Gordon Research Conference on Molecular Evolution

Colony Harbortown Marina Resort
Ventura, California
January 28 - February 2, 1996

William R. Atchley, Chair
Walter M. Fitch, Vice-Chair

This Conference is supported by the Alfred Sloan Foundation, the National Science Foundation, the Army Research Office and the Center for Quantitative Genetics at North Carolina State University.

Program for Gordon Conference in Molecular Evolution

Session Speaker	Topic	Discussion Leader	Time
Sunday evening, January 28			
Opening Session		Bill Atchley	7:30-7:45
Early Evolution		Bill Schopf	7:45-8:00
Ford Doolittle	Rooting the tree of life		8:00-8:45
Masami Hasegawa	Origin and early evolution of eukaryotes		9:00-9:45
Monday morning, January 29			
Genome and Organelle Evolution		Dick Hudson	8:30-8:45
Priscilla Tucker	Evolution of the Mammalian Y-chromosome	photo	8:45-9:30
Jeff Palmer	Transfer of organelle genes to the nucleus		9:45-10:30
	Group Photo		**10:50-11:10**
John Doebley	Molecular basis of morphological evolution		11:20-12:05
Monday evening, January 29			
Statistical inference		Masatoshi Nei	7:30-7:45
Nick Goldman	Mathematical modelling of molecular evolution		7:45-8:30
Wen-Hsiung Li	Bootstrap techniques in phylogenetic analyses		8:45-9:30
Tuesday morning, January 30			
Evolution of development		Andy Clark	8:45-9:00
Rudy Raff	Link between development and evolution		9:00-9:45
Diethard Tautz	Evolution of gene networks		10:00-10:45
William Atchley	Evolution of helix-loop-helix transcription factors		11:20-12:05
Tuesday evening, January 30			
Experimental phylogenetics		Walter Fitch	7:30-7:45
J J Bull	Experimental phylogenetics		7:45-8:45
Poster session			8:45-10:00
Wednesday morning, January 31			
Molecular Population Genetics		Morris Goodman	8:45-9:00
Maryellen Ruvolo	Inferring primate evolution with DNA sequence data		9:00-9:45
Charles Aquadro	Determinants of genomic diversity: impact of recombination		10:00-10:45
Dennis Powers	Evolution of gene structure and expression in natural populations		11:20-12:05
Wednesday evening, January 31			
Non-tree-like evolution		Roger Milkman	7:30-7:45
Walter Fitch	Reticulation in HIV evolution		7:45-8:30
Andreas Dress	Phylogenetic networks: Theory, software and visualization		8:45-9:30
Thursday morning, February 1			
Viral evolution		Margaret Kidwell	8:45-9:00
Paul Sharp	Origins and evolution of AIDS viruses		9:00-9:45
Susan Wessler	Transposible elements and the evolution of gene expression		10:00-10:45
Walter Gilbert	New Arguments for Old Introns		11:20-12:05
Thursday evening, February 1			
Pattern and Function		Jeff Thorne	7:30-7:45
Nancy Maizels	Phylogeny by function: origin of tRNA is in replication		7:45-8:30
Marcella McClure	Identifying patterns in distantly related proteins		8:45-9:30

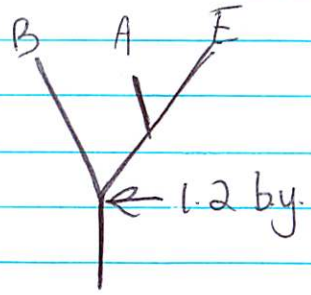
There will be a 15 minute question and discussion period following each talk, In the mornings, there will be a coffee break from 11:00 - 11:20 am (between the second and third talks).

1/28/96

Molecular Evolution Gordon Conference

Bill Schopf

- R. Doolittle et al Science
  - suggest eukaryotic & bacterial common ancestor was ~1.2 billion years ago



- why he doesn't think so
  - Stromatolites ... were present
    - cyanobacterial fossils

WF Doolittle - Constructive Evolution

- Complex & necessary interaction betw. molecules can be established by strictly neutral means

- Evolutions big steps
  - incr. in complexity → usu. given some adaptationist explanation.

- Neutral alternative
  - but no intrinsic reason for direction

- Complexity ... Maynard Smith & Szathmar 1995

possible measures of complexity
- genome content
- cell types?
- body size
- interactions between molecules

THIS PROCESS IS DEPENDENT ON RELAXED SELECTION FOR SIMPLICITY

- w RNA ped example
  why not more complex in eubacteria

Neutral models for mec. in complexity

① RNA editing
- C→U editing

- imagine there is a C→U deaminase that f(x)'s on tRNA
- imagine it recognized a site w/ a U
- then if there is a U→C mutation at this site then these would be OK.

② Codon-capture & novel genetic codes Osawa – Evolution of the genetic code

③ A. Lambowitz

- one intron reqs. CYT-18 protein for splicing
- then other introns may be stabilized by prot-prot interactions or prot-RNA
- then these may lose splicing ability

④ peptidyl-transferase

⑤ gene duplication → each become dependent on diff. molecules

⑥ molecular drive

⑦ polite DNA · E. Zuckerkandl

M. Hasegawa - Origin & Evolution of Eukaryotes -- Rooting the tree of life

Uncertainties in assumptions

MOLPHY 2.2

133.58.12.20 /pub/molphy*

Protml

Advantages

- stochastic subst.
- robust (somewhat) to rate variation
- improves subst. model
- can evaluate total evidence from multiple genes

Amino-Acid substitution models

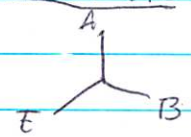
both assume equilib. freq are average freq.

- Dayhoff
- Jones, Taylor, Thornton (JTT)

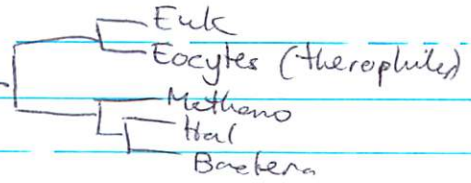
∴ can be improved by using actual aa freq. of each protein

Rooting tree of life

Woese

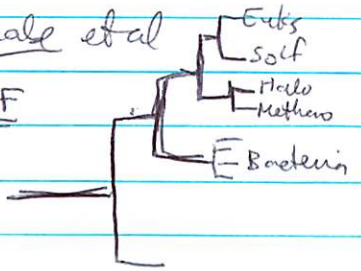


Lake



Iwabe et al

EF



ATPase

- showed sim. comparison
- but due to paralogous comparison

Brown & Doolittle

- AA tRNA synthetase
  - Ile
  - Val
  - Leu

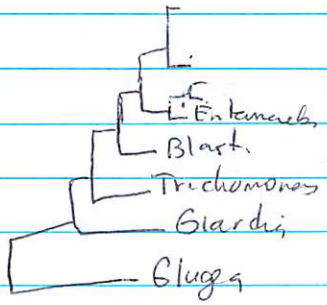
Early evolution of eukaryotes

- phylogenetic placement of amitochondrial eukaryotes would be v. useful

Problems w/ rRNA tree?
- GC content variation (e.g. GIARDIA ≅ 75%)

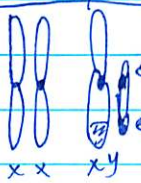
- ∴ use protein sequences

GIARDIA
- EF1-α
  - aa composition is not v. diff. from other EF1α despite GC bias



Priscilla Tucker - Evolution of sry (male sex determining locus)

Mammalian Sex Chromos

① 

- only small region of homology
- ∴ v. little recombination of Y chrom. specific genes

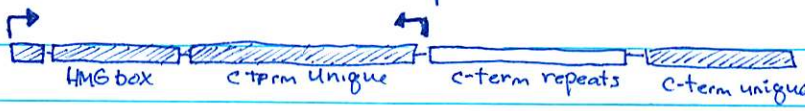
② appears to have evolved once in mammals
③ Y chromosome is late replicating
④ chromosome specific effects
- ⓐ hitch-hiking
- ⓑ background selection
- ⓒ Hell-Robertson effect
- ⓓ bottleneck effects
- ⓔ Muller's ratchet
- ⓕ male-driven molecular evolution

ⓐ ⓑ ⓒ more active in non-recombining regions

Other chrom. spec. effects
⑤ mutation

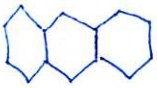
Sry

- involved in male tester development

SRY 

- studied SRY in Murine genera
- what is variation w/in & betw. species?
- v. low variation in unique sequence regions
- multiple copies on Y in some species
  - some variation among paralogs in C-term repeat regions

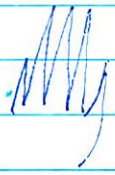
6



Explanation

GAGAGA

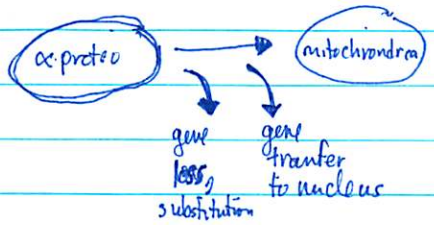
GAGA



Constructive & Destructive Neutral Evolution

J. Palmer - Tranfer of Organellar Genes

Mitochondria



When did transfer occur?

- great variation in distribution / # of genes in diff. organelles

Gene Substitution

chloroplast transfer:

tufA	glu syn
Fe-SOD	Cu-Zn SOD
F-1,6-BP	mal. dehyd
GAPDH	TPI
PGI	aldolase

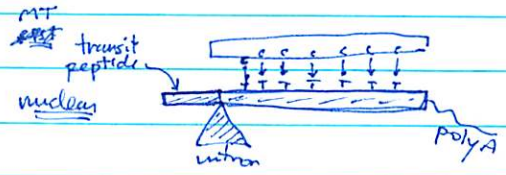
} CPST substitution

Gene Transfer

- probably not still going on in yeast & animal mt DNA because of change in genetic code

Recent Transfers?

- probe southerns w/ mt or CPST probes
- CoxII lost in some legumes nuclei
  - many C→T changes in nuclear gene
  - (the mt gene would be edited)
  - most likely edited message was transfered
- some have it in nucleus, some in mt

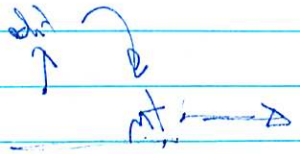


How much is this due to C→T changes.

CoxII & Cob

- longer branch lengths in genes after transfer to nucleus

RNA editing
C
↓
U



John Doebley - Molecular Basis of Morph. Evolution in Maize

Teosinte & Maize
- same biological species
- allows genetic analysis

Color
- teosinte doesn't have colored kernels, but does make anthocyanin
- Anthocyanin pathway
  Enzymatic vs. regulator genes

- Tester crosses
  - teosinte has dominant f(x) alleles for all enzymatic loci

  - R1 -- teosinte has recessive alleles, ∴ crosses w/ mutants have no kernal anthocyanin

  - C1 --- teosinte also recessive

C1
- regulated by VP1 (VP1 activates C1)
- hypothesis
  - started w/ two indep. pathways — VP1 → seed mature; C1 → anthocyanin
  - then these fused by VP1 activating C1 w/ acquisition of duplication in C1 promoter
- but duplication thought to cause this was present in teosinte

Gene exchange?

HKA test ... nothing significant

C1.
- lowest ⊖ value for any maize locus

Paramutation
- one allele heritably alters another allele in heterozygotes at high frequency

maybe due to methylation state

Morphological Differences

QTL mapping

maize × teosinte → F1 → F2

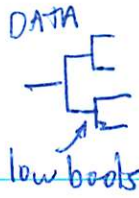
- many phenotypes mapped to long arm chromosome
- submapped to region
- this region contained teosinte branched gene

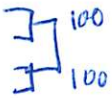
- used mutator line cross

mu × tsb

↓

wt tsb or (mu tsb) = mutant

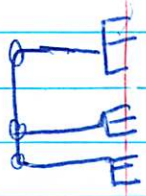




M.Nei

Reconstruction of Trees

1) Estimation of branch lengths & w/ topology
2) Estimation of topology



- Rannala & Yang (manuscript)
  - ~~Bayesian~~
  - treats tree as random variable

3) Does a "sophisticated" substitution model improve phyl. reconstruction

4) Does the substitutional model remain the same over time

N. Goldman:

- what else is there in aligned sequences?

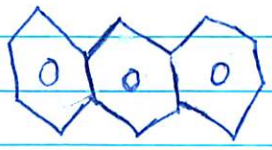
1. Substitution matrices
2. but
   - diff. betw. codon positions
   - diff betw. positions
3. Codon-based models

New Stuff?

1. Robustness (diff. models)
2. Confidence

Testing Models

① Test a goodness of fit - are certain parameters important?
- transition - transversion ?



Future things?
① invariant sites
② 2ary structure
③ 3D structure
④ convergence & parallel changes
⑤ coding constraints

W-H Li - Bootstraps

Smaller sequences ... more affected by variance w/ many parameters

Bootstrap
- what do % values mean

- boot % is less that prob. that tree is correct
- this underestimation incr. w/ more taxa

- with more taxa then less likely to get a particular pattern by random

- the important parameter is #. of alternative trees

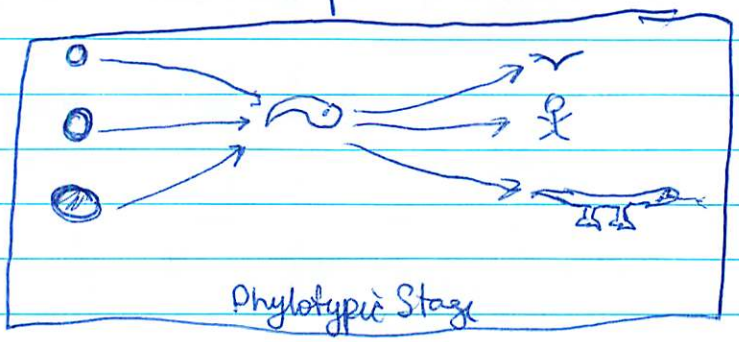
A Clark

Homeotic genes

Evolutionary genetics of development
Evolutionary of body plan

R. Raff

- Evolution of Development



- suggests that the phylotypic stage is the stage with the most protein-protein & other macromolecule interactions
  - like WFD "constructive evolution"

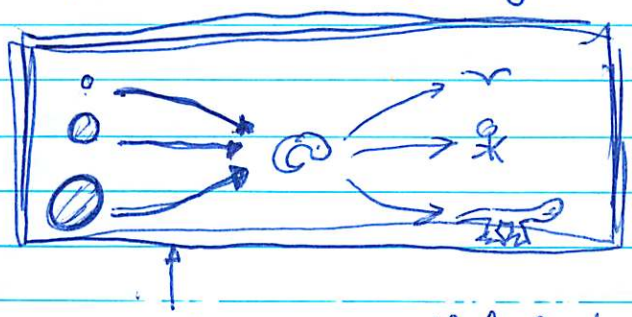
- Sea Urchins
  - two relatively closely related species
  - start out dvlp. v. differently
  - ~~evolution of phy~~

Evolution of phyla level body plan

D. Tautz - Evolution of Gene Networks (in embryos)

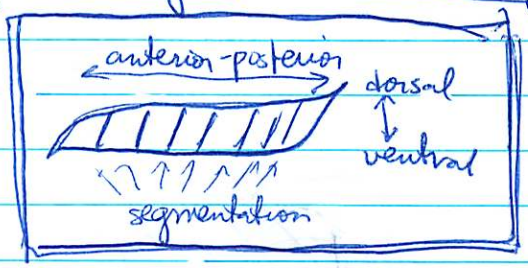
Molecular Comparative Embryology

- defintion of homology may be diff. from other cases



- are the processes that lead to phylotypic stage similar betw. species?

- segmentation genes (bauplan genes) → homeotic genes → cellular target genes



- embryo —(mostly tx. factors)→ bauplan → cell fates

- the tx. factors form gradients -- these work well in Drosophila embryo because all one cell essentially

- to remember
  ① lots of redundancy
  ② factors are expressed in phases

Many bauplan genes have similar f(x) to gene homologs in other species

∴ Regulatory gene networks appear to be highly conserved

But Not all genes retain their functions

↳ Depends on how you define f(x).

FAST EVOLVING GENES

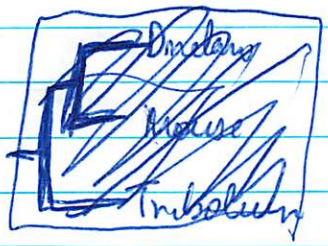
① 100 randomly chosen embryonic cDNAs from D. melano

↓

hybridize to gDNA from flies, house fly, beetles

↓

clone from close relatives (D. yakuba)



Bob Atchley - Evolution of HLH proteins

Common tx. factor motfs
- Zn finger
- HTH
- HLH
- leucine zippers
- steroid receptors

HLH motif

- highly conserved: found in yeast → mammals
- often function in tx. activators
- examples
  - many many mod
    - Myc, Max, Mad
    - MyoD
    - E12, ITF1
    - daughterless
- phylogenetic analysis

Cell Proliferation & Differentiation

Myc complex

-

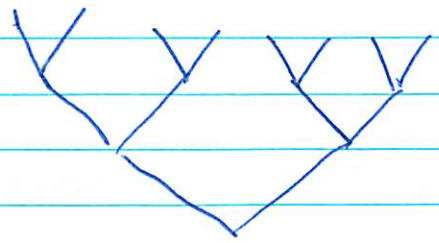
Jim Bull - Experimental Phylogenetics

Evolving bacteriophage in the lab

John H - never seen real data

1) Creating a known phylogeny

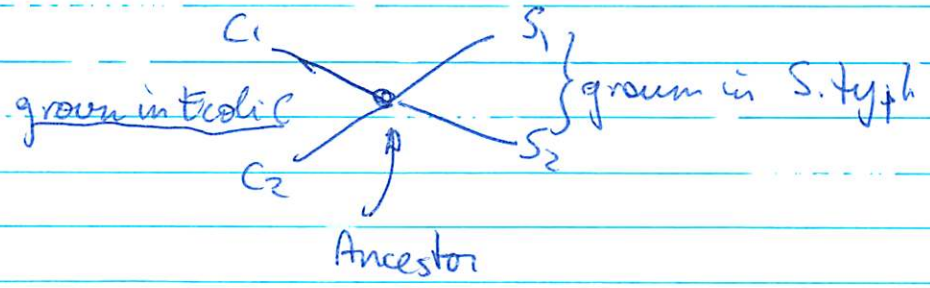
SHOULD USE MIXED POPULATIONS AT NODES



- 1st phylogeny was symmetric
- grown in presence of mutagen
- restriction maps
- all methods gave right tree
- reconstruction of ancestral states v. good

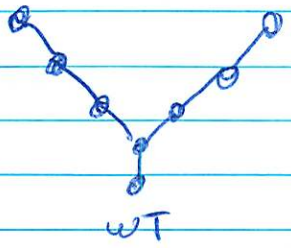
2) Convergent evolution in ϕX174

- grown in phage chemostat
- adaptation to high T°



- 40% of sites that change were convergent

3) Convergence in T7



6x trees like this

- grown in mutagen
- in every phylogeny get specific deletion

4) Antisense

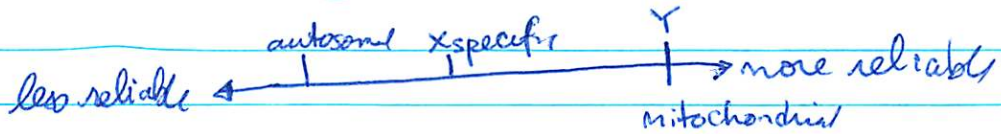
virus
↓
add antisense
↓
evolves resistance?
↓
add new antisense

Morris Goodman - Globin Genes

Mary-Ellen Ruvolo
- Neanderthal is an issue solved?

Using Multiple Genes

Single Gene Problems
① ancestral polymorphism
- how likely are you to get gene-tree ≠ species tree mismatches? see Takahata & Nei



Likelihood Ratio Method
- exclude some datasets
① those w/ ↓ trichotomy
② those w/ gene conversion
- can lead to lots of homoplasy
③ those w/ ancestral polymorphic

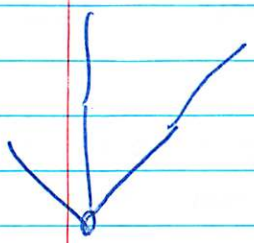
- combine closely linked genes
(Make sure to compare chimps, gorilla too)

Support Human-Chimp
- 11 gene sets
Support Gorilla Chimp
- 2 data sets
Support Gorilla-Human
- 1 data set

GGPDH

- Miyata et al 1987 (sperm vs. egg mutation)

If there was a trichotomy. plus uneven branch lengths



what about recombination being dependent on heterozygosity

Do these models assume recomb. in a particular region is constant over time. Variation might be important

Chip Aquadro

Recombination & Variation

At equilibrium
- neutral = heterozygosity = $4N_e u$
- selection = reduces heterozyg.
  - range of affect depends on recomb.

D. melanogaster
- can calculate rate of recomb. per physical unit
- highly variable
- levels of variability (w/in population) strongly correlated to recombination/physical unit

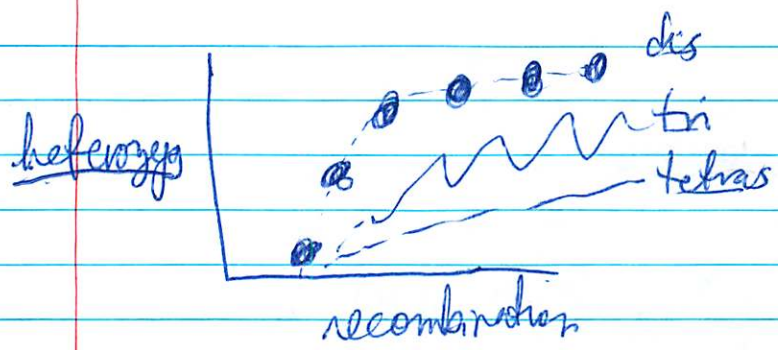
Explanations

both predict you should have a correlation betw. recomb & divergence betw. species

1. genes in regions of low recomb. are f(x) constrained
2. recombination is mutagenic
3. selective sweeps + hitchhiking
4. background selection against deleterious mutations

This is not seen

Background Selection Model
- only those regions free of deleterious mutations will ~~accumulate~~ persist in a population



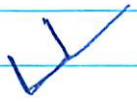
Codon Bias } may also be effected by
GC
recombination

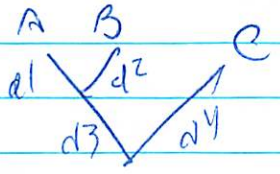
Dennis Powers

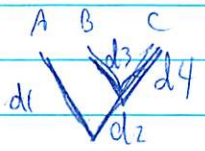
Selection?

Why fish?

tx induces recomb







$A'C' = d1 + d2 + d4$

$A'B' = d1 + d2$

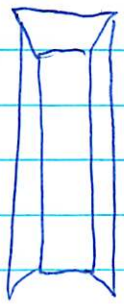
$B'C' = d2 + d3 + d4$

$AC = d1 + d2 + d4$

$AB = d1 + d2 + d3$

$BC = d3 + d4$

$A'C'$

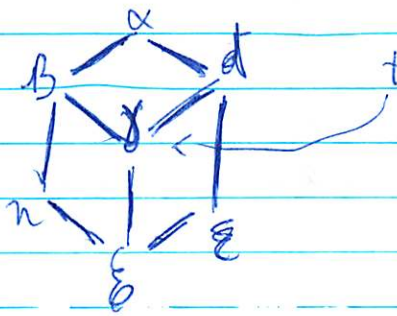


WFitch - Networks

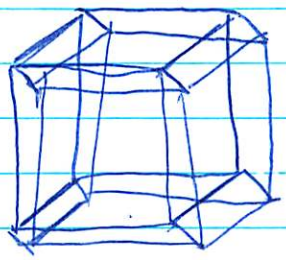
what to do if you have multiple equally "good" trees

① consensus
- strict-rule
- semi-strict → (keep those patterns that are not contradictory w/ others)
- majority rule

But consensus trees can be a problem:
- lose information



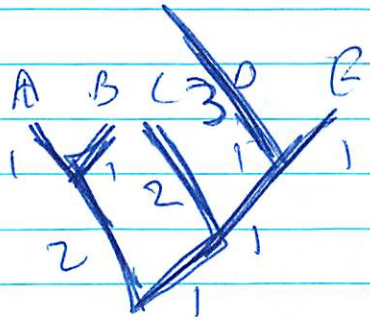
this sequence could arise by multiple paths



- sometimes this deals w/ parallel changes and sometimes it deals w/ ambiguity

- Networks can represent consensus trees w/o including additional trees

	A	B	C	D	E	
A	0	2	6	6	6	5
B	2	0	6	6	6	5
C	6	6	0	4	4	5
D	6	6	4	0	2	4.5
E	6	6	4	2	0	4.5



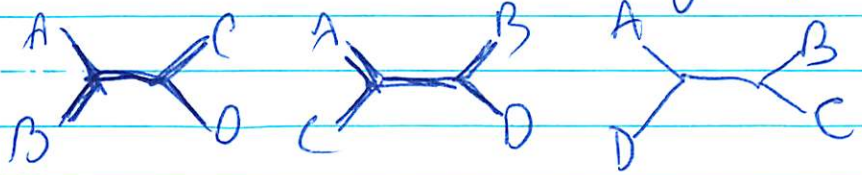
A	0	
B	0	
C	4	0
D	4	0
E	4	0
	12	

Networks – Split Decomposition

Represent non-treelike patterns when they are there

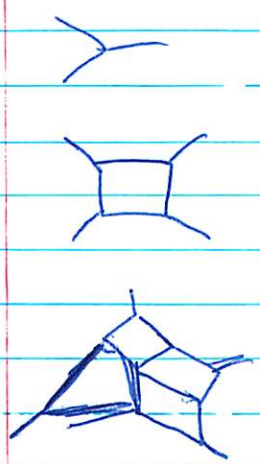
Buneman's Global Principle for trees

Suggest that w/ 4 taxa tree rather than trying to determine correct topology – to determine the least likely topology

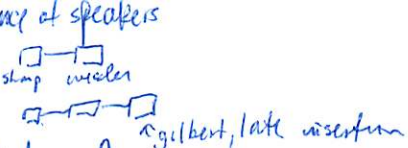


Most methods would pick 1, 2 or 3.
He suggests rejecting 1, 2 or 3!

This is similar to fourier analysis which decomposes periodic signals



sequence of speakers



M.K. dwell

Eukaryotic transposable elements

① Class I = retro-elements
② Class II = DNA elements

Link betw. retroelements & viruses

Evolution of TPONs

Mode of evolution
① Ancient/modern?
② Phylogeny
③ Mode of transmission

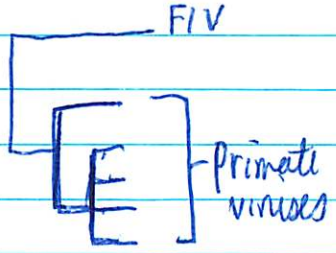
Host-element relationships
① Strictly parasitic
② Mutualistic
③ Sequential parasitic/mutualistic

Gene conversion/recombination frequency?
Host genes & cross-species transfer?

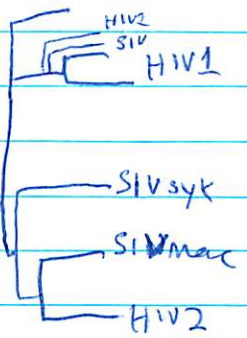
P. Sharp • Evolution of HIV

HIV evolution

1. groups w/ Lentiviruses
2. v. high rate
3. appear to have been many cross-species transfers
4. suggests no good molecular clock and that to get times of evolution of the virus to find regions of the tree w/ no cross-species events



is there a tendency of recombinants in particular regions?

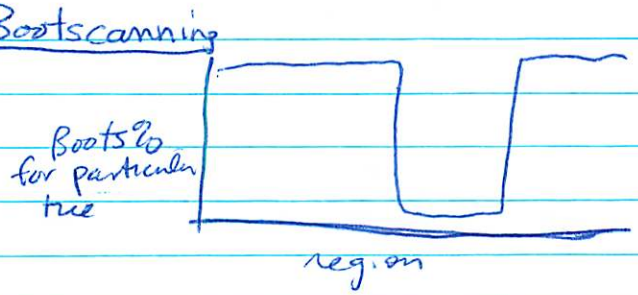


HIV2

Recombination in Retros/HIV

- compare phylogenies of diff. genes from same virus
- for this to occur an individual must be infected by two strains

Bootscanning



Sue Wessler

Are tpons useful?
- McClintock suggested TPONs were "controllers"

TPase
Junkyard
Unity

Higher plants

- much of this is "junk"
- but how much is dead?

Two Genes
R = tx. activator
Waxy =

① Analyze mutant alleles
- how do TPON's alter expression

② TPON element families

Waxy Mutations
- isolate spontaneous mutations
- many insertions are stable because they do not encode tpase and thus only hop when tpase is expressed elsewhere

- many mutation/insertions do not have much phenotypic effect

Retrotransposons in Waxy
- many are spliced
- but not all the splicing is accurate
- some differences betw. different tissues

Ds elements transpose locally

TPONs affect genes by excision too

Wax K - jumped into waxy
- used as query sequence
- found many plant genes w/ remnants

only in grasses [Tourists (>10000 copies)
- similar to heartbreaker (1000s of copies - - maybe?)

Stowaway - in monocots & dicots

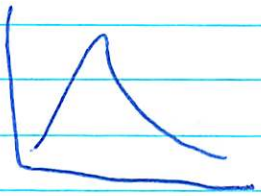
Both prefer to insert in TA regions

MITES
miniature-inverted-repeat-transposable-element

W. Gilbert

Exon shuffling - by more space ... get more recomb. making

Exon spectra



- appears that exons are reasonably correlated in size

Late	Data	Early
gain	Phylogy - not in bacteria - but in verts	loss
	introns at sim. positions	
disagree w/ data	Correlation of introns w/ 3D structure	agree w/ data

Other stuff

- introns can be lost by rev. tx.

- suggest correlation betw. modules & exons
- but if splicing is inaccurate then those insertions into ~~exon~~ modules might be more deleterious

Intron Phase

- suggests that late insertion would not prefer any phase"

- suggests exon shuffling favors in phase exons

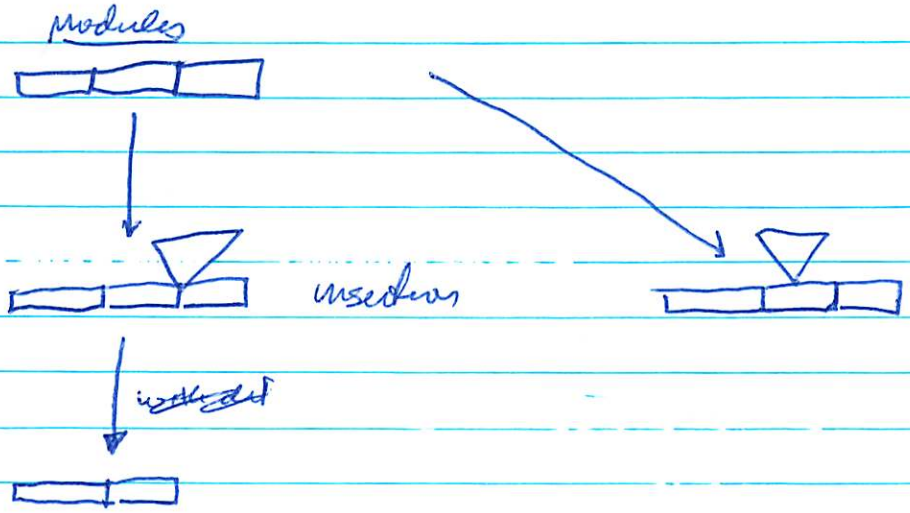
but

1. selection agst. out of phase
2. GC bias
3. should get homologous introns in same position

Models

- are introns in linkers of proteins

		nt
Size of exons	14 aa	: 42
	26 aa	: 78
	37 aa	: 111



If splicing is not-complete then the one at ~~upon~~ module border might be less damaging.

Jeff Thorne

Aligning sequences
- evolutionary framework
- statistical basis
- feasible -

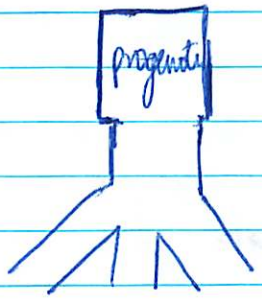
→ Using a tree to aid alignment
- Higgins
- Hein
- Feng & Doolittle
- Sankoff et al

Statistical basis
- Lloyd Allison
  - integrate over all alignments allow each to contribute

- PNAS 91:1059 JMB 235:1501 Science 262:208

Uses of evolution
- Constraint identification
- multiple seq. align
- motifs
- structure predict
- database search

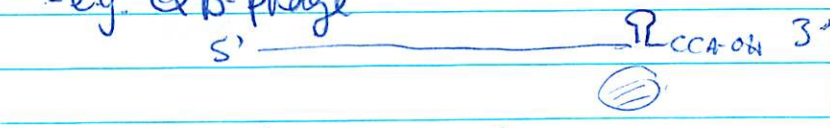
N. Mavyels



Molecular Fossils
- catalytic RNA
- suggests that those molecules that interact w/ lots of other molecules are MOST likely to remain as molecular fossils

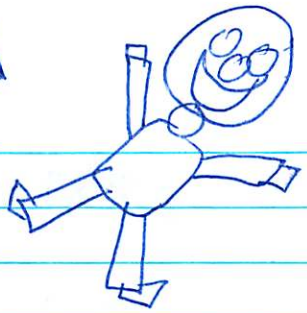
NOW I'D LIKE TO GO BACK TO A SIMPLER TIME

RNA Genomes - Replication Problems
① specificity
② telomere?
- circular
- special replication mechanism
- extra stuff at end that doesn't matter
- telomerase
- tag end
- e.g. Qβ-phage



- replicase binds at end
- CCA works like telomere because there is a CCA adding enzyme
- also in mitochondrial plasmids




Contemporary tRNAs

↑ added by CCA addy enzyme

RNA component of RNAse P cleaves 5' leader from tRNA precursor

A. Leambourty

- Neurospora plasmid (Mauriceville)
- DS DNA plasmid
- replicate by rolling circle tx. w/ long RNA w/ multiple copies

DNA
↓
Multipl RNAs
↓
cleave
↓
rev. txase

How did tRNA get involved in translation

- suggests aas may have been used to change tRNA structure to allow better priming.

tRNA in replication → tRNA in translation

↓

- maybe aa tRNA synthesases evolved to charge tRNA for replication

↓

translation

May explain

① diversity of tRNA syn. structure

Suggests ~~tRNA~~ tRNA may have evolved of two domain

CCA Adding enzyme

- some species add CCA
- some don't (CCA is already encoded)

M. McClure

- Reverse Transcriptase
  - highly divergent
  - ∴ needed better alignment algorithm

How well do multiple alignment programs do?
- every one uses pairwise alignments

Hidden Markov Model

- ordered series of motifs
- Model length
- Training Set Size
- Distribution of Similarity in training set
- Observed frequency of aa

dUTPase

- only found in two distinct retro-elements
- OS sequences

dUMP
↑ dUTPase
thymine ← dUTP
↓
uracil in DNA ← deamination
↓ ung
death