

Najib Ishaq

Reading 1: Myers Paper

This paper is a review and history of partial-sequence search algorithms and, particularly, BLAST. It starts by talking about a fast, approximate search for short, partial matches in a long sequence and then filtering the matches to remove mistakes. It explains how this hybrid method is often several orders of magnitude better than either on its own.

The core of the idea is to divide the search query into small k-mers, searching for matches to the k-mers, and finally putting the matches together. Matches are reported as hits if they score high enough. The question then becomes, what k-mer size results in the best expected-time performance.

The paper then dives into proving a recurrence relation for the neighborhood size around search queries. With this in hand, the algorithm can be proven to be bounded by a sublinear, concave, monotone increasing function.

The author then lays out pseudocode for generating these tight neighborhoods and argues for the bound on the total expected running time for BLAST.