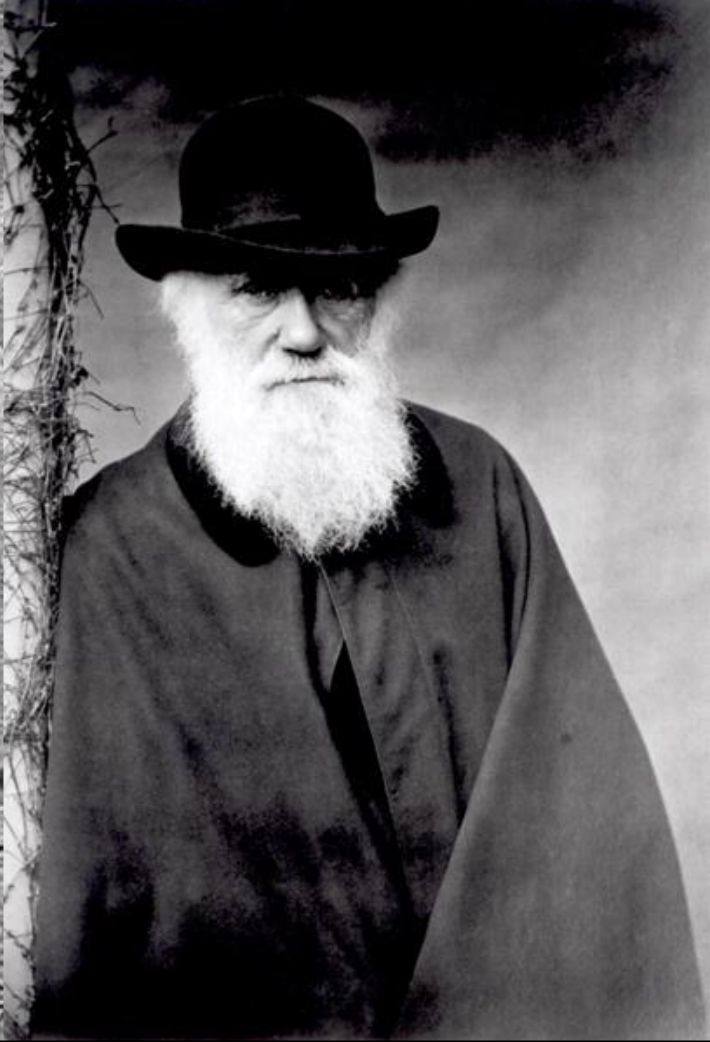


When Michel met Charles:
turning evolutionary biology into a network science

E. BAPTESTE

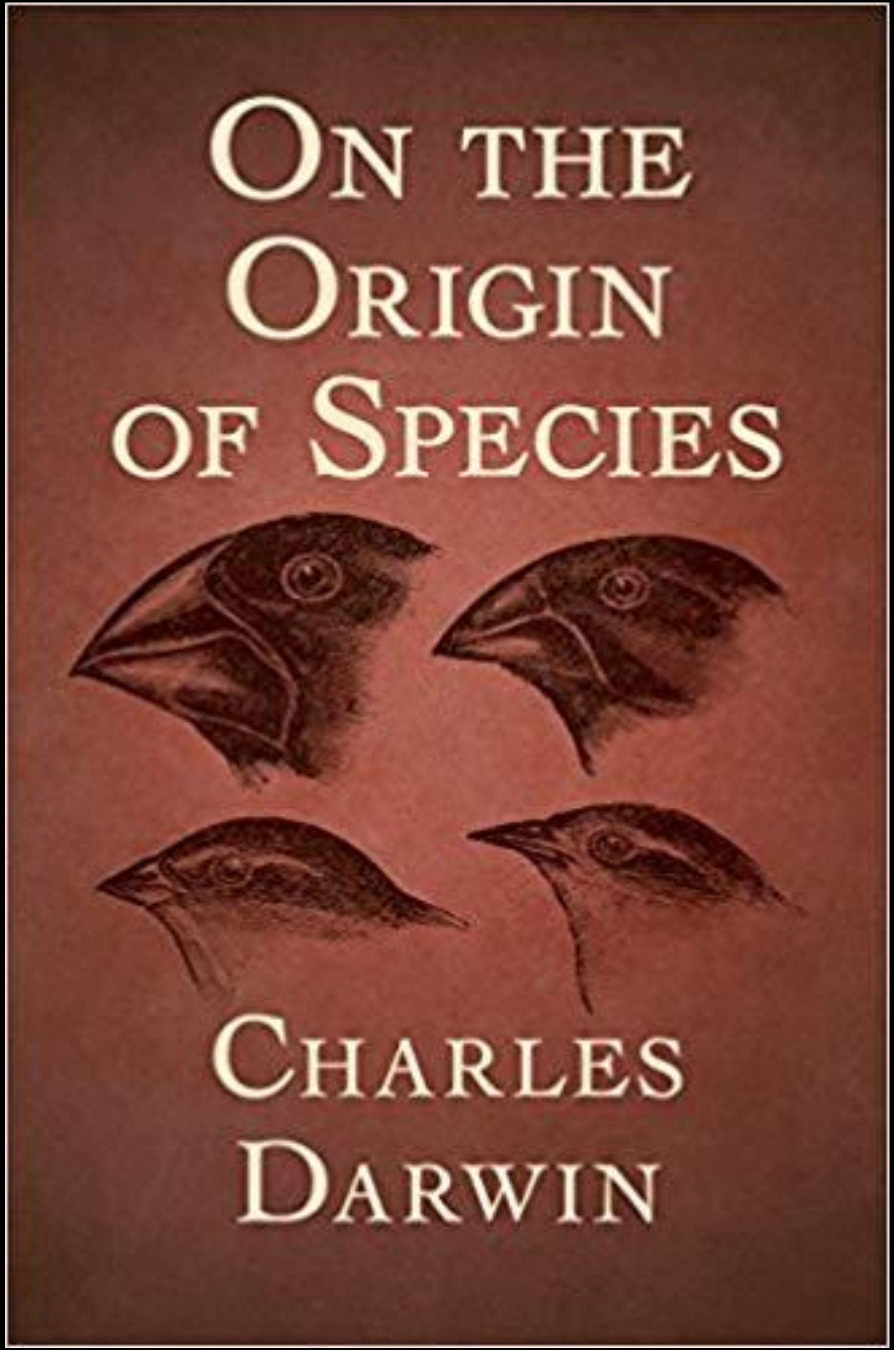
Take home message:

Evolution can be better described and analyzed using networks than trees, therefore evolutionary biologists need graph theoreticians.

Evolutionary biologists search what processes produced the diversity of life forms, over time.

e.g. why so many different moustache-bearing organisms?

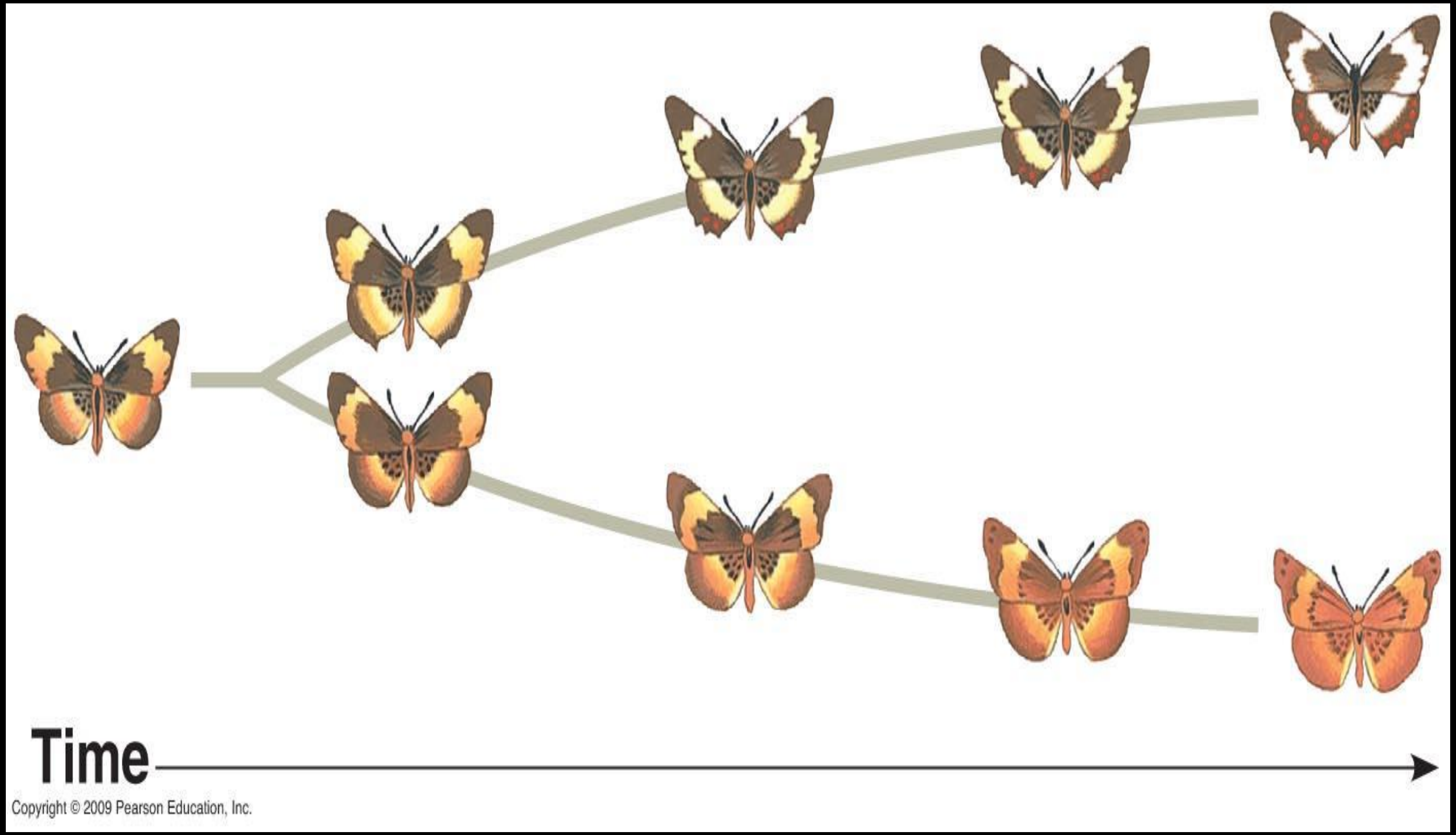
A great answer to this question came from Charles Darwin.



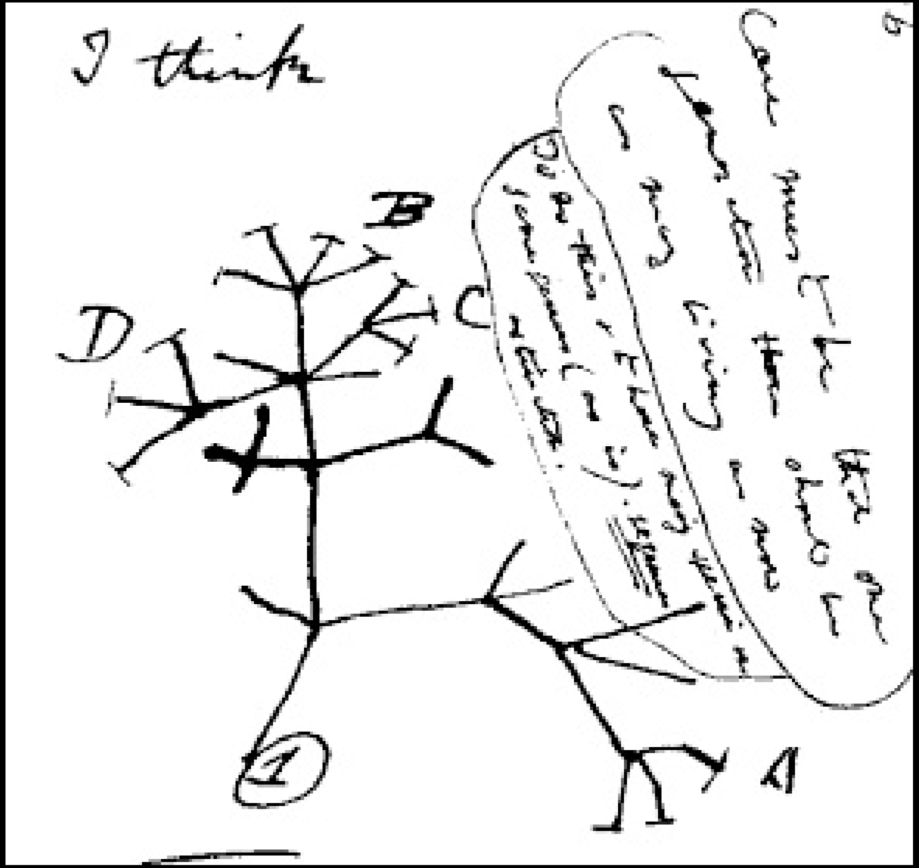
1859

Essentially, his book presented:

- **1 process**: descent with modification
- **3 conditions** for its realization (variation, inheritance, differential fitness)
- **2 bold hypotheses**: ***natural selection*** + ***tree of life***



Copyright © 2009 Pearson Education, Inc.

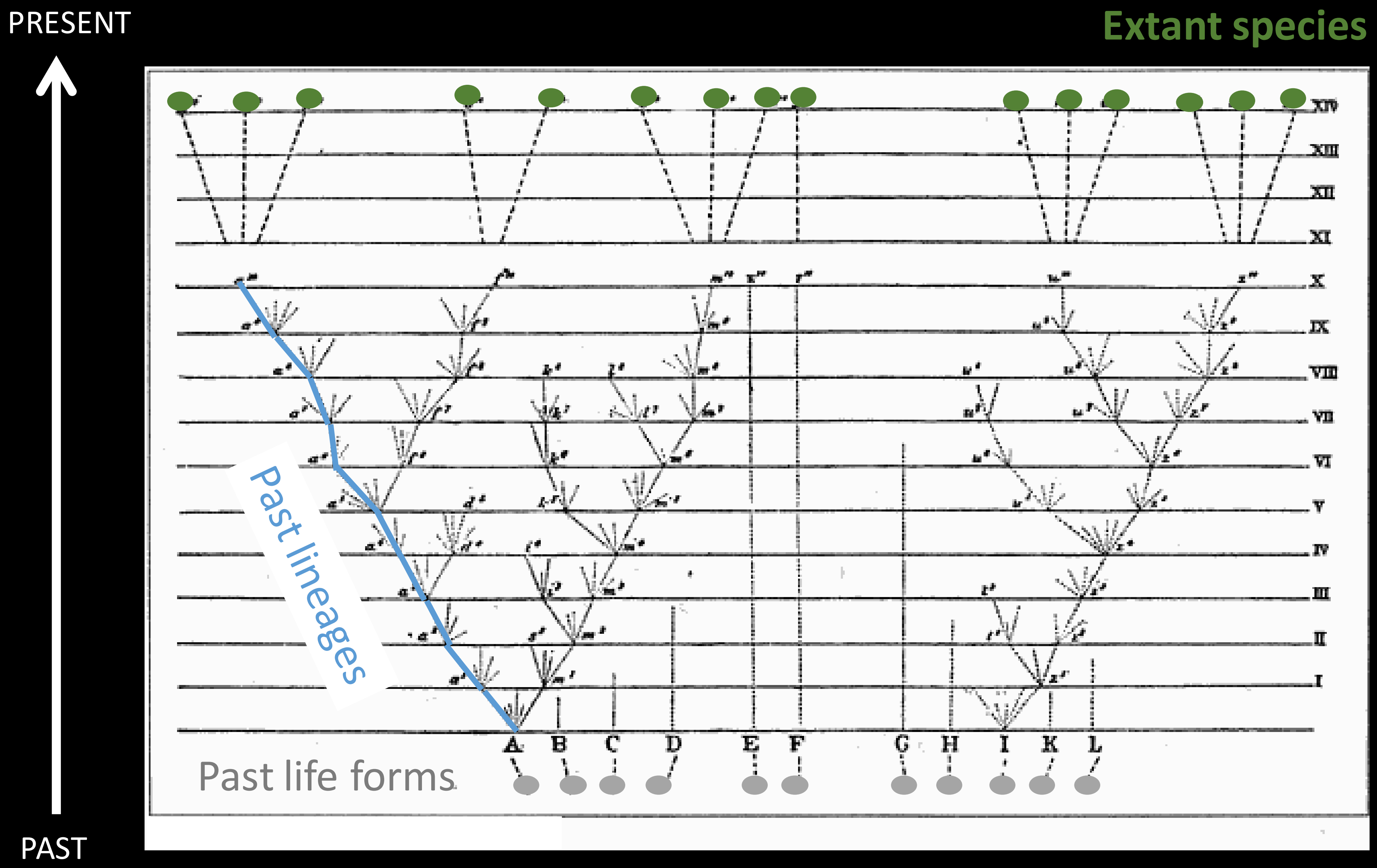


This process is often popularized as the **survival of the fittest**

- the **production of (advantageous) variation**
- the **transmission of that (advantageous) variation to offsprings**
- An **increased ability of organisms with advantageous variations to produce more offsprings**,

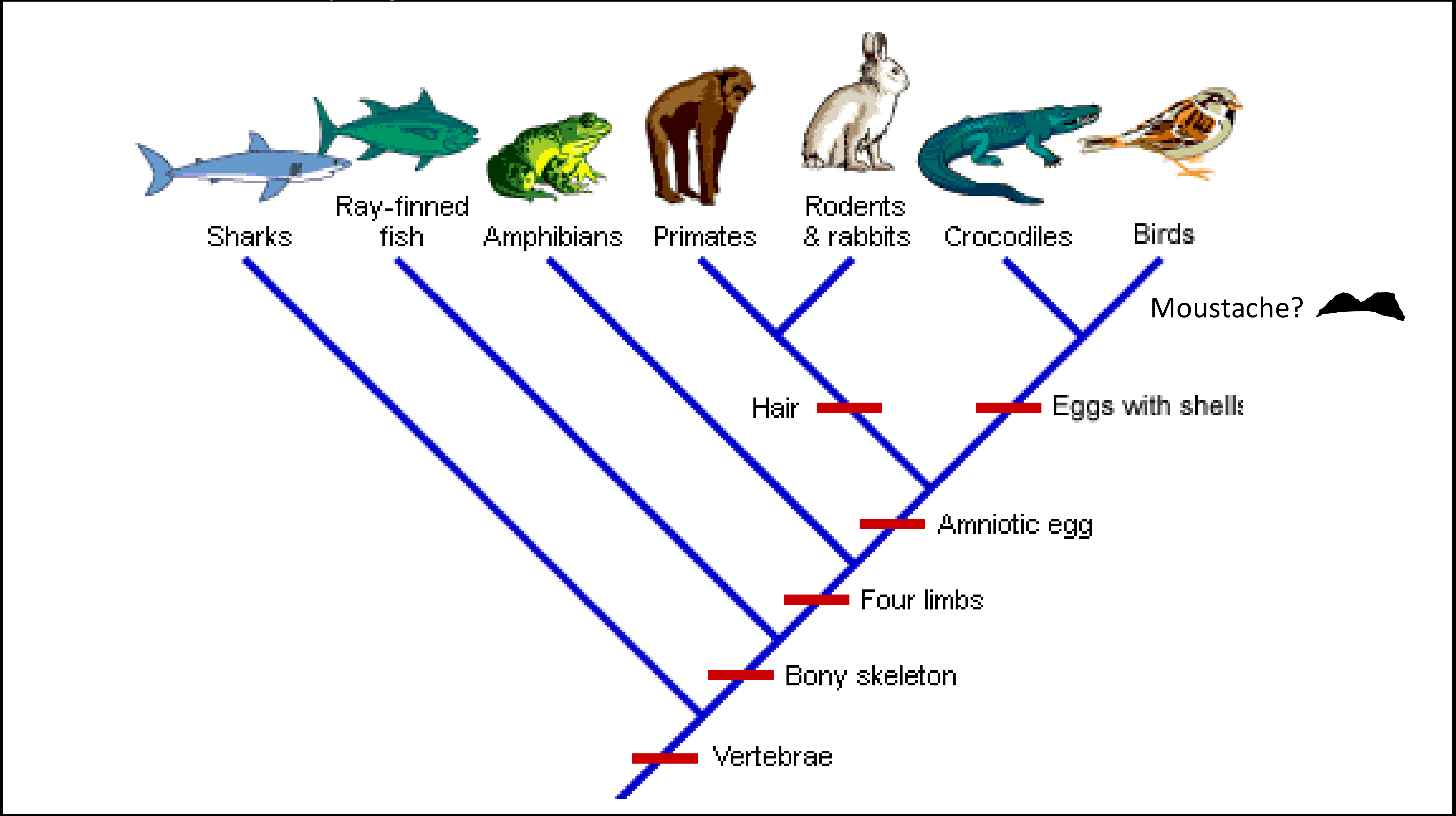
So that, over generations, **the frequency of more fit organisms would increase in a population.**

Darwin extrapolated this logic to explain the evolution of all **organismal** lineages on Earth.

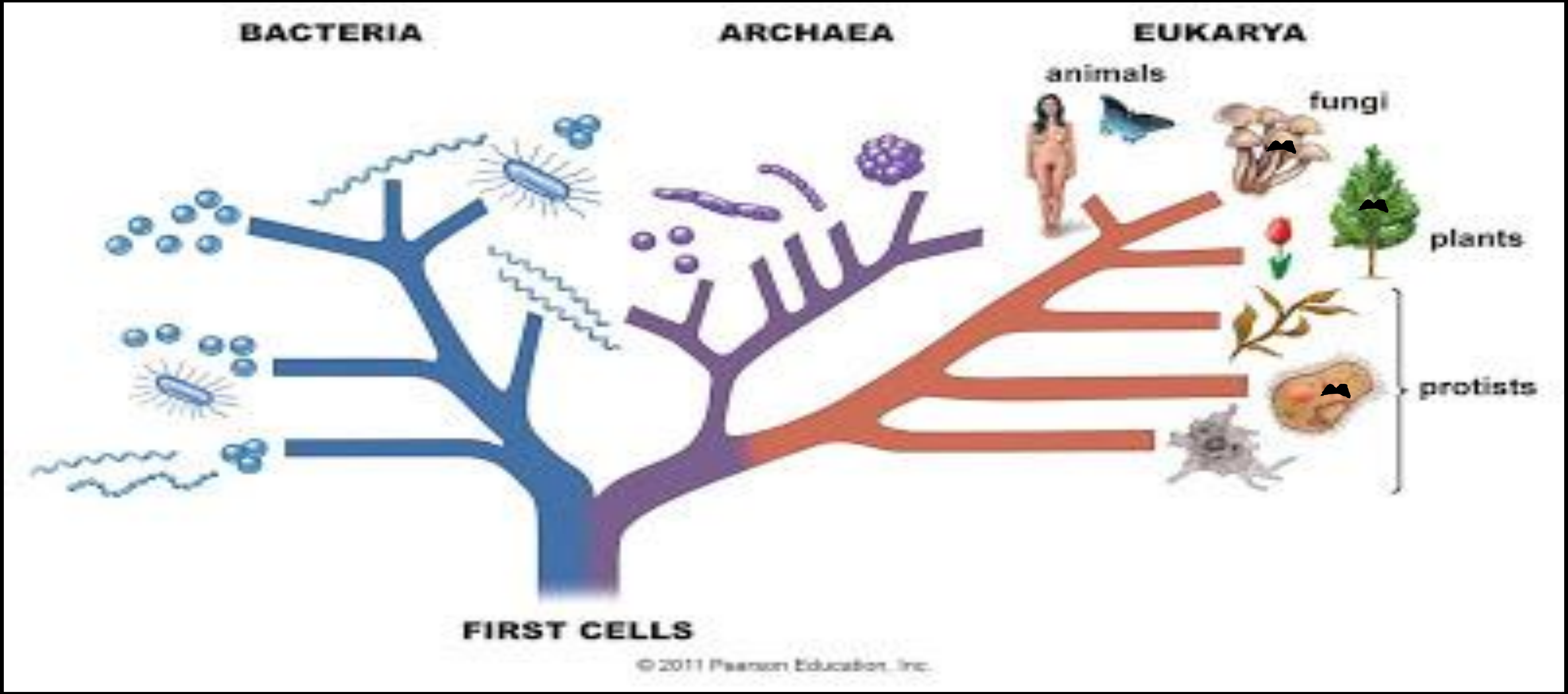


C. Darwin 1859

Since then, a popular way to infer the branching order of organismal lineages consist in identifying the shared derived traits inherited from an ancestor.

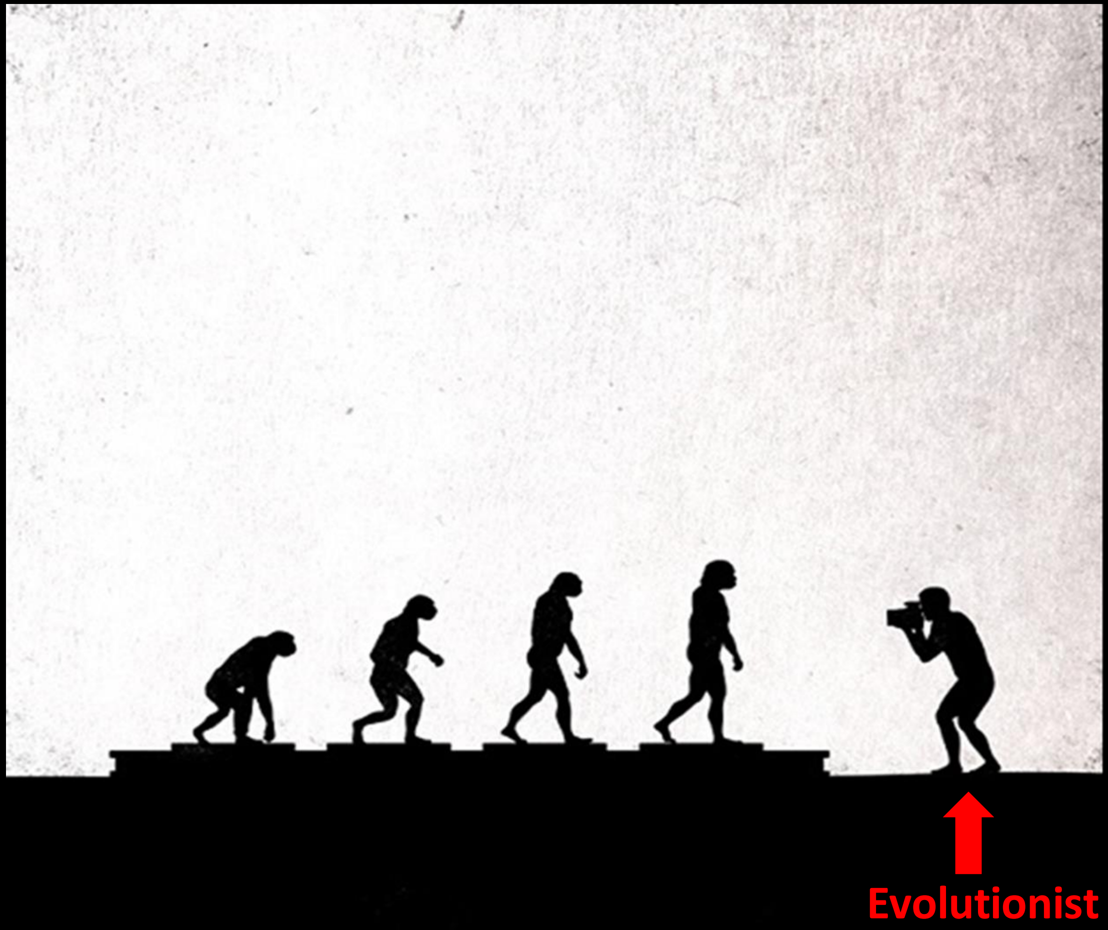


This research program culminated with the reconstruction of a Tree of life with **3 Domains**

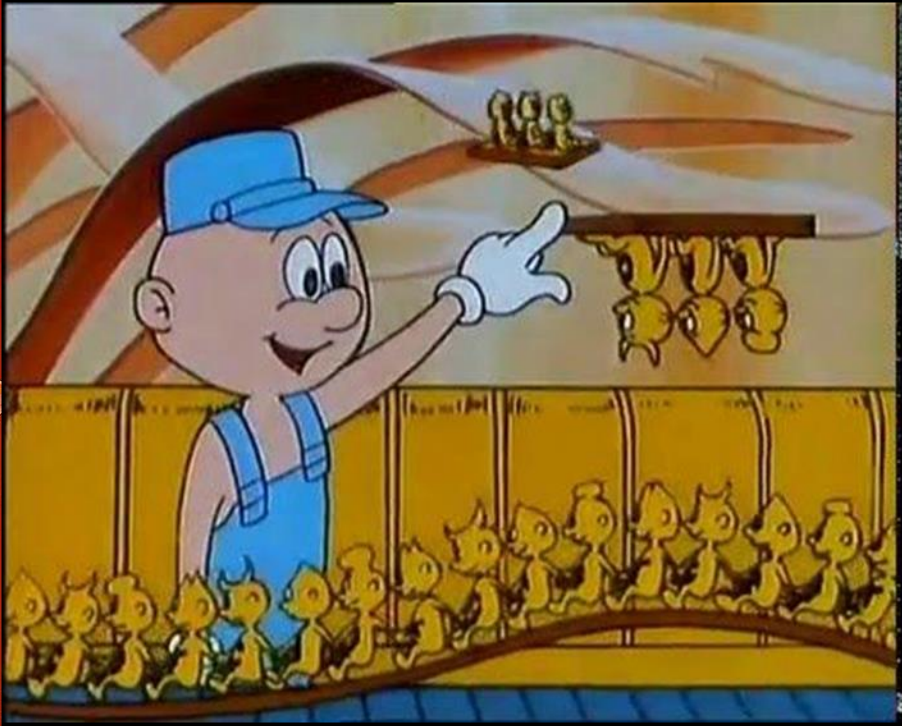


© 2011 Pearson Education, Inc.

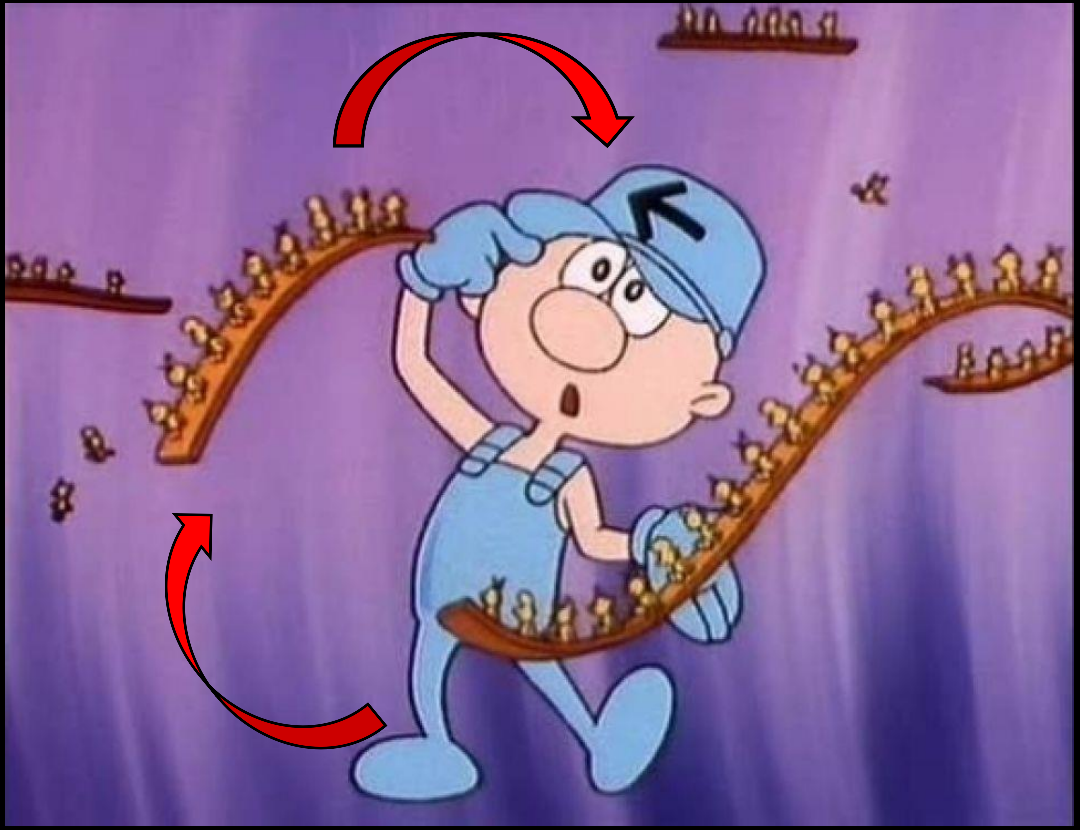
This focus however is quite specific: evolution is seen firstly via the prism of relatedness



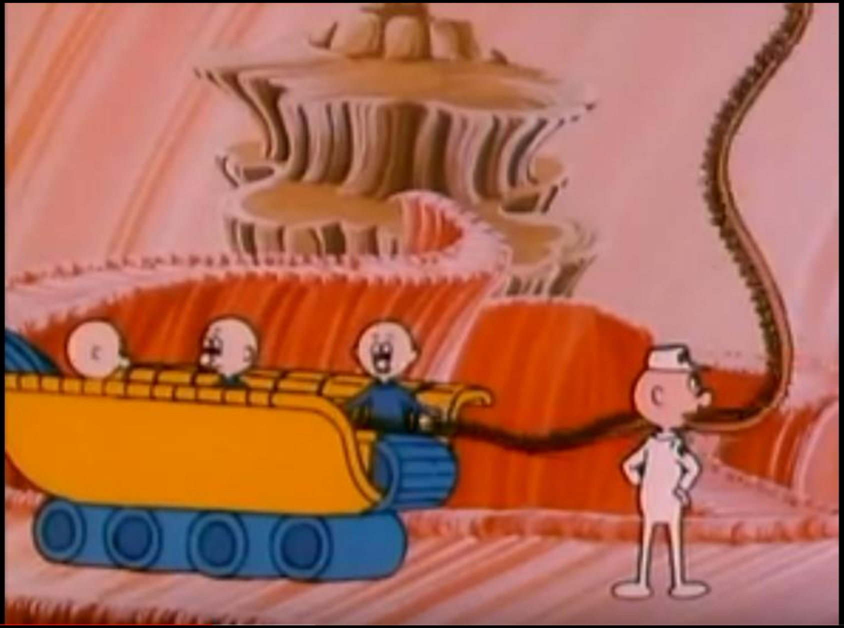
Yet, the biological world is not only about relatedness: it is also about organisational complexity



And explaining the evolution of biological organisations is a different, broader issue than inferring organismal relatedness.



For example, it takes an organisation to express a gene and produce a protein at the right time



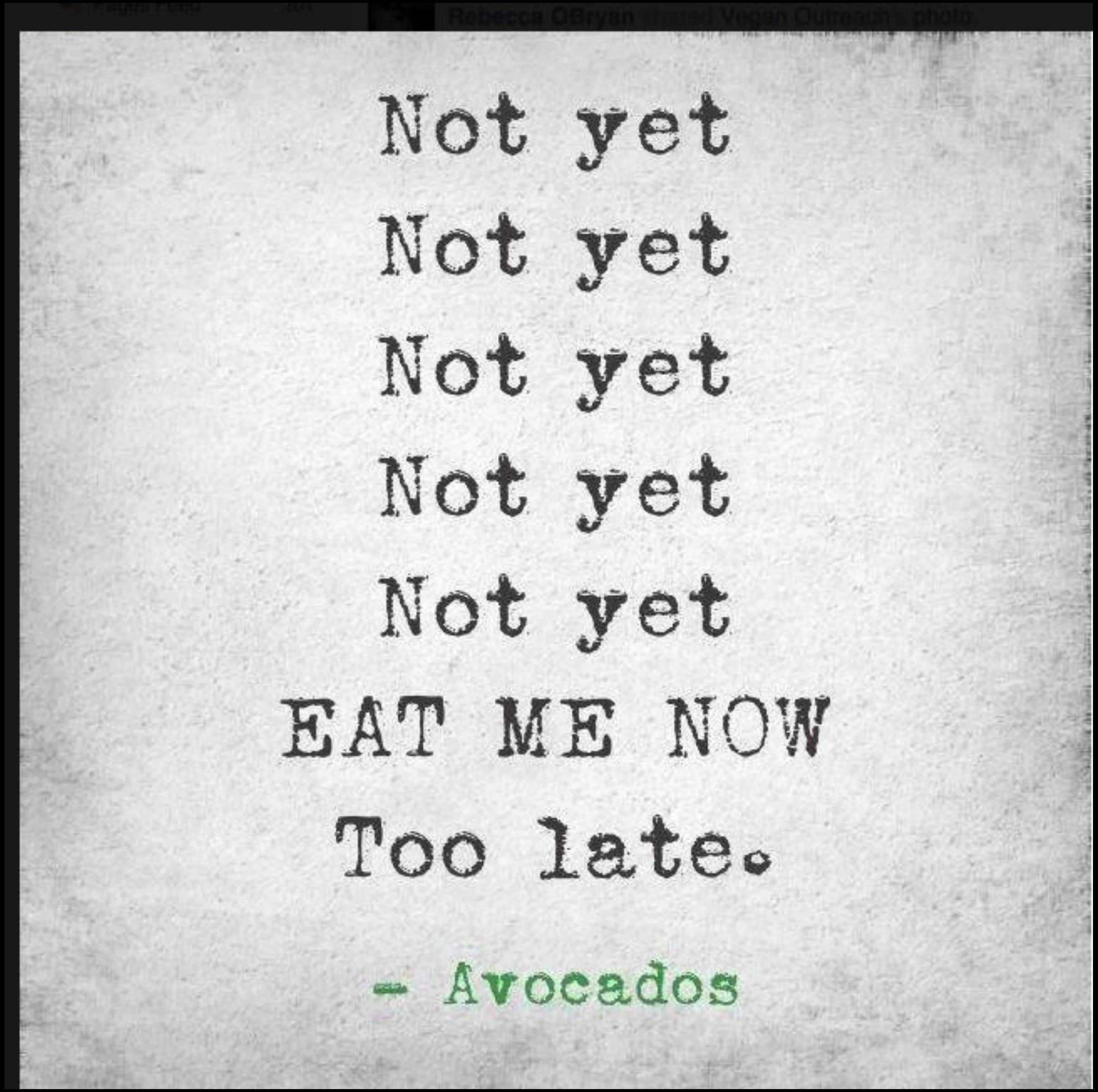
DNA

↓

RNA

↓

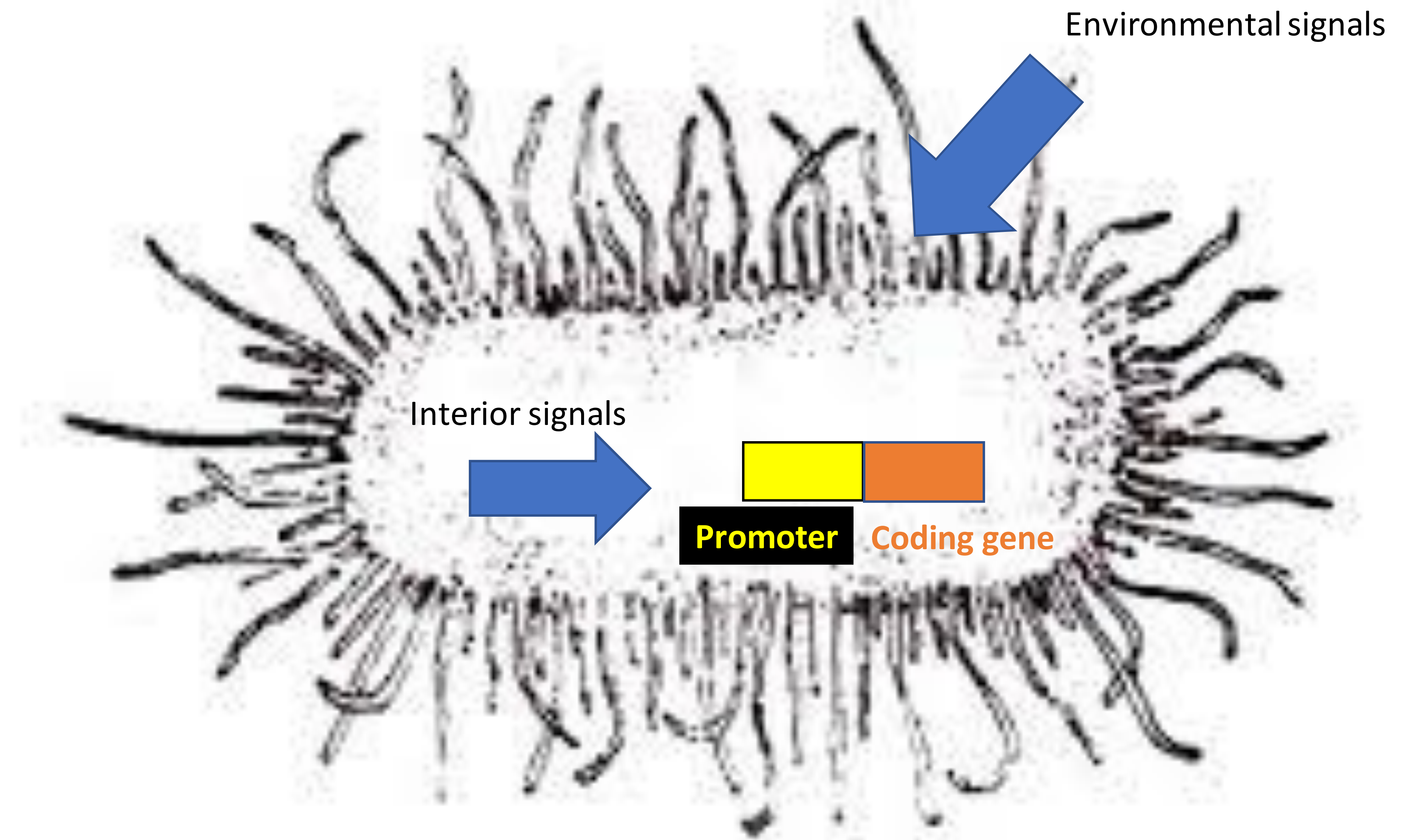
Proteins



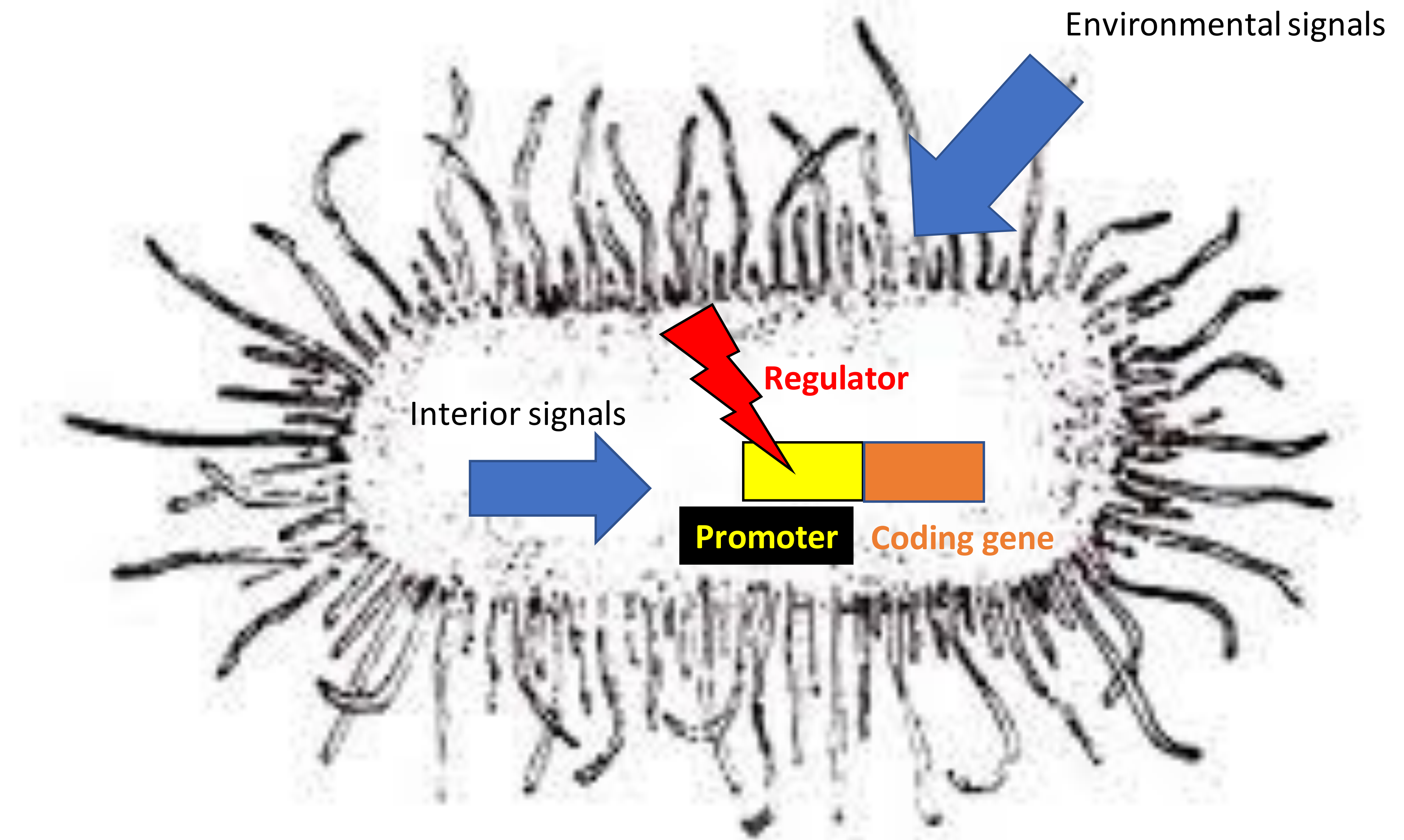
Not yet
Not yet
Not yet
Not yet
Not yet
EAT ME NOW
Too late.

- Avocados

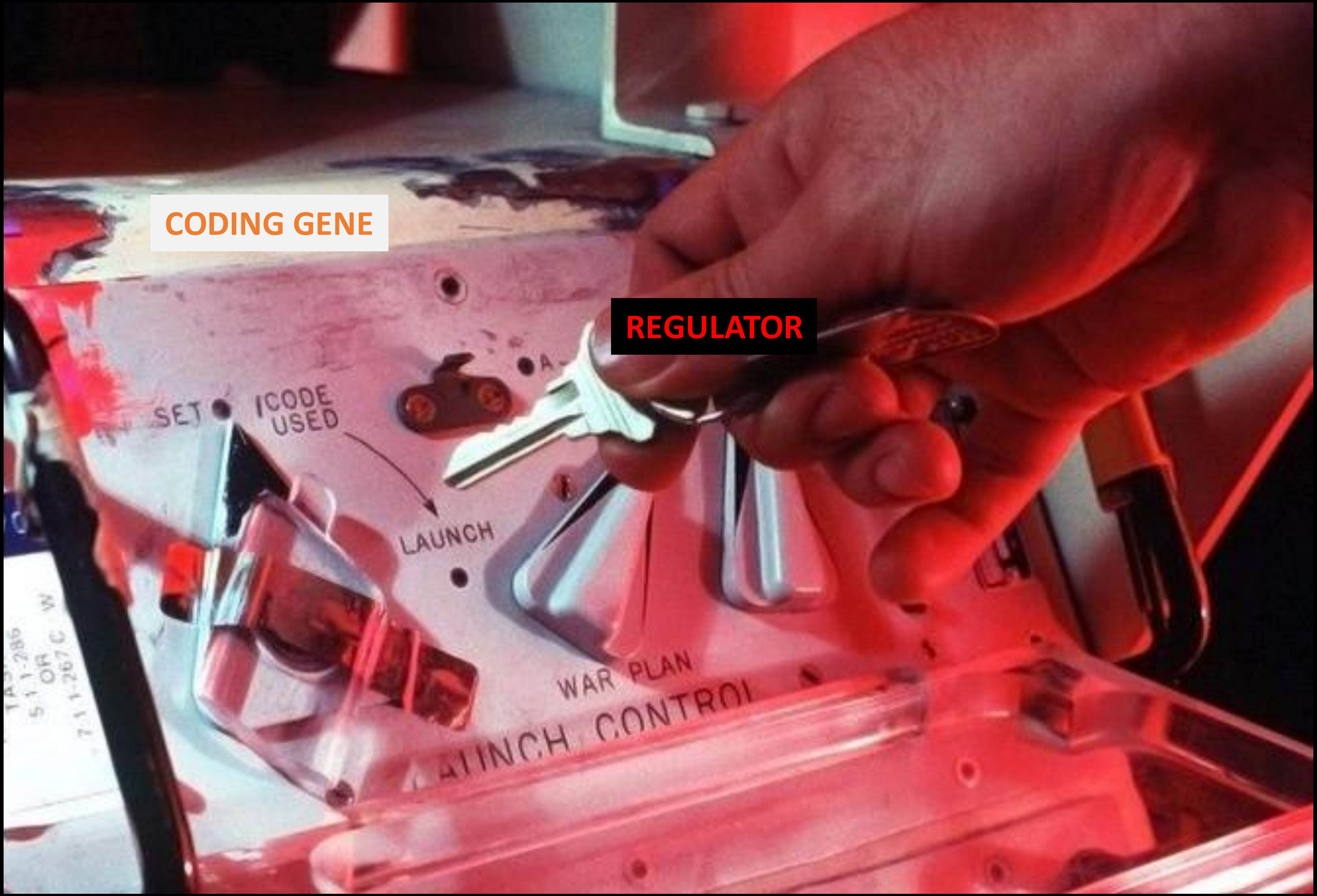
Gene expression depends from promoters.



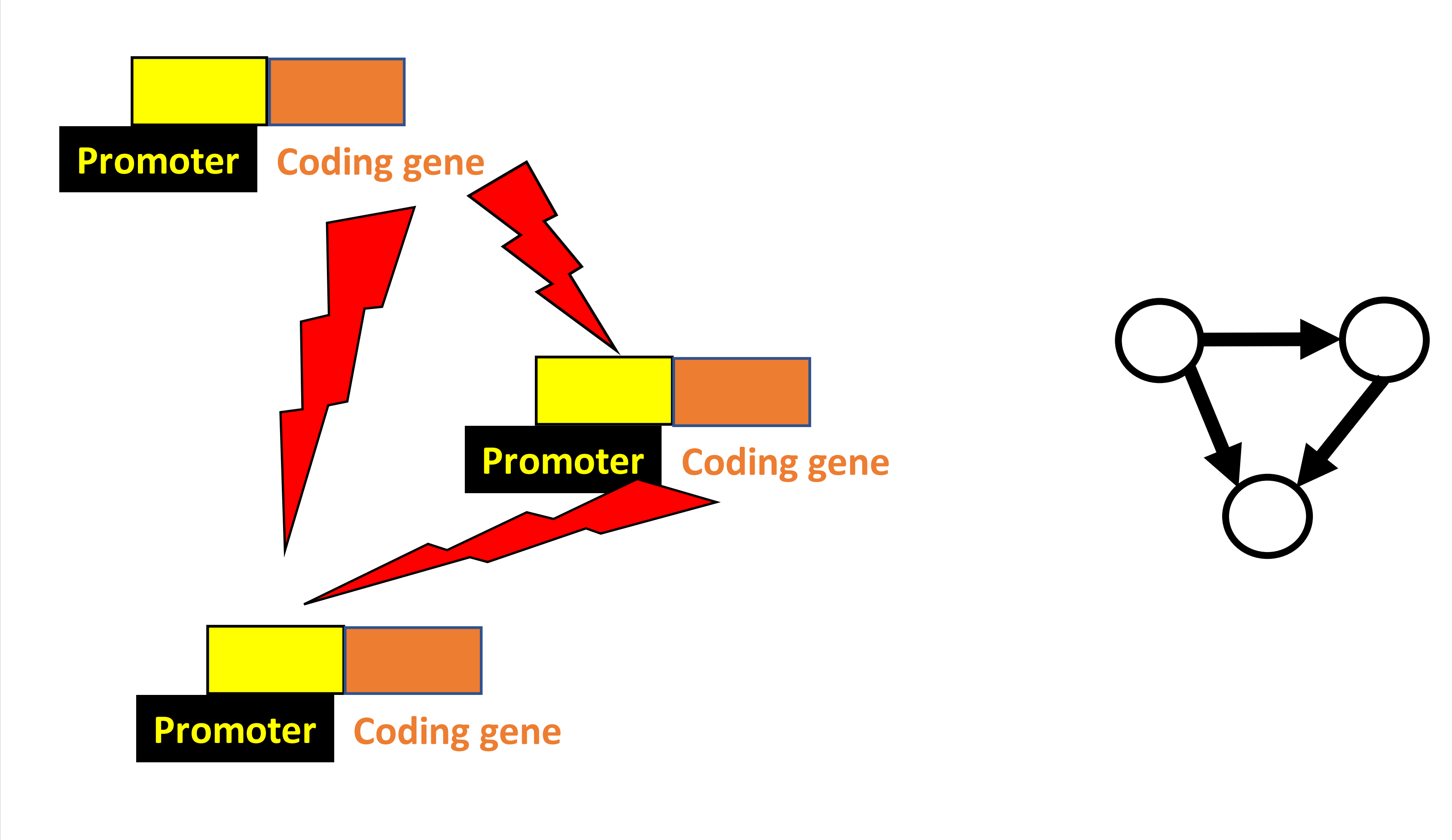
Gene expression depends from promoters and from regulators.



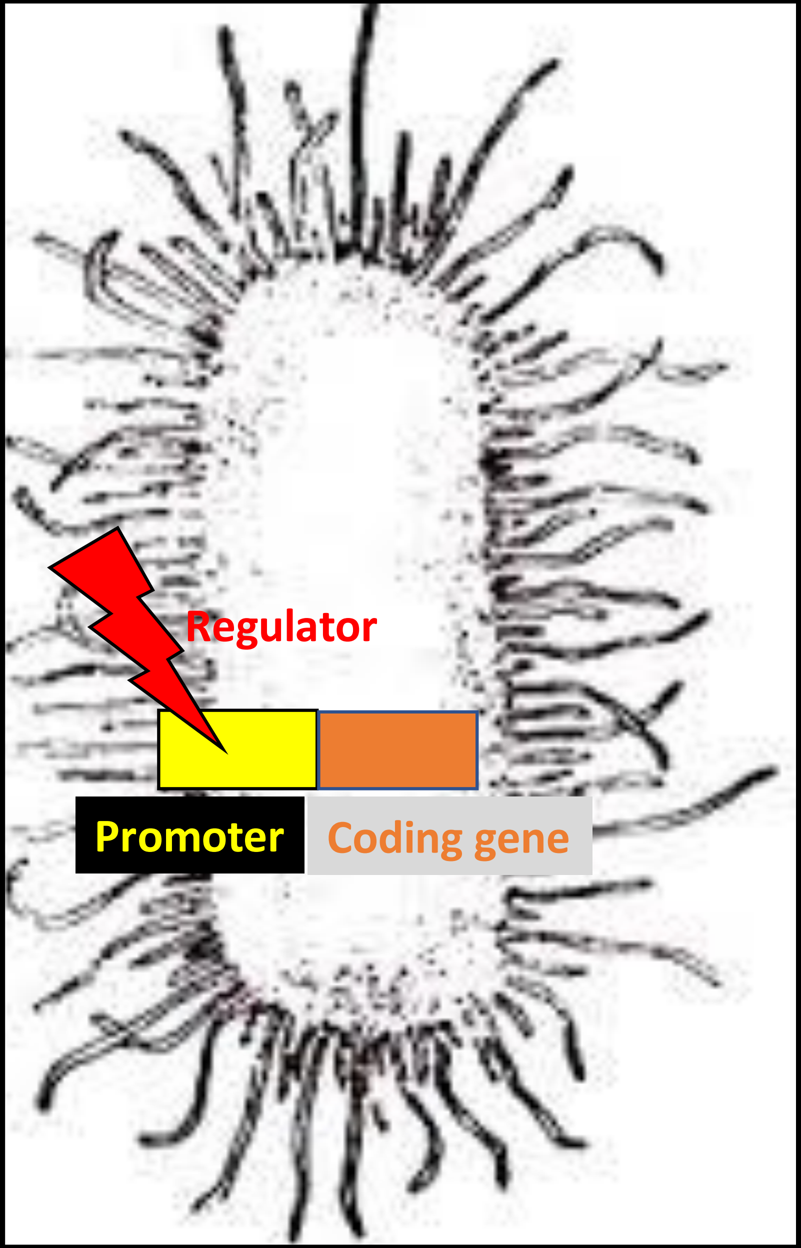
Interestingly, even regulators are regulated.

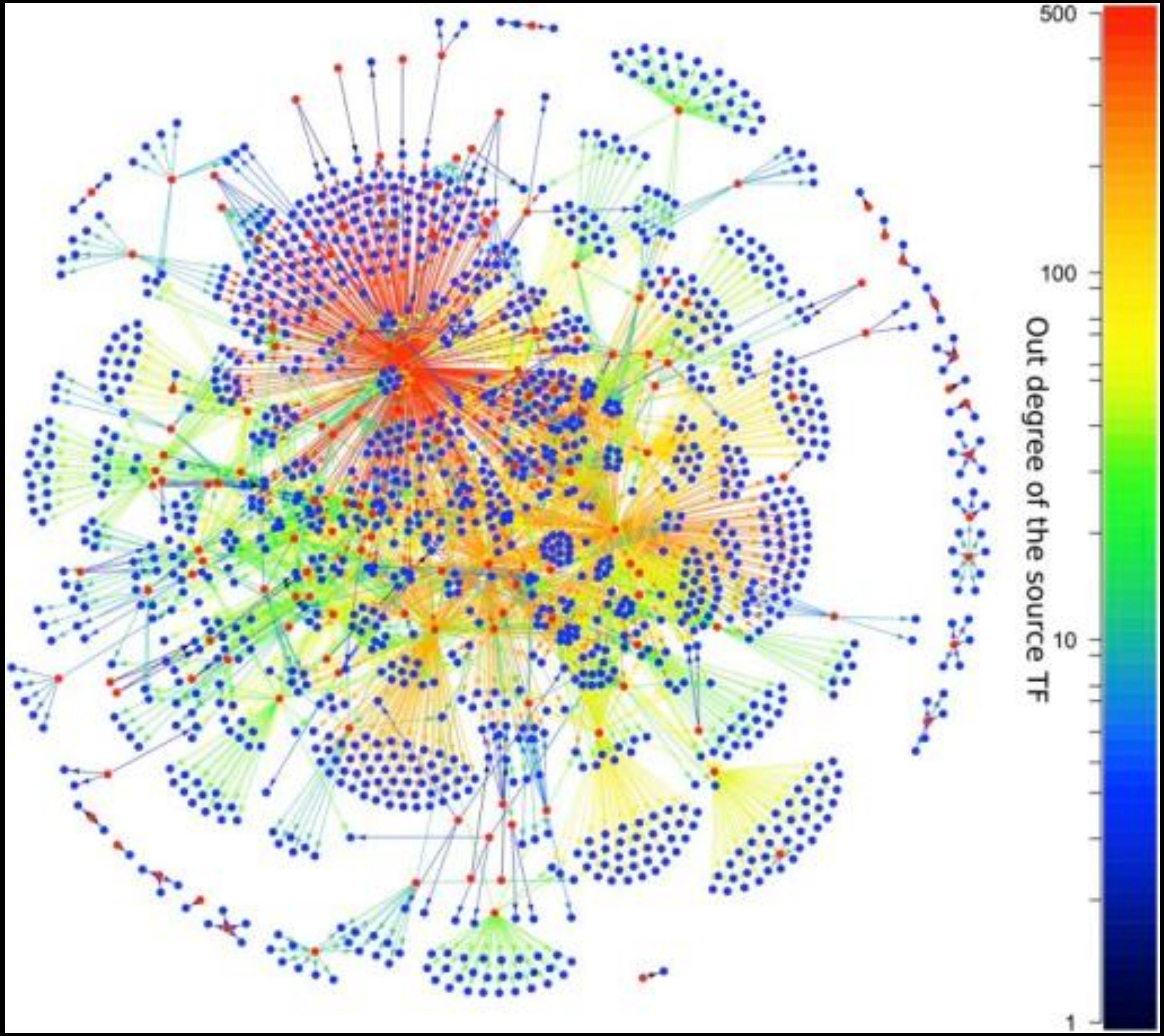


This phenomenon is modeled as a gene regulatory network (GRN).



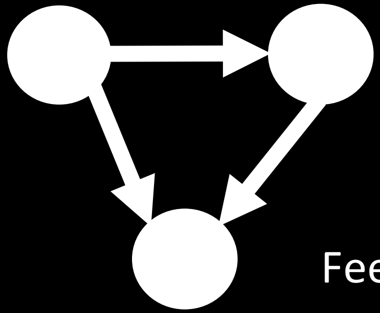
The gene regulatory network of *E. coli*





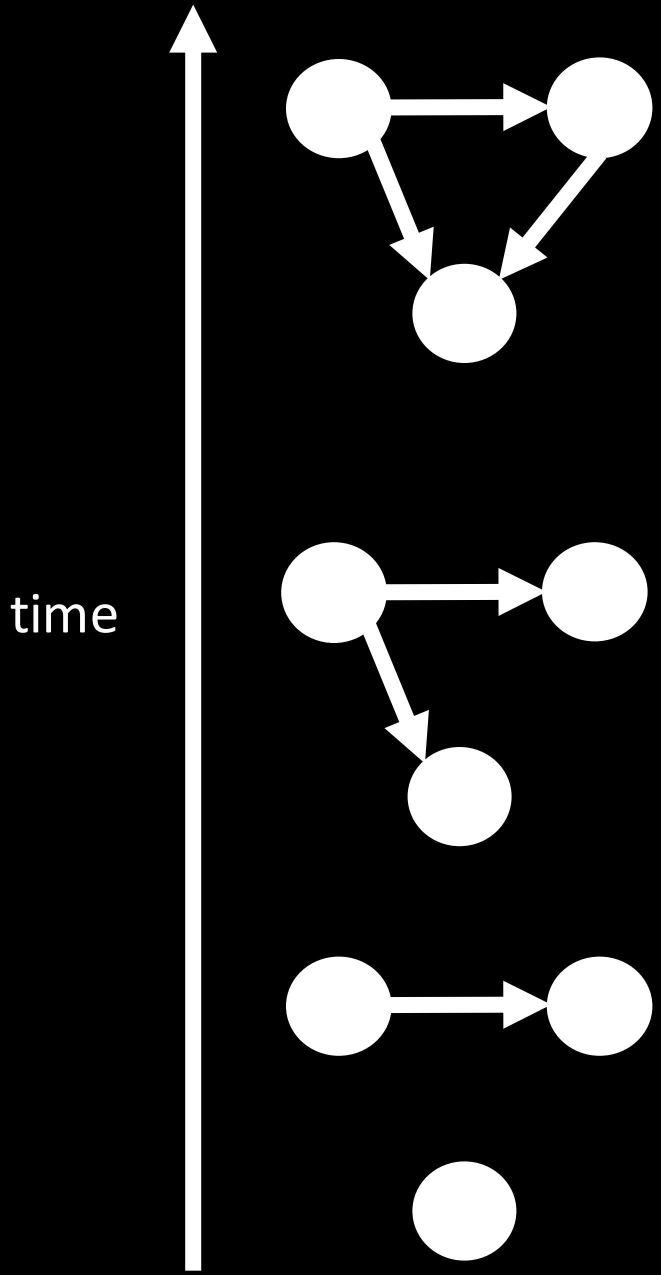
Regulators Other genes

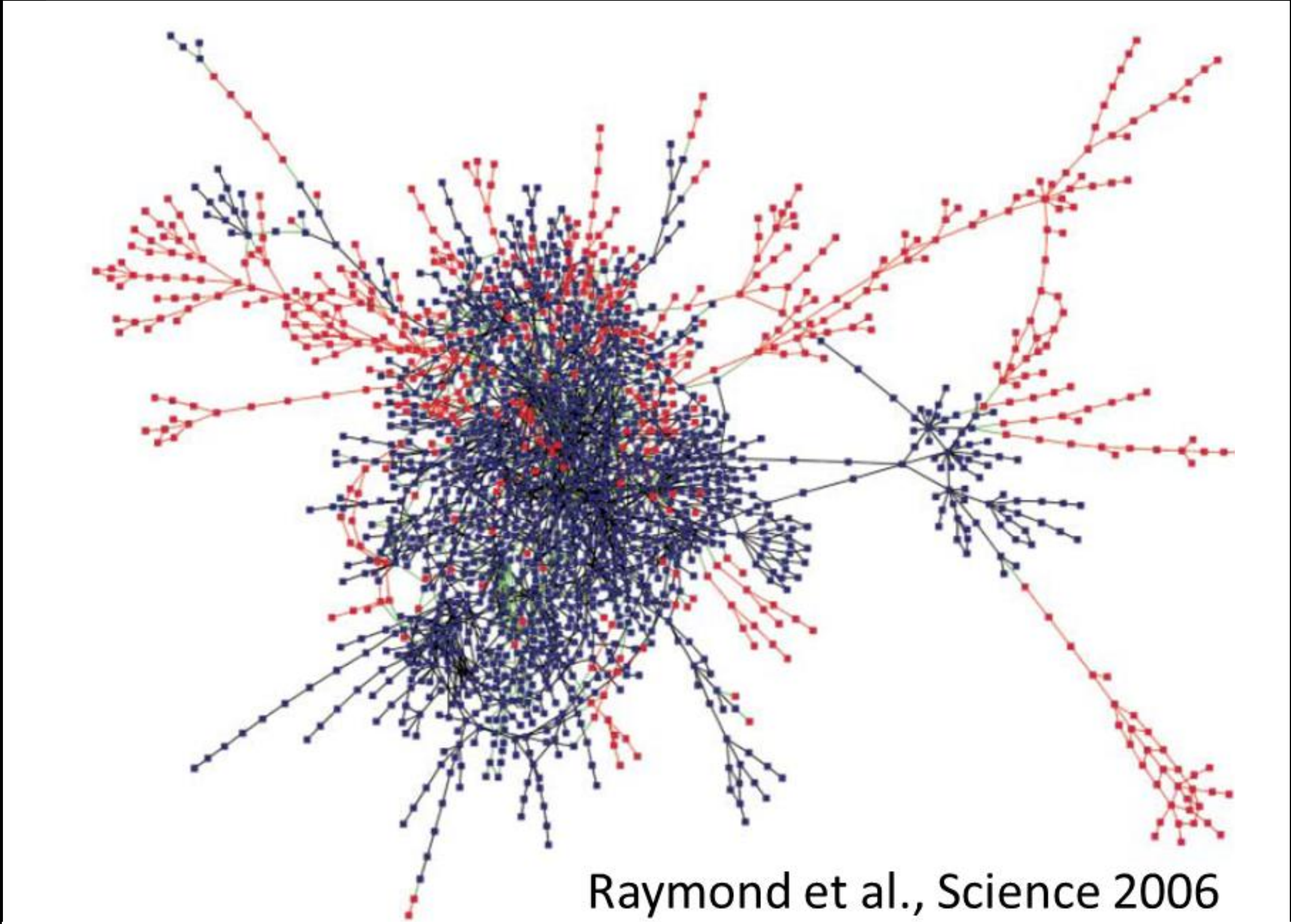
Topological analyses of such networks unravel informative patterns.



Feed forward loop

How architectures of the networks making the cells evolved: a new issue in evolutionary biology.





Raymond et al., Science 2006

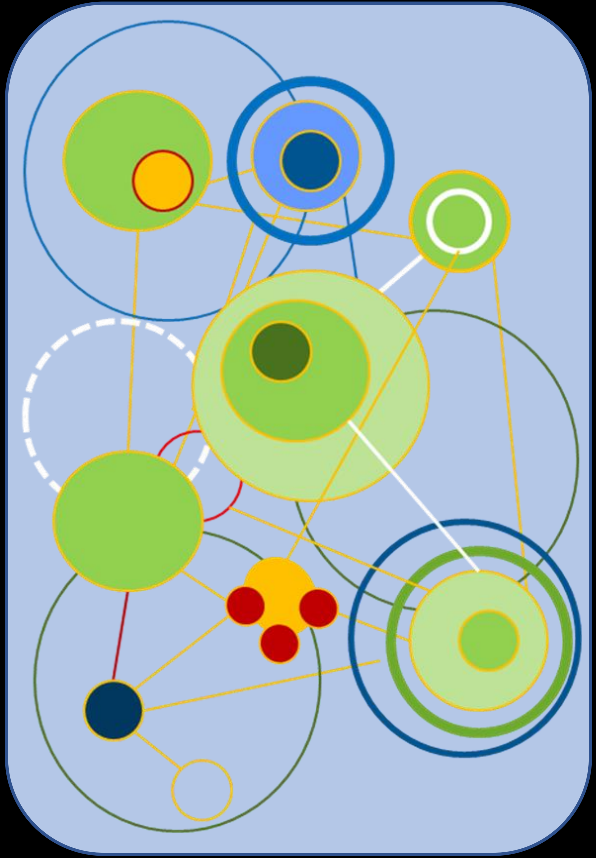
Metabolic network of cells Anoxic metabolites +reactions Oxic metabolites +reactions

Likewise, beyond the organismal scale: explaining the evolution of ecosystems is also a different, broader issue than inferring species relatedness.

Image credit: Davide Bonadonna.

Rather than outcomes of a tree-like evolution, network allows to consider biological entities as:

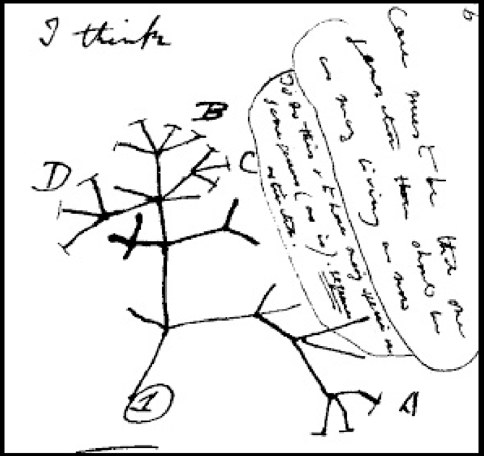
- (transient) <u>outcomes of various dynamic interactions</u>

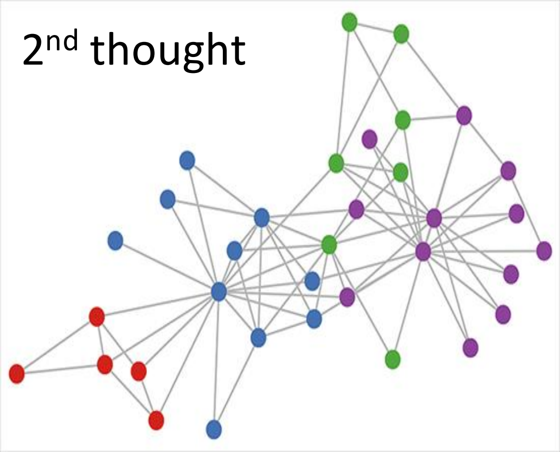


- <u>components of dynamic systems</u>

While trees target lineage relatedness/simple aspects of evolution, networks can also study organisations/complex aspects of evolution

- Networks offer alternative explanations of the origins of traits
- Network can describe unexpected evolved phenomena
- Networks have a potential to expand the scope of the evolutionary theory

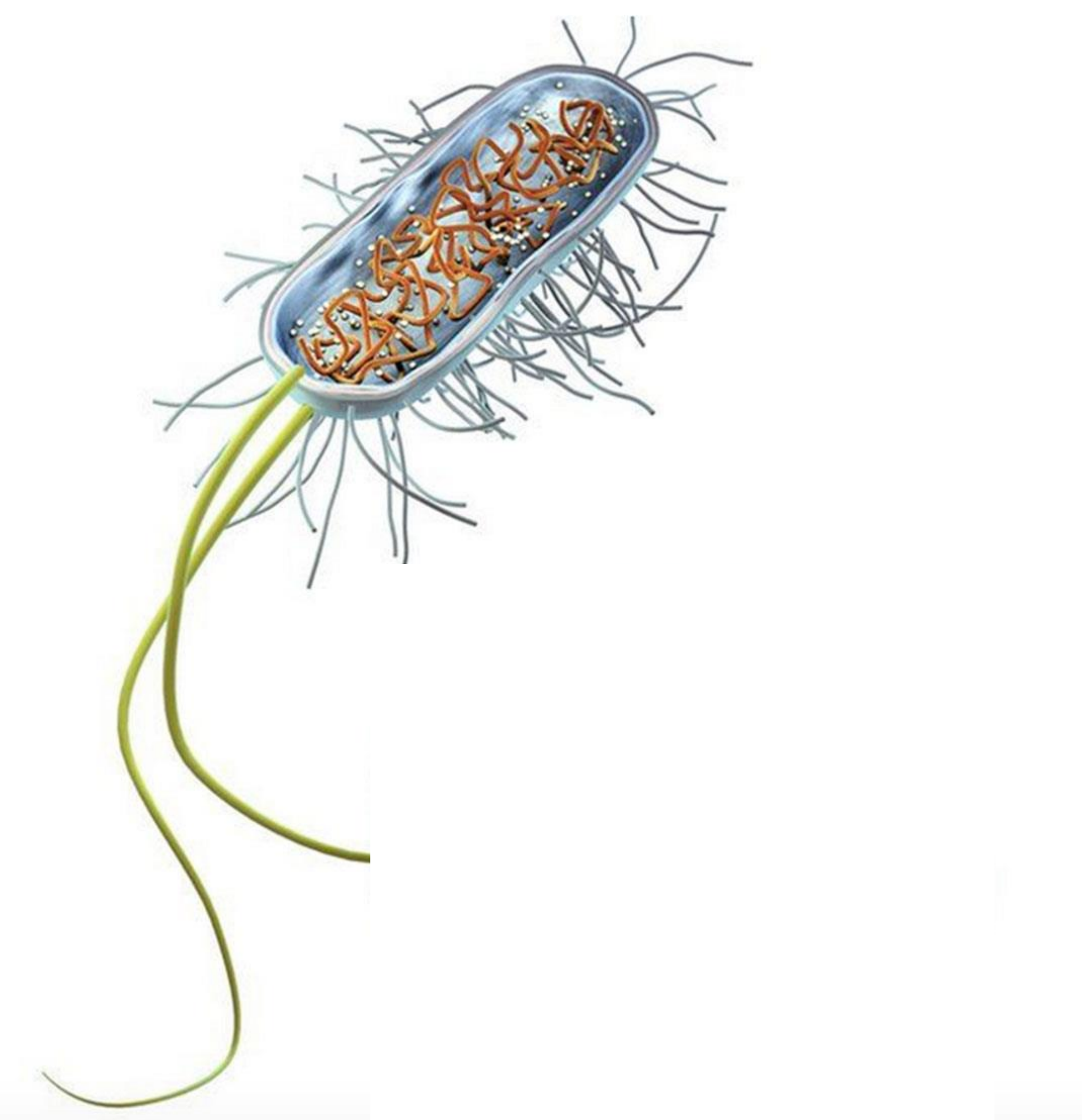




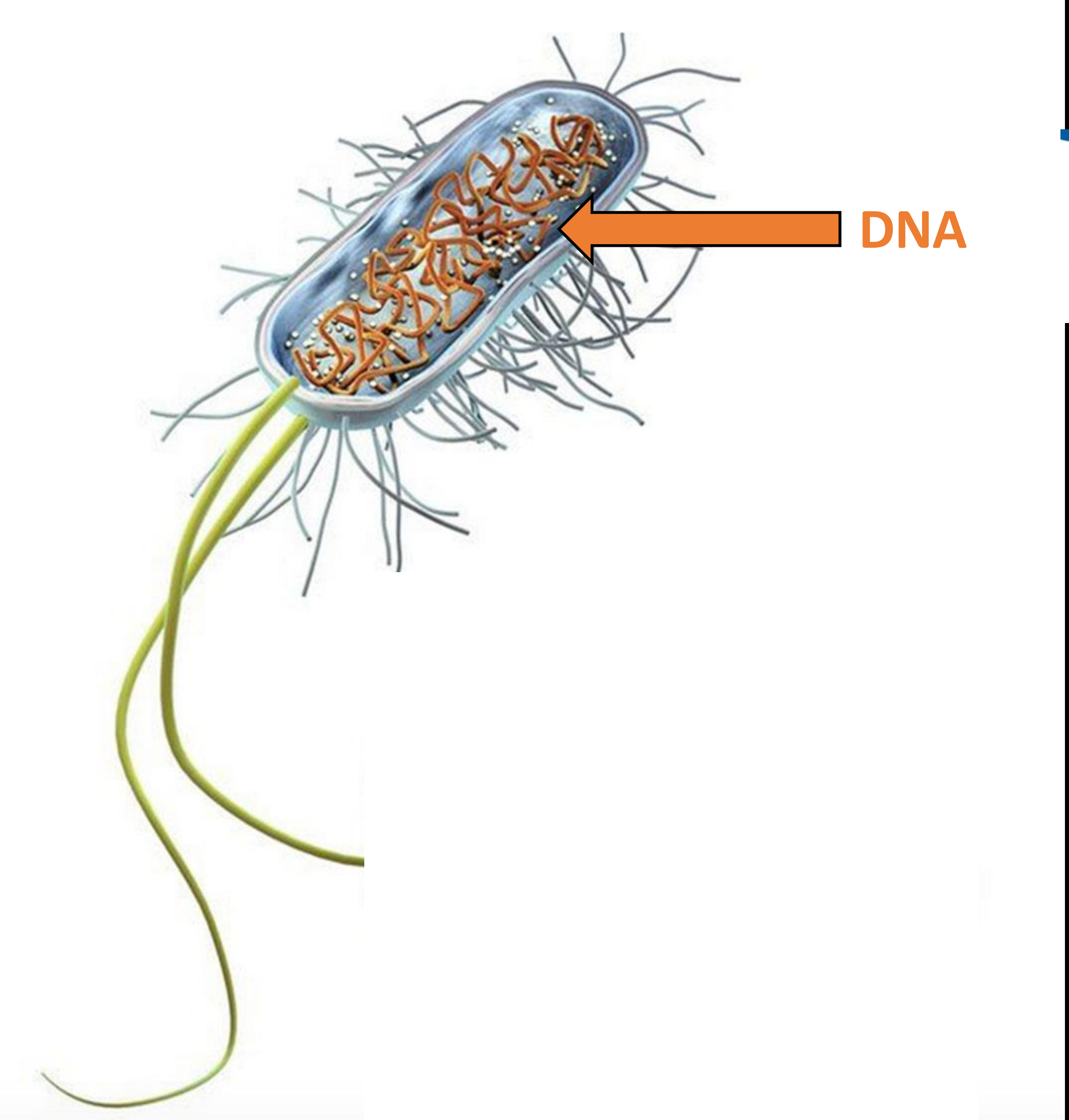
- On the origins of traits: the case of genes

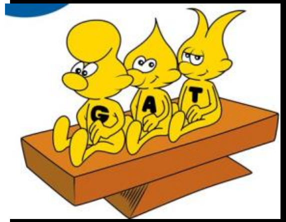
Where do novel genes come from?

In asexual lineages, the standard evolutionary theory would suggest that new genes derive from former genes by point mutations.

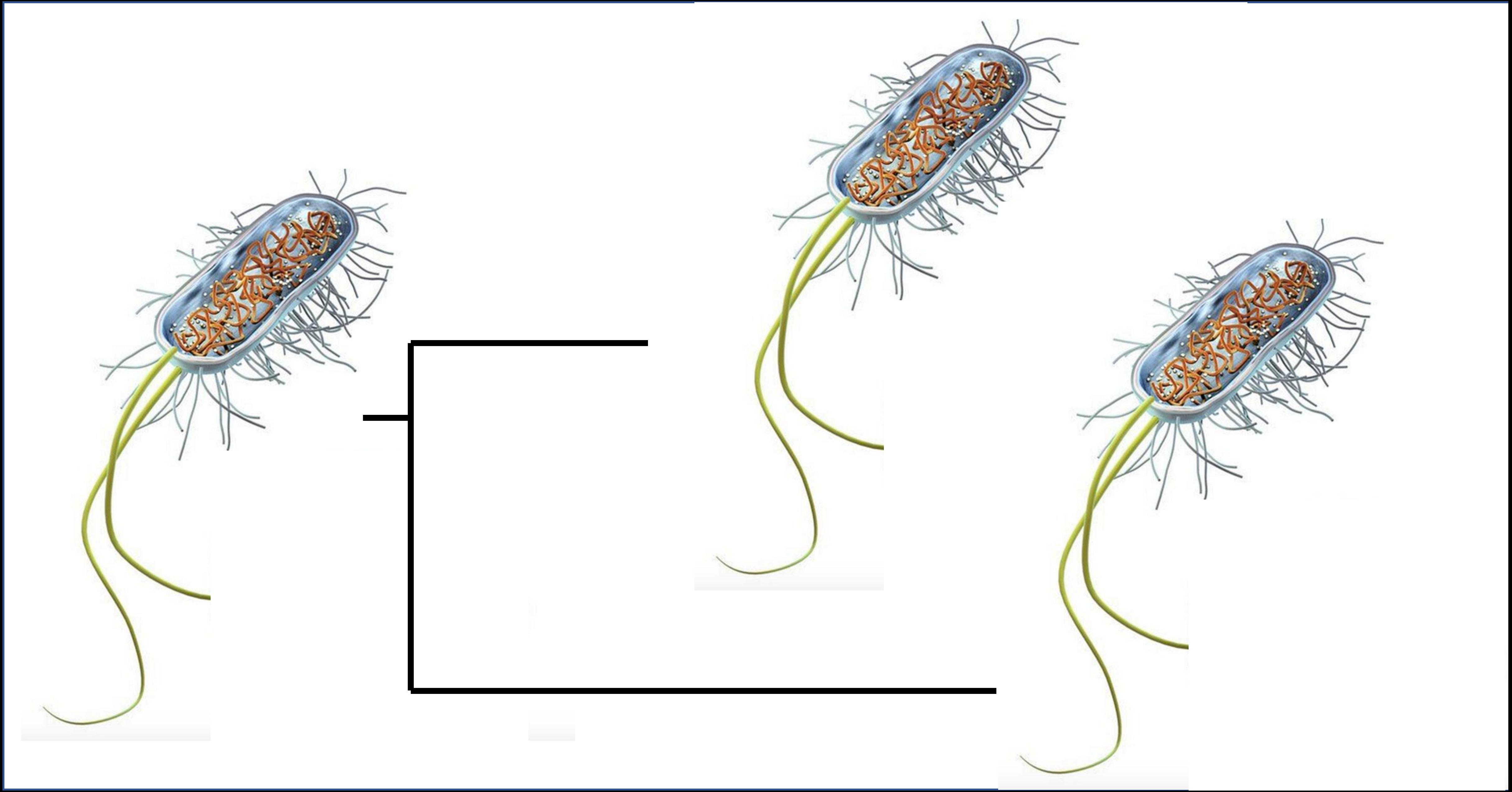


In asexual lineages, the standard evolutionary theory would suggest that new genes derive from former genes by point mutations.

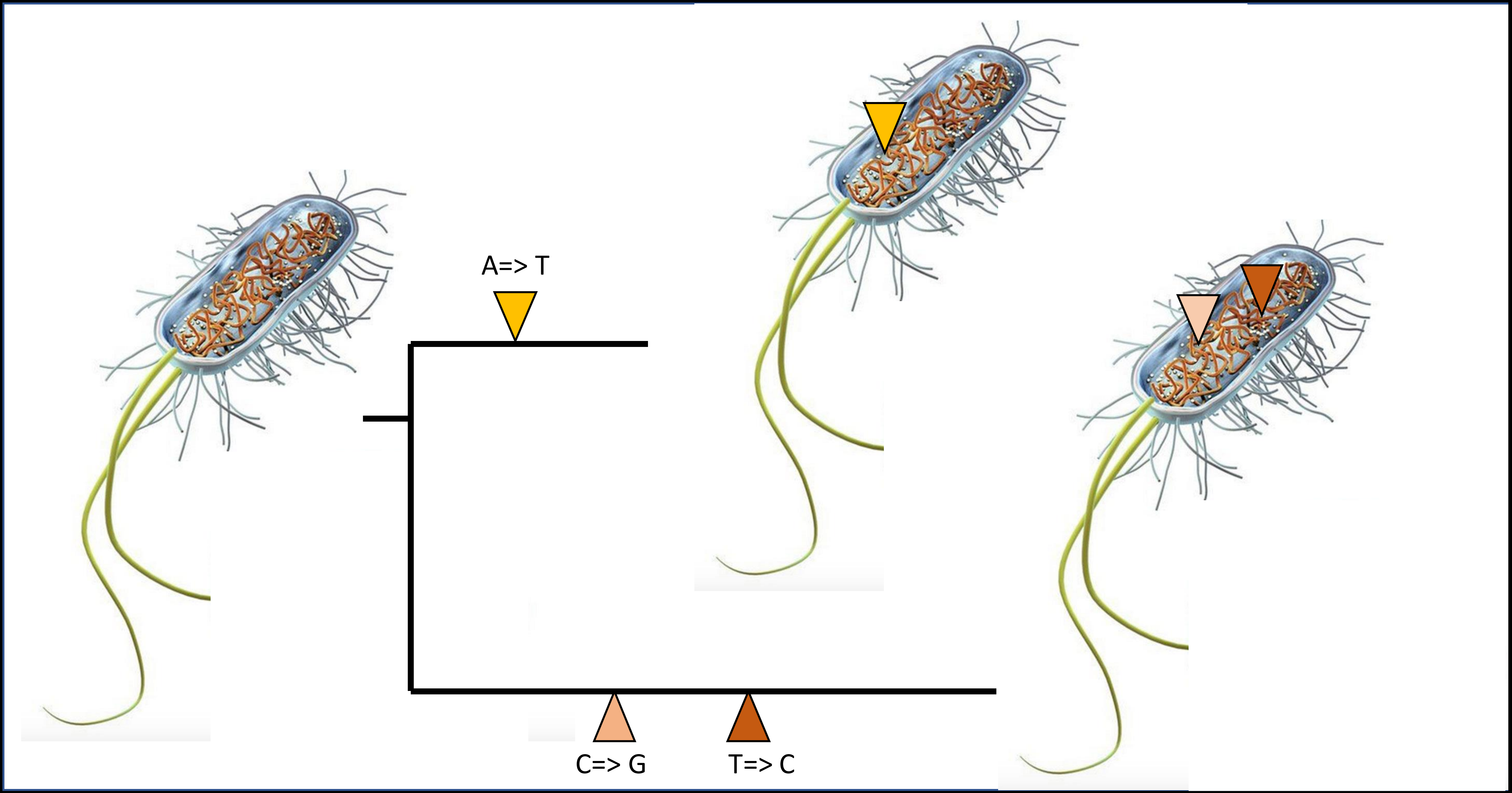




One may indeed think of such bacteria as entities ideally evolving by a tree-like process.



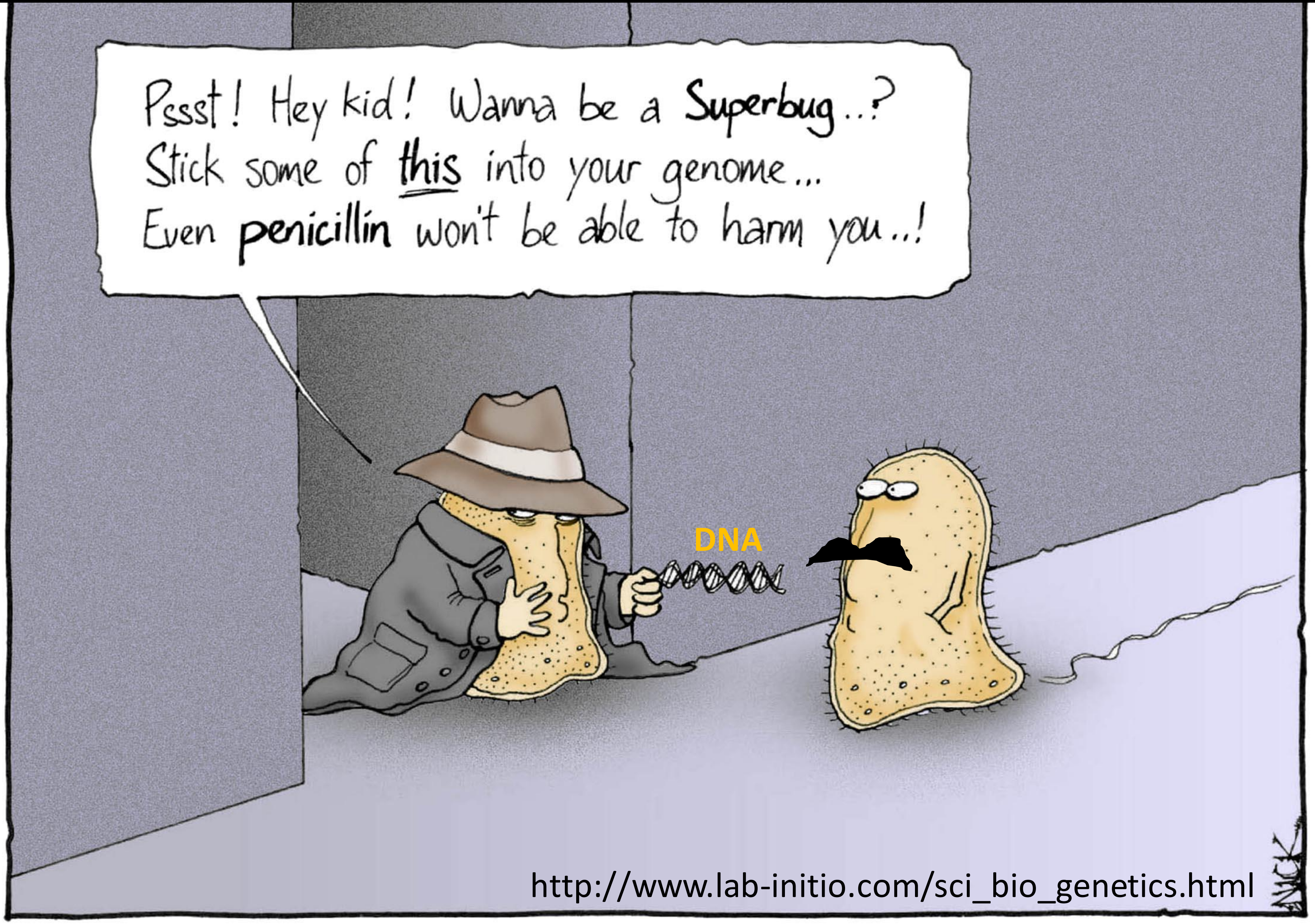
One may indeed think of such bacteria as entities ideally evolving by a tree-like process.



Yet, even for such taxa, network-thinking brings forward additional processes:

- **Horizontal Gene Transfer** (at odds with standard theory, since inheritance does not only go from one ancestor to its descendants!)
- **Domain combinations** (highligthing the reticulate evolution of the genes within organisms)

HGT is the process by which an organism receives genes from a neighbour, rather than from its immediate ancestor.

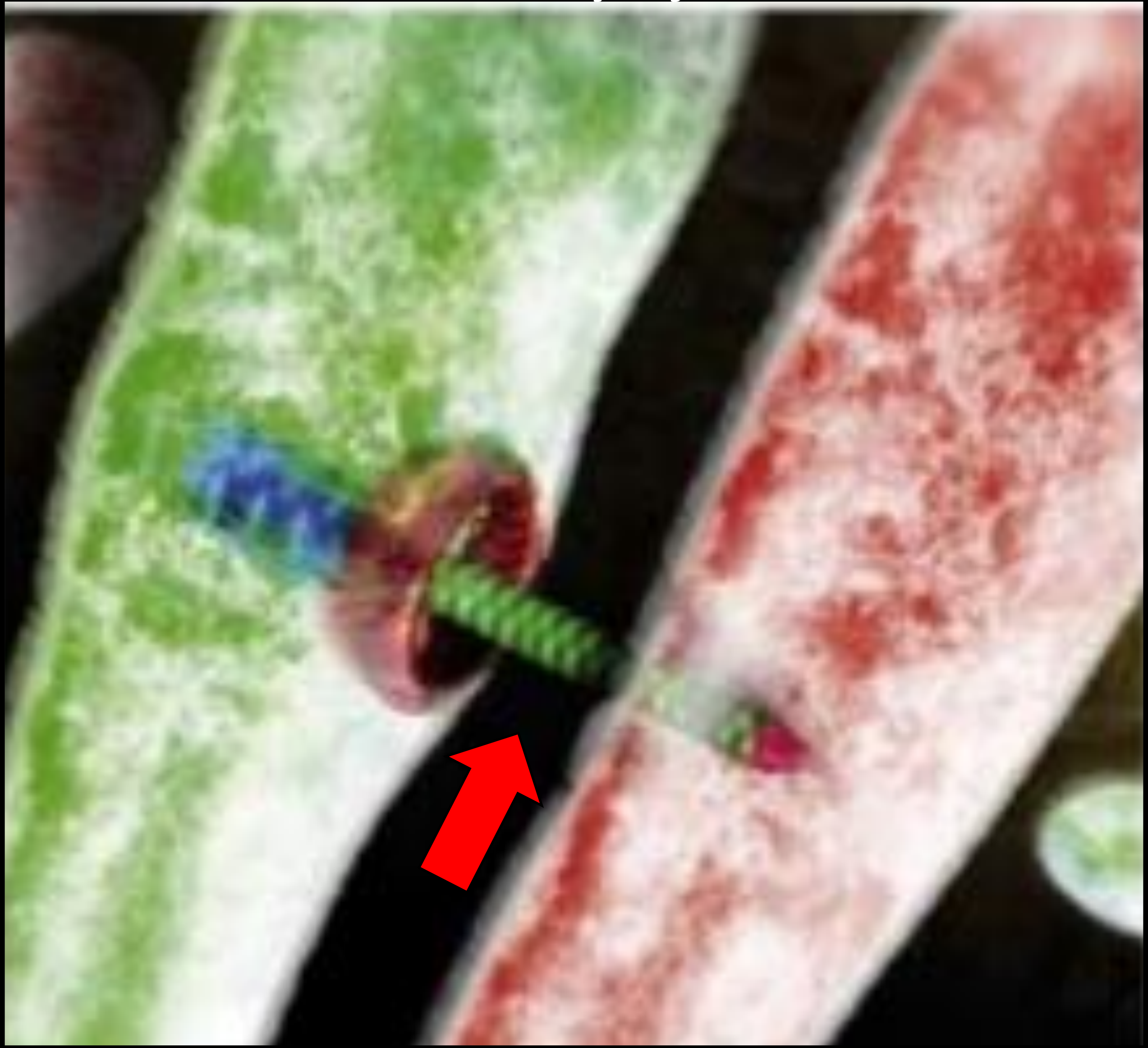


http://www.lab-initio.com/sci_bio_genetics.html

It was on a short-cut through the hospital kitchens that Albert was first approached by a member of the Antibiotic Resistance.

There are many roads to HGT

One cell kills another cell by injection of toxic substance

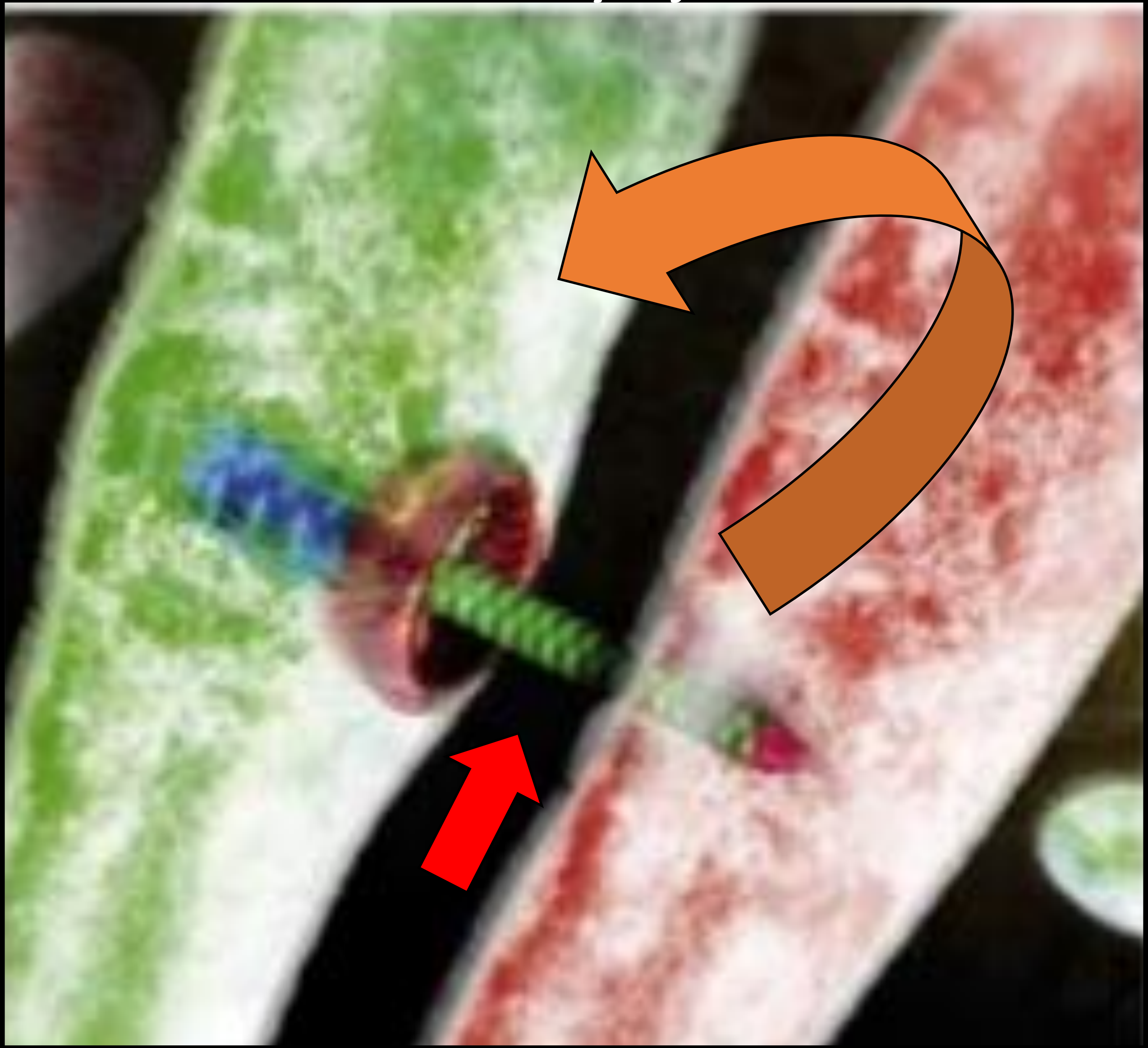


Type VI secretion system

Borgeaux, Science, 2015
Thomas, Mbio, 2017

There are many roads to HGT

One cell kills another cell by injection of toxic substance

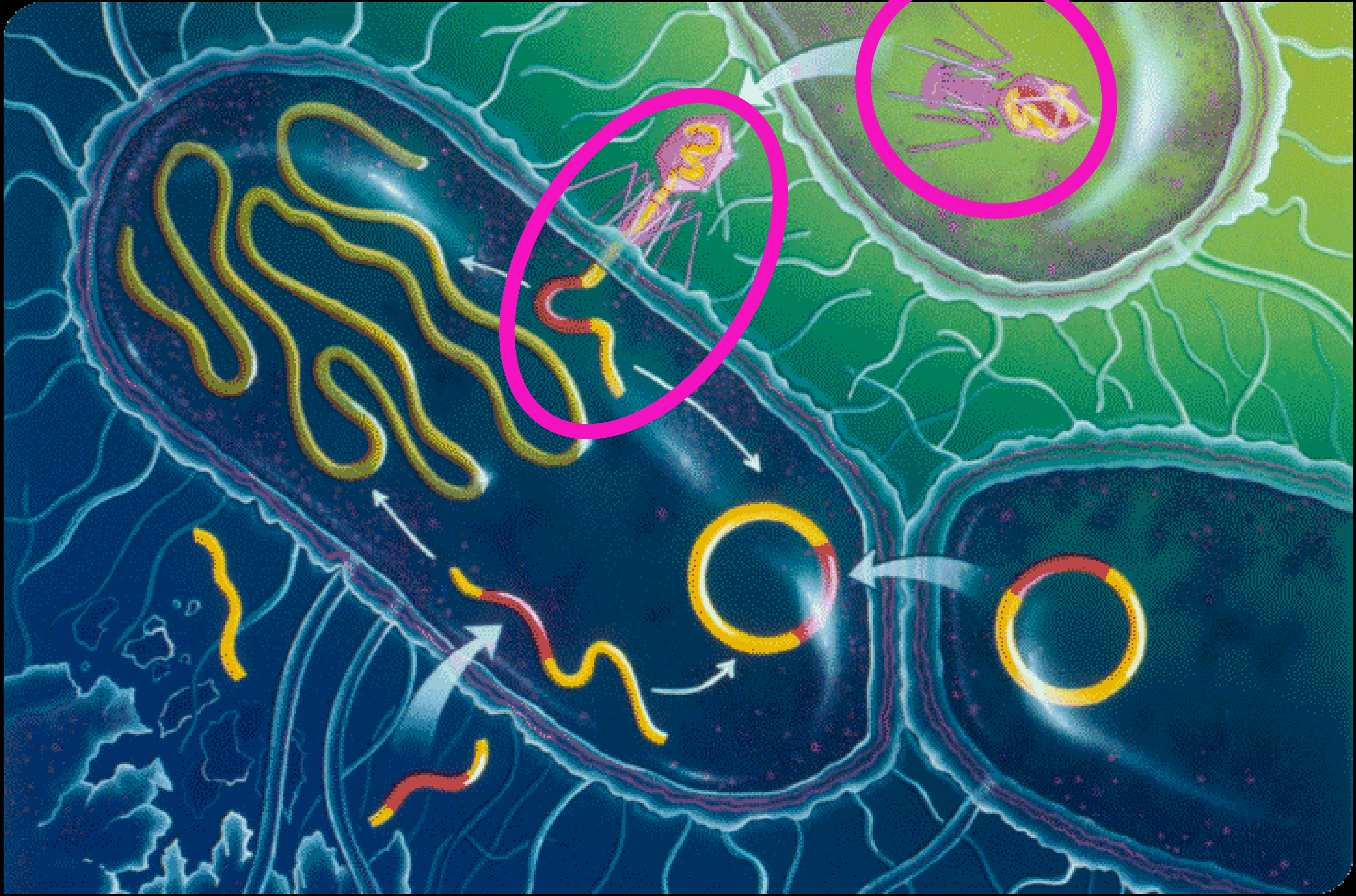


Type VI secretion system

Then the killer steals the DNA of its victim, and integrates bit of it in its own DNA

Borgeaux, Science, 2015
Thomas, Mbio, 2017

Horizontal Gene Transfer can also be mediated by mobile elements



- Viruses

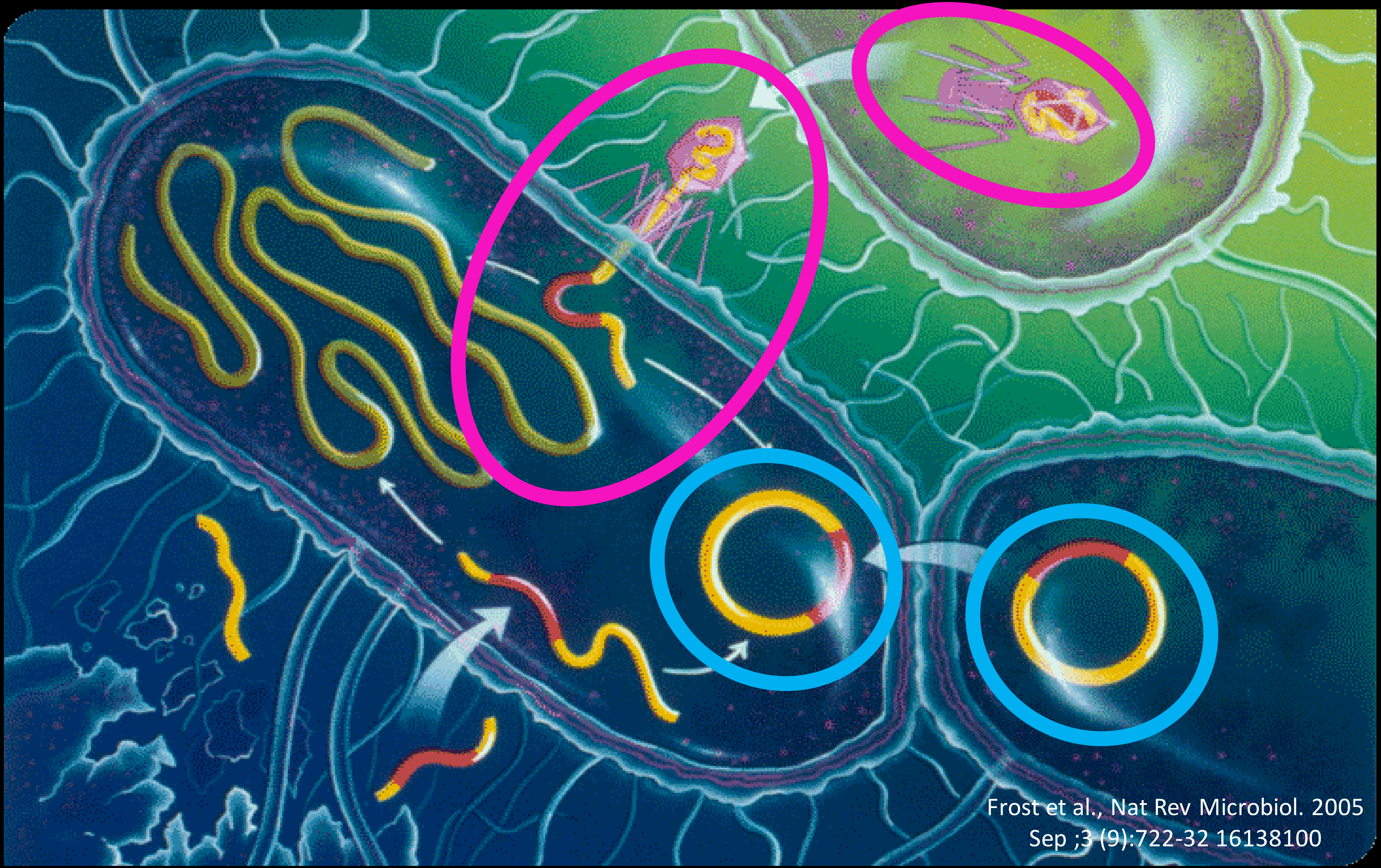
Frost et al., Nat Rev Microbiol. 2005 Sep ;3 (9):722-32 16138100

Horizontal Gene Transfer can also be mediated by mobile elements

- **Viruses**
- **Plasmids**

Frost et al., Nat Rev Microbiol. 2005 Sep ;3 (9):722-32 16138100

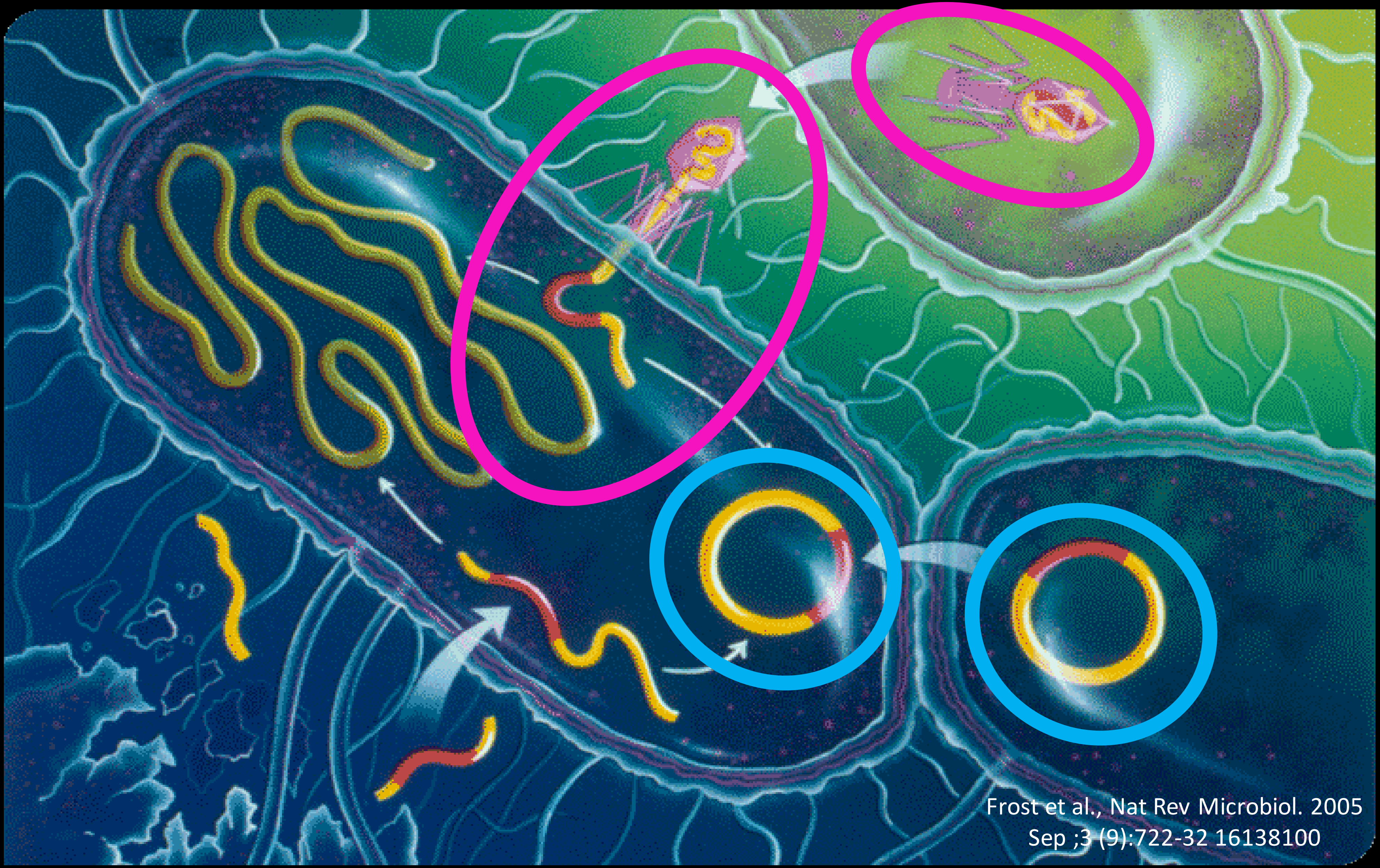
Numerous, abundant mobile genetic elements do not only move DNA between cells…



Frost et al., Nat Rev Microbiol. 2005 Sep ;3 (9):722-32 16138100

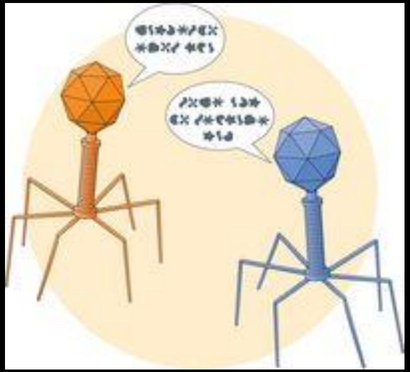
- 10^{30} viral particles realize 10^{25} infections/sec
- Plasmids move 1kb to a few Mb of DNA

Moreover, mobile elements can interact together in many ways



Frost et al., Nat Rev Microbiol. 2005 Sep ;3 (9):722-32 16138100

In 2017, it was discovered that viruses communicate when infecting cells!



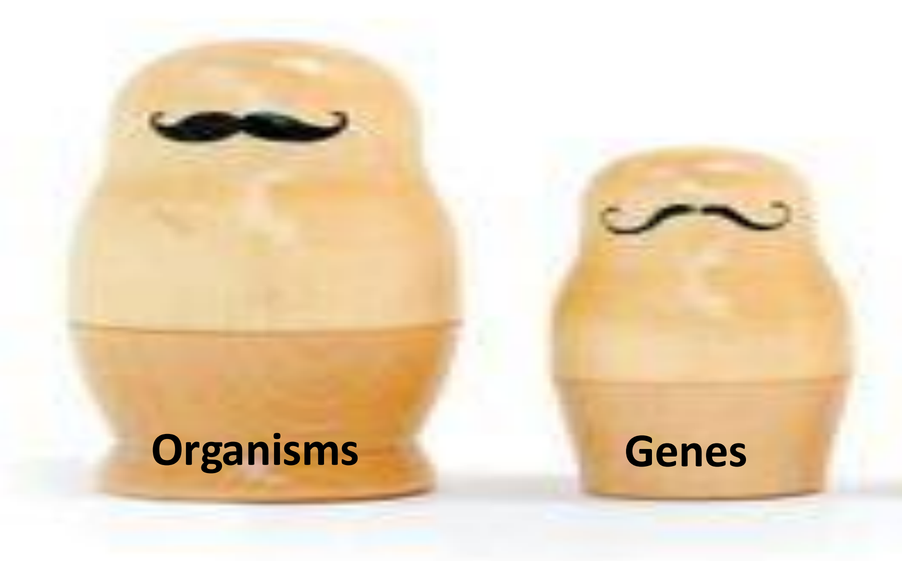
Erez *et al.*, Nature

***Arbitrium*, the language of viruses, determining whether they will hide in cells or kill them**

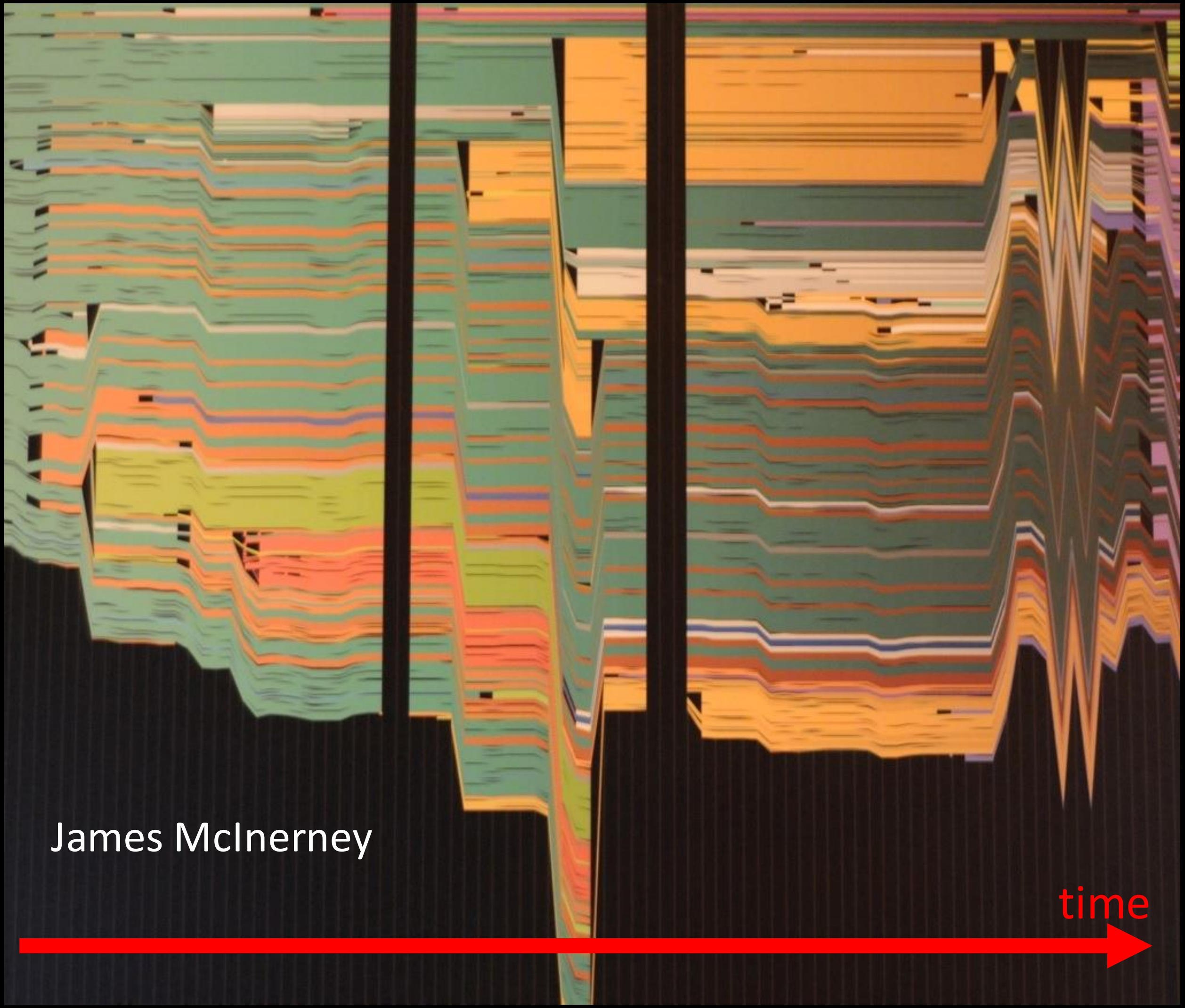
Another example, some viruses are superspreaders.

Keen *et al.* Mbio. 2016

Because organismal lineages and genes lineages can follow different roads, evolution must be studied at least at 2 levels.

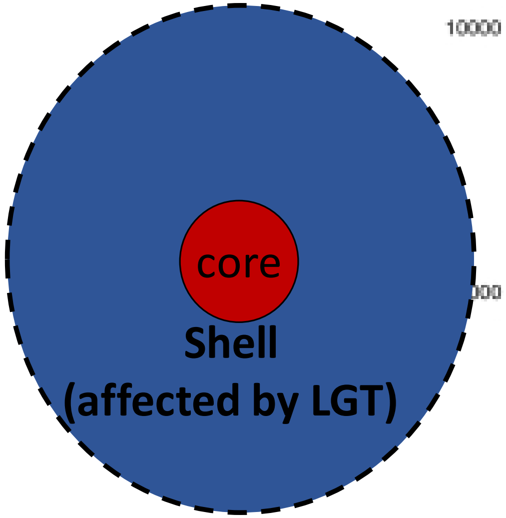


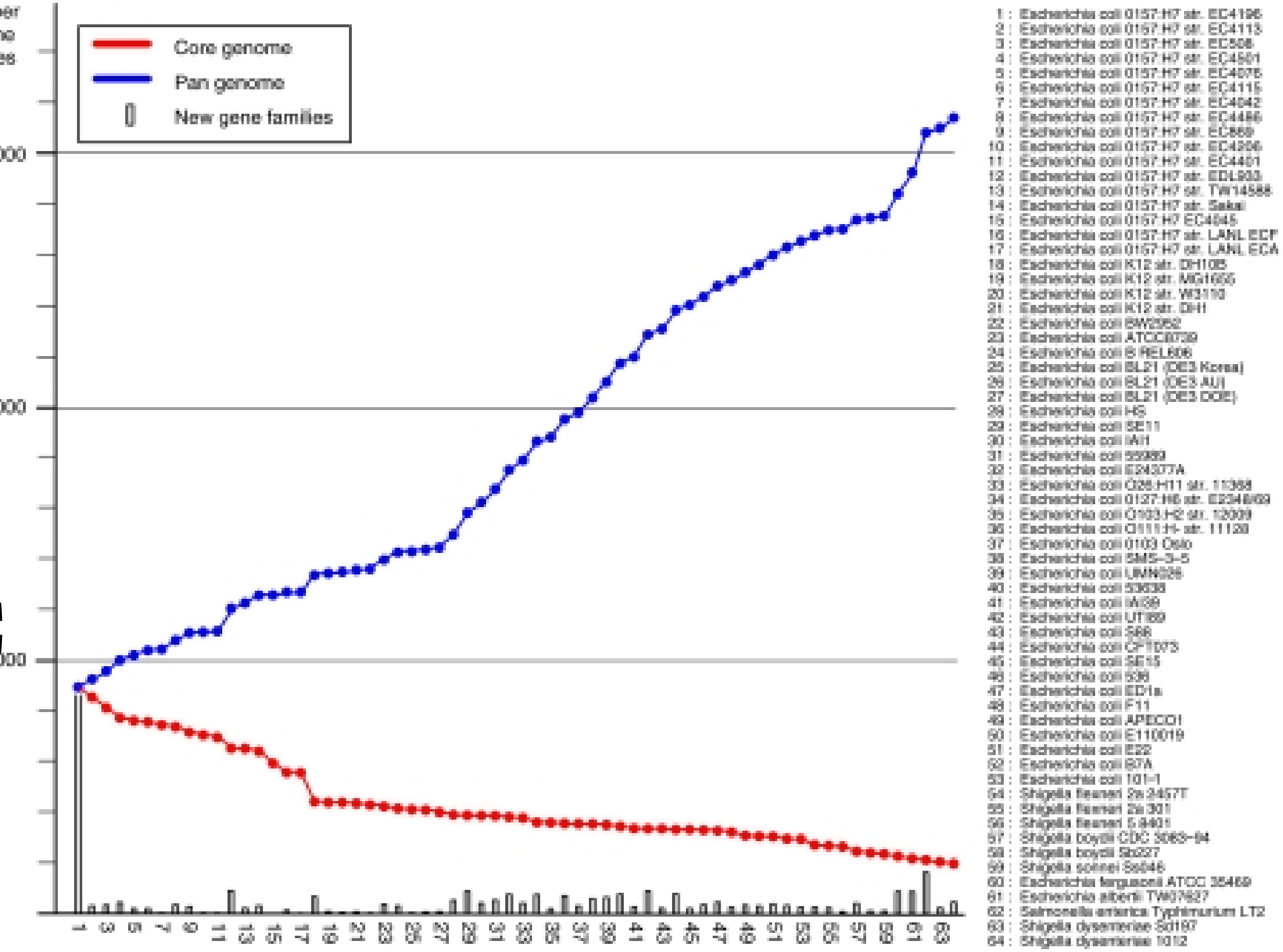
HGT means that variation and transmission also occur from outside the lineage, producing mosaic organisms



Most typically, producing bacterial groups with pangenomes.

Core genome to *E. coli* = 6%



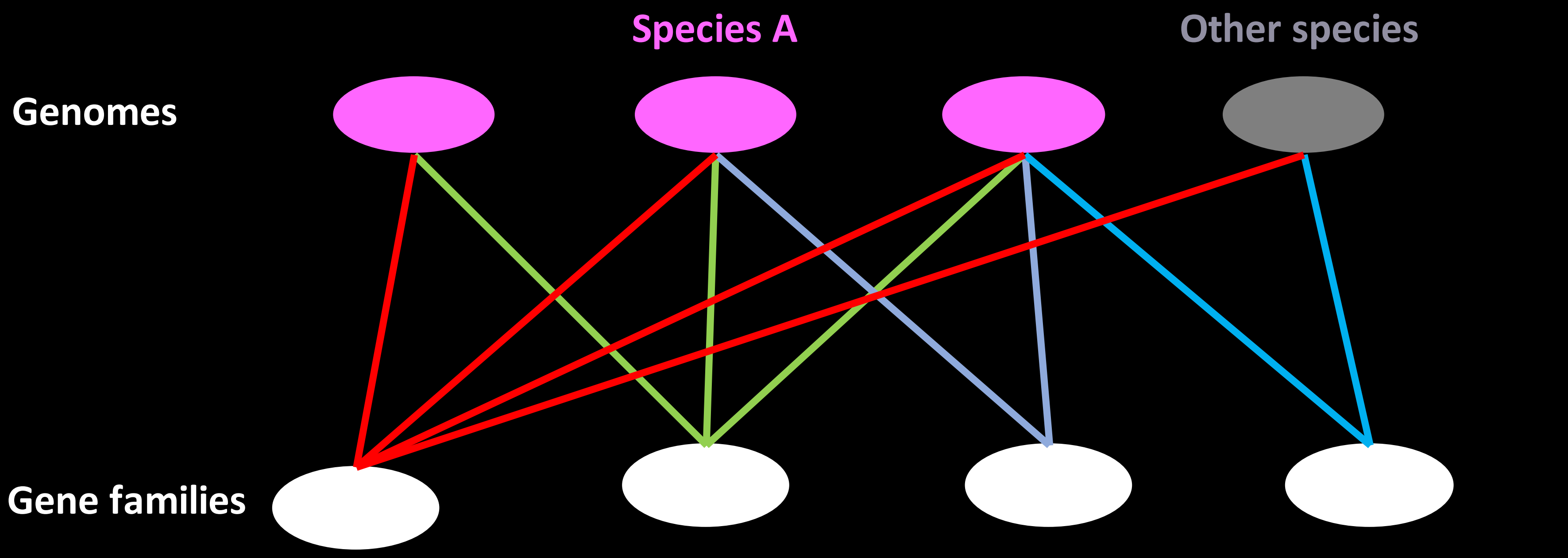


Lukjancenko O. et al. Microb Ecol. 2010

In the lab only 61 genes out of 246 065 cannot be transferred to *E. coli.*

Sorek et al., Science. 2007

Bipartite graphs are a practical way to characterize pangenomes for very large datasets



Species A

Other species

Genomes

Gene families

A core gene family

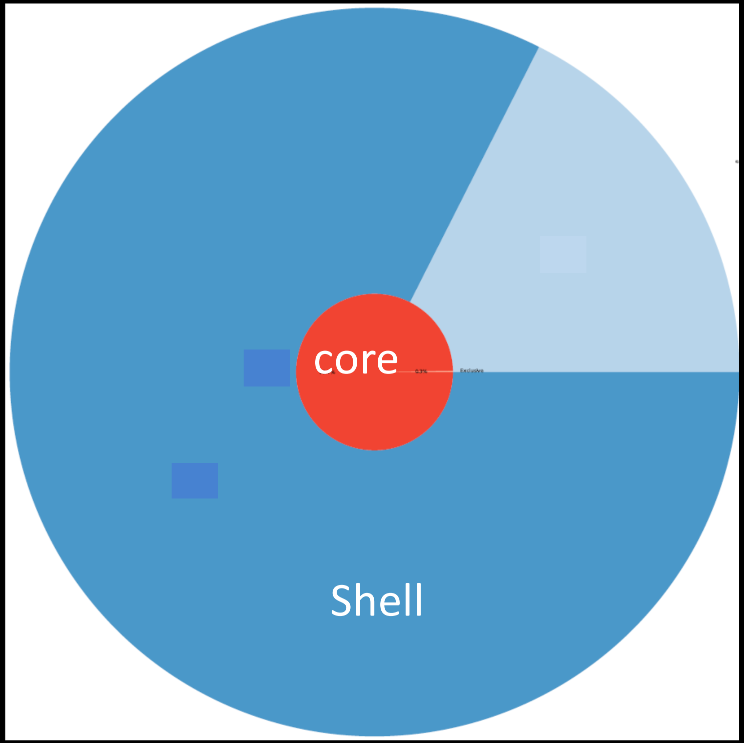
A core & exclusive gene family

Shell & exclusive gene family

Shell gene family

-Diagnostic feature
-Original innovation

- Original innovations

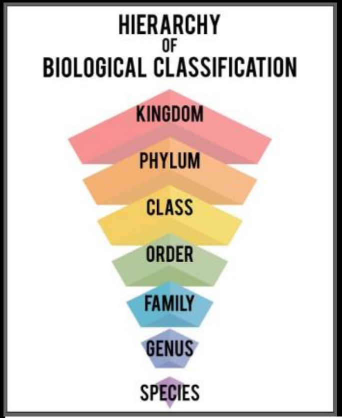


Watson *et al.* in prep

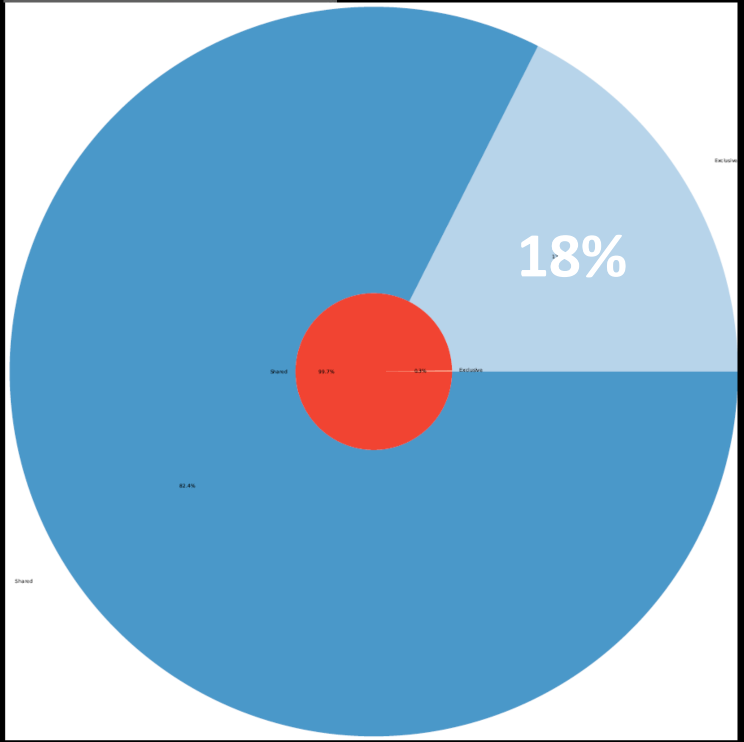
Corel et al. Trends in Micro 2016

In addition to gene transfer, these networks highlight *de novo* gene creation and the high turn-over of genetic material within lineages of genomes

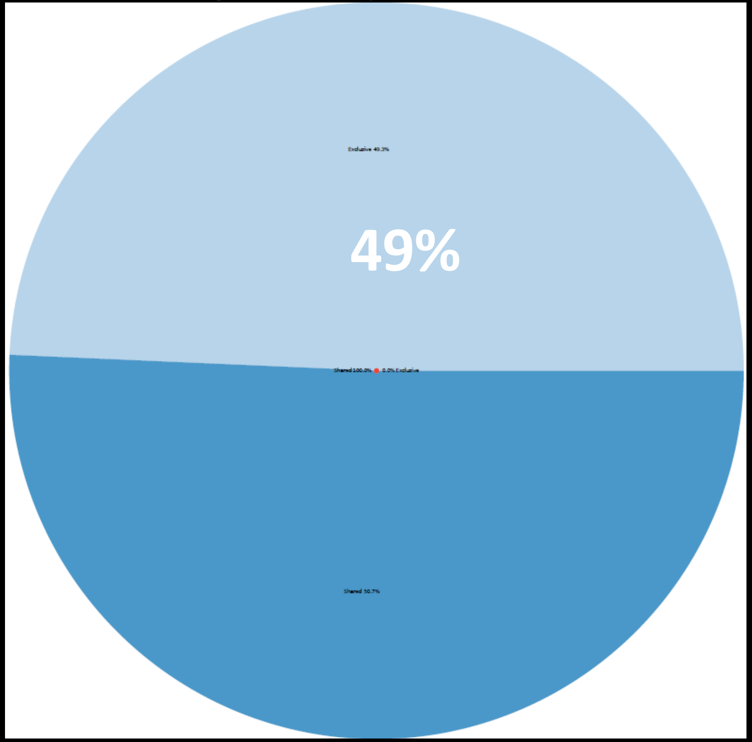
Watson *et al.* in prep



In *E. coli*

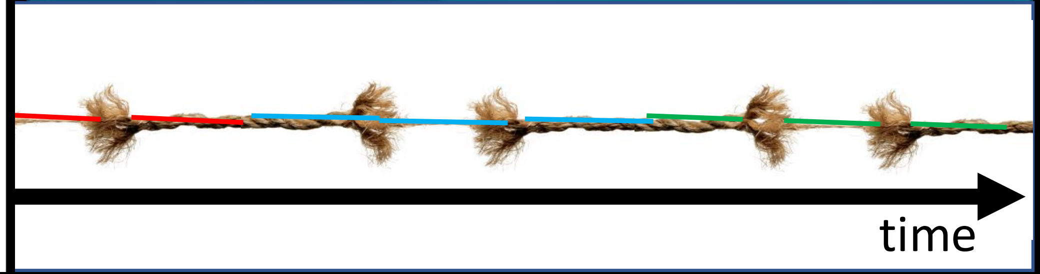


In all gamma-proteobacteria



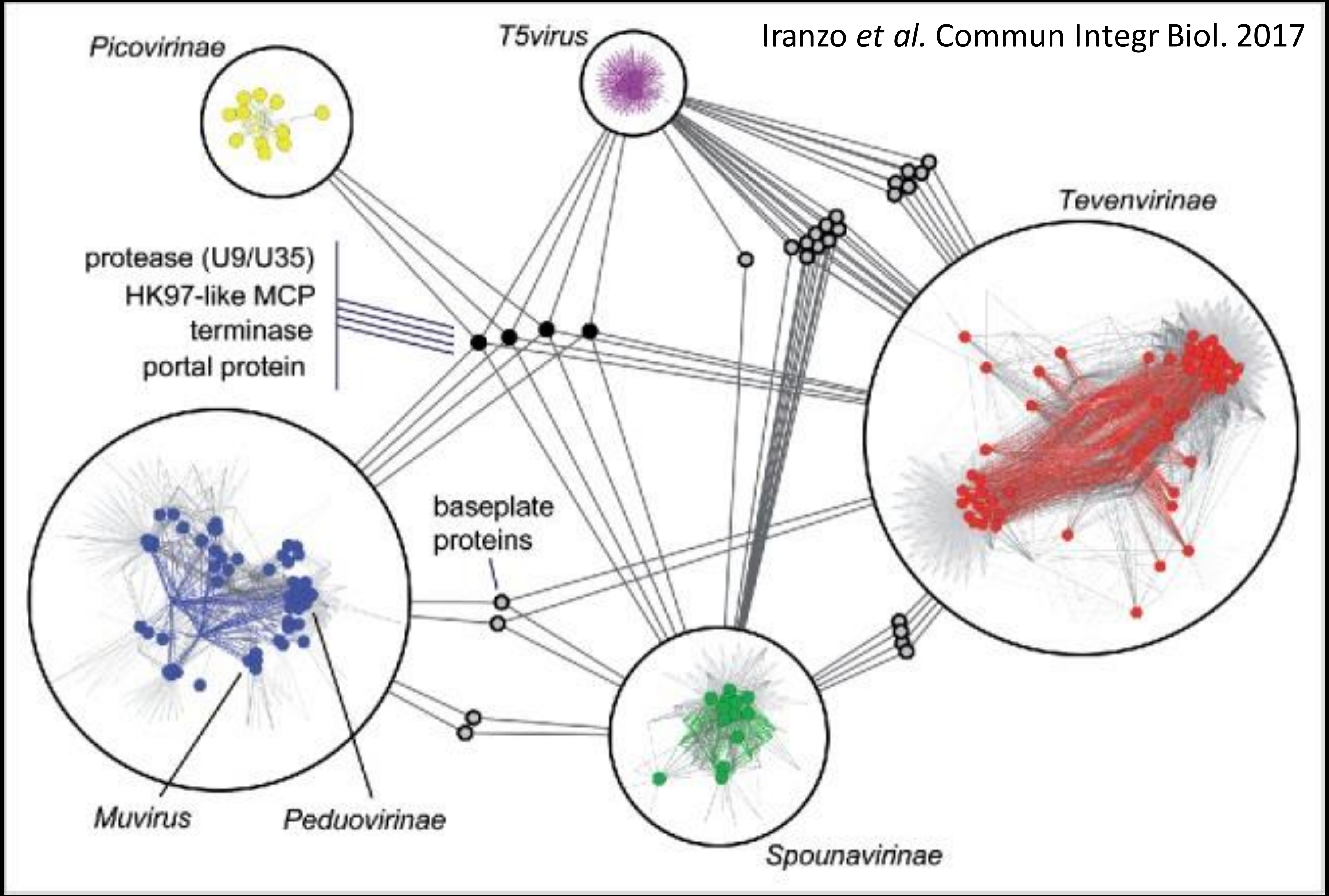
TURN OVER

TABULA RASA?



Historical continuity **with little (without?) similarity between ancestors and descendants.**

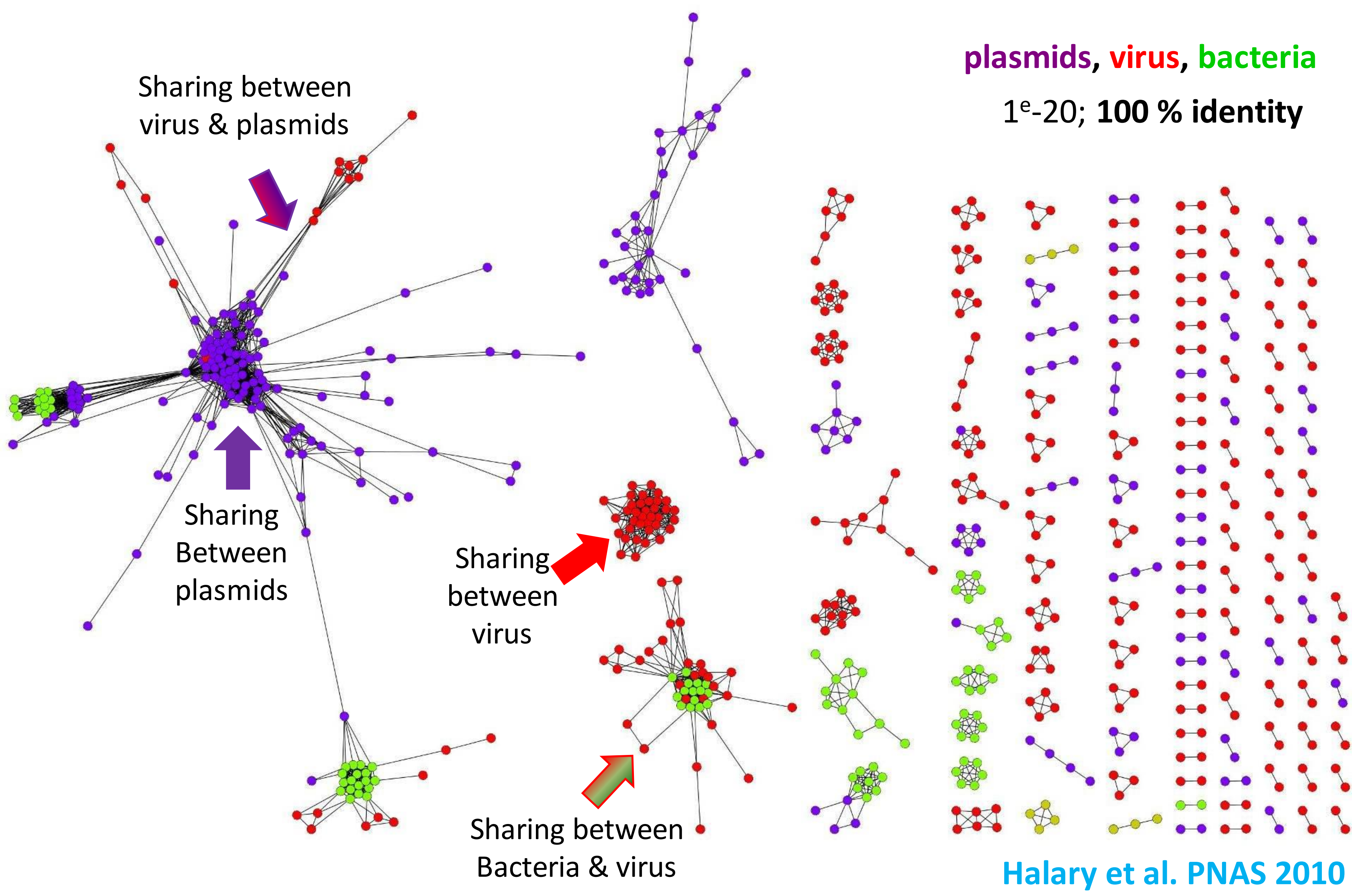
Furthermore, mobile genetic elements also exchange genes together.



Iranzo *et al.* Commun Integr Biol. 2017

These genetic interactions justify to **classify viruses and explain their evolution using networks.**

Likewise, gene transfer supports the use of networks, including unrelated entities, to study the evolution of microbial social life



plasmids, virus, bacteria

1e-20; **100 % identity**

Sharing between virus & plasmids

Sharing Between plasmids

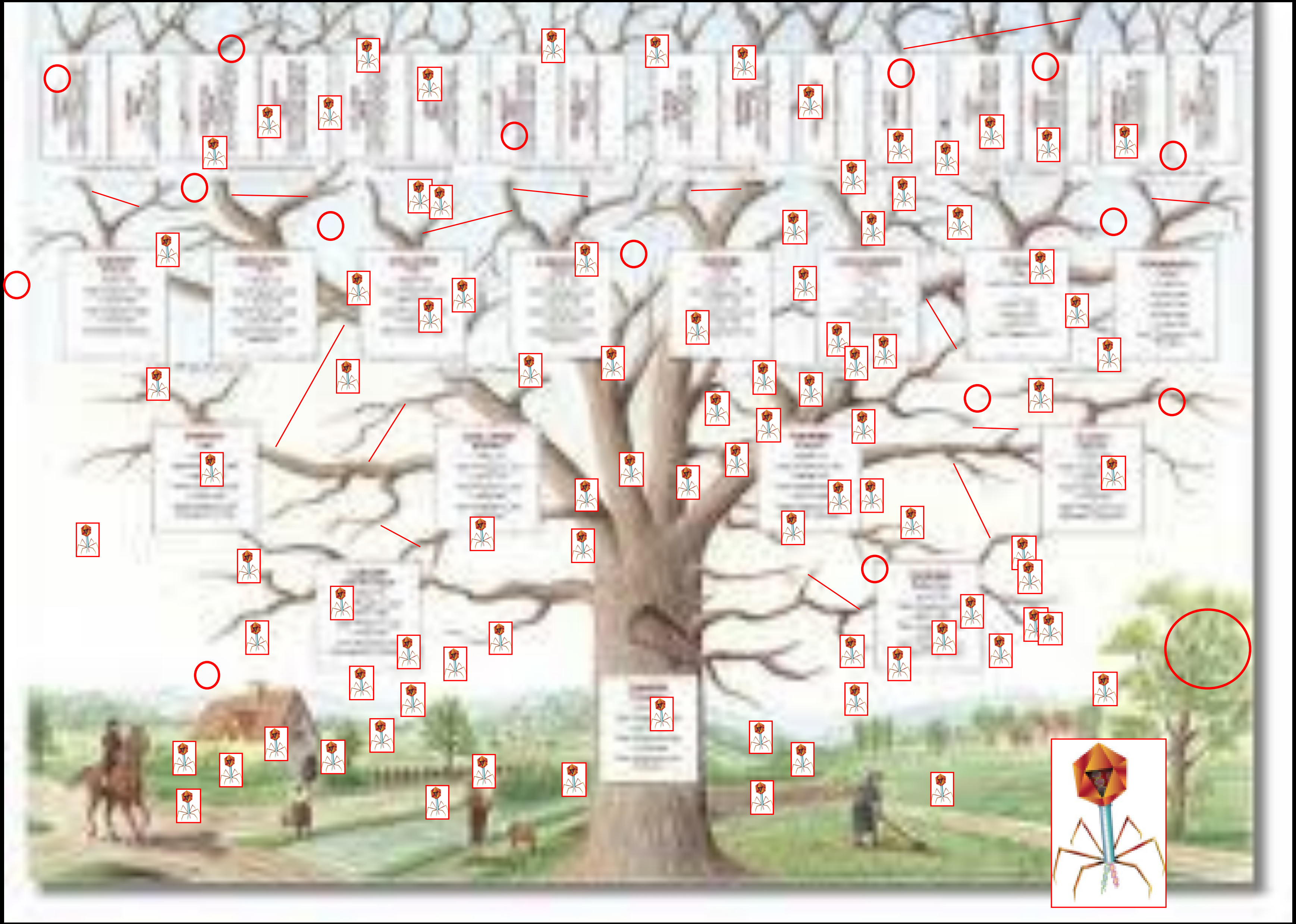
Sharing between virus

Sharing between Bacteria & virus

Halary et al. PNAS 2010

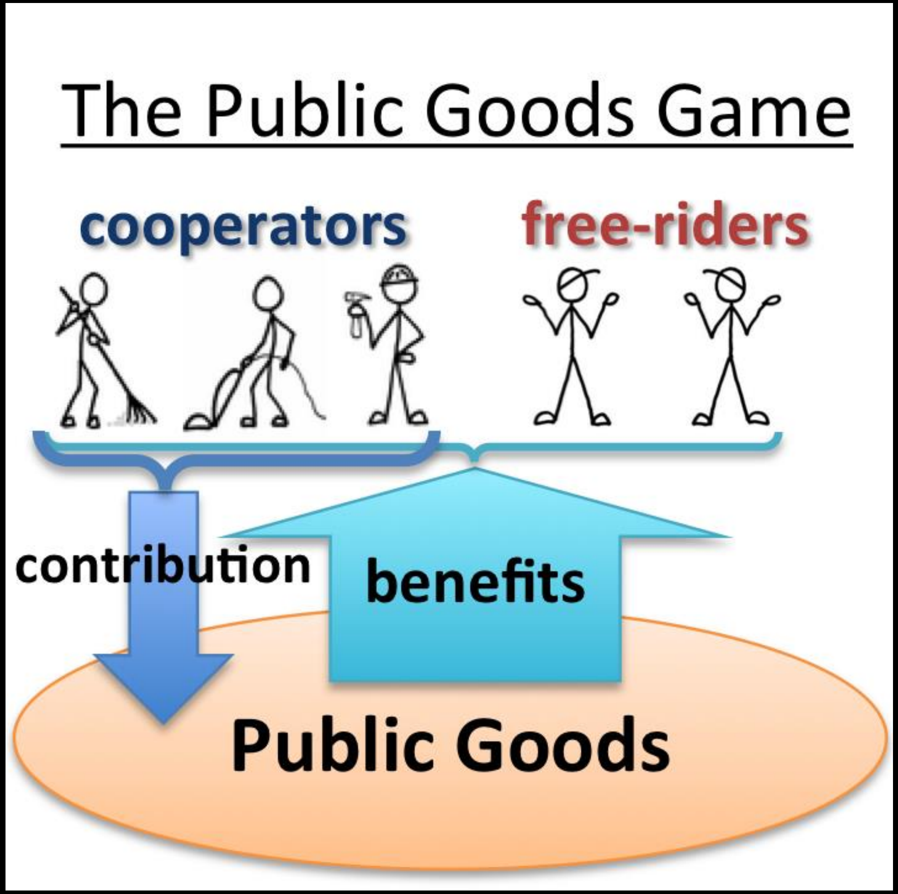
Network allows the study of evolving objects that are genealogically related

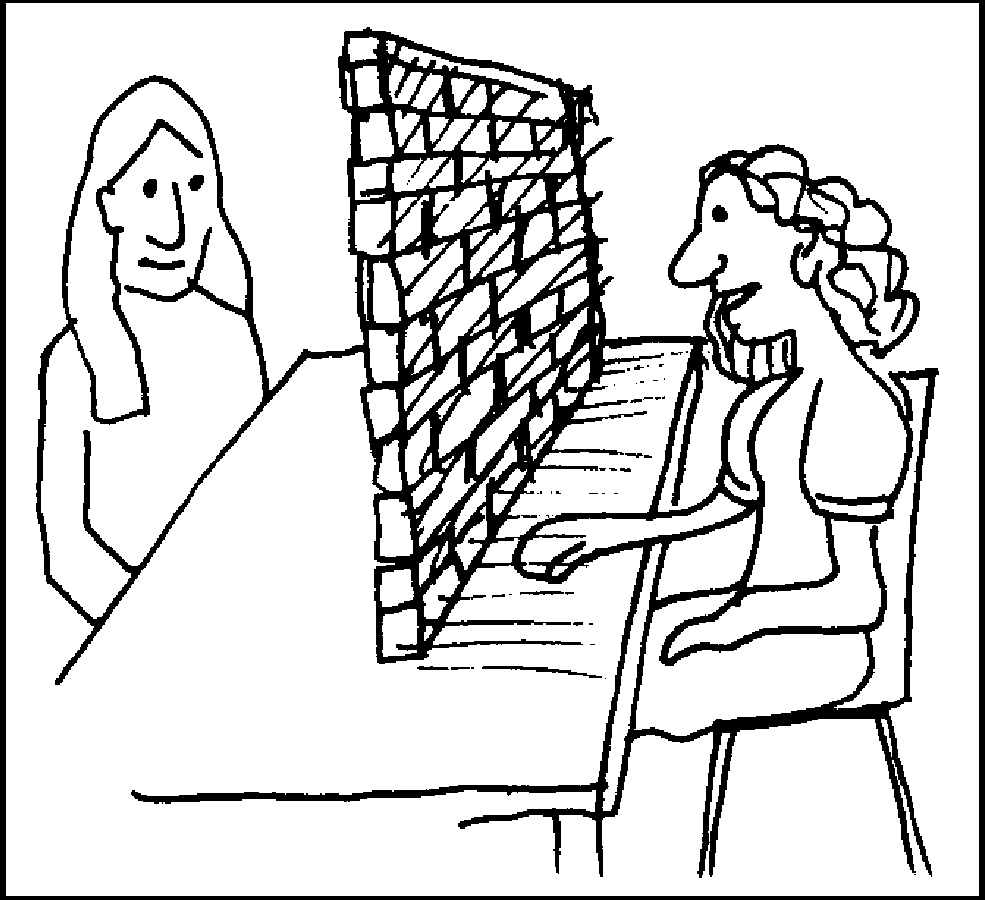
- One genealogical tree cannot represent all the evolutionary history



Thus, network approaches bring forward evolutionary questions that may go beyond relatedness

- **What partnerships** ?
  (clubs of genomes enjoying public genetic goods)
- **What barriers** to transmission ?

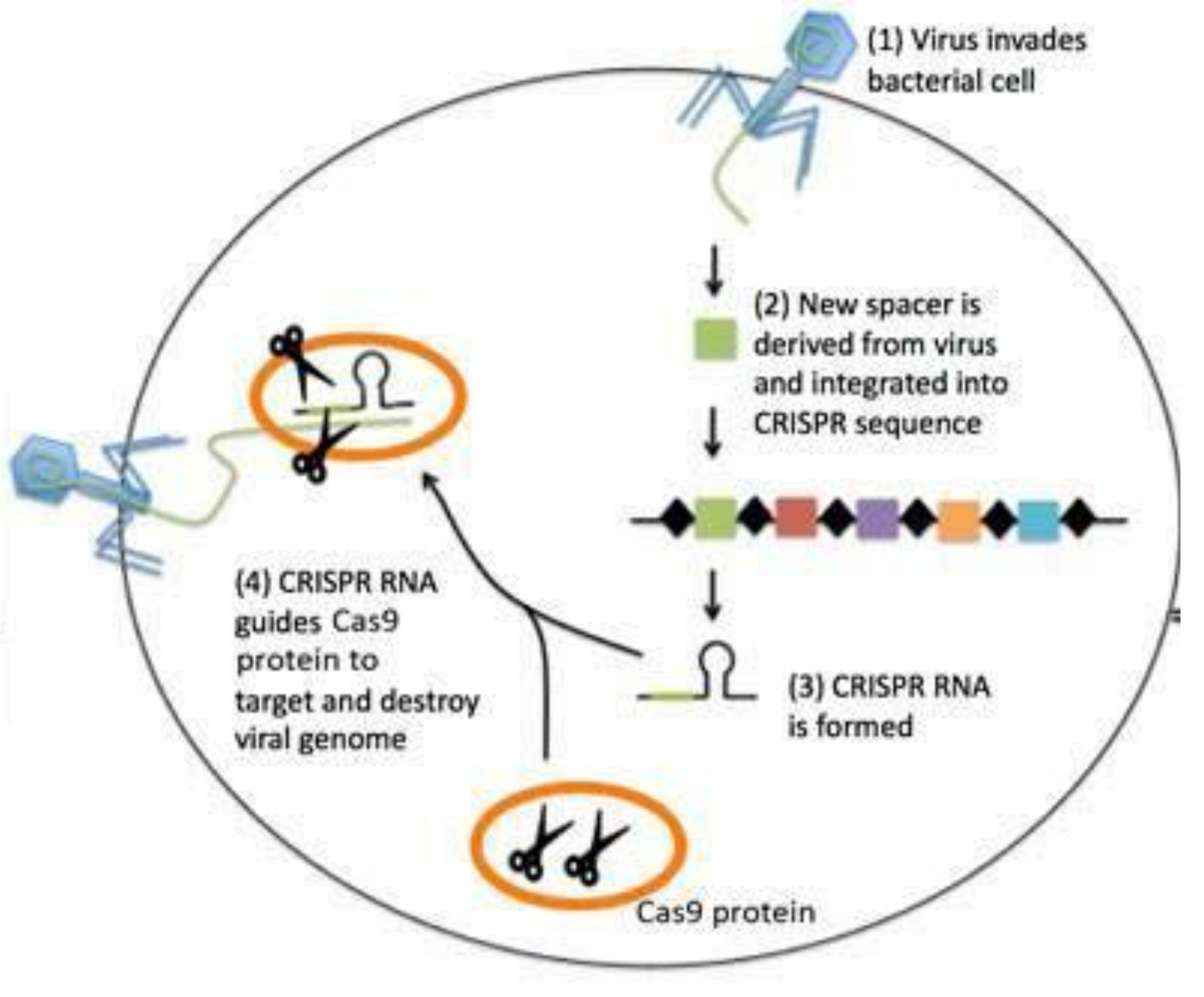




McInerney *et al.* Biology Direct 2011, Bapteste *et al.*, PNAS 2012.

For example, it predicts and explains the evolution of various types of barriers to gene transfer

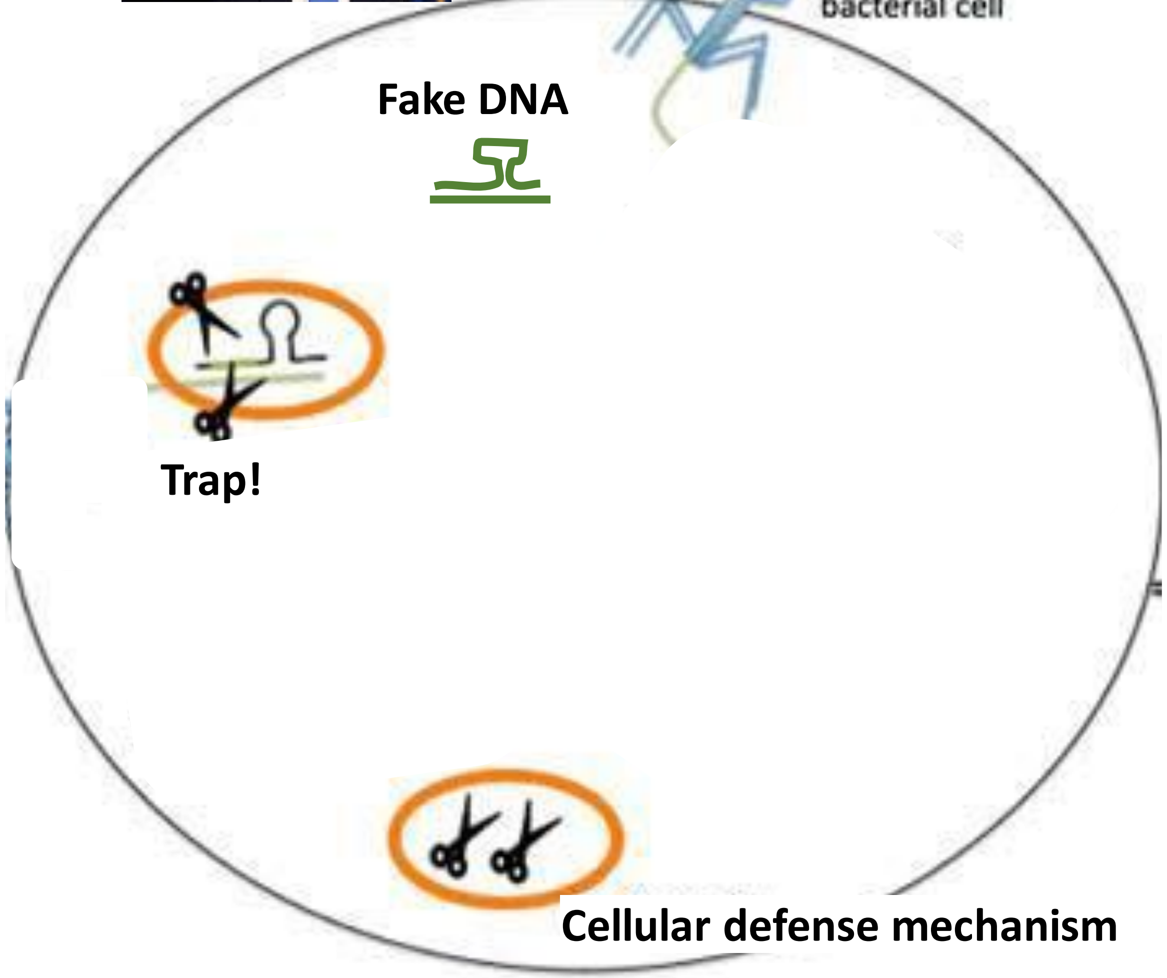
- CRISPR-Cas



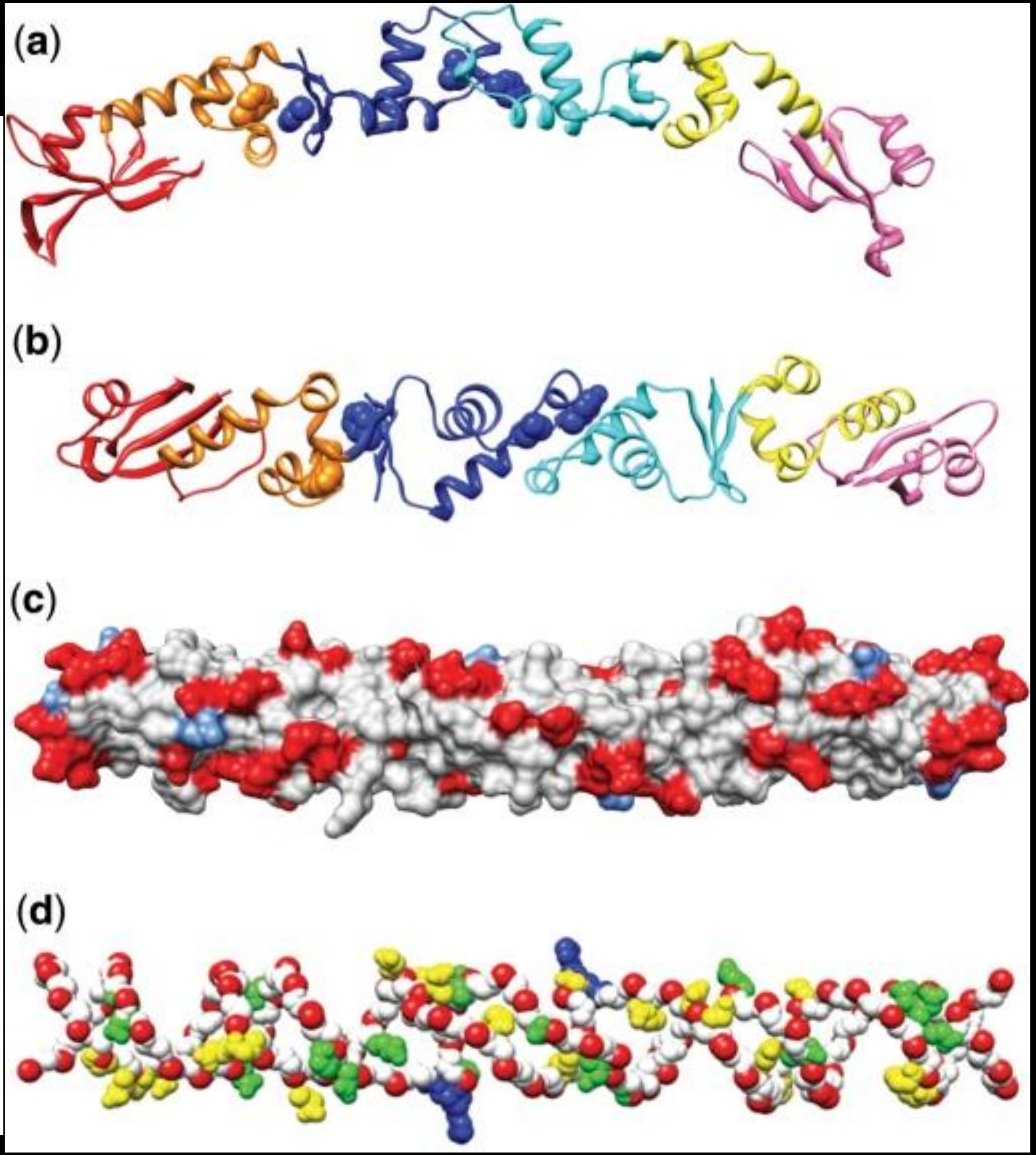
http://www.i-sis.org.uk/CRISPR_too_fast_for_comfort.php

For example, it predicts and explains the evolution of various types of barriers to gene transfer

- Fake DNA

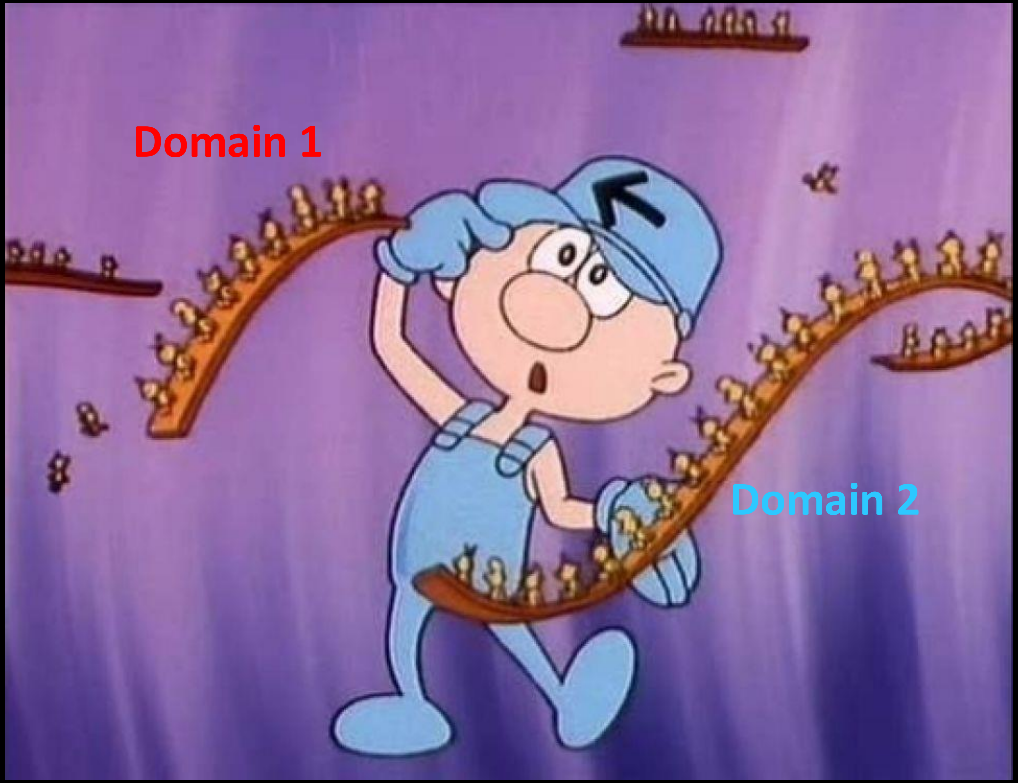


http://www.i-sis.org.uk/CRISPR_too_fast_for_comfort.php

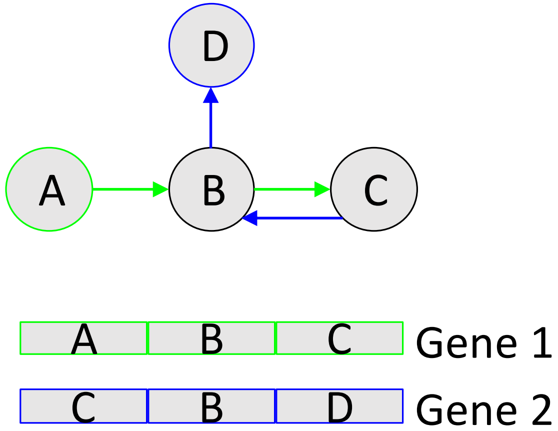


MacMahon et al. NAR 2009; Walkinshaw et al. Mol Cell. 2002

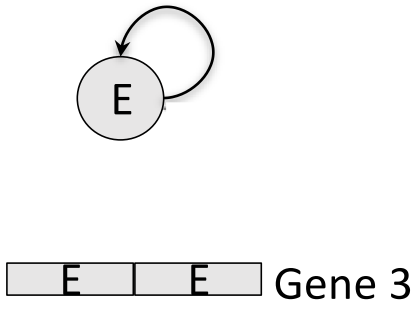
Domain combinations is another source of genetic variation, analyzable with networks



Basic structure of the network with nodes (domains) connected by an edge when they are found in the same protein.

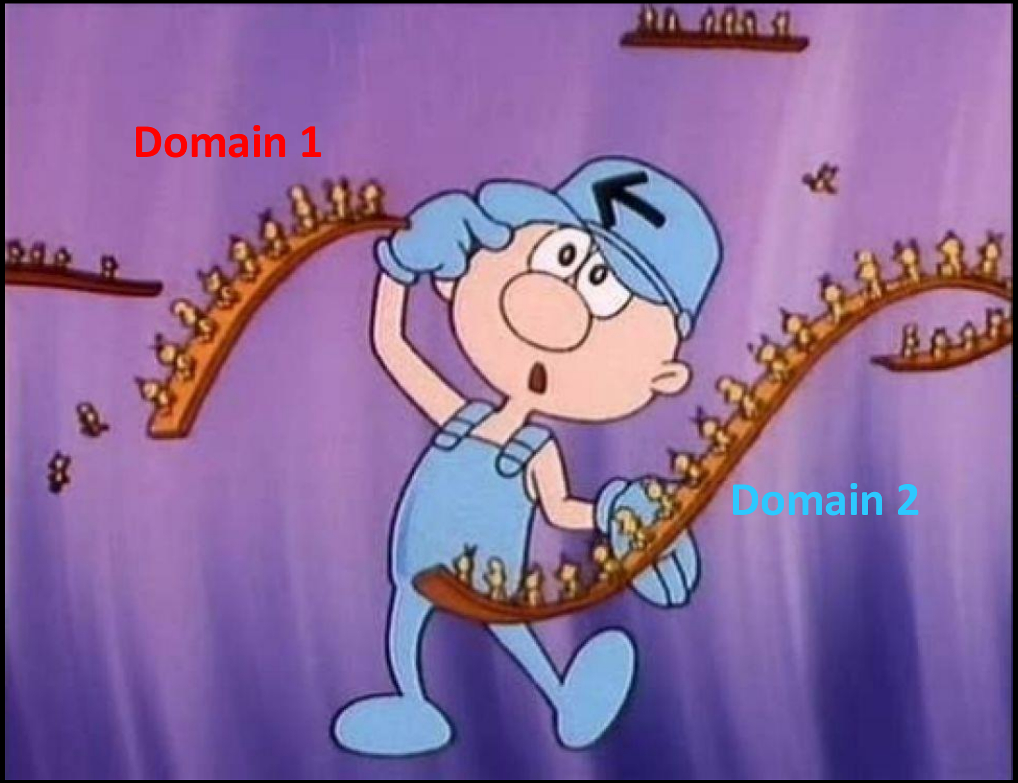


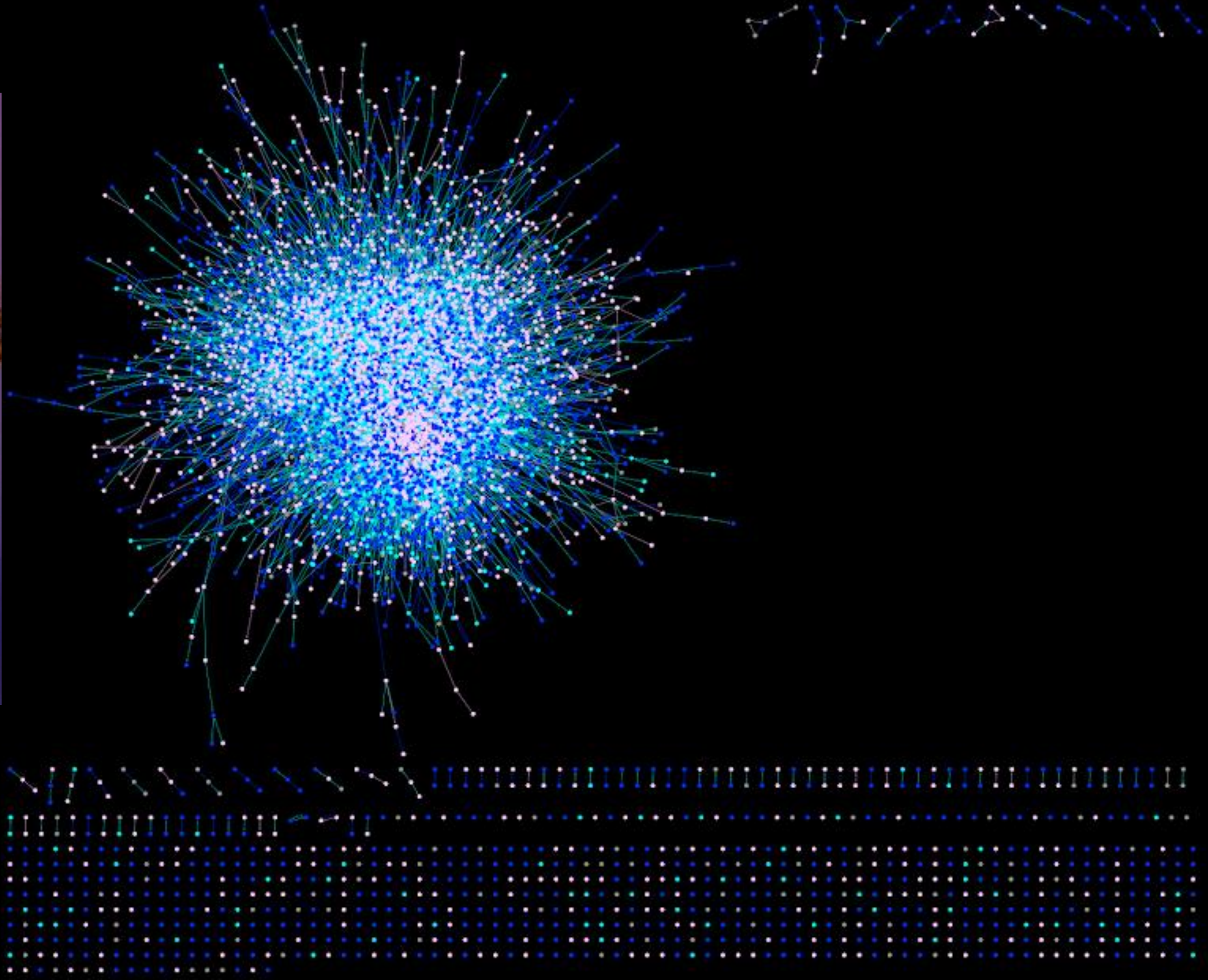
Basic directed structure of the network



Tandem repeats represented as self-loops

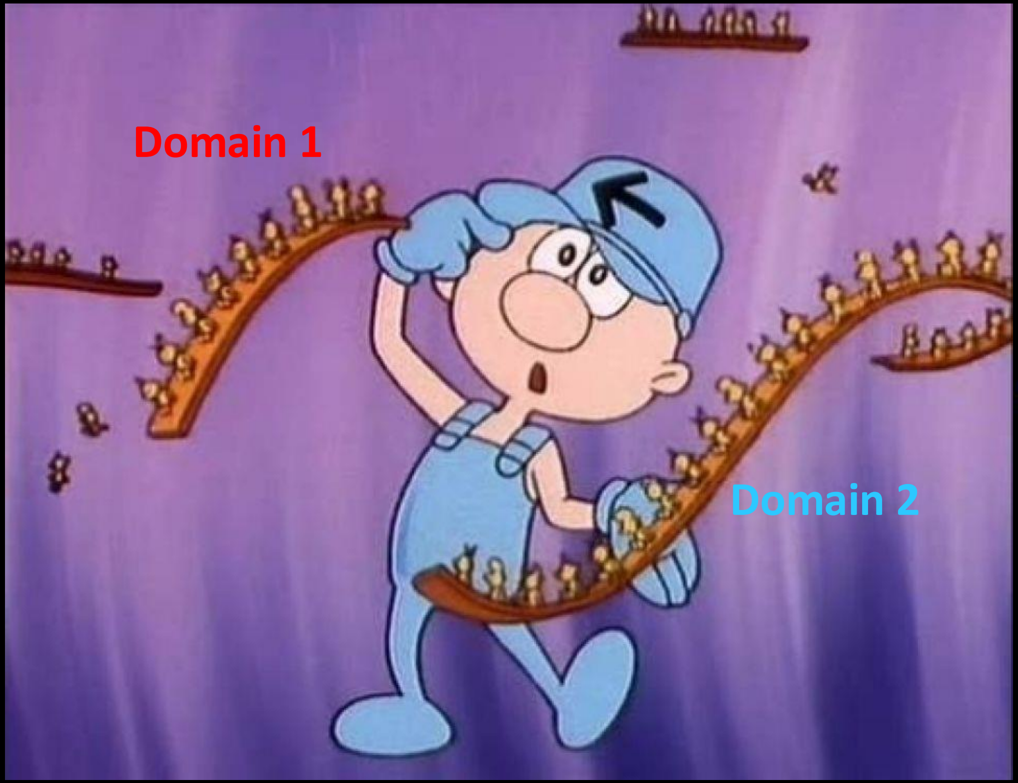
Bacterial genes have been significantly remodeled : >68 % of domains combinations (turquoise edges) did not exist in the first bacteria





Watson *et al.*, in prep

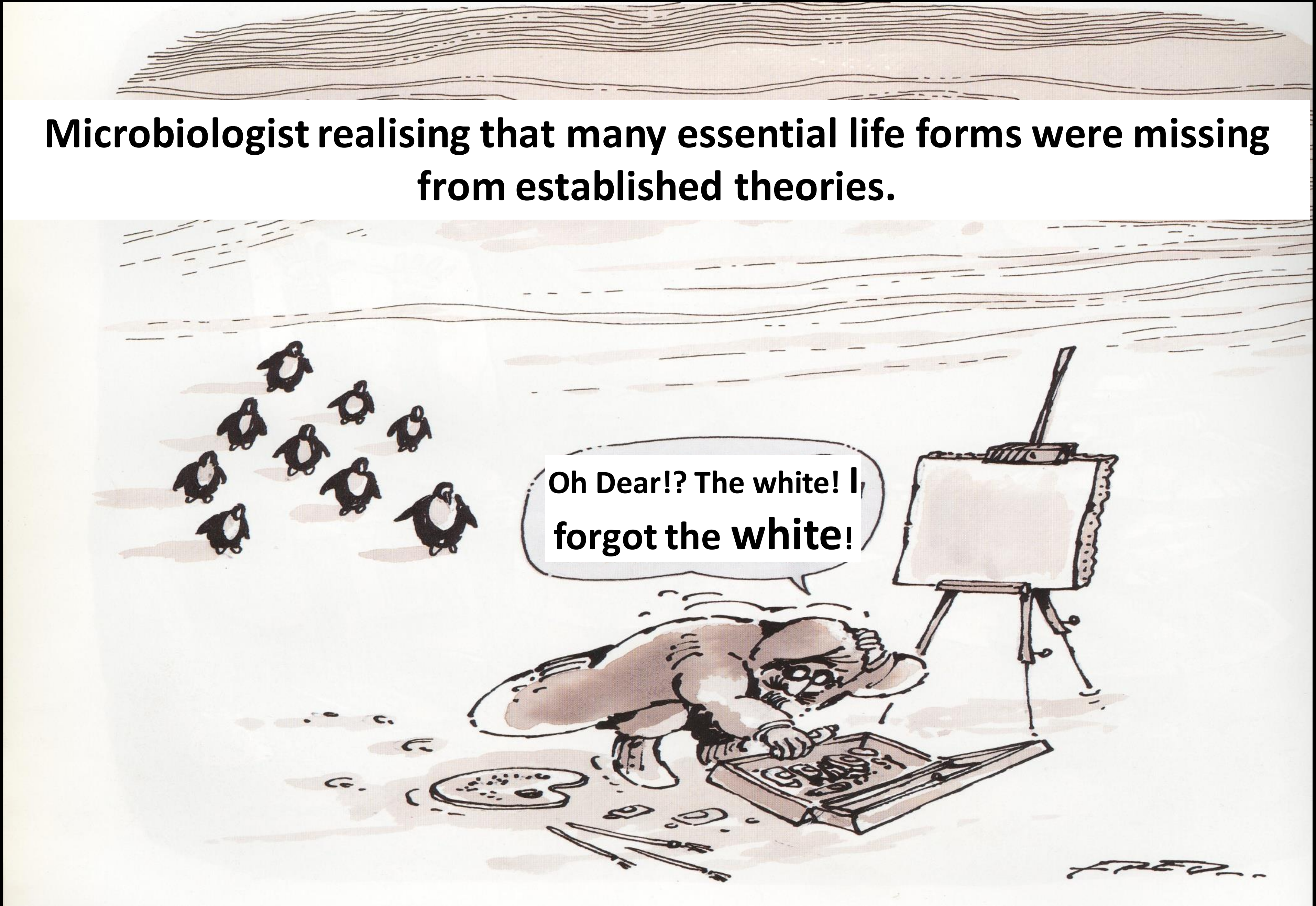
Bacterial genes have been significantly remodeled : >68 % of domains combinations (turquoise edges) did not exist in the first bacteria



Gene lineages, like organismal lineages, appear mosaic.

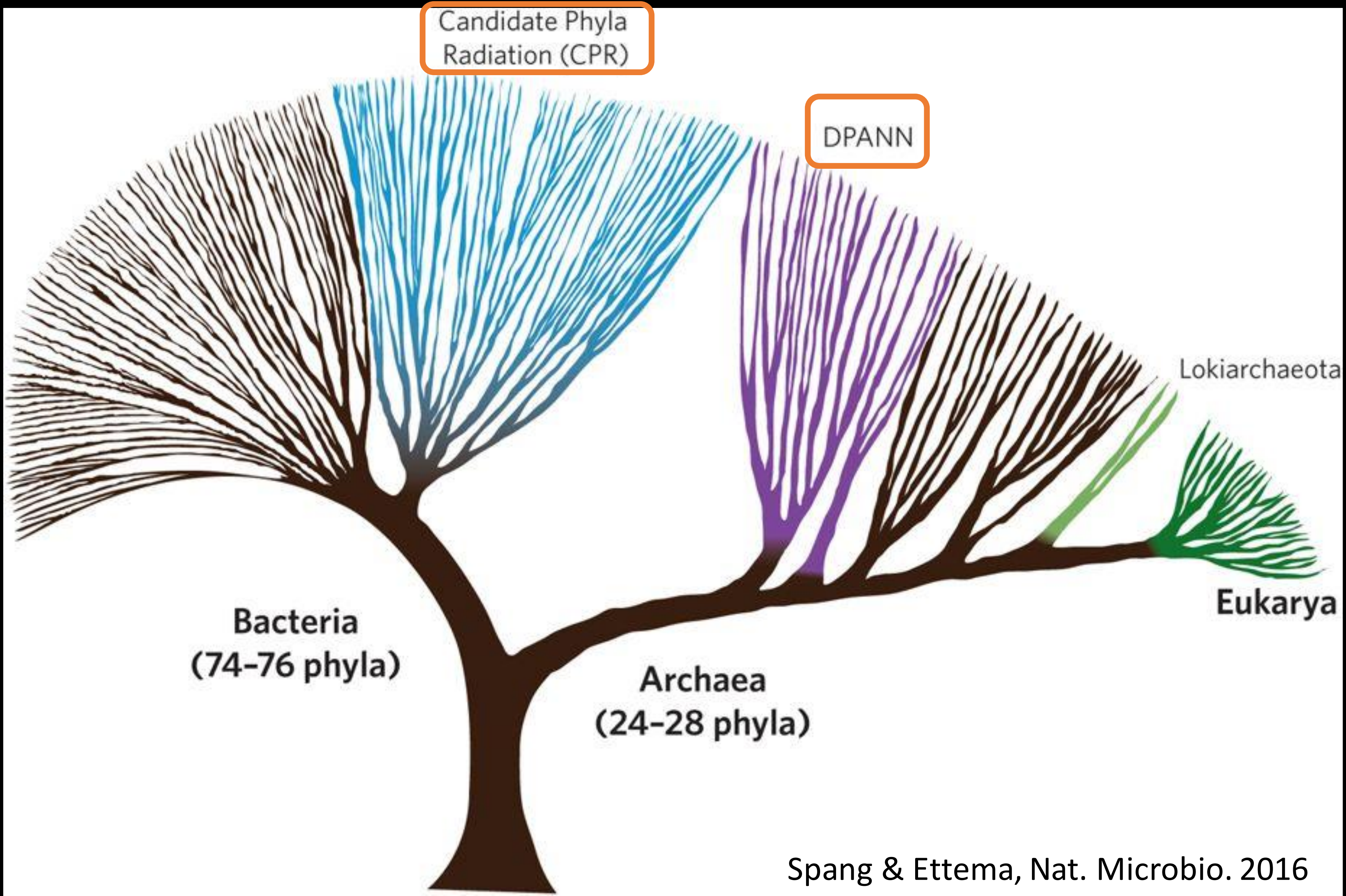
Watson *et al.*, in prep

- On the origins of traits: ultra-small cells with ultra-small genomes



Microbiologist realising that many essential life forms were missing from established theories.

Ultra-small cells with ultra-small genomes constitute a large part of biodiversity



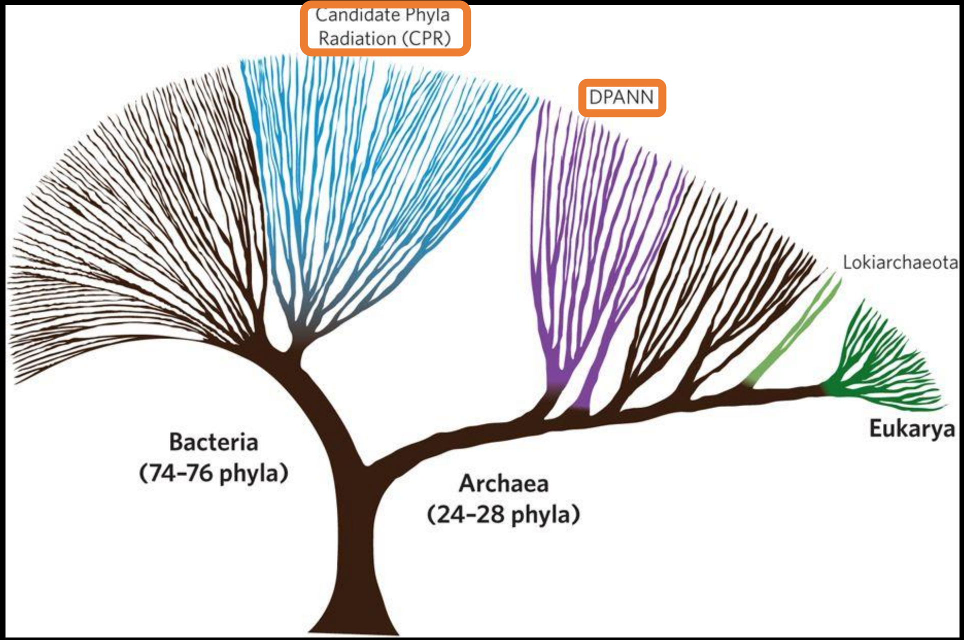
Spang & Ettema, Nat. Microbio. 2016

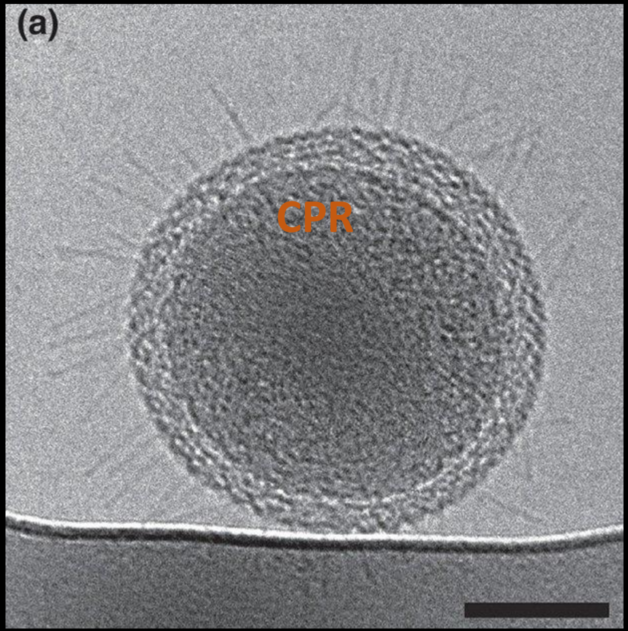
And largely they have been discovered only in the past 3 years!

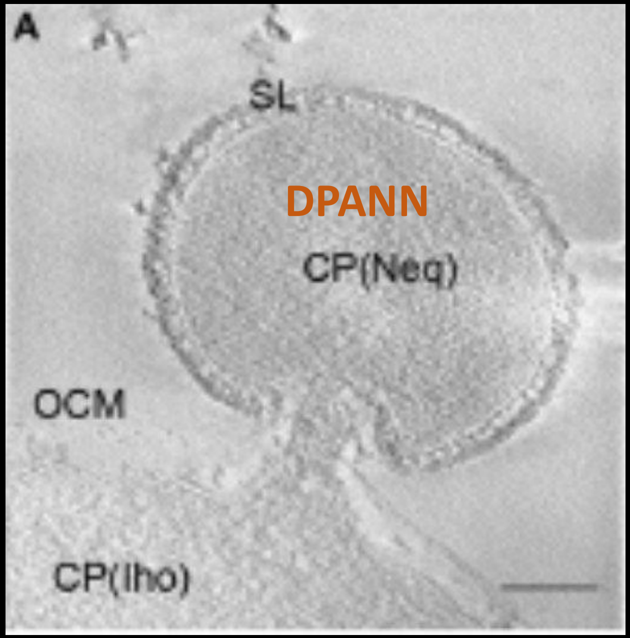
These ultra-small cells however cannot live on their own.

- Lack of biosynthesis pathways for nucleotides, lipids, and most amino acids.
  - Suggest a dependence on other community members

Castelle et al. Cell 2018



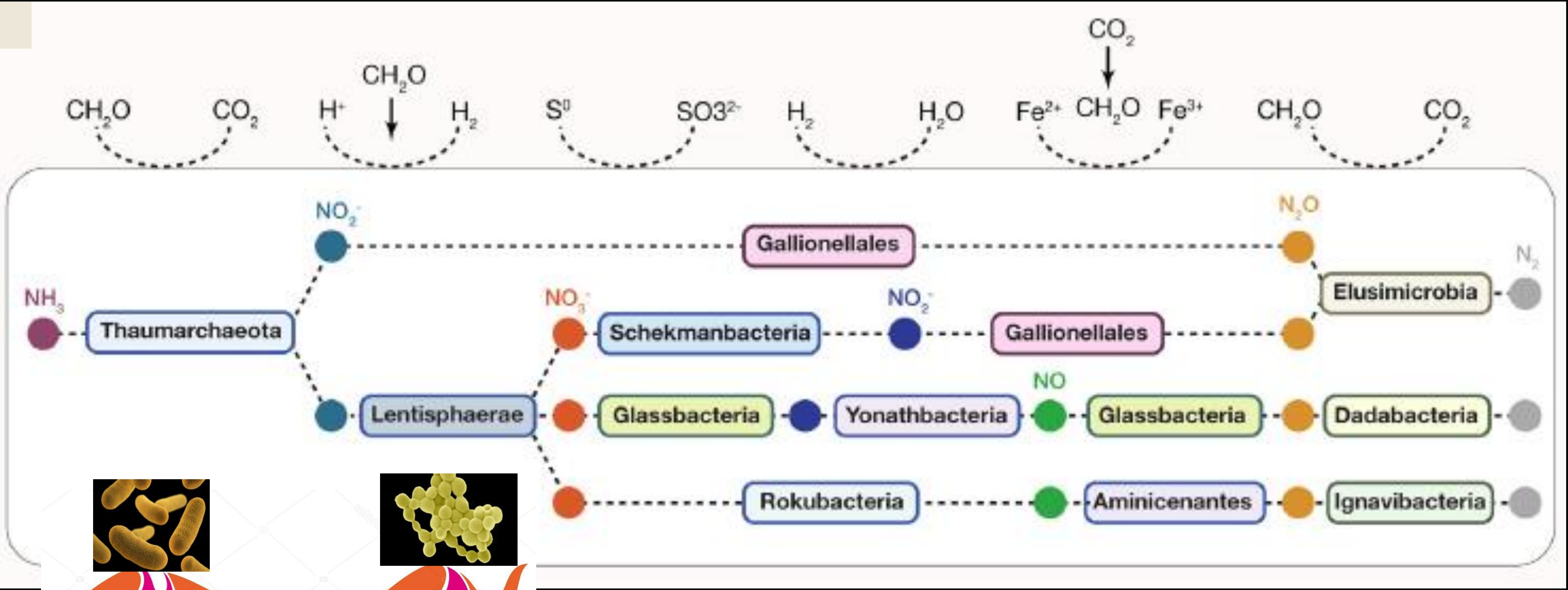


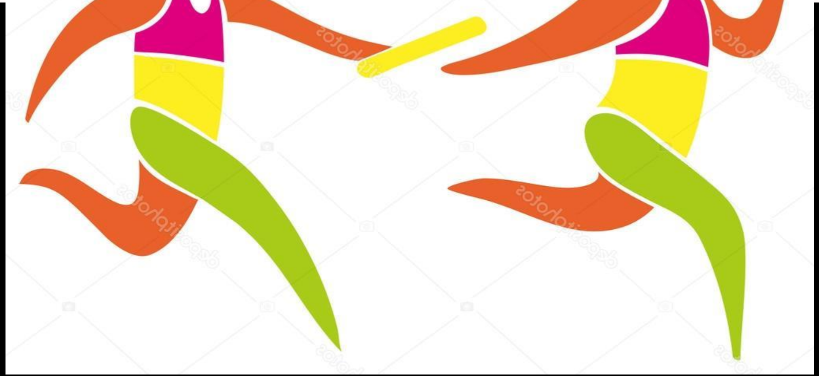


'Natural selection cannot possibly produce any modification in any one species exclusively for the good of another species.'

Darwin

They take part to communical geochemical pathways via metabolic handoffs

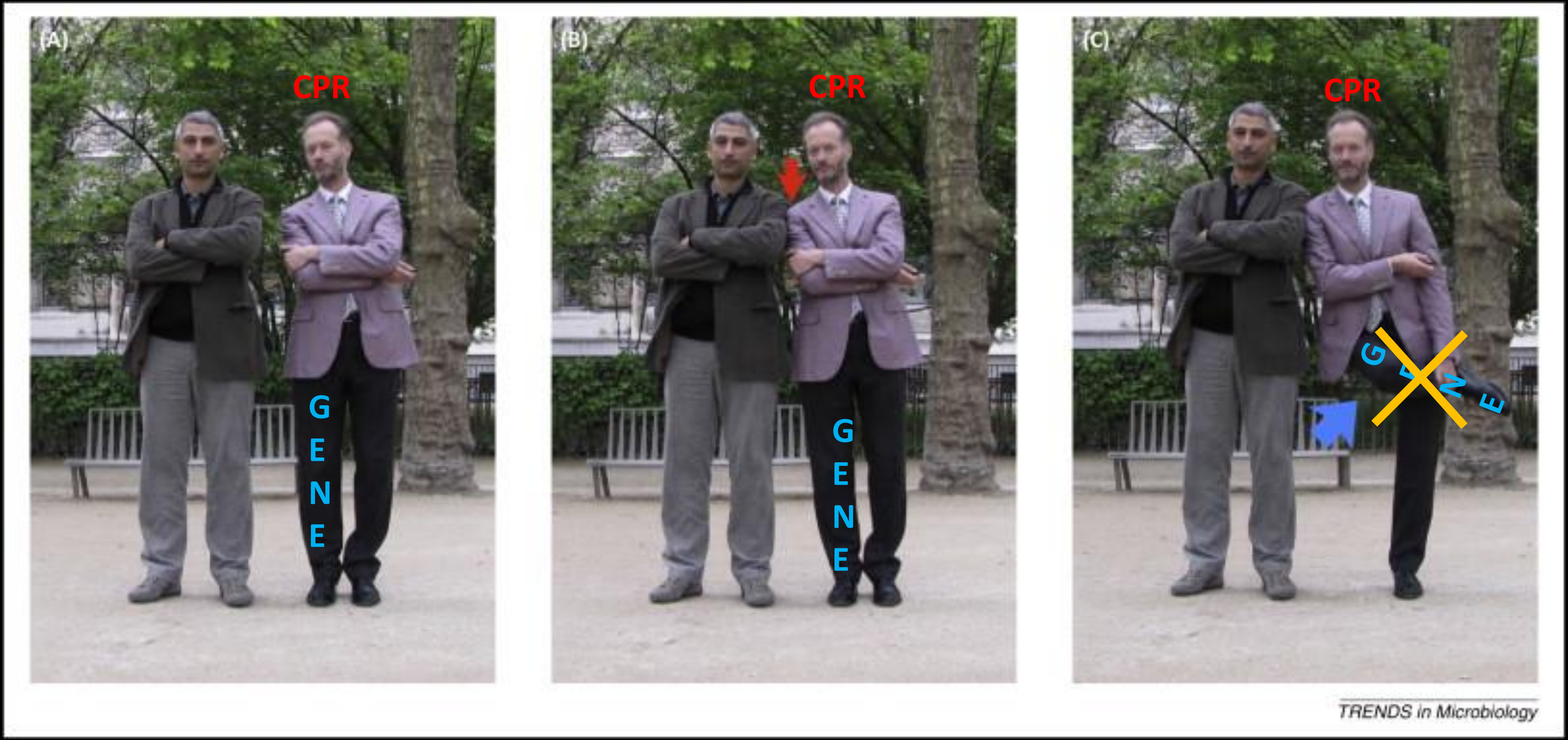




Castelle et al. Cell 2018

Network thinking explains the evolution of these ultrasmall bugs, too small to live on their own.

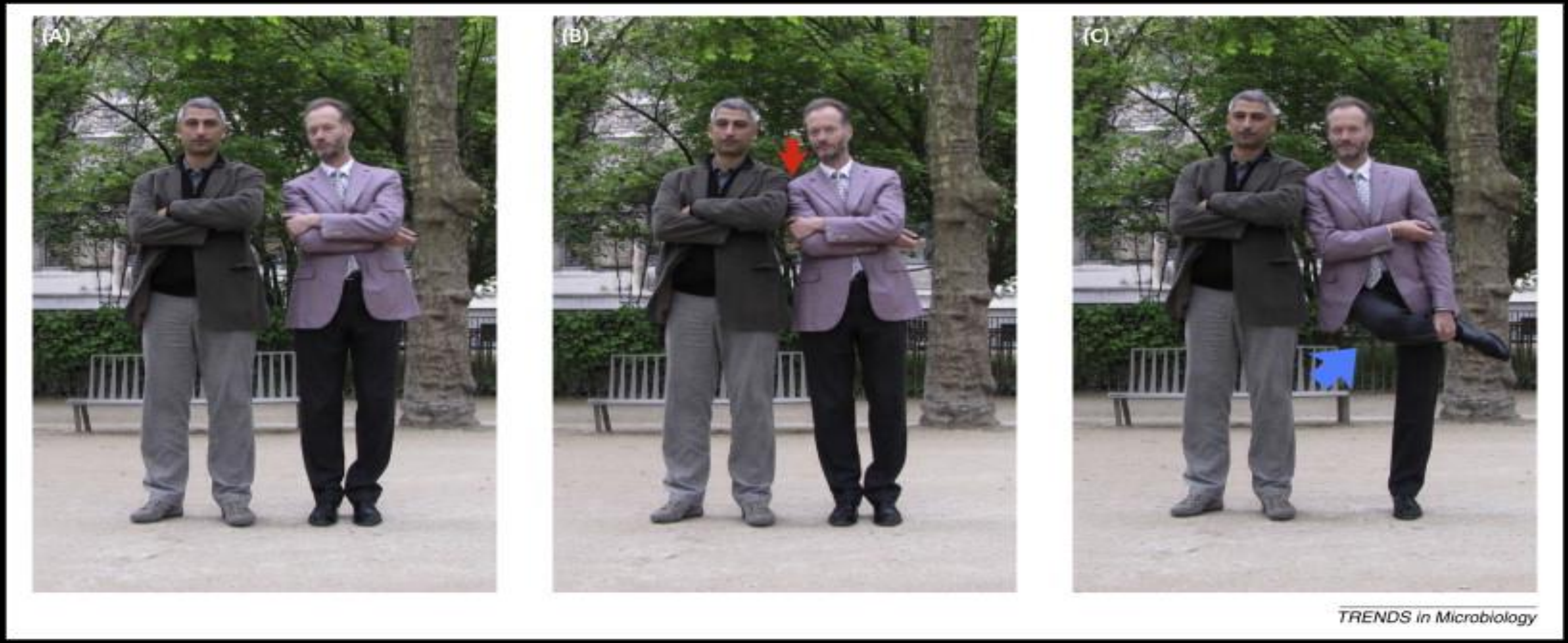
Ultrasmall cells have lost some of their genes due to their interactions with other bacteria



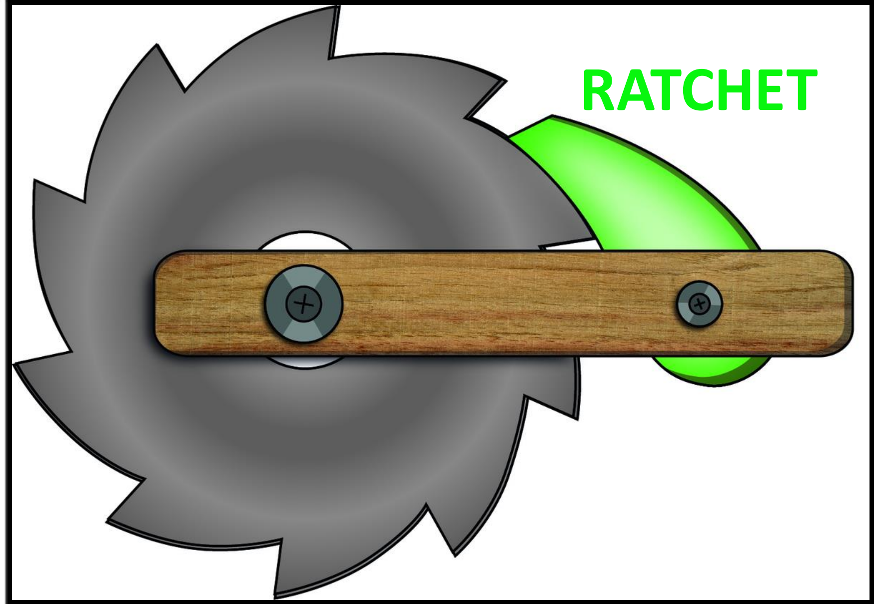
(Sélosse et al. Trends in Micro., 2014)

A process of **constructive neutral evolution**.

Such dependences are difficult to reverse- hence it can be predicted that complex microbial community, with non autonomous cells, evolve over time.



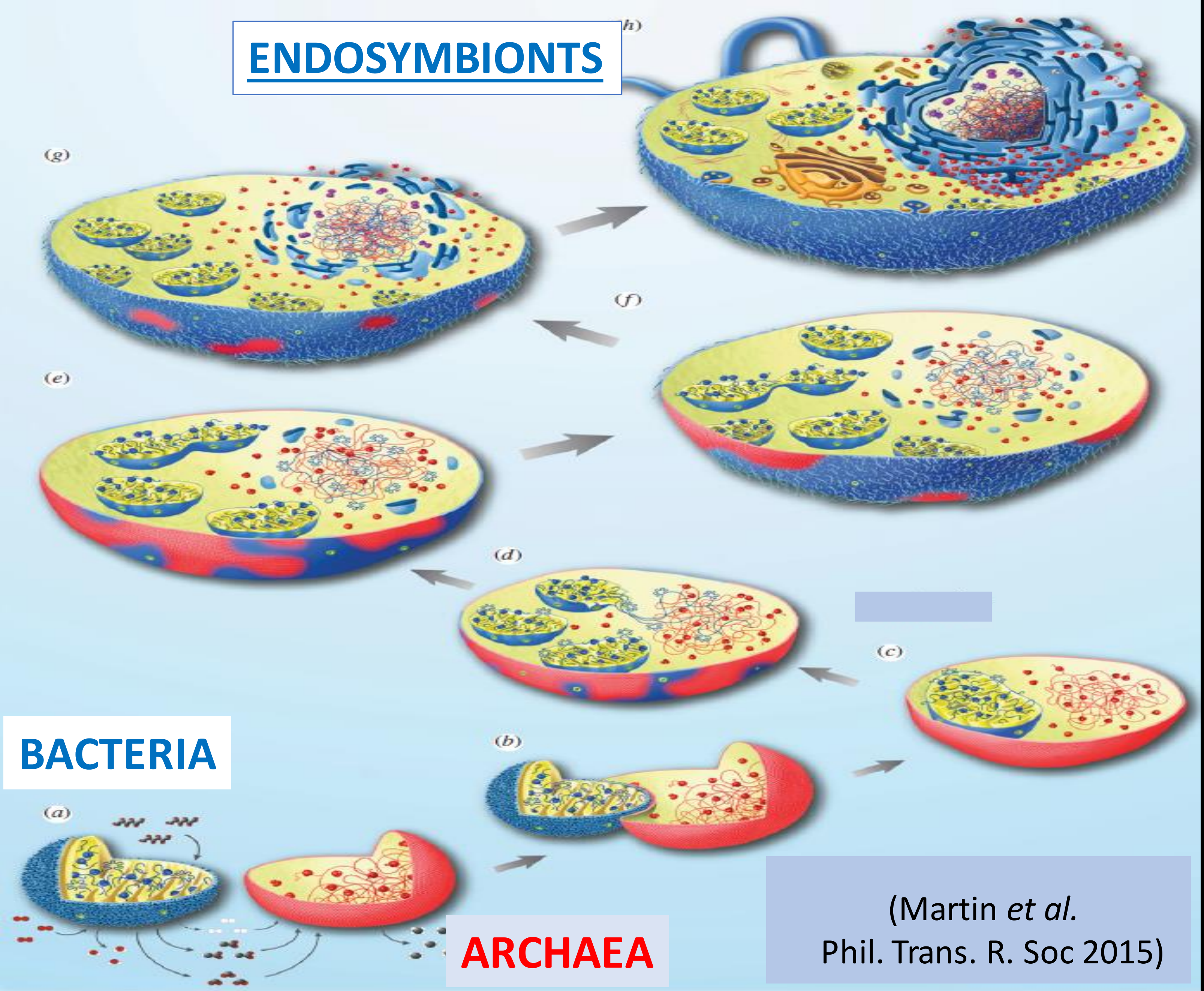
PRESUPPRESSION



(Gray & Doolittle, Science, 2010)

This type of explanation contrasts with more classic views.

On the origins of cells: endosymbioses, cellular dependences that evolved much further

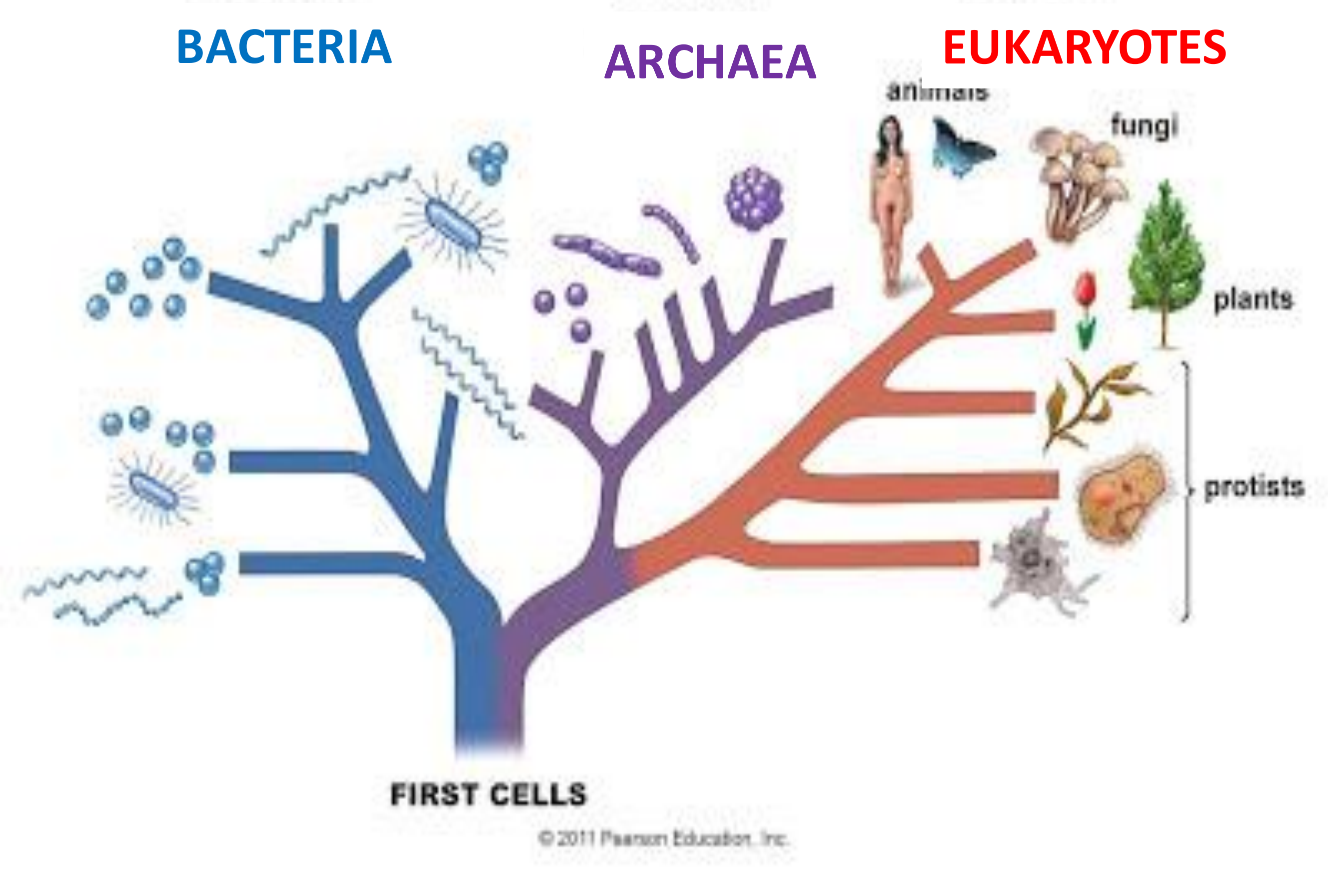


(Martin *et al.*
Phil. Trans. R. Soc 2015)

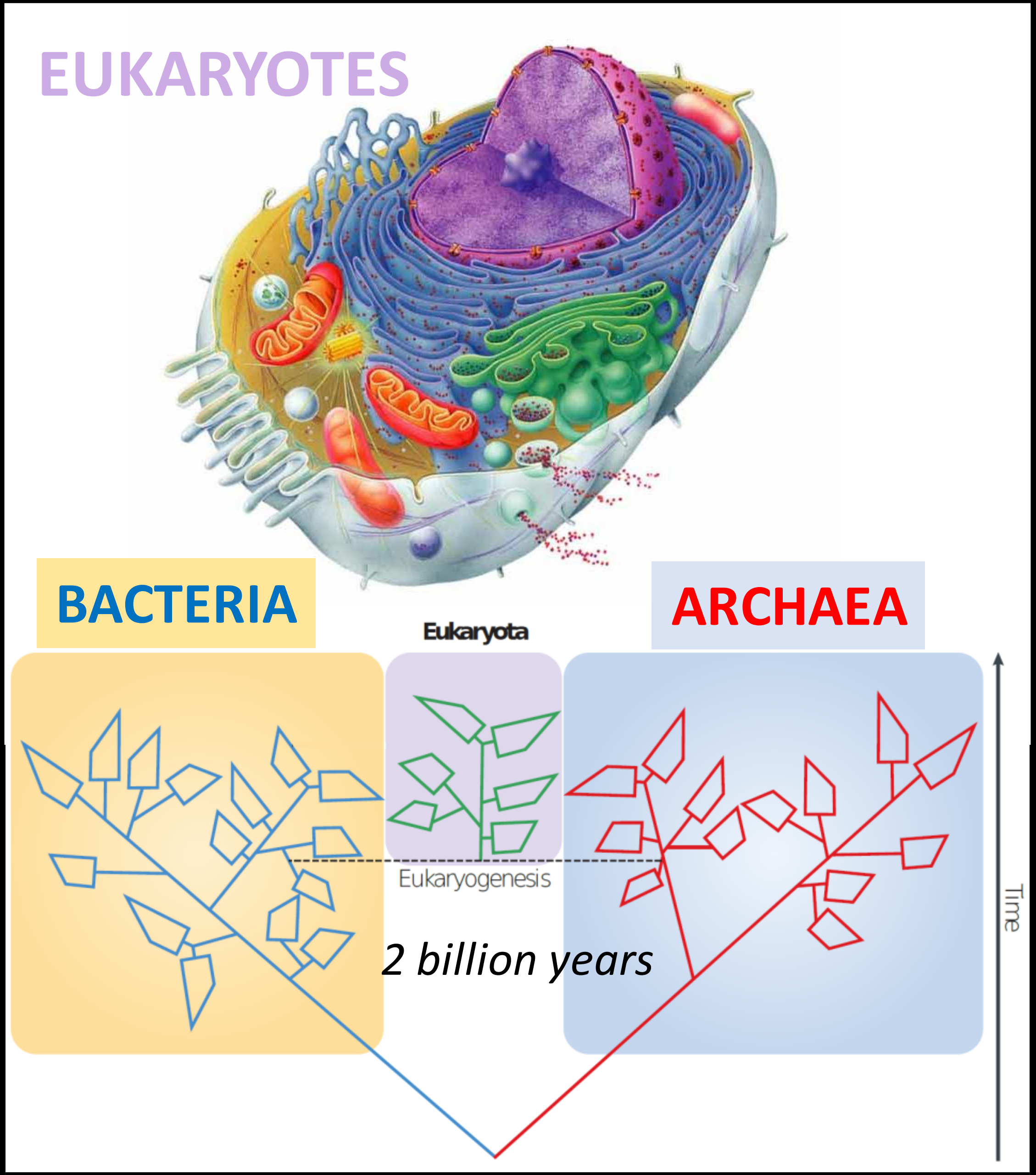
HOST:

EUKARYOTES

Network thinking supports a scenario at odds with a tree-like view

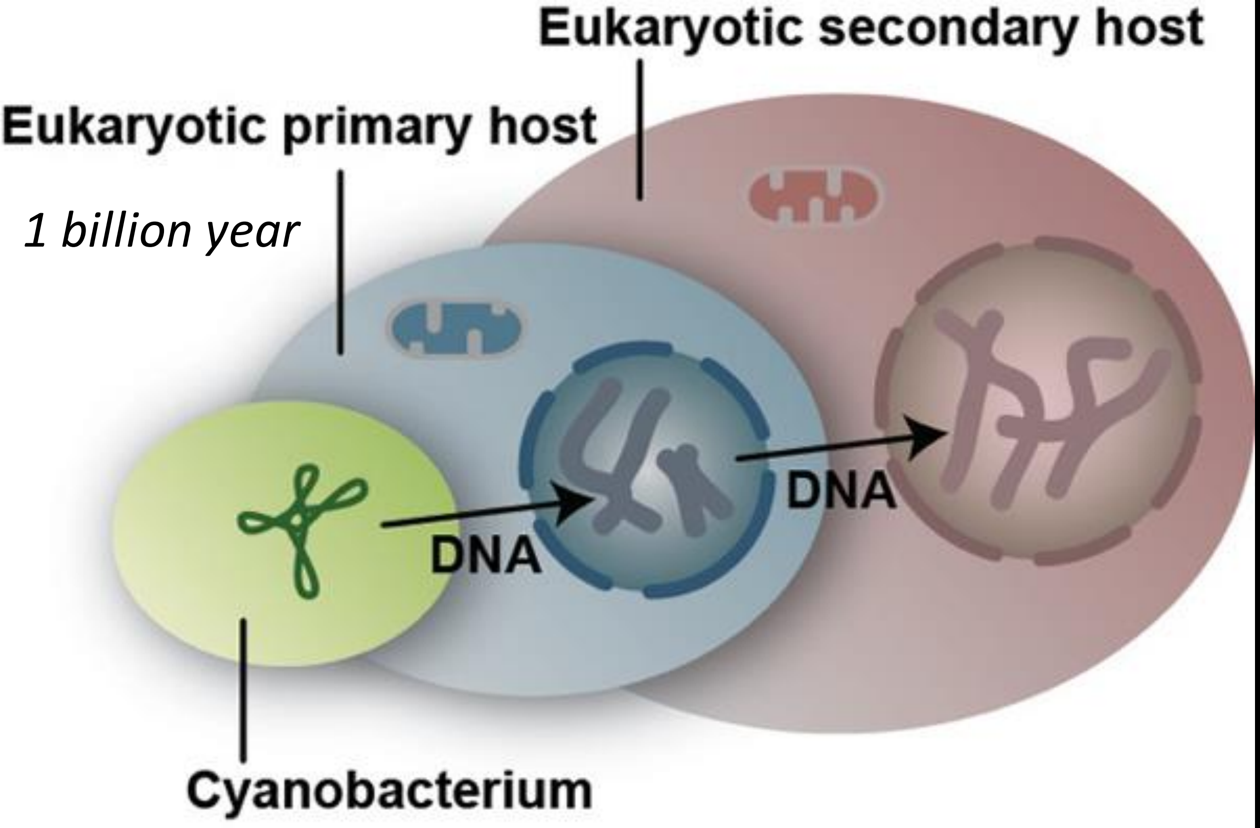


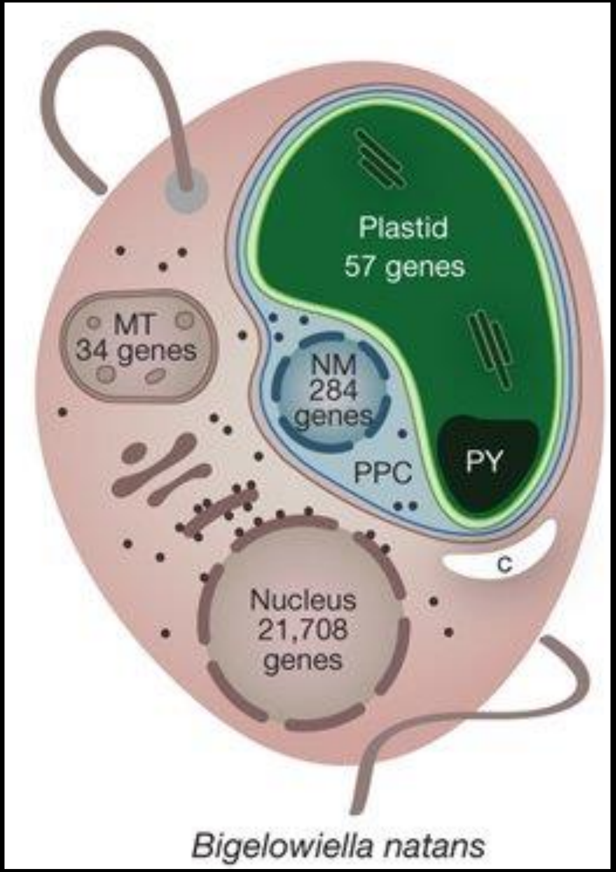
© 2011 Pearson Education, Inc.



Additional primary, secondary & tertiary endosymbioses are well documented in eukaryotes

➡ Algae, plants (Photosynthetic eukaryotes)





Endosymbiosis implied not only cell-level reticulating processes but also intracellular & gene level-reticulating processes

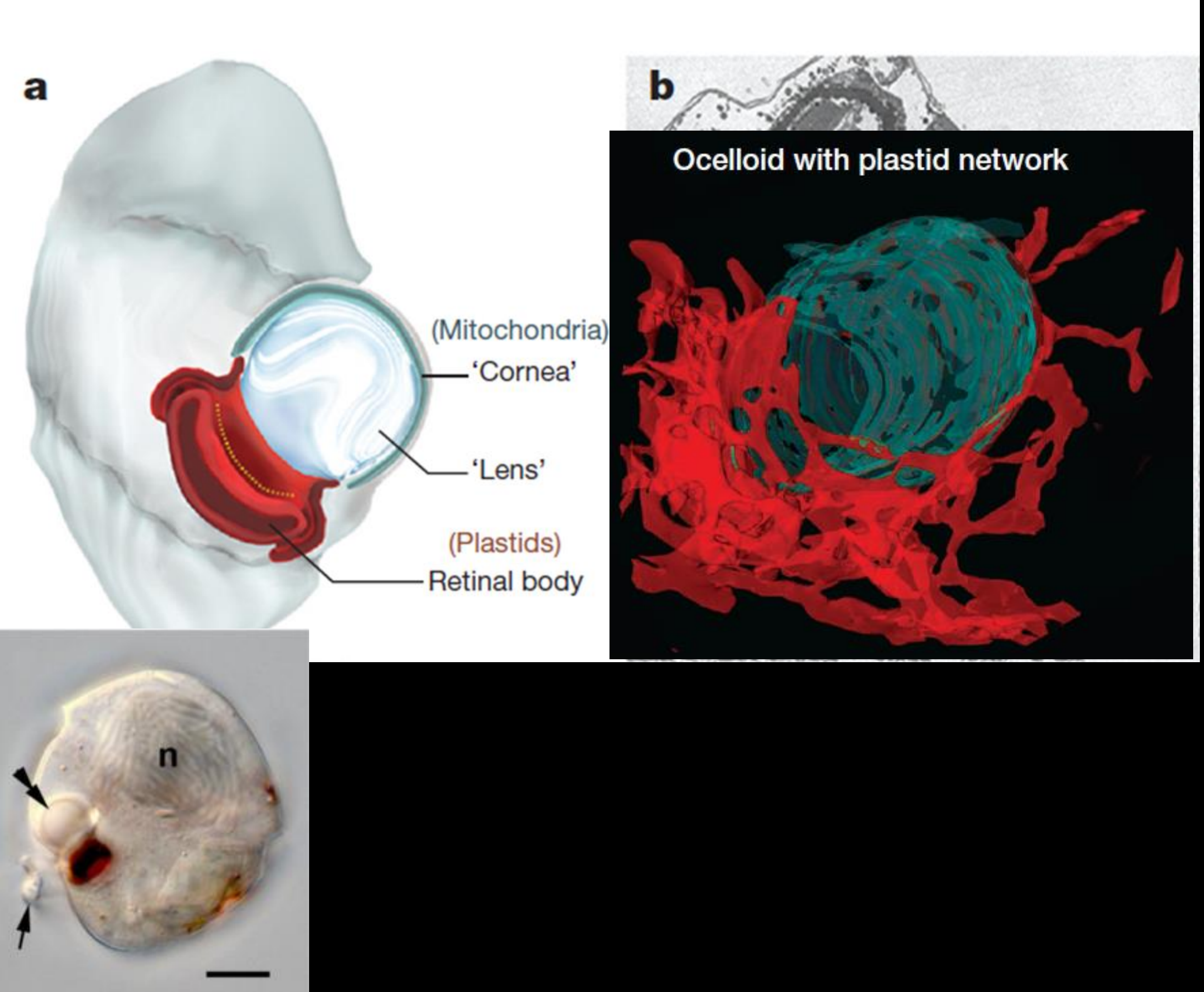
ARC BAC EUK1 EUK2

Competition between traits ?

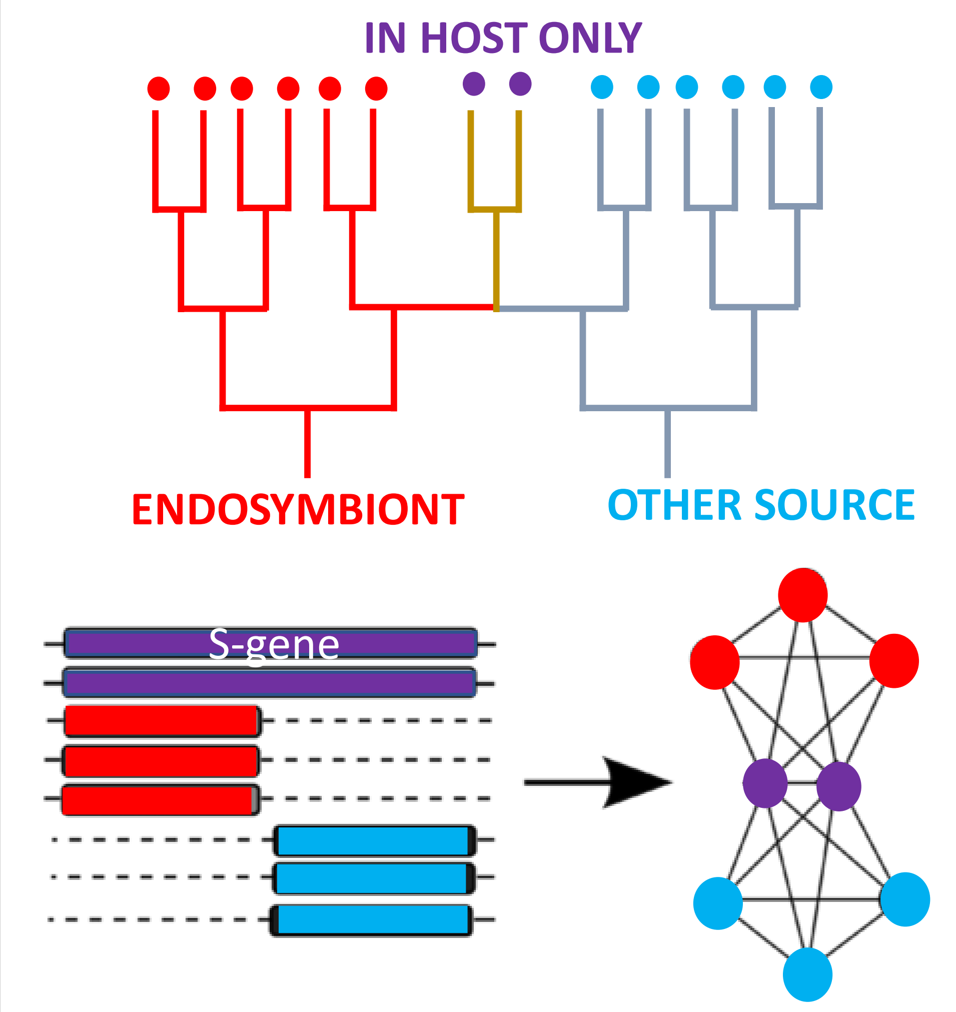
Cooperation between traits?

Chimerical systems within eukaryotes

- ‘The microbial eye’ (*Nematodinium*)



- ‘The Symbiogenetic-genes’

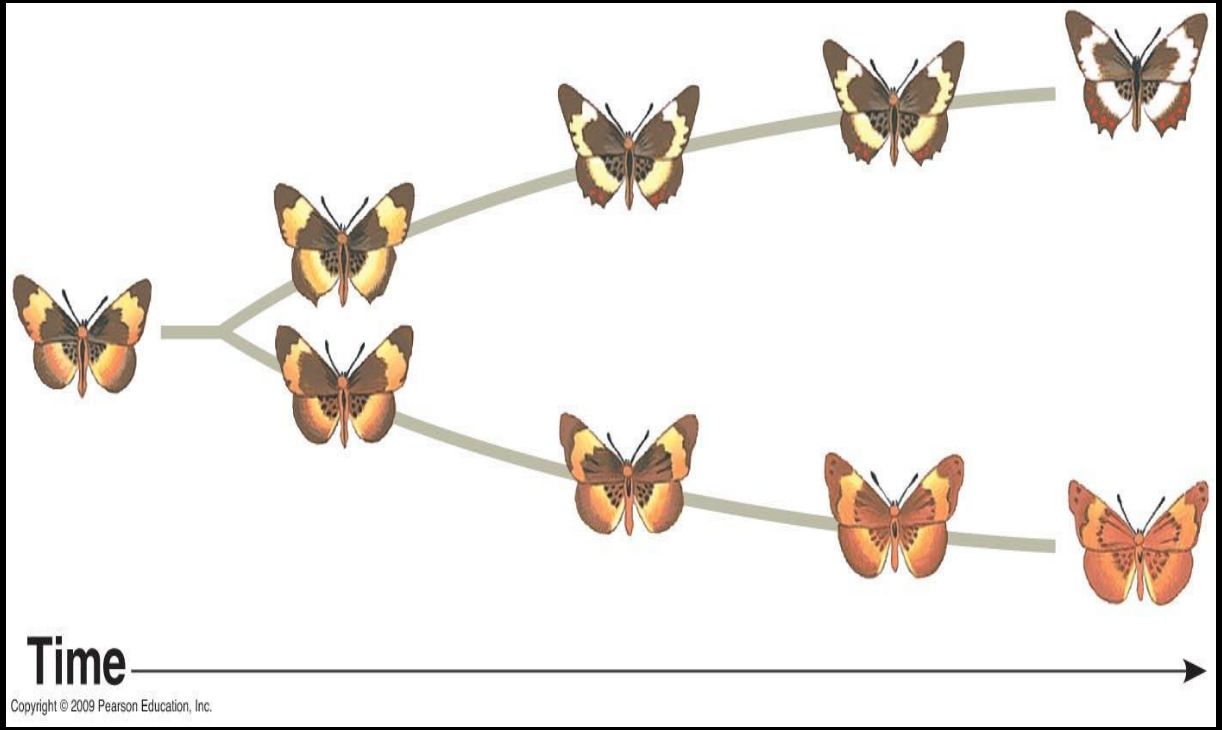


Jachiet et al. Bioinformatics. 2013;29(7):837-44.
Méheust et al. PNAS 2016; BMC Bio 2018

Such (nested) reticulate evolution was not anticipated by the classic theory

Each lineage has its own autonomous fate

Some lineages fuse



≠

Success: let **offsprings similar to parents**

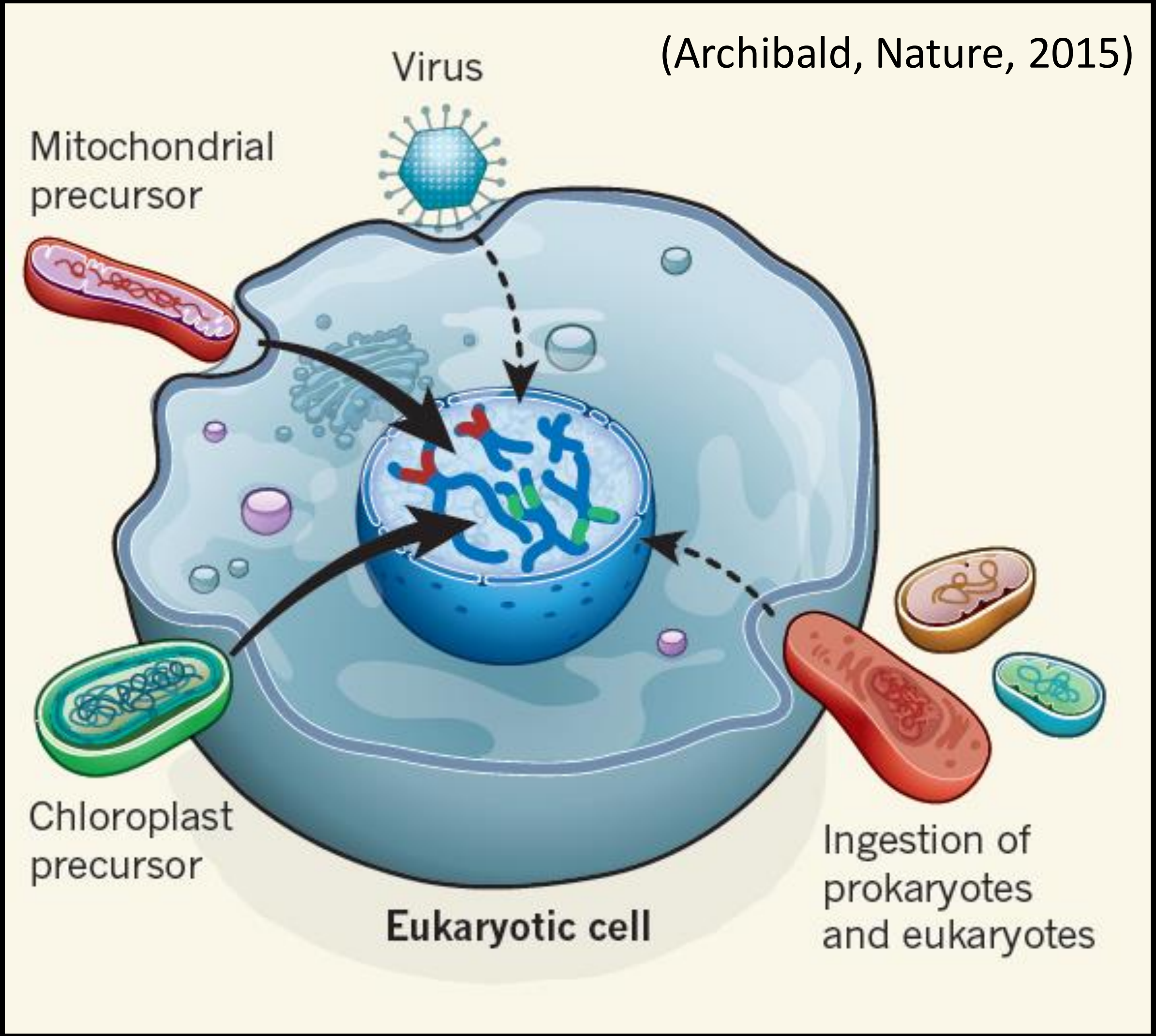
Offsprings can be very different from parents

Innovation by association

Dietel *et al.* TIM 2018

Since chimera are everywhere, naïve applications of Ockham's razor are questioned.

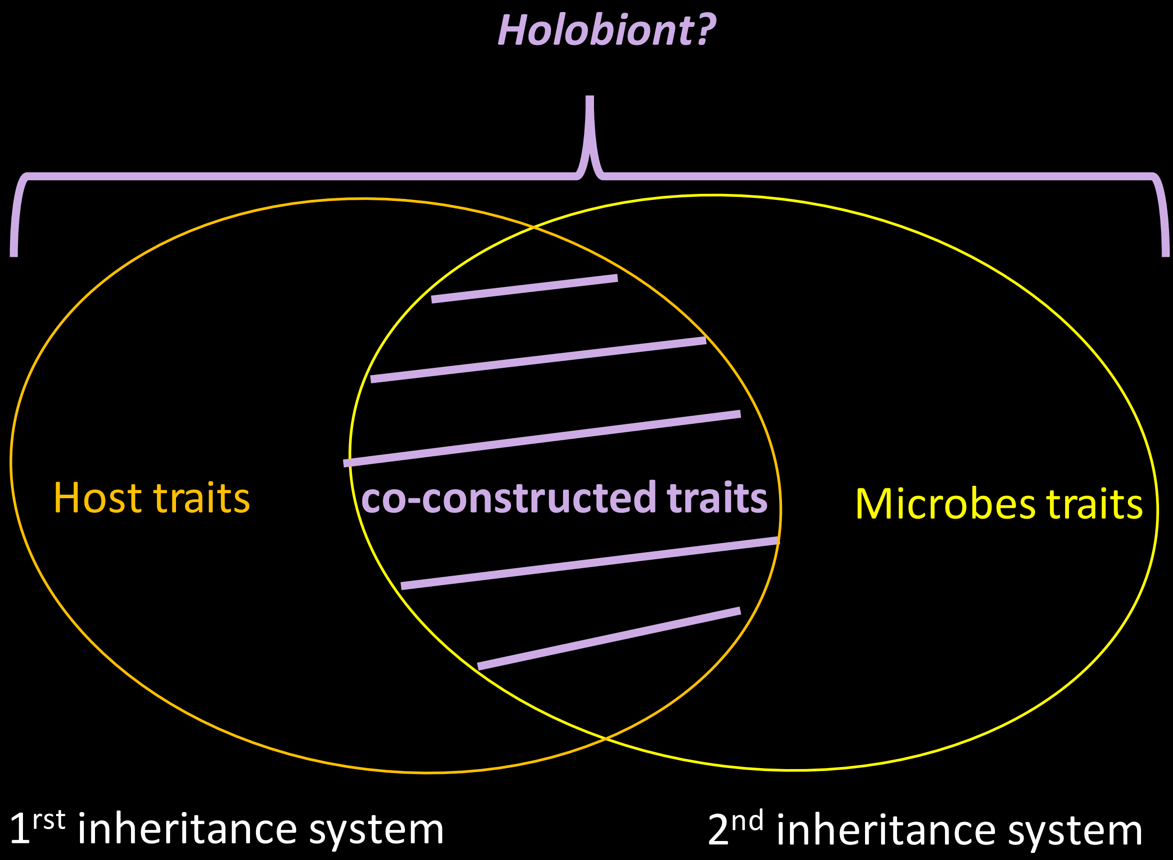
- « Beings should not be needlessly multiplied. »



(Archibald, Nature, 2015)

- **Organismal traits can be/ often are co-constructed!**

On the origins of traits: animals & plants as co-constructions?


Holobiont?

Host traits

co-constructed traits

Microbes traits

1rst inheritance system

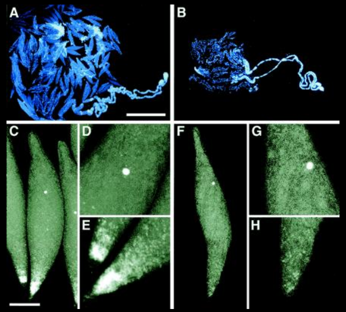
2nd inheritance system

Holobiont = a proposed association of microbes and an host, which could be selected by natural selection

Many documented cases of microbial impact on animal traits

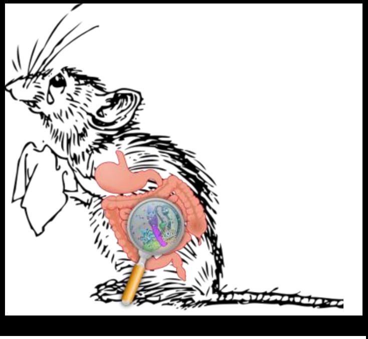
- Metamorphosis invertebrates

- Ovary formation (*Asobara wasp)*



- Antero–posterior axis (*Brugia malayi)*

- Complete gut (souris)

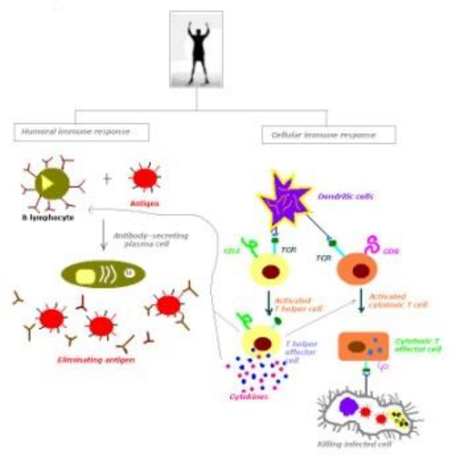


- Light Organ (bobtail squid)

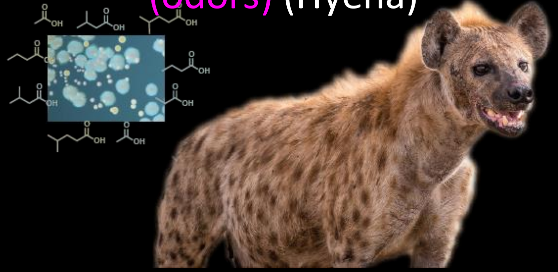
- Germination *Orchids*

- Vertebrate Immune System

Moreover, animal immune system co-constructs the microbiota (e.g. with antimicrobial peptides)



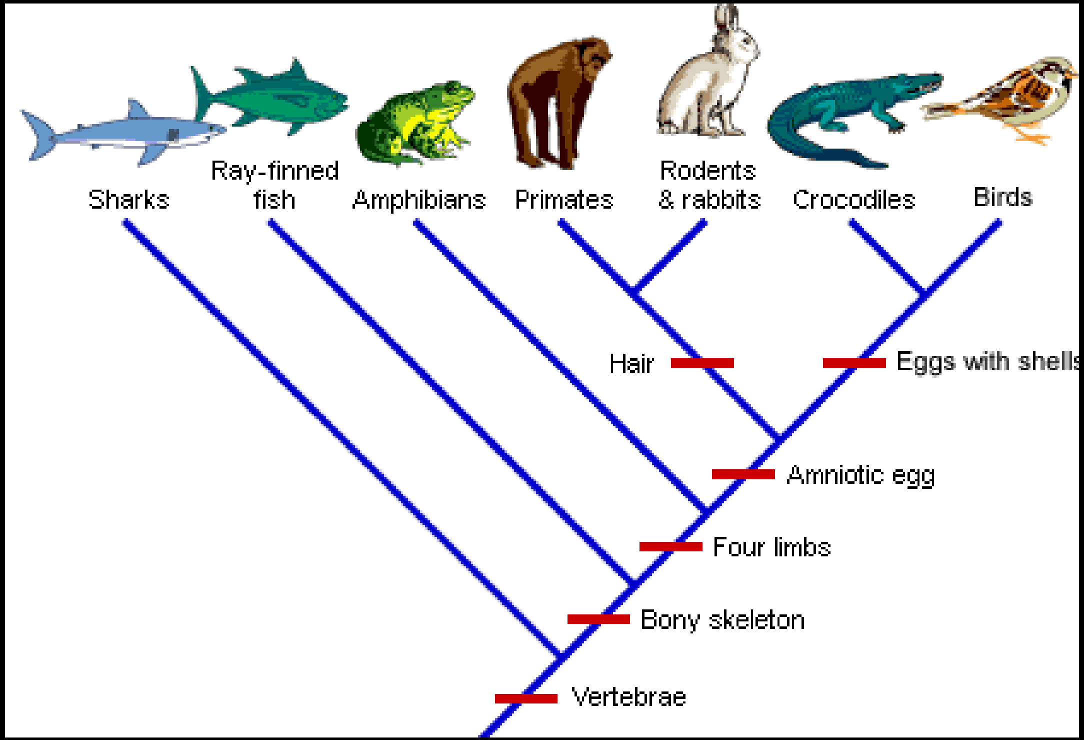
- Social relationships (odors) (Hyena)



- Mating (*Drosophila*)
- Speciation (*Nasonia*)

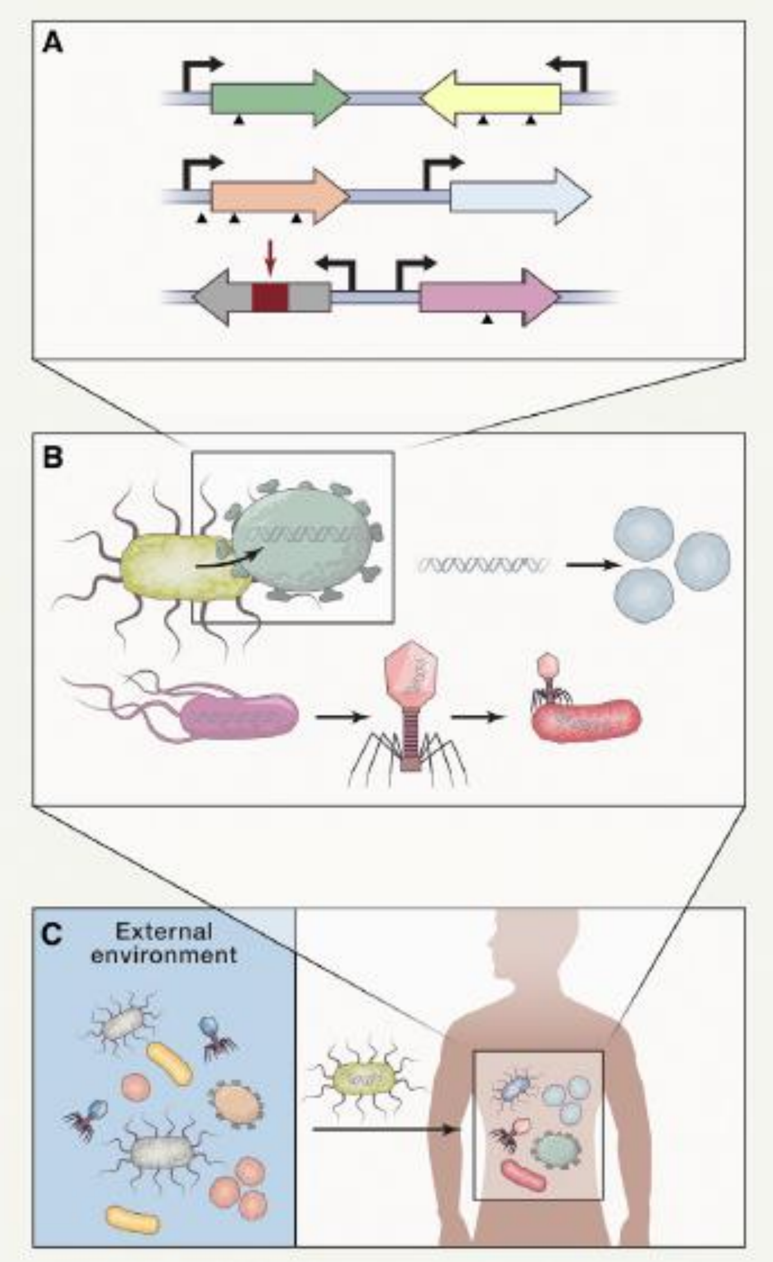
Gilbert, *Nat Rev. Genet* 2015
Theis *et al.* PNAS 2013
Brucker & Bordenstein Science 2012

On the origins of traits: animals & plants as co-constructions?

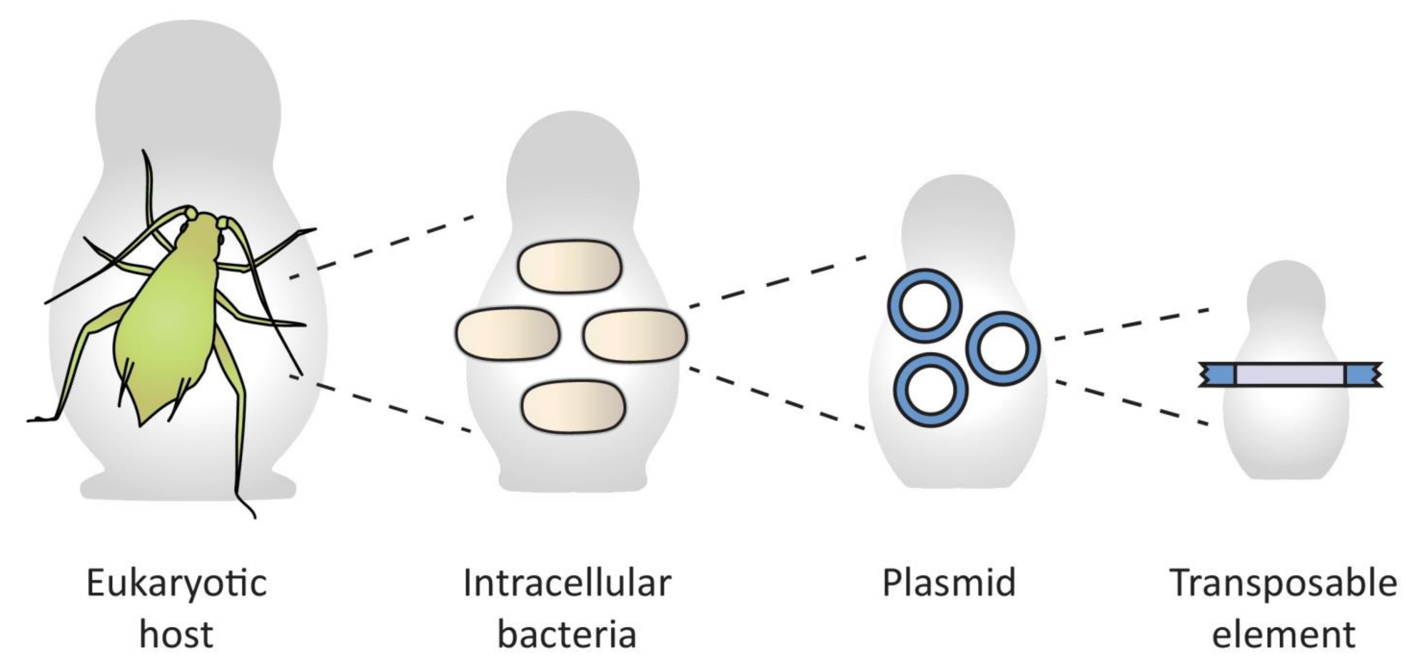


A classic phylogeny of host lineages would not account for the processes responsible for co-constructed traits.

Multipartite graphs however could highlight and help analyzing multilevel associations



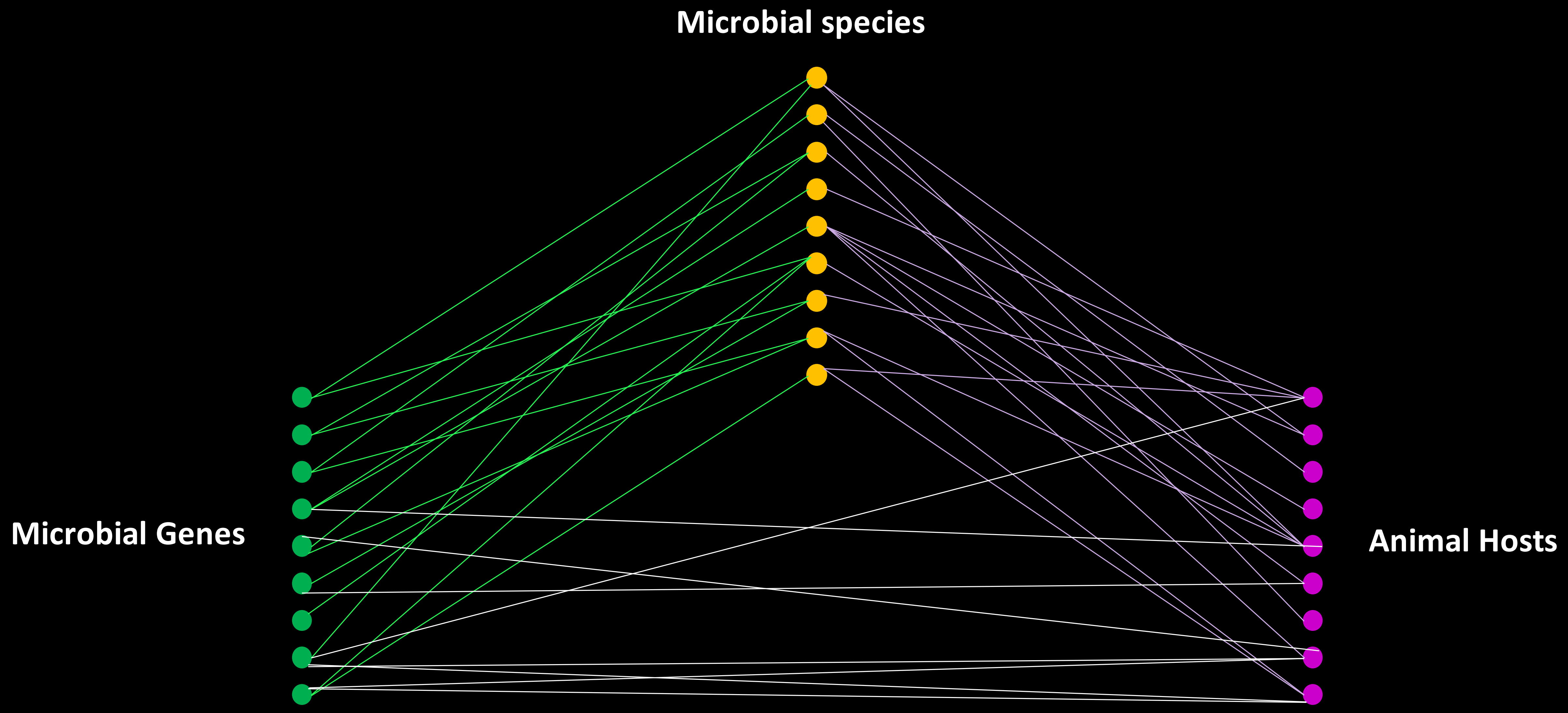
Ferreiro et al. cell 2018



Dietel TIM 2018

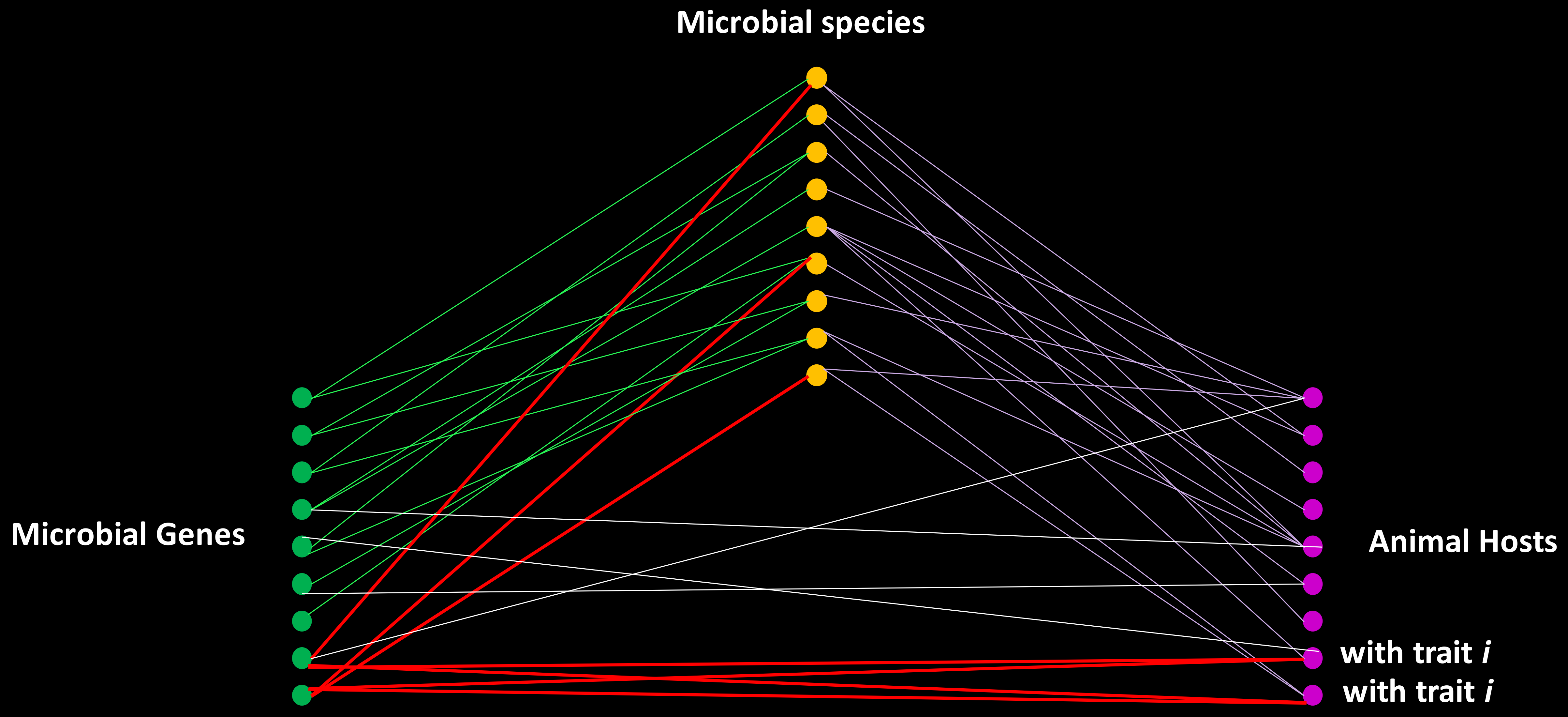
Multipartite graphs however could highlight and help analyzing multilevel associations

Corel et al. GBE 2018



Multipartite graphs however could highlight and help analyzing multilevel associations

Corel et al. GBE 2018



Microbial species

Microbial Genes

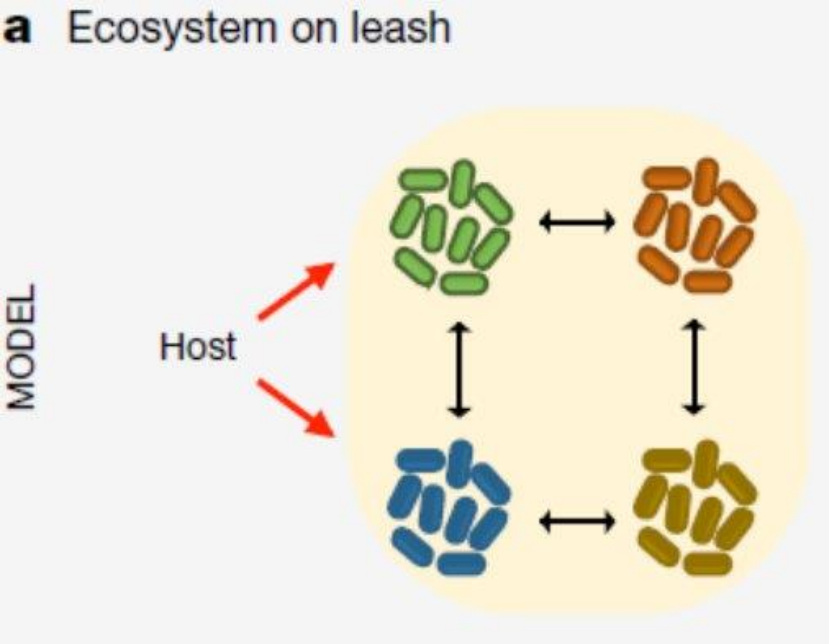
Animal Hosts

with trait *i*

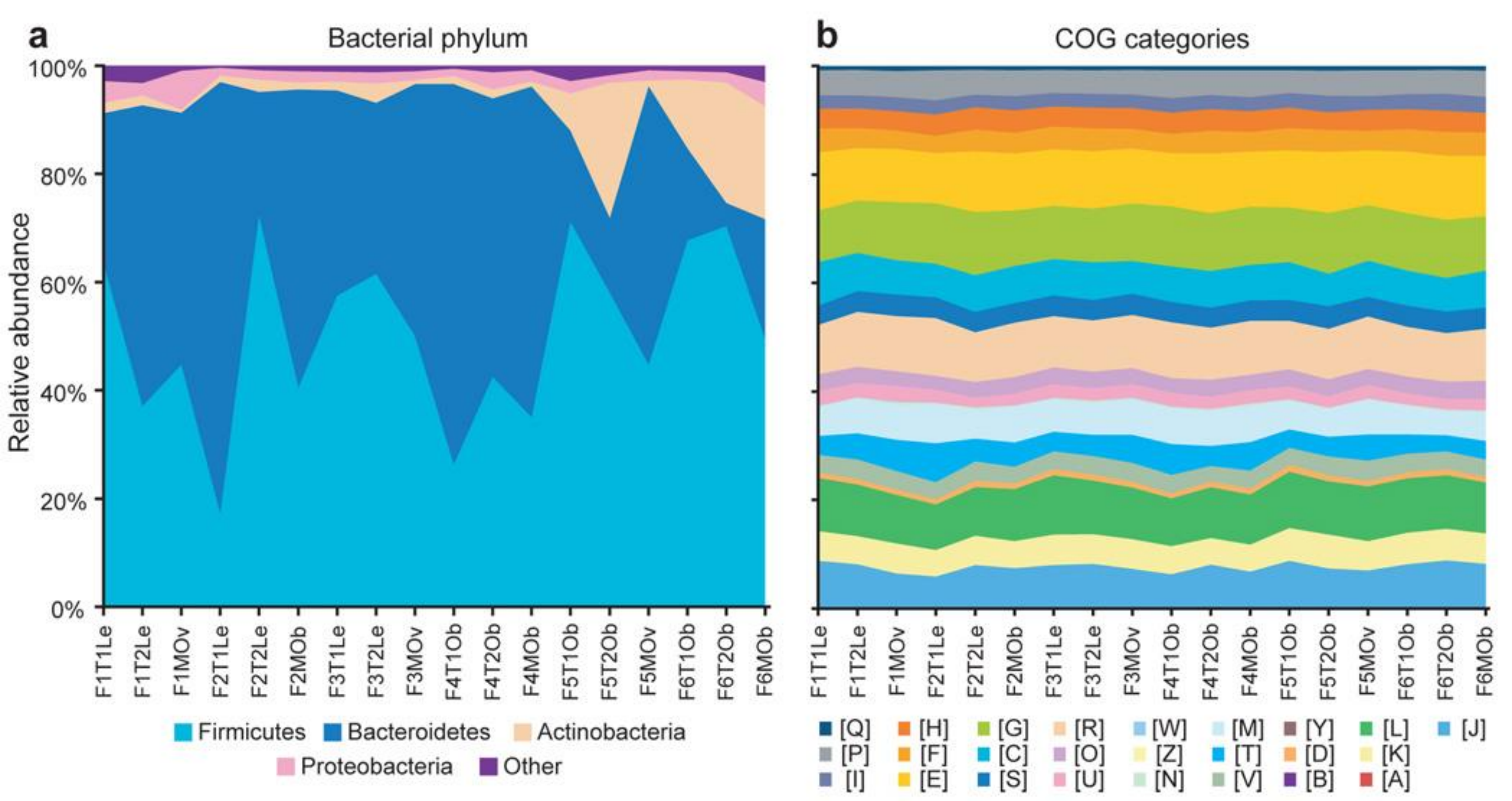
with trait *i*

e.g. genes that may be adaptive for animal hosts, irrespective of what bacteria carry them.

The song not the singer: selection by an host of the functions of an ecosystem



Foster *et al.* Nature 2017

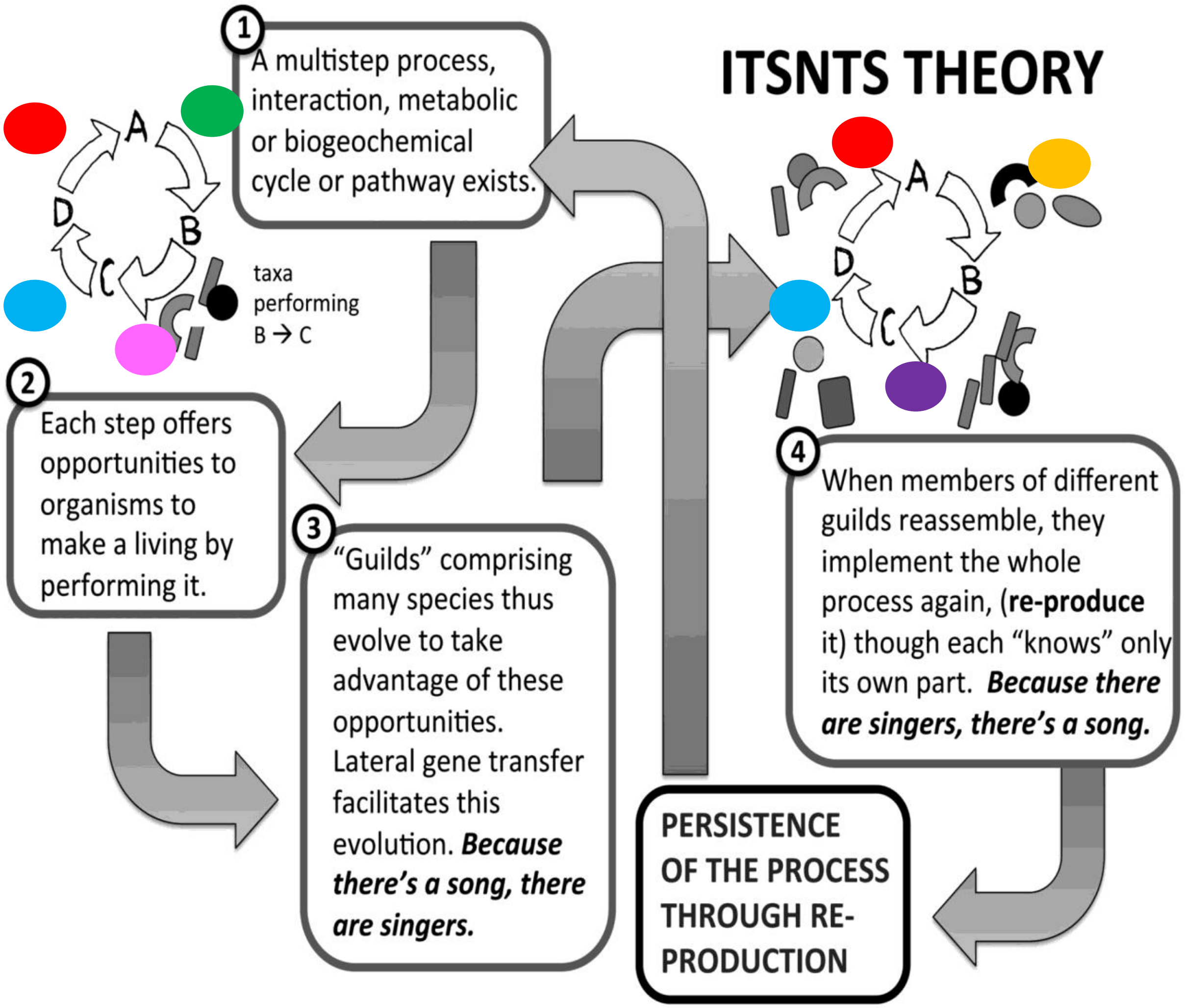


Turnbaugh *et al.* Nature 2009; Dinsdale et al. Nature, 2008; Doolittle & Zhaxybayeva, Bioscience, 2010 ; Yachi & Loreau, PNAS 1999

An ecosystem can get selected and evolve. This departs from a classic organismal-centered perspective on evolution.

A network representation also appears increasingly needed to account for ecosystem evolution

ITSNTS



① A multistep process, interaction, metabolic or biogeochemical cycle or pathway exists.

② Each step offers opportunities to organisms to make a living by performing it.

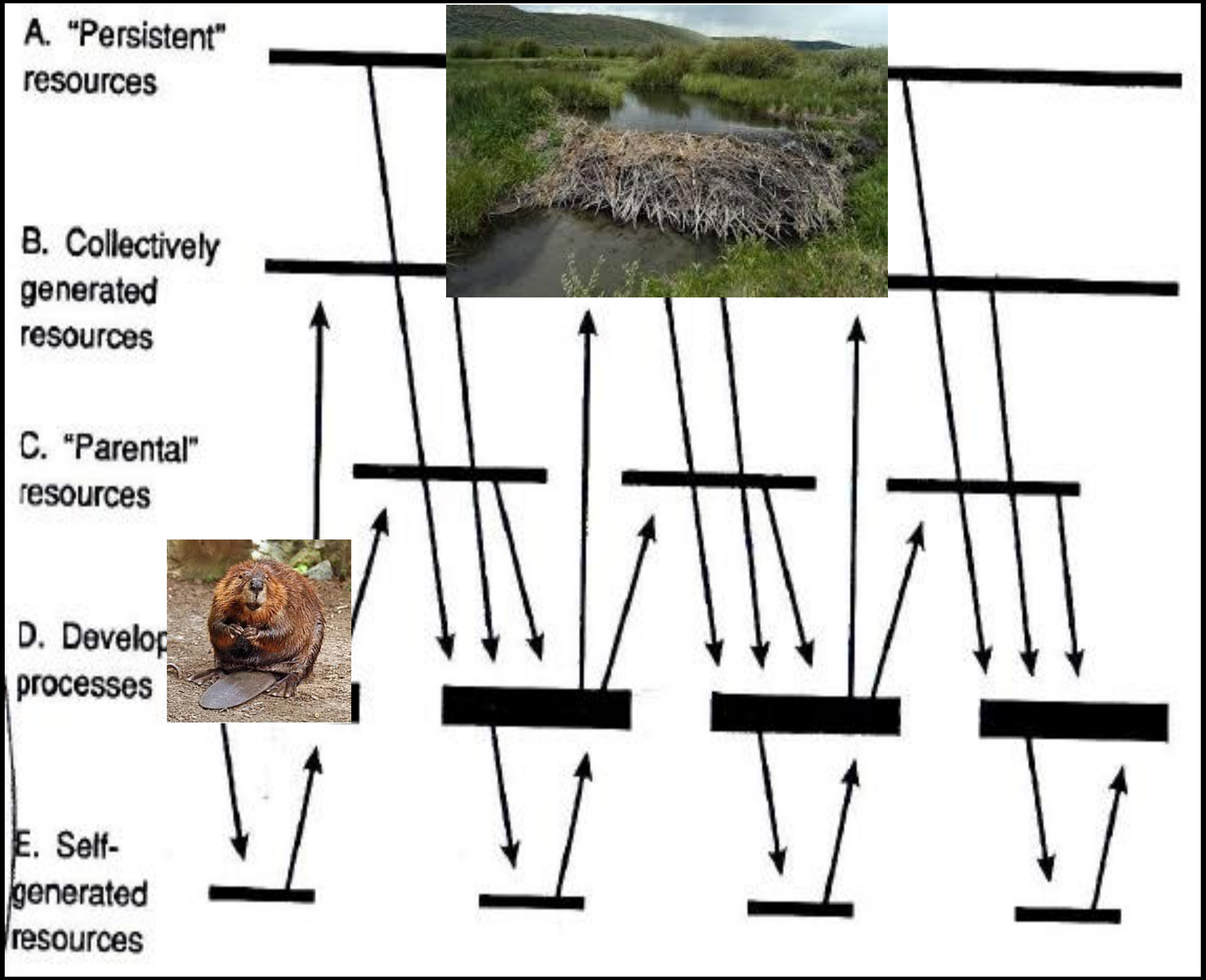
③ "Guilds" comprising many species thus evolve to take advantage of these opportunities. Lateral gene transfer facilitates this evolution. ***Because there's a song, there are singers.***

④ When members of different guilds reassemble, they implement the whole process again, (**re-produce** it) though each "knows" only its own part. ***Because there are singers, there's a song.***

PERSISTENCE OF THE PROCESS THROUGH RE-PRODUCTION

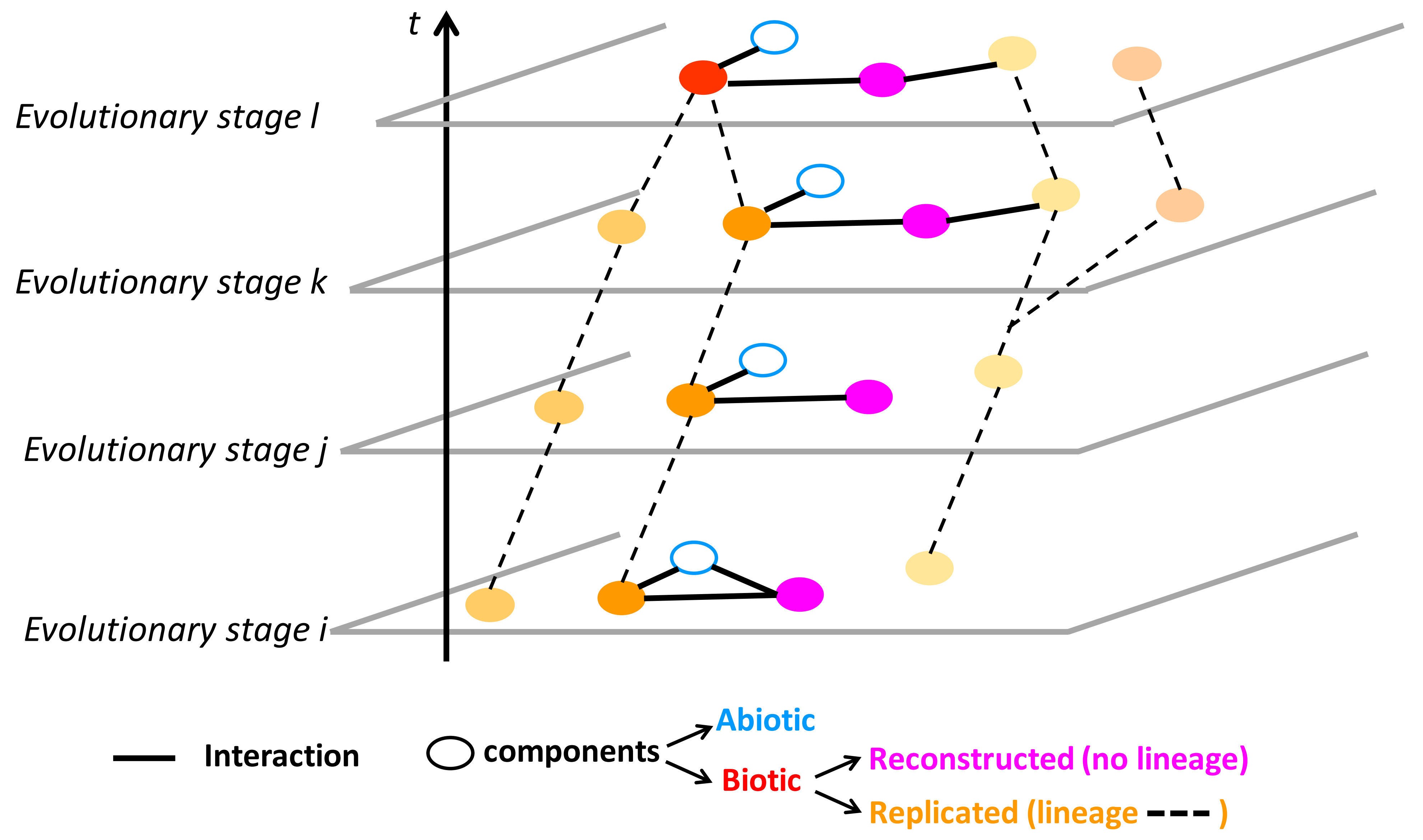
Doolittle & Inkpen PNAS 2018

DST



Griffiths & Gray, 1994

Evolutionary biology as a science of dynamic interaction networks



Interaction

components → Abiotic

→ Biotic → Reconstructed (no lineage)

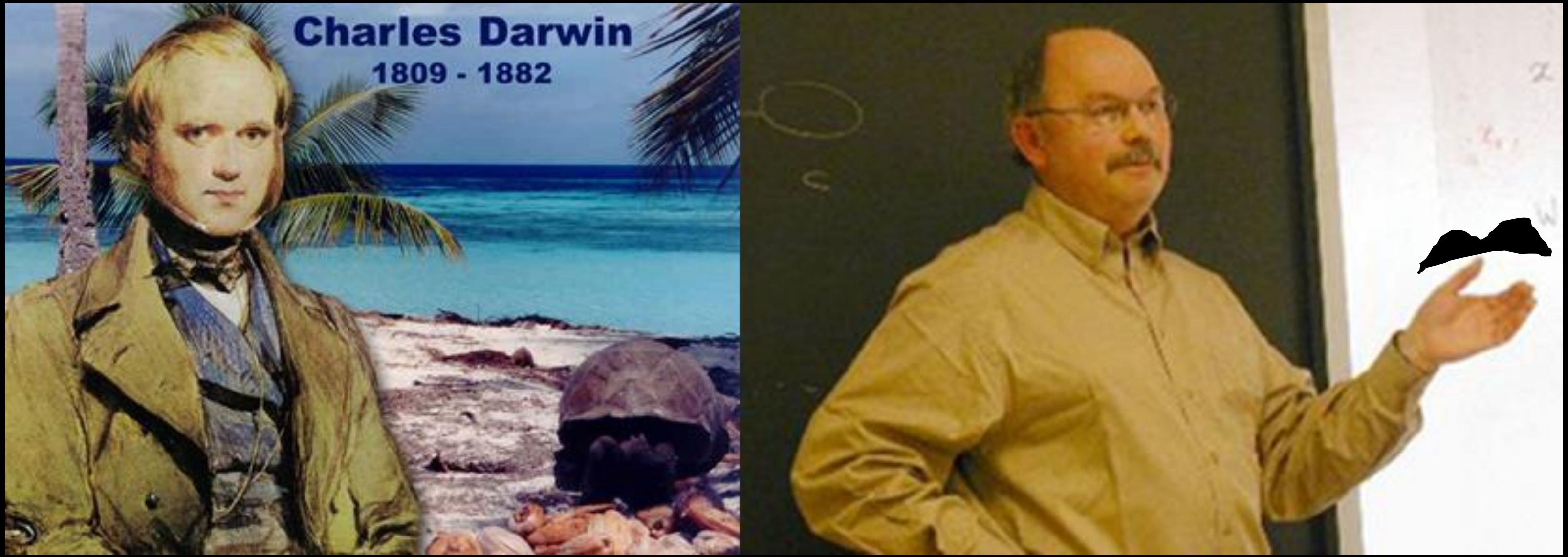
→ Replicated (lineage - - -)

Bapteste & Huneman, BMC Bio. 2018

I have shown many little networks.

Networks are very abstract tools suited for biology.

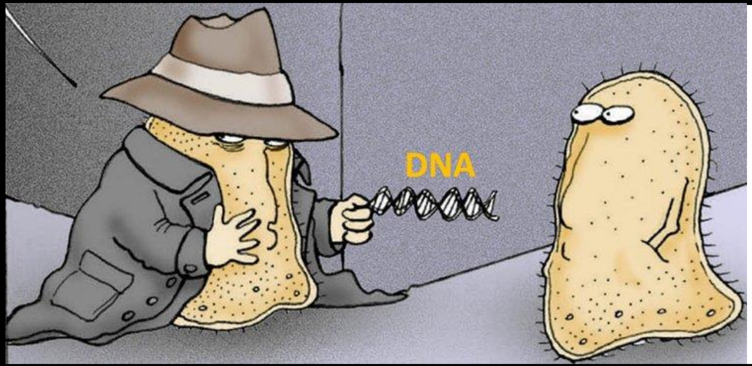
Because such networks will soon be everywhere in biology, there is an increasing room for collaboration between graph theoricians and evolutionary biologists.

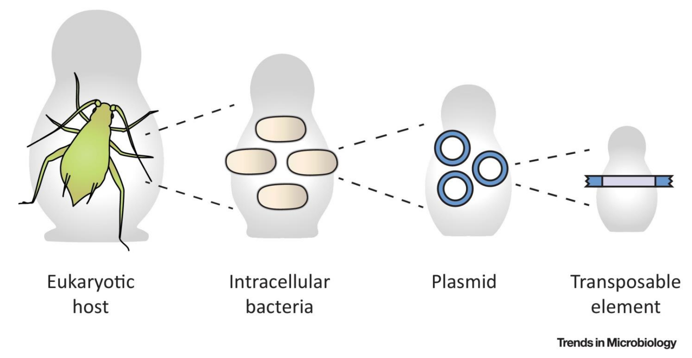


From this, conceptual changes could arise because explicititing the reticulate aspects of life interfere with the 3 evolutionary conditions of the classic theory

1) **Production of (advantageous) variation**

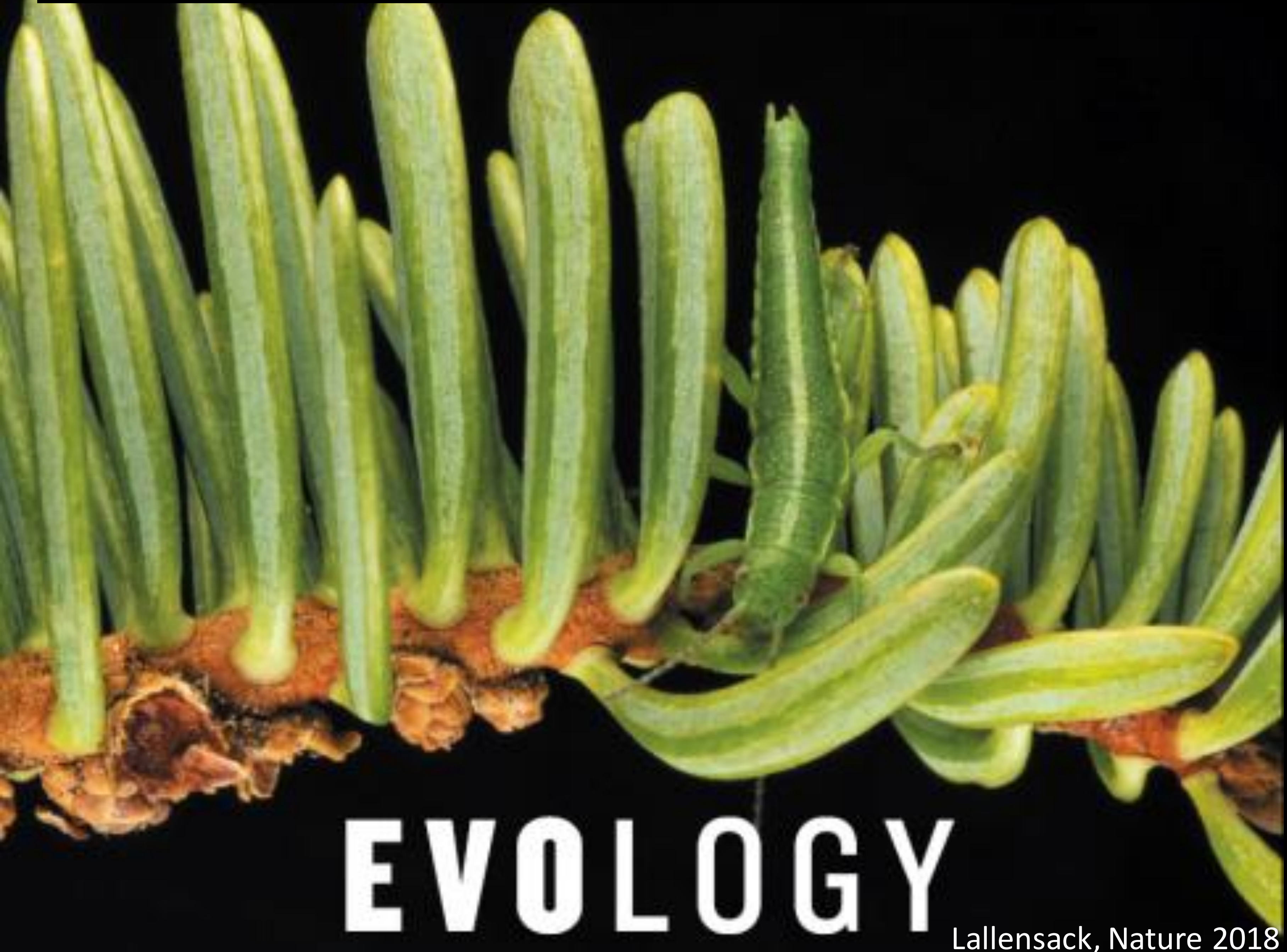

2) **Transmission of that (advantageous) variation to offsprings**





3) **Fitness:** increased ability of organisms with advantageous variations to produce more offsprings


Fitness is a relational property: natural selection depends on the ecological network



Lallensack, Nature 2018

Farkas *et al.* **relocate** 1,500 **green or striped stick insects** so that some insects' coloration clashed with their new home.

Suddenly maladapted, these insects became targets for hungry birds, and that caused a **domino effect**[1]. Birds drawn to bushes with mismatched stick insects stuck around to eat other residents, such as caterpillars and beetles, stripping some plants clean.
"It affects the entire community."

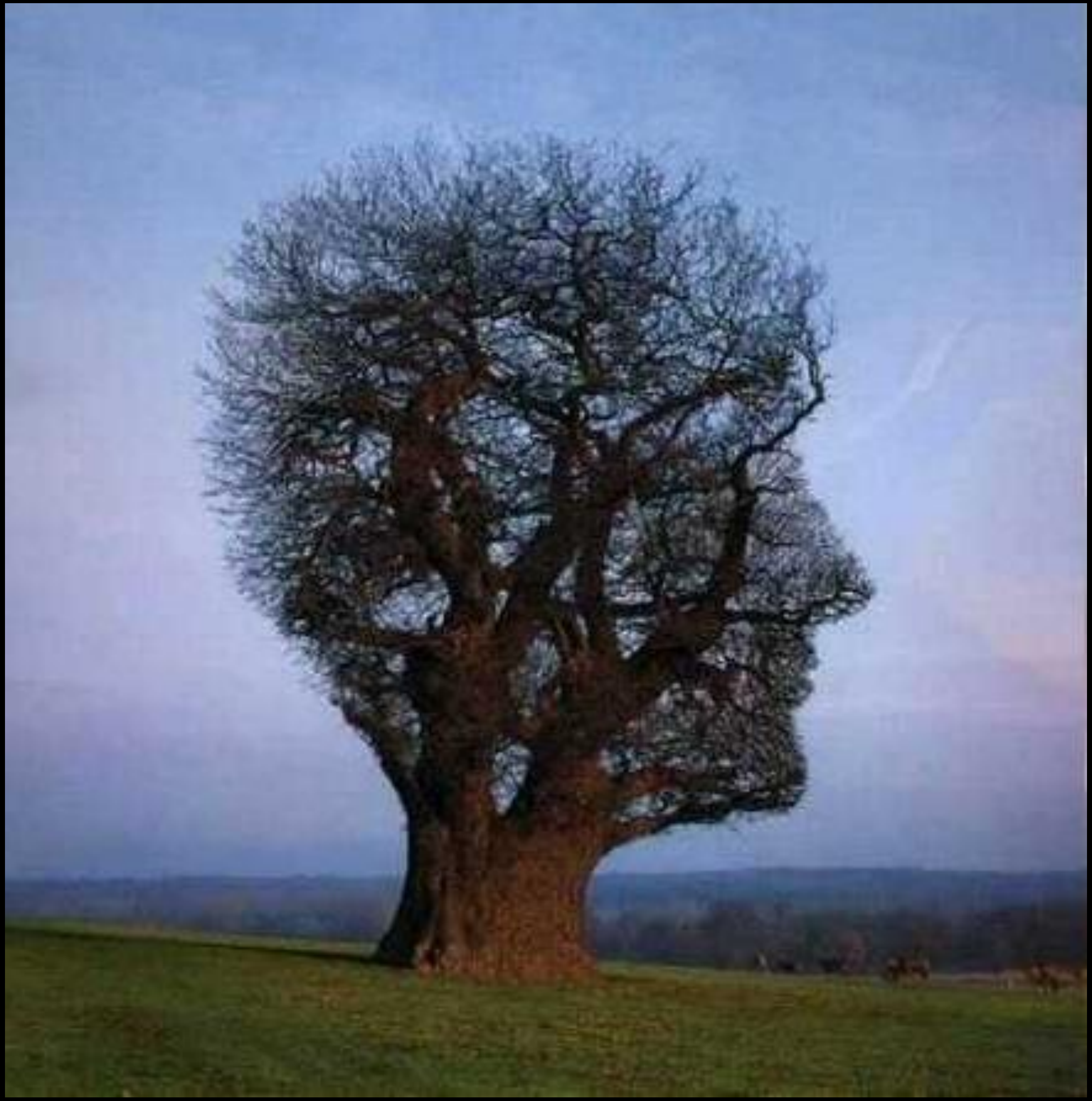
All this happened because of an out-of-place evolutionary trait.

Networks could go with important theoretical transitions in evolutionary biology

Evolutionary themes	With a phylogenetic tree	With networks
Studied objects	One lineage: organisms or genes	**Multilevel**
Studied processes	Vertical descent	**Vertical and horizontal descent**
Evolutionary perspective	Phylogenetic	**Diversity of interactions**
Explanatory Scope	Living systems	**Living Systems and Ecosystems**
Knowledge	Taxonomy & classification	**Processes**
Structure of the theory	Core founding disciplines	**Broader set of disciplines, including novel hybrid research fields**

Jordane Saget

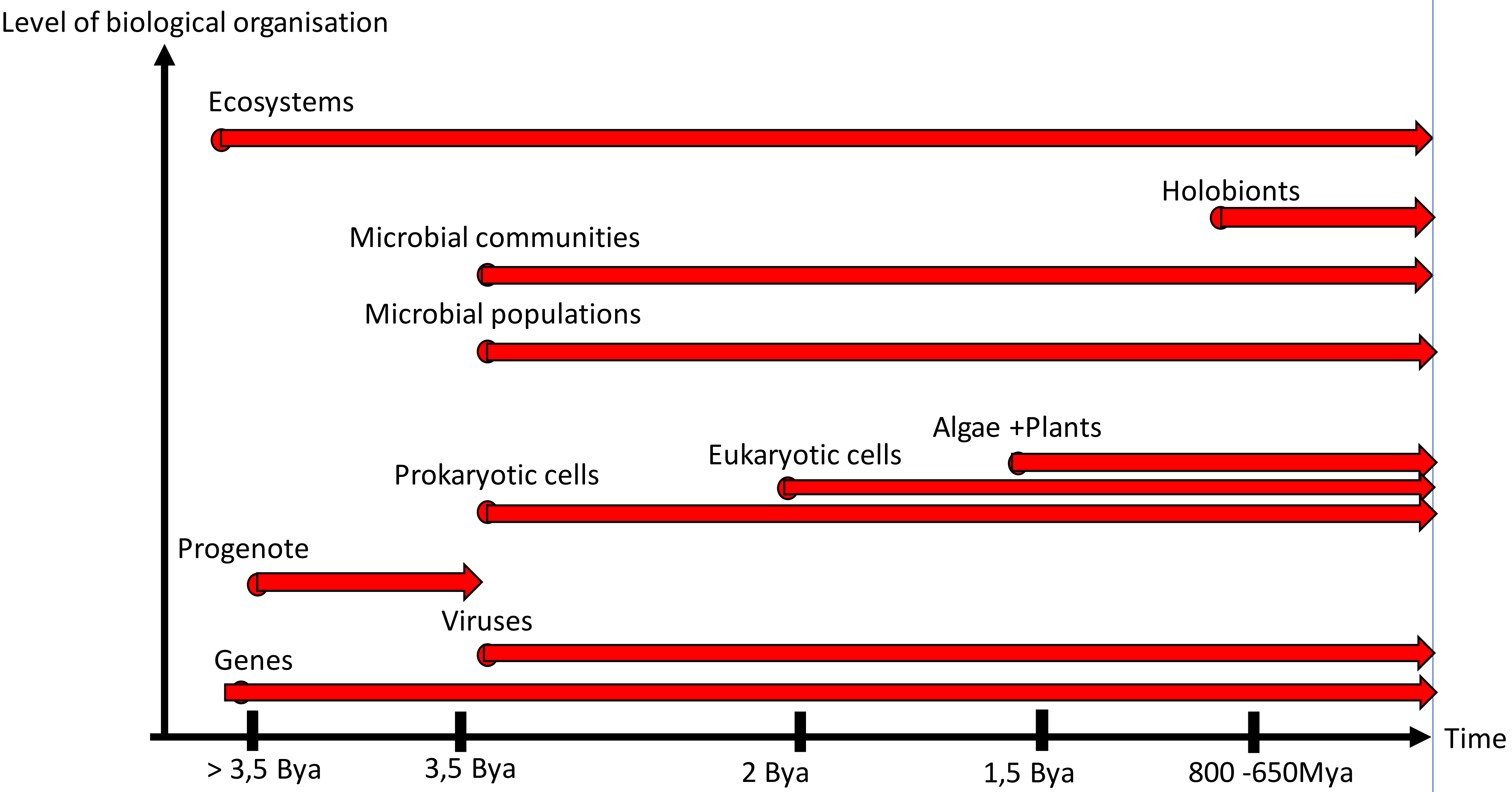
Thanks so much, Michel.



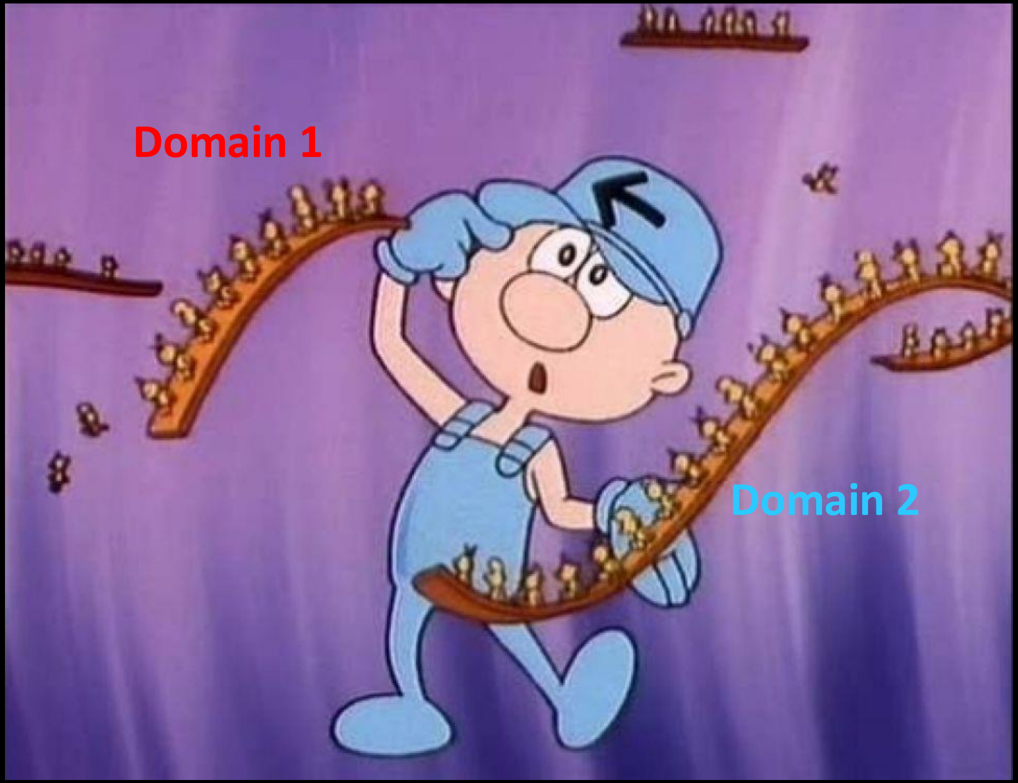
And thanks to all my great colleagues:

Phil Lopez, Ed Corel, Andrew Watson, Raph Méheust, Jananan Pathmanathan, PA Jachiet, Jessica Leigh, Seb Halary, Lucie Bittner, Romain Lannes, Ford Doolittle…

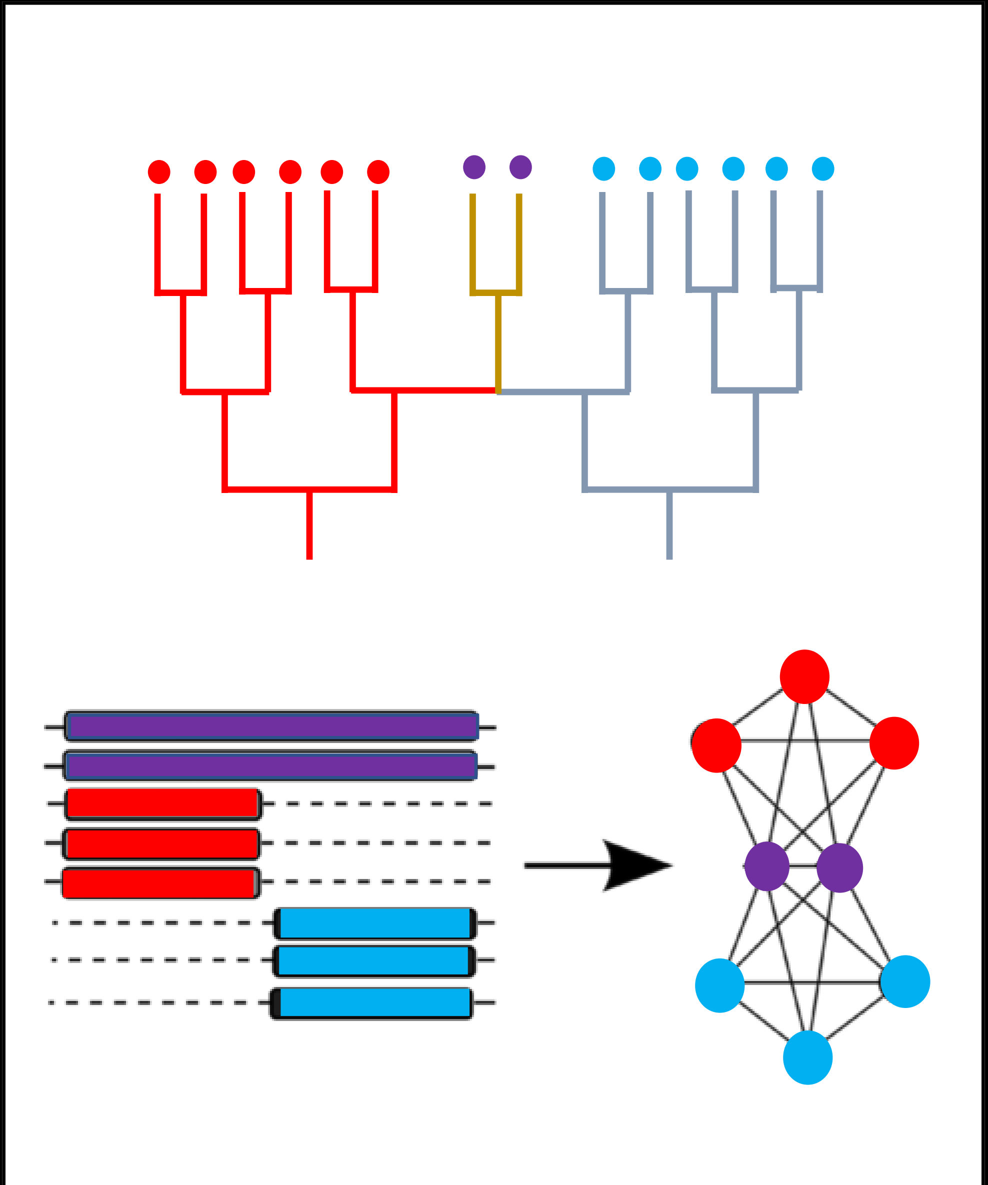
In biology, interactions and networks are and have always been everywhere.



Domain combinations is another source of genetic variation, analyzable with networks



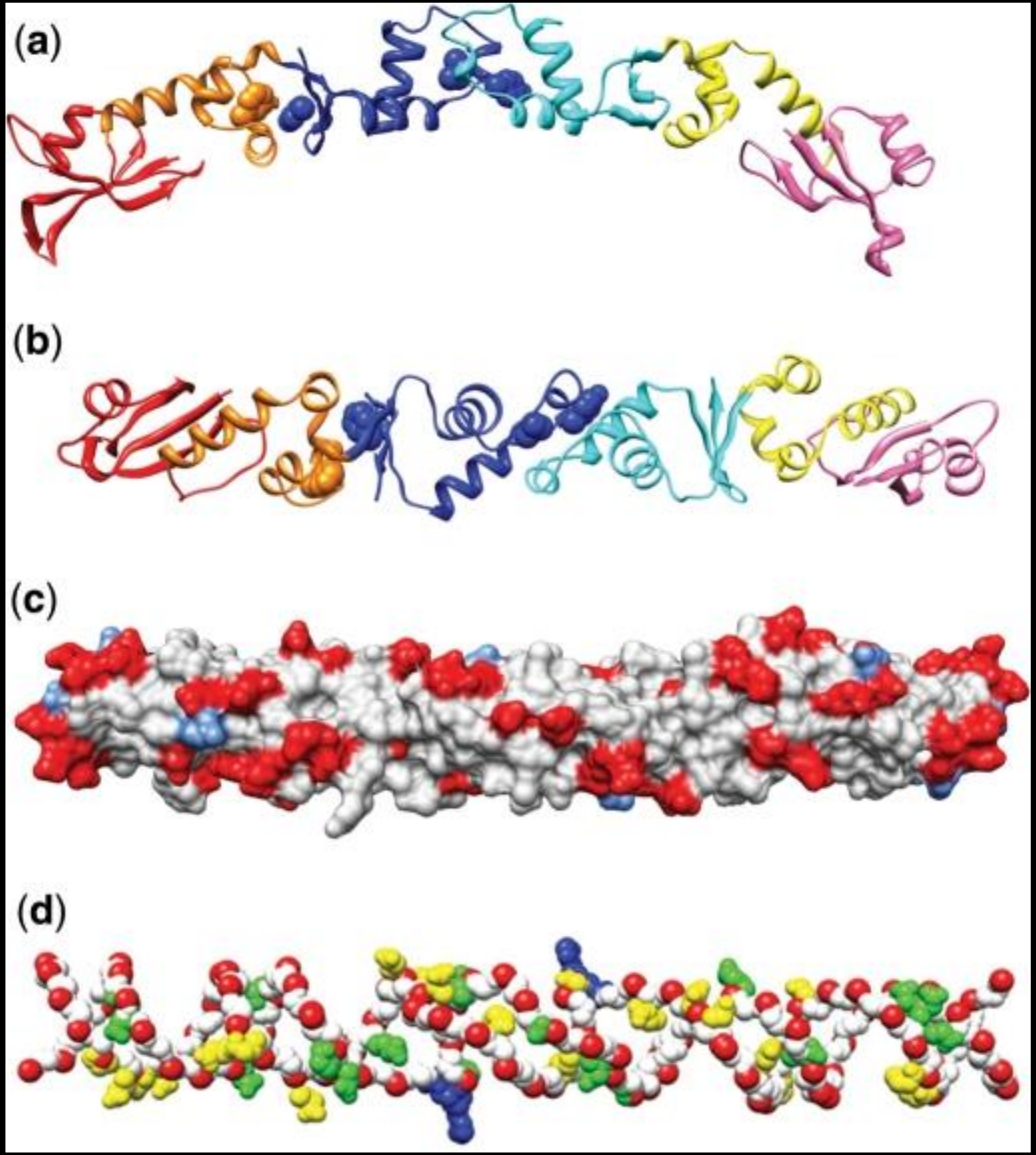
Gene lineages, like organismal lineages, appear mosaic.



Jachiet et al. Bioinformatics. 2013;29(7):837-44.

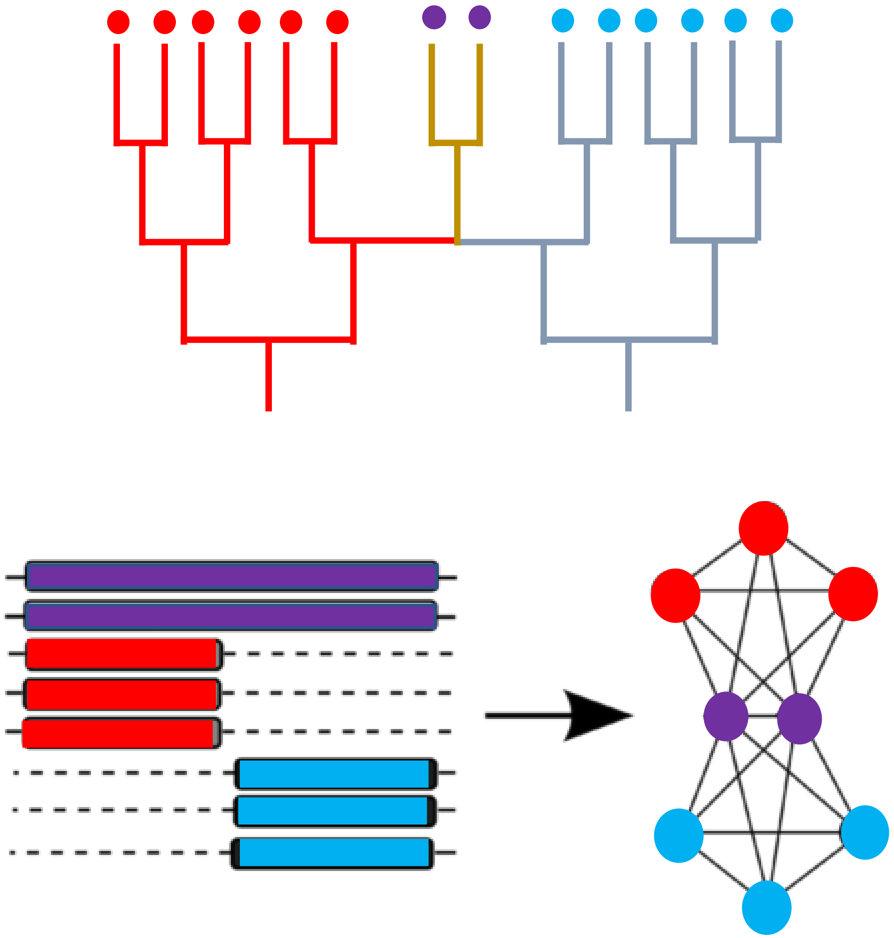
For example, it predicts and explains the evolution of various types of barriers to gene transfer

- **Fake DNA**

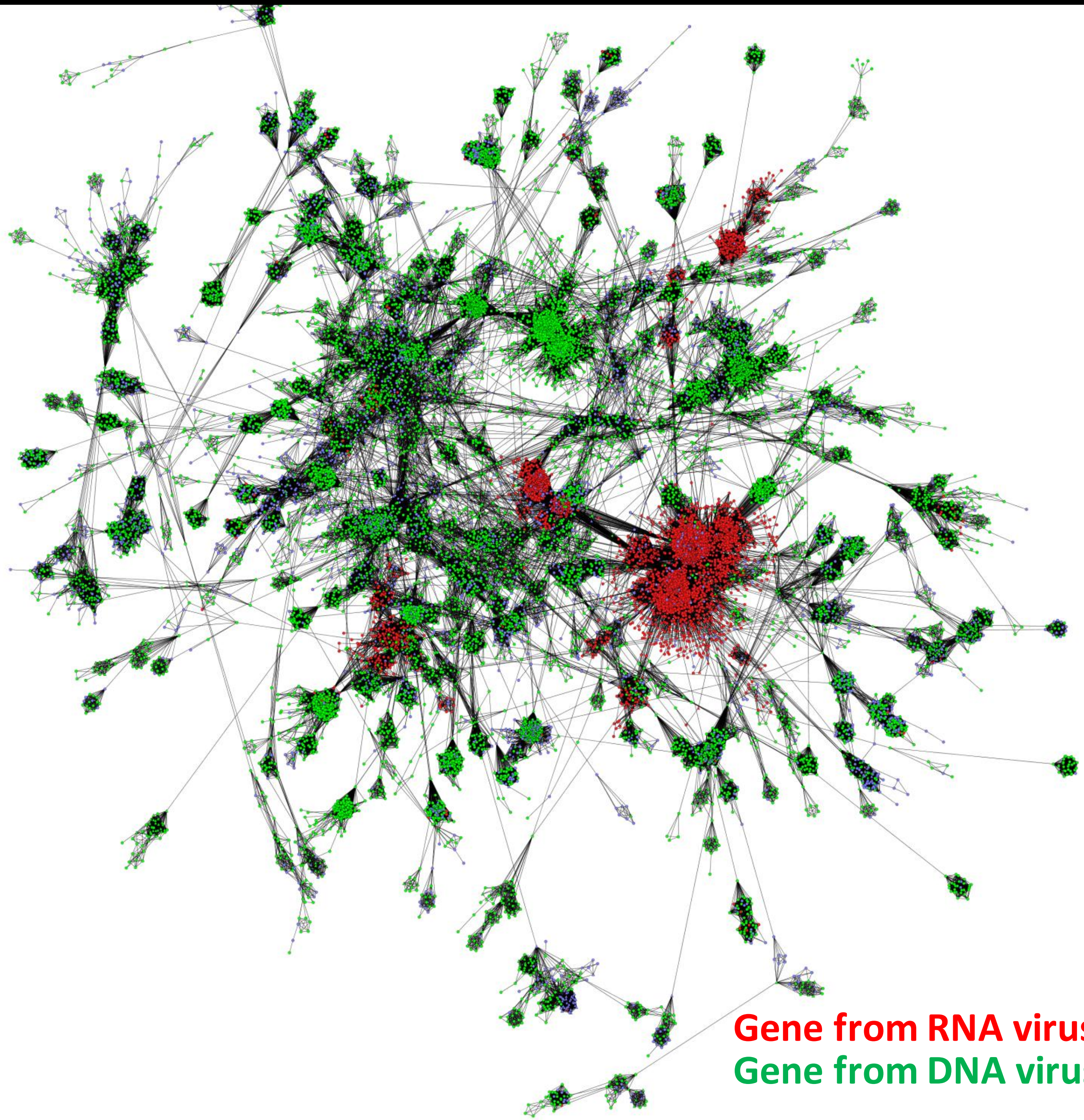


MacMahon et al. NAR 2009; Walkinshaw et al. Mol Cell. 2002

Domain combinations is another source of genetic variation, analyzable with networks

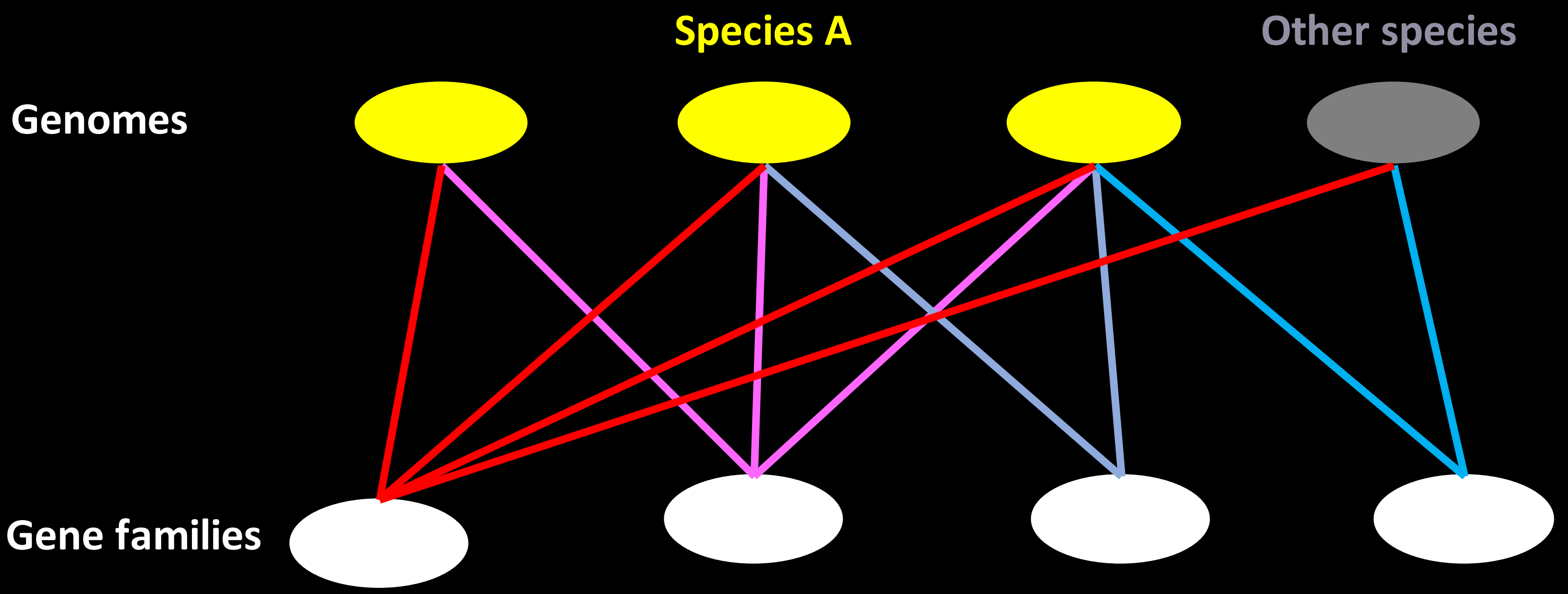


Jachiet et al. Bioinformatics. 2013;29(7):837-44.



Gene from RNA virus
Gene from DNA virus

Bipartite graphs are a practical way to characterize pangenomes for very large datasets



Species A

Other species

Genomes

Gene families

A core gene family

A core & exclusive gene family

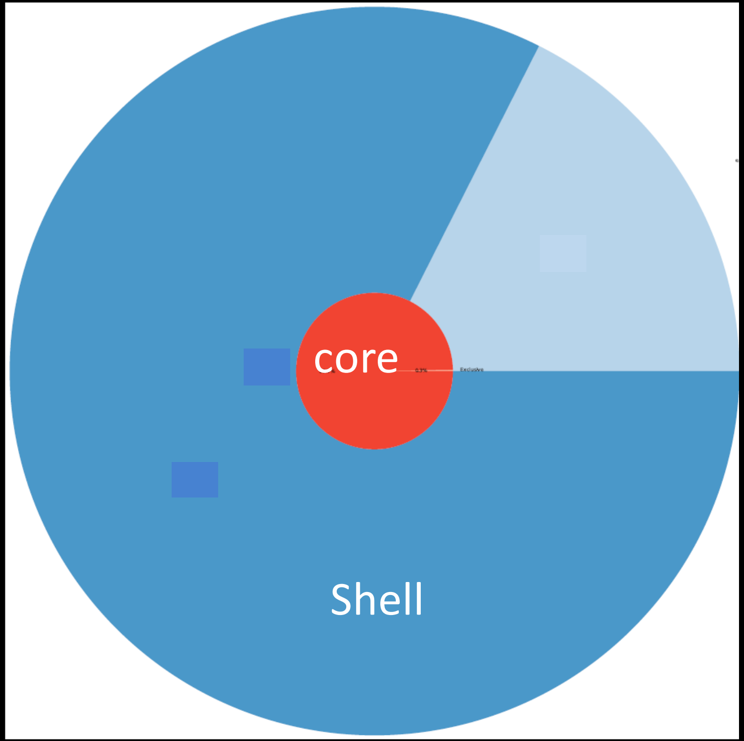
-Diagnostic feature
-Original innovation

Shell & exclusive gene family

- Original innovations

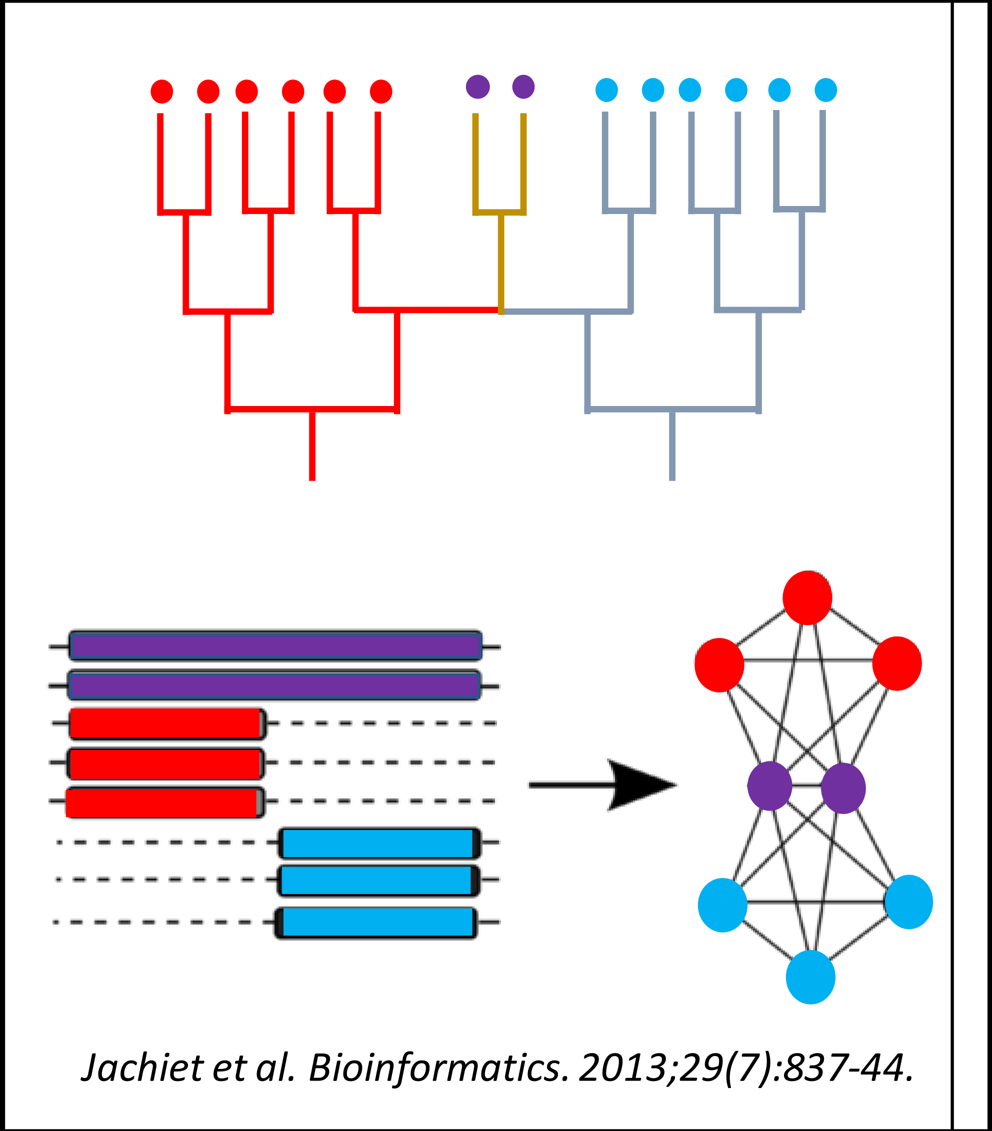
Shell gene family

Corel et al. Trends in Micro 2016



Watson *et al.* in prep

Domain combinations is another source of genetic variation, analyzable with networks



Jachiet et al. Bioinformatics. 2013;29(7):837-44.

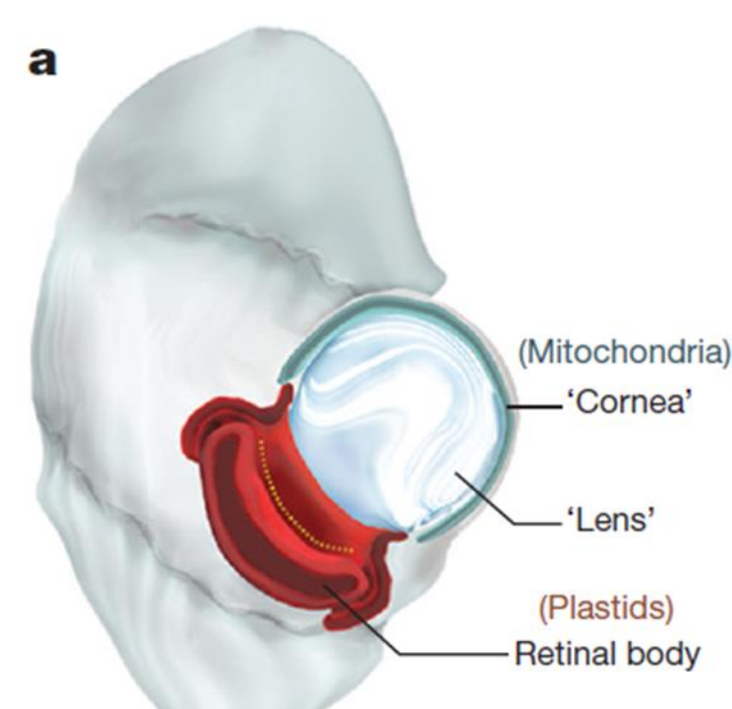
Gene from RNA virus

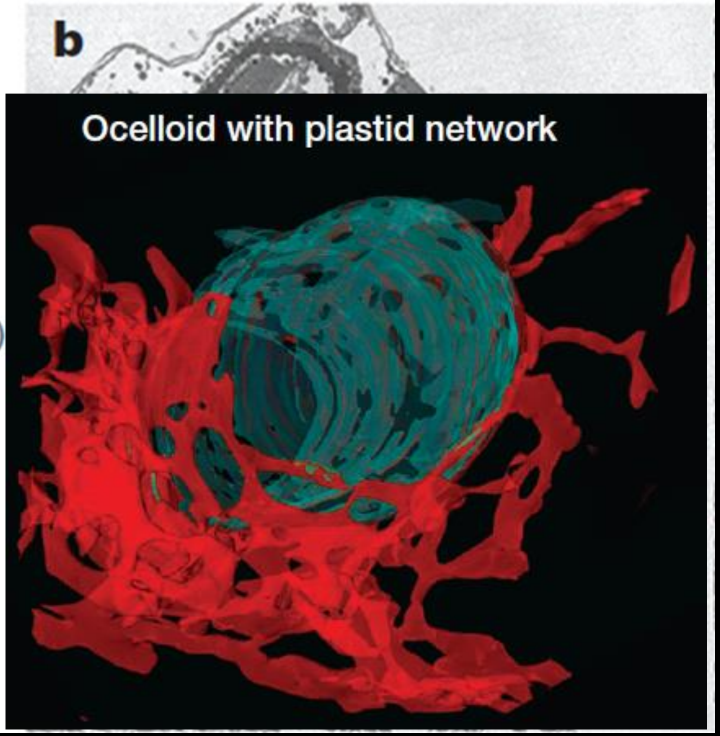
Gene from DNA virus

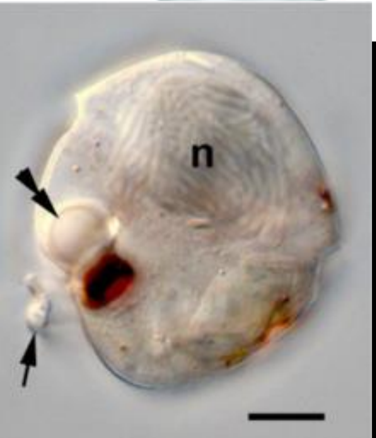
Gene lineages, like organismal lineages, appear mosaic.

➡Chimerical systems within eukaryotes

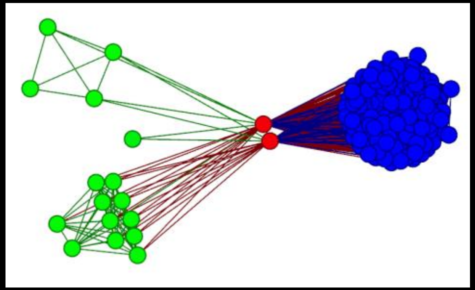
- ‘The microbial eye’ (*Nematodinium*)



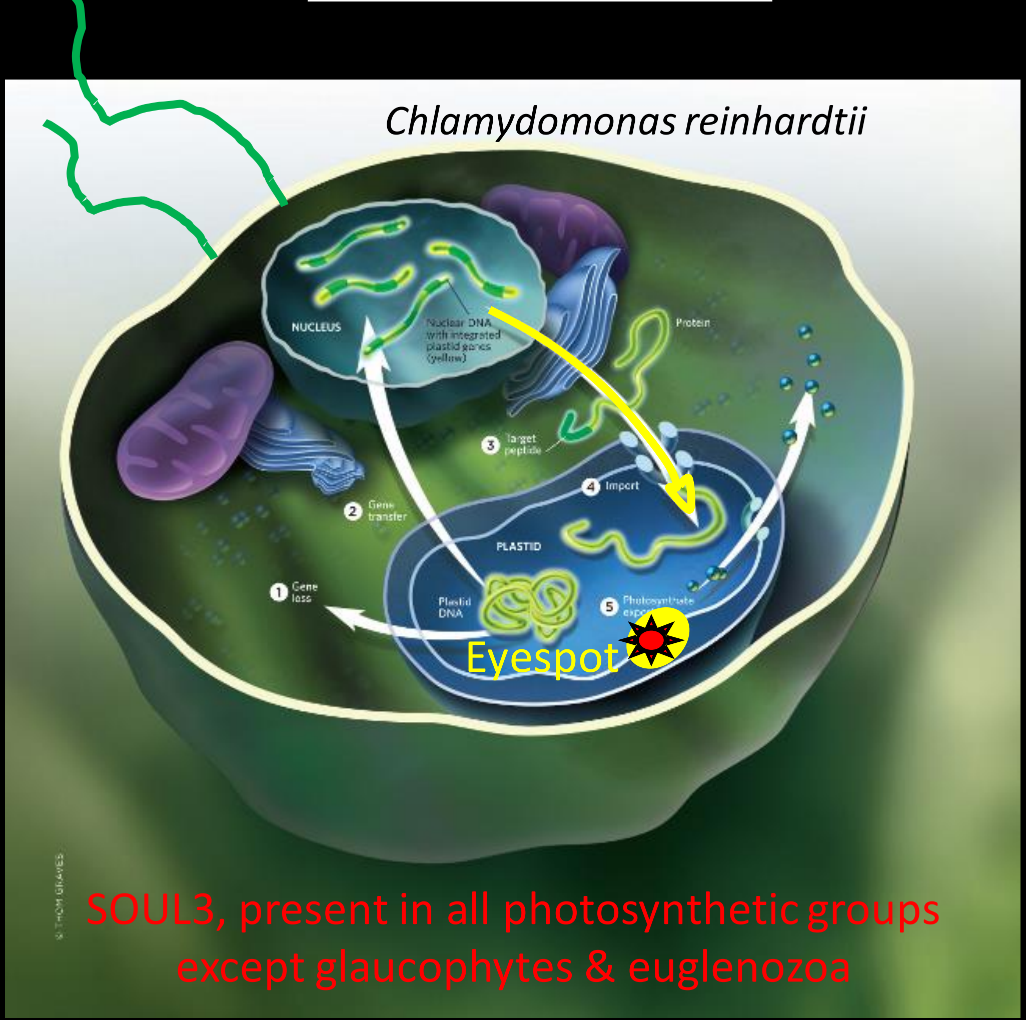




- ‘The Symbiogenetic-genes’



Chlamydomonas reinhardtii



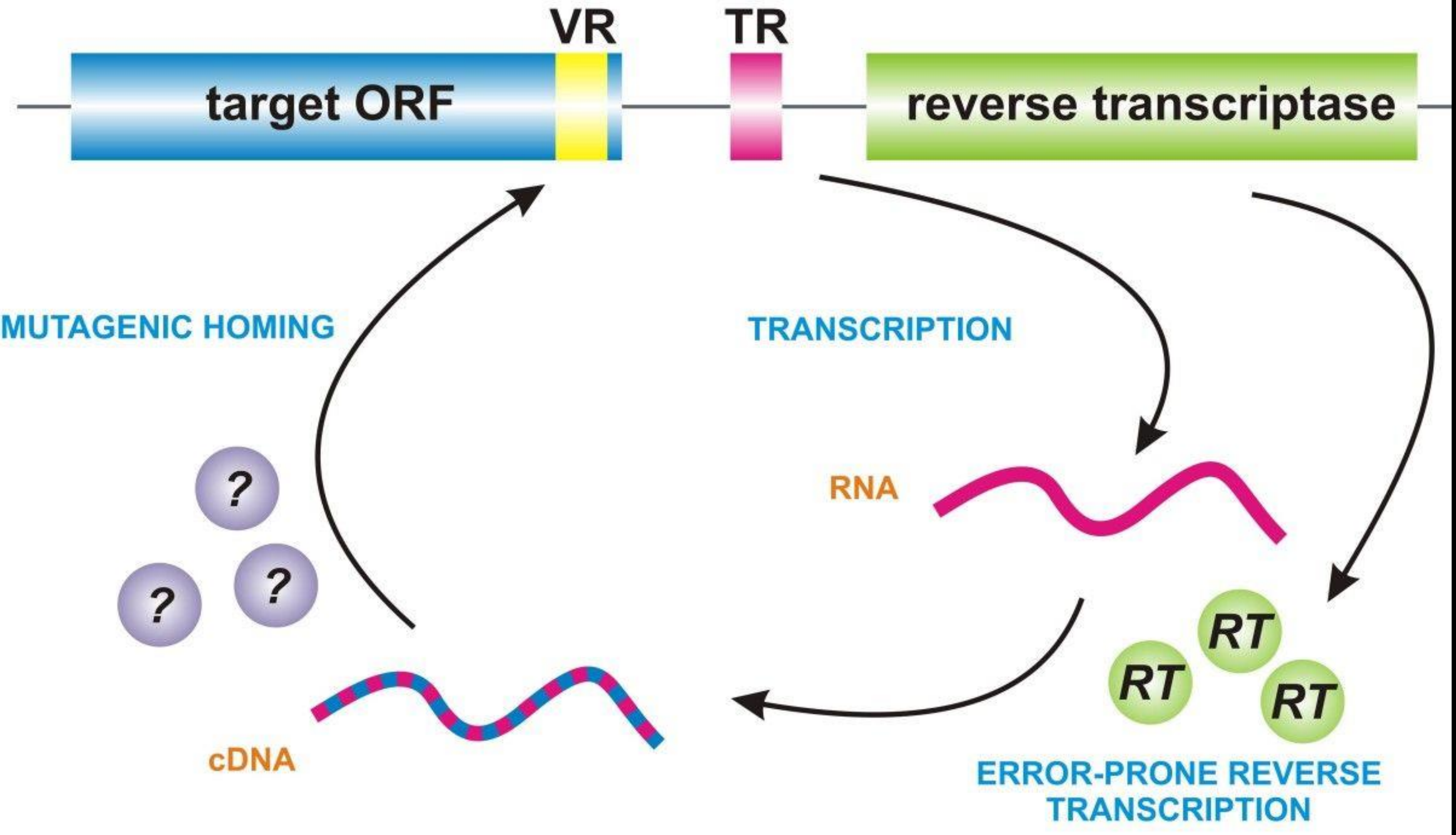
SOUL3, present in all photosynthetic groups except glaucophytes & euglenozoa

Moreover, CPR and DPANN are enriched in transferable Diversity Generating Retroelements

- Host dependence induces environmental stresses: limited energy and nutrients, + interactions with other cells.
- CPR/DPANN might adapt to these stresses by diversifying their genes using DGR.

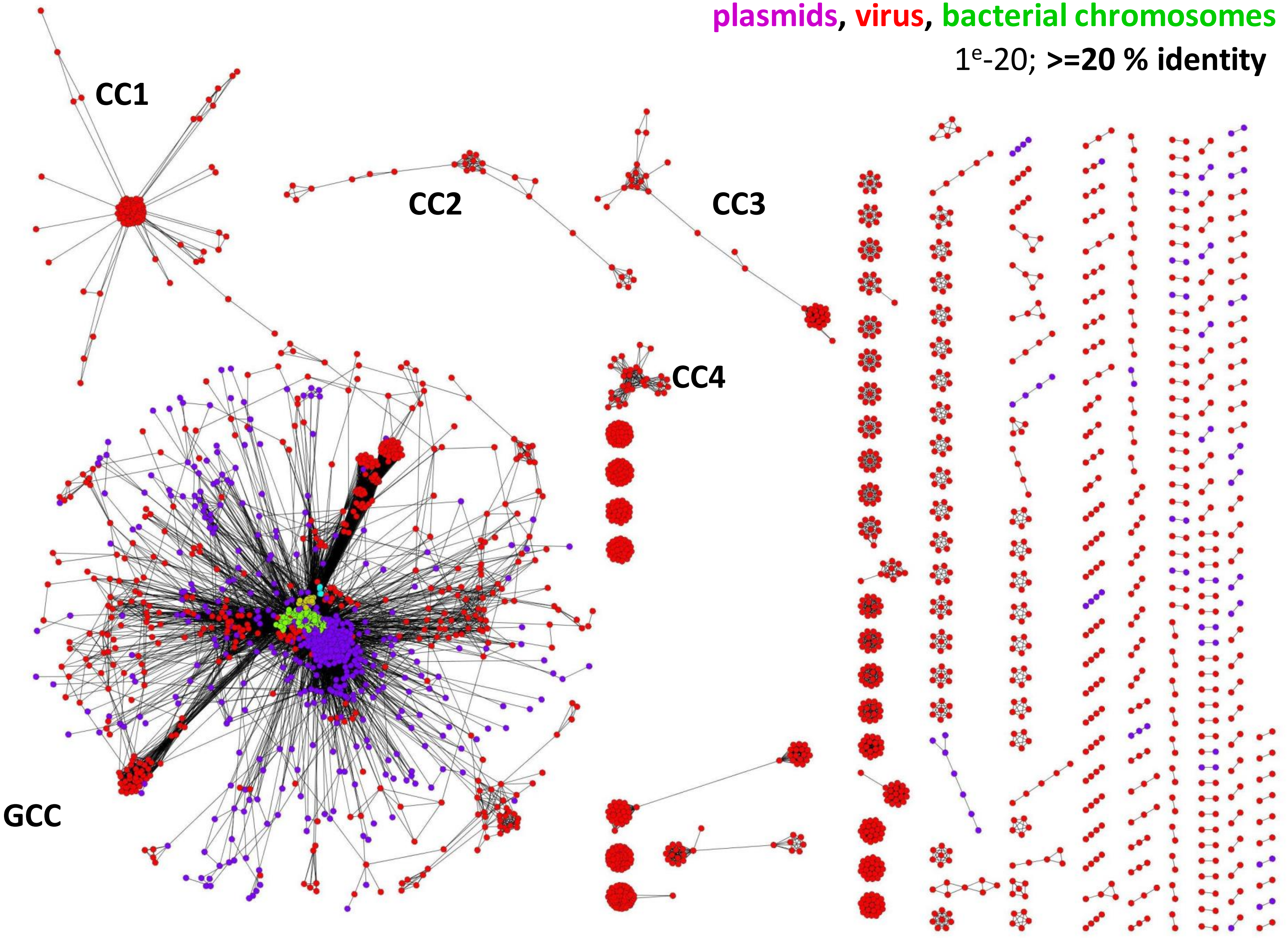
Paul *et al.* Nat Micro 2017

The DGR mechanism



*In front of an 'A', the enzymes use whatever nucleotide is at hand

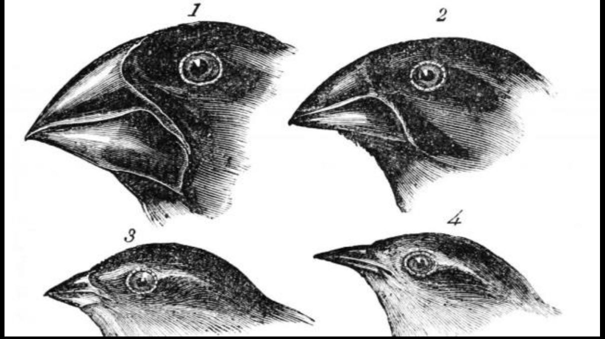
In gene-sharing networks, genetic worlds suggest isolated transmission groups among microbes



Hybrids are turning up everywhere, questioning the definition of species

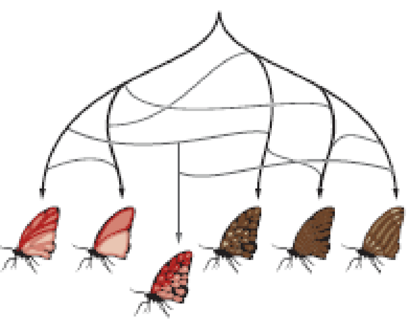
Pennisi, science 2016

4-10 % of the plants



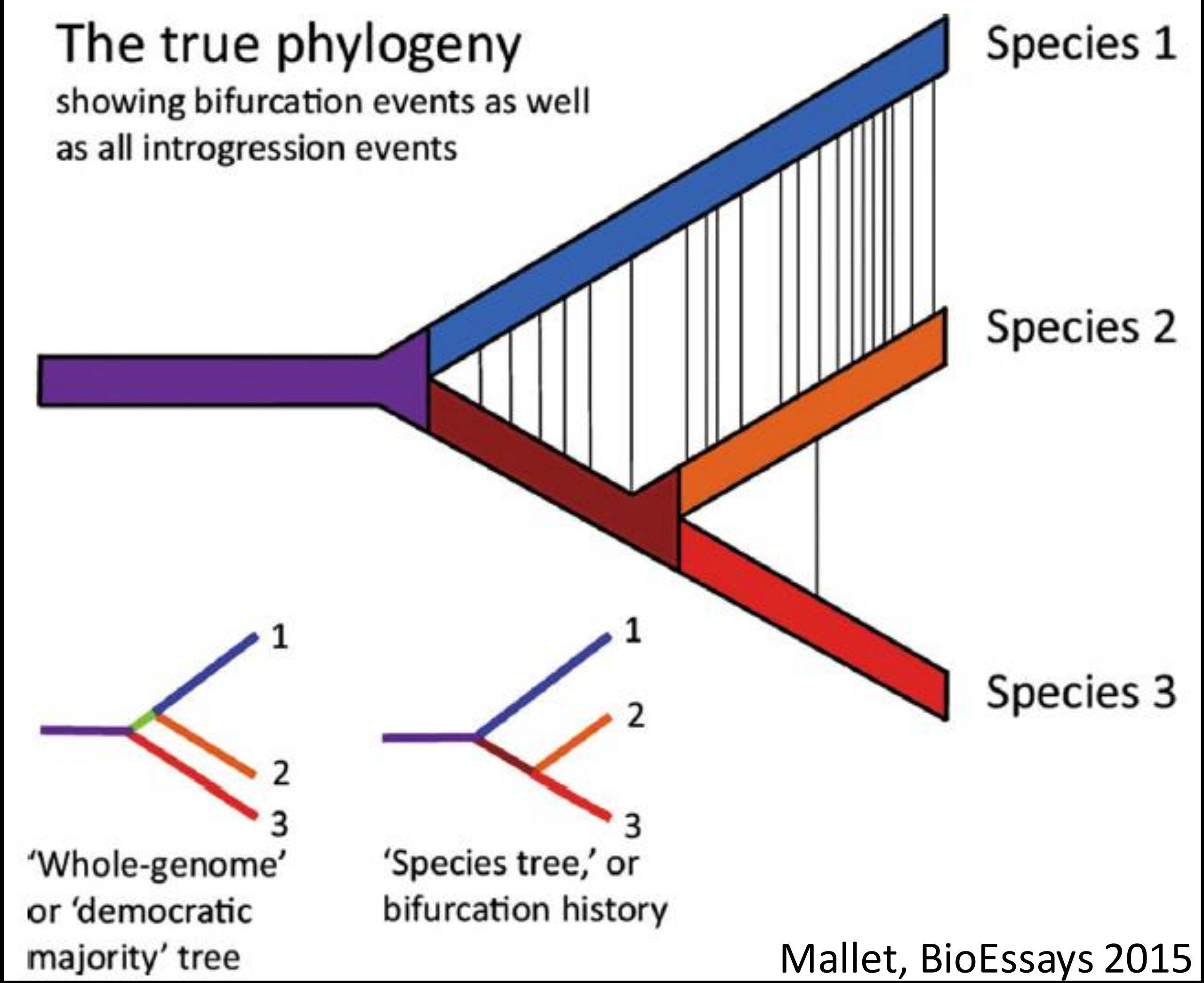
10 % of the birds

Grolar (Grizzly/Polar bear)



>30 % of the *Heliconius*

Transfer of Insecticide resistance

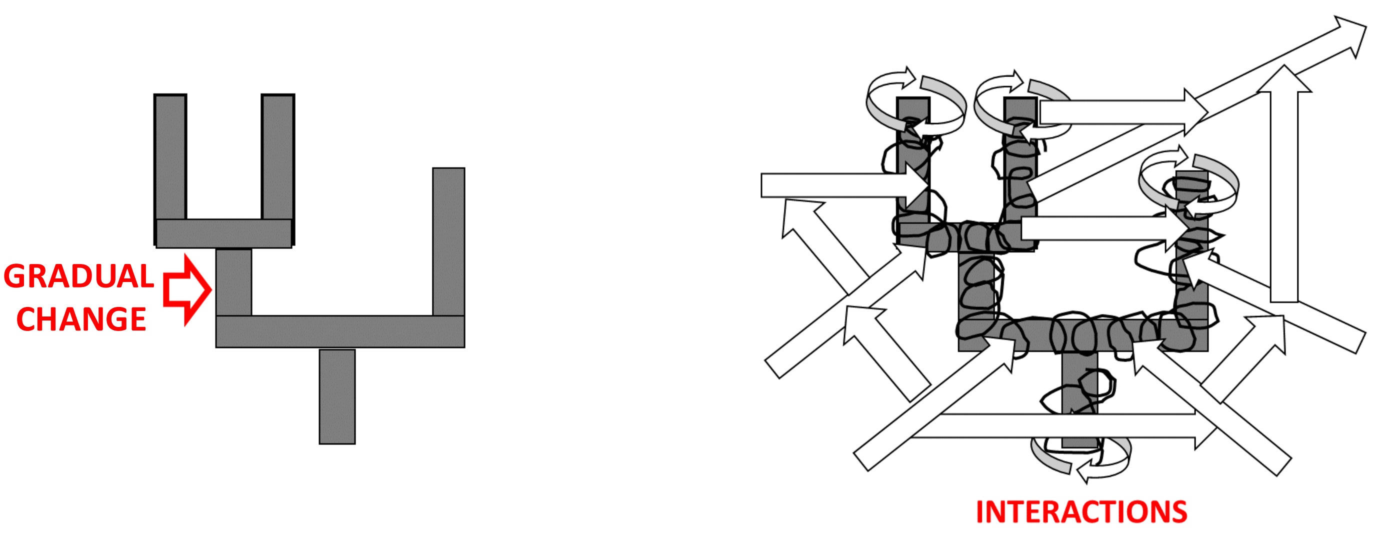


Mallet, BioEssays 2015

« Hybridization and introgression may occur among non sister-species as well as between sister-species, especially during adaptive radations. […] The origins of traits, and the genes that determine them can have very different histories from that of the species tree. »

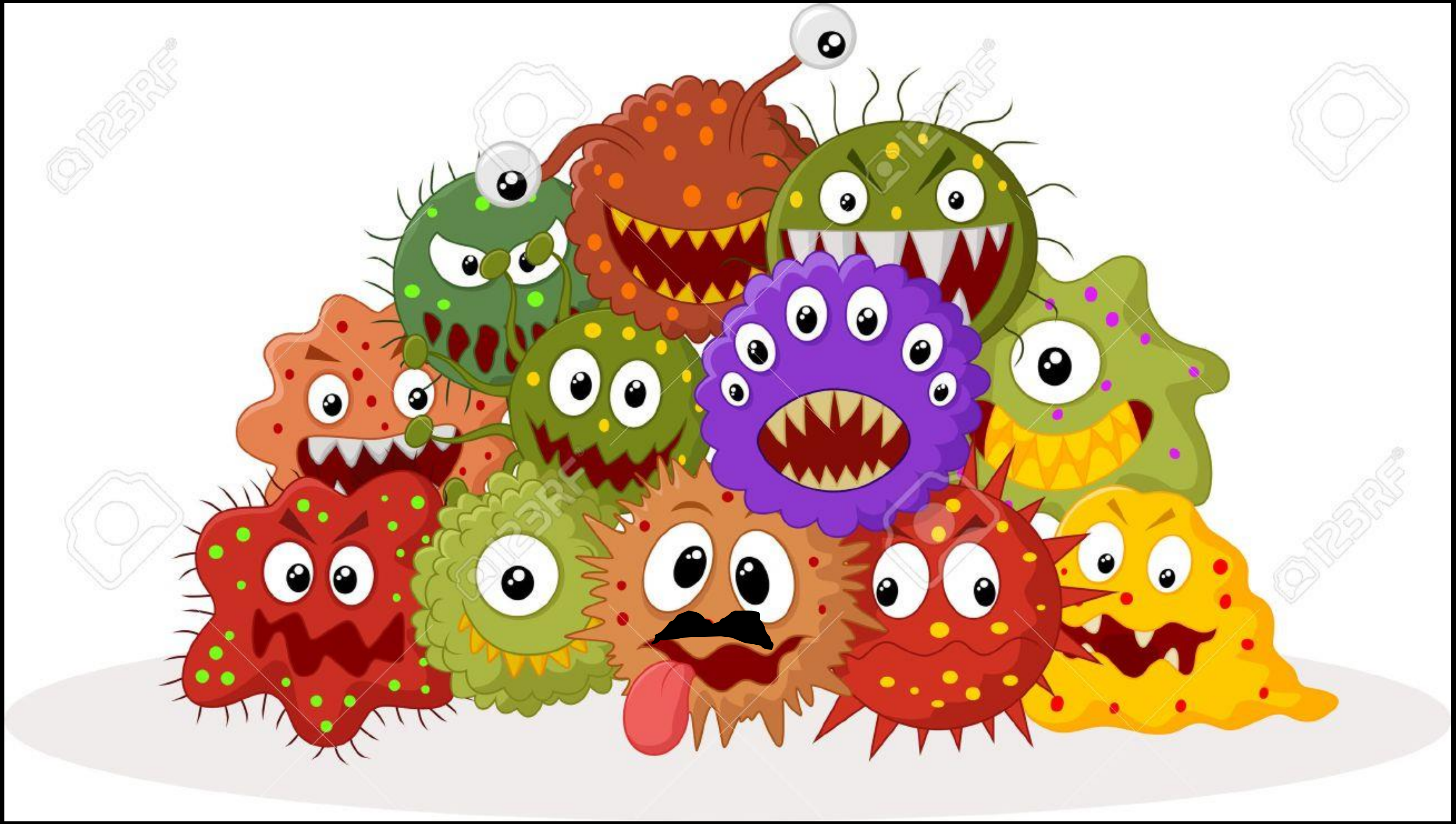
Networks could go with important theoretical transitions in evolutionary biology

- From an organismal or a gene-centered view to a multilevel view
- From a focus on vertical inheritance to an interest for both vertical and introgressive descent
- From a phylogenetic perspective to a comprehensive study of various biological relationships



- From a biotic only framework only to a biotic + abiotic framework
- From a taxonomic-based view to a processual perspective
- From a set of core founder disciplines to a broader set of disciplines, including novel hybrid research fields

- On the origins of traits: ultra-small cells with ultra-small genomes



Known bacterial & archaeal lineages

New bacterial & archaeal lineages from the environment