

Sara Kemmler 5760949
Robin Bonkaß 5769588

1	2	3	4	Σ

Übungsblatt Nr. 05

(Abgabetermin 02.06.22)

Aufgabe 1

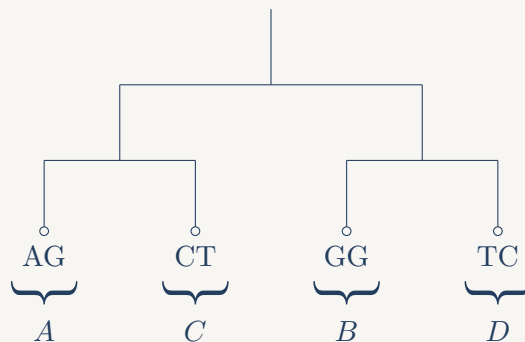
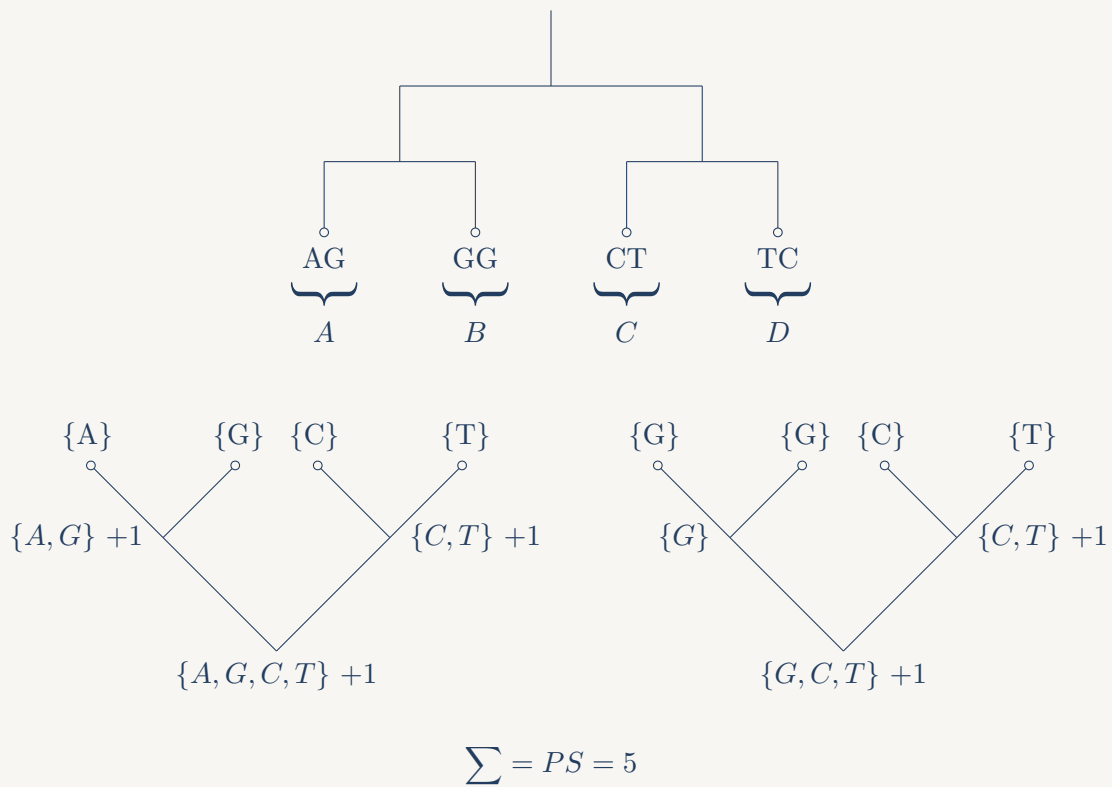
Seien die Sequenzen A, B, C, D gegeben durch:

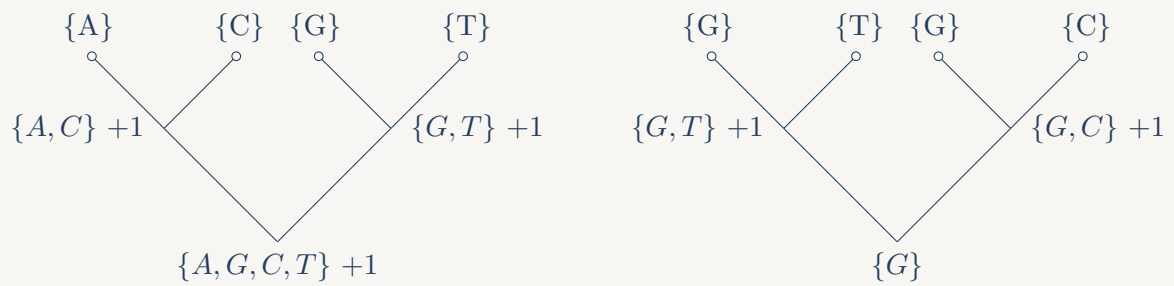
$A : AG$

$B : GG$

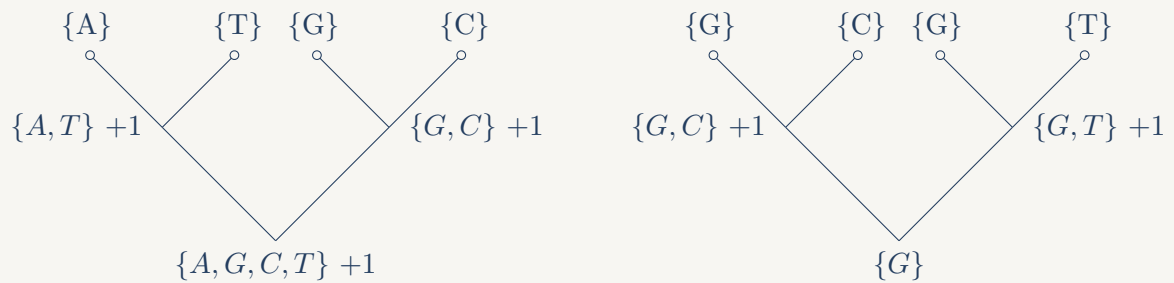
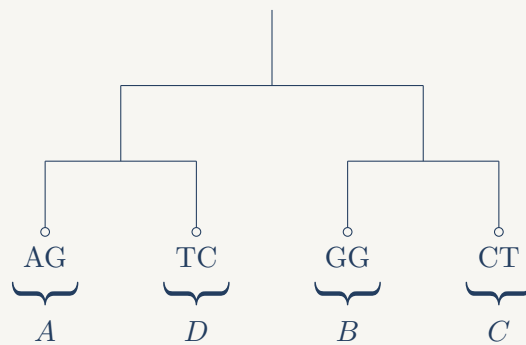
$C : CT$

$D : TC$





$$\sum = PS = 5$$



$$\sum = PS = 5$$

Somit haben alle drei Bäume für die definierten Sequenzen einen respective parsimony score von 5

Aufgabe 2

Seien die Sequenzen gegeben:

a_1 : TTC

a_2 : CGC

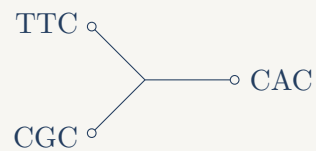
a_3 : CAC

a_4 : TCC

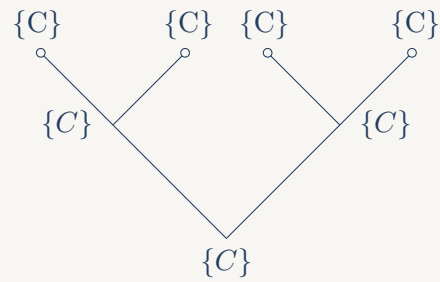
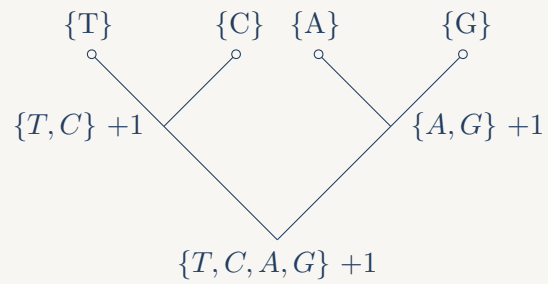
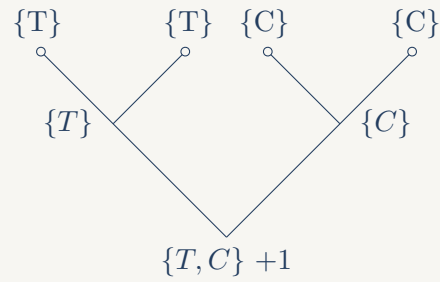
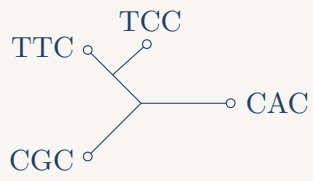
a_5 : GTC

Gesucht ist der maximum parsimony tree.

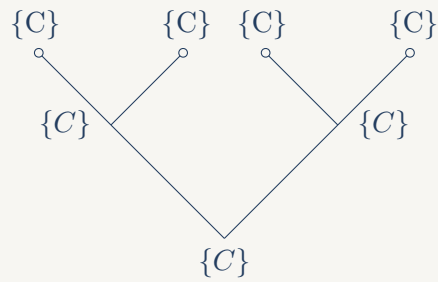
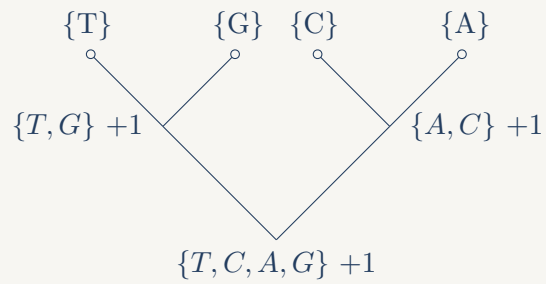
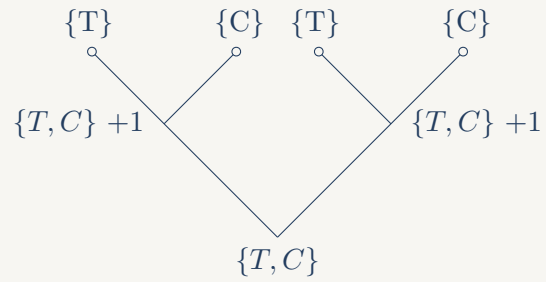
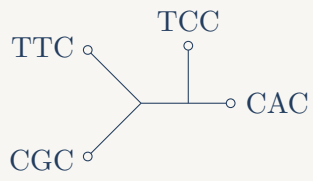
Konstruiert wird zunächst der Baum, welcher die ersten drei Sequenzen a_1, a_2 und a_3 enthält.



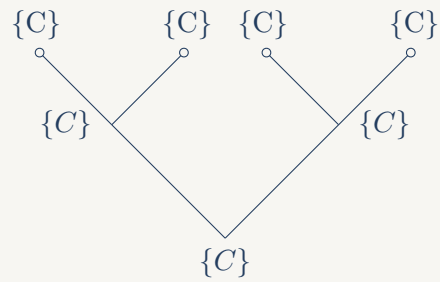
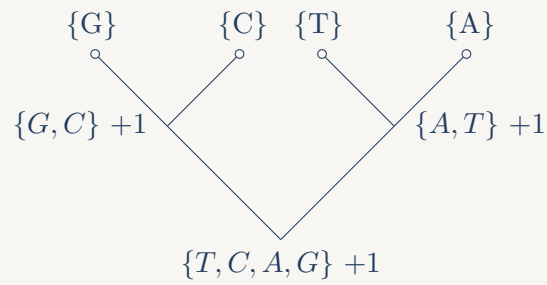
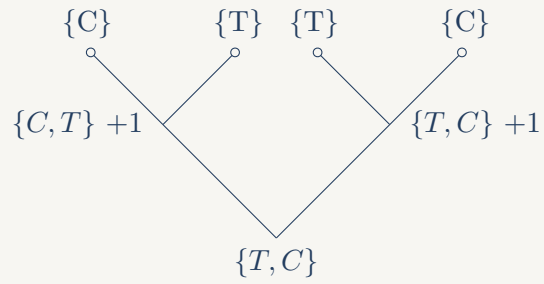
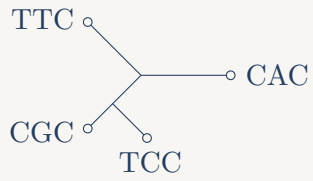
Nun wird die Sequenz a_4 jeweils an den existierenden Kanten eingefügt und der Parsimony Score berechnet.



$$\Sigma = +4$$

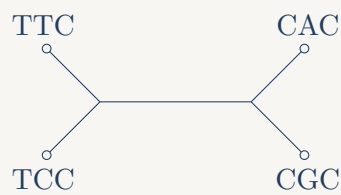


$$\Sigma = +5$$

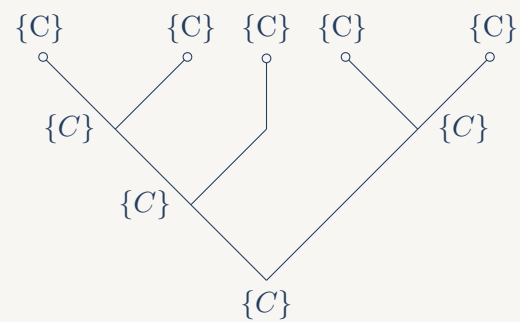
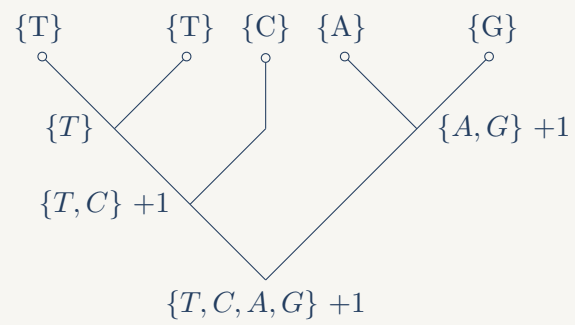
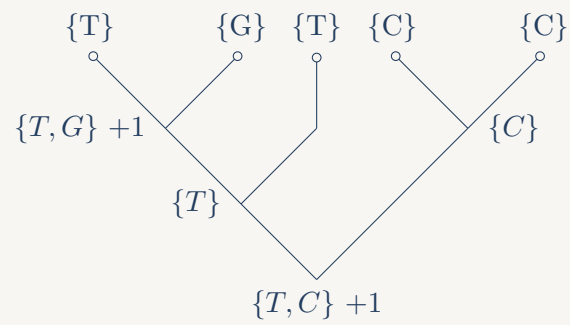
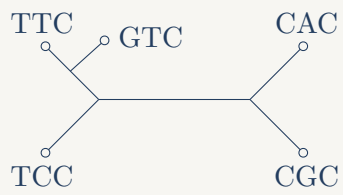


$$\Sigma = +5$$

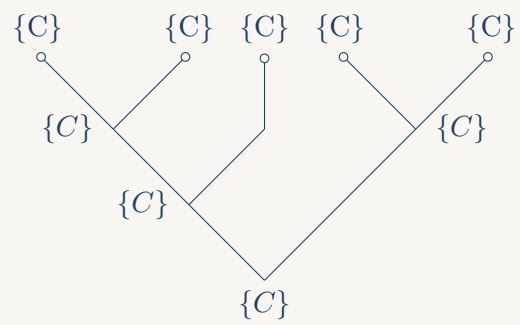
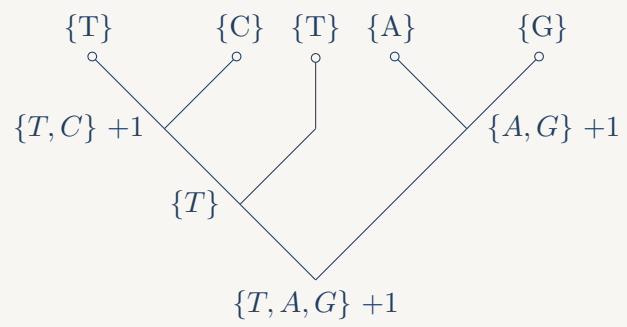
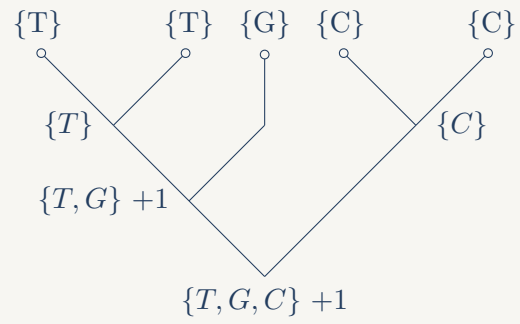
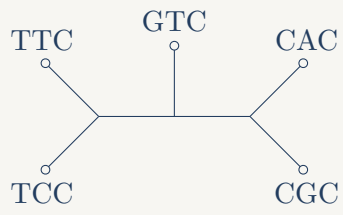
Der erste Baum hat den kleinsten Score:



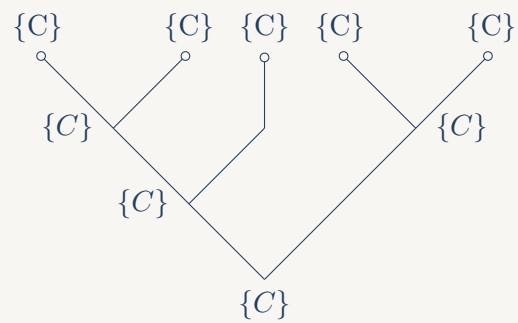
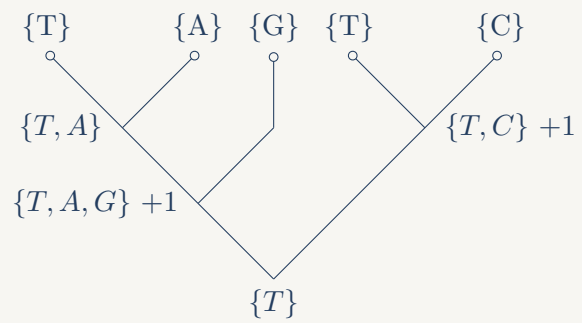
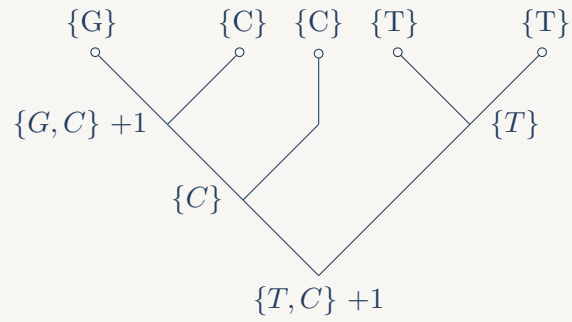
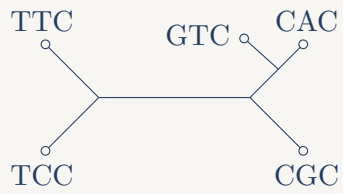
Nun wird dieser Baum verwendet, um a_5 einzufügen:



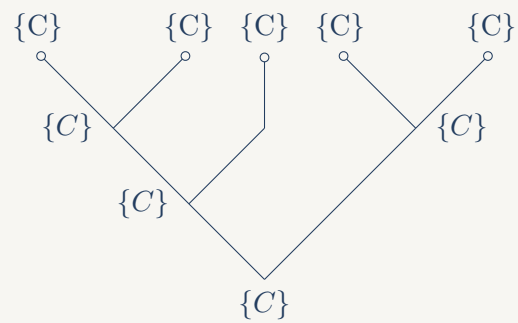
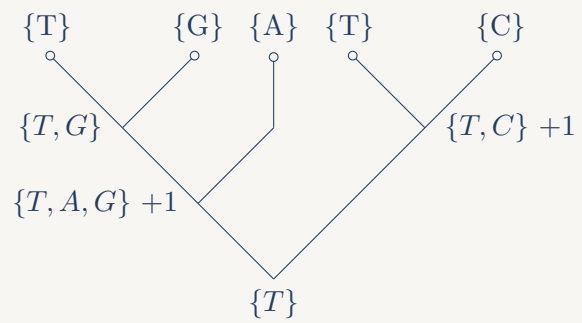
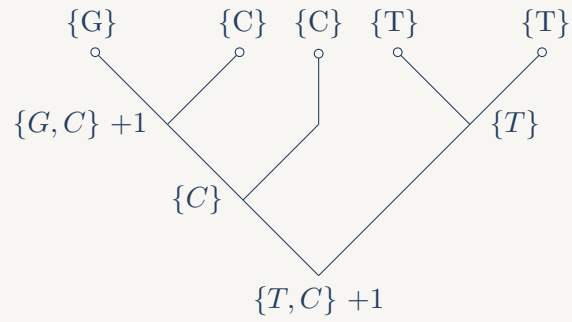
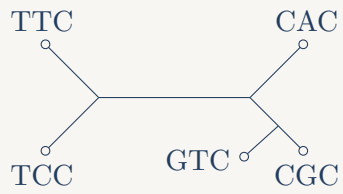
$$\Sigma = +5$$



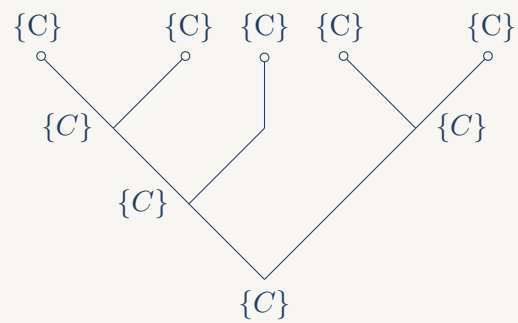
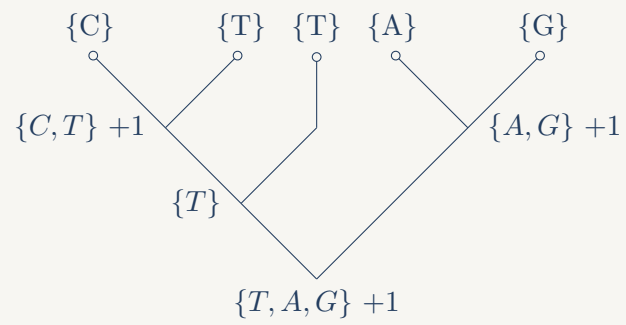
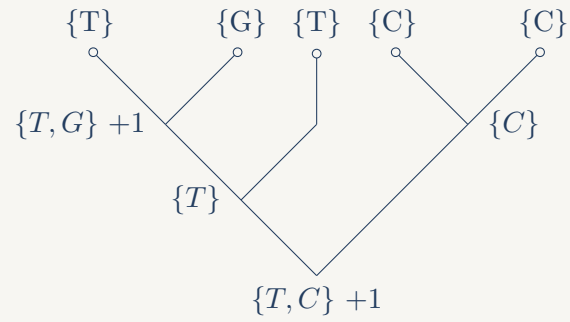
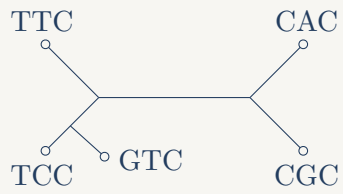
$$\Sigma = +5$$



$$\Sigma = +5$$

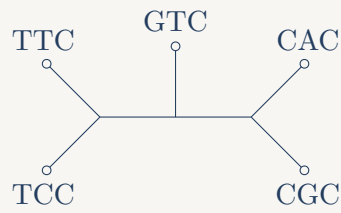


$$\Sigma = +5$$

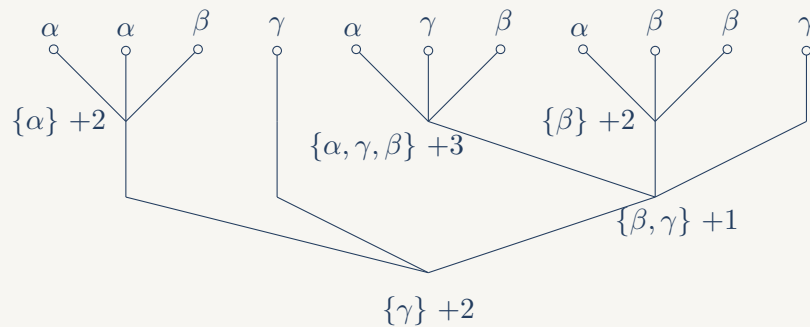


$$\Sigma = +5$$

Da jeder dieser Bäume den gleichen Score besitzt, kann nun einer zufällig ausgewählt werden. Der maximum parsimony tree mit dem Score 5 ist:



Aufgabe 3



Dieser entstandene Baum hat einen Parsimony-Score von 10. Folgender abgeänderter Fitch Algorithmus wurde für obigen Baum verwendet:

Algorithm 1 Forward Pass - ParsimonyScore(v)

Input: A ternary phylogenetic tree T , a state $c(w)$ for each leaf w of T

Output: The parsimony score $PS(T, c)$ for T and c

```

1: Set  $PS(T, c) = 0$ 
2: for all leaf nodes  $v \in T$  do
3:   set  $F(v) = \{c(v)\}$ 
4: end for
5: for each node  $v \in T \neq$  leaf, in bottom-up order do
6:    $C = \{w_1, \dots, w_n\}$  where  $w_i$  children of  $v$ 
7:   for all residues  $r_i$  in  $\Sigma$  do
8:      $x_i = 0$ 
9:     for all  $w_j$  in  $C$  do
10:      if  $r_i \in w_j$  then
11:         $x_i ++$ 
12:      end if
13:    end for
14:  end for
15:  Set  $F(v) = \{r_i \mid x_i \text{ maximum of } \{x_1, \dots, x_n\}\}$ 
16:  for all unsorted pairs  $(w_i, w_j) \in C \times C$  do
17:    if  $w_i \cap w_j = \emptyset$  then
18:       $PS(T, c) = PS(T, c) + 1$ 
19:    else
20:       $PS(T, c) = PS(T, c)$ 
21:    end if
22:  end for
23: end for
24: return score  $PS(T, c)$ 
  
```

Aufgabe 4

Der Code, der sich im file: `Sara_Kemmler_Robin_Bonkass_A5.py` findet, kann mit folgendem Befehl ausgeführt werden:

```
python3 Sara_Kemmler_Robin_Bonkass_A5.py -f1 distances_original.dist -f2 distances_tree1.dist -f3 distances_tree2.dist
```