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?
• 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 appearances.

© 2008 Nature Education All rights reserved.


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
The Evolution of Synapses

Greater signalling complexity, MASC gene family expansion, NMDA receptor duplication, MAGUK duplication

Stargazin, LIMK

Excitatory glutamate receptors (NMDA and AMPA), kainate receptors, K+ channels, neurelin, CASK, erbin

GABA receptors, metabotropic glutamate receptors, CaMKII, KIR channel, NOS, SynGAP, S-SCAM, Homer, GkAP, GRIP, CRIP1, agrin, MuSK, ankyrin, neurexin, NCAM

Cadherins, ephrin receptors, receptor and non-receptor tyrosine kinases, Dlg (MAGUK), Shank, calpain, spectrin, PDZ binding proteins

MASC downstream signaling components, PKC, PMCA, NF1, calmodulin, calcineurin

Nature Reviews | Neuroscience
Hierarchically nested subgroups

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).
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.
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

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

Data are variable!
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?
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.
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 on Coulomb'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/

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.

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.”

The Simulation Cycle

Collect facts
Observation
Experiments

Run simulation
Make predictions

Build/modify simulation based on hypotheses you believe explain data
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

Modeling collective animal behavior

- [http://www.wired.com/2013/03/powers-of-swarms/all/](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