

Bioinformática: Formação Acadêmica e Plataformas com Softwares e Ferramentas



ISSN 1518-4277
Dezembro, 2010

*Empresa Brasileira de Pesquisa Agropecuária
Centro Nacional de Pesquisa de Milho e Sorgo
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 109

Bioinformática: Formação Acadêmica e Plataformas com Softwares e Ferramentas

Roberto Willians Noda

Embrapa Milho e Sorgo
Sete Lagoas, MG
2010

Exemplares desta publicação podem ser adquiridos na:

Embrapa Milho e Sorgo

Rod. MG 424 Km 45

Caixa Postal 151

CEP 35701-970 Sete Lagoas, MG

Fone: (31) 3027-1100

Fax: (31) 3027-1188

Home page: www.cnpms.embrapa.br

E-mail: sac@cnpms.embrapa.br

Comitê de Publicações da Unidade

Presidente: Antônio Carlos de Oliveira

Secretário-Executivo: Elena Charlotte Landau

Membros: Flávio Dessaune Tardin, Eliane Aparecida Gomes, Paulo

Afonso Viana, João Herbert Moreira Viana, Guilherme Ferreira

Viana e Rosângela Lacerda de Castro

Supervisão editorial: Adriana Noce

Revisão de texto: Antonio Claudio da Silva Barros

Normalização bibliográfica: Rosângela Lacerda de Castro

Tratamento de ilustrações: Alexandre Esteves Neves

Editoração eletrônica: Alexandre Esteves Neves

Foto(s) da capa: Maria Luiza Noda

1ª edição

1ª impressão (2010): on line

Todos os direitos reservados

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Milho e Sorgo

Noda, Roberto Willians.

Bioinformática: formação acadêmica e plataformas com softwares e ferramentas / Roberto Willians Noda. -- Sete Lagoas: Embrapa Milho e Sorgo, 2010.

20 p. -- (Documentos / Embrapa Milho e Sorgo, ISSN 1518-4277; 109).

1. Computação. 2. Tecnologia da informação. 3. Biologia computacional. 4. Gerenciamento. I. Título. II. Série.

CDD 005 (21. ed.)

© Embrapa 2010

Autor

Roberto Willians Noda

Biólogo, Ph.D. em Agronomia (Genética e Melhoramento de Plantas), Pesquisador da Embrapa Milho e Sorgo, Sete Lagoas - MG, roberto.noda@cnpms.embrapa.br

Sumário

Introdução	7
Formação em bioinformática	8
Cursos de curta duração	9
Graduações em bioinformática e informática biomédica	9
Programas de pós-graduação em bioinformática e alguns programas que possuem área de pesquisa em bioinformática	10
Plataformas com softwares e ferramentas de bioinformática	10
Projetos disponibilizados como Live CDs/DVDs	12
VLinux	12
BioKnoppix	12
Vigyaan	12
Quantian	13
DNALinux	13
BioSLAX	13
Bio-Linux	13
Sistemas completos	14
BioBrew	14
Bio-Linux	14
Pacotes e repositórios	14
Repositório RedHat/CentOS/Fedora e derivados (RPM)	15
Repositório Debian (APT)	15
Conclusões	15
Referências	16

Bioinformática: Formação Acadêmica e Plataformas com Softwares e Ferramentas

Roberto Willians Noda

Introdução

Há muitas definições para [Bioinformática](#)¹; o Programa Interunidades de Pós-Graduação em Bioinformática da Universidade de São Paulo (USP) (<http://www.ime.usp.br/posbioinfo>), descreve-a como o estudo da aplicação de técnicas computacionais e matemáticas à geração e ao gerenciamento de informação pela/para Biologia e/ou [Biotecnologia](#).

Para o Programa de Doutorado em Bioinformática da Universidade Federal de Minas Gerais (UFMG) (<http://www.pgbioinfo.icb.ufmg.br>), a Bioinformática é uma área estratégica no desenvolvimento de projetos multidisciplinares envolvendo as Ciências Biológicas e da Computação e para a formação de profissionais que possam contribuir com a inclusão do País na fronteira moderna da pesquisa mundial. Bioinformática contempla áreas tão distintas quanto a análise de sequências de [DNA](#) e [proteínas](#), o estudo de estruturas tridimensionais de moléculas, a visualização e a análise de imagens e sinais biológicos.

Na Universidade Federal do Paraná (FPR), o Programa de Pós-Graduação em Bioinformática (<http://www.bioinfo.ufpr.br/mod/resource/view.php?id=113>) caracteriza a Bioinformática como a aplicação da tecnologia de informação

¹Para facilitar a compreensão do texto, vários termos terão links com explicações sobre eles na [Wikipedia](#).

ao gerenciamento de dados biológicos, sendo essencialmente uma ciência biológica. Os especialistas em Bioinformática são os criadores das ferramentas, e é fundamental que entendam os problemas biológicos tanto quanto as soluções computacionais. A pesquisa em Bioinformática e [Biologia Computacional](#) pode incluir desde a abstração de um sistema biológico em um modelo matemático, o desenvolvimento de bancos de dados e ferramentas Web para acesso a dados até a implementação de novos algoritmos de análise.

A Associação Brasileira de Bioinformática e Biologia Computacional (2009) (<http://lgmb.fmrp.usp.br/ab3c>), que é uma sociedade científica dedicada ao avanço do entendimento de seres vivos por métodos formais, acredita que o estudo sistemático de fenômenos biológicos através de métodos multidisciplinares quantitativos de larga escala é uma revolução científica, que, muito provavelmente, terá um enorme impacto econômico durante o século 21.

Os termos Bioinformática e Biologia Computacional são considerados sinônimos por alguns autores, mas há definições que determinam a Bioinformática como sendo o ramo da Biologia Computacional responsável pelas análises dos dados de sequências de genomas, transcritomas e proteomas, fazendo com que o termo Biologia Computacional seja mais amplo, e sendo definido como qualquer aplicação da computação na área biológica.

Formação em bioinformática

A seguir, apresentamos *links* para cursos de curta duração com oferecimento regular e cursos de graduação em Bioinformática ou áreas correlatas. Há, também, no Quadro 1, três dos Programas de Pós-Graduação em Bioinformática existentes no Brasil, três dos cursos que contêm áreas de pesquisa em Bioinformática e informações sobre Programas de Pós-Graduação Recomendados pela Capes, constando a Grande Área, a Área, o Programa (com *link* para mais informações sobre cada Programa no site da Capes), a Instituição de Ensino Supe-

rior, a Unidade Federativa e o Conceito na avaliação da Capes para os níveis de Mestrado e Doutorado, quando presentes. Lembrando que, para ser recomendado, o Programa deve receber nota maior ou igual a 3 e as notas variam de 1 a 7.

Cursos de curta duração

Curso de Verão em Bioinformática

<http://lgmb.fmrp.usp.br/cvbioinfo>

Departamento de Genética – Faculdade de Medicina de Ribeirão Preto (FMRP)
Universidade de São Paulo (USP)

Curso de Verão em Bioinformática

<http://www.ime.usp.br/posbioinfo>

Programa Interunidades de Pós-Graduação em Bioinformática
Universidade de São Paulo (USP)

Curso de Inverno de Bioinformática

<http://gbi.fmrp.usp.br/cursodeinverno>

Grupo de Bioinformática - Faculdade de Medicina de Ribeirão Preto (FMRP)
Universidade de São Paulo (USP)

Graduações em bioinformática e informática médica

Bioinformática

http://www.cruzeirosul.edu.br/content/graduacao_detalhe.aspx?curso=56

Universidade Cruzeiro do Sul

Informática Biomédica

http://www.fmrp.usp.br/cg/novo/index.php?option=com_content&view=article&id=115&Itemid=91

Faculdade de Medicina de Ribeirão Preto (FMRP) - Universidade de São Paulo (USP)

Programas de pós-graduação em bioinformática e alguns programas que possuem área de pesquisa em bioinformática

Quadro 1. Principais programas de Pós-graduação em Bioinformática ou contendo linhas de pesquisa em Bioinformática, no Brasil.

Grande área	Área	Programa*	IES	UF	Conceito	
					M	D
Ciências Biológicas	Genética	Bioinformática	UFMG	MG	-	5
Ciências Biológicas	Genética	Bioinformática	USP	SP	5	5
Multidisciplinar	Interdisciplinar	Bioinformática	UFPR	PR	3	-
Multidisciplinar	Interdisciplinar	Biologia Computacional e Sistemas	FIOCRUZ	RJ	4	4
Multidisciplinar	Interdisciplinar	Modelagem Computacional	LNCC	RJ	5	5
Ciências Exatas e da Terra	Ciência da Computação	Ciência da Computação	UNICAMP	SP	5	5

IES – Instituição de Ensino Superior, M - Mestrado Acadêmico, D – Doutorado

* *Links* com informações sobre cada Programa no site da CAPES

Fonte: Cursos Recomendados pela CAPES (<http://www.capes.gov.br/cursos-recomendados>), consultado em dezembro de 2010.

Plataformas com softwares e ferramentas de bioinformática

Como a maioria dos softwares e ferramentas de Bioinformática são elaborados para funcionar em uma plataforma baseada em [Linux](#), a maneira mais fácil de ter o seu primeiro [servidor](#) de Bioinformática é utilizando alguma [Distribuição Linux](#) como [sistema operacional](#).

O sistema operacional é o software responsável por gerenciar o [hardware](#), servindo de “interlocutor” entre o usuário e a máquina, e possibilitando que outros softwares (os [aplicativos](#)) utilizem o hardware para execu-

tar os comandos do usuário. São exemplos de sistemas operacionais o [Windows](#), o [MacOS](#), as várias [distribuições de GNU/Linux](#), o [Unix](#) e outros. Os editores de texto, os editores de imagens, os navegadores de internet, os players de vídeo e muitos outros são os softwares aplicativos, aqueles que executam funções específicas, tendo por objetivo o desempenho de tarefas práticas.

O termo Linux é utilizado para designar o núcleo (kernel) do sistema operacional, mas também tem sido usado para se referir aos sistemas operacionais baseados no núcleo Linux adicionados de outros programas. Qualquer usuário pode utilizar, estudar, modificar e distribuir o Linux, pois ele está sob a [licença GLP](#) (General Public License - Licença Pública Geral). As distribuições Linux têm se tornado cada vez mais populares devido ao uso de interfaces gráficas mais amigáveis, facilidades de instalação e manutenção, estabilidade e robustez.

Para designar os projetos com o objetivo de disponibilizar softwares e ferramentas de Bioinformática em uma plataforma Linux, tem-se utilizado genericamente o termo [BioLinux](#). Facilitar e divulgar o uso da Bioinformática é outro grande objetivo desses projetos e, para tanto, eles fornecem os softwares e as ferramentas de Bioinformática em [repositórios](#), em sistemas completos para instalação ou para executar diretamente de CD ou DVD (sem instalação – [Live CD/DVD](#)), além de construírem sistema de ajuda aos usuários.

Embora não sejam a melhor forma de utilização da bioinformática, pelas limitações de configuração oriundas do funcionamento a partir do CD/DVD, os Live CDs/DVDs podem ser uma forma interessante de se iniciar na Bioinformática e/ou fazer análises que não requerem muita capacidade computacional.

É possível encontrar alguns BioLinux disponíveis na internet, porém poucos projetos ainda contam com versões atualizadas e, na Bioinformática, assim como em quaisquer sistemas computacionais, as versões mais novas costumam passar por processos de melhoria no desempenho, incorporações de novos procedimentos e correções de erros,

sendo altamente recomendável fazer as análises com as versões mais atuais dos softwares, ferramentas e sistemas operacionais.

Projetos disponibilizados como Live CDs/DVDs

Lista de projetos BioLinux com descrições (adaptadas e traduzidas para a Língua Portuguesa) obtidas dos sites que os disponibilizam:

VLinux

Disponível em: <http://bioinformatics.org/vlinux>

O sistema de trabalho de Bioinformática VLinux é uma distribuição Linux para Bioinformática. É fácil de usar, não requer instalação, é uma distribuição em CD baseada no [Knoppix](#) 3.3. Ele inclui uma variedade de pacotes de análises de sequências e estruturas. É um produto de código aberto lançado sob a licença GNU/GPL. Última versão: 2004.

BioKnoppix

Informações disponíveis em <http://bioknoppix.hpcf.upr.edu>, mas com o download do sistema indisponível.

Bioknoppix é uma distribuição customizada do Knoppix Linux Live CD. Com esta distribuição, basta inicializar a partir do CD e você terá uma distribuição Linux totalmente funcional com aplicativos de código aberto orientados para biólogos moleculares. Última versão: 2004.

Vigyaan

Disponível em: <http://www.vigyaancd.org>

Vigyaan é uma plataforma de trabalho eletrônico para bioinformática, biologia computacional e química computacional. Ele foi projetado para atender às necessidades de novatos e peritos. VigyaanCD é um live CD de Linux que contém todos os softwares necessários para inicializar o computador e para uso de modelagem. VigyaanCD v1.0 é baseado no Knoppix v3.7. Última versão: 2005.

Quantian

Disponível em: <http://dirk.eddelbuettel.com/quantian.html>

Quantian é uma distribuição Knoppix/Debian adaptada para análises numéricas e quantitativas. A autoconfiguração e a inicialização direta-

mente do CD/DVD transforma qualquer computador em uma estação de trabalho Linux completa. Última versão: 2006.

DNALinux

Disponível em: <http://www.dnalinix.com>

DNALinux é uma máquina virtual com softwares para bioinformática pré-instalados. Os programas de uso comum em Bioinformática estão incluídos no DNALinux. Usando o VMware é possível rodar o Linux e o Windows lado a lado. DNALinux é baseada no Xubuntu 8.04. Última versão: 2009 (BASSI; GONZALEZ, 2010) .

BioSLAX

Disponível em: <http://www.bioslax.com>

BioSLAX é um Live CD/DVD novo com suíte de ferramentas de bioinformática. Inicializando em qualquer computador, este Live CD/DVD é executado com base no [SLAX](#), uma distribuição Linux baseada no [Slackware](#) Linux. Última versão: 2010 (BIOSLAX, 2010).

Bio-Linux

Disponível em: <http://nebc.nox.ac.uk/tools/bio-linux>

O Bio-Linux 6.0 é uma estação de trabalho de bioinformática completa, poderosa, configurável e fácil de manter. Bio-Linux oferece mais de 500 programas de bioinformática em uma base Linux Ubuntu 10.04. Há um menu gráfico para os programas de bioinformática, bem como um acesso rápido ao sistema de documentação de Bioinformática do Bio-Linux e a amostras de dados úteis para testar os programas. Você também pode instalar pacotes de Bio-Linux para lidar com tipos de dados de nova geração de sequenciamento. Você pode instalar o Bio-Linux em sua máquina ou executar a partir do Live DVD ou de um [dispositivo de memória flash USB](#). Última versão: 2010.

Sistemas completos

Sistemas completos, instalados e funcionando a partir do [disco rígido](#), são a melhor opção para se ter o primeiro servidor de Bioinformática, pois a instalação completa do Linux permite muito mais flexibilidade e poder computacional.

BioBrew

Disponível em: <http://bioinformatics.org/biobrew>

O BioBrew é uma coleção de aplicativos de código aberto para os pesquisadores da área biológica e é um projeto da bioinformatics.org. O BioBrew Roll for Rocks pode ser usado para criar o Rocks/BioBrew Linux, uma distribuição personalizada para bioinformática em clusters e em estação: ele automatiza a instalação de cluster, incluindo todos os softwares de [HPC](#) necessários a um cluster e contém os aplicativos mais populares de Bioinformática. As distribuições Rocks (e BioBrew RPMs) são compatíveis com as distribuições CentOS e Red Hat Enterprise Linux. Última versão: 2006. (BIOBREW, 2010)

Bio-Linux

Disponível em: <http://nebc.nox.ac.uk/tools/bio-linux>

Como anteriormente visto (Item 7, da Lista de Live CDs/DVDs), pode ser instalado em sua máquina ou executado a partir do Live DVD ou de um dispositivo de memória flash USB. Última versão: 2010.

Pacotes e repositórios

Quem possui um sistema operacional Linux funcionando pode, optar por somente adicionar pacotes, disponíveis em repositórios, ao seu sistema. A instalação de pacotes do sistema [RPM](#), para os usuários RedHat e derivados, e do sistema [APT](#), para a comunidade Debian, facilita a construção e manutenção de sistemas Linux, pois os pacotes, com seus plugins e acessórios, são facilmente instalados e todas as dependências de outros pacotes e bibliotecas são checados. Segue uma lista de repositórios de acordo com a distribuição Linux.

Repositório RedHat/CentOS/Fedora e derivados (RPM):

<http://ftp.bioinformatics.org/pub/biobrew>

<ftp://fr.rpmsfind.net/linux>

<http://informatics.umdj.edu/BioRPMs>

Repositório Debian (APT):

<http://packages.debian.org/lenny>

<http://nebc.nox.ac.uk/bio-linux/dists/unstable/bio-linux>

Conclusões

Pela disponibilidade de uma versão atualizada e do uso possível como Live DVD ou sistema completo, o BioLinux, disponível em: <http://nebc.nox.ac.uk/tools/bio-linux>, atualmente é a melhor opção, principalmente para usuários iniciantes em Bioinformática, pois a versão Live DVD permite avaliar o sistema operacional baseado no Ubuntu e todas os softwares e as ferramentas de Bioinformática sem qualquer necessidade de instalação. E, após a familiarização com o sistema, o usuário pode instalá-lo de forma completa e desfrutar do seu primeiro servidor de Bioinformática.

Cada usuário experiente em sistemas Linux e cada bioinformata tem sua distribuição Linux preferida, e não pretendemos modificar preferências. Pretendemos, sim, estimular os usuários iniciantes a utilizarem, da maneira mais ampla e acessível possível, os softwares e ferramentas de Bioinformática.

Porém, os usuários experientes também podem ganhar com as facilidades proporcionadas por esses projetos, seja para adicionar e atualizar softwares de Bioinformática mais facilmente nos seus sistemas, seja pela possibilidade de uso desses sistemas em treinamentos, sem que seja necessário instalar nada, executando o sistema em Live DVD.

A Bioinformática é uma área de pesquisa nova e carente de profissionais, por isso relacionamos algumas oportunidades de treinamentos. E cabe ressaltar que o treinamento de usuários iniciantes deve ser encarado como uma das obrigações dos usuários experientes, pois quanto mais usuários forem treinados em Bioinformática, mais a área crescerá, produzindo mais soluções e mais conhecimentos, tornando-a cada vez maior e mais importante.

Referências

AGARWAL, P.K. **Vigyaan bio/chemical software Workbench**. Disponível em: <<http://www.vigyaancd.org>>. Acesso em: dez. 2010.

ASSOCIAÇÃO BRASILEIRA DE BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL. **Sobre AB3C**. Disponível em: <<http://lgmb.fmrp.usp.br/ab3c>>. Acesso em: dez. 2010.

BASSI, S.; GONZALEZ, V. **DNALinux: a Linux solution for bioinformatics**. Disponível em: <<http://www.dnalinix.com>>. Acesso em: dez. 2010.

BIOBREW: bioinformatics packages. Disponível em: <<http://bioinformatics.org/biobrew>>. Acesso em: dez. 2010.

BIOSLAX: bioinformatics LiveCD Suite. Singapore: National University of Singapore, 2010. Disponível em: <<http://www.bioslax.com>>. Acessado em: dez. 2010.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Cursos recomendados e reconhecidos**. Disponível em: <<http://www.capes.gov.br/cursos-recomendados>>. Acesso em: dez. 2010.

EDDELBUETTEL, D. **The Quantian Scientific Computing Environment**. Disponível em: <<http://dirk.eddelbuettel.com/quantian.html>>. Acesso em: dez. 2010.

GIULIATTI, S. **Informática biomédica**. Ribeirão Preto: Faculdade de Medicina, Universidade de São Paulo, 2009. Disponível em: <http://www.fmrp.usp.br/cg/novo/index.php?option=com_content&view=article&id=115&Itemid=91>. Acesso em: dez. 2010.

NERC. **Environmental Bioinformatics Centre**. Bio-Linux. Disponível em: <<http://nebc.nox.ac.uk/tools/bio-linux>>. Acesso em: dez. 2010.

RODRIGUEZ, C. M. **BioKnoppix**. Disponível em: <<http://bioknoppix.hpcf.upr.edu>>. Acesso em: dez. 2010.

TIWARI, B. **Gearing up for Bioinformatics**. CBHD Newsletter, n. 34, 3 Mar. 2005. Disponível em: <http://gchelpdesk.ualberta.ca/news/03mar05/cbhd_news_03mar05.php>. Acesso em: dez. 2010.

UNIVERSIDADE CRUZEIRO DO SUL. **Bioinformática**. Disponível em: <http://www.cruzeirosul.edu.br/content/graduacao_detalle.aspx?curso=56>. Acesso em: dez. 2010.

UNIVERSIDADE DE SÃO PAULO. Faculdade de Medicina de Ribeirão Preto. Departamento de Genética. Curso de inverno de bioinformática. Disponível em: <<http://gbi.fmrp.usp.br/cursodeinverno>>. Acesso em: dez. 2010.

UNIVERSIDADE DE SÃO PAULO. Faculdade de Medicina de Ribeirão Preto. Departamento de Genética. Curso de verão em bioinformática. Disponível em: <<http://lgmb.fmrp.usp.br/cvbioinfo>>. Acesso em: dez. 2010.

UNIVERSIDADE DE SÃO PAULO. Programa Interunidades de Pós-Graduação em Bioinformática. Apresentação. Disponível em: <<http://www.ime.usp.br/posbioinfo>>. Acesso em: dez. 2010.

UNIVERSIDADE DE SÃO PAULO. Programa Interunidades de Pós-Graduação em Bioinformática. Cursos. Disponível em: <<http://www.ime.usp.br/posbioinfo>>. Acesso em: dez. 2010.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. Instituto de Ciências Biológica. Departamento de Bioquímica e Imunologia. Programa de doutorado em bioinformática. Disponível em: <<http://www.pgbioinfo.icb.ufmg.br>>. Acesso em: dez. 2010.

UNIVERSIDADE FEDERAL DO PARANÁ. Programa de Pós-Graduação em Bioinformática. Justificativa. Disponível em : <<http://www.bioinfo>

ufpr.br/mod/resource/view.php?id = 113 > . Acesso em: dez. 2010.

VIMALKUMAR, V. **VLinux Bioinformatics Workbench**. Disponível em: <<http://bioinformatics.org/vlinux>> . Acesso em: dez. 2010.

WIKIPEDIA. **BioLinux**. Disponível em: <<http://en.wikipedia.org/wiki/Bio-Linux>> . Acesso em: dez. 2010.

WIKIPEDIA: a enciclopédia livre: página principal. Disponível em: <http://pt.wikipedia.org/wiki/Main_Page> . Acesso em: dez. 2010.

Embrapa

Milho e Sorgo

**Ministério da
Agricultura, Pecuária
e Abastecimento**

