**Instituto de Ciências Biomédicas / Universidade de São Paulo**

**Elisabete José Vicente**

pasted-image.tiff

**Introdução à Bioinformática - do DNA à proteína (Roteiro de aulas práticas)**

pasted-image.tiff

**Fig**: **Dogma central da Biologia:** O fluxo de informação genética do [DNA](https://www.ncbi.nlm.nih.gov/books/n/mboc4/A4754/def-item/A5084/) para o [RNA](https://www.ncbi.nlm.nih.gov/books/n/mboc4/A4754/def-item/A5756/) (**transcrição**) e do RNA para a [protein](https://www.ncbi.nlm.nih.gov/books/n/mboc4/A4754/def-item/A5688/)**a** (**tradução**) ocorre em todas as células vivas. [Copyright](https://www.ncbi.nlm.nih.gov/books/about/copyright/) © 2002, Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter; Copyright © 1983, 1989, 1994, Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson

(Fonte: <https://www.ncbi.nlm.nih.gov/books/NBK21050/figure/A974/> )

**Reutilização deste material:** Salvo indicação em contrário, os conteúdos podem ser reutilizados não comercialmente sem o pedido de permissão, lembrando sempre de fazer a devida citação.

**CapÍtulo 3: PESQUISA DE HOMOLOGOS (“homology searches”)**

1. **O que genes e proteínas estão fazendo nos diferentes organismos**

Os cientistas acumularam dados sobre genes e proteínas por décadas, e agora tudo o que está disponível para você por meio de bancos de dados públicos. Durante as seguintes atividades você aprenderá sobre os bancos de dados e as ferramentas de bioinformática que podemos usar para consultá-los. Faremos isso cobrindo alguns tópicos:

- Para pesquisas de **homologia**, usaremos **similaridade** entre sequências para extrapolar a função. Também usaremos ferramentas de bioinformática, como o BLAST, para procurar sequências. Em bancos de dados biológicos, você se familiarizará com conjuntos de dados complexos e aprenderá a navegar com facilidade;

- Em sequência para funcionar, iremos **anotar** as proteínas de forma a atribuir uma função.

**2. Pesquisas de homologia**

**2.1. Semelhante não é o mesmo, mas ajuda!**

**pasted-image.tiff**

Neste capítulo, você será apresentado ao conceito de **anotação** e de **homologia**, uma maneira de identificar sequências semelhantes entre si.

Copos e canecas podem ser bem diferentes, mas todos servem a um propósito similar.

Imagine que você nunca tenha visto uma caneca de café antes. No entanto, você bebeu água de um copo, um copo de papel e até usou as mãos como recipiente. Quando você vê um "item desconhecido" (ou seja, caneca de café), apesar de sua forma desconhecida, com um identificador saindo de um lado, por causa da forma deste item você pode inferir que ele pode (ou pelo menos poderia) ser usado para o mesma finalidade de segurar ou beber um líquido. Muito intuitivo, certo? Atribuir a mesma função a coisas semelhantes é inato à natureza humana. O mesmo processo pode ser usado para sequências de proteínas.

Agora imagine que temos uma grande coleção de sequências de proteínas (por exemplo, 1.000), sendo que apenas 20 das quais sabemos a função. Podemos usar uma abordagem semelhante à usada com a caneca de café para inferir suas funções: sequências que parecem as mesmas têm uma boa chance de fazer a mesma coisa.

Isso é chamado de “**anotação por homologia”** este princípio que permite aos cientistas usar a **similaridade** para inferir a **função** baseando-se na conservação de uma determinada sequência com pequenas variações ao longo da evolução. Em termos gerais, quanto mais semelhantes forem duas sequências, maior a probabilidade de elas estarem relacionadas. Consequentemente, a **anotação por homologia** é baseada na **comparação de DNA** ou **proteínas** no nível de sequência - isto é, **comparando a similaridade de sequências** nucleotídicas ou de aminoácidos entre sequências relacionadas conhecidas.

Sequências de proteínas que conferem função são frequentemente encontradas em blocos de conservação chamados **domínios de proteínas**. Estas regiões têm uma estrutura ou **motivo tridimensional** definido (forma) que pode funcionar e evoluir independentemente do resto da sequência proteica. Esses blocos de conservação são encontrados em proteínas ao longo da natureza, e qualquer sequência de proteínas pode ter mais de um domínio de proteína.

A chave para usar a similaridade de motivo para inferir a função baseia-se no princípio de que, quando duas **proteínas têm uma função conservada**, embora sua **semelhança de sequência no nível de aminoácidos possa ser perdida,** sua **conservação deve permanecer.**

No entanto, como em todas as regras há **exceções** e **é possível** que **duas sequências ou motivos semelhantes entre si tenham diferentes funções** em **diferentes organismos** ou mesmo em **diferentes compartimentos** **da mesma célula**.

Portanto, é importante lembrar que a inferência de função é apenas uma projeção de sua função. Portanto, é comum ver nomes de proteínas ou descrições funcionais acompanhadas das palavras “**putativo**” ou “**potencial**”. Para ter certeza da função de uma proteína, ela deve ser confirmada experimentalmente.

Para uma visão mais aprofundada de como esses bancos de dados secundários podem ajudar na anotação de genomas completos é recomendado a leitura do artigo de revisão “Anotação de função de proteína por inferência baseada em sua homologia” de Loewenstein et al. 2009.

* Loewenstein et al. **Protein function annotation by homology-based inference**. Genome Biology, 2009, 10:207 (<http://genomebiology.com/2009/10/2/207>)

Obs:

Nas próximas seções, você aprenderá sobre ferramentas e bancos de dados que usam semelhanças entre sequências e motivos para ajudar a atribuir uma função à uma proteína.

**2.2. BLAST é uma ferramenta para anotação de homologia**

**A “Ferramenta básica de pesquisa de alinhamento local”** ou, **BLAST** (**Basic Local Alignment Search Tool**), **é uma ferramenta usada para encontrar sequências similares em bancos de dados biológicos. (Link BLAST:** [**https://blast.ncbi.nlm.nih.gov/Blast.cgi**](https://blast.ncbi.nlm.nih.gov/Blast.cgi)**)**

O **BLAST** é uma das ferramentas mais usadas para procurar sequências semelhantes. A maioria dos bancos de dados biológicos tem um servidor BLAST para pesquisa de seus conjuntos de dados. O BLAST é um programa de busca que é capaz de comparar rapidamente uma sequência de consulta com centenas a milhões de sequências.

O BLAST executa três etapas:

- 1ª: "corta" a sequência de consulta em pequenas "palavras" de tipicamente 3-4 aminoácidos para sequências de proteínas, ou 10-12 nucleotídeos para sequências de DNA (Fig. 1A);

- 2º: Usa essas palavras curtas para procurar correspondências perfeitas em todas as entradas no banco de dados (Fig. 1B);

- 3º: Quando uma correspondência é encontrada, o programa tenta estender o alinhamento, comparando as letras consecutivas. Para cada novo par de letras, avalia **se apresentam boa coincidência “a good match**” (Fig. 1C). **Se apresentam boa coincidência**, **a pontuação é aumentada**; e, se **a coincidência for ruim**, a **pontuação será reduzida**. Então, a tabela de pontuação para cada par de aminoácidos ou nucleotídeos é pré-computada e incorporada ao algoritmo BLAST.

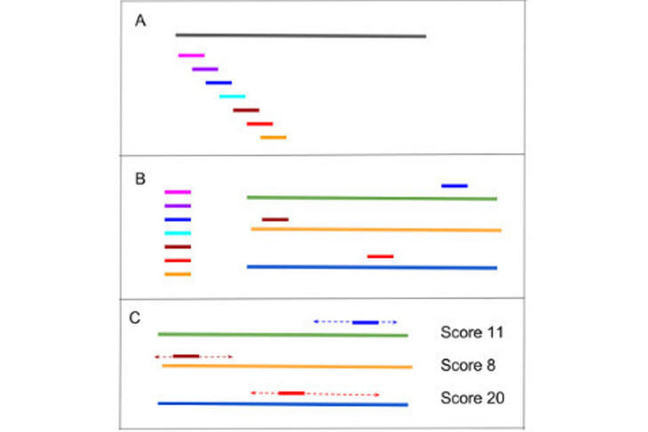


Figura: As três etapas de um alinhamento BLAST (simplificado). A) corte da sequência a ser consultada (“**query**”) em palavras curtas; B) encontrar correspondências (“**matches**”) exatas no banco de dados; C) estende as correspondências e atribui uma pontuação (“**a score**”).

**A etapa de extensão continuará até que a pontuação geral caia abaixo de um determinado valor. Nesse ponto, a etapa de extensão é interrompida** e o alinhamento é gravado **com esta pontuação. Os resultados são apresentados como uma lista de alinhamentos com pontuações associadas.** Os **alinhamentos com as pontuações mais altas são provavelmente correspondências verdadeiras ou homólogos da sequência de consulta.**

Outros parâmetros de resultado são relatados, como: **E-value** (valor da expectativa) e “**percentage identity”** (porcentagem de identidade):

- O **E-value** descreve o **número de ocorrências** (“**hits”)** que podem ser encontradas por acaso, considerando o **comprimento da sequência** e **o tamanho do banco de dados**. **Quanto menor o E-value, maiores as chances de que o resultado não seja devido ao acaso;**

- Porcentagem de identidade (“**percentage identity”)**.

**BLAST – Os diferentes tipos de programas BLAST**

Existem diferentes tipos de **programas BLAST** dependendo se a consulta é de uma sequência de nucleotídeos ou de proteína e, também, depende da natureza do banco de dados em que se está pesquisando (nucleotídeos ou proteína):

**** Quando a **consulta** (**“query”**) e o **banco de dados** são do mesmo **tipo:**

- **BLASTn**: Consulta por correspondências de uma sequência de nucleotídeos em um banco de dados de nucleotídeos;

- **BLASTp**: Consulta por correspondências de uma sequência de proteína em um banco de dados de proteínas.

**** Quando a **consulta** (**“query”**) e o **banco de dados** são de **tipos diferentes?**

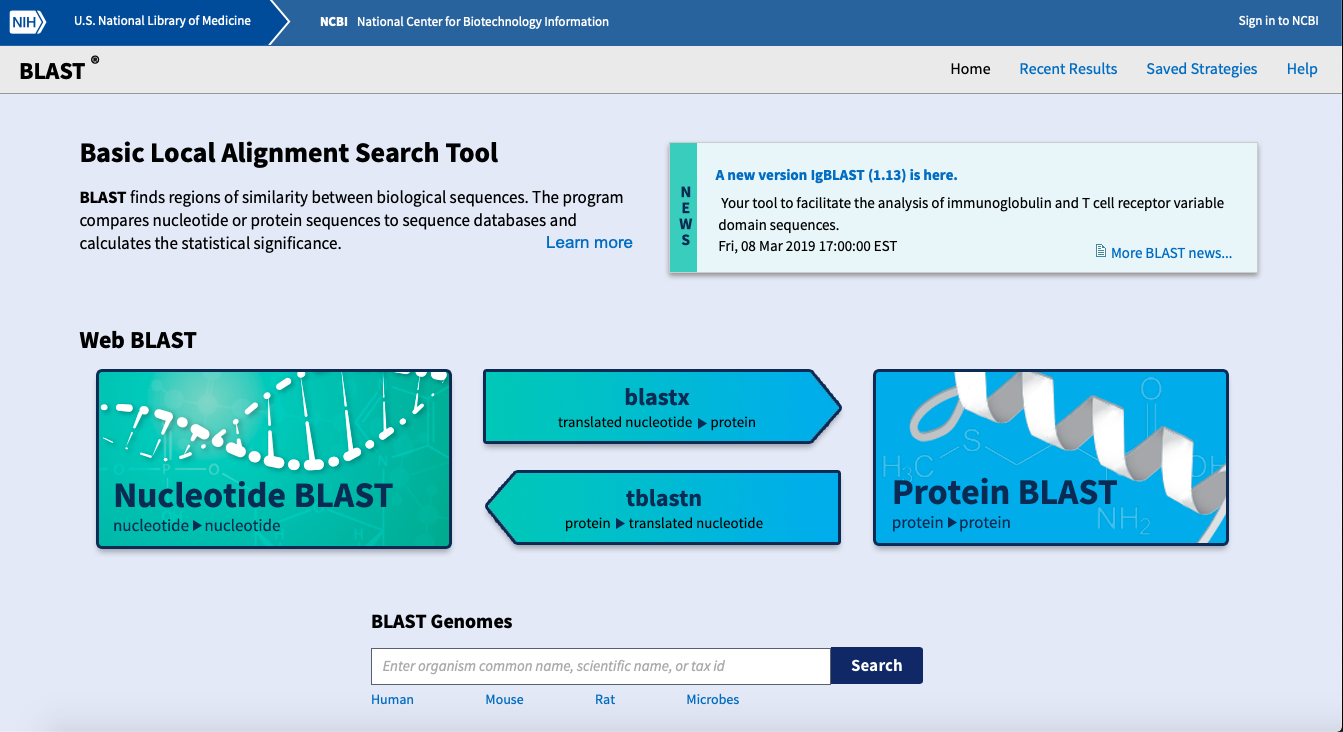
Para fazer um alinhamento, o “**query**” e o “**subject**” (banco) precisam ser da mesma natureza. Então, o que fazer quando o “**query”** (consulta) é uma sequência de **proteína** e a pesquisa visa encontrar as melhores correspondências em um **banco de dados de nucleotídeos** “(**subject**”), ou vice-versa?

De modo muito útil, **o BLAST pode traduzir todas as entradas no banco de dados de nucleotídeos em sequências de proteínas**, sendo que cada sequência pode ser traduzida nos 6 quadros possíveis, também, o BLAST **pode traduzir** o **“query”** nos 6 quadros possíveis. Nestes casos, devem ser utilizados os seguintes tipos de BLAST:

- **tBLASTn**: Para pesquisa de uma sequência de **proteína** (**“query”)** em um **banco de nucleotídeos**; pois, estará sendo usado o **"banco de dados traduzido"** como “**subject**”;

- **BLASTx**: Para pesquisa de uma sequência de **nucleotídeos (“query”)** em um **banco de dados de proteínas** (“**subject**”); pois, estará sendo usado o **"“query"** traduzido nos 6 quadros possíveis.

**2.3. Uso do BLAST (Basic Local Alignment Search Tool)**

**O** [BLAST](https://blast.ncbi.nlm.nih.gov/Blast.cgi) é uma ferramenta poderosa usada para pesquisar um banco de dados de sequências de DNA ou de proteínas, a fim de encontrar “hits” ("ocorrências") semelhantes a uma sequência “**query” (**consultada). O BLAST é usado para vários propósitos, incluindo a inferência de **uma possível função de uma proteína.** O website do [NCBI](https://www.ncbi.nlm.nih.gov/) inclui um servidor BLAST muito fácil de usar, através do qual se pode pesquisar os bancos de dados de sequências de **nucleotídeos** e de **proteínas** do NCBI, cujo link é: [https://blast.ncbi.nlm.nih.gov/Blast.cgi#](https://blast.ncbi.nlm.nih.gov/Blast.cgi%23) **.**

****

**VIDEO**

Neste tutorial em vídeo, é demonstrado o uso da ferramenta de busca de similaridade BLAST, hospedado pelo NCBI.

link:[https://blast.ncbi.nlm.nih.gov/Blast.cgi#](https://blast.ncbi.nlm.nih.gov/Blast.cgi%23))

**0:05.**

Neste vídeo, aprenderemos sobre as funções básicas do BLAST. Para isso, vá para o seu navegador favorito e digite **BLAST NCBI**.

**0:27**

Você acessará a página de destino, onde verá quatro programas de pesquisa diferentes. Neste momento, para essa instância, usaremos o **Nucleotídeo BLAST**. Na página de pesquisa, você encontrará três seções diferentes: A primeira é (“**Entry Query Sequence**”) a entrada da sequência a ser consulta; a segunda é (“**Choose Search Set**”) um conjunto de pesquisa de escolha; e a terceira é (“**Programme Selection**”) uma seleção de programa. Agora, usaremos o número de acesso, mas você também poderá usar a sequência **FASTA** ou um **arquivo salvo anteriormente**. Para isso, você usará (“**Choose File function**”) a função Escolher arquivo. Neste momento, vamos usar a sequência que usamos antes, cujo **número de acesso** é **X81322**. No “**Choose Search** **Set”** há três bancos de dados diferentes que você pode usar: o humano, o mouse e others.

**1:13**

Neste exemplo, vamos usar o outros, (“**others**”) que é o banco de dados não redundante, que inclui todas as sequências que foram depositadas no repositório do BLAST.

**1:25**

Nesta seção, para o organismo, você pode incluir ou excluir um determinado táxon. Por exemplo, se eu optar por incluir apenas bactérias, ele só pesquisará nos resultados das bactérias. Você pode optar por incluir, deixando esta caixa em branco ou marcá-la para excluir. Por enquanto, vamos apenas **optar por excluir**. Se você se mover para a terceira seção, que é (“**Programme Selection**”) a Seleção de Programa, você encontrará três modos diferentes de BLAST - as **sequências altamente similares**, as **mais diferentes**, e um **pouco similares**. O que é importante notar aqui é que **sequências altamente similares**, que são o **Megablast**, usam **95% ou mais sequências semelhantes**.

**2:09**

Para obter mais informações sobre os tipos de BLAST, basta clicar em (“**Information tab**”) “Guia de Informações” e você obterá todas as informações para todas as três versões diferentes do BLAST. Devemos então revisar as informações que colocamos. No topo está nossa sequência de entrada que, como você pode ver, é preenchida automaticamente com o gene *hpcC*. Em seguida, usamos o banco de dados não redundante e um ID de taxa de bactérias. Nesse exemplo, como estamos usando um gene de *E. coli*, não queremos incluir as pesquisas de *E. coli* e, na verdade, procurar outros organismos que possam ter o mesmo gene. Então, aqui, vamos digitar *E. coli*, e um menu “dropdown” irá aparecer com um ID de táxon, 562. E neste caso, vamos excluir os resultados.

**3:03**

Agora, vamos clicar BLAST. Você será levado para a página de pesquisa de função. Enquanto aguardamos, este é bom momento para ressaltar que o BLAST se tornou uma ferramenta realmente complexa, e você pode ficar à vontade para usar quaisquer opções diferentes para observar diferentes funções de busca. Chegando o resultado da pesquisa. O que você verá no topo é o que você pesquisou exatamente. Neste caso, o gene de *E. coli hpc*, a molécula, o comprimento do gene. À direita, você verá qual banco de dados que foi utilizado e várias outras informações. Em seguida é o resumo gráfico. No topo está a chave, à qual voltaremos mais tarde.

**3:43**

A próxima é a caixa turquesa, que representa o comprimento de seu “nucleotide query” (consulta de nucleotídeo), que é representada por 1.500 nucleotídeos, aproximadamente. E você pode ver a partir do comprimento, que é aproximadamente 1.500 nucleotídeos. Abaixo, você pode ver as barras vermelhas e estas indicam suas pontuações de alinhamento. Se você voltar para a chave, verá números associados às caixas. Aqui, em nossos resultados, a maioria das nossas sequências - se não todas – estão acima de 200. E, neste caso, **quanto maior o “score”** (pontuação), **melhor é o alinhamento**. Se, então, minimizarmos isso, a próxima etapa será uma guia Descrições. Aqui, você verá os diferentes organismos que foram semelhantes à sequência de pesquisa.

**4:26**

No lado direito, você verá vários números, que incluem seu “max score” (pontuação máxima), “total score” (pontuação total), “query coverage” (cobertura de consulta), E-value, “percentage identity” entre o “query” (consulta) e o “subject” (banco de busca). O que é importante notar aqui é o “E-value”. O “**E-value**”, ou valor esperado, **é a porcentagem de identidade** entre o “query” e o “subject”, quantificando a chance de que essa correspondência tenha sido encontrada por acaso, considerando o tamanho da sequência pesquisada e o tamanho do banco de dados consultado. **Quanto menor o “E-value”**, o resultado tem melhor qualidade. Minimizando isto, você pode ir para a “Alignment section” (seção Alinhamento). No topo está o que sua sequência “query” combinou e abaixo estão duas sequências – o seu “query” e o “subject” sendo a entrada de *Shigella*.

**5:09**

Aqui também é uma representação gráfica, onde se pode ver linhas de correspondências, que significa **sintenia** ou **correspondência**, e onde não há correspondência há uma lacuna. Se você quiser explorar mais do seu gene, pode sempre clicar no botão Editar e reenviar, e ele retornará para a página Pesquisa da “query sequence” (sequência a ser consultada). Aqui você pode alterar diferentes parâmetros. Por exemplo, você pode optar por incluir *E. coli* desta vez ou, escolher um banco de dados maior de bactérias e ver quais resultados surgem.

Nesta demonstração do BLAST, usamos um gene conhecido para analisar diferentes bancos de dados e ver os resultados correspondentes ou não correspondentes.

**2.4. Mergulhando no BLAST. Parte 1 - passo a passo.**

Neste tópico, você aprenderá como fazer uma pesquisa no BLAST. Você será guiado passo-a-passo pelo processo de envio de uma pesquisa BLAST usando o servidor **NCBI BLAST**.

Nesta primeira tarefa, você usará um número de acesso para recuperar uma sequência de nucleotídeos ou de uma proteína do **Genbank** e usará o **BLAST** para encontrar sequências similares dentro dos bancos de dados especificados.

**A.**

1. Vá para o site **Genbank/NCBI**, primeira página ( <https://www.ncbi.nlm.nih.gov/> )
2. No “**search box**” (caixa de pesquisa), escreva “**NP\_458691.1**. Escolha a opção “All Databases” no menu “drop down”. Isso permitirá que a pesquisa seja realizada em vários bancos de dados e não ficará restrita a bancos de dados de nucleotídeos ou de proteínas. Isto é importante porque, até o momento, não sabemos se o NP\_458691.1 é uma entrada de DNA ou de proteína.
3. Isto permitirá que a pesquisa seja realizada em muitos bancos de dados e não ficará restrita a bancos de dados de nucleotídeos ou de proteínas, não sabemos se o NP\_458691.1 é uma entrada de DNA ou proteína.
4. Clique em Pesquisar. Que tipo de resultados são mostrados? Dica: haverá um link direto para a entrada que corresponde ao número de acesso NP\_458691.1, bem como uma entrada na categoria “Proteína”.
5. Navegue até a entrada clicando em “**proteína secretada associada à invasão**” [ou use este link](https://www.ncbi.nlm.nih.gov/protein/16763074/). Esta entrada é de uma sequência de nucleotídeos ou de proteínas?
6. Agora você executará uma pesquisa do BLAST para procurar uma sequência semelhante. Primeiro, você terá que recuperar a sequência. Para fazer isso, vá para “Send to” ( “Enviar para”) no canto superior direito, clique no menu de busca, “select File” e, em formato, selecione “**FASTA**”.
7. Baixe a sequência e abra o arquivo em um editor de texto.
8. Navegue até o servidor do **BLAST** ( [https://blast.ncbi.nlm.nih.gov/Blast.cgi#](https://blast.ncbi.nlm.nih.gov/Blast.cgi%23) **)**
9. Qual tipo de programa BLAST você deve usar com base no tipo de “**query”** (nucleotídeo ou proteína) e no tipo de banco de dados, nucleotídeo ou / proteína? Selecione uma das opções.
10. Cole a sequência “**query”** no “search box” (caixa de pesquisa). Deixe todas as opções como mostrado como padrão, mas clique em “Show results in a new window” ("Mostrar resultados em uma nova janela") encontrado na parte inferior da página ao lado do botão BLAST. Agora clique em BLAST e aguarde os resultados.
11. Estude os resultados. Quais organismos são mostrados que têm sequências semelhantes ao seu “**query”**? Quão semelhantes são essas sequências com o seu “**query”**? Dica: compare “**E-values”** e “**Scores”.**

**B.**

1. Vamos repetir a pesquisa o BLAST e alterar alguns parâmetros. Você pode editar sua pesquisa clicando no link “**Edit and Resubmit**” (“Editar e reenviar”) no canto superior esquerdo ou retornando à guia original e editando os parâmetros.
2. Vamos alterar os parâmetros do banco de dados. Vamos limitar nossa pesquisa à espécie ***Yersinia***. Para fazer isso, digite “***Yersinia***” na caixa "**Organism**" dentro da seção "**Choose Search Set**". Execute a pesquisa do BLAST.
3. Estude os resultados. Há “good matches” (bons resultados) dentro da espécie *Yersinia*?
4. Vamos repetir a pesquisa, mas desta vez limitando os resultados à espécie ***Shigella***.
5. Estude os resultados. Há “good matches” (bons resultados) em ***Shigella***?
6. Repita a pesquisa do BLAST, agora **excluindo todas as enterobactérias**. Para fazer isso, digite "**enterobacteria**" dentro da caixa “Choose Search Set” na seção "**Choose Search Set**" e marque a opção "**Excluir**". Execute a pesquisa do BLAST. Quais são suas descobertas? Concentre-se nas entradas não ***Salmonella***.

**C.**

18. Com base nos resultados combinados, o que você pode concluir sobre a presença de NP\_458691.1 em outras bactérias?

**D.**

1. Repita a análise **usando bancos de dados de nucleotídeos** em vez de **bancos de dados de proteínas**.

sugestão: Vá para a página inicial do BLAST e procure por um programa que permita usar um “**query”** (consulta) de **proteína** contra um **banco de dados de nucleotídeos traduzidos**. Como os resultados se comparam à primeira parte desta análise? Limite / exclua espécies de seus resultados como antes.

**…………………………………………………………………………………………………………..**

**Questão 1:**

Verifique se as afirmações abaixo estão corretas:

(**BLASTp**: “**query”** e “**database”** são proteínas.

1. Melhor alinhamento para query **sem organismo**: score é 495, e-value é 4e-177, percent identity 100.00% para ***Salmonella enterica***.

2. Melhor alinhamento para query com **Yersinia**: "Proteína hipotética” (*Yersinia* sp. CFS 1934); score é 157, e-value 5e-50, identidade 38.66%.

3. Melhor alinhamento para query com ***Shigella***: sem resultado.

4. Melhor alinhamento para query com **exclusão** de **enterobactérias**: ***Chromobacterium violaceum***: score é 90,9, e-value 3e-18, identity 31,05%.

**Questão 2:**

**tBLASTn**

1. ***Salmonella*** ***enterica* subsp**.: Score 493, e-value 2e-176, identity is 99.58%.

2. ***Yersinia*** ***sp***: score is 157, e value 2e-42, identity 38.24%.

3. ***Shigella***: no result.

4. ***Chromobacterium violaceum***: score 88.6, e- value 3e-16, identity 30.53%.

**…………………………………………………………………………………………………………..**

**2..5. Mergulhando no BLAST. Parte 2 - Investigue e Discuta**

**Você recebe uma sequência “query” e é solicitado a executar uma pesquisa usando o BLAST.**

Clique e tente o máximo de recursos possível, mesmo que estejam incorretos

1. Use **KJE42419** para **encontrar sequências de proteínas similares** usando um tipo de programa **BLAST compatível**. Investigue os resultados com “**scores**” “**melhores**” e “**piores**”. Por exemplo, observe as espécies de bactérias e o nome dos gêneros.
2. Este código é o mesmo ou diferente do “query”?
3. Como você filtraria as correspondências de baixa pontuação (”low-scoring”) em pesquisas futuras?
4. Há mais de uma maneira de filtrar resultados de baixo “score”?
5. Discuta suas opiniões com outros colegas.
6. Em qual “**E-value**” ou “**score”** você faria a linha de corte, “threshold”, do limiar para acertos significativos e por quê?
7. Compartilhe este valor-limite "limiar para acertos significativos", “**E-value**” ou “**score”,** com outras pessoas na área de comentários e partilhe o motivo da sua escolha.
8. Como o **seu valor escolhido** de “**E-value**” se compara com os dos demais colegas?
9. Dê um “**Like**” nas escolhas e justificativas de pelo menos 1 dos colegas.
10. Diga se o “**E-value**” ou o “**score”** que você escolheu dar o **‘Like’** é o mesmo ou é diferente daquele de sua escolha original.
11. Diga se estes comentários reforçaram ou mudaram seus pensamentos originais e escolha do “**E-value**” ou do “**score”**?

**…………………………………………………………………………………………………………..**

****

**Questão** **3:**

Qual das seguintes afirmações NÃO é relevante para o processo de anotação por homologia? Escolha UMA opção:

a) Resultados do BLAST

b) Estrutura 3D da proteína

c) accession number da sequência

d) Conservação de sequencias de amino ácidos

**Questão 4**

Quais desses valores fornecidos como resultados de uma pesquisa do BLAST serão alterados se o tamanho do banco de dados for alterado?

a) Percentagem de identidade

b) E-value

c) Score

d) A porcentagem de cobertura dos resultados do BLAST depende do “query” e do “subject”

e não depende do tamanho ou da natureza do banco de dados

- Percentage identity

**- E-value**

- Score

- The percentage of coverage of a BLAST results depends on the and do not depend on the size or nature of the database

**Questão 5**

Select all the answers you think are correct.

Na página de submissão do BLAST, qual das seguintes opções pode ser usada para inserir a sequência de consulta? Selecione todas as respostas que você acha corretas.

a) colar uma sequência FASTA

b) inserir (“upload”) um arquivo de sequência FASTA

c) o nome da proteína

d) o número de acesso da proteína

**Respostas Corretas:**

1. c: Um número de acesso ou identificador não está relacionado à função da proteína.
2. b) O “E-value” depende, entre outros fatores, do tamanho do banco de dados.
3. a, b, d.

**…………………………………………………………………………………………………**

Creative Commons https://pixabay.com/en/dna-double-helix-science-rna-296744/