

Two levels of genetic evolution, both running in the course of time, can be distinguished. The first one concerns changes of sequences at small scale with mutational events limited to bases mutations and insertion/deletions. At the level of genomes, evolutionary events involve large segments of chromosomes which are “rearranged”: transposed, translocated, reversed, duplicated etc. Evolution is then described in terms of chromosomal organization, that is the order in which the genes occur along the sequences. So computational approaches for comparing two genomes need generally two stages. The first stage, which relies to the evolution at small scale, consists in determining clusters of orthologous (related) genes in the genomes. After this stage, a genome is represented by the sequence of the cluster idents of its genes in the order in which they appear. Then comparison techniques dealing with rearrangements such as sorting by reversal or finding conserved segments (synthenies) can be applied.

A concern with the preceding type of approaches is that the identification of the clusters of orthologous is only based upon the similarity between genes and is totally disconnected from their respective neighbourhood in the genomes. We propose an extension of the N-map method, a pairwise asymmetrical approach which avoids this drawback. Genomes being still represented as sequences of genes, the extended N-map method remains to find the best way of partitioning the first sequence of genes into N parts and align each part against the second sequence, in order to maximize the sum of the alignment scores. Alignment of parts is computed using standard gap penalty and additive similarity score between genes (derived from the corresponding proteins similarity). We show how to compute the extended N -map of two genomes s and t with time complexity $O(|s| \times |t| \times N)$ using $O(|s| + |t| \cdot N)$ memory space, where $|\cdot|$ design the number of genes of the genome. The choice of the most relevant number of parts for comparing two given sequences of genes is done with regards to the empirical distributions of the total scores of extended N -maps (with N in a reasonable range) obtained by shuffling the first sequence of genes. Finally the approach is illustrated with comparisons of pairs of genomes