

Referências Bibliográficas

- [1] S. Altschul, W. Gish, W. Miller, E. Myers, D. Lipman; A Basic Local Alignment Search Tool; Journal of Molecular Biology v.215, p.403-410, 1990; 1
- [2] R. Bjornson, A. Sherman, S. Weston, N. Willard, J. Wing. TURBOBLAST: A parallel implementation of BLAST built on the TURBOHUB; In Proceedings of the International Parallel and Distributed Processing Symposium, p.183-190, 2002; 1, 4.2.1
- [3] D. Benson, I. Karsch-Mizrachi, D. Lipman, J. Ostell, D. Wheeler; GenBank; Nucleic Acids Research, v.31, pp.23-27, 2003; 2.1.2, 6.2
- [4] NCBI BLAST, <http://www.ncbi.nlm.nih.gov/BLAST>; acessado em 07/07/2007; 6.2
- [5] P. Carvalho, R. Glória, A. Miranda, W. Degreve; Squid - a simple bioinformatics grid; BMC Bioinformatics, v.6, p.197-204, doi:10.1186/1471-2105-6-19; 2005; 2.4, 4.1
- [6] E. Chi, E. Shoop, J. Carlis, E. Retzel, J. Riedl; Efficiency of Shared-Memory Multiprocessors for a Genetic Sequence Similarity Search Algorithm; Technical Report TR97-005, University of Minnesota Computer Science Department, 1997; 1, 2.1.1, 3.2, 4
- [7] G. Cochrane, P. Aldebert, N. Althorpe, M. Andersson, W. Baker, K. Bates, S. Bhattacharyya, P. Browne, A. Broek, M. Casto, K. Duggan, R. Eberhardt, N. Faruque, J. Gamble, T. Kulikova, C. Lee, R. Leinonen, Q. Lin, V. Lombard, R. Lopez, M. McHale, H. McWilliam, G. Mukherjee, F. Nardone, M. Pastor, S. Sobhany, P. Stoehr, K. Tzouvara, R. Vaughan, W. Zhu, R. Apweiler; EMBL Nucleotide Sequence Database: developments in 2005; Nucleic Acids Research v.34, p.10-15, 2006; 1
- [8] T. Cormen, C. Leiserson, R. Rivest, C. Stein; Algoritmos: Teoria e Prática; Editora Campus, 2002; 2.2
- [9] R. Costa; Alocação de Dados e Distribuição de Carga para Execução Paralela da Estratégia BLAST de Comparação de Sequências; Dissertação de Mestrado, Departamento de Informática, PUC-Rio, 2002; 1, 2.1.1, 3.2, 4.1, 5.1.5, 5.2.3

- [10] R. Costa, S. Lifschitz; Database allocation strategies for parallel BLAST evaluation on clusters; *Distributed and Parallel Databases*, v.13, p.99-127, 2003; 1, 2, 4.1, 4.2.1
- [11] A. Darling, L. Carey, W. Feng; The Design, Implementation, and Evaluation of mpiBLAST; *ClusterWorld Conference & Expo and the 4th International conference on Linux Clusters: The HP Revolution 2003*; LA-UR 03-2862, 2003. 1, 2, 2.1.1, 2.1.2, 4.2.2, 5.1.4, 5.1.5, 5.2.3
- [12] A. Deshpande, D. Richards, W. Pearson; A platform for biological sequence comparison on parallel computers; *Computer Applications in the Biosciences*, v.7, n.2, p.237-247, 1991; 2.1.1, 3.2
- [13] D. DeWitt, J. Gray; *Parallel Database Systems: The Future of High Performance Database Systems*; *Communications of ACM* 35, v.6, p.85-98, 1992;
- [14] R. Elmasri, S. Navathe; *Fundamentals of Database Systems*; 4nd Edition. Addison-Wesley, 2004; 5.1.3
- [15] I. Foster; *Designing and Building Parallel Programs*; Addison Wesley, 2005; 2.1, 2.1.1, 2.1.1, 6.1
- [16] J. Grant, R. Dunbrack, F. Manion, M. Ochs; BeoBLAST: Distributed BLAST and PSI-BLAST on Beowulf Cluster; *Bioinformatics Applications Notes*, v.18, p.765-766, 2002; 1, 2, 4.1
- [17] H. Kim, D. Han; Hyper-BLAST: A Parallelized BLAST on Cluster System; *International Conference on Computational Science, Lecture Notes In Computer Science*, v.2569, p.213-222, 2003; 1, 4.2.1, 4.2.2
- [18] H. Kim, H. Kim, K. Han; Performance Evaluation of BLAST on SMP Machines; *The Fourth International Symposium on Parallel and Distributed Processing and Applications, Lecture Notes in Computer Science*, pp 668-676, 2006; 1
- [19] I. Korf, M. Yandell, J. Bedell; *An Essential Guide to the Basic Local Alignment Search Tool*; O' Reilly & Associates, Inc., 2003 2.1.1, 3.1, 3.3.1, 3.3.4, 5.1.4
- [20] H. Lin, X. Ma, P. Chandramohan, A. Geist, N. Samatova; Efficient data access for parallel BLAST; *International Parallel and Distributed Processing Symposium*, p.72-82, 2005; 2.1.2, 4.2.2
- [21] D. Mathog; Parallel BLAST on split database; *Bioinformatics*, v.19, p.1864-1866, 2003; 1, 2, 2.2, 3.2, 4.2.1

- [22] B. Miranda, A. Lima, P. Valduriez, M. Mattoso; Apuama: Combining Intra-query and Inter-query Parallelism in a Database Cluster; Currents Trends in Database Technology, Lecture Notes In Computer Science; p.649-661, 2006; 4.2.2
- [23] K. Muriki, K. Underwood, R. Sass; RC-BLAST: Towards a Portable, Cost-Effective Open Source Hardware Implementation; Proceeding of the 19th IEEE International Parallel and Distributed Processing Symposium, v.8, p.196-198, 2005; 4.2
- [24] M. Noronha; Implementação e avaliação de desempenho de um driver para gerência de E/S em aplicações de bioinformática; Dissertação de Mestrado, Departamento de Informática da PUC-Rio, 2006; 2.1.2, 4, 4.2.2, 7.3
- [25] C. Oehmen, J. Nieplocha; ScalaBLAST: A Scalable Implementation of BLAST for High-Performance Data-Intensive Bioinformatics Analysis; IEEE Transactions of Parallel and Distributed Systems, v.17, p.740-749, 2006; 1, 2, 2, 2.1.2, 4.2.2
- [26] M. Ozsu, P. Valduriez; Princípios de Sistemas de Bancos de Dados Distribuídos; 2nd Edition, Prentice-Hall, 1999; 1, 2, 4.2, 5.0.1
- [27] Message Passing Interface Forum, <http://www.mpi-forum.org/>; acessado em 07/07/2007;
- [28] K. Pedretti, T. Casavant, R. Braun, T. Scheetz, C. Birkett, C. Roberts; Three Complementary Approaches to Parallelization of Local BLAST Service on Workstation Clusters; Proceedings of the 5th International Conference on Parallel Computing Technologies, Lecture Notes In Computer Science, v.1662, p.271-282, 1999; 2, 2.2, 4.1, 4.2.1
- [29] Y. Qi, F. Lin; Parallelization of the BLAST algorithm; Cellular e Molecular Biology Letters, v.10, p.281-285, 2005; 4.2.2
- [30] M. Rieffel, T. Gill, W. White; PARACEL BLAST; <http://www.paracel.com>; acessado em 07/07/2007; 1, 4.1, 4.2.2
- [31] J. Rosa; Um Estudo de Compactação de Dados para Biosseqüências; Dissertação de Mestrado, Departamento de Informática, PUC-Rio, 2006; 2.1.2, 4
- [32] I. Rossi, D. Machignoli, D. Medini, R. Beltrami, C. Donati, P. Fariselli, A. Covacci, R. Casadio; A simple yet effective implementation of a parallel BLAST for computer clusters; Virtual Conference on Genomics and Bioinformatics Papers, v.1, p.10-14, 2002; 3.1, 4.2.1, 5.1.4

- [33] D. Sousa, S. Lifschitz; Avaliação do E-value para Execução do BLAST sobre Bases de Dados Fragmentadas; Série Monografias em Ciência da Computação MCC 17/07 Rio de Janeiro: Departamento de Informática, PUC-Rio, 2007; 2.2, 2.3, 5.1.4
- [34] H. Taki, G. Utard; MPI-IO on a Parallel File System for Cluster of Workstations; Proceedings of the 1st IEEE Computer Society International Workshop on Cluster Computing, p.150, 1999; 4.2.2
- [35] O. Trelles-Salazar, E. Zapata, J. Carazo; On a efficient parallelization of exhaustive sequence comparison algorithms on message passing architectures; Computer Applications in the Biosciences, v.10 n.5, p.509-511, 1994; 2, 2.1.1, 4.1
- [36] C. Wang, E. Lefkowitz; SS-Wrapper: a package of wrapper applications for similarity searches on Linux clusters; BMC Bioinformatics, v.5, p.171-180, 2004; 1, 4, 4.1
- [37] J. Wang, Q. Mu; SOAP-HT-BLAST: High throughput BLAST based on web services; Bioinformatics, v.19, p.1863-1864, 2003; 2, 4.1
- [38] WU-BLAST, <http://blast.wustl.edu/>; acessado em 07/07/2007;
- [39] Y. Zhu, H. Jiang, X. Quin, D. Swanson; A Case Study of Parallel I/O for Biological Sequence Search on Linux Clusters; International Conference on Cluster Computing, p.308-315, 2003; 4.2.2

A

Descrição do Agrupamento de Computadores e Programas utilizados

Nós temos utilizado um agrupamento homogêneo formado atualmente por 32 estações de trabalho com CPUs Intel Pentium 4, 249 Mb de RAM, 20 GB de disco local e relógio de 1715 Mhz. O sistema operacional instalado é o Red Hat Linux versão 9 *kernel* 2.4.20-31.9. Todas as estações de trabalho estão conectadas através de um switch Ethernet IBM8274.

O código das ferramentas sugeridas foram desenvolvidos em Linguagem C, compilado em GNU gcc 3.2.2. A interface paralela utilizada foi LAM-MPI (lam-7.0.6).

Em nossos testes foi utilizado a ferramenta NCBI BLAST, versão 2.2.26, para executar a comparação entre as seqüências. Nossas seqüências de consulta foram obtidas a partir do repositório de domínio público: http://racex00.tamu.edu/Amel_pre_release2_OGS_cds.fa, acessado em 06/07/2007.