



WINNER



2004 R&D 100 Awards Winner

mpiBLAST: A High-Speed Software Catalyst for Genetic Research

BLAST, an open-source software package distributed by the National Center for Biotechnology Information, has become the ubiquitous genomic-sequencing tool in molecular biology. With mpiBLAST, our open-source parallelization of BLAST, we have dramatically enhanced BLAST's throughput and minimized its response time. The mpiBLAST software uses a new process known as in-memory database segmentation, in which a database is chopped into memory-sized pieces so that each compute node searches only a distinct portion of the database. When each portion has been searched, the message-passing interface (mpi) handles the communication to merge the results from each compute node. Thus, a search of a 300-kilobyte query that took 1,346 minutes (22.4 hours) using BLAST takes only a few minutes with mpiBLAST.

Applications

- Enables quick identification of previously unknown viruses, e.g., the SARS coronavirus, West Nile virus, and different strains of AIDS and cancer
- Accelerates complete genomic sequencing of organisms and drug discovery, reducing bioterrorism threats via pathogen detection and identification
- Contributes to other medical efforts, e.g., phylogenetic profiling and pairwise genome alignment
- Provides a more effective data-mining technique, e.g., technique could help identify and correlate intelligence and reconnaissance information and parallelize Internet search engines