# Did Networks and Community Dynamics Drive the Rapid Evolution of Early Life?

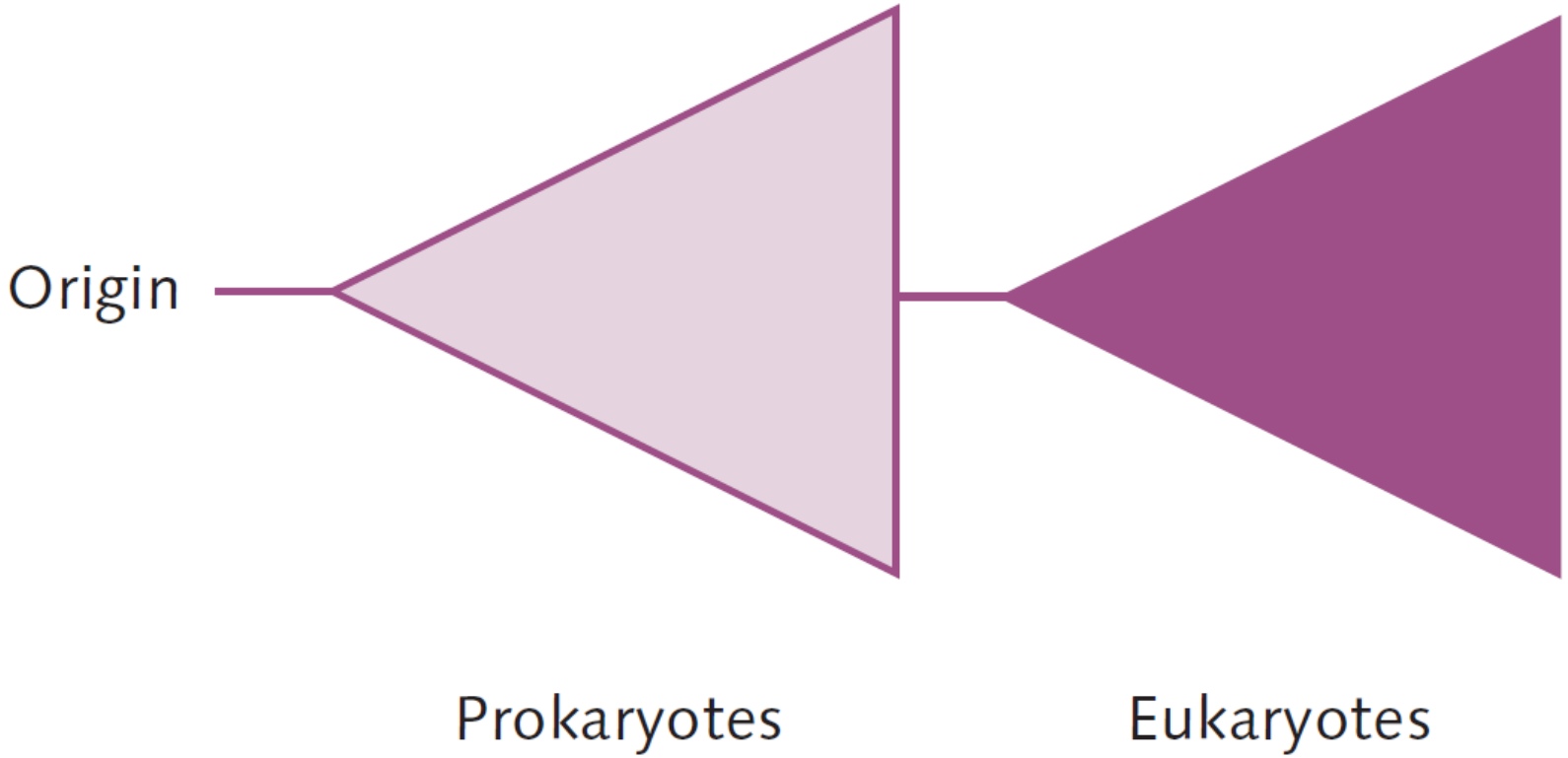
Nigel Goldenfeld

Department of Physics

Institute for Genomic Biology

University of Illinois at Urbana-Champaign

Once upon a time …

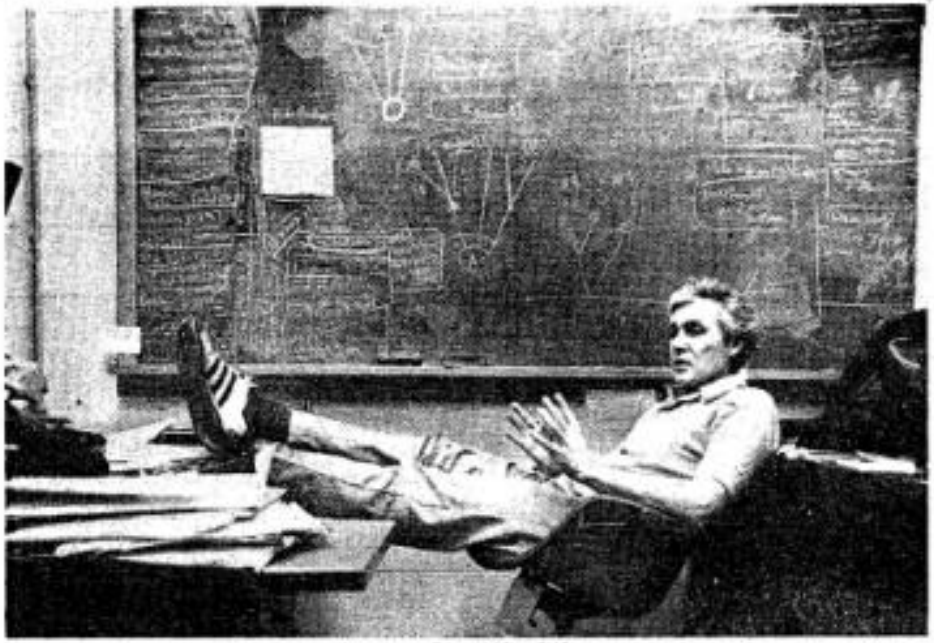



Pace 2009

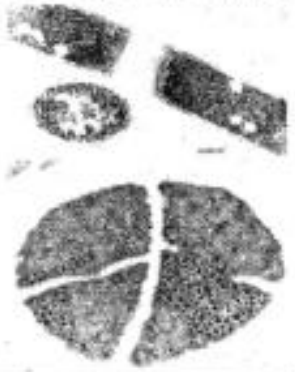
"All the News That's Fit to Print"

# The New York Times

VOL. CXXVII..No. 43,748 NEW YORK, THURSDAY, NOVEMBER 3, 1977



Dr. Carl R. Woese, leader of research team, in his office at the University of Illinois. Photo at right shows the newly discovered microorganism: top, a chain of two organisms, each one-thousandth-of-a-millimeter long; center, a cross section of the chain; bottom, an organism dividing into four cells.



## *Scientists Discover a Form of Life That Predates Higher Organisms*

By RICHARD D. LYONS

Special to The New York Times

URBANA, Ill., Nov. 2—Scientists studying the evolution of primitive organisms reported today the existence of a separate form of life that is hard to find in nature. They described it as a "third kingdom" of living material, composed of ancestral cells that abhor oxygen, digest carbon dioxide and produce methane.

The group investigating the evolution of microorganisms.

The genetic tracking efforts of the scientific group, which spanned five years, were made public today by two of the Federal agencies that supported the research, the National Aeronautics and Space Administration and the Nation-

## U. N. COUNCIL AGREES ON ARMS-BAN TERMS AGAINST SOUTH AFRICA

### Black Nations Accept Revised Draft —Embargo May Be Acted On as Early as Tomorrow

By KATHLEEN TELTSCH

Special to The New York Times

UNITED NATIONS, N.Y., Nov. 2—The 15 members of the Security Council agreed at a private meeting today on the terms of a revised proposal for a mandatory embargo on arms sales to South Africa to induce it to change its repressive racial practices.

Except for some minor details, the Council members agreed on the text of a resolution calling for such an embargo, and this was accepted later by the 49-nation African bloc of the United Nations, which had pressed unsuccessfully for economic sanctions.

The agreement today means that the resolution is likely to be approved without a vote as a statement of consensus by the Council, possibly as early as Friday.

**First Such Step Against a Member**

This would be the first time that the Council has imposed the punitive measure of sanctions against a United Nations member.

The new text goes part way toward meeting the position of the 49 African countries, which brought a complaint to the Council after South Africa's severe crackdown on Oct. 19 against black organizations and individuals and their supporters.

The draft resolution calls on all countries, including those not in the United Nations, to "cease forthwith" any provision to South Africa of arms, ammunition of all types, military vehicles and equipment and spare parts. In response to demands by African countries, the revised text included calls for a review

## *Vance Welcomes Offer by Soviet As 'Major Step'*

By BERNARD GWERTZMAN

Special to The New York Times

WASHINGTON, Nov. 2—Secretary of State Cyrus R. Vance said today that Leonid I. Brezhnev's proposal for a moratorium on all underground nuclear detonations — for peaceful uses as well as weapons tests—marked "a major step forward" toward a comprehensive test ban, but that differences persisted on the duration of such an accord.

At a news conference, Mr. Vance underscored Washington's satisfaction with the proposal. But reflecting the ambiguous state of Soviet-American relations, Mr. Vance also confirmed a report in The New York Times that the Administration had been urging the Russians not to proceed with pending trials of dissidents lest they could harm overall relations.

**'A Mixed Set of Factors'**

"Let me say that the relationships between ourselves and the Soviet Union are always a mixed set of factors," he said. "We have areas in which we may be making progress; there are other areas in which we may be standing still; and there are still other areas in which we may be retrogressing. And today is like any other time in that there are all of these different kinds of currents and crosscurrents flowing in our relationships."

On the whole, there has been an improvement in relationships over the last

Continued on Page A2, Column 1

## CARTER MAKES PLEA TO JEWS ON MIDEAST

### He Calls on Leaders to Support

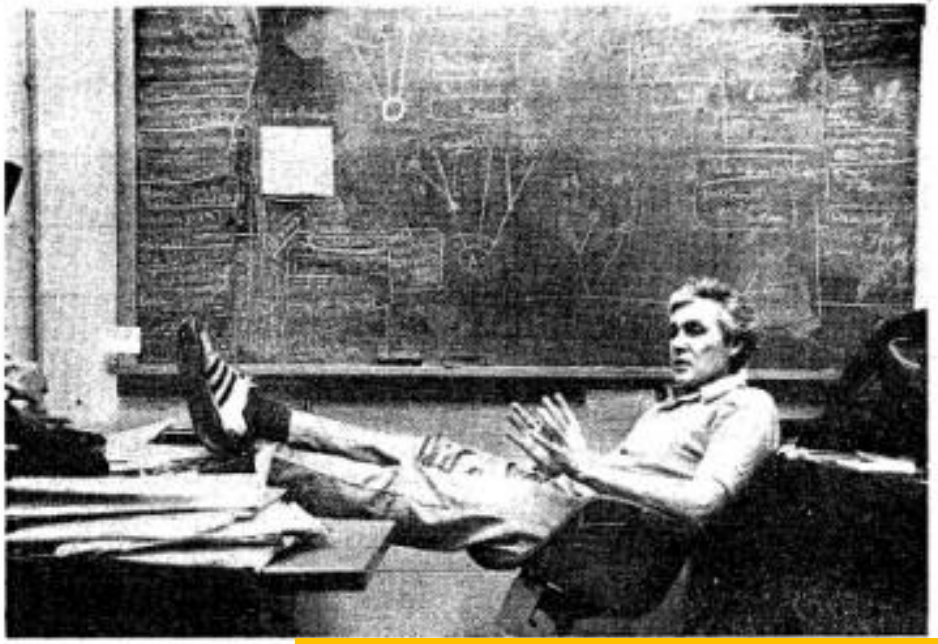
6

## CONVICTION IS UPHELD

"All the News That's Fit to Print"

# The New York Times

VOL. CXXVII..No. 43,748 NEW YORK, THURSDAY, NOVEMBER 3, 1977



Dr. Carl R. Woese, leader of research
Illinois. Photo at right shows the
chain of two organisms, each one-th
cross section of the chain; bottom,

## Carl R. Woese 1928-2012

## Scientists Discove That Predates

By RICHA

URBANA, Ill., Nov. 2—Scientists study
ing the evolution of primitive organisms
reported today the existence of a separate
form of life that is hard to find in nature
They described it as a "third kingdom"
of living material, composed of ancestral
cells that abhor oxygen, digest carbon
dioxide and produce methane.

years, were made public today by two
of the Federal agencies that supported
the research, the National Aeronautics
and Space Administration and the Nation-

## CONVICTION IS UPHELD

## U. N. COUNCIL AGREES ON ARMS-BAN TERMS AGAINST SOUTH AFRICA

### Black Nations Accept Revised Draft —Embargo May Be Acted On as Early as Tomorrow

By KATHLEEN TELTSCH

UNITED NATIONS, N.Y., Nov. 2—The 15 members of the Security Council agreed at a private meeting today on the terms of a revised proposal for a mandatory embargo on arms sales to South Africa to induce it to change its repressive racial practices.

Except for some minor details, the Council members agreed on the text of a resolution calling for such an embargo, and this was accepted later by the 49-nation African bloc of the United Nations, which had pressed unsuccessfully for economic sanctions.

tion of all types, military vehicles and equipment and spare parts. In response to demands by African countries, the revised text included calls for a review

## *Vance Welcomes Offer by Soviet As 'Major Step'*

By BERNARD GWERTZMAN

WASHINGTON, Nov. 2—Secretary of State Cyrus R. Vance said today that Leonid I. Brezhnev's proposal for a moratorium on all underground nuclear detonations — for peaceful uses as well as weapons tests—marked "a major step forward" toward a comprehensive test ban, but that differences persisted on the duration of such an accord.

At a news conference, Mr. Vance underscored Washington's satisfaction with the proposal. But reflecting the ambiguous state of Soviet-American relations, Mr. Vance also confirmed a report in The New York Times that the Administration had been urging the Russians not to proceed with pending trials of dissidents lest they could harm overall relations.

**'A Mixed Set of Factors'**

"Let me say that the relationships between ourselves and the Soviet Union are always a mixed set of factors," he said. "We have areas in which we may be making progress; there are other areas in which we may be standing still and there are still other areas in which we may be retrogressing. And today is like any other time in that there are all of these different kinds of currents and crosscurrents flowing in our relationships."

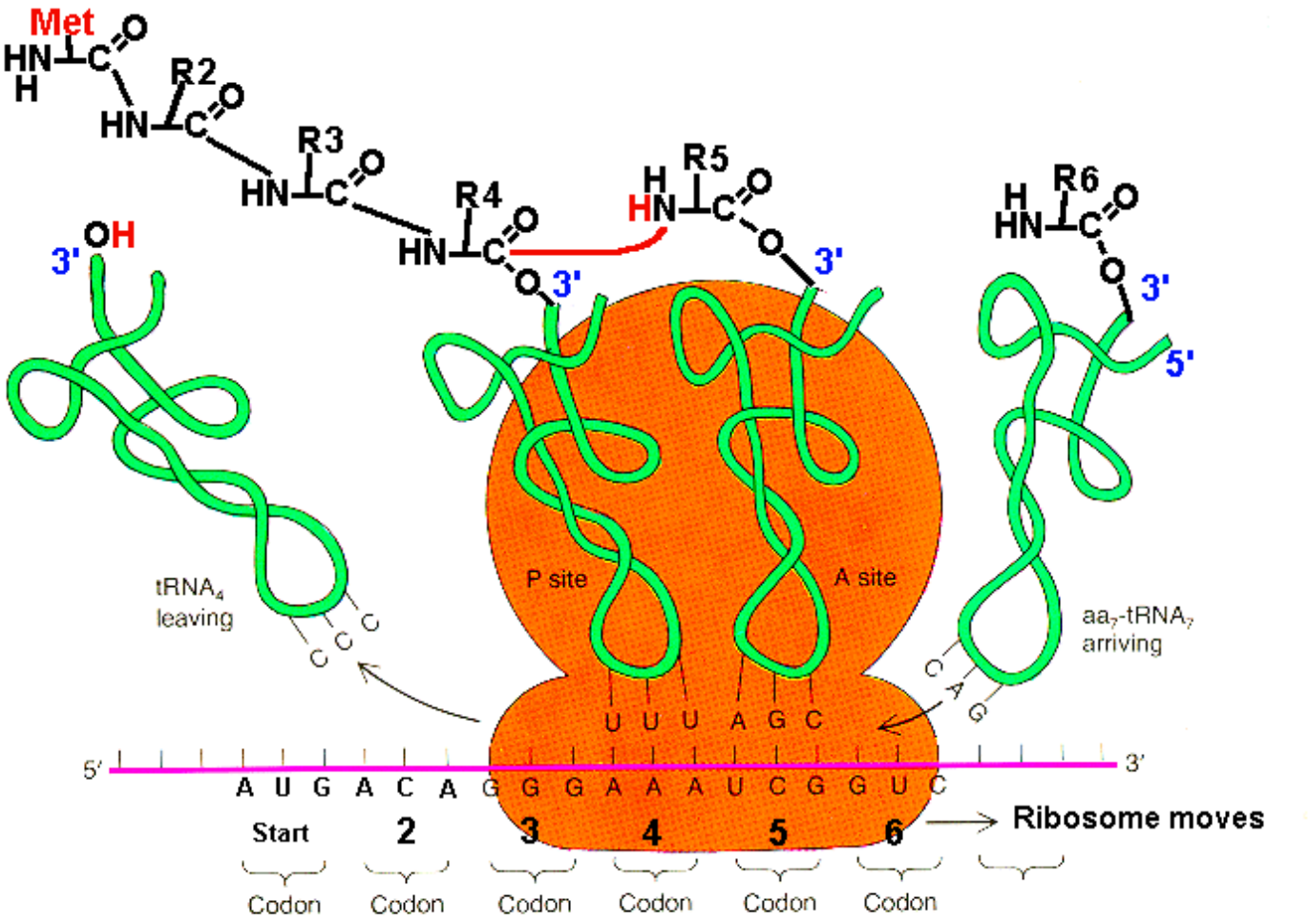
On the whole, there has been an improvement in relationships over the last

Continued on Page A2, Column 1

## CARTER MAKES PLEA TO JEWS ON MIDEAST

### He Calls on Leaders to Support

# Translation

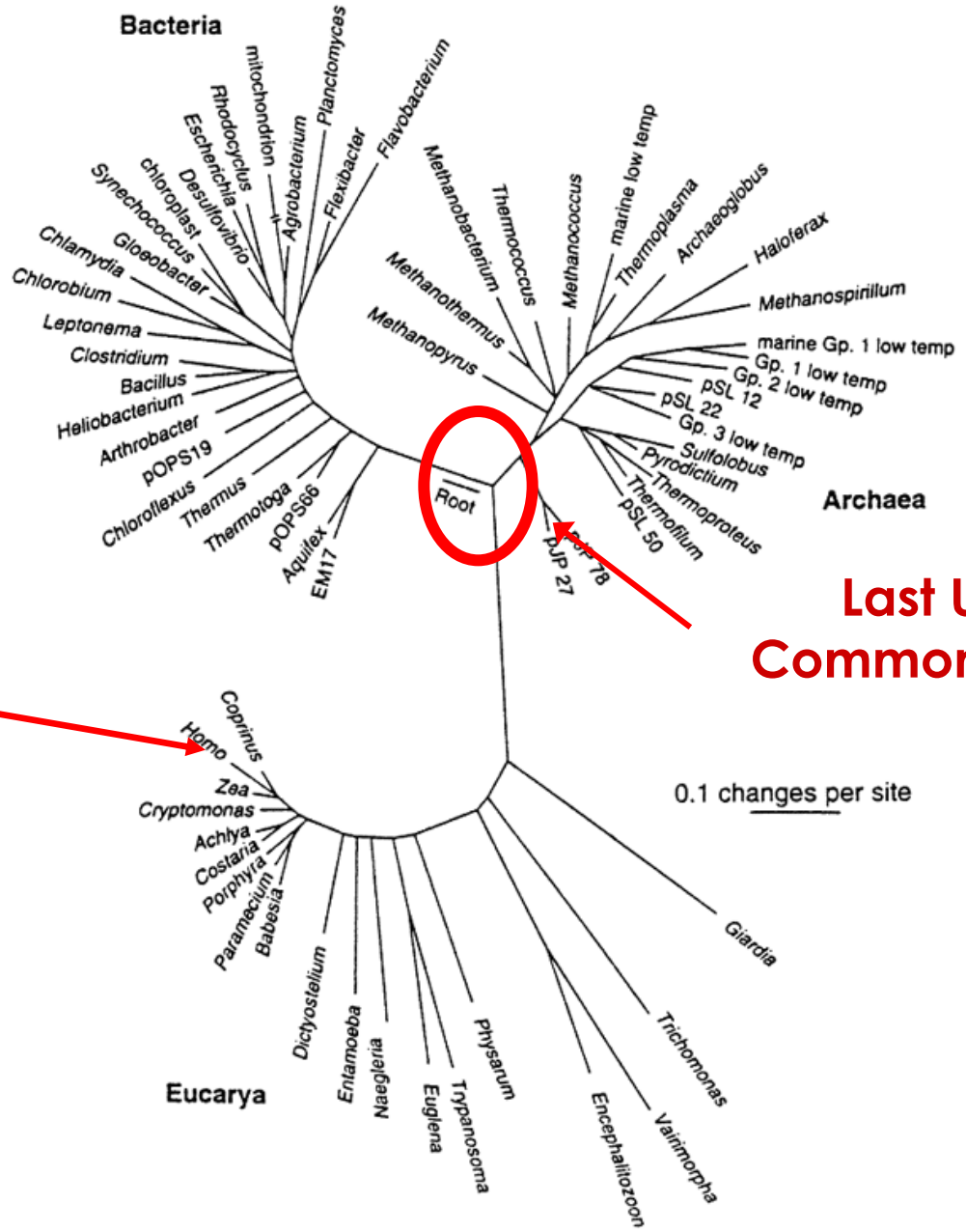



Modified from Griffiths et al., AN INTRODUCTION TO GENETIC ANALYSIS, 6th Ed., W.H. Freeman & Co., 1996.

# Molecular phylogeny


- Fragments of 16SrRNA gene for different species
- Strong similarities, but also differences
- Differences reflect divergent evolutionary history
- “Edit path” between sequences => evolutionary history of organisms

# The Tree of Life




**You are here**

**Last Universal Common Ancestor ?**

*- NR Pace, Science 1997*

*Proc. Natl. Acad. Sci. USA*
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

# Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

The first argument concerns the stability of the general phenotypes. The general eubacterial phenotype has been stable for at least 3 billion years—i.e., the apparent age of blue-green algae (31). The methanogenic phenotype seems to be at least this old in that branchings within the two urkingdoms are comparably deep (see Table 1). The time available to form each phenotype (from their common ancestor) is then short by comparison, which seems paradoxical in that the two phenotypes are so fundamentally different. We think that this ostensible paradox implies that the common ancestor in this case was not a prokaryote. It was a far simpler entity; it probably did not evolve at the "slow" rate characteristic of prokaryotes; it did not possess many of the features possessed by prokaryotes, and so these evolved independently and differently in separate lines of descent.

*Proc. Natl. Acad. Sci. USA*
Vol. 74, No. 11, pp. 5088–5090, November 1977
Evolution

# Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

The first argument concerns the stability of the general phenotypes. The general eubacterial phenotype has been stable for at least 3 billion years—i.e., the apparent age of blue-green algae (31). The methanogenic phenotype seems to be at least this old in that branchings within the two urkingdoms are comparably deep (see Table 1). The time available to form each phenotype (from their common ancestor) is then short by comparison, which seems paradoxical in that the two phenotypes are so fundamentally different. We think that this ostensible paradox implies that the common ancestor in this case was not a prokaryote. It was a far simpler entity; it probably did not evolve at the "slow" rate characteristic of prokaryotes; it did not possess many of the features possessed by prokaryotes, and so these evolved independently and differently in separate lines of descent.

# The Tree of Life

- The existence of the tree shows that all life on Earth is related.
  - The root of the "big tree" is about 3.5-3.8 billion years ago.
- Why can't we see further back in time?
  - Is it because of lack of phylogenetic resolution?
  - Or is there a fundamental reason that we lose the scent of life?
- The existence of a tree is not mandatory.
  - It arises from the vertical descent of individual lineages, genes being transferred to successive generations with variation.

# Darwinian evolution = population genetics

- Origin of species
- Change in frequency of gene alleles in a population

# Darwinian evolution = population genetics

- Origin of species
- Change in frequency of gene alleles in a population

- But what was the character of evolution before the Last Universal Common Ancestor?
  - No species
  - No genes
  - What drives the evolution of complexity?
  - Why was the process of evolution in early life so fast?
  - Is this process manifested and connected with evolution today?

# Life is Physics: Evolution as a Collective Phenomenon Far From Equilibrium

Nigel Goldenfeld[1] and Carl Woese[1,2]

## Abstract

Evolution is the fundamental physical process that gives rise to biological phenomena. Yet it is widely treated as a subset of population genetics, and thus its scope is artificially limited. As a result, the key issues of how rapidly evolution occurs and its coupling to ecology have not been satisfactorily addressed and formulated. The lack of widespread appreciation for, and understanding of, the evolutionary process has arguably retarded the development of biology as a science, with disastrous consequences for its applications to medicine, ecology, and the global environment. This review focuses on evolution as a problem in nonequilibrium statistical mechanics, where the key dynamical modes are collective, as evidenced by the plethora of mobile genetic elements whose role in shaping evolution has been revealed by modern genomic surveys. We discuss how condensed matter physics concepts might provide a useful perspective in evolutionary biology, the conceptual failings of the modern evolutionary synthesis, the open-ended growth of complexity, and the quintessentially self-referential nature of evolutionary dynamics.

Annu. Rev. Condens. Matter Phys. 2011. 2:375–99



# 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


KAPLAN

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.**

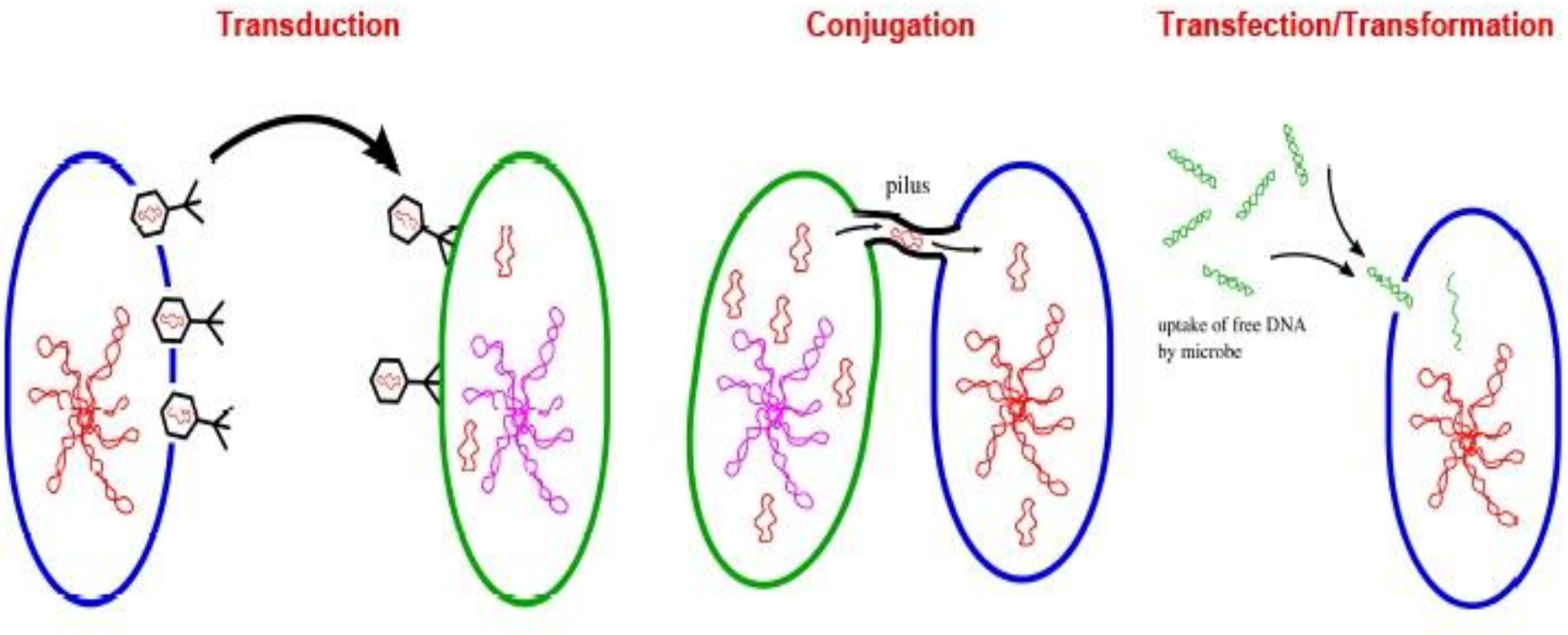
FURTHER READING
Frigaard, N., Martinez, A., Mincer, T. & DeLong, E. *Nature* **439**, 847–850 (2006).
Sullivan, M. *et al. PLoS Biol.* **4**, e234 (2006).
Pedulla, M. *et al. Cell* **113**, 171–182 (2003).
Vetsigian, K., Woese, C. & Goldenfeld, N. *Proc. Natl Acad. Sci. USA* **103**, 10696–10701 (2006).

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

CONNECTIONS

Are there more general modes of evolution than vertical Darwinian evolution today?

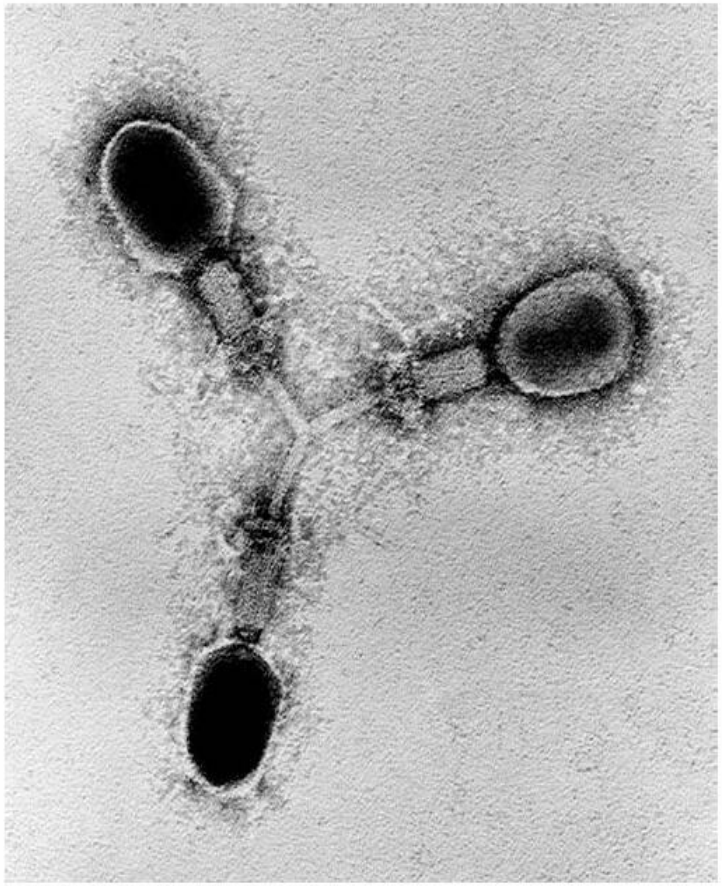
# Horizontal gene transfer




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

# Gene transfer between host and virus

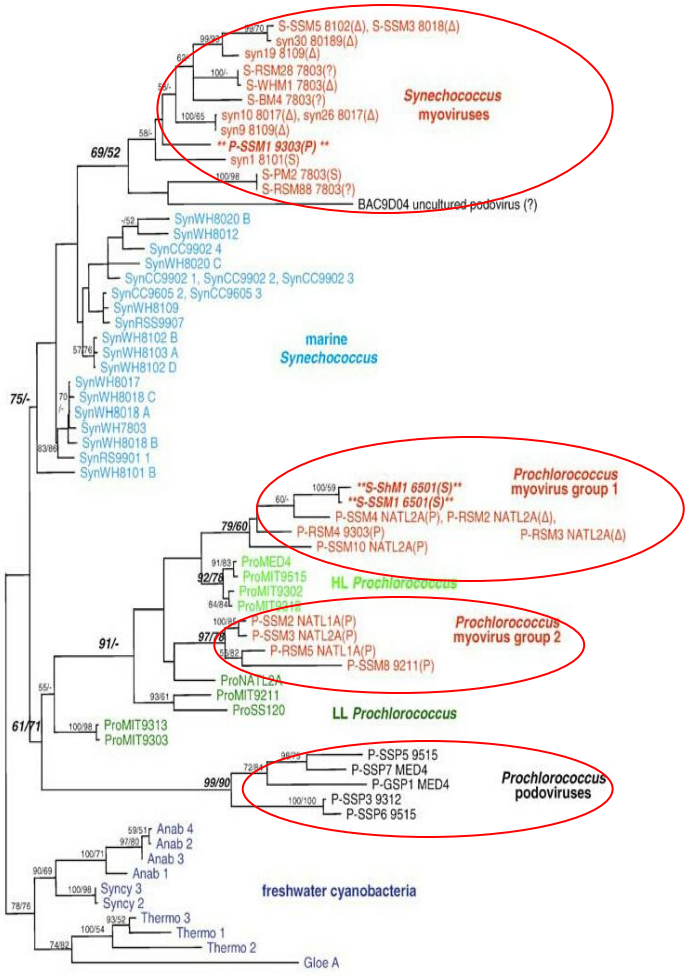
Sullivan et al. , PlosBiol (2006)



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 …)

# 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 …)

**Take-home message: the "Cloud" was invented 3-4 billion years ago by microbes**

# Patterns in the genetic code

# 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 |
| | Met | | Lys | Arg | A |
| | | | | | 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<br>---<br>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 |
| | Met | | Lys | Arg | A |
| | | | | | 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 |

# 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 |
| | Met | | Lys | Arg | A |
| | | | | | G |
| G | Val | Ala | Asp | Gly | U |
| | | | | | C |
| | | | Glu | | A |
| | | | | | G |

# Polar requirement

- Polar requirement is a counterplay between two tendencies of amino acids
  - Polar interaction of ring N on bases with polar part of amino acid
  - Non-polar interaction between organic parts of base with amino acid

- In 1965-1966 Carl Woese and colleagues devised a way to quantify the chemical properties of amino acids, and called their measure "polar requirement"

# Polar requirement

- Woese et al (1966) document interactions between amino acids and bases to see if these had influenced the code.
- Explored chromatography of amino acids in water-pyridine mixtures.
  - Separation not sufficient evidence – amino acids could just be moving with water
  - Explore trend of motion of amino acids with water concentration!
- $R_M$ measures mobility of amino acid
- Polar requirement is slope of log $R_M$ vs. Mole % $H_2O$

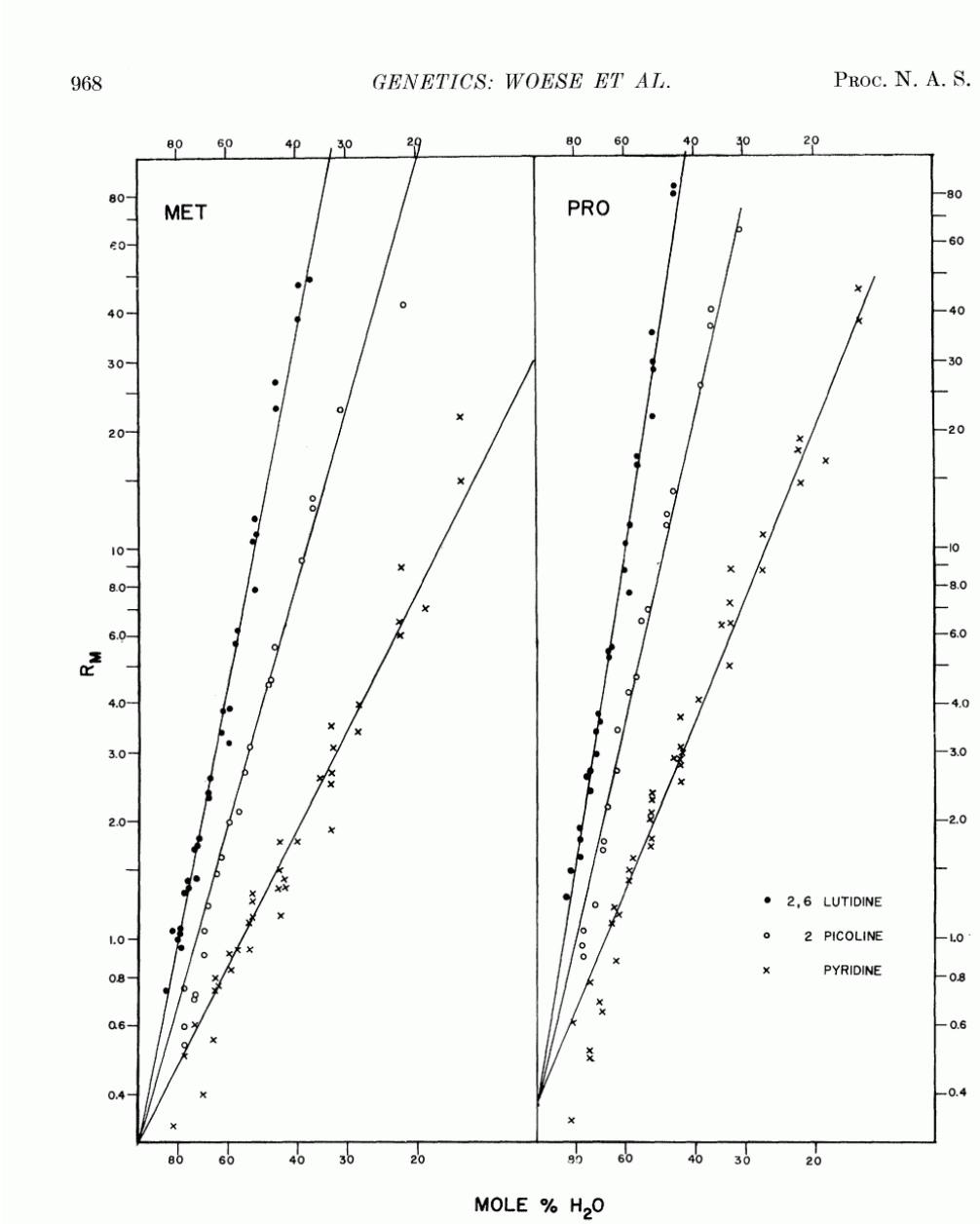



FIG. 1.—Log amino acid $R_M$ vs. log mole fraction $H_2O$ in the chromatography solvent.

## 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 |
| | Met 5.3 | | Lys 10.1 | Arg | A |
| | | | | | G |
| G | Val | Ala | Asp 13.0 | Gly | U |
| | | | | | C |
| | | | Glu 12.5 | | A |
| | | | | | G |

The genetic code is not just universal …
it’s nearly optimal in minimizing errors

# 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

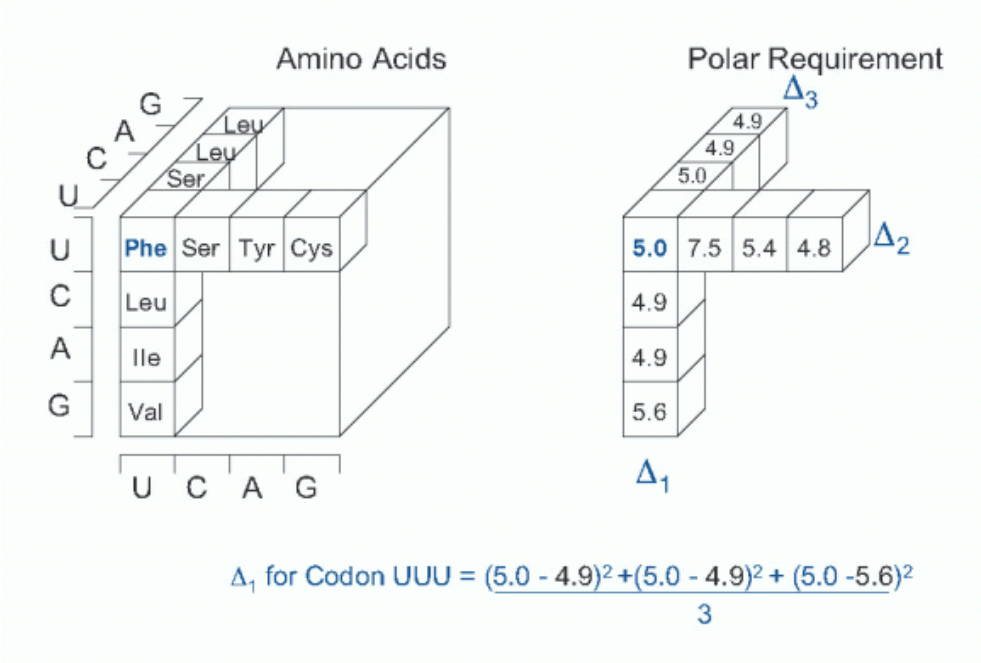
| | U | C | A | G | |
|---|---|---|---|---|---|
| U | Phe | Ser | Tyr | Cys | U |
| | | | | | C |
| | Leu | | STOP | STOP | A |
| | | | | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | | | | | C |
| | | | Gln | | A |
| | | | | | G |
| A | Ile | Thr | Asn | Ser | U |
| | | | | | C |
| | | | Lys | Arg | 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
- $20! \sim 10^{18}$ possible codes

# Simulated genetic codes

- 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

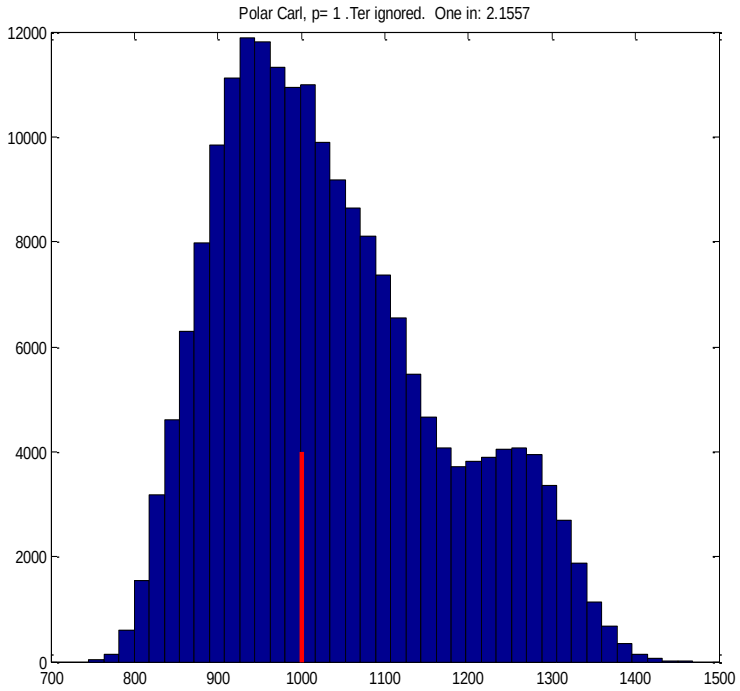


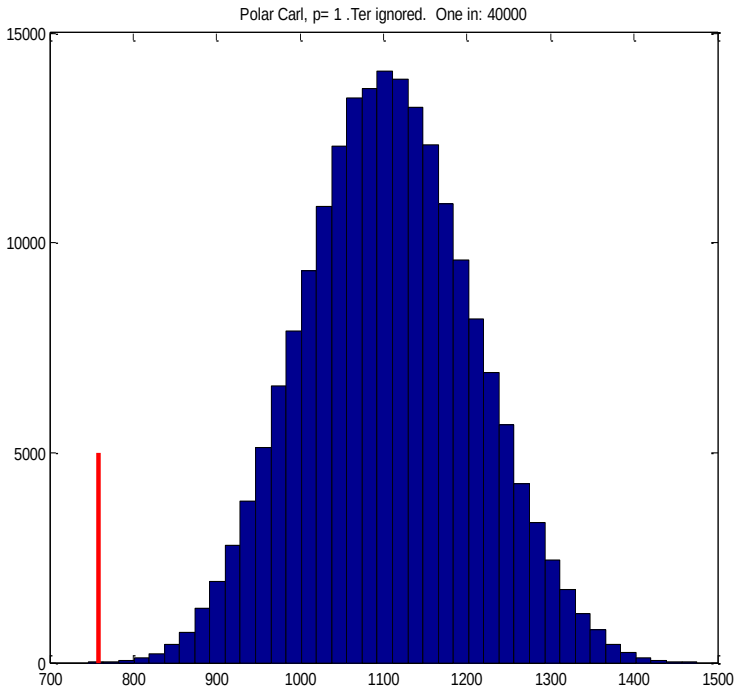

Knight Ph.D thesis (2001)

# Simulated genetic codes

Naïve expectation – current code is frozen accident

Actual result – current code is not a frozen accident




Probability distribution

Code property

# Optimality of the genetic code with respect to the polar requirement

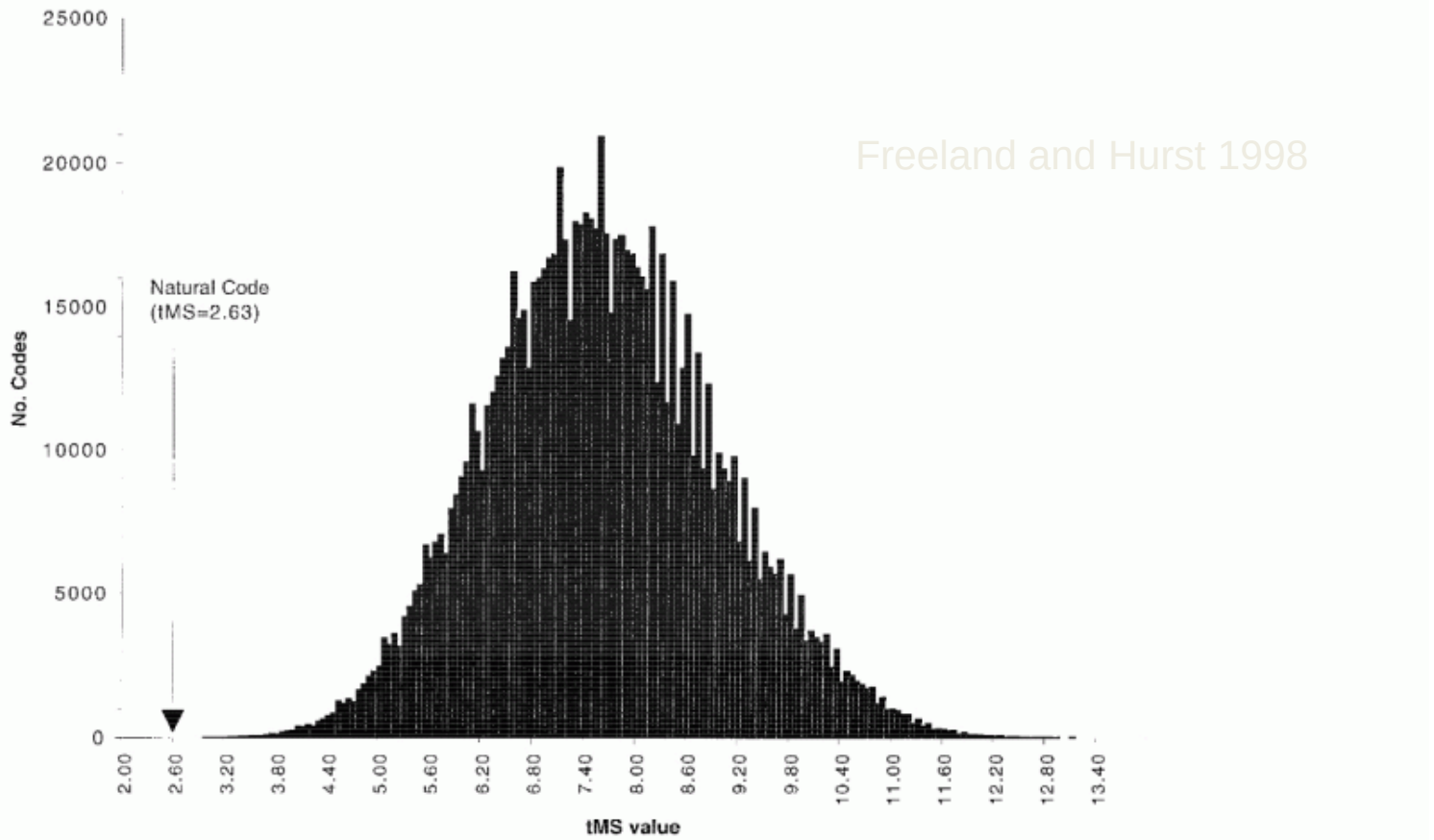



**Fig. 7.** Frequency distribution for the tMS0 (=MS0 adjusted for mistranslation parameters) values obtained from 1 million randomly generated variants of the natural genetic code. The *X* axis gives a particular range of categories of MS values, and the *Y* axis gives the number of random variant codes 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 $(19 \pm 4) \times 10^{-8}$

# Mechanisms for evolution of the genetic code

# The puzzle

- The genetic code is a “complex trait”.
- The genetic code has evolved.
- Yet, it is universal … Why isn’t there a diversity of codes?
- Is the universality accidental or is it there for a deep reason?
- What does it tell us about early evolution?

- **Suggestion: The code is special because it is an “innovation-sharing protocol”**

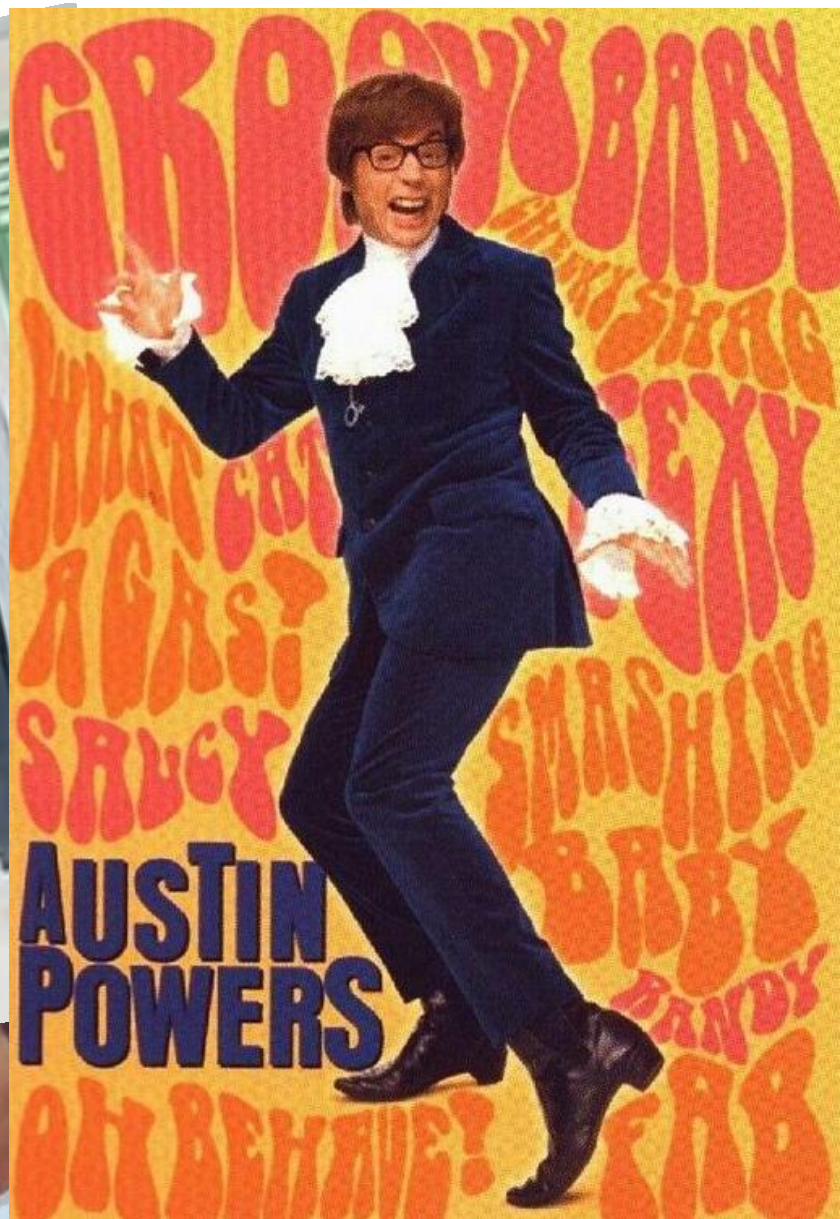
  - **Translation is a unique cellular function, part of the “operating system” of the cell.**

# How can a code evolve?

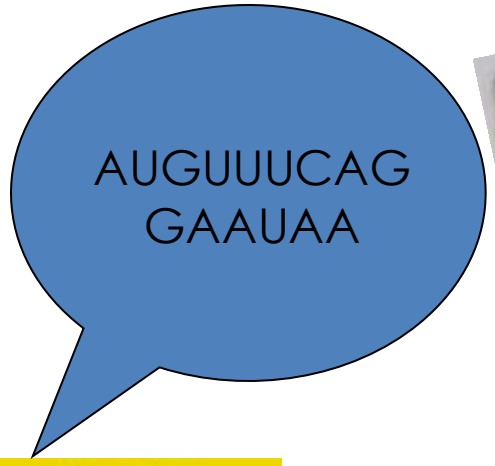
http://xtremzik1.free.fr/Pochettes/Austin_Powers-front.jpg

http://wesclark.com/am/spy_game.html



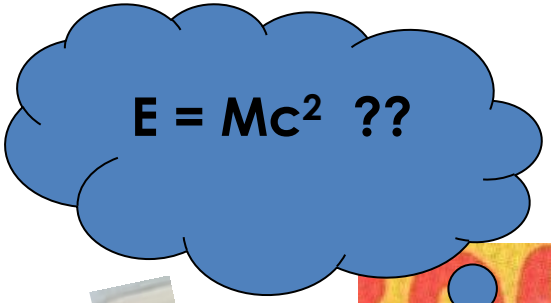
# How can a code evolve?

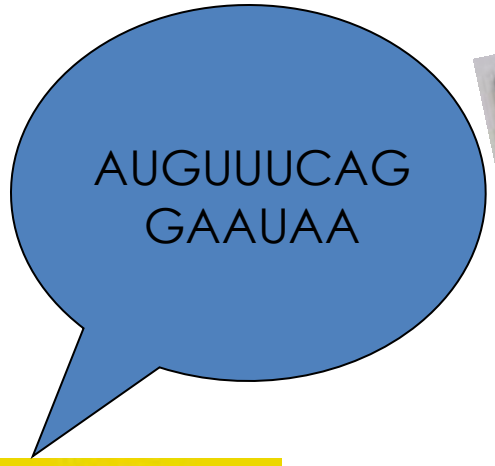





- Only a single message

# How can a code evolve?





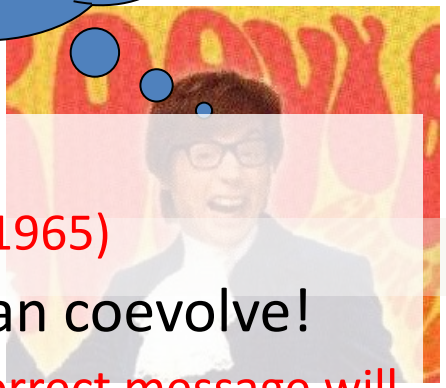
- Change (evolve) the code => message is garbled

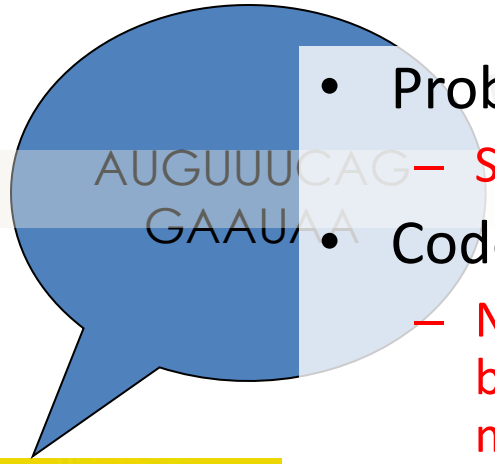
# How can a code evolve?

Prob(Phe Gln Glu) = 0.9

AUGUUUCAG
GAAUAA

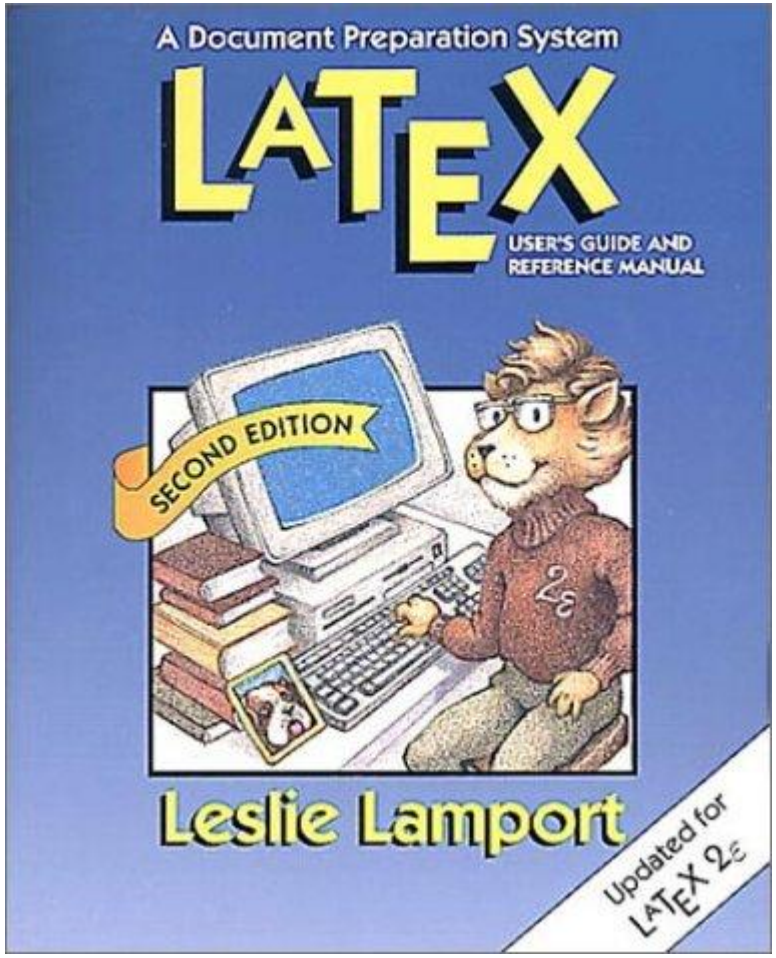

# How can a code evolve?

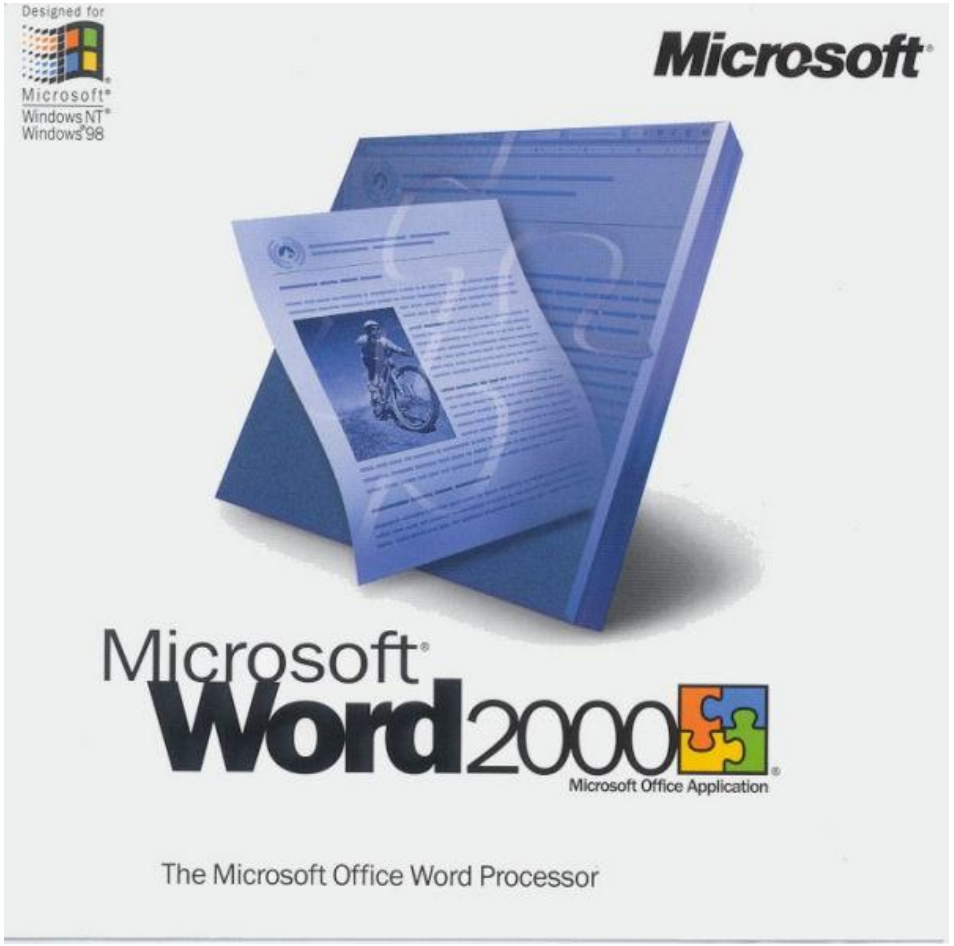



- Probabilistic code book
  - Statistical proteins (Woese 1965)
- Code book and message can coevolve!
  - Non-zero probability that correct message will be interpreted, because there is not a single message but an ensemble or probability distribution of messages
  - Refinement of code and greater accuracy demanded for translation

# Selection pressure influences code dynamics

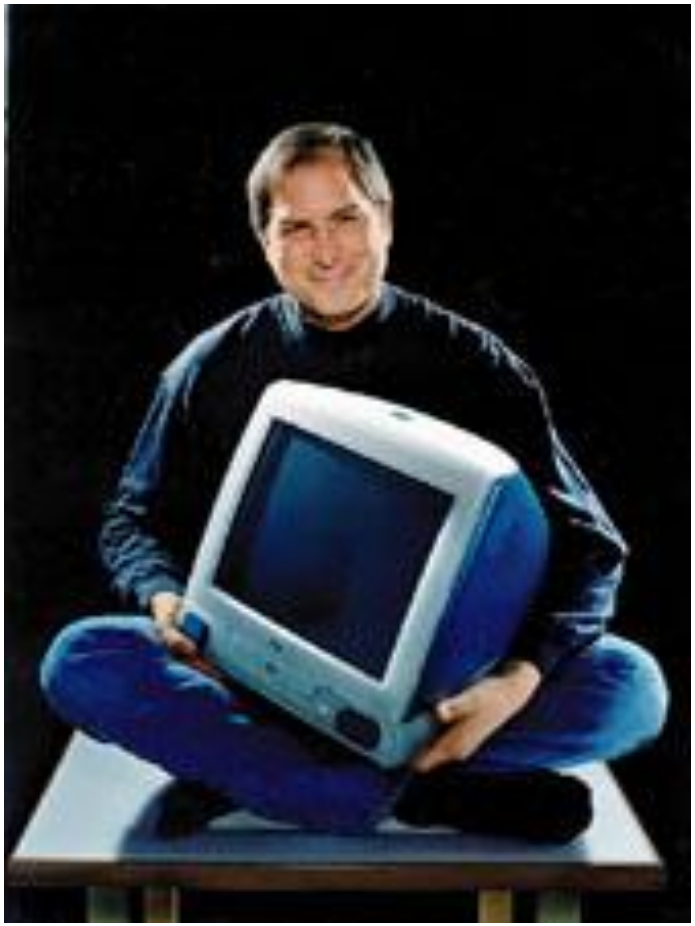



VS.



# Competition between innovation-sharing protocols


http://alt1040.com/uploads/bill-gates-1983.jpg



http://www.henno.com/jobs.jpg

- Community with access to the broadest range of innovations has an evolutionary advantage

A problem has been detected and Windows has been shut down to prevent damage
to your computer.

The problem seems to be caused by the following file: SPCMDCON.SYS

PAGE_FAULT_IN_NONPAGED_AREA

If this is the first time you've seen this Stop error screen,
restart your computer. If this screen appears again, follow
these steps:

Check to make sure any new hardware or software is properly installed.
If this is a new installation, ask your hardware or software manufacturer
for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware
or software. Disable BIOS memory options such as caching or shadowing.
If you need to use Safe Mode to remove or disable components, restart
your computer, press F8 to select Advanced Startup Options, and then
select Safe Mode.

Technical information:

*** STOP: 0x00000050 (0xFD3094C2,0x00000001,0xFBFE7617,0x00000000)

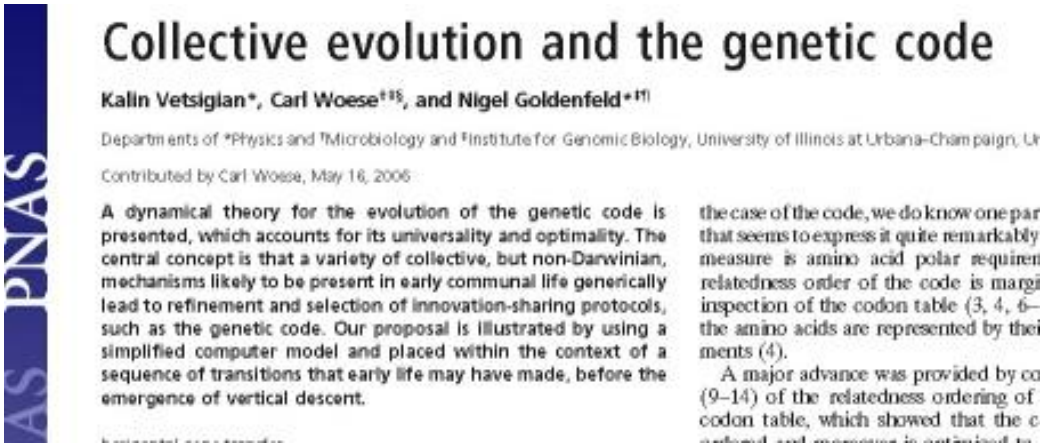

***  SPCMDCON.SYS - Address FBFE7617 base at FBFE5000, DateStamp 3d6dd67c

# 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

# Three mechanisms for universality

- Evolutionary scenario combining the three mechanisms
  - Competition between innovation-sharing protocols
  - HGT of protein coding regions
  - Genetic exchange of translational components
- Model based on work of Ardell and Sella (2002), but with HGT, tRNA population dynamics.



PNAS

## Collective evolution and the genetic code

Kalin Vetsigian*, Carl Woese*†§, and Nigel Goldenfeld*‡¶

Departments of *Physics and †Microbiology and ‡Institute for Genomic Biology, University of Illinois at Urbana–Champaign, U

Contributed by Carl Woese, May 16, 2006

A dynamical theory for the evolution of the genetic code is presented, which accounts for its universality and optimality. The central concept is that a variety of collective, but non-Darwinian, mechanisms likely to be present in early communal life generically lead to refinement and selection of innovation-sharing protocols, such as the genetic code. Our proposal is illustrated by using a simplified computer model and placed within the context of a sequence of transitions that early life may have made, before the emergence of vertical descent.

the case of the code, we do know one par
that seems to express it quite remarkably
measure is amino acid polar requiren
relatedness order of the code is margi
inspection of the codon table (3, 4, 6–
the amino acids are represented by thei
ments (4).

A major advance was provided by co
(9–14) of the relatedness ordering of
codon table, which showed that the c

# 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!**

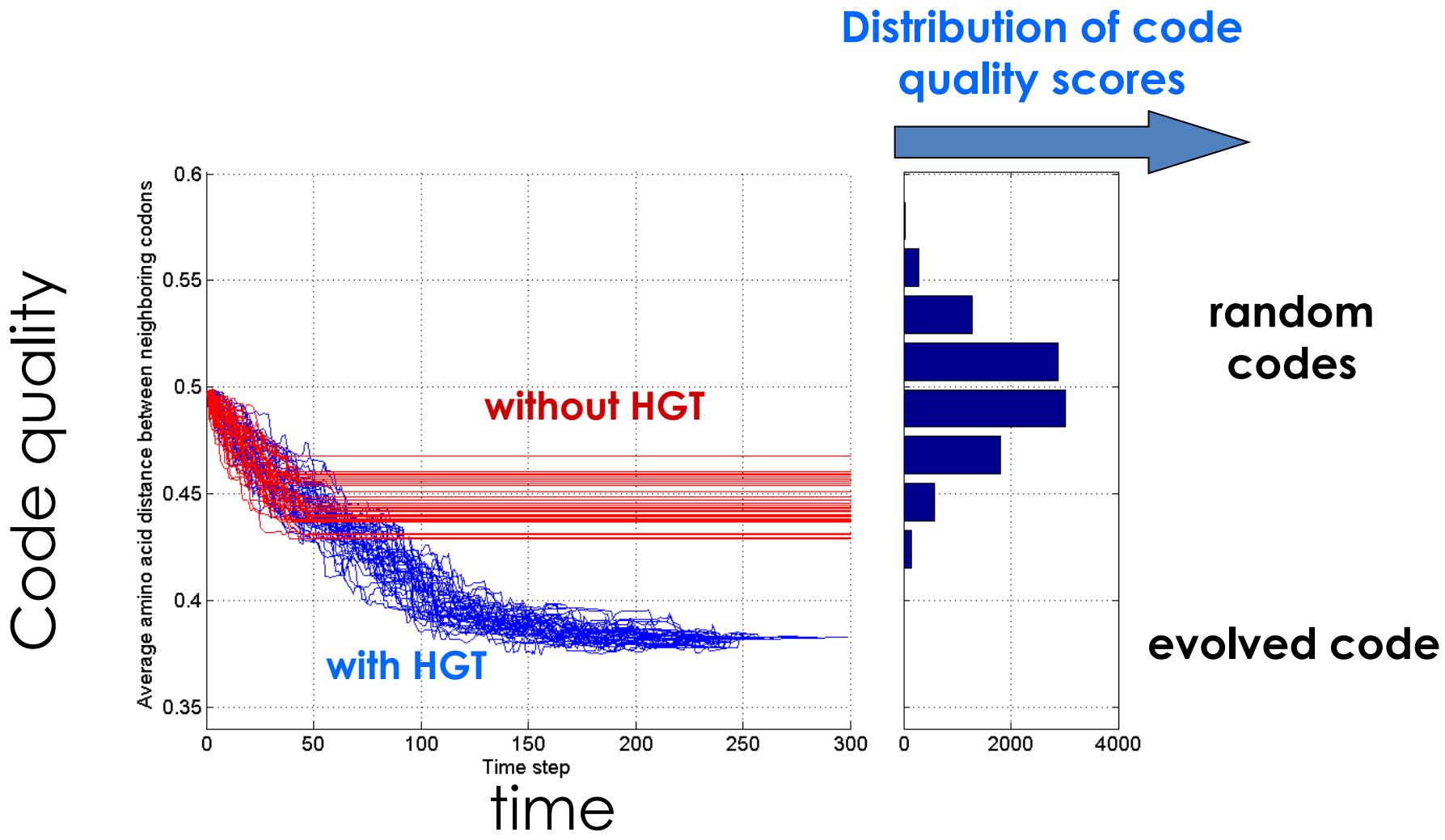
# What do we measure?

- 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

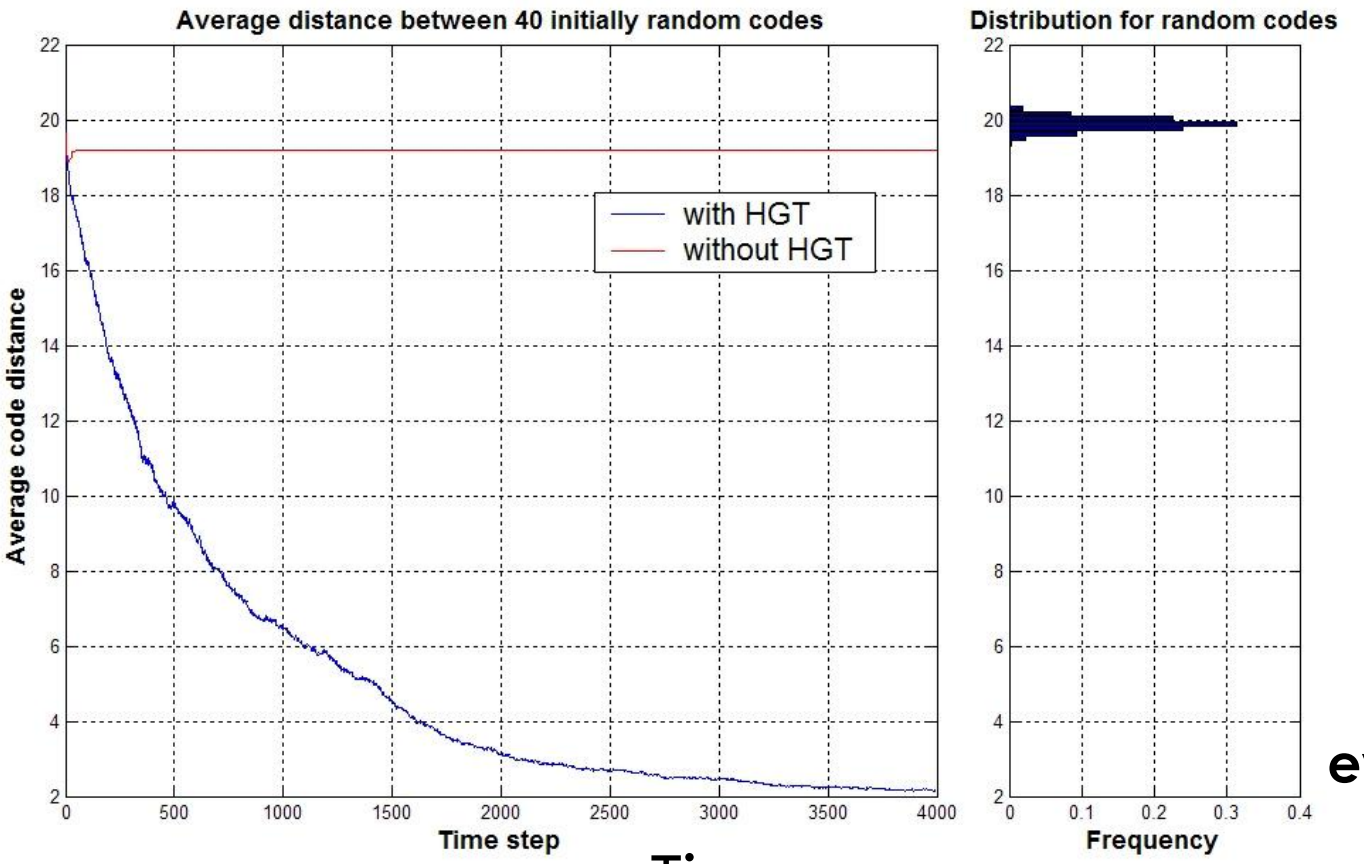



**HGT leads to optimality**

# 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** |

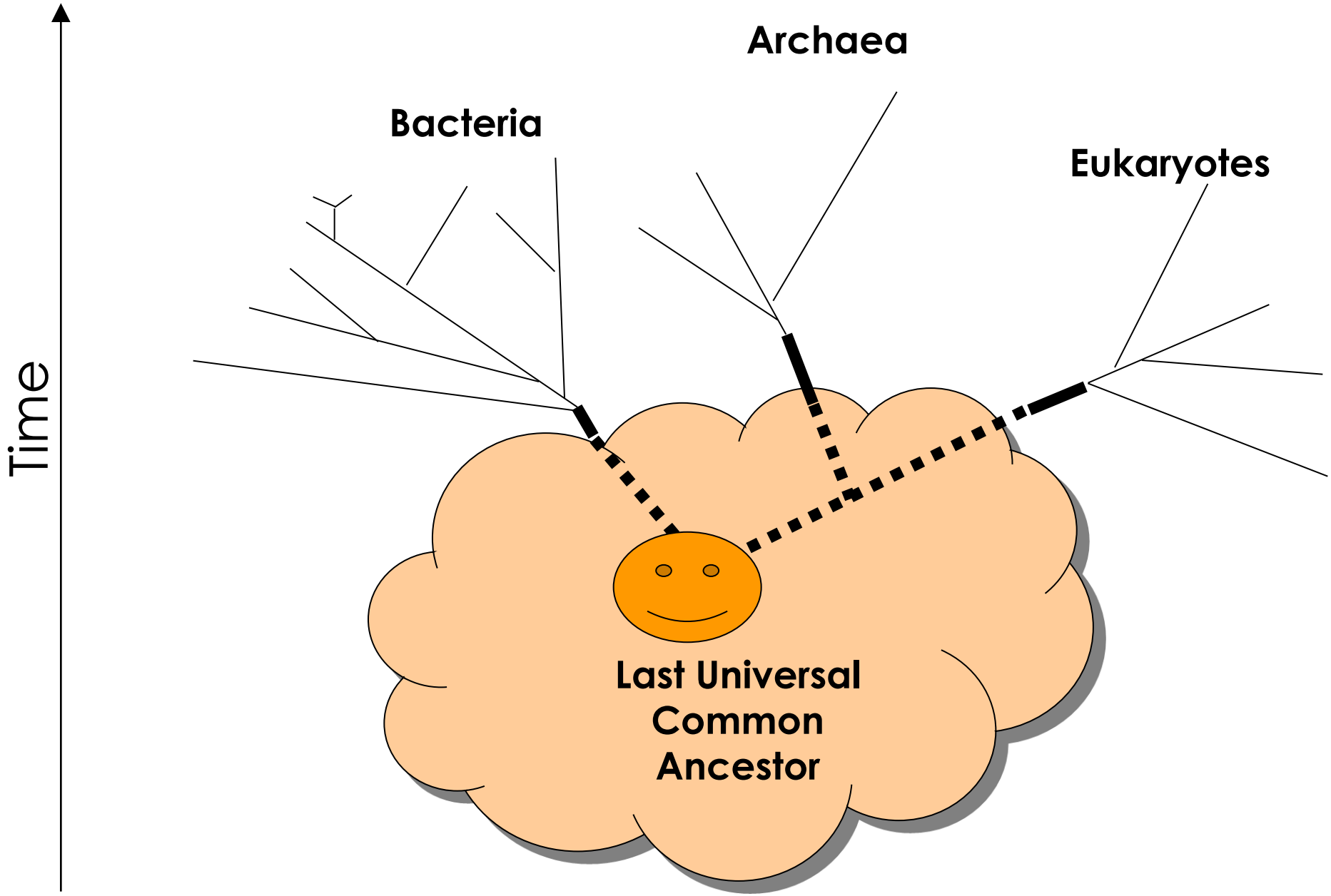
| | | | |
|---|---|---|---|
| Compositions | Membrane domain of multi-spanning proteins | NAKH920108 | 12736 |
| | Membrane domain of single-spanning proteins | NAKH920105 | 54367 |
| | Sheet propensity | KANM800102 | 12730 |
| | Helix propensity | KANM800101 | 47764 |
| | Beta-strand indices for alpha/beta-proteins | 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** |

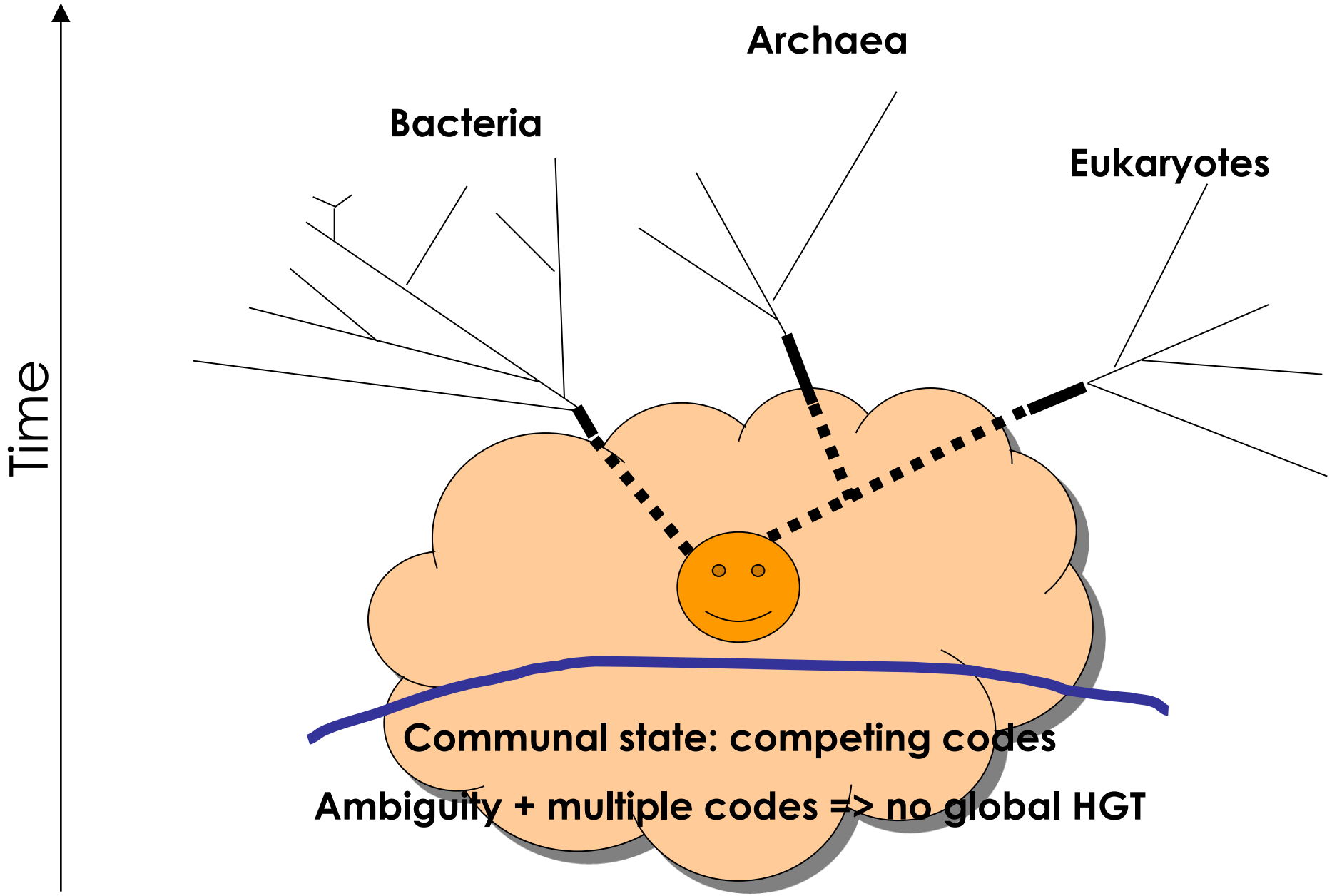
# Is the polar requirement special?

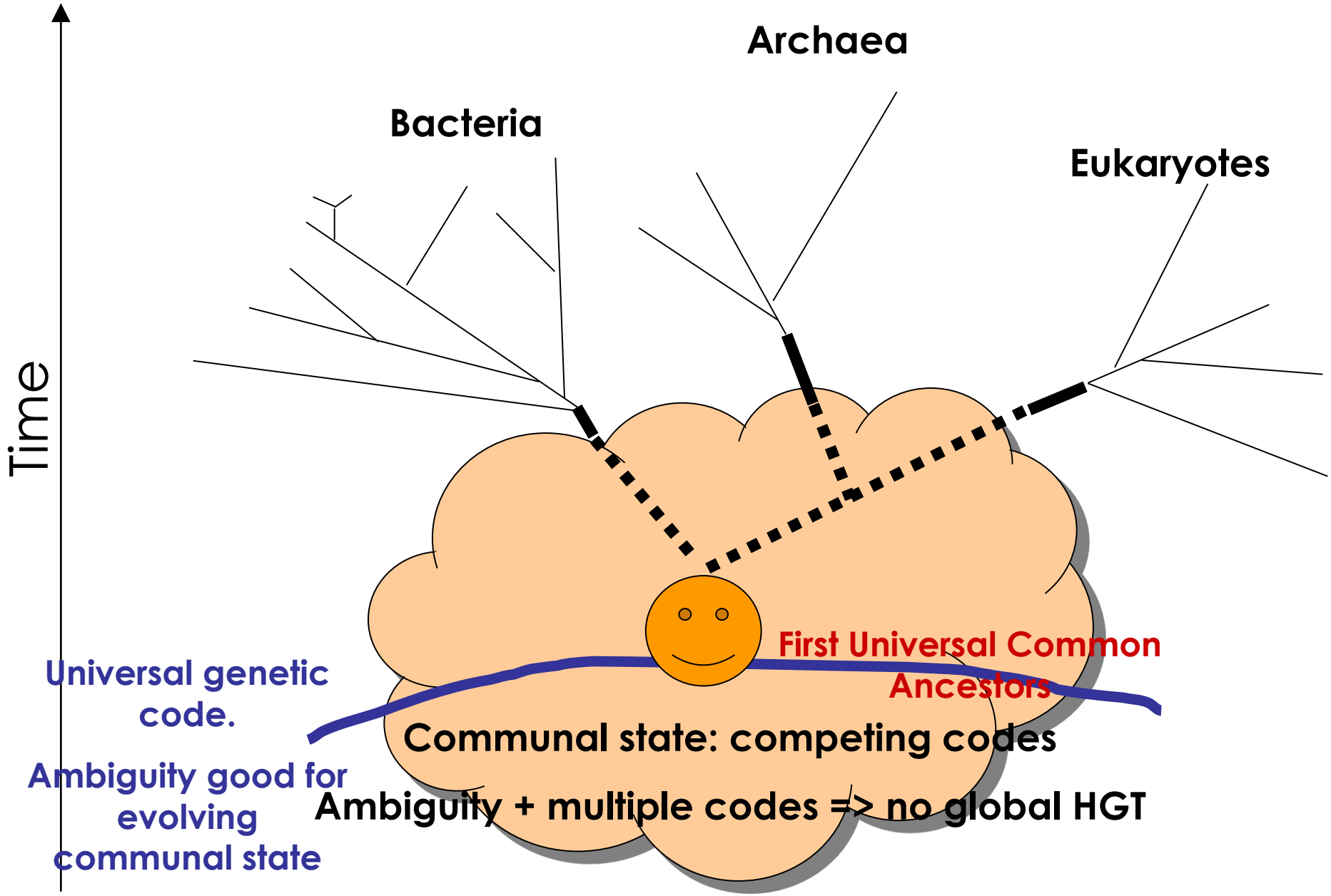
- Polar requirement & Grantham polarity are free amino acid properties – hence prebiotic
  - Canonical code always optimal
- Early peptide/Modern peptide
  - Canonical code virtually never optimal
- Optimality of polar requirement suggests genetic code is a relic of very early life!

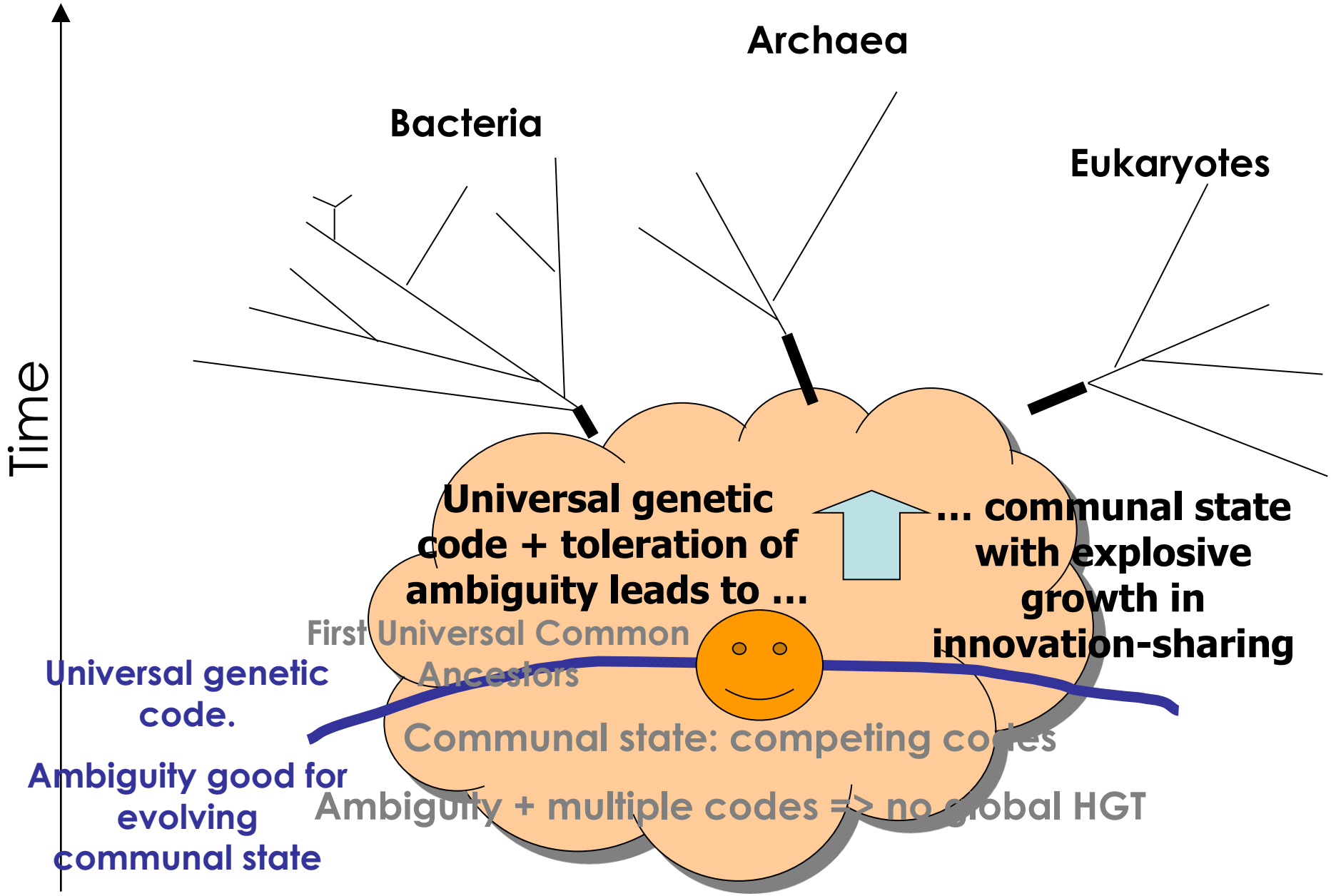
# The phase diagram of life …

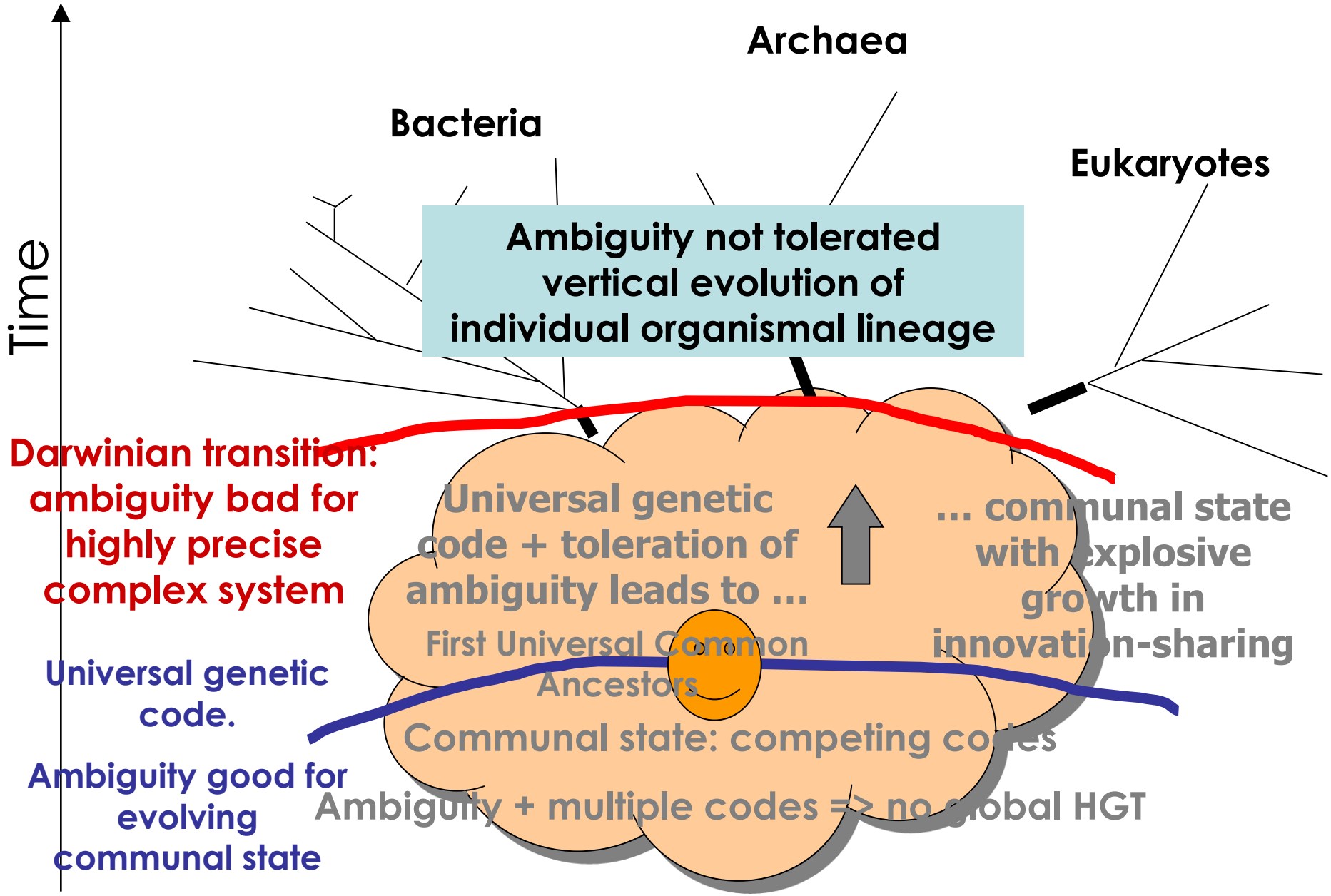
… as inferred from the collective dynamics of innovation-sharing protocols

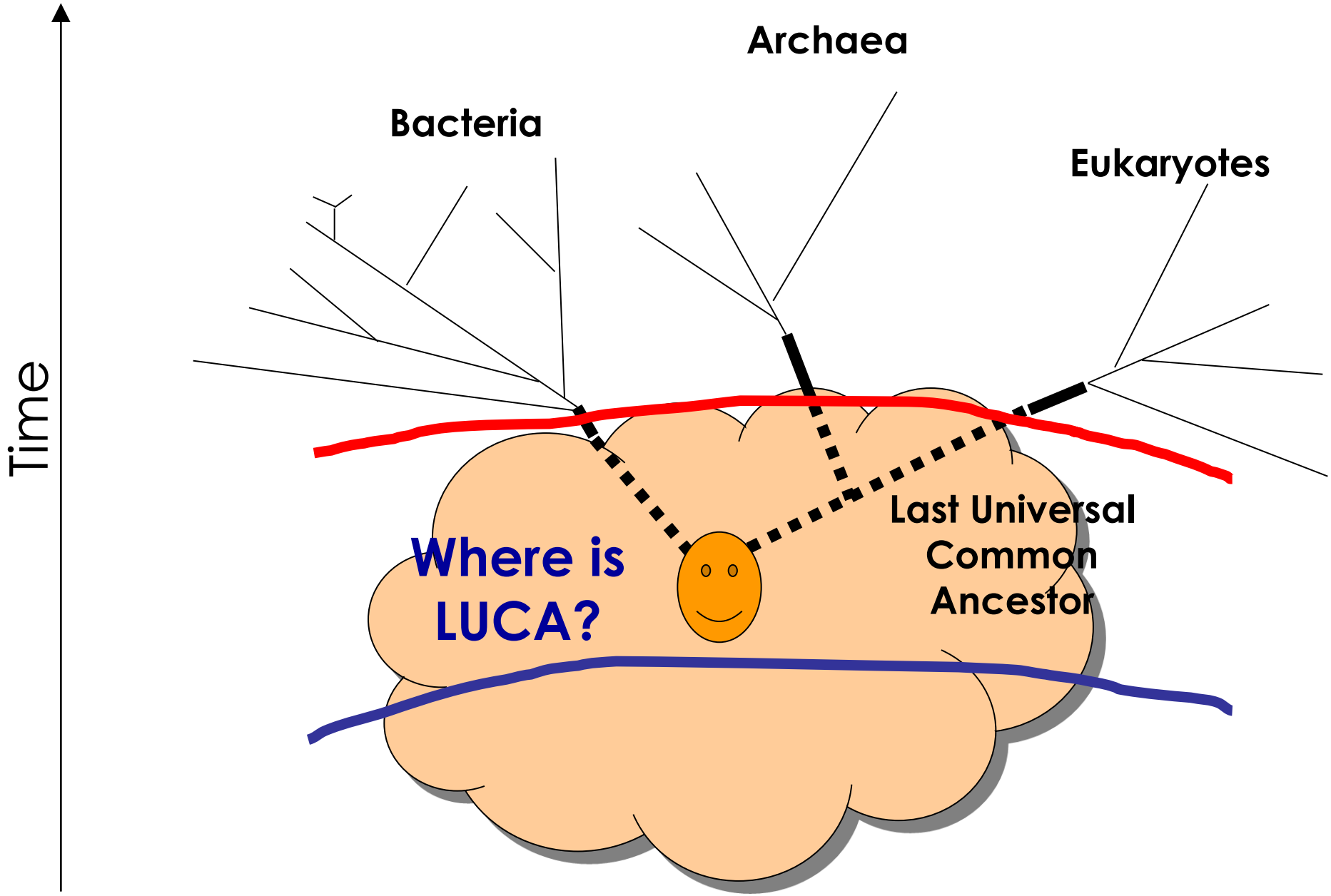

Archaea
Bacteria
Eukaryotes
Time
Last Universal
Common
Ancestor

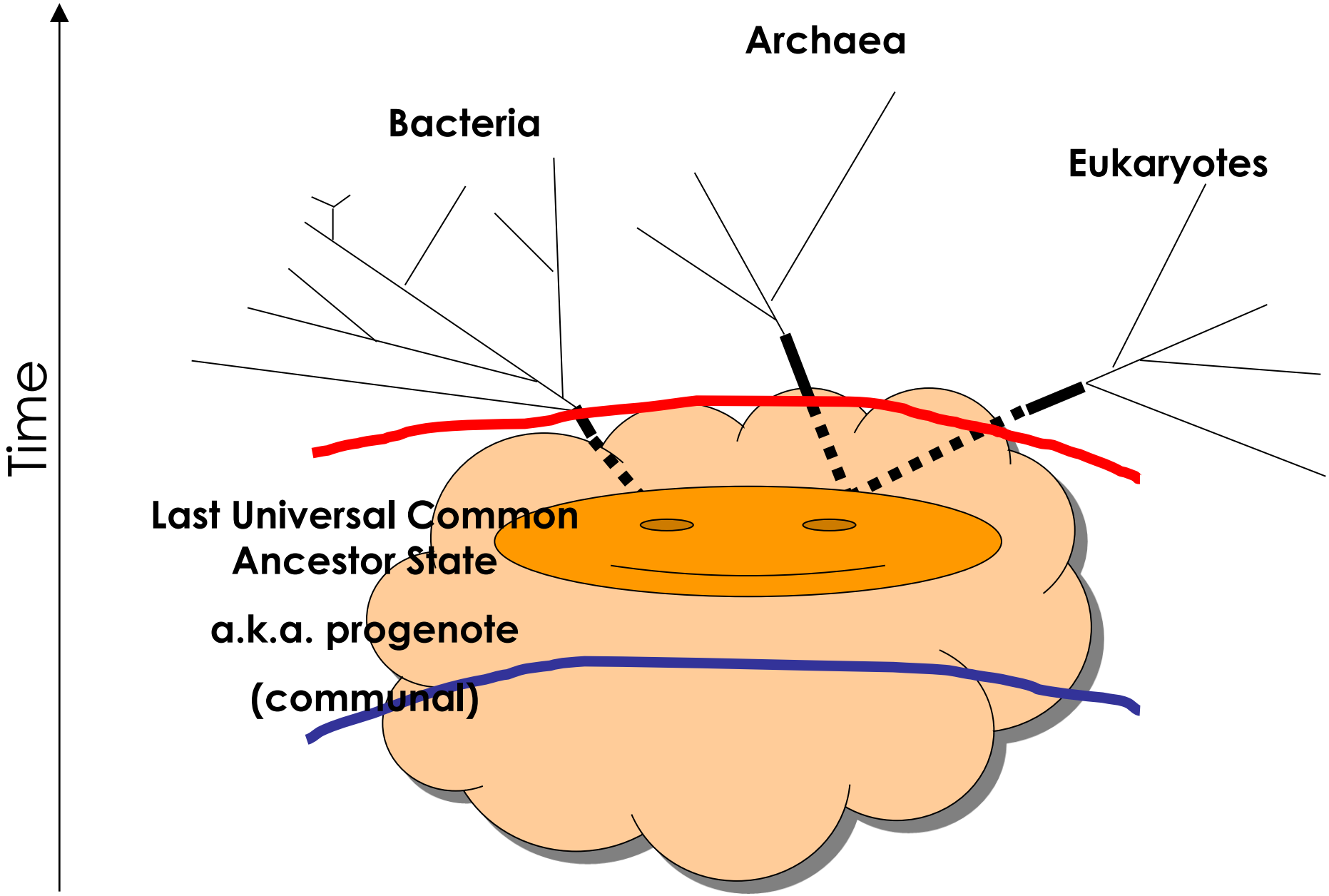

Archaea
Bacteria
Eukaryotes
Time
Communal state: competing codes
Ambiguity + multiple codes => no global HGT

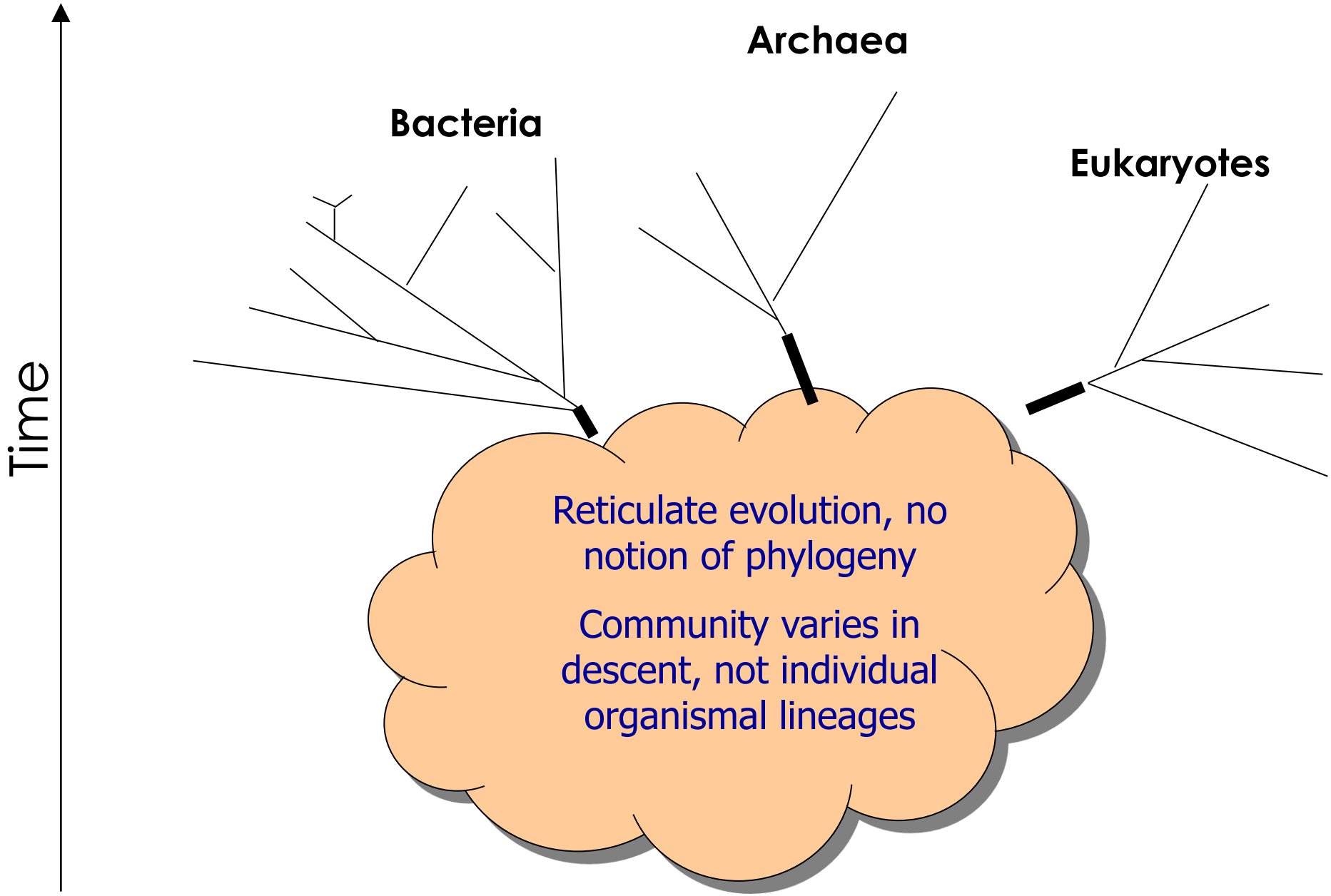

Archaea
Bacteria
Eukaryotes
Time
First Universal Common Ancestors
Universal genetic code.
Communal state: competing codes
Ambiguity good for evolving communal state
Ambiguity + multiple codes => no global HGT


Archaea
Bacteria
Eukaryotes
Time
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


Archaea
Bacteria
Eukaryotes
Time
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


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

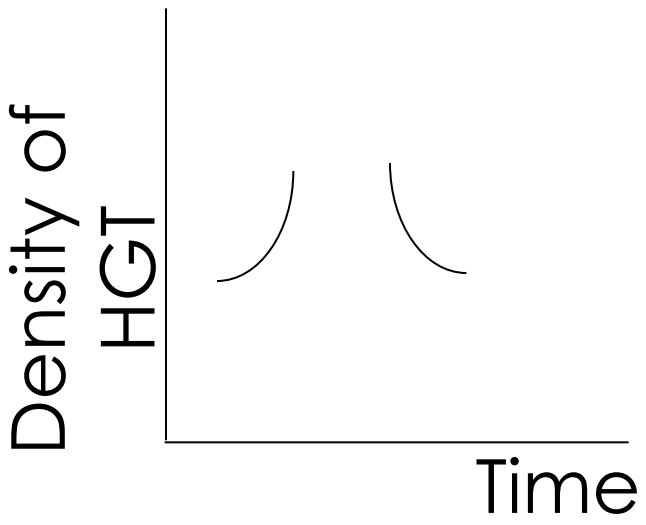

Archaea
Bacteria
Eukaryotes
Time
Last Universal Common Ancestor State
a.k.a. progenote
(communal)

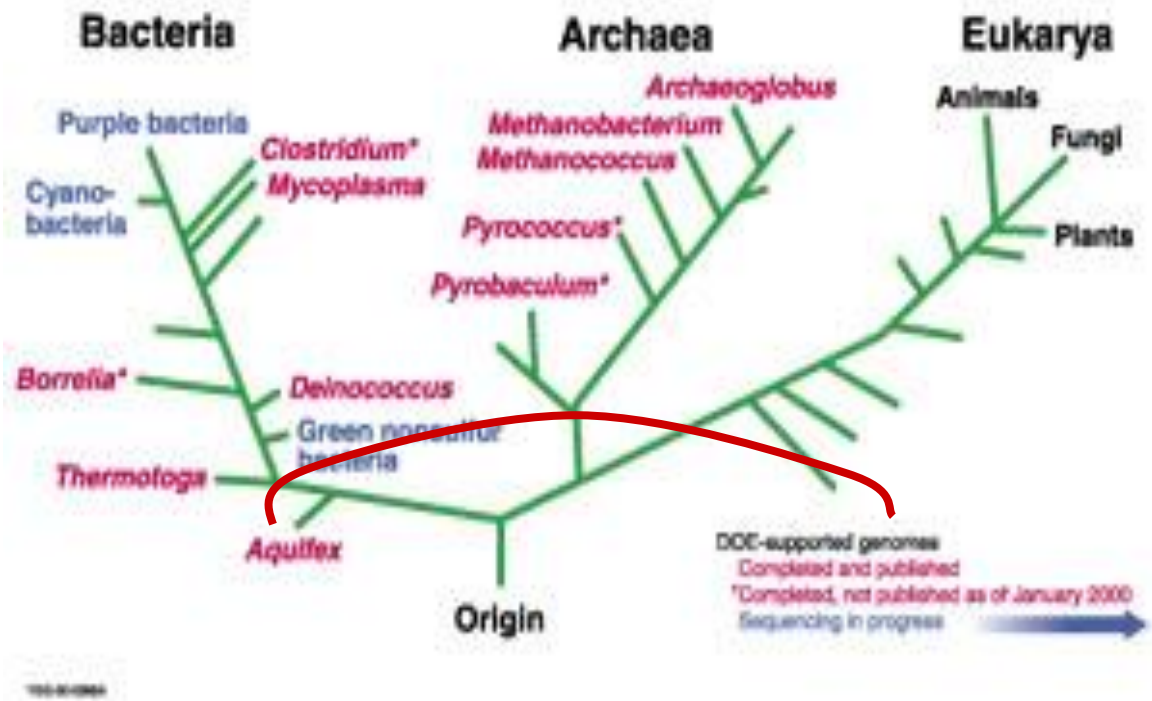

Archaea
Bacteria
Eukaryotes
Time
Reticulate evolution, no notion of phylogeny
Community varies in descent, not individual organismal lineages

# Predictions and future work

# Experimental predictions

- Collective phase may explain the origin of genes, viruses, gene transfer agents
- Density of HGT events may show near-transitional behavior
- Map density of HGT events onto the ribosomal phylogenetic tree

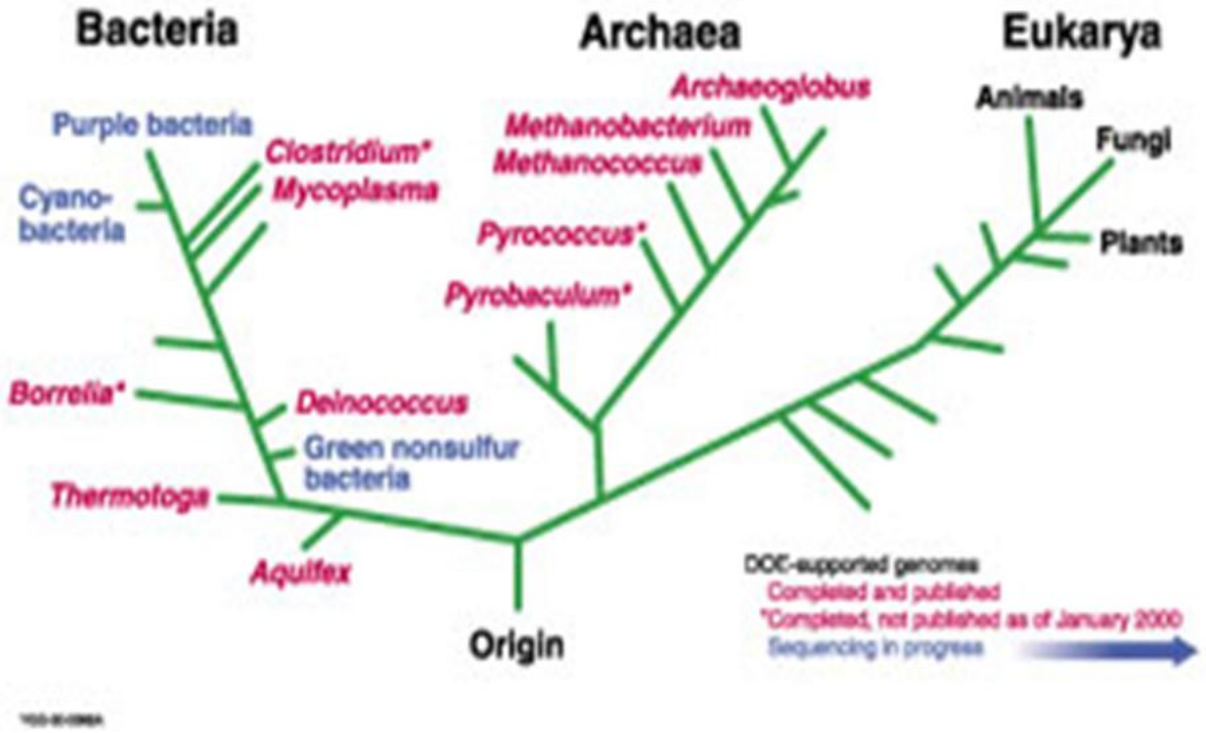


# Windows on the progenote

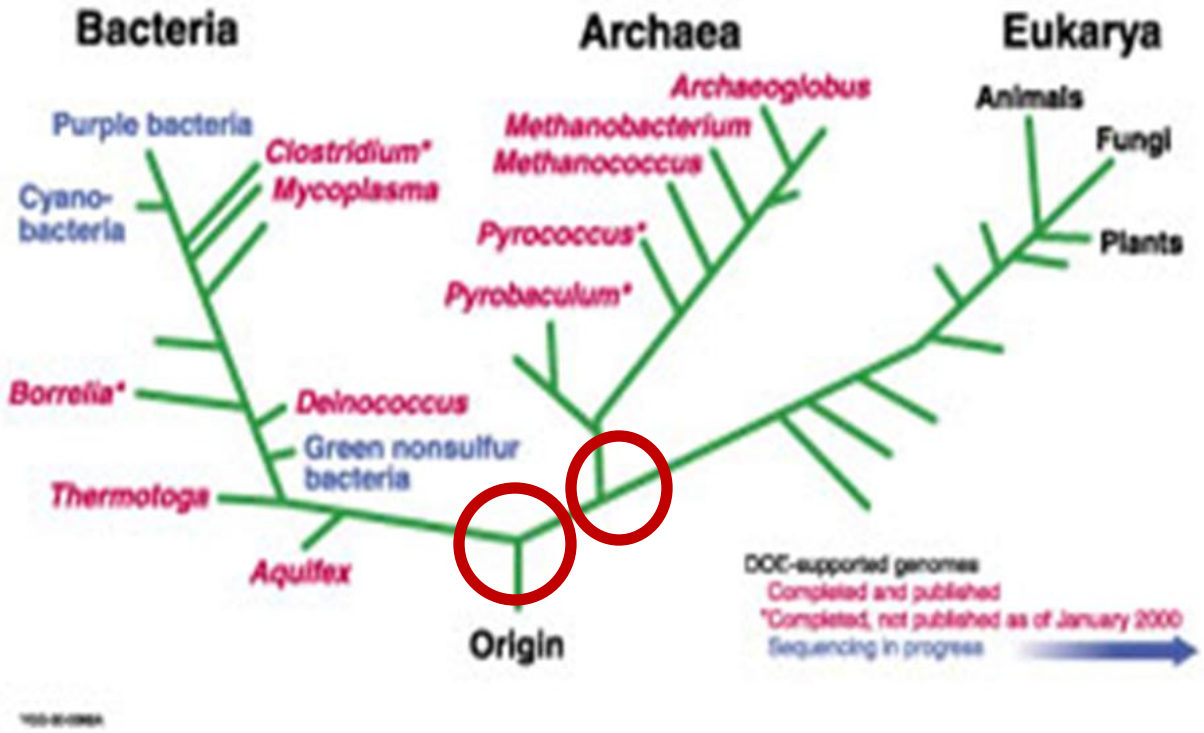
- Underlying idea: look at core cellular machinery that will be most reflective of collective effects
  - Amino-acyl-tRNA synthetases, the least conserved elements of translational machinery
  - Rich structure: two classes, complex relatedness groups
  - Deviations from canonical phylogenetic pattern of rRNA, elongation factors, transcriptional machinery
- Evidence for ancient HGT events?
- **Action plan: comparative phylogenomics of these ancient proteins to look for HGT events, map out the density of these correlated to rRNA phylogeny; extend to other proteins; attempt to understand the order in which transitions of evolutionary structure took place**

# Breakdown of the progenote state and the transition to vertical evolution



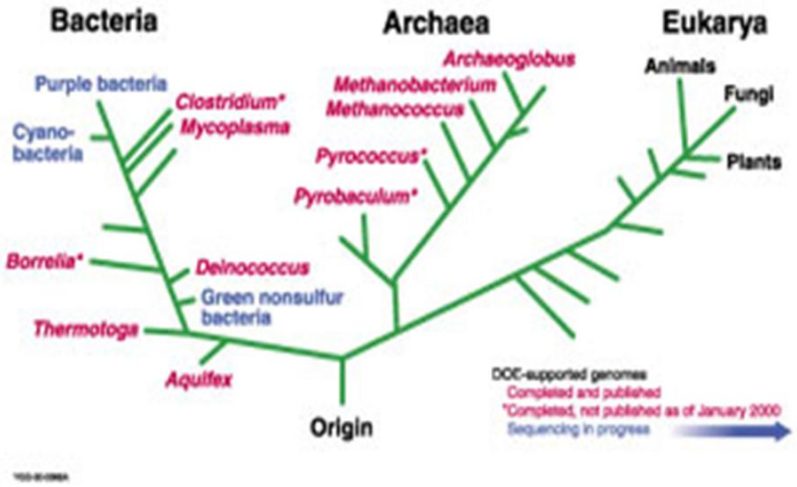

- Theory suggests that a progenote state is an inevitable feature of the growth of complexity. Why does it break down to vertical evolution?

# Breakdown of the progenote state and the transition to vertical evolution




- Two “last common ancestors” to characterize and compare!

# Breakdown of the progenote state and the transition to vertical evolution




- Theory suggests that a progenote state is an inevitable feature of the growth of complexity. Why does it break down to vertical evolution?

## Action plan

- Theoretical models of the instabilities of evolving communities of innovation-sharing organisms analyzed as a “pattern formation” in gene space
- Search for components of cells that have not crossed the Darwinian threshold. Such processes would be more amenable to change, e.g. incorporation of functional homologs from different domain. We will try this experimentally, e.g. interchangeability of sliding clamps between bacteria and Archaea.
  - Comparison with components that are more rigid and frozen in (e.g. ribosome itself).
  - Characterize the diversity of mechanisms for translation, transcription, replication using bioinformatics.
- Characterization of the core cellular machinery of last common ancestor of Eukaryotes and Archaea. Comparison with bacteria.

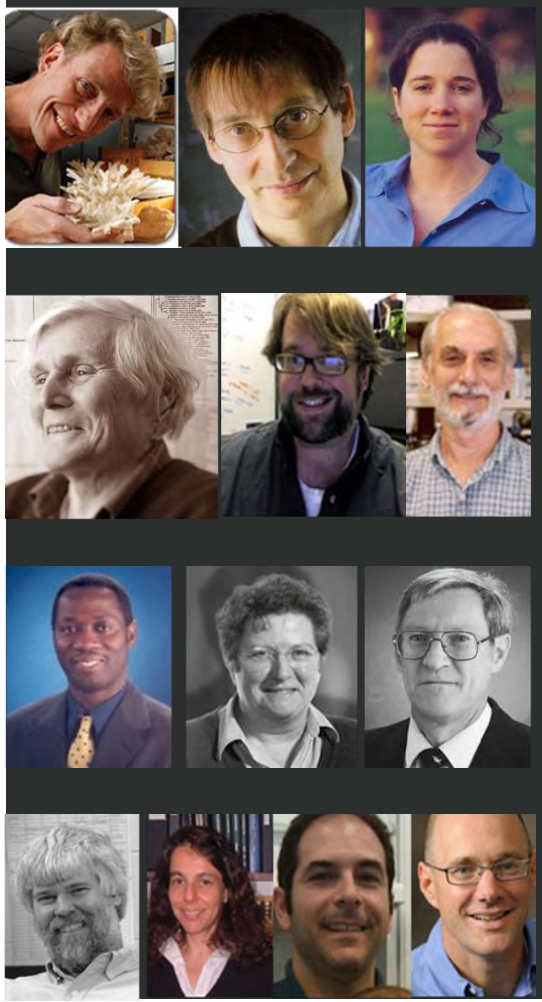
# Big Questions

- What are the universal principles governing evolving matter and the existence of the phenomenon of life?
- What evolutionary dynamics allowed life to go from nothing to LUCA in < 1 billion years?
- Is there evidence for the progenote state embedded in genomes?
- How did the progenote state break down > 3.8 billion years ago?
- What determines the speed of evolution? Collective effects? Environmental stress?

# Conclusions

- The genetic code is universal and optimal in the sense of minimizing errors
- These properties do not follow from vertical Darwinian evolution but can be the outcome of a collective phase of life (the progenote)
- Other remnants of this state are likely buried in ancient genomes

# Institute for Universal Biology



- One of two new research groups joining NASA Astrobiology Institute (NAI)

BIG QUESTIONS:

- Why does life exist?
- How does it arise in different environments and planets?
- How did life evolve before there were genes, species, individual organisms and cells? Clearly not Darwinian!
- What was the nature of evolution at this early time?

BIG ANSWERS

- Build a “Hubble telescope for genes”, exploring deep evolutionary time
- Seek signatures of early collective states of life occurring before individual organisms on earth

- Highly diverse research team includes fields of microbiology, geobiology, computational chemistry, genomics, physics and engineering. This research could only be done at UIUC.
- Significant outreach component - new middle school teacher partnership, web-based video series, massive online open astrobiology course (pending Coursera inclusion).