

912

Land Ultrag # 2934880

QUESTION 1. (10 marks)

Suppose that the sequences shown below are part of the myosin-16 (*MYH16*) gene from various primates (adapted from Stedman, *Nature* 428:415, 2004). Dots represent positions identical to the macaque sequence. Use the UPGMA method to reconstruct a phylogenetic tree for *MYH16*. Give an **interpretation** of your analysis, and discuss any **hypotheses** that are raised. Note that there is a myosin multi-gene family in vertebrates.

frankly => pseudogene..

M	Macaque	TCCACAGCACTGTACCCCATTTTGTCCGCTGTATTGTGCCCATGAGTTTAAGCAGTCAG	
C	ChimpanzeeT.....A.C.....C..C.....T.GA.....A..C.....G..T.	⊙ same
O	OrangutanT.....C.C.....T..C.....T.....T.GA.....A..C.....A..G.	
H	HumanT.....A.C.....C.T.A.....T.....T.GG.....G.....A..C.....G..T.	
G	GorillaT.....C.C.....C.....T.GA.....T.....G..C.....G..G.	
PC	Pygmy chimpT.....A.C.....C..C.....T.GA.....A..C.....G..T.	⊙

	M	C/PC	O	H
C/PC	12	X	X	X
O	14	7	X	X
H	16	6	12	-X
G	12	5	6	10

assuming there are the differences:

shortest distances C/PC = same sequence for this stretch.

⊙ $d(G, C/PC) = 5/2 = 2.5$

⊙ $d(G, C/PC) \circ = \frac{d(G, O) + d(C/PC, O)}{2}$
 $= \frac{6 + 7}{2}$
 $= 6.5$

$d(G, C/PC) \circ = 3.25$

2.5
- positive selection
- (don't) frequency of mutation
- common stage
- random

$d(G, C/PC, O) \circ = \frac{10 + 6 + 12}{3} = 9.33$

$d(G, C/PC, O) \circ = 9.33/2 = 4.67$

⊙ $d(G, C/PC, O, H) \circ = \frac{12 + 12 + 14 + 16}{4} = 13.5$

$d(G, C/PC, O, H) \circ = 13.5/2 = 6.75$

for C/PC
H
G
O
=> you forget to compare the 2 bases

Interpretation:

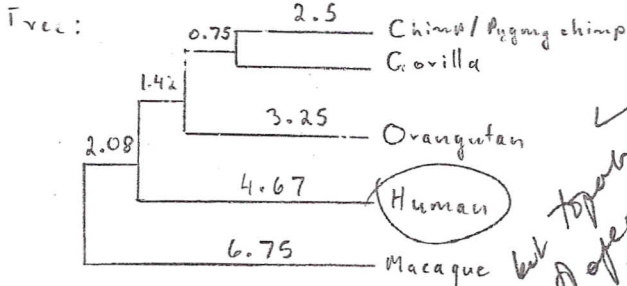
- C + PC have identical sequence for this stretch, therefore they are related.

- Gorilla is closely related to C + PC + the branch length infers number of changes which have occurred

H C - The other species follow with relatedness, gets lower as we move down tree.

F G - The sequence or stretch (specific) which was compared may not be representative of actual relationships.

- The nucleotide sequence indicated on top shows are relatively high % identity between all of the species, but



but topology of species tree?

tree is not to scale

hypotheses

Macaque is more distantly related to these other species due to functional restraints.

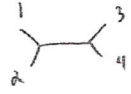
due to presence of myosin multi gene

QUESTION 2 (10 marks total)

a) (6 marks) Suppose that the protein X coding sequences below are from four different copies of gene X in tobacco. Use the maximum parsimony method to infer their potential relationships. In your answer, include your *rationale* and also a *ranking* of sites with respect to their informational content.

if one differs still up shot branch by the

- 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
- ① Sequence Xa ... TTG ATA ACT AAG TAC ACT ACC ATA ACC ACC TTA ACC CAA ...
 - ② Sequence Xb ... TTG ATT ACA AAG TAT GCG ACA ATA GCT ATT CTA ACT CAT ...
 - ③ Sequence Xc ... TTG ACA GCG AAG TAT ACC ACC ATA GCT ACC CTC ACC CAT ...
 - ④ Sequence Xd ... CTG ATA GCC AAG CAC ACT ACA ATA ACC ACC CTA ACT TAA ...

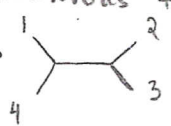


site	tree 1	tree 2	tree 3
7			
15, 25, 27, 39 Call have same pattern: example for site 15 shown			
21, 36 [same pattern as site 27 shown]			
Total	13	12	<u>10</u>

Sites # 7, 15, 21, 25, 27, 36 & 39 are all information sites because they have support in 2 of the 4 being same & 2 of them being different.

all same gives rise of ancestral strain

- Tree # 3 is the most parsimonious tree as it requires the fewest number of changes in its relationships.



b) (4 marks) For ONE of the following, explain how it might influence maximum parsimony analysis:

- i) Using amino acid sequences instead of nucleotide sequences
- ii) Homoplasy

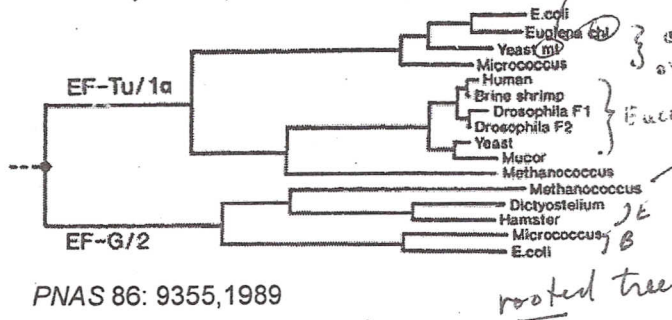
(AA)
Amino acid sequences: Because nucleotide sequences can undergo synonymous & nonsynonymous substitutions, the results might be skewed. The degeneracy of the genetic code table could influence actual differences which will not be picked up using AA, but will be picked up using nucleotides. The correlation between relationships of homologous sequences / organisms could be compromised & therefore, corrections would have to be made to account for these differences. An organism may appear to be closely related to another organism @ the amino acid level as opposed at the nucleotide level.

35

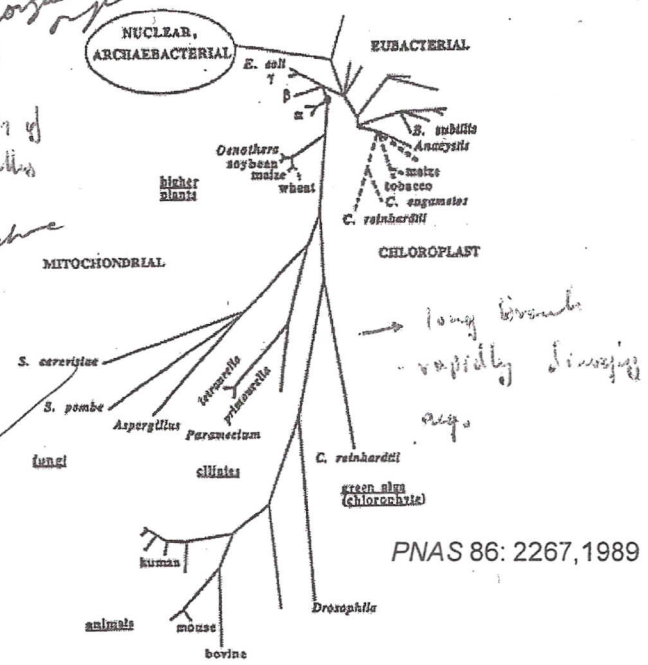
QUESTION 3. (10 marks)

For **ONE** of the two figures below, explain (i) what the figure shows, (ii) how it illustrates certain fundamental evolutionary concepts and (iii) how/why molecular approaches have been useful in addressing this issue.

A rooting of universal tree.



B



using a cladogram
The figure shows an analysis of various organisms & their relationships to each other based on elongation factor sequences. The reason this

possible is because the relative evolution of that sequence is conserved and all organisms have those factors. → duplication in eukaryotes so one copy could be used to root the universal tree

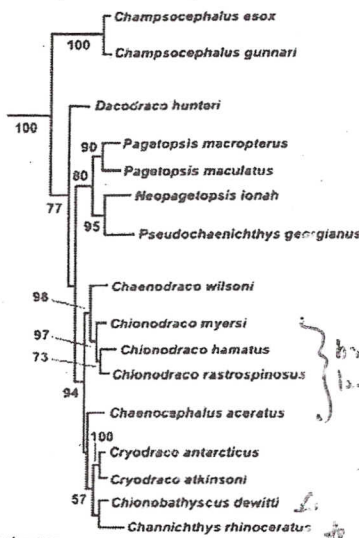
- It shows that the sequences of EF-Tu/1A is more closely related among these organisms as opposed to the sequences of EF-G/2. The comparison also shows a closer relationship between archaea + eucarya as opposed to bacteria.

(b) Molecular data is better because it is quantitative, ~~more~~ more abundant rather than morphological characteristics. Molecular data evolves at a generally more constant rate opposed to morphological & physiological characteristics. By the same regard, using molecular data makes it easier to confer homology or determine homology. Another point is that using molecular data allows us to draw relationships between distantly related organisms.

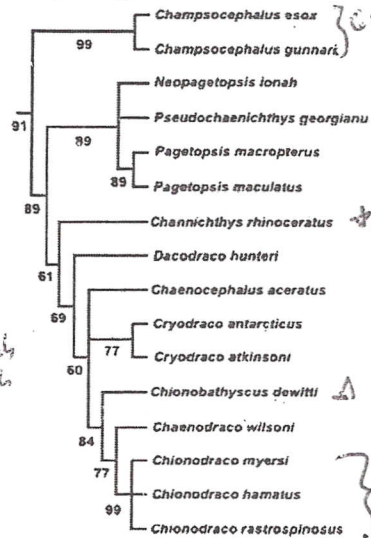
QUESTION 4. (10 marks)

This figure is from a paper (Near et al. *Mol. Phylogenet. Evol.* 28:87, 2003) dealing with phylogenetic relationships of Antarctic icefishes. Give an **interpretation** of these data and propose additional **molecular evolutionary studies** that could be done to further address these issues of icefish phylogeny. Note: mtDNA = mitochondrial DNA; ND2 = NADH dehydrogenase subunit 2; 16S = LSU rRNA.

mtDNA (ND2 and 16S)



Morphology



↑ of changes
are been
shown for
mtDNA @
molecular
evol.

Additional studies:

- sequencing of mtDNA to ensure % identity & homology
- protein analysis & expression levels of structural proteins & RNA such as tRNA, rRNA.
- DNA analysis using RFLP, AFLP & DNA-DNA hybridization.

we know all these fish share a common ancestor.

flipped clades are not differences.

Interpretation

- Dendrogram on left uses mitochondrial DNA, which is a molecular analysis tool. This allows for easier assessment of homology, indicated by relatively good bootstrap values. The confidence levels in both are quite high.
- Dendrogram on right uses morphology as the trait. For morphology, there are branches which have been collapsed, those branches have not been collapsed in mtDNA. [could be homoplasy / convergent evolution]
- The relatively high substitution of nucleotides in mitochondria is a useful tool to determine relationship as homology is closely related organisms, therefore it can be concluded that these organisms are closely related.
- It should be noted that the topology in both trees is relatively the same but bootstrap values for molecular data have higher confidences [in most instances]
- Changes in mitochondrial information could also point to functional constraints as in oxygen levels, blood content (globin genes etc.), characteristics that are not measurable by morphology alone. relative rate test → maybe (image) evolve faster than other.
- The order of the species & their relationships to each other are different in mtDNA & morphology data.
- In both cases, there is no particular outgroup & in both cases a common ancestor seems likely.