



TDWG 2014

Biodiversity
Information
Standards



Jönköping
Sweden

Genomics and Metagenomics: What is the role of the Darwin Core?

Discussion Session

To what extent will the increased use of genomic and metagenomic technologies affect our understanding and documentation of biodiversity?

<https://docs.google.com/document/d/1y8-lylIE780FuEicHixVA0Btz4DcGuh4s5i6CMloUjQ/>

Genomics and Metagenomics: What is the role of the Darwin Core?

Robert J. Robbins
RJR8222@gmail.com

Genomics and Metagenomics: What is the role of the Darwin Core?

Robert J. Robbins

RJR8222@gmail.com

➔ (<http://www.rj-robbins.com/presentations/TDWG-2014>) ➔

Good Advice:

Always design a thing by considering
it in its next larger context...

- Eliel Saarinen

Abstract:

The most fundamental unit of traditional biodiversity - the individual organism (defined as a physically connected, multi-cellular aggregation, with all of the cells clonally derived from one ancestral cell) - has no parallel in the world of prokaryotic biology. Yet recent advances (metagenomics tools, etc) have shown that about half of the world's biomass and by far most of its physiological (as opposed to morphological, or mechanical) biodiversity occurs in the prokaryotic realm. Other work is establishing that symbiosis (in the sense of the bio-outsourcing of some key biological functions) is far more common than traditionally recognized. In fact, because horizontal gene transfer (HGT) occurs without regard for species boundaries, it is now becoming clear that some microbial genes are more attributes of a particular ecosystem than of a particular individual or species. (This is why, for example, the same "cassette" of pathology genes is often found spreading across multiple bacterial species in hospital settings.)

All of the above suggests that an understanding of biodiversity is incomplete, and significantly so, without the inclusion of microbial biodiversity, both as a component of ecosystems ("free living" microbial communities) and also at the level of full understanding of macroscale organisms (as influenced by the presence, make-up, physiology, and function of their associated microbiomes).

From the perspective of biodiversity informatics, the complete addition of a microbial component using current methods and schemas is not possible. This session will begin with an overview of the gaps in today's biodiversity information systems which preclude an accurate understanding of patterns of life on Earth, and how those patterns are changing. This overview will be followed by a discussion around possible roles for Darwin Core in helping to fill these gaps, and how the standard might have to evolve in order to fulfill these roles.

Genomics/Metagenomics and Biodiversity:

- What is “biodiversity” and how should biodiversity information be managed?
 - Classical (specimen-based) biodiversity.
 - Genomics and molecular biodiversity.
- Metagenomics reveals biological dark matter.
 - Two trees / two super-niches / one biosphere.
 - Biological dark matter is weird.
- The challenge for biodiversity informatics.
 - Classical biodiversity contains several key assumptions about the fundamental nature of biological systems.
 - Virtually none of those fundamental assumptions apply to biological dark matter.

Genomics/Metagenomics and Biodiversity:

- The vanishing individual.
 - Microbial ecology and prokaryotic systems.
 - Pervasive symbiosis (we are all lichens, now).
 - The holobiont as major functional unit in macro-scale biology.
- Evolution and a post-modern synthesis.
 - Prokaryotes are different.
 - No simple tree of life.
- Reality is not negotiable.
 - Examples from genetics & genomics.
 - The future of biodiversity.
- Welcome to the world of 21st Century biodiversity.

What is “biodiversity” and how should biodiversity information be managed?

Classical (specimen-based) biodiversity.

Traditional biodiversity:

- A fundamental data object in traditional biodiversity is the assertion of an occurrence of a particular (known, named, and classified) organism at some point in time and space.

Multiple Choice:

To be included usefully into a biodiversity information system, a data record **MUST** include detailed information about

1. Organism(s)
2. Specimen(s)
3. Name(s)
4. Taxon (Taxa)
5. All of the above
6. Only some of the above
7. None of the above

What is “biodiversity” and how should biodiversity information be managed?

Genomics and molecular biodiversity.

Why genomics and biodiversity?

- Biodiversity is less a field of biology than a perspective (that of variance) into biology.
- Diversity is a sine qua non of biology; no diversity, no evolution.
- Genetics / genomics are equally central to biology – genetics is the study of the hereditary machinery, the basis of heritable variation, the raw material for evolution, the ultimate source of biodiversity.

Why genomics and biodiversity?

- Biodiversity is less a field of biology than a perspective

The connection between genomics and biodiversity seems obvious and profound.

evolution, the ultimate source of biodiversity.

The Basics:

- As sequencing gets cheaper, its practical applicability to biodiversity will increase.

The Basics:

- As sequencing gets cheaper, its practical applicability to biodiversity will increase.
- Barcode-type data are useful in a diversity-diagnostic sense.

The Basics:

- As sequencing gets cheaper, its practical applicability to biodiversity will increase.
- Barcode-type data are useful in a diversity-diagnostic sense.
- Metagenomic tools allow a broad diversity assessment in a single test.

The Basics:

- As sequencing gets cheaper, its practical applicability to biodiversity will increase.
- Barcode-type data are useful in a diversity-diagnostic sense.
- Metagenomic tools allow a broad diversity assessment in a single test.
- We may need to consider accepting the inferred occurrence of an organism at a point in space time, based on molecular data alone.

The Basics:

- Geo-referenced genetic data can provide evidence of patterns of origin and distribution of new genes.

The Basics:

- Geo-referenced genetic data can provide evidence of patterns of origin and distribution of new genes.
- Should species-abundance maps be extended to include geno-clines?

The Basics:

- Geo-referenced genetic data can provide evidence of patterns of origin and distribution of new genes.
- Should species-abundance maps be extended to include geno-clines?
- and so on...

Metagenomics reveals biological dark matter.

Two trees / two super-niches / one biosphere.

Different Trees

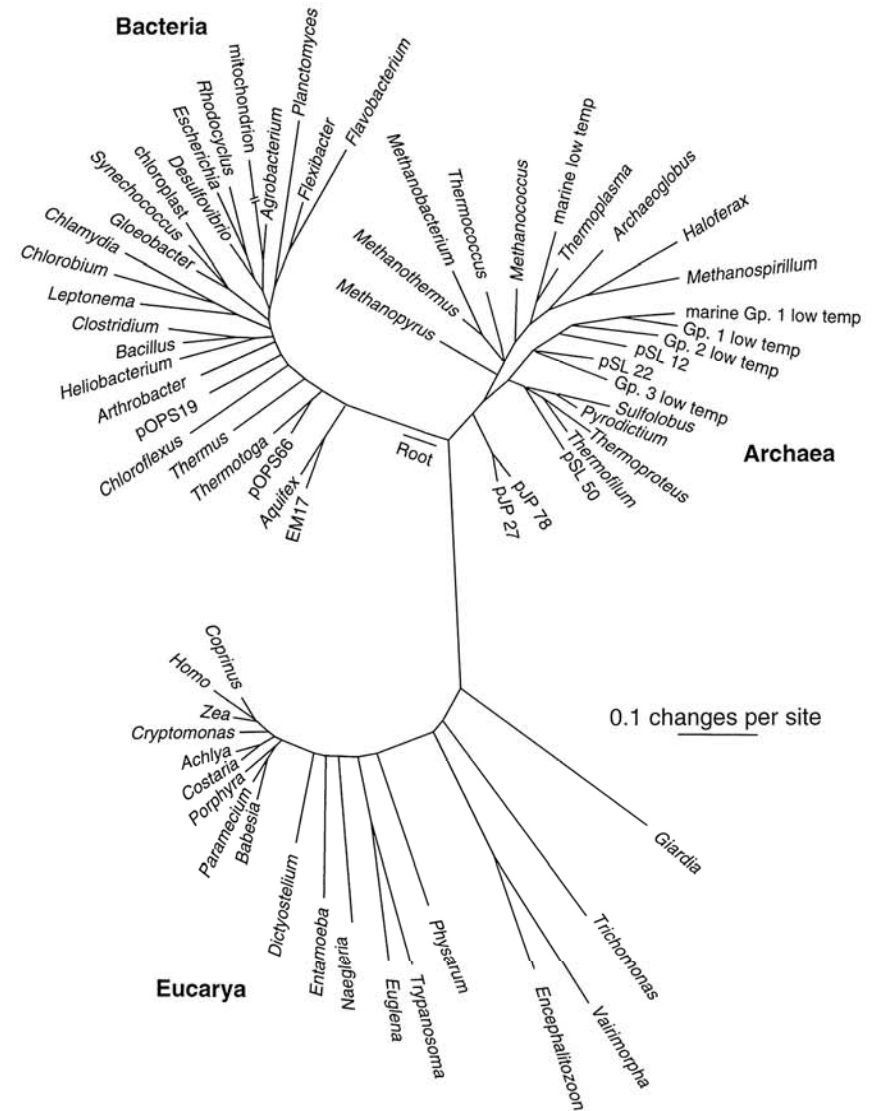


A fairly familiar-looking tree, albeit drawn a little more metaphorically than precisely.

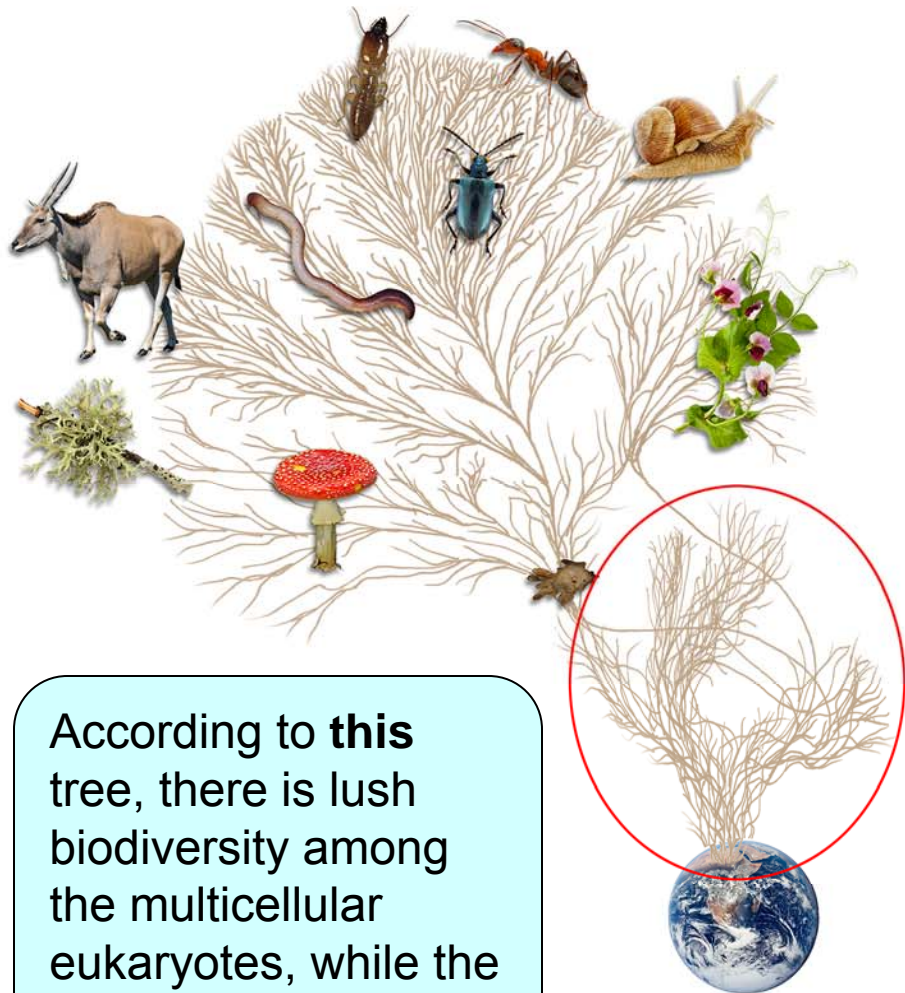
Different Trees

Another tree, this one based on comparisons of rRNA short sub-unit sequences.

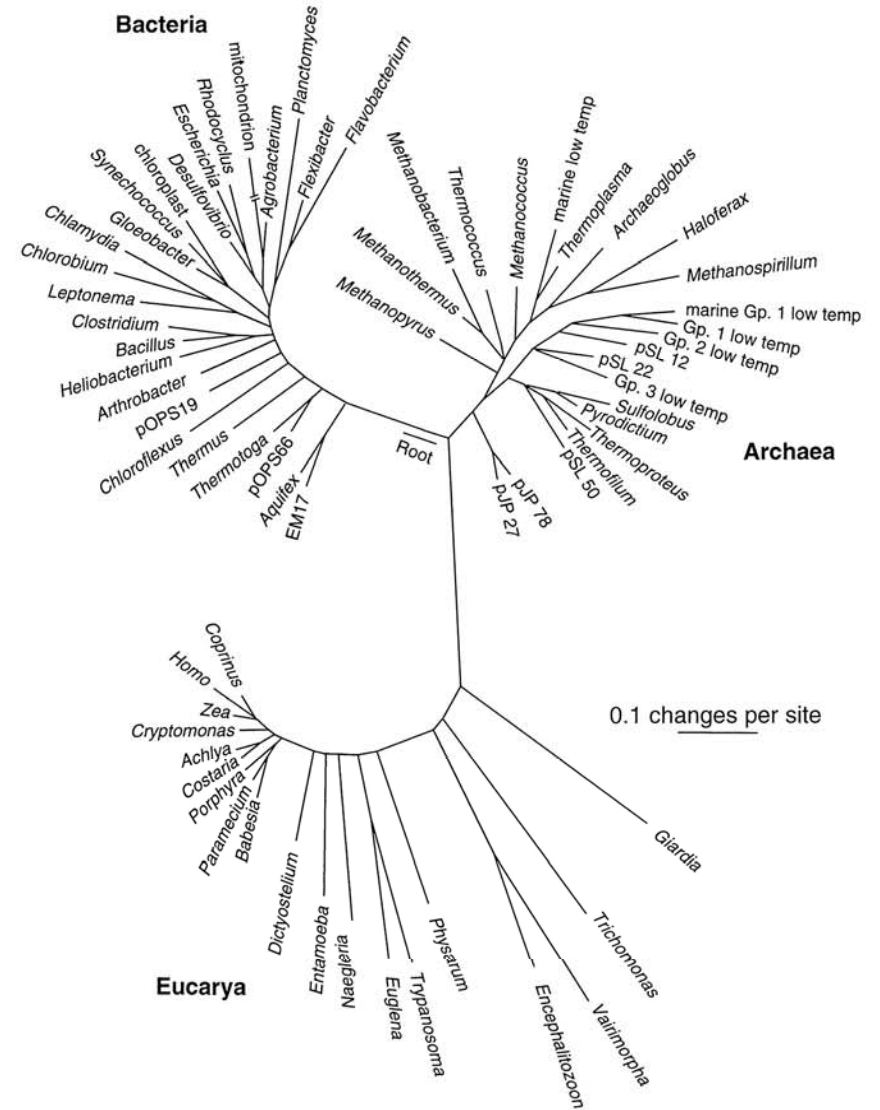
Branch length reflects actual divergence of sequence.



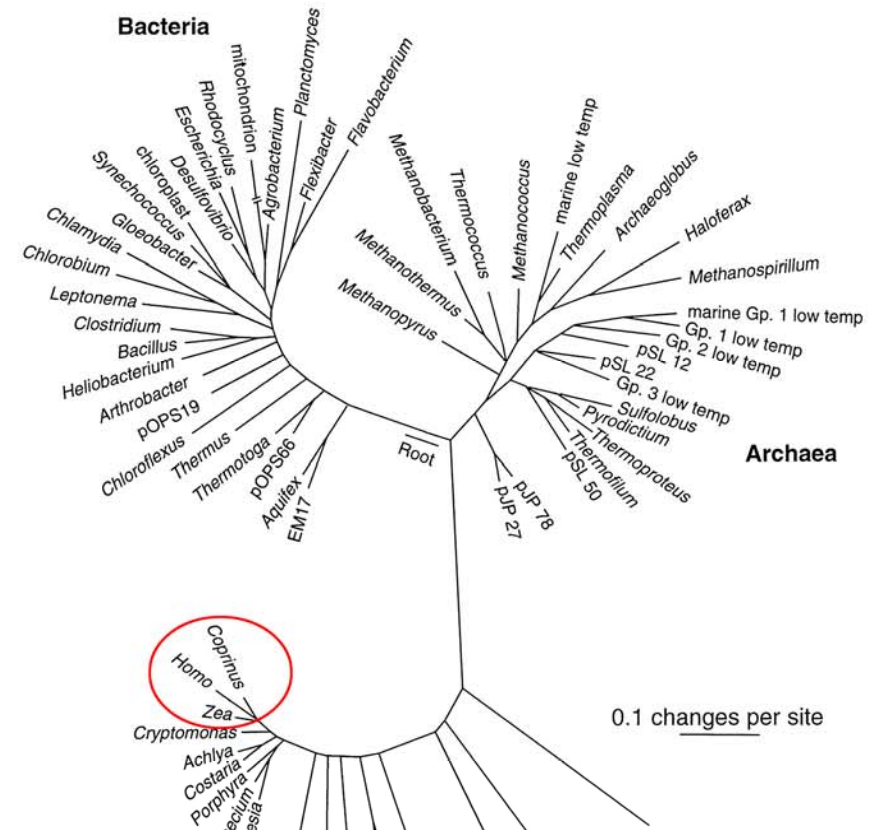
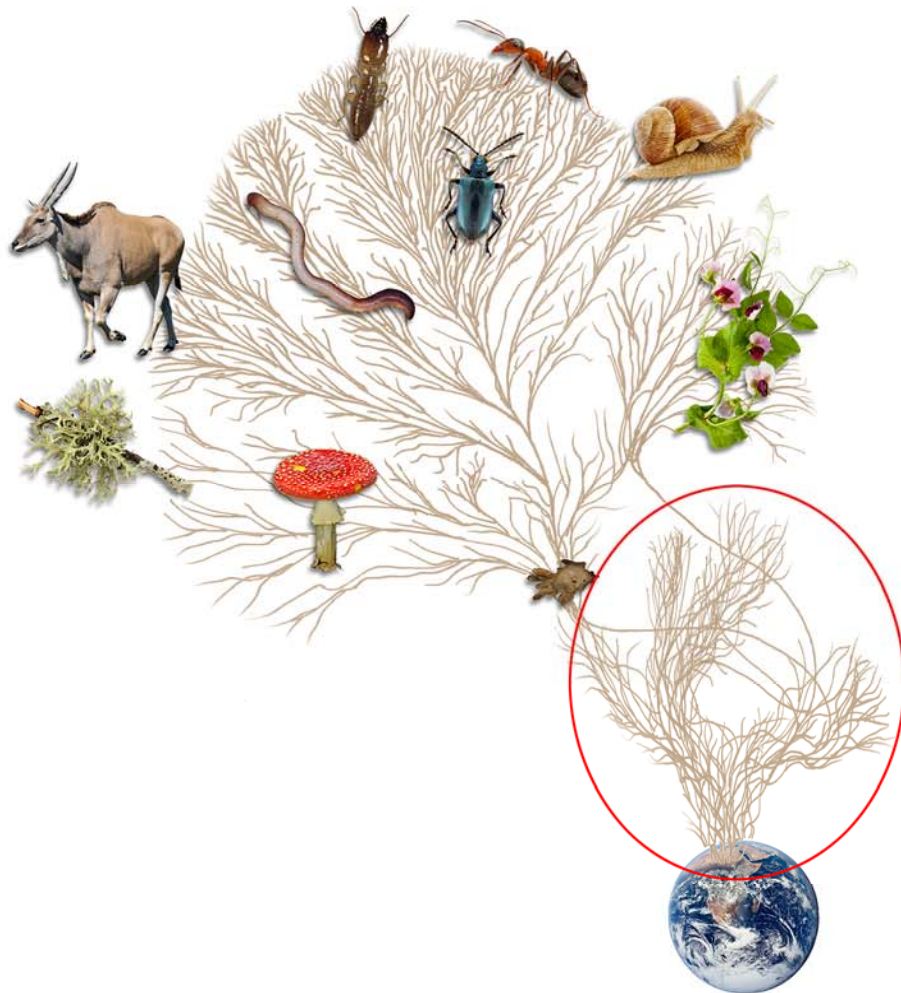
Different Trees



According to **this** tree, there is lush biodiversity among the multicellular eukaryotes, while the microbes seem to be pretty much the same.



Different Trees

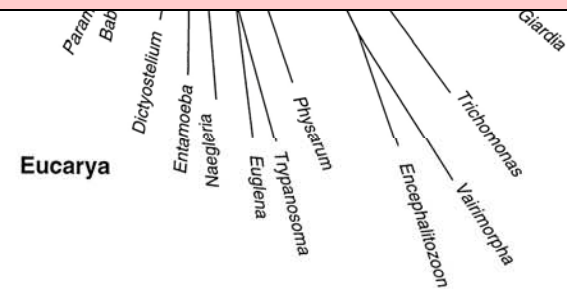
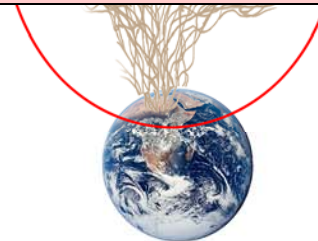


According to **this** tree, there is tremendous biodiversity among the prokaryotes, while the differences among the multicellular eukaryotes (MCEs) barely qualify as variations on a theme.

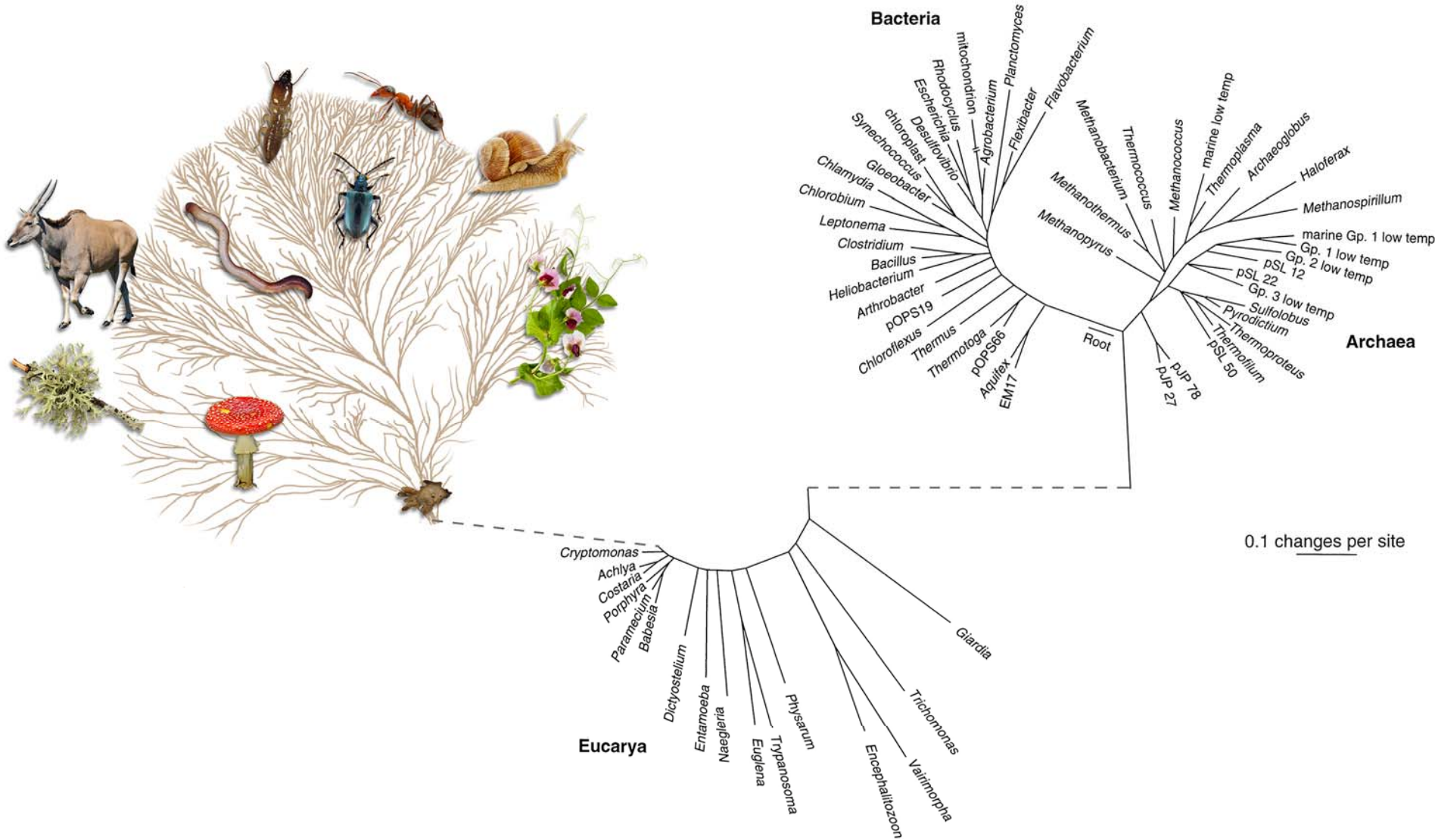
Different Trees



Surely these two trees cannot both be right. At least one must be a significant misrepresentation.

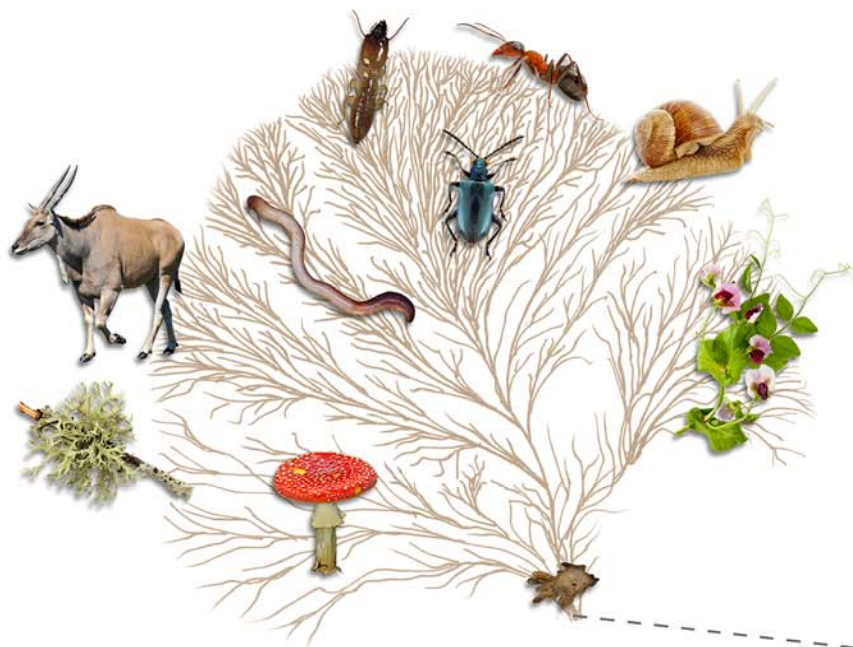


Different Mega-Niches

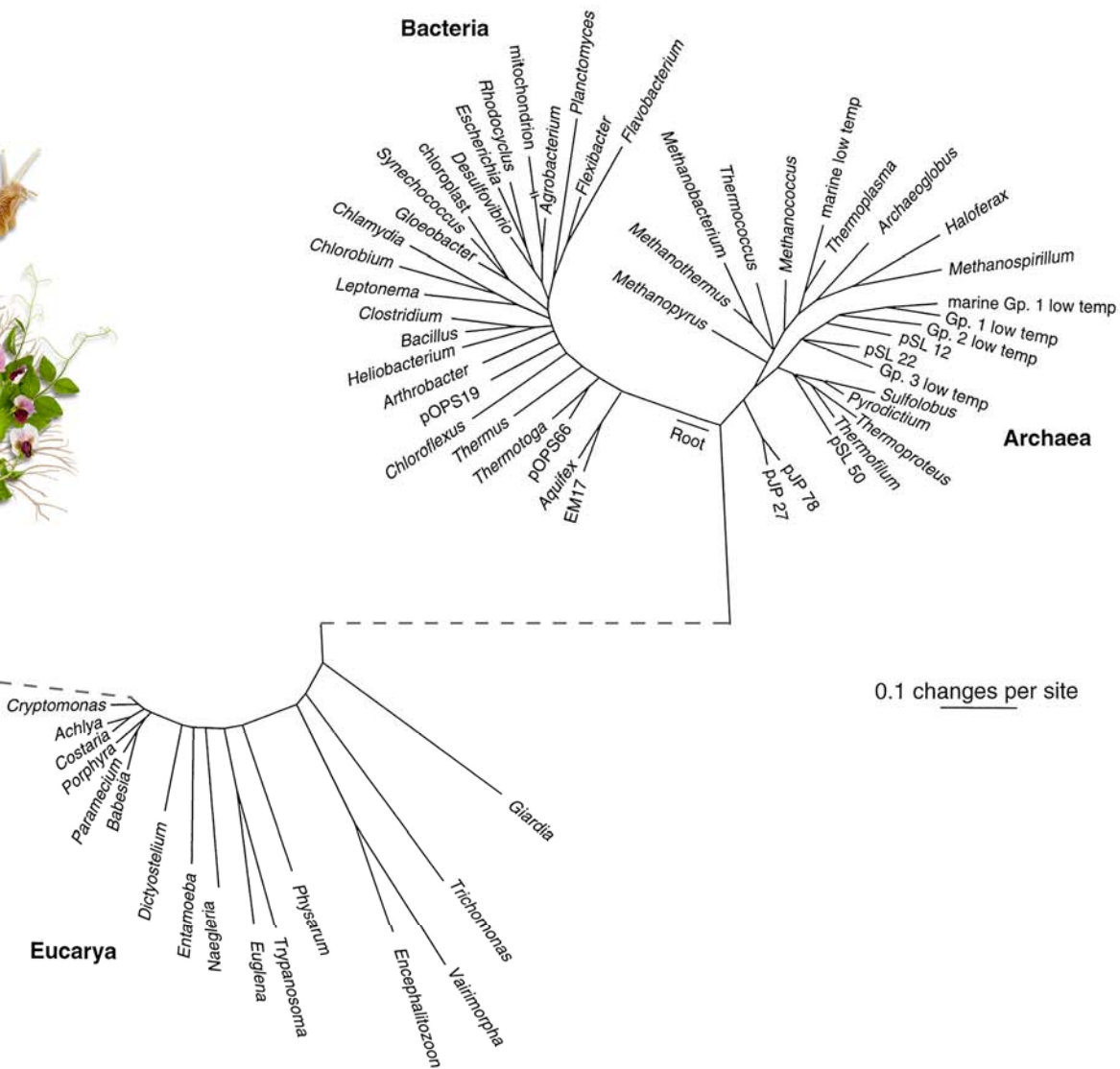


Different Mega-Niches

Morphological Mega-Niche

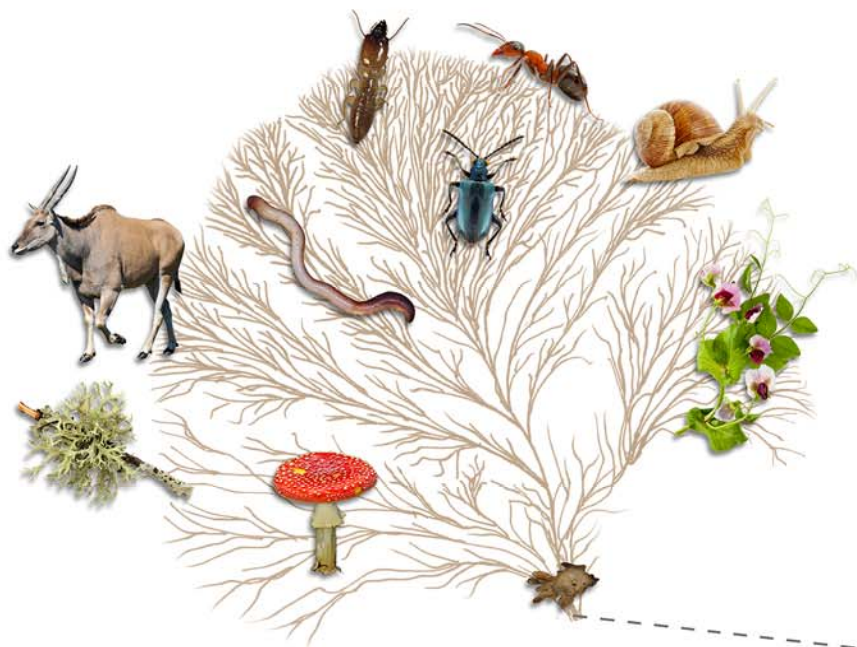


Diversity among multi-cellular eukaryotes (MCEs) is largely morphological. From a morphological perspective, all prokaryotes look pretty much alike.



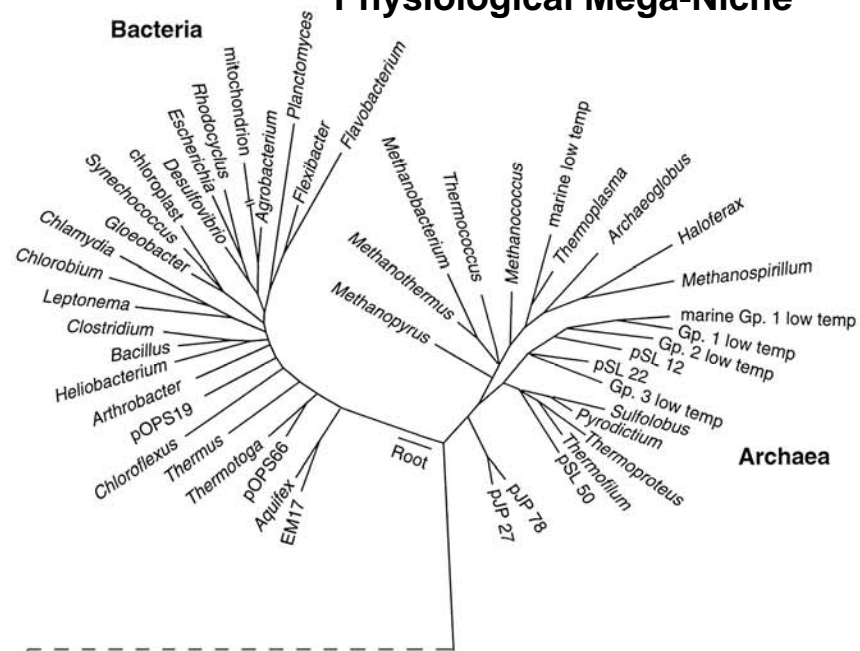
Different Mega-Niches

Morphological Mega-Niche



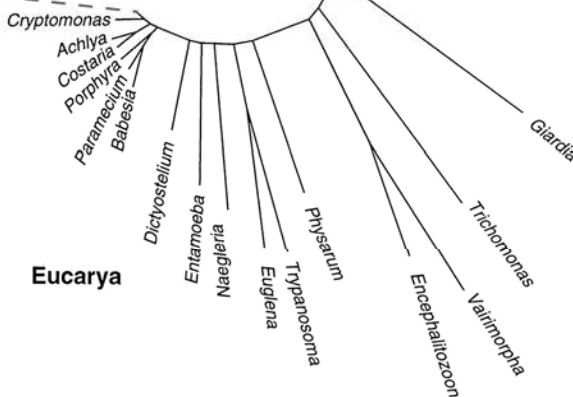
Diversity among multi-cellular eukaryotes (MCEs) is largely morphological. From a morphological perspective, all prokaryotes look pretty much alike.

Physiological Mega-Niche



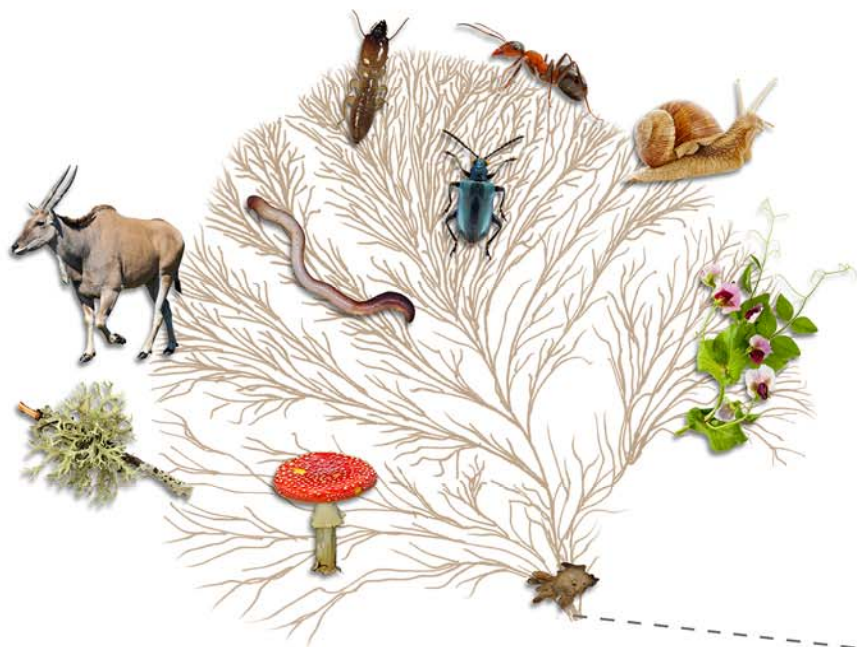
0.1 changes per site

Diversity among prokaryotes is largely physiological. From a physiological perspective, all MCEs are in fact just variations on a theme.



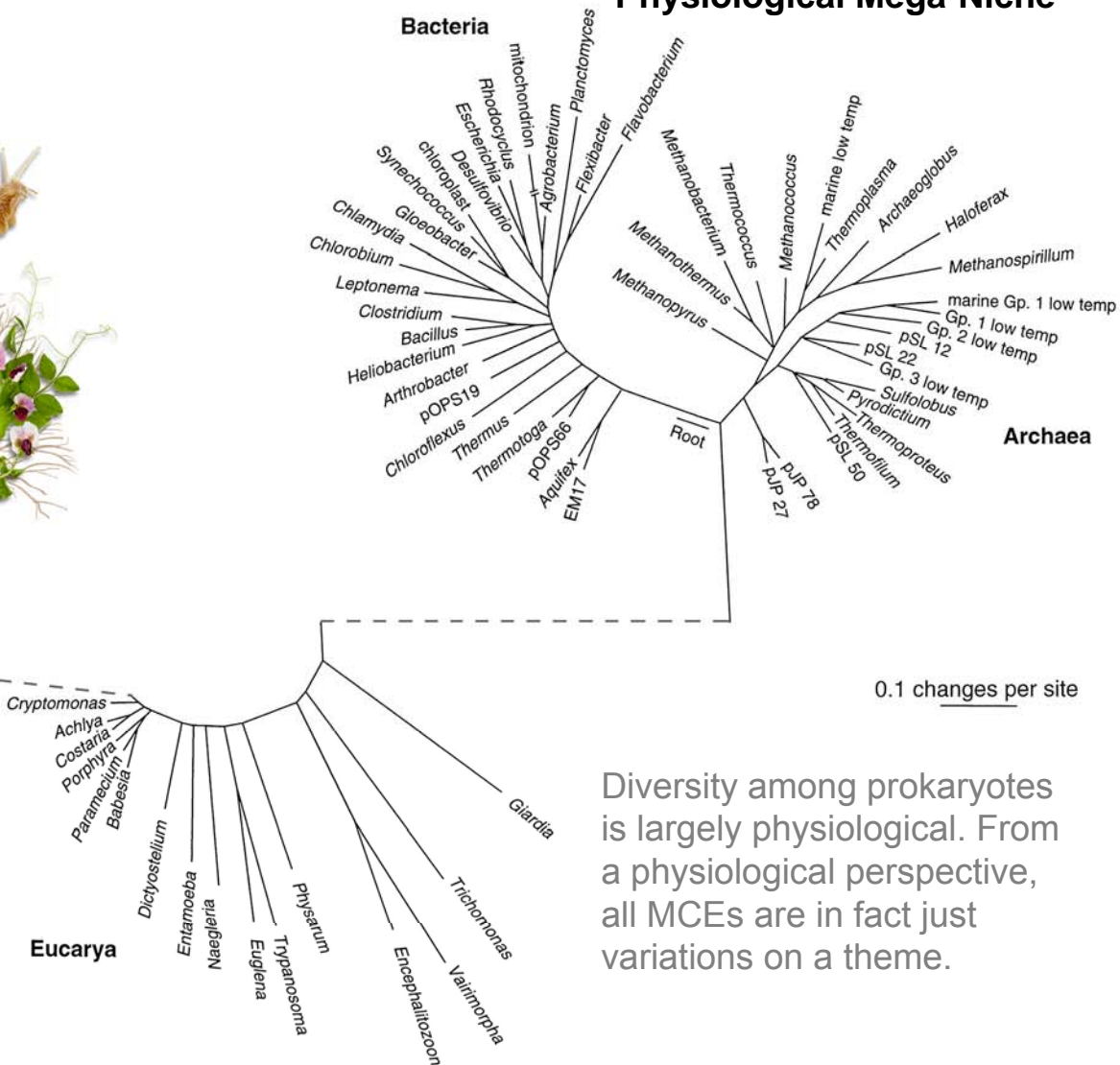
Different Mega-Niches

Morphological Mega-Niche



Diversity among multi-cellular eukaryotes (MCEs) is largely morphological. From a morphological perspective, all prokaryotes look pretty much alike.

Physiological Mega-Niche



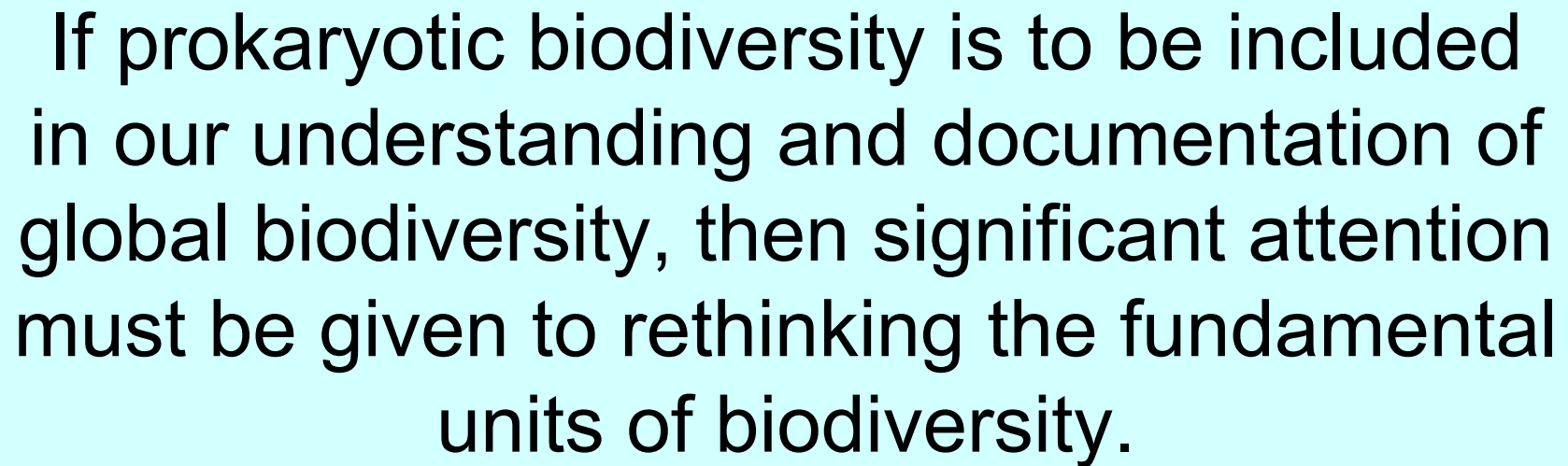
Diversity among prokaryotes is largely physiological. From a physiological perspective, all MCEs are in fact just variations on a theme.

Developing a model of biodiversity that includes both will require a substantial rethink. Prokaryotes are profoundly different from MCEs. Many “fundamental” concepts, based on the study of MCEs, simply do not apply to prokaryotes. The relationship is similar to that between classical and quantum physics.

Different Mega-Niches

Morphological Mega-Niche

Physiological Mega-Niche



If prokaryotic biodiversity is to be included in our understanding and documentation of global biodiversity, then significant attention must be given to rethinking the fundamental units of biodiversity.

Developing a model of biodiversity that includes both will require a substantial rethink. Prokaryotes are profoundly different from MCEs. Many “fundamental” concepts, based on the study of MCEs, simply do not apply to prokaryotes. The relationship is similar to that between classical and quantum physics.

Metagenomics reveals biological dark matter.

Biological dark matter is weird.

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

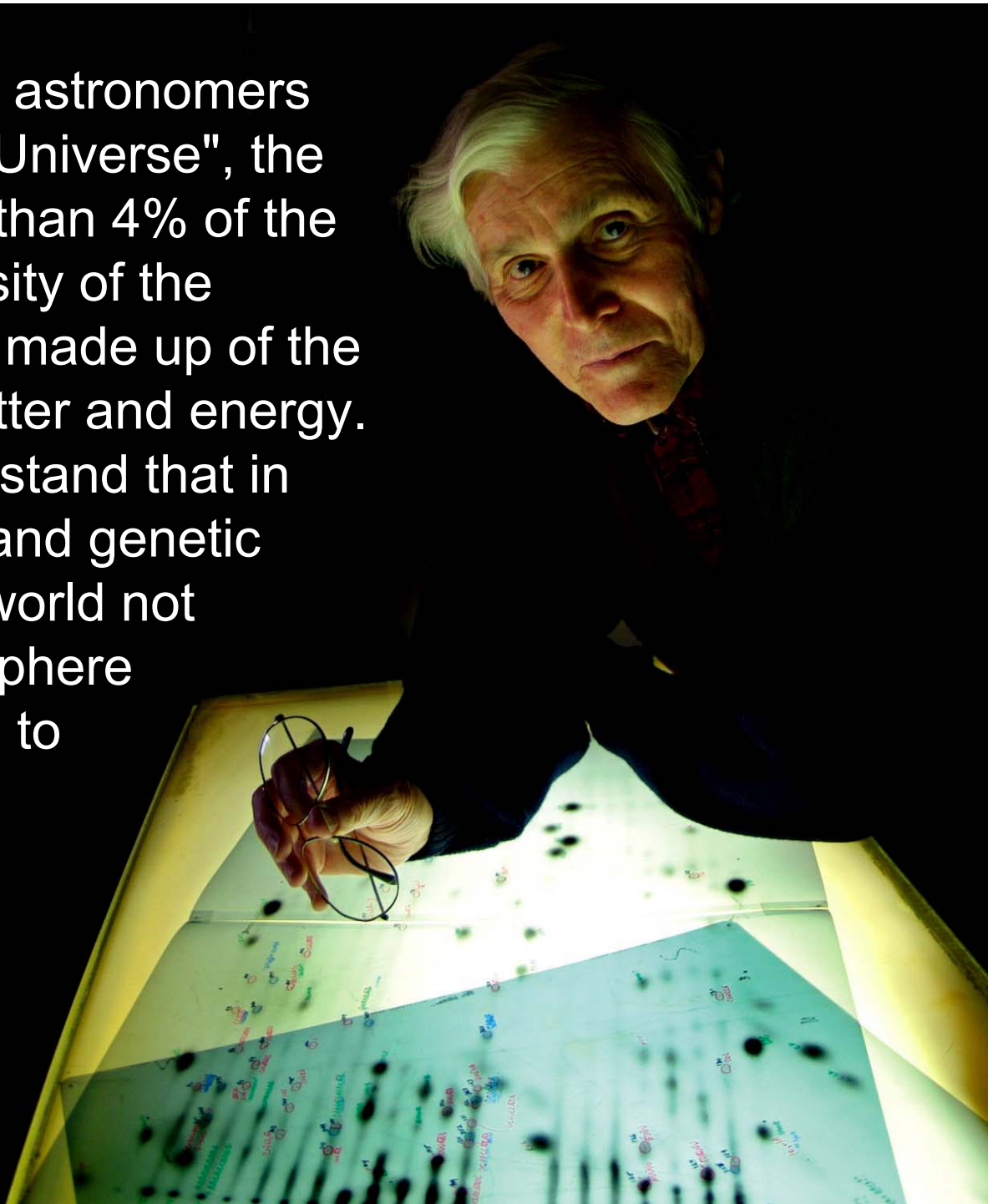
- Metagenomics tools are showing that, compared to macro-scale organisms, the diversity of microbial communities is staggering.

Intra-species bacterial genetic diversity is greater than that among the great apes; intra-genus bacterial diversity is greater than that among all the mammalia.

- Metagenomics tools are showing that a full understanding of macro-scale organisms will depend on an understanding of their interactions with their associated microbiomes.

Understanding how different ants optimize nutrient acquisition, and thus how they function in their niches, depends on an understanding of their associated gut microbiomes.

We now know that what astronomers used to think of as "the Universe", the visible universe, is less than 4% of the total matter/energy density of the universe, the rest being made up of the still mysterious dark matter and energy. Similarly, we now understand that in terms of both numbers and genetic diversity, the microbial world not only dominates the biosphere but is almost impossible to sample properly.





We now know that what astronomers used to think of as "the Universe", the visible universe, is less than 4% of the total matter/energy density of the universe, the rest being made up of the still mysterious dark matter and energy. Similarly, we now understand that in terms of both numbers and genetic

Understanding this biological dark matter must be a top goal for 21st century biology. In the 20th century we found that classical physics was only an approximation of reality – an incredibly useful approximation, but an approximation nonetheless. So, too, with classical biology.

Prokaryotes operate on a time
scale that is markedly different
than MCEs

Parlor game: fill in the blanks:

Since _____

there have been as many bacterial generations
as there would have been human generations

since _____

Parlor game: fill in the blanks:

Since ***the beginning of this month***

there have been as many bacterial generations
as there would have been human generations

since ***the end of the ice age.***

Parlor game: fill in the blanks:

Since *July 4th*

there have been as many bacterial generations
as there would have been human generations

since *Neanderthals were common in Europe.*

Parlor game: fill in the blanks:

Since ***TDWG in Beijing***

there have been as many bacterial generations
as there would have been human generations

since ***the first Neanderthals in Europe.***

Parlor game: fill in the blanks:

Since ***Bush invaded Iraq (2003)***

there have been as many bacterial generations
as there would have been human generations

since ***the appearance of the first hominids.***

Parlor game: fill in the blanks:

Since ***Reagan was elected president (1980)***

there have been as many bacterial generations
as there would have been human generations

since ***the divergence of chimps and humans.***

Parlor game: fill in the blanks:

Since ***Robert Hooke drew the first cell (1643)***

there have been as many bacterial generations
as there would have been human generations

since ***the K-T boundary.***

Parlor game: fill in the blanks:

Since *Tutankhamen was pharoah*

there have been as many bacterial generations
as there would have been human generations

since *the beginning of the Cambrian.*

Prokaryotes have a
pan-genome

Pan-genome:

- Prokaryotes have highly flexible genomes: of all of the genes ever found in some *E. coli*, no more than 60% are ever found in one *E. coli*.

Pan-genome:

- Prokaryotes have highly flexible genomes: of all of the genes ever found in some *E. coli*, no more than 60% are ever found in one *E. coli*.
- Prokaryotes acquire and lose “hereditary” material throughout their lifetime.

Pan-genome:

- Prokaryotes have highly flexible genomes: of all of the genes ever found in some *E. coli*, no more than 60% are ever found in one *E. coli*.
- Prokaryotes acquire and lose “hereditary” material throughout their lifetime.
- Some “optional” genes seem to occur in any species in a particular environment.

Pan-genome:

- Prokaryotes have highly flexible genomes: of all of the genes ever found in some *E. coli*, no more than 60% are ever found in one *E. coli*.
- Prokaryotes acquire and lose “hereditary” material throughout their lifetime.
- Some “optional” genes seem to occur in any species in a particular environment.
- In prokaryotes, reproduction and inheritance are decoupled.

The challenge for biodiversity informatics.

Classical biodiversity contains several key assumptions about the fundamental nature of biological systems.

MCE Assumptions:

- Organisms maintain stable genomes.
- Reproduction and inheritance are completely coupled.
- The biosphere can be divided up into populations of interbreeding individuals.
- Because all genes in a genome are inherited at once, all genes in a species share the same phylogenetic history.
- Individual organisms are multi-cellular aggregates of differentiated somatic cells.

The challenge for biodiversity informatics.

Virtually none of those fundamental assumptions apply to biological dark matter.

The vanishing individual.

Microbial ecology and prokaryotic systems.

The vanishing individual.

Pervasive symbiosis (we are all lichens, now).

The traditional biologist's view of life, first stated by Aristotle, starts by looking at individual organisms and asking what properties they have in common. (p 5)

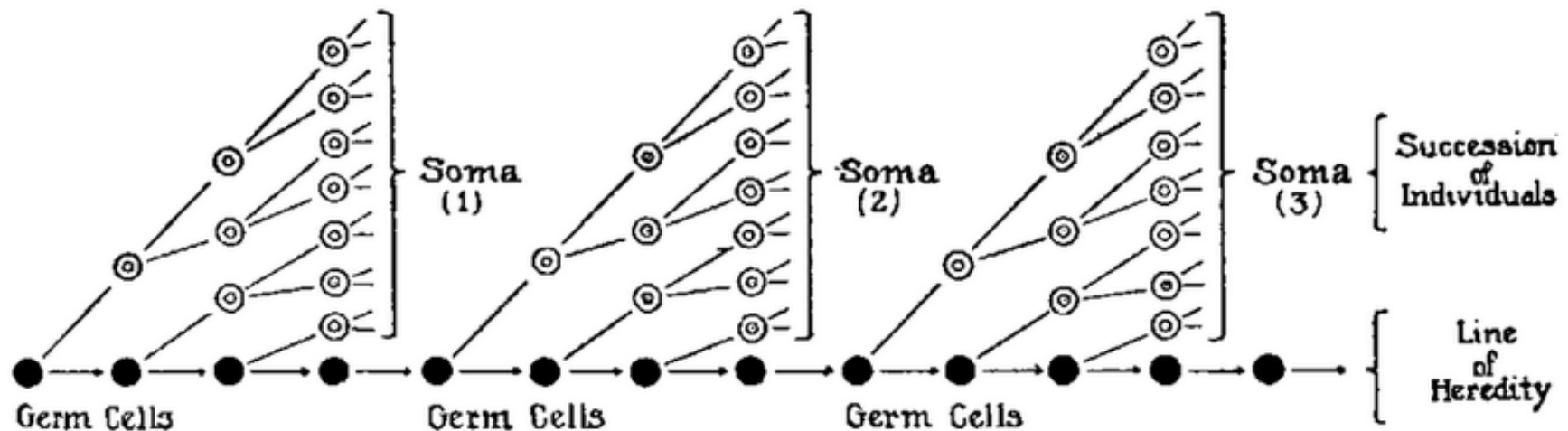
Morowitz, Harold J. 1992. *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis*. New Haven: Yale University Press.

The subjects of classification are organisms and the subjects of taxonomy are classifications. (p11)

It seems obvious ... that the real unit in nature, the one thing that is usually completely objective in spite of some marginal cases, is the individual organism. (p. 18)

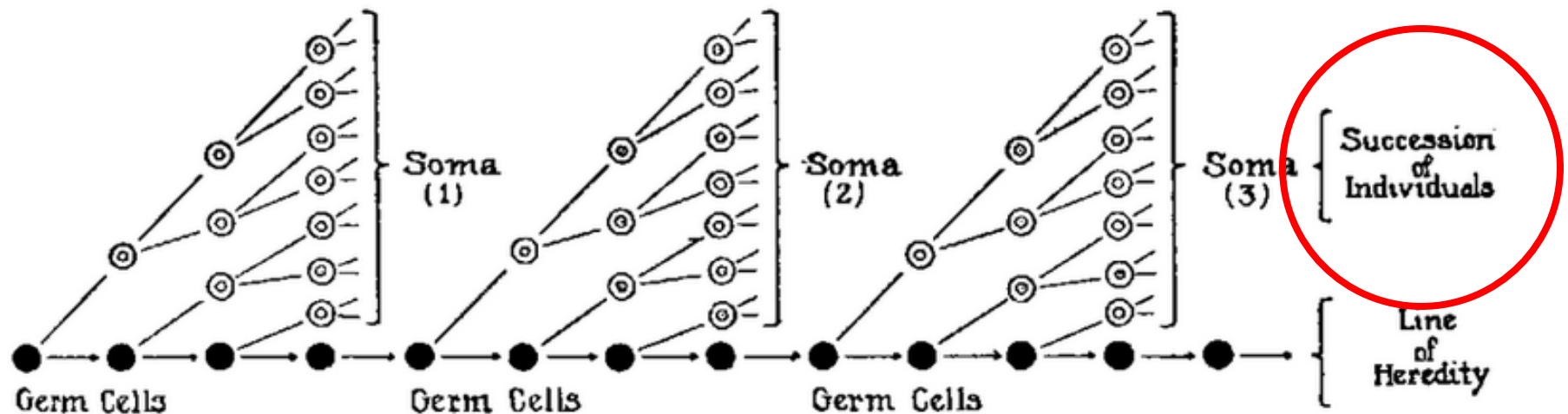
George Gaylord Simpson. 1961. *Principles of Animal Taxonomy*. New York: Columbia University Press.

Weismann's Germ Plasm Theory



Organisms that best satisfy the notion of exhibiting “completely objective” individuals are animals that follow a Weismannian pattern of development – that is, an early sequestering of a separate germ line, with a complete logical and physical separation of somatic and germ tissue. Such animals begin as a zygote, then develop mitotically into a multicellular adult that, with luck, lives to adulthood and reproduces via the meiotic production of gametes.

Weismann's Germ Plasm Theory



Organisms that best satisfy the notion of exhibiting “completely objective” individuals are animals that follow a Weismannian pattern of development – that is, an early sequestering of a separate germ line, with a complete logical and physical separation of somatic and germ tissue. Such animals begin as a zygote, then develop mitotically into a multicellular adult that, with luck, lives to adulthood and reproduces via the meiotic production of gametes.

Weismann's Germ Plasm Theory

On the Weismannian view, differentiated somatic cells make up the **individual**.

The germ cells represent an independent line of potentially immortal, but non-differentiated cells that carry hereditary information.

line, with a complete logical and physical separation of somatic and germ tissue. Such animals begin as a zygote, then develop mitotically into a multicellular adult that, with luck, lives to adulthood and reproduces via the meiotic production of gametes.

Weismann's Germ Plasm Theory

On the Weismannian view, differentiated somatic cells make up the **individual**.

The germ cells represent an independent line of potentially immortal, but non-differentiated cells that carry hereditary information.

Prokaryotic organisms have no soma.

line, with a complete logical and physical separation of somatic and germ tissue. Such animals begin as a zygote, then develop mitotically into a multicellular adult that, with luck, lives to adulthood and reproduces via the meiotic production of gametes.

The subjects of classification are organisms and the subjects of taxonomy are classifications. (p11)

It seems obvious ... that the real unit in nature, the one thing that is usually completely objective **in spite of some marginal cases**, is the individual organism. (p. 18)

George Gaylord Simpson. 1961. *Principles of Animal Taxonomy*. New York: Columbia University Press.

The subjects of classification are
organisms. Taxonomy
are classified.

Not very many...

It seems obvious ... that the real unit in
nature, the one thing that is usually
completely objective in spite of **some**
marginal cases, is the individual
organism. (p. 18)

George Gaylord Simpson. 1961. *Principles of Animal Taxonomy*. New
York: Columbia University Press.

The subjects of classification are
organisms and their taxonomy
are classified

Not very important...

It seems obvious ... that the real unit in
nature, the one thing that is usually
completely objective in spite of some
marginal cases, is the individual
organism. (p. 18)

George Gaylord Simpson. 1961. *Principles of Animal Taxonomy*. New
York: Columbia University Press.

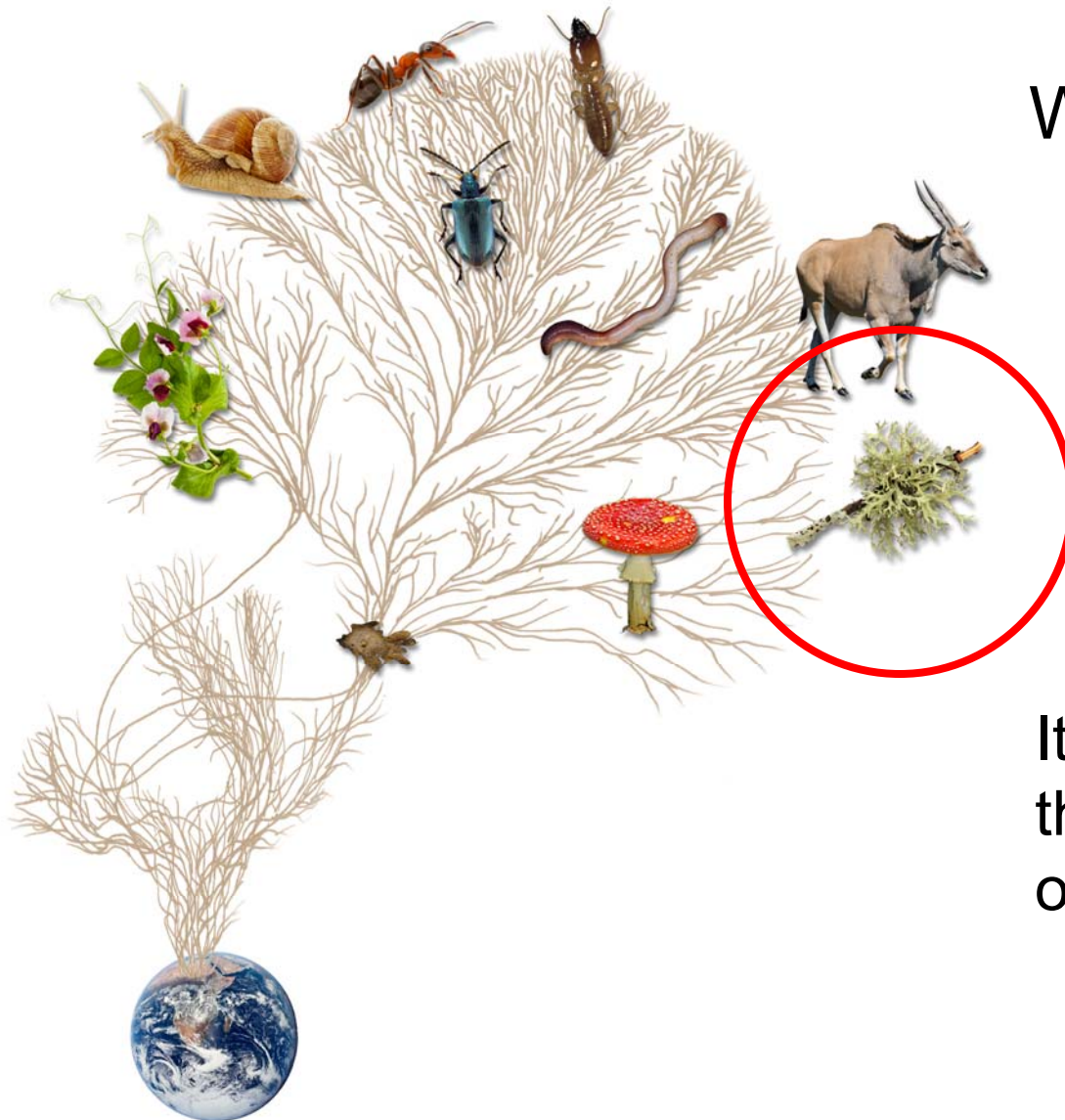
The su
organis
are cla
It seem
nature, the one thing that is usually
completely objective in spite of some
marginal cases, is the individual
organism. (p. 18)

Without doubt...
Indubitably...
Really ? ! ?

onomy
unit in

George Gaylord Simpson. 1961. *Principles of Animal Taxonomy*. New York: Columbia University Press.

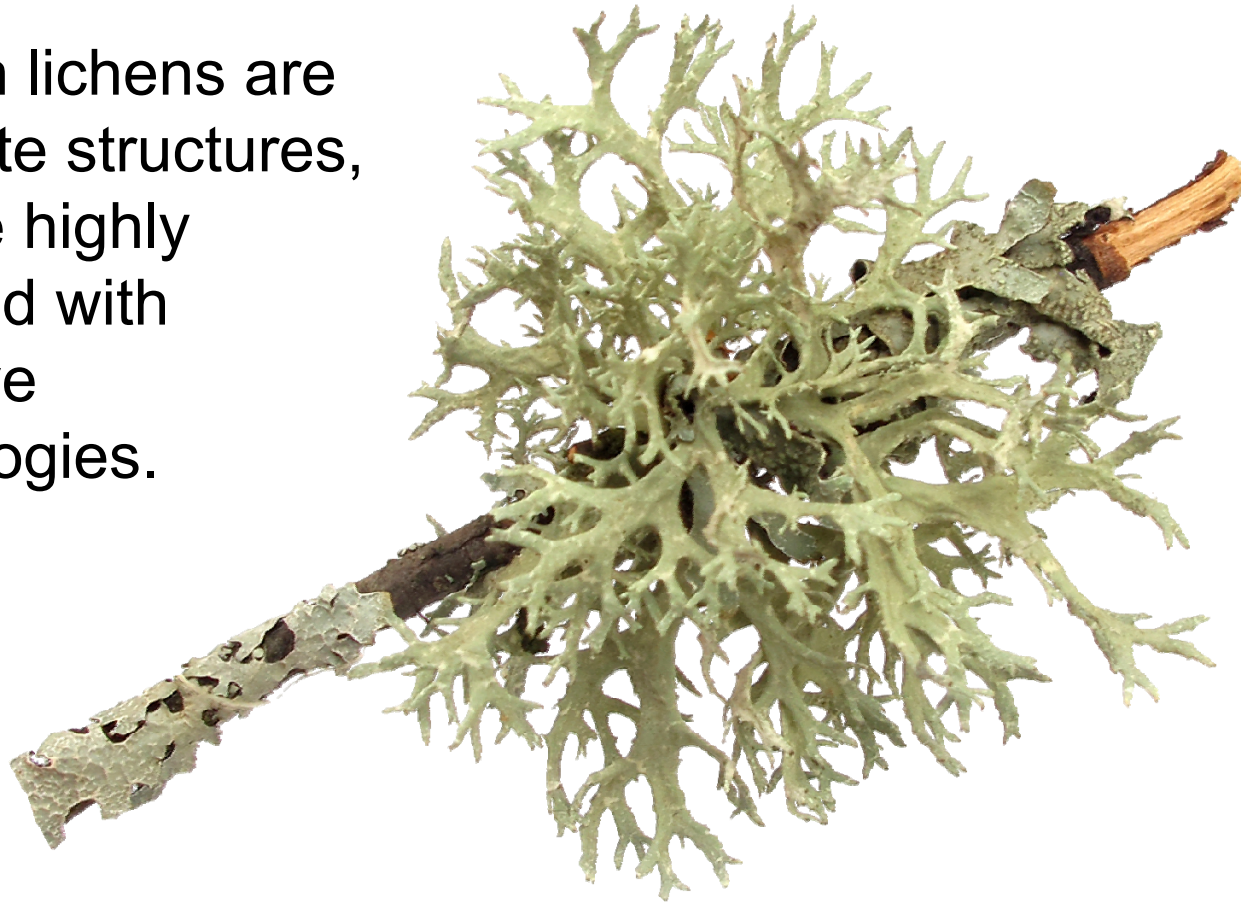
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



What about lichens?

It is not easy to apply
the concept of individual
organism to a lichen.

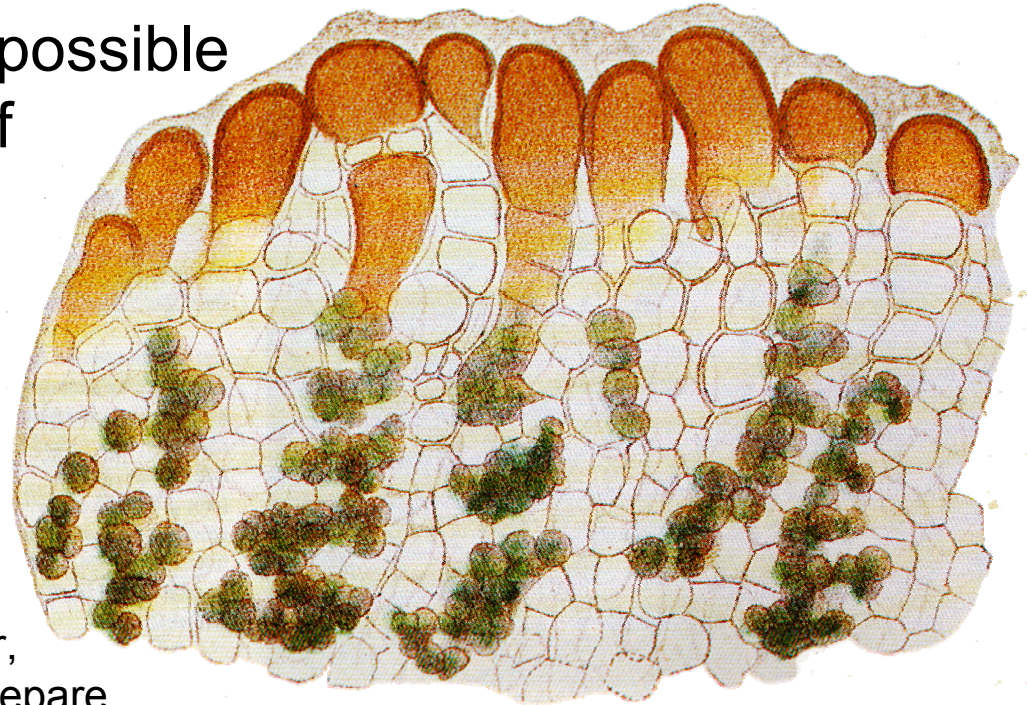
Although lichens are composite structures, most are highly organized with distinctive morphologies.



Without a detailed microscopical examination, most lichens appear to be single entities.

At the cellular level, it is possible to detect the presence of algae, embedded in the mycobiont tissue.

[L]ichens are not simple plants, not ordinary individuals in the ordinary sense of the word; they are, rather, colonies, which consist of hundreds of thousands of individuals, of which, however, one alone plays the master, while the rest, forever imprisoned, prepare the nutriment for themselves and their master. This fungus is a fungus of the class Ascomycetes, a parasite which is accustomed to live upon others' work. Its slaves are green algae, which it has sought out, or indeed caught hold of, and compelled into its service. It surrounds them as a spider its prey, with a fibrous net of narrow meshes, which is gradually converted into an impenetrable covering, but while the spider sucks its prey and leaves it dead, the fungus incites the algae found in its net to more rapid activity, even to more vigorous increase.



Simon Schwendener (1869)

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

At the cellular level, it is possible to detect the presence of algae, embedded in the mycobiont tissue.



Schwendener's prose is dramatic, but makes the key point: a lichen is a living unit, but is not an "individual" as classically conceived. Nor can it be decomposed into individuals without giving up both its essence and its viability.

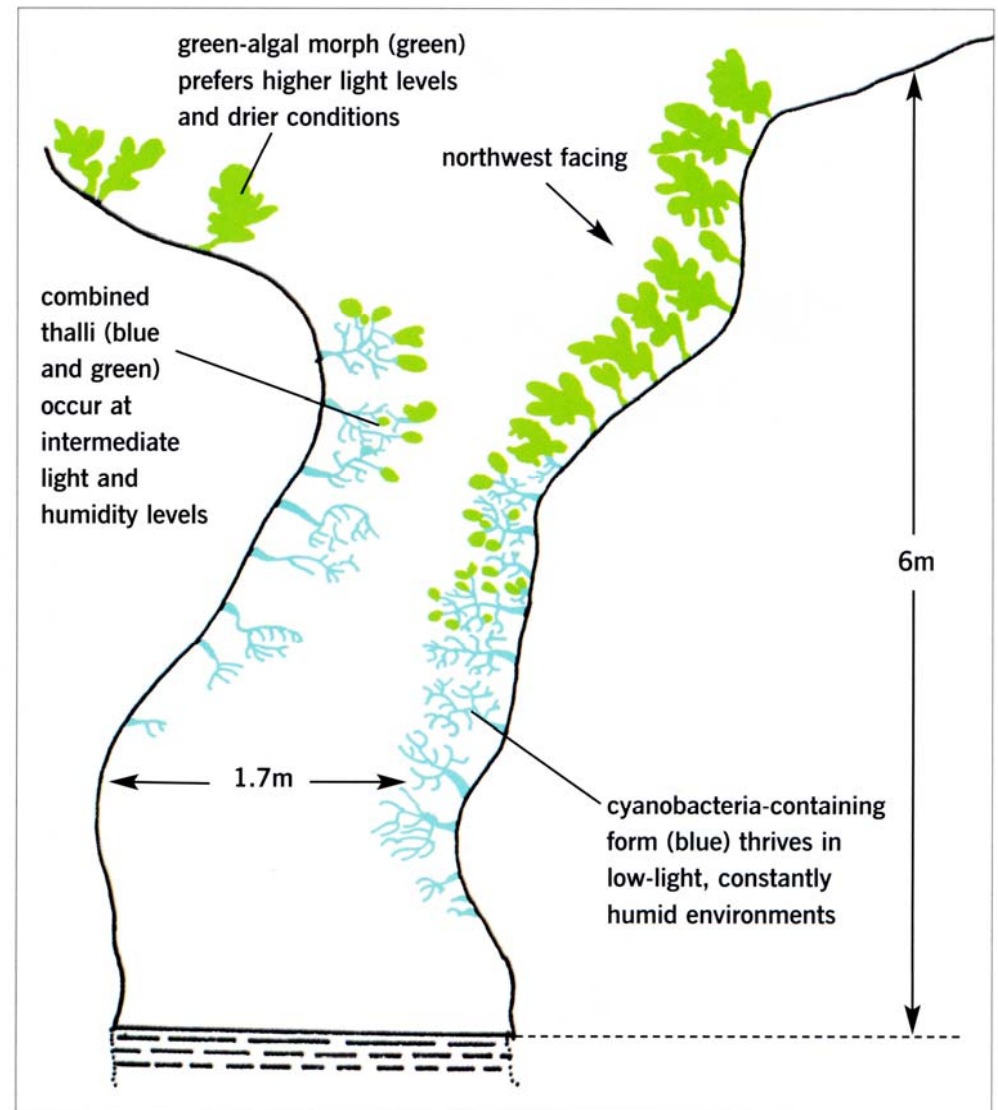
A lichen is a composite organism that cannot be subdivided into "individuals" and remain living. How does this square with the idea of the "individual" as the true "fundamental unit" of nature?

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

The same fungal partner can combine with different algal partners, to produce phenotypically distinct lichens, that occupy different niches.

In the right circumstances, they may be found in a gradient.

Are these two different lichens, or ...



RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

The same fungal partner can combine with different algal partners, to produce phenotypically distinct lichens,

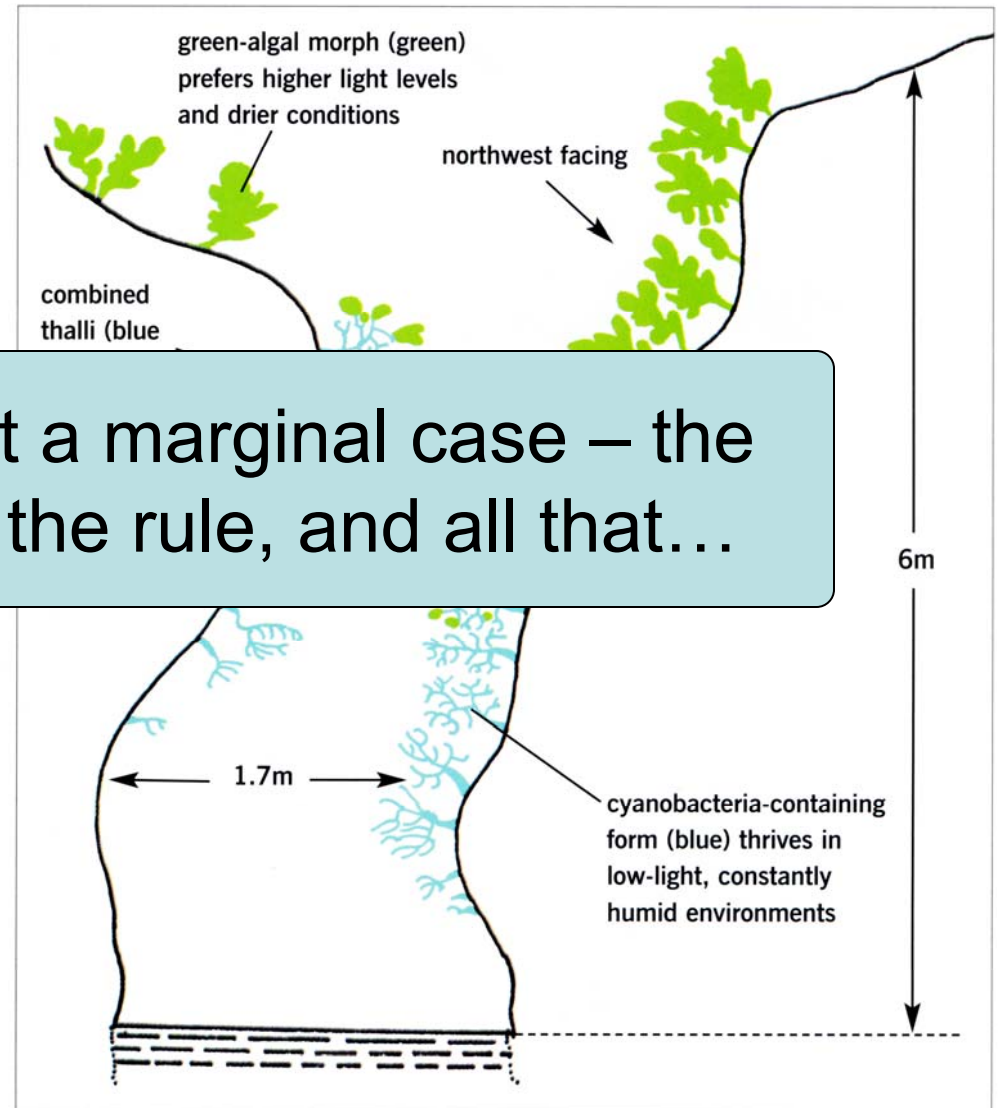
that

In the

they may be found in a

gradient.

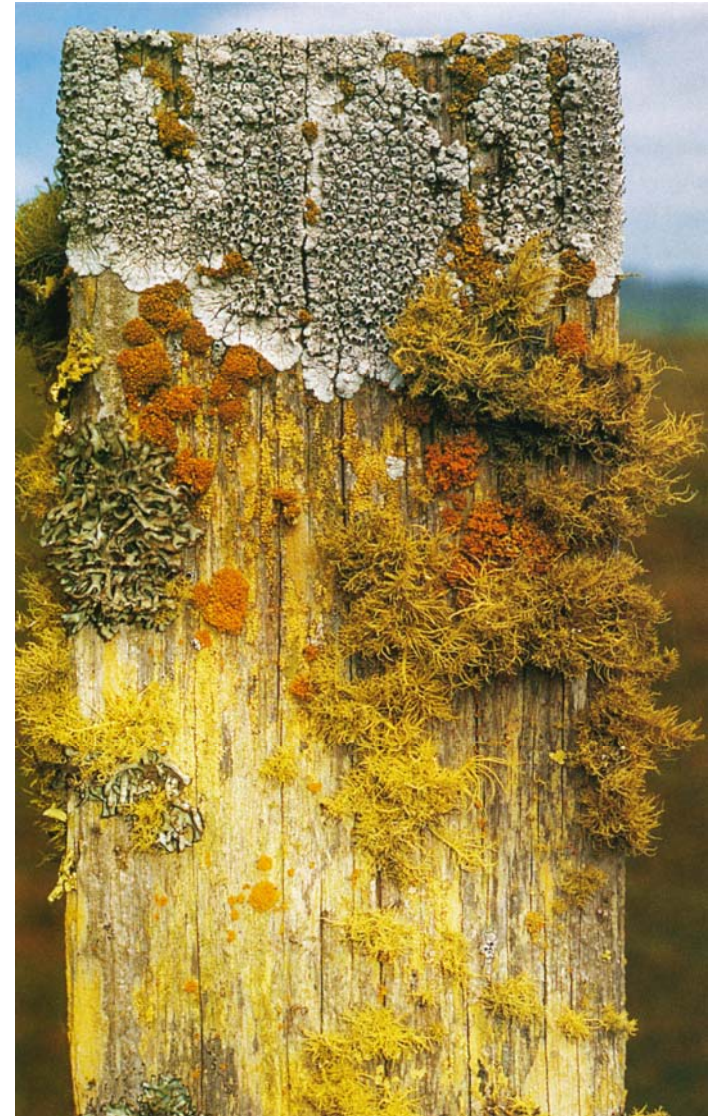
Are these two different lichens, or ...



Maybe lichens are just a marginal case – the exception that proves the rule, and all that...

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

But they are certainly not rare...



RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



What about termites?

They are a critically important, sometimes dominant, species in many ecosystems, yet they cannot exist without their gut symbionts.

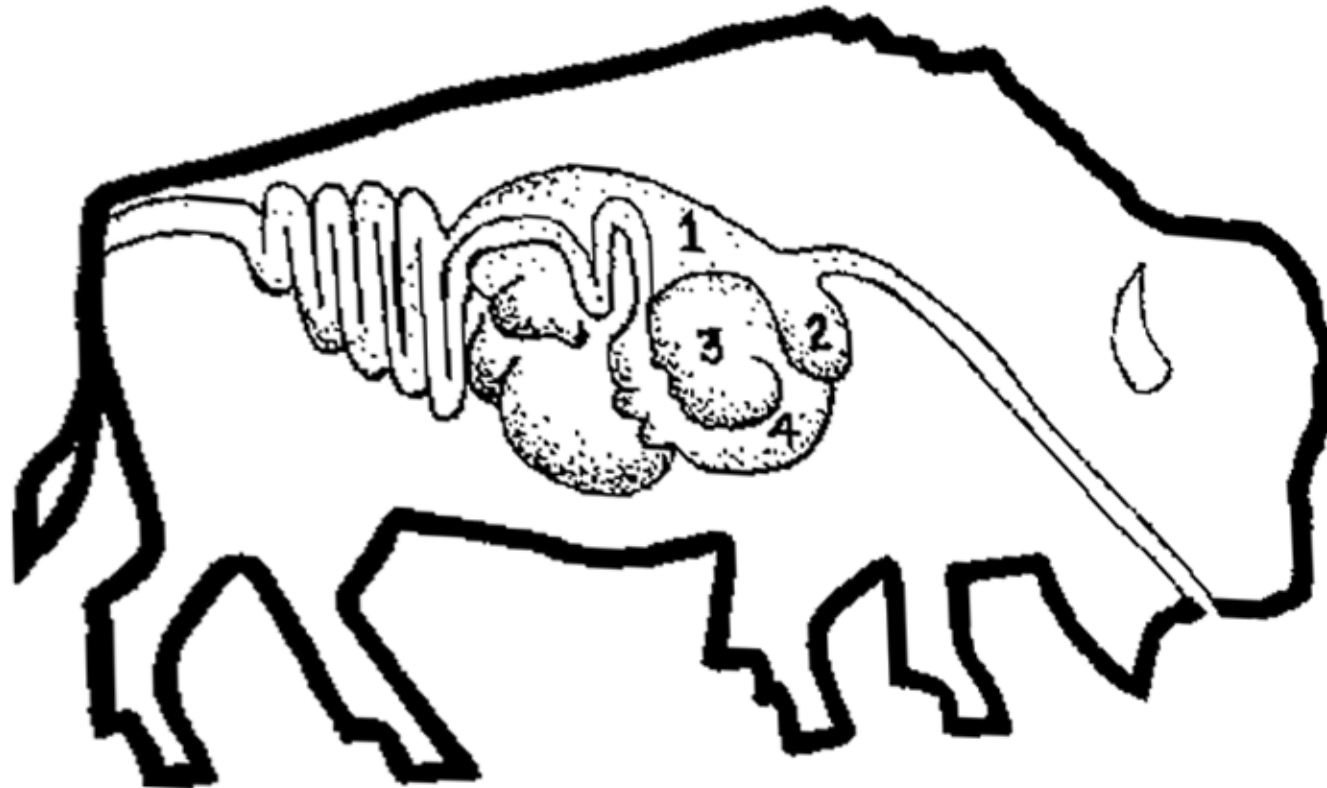
Can we really dismiss termites as a rare, marginal case?

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



How about mammals? Surely they exemplify the idealized rugged individualism of autonomous organisms – the fundamental (and completely objective) unit of nature...

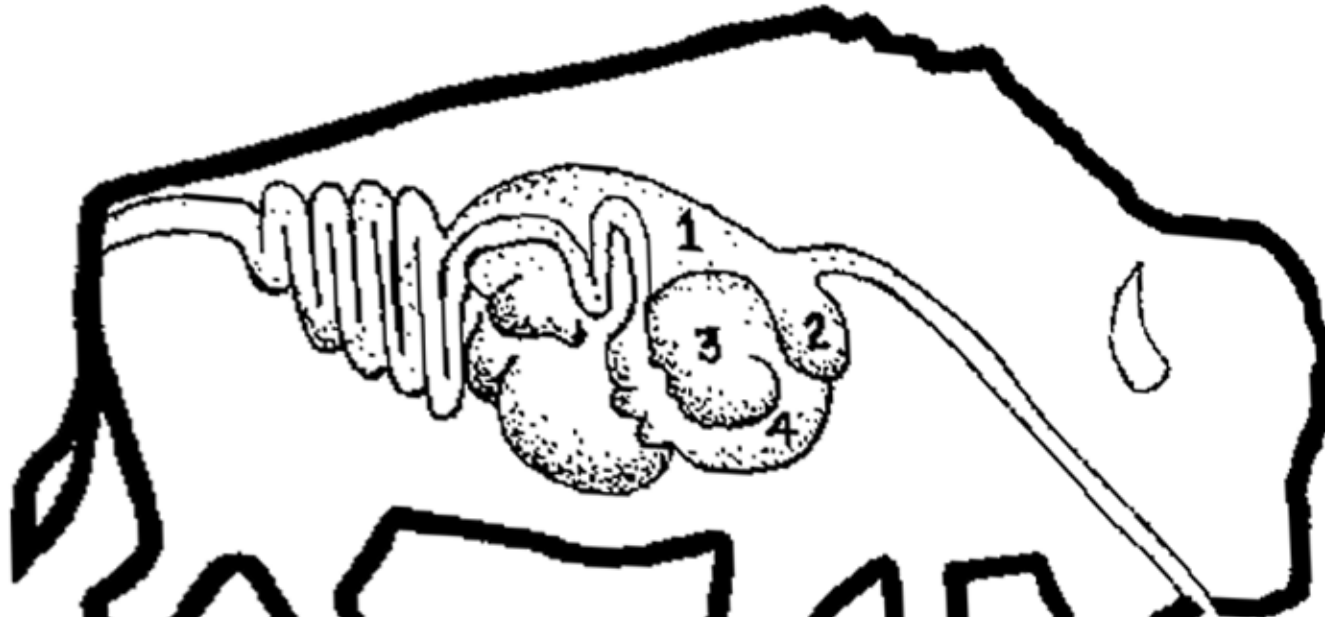
But wait, bison are ruminants



and cannot digest their food without
the assistance of their gut flora.

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

But wait, bison are ruminants



A ruminant is just as much a composite organism as a lichen. The alleged “individual” – whether mycobiont or buffalo – cannot obtain nutrients, and thus cannot live, without its microbial partners.

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

Ruminants are keystone species in many grassland ecosystems.

It is IMPOSSIBLE to equate “keystone species” with “marginal case”.

A ruminant is just as much a composite organism as a lichen. The alleged “individual” – whether mycobiont or buffalo – cannot obtain nutrients, and thus cannot live, without its microbial partners.

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



OK, people then.

Surely human beings aren't just hopped up lichens...

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



8 June 2012

14 June 2012

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

First results
from the Human
Microbiome
Project highlight
the healthy
variation in
our microbial
selves

PAGES 194, 207 & 215

FELLOW TRAVELLERS

MATERIALS

FEELING THE PRESSURE

The quest for convincingly
metallic hydrogen

PAGE 174

CLIMATE CHANGE

GET USED TO UNCERTAINTY

Climate modelling
faces its limits

PAGE 193

SOLAR SYSTEM

TITAN'S ELUSIVE METHANE

Tropical lakes on Saturn's
enigmatic moon?

PAGE 237

NATURE.COM/NATURE

14 June 2012



NEWS & VIEWS

MICROBIOLOGY

Learning about who we are

Microbial inhabitants outnumber our body's own cells by about ten to one. These residents have become the subject of intensive research, which is beginning to elucidate their roles in health and disease. [SEE ARTICLES P.207 & P.215](#)

DAVID A. RELMAN

The dawn of the twenty-first century has seen the emergence of a major theme in biomedical research: the molecular and genetic basis of what it is to be human. Surprisingly, it turns out that we owe much of our biology and our individuality to the microbes that live on and in our bodies — a realization that promises to radically alter the principles and practice of medicine, public health and basic science. It is therefore appropriate that ever more research is focused on these microbes and their genes, which together are known as the human microbiome¹. In this issue, the Human Microbiome Project Consortium^{2,3} publishes the most extensive catalogue yet of organisms and genes pertaining to our microbiomes.

The first observations of indigenous human microbiota were published more than 300 years ago, soon after the invention of the microscope. Today's view of the microbial world has been radically improved by DNA-sequencing technology. In the wake of the Human Genome Project, calls were issued⁴ for enhanced efforts to be made to characterize the 'second human genome' — the human microbiome. At the end of 2007, the US National Institutes of Health (NIH) launched the Human Microbiome Project (HMP) and, in early 2008, the European Commission and China initiated the Metagenomics of the Human Intestinal Tract (MetaHIT) project. Other countries have begun similar ventures, motivated in part by an interest in better defining their biological heritage.

Two studies, by Huttenhower *et al.*² (page 207) and Methé *et al.*³ (page 215), together with 15 other papers^{5,6} that are being published simultaneously elsewhere, comprise the first reports of the HMP Consortium research groups. The primary data, as described by Methé and colleagues³, were derived from samples collected from 242 healthy adults in the United States, at 15 (for males) or 18 (for females) body sites — from the skin, nose, mouth, throat, vagina and faeces (to represent the distal gastrointestinal tract). Each person was sampled up to three times over 22 months, generating a total of 11,174 samples.

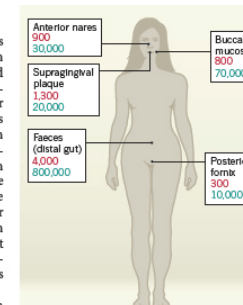


Figure 1 | Variation in diversity. Researchers of the Human Microbiome Project are studying the microbial inhabitants of the human body, using samples taken from 242 healthy adults at 15 (for males) or 18 (for females) body sites — from the skin (four sites), mouth and throat (nine sites), vagina (three sites), nostrils and faeces (to represent the distal gastrointestinal tract). Huttenhower *et al.*² and Methé *et al.*³ have estimated the number of microbial species and their genes in these samples, and found substantial variation in microbial community composition at different body habitats. The two groups used different counting methodologies, and their numbers vary accordingly, such that exact figures are not available. However, crude estimations⁷ of number of microbial species (red) and number of microbial genes (blue) are shown, for examples of sites containing high species diversity, such as the gastrointestinal tract and teeth (supragingival plaque); sites with intermediate diversity, such as the inside of the cheek (buccal mucosa) and nostrils (anterior nares); and sites with lower diversity, such as the vaginal posterior fornix. The authors also found substantial variation in both the diversity and the composition of the microbial communities at different sites within the same general body region.

The consortium researchers obtained the nucleotide sequence of the small-subunit ribosomal RNA — a molecule found in all cellular life — from microorganisms in 5,177 of these samples². These sequences are commonly

used to infer the genetic relationships between organisms. The researchers also surveyed the genomes of the microbes in 681 of the samples² using a shotgun sequencing approach, which generates random sequences (reads) from a complex pool of DNA molecules. The reads are then assembled on the basis of overlapping sequence similarity, allowing researchers to identify genes and to predict the functions of the proteins that they encode.

The investigators mapped their reads to all available microbial and viral genome sequences to assess community composition — the different types of microbes and their relative abundance — at the various body sites. The researchers also determined the whole-genome sequences of about 800 bacterial strains isolated from humans (from a planned total of 3,000); these sequences have been placed in public databases and can be used as reference genomes for comparative purposes. The consortium authors conclude^{2,3} that they have identified the majority of the common microbial taxa and their genes present in these 242 healthy humans.

One of the great strengths of the HMP is that samples were collected simultaneously from multiple body habitats of the same individuals. This allowed Huttenhower *et al.*² to discover that taxonomic and genetic diversity were greatest in tooth and stool samples, intermediate in skin samples and on the inside surface of the cheek, and lowest in vaginal samples^{2,3} (Fig. 1). The researchers report that each habitat is characterized by a small number of highly abundant 'signature' taxa, but that the relative representation of taxa and genes in each habitat varies considerably between individuals. In most samples, high-abundance taxa are accompanied by low-abundance taxa from the same genus, suggesting that within-community niche specialization occurs. These findings confirm those of an earlier study⁷, which demonstrated that body habitat accounts for much of the variation in bacterial community composition. Although there is clear evidence for individuality in people's microbiome compositions, the limited temporal scope of the HMP data set prevents a robust analysis of how these communities change over time.

As shown previously for faecal samples⁸,

194 | NATURE | VOL 486 | 14 JUNE 2012

© 2012 Macmillan Publishers Limited. All rights reserved.

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

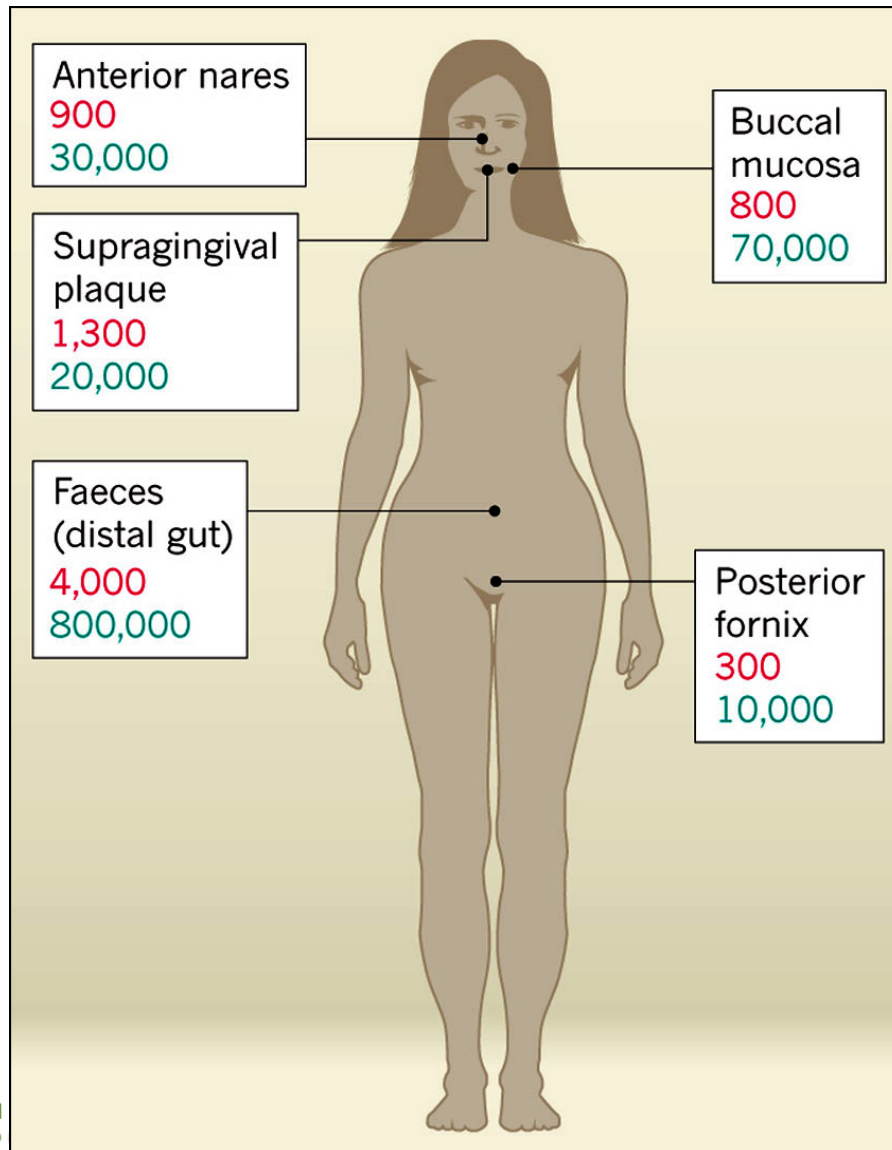


Figure 1 Variation in diversity. Researchers of the Human Microbiome Project are studying the microbial inhabitants of the human body, using samples taken from 242 healthy adults at 15 (for males) or 18 (for females) body sites — from the skin (four sites), mouth and throat (nine sites), vagina (three sites), nostrils and faeces (to represent the distal gastrointestinal tract). Huttenhower *et al.* and Methé *et al.* have estimated the number of microbial species and their genes in these samples, and found substantial variation in microbial community composition at different body habitats. The two groups used different counting methodologies, and their numbers vary accordingly, such that exact figures are not available. However, crude estimations of number of microbial species (red) and number of microbial genes (blue) are shown for examples of: sites containing high species diversity, such as the gastrointestinal tract and teeth (supragingival plaque); sites with intermediate diversity, such as the inside of the cheek (buccal mucosa) and nostrils (anterior nares); and sites with lower diversity, such as the vaginal posterior fornix. The authors also found substantial variation in both the diversity and the composition of the microbial communities at different sites within the same general body region.

The *vanishing* individual.

The holobiont is the major functional unit
in macro-scale biology.

Evolution and a post-modern synthesis.

Evolutionary basics

Nothing in Biology Makes Sense Except in the Light of Evolution

Nothing in Biology Makes Sense Except in the Light of Evolution

Without a quantifiable, underlying model of
heredity, evolution is only a metaphor, with
no explanatory power.

Explanatory Evolution (modern synthesis):

- Individuals of generation n acquire genes — *from parent(s) as a one time event*
- Individuals of gen n change their genes — *via mutation, a very rare saltational event*
- Individuals of gen n either survive to reproduce their genes or not — *affected by fitness (selection) or random events (drift)*
- Individuals of gen $n+1$ acquire genes — *from parent(s) as a one time event*

Metaphorical Evolution (Dawkins' memes):

- Individuals of generation n are exposed to memes which they may or may not acquire — *from parent(s) and environment as a continuous process*
- Individuals of gen n change their memes — *in a continuous and non-saltational manner*
- Individuals of gen n may lose some memes entirely and may also create some entirely new memes
- Individuals of gen n either reproduce their memes or not — *affected by their influence and persuasiveness*
- Individuals of gen $n+1$ are exposed to memes which they may or may not acquire — *from parent(s) and environment as a continuous process*

Evolution and a post-modern synthesis.

The assumption of a tree of life

Assumptions:

- An organism's traits are controlled (affected) by its underlying hereditary material.
- The hereditary material is acquired from an organism's parent(s) at the moment of reproduction.
- During an organism's life it neither gains nor loses hereditary material, except through mutation.
- Differences in traits (controlled by hereditary material) can affect the likelihood of surviving to reproduce across different individuals

Conclusions:

- Evolution will occur.
- All organisms may be placed into a *tree of life*.

Corollary:

- All *natural* classification systems will be based upon relative locations in the tree of life.

420

CLASSIFICATION.

CHAP. XIII.

All the foregoing rules and aids and difficulties in classification are explained, if I do not greatly deceive myself, on the view that the natural system is founded on descent with modification; that the characters which naturalists consider as showing true affinity between any two or more species, are those which have been inherited from a common parent, and, in so far, all true classification is genealogical; that community of descent is the hidden bond which naturalists have been unconsciously seeking, and not some unknown plan of creation, or the enunciation of general propositions, and the mere putting together and separating objects more or less alike.

Darwin, *Origin of Species*

420

CLASSIFICATION.

CHAP. XIII.

All the foregoing rules and aids and difficulties in classification are explained, if I do not greatly deceive myself, on the view that the natural system is founded on descent with modification; that the characters which naturalists consider as showing true affinity between any two or more species, are those which have been inherited from a common parent, and, in so far, all true classification is genealogical; that community of descent is the hidden bond which naturalists have been unconsciously seeking, and not some unknown plan of creation, or the enunciation of general propositions, and the mere putting together and separating objects more or less alike.

Darwin, *Origin of Species*

420

CLASSIFICATION.

CHAP. XIII.

All the foregoing rules and aids and difficulties in classification are explained, if I do not greatly deceive myself, on the view that the natural system is founded on descent with modification; that the characters which naturalists consider as showing true affinity between any two or more species, are those which have been inherited from a common parent, and, in so far, all true classification is genealogical; that community of descent is the hidden bond which naturalists have been unconsciously seeking, and not some unknown plan of creation, or the enunciation of general propositions, and the mere putting together and separating objects more or less alike.

Darwin, *Origin of Species*

420

CLASSIFICATION.

CHAP. XIII.

All the foregoing rules and aids and difficulties in classification are explained, if I do not greatly deceive myself, on the view that the natural system is founded on descent with modification; that the characters which naturalists consider as showing true affinity between any two or more species, are those which have been inherited from a common parent, and, in so far, all true classification is genealogical; that community of descent is the hidden bond which naturalists have been unconsciously seeking, and not some unknown plan of creation, or the enunciation of general propositions, and the mere putting together and separating objects more or less alike.

Darwin, *Origin of Species*

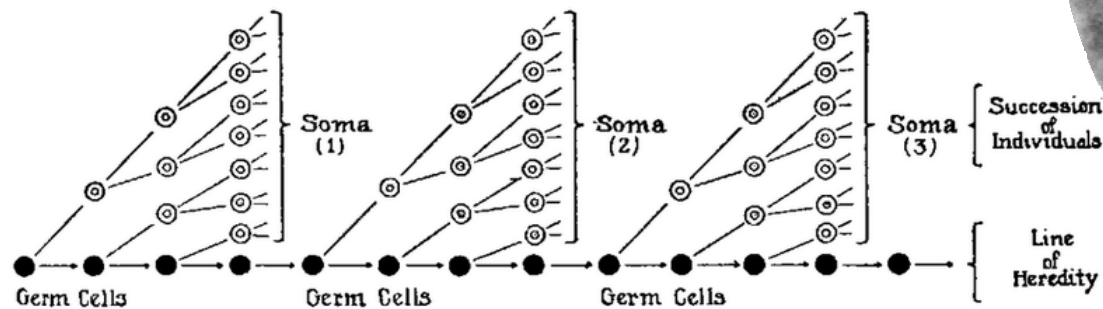
Evolution and a post-modern synthesis.

Review of modern synthesis



Darwin knew that his model required some hereditary mechanism that could supply the variation upon which selection could work, but which would also be resistant to dilution through “blending.” He never developed a working model of his own, and some of his provisional ideas flirted with Lamarckism.

August Weismann's work on the germ-plasm theory assumed the hereditary stuff was in the cell nucleus and showed how this ruled out Lamarckian-style inheritance.



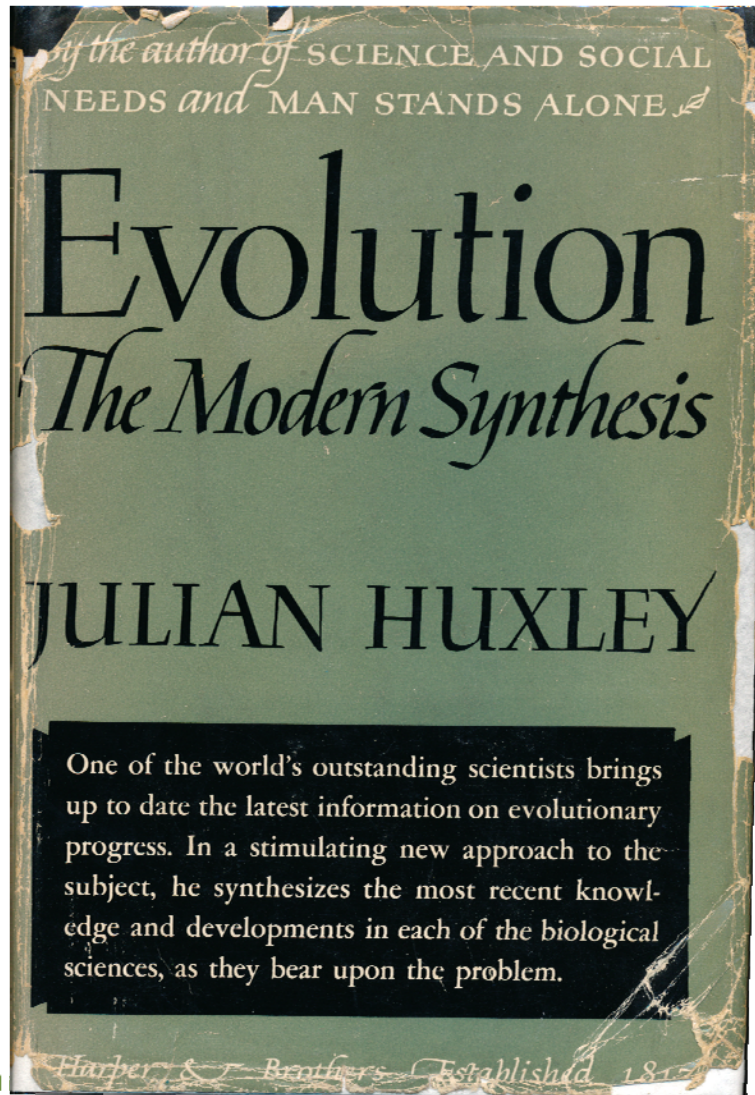
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM



George Romanes thought Weismann's germ-plasm work ruling out the inheritance of acquired characteristics was an important extension to Darwin's own thinking, and so coined the phrase neo-Darwinism to describe this improved evolutionary model.

Mendel's work could have provided the hereditary model, but his work was unknown to Darwin (and unknown to most of science) until 1900, when the rediscovery of his work triggered an explosion of new research, establishing the field of classical genetics.






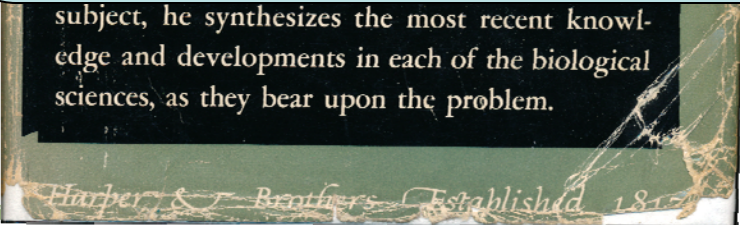
The combination of neo-Darwinism and Mendelism produced

The Modern Synthesis

which has provided the intellectual foundation of most evolutionary thought from 1940 to the present.

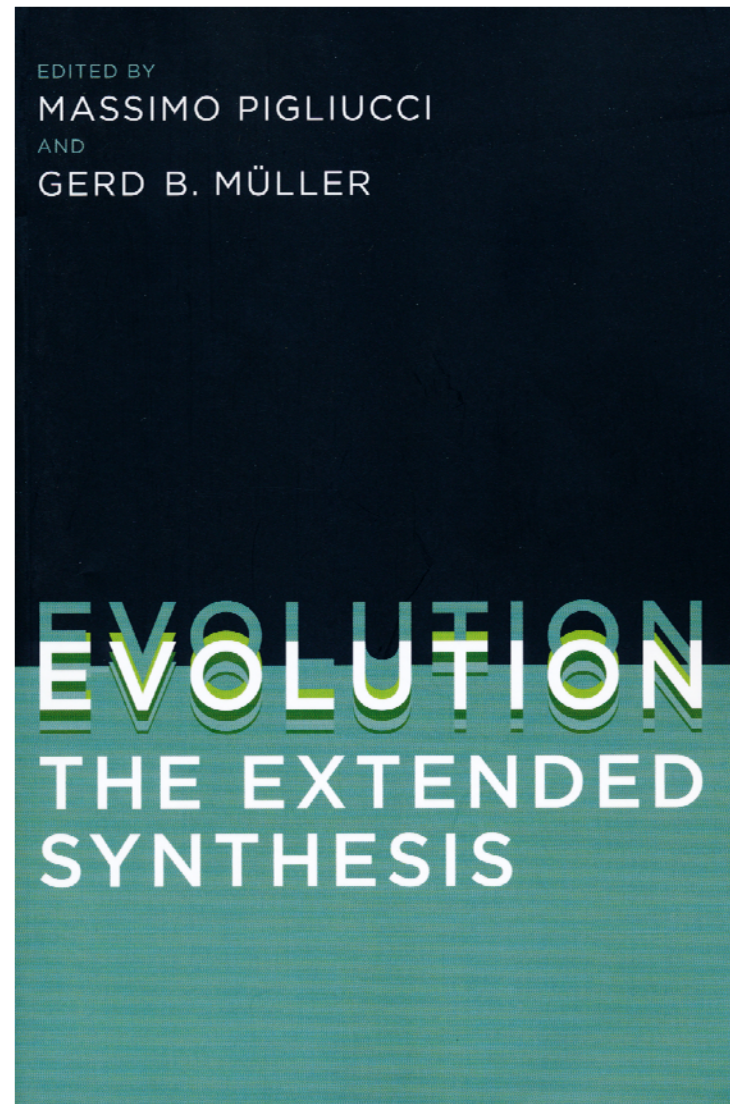


Note that The Modern Synthesis was completed before Watson and Crick worked out the structure of DNA and before any tools of molecular biology were available to address problems of heredity, development, or evolution.



subject, he synthesizes the most recent knowl-
edge and developments in each of the biological
sciences, as they bear upon the problem.

Today, some researchers are attempting to integrate newer findings from genomics and other fields to yield an improved and extended synthesis, suitable for 21st-century biology.

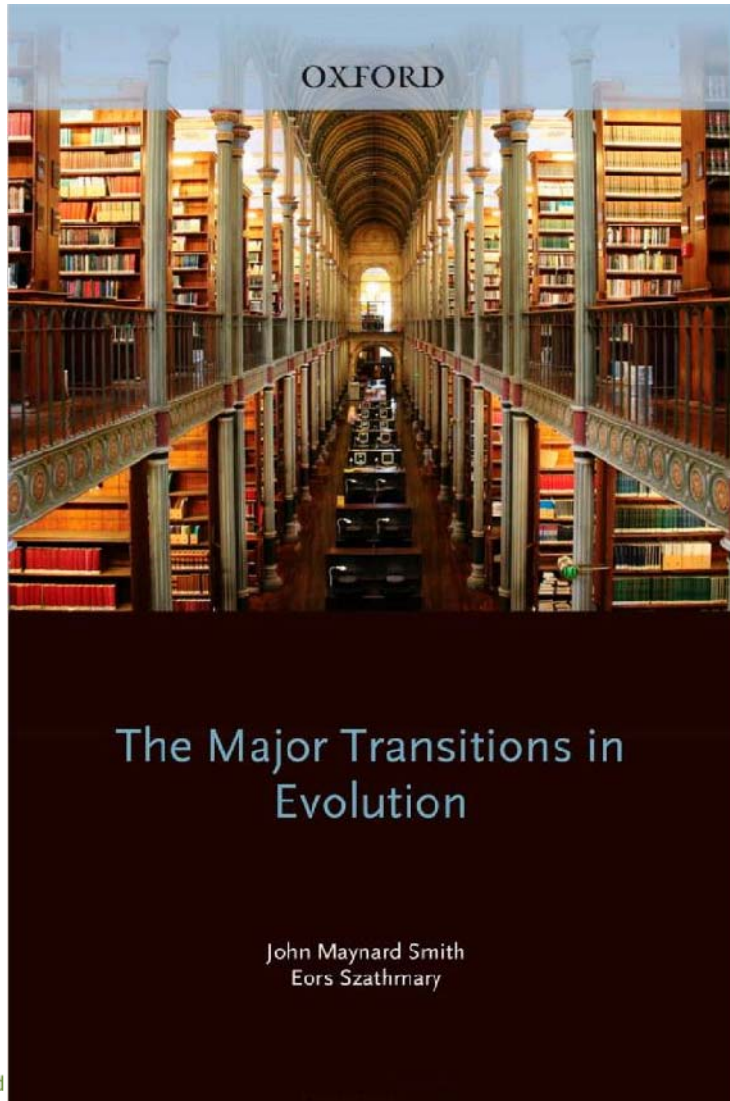


EDITED BY

In planning for the future, TDWG would be well advised to attend carefully to these newly emerging evolutionary concepts.

Evidence for new complexities and subtleties is growing, while some earlier fundamental assumptions are proving to be wrong.

The possibility of significant extensions to our basic notions of organism and species seems not far off.



In the history of life on Earth, several major transitions have occurred.

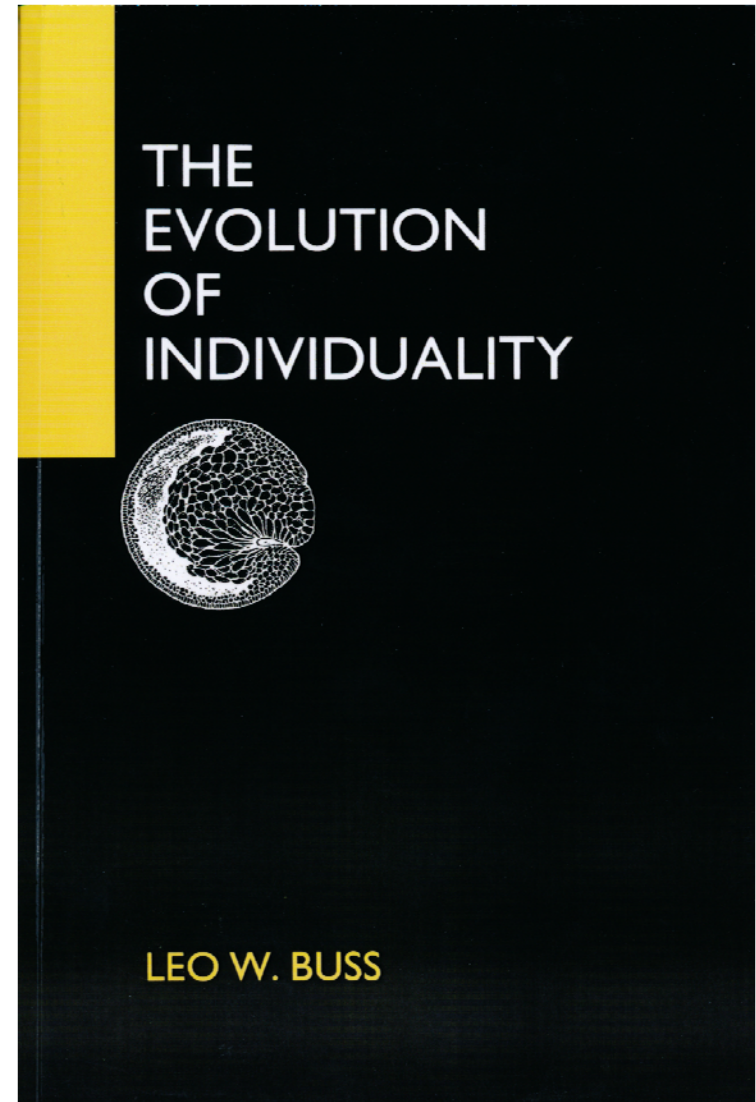
These transitions were significant enough to change the nature of the evolutionary process itself, making it impossible to apply assumptions and analyses from one side of a transition to the other.

Replicating molecules → Populations of molecules
Independent replicators → Chromosomes
RNA → DNA
Prokaryotes → Eukaryotes
Asexual clones → Sexual populations
Unicellularity → Multicellularity
Solitary individuals → Colonies
Primate societies → Human societies (language)

For more than 80% of the time life has been evolving on Earth, multicellular “individuals” did not exist.

Even now, they occur in only a handful of top-level taxa.

Thus, making the “individual” the centerpiece for understanding evolution and for classifying life on Earth seems problematic.



RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

For more than 80% of the

Attempting to understand microbial communities by thinking of them as a bunch of little bitty mice is an activity that falls on a continuum somewhere between fruitless and just plain wrong.

standing evolution and for classifying life on Earth seems problematic.

LEO W. BUSS

Evolution and a post-modern synthesis.

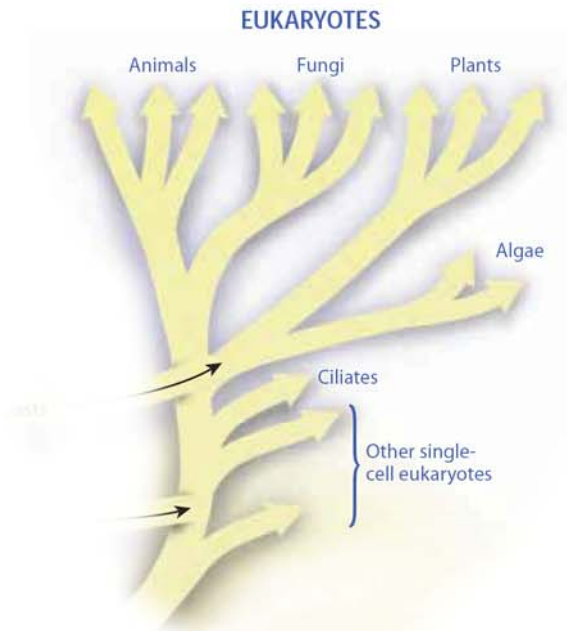
Prokaryotes are different.

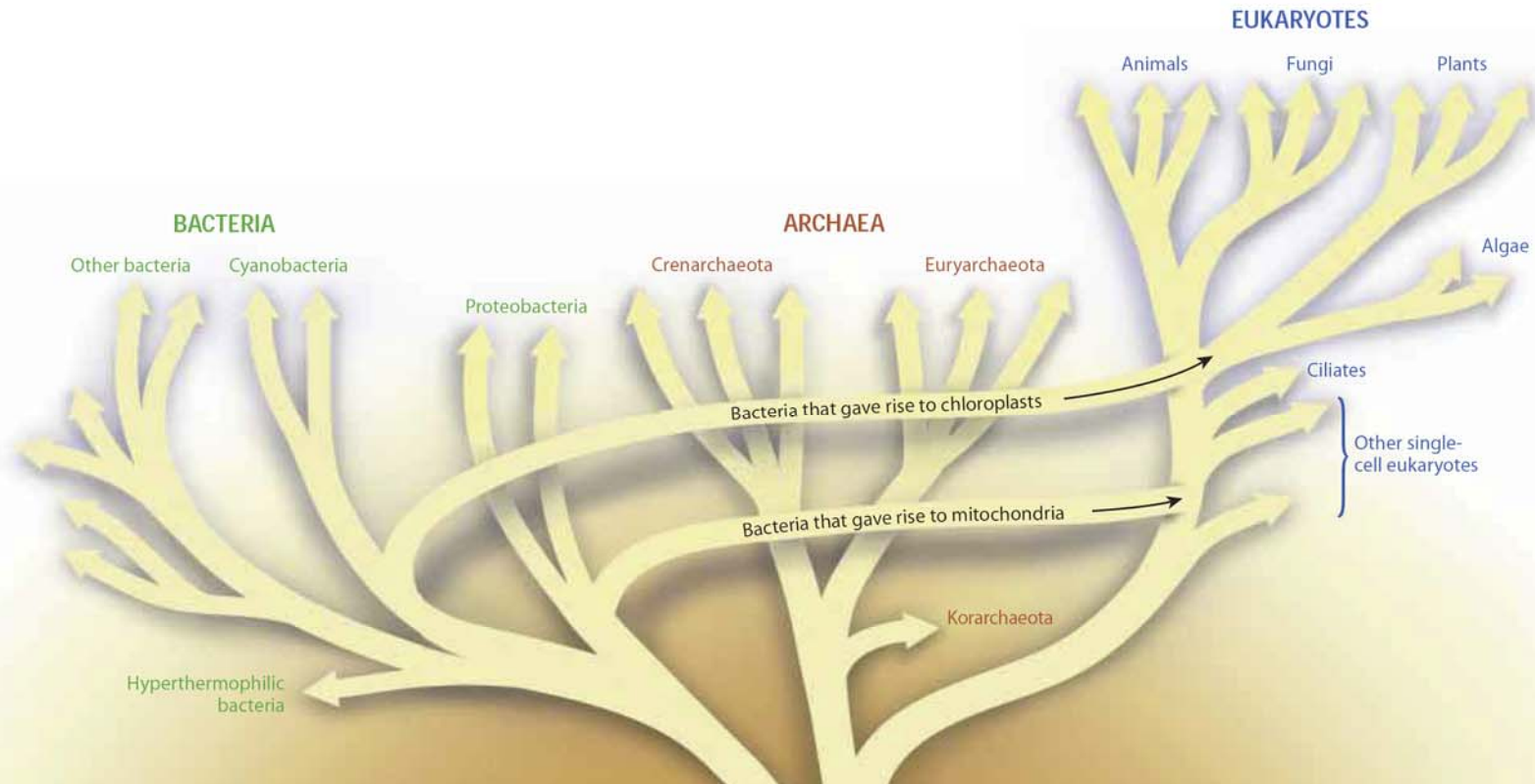
Evolution and a post-modern synthesis.

No simple tree of life.

When is a Tree not a Tree?

Most multicellular eukaryotic taxa can be arranged in a tree-like configuration, but when we include the origin of intra-cellular organelles things get more complicated...

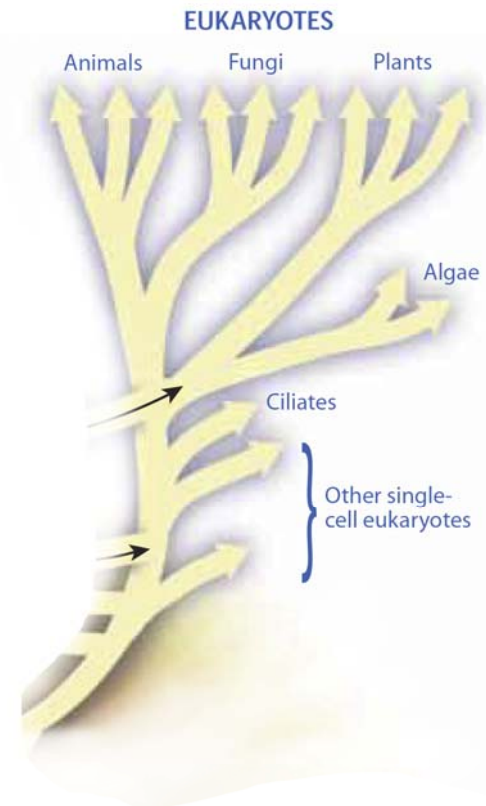




The origin of mitochondria, chloroplasts, and several other eukaryotic cell inclusions through endosymbiosis means that, technically speaking at least, not only are eukaryotic taxa polyphyletic, so are eukaryotic “individuals”.

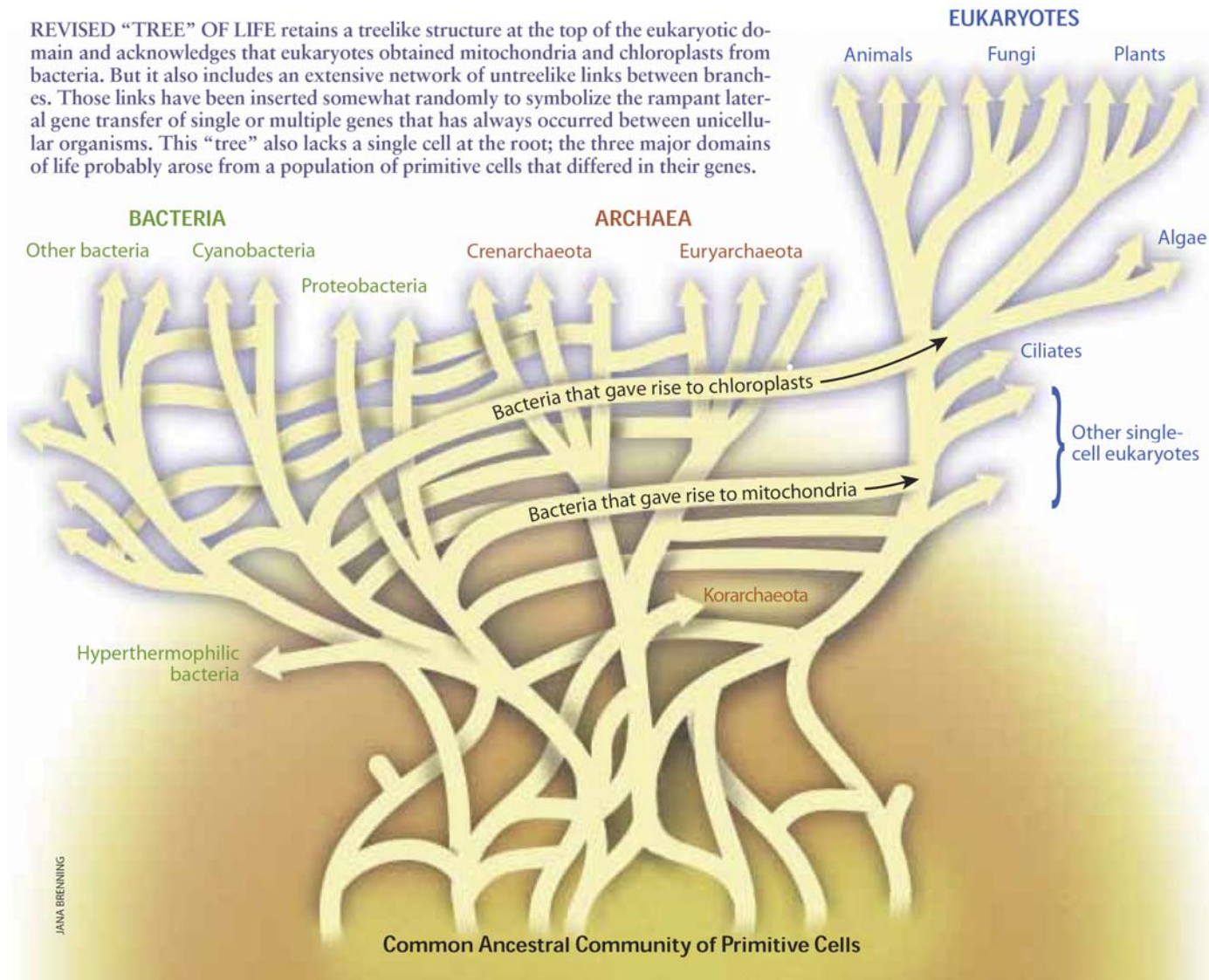
In addition...

We now know that horizontal gene transfer (HGT) occurs regularly among prokaryotes and most likely was the dominant form of inheritance during the early evolution of life on Earth.



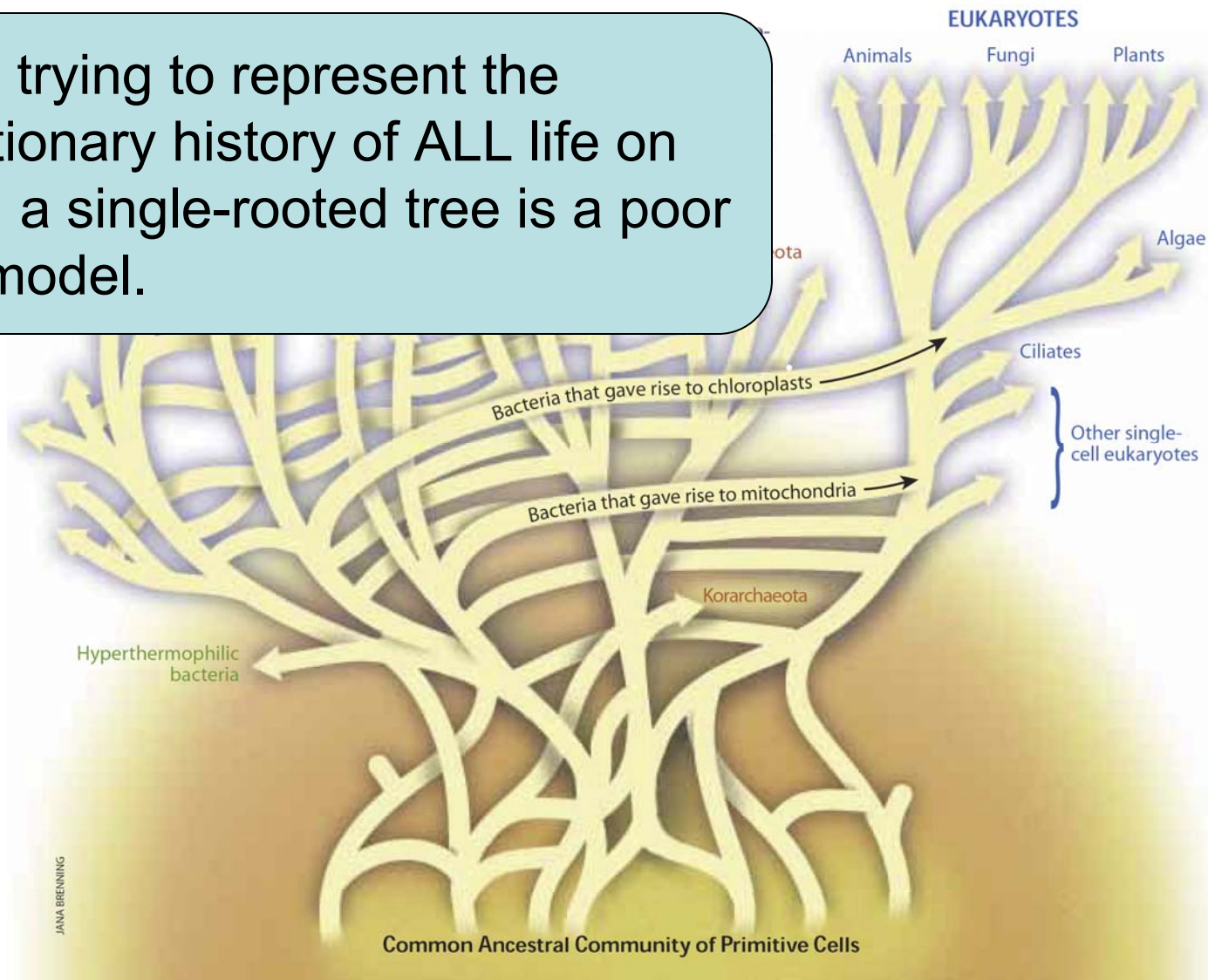
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

REVISED “TREE” OF LIFE retains a treelike structure at the top of the eukaryotic domain and acknowledges that eukaryotes obtained mitochondria and chloroplasts from bacteria. But it also includes an extensive network of untrelike links between branches. Those links have been inserted somewhat randomly to symbolize the rampant lateral gene transfer of single or multiple genes that has always occurred between unicellular organisms. This “tree” also lacks a single cell at the root; the three major domains of life probably arose from a population of primitive cells that differed in their genes.



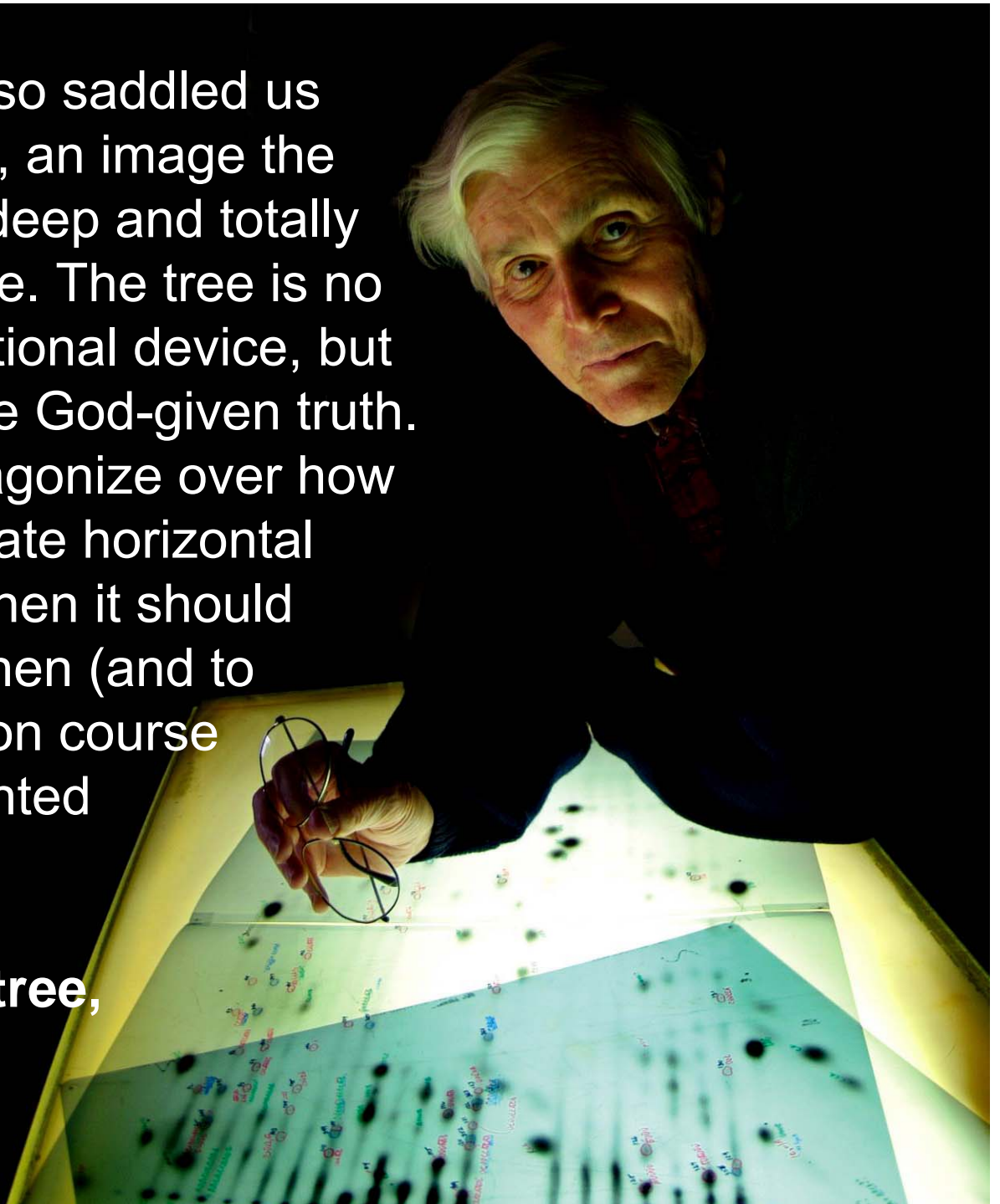
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

When trying to represent the evolutionary history of ALL life on Earth, a single-rooted tree is a poor data model.



Classical biology has also saddled us with a phylogenetic tree, an image the biologist invests with a deep and totally unwarranted significance. The tree is no more than a representational device, but to the biologist it is some God-given truth. Thus, for example, we agonize over how the tree can accommodate horizontal gene transfer events, when it should simply be a matter of when (and to what extent) the evolution course can be usefully represented by a tree diagram:

**Evolution defines the tree,
not the reverse.**



Reality is not negotiable.

Examples from genetics & genomics.

Before Molecular Biology:

Genes are the fundamental units of mutation, recombination, and heredity; they are arranged on the chromosomes like beads on a string.

Before Molecular Biology:

Genes are the fundamental units of mutation, recombination, and heredity; they are arranged on the chromosomes like beads on a string.

After Molecular Biology:

No fundamental units, no beads, and no string...

Reality is not negotiable.

The future of biodiversity.

Before Metagenomics:

Individual organisms are the fundamental units of biodiversity; their evolutionary history can be explained by arranging them into groups, with the groups composed into a single-rooted tree.

Before Metagenomics:

Individual organisms are the fundamental units of biodiversity; their evolutionary history can be explained by arranging them into groups, with the groups composed into a single-rooted tree.

After Metagenomics

No “completely objective” individuals, no one true tree.

**This is fundamentally a
microbial world.**

Microbial World:

- In addition to being ubiquitous and abundant on and in every macroscale organism, prokaryotes occur in every imaginable environment (and maybe a few more).

Microbial World:

- In addition to being ubiquitous and abundant on and in every macroscale organism, prokaryotes occur in every imaginable environment (and maybe a few more).
- They are abundant (more bacteria in a bucket of seawater than there are mammals in Africa).

Microbial World:

- In addition to being ubiquitous and abundant on and in every macroscale organism, prokaryotes occur in every imaginable environment (and maybe a few more).
- They are abundant (more bacteria in a bucket of seawater than there are mammals in Africa).
- They are locally diverse (1 g of soil contains 10^7 - 10^9 prokaryotic cells, with 2,000–18,000 different genomes).

Microbial World:

To repeat: That's 2,000 to 18,000
separate "species" in a
teaspoon of soil...

different genomes).

Microbial World:

To repeat: That's 2,000 to 18,000 separate "species" in a teaspoon of soil...

Reminder: There are about 10,000 species of birds on the entire Earth.

different genomes).

Bottom Line:

- This is fundamentally a microbial biosphere. Half of the biomass and most of the diversity occur in microbes.

Bottom Line:

- This is fundamentally a microbial biosphere. Half of the biomass and most of the diversity occur in microbes.
- Microbes occur in free-living communities and also in very tight, functional associated with ALL multi-cellular organisms.

Bottom Line:

- This is fundamentally a microbial biosphere. Half of the biomass and most of the diversity occur in microbes.
- Microbes occur in free-living communities and also in very tight, functional associated with ALL multi-cellular organisms.
- The notion of individual organisms, as fundamental units in nature, is not objective “truth” – instead it is, at best, a useful approximation.

Bottom Line:

This is fundamentally a microbial biosphere

From the perspective of community biology (which arguably is synonymous with “biology”), the “individual” is a reductionist abstraction.

It is useful in the way “assume a spherical cow” is useful in biophysics – it simplifies the analysis, but at some cost to its correspondence with reality.

mental units in nature, is not objective “truth” – instead it is, at best, a useful approximation.

The idea that the whole may be understood by understanding all of its parts is the conceit of reductionism.

The idea that the whole may be understood by understanding all of its parts is the conceit of reductionism.

The idea that the whole may be understood by understanding a few (or only one) of its parts is simply wrong.

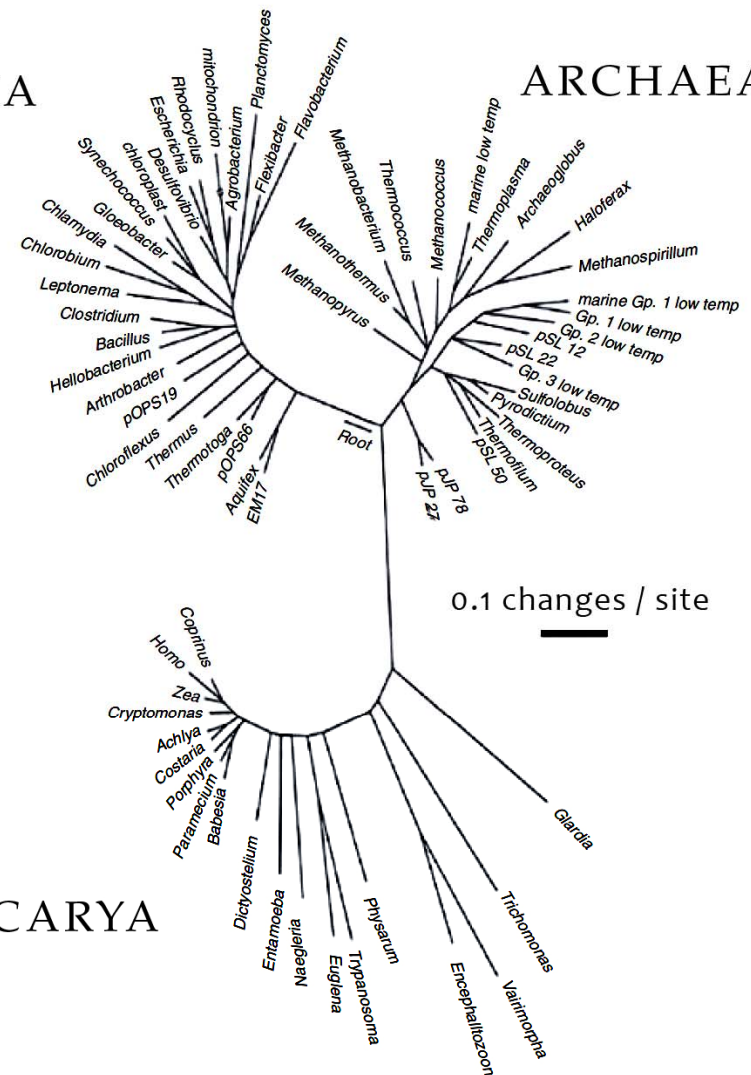
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

If the goal of biodiversity studies is to understand all of the diversity in the Earth's biosphere...

BACTERIA

ARCHAEA

EUCARYA



RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

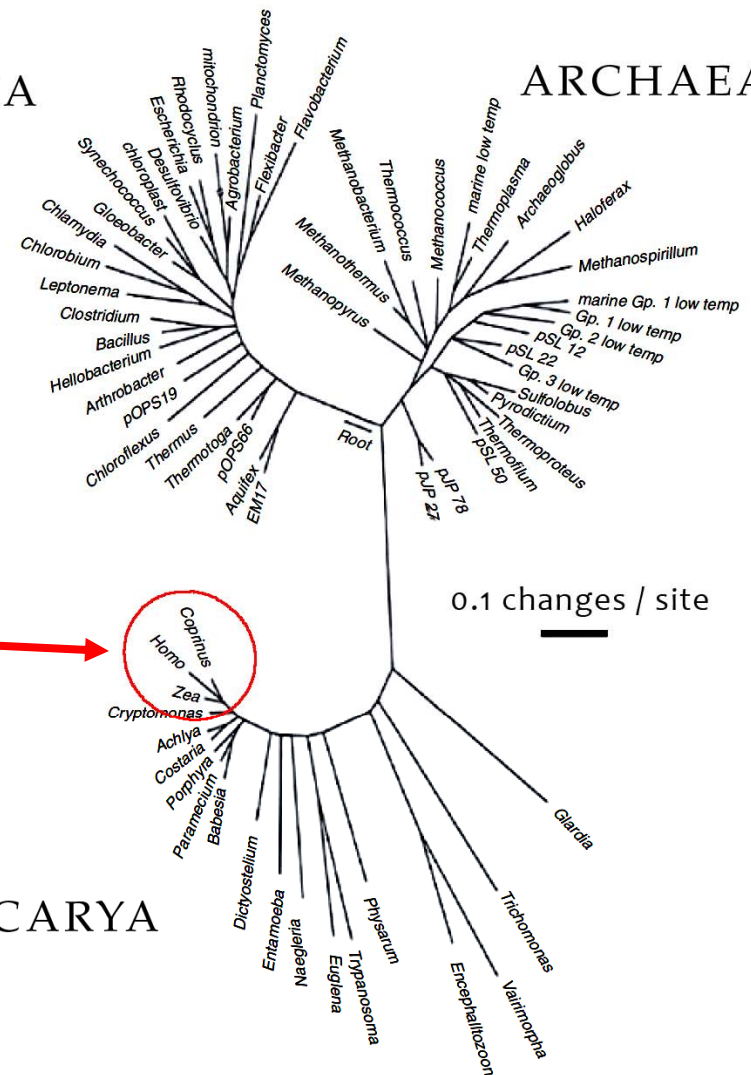
If the goal of biodiversity studies is to understand all of the diversity in the Earth's biosphere...

Then the notion that we can accomplish that goal only by looking here is just plain wrong.

BACTERIA

ARCHAEA

EUCARYA



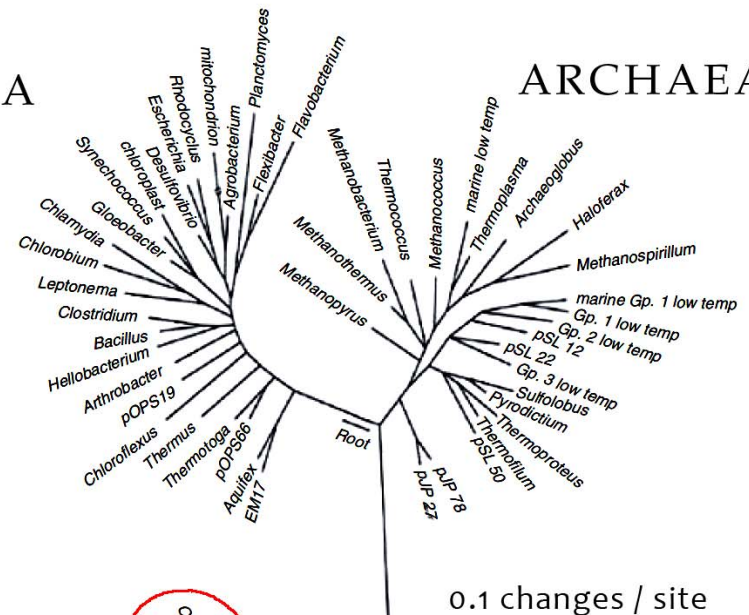
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

If the goal of biodiversity studies is to understand all of the diversity in the Earth's biosphere...

Then the notion that we can accomplish that goal only by looking here

BACTERIA

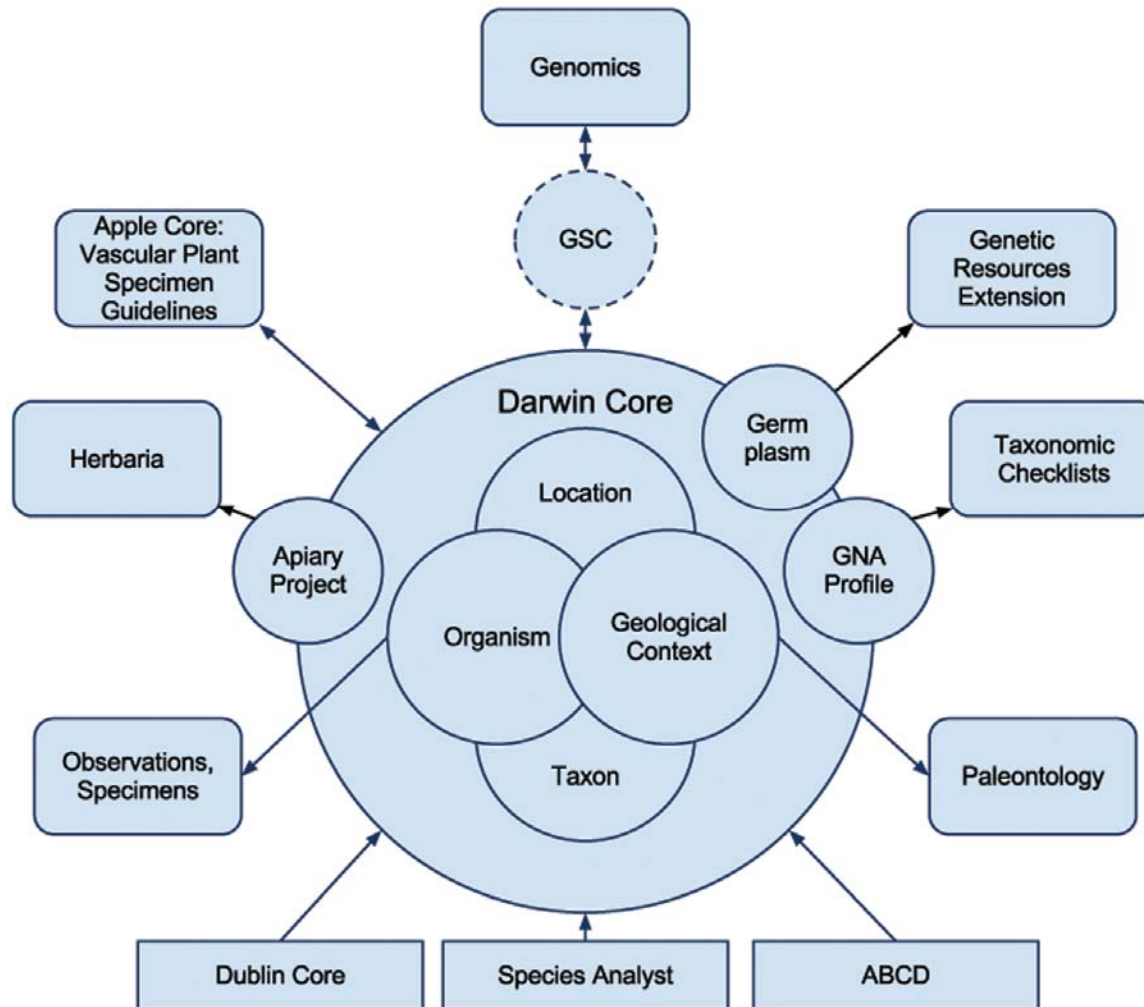
ARCHAEA



To maintain relevancy, to deal with the non-negotiable aspects of nature, 21st century biodiversity studies **MUST** include a large and growing commitment to understanding microbial biodiversity.

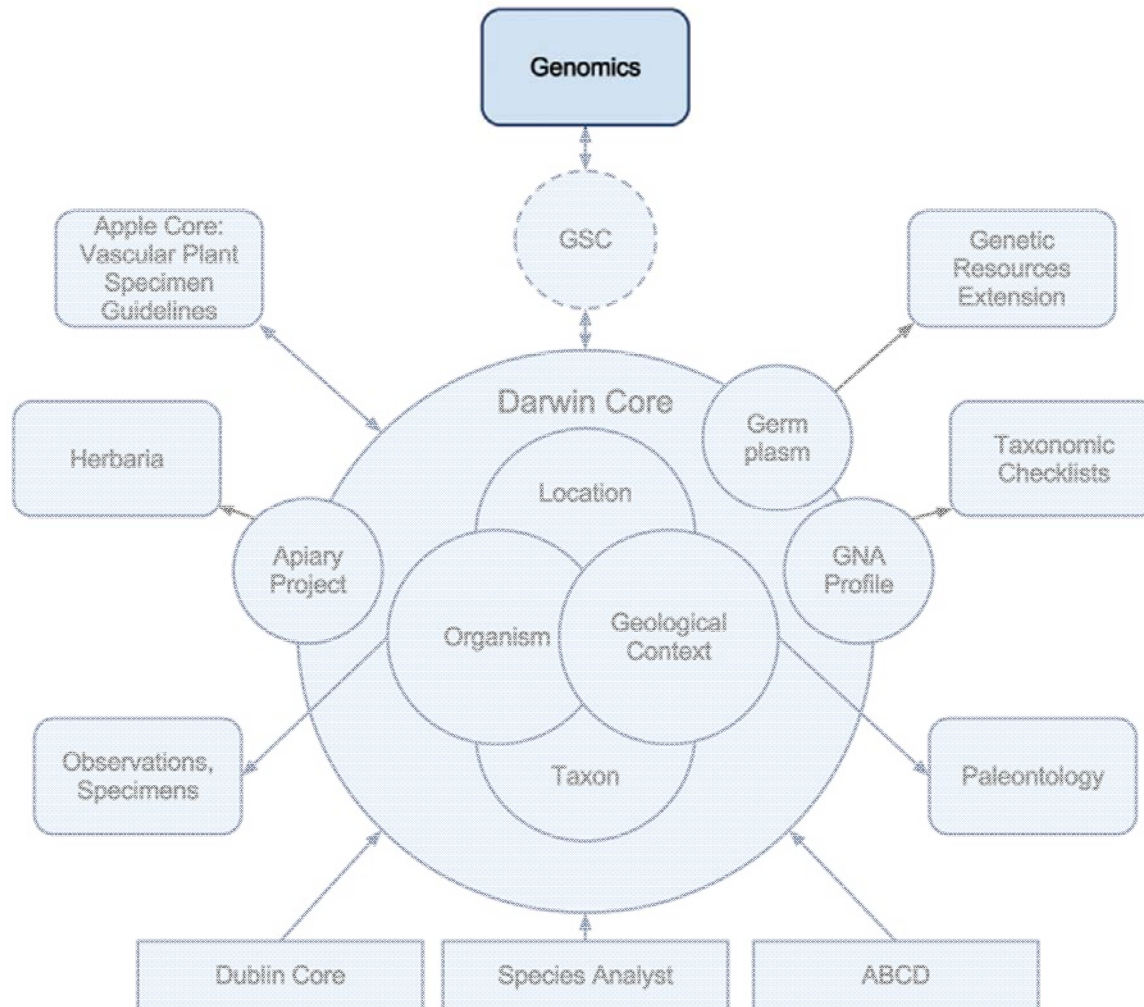
RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

DwC:



RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

DwC:



**Welcome to the world of 21st
century biodiversity.**

RESEARCH COORDINATING NETWORK FOR THE GENOMIC STANDARDS CONSORTIUM

Robert J. Robbins
RJR8222@gmail.com

END

 (<http://www.rj-robbins.com/presentations/TDWG-2014>) 