



The high-performance computing platform we use enables BioVelocity to use a brute force index, built out of all possible base pair sequences of various k-mer lengths. A k-mer is a DNA sequence substring contained within a larger biological sequence and it is key for assembling DNA sequences. This highly effective “brute force” index is used to map against thousands of references and allows for quick alignment of the k-mers simultaneously.

BioVelocity involves simple alignment, where the reads align to the reference genome from start to finish of the read length. A Needleman-Wunch algorithm (this algorithm is widely used and also known as the optimal matching algorithm and the global alignment technique) was also added to the BioVelocity pipeline to facilitate the selection of indels (indels are the insertion or deletion of bases in the genome).

BioVelocity read mapping conducts global pairwise alignment; without skipping any bases. The k-mer index searches the entire reference genome globally to find a matching sequence that is complementary with a read. Local pairwise alignment aligns the reads on the reference genomes at random positions where the k-mer indexes are matched with the reference genomes.