ABSTRACT

DNA samples are found in fragments, obtained in traces of a crime scene, collected from hair or blood samples, for genetic or paternity tests. To identify whether this fragment belongs or not to a given DNA sequence it is necessary to compare it with a determined sequence which usually come from a database, for instance, to point asuspect. To this end, we need an efficient tool to perform the alignment of the DNA sequence found with the ones stored in the database. The alignment of DNA sequences, which is a field of bioinformatics that helps to understand the relationship between genetic sequences and their functional relationships and parenting. This task is often performed by software that scan clusters of databases, which requires high computing effort, thus increasing the cost of DNA sequences alignment projects. This work presents a parallel hardware architecture, for BLAST algorithm, to DNA pairwise alignment. This is the original version of the BLAST algorithm, that resulted in several other versions. The BLAST algorithm is a heuristic method and is the fastest algorithm for sequence alignment. The strategy of BLAST is to divide the sequences into smaller subsequences of size w. After making comparisons in these subsequences, algorithm steps analyzes only the subsequences that are identical. Thus, reducing the number of tests and combinations needed to perform the alignment. For each identical sequence found, three steps are followed by the algorithm: seeding, extension and evaluation. The proposed hardware architecture is based on the characteristics of the algorithm to implement a fully parallel hardware, where the basic steps of BLAST are pipelined. The proposed architecture was implemented in FPGA and the results show a comparison between the area occupied, number of cycles and maximum frequency of operation permitted, as a function of alignment parameters. The result is a hardware architecture in reconfigurable logic, scalable, efficient and with low cost, capable of aligning the pairs of sequences using BLAST algorithm.

Keywords: DNA Matching. Bioinformatics. Hardware. Parallel Architecture. Heuristic.