TITLE OF INVENTION: [METHOD FOR THE COMPRESSION OF GENOME SEQUENCE DATA]

ABSTRACT

The invention relates to a reference-based method for the compression of genome sequence data produced by a sequencing machine. The sequences of nucleotides or bases, that have been previously aligned to a reference sequence, are determined to be perfectly mapped, imperfectly mapped or unmapped with the reference sequence; and then coded according to said determination. The determining step comprises comparing, for each imperfectly mapped sequence, the number of mismatches between said sequence and the reference sequence with a reference threshold value, and encoding the imperfectly mapped sequences according to distinct encoding processes, depending on the result of said comparison method for the compression of genome sequence data produced by a sequencing machine.

[Fig. 1.]