NAS - Genetics + Origin of Species 1.31.97

Francisco Ayala - Intro + Sales of Book

Wally Gilbert - Origin of Genes

Introns Early } says both sides use same data to boost their
Introns Late } argument

1. Phylogeny - bacteria no introns
2. Homologous genes - intron position } Data - but interpretation
3. Correlation w/ 3D position } is up for grabs

Exon shuffling

- says introns should have "dramatic" impact of recombination rate - and allows mixing of different pieces.
- Exon theory of genes
- suggest that proteins are made up of "words" and it is easier to make a sentence out of words than random letters.

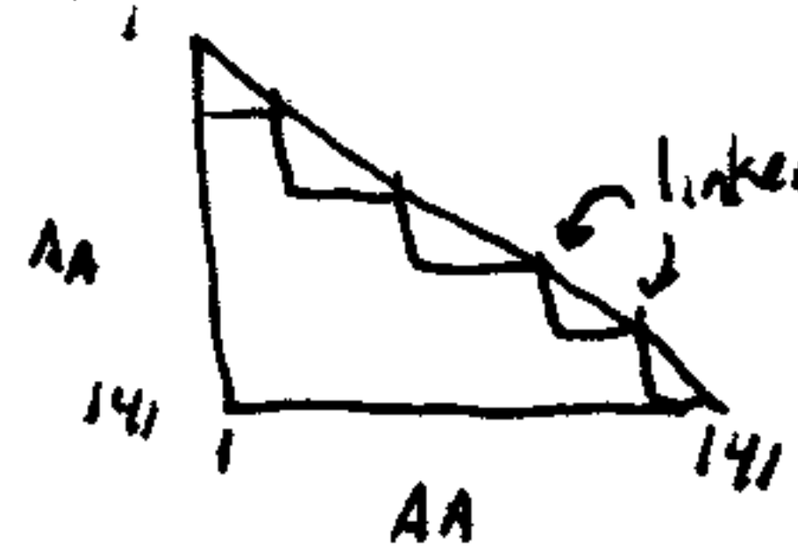
∴ would expect exons to correlate w/ structure

Intron Phase

- position of intron w/in codon
- suggests introns would not be in phase if introns inserted late.
- suggest separate exons should be in phase if exon shuffling important.
- in data ... there is an excess of symmetrical exons

Protein modules

- introns early predicts exons should represent "modules" in 3-D space.
- but hard to determine where the boundaries are.
- but if splice signals are important... phase may affect splicing
- but selection after insertion could affect position
- he suggests that introns late couldn't explain this w/o excess biochemical pleading
- are introns correlated at all w/ particular amino acids or flanking n.t. sequence
- are introns correlated w/ degree of aa conser



linker regions - positions between groups of aa that are very close together.

So.. data looks like there is a bias ... what else could

No sequence bias
No aa bias
No surface bias
No AGG or AGGT bias

} no obvious bias in regions near linkers

"Ancient introns" .. conserved positions → but if introns stay at same position must be some selection. So... why couldn't this selection have driven insertion of introns to these positions

Linker regions

~15 aa	21A°
~22 aa	28A°
~30 aa	33A°

with your ancient introns must need some selection to keep them in position so why couldn't late insertions do this

Lewontin suggests we need to know more information about splicing signals to know if there is residual bias.

M. Perugrino asked about whether splicing failure could explain phase correlations.

Margaret Kidwell - TPON variation in animals + plants

Dobzhansky well aware of ionizing radiation as a mutagen but not aware of tpons role as mutagen.

A. Intro
B. Types of variation caused by TPON's
C. Co-evolution of TEs + host

Transposons

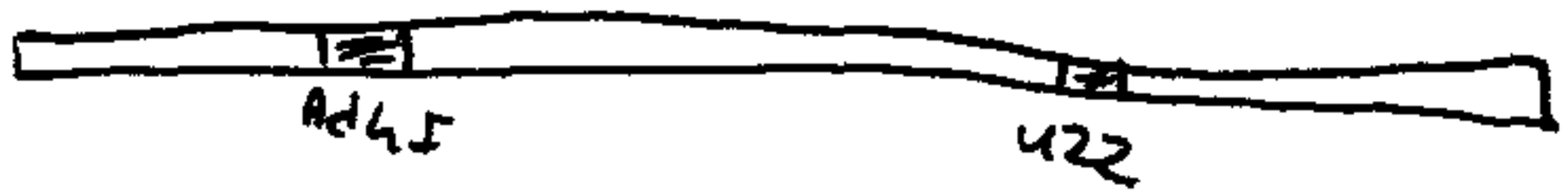
Retros = Class I
DNA-DNA = Class II
Wesslers = Mixtes

} A

Autonomous vs. non-autonomous

Class I Autonomous - code a r.t.
Class II - Autonomous - code a tpase

Maize



Types of TE Induced Variation

1. Coding seq. altered
2. Non-coding altered
3. Mediation of molecular events (recombination)
4. Quantitative variability
5. Genome size
6. Hybrid dysgenesis

1. Majority of TE induced mutations are deleterious
2.
3. Excisions can leave footprints
4. Linkage of mutator + Mutation

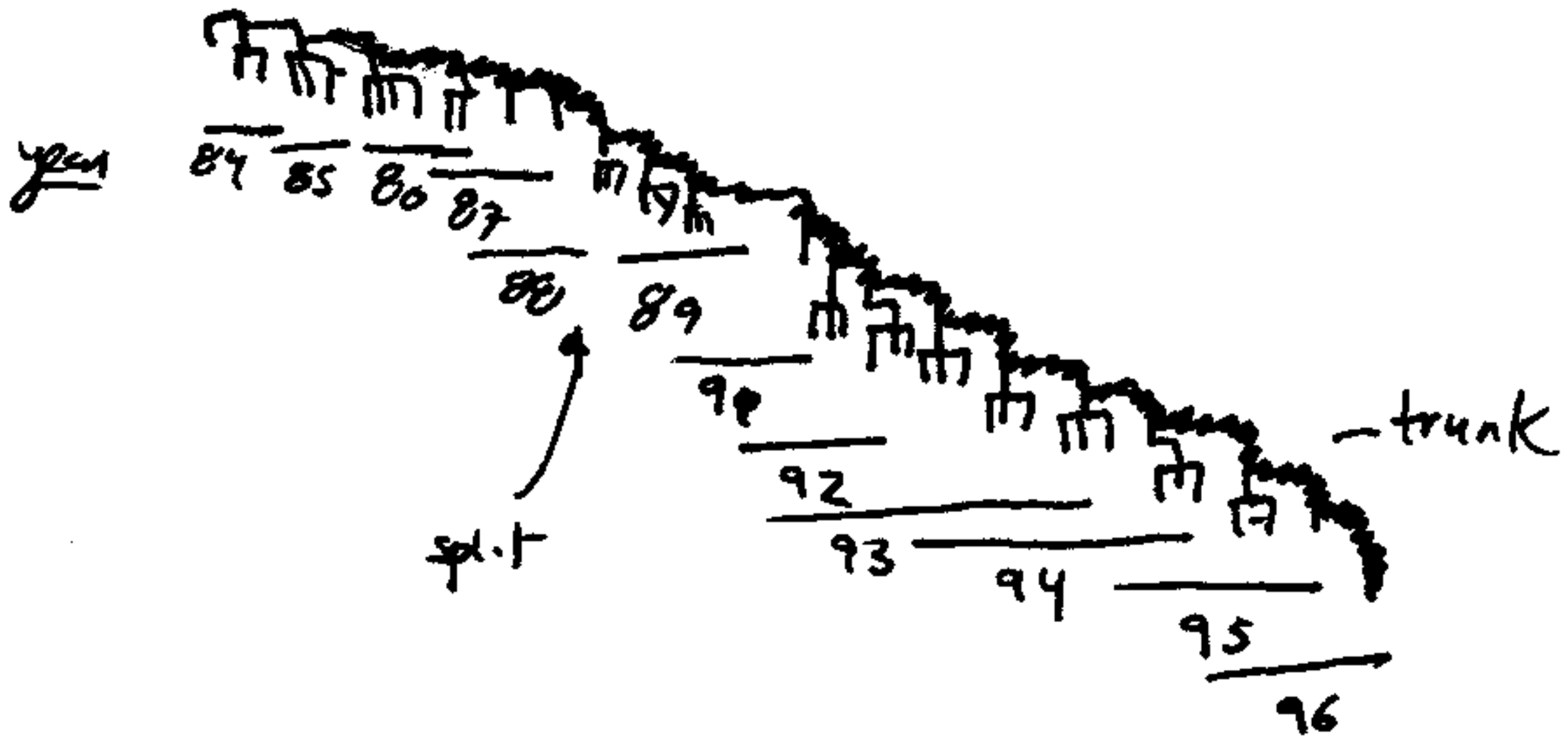
how many TEs are ancient and how many are new

TE "Uses" in hosts

① new tissue specific expression
② regulation
③ new introns
④ telomeres (in Drosoph. la)
⑤ DSB repair
⑥

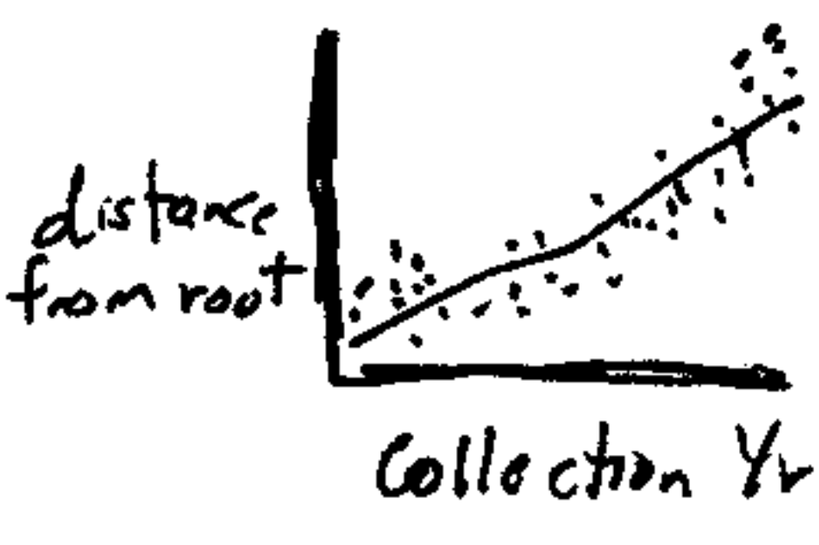
Walter Fitch - Influenzae - New Stuff

- 254 HA1 gene sequences
- isolated 1983-1996
- align w/ no gaps
- know dates of many, growth mechanism of 224

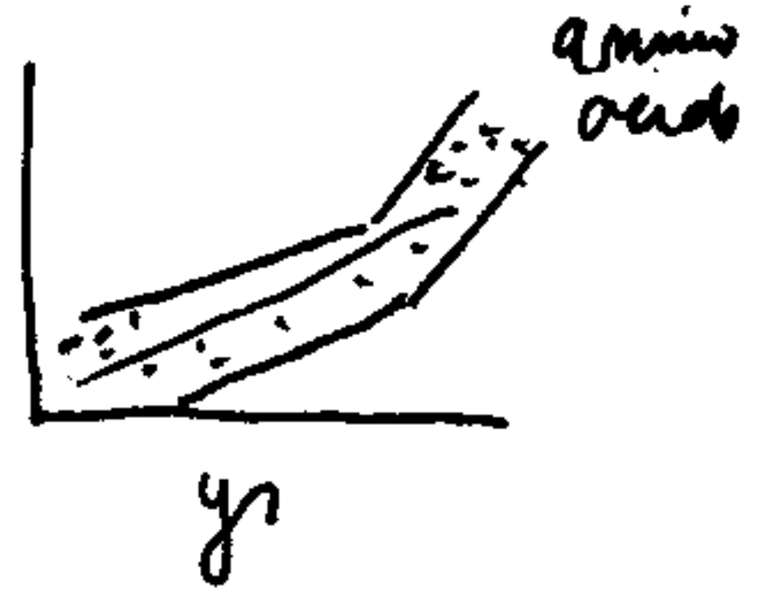


trunk = backbone
twigs = between
tips =

Should take equiv # of total branch length... for tips vs trunk.



nucleotides



- Curves have and encreased slope
- But thinks larger data set may be explain

Replacement substitutions

Codon position	Obs	Expected
1	217	224
2	254	242
3	77	81.4

most proteins... Δ's in pos. 1 usually in excess ... prob. bec. Δ's are less damaging.

	#	observed Δs		expected
Trunk		46	<	61
Twig		119	<	210
Tip		380	>>	274

} Trunks excess in twigs may be due to growth in eggp.

But.. literature says you should do this in kidney cells

Why so many Δ's in tips

① passaging - new mutation
② parsimony bad (can't infer older changes)
③ passaging -- selects
④ selection of strains to use biased (clearly is biased -- only look at different ones)
⑤ unforseen biology

Acctran ... leads to Δ's in deep branches
Deltran ... leads to Δ's in tips

of substitutions

- six codons are hypervariable
- most of codons on trunk are invariant
- most of tips are variable

Silent vs. non-silent

- calculate s. bit
- calculate non-silent } over whole gen

- then look at specific sites ... use overall rates to calculated expected # of changes and then compare to real values

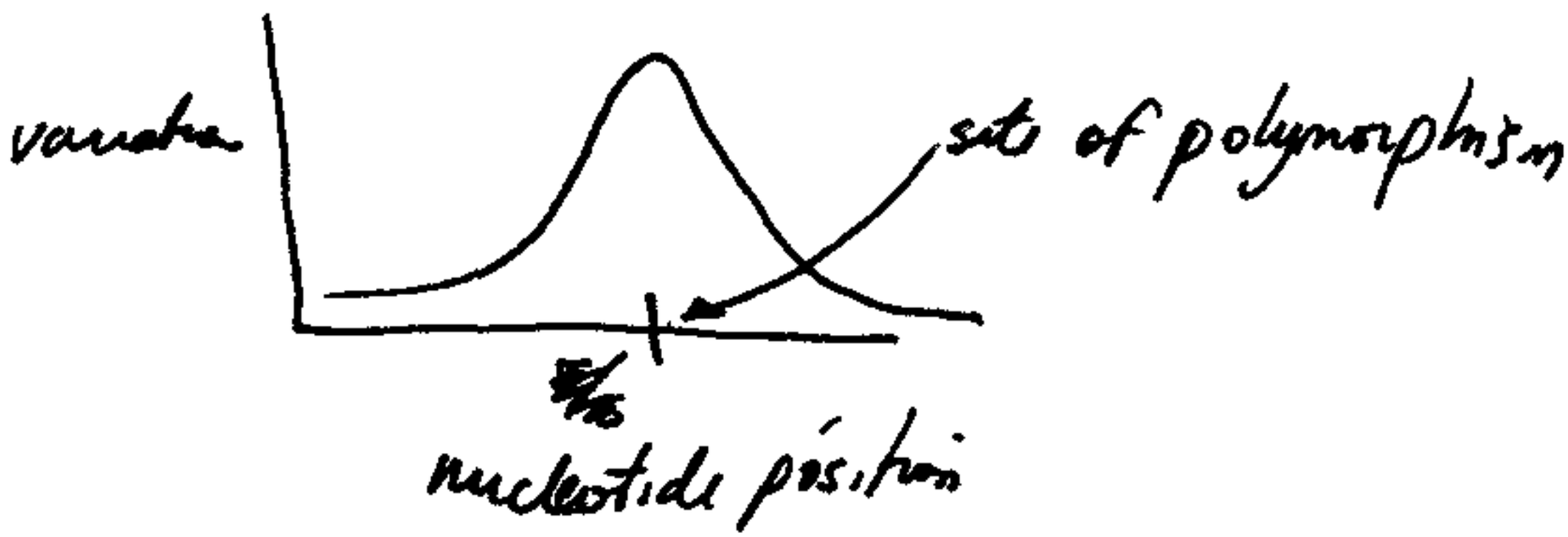
Dick Hudson -

Models to explain nucleotide polymorphisms

- "mostly neutral" model
- or, few does selection a sparsely distributed sites affect polymorphisms at other sites

for balanced polymorphisms -
- selection will maintain alleles
- but linked loci to these sites could continue to diverge indefinitely

e.g. chromosome rearrangements



2.f - Adh

Kreitman + Hudson '91

A selected loci will drive variation down

Candidate Loci:

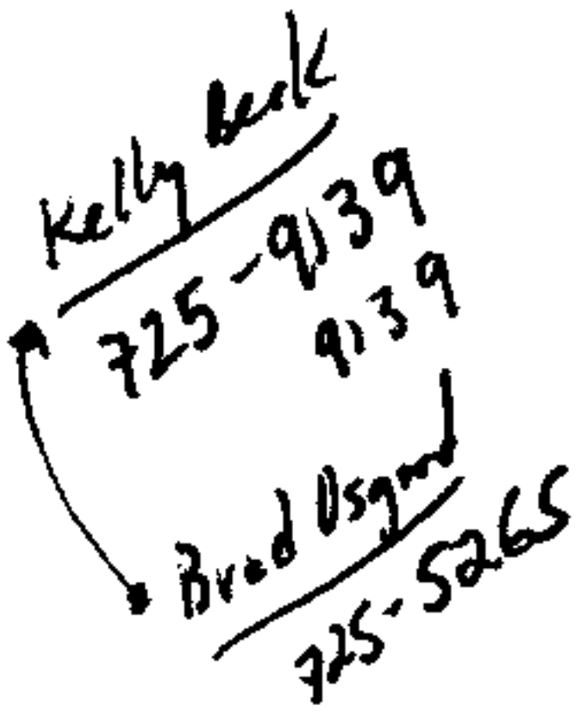
SOD - many known polymorphisms

- Slow vs. fast allele

-all slow = 1 sequence

-many fast alleles

Suggests



728-8222

Andy Clark

Shared polymorphisms

- at highly variable sites... between species comparisons should have high # polymorphic sites even by chance.

- tree 2
  100 18
  120 2
  130

← tested prob. that get 18 shared sites betw. these two species... but chose only one that is extreme

What is prob. that 1 sample ~~of~~ three Samples of normal curves is outside normal?

Wyatt Anderson

Dick Lewontin - Is Pop. Gen an Experimental Science (it is if you are a graduate stude

Intro by John Moor

- When he was invited he told Ayala he wasn't sure NAS still in business... but he went to Lobiany and found their magazine.

Says pop-gen doesn't work by hypothesis. It collects data and then later pretends these are the interesting questions

- Seasonal Δ's in chrom. inversions in D. pseudo + others
- So — how do experiment?
  - put in cages and measure Δ's
  - but had to work to find ♀♀ in cages
    that Δ's occurred ... 18=no, 21=no, 25=labs

Says that finding this T_b is NOT an experiment but is an example of how you can get results LIKE field.

Consistently take glucose on + off

So... how do this in Nature?
- need to measure fitness

3 pieces of literature that had big influence

① chromosomal inversions

② isozyme polymorphisms
- but when done carefully in the lab don't get large, consistent results
- or, in ADH in flies... it works, but may not relate to nature.

③ nucleotide polymorphism
- e.g. Adh in Drosoph. la... many silent Δ's but almost no aa Δ's

But how can all these Δ's be selected against?

Statistics

Dobzhansky said statistics is a way to make bad Data look good.

Custom Usage

John Avise - Species Concept

Biological Species Concept - focus on reproduction isolation

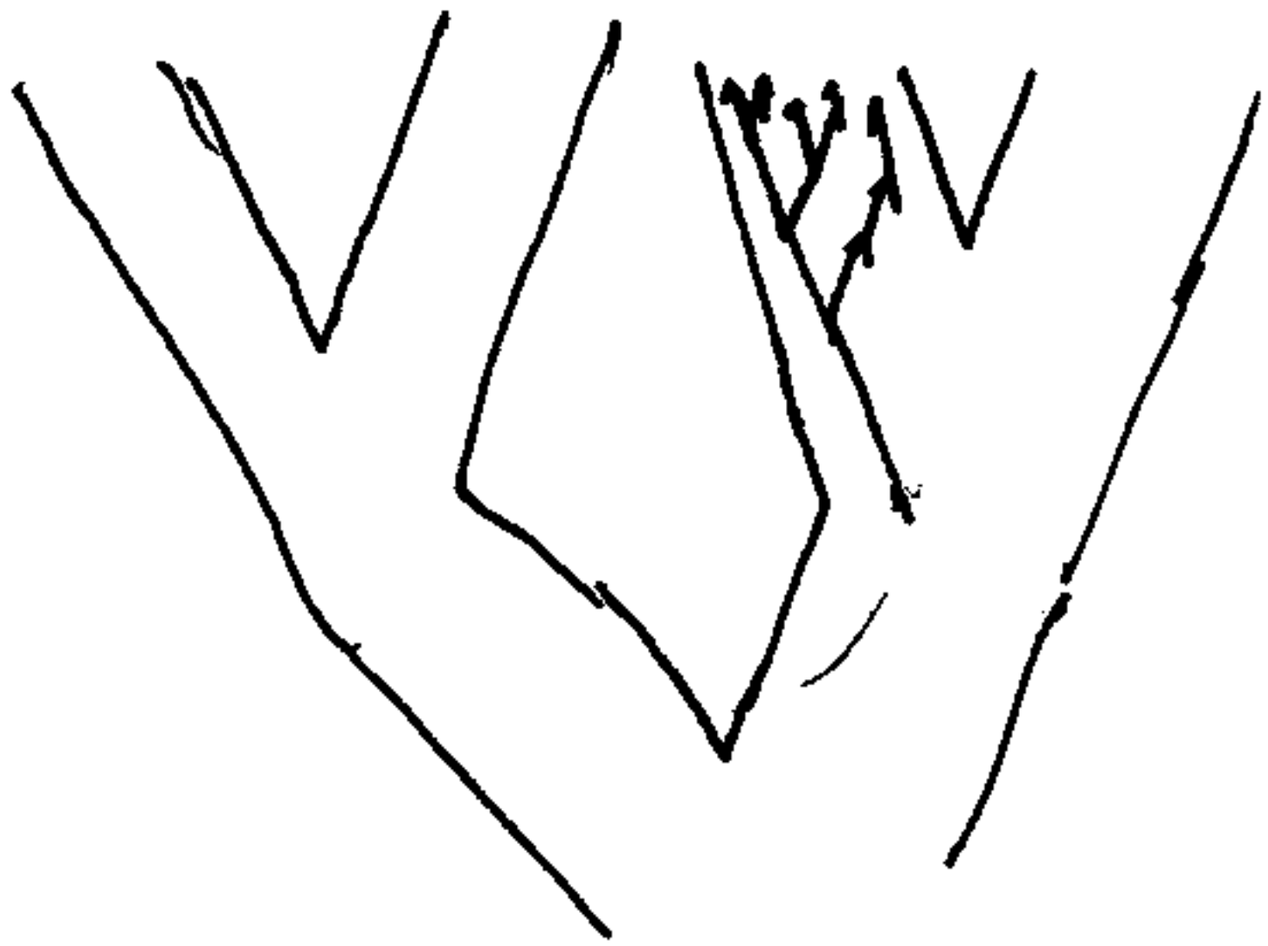
Phylogenetic Species Concept - focus on clades/classification

Questions

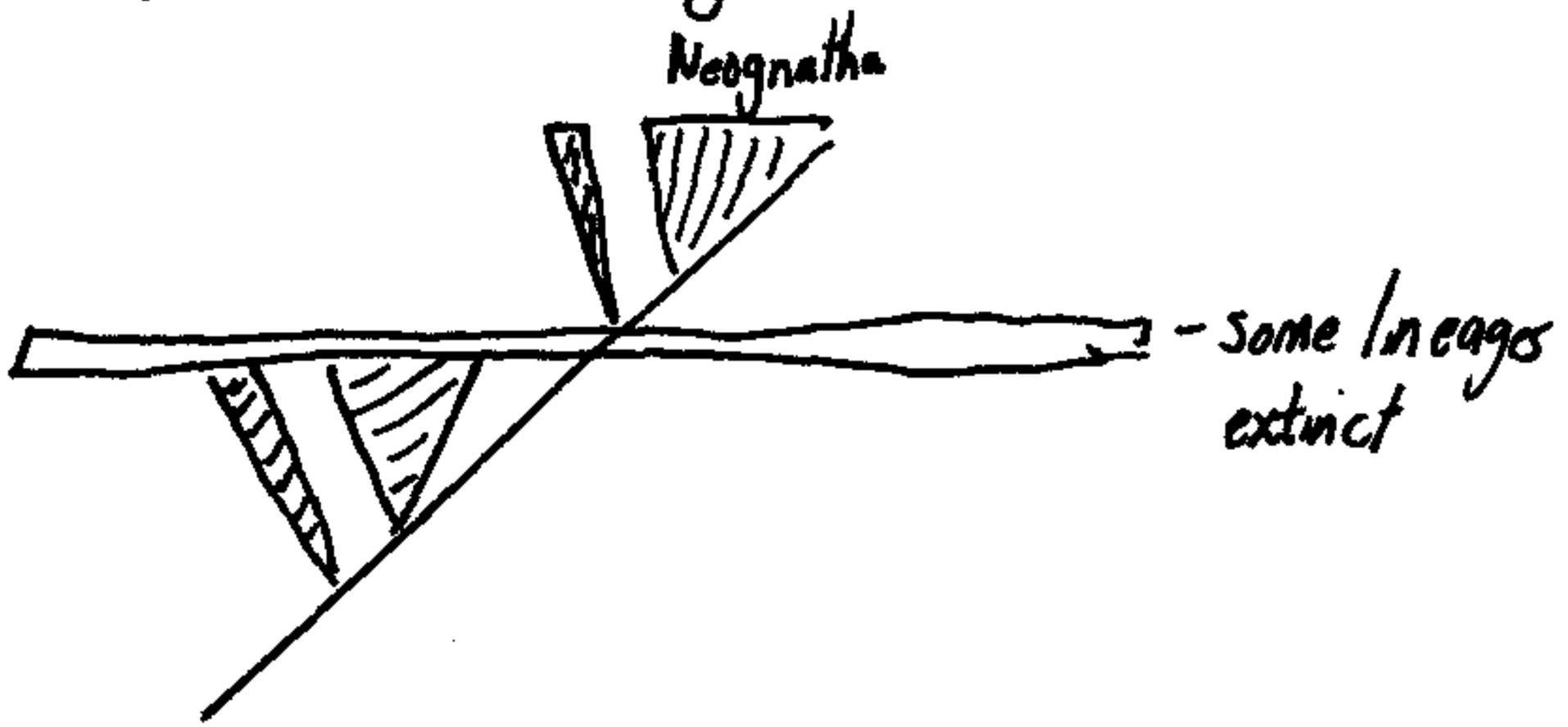
1. Are they really different?
2. Is PSC useful
3. Can they be combined

Allelic genealogies

- allele = length of DNA that has not undergone recombination in the ecological history of taxa under consideration

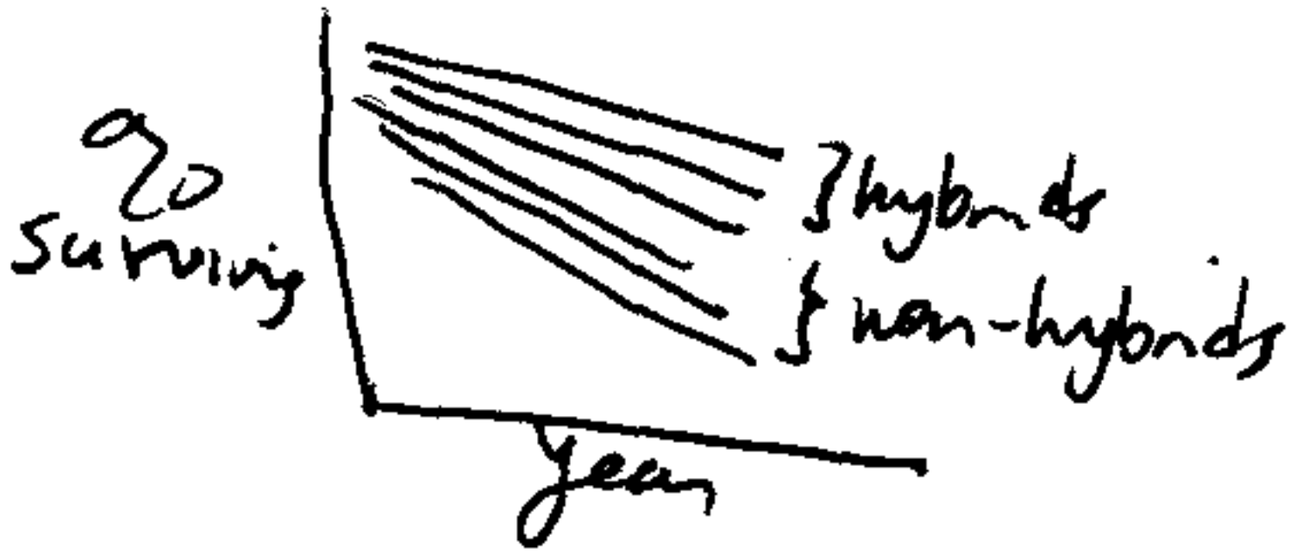


Peter Grant: Genetics + Origin of Bird Species



Dobzhansk + speciation
① takes long time
② begins w/ genetic diff. of allopatric
③ Post-zygotic isolation

Darwin's Finches



Not much stat. sign. but clearly no less survival in hybrids

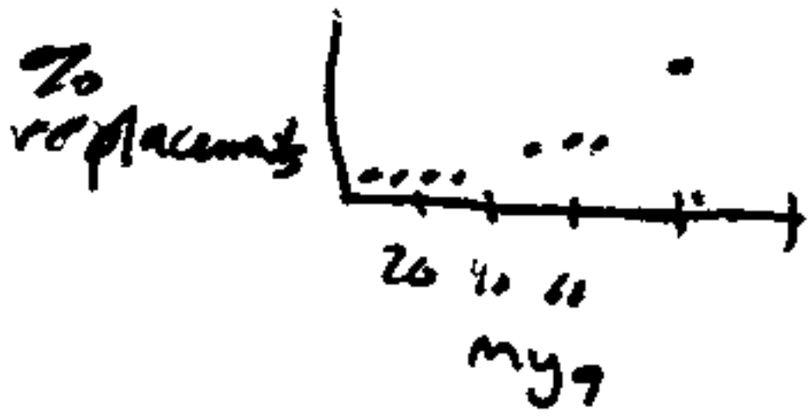
F. Ayala

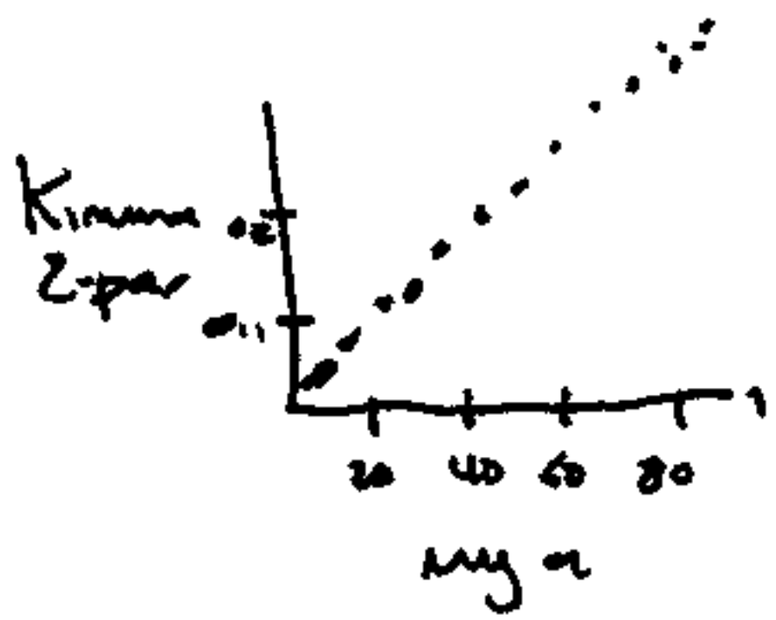
Sequence evolution

GAPDH } substitution rates
SOD

Molecular Clock
- first formalized by Zuckerkandl + Pauling 1965
- 1st real test was Fitch + Margoliash 1967
- Kimura 1968 -
  - if neutral theory is correct, substitutions should be poisson process - ∴ mean ≅ variance
- This is what Ayala is trying to test

GAPDH
- sequenced gene in many species





Rates of aa subst.

1. within Drosophila

Jeff Powell

King + Jukes 1968 "As far as is known, synonymous mutations are truly neutral with respect to natural selection."

Codon Usage Bias

- deviation from random at degenerate sites
- phylogenetic persistance
  - can be specific for certain genes

How measure?

Effective # of codons = highly biased genes have low ENC

When a gene is biased in one species ... it tends to be in other species too.

4 fold sites... correlate codon usage bias
w/ nt in wobble position

Ellie: as genes become more biased... C tends
to be more in wobble position

in bacteria: T is correlated w/ degree of
codon bias

Most as contribute to codon usage bias...
exception is Asp

Causes of Codon Usage Bias

① <u>mutation bias</u>

e.g. isochores in warm blooded mammals
-Genes in A/T isochores use A/T in wobbles
-Genes in C/G isochores use G/C in Wobbles

- but in Drosophila ... doesn't think mutation bias causes codon bias in Drosophila
  - if mutation bias did these -- he thinks it should affect all aa
  - suggests mutation bias in Drosophila is A/T biased

- level of codon usage bias incr. w/ incr. recombination (regions of high recomb. should have more specific selection).

⑤ <u>Selection</u>

- in bacteria - isoaccepting tRNAs are at diff. abundance.

- This would incr. speed + accuracy of translation

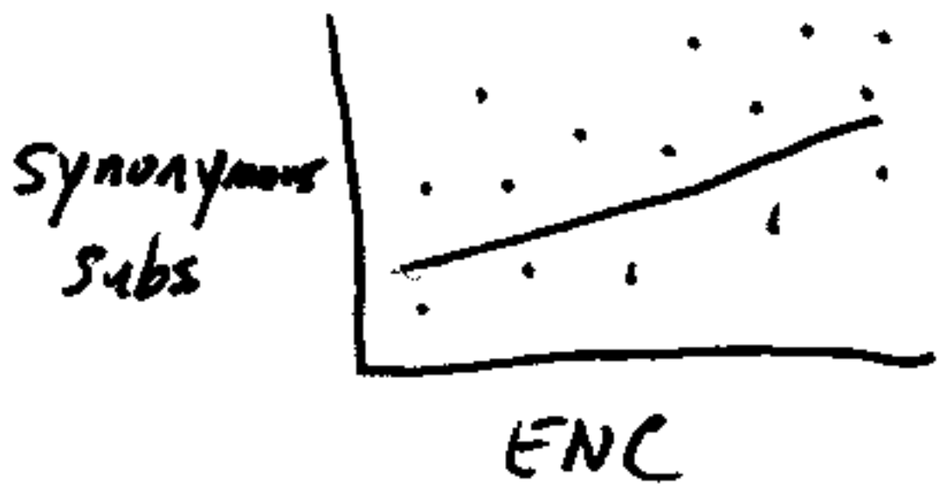
Highly expressed genes tend to be most biased

Shorter genes tend to be the more biased
- suggests that in shorter genes affects on single tRNA "problems" have a larger % effect than in longer gene

Suggests <u>Hiroshi's</u> study showing Accuracy is important could be due to <u>speed</u> affects because slower translation should be more inaccurate.

Effects of Codon Bias

Between species



(melanogaster vs pseudoobs)

within species

} this says higher biased genes have higher rates of syn subst.

Michael Clegg: Rubisco Evolution

5×10^{-9} subs/site/year = syn. rate in plant nucleus

1×10^{-9} " " " = " " " chloroplast

Evolution of Plant Nuclear Genes

rbcS = small subunit

= 4-8 copies per species

M. Nei - Evolution by the birth-and-death processes in multigene families of the vertebrate immune system

1. Concerted evolution
   a. Genes producing large quant. of RNA or protein
   b. MHC + Ig genes?
2. Birth + Death model

MHC loci:
- excessive # of polymorphisms - he argues is due to selection (e.g. ds/dn) but others argue gene conversion important