Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye

Algorithms in Comparative Genomics
Summer 2018

Exercises

Number 5, return 2018 June 15

1. Given two signed multichromosomal genomes

$$A \;=\; [\,7\;\;9\;\;8\,]\;\;[\,2\;\;1\;\;3\;\;5\;\;4\;\;6\;\;-12\;\;11\;\;10\,]$$

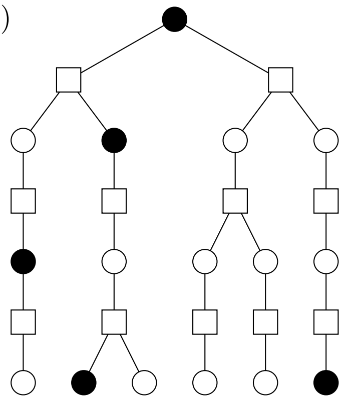
and

$$B \;=\; [\,1\;\;2\;\;3\;\;4\;\;5\;\;6\;\;7\;\;8\;\;9\,]\;\;[\,10\;\;11\;\;12\,].$$

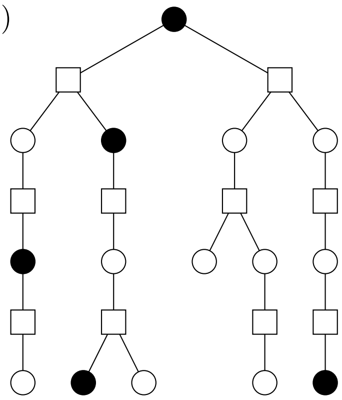
   (a) What is the breakpoint distance between A and B?
   (b) Draw the adjacency graph of A and B.
   (c) What is the DCJ distance between A and B?
   (d) Identify the components of A with respect to B and draw the component tree.
   (e) Calculate the genomic distance.

2. Calculate the optimal cost of the following four trees (filled circles = oriented components, open circles = unoriented real components, crossed circles = unoriented semi-real components):

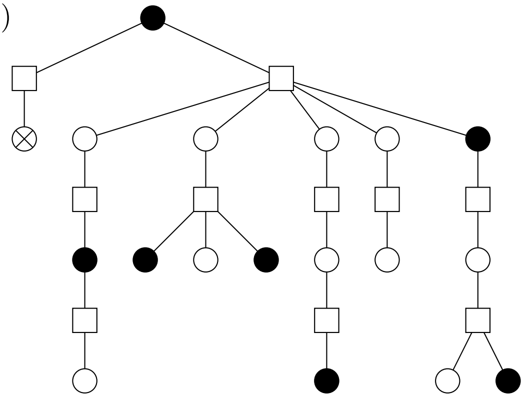
(a)



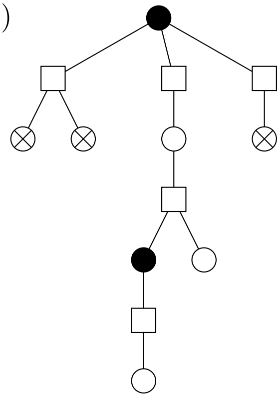
(b)



(c)



(d)



3. Another example for a rearrangement distance is the inversion-indel model, combining inversions and indels. Again, the distance can be phrased in the form

   $$d_{\text{INV-indel}}(A, B) = d_{\text{DCJ}}(A, B) + t_{\text{INV-indel}}(A, B),$$

   where $t_{\text{INV-indel}}$ is the cost of an optimal cover of a component tree. This time, however, the good and bad vertices of the component tree can in addition have a blue and/or an orange dot.

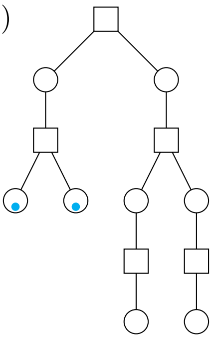
   The cost of a path is

   - 1 if the path is short;
   - 1 if the path is long and its endpoints share a dot of the same color;
   - 2 otherwise (long path with no same dot).

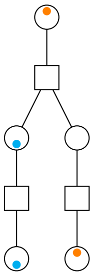
   Details can be found in the following paper: Willing, Zaccharia, Braga, Stoye. On the inversion-indel distance. BMC Bioinformatics 14(Suppl 15):S3, 2013.

   (a) What is the optimal cover cost of the following two trees:

   (a.i)

   (a.ii)

   (b) Give an example showing that the number of leaves of the different types (short with no dot, orange dot, blue dot, both dots; long with no dot, orange dot, blue dot, both dots) alone does not suffice to give the cost of a tree. In other words: Give two trees with the same number of leaves of each type, that have different cover costs.