Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye
Dr. Marília D. V. Braga

Algorithms in Genome Research
Winter 2018/2019

Exercises

Number 12, Discussion: 2019 February 08

1. The DCJ, SCJ and breakpoint distances between two genomes A and B are given by

$$d_{\mathrm{DCJ}} = n - c - \frac{p_{AB}}{2}, \qquad d_{\mathrm{SCJ}} = 2n - 2c_2 - p \qquad \text{and} \qquad d_{\mathrm{BP}} = n - a - t/2,$$

where n is the number of genes; c is the number of cycles, c_2 the number of 2-cycles, p the number of paths and p_{AB} the number of AB-paths in the adjacency graph; and a is the number of adjacencies and t the number of telomeres common to A and B.

Theoretical bounds for the SCJ distance with respect to the breakpoint and to the DCJ distance are

$$d_{\mathrm{BP}}(A, B) \quad \leq \quad d_{\mathrm{SCJ}}(A, B) \quad \leq \quad 2\, d_{\mathrm{BP}}(A, B)$$

and

$$d_{\mathrm{DCJ}}(A, B) \quad \leq \quad d_{\mathrm{SCJ}}(A, B) \quad \leq \quad 4\, d_{\mathrm{DCJ}}(A, B).$$

Give non-trivial examples showing that these bounds are tight.

2. Given two mixed genomes

$$A \;=\; [\,3\ 4\ -8\,]\ \ [\,2\ 1\ 7\ 5\ 6\,]\ \ (\,11\ 12\ -13\ 9\ 10\,)\ \ [\,-15\ -14\,]$$

and

$$B \;=\; [\,1\ 2\ 3\ 4\ 5\,]\ \ (\,6\ 7\,)\ \ [\,8\ 9\ 10\ 11\ 12\ 13\ 14\ 15\,],$$

where round parentheses $(\ldots)$ indicate circular and square brackets $[\ldots]$ indicate linear chromosomes.

  (a) What is the breakpoint distance between A and B?

  (b) What is the DCJ distance between A and B?

  (c) What is the SCJ distance between A and B?

  (d) Identify the operations that explain the differences between the distances (if any).

3. Given three unichromosomal linear genomes

$$A \;=\; [\,1\ 2\ 3\ 4\ 5\,], \quad B \;=\; [\,2\ 1\ -5\ -4\ 3\,], \quad C \;=\; [\,2\ 3\ 1\ -5\ -4\,].$$

Compute an SCJ median of A, B and C. (Note that the median does not necessarily need to be unichromosomal linear.)