



UNIVERSIDADE FEDERAL DA BAHIA

INSTITUTO DE BIOLOGIA

CURSO DE CIÊNCIAS BIOLÓGICAS

ANÁLISE DE SEQUÊNCIAS MULTILOCUS (MLSA) COMO FERRAMENTA AUXILIAR
PARA A CLASSIFICAÇÃO DE ESPÉCIES PATOGÊNICAS EMERGENTES DO GÊNERO

Corynebacterium

por

DANIELE ALMEIDA ALVES

Trabalho de conclusão de curso apresentado ao
Instituto de Biologia da Universidade Federal
Bahia como requisito para obtenção do grau de
Bacharel em Ciências Biológicas

Salvador, BA

2018

DANIELE ALMEIDA ALVES

ANÁLISE DE SEQUÊNCIAS MULTILOCUS (MLSA) COMO FERRAMENTA AUXILIAR
PARA A CLASSIFICAÇÃO DE ESPÉCIES PATOGÊNICAS EMERGENTES DO GÊNERO
Corynebacterium

Trabalho de conclusão de curso apresentado ao
Instituto de Biologia da Universidade Federal
Bahia como requisito para obtenção do grau de
Bacharel em Ciências Biológicas

Orientador: Dr. Luis Gustavo Carvalho Pacheco
(UFBA)

Co-orientadora: M.a Carolina Silva Santos
(UFBA)

Salvador, BA

2018

Data da defesa: 01 de Agosto de 2018

Banca examinadora

Prof. Dr. Luis G. C. Pacheco

Instituto de Ciências da Saúde / Universidade Federal da Bahia

Prof. Dr. Pedro Milet Meirelles

Universidade Federal da Bahia

Dra. Brena Mota Moitinho Sant'Anna

Universidade Federal da Bahia

AGRADECIMENTOS

À Universidade Federal da Bahia pelas oportunidades conferidas a mim durante os anos de graduação;

Ao meu orientador, Professor Doutor Luis Pacheco, pela recepção, ensinamentos e confiança nessa etapa tão importante na minha carreira profissional. Obrigada por acreditar em mim!

À minha co-orientadora, M.^a Carolina Silva Santos, pelas contribuições e disponibilidade em me ajudar sempre que necessário;

Ao Grupo de Estudos em Genômica Funcional e Biologia Sintética (GenoFun), grupo de pesquisa coordenado pelo Professor Doutor Luis Pacheco no qual estou inserida, e aos meus colegas do Grupo;

Ao Laboratório de Bioinformática e Ecologia Microbiana pela prestatividade;

Ao professor Doutor Pedro Milet Meirelles, pela concessão do espaço físico para realização do trabalho, contribuições para o entendimento das ferramentas de bioinformática utilizadas e por aceitar o convite para compor a banca examinadora;

À professora Dra. Brena Mota Moitinho Sant'Anna por também ter aceitado o convite para compor a banca de avaliação final;

Ao professor Doutor Domingos Benicio Oliveira Silva Cardoso por toda contribuição;

Ao professor Doutor Gilberto Cafezeiro Bomfim e a Doutora Ticiano Carvalho Pereira por todo apoio prestado durante a graduação;

Aos que acreditaram sempre em mim, incentivando e apoiando cada momento dessa trajetória, especialmente a Érico Vinício;

À Bila Brandão, por me ajudar a compreender meu verdadeiro potencial;

À Dejanete Rodrigues, José Carlos Sampaio e Adriel Inácio por me ajudarem a seguir em busca da realização dos meus sonhos.

RESUMO

Espécies não diftéricas do gênero *Corynebacterium* têm emergido como patógenos oportunistas, estando amplamente distribuídos no meio ambiente. A identificação destes patógenos é um desafio e está se tornando cada vez mais difícil distingui-las com base em seus perfis bioquímicos, pois são necessários diversos testes bioquímicos que não estão disponíveis no Api Coryne, um sistema padronizado para a identificação de bactérias corineformes. Os métodos moleculares apresentam valor complementar, porém não são práticos para uso rotineiro nos laboratórios, e há relatos que espécies intimamente relacionadas não podem ser adequadamente diferenciadas pelas técnicas moleculares. Buscando uma melhor resolução entre as espécies dentro de um gênero, a Análise de Sequências Multilocus (MLSA) é utilizada para este fim. Nesse contexto, utilizamos a MLSA como ferramenta auxiliar para a classificação de espécies patogênicas emergentes do gênero *Corynebacterium* usando um conjunto de genes codificadores de proteínas presentes nos isolados medicamente relevantes. Os alinhamentos foram editados manualmente, permitindo uma melhor análise das sequências. Ao identificar sequências com baixa similaridade, utilizamos uma ferramenta on-line que analisa os arquivos de sequenciamento de forma rápida e com baixas taxas de erro. Isso permitiu a confirmação de uma identificação errada para o isolado *C. xerosis* BMi9815. Foi realizada uma pré-análise para uma avaliação das árvores baseadas em genes individuais, gerando 11 árvores que foram analisadas individualmente buscando inconsistências entre os ramos, observando os agrupamentos entre as espécies que já são conhecidas por serem bem definidas e os agrupamentos entre espécies diferentes. Em um alinhamento final foram concatenados 8 genes (total de 10.608 pb) obtidos de cada uma das 47 sequências genômicas completas, para posterior realização da análise filogenética por máxima verossimilhança utilizando o Smart Model Selection, e uma análise de split decomposition e construção de SplitsTree.. As nossas análises também permitiram observar 2 clados bem definidos que incluem linhagens de *C. minutissimum* e *C. aurimucosum* apoiando a hipótese de que essas espécies não são bem definidas taxonomicamente. Os resultados permitem sugerir a necessidade de uma melhor definição dessas espécies pelos métodos taxonômicos clássicos associados a esses novos métodos moleculares, bem como uma reclassificação de ambas espécies e, potencialmente, descrever novas espécies.

ABSTRACT

Non-diphtheria species of the genus *Corynebacterium* have emerged as opportunistic pathogens, being widely distributed in the environment. The identification of these pathogens is a challenge and it is becoming increasingly difficult to distinguish them based on their biochemical profiles because a number of biochemical tests are required which are not available in Api Coryne, a standardized system for the identification of coryneform bacteria. Molecular methods have complementary value, but are not practical for routine use in laboratories, and there are reports that closely related species can't be adequately differentiated by molecular techniques. Searching for a better resolution among species within a genus, Multilocus Sequence Analysis (MLSA) is used for this purpose. In this context, we used the MLSA as an auxiliary tool for the classification of emerging pathogenic species of the genus *Corynebacterium* using a set of protein encoding genes present in the medically relevant isolates. The alignments were edited manually, allowing a better analysis of the sequences. By identifying sequences with low similarity, we use an online tool that analyzes the sequencing files quickly and with low error rates. This allowed the confirmation of a misidentification for isolate *C. xerosis* BMi9815. A pre-analysis was performed for an evaluation of trees based on individual genes, generating 11 trees that were individually analyzed for inconsistencies between the branches, observing the groupings of the species that are already known to be well defined and the groupings between different species. Eight genes (total of 10,608 bp) obtained from each of the 47 complete genomic sequences were concatenated for later performance of the maximum likelihood phylogenetic analysis using the Smart Model Selection, and a split decomposition analysis and SplitsTree construction. Our analyzes also allowed us to observe 2 well defined clades that include *C. minutissimum* and *C. aurimucosum* lines supporting the hypothesis that these species are not well defined taxonomically. The results allow to suggest the need for a better definition of these species by the classical taxonomic methods associated to these new molecular methods, as well as a reclassification of both species and potentially to describe new species

LISTA DE IMAGENS

- Figura 1 Fluxograma geral da metodologia utilizada neste trabalho.17
- Figura 2 Árvore filogenética com base nas sequências do gene codificador da proteína ribossomal S17 pelo método Neighbor-Joining, com 100 réplicas de bootstrap.26
- Figura 3 Árvore filogenética com base nas sequências do gene codificador da subunidade beta da enzima RNA polimerase pelo método Neighbor-Joining, com 100 réplicas de bootstrap27
- Figura 4 Árvore filogenética com base nas sequências concatenadas dos genes *DNAK*; *GyrB*; *L4*; *L16*; *RECA*; *RPOB*; *S3*; *S19*.28
- Figura 5 SplitTree baseada em oito sequências gênicas concatenadas. Rede filogenética foi gerada usando o programa SplitsTree4.29

LISTA DE TABELAS

Tabela 1 Espécies e linhagens selecionadas para o estudo.	18
Tabela 2 Genes usados nesse trabalho, produto genético e tamanho de cada gene.	20

LISTA DE ABREVIATURA E SÍMBOLOS

16S rRNA	Gene codificador da subunidade 16S do RNA ribossômico
ATCC	<i>American Type Culture Collection</i>
MALDI	Matrix Assisted Laser Desorption Ionization
DNA	Ácido Desoxirribonucléico
PCR	Reação em cadeia da polimerase
mPCR	Multiplex PCR
RAST	Rapid Annotations using Subsystems Technology
rRNA	RNA ribossômico
RpoB	Gene codificador da subunidade beta da enzima RNA Polimerase
SMS	Smart Model Selection
TOF	Time of flight

ÍNDICE

1. Introdução.....	11
2. Justificativa.....	15
3. Objetivos.....	16
3.1 Geral.....	16
3.2 Específicos.....	16
4. Materiais e métodos.....	17
4.1 Seleção das espécies e obtenção das sequências genômicas completas:.....	18
4.2 Seleção dos genes:.....	20
4.3 Alinhamento das sequências:.....	21
5. Resultados.....	23
6. Discussão.....	30
7. Conclusões.....	33
8. Referências Bibliográficas.....	34

1. Introdução

Os bacilos Gram-positivos que não formam esporos constituem um grupo diverso de bactérias aeróbias e anaeróbias. Os organismos pertencentes ao gênero *Corynebacterium* são comuns na microbiota normal da pele e nas mucosas dos seres humanos, e assim, são frequentemente contaminantes de amostras clínicas submetidas a avaliação de diagnóstico (Brooks et al.,2014). O patógeno mais significativo desse grupo continua sendo o *Corynebacterium diphtheriae*, principal causa da difteria, uma doença que embora tenha desaparecido dos países desenvolvidos após a implementação da vacinação, é necessária uma vigilância contínua uma vez que novos casos continuam sendo relatados em várias regiões (Forbes et al, 2017).

As espécies não diftéricas deste gênero, também chamados de differóides têm emergido como patógenos oportunistas, estando amplamente distribuídos no meio ambiente. Além disso, são conhecidos por seu comportamento comensal na pele e mucosas, se tornando um foco de atenção, pois há relatos crescentes de seu isolamento em pacientes com várias infecções (Rizvi et al, 2013).

A identificação de *Corynebacterium* spp. não diftéricas é um desafio e está se tornando cada vez mais difícil distingui-las com base em seus perfis bioquímicos, pois estes requerem diversos testes bioquímicos que não estão disponíveis no sistema Api Coryne, um sistema padronizado para a identificação de bactérias corineformes que utiliza 21 testes bioquímicos miniaturizados para detecção de atividades enzimáticas e fermentação de carboidratos, além de uma base de dados on-line específica, gerando resultados após 24 horas (API Coryne™, Guia do Usuário. BioMérieux) (Bernard, 2012).

A identificação inequívoca das espécies patogênicas pode ser dificultada devido a grande variedade de perfis bioquímicos comumente relatados para essas espécies, incluindo *Corynebacterium striatum*, *Corynebacterium amycolatum*, *Corynebacterium minutissimum* e *Corynebacterium xerosis*. O nosso grupo de pesquisa realizou recentemente uma revisão abrangente da literatura com uma abordagem de bioinformática com o propósito de identificar a base genômica que contribui para as variabilidades bioquímicas observadas nos métodos de identificação fenotípica dessas bactérias. Essa revisão foi baseada na reconstrução de seis reações / vias bioquímicas específicas em 33 sequências genômicas inteiras recentemente liberadas, utilizando

dados recuperados de bancos de dados curados associados a pesquisas de homologia para confirmar resultados específicos (Santos et al., 2018).

Os bancos de dados associados são atualizados com pouca frequência e não fornecem discriminação suficiente entre as espécies recém-descritas. As análises manuais consomem tempo e exigem grande quantidade de material biológico, o que pode ser uma desvantagem. Métodos moleculares, como sequenciamento dos genes *rpoB* e *16S rDNA*, demonstraram ter valor complementar, mas não são práticos para uso rotineiro devido ao seu alto custo. E ainda há relatos que essas espécies intimamente relacionadas não podem ser adequadamente diferenciadas por técnicas moleculares (Bernard, 2012).

Sabendo que algumas espécies de *Corynebacterium* spp. são difíceis de identificar e requerem testes moleculares adicionais, foi desenvolvido também pelo nosso grupo de pesquisa um método de PCR multiplex (mPCR) baseado em genes específicos para cada espécie para uma diferenciação eficiente de isolados clínicos de *C. striatum*, *C. amycolatum* e *C. xerosis*. Fazendo uma comparação com trabalhos realizados anteriormente usando PCR para a identificação dessas corinebactérias, foi mostrada a necessidade de análises complementares, bem como foi ilustrada a utilidade do recém-desenvolvido mPCR com *primers* específicos como uma abordagem complementar rápida para a resolução de identificações ambíguas ou incorretas de espécies de *C. striatum* / *amycolatum* / *xerosis*. Além disso foi sugerida a disponibilidade de um ensaio de mPCR que também poderia incluir a espécie *C. minutissimum* (Santos, C. et. al, 2017).

Outro teste utilizado atualmente na identificação microbiana clínica é a espectrometria de massas do tipo MALDI-TOF. Mesmo essa abordagem sendo considerada rápida, barata e precisa, as avaliações recentemente publicadas mostraram identificações ambíguas e limitações na variedade de espécies nos bancos de dados, que também não são atualizados com frequência adequada (Alibi et al, 2015; Bernard, 2012).

Os métodos genotípicos estão desempenhando um papel fundamental na identificação e classificação filogenética em vários níveis taxonômicos. O sequenciamento do genoma total é considerado o método mais confiável para identificação da espécie estudada. A identificação de genes espécie específicos é uma das vantagens apresentadas por esse método. Porém seu alto custo de aquisição dos equipamentos necessários e necessidade de qualificação dos responsáveis pela execução e

interpretação dos testes restringe tal teste a laboratórios de referência (Glaeser & Kämpfer, 2015; Köser et al., 2012).

O gene codificador do RNA ribossômico 16S (16S rRNA) foi, e ainda é, aplicado como gene marcador universal para análise evolutiva de bactérias cultiváveis e não cultiváveis. No entanto, o 16S é bem conhecido por sua limitada resolução filogenética que dificulta a utilidade em nível de espécie ou subespécie (Fox et al., 1992). Sequências genômicas são recomendadas nas análises que tem como objetivo propósitos taxonômicos, em vez da utilização de hibridização convencional DNA-DNA e da filogenia baseada no 16S rRNA. A filogenia baseada no genoma, também chamada filogenômica, é inferida usando um conjunto de genes em vez de um único gene. Esses genes são definidos como cópias simples, homólogas e universalmente presentes no grupo alvo, que podem ser qualquer classificação taxonômica (Na et al., 2018)

O método atualmente aplicado na taxonomia procariótica, considerando fragmentos internos de vários genes codificadores de proteínas, é chamado de análise de sequência multilocus (MLSA, do inglês - multilocus sequence analysis) (Gevers et al., 2005). Na maioria das vezes, uma nova linhagem é primeiro alocada filogeneticamente com base na análise do gene 16S rRNA a nível de gênero. Buscando uma melhor resolução entre as espécies dentro de um gênero o MLSA é cada vez mais aplicado para este fim. Fragmentos internos de genes codificadores de proteínas são sequenciados e posteriormente usados para calcular árvores filogenéticas. Como cada filogenia da sequência gênica única reflete apenas a evolução desse único gene, o que novamente pode não refletir a relação filogenética "verdadeira", fragmentos de DNA de várias sequências gênicas são concatenados e árvores filogenéticas são calculadas com base nisso. Assume-se que as árvores baseadas nas sequências alinhadas concatenadas refletem melhor a relação "verdadeira" dos táxons bacterianos (Glaeser & Kämpfer, 2015).

Vários estudos relatam que foram feitas análises da filogenia do gênero *Corynebacterium*, mas apenas através de um número limitado de linhagens. Entre esses, foi feito um estudo de 56 espécies (Khamis et al., 2004) que analisou a filogenia examinada em dois genes separados (16S e *rpoB*) para tentar identificar o melhor método para definir espécies do gênero. Um estudo posterior dos mesmos autores mostrou que uma sequência parcial de *rpoB* forneceu o melhor método para identificação de corinebactérias. Eles sugerem que uma forte análise filogenética em todo o gênero

usando uma análise mais ampla, incluindo muitos outros genes, resolveria algumas das diferenças observadas entre os estudos anteriores e forneceria mais informações sobre a evolução do gênero (Oliveira et al., 2017).

Nesse contexto, o presente trabalho se propõe estudar a MLSA como ferramenta auxiliar para a classificação de espécies patogênicas emergentes do gênero *Corynebacterium* usando um conjunto de genes codificadores de proteínas presentes nos isolados medicamente relevantes, levando em consideração a disponibilidade das sequências nos bancos de dados.

2. Justificativa

Estudos demonstram as dificuldades em identificar de forma definitiva as bactérias patogênicas emergentes do gênero *Corynebacterium*. Entre essas, está a dificuldade em decidir corretamente em tempo hábil se a recuperação dessas bactérias implica contaminação ou tem relevância clínica, já que fazem parte da microbiota normal da pele. Portanto, recomenda-se que os médicos tentem identificar qualquer um dos táxons descritos como medicamente relevantes ao nível de gênero e espécie, se o organismo estiver isolado de locais do corpo normalmente estéreis, por exemplo, hemocultura. O número de espécies do gênero *Corynebacterium* identificadas como causadoras de infecções oportunistas e hospitalares em humanos vem crescendo bastante atualmente e a importância das corinebactérias para os microbiologistas clínicos e profissionais da saúde aumentou muito nos últimos anos.

Sabendo que a identificação baseada apenas em testes fenotípicos tem como consequência a frequente identificação ambígua desses patógenos, e alguns testes moleculares são inviáveis para uso rotineiro em laboratórios, tanto pelo seu alto custo quanto os relatos que espécies intimamente relacionadas não podem ser adequadamente diferenciadas, análises envolvendo atuais ferramentas de bioinformática permitem a identificação de forma mais prática.

Diante disso, espera-se que esse trabalho contribua para um melhor entendimento da Análise de Sequências Multilocus (MLSA) como ferramenta auxiliar para a classificação dessas espécies.

3. Objetivos

3.1 Geral

Avaliar a Análise de Sequências Multilocus (MLSA) como ferramenta auxiliar para a classificação de espécies patogênicas emergentes do gênero *Corynebacterium* a fim de estabelecer limites de separação entre as espécies taxonomicamente confusas.

3.2 Específicos

- Selecionar as espécies e isolados descritos como patogênicos emergentes pertencentes ao gênero *Corynebacterium*;
- Identificar um conjunto de genes *housekeeping* compartilhados pelos vários isolados;
- Realizar os alinhamentos das sequências individuais para cada gene para inferir quais deles refletem uma melhor resolução filogenética;
- Escolher o conjunto de genes para análises filogenéticas por máxima verossimilhança e de matriz de dissimilaridades por *split decomposition*;
- Verificar a viabilidade da MLSA como ferramenta complementar aos testes já amplamente utilizados para identificação de espécies patogênicas do gênero estudado.

4. Materiais e métodos

A metodologia utilizada nesse trabalho está esquematizada no fluxograma abaixo e está descrita após a figura abaixo

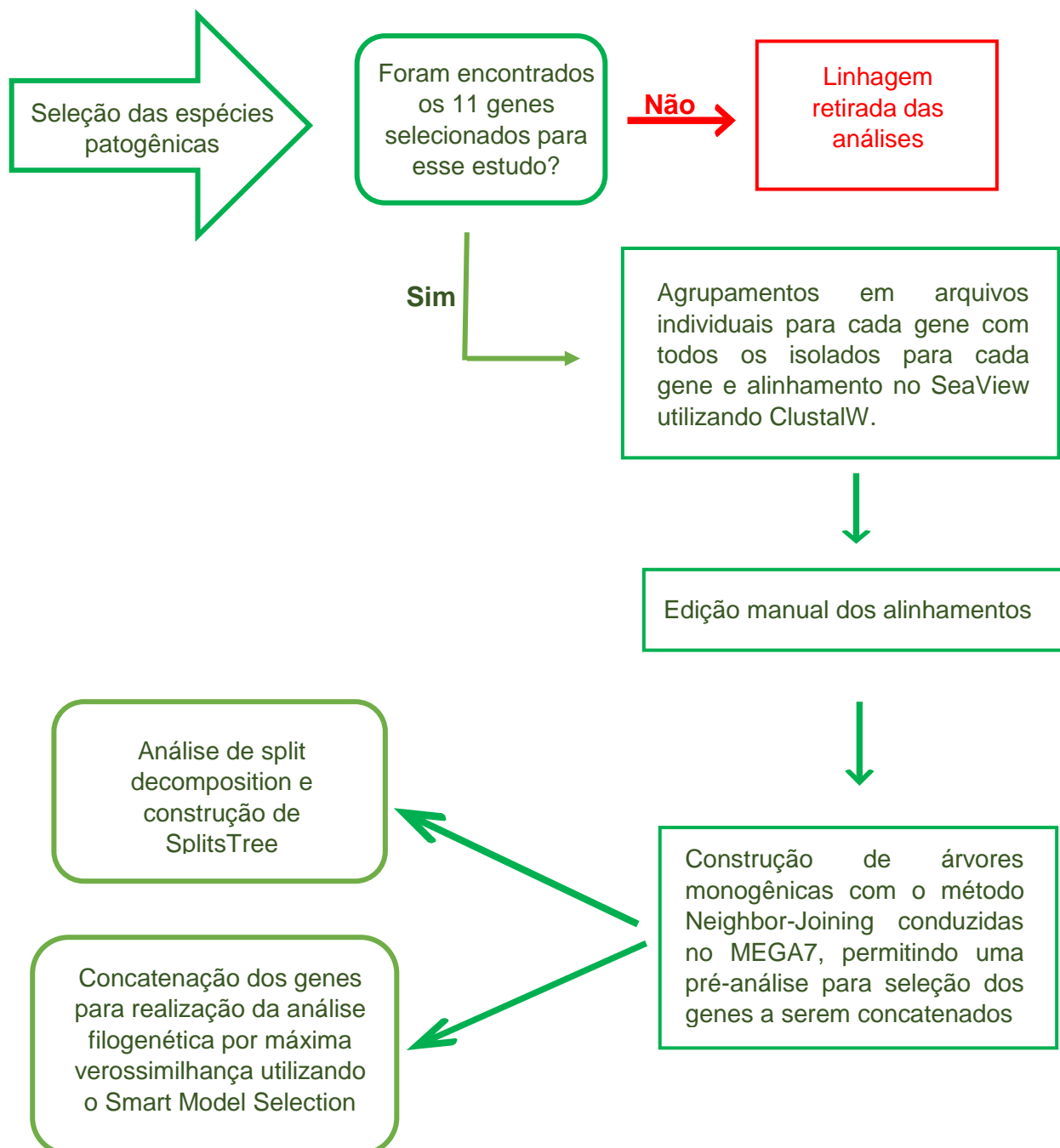


Figura 1 Fluxograma geral da metodologia utilizada neste trabalho. O fluxograma descreve o passo-a-passo das etapas realizadas nas etapas de seleção dos isolados e dos genes, no alinhamento das sequências e nas construções de árvores e rede filogenética.

4.1 Seleção das espécies e obtenção das sequências genômicas completas:

Os isolados bacterianos utilizados nesse trabalho foram previamente obtidos por Martins et al. (2009), tendo sido já caracterizados por testes fenotípicos padrão. Foram selecionadas 13 espécies correspondendo a 49 linhagens de corinebactérias patogênicas emergentes, conforme descrição na Tabela 1. Essa seleção foi baseada na relevância médica e disponibilidade de sequência genômica completa ou rascunho depositadas em banco de dados.

Tabela 1 Espécies e linhagens selecionadas para o estudo.

Espécie	Linhagens
<i>Corynebacterium accolens</i>	<i>C. accolens</i> AH4003
	<i>C. accolens</i> ATCC49725
	<i>C. accolens</i> ATCC49726
<i>Corynebacterium afermentans</i>	<i>C. afermentans</i> DSM44280
	<i>C. afermentans</i> LCDC880199
<i>Corynebacterium amycolatum</i>	<i>C. amycolatum</i> ICIS9
	<i>C. amycolatum</i> ICIS53
	<i>C. amycolatum</i> SK46
	<i>C. amycolatum</i> UMB0042
<i>Corynebacterium aurimucosum</i>	<i>C. aurimucosum</i> 118CAUR
	<i>C. aurimucosum</i> 620CAUR
	<i>C. aurimucosum</i> 911CAUR
	<i>C. aurimucosum</i> 945CAUR
	<i>C. aurimucosum</i> 1237CAUR
	<i>C. aurimucosum</i> ATCC700975
<i>Corynebacterium coyleae</i>	<i>C. coyleae</i> DSM44184
<i>Corynebacterium freneyi</i>	<i>C. freneyi</i> DNF00450

<i>Corynebacterium minutissimum</i>	<i>C. minutissimum</i> 1941 <i>C. minutissimum</i> ATCC23348 <i>C. minutissimum</i> NBRC15361
<i>Corynebacterium propinquum</i>	<i>C. propinquum</i> 143_CAUR <i>C. propinquum</i> 149_CJEI <i>C. propinquum</i> 1274_CJEI <i>C. propinquum</i> DSM44285
<i>Corynebacterium pseudodiphtheriticum</i>	<i>C. pseudodiphtheriticum</i> 090104 <i>C. pseudodiphtheriticum</i> DSM44287
<i>Corynebacterium simulans</i>	<i>C. simulans</i> 1B08 <i>C. simulans</i> PES1 <i>C. simulans</i> Wattiau
<i>Corynebacterium singulare</i>	<i>C. singulare</i> IBSB52218
<i>Corynebacterium striatum</i>	<i>C. striatum</i> 215 <i>C. striatum</i> 216 <i>C. striatum</i> 542CAUR <i>C. striatum</i> 587CAUR <i>C. striatum</i> 797CAUR <i>C. striatum</i> 962CAUR <i>C. striatum</i> 963CAUR <i>C. striatum</i> 1327CAUR <i>C. striatum</i> 1329CAUR <i>C. striatum</i> 1961 <i>C. striatum</i> 2023 <i>C. striatum</i> 2230 <i>C. striatum</i> 2245 <i>C. striatum</i> 2247 <i>C. striatum</i> KC-Na-01 <i>C. striatum</i> ATCC6940

Corynebacterium xerosis

C. xerosis373T

C. xerosisBmi9815

C. xerosisNBRC16721

4.2 Seleção dos genes:

Inicialmente foi selecionado um conjunto de 11 genes para as análises. Essa seleção foi fundamentada nas recomendações descritas por Lang et al. 2013, que estudou a filogenia de genomas bacterianos usando genes conservados sugerindo um conjunto de 24 genes de cópia única e onipresentes. Dentre os 11 genes selecionados para a análise desse trabalho, 5 já foram utilizados por Pal et al., 2018 em seu trabalho de genômica evolutiva comparativa de *Corynebacterium*. Kämpfer e Glaeser, 2015 sugerem que os estudos de MLSA utilizem sequências de genes que codificam proteínas com funções conservadas ('housekeeping genes') gerando árvores filogenéticas e, posteriormente, deduzindo filogenias. A Tabela 2 apresenta informações sobre estes genes.

Tabela 2 Genes usados nesse trabalho.

Gene	Descrição	Tamanho do gene (pb)
<i>rplD</i>	50S proteína ribossomal L4	663pb
<i>rplN</i>	50S proteína ribossomal L14	369pb
<i>rplP</i>	50S proteína ribossomal L16	417pb
<i>recA</i>	Recombinase recA	1134pb
<i>rpoB</i>	Subunidade beta de RNA polimerase	3477pb
<i>rpsC</i>	30S proteína ribossomal S3	747pb
<i>rpsL</i>	30S proteína ribossomal S12	372pb
<i>rpsQ</i>	30S proteína ribossomal S17	285pb
<i>rpsS</i>	30S proteína ribossomal S19	279pb
<i>dnaK</i>	Chaperona molecular DNAK	1845pb
<i>gyrB</i>	Subunidade beta de DNA girasse	2055pb

4.3 Alinhamento das sequências:

Foi selecionado um total de 49 sequências extraídas a partir da anotação realizada no servidor RAST (Rapid Annotations using Subsystems Technology) para as 13 espécies. Após a coleta dessas sequências, foi criado um arquivo agrupando as sequências de todas as linhagens para cada gene. Posteriormente foi realizado um alinhamento múltiplo no programa SeaView Versão 4 (Gouy et al., 2010), utilizando o programa externo ClustalW versão 2 (Larkin et al. 2007). O SeaView é um editor de alinhamentos de múltiplas sequências que permite adicionar ou remover uma ou várias lacunas em uma ou várias sequências simultaneamente. O SeaView conta com programas externos para executar vários alinhamentos de sequência. (Gouy, Guindon, & Gascuel, 2010). Nesse trabalho foi usado ClustalW versão 2, programa que é executado com seu valor de parâmetro padrão, anteriormente escolhido por seus autores. O SeaView também permite que as sequências de DNA codificadoras de proteínas não alinhadas sejam traduzidas para proteína, alinhadas e exibidas de novo como sequências de DNA, um procedimento que produz alinhamentos de sequências de codificação mais realistas, quando comparado ao alinhamento no nível dos nucleotídeos. Com isso cada alinhamento foi melhorado por edição manual.

4.4 Construção de árvores monogênicas

Foram inferidas árvores monogênicas utilizando o método Neighbor-Joining. Todas as posições contendo lacunas e dados perdidos foram eliminadas. Essas análises evolutivas foram conduzidas no MEGA7. Essas árvores monogênicas permitiram realizar uma pré-análise para uma avaliação das árvores baseadas em genes individuais. Foram geradas 11 árvores que foram observadas individualmente buscando inconsistências entre os ramos, observando os agrupamentos entre as espécies que já são conhecidas por serem bem definidas e os agrupamentos entre espécies diferentes. Após essas análises 8 genes foram selecionados para a análise por máxima verossimilhança e foram concatenados na seguinte ordem: *dnaK*; *gyrB*; *L4*; *L16*; *recA*; *rpoB*; *S3*; *S19*, com um total de 10608 pb.

4.5 Análise filogenética por máxima verossimilhança utilizando o Smart Model Selection

Em um alinhamento final, o arquivo com os genes *dnaK*; *gyrB*; *L4*; *L16*; *recA*; *rpoB*; *S3*; *S19* foi carregado no ambiente PhyML. O software Smart Model Selection (SMS) é implementado neste ambiente estando disponível no servidor web <http://www.atgc-montpellier.fr/phyml/>. Essa ferramenta usa estratégias para evitar o teste de todos os modelos e opções, simplificando alguns cálculos para economizar tempo de computação. SMS é um algoritmo simples, rápido e preciso para estimar grandes filogenias por máxima verossimilhança (Lefort, Longueville, & Gascuel, 2017). O arquivo gerado para análise foi exportado para o formato PHYLIP, sendo esse arquivo usado para a análise por verossimilhança, adicionado ao servidor web, utilizando as configurações padrão com o critério de seleção para substituição de modelos o Critério de Informação de Akaike (AIC).

Redes Filogenéticas: com as sequências concatenadas foi construída uma SplitsTree, utilizando o software SplitsTree4, principal aplicativo para o cálculo de redes filogenéticas não enraizadas a partir de dados de sequências moleculares. Esse programa calcula uma árvore ou rede filogenética usando métodos como decomposição dividida, rede vizinha, rede de consenso, métodos de super redes ou métodos para computar redes de hibridação ou de recombinação simples (Huson & Bryant, 2006). Nesse trabalho foi feita a análise de Split decomposition e construção da SplitsTree a partir do arquivo carregado nesse aplicativo contendo os genes *dnaK*; *gyrB*; *L4*; *L16*; *recA*; *rpoB*; *S3*; *S19* concatenados.

5. Resultados

- **Seleção das espécies**

No início desse trabalho foi feita a busca de 11 genes para cada uma das 49 linhagens entre as bactérias selecionadas do gênero *Corynebacterium* spp. Ao realizar a pesquisa desses genes no RAST, não foi possível encontrar todos os genes para o isolado *C. striatum* 2245. Levando em consideração essa indisponibilidade, retiramos esse isolado das análises. O isolado *C. xerosis* BMI9815, apresentou uma sequência muito diferente dos demais ao ser adicionado ao alinhamento. Após tentar utilizar a sequência reversa e complementar e ainda perceber claramente uma diferença nessa sequência foi feita uma análise do genoma com o GenomePeek, ferramenta on-line que analisa os arquivos de sequenciamento de genoma único e de metagenoma, de forma rápida e com baixas taxas de erro (McNair & Edwards, 2015) que confirmou uma identificação errada. Esse isolado foi identificado como *Brachybacterium tyrofermentans*, sendo esse um erro do genoma depositado no NCBI.

- **Árvores monogênicas (Neighbor-Joining)**

A análise foi realizada aplicando o método Neighbor-Joining, com valores bootstrap de 100 réplicas. Esse método é fundamentado no agrupamento de vizinhos e tem como princípio encontrar pares de unidades taxonômicas operacionais que minimizem o comprimento total do ramo em cada estágio de agrupamento. As figuras 2 e 3 apresentam as árvores filogenéticas obtidas com os genes codificadores da proteína ribossomal S17 e da subunidade beta da enzima RNA polimerase, respectivamente. Na figura 2 é possível observar que não há um bom agrupamento para isolados de *C. striatum*, grupo considerado bem definido. Assim como os isolados da espécie *C. accolens* AH4003 e *C. simulans* que nessa árvore monogênica apresenta uma relação mais próxima com os isolados de *C. striatum*. Na figura 2 está representada a árvore monogênica correspondente ao gene *ropB*, considerado o gene alvo secundário ao gene 16S, sendo reconhecido por ser um melhor método para identificação de corinebactérias. Conforme mencionado anteriormente o gene S17 foi retirado da concatenação final por apresentar agrupamentos confusos. Diferentemente, o *rpoB* foi mantido para a concatenação, pois mesmo em uma árvore monogênica é possível observar que os isolados de *C. striatum* se

agrupam corretamente. As árvores monogênicas das figuras 1 e 2 são apresentadas como um parâmetro entre o gene *rpoB*, que é reconhecidamente considerada a abordagem mais utilizada para a caracterização de espécies de *Corynebacterium* e o gene S17, que mesmo sendo um gene codificador de proteínas não apresentou um bom agrupamento entre os isolados que já são esperados.

Como podemos observar na Figura 2 a árvore monogênica não representa uma filogenia bem resolvida, quando a comparamos com análise filogenética por verossimilhança usando um conjunto de genes. Diferenças nas topologias das árvores são evidentes. Nessa figura os isolados da espécie *C. striatum* não estão claramente agrupadas, sendo esse um grupo conservado. Também é possível perceber que os isolados de *C. xerosis* e *C. freneyi* não estão claramente separados na árvore monogênica, diferente da filogenia com um conjunto de oito genes concatenados que os mostra como grupos diferentes. As árvores monogênicas foram analisadas uma a uma, buscando inconsistências para decidir quais genes seriam incluídos na concatenação final. Podemos observar que na figura 3 *C. striatum* se agrupam corretamente por *rpoB*, o que não aconteceu com o gene S17. Diante disso, as sequências nucleotídicas dos genes codificadores das proteínas S17 (representado a filogenia na figura 2), L14, e S12 foram excluídos.

- **Filogenia baseada nos critérios de Verossimilhança**

A análise por verossimilhança foi realizada utilizando o software Smart Model Selection (SMS). A figura 4 representa a árvore correspondente a essa análise, demonstrando um melhor agrupamento de alguns isolados que já são conhecidos como bem definidos, como os isolados da espécie *C. striatum*, quando comparadas às árvores monogênicas em que alguns clados não agruparam corretamente esses isolados. Isso permitiu confirmar que a MLSA pode ser uma ferramenta auxiliar para estabelecer limites de separação entre as espécies patogênicas emergentes, bem como as taxonomicamente confusas. O melhor modelo para essas análises selecionado pelo SMS foi o GTR, considerado o mais complexo e admite seis diferentes parâmetros, que são as seis possibilidades de substituição entre as bases, além de levar em conta as desigualdades na composição de nucleotídeos entre as sequências.

- **Redes filogenéticas:**

A SplitsTree construída usando SplitsTree versão 4 é apresentada na figura 5. Nesta figura podemos perceber uma relação próxima entre as linhagens de *C. minutissimum* e *C. aurimucosum*, que são consideradas taxonomicamente confusas. Essas linhagens não apresentaram uma separação clara entre os ramos. Entre essas linhagens também é possível perceber que o isolado *C. singulare* IBSB52218 apresenta uma relação próxima aos isolados de *C. aurimucosum* e *C. minutissimum*. Os ramos que representam essas relações está destacado em vermelho. Nessa mesma figura também é possível observar o agrupamento entre *C. acollens* AH4003 com os isolados de *C. simulans* em um ramo do lado oposto aos outros isolados depositados como mesma espécie. Os ramos que apresentam essas relações estão destacados em azul.

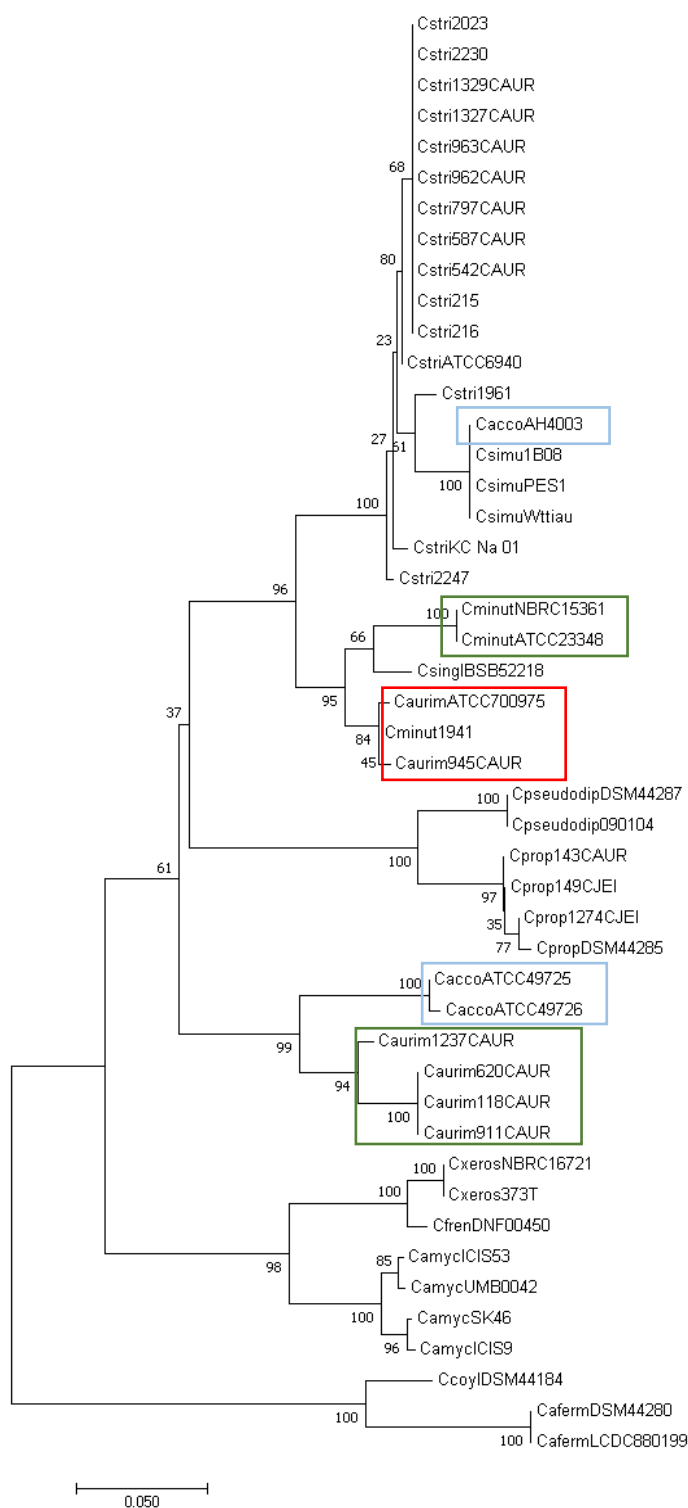


Figura 2 Árvore filogenética com base nas seqüências do gene codificador da proteína ribossomal S17 pelo método Neighbor-Joining, com 100 réplicas de bootstrap. A análise envolveu 47 seqüências nucleotídicas. Os isolados que não foram bem agrupados estão destacados. Em azul é mostrado o isolado *C. accolens* AH4003 em um clado diferente dos outros isolados dessa espécie e o agrupamento que não está bem definido para os isolados da espécie *C. striatum* incluindo os isolados *C. accolens* AH4003 e *C. simulans*. Os isolados *C. minutissimum* 1941, *C. aurimucosum* ATCC700975 e *C. aurimucosum* 945CAUR destacados em vermelho formam um clado não associado aos outros isolados dessas espécies, conforme os clados destacados em verde.

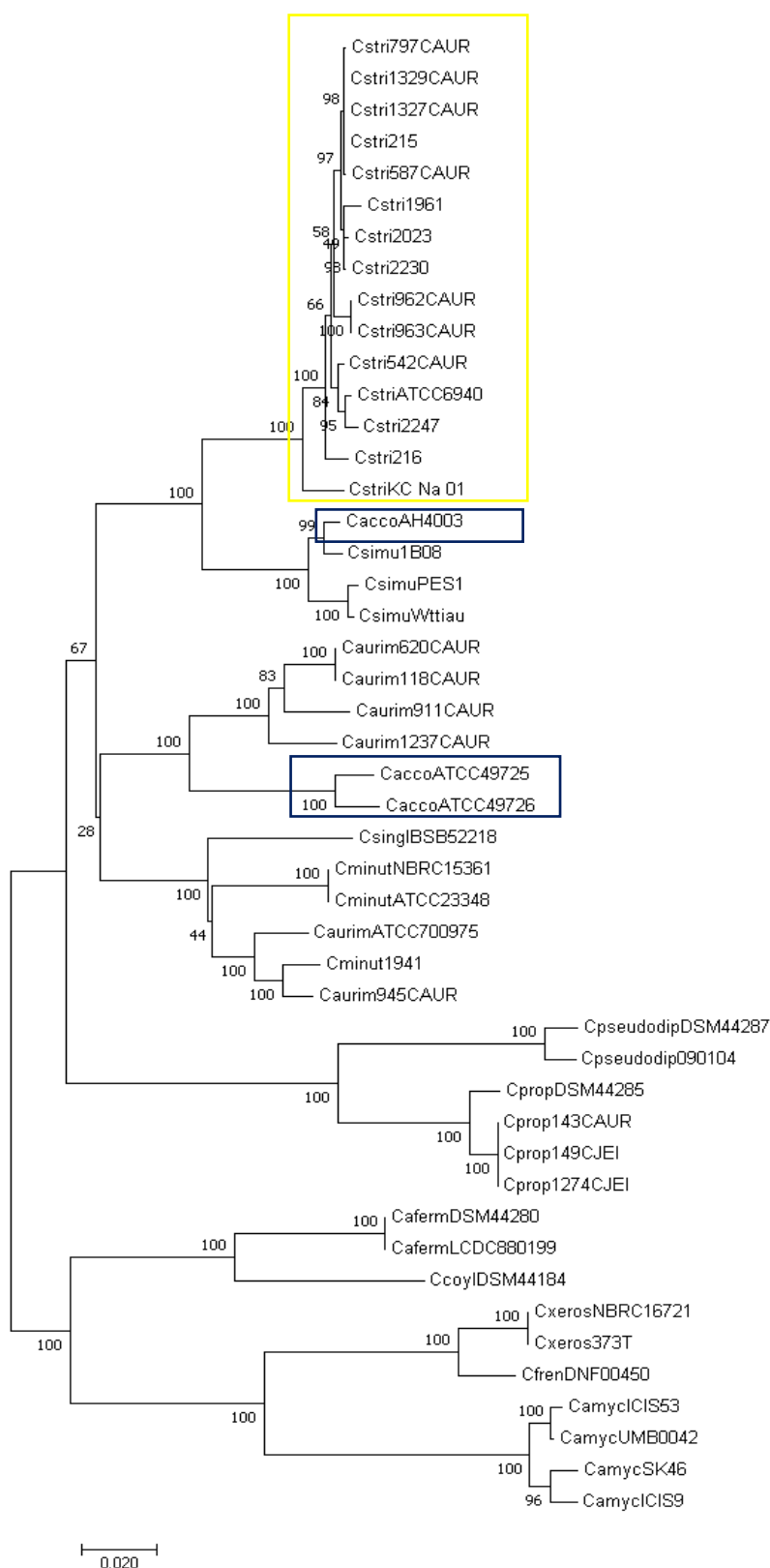


Figura 3 Árvore filogenética com base nas sequências do gene codificador da subunidade beta da enzima RNA polimerase pelo método Neighbor-Joining, com 100 réplicas de bootstrap. A análise envolveu 47 sequências nucleotídicas. O clado destacado em amarelo representa o agrupamento bem definido para os isolados da espécie *C. striatum* resultante da análise envolvendo o gene *rpoB*. Em azul estão destacados os isolados *C. accolens* em clados separados.

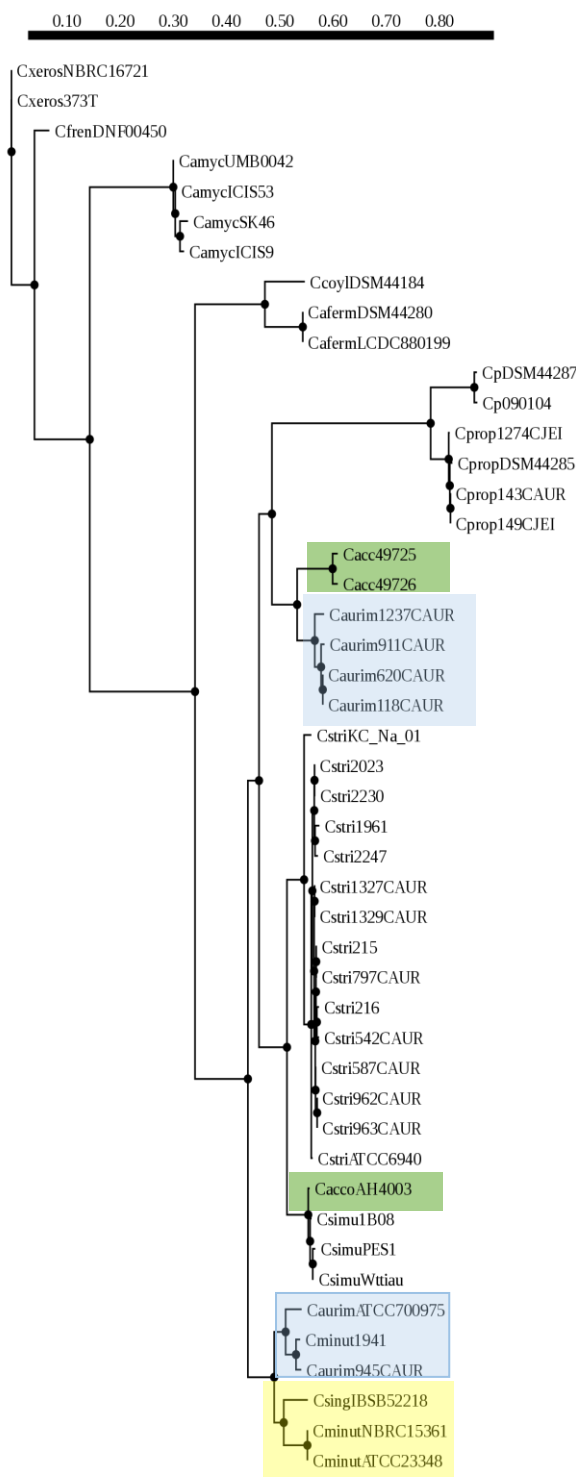


Figura 4 Árvore filogenética com base nas sequências concatenadas dos genes *DNAK*; *GyrB*; *L4*; *L16*; *RECA*; *RPOB*; *S3*; *S19*. A árvore foi construída no ambiente PhyML utilizando o SMS que selecionou o modelo GTR para essas análises. Os clados destacados em azul representam clados não associados para isolados depositados como uma mesma espécie, mostrando que algumas espécies não são taxonomicamente bem definidas. O isolado *C. accolens* AH4003, destacado em verde, não formou um clado associado com os outros isolados dessa mesma espécie, e formou um clado com os isolados da espécie *C. simulans*. Nessa análise também é possível observar a separação entre o isolado *C. singulare* IBSB52218 e *C. minutissimum* diferente das análises monogênicas.

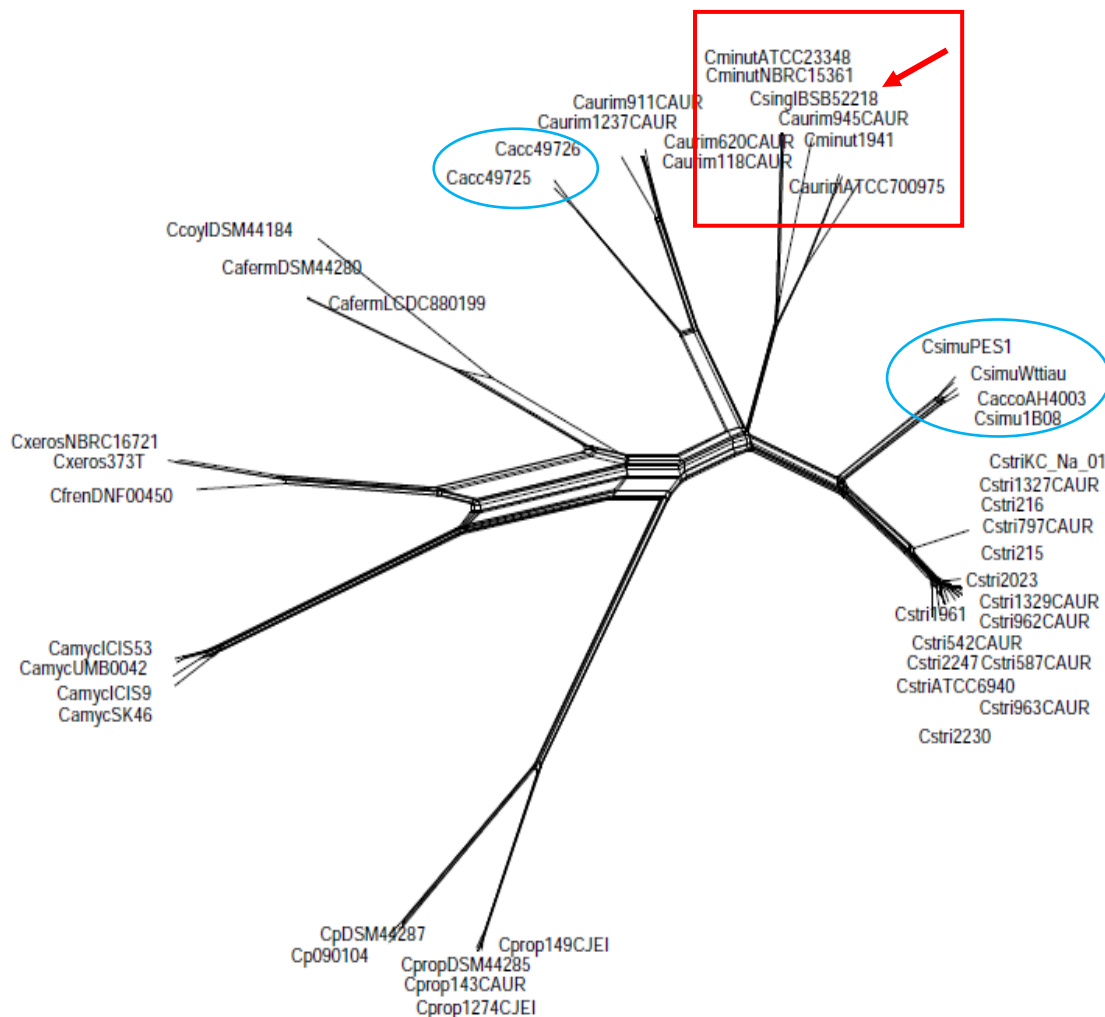


Figura 5 SplitTree baseada em oito seqüências gênicas concatenadas. Rede filogenética foi gerada usando o programa SplitsTree4. O isolado *C. accolens* AH4003 foi agrupada em um clado não associado aos outros isolados da espécie, destacado em azul à direita da rede. *C. singulare* IBSB52218 apresenta uma relação próxima aos isolados de *C. aurimucosum* e *C. minutissimum*, destacado em vermelho.

6. Discussão

Neste trabalho buscamos avaliar a análise de sequências multilocus como uma ferramenta auxiliar para a classificação de espécies patogênicas emergentes do gênero *Corynebacterium*. Vários estudos têm tentado resolver este problema, tendo em vista a reconhecida dificuldade em identificá-las apenas com testes que avaliam os perfis bioquímicos, pois as corinebactérias apresentam uma ampla variedade de processos metabólicos que não estão disponíveis no sistema Api Coryne. Além disso a infrequente atualização de dados não fornece discriminação suficiente entre as espécies recém-descritas (Alibi, 2015). Outro ponto importante que corrobora a limitação desse teste é a sua rara avaliação, já que nos últimos dez anos apenas duas avaliações foram publicadas (Bernard, 2012).

Diante de tais limitações, testes complementares são recomendados para a identificação definitiva dessas espécies. Os métodos baseados em espectrometria de massas do tipo MALDI-TOF são muito rápidos e foram desenvolvidos especificamente para identificação bacteriana. No entanto, há relatos de identificação ambígua e limitações na variedade de espécies em seus bancos de dados. As avaliações do MALDI-TOF usando o sistema Bruker Biotyper mostraram que não é possível diferenciar de maneira confiável *C. aurimucosum* de *C. minutissimum* e *C. minutissimum* de *C. singulare*. Assim como é reconhecido que a identificação diferencial de *C. striatum*, *C. amycolatum* e *C. xerosis* por testes bioquímicos permanece difícil. Várias identificações erradas foram relatadas anteriormente. Isso se deve ao fato destas espécies compartilharem muitas características fenotípicas e apresentar poucas diferenças bioquímicas, mesmo sendo geneticamente diferentes (Alibi et al., 2015).

Os métodos moleculares são recomendados para identificação definitiva de forma que os resultados bioquímicos corroborem a caracterização genética em comparação com os algoritmos de identificação comumente utilizados. Porém métodos moleculares, como sequenciamento do gene *rpoB* e 16S rRNA não são práticos para uso rotineiro devido ao seu alto custo. E ainda há relatos que essas espécies intimamente relacionadas não podem ser adequadamente diferenciadas por técnicas moleculares baseadas em sequenciamento de genes ortólogos únicos.

No estudo realizado por Alibi et al, 2015 foram avaliadas 97 linhagens recuperadas de isolados clínicos. Com suas análises eles confirmaram que houve um acordo entre Api Coryne e MALDI-TOF-MS em 88,65% dos casos, que é quase semelhante ao relatado por outros autores. E dentre as limitações encontradas nos testes bioquímicos, é reconhecido que o número limitado de substratos nos painéis favorece a identificação errônea, já que não fornecem um número suficiente de características únicas para diferenciar entre todos os taxóons relevantes.

Como já foi mencionado anteriormente as análises filogenéticas com genes codificadores de proteínas podem ser realizadas para obter uma resolução filogenética mais alta. Alguns autores demonstraram que MLSA pode resolver problemas filogenéticos entre espécies intimamente relacionadas que não podem ser diferenciadas com base no 16S rRNA. O método utilizando um conjunto de genes minimizam o viés gerado em histórias evolutivas com um único gene. Geralmente são analisados aqueles que codificam subunidades de enzimas presentes em todos os táxons, como por exemplo, a subunidade da DNA gyrase (*gyrB*), a subunidade beta da RNA polimerase (*rpoB*), recombinase A (*recA*), entre outros (Glaeser & Kämpfer, 2015).

Mesmo o gene codificador *rpoB* sendo recomendado para identificação dessas bactérias é possível observar que não há uma boa resolução filogenética semelhante a análise usando um conjunto de genes quando se compara os isolados *C. xerosis* NBRC16721; *C. xerosis* 373T e *C. freneyi* DNF00450. Também é possível perceber que o isolado *C. aurimucosum* ATCC700975 encontra-se mais próximo aos isolados de *C. minutissimum* do que os outros isolados dessa linhagem.

Os programas filogenéticos atuais fornecem uma ampla variedade de modelos para representar tanto a variabilidade de taxas entre locais quanto o processo de substituição. Para selecionar entre esses muitos modelos, os critérios estatísticos são usados para encontrar a melhor relação modelo probabilidade/complexidade. Uma abordagem simples e padrão é testar todos os modelos e depois selecionar o melhor (Lefort et al, 2017). Sabendo que o SMS é apresentado como uma ferramenta que simplifica alguns cálculos para economizar tempo de computação e seleciona o melhor modelo, utilizamos essa ferramenta para inferir a história evolutiva dos isolados selecionados para esse estudo.

Com a análise de verossimilhança é possível sugerir o isolado *C. accolens* AH4003 como *C. simulans*, devido à proximidade dos ramos. Comparando as árvores monogênicas, é

claramente notável uma resolução filogenética mais robusta dos táxons na árvore de verossimilhança. Podemos observar uma melhor separação entre as espécies. É importante notar também que há claramente dois clados bem definidos que incluem linhagens de *C. minutissimum* e *C. aurimucosum*. Isso suporta a hipótese de que essas espécies não são bem definidas taxonomicamente. Esse clados estão destacados em azul e amarelo na figura 4. Esse mesmo fica claro na SplitsTree, que também está destacado em azul, representado pela figura 5.

Nesse trabalho também foi utilizado o programa SplitsTree4 para inferir uma rede filogenética. Esse é o principal aplicativo para o cálculo de redes filogenéticas não enraizadas a partir de dados de sequências moleculares, que permite mais uma vez observar *C. accolens* AH4003 mais próximo aos isolados da espécie *C. simulans* do que as outras espécies que são diferentes. Também é possível perceber a diferença entre *C. xerosis* e *C. freneyi*, bem como é possível observar o isolado *C. aurimucosum* ATCC 700975 separado dos outros isolados dessa espécie, e a relação muito próxima os isolados *C. aurimucosum* 945CAUR e *C. minutissimum* 1941.

7. Conclusões

- A identificação errada do isolado *C. xerosis* BMi9815 permitiu confirmar que há erro do genoma depositado no NCBI;
- A análise usando um conjunto de genes *housekeeping* permitiu a confirmação de uma melhor árvore filogenética devido ao melhor agrupamento de isolados pertencentes a uma mesma espécie quando comparadas às árvores monogênicas em que alguns clados não agruparam corretamente os grupos que são conhecidos como bem definidos como *C. striatum*;
- A análise por Máxima Verossimilhança permite possível sugerir uma reclassificação para *C. accolens* AH4003 como *C. simulans*, diante do agrupamento apresentado nas árvores filogenéticas e outros resultados ainda não publicados pelo nosso grupo de pesquisa, como a análise por identidade média de nucleotídeos (ANI);
- A MLSA permitiu inferir que os isolados de *C. aurimucosum* e *C. minutissimum* não são taxonomicamente bem definidas e conseqüentemente dificulta a identificação na microbiologia clínica;
- É possível sugerir uma reclassificação taxonômica de ambas as espécies por métodos taxonômicos clássicos associados a novos métodos moleculares.

8. Referências Bibliográficas

- ALIBI, S., Ferjani, A., Gaillot, O., Marzouk, M., Courcol, R., & Boukadida, J. (2015). Identification of clinically relevant *Corynebacterium* strains by Api Coryne, MALDI-TOF-mass spectrometry and molecular approaches. **Pathologie Biologie**, 63(4–5), 153–157. <https://doi.org/10.1016/j.patbio.2015.07.007>
- AL-SAARI, N., Gao, F., Rohul, A. A. K. M., Sato, K., Sato, K., Mino, S., Sawabe, T. (2015). Advanced microbial taxonomy combined with genome-based-approaches reveals that *Vibrio astriarenae* sp. Nov., an agarolytic marine bacterium, forms a new clade in Vibrionaceae. **PLoS ONE**, 10(8), 1–17. <https://doi.org/10.1371/journal.pone.0136279>
- BERNARD, K. (2012). The genus *Corynebacterium* and other medically relevant coryneform-like bacteria. **Journal of Clinical Microbiology**, 50(10), 3152–3158. <https://doi.org/10.1128/JCM.00796-12>
- BROOKS, G.F. Microbiologia médica, 26a ed. Porto Alegre: AMGH, 2014.
- FORBES, B. A. (2017). “Did I Hear You Correctly? The Organism Identified Was *Corynebacterium diphtheriae*?” **Clinical Microbiology Newsletter**, 39(5), 35–41. <https://doi.org/10.1016/j.clinmicnews.2017.02.001>
- FOX, G.E., Wisotzkey, J.D., and Jurtshuk, P.J. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. **Int. J. Syst. Bacteriol.** 42, 166–170.
- GEVERS, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J. (2005) Re-evaluating prokaryotic species. **Nat. Rev. Microbiol.** 3, 733–739.
- GLAESER, S. P., & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. **Systematic and Applied Microbiology**, 38(4), 237–245. <https://doi.org/10.1016/j.syapm.2015.03.007>
- GOUY, M., Guindon, S., & Gascuel, O. (2010). Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. **Molecular Biology and Evolution**, 27(2), 221–224. <https://doi.org/10.1093/molbev/msp259>

- HUSON, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. **Molecular Biology and Evolution**, 23(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- KÖSER, C. U., Ellington, M. J., Cartwright, E. J. P., Gillespie, S. H., Brown, N. M., Farrington, M., ... Peacock, S. J. (2012). Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. **PLoS Pathogens**, 8(8). <https://doi.org/10.1371/journal.ppat.1002824>
- LARKIN MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
- LANG, J. M., Darling, A. E., & Eisen, J. A. (2013). Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes: Supertrees and Supermatrices. **PLoS ONE**, 8(4). <https://doi.org/10.1371/journal.pone.0062510>
- LEFORT, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. **Molecular Biology and Evolution**, 34(9), 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- MCNAIR, K., & Edwards, R. A. (2015). GenomePeek—an online tool for prokaryotic genome and metagenome analysis. **PeerJ**, 3, e1025. <https://doi.org/10.7717/peerj.1025>
- SEONG-In NA, Kim, Y. O., Yoon, S. H., Ha, S. min, Baek, I., & Chun, J. (2018). UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. **Journal of Microbiology**, 56(4), 281–285. <https://doi.org/10.1007/s12275-018-8014-6>
- OLIVEIRA, A., Oliveira, L. C., Aburjaile, F., Benevides, L., Tiwari, S., Jamal, S. B., Wattam, A. R. (2017). Insight of genus *Corynebacterium*: Ascertaining the role of pathogenic and non-pathogenic species. **Frontiers in Microbiology**, 8(OCT). <https://doi.org/10.3389/fmicb.2017.01937>
- PAL, S., Sarkar, I., Roy, A., Mohapatra, P. K. D., Mondal, K. C., & Sen, A. (2018). Comparative evolutionary genomics of *Corynebacterium* with special reference to codon and amino acid usage diversities. **Genetica**, 146(1), 13–27
- RIZVI, M., Rizvi, M. W., Shaheen, Sultan, A., Khan, F., Shukla, I., & Malik, A. (2013). Emergence of coryneform bacteria as pathogens in nosocomial surgical site infections in a

tertiary care hospital of North India. **Journal of Infection and Public Health**, 6(4), 283–288. <https://doi.org/10.1016/j.jiph.2013.01.005>

SANTOS, A. S., Ramos, R. T., Silva, A., Hirata, R., Mattos-Guaraldi, A. L., Meyer, R., Pacheco, L. G. C. (2018). Searching whole genome sequences for biochemical identification features of emerging and reemerging pathogenic *Corynebacterium* species. **Functional and Integrative Genomics**, 1–18. <https://doi.org/10.1007/s10142-018-0610-3>

SANTOS, C. S., Ramos, J. N., Vieira, V. V., Pinheiro, C. S., Meyer, R., Alcantara-Neves, N. M., Pacheco, L. G. C. (2017). Efficient differentiation of *Corynebacterium striatum*, *Corynebacterium amycolatum* and *Corynebacterium xerosis* clinical isolates by multiplex PCR using novel species-specific primers. **Journal of Microbiological Methods**, 142(August), 33–35. <https://doi.org/10.1016/j.mimet.2017.09.002>