



PESQUISA DE SIMILARIDADE DE CADEIAS DE DNA E PROTEÍNAS USANDO PROCESSAMENTO PARALELO

Luciano Machado Granella (Bolsista SAE/UNICAMP) e Prof. Dr. Ricardo de Oliveira Anido (Orientador), Instituto de Computação - IC, UNICAMP

A descoberta de homologia entre seqüências de aminoácidos ajuda a determinar a que família de proteínas pertence a nova proteína e pode dar as primeiras pistas sobre a funcionalidade desta nova proteína. Nos últimos anos, os bancos de dados de seqüências de proteínas e DNA cresceram muito em tamanho e em importância porque ficou mais fácil encontrar seqüências homólogas, mas por outro lado, o trabalho computacional cresceu também. A ferramenta BLAST (Basic Local Alignment Search Tool) é a mais usada atualmente por pesquisadores no mundo todo, mas a ausência de uma versão que faça uso de computação distribuída e paralela torna árdua a busca de similaridades em banco de dados de proteínas e DNA. O método usado neste projeto para suprir esta necessidade foi o desenvolvimento de um script em PERL que divide o conjunto de seqüências de DNA ou aminoácidos e também os distribui entre os computadores encarregados de realizar o processamento e coletar os resultados finais, usando uma rede de alta velocidade para a comunicação entre tais computadores. Esta implementação permitiu que uma busca que consumiria cinco dias de processamento num único computador fosse realizada em apenas sete horas por um conjunto de 60 computadores comuns.

Processamento paralelo - BLAST - Homologia