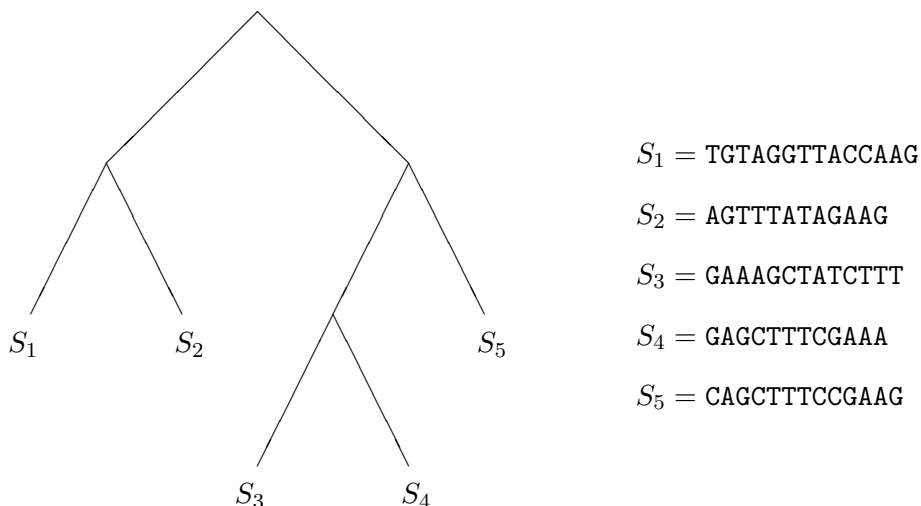


Algorithms in Genome Research
Winter 2012/2013

Exercises

Number 10, Discussion: 11. January 2013

1. What kind of signals does one typically expect in non-protein coding DNA regions?
Are there differences between prokaryotes and eukaryotes?
2. Remember the details of the algorithm from the Sequence Analysis class that allows to find maximal repeats using a suffix tree. (If you did not attend that class, you can find the details of the algorithm in Section 7.12.1 of Dan Gusfield's book *Algorithms on Strings, Trees, and Sequences*, Cambridge University Press 1997.)
 - (a) Explain the main idea to efficiently test for left-maximality.
 - (b) Try and discuss how a suffix tree could be used to find *tandem repeats*, i.e. repeats where both copies follow each other immediately.
3. Read the paper of Blanchette et al. *Algorithms for Phylogenetic Footprinting*. Explain how the running time of the algorithm can be reduced from $O(n \cdot k \cdot (4^{2k} + l))$ to $O(n \cdot k \cdot (4^k + l))$.
4. What is the application scenario of phylogenetic footprinting?
5. 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.)