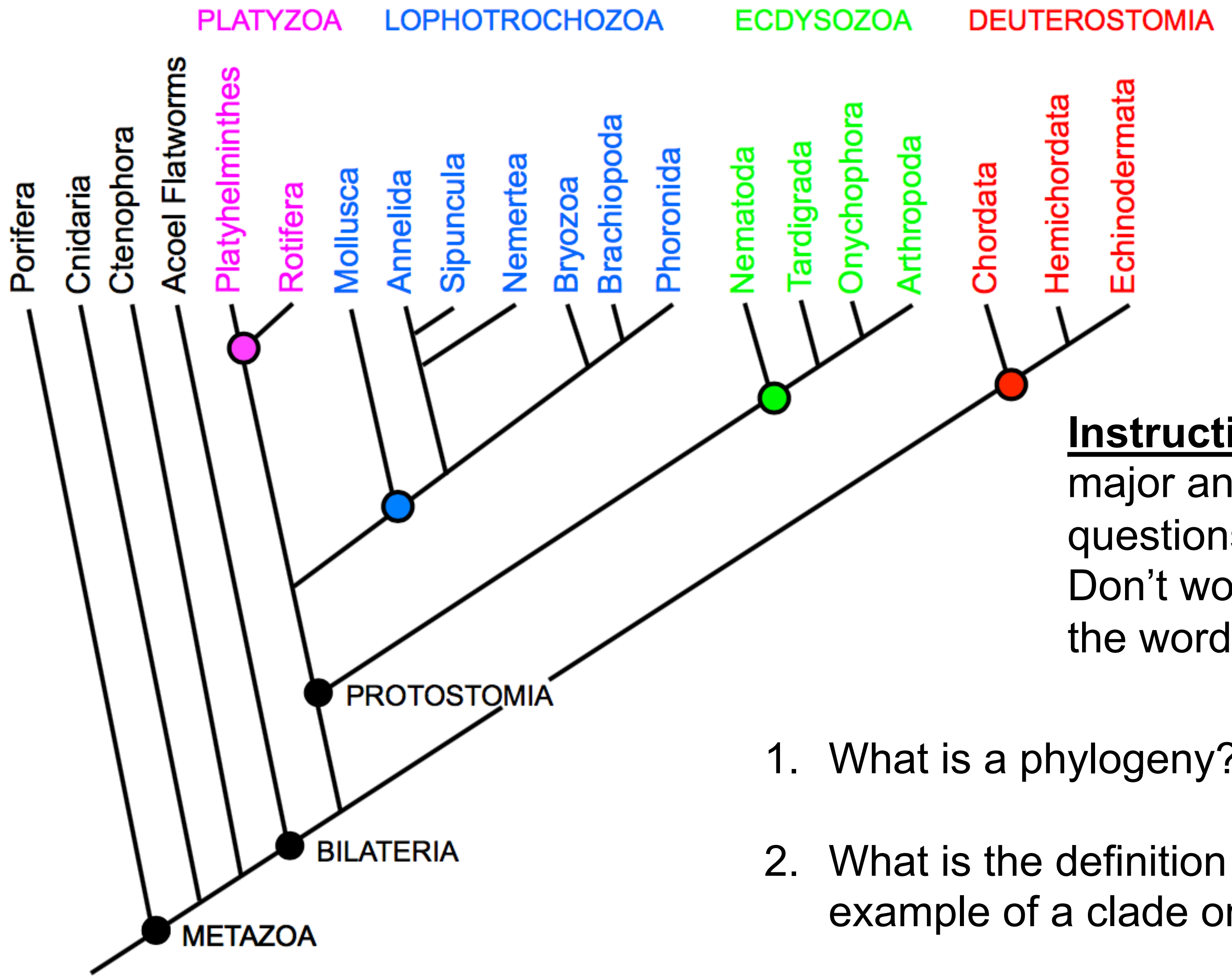
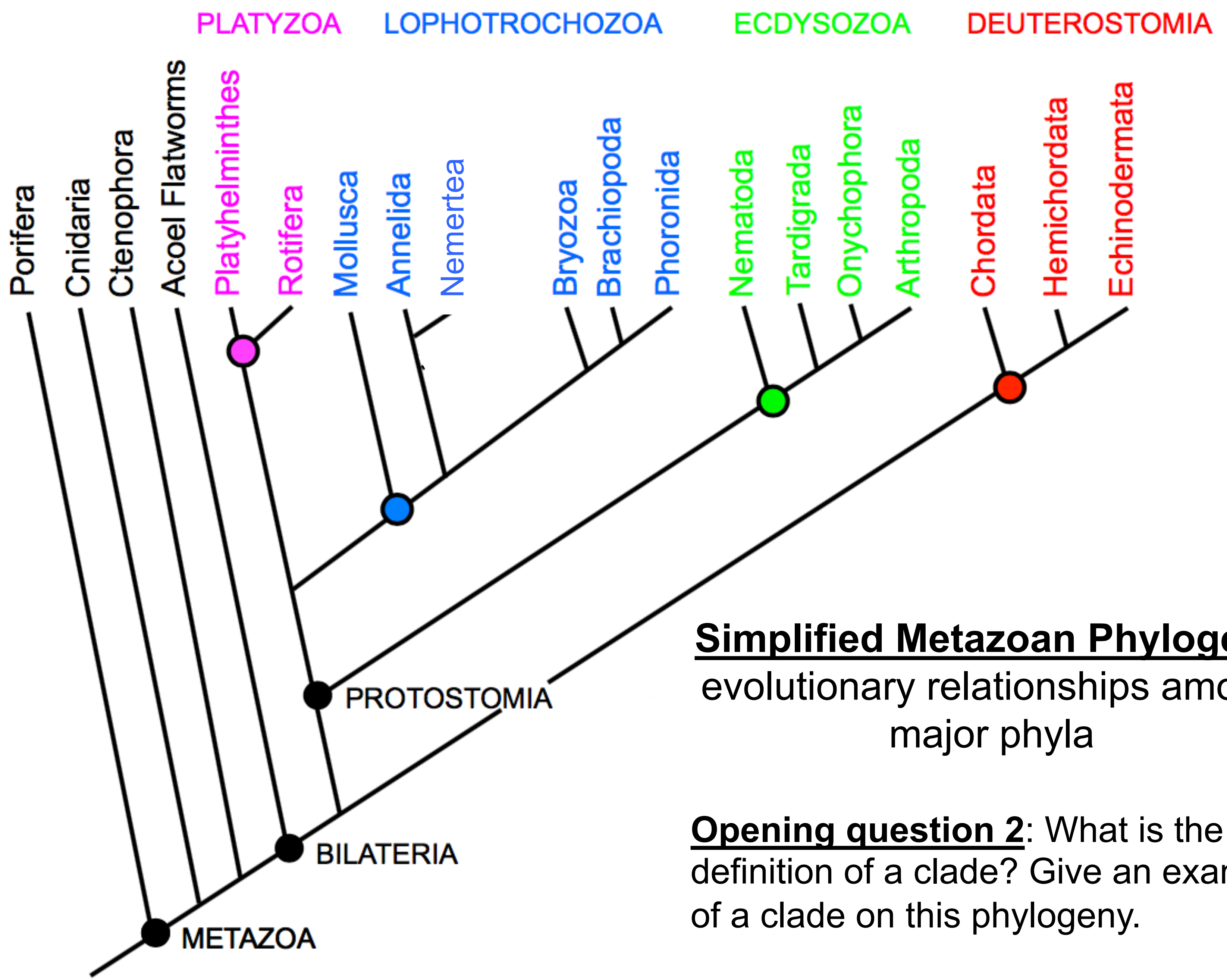
# Opening Thought Questions – 2/6




**Instructions**: This is a phylogeny of the major animal phyla. Think about these questions and write a brief response. Don't worry if you don't understand all of the words or names on this slide!

1. What is a phylogeny?

2. What is the definition of a clade? Give an example of a clade on this phylogeny.



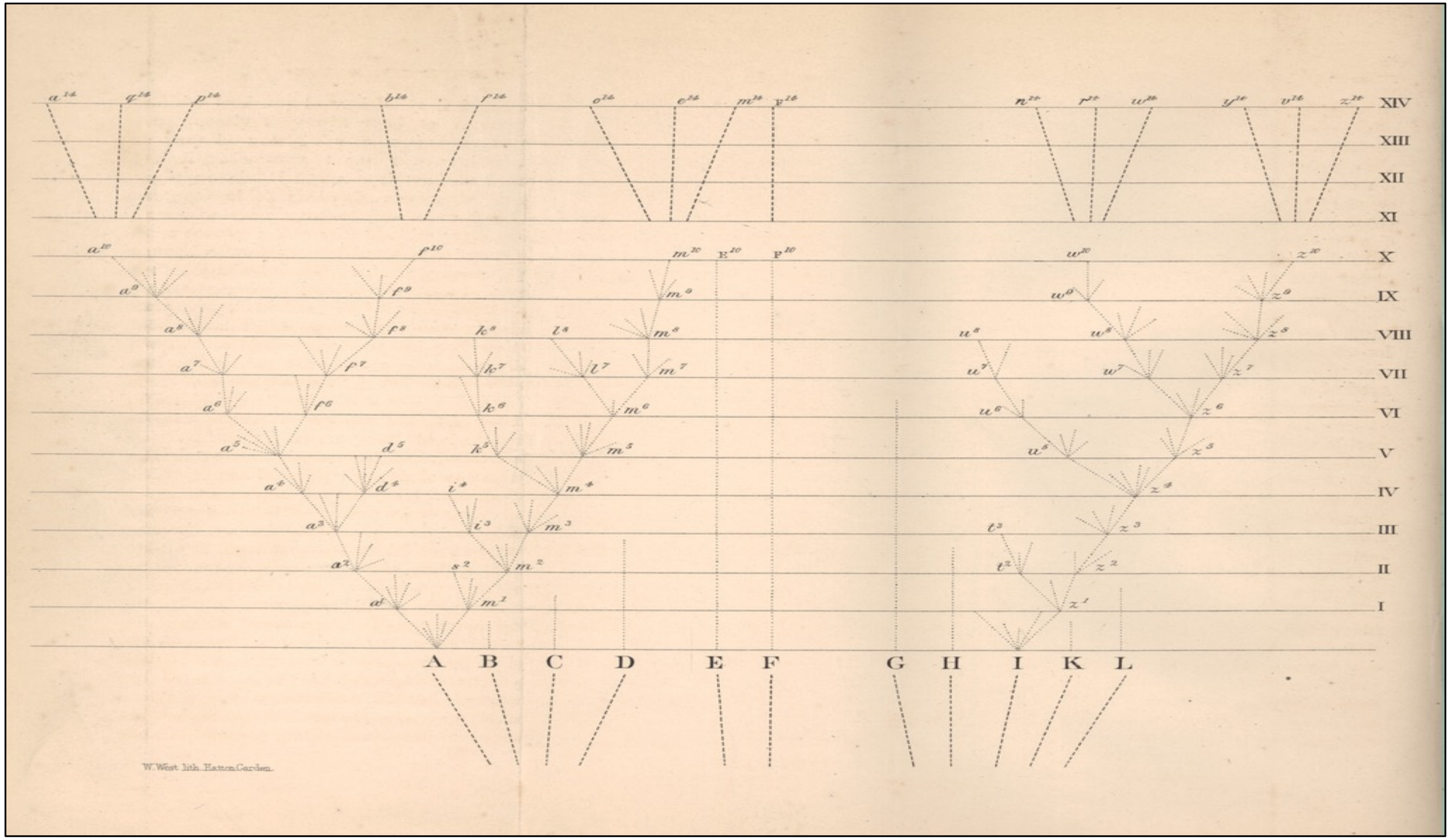

## <u>**Simplified Metazoan Phylogeny**</u>

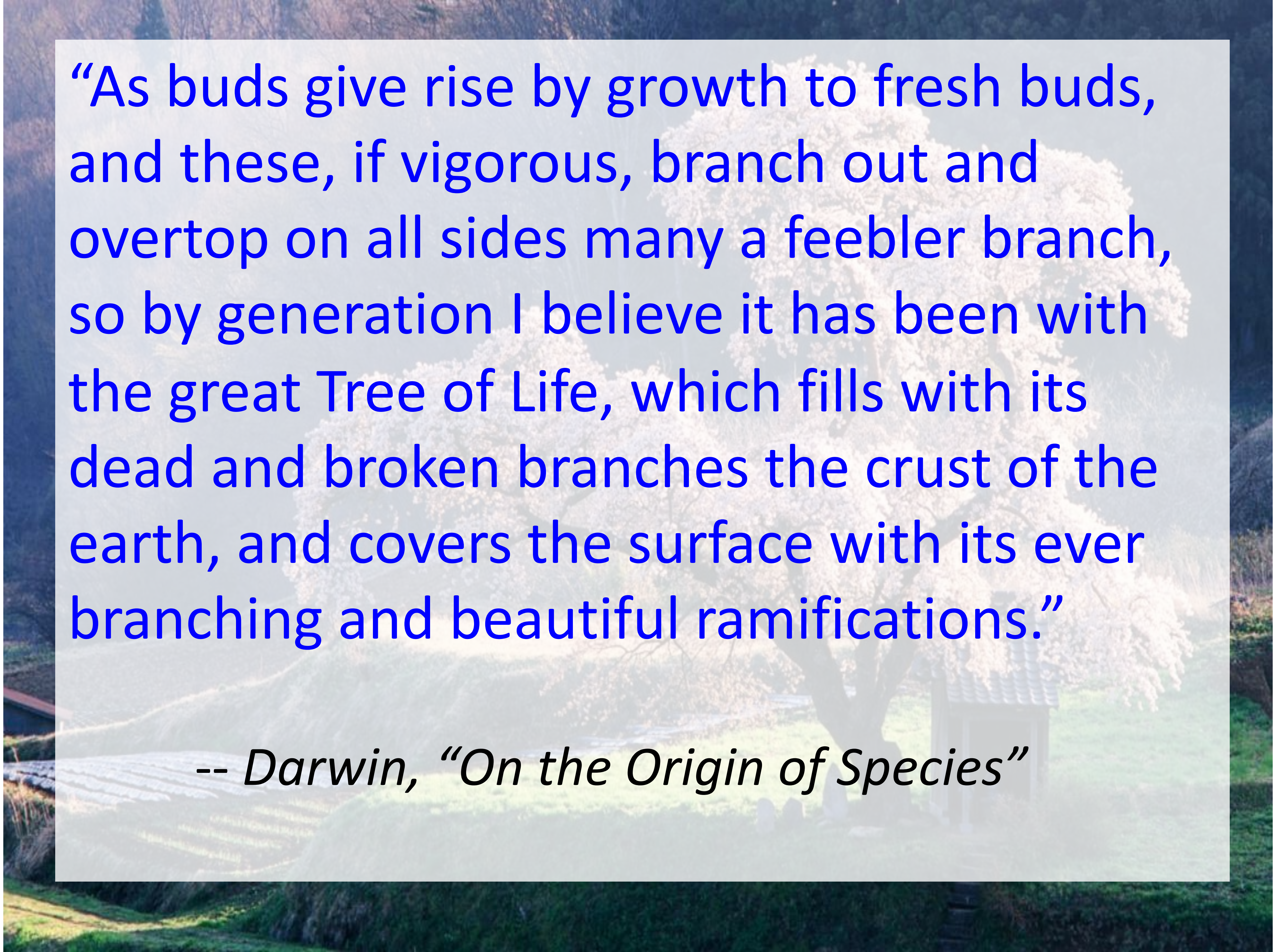
evolutionary relationships among major phyla

<u>**Opening question 2**</u>: What is the definition of a clade? Give an example of a clade on this phylogeny.

# Charles Darwin: 1859

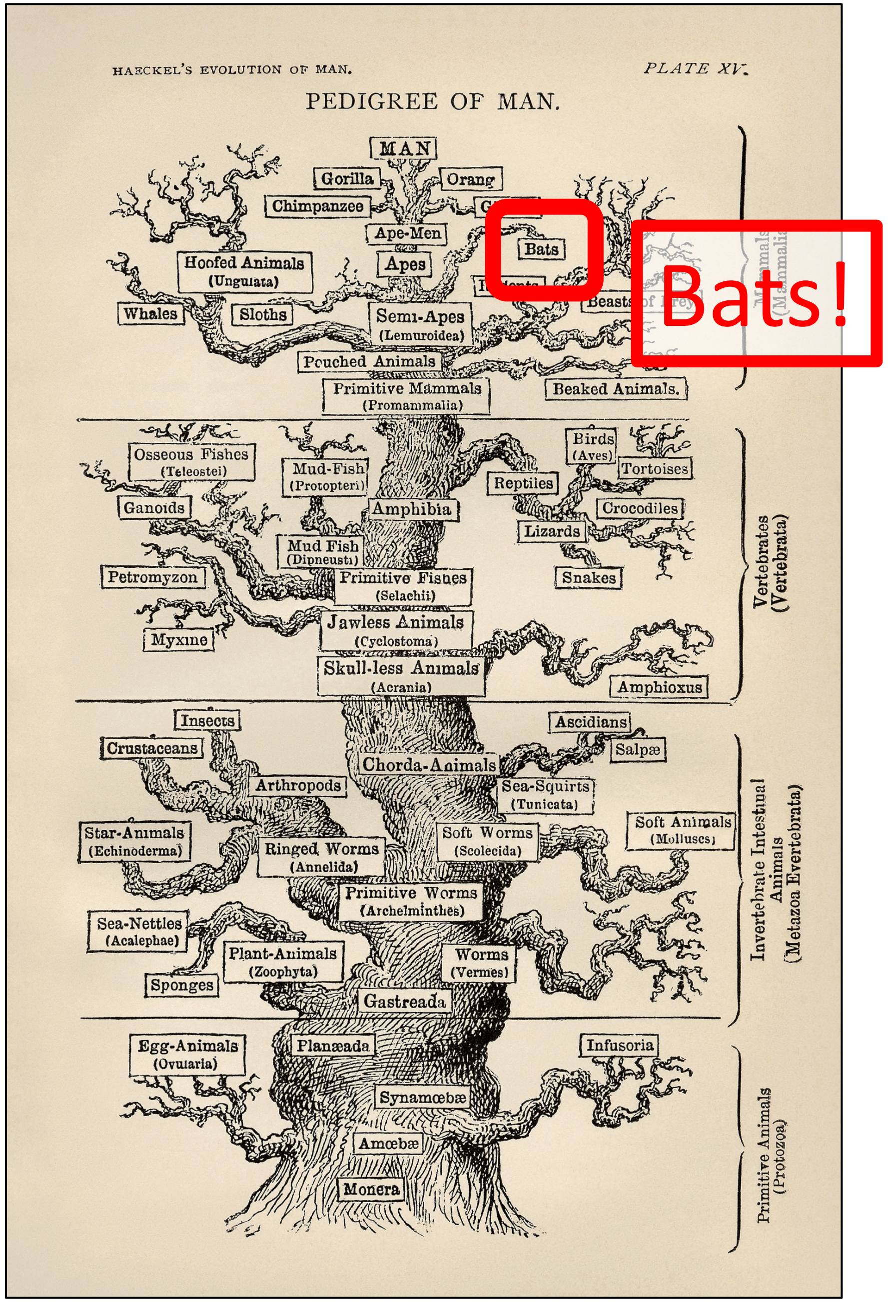
"On the Origin of Species"

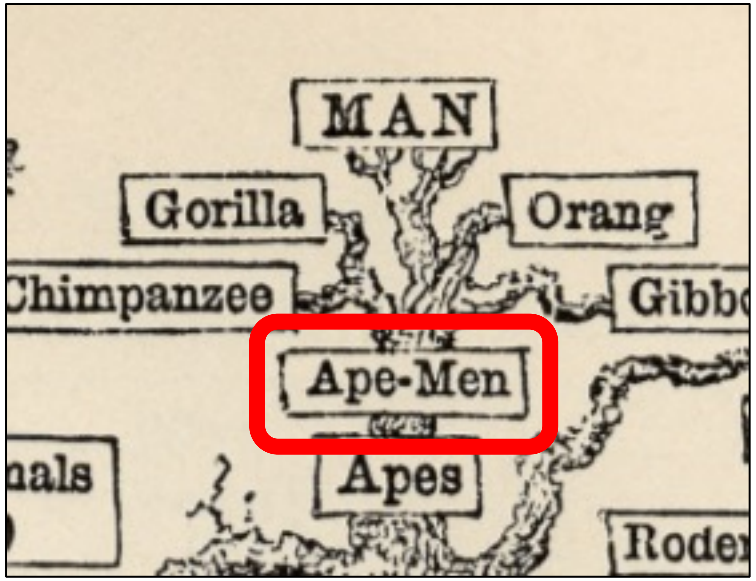



"As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications."
*-- Darwin, "On the Origin of Species"*

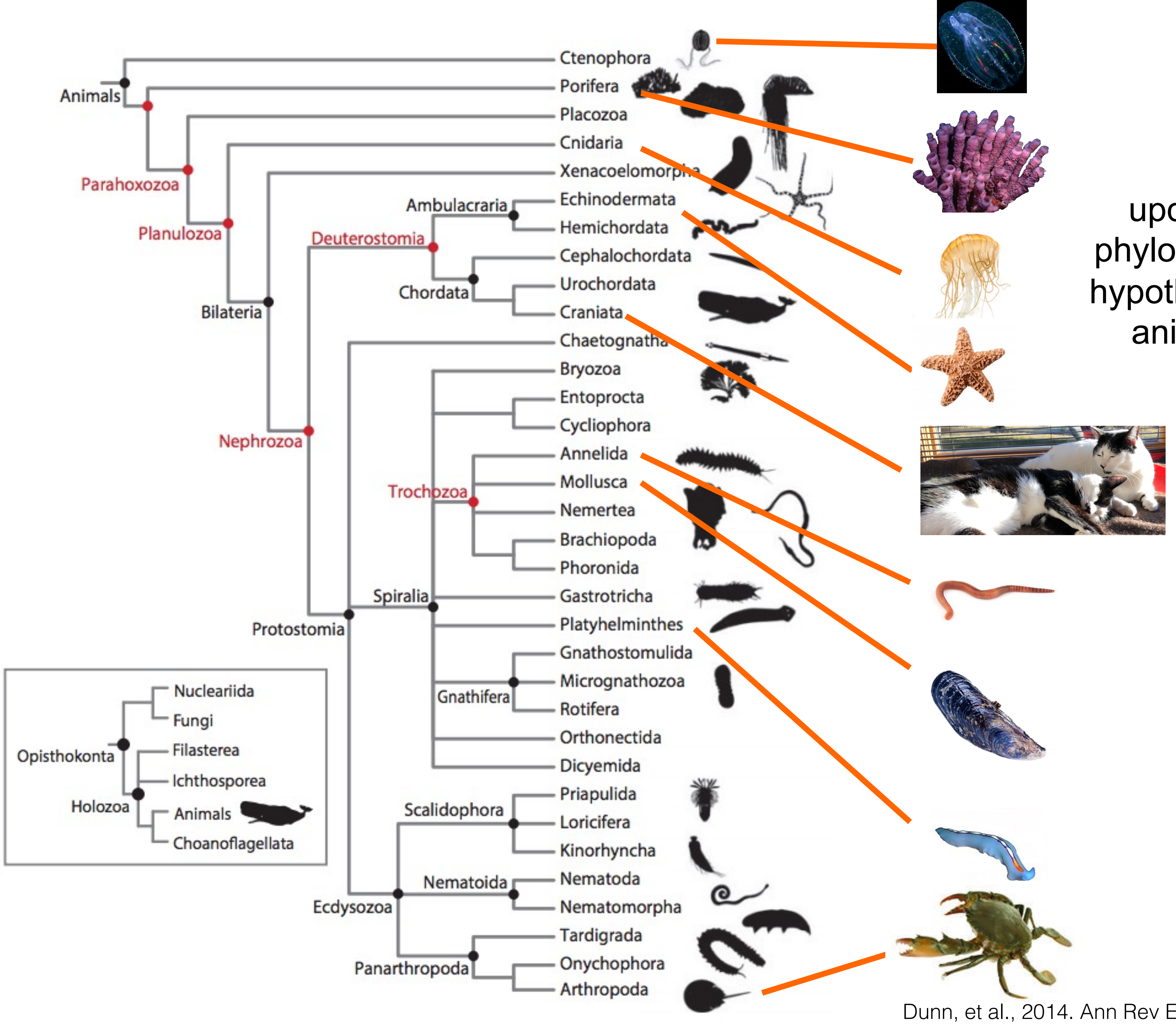
# Ernst Haeckel 1879

- “The Evolution of Man”



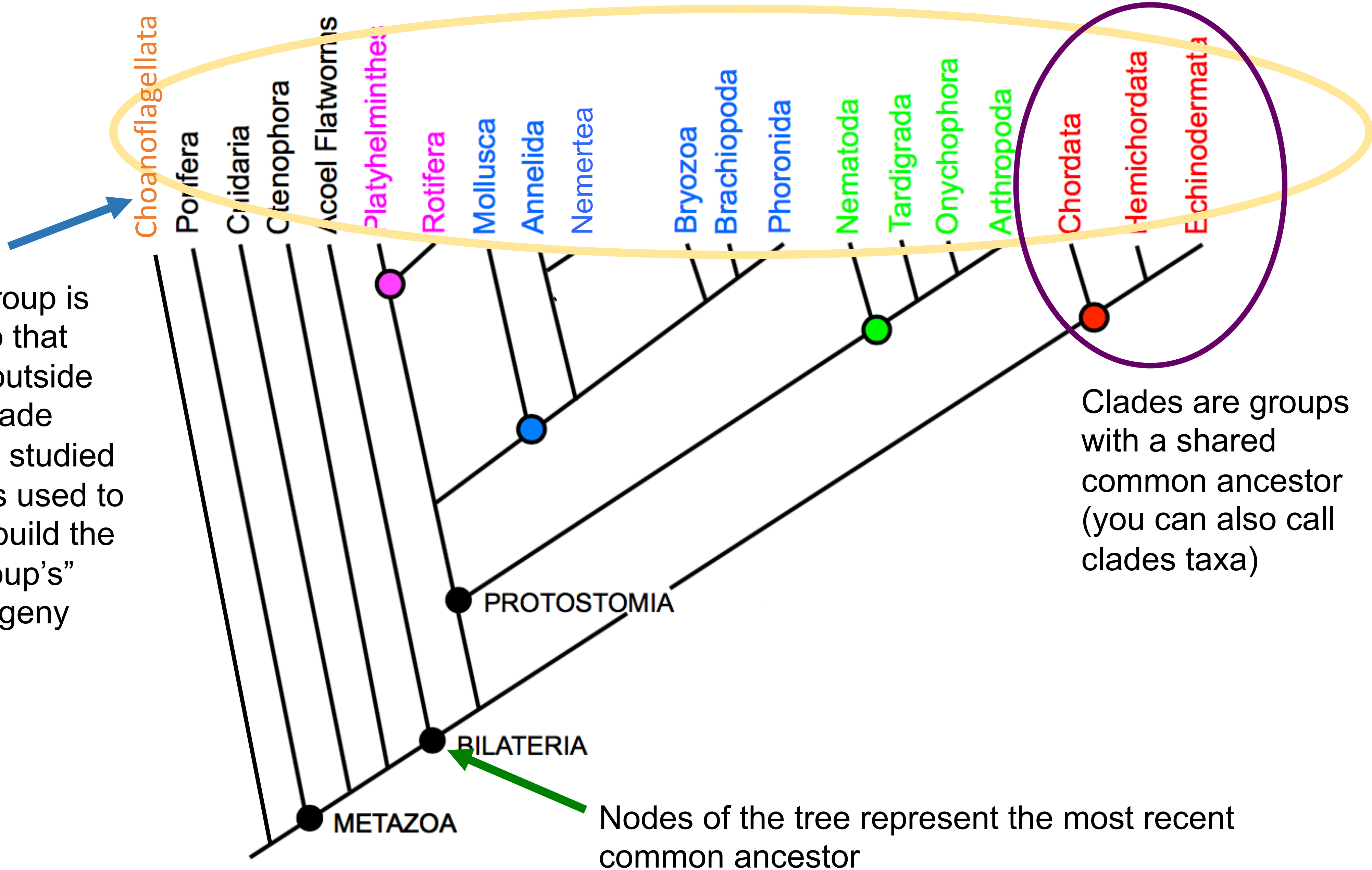
# Phylogenetic Trees




updated phylogenetic hypothesis of animals

Dunn, et al., 2014. Ann Rev Evol Ecol System.

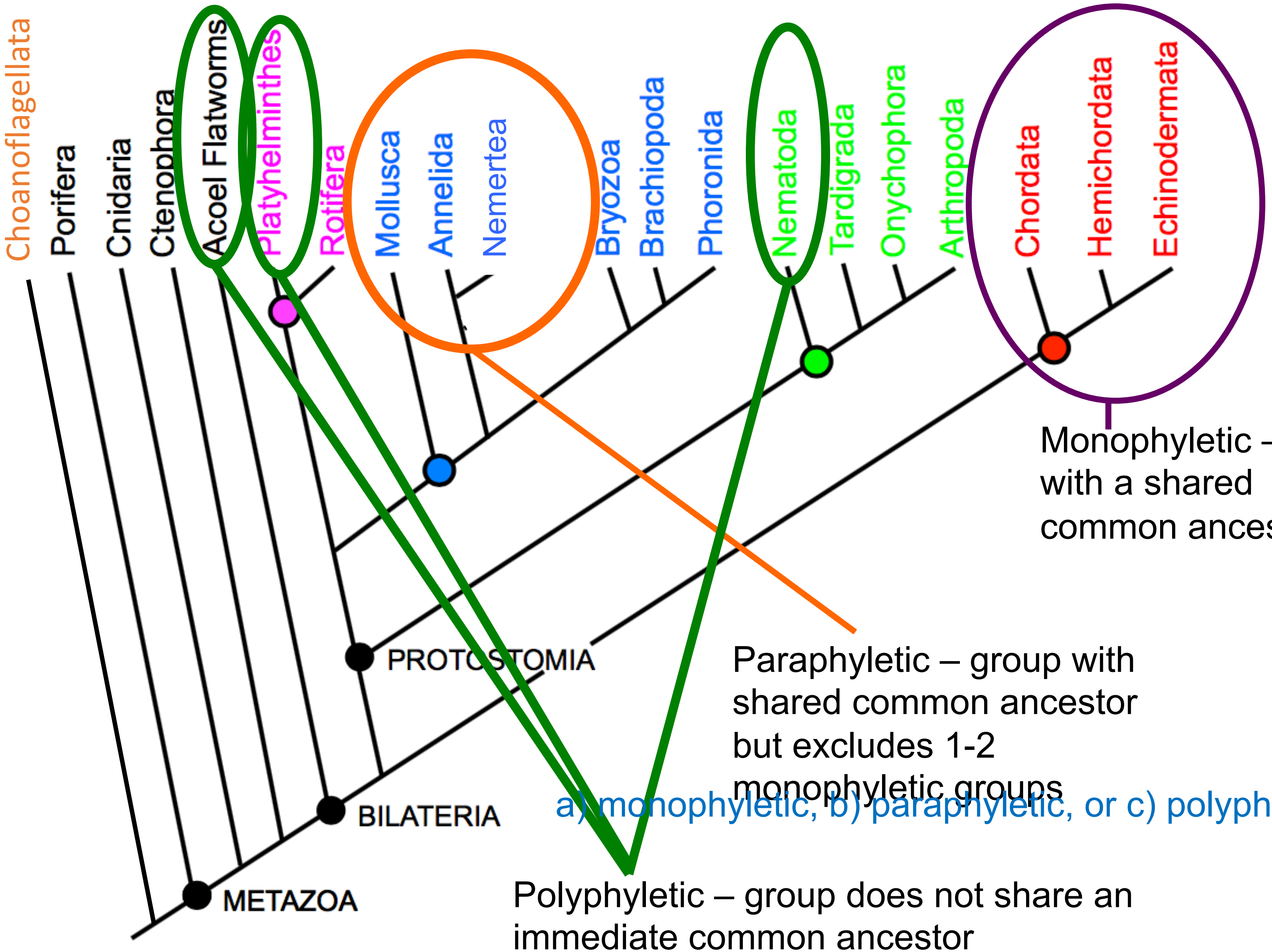
Tips of the tree are taxonomic groups (operational taxonomic units – OTU's)




Outgroup is group that falls outside the clade being studied and is used to help build the "ingroup's" phylogeny

Clades are groups with a shared common ancestor (you can also call clades taxa)

Nodes of the tree represent the most recent common ancestor



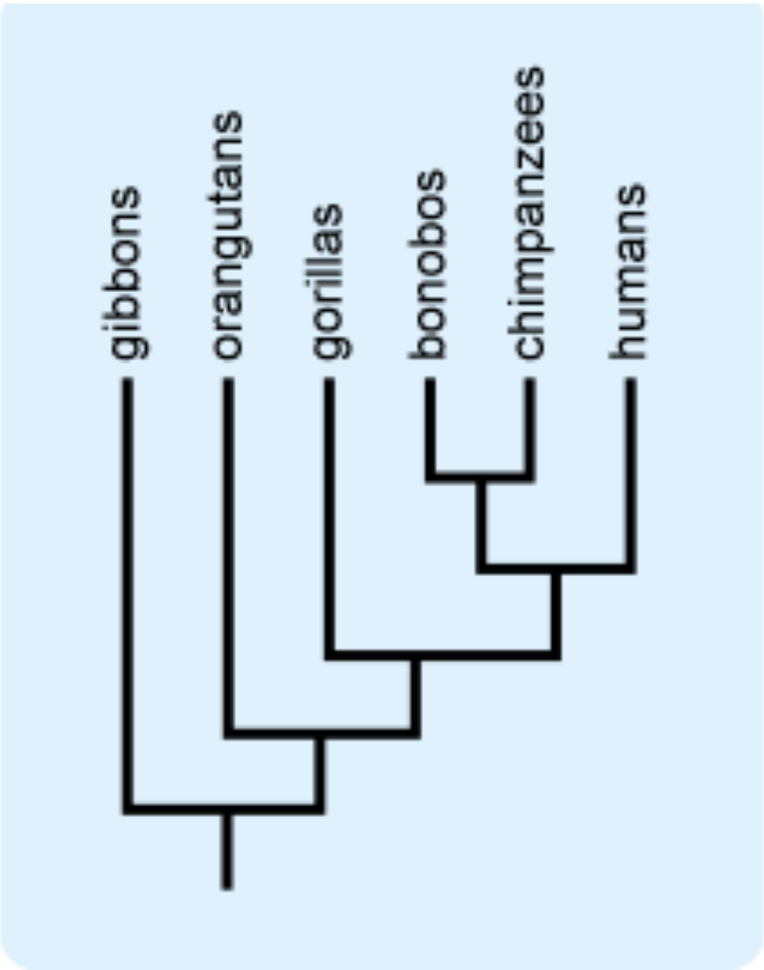

Monophyletic – group with a shared common ancestor

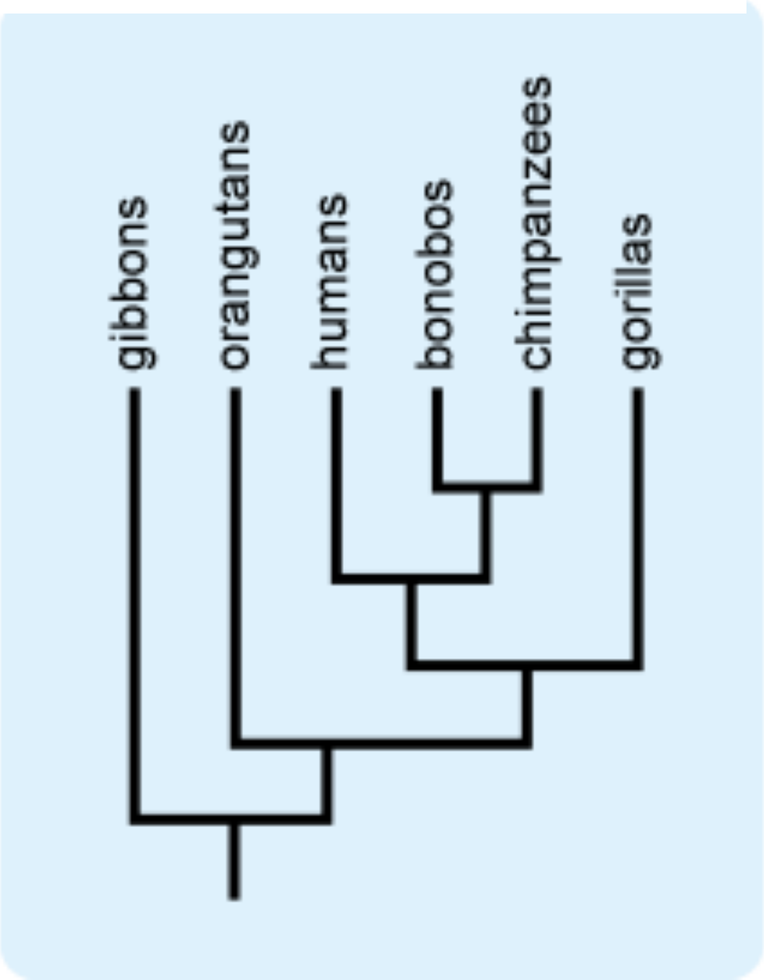
Paraphyletic – group with shared common ancestor but excludes 1-2 monophyletic groups

a) monophyletic, b) paraphyletic, or c) polyphyletic?

Polyphyletic – group does not share an immediate common ancestor

# Flipping nodes: these trees are the same



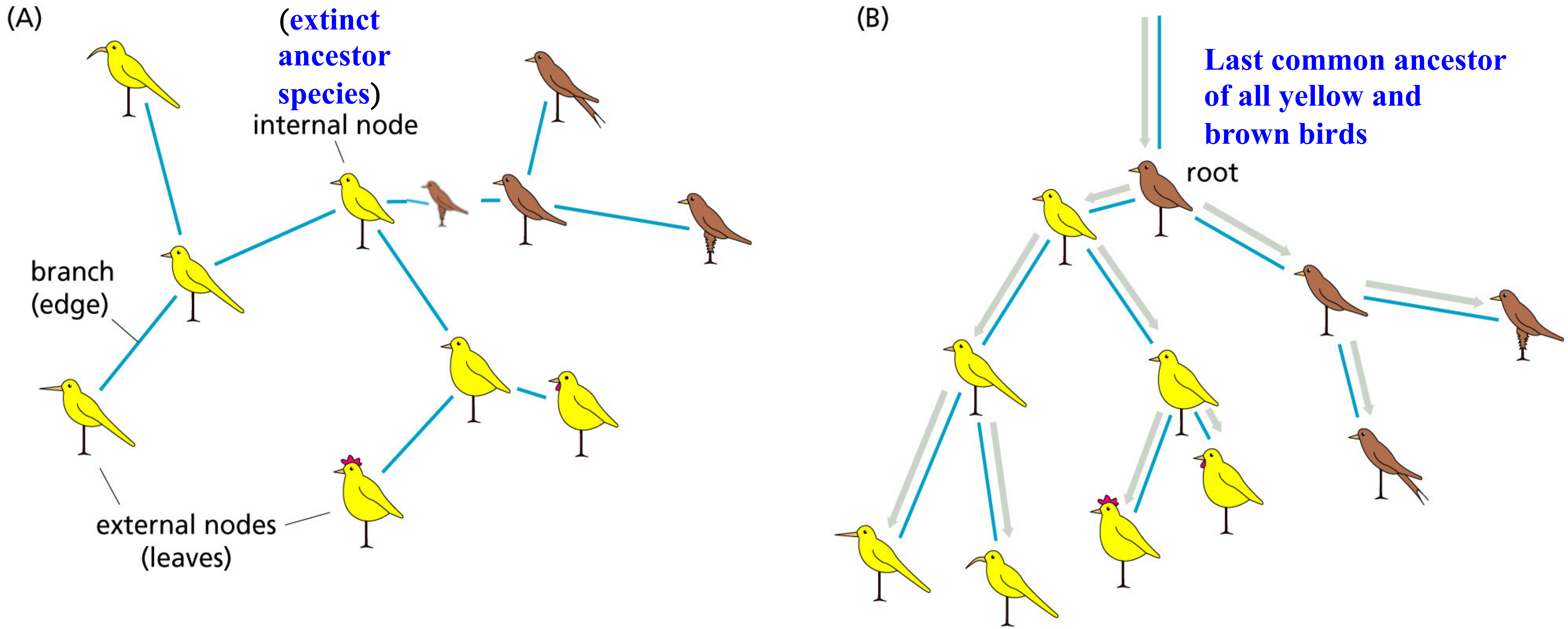




Understanding Evolution: evolution.berkeley.edu

# Terminology

- OTUs or leaf nodes – the tree's input sequences
- Clade/taxon (plural = taxa) – an internal node, plus everything that descends from that node
- Outgroup – is a group used in phylogenetic analyses and falls outside the clade being studied. All members of the group being studied are more closely related to each other than to the outgroup.

# Unrooted and Rooted Trees

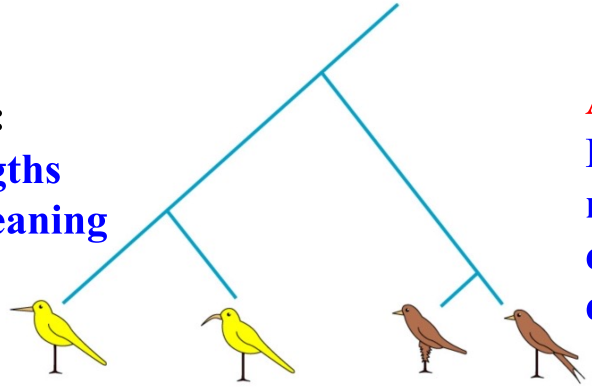



**Unrooted: cannot tell**

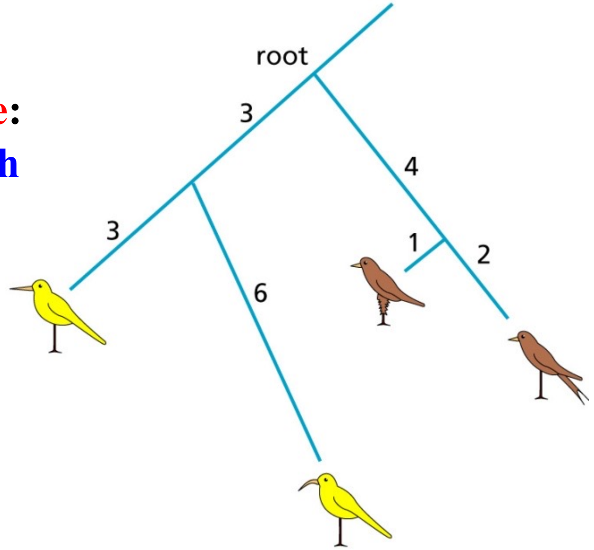
**Have yellow birds evolved from brown birds, or brown birds from yellow birds?**

**Rooted: can tell**

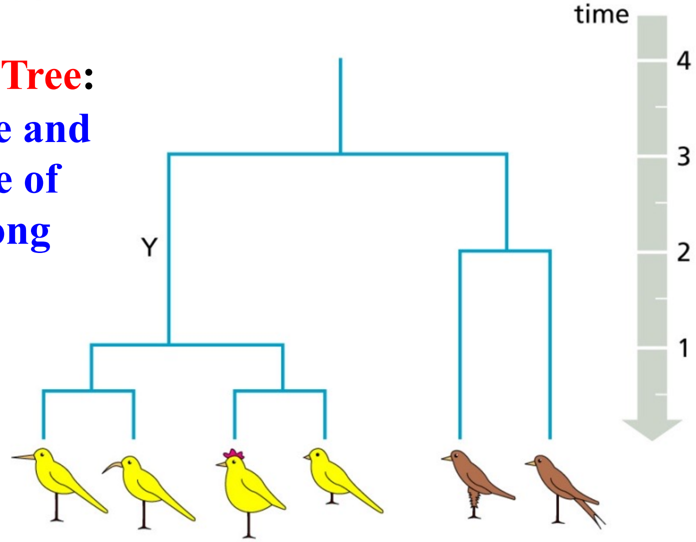
**Cladogram**:
**Branch lengths carry no meaning**



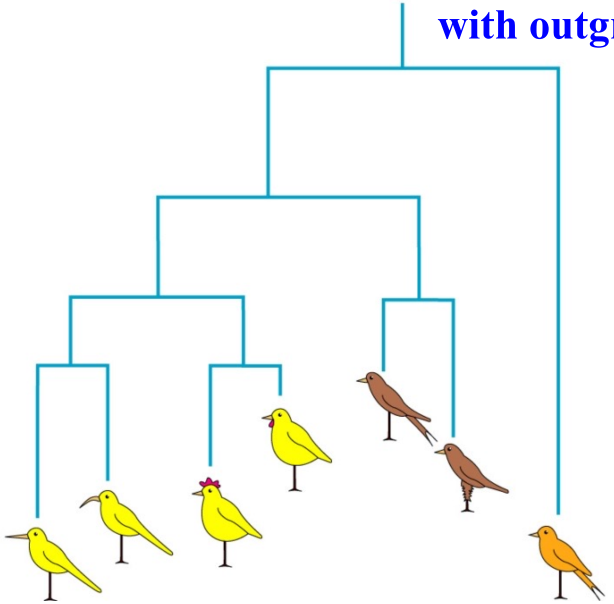
**Additive Tree**:
**Branch length measures evolutionary divergence**



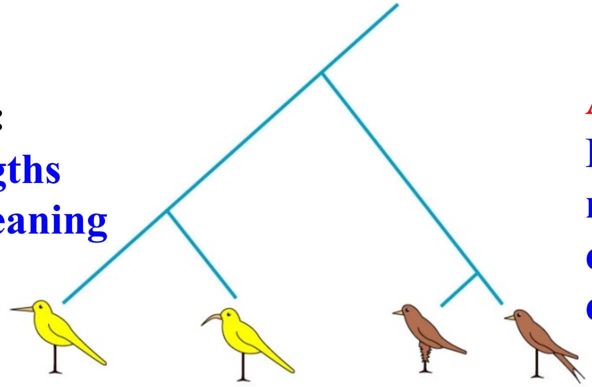

**Ultrametric Tree**:
**Additive tree and constant rate of mutation along branches**



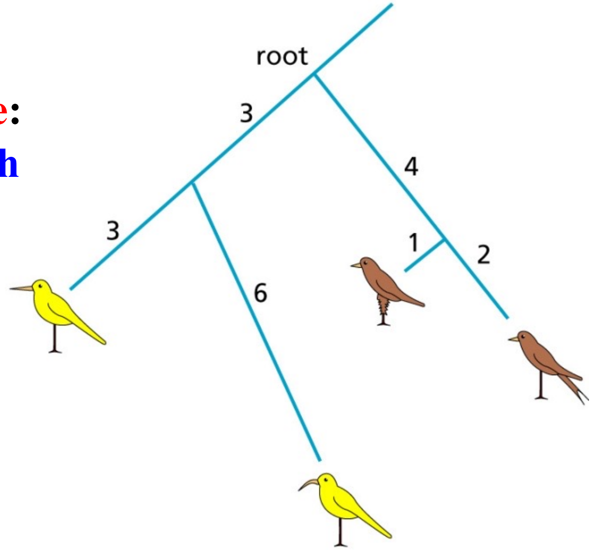

**Additive Tree**:
**with outgroup**



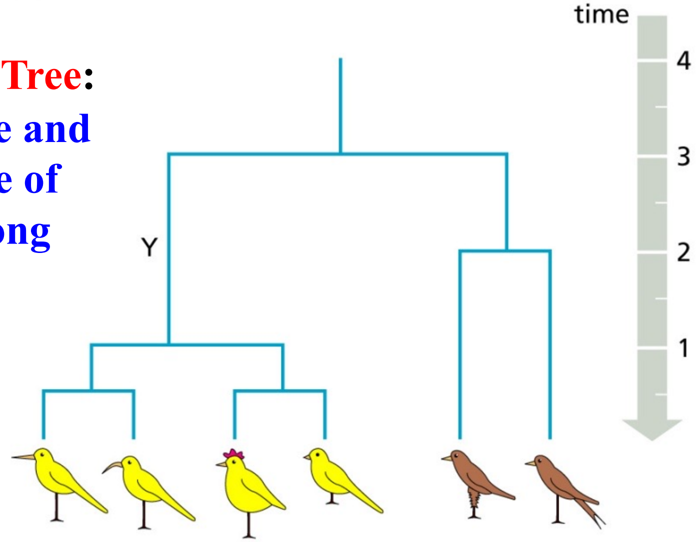
**Cladogram**: **Branch lengths carry no meaning**



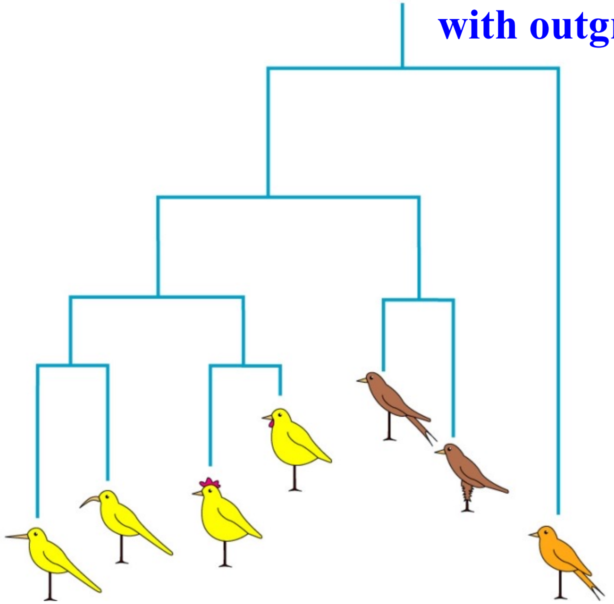
**Additive Tree**: **Branch length measures evolutionary divergence**



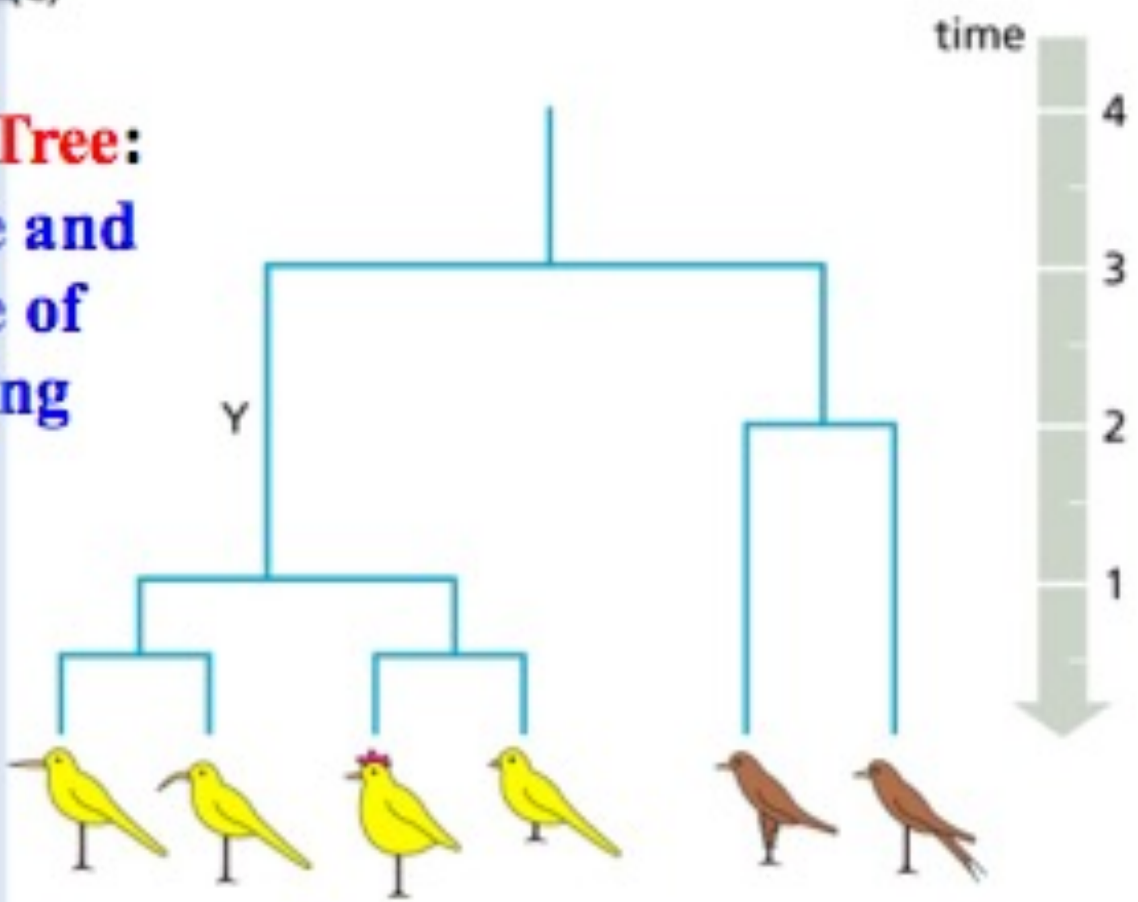

**Ultrametric Tree**: **Additive tree and constant rate of mutation along branches**




**Additive Tree**: **with outgroup**



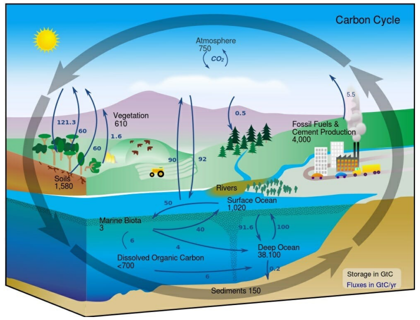
**Ultrametric Tree:**
**Additive tree and constant rate of mutation along branches**

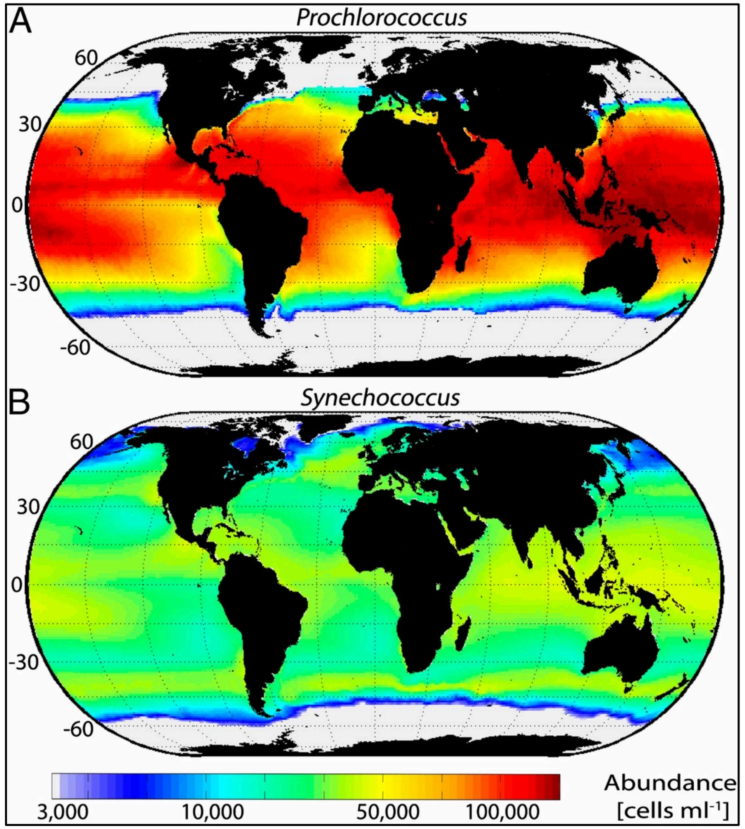


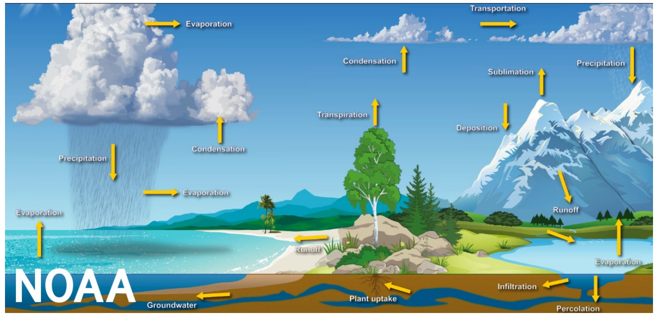

The *Ultrametric property:*
Constant distance from any OTU to root

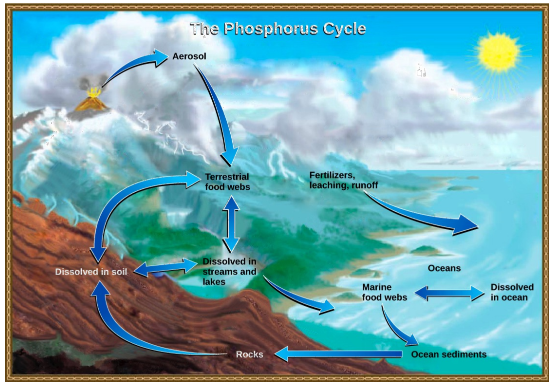
# Ecology studies abundance and distribution of organisms

- It does this by studying movement and transformation of resources



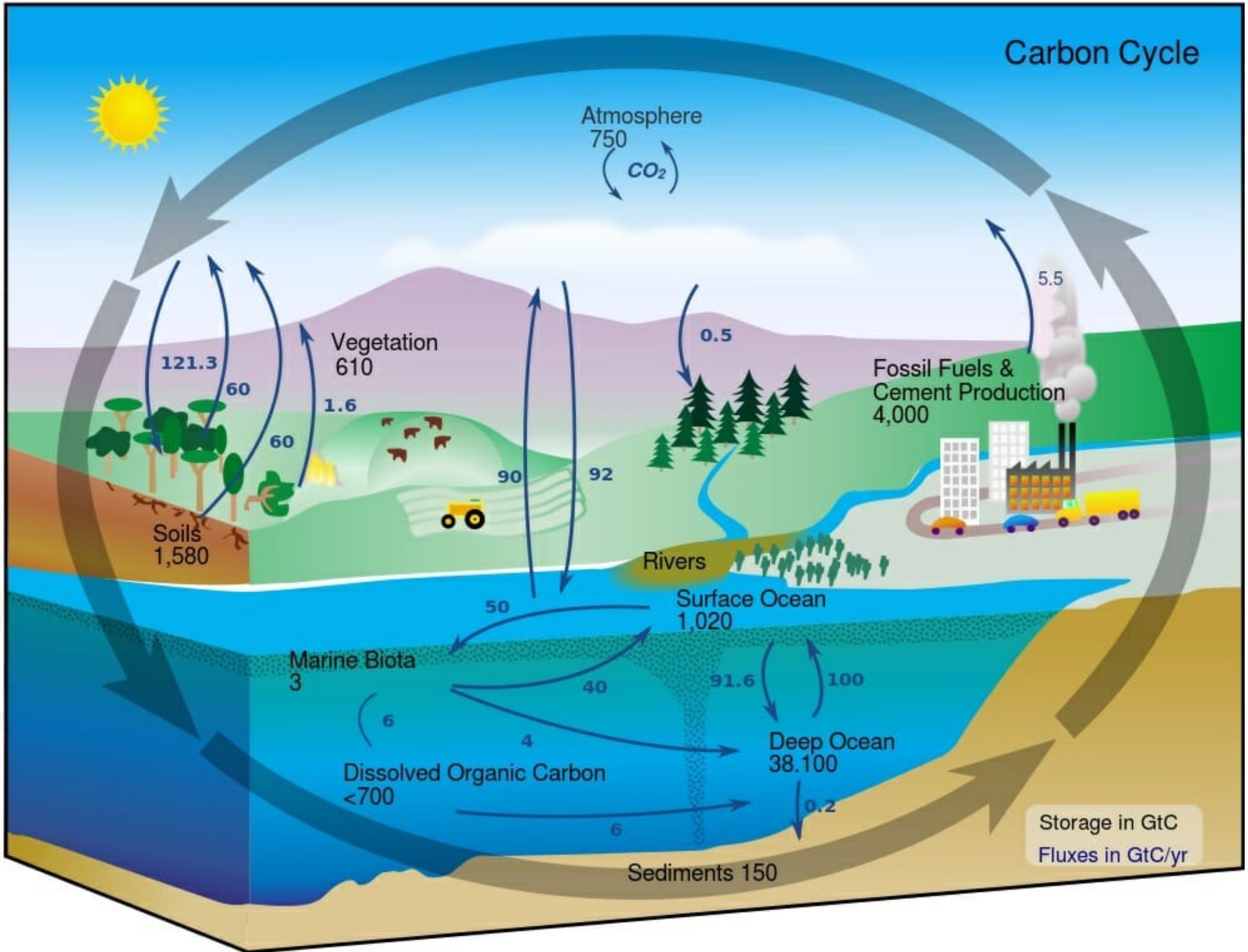



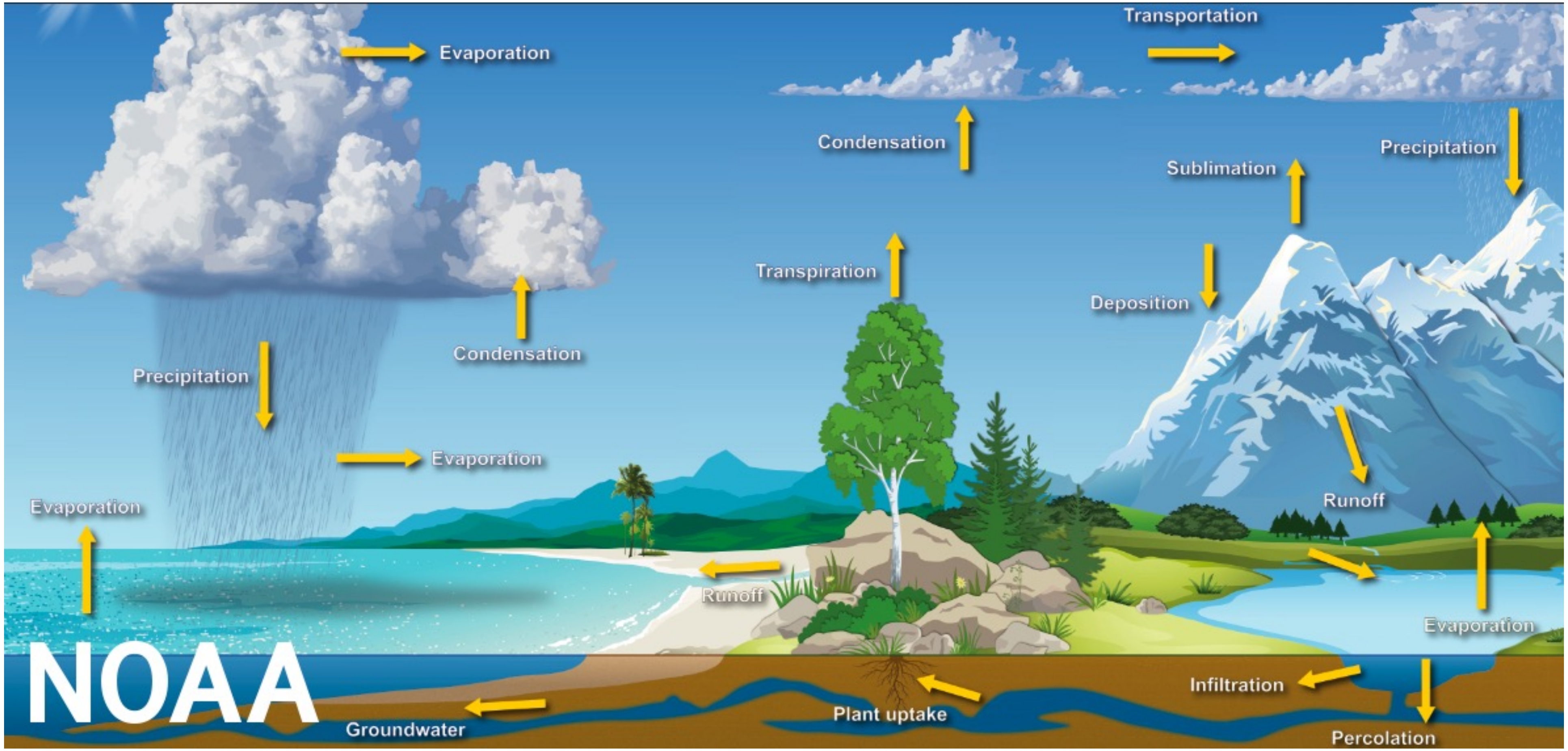
# Ecology studies movement and transformation of resources

- E.g. Carbon
  - $CO_2$, biomass, dissolved organic carbon (DOC), particulate organic carbon (POC) …
- E.g. Water
  - Oceans, lakes, rivers, clouds, sea ice …
- E.g. Phosphorus
  - Fertilizer, sediments, minerals, bones …
- E.g. Nitrogen
  - $N_2(g)$, ammonia, nitrate, nitrate …


Carbon Cycle
Atmosphere
750
$CO_2$
121.3
60
60
1.6
Vegetation
610
90
92
0.5
5.5
Fossil Fuels &
Cement Production
4,000
Soils
1,580
Rivers
Surface Ocean
1,020
50
Marine Biota
3
40
91.6
100
6
4
Deep Ocean
38.100
Dissolved Organic Carbon
<700
6
0.2
Sediments 150
Storage in GtC
Fluxes in GtC/yr

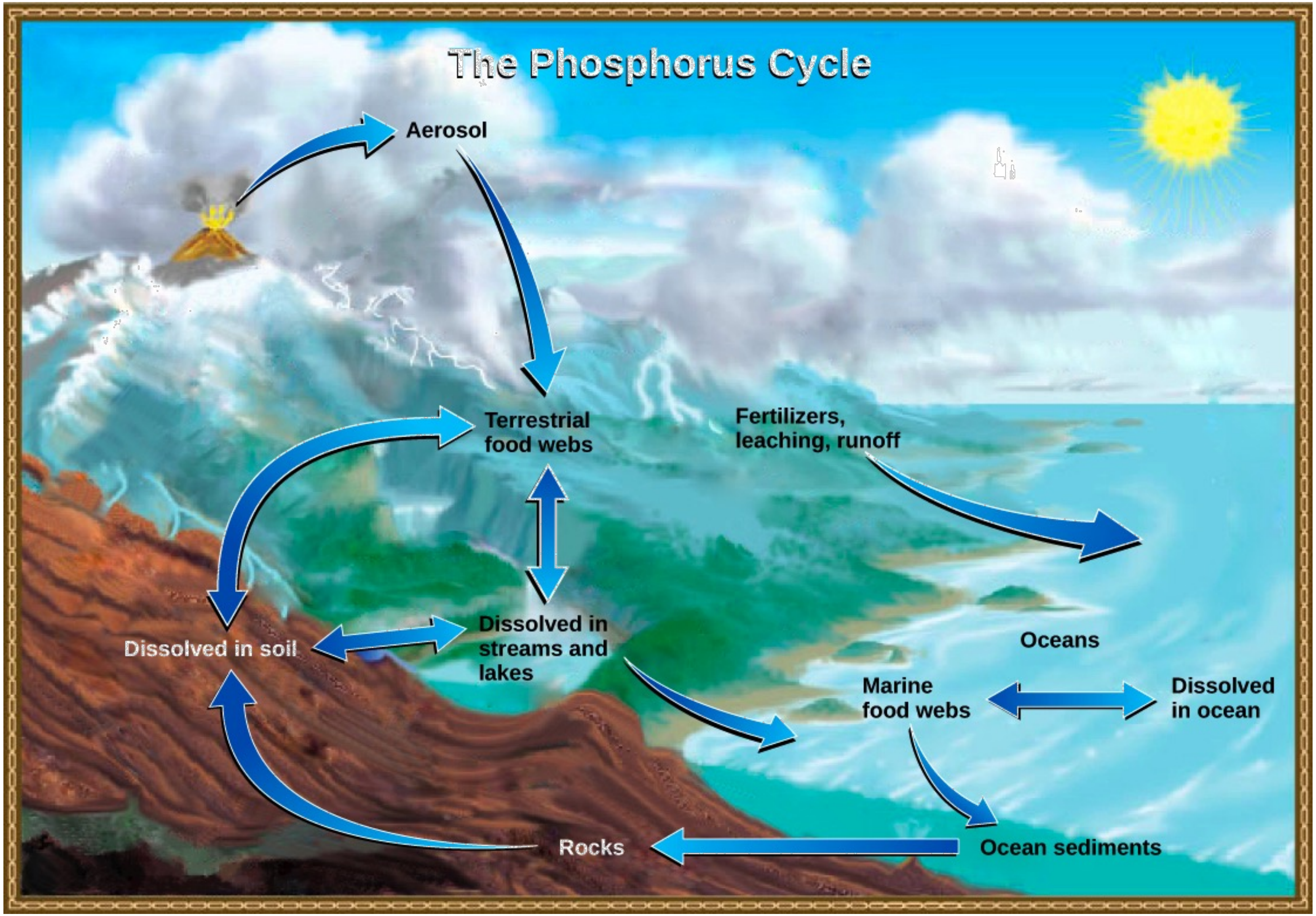
# Ecology studies movement and transformation of resources

- E.g. Carbon
  - $CO_2$, biomass, dissolved organic carbon (DOC), particulate organic carbon (POC) …
- E.g. Water
  - Oceans, lakes, rivers, clouds, sea ice …
- E.g. Phosphorus
  - Fertilizer, sediments, minerals, bones …
- E.g. Nitrogen
  - $N_2(g)$, ammonia, nitrate, nitrate …


Evaporation
Transportation
Condensation
Precipitation
Sublimation
Transpiration
Deposition
Condensation
Precipitation
Evaporation
Evaporation
Runoff
Runoff
Evaporation
NOAA
Infiltration
Plant uptake
Groundwater
Percolation

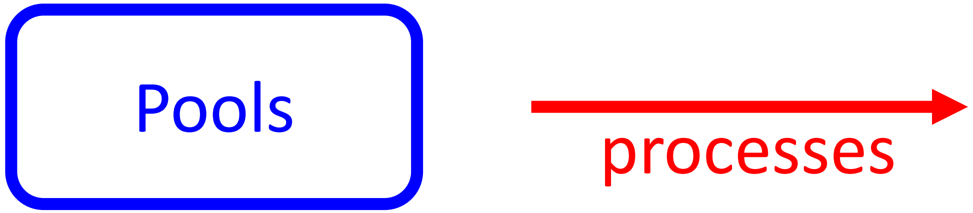
# Ecology studies movement and transformation of resources

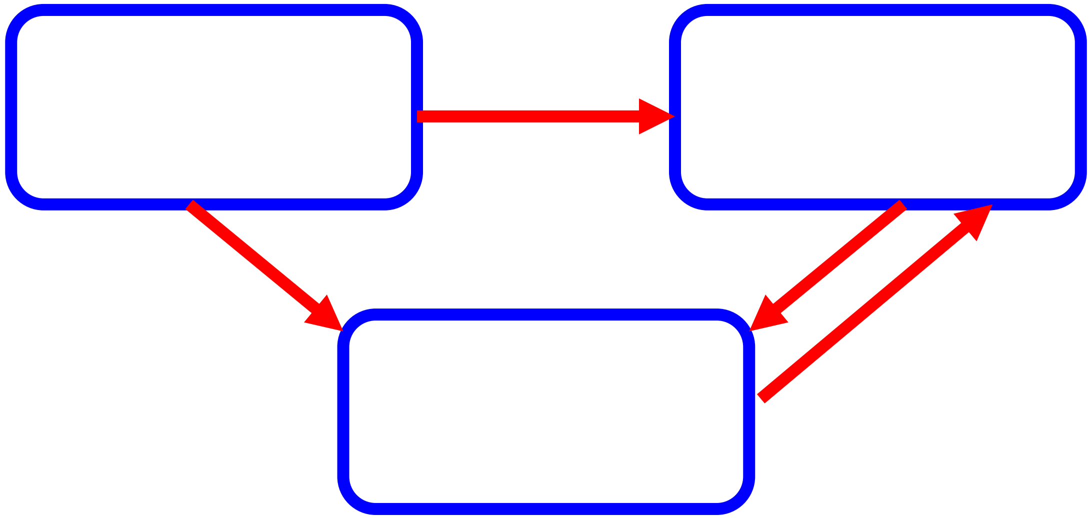
- E.g. Carbon
  - $CO_2$, biomass, dissolved organic carbon (DOC), particulate organic carbon (POC) …
- E.g. Water
  - Oceans, lakes, rivers, clouds, sea ice …
- E.g. Phosphorus
  - Fertilizer, sediments, minerals, bones …
- E.g. Nitrogen
  - $N_2(g)$, ammonia, nitrate, nitrate …




Khan Academy

# Generally…



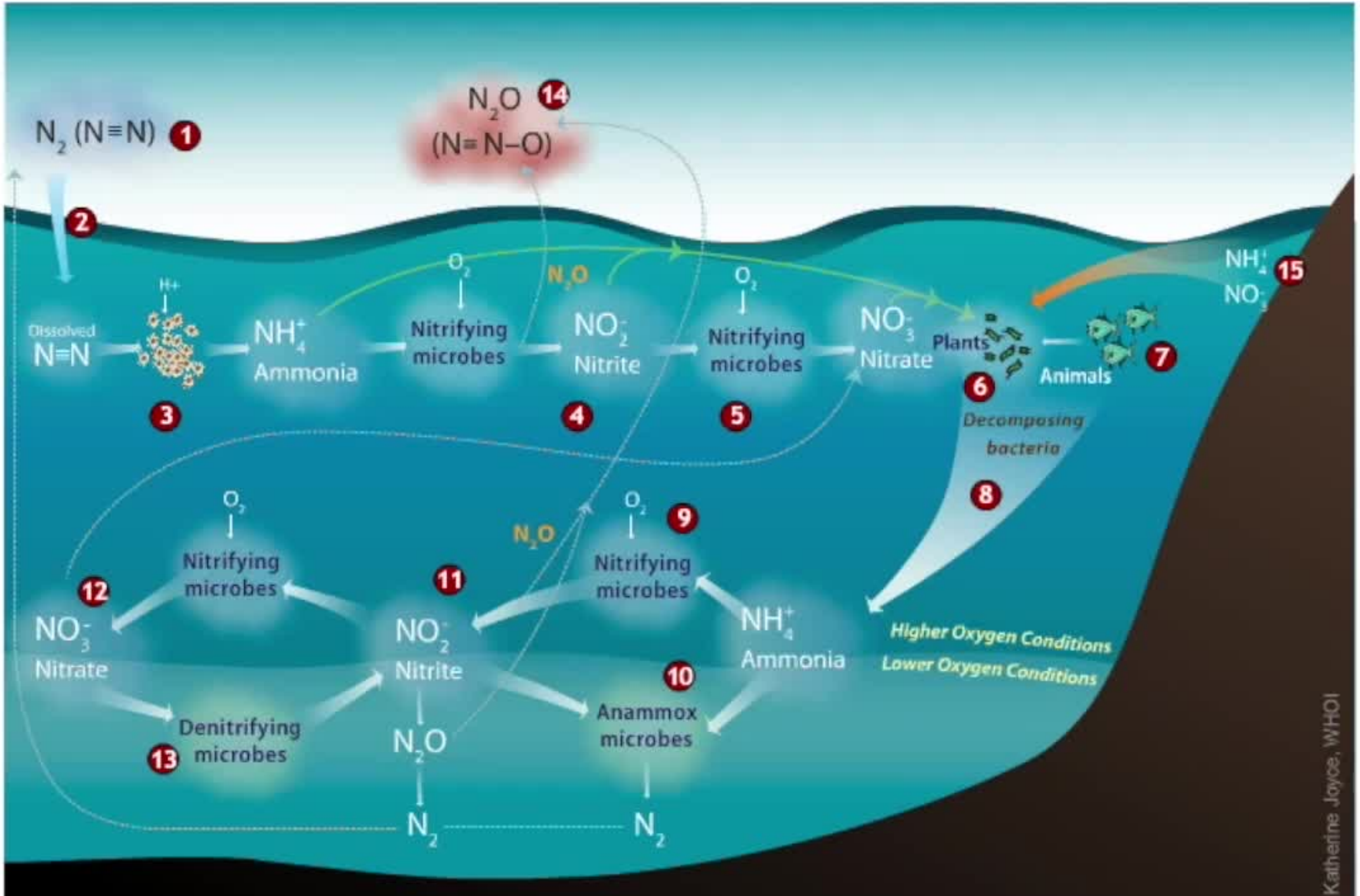



- Process rates are complicated functions of many variables
  - Which vary over time
  - Partial differential equations

# Ecology studies movement and transformation of resources

- E.g. Carbon
  - $CO_2$, biomass, dissolved organic carbon (DOC), particulate organic carbon (POC) …
- E.g. Water
  - Oceans, lakes, rivers, clouds, sea ice …
- E.g. Phosphorus
  - Fertilizer, sediments, minerals, bones …
- E.g. Nitrogen
  - $N_2(g)$, ammonia, nitrate, nitrate …

# The Marine Nitrogen Cycle



# If you understand the arrows, you can derive differential equations

$$\sigma^2_{N^*} = \left(\frac{\partial N^*}{\partial N}\sigma_N\right)^2 + \left(\frac{\partial N^*}{\partial P}\sigma_P\right)^2 + \left(\frac{\partial N^*}{\partial r^{N:P}_{\text{nitr}}}\sigma_{r^{N:P}_{\text{nitr}}}\right)^2 + \left(\frac{\partial N^*}{\partial r^{N:P}_{\text{denitr}}}\sigma_{r^{N:P}_{\text{denitr}}}\right)^2$$

$$= \left(\frac{r^{N:P}_{\text{denitr}}}{r^{N:P}_{\text{denitr}} - r^{N:P}_{\text{nitr}}}\,\sigma_N\right)^2 + \left(\frac{r^{N:P}_{\text{denitr}}}{r^{N:P}_{\text{denitr}} - r^{N:P}_{\text{nitr}}}\,r^{N:P}_{\text{nitr}}\,\sigma_P\right)^2 + \left(\frac{r^{N:P}_{\text{denitr}}}{r^{N:P}_{\text{denitr}} - r^{N:P}_{\text{nitr}}}\right.$$

$$\left(-P + \frac{N - r^{N:P}_{\text{nitr}}P + \text{const}}{r^{N:P}_{\text{denitr}} - r^{N:P}_{\text{nitr}}}\right)\sigma_{r^{N:P}_{\text{nitr}}}\Bigg)^2 + \left(\left(N - r^{N:P}_{\text{nitr}}P + \text{const}\right)\frac{-r^{N:P}_{\text{nitr}}}{\left(r^{N:P}_{\text{denitr}} - r^{N:P}_{\text{nitr}}\right)^2}\,\sigma_{r^{N:P}_{\text{denitr}}}\right)^2.$$

GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 11, NO. 2, PAGES 235–266, JUNE 1997

**Global patterns of marine nitrogen fixation and denitrification**

Nicolas Gruber
Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland

Jorge L. Sarmiento
Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, New Jersey

$\partial$ Means partial derivative, i.e. life just got very hard

# If you understand the arrows, you can derive differential equations

$$\Theta(x; x_c,\ \lambda) \equiv \frac{1}{2}\left(1 - \tanh\left(\frac{x - x_c}{\lambda}\right)\right)$$

Trig???
Seriously???

nature > articles > article

Article | Published: 13 February 2019

**Convergent estimates of marine nitrogen fixation**

Wei-Lei Wang, J. Keith Moore, Adam C. Martiny & François W. Primeau

*Nature* **566**, 205–211 (2019) | Cite this article

**12k** Accesses | **109** Citations | **65** Altmetric | Metrics

*Change over time = a rate, we do those with derivatives → PDEs*

# So here's the thing about PDEs

- Best case: very very hard
- Worst case: impossible
- Or run computer models

# So here's the thing about PDEs

- Best case: very very hard
- Worst case: impossible
- Or run computer models
  - The only practical way
  - Estimate initial quantity in each pool at some starting time
  - Compute transfer in/out of each pool during next simulated day/hour/minute/??? using the PDEs
    - Repeat as long as you like
  - Always some error, which increases over time
  - But good computer models are very good


10^24 stars
10^29 bacteria

# Bacteria in the ocean

# Bacteria in the ocean

- Average concentration: 10^6 cells / ml (Whitman et al. 1998)
- Higher near land (nutrient runoff), near surface (photosynthesis)
- Lower at depth (no sunlight), in oligotrophic open ocean

# Bacteria in the ocean

- Average concentration: 10^6 cells / ml (Whitman et al. 1998)
- Higher near land (nutrient runoff), near surface (photosynthesis)
- Lower at depth (no sunlight), in oligotrophic open ocean

"The moment you sample ocean water, you have a big-data problem"
- PH

# But this is relatively new knowledge

- The old belief:
  - Green / turquoise / light blue = chlorophyll = life
    - Only in shallow water, near land
    - Cyanobacteria = phylum of photosynthesizing bacteria
    - Cyan: R=0, G=255, B=255
  - Dark blue = desert
    - The open ocean
    - Indigo: R=0, G=65, B=106

# But this is relatively new knowledge

- The old belief:
  - Green / turquoise / light blue = chlorophyll = life
    - Only in shallow water, near land
    - Cyanobacteria = the phylum of photosynthesizing bacteria
    - Cyan: R=0, G=255, B=255
  - Dark blue = desert
    - The open ocean
    - Indigo: R= 0, G=65, B=106

# But this is relatively new knowledge

- The old belief:
  - Green / turquoise / light blue = chlorophyll = life
    - Only in shallow water, near land
    - <u>Cyano</u>bacteria = the phylum of photosynthesizing bacteria
    - Cyan: R=0, G=255, B=255
  - Dark blue = desert
    - The open ocean
    - Indigo: R= 0, G=65, B=106

# But this is relatively new knowledge

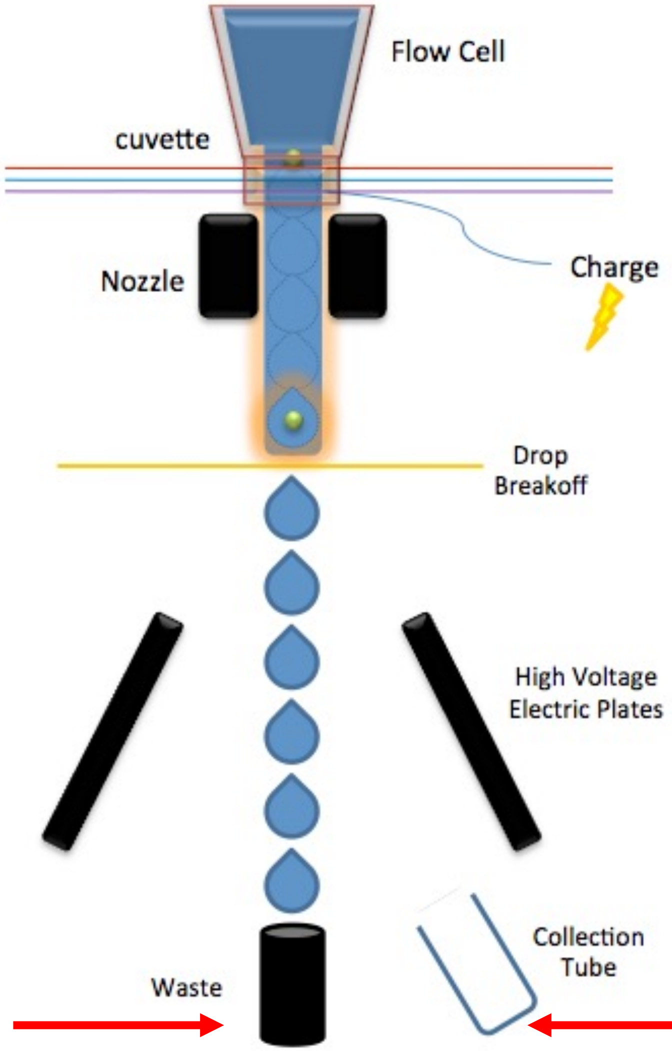
- The old belief:
  - Green / turquoise / light blue = chlorophyll = life
    - Only in shallow water, near land
    - Cyanobacteria = the phylum of photosynthesizing bacteria
    - Cyan: R=0, G=255, B=255
  - Dark blue = desert
    - The open ocean
    - Indigo: R= 0, G=65, B=106


TAHITI
REVA DREAMS

# Sallie (Penny) Chisholm

https://www.ted.com/talks/penny_chisholm_the_tiny_creature_that_secretly_powers_the_planet#t-985653

# Flow Cytometers can *separate* cells based on size/shape/color

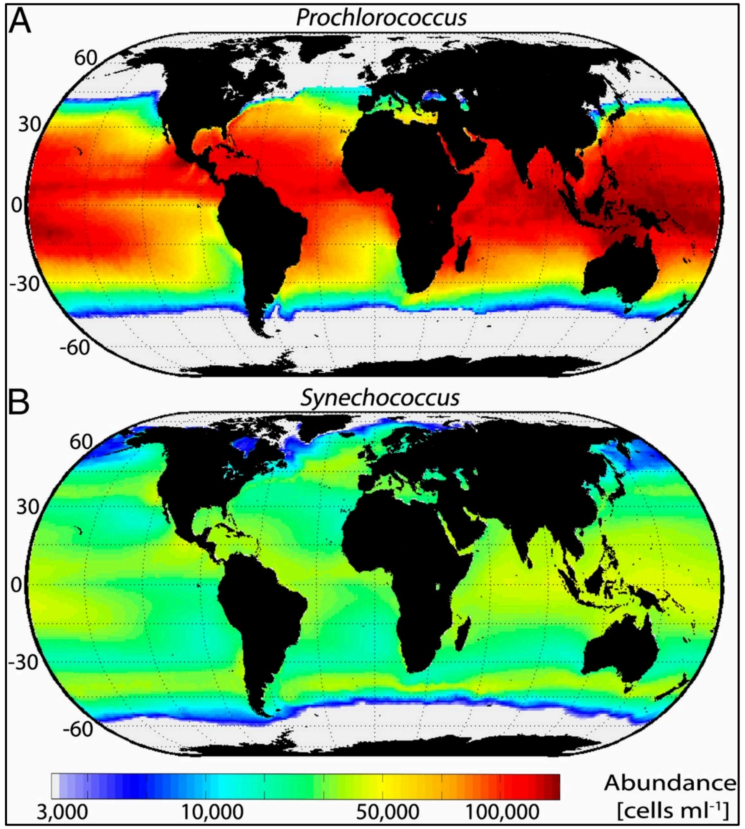



What you don't want

What you want

# Prof. Chisholm’s great discovery

- Flow cytometers work on ships
- Genus *Prochlorococcus*
- A tiny cyanobacterium (photosynthesizer)
- Found in much of the open ocean except high northern and southern latitudes
- (Synechococcus: the previously known open-ocean tiny cyano … note much lower abundance)




Flombaum et al. PNAS 2013
Link to article in note for this slide

# So for a long time

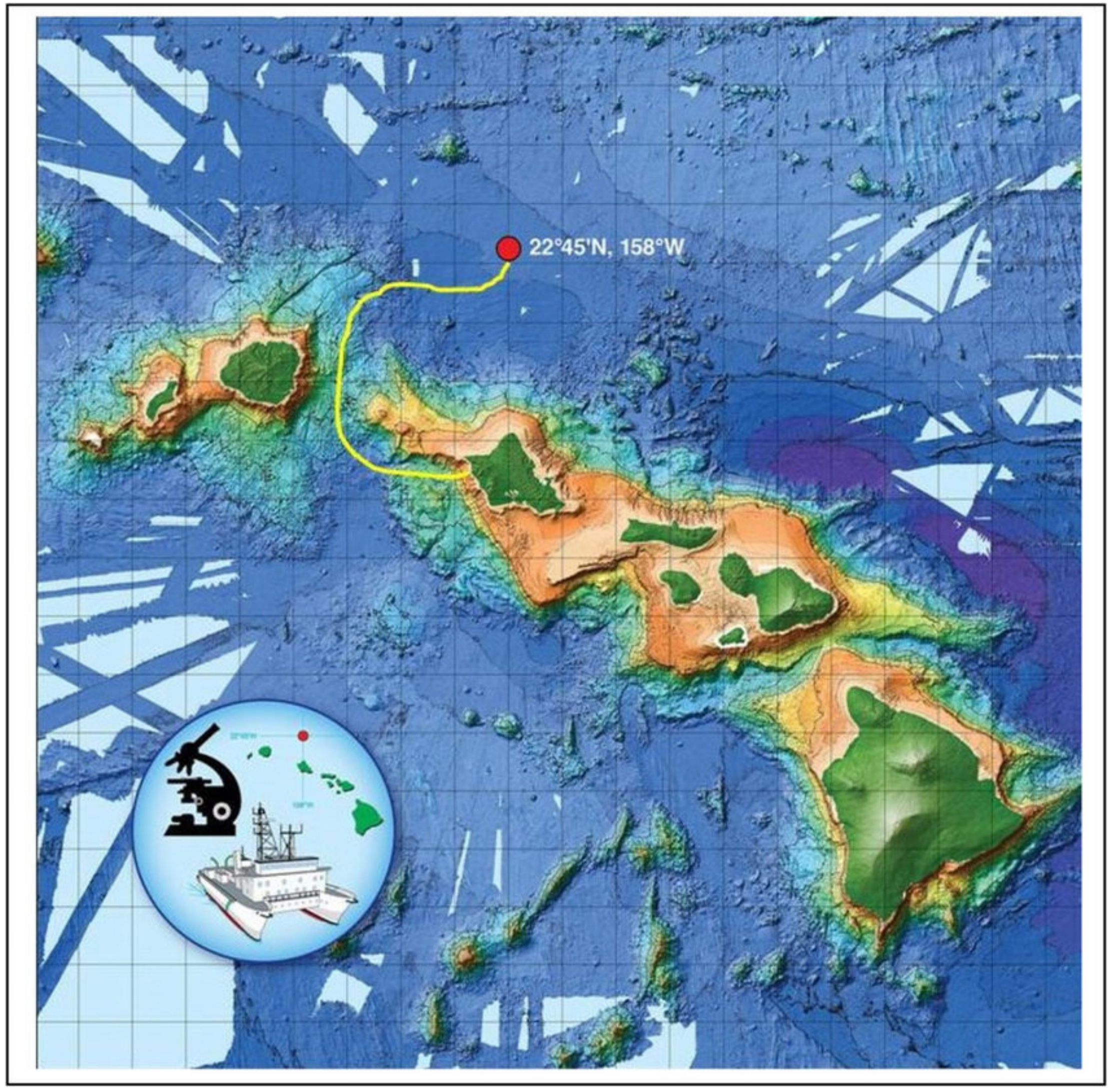
- *Prochlorococcus* and *Synechococcus* were believed to be the main open-ocean bacteria.
- But the ecological budget didn't balance.
- We need ***metagenomic*** studies to figure it out


1μ

$500{,}000{,}000\ km^2$


500,000,000 km$^2$
.000000000004 m$^2$

# A metagenomic search for marine nitrogen fixers at Station Aloha



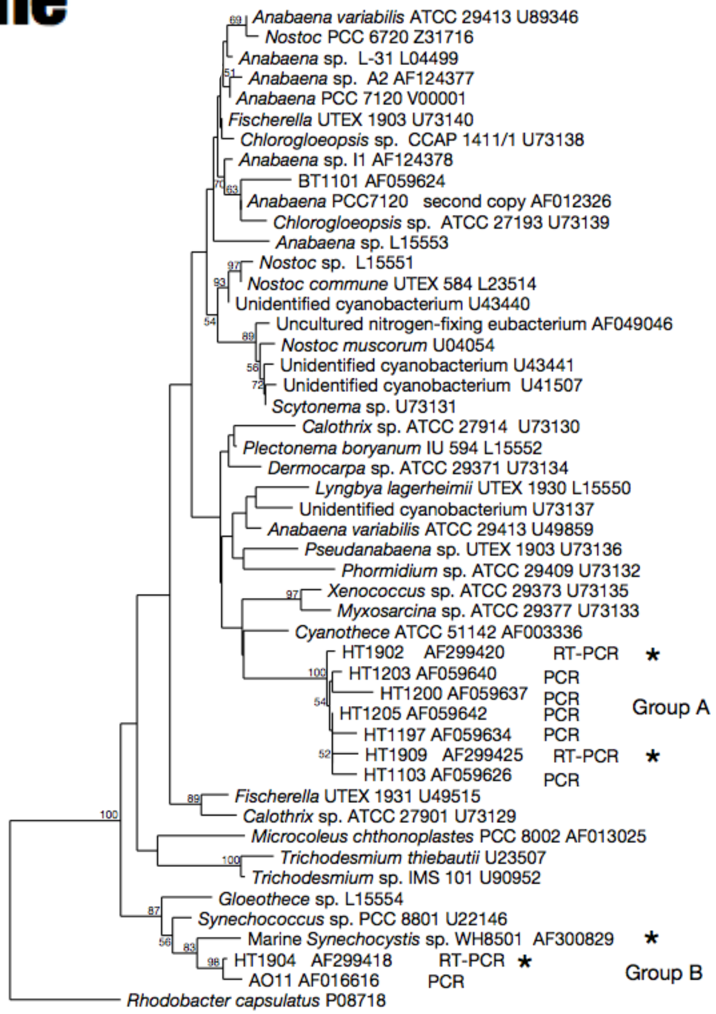

- Sample oligotrophic ocean water
- PCR with *nifH* primers

# Discovery of unicellular marine nitrogen fixers

## Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean

**Jonathan P. Zehr*, John B. Waterbury†, Patricia J. Turner*, Joseph P. Montoya‡, Enoma Omoregie*, Grieg F. Steward*, Andrew Hansen§ & David M. Karl§**

-- Nature, 2001



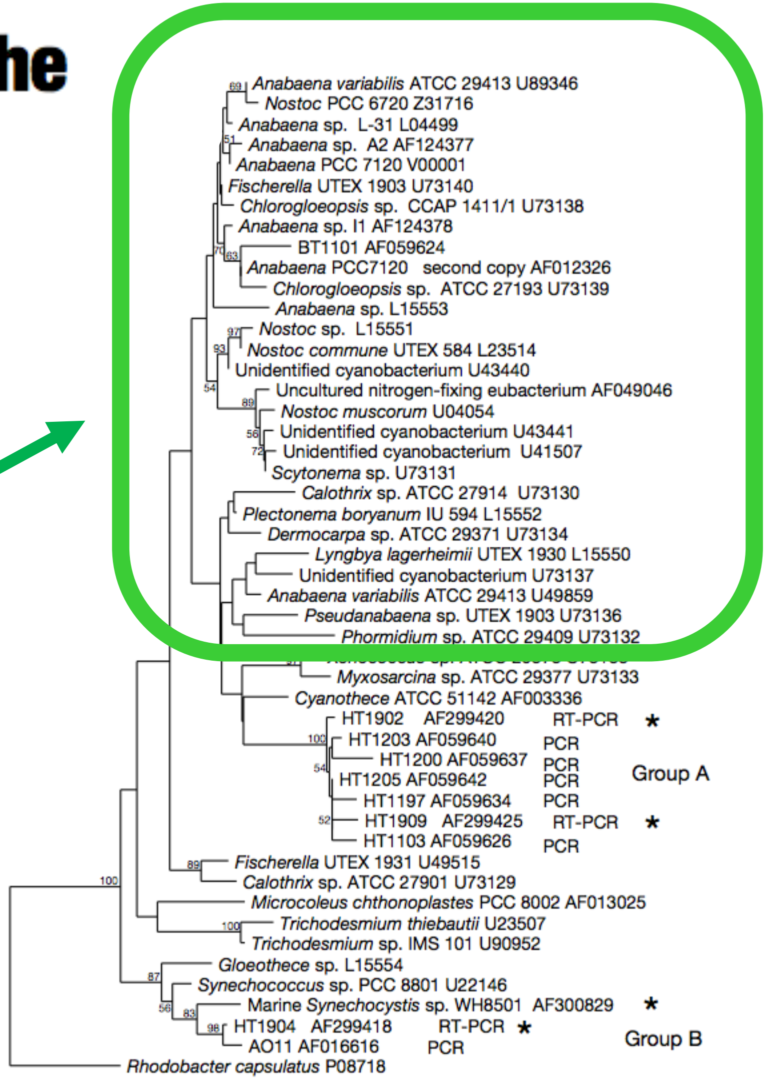
# Discovery of unicellular marine nitrogen fixers

**Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean**

**Jonathan P. Zehr*, John B. Waterbury†, Patricia J. Turner*, Joseph P. Montoya‡, Enoma Omoregie*, Grieg F. Steward*, Andrew Hansen§ & David M. Karl§**

-- Nature, 2001

Already known to science (note the scientific names)



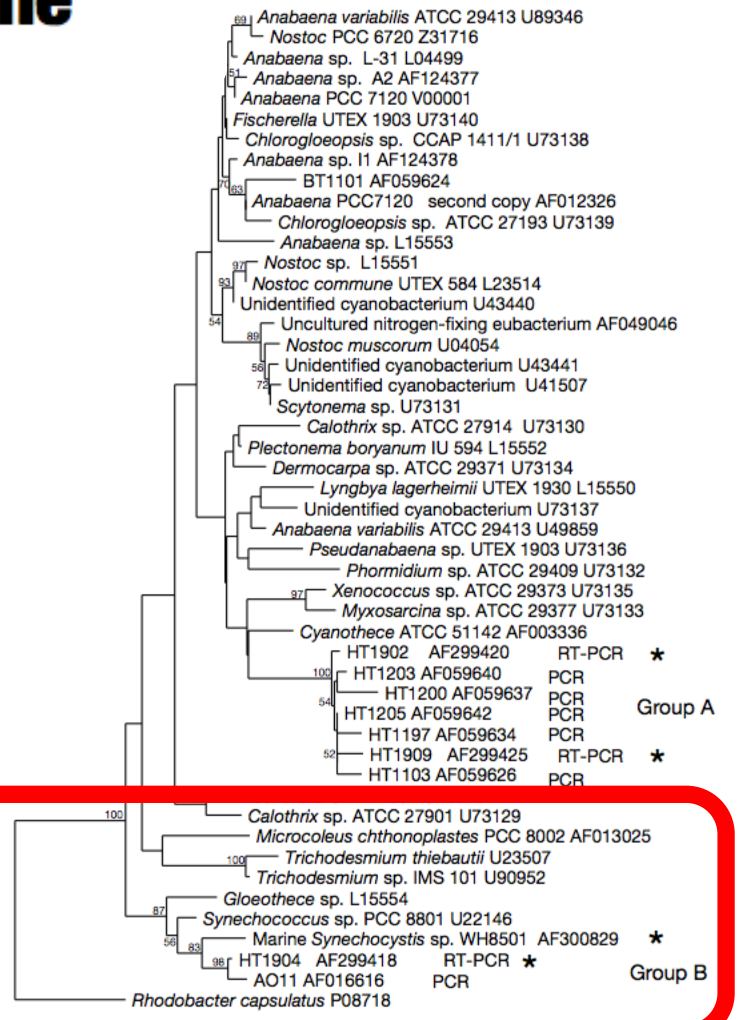
# Discovery of unicellular marine nitrogen fixers

**Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean**

**Jonathan P. Zehr*, John B. Waterbury†, Patricia J. Turner*, Joseph P. Montoya‡, Enoma Omoregie*, Grieg F. Steward*, Andrew Hansen§ & David M. Karl§**

-- Nature, 2001

Also known
to science

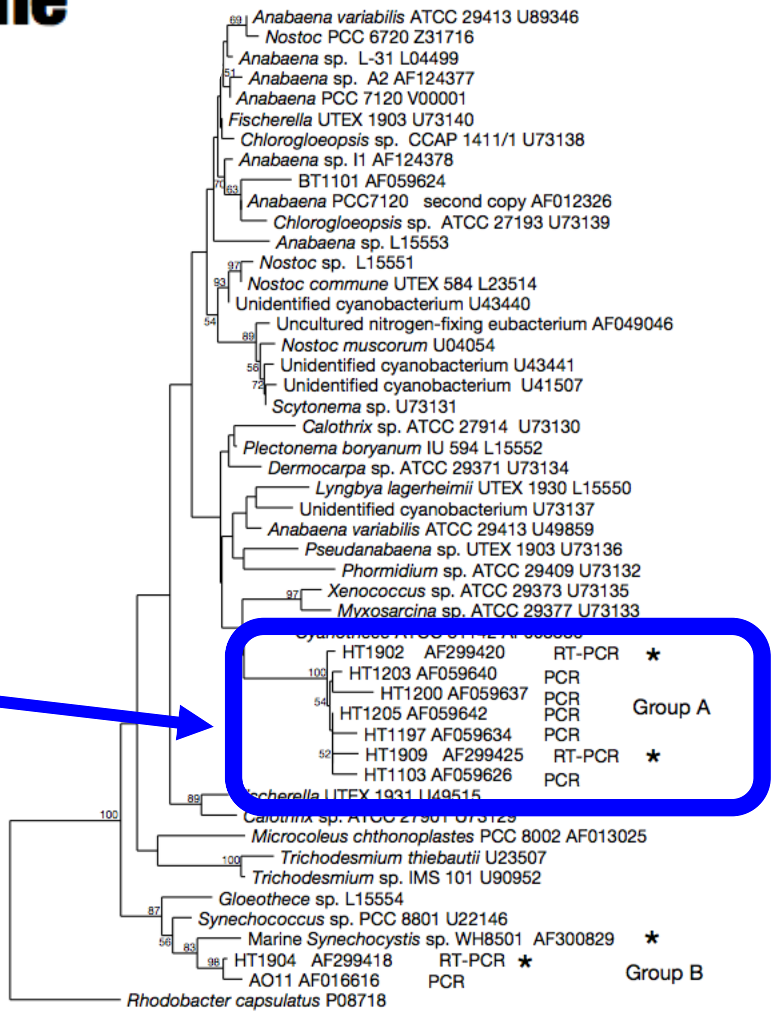


# Discovery of unicellular marine nitrogen fixers

## Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean

**Jonathan P. Zehr*, John B. Waterbury†, Patricia J. Turner*, Joseph P. Montoya‡, Enoma Omoregie*, Grieg F. Steward*, Andrew Hansen§ & David M. Karl§**

-- Nature, 2001




Unknown to science
(no exact BLAST hits)

“Group A”

# Discovery of unicellular marine nitrogen fixers

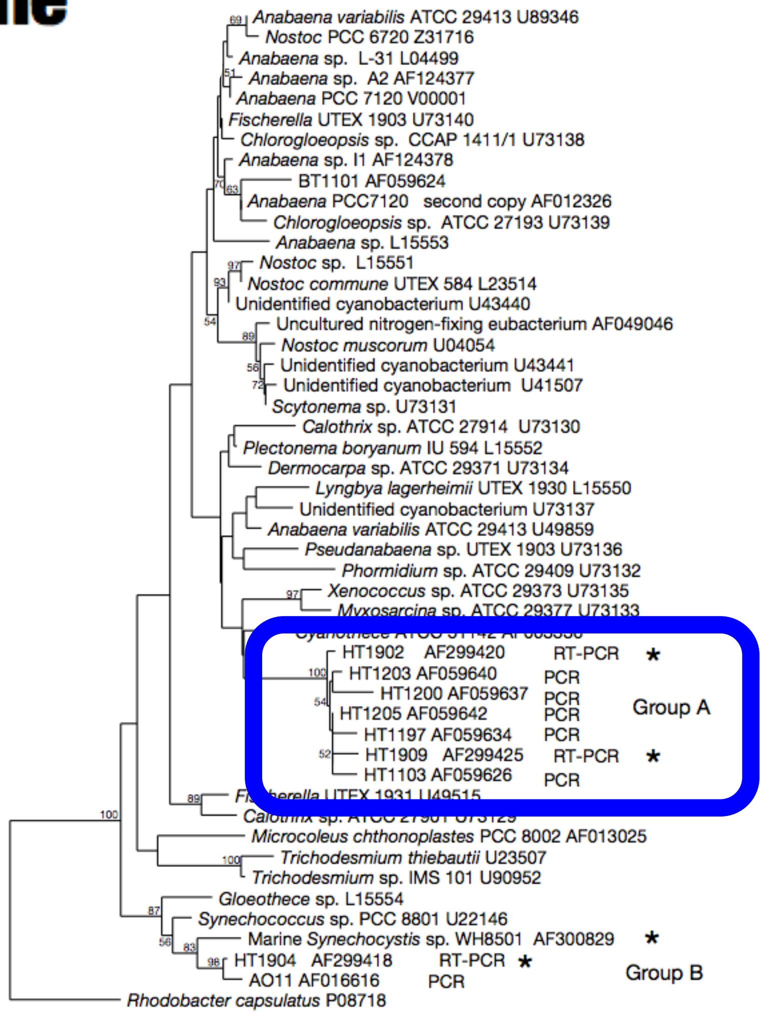
**Unicellular cyanobacteria fix $N_2$ in the subtropical North Pacific Ocean**

**Jonathan P. Zehr*, John B. Waterbury†, Patricia J. Turner*, Joseph P. Montoya‡, Enoma Omoregie*, Grieg F. Steward*, Andrew Hansen§ & David M. Karl§**

-- Nature, 2001

# UCYN-A

Unicellular Cyanobacteria, Group A



# Ok, now what?

**Questions**

- What's the ***genetic potential*** of UCYN-A?
- Is UCYN-A widespread?
- If it's widespread, why was it unknown for so long?

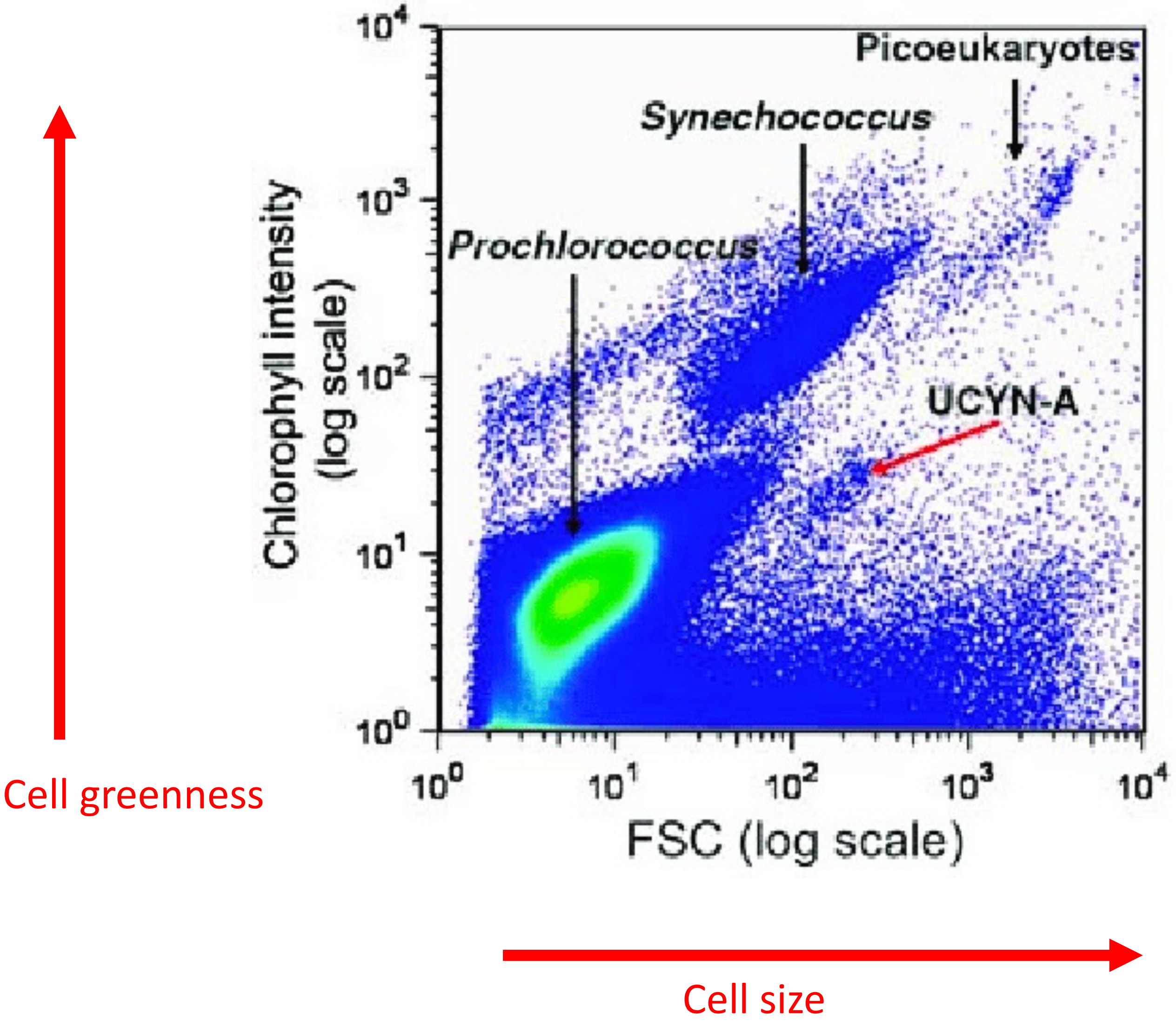
**Techniques**

- Assembly and annotation.
- Look in different environments (e.g. coastal)
- ???????????

# The genetic potential of UCYN-A

1. Get some cells (so far only had *nifH* molecules)
2. Extract and amplify all the DNA
3. Assemble
4. Annotate

# Step 1: Get some cells




This is a *cytogram*

1.A Generate cytogram
1.B Exhaustive PCR search for UCYN-A *nifH*
1.C Isolate UCYN-A cells

Zehr & al, 2008

# Step 2: Extract and amplify all the DNA

- Sequence tech, especially at that time, couldn't determine nucleotide sequences longer that a few hundred base pairs (bp).
- Microbial chromosomes are at least 1M bp.
- There are "restriction" enzymes that cut DNA into right-sized chunks.
- What if you use them, and sequence every chunk? (pretend chunk size = 5).

AGTTC | GCGAT | TACTA | GGTAT | AACTA | TCTAG | GGGAA

↓

GCGAT
AGTTC
GGTAT
GGGAA
TCTAG
AACTA
TACTA

# Step 2: Extract and amplify all the DNA

AGTTC GCGAT TACTA GGTAT AACTA TCTAG GGGAA

↓

GCGAT
AGTTC
GGTAT
GGGAA
TCTAG
AACTA
TACTA

Individually sequenced small molecules, reported in random order by sequencing machine

- What's the right order?
- The data gives no clues.
- There are 5,039 wrong orders *just in this tiny example!*
- What if you use them, and sequence every chunk? (pretend chunk size = 5)
- There has to be a better way

# Step 2, the better way: Shotgun Sequencing

Shotgun pellets fragment targets in random patterns and sizes