Biological modeling from a condensed matter perspective

Nigel Goldenfeld
Department of Physics
Institute for Genomic Biology
University of Illinois at Urbana-Champaign

“Ask not what physics can do for biology; ask what biology can do for physics”

Stanislaw Ulam

Outline

- What physics can do for biology
  - How biologists are actively seeking input from physicists
  - What problem areas are being ignored by physicists
  - Examples

- What biology can do for physics
  - Why biological problems are of interest to physicists
  - Is there “new physics” in biology?

- Challenges and obstacles to the participation of physicists in biology

My perspective

- Unifying theme of condensed matter physics is arguably collective phenomena
    - Superfluidity in He II as well as neutron stars
    - BCS pairing in superconductors and QCD
    - Bose-Einstein condensation in atomic gases
    - Turbulence in water and in quark-gluon plasma

- Key difference between condensed matter and biological physics: importance of problem choice
    - Condensed matter: typically a few problems of over-riding importance at any one time
    - Biological matter: no clear consensus on a few hot topics
    - Implication: problem selection is a key “success factor” in biological physics

- My perspective: look for biology problems dominated by collective effects
    - Generally overlooked by the biophysics community (focus more on single-molecule biophysics)
    - Condensed matter physicists have something new to offer biologists

My perspective

- Condensed matter physics
  - Statistical physics of polymers
  - Nonequilibrium pattern dynamics, materials processing
  - Renormalization group far from equilibrium
  - Phase transition kinetics and topological defects
  - High temperature superconductivity (d-wave pairing and critical phenomena)
  - Fluid turbulence
- Biological physics
  - Scaling laws in ecology
  - Dynamics of speciation in microbes
  - Evolution of the genetic code
  - Ecology of microbes at geothermal hot springs
  - Phenotype variation in biofilms
  - Nonlinear dynamics of genomes over evolutionary time
  - Evolution of the rare biosphere

Outline

- What physics can do for biology
  - How biologists are actively seeking input from physicists
  - What problem areas are being ignored by physicists

- What biology can do for physics
  - Why biological problems are of interest to physicists
  - Is there “new physics” in biology?

- Challenges and obstacles to the participation of physicists in biology

Biocomplexity

- What is “complexity”?
  - Complexity = structure + large fluctuations
- The big idea: research on the individual components of complex systems provides only limited information about the behavior of the systems themselves
- Biocomplexity arises from interplay of complex biological, physical and even social systems

What is biocomplexity?

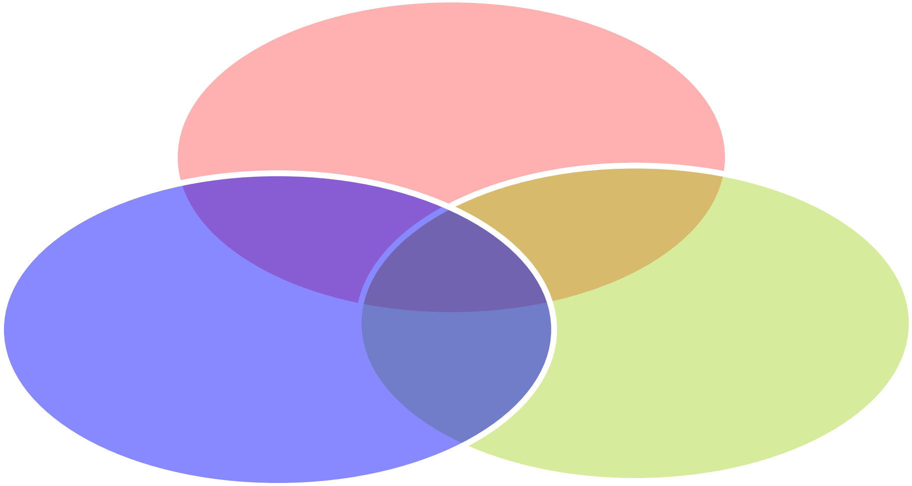
NSF Solicitation 00-22

- “Biocomplexity arises from dynamics spanning several levels within a system, between systems, and/or across multiple spatial (microns to thousands of kilometers) and temporal (nanoseconds to eons) scales.
- This special competition will specifically support Research Projects which directly explore nonlinearities, chaotic behavior, emergent phenomena or feedbacks within and between systems and/or integrate across multiple components or scales of time and space in order to better understand and predict the dynamic behavior of systems.”

Why do we care?

- **Steven Jay Gould, New York Times, Feb 19, 2001**
  - “Homo sapiens possesses between 30,000 and 40,000 genes... In other words, our bodies develop under the directing influence of only half again as many genes as the tiny roundworm ....”
  - “The collapse of the doctrine of one gene for one protein, and one direction of causal flow from basic codes to elaborate totality, marks the failure of reductionism for the complex system that we call biology.”
  - “First, the key to complexity is not more genes, but more combinations and interactions generated by fewer units of code — and many of these interactions (as emergent properties, to use the technical jargon) must be explained at the level of their appearance, for they cannot be predicted from the separate underlying parts alone.”

Examples of biocomplexity



- Commonality
  - Modelling on many scales of space and time
  - Interconnectedness – networked dynamical elements
  - Feedback and other cooperative dynamics

- Microbial ecology
  - Microbe-microbe interactions
  - Microbe-geochemical cycles
  - Feedback: microbes modify environment, environment influences microbial population biology
    - Which dynamics are key?
- Systems biology
  - Cell cycle, metabolic cycles, …
  - Interactions between reactants give coupled rate equations, diffusion
    - Which dynamics are key?
- Evolution and emergence
  - Agent-based modelling of competing or co-evolving organisms, e.g. formation of biochemical niches
    - Collective dynamics
    - Which dynamics are key?

Biology from a physics perspective

Biologists are interested in what makes organisms different.

Physicists are interested in what makes them all the same.

Biology from a physics perspective

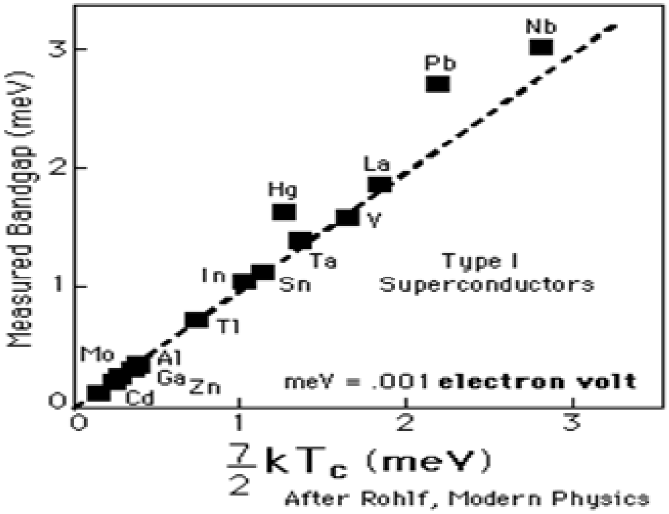
Biologists are interested in what makes organisms different.

Physicists are interested in what makes them all the same.

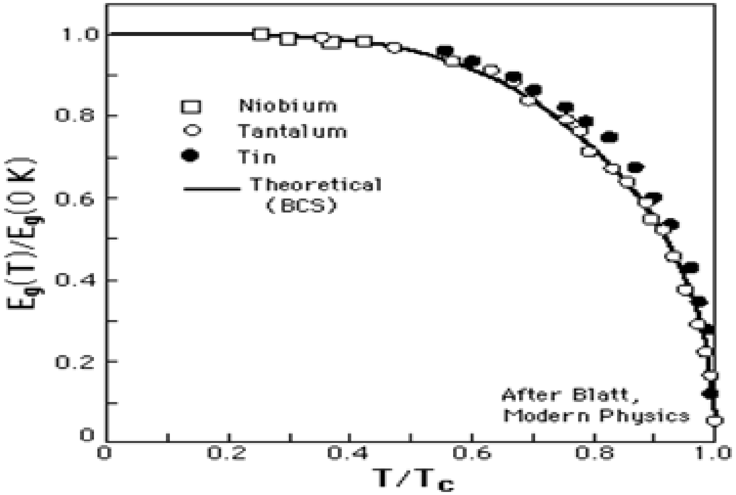
- **This means that biologists and physicists tend to have complimentary approaches to biological phenomena**
  - Condensed matter physics: “That which is not forbidden is mandatory”
    - Look for interesting processes and interactions, then ask how they are realized in the world of atoms
  - Biology: see what’s there, understanding is *ad hoc* and discovery-based
    - Little real tradition of interplay between experiment and theory to advance understanding.

Superconductivity – it's not about the atoms

- **Example: Bardeen-Cooper-Schrieffer theory of superconductivity**
  - Most successful many-body theory
  - Spectacular agreement with experiment:
    - Universal ratios: $\Delta/k_BT_c = 3.53$
    - Universal functions: $\Delta(T)$
  - Predictions same for all materials to which the theory applies

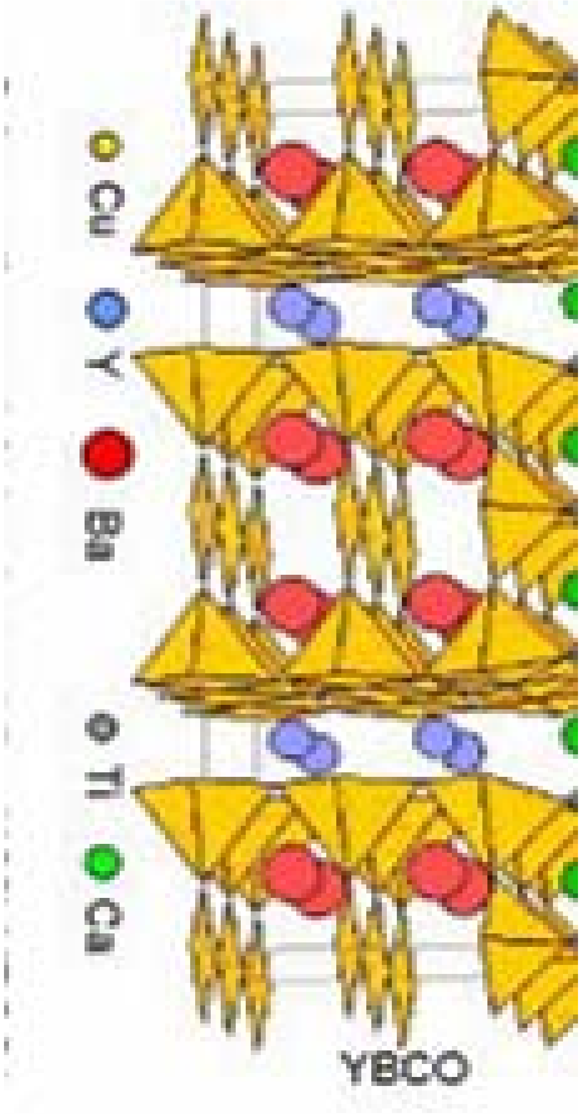


After Rohlf, Modern Physics



After Blatt, Modern Physics

Superconductivity – it's not about the atoms



J. Eckstein

- **BCS model leaves out much physics and all chemistry**
  - No atoms, no phonons, no electronic band structure
  - Just hopping electrons that can pair
  - Cooperativity is the essence!

- **But poor ability to predict T_c.**
  - Must ask what is the regime of validity of the theory

Superconductivity – it's not about the atoms

- **Moral of the story:**
  - Focus on the process, not the realization
    - Process = quantum dynamics of correlated electrons
    - Realization = Y-Ba-Cu-O atoms that you need to mix up in the lab to create something that manifests the process
  - Cannot explain superconductivity by just thinking about atoms: must understand their cooperative behavior.
- **Focus on what really matters!**

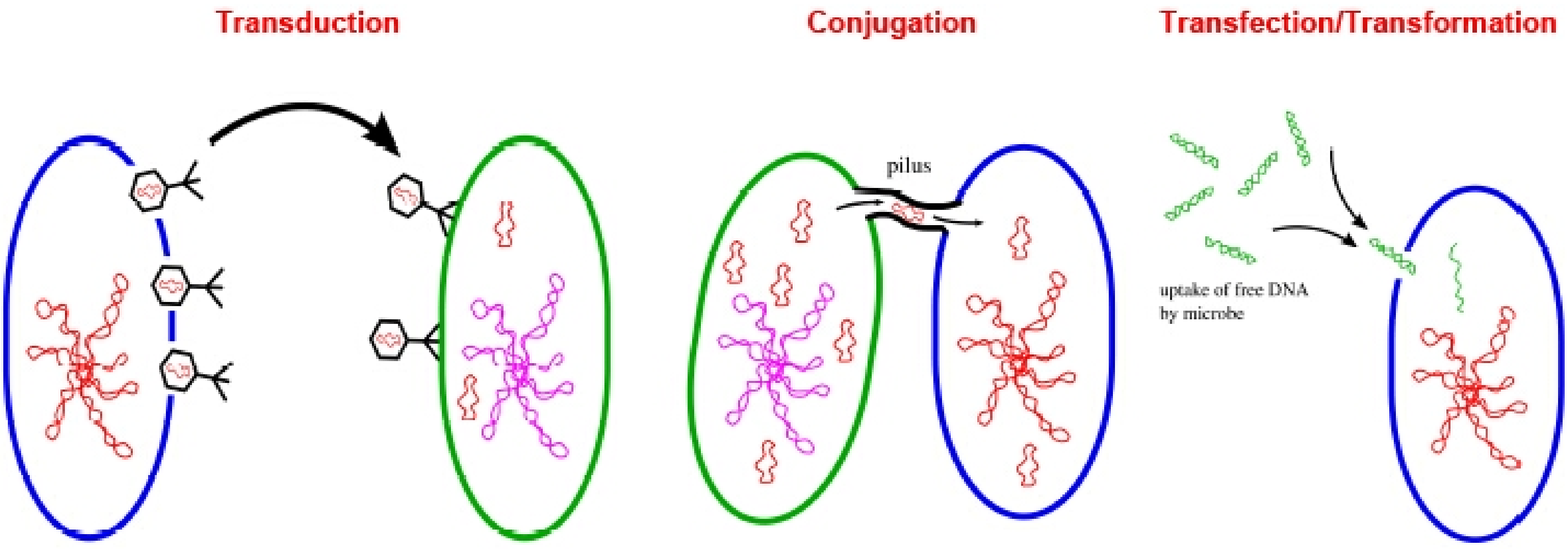
Life – it's not about the molecules

- **Moral of the story:**
  - Focus on the process, not the realization
    - Process = evolution
    - Realization = C, H, N, O, P, ... atoms that you need to mix up in the lab to create something that manifests the process
  - Cannot explain life by just thinking about atoms: must understand their cooperative behavior.
- **Focus on what really matters!**

What matters for life?

Cooperative effects

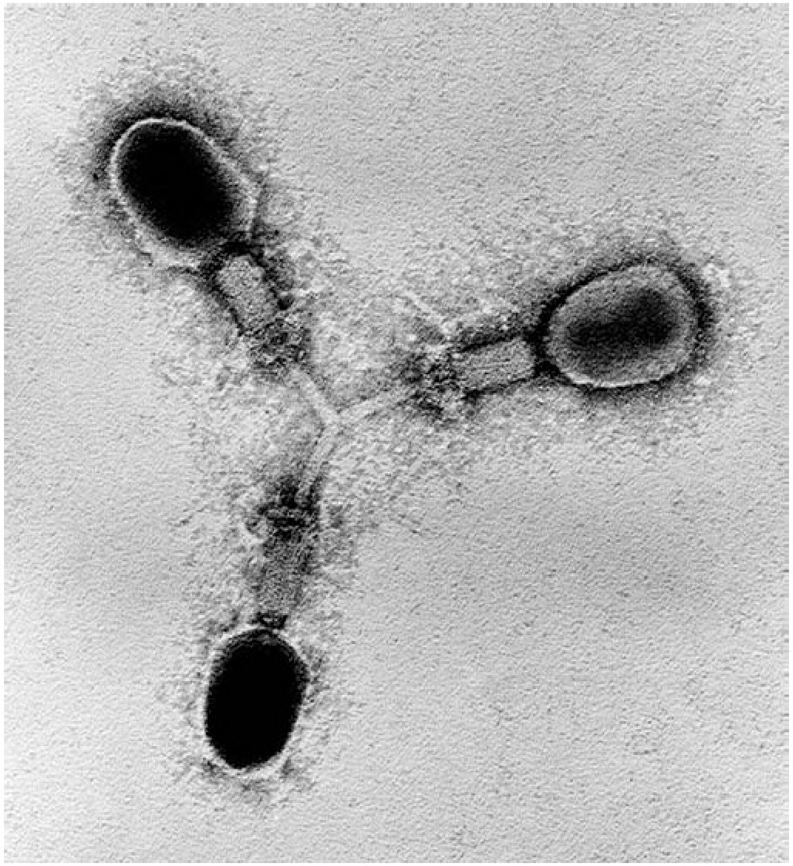
Horizontal gene transfer



Microbes can do this … but what happens when they all do it?

Gene transfer between host and virus

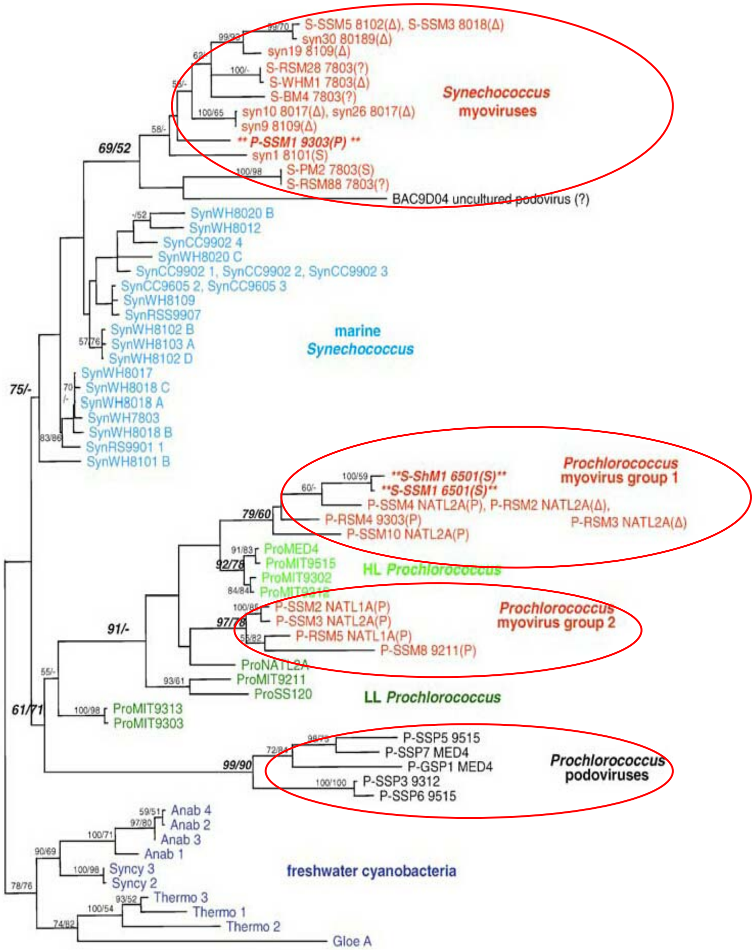
Sullivan et al. , PlosBiol (2006)



DOI: 10.1371/journal.pbio.0040264.g001

Cyanophages—viruses that infect photosynthetic marine bacteria—not only possess genes for photosynthesis but also exchange genetic material with their cyanobacterial hosts.

Hill, PlosBiol (2006)



PsbA gene acquired by phage

Phylogeny of psbA gene in cultured cyanobacteria and cyanophages

Is there a benefit to microbes of viruses?

“Therefore, mounting evidence indicates that host-like genes acquired by phages undergo a period of diversification in phage genomes and serve as a genetic reservoir for their hosts. Thus, a complex picture of overlapping phage and host gene pools emerges, where genetic exchange across these pools leads to evolutionary change for host and phage. Fully understanding the mechanisms of microbial and phage coevolution clearly requires an improvement in our ability to quantify horizontal gene transfer at the whole and partial gene level and in our ability to accurately estimate the relative fluxes into and out of these pools.” (Sullivan et al. 2006)

Yes: microbe-phage interactions create a global reservoir of photosynthetic genes, benefiting both microbes and phages. **(E. Anderson (1966), N. Anderson (1970), S. Sonea (1988, 2001), M. Syvanen (1984) & many others, including L. Villareal, Weinbauer, Ochman, Lawrence, Groisman, Hatfull, Hendrix, Brussow ...)**

Spread of antibiotic resistance genes

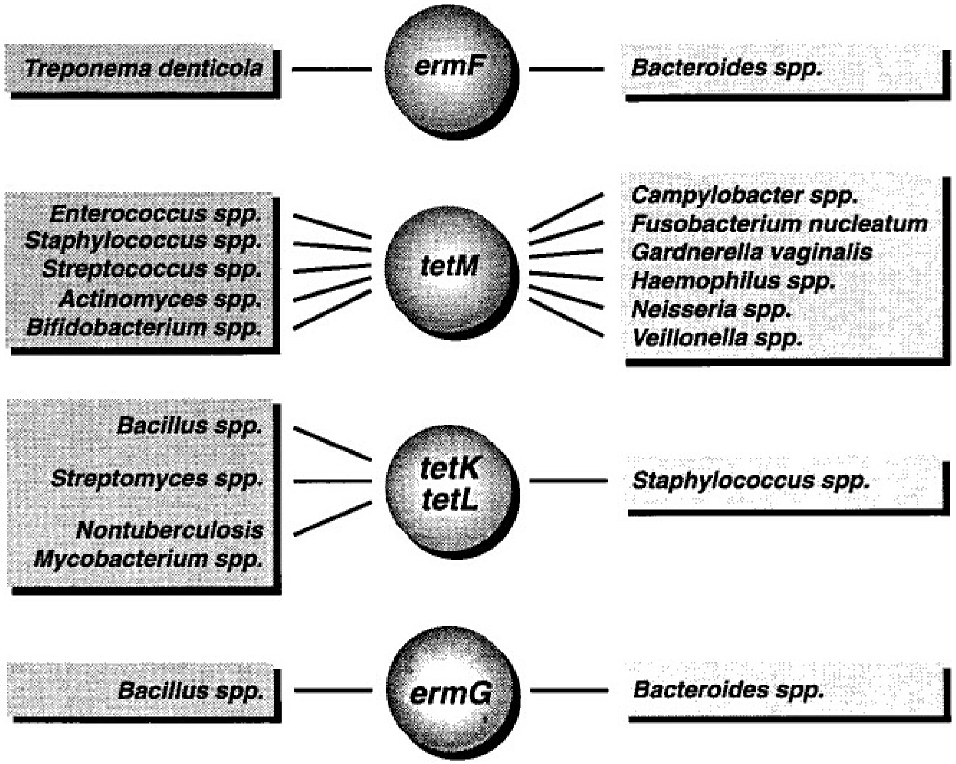


FIG. 1. Evidence that broad-host-range transfer of resistance genes occurs readily in nature (18–20). In each of the cases depicted here, virtually identical copies of the gene in the circle have been found in the species listed in the boxes connected to the circle. Such examples do not prove direct transfer between the species listed but indicate that there is some way for genes to move, whether directly or indirectly, between these genera.

Salyers & Amabile-Cuevas (1997)

- **Virtually identical copies of resistance genes found in distantly related bacteria**
  - Genes are being expressed
- **Genes cross species and phylum boundaries**
  - Gram-positive/enteric
  - Bacteroides/enteric
- **Genes cross physical locations**
  - Bacteroides spp. (colon)/Bacillus spp. (soil)

NG and C. Woese, Nature **445**, 369 (2007)

Fully-referenced version can be downloaded from the quantitative biology arxiv q-bio.PE/0702015 or from

http://guava.physics.uiuc.edu

Biology's next revolution

The emerging picture of microbes as gene-swapping collectives demands a revision of such concepts as organism, species and evolution itself.

Nigel Goldenfeld and Carl Woese

One of the most fundamental patterns of scientific discovery is the revolution in thought that accompanies a new body of data. Satellite-based astronomy has, during the past decade, overthrown our most cherished ideas of cosmology, especially those relating to the size, dynamics and composition of the Universe.

Similarly, the convergence of fresh theoretical ideas in evolution and the coming avalanche of genomic data will profoundly alter our understanding of the biosphere — and is likely to lead to revision of concepts such as species, organism and evolution. Here we explain why we foresee such a dramatic transformation, and why we believe the molecular reductionism that dominated twentieth-century biology will be superseded by an interdisciplinary approach that embraces collective phenomena.

The place to start is horizontal gene transfer (HGT), the non-genealogical transfer of genetic material from one organism to another — such as from one bacterium to another or from viruses to bacteria. Among microbes, HGT is pervasive and powerful — for example, in accelerating the spread of antibiotic resistance. Owing to HGT, it is not a good approximation to regard microbes as organisms dominated by individual characteristics. In fact, their communications by genetic or quorum-sensing channels indicate that microbial behaviour must be understood as predominantly cooperative.

In the wild, microbes form communities, invade biochemical niches and partake in biogeochemical cycles. The available studies strongly indicate that microbes absorb and discard genes as needed, in response to their environment. Rather than discrete genomes, we see a continuum of genomic possibilities, which casts doubt on the validity of the concept of a 'species' when extended into the microbial realm. The uselessness of the species concept is inherent in the recent forays into metagenomics — the study of genomes recovered from natural samples as opposed to clonal cultures. For example, studies of the spatial distribution of rhodopsin genes in marine microbes suggest such genes are 'cosmopolitan', wandering among bacteria (or archaea) as environmental pressures dictate.

Equally exciting is the realization that viruses have a fundamental role in the biosphere, in both immediate and long-term evolutionary senses. Recent work suggests that viruses are an important repository and memory of a community's genetic information, contributing to the system's evolutionary dynamics and stability. This is hinted at, for example, by prophage induction, in which viruses latent in cells can become activated by environmental influences. The ensuing destruction of the cell and viral replication is a potent mechanism for the dispersal of host and viral genes.

It is becoming clear that microorganisms have a remarkable ability to reconstruct their genomes in the face of dire environmental stresses, and that in some cases their collective interactions with viruses may be crucial to this. In such a situation, how valid is the very concept of an organism in isolation? It seems that there is a continuity of energy flux and informational transfer from the genome up through cells, community, virosphere and environment. We would go so far as to suggest that a defining characteristic of life is the strong dependency on flux from the environment — be it of energy, chemicals, metabolites or genes.

Nowhere are the implications of collective phenomena, mediated by HGT, so pervasive and important as in evolution. A computer scientist might term the cell's translational apparatus (used to convert genetic information to proteins) an 'operating system', by which all innovation is communicated and realized. The fundamental role of translation, represented in particular by the genetic code, is shown by the clearly documented optimization of the code. Its special role in any form of life leads to the striking prediction that early life evolved in a lamarckian way, with vertical descent marginalized by the more powerful early forms of HGT.

Refinement through the horizontal sharing of genetic innovations would have triggered an explosion of genetic novelty, until the level of complexity required a transition to the current era of vertical evolution. Thus, we regard as regrettable the conventional concatenation of Darwin's name with evolution, because other modalities must also be considered.

This is an extraordinary time for biology, because the perspective we have indicated places biology within a context that must necessarily engage other disciplines more strongly aware of the importance of collective phenomena. Questions suggested by the generic energy, information and gene flows to which we have alluded will probably require resolution in the spirit of statistical mechanics and dynamical systems theory. In time, the current approach of post-hoc modelling will be replaced by interplay between quantitative prediction and experimental test, nowadays more characteristic of the physical sciences.

Sometimes, language expresses ignorance rather than knowledge, as in the case of the word 'prokaryote', now superseded by the terms archaea and bacteria. We foresee that in biology, new concepts will require a new language, grounded in mathematics and the discoveries emerging from the data we have highlighted. During an earlier revolution, Antoine Lavoisier observed that scientific progress, like evolution, must overcome a challenge of communication: "We cannot improve the language of any science without at the same time improving the science itself; neither can we, on the other hand, improve a science without improving the language or nomenclature which belongs to it." Biology is about to meet this challenge. ■

Nigel Goldenfeld is in the Department of Physics and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA. Carl Woese is in the Department of Microbiology and Institute for Genomic Biology, 601 South Goodwin Avenue, Urbana, Illinois 61801, USA.

FURTHER READING

Frigaard, N., Martinez, A., Mincer, T. & DeLong, E. *Nature* **439**, 847–850 (2006).
Sullivan, M. *et al. PLoS Biol.* **4**, e234 (2006).
Pedulla, M. *et al. Cell* **113**, 171–182 (2003).
Vetsigian, K., Woese, C. & Goldenfeld, N. *Proc. Natl Acad. Sci. USA* **103**, 10696–10701 (2006).

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

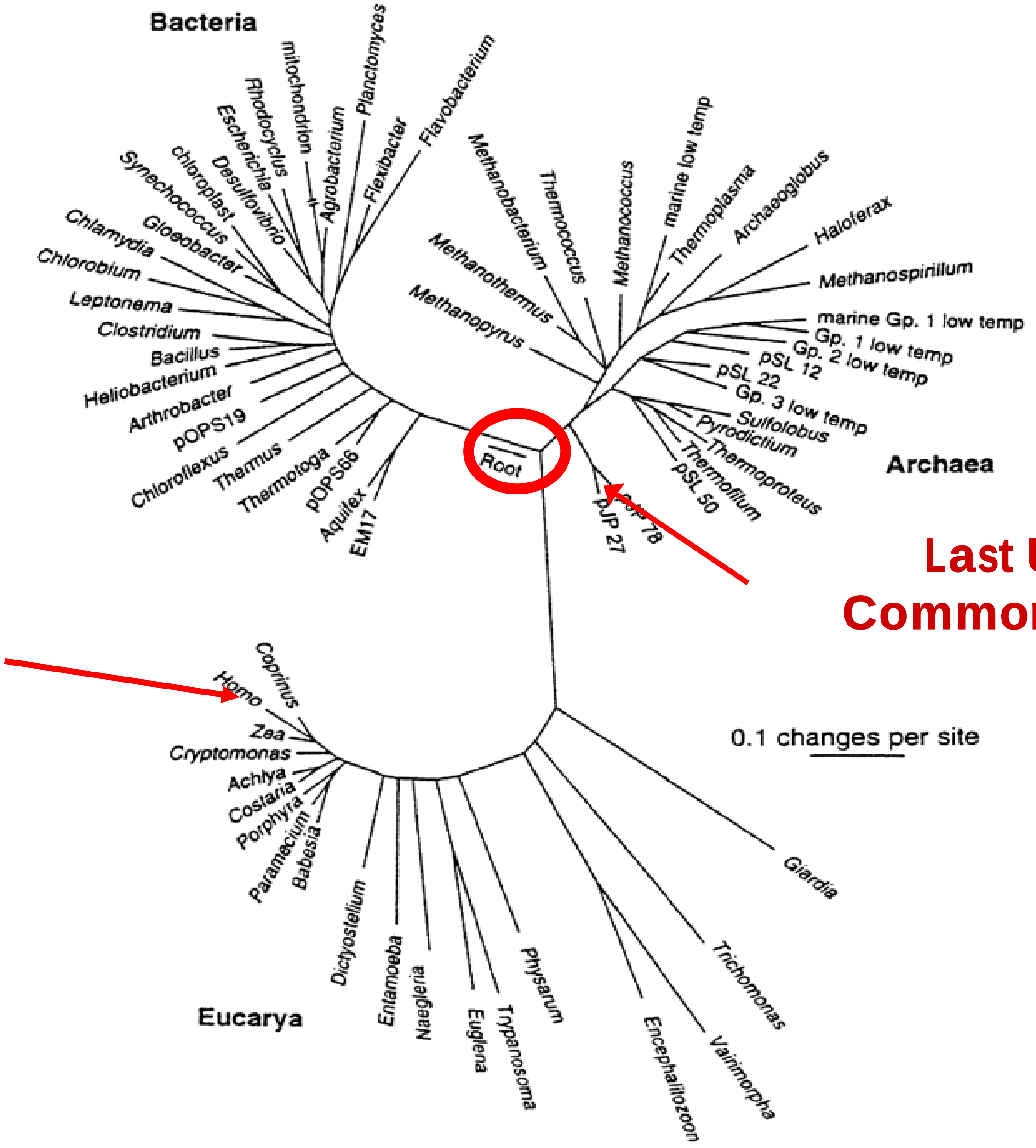
KAPUSTA

Statistical mechanics of genes: impact on evolution

1. The genetic code

K. Vetsigian, C.R. Woese and Nigel Goldenfeld. Communal evolution of the genetic code. *Proc. Natl. Acad. Sci.* 103, 10696-10701 (2006)

The Tree of Life



You are here

Last Universal Common Ancestor ?

- NR Pace, Science 1997

If we rebooted life on Earth, would we have the same genetic code?

Yes! (or very similar)

The canonical genetic code

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

The yeast (mitochondrial) genetic code

Code still evolving (slightly).

Recent changes, may involve modern translation machinery, so mechanism may be different from that before last common ancestor.

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Thr	Pro	His	▨	U
					C
			Gln		A
					G
A	Ile --------- Met	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

Degeneracy

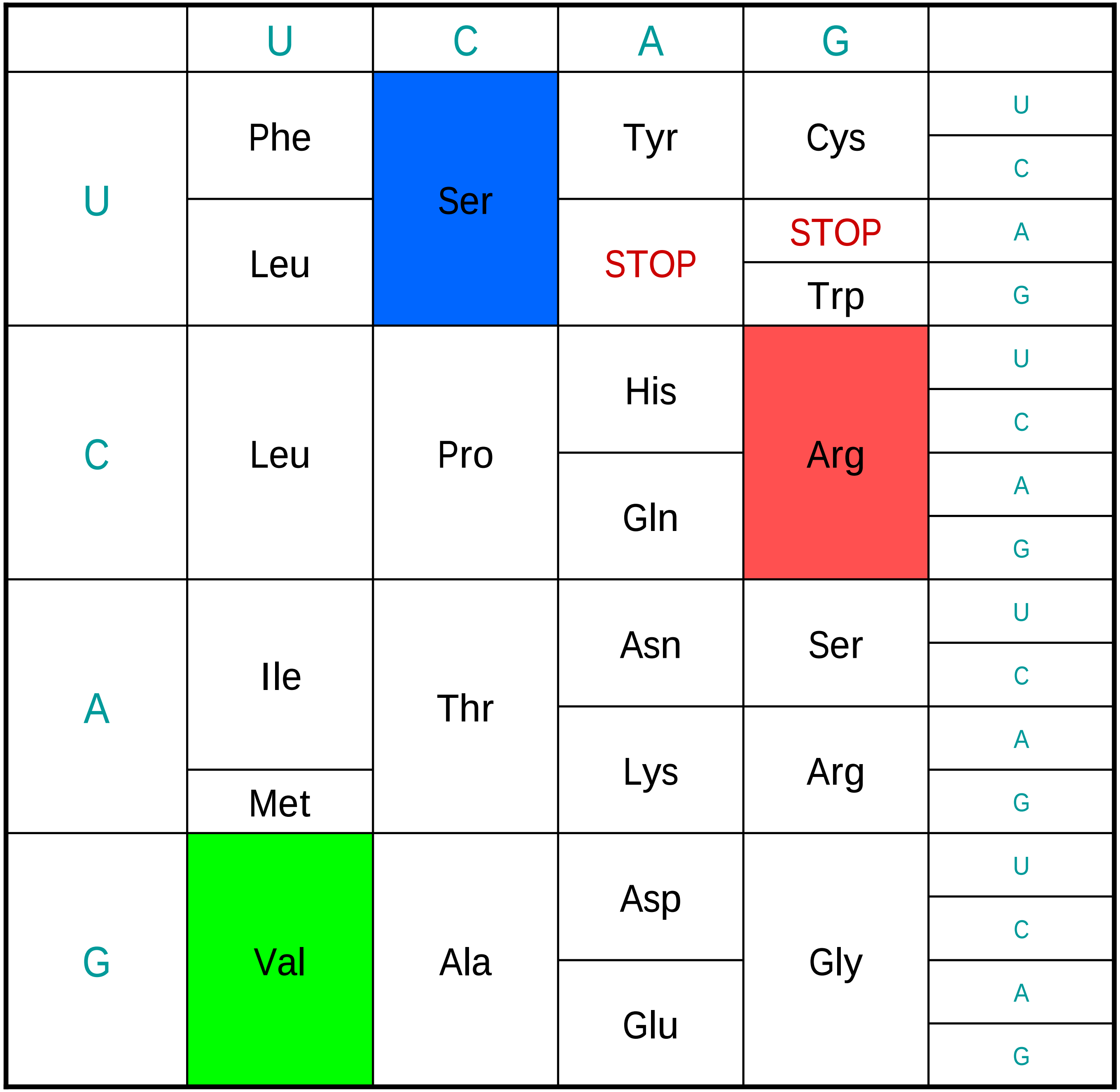
Met & Trp are only amino acids with one codon

Sonneborn (1965), Woese (1965), Zuckerkandl and Pauling (1965)

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

Clustering

Amino acids not scattered randomly but occur in blocks



	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

GC content related to degeneracy

GC binding stronger than AU. First doublets are G,C gives 4-codon block.

Stronger binding to cognate anti-codon implies no need for third codon to provide discrimination?

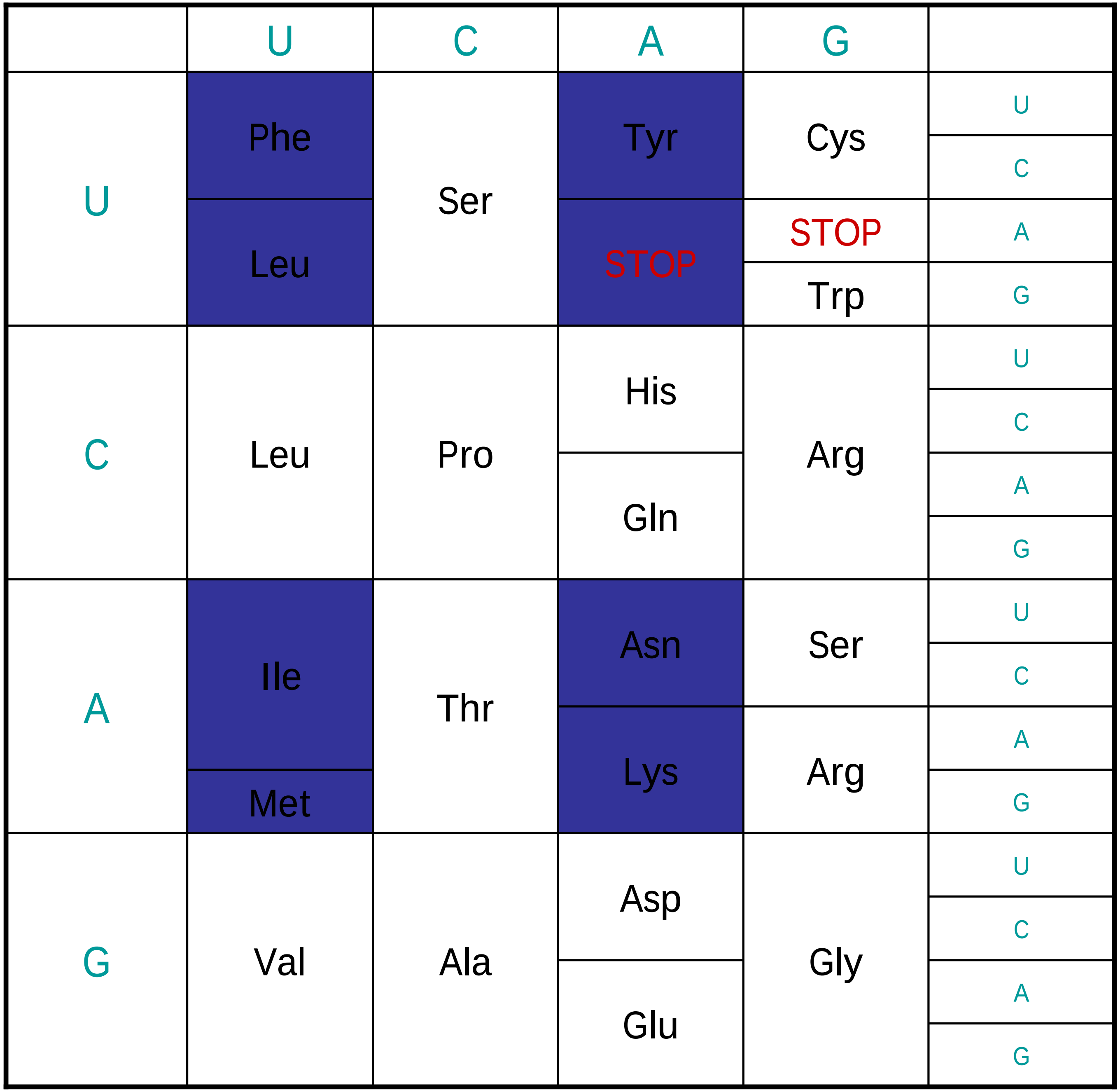
Lagerkvist (1978)

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

GC content related to degeneracy

GC binding stronger than AU. First doublets are A,U gives split blocks.

Weaker binding to cognate anti-codons means third codon needed for discrimination?



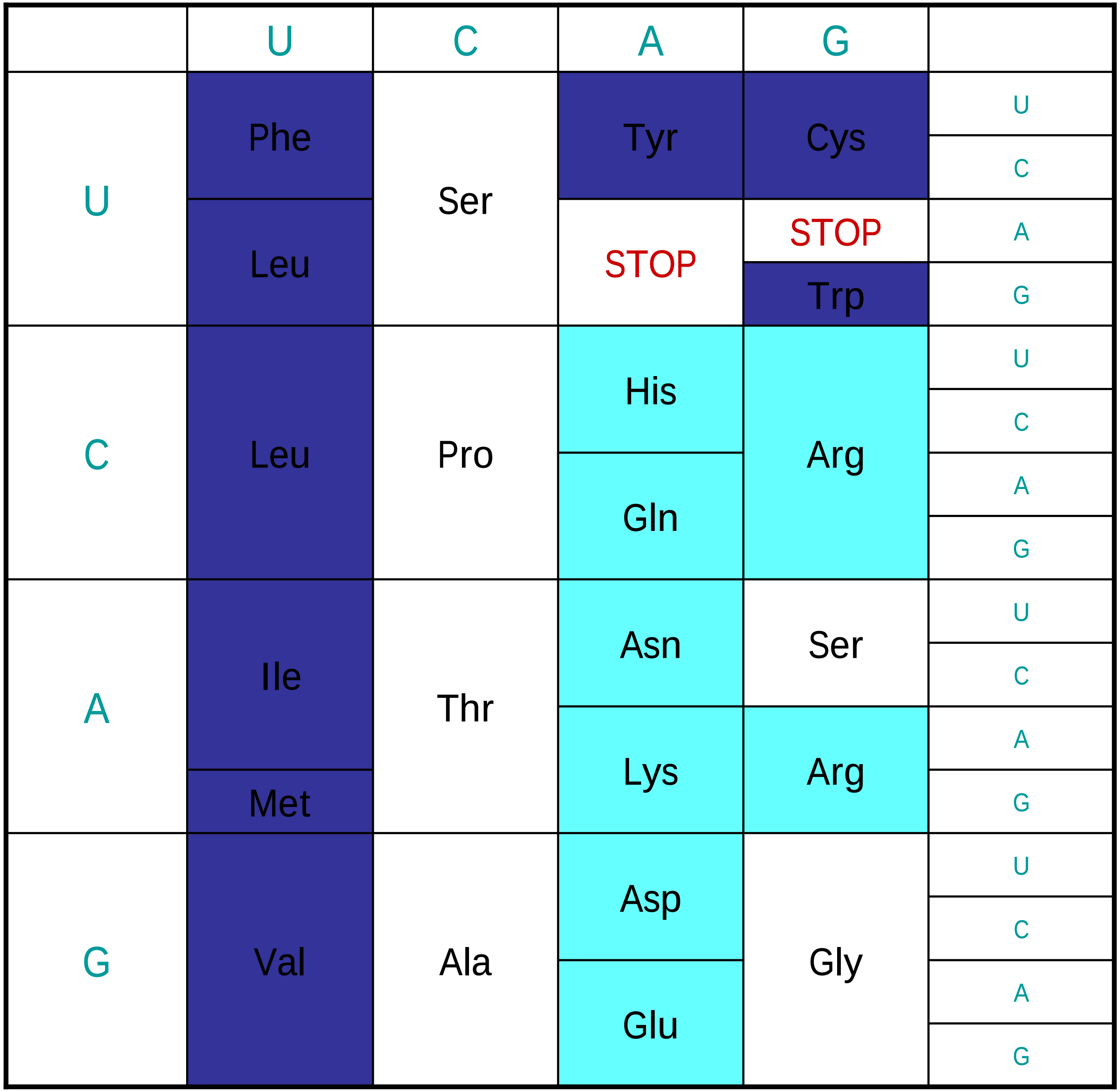
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

Hydrophobicity

Most hydrophobic amino acids are Phe, Leu, Ile, Met and Val.

Most hydrophilic amino acids are His, Gln, Asn, Lys, Asp, Glu.

Amino acids with complementary anti-codons tend to have opposite hydrophobicity.

Woese (1965), Volkenstein (1966)

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

Polar requirement

Amino acids with shared doublet have similar “polar requirement” – a quantification of amino acid-pyridine affinity.

(Woese et al. 1966)

	U	C	A	G	
U	Phe 5.0	Ser	Tyr	Cys	U
					C
	Leu 4.9		STOP	STOP	A
				Trp	G
C	Leu	Pro	His 8.4	Arg	U
					C
			Gln 8.6		A
					G
A	Ile 4.9	Thr	Asn 10.0	Ser	U
					C
			Lys 10.1	Arg	A
	Met 5.3				G
G	Val	Ala	Asp 13.0	Gly	U
					C
			Glu 12.5		A
					G

Optimality of the code

- Does the genetic code minimise errors?
  - Point mutations tend to substitute similar amino acids? (Sonneborn 1965)
  - Errors in translation tend to substitute similar amino acids? (Woese 1965)
- How can we explore such issues, when we have only one universal code?
  - Computer simulation! (Alff-Steinberger 1969)

Simulated genetic codes

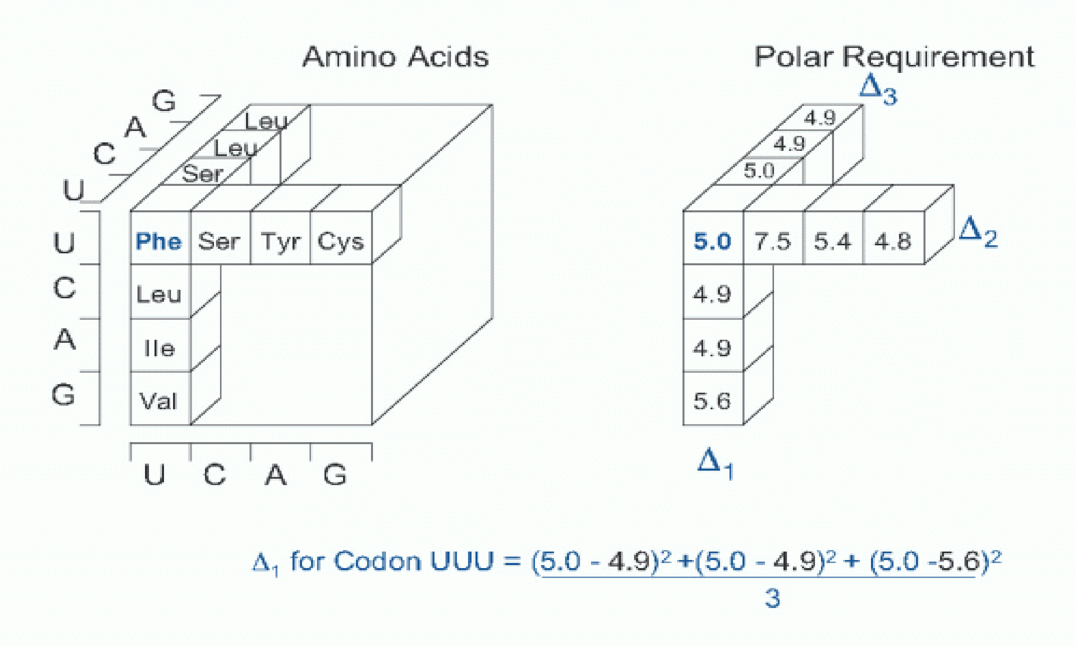
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Leu	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Asn	Ser	U
					C
			Lys	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

	U	C	A	G	
U	Leu	Ser	Tyr	Cys	U
					C
	Leu		STOP	STOP	A
				Trp	G
C	Phe	Pro	His	Arg	U
					C
			Gln		A
					G
A	Ile	Thr	Lys	Ser	U
					C
			Asn	Arg	A
	Met				G
G	Val	Ala	Asp	Gly	U
					C
			Glu		A
					G

- Permute labels – new codes with same pattern of degeneracy
- 20! ~ 10^{18} possible codes

Score codes on their tolerance to mutations or errors

- Basic idea: generate by Monte Carlo simulation a large number of simulated genetic codes
- For each code, score the effect of point substitutions in 1st, 2nd & 3rd codon positions, summed over the whole code
- Plot a histogram of the scores obtained
- Compare with the canonical genetic code

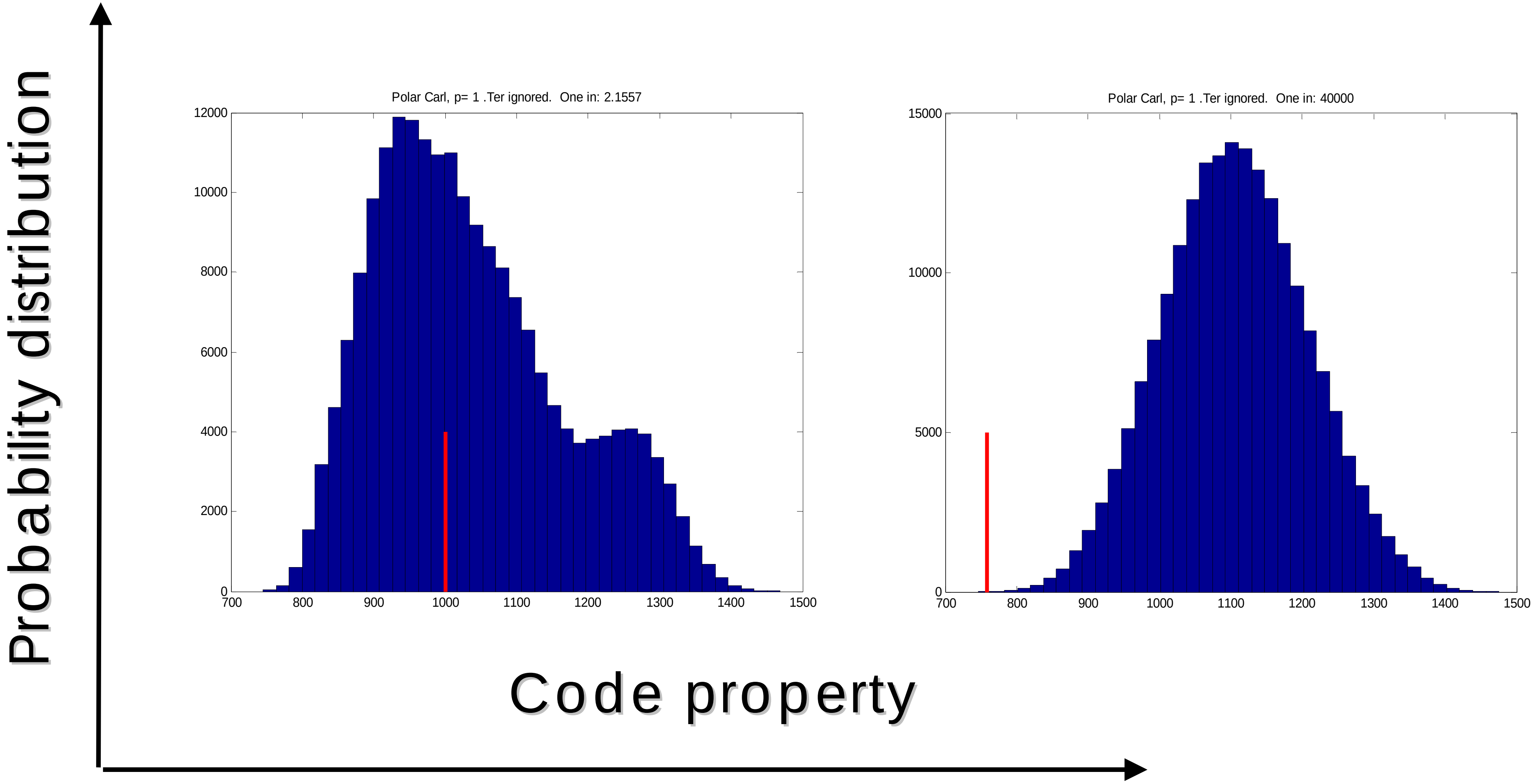


Knight Ph.D thesis (2001)

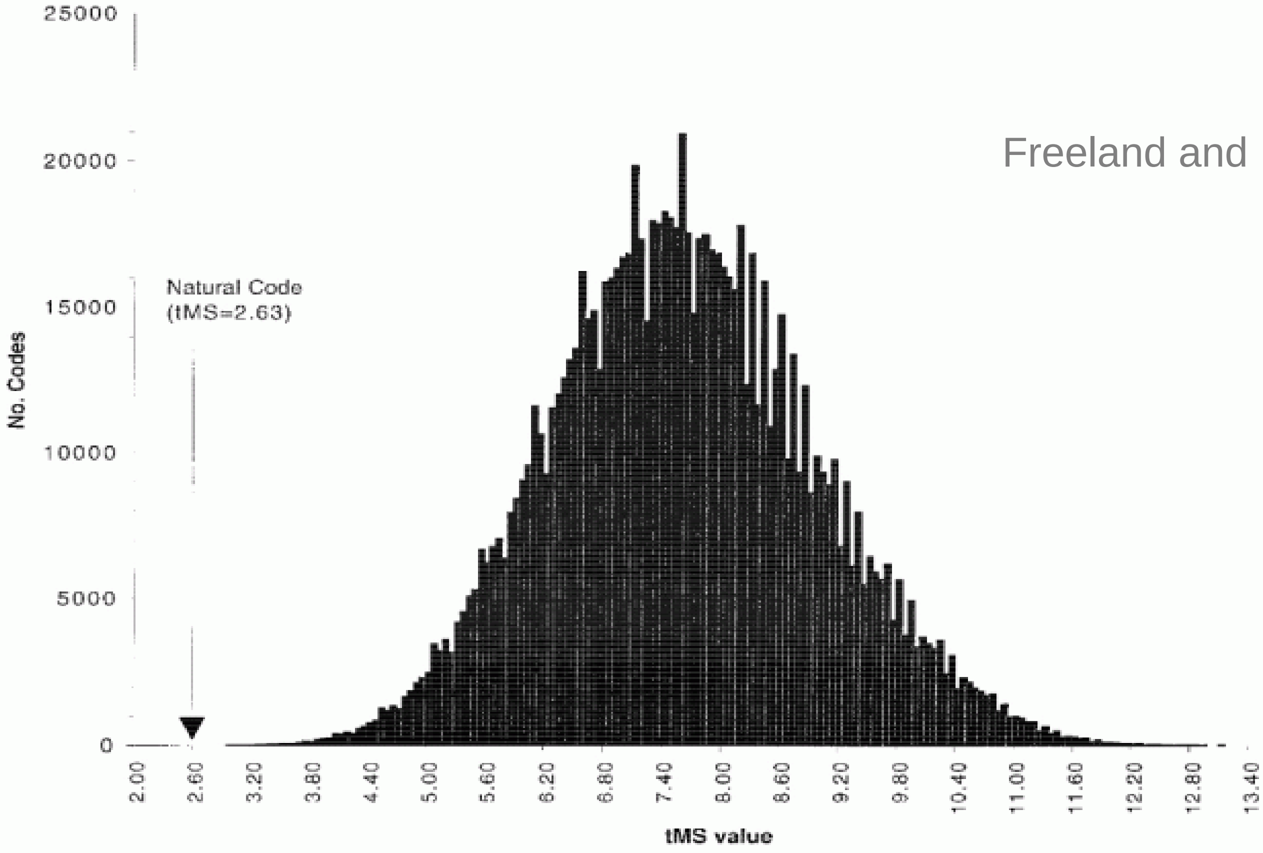
Simulated genetic codes

Naïve expectation – current code is frozen accident

Actual result – current code is not a frozen accident



Optimality of the genetic code with respect to the polar requirement



Freeland and Hurst 1998

Fig. 7. Frequency distribution for the tMS0 (=MS0 adjusted for mistranslation parameters) values obtained from 1 million randomly generated variants of the natural genetic code. The *X* axis gives a particular range of categories of MS values, and the *Y* axis gives the number of random variant codes generated with an MS value in that category (from a sample of 1 million random variant codes tested). In addition, the arrow indicates the category into which the tMS0 calculation for the natural code falls: the cumulative frequency to the left of this arrow therefore indicates the proportion of more conservative codes found among the random variants. This cumulative frequency is in fact 1 (i.e., only 1 of the 1 million variants had a lower tMS value), indicating that under our quantification of mistranslation parameters, the probability of a code as efficient as or more efficient than the natural code evolving by chance alone is 0.000001.

Optimality of the genetic code with respect to the polar requirement

- Haig and Hurst (1991) simulated 10,000 genetic codes.
  - Only 2 were better than the genetic code in minimising errors

- Freeland and Hurst (1998) simulated 1,000,000 genetic codes
  - Weight transition and transversion differently
    - Transitions: purine-purine (A,G); transversion: purine-pyridine (U,C)

- Butler et al. (2007) extended the analysis to show that the probability of finding a more optimal code than the canonical one is $(26 \pm 1.6) \times 10^{-7}$
  - Also developed a theoretical measure based on the radial correlation function of amino acid in water-pyridine mixtures
  - This computational polar requirement gives an optimality probability of $(19 \pm 4) \times 10^{-8}$

Popularity contest

- Genetic code is not just one more trait, it is an innovation-sharing protocol.
- The more users a code has, the more beneficial traits are discovered and distributed
- Organisms having more popular codes are
  - more protected against invasions from organisms having different codes
  - more likely to invade other niches
- The most popular code wins
  - Not the most optimal code!
- *Universality* is the only stable solution

Simulations of code evolution

Coevolution model

- Asexual population
- Phenotype of individuals is distribution of proteins
  - Fitness is a function of the phenotype
- Proteins obtained by translating genome with code, with errors
- Individual reproduction rate function of fitness
- Messages change faster than codes:
  - Quasi-static equilibrium: codon usage equilibrates to code
  - Mutate code
  - Mutant code with higher fitness than existing code with existing message can invade the population
- **Hence, code can evolve due to selection at the phenotype!**

What do we measure?

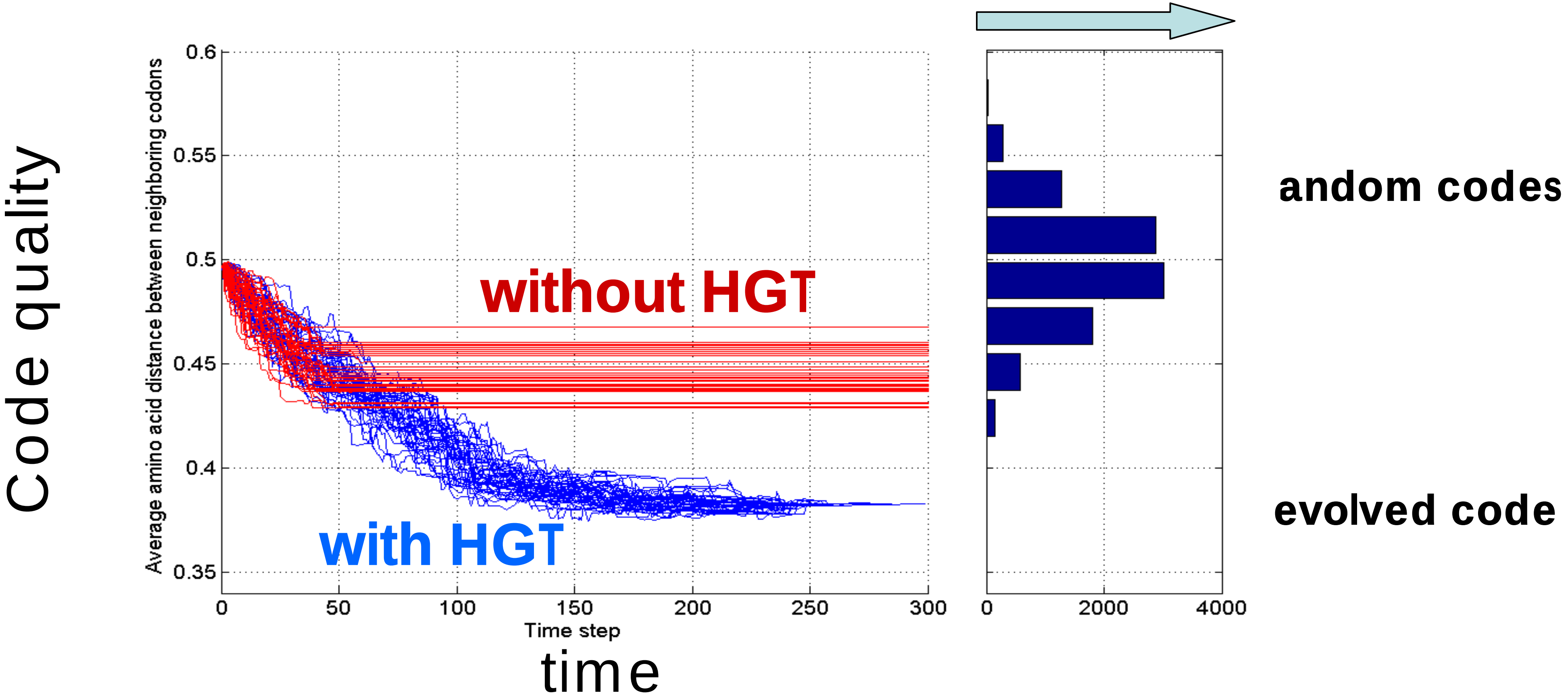
- We are trying to understand the universality and optimality of the genetic code, so need proxies for these characteristics.

- Optimality: reflects "error-minimization" aspects of the code.
  - Average amino acid distance between neighbouring codons

- Universality: how many codes present in the population, and how different are they?
  - Average distance between codes in the simulation

Evolution of code quality

Distribution of code quality scores

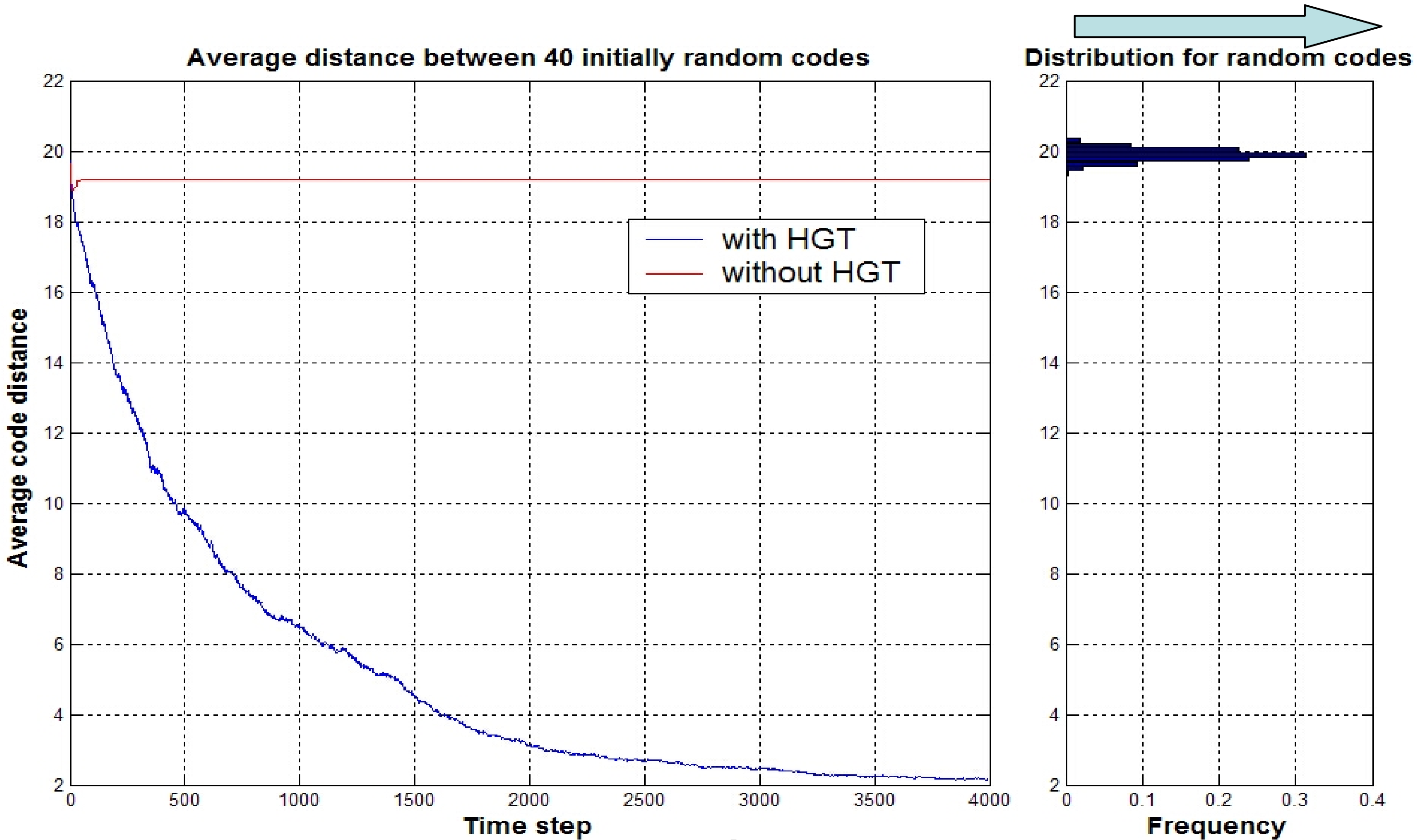


HGT leads to optimality

Evolution of code distances

Distribution of code distances

Average code distance



random codes

evolved code

Time

HGT leads to universality

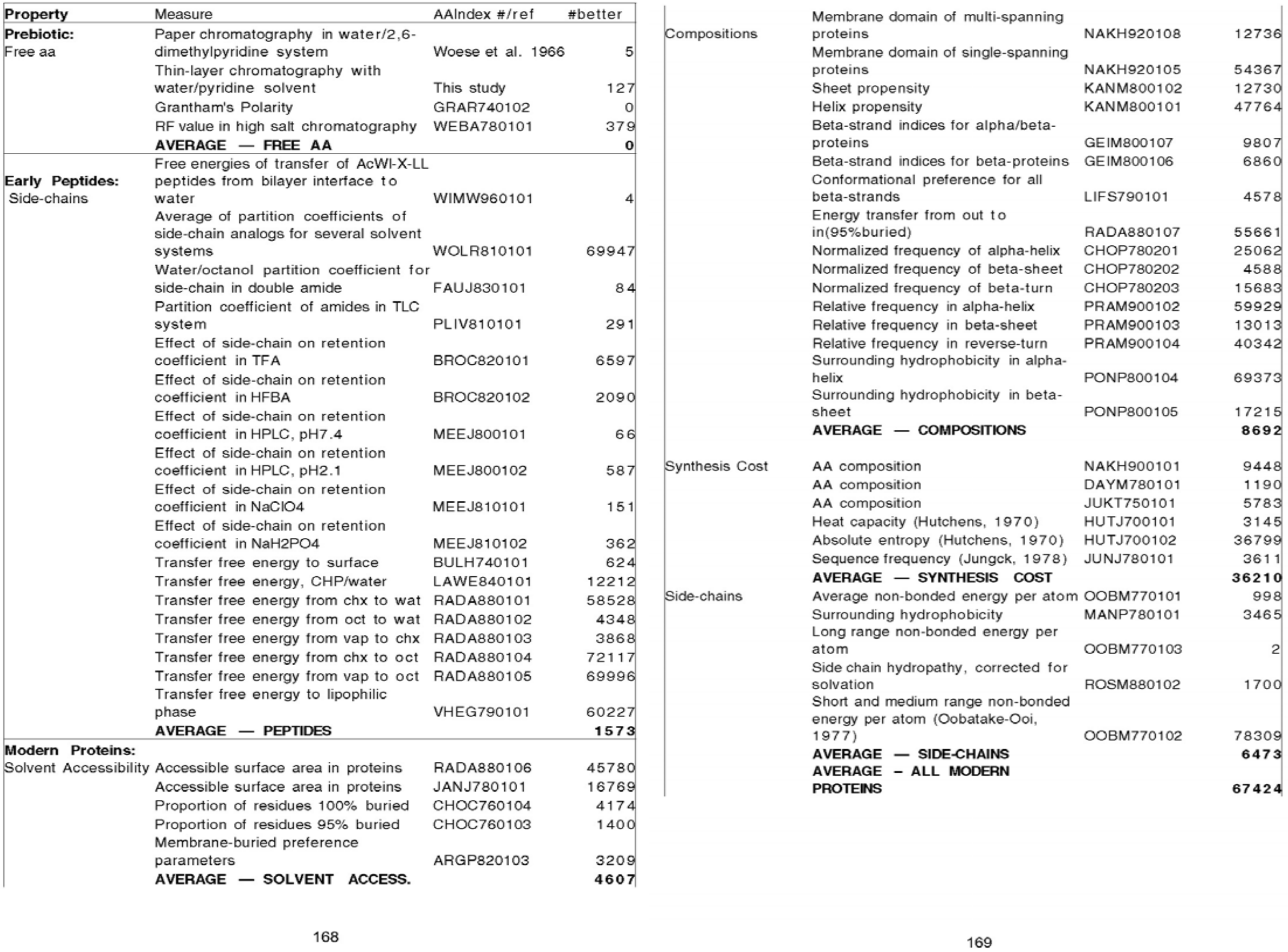
Implications for early life

Polar requirement sees further back in time than sequence phylogeny

Is the polar requirement special?

- Knight (2001) tested for optimality for a variety of other amino acid properties

- Hydropathy (side chain hydrophibicity) doesn’t work, for example. (Haig & Hurst 1991).



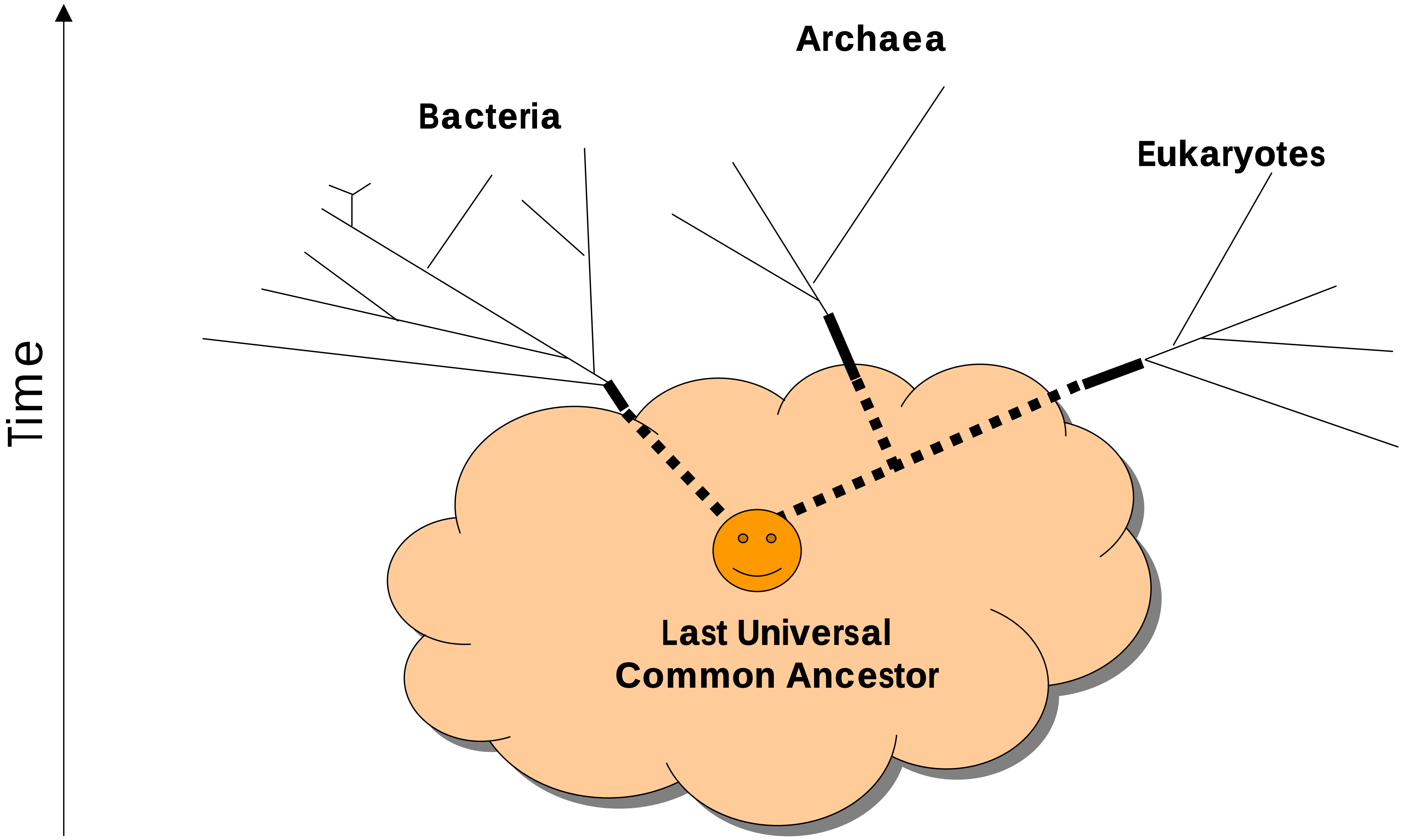
Property	Measure	AAIndex #/ref	#better
Prebiotic: Free aa	Paper chromatography in water/2,6-dimethylpyridine system	Woese et al. 1966	5
	Thin-layer chromatography with water/pyridine solvent	This study	127
	Grantham's Polarity	GRAR740102	0
	RF value in high salt chromatography	WEBA780101	379
	AVERAGE — FREE AA		**0**
Early Peptides: Side-chains	Free energies of transfer of AcWl-X-LL peptides from bilayer interface to water	WIMW960101	4
	Average of partition coefficients of side-chain analogs for several solvent systems	WOLR810101	69947
	Water/octanol partition coefficient for side-chain in double amide	FAUJ830101	84
	Partition coefficient of amides in TLC system	PLIV810101	291
	Effect of side-chain on retention coefficient in TFA	BROC820101	6597
	Effect of side-chain on retention coefficient in HFBA	BROC820102	2090
	Effect of side-chain on retention coefficient in HPLC, pH7.4	MEEJ800101	66
	Effect of side-chain on retention coefficient in HPLC, pH2.1	MEEJ800102	587
	Effect of side-chain on retention coefficient in NaClO4	MEEJ810101	151
	Effect of side-chain on retention coefficient in NaH2PO4	MEEJ810102	362
	Transfer free energy to surface	BULH740101	624
	Transfer free energy, CHP/water	LAWE840101	12212
	Transfer free energy from chx to wat	RADA880101	58528
	Transfer free energy from oct to wat	RADA880102	4348
	Transfer free energy from vap to chx	RADA880103	3868
	Transfer free energy from chx to oct	RADA880104	72117
	Transfer free energy from vap to oct	RADA880105	69996
	Transfer free energy to lipophilic phase	VHEG790101	60227
	AVERAGE — PEPTIDES		**1573**
Modern Proteins: Solvent Accessibility	Accessible surface area in proteins	RADA880106	45780
	Accessible surface area in proteins	JANJ780101	16769
	Proportion of residues 100% buried	CHOC760104	4174
	Proportion of residues 95% buried	CHOC760103	1400
	Membrane-buried preference parameters	ARGP820103	3209
	AVERAGE — SOLVENT ACCESS.		**4607**
Compositions	Membrane domain of multi-spanning proteins	NAKH920108	12736
	Membrane domain of single-spanning proteins	NAKH920105	54367
	Sheet propensity	KANM800102	12730
	Helix propensity	KANM800101	47764
	Beta-strand indices for alpha/beta-proteins	GEIM800107	9807
	Beta-strand indices for beta-proteins	GEIM800106	6860
	Conformational preference for all beta-strands	LIFS790101	4578
	Energy transfer from out to in(95%buried)	RADA880107	55661
	Normalized frequency of alpha-helix	CHOP780201	25062
	Normalized frequency of beta-sheet	CHOP780202	4588
	Normalized frequency of beta-turn	CHOP780203	15683
	Relative frequency in alpha-helix	PRAM900102	59929
	Relative frequency in beta-sheet	PRAM900103	13013
	Relative frequency in reverse-turn	PRAM900104	40342
	Surrounding hydrophobicity in alpha-helix	PONP800104	69373
	Surrounding hydrophobicity in beta-sheet	PONP800105	17215
	AVERAGE — COMPOSITIONS		**8692**
Synthesis Cost	AA composition	NAKH900101	9448
	AA composition	DAYM780101	1190
	AA composition	JUKT750101	5783
	Heat capacity (Hutchens, 1970)	HUTJ700101	3145
	Absolute entropy (Hutchens, 1970)	HUTJ700102	36799
	Sequence frequency (Jungck, 1978)	JUNJ780101	3611
	AVERAGE — SYNTHESIS COST		**36210**
Side-chains	Average non-bonded energy per atom	OOBM770101	998
	Surrounding hydrophobicity	MANP780101	3465
	Long range non-bonded energy per atom	OOBM770103	2
	Side chain hydropathy, corrected for solvation	ROSM880102	1700
	Short and medium range non-bonded energy per atom (Oobatake-Ooi, 1977)	OOBM770102	78309
	AVERAGE — SIDE-CHAINS		**6473**
	AVERAGE – ALL MODERN PROTEINS		**67424**

168

169

Implications for early life

Generic phase diagram for evolutionary dynamics



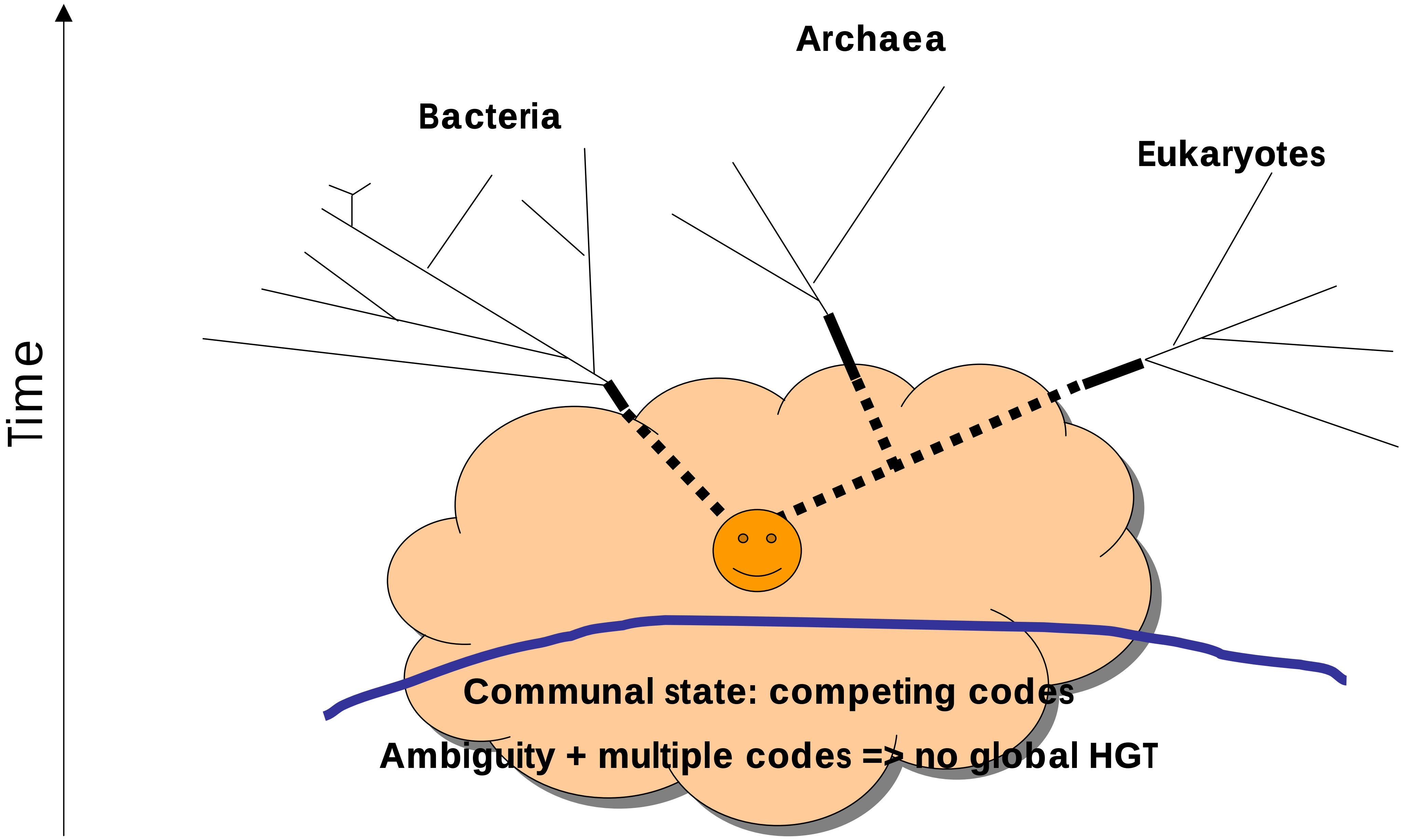
Time

Bacteria

Archaea

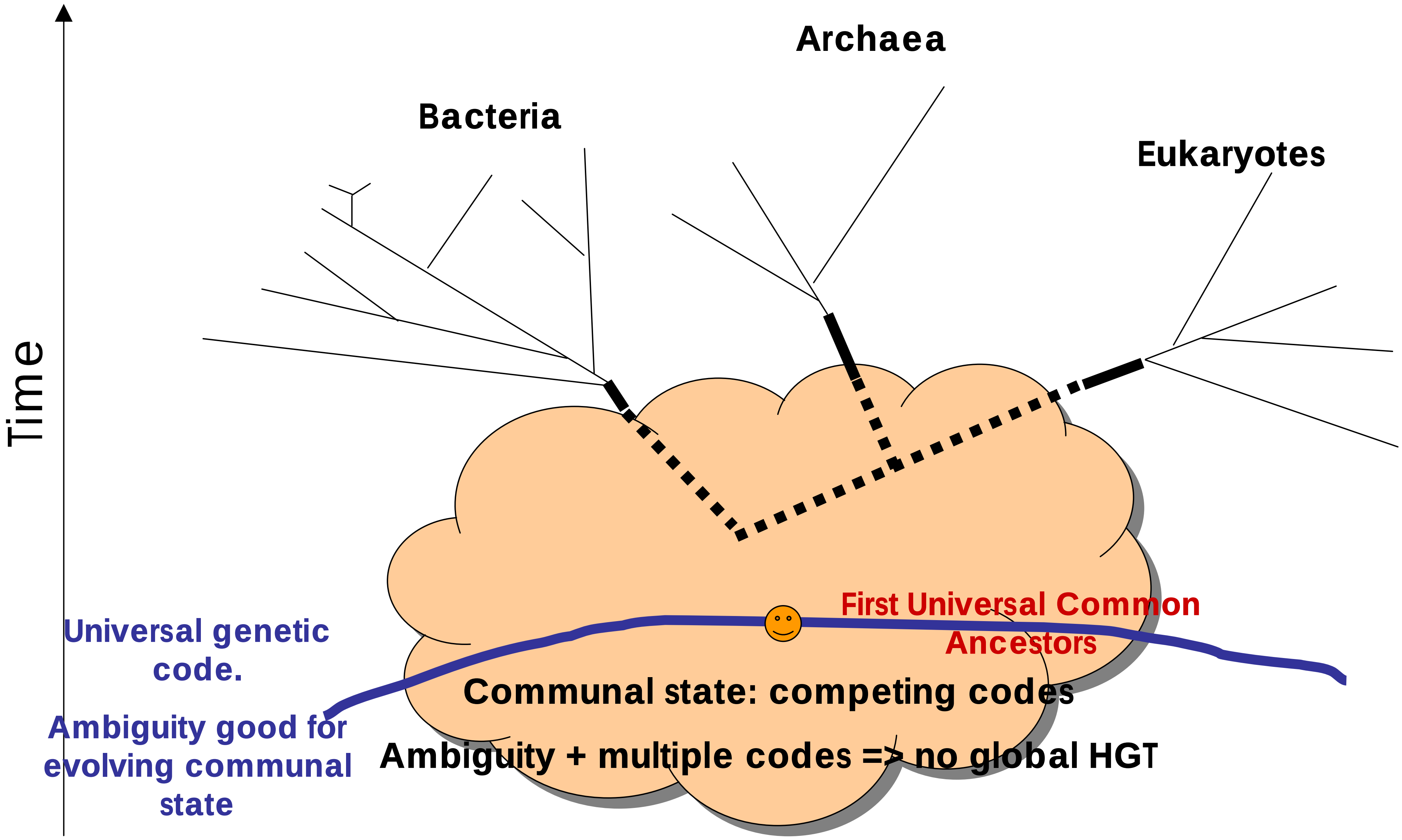
Eukaryotes

Last Universal
Common Ancestor



Communal state: competing codes

Ambiguity + multiple codes => no global HGT



Time

Archaea

Bacteria

Eukaryotes

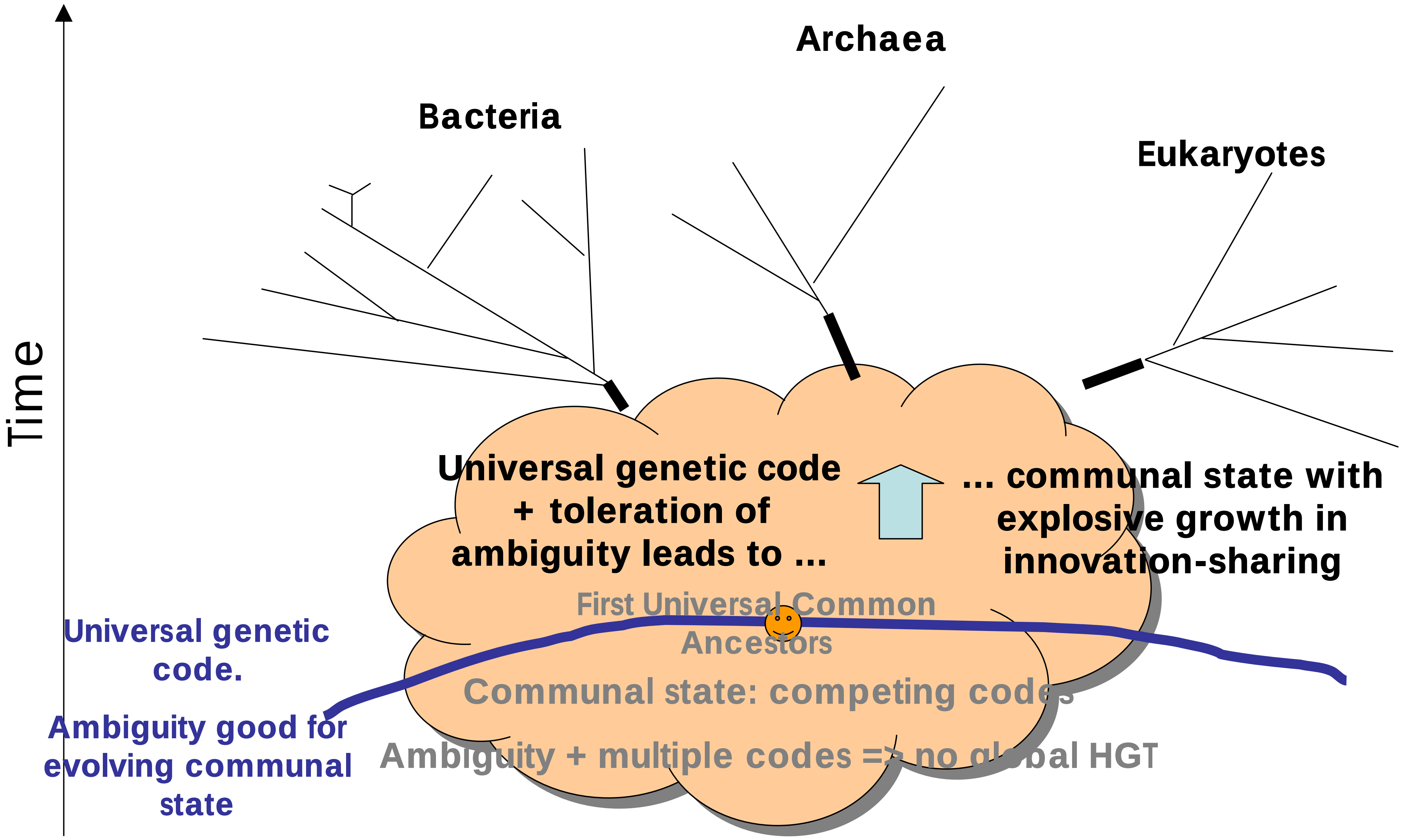
First Universal Common Ancestors

Universal genetic code.

Communal state: competing codes

Ambiguity good for evolving communal state

Ambiguity + multiple codes => no global HGT



Universal genetic code + toleration of ambiguity leads to ...

... communal state with explosive growth in innovation-sharing

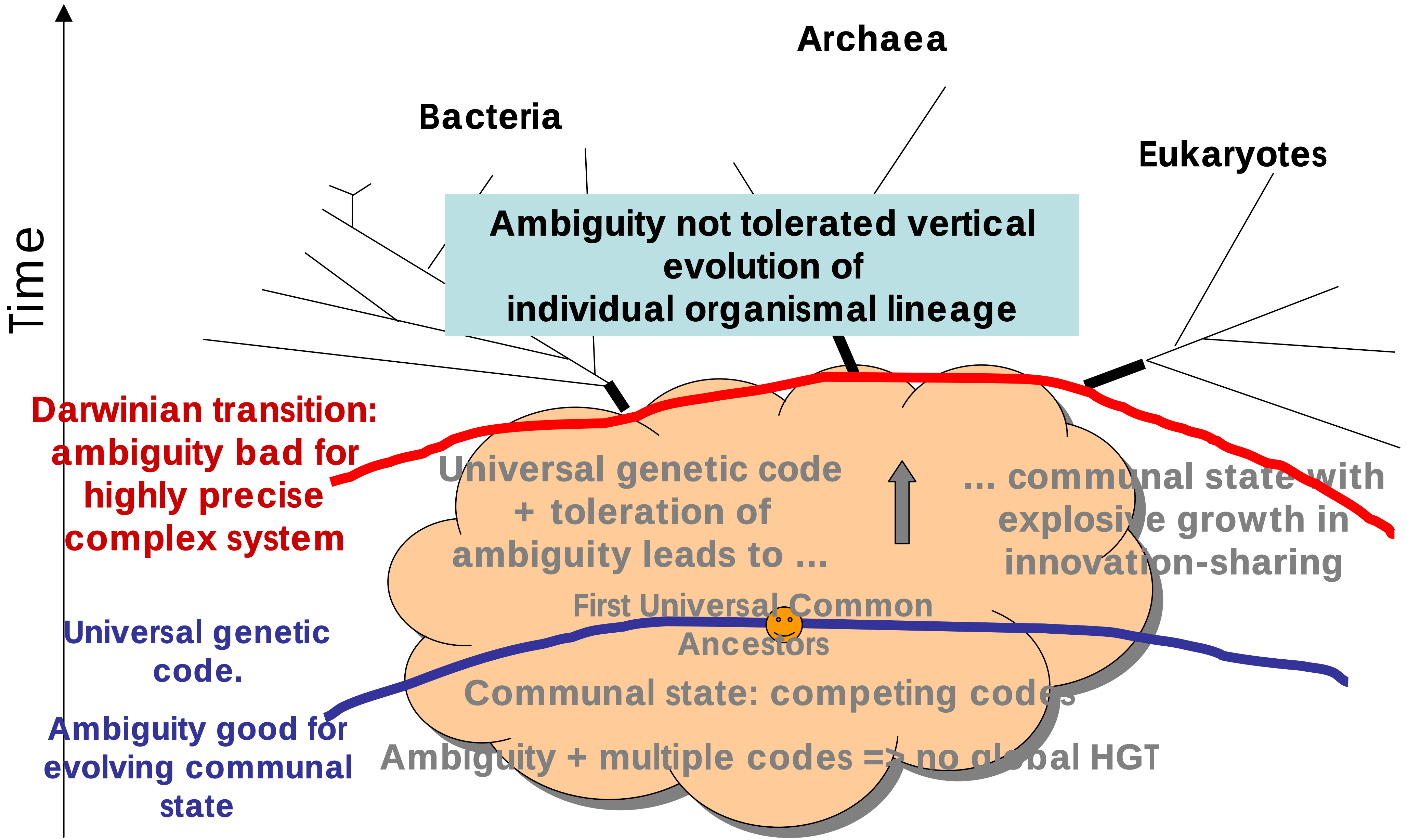
First Universal Common Ancestors

Communal state: competing codes

Ambiguity + multiple codes => no global HGT

Universal genetic code.

Ambiguity good for evolving communal state



Ambiguity not tolerated vertical evolution of individual organismal lineage

Darwinian transition: ambiguity bad for highly precise complex system

Universal genetic code + toleration of ambiguity leads to ...

... communal state with explosive growth in innovation-sharing

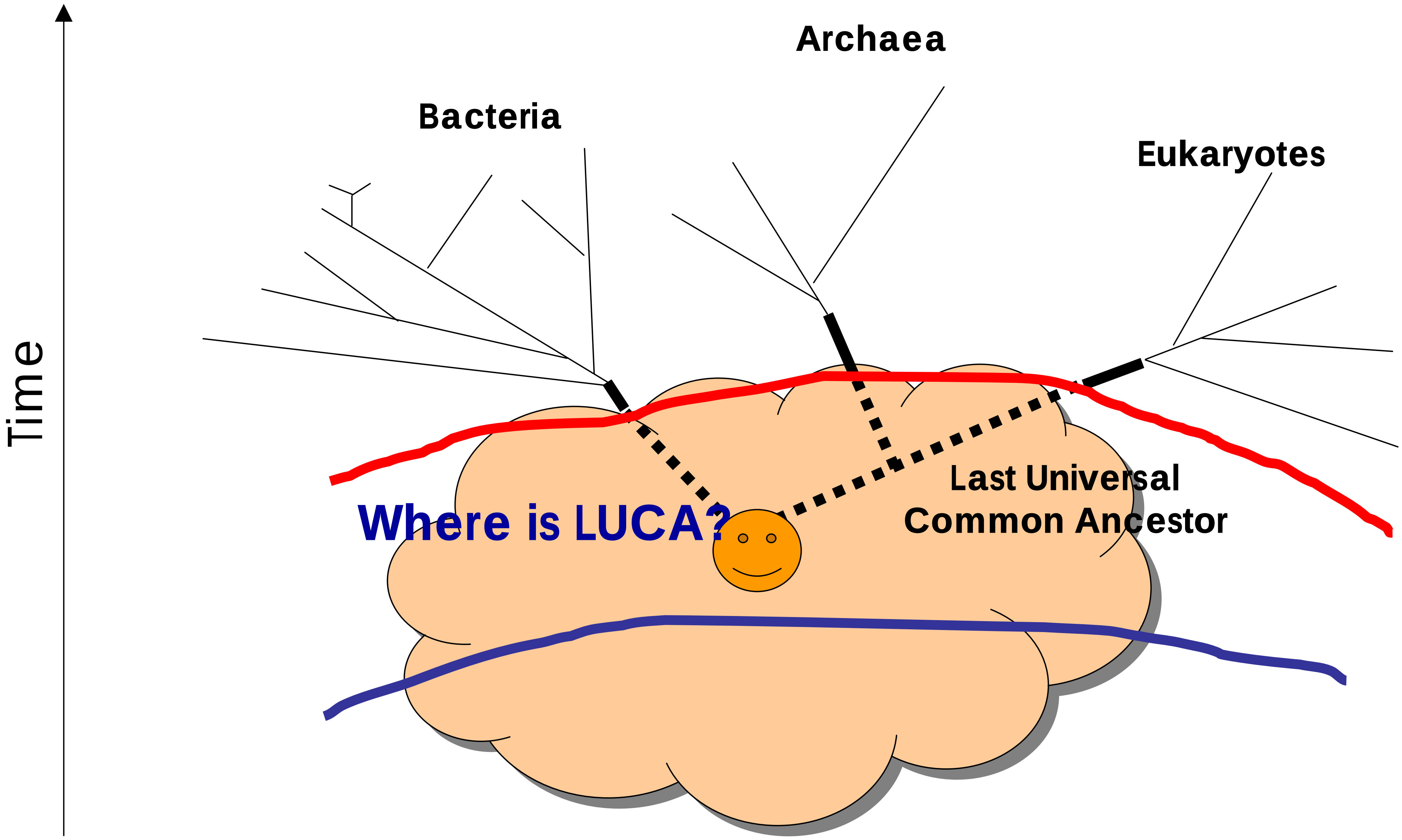
First Universal Common Ancestors

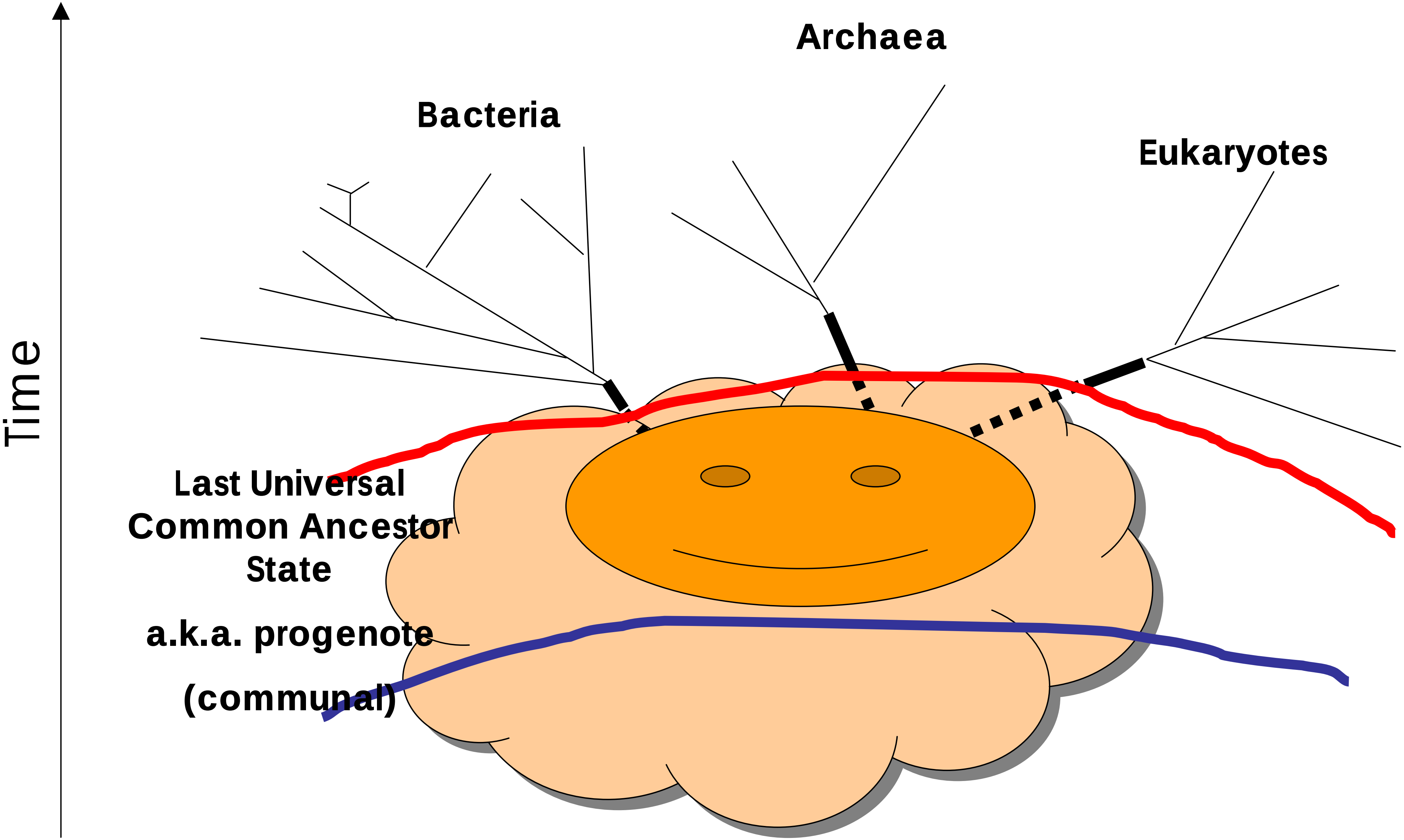
Universal genetic code.

Communal state: competing codes

Ambiguity good for evolving communal state

Ambiguity + multiple codes => no global HGT


Time
Archaea
Bacteria
Eukaryotes
Last Universal Common Ancestor
Where is LUCA?



Time

Bacteria

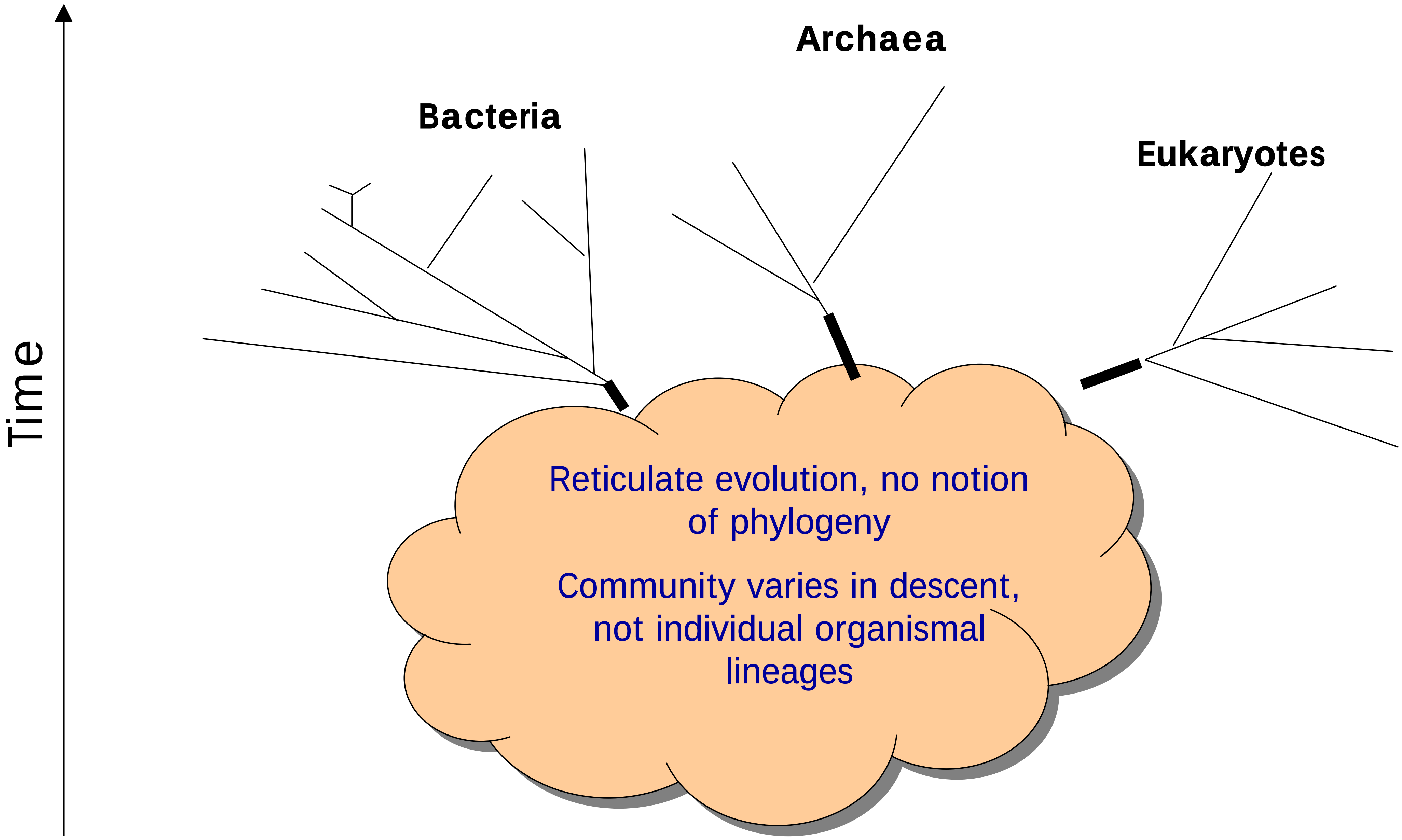
Archaea

Eukaryotes

Last Universal Common Ancestor State

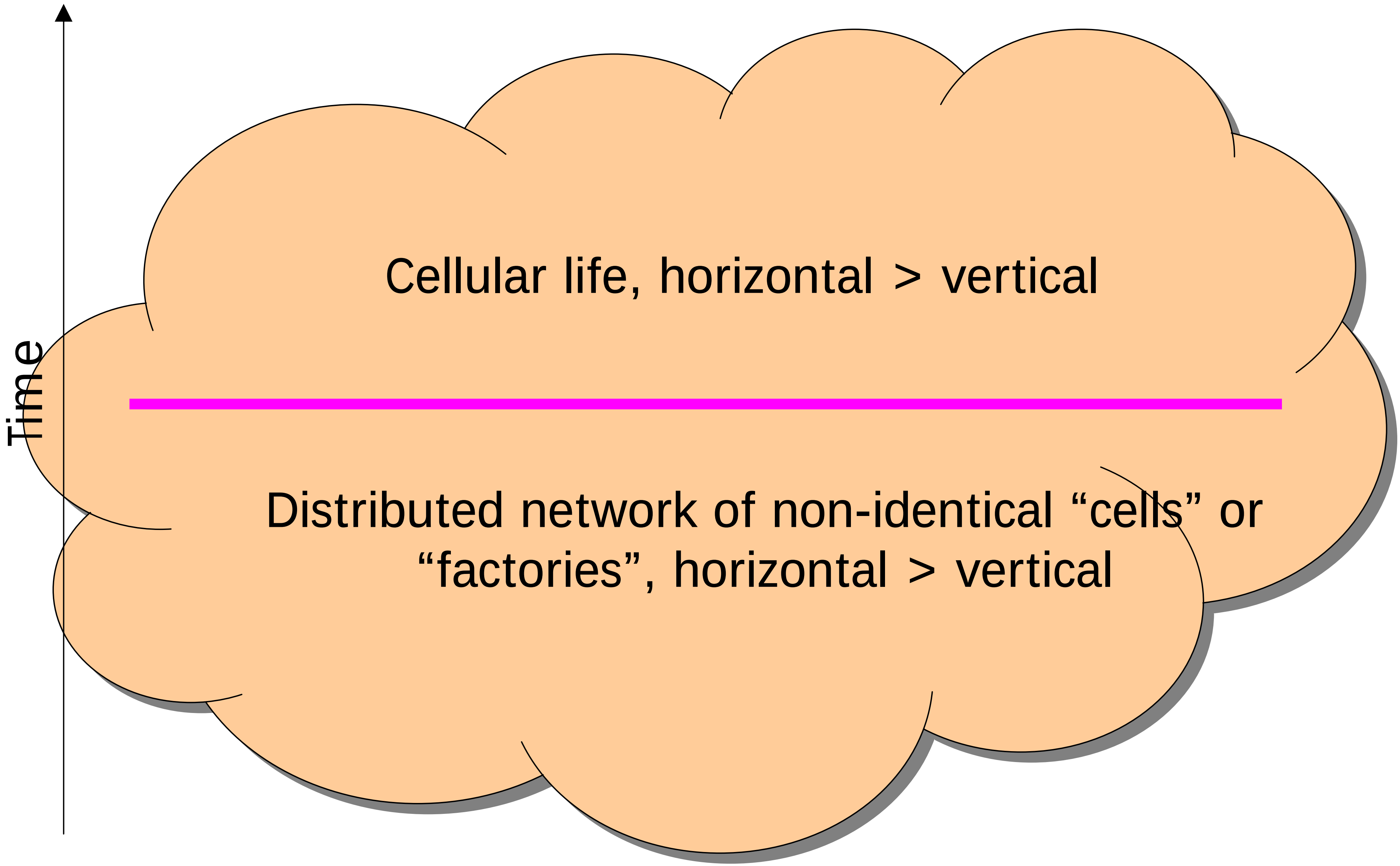
a.k.a. progenote

(communal)



Reticulate evolution, no notion of phylogeny

Community varies in descent, not individual organismal lineages



Cellular life, horizontal > vertical

Distributed network of non-identical “cells” or “factories”, horizontal > vertical

Conclusion of this study

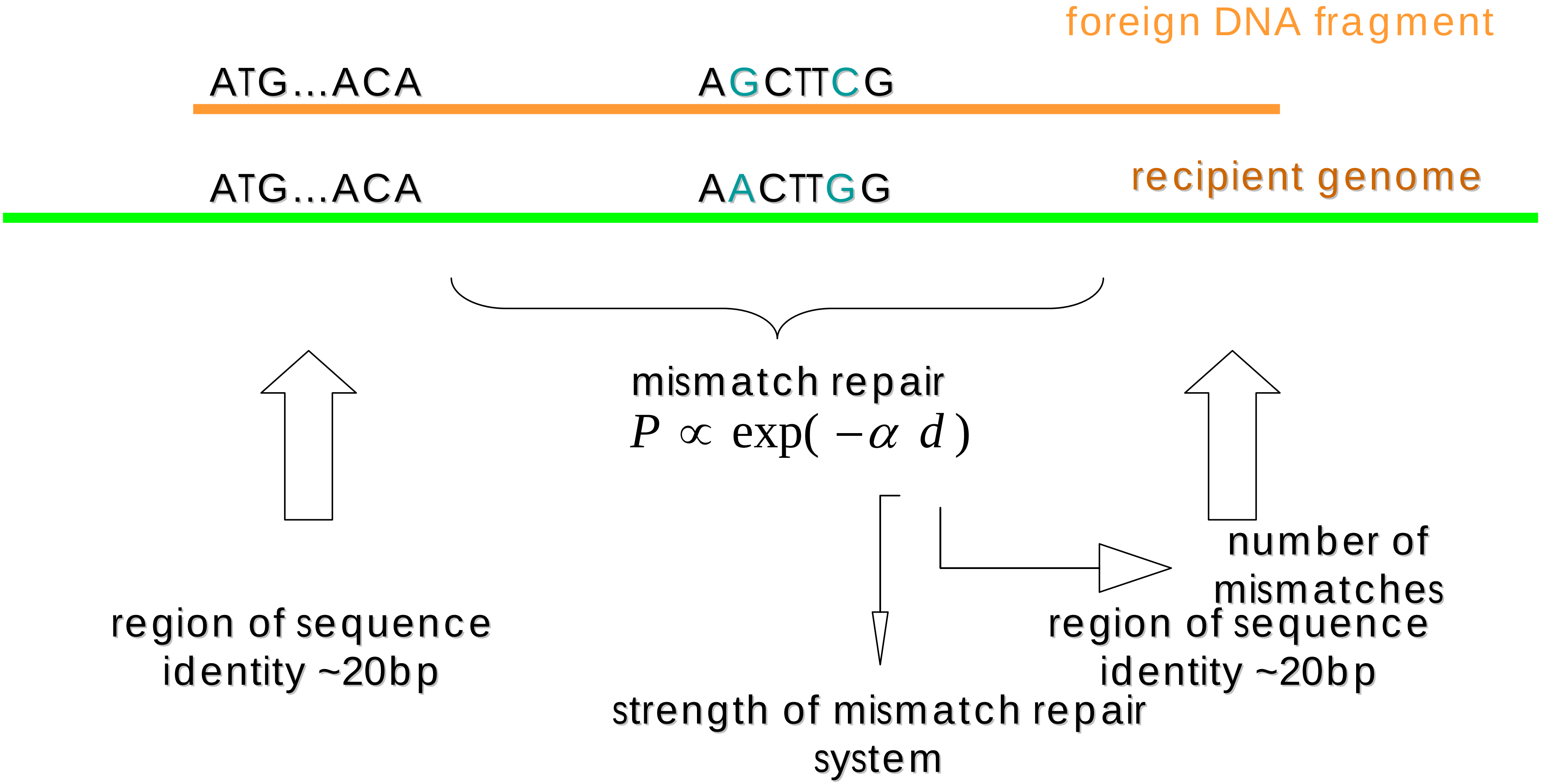
- Dynamics of genetic code coevolution with statistical proteins and rampant horizontal gene transfer can explain universality and optimality of the canonical genetic code
- Polar requirement is a crucial clue about prebiotic life
- Three phases of evolving life:
  - Weak communal evolution (no unified innovation-sharing protocol)
  - Strong communal evolution (unified innovation-sharing protocol)
  - Individual evolution, vertical descent (Darwinian)

Statistical mechanics of genes: impact on evolution

2. Microbial speciation

K. Vetsigian and N. Goldenfeld. Global divergence of microbial genome sequences mediated by propagating fronts. *Proc. Natl. Acad. Sci. (USA)*., 102, 7332–7337 (2005)

Homologous recombination



foreign DNA fragment

ATG…ACA AGCTTCG

ATG…ACA AACTTGG

recipient genome

mismatch repair

$$P \propto \exp(-\alpha\ d)$$

region of sequence identity ~20bp

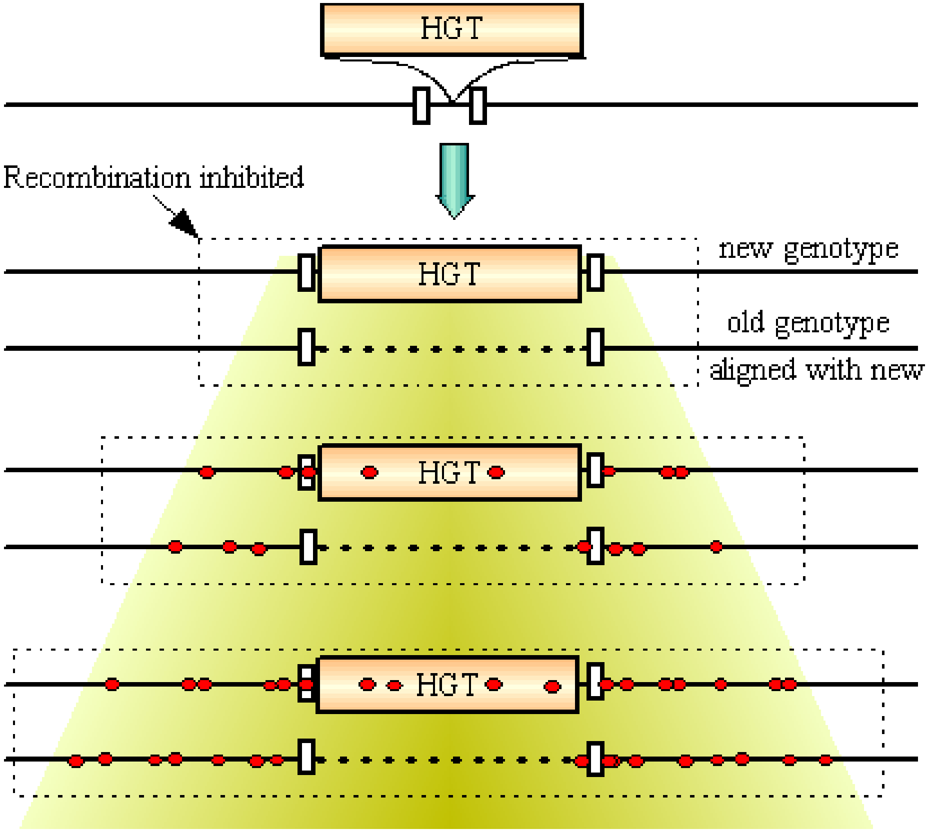
number of mismatches

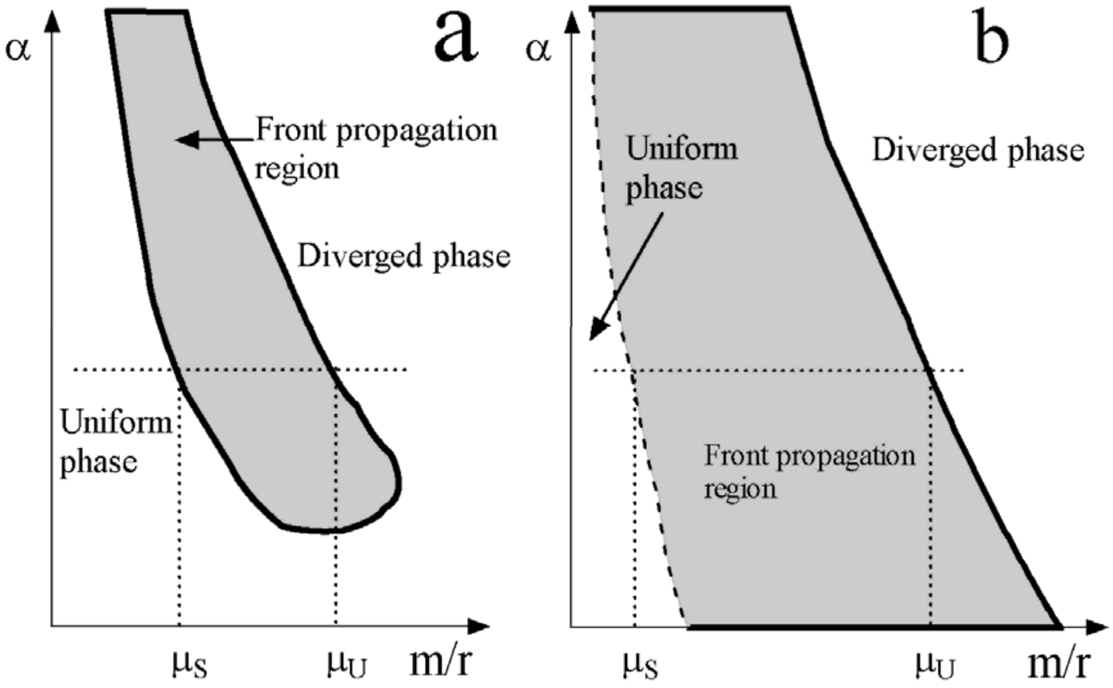
region of sequence identity ~20bp

strength of mismatch repair system

Front propagation in microbial genomes

Competition between mutation and recombination → phase transition





Speciation in absence of selection is possible!
Mechanism is observed in *Bacillus* …

Classify bacterial modes of evolution based on homologous recombination details

- **Bacteria requiring <u>one end</u> sequence identity**
  - Diversification fronts unlikely
  - Partial genetic isolation stable (no "species")
- **Bacteria requiring <u>two end</u> sequence identity**
  - Diversification fronts likely
  - Partial genetic isolation unstable – leads to global isolation (well defined "species")

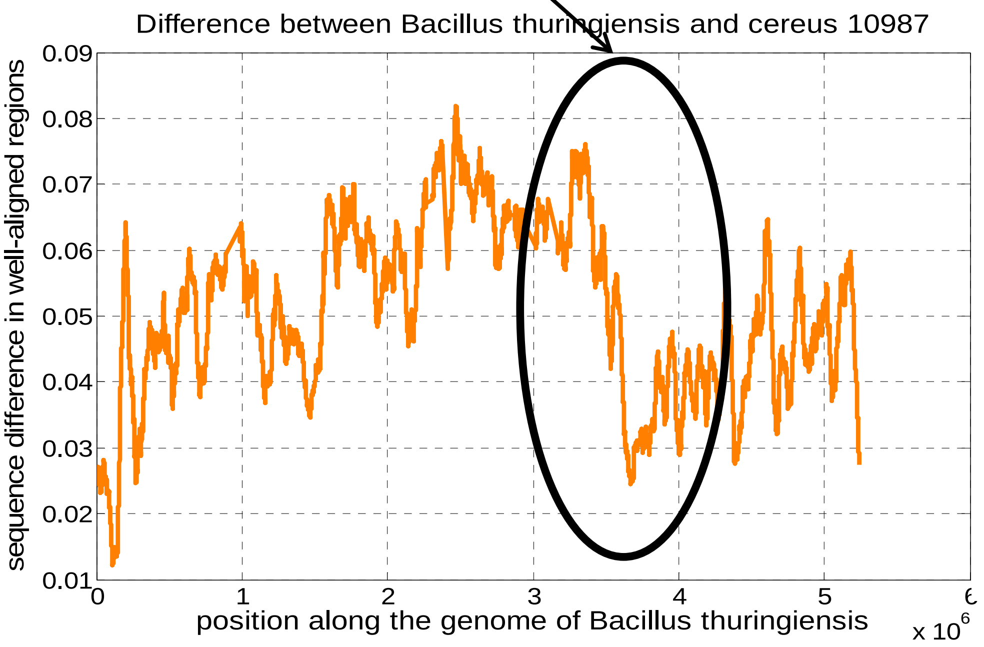
Refine classification by learning more about the details of homologous recombination and examining their relevance

Comparative genomic profile

Bacillus cereus group

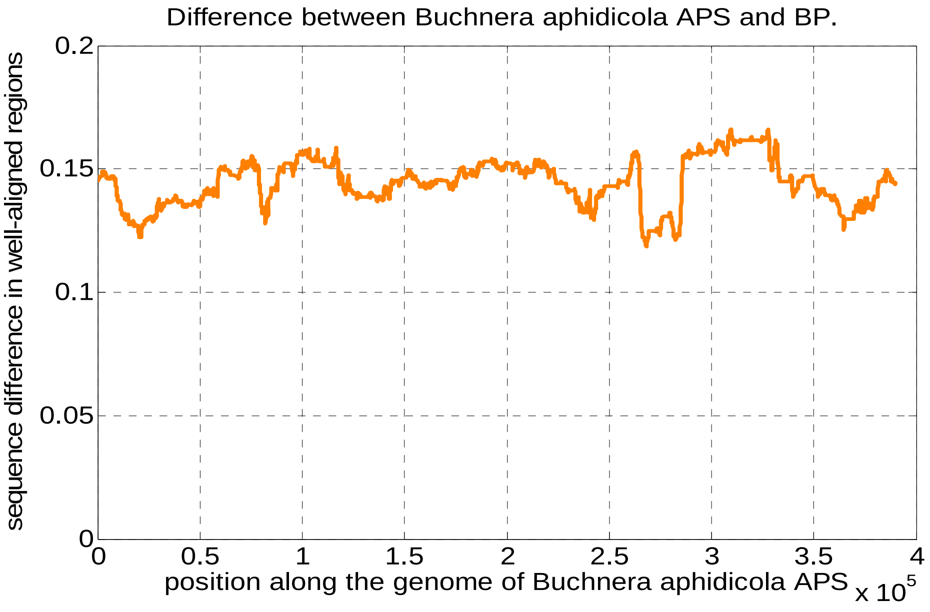
Front predicted

candidate for diversification front



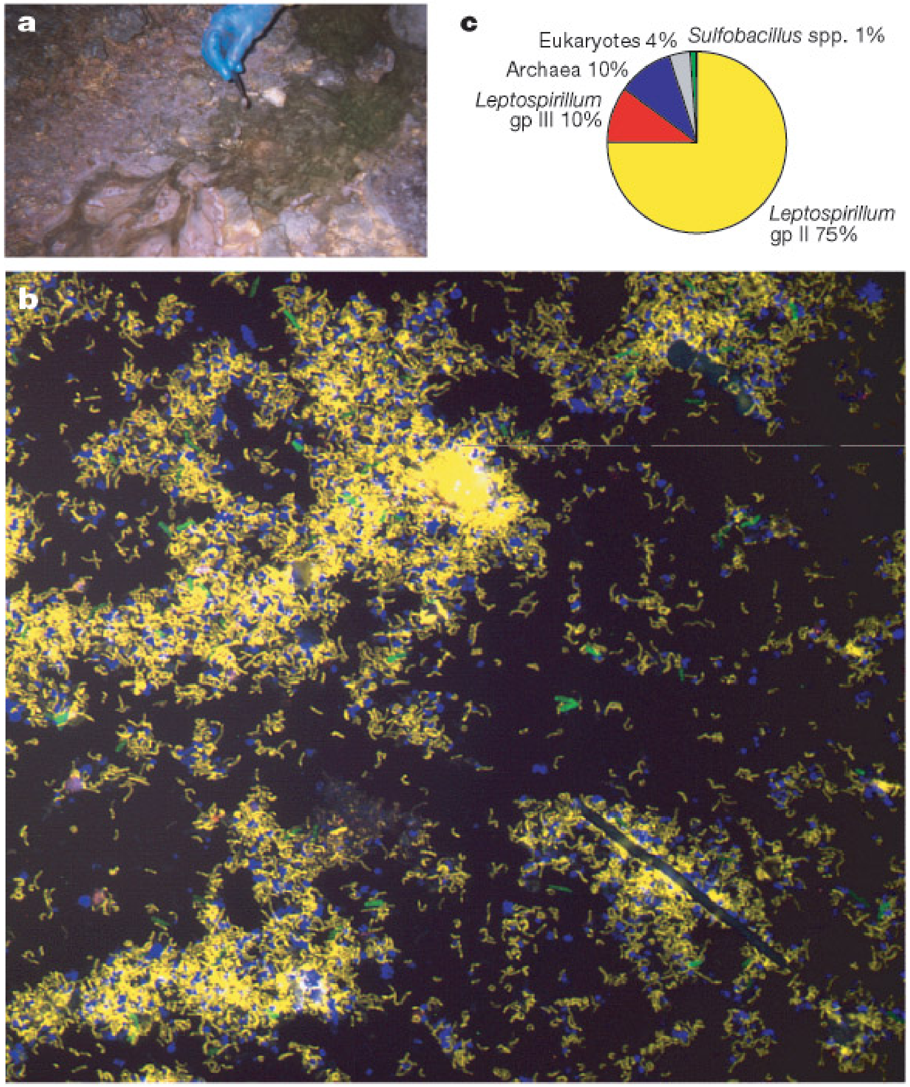
Buchnera aphidicola

Front not predicted

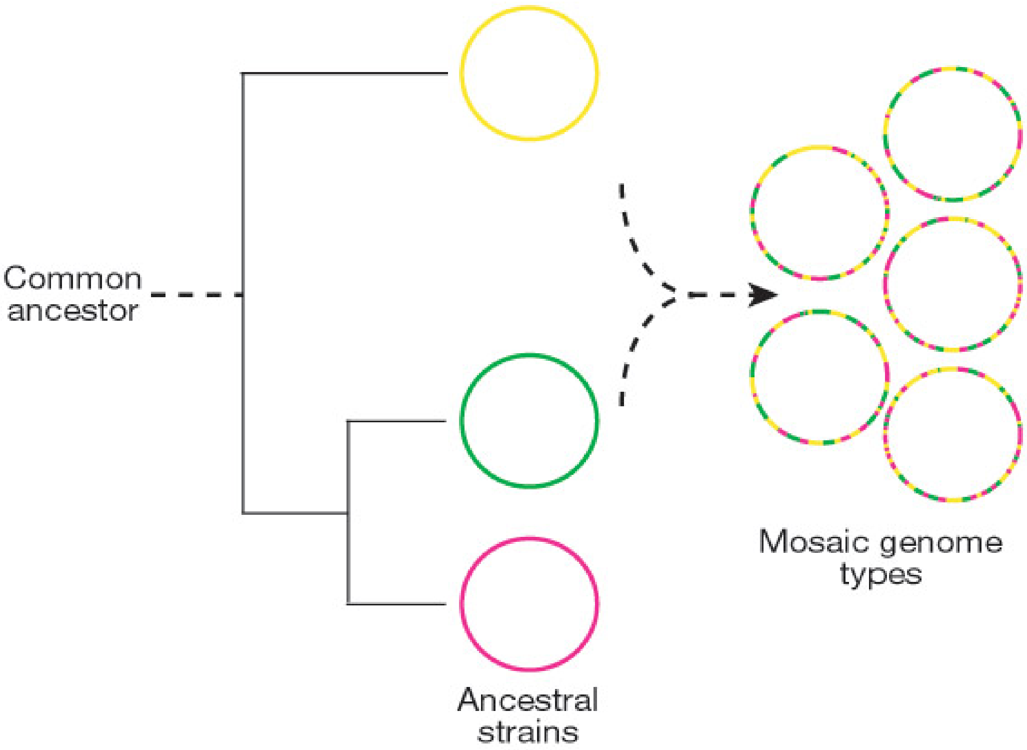


Recent study of extreme environment

- **Tyson et al (2004) examined acid mine drainage.**
  - Abandoned mine with FeS_2
  - Microbe interactions -> sulphuric acid (pH ~ 0.5)
  - Extreme environment -> low diversity
    - 2 abundant organisms (*Ferroplasma* and *Leptospirillum*)
    - Whole genome reconstruction achievable by shotgun sequencing



Mosaicism in microbial genomes



- Mosaic structure of genome suggests microbes swapped intact large quantities of genetic material, from three closely-related ancestral genomes

- Could it be the other way round? Could the mosaic structure and the “ancestral genomes” actually have descended from a homogeneous community?

  Spontaneous speciation?

- Ecology and evolution become intertwined

Conclusion from this study

- **Predict that horizontal transfer events can initiate or nucleate diversification fronts leading to speciation that propagate along microbial genomes over evolutionary time**
  - Whether this occurs depends upon the way in which alien DNA is incorporated into the chromosome
- **Fronts observed where theory predicts and not observed in microbes where theory predicts they should not arise**
- **Predict a mosaic structure of genomes that is observed and is otherwise puzzling**

Statistical mechanics of genes: impact on evolution

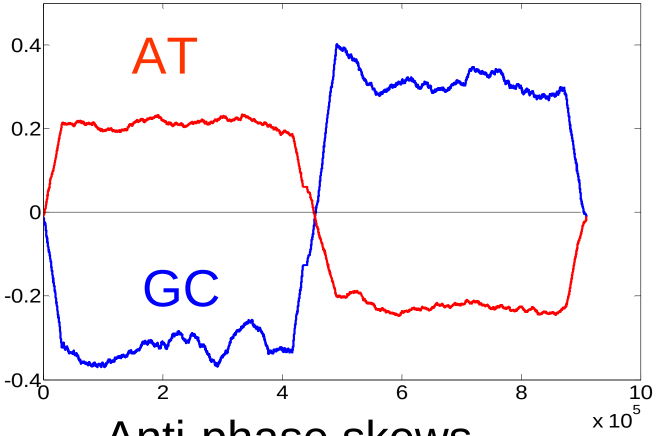
3. Origin of genome biases

K. Vetsigian and N. Goldenfeld. Genome rhetoric and the emergence of compositional bias. *Proc. Natl. Acad. Sci. (USA)* (2008)

The puzzle of genome bias

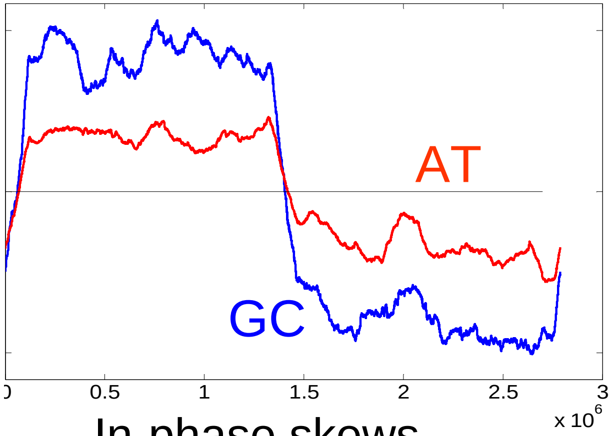
- GC content **(Ex.: *Streptomyces coelicolor*** 72 %, ***Arcobacter butzleri*** 27%)
- Codon usage
- AT skew **= (A-T) / (A+T),** GC skew **= (G-C) / (G+C)**

Borrelia burgdorferi

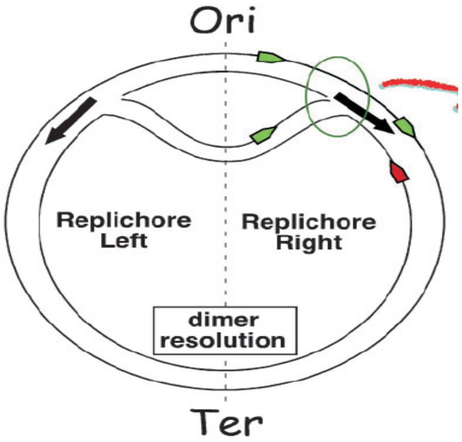


Anti-phase skews

Clostridium tetani



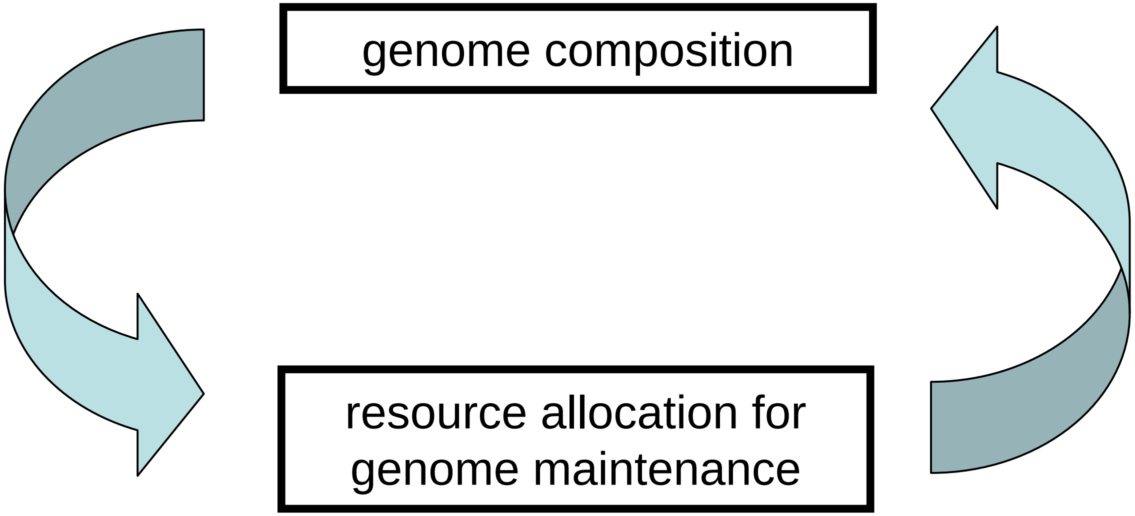
In-phase skews



  - All possible combinations of positive and negative skews are observed
  - Proposed biological explanations only lead to one sign of the skews

Emergent genome editing: Systems biology meets evolution

System level evolutionary dynamics of genome *maintenance* processes directs genome *evolution*

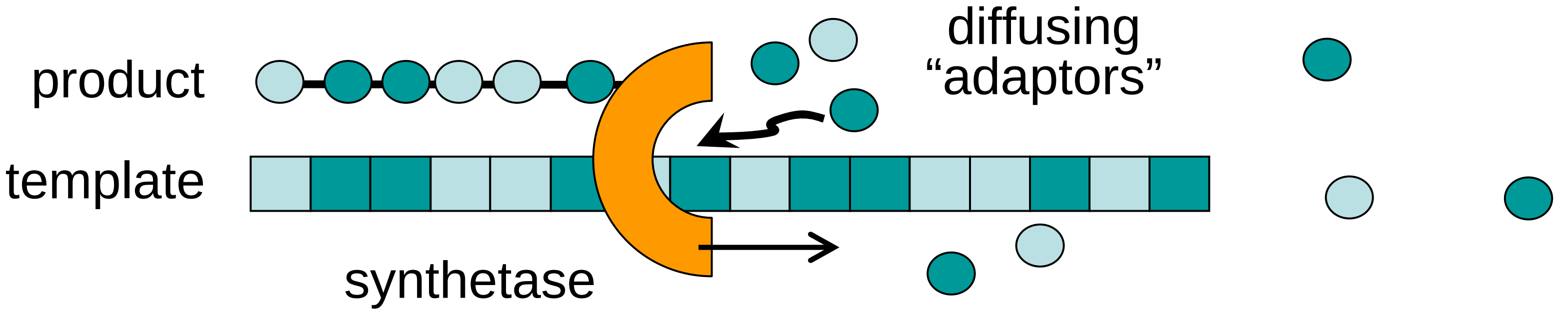


genome composition

resource allocation for genome maintenance

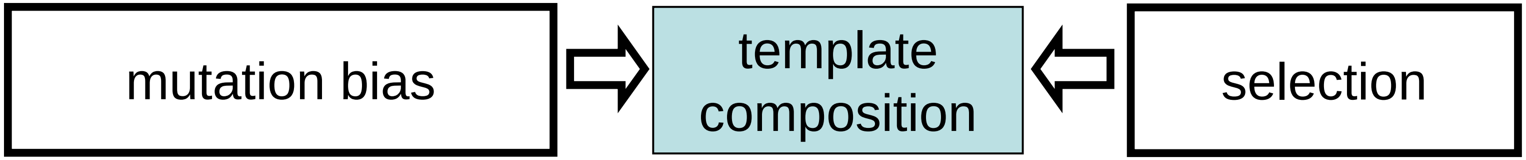
Template-directed synthesis:

the least common denominator of replication, transcription and translation



process	template letters	adaptors	synthetase	encoding rule
translation	codons	tRNAs	ribosome	Genetic code
transcription	A,G,C,T	NTPs	RNA polymerase	Watson-Crick pairing
replication	A,G,C,T	dNTPs	DNA polymerase	Watson-Crick pairing

Mutation-selection theory of template-directed synthesis

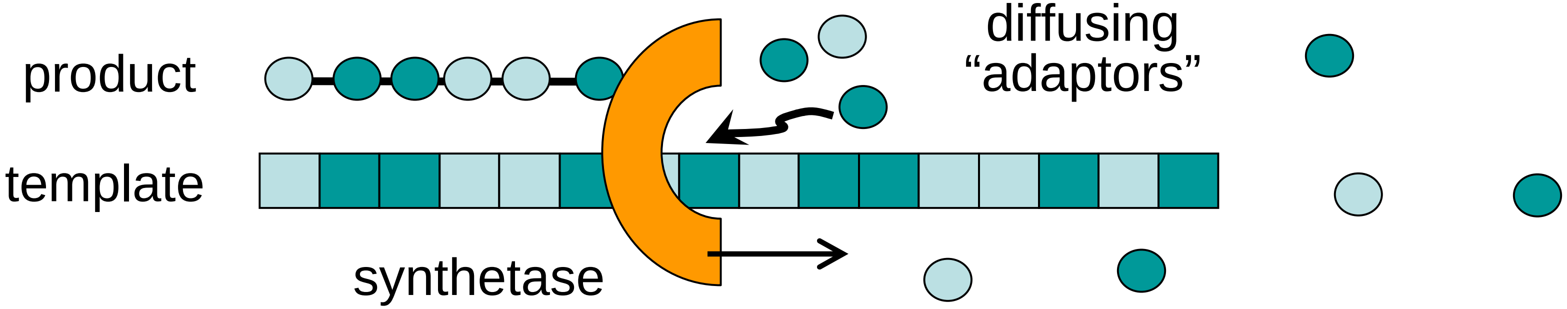


- **Example: Different tRNA species and abundances in different organisms *select* for different codon usage**
  - More abundant tRNAs select for their codons (to optimize translational efficiency)

- **But why are mutation and selection different in the first place (given the universal function + structure of information processing)?**

- **Such a model requires different *ad hoc* mechanisms for each bias**

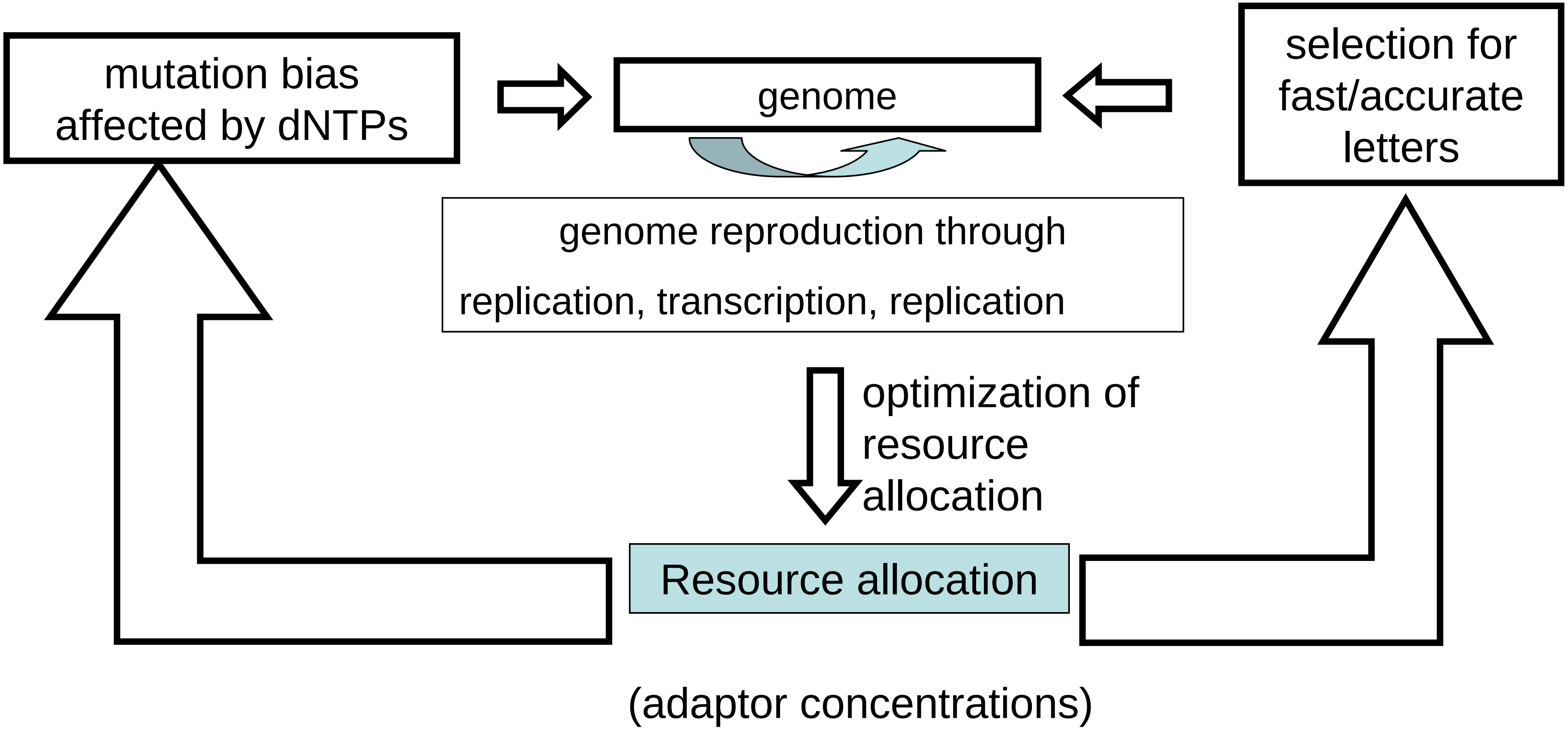
Speed and accuracy are affected by the adaptor concentrations



	at ■ :	at □ :
diffusion:	wait time ~ **1 / [●]**	wait time ~ **1 / [○]**
competition:	error rate ~ **[○] / [●]**	error rate ~ **[●] / [○]**

- **Speed and accuracy of synthesis – want higher abundance of common adaptors**
- **Fluctuations in concentrations irrelevant at high mutation rates, but at low mutation rates, instabilities cause symmetry breaking**

Combining mutation-selection framework with resource optimization



mutation bias
affected by dNTPs

genome

selection for
fast/accurate
letters

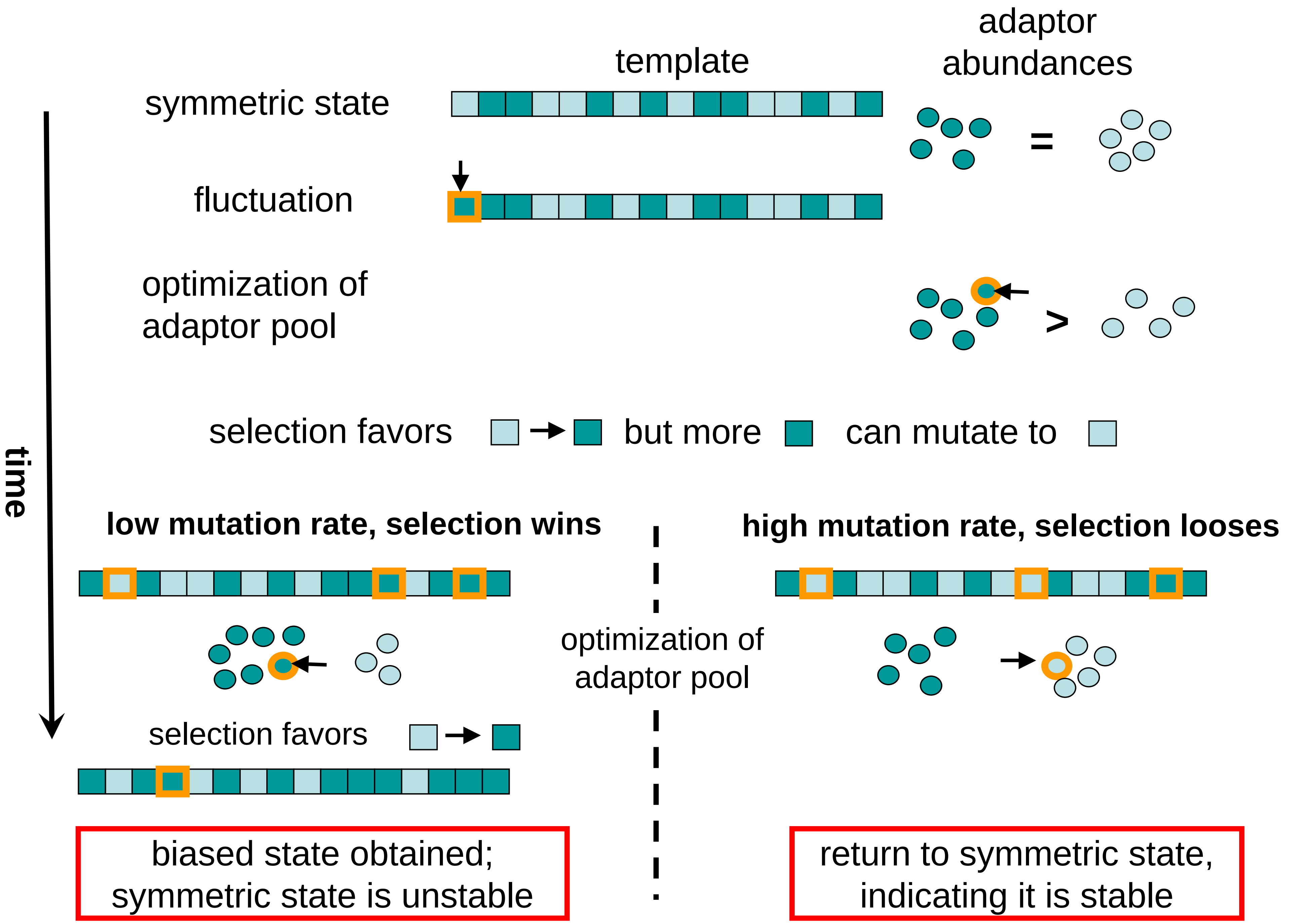
genome reproduction through
replication, transcription, replication

optimization of
resource
allocation

Resource allocation

(adaptor concentrations)

Feedback loops generate evolutionary instabilities



template

adaptor abundances

symmetric state

=

fluctuation

optimization of adaptor pool

\>

selection favors ⬜ → 🟦 but more 🟦 can mutate to ⬜

time

low mutation rate, selection wins

high mutation rate, selection looses

optimization of adaptor pool

selection favors ⬜ → 🟦

biased state obtained;
symmetric state is unstable

return to symmetric state,
indicating it is stable

Conclusion from this study

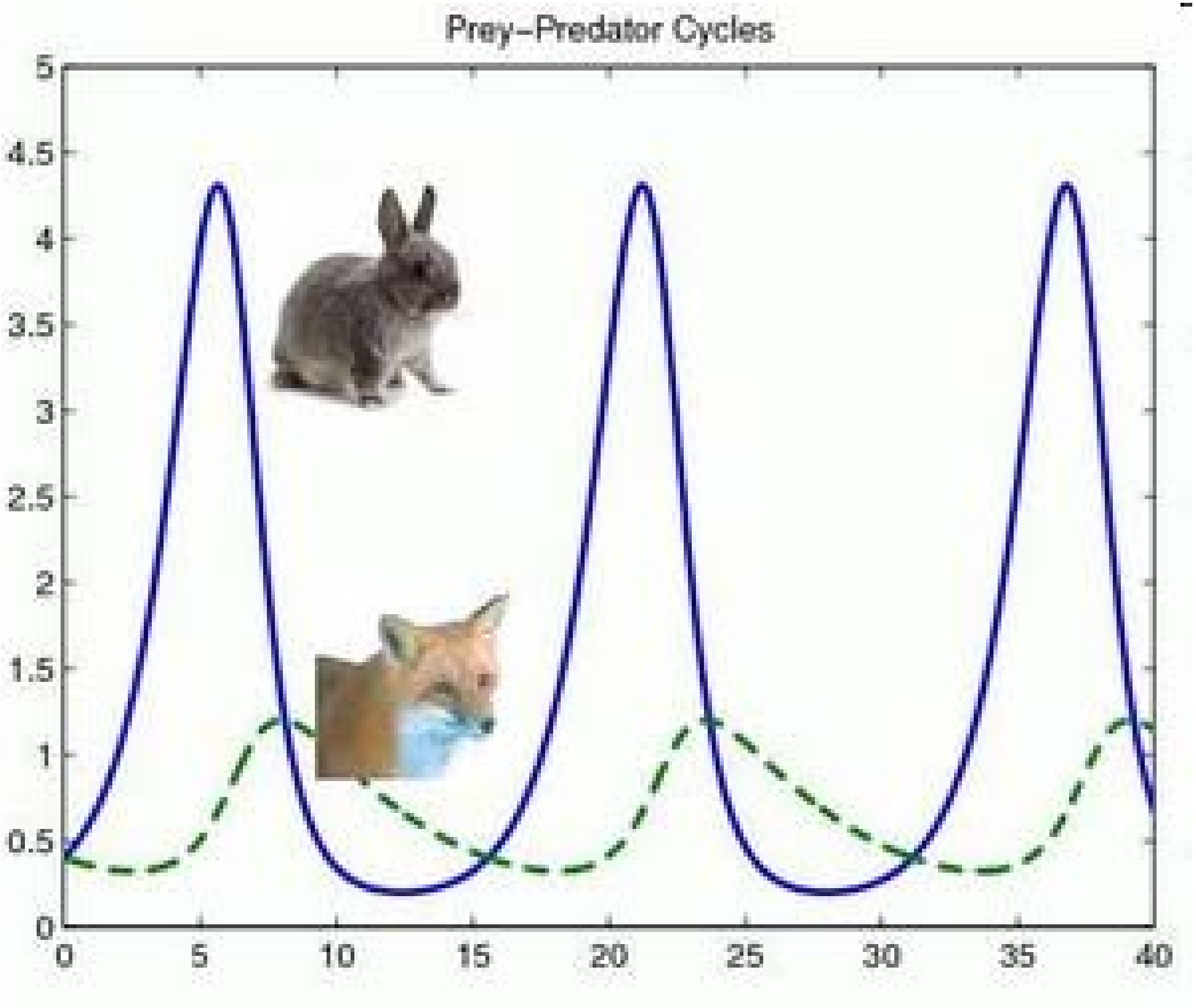
- **Theory predicts that genome composition should not be uniform**
- **Universal selection towards bias but not its direction**
- **Classify patterns of diversity based on the patterns of stable solutions**
- **Implications for evolution of cell machinery**

Statistical mechanics of ecology

1. Predator-Prey fluctuations

A.J. McKane and T. Newman. Predator-Prey Cycles from Resonant Amplification of Demographic Stochasticity. Phys. Rev. Lett. 94, 218102 (2005)

Predator-Prey Cycles



Predator-Prey Cycles

Predators

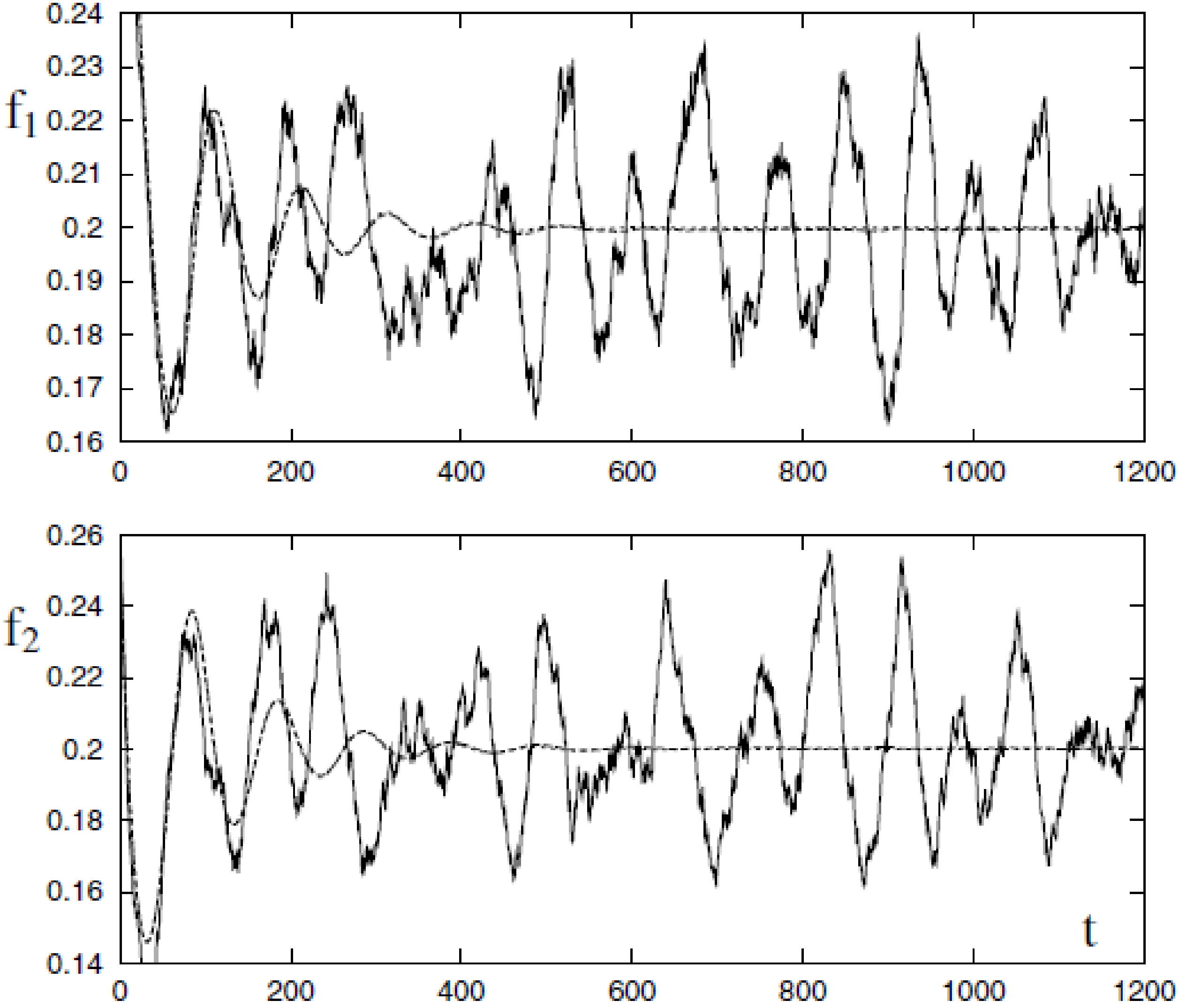
$$\frac{df_1}{dt} = n(f_2)f_1 - \mu f_1$$

Prey

$$\frac{df_2}{dt} = rf_2\left(1 - \frac{f_2}{K}\right) - g(f_2)f_1.$$

Problem: these equations do not have oscillatory solutions!

Predator-Prey Cycles in Simulation or Field



Solution to paradox

- **McKane and Newman showed that population dynamics equations used by ecologists are mean field equations**
- **Fluctuations ~ O(1/population size) are amplified and give rise to periodic oscillations.**
- **Can see this in small samples in the wild or in computer simulations**

Conclusion to this study

- **Simplistic modeling may not capture subtle fluctuation effects, spatial correlations**

- **Many fundamental phenomena overlooked**

- **Application: chemical reactions in the cell**
- **Application: ecology**

Outline

- **What physics can do for biology**
  - How biologists are actively seeking input from physicists
  - What problem areas are being ignored by physicists
  - Examples

- **What biology can do for physics**
  - Why biological problems are of interest to physicists
  - Is there "new physics" in biology?

- **Challenges and obstacles to the participation of physicists in biology**

Why are physicists going into biology?

- **Physics tools create new technologies for manipulation of biological systems at smallest scales**
  - Single-molecule biophysics
  - Physics Frontier Center for the Physics of Living Cells (UI) will explore single event processes in live cells!

- **New regimes for statistical mechanics, especially non-equilibrium**
  - N large but not infinite
  - Fluctuations dominant
  - How do biological machines function in such a noisy environment?

Are there “new physical laws”?

- **Schrodinger, Delbruck, … earliest physicists to work in biology**
  - Partial motivation was to find new physical laws!
- **Where would such laws lie?**
  - My answer: emergence and existence of life itself
  - How does matter self-organize hierarchically to create replicating, evolvable structures?
  - How do molecules come to life?

Outline

- **What physics can do for biology**
  - How biologists are actively seeking input from physicists
  - What problem areas are being ignored by physicists
  - Examples

- **What biology can do for physics**
  - Why biological problems are of interest to physicists
  - Is there “new physics” in biology?

- **Challenges and obstacles to the participation of physicists in biology**

Funding

- **Who will fund physicists doing biology?**
  - Problem: generic understanding not appreciated by agencies such as NIH, which are disease-focused
    - Solution: collaborate with biologists to answer focused questions. This is unsatisfactory for some, if the goal is basic understanding, not applied understanding
  - NSF-Biology, but biologists are the reviewers ...
  - NSF-DMR
    - Much research by condensed matter physicists is funded through DMR. Is this appropriate?
- **Needed: separate component of NSF-Biology with reviewer community of quantitative biologists**
  - Encouraging signs: many NSF solicitations now require collaborations between biologists and physicists/mathematicians

Publishing

- **Physicists addressing direct biological problems are being successful in publishing in high-impact journals**
  - E.g. observations of molecular motors, replication dynamics etc.

- **Theoretical work mostly in Phys. Rev. E**

- **High-impact journals not always able to assess work in new fields, such as biological complexity**

Is biological physics = physics?

- **Difficult for physics departments to assess faculty with careers in biological physics**

- **Training of students?**
  - Does a student working on modeling cell membrane really need to pass the same Qualifying Exam as a nuclear physicist?

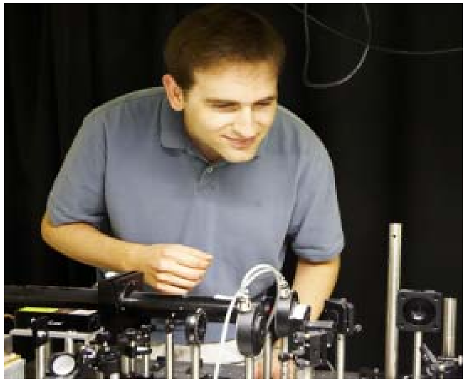
  - By working right away in biology, are physics students missing out on learning how to think like a physicist?

Opportunities

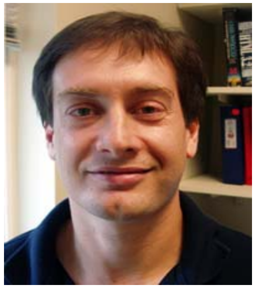
Training future biological physicists

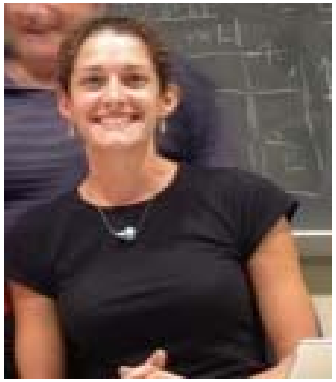
- Meet PFC current trainees (~30% women; Physics Dept. has 13% women)






- produced > 20 professors



Ivan Rasnik (Ha)
Emory U.
NSF CAREER award

Thorsten Ritz
(Schulten)
UC Irvine

Ahmet Yildiz
(Selvin)
MIT, Berkeley,..?

Rommie Amaro
(Luthey-Schulten)
UC Irvine

Opportunities for biological physics

- **The “new condensed matter physics” – rapid pace of discovery with genomics, single-molecule techniques, metagenomics (sampling of environmental DNA), understanding of translation, replication, … many new rich phenomena**

- **Fundamental intellectual challenges: origin of life. This is a physics question, not a biological or chemistry problem!**