**Computational Biology HS 2021**

Rough Exam questions (from memory) Aswin

Question 1: True or False type questions (very easy)

* A tree was given and various mutations were marked on its branches. Questions asked were like which mutations are in which species, which ones formed a clade, etc.
* A Smith Waterman score matrix was given and T/F questions about its alignment.

Question 2: COVID 19 in Zurich is modelled using a tree

* What do ‘birth’ events mean w.r.t this tree?
* If an infected individual recovers from Covid within 10 days, what is the value of death rate in days-1?
* What is reproductive number? How is reproductive number written in terms of birth and death rates? Explain the logic.
* Given a covid transmission tree. Why might we not be able to sample all birth events in the tree?
* If you analysed covid cases by jointly analysing sequences from Zurich and Basel, how would that alter your estimates of birth and death rates? How would you alternatively model to get correct estimates?

Question 3: For all the tree generation methods UPGMA, Parsimony, Maximum Likelihood and Bayesian Inference, explain where exactly do we use the aligned sequence data in each of these methods. Also give one advantage and one disadvantage of each method

Question 4: Phylogenetic Tree and Bayesian Inference

A figure of a tree like below is given. Nodes and branch lengths are labelled.



* If P(Xi | Xj , t) is the probability of nucleotide substitution along a branch, write the formula for probability of observing the nucleotides at root 7.
* Given the equilibrium distribution of nucleotides pi, write the formula for likelihood of nucleotides at the root 7.
* Write the formula for the total likelihood of the whole tree.
* Say you want to perform Bayesian inference using the above information. Write the formula for phylogenetic posterior. Write one sentence about each term.
* Which term can you omit when doing MCMC simulations and why?

Question 5: Coalescent Theory

* Figure of a coalescent process is given with N haploids. Derive the formula that 2 of these coalesce in the 4th generation.
* Using the coalescent process, a tree (figure given) is built. Tree has 3 leaves. What is the probability density of the tree?
* Say you estimated the tree based on the coalescent with some certainty. When doing the phylodynamic analysis to estimate Ng, your estimate can be wrong even though tree is inferred correctly. Reasons?