

**EMAGEN: An Efficient Approach to Multiple Whole Genome Alignment*****Deogun, Jitender S., Zhang, Jingyi, Ma, Fangrui******Department of Computer Science and Engineering, University of Nebraska, Lincoln, NE, USA***

Following advances in biotechnology, many new whole genome sequences are becoming available every year. A lot of useful information can be derived from the alignment and comparison of different genomes. However, most of the current research focuses on pairwise genome alignment, and only a few available applications can efficiently align multiple genomes. In this paper, we present an efficient approach to align closely related multiple whole genomes, combining suffix arrays, graph theoretic formulation and existing tools for gap (short sequence) alignment. Our approach first finds a maximum set of aligned conserved regions among multiple whole genomes, and then aligns the gaps between consecutive conserved regions with Clustal W.

We present two methods to find the maximum set of aligned conserved regions among whole genomes. In first method, called Direct Matching (DM), multiple whole genomes are aligned with their DNA sequences. However, because most parts of prokaryotic genomes are encoded regions, we introduce second method, Functional Matching (FM), to especially align multiple prokaryotic genomes with their concatenated protein sequences. We present experimental results for both methods and give the analysis of the results. The FM method generates much better results for less closely related prokaryotic genomes than DM method. It outputs more and longer conserved regions, which conveys more accurate and detailed information about the conservation and inheritance of genomes, and generates more detailed alignments.