Brazilian EMBnet Node: progress Report

Ana Tereza Vasconcelos[1], Wim M. Degrave[2], Goran Neshich[3]

[1]Laboratório Nacional de Computação, Científica Laboratório de Bioinformática, Quitandinha Petrópolis, Rio de Janeiro (Brazil)
[2]Oswaldo Cruz Institute (IOC), FIOCRUZ, Rio de Janeiro, Brazil
[3]Brazilian Agricultural Research Corporation (EMBRAPA) and UNICAMP's Department of Biology, Campinas, Brazil

Mission

The Brazilian EMBnet node conducts research and development in Bioinformatics and Computational Biology, with emphasis on creating and applying computational and mathematical methods and models for solving biological problems. The Brazilian node is formed by a network of three institutions: The National Laboratory for Scientific Computation (LNCC - Petropolis), the Oswaldo Cruz Foundation (Fiocruz – Rio de Janeiro) and Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA - Campinas). The network maintains and develops databases and tools in bioinformatics and computational biology to supply the needs of thematic networks and national and international collaborative projects, organizes training courses on several levels, and promotes technology and innovation.

National laboratory for scientific computation - LNCC, Laboratory for Bioinformatics

Dr. Ana Tereza Vasconcelos.

Infrastructure

Computational

The LNCC, one of the National Institutes of the Ministry of Science and Technology, has at present the following available computational resources: Sunfire 6800 with 24 processors and 24 Gb of memory, SunFire 3800 with 4 processors and 32 Gb of memory, SGI Challenger with 8 processors and 2 Gb of memory and 2 Sun Enterprise 450, offering a set of tools with modern technology that is up-to-date and ready for the development of applications that demand high levels of computational and scientific resources. The computational resources of the LNCC also include 90 Unix workstations (IBM, Silicon Graphics, Sun and Linux), 350 PCs and 100 printers. The external network of the LNCC is made of two links, one of 34 Mbps (megabits/second) to the POP-Rio de Janeiro of the RNP which is operated by the LNCC in its old headquarters and another of 2 Mbps to the REDERIO. The rate of use of the two links (Rio de Janeiro-Petrópolis) is of approximately 50%. Expansion of the links with the REDERIO aiming at the interconnection with the REMAV-Rio de Janeiro (High-Speed Metropolitan Network) is under study. Two communication servers for dialed access, each having 30 digital lines (total 60 lines) should also be at our disposal and will be located in Petrópolis and in Rio de Janeiro (POP-Rio de Janeiro).

The platform of the internal network of the LNCC is composed at present of 2 CISCO Catalyst switches, model 6509, interconnected to 4 Gbps, interconnecting two clusters of Catalyst switches, model XL-2909, with FEC connections of 800 Mbps in each cluster. The master switches of each cluster have 2 expansion slots available, besides several 10/100 Mbps ports reserved for expansions. In total, the LNCC has approximately 500 10/100 Mbps ports in the clusters of switches interconnecting the workstations of its technical/scientific staff. The cabling is certified and warranted by Lucent Tecnologies for a period of 15 years. The switch connections are made of fiber optics and the links from these to the rooms (stations) are made in category 7 twisted-pair cables.

Genomic

The Computational Genomics Unit Darcy Fontoura de Almeida is associated to the Laboratory of Bioinformatics of the National Laboratory of Scientific Computation – LNCC. This Unit, coordinated by Ana Tereza Ribeiro de Vasconcelos, has the purpose of integrating the activities of high-throughput DNA sequencing and bioinformatics into a single center, thus allowing for the best possible use of the data generated by the new 454 GS FLX sequencer of Roche. Inaugurated in September 19, 2008, the Computational Genomics Unit is a center of excellence of national reference in high-throughput sequencing. At present, the 454 GS FLX sequencer is the only one in South America that follows all the specifications of the Roche manufacturer. The laboratory can also count with an Agilent Bionalyzer 2100, a Nanodrop 3000 fluorometer, a Genomic Solutions HydroShear, a Qiagen TissueLyser, centrifuges, a Beckman Coulter Z1, Veriti thermocyclers and other support equipment.

Projects with financial support

2008 - Actual: Genômica Computacional e o Seqüenciamento Parcial do Genoma de Trypanosoma Cruzi

Financial support: Fundação Carlos Chagas Filho de Amparo à Pesq. do Estado do Rio de Janeiro-FAPERJ

2008 - Actual: apoio para a manutenção e instalação da unidade multiusuário de genômica computacional

Financial support: Fundação Carlos Chagas Filho de Amparo à Pesq. do Estado do Rio de Janeiro-FAPERJ

2008 - Actual: Rede Sul Americana e Iberoamericana de Bioinformática (Red SurAmericana e Iberoamericana de Bioinformatica)

Financial support: Nacional de Desenvolvimento Científico e Tecnológico-CNPq

2008 - Actual: Rede Nacional de Sequenciamento de DNA - Projeto Genoma Brasileiro: Determinação de Genomas Relevantes para a Saúde Humana

Financial support : Ministério da Ciência e Tecnologia-MCT, Ministério da Saúde-MS

2008 - Actual: Rede Brasileira de Pesquisas sobre o Câncer - RBPC

Financial support Ministério da Saúde-MS e Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq

2007 - Actual: Biotecnologia - Insumos para Genômica e Proteômica

Financial support: Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq

2007 - Actual: Prospecção de novos genes com potencial biotecnológico

Financial support: Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq

2006 - Actual: Estudo multicêntrico para caracterização molecular das hemofilias A e B e determinação do estado de portador de hemofilia no Brasil

Financial support : Ministério da Saúde-MS

2006 - 2008: Brazilian Microbiological Resource Center (BMRC)

Financial support: Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, Empresa Brasileira de Pesquisa Agropecuária-Centro Nac. de Pesq. de Soja-EMBRAPA SOJA

2006 - Actual: CTpedia database

Financial support: Ludwig Institute for Câncer research

2004 - Actual: HAMAP BRAZIL - PAthogenic Proteins Annotation Project

Financial support: Swiss Institute for Bioinformatics

2004 - 2008: Projeto Genomica comparativa de Xylella fastidiosa

Financial support : Ministério da Ciência e Tecnologia-MCT, Universidade de São Paulo-USP

2004 - Actual: Fixadores de Nitrogenio

Financial support: Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq, Empresa Brasileira de Pesquisa Agropecuária-Centro Nac. de Pesq. de Soja-EMBRAPA SOJA

Courses

- Genômica funcional de microrganismos patogênicos, 2009.
- Genômica e Bioinformática, 2008.
- Bioinformática I - Banco de dados do ponto de vista biológico , 2007.
- Tópicos Especiais em Genética II -Genômica Comparativa , 2007.
- Análise e Comparação de Genomas - Procariotos , 2006.
- Bioinformática I - Banco de Dados do Ponto de Vista Biológico , 2006.

Publications

1. PINEROGONZALEZ, J, CARRILLOFARNES, O, VASCONCELOS, A, GONZALEZPEREZ, A, VASCONCELOS, A. T. R. Conservation of key members in the course of the evolution of the insulin signaling pathway. Biosystems. , v.95, p.7 - 16, 2009.
2. Almeida, L. G., Sakabe, N. J., deOliveira, A. R., SILVA, M. C. C., MUNDSTEIN, A. S., Cohen, T., Chen, Y.-T., Chua, R., Gurung, S., Gnjatic, S., Jungbluth, A. A., Caballero, O. L., Bairoch, A., Kiesler, E., White, S. L., Simpson, A. J. G., Old, L. J., Camargo, A. A., VASCONCELOS, A. T. R. CTdatabase: a knowledgebase of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Research. , v.37, p.D816 - D819, 2009.

3. VEIGA, D. F., DEUS, H. F., C A, VASCONCELOS, A. T. R., ALMEIDA, J. S.DASMiner: discovering and integrating data from DAS sources. BMC Systems Biology. , v.1, p.1 - , 2009.
4. Gonzalez Perez, Abel, Espinosa Angarica, Vladimir, Collado-Vides, Julio, VASCONCELOS, A. T. R.From sequence to dynamics: the effects of transcription factor and polymerase concentration changes on activated and repressed promoters. BMC Molecular Biology. , v.10, p.92 - , 2009.
5. Cristiane C Thompson, Ana Carolina P Vicente, Rangel Souza, VASCONCELOS, A. T. R., Tammi Vesth, Nelson Alves Jr,, Tetsuya Iida, Fabiano L. Thompson Genomic taxonomy of vibrios. BMC Evolutionary Biology (Online). , v.1, p.1 - 10, 2009.
6. Pinto, Fabiana G. S., Chueire, Ligia M. O., Vasconcelos, Ana Tereza R., Nicolás, Marisa F., Almeida, Luiz G. P., Souza, Rangel C., Menna, Pâmela, Barcellos, Fernando G., Megías, Manuel, HUNGRIA, Mariângela Novel genes related to nodulation, secretion systems, and surface structures revealed by a genome draft of Rhizobium tropici strain PRF 81. Functional & Integrative Genomics. , p.1 - 8, 2009.
7. BARRETO, K. S., TORRES, A. R., BARRETO, M. R., VASCONCELOS, A. T. R., ASTOLFO-FILHO, Spartaco, HUNGRIA, Mariângela Diversity in antifungal activity of strains of Chromobacterium violaceum from the Brazilian Amazon.. Journal of Industrial Microbiology and Biotechnology. , v.1, p.10.1007/s10295 - , 2008.
8. Freire, Pablo, Vilela, Marco, Deus, Helena, Kim, Yong-Wan, Koul, Dimpy, Colman, Howard, Aldape, Kenneth D., Bogler, Oliver, Yung, W. K. Alfred, Coombes, Kevin, Mills, Gordon B., Vasconcelos, Ana T., Califano, Andrea, VASCONCELOS, A. T. R., Almeida, Jonas S. Exploratory Analysis of the Copy Number Alterations in Glioblastoma Multiforme. plos one. , v.3, p.e4076 - , 2008.
9. GODOY, L, VASCONCELOS, A. T. R., CHUEIRE, L, SOUZA, R, NICOLAS, M, BARCELLOS, F, HUNGRIA, M Genomic panorama of Bradyrhizobium japonicum CPAC 15, a commercial inoculant strain largely established in Brazilian soils and belonging to the same serogroup as USDA 123. Soil Biology & Biochemistry. , p.1 - 11, 2008.
10. GONZALEZ, A., Gonzalez-Gonzalez E., ESPINOSA, V., VASCONCELOS, A. T. R., COLLADO-VIDES, J. Impact of Transcription Units rearrangement on the evolution of the regulatory network of gamma-proteobacteria. BMC Genomics. , v.9, p.1 - 18, 2008.
11. de Mello Varani, Alessandro, SOUZA, Rangel Celso, Nakaya, Helder I., de Lima, Wanessa Cristina, Paula de Almeida, Luiz Gonzaga, Kitajima, Elliot Watanabe, Chen, Jianchi, Civerolo, Edwin, Vasconcelos, Ana Tereza Ribeiro, Van Sluys, Marie-Anne Origins of the Xylella fastidiosa Prophage-Like Regions and Their Impact in Genome Differentiation. plos one. , v.3, p.e4059 - , 2008.
12. VILELA, M. A., CHOU, I., VINGA, S., VASCONCELOS, A. T. R., Eberhard O. Voit, ALMEIDA, J. S. Parameter optimization in S-system models. BMC systems biology. , v.2, p.1752-0509-2-35 - , 2008.
13. Veiga, Diogo FT, Vicente, Fábio FR, Nicolás, Marisa F, Vasconcelos, Ana Tereza R Predicting transcriptional regulatory interactions with artificial neural networks applied to E. coli multidrug resistance efflux pumps. BMC Microbiology (Online). , v.8, p.101 - , 2008.
14. Espinosa Angarica, Vladimir, Gonzalez Perez, Abel, Vasconcelos, Ana T, Collado-Vides, Julio, Contreras-Moreira, Bruno, VASCONCELOS, A. T. R.Prediction of TF target sites based on atomistic models of protein-DNA complexes. BMC Bioinformatics. , v.9, p.436 - , 2008.
15. Bernardes, Juliana S, Fernandez, Jorge H, Vasconcelos, Ana Tereza R Structural descriptor database: a new tool for sequence based functional site prediction. BMC Bioinformatics. , v.9, p.492 - , 2008.
16. GONZALEZ, J. P., FARNES, O. C., VASCONCELOS, A. T. R., GONZALEZ, A.The impact of the emergence of IRS molecules on the evolution of the insulin signaling pathway. Biosystems. , v.prelo, p.01 - 10, 2008.
17. VILELA, M. A. M., BORGES, Carlos Cristiano, VINGA, S., VASCONCELOS, A. T. R., Santos Helena, Eberhard O. Voit, ALMEIDA, J. S. Automated Smoother for the Numerical Decoupling of Dynamics Models. BMC Bioinformatics. , v.8, p.1 - 8, 2007.
18. ALarCON, F., VASCONCELOS, A. T. R., lucia Yim, ZAHA, Arnaldo Genes involved in cell division in mycoplasmas. Genetics and Molecular Biology. , v.30, p.174 - 181, 2007.
19. VEIGA, D. F., VICENTE, F.F.R., FUENTE A, L., MAIA, M. A. G. M., VASCONCELOS, A. T. R. Genome-wide partialcorrelation analysis of Escherichia coli Microarray Data. Genetics and Molecular Research. , v.6, p.730 - 742, 2007.
20. BAREINBOIM, Elias, VASCONCELOS, A. T. R., SILVA, Jôao Carlos Pereira da Grammatical inference applied to linguistic modeling of biological. RECIIS. , v.2, p.329 - 333, 2007.
21. Souza, Rangel Celso, ALMEIDA, Darcy Fontoura de, ZAHA, Arnaldo, Morais, David Anderson de Lima, Vasconcelos, Ana Tereza Ribeiro de In search of essentiality: Mollicute-specific genes shared by twelve genomes. Genetics and Molecular Biology. , v.30, p.169 - 173, 2007.
22. THOMPSON, F., Bruno Gomez-Gil, VASCONCELOS, A. T. R., Tomoo Sawabe Multilocus sequence analysis reveals that Vibrio harveyi and V. camp-

bellii form distinct species.. Applied Enviromental Microbiology. , v.13, p.4279 - 4285, 2007.

23. Brocchi, Marcelo, Vasconcelos, Ana Tereza Ribeiro de, ZAHA, Arnaldo Restriction-modification systems in Mycoplasma spp. Genetics and Molecular Biology. , v.30, p.236 - 244, 2007.
24. GONZALEZ, A., ESPINOSA, V., VASCONCELOS, A. T. R., COLLADO-VIDES, J. Tractor _ DB (version 2.0): a database of regulatory interactions in. Nucleic Acids Research. , v.35, p.D132 - D136., 2007.

Books chapters

- ALMEIDA, Luiz Gonzaga Paula, de Vasconcelos, A. T. R., MAIA, M. A. G. M. A Simple and Fast Term Selection Procedure for Text Clustering In: Intelligent Text Categorization and Clustering ed.Heidelberg : Springer Berlin /, 2008, v.164, p. 47-64.
- WANDERLEY, M. F. B., SILVA, Jõao Carlos Pereira da, BORGES, Carlos Cristiano, de Vasconcelos, A. T. R. Application of Genetic Algorithms to the Genetic Regulation Problem In: Advances in Bioinformatics and Computational Biology ed.Heidelberg : Springer Berlin /, 2008, v.5167, p. 140-151.
- VARANI, A. M., LIMA, Wanessa Cristima de, MOREIRA, L., OLIVEIRA, M., SOUZA, Rangel Celso, CIVEROLO, E., VASCONCELOS, A. T. R., SLUYS, Marie Anne Van. Common Genes and Genomic Breaks: A Detailed Case Study of the Xylella fastidiosa Genome Backbone and Evolutionary Insights In: Plant Pathogenic Bacteria: Genomics and Molecular Biology ed.Reading : Inglaterra, 2008

Services

- EMBnet node
- Expasy Mirror
- CTdatabase
- Brazilian Microbiological Resource
- Mamibase
- Tractor DB
- Structural Descriptor DataBase
- SABIA – Software for automatic Bacterial Anottation

Fundação oswaldo cruz – FIOCRUZ, Platform for Bioinformatics, and Laboratory for Functional Genomics and Bioinformatics

Dr. Wim Degrave

The EMBnet node activities at Fiocruz are assured by the institutional Bioinformatics Platform, with support from the VPPLR-PDTIS program-RPT4A and by the IOC - Functional Genomics and Bioinformatics Unit and support from the Program for Scientific Computing, and the Fiocruz Network.

Team members:

- Wim M. Degrave
- Antonio Basilio de Miranda
- Thomas Dan Otto (currently at the Sanger Institute, UK)
- Fábio F. Mota - Technologist
- Mark Catanho - PhD student in Cellular and Molecular Biology
- Ana Carolina Guimarães - PhD student in Cellular and Molecular Biology
- Flávio Engelke - Master student in Biomedical Sciences

Activities

Bioinformatics services; support for genomics and proteomics platforms at Fiocruz, genome sequencing projects, software and application development; installation and upgrading of software; construction, implementation and updating of databases; design and maintenance of information services; organization of training courses and on-line training, research projects in comparative genomics, evolutionary biology and genome wide metabolic analysis, drug development in neglected diseases.
The node aims to:

- provide the environment and support in bioinformatics (biological data processing, access to genetic databases, creating and maintaining databases for proteomic analysis) and support for special applications such as molecular modeling, assembly and genome analysis, support for proteomics,
- organize hands-on training courses and on-line training to users, mostly within graduate and postgraduate programs;
- contribute to specific research projects through software and database development;
- disseminate bioinformatics as a tool and as a research and development discipline. The Bioinformatics node contributes to improvement of public health and the development of new technologies and tools;
- generate a potential economic impact, because it contributes to the patentability in research projects and innovations, and has captured external resources for this purpose.

Infrastructure

The main infrastructure of the unit is currently comprised of a dozen of smaller dedicated servers. Two larger servers are to be included in 2010. Fiocruz has an extensive network of fiber optics, linking several thousands of PCs in the different Institutes that comprise the Foundation, and is connected to the RNP and REDERIO through high speed links. Fiocruz counts with several additional bioinformatics groups performing research and development in fields such as genomics, statistics and epidemiology, molecular modeling, georeferencing, systems biology etc., and counts with post-graduate courses in Computational and Systems Biology.

Special Services offered:

- Bioinformatics databases and applications
- Genome assembly
- Web servers
- General sequence analysis
- Proteome analysis
- Data processing

The most common software packages for sequence assembly and database are available.

Products developed by the team of the platform:

- BioParser * - Analyzer/parser for all varieties of BLAST and FASTA, with support for versions of BLAST with and without gaps.
- SQUID * - Friendly local grid environment for the use of BLAST and FASTA programs.
- GenoMycDB - Database for information related to the genome and proteome of mycobacteria
- REReP - Method to facilitate the assembly of genomes, based on the detection and filtering of seqs. repetitive (applicable to data obtained by the method of Sanger and probably pirosequenciamento)
- AnEnPi - Tool for clustering, similarity search, identification of cases of functional analogy and reconstruction of metabolic pathways.
- ProteinWorldDB - Database indexes of similarity between protein sequences of hundreds of genomes – http://www.proteinworlddb.org

Courses

- Computational analysis of sequence and protein (IOC 26051)
- Origin, Structure and Evolution of prokaryotic genomes (IOC 26052)

Recent new collaborations

- Analysis of the genome of Streptococcus pneumoniae, in collaboration with BioManguinhos (Dr. Marco Medeiros)
- Development of a multiplex PCR for distinguishing species of the genus Wolbachia, Ehrlichia, Rickettsia and Anaplasma, in collaboration with Dr. Agnes Rossi (top Mar/2009 - Ready for testing on bench)
- Analysis of genes of Vibrio mimicus, in collaboration with Dr. Ana Carolina Vicente (top Sep/2009)

Publications

1. Buschiazzo A, Goytia M, Schaeffer F, Degrave W, Shepard W, Grégoire C,Chamond N, Cosson A, Berneman A, Coatnoan N, Alzari PM, Minoprio P. Crystal structure, catalytic mechanism, and mitogenic properties of Trypanosoma cruzi proline racemase. Proc Natl Acad Sci U S A. 2006 Feb 7;103(6):1705-10.
2. Carvalho PC, Fischer JS, Chen EI, Domont GB, Carvalho MG, Degrave WM, Yates JR3rd, Barbosa VC. GO Explorer: A gene-ontology tool to aid in the interpretation of shotgun proteomics data. Proteome Sci. 2009 Feb 24;7:6.
3. Carvalho PC, Carvalho Mda G, Degrave W, Lilla S, De Nucci G, Fonseca R, Spector N, Musacchio J, Domont GB. Differential protein expression patterns obtained by mass spectrometry can aid in the diagnosis of Hodgkin's disease. J Exp Ther Oncol. 2007;6(2):137-45.
4. Carvalho PC, Freitas SS, Lima AB, Barros M, Bittencourt I, Degrave W, Cordovil I, Fonseca R, Carvalho MG, Moura Neto RS, Cabello PH. Personalized diagnosis by cached solutions with hypertension as a study model. Genet Mol Res. 2006 Dec 18;5(4):856-67.
5. Catanho M, Mascarenhas D, Degrave W, de Miranda AB (2006). "BioParser: a tool for processing of sequence similarity analysis reports". Appl Bioinformatics. 5(1):49-53.
6. Catanho M, Mascarenhas D, Degrave W, Miranda AB. (2006). "GenoMycDB: a database for comparative analysis of mycobacterial genes and genomes". Genet Mol Res. 2006 Mar 31;5(1):115-26.
7. Guimarães AC, Otto TD, Alves-Ferreira M, Miranda AB, Degrave WM. In silico reconstruction of the amino acid metabolic pathways of Trypanosoma cruzi. Genet Mol Res. 2008 Sep 23;7(3):872-82.
8. Noronha MF, Lifschitz S, de Miranda AB. (2008). "A Practical Evaluation of BioProvider". Lecture Notes in Computer Science 5167, p. 174-177.
9. Otto TD, Gomes LH, Alves-Ferreira M , Miranda AB, Degrave WM. (2008). "ReRep: Computational detection of repetitive sequences in genome survey sequences (GSS)". BMC Bioinformatics, 9:366. doi:10.1186/1471-2105-9-366.
10. Otto TD, Guimarães AC, Degrave WM, de Miranda AB. (2008a). "AnEnPi: identification and annotation of analogous enzymes". BMC Bioinformatics Dec 17;9:544.
11. Otto TD, Vasconcellos EA, Gomes LH, Moreira AS, Degrave WM, Mendonça-Lima L, Alves-Ferreira

M. (2008b). "ChromaPipe: a pipeline for analysis, quality control and management for a DNA sequencing facility". Genet Mol Res. Sep 23;7(3):861-71.

The empresa brasileira de pesquisa agropecuária-EMPRAPA, Laboratory for Computational Biology

Dr. Goran Neshich

Embrapa, through its laboratory for Computational Biology, has a long record of offering services to academic partners through the internet using its experience and knowledge in maintaining its own product STING. It mirrors also public databases such as PDB, Uniprot, Prosite, HSSP, DSSP, ProTherm etc. while maintaining STING mirrors at 5 continents. Embrapa´s activities include intensive service, education and development and involves students as well as experienced colleagues both from Brazil and from Latin America.

Infrastructure

The lab is located in an environment with plenty of space for students, researchers, computer space, dedicated server space, dedicated space for training. Currently, the hardware infrastructure is going through an extensive renewal and new machines are being installed, replacing old (2001) acquired servers and PC stations. Expanded storage space is being acquired to aid in the ever growing problem of expanding disc space for DB and their back-ups. We have SUN, SGI and Dell clusters, totaling at about 60 CPUs, while total storage space around 15 Tb in separate servers. Around 15 Linux/Windows dual system PCs are dedicated for student work and other 20 are dedicated for training only. Due to infrastructure updating, lab reconstruction and team renewal, the lab stopped temporarily to offer general EMBnet services until all pending issues are resolved, in order to provide for better services.

Resources

We are restoring and expanding at the same time the STING and its database, transforming it into a federative contribution platform. We would like to offer to the EMBnet not only the new STING but also our experience in upgrading it, maintaining it, mirroring it and using it for educational purposes.

Education

We run a course for two major universities and their program for bioinformatics – UFMG and Unicamp. Both are well attended and teach mostly structural computational biology, but also some tools and DBs from sequence – one dimensional world.

During the last three years we offered total of 3 courses for more than 50 students, mainly covering material from structural computational biology and structural bioinformatics.

Database construction

We constructed a first Latin American database that was registered in the NAR DB issue. Since then, we aggregated many parameters into that same STING _ DB, making it the largest of its kind available for access over the web. Currently this database contains more than 28.5•10^9 registers (61,000 PDB files, ~130,000 chains, ~300AA/chain, 731 descriptors/AA).

Software development

We also published and posted on the Web STING suite of software programs for educational and analytical purposes. Analysis STING protocols are designed for routine use and can generate advanced reports about structure, sequence, function, stability and binding of proteins and their ligands.

Projects

- Study of Macromolecular Communication in Homo and Hetero complexes through their interfaces Unicamp-IB+Embrapa/CNPTIA. Large scale protein function prediction tools" Genoscope, France + mbrapa/CNPTIA
- "TargetsDB - Base de dados de alvos terapêuticos validados" UFMG + Embrapa/CNPTIA
- Automatic prediction of protein-protein interfaces based on a novel hydrophobicity index studies Unicamp-IB + Embrapa/CNPTIA
- Free Bioinformatics Technology consolidation and application in Biomedicine (FreeBIT) Red Iberoamericana de Bioinformatica + Embrapa/CNPTIA
- "Druggable proteins: Identification of potential therapeutic targets for development of agrochemicals, veterinary and medical drugs and vaccines for treating plant and animal diseases important for agriculture and live stock" UFMG+USP+UNICAMP+UNIFEI+EMBRAPA. GenoProtPlus SUN Computers e EMBRAPA

- Molecular modeling and structural analysis of the protein twitching motility a product of XF1633 gene of Xylella fastidiosa. EMBRAPA - CNPTIA

Publications

1. Jorge H. Fernandez, Marcia O. Mello, Leticia Galgaro, Aparecida S. Tanaka, Marcio C. Silva-Filho, Goran Neshich "PROTEINASE INHIBITION USING SMALL BOWMAN-BIRK TYPE STRUCTURES." Genet. Mol. Res. 6 (4): 846-858 (2007)
2. R.C. Melo, C. Ribeiro, C.S. Murray, C.J.M. Veloso, C.H.. da Silveira, Goran Neshich, W. Meira Jr, R.L. Carceroni and M.M. Santoro:"Finding protein-protein interactions by contact-maps matching."Genet. Mol. Res. 6 (4): 946-963 (2007)
3. Walter Rocchia and Goran Neshich:"Electrostatic Potential Calculation for biomolecules - creating a database of pre-calculated values reported on a per residue basis for all PDB protein structures."Genet. Mol. Res. 6 (4): 923-936 (2007)
4. Stanley R. M. Oliveira, Gustavo V. Almeida, Kassyus R. R. Souza, Diego N. Rodrigues, Paula R. Kuser-Falcão, Michel E. B.Yamagishi, Edgard H. Santos, Fábio D. Vieira, José G. Jardine and Goran Neshich:"STING _ RDB: A relational database of structural parameters for protein analysis with support for Data Warehousing and Data Mining."Genet. Mol. Res. 6 (4): 911-922 (2007)
5. NESHICH, G:"Computational Biology in Brazil."PLoS Comput Biol. Oct; v3 (issue 10, e185), 2007.
6. MOUTRAN, Alexandre; BALAN, Andrea; PEREZ, Carolina Santacruz; FERREIRA, Rita Café; RAMOS, Carlos; FERREIRA, Luís Carlos Souza; NESHICH, Goran: "Crystallographic structure and substrate-binding interactions of the 3 molybdate-binding protein of the phytopathogen 4 Xanthomonas axonopodis pv. citri."
7. R. C. Togawa, C. Ribeiro, I. Mazoni, T. Pelligrinelli, and NESHICH, Goran: " The Table of Interface Forming Residues as the Specificity Indicator for Serine Proteases Bound to Different Inhibitors "Accepted: BIOCOMP 08, 2008.
8. Carlos H. da Silveira, Douglas E. V. Pires , Raquel C. Melo , Cristina Ribeiro , Caio J. M. Veloso , Julio C. D. Lopes , Wagner Meira Jr, Goran Neshich , Carlos H. I. Ramos , Raul Habesch, Marcelo M. Santoro: " Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. PROTEINS: Structure, Function, and Bioinformatics,2008 Aug 14;74(3):727-743. 2008
9. Janaina Gomide ; Raquel Melo Minardi ; Marcos Augusto dos Santos ; Wagner Meira Jr. ; Julio Cesar Dias Lopes ; SANTORO, Marcelo; NESHICH, Goran:" Using Linear Algebra for Protein Structural Comparison and Classification."Genet. Mol. Res., 2009.

Chilean EMBnet node: progress report

J. Cristian Salgado
Universidad de Chile, DCC - Escuela de Ingeniería, Dept. of Computer Science, Santiago, Chile

In the past three years we have held a large number of activities related to bioinformatics and Systems Biology, which have had a big impact on the local scientific community. We have organized two mathematical modeling courses (1-2 weeks) aimed at the national biological research community.;

- the first one, Mathematical Modelling of Biological Systems, was offered in 2008 by local instructors with a solid background on biology, mathematics and computer science and was attended by over 70 participants;
- the second was an advanced course on Mathematical Models in Biology, was given by Benoit Perthame from Laboratorie J.-L. Lions, UPMC/INRIA, and it was attended by over 20 participants;
- in addition, we are planning on extending these outreach activities and offer a regional Latin American Systems Biology course next year (2010).

During this period, we have been awarded a major national grant (1 M$ USD/yr) for the creation of the Institute for Cell Dynamics and Biotechnology: A Centre for Systems Biology (http://www.icdb.uchile.cl/icdb). This institute is composed by scientists whose background and area of expertise is very diverse. There has been a tremendous amount of cross-fertilization between the scientists participating in this institute, leading to a large number of collaborations, such as: