Uma imagem com texto

Descrição gerada automaticamente

**Licenciatura em Bioinformática**

(2ºAno 2ºSemestre)

**Análise de Sequências Biológicas**

**Molecular Phylogenetic Analysis of Tulipa Liliacea**

Inês Baptista, 202100917

João Ribeiro, 202101169

Margarida Santos, 202100915

Pedro Pacheco, 202100957

**2022/2023**

**Índice**

[**Introdução** 3](#_Toc134305025)

[1. Problema biológico e Objetivos 3](#_Toc134305026)

[2. Métodos Utilizados 3](#_Toc134305027)

[3. Conclusões 4](#_Toc134305028)

[**Materiais e Métodos** 5](#_Toc134305029)

[**Resultados** 6](#_Toc134305030)

[**Discussão** 10](#_Toc134305031)

[**Referências Bibliográficas** 11](#_Toc134305032)

# **Introdução**

## Problema biológico e Objetivos

Este estudo tem como objetivo analisar e investigar as árvores filogenéticas de espécies pouco comuns de *Tulipas* da ReservaNatural de Aksu-Zhabagly juntamente com a sua identificação morfológica e código de barras de ADN. As secções do espaçador interno transcrito (ITS) ribossómico nuclear são um dos marcadores de ADN mais utilizados na codificação de ADN e na investigação filogenética (Today, s.d.). A Reserva Natural é a reserva natural mais antiga da Ásia Central e alberga 1312 espécies de plantas vasculares, 44 das quais estão classificadas como ameaçadas ou em perigo no livro de dados vermelhos do Cazaquistão.

O principal objetivo deste estudo é obter a identificação correta das espécies e a avaliação das suas relações filogenéticas. Ao longo dos anos, o número de espécies do género *Tulipa* oscilou entre 81 e 100 espéciessendo a sua classificação ter sido alterada várias vezes concentrando-se mais em características morfológicas ou citogenéticas. Inicialmente, para começar a resolver os problemas de identificação, os cientistas utilizaram uma comparação de caracteres moleculares como alternativa à identificação morfológica. (Today, s.d.).

Neste paper é descrito o estudo de 17 espécies raras de túlipas (*T. greigii*, *T. gréigii* var. Red-Yellow, *T. kaufmanniana, Hybrid of T. kaufmanniana/T. greigii, T. kaufmanniana, T. turkestanica, T. bifloriformis)* encontradas no AZNR (Aksu-Zhabagly Nature Reserve) onde são determinadas a diversidade genética e são estaabelacidas relações entre estas espécies que de outra forma não seriam possíveis. A identificação do barcode foi feita pela sequenciação da região ITS (internal transcribed spacer) do DNA nuclear e dos gener matK e ycf1b do DNA plasmídeo que foram isolados das tulipas.

O código de barras de ADN podem ser utilizados repetidamente para a identificação baseada no ADN surgiu como uma técnica útil para aqueles que não possuem características físicas distintas. O método PCR foi utilizado e feito um perfil de repetição de sequências simples (ISSR) para estudar a *Tulipa* do Irão, enquanto a técnica de polimorfismo de comprimento de fragmentos amplificados (AFLP) foi utilizada para examinar a Tulipa orphanidea L. da Turquia. (Today, s.d.).

## Métodos Utilizados

**Um dos métodos utilizados foi o isolamento de DNA-PCR e sequenciamento de Sanger.**

Este consiste na extração de uma quantidade de DNA de folhas congeladas de tulipas utilizando o método de CTAB. As amostras de tecido foram recolhidas num tubo de microcentrifugadora sendo em seguida colocadas nos adaptadores e pulverizadas. Para cada tudo, foi adicionado um tampão de extração pré-aquecido, que se completou até 1mL, e de seguida agitou-se no vortex para ficar bem misturado. Depois, as amostras foram incubadas durante 1 hora a 65ºC tendo-se adicionado um volume igual de álcool isoamílico/clorofórmico, tendo sido novamente agitado no vórtex antes de se centrifugar os tubos à velocidade máxima durante 5 minutos. Transferiu-se a camada aquosa superior para um novo tudo de microcentrifugação de 2 mL contento 600 μl de 2-propanol, agitou-se bem e centrifugou-se os tubos à velocidade máxima durante 2 minutos. O sobrenadante foi removido e o sedimento levado com EtOH, por inversões suaves. O passo seguinte foi centrifugar a alta velocidade 2 minutos o etanol. O ADN foi imediatamente dissolvido com a utilização da RNAse A a 55ºC durante 10-30 minutos. A qualidade do ADN foi seguidamente determinada por eletroforese em agarose e a sua concentração determinada utilizando um espectrofotómetro. O par de primers foi utilizado para a amplificação da região do ITS nuclear feita por PCR efetuada num termociclador nas seguintes condições: passo inicial de desnaturação a 95ºC durante 5 minutos, seguida de 30 amplificações a 95ºC durante 45 segundos, a 55ºC durante 30 segundos e a 72ºC durante 30 segundos, seguidas de uma extensão final de 72ºC durante 3 minutos. Os produtos de PCR foram analisados por eletroforese em gel com tampão TAE 1x e limpos. Os produtos PCR purificados foram sequenciados em ambas as direções utilizando os primers utilizados para a amplificação.

Os cromatogramas após a sequenciação foram analisados e editados com o software SnapGene. As sequências editadas foram analisadas por buscas BLAST para análise preliminar. O alinhamento de sequências múltiplas foi efetuado pelo programa Mega 11. Sequências para região ITS, genes matK e ysf1 para T. lehmanniana, T. clusiana, T. chrysantha, T. turkestanica, T. tarda, T. sprengeri, T. cretica, T. uniflora, T. iliensis, T. silvestris, T. humilis, Lilium sulphureum, L. pumilum foram extraídas do GenBank.

## Conclusões

O estudo feito através de DNA barcoding permitiu algumas conclusões sobre estas espécies. O gene MatK foi encontrado nas diferentes espécies analisadas como permanecendo igual em todas, existindo apenas uma substituição genética relacionada com outra espécie, *T. turkestanica*. Por sua vez, o seu gene ycf1b evoluiu mais lentamente nas espécies de tulipas do que nos outros membros da família Liliceae, facto que se prevê dever-se à baixa taxa de recombinações genéticas de tulipas graças à sua incompatibilidade com outras plantas. Por último, a região ITA do DNA nuclear permitiu conclusões mais definitivas, a par com o DNA plastídeo, permitindo estabelecer conclusões fortes entre as relações evolucionárias das diferentes espécies de *Tulipa*.

# **Materiais e Métodos**

Foram extraídos os genes do NCBI recorrendo a um programa disponibilizado no [GitHub](https://github.com/joao-ribeirooo/Python-program), depois realizou-se vários alinhamentos para o gene ITS, ycf1 e matK recorrendo a função automática do MAFFT (v7.490) para realizar o melhor alinhamento possível.

Usando cada sequência alinhada, inferimos a relação filogenética utilizando Maximum Likelihood e Bayesian Inference. Para as árvores de Maximum Likelihood foi usado o RaxML-NG(v1.1) com o modelo TIM3+G4 para o gene ITS, HKY para o gene ycf1 e TVM para o gene matK. Estes modelos foram obtidos pela utilização do ModelTest-NG(v0.1.7) em que foram determinados utilizando o critério Akaike(AICc). Os valores de bootstrap foram também computados usando o RaxML-NG(v1.1) com 10.000, 1.000 e 30.000 replicações para os genes ITS, ycf1 e matK, respetivamente. Para as árvores de Bayesian Inference foi usado o MrBayes(v3.2.7a) e feito 1.000.000 de gerações de MCMC(Markov Chain Monte Carlo) para ITS e ycf1 e 10.000.000 para o gene matK. No final para observar as árvores foi utilizado o figtree(v1.4.4). O código para recriar as árvores está disponibilizado no [GitHub](https://github.com/joao-ribeirooo/Codigo-das-arvores).

# **Resultados**

Na ***figura 1*** está representada a árvore da região ITS pelo método de Maximum Likelihood. O gene ITS é uma região do DNS presente no núcleo das células que está presente, neste caso, nas *Tulipas* que queremos estudar. Podemos dizer que todas as espécies descenderam de um ancestral comum, porém a espécie *Lilium Sulphureum* encontra-se mais afastada das restantes espécies de Tulipas, sendo considerada então o *outgroup.*

Pela análise da árvore conseguimos perceber que desde a espécie “*Tulipa Tarda*” até à “*Tulipa Clusiana*”, estas representam uma maior similaridade entre elas uma vez que partilham o mesmo nó na árvore sendo do mesmo ancestral comum que originou todas as outras espécies existentes com valor de bootstrap 69. Já as espécies desde a “*Tulipa Hybrid Cultivar AZHNR*” até “*Tulipa Greigii Red Yellow*” também descendem de um mesmo nó, indicado com o valor de bootstrap 71, sendo então originadas as espécies indicadas. Com isto, conseguimos concluir que existem 4 grupos de espécies havendo uma espécie outgroup do ancestral comum inicial (“*Tulipa greigii”*) e outro outgroup dentro do resto das espécies que se forem evidenciando. Comparando com o paper original (árvore A), as árvores adequam-se e apresentam semelhanças, apesar da ordem e maneira de apresentação dos ramos e nós ser um pouco diferente do que foi realizado por nós.

Uma imagem com texto, diagrama, Paralelo, Esquema

Descrição gerada automaticamente

**Figura 1** – Tree ITS MLA A (A corresponde à letra da árvore correspondente do paper original)

Na ***figura 2*** está representada a árvore do gene ycf1b pelo método de Maximum Likelihood. Este método é uma técnica utilizada para construir árvores filogenéticas que representam as relações evolutivas entre diferentes espécies, neste caso, de plantas utilizando um modelo probabilístico para inferir a filogenia de uma amostra de sequências genéticas. (Science, s.d.) A árvore filogenética obtida por meio da análise de Maximim Likelihood é a hipótese de filogenia que maximiza a probabilidade de observar as sequências de DNA observadas.

É possível observar que apenas a espécie “*Tulipa humilis var pulchella*” difere mais das restantes espécies, sendo esta então o *outgroup*. Temos presente um nó, com o valor de bootstrap 98, onde estão apenas duas espécies “*Tulipa turkestanica*” e “*Tulipa bifloriformis*”, que indica que estas são mais parecidas entre si e que se divergiram assim mais recentemente. Num outro nó, que vai desde a espécie “*Tulipa greigii*” até “*Lilium pumilum*”, é possível afirmar que estas espécies se encontram todas mais relacionadas entre elas do que com as restantes espécies de *Tulipas*. A partir desta análise conseguimos perceber quais são as espécies mais relacionadas uma com as outras e como é que elas evoluíram ao longo do tempo, os possíveis eventos de especiação, onde novas espécies podem originar-se a partir de ancestrais comuns e qual a importância do gene ycf1b teve na evolução das plantas.

A interpretação que fazemos comparando com o paper original (árvore E) é que existem diferenças visíveis entre cada um. No paper original, a árvore divide-se em apenas dois grandes grupos sem haver grandes outgroups, enquando que na análise das nossas árvores encontraamos dois outgroups que se evidenciam logo visualizando depois o resto das divergências.

Uma imagem com texto, diagrama, captura de ecrã, file

Descrição gerada automaticamente

**Figura 2** - Tree ycf1b ML E(E corresponde à letra da árvore correspondente do paper original)

A ***figura 3*** representa a árvore da região ITS pelo método de Inferência Bayesiana. A Inferência Bayesiana é um método que vai estimar a probabilidade de uma hipótese de filogenia ser verdadeira com base nos dados observados, sendo especialmente útil quando a amostra de sequência é grande e/ou quando as sequências são muito divergentes (como vamos poder observar na *figura 3*). É também possível

São identificados dois outgroup pelas espécies “*Tulipa gr*” e “*Tulipa hy*”. Podemos destacar um grande grupo de espécies entre “*Tulipa cl*” e a “*Lilium su*” que nos diz que estas são mais parecidas geneticamente entre elas do que com as restantes espécies. As mutações ocorridas fizeram com que se divergissem da maneira que a árvore é apresentada.

Comparando com o paper original (árvore B) as diferenças são muito significativas em relação a cada árvore mostrada. No paper original mostra apenas um outgroup pertencendo a “*Lilium sulphureum*” enquanto que na nossa árvore temos dois outgroups mais visíveis a partir do primeiro ancestral comum. Na nossa árvore existe uma maior divergência das espécies não havendo muitas relações entre elas, ou seja, há muitos ramos em que se mostra a pouca intimidade de relação entre as espécies, enquanto que no paper originial ocorre grandes parecenças entre as espécies.

Uma imagem com texto, diagrama, Paralelo, file

Descrição gerada automaticamente

**Figura 3**-Tree ITS Bayesian Inference B(B corresponde à letra da árvore correspondente do pper original)

Na investigação do gene ycf1b a partir do método de construção Bayesian inference e a utilização de sequências de diferentes espécies de tulipas (retiradas do GeneBank) e *Lilium pumilum* escolhido como grupo exterior, mas pertencendo à mesma à família *Liliaceae* chegamos à árvore que podemos ver na ***figura 4.***

Podemos perceber que uma grande quantidade de amostras de tulipas ficaram agrupadas num único ramo, mas podemos reparar que nesse mesmo ramo existe uma *Tulipa* pertencente ao grupo exterior que partilha o mesmo ancestral que as diferentes espécies de tulipas, mas como essa existem mais mas há uma exceção, o isolamento de uma tulipa pertencente ao grupo exterior que está isolada. Com isto podemos concluir que não existiu uma grande divergência na evolução do mesmo gene nestas espécies de ambos os grupos de estudo. Claro que estas conclusões são só aplicáveis na utilização deste método de Bayesian Inference e com o respetivo gene e estes grupos em estudo, pois as conclusões em relação a estes dois grupos de estudo podem eventualmente mudar com a mudança de uma destas importantes variáveis.

As evidentes diferenças entre as árvores do paper original (F) e as nossas árvores são bastantes visíveis. Os níveis de confiança das nossas árvores são normalmente cerca de 1, o que significa ter um bom nível de confiança para as suas análises e parencenças. Enquanto que no paper original (árvore F) os níveis de bootstrap já diferenciam um pouco e não existem ramos tão lineares como na nossa árvore. Estes resultados indicam que o gene ycf1b evoluiu minimamente.

Uma imagem com texto, diagrama, Paralelo, file

Descrição gerada automaticamente

**Figura 4** - Bayesian Inference F do gene ycf1b (F significa a letra correspondente ao paper original)

A árvore filogenética resultante da análise do gene MatK utilizando o método Maximum Likelihood é composta por dois grupos visíveis. Um outgroup pertencente à espécie *Lilium Pumilum* e um outro que corresponde à divergência e caminhos que as restantes espécies tomaram através do gene MatK. A partir do mesmo ancestral comum do grupo de espécies acima da árvore surgiram dois grupos, a espécie “*Tulipa Uniflora*” e as restantes espécies “*Tulipa Clusiana*” até “*Tulipa Bifluriformis*”.

Cada ramo representa as diferentes espécies de plantas e cada nó representa os pontos de divergência evolutiva. Os ramos mais curtos representam as espécies que estão mais relacionada entre si, enquanto que os ramos mais longos representam espécies mais distantes. A posição a que cada espécie está na árvore filogenética, indica a ordem cronológica da divergência evolutiva, ou seja, as espécies mais próximas são aquelas que estão mais relacionadas e as espécies mais distantes são as que divergiram há mais tempo.

Comparando com o paper original (árvore C), a nossa árvore mostra menos as mesmas evoluções e divergências o que nos leva a concluir que a sequências MatK é altamente conservada em todas as diferentes espécies de *Tulipa*.

Uma imagem com diagrama, texto, Retângulo, file

Descrição gerada automaticamente

**Figura 5** - Tree MatK Maximum LikeLihood C (C significa a letra da árvore do paper original correspondente)

Como já dito anteriormente, a inferência Bayesiana utiliza um método que é usado para representar asw relações evolutivas entre as diferentes espécies. A partir desta árvore conseguimos visualizar a linearidade dos ramos das espécies que ocorre na ***Figura 6,*** o que significa que não houve divergências ou relações evolutivas do gene MatK no método Bayesian Inference. As distâncias evolutivas foram iguais entre as espécies, a similaridade entre elas foram iguais indicamente menos distância evolutiva fornecendo uma tentativa de estimativa da incerteza nas relações filogenéticas inferidas. Apesar disto, o valor 1 que significa o bootstrap significa que há uma maior probabilidade da hipótese de filogenia ser verdadeira.

Comparando com o paper original (árvore D), as diferentes são visíveis não havendo de todo esta linearidade dos ramos e valores de bootstrap tão confiáveis. Estes resultados a partir do paper original indicam que o gene MatK é altamente conservada em todas as diferentes espécies.

Uma imagem com texto, captura de ecrã, file, Paralelo

Descrição gerada automaticamente

**Figura 6** - Tree MatK Bayesian Inference D (D significa a letra correspondente à árvore do paper original)

# **Discussão**

O estudo foi capaz de fazer contribuições significativas para a compreensão da filogenia das espécies de *Tulipa*. (Today, s.d.). A identificação correta das espécies de *Tulipas* é importante para o estudo da diversidade biológica da região, identificação e evolução das *Tulipas* bem como para a conservação e gestão adequada dos recursos naturais incluindo a história da diversificação das espécies e a origem geográfica das diferentes linhagens.

As análises filogenéticas moleculares usaram o DNA das plantas para inferir as relações evolutivas entre as diferentes espécies de *Tulipas* permitindo assim comparar as sequências de DNA das espécies coletadas com as sequências conhecidas de outras espécies de *Tulipas* e determinar com maior precisão a identidade das espécies coletadas. A análise filogenética mostra que o clado se divide em 2 subclados.

O gene MatK é conservado em todas as diferentes espécies de *Tulipa* existindo uma substituição genética conhecida que está relacionada com outra espécie na espécie *T. turkestanica.* (Today, s.d.). Esta substituição genética pode ser usada como um marcador para distinguir *T. turkestanica* de outras espécies de *Tulipas.* (GeneCards, s.d.)O gene ycf1b evoluir mais lentamente nas espécies de tulipas do que noutros membros da família Liliaceae devido à falta de recombinação de genes em resultado da sua auto-incompatibilidade com outras espécies de plantas. Essa auto-incompatibilidade vai evitar a fertilização cruzada entre indivíduos geneticamente diferentes da mesma espécie ou de espécies diferentes impedindo, neste caso, a troca de material genético entre as espécies incluindo o gene trabalhado e estudado, ycf1b.(Reports, s.d.) Isto vai fazer com que haja uma redução na diversidade genética e à fixação de alelos numa população que vai levar a uma evolução lenta do gene, ocorrendo também mutações que prejudicam a função do gene. O artigo original utilizou técnicas de análise filogenética molecular para investigar as relações evolutivas entre as espécies de *Tulipas* do AZNR. As análises foram realizadas a partir de DNA extraído das amostras de plantas coletadas para depois conseguir analisar as árvores pelos diferentes métodos como Máxima Verossimilhança e Inferência Bayesiana tal como na análise que realizámos a partir das árvores apresentadas.

Para finalizar:

* As análises filogenéticas moleculares permitiram compreender as relações evolutivas entre as *Tulipas* da ReservaNatural de Aksu-Zhabagly e de outras espécies de outras regiões geográficas;
* Foi constatado que as espécies da família *Liliacea* apresentam uma diversidade genética moderada (apesar de sofrerem algumas mutações e divergirem de ancestrais comuns diferentes) e uma estruturação populacional significativa;
* A partir das extrações de DNA e depois de realizar as árvores filogenéticas conseguimos verificar os outgroup, bootstrap e os ancestrais comuns de cada espécies verificando também as divergências ou possíveis mutações nas espécies.

# **Referências Bibliográficas**

Aksu-Zhabagly. (s.d.). *Molecular Phylogenetic Analysis*. Obtido de https://www.researchgate.net/publication/369371898\_Molecular\_phylogenetic\_analysis\_of\_Tulipa\_Liliaceae\_from\_Aksu-Zhabagly\_Nature\_Reserve.

GeneCards. (s.d.). *GeneCards - MatK*. Obtido de https://www.genecards.org/cgi-bin/carddisp.pl?gene=MATK.

Reports, S. (s.d.). *ycf1, the most promising plastid DNA barcode of land plants*. Obtido de https://www.nature.com/articles/srep08348.

Science, T. D. (s.d.). *Maximum Likelihood*. Obtido de https://towardsdatascience.com/probability-concepts-explained-maximum-likelihood-estimation-c7b4342fdbb1.

Today, P. S. (s.d.). *Molecular Phlylogenetic Analysis*. Obtido de https://www.researchgate.net/publication/369371898\_Molecular\_phylogenetic\_analysis\_of\_Tulipa\_Liliaceae\_from\_Aksu-Zhabagly\_Nature\_Reserve.