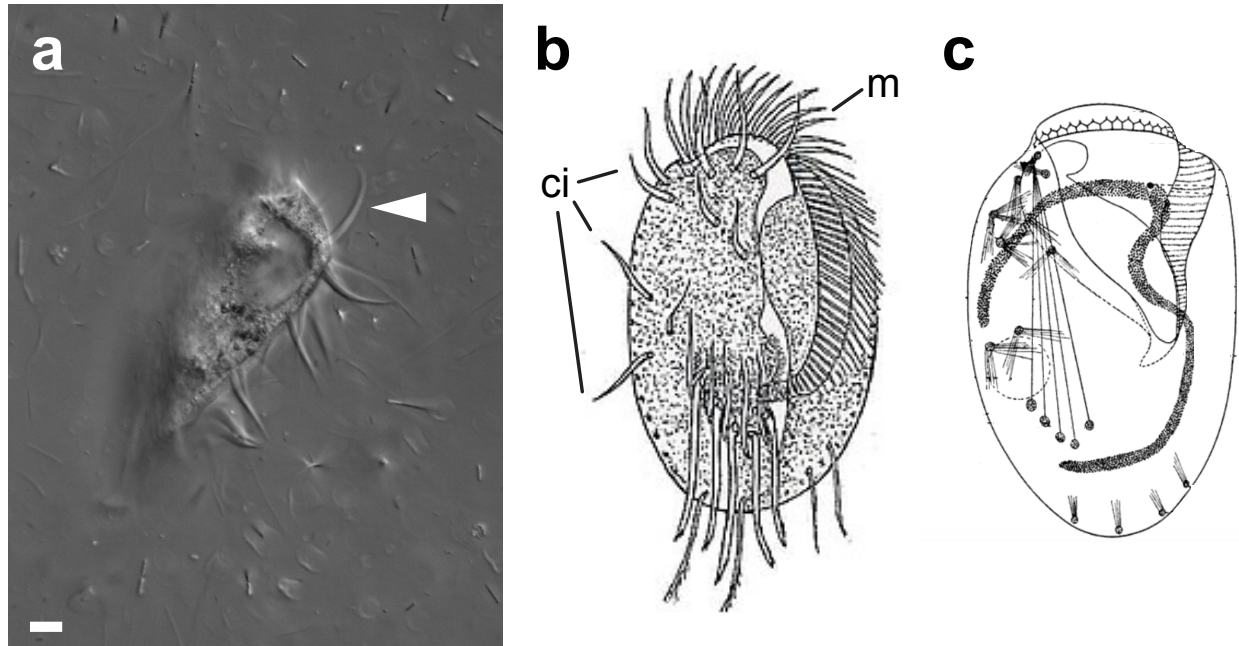
560

561   **Figure 1.** *Euplotes* exhibits highly polarized, complex cellular architecture and walks

562 across surfaces using microtubule-based organelles called cirri, some of which are physically

563 linked. Scale bar is 10 μm. **a**, A single *Euplotes eurystomus* cell in profile displays its ventral

564 cirri, which are used for walking locomotion across surfaces (arrowhead indicates a single cirrus

565 stretching out from the cell). **b**, A drawing of a *Euplotes* cell, viewed from the ventral surface,

566 highlighting the complex, asymmetric structure of the cell. Notable features include the cirri (ci)

567 and the membranellar band (m), wrapping from the top of the cell to the center, which is used to

568 generate a feeding current to draw in prey items. Drawing adapted and obtained from Wikimedia

569 Commons, from original source [87]. **c**, A drawing of a *Euplotes* cell, highlighting the fiber system

570 associated with the cirri, historically referred to as the neuromotor apparatus. Drawing adapted

571 from [88].

26

572
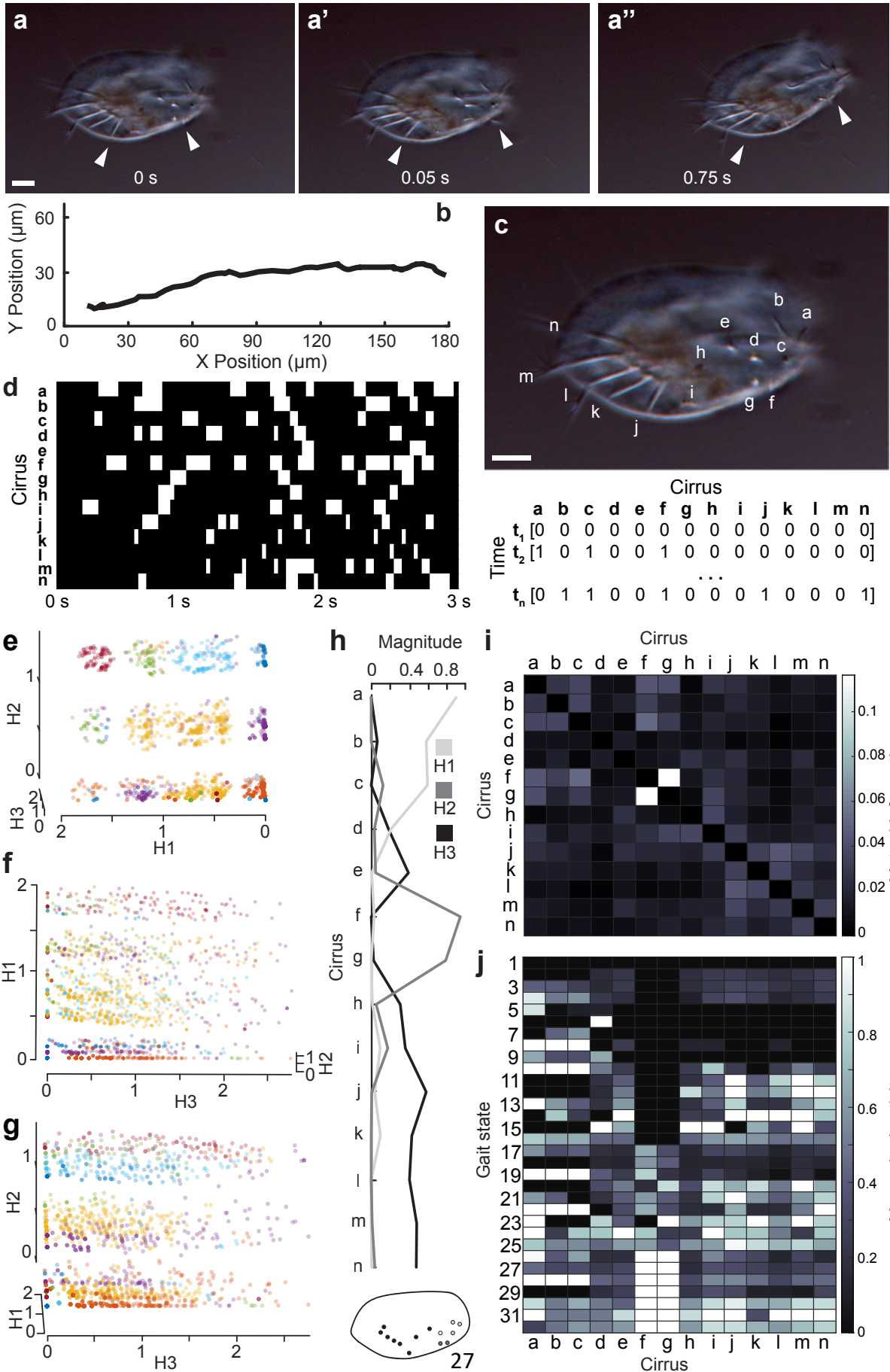
573   **Figure 2.** The gait of *Euplotes* can be described in a discrete, reduced state space with gait

574   states corresponding to identifiable patterns of cirral activity. **a-a''**, The movements of cirri

575   during walking locomotion are clearly visible by brightfield microscopy by focusing on a plane

576   at the surface of the coverslip on which cells are walking. Three snapshots depict different time

577   points during a single walking trajectory, and white arrowheads indicate cirri. In the panels from

578   left to right, the cirrus indicated by the arrowhead on the left is stationary, stationary, and then

579   moving, and the cirrus indicated by the arrowhead on the right is stationary, moving, and then

580   stationary. Scale bar is 15 μm. **b**, The trajectory of a cell during a single recorded trajectory as

581   the cell walked across a coverslip from left to right. The cell position was manually tracked in

582   each frame. **c**, The scheme for encoding cirral dynamics during walking involved labeling each

583   of the 14 distinguishable ventral cirri (a-n), and recording cirral activity in each frame,

584   corresponding to timepoints $(t_1,...,t_n)$, of recordings of walking cells as a 14-bit binary vector.

585   Each entry in each vector is given a value of either 0 if the cirrus is not moving and in contact

586   with the coverslip or 1 if the cirrus is moving. Scale bar is 15 μm. **d**, Representative visualization

587   of cirral dynamics for a single trajectory of a walking cell. These dynamics correspond to the

588   walking trajectory in **b**. Each row corresponds to a cirrus and each column is a single video

589   frame. White denotes cirral activity, a value of 1, in the vector encoding of dynamics from **c**.

590   Note the dynamical complexity and discrete, stepping-like nature of cirral movements. **e-g**,

591   Three roughly orthogonal views of a plot displays the structure of all recorded cirral dynamics

592   encoded as in Figure 2C from 13 cells over 2343 timepoints in a reduced state space obtained by

593   non-negative matrix factorization (NMF). Axes correspond to the components of the NMF (H1,

594   H2, H3), and each point is a single timepoint. Randomized colors highlight the 32 clusters

595   identified using the density-based spatial clustering of applications with noise (DBSCAN)

596   algorithm [66]. We refer to these clusters as gait states, and they correspond to unique

597   configurations of cirral activity during walking locomotion (see panel F). **h**, Plot of the

598   magnitudes associated with each cirrus corresponding to the components of the NMF of cirral

599   dynamics shows distinct contributions from spatially distinct groups of cirri. Component H1, for

600   example, is associated with activity in cirri a, b, and c. The tracing of a cell including the position

601   of cirri has the same color map as the plot above and shows the grouping of the cirri

602   corresponding to each component. **i**, A heatmap of mutual information between all pairs of cirri

603   shows that correlations in cirral activity correspond to the NMF components displayed in **d**. For

604   example, cirri a, b, and c share mutual information with one another and are the cirri contributing

605   to component H1. **j**, A heatmap representation of the cirral activity associated with each of the 32

606   gait states. Values for each cirrus are the mean over all instances of the gait state. Note that each

607   gait state has a unique signature of cirral activity.

608
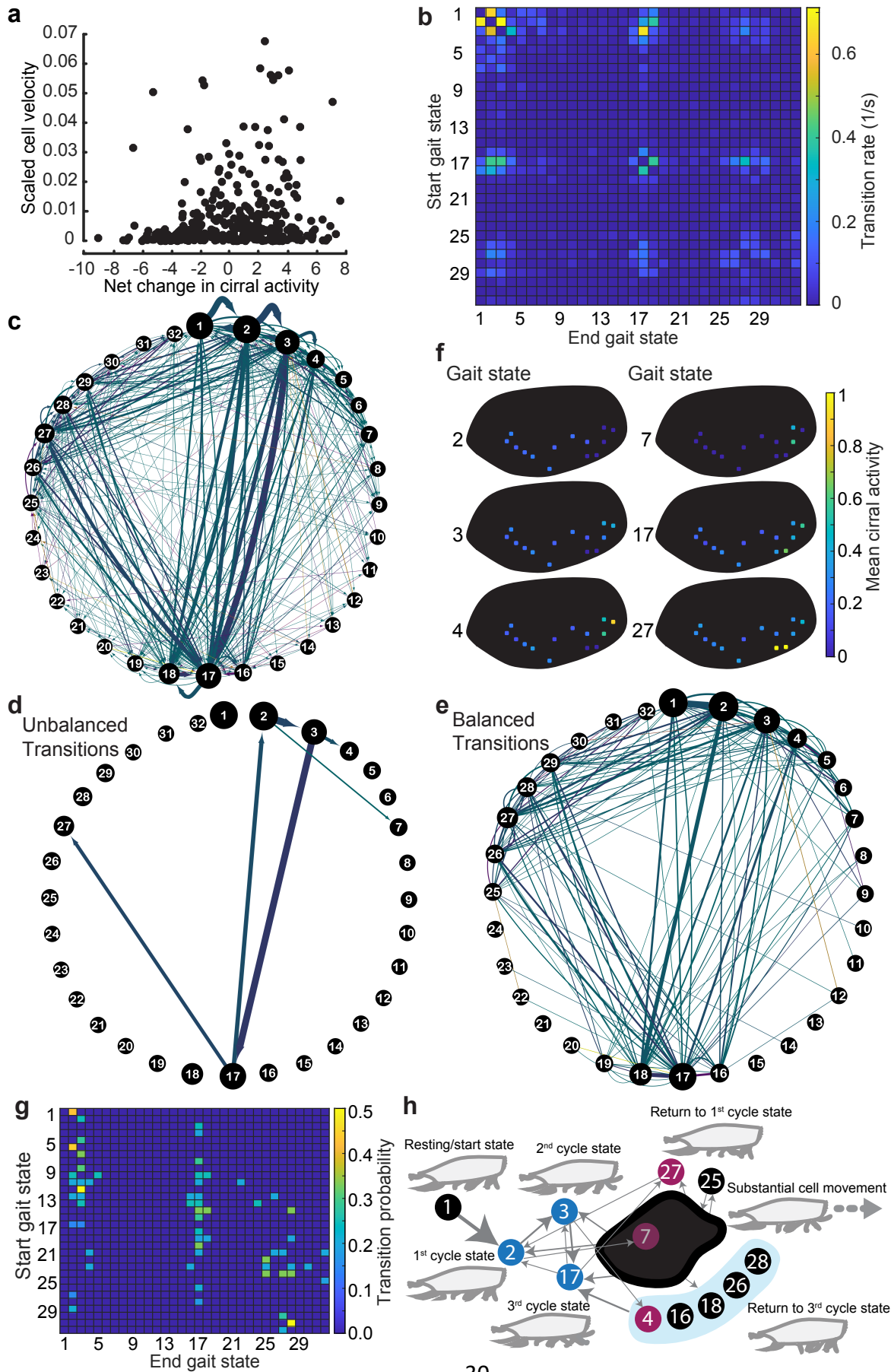
610   **Figure 3.** *Euplotes* walks with a cyclic stochastic gait exhibiting broken detailed balance,

611   stereotypy, and state machine-like dynamics. All data is pooled from the walking trajectories of

612   13 different cells over 2343 timepoints and 1423 pairwise gait state transitions. **a**, A plot of the

613   mean net change in cirral activity versus the net scaled cell velocity associated with all

614   transitions between the 32 gait states identified in Fig. 3 shows that the change in number of

615   active cirri is not strongly correlated with cell velocity (R^2=0.03). Cell velocities were obtained

616   from manually tracked walking trajectories and then scaled by dividing frame to frame

617   displacements for each trajectory by the length of the cell being tracked and also dividing by the

618   average frequency of cirral inactivity. Scaling provided a non-dimensional velocity scaled by

619   natural units of the system. Note that at low Reynolds number, velocity should be proportional to

620   force [44], so this plot also reflects the net walking force generated by the cell. Net change in cirral

621   activity was computed using the data presented in Fig. 3F. Note that the largest velocities are

622   associated with small negative and small to moderate positive net changes in cirral activity. **b**,

623   The transition matrix of all gait state transitions, with rows representing the starting state and

624   columns indicating the ending state, exhibits broken detailed balance. Rates were estimated by

625   dividing the total number of observed transitions between each state pair and dividing by the

626   total time observed. Under detailed balance or equilibrium conditions, transitions from one state

627   to another should be balanced by reverse transitions. Lack of this kind of reversibility, as seen by

628   the lack of symmetry of the heatmap across the diagonal, indicates broken detailed balance and

629   non-equilibrium dynamics. **c**, A directed graph representation of all gait state transitions. Nodes

630   correspond to the 32 gait states, and node sizes are scaled by the proportion of total time cells

631   spent in each state. Directed edges are represented by arrows between nodes and signify state

632   transitions. The size of the arrows is scaled by transition rates as in **b**. Edge color represents

31

633   scaled cell velocity as in **a**, with cool colors (more cyan) representing lower velocity, and

634   warmer colors (more magenta) representing higher velocity. **d**, A subset of transitions visualized

635   as in **c** shows the restricted and relatively high frequency nature of unbalanced, non-equilibrium-

636   like transitions. Only transitions that were observed to happen more than one time and exhibiting

637   a significant difference between forward and reverse transitions (p<0.05 by binomial test, see

638   Methods for more details) are displayed. **e**, A subset of transitions, similarly to panel E, except

639   that only the balanced transitions, lacking a significant difference between forward and reverse

640   transitions (p<0.05 by binomial test) are displayed, also show a complex and widespread

641   structure, this time of balanced, equilibrium transitions. Note that the majority of transitions

642   associated with high cell velocity involve equilibrium-like dynamics. **f,** Examples illustrating the

643   spatial organization of cirral activity corresponding to gait states. Some states, such as 7,

644   correspond to activity in spatially discrete groups of cirri, while others, such as 17, correspond to

645   cirral activity across the cell. The gait states displayed here are those involved in unbalanced

646   transitions. **g**, A heatmap of transition probabilities between states, showing only the most

647   probable transitions from a given state with all others set to zero, shows distinct structure. In

648   cases were multiple state transitions from a state were tied for the highest probability, all of these

649   transitions are displayed. Fewer than half of the total states are recipients of multiple high

650   probability transitions, and many states are the recipients of no high probability transitions. **h**, A

651   representation of functional states and transitions between them highlights the machine-like

652   nature of the gait of *Euplotes*. Gait states are represented as circles with numerical labels. Blue

653   circles represent states that are both recipients and sources of unbalanced transitions as identified

654   in **d** and constitute the three cycle states. Red circles represent states that are recipients but not

655   sources of unbalanced transitions as identified in **d**. Black circles correspond to gait states that

656   are associated only with balanced transitions as in **e**. States receiving no more than one unique

657   high probability transitions from states with only a single highest as identified in **g** were grouped

658   together into a compound state represented by the dark gray blob. The blue background behind

659   states 4, 16, 18, 26, and 28 indicates that these states all share the same highest probability

660   transitions between states identified in this panel, and thus, the group constitutes a single

661   compound functional state. Arrows represent the highest probability transitions between the

662   states, including compound states composed of multiple gait states as identified in Fig. 2 and 3

663   (dark gray blob and blue background). Gait state 1 is also depicted, as it is the state in which

664   cells spent the most time over all walking trajectories and also is uniquely the state from which

665   cells begin walking. Cells also frequently return to the state during walking. Further, transitions

666   from gait state 2 from gait state 1 constitute the single highest frequency transition. Together, all

667   identified states in this panel constitute functional states. Arrows represent the most probable

668   transitions between functional states, and all unbalanced transitions are also represented with size

669   scaled by their proportional probability compared to all other transitions emanating from the

670   source functional state. Cartoons are a walking cell in profile with cirri in a configuration

671   representative of the corresponding functional state. Labels refer to the apparent functional role

672   of states and their associated transitions. Beginning from gait state 1, the resting/start state, cells

673   are most likely to follow transitions from gait state 2 to 3 to 17 at which point cells are likely to

674   enter the functional state associated with substantial cell movement involving variable balanced

675   transitions between a number of gait states. Transitions are then likely to lead back toward the

676   cycle states. Note that while this representation of gait dynamics highlights the most probable

677   transitions, substantial variability, primarily involving reversible transitions, occurs during

678   walking trajectories.
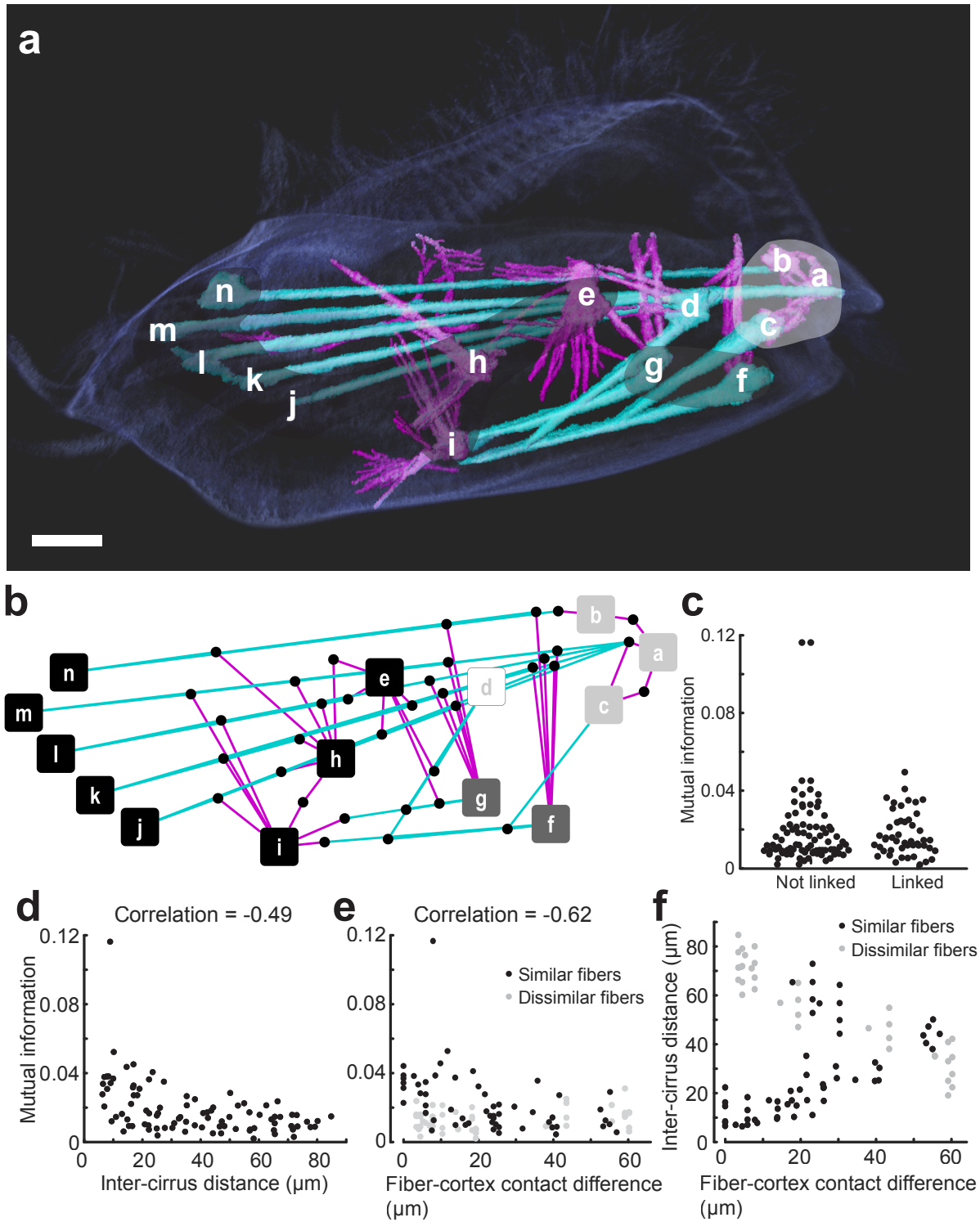
Figure 4. The fiber system of *Euplotes* is complex, interconnected, and constrains

mechanistic hypotheses of gait coordination. **a**, The SiR-tubulin labeled cell (faint, dark blue)

682   was imaged by confocal microscopy, and a 3D reconstruction as obtained from serial confocal

683   slices. Fibers were manually traced in each slice using TrakEM2 in FIJI. Two morphologically

684   distinct classes of fibers were observed and are indicated as follows: thick, linear fibers are cyan

685   and thinner, filamentous fibers are magenta (see Fig. S6 for raw image data). Fibers emanate

686   from the base of each cirrus and form a connected network between all cirri. The base of each

687   cirrus is indicated by corresponding letters (as in Fig. 2c). Gray shading indicates the dynamical

688   groups identified by dimensionality reduction and follows the same color scheme as in Fig. 3d.

689   Scale bar is 10 μm. **b**, A graph representation of fiber-fiber connections illustrates the complex

690   and interconnected nature of cirrus associated fiber topology. Nodes correspond to the cirri to

691   which each fiber system is associated, and edges indicate connections between fiber systems.

692   Colors of nodes indicate the same groups as in **a**, and colors of edges indicate the types of fibers

693   connecting to one another, cyan for thick fiber connections, magenta for thin fiber connections,

694   and purple for thick to thin fiber connections. **c**, Pairs of cirri that are linked by fiber-fiber

695   contacts show no statistically significant difference in mutual information compared to those

696   lacking fiber-fiber contacts. The plot displays mutual information between all pairs of cirri

697   grouped by the absence (Not linked) or presence (Linked) of associated fiber-fiber connections.

698   Statistical significance was evaluated by the Wilcoxon rank sum test. Note that when pairs of

699   cirri are grouped by fiber-fiber connection type, we do observe an increase in mutual information

700   for cirri associated with only thin fiber-fiber connection and only thick fiber-fiber connections

701   compared to those lacking fiber-fiber connections (see Fig. S6). **d**, A plot of mutual information

702   as a function of inter-cirrus distance displays negative correlation, with a Spearman correlation

703   coefficient of -0.49 (p<0.001). Plotted values are defined with respect to pairs of cirri. **e**, A plot

704   of mutual information as a function of fiber-cortex contact distance grouped by fiber type

705   similarity and lack thereof displays negative correlation, with a Spearman correlation coefficient

706   of -0.62 (p<0.001) for pairs of cirri with similar fiber types and no significant correlation for

707   those with dissimilar fiber types. Similarity of fiber types is defined in terms of sharing at least

708   some fiber types as defined in **a**. Fiber-cortex contact difference is measured by the mean cross

709   nearest neighbor distance (see Methods) for all fiber-cortex contact points associated with each

710   cirrus. The negative correlation values from the data plotted in **d** and **e** indicate that cirri that are

711   closer to one another and also cirri with fiber-cortex contacts in nearby regions of the cell tend to

712   have higher mutual information, and indeed cirri that are both close to one another and with

713   similar patterns of fiber-cortex contacts display the highest mutual information. **f**, A plot of fiber-

714   cortex contact difference versus inter-cirrus difference (as in panels **d** and **e**) illustrates that

715   nearby cirri tend to have similar associated fiber-cortex contacts, highlighting that nearby cirri

716   with similar fiber-cortex contacts share the most mutual information.

717

718 **Methods**

719 *Cell husbandry*

720 Cultures of *Euplotes eurystomus* were obtained from Carolina Biological Supply Company (Item

721 #131480). Individual cells were isolated from cultures, which contained other protists and

722 meiofauna, by pipetting and placed in non-treated 6-well plates (Thermo Fischer Scientific 08-

723 772-49) containing spring water taken from cultures. Cells were kept in wells for no longer than

724 five days before imaging, and if cells were to be kept for longer than 48 hours, wells containing

725 cells were supplemented with 1% Cereal Grass Medium [1] (from Thermo Fischer Scientific

726 S25242) to prevent depletion of a population of prey bacteria and otherwise maintain *Euplotes*

727 under constant growth conditions.

728

729 *Live cell brightfield microscopy*

730 Cells were concentrated by centrifugation (500×g for 5 min) and resuspended either in 0.5 mL of

731 spring water in coverglass bottomed FluoroDishes (World Precision Instruments FD35-100) or

732 in 0.2 mL spring water on a coverslip (FisherScientific, 12-545-D) for imaging. No more than

733 three cells were kept in 0.5 mL imaging samples and only one cell was ever kept in 0.2 mL

734 imaging samples in order to minimize cell-cell interactions. Cells were observed to exhibit

735 spontaneous walking activity on coverglass. Walking cells in FluoroDishes were imaged under

736 brightfield illumination using a Zeiss Z.1 Observer and Hamamatsu Orca Flash 4.0 V2 CMOS

737 camera (C11440-22CU) with a 20x, 0.8 NA Plan-Apochromat (Zeiss) objective. Cells on

738 coverslips were imaged under brightfield illumination with coverslips inverted over a well

739 containing a small amount of distilled water to reduce evaporation using a Zeiss Axio Zoom.V16

740 and a PCO pco.dimax S1 camera. Importantly, in both imaging systems, the focal plane was set

741   at the interface between cirri of walking cells and the glass surface upon which they were

742   walking. Images were acquired at 0.033 seconds per frame with a 0.005 second exposure in order

743   to capture all cirral dynamics during walking with minimal blur.

744

745   ***Quantification of walking dynamics***

746   Movies of walking cells were viewed using FIJI [2]. Movement of cirri, or lack thereof was

747   clearly visible in each movie frame (see Fig. 2a and Movie S1). The dynamical state of each

748   cirrus in each movie frame was manually annotated. For each frame, each cirrus received a label

749   of "1" if the cirrus was in motion and "0" if the cirrus was not moving and in contact with the

750   coverslip. Motion of cirri was evident in terms of a change in cirrus shape or tip position often in

751   addition to blur due to motion during image acquisition or position out of the focal plane (see

752   Fig. 2a and Movie S1). While only slowly walking cells were recorded, sometimes cells

753   nevertheless exhibit brief, spontaneous departures from slow walking during the course of movie

754   acquisition. Any frame in which the movement of the cell and/or cirri were too fast to be

755   resolved, such as during spontaneous escape responses[3] (Movie S2), was excluded from analysis

756   such that some videos were split into a number of separate continuous sequences. Thus, each

757   movie frame associated with a particular time point in the walking trajectory, with the exception

758   of those excluded from analysis as described, yielded a corresponding 14-element binary vector

759   encoding the motility state of the cell in terms of the movement of cirri. Cell movement was

760   tracked using the manual tracking feature of the TrackMate plugin in FIJI [4]. The center of each

761   cell was used as the reference feature for tracking. We analyzed the walking dynamics of 13

762   different cells.

763

764   *Dimensionality reduction*

765   Dimensionality reduction was performed by non-negative matrix factorization (NMF)

766   implemented in MATLAB release 2019b (Mathworks, Natick). NMF was chosen as a

767   dimensionality reduction technique to allow us to obtain a reduced, sparse, and interpretable

768   representation of walking dynamics. Because NMF derives non-negative factors, the basis

769   vectors in NMF space correspond to patterns of cirral activity. NMF involves factoring data, A,

770   an n by m matrix, into non-negative factors W, an n by k matrix, and H, a k by m matrix where

771   the product $W*H$ approximates A. To determine the appropriate number of dimensions or rank, k,

772   that are necessary to accurately represent the data without overfitting, we performed cross-

773   validation by imputation with random holdouts [5,6], also implemented in MATLAB. We randomly

774   held out 15% of our walking dynamics data, performed NMF for a given k, and then used the

775   NMF reconstruction $W*H$, to update the missing data entries. This process of updating is known

776   as imputation, and we repeated the imputation process 50 times, by which point the imputed

777   values had stabilize, to obtain a final NMF reconstruction. We then computed the root mean

778   squared residual (RMSR) between the final NMF reconstruction, $W*H$, and our dataset, A. We

779   performed this entire process 100 times for each value of k. As is generally the case for NMF, we

780   observed a monotonic decrease in reconstruction error with increasing k without performing the

781   imputation procedure [7] (Fig. S2a). In contrast to this trend, we observed an increase in RMSR of

782   imputed values with increasing k indicating overfitting [5] (Fig. S2b). We chose $k=3$ because this

783   value was the highest value before a notable increase in imputation error (Fig. S2b), which

784   would indicate overfitting [5,6]. Thus, our choice of rank 3 selects the lowest rank approximation

785   that captures structure of the dataset without overfitting that structure. Further, our choice

786   facilitated the visual inspection of the structure of data in the reduced dimensional

787   reconstruction.

788   Finally, we noted that for our chosen value of k, due to the stochastic nature of the NMF

789   algorithm, which involves a random initialization step, we obtained slightly different solutions

790   for different iterations [5]. In order to choose the best reduced dimensional approximation,

791   therefore, we performed NMF 500 times and chose the particular solution corresponding to the

792   lowest RMSR compared to our dataset.

793

794   ***Clustering***

795   Clustering on the dataset obtained using NMF was performed by density-based spatial

796   clustering of applications with noise (DBSCAN) algorithm [8] implemented in MATLAB release

797   2019b (Mathworks, Natick). Structure in NMF space was clearly visible (Fig. 2e-g), and

798   DBSCAN using a Euclidean distance metric, was initially chosen as a clustering method because

799   it yielded qualitatively good partitioning of the data. The DBSCAN algorithm involves stochastic

800   search within neighborhoods of a given radius ε around datapoints, and points with a minimum

801   number of neighbors, n, within their neighborhood are grouped as belonging to the same cluster,

802   leaving two free parameters to determine. We set ε by first using the

803   clusterDBSCAN.estimateEpsilon function in MATLAB (release 2020b, Phased Array System

804   Toolbox), which yielded a value of 0.15. We next set about determining the minimum neighbor

805   number, n. To do so, we computed the average Silhouette coefficient, a commonly used measure

806   of clustering quality that indicates how well-separated clusters are [9], for various values of n. The

807   results of this analysis are plotted in Fig. S3. Higher Silhouette coefficients indicate better

808   clustering, and we found that a value of $n=8$ maximized the mean Silhouette coefficient (Fig.

40

809   S3a). We also noted, however, that for this value, many datapoints were found to be outliers, not

810   belonging to any cluster due to having too few points within a distance of ε. Fig. S3b displays

811   percentage of datapoints found to be outliers as a function of n. In order to avoid categorizing

812   more than 5% of datapoints as outliers, we chose to settle on $n=4$, which does not have a

813   significantly different mean Silhouette coefficient compared to any of the others in the range

814    $n=2$-7. This choice was further supported by the fact that major clusters involving more than 5

815   datapoints identified with $n=8$ were also identified with $n=4$.

816   Although this set of parameters gave qualitatively and quantitatively reasonable

817   clustering results, we sought to further refine our clusters and to further reduce the outlier

818   datapoints. We noted the obvious partitioning of the NMF dataset into three groups along the H2

819   axis (Fig. 2e). We found the previously determined parameter values to yield good clustering for

820   the top and middle partitions (H2≤1.1 and 0.2<H2<1.1), with no outliers. For the lower partition

821   (H2≤0.2), however, we found that we were able to improve clustering by using ε=0.1182. With

822   this updated value, we found no statistically significant change in Silhouette coefficient and

823   reduced outliers to 0%. The clusters obtained by this process constituted the identification of the

824   32 gait states. We note here that the problem of determining the true or optimal number of

825   clusters is an unresolved problem [10], and we note that we have followed standard methods to

826   determine cluster number, and we found that our key results do not depend sensitively on the

827   precise number of clusters identified (see following section and Fig. S8 for more details).

828

829   ***State transition analysis***

830   Following dimensionality reduction and clustering to identify gait states, we proceeded to

831   characterize state transition dynamics. For each cell trajectory, we identified all unique gait state

832   transitions for a total of 1423 unique pairwise transitions over the cumulative 2343 video frames

833   for 77.14 s of recording. We computed empirical transition rates between states as the total

834   number of observed transitions divided by the total time of observation. In order to determine

835   which transitions were balanced and which were unbalanced, we followed Chang and Marshall

836   [11], and performed binomial tests of statistical significance. Assuming a system at equilibrium,

837   with all transitions obeying detailed balance, we expect to observe some deviation from exactly

838   reciprocal transitions and can calculate the probability of observing a given set of ratios given

839   underlying probabilities of forward and reverse transitions. The binomial probability of

840   observing a set of transitions with known forward and reverse probabilities is given by

841
$$P(X = f) = \binom{n}{f} p_{forward}^{f} p_{reverse}^{n-f}$$

842   where $\binom{n}{f} = \frac{n!}{f!(n-f)!}$ is the choose function, f is the number of forward transitions, n is the total

843   number of transitions (such that n-f is the number of reverse transitions), and the probabilities

844    $p_{forward}$ and $p_{reverse}$ are the forward and reverse probabilities. Considering only the set of

845   transitions involving a specific pair of states, and calculating the probability that a transition

846   between those states is either in the forward or reverse direction, the values of forward and

847   reverse probabilities in the balanced case must be equal such that $p_{forward} = p_{reverse} = 0.5$.

848   With an α level of 0.05, we then considered reciprocal transition pairs with binomial

849   probabilities less than 0.05 to be significantly unbalanced. Figure S9 displays the binomial

850   probabilities associated with all transitions.

851   In order to calculate the estimated entropy production rate, we followed Wan and

852   Goldstein [12], where the entropy production rate is defined as

853
$$\dot{S} = \frac{1}{2} \sum_{i \neq j} J_{ij} A_{ij}$$

854

855   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_l are the

856   probabilities of being in state l at steady state and the k_{ij} are the transition rates between states i

857   and j. We estimate the state occupancy probabilities p_l as $\frac{T_l}{T_{Total}}$, where T_l is the amount of time

858   spent in state l over all trajectories and T_{Total} is the total recorded time, and the rates k_{ij} as $\frac{N_{ij}}{T_i}$,

859   where N_{ij} is the total number of observed transitions from state i to state j and T_i is the total time

860   spent in state i. To avoid $k_{ji} = 0$, we let $k_{ji} = \frac{1}{p_j T_{max}}$ where $T_{max} = 11.55$ seconds is the

861   maximum duration of any single recorded walking trajectory.

862   In the course of our state transition analysis, we also checked whether the waiting times

863   between instances of each state might be non-exponentially distributed, with exponential

864   distributions indicative of an embedded Markov process or possibly self-organized criticality

865   [13,14]. Using the Lilliefors test implemented in Matlab, we found that in general, waiting times

866   were not exponentially distributed, although states 2, 3, 6, 16, 17, 18, 25, 27, 28, 32 were found

867   to have waiting times consistent with exponential distributions with Benjamini-Hochburg

868   corrected p-values of 0.046, 0.046, 0.022, 0.008, 0.046, 0.017, 0.046, 0.0081, 0.0046, 0.0046

869   respectively. Interestingly, none of the waiting times between the movements of individual cirri

870   were found to be consistent with exponential distributions. These results are consistent with

871   mechanisms constraining the temporal dynamics of cirri and state transitions.

872   In order to begin evaluating whether state transitions obeyed the Markov property, where

873   the transition probabilities from one state to the next are completely determined by current state

874   [15,16], we estimated the transition matrix for walking dynamics, consisting of the transition

875   probabilities between all states. We estimate the transition probability from state i to state j as

43

876    $p_{ij} = \frac{N_{ij}}{\sum_k N_{ik}}$ such that $\sum_i p_{ij} = 1$. The entries of the transition matrix, P, are these transition

877   probabilities with indices i for rows and j for columns. If gait state transitions obeyed the Markov

878   property, we expect that the product of the transition matrix with itself, P^2, would be equivalent

879   to the two-step transition matrix where transition probabilities are computed as before except that

880   state j is the state to which i has transitioned after an intervening transition. Figure S5 displays

881   the results of this analysis showing that the two matrices show substantial quantitative and

882   qualitative differences. Although these results strongly suggest violation of the Markov property,

883   we applied the Billingsley test for a more statistically rigorous evaluation [17,18]. This test was

884   implemented and performed in Matlab (release 2019b). The Billingsley test gives a χ^2 metric

885   with M^2-$2M$ degrees of freedom given by

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

887   where R_{ij}, the independent trials probability matrix, is given by

888    $R_{ij} = \sum_{k=1}^{M} N_{kj} / \left(\sum_{h=1}^{M} \sum_{l=1}^{M} N_{hl} - \sum_{k=1}^{M} N_{ik}\right).$

889   Importantly, we also noted that the key qualitative results of our state transition analysis

890   are robust to the details of clustering results. In particular, we find that strongly unbalanced

891   transitions and violation of the Markov property exist for a range of clustering parameters.

892   Figure S8 displays the transition matrices for different clustering results.

893

894   ***Confocal microscopy***

895   Cells were prepared for imaging and placed into a FluoroDish as described in the Live Cell

896   Brightfield Microscopy section. Cells were then labeled with SiR-tubulin (Spirochrome provided

897   by Cytoskeleton, Inc, CY-SC002) at 1 µM concentration. Cells were imaged using a Zeiss LSM

898   880 AxioExaminer and a 40x, 1.2 NA C-Apochromat water immersion objective (Zeiss) and

899   excitation provided by a 633 nm laser (Zeiss). Only one full confocal z-stack of a complete cell

900   was obtained during imaging to avoid effects of photodamage.

901

902   *Fiber reconstruction and analysis*

903   The image stack resulting from confocal imaging was first aligned in FIJI using the

904   StackReg plugin [2]. Next, fibers were manually segmented in each of the aligned z-stack images

905   using the TrakEM2 plugin in FIJI [19,20]. Thick and thin fibers (Fig. 5a) were morphologically

906   distinguished, with thick fibers having a diameter of no less than 5 µm at the thinnest point.

907   Fibers were traced from their distal termini to their convergences at the base of the cirri with

908   which they were associated. Following segmentation, 3D surfaces were reconstructed in

909   TrakEM2. Inter-fiber contacts were then found by inspection of 3D reconstructions and verified

910   by examining individual z-stack frames to confirm intersections between fibers.

911

912   **References**

913   1.   King, N., Young, S. L., Abedin, M., Carr, M. & Leadbeater, B. S. C. Starting and

914   maintaining Monosiga brevicollis cultures. *Cold Spring Harb. Protoc.* **4**, (2009).

915   2.   Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat.*

916   *Methods* **9**, 676–682 (2012).

917   3.   Ricci, N., Giannetti, R. & Miceli, C. The ethogram of Euplotes crassus (ciliata,

918   hypotrichida): I. The wild type. *Eur. J. Protistol.* **23**, 129–140 (1988).

919   4.   Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle

920   tracking. *Methods* **115**, 80–90 (2017).

921   5.   Kanagal, B. & Sindhwani, V. Rank Selection in Low-rank Matrix Approximations: A

922   Study of Cross-Validation for NMFs. in *Advances in Neural Information Processing*

923   *Systems* **23**, (2010).

924   6.   Owen, A. B. & Perry, P. O. Bi-Cross-Validation of the SVD and the Nonnegative Matrix

925   Factorization. *Source Ann. Appl. Stat.* **3**, 564–594 (2009).

926   7.   Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and

927   Principal Components Models. *Technometrics* **20**, 397–405 (1978).

928   8.   Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. *A Density-Based Algorithm for Discovering*

929   *Clusters in Large Spatial Databases with Noise*. (1996).

930   9.   Ronan, T., Qi, Z. & Naegle, K. M. Avoiding common pitfalls when clustering biological

931   data. *Science Signaling* **9**, re6–re6 (2016).

932   10.   Kawamoto, T. & Kabashima, Y. Cross-validation estimate of the number of clusters in a

933   network. *Sci. Rep.* **7**, 1–17 (2017).

934   11.   Chang, A. Y. & Marshall, W. F. Dynamics of living cells in a cytomorphological state

935   space. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21556–21562 (2019).

936   12.   Wan, K. Y. & Goldstein, R. E. Time Irreversibility and Criticality in the Motility of a

937   Flagellate Microorganism. *Phys. Rev. Lett.* **121**, 58103 (2018).

938   13.   Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: An explanation of the 1/f

939   noise. *Phys. Rev. Lett.* **59**, 381–384 (1987).

940   14.   Sánchez, R., Newman, D. E. & Carreras, B. A. Waiting-Time Statistics of Self-Organized-

941   Criticality Systems. *Phys. Rev. Lett.* **88**, 4 (2002).

942   15.   Markov, A. A. Extension of the Limit Theorems of Probability Theory to a Sum of

943   Variables Connected in a Chain. *Notes Imp. Acad. Sci. St. Petersbg. VIII Ser. Physio-*

944   *Mathematical Coll.* **22**, (1907).

945   16.   Gagniuc, P. *Markov Chains: From Theory to Implementation and Experimentation*.

946   (Wiley, 2017).

947   17.   Billingsley, P. *Statistical Methods in Markov Chains*. *The Annals of Mathematical*

948   *Statistics* **32**, 12–40 (Institute of Mathematical Statistics, 1961).

949   18.   Hiscott, R. N. *Chi-Square Tests for Markov Chain Analysis*. *Mathematical Geology* **13**,

950   (1981).

951   19.   Cardona, A. *et al.* TrakEM2 Software for Neural Circuit Reconstruction. *PLoS One* **7**,

952   e38011 (2012).

953   20.   Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration

954   based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).

955

966   **Author contributions**

967   BTL, JDP, and WFM conceived of the study and developed analyses. BTL collected data. BTL,

968   JG, JDP, and WFM interpreted the data. BTL and JG performed analyses. BTL and WFM wrote

969   the paper.

970

971   **Competing interests**

972   The authors declare no competing interests.

973

974   **Supplementary Information** is available for this paper including Figures S1-S9 and Movies S1

975   and S2.