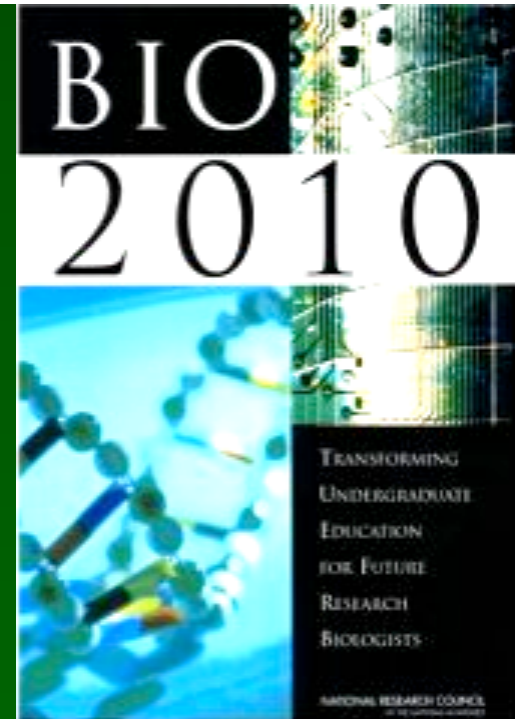


$$\begin{aligned} Imm. &= \frac{P - S}{d} \\ Ext. &= \frac{S}{a} \end{aligned}$$

BIO 2010: Transforming Undergraduate Education for Future Research Biologists

National Research Council (2003)



Recommendation #1:

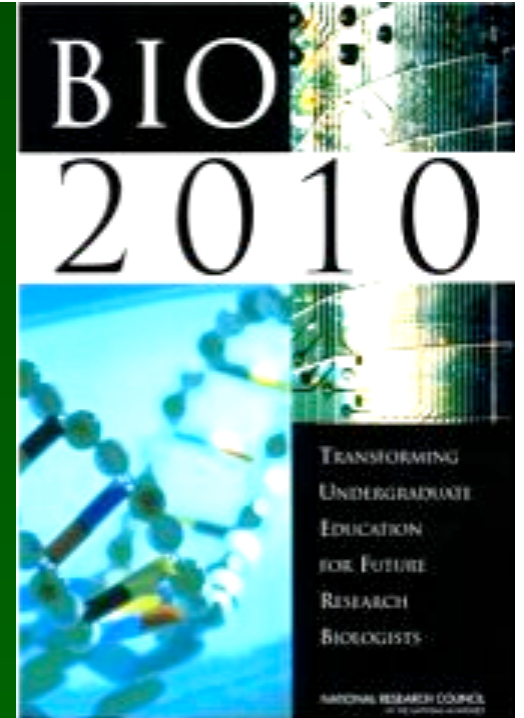
“Those selecting the new approaches should consider the importance of mathematics...”

Recommendation #2:

“Concepts, examples, and techniques from mathematics...should be included in biology courses. ...Faculty in biology, mathematics, and physical sciences must work collaboratively to find ways of integrating mathematics...into life science courses...”

BIO 2010: Transforming Undergraduate Education for Future Research Biologists

National Research Council (2003)



Specific strategies:

- A strong interdisciplinary curriculum that includes physical science, information technology, and math.
- Early opportunities for independent research.
- Meaningful laboratory experiences.

The ESTEEM Project Homepage

<http://bioquest.org/esteem>

55 modules: Broad range of topics and data sets



The screenshot shows the homepage of the Biological ESTEEM Project. The browser address bar displays <http://bioquest.org/esteem/index.php>. The page features a navigation bar with links: ESTEEM Home, Modules (highlighted with a red box), About, Contribute, Editorial Board, and Search. The main content area is titled "The Biological ESTEEM Collection: A project of the BioQUEST Curriculum Consortium". It describes the "Three E's of the Collection": Excel, Exploratory, and Mathematical. The "Excel" section explains that Microsoft's spreadsheet software was chosen for data collection and analysis. The "Exploratory" section describes the use of Excel for developing applications involving matrix algebra, statistics, and differential equations. A photo of student developer Annelise Myers is shown, along with a quote from her. The "Mathematical" section mentions the use of Excel for developing applications involving matrix algebra, statistics, and differential equations. The right sidebar features the "Math DL" logo and a section titled "Digital Classroom Resources" which provides a select collection of free online learning materials. Below this is a section titled "Implementing NRC Bio 2010's Recommendations for More Mathematics in Undergraduate Biology Education" which discusses the need for more mathematics in biology education.

Biological ESTEEM: Excel Simulations and Tools for Exploratory, Experiential Mathematics

<http://bioquest.org/esteem/index.php> Google

BioQUEST Curriculum Consortium

biological^(es+eem)

Excel Simulations and Tools for Exploratory, Experiential Mathematics

MAA Online
The Mathematical Association of America

ESTEEM Home **Modules** About Contribute Editorial Board Search

The Biological ESTEEM Collection:

A project of the BioQUEST Curriculum Consortium

The Three E's of the Collection:

Excel Microsoft's spreadsheet software Excel was chosen as a general development environment for the ESTEEM project because most biologists and mathematicians have it on their desktop computers, use it at least minimally for data collection, and find it fairly easy to operate. In addition, Excel is powerful enough to develop applications that involve matrix algebra, statistics, finite difference equations, and simple ordinary differential equations.

Exploratory Since parameters are so easy to change in Excel and it is so easy to import data from diverse and heterogeneous resources, our modules are intended to be adoptable, adaptable, extensible, flexible, and utilitarian for students who are engaged in a variety of biology and mathematics courses. We often have built templates that are easily employed for major modification to modules or easy mimicry to include new data sets or additional complication to current models.

Student developer, Annelise Myers, presents her preliminary work at the 2005 BioQUEST Summer Workshop

Math DL
The MAA Mathematical Sciences Digital Library

Digital Classroom Resources

The Digital Classroom Resources (DCR) provides a select collection of free online learning materials which are available through the site. These materials have been classroom tested and peer reviewed.

MAA Digital Classroom Resources Editor

Doug Ensley
Department of Mathematics
Shippensburg University
Shippensburg, PA 17257

Copyright 2006, All Rights Reserved, The Mathematical Association of America.

Implementing NRC Bio 2010's Recommendations for More Mathematics in Undergraduate Biology Education

In 2002, the National Research Council made eight major recommendations for the improvement of undergraduate biology education in its publication: [BIO2010: Transforming Undergraduate Education for Future Research Biologists](#). The first two of these recommendations both emphasized the need for additional attention to the inclusion of more mathematics:

"It is important that all students understand the growing relevance of quantitative science in addressing life-science questions. Thus, a better integration of quantitative applications in biology would not only enhance life science education for all students, but also decrease the chances that

Module Main Page

- Screenshots & brief description
- Mathematical expression
- Research articles & primary data
- User manual & curriculum materials (in progress)
- Downloadable Excel sheet

Biological ESTEEM: Excel Simulations and Tools for Exploratory, Experiential Mathematics

http://bioquest.org/esteem/esteem_details.php?product_id=211

Google

BioQUEST Curriculum Consortium

biological^(es+eem)

Excel Simulations and Tools for Exploratory, Experiential Mathematics

MAA Online
The Mathematical Association of America

ESTEEM Home Modules About Contribute Editorial Board Search

Island Biogeography



This excel workbook demonstrates the principles of the MacArthur-Wilson theory of Island Biogeography. It allows the user to define the mainland species pool, area of the island, and distance of the island from the mainland. Graphical output included species richness equilibrium at varying island size and distance. The workbook also allows the user to calculate a species-area function for data entered into the data input page. Several datasets on island area and species richness are included for various types of islands and species. Variables and formulas are defined in the accompanying tutorial.

Fundamental Mathematical Expression

$$y = a x^b$$

$$\log y = \log a + \log b$$

(compare with $y = b + mx$)

Logarithms, Exponents, Power Function

Equation's Author(s)




Robert H. MacArthur and Edward O. Wilson

Source

Author(s):
John Jungck, Beloit College
Jennifer Spangenberg, Beloit College

Published by: BioQUEST Curriculum Consortium

OS: win98, win2000, winXP

User Manuals and Curricular Materials

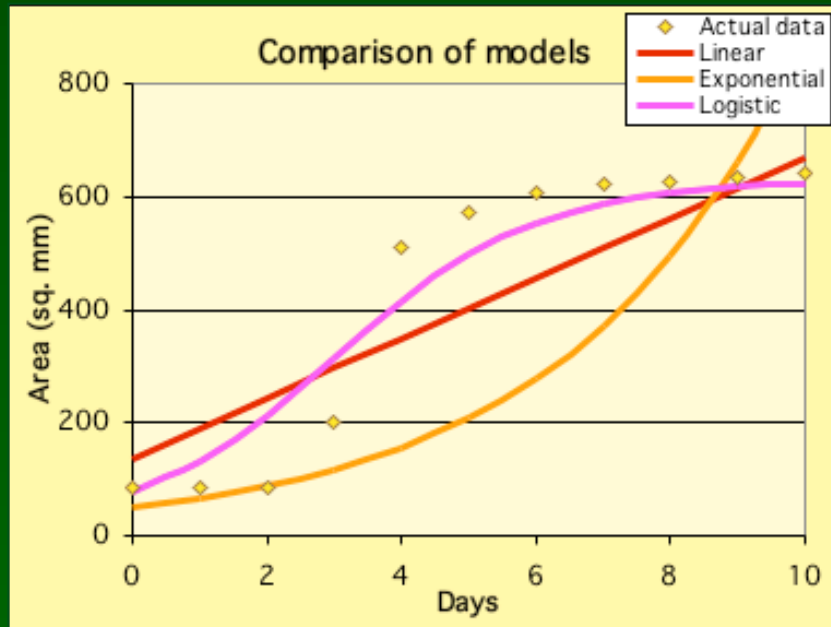
Download

- [IslandBiogeography.xls](#)

Primary Reference

MacArthur R H, Wilson E O (1967) *The Theory of Island Biogeography*. Princeton: Princeton University Press.

Population growth (Continuous Growth Models)



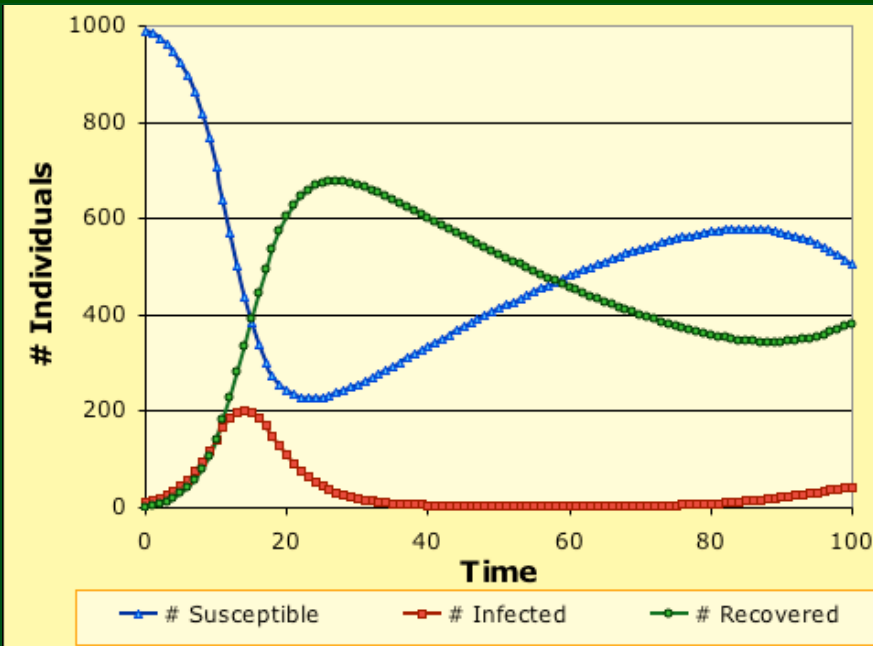
Bioinformatics (Pairwise Alignment)

			A	T	C	G	C	T	T							
		0	-7	-14	-21	-28	-35	-42	-49							
		###	###	-14	###	###	###	###	###							
C		-4	-14	-11	-21	-9	-28	-25	-35	-23	-42	-39	-49	-46	-56	
		-7	-14	-4	-11	-11	-18	-9	-16	-16	-23	-23	-30	-30	-37	-37
		###	###	###	###	-9	###	###	###	###	###	###	###	###	###	
G		-11	-11	-8	-18	-15	-16	-4	-23	-20	-30	-27	-37	-34	-44	
		-14	-21	-11	-18	-8	-15	-15	-22	-4	-11	-11	-18	-18	-25	-25
		###	###	###	###	###	###	-4	###	###	###	###	###	###	###	
C		-18	-18	-15	-15	-3	-22	-19	-11	1	-18	-15	-25	-22	-32	
		-21	-28	-18	-25	-15	-22	-3	-10	-10	-17	1	-6	-6	-13	-13
		###	###	###	###	###	###	###	###	1	###	###	###	###	###	
T		-25	-25	-13	-22	-19	-10	-7	-17	-14	-6	6	-13	-1	-20	
		-28	-35	-25	-32	-13	-20	-10	-17	-7	-14	-6	-13	6	-1	-1
		###	###	###	###	###	###	###	###	###	###	6	###	###	###	
C		-32	-32	-29	-20	-8	-17	-14	-14	-2	-13	-10	-1	2	-8	
		-35	-42	-32	-39	-20	-27	-8	-15	-14	-21	-2	-9	-1	-8	2
		###	###	###	###	###	###	###	###	###	###	###	-1	###	###	
T		-39	-39	-27	-27	-24	-15	-12	-21	-18	-9	3	-8	4	-5	
		-42	-49	-39	-46	-27	-34	-15	-22	-12	-19	-9	-16	3	-4	4
		###	###	###	###	###	###	###	###	###	###	###	###	###	###	4

Gene regulation (Operon)

Strand	Upstream Regulator		Promoter			Operator		<i>lacZ</i> (Galactosidase)	<i>lacY</i> (Permease)	<i>lacA</i> (Transacetylase)	Sugars present in medium	
	I (+/-/s)		P (+/-)			O (+/c)		Z (+/-)	Y (+/-)	A (+/-)	Lactose?	Glucose?
F-	<div>+ (Wild-type)</div>		<div>+ (Wild-type)</div>			<div>c (Constitutive)</div>		<div>+ (Wild-type)</div>	<div>- (Nonfunct.)</div>	<div>+ (Wild-type)</div>	<div>Absent</div>	<div>Absent</div>
F'	<div>- (Nonfunctional)</div>		<div>+ (Wild-type)</div>			<div>+ (Wild-type)</div>		<div>- (Nonfunct.)</div>	<div>+ (Wild-type)</div>	<div>- (Nonfunct.)</div>		
Strand	Super-repressing repressor produced?	Normally active repressor produced?	cAMP levels?	cAMP/CAP complex binds to promoter?	RNA polymerase binds to promoter?	Repressor binds to operator?	Polymerase can transcribe?	Galactosidase?	Permease?	Transacetylase?		
F-	FALSE	TRUE	High	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE		
F'	FALSE	TRUE	High	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE		

Epidemiology (*SIR Model*)

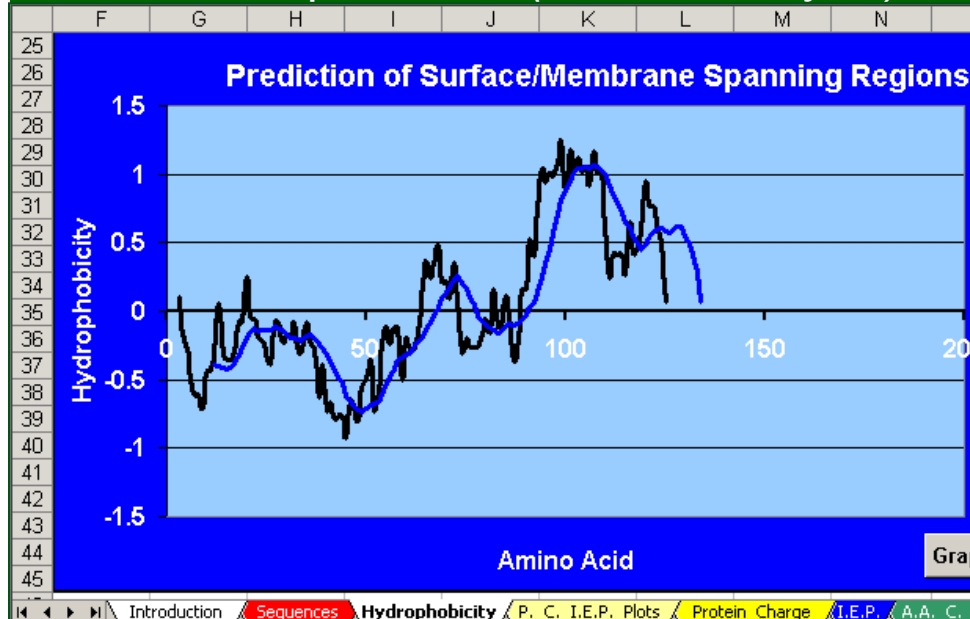


Phylogenetics (*Evo/Seq*)

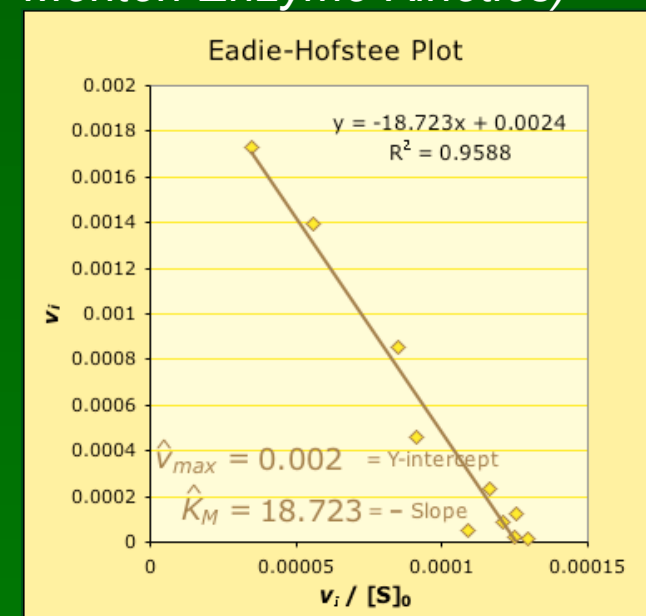
Distance matrix: Nucleotide Sequences							
	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6	Seq. 7
Seq. 1	0	10	2	6	10	10	12
Seq. 2		0	10	10	8	4	12
Seq. 3			0	6	10	10	12
Seq. 4				0	10	10	12
Seq. 5					0	8	12
Seq. 6						0	12
Seq. 7							0

Distance matrix: Amino Acid Sequences							
	Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6	Seq. 7
Seq. 1	0	6	2	3	4	7	8
Seq. 2		0	6	7	4	3	10
Seq. 3			0	3	4	7	8
Seq. 4				0	5	8	9
Seq. 5					0	5	8
Seq. 6						0	11
Seq. 7							0

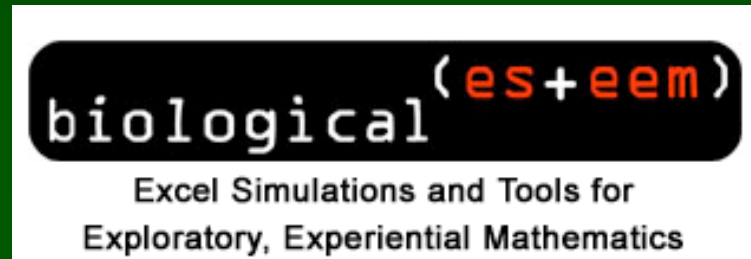
Structure prediction (*Protein Analysis*)



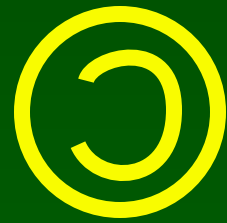
Biochemistry (*Michaelis-Menten Enzyme Kinetics*)



Unpacking “ESTEEM”



- *Excel*: ubiquitous, easy, flexible, non-intimidating
- Exploratory: apply to real-world data; extend & improve
- Experiential: students engage directly with the math



Copyleft

Users may freely

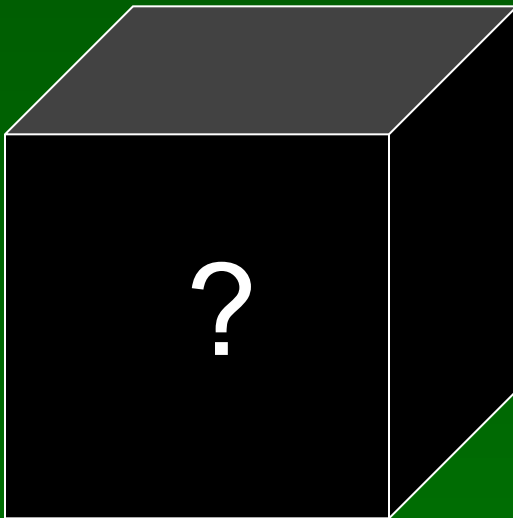
- download
- use
- modify
- share

the software, *w/
proper attribution*

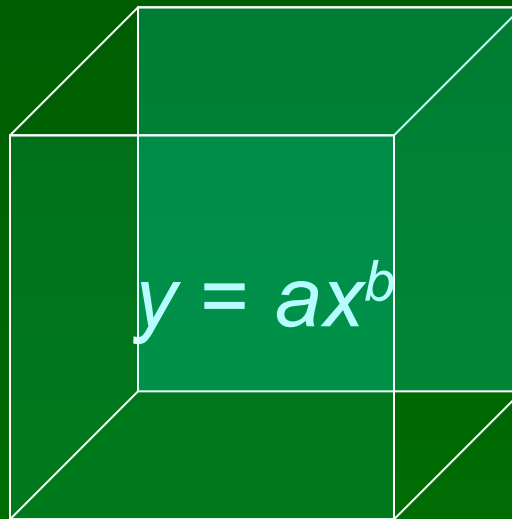
More info available at *Free
Software Foundation* website

Three Boxes

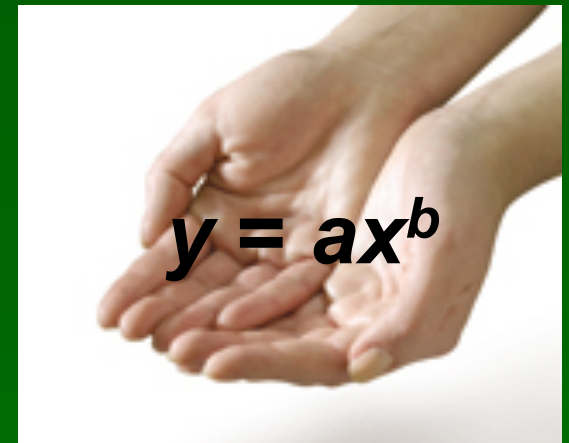
How do students interact with the mathematical model underlying the biology?



Black box:
Hide the model



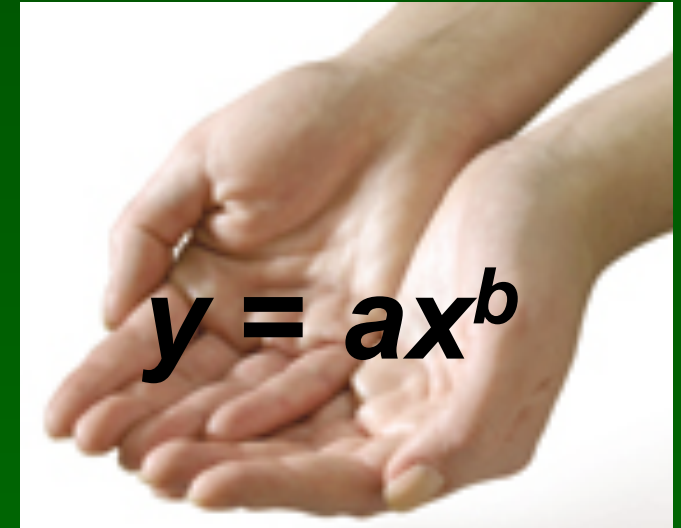
Glass box:
Study the model



No box:
Build the model!

1. The No-Box Approach

- a. Students try to build a **conceptual model** of how a particular biological system works.
- b. They then **translate** that understanding into simple mathematical expressions and implement the model in *Excel*.



Examples: [Island Biogeography](#), SIR

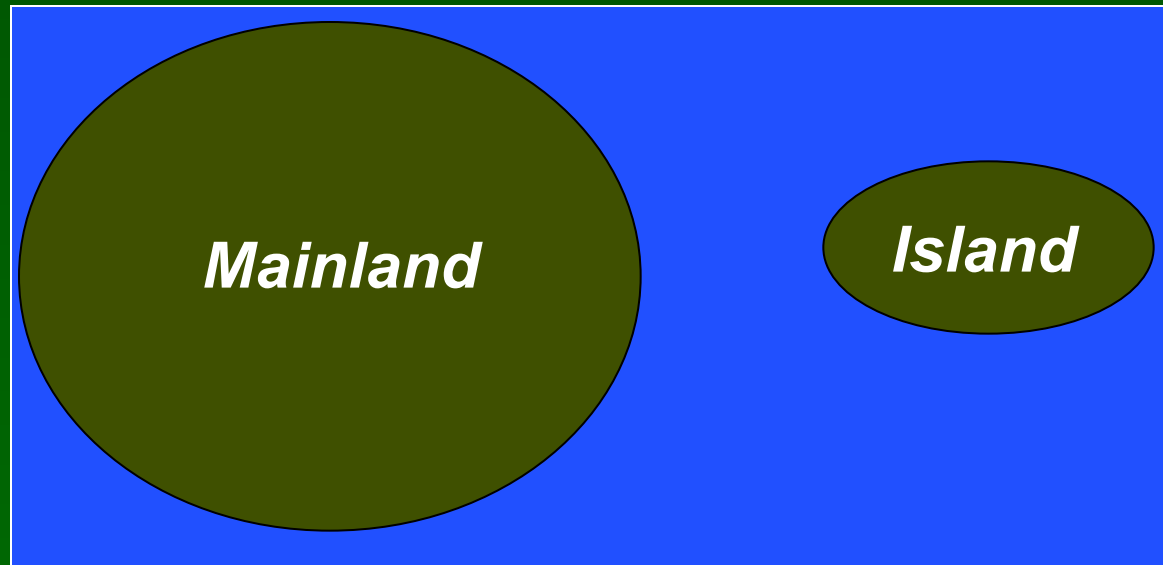
Island Biogeography

(J.A. Spangenberg & J.R. Jungck)

- Survey of plant species on islands off British mainland
- Which islands would you expect to have most species, and why?

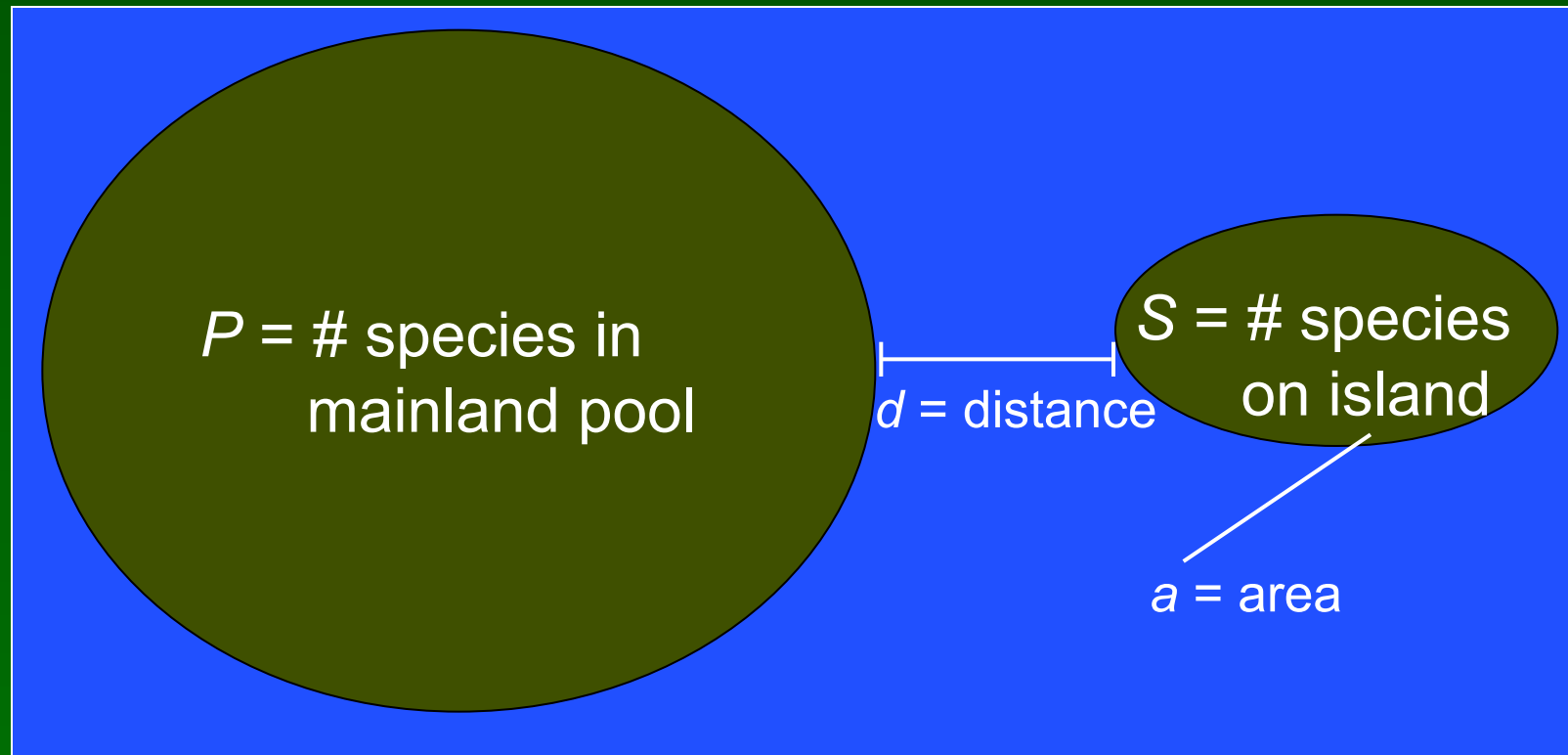


Island Biogeography



1. What parameters most strongly influence immigration of species to the island or local extinction of species already there?

Island Biogeography

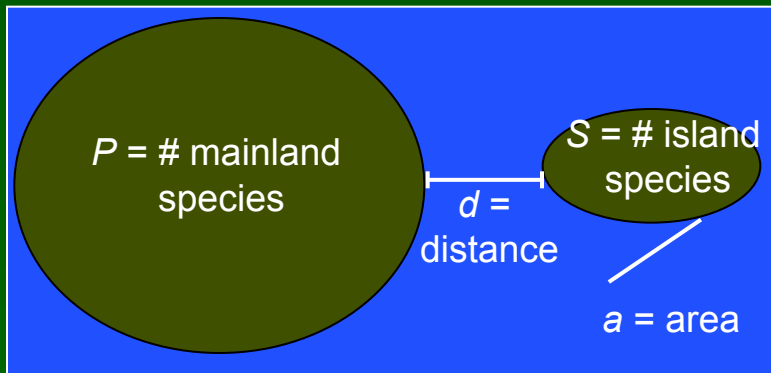


2. For each parameter, predict its *general effect* (+, 0, –) on immigration.

Then do the same for extinction.

Island Biogeography

My predictions:



Parameter	Effect on immigration	Effect on extinction
P	+	0
S	—	+
d	—	0
a	0	—

Individual students' predictions may differ, depending on their assumptions

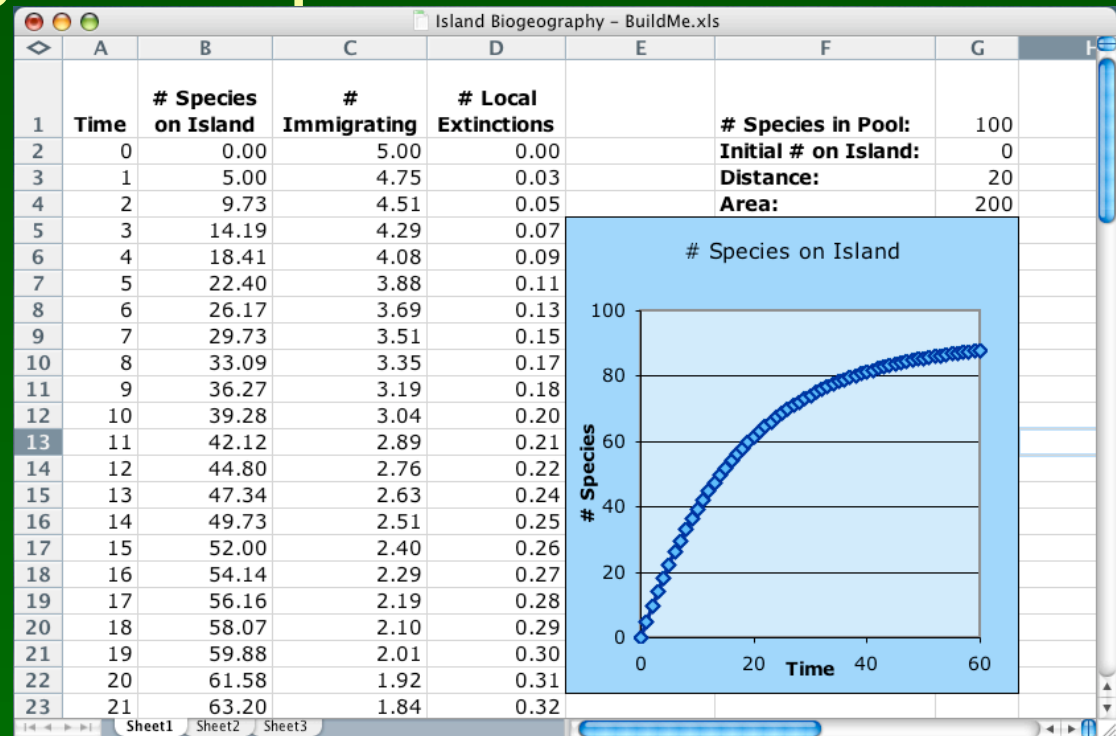
3. Write equations for immigration & extinction rates that reflect the behavior you predicted.

Island Biogeography: Building a Simple Model

Model:

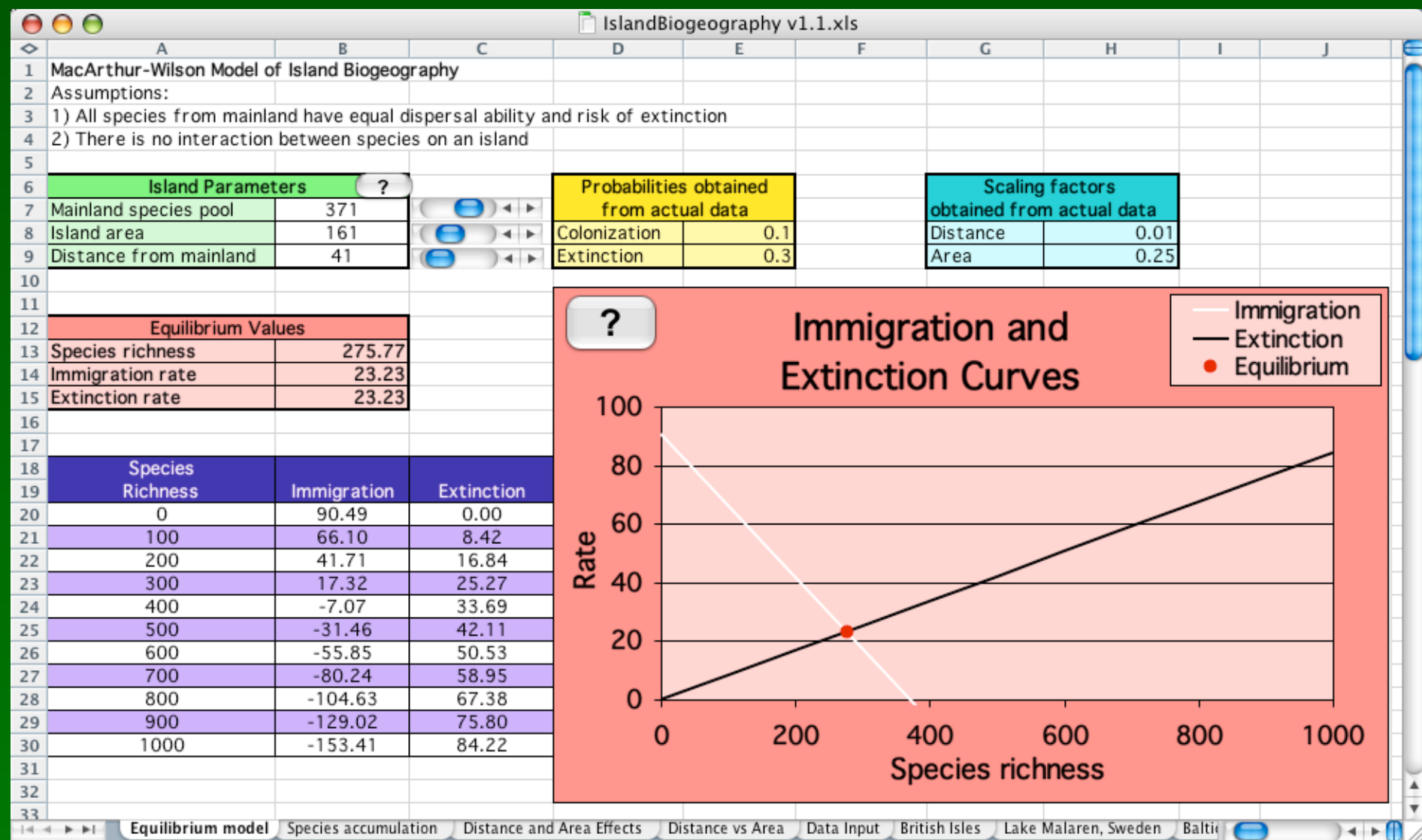
$$\text{Immigration} = \frac{P - S}{d}$$

$$\text{Extinction} = \frac{S}{a}$$



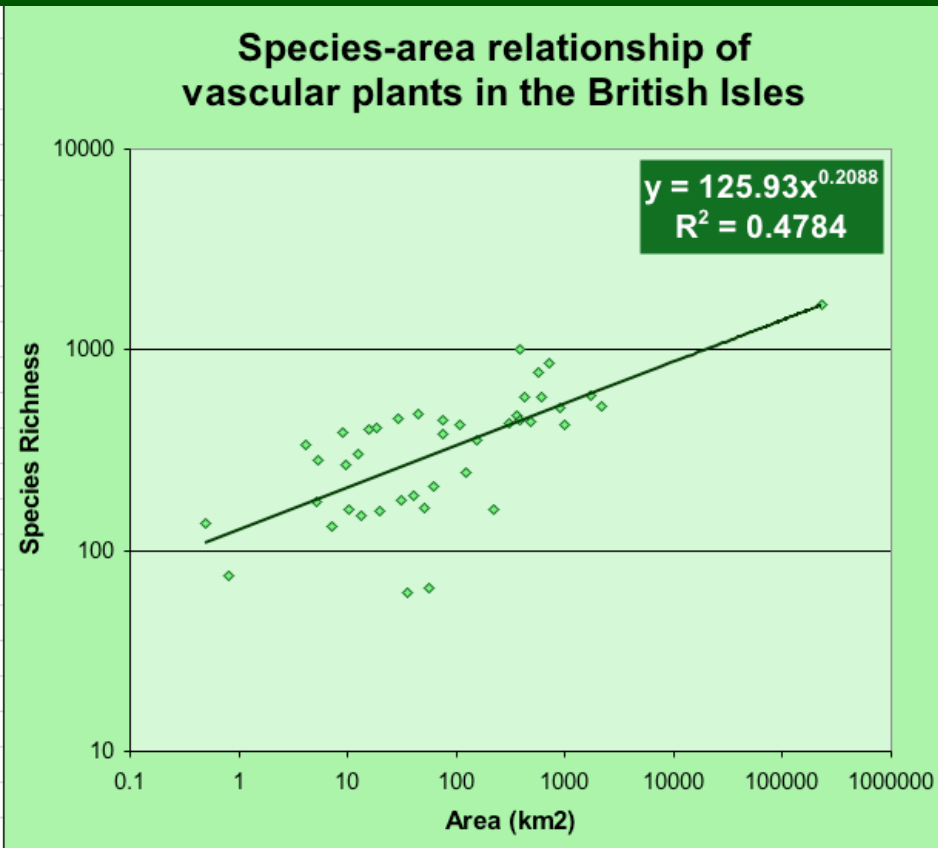
- How can we calculate equilibrium species richness \hat{S} ?
(Set immigration = extinction, solve for S)
- How do island area & distance from mainland affect \hat{S} ?

Beyond the No-Box Approach: Using the prebuilt *Excel* workbook



Data Set: British Isles

Island	Area (km ²)	Richness
May	0.5	137
Ailsa	0.8	75
Lundy	4.1	338
Fair	5.2	174
Muck	5.4	284
N. Ronaldsay	7.3	131
Iona	9.1	388
Mingulay	9.6	269
Rona	10.4	159
Canna	12.7	300
Foula	13.5	149
Gigha	15.5	401
Barra	18.4	409
Whalsay	19.7	158
Eigg	29	453
Bressay	31.1	177
Stronsay	35.2	62
Fetlar	40.9	189
Colonsay	44.8	482
Sanday	50.2	162
Westray	55.4	65
S. Ronaldsay	60.9	207
Coll	74.1	443



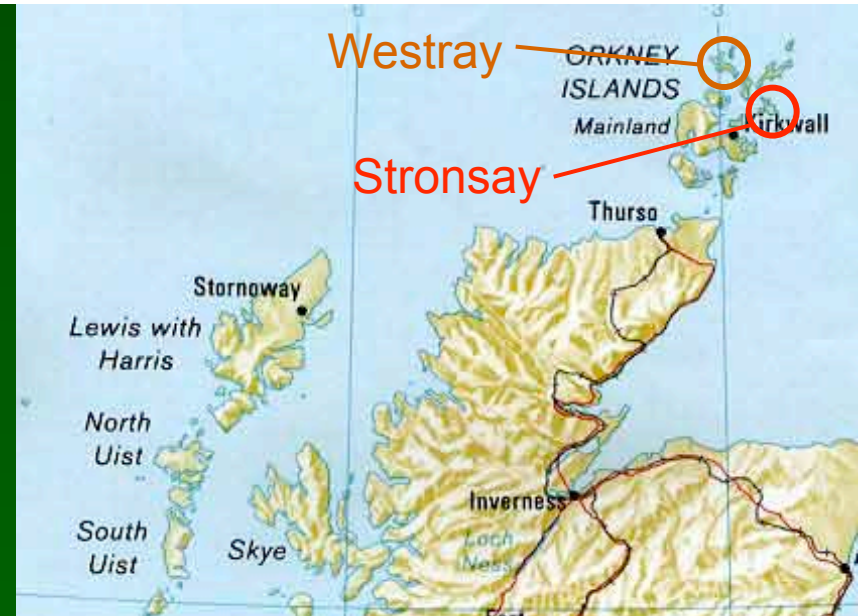
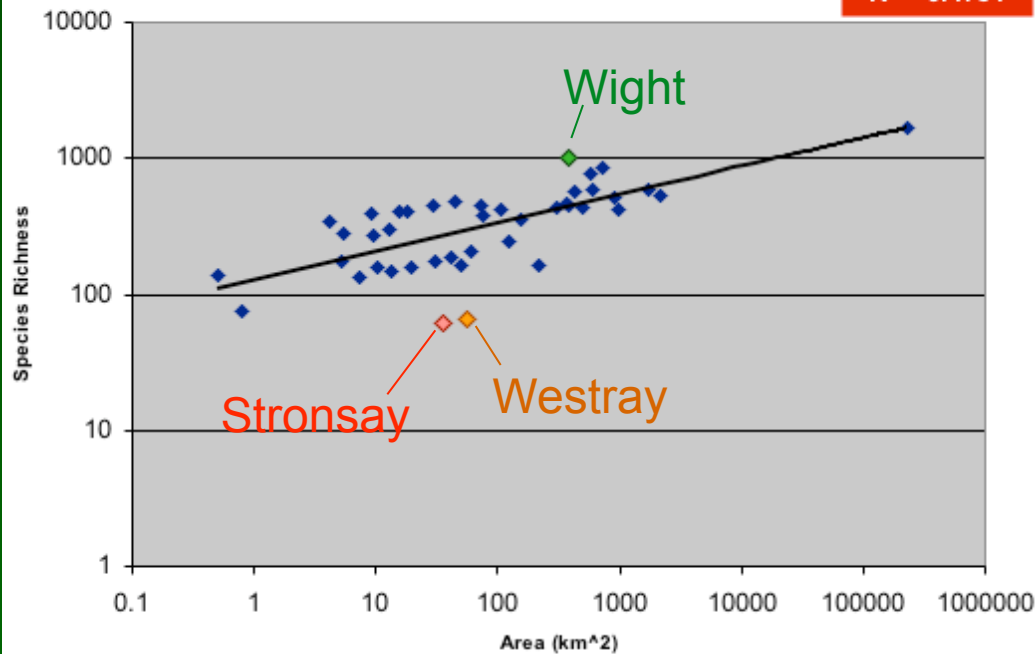
Questions for further investigation:

- How much does the relationship depend on 1-2 individual data points?
- Why are some individual data points so far from the trendline?
Can we substantially improve the model by adding more factors?

Species-Area relationships of vascular plants in the British Isles

$$y = 125.93x^{0.2089}$$

$$R^2 = 0.4784$$



Low outliers:

- high latitude (subarctic)
- distant from mainland

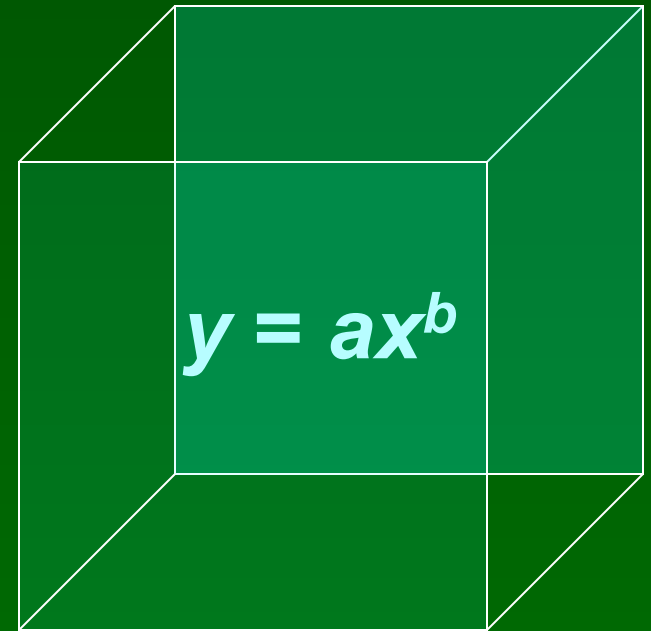
High outliers:

- low latitude (temperate)
- close to mainland



2. The Glass-Box Approach

- a. Students analyze prebuilt software tools to understand specific relationships and techniques (e.g., logistic population growth, prediction of 2° protein structure)
- b. This involves exploring the inner workings of a model's equations or algorithms.



Examples: *Continuous Growth Models, [EvolSeq](#), Mark/Recapture, Protein Analysis, Split Decomp*

EvoSeq

Where do trees like this come from?



EvolSeq

Ribulose 1,5-bisphosphate carboxylase (rubisco)

<i>Zea</i>	AAAGAAGACCACAGGGGCCCTGCTGGAGATGAAGGCCCAACC
<i>Arabidopsis</i>	AAAGAAGACCACGGAAGGCCCTGCTGGAGCTGAAGGCCCAACC
<i>Sphagnum</i>	AAAGAAGACCACGGAAGGCCCTGCTGGAGCTGAAGGCCCAACC
<i>Picea</i>	AGAGAAGACCAAGGAAGGCCCTCCTGGAGCTGAAGGCCCAACC
<i>Cyathea</i>	AGAGAAGACCAAAAGAGGCCCTGCTGGAGCTGAAGGCCGAGCCC

Task 1:

Calculate the evolutionary distance
between each pair of sequences.

(Treat all differences alike, or assign greater weight to transversions?
Depends on your assumptions re. molecular evolution!)

# of sequences:	7	(must be between 2 and 20 inclusive)	Distance matrix display			
			<input type="radio"/> Upper triangular <input checked="" type="radio"/> Symmetric			

Distance matrix: Nucleotide Sequences

	<u>Seq. A</u>	<u>Seq. B</u>	<u>Seq. C</u>	<u>Seq. D</u>	<u>Seq. E</u>	<u>Seq. F</u>	<u>Seq. G</u>
Seq. A	0	8	12	12	10	12	10
Seq. B	8	0	12	12	10	12	10
Seq. C	12	12	0	4	12	4	12
Seq. D	12	12	4	0	12	2	12
Seq. E	10	10	12	12	0	12	6
Seq. F	12	12	4	2	12	0	12
Seq. G	10	10	12	12	6	12	0

Task 2:

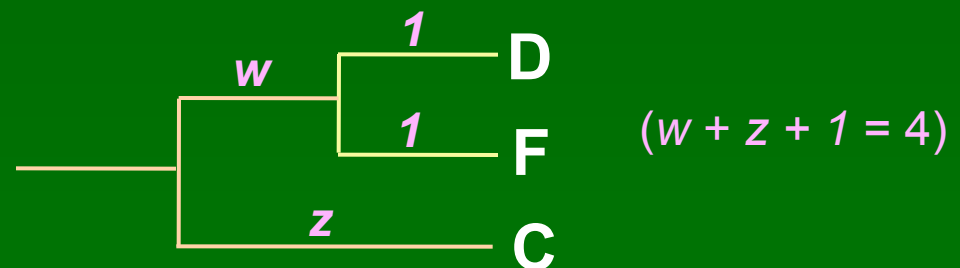
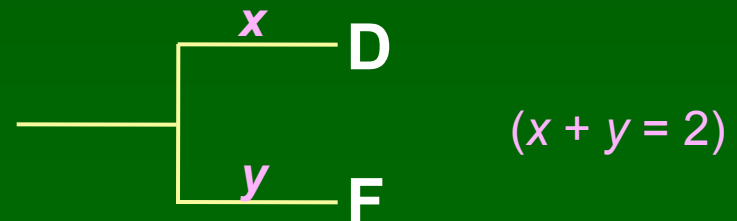
Construct a phylogenetic tree for these seven taxa.

BONUS: Include branchlengths!

Distance matrix: Nucleotide Sequences							
	<u>Seq. A</u>	<u>Seq. B</u>	<u>Seq. C</u>	<u>Seq. D</u>	<u>Seq. E</u>	<u>Seq. F</u>	<u>Seq. G</u>
Seq. A	0	8	12	12	10	12	10
Seq. B	8	0	12	12	10	12	10
Seq. C	12	12	0	4	12	4	12
Seq. D	12	12	4	0	12	2	12
Seq. E	10	10	12	12	0	12	6
Seq. F	12	12	4	2	12	0	12
Seq. G	10	10	12	12	6	12	0

Strategy 1: cherry-picking

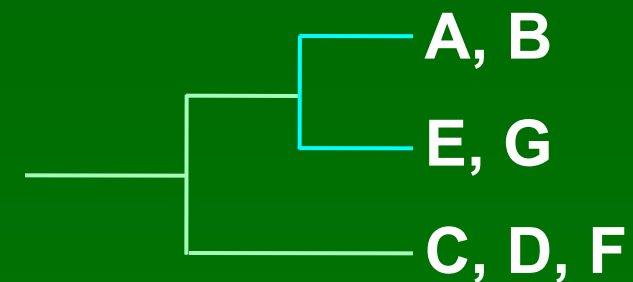
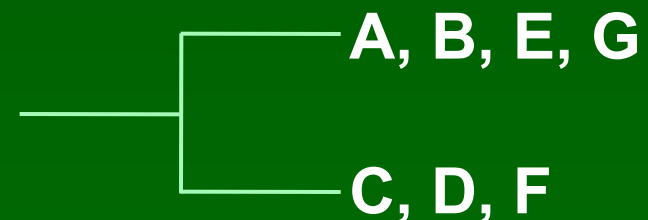
- Unite the **2 most similar taxa**
- Assign **branchlengths** that fit all pairwise distances
- Repeat



Distance matrix: Nucleotide Sequences							
	<u>Seq. A</u>	<u>Seq. B</u>	<u>Seq. C</u>	<u>Seq. D</u>	<u>Seq. E</u>	<u>Seq. F</u>	<u>Seq. G</u>
Seq. A	0	8	12	12	10	12	10
Seq. B	8	0	12	12	10	12	10
Seq. C	12	12	0	4	12	4	12
Seq. D	12	12	4	0	12	2	12
Seq. E	10	10	12	12	0	12	6
Seq. F	12	12	4	2	12	0	12
Seq. G	10	10	12	12	6	12	0

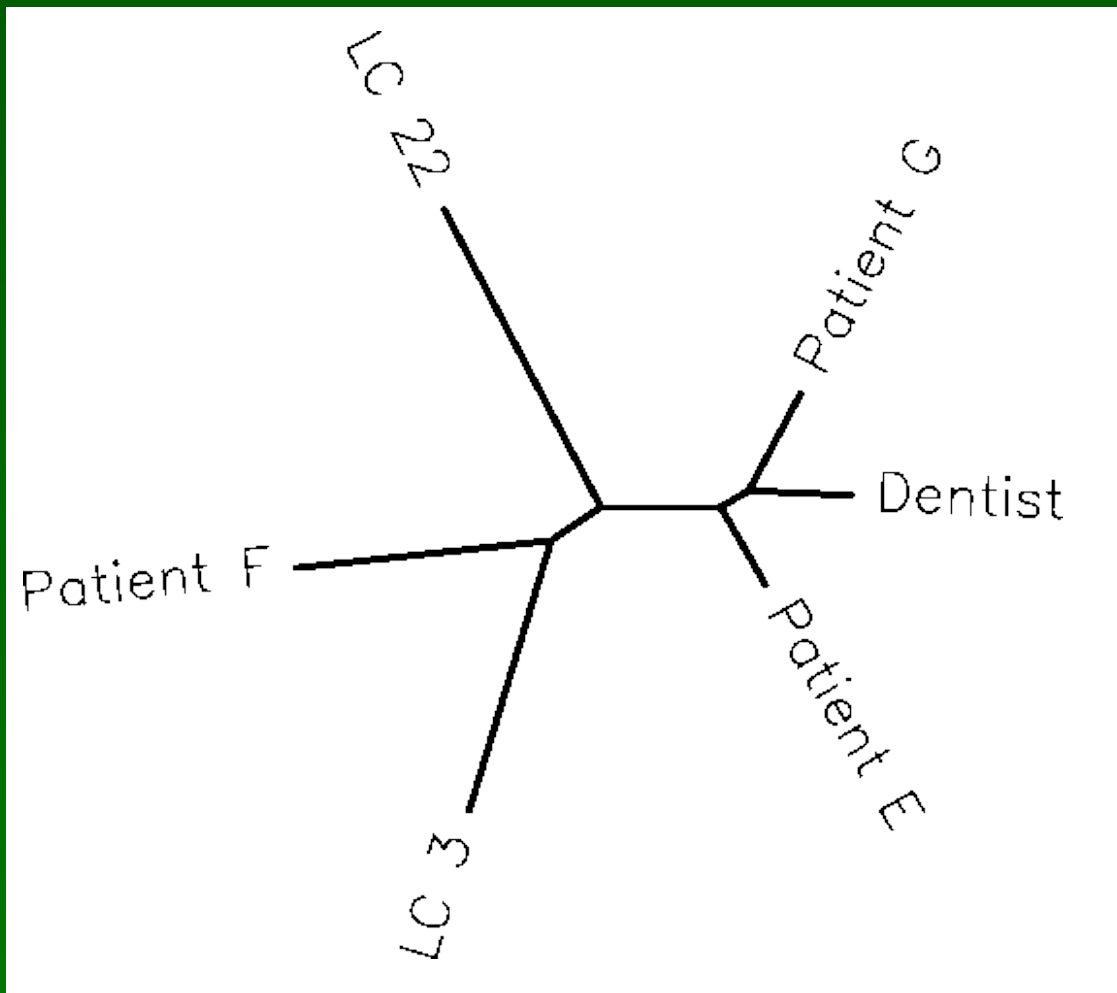
Strategy 2: splitting

- Divide taxa into 2 groups separated by largest distance
- Repeat
- At end, assign branchlengths that fit all pairwise distances



SplitDecomp

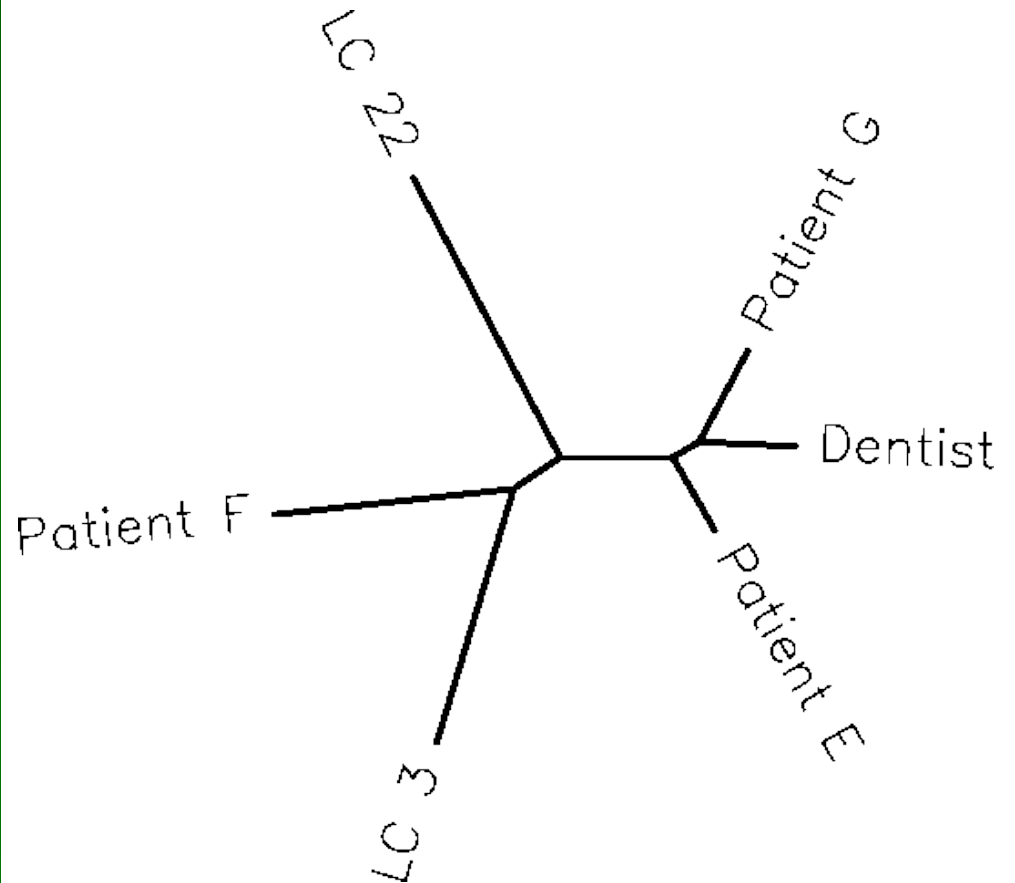
How much evidence supports a given branch?



Phylogenetics in the Courtroom

- In 1990, a Florida woman contracted HIV. Only known risk factor: invasive dental surgery by an HIV-positive dentist.
- HIV evolves so fast that viral DNA differs within and between indiv. patients!
- Examined sequences from dentist, several HIV+ patients, and local controls:

Interpretation? Significance?

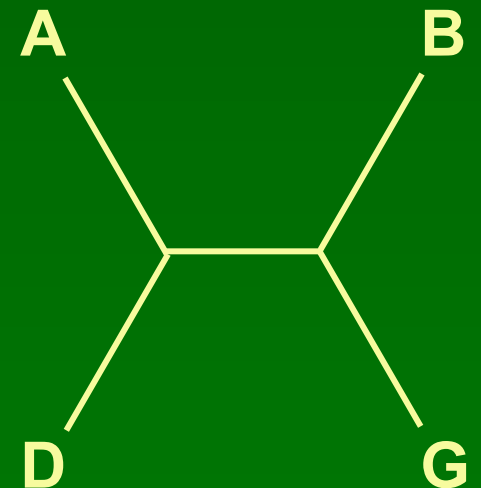
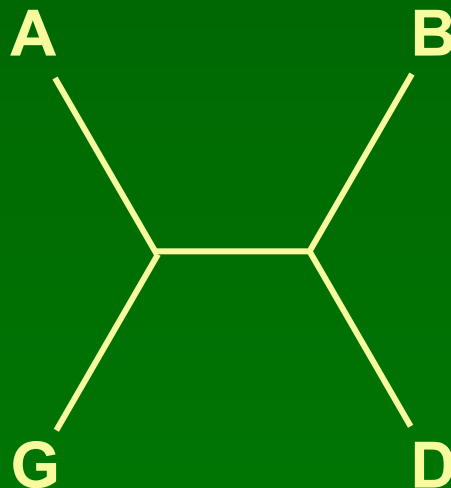
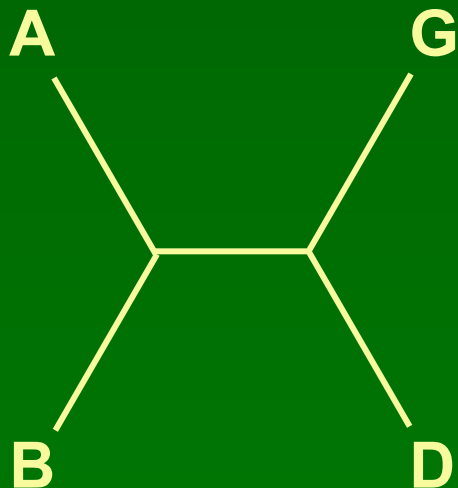


Split Decomposition:

Evaluate Evidence for Internal Branches

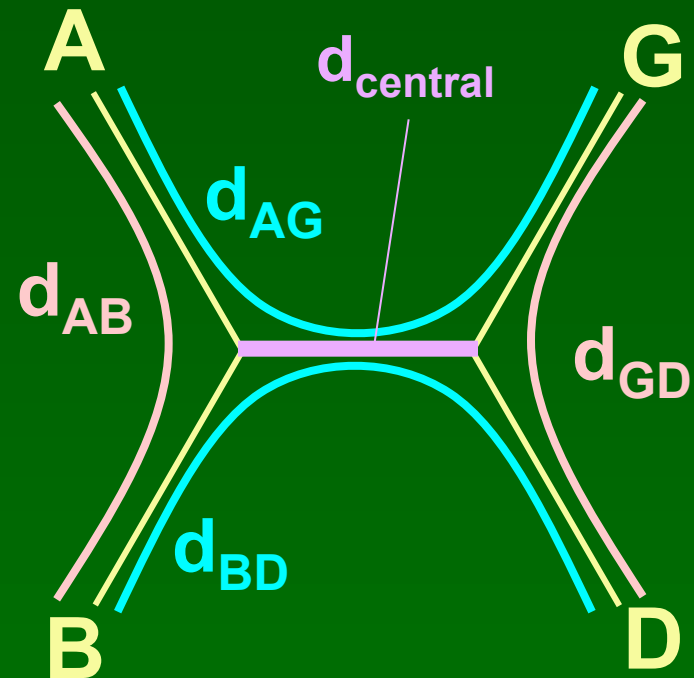
Alpha	AAGAAGCAGGGGCTGGAGATGAAGGCCCATG
Beta	AAGAAGCGGAGGGCTGGAGCTGAAGGCCCATG
Gamma	AAGAAGCGGAGGGTGGAGCTGAAGGCCCACG
Delta	GAGAAGAGGAGCCTGGAGCTGAAGGCCCACA

Three possible unrooted trees:



Alpha	AAGAAGCAGGGGCTGGAGATGAAGGCCCATG	} d_{AB}
Beta	AAGAAGCGGAGGCTGGAGCTGAAGGCCCATG	
Gamma	AAGAAGCGGAGGGTGGAGCTGAAGGCCCACG	} d_{GD}
Delta	GAGAAGAGGAGCCTGGAGCTGAAGGCCACACA	

How can we calculate
the length of the
central branch?



$$d_{\text{central}} = \frac{d_{AG} + d_{BD} - (d_{AB} + d_{CD})}{2}$$

ONLY for the
correct tree!!!

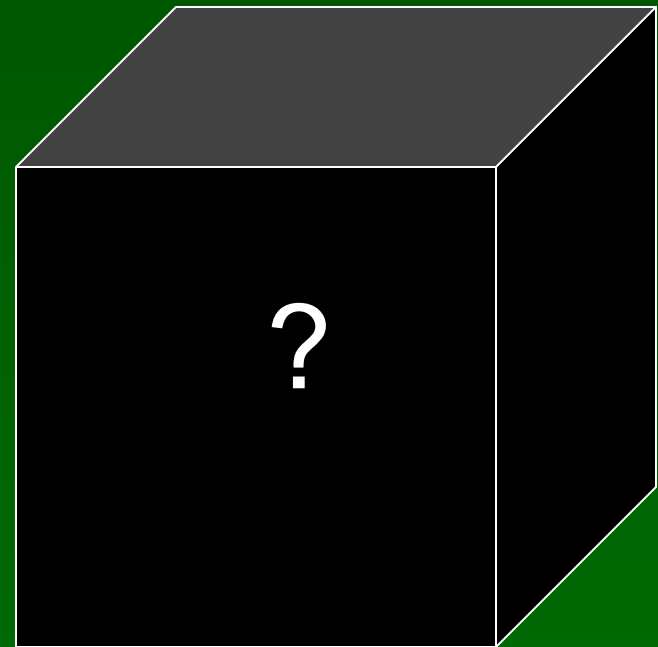
2. SplitDecomp

<i>Distance matrix</i>					
		<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>
		Dentist	Patient E	Local control 3	Local control 22
A)	Dentist		14	43	40
B)	Patient E			43	37
C)	Local control 3				44
D)	Local control 22				

<i>Split decomposition</i>				
Putative tree	Split index 1	Split index 2	Four-point condition	
((A,B), (C,D))	11	12.5	58	$=d_{AB} + d_{CD}$
((A,C), (B,D))	-11	1.5	80	$=d_{AC} + d_{BD}$
((A,D), (B,C))	-12.5	-1.5	83	$=d_{AD} + d_{BC}$

3. The Black-Box Approach

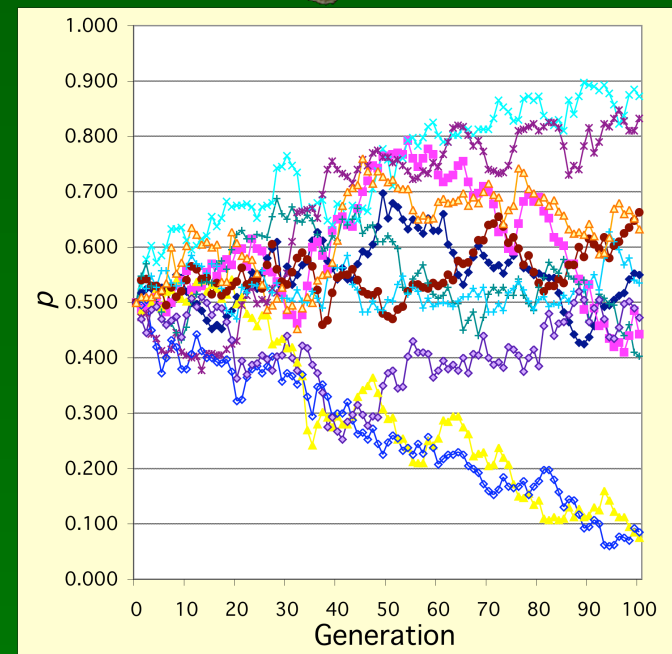
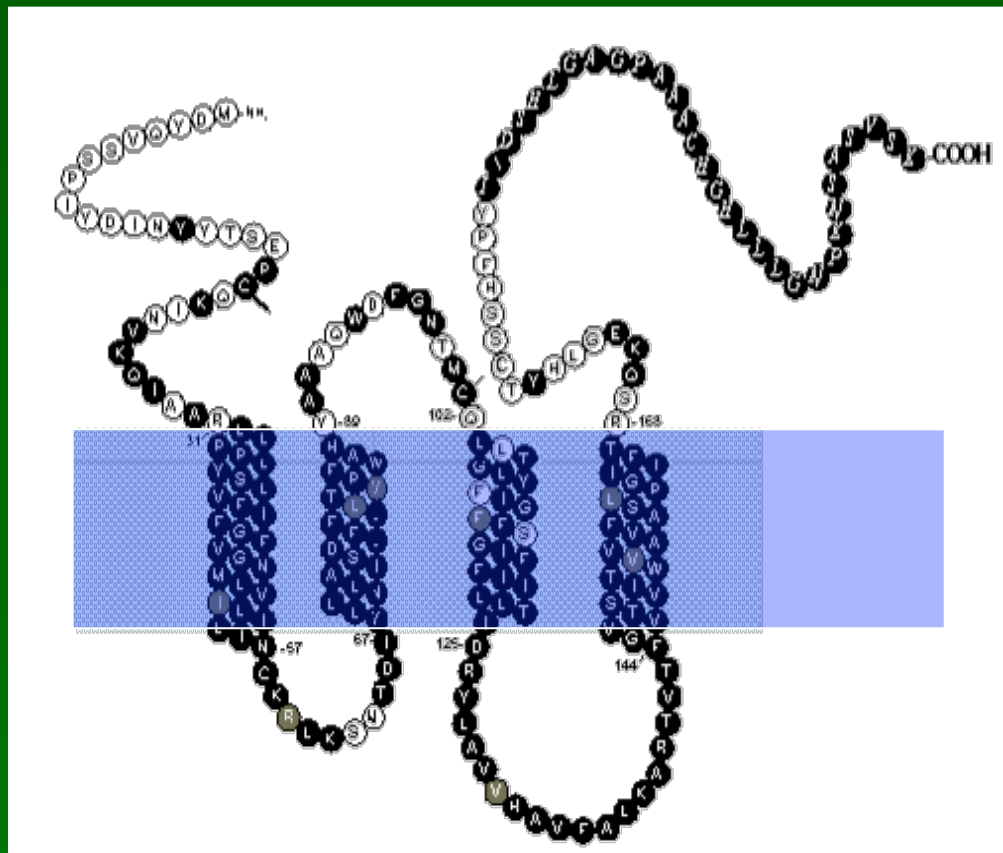
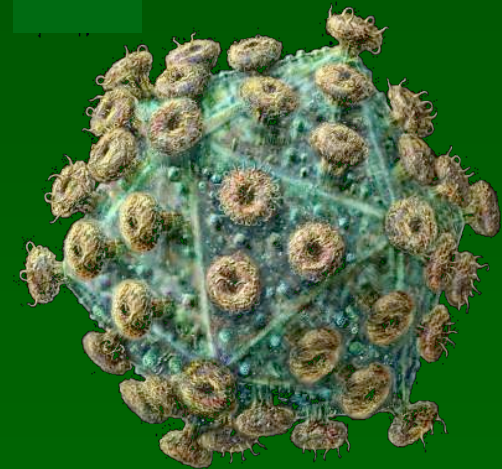
- a. Students use prebuilt software tools to investigate open-ended biological questions.
- b. This may involve determining model's sensitivity to specific parameters, analyzing range of variation in stochastic models, etc.



Examples: *3D Fractal-Tree*, [Deme](#), *javaBenzer*,
Michaelis-Menten Enzyme Kinetics

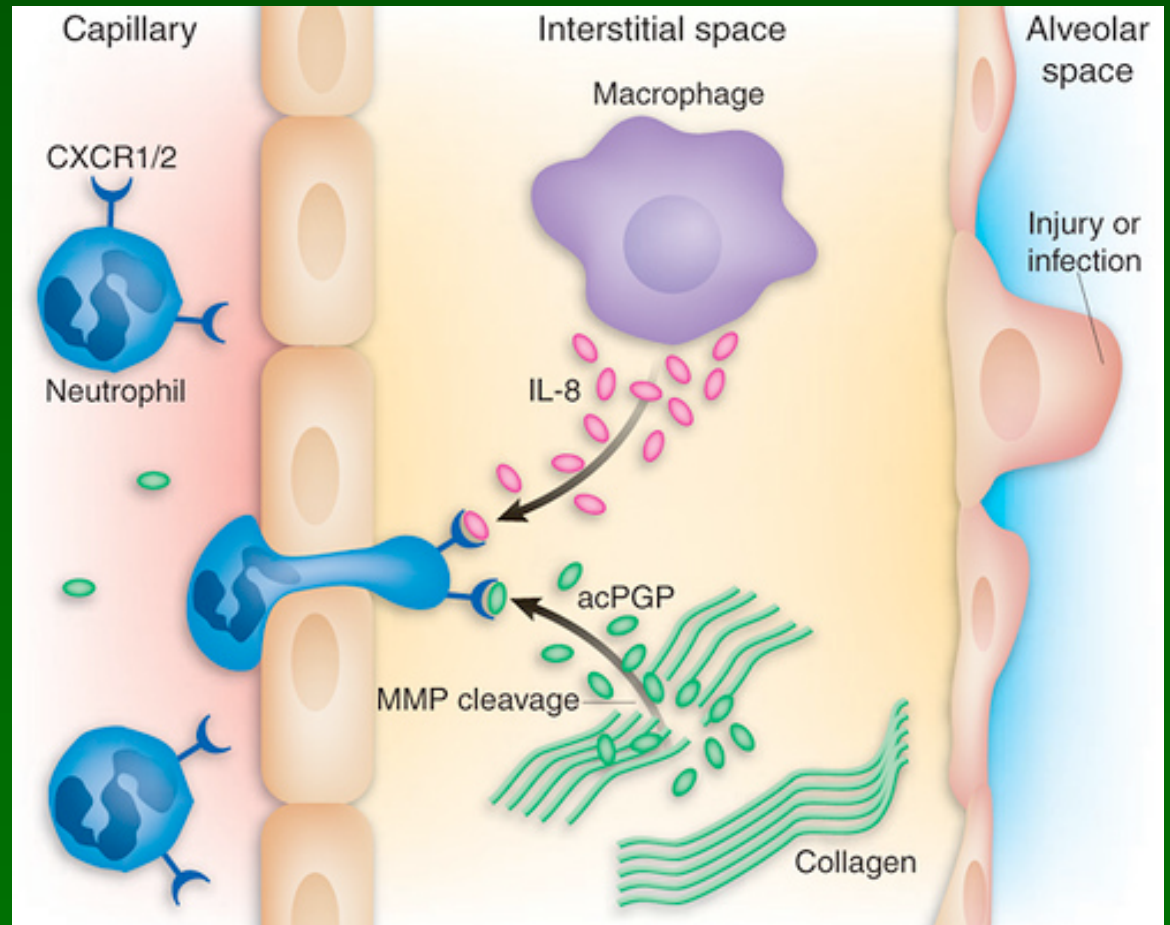
The Case of the Protective Protein

(under review at *Evolution Education & Outreach* as of 8/3/2009)



Chemokines

- A family of small proteins secreted by cells to control migration of nearby cells (e.g. during tissue development or immune response)

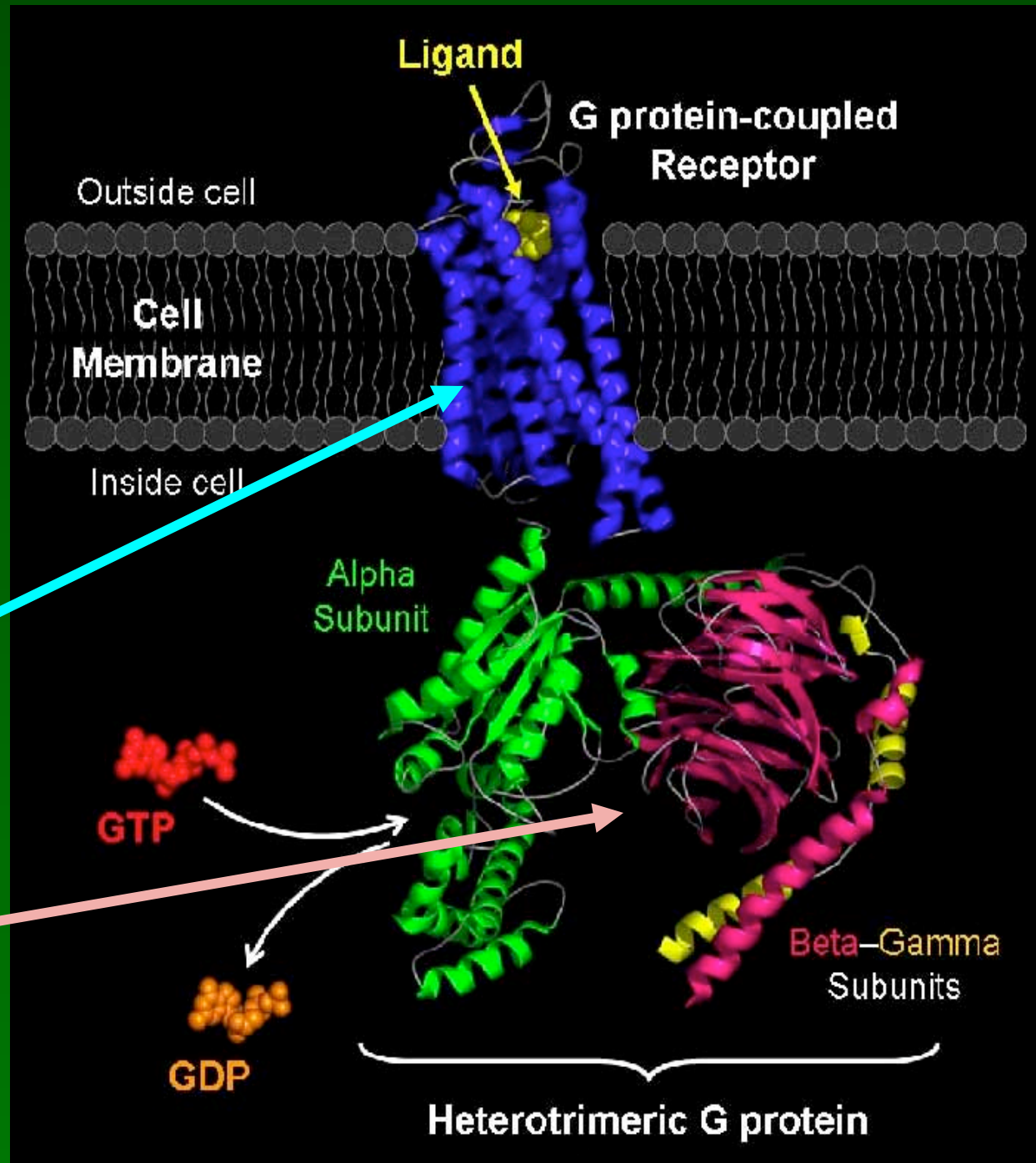


- At least 47 different chemokines known: vary in specific function.

Chemokine Receptors

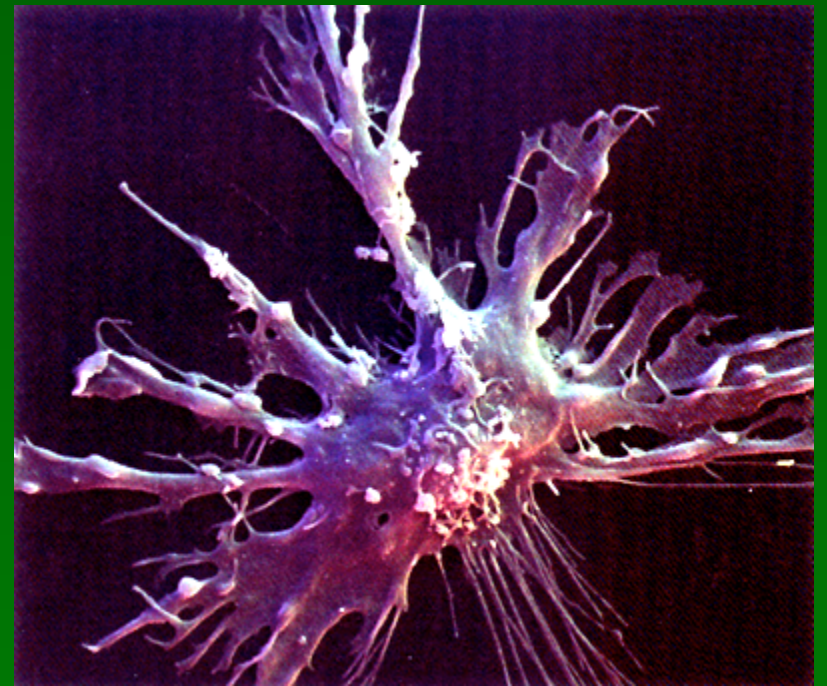
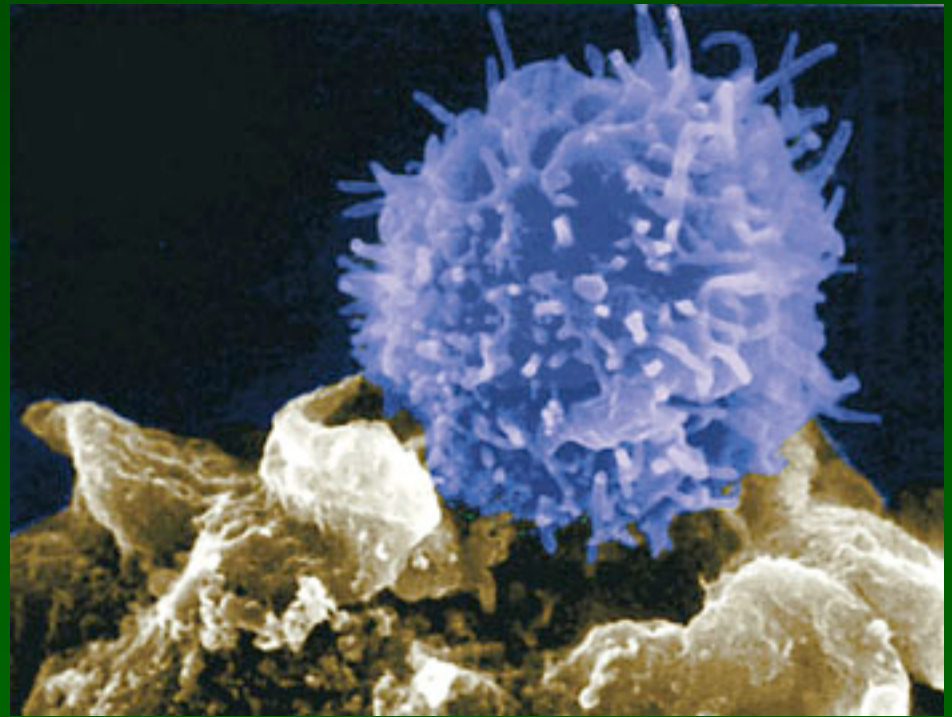
All known chemokine receptors share the same basic structure:

- 7 helices that span plasma membrane
- Signal transduction via a cytoplasmic G protein



CCR5

- Expressed on many immune cells: T cells, macrophages, and dendritic cells.
- Exact function unclear:
no apparent effect detected in individuals without functional copies
- **Used by HIV as co-receptor for binding to and entry into host cells!**



Life Cycle of HIV

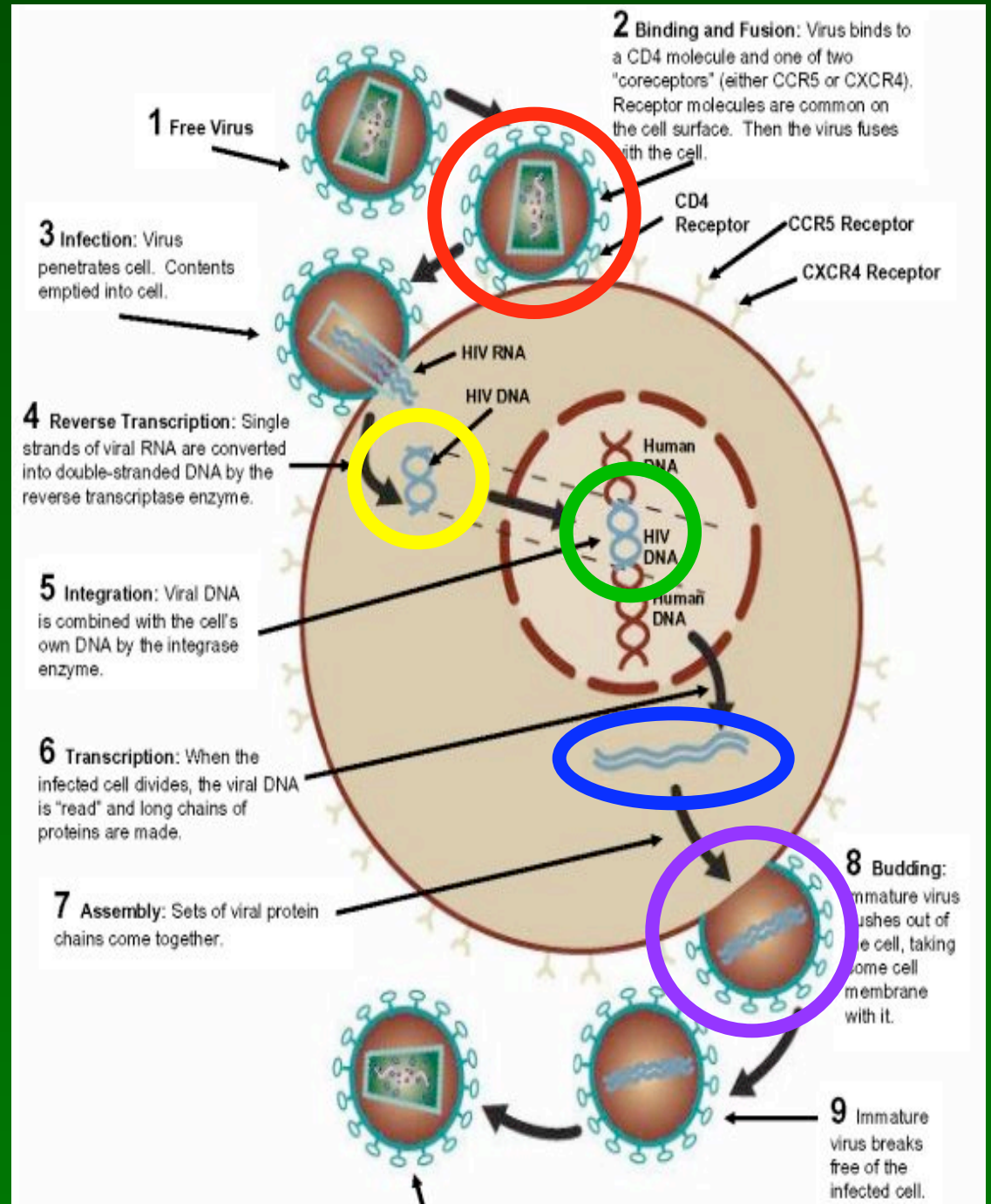
1. Virus docks with receptors on host cell (CD4 + co-receptor)

2. Reverse transcription: viral RNA → DNA

3. Viral DNA inserts into host's DNA

4. Viral RNA transcribed & proteins assembled

5. New virions bud from host cell, killing it



CCR5Δ32

- Allele of CCR5 gene, first described in 1996.

CCR5+ allele (wild-type)

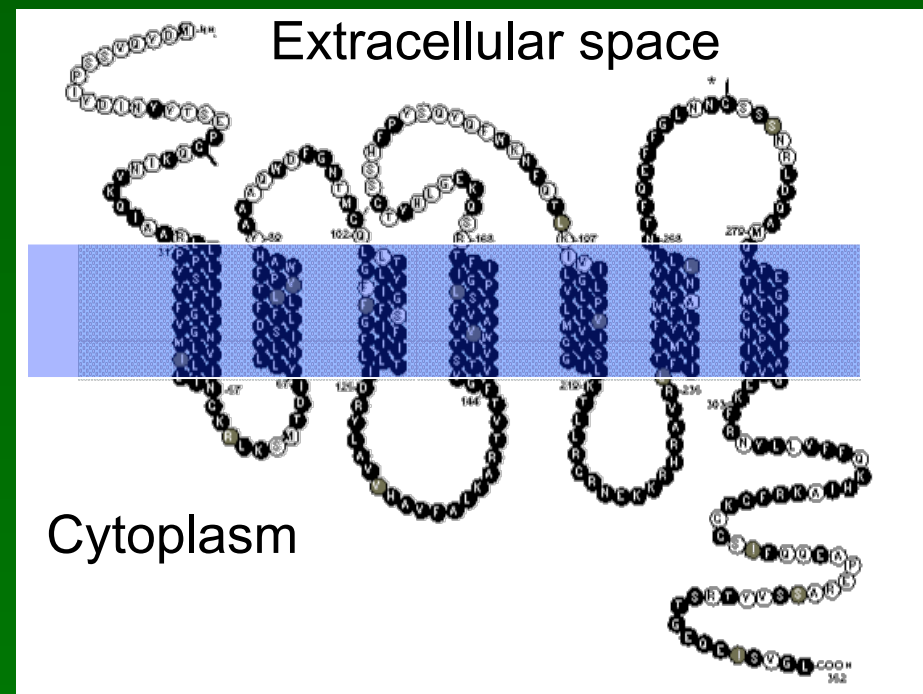
5' . . TTT CCA TAC AGT CAG TAT CAA TTC TGG AAG AAT TTC
CAG ACA TTA AAG ATA GTC ATC TTG GGG CTG GTC CTG . . . 3'

Protein:

FPYSQYQFWKNFQTLKIVILGLVPLLMVICYSGILKTLLRCRNEKKR...

Normal CCR5 protein:

- 7 transmembrane helices
- Ligand-binding extracellular domain
- G-protein-binding cytoplasmic domain



CCR5 Δ 32

- Δ 32 allele has a deletion of 32 bp, causing a frameshift and premature termination of the protein.

CCR5 Δ 32 allele (mutant)

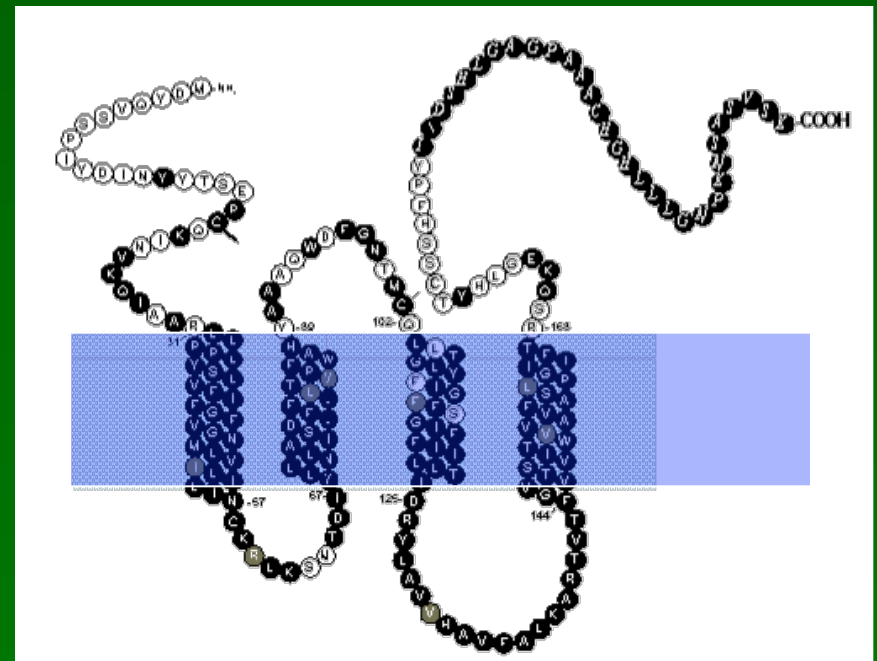
5' ..TTT CCA TAC ~~AGTCAGTATCAATTCTGGAAGAATTCCAGACATT~~
AAA GAT AGT CAT CTT GGG GCT GGT CCT GCC GCT GCT... 3'

Protein:

FPYIKDSHLGAGPAAACHGHLLLGNPKNSASVSK*

Mutant protein:

- Only 4 transmembrane helices
- No G-protein-binding cytoplasmic domain
- Nonfunctional both as a chemokine receptor and as an HIV-coreceptor



CCR5 Δ 32 and HIV

CCR5 genotype dramatically affects both risk of contracting HIV and course of the infection.

CCR5 Genotype	Risk of Contracting HIV	Mean Time to Onset of AIDS Symptoms
+ / +	standard	~ 10 years
+ / Δ 32	standard?	12 – 13 years
Δ 32 / Δ 32	virtually zero	unknown

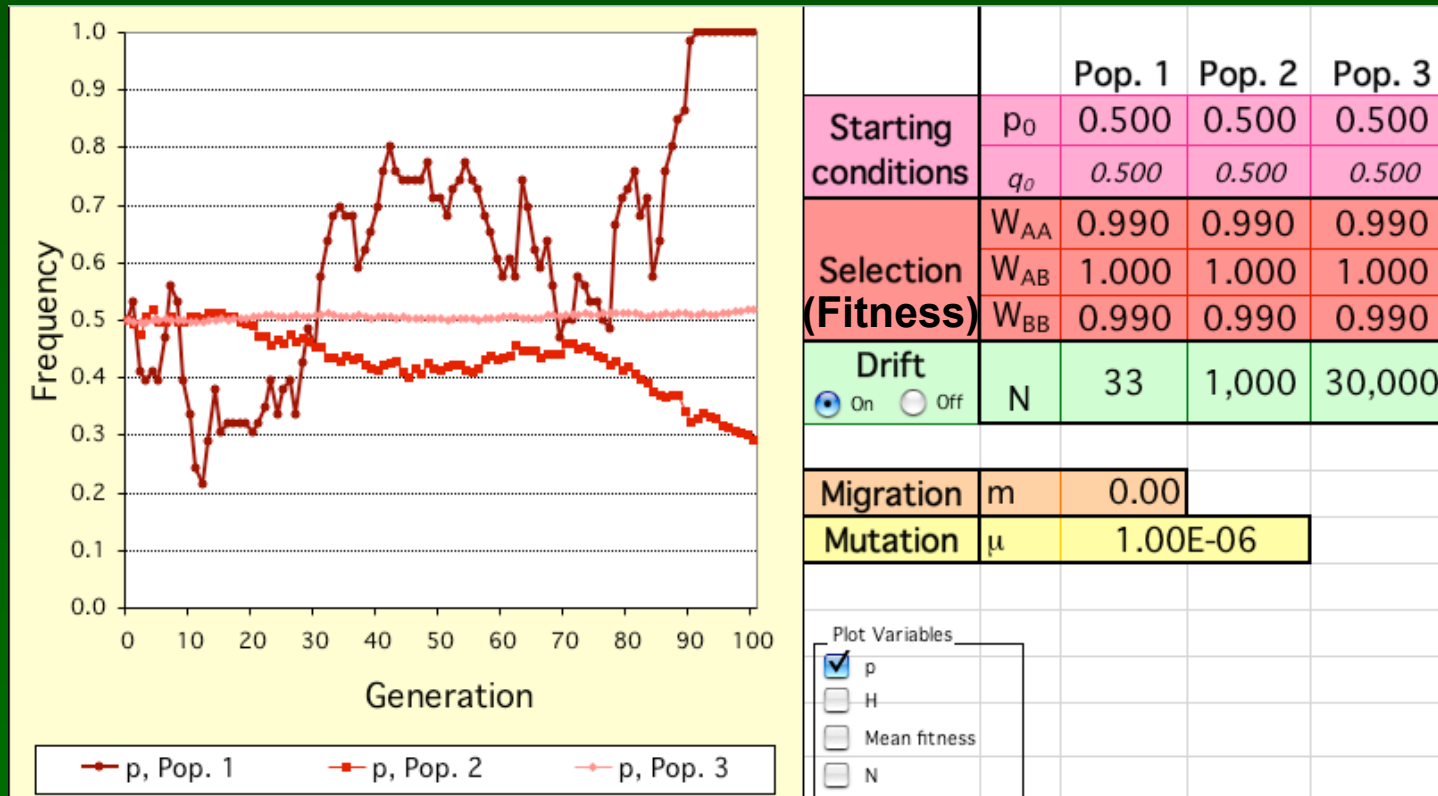
Hypothesis 1: $\Delta 32$ is neutral.

- Initial estimates based on linkage disequilibrium:
 $\Delta 32$ arose ~700 years ago (28 generations).

Stephens *et al.* (1998). *Amer. J. Hum. Gen.* 62: 1507–15

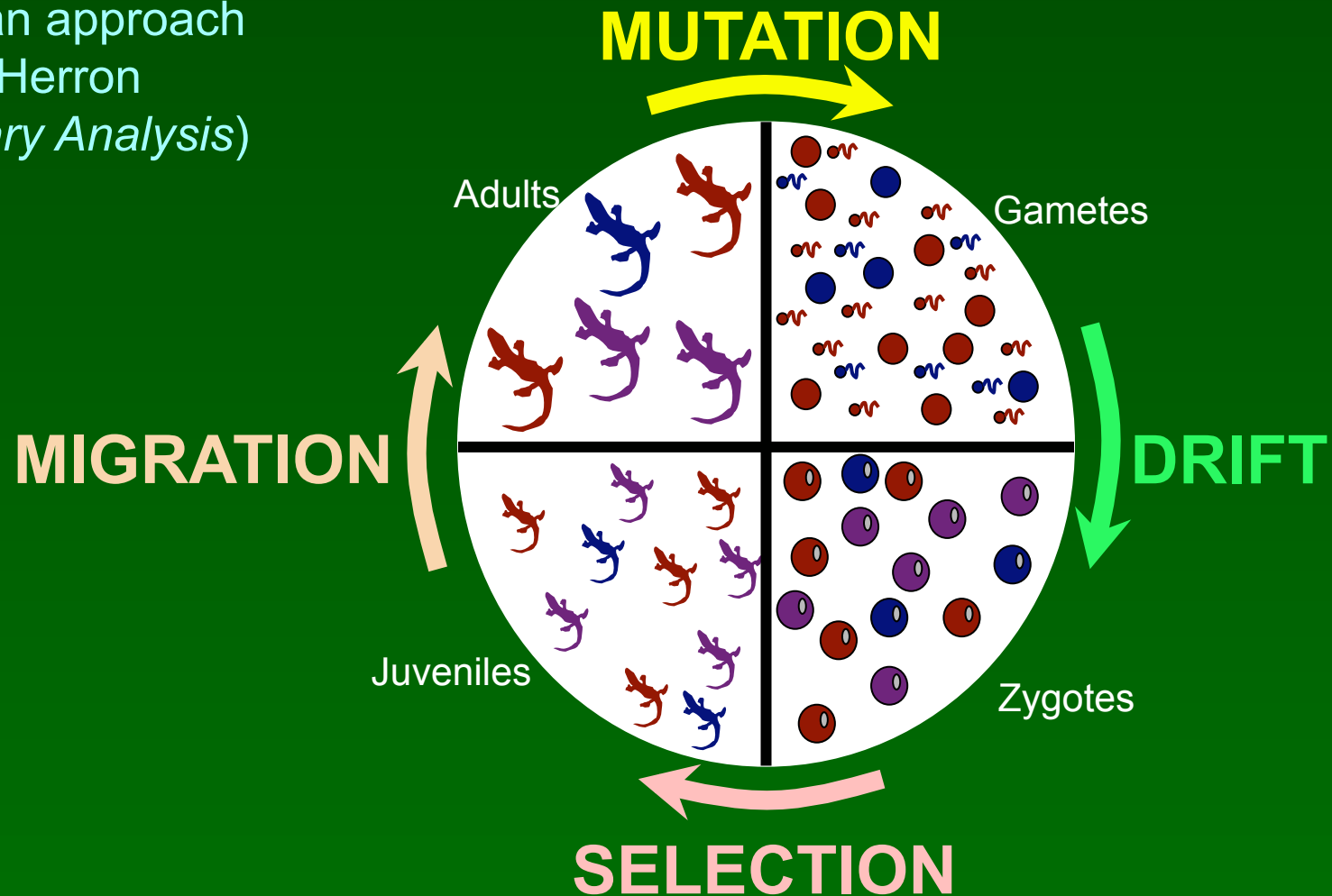
Is it plausible that a new neutral mutation could reach 10% frequency in only 28 generations?

ESTEEM Module: *Deme*



- Models evolution of up to 3 pops. simultaneously
- User controls four evol. forces:
 1. Selection
 2. Drift (Pop. size)
 3. Migration
 4. Mutation

Following an approach
used by J. Herron
(*Evolutionary Analysis*)



- Each generation, *Deme* steps through the life cycle
- Solves for genotype/allele frequencies at each step using standard equations from population genetics

Hypothesis 1: $\Delta 32$ is neutral.

Procedure (Prospective Model)

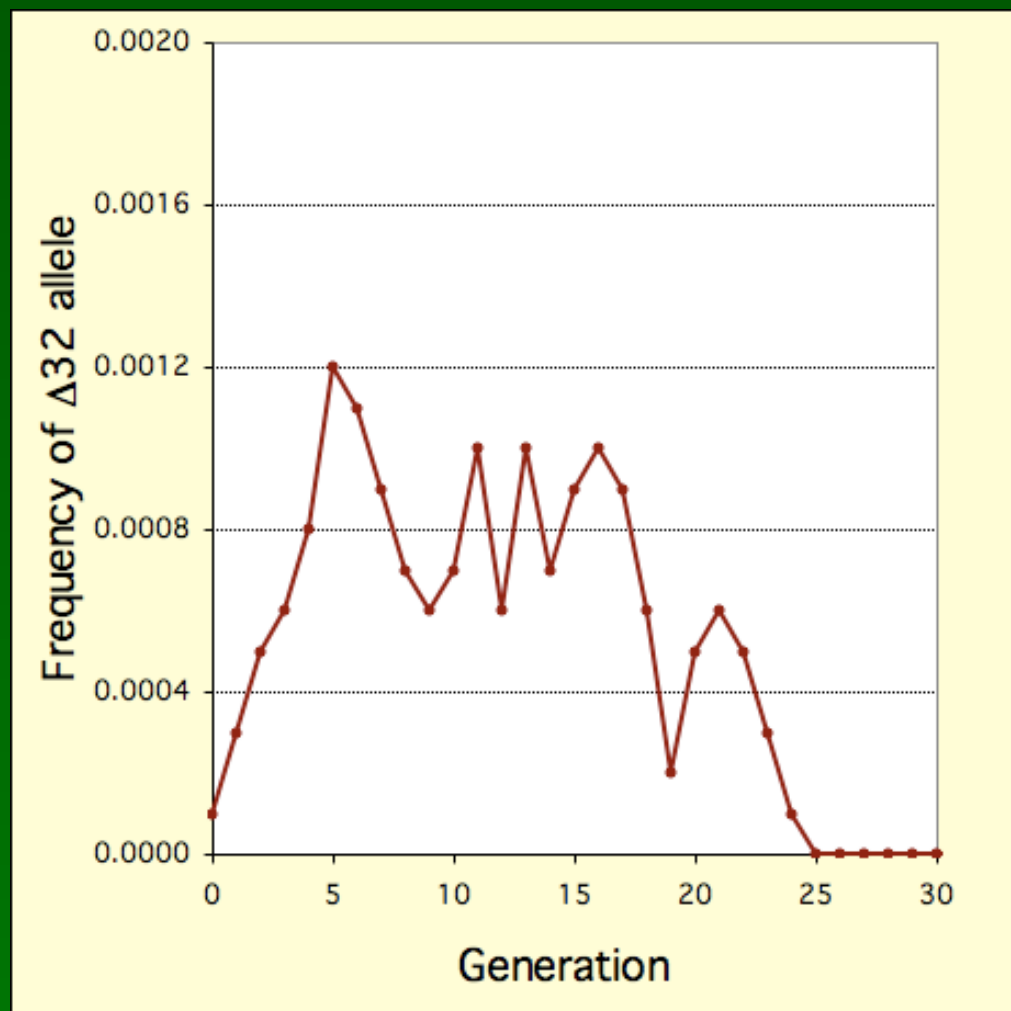
- Set $N_e = 5000$ (following Stephens *et al.*)
- Set $p_0 = 1 / 2N_e$ (one copy of $\Delta 32$)
- Set $W_{\Delta\Delta} = W_{+\Delta} = W_{++} = 1$ (neutrality)
- Calculate time to reach $p = 0.10$

What do you predict will happen to
the frequency of $\Delta 32$ over time?

Hypothesis 1: $\Delta 32$ is neutral.

Problem:

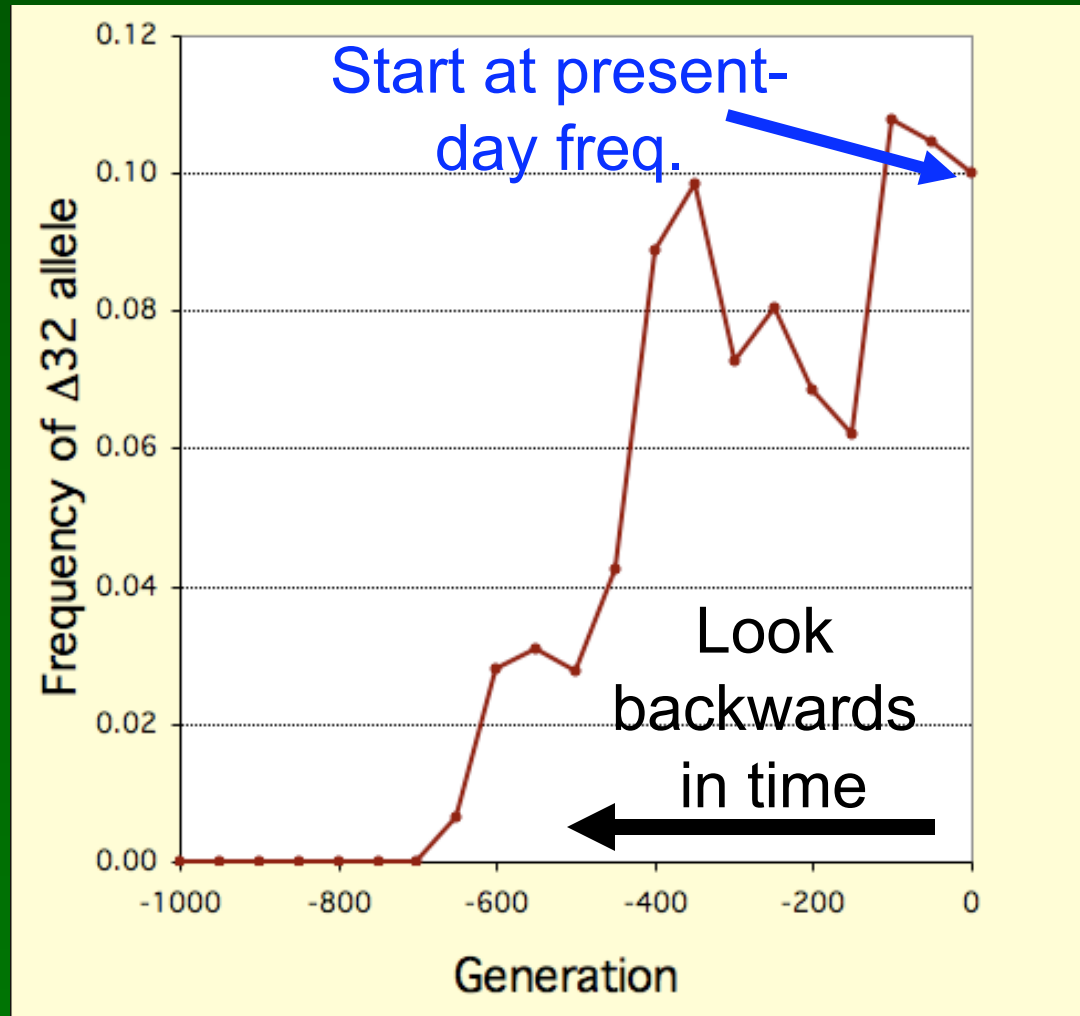
- In ~99% of trials, $\Delta 32$ is lost before reaching $p = 0.10$ — why?
- Our question: *given* that $\Delta 32$ *does* eventually reach $p = 0.10$, how long does that process take?
- Either run 100 trials to get 1 usable result, or reframe the problem.



Hypothesis 1: $\Delta 32$ is neutral.

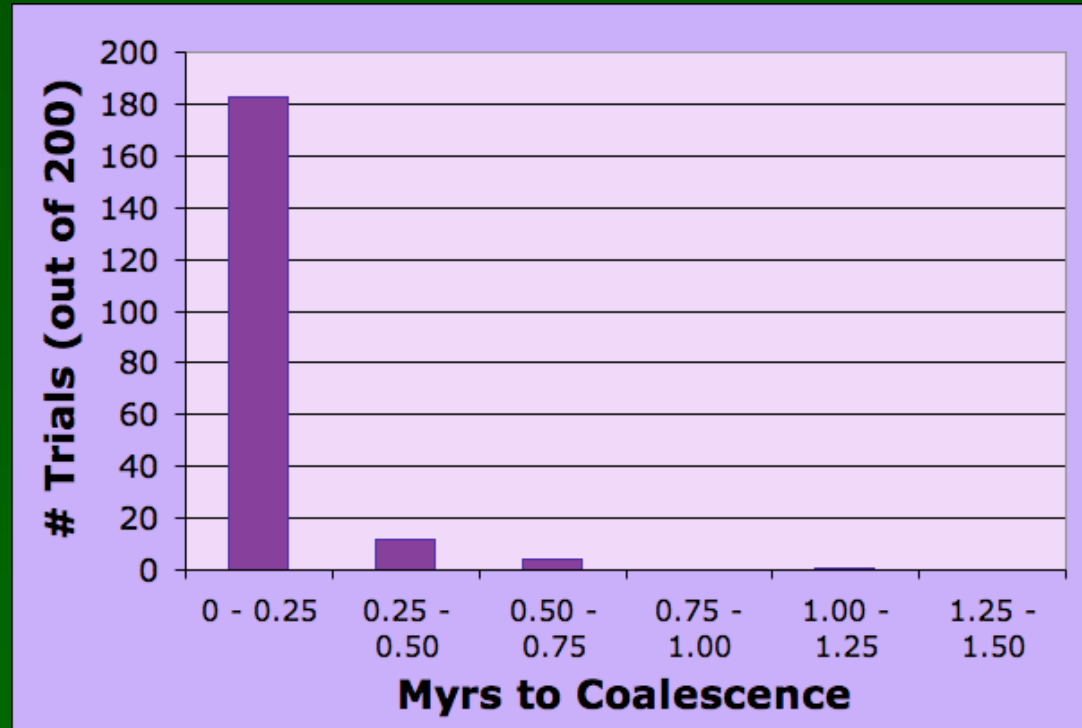
Procedure (Coalescent model)

- Set $p_0 = 0.10$
(observed frequency)
- **Invert** the standard
pop. gen. equations
- Calculate time to
reach $p = 0$



Hypothesis 1: $\Delta 32$ is neutral.

Mean	103,794 y
Std. Dev.	137,879 y



Estimated age
based on LD : 700 y

95% CI under
neutrality : 11,725 –
421,100 y

Assuming LD estimate
is accurate, **reject**
hypoth. of neutrality.

Hypothesis 2:

Historical selective advantage for $\Delta 32$.

Assume $W_{\Delta\Delta} = 1$; $W_{+\Delta} = 1 - s/2$; $W_{++} = 1 - s$
(additive genetic effect)

Question:
What values of s are consistent
with reaching $p = 0.10$ in 700 y?

Hypothesis 2:

Historical selective advantage for $\Delta 32$.

s	Mean age (y)	Variance	95% CI
0.10	2457.4	688.3	1600 - 1400
0.20	1264.4	131.1	825 - 1975
0.25	1001.3	78.8	675 - 1500
0.30	817.1	49.3	550 - 1250
0.35	682.0	28.9	475 - 1000
0.40	573.1	27.0	400 - 850
0.45	491.1	17.0	350 - 750
0.50	417.4	11.3	300 - 600

Hypothesis 2:

Historical selective advantage for $\Delta 32$.

What would $s = 0.25 - 0.45$
mean in the real world?

What could cause this intensity of selection in
Caucasians over the past 700 years?

Is CCR5 Δ 32 a genetic legacy of epidemics past?

“...the cumulative results point to a **selective sweep** and to one with **enormous selective mortality within historic times**, perhaps mediated by a **widespread epidemic**.”

“The bubonic plague, which claimed the lives of 25%–33% of Europeans during the Black Death...and which has had multiple outbreaks before and since, is an obvious candidate... If the mechanism of *Yersinia*-induced macrophage apoptosis involved macrophage chemokine receptor 5, the CCR5- Δ 32 mutation would be an attractive candidate for a strong selective pressure 600–700 years ago.”

—Stephens *et al.*, 1998

Take-home message for students

- “Stochastic” \neq “totally random”!
We can still narrow the probable outcomes or likely parameter values to some range.
- Even fairly simple mathematical models can help us investigate important (and cool!) questions in biology.
- All scientific conclusions are provisional.
New genetic maps of $\Delta 32$: ~ 5000 yrs. old.
Redo analysis: consistent with drift alone!