



Análise de variantes de *M. tuberculosis* para iniciantes

Tutorial adaptado de: [Link original](#)

Autores: Peter van Heusden, Simon Gladman, Thoba Lose

Data: 25/07/2020

Adaptação: Nuno S. Osório

Data da adaptação: 28/10/2024

Introdução

A tuberculose (TB) é uma doença infecciosa causada pela bactéria *Mycobacterium tuberculosis* (Mtb). A análise de variantes genómicas desta bactéria pode ajudar a identificar mutações que levam à resistência a medicamentos e auxiliar na vigilância de surtos. Neste tutorial, você aprenderá a identificar essas variantes usando ferramentas bioinformáticas no Galaxy.

Objetivos

- Obter e preparar os dados de sequenciação para análise.
- Realizar controle de qualidade dos dados.
- Detectar e anotar variantes genómicas.
- Interpretar os resultados para entender a resistência a medicamentos.

Perguntas

No final deverá ser capaz de responder a estas questões pela análise dos resultados obtidos:

1. Qual será a linhagem mais provável da amostra testada?
2. Foram encontradas potenciais resistências a antibióticos?

Requisitos

Este tutorial assume que você já possui uma conta no Galaxy e está minimamente familiarizado com o ambiente Galaxy. Caso contrário, veja o [Guia de Introdução ao Galaxy](#) antes de começar.

Passo 1: Obtenção dos Dados

Vamos começar obtendo os ficheiros de dados. Para isso, siga as etapas abaixo:

1. **Vá até à sua conta Galaxy.** Se não tiver uma conta, crie uma gratuitamente em usegalaxy.org ou usegalaxy.eu

2. Carregue os seguintes ficheiros de dados para o Galaxy, copiando os links para a ferramenta de importação de dados:

https://zenodo.org/record/3960260/files/004-2_1.fastq.gz

https://zenodo.org/record/3960260/files/004-2_2.fastq.gz

https://zenodo.org/record/3960260/files/Mycobacterium_tuberculosis_ancestral_reference.gbk

https://zenodo.org/record/3960260/files/Mycobacterium_tuberculosis_h37rv.ASM19595v2.45.chromosome.Chromosome.gff3

3. Na interface do Galaxy, vá para "**Upload Data**", selecione "**Paste/Fetch Data**" e cole os links e carregue "**Start**". Certifique-se de que os ficheiros são importados corretamente.

Passo 2: Controle de Qualidade dos Dados (FastQC)

Antes de proceder à análise, é essencial verificar a qualidade dos dados de sequenciação. Isso ajuda a identificar problemas como baixa qualidade ou adaptadores remanescentes que podem afetar os resultados. Vamos usar o **FastQC** para isso.

Etapas detalhadas:

1. No Galaxy, pesquise por **FastQC** na barra de ferramentas.
2. Selecione a ferramenta **FastQC Read Quality Reports** nos resultados.
3. No campo "**Short read data from your current history**", selecione os ficheiros FASTQ que carregou (as suas leituras sequenciadas forward e reverse).
4. Deixe as opções como padrão para o resto dos campos e clique em **Executar**.

Após a execução, o FastQC gerará relatórios visuais para cada ficheiro. Verifique os relatórios para ver se as leituras têm qualidade suficiente para análise. Se houver problemas, como baixa qualidade ou presença de adaptadores, pode ser necessário aplicar ferramentas de limpeza de dados antes de prosseguir.

Passo 3: Alinhamento das Leituras

Depois de verificar a qualidade dos dados, o próximo passo é alinhar as leituras de sequenciação com o genoma de referência. Para isso, usamos o **Snippy**, que detecta variantes ao comparar as suas leituras com o genoma de referência.

Etapas detalhadas:

1. Pesquise por **Snippy** no Galaxy.
2. Abra a ferramenta **Snippy**.
3. Nas opções:
 - Selecione as leituras forward e reverse que você verificou no FastQC.
 - Carregue o genoma de referência (ficheiro .gbk) que importou.
4. Deixe os outros parâmetros como padrão e clique em **Executar**.

O Snippy analisará as suas leituras e retornará uma lista de variantes detectadas (SNPs), que poderão ser exploradas nos próximos passos.

Passo 4: Filtragem e Anotação de Variantes

Agora que temos uma lista de variantes, o próximo passo é filtrá-las e anotá-las. Filtrar significa remover variantes que possam ser ruído, como aquelas com baixa cobertura. A anotação atribui significado às variantes, associando-as a genes e fenótipos específicos (por exemplo, resistência a medicamentos).

Etapas detalhadas:

1. Pesquise pela ferramenta **SnpEff** no Galaxy.
2. Abra a ferramenta e carregue as variantes geradas pelo Snippy.
3. Carregue o ficheiro de anotação de referência que importou anteriormente (.gff3).
4. Clique em **Executar (“Run Tool”)** para anotar as variantes.

O resultado incluirá uma lista de variantes anotadas com informações sobre genes, impacto, e possível resistência a medicamentos.

Passo 5: Filtragem Adicional de Variantes e Perfil de Resistência a Medicamentos

Para obter uma lista confiável de variantes, é necessário realizar uma filtragem adicional, pois o genoma de *M. tuberculosis* possui regiões difíceis de mapear, como os genes PE/PPE/PGRS e as sequências de inserção (IS). Inserções e deleções (indels) podem também causar variantes de nucleotídeo único (SNVs) falsas perto das regiões indel. Finalmente, regiões com baixa cobertura devido a deleções ou alta concentração de GC podem ter chamadas de variantes de baixa confiança. A ferramenta TB Variant Filter ajuda a filtrar variantes com base nesses critérios.

Etapas detalhadas:

1. No Galaxy, pesquise pela ferramenta **TB Variant Filter**.
2. Abra a ferramenta **TB Variant Filter (versão Galaxy 0.4.0+galaxy0)**.
3. Nas opções:
 - Em “VCF file to be filter”, selecione os ficheiros VCF gerados pelo Snippy (dados XX).
 - Em “Filters to apply”, selecione “Filter variants by region” e “Filter sites by read alignment depth”.
4. Clique em **Executar (“Run Tool”)** para aplicar os filtros.

Após a execução do filtro, um novo ficheiro VCF será gerado, contendo apenas as variantes de alta confiança. Esse ficheiro poderá ser usado para uma análise mais confiável da resistência a medicamentos.

Passo 6: Identificar mutações de resistência a antibióticos

Podemos usar a ferramenta **TB Profiler** para gerar um relatório que inclui mutações associadas a resistência a antibióticos e SNPs filogeográficos.

Etapas detalhadas:

1. No Galaxy, pesquise pela ferramenta **TB-Profiler Profile**
2. No **Input File Type**:
 - Adicione as reads fastq 1 e a 2.
3. Clique em **Executar (“Run Tool”)** para gerar o relatório.

O relatório gerado inclui informações sobre a linhagem final da amostra e a presença de resistências a medicamentos.

Conclusão

Este tutorial guiou-o por uma análise de variantes de *M. tuberculosis*, desde a obtenção de dados, controle de qualidade, até a anotação e interpretação de variantes. Para mais detalhes, explore outras ferramentas no Galaxy ou veja o tutorial completo. Este tutorial foi traduzido e simplificado.