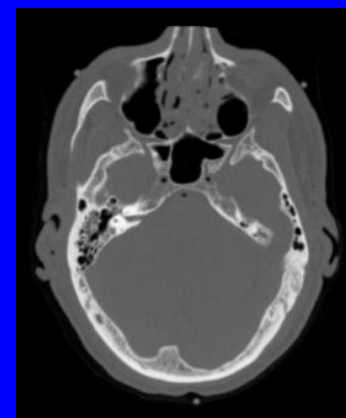
Mathematics Is Biology’s Next Microscope, Only Better; Biology Is Mathematics’ Next Physics, Only Better

Joel E. Cohen
Laboratory of Populations
Rockefeller & Columbia Universities

10 July 2003

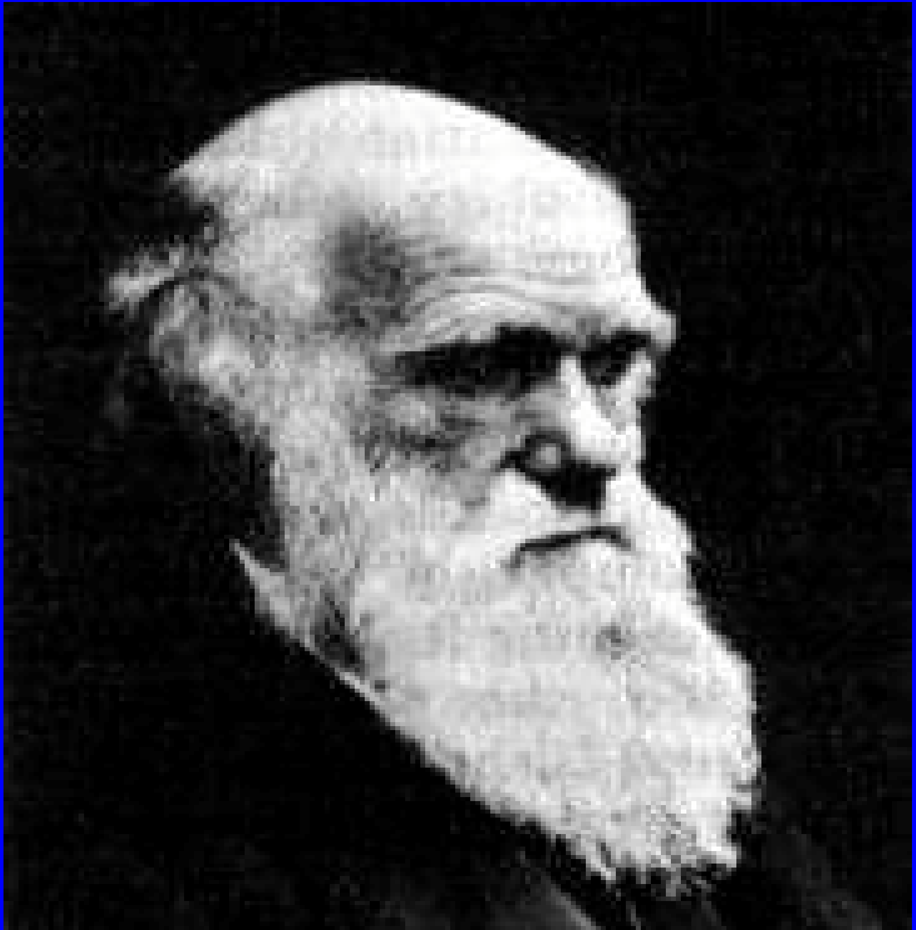
Mathematics Is Biology's Next Microscope, Only Better



- Microscope (late 17th C.) caused revolution in biology by revealing otherwise invisible, & previously unsuspected, worlds.
- Mathematics (broadly interpreted) can reveal otherwise invisible worlds in all kinds of data, not only optical.

"Mathematics seems to endow one with something like a new sense."

Charles Darwin



Biology Is Mathematics' Next Physics, Only Better

Newton

Mendel

- Physics stimulated enormous advances in mathematics, e.g. geometry, calculus.
- Biology can stimulate creation of new realms of mathematics.
- Is living nature qualitatively more heterogeneous than non-living nature?

Outline

- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
- Future
  - potential problems
  - opportunities

Outline

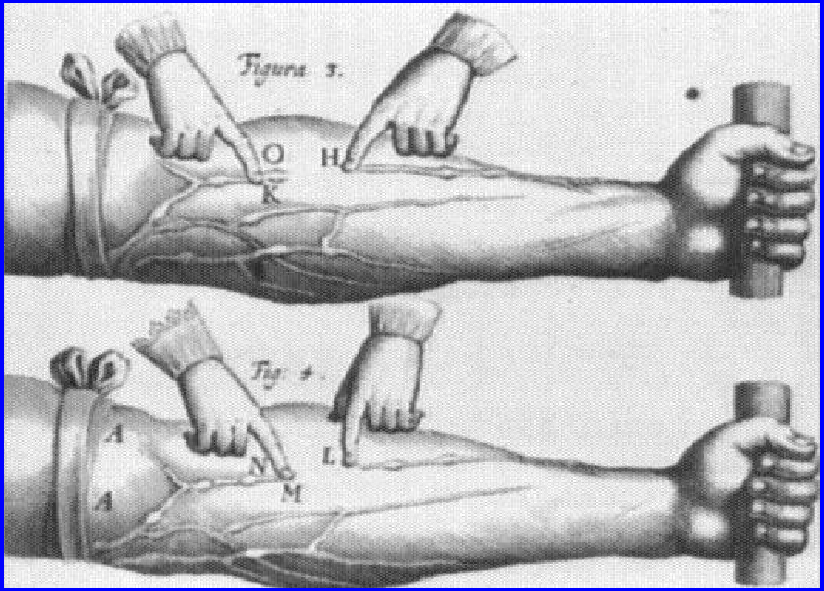
- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
- Future
  - potential problems
  - opportunities

Biology since 1500

- External factor: Columbian exchange
  (New World species, foods, diseases)
- Technical progress
  - quantitation (Harvey 1615 [1628])
  - microscope (Leeuwenhoek 1660-1700)
  - chemistry (Liebig 1855)
- Conceptual progress
  - anatomical observation (Vesalius 1543)
  - cell theory (Schleiden Schwann Virchow Weissman 1838-80)
  - evolution (Darwin Wallace 1859)
  - genes (Mendel Correns von Seysenegg deVries 1865-1900)

William Harvey (1578-1659)

Exercitatio Anatomica De Motu Cordis et Sanguinis In Animalibus (1628)



Harvey showed that blood circulates

- Galen (2d C.): blood ebbs & flows, pumped by arteries; heart is passive.
- Harvey (1615): heart & veins have 1-way valves; flow is unidirectional.
- Left ventricle (dead) holds 2 oz. ~ 60 ml.
- >1/8-1/4 of blood is expelled per stroke.
- Heart beats 60-100 times/minute.
- $\therefore$60ml x 1/8 x 60 beats/min x 60 = 27 l/h.
- Average human has 5.5 l blood.

[Marcello Malpighi (1628-94) saw capillaries.]

Outline

- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
- Future
  - potential problems
  - opportunities

Mathematics since 1500

- Geometry & topology
  - analytic geometry (Descartes 1637)
  - non-Euclidean geometries (1823-1830)
- Algebra
  - roots of equations, group theory, symmetry
  - linear algebra (19th-20th C.)
- Analysis: *modern mathematical thought*
  - calculus: theory of limits (Newton 1666 Leibniz)
  - probability (Pascal Fermat 1654 DeMoivre Laplace Gauss 1733-1809 normal curve)
- External factor: computers (war, business)

Mathematics arising from biological problems (1)

- Age structure of stable populations (Euler 1760)
- Logistic equation for limited population growth (Verhulst 1838)
- Branching processes, extinction of family names (Galton 1889)
- Correlation (K. Pearson 1903)

Mathematics arising from biological problems (2)

- Markov chains, statistics of language (Markov 1906)
- Hardy-Weinberg equilibrium (1908)
- Analysis of variance, design of agricultural experiments (Fisher 1920s)
- Dynamics of interacting species (Lotka 1922 Volterra 1926-37)

Mathematics arising from biological problems (3)

- Birth process (Yule 1925), birth and death process (D.G. Kendall 1948)
- Traveling waves in genetics (Fisher; Kolmogorov Petrovsky Piscounov 1937)
- Game theory (von Neumann 1944)
- Distribution for estimating bacterial mutation rates (Luria Delbruck 1943)
- Morphogenesis (Turing 1952)

Mathematics arising from biological problems (4)

- Diffusion equation for gene frequencies (Kimura 1954)
- Circular interval graphs, genetic fine structure (Benzer 1959)
- Threshold functions of random graphs, models of communication networks or “even of organic structures of living matter” (Erdös-Rényi 1960)

Mathematics arising from biological problems (5)

- Sampling formula for haplotype frequencies (Ewens 1972)
- Coalescent, genealogy of populations (Kingman 1982)

Outline

- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
- Future
  - potential problems
  - opportunities

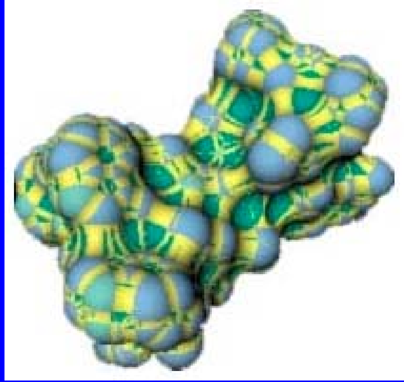
The biological landscape

Questions

- Structure(s)
  - How is it built?
- Function(s)
  - How does it work?
- Pathology(ies)
  - What goes wrong?
- Repair(s)
  - How is it fixed?
- Origin(s)
  - How did it begin?

Domains

- Molecules
- Cells
- Tissues
- Organs
- Individuals
- Populations
- Communities, ecosystems
- Biosphere



The applied mathematical landscape

- Data structures
- Algorithms
  - analyze data
  - analyze models
- Theories & models, including all pure math
  - analyze data
  - analyze ideas
- Computers & software
  - embody mathematical knowledge
  - interface with humans (vision, speech)
  - compute

The landscape of biology and mathematics

The landscape of research in mathematics and biology contains all combinations of a problem from the matrix of biological problems and problem areas from applied mathematics.

Outline

- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
    - gene expression & molecular pharmacology
    - food webs, body size & abundance
- Future
  - potential problems
  - opportunities

Example 1: integrate gene expression & molecular pharmacology

- Scherf U, 16 others. A gene expression database for the molecular pharmacology of cancer. Nature Genetics 2000 Mar; 24(3):236-44.

Results

- "used cDNA microarrays to assess gene expression profiles in 60 human cancer cell lines of the National Cancer Institute's drug discovery program."
- "link the bioinformatics with chemoinformatics by correlating gene expression and drug activity patterns in the 60 cell types. Clustering the cells on the basis of gene expression yields a picture very different from that obtained when the cells are clustered on the basis of their response to drugs."

Data http://discover.nci.nih.gov/arraytools/

Drug activity matrix A

1400 drugs x 60 cell lines from human cancers of various organs (including 118 drugs with "known mechanism of action")

a_{dc} = activity of drug d in suppressing growth of cell line c

a_{dc} = sensitivity of cell line c to drug d

Gene expression matrix T ("T" for target)

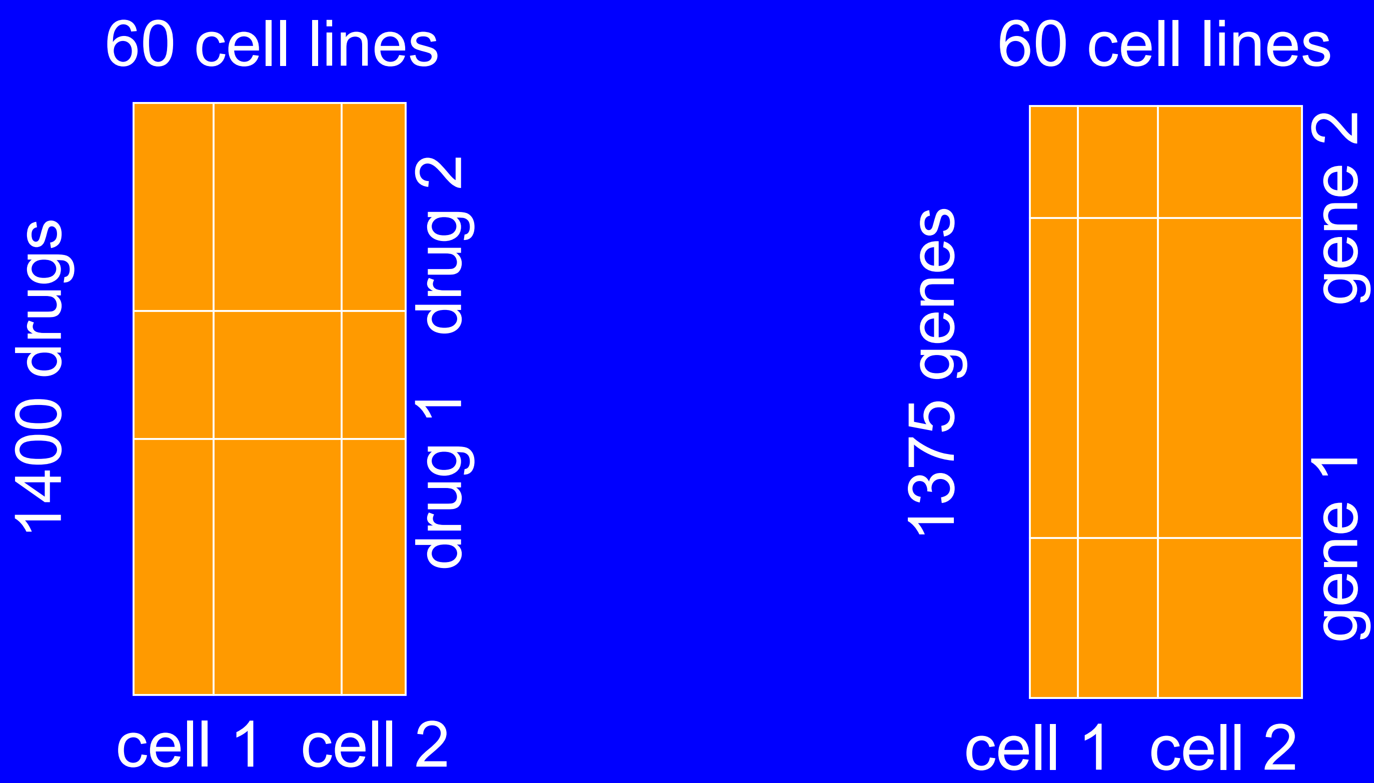
1375 genes x same 60 cell lines

t_{gc} = relative abundance of mRNA transcript of gene g in cell line c

t_{gc} = cell line c's expression of gene g

Correlation & clustering

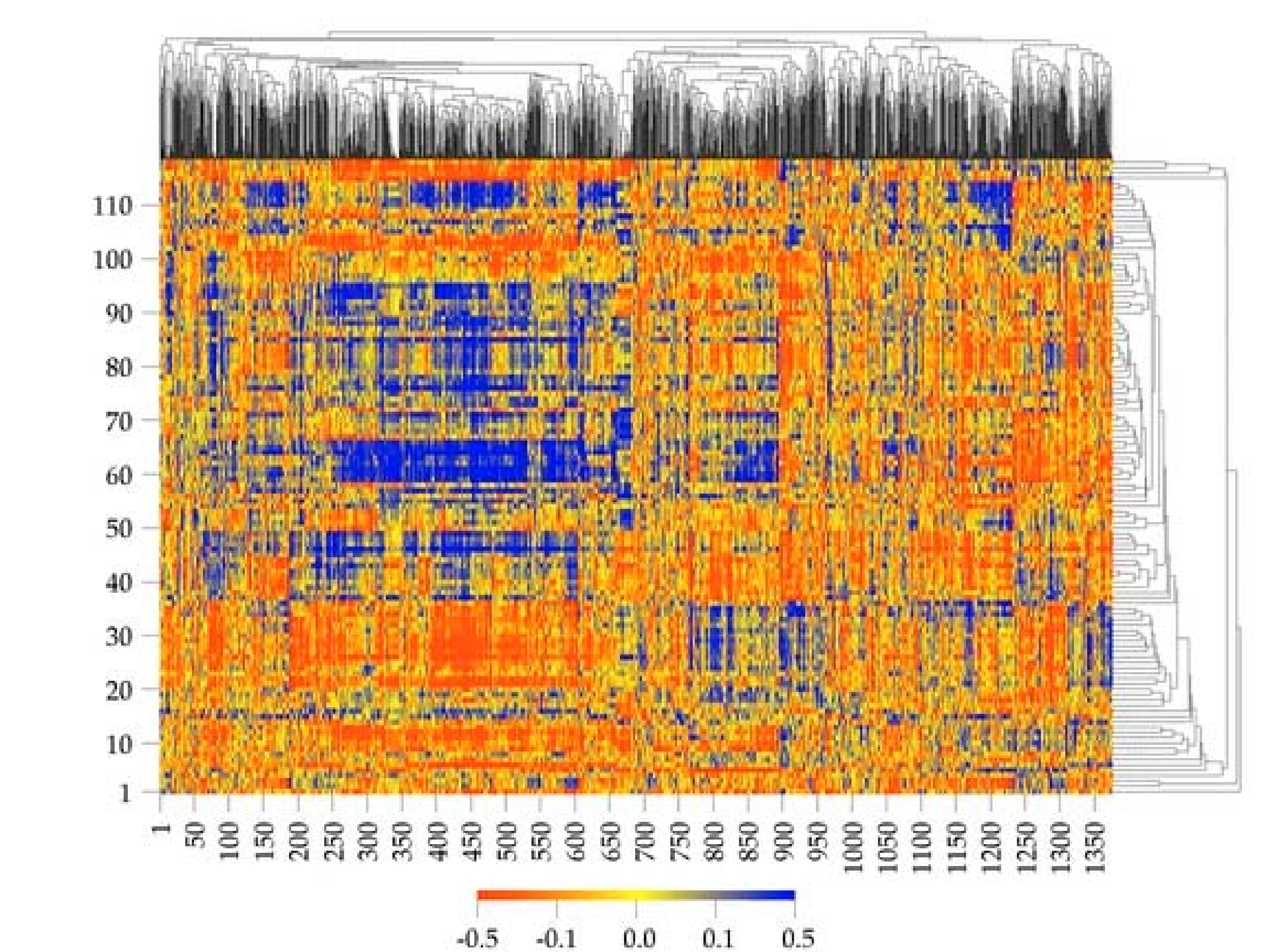
A = activity matrix T = expression matrix



A(cell 1, cell 2). T(cell 1, cell 2). Correlate & cluster. Different.
A(drug 1, drug 2). T(gene 1, gene 2). Correlate & cluster.
(A drug d, T gene g). Correlate.

Clustered image map

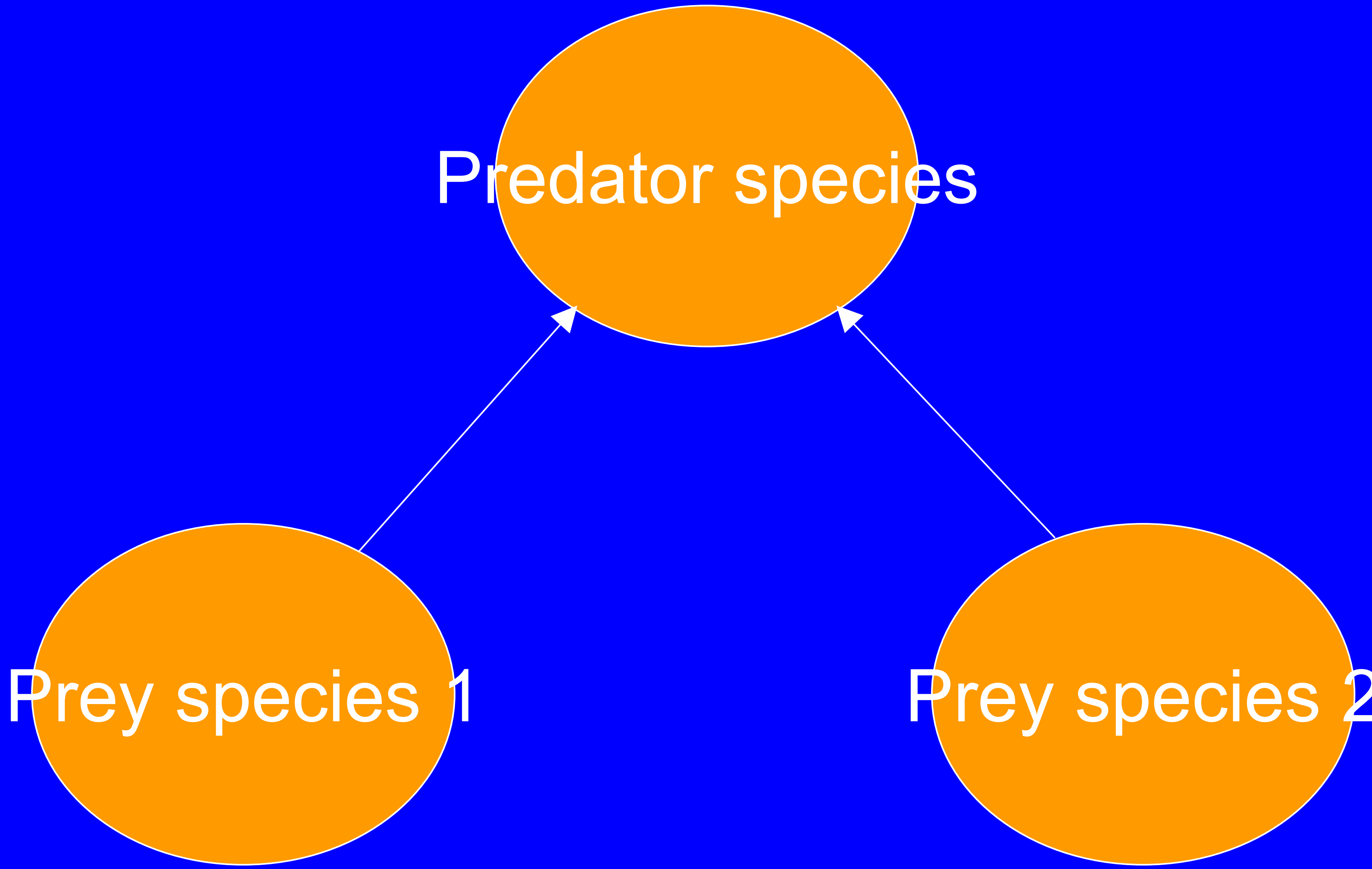
gene expression-drug activity correlations plotted as a function of clustered genes (x-axis) and clustered drugs (y-axis)



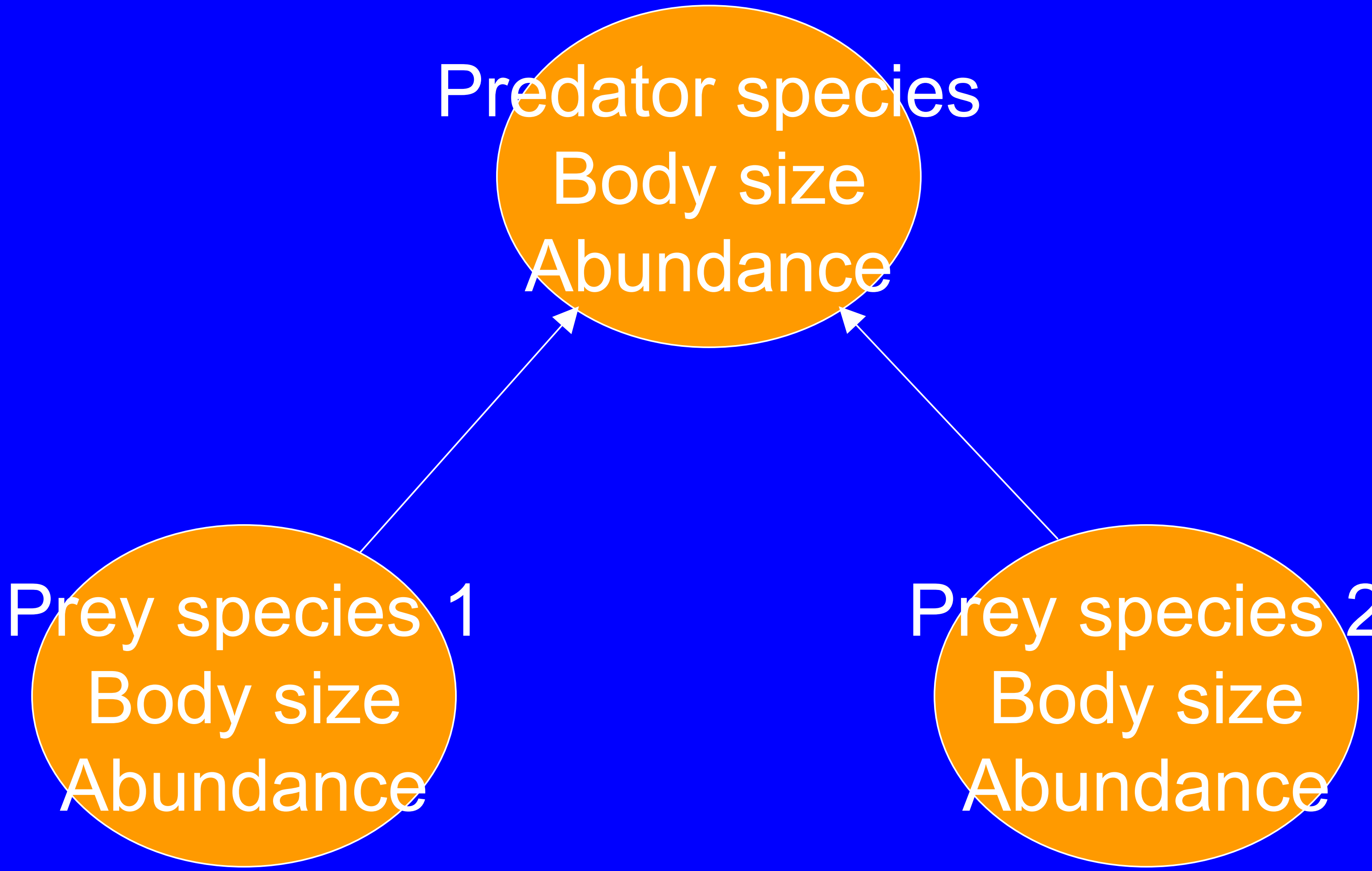
Example 2: integrate food webs (attribute of ecological communities) with body size (attribute of individuals) & abundance (attribute of populations)

Joel E. Cohen, Tomas Jonsson, Stephen R. Carpenter, Ecological community description using the food web, species abundance, and body size. *Proc. National Acad. Sci. USA* 100(4):1781-1786, 18 February 2003

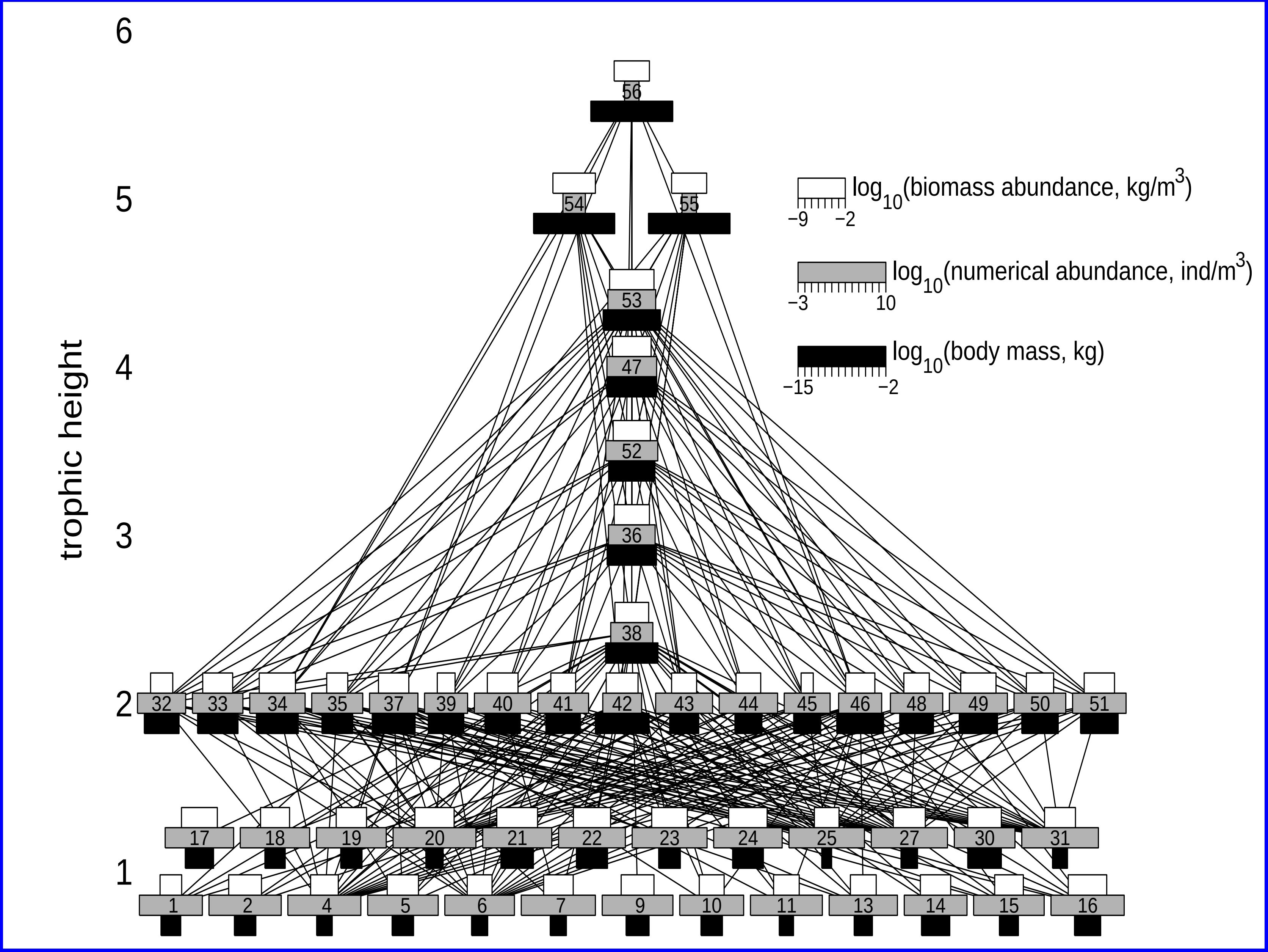
Old: Food web with feeding only



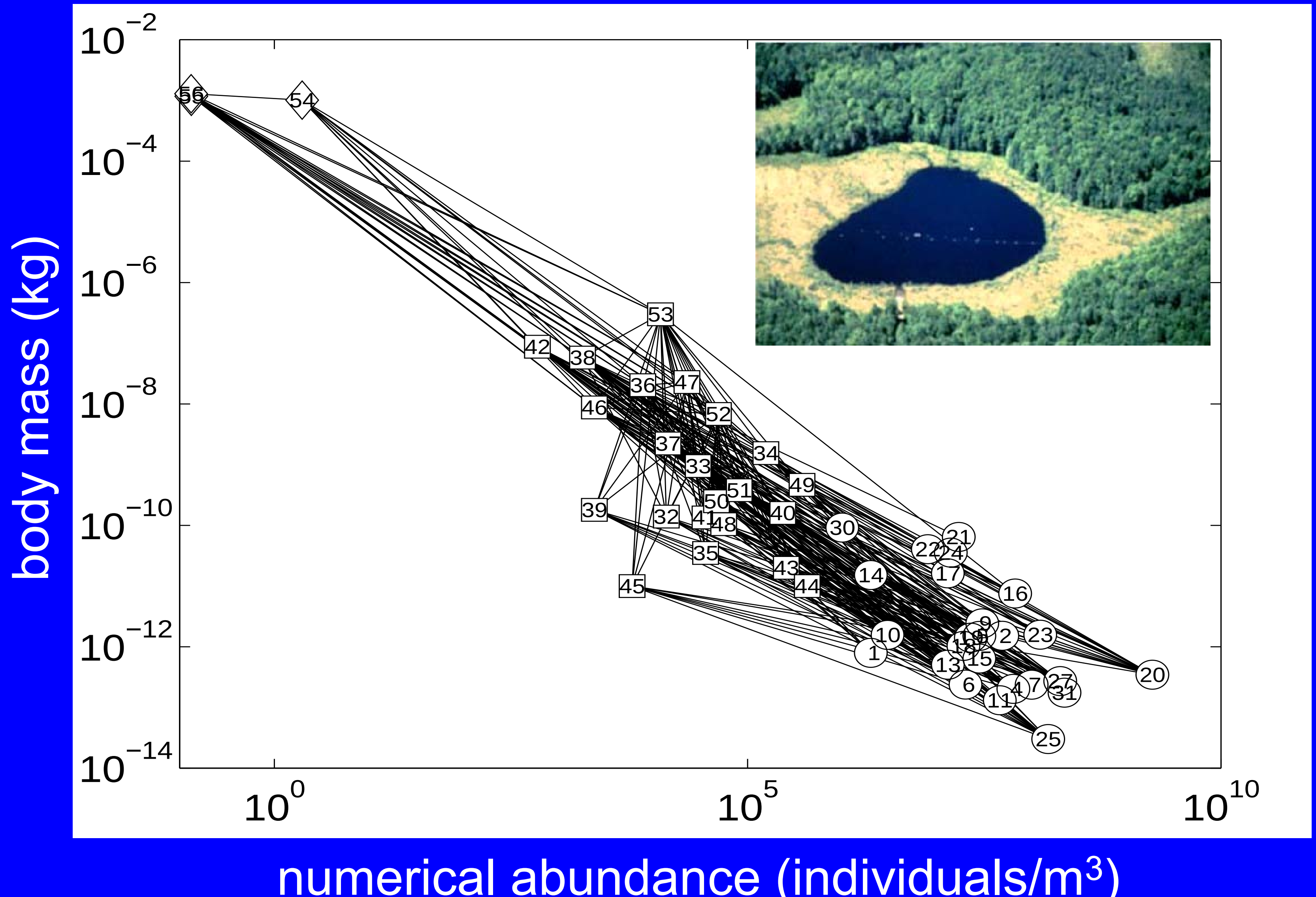
New: Food web, body size, abundance



Tuesday Lake 1984



Food web, body mass & numerical abundance 1984

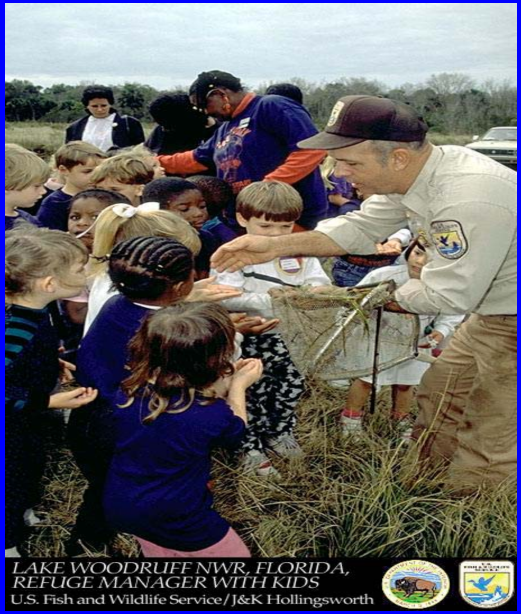


Outline

- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
- Future
  - potential problems
  - opportunities

Potential problems

- Educating scientists
- Intellectual property
- National security



Potential problems (1) educating scientists

- K-16 education in math, science
  - U.S. math & science education falls behind that of other industrial nations.
- graduate, postdoc cross-training
- diversity: gender, ethnicity, nationality (visas)
- educating peer reviewers to approve exploratory research of high quality

Potential problems (2) intellectual property

- Science as a potlatch culture
- Bayh-Dole 1980
  - Rai & Eisenberg, *Amer. Scientist* 91:52, 2003
- Tragedy of anti-commons: Madey v. Duke
  - Duke Petition for Writ of Certiorari to U.S. Supreme Court: “The possibility that the patent system could stifle or even stymie the progress of biotechnology and other important fields of research is both real and profound.”

Potential problems (3) national security

- Win-win domains
  - EPA biowarfare monitoring
  - foot & mouth disease
  - smallpox inoculation strategy
- Good (openness) vs. good (defense)
- Good (privacy) vs. good (security): databases, biomarkers (SNPs)
- "Sensitive but unclassified" information
  - who does the research? (non-US?)
  - with what publication rights/obligations?

Outline

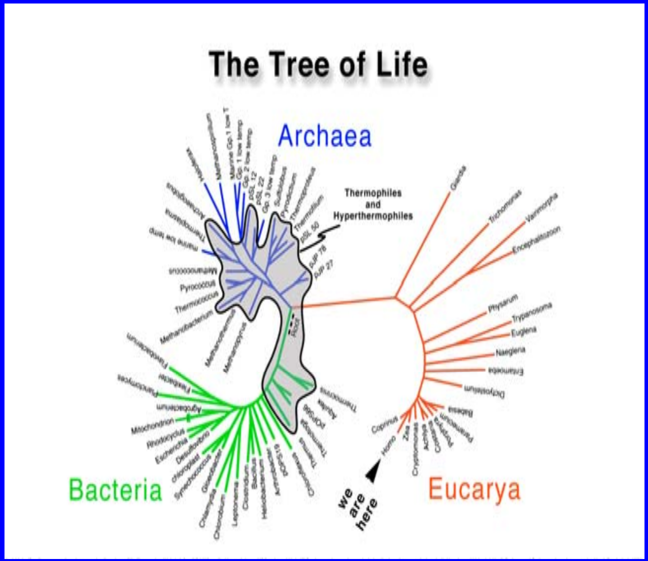
- Past
  - biology
  - mathematics
- Present
  - landscapes of biology & applied math
  - examples
- Future
  - potential problems
  - opportunities

Mathematics can help biologists grasp problems that are otherwise

- Too big
  - biosphere
- Too slow
  - macro evolution
- Too remote in time
  - early extinctions
- Too complex
  - brain
- Too small
  - molecular structure
- Too fast
  - photosynthesis
- Too remote in space
  - life at extremes
- Too dangerous or unethical
  - epidemiology of infectious agents

Biological challenges (1)

- Understand cells, their diversity within & between organisms, & their interactions with biotic & abiotic environments
- Understand brain, behavior & emotion
  - Why do or don't people have children?
- Replace tree of life by network to represent lateral transfers
  - genes
  - genomes
  - prions



Biological challenges (2)

- Couple atmospheric, terrestrial & aquatic biosphere with global physicochemical processes

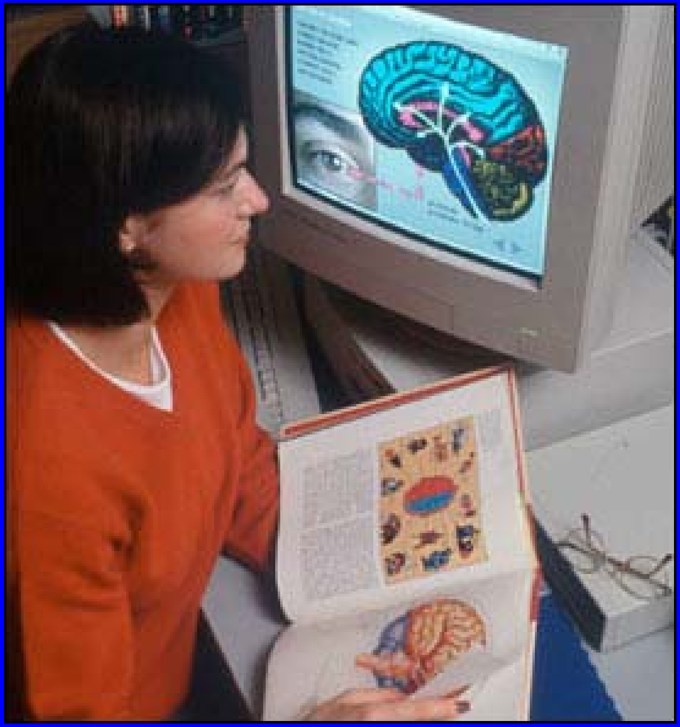
- Monitor living systems to detect large deviations
  - epidemics natural or induced
  - physiological or ecological pathologies

Mathematical challenges (1)

- Understand computation: gain insight & prove theorems from numerical computation & agent-based models
- Model multilevel systems, e.g., cells in people in human communities in physical, chemical, & biotic ecologies
- Understand uncertainty & risk by integration of frequentist, Bayesian, subjective & other theories of probability

Mathematical challenges (2)

- Understand data mining, simultaneous inference (beyond Bonferroni)
- Set standards for clarity, performance, publication & permanence of software & computational results



Thank you.

Your thoughts?

cohen@rockefeller.edu

Acknowledgments

I am grateful for helpful suggestions from Mary Clutter, Charles Delwiche, Bruce A. Fuchs, Yonatan Grad, Alan Hastings, Kevin Lauderdale, Zaida Luthey-Schulten, Daniel C. Reuman, Noah Rosenberg, Michael Pearson and Samuel Scheiner, support from U.S. National Science Foundation grant DEB 9981552, and the hospitality of Mr. and Mrs. William T. Golden during this work. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.