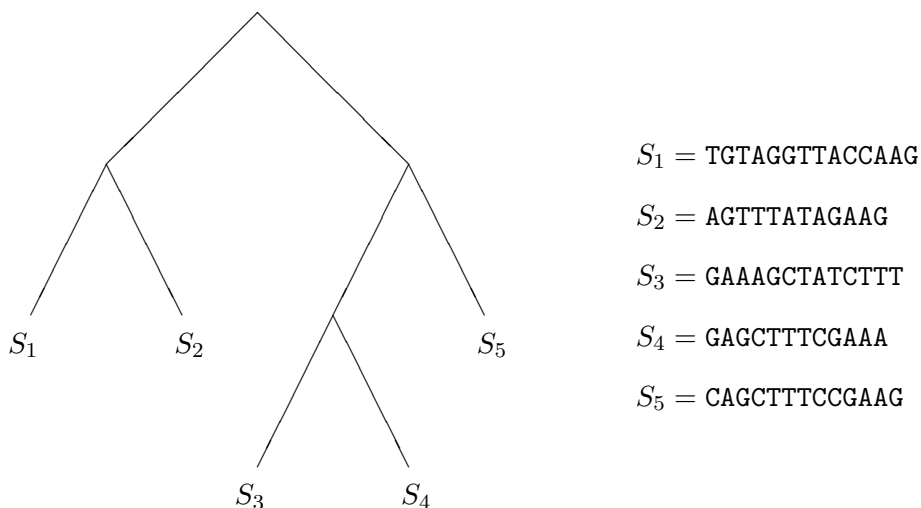


Algorithms in Genome Research
Winter 2021/2022

Exercises

Number 4, Discussion: 2021 December 10

1. What is the mRNA Codon used in the incorporation of Selenocysteine? Does this codon also code for something else? How is this possible?
2. Which types of information may be used by a prokaryotic gene finder?
What are the algorithmic/computational techniques that are employed?
3. What are the main issues why eukaryotic gene finding is more difficult than prokaryotic gene finding?
What are the algorithmic/computational techniques that are used in order to account for this increased difficulty?
4. Discuss the issue of obtaining test data sets for HMM training in gene finding.
5. What kind of signals does one typically expect in non-protein coding DNA regions?
Are there differences between prokaryotes and eukaryotes?
6. What is the application scenario of phylogenetic footprinting?
7. Find all phylogenetic footprints of length $k = 4$ with a parsimony score of at most $d = 2$ in the following tree T :



(Hint: You may implement the algorithm or try it manually. The latter should not take longer than 30 minutes.)