

Algorithms in Bioinformatics II, SS2003

Assignment sheet # 10

Daniel Huson

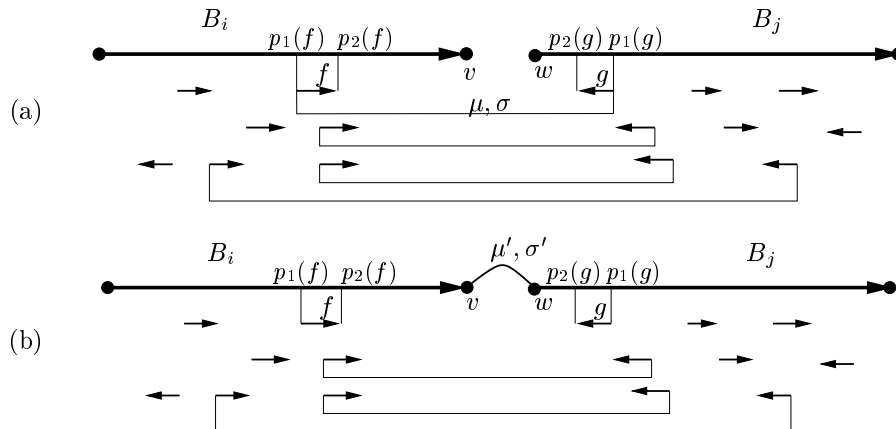
June 29, 2003

1 The contig-mate graph (2 points)

Construct the contig-mate graph for the following data: We are given four contigs $B_1 \dots B_4$, of length 3000, 5000, 2000 and 2000, respectively. Moreover, we are given 6 pairs of mates $\{f_1, g_1\}, \{f_2, g_2\}, \dots$, the first four coming from one library, with mean length $\mu = 2kb$ and standard deviation $\sigma = 200bp$, and the others coming from a second library, with mean length $\mu = 10kb$ and standard deviation $\sigma = 1kb$. For each read in the following table, we list which contig it hits and where (positions p_1 and p_2 in the contig to which the begin and end of the read aligns). All reads have length 500.

i	f_i		g_i		Library	
	contig	$p_1 - p_2$	contig	$p_1 - p_2$	μ	σ
1	B_1	500 - 0	B_3	1500 - 1000	$2k$	100
2	B_1	1000 - 500	B_3	1000 - 500	$2k$	100
3	B_2	4000 - 4500	B_4	1500 - 2000	$2k$	100
4	B_2	4500 - 5000	B_4	1000 - 1500	$2k$	100
5	B_2	500 - 1000	B_3	1500 - 2000	$10k$	$1k$
6	B_2	2500 - 3000	B_1	1500 - 1000	$10k$	$1k$

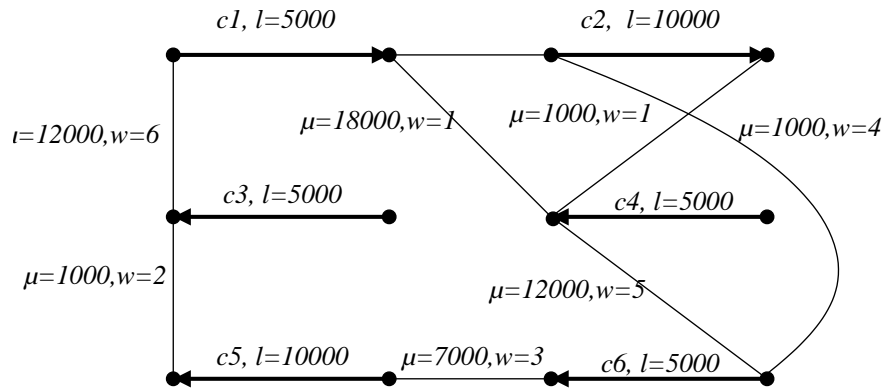
To construct the contig-mate graph, represent each contig by an oriented contig edge and then define a mate edge between nodes v and w say, if a mate pair f, g exists that spans the gap between v and w , as indicated here:



2 Running the greedy path-merging algorithm (3 points)

Given the following contig-mate graph G , illustrate step-by-step how the greedy path-merging algorithm proceeds to compute a scaffolding for this graph. In each step, indicate which mate-pair edge is currently

active and highlight the currently *selected* edges. Indicate which inferred edges (if any) are inserted into the graph.



3 New Trends in Phylogenetics and Genomics (5 points)

A ZBIT-workshop on phylogenetics on *New Trends in Phylogenetics and Genomics* is being held from 1.-4.7.2003 at the Max-Planck Haus. This should be an interesting meeting. Instead of attending the usual lecture on Wednesday, you are requested to attend some of the talks instead. In particular, you should attend the talk by Mike Steel on Wednesday 10:30-11:15h.

This assignment is designed to encourage you to attend some of the other talks, too.

Please write a one page report on two talks that you attended at the workshop. It should include, but not be restricted to:

- Name of the speaker and title of the talk.
- Context and motivation for the work presented.
- Emphasis of the talk: biology, computer science or mathematics?
- What was the main problem addressed?
- How would you categorize this talk: new ideas, research in progress or final results?
- Strengths and weaknesses of the work.
- Strengths and weaknesses of the presentation.
- Applications of the results?

Assignments due: **Monday, July 6, 10am**