Dauer fate in a *Caenorhabditis elegans* Boolean network model

Alekhya Kandoor [1], **Janna Fierst** [Corresp. 2]

[1] Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States

[2] Biomolecular Sciences Institute and Department of Biology, Florida International University, Miami, Florida, United States

Corresponding Author: Janna Fierst
Email address: jfierst@fiu.edu

Cellular fates are determined by genes interacting across large, complex biological networks. A critical question is how to identify causal relationships spanning distinct signaling pathways and underlying organismal phenotypes. Here, we address this question by constructing a Boolean model of a well-studied developmental network and analyzing information flows through the system. Depending on environmental signals *Caenorhabditis elegans* develop normally to sexual maturity or enter a reproductively delayed, developmentally quiescent 'dauer' state, progressing to maturity when the environment changes. The developmental network that starts with environmental signal and ends in the dauer/no dauer fate involves genes across 4 signaling pathways including cyclic GMP, Insulin/IGF-1, TGF-β and steroid hormone synthesis. We identified three stable motifs leading to normal development, each composed of genes interacting across the Insulin/IGF-1, TGF-Beta and steroid hormone synthesis pathways. Three genes known to influence dauer fate, *daf-2*, *daf-7* and *hsf-1*, acted as driver nodes in the system. Using causal logic analysis, we identified a five gene cyclic subgraph integrating the information flow from environmental signal to dauer fate. Perturbation analysis showed that a multifactorial insulin profile determined the stable motifs the system entered and interacted with *daf-12* as the switchpoint driving the dauer/no dauer fate. Our results show that complex organismal systems can be distilled into abstract representations that permit full characterization of the causal relationships driving developmental fates. Analyzing organismal systems from this perspective of logic and function has important implications for studies examining the evolution and conservation of signaling pathways.

1   Dauer fate in a *Caenorhabditis elegans* Boolean network model

2   Alekhya Abhiram Kandoor[1] and Janna L. Fierst[2*]

4   [1] Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA

5   [2] Biomolecular Sciences Institute and Department of Biological Sciences, Florida International University,

6   Miami, FL, USA

7   * Author for correspondence: jfierst@fiu.edu

9   **Abstract**

10   Cellular fates are determined by genes interacting across large, complex biological networks. A critical

11   question is how to identify causal relationships spanning distinct signaling pathways and underlying

12   organismal phenotypes. Here, we address this question by constructing a Boolean model of a well-

13   studied developmental network and analyzing information flows through the system. Depending on

14   environmental signals *Caenorhabditis elegans* develop normally to sexual maturity or enter a

15   reproductively delayed, developmentally quiescent 'dauer' state, progressing to maturity when the

16   environment changes. The developmental network that starts with environmental signal and ends in the

17   dauer/no dauer fate involves genes across 4 signaling pathways including cyclic GMP, Insulin/IGF-1, TGF-

18   β and steroid hormone synthesis. We identified three stable motifs leading to normal development,

19   each composed of genes interacting across the Insulin/IGF-1, TGF-Beta and steroid hormone synthesis

20   pathways. Three genes known to influence dauer fate, *daf-2*, *daf-7* and *hsf-1*, acted as driver nodes in

21   the system. Using causal logic analysis, we identified a five gene cyclic subgraph integrating the

22   information flow from environmental signal to dauer fate. Perturbation analysis showed that a

23   multifactorial insulin profile determined the stable motifs the system entered and interacted with *daf-*

24   *12* as the switchpoint driving the dauer/no dauer fate. Our results show that complex organismal

25   systems can be distilled into abstract representations that permit full characterization of the causal

26   relationships driving developmental fates. Analyzing organismal systems from this perspective of logic

27   and function has important implications for studies examining the evolution and conservation of

28   signaling pathways.

29

30

31

32   **Introduction**

33 Cellular states, developmental pathways and organismal phenotypes are determined by interactions

34 between biomolecules. These biological networks are large and complex, involving genes, proteins and

35 regulatory molecules spanning diverse pathways and levels of biological organization [1]. Data-based

36 network inference often implicates large numbers of biomolecules through correlation-based methods

37 [2] but network modeling and analysis indicate causal relationships are the key to controlling

38 information flow and determining cellular fates [3]. A fundamental question is how to identify the

39 critical, causal relationships within complex biological systems and between distant nodes in a network

40 [4]. Here, we construct a Boolean model of a well-studied model system and use network analysis

41 techniques to identify the core interactions responsible for the phenotype.

42

43 *C. elegans* is a small (~1mm), free-living nematode worm that has been used as a model organism in

44 biological research for over 50 years [5]. Populations are androdioecious, composed of self-fertile

45 hermaphrodites and a small number of males, resulting in straightforward laboratory culture and rapid

46 propagation. *C. elegans* are transparent, facilitating microscope study, and eutelic, meaning that all

47 adults have the same number of somatic cells. In *C. elegans* hermaphrodites there are 959 somatic cells

48 while males have 1031 [6]. The entire life cycle can be completed within 3 days and these factors have

49 led to the establishment of *C. elegans* as an important model organism for studying development,

50 genetics and essential biological processes. *C. elegans* was the first metazoan to have its genome fully

51 sequenced [7, 8] and the long history of *C. elegans* research has provided a wealth of genetic, genomic

52 and developmental understanding of this small worm [9-12].

53

54 One of the most studied developmental processes in *C. elegans* is dauer [13, 14]. The small worms are

55 vulnerable to environmental conditions including lack of food or moisture [15] and high temperature

56 [16]. *C. elegans* senses its environment through chemoreceptors that perceive biomolecules including

57 pheromones produced by other worms [17]. In low-stress environments the worms progress through

58 discrete larval stages L1-L4, to adulthood and reproductive maturity [13]. Environmental triggers can

59 activate a developmental switch at the L1 stage and lead to dauer arrest [18]. Dauer worms can move

60 but are developmentally quiescent and survive without feeding for periods of up to 4 months.

61

62 *C. elegans* can be forced into the dauer fate with high pheromone concentrations, high temperature or

63 starvation in the lab but the signals can also interact quantitatively in the worm's natural environment

64 [17]. The dauer pheromone itself is a complex mix of ascarosides [19, 20] with natural variation in

65   production, composition and response between populations [21], and species [22]. *C. elegans* strains

66   have substantial, quantitative natural variation for dauer entry [23], [24] and exit [25] but for the

67   individual worm dauer is an irreversible developmental decision. Undergoing dauer has lasting effects

68   on worm reproduction [26], physiology [27] and locomotion [28].

69

70   Dauer is a unique, discrete developmental switch and the molecular basis has been studied extensively.

71   Over 100 genes across 4 signaling pathways (cyclic GMP, Insulin/IGF-1, TGF-β and steroid hormone

72   synthesis) have been implicated as influencing dauer [13, 29]. Despite extensive descriptive

73   characterization, we lack a robust understanding of the interplay between these pathways and the

74   causal links that drive the system to one fate or another. Here, we study a Boolean model and identify

75   the critical interactions determining dauer fate.

76

77   Simplified representations of complex biological networks can accurately distill systems down to

78   essential interactions [30]. Boolean networks are graphs where nodes represent genes and edges are

79   regulatory interactions defined as Boolean functions or logic gates, for example the operators 'AND' or

80   'OR' [31]. Boolean models of biological processes represent states as discrete 'on' or 'off' [32] and

81   accurately represent the switch-like nature of dauer induction. Regulatory interactions are threshold-

82   gated, qualitative relationships. Here, we use our Boolean network to emulate cellular behavior

83   observed in experimental settings and systematically perturb the system to discover causal links within

84   the larger complex network. We identify the critical interactions and combinations of gene states that

85   determine dauer fate. Evolutionary searches of pathway conservation and divergence typically examine

86   gene-by-gene, sequence level homology [33]. Here, we suggest that causal logic and function may be

87   the true phenotypes underlying the evolution of signaling pathways.

88

89   **Methods**

90   ***Constructing the dauer network***

91   We first extracted 85 dauer-involved genes described in the WormBook 'Dauer' chapter [13]. For each of

92   these genes, we then searched on WormBase [53] for genes known to interact with these dauer genes

93   and genes known to act in the same pathways as these dauer genes. We focused on regulatory

94   interactions and included those for which we identified information regarding the direction of the

95   interaction, for example reported relationships between effector proteins and targets. We did not

96   include genetic interactions or genetic masking/enhancing/suppressing effects between genes. Some of

97   our gene pairs may have reported genetic interactions along with regulatory interactions but we focused

98   on regulatory interactions for constructing our model.

99

100   We next surveyed the literature for regulatory interactions within this gene and pathway set and

101   created a filtered set of dauer genes and main-effect regulatory interactions (Supplementary Table 1).

102   For example, of the 40 insulin signaling genes in the *C. elegans* genome few have been reported to act

103   independently in the Insulin/IGF-1 pathway [5]. We condensed the larger dauer-involved set down to 26

104   genes, each part of the four dauer-inducing pathways represented as nodes in our Boolean model, and

105   connected by 34 edges. We included the nodes 'pheromone' [34] and '*cmk-1*' (a food receptor [35]) as

106   inputs to the system. These interactions were mainly obtained from the WormBase database [53] and

107   verified with the scientific publications cited on WormBase (a full list of nodes, interactions, and

108   citations is given in Supplementary Table 1).

109

110   We used this regulatory interaction information for each gene to formulate Boolean functions

111   (Supplementary Table 2). We combined literature-reported regulations into a preliminary model and

112   used results obtained from simulations and perturbations to fine-tune the Boolean functions for each

113   node until we were able to mimic the expected behavior of the system. For example, multiple articles

114   reported that *daf-16* was activated either when *akt* was repressed [54-57] or when *daf-12* was activated

115   [58, 59]. Simulated conditions for normal adult development should therefore result in *daf-16*

116   repression and no dauer activation. To achieve this known output for *daf-16* we assigned an 'OR'

117   operator for these two regulators. In some cases, we were able to extract the relationship between

118   regulatory nodes from the literature. This approach has been similarly demonstrated in previous works

119   to gain qualitative insights using discrete networks [60, 61].

120

121   ***Studying system trajectories***

122   The initial conditions used to generate the trajectories of the system's dynamics were based on well-

123   known environmental cues such as food availability (*cmk-1*) and dauer pheromone (pher) detected by *C.*

124   *elegans* [17, 35]. Favorable conditions allowed the organism to develop into an adult ("no dauer") and

125   unfavorable conditions into "dauer" [34]. We set the initial conditions as *cmk-1* = 1 and pher = 0 for

126   normal development (here, dauer = 0) and *cmk-1* = 0 and pher = 1 for dauer fate (here, dauer = 1). We

127   monitored the dynamics and convergence of the system based on the state of the 'dauer' output node

128   (trajectories for each simulation are shown in Supplementary Figure 2). Temperature is another well-

129 known regulator that induces dauer formation. We chose to focus here on dauer as a response to

130 resource availability and our model does not include temperature as an input.

131

132 We performed simulations with asynchronous updating using boolean2 software [62]. In order to

133 explore system dynamics, the Boolean functions for each gene needs to be updated at each timestep to

134 trace out the trajectory of the system. These transitions for the Boolean functions were carried out

135 using an asynchronous scheme that randomly assigned the states of the nodes. Biological events do not

136 occur at the same time and the timescale at which the events occur can be diverse. Here, we

137 represented this with an asynchronous update scenario. In this framework a collector class averaged the

138 states of nodes over multiple runs for simulating a system with a non-deterministic updating scheme.

139 We performed 100 simulations for each set of initial conditions and for every iteration, the collector

140 averaged the state of each node generated for 1000 steps. In all our simulations, the nodes reached

141 equilibrium after approximately 15 iterations (trajectories are shown in Supplementary Figure 2). Nodes

142 with states > 0.5 were considered 'ON' (as indicated by [62]) and nodes with states <0.5 were

143 considered 'OFF'.

144

145 ***Validating the model***

146 To validate the model, we checked whether the model has the ability to reproduce known experimental

147 outcomes. Mutant phenotypes in *C. elegans* that lead to either adult or dauer formation are well

148 studied [13]. We used published experimental evidence to check whether our model could reproduce

149 similar results. For example, dauer-constitutive (Daf-c) mutants lead to dauer formation in conditions

150 that do not favor dauer while Dauer-defective (Daf-d) mutants will not develop into dauer larvae [18,

151 40-42]. We knocked down known Daf-c and Daf-d genes to validate that our dauer Boolean model

152 captured essential wild type and mutant phenotypes (Supplementary Figure 1b).

153

154 Our model simulations helped us identify the attractor space and stable motifs. A Boolean model will

155 eventually settle down into a static state where all the nodes are either 'on' or 'off' and remain in that

156 state through further iterations. This state of the model is called an attractor and in biological modeling

157 attractors depict specific biological phenotypes. Within this, a set of nodes within the network forms

158 cycles which stay in specific states, and these are called stable motifs. We can represent a network as a

159 reduced system using the nodes that are critical for controlling the stable motifs. We can then study the

160   model dynamics using the reduced form and can make predictions using simpler representations of the

161   model.

162

163   Attractors of Boolean networks represent biological phenotypes, here the 'dauer' and 'no dauer' fates.

164   We explored the set of attractors and stable motifs for our model by iterating through all possible

165   combinations of initial conditions and perturbations using pystablemotifs software [36, 37]. A

166   perturbation of a node in our experiments refers to reversing the existing state of a node and specify

167   that along with the initial conditions set. The network diagrams were created using Cytoscape [63].

168   Code, simulations and scripts are available at https://github.com/alekhyaa2/Dauer-Boolean-Model.

169

170   **Results**

171   ***Constructing a dauer Boolean network***

172   Our exhaustive literature survey identified essential regulatory interactions among dauer pathway

173   genes. From these we constructed a Boolean network with 26 nodes and 34 interactions (Figure 1a;

174   Supplementary Table 1). The nodes 'pheromone' [34] and '*cmk-1*' (a food receptor [35]) were input

175   signals that allowed the model to converge onto the dauer/no dauer output node.

176

177   ***Validating the model by replicating Daf-c and Daf-d phenotypes***

178   In *C. elegans* mutant strains can be categorized as dauer-constitutive (Daf-c), genotypes which produce

179   dauer arrest even in the absence of pheromone and presence of food, and Dauer-defective (Daf-d),

180   genotypes which do not enter dauer arrest under typical dauer-favorable conditions (i.e., pheromone

181   present and/or food limited) [18, 40-42]. We validated our model by knocking down known Daf-c genes

182   (*daf-2*, *daf-7*, *daf-11*, and *daf-9*) separately under initial conditions (*cmk-1* = 1 and pher = 0) that do not

183   typically favor dauer and found that these mutants followed the Daf-c phenotype and entered dauer

184   arrest. We also knocked down known Daf-d mutant genes (*daf-12* and *daf-16*) under conditions that

185   typically favor dauer (here, *cmk-1* = 0 and pher = 1) and found these replicated the Daf-d phenotype,

186   failing to enter dauer arrest.

187

188   ***Identifying stable motifs***

189   A network attractor is a set of gene states the system can stabilize in [36]. The dauer Boolean network

190   had 10 different attractor states (Supplementary Figure 1) with 1 resulting in developmental progression

191   to adult, 8 resulting in the dauer fate and 1 resulting in either dauer or no dauer depending on system

192 conditions [37]. Although an attractor represents a potential stable point, not all attractors are

193 accessible from all states. To further analyze system stability, we divided the total network into small

194 subgraphs [3, 38]. We identified three Stable Motifs (SMs) in the developmental progression to adult

195 (Figure 1b-d). Each of these encompassed genes from at least two different pathways. Stable Motif 1

196 (SM1; Figure 1b) spanned 10 genes across the IGF-1/insulin signaling, TGF-β and Steroid hormone

197 pathways [13]. SM2 and SM3 involved 4 and 8 genes, respectively, across the IGF-1/insulin signaling and

198 TGF-β pathways. *daf-2*, *daf-7* and *hsf-1* (heat shock protein) were common between all 3 stable motifs

199 with *hsf-1* acting as a crucial node where different pathways intersect. This result has been shown by

200 previous experimental work [39]. Each of the SMs involved a different insulin signaling gene, *ins-1*, *ins-7*

201 or *ins-18*. We did not identify any SMs for dauer formation (the A3 attractor; Supplementary Figure 1).

202

203 ***Identifying driver nodes in the SMs***

204 We simulated gene perturbations to identify the driver nodes for each of our identified SMs [37].

205 Changing the existing state of the critical nodes shared between the stable motifs (i.e., *daf-2* or *daf-7*

206 downregulation or *hsf-1* upregulation) prevented the system from entering into any of the SMs and

207 pushed it to one of the dauer attractors. Perturbation analysis identified the critical causal interactions

208 in the network as SM2, encompassing *daf-2*, *daf-7*, *ins-7* and *hsf-1*. Perturbation of *daf-2*, *daf-7* or *hsf-1*

209 genes prohibited the system from stabilizing in the SM and the attractors corresponded to each of these

210 nodes. We concluded that *daf-2*, *daf-7* and *hsf-1* were the driver nodes for the SMs. When we perturbed

211 the other nodes of these three stable motifs, the system converged onto either dauer or no dauer via

212 different stable motifs. (Supplementary Table 3). But in case of *daf-2*, *daf-7* and *hsf-1* perturbations, the

213 system did not enter into any of the stable motifs, instead settling into the attractors of those

214 corresponding nodes.

215

216 Perturbations of *daf-2*, *daf-7* and *hsf-1* resembled the Daf-c mutant system. When the model was

217 started with dauer-favorable conditions, the system did not have any stable motifs. The same lack of

218 stable motifs occurred with Daf-c mutant phenotypes. Under both dauer-favorable and Daf-c mutant

219 conditions (here, related gene perturbation and/or favorable initial conditions), the system did not have

220 any cyclic subgraphs facilitating control. Instead, the driver nodes decided the fate of the organism in

221 expressing a wildtype or a Daf-c mutant phenotype.

222

223 ***Analyzing causal interactions in the dauer Boolean network***

224   We identified a reduced network representing causal information flow [43] from environmental input to

225   dauer fate (Figure 2a). The reduced network contained a cyclic subgraph (Figure 2b) similar to SM2,

226   containing *daf-2*, *daf-7*, *ins-7* and *hsf-1*. However, it included a necessary inhibitory regulation of *daf-9*

227   by *daf-2*. The dynamics of the entire system can be summarized as input to this subgraph. The full

228   reduced network (Figure 2c) included 12 nodes and 13 edges.

229

230   ***Insulin signaling drives the binary dauer/no dauer fate***

231   Gene perturbations revealed an additional 7 SMs (SM4-10; Supplementary Table 3). For each of the

232   perturbations performed, the system could end up in one of 10 different attractors but few of these

233   could be reached from stable conditions with active stable motifs. Activating 'dauer' conditions including

234   low food and pheromone presence resulted in dauer fate and the A3 attractor (Figure 3a) and available

235   food with pheromone absence resulted in the system reaching the A7 attractor through SMs 1, 2 or 3

236   (Figure 3b).

237

238   We identified insulin-dependent dauer arrest attractors across the perturbations. Activation of *ins-1* and

239   *ins-18* resulted in the dauer fate through SMs 4-10 but inactivation of either pushed the system to

240   normal development (Figure 3c). Together with *ins-7* the activation/inactivation profile of these genes

241   pushed the system to normal development through SMs 1, 2 or 3 (Figure 1b-d). Similarly, *daf-12*

242   activation resulted in dauer arrest with the insulin profile determining the SM the system entered.

243   When the antagonistic insulins (*ins-1* and *ins-18*) were activated the system stabilized into SM 8 and

244   when *ins-1* was turned off the system stabilized in either SM 2 or 3 (Figure 3d).

245

246   We found that the system must travel through SMs 1, 2 or 3 to end up in normal developmental

247   progression and avoid the dauer fate. To enter dauer arrest the system must travel through one of SMs

248   4-10 with *ins-1* and *ins-18* activated and *ins-7* inactivated. In this model insulin signaling controls the

249   transitions between normal development and dauer in the presence of gene perturbations. The insulin

250   genes act as decision makers for the convergence of these dauer pathways.

251

252   ***Identifying binary states across dauer pathways***

253   For normal development we identified one strongly correlated module, one strongly anticorrelated

254   module and a third set of genes with quantitative relationships across the SMs (Figure 4). For dauer

255   arrest we identified two strongly correlated modules and two strongly anticorrelated modules. These

256   modules span the 4 signaling pathways involved in dauer arrest and indicate co-regulation and

257   correlative relationships between distant genes in the network.

258

259   **Discussion**

260   We have constructed a Boolean model of the dauer arrest network and identified the critical, causal

261   interactions driving the system between dauer fate and normal development. Although the network

262   contains 26 genes and 34 interactions, the system can be driven between the dauer/no dauer fate by

263   controlling a few critical nodes. Importantly, the switchpoints for this system require specific

264   combinations of gene states behaving synergistically or antagonistically. The results of our model

265   demonstrate that complex systems can be controlled through a few critical nodes and connections,

266   echoing both control theory [3, 44] and long-standing theoretical work on biological systems [45].

267   Evolutionary studies of signaling pathway divergence and conservation have focused on individual genes

268   and sequence-level changes [33]. Our results demonstrate that diffuse, correlative interactions have no

269   impact on system stability or developmental fate. We suggest that system-level function and dynamics

270   are, instead, the true evolutionary targets of selection.

271

272   Many of the results of our study mirror experimental dauer investigations. For example, a systematic

273   analysis of mutants across insulin signaling genes in *C. elegans* identified roles in lifespan, reproduction,

274   pathogen resistance, thermotolerance and dauer entry and exit [46]. Here, by constructing a network

275   that encompassed multiple signaling pathways we were able to analyze the role of these insulin

276   signaling genes and robustly identify the interaction between insulin signaling, heat shock response and

277   daf-sensitive genes that determines dauer fate. Our results demonstrate that dauer signaling pathways

278   find different routes to transduce the signals from the environment to developmental decision.

279

280   Over 100 genes have been described as influencing the dauer fate in *C. elegans* [13, 29]. However, many

281   of these have been identified through mutant analyses or total loss of function alleles [47]. Genes may

282   be capable of influencing the dauer state through loss-of-function or extreme phenotypes but this may

283   not translate to a gene's involvment in dauer fate in natural populations. We focused on the genes and

284   interactions determined essential to the dauer network. Distilling these further into causal interactions

285   revealed a core set of genes and proteins controlling dauer fate. Transcriptomic studies have identified

286   large, co-regulated, correlated modules associated with dauer mutants, entry and exit [46]. Our results

287   show that these co-regulated modules also occur in our model (Figure 4) but their expression and

288   function are correlated consequences of developmental decision-making occurring at a few critical

289   points in the system.

290

291   *C. elegans* is a microscopic soil-dwelling organism and extremely dependent on environmental

292   conditions. Some type of developmental arrest or metabolic stasis occurs in many small organisms with

293   similar environmental dependence. For example, tardigrades (Phylum Tardigrada) are microscopic

294   invertebrates closely related to nematodes and can enter both ametabolic and anhydrous states where

295   they extrude all water and remain in suspended animation for multiple years [48]. Despite the ubiquity

296   of developmental and metabolic stasis, there is little understanding of the genetics and regulatory

297   systems underlying these processes [49]. Evolutionary studies of developmental arrest across

298   nematodes have used dauer genes and pathways as a tool to construct related models. For example, the

299   parasitic nematode *Haemonchus contortus* transitions from both its L3 and L4 stages depending on cues

300   from the host's gut environment [50]. A dauer model based on *C. elegans* identified 61 gene

301   homologues with varying levels of conservation, each of which may be a potential target for life cycle

302   disruption and anti-parasitic intervention [51]. Alternatively, many of these genes may not be involved

303   in developmental arrest in *H. contortus*. They may have pleiotropic functions in *C. elegans* or have been

304   co-opted for other organismal phenotypes.

305

306   Multiple studies support this scenario of conserved dauer phenotype and divergent gene and sequence

307   evolution. A study of dauer-related genes and pathways within Phylum Nematoda found a high number

308   of *Caenorhabditis* and *C. elegans*-specific dauer genes and an evolutionarily recent co-option of TGF-β

309   signaling into the dauer pathway [29]. Although the insulin signaling pathway is broadly conserved,

310   outside of *Caenorhabditis* the insulin genes examined showed little evolutionary conservation. In the

311   nematode *Pristionchus pacificus* dauer pheromones produce dauer entry in individuals of other

312   genotypes, suggesting the worms are using dauer signaling in intraspecific competition [22]. Natural

313   variation in *P. pacificus* dauer is partially explained by copy number variation in *dauerless*, an 'orphan'

314   gene that lacks homologs in related species [52]. These examples of near-scale divergence suggest that

315   although dauer itself is a conserved developmental fate, the individual genes underlying dauer may not

316   be essential targets of selection.

317

318   Our results suggest that system-level properties responsible for developmental fate may be the true

319   evolutionary phenotypes under selection. The small number of genes and interactions underlying stable

320   motifs and the causal logic backbones may be subject to higher levels of evolutionary conservation, or

321   their system 'role' may be conserved. Laboratory genetic manipulations within *C. elegans* and across

322   nematodes, focusing on the key genes and interactions identified here, would be a tractable, exciting

323   avenue for future study. For example, within *C. elegans* there are at least 40 insulin signaling homologs

324   [5]. Divergence, loss and duplication of individual genes may not be as significant as the role the

325   homologs together play in dauer signaling.

326

327   This work has important future applications. The model can be used to study how different cellular

328   responses work in cohort with dauer pathways by extending the system to encompass genes and

329   interactions involved in disparate pathways including the cell cycle and immune response. Expanding

330   this framework can provide a tractable system to test experimental hypotheses that are currently

331   challenging, labor-intensive and beyond the scope of traditional laboratory investigations. Viewing the

332   conservation and divergence of signaling pathways through the lens of causality and system-level

333   function can provide a robust framework for understanding the evolution of biological networks.

334

335

336

337   **Acknowledgements**

338   JLF and AAK were supported by National Science Foundation grants EF-1921562 and DEB1941854 and

339   NIGMS award R35GM147245 to JLF.

340 **References**
341
342 1. Weng, G., U.S. Bhalla, and R. Iyengar, *Complexity in biological signaling systems.*
343   Science, 1999. **284**(5411): p. 92-6.
344 2. Stolovitzky, G., D. Monroe, and A. Califano, *Dialogue on reverse-engineering*
345   *assessment and methods: the DREAM of high-throughput pathway inference.* Ann N Y
346   Acad Sci, 2007. **1115**: p. 1-22.
347 3. Zanudo, J.G. and R. Albert, *Cell fate reprogramming by control of intracellular network*
348   *dynamics.* PLoS Comput Biol, 2015. **11**(4): p. e1004193.
349 4. Lawyer, G., *Understanding the influence of all nodes in a network.* Sci Rep, 2015. **5**: p.
350   8665.
351 5. Girard, L.R., et al., *WormBook: the online review of Caenorhabditis elegans biology.*
352   Nucleic Acids Research, 2007. **35**: p. D472-D475.
353 6. Corsi, A., B. Wightman, and M. Chalfie, *A Transparent Window into Biology: A Primer*
354   *on Caenorhabditis elegans.* Genetics, 2015. **200**(2): p. 387-407.
355 7. Consortium, C.e.S., *Genome sequence of the nematode C. elegans.* Science, 1998.
356   **282**(5396): p. 2012-8.
357 8. Waterston, R.H., J.E. Sulston, and A.R. Coulson, *The Genome*, in *C. elegans II*, nd, et al.,
358   Editors. 1997: Cold Spring Harbor (NY).
359 9. Riddle, D.L., et al., *Introduction to C. elegans*, in *C. elegans II*, nd, et al., Editors. 1997:
360   Cold Spring Harbor (NY).
361 10. Kemphues, K.J. and S. Strome, *Fertilization and Establishment of Polarity in the*
362   *Embryo*, in *C. elegans II*, nd, et al., Editors. 1997: Cold Spring Harbor (NY).
363 11. Schedl, T., *Developmental Genetics of the Germ Line*, in *C. elegans II*, nd, et al., Editors.
364   1997: Cold Spring Harbor (NY).
365 12. Schnabel, R. and J.R. Priess, *Specification of Cell Fates in the Early Embryo*, in *C.*
366   *elegans II*, nd, et al., Editors. 1997: Cold Spring Harbor (NY).
367 13. Hu, P.J., *Dauer.* WormBook.
368 14. Cassada, R.C. and R.L. Russell, *Dauerlarva, a Post-Embryonic Developmental Variant*
369   *of Nematode Caenorhabditis-Elegans.* Developmental Biology, 1975. **46**(2): p. 326-342.
370 15. Golden, J.W. and D.L. Riddle, *The Caenorhabditis-Elegans Dauer Larva -*
371   *Developmental Effects of Pheromone, Food, and Temperature.* Developmental Biology,
372   1984. **102**(2): p. 368-378.
373 16. Ailion, M. and J.H. Thomas, *Dauer formation induced by high temperatures in*
374   *Caenorhabditis elegans.* Genetics, 2000. **156**(3): p. 1047-1067.
375 17. Jeong, P.Y., et al., *Chemical structure and biological activity of the Caenorhabditis*
376   *elegans dauer-inducing pheromone.* Nature, 2005. **433**(7025): p. 541-5.
377 18. Vowels, J.J. and J.H. Thomas, *Genetic analysis of chemosensory control of dauer*
378   *formation in Caenorhabditis elegans.* Genetics, 1992. **130**(1): p. 105-23.
379 19. Butcher, R., et al., *Small-molecule pheromones that control dauer development in*
380   *Caenorhabditis elegans.* Nature Chemical Biology, 2007. **3**(7): p. 420-422.
381 20. Butcher, R., et al., *A potent dauer pheromone component in Caenorhabditis elegans that*
382   *acts synergistically with other components.* Proceedings of the National Academy of
383   Sciences of the United States of America, 2008. **105**(38): p. 14288-14292.
384 21. Viney, M. and S. Harvey, *Reimagining pheromone signalling in the model nematode*
385   *Caenorhabditis elegans.* Plos Genetics, 2017. **13**(11).

386 22. Mayer, M. and R. Sommer, *Natural variation in Pristionchus pacificus dauer formation*
387   *reveals cross-preference rather than self-preference of nematode dauer pheromones.*
388   Proceedings of the Royal Society B-Biological Sciences, 2011. **278**(1719): p. 2784-2790.
389 23. Harvey, S., A. Shorto, and M. Viney, *Quantitative genetic analysis of life-history traits of*
390   *Caenorhabditis elegans in stressful environments.* Bmc Evolutionary Biology, 2008. **8**.
391 24. Viney, M., M. Gardner, and J. Jackson, *Variation in Caenorhabditis elegans dauer larva*
392   *formation.* Development Growth & Differentiation, 2003. **45**(4): p. 389-396.
393 25. Bubrig, L.T., J.M. Sutton, and J.L. Fierst, *Caenorhabditis elegans dauers vary recovery*
394   *in response to bacteria from natural habitat.* Ecology and Evolution, 2020. **10**(18): p.
395   9886-9895.
396 26. Ow, M.C. and S.E. Hall, *Starvation memory resulting in reproductive plasticity is*
397   *conserved in some Caenorhabditis elegans wild isolates.* MicroPubl Biol, 2020. **2020**.
398 27. Ow, M.C., A.M. Nichitean, and S.E. Hall, *Somatic aging pathways regulate reproductive*
399   *plasticity in Caenorhabditis elegans.* Elife, 2021. **10**.
400 28. Pradhan, S., et al., *Environmental Programming of Adult Foraging Behavior in C.*
401   *elegans.* Curr Biol, 2019. **29**(17): p. 2867-2879 e4.
402 29. Gilabert, A., et al., *Expanding the view on the evolution of the nematode dauer signalling*
403   *pathways: refinement through gene gain and pathway co-option.* BMC Genomics, 2016.
404   **17**: p. 476.
405 30. Bornholdt, S., *Less is more in modeling large genetic networks.* Science, 2005.
406   **310**(5747): p. 449-+.
407 31. Bornholdt, S., *Boolean network models of cellular regulation: prospects and limitations.*
408   Journal of The Royal Society Interface, 2008. **5**(Suppl 1): p. S85.
409 32. Wang, R., A. Saadatpour, and R. Albert, *Boolean modeling in systems biology: an*
410   *overview of methodology and applications.* Physical Biology, 2012. **9**(5).
411 33. Pires-daSilva, A. and R. Sommer, *The evolution of signalling pathways in animal*
412   *development.* Nature Reviews Genetics, 2003. **4**(1): p. 39-49.
413 34. Golden, J.W. and D.L. Riddle, *The Caenorhabditis elegans dauer larva: developmental*
414   *effects of pheromone, food, and temperature.* Dev Biol, 1984. **102**(2): p. 368-78.
415 35. Neal, S.J., et al., *Feeding state-dependent regulation of developmental plasticity via*
416   *CaMKI and neuroendocrine signaling.* Elife, 2015. **4**.
417 36. Rozum, J.C., et al., *Parity and time reversal elucidate both decision-making in empirical*
418   *models and attractor scaling in critical Boolean networks.* Sci Adv, 2021. **7**(29).
419 37. Rozum, J.C., et al., *pystablemotifs: Python library for attractor identification and control*
420   *in Boolean networks.* Bioinformatics, 2021.
421 38. Zanudo, J.G. and R. Albert, *An effective network reduction approach to find the*
422   *dynamical repertoire of discrete dynamic networks.* Chaos, 2013. **23**(2): p. 025111.
423 39. Barna, J., et al., *Heat shock factor-1 intertwines insulin/IGF-1, TGF-beta and cGMP*
424   *signaling to control development and aging.* BMC Dev Biol, 2012. **12**: p. 32.
425 40. Vowels, J.J. and J.H. Thomas, *Multiple chemosensory defects in daf-11 and daf-21*
426   *mutants of Caenorhabditis elegans.* Genetics, 1994. **138**(2): p. 303-16.
427 41. Gottlieb, S. and G. Ruvkun, *daf-2, daf-16 and daf-23: genetically interacting genes*
428   *controlling Dauer formation in Caenorhabditis elegans.* Genetics, 1994. **137**(1): p. 107-
429   20.
430 42. Albert, P.S. and D.L. Riddle, *Mutants of Caenorhabditis elegans that form dauer-like*
431   *larvae.* Dev Biol, 1988. **126**(2): p. 270-93.

432   43.   Maheshwari, P. and R. Albert, *A framework to find the logic backbone of a biological*
433   *network.* BMC Syst Biol, 2017. **11**(1): p. 122.

434   44.   Zanudo, J.G.T., G. Yang, and R. Albert, *Structure-based control of complex networks*
435   *with nonlinear dynamics.* Proc Natl Acad Sci U S A, 2017. **114**(28): p. 7234-7239.

436   45.   Riedl, R., *SYSTEMS-ANALYTICAL APPROACH TO MACRO-EVOLUTIONARY*
437   *PHENOMENA.* Quarterly Review of Biology, 1977. **52**(4): p. 351-370.

438   46.   Fernandes de Abreu, D.A., et al., *An insulin-to-insulin regulatory network orchestrates*
439   *phenotypic specificity in development and physiology.* PLoS Genet, 2014. **10**(3): p.
440   e1004225.

441   47.   Fay, D., *Genetic mapping and manipulation: chapter 1--Introduction and basics.*
442   WormBook, 2006: p. 1-12.

443   48.   Keilin, D., *The problem of anabiosis and latent life: history and current concept.*
444   Proceedings of the Royal Society Series B, 1959. **150**: p. 149-191.

445   49.   Kalimari, M., et al., *Comparative transcriptomics suggest unique molecular adaptations*
446   *within tardigrade lineages.* BMC Genomics, 2019. **20**: p. 607.

447   50.   Gibbs, H., *Hypobiosis in parasitic nematodes— an update.* Advances in Parasitology,
448   1986. **25**: p. 129-74.

449   51.   Ma, G., et al., *Dauer signalling pathway model for Haemonchus contortus.* Parasites &
450   Vectors, 2019. **12**.

451   52.   Mayer, M., et al., *The Orphan Gene dauerless Regulates Dauer Development and*
452   *Intraspecific Competition in Nematodes by Copy Number Variation.* PLoS Genetics,
453   2015. **11**(6): p. e1005146.

454   53.   Harris, T.W., et al., *WormBase: a modern Model Organism Information Resource.*
455   Nucleic Acids Research, 2019. **48**(D1): p. D762-D767.

456   54.   Henderson, S.T. and T.E. Johnson, *daf-16 integrates developmental and environmental*
457   *inputs to mediate aging in the nematode Caenorhabditis elegans.* Curr Biol, 2001.
458   **11**(24): p. 1975-80.

459   55.   Lin, K., et al., *Regulation of the Caenorhabditis elegans longevity protein DAF-16 by*
460   *insulin/IGF-1 and germline signaling.* Nat Genet, 2001. **28**(2): p. 139-45.

461   56.   Lee, R.Y., J. Hench, and G. Ruvkun, *Regulation of C. elegans DAF-16 and its human*
462   *ortholog FKHRL1 by the daf-2 insulin-like signaling pathway.* Curr Biol, 2001. **11**(24): p.
463   1950-7.

464   57.   Paradis, S. and G. Ruvkun, *Caenorhabditis elegans Akt/PKB transduces insulin receptor-*
465   *like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor.* Genes Dev, 1998.
466   **12**(16): p. 2488-98.

467   58.   Matyash, V., et al., *Sterol-derived hormone(s) controls entry into diapause in*
468   *Caenorhabditis elegans by consecutive activation of DAF-12 and DAF-16.* PLoS Biol,
469   2004. **2**(10): p. e280.

470   59.   Dowell, P., et al., *Convergence of peroxisome proliferator-activated receptor gamma and*
471   *Foxo1 signaling pathways.* J Biol Chem, 2003. **278**(46): p. 45485-91.

472   60.   Hopfensitz, M., et al., *Attractors in Boolean networks: a tutorial.* Computational
473   Statistics, 2013. **28**(1): p. 19-36.

474   61.   Bornholdt, S., *Boolean network models of cellular regulation: prospects and limitations.*
475   J R Soc Interface, 2008. **5 Suppl 1**: p. S85-94.

476   62.   Albert, I., et al., *Boolean network simulations for life scientists.* Source Code Biol Med,
477   2008. **3**: p. 16.

478 63.   Shannon, P., et al., *Cytoscape: a software environment for integrated models of*
479   *biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
480

Figure 1

Figure 1 Models representing the four major dauer inducing pathways and the stable motifs of this network.

(a) The dauer Boolean network included 26 nodes and 34 edges. Each node represented a molecular entity (gene, receptor, protein, etc.) involved in signal transduction pathways. Edges indicated regulatory interactions between the nodes that activated (signified by blue arrow heads) or inhibited (signified by red 'T' edges). The four dauer inducing pathways were represented by four different colors. Here, *cmk-1* and pheromone were input signals which when turned on and off, decided the dauer fate. Nodes *srbc* (pheromone receptor) and *hsf-1* (heat shock protein) did not belong to any of the four dauer pathways. (b), (c) and (d) Subgraphs that were part of the cyclic SMs 1, 2 and 3 respectively, into which the dauer Boolean network stabilizes. The nodes in red text were inhibited and nodes in black text were activated.
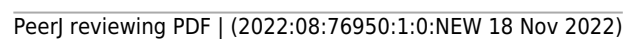
Figure 2

Figure 2 The causal logic backbone of the dauer Boolean network

(a) The edges were assigned causal logic (blue = necessary, red = sufficient and black = sufficient and necessary). Nodes filled grey with dashed borders were input signals where grey filled nodes were part of cyclic subgraphs (SMs) and solid borders indicated output nodes. (b) The logic backbone derived from identifying subgraphs from input to motif nodes and motif nodes to output signals. Activation of pher and *ncr* was sufficient to reach the motif and *cmk-1* was necessary. The motif stabilization was sufficient to control dauer fate. (c) The reduced dauer logic network consisted of 12 nodes and 13 edges. *Daf-9* and *daf-12* entered a cyclic feedback loop. The insulin factors involved in different endocrine pathways were important points of the network regulating *daf-2* activity and present in the reduced network.
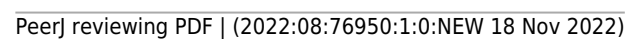
Figure 3

Figure 3 Model transitions from initial conditions to attractors.

(a) Under unfavourable conditions, the system eventually reached the "Dauer" attractor A3 without settling in any of the cyclic motifs. (b) Under favorable conditions (*cmk-1* is on and pheromone is off), the network stabilized in either of the three cyclic SMs which led to the attractor A7 and the "No dauer" phenotype. (c) When a node of SM 1, 2 or 3 was perturbed the system stabilized in different SMs and converged on either dauer arrest or normal development. (d) *daf-12* perturbation resulted in dauer arrest through SMs 2, 3 or 8. *daf-12* presence, irrespective of initial conditions, was sufficient for dauer formation. SMs 4-10 and the attractors of the perturbed node simulations are given in the Supplementary Materials.
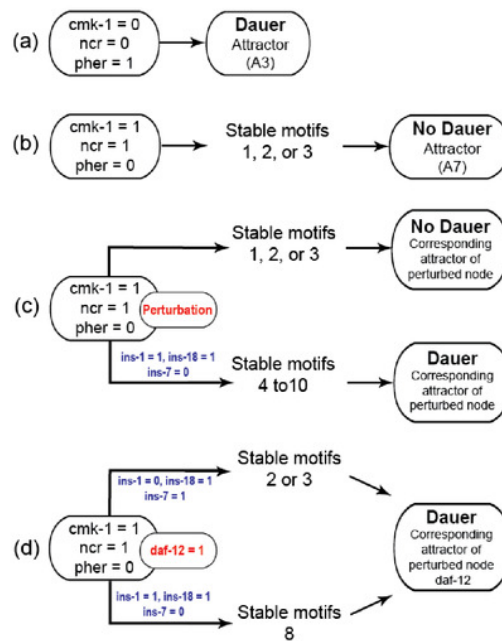
Figure 4

Figure 4 Correlation plots as a function of hamming distance for every node of the attractors converging to either "dauer" or "no dauer" when the nodes of the SMs were perturbed.

Hamming distance = 0 implied that the two nodes were in the same state and as the distance value increased the states in which these two nodes exist in the attractors varied. From the dauer=1 correlation plot we observed clusters of nodes that existed in same state across the four dauer inducing pathways. *Ins-1* and *ins-18* were positively and negatively correlated when attractors converged on dauer and no dauer phenotypes, respectively.