Università degli Studi di Verona

Modelli Biologici Discreti
a.a. 2014/2015

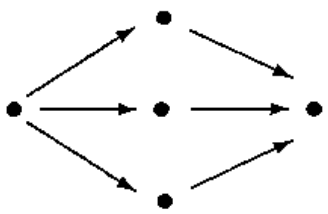
Docente: Zsuzsanna Lipták

Verona, 5 Nov. 2014

Exercises on Fragment Assembly

(Ex. 5-11 are from the book "Introduction to Computational Molecular Biology" by João Setubal and João Meidanis, 1997, ch. 4.)

1. What is the reverse complement of `ACCGTCG`?
2. Give an example for an rc-palindrome.
3. Produce an example for $\mathcal{F}$ where the shortest common superstring is not unique.
4. p. 27 (updated version: p. 28) of the lecture slides: Prove that $d_s(u, v) \leq |u|$ for any two strings u, v. Give an example where $d_s(u, v) \neq d_s(v, u)$.
5. What is the minimum, maximum, and average coverage for the layout on p. 7 of the lecture slides?
6. What is the smallest value for ϵ such that the layout on p. 10 is valid under the Reconstruction model?
7. Let $\mathcal{F} = \{\texttt{ATC}, \texttt{TCG}, \texttt{AACG}\}$. Find the best layout for this collection according to the Reconstruction model with $\epsilon = 0.1$ and $\epsilon = 0.25$. Be sure to consider reverse complements.
8. Construct the overlap graph for $\mathcal{F} = \{\texttt{AAA}, \texttt{TTA}, \texttt{ATA}\}$. Find a shortest common superstring for this collection.
9. Find sequences that give rise to the following overlap graph, where only edges with positive weights are shown. The particular weights you come up with not important as long as they are strictly positive.



10. Let $\mathcal{F} = \{\texttt{TCCCTACTT}, \texttt{AATCCGGTT}, \texttt{GACATCGGT}\}$. Finid the best set of contigs for this collection according to the Multicontig model with $\epsilon = 0.3$ and $t = 5$.

11. Let A, B, C, D, E, X, Y, and Z be blocks in a target DNA sequence. Suppose that these blocks are larger than any fragment that can be sampled from the trget molecule. Fill in the blanks below with letters X, Y, or Z so that this sequence becomes inherently ambiguous to assemble.

$$\text{target} = A \quad X \quad B \quad \ldots \quad C \quad Y \quad D \quad \ldots \quad E \quad Z$$

12. Find an example where the Greedy Algorithm does not produce a shortest superstring.