

Home Work 1

Introduction to Bioinformatics Algorithms, Winter 2011

Given on Thursday, January 6, 2011. Due on Wednesday, January 12. (E.mail to TA).

State your answers legibly and concisely. Your solutions will be graded on correctness, elegance, **clarity** and originality. Your proofs should **avoid getting bogged down in too much detail**. Your algorithms should be as efficient as possible (inefficient algorithms will result in lower grades). Please note that the work handed in must be your own. No late assignments will be accepted without prior arrangement. Each problem is 20 points.

Problem 1. Write a program for computing the 2-break distance between genomes (consisting of circular chromosomes). Use the program to answer the following questions:

- Given circular permutations $+1, +3, -7, -6, +4, +2, -5$ and $+4, -7, +3, -6, +1, -5, +2$ find the minimum number of 2-breaks required to transform one permutation into another. Output a shortest series of 2-breaks transforming one genome into another and compute the number of such minimum transformations.
- What is the expected 2-break distance $distance(n)$ between two *random* genomes on n genes? Use the program you wrote to construct the plot of $distance(n)$ for $10 \leq n \leq 100$ using simulations. You will need to rigorously define the notion of a random genome and write a program for generating random genomes.
- Find the *reversal distance* between the linear versions of these permutations (feel free to use <http://grimm.bioprotects.org/GRIMM>). Find minimal number of breakpoint re-uses in the transformations of $+1, +3, -7, -6, +4, +2, -5$ into $+4, -7, +1, +3, -6, -5, +2$. Please note that this problem asks you to find the reversal distance rather than the 2-break distance (you don't need the program you wrote to answer this question).

Problem 2. Given k extant genomes, reconstruct the evolutionary history of their rearrangements. **Warning: the previous sentence is not a well-defined problem formulation.**

- Transform this problem into a well-defined computational problem.
- It is known that reversals are more common than translocations, fusions, and fissions. Let's assume that reversals, translocations, fusions, and fissions account for 60%, 30%, 5% and 5% of all rearrangements correspondingly. Change the problem formulation to take advantage of this new piece of information.

Problem 3. You are given two relatively short (let's say 1000 nucleotides in length) segments of DNA from human and mouse and told that, in addition to mutations, insertions, and deletions of individual nucleotides, there was a *single* reversal between these two segments. You are asked to find the breakpoints of this reversal. **Warning: the previous two sentences do not represent a well-defined problem formulation.**

- Transform this problem into a well-defined computational problem.

- Suggest an algorithm for solving this problem and estimate its computational complexity.

Problem 4. Given a tumor karyotype (e.g., MCF7 chromosome painting image shown in the class), reconstruct the set of rearrangements that produced this karyotype. Transform this (vaguely defined) question into a well-defined computational problem.

Problem 5. Reconstructing Phylogenetic Trees. Let $P_1 \dots P_n$ be n signed genomes with circular chromosomes that evolved by some unknown 2-breaks according to an unknown evolutionary tree. Without loss of generality, we assume that there was at least one 2-break on every branch of the tree.

One may classify a 2-break as independent if it creates exactly two new breakpoints (does not reuse breakpoints). We call the rearrangement process independent if all its 2-breaks are independent. Prove the following theorem and devise an algorithm for evolutionary tree reconstruction in the case of genomes evolved by independent 2-breaks.

Theorem. If n genomes of length m are produced by independent 2-breaks, then both the correct evolutionary tree for these genomes and the ancestral architectures of all its branching vertices may be reconstructed in polynomial time.