

**Instituto de Química – Universidade de São Paulo – professor João Carlos Setubal**

**Roteiro para aula prática de bioinformática da disciplina QBQ102**

**Lista de genomas (são todos de bactérias ou arqueias)**

*Azotobacter vinelandii*

*Xanthomonas citri*

*Bacillus subtilis*

*Xylella fastidiosa 9a5c*

*Archaeoglobus fulgidus*

*Sulfolobus solfataricus*

*Thermotoga maritima*

*Corynebacterium glutamicum*

*Clostridium acetobutylicum*

*Escherichia coli K12*

*Pseudomonas aeruginosa*

*Sinorhizobium meliloti*

*Rickettsia prowazekii*

*Chlamydia trachomatis*

*Treponema pallidum*

*Mycoplasma pneumoniae*

*Leifsonia xyli*

*Leptospira interrogans*

*Mycobacterium tuberculosis H37Rv*

*Methanococcus jannaschii*

*Thermoplasma volcanium*

*Thermobispora bispora*

## Como buscar um genoma

Na home page do NCBI (<http://www.ncbi.nlm.nih.gov>) escolha 'genome' no menu da caixa de busca, e digite o genoma que você quer buscar, *atentando para não errar a grafia*.

Na página desse genoma, encontre o link **browse the list** e clique nele.

Vai aparecer uma tabela (A) com vários genomas dessa espécie. Fixe-se no primeiro genoma da tabela.

Para poder navegar no genoma, clique no link da coluna **replicons** que começa com **NC** e que seja de um cromossomo

Na próxima página, clique no link **Graphics**

Para poder ver a lista de proteínas desse genoma, na tabela (A) clique no link da coluna **proteins** (que é um número)

## Lista de genes a serem buscados (com ctrl-f no navegador)

### Replicação

DNA polymerase  
primase  
gyrase  
helicase  
ligase

### Transcrição

RNA polymerase  
transcription factor

### Tradução

<Aminoacyl>-tRNA synthetase sendo <Aminoacyl> = Histidyl, Methyonil, Seryl, Leucyl, Valyl, etc

ribosomal protein (há diferentes tipos)

## BLAST de genes importantes em processos básicos de biologia molecular

Nesta parte, cada aluno deverá realizar BLASTp com as sequências dos genes (em aminoácidos, ou seja, as sequências das **proteínas**) que achou na parte anterior contra os mesmos genes de outro genoma (sua escolha), e reportar os alinhamentos encontrados.

De posse da sequência de aminoácidos de um gene (em formato FASTA) de um genoma (o *genoma consulta*), acesse a página de BLASTp (faça google blast, ou procure na aba *sequence analysis* na home page do NCBI).

Insira a sequência na caixa **Enter Query Sequence**

No quadro **Choose Search Set** e no subquadro **Organism** digite o nome do outro organismo contra o qual você vai rodar o BLAST. Esse será o *genoma alvo*.

Em seguida rode BLAST apertando o botão.

### **Relatório**

**Deve ser um relatório por grupo de alunos num determinado computador.**

**Num único email para o professor, informe no começo do email os nomes dos alunos que constituem a equipe.**

Apresente uma lista de pelo menos 3 genes com as seguintes informações para cada gene:

nome do organismo / genoma utilizado

locus tag do gene

coordenadas genômicas e a fita (mais ou menos)

nome do produto desse gene

sequência em aminoácidos

Em seguida descreva em seu relatório o resultado da busca por BLAST para cada um dos genes escolhidos. Inclua informações apenas do melhor alinhamento obtido. Para cada alinhamento (de cada um dos 3 genes) inclua as seguintes informações:

qual o gene consulta (pode ser pelo locus tag já informado)

Nome do genoma alvo

O alinhamento

**Inclua comentários** sobre a *qualidade* e *cobertura* desse alinhamento. Cobertura significa verificar quanto das sequências do gene consulta e do gene alvo participam efetivamente do alinhamento. A qualidade é dada pelo e-value e pelo porcentual de identidade.

Caso a busca por BLAST não resulte em alinhamentos, vc deve variar o genoma alvo ou o gene consulta. Todos os alinhamentos reportados tem que ter e-values melhores do que  $1e-5$ .