Trees, Networks, and Simulation in Biology

from a biologist's perspective

Learning Outcomes

After this class, you will

- Have insight into some key questions and problems biologists are interested in
- Be familiar with examples of three important ways in which biologists visually represent data
- Appreciate how changes in visual representation can give insight (or distort the data)
- Be aware of the noisiness of biological data; biological rules have many exceptions, and data sets can be huge

What are some important biological questions?

Questions biologists ask

- What are the principles governing biological systems?
- How is something (a gene, an organism) evolutionarily related to something else?
- How does something (molecule, protein complex, cell, organism) interact with something else?

Biology is highly descriptive

- Rules are fuzzy and have many exceptions
- Visual representations are key to conveying ideas
- Data are highly noisy
- Datasets can be HUGE
- Goal is to elucidate mechanisms from all those data

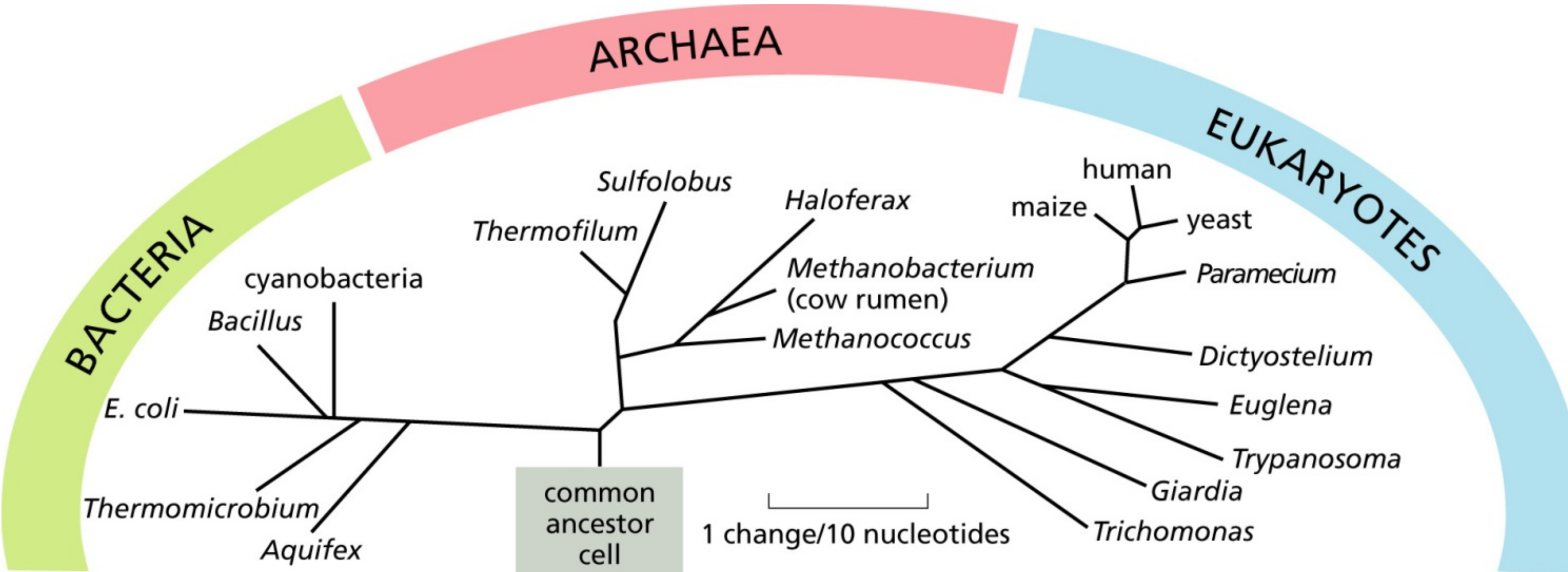
Trees

Questions biologists ask

- What are the principles governing biological systems?
- **How is something (a gene, an organism) evolutionarily related to something else?**
- How does something (molecule, protein complex, cell, organism) interact with something else?

Why is the theory of evolution important?

Evolution makes bioinformatics possible



- All life arose from a common ancestor
- Leaves of tree are existing species or sequences
- Internal nodes are hypothetical common ancestors
- Length of branches is amount of change between ancestor and descendent (for trees using gene sequences as data; in many cases, length is not meaningful)

Reading phylogenetic trees

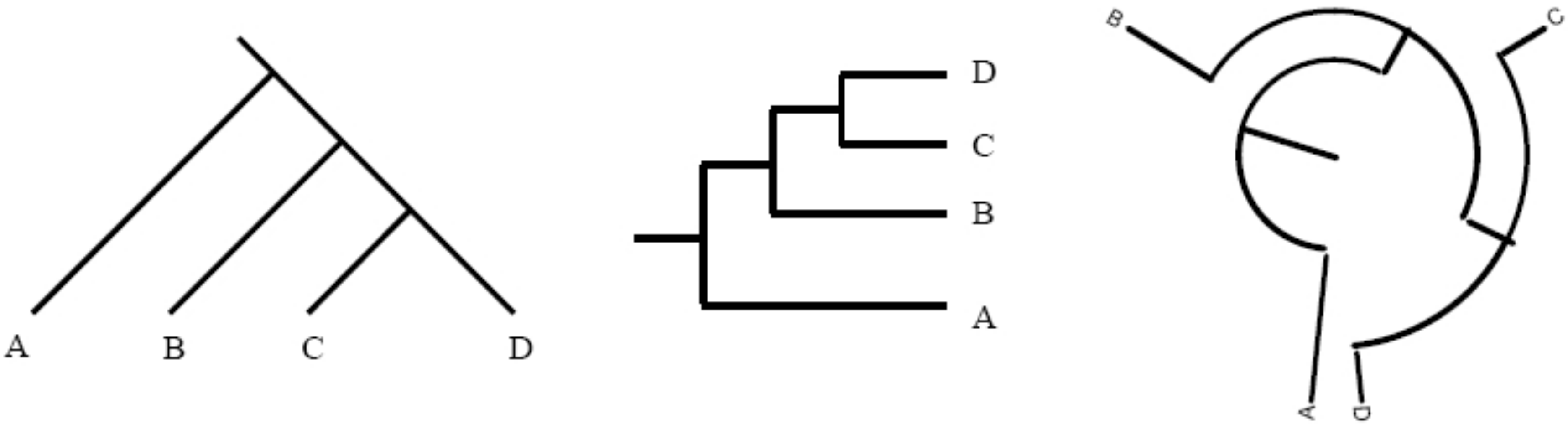
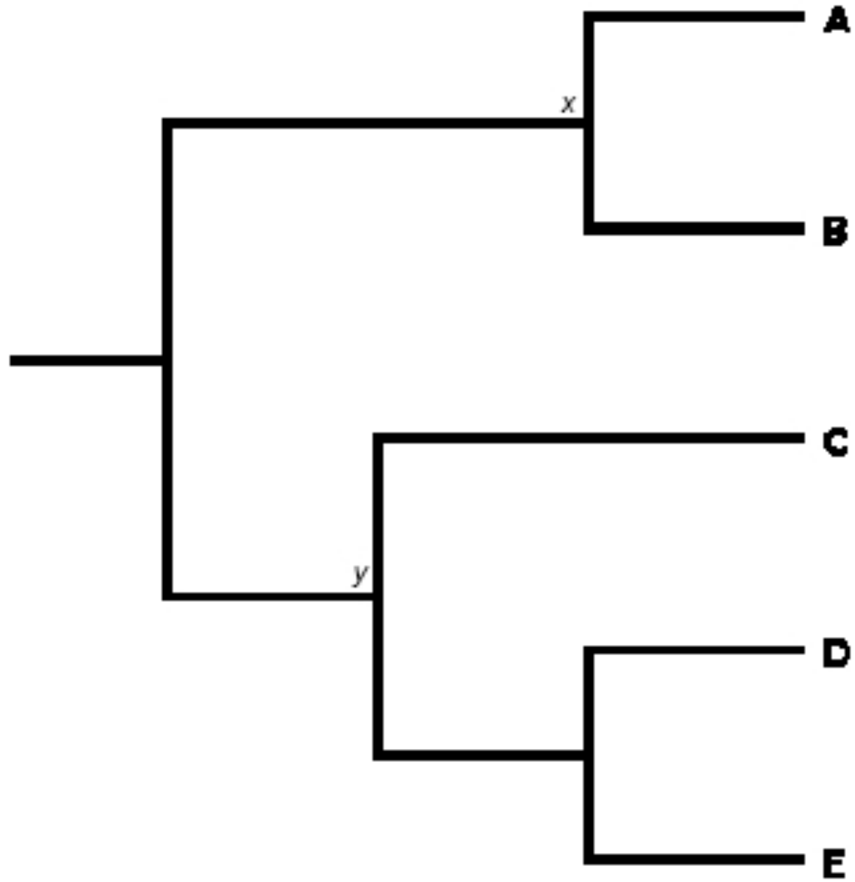


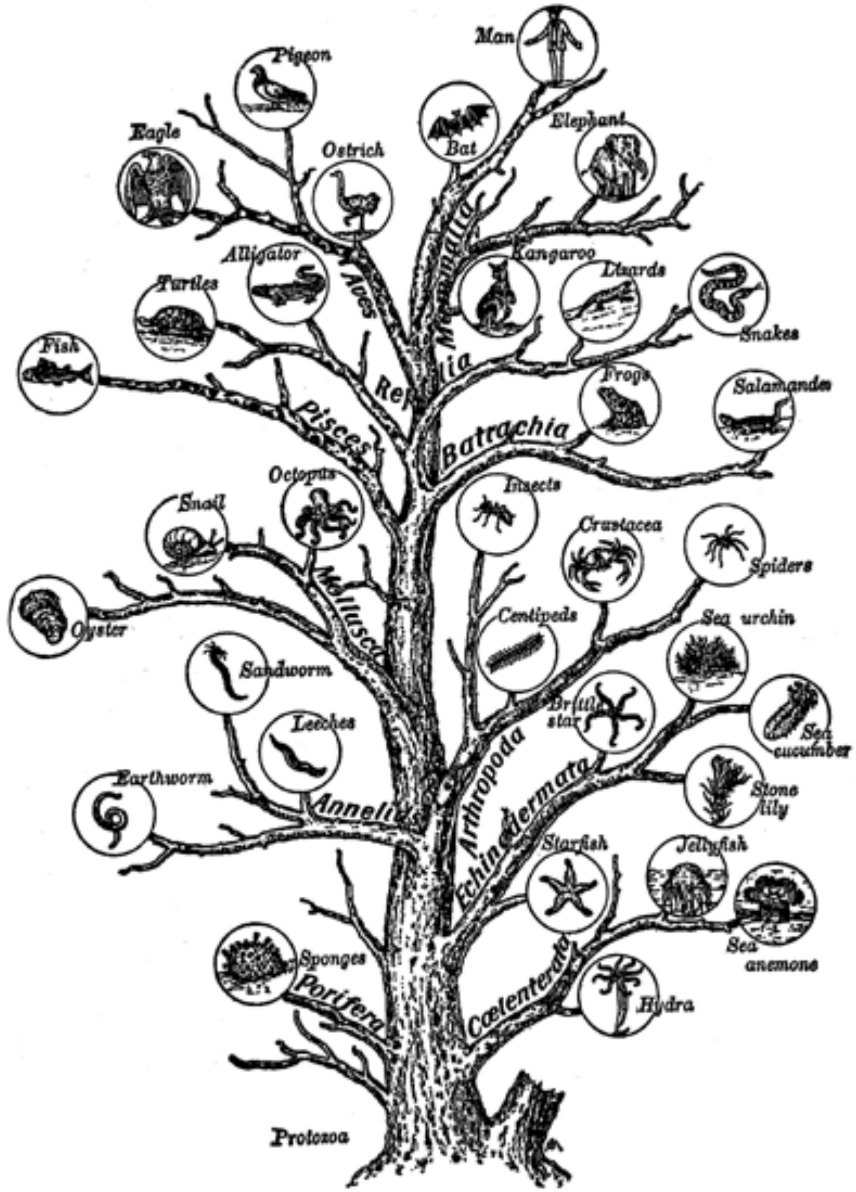
Figure 6 : Types of phylogenetic trees.
These trees depict equivalent relationships, despite having different appearanc

Baum, D. (2008) Reading a phylogenetic tree: The meaning of monophyletic groups. *Nature Education* 1(1):190

http://www.nature.com/scitable/topicpage/reading-a-phylogenetic-tree-the-meaning-of-41956#

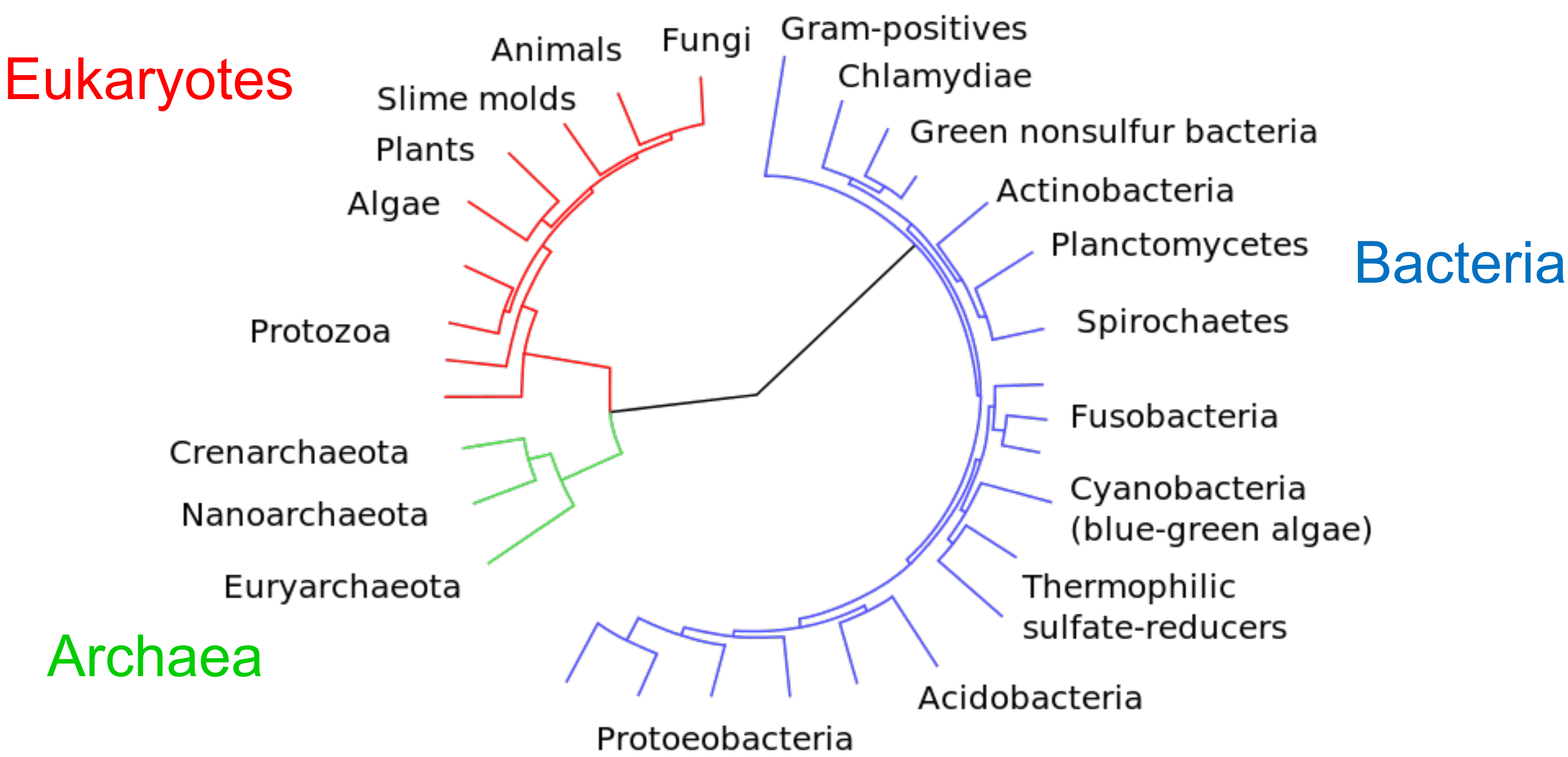


- **Figure 8: Trees contain information on the relative timing of nodes only when the nodes are on the same path from the root (i.e., when one node is a descendant of another).**
- In this tree, nodes x and y are not on the same path, so we cannot tell whether the ancestral organisms in node x lived before or after those in node y.
- In some case, branch length DOES mean something – e.g. changes / 10 nucleotides



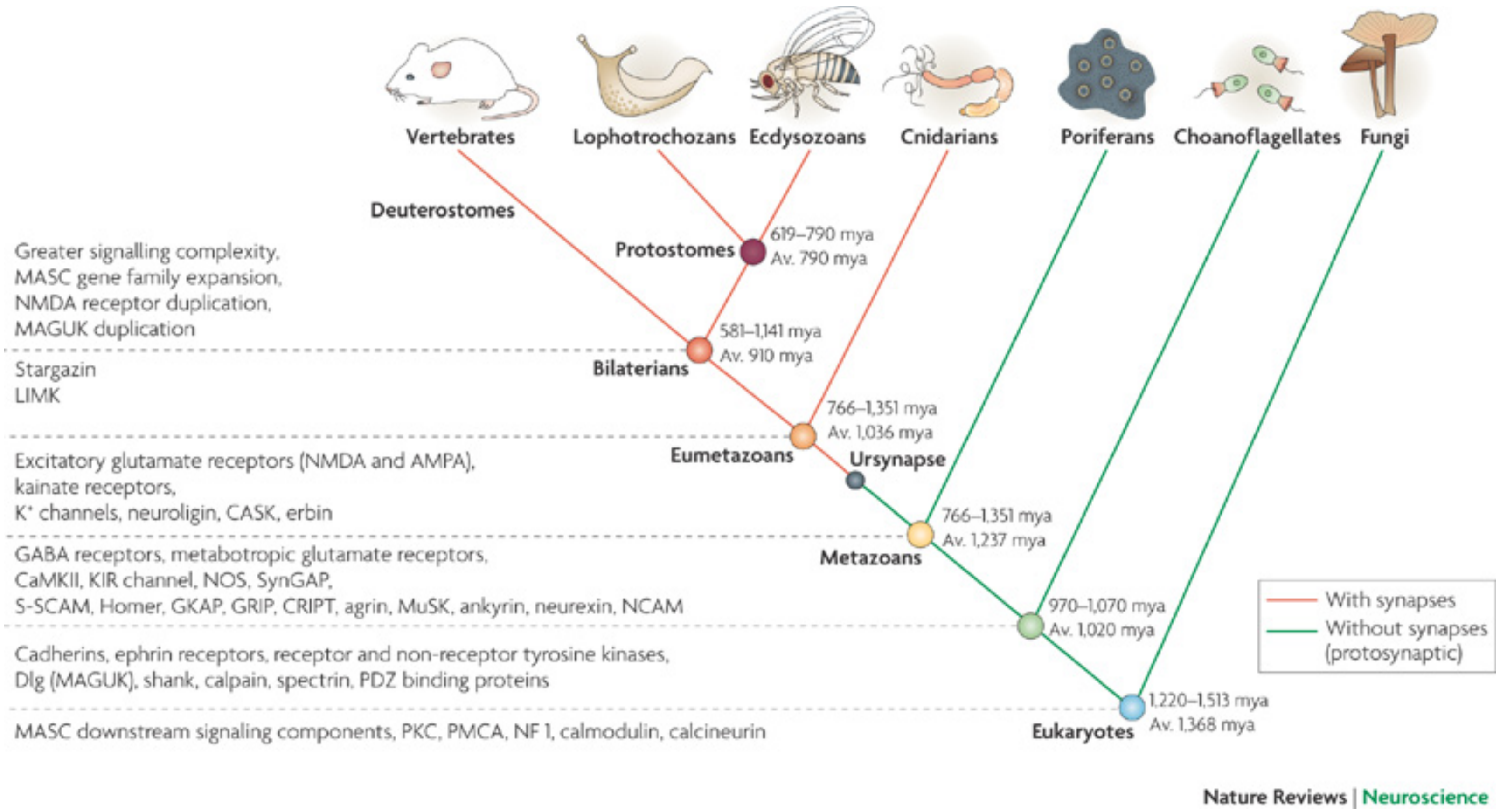
Man
Pigeon
Bat
Elephant
Eagle
Ostrich
Mammalia
Aves
Alligator
Kangaroo
Lizards
Turtles
Snakes
Fish
Reptilia
Frogs
Salamander
Pisces
Batrachia
Octopus
Insects
Snail
Crustacea
Spiders
Mollusca
Oyster
Centipeds
Sea urchin
Sandworm
Brittle star
Sea cucumber
Leeches
Arthropoda
Earthworm
Stone lily
Annelida
Echinodermata
Starfish
Jellyfish
Sea anemone
Sponges
Porifera
Hydra
Cœlenterata
Protozoa

Tree of life from sequenced genomes



A highly resolved Tree of Life, based on completely sequenced genomes [1]. The image was generated using iTOL: Interactive Tree Of Life [2], an online phylogenetic tree viewer and Tree Of Life resource. Eukaryotes are colored red, archaea green and bacteria blue. PNG image traced by hand to produce SVG version. 1. ↑ Ciccarelli, FD (2006). "Toward automatic reconstruction of a highly resolved tree of life." (Pubmed). Science 311(5765): 1283-7. 2. ↑ Letunic, I (2007). "Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.". Bioinformatics 23(1): 127-8.

The Evolution of Synapses

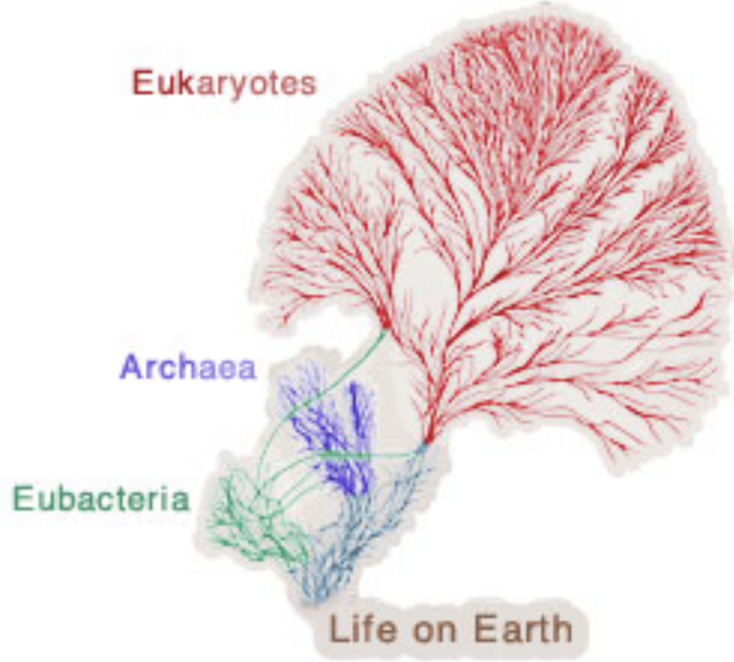


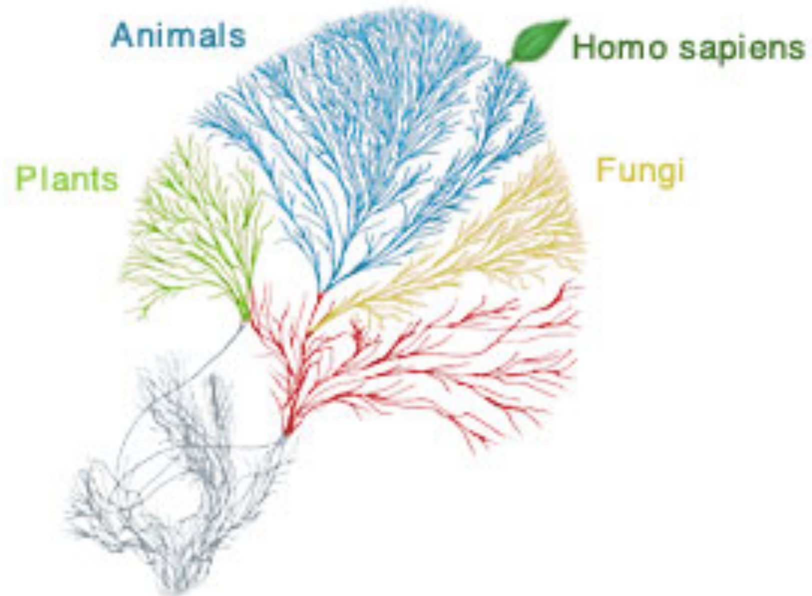
With synapses

Without synapses

Hierarchically nested subgroups

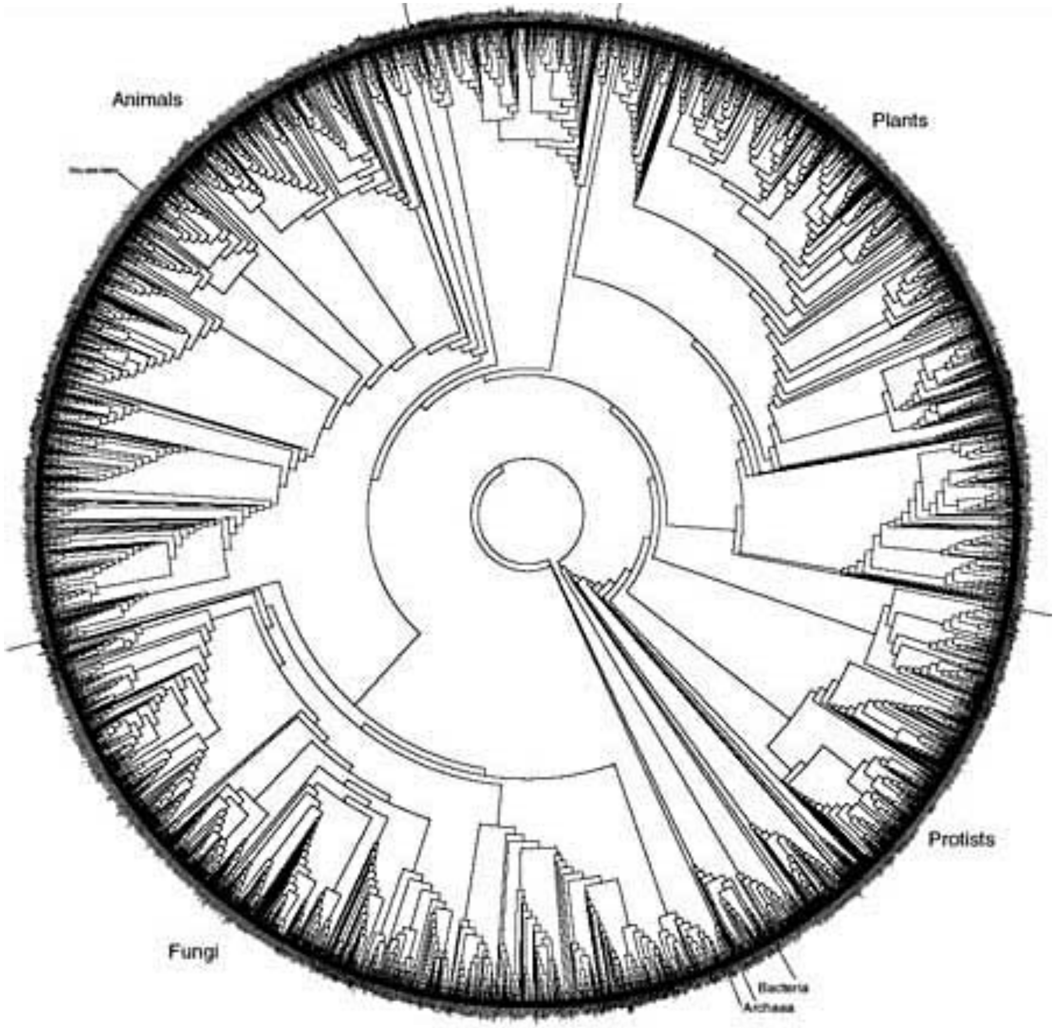
http://tolweb.org/tree/home.pages/structure.html



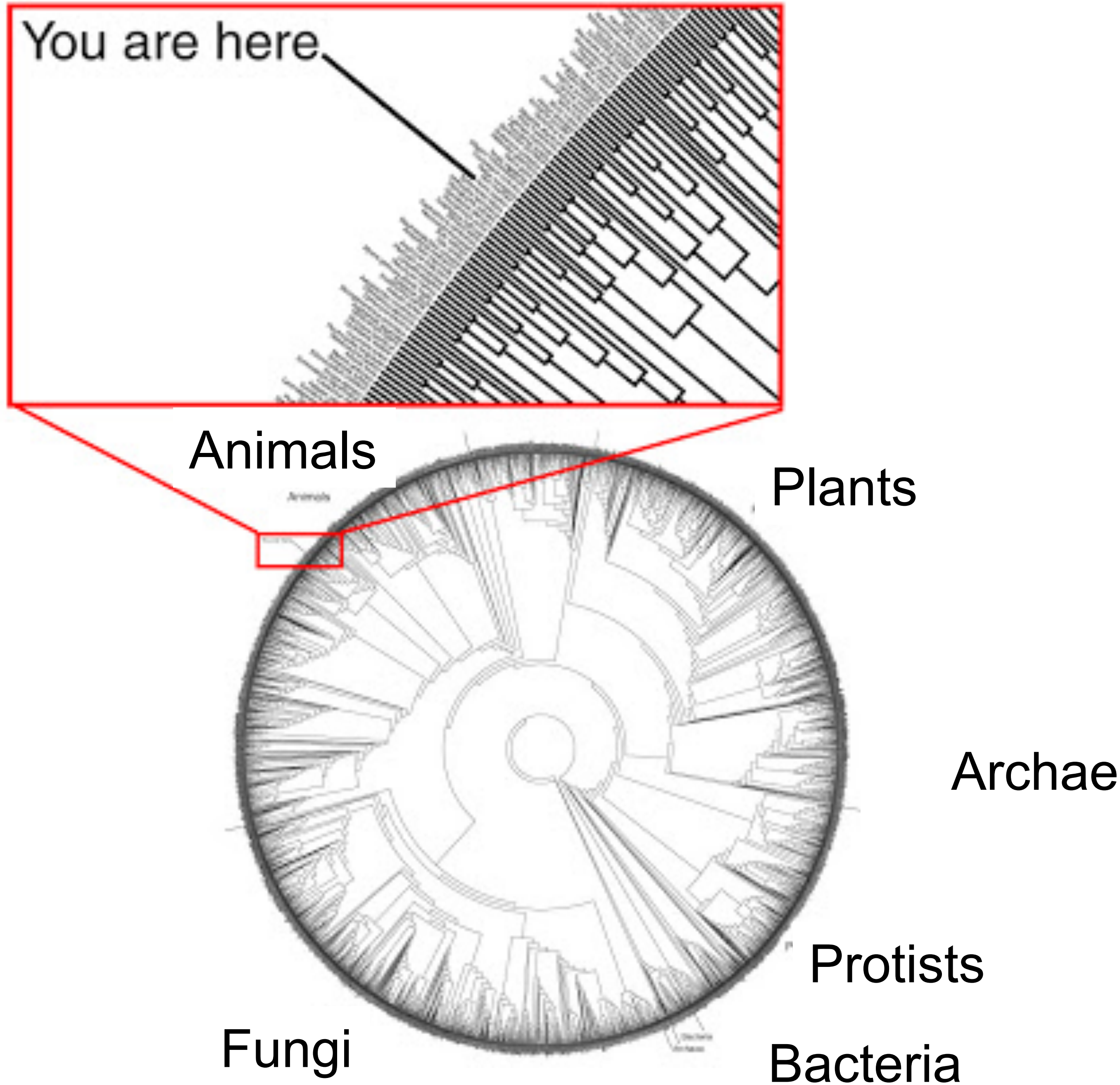


Life on Earth can be divided into a series of hierarchically nested subgroups, starting at the root of all life and ending at the tips in groups that cannot be further subdivided into distinct genetic lineages, e.g., *Homo sapiens* (humans).

http://www.utexas.edu/features/2008/03/03/tree/

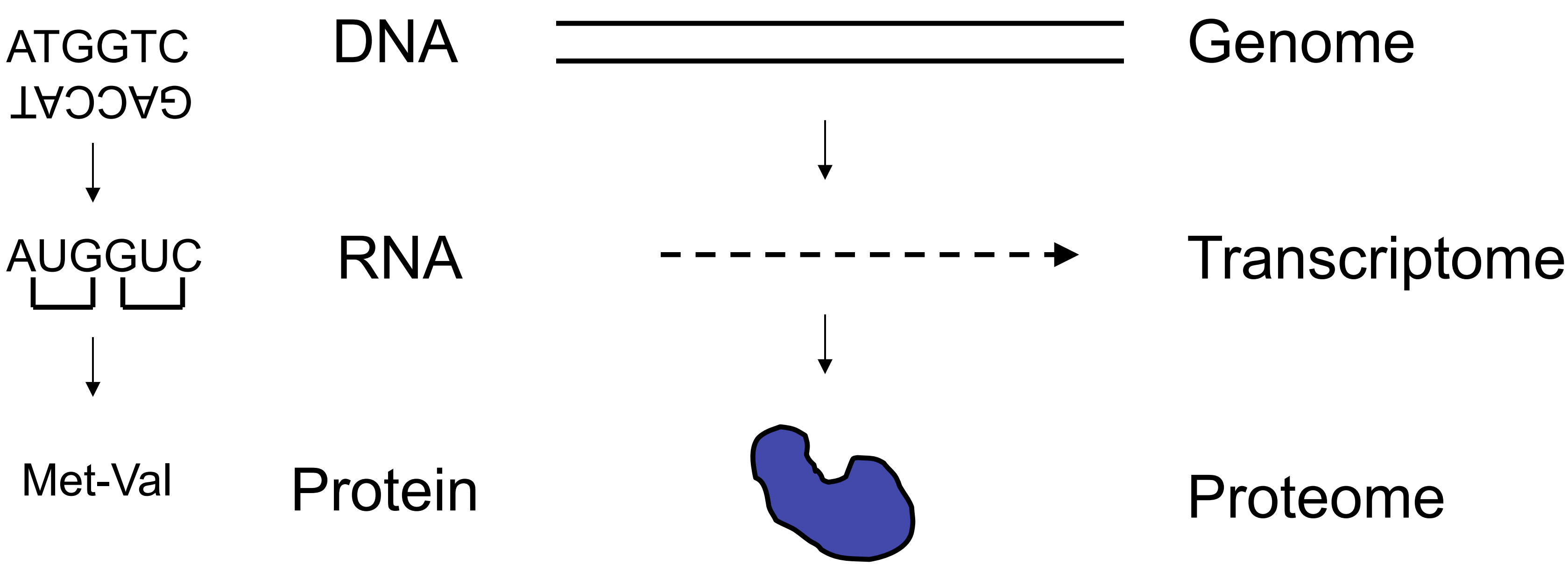


This phylogenetic tree, created by David Hillis, Derreck Zwickil and Robin Gutell, depicts the evolutionary relationships of about 3,000 species throughout the Tree of Life. Less than 1 percent of known species are depicted.


You are here
Animals
Plants
Archae
Protists
Fungi
Bacteria

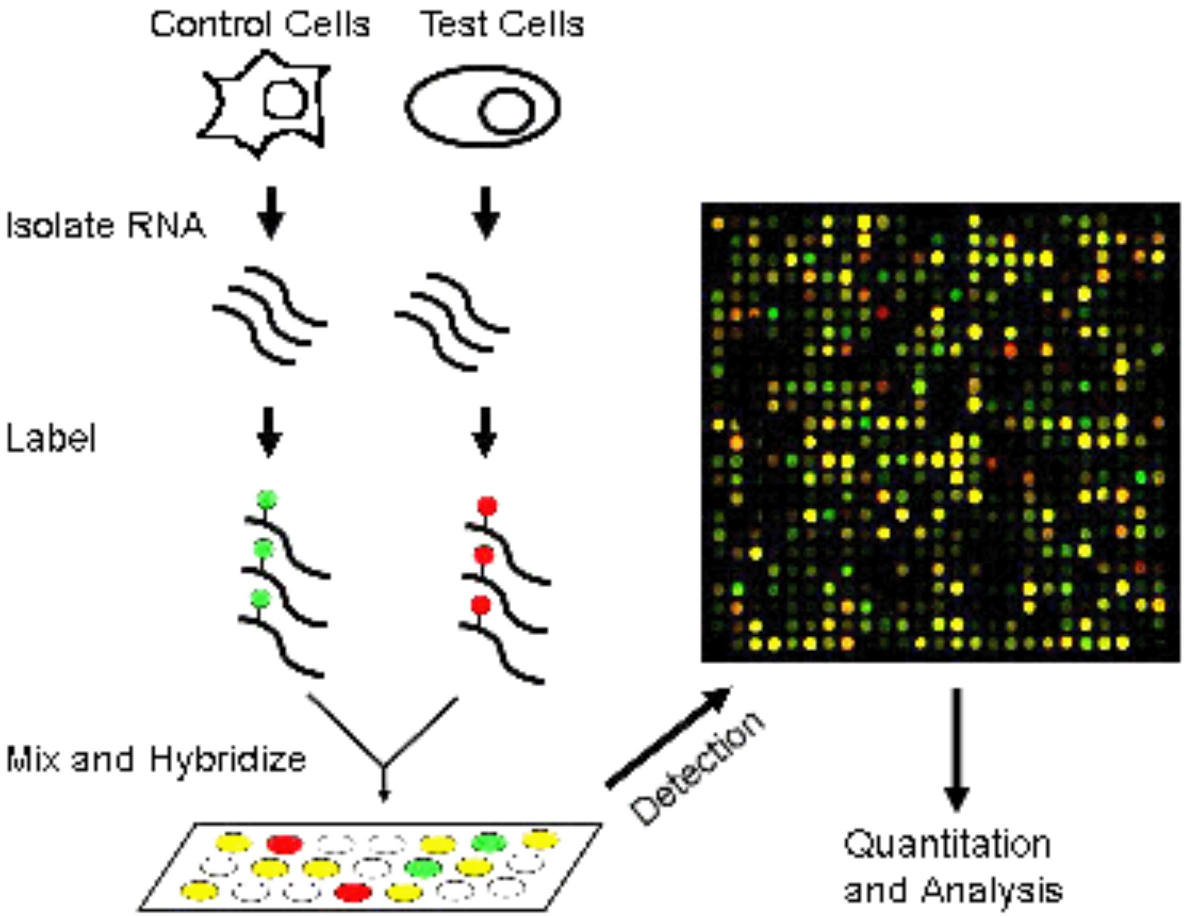
Another tree question

The Central Dogma of Molecular Biology: Genes Encode Proteins



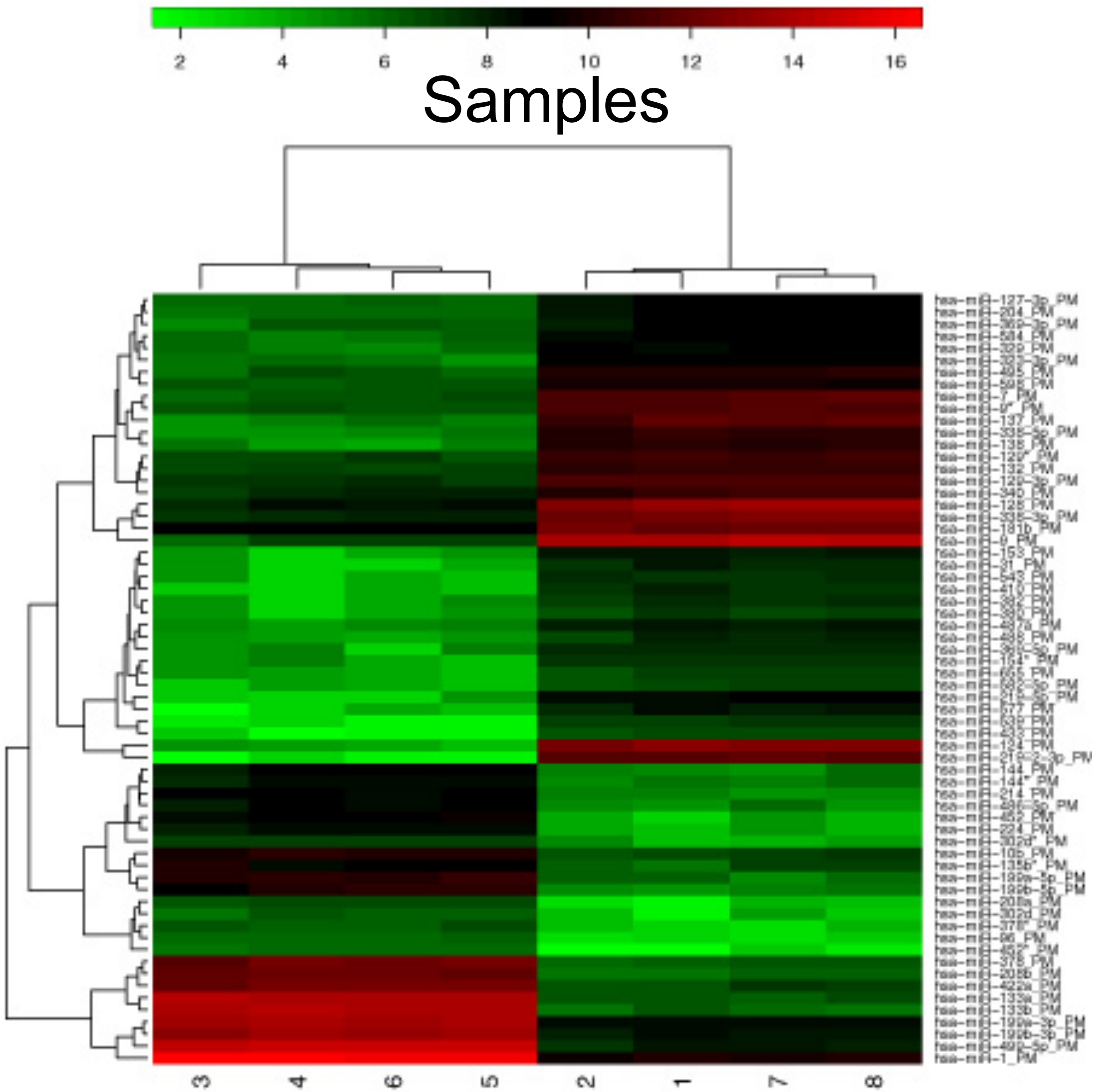
- All cells in an organism have the same DNA
- Do all cells in an organism have the same proteins?
- What genes are activated in response to a stimulus? (e.g. immune system)
- What genes are turned on (or off) differently in a tumor vs. a normal cell?

A microarray experiment



Data are variable!

- http://azcc.arizona.edu/research/shared-resources/gsr/services



Genes (mRNAs)

What are the trees telling you?

http://www.cbioc.com/en/services/bioinformatics-services/

Networks

Questions biologists ask

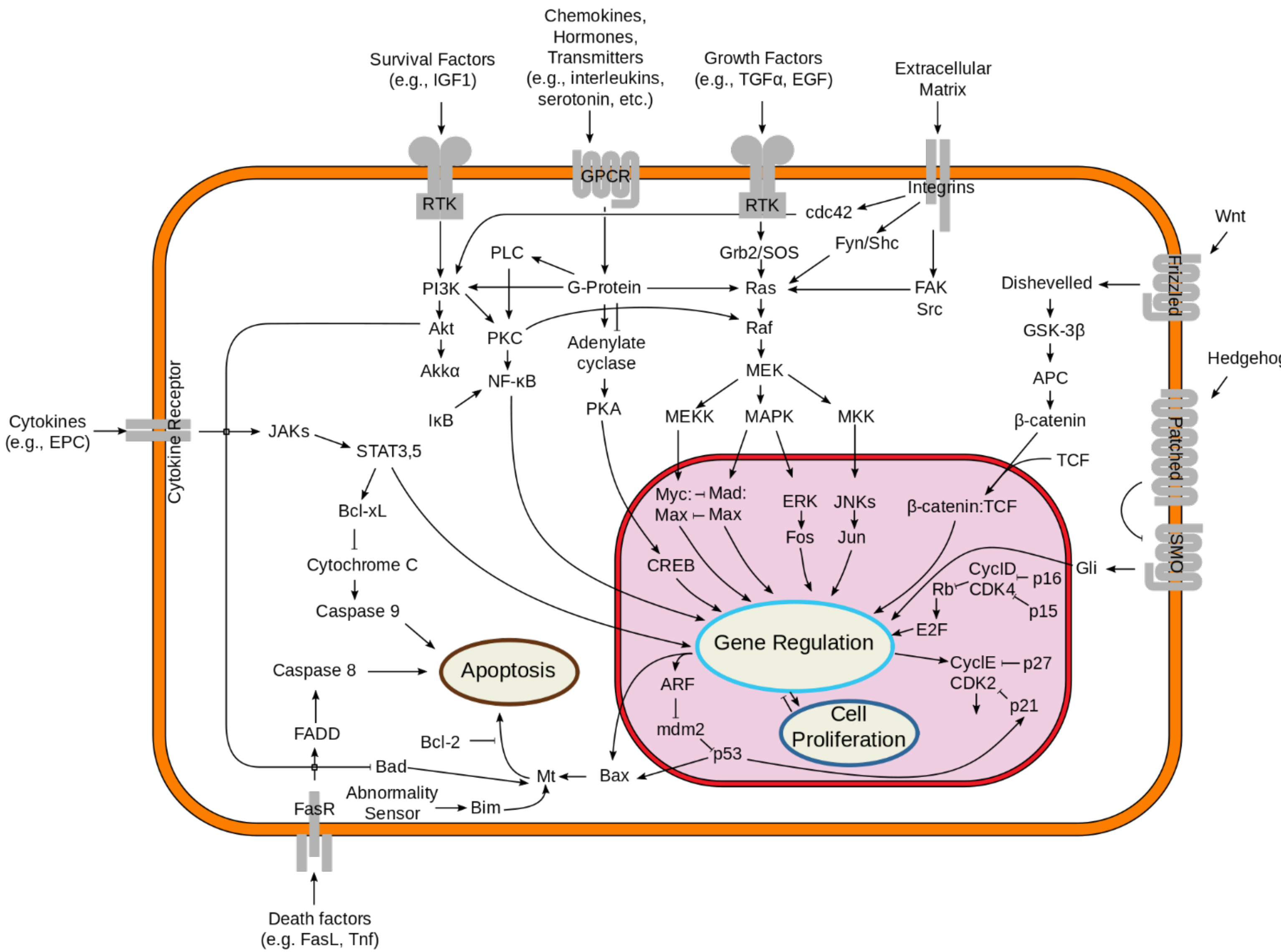
- What are the principles governing biological systems?
- How is something (a gene, an organism) evolutionarily related to something else?
- **How does something (molecule, protein complex, cell, organism) interact with something else?**

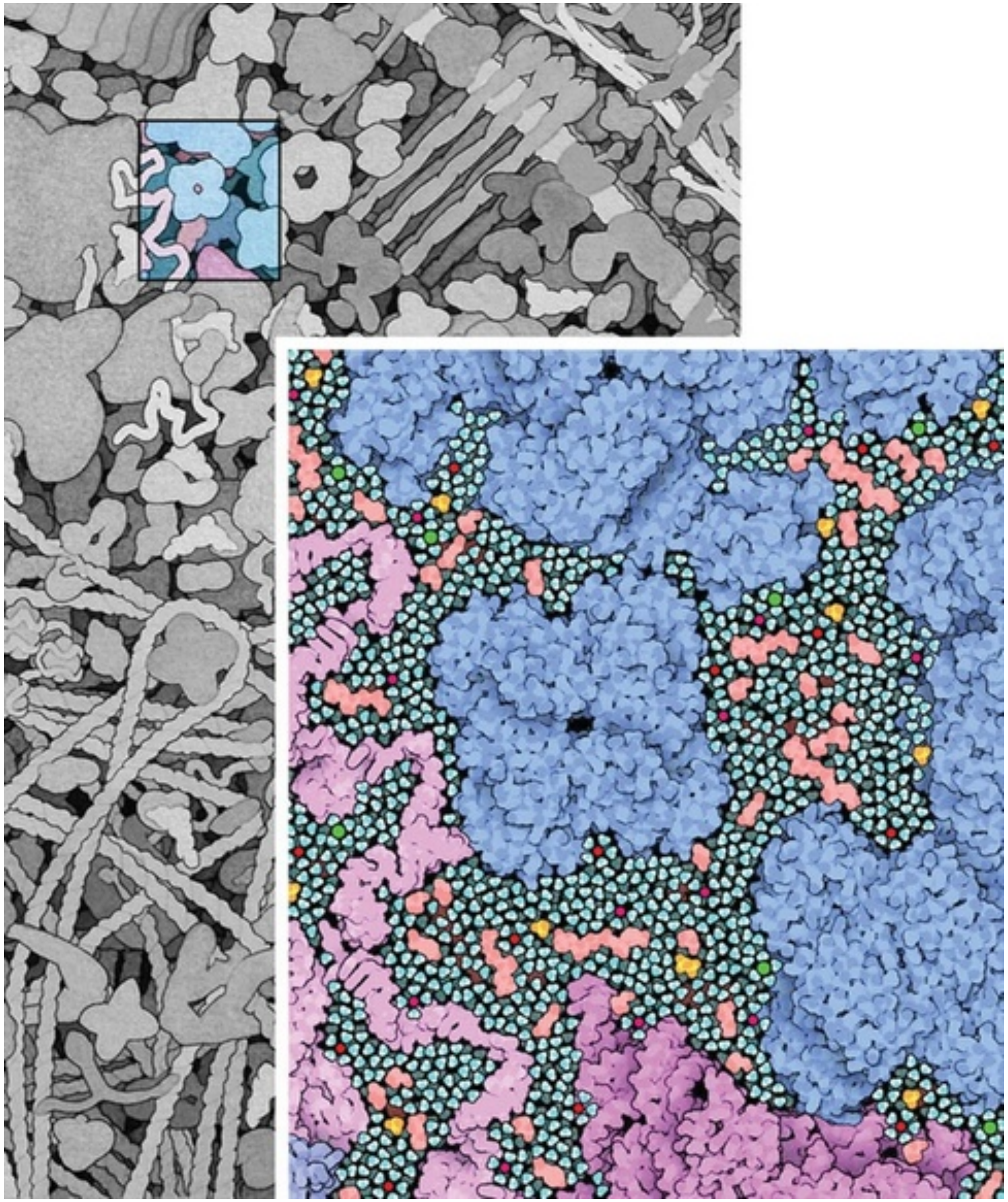
Networks in biology

- Protein-protein interaction
- Gene regulatory (protein-DNA)
- Gene co-expression
- Metabolic (biochemical reaction)
- Signaling
- Neural networks
- Food webs
- Ecological networks

How do cells respond to stimuli?

Signal Transduction Pathways





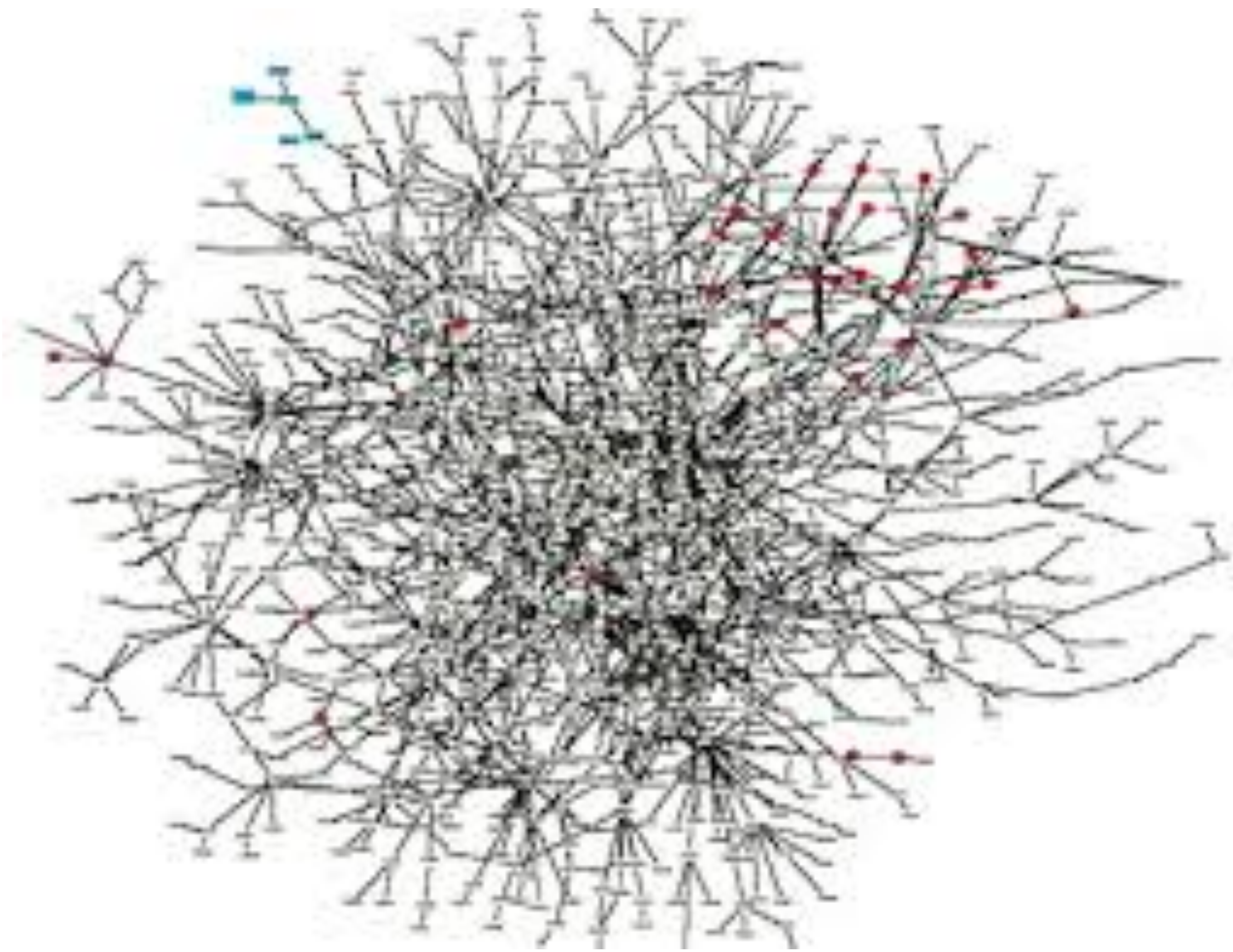
Proteins are densely packed in cells

Little blue guys are water molecules

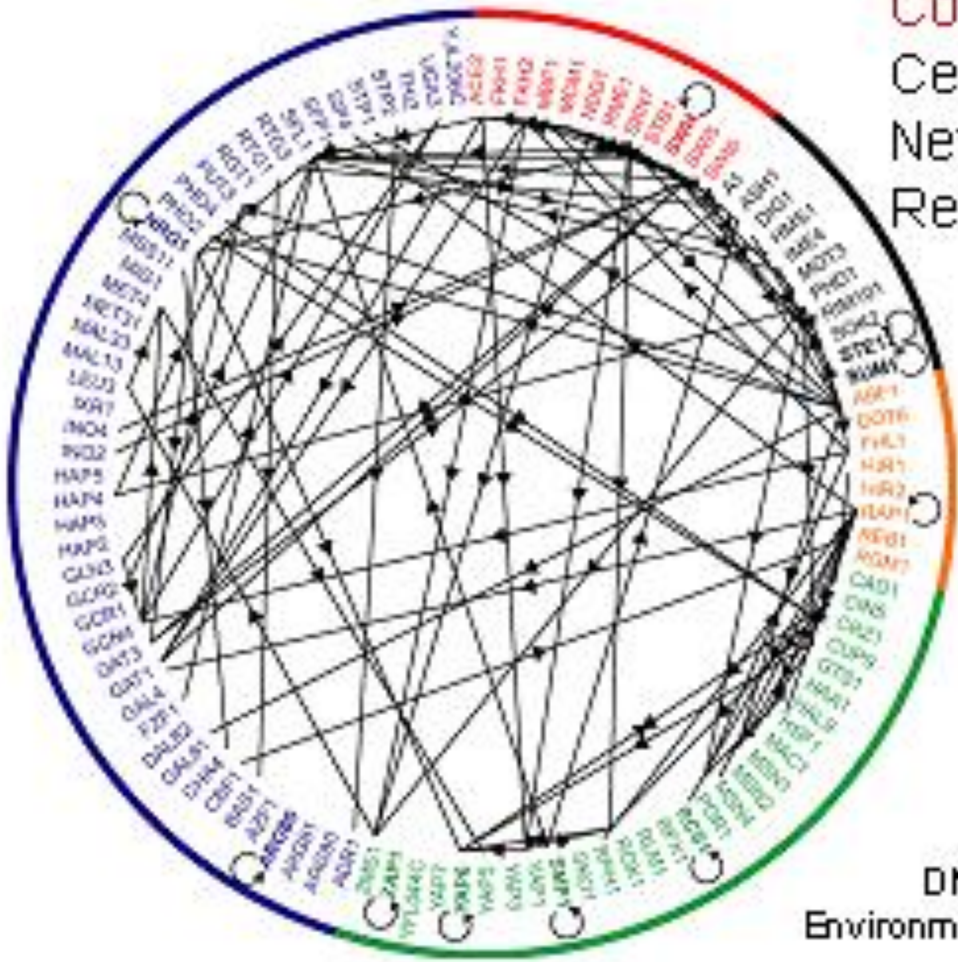
http://www.nature.com/nchembio/journal/v7/n6/images_article/nchembio.575-F1.jpg
https://www.youtube.com/watch?v=uHeTQLNFTgU

What kinds of data support the existence of pathways?

Protein-protein interaction networks: the classic hairball

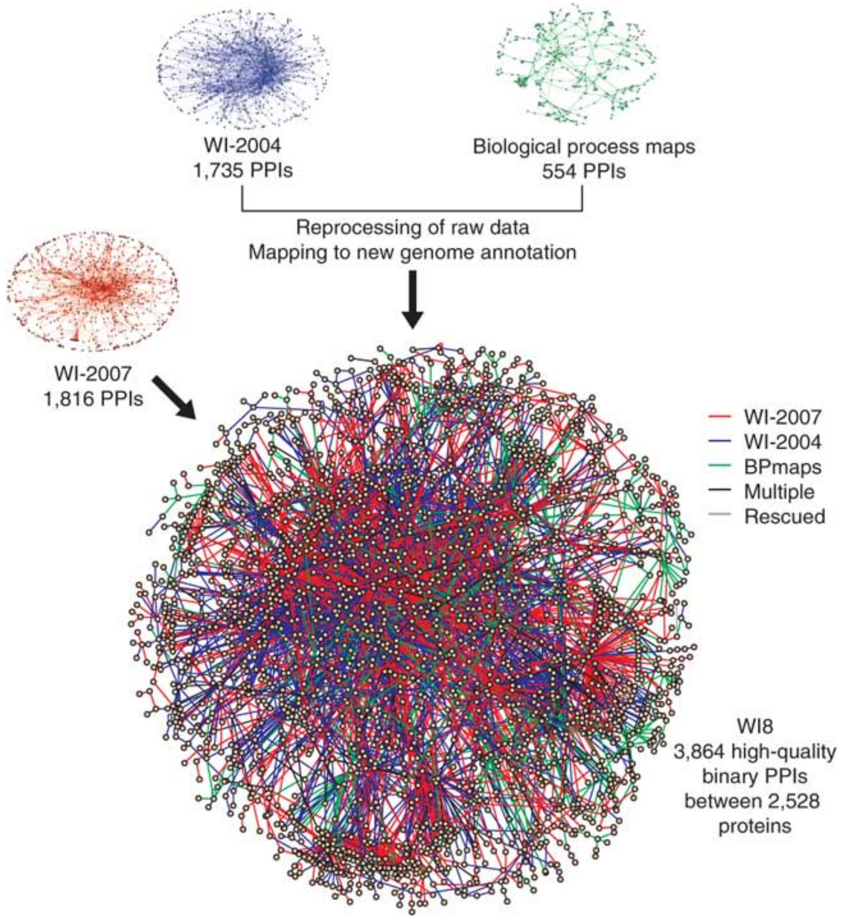


Coordinate Control of Cellular Functions: Networks of Regulators Regulating Regulators



http://systemsbiology.case.edu/projects/Signal%20Transduction.shtml

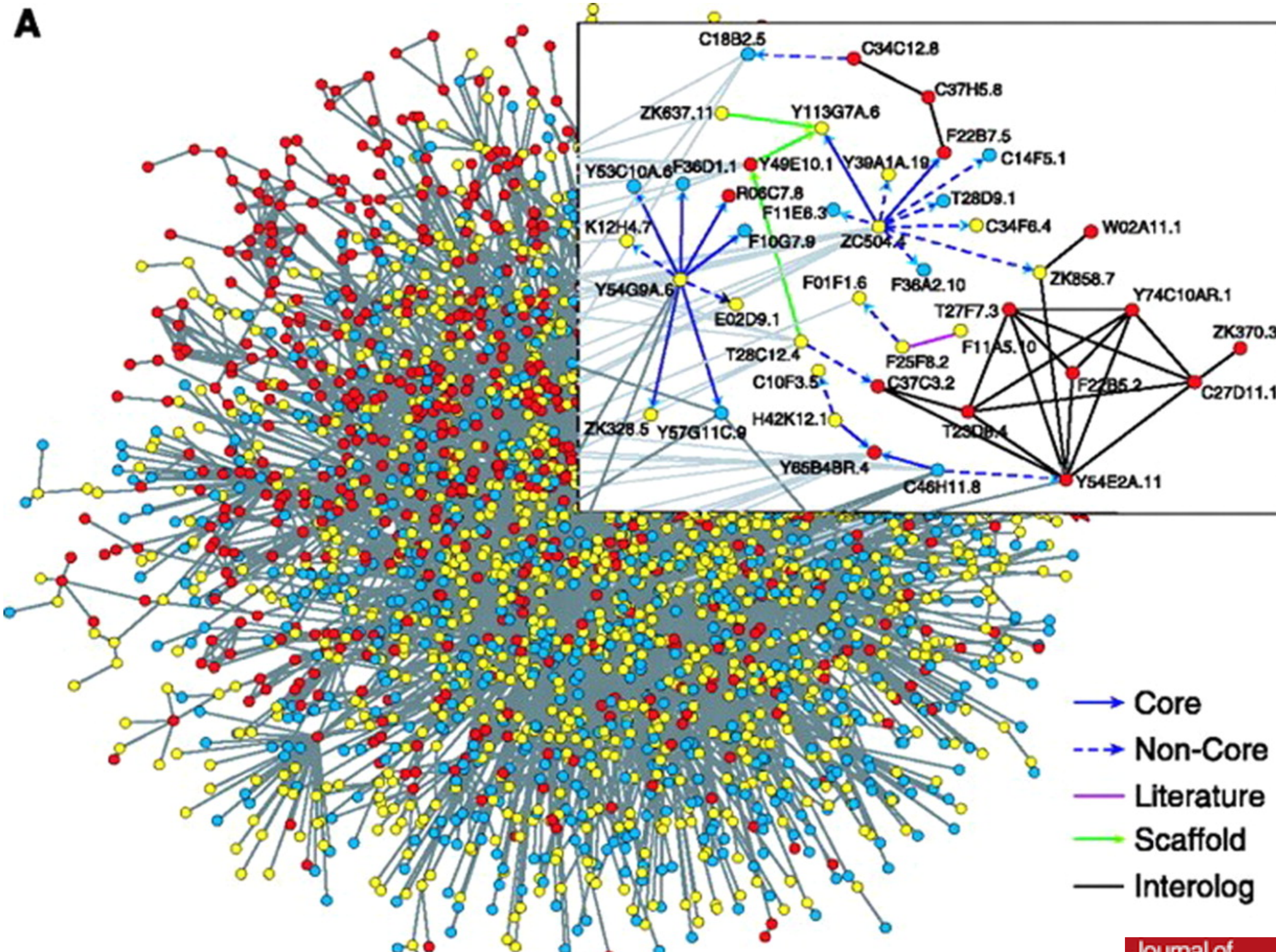
C. Elegans interactome



The color of the edge indicates the data set of origin: WI-2007, red; WI-2004, blue; biological process maps, green. Edges corresponding to more than one of these evidence types are shown in black, and edges corresponding to 'rescued' interactions—that is, supported by at least two lower-confidence pieces of evidence—in gray. Only the main giant component of the network (connected subgraph that contains the majority of the entire network's nodes) is shown.

Nature Methods **6**, 47 - 54 (2009)

C. elegans protein interaction network.

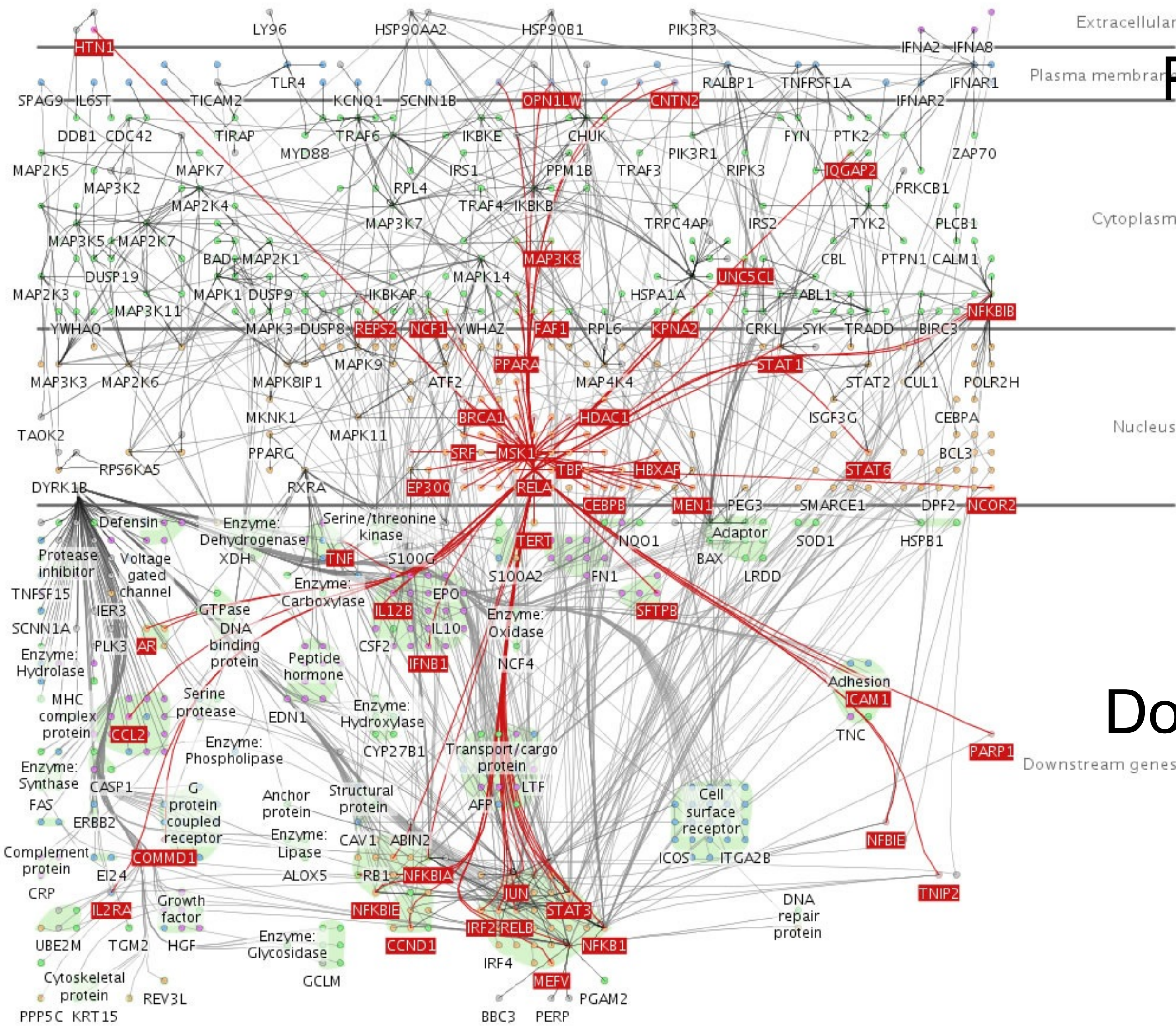


ancient, red; multicellular, yellow; and worm, blue.

Albert R J Cell Sci 2005;118:4947-4957

Journal of Cell Science
jcs.biologists.org

Networks including cell location



Extracellular

Plasma membrane

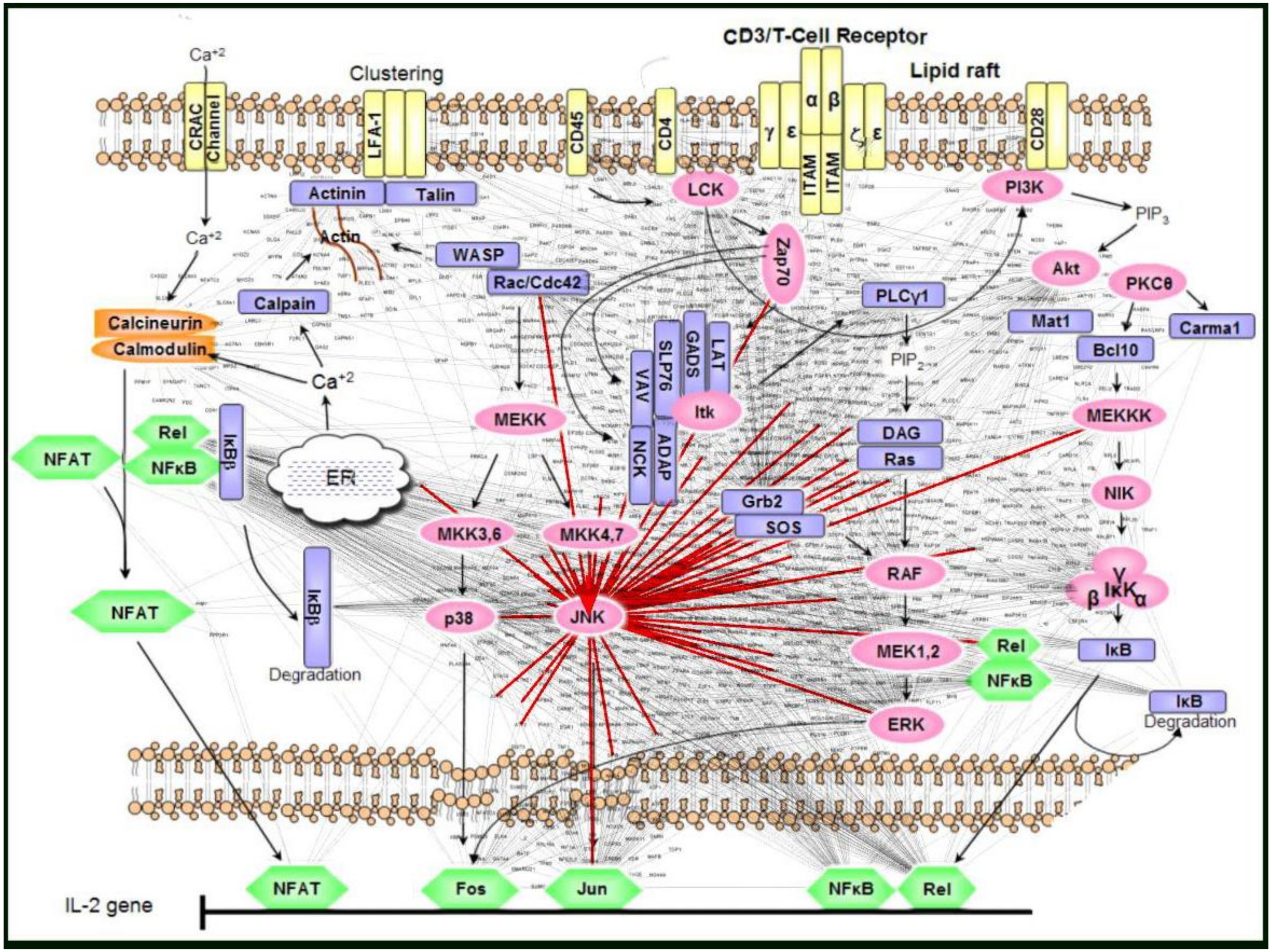
Cytoplasm

Nucleus

Downstream genes

- Cerebral plug-in, Cytoscape

Cell location + schematic shape

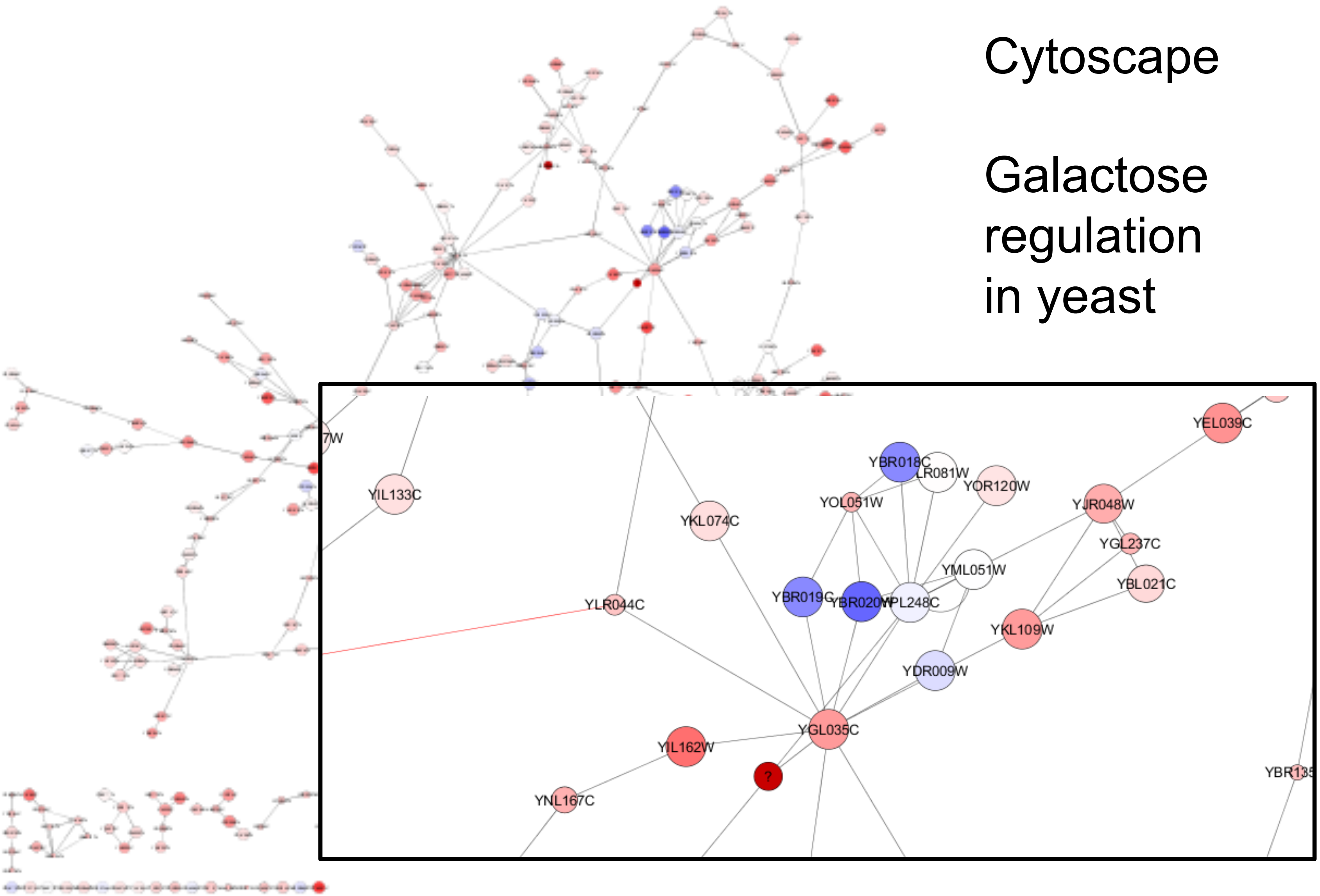


- http://vis.cs.brown.edu/areas/projects/proteins.html

PPI + expression + significance

Cytoscape

Galactose regulation in yeast

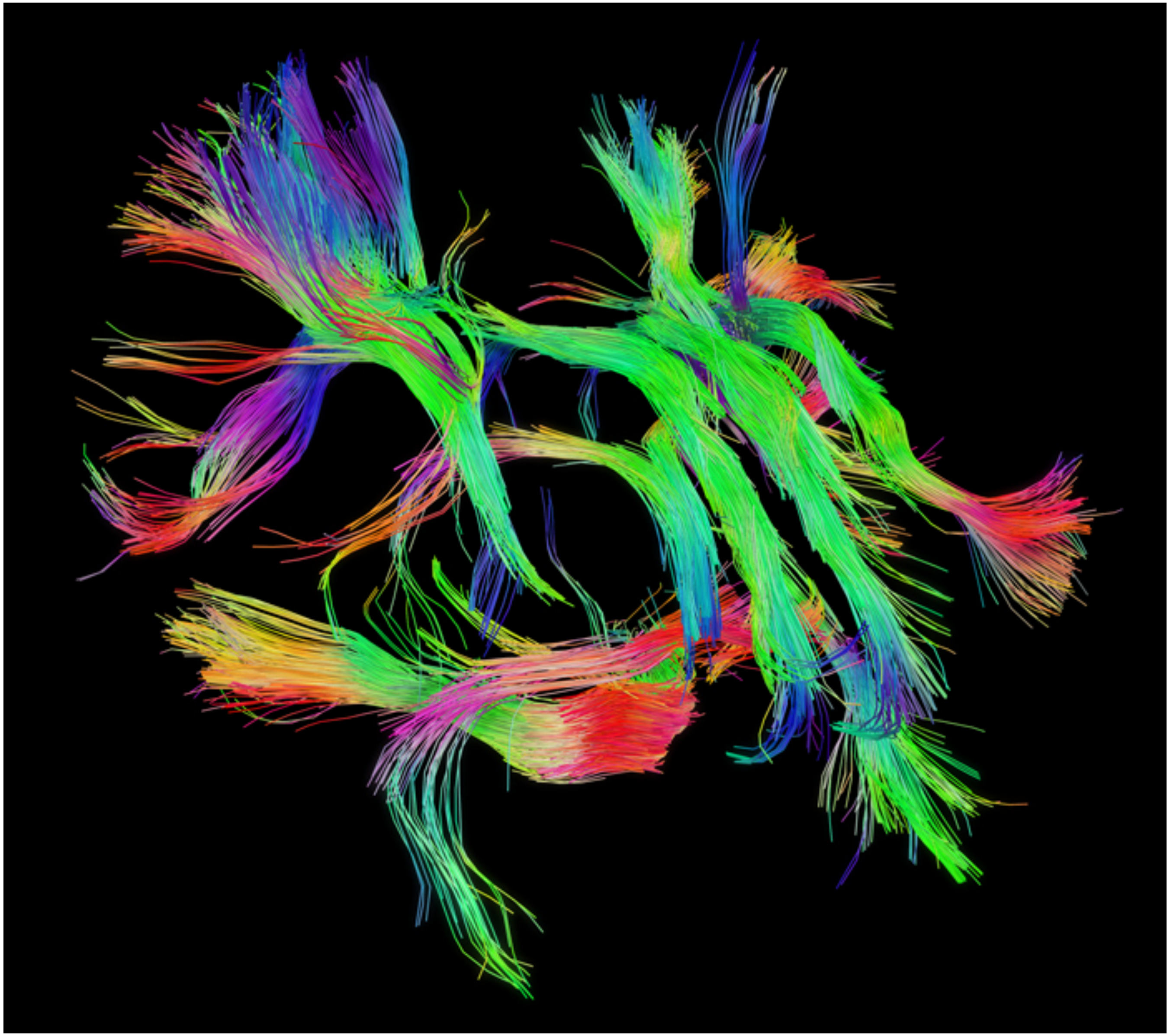


Force-directed graphs (Wikipedia)

- Force-directed graph drawing algorithms assign forces among the set of edges and the set of nodes of a graph drawing. Typically, spring-like attractive forces based on Hooke's law are used to attract pairs of endpoints of the graph's edges towards each other, while simultaneously repulsive forces like those of electrically charged particles based onCoulomb's law are used to separate all pairs of nodes. In equilibrium states for this system of forces, the edges tend to have uniform length (because of the spring forces), and nodes that are not connected by an edge tend to be drawn further apart (because of the electrical repulsion). Edge attraction and vertex repulsion forces may be defined using functions that are not based on the physical behavior of springs and particles; for instance, some force-directed systems use springs whose attractive force is logarithmic rather than linear.

The Human Connectome Project

- Navigate the brain in a way that was never before possible; fly through major brain pathways, compare essential circuits, zoom into a region to explore the cells that comprise it, and the functions that depend on it.
- The Human Connectome Project aims to provide an unparalleled compilation of neural data, an interface to graphically navigate this data and the opportunity to achieve never before realized conclusions about the living human brain.
- http://www.humanconnectomeproject.org/



White matter fiber architecture from the Connectome Scanner dataset. The fibers are color-coded by direction: red = left-right, green = anterior-posterior, blue = ascending-descending (RGB=XYZ). www.humanconnectomeproject.org

C. elegans connectome

http://wormwiring.org/

http://wormwiring.org/data/dataTools/neuronPartners.html

Simulation

Questions biologists ask

- **What are the principles governing biological systems?**
- How is something (a gene, an organism) evolutionarily related to something else?
- How does something (molecule, protein complex, cell, organism) interact with something else?

- “I began to view the simulations as an extension of my brain,” Couzin says. “By allowing the computer to help me think, I could develop my intuition of how these systems worked.”

- Science, in general, is a lot better at breaking complex things into tiny parts than it is at figuring out how tiny parts turn into complex things.

- Ed Yong, How the Science of Swarms Can Help Us Fight Cancer and Predict the Future. Wired. 3/19/13.

Why Use Models?

"Models let us put the information that we have together in a rational, orderly way to make predictions about the future. Without them, all we can do is guess."

-- Donald DiAngelis, ecologist, U Miami

"A quantitative model is a tool that has to fit an experimental study, and the model's value should be judged not by how complex and detailed it is, but by what could be learned from it."

-- Mogilner et al. (2012). Science 336: 175.

The Simulation Cycle

Collect facts
Observation
Experiments

↘

Build/modify
simulation based
on hypotheses
you believe
explain data

←

Run simulation
Make predictions

↗

Types of models

- ODEs: changes in concentration over time
- PDEs: changes in concentration over time and space
- Stochastic simulations: molecules as random numbers
- Agent-based simulations: tracking individuals
- Boolean networks
- Bayesian models
- Network analysis

Mogilner et al. (2012) Science 336: 175-179.

Modeling collective animal behavior

- http://www.wired.com/2013/03/powers-of-swarms/all/
- Iain Couzin
- Models with simple rules for agents that results in complex behavior of groups
- Visualization piece of the model is critical to see if agents are ‘acting like’ the real organisms (biologists love visual representations)

Locusts and swarms

How does a small, chaotic group of stupid locusts turn into a cloud of millions, united in one purpose?

At a certain density, the bugs would shift to cohesive, aligned clusters. And at a second critical point, the clusters would become a single marching army.

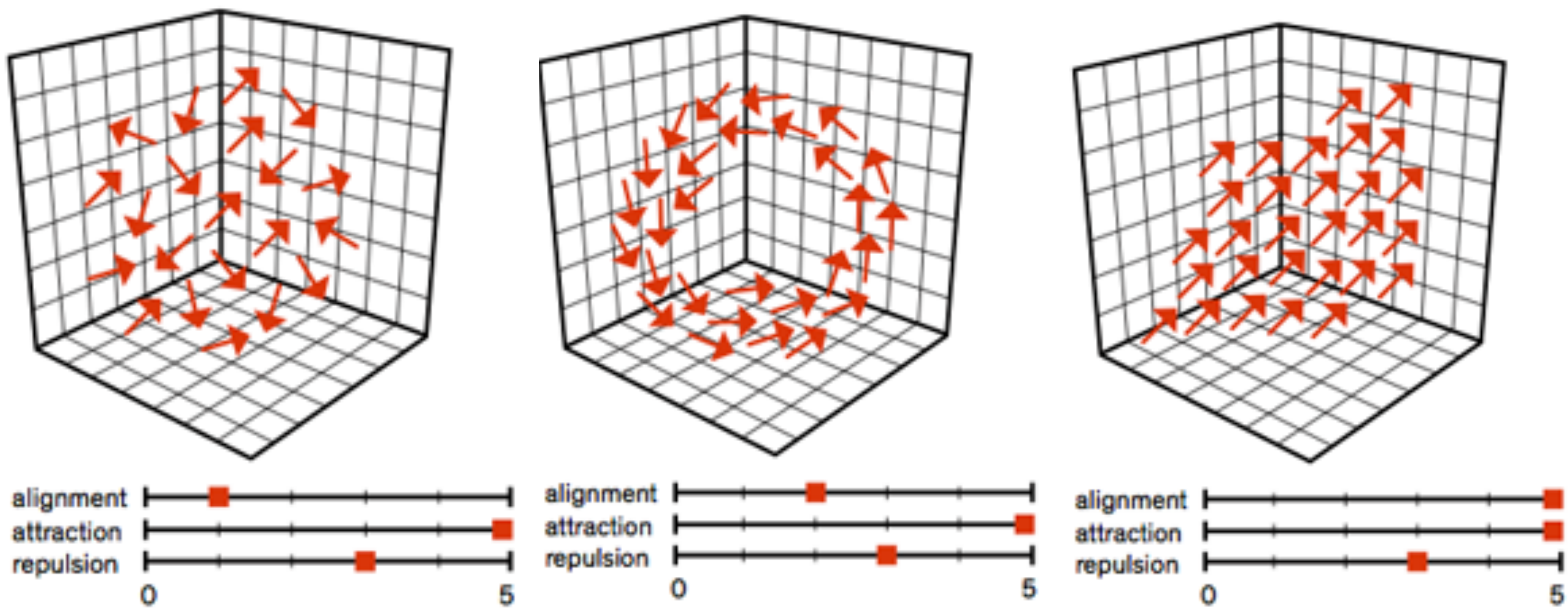
- The locusts were biting each other if they got too close.
- Cannibalism, not cooperation, was aligning the swarm.
- Couzin followed up on a **prediction** of the model: if you cut the nerve in the abdomen that lets locusts feel bites from behind, you completely remove their capacity to swarm – **verifying** the model

- Behavior that seems impossibly complex can have disarmingly simple foundations.

Boids: Three simple rules

- Move toward the average position of your neighbors (attraction)
- Keep some distance from them (repulsion)
- Align with their average heading (alignment)

Emergent behavior



NOT the wisdom of crowds

- Shiners congregate (‘hide out’) in dark patches
- Do they search out darkness and tell each other where to find it?
- Shiners slow down when they hit dark patches
- When a disorganized group of shiners hits a dark patch, fish on the edge decelerate and the entire group swivels into darkness.
- None of the shiners are purposefully swimming toward anything. The crowd has no wisdom to cobble together.

Conclusions

- Biology is complex and noisy!
- Visualizations are incredibly important and useful tools to analyze biological data
- Lots of ongoing visualization work
- Talk with domain experts when designing your visualizations