



## INTRODUÇÃO À BIOINFORMÁTICA

Rodrigo Mendes

<http://rodrigomendes.tripod.com>

A Bioinformática é a união da Biologia e da Computação envolvendo diversas linhas do conhecimento, tais como: biologia molecular, matemática, estatística, engenharia de *softwares* e ciência da computação. A Bioinformática gera diversos *softwares* e modelos que auxiliam o estudo e análise de dados moleculares. Segue um roteiro de tarefas para o conhecimento de algumas ferramentas de estudos moleculares *in silico*.

• **DIGESTÃO VIRTUAL DE FRAGMENTOS DE DNA** • <http://www.restrictionmapper.org>

Colocar a seqüência de interesse no campo **Sequence Info** e selecionar os parâmetros nas áreas **Conformation** e **Include**. Proceder com a análise clicando em **Map Sites** e **Virtual Digest**.

• **NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION** • <http://www.ncbi.nlm.nih.gov/>

**About NCBI** apresentação do *site*.

**TaxBrowser** verificar a classificação taxonômica de um organismo. Segue exemplos.

1. Deixar campo **Search for** vazio e clicar em **Go**, então surgirão *links* para o nível máximo de classificação dos seres vivos e para uma lista de alguns organismos mais utilizados em pesquisas moleculares.
2. Digitar *Homo sapiens* no campo **Search for** e então verificar saída e *links* complementares.

**Map Viewer** (na área **Hot Spots**) verificar e fazer buscas de cromossomos, genes e seqüências nos organismos com o genoma seqüenciado. Para visualizar os organismos como lista clique em **Switch to List View** ou como gráfico clique em **Switch to Graphical View**. Segue exemplos.

1. Selecionar **Homo sapiens (human)** no link **Mammals**. Surgirá o idiograma do genoma humano. Clicar no cromossomo **Y**. Surgirá o cromossomo em detalhes. Clicar no gene **USP9Y** e verificar estrutura, contexto genômico e comentários. Verificar informações bibliográficas clicando nos *links PubMed*. Verificar seqüência de nucleotídeos ou aminoácidos na área **NCBI Reference Sequences (RefSeq)** no link **NM\_004654**.
2. Selecionar **Arabidopsis thaliana (thale cress)** no link **Plants**. Surgirá o idiograma do genoma de *A. thaliana*. Clicar em **CHL**. Sugirá o genoma cloroplastidial da planta com *links* para as diferentes regiões.

**Structure** verificar estruturas protéicas. Segue exemplo.

1. Digitar **aconitase** no campo **Search Entrez (Structure) for**. Surgirá uma lista de seqüências, clicar no primeiro link **1:\_\_\_**. Surgirá a descrição da seqüência e a possibilidade de gerar a estrutura tridimensional da proteína selecionando os parâmetros de visualização e clicando-se em **View 3D Structure** (nota: para a visualização da estrutura 3D o computador deverá possuir o *software Get Cn3D 4.1!* - link disponível para *download*).

**Blast** verificar a similaridade de uma seqüência de interesse com seqüências depositadas no banco de dados GenBank. Segue exemplos.

1. Copiar seqüência de nucleotídeos do gene **USP9Y** utilizado no exemplo 1 do **Map Viewer**. No campo **Nucleotide** clicar em **Nucleotide-nucleotide BLAST (blastn)**. Sugirá tela para o BLAST. Colar seqüência de interesse no campo **Search**. Clicar no botão **BLAST!**. Surgirá nova janela, clicar no botão **FORMAT!** Aguardar resultado. Surgirá a janela de resultados apresentando as seqüências mais similares do banco de dados.
2. Fazer um BLAST digitando uma seqüência aleatória de 20 nucleotídeos. Provavelmente haverá nenhuma ou baixa similaridade de seqüências, o que indica que a seqüência dos nucleotídeos no DNA não é aleatória.

• **CONSTRUÇÃO DE PRIMERS *Primer3*** • <http://frodo.wi.mit.edu>

Colocar a seqüência de interesse (a qual se deseja detectar/amplificar pela PCR utilizando-se os *primers* a serem construídos) no campo de busca. Buscar seqüência de *Escherichia coli* no NCBI, digitar **AY669291** no campo **Search (Nucleotide) for** e **Go**, e então copiar seqüência. São diversos os parâmetros para a obtenção de *primers* de acordo com o interesse, neste caso utilizar o *default* (não alterar os parâmetros) e clicar em **Pick Primers**. Surgirá uma nova janela indicando as características, posição e constituição dos *primers forward* e *reverse*.

• **ALINHAMENTO MÚLTIPLO DE SEQÜÊNCIAS E CONSTRUÇÃO DE ÁRVORE FILOGENÉTICA**

**ClustalW** • <http://www.ebi.ac.uk/clustalw> (Alinhamento e Cladograma)

Digitar seu *e-mail* no campo **YOUR EMAIL**. Utilizar os parâmetros de acordo com o *default*. Colocar as seqüências de interesse no campo **Enter or Paste a set of Sequences in any supported format** e clicar em **Run** (nota: podem ser utilizadas seqüências de nucleotídeos ou aminoácidos). Aguardar resultado e visualizar cladograma de distâncias. Utilizar o exemplo do tutorial no *site* <http://www.ebi.ac.uk/2can/tutorials/protein/clustalw1.html>.

**GeneBee** • [http://www.genebee.msu.su/services/phree\\_reduced.html](http://www.genebee.msu.su/services/phree_reduced.html) (Cladograma)

Colocar as seqüências alinhadas no campo **Alignment** (nota: podem ser utilizadas seqüências de nucleotídeos ou aminoácidos). Na área **Tree and Picture Options** selecione **Phylogram** e **(Bootstrap) Yes**. Coloque seu *e-mail* no campo **Your E-mail address** e clique em **Enviar Consulta**. Em instantes surgirá uma nova tela com os resultados, os quais também serão enviados automaticamente para seu *e-mail*. Utilizar o exemplo do campo **Alignment Example**.

**ATIVIDADES**

1. Escolher um organismo que tenha seu genoma completamente seqüenciado, citar sua classificação taxonômica completa, buscar um gene neste organismo, descrever sua localização no genoma e encontrar um trabalho científico (*paper*) relacionado ao gene.
2. Buscar seqüências nucleotídicas do rDNA de 3 organismos eucariotos e 3 procariotos, fazer o alinhamento múltiplo e construir a árvore filogenética.

**UMA HISTÓRIA REAL**

No ano de 1998 em Lafayette, Louisiana EUA, ocorreu um caso de alta repercussão. Uma mulher denunciou que seu ex-amante, um médico, deliberadamente lhe havia injetado sangue contaminado com HIV. Ela não havia percebido que tinha sido injetada com sangue e muito menos sabia de quem era o sangue contaminado. Fez a denúncia após alguns meses quando adoeceu manifestando os sintomas da AIDS. Registros mostraram que o médico havia retirado sangue de um paciente com HIV positivo no dia em que ela foi injetada, mas não havia registros de sua injeção e nenhuma testemunha. Então, como sua história poderia ser provada?

Árvores filogenéticas são a melhor evidência científica para um caso como este. Existe uma taxa de mutação do vírus HIV mesmo dentro de um indivíduo. Se uma pessoa contamina uma outra pessoa, existirá poucas diferenças entre o vírus no doador e no receptor. Como o vírus caminha de pessoa para pessoa, ele se mantém alterando, ou seja, acumulando mutações tornando-se mais diferente com o passar do tempo. Então, as seqüências nucleotídicas do HIV em dois indivíduos que contraíram o vírus de duas pessoas diferentes serão bem diferentes. Assim, se a história contada pela mulher fosse verdadeira, seu vírus seria muito semelhante ao vírus do paciente, que teve seu sangue retirado pelo médico e muito diferente de outros vírus de outras pessoas que viviam em Lafayette. E foi exatamente o que a árvore evolucionária mostrou, seu vírus parecia ter vindo do paciente do médico acusado e mostrava muita diferença em relação aos outros vírus de outras pessoas da cidade. Considerando que não havia nenhuma outra explicação para a mulher ter se contaminado com o sangue do paciente, a análise evolucionária do vírus apoiou sua história. (nota: este foi o primeiro caso no qual se usou filogenia no tribunal americano)