Nanopore automata

Ian Holmes[1,2,*]

1 Lawrence Berkeley National Laboratory, Berkeley, CA, USA

2 Department of Bioengineering, University of California, Berkeley, CA, USA

Contents

1 Abstract

State machine algorithms for aligning Nanopore read pairs.

2 Specification

2.1 Parameterization algorithm

Given the following inputs

- Reference genome (FASTA)

- Segment-called reads (FAST5/HDF5)

Perform the following steps

- Perform Baum-Welch to fit a rich model

Rich model incorporates segment statistics.

2.2 Reference search algorithm

Given the following inputs

- Segment-called reads (FAST5/HDF5)

- Parameterized rich model

Perform the following steps

- Prefilter read pairs using k-mer threshold heuristics

- Perform Viterbi alignment using intersection of two rich models

2.3   Implementation

2.4   Evaluation

3   Methods

3.1   Model

3.2   Parameterization

3.2.1   Baum-Welch algorithm

3.3   Search

3.3.1   Prefiltering heuristic

(Omit this step in first implementation?)

- Build database of k-mer frequencies across all reads

- Compute likelihood of each read using a "bag of words" model

- Compute likelihood of each read-pair using a "bag of paired words" model

- Compute posterior probability of a match

3.3.2   Viterbi algorithm

4 Results

5 Discussion

6   Acknowledgments

7 Figure Legends