

Referências Bibliográficas

- Achard, F.; Vaysseix, G.; Barillot, E.: XML, Bioinformatics and Data Integration. *Bioinformatics*, Vol. 17 (2), pp. 115-125, 2001.
- Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., and Lipman D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1;25(17):3389-402, 1997.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J.: Basic local alignment search tool. *Journal Molecular Biology*, 215(3), pp. 403-410, 1990.
- Andersson A. and Nilsson S.: Efficient implementation of suffix trees. *Software-Practice & Experience*. 25(2), pp. 129-141, 1995.
- Belleau F., Nolin M. A., Tourigny N., Rigault P. and Morissette J.: Bio2RDF: towards a mashup to build bioinformatics knowledge systems. *Journal Biomedical Informatics*. 41(5), pp.706-716, 2008.
- Benson, D.A.; Mizrachi, I.K.; Lipman, D.J.; Ostell, J. and Wheeler, D.L.: Genbank. *Nucleic Acids Research*, vol. 35, Database issue D21-D25, 2007.
- bioDBnet. bioBDnet: biological DataBase network. Disponível em: <<http://biodbnet.abcc.ncifcrf.gov>>. Acesso em Abril de 2012.
- Bioinformatics Factsheet - NCBI – NIH: Revised: March 29, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>>. Acesso em Dezembro de 2011.
- BLASTgres. BLASTgres: Biosequence data management in PostgreSQL. Disponível em: <<http://www.blastgres.org>>. Acesso em Outubro de 2010.
- Bork P., Koonin E. V.: Predicting functions from protein sequences--where are the bottlenecks? *Nat Genet*, 18(4):313-8, 1998.
- Brosch R., Pym A.S., Gordon S. V., and Cole S. T.: The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol*, 9(9):452-8, 2001.

- Catanho M.; Degrave W.; Miranda A. B.: Análise Comparativa de Genomas Procarióticos. *Biotecnologia Ciência & Desenvolvimento* (Online), v. 37, p. 20-29, 2007.
- Catanho M.; Mascarenhas D.; Degrave W.; Miranda A. B.: BioParser: a tool for processing of sequence similarity analysis reports. *Appl Bioinformatics*, 5(1), pp. 49–53, 2006.
- Chargaff E.: Some recent studies on the composition and structure of nucleic acids. *Journal of Cellular and Comparative Physiology. Supplement: Research Conference for Biology and Medicine of The Atomic Energy Commission*. 38(S1), pages 41-59, 1951.
- Chargaff E.: Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* 6 (6), pp. 201- 209, 1950.
- Cheung C. F., Yu J. X. and Lu H.: Constructing Suffix Tree for Gigabyte Sequences with Megabyte Memory. *IEEE Transactions on Knowledge and Data Engineering*. 17(1), pp. 90-105, 2005.
- Codd E. F.: A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*. 13(6), pp. 377-387, 1970.
- Costa, R.L.C., Lifschitz, S. Database allocation strategies for parallel blast evaluation on clusters. *Distributed and Parallel Databases*, 13(1), pp. 99--127, 2003.
- Crick, F. H. C.: On Protein Synthesis. *Symp. Soc. Exp. Biol. XII*, pp. 139-163, 1958.
- Crick, F. H. C.: Central Dogma of Molecular Biology. *Nature*. 227, pp. 561-563, 1970.
- Dada J. O., Spasic I., Paton N. W. and Mendes P.: SBRML: a markup language for associating systems biology data with models. *Bioinformatics*. 26(7), pp. 932-938, 2010.
- Dayhoff, M.O., Schwartz, R. and Orcutt, B.C.: A model of Evolutionary Change in Proteins. *Atlas of protein sequence and structure*. v.5, supplement 3 ed. Nat. Biomed. Res. Found. pp. 345-358, 1978.

- Dias, T. L.; Câmara, G.; Davis JR., C. A.: Modelos Espaço-Temporais. In: Marco Antônio Casanova; Gilberto Câmara; Clodoveu Augusto Davis Júnior; Lúbia Vinhas; Gilberto Ribeiro de Queiroz. (Org.). Bancos de Dados Geográficos. Curitiba (PR): EspaçoGeo, 2005, v. único, p. 147-180.
- Eilbeck K, Brass A, Paton N, Hodgman C.: INTERACT: an object oriented protein-protein interaction database. *Proc Int Conf Intell Syst Mol Biol.* pp. 87-94, 1999.
- Elmasri, R.; Navathe, S.: Sistemas de Banco de Dados: Fundamentos e Aplicações. 4. ed. Rio de Janeiro: Pearson Education, 2005.
- Eltabakh M. Y., Hon W.-K., Shah R., Aref W. and Vitter J. S.: The SBC-tree: An Index for Run-Length Compressed Sequences. *Proceedings of the 11th International Conference on Extending Database Technology (EDBT)*, pp. 523-534, 2008.
- Eltabakh M. Y., Ouzzani M., Aref W. G., Elmagarmid A. K., Silva Y. L., Arshad M. U., Salt D., Baxter I.: Managing Biological Data using bdbms. *Proceedings of 24th IEEE International Conference on Data Engineering (ICDE)*, pp 1600-1603, 2008.
- Eltabakh M. Y., Ouzzani M. and Aref W G.: bdbms: A Database Management System for Biological Data. *Proceedings of Third Biennial Conference on Innovative Data Systems Research (CIDR)*, 2007.
- Eltabakh M. Y., Eltarras R. H. and Aref W. G.: Space-partitioning trees in postgresql: Realization and performance. *Proceedings of the 22nd International Conference on Data Engineering (ICDE)*, IEEE Computer Society, pp. 100- 111, 2006.
- Entrez Gene - NCBI Entrez Gene, Disponível em: <<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>>. Acesso em Outubro de 2010.
- Ferragina P. and Grossi R.: The string B-tree: a new data structure for string search in external memory and its applications. *Journal of the ACM*, 46(2), pp. 236-280, 1999.
- Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR and Bateman A.: The Pfam protein families database. *Nucleic acids research*, 38(Database issue), pp. 211-222, 2010.

- Goble C. and Stevens R.: State of the nation in data integration for bioinformatics. *Journal of Biomedical Informatics*. 41(5), pp 687-693, 2008.
- Graves, M.; Bergeman, E. R.; Lawrence, C. B. A Graph-Theoretic Data Model for Genome Mapping Databases. Hawaii International Conference on System Sciences (HICSS) - 28, Biotechnology Computing Track. January, 1995.
- Henikoff, S.; Henikoff, J.G.: Amino Acid Substitution Matrices from Protein Blocks. *PNAS* 89 (22), pp. 10915-10919, 1992.
- Hsiao R., Parker D. S. and Yang H.: Support for BioIndexing in BLASTgres. *Lecture Notes in Computer Science - Data Integration in the Life Sciences*. Vol. 3615, pp. 284-287, 2005.
- Hunt, E., Atkinson, M. P. and Irving, R. W.: Database indexing for large dna and protein sequence collections. *The International Journal on Very Large Data Bases (VLDB)*, 11(3), pp. 256-271, 2002.
- Hunt E., Atkinson M. P. and Irving R.W.: A Database Index to Large Biological Sequences. *In Proceedings of the 27th International Conference on Very Large Data Bases (VLDB)*, pp. 139-148, 2001.
- Irving R.W. and Love L.: The Suffix Binary Search Tree and Suffix AVL Tree. Technical Report TR-2000-54, Univ. of Glasgow, Dept. of Computing Science, 2000.
- Jagadish H. V. and Olken F.: Database management for life sciences research. *ACM SIGMOD Record*, 33(2), pp. 15-20, 2004.
- Jagadish H. V., Koudas N. and Srivastava D.: On effective multi-dimensional indexing for strings. *ACM SIGMOD Record*, 29(2), pp. 403-414, 2000.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M.; KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Research*, 40(Database issue), pp. 109-114, 2012.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), pp. 27-30, 2000.
- Keet, C.M. Biological data and conceptual modelling methods. *Journal of Conceptual Modeling*, Issue 29, 2003.
- Köhler, J.: Integration of Life Science Databases. *DDT: BIOSILICO*, 2(2), pp. 61-69, 2004.

- Koonin, E. V.: Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*, pp. 309-38, 2005.
- Kröger, P. Molecular Biology Data: Database Overview, Modeling Issues, and Perspectives. Thesis Draft, INSTITUT FÜR INFORMATIK – Ludwig Maximilians Universität München (supervisor: Prof. Dr. François Bry and PD Dr. Rolf Backofen), May 2001.
- Lifschitz S.: Gerenciadores de Dados Biológicos: Genéricos ou Ad-hoc? *Anais: XXXIV Seminário Integrado de Software e Hardware (SEMISH)*, pp. 2085-2099, 2007.
- Lifschitz, S., Mauro, R.C., An i/o device driver for bioinformatics tools: the case for blast. *Genetics and Molecular Research (GMR)*, 4(1):563--570, 2005
- Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. *Science*, 22;227(4693):1435-41, 1985.
- Lodish, H., Berk, A., Zipursky, L. S., Matsudaira, P., Baltimore, D., Darnell, J. *Molecular Cell Biology*, 6th Edition. W H Freeman & Co., 2007.
- Luchtan M., Warade C., Weatherly D. B., Degrave W. M., Tarleton R. L. and Kissinger J. C.: TcruziDB: an integrated Trypanosoma cruzi genome resource. *Nucleic Acids Research*. Vol. 32, Database issue D344-D346, 2004.
- Luscombe N. M., Greenbaum D. and Gerstein M.: What is bioinformatics? An introduction and overview. *International Medical Informatics Association*, pp. 83-100, 2001.
- Macedo J.A.F de, Lifschitz S., Porto F., Picouet P., de Miranda A.B., Otto, T.D. Towards a Conceptual Modeling Language for Biological Domains, Brazilian Symposium in Bioinformatics (poster proceedings), pp 128-137, 2007.
- Macedo J.A.F de, Lifschitz S., Porto F., Picouet P.. Dealing with Some Conceptual Data Model Requirements for Biological Domains, International Symposium on Bioinformatics and Life Sciences Computing, pp 651-656, 2007.
- Markowitz, V. M.; Chen, I. A.; Kosky, A. S.; Szeto, E. Facilities for Exploring Molecular Biology Databases on the Web: A Comparative Study. 2nd Pacific Symposium on Biocomputing, Santa Fe, New Mexico, 1997.
- Manber U. and Myers G.: Suffix arrays: a new method for on-line string searches. *SIAM Journal of Computer*. 22(5), pp. 935-948, 1993.

- McDermott J, Samudrala R. The Bioverse: An object-oriented genomic database and webserver written in Python. In Proceedings of the conference on Objects in Bio- & Chem-Informatics, 2002
- Mudunuri U., Che A., Yi M. and Stephens R. M.: bioDBnet: the biological database network. *Bioinformatics*. 25(4), pp. 555-556, 2009.
- NAR: *Nucleic Acids Research*. Disponível em: < <http://nar.oxfordjournals.org/>>. Acesso em Fevereiro 2012.
- Navarro G., Sutinen E., Tanninen J. and Tarhio J.: Indexing Text with Approximate q-grams. *Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching*. Vol. 1848/2000, pp. 350-363, 2000.
- NCBI Annotation Information - Information on NCBI genome annotation methods, Disponível em: <<http://www.ncbi.nlm.nih.gov/genome/guide/build.shtml>>. Acesso em Janeiro de 2012.
- Neteler, M. and Mitásová, H.: Open Source GIS: A GRASS GIS Approach. Third Edition. *The International Series in Engineering and Computer Science*: Vol. 773, pp. 406, 2008.
- Noronha, M.F. 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 (orientação: Sérgio Lifschitz), Setembro 2006.
- Otto T. D., Catanho M., Tristão C., Bezerra M., Fernandes R. M., Elias G. S., Scaglia A. C., Bovermann B., Berstis V., Lifschitz S., Miranda A. B. de and Degraeve W.: ProteinWorldDB: querying radical pairwise alignments among protein sets from complete genomes. *Bioinformatics*. 26(5), pp 705-707, 2010.
- Pearson W.R.: Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. *Genomics*, 11(3), pp. 635-650, 1991.
- PostgreSQL - PostgreSQL 9.1.2 Documentation, Disponível em: <<http://www.postgresql.org/docs/9.1/>>. Acesso em Fevereiro de 2012.
- Puiu D., Enomoto S., Buck G. A., Abrahamsen M. S. and Kissinger J. C.: CryptoDB: the Cryptosporidium genome resource. *Nucleic Acids Research*, Vol. 32, Database issue D329-D331, 2004.

Punta M., Coggill P.C., Eberhardt R.Y., Mistry J., Tate J., Boursnell C., Pang N., Forslund K., Ceric G., Clements J., Heger A., Holm L., Sonnhammer E.L.L., Eddy S.R., Bateman A., Finn R.D.: The Pfam protein families database. *Nucleic Acids Research*, 40(Database Issue), pp.290-301, 2012.

RefSeq - NCBI Reference Sequence (RefSeq), Disponível em: <<http://www.ncbi.nlm.nih.gov/RefSeq/>>. Acesso em Junho de 2012.

Russo, C. A. M.: Como Escolher Genes para Problemas Filogenéticos Específicos?. In: Sergio Russo Matioli (Org.). *Genética Molecular e Evolução*. 2a ed. SAO PAULO: HOLOS, v. 1, p. 130-136, 2009.

Seibel, L.F.B, Macedo, J.A.F., Lemos, M., Lifschitz, S., Miranda, A.B., Alves, M., Degrave, W., A conceptual model for molecular biology information. II Workshop Brasileiro de Bioinformática (WOB), pp 47--56, 2003

Seibel, L. F. B. *Bio-AXS: Uma Arquitetura para Integração de Fontes de Dados e Aplicações de Biologia Molecular*. Tese de Doutorado. Departamento de Informática da PUC/RJ. 2000.

Seltzer, M.: Beyond relational databases. *ACM Communications*, 51(7), 2008.

Shin, D. Comparative Study of Relational and Object-Oriented Modelings of Genomic Data. Hawaii International Conference on System Sciences (HICSS), Hawaii, USA. Institute of Electrical and Electronics Engineers, Inc., 1995.

Shui, W. M., Wong R. K., Graham S. C., Lee L. K., Church W. B.: A New Approach to Protein Structure and Function Analysis Using Semi-Structured Databases. *Conferences in Research and Practice in Information Technology Series - Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics*. Vol. 19, pp. 61-69, 2003

Simonetto E. O. and Ruiz D.D.: A proposal model to incorporation of temporal aspects to the relational DBMS using active databases concept. *Proceedings of IEEE fourth international workshop on databases and information systems*. Vol. 1, pp. 52-59, 2000.

Smedley D., Swertz M. A., Wolstencroft K., Proctor G., Zouberakis M., Bard J., Hancock J. M. and Schofield P.: Solutions for data integration in functional genomics: a critical assessment and case study. *PubMed - Briefings in Bioinformatics*. 9(6), pp. 532-544, 2008.

Smith T.F., Waterman M.S.: Comparison of Biosequences. *Adv. Appl. Math.* 2, 482-9, 1981.

Souza, D. Estratégias de fragmentação mista para bancos de dados de seqüências genômicas, Dissertação de Mestrado, Departamento de Informática da PUC-Rio (orientação: Sérgio Lifschitz), Julho 2007.

Stein, L. D.: Integrating Biological Databases. *Nature Review Genetics*, 4(5), pp. 337-345, 2004

Stephens S. M., Chen J. Y., Davidson M. G., Thomas S. and Trute B M.: Oracle Database 10g: a platform for BLAST search and Regular Expression pattern matching in life sciences. *Nucleic Acids Research*, Vol. 33, Database issue D675-D679, 2005.

The Gene Ontology Consortium. The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Research*, 38(Database issue), pp. 331-335, 2010.

The Gene Ontology Consortium. The Gene Ontology project in 2008. *Nucleic Acids Research*, 36(Database issue), pp. 440-444, 2008.

The NCBI Handbook, Disponível em: <<http://www.ncbi.nlm.nih.gov/books/NBK21101/>>. Acesso em Janeiro de 2012.

Topaloglou T.: Biological data management: Research, practice and opportunities. *In Proceedings of International Conference on Very Large Data Bases (VLDB)*, pp. 1233-1236, 2004.

UniProt - UniProtKB, Protein knowledgebase. Disponível em: <<http://www.uniprot.org/>>. Acesso em Junho de 2012.

Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Kourtesis P, Ledley RS, Suzek BE, Vinayaka CR, Zhang J, Barker WC.: The Protein Information Resource. *Nucleic Acids Res.* Vol. 31(1), pp. 345-347, 2003.