Instructions for the BIO 332 final assignment.

Files are available on the student portal. The PC-lab (where we did the practicals) is available for your work. Unfortunately, there is a shortage of extra keys, but you may usually find people who are willing to let you in. One key may be borrowed (overnight?) from former Department of Botany. Contact:

Oddfrid T.K. Førland
Studiekonsulent / Study counsellor
Institutt for biologi / Department of biology
P.O. Box 7800, N-5020 Bergen
Besøksadresse / Visiting address: Allégate 41
Tlf. +47 55 58 22 24. Fax. +47 55 58 96 67

In case of technical problems, contact Arild Breistøl (zoology 1st floor) (tlf. 82233) or E.Willassen (tlf. 82901).

When to submit:
Final deadline on December 1 (2004).

How to submit:
Submit by Email to Oddfrid.Forland@bio.uib.no

She will collect your files in electronic folders marked with your candidate number so that material is anonymous when evaluated and marked.

What to submit:
1) Your name and candidate number
2) PAUP* script files and files produced by the scripts.
3) A report with text, tables, and figures (trees) presenting the methods, results, your interpretations, conclusions, and possible references to literature. The final part of the report must contain a table with the following format:

Supplementary files

File number	contents	File name
1	Paup script	Myfile_1.nex
2	Results model testing	modeltest.txt

Please refer to the file numbers in the text if you want to document details of your work, for instance, ' – the GTR+G model was selected based on results (2) from the hierarchical testing with Modeltest – '

UoB Nov.16. 2004. E.Willassen

BIO332 Final assignment.

November 2004 (E.Willassen)

A key issue in public interest about evolutionary questions is the relationship between humans and the great apes. (Some background information may be found in the file *human_evol.pdf*, Willassen 2004: lecture notes from BIO210). For this assignment, you will examine some of the evidence for the relationships of the hominids by analyzing two data sets of mitochondrial sequences.

Data set 1 (Hayasaka et al. 1988) includes two mitochondrial protein coding genes and three tRNAs from twelve primates. The range of each gene is defined in the available nexus data file (***Data1.nex***).

Data set 2 (***Data2.nex***) is an alignment of additional mitochondrial tRNAs from some of the same taxa.

Use the text book and additional material with your previous knowledge from the course. Check the PAUP* and MrBayes manuals to find additional information on commands.

Analyze Data set 1

Write a *PAUP script** that does the following when executed:

1) Logs the run
2) computes uncorrected distance (p-distance) for each gene so that you may subsequently compute average p-distance in each gene by importing the results to Excel.
3) computes total pair wise (uncorrected) differences between the sequences and also computes the numbers of transitions and transversions in 1st, 2nd, and 3rd positions of the protein coding genes.
4) computes "empirical" base frequencies for the alignment.
5) computes MP tree(s) and saves the MP tree(s) to a file.
6) reconstructs character state changes on the branch representing the most recent ancestor of Homo_sapiens and Pan. [hint: describetrees /apolist=yes].
7) Computes and saves a strict consensus tree to indicate unresolved nodes.
8) computes bootstrap support and saves the bootstrap consensus tree to a file.

Presentation:

Present the results from 2) and 4) in tables. 3) Plot transitions and transversions (y-axis) in each codon position versus absolute pair wise differences (x-axis). Present the numbers of substitution types (C<->T, etc) in the most recent common ancestor (MRCA) of Homo and Pan in a table. Present your strict consensus tree in your report [hint: export graphic from Treeview or Mesquite?]

Interpretation:

Is higher sequence divergence indicated in particular parts of the alignment?
Is saturation indicated to a larger degree in any of the codon positions, and if so, how would you explain that?
Is there a bias in nucleotide frequencies?
Do the results from 6) indicate equal rates of nucleotide substitution types?
Does this exploration of the data suggest what sort of properties we would require of a model to be used with this data set for phylogeny reconstruction with distance or ML methods?

Prepare, execute, and log scripts for ML analysis of Data set 1

9) use the script *ModelblockPAUPb10.nex* and hierarchical log-likelihood ratio testing with *Modeltest* to find an evolutionary model that describes all data best under the ML criterion.
10) Use the model without the parameter estimates from the model testing [hint: see how your model is phrased in Modelblock], and compute the

likelihood scores for your MP trees. [Hint: by adding the option khtest=normal to your commands, the Kishino-Hasegawa test will tell you whether one tree is significantly better than alternative trees].

11) Compute the maximum likelihood tree with the parameter estimates suggested by model test. Save the tree.

Analyze Data set 1 with MrBayes

12) Find a suitable model for the data by either adapting the most similar model to the one used with ML above [hint: see file 'ML_models…' by J.Nylander], or by running *MrModeltest* .
13) Apply the model for Bayesian estimation of phylogeny.
14) Use *Tracer* to decide when likelihood estimates (and preferably other parameters) are in equilibrium, and effective sample size (ESS) is sufficient. (Make sure that you have a large number of trees for the computation of posterior probabilities on branches.)
15) Present your tree with branch support in your report.

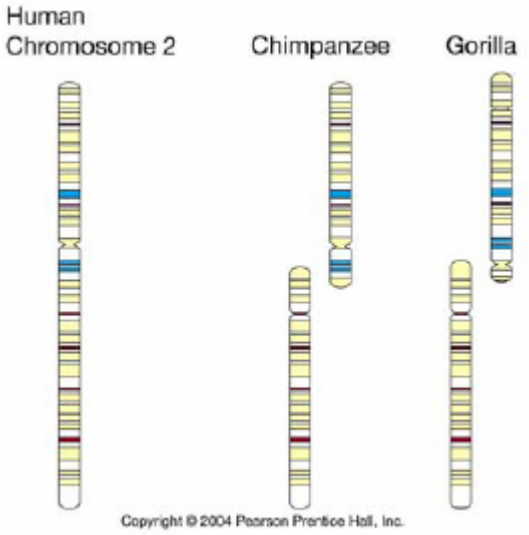
Analyze Data set 2

16) Use you knowledge and skills to decide whether a phylogeny reconstructed from Data set 2 is congruent with results obtained with Data set 1.

Sum up your analyzes with respect to the question of monophyly human-chimps.

How would you explain this striking similarity: that the homologous sites of the human chromosome 2 are found on two separate chromosomes in chimps and gorillas?

What sort of information would you need in order to decide whether these characteristics are actually in conflict with the hypothesis of humans and chimps as sister groups?



Dating the diversification of hominids

17) Use Dataset 1 and load your previously achieved unrooted ML tree to memory
18) Estimate likelihood score, parameters, and branch lengths for the tree under the chosen model.
19) print the tree with branch lengths to the log. [hint: describetree `/ plot=phbrlens=yes].` Do the terminal branches indicate an ultrametric tree? [hint: see textbook]
20) Root the tree with *Tarsius* as outgroup.
21) Estimate likelihood score, parameters, and branch lengths for the tree under a molecular clock constraint [hint: clock=yes].

22) print the tree with branch lengths to the log.
23) Use the log likelihood ratio test to decide whether evolution in the unconstrained ML tree significantly deviates from a molecular clock. **Note: The degrees of freedom for the test are N-2 (Not N-1 !)** (N=#taxa). (The ultrametric tree is the null hypothesis.)

A 'standard molecular clock' for animal mitochondrial DNA is 2% nucleotide divergence per million years, i.e. a substitution rate of 0.01 nucleotides per nucleotide per million years. Why is the divergence rate two times the evolutionary rate?

24) If the ultrametric tree can be used to model the evolution of the primates, use the evolutionary rate and branch lengths to date the nodes in the tree.

In 2002, about 6-7 million year fossil remains of a hominid species called *Sahelanthropus tchadensis* were discovered in Tchad (Nature 418, 145–151). It has been suggested that *S. tchadensis* represents the MRCA of chimps and humans. Other researchers claim that these fossils have characteristics that are more gorilla-like. How does your molecular dating contribute to this discussion?

Reference

Hayasaka, K., T. Gojobori, and S. Horai. 1988. Molecular phylogeny and evolution of primate mitochondrial DNA. Mol. Biol. Evol., 5:626-644)

Extra files

Data1.nex
Data2.nex
ModelblockPAUPb10.nex
(There is a bugged version of Modelblock out there. Make sure to use this file)
Human_evol.pdf
Paup.pdf
Mrbayes3.pdf