

## Prática de Bioinformática – Identificação taxonômica com o programa BLAST

### Introdução

A taxonomia de bactérias baseada em características fenotípicas morfológicas, bioquímicas e fisiológicas apresenta diversas limitações. Por outro lado o uso de sequências biológicas de marcadores genéticos tem permitido a identificação taxonômica e análise de relações evolutivas (filogenia) de forma mais ampla e acurada. O principal marcador genético utilizado para este fim, atualmente, é o gene que codifica para o RNA ribossomal de 16S, identificado como **gene 16S rRNA**. Este gene gera uma molécula de RNA na célula que se liga a diversas proteínas para formar a subunidade menor do ribossomo. Dois outros genes, identificados como 23S rRNA e 5S rRNA, também geram moléculas de RNA que se ligam a proteínas para formar a subunidade maior do ribossomo. A molécula do ribossomo (formado pela união das subunidades maior e menor) participa do processo de tradução, decodificando a informação nos RNA mensageiros (mRNA) em sequências de aminoácidos das proteínas.

Os genes rRNA estão distribuídos universalmente entre as células procarióticas (domínio das bactérias e arqueias) e possuem equivalentes nas células eucarióticas. Além disto, estes são genes com a sequência de DNA conservada ente os diferentes níveis taxonômicos. Quanto maior a distância evolutiva entre os organismos, maior a diferença de bases nas sequências de DNA dos genes rRNA destes organismos. Os três genes, normalmente, ocorrem juntos em um operon no genoma dos procariotos e possuem cerca de 1.500pb (16S rRNA), 3.000pb (23S rRNA) e 120pb (5S rRNA). Destes, o gene 16S rRNA é o mais amplamente utilizado para identificação taxonômica e análises de relações evolutivas entre organismos procarióticos. Este gene possui regiões altamente conservadas, mesmo entre organismo distantes evolutivamente, intercaladas por regiões variáveis/hipervariáveis. As regiões conservadas permitem gerar iniciadores universais (que flanqueiam as regiões variáveis) que podem ser usados para amplificação do gene por técnica de PCR a partir de DNA de organismos conhecido ou mesmo desconhecidos, enquanto as regiões variáveis (nove nos procariotos, identificadas de V1 a V9) permitem a diferenciação taxonômica a partir da comparação das sequências de DNA.

Atualmente existem bancos de dados dedicados, principalmente, ao gene 16S rRNA, sendo os principais RDP ([rdp.cme.msu.edu](http://rdp.cme.msu.edu)), Silva ([www.arb-silva.de](http://www.arb-silva.de)) e rrnDB ([rrnodb.umms.med.umich.edu](http://rrnodb.umms.med.umich.edu)), mas sequências deste gene também podem ser obtidas a partir dos bancos de dados ordinários, como o banco de dados de nucleotídeos do NCBI ([www.ncbi.nlm.nih.gov/nucleotide](http://www.ncbi.nlm.nih.gov/nucleotide)). Estes bancos de dados disponibilizam as sequências de DNA, relacionam a sequências de DNA a taxonomia, facilitam a navegação na hierarquia taxonômica e disponibilizam ferramentas de análise a partir das sequências de DNA.

### Objetivo

Identificação taxonômica de bactéria a partir de sequência de DNA do gene marcador 16S rRNA

### Problema prático

Um laboratório de microbiologia veterinária isolou uma bactéria a partir de uma amostra de líquido sinovial da articulação do joelho de frangos de corte, as aves apresentavam depressão, prostração e inflamação da articulação. O DNA da bactéria isolada foi extraído, o gene 16S rRNA foi amplificado utilizando um par de iniciadores universais e, posteriormente, o amplificado foi submetido ao sequenciamento de DNA. A sequência que foi obtida é mostrada abaixo, no item “**Conjunto de dados**”.

### Procedimento prático

#### Análise 1

1. Acesse a ferramenta **BLAST/NCBI on line** no endereço abaixo  
**<http://blast.ncbi.nlm.nih.gov>**  
Clique no quadro verde com a opção “**Nucleotide BLAST**”  
Uma nova janela abrirá com um formulário para a análise do **BLAST**

#### 2. Análise da sequência

1. Copie a sequência problema, fornecida, e cole na página de formulário do **BLAST**, no campo “**Enter accession number(s), gi(s), or FASTA sequence(s)**”

Centro Politécnico – Caixa Postal 19046 – CEP: 81531-990 – Curitiba/PR  
Telefones: (41) 3361-1660 – Fax (41) 3266-2042 – [bioqui@ufpr.br](mailto:bioqui@ufpr.br) / [npca@ufpr.br](mailto:npca@ufpr.br)  
[www.ufpr.br](http://www.ufpr.br)

2. No quadro “**Choose Search Set → Database**”  
→ clique no botão e no *menu* que aparecerá selecione “**16S ribosomal RNA sequences (Bacteria and Archaea)**”
3. No final da página, ao lado do botão “**BLAST**”  
→ selecione a opção “**Show results in a new window**”  
→ clique no botão “**BLAST**” para realizar a análise  
OBS.: O resultado aparecerá em uma nova aba no navegador
3. Análise dos resultados
  1. Verifique a lista de organismos encontrada
  2. Role a página até o alinhamento para o primeiro resultado  
Qual o comprimento do alinhamento?  
Qual a porcentagem de identidade entre as sequências?  
Quantas diferenças de bases existe entre as duas sequências? Identifique algumas.  
O alinhamento cobre todo o comprimento de ambas as sequências?  
A direção do alinhamento entre as duas sequências é a mesma?
  3. Volte a tabela de resultados e clique no *link* para a última coluna do primeiro organismo  
OBS.: A sequência do banco de dados será mostrada em uma nova página  
No menu à direita, clique no *link* “**Taxonomy**”  
Na nova página aberta, clique no *link* para o nome do organismo  
Qual a classificação completa do organismo nos diferentes níveis taxonômicos?
  4. Volte ao início da página e em “**Other reports:**”, clique no *link* “[**Taxonomy reports**]”  
OBS.: O resultado será mostrado em uma nova aba no navegador em uma lista taxonômica  
Quanto gêneros de bactérias estão representados?  
A qual(is) grupo(s) taxonômico(s) pertencem estes gêneros?  
Para qual gênero ocorreu o maior número de resultados?
  5. Volte a página de resultados e em “**Other reports:**”, clique no *link* “[**Distance tree of results**]”  
OBS.: O resultado será mostrado em uma nova aba do navegador em uma árvore filogenética  
Em “**Sequence Label**”, clique no botão e selecione “**Taxonomic Name (if available)**”  
Identifique a sequência problema e qual(ais) organismo(s) está(ão) mais próximo(s)  
Coloque o ponteiro do *mouse* sobre o triângulo verde mais próximo da sequência problema  
No *menu* que será mostrado, selecione “**Expand**”  
Os organismos são os mesmos mostrados na página inicial de resultados?  
O agrupamento observado na árvore tem coerência taxonômica?  
Clique no botão “**Tools**”, na parte superior direita do quadro da árvore, e selecione no *menu* “**Layout**” e teste as visualizações para “**Radial tree**” e “**Circular tree**”

## Análise 2

4. Análise da sequência
  1. Volte a página de formulário do **BLAST**
  2. Em “**Database**”, escolha “**16S ribosomal RNA sequences (Bacteria and Archaea)**”
  3. No campo “**Organism**”, inclua “**Firmicutes**” (selecione a partir do *menu* que será mostrado ao digitar)
  4. Selecione a opção “**Exclude**”
  5. Em “**Limit to**”, marque a opção “**Sequences from type material**”
  6. No final da página, clique no botão “**+ Algorithm parameters**”
  7. Nas opções que serão mostradas, em “**General Parameters → Max target sequences**”, selecione o valor “**500**”
  8. Clique no botão “**BLAST**” para fazer a análise

## 5. Análise do resultado

1. Verifique a lista de organismos na tabela de resultados
2. Clique no *link* para a descrição do primeiro organismo na tabela de resultado  
Qual o organismo, o comprimento do alinhamento e a identidade?
3. Volte na tabela de resultados e clique no *link* para a descrição do último organismo  
Qual o organismo, o comprimento do alinhamento e a identidade?  
Compare estes resultados com a primeira análise do **BLAST**
4. Volte ao início da página e clique no *link* para “[**Taxonomy reports**]”  
Quais grupos taxonômicos estão representados?  
Volte a página de resultados e clique no *link* para “[**Distance tree of results**]”  
Análise a árvore e compare com aquela obtida para a primeira análise do **BLAST**

## Material/Programas

BLAST. Disponível *on line* <<http://blast.ncbi.nlm.nih.gov>>

## Conjunto de dados

Sequência do gene 16S rRNA de organismo desconhecido.

>16S\_rRNA\_unknown\_organism

```
aggatgaacgctggcggcgtgcctaatacatgcaagtcgagcgaacggacgagaagcttgccttctctgatg
ttagcggcggacgggtgagtaaacacgtggataacctataagactgggataacttcgggaaaccggag
ctaataccggataaatatgttgaaccgcatgggtcaaaagtgaagacggcttgcgtgacttatagatgg
atccgcgctgcattagctagttggttaaggaacggcttaccaggcaacgatgcatagccgacctgagagg
gtgatcggccacactggaactgagacacggctccagactcctacgggaggcagcagtagggaatcttccgca
atgggcgaaagcctgacggagcaagccgctgagtgatgaaggcttccggatcgtaaaactctgttattag
ggaagaacatatgtgtaagtaactgtgcacatcttgacgggtacctaatacagaagccacggctaactacgt
gccagcagccgcggtaatacgtaggtaggcaagcgttatccggaattattgggcgtaagcgcgctaggcg
gttttttaagtctgatgtgaaagcccacggctcaaccgtggagggtcattggaactggaaaacttgagtg
cagaagaggaaagtggaattccatgtgtagcggtgaaatgCGCagagatatggaggaacaccagtgccgaa
ggcgactttctggctctgtaactgacgctgatgtgCGaaagcgtggggatcaaacaggattagataccctgg
tagtccacgccgtaaacgatgagtgcctaagtgttagggggtttccgccccttagtgctgcagctaacgcat
taagcactccgcctggggagtacgaccgcaaggttgaaactcaaggaattgacggggaccgcacaagcg
gtggagcatgtggtttaattcgaagcaacgcgaagaacctaccaatttgacatcctttgacaactctag
agatagagccttccccttcgggggacaaagtgacaggtggtgcatggttgctcgtcagctcgtgctgtaga
tggtgggttaagtcccgcacgagcgaacccttaagcttagttgcatcattaaagtgggcactcctaagt
tgctgcccgtgacaaaccggaggaaggtggggatgacgtcaaatcatcatgccccttatgatttgggctac
acacgtgctacaatggacaatacaaggcagcgaaccgcgaggtcaagcaaatcccataaagttgttct
cagttcggattgtagctgcaaccgactacatgaagctggaatcgctagtaatcgtagatcagcatgctac
gggtaatacgttcccgggtctgtacacaccgctcaccacgagagtttgtaaccccgaagccggtg
gagtaccttttaggagctagccgtcgaaggtgggacaaatgattgggtgaagtcgtaacagct
```

## Referências