

Restriction maps

MPA-PRG: Programming in Bioinformatics

Exercise 6

Introduction

- restriction motifs
 - specific recognition sites on DNA strand recognized by restriction endonuclease (RE) that cleaves DNA into fragments
- Double Digest Problem (DDP)
 - digestion of a single sequence by two different enzymes
- Partial Digest Problem (PDP)
 - incomplete (partial) digestion of a single sequence by a single enzyme

DDP

- assume sets of fragment sizes:
 - $\Delta X_A = \{a_1, a_2, \dots, a_m\}$ – products generated by digestion with A
 - $\Delta X_B = \{b_1, b_2, \dots, b_n\}$ – products generated by digestion with B
 - $\Delta X_{AB} = \{c_1, c_2, \dots, c_{m+n-1}\}$ – products generated by digestion with A + B
- enumeration of all possible combinations of fragment arrangements for both enzymes
- for each combination:
 - take the fragments sizes
 - create position maps
 - combine positions
 - calculate successive differences
 - sort differences
 - compare results with products of double digestion (ΔX_{AB})

Task - DDP

- In R, create a function for the brute-force DDP algorithm for one possible arrangement of fragments.
- Modify it to work with all possible arrangements of fragments.

PDP

- assume a multiset of fragment sizes ΔX and we search for a set of positions $X = \{x_1 = 0, x_2, \dots, x_n\}$, that generates $\Delta X = \{x_j - x_i: 1 \leq i < j \leq n\}$

- the number of restriction sites:

$$\frac{N(N-1)}{2} = m$$

N – the number of restriction sites, m – the number of fragments in the multiset ΔX

- $x_1 = 0$
- find maximum value in ΔX = corresponds to the longest possible fragment (upper right corner of the previous table)
- deduce possible position of restriction site
- remove fragments sizes from ΔX that are generated by this position
- repeat until all fragments are exhausted

Task - PDP

- In R, implement a recursive algorithm for the Partial Digest Problem (PDP) according to the pseudocode.