Exploring Hierarchical Evolution with an Artificial Protocell (Harder than it might seem!)

Barry McMullin
Dublin City University
Barry.McMullin@dcu.ie

3rd Workshop on MAS in Biology at meso or macroscopic scales
LaBRI , Bordeaux, 17 June 2010
(FoilTEX Presentation)

Acknowledgements

- DCU Alife Lab:
  - Ciarán Kelly
  - Darragh O'Brien
  - George G. Mitchell
  - James Decraene
- Financial support under EU FP6 Integrated Project PACE
  Contract number: 002035 www.istpace.org

What is a *protocell*?

Three (integrated, co-dependent) sub-systems:

- Metabolism
- Containment (membrane)
- Information (heredity?)

The models presented in this talk involve only Information and Containment (sort-of). So strictly only "proto-protocells"!

Template-Replicator Worlds

- Abstraction of RNA world
- Binary string "molecules" (variable length)
- Well stirred reactor (random collisions, or "simplex space")
- Replicase function only (**why?**):
  - S: substrate (= template)
  - E: enzyme (= replicase)
  - $E + S \rightarrow E + 2S$
- Investigate two distinct evolutionary scenarios:
  - Pure molecular, single level, selection: Fixed-size buffered reactor (displace random molecule); *or*
  - Hierarchical, coupled, protocell level selection: no molecular displacement, cells grow up to threshold size, then split (random assortment) and displace random cell
  - Assume reaction rates are either 1 (binds) or 0 (doesn't bind) (**why???**)

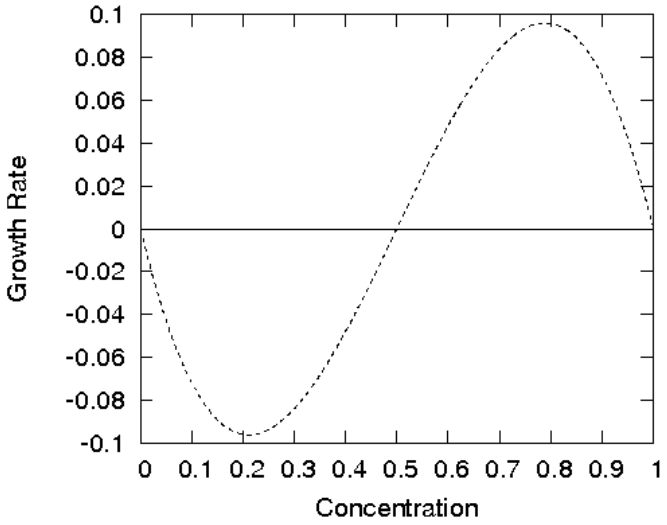
Replicase Binding Rule (Mark-1)

- Need a rule for whether any given E can bind to (and replicate) any given S
- Simple idea: bind if E and S are "similar"
- Simplistic idea: bind if E and S are identical
- (All-or-nothing "deterministic" binding ... discuss!)
- $\Rightarrow$ *Everything* is an independent *self*-replicase (replicator)! ☺
- Survival of the fittest (molecular species):
  - *Fidelity, Fecundity, Longevity!?*
  - Well . . . only *Fidelity!* in our case?

Not quite so simple!

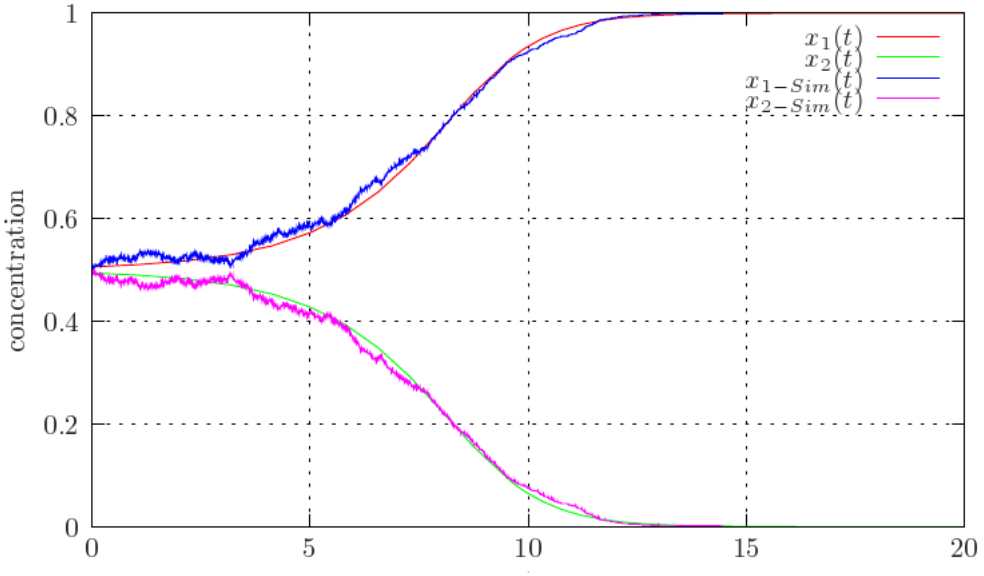
- $2S \rightarrow 3S$ (*not* $S \rightarrow 2S$)
- Degenerate, 1-element, hypercycle
- Hyperbolic (super-exponential) growth
- *Survival of the common*
- (Ignoring replication error/mutation for now...)

In pictures



Hyperbolic growth: Positive frequency dependence

In pictures



Survival of the common

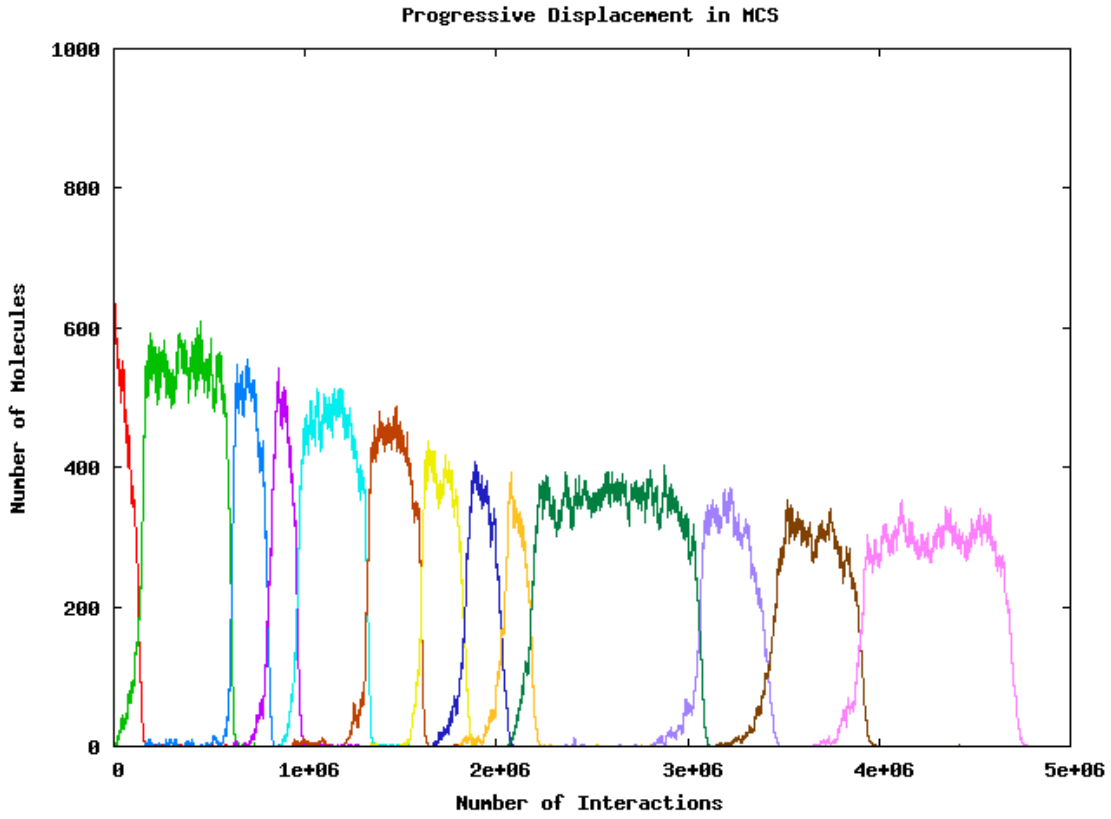
Now add mutation?

- Per bit error rate: m
- Mainly (99%) bit-flip, but occasional (1%) insert/delete
- Per molecule (length n) error rate: $M = 1 - (1 - m)^n$
- Every mutant is also a self-replicase
- *But* hyperbolic growth makes invasion very difficult
- Steady state mutant population of $1 - M$
- Dominant species can stay "dominant" even at high M (e.g., $M > 0.5$)
- So far, so boring: no molecular level "evolution"; no point in even examining protocell level evolution!

Replicase Binding Rule: Mark-2

- Let's (minimally) generalise the binding rule to allow *some* cases of "cross-replication"
- Bind (and replicate) if E is a *substring* of S
- Introduces parasitism/exploitation
  - Say P longer than H
  - Then H replicates P but P does not reciprocate
  - P is a (facultative) parasite of H
  - P can invade and displace H from "arbitrarily" small concentration (even just 1 molecule, albeit stochastically): overrides previous "survival of the common"

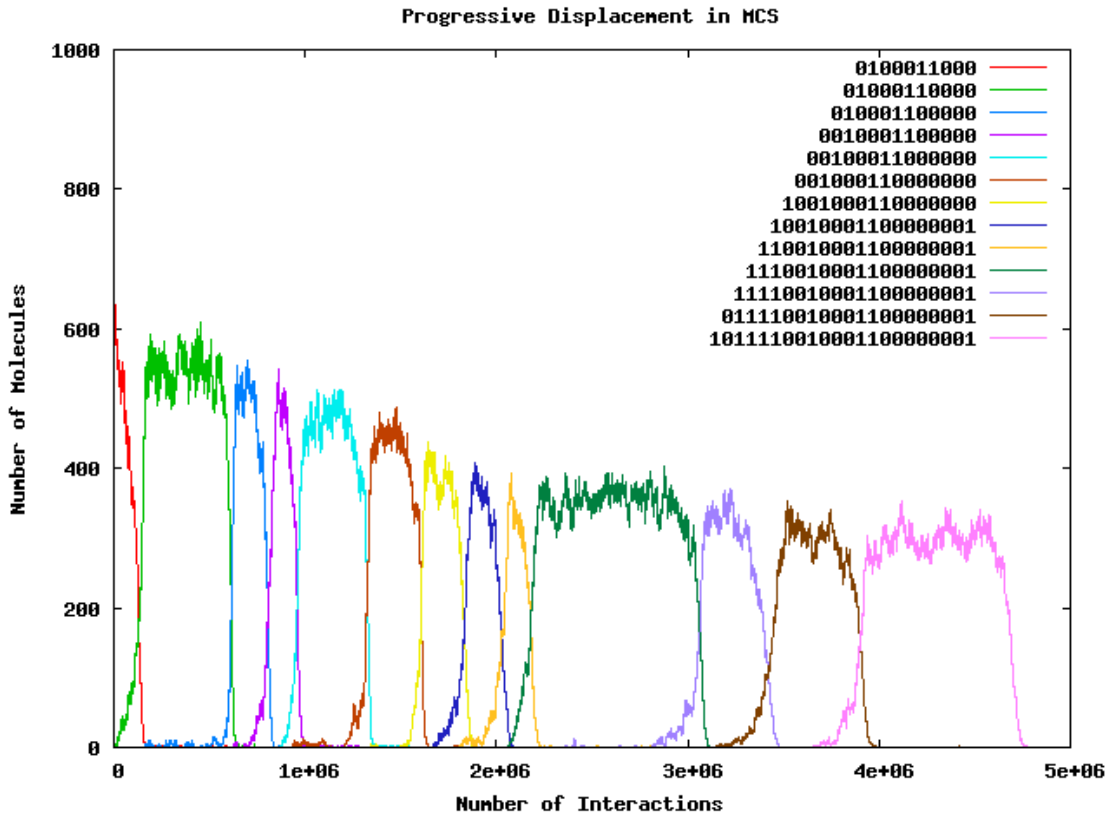
Macro-evolutionary Dynamic



Macro-evolutionary Dynamic

- Succession of micro-evolutionary, Darwinian (?), selection events
- Climbing mount improbable, just as expected?
- Well ... no!

Macro-evolutionary Dynamic



Macro-evolutionary Dynamic

- Displacements always by parasitic, incrementally longer strings.
- Increased length means increased error rate (M); i.e. reduced fidelity.
- (Reduced fidelity also implies higher mutational load, so lower steady state concentration and thus lower average fecundity; albeit exactly compensated by increasing longevity in quasi-steady-state...)
- So (Dawkinsian) "fitness" actually *decreases* monotonically!

Macro-evolutionary Dynamic

- At the very least, it underlines that the Dawkins' slogan ("Longevity, Fecundity, Fidelity") is somewhat (very?) simplistic.
- We suggest that it was worthwhile to isolate and characterise this phenomenon clearly *before* adding additional complications.
- But best of all: it immediately offers *a simple candidate problem for "solution" by protocell level selection.*

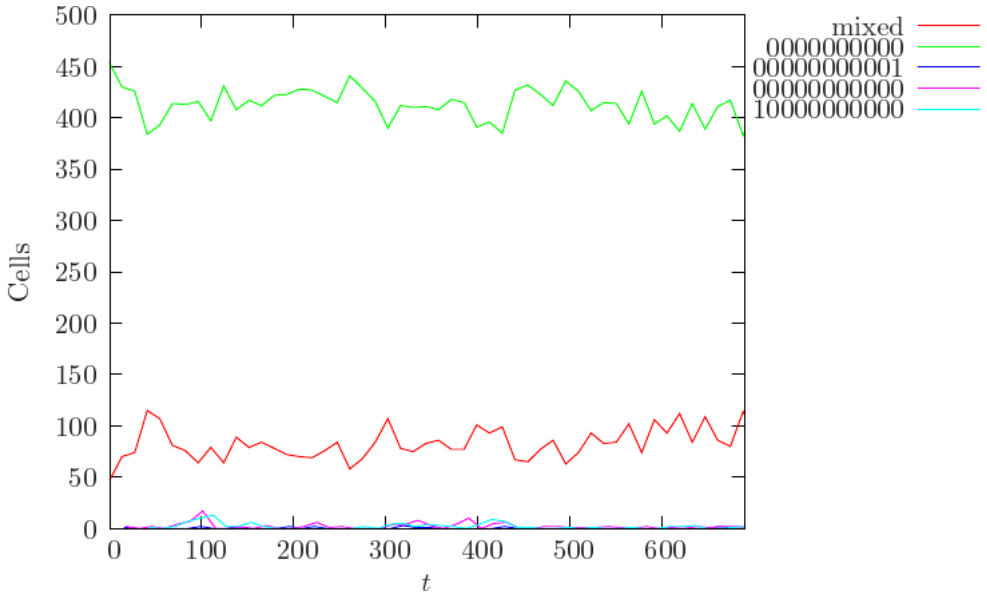
(Proto-)cells to the Rescue!

- So we introduce compartmentalisation, i.e., a hierarchical protocell level of selection ("major transition")
- Reprise:
  - Compartment is no longer a fixed-size buffered reactor: instead it just grows bigger (it is a "cell")
  - At a critical (ad hoc) size, cell reproduces by fission (random distribution of molecules to offspring cells)
  - Fixed size population of *cells* (so each fission event also displaces a random cell)
  - Selection at molecular level now interacts with selection at cell level

Prediction?

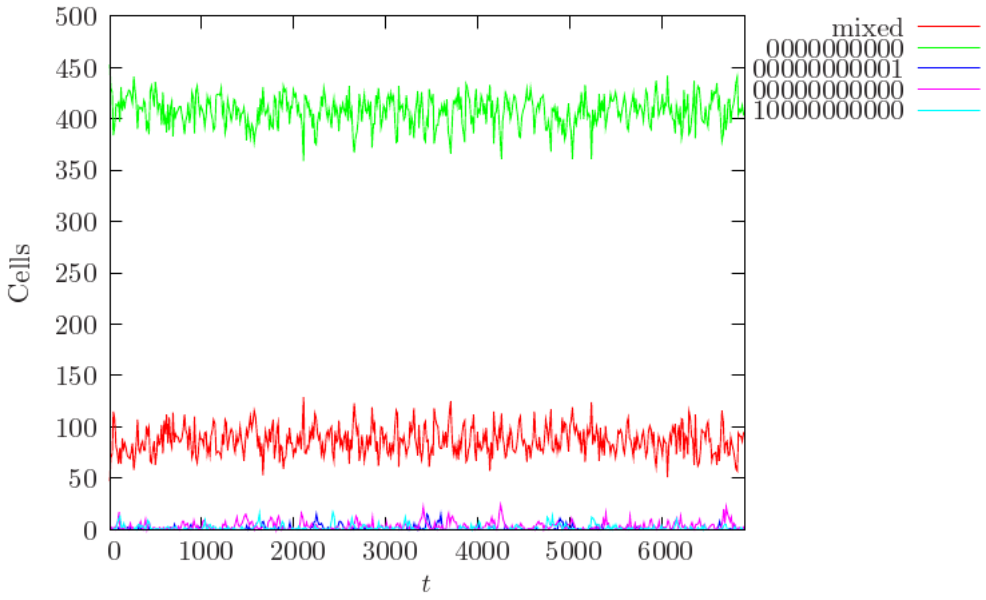
- Molecular level selection tends toward longer molecular species
- Cell level selection tends toward (domination by) shorter molecular species
- *Stalemate!*

Test? (short run)



Cell Level Selection: Evolutionary Stalemate

Test? (long run)



Cell Level Selection: Evolutionary Stalemate

Little steps. . .

- Could trivially reverse the Mark-2 binding rule to be "superstring" rather than "substring":
  - molecular and cellular selection exactly aligned (rather than opposed)
  - "evolves" (collapses!) to either `0` or `1`
  - but has no more general interest. . .
- Better idea: generalise just enough to allow *both* longer and shorter faculative parasites (relative to their hosts)
- Pure molecular level selection should show random shifts, both positive and negative, in the length of the dominant molecular species
- Protocell level selection would then, presumably, pick out the (cell-level) mutants dominated by shorter molecular species
- Macroevolutionary trend toward cells dominated by progressively shorter molecular species (due to increased molecular fidelity)

Optional Detour: Molecular Ecology 101!

- We have mentioned just two "ecological" relationships so far:
  - two independent self-replicases (single component hypercycles, survival of the common)
  - facultative parasitism (rapid selective displacement; invasion from rarity)
- But does that exhaust the possibilities? Surely not!

The Space of (Toy!) Replicase Dynamics

- What is the space of possible distinct dynamics in a fixed-size buffered reactor containing just two species of (potential) replicases?
- Toy version:
  - we are still assuming replication as the only enzymatic function; *and*
  - we neglect replication error/mutation error (otherwise the simplification to two species would immediately be violated; but the hope is that we will reintroduce mutation later as a rare perturbation to the "underlying", "intrinsic" dynamics); *and*
  - we are still assuming all rate constants are either 0 or 1; *but*
  - at least with all those simplifications, the space of distinct dynamic behaviours should be pretty manageable to review! ☺

The Space of (Toy!) Replicase Dynamics

- Each possible case is represented by a reaction matrix:

$$\begin{bmatrix} r_{XX} & r_{XY} \\ r_{YX} & r_{YY} \end{bmatrix}$$

- r_{AB} denotes the rate constant when applying A as a replicase to template B:
  - If A can bind to B then it replicates it, $A + B \longrightarrow A + 2B$. The rate constant is 1.
  - Conversely, if A cannot bind to B, the collision is elastic and the rate constant is 0.
  - Note that $r_{AB} \neq r_{BA}$ in general.

The Space of (Toy!) Replicase Dynamics

- Shortcuts interpretations:
  - $r_{XX} = 1$: X is SR
  - $r_{YY} = 1$: Y is SR
  - $r_{XY} = 1$: Y is a parasite of X
  - $r_{YX} = 1$: X is a parasite of Y
- Parasite + SR $\Rightarrow$ *facultative*
- Parasite + non-SR $\Rightarrow$ *obligate*

The Space of (Toy!) Replicase Dynamics

- For the fixed-size buffered reactor we have $x + y = 1$
- Given a reaction matrix we can then immediately write an approximate continuous ODE:

$$\dot{x} = (x^2 r_{XX} + xy r_{YX})y$$

- Fixed points: $\dot{x} = 0$

The Space of (Toy!) Replicase Dynamics

- Four reaction rates in the matrix
- Each is either 1 or 0
- $\Rightarrow$ exactly 16 matrices
- But some of these are identical modulo the arbitrary labelling of the molecular species
- By inspection, there are exactly 10 distinct possible dynamics
- We concentrate on:
  - "Incumbent", dominant, SR, X
  - In which cases will Y take over (invade from rarity)?

The Space of (Toy!) Replicase Dynamics

- Five distinct cases still remain (arbitrary numbering!):
  - Class 9: Independent self-replicases, incumbent advantage, survival of the common, no invasion.
  - Class 6: Y facultative parasite, rapid invasion and displacement from rarity.
  - Class 2: Y sterile; Y rapidly eliminated.
  - Classes 5, 7: Y non-SR, X facultative parasite; Y rapidly eliminated.
  - Classes 4, 1: Y obligate parasite of X; or both are facultative parasites of each other. $\dot{x} = \dot{y} = 0 \forall x, y$. Displacement by random drift theoretically possible, but large incumbent advantage (proportional to molecular population size), so only very occasional invasion from rarity. Can be safely neglected (?).
- So only two "original" cases (classes 6 and 9) really matter. Phew! [. . . end of detour!]

Replicase Binding Rules: Mark-3+?

- Have tentatively explored a series of "richer" binding rules
- All "inspired" by a notion of molecular "folding":
  - Primary structure: binary; inactive template conformation
  - Secondary structure: dibits or tribits code for symbols in "pattern language", active enzyme/replicase conformation
  - Spontaneous, rapid, folding/unfolding, $50 : 50$ probability of both conformations (very *un*-molecular?)

Replicase Binding Rule: Mark-3

- Reporting results for just one "trivial" (hah!) version. . .
- Pattern language contains just two symbols denoting single-bit literal matches.
- Still many possible choices!

Replicase Binding Rule: Mark-3

dibit	function	description
00	L	match literal 0
01	L	match literal 0
10	H	match literal 1
11	H	match literal 1

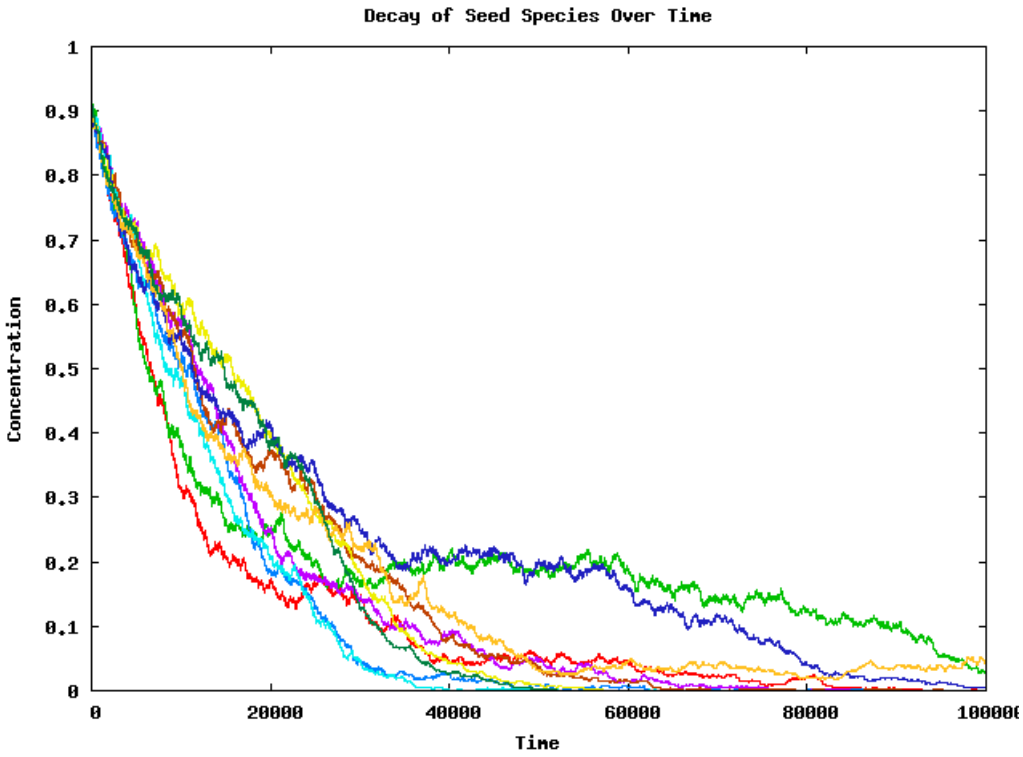
- We have verified:
  - Class 6 (Facultative Parasites), both lengthening and shortening, generally exist.
  - All other classes at least possible.

So: Did it "work"?

So: Did it "work"?

no

So: Did it "work"?

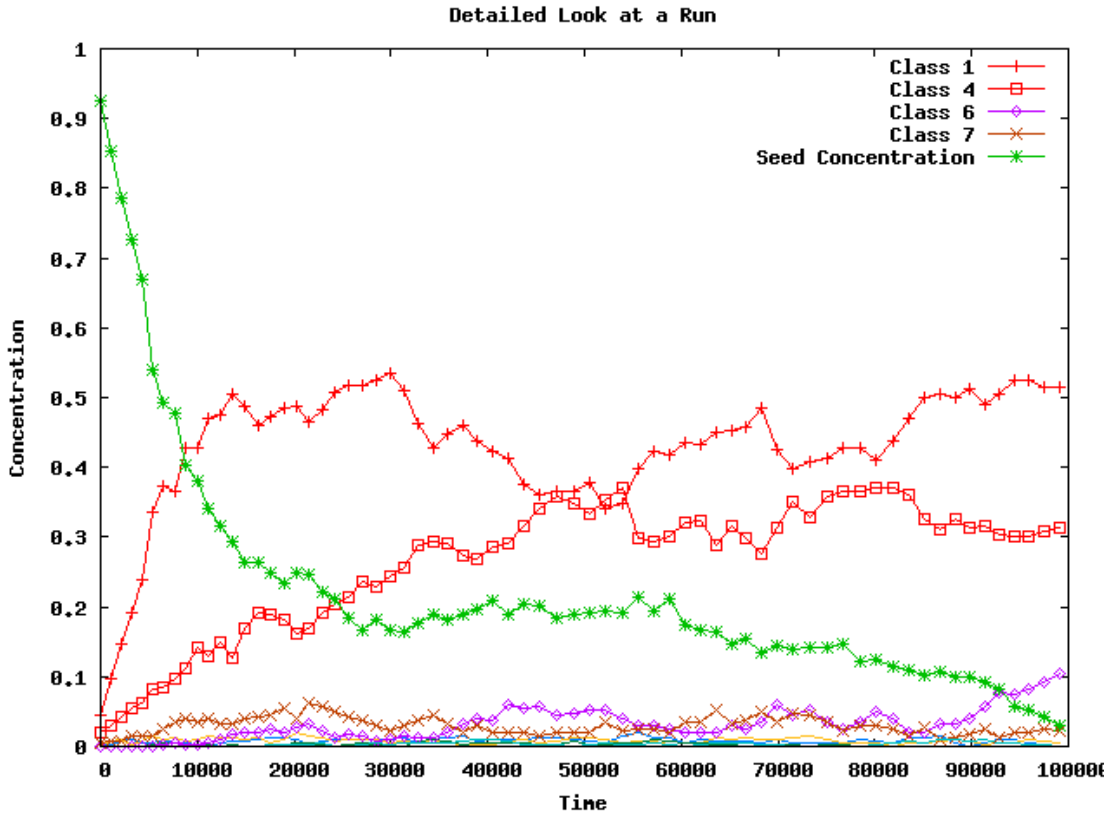


Collapse of dominant incumbent SR molecular species

Postmortem 1: Regression testing

- Locate an example of each reaction dynamics class.
- Seed reactor with just these two species, one already "dominant".
- Set replication error to zero (so that no other species will disrupt the dynamics).
- **Result:** in isolation, all classes behave as originally predicted.

Did it "work"? A closer look. . .



Postmortem 2

- Clearly one (or more) class 1 molecular species immediately and rapidly grow in concentration, at the cost of the original seed species.
- Two species are class 1 relative to each other if each replicates both itself and the other. In effect they cannot tell each other apart.
- But the predicted dynamic, on injection of such a species, was random drift ($\dot{x} = \dot{y} = 0$), apparently strongly favouring the already dominant incumbent.
- This dynamic *is* observed for class 1 pairs of species in isolation: so what is different in the "real" situation?

Postmortem 3 (20:20 Hindsight!)

- The original analysis (for classes 1 and 4) assumed an *unbiased* random drift between exactly two species . . .
- That is, replication error/mutation error should just add a little more noise to the underlying dynamics of drift.
- In fact, with this particular binding rule, for any SR species of length n, we estimate there will be c. $2^{(n/2)}$ other species within one mutational step which are all class 1 relative to it (and each other).
- So *aggregate* mutational outflow from the initially incumbent species will be (much) higher than inflow (by a factor of c. $30:1$ or more for $n = 10$).
- So mutation in this case is highly biased away from any single species to the ensemble of its class-1 mutants.
- After this immediate diversification, other effects will progressively contribute, including mutational bias toward a still larger ensemble including class 4 mutants.

Postmortem 4: A glimmer of light?

- While the details vary with other binding rules, this phenomenon was fairly robust across the (rather arbitrary) rules investigated.
- Reducing the absolute mutation rate ("within reason") simply slows down the decay of the original seed; but is not expected to change the ultimate outcome (at least in the single fixed-size buffered reactor case).
- Initial diversification into a set of fully class-1-connected species has no effect on overall reaction rate.
- But further diversification to only partially class-1-connected mutants and, especially, to class-4-connected mutants significantly impairs overall reaction rate.

What Next?

- General theoretical analysis seems difficult (?).
- In certain parameter ranges, hierarchical, protocell-level selection should still be able to limit the diversification (beyond the core, fully class-1-connected network associated with any SR-molecular species). This needs to be tested.
- *If* so: we may yet recover the intended, slightly richer, macro-evolutionary dynamic at the protocell population level.
- We would still like to investigate evolution of more "interesting" protocell function (e.g., cell cycle control?). (Need to expand the repertoire of molecular function ...)

Back to the "real" world?

Self-Sustained Replication of an RNA Enzyme

Tracey A. Lincoln and Gerald F. Joyce*

An RNA enzyme that catalyzes the RNA-templated joining of RNA was converted to a format whereby two enzymes catalyze each other's synthesis from a total of four oligonucleotide substrates. These cross-replicating RNA enzymes undergo self-sustained exponential amplification in the absence of proteins or other biological materials. Amplification occurs with a doubling time of about 1 hour and can be continued indefinitely. Populations of various cross-replicating enzymes were constructed and allowed to compete for a common pool of substrates, during which recombinant replicators arose and grew to dominate the population. These replicating RNA enzymes can serve as an experimental model of a genetic system. Many such model systems could be constructed, allowing different selective outcomes to be related to the underlying properties of the genetic system.

www.sciencemag.org **SCIENCE** VOL 323 27 FEBRUARY 2009

Conclusions

. . . while the choice [of the "elementary parts"] is enormously important and absolutely basic for the application of the axiomatic method, this choice is neither rigorously justifiable nor humanly unambiguously justifiable. All one can do is to try to submit a system which will stand up under common sense criteria.

— John von Neumann (*Theory and Organization of Complicated Automata*, Illinois, 1949)

Conclusions

Evolution is more complicated than you think.

Conclusions

Apply Hofstader's Law!

Related Online Resources

- Presentation slides:
  - `http://www.eeng.dcu.ie/~alife/talks/meso-2010/`
- DCU Alife Laboratory:
  - `http://www.eeng.dcu.ie/~alife/`
- Rince (National Engineering Research Institute):
  - `http://www.rince.ie/`

Copyright

Copyright