Biological networks

18.S995 - L18

dunkel@math.mit.edu

Examples

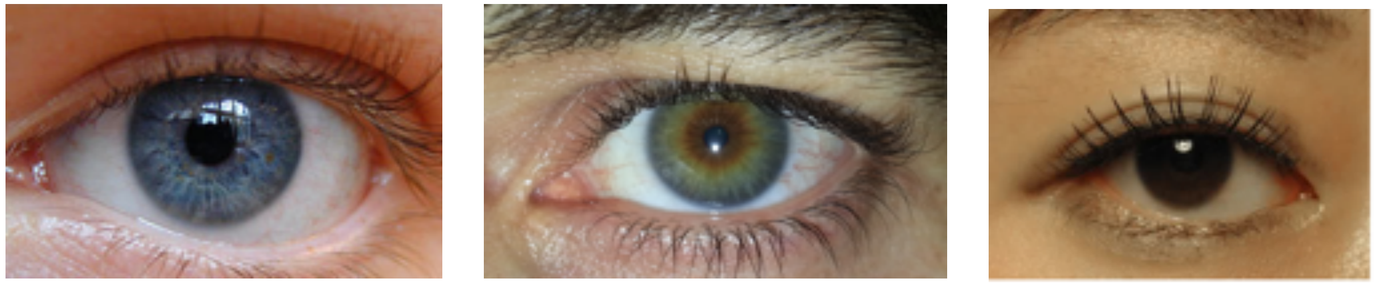
- gene regulatory networks
- neur(on)al networks
- structural networks (cytoskeleton, gels, …)
- hydrodynamic transport networks (blood vessels, slime molds, leaves, trees)
- tree structures (phylogenetic, cell lineages)
- ecological network

Gene regulatory networks

Gene facts

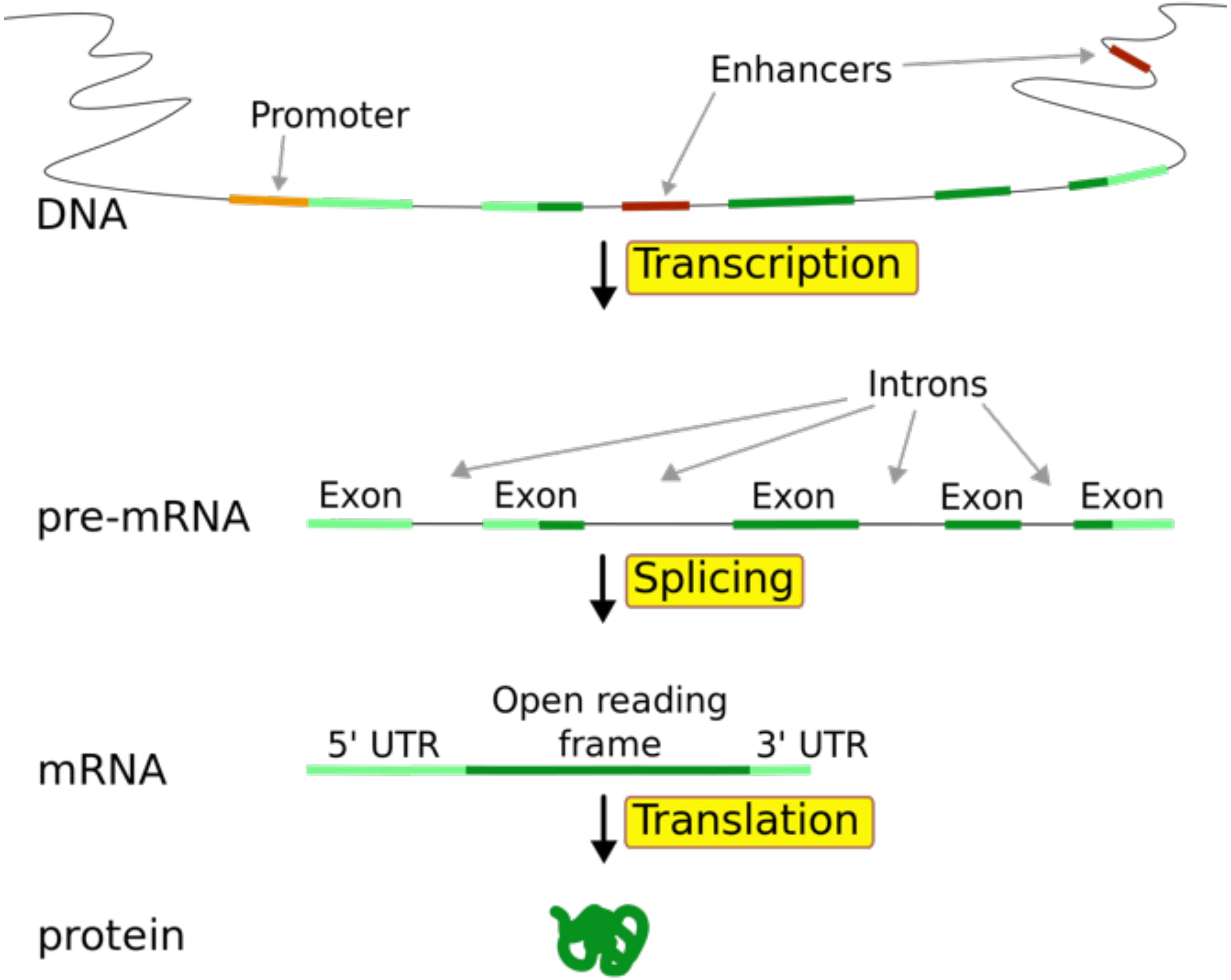
- gene = basic physical and functional unit of heredity
- 20,000-25,000 human genes, length from a few100 bp to 2 Mbp
- <1% different between humans
- alleles = alternate forms of the same gene

Courtesy: National Human Genome Research Institute



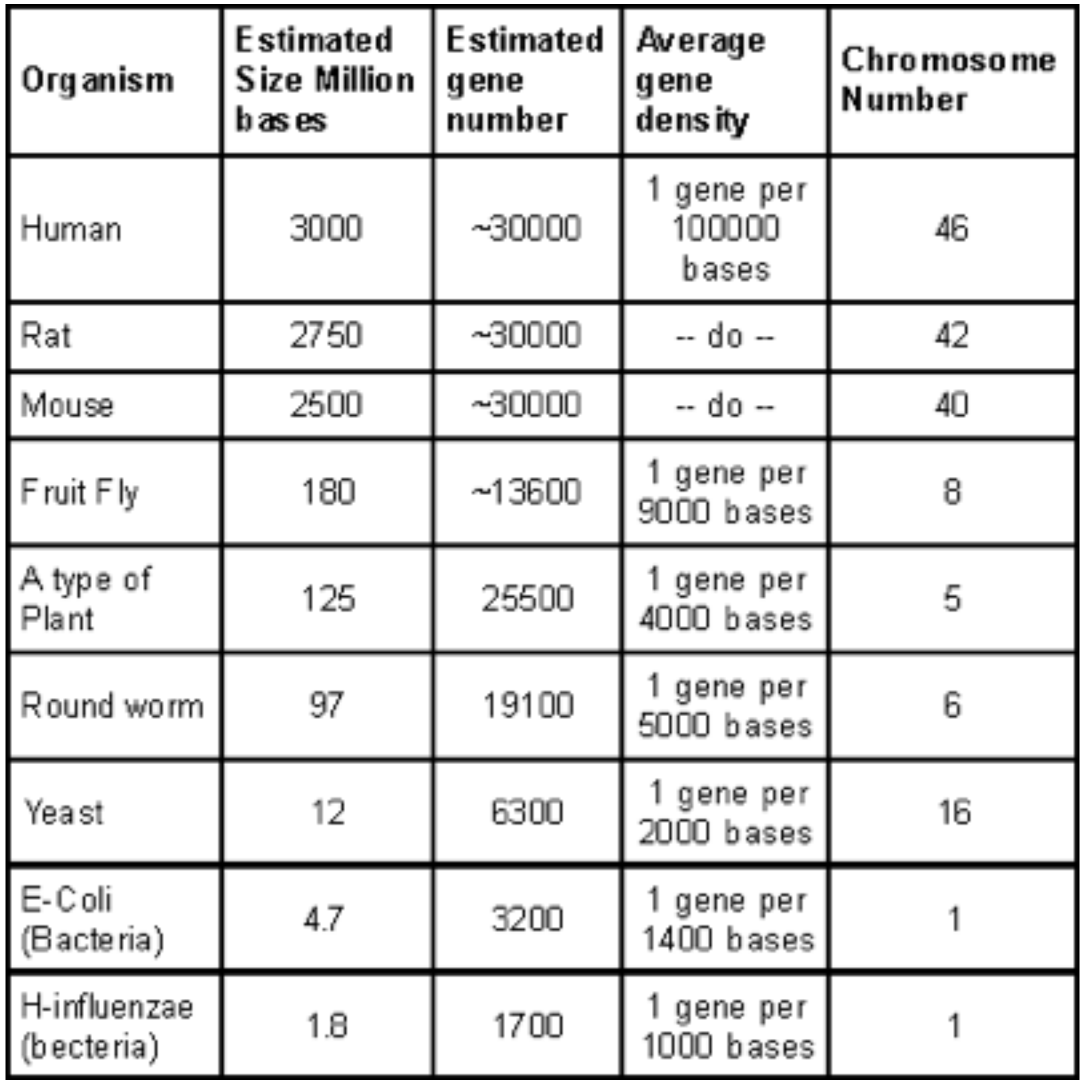
www.genome.gov

dunkel@math.mit.edu

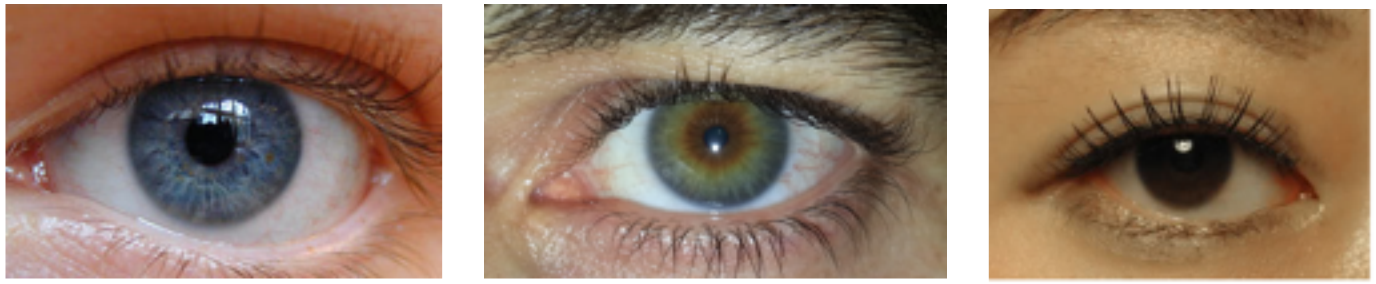

Enhancers
Promoter
DNA
Transcription
Introns
pre-mRNA
Exon
Exon
Exon
Exon
Exon
Splicing
Open reading frame
5' UTR
3' UTR
mRNA
Translation
protein

Gene facts

- gene = basic physical and functional unit of heredity
- 20,000-25,000 human genes, length from a few100 bp to 2 Mbp
- <1% different between humans
- alleles = alternate forms of the same gene



Organism	Estimated Size Million bases	Estimated gene number	Average gene density	Chromosome Number
Human	3000	~30000	1 gene per 100000 bases	46
Rat	2750	~30000	-- do --	42
Mouse	2500	~30000	-- do --	40
Fruit Fly	180	~13600	1 gene per 9000 bases	8
A type of Plant	125	25500	1 gene per 4000 bases	5
Round worm	97	19100	1 gene per 5000 bases	6
Yeast	12	6300	1 gene per 2000 bases	16
E-Coli (Bacteria)	4.7	3200	1 gene per 1400 bases	1
H-influenzae (becteria)	1.8	1700	1 gene per 1000 bases	1



dunkel@math.mit.edu

Human Genome Project

Completed in 2003, the Human Genome Project (HGP) was a 13-year project coordinated by the U.S. Department of Energy (DOE) and the National Institutes of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China, and others. Project goals were to

- *identify* all the approximately 20,500 genes in human DNA,
- *determine* the sequences of the 3 billion chemical base pairs that make up human DNA,
- *store* this information in databases,
- *improve* tools for data analysis,
- *transfer* related technologies to the private sector, and
- *address* the ethical, legal, and social issues (ELSI) that may arise from the project.

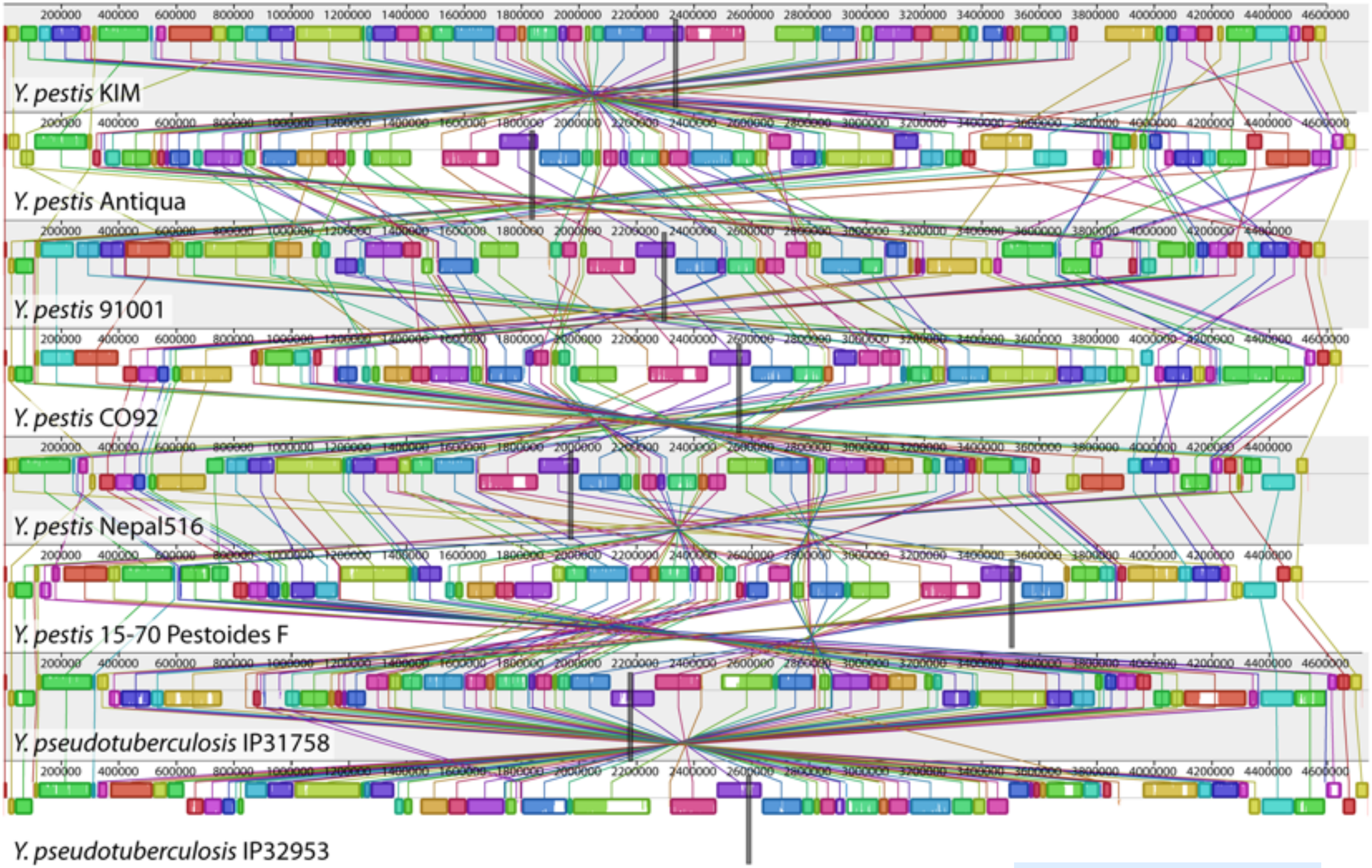
Though the HGP is finished, analyses of the data will continue for many years.

http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml

dunkel@math.mit.edu

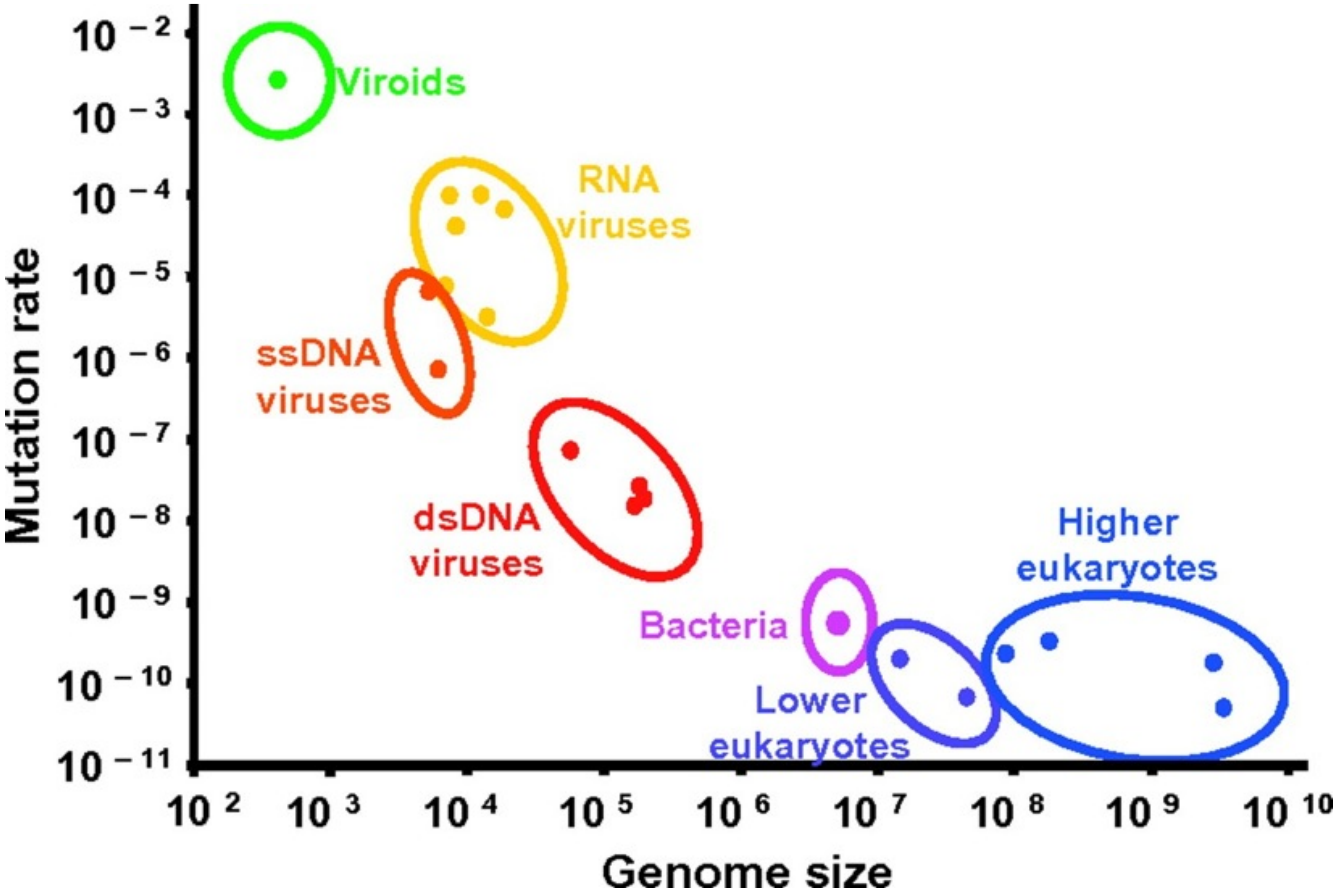
Comparative genomics

8 Yersina bacteria



PLOS Genetics **4** (7): e1000128

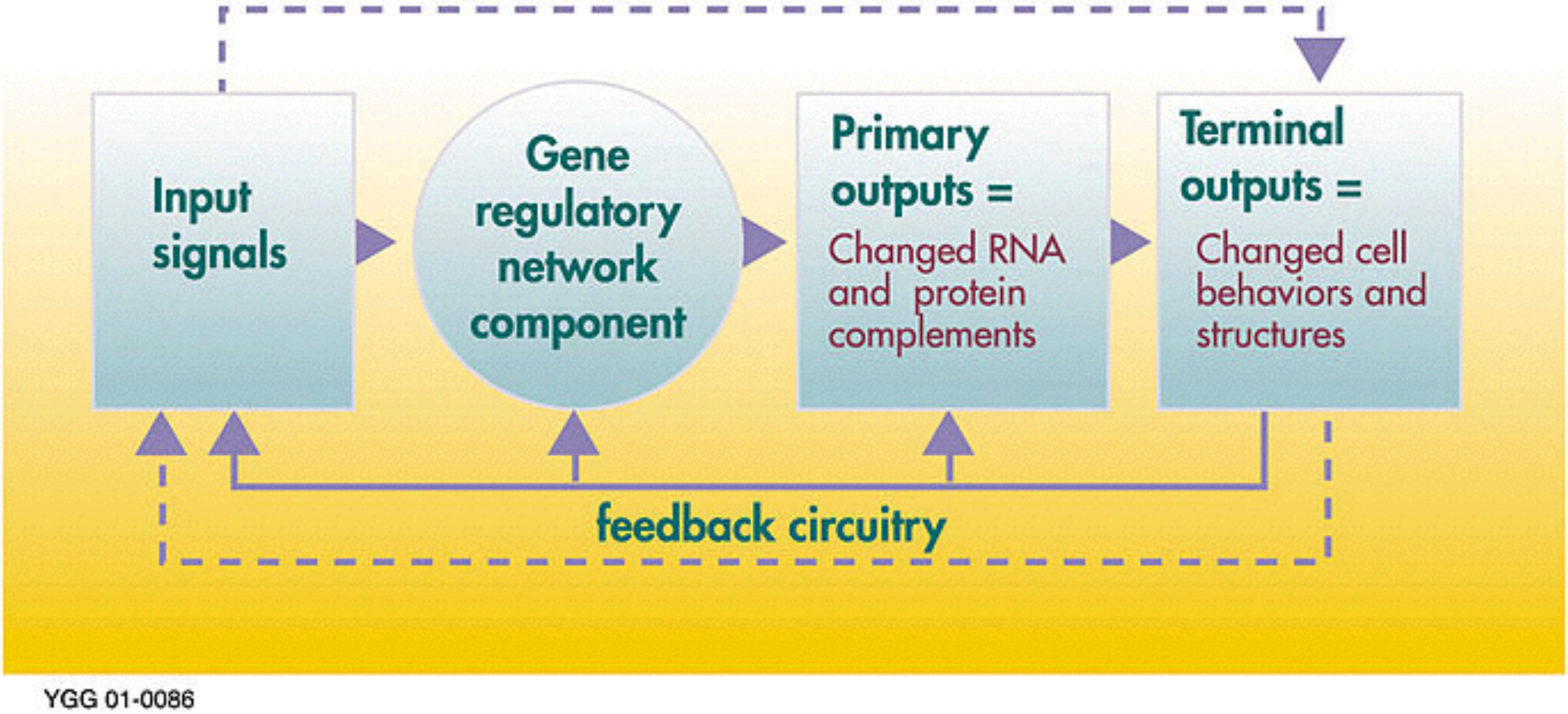
Relationship between mutation rate per nucleotide site and genome size for different genomic systems including viruses. [Reproduced with permission from ref. 19 (Copyright 2009, AAAS).].



Holmes PNAS (2010) 107:1742-1746

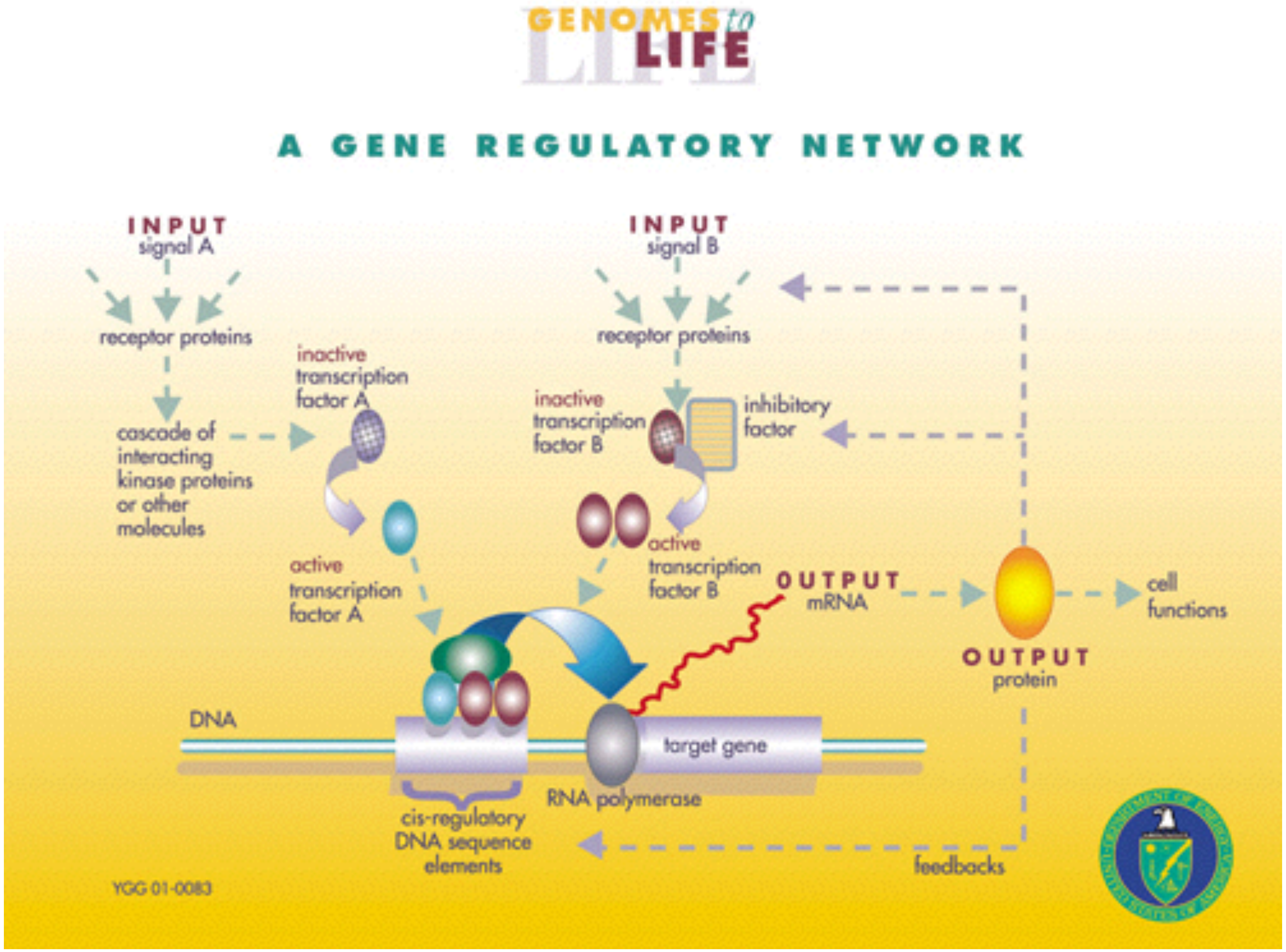
PNAS

Gene regulatory networks



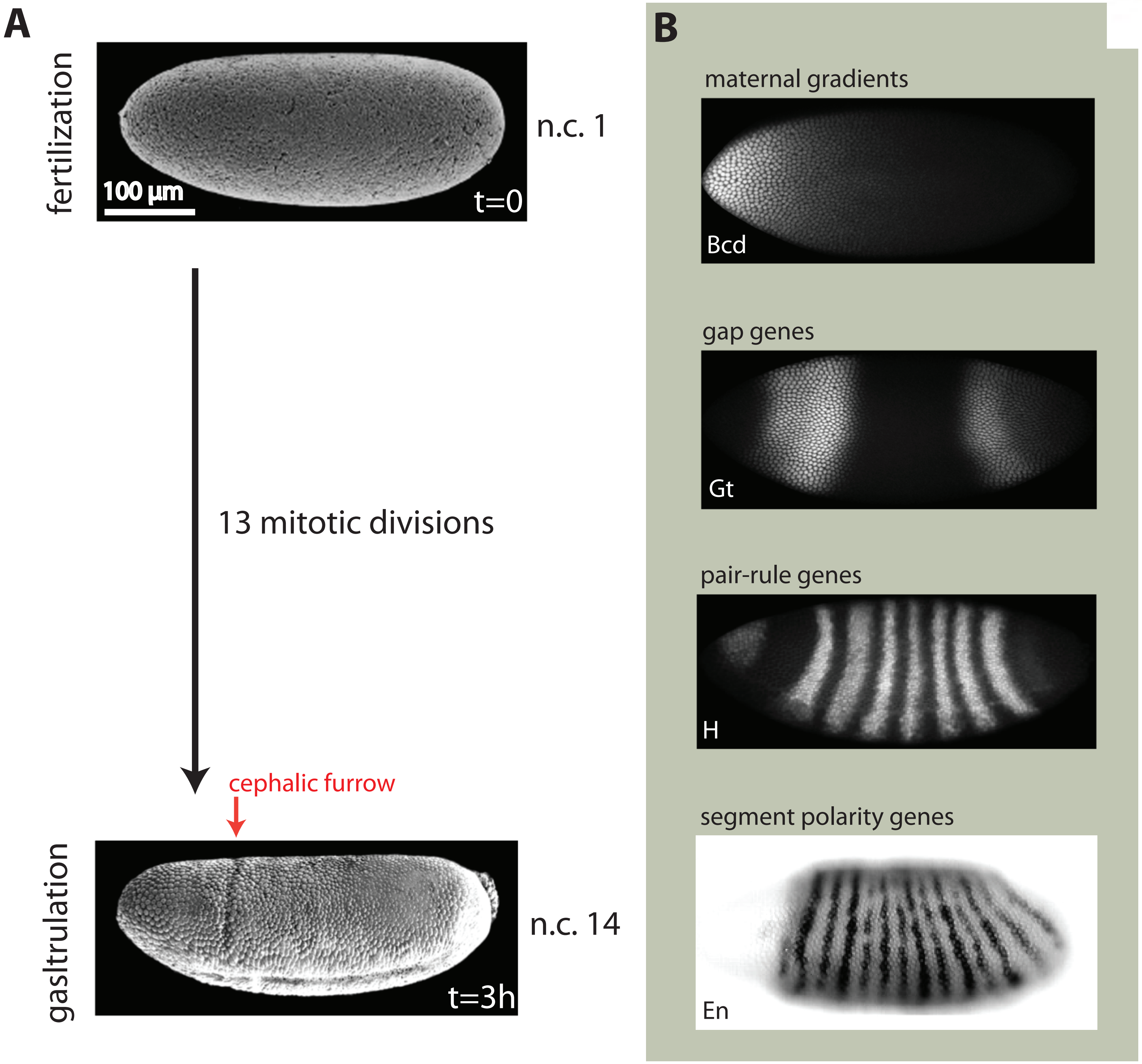
dunkel@math.mit.edu

Gene regulatory networks

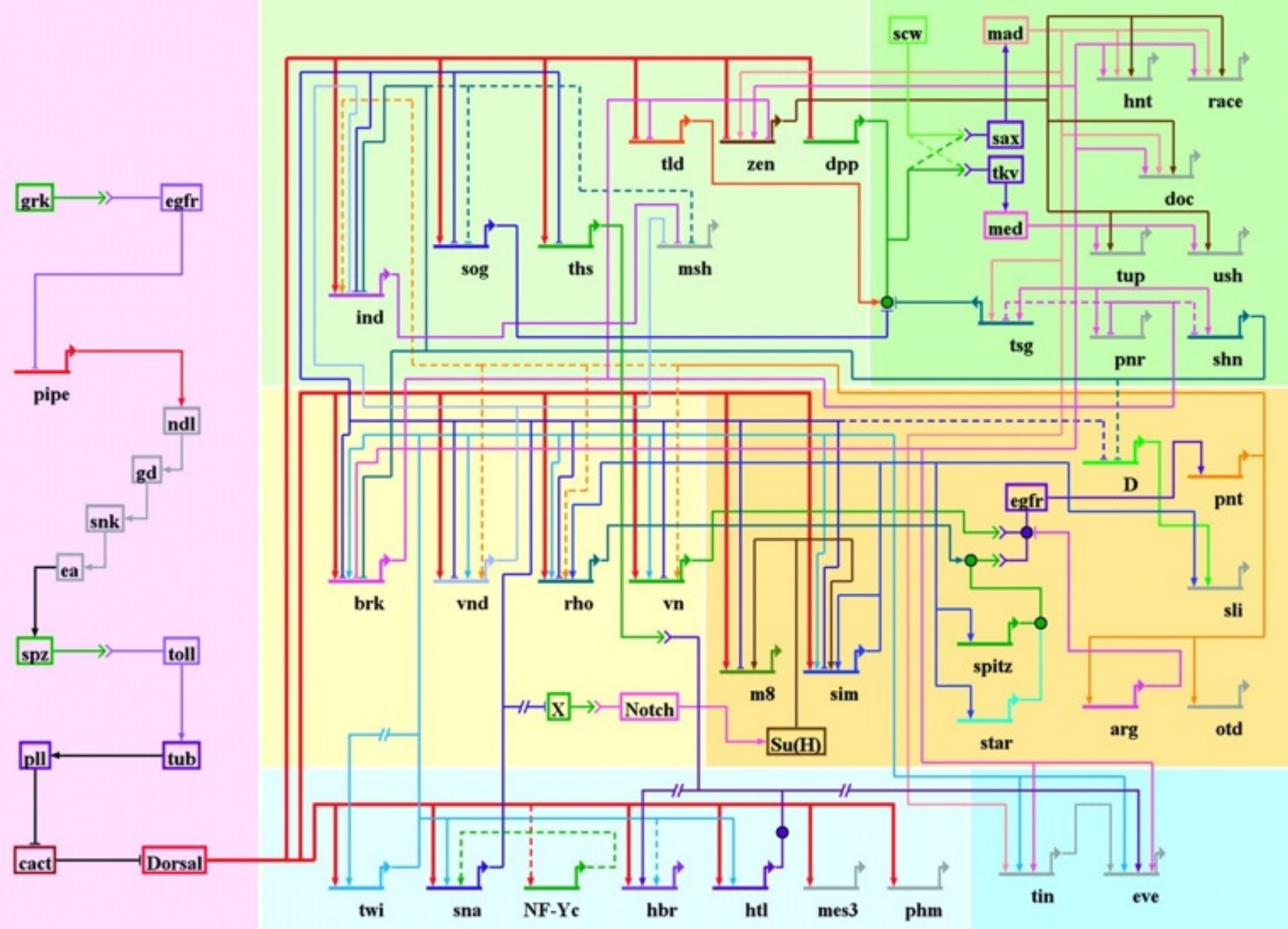


dunkel@math.mit.edu

Example: Drosophila embryo



The dorsal–ventral GRN in Drosophila



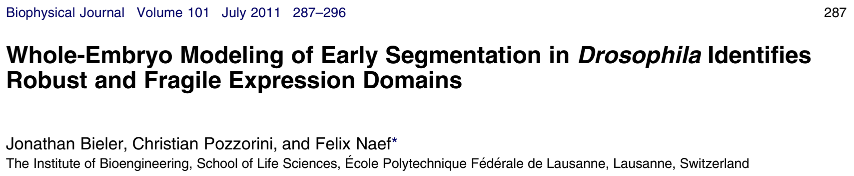
Levine and Davidson PNAS (2005) 102, 4936

Gap gene network

Molecular Systems Biology 9; Article number 639; doi:10.1038/msb.2012.72
Citation: ***Molecular Systems Biology*** **9:639**
www.molecularsystemsbiology.com

Accurate measurements of dynamics and reproducibility in small genetic networks

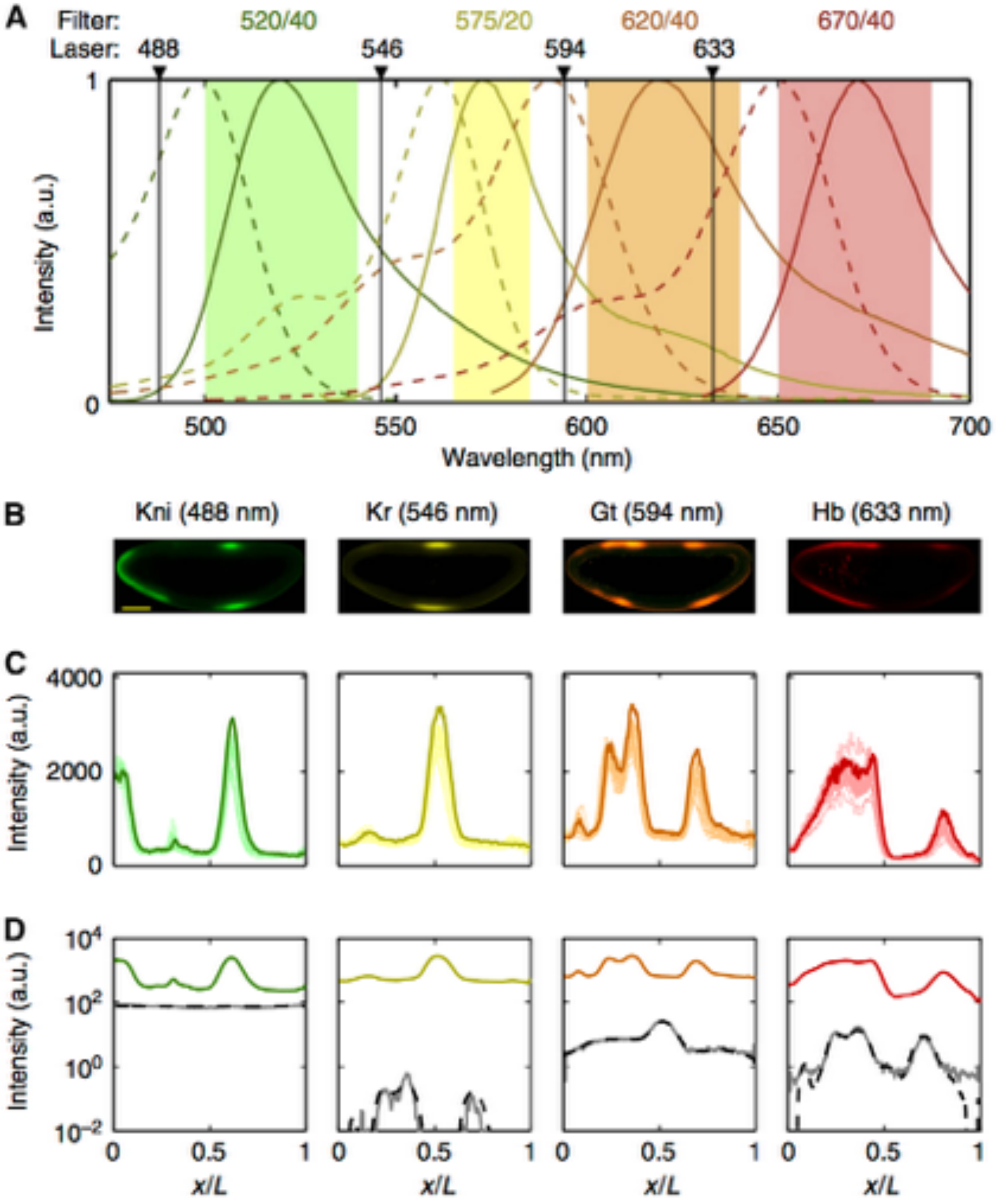
Julien O Dubuis[1,2], Reba Samanta[3] and Thomas Gregor[1,2,*]

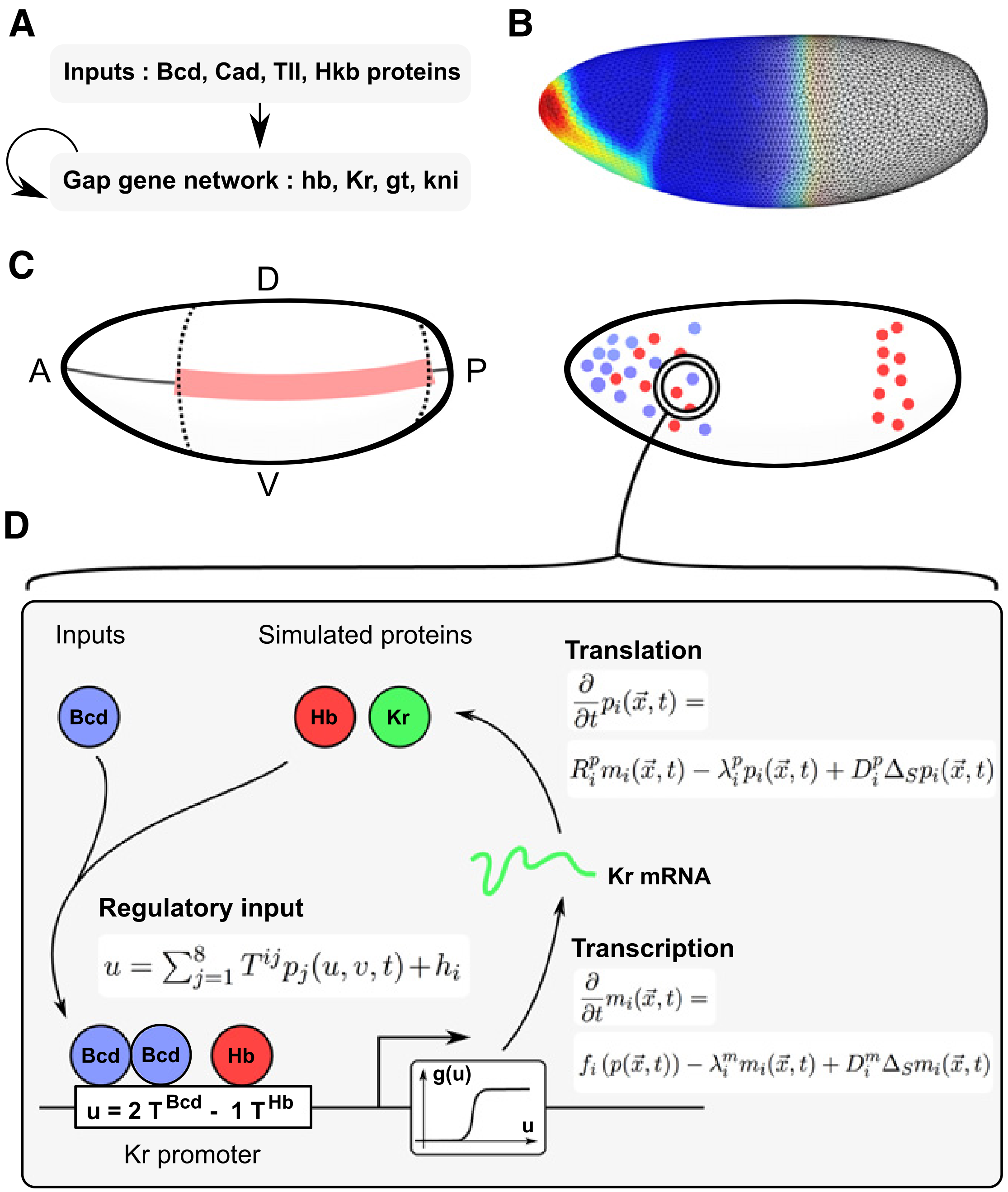


Biophysical Journal Volume 101 July 2011 287–296 287

Whole-Embryo Modeling of Early Segmentation in *Drosophila* Identifies Robust and Fragile Expression Domains

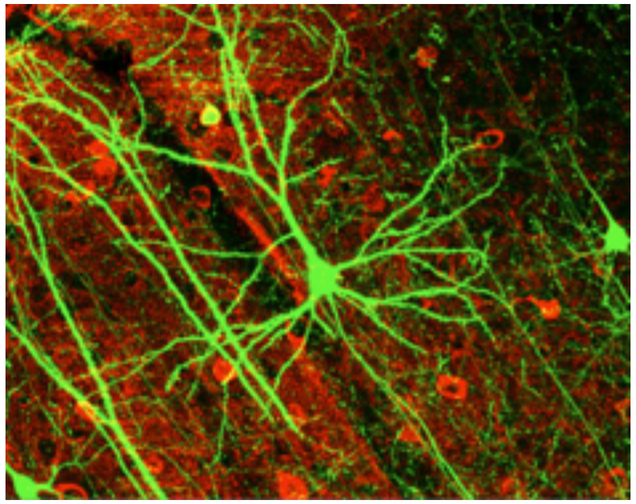
Jonathan Bieler, Christian Pozzorini, and Felix Naef*
The Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

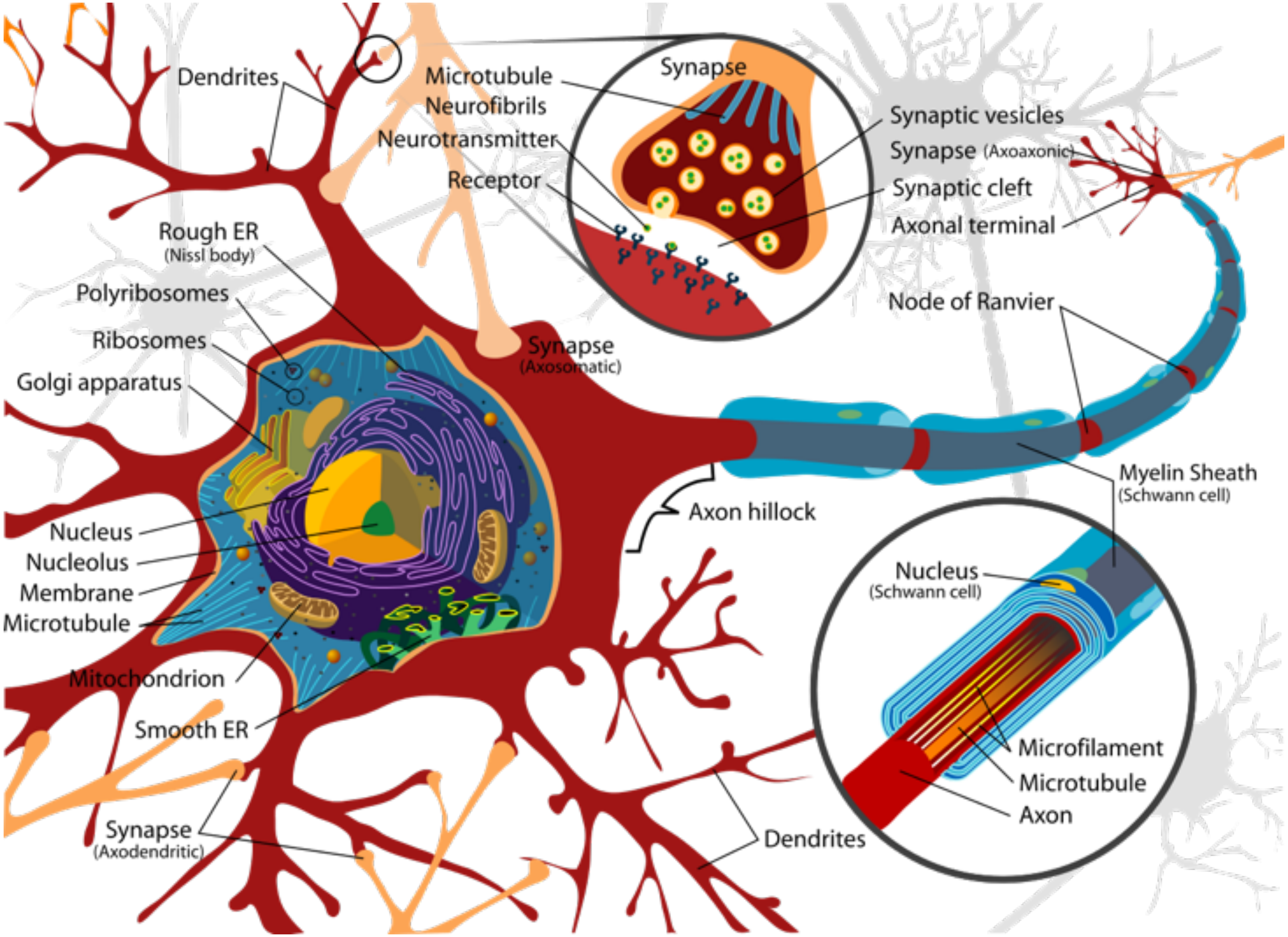


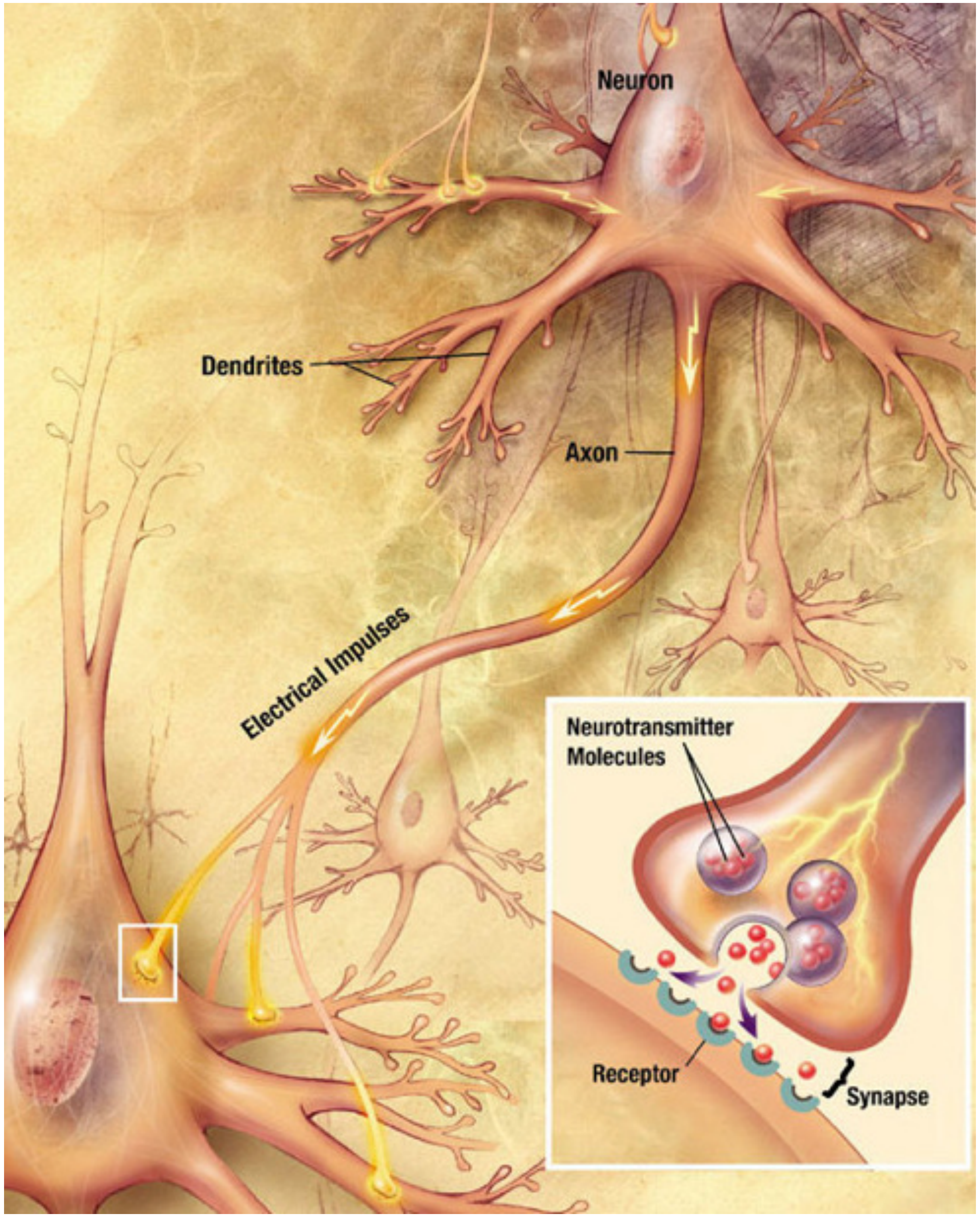


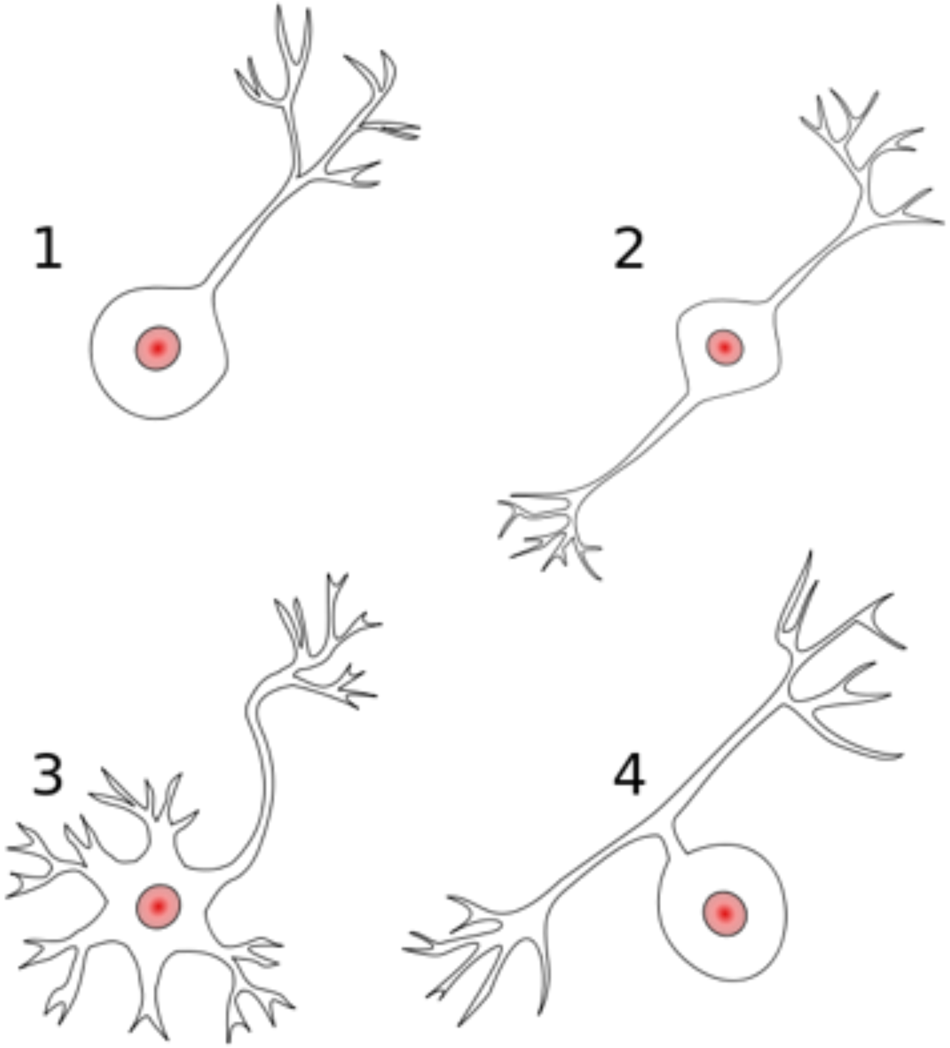
Neur(on)al networks

Image of pyramidal neurons in mouse cerebral cortex expressing green fluorescent protein. The red staining indicates GABAergic interneurons









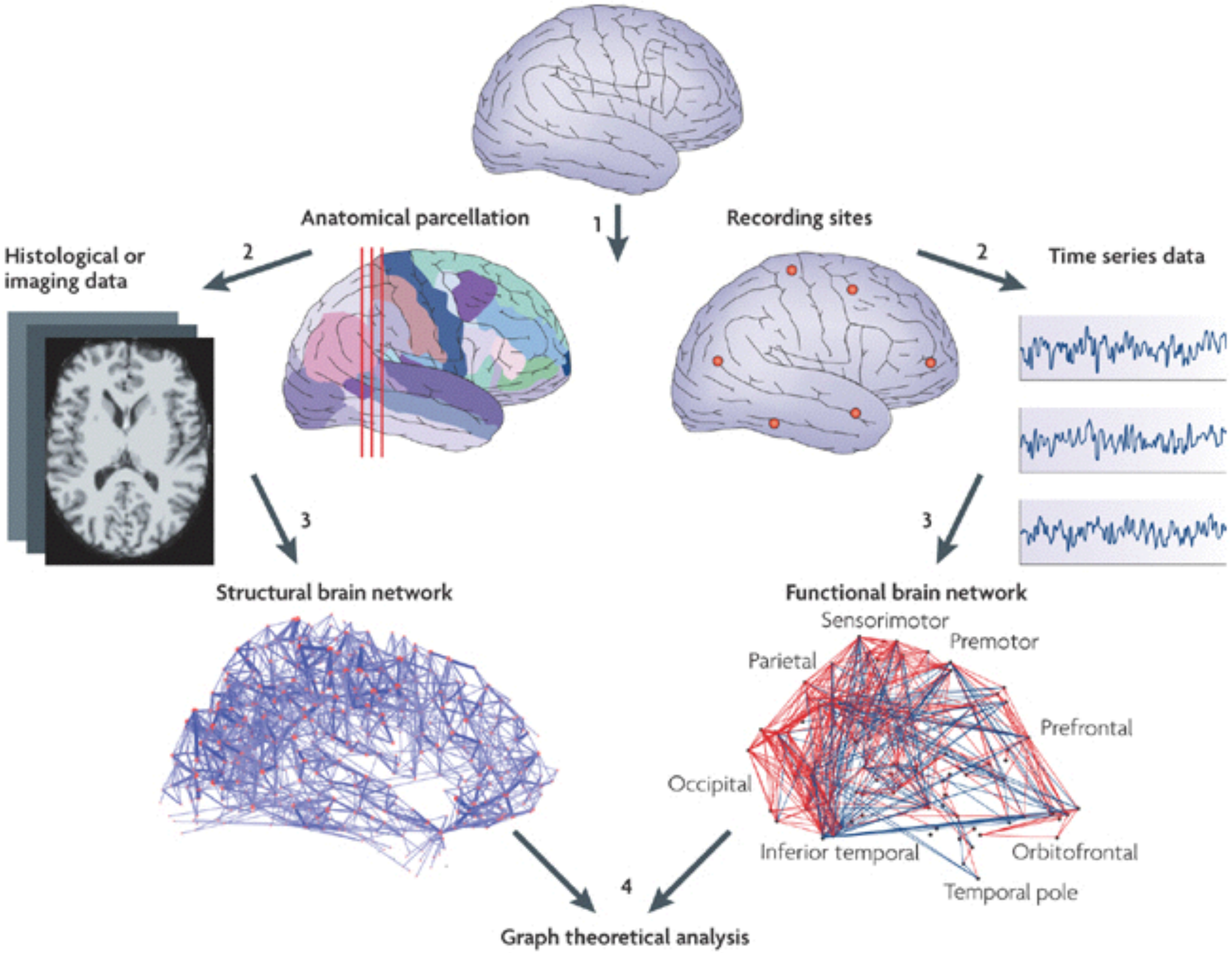
Different kinds of neurons:

1 Unipolar neuron

2 Bipolar neuron

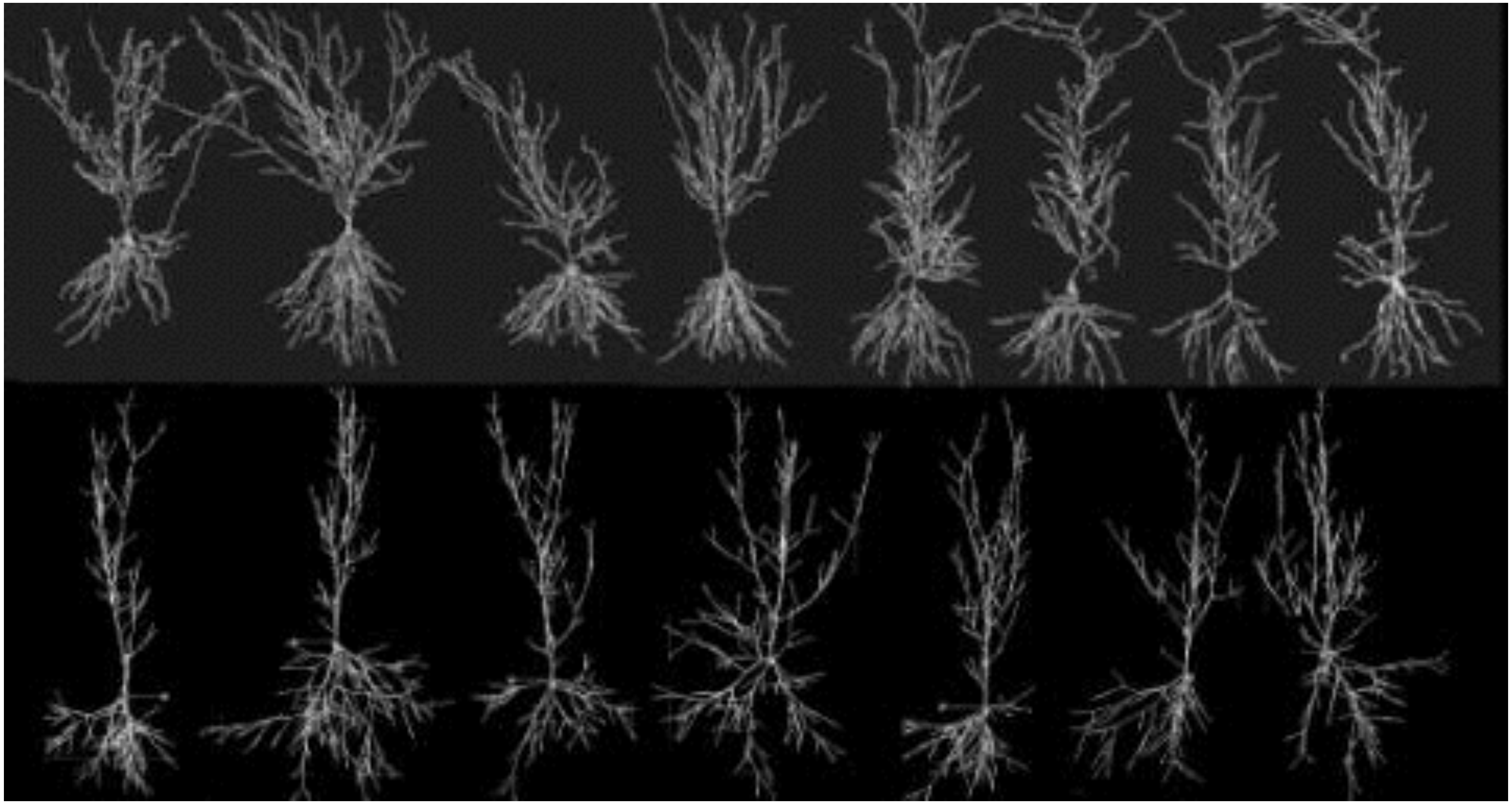
3 Multipolar neuron

4 Pseudounipolar neuron



Nature Reviews | Neuroscience

Ed Bullmore & Olaf Sporns
Nature Reviews Neuroscience **10**, 186-198 (March 2009)



Real hippocampal pyramidal cells (top) obtained from an experimental archive are used to extract fundamental parameters and generate virtual neurons (bottom), with Tamori’s algorithm. No two virtual or real neurons are ever identical, though they belong to a recognizeable morphological class.

http://krasnow1.gmu.edu/cn3/L-Neuron/HTM/paper.htm

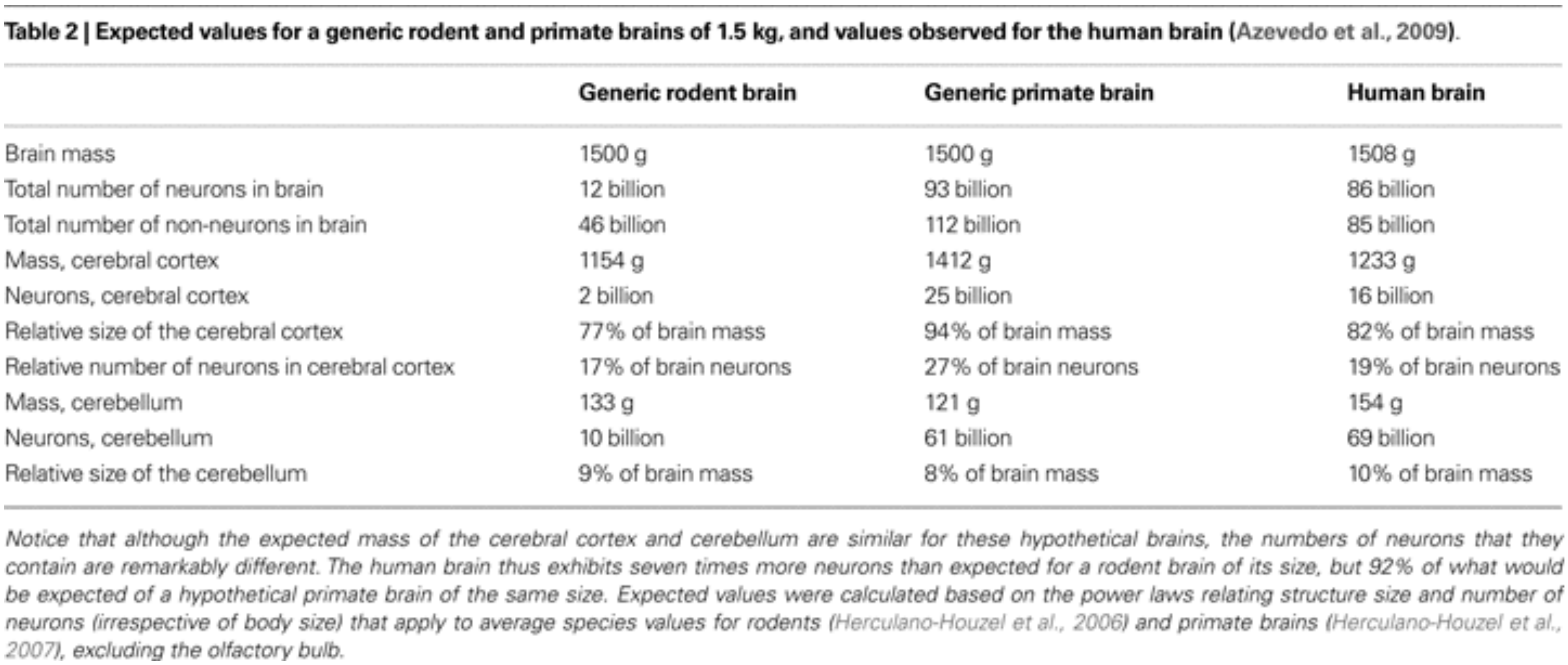
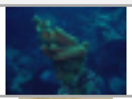
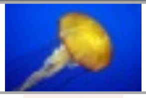
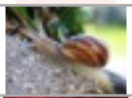
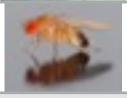
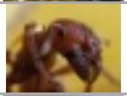
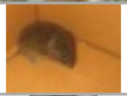
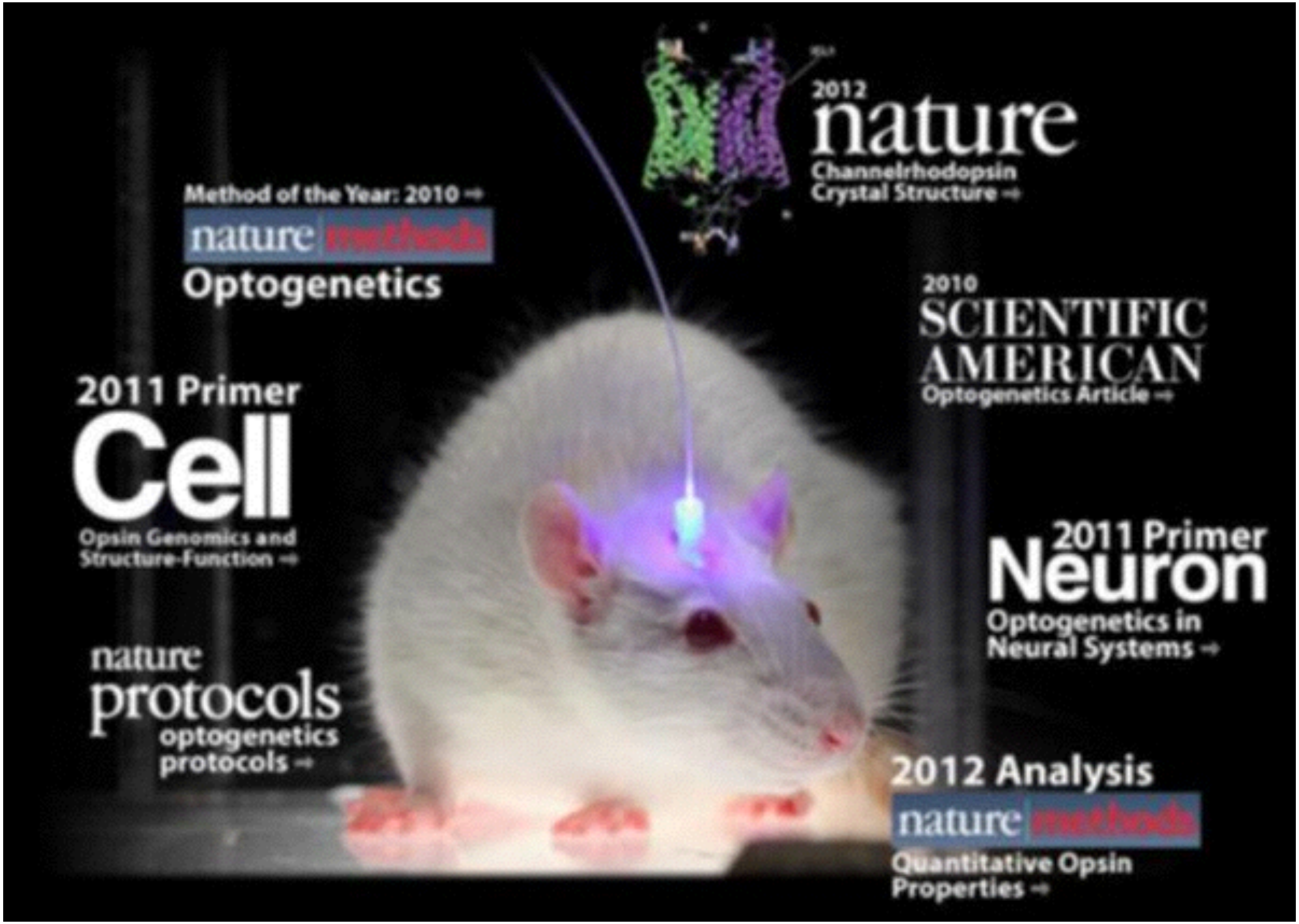


Table 2 | Expected values for a generic rodent and primate brains of 1.5 kg, and values observed for the human brain (Azevedo et al., 2009).

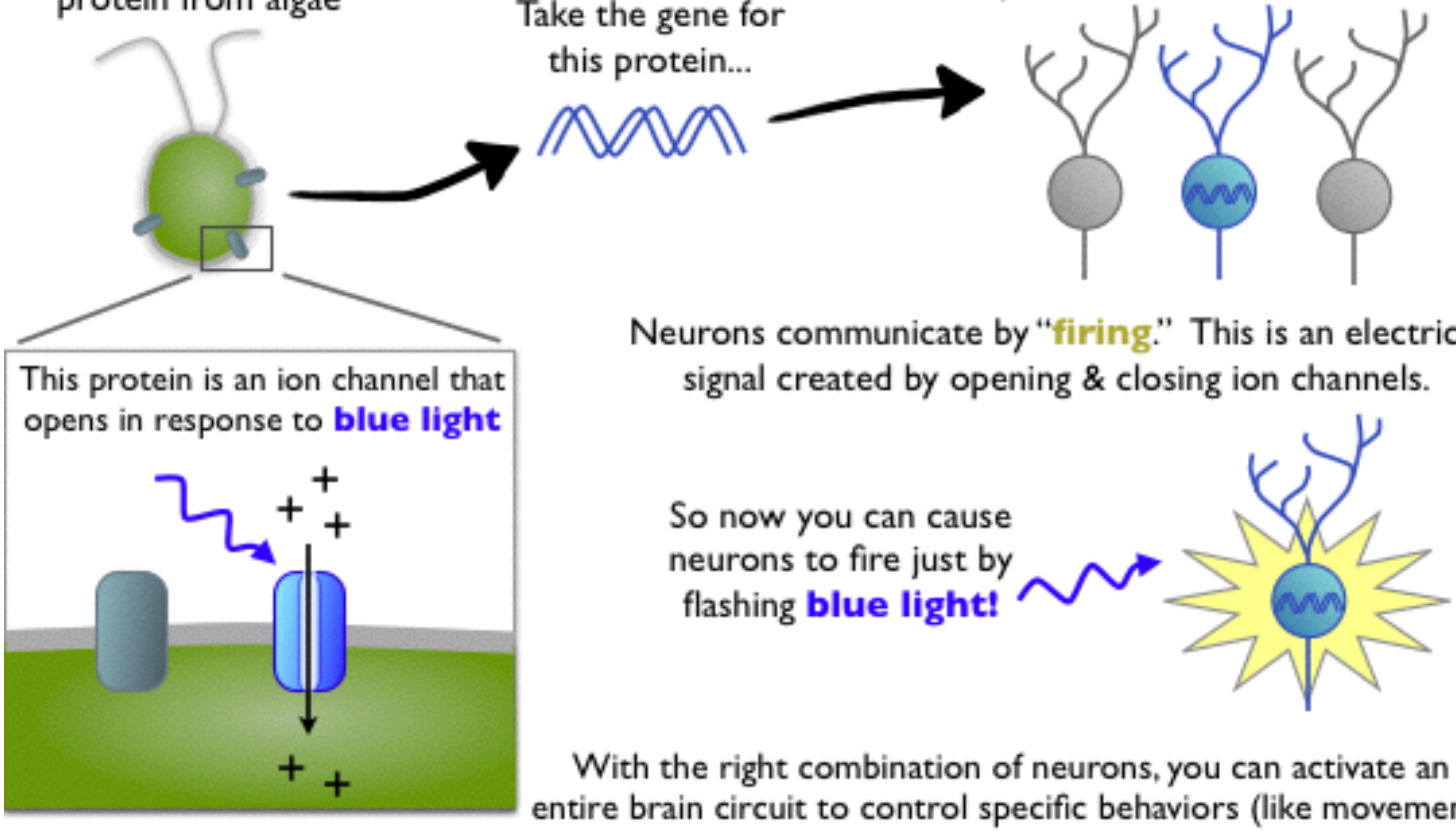
	Generic rodent brain	Generic primate brain	Human brain
Brain mass	1500 g	1500 g	1508 g
Total number of neurons in brain	12 billion	93 billion	86 billion
Total number of non-neurons in brain	46 billion	112 billion	85 billion
Mass, cerebral cortex	1154 g	1412 g	1233 g
Neurons, cerebral cortex	2 billion	25 billion	16 billion
Relative size of the cerebral cortex	77% of brain mass	94% of brain mass	82% of brain mass
Relative number of neurons in cerebral cortex	17% of brain neurons	27% of brain neurons	19% of brain neurons
Mass, cerebellum	133 g	121 g	154 g
Neurons, cerebellum	10 billion	61 billion	69 billion
Relative size of the cerebellum	9% of brain mass	8% of brain mass	10% of brain mass

Notice that although the expected mass of the cerebral cortex and cerebellum are similar for these hypothetical brains, the numbers of neurons that they contain are remarkably different. The human brain thus exhibits seven times more neurons than expected for a rodent brain of its size, but 92% of what would be expected of a hypothetical primate brain of the same size. Expected values were calculated based on the power laws relating structure size and number of neurons (irrespective of body size) that apply to average species values for rodents (Herculano-Houzel et al., 2006) and primate brains (Herculano-Houzel et al., 2007), excluding the olfactory bulb.

Name	Neurons in the brain/whole nervous system	Details	Image	Source
Sponge	0			
Trichoplax	0			[1]
Caenorhabditis elegans (roundworm)	302	~ 7,500 synapses		[2]
Jellyfish	800			[3]
Medicinal leech	10,000			[4]
Pond snail	11,000			[5]
Sea slug	18,000			[6]
Fruit fly	100,000	~ 10^{7}		[7]
Larval zebrafish	100,000			[8]
Lobster	100,000			[9]
Ant	250,000	Varies per species		[11]
Honey bee	960,000	~ 10^{9}		[12]
Cockroach	1,000,000			[13]
Adult zebrafish	~10,000,000 cells (neurons + other)			[14]
Frog	16,000,000			[15]
Mouse	71,000,000	~ 10^{11}		[16]
Rat	200,000,000	4.48 × 10		[3]
Octopus	300,000,000			[17]
Human	86,000,000,000	For average adult; 10		[18][19][20]


2012
nature
Channelrhodopsin
Crystal Structure ⇒
Method of the Year: 2010 ⇒
nature methods
Optogenetics
2010
SCIENTIFIC
AMERICAN
Optogenetics Article ⇒
2011 Primer
Cell
Opsin Genomics and
Structure-Function ⇒
2011 Primer
Neuron
Optogenetics in
Neural Systems ⇒
nature
protocols
optogenetics
protocols ⇒
2012 Analysis
nature methods
Quantitative Opsin
Properties ⇒

How optogenetics works



A light-sensitive protein from algae

Take the gene for this protein...

... and insert the DNA into specific neurons in the brain

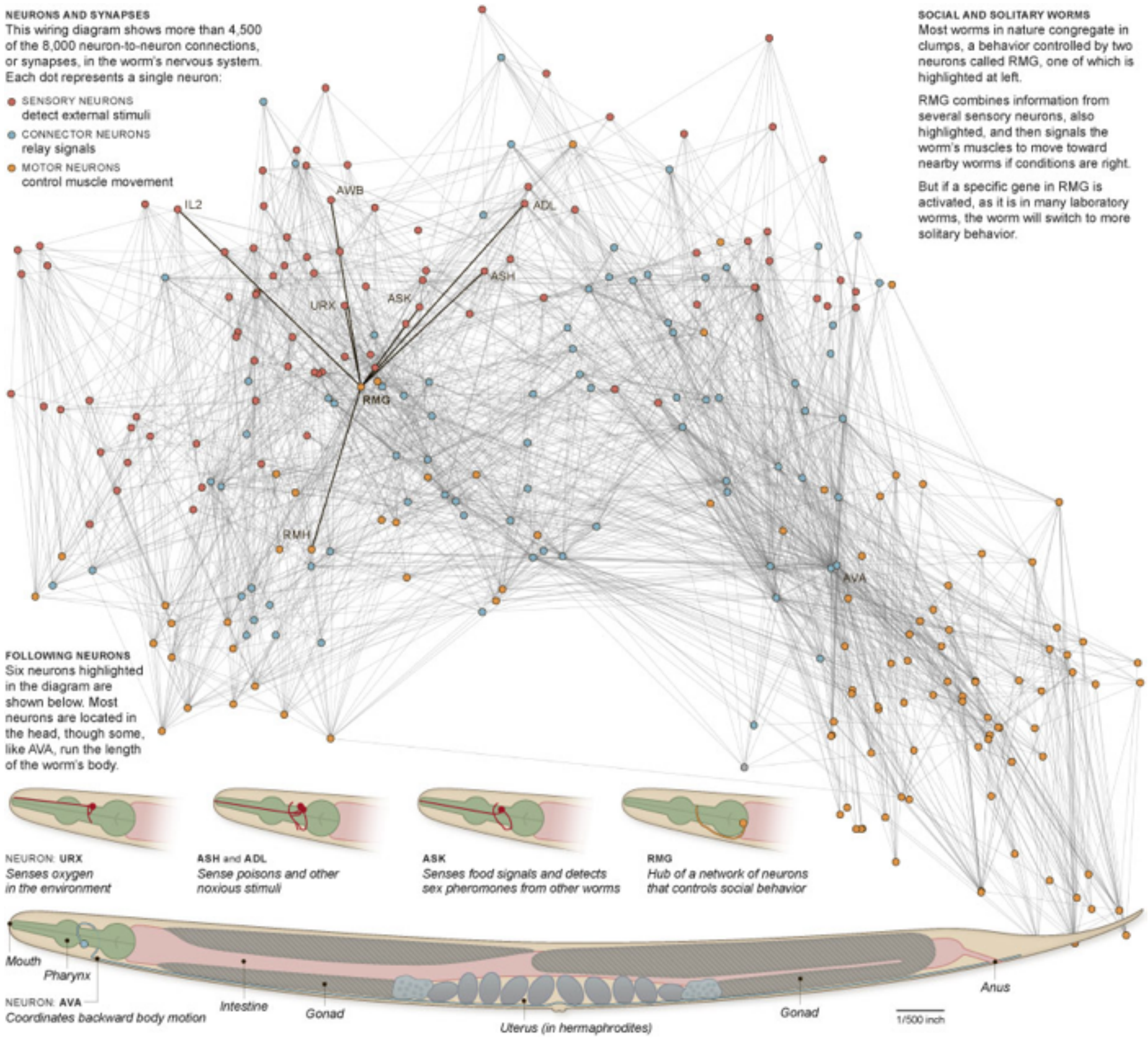
This protein is an ion channel that opens in response to **blue light**

Neurons communicate by "**firing**." This is an electrical signal created by opening & closing ion channels.

So now you can cause neurons to fire just by flashing **blue light!**

With the right combination of neurons, you can activate an entire brain circuit to control specific behaviors (like movement)

C elegans (302 neurons)



NEURONS AND SYNAPSES
This wiring diagram shows more than 4,500 of the 8,000 neuron-to-neuron connections, or synapses, in the worm's nervous system. Each dot represents a single neuron:

- 🟥 SENSORY NEURONS detect external stimuli
- 🟦 CONNECTOR NEURONS relay signals
- 🟧 MOTOR NEURONS control muscle movement

SOCIAL AND SOLITARY WORMS
Most worms in nature congregate in clumps, a behavior controlled by two neurons called RMG, one of which is highlighted at left.

RMG combines information from several sensory neurons, also highlighted, and then signals the worm's muscles to move toward nearby worms if conditions are right.

But if a specific gene in RMG is activated, as it is in many laboratory worms, the worm will switch to more solitary behavior.

FOLLOWING NEURONS
Six neurons highlighted in the diagram are shown below. Most neurons are located in the head, though some, like AVA, run the length of the worm's body.

NEURON: **URX**
Senses oxygen in the environment

ASH and **ADL**
Sense poisons and other noxious stimuli

ASK
Senses food signals and detects sex pheromones from other worms

RMG
Hub of a network of neurons that controls social behavior

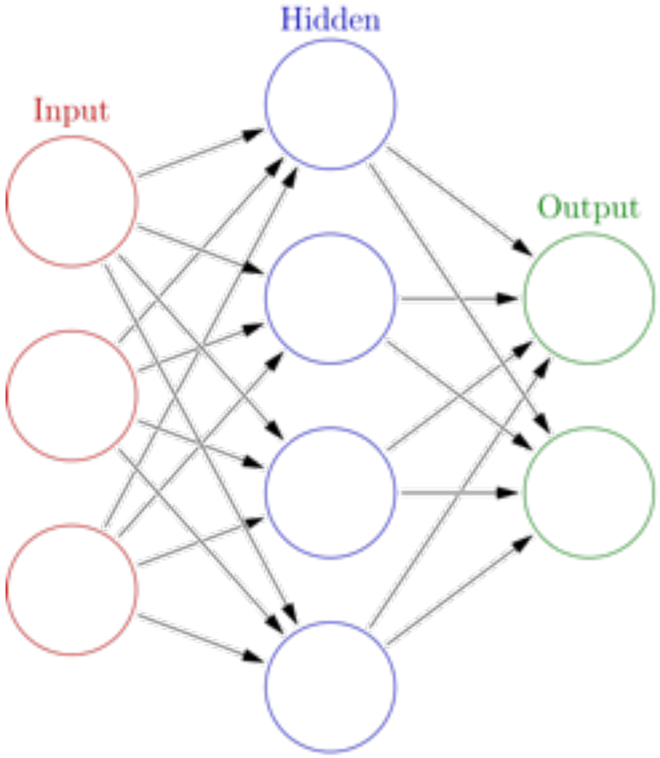
NEURON: **AVA**
Coordinates backward body motion

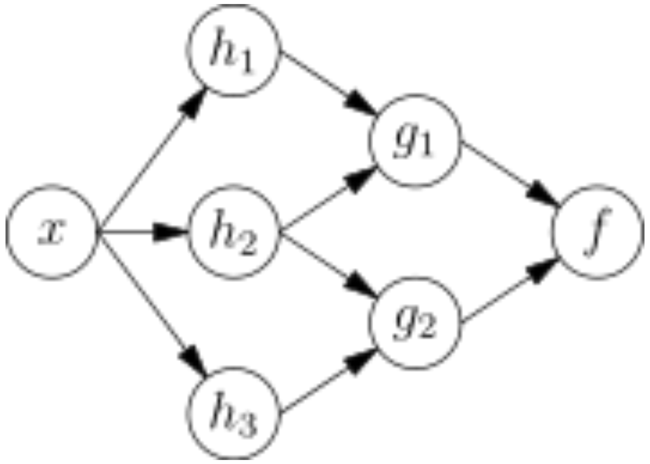
https://www.youtube.com/watch?v=I64X7vHSHOE

Artificial neural networks (ANNs)

An ANN is typically defined by three types of parameters:

1. The interconnection pattern between the different layers of neurons
2. The learning process for updating the weights of the interconnections
3. The activation function that converts a neuron's weighted input to its output activation.



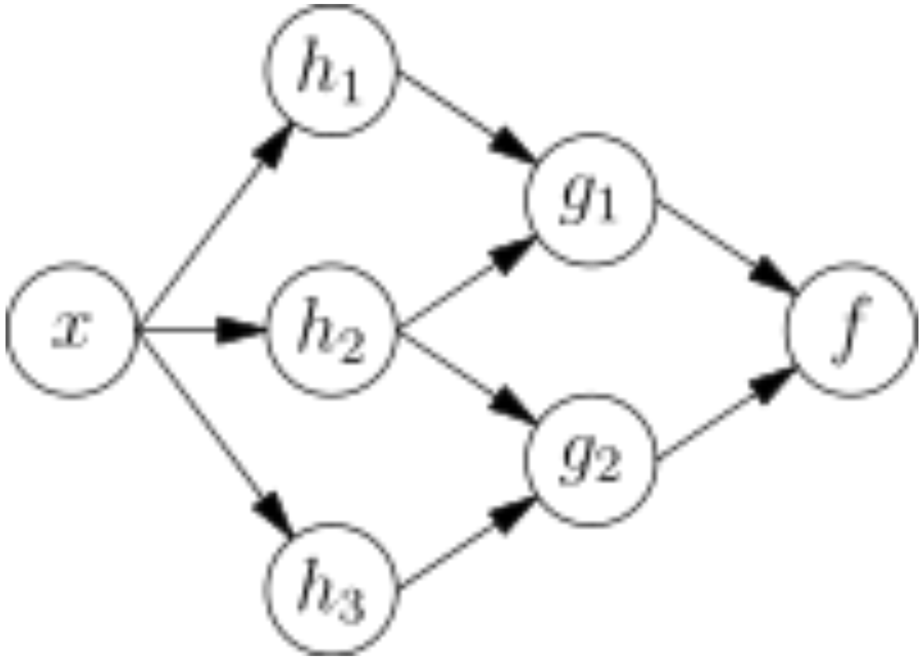


Mathematically, a neuron's network function ***f*** is defined as a composition of other functions ***h*,*g***

Learning

Given a specific *task* to solve, and a *class* of functions **F={*f*}**, learning means using a set of *observations* to find ***f**** in **F** which solves the task in some *optimal* sense.

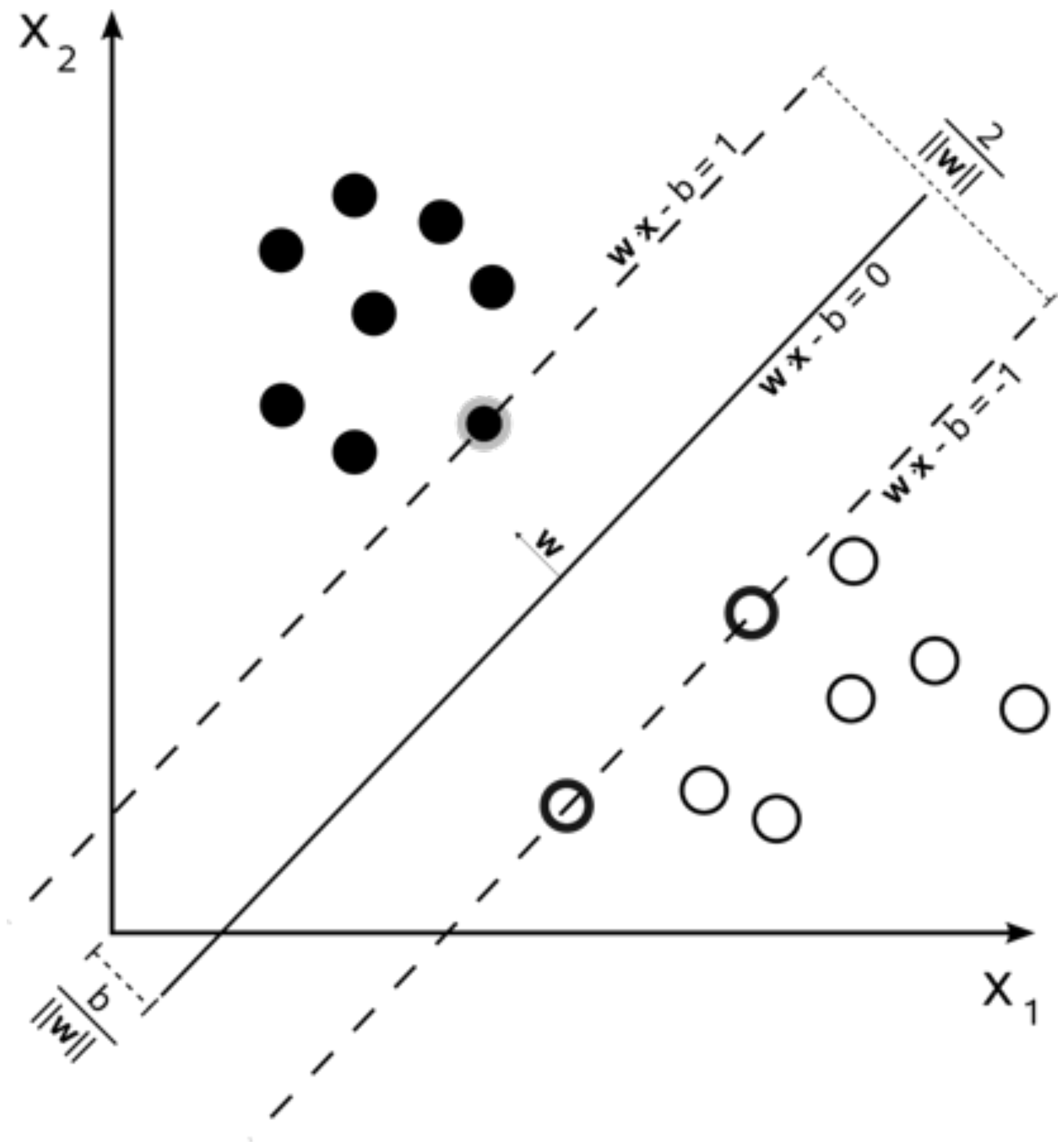
Needs real-valued typically convex cost function ***C(f)*** to determine what's optimal



Example: Support vector machines

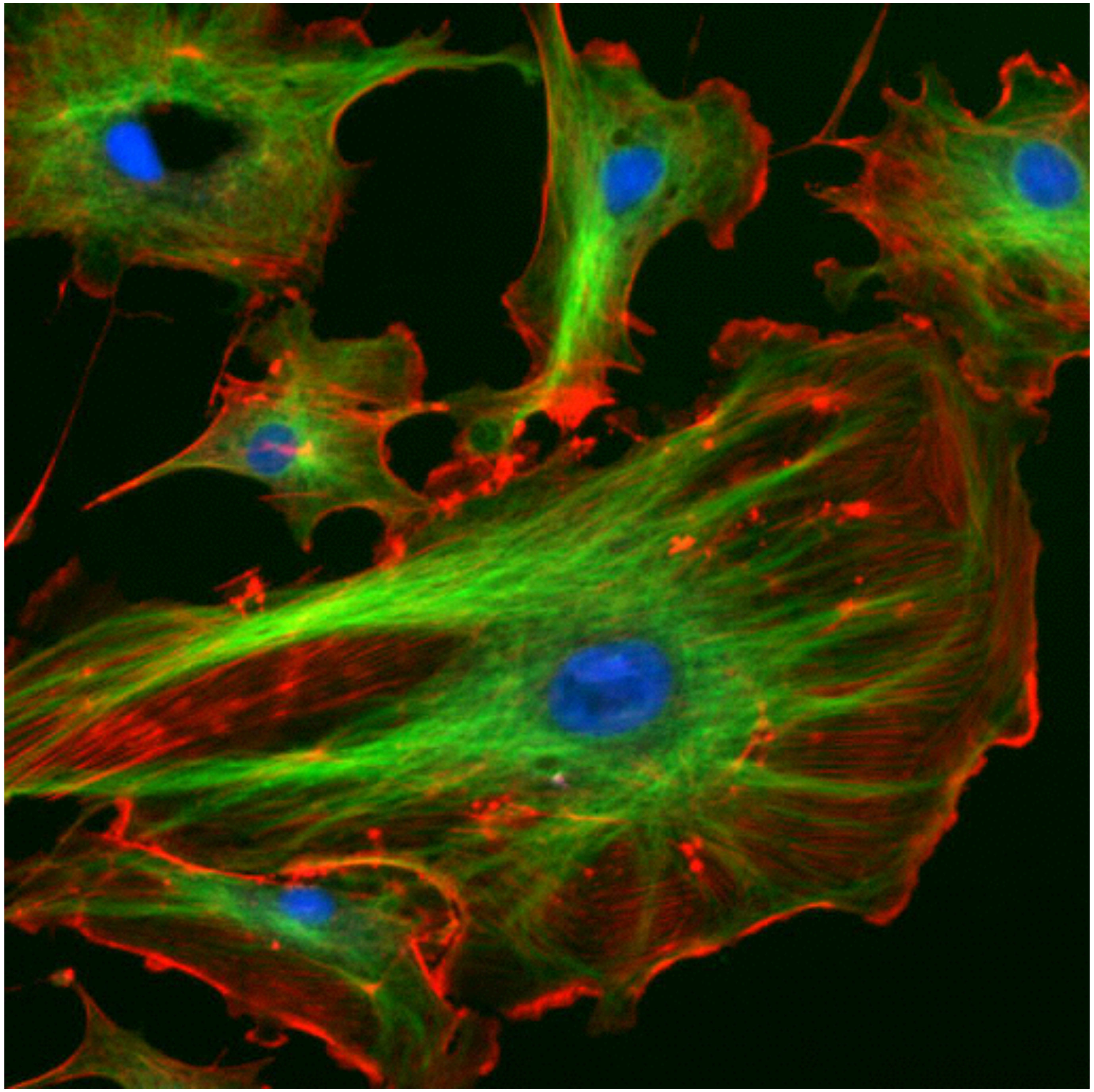
Choose the hyperplane so that the distance from it to the nearest data point on each side is maximized.

If such a hyperplane exists, it is known as the *maximum-margin hyperplane* and the linear classifier it defines is known as a *maximum margin classifier*; or equivalently, the *perceptron of optimal stability.*



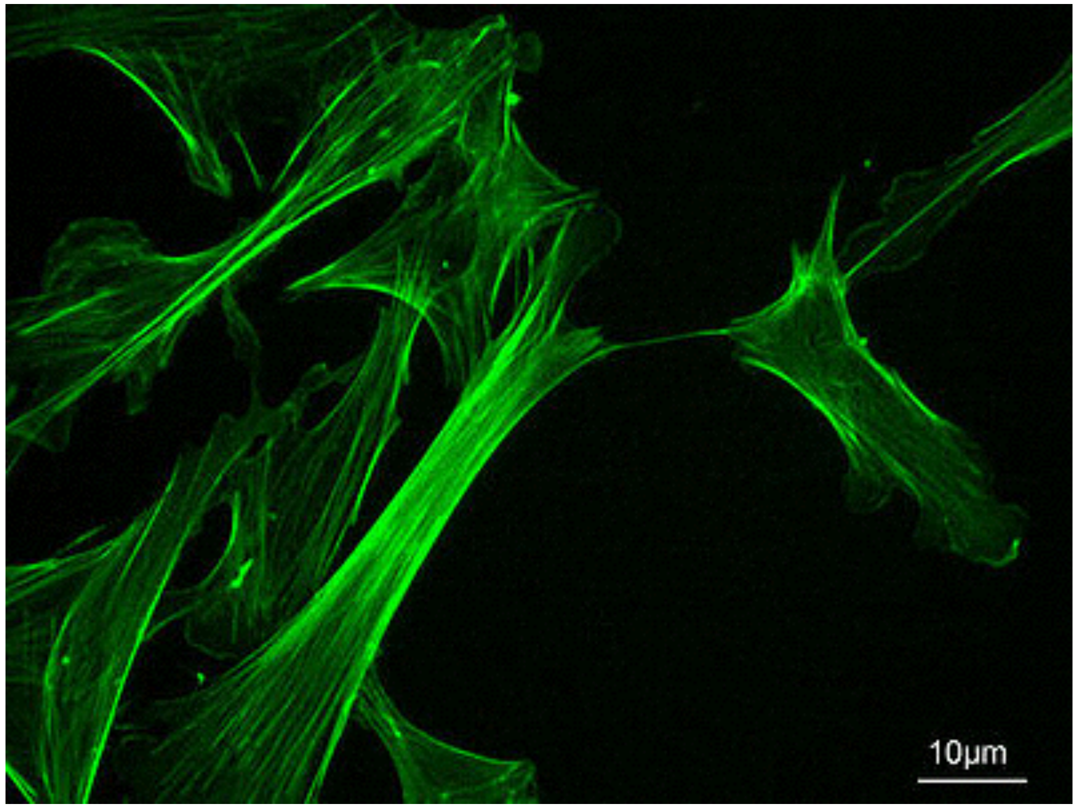
Structural networks

Cytoskeleton



The eukaryotic cytoskeleton. Actin filaments are shown in red, microtubules are in green, and the nuclei are in blue.

http://rsb.info.nih.gov/ij/images/



Actin cytoskeleton of mouse embryo fibroblasts, stained with phalloidin

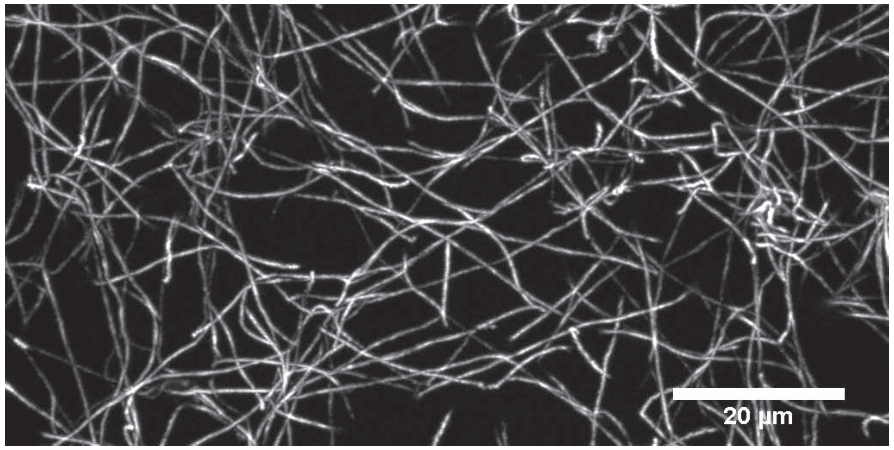
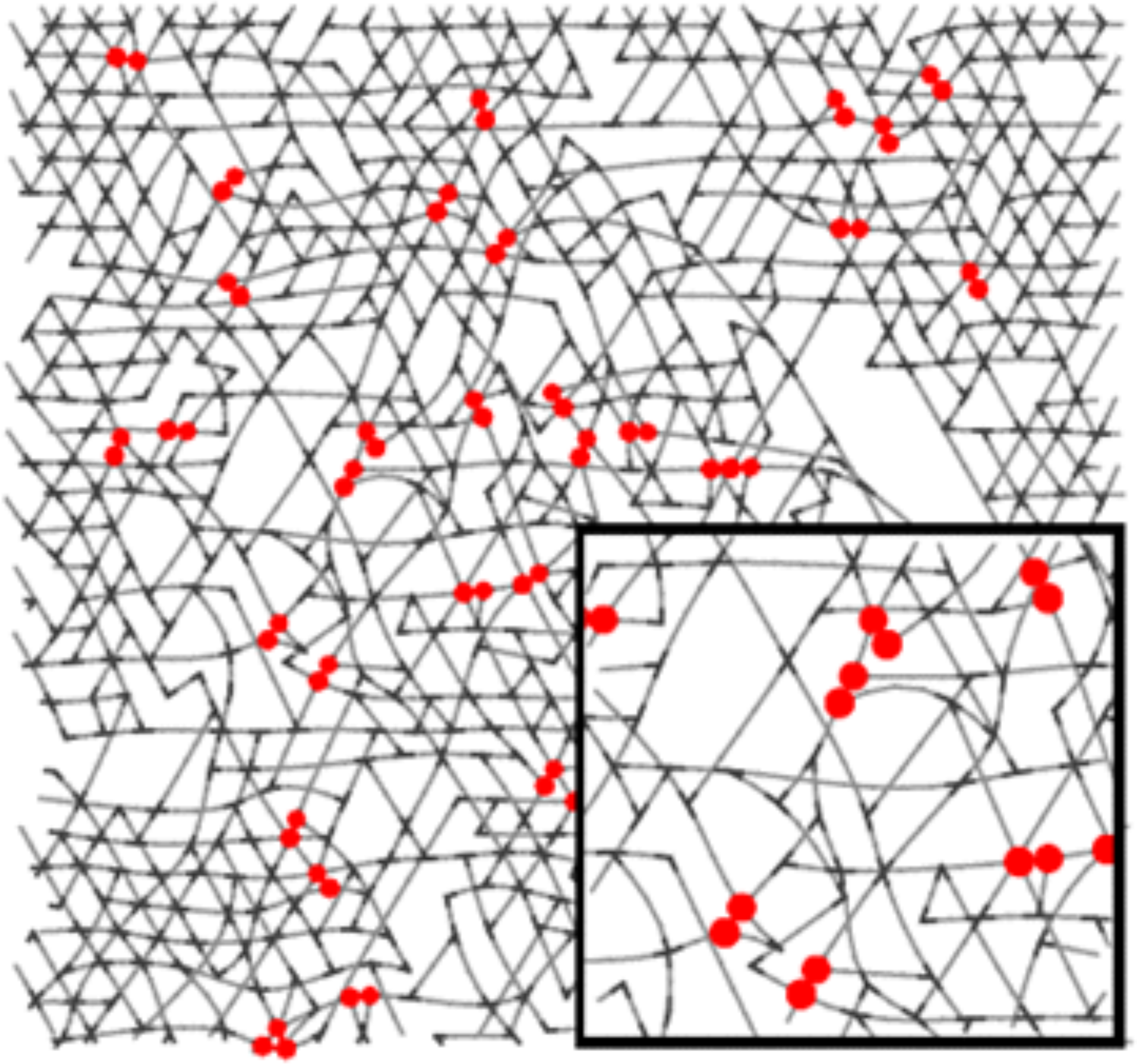


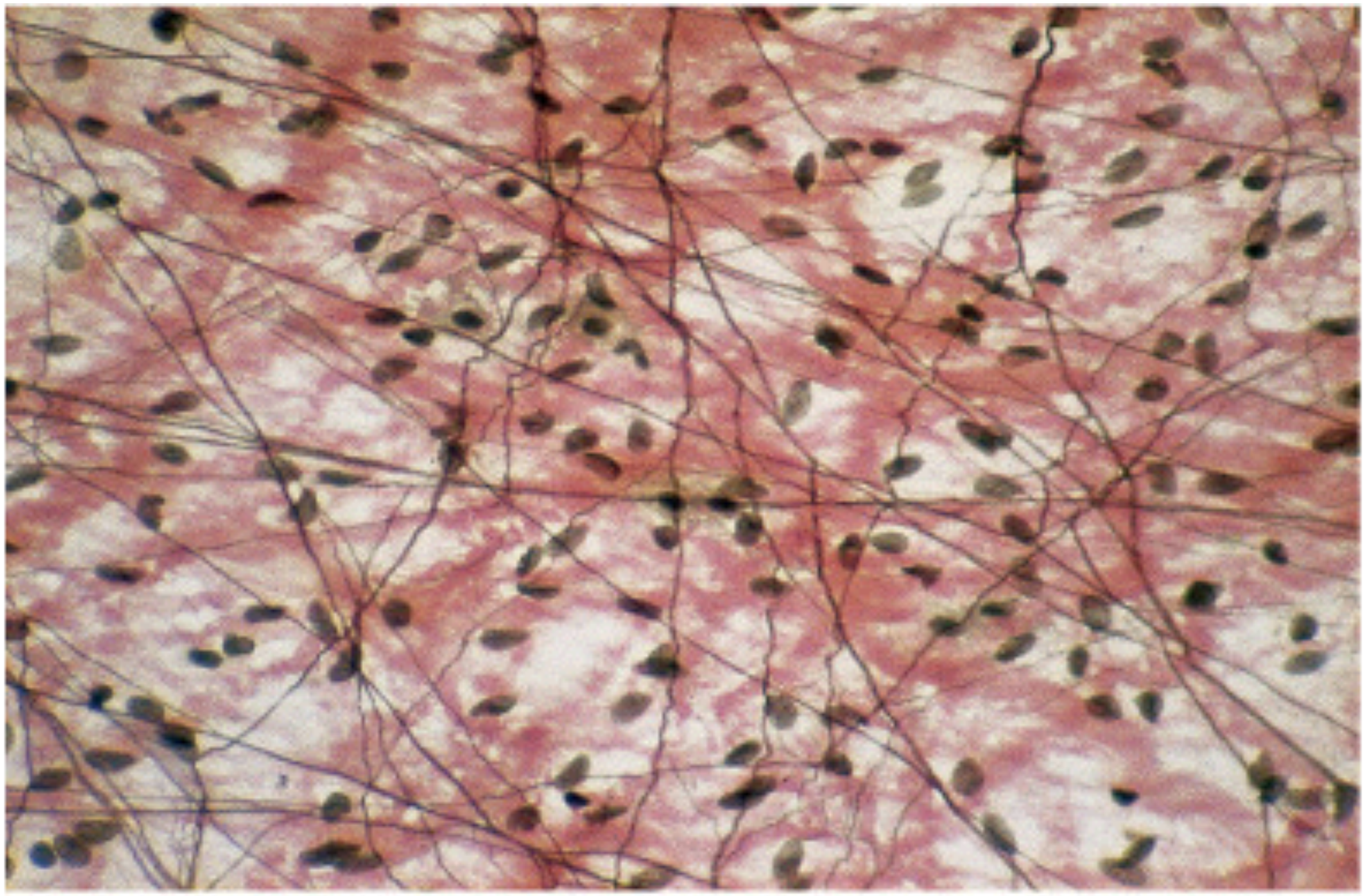
FIG. 2 Confocal microscopy image of a fluorescently labeled collagen network with a concentration of 0.4 mg/ml. Courtesy of Stefan Münster (Erlangen-Nurnberg).

Molecular motors stiffen non-affine semiflexible polymer networks



see also Brodersz & MacKintosh, Rev Mod Phys (2014)

Connective and supporting tissues

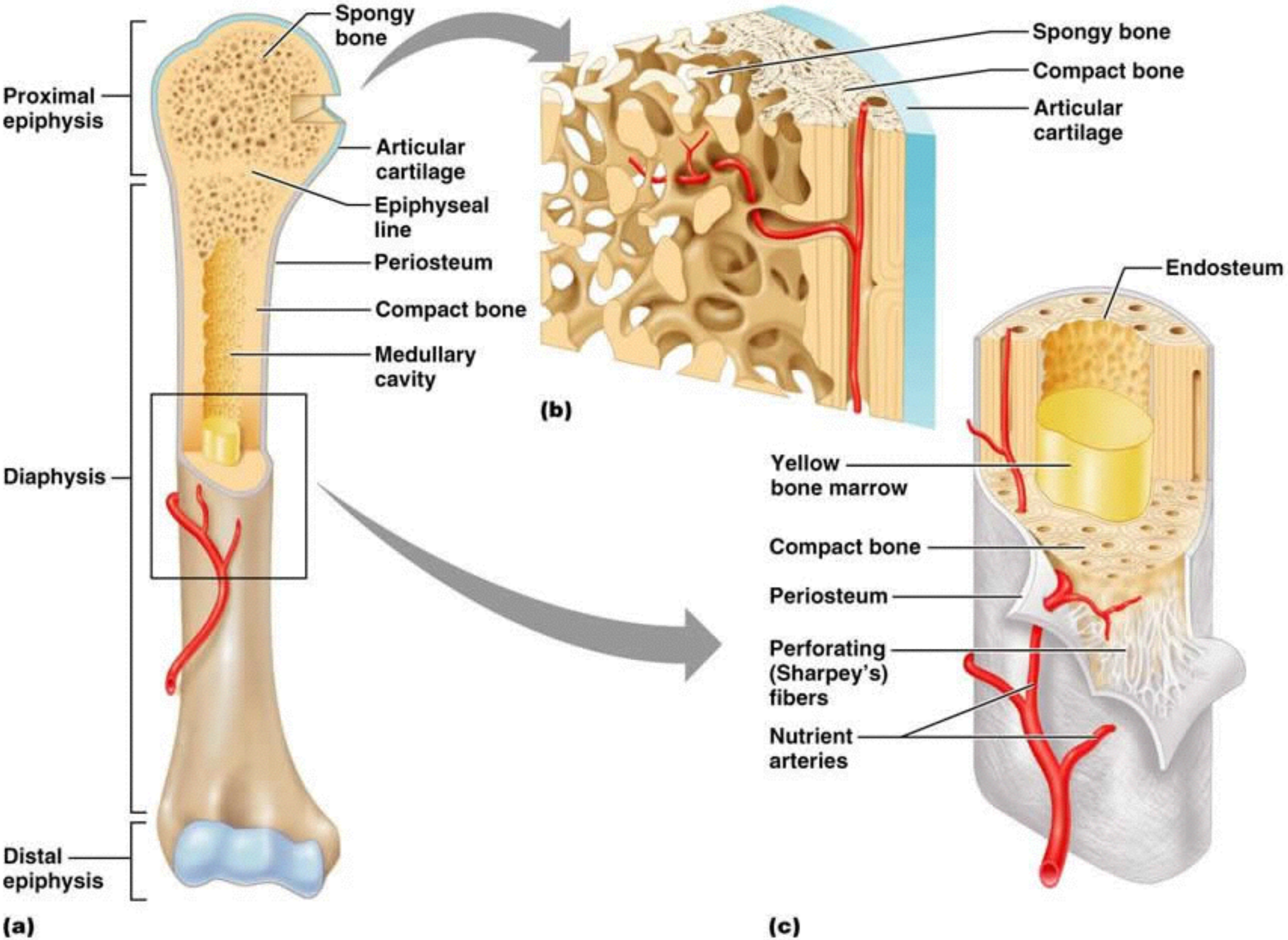


© Elsevier Ltd 2005. Standring: Gray's Anatomy 39e - www.graysanatomyonline.com

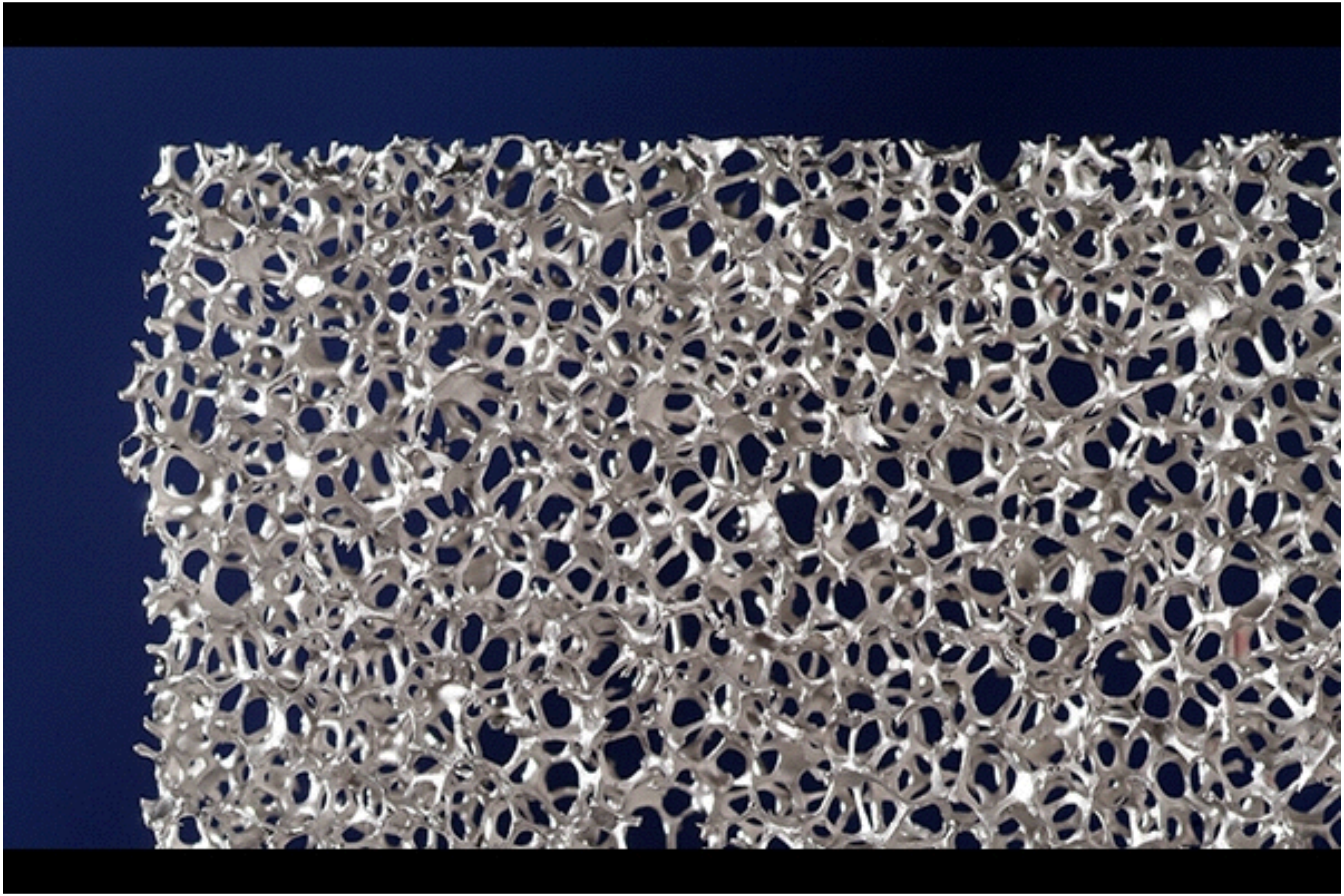
The connective tissues are defined as those composed predominantly of intercellular material, the extracellular matrix, which is secreted mainly by the connective tissue cells. The cells are therefore usually widely separated by their matrix, which is composed of fibrous proteins and a relatively amorphous ground substance.

Bone structures

Coloured scanning electron micrograph (SEM) of cancellous (spongy) bone. This tissue, found in the interior of bones, is characterised by a honeycomb arrangement of trabeculae (columns) and spaces.

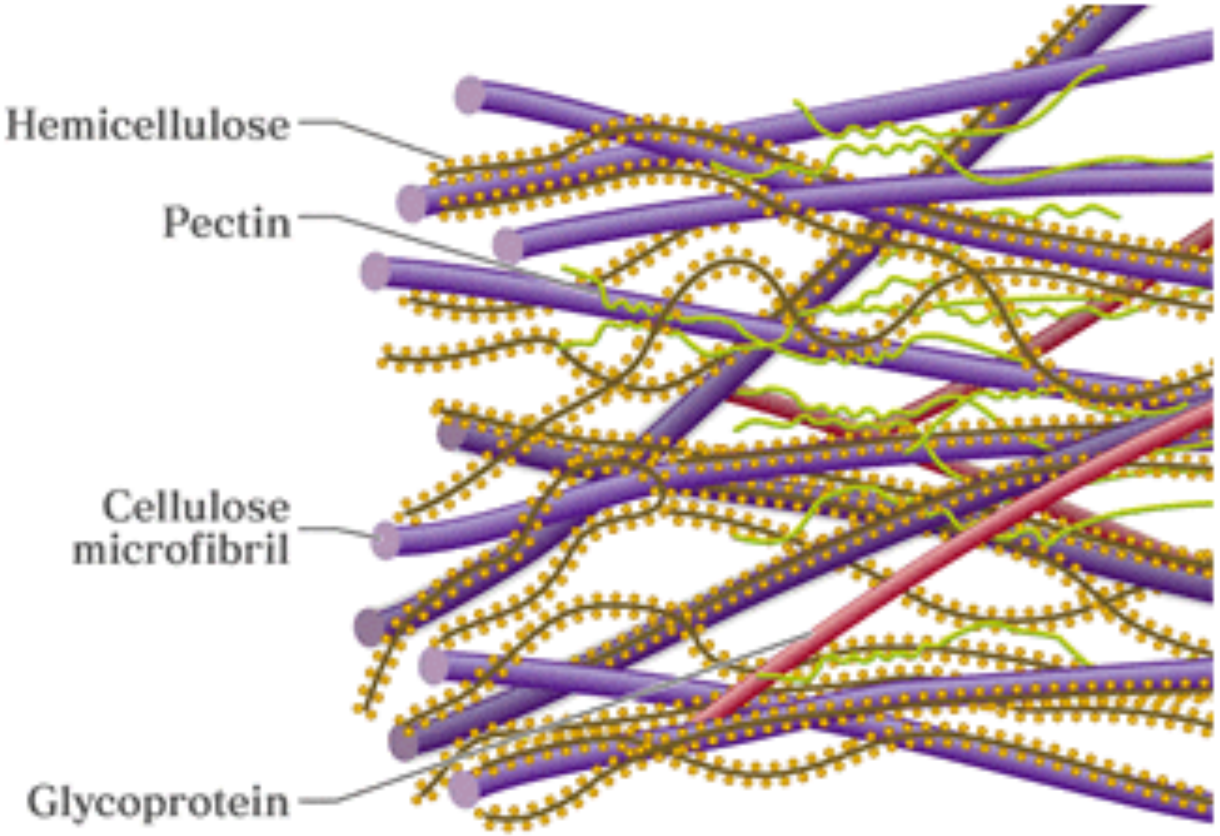

Spongy bone
Proximal epiphysis
Articular cartilage
Epiphyseal line
Periosteum
Compact bone
Medullary cavity
Diaphysis
Distal epiphysis
(a)
Spongy bone
Compact bone
Articular cartilage
(b)
Endosteum
Yellow bone marrow
Compact bone
Periosteum
Perforating (Sharpey's) fibers
Nutrient arteries
(c)

Metallic foam structures as a bone substitute



Fraunhofer Institute for Manufacturing Technology and Advanced Materials IFAM

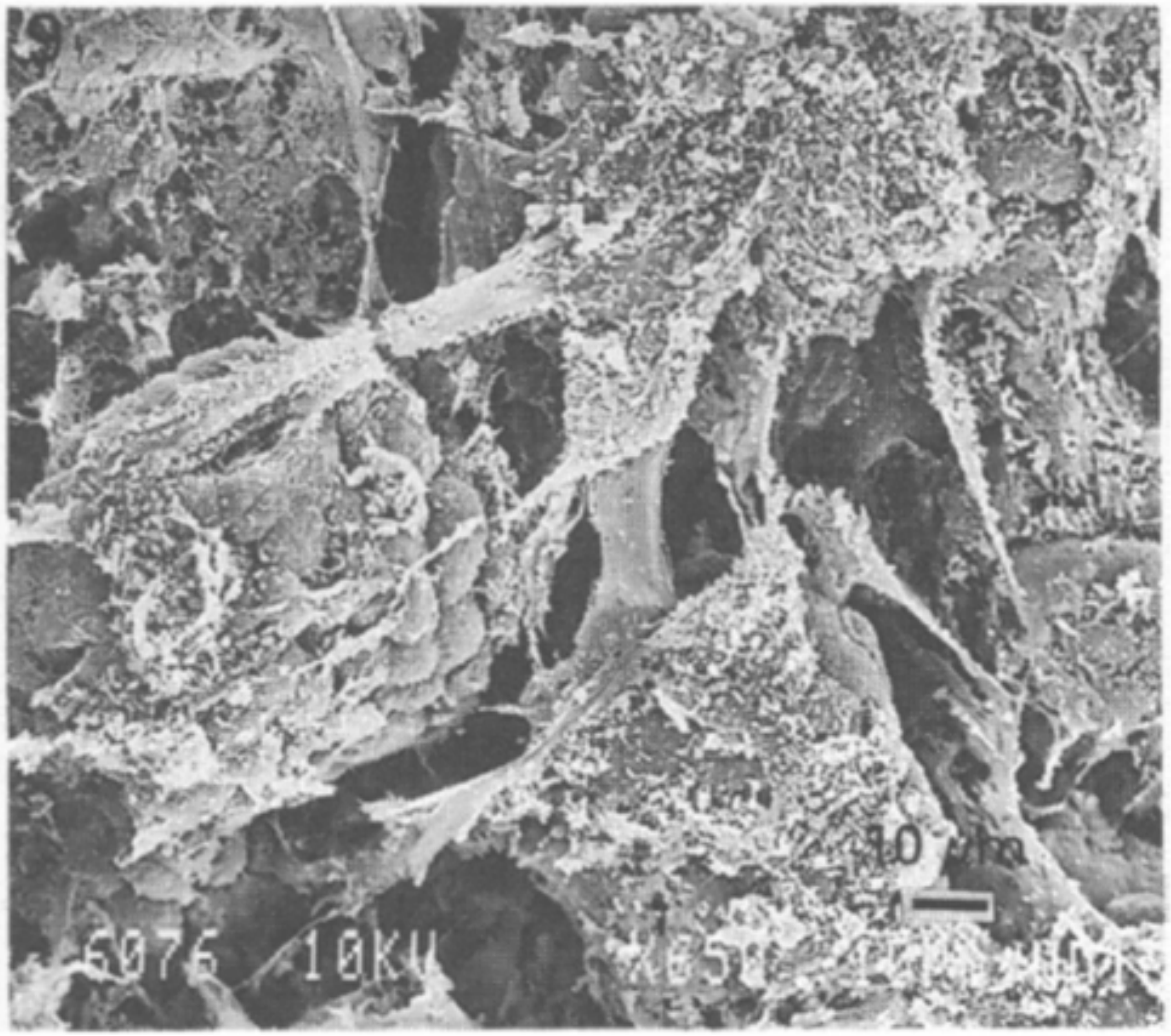
Inter-cellular matrix of plants



+ ZOOM

Mucus

Scanning electron microscopy image of rabbit stomach mucosa showing surface cells and numerous bundles of fiber-like mucus strands.
Scale bar 10 µm. Reproduced from Nunn et al. (51) with permission from Wiley.

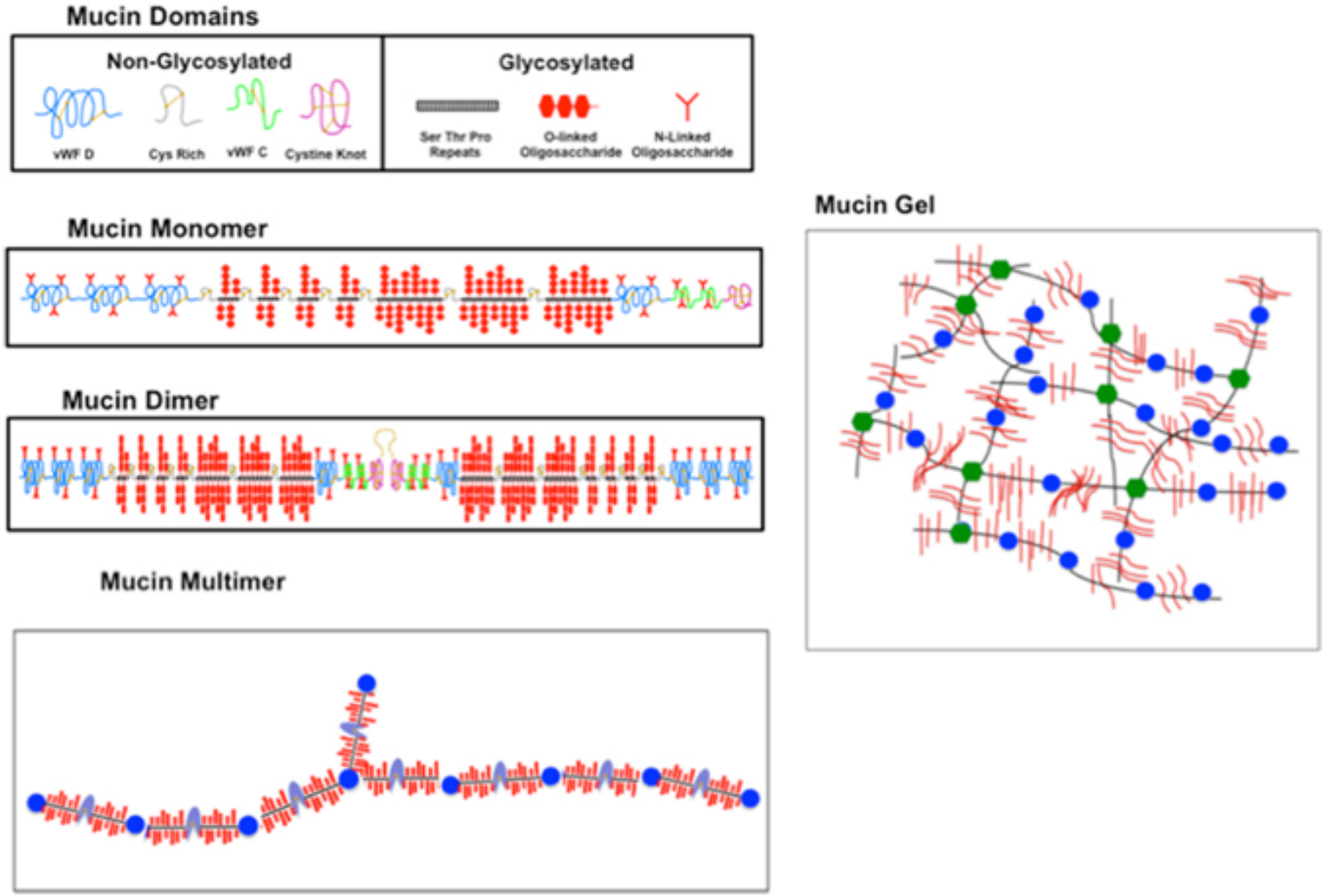


The influence of mucus microstructure and rheology in *Helicobacter pylori* infection

Front. Immunol., 10 October 2013 | doi: 10.3389/fimmu.2013.00310

Mucus

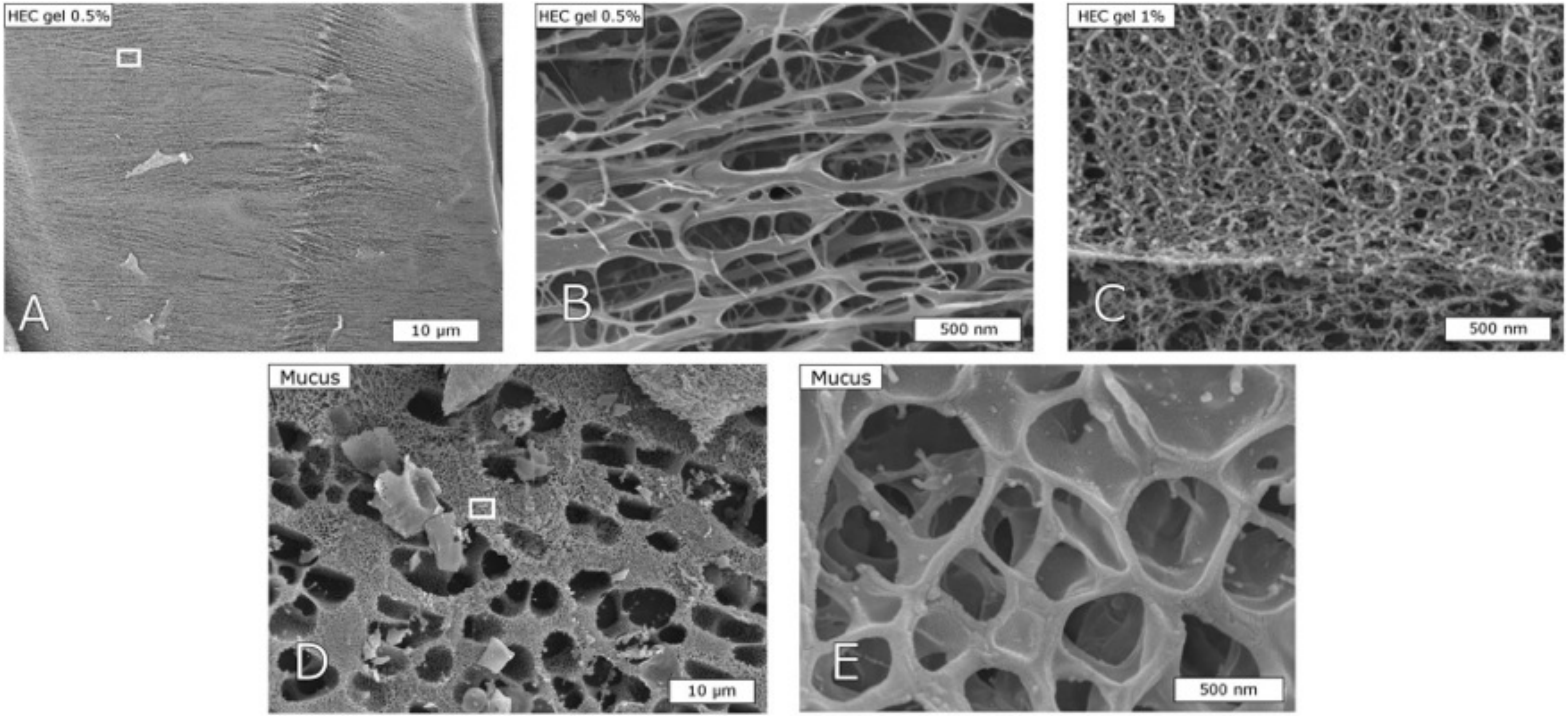
Scanning electron microscopy image of rabbit stomach mucosa showing surface cells and numerous bundles of fiber-like mucus strands. Scale bar 10 μm. Reproduced from Nunn et al. (51) with permission from Wiley.



The influence of mucus microstructure and rheology in *Helicobacter pylori* infection

Front. Immunol., 10 October 2013 | doi: 10.3389/fimmu.2013.00310

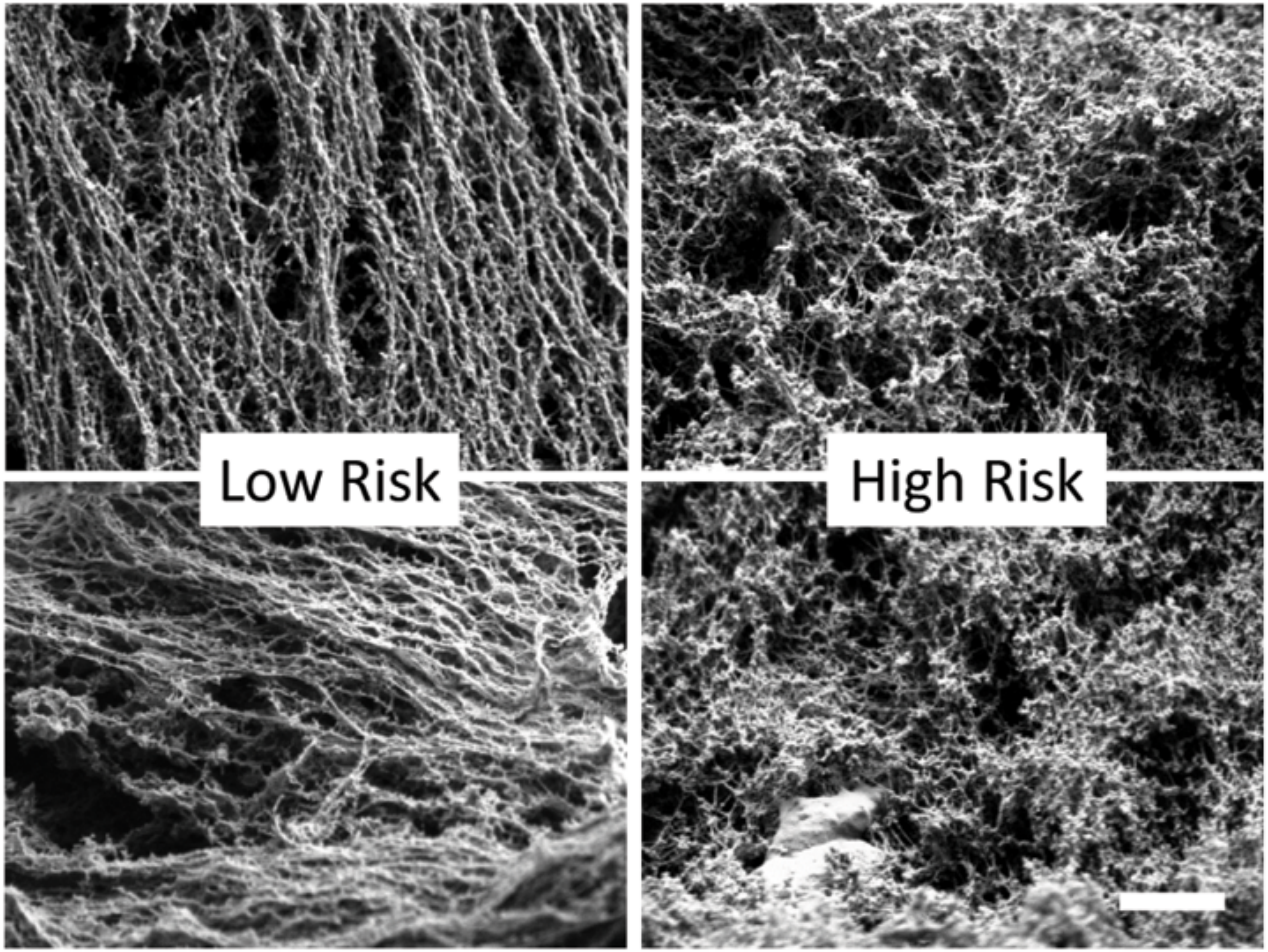
Representative cryo-SEM images of HEC hydrogels (A–C) and mucus (D and E).



Optical tweezers reveal relationship between microstructure and nanoparticle penetration of pulmonary mucus

Kirch J et al. PNAS 2012;109:18355-18360

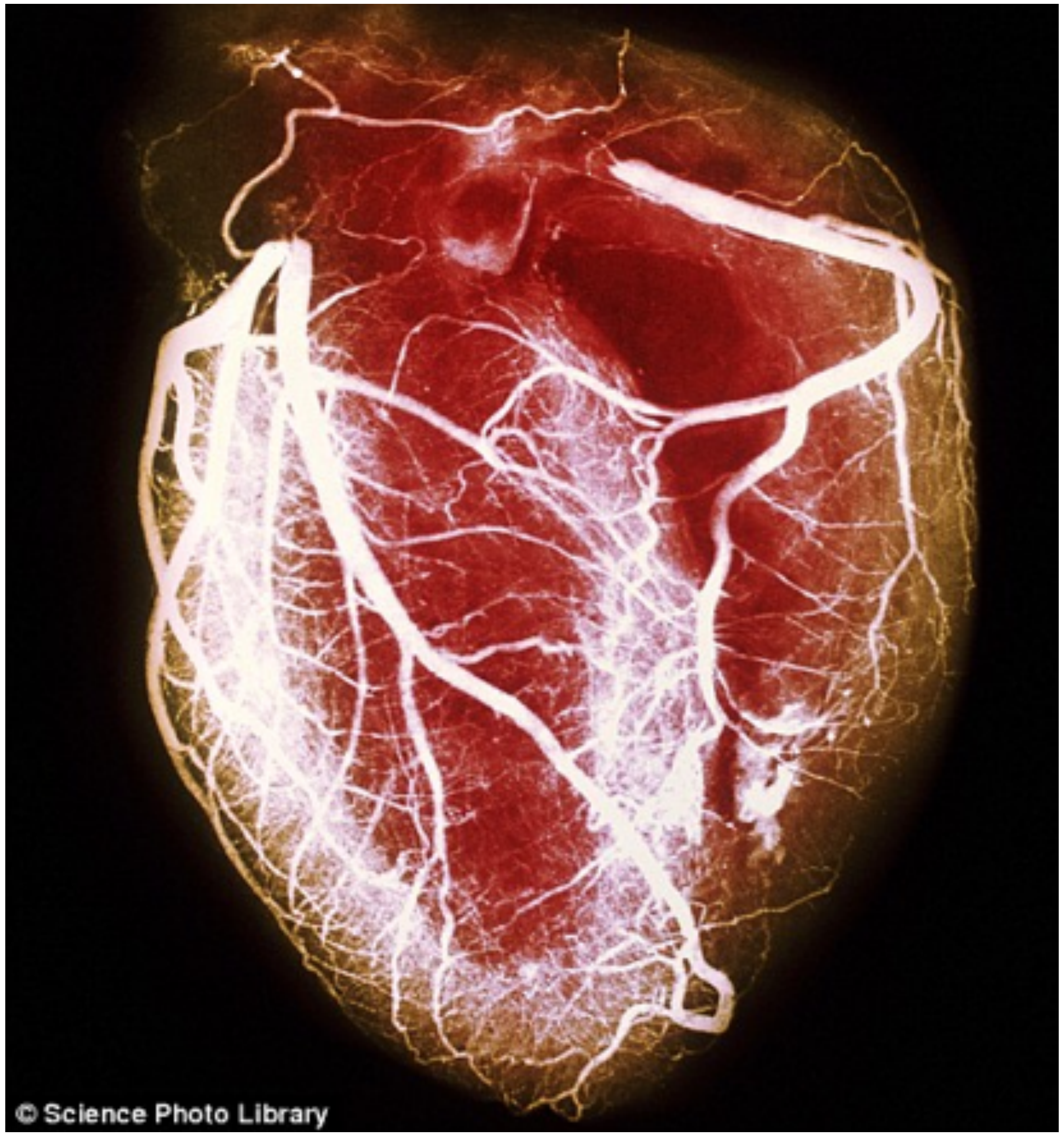
Cervical Mucus Properties Stratify Risk for Preterm Birth

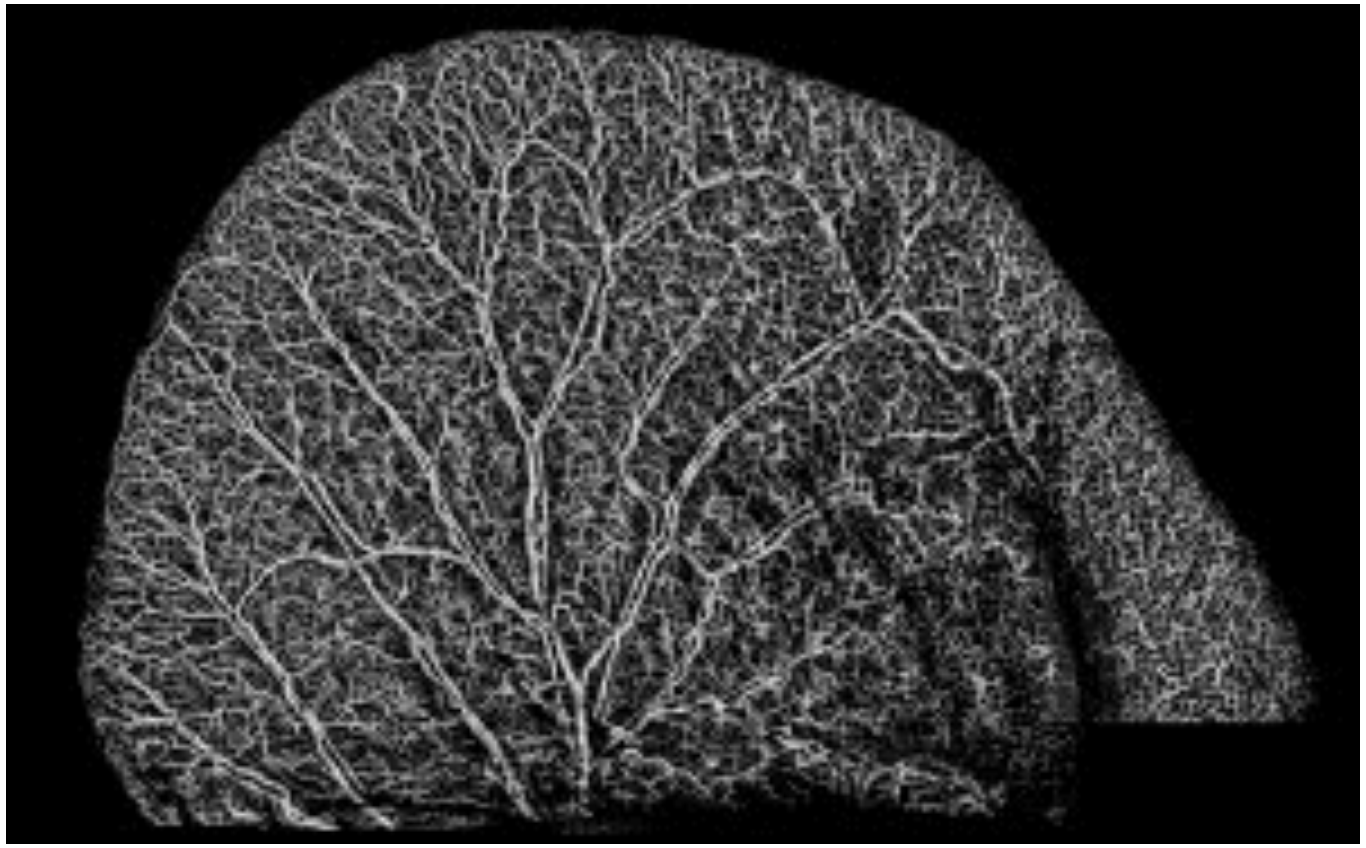


DOI: 10.1371/journal.pone.0069528

Transport networks

Blood vessels

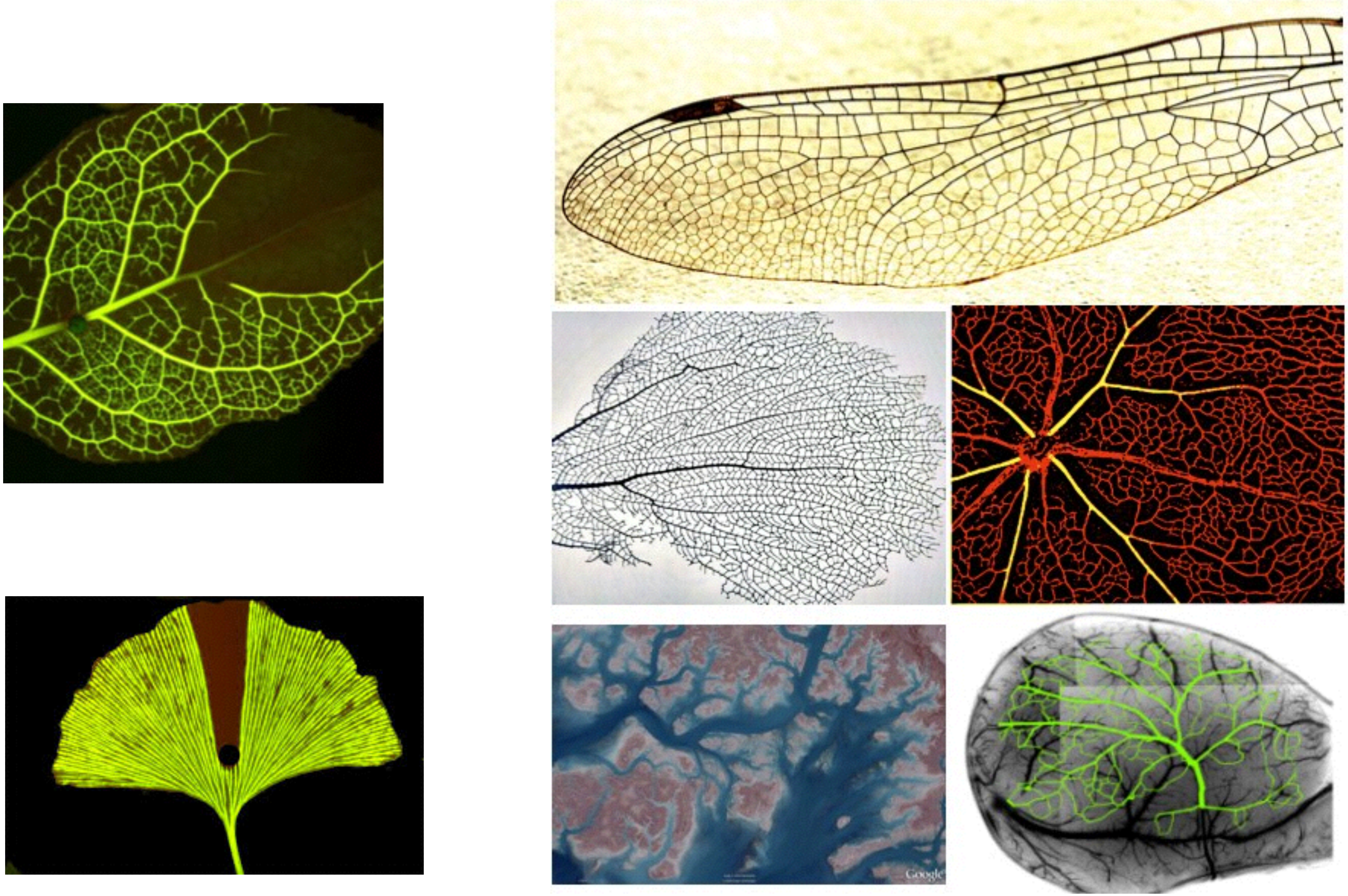




High-resolution microvasculature of a mouse ear from optical imaging microangiography. ***Siavash Yousefi, U of Washington***

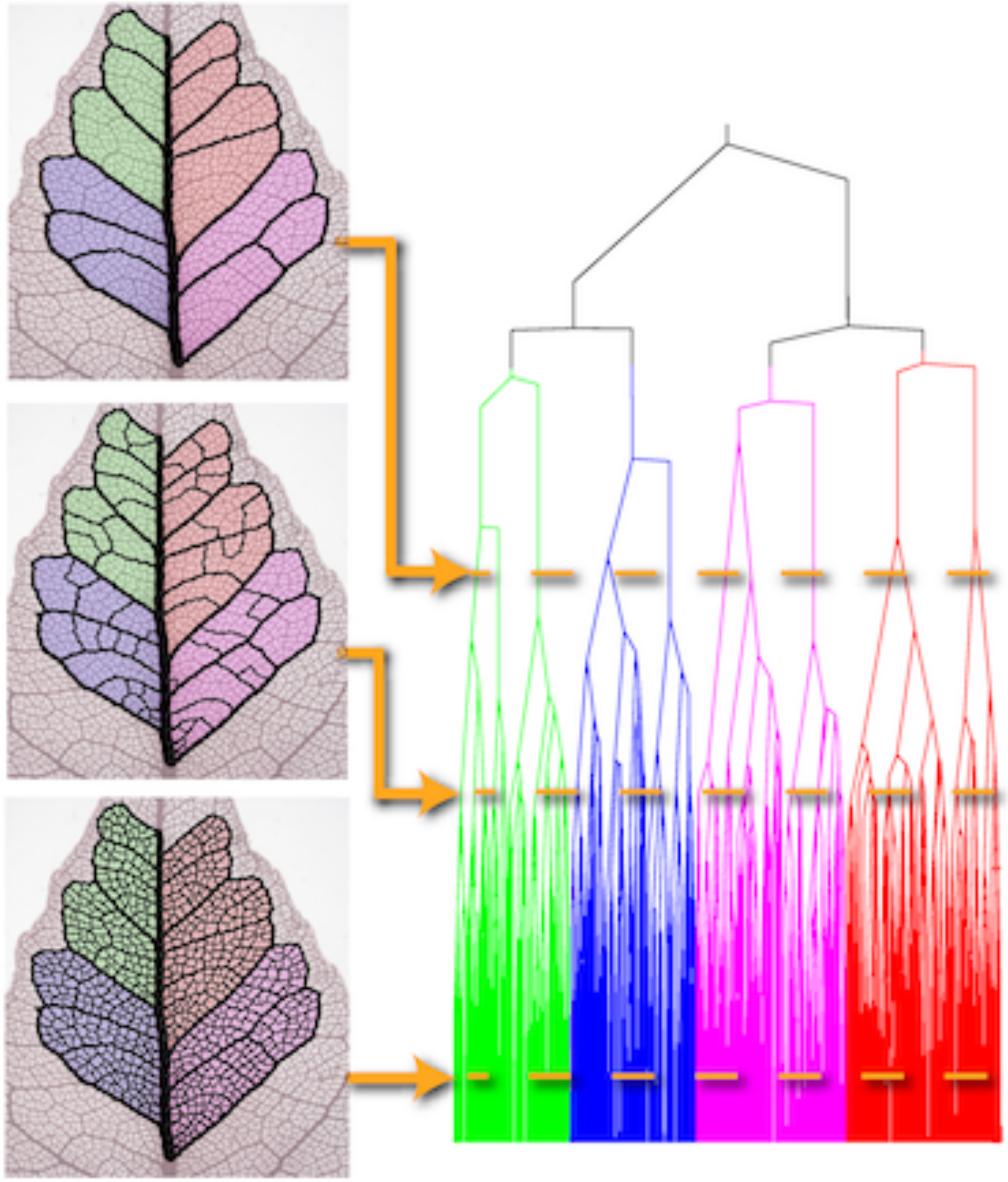
~25,000 miles of capillaries in an adult, each with an individual length of about 1 mm

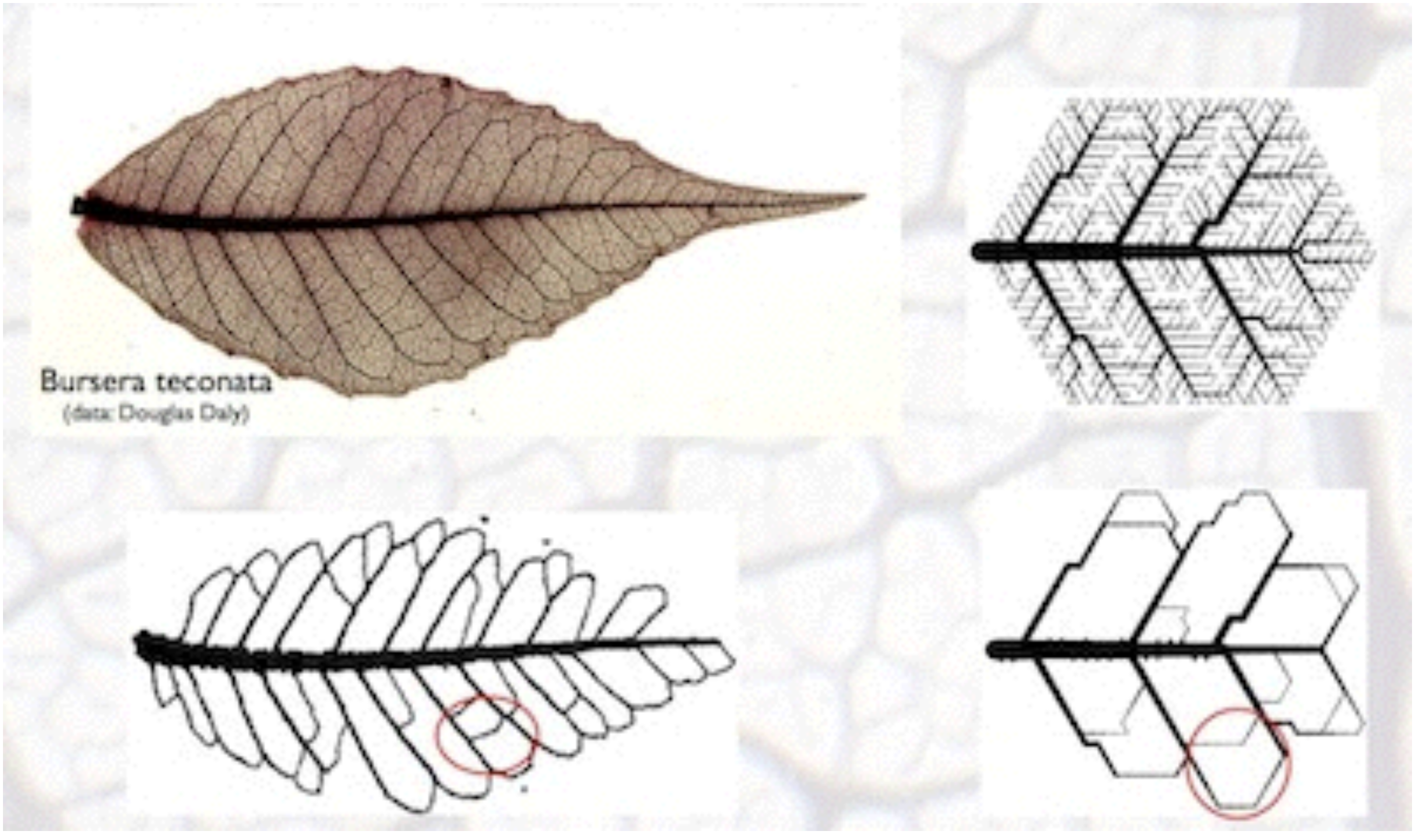
Liver vessels



Katifori lab, MPI Goettingen

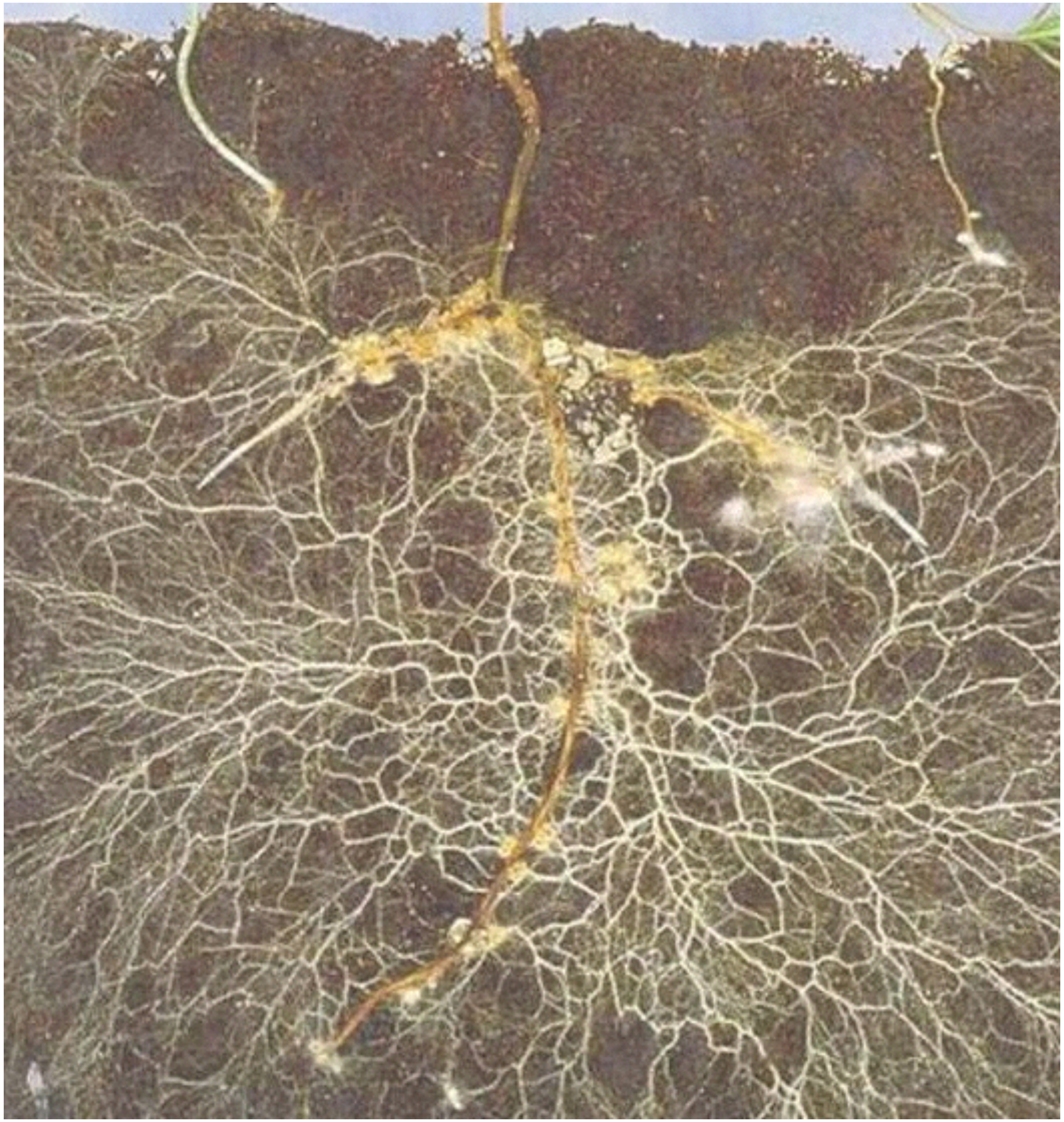
Network topology & optimization



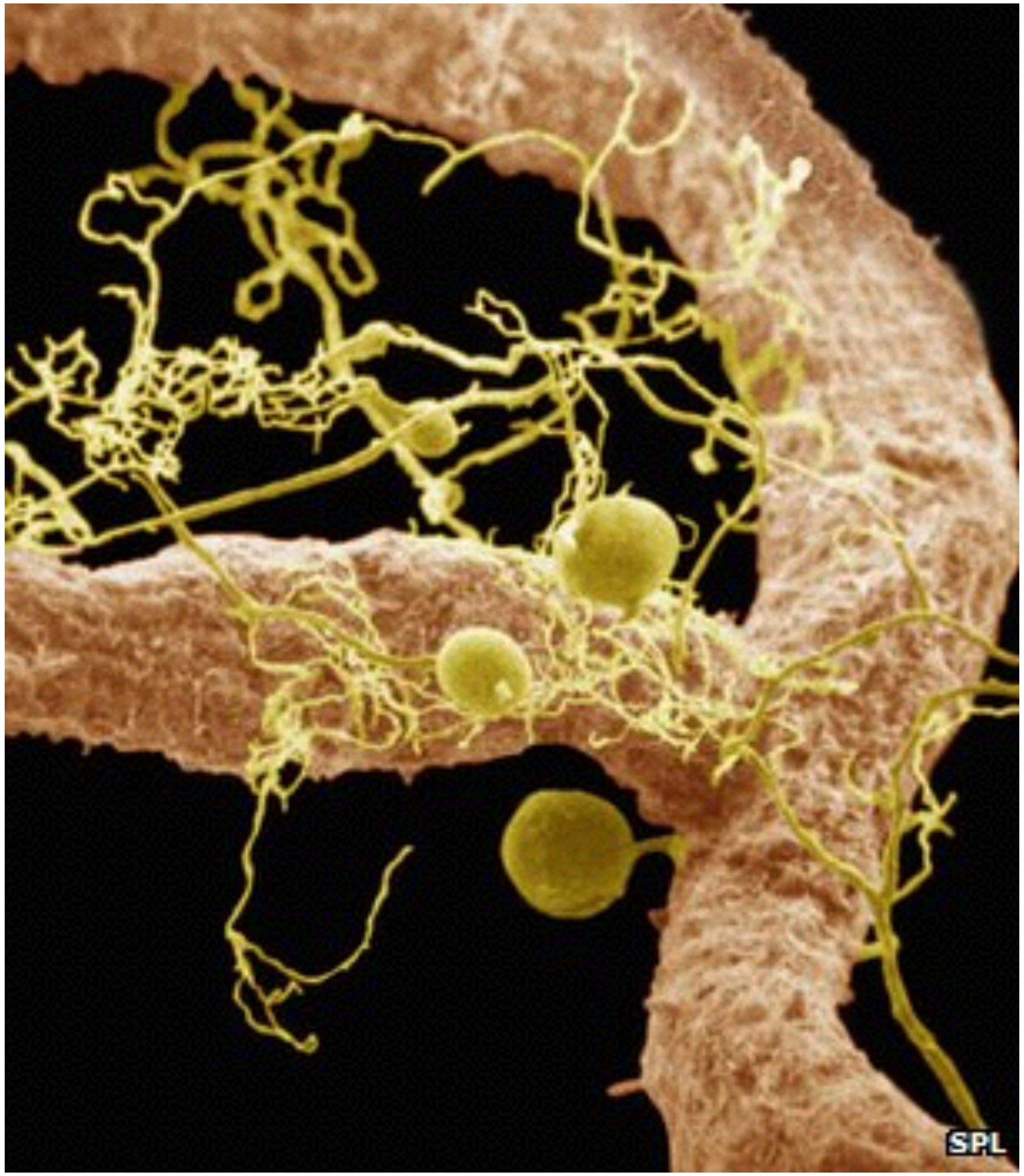


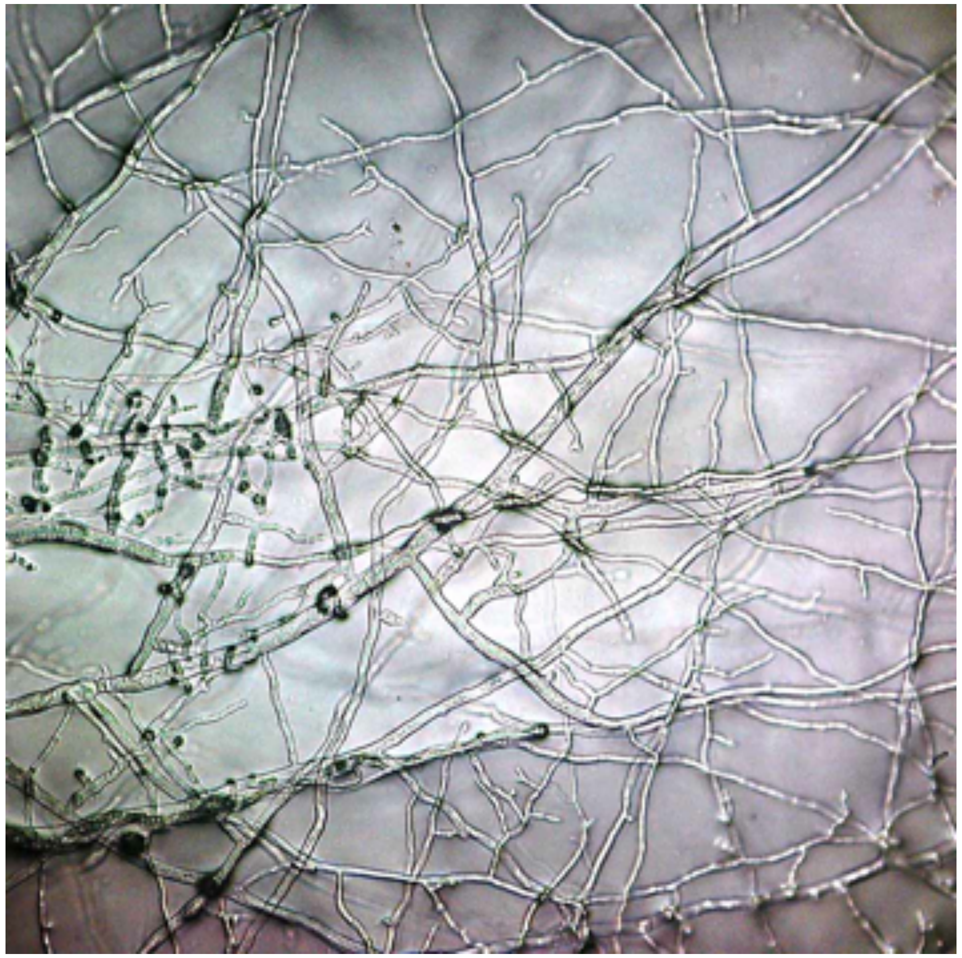
Katifori lab, MPI Goettingen

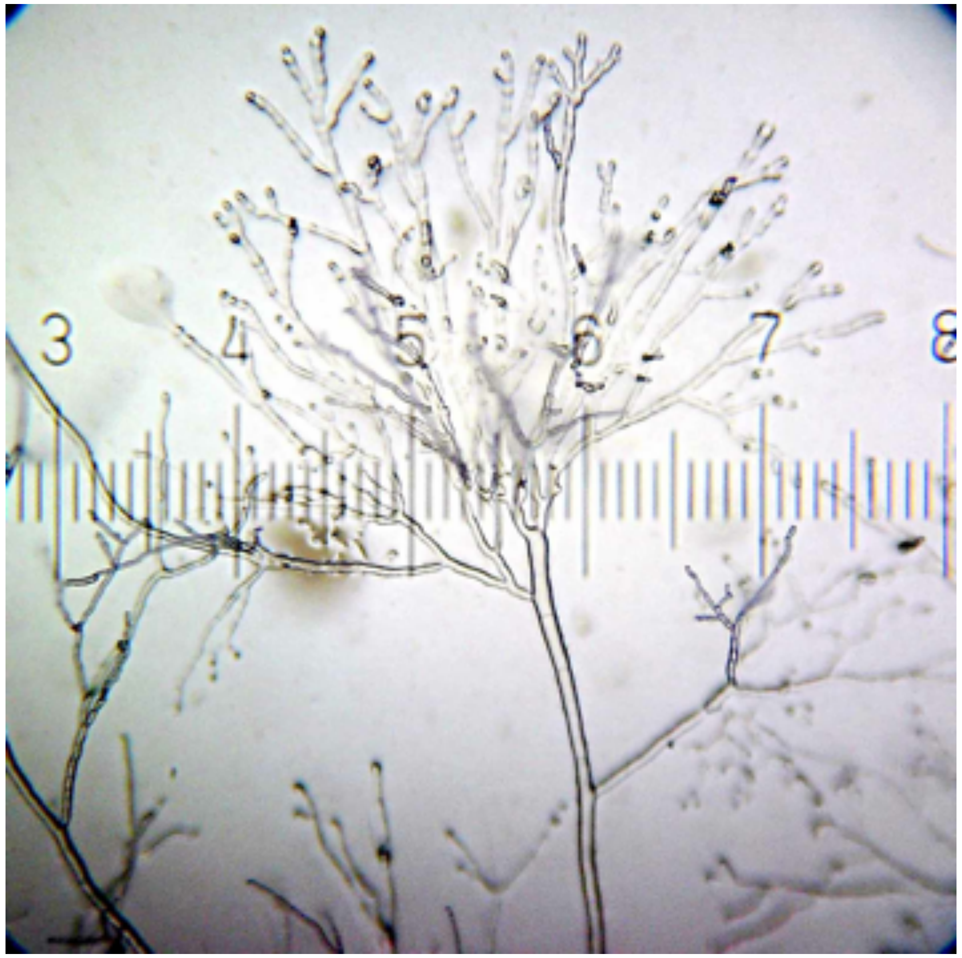
Root systems



Funghi







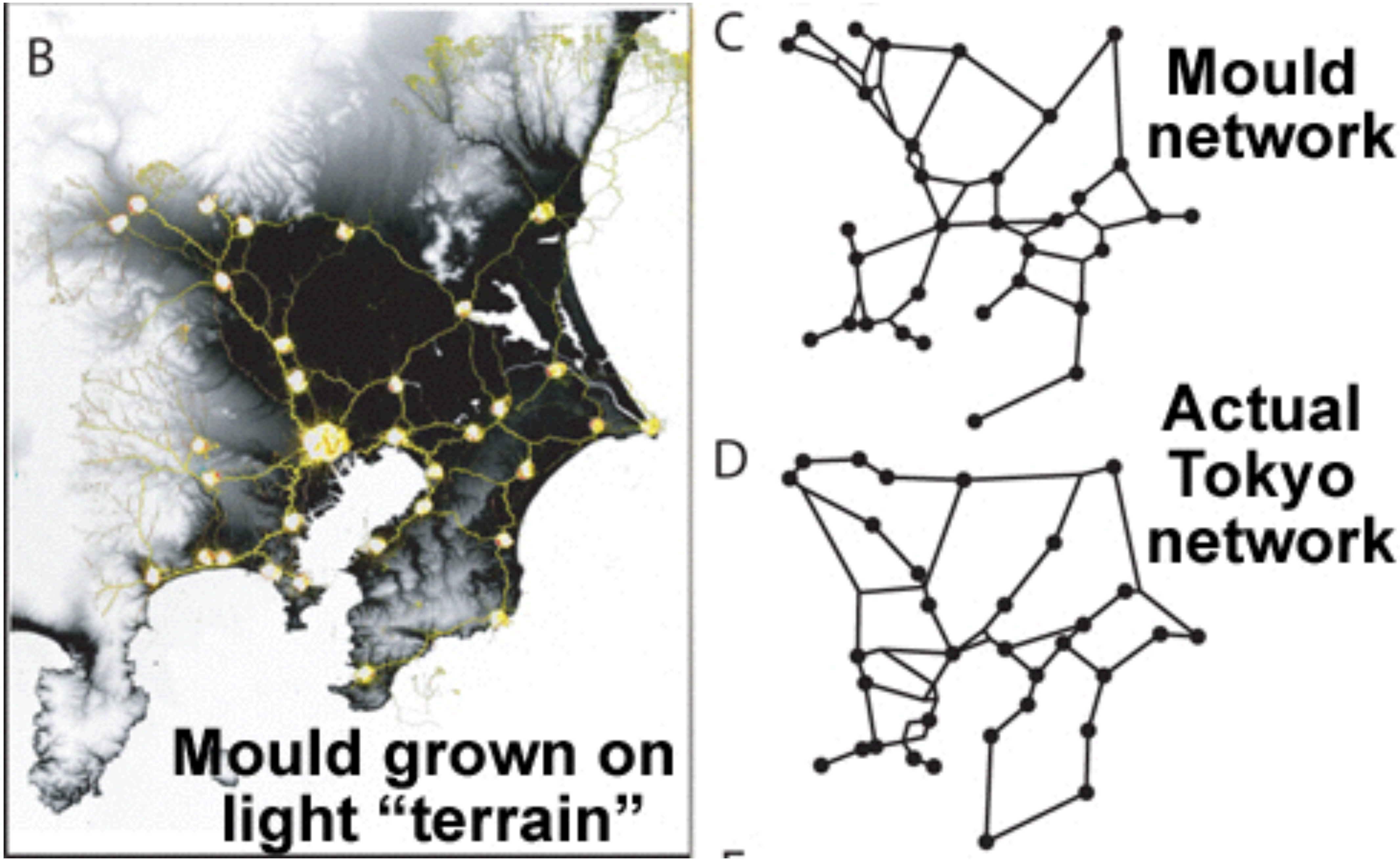
Mycorrhizae are mutualistic - they both need and are needed by the plants whose roots they inhabit

http://www.bbc.com/news/science-environment-22462855

Tokyo rail network by *Physarum plasmodium*

Tero *et al* (2010) Science

dunkel@math.mit.edu



Mould grown on light "terrain"

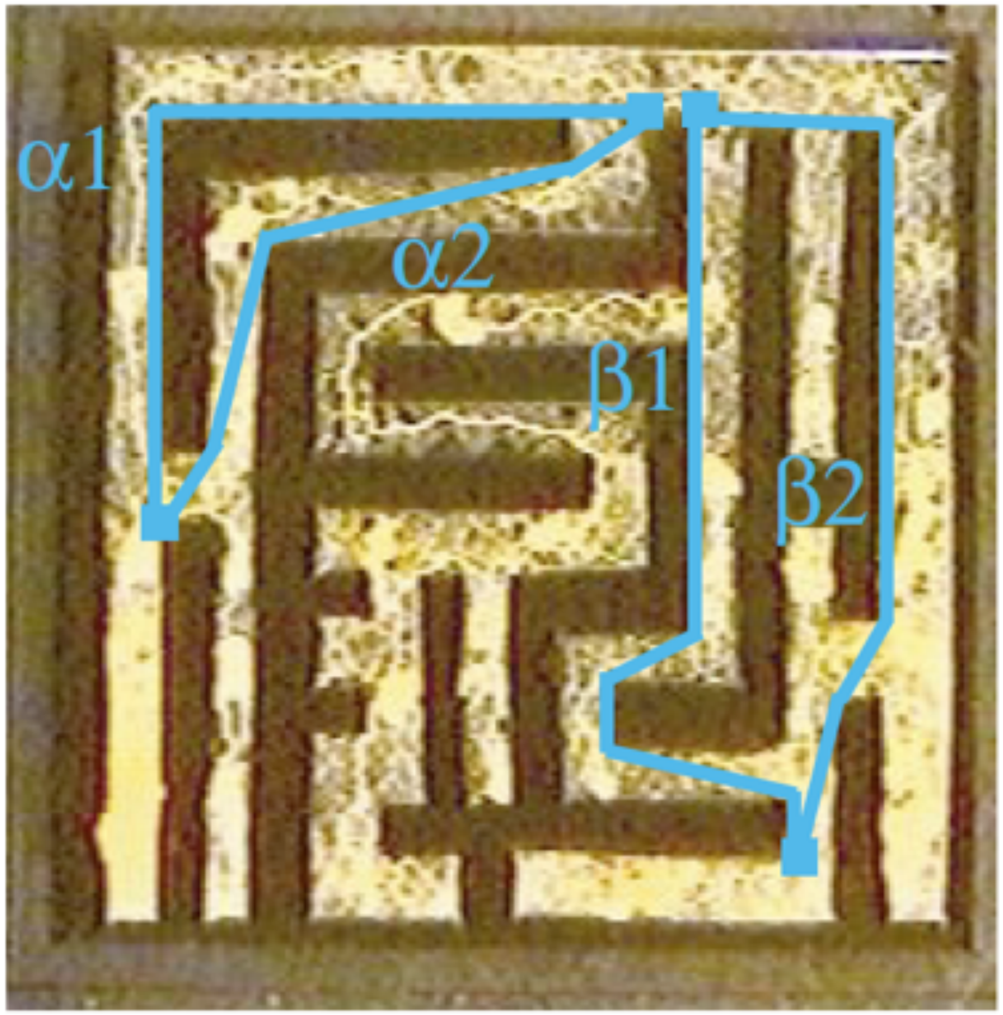
Mould network

Actual Tokyo network

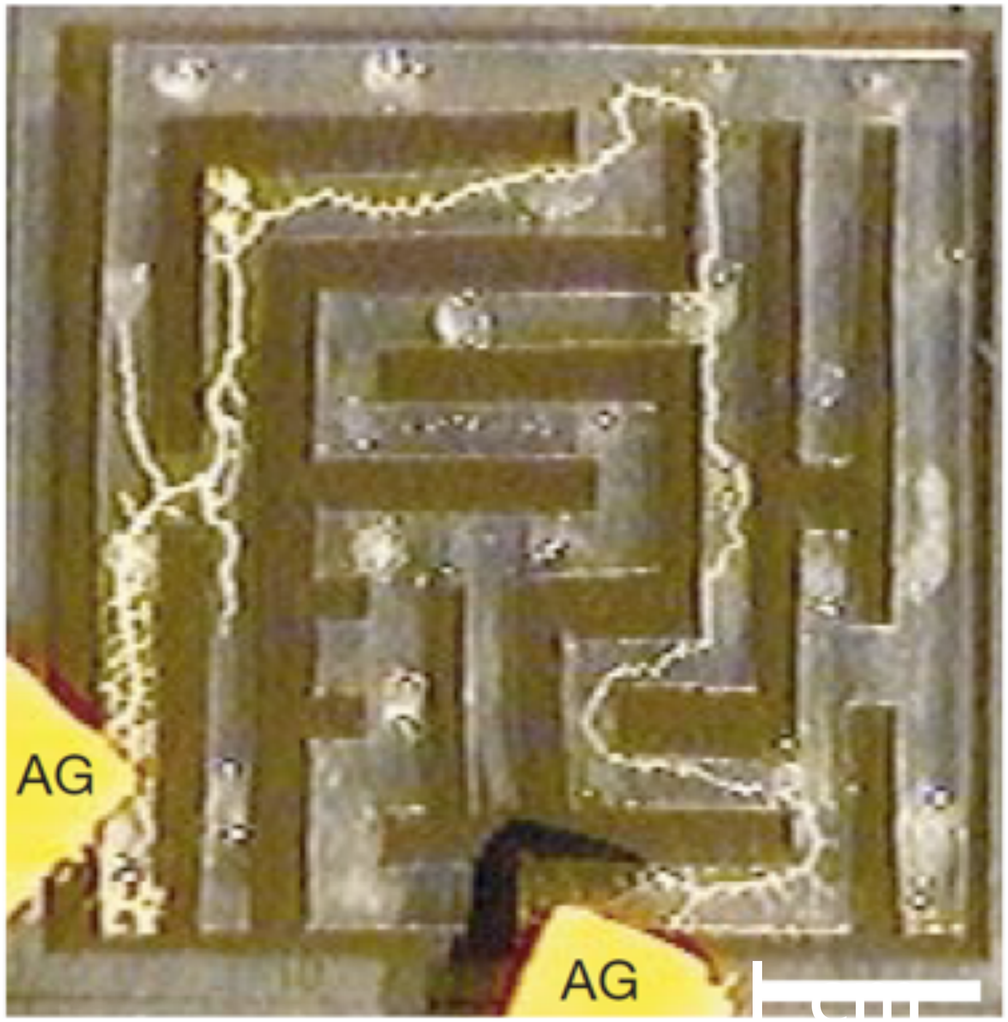
Tero *et al* (2010) Science

dunkel@math.mit.edu

Maze solving



time = 0

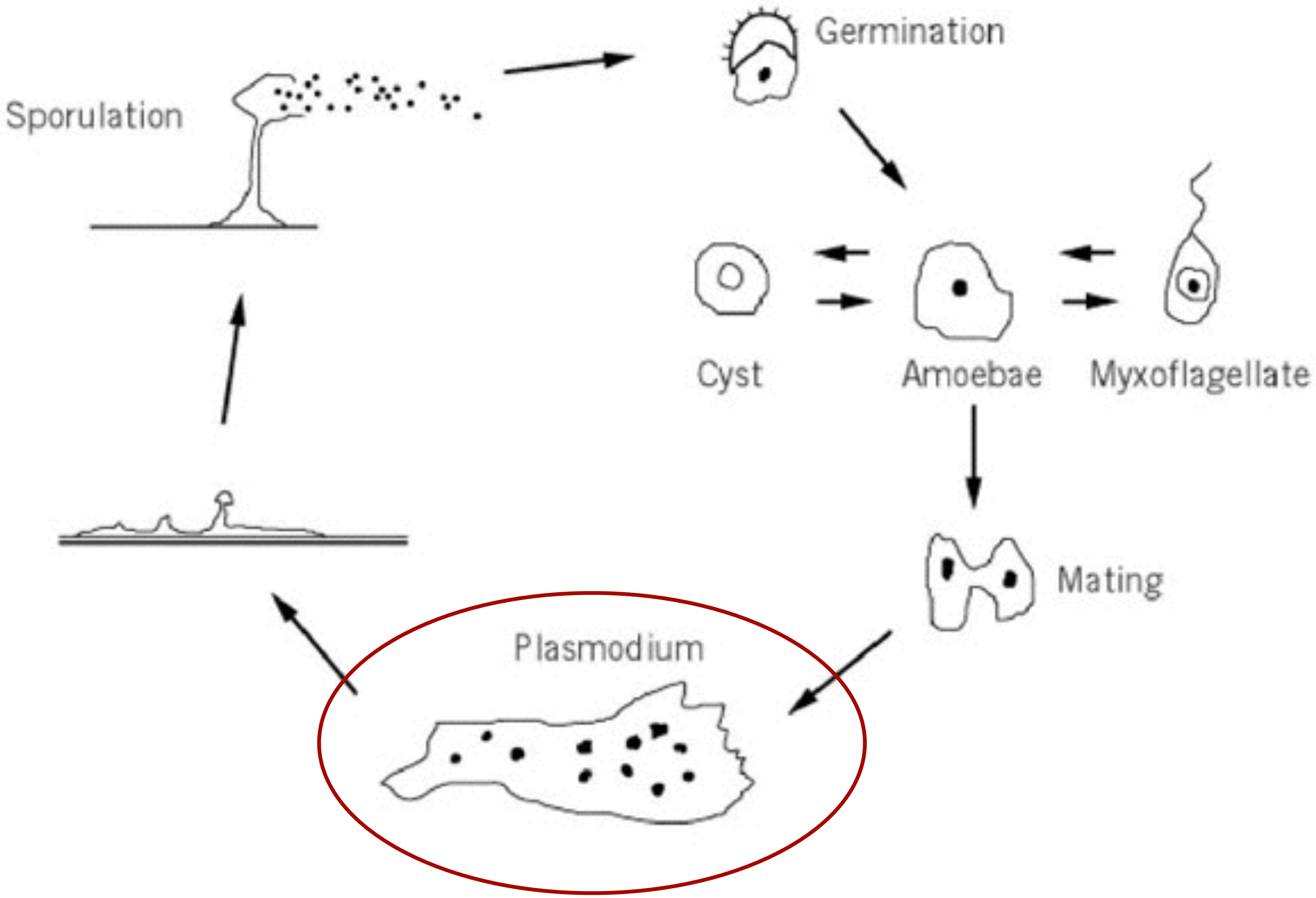


after 8 hours

Nakagaki *et al* (2000) Nature

dunkel@math.mit.edu

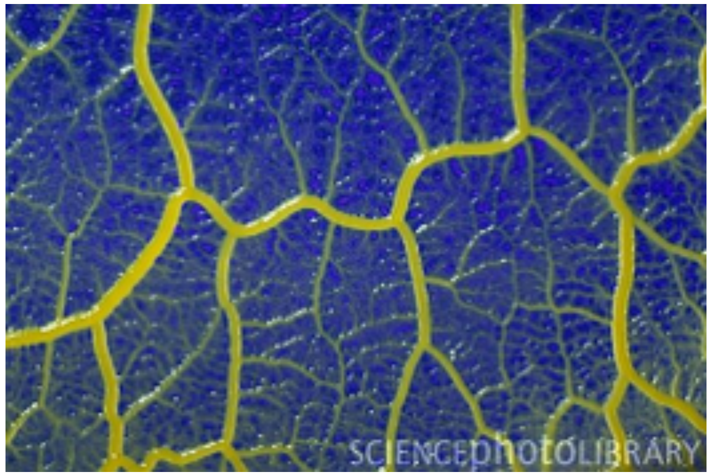
Physarum developmental cycle



dunkel@math.mit.edu

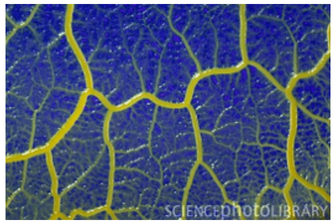
How does it work ?

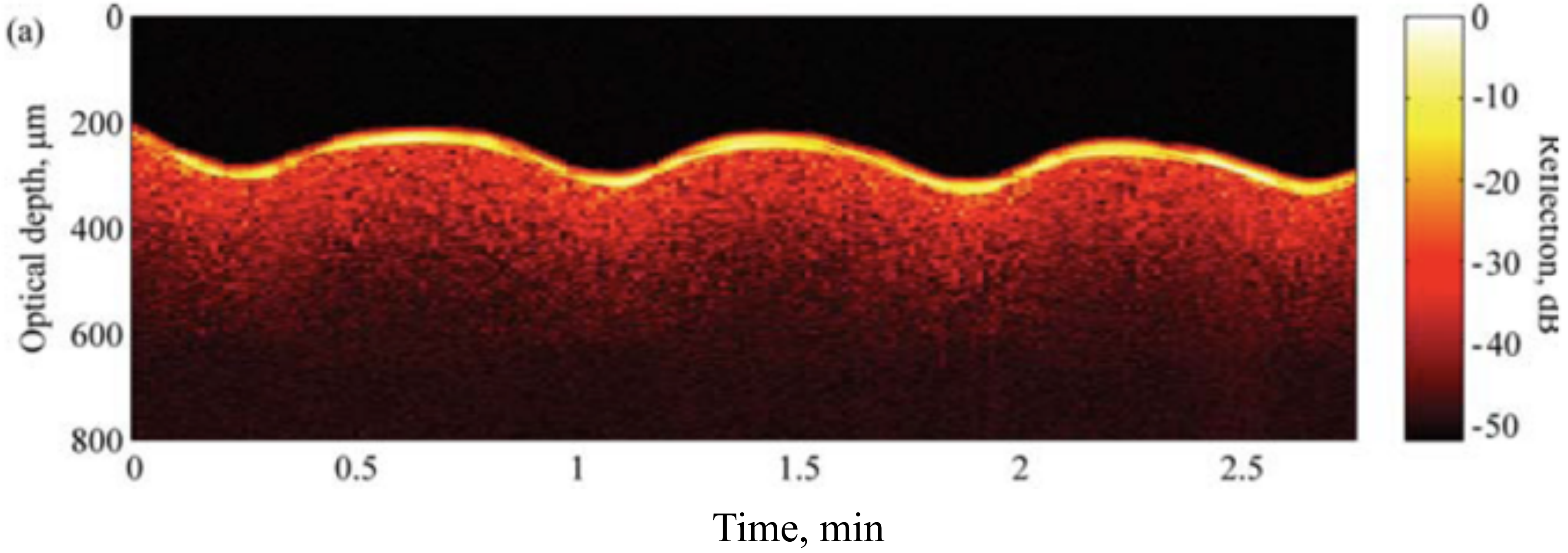
- acto-myosin contractions
- hydrodynamics
- noise (?)
- feedback (?)

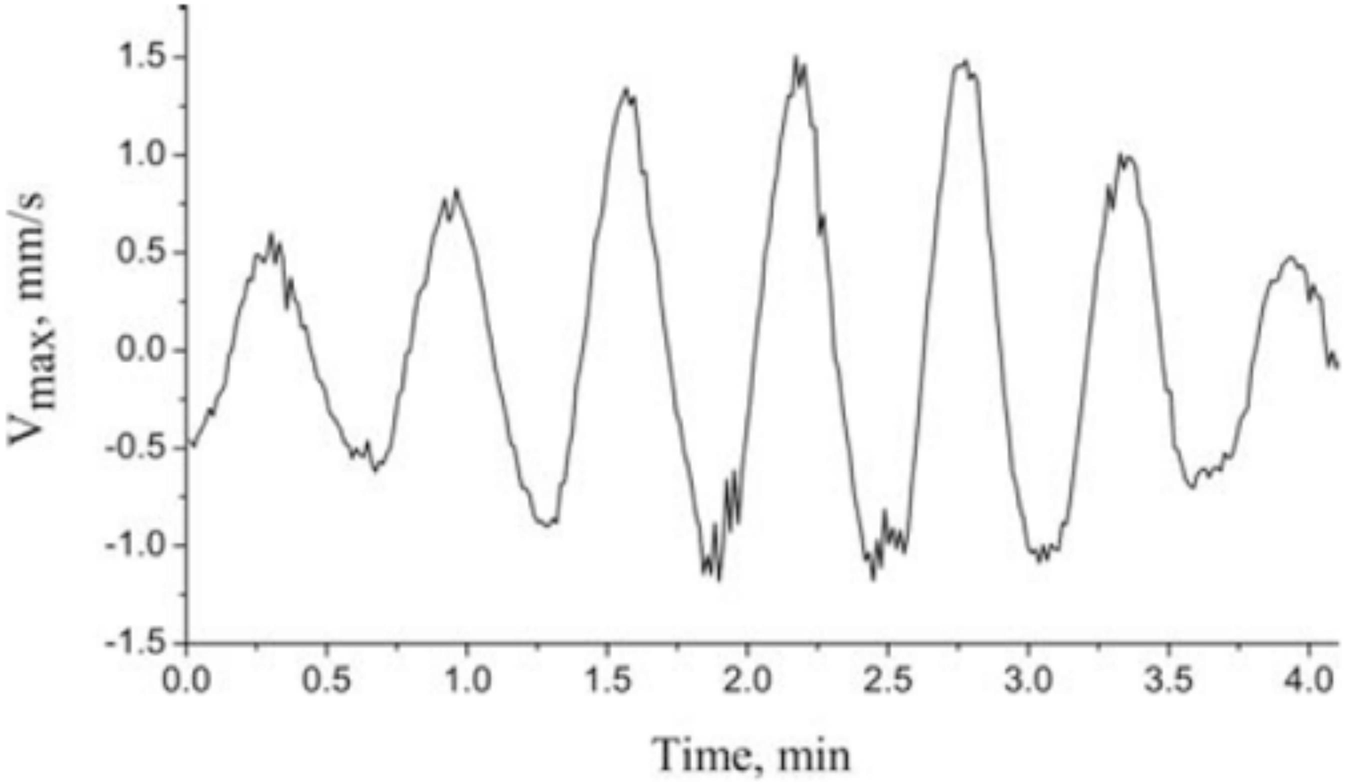


dunkel@math.mit.edu

‘Shuttle streaming’



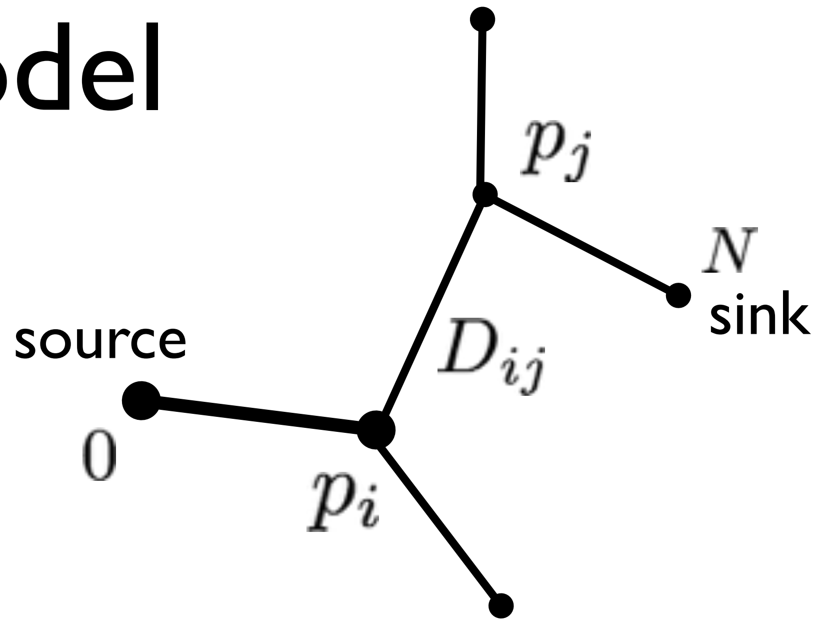


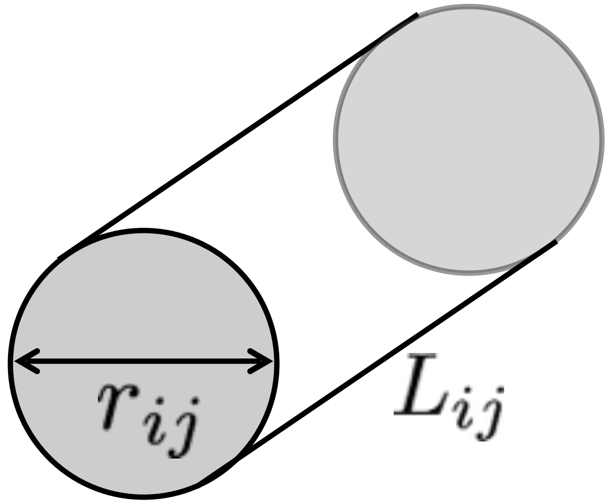


Bykov *et al* (2009)
J Biophoton

dunkel@math.mit.edu

‘Standard’ network model





flux

$$Q_{ij} = \frac{D_{ij}}{L_{ij}}(p_i - p_j)$$

conductivity

$$D_{ij} = \frac{\pi r_{ij}^4}{8\eta}$$

Kirchhoff’s law

$$\sum_{i \in \mathcal{N}(j)} Q_{ij} = \begin{cases} -I_0, & j = 1 \\ 0 & j = 2, \ldots N-1 \\ I_0, & j = N \end{cases}$$

dynamics

$$\frac{d}{dt} D_{ij} = |Q_{ij}| - rD_{ij}$$

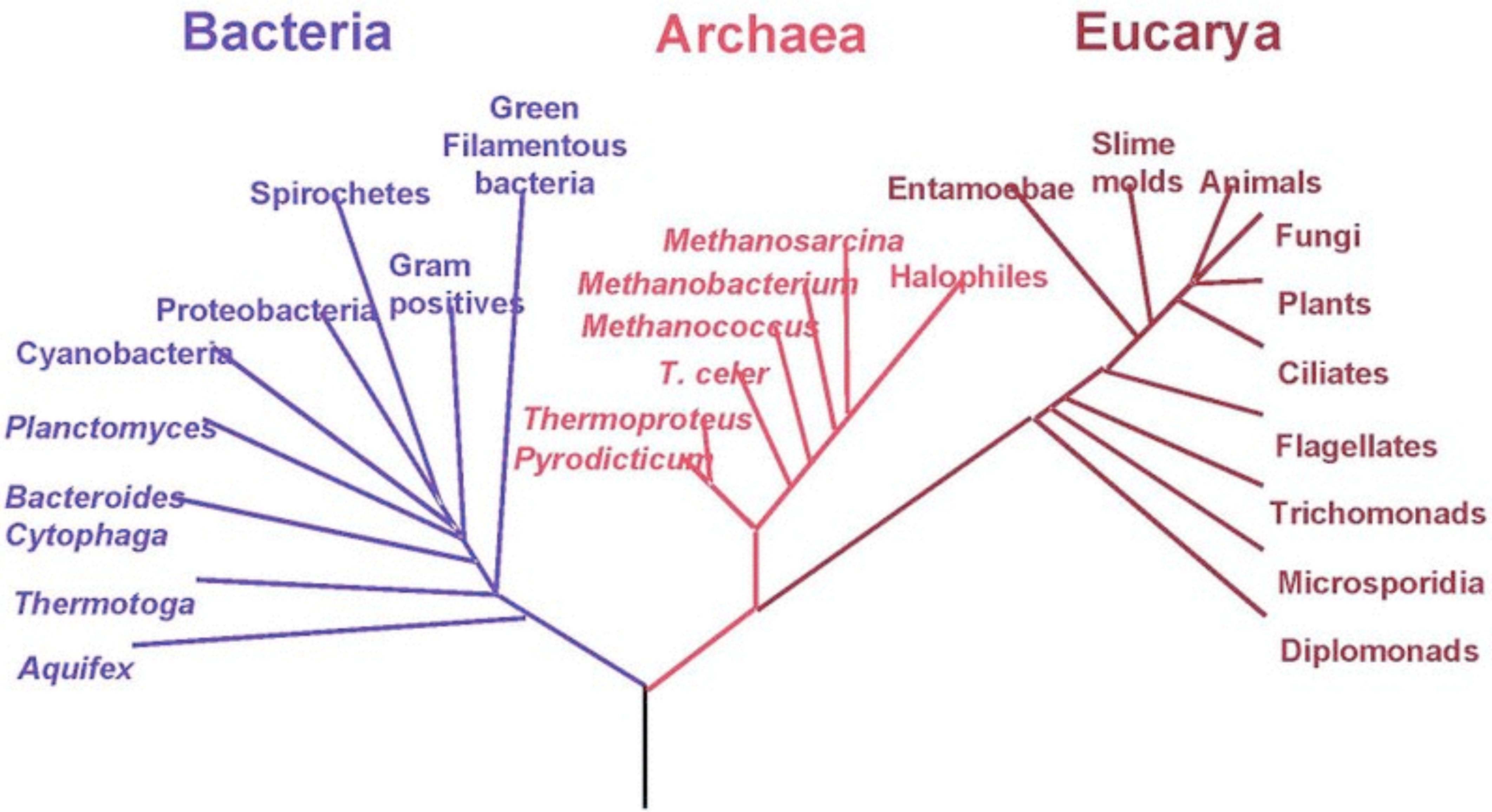
Nakagaki et al (2007) PRL

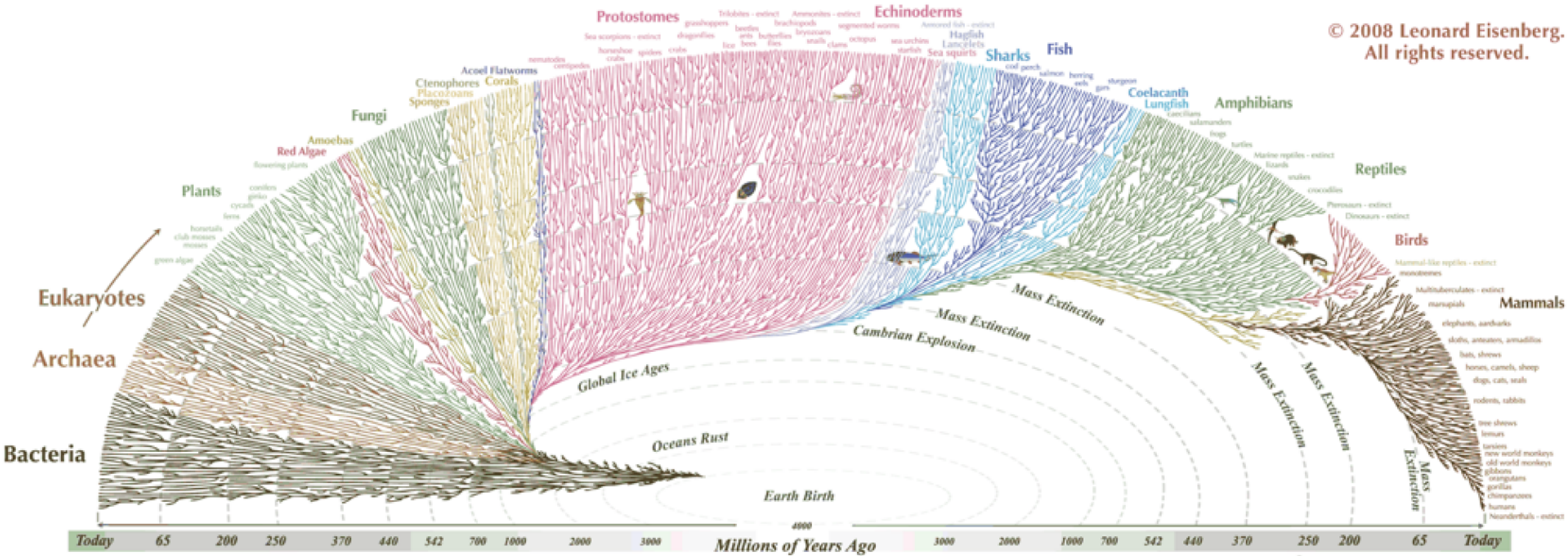
Bonifaci et al (2012) SODA

dunkel@math.mit.edu

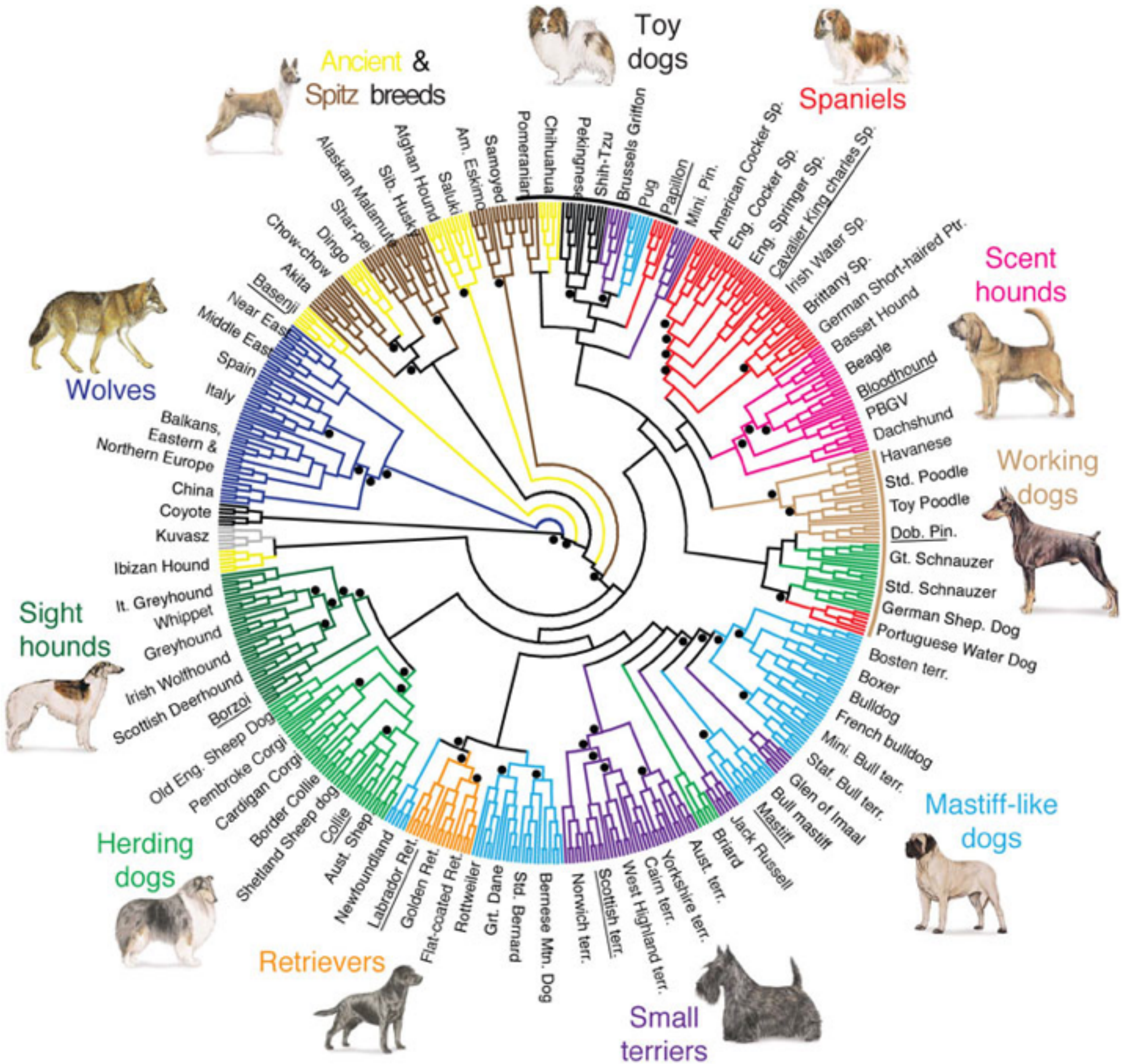
Tree structures

Phylogenetic Tree of Life





All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct



Phylogenetic genome mapping of dog breeds and their ancestors. Elaine A. Ostrander and Robert K. Wayne

Single cell studies

[redacted]

Cherry Tree Seasons

Springtime

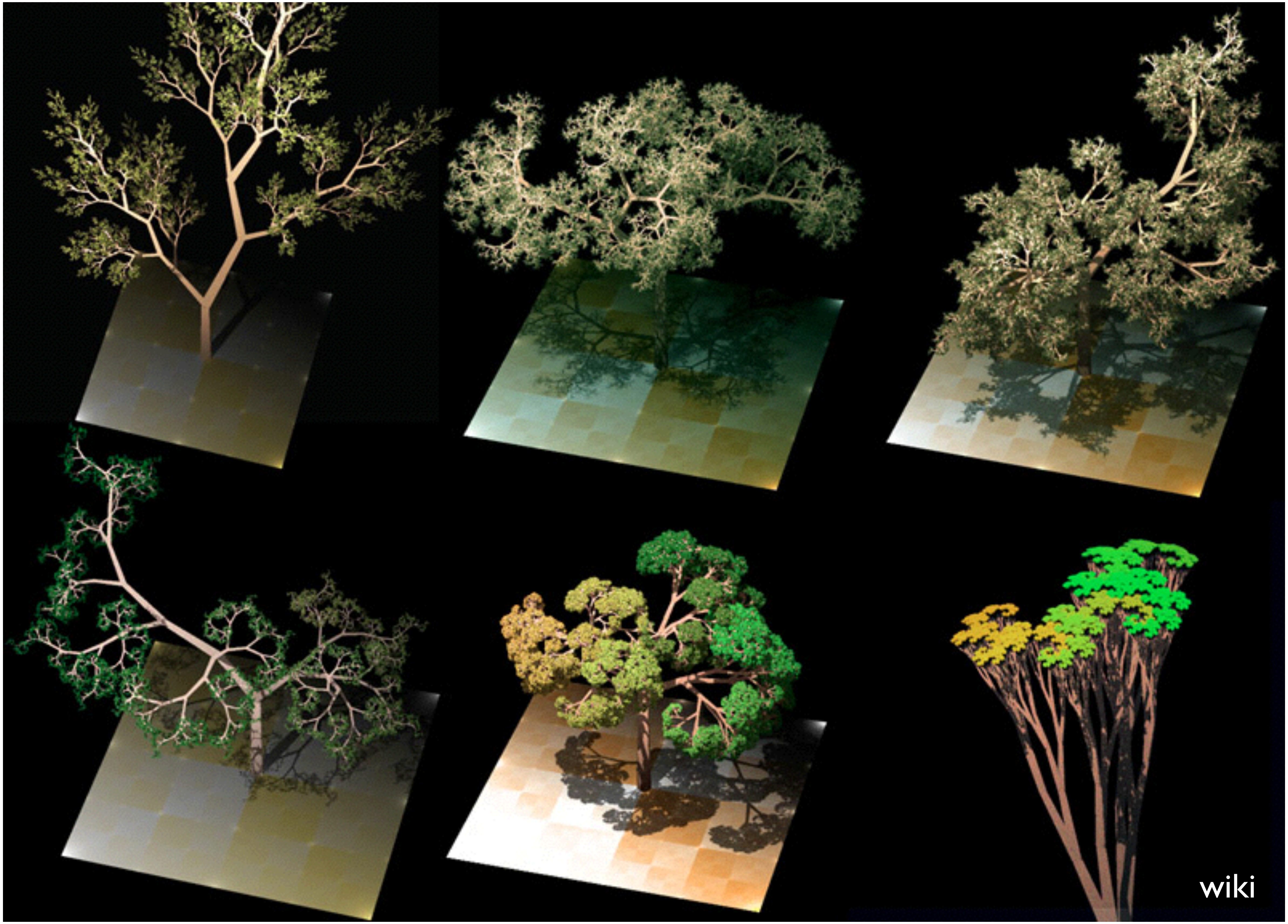
Summertime

Autumntime

Wintertime

L-Systems

http://algorithmicbotany.org/



wiki

L-Systems

Example 3: Cantor dust

variables : A B
constants : none
start : A {starting character string}
rules : (A → ABA), (B → BBB)
Let *A* mean "draw forward" and *B* mean "move forward".

This produces the famous Cantor's fractal set on a real straight line **R**.

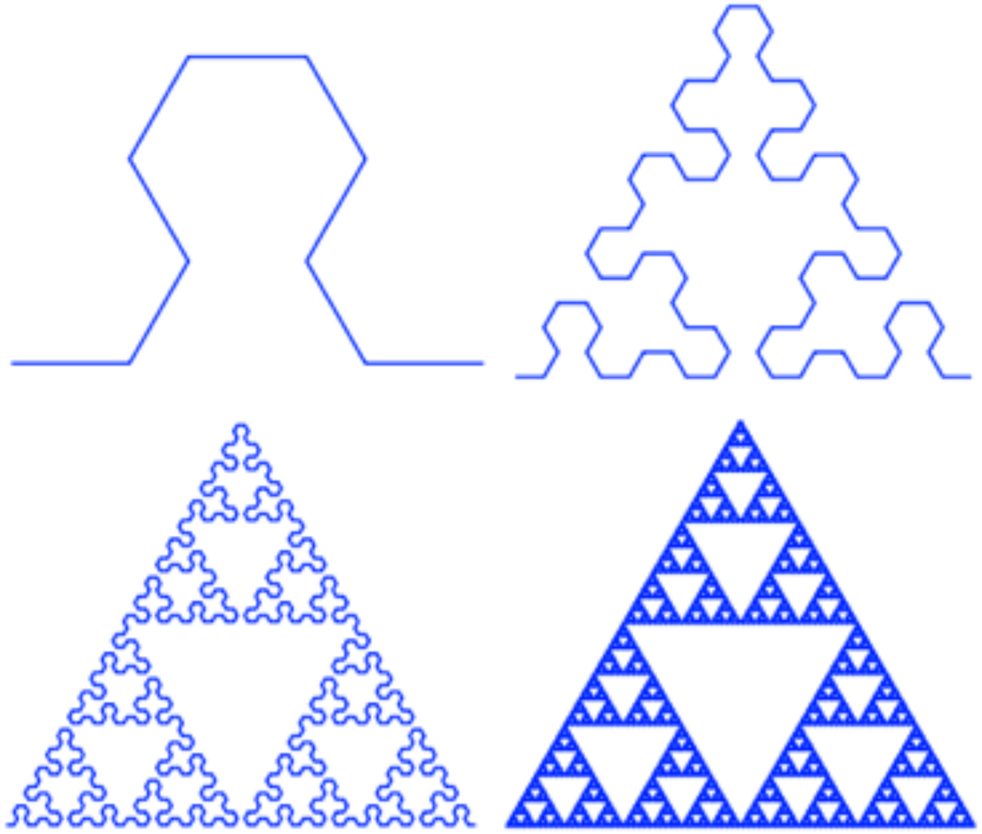
L-Systems

Example 5: Sierpinski triangle

The Sierpinski triangle drawn using an L-system.

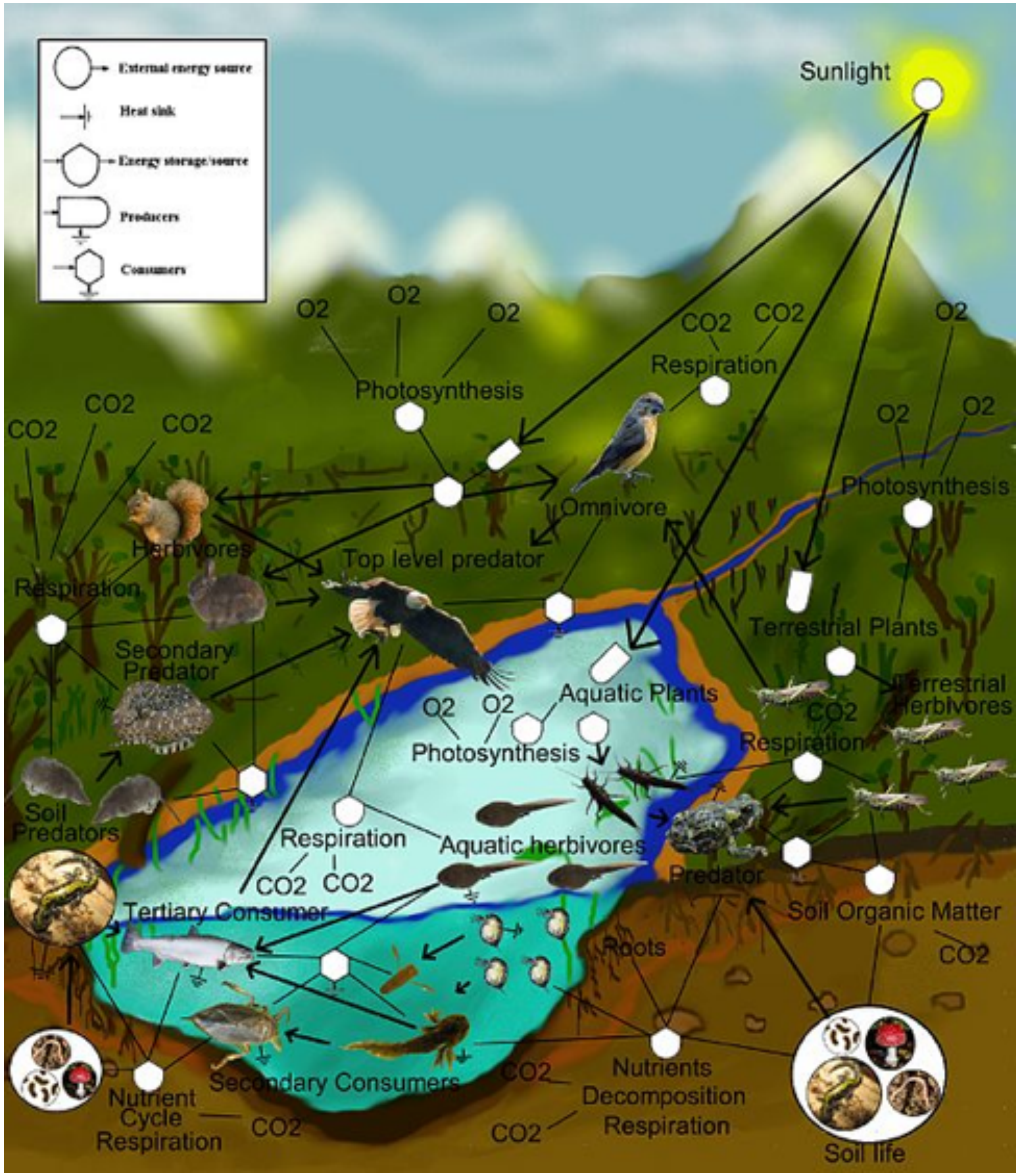
variables : A B
constants : + −
start : A
rules : (A → B−A−B), (B → A+B+A)
angle : 60°

Here, A and B both mean "draw forward", + means "turn left by angle", and − means "turn right by angle". The angle changes sign at each iteration so that the base of the triangular shapes are always in the bottom (otherwise the bases would alternate between top and bottom).

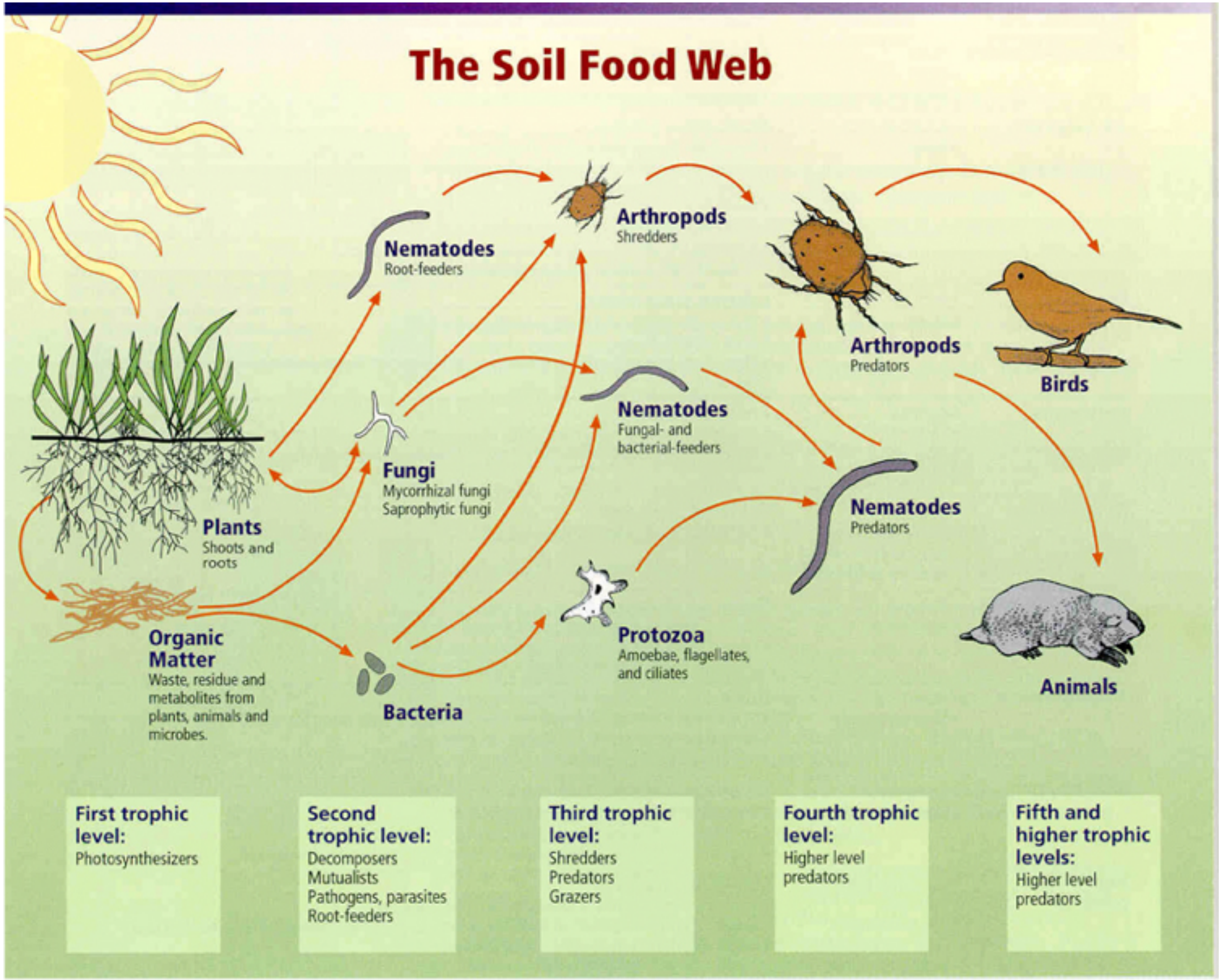


Ecological networks

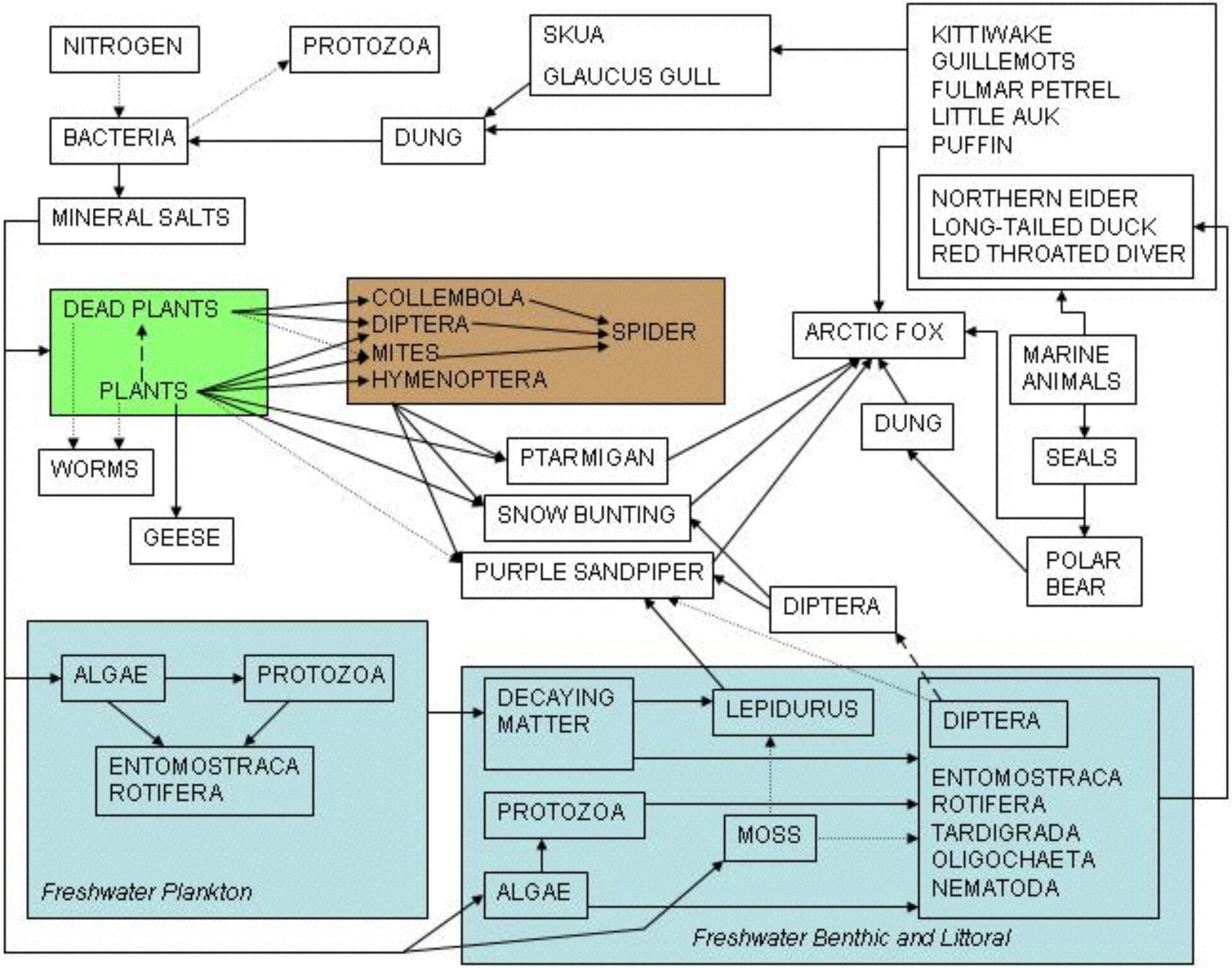
Food web



dunkel@math.mit.edu

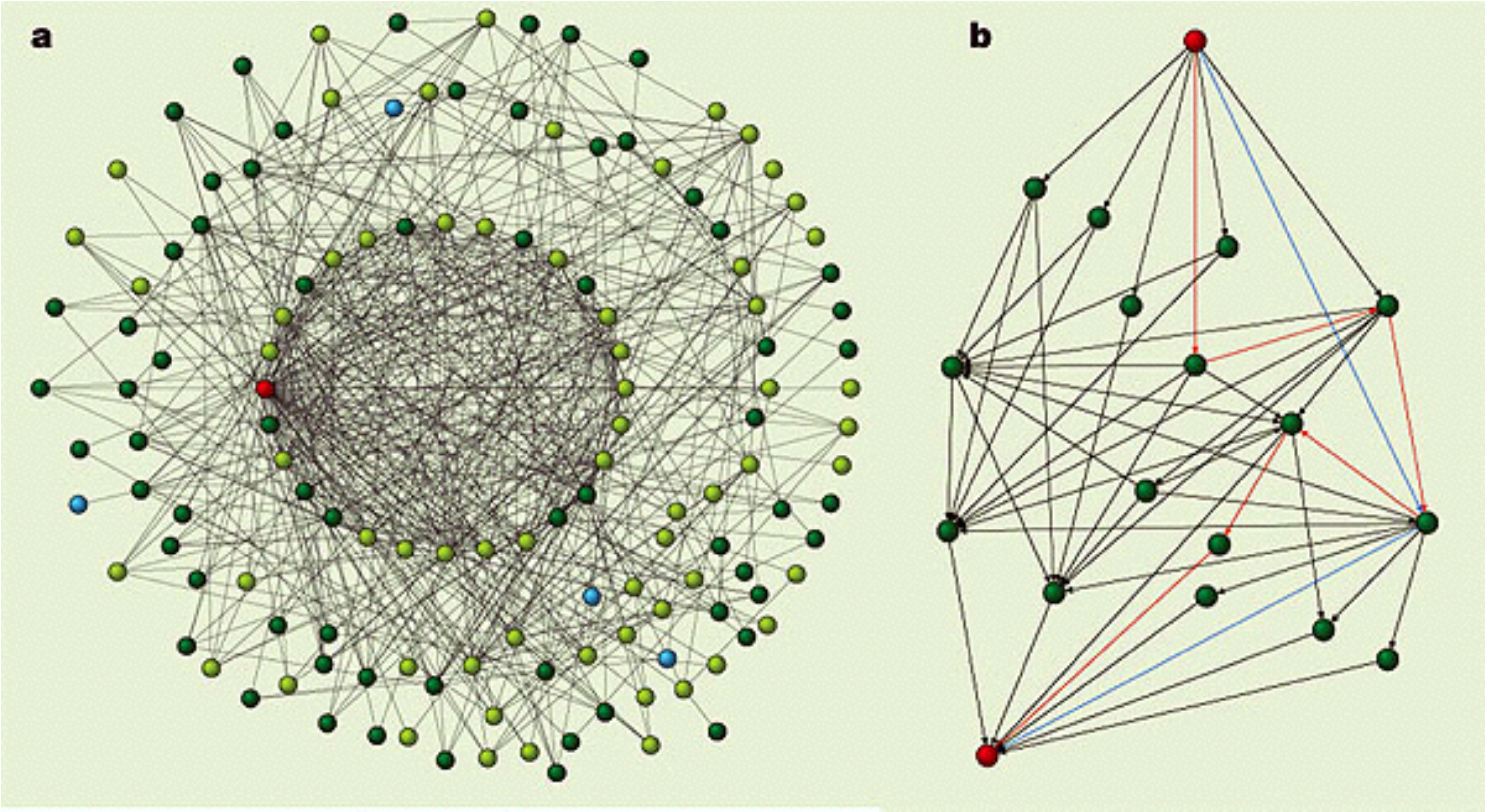


Relationships between soil food web, plants, organic matter, and birds and mammals
Image courtesy of USDA Natural Resources Conservation Service
http://soils.usda.gov/sqi/soil_quality/soil_biology/soil_food_web.html.


NITROGEN
PROTOZOA
BACTERIA
MINERAL SALTS
DUNG
SKUA
GLAUCUS GULL
KITTIWAKE
GUILLEMOTS
FULMAR PETREL
LITTLE AUK
PUFFIN
NORTHERN EIDER
LONG-TAILED DUCK
RED THROATED DIVER
DEAD PLANTS
PLANTS
COLLEMBOLA
DIPTERA
MITES
HYMENOPTERA
SPIDER
ARCTIC FOX
MARINE ANIMALS
WORMS
GEESE
PTARMIGAN
SNOW BUNTING
PURPLE SANDPIPER
DUNG
SEALS
POLAR BEAR
DIPTERA
ALGAE
PROTOZOA
ENTOMOSTRACA
ROTIFERA
Freshwater Plankton
DECAYING MATTER
LEPIDURUS
DIPTERA
PROTOZOA
MOSS
ALGAE
ENTOMOSTRACA
ROTIFERA
TARDIGRADA
OLIGOCHAETA
NEMATODA
Freshwater Benthic and Littoral

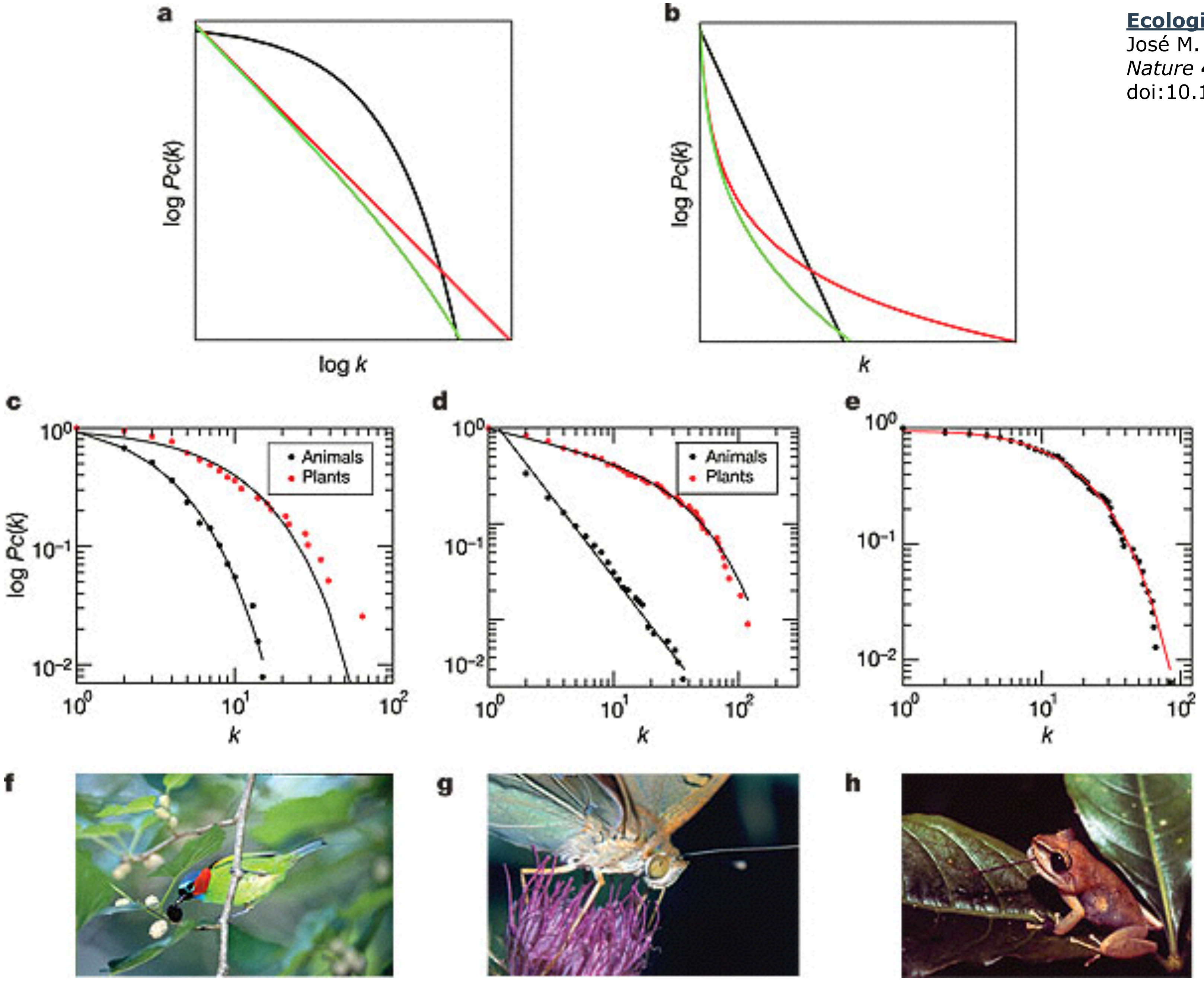
Shortest paths in a complex food web

Ecological networks and their fragility
José M. Montoya, Stuart L. Pimm and Ricard V. Solé
Nature **442**, 259-264(20 July 2006)
doi:10.1038/nature04927



The Ythan estuary food web[66,73]. **a**, Node colour indicates the length of the shortest path linking the most connected species (the flounder *Platichthys flesus*, in red) and each other species from the network. (The trophic direction of the links—what eats what—is ignored). Dark green, species are one link apart; light green, two links; and blue, three links. The central circle represents the densest sub-web[75], which consists of 28 species with at least 7 links with the rest of the species from the sub-web. This sub-web contributes most to the observed clustering. **b**, Food chains between basal species of *Enteromorpha* (red node at the bottom) and the top predator, the cormorant *Phalacrocorax carbo* (red node at the top). Links corresponding to the shortest path connecting them are in blue (2-links), and those corresponding to the longest food chain between these two species are in red (6-links).

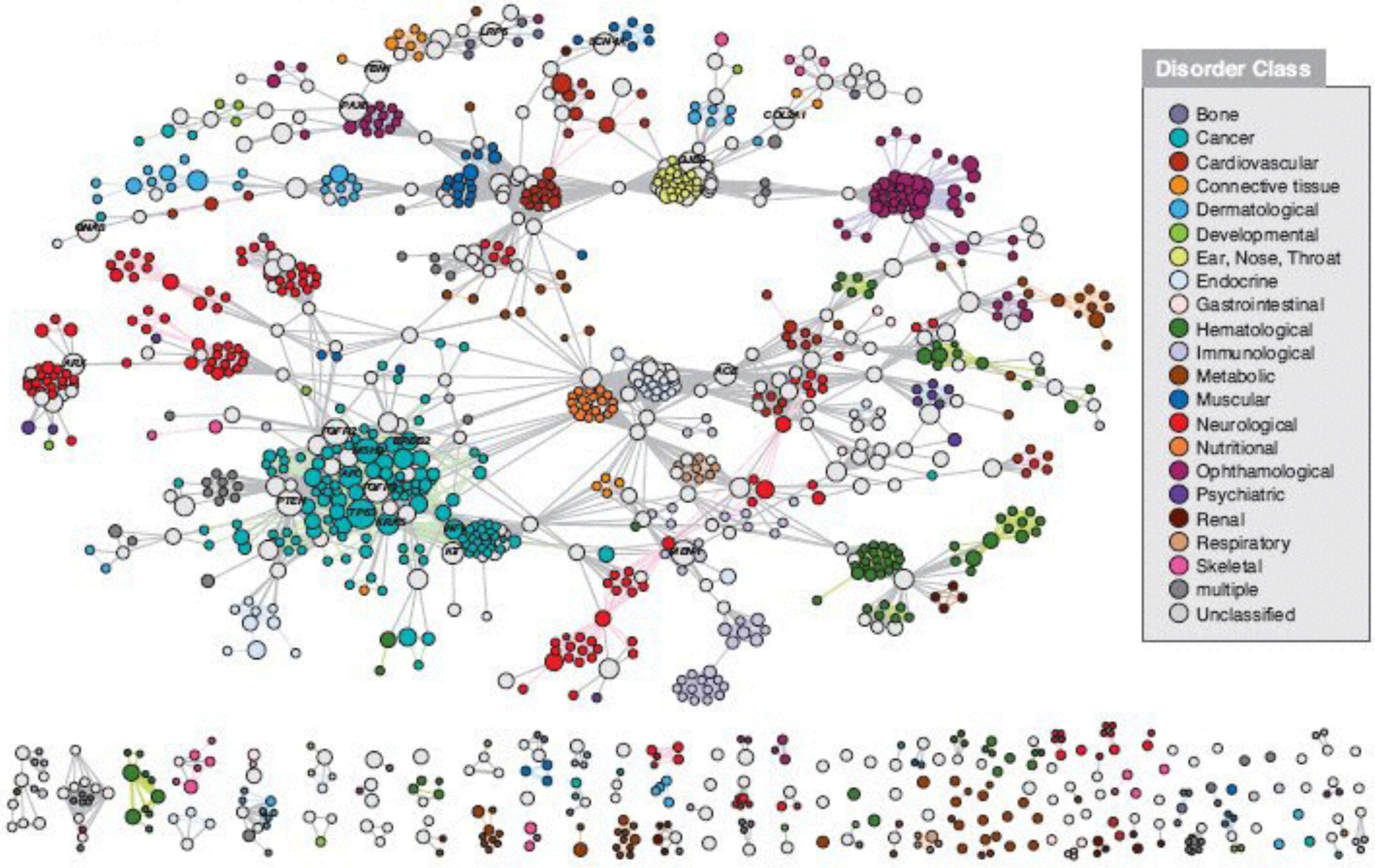
Ecological networks and their fragility
José M. Montoya, Stuart L. Pimm and Ricard V. Solé
Nature **442**, 259-264(20 July 2006)
doi:10.1038/nature04927



a–**e**, The cumulative probabilities *Pc*(*k*), for *k*, where *P*(*k*) is the probability a species has *k* links to other species, and is given by $P(k)k^{-}e^{-k/}$ where $e^{-k/}$ introduces a cut-off at some characteristic scale . Panels **a** (log–log) and **b** (log–linear) show three different modelled networks. Black lines, single-scale networks; when is very small, the distribution has a fast decaying tail, typically exponential, $P(k)e^{-k/}$. Green lines, truncated scale-free networks; these correspond to intermediate values of where the distribution has a power law regime followed by a sharp cut-off, with an exponential decay of the tail. Red lines, scale-free networks; for large values of the number of connections per species decays as a power law, $P(k)k^{-}$, a function with a relatively 'fat tail'. **c**–**e**, Experimental data (filled circles) and best fits (lines); **c**, a frugivore–plant web[76]; **d**, a pollinator–plant web[77]; and **e**, the food web from El Verde rainforest[78]. In **c** and **d**, for red circles, *k* is the number of plants species visited by an animal, and for black circles, *k* is the number of pollinator species visiting each plant species. In **e**, we sum prey–predator links and predator–prey links for each species. Best fits to the data in **c**–**e** are as follows: **c**, animals, exponential, $P(k) = e^{-k/3.998}$; plants, truncated power law, $P(k) = k^{-0.013}e^{-k/11.22}$; **d**, animals, power law, $P(k) = k^{-1.512}$; plants, truncated power law, $P(k) = k^{-0.2822}e^{-k/42.55}$; **e**, exponential, $P(k) = e^{-k/8.861}$. **f**–**h**, Photographs of a frugivore–plant (**f**), insect–flower (**g**) and predator–prey (**h**) interaction of webs depicted in **c**–**e**, respectively.

Genome-Wide Association Studies and Human Disease Networks

By: Leslie Pray, Ph.D. © 2008 Nature Education



Disease gene network (DGN).
In the DGN, each node is a gene, with two genes being connected if they are implicated in the same disorder. The size of each node is proportional to the number of disorders in which the gene is implicated. Nodes representing genes with links to multiple classes are colored dark grey, whereas unclassified genes are colored light grey. Genes associated with more than five disorders, and those mentioned in the text, are indicated with the gene symbol. Only nodes with at least one link are shown.© 2008 National Academy of Sciences, USA Goh, K.-I. et al. Th human disease network. **Proceedings of the National Academy of Sciences 104, 8685-8690 (2008).** All rights reserved.