# CSCI 451 Homework 4

Samuel Shissler, Brendan Tracey, and John Trapp

October 31, 2017

## Exercise 19

A

*Remark.* The suffix tree $T$ must contain $\alpha$. If $T$ can contain $\beta$ but not also contain $\alpha$ then the statement is disproved.

*Proof.* If $\alpha =$ "$abcd$" and $\beta =$ "$abc$", then in order for $\beta$ to exist in the tree (assuming it does not exists in some other, removed part), $\alpha$ will have to be segmented and thus no longer in the tree. □

B *Proof.* Because the suffix tree works backwards (suffixes), there should always be an internal node with path $\alpha$ because $\alpha$ is the suffix of $xy\alpha$. □

## Exercise 20

| i | SA[i] | Suffix | SA[i-1] |
|---|---|---|---|
| 1 | 8 | $ | a |
| 2 | 1 | acgtcga$ | $ |
| 3 | 7 | a$ | g |
| 4 | 2 | cgtcga$ | a |
| 5 | 5 | cga$ | t |
| 6 | 3 | gtcga$ | c |
| 7 | 6 | ga$ | c |
| 8 | 4 | tcga$ | g |

The BW string is a$gatccg.

## Exercise 6.9

BAliBASE Alignment of BB11001

$1aab\_ \quad ---GKGDPKKPRGKMSSYAFFVQTSREEHKKKHPDASVNFSEFSKKCSERWKT$
$1j46\_A \; -------MQDRVKRPMNAFIVWSRDQRRKMALENP--RMRNSEISKQLGYQW$
$1k99\_A \; MKKLKKHPDFPKKPLTPYFRFFMEKRAKYAKLHP--EMSNLDLTKILSKKYK$
$2lef\_A \; ---------MHIKKPLNAFMLYMKEMRANVVAEST--LKESAAINQILGRRW$

MUSCLE ClustalW Alignment

$2lef\_A$ $---------MHIKKPLNAFMLYMKEMRANVVAES-TLK-ESAAINQILGRRW$
$1j46\_A$ $-------MQDRVKRPMNAFIVWSRDQRRKMALENPRMR-NS-EISKQLGYQW$
$1k99\_A$ $MKKLKKHPDFPKKPLTPYFRFFMEKRAKYAKLHPEMS-NL-DLTKILSKKYK$
$1aab\_$ $---GKGDPKKPRGKMSSYAFFVQTSREEHKKKHPDASVNFSEFSKKCSERWKT$

As you can see[1], the results look similar, but they are not quite the same. For example, sequence 2lef_A has "-TLK-" in the MUSCLE and "T–LK" in the BAliSASE.

[1]The authors would like to apologize for the entire sequences not fitting. We can't find a good way to display it in LaTeX...