

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA
FILHO” CAMPUS BAURU**

ANNA JULIA CHEN

GENÔMICA POPULACIONAL DO FITOPATÓGENO OBRIGATÓRIO
Candidatus Liberibacter asiaticus

BAURU
2025

Anna Julia Chen

GENÔMICA POPULACIONAL DO FITOPATÓGENO OBRIGATÓRIO
Candidatus Liberibacter asiaticus

Trabalho de conclusão de curso, apresentado à Faculdade de Ciências, Campus de Bauru, Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte das exigências para a obtenção do título de Bacharelado em Ciências Biológicas.

Orientador: Prof. Dr. Filipe Pereira Matteoli

BAURU
2025

C518g	Chen, Anna Julia Genômica populacional do fitopatógeno obrigatório <i>Candidatus Liberibacter asiaticus</i> / Anna Julia Chen. -- Bauru, 2025 44 p. : tabs., fotos, mapas Trabalho de conclusão de curso (Bacharelado - Ciências Biológicas) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências, Bauru Orientador: Filipe Pereira Matteoli 1. Microbiologia. 2. Bioinformática. 3. Bactéria. 4. Fitopatógeno. I. Título.
-------	---

ANNA JULIA CHEN

GENÔMICA POPULACIONAL DO FITOPATÓGENO OBRIGATÓRIO
Candidatus Liberibacter asiaticus

Trabalho de conclusão de curso, apresentado à Faculdade de Ciências, Campus de Bauru, Universidade Estadual “Júlio de Mesquita Filho”, como parte das exigências para a obtenção do título de Bacharelado em Ciências Biológicas.

Bauru, 28 de Novembro de 2025.

BANCA EXAMINADORA

Prof. Dr. Filipe Pereira Matteoli
Faculdade de Ciências - UNESP

Prof. Dr. Ricardo Utsunomia
Faculdade de Ciências - UNESP

Dr. Francisnei Pedrosa Silva
Afiliações

AGRADECIMENTOS

À minha mãe, que durante toda minha vida e, principalmente, nesses quatro anos de faculdade, me apoiou, sonhou comigo e me deu todo o suporte para que eu pudesse chegar até aqui. Agradeço também à minha família, meu pai e meus irmãos, que acompanharam essa jornada com muito carinho.

Aos professores que me proporcionaram todo o conhecimento e experiências que muitas vezes ultrapassaram a sala de aula e me trouxeram com muito êxito, até este momento, fazendo-me ser apaixonada pela Biologia cada dia mais.

Ao meu orientador, Filipe, que durante o desenvolvimento desse trabalho me ensinou muito além da Microbiologia, me apoiou e me fez crescer tanto como pessoa, como profissional.

Aos meus amigos que fiz nessa caminhada da graduação em Bauru, que me apoiaram todos os dias, dentro e fora da faculdade, e estiveram comigo em todos os momentos. E meus amigos de São Paulo, de quem tão cedo precisei me despedir, mas que nunca deixaram de se fazer presentes.

À UNESP e à cidade de Bauru, que me proporcionaram um lar, experiências, aprendizados e conhecimentos, pelos quais serei grata para sempre

À todos que compartilharam, com amor, esses anos tão importantes da minha vida.

“Você não percebeu que você é o único
representante do seu sonho na face da Terra?”
— *Emicida, participação de Rael*

RESUMO

As bactérias ambientais desenvolvem múltiplas relações ecológicas com as plantas, desde a fitopatogenia até a promoção do crescimento vegetal. *Candidatus Liberibacter asiaticus* (CLas) é uma bactéria ambiental, não cultivável por ser um patógeno intracelular obrigatório. Nos citros, CLas é o agente etiológico da doença Huanglongbing (HLB), um flagelo que causa impactos significativos aos produtores desta importante *commodity* agrícola. A contaminação das plantas ocorre via hospedeiro invertebrado, o psílido *Diaphorina citri*. Apesar de amplamente conhecida quanto aos seus efeitos devastadores nos citros, o manejo do HLB é feito majoritariamente por práticas que visam conter o contágio ou a dispersão do vetor, uma vez que pouco se sabe sobre os mecanismos moleculares que possibilitem o controle da infecção após a colonização da planta. Nos últimos anos, a popularização dos métodos de sequenciamento de reads longos tornou possível a obtenção de genomas montados a partir de metagenomas (MAGs) de qualidade superior. Desde então, análises bioinformática de MAGs de várias espécies de *Candidatus Liberibacter* em associação com genomas da espécie cultivável *Liberibacter crescens* estão desvendando aspectos sobre a biologia e evolução deste elusivo grupo de procariotos. Apesar dos avanços, nenhum estudo se aprofundou nas características genômicas populacionais do principal fitopatógeno do grupo, a bactéria CLas. Os dados genômicos obtidos a partir do NCBI totalizou um *dataset* de 179 *assemblies*, destes, 37 estão presente no RefSeq e 142 no GenBank. A anotação dos genes foi feita pelo software Bakta. A partir dos genomas do RefSeq foi determinado o pangenoma da espécie, no qual identificou-se a presença de 1252 genes, nos quais 944 estão presentes no core e 308 são acessórios. Essas informações possibilitaram a reconstrução da árvore filogenética desses 37 *assemblies*, realizada por máxima verossimilhança a partir dos SNPs do *core* genoma. A tentativa de uma estrutura de população não forneceu uma árvore filogenética de confiança, feita a partir dos SNPs dos grupos ortólogos dos *assemblies* do RefSeq junto com os do GenBank. Por isso, foi obtido um esquema alélico de core genoma (cgMLST) e uma Minimal Spanning Tree que revelou três grupos clonais bem delineados. Os resultados obtidos contribuem para o entendimento da estrutura genômica de *Candidatus Liberibacter asiaticus* e fornecem subsídios para futuros estudos comparativos e evolutivos.

ABSTRACT

Environmental bacteria develop multiple ecological relationships with plants, ranging from phytopathogenesis to plant growth promotion. *Candidatus Liberibacter asiaticus* (CLas) is an environmental bacterium that is unable to be cultivated due to its obligate intracellular lifestyle. In citrus, CLas is the etiological agent of Huanglongbing disease (HLB), a scourge that causes significant impacts to producers of this important agricultural commodity. It infects plants via an invertebrate host, the psyllid *Diaphorina citri*. Although HLB is widely known for its devastating effects on citrus, it is mostly managed through practices aimed at containing the infection or dispersal of the insect vector as little is known about the molecular mechanisms involved in the plant infection. In recent years, due to the popularization of long-read sequencing methods, it has become possible to obtain high-quality metagenome-assembled genomes (MAGs). Since then, bioinformatic analyses of MAGs from different *Candidatus Liberibacter* species, along with genomes from the culturable species *Liberibacter crescens*, are unraveling aspects of the biology and evolution of this elusive group of prokaryotes. Despite these advances, no study has addressed the population genomics features of the group's major phytopathogen, the CLas bacterium. Genomic data obtained from NCBI totaled a dataset of 179 assemblies, of which 37 were from RefSeq and 142 from GenBank. Gene annotation was performed using the Bakta software. From the RefSeq genomes, the species' pangenome was determined, in which 1,252 genes were identified, of which 944 are present in the core and 308 are accessory. This information enabled the reconstruction of the phylogenetic tree of these 37 assemblies, performed by maximum likelihood based on SNPs from the core genome. The attempt to infer a population structure did not yield a reliable phylogenetic tree, generated from SNPs of orthologous groups of the RefSeq assemblies together with those from GenBank. Therefore, a core genome allelic scheme (cgMLST) and a Minimal Spanning Tree were obtained, which revealed three well-defined clonal groups. The results obtained contribute to the understanding of the genomic structure of *Candidatus Liberibacter asiaticus* and provide support for future comparative and evolutionary studies.

LISTA DE FIGURAS

Figura 1. Mapa da produção de laranja - valor em reais	13
Figura 2. Laranja com a doença HLB	14
Figura 3. <i>Diaphorina citri</i>	16
Figura 4. Matriz de presença e ausência do pangenoma	27
Figura 5. Curva de acumulação do Pangenoma	28
Figura 6. Árvore filogenética de CLas dos <i>assemblies</i> do RefSeq	29
Figura 7. SNPs em comparação ao tamanho de cada gene	30
Figura 8. Árvore filogenética de CLas dos <i>assemblies</i> do GenBank	31
Figura 9. Distribuição do número de alelos por loci.	32
Figura 10. Esquemas de cgMLST	33
Figura 11. <i>Minimum Spanning Tree</i>	34

LISTA DE TABELAS

Tabela 1. Genomas de <i>Candidatus Liberibacter asiaticus</i> disponíveis no Refseq	23
--	----

LISTA DE ABREVIATURAS E SIGLAS

CLas	<i>Candidatus Liberibacter asiaticus</i>
Ca. Liberibacter	<i>Candidatus Liberibacter</i>
<i>D. Citri</i>	<i>Diaphorina citri</i>
HLB	<i>Huanglongbing</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
Indels	Inserção e Deleção
<i>L. crescens</i>	<i>Liberibacter crescens</i>
MAGs	<i>Metagenome-assembled genomes</i>
NCBI	National Center for Biotechnology Information
PIB	Produto Interno Bruto
PNCHLB	Programa Nacional de Prevenção e Controle do HLB
SNPs	<i>Single Nucleotide Polymorphisms</i>
TRNs	<i>Tandem Repeat Numbers</i>
<i>X. citri</i> pv. <i>citri</i>	<i>Xanthomonas citri</i> patovar <i>citri</i>
<i>X. fastidiosa</i>	<i>Xylella fastidiosa</i>
USDA	<i>United States Department of Agriculture</i>

SUMÁRIO

1. INTRODUÇÃO.....	12
1.1 Citricultura e <i>Huanglongbing</i>	12
1.2 Genômica de fitopatógenos.....	16
1.3 Genômica de <i>Candidatus Liberibacter asiaticus</i> (CLas)	18
2. JUSTIFICATIVA	21
3. OBJETIVO GERAL.....	21
3.1 Objetivos específicos	21
4. MATERIAIS E MÉTODOS.....	22
4.1 Obtenção dos genomas e anotação de genes.....	22
4.2 Pangenoma de <i>Candidatus Liberibacter asiaticus</i>	24
4.3 Árvore filogenética de <i>Candidatus Liberibacter asiaticus</i>	24
4.4 Estrutura populacional	24
4.4.1 Obtenção dos genes ortólogos mais informativos	25
4.5 Construção do esquema alélico e <i>Minimal Spanning Tree</i>	25
5. RESULTADOS	26
6. DISCUSSÃO	35
7. CONCLUSÃO.....	38
REFERÊNCIAS	38

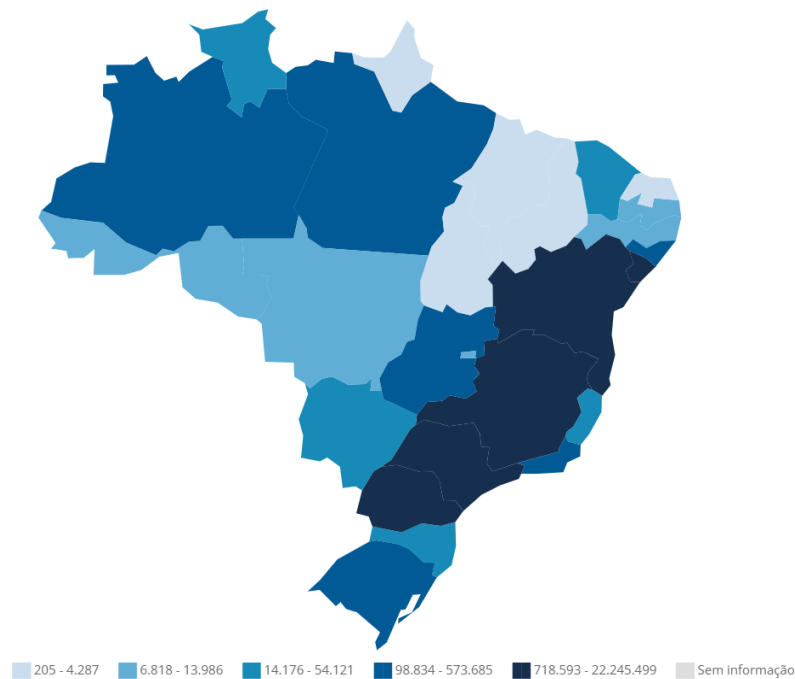
1. INTRODUÇÃO

1.1 Citricultura e *Huanglongbing*

A citricultura é uma das cadeias agrícolas mais relevantes do mundo, que movimenta bilhões de dólares anualmente sustentando economias locais em regiões tropicais e subtropicais. Laranjas, limões, tangerinas bem como pomelos estão entre as frutas mais produzidas e comercializadas globalmente, constituem fonte essencial de renda agrícola, emprego rural, além disso, matérias-primas para as indústrias de sucos, aromas assim como cosméticos. A produção mundial concentra-se principalmente em países como Brasil, Estados Unidos, China, Índia e México, que juntos respondem pela maior parte do mercado global de frutas cítricas bem como derivados (Barbieri et al., 2023). Porém, a ampla distribuição geográfica das culturas associada à baixa variabilidade genética dos cultivares tornam o setor altamente vulnerável a doenças de difícil controle, como o *Huanglongbing* (HLB), que ameaça a sustentabilidade econômica e ambiental da citricultura em escala global.

Nos Estados Unidos, o HLB provocou queda drástica na produção e na área citrícola, especialmente na Flórida, onde a produção de laranja caiu 90% desde o descobrimento da epidemia neste país em 2005 até 2022 (Graham et al., 2024). Na China, o HLB também teve impacto massivo: estudos junto com levantamentos regionais registram a remoção de milhões de árvores infectadas em províncias como Guangxi e Gannan (Liu et al., 2024). No Brasil, o agronegócio possui um papel fundamental na balança comercial do país, o Produto Interno Bruto (PIB) das atividades agrícolas foi de R\$ 2,45 trilhões em 2024 (Barros et al., 2024). Dentre as commodities agrícolas, a citricultura possui destaque, segundo o relatório do *United States Department of Agriculture*, a produção global de laranjas na safra de 2019/2020 foi estimada em 47,5 milhões de toneladas, das quais aproximadamente 15,1 milhões foram provenientes do Brasil, o que confere ao país o *status* de maior produtor global de laranjas (USDA, 2025). Os principais estados produtores são Paraná, São Paulo e Minas Gerais (Figura 1), segundo um levantamento feito pelo IBGE (2025) esses estados produziram mais de 24 milhões de reais em 2024, com a maior produção desde os últimos 4 anos.

Figura 1. Mapa da produção de laranja - valor em reais



Legenda. Mapa do Brasil indicando o valor da produção de laranjas no Brasil por Estados em 2024.

Fonte: IBGE (2024)

O HLB, amplamente conhecido como *Greening* foi descrito como fitopatologia em 1919 na China, seu nome significa “doença do ramo amarelo” (Figura 2). No Brasil, a doença passou a ganhar notoriedade após a primeira identificação de *Candidatus Liberibacter asiaticus* (CLas) em 2004, por meio de reação em cadeia da polimerase (PCR) em folhas de *Citrus sinensis* cvs. Hamlin e Valencia obtidas em Araraquara – SP (Coletta-Filho et al., 2004). Subsequentemente, 208 municípios dos estados de São Paulo, Minas Gerais e Paraná notificaram a ocorrência da doença, o que prontificou o Fundo de Defesa da Citricultura (Fundecitrus) a desenvolver programas direcionados ao combate do HLB, por meio de serviços gratuitos de diagnósticos por PCR, campanhas de conscientização, otimização de estratégias de manejo de citros e controle do inseto vetor de acordo com a Secretaria de Agricultura e Abastecimento do Estado de São Paulo (SÃO PAULO, 2023). O incremento do esforço de diagnóstico mostrou aumento na proporção de plantas infectadas com CLas (Lopes et al., 2009).

Figura 2. Laranja com a doença HLB



Legenda. Laranja contaminada com *Candidatus Liberibacter asiaticus*, manchas amareladas por sua casca. Fonte: Jornal Ibiá, 2022

O HLB atualmente é encontrado nos estados do Mato Grosso do Sul, Minas Gerais, Paraná, Santa Catarina e São Paulo; o que o torna uma preocupação nacional e que prontificou a publicação de uma portaria da Secretaria de Defesa Agropecuária do Ministério da Agricultura e Pecuária (BRASIL, 2025). A portaria Nº 1326 institui o Programa Nacional de Prevenção e Controle do HLB (PNCHLB) que é em essência um programa fitossanitário. O documento caracteriza os agentes etiológicos desta como CLas e *Candidatus Liberibacter americanus*. Além disso, a portaria estabelece quatro eixos principais de ação: i. controle dos estabelecimentos fornecedores de mudas de plantas hospedeiras, ii. inspeções fitossanitárias em imóveis rurais e iii. monitoramento do inseto vetor em áreas de risco, iv. restrições de transporte e manejo de *Murraya paniculata* (popularmente conhecida como murta-de-cheiro).

O primeiro eixo institui que os viveiros e fornecedores de material de propagação devem cumprir exigências rigorosas de cadastro, rastreabilidade e verificação periódica, de modo a assegurar que as mudas comercializadas estejam livres das bactérias causadoras do HLB. Já no segundo, o foco é nas propriedades rurais que cultivam plantas-hospedeiras, a norma exige inspeções regulares, com a obrigação de identificar e remover plantas com sintomas suspeitos e a manutenção de cadastro georreferenciado para controle sanitário mais eficaz. Já o terceiro eixo almeja o monitoramento do inseto vetor em áreas de risco (com infecção ativa para o HLB) e que Unidades Federativas sem registro mantenham levantamento anual do vetor *Diaphorina*

citri em zonas consideradas de risco, com ampla cobertura geográfica e sistema de armadilhas para captura e análise do inseto. Por fim, o quarto eixo restringe o transporte e o manejo de *Murraya paniculata*, com proibição ou limitação de seu transporte interestadual ou intermunicipal nas áreas com HLB ativo, devido ao uso desta planta ornamental como hospedeira alternativa de *D. citri*, o que pode contribuir para a propagação da doença (Lestiyani et al., 2024).

O monitoramento e prevenção são importantes pois os sintomas característicos do HLB tardam meses a aparecer, entre estes são estão a formação de manchas nas folhas denominado padrão mosqueado clorótico, perda de folhas, tamanho assimétrico dos frutos, o que afeta o sabor e a textura, além da queda prematura, o que causa expressiva redução na produtividade e na qualidade dos frutos; além disso, o florescimento fora da época e pode levar a planta à morte (Nehela; Killiny, 2020). Desde 2016, bactericidas contendo oxitetraciclina são utilizados em aplicação floria via *spray* como forma de controle da infecção (Li et al., 2019), mais recentemente, tratamentos injetáveis também estão sendo testados (Roldán; Stelinski; Pelz-Stelinski, 2025).

A infecção da bactéria na planta pode ocorrer pela enxertia de tecidos infectados, porém a via principal é pelo vetor psíldeo *D. citri*, um inseto da ordem Hemiptera, comumente chamado de Psíldeo-asiático-dos-citros (Figura 3), a contaminação ocorre quando o vetor se alimenta do floema da planta hospedeira, há preferência pelas folhas mais jovens pois é onde ocorre a alimentação e reprodução desse inseto (Carmo-Sousa; Cortés; Lopes, 2020). *D. citri* se desenvolve e reproduz em climas tropicais e subtropicais, apesar disso, um experimento conduzido demonstrou que os insetos adultos conseguem sobreviver a temperaturas mais baixas, essa sobrevivência explica a sua ampla distribuição geográfica (EUROPEAN FOOD SAFETY AUTHORITY, 2021). O primeiro registro de *D. citri* ocorreu na China em 1934 (Wu et al., 2025). No Brasil, a primeira descrição deste inseto ocorreu em 1942 (Costa Lima, A. M., 1942), atualmente, a maior densidade populacional do psíldeo foi encontrada em municípios da Bahia, em contrapartida, nos pomares dos estados do Amazonas e Pará, que não são endêmicos para o HLB, este inseto vetor não foi reportado até o momento (Noronha et al., 2019).

Figura 3. *Diaphorina citri*



Legenda. *Diaphorina citri*, inseto vetor da doença HLB em cima de um ramo de planta. Foto: Nilton Fritzons Sanches. Fonte: Revista Cultivar (2025) Embrapa, 2018

No psílídeo, CLas se aloja no interior de enterócitos, onde induz a formação de corpúsculos associados ao retículo endoplasmático. Esses corpos intracelulares fundem-se a vacúolos contendo as bactérias endocitadas, o que produz compartimentos modificados que favorecem a replicação bacteriana. Esse processo também provoca alterações na organização do núcleo e na arquitetura geral das células epiteliais do intestino do inseto (Ghanim et al., 2017). Interessantemente, a biologia celular da infecção de CLas em citros é distinta da observada em *Candidatus Liberibacter solanacearum* com seu hospedeiro psílídeo (*Bactericera trigonica*), visto que esta última bactéria induz autofagia dos enterócitos a partir da interação dos corpúsculos (Sarkar; Jassar; Ghanim, 2023); o que indica diversidade entre os mecanismos de patogênese intracelular no âmbito do complexo *Candidatus Liberibacter*.

1.2 Genômica de fitopatógenos

A genômica é uma ferramenta importante no estudo da biologia de fitopatógenos bacterianos, pois permite a compreensão detalhada dos mecanismos de adaptação, virulência e interação dos fitopatógenos com os hospedeiros vegetais. O sequenciamento e a análise comparativa de genomas podem revelar genes e ilhas genéticas associadas à patogenicidade, resistência a estresses e capacidade de colonização, além de fornecerem subsídios para a identificação de marcadores moleculares e alvos para controle biotecnológico (Goss, 2015;

Nizamani et al., 2023). Esses avanços possibilitam a reconstrução de rotas metabólicas, monitoramento de variantes emergentes e o entendimento das bases genéticas de fenótipos relevantes, como especificidade de hospedeiro e evasão da resposta imune vegetal.

Análises comparativas mostraram que alguns isolados do fitopatógeno de rosáceas, *Erwinia amylovora*, possuem o sistema de secreção tipo IV (T4SS) codificados por plasmídeos, os quais podem contribuir para a transferência horizontal de genes e modular a interação com a planta hospedeira (Parcey et al., 2020). Além disso, múltiplos sistemas de secreção estão envolvidos na virulência de fitopatógenos, incluindo o sistema tipo III (T3SS), essencial para a injeção de efetores que suprimem respostas de defesa da planta, e o tipo VI (T6SS), associado à competição bacteriana e colonização epifítica (Filloux, 2022). O genoma de *E. amylovora* também abriga ilhas de patogenicidade, genes de síntese de exopolissacarídeos como *ams* e *lev*, e reguladores de *quorum sensing*, que atuam para a formação de biofilmes e desenvolvimento dos sintomas característicos da doença do fogo-bacteriano (Rezzonico et al., 2024). Análises comparativas entre cepas de diferentes hospedeiros e regiões geográficas indicam variação significativa em plasmídeos e elementos móveis, sugerindo que a plasticidade genômica desempenha papel central na adaptação e evolução desse fitopatógeno.

Outro grupo complexo de fitopatógenos, o gênero *Xanthomonas*, alberga múltiplas espécies capazes de infectar cana-de-açúcar, feijão, mandioca, repolho, banana, frutas cítricas, tomate, pimenta e arroz. Neste gênero, as mais de 35 espécies se agrupam em subespécies ou patovares, com ampla diversidade genômica (Okoh et al., 2024). *Xanthomonas citri* pv. *citri* é um importante patógeno de citros, capaz de causar o cancro cítrico. Estudos genômicos populacionais demonstraram que há três grupos populacionais distintos de *X. citri* pv. *citri*, A, A* e Aw, os quais diferem em seu espectro de hospedeiros e forma de elicitação da resposta imune da planta. Além disso, revelou-se que o genoma acessório desta espécie, possui genes que codificam T4SS e genes de resistência a metais pesados que diferem entre os grupos populacionais, e podem configurar-se como alvos moleculares para a rápida diferenciação entre estes grupos (Webster; Bogema; Chapman, 2020).

Xylella fastidiosa é um dos fitopatógenos de maior relevância agrícola, causador de doenças devastadoras em culturas como café, citrus e videira, sendo responsável por perdas econômicas significativas no Brasil. Desde seu sequenciamento (Simpson et al., 2000), a espécie tornou-se simbólica por representar a primeira bactéria a ter seu genoma sequenciado no país, possibilitando a expansão dos estudos genômicos em fitopatologia. Pesquisas subsequentes demonstraram que a diversidade genética de *X. fastidiosa* está fortemente associada à presença de elementos genéticos móveis, como plasmídeos e outros elementos

genômicos transferíveis (Pierry et al., 2020), que podem influenciar a adaptação a diferentes hospedeiros e a virulência. Além disso, características metabólicas únicas da bactéria contribuem para seu crescimento lento e fastidioso, dificultando seu cultivo em condições laboratoriais e impactando o desenvolvimento de estratégias de controle (Gerlin et al., 2020).

Enquanto a absoluta maioria dos fitopatógenos são bactérias ambientais cultiváveis, o próprio gênero *Liberibacter* alberga uma variedade de outros fitopatógenos não cultiváveis. Uma análise filogenética ampla por meio de 345 genes ortólogos de CLas, outras espécies não cultiváveis, *Liberibacter crescens*, a única espécie cultivável do clado, e outras *Rhizobiales*; demonstrou que as espécies de *L. crescens* e *Ca. Liberibacter* se separaram de *Rhizobiales* em um processo de evolução passo-a-passo há cerca de 250–370 Ma. A separação das espécies fitopatogênicas ocorreu posteriormente há cerca de 132 Ma; os processos de especiação continuaram de forma sucessiva e o ancestral de CLas sofreu especiação há 63 Ma. Uma vez que os padrões de fitopatogenicidade variam de acordo com as espécies de *Ca. Liberibacter* foi possível inferir que o fenótipo fitopatogênico foi adquirido por eventos de perda/ganho de genes (Thapa et al., 2020).

1.3 Genômica de *Candidatus Liberibacter asiaticus* (CLas)

O *Candidatus Liberibacter asiaticus* é uma bactéria Gram-negativa, intracelular obrigatória e fitopatogênica (Achor et al., 2020), pertencente à família *Rhizobiaceae*, classe Alphaproteobacteria e filo Pseudomonadota. O status de táxon candidato relaciona-se com a sistemática de bactérias, que condiciona a descrição de novas espécies ao isolamento e cultivo *in vitro*, em meio axênico. CLas caracteriza-se pelo comportamento fastidioso, ou seja, requer altos níveis nutricionais, além de condições restritas de temperatura, pH e pressão osmótica o que impossibilita seu cultivo em meios de cultura sintéticos (Andrade; Li; Wang, 2020).

Enquanto a maioria dos fitopatógenos ambientais é dotado de ampla variabilidade genética, com elementos gênicos associados aos grupos populacionais, CLas possui um genoma compacto com cerca de 1,2 Mb e pouco mais de mil genes codificantes. Dentre estes, alguns genes determinantes para o fenótipo fitopatogênico, são oriundos de processos de transferência gênica genes codificantes de toxinas desconhecidas, homólogas às proteínas efetoras EspG/VirA, fosfolipase YdjM, e a endonuclease/exonuclease secretada (EEP) (Tan et al., 2021).

Em outubro de 2025, existem 187 genomas de CLas disponíveis no GenBank, todos são *metagenome-assembled genomes* (MAGs). A metagenômica é uma técnica de sequenciamento

de DNA diretamente a partir de amostras ambientais cuja montagem posterior dos *reads* permite o estudo de populações microbianas de forma independente do cultivo (Setubal, 2021). Após o sequenciamento, os *reads* são montados em *contigs* (fragmentos contíguos de DNA), agrupados em *bins* que representam genomas parciais ou completos; estes últimos são denominados MAGs (Kim; Pongpanich; Porntaveetus, 2024).

Os estudos envolvendo MAGs são uma marca da microbiologia ambiental pós sequenciamento de 3ª geração, ou *long-read sequencing*, ajudaram a compreender inúmeros processos mediados por microrganismos não cultiváveis. A depender da plataforma para sequenciamento o comprimento da leitura pode conter de 10–20 Kb, o que confere cobertura mais ampla e produz *assemblies* menos fragmentados. Os *reads* longos otimizam a detecção de variantes nos genomas como indels, variações estruturais, rearranjos complexos, e também são frequentemente utilizados associados aos *reads* curtos em *assemblies híbridos* (Kim; Pongpanich; Porntaveetus, 2024).

A primeira cepa de CLas do Brasil sequenciada foi a 9PA, o *assembly* é composto por 157 contigs e com a integração de outros dados da mesma sequência de metagenomas que fecharam todas as lacunas no genoma. Resultando na montagem bem-sucedida de um cromossomo circular com 1,231 Mb e teor de GC de 36,5% (Dutra et al., 2024). Um outro estudo sequenciou e montou genomas completos de 135 cepas de CLas de dez províncias de citros na China, os *assemblies* produzidos variaram de 1,221 Mb a 1,308 Mb (Zheng et al., 2024).

Os esforços de sequenciamento genômico revelaram que os genomas de CLas são caracterizados pela ausência de inúmeros genes essenciais ao metabolismo procariótico, o que indica dependência metabólica em relação aos seus hospedeiros para a obtenção de aminoácidos e ATP (Labbé et al., 2025). Outra peculiaridade importante é a prevalência de regiões horizontalmente transferidas por profagos, que podem representar até 11% do tamanho total dos genomas e que albergam maior variabilidade genética (Thapa et al., 2020). Atualmente três fagos dsDNA da família *Podoviridae* foram descritos em diversos genomas: Type 1 (SC1-like ou P-YN-1-like), Type 2 (SC2-like ou P-GD-2-like), e Type 3 (P-JXGC-3-like) (Zhang et al., 2011; Zheng et al., 2018). Além disso, outros profagos foram reportados em *assemblies* específicos. Análises transcriptômicas evidenciaram que cepas de CLas contendo o fago P-GD-2 se replicam mais rapidamente e apresentam maior virulência do que cepas contendo o fago P-YN-1. Estimativas de número de cópias dos fagos por célula indicam que P-YN-1 produz múltiplas cópias e expressa transcritos associados ao ciclo lítico, ao passo que P-GD-2 possui uma única cópia e expressa genes de lisogenia (Zheng et al., 2023). Em conjunto esses dados

revelam que a tipagem do fago está intrinsicamente correlacionada com o perfil de patogenicidade exibido pelas cepas de CLas.

Os primeiros estudos de diversidade genética de CLas basearam-se em marcadores simples, como o número de repetições em tandem (*tandem repeat numbers*, TRNs) na região CLIBASIA_01645 que permitiram distinguir populações de diferentes origens geográficas, como Guangdong (China) e Flórida (EUA) (Chen et al., 2010). No estado americano, dois genótipos com TRN < 10 foram encontrados, um restrito à Flórida central e outro distribuído em todo estado, já na China as variações foram heterogêneas e não correlacionadas com a geografia (Katoh et al., 2011; Ma et al., 2014). Já Wang et al., (2012) identificaram profundo mosaïcismo genético em outra região, o locus CLIBASIA_05640 a CLIBASIA_05650 e na região intergênica adjacente.

A incongruência de um marcador adequado e a estreita relação com os fagos, levaram os pesquisadores a correlacionarem o perfil de fagos com a estrutura populacional desta bactéria. Através do sequenciamento de Sanger, (You et al., 2024), analisaram a diversidade genética de três loci para analisar a diversidade genética de profagos em 500 amostras de CLas na China. Os resultados mostraram que o profago Type 2 (P-GD-2) é o mais prevalente, presente em 55% das amostras. Além disso, as populações que possuem o profago Type 1 (P-YN-1) se mantêm estáveis, enquanto as que possuem os profagos Type 2 (P-GD-2) e Type 3 (P-JXGC-3) se expandem.

A caracterização genética de *Candidatus Liberibacter asiaticus* também tem sido tradicionalmente realizada por meio da amplificação de genes conservados, utilizados como marcadores em estudos de diversidade populacional e filogenia. Entre os principais alvos empregados destacam-se os genes 16S rRNA, 23S rRNA *intergenic spacer region*, omp (*outer membrane protein*), nrdB (*deoxy-ribonucleotide reductase*), β -operon, e o cluster gênico tufB–secE–nusG–rplKAJL–rpoB (Gao et al., 2022). Embora amplamente utilizados, esses marcadores apresentam limitações na detecção de regiões genômicas hipervariáveis, o que restringe a compreensão da diversidade total do patógeno.

O único estudo genômico amplo que não restringiu previamente a detecção de variabilidade genética em uma região foi realizado por (Gao et al., 2022). O estudo obteve as regiões variáveis por alinhamento de 34 *assemblies* a uma referência (a cepa GXPSY), *accession* CP004005). Após filtrar os 6012 polimorfismos de nucleotídeo (*single nucleotide polymorphisms* - SNPs) único encontrados, quatro regiões genômicas hipervariáveis, duas de origem cromossômica e duas de origem viral, foram utilizadas como alvos de *quantitative* PCR (qPCR) para analisar 1788 amostras de diferentes citros sintomáticos ou não para HLB em

regiões da China. Os resultados não encontraram associação entre a diversidade genética das regiões genômicas hipervariáveis e a localização geográfica, sintomas do HLB ou variedade de citro afetada.

Assim, mesmo com sucessivos avanços na genômica de CLas, ainda não foi obtida uma estrutura populacional capaz de descrever satisfatoriamente a filogeografia desta bactéria. Por isso, este trabalho almeja obter a estrutura populacional de CLas por meio de análises genômicas.

2. JUSTIFICATIVA

Diante das informações disponíveis, observa-se que *Candidatus Liberibacter asiaticus* (CLas) é um fitopatógeno de interesse agrícola com relevância internacional. Por isso, torna-se cada vez mais importante estudar e entender melhor os processos evolutivos intra-espécie. Dessa forma, este trabalho de conclusão de curso situa-se na fronteira do conhecimento devido ao uso de genômica comparativa de metagenomas de CLas para estudar o pangenoma e a estrutura populacional desta bactéria.

3. OBJETIVO GERAL

Propõe-se no presente estudo determinar o pangenoma, a filogenia e a estrutura populacional de *Candidatus Liberibacter asiaticus*, a fim de compreender a diversidade genética e os padrões epidemiológicos dessa bactéria, contribuindo para o entendimento de sua adaptação e disseminação em diferentes regiões citrícolas do mundo.

3.1 Objetivos específicos

- Avaliar a qualidade e a completude dos metagenome-assembled genomes (MAGs) de *Candidatus Liberibacter asiaticus*
- Determinar e estudar o pangenoma da espécie, identificar o *core genes* e os genes acessórios
- Realizar uma filogenia robusta com base no *core* genoma
- Delimitar agrupamentos intra-espécie por meio de análises de estrutura populacional
- Correlacionar os agrupamentos genômicos com informações geográficas disponíveis.

4. MATERIAIS E MÉTODOS

4.1 Obtenção dos genomas e anotação de genes

Os *assemblies* de metagenomas foram obtidos a partir do *dataset do National Center for Biotechnology International* (NCBI), para o *download* utilizou-se a *flag* para baixar apenas os genomas depositados no banco de dados do Refseq do NCBI (Goldfarb et al., 2025).

Após a verificação acerca dos metadados, construiu-se o conjunto de dados (Tabela 1), composta por 37 genomas de *Candidatus Liberibacter asiaticus*, dos quais quinze estão montados em apenas um *contig*, o que os classifica como completos. A maioria oriunda da China e dos Estados Unidos da América, com apenas oito genomas de outras localidades, incluindo uma do Brasil. Após a análise dos metadados, foi escolhido o *software* Bakta (Beyvers et al., 2025) para a anotação dos genes, com a opção *–meta* que serve para genomas provenientes de metagenomas.

Tabela 1. Genomas de *Candidatus Liberibacter asiaticus* disponíveis no Refseq

Cepa	Nível de montagem	Contigs	Sequenciamento	Local de Isolamento
A4	Completo	1	Illumina MiSeq	China: Guangdong
JRPAMB1	Completo	1	PacBio Sequel	USA: Florida
9PA	Completo	1	Illumina NextSeq	Brazil: São Paulo, Matão
CoFLP1	Completo	1	Illumina HiSeq	Colombia: Municipio Dibulla
TaiYZ2	Completo	1	Illumina HiSeq	Thailand: Songkhla
PGD	Completo	1	Illumina HiSeq	China: Guangdong
PYN	Completo	1	Illumina HiSeq	China: Yunnan
GDCZ	Completo	1	Illumina HiSeq; PacBio Sequel	China: Guangdong
YNBC	Completo	1	Illumina HiSeq	China: Yunnan
GDCH12	Completo	1	Illumina HiSeq	China: Guangdong
psy62	Completo	1	-	USA: Florida
gxpsy	Completo	1	Illumina HiSeq	USA: Washington D.C
FL17	Contig	3	Illumina MiSeq	USA: Florida
Ishi-1	Completo	1	Illumina GAIIx; HiSeqTM 2000	Japan: Okinawa, Ishigaki
YCPsy	Contig	9	Illumina MiSeq	China: Guangdong
JXGC	Completo	1	Illumina HiSeq	China: Jiangxi
AHCA1	Chromosome	21	Illumina HiSeq	USA: Anaheim, California
YNJS7C	Contig	3	Illumina HiSeq	China: Yunnan
PA19	Contig	5	Illumina HiSeq	Pakistan
ReuSP1	Completo	1	Illumina HiSeq	France: La Reunion
SGCA5	Contig	56	Illumina MiSeq	USA: California, San Gabriel
DUR1TX1	Contig	59	Illumina HiSeq	USA: Texas
Mex8	Contig	60	Illumina HiSeq	Mexico: Mexicali
SGCA16	Contig	35	Illumina HiSeq	USA: San Gabriel
LBR19TX2	Contig	28	Illumina HiSeq	USA: Texas
LBR23TX5	Contig	30	Illumina HiSeq	USA: Texas
YNXP-1	Contig	43	Illumina HiSeq	China: YunNan
JXGZ-1	Contig	43	Illumina HiSeq	China: JiangXi
CA-Root	Scaffold	39	Illumina MiSeq	USA: California, Anaheim
PA20	Contig	7	Illumina HiSeq	Pakistan
DUR2TX1	Contig	160	Illumina HiSeq	USA: Texas
GFR3TX3	Contig	80	Illumina HiSeq	USA: Texas
CHUC	Contig	69	Illumina NextSeq	China
CRCFL16	Contig	232	Illumina MiSeq	USA: Florida
MFL16	Contig	135	Illumina HiSeq	USA: Florida
HHCA16	Contig	97	Illumina HiSeq	USA: Hacienda Heights
A-SBCA19	Contig	81	Illumina NextSeq; HiSeq	USA: California

Legenda: Caracterização inicial dos genomas de CLas disponíveis pelo NCBI.

4.2 Pangenoma de *Candidatus Liberibacter asiaticus*

Para a construção do pangenoma, foi utilizado o *software* Panaroo (Tonkin-Hill et al., 2020), que permite a geração e visualização de gráficos de redes de pangenomas, possibilitando análises detalhadas dos genomas central e acessório. A análise foi conduzida no modo de limpeza moderado, o qual corrige sobreposições e erros comuns de anotação. Foram removidos genes inválidos e os válidos foram agrupados em famílias gênicas utilizando um *threshold* de 95% de identidade. A execução foi realizada no modo de análise correspondente ao pangenoma completo. Para construção da visualização da matriz de presença e ausência foi utilizado o *software* Roary (Page et al., 2015).

4.3 Árvore filogenética de *Candidatus Liberibacter asiaticus*

A partir dos resultados obtidos com o Panaroo, foi possível realizar o alinhamento do *core genes* de *Candidatus Liberibacter asiaticus* utilizando o *software* MAFFT (Vences et al., 2022). Esse arquivo foi utilizado como entrada no *software* SNP-sites (Page et al., 2016), para a extração dos polimorfismos de nucleotídeo único (SNPs) — variações em uma única base do DNA, amplamente empregadas como marcadores moleculares em estudos evolutivos.

Com os SNPs do *core genes*, foi realizado um alinhamento múltiplo utilizando o *software* MUSCLE (Edgar, 2004), seguido da inferência filogenética por máxima verossimilhança no *software* IQ-TREE version 2.4.0 (Minh et al., 2020). A análise foi conduzida com os modos -MFP (Model Finder Plus) e -ASC, este último aplicado quando todas as posições são variáveis. Foram realizadas 1000 replicações de *bootstrap* ultrarrápido (-B 1000) e 1000 replicações do teste *Approximate Likelihood Ratio Test* (alrt 1000), ambos empregados para avaliar o suporte dos ramos. O parâmetro nt AUTO foi utilizado para permitir que o programa determinasse automaticamente o número de *threads*, a fim de otimizar o processamento. Para gerar a imagem foi utilizado o site iTOL (Letunic; Bork, 2024).

4.4 Estrutura populacional

Para a análise da estrutura populacional, foram obtidos os genomas adicionais disponíveis no NCBI, totalizando 148 genomas. O acesso aos *assemblies* foi realizado por meio do *download* do *dataset* correspondente, excluindo-se aqueles pertencentes à base RefSeq, de modo a evitar redundâncias.

O controle de qualidade dos genomas foi realizado com o *software* CheckM (Parks et al., 2015), mantidos apenas os genomas com completude superior a 90%, resultando em 142 genomas para as etapas seguintes. As anotações genômicas desses 142 genomas foram

realizadas com o *software* Bakta, utilizando a opção --meta, a fim de padronizar o processo de anotação entre todos os genomas e permitir a detecção de genes em dados metagenômicos. O CheckM utiliza o Prodigal (Hyatt et al., 2010) para a predição de genes durante a estimativa de completude e contaminação genômica.

4.4.1 Obtenção dos genes ortólogos mais informativos

Para a identificação dos genes ortólogos no primeiro conjunto de dados (genomas do RefSeq), composto por 37 genomas, foi utilizado o *software* ProteinOrtho (Lechner et al., 2011). Esse programa detectou 1230 grupos ortólogos preditos, dos quais foram selecionados 769 presentes em todos os 37 genomas.

As sequências dos genes codificantes dos 769 grupos ortólogos selecionados foram submetidas a um alinhamento múltiplo das sequências de nucleotídeos por meio do *software* MUSCLE, em seguida, processados no SNP-sites (Page et al., 2016) para a identificação dos polimorfismos de nucleotídeo único (SNPs). Posteriormente, foram filtrados os grupos ortólogos que apresentaram seis ou mais SNPs.

Posteriormente, as sequências dos 12 grupos ortólogos com mais SNPs da cepa GCF_030585885 foram usada como *query* em BLASTNs (Chen et al., 2015) para recuperar as sequências correspondentes nos 142 *assemblies* oriundos do Genbank, que foram concatenados e utilizados como banco de dados para a *subject*, com um *evaluate* de $1e-10$. Após os resultados, foram filtrado as *subjects* que estavam duplicadas, deixado apenas as que possuíam maior alinhamento, pelo id das *subjects* e com as sequências dos 12 grupos ortólogos completos foi feito um alinhamento múltiplo pelo MUSCLE (Edgar, 2004) e depois, os alinhamentos foram concatenados pelo FASConCAT-G_v1.06.1.pl (Kück; Longo, 2014). Esse alinhamento foi utilizado para reconstruir a árvore filogenética no IQTREE 2.4.0 (Minh et al., 2020) e visualizado no iTOL (Letunic; Bork, 2024).

4.5 Construção do esquema alélico e Minimal Spanning Tree

Um esquema alélico de genoma inteiro (wgMLST) foi obtido para o conjunto total dos *assemblies* de CLas (179 *assemblies*) por meio do software chewBBACA (Silva et al., 2018), que implementa a possibilidade de detecção das variantes dos loci (alelos) por *gene-by-gene*. O esquema alélico obtido foi formado por 1130 loci e 4532 alelos, dos quais dois loci foram removidos devido à presença de parálogos, posteriormente o conjunto total de loci (wgMLST), foi processado para obtenção de loci presentes em 95% dos *assemblies* (cgMLST); um total de

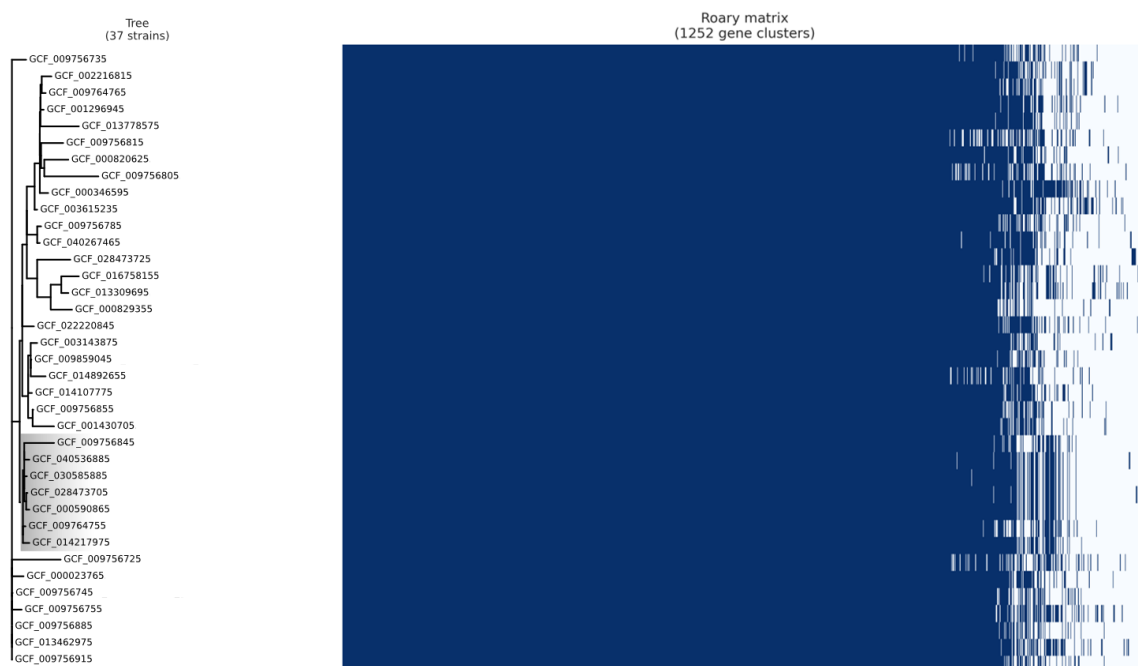
1128 loci foram retidos e utilizados na construção de uma *Minimal Spanning Tree* com o auxílio do PHYLOViZ Online (Ribeiro-Gonçalves et al., 2016).

5. RESULTADOS

As análises iniciais foram realizadas com os 37 genomas presentes no banco de dados RefSeq. As sequências de proteínas preditas desses genomas foram utilizadas para a construção do pangenoma de *Candidatus Liberibacter asiaticus*, o qual consistiu em 1252 famílias gênicas. O *core* genoma, formado pelos genes presentes em todos os 37 *assemblies*, possui 944 genes, o que representa 75,4% do pangenoma. Os genes que fazem parte de mais do que um, mas não de todos os *assemblies* pertencem ao genoma acessório, e representam 24,6% do pangenoma total.

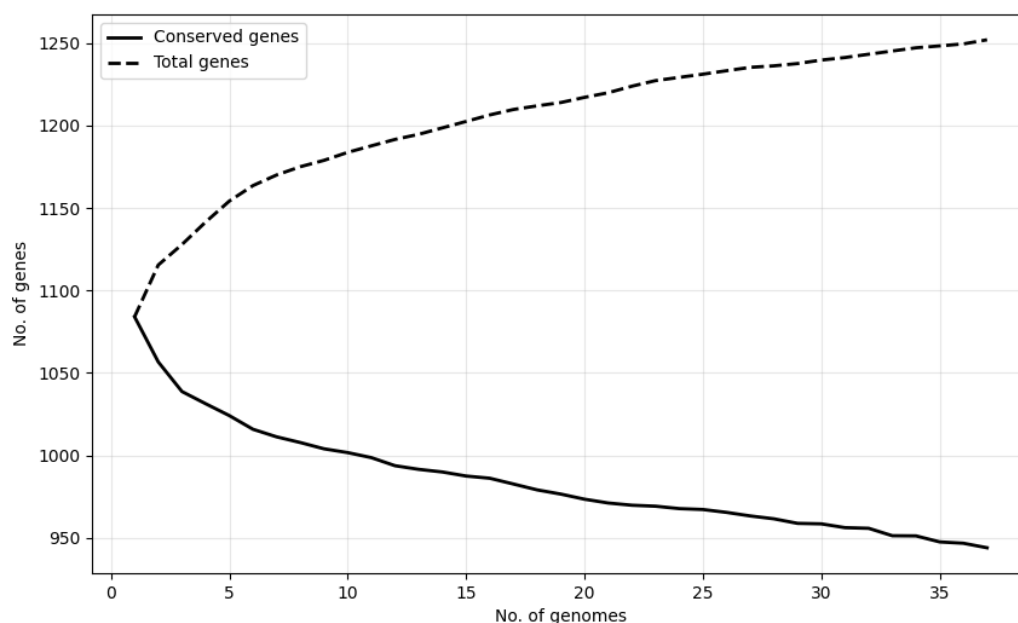
O genoma acessório é subdividido de acordo com o grau de presença dos genes que o constituem no *dataset*, enquanto os *cloud* genes são àqueles presentes em <15% dos *assemblies*, os *shell* genes estão presentes em 15% a 99% do *dataset* e o *softcore* presente em uma grande maioria entre 95% a 99% mas não em todos. A distribuição do genoma acessório no grupo *shell* foi de 125 genes, que representa 10% dos genes acessórios, 115 genes no grupo *cloud*, que corresponde a 9,2% e 68 genes no *softcore*.

Para visualizar o pangenoma e a distribuição dos genes nas cepas de CLas, a presença e ausência das famílias gênicas foi mapeada em uma árvore filogenética produzida por Máxima Verossimilhança pelo IQTREE2 (Figura 4). As famílias gênicas presentes estão representadas na cor azul, enquanto as áreas sem cor representam a ausência de uma família gênica do pangenoma. O clado destacado pelo sombreado cinza, composto por 7 genomas, denota maior homogeneidade nos genes acessórios.

Figura 4. Matriz de presença e ausência do pangenoma

Legenda. Distribuição da presença e ausência de 1252 *cluster* gênicos distribuídos em 37 *assemblies* de *Candidatus Liberibacter asiaticus*, azul indica presença e o branco ausência. Árvore filogenética construída a partir de *core genes*.

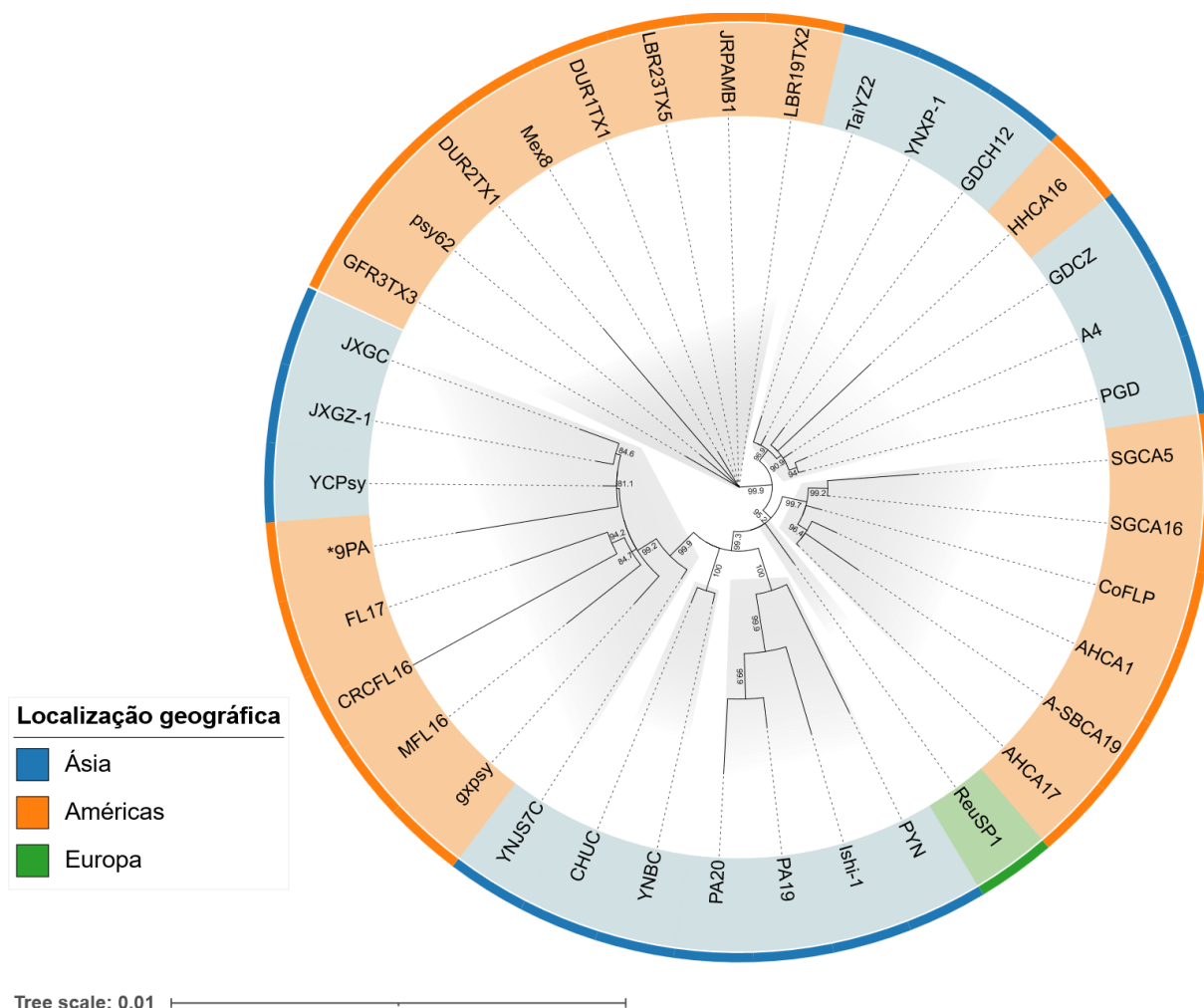
Para visualizar o comportamento do pangenoma e do core genoma durante sua construção, e inferir o grau de adequação do *dataset* utilizado, foram construídas curvas de acumulação do pan- e do core genoma (Figura 5). As curvas revelaram que medida que novos genomas foram adicionados, o número de genes totais (curva tracejada) apresentou um aumento gradual, ao passo que o número de *core genes* (curva sólida) apresentou redução. Porém, uma observação cautelosa da curva tracejada revela que a partir do 15º genoma, até o fim do *dataset*, somente 50 novos genes foram adicionados ao pangenoma; o que indica estabilização da curva e a caracterização de um pangenoma fechado, que provavelmente estabilizaria com um *dataset* maior. Os *shell*, *cloud* e genes únicos ainda serão estudados detalhadamente.

Figura 5. Curva de acumulação do Pangenoma

Legenda. Curva de acumulação do pangenoma e do *core genome* de CLas com base nos 37 *assemblies*. A linha tracejada representa o *core genes*, à medida que adiciona novos genomas, o valor total diminui. A curva da linha contínua representa os genes totais conforme a adição de novos genomas

A árvore filogenética por máxima verossimilhança obtida pelo IQTREE permitiu inferir que os genomas de CLas dividem-se em cinco clusters populacionais (Figura 6). Os ramos iniciais composto por genomas vindo da América não se clusterizaram e a cepa DUR2TX1 possui um destaque no seu ramo indicando uma maior quantidade de eventos de inserção e deleção de genes comparado aos outros. Em seguida, o clado presente majoritariamente por cepas vindas da Ásia com apenas uma cepa vinda da América HHCA 16, mais especificamente vinda dos Estados Unidos, também possui um ramo longo. O próximo clado se destaca com todas as cepas originadas da América, seguido por uma cepa sem nenhum *cluster*, oriunda da Europa, especificamente da França. O clado seguinte possui uma sequência de dois clusters de genomas vindos da Ásia e finalizando com o último clado com cepas vinda dos dois continentes, há um destaque com um asterisco na cepa 9PA, vinda do Brasil, esta e a CRCFL 16 possuem tamanhos de ramos longos comparado aos outros.

Figura 6. Árvore filogenética de CLas dos *assemblies* do RefSeq



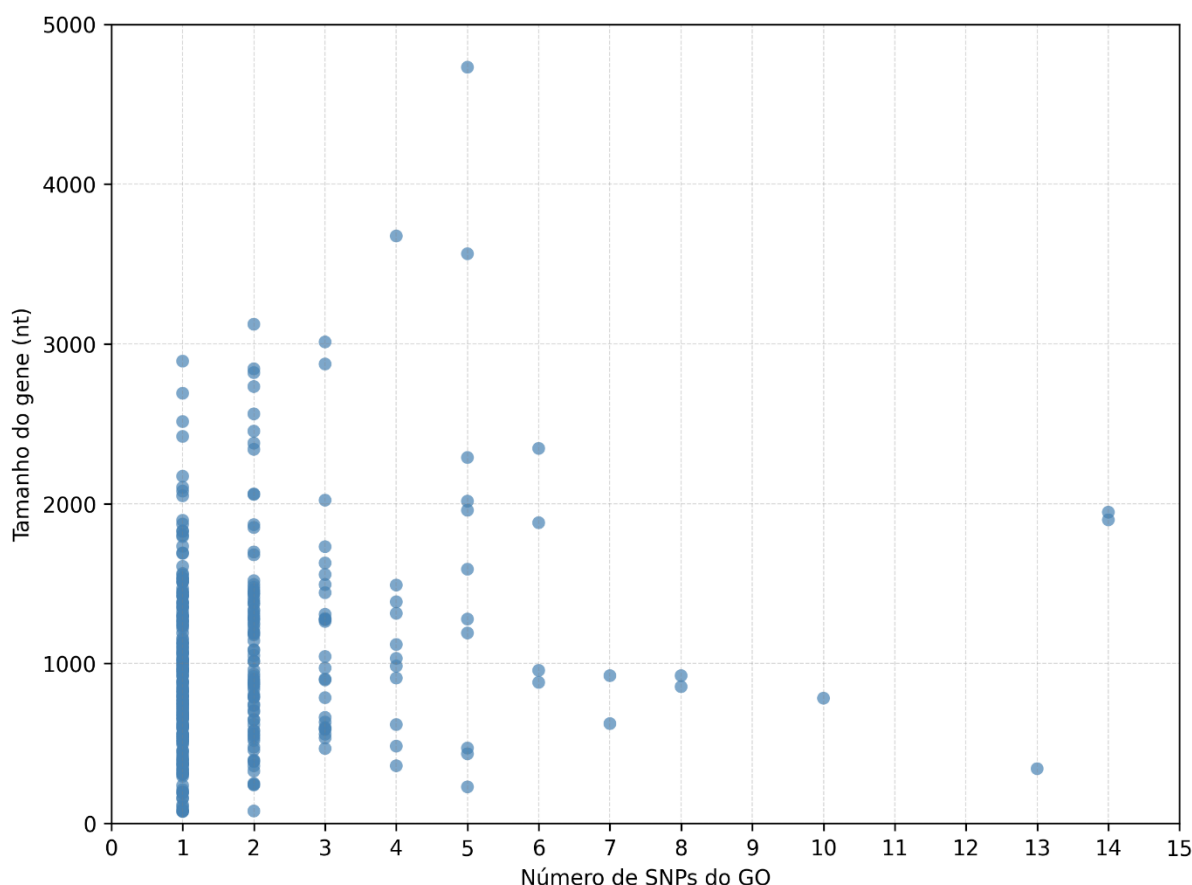
Legenda. Árvore filogenética dos 37 genomas de CLas, as cores indicam a localização geográfica de origem das cepas. A filogenia foi inferida por máxima verossimilhança a partir dos SNPs do *core genome*. Valores internos representam *bootstrap* 1000 replicatas. A cepa *9PA com o destaque em asterisco, representa a única vinda do Brasil.

Os genes ortólogos são àqueles gerados por processos de especiação, por isso são interessantes para estudos filogenéticos, uma vez que a história evolutiva reflete desses genes tende a refletir de forma fidedigna a história evolutiva da própria espécie. Por isso, foi feita a predição dos genes ortólogos para os 37 *assemblies* do RefSeq, os genes dos grupos ortólogos preditos foram alinhados, e o número de SNPs obtidos contabilizados, a fim de obter-se os grupos ortólogos mais variáveis (i.e., com maior número de SNPs).

Na identificação dos genes ortólogos, o arquivo gerado pelo ProteinOrtho apresentou 1230 grupos ortólogos (GO), dos quais foram selecionados apenas aqueles presentes nos 37 genomas, resultando em 769 grupos ortólogos, dos quais foram extraídos os SNPs.

A partir da análise dos SNPs, foram identificados 325 genes (Figura 7) que apresentaram alguma posição variável. Em seguida, realizou-se uma filtragem dos GO com maior número de SNPs, totalizando 12 GO selecionados. Dentre esses, dois GO apresentaram 14 SNPs (o valor máximo observado), um GO apresentou 13 SNPs, um GO apresentou 10 SNPs, dois GO apresentaram 8 SNPs, dois GO apresentaram 7 SNPs, e quatro GO apresentaram 6 SNPs.

Figura 7. SNPs em comparação ao tamanho de cada gene



Legenda. *Scatter plot* da distribuição dos SNPs dos genes ortólogos em função do tamanho dos genes de CLas. Cada ponto azul representa um grupo ortólogo e indica a quantidade de SNPs identificados em função do comprimento do grupo ortólogo em pares de bases (bp).

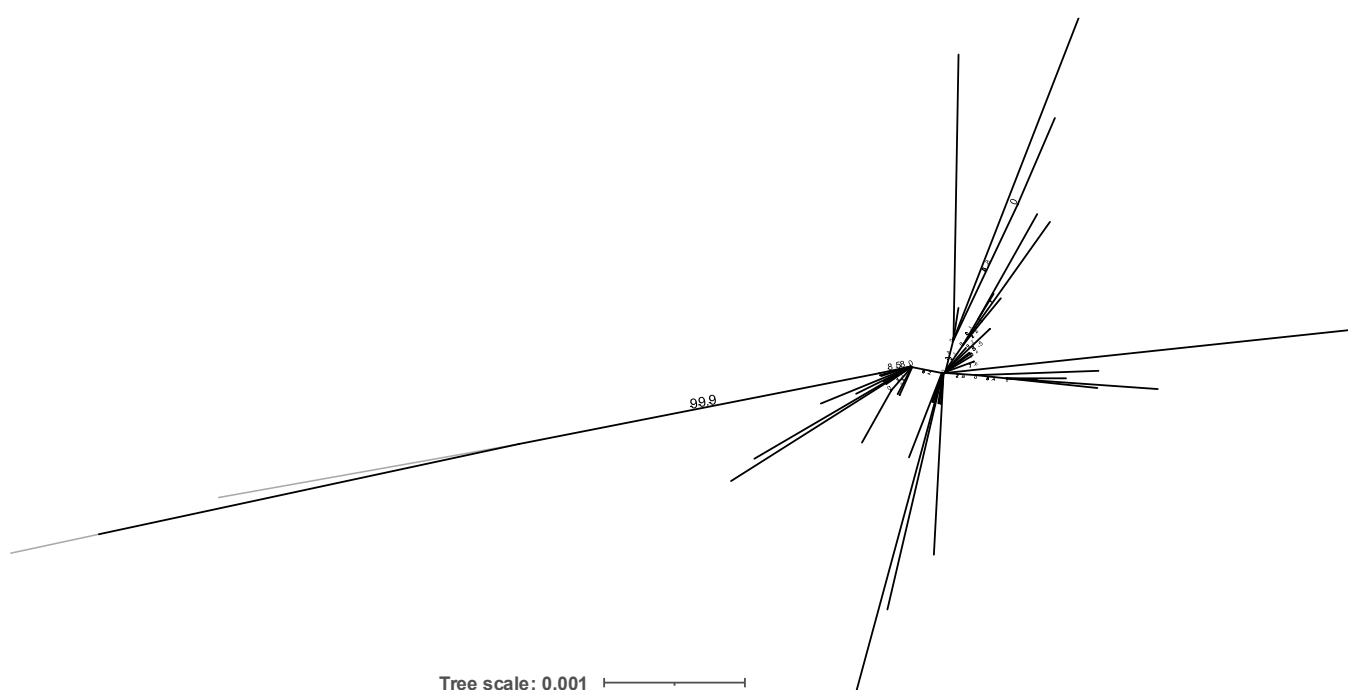
Os grupos ortólogos com 6 ou mais SNPs foram utilizados em BLASTNs para recuperar as sequências correspondentes nos 142 genomas adicionais conforme o item 4.4.1, a fim de obter uma filogenia ampla da espécie a partir de um conjunto restrito de marcadores moleculares informativos.

Através da análise feita, foram encontrados 142 *hits* em sete dos doze arquivos de genes ortólogos, e em apenas um arquivo 141, 143 e 144 *hits*, todos com mais de 98% de identidade, os *hits* duplicados foram filtrados pelo tamanho do alinhamento. Esse resultado então foi alinhado com as sequências correspondentes dos 12 grupos ortólogos e concatenados, para

assim ser possível reconstruir a árvore filogenética com um dataset ampliado para um total de 179 *assemblies*.

A árvore filogenética inferida pelo método de máxima verossimilhança no IQ-TREE (Figura 8) apresentou baixa resolução, com valores de *bootstrap* próximos ou iguais a zero para a maioria dos ramos, o que impossibilitou uma interpretação confiável dos agrupamentos. Esses resultados indicam que o conjunto de SNPs utilizado não foi suficiente para sustentar uma estrutura filogenética consistente, possivelmente devido à baixa variabilidade genética entre os genomas analisados.

Figura 8. Árvore filogenética de CLas dos *assemblies* do GenBank



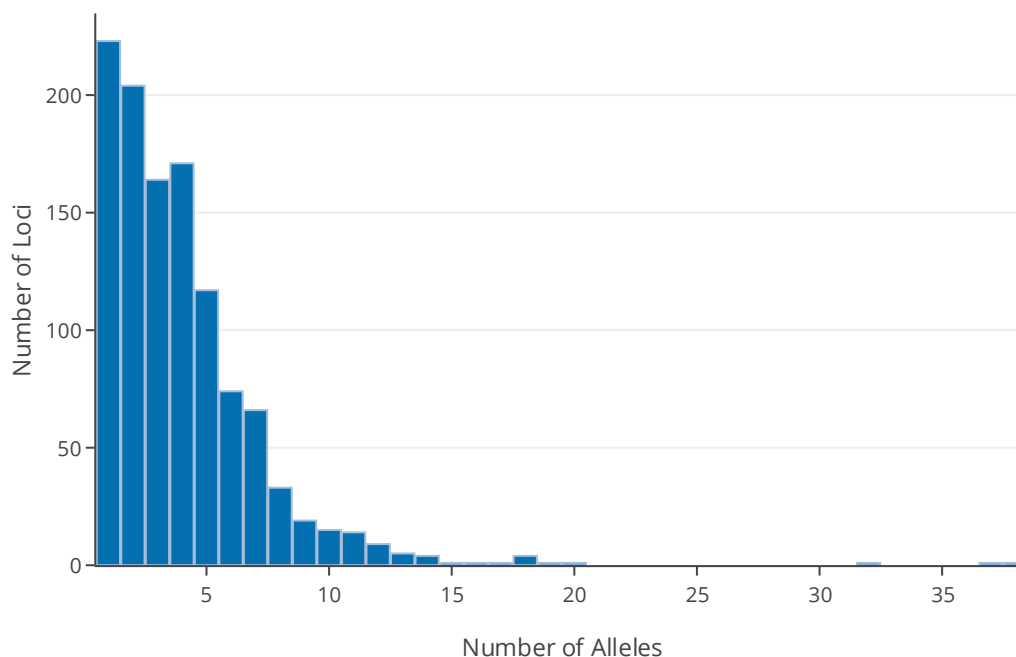
Legenda. Árvore filogenética dos 179 genomas de CLas. A filogenia foi inferida por máxima verossimilhança.

Assim, optou-se pela adoção de uma abordagem alternativa baseada em esquemas alélicos realizada pelo chewBBACA, esse programa permite avaliar a diversidade genômica por meio da presença e variação de alelos em múltiplos loci, fornecendo maior poder discriminatório para a análise populacional, principalmente em espécies ou linhagens com pouca variabilidade genética.

A análise inicial determinou os perfis alélicos dos 179 genomas selecionados realizada pelo método *gene-by-gene*, um alelo é uma variante específica de sequência que ocorre em um determinado locus. Foram encontrados 1130 loci com 4532 alelos, a ampla maioria dos loci

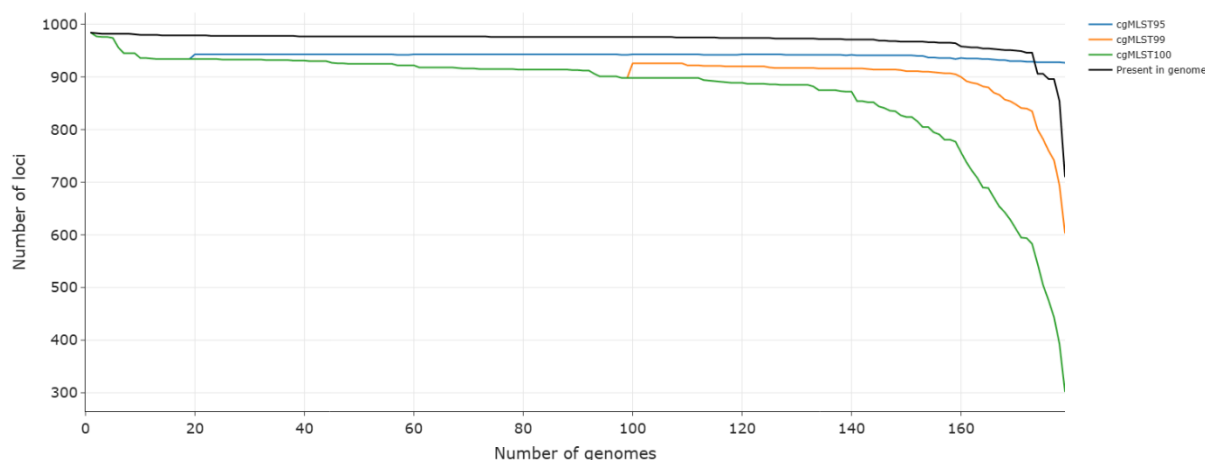
possuem < 10 alelos, na verdade 223 loci possuem apenas um alelo, ou seja, apresentam sequências idênticas (Figura 9). Na média, há 4,01 alelos por loci, esses valores reforçam o baixo polimorfismo em CLas.

Figura 9. Distribuição do número de alelos por loci.



Legenda. Gráfico em barras do número de loci (eixo y) pela quantidade de alelos (eixo x) entre os 179 *assemblies* de CLas

Após a obtenção do esquema alélico, dois loci parálogos foram removidos, resultando em 1128 loci. Esses loci foram subsequentemente filtrados para a obtenção dos loci que constituem o *core genome multilocus sequence typing* (cgMLST) (Figura 10), esse esquema é definido como o conjunto de loci presentes em todas as cepas analisadas ou em $>95\%$ das cepas, a depender das limitações de sequenciamento e montagem. O resultado do gráfico do esquema de cgMLST apresentou uma curva mais estável para o cgMLST95 que possui um *threshold* um pouco menor em comparação aos de cgMLST100 e cgMLST99.

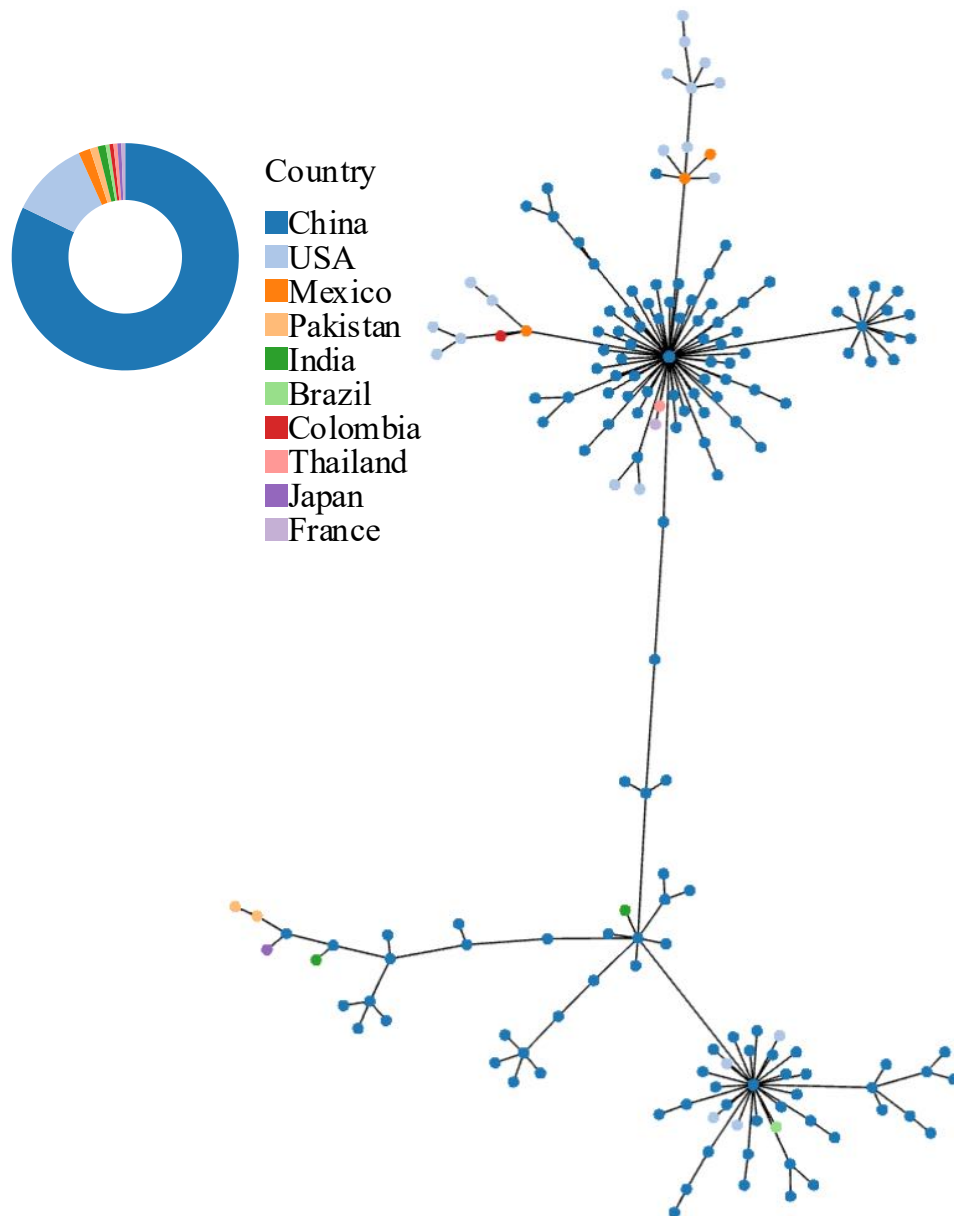
Figura 10. Esquemas de cgMLST

Legenda. Curvas dos esquemas cgMLST para os 179 *assemblies* de CLas. O eixo X mostra o número de genomas incluídos na análise e o eixo Y indica o número de loci compartilhados. As curvas representam diferentes *thresholds* de inclusão de loci no esquema: cgMLST95 (presente em $\geq 95\%$ dos genomas), cgMLST99 ($\geq 99\%$) e cgMLST100 (100%). A linha preta indica o número total de loci presentes em cada genoma individual.

Estabelecido o *threshold* de 95% de presença no cgMLST, foi possível calcular uma *Minimum Spanning Tree* (MST) (Figura 11), cada nó representa um isolado de localidades diferentes, a maioria oriunda da China, seguida pelo Estados Unidos da América. A árvore apresenta dois principais *clusters*, um à esquerda e um à direita; distantes entre si. Essa distância, indica que as cepas da China, predominantes no dataset, não são homogêneas quanto ao seu cgMLST. No cluster acima, observa-se a presença de subclusters curtos e maior diversidade geográfica. Também nota-se a emergência de outro grupo clonal derivado, formado exclusivamente por cepas chinesas. Já o cluster abaixo é menos homogêneo, os genomas estão distantes, denotando maior variabilidade nos esquemas alélicos. Os genomas situados entre os dois clusters são possíveis ancestrais e podem ser utilizados para enraizar árvores filogenéticas intra-espécie de CLas.

Além disso, os isolados da Tailândia e França, Colômbia e os do México também se agruparam no cluster acima da MST com alguns dos EUA, isso indica uma maior similaridade entre essas cepas vindas das Américas. Ao passo que no *cluster* abaixo além dos clones da China e dos EUA, está presente também a única cepa do Brasil e cepas de outros países do oriente, como Paquistão, Japão e Índia.

Figura 11. Minimal Spanning Tree



Legenda. Minimal Spanning Tree baseada no cgMLST 95, mostra as relações genéticas entre as cepas analisadas. Cada nó representa uma cepa e as cores indicam a localização geográfica da origem do genoma.

6. DISCUSSÃO

Existem poucos táxons fitopatogênicos que são intracelulares obrigatórios fora do clado *Candidatus Liberibacter*, o mais conhecido são os fitoplasmas, bactérias pertencentes ao gênero *Candidatus Phytoplasma*, filo *Mycoplasmata* (antigo *Tenericutes*). Ao contrário de *Candidatus Liberibacter*, fitoplasmas infectam uma ampla gama de hospedeiros vegetais, incluindo culturas alimentares, frutíferas, oleaginosas, ornamentais e até plantas daninhas; as infecções também são mediadas por insetos, porém os fitoplasmas possuem múltiplos vetores conhecidos, como cicalídeos (ordem Hemiptera, família Cicadellidae ou *leafhoppers*), fulgorídeos (ordem Hemiptera, família Fulgoroidea ou *planthoppers*) e psílídeos (Wang et al., 2024).

Além dos fitopatógenos, diversas bactérias que infectam animais também exibem um estilo de vida intracelular obrigatório, resultado intensa coevolução com seus hospedeiros. Entre os exemplos mais estudados estão espécies dos gêneros *Rickettsia*, *Anaplasma* e *Ehrlichia*, que parasitam células sanguíneas e endoteliais, e as *Chlamydia*, cujo ciclo de vida envolve estágios alternados entre formas infecciosas e replicativas. Entretanto, há patógenos intracelulares obrigatórios cultiváveis, ainda que sejam fastidiosos, como *Leptospira* e *Borrelia* (Coers; Newton; Carlyon, 2021). Assim, ainda que existam táxons intracelulares obrigatórios não cultiváveis, as duas características não são obrigatoriamente concorrentes.

O estilo de vida intracelular obrigatório parece decorrer de processos evolutivos de redução de genoma, nos quais genes não essenciais são perdidos em favor de maior adaptação de nicho. Georgiades e Raoult (2011) notaram, em um influente trabalho, que 11 de 12 dos principais patógenos bacterianos humanos possuíam genomas significativamente menores que táxons filogeneticamente próximos. Se considerarmos o genoma cromossômico de *Escherichia coli*, com 4,5 Mb, como o genoma bacteriano padrão, o tamanho do genoma das bactérias intracelulares obrigatórias chama ainda mais a atenção devido ao tamanho reduzido: fitoplasmas (~0,5 a 0,96 Mb), *Borrelia burgdorferi* (~0,91 Mb) e *Candidatus Liberibacter asiaticus* (~1,23 Mb). Cabe ressaltar que a redução genômica é um processo evolutivo que ocorre em outras especializações ecológicas que não a patogênese, como por exemplo em bactérias simbióticas (McCutcheon; Moran, 2012).

A compreensão das forças evolutivas por trás do desenvolvimento da patogênese é importante para o desenvolvimento de estratégias que permitam mitigar os danos causados pelos patógenos. No caso de CLas, a análise do pangenoma de *Candidatus Liberibacter asiaticus* mostrou que 75,4% do genoma corresponde aos *core genes*, enquanto 24,6% dos genes

compõem o genoma acessório, indicando baixa variabilidade genética entre as cepas analisadas. Essa homogeneidade é esperada em função do estilo de vida intracelular obrigatório da bactéria, que habita exclusivamente o seu vetor, *Diaphorina citri*, ou os citros hospedeiros, reduzindo a necessidade de um grande repertório de genes acessórios.

O trabalho mais amplo de genômica de CLas, obteve 135 novos MAGs de diferentes regiões do sul da China (Zheng et al., 2024), no que tange aos genes preditos, os genomas tiveram de 1,104 a 1,211 genes, o que indicou baixa variabilidade genética nesta espécie. Neste trabalho, confirmamos essa hipótese por meio da construção do pangenoma que revelou, pela primeira vez, a magnitude da conservação gênica desta espécie, com cerca de 75% do pangenoma conservado em isolados dispersos ao redor do mundo, e transmissíveis por diferentes espécies de psílídeos. A curva de acumulação do pangenoma se mostrou fracamente ascendente (linha pontilhada – Figura 5), com cerca de 4 novos genes adicionados a cada novo genoma, o que indica que ainda há diversidade genética a ser descoberta, motivando-nos a repetir o pangenoma para o dataset expandido para todos os 179 *assemblies* (incluindo os oriundos do GenBank).

A reconstrução de uma árvore filogenética por máxima verossimilhança a partir do alinhamento do core genoma, apresentou topologia condizente com a origem geográfica em alguns ramos, porém, enquanto Zheng et al (2024), reportou a presença de dois clusters de CLas intra-China, nossos dados revelaram a presença de outros grupos, incluindo um clado formado majoritariamente por genomas dos EUA, com uma cepa da Colômbia (com os *assemblies* SGCA5, SGCA16, CoFLP, AHCA1, A-SBCA19, AHCA17). Além disso, a única cepa proveniente da Europa não se agrupou a nenhum dos cladros, sugerindo que as cepas desse continente apresentam maior distinção genética em relação às demais, hipótese que carece de dados para confirmação. A ausência de uma clara delimitação geográfica entre os cladros pode estar relacionada a fatores como o intenso comércio internacional de mudas e frutos cítricos, que favorece o fluxo gênico entre populações. Além disso, nas plantações comerciais de citros, a reprodução vegetativa (clonagem) é amplamente utilizada, o que reduz a variabilidade genética dos hospedeiros e pode contribuir para a dispersão de cepas geneticamente semelhantes em diferentes regiões.

A quantidade de SNPs nos grupos ortólogos obtidas neste trabalho foi de ordens de magnitude menor do que a obtida por Zheng et al. (2024). Enquanto obtivemos 325 SNPs, no referido trabalho 5090 SNPs foram obtidos. Dois motivos explicam essa discrepância, o dataset e a metodologia de SNP calling; enquanto aqui foram analisados 37 *assemblies* do RefSeq e o SNP calling foi realizado por alinhamento múltiplo e posterior detecção das posições variáveis,

pelo *software* SNP-sites, Zheng et al. (2024) realizou a detecção utilizando múltiplos MAGs de baixa qualidade e por meio de alinhamento ao cromossomo da cepa A4 e aos genomas dos profagos P-YNBC-1, P-A4-2 e P-JXGC-3, por meio do *software* snippy.

A tentativa aqui realizada de obtenção grupos ortólogos com maior quantidade de SNPs para construção de uma filogenia robusta com todos os 179 *assemblies* não produziu uma árvore filogenética confiável. Por isso, almejamos a construção de um esquema alélico, realizado a partir do *software* chewBBACA, uma ferramenta que cria e avalia esquemas, além de resultados de MLST, de genoma central e completo (cg/wgMLST). Tradicionalmente, o MLST *schema* é construído a partir de um número reduzido de genes “housekeeping” (geralmente sete), enquanto o cgMLST expande essa abordagem para milhares de loci, oferecendo maior resolução para estudos populacionais e filogenéticos ao custo da confiabilidade dos loci. Diferente de métodos baseados apenas em SNPs, a abordagem *gene-by-gene* avalia variações em cada gene homólogo (locus), gerando perfis alélicos que podem ser comparados entre diferentes cepas. Assim, ao empregar o chewBBACA, foi possível definir um conjunto consistente de loci compartilhados entre as amostras, permitindo a construção de um esquema cgMLST representativo e, conseqüentemente, a inferência de relações genéticas mais confiáveis entre os 179 genomas analisados.

O esquema alélico demonstrou novamente a baixa variabilidade genética da espécie, com 1130 loci e 4532 alelos, uma média de variação de 4 alelos por loci. Além disso a *Minimum Spanning Tree* ressaltou o intenso comércio das plantas, uma vez que apresentou diversos clones da China, relacionado com o dos EUA. Ainda assim houve uma delimitação em relação as cepas vindas da Colômbia e do México, que se relacionaram em um dos ramos com os isolados dos EUA. MSTs construídas a partir do cgMLST foram utilizadas, de forma direta, para determinar patovares, subsespécies e patotipos em *Xanthomonas citri* (Ragupathy et al., 2023). Apesar de a inspeção visual da MST sugerir a presença de três clusters populacionais principais, esta interpretação ainda é preliminar. A identificação de estruturas populacionais em esquemas cgMLST exige métodos de clusterização específicos, capazes de considerar diferenças alélicas de forma estatística e reprodutível. Diversos algoritmos têm sido empregados para este fim, como goeBURST, hierarchical clustering, MSTreeV2 e pHierCC, cada um incorporando regras distintas para definição de grupos a partir das distâncias alélicas. Esses métodos estão reunidos na ferramenta ReporTree (Mixão et al., 2023), que futuramente será utilizada para realizar uma detecção robusta e quantitativa dos clusters presentes nos genomas de *Candidatus Liberibacter asiaticus*.

7. CONCLUSÃO

Os resultados obtidos indicam que *Candidatus Liberibacter asiaticus* apresenta baixa variabilidade genômica, conforme evidenciado pelas análises de pangenoma e do esquema alélico, caracterizando um pangenoma fechado e uma alta conservação gênica, refletindo a natureza clonal e da espécie entre as cepas analisadas. Aspectos relevantes como o genoma acessório, as funções dos genes com maior quantidade de SNPs e a inferência de clusters populacionais por algoritmos adequados ainda serão estudados. Dessa forma, os resultados obtidos ampliam o conhecimento sobre a estrutura genômica e a conservação genética de *Candidatus Liberibacter asiaticus*, contribuindo para a compreensão de sua biologia, evolução e impacto fitopatogênico.

REFERÊNCIAS

ACHOR, Diann *et al.* Dynamics of *Candidatus Liberibacter asiaticus* Movement and Sieve-Pore Plugging in Citrus Sink Cells. **Plant Physiology**, v. 182, n. 2, p. 882–891, fev. 2020.

ANDRADE, Maxuel; LI, Jinyun; WANG, Nian. *Candidatus Liberibacter asiaticus*: virulence traits and control strategies. **Tropical Plant Pathology**, v. 45, n. 3, p. 285–297, 1 jun. 2020.

BARBIERI, Hellen Bertoletti *et al.* An overview of the most threatening diseases that affect worldwide citriculture: Main features, diagnose, and current control strategies. **Frontiers in Natural Products**, v. 2, 24 jan. 2023.

BARROS, Geraldo Sant’Ana de Camargo *et al.* **PIB do Agronegócio Brasileiro**. Desenvolvimento de Sites. Disponível em: <<https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx>>. Acesso em: 1 out. 2024.

BEYVERS, Sebastian *et al.* Bakta Web – rapid and standardized genome annotation on scalable infrastructures. **Nucleic Acids Research**, v. 53, n. W1, p. W51–W56, 7 jul. 2025.

BRASIL. Ministério da Agricultura e Pecuária. Secretaria de Defesa Agropecuária. Portaria nº 1.326, de 4 de julho de 2025. Institui o Programa Nacional de Prevenção e Controle à doença denominada Huanglongbing- PNCHLB e os critérios e procedimentos para classificação e manutenção do status fitossanitário das Unidades Federativas e as medidas de prevenção e controle da doença, no território nacional. *Diário Oficial da União: seção 1*, Brasília, DF, n. 128, p. 14–17, 7 jul. 2025. Disponível em: <https://www.in.gov.br/web/dou/-/portaria-sda/mapa-n-1.326-de-4-de-julho-de-2025-640269613>. Acesso em: 1 nov. 2025.

CARMO-SOUSA, Michele; CORTÉS, Mayerli Tatiana Borbón; LOPES, João Roberto Spotti. Understanding psyllid transmission of *Candidatus Liberibacter* as a basis for managing huanglongbing. **Tropical Plant Pathology**, v. 45, n. 6, p. 572–585, 1 dez. 2020.

CHEN, J. *et al.* Guangdong and Florida Populations of ‘*Candidatus Liberibacter asiaticus*’ Distinguished by a Genomic Locus with Short Tandem Repeats. **Phytopathology®**, v. 100, n. 6, p. 567–572, jun. 2010.

CHEN, Ying *et al.* High speed BLASTN: an accelerated MegaBLAST search tool. **Nucleic Acids Research**, v. 43, n. 16, p. 7762–7768, 18 set. 2015.

COERS, Jörn; NEWTON, Hayley J.; CARLYON, Jason A. Can’t live outside you: a thematic issue on obligate intracellular bacterial pathogens. **Pathogens and Disease**, v. 79, n. 9, p. ftab054, 9 dez. 2021.

COLETTA-FILHO, H. D. *et al.* First Report of the Causal Agent of Huanglongbing (“*Candidatus Liberibacter asiaticus*”) in Brazil. **Plant Disease**, v. 88, n. 12, p. 1382–1382, dez. 2004.

COSTA LIMA, A. M. **Insetos Do Brasil. Tomo 03. Homópteros**. Rio de Janeiro: Escola Nacional de Agronomia, 1942.

DE OLIVEIRA, Roberto Pedroso *et al.* Tecnologias para Prevenção e Manejo do Huanglongbing (HLB) em Polos de Citricultura de Base Familiar. Embrapa Clima Temperado. Documentos, 281. n. Embrapa Clima Temperado, p. 28, 2018.

DUTRA, Michele FS *et al.* The complete genome sequence of “*Candidatus Liberibacter asiaticus*” strain 9PA and the characterization of field strains in the Brazilian citriculture. **mSphere**, v. 9, n. 12, p. e00376-24, 11 nov. 2024.

EDGAR, Robert C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 1 mar. 2004.

EW, Reinaldo Alnei. **Citricultura em alerta: a “Greening” já está na Fronteira - Jornal Ibiá**. Disponível em: <<https://jornalibia.com.br/destaques/citricultura-em-alerta-a-greening-ja-esta-na-fronteira/>>. Acesso em: 26 out. 2025.

FILLOUX, Alain. Bacterial protein secretion systems: Game of types. **Microbiology**, v. 168, n. 5, p. 001193, 2022.

GAO, Fanglan *et al.* Genetic Diversity of “*Candidatus Liberibacter asiaticus*” Based on Four Hypervariable Genomic Regions in China. **Microbiology Spectrum**, v. 10, n. 6, p. e02622-22, 21 nov. 2022.

GEORGIADIS, Kalliopi; RAOULT, Didier. Genomes of the Most Dangerous Epidemic Bacteria Have a Virulence Repertoire Characterized by Fewer Genes but More Toxin-Antitoxin Modules. **PLOS ONE**, v. 6, n. 3, p. e17962, 18 mar. 2011.

GERLIN, Léo *et al.* Genome-Scale Investigation of the Metabolic Determinants Generating Bacterial Fastidious Growth. **mSystems**, v. 5, n. 2, p. 10.1128/msystems.00698-19, 31 mar. 2020.

GHANIM, Murad *et al.* ‘*Candidatus Liberibacter asiaticus*’ Accumulates inside Endoplasmic Reticulum Associated Vacuoles in the Gut Cells of *Diaphorina citri*. **Scientific Reports**, v. 7, n. 1, p. 16945, 5 dez. 2017.

GOLDFARB, Tamara *et al.* NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. **Nucleic Acids Research**, v. 53, n. D1, p. D243–D257, 6 jan. 2025.

GOSS, Erica M. Genome-enabled analysis of plant-pathogen migration. **Annual Review of Phytopathology**, v. 53, p. 121–135, 2015.

GRAHAM, James H. *et al.* Management of Huanglongbing of Citrus: Lessons from São Paulo and Florida. **Annual Review of Phytopathology**, v. 62, n. Volume 62, 2024, p. 243–262, 9 set. 2024.

HEALTH (PLH), EFSA Panel on Plant *et al.* Pest categorisation of *Diaphorina citri*. **EFSA Journal**, v. 19, n. 1, p. e06357, 2021.

HYATT, Doug *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, v. 11, n. 1, p. 119, 8 mar. 2010.

KATOH, Hiroshi *et al.* Differentiation of “*Candidatus Liberibacter asiaticus*” Isolates by Variable-Number Tandem-Repeat Analysis. **Applied and Environmental Microbiology**, v. 77, n. 5, p. 1910–1917, mar. 2011.

KIM, Chankyung; PONGPANICH, Monnat; PORNTAVEETUS, Thantrira. Unraveling metagenomics through long-read sequencing: a comprehensive review. **Journal of Translational Medicine**, v. 22, n. 1, p. 111, 28 jan. 2024.

KÜCK, Patrick; LONGO, Gary C. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. **Frontiers in Zoology**, v. 11, n. 1, p. 81, 18 nov. 2014.

LABBÉ, Frédéric *et al.* Genomic footprint of a shared Type 5 prophage in “*Candidatus Liberibacter asiaticus*” and “*Candidatus Liberibacter africanus*,” two destructive bacterial pathogens of citrus Huanglongbing. **Applied and Environmental Microbiology**, v. 0, n. 0, p. e01072-25, 15 out. 2025.

LECHNER, Marcus *et al.* Proteinortho: Detection of (Co-)orthologs in large-scale analysis. **BMC Bioinformatics**, v. 12, n. 1, p. 124, 28 abr. 2011.

LESTIYANI, Ayu *et al.* Natural Infection of *Murraya paniculata* and *Murraya sumatrana* with ‘*Candidatus Liberibacter asiaticus*’ in Java. **Plant Disease**, v. 108, n. 9, p. 2760–2770, set. 2024.

LETUNIC, Ivica; BORK, Peer. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. **Nucleic Acids Research**, v. 52, n. W1, p. W78–W82, 5 jul. 2024.

LI, Jinyun *et al.* The in Planta Effective Concentration of Oxytetracycline Against ‘*Candidatus Liberibacter asiaticus*’ for Suppression of Citrus Huanglongbing. **Phytopathology®**, v. 109, n. 12, p. 2046–2054, dez. 2019.

LIU, Yujiang *et al.* Modeling the Transmission Dynamics and Optimal Control Strategy for Huanglongbing. **Mathematics**, v. 12, n. 17, p. 2648, jan. 2024.

LOPES, S. A. *et al.* Graft transmission efficiencies and multiplication of “*Candidatus Liberibacter americanus*” and “*ca. Liberibacter asiaticus*” in citrus plants. **Phytopathology**, v. 99, n. 3, p. 301–306, mar. 2009.

MA, W. *et al.* Population Structures of ‘*Candidatus Liberibacter asiaticus*’ in Southern China. **Phytopathology®**, v. 104, n. 2, p. 158–162, fev. 2014.

MCCUTCHEON, John P.; MORAN, Nancy A. Extreme genome reduction in symbiotic bacteria. **Nature Reviews Microbiology**, v. 10, n. 1, p. 13–26, jan. 2012.

MINH, Bui Quang *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. **Molecular Biology and Evolution**, v. 37, n. 5, p. 1530–1534, 1 maio 2020.

MIXÃO, Verónica *et al.* ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data. **Genome Medicine**, v. 15, n. 1, p. 43, 15 jun. 2023.

NEHELA, Yasser; KILLINY, Nabil. Revisiting the Complex Pathosystem of Huanglongbing: Deciphering the Role of Citrus Metabolites in Symptom Development. **Metabolites**, v. 10, n. 10, p. 409, out. 2020.

NIZAMANI, Mir Muhammad *et al.* High-throughput sequencing in plant disease management: a comprehensive review of benefits, challenges, and future perspectives. **Phytopathology Research**, v. 5, n. 1, p. 44, 12 set. 2023.

NORONHA, Aloysia Cristina da Silva *et al.* Occurrence and distribution of *Diaphorina citri* (Hemiptera: Liviidae) and *Tamarixia radiata* (Hymenoptera: Eulophidae) in Pará state, Brazil. **Arquivos do Instituto Biológico**, v. 86, p. e0302018, 2019.

OKOH, Efenai B. *et al.* A Multilocus Sequence Typing Scheme for Rapid Identification of *Xanthomonas citri* Based on Whole-Genome Sequencing Data. **Phytopathology®**, 11 jul. 2024.

PAGE, Andrew J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. **Bioinformatics**, v. 31, n. 22, p. 3691–3693, 15 nov. 2015.

PAGE, Andrew J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. **Microbial Genomics**, v. 2, n. 4, p. e000056, 2016.

PARCEY, Michael *et al.* Comparative genomic analysis of *Erwinia amylovora* reveals novel insights in phylogenetic arrangement, plasmid diversity, and streptomycin resistance. **Genomics**, v. 112, n. 5, p. 3762–3772, 1 set. 2020.

PARKS, Donovan H. *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **Genome Research**, v. 25, n. 7, p. 1043–1055, 7 jan. 2015.

PIERRY, Paulo Marques *et al.* Genetic Diversity of *Xylella fastidiosa* Plasmids Assessed by Comparative Genomics. **Tropical Plant Pathology**, v. 45, n. 3, p. 342–360, jun. 2020.

RAGUPATHY, R. *et al.* Core-Genome Multilocus Sequence Typing for Epidemiological and Evolutionary Analyses of Phytopathogenic *Xanthomonas citri*. **Applied and Environmental Microbiology**, v. 89, n. 5, p. e02101-22, 17 abr. 2023.

REZZONICO, Fabio *et al.* Burning questions for fire blight research: I. Genomics and evolution of *Erwinia amylovora* and analyses of host-pathogen interactions. **Journal of Plant Pathology**, v. 106, n. 3, p. 797–810, 1 ago. 2024.

RIBEIRO-GONÇALVES, Bruno *et al.* PHYLOViZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. **Nucleic Acids Research**, v. 44, n. W1, p. W246–W251, 8 jul. 2016.

ROLDÁN, Erik L.; STELINSKI, Lukasz L.; PELZ-STELINSKI, Kirsten S. Evaluation of tree-injected oxytetracycline and antisense oligonucleotides targeting *Candidatus Liberibacter asiaticus* in citrus. **Pest Management Science**, v. 81, n. 3, p. 1487–1500, 2025.

SÃO PAULO. Secretaria de Agricultura e do Abastecimento. *Huanglongbing (HLB), greening dos citros*. Informativo Defesa AgroSP n.º 027, out. 2023. São Paulo: Secretaria de Agricultura e do Abastecimento do Estado de São Paulo, 31 out. 2023. Disponível em: <https://www.defesa.agricultura.sp.gov.br/informativo/defesa-agrosp-no-027-outubro2023/huanglongbing-hlb-greening-dos-citros/>. Acesso em: 09 nov. 2025.

SARKAR, Poulami; JASSAR, Ola; GHANIM, Murad. The plant pathogenic bacterium *Candidatus Liberibacter solanacearum* induces calcium-regulated autophagy in midgut cells of its insect vector *Bactericera trigonica*. **Microbiology Spectrum**, v. 11, n. 5, p. e01301-23, 28 set. 2023.

SETUBAL, João C. Metagenome-assembled genomes: concepts, analogies, and challenges. **Biophysical Reviews**, v. 13, n. 6, p. 905–909, 1 dez. 2021.

SILVA, Mickael *et al.* chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. **Microbial Genomics**, v. 4, n. 3, p. e000166, 2018.

SIMPSON, A. J. G. *et al.* The genome sequence of the plant pathogen *Xylella fastidiosa*. **Nature**, v. 406, n. 6792, p. 151–157, 13 jul. 2000.

TAN, Yongjun *et al.* Comparative Phylogenomic Analysis Reveals Evolutionary Genomic Changes and Novel Toxin Families in Endophytic *Liberibacter* Pathogens. **Microbiology Spectrum**, v. 9, n. 2, p. e00509-21, 15 set. 2021.

THAPA, Shree P. *et al.* Genome-wide analyses of *Liberibacter* species provides insights into evolution, phylogenetic relationships, and virulence factors. **Molecular Plant Pathology**, v. 21, n. 5, p. 716–731, 2020a.

THAPA, Shree P. *et al.* Genome-wide analyses of *Liberibacter* species provides insights into evolution, phylogenetic relationships, and virulence factors. **Molecular Plant Pathology**, v. 21, n. 5, p. 716–731, maio 2020b.

TONKIN-HILL, Gerry *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. **Genome Biology**, v. 21, n. 1, p. 180, 22 jul. 2020.

VENCES, Miguel *et al.* Concatenator, a user-friendly program to concatenate DNA sequences, implementing graphical user interfaces for MAFFT and FastTree. **Bioinformatics Advances**, v. 2, n. 1, p. vbac050, 1 jan. 2022.

WANG, Ruotong *et al.* Phytoplasma: A plant pathogen that cannot be ignored in agricultural production—Research progress and outlook. **Molecular Plant Pathology**, v. 25, n. 2, p. e13437, 23 fev. 2024.

WANG, Xuefeng *et al.* Molecular characterization of a mosaic locus in the genome of “Candidatus Liberibacter asiaticus”. **BMC Microbiology**, v. 12, n. 1, p. 18, 26 jan. 2012.

WEBSTER, John; BOGEMA, Daniel; CHAPMAN, Toni A. Comparative Genomics of *Xanthomonas citri* pv. *citri* A* Pathotype Reveals Three Distinct Clades with Varying Plasmid Distribution. **Microorganisms**, v. 8, n. 12, p. 1947, dez. 2020.

WU, Fengnian *et al.* Tracking the geographical distribution of the Asian citrus psyllid *Diaphorina citri* throughout China using mitogenomes and endosymbionts. **Journal of Pest Science**, v. 98, n. 2, p. 1173–1185, 1 mar. 2025.

YOU, Ping *et al.* Potential habitat suitability of *Candidatus Liberibacter asiaticus* and genetic diversity of its prophages across China. **Microbiology Spectrum**, v. 12, n. 11, p. e00633-24, 24 set. 2024.

ZHANG, Shujian *et al.* “Ca. *Liberibacter asiaticus*” carries an excision plasmid prophage and a chromosomally integrated prophage that becomes lytic in plant infections. **Molecular plant-microbe interactions: MPMI**, v. 24, n. 4, p. 458–468, abr. 2011.

ZHENG, Yongqin *et al.* Whole genome sequences of 135 “*Candidatus Liberibacter asiaticus*” strains from China. **Scientific Data**, v. 11, n. 1, p. 1018, 19 set. 2024.

ZHENG, Zheng *et al.* A Type 3 Prophage of “*Candidatus Liberibacter asiaticus*” Carrying a Restriction-Modification System. **Phytopathology**, v. 108, n. 4, p. 454–461, abr. 2018.