Evolution of complexity from the perspective of learning and the origin of life

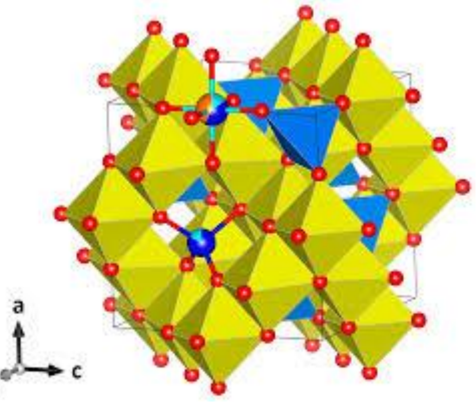
Eugene V. Koonin

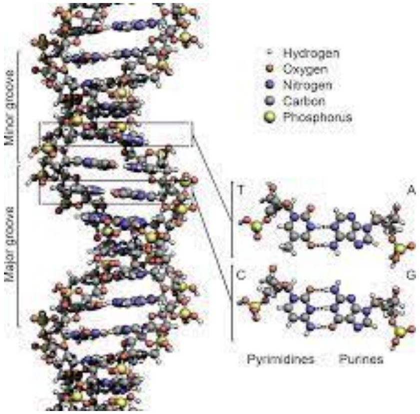
National Center for Biotechnology Information, NLM, NIH

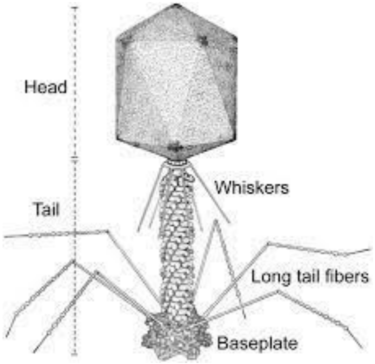
Evolution of Complexity from the Statistical Physics Perspective

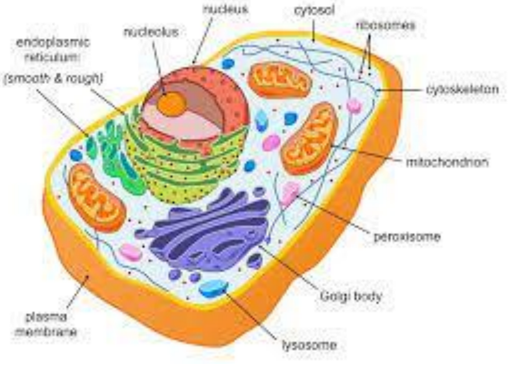
July 1 2022, Yerevan

The unprecedented complexity of life









complexity

S. Lloyd, “Measures of complexity: a non-exhaustive list” – 40 different definitions

“I know it when I see it”
Justice Potter Stuart on definition of obscenity (1964)

Evolution of complexity, multilevel learning and thermodynamics

- Wolf YI, Katsnelson MI, Koonin EV. **Physical foundations of biological complexity.** *Proc Natl Acad Sci U S A*. 2018 Sep 11;115(37):E8678-E8687

- Vanchurin V. 2020. **The World as a Neural Network.** *Entropy* 22(11):1210

- Vanchurin V, Wolf YI, Katsnelson MI, Koonin EV. **Toward a theory of evolution as multilevel learning**. *Proc Natl Acad Sci U S A. 2022 Feb 8;119(6):e2120037119.*

- Vanchurin V, Wolf YI, Koonin EV. Katsnelson MI. **Thermodynamics of evolution and the origin of life.** *Proc Natl Acad Sci U S A. 2022 Feb 8;119(6):e2120042119*

What is life? What do we want to explain? The signatures of biology

- 1. **Distinct information processing units** – cells, organisms. Selection for persistence/stability.
- 2. **Frustration** – conflicting objectives on different scales – e.g., individual cells vs multicellular organism – major driver of complexity evolution – general physical phenomenon, e.g. origin of patterns in spin glasses (Wolf et al 2018).
- 3. **Multilevel hierarchy** of scales/multilevel selection.
- 4. **Near optimality** – stochastic optimization – local minima on complex, rugged fitness landscape
- 5. **Diversity of near optimal solutions** - rugged fitness landscapes

Vanchurin, Wolf, Katsnelson, EK, PNAS 2022

What is life? What do we want to explain? The signatures of biology

- 6. **Separation of genotype from phenotype** – digital vs analogue information – feedback, asymmetric information flow (Central Dogma).
- 7. **Replication of digital information carriers.**
- 8. **Natural (biological) selection** – predicated on #1, #6, #7.
- 9. **Parasites – host-parasite coevolution** – emergence of parasites is inevitable and promotes complexity
- 10. **Programmed death** – general feature of all cells but occurs at other levels as well

Vanchurin, Wolf, Katsnelson, EK, PNAS 2022

More general principles of system evolution/ learning to explain complexity including biology

- P1. **Loss function (optimization)**. In any evolving/learning system, there exists a loss function of time-dependent variables that is minimized during evolution.
- P2. Hierarchy of scales. Evolving systems encompass multiple dynamical variables that change on different temporal scales.
- P3. Frequency gaps. Dynamical variables are split among distinct levels of organization separated by sufficiently wide frequency gaps - substantially different characteristic time scales/change rate.
- P4. Renormalizability. Across the entire range of organization levels of evolving systems, a statistical description of faster-changing (higher-frequency) variables is feasible through the slower-changing (lower-frequency) variables.

unique to life

- P5. Extension. Evolving systems have the capacity to recruit additional variables that can be utilized to sustain the system and the ability to exclude variables that could destabilize the system.
- P6. Replication and elimination. Evolving systems replicate and eliminate information-processing units (IPUs) on every level of organization.
- P7. Information flow. In evolving systems, slower-changing levels absorb information from faster-changing levels during learning and pass information down to the faster levels for prediction of the state of the environment and the system itself.

Vanchurin, Wolf, Katsnelson, EK, PNAS 2022

General evolutionary principles follow from learning dynamics/optimization

multilevel learning → **multilevel selection**

During system evolution, nearly constant variables become adaptable, yielding additional levels of learning – growing complexity

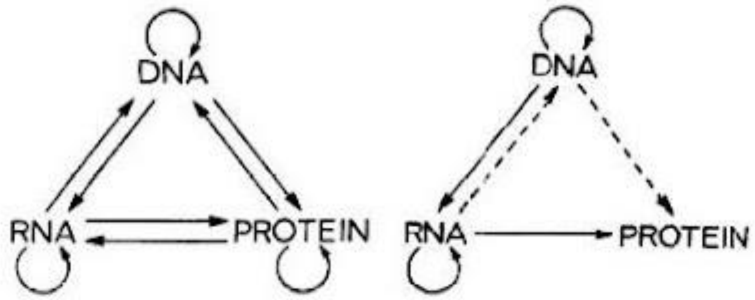
Renormalizability (P4) – same principles and equations (e.g. those for loss function) apply at all levels

Vanchurin, Wolf, Katsnelson, EK, PNAS 2022

Asymmetrical information flow – generalized Central Dogma of Molecular Biology (P7)

- 1) **Environment prediction phase**:
  fast information flow from slow changing variables (genome) to fast-changing variables (phenotype), through multiple layers
- 2) **Environment learning phase**: slow information flow from fast-changing to slow-changing variables – not a microscopic reversal of prediction phase (mutation-selection)
- **Slow variables have to be largely independent from fast variables to determine temporally stable rules that are essential for learning/evolution – separation between long-term (nucleic acids) and short term (proteins) memory - hence Generalized Central Dogma**

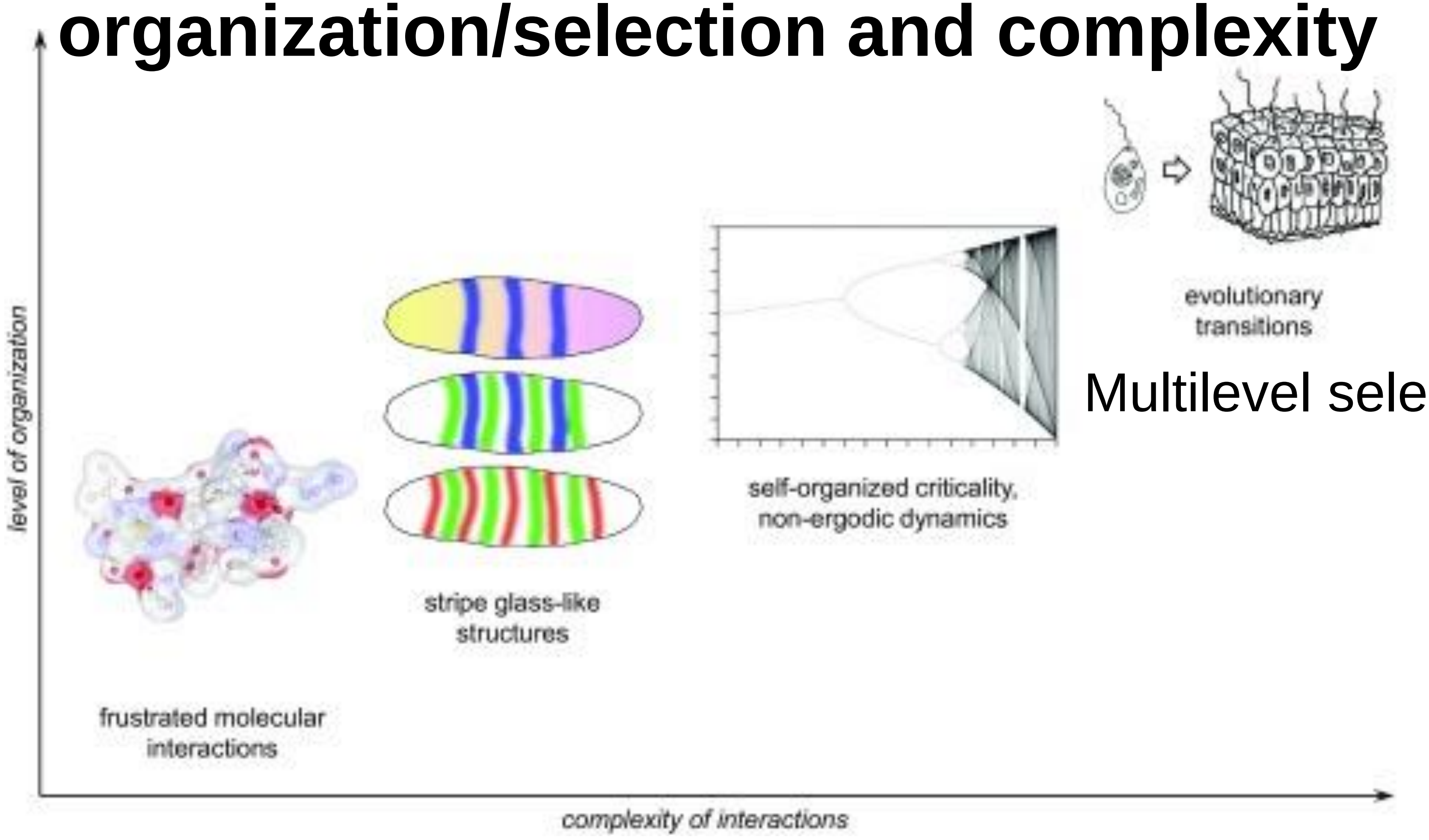
[informal explanation – formal derivation given in Vanchurin et al. PNAS 2022;
recall Takeuchi's
talk]



Crick, Nature 1970

Frustration: different learning objectives at different levels

From frustrated states to new levels of organization/selection and complexity



Multilevel selection

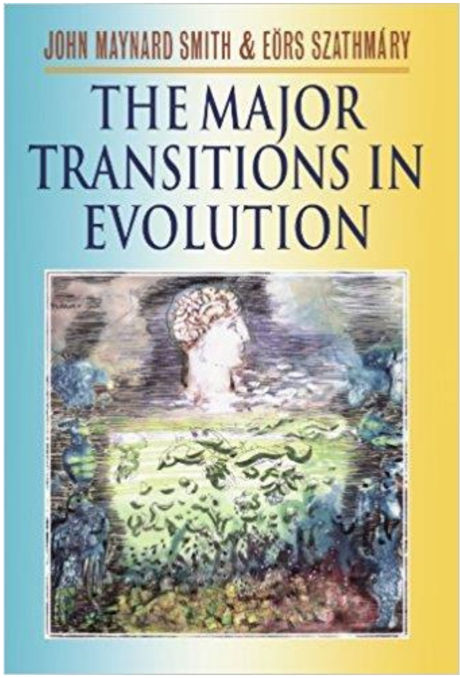
Wolf, Katsnelson, Koonin, PNAS 2018

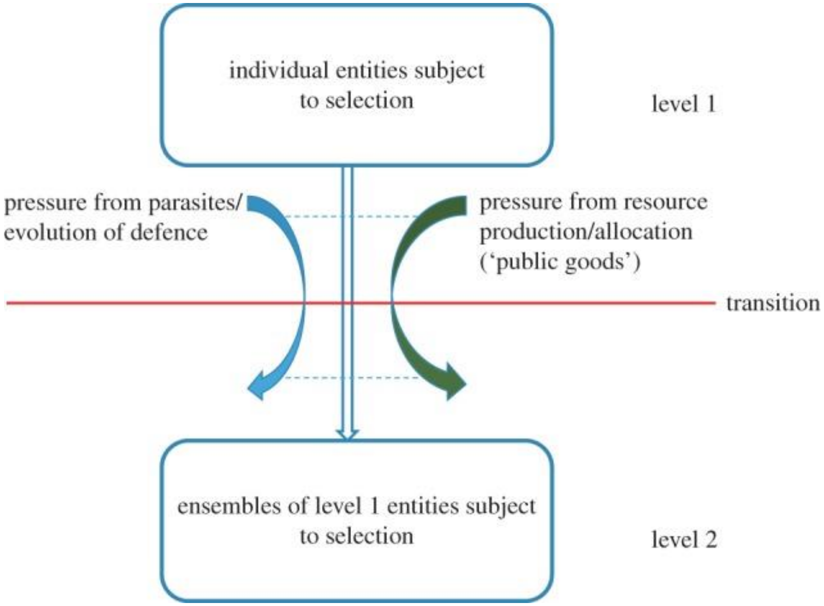
Major transitions in evolution

Evolutionary Transition in Individuality:

New levels of selection – frustration begets cooperation

- Origin of cells
- Eukaryotic cells
- Multicellularity
- Eusociality
- Society





Koonin Phil Trans Royal Soc 2016

Phenomenology of major transitions in evolution including origin of life

Origin of life: a phase transition that gave rise to a distinct, highly efficient form of learning - learning algorithm known as ***natural selection***

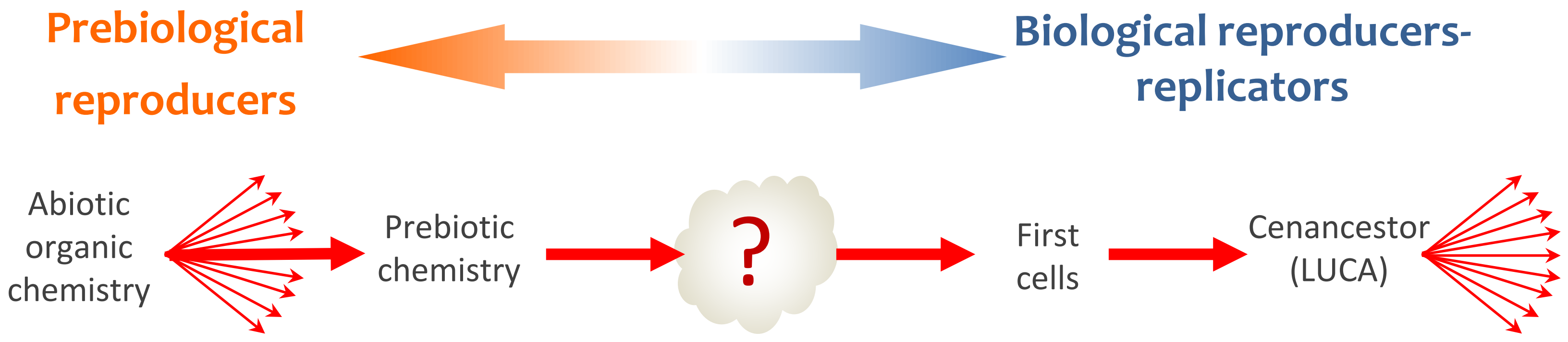
Transition occurs when

$$\Omega_p(\mathcal{T}_c, \mathcal{M}_c) = \Omega_b(T_c, \mu_c)$$

That is, grand potential of an ensemble of molecules becomes equal to grand potential of a biological system
$\mathcal{T}_c, \mathcal{M}_c$ - temperature, chemical potential;
T_c, μ_c – “evolutionary temperature”, evolutionary potential [briefest formulation, details in Vanchurin et al, PNAS 2022; recall Vitaly Vanchurin’s talk]

Later transitions can be described analogously

Origin of Life



Prebiological reproducers

Biological reproducers-replicators

Abiotic organic chemistry

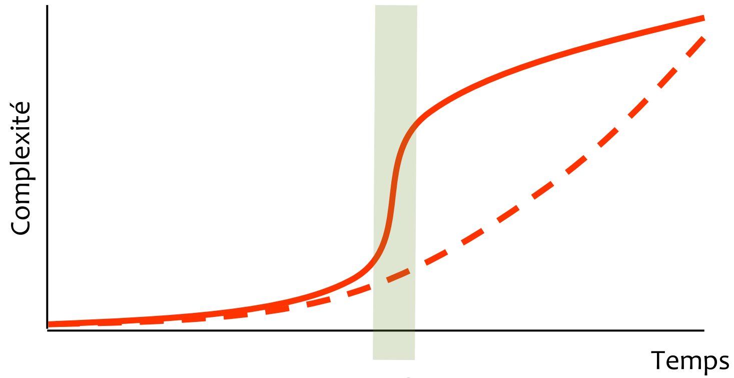
Prebiotic chemistry

?

First cells

Cenancestor (LUCA)

Transition from pre-life to life



Life

$$\Omega_p (\mathcal{T}_c,\mathcal{M}_c)=\Omega_b (T_c,\mu_c)$$

Courtesy of Puri Lopez-Garcia

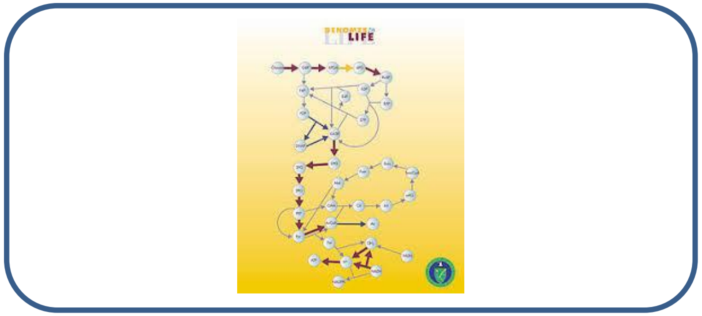
Puri Lopez-Garcia
Univ Paris-Saclay

Origins of cooperation and conflict: A mutualistic scenario for the origin of life under the multilevel learning perspective

- Two principal types of propagating biological entities
- 1. **Reproducers** – physical structure reproducing, genome replication insufficient – **cells**
- 2. **Replicators** – genome replication only – **mobile genetic elements including viruses but also cellular genomes**
- -***Origin of life: mutualistic merger of reproducers and replicators***

Origins of cooperation and conflict: A mutualistic scenario for the origin of life

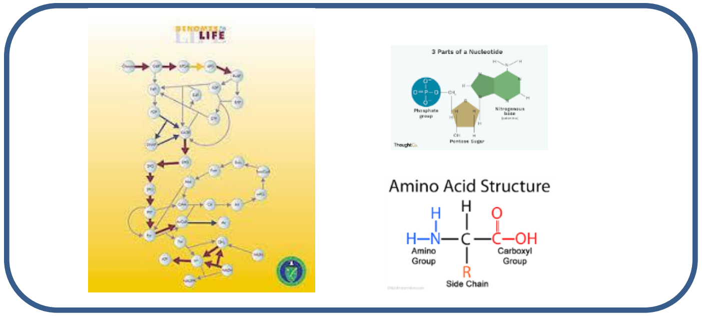
Learning:
Non-trainable variables: ***x(e), x(o)***
Trainable variables: ***q(a), q(n)*** – traits of reproducers

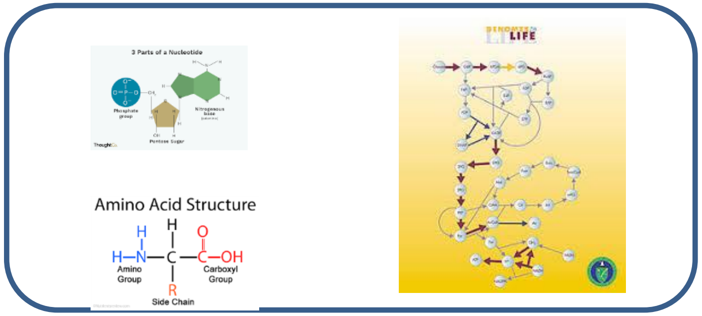


Pre-life
Proto-metabolic networks within vesicles/ compartments

Selection/learning without genomes:
Survival of the **most persistent**

Evolution of 'pure' reproducers

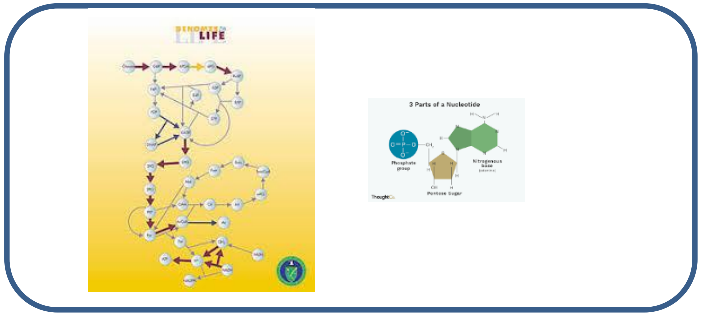




Nucleotides (and amino acids?) as cofactors:
Driving force for nucleotide (and amino acid?) accumulation

Origins of cooperation and conflict:
A mutualistic scenario for the origin of life

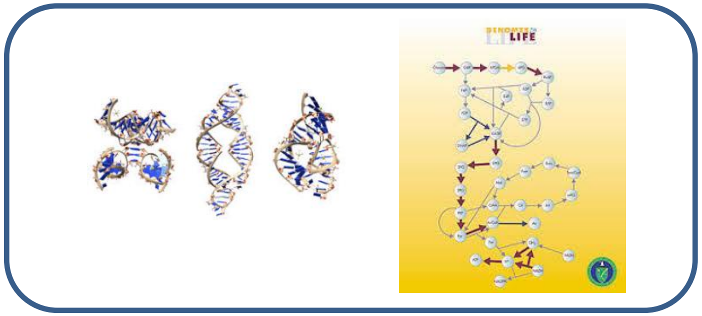
Learning: nucleotide accumulation enables polymerization:
Ribozyme catalysts
Including aminoacylating ribozymes

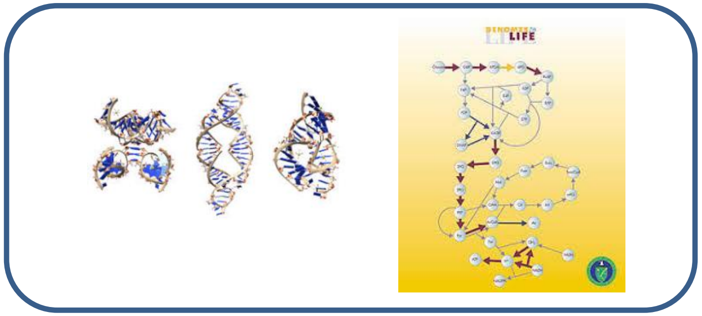


Pre-life

Selection without genomes:
Survival of the **most persistent**

Evolution of ‘pure’ reproducers

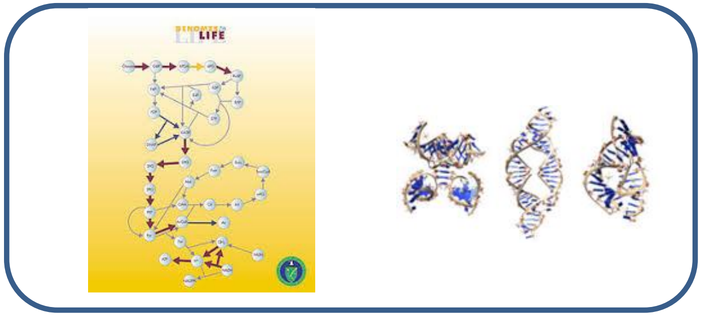




Concomitant non-templated peptide synthesis

Origins of cooperation and conflict:
A mutualistic scenario for the origin of life

nucleotide accumulation enables Polymerization: *ribozyme catalysts*

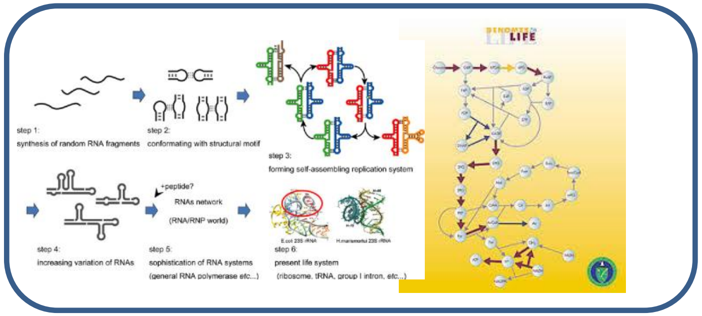


Pre-life

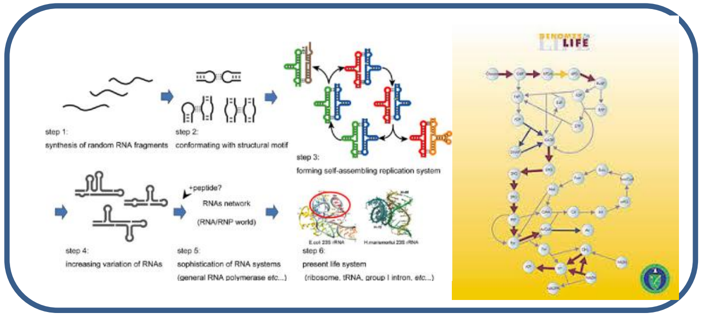
Learning - Origin of replicators:
Memorizing sequences of efficient ribozymes – selection for replication

Selection without genomes:
Survival of the most persistent

Concomitant non-templated peptide synthesis

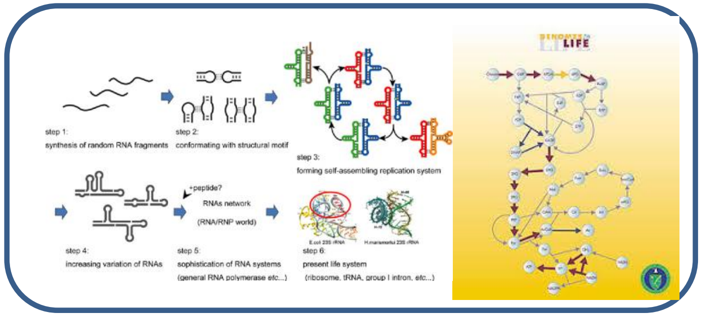


Proto-Life: Memory of the system state stored in RNA sequences



Origins of cooperation and conflict:
A mutualistic scenario for the origin of life

Emergence of **multilevel learning**:
New class of slow variables ***q(c):***
genomes/long term memory



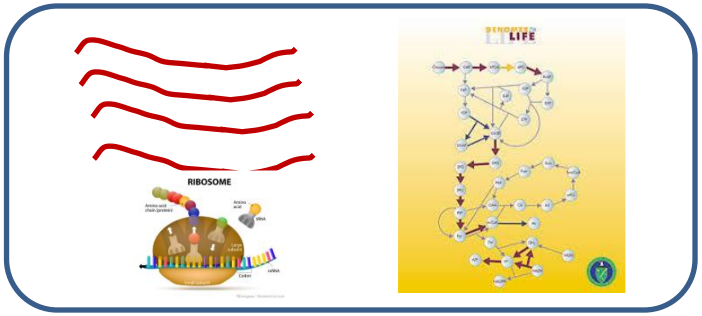
**Proto-
Life:
Memory
Of the
system
state**

- Origin of genomes:
  survival of the fittest
- Concomitant origin of translation – enzymes

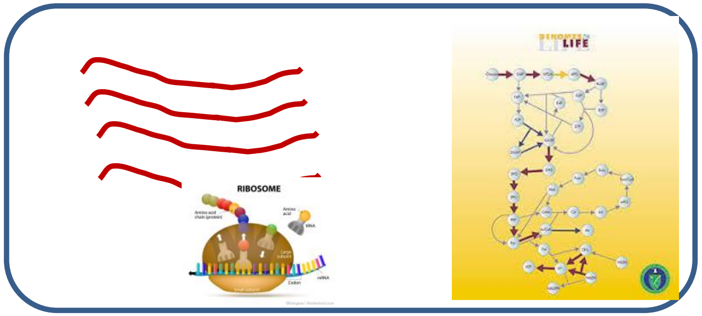
Reproducer components encoded by replicators

Evolution of replicators:
replicator-carrying protocells – **reproducer-replicator mergers** - outcompete those lacking replicators

Mutualistic relationship/coevolution between replicators and reproducers



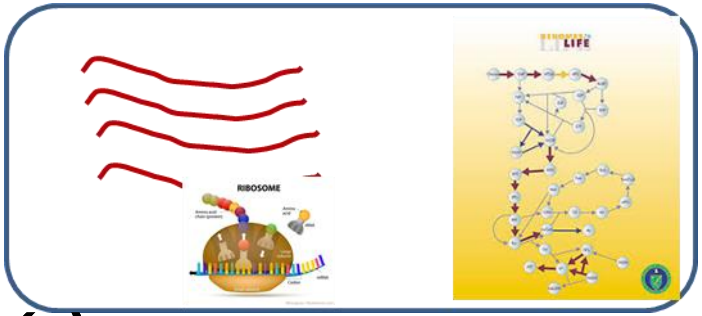
Life



Origins of cooperation and conflict: A mutualistic scenario for the origin of life

Sustained evolution enabled by replicators encoding components of reproducers:

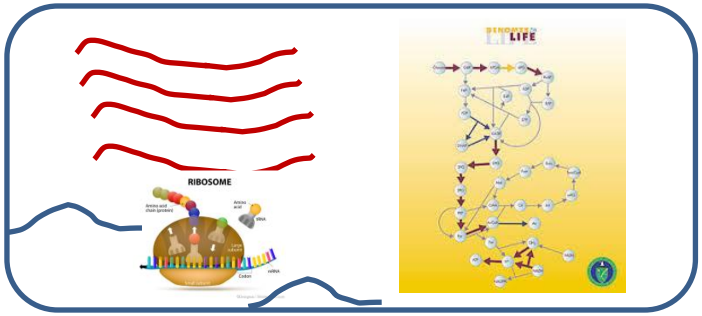
Stable core: ***q(a)*** ⟷ ***q(c)***



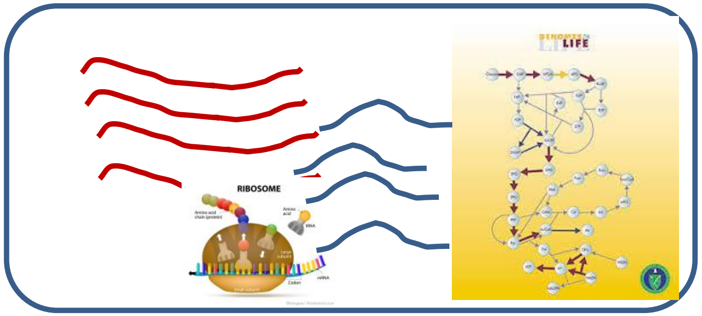
Early Life

Origin of parasites and biological conflict

Split of replicators into cooperators (cellular genomes) and parasites – concomitant with origin of efficient replication



Life

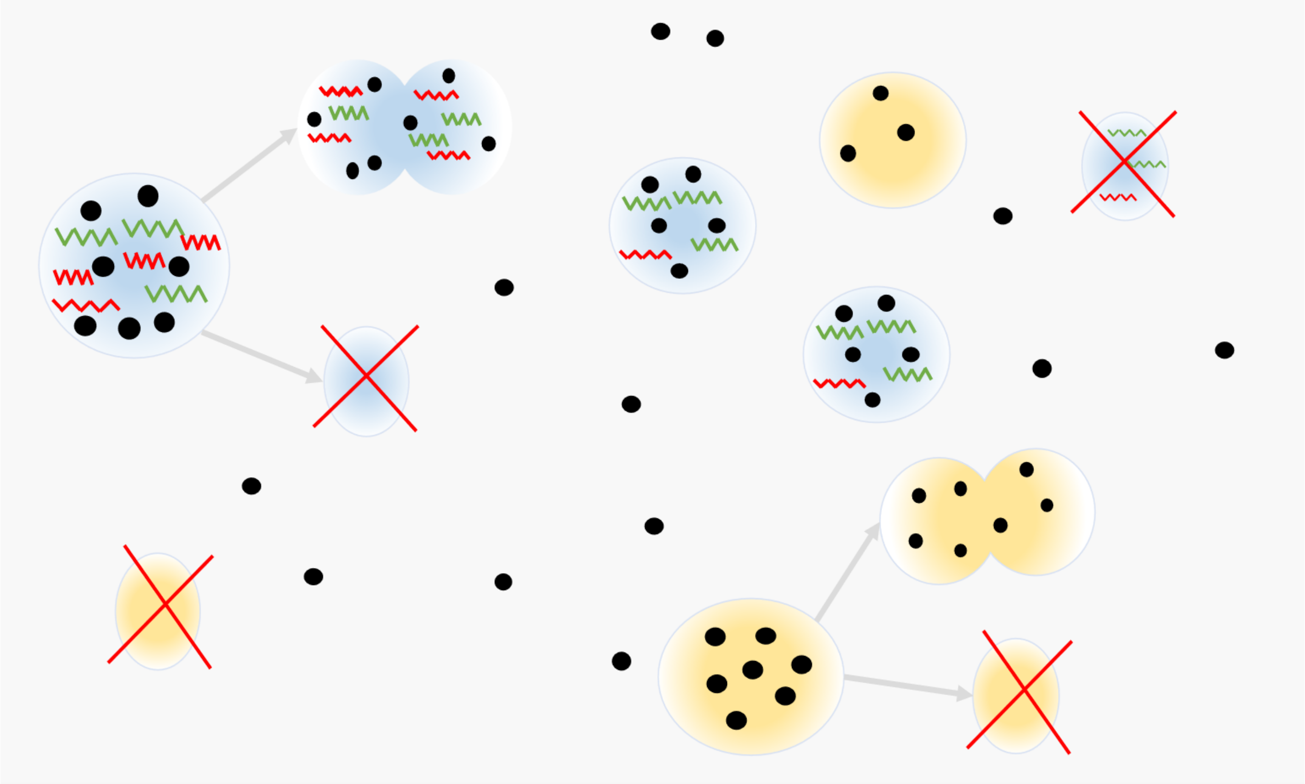


A mathematical model of the origin of life through a mutualistic symbiosis of reproducers and replicators: conditions for the origin of genomes

Babajanian, Wolf, Allahverdyan, Khachatryan, Lopez-Garcia, EK, in preparation

Conditions for the origin of genomes: competition between protocells containing and lacking genetic elements

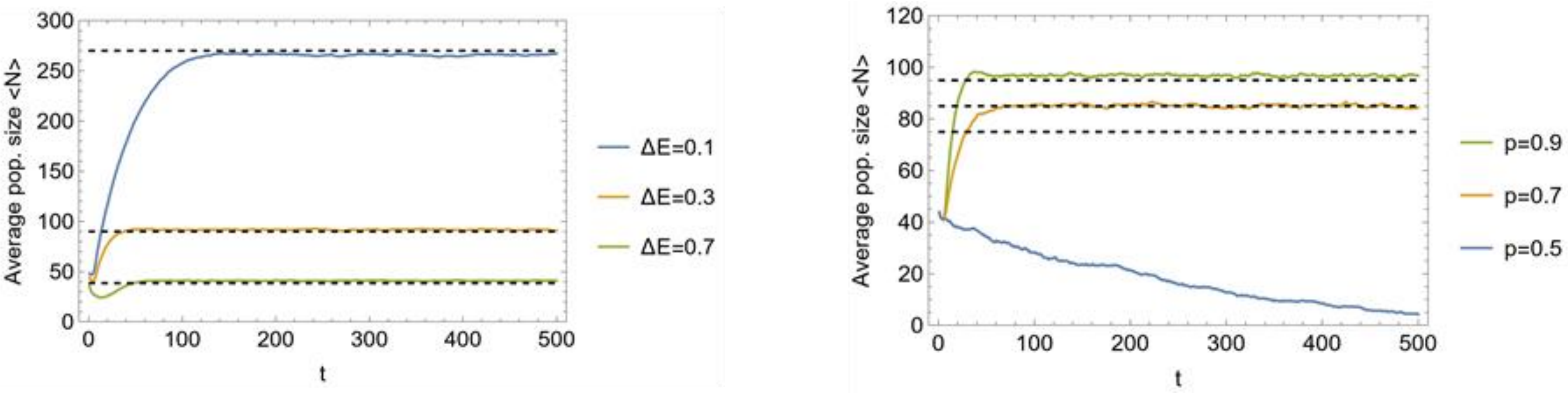
Agent-based model



yellow: protocells without genetic elements
blue: protocells containing genetic elements
black circles: resources
green: genetic elements – cooperators
red: genetic elements – parasites/defectors

Pre-life: evolution of “pure” reproducers, no replicators

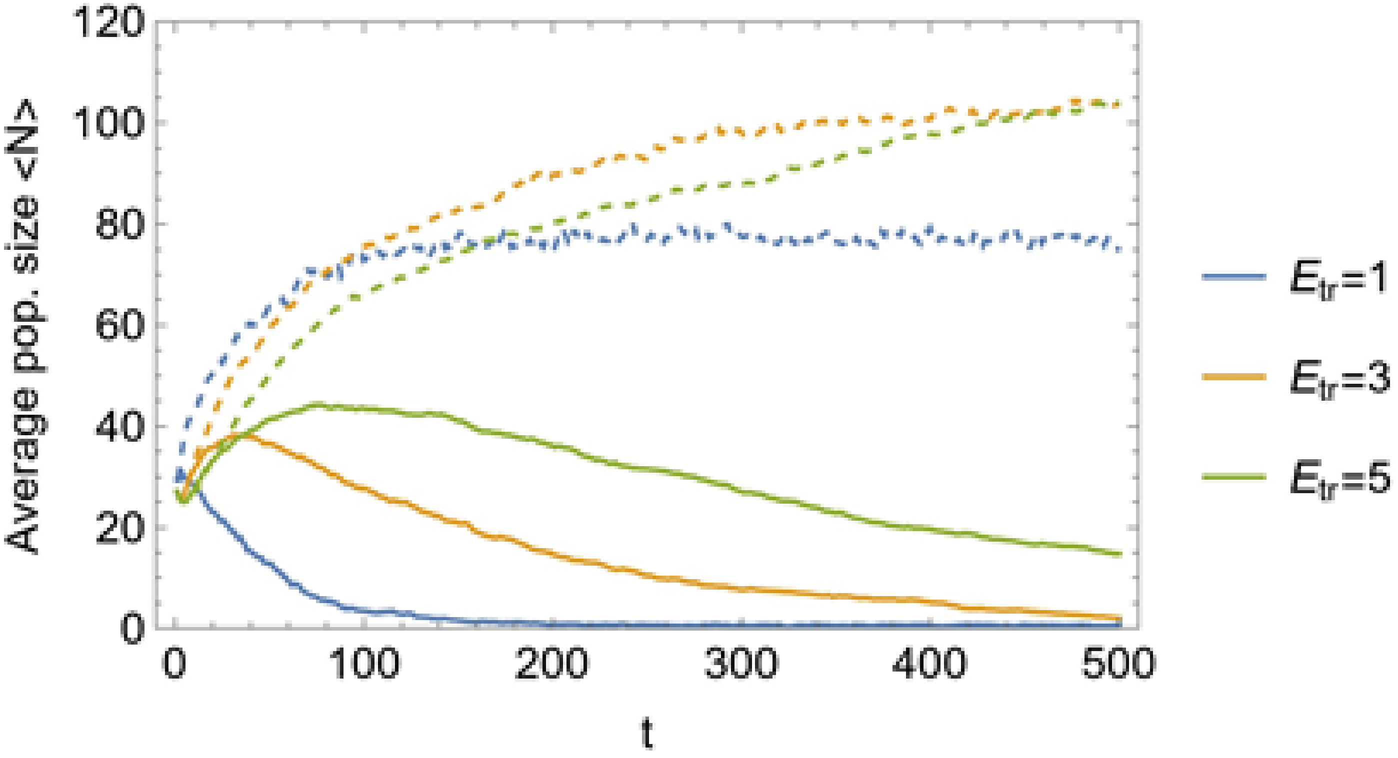
Protocells harbor networks of proto-metabolic reactions and divide randomly



Protocell population growth depends on housekeeping cost ΔE and successful division probability *p.*

Learning/Selection to decrease ΔE and increase *p*

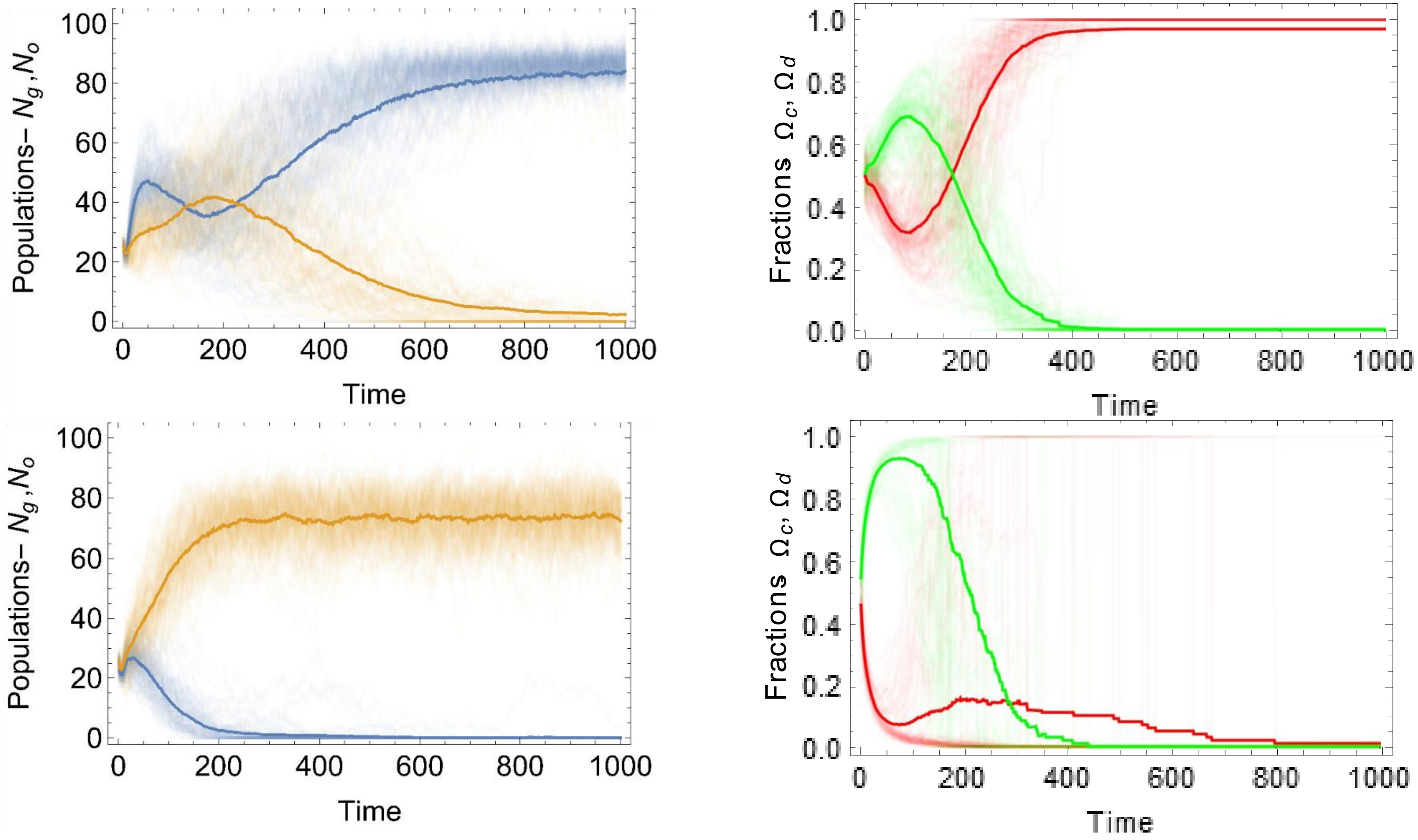
Symmetrically dividing protocells outcompete randomly dividing ones



Competition between protocells with and without replicators: Chance for cooperators-only protocells to emerge via random division

K=1

K=10



K, mean replication rate per round of resource consumption by protocells

🟦 replicators

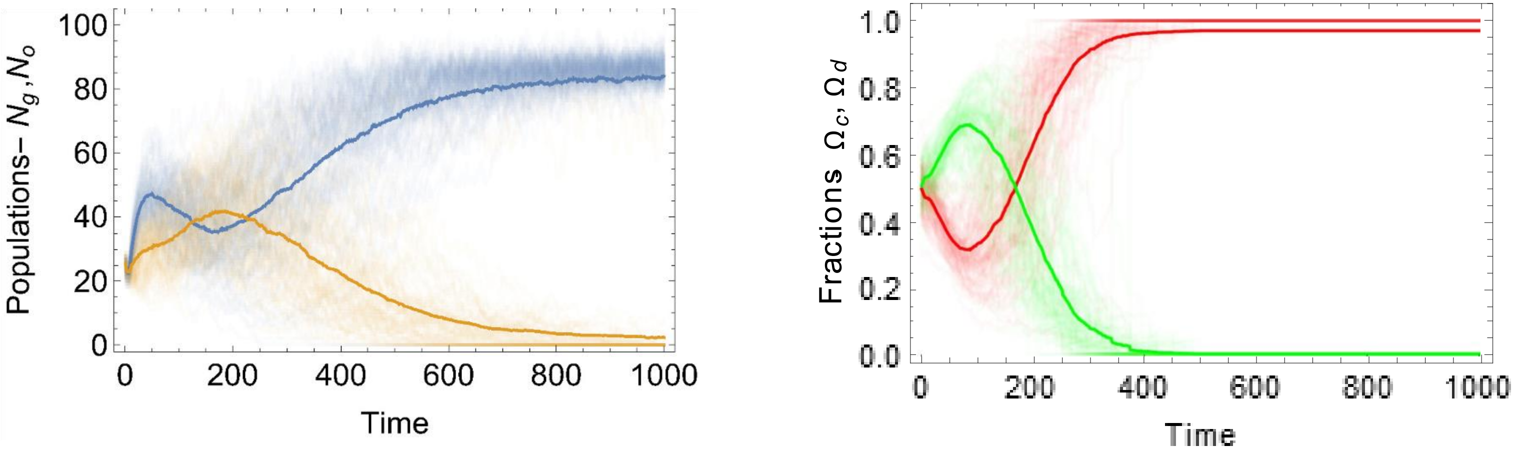
🟨 no replicators

🟥 cooperators

🟩 parasites/defectors

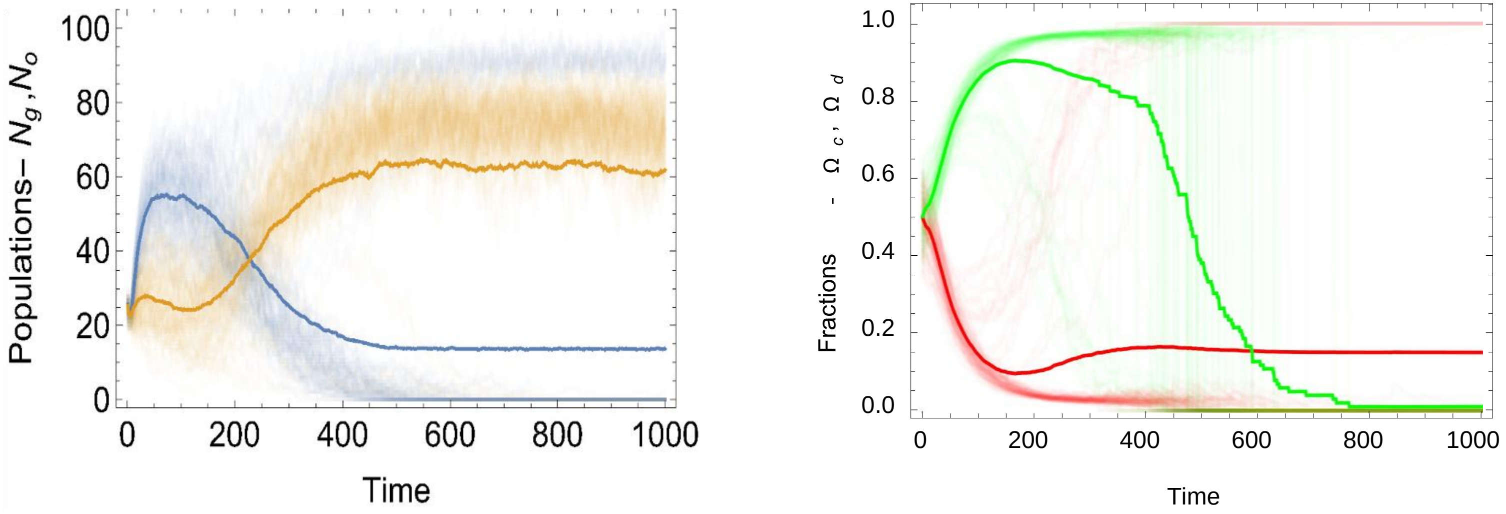
Random division outcompetes symmetrical division for protocells containing both cooperator and parasite replicators

Random division



K=1

Symmetrical division



replicators

no replicators

cooperators

Parasites/defectors

Implications for origin of life

- **Frustration** between replicator and reproducer-level selection – resulting in selection for coordination of replication and division rates
- Advantage of **stochastic division** in the presence of cooperators and parasites – provides for emergence of cooperators-only protocells
- Redundancy of replication machinery in cooperators-only protocells – **origin of large genomes**
- **Symmetrical division** advantageous once cooperators-only protocells emerge
- Vulnerability to parasite invasion – **early emergence of defense in protocells**

Take home…

- The principal features of life can be derived from the formalism of multilevel learning
- Major transitions in evolution: genuine, physical phase transitions
- The key transition – origin of life (cells) – “mutual learning” of reproducers and replicators – mutualistic symbiosis, coordination of genome reproduction and protocell division as a condition for fixation of genomes in evolution

Contributors

- Vitaly Vanchurin, NCBI, Duluth Institute for Advanced Study, Duluth, MN
- Mikhail Katsnelson, Radboud U, Nijmegen, Netherlands
- Yuri Wolf, NCBI
- Armen Allakhverdyan, Alikhanyan National Laboratory, Yerevan
- Andranik Khachatryan, Alikhanyan National Laboratory, Yerevan
- Purificacion Lopez-Garcia, U Paris-Saclay, Gif-sur-Yvette, France
- Sanasar Babajanian, NCBI