

# Contribuições do uso de *Data Warehouse* para o Armazenamento de Dados em Bioinformática

Luiz Fernando Picolo<sup>1</sup>

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul (IFMS)  
Nova Andradina – MS – Brasil  
luizpicolo@luizpicolo.com.br

**Resumo.** *Com o aumento das informações provenientes das áreas biológicas, a Biologia e a informática encontraram um ponto em comum, originando assim, a Bioinformática. Este novo ramo da ciência tem como objetivo utilizar o poder computacional para organizar e estruturar os dados de forma compreensível. Neste sentido, as tecnologias computacionais são utilizadas para facilitar o processo de consulta, cruzamentos e agregações de dados. Neste artigo, busca-se demonstrar os Data Warehouses e suas contribuições para o armazenamento de dados em Bioinformática, por serem uma longa série de visões de dados tiradas ao longo do tempo, proporcionando aos pesquisadores informações para a correlação entre os dados.*

**Palavras-chave:** *Bioinformática, Data Warehouse.*

## 1. Introdução

Na sociedade contemporânea a tecnologia não é algo distante ou somente acessível aos grandes pensadores científicos ou empresas, ela está presentes nos afazeres comuns de todos. A utilização do celular, a televisão, ou o simples fato de se consultar a internet, nada mas são do que exemplos do crescimento tecnológico que se vive nos tempos atuais. Assim, a tecnologia gerou mudanças que foram incorporadas e são utilizadas de forma quase transparente pela sociedade.

Para Mattar Neto (2005), outro fato que merece atenção, e que sofreu mudanças devido a computação, são os conceitos relacionados ao acúmulo de informações. Para o autor, acumular informações não deve ser compreendido como sinais de sabedoria, pois na era da informática, todas as informações são armazenadas, disponibilizadas e transmitidas por poderosos bancos de dados, os quais fazem parte deste cabedal tecnológico da contemporaneidade vigente. Para ele, a habilidade de estabelecer conexões entre as informações é o fato primordial para a geração de informações relevantes sobre o tema que é pesquisado. Neste sentido, como afirma Mattar Neto (2005), a informática caracteriza-se como um nova tecnologia intelectual que potencializa não apenas a memória, mas também diferentes formas de raciocínio e perspectivas sobre a informação. Logo, tratar informações e não somente acumulá-las tornou-se a forma mais eficaz de se obter resultados satisfatórios.

Como na sociedade, e análoga as ideias de Mattar Neto (2005), as ciências

biológicas também estão incorporadas a tecnologia e ao tratamento de suas informações. Segundo Rocha (2011) com o avanço da biologia e da informática, ambas ciências encontraram entre si um ponto em comum, ou seja, esta nova ciência que seria criada, surgia a partir da necessidade da biologia de não apenas armazenar os dados gerados, mas sim de tratar tais informações com equipamentos de alta precisão. Tal ciência que surgia em meados dos anos 90, seria chamada de Bioinformática. Por este motivo, há esta similaridade com os conceitos de Mattar Neto (2005), sendo que, a Bioinformática busca utilizar o poder computacional para armazenar e trata as informações obtendo assim, novos e relevantes dados.

Dentro deste contexto, a tecnologia dos *Data Warehouses* apresenta uma alternativa para o tratamento das informações em Bioinformática. Um Data Warehouse, como ferramenta computacional, absorve todos estes conceitos apresentados por Mattar Neto (2005), ou seja, de não ser somente uma forma de armazenagem, mas sim uma ferramenta que amplia as possibilidades e auxilia no cruzamento de informações e tomada de decisões.

Neste sentido, este artigo tem como objetivo demonstrar um dos principais esforços computacionais para sanar alguns destes problemas envolvidos ao armazenamento e ao tratamento das informações. Portanto, para uma compreensão mais minuciosa, tratar-se-á especificadamente sobre os *Data Warehouses* e suas contribuições para o armazenamento de informações em Bioinformática.

## **2. Revisão Teórica**

Ao adentrar a teoria que embasa os meios pelos quais a informática se liga as pesquisas biológicas, torna-se claro que a biologia, nos dias atuais, seria impossível sem os recursos que a computação proporciona, visto que, toda informação produzida, como no caso das pesquisas genômicas, necessitam ser armazenadas, tratadas e, posteriormente, analisadas. Logo, com o avanço de ambas as partes, surge um ponto em comum que une as duas ciências originalizando um novo ramo de pesquisa, a Bioinformática.

A Bioinformática pode ser compreendida como uma área multidisciplinar. Segundo Queiroz (2012), este ramo da ciência é um campo de pesquisa que está em rápido crescimento sendo desenvolvida para atender as necessidades da manipulação de grandes volumes de dados genéticos e bioquímicos, os quais, originam-se a partir dos esforços individuais de vários pesquisadores. Sua abrangência se faz em diversos aspectos e em vários campos científicos como: na física, na química, na biologia, ciência da computação, entre outros.

Outra característica da Bioinformática definiu-se por ser uma modalidade que envolve aspectos de aquisição, processamento, armazenamento, distribuição, análise e interpretação das informações. Assim, “a Bioinformática [...] tem como objetivo desenvolver e aplicar técnicas computacionais no estudo da genética, da biologia molecular e da bioquímica” (WIECZOREK e LEAL, 2006, p. 01) buscando a compreensão e a relação das diversas informações fracionadas e oriundas destas diversas áreas biológicas. Portanto, “a Bioinformática usa o poder computacional para catalogar, organizar e estruturar estas informações em uma entidade compreensiva e

extremamente importante para a biologia” (ROCHA, 2011, p. 03).

Outro fato interessante que impõe o uso de ferramentas computacionais é a grande escala de dados originados das pesquisas realizadas por meio do projeto genoma<sup>1</sup>. Com o início deste projeto em 1990, o volume de informações derivadas do conhecimento genômico tornou-se exponencialmente crescente. Um exemplo, e consequentemente confirmação deste fato, foi a publicação do genoma humano o qual contém cerca de 30 mil genes, ou seja, a sequência completa de um indivíduo tem cerca de 3 bilhões de bancos de DNA. No *Sanger Center*<sup>2</sup> já existem algumas dezenas de *terabytes* armazenados com uma estimativa de crescimento, e que, segundo Lifchitz (2006), podem crescer na ordem de um *petabyte* por ano. Outro exemplo deste crescimento absurdo de dados genômicos é o *GenBank*<sup>3</sup> o qual armazenava quase 15 bilhões de dados e, em Janeiro de 2011, passou a conter quase 130 bilhões, seguindo assim em ritmo de crescimento constante.

Com tamanha quantidade de informações necessitou-se de ferramentas computacionais que pudessem sistematicamente armazenar tal quantidade de dados para que os mesmos se tornassem fontes de estudos.

Neste sentido, Felix et al (2002, p. 05) expõem que:

o grande volume de informação gerado pelos projetos de análise de transcriptomas<sup>4</sup> tem tornado cada vez mais complexo o armazenamento e a análise dos dados. Para contornar tal dificuldade, devem ser implementados, bancos de dados, que disponibilizem, de modo confiável, os dados e ferramentas de análise. (FELIX, 2002, p. 05)

Justamente neste ponto, apoiando-se na afirmação acima, pode-se firmar que há a presença de bancos de dados robustos e sistemas gerenciadores de bancos de dados (SGBDs) que possam suportar tamanha grandeza de informações. Além dos softwares, o desafio apresentado pela Bioinformática é encontrar a melhor forma de armazenamento e de pesquisa (SQL<sup>5</sup>) para os dados gerados.

## 2.1. Banco de dados

Para que o entendimento de alguns conceitos que serão utilizados no decorrer deste artigo sejam claras, neste momento, deve-se compreender alguns conceitos. Primeiramente far-se-á a explicação sobre o que são bancos de dados e uma breve explicação sobre os sistemas utilizados para o seu gerenciamento.

---

1 O genoma (conjunto de genes de uma espécie) está contido na área da ciência denominada genética, que é responsável pelo estudo da reprodução, herança, variação e de aspectos relacionados à descendência. Disponível em: <<http://educar.sc.usp.br/licenciatura/2001/genoma/genoma.html>>. Acessado em: 05 Dez 2012

2 Centro de Pesquisa e de sequenciamento de DNA britânico. <<http://www.sanger.ac.uk/>>

3 GenBank é [...] uma coleção anotada de todas as sequências de DNA publicamente disponíveis. Disponíveis em: <<http://www.ncbi.nlm.nih.gov/genbank/>>. Acessado em: 20 Dez 2012

4 Para maiores informações visite: <http://www.ime.usp.br/posbioinfo/cv2011/eduardoreis.pdf> Acessado em: 20 Dez 2012

5 Do inglês, Structured Query Language ou Linguagem de Consulta Estruturada.

Lifschitz (2006, p. 03) afirma que:

um sistema de banco de dados é [...] parte de um sistema de informação, o qual inclui todos os componentes que são envolvidos na organização dos recursos. Esses incluem os próprios dados, o software do SGBD, o hardware do sistema, a mídia de armazenamento e os aplicativos que acessam e atualizam os dados. Desta forma, o sistema de banco de dados é parte importante da organização de um sistema de informação. Por isto, o ciclo de vida um sistema de informação deve também incluir o ciclo de vida de um banco de dados. (LIFSCHITZ, 2006, p. 03)

Bancos de dados, segundo Takai, Italiano e Ferreira (2005), são a coleção de dados logicamente relacionados que contém algum significado, ou seja, elas são estruturas construídas e posteriormente preenchidas para um propósito específico. Outra característica apontada pelos autores é que os bancos de dados representam aspectos do mundo real, ou como os próprios dizem, o “míni mundo”, e toda alteração no “míni mundo” provoca alteração em sua base de dados. Logo, a afirmação de Lifschitz (2006) neste ponto entra em concordância com a de Taki, Italiano E Ferreira (2005), em que ambos os autores atribuem aos bancos de dados a importância na organização e na posterior utilização dos dados, gerando assim, benefícios para o objetivo que o mesmo foi criado.

Já os “Sistemas Gerenciadores de banco de Dados ou (SGBDs) são uma coleção de programas que permitem aos usuários criarem e manipularem um banco de dados” (TAKAI, ITALIANO E FERREIRA, 2005, p. 05) por meio de uma interface de alto nível. Ao contrário da proposta feita para os bancos de dados, ou seja, a de ser estruturado para determinado fim, o SGBD tem propósito mais genérico direcionado à facilitar o processo de definir, construir e manipular bancos de dados de diversas aplicações. Segundo Takai, Italiano e Ferreira (2005) com o tempo, os SGBDs passaram a utilizar diferentes formas de representação ou modelos de dados para descrever a estrutura das informações contidas em seus bancos de dados. Assim, o objetivo de se projetar um banco de dados é acomodar informações para que usuários em potencial possam usufruir dos dados que o mesmo contém, de forma ágil e objetiva.

### **3. Objetivos**

O presente artigo tem por objetivo principal compreender o uso dos *Data Warehouses* para o armazenamento das informações provenientes da Bioinformática atentando assim às discussões relacionadas aos aspectos deste ramo científico e, em particular, para o auxílio do mesmo para futuras tomadas de decisões e conexão entre informações.

Para a melhor compreensão do objetivo geral, foram definidos alguns objetivos específicos: Compreender os relevantes que fazem a conexão entre a ciência biológica e a computação; Demonstrar a contribuição dos *Data Warehouses* para a tomada de decisões em Bioinformática; Descobrir como estão sendo estruturados os *Data Warehouses* para atender os requisitos da Bioinformática.

Consequentemente, como todo conhecimento criado e disponibilizado, mesmo sendo uma pequena fatia do imenso patrimônio de pesquisas realizadas na área, é inata a contribuição para futuras pesquisas em Bioinformática relacionadas ao objetivo

proposto.

#### 4. Percurso Metodológico

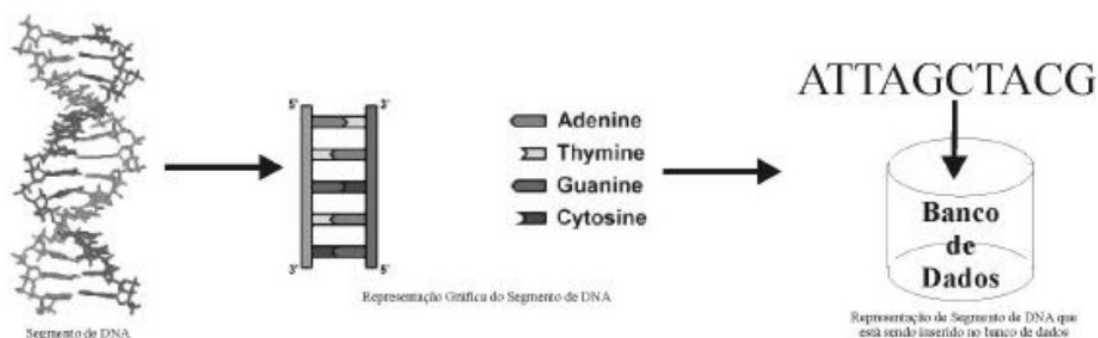
A metodologia, com base nos conceitos de Gil (2010), adotada para a elaboração deste artigo foi constituída da pesquisa bibliográfica. Segundo o autor, toda pesquisa acadêmico requer em determinados momentos a realização do trabalho caracterizado como pesquisa bibliográfica. A maioria das pesquisas são realizadas com base principalmente em materiais obtidos em fontes bibliográficas, permitindo ao investigador a cobertura de uma gama de fenômenos maior que o mesmo poderia pesquisar diretamente.

A primeira etapa, a qual foi dividida em duas sub-etapas, foi a busca de fontes referentes ao tema proposto. Em primeiro momento foi feita a abordagem dos estudos referentes a Bioinformática e banco de dados, buscando a relação entre estas duas linhas de pesquisa. Logo após, realizou-se a busca de textos sobre os conceitos de *Data Warehouses* e como os mesmos estavam sendo utilizados e estruturados para o armazenamento dos dados.

A segunda etapa se deu a leitura do material e o fichamento do mesmo. Após o fichamento se fez a organização lógica dos assunto. A última etapa foi o relacionamento de todas as ideias e a criação deste artigo.

#### 5. Análise dos Resultados Obtidos

Um problema a ser superado ao se tratar sobre banco de dados para Bioinformática é que bancos de dados, segundo Wieczorek e Leal (2006), têm sido em grande parte usados para administrar dados empresariais, números simples, caráter ou datas. Poucos bancos de dados tiveram uma habilidade nativa para lidar com dados complexos, como dados multimídia, texto, dados espaciais, ou dados genéticos (sucessão de genes). Neste sentido, a busca, a comparação e o armazenamento, como pode ser visualizado na Figura 1, de grandes cadeias de caracteres se torna difícil de ser executada.



**Figura 1: Etapas realizadas para armazenar um segmento (sucessão de genes) de DNA em um banco de dados. (WIECZOREK e LEAL, 2006, p. 04)**

Banerjee (2000) relata em suas explanações sobre plataformas de banco de

dados para Bioinformática que, conseqüentemente, todos os genes serão encontrados, não por ser somente um fato, mas uma tendência com o volume de dados gerados. Entretanto, muitos dos desenvolvimentos paralelos a estas descobertas, as quais acarretarão mudanças benígnas na área da medicina, como a descoberta de doenças hereditárias ou a cura de males já existentes por exemplo, não serão feitas dentro de laboratórios biológicos, mas em plataformas computacionais de alto desempenho, com sistemas de armazenamento em massa que conterão os dados genômicos, bancos de dados para pesquisar os dados, identificando semelhanças e padrões, bem como a integração software para a unificação das fatias de conhecimento desenvolvidas nas diversas instituições.

Para ele, existem quatro tecnologias que tendem a resolver os problemas referentes ao armazenamento de dados biológicos, as quais são: tecnologias de *Warehousing*<sup>6</sup> para dados em padrões genéticos; a arquitetura de extensibilidade para armazenamento de dados nativamente e execução de estruturas de procura; tecnologias de integração de dados para habilitar questões heterogêneas por fontes biológicas distribuídas, e tecnologias de portal de Internet que possam permitir a publicação das informações de pesquisas. Sendo que, para este artigo, se focará apenas nas tecnologias de *Warehousing* para dados em padrões genéticos.

Segundo Takai, Italiano e Ferreira (2005) um *Warehousing* pode ser definido como uma técnica utilizada para recuperação e integração de dados a partir de fontes distribuídas, autônomas e, possivelmente, heterogêneas. Estes dados são armazenados em um grande depósito chamado de *Data Warehouse*. Um *Data Warehouse* sumaria os dados que são organizados em dimensões, disponibilizando-os para consultas e análises por meio de aplicações OLAP (*On-Line Analytical Processing* ou processamento analítico online) e sistemas de suporte à decisão.

Para Bongioiolo (2006) uma diferencial que se destaca entre um *Data Warehouse* e um banco de dados transacional está no fato de que o segundo, como a própria denominação informa, armazena dados das transações diárias nas quais são executadas operações pré-definidas. Por este motivo os dados podem sofrer mudanças. Já um *Data Warehouse* armazena dados para análise e tomada de decisão. Isso pode envolver consultas que variam em sua complexidade, pois o mesmo contém dados históricos de anos.

Logo, algumas características necessárias para este ambiente de armazenamentos são:

- Integração de dados de múltiplas fontes;
- Facilitar o processo de consulta, cruzamentos e agregações de dados;
- Ser flexível para suportar grandes mudanças e quantidade de informação; dentre outras;
- Estar disponível para o acesso para que os usuários possam interagir com os

---

6 O termo *Warehousing* não possui tradução adequada para o português.

dados de forma rápida e objetiva;

Desta forma, pode-se afirmar que “os dados armazenados no *Data Warehouse* são, para todos os propósitos práticos, uma longa série de visões do banco de dados, tiradas ao longo do tempo” (TAKAI, ITALIANO E FERREIRA, 2005, p. 89). Neste sentido, os *Data Warehouses* proporcionam tanto à Bioinformática como para empresas, uma base concreta de agregação de dados para futuras análises. Os dados contidos em um *Data Warehouse* são organizados e armazenados com a perspectiva de longo prazo, visando desta forma a realização de comparações e pesquisas nas informações contidas. Outra característica básica está na sua integração de dados. Como as informações provenientes da Bioinformática não são oriundas apenas de uma fonte e sim de uma aglomeração de dados vindouros de laboratórios biológicos espalhados em vários centros universitários e laboratórios de pesquisa, a integração de dados proporcionada por um *Data warehouse* uma infra estrutura que permite criar e manter uma visão consistente de várias fontes de dados autônomas.

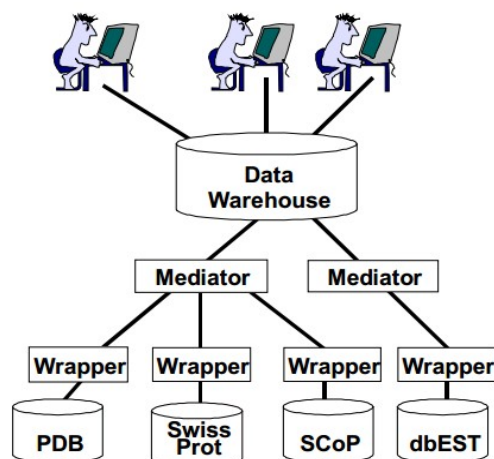
Entretanto, mesmo sendo bem aceita esta abordagem para o tratamento dos problemas da integração de dados vindouros de fontes de dados distribuídas e heterogêneas, Seibel (2000) aponta que esta forma de integração também não atende à integração de informações biológicas com relação à atualização dos esquemas e também com relação à atualização das instâncias de dados (os pesquisadores precisam ter acesso aos dados mais recentes).

Logo, baseando-se na afirmação já citada, Seibel (2000) propõe uma complementação a abordagem do *Data Warehouse*, a qual pode ser visualizada na Figura 2.

Uma arquitetura baseada na abordagem *wrapper-mediador* difere fundamentalmente da abordagem *Data Warehouse* pelo fato de que os dados, na primeira, não são materializados. Porém, estas abordagens podem ser complementares pois o mediador pode ser usado como fonte de dados para a implementação do *Data Warehouse* (SEIBEL, 2000).

Em 1996 o *Lawrence Livermore National Laboratory* deu início ao projeto *DataFoundry* para a criação de um *Data Warehouse* para o armazenamento de dados genômicos. Segundo Critchlow (2000), para a criação de um *Data Warehouse* deve-se criar uma estrutura com alto grau de flexibilidade para lidar com a natureza dinâmica do domínio. Ao contrário de aplicações comerciais, fontes de dados científicos, dando ênfase a dados genômicos, são extremamente dinâmicas e, em sua grande maioria, oriundas de diversas localidades e banco de dados diferentes. Logo, para sanar este problema, uma das camadas presentes em um *Data Warehouse* são os *Wrappers*, os quais permitem “o acesso às fontes de dados distribuídas e heterogêneas e promovem a transformação entre o modelo de dados da fonte e o modelo de dados interno” (CÔCO, 2005, p. 14), e outra são os *Mediators* (Mediadores) que oferecem o acoplamento entre provedores e consumidores de dados por envio de serviços que tratam a questão semântica e sintaxe entre os objetos dos sistemas locais. Este sistema fornece uma visão uniforme dos recursos que estão disponíveis para usuários, permitindo manipular esses recursos por meio de uma linguagem de alto nível. Sempre que uma fonte de dados

muda suas informações, os *Wrappers* e os *Mediators*, como mostra a Figura 2, devem ser atualizados para que estas modificações sejam espelhadas no *Data Warehouse*.



**Figura2: Arquitetura de um *Data Warehouse*.**  
(CRITCHLOW et al, 2000, p.08)

Wieczorek e Leal (2006) expõe que, além dos vários problemas e soluções já citados, ainda há grandes desafios para a introdução desta abordagem, pois deve-se manter um *Data Warehouse* extremamente funcional, mesmo integrando várias fontes de dados que sofram mudanças constantemente. Contudo, várias pesquisados se debruçam sobre o assunto para criar soluções para tal situação.

O *Data Warehouse*, como foi afirmado por Banerjee (2000), é uma tendência para o armazenamento de dados oriundos das pesquisas genômicas, pois os dados gerados não são armazenados apenas em um banco de dados, mas espalhados em diversas instituições de pesquisas espalhadas pelo mundo. “Com a proliferação destes bancos de dados públicos que tem sido observado nos últimos anos, os pesquisadores precisam de uma forma fácil e intuitiva de acessar todos estes dados” (BONGIOLO, 2006, p. 85). Logo, baseando-se nestas informações, nada mais plausível que, com as cargas e recargas feitas em um *Data Warehouse*, todos os laboratórios de pesquisas possam ter em mãos uma cópia das informações para que os mesmos possam realizar suas pesquisas mediante aos dados coletados. Outro fato que justifica a utilização dos *Data Warehouse* é que:

consultar [...] dados de forma distribuída tem seus problemas e limitações, como restrições de recursos do servidor remoto, preocupação com a segurança dos dados [...] pela rede, além da logística inerente a consulta de bancos distribuídos. Num ambiente assim, consultas distribuídas que gerem um grande volume de dados são difíceis de processar, além de requerer que consultas complexas integrem dados heterogêneos de fontes distintas. (BONGIOLO, 2006, p. 85)

Além da limitação citada, pode-se incluir a possibilidade de erros ao se retornar



certa quantidade de dados. Como as informações estão distribuídos em servidores remotos, a disponibilidade dos mesmos não pode ser garantida e, uma consulta pode ser falha devido a esta circunstância. Assim um Data Warehouse é uma ferramenta importante para a Bioinformática, pois os mesmos, segundo Bongioiolo (2006), são construídos para que tais dados sejam armazenados e acessados de forma a não serem limitados por tabelas e linhas estritamente relacionais. Outro adendo é que, o maior interesse em utilizar os *Data Warehouses* esta também no fato que, além de integrar vários bancos heterogêneos, os mesmo convertem e limpam seus dados e implementam ferramentas que facilitam a navegação pelos mesmos, permitindo cruzamentos e agregações.

## 6. Considerações Finais

Este artigo procurou evidenciar como os *Data Warehouses* contribuem para o armazenamento e a pesquisa dos dados em Bioinformática. Com esta ferramenta tecnológica, os pesquisadores podem fazer comparações, análises e cruzamentos entre várias informações, as quais são tão importantes para as pesquisas biológicas, pois, por meio deste cruzamentos de dados é que se obtêm resultados que levarão os pesquisadores a descobertas para cura de doenças e prevenção de males hereditários.

Para trabalhos futuros pode-se mencionar, na linha de pesquisa sobre banco de dados, como estão sendo utilizados os bancos de dados para o armazenamento das informações provenientes das pesquisas feitas no Brasil, sendo que, este artigo, buscou demonstrar um panorama geral. Assim, uma pesquisa que volte a visão para o Brasil, seria de grande importância para futuras discussões sobre o assunto.

## Referências

- BANERJEE, Sandeepan. **A Database Platform for Bioinformatics**. Oracle Corporation, Redwood Shores, 2000. Disponível em: <[http://www.grandpoohbah.net/Sandeepan/VLDB\\_Bioinformatics.pdf](http://www.grandpoohbah.net/Sandeepan/VLDB_Bioinformatics.pdf)> Acessado em: 05 Dez 2012.
- BONGIOLO, Elisângela. **Análise panorâmica da Bioinformática no Brasil: Proposta da gestão de pessoas para os laboratórios de pesquisa**. Criciúma, 2006. Disponível em: <<http://www.bib.unesc.net/biblioteca/sumario/00002E/00002E38.pdf>> Acessado em: 20 Dez 2012
- CÔCO, Tatiana Mara. **Implementando wrappers xml e relacional para o Codims**. 2005. Disponível em: <[http://codims.lprm.inf.ufes.br/publicacoes/Monografia\\_Tatiana.pdf](http://codims.lprm.inf.ufes.br/publicacoes/Monografia_Tatiana.pdf)> Acessado em: 09 Dez 2012.
- CRITCHLOW, Terence.; MUSICK, Ron.; SLEZAK, Tom. **An Overview of Bioinformatics Research at Lawrence Livermore National Laboratory**. 2000. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.939&rep=rep1&type=pdf>> Acessado em: 20 Dez 2012.

FELIX, Juliana de Maria.; DRUMMOND, Rodrigo Duarte.; NOGUEIRA, Fábio Tebaldi Silveira.; JUNIOR, Vicente Eugenio de Rosa. JORGE, Renato Atílio.; ARRUDA, Paulo.; MENOSSI, Marcelo. **Genoma Funcional**. 2002. Disponível em: <<http://ipe.cbmeg.unicamp.br/cafe/pub/felix-gf.pdf>> Acessado em: 10 Dez 2012.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 5. Ed. São Paulo: Atlas, 2010.

LIFSCHITZ, Sérgio. **Algumas Pesquisas em Bancos de Dados e Bioinformática**. PUC-Rio, 2006. Disponível em: <<http://www.natalnet.br/sbc2006/pdf/arq0271.pdf>> Acessado em: 03 Dez 2012.

MATTAR NETO, João Augusto. **Metodologia Científica na Era da Informática**. 2. ed. São Paulo: Saraiva, 2005.

QUEIROZ, Alexandre. **Apostila de Introdução à Bioinformática**. Rio Grande do Norte: UFRN, 2002. Disponível em: <[http://genfis40.esalq.usp.br/genfis/index.php?option=com\\_phocadownload&view=category&download=10:introducao-a-bioinformatica&id=7:apostilas-e-artigos&Itemid=68](http://genfis40.esalq.usp.br/genfis/index.php?option=com_phocadownload&view=category&download=10:introducao-a-bioinformatica&id=7:apostilas-e-artigos&Itemid=68)> Acessado em: 11 Dez 2012.

ROCHA, Cícero Pinho. **Banco de dados em bioinformática**. UESPI, 2011. Disponível em: <[http://artigocientifico.uol.com.br/uploads/artc\\_1194549401\\_86.pdf](http://artigocientifico.uol.com.br/uploads/artc_1194549401_86.pdf)> Acessado em: 14 Dez 2012.

SEIBEL, Luiz Fernando Bessa. **Bio-AXS: Uma Arquitetura para Integração de Fontes de Dados e Aplicações de Biologia Molecular**. Disponível em: <[ftp://ftp.inf.puc-rio.br/pub/docs/theses/02\\_PhD\\_seibel.pdf](ftp://ftp.inf.puc-rio.br/pub/docs/theses/02_PhD_seibel.pdf)> Acessado em: 01 Dez 2012.

TAKAI, Osvaldo Kotaro.; ITALIANO, Isabel Cristina.; FERREIRA, João Eduardo. **Introdução a Banco de Dados**. São Paulo: USP, 2005. Disponível em: <<http://www.ime.usp.br/~jef/apostila.pdf>> Acessado em: 17 Dez 2012.

WIECZOREK, Emilio Mario.; LEAL, Eduardo. **Caminhos e Tendências do uso de Banco de Dados em Bioinformática**. CEULP, 2006. Disponível em: <<ftp://www.ufv.br/DBG/material%20curso%20bioinfo/Leitura%20Complementar/artigos/Caminhos%20e%20Tend%EAncias%20do%20uso%20de%20Banco%20de%20Dados%20em%20Bioinform%20tica.pdf>> Acessado em: 10 Dez 2012.