Bielefeld University
Faculty of Technology

Dr. Daniel Dörr

Algorithms for Genome Rearrangement
Summer 2017

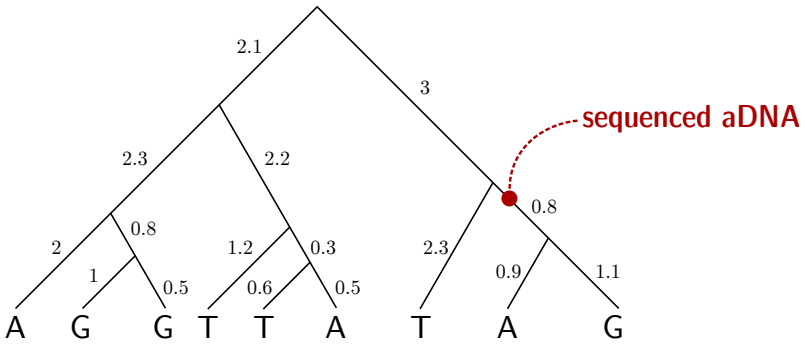
Exercises

Exercise 12, 14.06.2017

As a last step for reconstructing ancestral genomes with ancient DNA and genome data from extant genomes, sequence gaps are closed between inferred ancestral markers. For this step, the algorithm AGapES (Luhmann *et al.*, 2017) computes an alignment between the "gap" sequences conserved in extant genomes. It then uses this alignment to infer a *template sequence*. For each site of the alignment, one can use the Sankoff-Rousseau algorithm to infer the nucleotide character at the corresponding position in the template. Thus, rather than inferring adjacencies as shown in the lecture, this time nucleotide bases are inferred. Consequently, instead of having two states that indicate presence or absence of adjacencies, one has four states A, C, G, and T. For the transition costs between states we will make use of a common model of sequence evolution. This model differentiates between substitution rates within (so-called *transitions*) and between (so-called *transversions*) of the groups of purines (A, G) and pyrimidines (C, T). The following transition cost matrix for A, C, G, T states is given, where $l(u, v)$ is the branch length between the two vertices of the phylogeny along which the transition will occur: (6 P)

$d(s,t)$	$\boldsymbol{A}$	$\boldsymbol{G}$	$\boldsymbol{C}$	$\boldsymbol{T}$
$\boldsymbol{A}$	0	$0.8/l(u,v)$	$1/l(u,v)$	$1/l(u,v)$
$\boldsymbol{G}$	$0.8/l(u,v)$	0	$1/l(u,v)$	$1/l(u,v)$
$\boldsymbol{C}$	$1/l(u,v)$	$1/l(u,v)$	0	$0.8/l(u,v)$
$\boldsymbol{T}$	$1/l(u,v)$	$1/l(u,v)$	$0.8/l(u,v)$	0

Compute the state assignment of all internal vertices for the following phylogeny under the assumption that we aim to infer a template for an aDNA sequence at the indicated position.



Hand in solutions before the tutorial on 21.07.2017