



Fast Implementation of e-PCR on Cray SV1 System

Tal, J.¹, Lapidot, M.² and Safran, M.³

¹ Cray LTD., Herzliya, Israel

² Department of Molecular Genetics, Weizmann Institute of Science

³ Bioinformatics and Biological Computing Unit, Biological Services, Weizmann Institute of Science

The highly specific and sensitive PCR process provides the basis for sequence-tagged sites (STSs), unique landmarks that have been used widely in the construction of genetic and physical maps of the human genome. Electronic PCR (e-PCR) refers to the process of recovering these unique sites in DNA sequences by searching for subsequences that closely match the PCR primers and have the correct order.

The Weizmann Institute's GeneCards/UDB project uses the e-PCR program from NCBI to search for known STS primer sequences in contigs, in order to produce a more precise positioning of markers, genes, and EST clusters in sequenced regions.

The standard implementation of e-PCR on a Sun/Solaris platform takes a few days to complete the search for primers sequences within the whole human genome.

A fast approach to e-PCR was implemented on a Cray-SV1 Supercomputer. The Cray SV1 Parallel Vector architecture has special hardware features that are used to boost the performance of the STS search procedure by an order of magnitude.

Performance of sample runs shows a boost of ~8X in cpu performance between our fast approach on Cray SV1@300MHz and SGI origin2000@400MHz, when no mismatches are allowed. This ratio grows when mismatches are allowed. When running with a threshold of two mismatches a ratio of ~28X was measured. The performance ratio of the new Cray code relatively to the standard code running on the Cray is ~7X with no mismatches and ~13X with two mismatches.