BWA-MEM tool to align sequence reads, clone sequences, and assembly contigs.

Technology used:-BWA-MEM

Tool name:-BWA-MEM

Tool name:-BWA-MEM

Approach:-

1. Aligning a single query sequence:- 1.1 Seeding and re-seeding 1.2 Chaining and chain filtering 1.3 Seed extension

|  |  |
| --- | --- |
|  |  |

1. Pair-end mapping:-2.1Rescuing missing hits 2.2 Pairing

Implementation details:-

1.1BWA-MEM follows the canonical seed-and-extend paradigm.

1.2 We call a group of seeds that are colinear and close to each other a chain. We greedily chain the seeds while seeding and then filter out short chains largely contained in a long chain and much worse than the long chain.

1.3  We rank a seed by the length of the chain it belongs to and then by the seed length. For each seed in the ranked list, from the best to the worst, we drop the seed if it is already contained in an

alignment found before, or extend the seed with a banded affine-gap-penalty dynamic programming (DP) if it potentially leads to a new alignment. BWA-MEM’s seed extension differs from the standard seed extension in two aspects.

* 1. Like BWA (Li and Durbin, 2009), BWA-MEM processes a batch of reads at a time.BWA-MEM prefers to pair the two ends; otherwise, it prefers the unpaired alignments. BWA-MEM chooses the pair (i, j) that maximizes Sij as the final alignments for both ends. This pairing strategy jointly considers the alignment scores, the insert size, and the possibility of chimeric read pairs.

Performance analysis:-

Accuracy:-On accuracy, NovoAlign is the best. BWA-MEM is close to NovoAlign for PE reads and is comparable to GEM and Cushaw2 for SE.

Speed:-On speed, BWA-MEM is similar to GEM and Bowtie2 for this data set but is about 6 times as fast as Bowtie2 and Cushaw2 for a 650bp long-read data set.

Advantage:-

Accuracy:-On accuracy, NovoAlign is the best. BWA-MEM is close to NovoAlign for PE reads and is comparable to GEM and Cushaw2 for SE.

Speed:-On speed, BWA-MEM is similar to GEM and Bowtie2 for this data set but is about 6 times as fast as Bowtie2 and Cushaw2 for a 650bp long-read data set.

We speculate BWA-MEM is more performant for longer reads firstly because of its advanced seeding algorithm, which identifies most standing seeds with one pass of the read, and secondly because

of banded dynamic programming, which guarantees linear time complexity in the length of query sequences. BWA-MEM and BWA-SW can also identify chimeric reads, a crucial feature for contig alignment but lacking in most NGS long-read mappers.

Disadvantage:-

Nucmer finished the alignment in 25 seconds, giving 105,505 substitution differences between the strains. BWA-MEM is slower. It finished the alignment in 131 seconds, reporting 104,321substitutions of which 102,241overlapping with the Nucmer output.