**Assignment 02**

**High Throughput Sequencing at your fingertips**

**Licenciatura em BioInformática**

**Docente:** Francisco Pina Martins

**UC:** Análise de Sequências Biológicas

**Realizado por:**

André Rosado Nº2940

Diogo Gonçalves Nº3081

Gonçalo Pedro Nº3091

Nuno Ferreira Nº2911

Tiago Matias Nº3089

# **Índice**

[**Índice** 2](#_Toc75551340)

[**Introdução** 3](#_Toc75551341)

[**Materiais e Métodos** 4](#_Toc75551342)

[**Resultados/Discussão** 7](#_Toc75551343)

[**Bibliografia** 9](#_Toc75551344)

# **Introdução**

No âmbito da unidade curricular de Análise de Sequências Biológicas, foi-nos proposto pelo docente a realização de uma tarefa/trabalho com o objetivo de aperfeiçoarmos as nossas habilidades e práticas tendo como base e princípios alguns métodos aprendidos nas aulas durante o semestre onde inclusive utilizamos alguns dos programas empregues no Assignment 1.

O principal objetivo deste trabalho é esclarecer algumas dúvidas associadas a um tipo de planta amazónica. Alguns exploradores encontraram este tipo de planta e repararam que de entre todas as suas características, a que mais se sobressaía era o facto de que alguns indivíduos possuíam flores particularmente malcheirosas (Pungents) enquanto a maioria deste tipo de planta não possuía nenhum odor percetível (Bland). Alguns destes exploradores afirmavam que este acontecimento se devia ao facto de que alguns indivíduos Pungent pertenciam a uma espécie diferente, enquanto outros diziam que se devia à variabilidade intraespecífica. Como tal, tendo em conta que o método de Sanger foi inconclusivo, os indivíduos Pungent e Bland foram sequenciados utilizando métodos RAD-Seq.

Neste trabalho, de modo a resolver o problema descrito, o grupo criou um plano de trabalho para uma abordagem mais elaborada e organizada. Decidimos então utilizar o método de High Throughput Sequencing (HTS).

Este método de High Throughput Sequencing é utilizado na maioria das vezes para estudos de expressões génicas. Esta técnica de HTS, como era antigamente chamada, era utilizada para sequenciar cópias de DNA expressos em pequenas amostras. Nos dias de hoje, o mesmo é feito, embora em muito maior escala, e agora chamado de RNA-seq que mostra os genes que são ativos numa determinada amostra biológica.

Posto isto, na realização deste trabalho, o conjunto de dados que utilizamos incluíam as espécies Pungent e Bland.

# **Materiais e Métodos**

Neste ponto do relatório vamos fazer uma descrição dos processos e do procedimento que realizamos. Inicialmente, para dar início à metodologia do trabalho, fizemos o download de um “pre-built High Throughput Sequencing (HTS) dataset” que nos foi fornecido pelo docente. Nestes ficheiros podemos encontrar todo o datatype necessário com que iríamos trabalhar.

Posteriormente descompactamos ambos os ficheiros e todos os documentos necessários para preparação do datatype.



Fig. 1 - Ficheiros descompactados e presentes na VM

De seguida, pelo facto de possuirmos os parâmetros padrões, tivemos de realizar algumas alterações e ajustes. Houve uma necessidade de mexer nos parâmetros de modo a permitir que conseguíssemos fazer correr o ipyrad. (Figura 2)

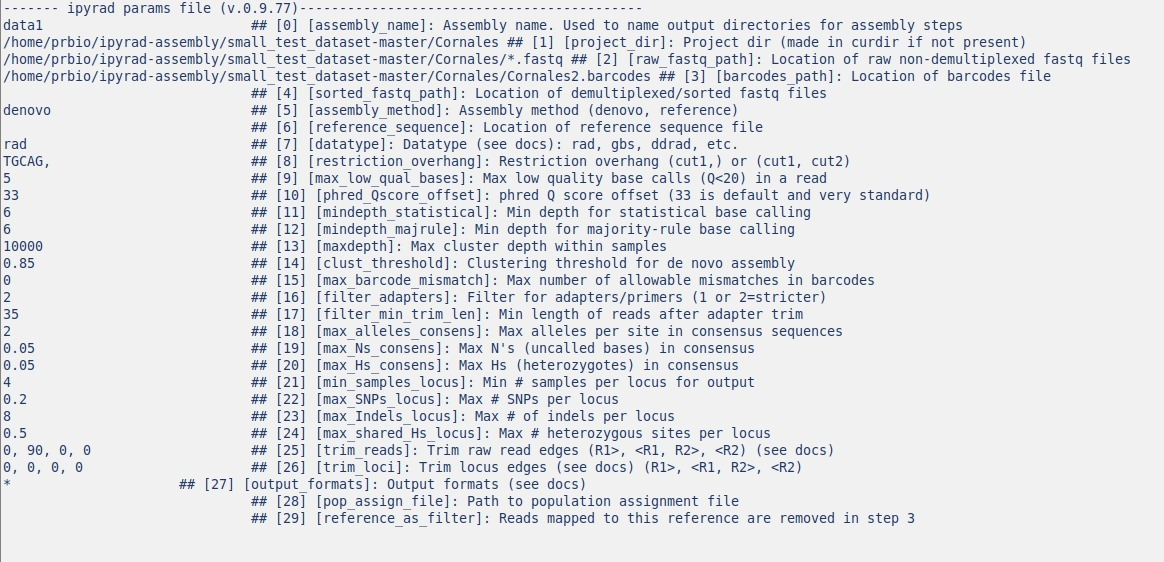


Fig. 2 – Ajustes nos parâmetros

Nesta situação, por possuirmos dois ficheiros fastq tivemos de usar o “\*.fastq” para utilizar os dois ficheiros como demonstra a figura 3.

Fig. 3 – Passo para utilizar os dois ficheiros (\*.fastq)

Posteriormente, procedemos à execução do ipyrad desde o passo dois até ao passo 7. Neste último fomos impedidos de continuar devido a um erro que nos informava que não havia taxa suficiente para os barcodes mencionados. Assim, de modo a resolver este problema, utilizamos um script de Python para ultrapassar a barreira.

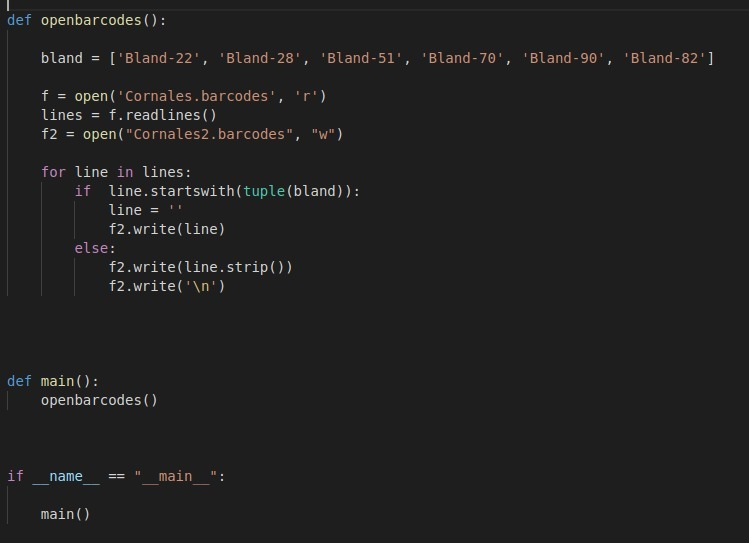


Fig. 4 – Script de Python para resolver o problema de taxa

Com este script de Python, pudemos criar um novo ficheiro de barcodes que apenas contém os barcodes necessários para a realização da tarefa.

De seguida, utilizamos a ferramenta Raxml para criar uma árvore e mais tarde utilizaremos o Figtree para visualizar o resultado obtido.

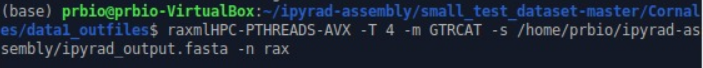


Fig. 5 – Processo de utilização do Raxml

Neste ponto, como apenas tínhamos utilizado um método, decidimos dar uso à ferramenta Mrbayes para obter uma outra árvore. Aqui utilizamos o ficheiro Nexus que obtivemos do ipyrad e de seguida, para simplificar o processo, usamos o bloco do Mrbayes para podermos corrê-lo.

Abrimos o terminal na pasta onde tínhamos o ficheiro Nexus que usamos na execução do processo do Mrbayes, colocando lá o resultado obtido.



Fig. 6 – Processo de execução do Mrbayes

Para finalizar, depois de todo este processo descrito, obtivemos com output um ficheiro com uma árvore, onde utilizamos a ferramenta Figtree para observar a mesma.

# **Resultados/Discussão**

Como não foi possível visualizarmos a nossa árvore no RAxML, decidimos utilizar a ferramenta Figtree para nos auxiliar na nossa visualização.

Nesta árvore é difícil distinguir, com facilidade, quais os ramos pertencentes á espécie Pungent dos ramos pertencentes á espécie Bland.



Fig. 7 - Árvore obtida através do figtree

Devido a esta dificuldade de distinção decidimos então criar uma árvore do tipo Bayesian mas desta vez aplicamos o método MrBayes.

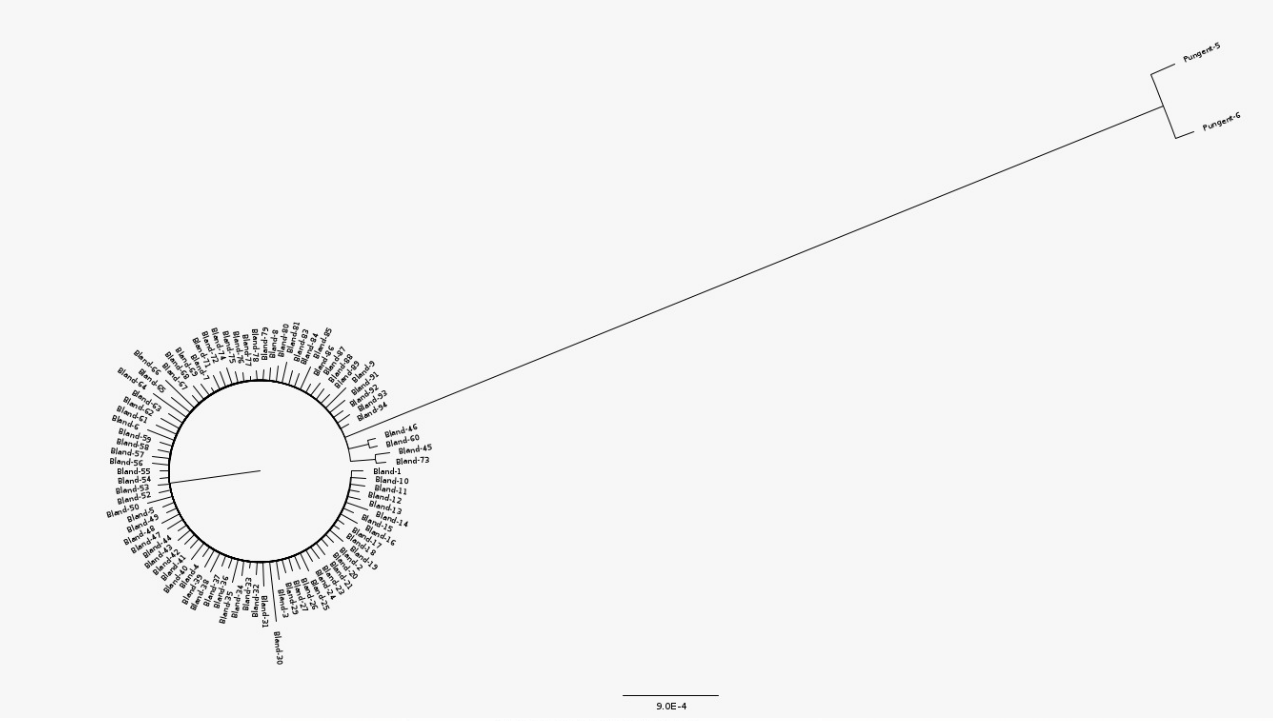


Fig. 8 - Árvore Bayesian com o método MrBayes

Nesta árvore, é possível verificar uma melhor visualização e distinção de quais os ramos pertencentes á espécie Pungent e quais os pertencentes á espécie Bland.

Como conclusão em resposta ao problema biológico, podemos verificar que após as dificuldades em distinguir quais os ramos que pertenciam à Bland e Pungent, conseguimos concluir que a espécie Bland obteve uma variedade genética muito mais elevada que a espécie Punget, pois como se pode observar na figura 2, a espécie Bland tem uma ramificação muito mais elevada em comparação com a espécie Pungent que tem uma ramificação muito reduzida.

# **Bibliografia**

* https://stuntspt.gitlab.io/asb2021/assignments/Assignment02.html
* <https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>
* <https://github.com/NBISweden/MrBayes/blob/develop/doc/manual/Manual_MrBayes_v3.2.pdf>
* <https://www.biostars.org/p/287346/>
* <http://www.jcsantosresearch.org/Class_2014_Spring_Comparative/pdf/week_6/Feb_12_2015_Chronograms.pdf>