

LIGA DE ENSINO DO RIO GRANDE DO NORTE
FACULDADE NATALENSE PARA O DESENVOLVIMENTO DO RIO
GRANDE DO NORTE
CURSO BACHARELADO EM SISTEMA DE INFORMAÇÃO

Bioinformática: Vantagens do uso de Banco de Dados na Biologia

Gabriel Martins de Souza

NATAL / RN

2010

GABRIEL MARTINS DE SOUZA

BIOINFORMÁTICA

Projeto de Monografia apresentado ao curso de Bacharelado em Sistemas de Informação da Faculdade Natalense para o Desenvolvimento do Rio Grande do Norte como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Joseane Pinheiro

NATAL / RN

2010

37299

Catálogo na Publicação – Biblioteca da FARN
Setor de Processos Técnicos

Souza, Gabriel Martins de
Vantagens do uso de Banco de dados na Biologia / Gabriel Martins
de Souza. – Natal: FARN, 2010.
21f.

Orientador: Joseane Pinheiro (MESTRE.)
Monografia (Bacharel em Sistema de Informação) – Faculdade
Natalense para o Desenvolvimento do Rio Grande do Norte.

1. Sistema de Informação – Monografia. 2. Banco de Dados –
Monografia. 3. Bioinformática – Monografia. I. Título.

RN/FARN/BC

CDU 004.65

BIBLIOTECA UNI-RN
004.65 S719b - RB: 37299 (M002107)
Título: Bioinformática : vantagens do uso de
Autoria: Souza, Gabriel, Martins de
2010, ex. 1

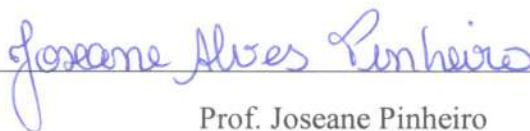
LIGA DE ENSINO DO RIO GRANDE DO NORTE
FACULDADE NATALENSE PARA O DESENVOLVIMENTO DO RIO GRANDE DO
NORTE

CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

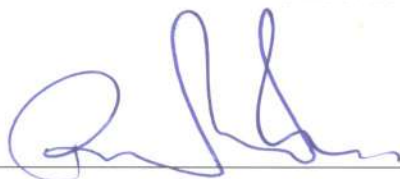
Vantagens do uso de Banco de Dados na Biologia

Gabriel Martins de Souza

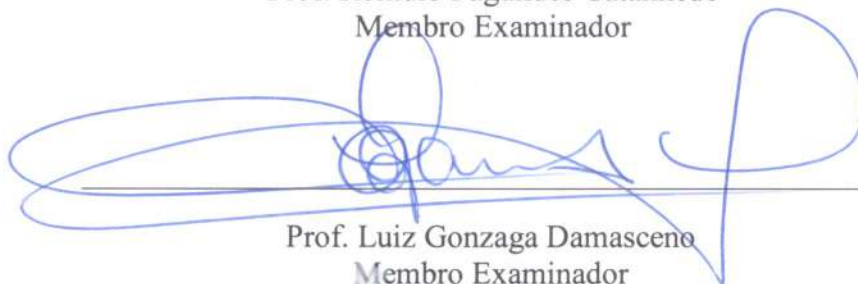
Monografia apresentada e aprovada em ____ de ____ de ____, pela banca examinadora
composta pelos seguintes membros:



Prof. Joseane Pinheiro
Orientador



Prof. Romulo Fagundes Catanhede
Membro Examinador



Prof. Luiz Gonzaga Damasceno
Membro Examinador

Natal/RN, ____ de ____ de ____.

À minha família dedico este trabalho, por sempre ter acreditado em minhas capacidades, principalmente aos meus pais, por ter me dado esta grande oportunidade.

AGRADECIMENTOS

Gostaria primeiramente de agradecer a Deus, por sempre ter me guiar para o caminho certo, dado a minha saúde e sabedoria para chegar até aqui .

A meu pai, por ter tornando muitos de meus sonhos possíveis e por ser um exemplo de homem a se espelhar.

À minha mãe, por ser uma mulher guerreira, batalhadora e por ter me guiado para o lado certo da vida, e por esta comigo nas horas mais difíceis da minha vida, me dando apoio e força pra conseguir vencer.

À minha orientadora, Joseane Pinheiro, por ter depositado uma grande confiança neste trabalho e por ter ajudado na realização do mesmo com críticas e sugestões.

A todos os professores e amigos que, direta ou indiretamente contribuíram para a conclusão deste trabalho.

Sumário

1	Introdução	3
2	Referencial Teórico.....	4
2.1	Histórias Da Bioinformática.....	4
2.2	Tecnologias.....	5
2.3	Projeto genoma	8
3	Metodologia	10
4	Análise dos Resultados	11
4.1	Bancos de Dados na Bioinformática.	11
4.2	GenBank:	13
4.3	PDB:	14
4.4	SWISS-PROT:.....	15
4.5	TrEMBL:	15
5	Conclusão	16
6	Referências Bibliográficas	17

Resumo

Com o passar dos anos a biologia deparou-se com os problemas de tratamento dos dados produzidos por suas pesquisas, no entanto com a evolução da informática ela encontrou uma poderosa aliada que ajudou a solucionar os problemas como armazenamento, organização e velocidade de processamento dos seus dados. Essa junção deu origem a uma nova e importante ciência denominada Bioinformática.

No decorrer da pesquisa mostrara a historia da junção da informática com a biologia, a forma como auxiliou com o tratamento de dados, as dificuldades enfrentadas pela bioinformática e principalmente a importância traga pelos bancos de dados na área, que sem eles seria praticamente impossível armazenarem e manipular de forma tão precisas os dados produzidos.

Palavras chaves: Biologia, bioinformática, banco de dados, Projeto genoma.

ABSTRACT

Over the years biology has encountered problems processing the data produced by their research, however with the evolution of computer she found a powerful ally who helped solve the problems such as storage, organization and processing speed of its dados. Essa junction gave rise to an important new science known as bioinformatics.

During the research had shown the story of the junction of computing and biology, the way helped with data processing, the difficulties faced by bioinformatics and especially the importance to bring the databases in the area, without which they would be practically impossible to store and handle so accurate data produced.

Keywords: biology, bioinformatics, database, genome project.

1 Introdução

Apesar de não ser um campo novo a bioinformática vem em rápido crescimento. Surgiu da necessidade de melhorar as formas de manipular grandes quantidades de dados genéticos e bioquímicos. Várias informações oriundas das diversas áreas da Biologia (tais como Biologia molecular, bioquímica estrutural, enzimologia, fisiologia e patologia), a bioinformática usa o poder computacional para catalogar, organizar e estruturar estas informações em uma entidade compreensiva e extremamente importante para a Biologia.

O boom da área foi o projeto de mapeamento do genoma de diversas espécies (genoma é o conjunto dos genes que regulam todo um organismo biológico). Sem a ajuda da Informática seria impossível uma pesquisa deste tipo, mesmo com organismos mais simples, como fungos.

O uso da informática para trabalhar com grande quantidade de dados não é coisa nova, foi uma das motivações de sua criação. A informática ajuda a biologia, a tratar com a grande quantidade de dados gerada em suas pesquisas, além disso, é muito útil pra fazer simulações.

Assim com a junção da informática com a biologia além de trazer vários benefícios como velocidade de processamento e armazenamento de genes, também contribuiu com a medicina, através da bioinformática hoje temos condições de descobrir curas para varias doenças cujo essas que levariam anos para serem descobertas hoje com a informática conseguimos reduzir esse tempo, por isso venho através deste projeto justificar a importância da bioinformática para o mundo mostrando assim seus benefícios e dificuldades enfrentadas por ela ao decorrer dos anos.

Nesta monografia mostra como os bancos de dados são utilizados pra auxiliar a biologia. Para isso estudou-se as aplicações de bioinformática e observou-se em que tecnologia é usada, avaliou-se como é usada a informática em aplicações da biologia e se estudou mais sobre os bancos de dados, por fim, procurou-se verificar as vantagens trazidas pela informática na biologia, no caso do uso dos bancos de dados.

2 Referencial Teórico

"A bioinformática é uma nova disciplina científica com raízes nas ciências da computação, na estatística e na biologia molecular. A bioinformática desenvolveu-se para enfrentar os resultados das iniciativas de seqüenciamento de genes, que produzem uma quantidade cada vez maior de dados sobre proteínas, DNA e RNA. Desse modo, os biólogos moleculares passaram a utilizar métodos estatísticos capazes de analisar grandes quantidades de dados biológicos, a predizer funções dos genes e a demonstrar relações entre genes e proteínas". (UNIVERSIDADE DE WAGENINGEN apud VOGT, 2003)

2.1 Histórias Da Bioinformática

A história da bioinformática começa na década de 1940, quando foi inventado o moderno computador digital. O primeiro computador moderno começou a funcionar por volta de 1946. Ele se chama digital, pois os dados são armazenados com alfabeto binário, os dígitos binários, zeros e uns. A operação também é digital toda em cima da lógica liga/desliga. (SETUBAL, 2003)

A descoberta de Watson e Crick, em 1953, de que o DNA é estruturado como hélice dupla, trouxe como desdobramento a Bioinformática. Essa descoberta mostrou que a informação genética é armazenada de forma digital, ou seja, a informação genética também é escrita com um alfabeto. Tal alfabeto, entretanto é quartanário diferente do digital do computador que é binário (zero e um). No alfabeto genético são utilizadas quatro letras: A, C, G e T.

Watson e Crick descobriram que a forma dos genes operarem até certo ponto também é digital, ou seja, os genes podem ser "ligados" ou "desligados".

Entretanto foram necessários 42 anos para que a união da biologia com a informática pudesse acontecer. Do lado a informática, a capacidade de armazenamento e processamento precisou aumentar e do lado da biologia, esse tempo foi gasto em entendimento de como o alfabeto podia ser usado para escrever palavras e frases.

Então, através de uma evolução que parece mais ou menos sincronizada, desembocamos em 1995, quando os computadores já estavam suficientemente poderosos para poder processar os milhões e milhões de letras que passaram a vir à luz. E assim nasceu a bioinformática, com a missão de ajudar-nos a entender a história que está escrita nesse livro da vida (SETUBAL, 2003).

O grande desenvolvimento da bioinformática ocorreu com a decodificação dos genes. E nesse ponto podemos destacar que a velocidade foi uma contribuição muito

importante na realização dessa tarefa.

[...]do ponto de vista da velocidade com que a tarefa pode ser realizada, que o Genoma Humano, inicialmente previsto, em 1987, quando foi concebido, para ser desenvolvido e concluído em 15 anos, acabou sendo antecipado, em cerca de 5 anos, tendo em 26 de junho de 2000 sido dado por decifrado em seus aspectos essenciais(VOGT, 2003)

Hoje em dia para termos noção da importância que a informática proporcionou ao desenvolvimento genômico, um gene, com cerca de 12 mil bases pode ter sua sequência decifrada em um minuto, enquanto há três anos atrás a mesma tarefa levaria vinte minutos e há vinte anos, cerca de um ano para ser decifrada.

No Brasil, desde 1999, a bioinformática vem sendo estudada. O primeiro trabalho nessa área foi do sequenciamento completo do DNA da bactéria *Xylella fastidiosa*, patógeno que causava prejuízos à cultura de cítricos, por causar a doença chamada amarelinho.

2.2 Tecnologias

A bioinformática é uma área interdisciplinar que envolve conhecimentos de engenharia de softwares, matemática, estatística, ciência da computação e biologia molecular. Assim iremos falar de algumas tecnologias da área de computação mais utilizada nas pesquisas de biologia.

a) Sistemas operacionais

Segundo TANENBAUM & WOODHULL, 2000, o programa de sistema mais fundamental é o sistema operacional, que controla todos os recursos do computador e fornece a base sobre a qual os programas aplicativos podem ser escritos.

Os sistemas operacionais são responsáveis pelo gerenciamento dos hardwares, ou seja, através dos sistemas operacionais conseguimos fazer funcionar todos componentes de hardwares.

Segundo PROCIMI, a maioria das aplicações de bioinformática são compiladas e distribuídas para a execução em plataformas derivadas do Unix. Esses sistemas são mais confiáveis e tem melhor qualidade no gerenciamento de grandes quantidades de dados. Atualmente, temos o Linux nas mais diversas versões, que é um Sistema Operacional baseado no UNIX e possui código gratuito. As telas iniciais do Unix e do Linux podem ser vistas na figura:



Figura 1: exemplo de telas dos sistemas operacionais UNIX e Linux.

Fonte: Pessoal

b) Linguagem de programação

Segundo Vinicius uma linguagem de programação é um conjunto finito de símbolos com os quais se escrevem programas de computador.

Os profissionais que trabalham na bioinformática, contam com programas aplicativos já desenvolvidos e a demanda por novos programas ainda é grande. Assim é necessário que um profissional seja capaz de desenvolver novos programas para lidar com os mais diversos problemas encontrados durante a análise de dados em biologia molecular.

Tais profissionais chamados de bioinformatas devem ter conhecimento sobre algum tipo de linguagem de programação. “Das mais utilizadas na área, podemos citar: Basic, pascal, C, C++, Java, COBOL e Fortran. Entretanto, a linguagem mais utilizada pelos bioinformatas é, sem sombra de dúvida, o PERL.” (PROCIMI, 2002)

BOX1 - Exemplo de programa PERL para obter a fita reversa-complementar a partir de uma sequência de DNA desejada.

```
#!/usr/bin/perl
# Sequência que se deseja utilizar
$meuDNA = 'TTCCGAGCCAATTGTATCAGTTGCCAATAG';
# Inverte a ordem da sequência de DNA
$RevCom = reverse $meuDNA;
# Troca as bases produzindo a fita complementar
$RevCom =~ tr/ACGT/TGCA/;
print "Minha sequência invertida é: \n $RevCom";
```

Figura2 : Exemplo de Programa em PERL

Fonte: PROSCIMI, 2002.

Na figura 2 vemos um exemplo de programa PERL, onde “A primeira linha é obrigatória e diz ao programa onde se encontra o interpretador PERL para que o mesmo possa achá-lo na hora de sua execução. As linhas seguintes que se iniciam com o Sinal de “#” representa linhas de comentários. As variáveis em PERL são seguidas \$ não precisam ser declaradas, cabe ao programador saber como e em que contexto devem ser utilizadas. Os comandos terminam sempre com ponto-e-vírgula e o sinal de “=~” esta relacionado a utilização de uma expressão regular.(PROSDOCIMI,2002

O PERL (Practical Extract and Re-port Language) é uma linguagem de programação, simples e muito rica, além de disponível gratuitamente. Foi criada por Larry Wall, originalmente para produzir relatórios de informações de erros, que a disponibilizou na Internet no espírito freeware, pensando que alguém pudesse achá-la útil. Ao longo dos anos esta linguagem conquistou milhares de adeptos e, através de várias colaborações recebidas para seu aprimoramento, o PERL é hoje conceituado como uma linguagem sofisticada, que possui como ponto forte a manipulação de texto, mas que, além disso, possui todas as características de uma linguagem de alto-nível genérica. É essa grande facilidade para a manipulação de texto que fez do PERL a linguagem mais utilizada no tratamento de dados de seqüências de DNA e proteínas. (PROSCIMI, 2003)

O PERL é uma linguagem onde suas funcionalidades são acrescidas por módulos, que são distribuídos gratuitamente. Existem módulos para inúmeras aplicações, vão desde aplicações de estáticas clássicas até aplicações 3D. Existem módulos do PERL que são especialmente destinados ao uso de aplicações em bioinformática. Destacando-se **IOPERL E BIOGRAPHICS**, esses dois módulos disponibilizam várias ferramentas úteis para aplicações nessa área.

Uma característica marcante do PERL é a boa interconectividade com banco de dados, essa linguagem atende muito bem a esta demanda, através da biblioteca PERL-DBI, que é um conjunto de módulos que fornece uma interface consistente para solução e integração com banco de dados. Por essa e várias outras características citadas é que o PERL é uma das linguagens preferidas pelos bioinformatas.

c) Banco de Dados

Banco de Dados é um conjunto de Dados inter-relacionados representando informações sobre um determinado domínio, cujo objetivo é registrar e manter informações consideradas significativas a esse domínio.

Devido a grande quantidade de informações que são produzidas atualmente principalmente em projetos Genoma, Transcriptoma e Proteoma. O uso dos bancos de dados, vem assumindo uma posição crescente e importante em projetos de bioinformática, devido sua capacidade de armazenar e organizar os dados. (Prosdocimi, 2002)

Existem diversos tipos de gerenciamento de banco de dados disponíveis no mercado atualmente como o MySQL, PostgreSQL, Microsoft SQLserver. Segundo Prosdocimi o mais utilizado dentre esses três citados pela comunidade acadêmica de projeto genoma, é o MySQL, devido a facilidade de manuseio, sua velocidade de acesso aos dados, por ser gratuito e possuir código aberto.

A comunidade científica tem gasto muito tempo e dinheiro na construção de banco de dados públicos.

Os bancos de dados armazenam seqüências nucleotídeos, de aminoácidos ou estruturas de proteínas, e são classificados em bancos de seqüências primários e secundários.

Os bancos primários são formados pela deposição direta de seqüências de nucleotídeos, aminoácidos ou estruturas protéicas, sem qualquer processamento ou análise. Os principais bancos de dados primários são o GenBank, o EBI (European Bioinformatics Institute), o DDBJ (DNA Data Bank of Japan) e o PDB (Protein Data Bank).

Os bancos de dados secundários, como o PIR (Protein Information Resource) ou o SWISS-PROT, são aqueles que derivam dos primários, ou seja, foram formados usando as informações depositadas nos bancos primários.

Os bancos de seqüências podem ser classificados como bancos estruturais ou funcionais. Os bancos estruturais mantêm dados relativos à estrutura de proteínas. Embora a seqüência de nucleotídeos, a seqüência de aminoácidos e a estrutura de proteína sejam formas diferentes de representar o produto de um dado gene, esses aspectos apresentam informações diferentes e são tratados por projetos diferentes, que resultam em bancos específicos.

Dos bancos funcionais, o KEGG (Kyoto Encyclopedia of Genes and Genomes) é um dos mais utilizados. Disponibiliza links para mapas metabólicos de organismos com genoma completamente ou parcialmente seqüenciado a partir de seqüências e de busca através de palavras-chave.

Atualmente a maioria das revistas exige que as seqüências identificadas pelos laboratórios sejam submetidas a um destes bancos antes mesmo da publicação do artigo.

2.3 Projeto genoma

O Projeto Genoma Humano (Human Genome Project, HGP) é uma das maiores façanhas da história da humanidade. Ele é traduzido como um esforço da pesquisa internacional para seqüenciar e mapear todos os genes dos seres humanos, que no seu conjunto é conhecido como genoma. (Giselda MK Cabello, MSc. Responsável pelo Projeto Fibrose Cística Laboratório de Genética Humana Departamento de Genética/IOC/FIOCRUZ)

O **Projeto genoma humano**, teve seu início por volta de 1990, projetado para durar 15 anos, tinha como **objetivos**: determinar a ordem, ou seqüência, de todas as bases do nosso DNA genômico; identificar e mapear os genes de todos os 23 pares de cromossomos humanos; armazenar essa informação em bancos de dados, desenvolver ferramentas eficientes para analisar esses dados e meios de usar esta informação para

estudo da biologia e da medicina.

O objetivo principal era construir uma serie de diagramas descritivos de cada cromossomo humano, mas para que o objetivo fosse realizado seria necessários dividir os cromossomos em fragmentos menores que possam ser propagados e caracterizados e depois ordenar os mesmos de forma a corresponderem suas respectivas posições nos cromossomos, ou seja, realizar o mapeamento.

Segundo Biello, 2008, ao seqüenciar vários genomas cientistas criaram um mapa da diversidade genética, que pode abrir caminhos para a possível descoberta de raízes genética das doenças.



Figura 3: 1.000 genomas
FONTE: BIELLO, 2008

O Brasil também participa do Projeto genoma, Apesar das tecnologias no Brasil ainda não serem avançadas, o Brasil esta investindo forte nessa área. Com a descoberta do seqüenciamento completo da bactéria *Xyella fastidiosa* patógeno que causa prejuízo a cultura de cítricos(popularmente chamada de doença do amarelinho), o Brasil ocupou um lugar de destaque com Projeto Genoma esse projeto foi tão bem sucedido que o ganhou um lugar de destaque internacional.

3 Metodologia

- Quanto à natureza:

A pesquisa é básica, pois, objetiva apenas discorrer sobre a bioinformática sem aplicação prática prevista. Além de procurar mostrar as vantagens e dificuldades encontradas no uso da informática na biologia.

- Quanto a forma de abordagem:

A pesquisa é qualitativa, segundo Gil, 1991, a pesquisa qualitativa considera que há uma relação dinâmica entre o mundo real e o sujeito, isto é, um vínculo indissociável entre o mundo objetivo e a subjetividade do sujeito que não pode ser traduzido em números. É uma pesquisa descritiva.

Nossa pesquisa irá mostrar o significado da bioinformática mostrando como essa área uniu os recursos da informática à biologia.

- Quanto aos objetivos

Quanto aos objetivos a pesquisa é exploratória e explicativa. Exploratória, pois, visa proporcionar maior conhecimento com o problema visando (a bioinformática) a torná-lo explícito. Envolve levantamento bibliográfico; análise de exemplos que estimulem a compreensão. Explicativa pois tem por fim aprofundar o conhecimento a cerca da realidade (biologia) procurando explicar a razão, o "porquê" das coisas (porque a biologia se associou a informática para criar a bioinformática)

- Quanto aos procedimentos técnicos

A pesquisa é bibliográfica, pois será desenvolvida a partir de material já publicado, constituído principalmente de artigos, e material disponibilizado na Internet. E documental, pois será elaborada a partir de materiais que não receberam tratamento analítico.

4 Análise dos Resultados

4.1 Bancos de Dados na Bioinformática.

Em Rocha, 2007 o pesquisador afirma que

“A quantidade de bancos de dados em Biologia Molecular vem crescendo exponencialmente nos últimos anos, e o aspecto funcional da bioinformática é a representação, o armazenamento e a distribuição de dados. Os objetivos destes bancos variam e podem ser utilizados para armazenar e disponibilizar bioseqüências, funções moleculares, estruturas de proteínas, modelos metabólicos, entre outros, oferecendo em alguns casos informações mais amplas, ou seja, cobrindo uma área mais significativa da Biologia, e em outros, informações menos detalhadas.”

Ainda em Rocha, 2007 no início do projeto genoma com o surgimento da bioinformática, as biosseqüências eram tratadas em cadeias de textos, e por este motivo, seu armazenamento eram realizados em cadeias de flats files (Arquivos textos).

Mas em nossa pesquisa nada se fala sobre como os bancos de dados biológicos eram elaborados e manipulados antes do computador. Supõem-se que deveria ser bastante difícil o trabalho com eles antes do advento do computador, se é que existiam.

Todas as pesquisas apontam o Projeto Genoma como o maior incentivador do uso da informática na biologia, bem como do uso dos bancos de dados. Falcão, 2005, afirma que:

“quando os pesquisadores isolam novas seqüências em laboratório, eles buscam obter a maior quantidade de informação possível sobre essas seqüências. Mas, devido ao tamanho das seqüências e a complexidade em analisá-las, esse processo foi se tornando cada vez mais difícil de ser realizado de forma manual”.

Foi assim que se passou a procurar métodos computacionais que pudessem ser usados para fazer análise das seqüências de DNA e identificação dos genes. Além de trabalhar com um número enorme de dados, as formas de armazenamento usadas até então não eram boas para o problema e as pesquisas na área começaram a torna-se cada vez mais difíceis implicando assim para o surgimento dos gerenciadores de Bancos de Dados específicos para área.

Em alguns casos, as informações biológicas são obtidas por análise computacional de outros bancos de dados; já em outros casos, através da literatura ou informações definidas por pesquisadores.

Pela importância do Banco de Dados nas pesquisas em biologia vamos apresentar uma abordagem geral sobre a importância destes bancos de dados biológicos, sua estruturação, organização e busca de informações nos principais bancos existentes na área de bioinformática.

Hoje na Bioinformática os Bancos de Dados estão automatizados. Entretanto ainda é um O grande desafio a melhor forma de armazenamento, para os dados gerados por projetos de pesquisa na área da biologia molecular, como o projeto genoma humano, que possui centenas de dados a espera para serem armazenados e tratados. A partir dessa necessidade de armazenar e tratar os dados vem surgindo uma enorme quantidade de banco de dados voltado pra essa área da biologia.

“Existem basicamente dois tipos de bancos de dados disponíveis para utilização na pesquisa de genes e proteínas” (Baxevanis & Ouellette 2001). Esses bancos são chamados respectivamente de primários e secundários, onde o banco primário é responsável por apresentar resultados de dados experimentais que são publicados com alguma interpretação (comunicação com outros bancos), mas não havendo uma análise aprofundada desses dados. É o exemplo do GenBank, EMBL e PDB (Protein Data Bank). Já os secundários ao contrário, são aqueles onde existe uma compilação (reunir várias informações de outros bancos) e interpretação dos dados de entrada por um ou mais cientistas, de tal forma que podem ser obtidos os mais representativos e interessantes. São eles COG, SWISS-PROT e o TrEMBL.

Segundo Falcão, 2005 de uma forma geral existem várias vantagens em se trabalhar usando banco de dados como:

- Velocidade/Eficiência: Os dados são obtidos e atualizados rapidamente no BD;
- Concorrência: diferentes aplicações/usuários podem compartilhar os mesmos dados;
- Redundância controlada: dados compartilhados não precisam ser replicados;
- Aplicabilidade: Diminuição no tempo de desenvolvimento da aplicação.

A grande dificuldade enfrentada quando se fala em banco de dados é que poucos têm capacidade para lidar com dados complexos, como dados multimídias, texto, dados espaciais, ou dados genéticos. A dificuldade desses dados está em serem controlados, como questões localização de genes em cadeias de DNA e sucessões de genes. A **figura**

4 mostra as etapas de armazenamento de segmentos (sucessões de genes) de DNA em um banco de dados.

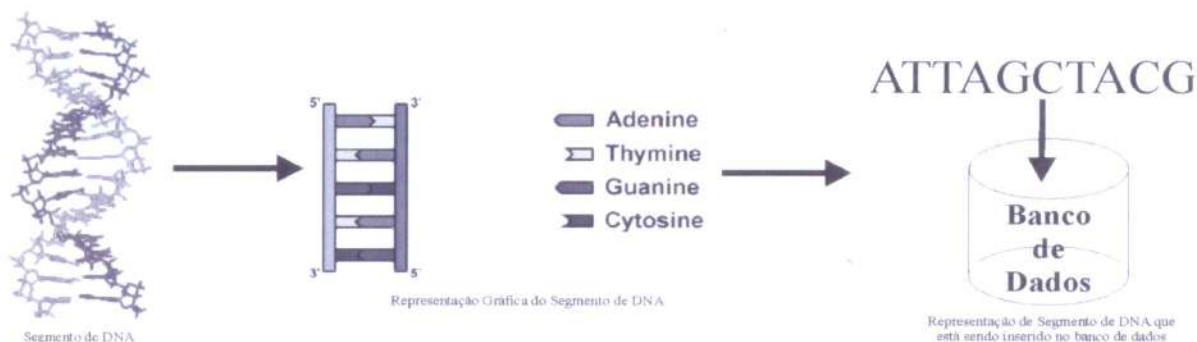


Figura 4. Etapas realizadas para armazenar um segmento de DNA em um banco de dados.Fonte: Wieczorek,2002

No Brasil a grande dificuldade hoje encontrada é o alto custo para implantar um sistema de gerenciamento de banco de dados, pois os sistemas hoje existentes no mercado são muito caros tornando viável apenas para grandes empresas. Além de não serem encontrados com facilidade no mercado.

Vamos falar agora sobre alguns exemplos de bancos de dados utilizados na área da bioinformática:

4.2 GenBank:

Segundo Falcão, 2005 “é o banco de dados de sequência de DNA mais conceituado hoje no mundo, mantido pelo National Center for Biotechnology Information (NCBI) da National Library of Medicine (NLM). Ele foi estabelecido em 1978 como um repositório central para os dados de sequência de DNA.”

O GenBank é mantido como uma combinação de arquivos de textos padronizados, bancos relacionais e arquivos que contém notação de Sintaxe Abstrata, (utiliza uma linguagem diferente da linguagem de programação, devido a necessidade de especificação de estruturas de dados de forma abstrata). A cada acesso ao GenBank é atribuído um identificador único pelo NCBI. Novos identificadores são relacionados às atualizações, só que o identificador da entidade original permanece inalterado para a proposta de arquivamento.

Usuários comuns não conseguem acessar as estruturas dos dados diretamente por meio de consultas ou outras funções. A maneira de se fazer consulta é fornecida por meio da aplicação *Entrez* que permite a busca por palavra-chave, a sequência e identificadores únicos do GENBANK por uma interface estática.

O GenBank não é um banco de dados curados ou seja as informações podem possuir inconsistência.

4.3 PDB:

“Criado em 1971 pelos Drs. Edgar Meyer e Walter Hamilton do Brookhaven National Laboratory, a gestão foi transferida em 1988 ao Research Collaboratory for Structural Bionformatics (RCSB). Rutgers University é a sede principal e estar atualmente dirigido por Helen M. Berman”. (Beman, 2006)

O Protein Data Bank consiste em uma organização que atua como centros de armazenamento, processamento e distribuição de dados. Esse é um banco de dados de estrutura de proteína e outras macromoléculas. Responsável por fornecer detalhes experimentais sobre a geração da estrutura, atribuições da estrutura, coordenações atômicas entre outras.

Segundo Ferreira, 2005 este banco de dados armazena também estruturas 3D de proteínas e Macromoléculas.

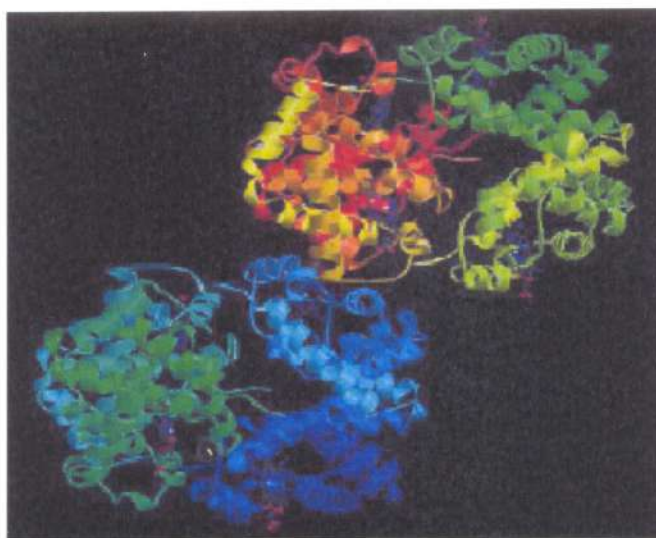


Figura 5: Proteína 3D

Fonte: Ferreira, 2005

4.4 SWISS-PROT:

Este banco foi criado em 1986 pelo Departamento de Bioquímica Médica da Universidade de Genebra e EMBL. É mantido pelo Swiss Institute of Bioinformatics (SIB) e EBI/EMBL.

O SWISS-PROT mantém um alto nível de anotações, como a descrição e a função da proteína, estrutura dos seus domínios, modificações pós-tradução, além de ter uma estrutura que facilita o acesso computacional a diferentes campos de informações.

4.5 TrEMBL:

O TrEMBL é um suplemento de anotações do SWISS-PROT nele está contido todas as informações de sequência de nucleotídeos do Embl ainda não disponível no SWISS-PROT. As entradas desse banco são menores e menos anotadas do que a do SWISS-PROT, mais são transferidas para o mesmo assim que uma anotação confiável seja disponível.

Os bancos de dados são essenciais para uma boa organização e armazenamentos dos dados, No decorrer do tempo percebeu que seria praticamente impossível armazenar e organizar essa enorme quantidade de dado produzido pelos genes, por isso se tornou tão importante os bancos de dados.

5 Conclusão

Após pesquisa realizada vimos que a informática esta presente nas mais diversas áreas do conhecimento inclusive na biologia. A junção da informática com a biologia deu origem a uma nova ciência denominada Bioinformática, trazendo com ela inúmeras vantagens como a velocidade de processamento de dados, armazenamento e organização de dados, entre outras, com isso vimos resultados que levariam anos para serem descobertos hoje levam meses e ate dias.

Ver-se também a grande importância e a evolução dos bancos de dados voltados para essa área, e que sem os bancos de dados seria impossível hoje a organização e armazenamento precisos das informações.

Os bancos de dados são devidos em dois tipos os primários e secundários, onde no primários os resultados publicados são experimentais. Já os secundários os resultados passam por uma análise mais aprofundada com isso obtém resultados mais interessantes e representativos.

Devido à grande quantidade de dados produzidos pelos projetos de bioinformáticas foram criados bancos de dados específicos onde estes bancos armazenam, organizam e disponibilizam essas informações na internet facilitando assim o acesso as informações para os bioinformátas.

Com isso concluímos que a informática é uma ferramenta extremamente poderosa e que aliada à biologia é essencial para o desenvolvimento de pesquisas da cura dos mais variados tipos de doenças, agilizando resultados e beneficiando o mundo.

6 Referências Bibliográficas

BIELLO, David. **Projeto dos 1.000 Genomas: expandindo o mapa da genética humana**. Scientific American Brasil. 28 de janeiro de 2008. Disponível em: <http://www2.uol.com.br/sciam/noticias/projeto_dos_1_000_genomas_expandindo_o_mapa_da_genetica_humana.html>

CABELLO, Giselda M.K.. **Genoma Humano: Conceito e Linguagem**. Laboratório de Genética Humana Departamento de Genética/IOC/FIOCRUZ. Disponível em: <<http://www.ghente.org/ciencia/genoma/index.htm>>. Acessado em: 20 de julho de 2010.

*CASTRO, Fábio de. **Bioinformática faz 10 anos no Brasil**. FAPESP, 2009.

PROSDOCIMI, Francisco et al. Bioinformática: Manual do Usuário. **Biotecnologia Ciência & Desenvolvimento**, Ano V, nº 29, novembro/dezembro, 2002. Disponível em: <<http://www.biotecnologia.com.br/edicoes/ed29.php>>

SETUBAL, João Carlos. A origem e o sentido da bioinformática. **Com Ciência: Revista Eletrônica de Jornalismo Científico**. SBPC. Agosto, 2003. Disponível em : <<http://www.comciencia.br/reportagens/bioinformatica/bio01.shtml>>

TANENBAUM, Andrew S.; WOODHULL, Albert S.. **Sistemas Operacionais: Projeto e Implementação**. Tradução Edson Furmankiewicz. 2ª Edição. Porto Alegre, Bookman, 2000.

VOGT, Carlos. Bioinformática, genes e inovação. **Com Ciência: Revista Eletrônica de Jornalismo Científico**. SBPC. Agosto, 2003. Disponível em : <<http://www.comciencia.br/reportagens/bioinformatica/bio01.shtml>>

Vinicius, Marcus Analista Técnico de Informática da SUSEP - Superintendência de Seguros Privados. Professor de Informática do Curso Debret, no Rio de Janeiro. Mestre em Engenharia de Sistemas e Computação pela COPPE-UFRJ. Abril de 2002 em: <http://www.vemconcursos.com/opiniaio/index.phtml?page_id=440>

Helen M. Berman». Rutgers Chemistry and Chemical Biology Department disponível

em:

<http://pt.wikilingue.com/es/Protein_Data_Bank#Refer.C3.AAncias>

WIECZOREK, Emilio Mario. Caminhos e Tendências do Uso de Bancos de Dados em Bioinformática. Prática em Sistemas de Informação I - Estágio. Palmas-TO, 2002

Ferreira, Rodrigues. Introdução a bioinformática. Centro de Biologia Genômica e Molecular (PUCRS), Instituto de Informática (UFRGS), Uberaba, junho de 2005.

ROCHA, Cícero Pinho. Banco de Dados em Bioinformática. Pós-graduação em Análise de Sistemas. Faculdade Piauiense, 2007

FALCÃO, Carlos Roberto et AL , Banco de Dados Biológicos. UNAMA – Universidade da Amazônia, 2005 Belém – PA – Brasil. Disponível no Site: <http://www.ead.informatica.com/capacita2/index.php?option=com_content&view=article&id=60:banco-de-dados-biologicos&catid=43:informatica&Itemid=77>.