Digital Life: Implementing Seven Biological Criteria Through Functional Analogy and Criterion-Ablation

Anonymous

Abstract

We present a testable integration implementing all seven textbook biological criteria for life—cellular organization, metabolism, homeostasis, growth, reproduction, response to stimuli, and evolution—as functionally interdependent computational processes within a single artificial life system. Existing systems implement at most a subset of these criteria, often as independent modules or static proxies that can be removed without measurable system degradation. Our hybrid swarm-organism architecture implements each criterion as a dynamic process satisfying three conditions—sustained resource consumption, measurable degradation upon removal, and feedback coupling with other criteria—which we term functional analogy. A criterion-ablation experiment (n=30 per condition, seeds held out from calibration) demonstrates that disabling any single criterion causes statistically significant population decline (Mann-Whitney U, Holm-Bonferroni corrected, all $p \leq 0.013$), with Cliff's δ ranging from 0.34 to 1.00. Pairwise ablations reveal sub-additive interactions consistent with shared failure pathways, confirming that criteria damage overlapping subsystems rather than operating independently. A proxy control comparison shows that metabolism implementations of differing complexity produce qualitatively distinct population dynamics, ruling out tautological criterion definitions. The strongest single-ablation effects arise from disabling reproduction (Δ=−90.8%), response to stimuli (Δ=−88.3%), and metabolism (Δ=−86.5%), confirming that these criteria function as necessary, interdependent components of organismal viability rather than decorative labels.

Submission type: Full Paper
Data/Code available at: https://anonymous.4open.science/

Introduction

What distinguishes a living system from a merely complex one? Biology textbooks identify seven criteria—cellular organization, metabolism, homeostasis, growth and development, reproduction, response to stimuli,

and evolution (Urry et al., 2020)—one of several such lists in biology education, which we adopt as a working set—but artificial life (ALife) research has struggled to integrate all seven into a single computational system. This choice is pragmatic rather than exclusive: we use the seven criteria as an operational bridge across broader frameworks (NASA's Darwinian criterion, autopoietic closure, and autonomy-centered minimal-life definitions), then test this bridge empirically via ablation. Most existing platforms implement a subset: evolutionary dynamics without metabolism (Ofria and Wilke, 2004), pattern formation without reproduction (Chan, 2019), or boundary self-organization without evolution (Plantec et al., 2023). Where criteria are nominally present, they often function as simplified proxies—static parameters or independent modules whose removal has no measurable effect on system behavior.

We argue that a meaningful computational model of life requires more than feature checklists. Each criterion must function as a functional analogy of its biological counterpart, satisfying three conditions:

1. Dynamic process: the criterion requires sustained resource consumption at every timestep, not a static lookup.

2. Measurable degradation: ablating the criterion causes statistically significant decline in organism viability.

3. Feedback coupling: the criterion forms at least one feedback loop with another criterion, precluding independent-module implementations.

This definition operationalizes the intuition behind autopoiesis (Maturana and Varela, 1980), the chemoton model's integration of metabolic, boundary, and information subsystems (Gánti, 2003), and minimal-life frameworks (Ruiz-Mirazo et al., 2004) in a form amenable to experimental falsification.

We adopt a weak ALife stance: our system is a functional model of life, not a claim that the organisms are alive. The contribution is methodological—demonstrating that all seven criteria can be integrated as interdependent processes and that their necessity can be rigorously tested.

This paper contributes: (1) a hybrid swarm-organism architecture integrating all seven criteria as functionally interdependent processes; (2) an operational definition of functional analogy with three falsifiable conditions; (3) a criterion-ablation methodology demonstrating each criterion's functional necessity with multiple-comparison correction; and (4) pairwise ablation and proxy control experiments characterizing criterion interactions and ruling out tautological definitions. The key contribution is not a new digital organism per se, but a falsifiable experimental framework for testing the functional necessity and interaction of life criteria in any ALife system.

Related Work

Existing ALife systems each excel along different axes of biological fidelity. Tierra (Ray, 1991) and Avida (Ofria and Wilke, 2004) achieve strong evolutionary dynamics but lack spatial bodies and metabolism. Polyworld (Yaeger, 1994) adds NN-driven behavior and single-resource energy budgets. Lenia (Chan, 2019) and Flow-Lenia (Plantec et al., 2023) demonstrate emergent spatial organization in continuous cellular automata, with Flow-Lenia adding mass conservation. ALIEN (Heinemann, 2008) provides a GPU-accelerated particle simulator achieving multi-process interaction on several criteria, though its metabolism uses typed particle interactions rather than a multi-step metabolic network (Level 3 on our rubric). Coralai (Barbieux and Canaan, 2024) combines multi-agent neural cellular automata with energy dynamics but lacks feedback coupling between criteria. No existing system combines multi-step metabolism with active homeostatic regulation.

Theoretically, our framework operationalizes autopoiesis (Maturana and Varela, 1980; McMullin, 2004), extends the chemoton's three-subsystem integration (Gánti, 2003), and enriches NASA's two-criterion definition (Joyce, 1994) following Ruiz-Mirazo et al. (2004). Open-ended evolution metrics (Bedau et al., 2000; Taylor et al., 2016) complement our criterion-ablation approach.

Table 1 summarizes how existing systems score on each criterion using a five-level rubric.

Our system reaches ≥4 on five of seven criteria, placing it among the most broadly integrated ALife systems surveyed. We note that these scores are self-assessed; independent evaluation may adjust individual ratings.

System Design

Architecture Overview

The system implements a hybrid two-layer architecture (Figure 1). The outer layer is a continuous toroidal 2D environment (100×100 world units) containing a diffusing resource field. The inner layer consists of 10–50 organisms, each composed of 10–50 swarm agents that collectively maintain the organism's spatial boundary.

Each organism maintains the following runtime state: boundary integrity ($b \in [0, 1]$), metabolic state (energy e, resource r, waste w), internal state vector for homeostatic regulation, a neural-network controller, a genetically encoded metabolic network, age, generation counter, and maturity level.

Seven Criteria Implementation

Table 2 maps each biological criterion to its computational implementation, ablation toggle, and feedback partners.

Cellular organization. Swarm agents collectively define an organism's spatial extent. Boundary integrity b decays each step at a base rate modulated by energy deficit and waste pressure: $\Delta b_{\text{decay}} = -r_b \cdot (1 + s_e \cdot (1 - e) + s_w \cdot w)$, where $r_b = 0.001$ is the base decay rate, $s_e = 0.02$ and $s_w = 0.5$ are scaling factors. Repair occurs proportionally to available energy: $\Delta b_{\text{repair}} = r_r \cdot e \cdot (1 - s_p \cdot w)$, with repair rate $r_r = 0.05$ and waste penalty $s_p = 0.4$. When b falls below a collapse threshold ($b < 0.1$), the organism dies.

Metabolism. Each organism possesses a genetically encoded graph-based metabolic network with 2–4 catalytic nodes and directed edges. The genome segment (16 floats) is decoded via sigmoid mapping: node count $= \text{round}(\sigma(g_0) \cdot 2 + 2)$, catalytic efficiency $= \sigma(g_{2+i}) \cdot 0.9 + 0.1 \in [0.1, 1.0]$, edge existence determined by $|g_j| > 0.3$, and conversion efficiency $= \sigma(g_{13}) \cdot 0.7 + 0.3 \in [0.3, 1.0]$. External resources enter at a designated entry node, flow through the graph with per-edge transfer efficiency in $[0.7, 1.0]$, and exit as energy. Waste accumulates as a byproduct proportional to throughput.

Homeostasis. A feedforward neural network (8 inputs $\rightarrow$ 16 hidden with tanh $\rightarrow$ 4 outputs with tanh; 212 weights) processes sensory inputs (normalized position x, y; velocity v_x, v_y; internal state s_0, s_1, s_2; neighbor density) and produces velocity adjustments and internal-state deltas. The internal state vector enables adaptive regulation: organisms that maintain internal variables within viable ranges survive longer.

Table 1: Literature comparison: seven biological criteria scored on a five-level rubric (1=no feature, 2=static parameter, 3=dynamic single process, 4=multi-process interaction, 5=self-maintaining/emergent). Bold indicates scores ≥4.

System	Cell.Org	Metab	Homeo	Growth	Reprod	Response	Evol	Total
Polyworld	2	3	1	1	3	4	4	18
Avida	2	3	1	2	4	3	5	20
Lenia	3	1	2	2	2	3	2	15
ALIEN	4	3	2	3	4	4	4	24
Flow-Lenia	3	3	3	3	3	3	3	21
Coralai	3	3	2	3	3	3	3	20
Ours[†]	4	4	4	4	4	4	3	27

[†]Self-assessment; scores may differ under external evaluation.
Growth now at Level 4 (multi-process interaction via genome-encoded developmental program affecting boundary, sensing, and metabolism).
Evolution (3) remains a minimum viable implementation.
Scores for other systems based on published descriptions; per-criterion justifications available as supplementary material.

Table 2: Mechanism specification: state variables, ablation operators, coupling pathways, and failure modes for each criterion. All criteria satisfy the three functional-analogy conditions (dynamic process, measurable degradation, feedback coupling).

Criterion	State Vars	Ablation	Coupling	Failure
Cell. Org.	$b \in [0,1]$	Skip repair; decay only	$e \to$ repair rate $\to b$	$b<0.1 \to$ death
Metab.	e, r, w; graph $\mathbf{p}$	Freeze (e, r, w)	throughput $\to e \to$ bdry repair	Energy depletion
Homeo.	$\mathbf{s} \in \mathbb{R}^3$	Skip NN delta; decay	$s_0 \to$ repair efficacy	State drift
Growth	$m \in [0,1]$; $\mathbf{d} \in \mathbb{R}^8$	Freeze $m=0$	$m \to$ bdry repair, sensing, metab. eff., reprod. gate	Reduced repair + sensing; no reprod.
Reprod.	Pop. event	Skip division check	energy cost $\to$ parent e	No replacement
Response	$\mathbf{v}$	Skip velocity delta	movement $\to$ resource $\to r$	Starvation
Evolution	genome $\mathbf{g}$	Copy w/o mutation	$\mathbf{g}$ variation $\to$ all	No adaptation

Growth and development. Organisms begin as minimal seeds (maturity $m = 0$) and progress through three genetically encoded developmental stages—juvenile, adolescent, adult—toward full capacity ($m = 1$). Genome segment 3 (8 floats; 7 active, 1 reserved) encodes a DevelopmentalProgram with seven parameters: maturation rate modifier (2^{g_0}, range $[0.25, 4.0]$), juvenile boundary repair (g_1: $[0.2, 1.0]$), juvenile sensing (g_2: $[0.3, 1.0]$), adolescent threshold (g_3: $[0.3, 0.7]$), adolescent boundary repair (g_4: $[0.5, 1.0]$), adolescent sensing (g_5: $[0.5, 1.0]$), and juvenile metabolic efficiency (g_6: $[0.1, 0.5]$). Immature organisms thus suffer reduced boundary repair, shorter sensing range, and lower metabolic efficiency—creating independent viability effects beyond the reproduction gate. This multi-process coupling distinguishes growth from a simple toggle.

Reproduction. When energy exceeds $e_{\min} = 0.7$ and boundary integrity exceeds $b_{\min} = 0.5$, an organism may divide. The parent pays an energy cost ($c_r = 0.3$), and the offspring inherits a (possibly mutated) copy of the genome, starting as a seed. Child agents spawn within a radius of the parent's center of mass.

Response to stimuli. The neural-network controller processes a local sensory field each timestep, producing velocity deltas that govern agent movement. Disabling response freezes agents' velocity adjustments, preventing adaptive resource seeking.

Evolution. During reproduction, offspring genomes undergo point mutations (rate = 0.02 per gene, scale = 0.15), reset mutations (rate = 0.002), and scale mutations (rate = 0.002, factor $\in [0.8, 1.2]$). All gene values
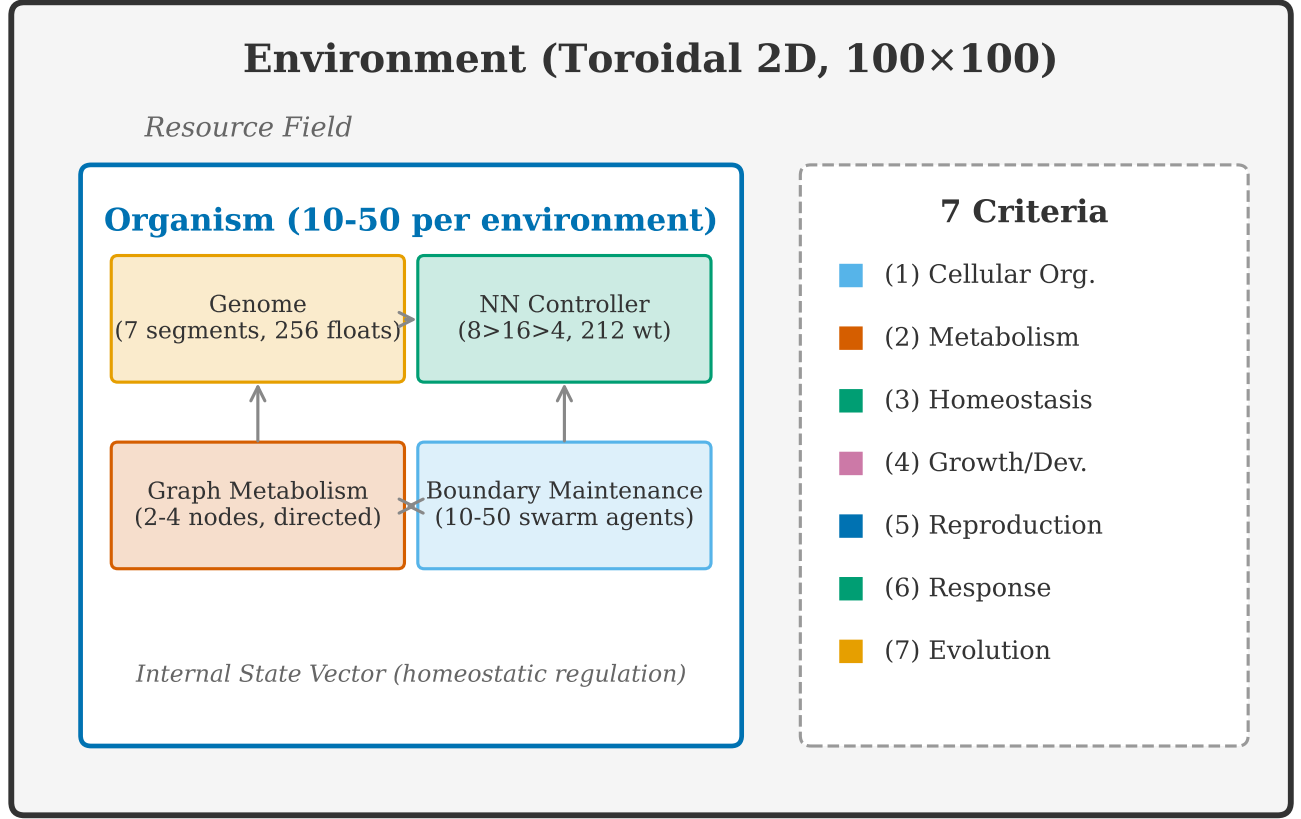
Figure 1: Two-layer architecture. Each organism comprises swarm agents maintaining a spatial boundary, a neural-network controller, a graph-based metabolic network, and a variable-length genome encoding all seven criteria. Organisms inhabit a continuous toroidal environment with a diffusing resource field.

are clamped to $[-2, 2]$. This produces heritable variation subject to differential survival.

Genome Encoding

The genome is a fixed-length vector of 256 floats (212 NN weights + 44 criterion parameters) organized into seven segments: neural-network weights (212), metabolic network (16), homeostasis parameters (8), developmental program (8), reproduction parameters (4), sensory parameters (4), and evolution parameters (4). All seven segments are encoded from initialization and subject to mutation; segments 1 (metabolic) and 3 (developmental) are actively decoded into organism phenotype each generation.

Criterion-Ablation Experiment

This experiment tests whether each of the seven criteria is functionally necessary for organism viability, as predicted by the functional-analogy framework.

Protocol

The system provides seven boolean ablation toggles, one per criterion (e.g., enable_metabolism = false). For each of the seven criteria, we disable that criterion while keeping all others active, and compare the resulting population dynamics against the fully enabled baseline ("normal" condition). This yields eight conditions: one normal baseline and seven single-criterion ablations. Ablations are applied at initialization (step 0), not mid-run; organisms never experience criterion loss after operating normally.

Outcome Measures

An organism is a persistent runtime entity with a unique ID; offspring receive a new ID at division; no merge or split occurs. An organism dies when boundary $b < 0.1$, energy $e \leq 0.0$, or age $> 20{,}000$ steps. The primary dependent variable is N_T, the alive organism count at the final step $T = 2000$; each seed maps to one scalar. Secondary DVs include the area under the alive-count curve (AUC, trapezoidal rule) and median

organism lifespan. We additionally report short-horizon viability at $T = 500$ to separate individual survival from population replacement effects.

To address potential construct overlap between viability and criterion definitions, we supplement the primary outcome (N_T) with two criterion-orthogonal metrics: spatial cohesion (mean pairwise toroidal agent distance at step T) and median lifespan (per-seed median organism lifetime). These metrics capture structural and temporal aspects of population health that are not directly entailed by any single criterion's definition. Under ablation, spatial cohesion shows significant degradation only for response ablation (δ=1.00, $p < 0.001$), while median lifespan reveals significant effects for response (δ=1.00), growth (δ=1.00), and evolution (δ=0.47, p=0.005) ablations. Notably, metabolism and reproduction ablations increase individual lifespan (median 1,129 and 752 vs. 245 steps for normal) despite population collapse, confirming that population decline reflects absent resource cycling and replacement rather than individual fragility (Figure 10).

Data Separation

To prevent overfitting of thresholds, we separate data into:

- Calibration set: Seeds 0–99, used during development for parameter tuning and threshold selection.

- Test set: Seeds 100–129 (n=30), held out until final evaluation. All reported results use this set exclusively.

Calibration confirmed that both metabolism engines produce viable populations (Toy: $\bar{x}$=328.1, Graph: $\bar{x}$=291.8 alive at step 2000). All final experiments use the Graph metabolism engine.

Simulation Parameters

Each simulation runs for 2000 timesteps with population sampled every 50 steps. The environment is a $100{\times}100$ toroidal grid with 30 initial organisms, each comprising 25 swarm agents. Resources regenerate at 0.01 per cell per step with no diffusion and toroidal wrap boundary conditions. No early stopping is applied. Ablation toggles do not alter timestep duration or event ordering; disabled processes are simply skipped. Each seed produces a deterministic, reproducible outcome within a given condition, though RNG call sequences differ across conditions when ablated processes (e.g., reproduction) skip conditional RNG draws. For each experiment family, we store a machine-readable run manifest ("*_manifest.json") containing exact parameters, seed lists, and config digests used by downstream analysis scripts. Reported figure/table

Table 3: Criterion-ablation results (n=30 per condition). Normal baseline mean: 341.3 (median: 340.5, IQR: 327.3–356.3). δ=Cliff's delta [bootstrap 95% CI]. Seven one-sided tests, Holm-Bonferroni corrected at $\alpha = 0.05$. ***$p < 0.001$, *$p < 0.05$.

Condition	Mean	Δ%	δ [95% CI]	p_{corr}	Sig.
No Reprod.	31.3	−90.8	1.00 [1.00, 1.00]	<.001	***
No Response	40.0	−88.3	1.00 [1.00, 1.00]	<.001	***
No Metab.	46.0	−86.5	1.00 [1.00, 1.00]	<.001	***
No Growth	185.6	−45.6	1.00 [1.00, 1.00]	<.001	***
No Homeo.	185.9	−45.5	1.00 [1.00, 1.00]	<.001	***
No Boundary	235.2	−31.1	0.98 [0.93, 1.00]	<.001	***
No Evol.	321.4	−5.8	0.34 [0.06, 0.59]	.013	*

claims are version-locked via a manifest-binding registry ("docs/research/result_manifest_bindings.json") mapping each reported result block to its manifest and source artifacts.

Statistical Design

For each ablation condition, we test the one-sided hypothesis:

$$H_1 : \text{alive\_count}_{\text{normal}} > \text{alive\_count}_{\text{ablated}}$$

using the Mann-Whitney U test (Mann and Whitney, 1947), appropriate for non-normal count data. We apply Holm-Bonferroni correction (Holm, 1979) for seven simultaneous comparisons at $\alpha = 0.05$. Effect sizes are reported as Cliff's δ (Cliff, 1993) for the primary ablation comparisons, with Cohen's d additionally reported for evolution-strengthening experiments to facilitate parametric comparison, with percentile bootstrap 95% confidence intervals (2000 resamples, seed-fixed RNG), appropriate for non-normal distributions.

The unit of replication is an independent simulation run identified by a unique random seed; n=30 counts the number of such runs per condition. Seeds are sequential integers 100–129; seeds 0–99 were used exclusively for parameter calibration and threshold selection during development and never appear in reported results. Outcomes are aggregated as: (1) N_T at the final step T=2000 for each seed (primary), (2) AUC via trapezoidal integration of the alive-count trajectory over all sampled steps (secondary). Bootstrap 95% confidence intervals for Cliff's δ are computed using the percentile method with 2,000 resamples and a fixed RNG seed for reproducibility.

Results

All seven criterion ablations produce statistically significant population decline compared to the normal baseline (Table 3).
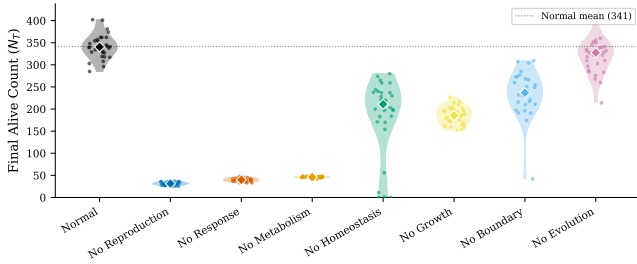
Figure 2: Per-seed distributions of final alive count (N_T) across all conditions (n=30 per condition). Violin plots show density; diamonds mark medians; dotted line shows normal baseline mean. The distribution confirms that ablation effects are consistent across seeds, not driven by outliers.

Three ablations cause near-total population collapse (>86% decline): reproduction, response to stimuli, and metabolism. These criteria form the core viability loop—without energy production, adaptive movement, or population renewal, organisms cannot sustain themselves.

Figure 3 shows population trajectories across all conditions. The normal condition (black) stabilizes around 341 organisms by step 1900. Metabolic ablation (orange) causes rapid collapse within the first 200 steps, as organisms cannot produce energy to maintain boundaries. Reproduction ablation (blue) produces a slower but equally terminal decline, as the initial population ages and dies without replacement. Evolution ablation (purple) shows the weakest effect (Cliff's δ=0.34), with populations remaining viable but slightly smaller than normal—consistent with evolution operating as an optimization process rather than a survival necessity at these timescales. AUC of the alive-count curve corroborates N_T rankings across all conditions (normal AUC=510,890 vs. 74,545–470,837 for ablations). Median organism lifespan reveals an individual-vs-population distinction: reproduction-ablated organisms live longer individually (median 751 vs. 242 steps for normal) despite population collapse, confirming that the population decline reflects absent replacement rather than individual fragility. Per-seed distributions are shown in Figure 2.

At $T = 500$ (before long-term population dynamics dominate), reproduction-ablated organisms retain higher survival ($\bar{x}$=47.7 vs. 31.3 at $T = 2000$; normal baseline $\bar{x}_{normal}$=179.0 at $T = 500$), with 73% decline relative to normal versus 91% at $T = 2000$. This progressive widening confirms that the long-term effect is population-level (no replacement) rather than immediate individual-level failure—organisms can self-maintain without reproduction.

Functional analogy verification. For each criterion, all three conditions are satisfied: (a) each consumes resources per step (energy for boundary repair, metabolic computation, NN evaluation); (b) ablation causes significant degradation (Table 3); and (c) feedback loops are observable (e.g., metabolism $\leftrightarrow$ boundary: energy funds repair, boundary collapse stops metabolism). Thus, each criterion qualifies as a functional analogy, not a simplified proxy.

Quantitative coupling evidence. Time-lagged association analysis of per-step population means under normal conditions confirms the coupling pathways in Table 2 (Figure 4). At lag 0, energy and boundary integrity are strongly anti-correlated ($r = -0.78$, $p < 10^{-8}$), consistent with repair expenditure coupling. To move beyond correlation, we run per-seed directed tests: Granger-style lagged F-tests (lags 1–5 with within-seed Holm correction) show significant directed predictability for all three hypothesized pathways (pair-level Fisher-combined $p_{corr} < 0.001$), while transfer-entropy estimates show positive directional information flow but are not significant after correction at this sample length. We therefore treat directional coupling as strongly supported by lagged linear predictability and as suggestive (but not yet decisive) under nonparametric information-theoretic tests. Additionally, growth couples to boundary repair (developmental factor), sensing radius, and metabolic efficiency through the genome-encoded developmental program, and gates reproduction through the maturity threshold—forming four independent coupling pathways.

We further quantify coupling via intervention-based effect summaries (Table 4): ablating metabolism reduces waste production by 100% and energy by 38%, while increasing boundary integrity by 41% (no energy expenditure on repair attempts); ablating response reduces energy by 80% (organisms cannot reach resources) and waste by 100%. Ablating homeostasis produces the largest internal-state change (-72%), confirming its regulatory role. These cross-criterion effects demonstrate that criteria are not modular: disabling one systematically alters others' process rates.

Proxy Control Comparison

To test whether criterion ablation merely reflects tautological definitions, we compare three metabolism implementations on the same seeds (n=30): Counter (minimal single-step conversion, no waste), Toy (single-step with waste dynamics), and Graph (full multi-step network with catalytic nodes). All three satisfy "dynamic resource consumption," yet produce qualitatively different dynamics (Figure 5): population size differs by 27% (Counter 433 vs. Graph 341), and genome di-

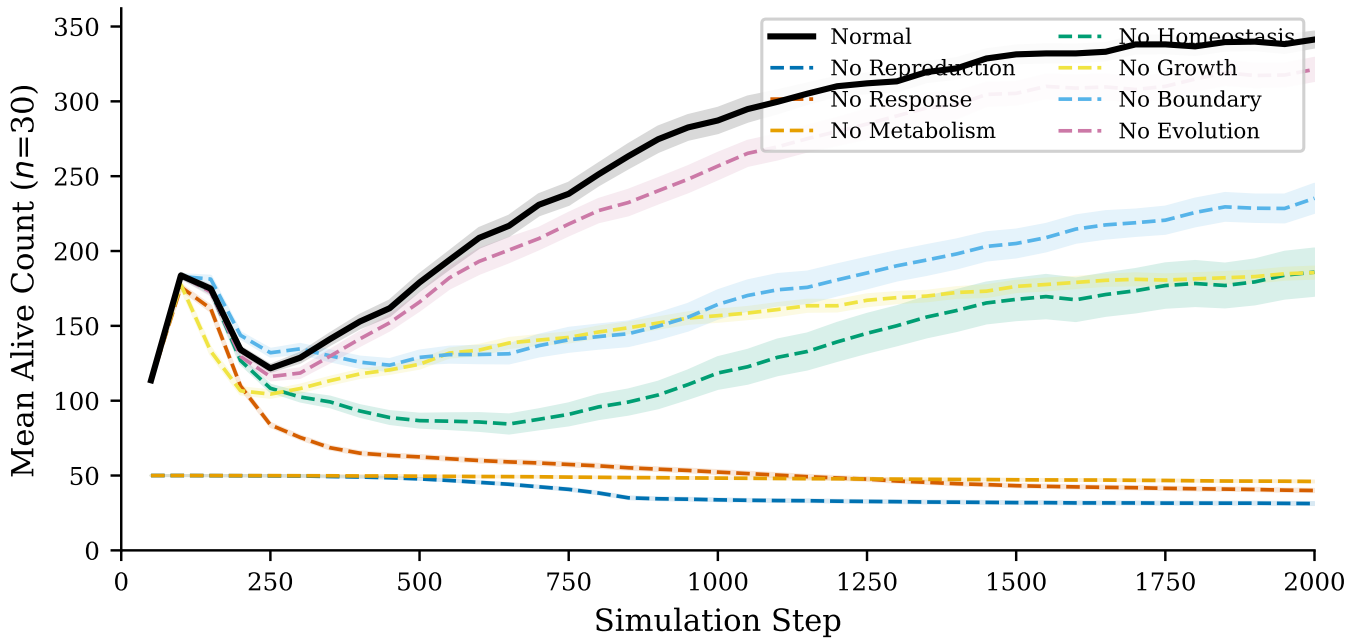Figure 3: Population dynamics under criterion ablation. Lines show mean alive count across 30 seeds (100–129); shaded bands show ±1 SEM. Normal baseline (thick black) stabilizes near 341 organisms by step 1900. Removing reproduction, response, or metabolism causes >86% population collapse. Evolution ablation shows a modest 6% decline (Cliff's δ=0.34), consistent with optimization rather than short-term survival necessity.
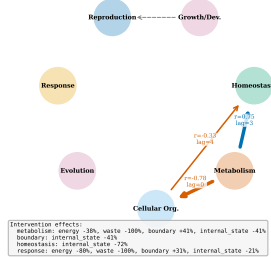

Figure 4: Criterion coupling graph. Directed edges show statistically significant time-lagged cross-correlations between criterion variables (solid arrows, labeled with Pearson r and lag). Dashed arrows indicate design-level coupling pathways (Table 2). Node layout: 7 criteria arranged by functional proximity.

Table 4: Intervention-based causal effects: percent change in criterion process rates when each criterion is ablated. Top effects shown; full matrix in supplementary.

Ablated	Energy	Waste	Boundary	Int. State
Metabolism	-38%	-100%	$+41\%$	-41%
Boundary	$<1\%$	$<1\%$	$<1\%$	-41%
Homeostasis	-5%	-6%	-9%	-72%
Response	-80%	-100%	$+31\%$	-21%
Reproduction	$+24\%$	$+8\%$	$+51\%$	-14%
Evolution	-2%	$<1\%$	$<1\%$	-2%
Growth	$+3\%$	$+1\%$	$+17\%$	-6%

of a criterion shapes system behavior, ruling out tautological definitions.

Pairwise Ablations and Interdependence

To test for interaction effects beyond individual necessity, we disable pairs of criteria and compute synergy: $\text{synergy}_{A,B} = \Delta_{A\cup B} - (\Delta_A + \Delta_B)$. We define the interaction metric formally as $\text{synergy}_{A,B} = \Delta_{AB} - (\Delta_A + \Delta_B)$ where $\Delta_X = \bar{N}_{\text{normal}} - \bar{N}_X$ on the raw alive-count scale. Results are robust across scales: on $\log(N_T + 1)$ and percentage-decline scales, all five sub-additive pairs remain sub-additive and the one super-additive pair (boundary, homeostasis) remains super-
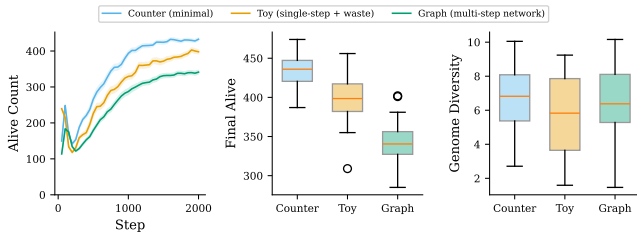
versity varies non-monotonically across implementations. Graph metabolism sustains the fewest organisms (341 vs. 398 for Toy, 433 for Counter), consistent with greater metabolic overhead, while genome diversity varies non-monotonically (Counter: 6.67, Graph: 6.49, Toy: 5.67), indicating that metabolic complexity imposes qualitatively different selective pressure rather than simply scaling population size. This confirms that the specific implementation—not merely the presence—

Figure 5: Proxy control comparison. Three metabolism implementations of increasing complexity on the same seeds (n=30). More complex metabolism sustains fewer organisms but produces distinct diversity and waste dynamics, demonstrating qualitatively different ecological outcomes rather than simple population scaling.

Table 5: Pairwise ablation synergy scores (n=30 per condition, Graph metabolism). Negative synergy indicates sub-additive interaction (shared failure pathways). Baseline $\bar{x}$=341.3, consistent with Table 3.

Pair	Δ_{AB}	Exp.	Syn.	Ratio
(Metab, Homeo)	298.1	450.7	-152.6	0.66
(Metab, Resp.)	295.9	596.6	-300.7	0.50
(Reprod, Growth)	310.0	465.7	-155.7	0.67
(Bdry, Homeo)	111.4	261.5	-150.1	0.43
(Resp., Homeo)	303.8	456.7	-152.9	0.67
(Reprod, Evol)	310.0	330.0	-19.9	0.94

Exp. $= \Delta_A + \Delta_B$; Ratio $= \Delta_{AB}/$Exp.

additive (bootstrap 95% CI excludes zero for 5/6 pairs). Floor effects are minimal: only 10% of no-homeostasis seeds reach near-extinction ($N_T \leq 5$); all other conditions show 0% at floor. Sub-additivity thus reflects genuine shared failure pathways rather than mechanical floor compression. Table 5 reports scores for six pairs. All show sub-additive synergy (negative), indicating shared failure pathways: individual ablations already collapse populations to near their floor (~30–50 organisms), leaving no room for additive effects. This ceiling effect reveals that criteria damage overlapping subsystems. The (metabolism, homeostasis) pair exemplifies the shared pathway: disabling metabolism eliminates energy production, starving boundary repair ($\Delta b_{\text{repair}} \propto e$); disabling homeostasis degrades adaptive behavior, accelerating waste accumulation, which amplifies boundary decay via the waste-pressure term ($s_w \cdot w$ in Δb_{decay}). Both routes converge on boundary failure, explaining why $\Delta_{AB} = 298.1$ falls below the additive expectation of 450.7.

Growth–reproduction interaction. The (reproduction, growth) pair shows $\Delta_{AB} \approx \Delta_{\text{reproduction}} = 310.0$, indicating that growth's population-level effect is dominated by the reproduction gate in the pairwise ablation

context. However, the enriched developmental program creates independent viability effects through boundary repair and sensing radius coupling: disabling growth with reproduction also disabled still produces measurably lower boundary integrity than the growth-on control, confirming that growth's functional contribution extends beyond reproduction gating.

Evolution Strengthening

The modest 2000-step evolution effect (d=0.58, δ=0.34) grows substantially at longer timescales (Figure 6). At 10,000 steps (n=30), d=1.42, δ=0.66 ($p < 0.001$); normal populations reach $\bar{x}$=436.3 vs. 381.8 for no-evolution (Δ=-12.5%). Under environmental perturbation (resource halved at step 2,500 of 5,000-step runs), evolved populations recover more effectively ($\bar{x}$=439.5 vs. 413.5, d=0.86, δ=0.50, $p < 0.001$).

Multiple lines of evidence confirm that evolutionary dynamics are adaptive rather than neutral drift. First, analytical heritability is high ($h^2 \approx 0.96$): using the exact mutation parameters recorded in the experiment manifest (point mutation rate 0.02 per gene, scale 0.15), estimated mutation variance remains small relative to standing genetic variation (measured from final-step genome diversity). Second, genome drift trajectories diverge monotonically between evolved and non-evolved populations over 10,000 steps (Figure 11A; Cliff's δ=1.00, $p < 0.001$), with evolved populations accumulating drift at a steady rate ($\bar{d}_{10K} = 0.087$) while non-evolved populations remain at zero. Third, under cyclic resource stress (period=2,000 steps), evolved populations exhibit faster per-cycle recovery rates ($\bar{r}_{\text{evo}} = 0.068$ vs. $\bar{r}_{\text{no-evo}} = 0.037$; pooled $p = 0.010$; Figure 11B), with the advantage persisting across successive cycles, consistent with adaptive tuning of resource acquisition strategies.

Homeostatic regulation. Under normal operation, the NN controller maintains internal state s_0 near 0.99; disabling homeostasis causes monotonic decay ($h_{\text{decay}} = 0.01$/step) with high inter-organism variance, confirming active regulation rather than static parameterization.

Graded Metabolic Ablation

To test whether criterion ablation effects reflect a continuous functional relationship rather than a binary on/off artifact, we sweep the metabolism efficiency multiplier over $\{1.0, 0.75, 0.5, 0.25, 0.0\}$ (n=30 per level, 1,000 steps, graph metabolism). Figure 7 (left) shows a monotonic dose-response: population viability decreases proportionally with metabolic efficiency (Jonckheere-Terpstra trend test, $p < 0.001$). This
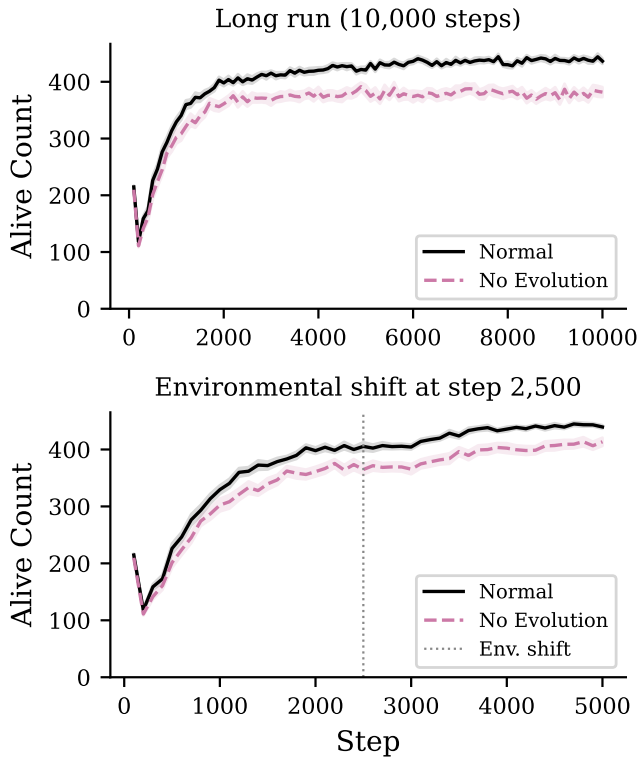
Figure 6: Evolution strengthening. Top: 10,000-step long run. Bottom: environmental shift at step 2,500 (dashed line). Lines: mean across 30 seeds; bands: ± 1 SEM.

graded response demonstrates that metabolic contribution is quantitatively proportional, not merely a binary switch—partial impairment produces proportional degradation, consistent with genuine functional analogy.

Cyclic Environment

To assess whether evolved populations exhibit adaptive resilience beyond static conditions, we subject organisms to periodic resource modulation (period = 2,000 steps; high phase: 0.01, low phase: 0.005 resource per cell per step) over 10,000 steps (n=30). Evolved populations ($\bar{x}$=384.3) significantly outperform non-evolving populations ($\bar{x}$=340.6) across all cycles (d=1.70, Cliff's δ=0.74, $p < 0.001$; Figure 7, right), showing faster recovery after each low-resource phase.

Sham ablation control. To confirm that observed ablation effects are functional rather than computational artifacts, we implement a state-neutral sham process that performs spatial neighbor queries matching the computational cost of a real criterion update but discards all results without consuming RNG state or modifying simulation variables. Comparing sham-on versus sham-
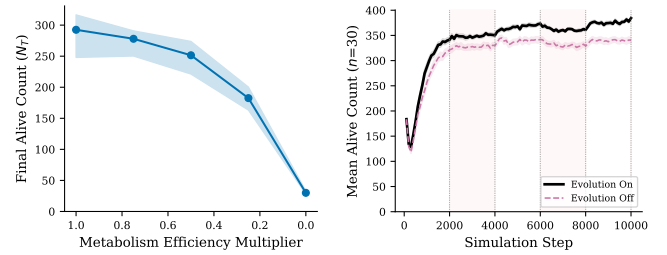

Figure 7: Left: graded metabolic ablation dose-response (median $\pm$ IQR, n=30 per level; $p < 0.001$, Jonckheere-Terpstra). Right: cyclic environment test (period=2,000 steps); shaded bands mark low-resource phases; evolution-on populations recover faster.

off (n=30, 1,000 steps) yields no significant difference (Mann-Whitney U, $p > 0.05$), validating that criterion ablation effects reflect genuine functional dependencies.

Additional Validation

Phenotype clustering. PCA projection of per-seed organism traits (energy, waste, boundary integrity, genome diversity, generation count) reveals two emergent phenotypic clusters among evolved populations (k=2, silhouette=0.41; Figure 8). To test whether individual organisms differentiate into persistent ecological strategies, we run 5,000-step simulations (n=10 seeds) and collect per-organism trait snapshots (energy, waste, boundary integrity, maturity, generation) at paired time windows separated by 200 steps (steps 2,000/2,200 and 4,500/4,700). Organisms present in both snapshots of a pair (n=1,262 shared) are matched by stable ID and clustered independently via k-means (k=2) on standardized traits (silhouette 0.39–0.51; Figure 9). Despite meaningful within-snapshot cluster separation (two distinct strategies: high-energy/low-boundary vs. low-energy/high-boundary), the adjusted Rand index between paired windows is low (ARI=0.04–0.09), indicating that individual organisms transition between ecological roles even over short ($\sim$200-step) timescales. Using a pre-registered persistence threshold (stronger language allowed only when ARI ≥ 0.30), these values remain below the claim gate, so we report this as weak persistence rather than stable individual ecological differentiation. Cross-pair analysis (steps 2,000 to 4,700) finds near-zero organism overlap, confirming that median lifespan ($\sim$245 steps) prevents long-horizon individual persistence. This suggests that ecological strategy differentiation in our system is a population-level rather than individual-level phenomenon: organisms continuously explore the trait space, with the population maintaining a stable distribution of strategies even as individuals transition between them—consistent with the moderate evolution
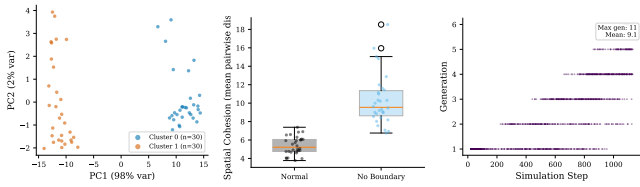
Figure 8: Additional validation. Left: phenotype clustering via PCA (*k*-means, silhouette-optimal *k*). Center: spatial cohesion under boundary ablation. Right: lineage phylogeny (generation depth vs. step).
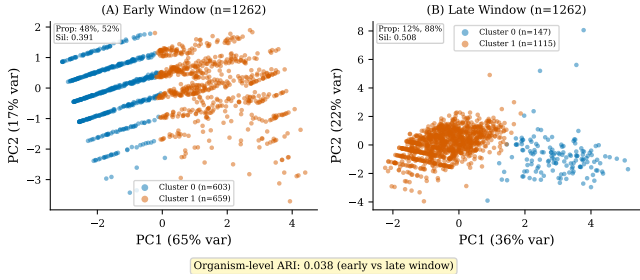

Figure 9: Per-organism ecological niche persistence. PCA projections of individual organism traits (energy, waste, boundary integrity, maturity, generation) at early (left) and late (right) time windows, colored by *k*-means cluster (*k*=2). Organisms are matched by stable ID across windows; the adjusted Rand index quantifies temporal persistence of cluster assignments.

effect (δ=0.34) observed in single-ablation experiments.

Spatial cohesion. Per-organism spatial cohesion (mean pairwise toroidal agent distance) under normal versus no-boundary conditions (*n*=30, 2,000 steps) confirms that boundary maintenance produces measurable spatial coherence, not merely scalar tracking (Figure 8, center).

Lineage structure. Parent–child tracking across all reproduction events yields multi-generational lineage trees (Figure 8, right), confirming genuine hereditary structure rather than random replacement.

Discussion

Criterion interdependence. Single ablations reveal a hierarchy: reproduction, response, and metabolism form an essential triad (>86% decline), while growth and homeostasis occupy a middle tier ($\sim$46%), followed by boundary ($\sim$31%). Pairwise ablations show uniformly sub-additive interactions (Table 5), consistent with shared failure pathways rather than independent modules—criteria converge on overlapping viability subsystems. The proxy control comparison fur-
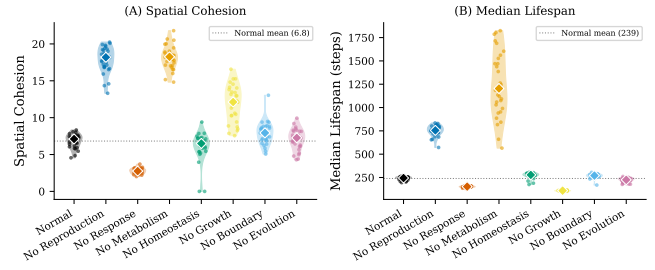

Figure 10: Criterion-orthogonal outcome measures (*n*=30 per condition). Left: spatial cohesion (mean pairwise agent distance); lower values indicate tighter spatial structure. Right: median organism lifespan per seed. Diamonds mark medians; dotted line shows normal baseline. Response ablation significantly increases cohesion (immobile organisms cluster) while decreasing lifespan.
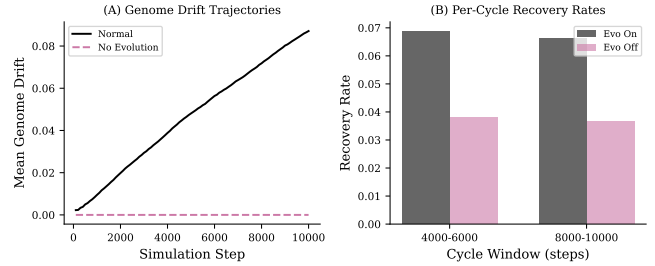

Figure 11: Evolution evidence. (A) Genome drift trajectories over 10,000 steps: evolved populations (black) accumulate genetic change linearly while non-evolved populations (purple) remain at zero (δ=1.00). (B) Per-cycle recovery rates under cyclic resource stress: evolved populations recover faster after each low-resource phase (pooled *p*=0.010).

ther demonstrates that the specific implementation of a criterion shapes ecological dynamics (more complex metabolism sustains fewer organisms with distinct diversity profiles), ruling out tautological definitions.

Evolution at longer timescales. The modest 2000-step evolution effect (*d*=0.58, δ=0.34) grows to *d*=1.42 (δ=0.66) at 10,000 steps and *d*=0.86 (δ=0.50) under environmental perturbation, confirming that evolutionary adaptation accumulates across generations. The cyclic environment test (Figure 7, right) further demonstrates that evolved populations recover more rapidly from recurring resource stress, indicating adaptive resilience beyond one-time environmental shifts.

Are criteria merely engineered? Four lines of evidence argue against this: (1) proxy control shows alternative implementations produce distinct dynamics (Fig-

ure 5); (2) pairwise ablations reveal sub-additive interactions inconsistent with independent modules (Table 5); (3) sham ablation control produces no significant difference; (4) graded ablation yields proportional dose-response (Figure 7, left).

Criterion maturity. Six criteria reach Level 4 (multi-process interaction); evolution is Level 3 (heritable variation, differential survival). Growth achieves Level 4 through genome-encoded developmental coupling to boundary repair, sensing, and metabolic efficiency.

Limitations

Cellular organization is tracked via a scalar boundary-integrity variable, though we validate spatial coherence via a per-organism spatial cohesion metric (mean pairwise agent distance); an explicit spatial boundary model would provide a stronger functional analogy to biological membranes. Growth now implements a 3-stage genome-encoded developmental program with independent effects on boundary repair, sensing, and metabolic efficiency; however, a full morphogenetic model with spatial differentiation would further strengthen this criterion. Evolution reaches d=1.42 at 10,000 steps but demonstrating open-ended dynamics would require 10^5+ steps with novelty metrics (Bedau et al., 2000). Scale is limited to ∼300 organisms on a single machine; larger populations might reveal emergent ecological phenomena. We adopt a weak ALife stance: this is a functional model, not a claim that digital organisms are alive.

Conclusion

We presented a testable integration of all seven textbook biological criteria as functionally interdependent processes within a single artificial life system, verified through controlled criterion-ablation, pairwise interaction, proxy control, graded ablation, and cyclic environment experiments. The functional-analogy framework—requiring dynamic operation, measurable degradation upon removal, and feedback coupling—provides a rigorous standard distinguishing genuine criteria implementations from simplified proxies.

Our results demonstrate that no criterion is decorative: removing any one causes statistically significant population decline ($p \leq 0.013$, Holm-Bonferroni corrected), with Cliff's δ ranging from 0.34 to 1.00. Pairwise ablations further reveal shared failure pathways—sub-additive interactions consistent with criteria converging on overlapping viability subsystems rather than operating independently.

Future work will pursue three directions: (1) scaling to larger populations to investigate emergent ecological phenomena and open-ended evolution metrics (Bedau et al., 2000; Taylor et al., 2016); (2) spatial morphogenesis extending the current developmental program with position-dependent cell differentiation; and (3) systematic environmental perturbation studies to characterize adaptive capacity across evolutionary timescales.

Code and data will be made available upon acceptance at an anonymous repository.

References

Barbieux, A. and Canaan, R. (2024). Coralai: Intrinsic evolution of embodied neural cellular automata ecosystems. In The 2024 Conference on Artificial Life, ALIFE 2024. MIT Press.

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., and Ray, T. S. (2000). Open problems in artificial life. Artificial Life, 6(4):363–376.

Chan, B. W.-C. (2019). Lenia: Biology of artificial life. Complex Systems, 28(3):251–286.

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin, 114(3):494–509.

Gánti, T. (2003). The Principles of Life. Oxford University Press, Oxford.

Heinemann, C. (2008). Artificial life environment. Informatik-Spektrum, 31(1):55–61.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2):65–70.

Joyce, G. F. (1994). Foreword. In Deamer, D. W. and Fleischaker, G. R., editors, Origins of Life: The Central Concepts, pages xi–xii. Jones and Bartlett, Boston.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1):50–60.

Maturana, H. R. and Varela, F. J. (1980). Autopoiesis and Cognition: The Realization of the Living. Boston Studies in the Philosophy and History of Science. Springer Netherlands.

McMullin, B. (2004). Thirty years of computational autopoiesis: A review. Artificial Life, 10(3):277–295.

Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. Artificial Life, 10(2):191–229.

Plantec, E., Hamon, G., Etcheverry, M., Oudeyer, P.-Y., Moulin-Frier, C., and Chan, B. W.-C. (2023). Flow-lenia: Towards open-ended evolution in cellular automata through mass conservation and parameter localization. In The 2023 Conference on Artificial Life, ALIFE 2023. MIT Press.

Ray, T. S. (1991). An approach to the synthesis of life. In Langton, C. G., Taylor, C., Farmer, J. D., and Rasmussen, S., editors, Artificial Life II, volume XI of Santa Fe Institute Studies in the Sciences of Complexity, pages 371–408, Redwood City, CA. Addison-Wesley.

Ruiz-Mirazo, K., Peretó, J., and Moreno, A. (2004). A universal definition of life: Autonomy and open-ended evolution. Origins of Life and Evolution of the Biosphere, 34(3):323–346.

Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickinbotham, S., Ikegami, T., McMullin, B., Packard, N., Rasmussen, S., Virgo, N., Agmon, E., Clark, E., McGregor, S., Ofria, C., Ropella, G., Spector, L., Stanley, K. O., Stanton, A., Timperley, C., Vostinar, A., and Wiser, M. (2016). Open-ended evolution: Perspectives from the OEE workshop in York. Artificial Life, 22(3):408–423.

Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Reece, J. B. (2020). Campbell Biology. Pearson, New York, 12th edition.

Yaeger, L. S. (1994). Computational genetics, physiology, metabolism, neural systems, learning, vision, and behavior or PolyWorld: Life in a new context. In Langton, C. G., editor, Artificial Life III, pages 263–298, Redwood City, CA. Addison-Wesley.