Pervasive biological conflicts and evolution of complexity

Eugene V. Koonin

National Center for Biotechnology Information, NLM, NIH

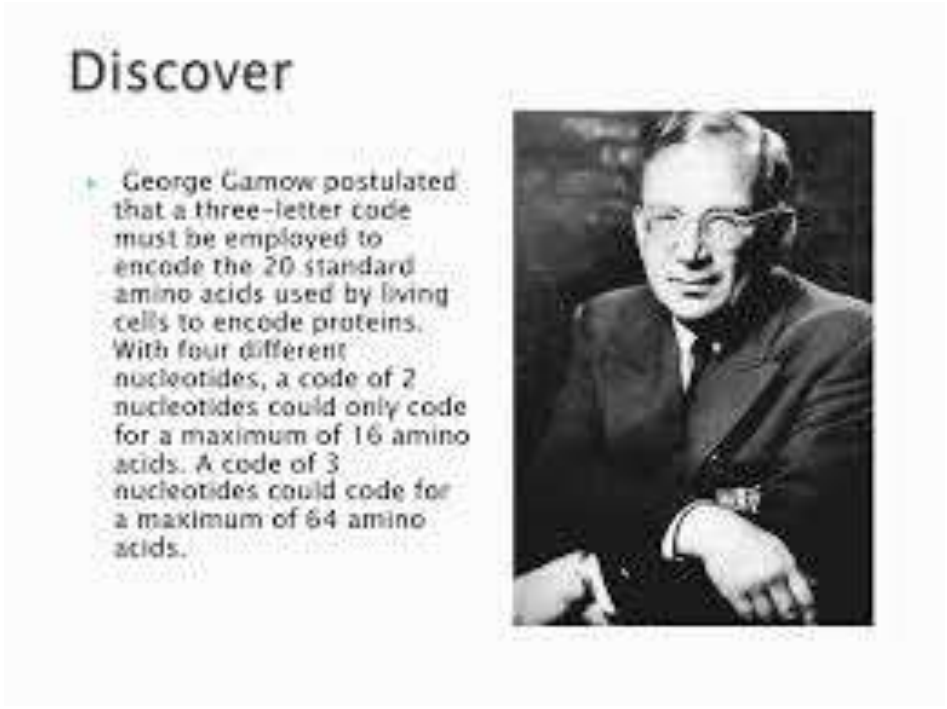
RASA, IX International Conference, Washington, DC, Nov 3, 2018

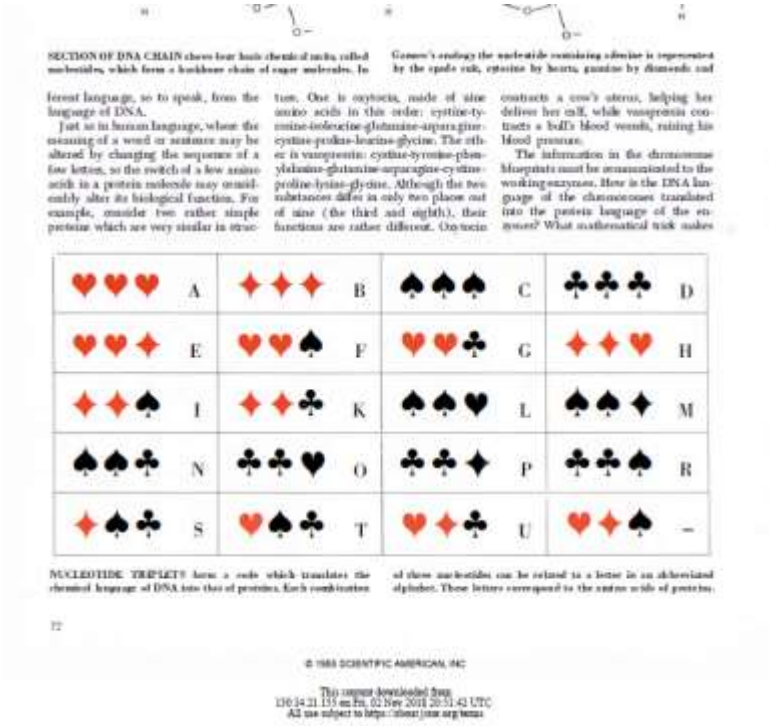
NCBI: Evolutionary Genomics Group

http://www.ncbi.nlm.nih.gov/research/groups/koonin/

Simple seminal ideas on the nature of the genetic code

The 1st, conceptual genetic code table





SECTION OF DNA CHAIN shows four basic chemical units, called nucleotides, which form a backbone chain of sugar molecules. In Gamow's analogy the nucleotide containing adenine is represented by the spade suit, cytosine by hearts, guanine by diamonds and

ferent language, so to speak, from the language of DNA.

Just as in human language, where the meaning of a word or sentence may be altered by changing the sequence of a few letters, so the switch of a few amino acids in a protein molecule may considerably alter its biological function. For example, consider two rather simple proteins which are very similar in structure. One is oxytocin, made of nine amino acids in this order: cystine-tyrosine-isoleucine-glutamine-asparagine-cystine-proline-leucine-glycine. The other is vasopressin: cystine-tyrosine-phenylalanine-glutamine-asparagine-cystine-proline-lysine-glycine. Although the two substances differ in only two places out of nine (the third and eighth), their functions are rather different. Oxytocin contracts a cow's uterus, helping her deliver her calf, while vasopressin contracts a bull's blood vessels, raising his blood pressure.

The information in the chromosome blueprints must be communicated to the working enzymes. How is the DNA language of the chromosomes translated into the protein language of the enzymes? What mathematical trick makes

♥♥♥ A	♦♦♦ B	♠♠♠ C	♣♣♣ D
♥♥♦ E	♥♥♠ F	♥♥♣ G	♦♦♥ H
♦♦♠ I	♦♦♣ K	♠♠♥ L	♠♠♦ M
♠♠♣ N	♣♣♥ O	♣♣♦ P	♣♣♠ R
♦♠♣ S	♥♠♣ T	♥♦♣ U	♥♦♠ –

NUCLEOTIDE TRIPLETS form a code which translates the chemical language of DNA into that of proteins. Each combination of three nucleotides can be related to a letter in an abbreviated alphabet. These letters correspond to the amino acids of proteins.

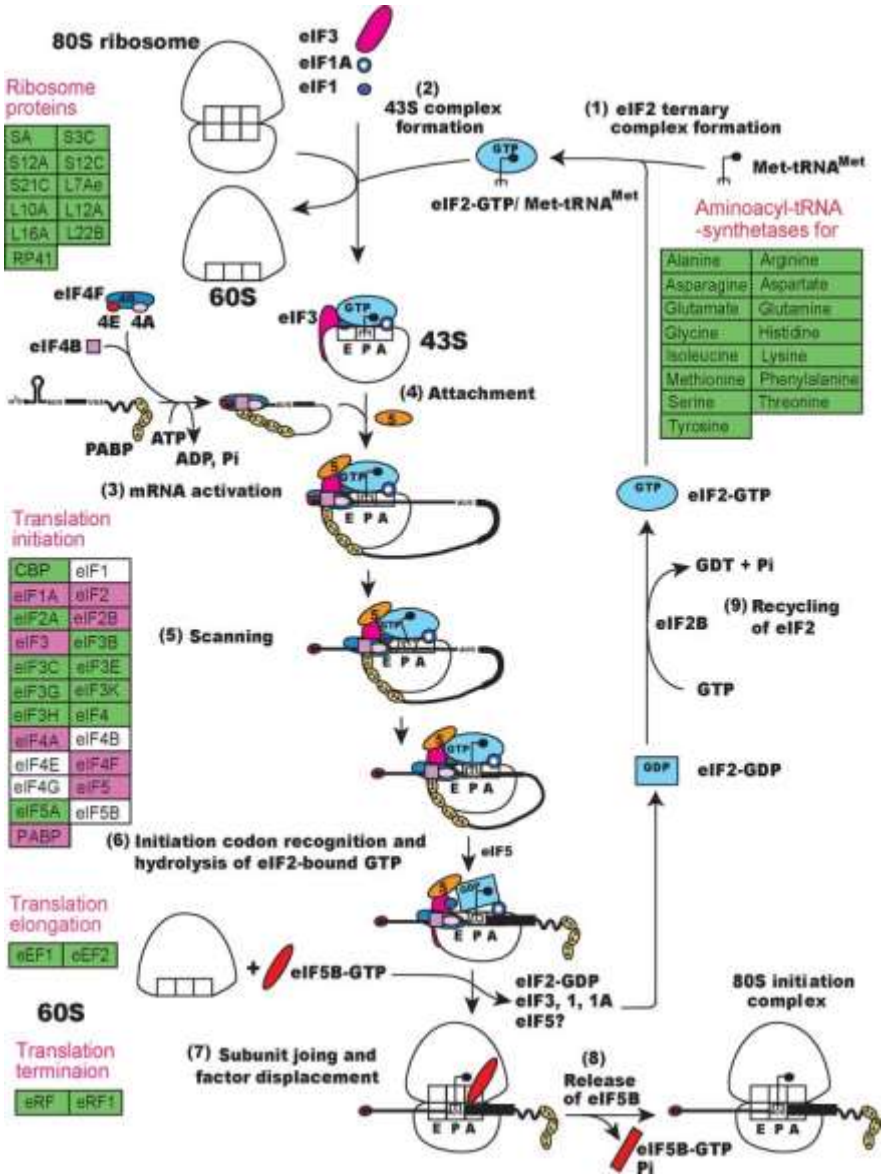
The 1st suggestion of a non-overlapping, triplet genetic code:

"…it appears more probable that the number of determining nucleotides exceed by a factor of 3 the number of amino acid residues in a synthesized protein, so that neighboring residues do not share determining nucleotides".

Information transfer in the Living Cell
Gamow, G,
Scientific American,
Oct 1955

Gamow G, Ycas M. STATISTICAL CORRELATION OF PROTEIN AND RIBONUCLEIC ACID COMPOSITION. Proc Natl Acad Sci U S A. 1955 Dec 15

Modern translation system - ~40 RNAs + ~100 proteins – epitome of “irreducible” molecular complexity



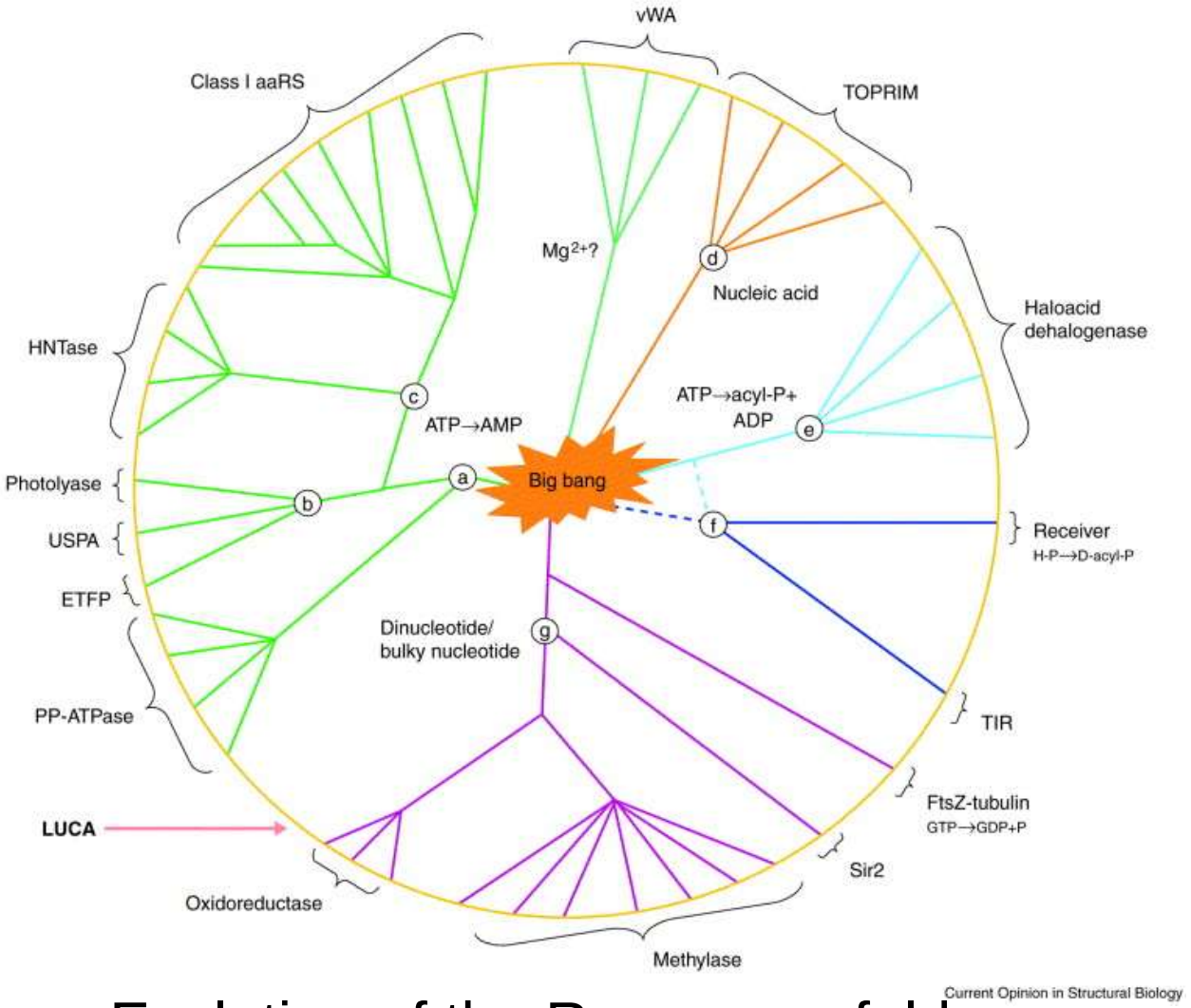
“Protein evolution paradox”:

Protein evolution requires high-fidelity translation which Itself depends on numerous proteins

Wolf, Koonin, Biol Direct 2007

“Protein evolution paradox”: extensive protein evolution before modern protein-based translation evolved

Koonin, Life 2017



Evolution of the Rossmann fold

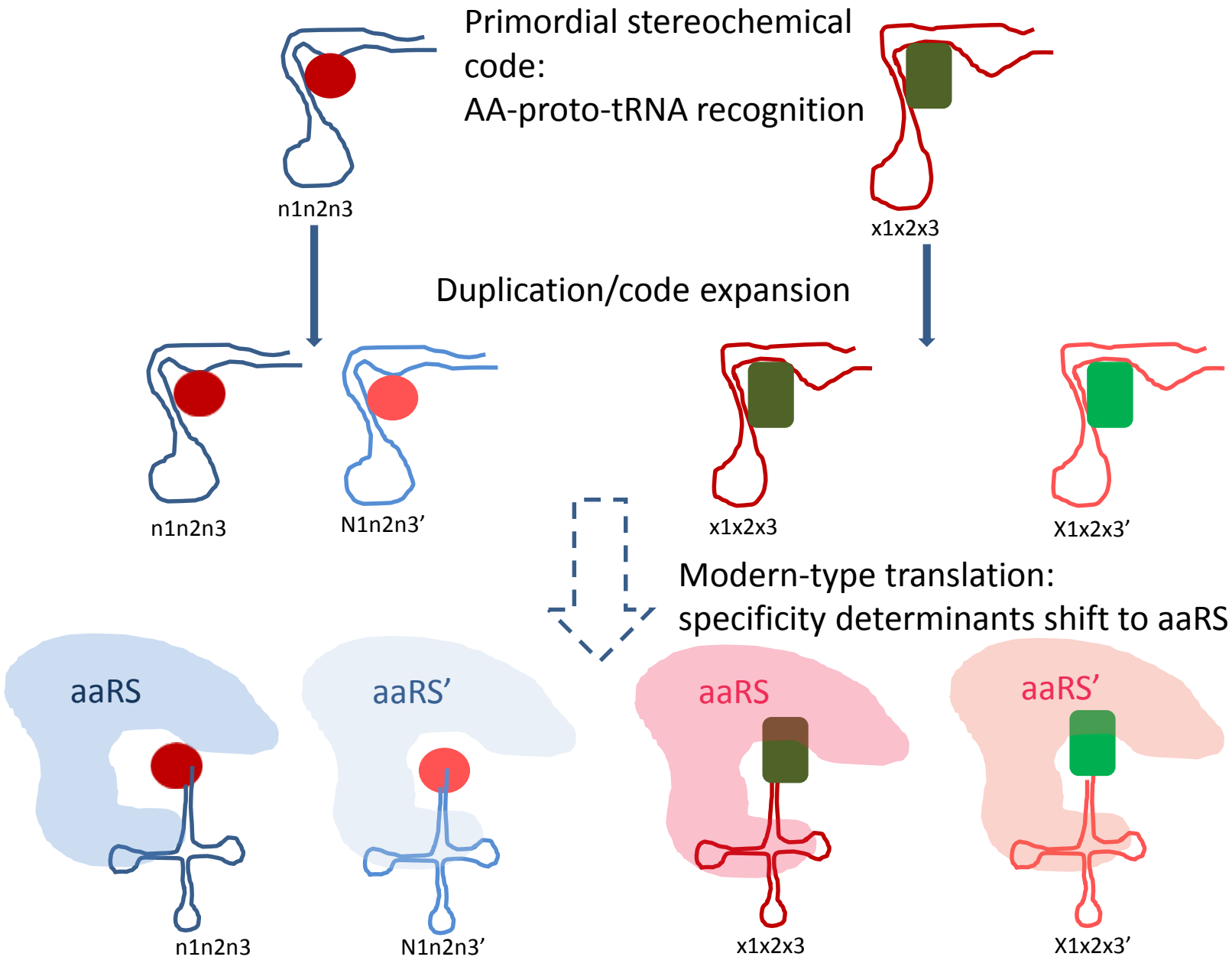
Aravind, Mazumder, Vasudevan, Koonin, Curr Opinion Struct Biol 2002

Resolution to the protein evolution paradox: RNA-based translation with sufficient fidelity

- RNA-based translation implies streochemical basis of the code – RNA-amino acid recognition
- Codon (anticodon)-amino acid interaction: none for some amino acids, inconsistent for others

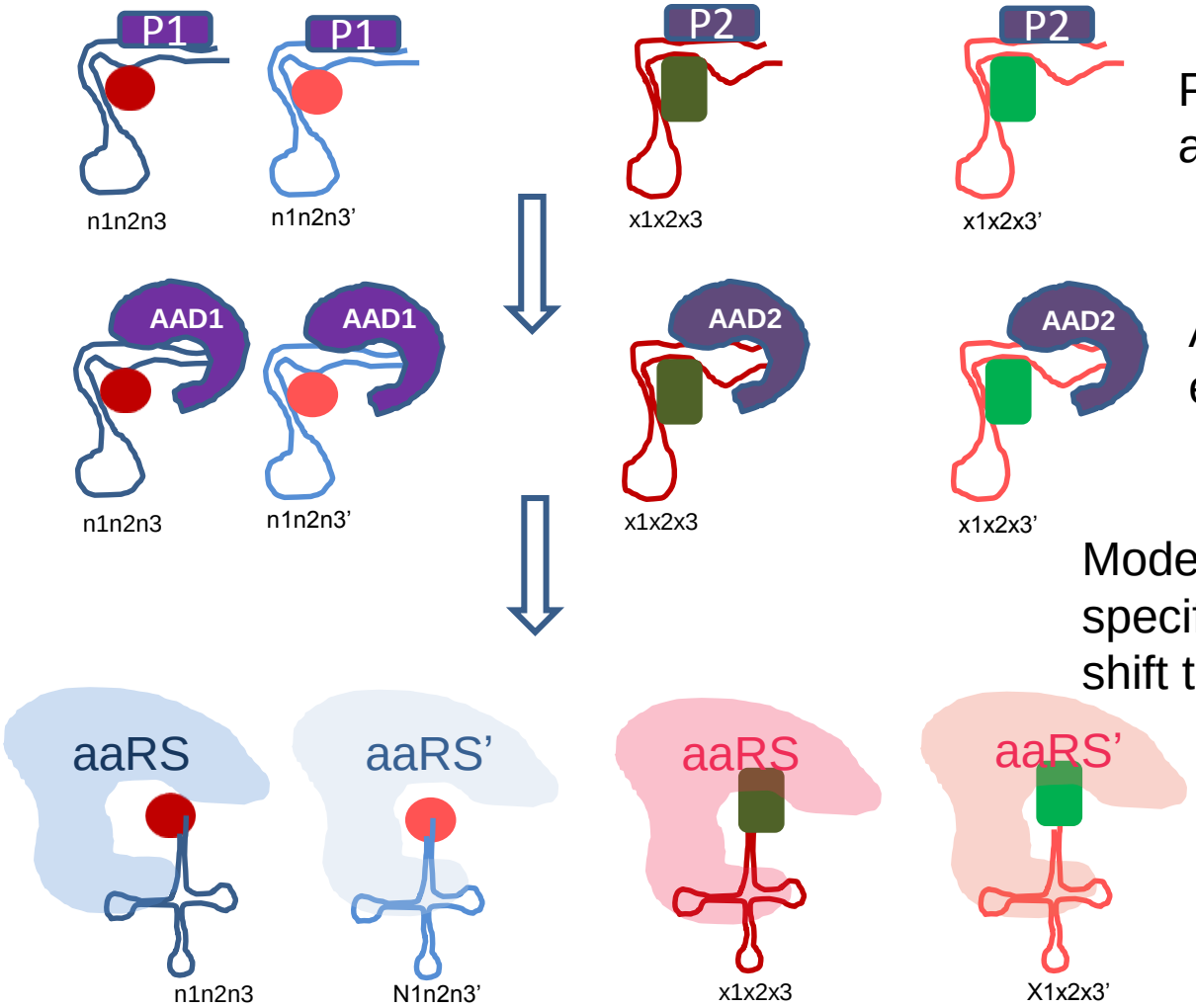
Koonin, Novozhilov, Annu Rev Genet 2017

A different primordial stereochemical code



Koonin, Novozhilov, Annu Rev Genet 2017

Switch to protein-based translation



Koonin, Life 2017

Biological and physical perspectives on the evolution of complexity; compatibility and complementarity

Towards physical principles of biological evolution

Mikhail I Katsnelson, Yuri I Wolf and Eugene V Koonin

Physica Scripta 2018

Biological systems reach organizational complexity that far exceeds the complexity of any known inanimate objects. Biological entities undoubtedly obey the laws of quantum physics and statistical mechanics. However, is modern physics sufficient to adequately describe, model and explain the evolution of biological complexity? Detailed parallels have been drawn between statistical thermodynamics and the population-genetic theory of biological evolution. Based on these parallels, **we outline new perspectives on biological innovation and major transitions in evolution, and introduce a biological equivalent of thermodynamic potential that reflects the innovation propensity of an evolving population**. Deep analogies have been suggested to also exist between the properties of biological entities and processes, and those of frustrated states in physics, such as glasses. Such systems are characterized by frustration whereby local state with minimal free energy conflict with the global minimum, resulting in "emergent phenomena". **We extend such analogies by examining frustration-type phenomena, such as conflicts between different levels of selection, in biological evolution. These frustration effects appear to drive the evolution of biological complexity.** We further address evolution in multidimensional fitness landscapes from the point of view of percolation theory and suggest that **percolation at level above the critical threshold dictates the tree-like evolution of complex organisms**. Taken together, these multiple connections between fundamental processes in physics and biology imply that construction of a meaningful physical theory of biological evolution might not be a futile effort.

Equivalencies between thermodynamics and evolution/population genetics

	Biological equivalent	
Thermodynamic variable	Sella, Hirsh. The application of statistical physics to evolutionary biology. PNAS 2005	Katsnelson, Wolf, Koonin. Towards physical principles of biological evolution. Physica Scripta 2018
Inverse temperature, $\beta=1/T$	Effective population size N_e	
Entropy per particle	Derived from the free fitness expression	Evolutionary information density: $D(N)= 1-H/N$
Free energy Hamiltonian	Minus log of fitness	-
Thermodynamic potential	Derived from the Hamiltonian by Gibbs formula	Evolutionary innovation potential: $dI= dt(dH/dt)/N_e$

Equivalencies between thermodynamics and evolution/population genetics

Innovation potential in evolution:

dI= dt(dH/dt)/Ne ~thermodynamic potential

where evolutionary entropy (analog of Shannon entropy for alignment of homologoous sequences) is:

$$H = \sum_{i=1}^{L} H_i = -\sum_{i=1}^{L} \sum_j f_{ij} \log f_{ij}$$

Evolutionary Information (~evolutionary conservation)

$$I(N) = N - \sum_{i=1}^{k} H(L_i),$$

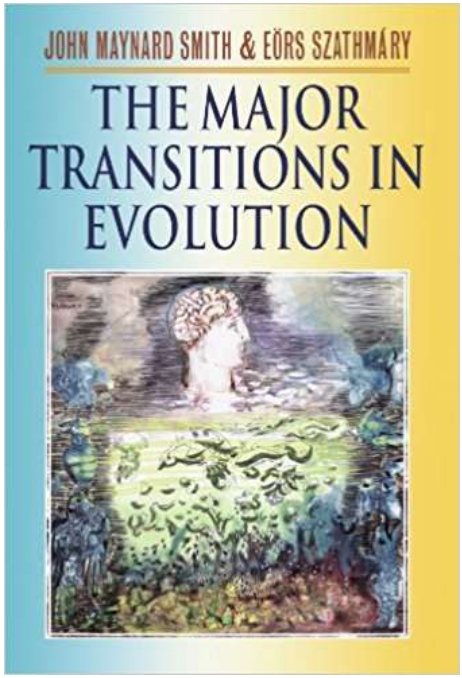
Evolutionary information density:

$D(N)= 1-H/N$

Major transitions in evolution

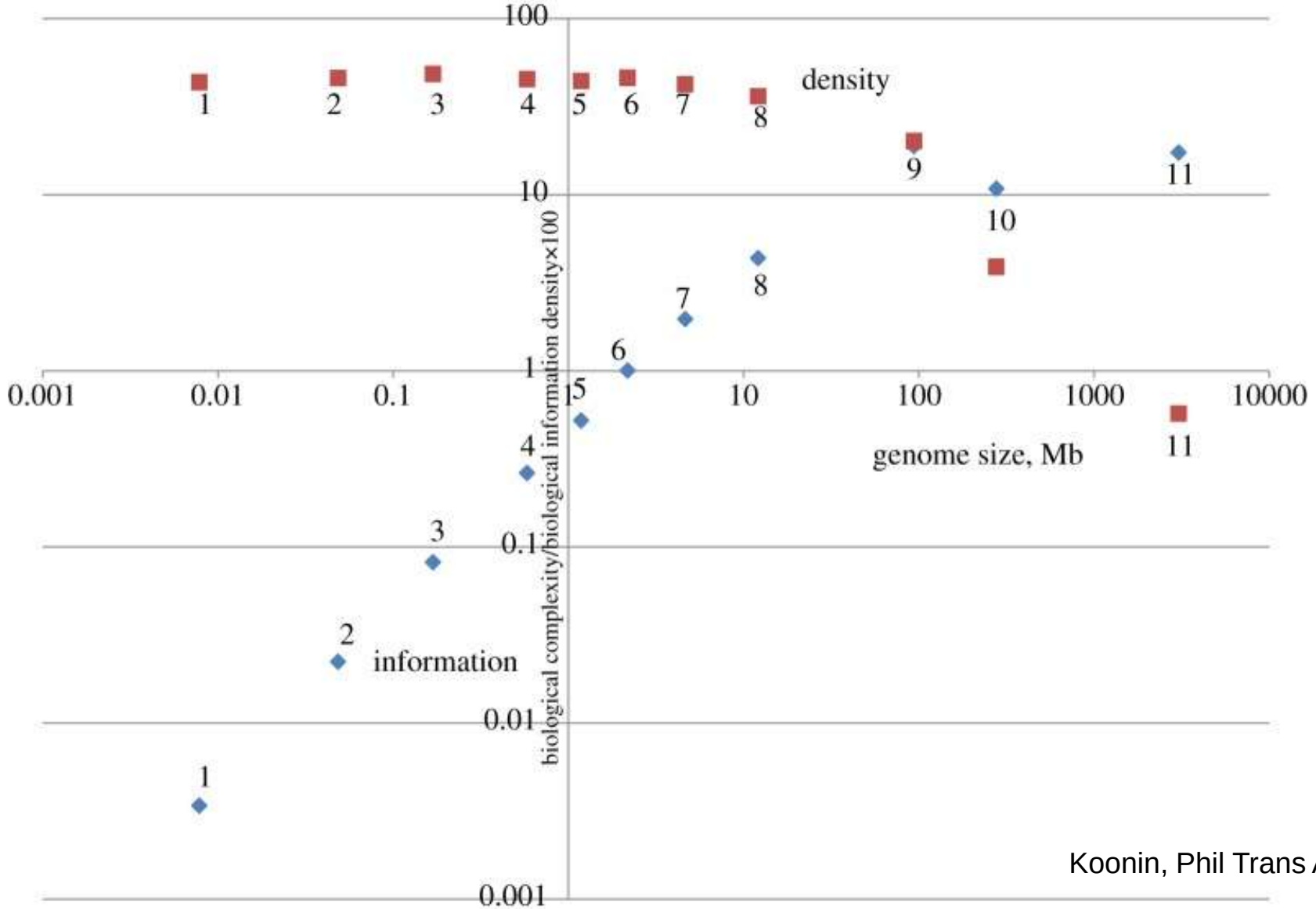
Evolutionary Transition in Individuality:
New levels of selection

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



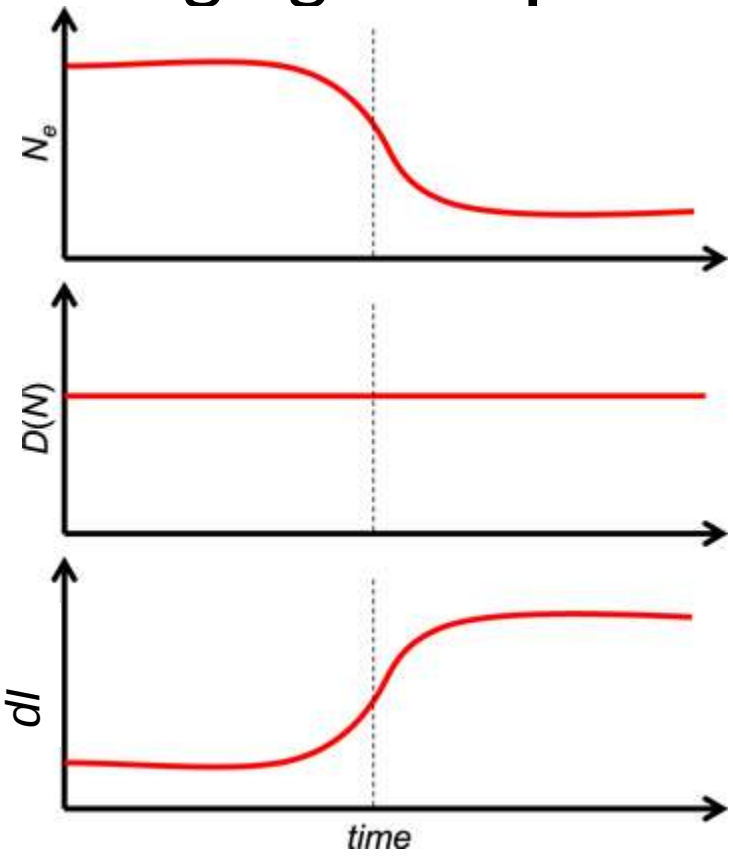
What happens to the key population/thermodynamic quantities during major transitions?

I(N) increases with genome size
D(N) remains ~constant over wide range of genome size, drops for large genomes



Koonin, Phil Trans A 2016

Major evolutionary transitions as adiabatic first-order transitions with constant entropy and changing temperature



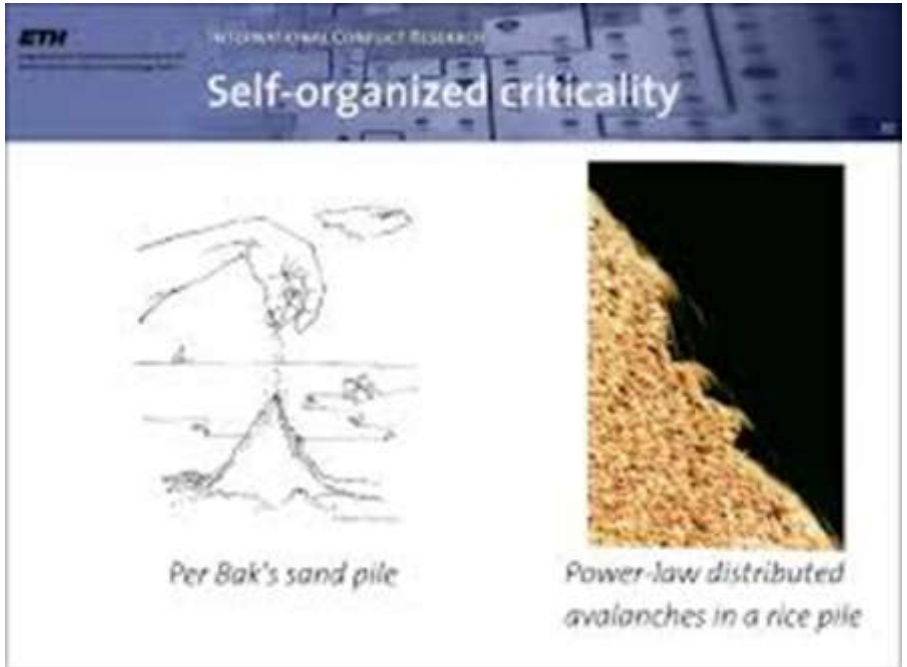
The three panels schematically show the changes in effective population size (top), evolutionary information density (middle panel) and evolutionary innovation potential (bottom) at a MTE that is shown by the vertical dotted line.

Katsnelson, Wolf, Koonin, Phsys Scripta 2018

Self-organized criticality behind the major transitions

Stefan Boettcher and Maya Paczuski
Exact Results for Spatiotemporal Correlations in a Self-Organized Critical Model of Punctuated Equilibrium, PRL 1996

- We introduce a self-organized critical model of punctuated equilibrium with many internal degrees of freedom (M) per site. We find exact solutions for $M \to \infty$ of cascade equations describing avalanche dynamics in the steady state. This proves the existence of simple power laws with critical exponents that verify general scaling relations for nonequilibrium phenomena. Punctuated equilibrium is described by a devil's staircase with a characteristic exponent $\tau_{first}=2-d/4$ where d is the spatial dimension.



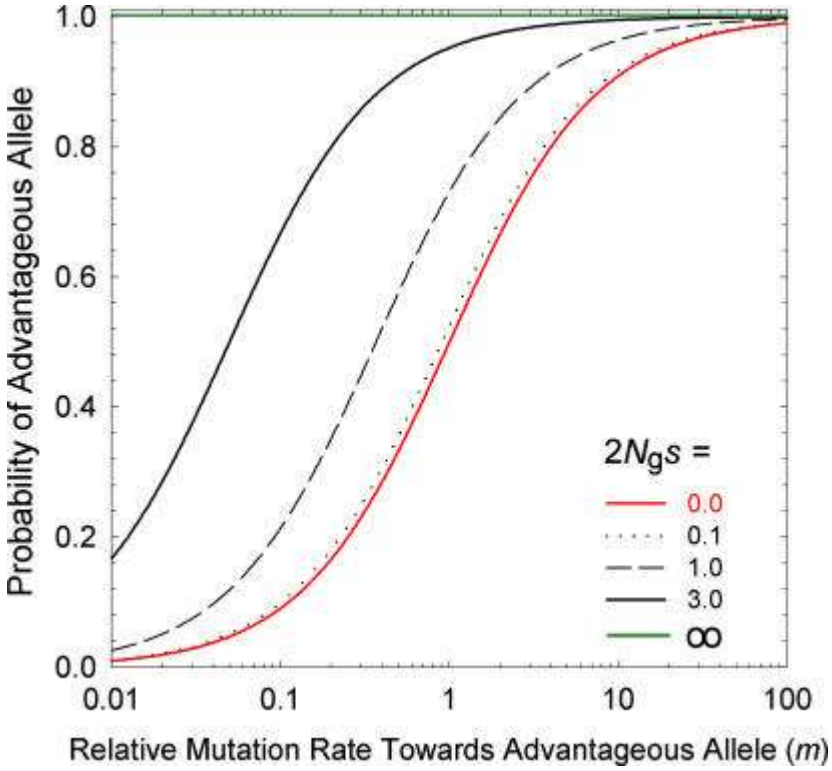
Non-adaptive origin of genomic complexity

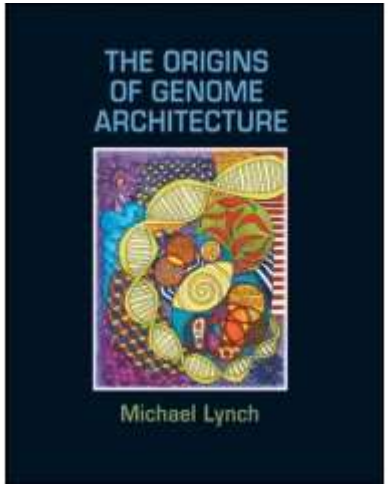
Lynch M, Conery JS.

The origins of genome complexity. Science. 2003 Nov 21;302(5649):1401-4.

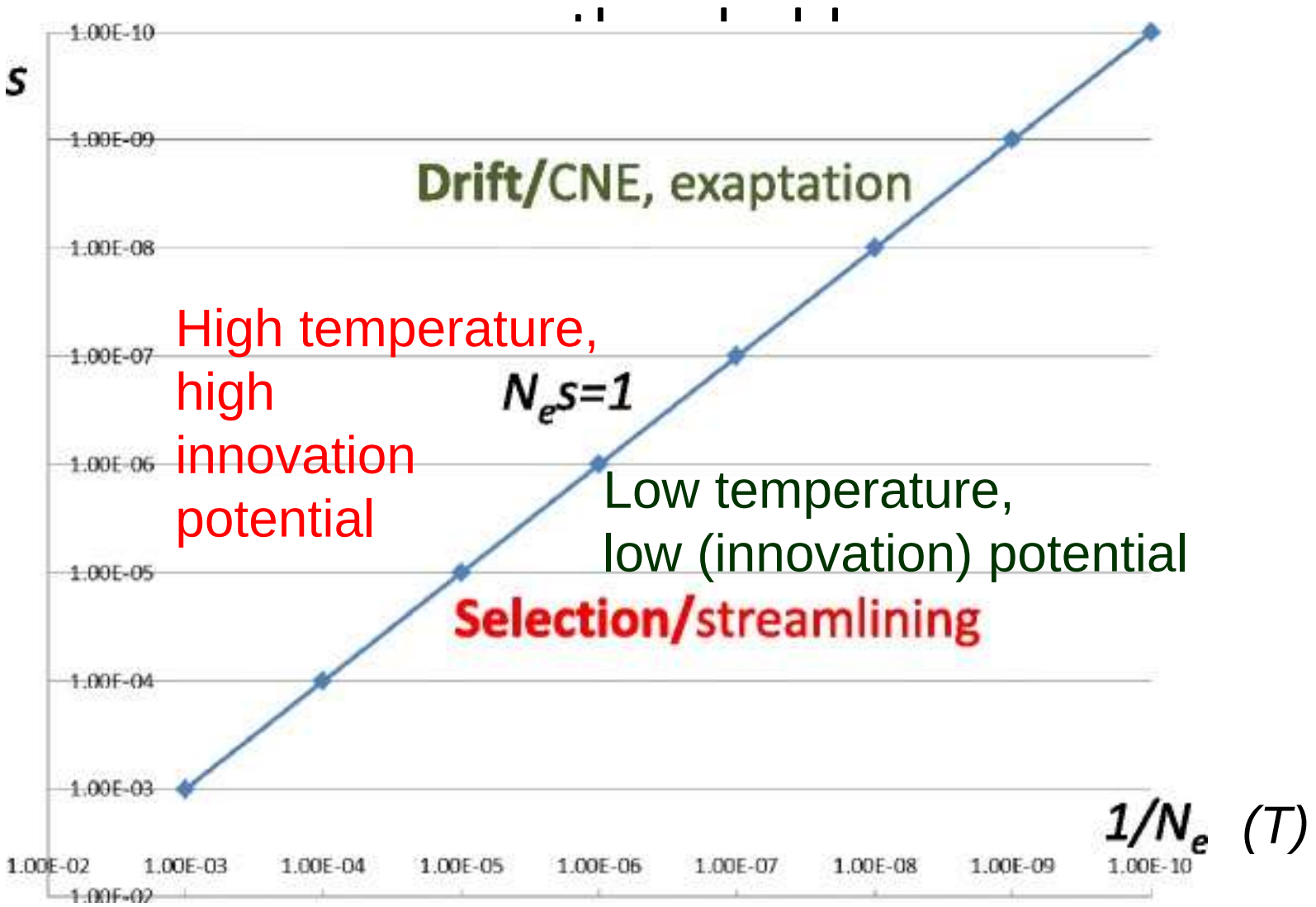
Complete genomic sequences from diverse phylogenetic lineages reveal notable increases in genome complexity from prokaryotes to multicellular eukaryotes. The changes include gradual increases in gene number, resulting from the retention of duplicate genes, and More abrupt increases in the abundance of spliceosomal introns and mobile genetic elements. We argue that many of these modifications emerged passively in response to the long-term population-size reductions that accompanied increases in organism size. According to this model, much of the restructuring of eukaryotic genomes was initiated by nonadaptive processes, and this in turn provided novel substrates for the secondary evolution of phenotypic complexity by natural selection. The enormous long-term effective population sizes of prokaryotes may impose a substantial barrier to the evolution of complex genomes and morphologies.

Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci U S A. 2007





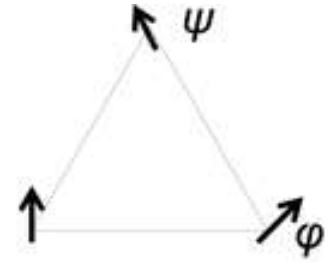
Not all mutations are visible to selection: drift

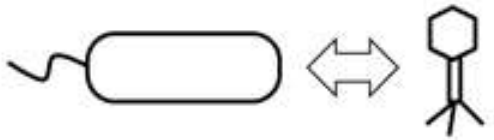


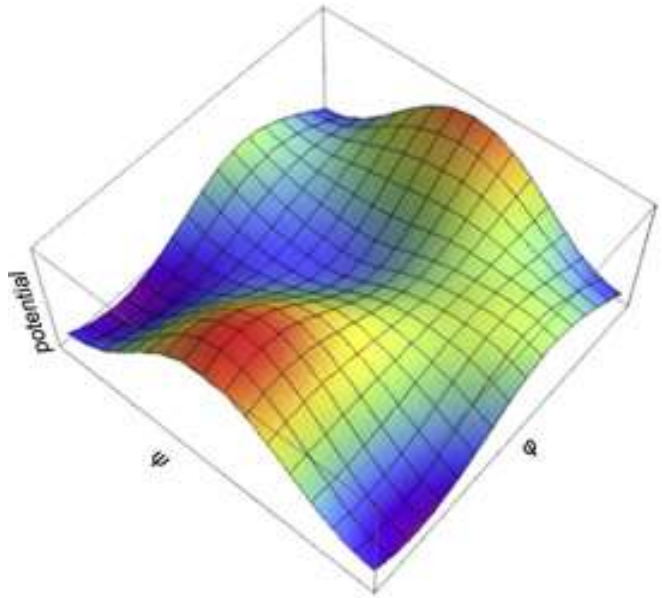
Splendor and misery of adaptation, or the importance of neutral null for understanding evolution.

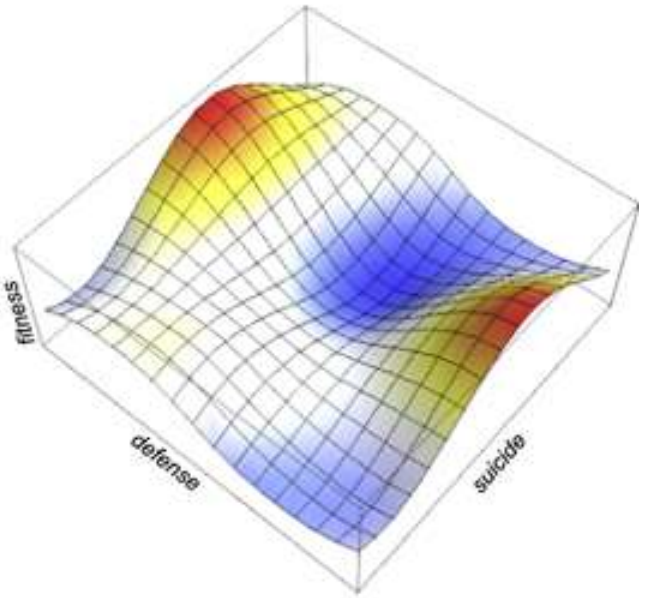
Koonin, BMC Biol 2016

Competing interactions and frustration in physical and biological systems









Potential ~ Fitness

Competing interactions drive evolution of complexity

Katsnelson, Wolf, Koonin, PNAS 2018

Frustration in a high-dimensional space leads to self-organized criticality

Self-organized criticality in glassy spin systems requires a diverging number of neighbors.

Andresen JC, Zhu Z, Andrist RS, Katzgraber HG, Dobrosavljević V, Zimanyi GT. PRL

We investigate the conditions required for general spin systems with frustration and disorder to display self-organized criticality, a property which so far has been established only for the fully connected infinite-range Sherrington-Kirkpatrick Ising spin-glass model [Phys. Rev. Lett. 83, 1034 (1999)]. Here, we study both avalanche and magnetization jump distributions triggered by an external magnetic field, as well as internal field distributions in the short-range Edwards-Anderson Ising spin glass for various space dimensions between 2 and 8, as well as the fixed-connectivity mean-field Viana-Bray model. Our numerical results, obtained on systems of unprecedented size, demonstrate that self-organized criticality is recovered only in the strict limit of a diverging number of neighbors and is not a generic property of spin-glass models in finite space dimensions.

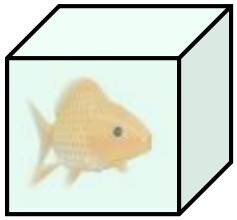
Competing interactions and frustrated states permeate biology

System	Frustration-producing elements (competing interactions)	Evolutionary consequences
Proteins	Hydrogen and Van der Waals bonds between side chains of monomers	Emergence of stable conformations and semi-regular patterns in protein structures
Gene regulation networks	Activators and repressors	Emergence of meta-stable expression patterns
Cells	Membranes and channels	Emergence of compartments and cellular machinery dependent on electrochemical gradients
Autonomous and semi-autonomous self-replicating genetic systems	Replicator and parasite genomes	Emergence of self vs nonself discrimination
Autonomous and semi-autonomous self-replicating genetic systems	Host cells and viruses	Emergence of infection mechanisms, defense and counter-defense systems, evolutionary arms race
Autonomous and semi-autonomous self-replicating genetic systems	Host cells and transposons	Emergence of intra-genomic DNA replication control; hotbeds of evolutionary innovation
Autonomous and semi-autonomous self-replicating genetic systems	Host cells and plasmids	Emergence of beneficial cargo genes, plasmid addiction systems, efficient gene exchange and transfer mechanisms

Competing interactions and frustrated states permeate biology

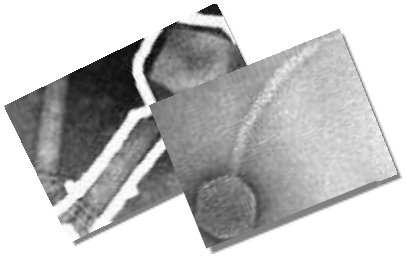
Communities of unicellular organisms	Individual cells	Emergence of information exchange and quorum sensing mechanisms; replication control apoptosis and multicellularity
Multicellular organisms	Soma and germline	Emergence of complex bodies and sexual reproduction
Populations	Individual members	Emergence of population-level cooperation; kin selection
Populations	Partners with unequal parental investment (males and females)	Emergence of sexual selection and sexual dimorphism
Biosphere	Species in different niches	Emergence of interspecies competition, host-parasite and predator-prey relationships, mutualism
Societies	…	…

Viruses are the dominant entities in the biosphere – physically and genetically – as shown by viral metagenomics – virome studies



1 cm^3 of seawater contains 10^6-10^9 virus particles

Suttle, C.A. (2005) *Nature* **437**:356



There are millions of diverse bacteriophage species in the water, soil, and gut

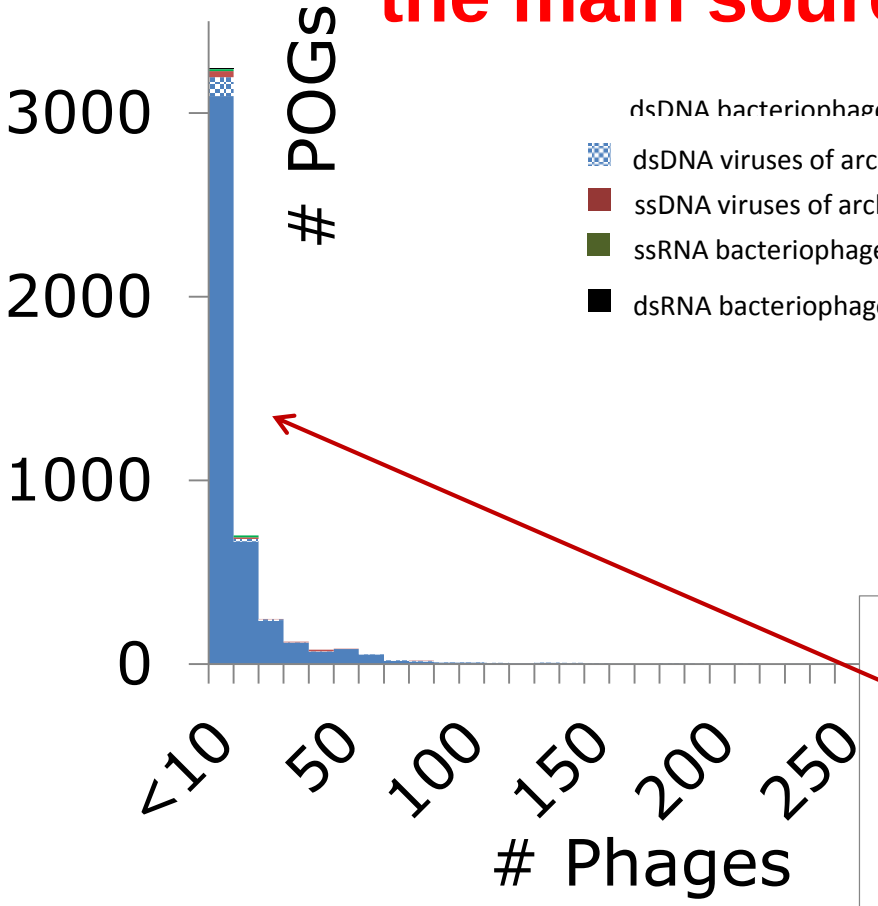
Edwards and Rohwer (2005) *Nat. Rev. Microbiol.* **3**:504

•*Viruses are the most abundant biological entities in the biosphere: 10-100 virus particles per cell in most habitats*

Human body: ~10^{14} human cells vs **~10^{15}** bacteria vs ~**10^{16}** viruses

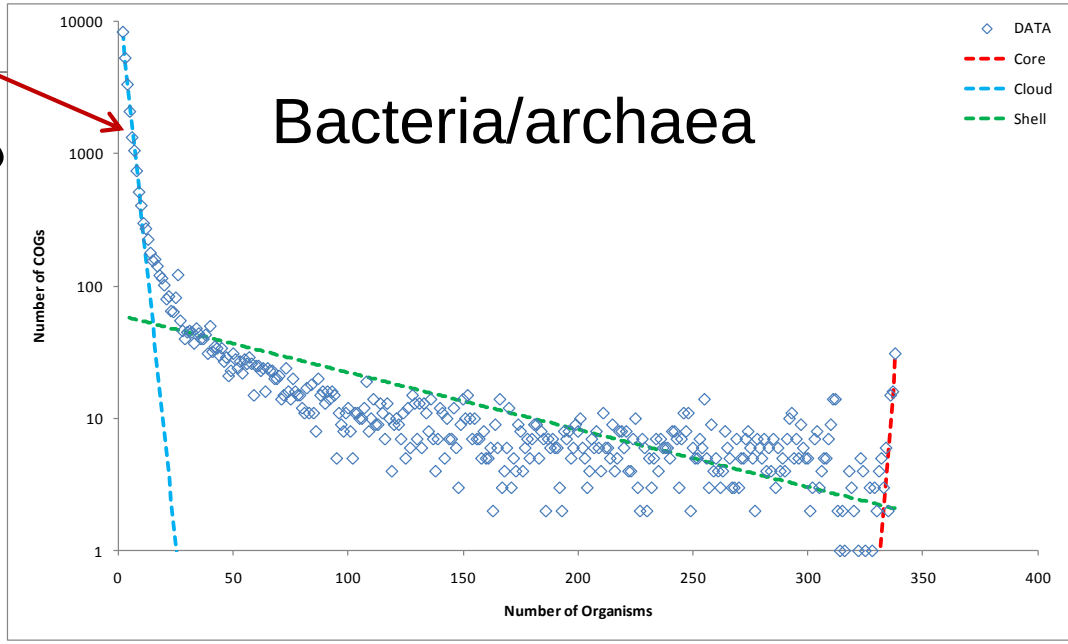
Earth-wide counts: ~5x10^{30} bacteria vs ~10^{32} viruses

Viruses: "Infinite" pool of genetic diversity – the main source of new genes on earth?



Kristensen et al. J. Bacteriol 2013

Gene frequency plots for viruses and cellular life forms are drastically different



Koonin, Wolf 2008

Inevitability of parasites

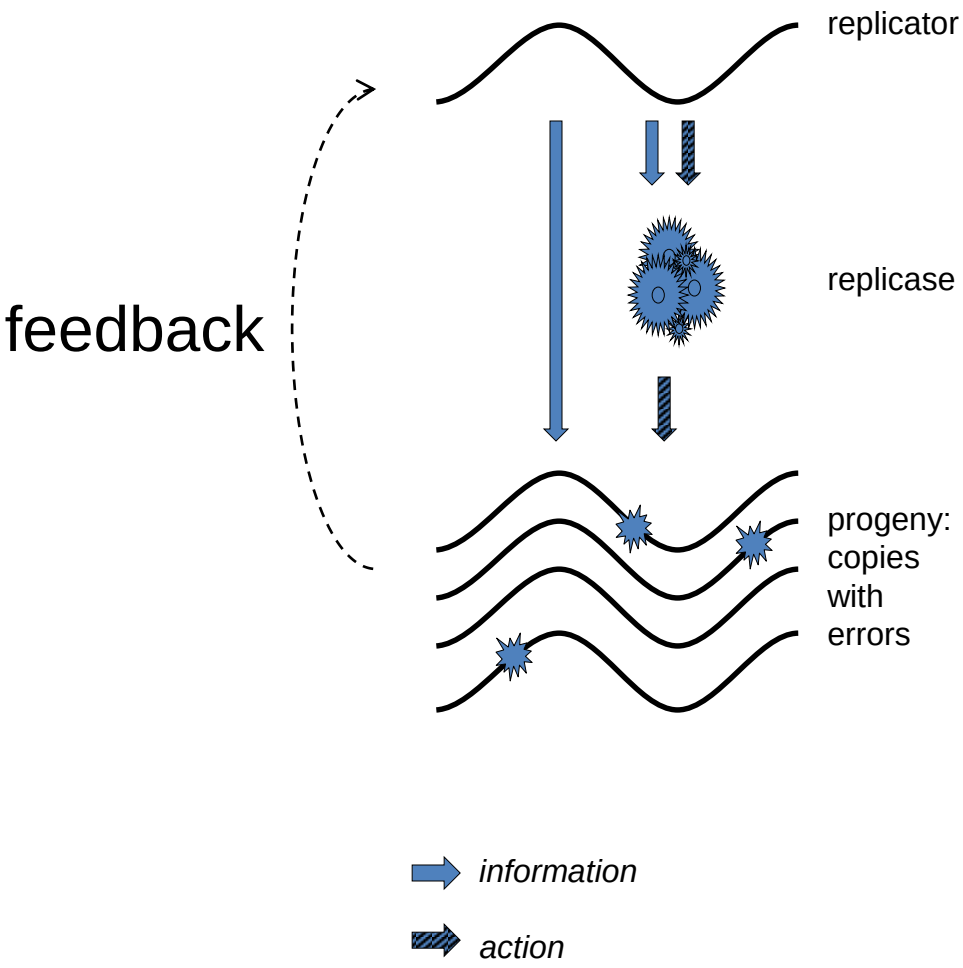
Parasites (cheaters) evolve in **ANY** replicator system (e.g. hypercycle): as soon as there is a resource to steal, such as a replicase, someone will find a way to live by stealing it

More formally: parasite-protected states of replicators are thermodynamically unstable

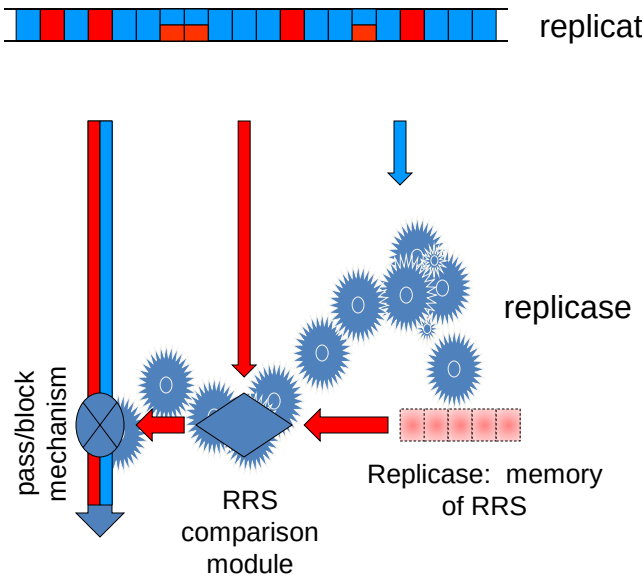
Maynard Smith J (1979) Hypercycles and the origin of life. Nature 280:445–446

Koonin, Wolf, Katsnelson. Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. Biol Direct 2017 http://www.biorxiv.org/content/early/2017/08/30/182410

The simplest replicator systems

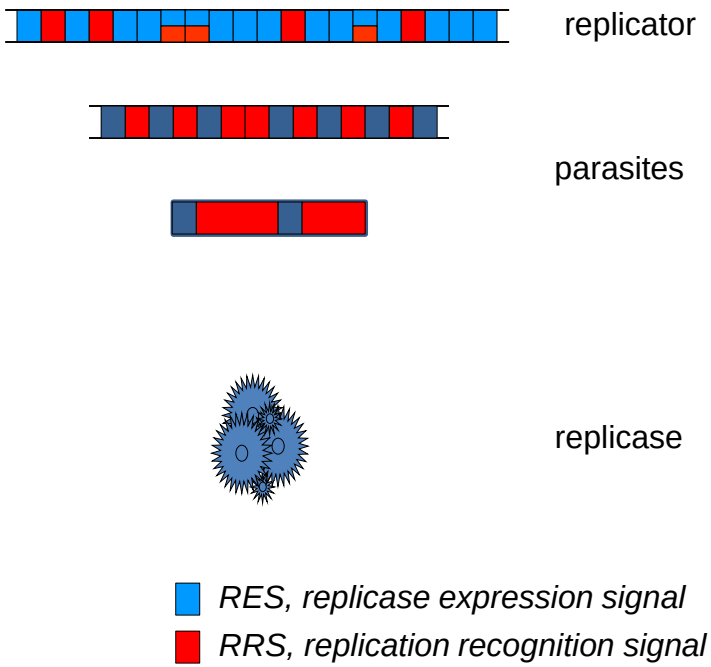


Koonin, Wolf, Katsnelson.
Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. Biol Direct 2017



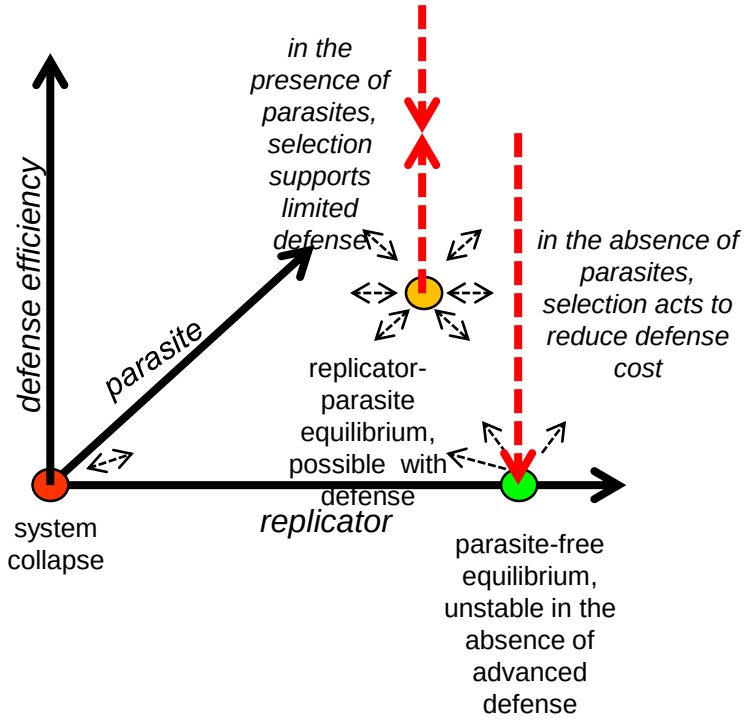
Koonin, Wolf, Katsnelson. Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. Biol Direct 2017

Parasites retain replicase recognitions signal but lose replication expression signal



Koonin, Wolf, Katsnelson. Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. Biol Direct 2017

Only host-parasite equilibria are stable



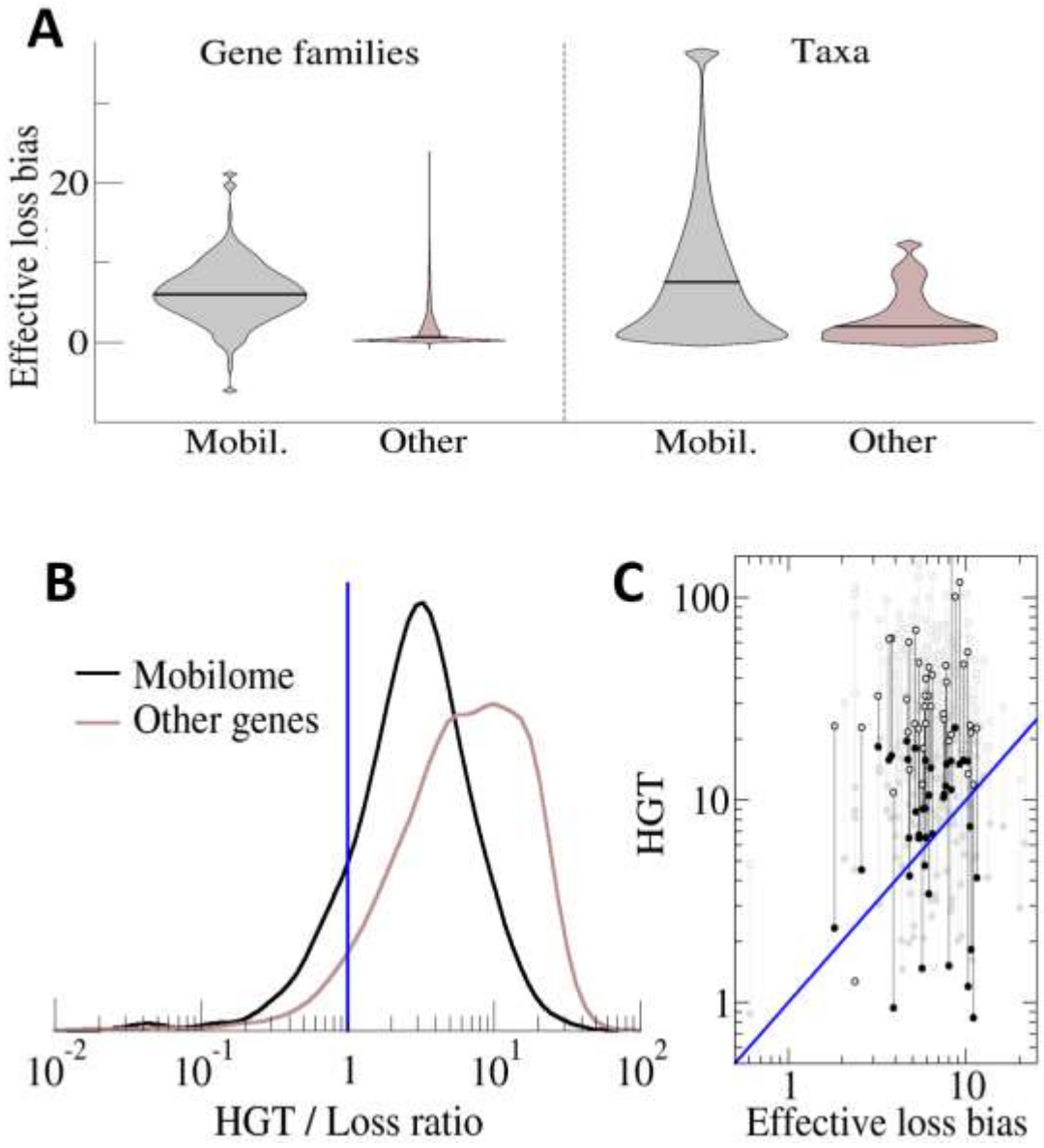
Koonin, Wolf, Katsnelson. Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states. Biol Direct 2017

Inevitability of genetic parasites/ selfish elements in the cellular world: is it possible to purge parasites?

- (Almost) all cellular life forms are hosts to diverse genetic parasites with various levels of autonomy including plasmids, transposons and viruses
- Theoretical modeling of the evolution of primordial replicators indicates that parasites ('cheaters') necessarily evolve in such systems and can be kept at bay primarily via compartmentalization
- Given the (near) ubiquity, abundance and diversity of genetic parasites, the question becomes pertinent: **are such parasites intrinsic to life?**
- At least in prokaryotes, the persistence of parasites is linked to the rate of horizontal gene transfer (HGT)
- Asexual clonal populations are doomed to extinction by accumulation of deleterious mutations: **Muller's ratchet**
- HGT > threshold - **escape from Muller's ratchet**
- HGT > threshold - **parasite persistence**
- $h^*_{MR} > h^*_{PP}$ - **no chance to purge parasitic selfish elements**

Takeuchi, Kaneko, Koonin G3 2014
Iranzo, Puigbo, Lobkovsky, Wolf, Koonin, GBE 2016

Testing the parasite persistence condition on 35 clusters of closely related bacterial and archaeal genomes



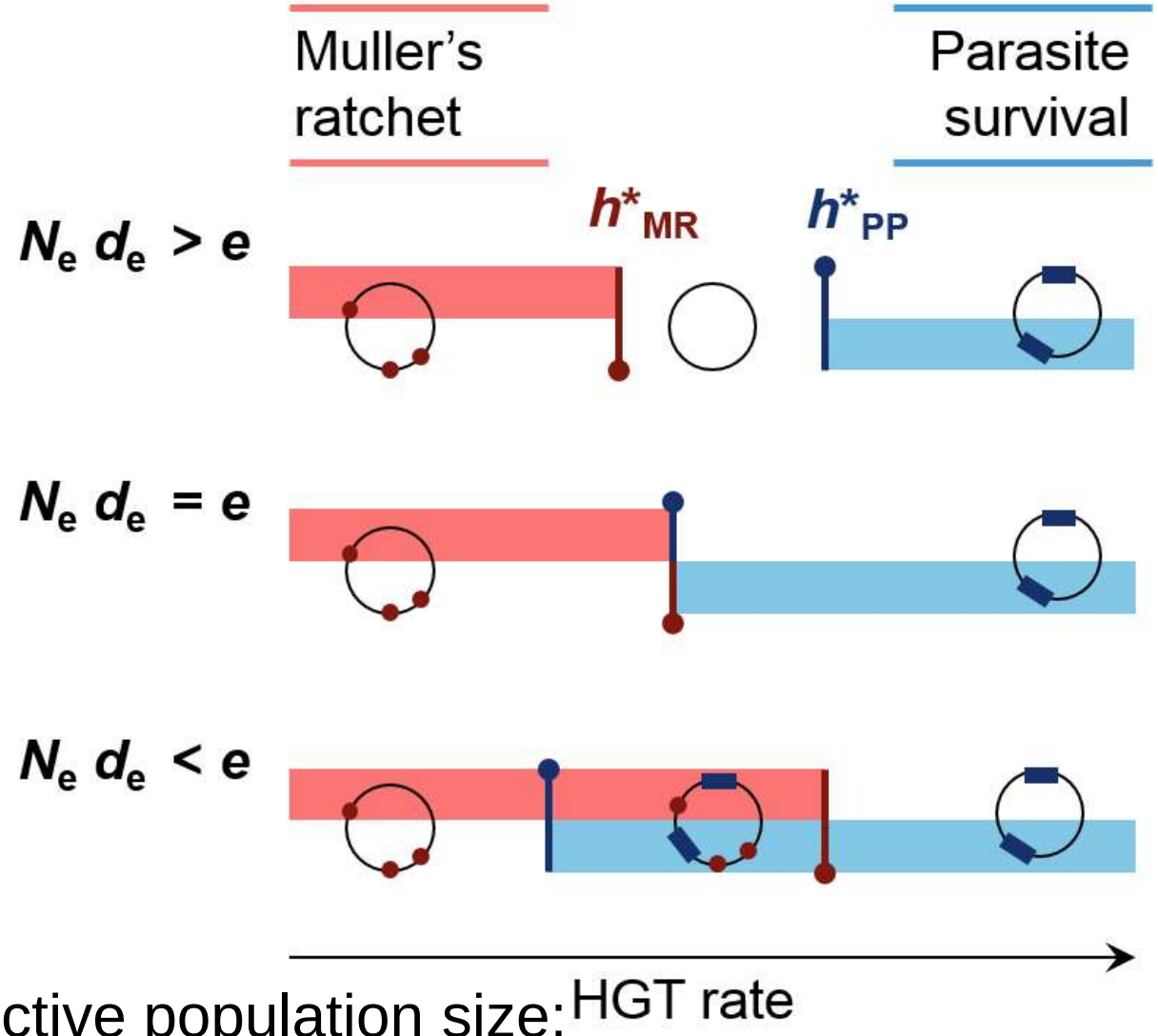
Persistence condition:

$h > d_e$

Holds for most microbes and most gene families

Puigbo et al BMC Biol 2014; Iranzo et al, GBE 2016

Between a rock and a hard place: Muller's ratchet vs parasite persistence – can microbes eliminate parasites by lowering HGT rate?



Ne, effective population size;
de, effective deletion bias

Iranzo et al GBE 2016

For most bacterial and archaeal genes, even those devoid of active proliferation capacity:

$$h^*_{MR} > h^*_{PP}$$

Thus, persistence of genetic parasites/selfish elements is the price to pay for maintenance of host genome integrity and hence inevitable; moreover, most genes tend to be "selfish"

Iranzo et al. GBE 2016

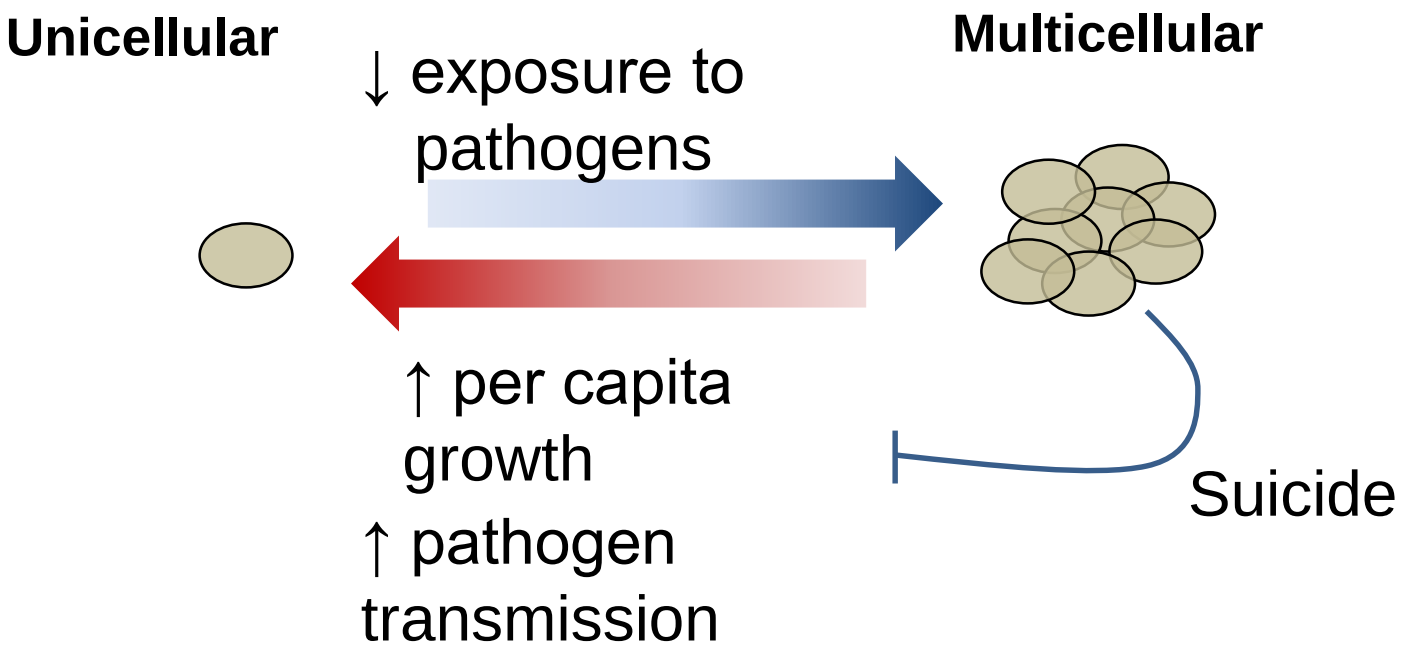
Parasites promote evolution of complexity and major evolutionary transitions
-Genetic parasites and evolution of multicellularity

Parasites drive evolution of complexity, in particular, compartmentalization

The class of replicators customarily called “parasites” is known to play important roles for the evolutionary dynamics of RNA-like replicator systems. Parasites are molecules that do not catalyze replication of other molecules but can be replicated by the catalysts, possibly at a faster rate than the catalysts themselves. **Under well-mixed conditions, the parasite can bring a replicator system to extinction by (over)exploiting catalysts**. Because of this inherent instability of RNA-like replicator systems against the parasite, it is necessary to consider spatial structure in the population of replicators and the discreteness of the population, which can prevent the extinction caused by parasites. Moreover, if extinction is prevented through spatial pattern formation, the parasite can contribute to the evolution of complexity in RNA-like replicator systems.

Takeuchi N, Hogeweg P. The role of complex formation and deleterious mutations for the stability of RNA-like replicator systems. J Mol Evol. 2007 Dec;65(6):668-86.

Selection pressures involved in pathogen-host interaction

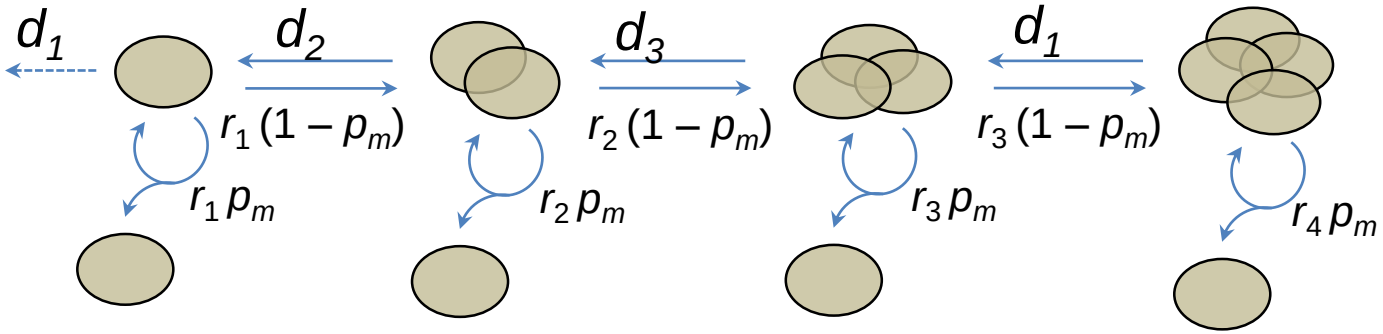


Iranzo et al. Cell Cycle 2015

Population model

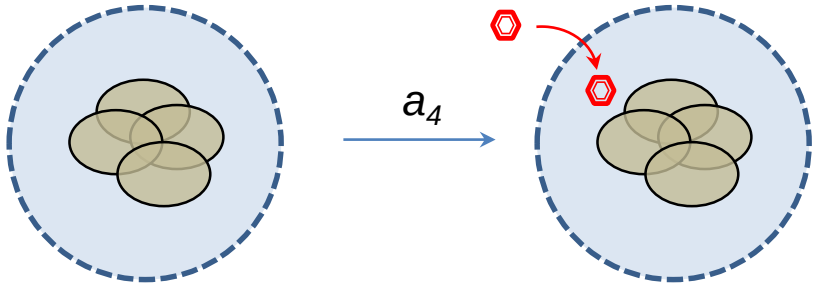
Replication / Death

$$d_k = k\,d$$
$$r_k = k^{\gamma} r$$
$$2/3 \leq \gamma \leq 1$$



Pathogen arrival

$$a_k = k^{\alpha} a$$
$$\alpha = 2/3$$

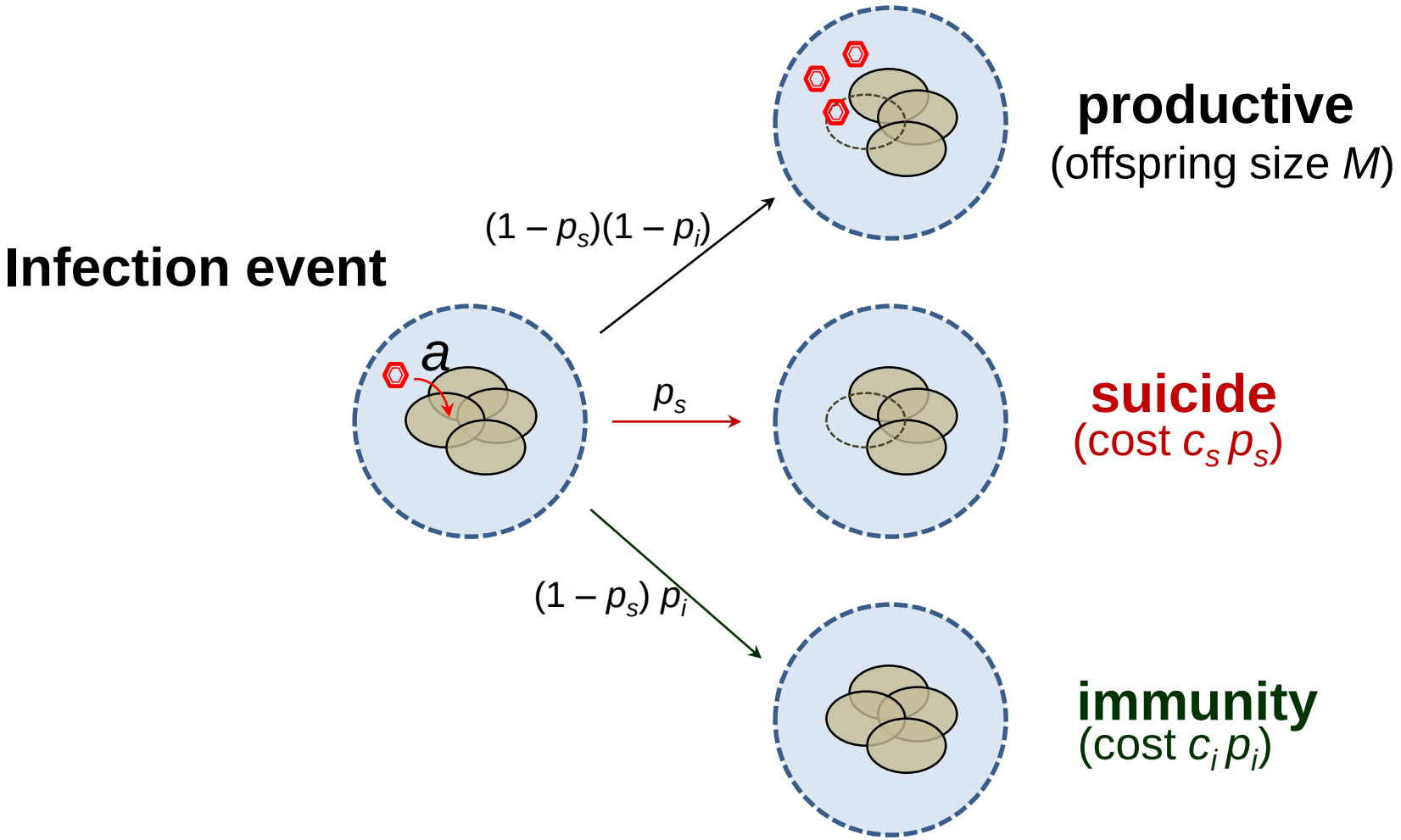


Tradeoff: Opposite effects of multicellularity on replication and pathogen exposure

- **Multicellularity protects against parasite through suicide**

Iranzo et al. Cell Cycle 2015

Population model



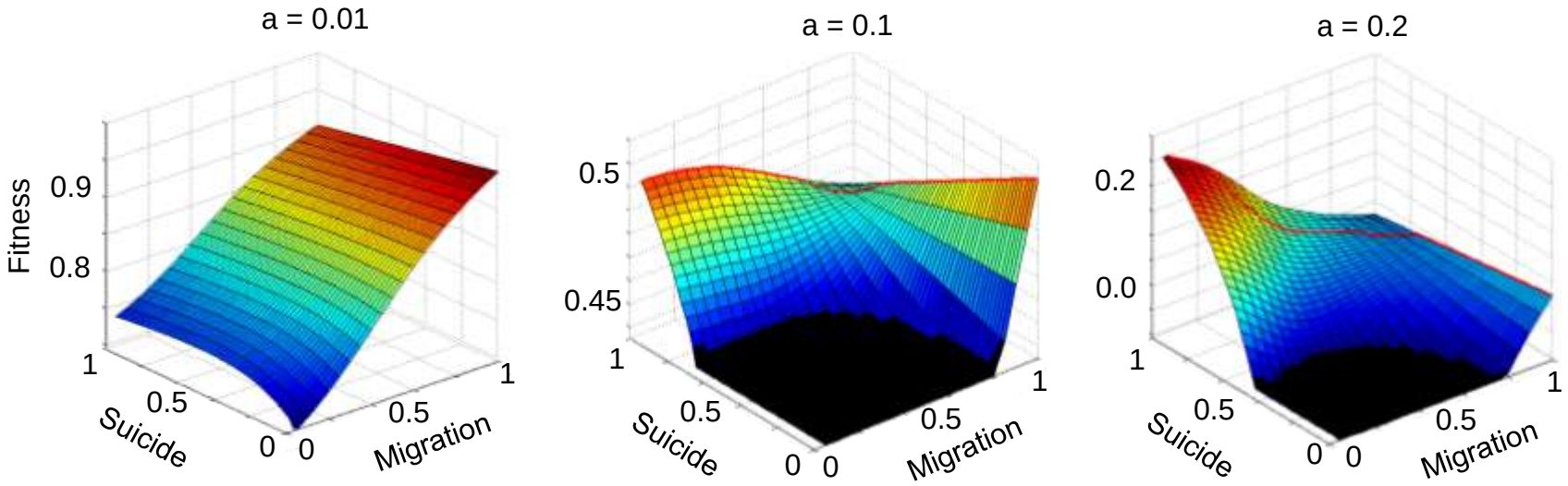
Iranzo et al. Cell Cycle 2015

Evolution of multicellularity/programmed cell suicide under the pressure of pathogens

increasing pathogen pressure

Singletons, no suicide

Clusters, suicide



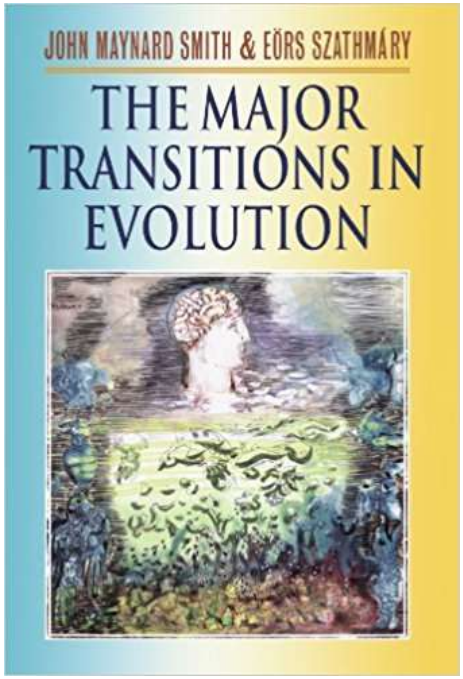
Under strong pathogen pressure, suicide/multicellularity provide the only path to high fitness

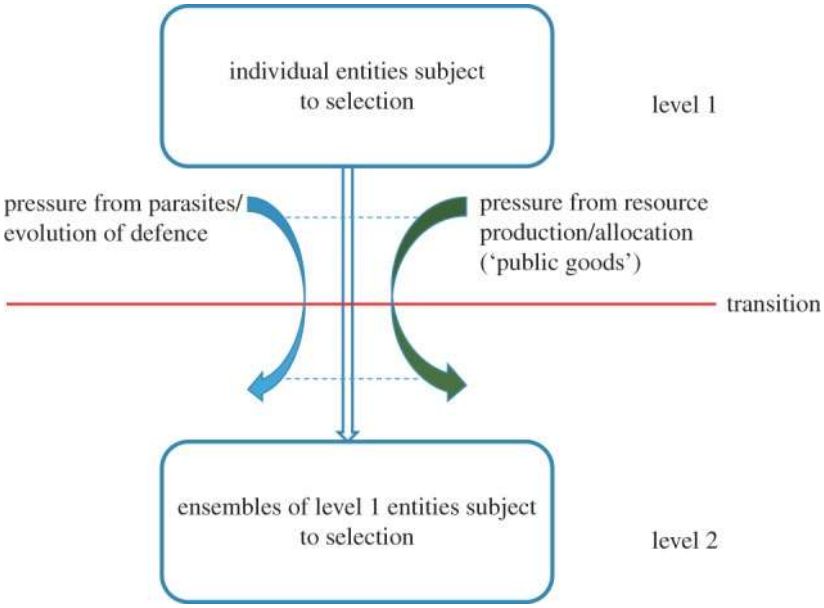
Iranzo et al. Cell Cycle 2015

Major transitions in evolution

Evolutionary Transition in Individuality:
New levels of selection

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





Koonin Phil Trans Royal Soc 2016

- **Evolution of complexity in physics and biology is driven by competing interactions that generate frustrated states in multiple dimensions and self-organized criticality**
- **Emergence and persistence of selfish genetic elements/viruses are inevitable in any evolving replicator system/population:**

  A major biological universal
- **Perennial parasite-host arms race drove evolution of cellular organisms including the first cells, multicellularity and most likely other major evolutionary transitions**
- **Emergence of new levels of complexity from simple physical principles… general physical theory of biology?**