UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

RAQUEL MOURA MACHADO

POLIPLOIDIA EM *Psidium cattleyanum* Sabine (Myrtaceae): IMPLICAÇÕES CITOGENÉTICAS E EVOLUTIVAS

POLYPLOIDY IN *Psidium cattleyanum* Sabine (Myrtaceae): CYTOGENETIC AND EVOLUTIONARY IMPLICATIONS

CAMPINAS
2021

RAQUEL MOURA MACHADO

POLIPLOIDIA EM *Psidium cattleyanum* Sabine (Myrtaceae): IMPLICAÇÕES CITOGENÉTICAS E EVOLUTIVAS

POLYPLOIDY IN *Psidium cattleyanum* Sabine (Myrtaceae): CYTOGENETIC AND EVOLUTIONARY IMPLICATIONS

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Doutora em Biologia vegetal

Thesis presented to the Institute of Biology at the State University of Campinas as part of the requirements for obtaining a PhD in Plant Biology

Orientadora: Profa. Dra. Eliana Regina Forni Martins

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELA ALUNA RAQUEL MOURA MACHADO E ORIENTADA PELA PROFA. DRA. ELIANA REGIRA FORNI MARTINS

CAMPINAS

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

M18p

Machado, Raquel Moura, 1990-
Poliploidia em *Psidium cattleyanum* Sabine (Myrtaceae) : implicações citogenéticas e evolutivas / Raquel Moura Machado. – Campinas, SP : [s.n.], 2021.

Orientador: Eliana Regina Forni Martins.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Tamanho de genoma. 2. Cariótipos. 3. Poliploide. I. Forni-Martins, Eliana Regina, 1957-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Polyploidy in *Psidium cattleyanum* Sabine (Myrtaceae) : cytogenetic and evolutionary implications
Palavras-chave em inglês:
Genome size
Karyotypes
Polyploidy
Área de concentração: Biologia Vegetal
Titulação: Doutora em Biologia Vegetal
Banca examinadora:
Eliana Regina Forni Martins [Orientador]
Ana Paula de Moraes
Carolyn Elinore Barnes Proença
Gabriela Silvina Speroni Gómez
Fabio Pinheiro
Data de defesa: 29-04-2021
Programa de Pós-Graduação: Biologia Vegetal

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-9603-5747
- Currículo Lattes do autor: http://lattes.cnpq.br/5627195467622523

Campinas, 29 de abril de 2021

COMISSÃO EXAMINADORA

Profa. Dra. Eliana Regina Forni Martins

Dra. Ana Paula de Moraes

Dra. Carolyn Elinore Barnes Proença

Dra. Gabriela Silvina Speroni Gómez

Prof. Dr. Fábio Pinheiro

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa de pós graduação em Biologia vegetal do Instituto de Biologia da Universidade Estadual de Campinas.

A todas as mulheres que já lutaram ou continuam lutando por igualdade e respeito, dedico.

AGRADECIMENTOS

A jornada de um doutorado é como escalar uma montanha, são quatro anos com altos e baixos, penhascos e paisagens deslumbrantes, assim fincar a bandeira no topo é uma das grandes conquistas de uma vida e, também, o começo de uma nova aventura. Mas como toda longa e difícil jornada, eu não teria conseguido sem apoio, então esse é o meu agradecimento.

Gostaria de agradecer a UNICAMP, ao programa de Pós-Graduação em Biologia Vegetal do Instituto de Biologia e ao Departamento de Biologia Vegetal (DBV) por fornecer a estrutura para a execução desse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Agradeço também Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo número 141938/2020-0, pela concessão da bolsa de estudos.

Agradeço a Profa. Dra. Eliana Regina Forni Martins, minha orientadora, por sempre me incentivar, mesmo nas ideias mais mirabolantes, durante o mestrado e doutorado, pela paciência e entusiasmo, por ser a referência de orientadora, profissional e ser humano que eu levarei para a vida.

Agradeço também aos professores e funcionários do DBV que diretamente ou indiretamente contribuíram para essa jornada, desde tomar um cafezinho junto até os ensinamentos partilhados. E aos gestores e funcionários das Unidades de Conservação e todas as pessoas que contribuíram com meus trabalhos de campo (uma lista imensa), agradeço.

Aos membros do exame de qualificação: Prof. Dr. Fábio Pinheiro, Dra. Prianda Laborda e Prof. Dr. Ricardo Augusto Lombello, da análise prévia: Prof. Dr. Fábio Pinheiro, Dra. Bárbara Leal e Profa. Dra. Ana Paula de Moraes, aos membros da banca de defesa: Profa. Dra. Eliana Regina Forni Martins, Profa. Dra. Ana Paula de Moraes Profa. Dra. Carolyn Elinore Barnes Proença, Profa. Dra. Gabriela Silvina Speroni Gómez e Prof. Dr. Fábio Pinheiro e suplentes: Dra. Amanda Teixeira Mesquita, Dra. Bárbara Simões Santos Leal e Profa. Dra. Samantha Koehler, obrigada pela leitura crítica dos textos e sugestões dadas para melhorar o meu trabalho.

Aos amigos do LEEG (Laboratório Ecologia e Evolutiva e Genômica de Plantas) por todos os bolos divididos, ensinamentos e conversas (dentro e fora do horário de trabalho), muito obrigada. Aos chefes do LEEG, Profa. Dra. Clarisse Palma da Silva e Prof. Dr. Fábio Pinheiro, obrigada por fornecerem a estrutura e suporte para a conclusão deste trabalho.

Agradeço a todos do Laboratório de Análise Genética e Molecular (LAGM) por me acolherem tão bem, pela paciência de repassar todos os ensinamentos e pelos bolos compartilhados. Em especial, agradeço Profa. Dra. Anete Pereira de Souza por ser uma colaboradora tão entusiasmada e por dar toda a estrutura e o suporte para que eu realizasse as análises de genética. Também agradeço a Dra. Fernanda Ancelmo e ao Dr. Fábio Matos por me serem meus mentores na parte prática de laboratório.

Agradeço a todos os amigos da Biologia Vegetal da Unicamp companheirismo, amizades e, principalmente, conversas pós almoços e festinhas, vocês foram essenciais para dar força nessa escalada. Aos amigos especiais que essa jornada me trouxe, os “TheMonios”, “Café da Tarde da Sofía”, “RepRegina” e do “Thatatatiticiannel”, vocês tornaram esse caminho mais leve ao longo desses anos, obrigada!

Um obrigado especial a minha família de São Paulo, que me acolheu com gentileza e respeito, me apoiando em todas as etapas desde que me conheceram. À minha família de Fortaleza, em especial minha mãe Ana, meu padrasto Junior e meu irmão Levi, o meu muito obrigada por me darem raízes, mas também asas, para que eu pudesse sair e realizar meus sonhos.

Para o meu companheiro de vida, André Vito Scatigna, agradeço imensamente, por todo apoio sempre (desde trabalho de campo até mudanças de vida). Obrigada por estar ao meu lado a cada passo dessa jornada. Ainda vamos escalar muitas montanhas juntos e manteremos em mente que ficar a bandeira no topo, apesar de difícil, sempre vale a pena.

RESUMO

A poliploidia é uma das principais forças que atuam na diversificação nas Angiospermas. Um grande esforço tem sido empregado em mensurar a extensão dos efeitos da poliploidia na evolução de grupos de plantas, no entanto poucos sistemas poliploides são bem estudados. Estudos que envolvem análise da distribuição espacial dos citótipos combinados com dados citogenéticos, de diversidade e estrutura genética de populações, podem elucidar questões relacionadas à evolução de grupos através da poliploidização. Com diversos níveis de ploidia relatados na literatura, *Psidium cattleyanum* Sabine (Myrtaceae) é um complexo poliploide natural de ampla distribuição geográfica, o que torna um modelo ideal para testar os efeitos da poliploidia na evolução de grupos de plantas. O principal objetivo da tese foi investigar os citótipos de *P. cattleyanum* sob os aspectos citogenético (Capítulo 1), genético (Capítulo 2 e 3) e ecológico (Capítulo 3). Usando técnicas de bandamento com fluorocromos, mapeamento físico dos genes do DNA ribossomal (18S e 5S) e estimativa do conteúdo de DNA no Capítulo 1, nós mostramos que em *P. cattleyanum* possui 13 citótipos (2n = 3x = 33 até 12x = 132), sendo quatro citótipos não relatados na literatura (2n = 3x = 33, 9x = 99, 10x = 110, 12x = 132) e que há um aumento linear de bandas CMA^{+}, sinais de DNAr e conteúdo de DNA de acordo com o nível de ploidia, porém nos citótipos com ploidia mais alta (11x e 12x) observamos rearranjos cromossômicos e o fenômeno de "*genome downsize"*. Também concluímos que há uma tendência dos níveis de ploidia mais altos distribuírem-se mais próximo do equador e confirmamos que o nível de ploidia não se relaciona com a cor do fruto. No Capítulo 2, usando três citótipos de *P. cattleyanum* (2n = 4x = 44, 8x = 88, 12x = 132) desenvolvemos e caracterizamos marcadores genômicos do tipo microssatélites e realizamos um estudo preliminar de genética de populações em uma amostragem reduzida, adicionalmente investigamos a transferência cruzada para outras espécies do gênero *Psidium*. Além disso no Capítulo 3, acessamos a diversidade genética e estrutura das populações naturais de *P. cattleyanum*, com uso de microssatélites. Realizamos análises de nicho climático para estimar a similaridade e divergência do nicho ambiental para cada citótipo e investigamos a contribuição relativa de variáveis geográficas e ambientais que moldaram a organização da diversidade genética em populações de *P. cattleyanum*. Nesse capítulo, observamos que a diversidade genética diminui com o aumento da ploidia, que as populações de *P. cattleyanum* estão diferenciadas de acordo com a ploidia e os citótipos possuem baixa similaridade e alta divergência de nicho, mostrando que citótipos podem agir como uma etapa intermediária na divergência entre linhagens.

Palavras-chave: tamanho de genoma, evolução cariotípica, citótipos, especiação poliploide.

ABSTRACT

Polyploidy is one of the main forces underlying the diversification of Angiosperms. Efforts have been made to measure the extent of the effects of polyploidy on the evolution of groups of plants, however few polyploid systems are well studied. Studies involving analysis of the spatial distribution of cytotypes combined with cytogenetic data, genetic diversity, and population structure, can elucidate issues related to the evolution of groups by polyploidization. With several levels of ploidy reported in the literature, *Psidium cattleyanum* Sabine is a natural polyploid complex with wide geographic distribution, making this species an ideal model to test the effects of polyploidy on the evolution of groups of plants. The main goal of the thesis was to investigate the cytotypes of *P. cattleyanum* on the cytogenetic (Chapter 1), genetic (Chapter 2 and 3) and ecological (Chapter 3) aspects. Using fluorochromes banding techniques, physical mapping of ribosomal DNA genes (18S and 5S), and estimation of DNA content, we showed that in *P. cattleyanum* there are four cytotypes not reported in the literature (2n = 33, 99, 110, 132) in Chapter 1. Also, we noticed a linear increase in CMA^+ bands, DNAr signals, and DNA content according to the ploidy level; however, in cytotypes with higher ploidy (11x and 12x) we observed chromosomic rearrangements and the phenomenon of genome downsize. We also concluded that there is a tendency for higher ploidy levels to be distributed closer to the equator and confirmed that the ploidy level is not related to the color of the fruit. In Chapter 2, using three cytotypes of *P. cattleyanum* (2n = 44, 88, 132) we developed and characterized microsatellite genomic markers, and carried out a preliminary study of population genetics in a small sample. Additionally, we investigate the cross-amplification to other species of the *Psidium*. Moreover, in Chapter 3, using microsatellite markers, we accessed the genetic diversity and structure of natural populations of *P. cattleyanum* and performed climatic niche analyzes to estimate the similarity and divergence of the environmental niche for each cytotype. We also investigated the contribution of geographic and environmental variables that shaped the organization of genetic diversity in populations of *P. cattleyanum*. In this final chapter, we observed that genetic diversity decreases with the increase in ploidy. The populations of *P. cattleyanum* are differentiated according to ploidy and the cytotypes have low similarity and high niche divergence, showing that cytotypes can act as an intermediate stage in the divergence between lineages.

Key words: genome size, karyotype evolution, cytotypes, polyploid speciation.

SUMÁRIO

Organização geral da tese

Uma breve revisão da literatura está apresentada na introdução geral, onde abordo os principais tópicos que serão discutidos nos três capítulos. No primeiro capítulo de minha tese, nós determinamos número cromossômico e o nível de ploidia de 27 localidades de ocorrência de *Psidium cattleyanum* e comparamos, com marcadores citogenéticos (bandamento CMA/DAPI e hibridação *in situ*), os citótipos ao longo da distribuição natural da espécie. Adicionalmente, estimamos o conteúdo de DNA das populações naturais por citometria de fluxo e observamos uma relação entre ploidia/conteúdo de DNA. Esse capítulo foi escrito na forma de manuscrito intitulado "Cytogenetics analysis and DNA content of *Psidium cattleyanum* Sabine (Myrtaceae): distribution of a polyploid complex at a broad scale" e será submetido à revista *Trees*.

No segundo capítulo, desenvolvemos e caracterizamos marcadores genômicos do tipo microssatélites para três citótipos de *P. cattleyanum* (2n=44, 88, 132), realizamos um estudo preliminar de genética de populações utilizando uma amostragem reduzida e investigamos a transferência cruzada para outras espécies do gênero *Psidium*. Esse capítulo foi escrito na forma de manuscrito intitulado "Population genetics of polyploid complex *Psidium cattleyanum* Sabine (Myrtaceae): preliminary analyses based on new species-specific microsatellite loci and extension to other species of the genus" e já foi publicado na revista *Biochemical Genetics* (https://doi.org/10.1007/s10528-020-10002-1).

No terceiro e último capítulo, avaliamos a diversidade genética e os padrões de estrutura genética em populações de *P. cattleyanum* usando microssatélites. Além disso, usamos a análise de nicho climático no clima atual para estimar a similaridade e divergência do nicho ambiental para cada citótipo. Adicionalmente, investigamos a contribuição relativa de variáveis geográficas e ambientais que moldaram a organização da diversidade genética em populações de *P. cattleyanum*. Assim, usando *P. cattleyanum* como modelo, testamos duas hipóteses relacionadas à especiação poliploide. Este capítulo também já foi escrito em forma de manuscrito intitulado "Population structure and intraspecific ecological niche divergence of *Psidium cattleyanum* points to evidence of a future divergent lineage" e será submetido para a revista *Annals of Botany*. Ao final, trago as principais contribuições da tese na forma de considerações finais.

Introdução geral

Poliploidia: avanços e controvérsias

A poliploidia tem intrigado cientistas há cerca de 100 anos (ver os trabalhos de Stebbins 1971, Grant 1981, entre outros), e refere-se a presença de múltiplos conjuntos cromossômicos por célula (Van de Peer et al. 2017, Fox et al. 2020). A poliploidização é um fenômeno ubíquo em diversas linhagens de plantas (Wendel 2015, Wood et al. 2009, Gao 2020). Com o avanço das técnicas de genômica, sequenciamento de genomas e transcriptomas inteiros, mostrou-se que a poliploidia está presente em diversos organismos, e mesmo os considerados diploides apresentam algum traço de pelo menos um ciclo de poliploidização ancestral (Wendel 2015, Van de Peer et al. 2017, Fox et al. 2020, Van de Peer et al. 2020). Cerca de 30% a 70% das angiospermas possuem ancestrais poliploides (Wood et al. 2009).

Eventos de duplicação do genoma inteiro (*whole-genome duplication*- WGD), que originam organismos poliploides, podem ser rastreados até a origem e diversificação de diversas linhagens de plantas (Wood et al. 2009, Wendel 2015, Van de Peer et al. 2017, Van de Peer et al. 2020). Porém os desafios para os poliploides recém-formados são muitos, pois as consequências de um evento de WGD vão desde instabilidade genômica até a exclusão do citótipo em minoria (*minority cytotype exclusion*- Levin 1975). Da mesma maneira que as múltiplas cópias podem tamponar o efeito de genes deletérios, várias mutações negativas podem se acumular, sendo desvantajoso ao nível populacional (Fox et al. 2020). Assim, esse tema tem levado cientistas a amplos debates: seria a poliploidia um beco sem saída evolutivo (Mayrose et al. 2011), ou a fonte de variação genética e fenotípica para a diversificação das espécies (Soltis et al. 2014)?

As WGD parecem estar relacionadas historicamente com períodos de extinção ou mudanças no clima global (Baduel et al. 2018, Van de Peer et al. 2020). Mesmo com todos os desafios impostos por um evento de WGD há evidências claras de diversos ciclos de poliploidização ao longo da evolução, que levaram a uma abundância de espécies poliploides na natureza (Baduel et al. 2018). Entretanto, apesar de a poliploidia estar presente em diversos grupos de eucariotos, a compreensão da dinâmica evolutiva e de herança dos poliploides é menos avançada se comparada à de espécies diploides (Jighly et al. 2020).

Uma vez que a frequência de eventos de poliploidia tende a aumentar em situações de estresse ou mudanças em condições ambientais (Baduel et al 2018, Van de Peer et al. 2020), acredita-se que a poliploidia será a forma mais importante de especiação de plantas nos próximos 500 anos, no contexto de aceleração das mudanças climáticas (Levin 2019, Gao 2020). A poliploidia é um fenômeno comum na natureza, no passado, presente e, possivelmente, no futuro, porém existem aspectos desse fenômeno que poderão ser melhor explorados com o avanço tecnológico (Fox et al. 2020).

Poliploidia: de célula a espécies

Tradicionalmente, os poliploides são divididos em duas vias de formação: a autopoliploidia e a alopoliploidia. Na via autopoliploide, ocorre a multiplicação de um mesmo genoma, enquanto na via alopoliploide há duplicação de um genoma originado de um evento de hibridação (Schifino-Wittmann e Agnol 2003). Fatores como as múltiplas origens da poliploidia e a possibilidade de haver fluxo gênico entre as populações diploides e as poliploides, bem como entre os poliploides de diferentes origens, mostram que a poliploidia, em plantas, é um processo dinâmico (Ramsey e Schemske 1998, Schifino-Wittmann 2004, Baniaga et al. 2020). Anteriormente, acreditava-se que os alopoliploides eram naturalmente formados em maiores taxas, porém, atualmente, sabe-se que autopoliploides são formados em taxas semelhantes aos alopoliploides, mas, por serem indistinguíveis dos parentais diploides, são muitas vezes subestimados (Schifino-Wittmann 2004, Soltis et al. 2007, Parisod et al. 2009).

Os eventos de poliploidização levam ao aumento do número cromossômico e do conteúdo de DNA. Estudos recentes também mostram que a presença de mais de um conjunto cromossômico por célula gera inúmeras alterações que vão desde à expressão diferenciada de genes a mudanças no tamanho das células e/ou fisiologia intracelular (revisado por Baduel et al. 2018); assim observa-se que poliploides tendem a responder diferentemente a fatores abióticos, por exemplo tolerância ao estresse (Fox et al. 2020), e bióticos, como interação com polinizadores (Rezende et al. 2020, Baniaga et al. 2020, Van de Peer et al. 2020).

Para que haja o estabelecimento e persistência de um neoploliploide, após um evento de WGD, esse novo organismo deve vencer barreiras tanto a curto quanto a longo prazo (Baduel et al. 2018). A súbita duplicação do número cromossômico leva imediatamente a interrupção do pareamento e segregação regular dos cromossomos

homólogos, interrompendo o processo reprodutivo (Baduel et al. 2018). Há uma forte associação entre poliploidia e reprodução assexuada (Van Drunen e Husband 2017, Baduel et al. 2018). E é, provavelmente, uma forma dos poliploides recém formados continuarem se propagando na população. Outra forte barreira ao estabelecimento e persistência de poliploides em uma população é a exclusão do citótipo minoritário (Levin 1975), assim a reprodução assexuada pode contornar os efeitos desses cruzamentos desbalanceados entre os progenitores diploides (Van Drunen e Husband 2017, Baduel et al. 2018).

Outra possibilidade de escapar da exclusão do citótipo minoritário é a diferenciação de nicho entre os diploides e os poliploides. Ainda não existe uma teoria unificada sobre as consequências ecológicas da poliploidia (Parisod e Broennimann 2016). Dados contratastes mostram que há grupos com citótipos poliploides que apresentam distribuição geográfica diferente e diferenciação de nicho, como em *Dianthus broteri* Boiss. & Reut. (López-Jurado et al. 2019), *Saxifraga rosacea* Moench (Decanter et al. 2020), enquanto outros grupos apresentaram padrões diferentes, como conservação e contração do nicho ambiental (revisado por Glennon et al. 2014). Dados recentes suportam a ideia de diferenciação de nicho climático entre diploides e poliploides e taxas de diferenciação mais rápidas (Baniaga et al. 2020).

Uma vez que na alopoliploidia inicialmente há a hibridação entre duas espécies, o conjunto extra de cromossomos promove um estado multialélico estável ao longo das gerações, gerando variabilidade genética (Baduel et al. 2018), já em autopoliploides essa maior diversidade genética vem da herança polissômica (Haldane 1930). O aumento dos alelos, que ocorre após o evento de poliploidização, pode aumentar a probabilidade de mascarar efeitos de genes deletérios e de adquirir mutações benéficas (Otto 2007, Baniaga et al. 2020).

A maior diversidade genética pode ter um impacto positivo na competitividade dos citótipos, permitindo a colonização de novos habitats (Otto e Whitton 2000, Soltis e Soltis 2000). Esse é o caso de espécies invasoras, onde já foi demostrado que a poliploidia contribui para o sucesso das invasões biológicas (Pandit et al. 2011, te Beest et al. 2011, Moura et al. 2021). Uma vez que condições ambientais adversas, altas ou baixas temperaturas, favorece a produção de gametas não reduzidos (Ramsey e Schemske 1998, Rice et al. 2019, Fox et al. 2020), é provável que a frequência de poliploides aumente em áreas com perturbações ambientais e condições ambientais extremas (Baduel et al. 2018, Rice et al. 2019, Fox et al. 2020).

Citótipos podem ser considerados como um passo intermediário na divergência entre linhagens (Briggs e Walters 1997, Otto e Whitton 2000). Porém a inclusão de múltiplos citótipos dentro da mesma espécie e a tradição de nomear uma espécie baseado em morfologia, são as principais razões por não percebermos espécies originadas por autopoliploidia (Soltis et al. 2007). Já alopoliploides que geralmente combinam atributos dos parentais diploides são mais facilmente reconhecidos como espécie (Soltis et al. 2007).

O estabelecimento de novas linhagens poliploides é um desafio com altos ganhos, mas também altas perdas para as espécies (Baduel et al. 2019), apesar disso, é inegável a influência da poliploidia desde células a ecossistemas (Fox et al. 2020). Devido à dificuldade de obtenção de dados empíricos diretos, mensurar a extensão dos efeitos da poliploidia pode ser difícil (Baduel et al. 2018). Como discutido por Marchant et al. (2016), poucos sistemas poliploides têm as relações evolutivas bem documentadas. Assim, complexos poliploides oferecem uma oportunidade de testar a influência dos eventos de poliploidização nos processos que governam a evolução por poliploidia e o papel da variação de ploidia na variabilidade genética, adaptação em novos ambientes e especiação (Kolár et al. 2017, Gao 2019, Baniaga et al. 2020).

O modelo de estudo: complexo poliploide da espécie *Psidium cattleyanum* Sabine

Psidium L. é um gênero neotropical da família Myrtaceae Juss. com cerca de 100 espécies (Tuler et al. 2021). É um gênero monofilético e com rápida taxa de diversificação (Vasconcelos et al. 2017). No Brasil ocorrem cerca de 60 espécies, distribuídas em diversas formações vegetacionais, com a maior porcentagem das espécies ocorrendo na Floresta Atlântica, Cerrado e Caatinga (Tuler et al. 2021).

Dentre os diversos atributos que explicam a ampla distribuição geográfica e o sucesso em colonizar diversificados ambientes de algumas espécies de *Psidium*, a poliploidia destaca-se como uma característica chave para a origem, diversificação e distribuição do gênero (Costa e Forni-Martins 2006, Marques et al. 2016, Vasconcelos et al. 2017, Tuler et al. 2019, Machado et al. 2020). *Psidium* possui registros poliploides variando de 2n = 22 = 2x até 132 = 12x (Costa e Forni-Martins 2006, Éder e Silva et al. 2007, Marques et al. 2016, Tuler et al. 2019, Machado et al. 2020, Machado et al. 2021-Capítulo 1), derivando do número cromossômico básico de Myrtaceae de x = 11 (Atchison 1947).

Psidium cattleyanum possui uma grande diversidade de níveis de ploidia, caracterizando-se como um complexo poliploide. Já foram relatados os números cromossômicos de 2n = 33, 44, 46, 48, 55, 58, 66, 77, 82, 88, 99, 110 e 132 (Atchison 1947, Hirano and Nakazone 1969, Singhal et al. 1985, Raseira and Raseira 1996, Costa and Forni-Martins 2006, Costa 2009, Vázques 2014, Souza et al. 2014, Souza-Pérez and Speroni 2017, Tuler et al. 2019, Machado et al. 2020, Machado et al. 2021- capítulo 1). Além dos diversos citótipos, *Psidium cattleyanum* apresenta diferenciação na cor do fruto, com um morfotipo amarelo e outro vermelho, e é distribuída naturalmente na Floresta Atlântica, na porção leste do Brasil, desde a Bahia até o Rio Grande do Sul, no norte da Argentina (Wikler 2007) e no Uruguai (Souza-Pérez e Speroni 2017). É uma das duas espécies de *Psidium* que é invasora, também é uma agressiva invasora de ilhas (Pedrosa-Macedo e Smith 2007, Wilker 2007).

Estudos citogenéticos pontuais foram realizados, no Brasil, por Raseira e Raseira (1996) que determinaram o número cromossômico para quatro populações de *P. cattleyanum* coletadas em diferentes regiões (sudoeste, planalto e litoral) no Rio Grande do Sul e no sul do Paraná, com 2n = 66 e 2n = 88 para o morfotipo vermelho e 2n = 66 para todos os indivíduos do morfotipo amarelo. Costa (et al. 2008, 2009) determinaram 2n = 66 para um indivíduo do morfotipo vermelho no Rio de Janeiro e 2n = 44 para um indivíduo cultivado em Campinas (SP) do morfotipo amarelo. O trabalho de Souza et al. (2014) estimou número cromossômico para 10 acessos de *P. cattleyanum*, relatando, além de citótipos já conhecidos (múltiplos de x = 11), citótipos que não seguem o número básico da família, como 2n = 46, 48, 58, 82.

No Uruguai, Vázques (2014) e Souza-Pérez e Speroni (2017) relataram dois citótipos de *P. cattleyanum* que ocorrem no Uruguai, em materiais em cultivo. O morfotipo amarelo possuem 2n = 77 e 2n = 88. No mesmo estudo foi relatado 2n = 77 para o morfotipo vermelho.

Estudos prévios de biologia reprodutiva, relataram erros na meiose durante a formação dos grãos de pólen, diminuindo a viabilidade polínica, o que sugere a formação de frutos por processo assexuado, explicando os citótipos ímpares já relatados para a espécie (Raseira e Raseira 1996). Souza-Pérez e Speroni (2017), usando técnicas de anatomia vegetal, detectaram a formação de frutos por via apomítica. Singhal et al. (1985) realizaram a análise de meiose em material com 2n= 77 em *P. coriaceum* (sinônimo de *P. cattleyanum*) e observaram muitas irregularidades, junto com alta esterilidade de pólen (80%). A origem da poliploidia permanece não resolvida. Vázques (2014) sugeriu que *P.*

cattleyanum amarelo, com 2n=88, possui origem autopoliploide, pois o padrão de bandas de CMA/DAPI (Cromomicina A3 e 4',6-diamidino-2-fenilindol, respectivamente) indica a repetição oito vezes de um genoma x=11.

Um estudo pioneiro em *Psidium* integrou diversidade genética (usando marcadores moleculares do tipo microssatélites), números cromossômicos, conteúdo de DNA e distribuição geográfica (Tuler et al. 2019). Os autores observaram que as espécies do gênero que eram poliploides possuíam maior distribuição geográfica e eram mais diversos. *Psidium cattleyanum* foi incluído nesse estudo e apresentou uma alta diversidade genética e ampla distribuição geográfica.

Psidium cattleyanum é uma espécie chave para investigar a influência dos eventos de poliploidização na origem e diversificação de grupos de plantas, principalmente do gênero *Psidium*. Apesar desse potencial, não há estudos que integram diversas áreas do conhecimento em *P. cattleyanum*, apenas estudos que abordam apenas uma área específica, como citogenética ou biologia reprodutiva.

Uma vez que as discussões sobre os mecanismos de formação e estabelecimento dos poliploides ganharam força nos últimos anos (Van de Peer et al. 2017, Baduel et al 2018, Levin 2019, Fox et al. 2020, Gao 2020, Van de Peer et al. 2020), estudos multidisciplinares são imprescindíveis para mensurar os efeitos da poliploidia desde células até ecossistemas (Fox et al. 2020). Enquanto o WGD é reconhecido como uma importante força evolutiva para grupos de plantas, os fatores genéticos e ecológicos que afetam a abundância de linhagens poliploides ainda são insuficientemente conhecidas (Parisod et al. 2010).

Além disso, a prática de incluir citótipos que são morfologicamente semelhantes em uma única espécie, pode obscurecer os *insights* sobre a evolução e especiação (Soltis et al. 2007), assim estudos multidisciplinares são essenciais para o avanço no conhecimento da evolução de espécies com múltiplos citótipos. Para isso, *P. cattleyanum* torna-se um modelo ideal para abordagens multidisciplinares que testam os efeitos da poliploidia em diversos aspectos da biologia dos organismos e populações.

Referências

As referências da Introdução Geral estão ao final da Tese.

Cytogenetics analysis and DNA content of *Psidium cattleyanum* Sabine (Myrtaceae): distribution of a polyploid complex at a broad scale

Raquel Moura Machado[1,*] and Eliana Regina Forni-Martins[2]

[1]Programa de Pós-graduação em Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Rua Monteiro Lobato, 255, 13083-862, Campinas, SP, Brazil.

[2]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Rua Monteiro Lobato, 255, 13083-862, Campinas, SP, Brazil.

*Corresponding author: raquelmouramachado@gmail.com

Authors' ORCID:

Raquel Moura Machado - 0000-0002-9603-5747

Eliana R. Forni-Martins - 0000-0002-8889-4955

Abstract

Polyploidy is a ubiquitous process in Angiosperms, however, few natural polyploid complexes in tropical regions have been well described. *Psidium cattleyanum* (Myrtaceae) is a neotropical fruit species with a wide distribution and several ploidy levels reported in the literature, being characterized as a polyploid complex. In this study, we provide cytogenetic data of *P. cattleyanum* (chromosome numbers, distribution and location of CMA/DAPI bands and rDNA sites, and the genome size), using flow cytometry and conventional and molecular cytogenetic techniques. Moreover, we analyzed the geographic distribution of *P. cattleyanum* cytotypes. For 106 specimens (75 new samples and 31 from literature), we observed nine cytotypes, with the most frequent being 2n = 7x= 77. We performed CMA/DAPI banding and FISH technique for eight and six cytotypes, respectively. CMA/DAPI bands and 18S and 5S rDNA sites increased with ploidy level, suggesting a multiplication of x = 11 genome. In flow cytometry analyses, the values of 2C were positivelly correlated to the ploidy, except the cytotypes 2n = 110 and 132, in which the values were relatively lower than expected. Our results support the occurrence of a polyploid series in *P. cattleyanum*, derived by basic number x = 11, and suggest the autopolyploid origin of the complex. The distribution of cytotypes indicated a trend of higher ploidy levels in lower latitudes. Therefore, *Psidium cattleyanum* is a

good model group to study the role of polyploidization events in the evolutionary processes of natural populations.

Keywords: genome size, karyotype evolution, cytotypes, autopolyploidy, araçá

Declarations

Funding- This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and FAEPEX/UNICAMP.

Conflicts of interest- The authors declare no conflicts of interest.

Ethics approval- This article does not include studies about humans or animals performed by any of the authors.

Authors' contributions- All authors contributed equally to this manuscript.

Key Message

Psidium cattleyanum, a neotropical polyploid complex, is a potential model group to investigate chromosomic rearrangements associated with polyploidization and the influence of polyploidy in species.

Introduction

The distribution of polyploid plants shows latitudinal trends across the globe with polyploid frequency increasing away from the equator (Rice et al. 2019). Actually, there are just a few natural polypoid systems are well described in the tropical region (Reis et al. 2014), e.g., *Lippia alba* (Mill.) N. E. Brown (Verbenaceae; Reis et al. 2014), *Zygopetalum mackayi* Hook. (Orchidaceae; Gomes et al. 2008), even though the tropical region has high species diversity, and polyploidy is frequently associated with plant diversification (Soltis et al. 2014; Levin 2019).

Polyploidy refers to the gain of at least one chromosome set per cell (te Beets et al. 2011, Van de Peer et al. 2017) and is a frequent phenomenon in plants, especially in Angiosperms (Wood et al. 2009; Dodsworth et al. 2016). Polyploidy generates genomic changes (from the chromosome number to the DNA content), epigenetics, in plant growth and physiology (te Beest et al. 2011; reviewed by Baduel et al. 2018). Studies suggest that polyploidization, especially autopolyploidy, will be the most important driver of plant speciation for the next 500 years in the context of accelerating climate change (Gao 2019; Levin 2019). Thus, polyploid complexes offer unique opportunities to obtain

empirical data to test processes that govern evolution by polyploidy and the role of polyploidization events in adaptation and speciation (Kolár et al. 2017).

Intraspecific ploidy variation is reported in many plant groups (Husband et al. 2013), including some species of *Psidium*, as demonstrate by Tuler et al. (2019) and Marques et al. (2016). Myrtaceae Juss. is a karyotypically homogeneous family with the basic chromosome number x = 11 (Atchison 1947). Within this family, however, the genus *Psidium* L. has a large number of polyploid records, with chromosome numbers ranging from 2n = 2x = 22 to 9x = 98 (Atchison 1947; Costa 2009; Costa and Forni-Martins 2006; Marques et al. 2016; Souza et al. 2014; Srivastava 1977; Vijayakumar and Subramanian 1985; Tuler et al. 2019; Machado et al. 2020).

Psidium cattleyanum Sabine has nine cytotypes reported in the literature: 2n = 4x = 44, 4x = 46, 4x = 48, 5x = 55, 5x = 58, 6x = 66, 7x =77, 8x = 82 to 8x = 88 (Atchison 1947; Hirano and Nakazone 1969; Singhal et al. 1985; Raseira and Raseira 1996; Costa and Forni-Martins 2006; Costa 2009; Vázques 2014; Souza et al. 2014; Souza-Pérez and Speroni 2017; Tuler et al. 2019; Machado et al. 2020). Despite the number of cytotypes already known, most of the cytogenetic analysis of this species is based only on counting and measuring chromosomes with conventional staining. Some efforts were made by Vázques (2014), using chromomicinA3 (CMA) and 4',6-diamidino-2-phenilindole (DAPI) fluorochromes banding in two cytotypes 2n = 77 and 2n = 88. Based on the band pattern, the author suggested that *P. cattleyanum* with 2n = 88 has an autopolyploid origin, by presenting an eight-fold genome repetition for a base chromosome number x = 11. However, she recommended mapping 18S and 5S ribosomal RNA genes using FISH (Fluorescent *in situ* hybridization) to confirm this hypothesis. Another aspect investigated in some cytotypes of *P. cattleyanum* was the genome size (Costa et al. 2008; Vázques 2014; Souza et al. 2014), for which it was found that 2C value shows a direct relationship with the ploidy level of the cytotypes. Although there are many ploidy levels reported in the literature, knowledge on the distribution patterns of heterochromatin and distribution of sequences of rDNA in the different cytotypes of *P. cattleyanum* are still very limited.

Psidium cattleyanum is an edible fruit plant, widely distributed, that occurs in areas of the Atlantic Forest, from the northeast of Brazil (Bahia) to the north of Uruguay (Normand and Habbib 2001; Sobral and Sousa 2007). It was introduced in Hawaii and in several parts of the world and has become a serious invader. The fruits are yellow or red (Sobral and Sousa 2007), characterizing two morphotypes, that are commonly known as Araçá-amarelo and Araçá-vermelho. Previous studies have failed to associate certain

levels of ploidy with one of the two fruit colors (yellow and red morphotype), in the south (2n = 6x = 66 and 2n = 8x= 88, Raseira and Raseira 1996) and southeast (2n = 4x = 44 and 2n = 6x = 66, Costa 2008, 2009) of Brazil, in Uruguay (2n = 7x = 77 and 2n = 8x = 88, Vázques 2014) and also in Hawaii (2n = 6x = 66, 2n = 7x = 77 and 2n = 8x = 88, Hirano and Nakasone 1969).

The reproductive strategies of *P. cattleyanum* includes autogamy, allogamy, apomixis and root sprout (Raseira and Raseira 1996; Franzon et al. 2009; Souza-Perez and Speroni 2017). The presence of polyploidy and the several reproductive strategies could potentially explain the success of *P. cattleyanum* to colonize different habitats (Machado et al. 2020).

The fact that *P. cattleyanum* has several polyploid records and morphological variation makes this species a great model to measure the extent of polyploidy effects in chromosome numbers, distribution and location of CMA/DAPI bands and rDNA sites, and DNA content. In this study, we investigated cytotype variability (chromosome numbers, distribution and location of CMA/DAPI bands and rDNA sites and the genome size) of *P. cattleyanum* in different localities along the Atlantic Forest, to provide cytogenetic data of this tropical polyploidy complex at a broad scale. Additionally, we discussed about polyploid origin in *P. cattleyanum*. Thus, we can address the following questions: (1) Are cytotypes associated with a specific fruit color morphotype and/or a latitudinal gradient? (2) Is the genome size related to the increasing of chromosome numbers? (3) Based on cytogenetic data, what is the polyploid origin of *P. cattleyanum* cytotypes?

Material and Methods

Plant material

We collected seeds and fruits of 75 *P. cattleyanum* individuals from 27 localities encompassing most of the species native distribution in Brazil (varying from one to eight individuals per locality): Rio Grande do Sul (RS), Santa Catarina (SC), Paraná (PR), São Paulo (SP), Rio de Janeiro (RJ), Espírito Santo (ES), and Bahia (BA). Vouchers of all collection sites were deposited in the Herbarium UEC of the University of Campinas (Table 1; Online Resource 1).

Table 1 Sampling *Psidium cattleyanum* used in this study voucher, and collection sites information.

Locality	Voucher	Longitude	Latitude
Porto Alegre-RS-Brazil	*R.M. Machado 05*	-53.759	-29.721
Uruguaiana-RS-Brazil	*A.V. Scatigna 670*	-57.088	-29.755
Caxias do Sul-RS-Brazil	*R.M. Machado 06*	-51.179	-29.168
Estrada para Blumenau-SC-Brazil	*R.M. Machado 07*	-48.842	-26.309
Imbituba-Praia Redmelha-SC-Brazil	*R.M. Machado 08*	-48.587	-26.803
Dr. Pedrinho-SC-Brazil	*R.M. Machado 09*	-49.509	-26.731
Curitiba-PR-Brazil	*R.M. Machado 1*	-49.233	-25.446
Paranaguá- Ilha do Mel-PR-Brazil	*R.M. Machado 10*	-48.304	-25.561
Pontal do sul-PR-Brazil	*R.M. Machado 11*	-48.45	-25.67
Ubatuba-Praia da Fazenda-SP-Brazil	*R.M. Machado 12*	-45.069	-23.43
S. L. do Paraitinga- PESM-SP-Brazil	*R.M. Machado 13*	-45.145	-23.336
Ubatuba-Vila de Picinguaba-SP-Brazil	*R.M. Machado 14*	-45.069	-23.43
Bertioga-SP-Brazil	*R.M. Machado 2*	-45.958	-23.778
Ilha Comprida-SP-Brazil	*R.M. Machado 14*	-47.701	-24.853
Ilha do Cardoso-SP-Brazil	*R.M. Machado 15*	-47.906	-25.066
Paraty- Pedra da Macela RJ-Brazil	*R.M. Machado 16*	-44.901	-23.138
Paraty- Praia de Trindade-RJ-Brazil	*R.M. Machado 17*	-44.708	-23.338
Paraty- Praia do Taquari-RJ-Brazil	*R.M. Machado 4*	-44.679	-23.064
Macaé- PARNAJurubatiba-RJ-Brazil	*R.M. Machado 18*	-41.468	-22.199
Itatiaia-RJ-Brazil	*R.M. Machado 19*	-44.619	-22.428
Guarapari-ES-Brazil	*R.M. Machado 20*	-40.421	-20.606
Pontal do Ipiranga- Linhares-ES-Brazil	*R.M. Machado 21*	-39.713	-19.189
Barra Nova- São Mateus-ES-Brazil	*R.M. Machado 22*	-39.739	-18.988
Itaúnas-ES-Brazil	*R.M. Machado 23*	-39.696	-18.401
Santa Teresa-ES-Brazil	*R.M. Machado 24*	-40.539	-19.97
Caravelas-BA-Brazil	*R.M. Machado 25*	-39.181	-17.728
Trancoso-BA-Brazil	*R.M. Machado 26*	-39.063	-16.47

Chromosome preparations

Root tip meristems germinated from seeds on moist filter paper in Petri dishes were collected and pretreated with 2 mM 8-hydroxyquinoline for 24 h at 8 ° C, fixed in Farmer's solution (ethyl alcohol: glacial acetic acid, 3:1) for about 24 h, and maintained at 18 ° C for further use. The 75 individuals of *P. cattleyanum* were selected for cytological preparations. The slides preparations were performed according to the

protocol described by Guerra and Souza (2002), after enzymatic digestion (1% macerozyme, 2% cellulase and 20% pectinase) for 12 minutes at 37 ° C. For each slide, root tips meristems were dissected in 45% glacial acetic acid and squashed between slide and coverslip; For coverslip remotion, the slides were immersed in liquid nitrogen for 15 min the coverslip was removed. The slides were stained using a solution of 1:1 (v/v) Vectashield® with DAPI. At least 3 slides for individual were analyzed and 5 metaphases for each slide were photographed under a microscope Olympus BX51 with a DP72 camera attached and images were captured using CellSens software (Olympus, Tokyo, Japan). Additionally, the chromosome numbers for *P. cattleyanum* were surveyed from literature data in databases- Chromosome Counts Database (CCDB; Rice et al. 2015) and Index to Plant Chromosome Numbers (IPCN), original articles, books, and academic thesis.

CMA/DAPI Banding

CMA/DAPI banding, which stains GC-rich (CMA) and AT-rich DNA regions, was performed using the Schweizer (1976) protocol with modifications. Slide preparations were stained with CMA_3 (0.5 mg/ml), covered with a coverslip, then kept in the dark for one hour. The slides were rinsed in distilled water and, afterwards, counterstained with DAPI (2 mg/ml), covered with a coverslip, and stored in the dark for half an hour. After washing in distilled water, slides were mounted in antifadent mountant solution with 12 µL of glycerol/McIlvaine/MgCl2 buffer and covered with a coverslip. Next, slides were stored in the dark at 37 ° C for at least five days for further analysis under the fluorescence microscope.

Molecular cytogenetic analyses

The two best slides from the CMA/DAPI banding for six cytotypes (Table 2) were selected for the application of the Fluorescence *in situ* hybridization technique (FISH). The sequences of ribosomal genes 5S and 18S were obtained by PCR reaction from genomic DNA of *P. cattleya*num of cytotype 2n = 4x = 44. For the 5S ribosomal DNA (rDNA) probe, 5S rDNA-3 and 5S rDNA-4 primers were used and, for 18S rDNA probe, the primers used were NS1 and NS4 (White et al. 1990; Kitamura et al. 2001). Probes were labeled by nick translation (Roche Biochemicals®, UK) using digoxigenin-11-dUTP (Roche Biochemicals®, UK) and biotin-14-dATP (Roche Biochemicals®, UK), respectively, and FISH was performed according to Schwarzacher and Heslop-Harrison

(2000) with modifications. Slides were treated with RNase for 1 h at 37 ° C fixed in 4% paraformaldehyde for 10 min at room temperature (RT) and dehydrated in 70% and 100% alcoholic series for 10 min each. The hybridization mixture was composed of 100% formamide, 10% dextran, 10% SDS, 2X SSC, and rDNA 5S and 18S probes (both 3.33 ng/µl). Post hybridization washes (with 72% of stringency) were performed in 0.1xSSC (SSC – standard saline citrate) at 42 ° C, 2xSSC at RT, 2xSSC at 42ºC, 4xSSC at RT and 4xSSC at 42ºC. Labeled probes were detected by antibodies linked to fluorochromes, thus, 5S rDNA was detected by anti-digoxygenin-Rhodamine (Roche Biochemicals®, UK) and 18S rDNA by Avidin-FITC (Roche Biochemicals®, UK). After, slides were mounted in DAPI/Vectashield (1: 1) (Vector®). Images with 18S (green) and 5S (red) signals were treated and merged using Adobe Photoshop CS (6 ed.).

Flow cytometry analyses

Specimens of *P. cattleyanum* were cultivated in a greenhouse and used for genome size (GS) measurements. Genome size of 43 individuals of *P. cattleyanum* were estimated by flow cytometry and associated with to the chromosome counts previously realized in this study (Online Resource 2). Individuals from the same locality were included in our flow cytometric measurements and direct chromosome counts to ensure correspondence between chromosome number and GS estimations (Online Resource 2).

Approximately 1 cm^2 of young leaf tissue of each individual was chopped with fresh leaf material from an internal standard (*Pisum sativum* "Ctirad" 2C = 9.09 pg; Dolezel et al. 1998) to prepare the nuclear suspensions according to Dolezel et al. (2007). About 1 mL of Ebihara buffer (Ebihara et al. 2005) was added in the extract of nuclei, then 1 mL of the nuclear suspension was filtered through a 40 µm nylon gauze (CellTrics, PARTEC) mixed with 25µL 1 mg/mL RNAse (Sigma) and digested at 37ºC for 10 min. A total of 500 µl of nuclear suspension was mixed with 25 µL 1 mg/mL propidium iodide, incubated at least 10 min and analyzed. The measurement was performed on a BD FACS Calibur flow cytometer, using acquisition software BD FACSCellquest Pro, and one run of 5,000 counts was made for each individual.

The graphs were analyzed using Flowing Software (Turku Bioscience). The estimation of the genome size was calculated using the linear relationship between fluorescence signals from stained nuclei of the unknown sample and the reference standard. To cover all the cytotypes, we decide consider in some samples with CV higher than 5% for two samples with 12x level. We used the nomenclature proposed by Leitch

et al. (1998) for genome size classification. Additionally, we surveyed the literature data to complement our data.

Statistical analysis

To test whether the DNA content is related to the increase of ploidy level, a Spearman correlation test was performed using the mean between each ploidy level (Table 2; Online Resource 2) after testing the normality of the data (using Shapiro-Wilk normality test). To understand the contribution of each chromosome set, the data was plotted on a graph with the expected values and the observed values of known cytotypes. In addition, to test whether the ploidy level is related to morphotype, and whether it is related to latitude, Spearman's correlation test was performed. Analyzes were carried out in the R environment, using the basic R stats package.

Ploidy level distribution map

The individuals analyzed both in cytogenetics and genome size were georeferenced and organized in an Excel spreadsheet with the respective geographical coordinates written in decimal degrees and plotted on a Brazilian shapefile using the QGIS program. 3.2 for making spatial distribution map.

Results

Chromosome counts, map of cytotypes distribution, and correlations

For 75 individuals analyzed in this study, chromosome counts showed nine cytotypes: 2n = 3x = 33, 2n = 4x = 44, 2n = 5x = 55, 2n = 6x = 66, 2n = 7x = 77, 2n = 8x = 88, 2n = 9x = 99, 2n = 10x = 110 e 2n = 12x = 132 (Fig. 1; Online Resource 1), confirming five cytotypes from literature (2n = 44, 55, 66, 77 and 88) and adding four new counts (2n = 33, 99, 110 and 132). We gathered 31 literature records for *P. cattleyanum,* totalizing 106 records of species. Based on our data together with the surveyed literature, the four most frequent cytotypes of *P. cattleyanum* were 7x, 6x, 4x and 8x, with 22%, 21%, 21%, 12% respectively (Fig. 2; Online Resource 1). We also detected two or more chromosome counts in the different seeds produced by one same individual.

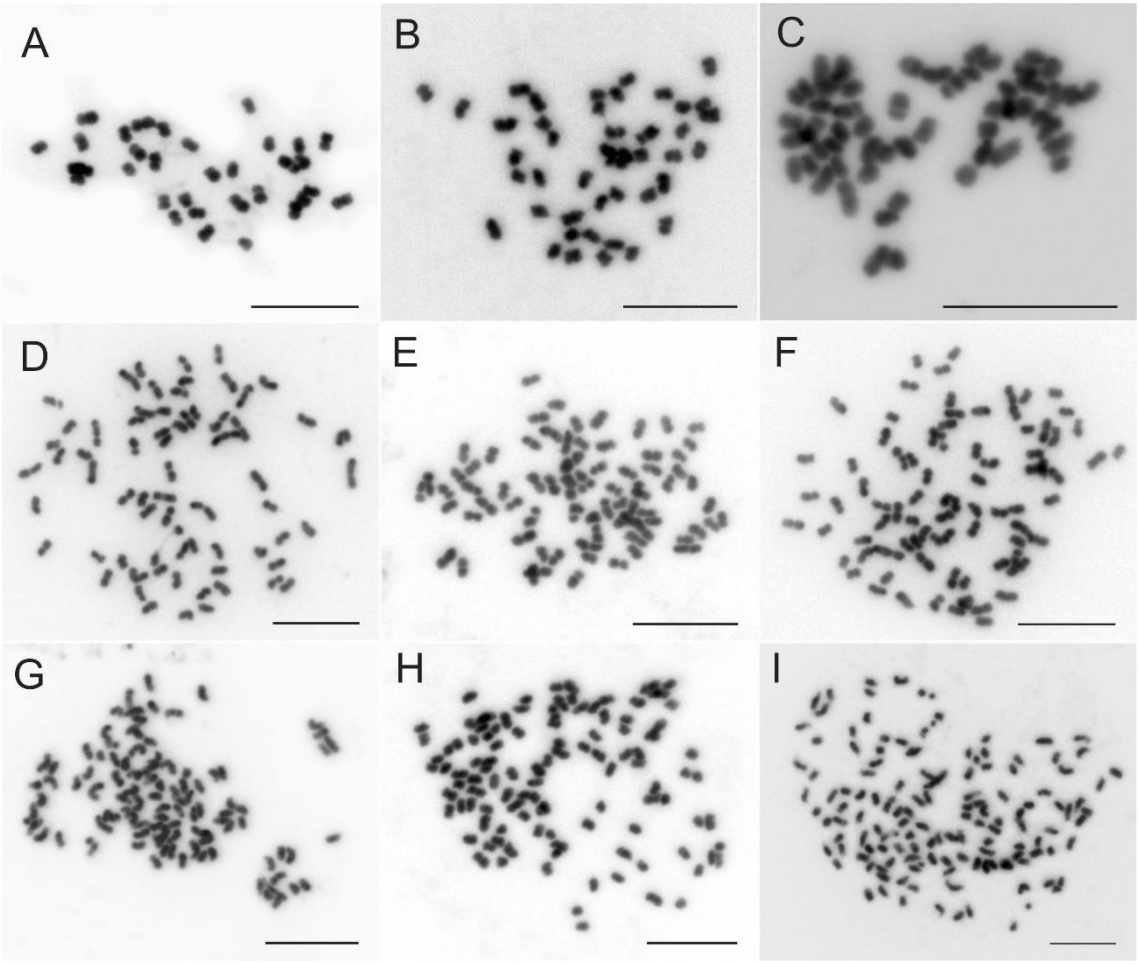


Fig. 1 Representative mitotic metaphases of *Psidium cattleyanum*: 2n = 3x = 33 (A), 2n = 4x = 44 (B), 2n = 5x= 55 (C), 2n = 6x = 66 (D), 2n = 7x = 77 (E), 2n = 8x = 88 (F), 2n = 9x = 99 (G), 2n = 10x = 110 (H) and 2n = 12x = 132 (I). Scale bar = 10 µm

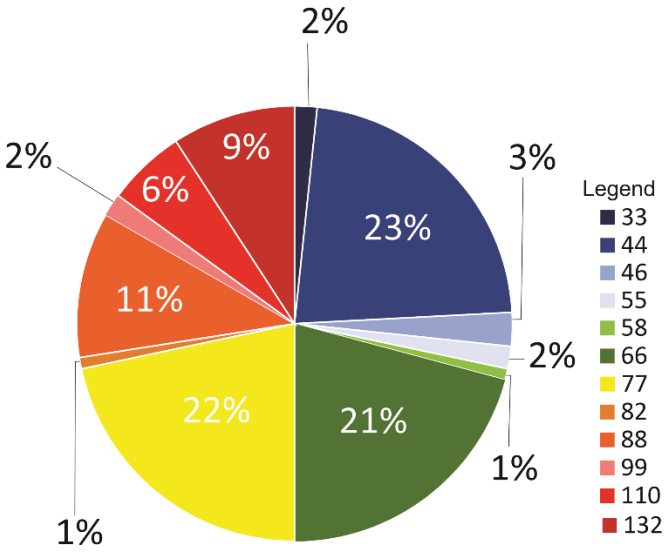


Fig. 2 Frequency graph (in percentage) of the 106 samples of *Psidium cattleyanum* analyzed. Each color represents a cytotype.

The geographic distribution of *P. cattleyanum* cytotypes for 27 localities is shown in Fig. 3. No specific geographic distribution pattern was observed for the cytotypes. There was no relationship between ploidy level and latitude (ρ = -0.33, p = 0.99). Visually, we observed a tendency to increase the frequency of larger ploidies at lower latitudes. We also observed more than one cytotype in some locations (mixed-ploidy populations) in Porto Alegre (RS), Santa Maria (RS), Dr. Pedrinho (SC), Curitiba (PR), S. L. do Paraitinga (SP), Paraty- Pedra da Macela (RJ), Itatiaia (RJ), and Trancoso (BA).

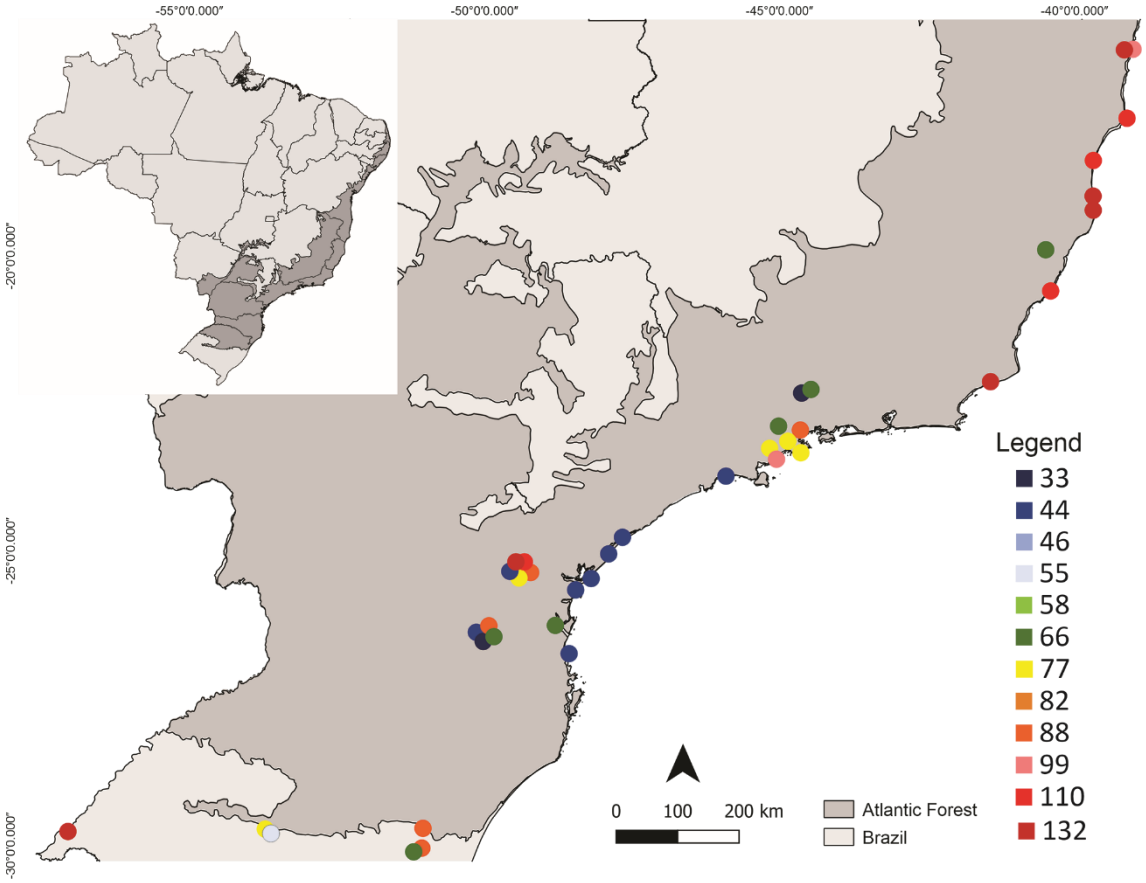


Fig. 3 Cytotype distribution of *Psidium cattleyanum* in Brazil. Each color represents a cytotype. Some localities have more than one cytotype.

Correlation between ploidy level and fruit color

We verified that the ploidy level was not related to the color of the fruit in *P. cattleyanum*. The same chromosome number was recorded in both red fruit and yellow fruit plants. The correlation test between ploidy and fruit color resulted in a weak correlation (ρ = 0.28, p< 0.01; Online Resource 1).

CMA/DAPI banding

In all cytotypes analyzed it was possible to identify bands with a strong $CMA^{+}/DAPI^{-}$ signal located in the region near to the centromere (Fig. 4; Table 1). In some preparations, we noticed that the chromosome arm with strong CMA^{+} band was away from the telomere region. The number of CMA^{+} bands increased as the number of monoploid sets increased, i.e., three bands in 3x cytotypes up to 12 bands in 12x cytotypes (Table 2; Fig 7. B). It was difficult to visualize the bands only in the 11x cytotype. Some $CMA^{-}/DAPI^{+}$ bands were observed, but these bands were inconspicuous, so we decided to omit them.

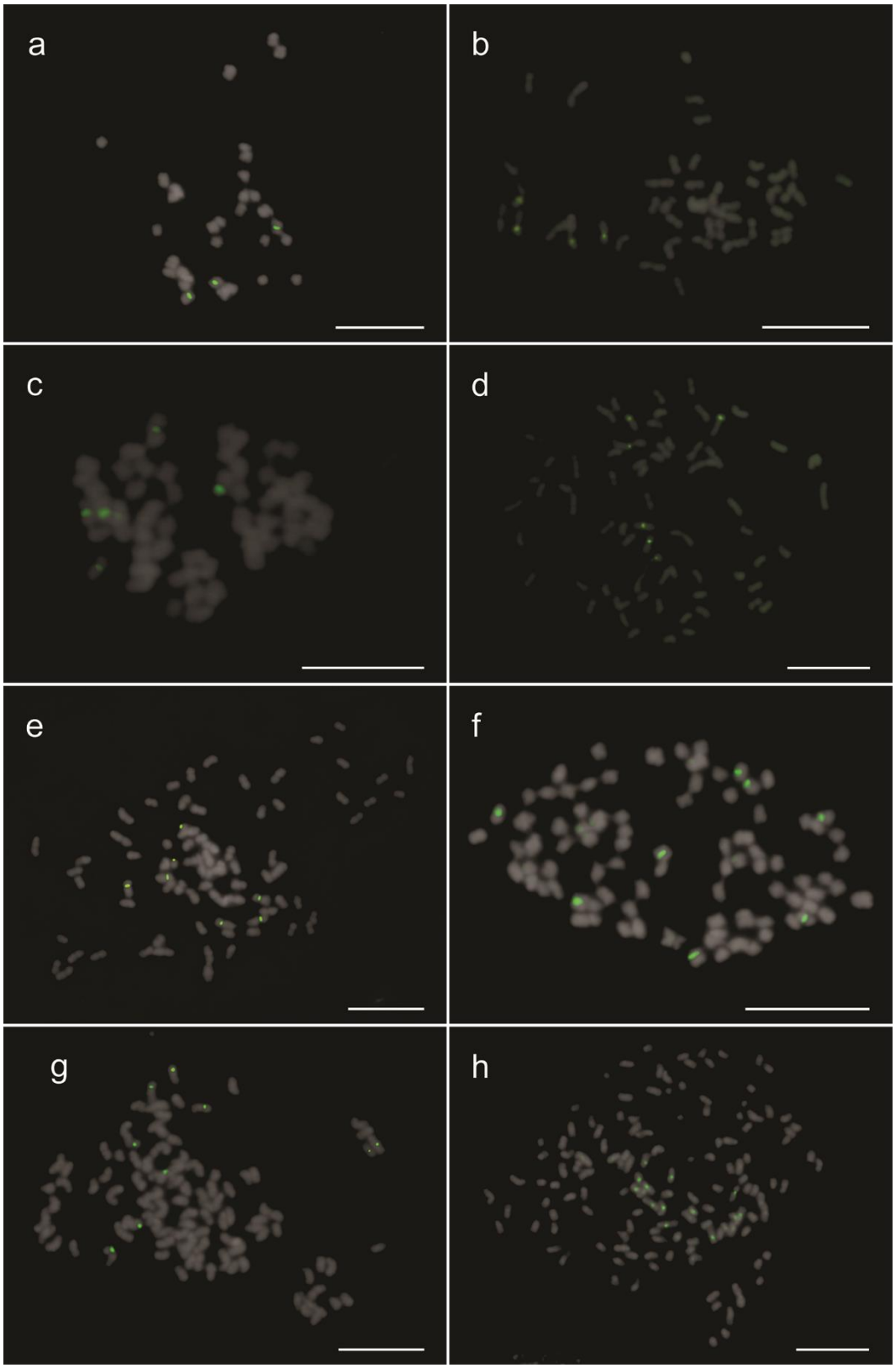


Fig. 4 Fluorescent CMA/DAPI banding in mitotic metaphase of eight cytotypes of *Psidium cattleyanum*: 2n = 3x = 33 (A), 2n = 4x = 44 (B), 2n = 5x = 55 (C), 2n = 6x = 66 (D), 2n = 7x = 77 (E), 2n = 8x = 88 (F), 2n = 9x = 99 (G) and 2n = 12x = 132 (H). CMA^{+}/$DAPI^{-}$ bands are green marks. Scale bar= 10 µm.

Molecular cytogenetics

The six cytotypes investigated with FISH, 2n= 33, 44, 66, 77, 88 e 132 (Fig. 5; Table 2), showed 5S rDNA sites in regions near to centromere (proximal). The 18S rDNA sites were also located in proximal region and coincided with the CMA^{+}/$DAPI^{-}$ bands. An exception to the observed pattern was an individual from Pontal do Sul (PR) (2n = 4x = 44; Fig. 5.B), in which not all CMA^{+}/$DAPI^{-}$ bands were coincident with the rDNA 18S signal location. Similar to the CMA^{+} bands, the number of 18S and 5S rDNA sites increased among the cytotypes according to the ploidy level (Fig. 5 and Fig. 7.C and Fig. 7.D; Table 2), suggesting a multiplication of x = 11 genome. The exception was the individual 2n = 12x = 132 that exhibit 12 18S rDNA sites and only 10 5S rDNA sites. The 18S and 5S rDNA sites occur in different chromosomes, except in dodecaploids (2n = 12x = 132), where the sites are syntenic.

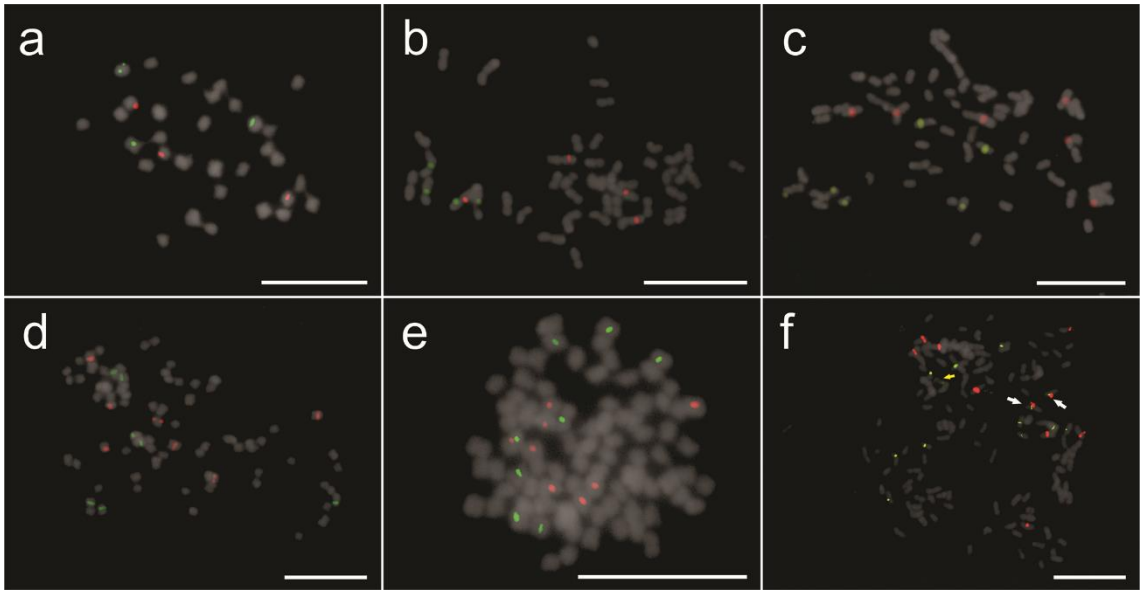


Fig. 5 Representative metaphases of six cytotypes of *Psidium cattleyanum*: (A) 2n = 3x = 33, (B) 2n = 4x= 44, (C) 2n = 6x = 66; (D) 2n = 7x = 77, (E) 2n = 8x = 88, and (F) 2n = 12x = 132 chromosomes. Green marks are 18S rDNA sites, and red marks are 5S rDNA sites. Scale bar = 10 µm. White arrows indicate the adjacent 5S and 45S rDNA sites and yellow arrow indicates a distended 45S rDNA site.

Table 2 Numbers of CMA^{+}/$DAPI^{-}$ bands and ribosomal sites for eight and six cytotypes of *Psidium cattleyanum*, respectively.

Cytotype	$CMA^{+}/DAPI^{-}$	18S	5S
2n= 33	3	3	3
2n=44	4	4	4
2n=55	5	-	-
2n=66	6	6	6
2n=77	7	7	7
2n=88	8	8	8
2n=99	9	-	-
2n=132	12	12	10

Flow cytometry analysis

Genome size estimates for 58 accessions (43 new measures and 15 from the literature) ranged from 1.53 to 6.03 pg (2C value- Online Resource 2). At the lowest ploidy levels (4x), the average DNA content was 2.18 pg, while in larger ploidy levels (12x) was 5.64 pg (Table 3). We observed a gradual and linear increase in the DNA content following the ploidy level up to 8x cytotype (Fig. 6 and Fig. 7.A), evidenced by a strong positive correlation ($\rho = 0.92$, $p < 0.01$). We noticed a downward trend in DNA content in the in larger ploidies (11x and 12x; Fig. 7A).

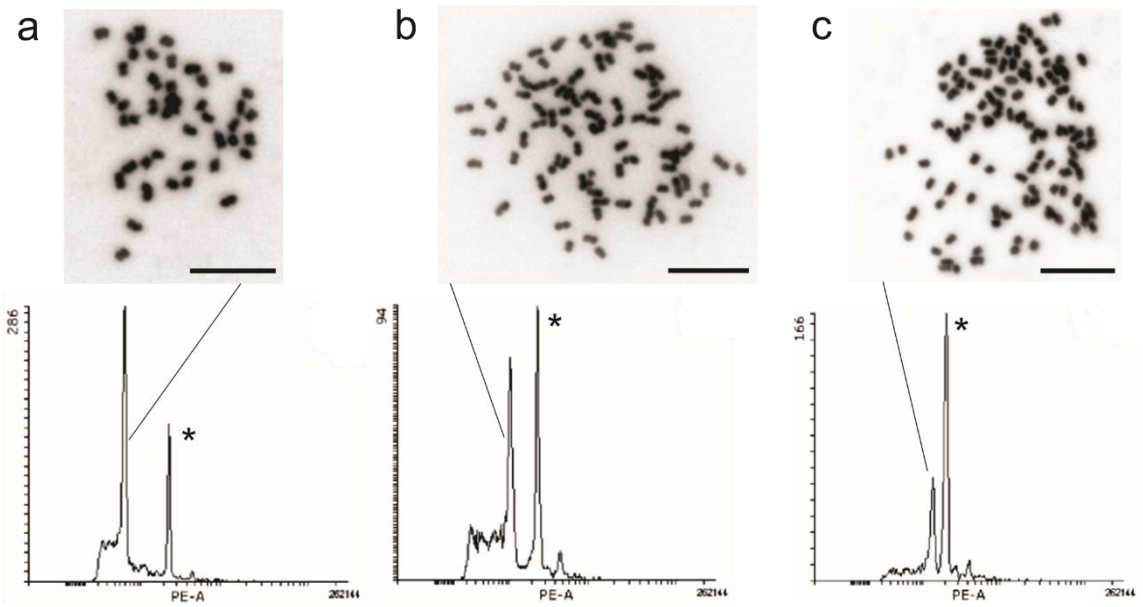


Fig. 6 6 Increased of DNA content according to the level of ploidy in cytotypes of Psidium cattleyanum. An approximation of the peak of Psidium cattleyanum (left peak in graphs)

in relation to the peak of the Pisum sativum (right peak) is observed with the increase in the chromosome number (a- 2n = 44, b- 2n = 88, c- 2n = 110). Scale bar- 10 µm

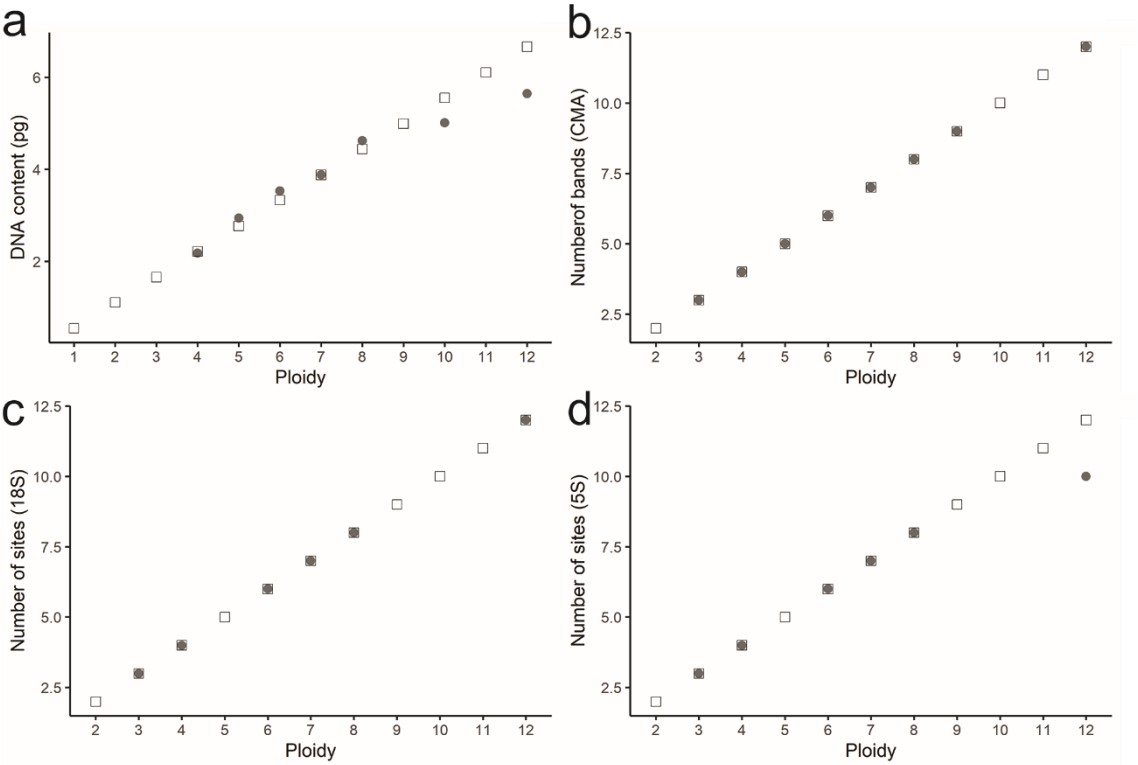


Fig. 7 Comparison between ploidy (x) and (A) DNA content, (B) numbers of CMA^{+} bands, (C) 18S rDNA sites and (D) 5S rDNA sites in *Psidium cattleyanum*. Squares represent the expected values according to a linear increase in ploidy level and gray circles represent observed data.

Table 3 Average of DNA 2C-value for 58 accessions of *Psidium cattleyanum* and standard deviation (SD).

Ploidy level (x)	Average of DNA 2C value (pg)+SD	Number of accessions
4x	2.18± 0.31	18
5x	2.94± 0.08	2
6x	3.53 ± 0.82	7
7x	3.88 ± 0.08	18
8x	4.62 ± 0.42	7

10x	5.01 ± 0.79	4
12x	5.64 ± 0.34	2

Discussion

This study is the first step towards a better understanding of the polyploidy evolution of *P. cattleyanum* polyploid complex, a complete karyotype characterization by using chromosome banding (CMA/DAPI), 5S and 18S ribosomal DNA physical mapping, and the genome size assessment. *Psidium cattleyanum* has been frequently described as a polyploid species, but our study reveals a larger cytogenetic variability in this taxon than previously known. Once the diploid level (2n=22) is rare in *Psidium* (Marques et al. 2016), the occurrence of the polyploid series in *P. cattleyanum* indicates that polyploidy plays an important role in the evolution and diversification of *Psidium*. The detection of genome downsizing and a structural variation in chromosomes suggests that a diploidization process is underway in cytotypes from this taxon. We also confirmed that specific ploidy level is not related to the fruit color, just as there is no apparent relationship between ploidy level and geographic distribution. We discuss these results below, explaining the patterns of chromosome number, bands CMA/DAPI, physical mapping of 18S and 5S rRNA genes and genome size found herein.

Psidium cattleyanum: a polyploid complex

Our results confirmed the occurrence of the polyploid series in *P. cattleyanum*, derived from x=11, following the base chromosome number for Myrtaceae (Atchinson 1947) and Myrteae tribe (Costa 2009). This study includes new occurrences for *P. cattleyanum*, 2n= 33, 99, 110 e 132, totaling 13 registered cytotypes (Atchison 1947; Hirano and Nakazone 1969; Singhal et al. 1985; Raseira and Raseira 1996; Costa and Forni-Martins 2006; Costa 2009; Vázques 2014; Souza et al. 2014; Tuler et al. 2019; Machado et al. 2020, and this study). The variation in chromosome number found here is probably the highest known for this genus (IPCN Chromosome Reports; Rice et al. 2015). The occurrence of cytotypes and polyploid species is a frequent event in clade Myrtoideae (Costa and Forni-Martins 2006; Costa 2009; Marques et al. 2016; Tuler et al. 2019; Machado et al. 2020). The diploid level remains unknow, so that 2n = 3x = 33 reported in this study is the smallest chromosome number for the species, while 2n = 12x = 132 is

the highest ploidy level reported to the entire genus. Previously, Éder-Silva et al. (2007) had reported 2n = 98 for *Psidium arborerum* Vell., also a *P. cattleyanum* synonym (Tuler et al. 2021).

The accessions with 2n = 77, 66, 44, and 88 chromosomes were the most common in number of records. According to Kolár et al. (2017), most of the mixed-ploidy species have two or three cytotypes. The fact that *P. cattleyanum* has more than three ploidies reinforces the hypothesis that polyploidy is the basis of diversification in *Psidium.* Mixed ploidy populations detected in *P. cattleyanum* is according with other studies for polyploids in Brazil, such as *Z. mackayi* (Gomes et al. 2008). Marhold et al. (2010) analyzed species of *Cardamine* (Brassicaceae), the species *C. yezoensis* has six cytotypes in Japan, predominantly 2n = 6x = 48, in addition to 8x, 10x, 11x and 12x. The latter authors found three mixed populations, with two to four cytotypes across the range studied. The detection of natural polyploidy series (e.g., Kharrat-Souissi et al. 2012; Reis et al. 2014; Tacuatiá et al. 2018; Afonso et al. 2020) is an important proof to support the idea that polyploidization is an ongoing process in evolution of plants (Soltis et al 2011; Baduel et al. 2019).

The different chromosome numbers in seeds observed from the same individual are probably associated with the union of unreduced gametes. Raseira and Raseira (1996) detected many meiotic irregularities and facultative apomixis in *P. cattleyanum* cytotypes, while Souza-Perez and Speroni (2017) observed reproduction apomixis in Uruguayan populations. Apomictic reproduction could also explain the large number of odd cytotypes in this species (Rye 1979), and why the most common record is 2n = 77. In *Eugenia* L. (Myrtaceae), some populations analyzed by Silveira et al. (2016) was 2n = 3x = 33 and 2n = 5x = 55. All populations of *Psidium* species analyzed by Tuler et al. (2019) were even-numbered cytotypes.

The lack of correspondence between ploidy level and geographic location (Latitude) is possibly because *P. cattleyanum* has been widely spread by humans. An analysis excluding individuals collected in cultivation may help to clarify this pattern. Additionally, the cytotypes 2n = 66, 77, and 88 has wide distribution range. *P. cattleyanum* was introduced in Hawaii and in different parts of the world, becoming a serious invader (Wilker 2007). The presence of several polyploid cytotypes and reproductive strategies probably facilitates the establishment and local adaptation of *P. cattleyanum*. The trend different spatial distribution patterns of polyploid and diploid populations in Myrtaceae was demonstrated by Silveira et al. (2016) and Tuler et al.

(2019) to *Eugenia* and *Psidium*, respectively. The authors of both studies analyzed natural populations excluding cultivated individuals. Other studies, such as Marhold et al. (2010), also showed the lack of relationship between different ploidies and geographic distribution in *C. yezoensis* (Brassicaceae) in Japan. However, cytotypes can show habitat preferences, as in *Jacobaeae vulgaris* (Asteraceae) (Hodálová et al. 2015) and *Eugenia punicifolia* (Kunth) DC. (Silveira et al. 2016).

The distribution trend of *P. cattleyanum* polyploid cytotypes also revealed a latitudinal gradient contrary to the pattern observed by Rice et al. (2019). In *P. cattleyanum*, the frequency of cytotypes with higher chromosomal numbers increased towards the equator (Fig. 3), showing that, when dealing with polyploid groups, the generation of new empirical data can challenge pre-existing concepts, especially in regions with high diversity, such as the Neotropical region. Future climatic niche modeling analyses of natural populations are necessary to understand the current distribution patterns of *P. cattleyanum* and may help explain the success of this great variation of cytotypes.

We also observed that there is no correspondence between ploidy level and color of fruit. It is in accordance with previous studies, in which the same chromosome number was found in plants with red and yellow fruits in Brazil (Raseira and Raseira 1996; Costa 2009, and this study), Uruguay (Vázques 2014), and in Hawaii (Hirano and Nakasone 1969).

Bands CMA/DAPI and rDNA 18S and 5S sites increase with ploidy level

In the seven *P. cattleyanum* cytotypes analyzed here, the $CMA^{+}/DAPI^{-}$ bands increased linearly with ploidy level. The same pattern was observed for the 2n = 88 and 2n = 77 cytotypes of Uruguayan populations (Vázques, 2014). The author detected eight and seven $CMA^{+}/DAPI^{-}$ proximal bands in individuals with 2n = 88 and 2n = 77, respectively. Indeed, in species with small chromosomes, such as *P. cattleyanum* (<3 um, Costa 2009; Vázques 2014), heterochromatic bands are preferably located in proximal regions (Guerra 2000).

In this study, all $CMA^{+}/DAPI^{-}$ bands were colocalized to rDNA 18S sites detected by FISH (except an 4x individual from Pontal do Sul), confirming the patterns discussed by Guerra (2000). The most characteristic heterochromatic regions detected with

fluorochromes (generally, CMA^+ and $DAPI^-$) were HC-NOR regions - which rDNA genes are located (Guerra 2000).

As the $CMA^+/DAPI^-$ bands, in *P. cattleyanum*, the 5S and 18S rDNA sites increase linearly with ploidy level (except 10 sites of 18S rDNA in 12x level). Similar rDNA patterns were observed in other taxa such as *Cenchrus ciliaris* L. (Poaceae) (Kharrat-Souissi et al 2012), *Lippia alba* (Reis et al 2014), and *Saccharum* species (Poaceae) (D'Hont, 1998). Considering the sugar cane (Saccharum ssp.), the current varieties also show a variety of chromosomal numbers (2n = 110 to 2n = 130), and the number of rDNA sites 18S and 5S increases according to the ploidy level suggesting euploid changes of x = 10 (D'Hont, 1998). Polyploidization events results in genetic and epigenetic changes, e.g., increase of number of rDNA sites, until full diploidization (Parisod et al. 2010).

Diploidization processes also explain the loss of two rDNA 5S sites in the cytotype 12x. Diploidization is a comprehensive term, including a variety of processes that transform a polyploid genome into a (pseudo) diploid (Mandáková et al. 2016). The transformations in the genome after the formation of a polyploid results is the elimination of specific genes or DNA sequences, including copies of rDNA18-5.8-26S and/or 5S, as has been reported in Triticeae (Poaceae), *Nicotiana* L. (Solanaceae), *Brassica* L. (Brassicaceae) and *Glycine* Willd. (Fabaceae) (reviewed by Wendel 2000). In general, when polyploidy is recent, it results in twice as many sites, but older polyploids often exhibit a number of "diploidized" rDNA sites (Roa and Guerra 2012).

Another evidence of structural rearrangements in the genome of the cytotype 2n = 132 was observed in adjacent rDNA 18S and 5S sites. In other *P. cattleyanum* cytotypes, all the rDNA 18S and 5S sites are located in different chromosomes. The idea that rDNA sites can change its number and position in the genome is well documented (Roa and Guerra 2012). Different mechanisms may be involved in the variation of the number and position of rDNA sites, both intra and interspecific, such as association with mobile elements, different chromosomal rearrangements, unequal genetic exchange, and gene transposition (Brown and O'Neill 2010; Roa and Guerra 2012; Roa and Guerra 2015).

A possible explanation to the separation of chromosome arms that carry HC-NOR bands ($CMA^+/DAPI^-$ proximal bands, corresponding to rDNA18S sites) in *P. cattleyanum* cytotypes is the presence of distended 45S rDNA sites regions. This phenomenon has also been observed in chromosomal studies for *P. cattleyanum* in Uruguayan populations (Vázques 2014). The occurrence of distended 45S rDNA sites regions affect genome

organization, and can lead to new chromosomal rearrangements (Brown and O'Neill 2010), like dysploidy. These events might help explain the chromosome counts in *P. cattleyanum* that do not follow the basic number of x = 11 (2n = 46, 48, 58 and 82) (Souza et al. 2014).

The fact that *P. cattleyanum* 18S rDNA sites were distributed in the proximal regions in all cytotypes differs from other taxa. Generally, the location of rDNA 18S-5.8S-26S sites follows a common distribution pattern in terminal regions of the short arm of chromosomes (Roa and Guerra 2012). Ribosomal DNA sites localized in proximal regions have already been shown in the genera *Arachis* L., *Vicia* L., *Trifolium* L., and *Lens* Mill. (Fabaceae) (Roa and Guerra 2012). Several factors can influence the dispersion of rDNA sites from the terminal regions of short arm to other chromosomal regions, such as translocations and association with transposable elements. The rDNA 5S sites are preferentially distributed in the proximal regions in species with small chromosomes (< 3 µm) (Roa and Guerra 2015), as *P. cattleyanum*.

According to current knowledge, the diploid state is known in a low percentage of species of the genus *Psidium* (Marques et al. 2016) and in some species only polyploid records are known, as is the case of *P. cattleyanum*. Polyploidy plays an important role in the diversification of the genus *Psidium*. Considering the cytogenetic findings of our study, *P. cattleyanum* is a good candidate model to investigate genomic rearrangements associated to polyploidization events, and the influence of polyploidy in evolution of the species.

DNA content of P. cattleyanum increases with ploidy level, but not linearly

As the CMA/DAPI bands and rDNA sites, we observed the increase of DNA content as the number of monoploid sets increased, but it was not exactly linear, especially in the cytotypes 2n = 110 and 132. Similar results were found in the autopolyploid complex *L. alba* (Reis et al. 2014), with the DNA content decreasing in cytotypes with higher ploidy. This pattern could be associated with genome downsizing phenomenon. Several authors have already reported the genome downsizing, including neotropical species such as *Sisyrinchium* L. (Iridaceae) (Tacuatiá et al. 2017). In genome downsize, the 2C mean value in higher ploidy levels is slightly smaller than expected, possibly due to the loss involved in the "accommodation" of the polyploid genome (Leitch and Bennett 2004).

Costa et al. (2008), studying the variation of the genome size in fleshy fruit of Myrtaceae, concluded that, for *Psidium*, the values of 2C were directly related to the basic number of x = 11 and increased linearly according to the level of ploidy (2n = 4x = 44 for *P. acutangulum* DC. and *P. cattleyanum*). Souza et al. (2014) performed counts and genome size estimates for eight species of the genus, including *P. cattleyanum*, concluding that the intraspecific variation in the value of 2C shows a direct relationship with the level of ploidy, however the author does not report whether this increase is linear or not. For the *P. cattleyanum* cytotypes from Uruguay, Vázques (2014) also reported that the value of 2C has a direct relation to the ploidy level determined for cytotypes 2n = 77 and 2n = 88. Mazzella et al. (2016), carrying out cross-pollination experiments in Uruguayan populations estimated five levels of ploidy (5x, 6x, 7x, 8x and 9x) through the quantification of DNA by flow cytometry in the seedlings obtained from the crosses. The authors reported that it is possible to deduce the level of ploidy by observing the results of cytometry, but the ratio of DNA content/level of ploidy is not specified. Angiosperms tend to present small genome size (Leitch et al. 1998), and *P. cattleyanum* follows this pattern (most of accessions had small and intermediate genome size).

Statistical analysis showed a strong correlation ($\rho = 0.92$, $p < 0.01$) between DNA content and ploidy level, with a tendency to increase DNA content when the ploidy level increases at least up to 8x (Fig. 7). Thus, flow cytometry can be used as a tool to indirectly estimate polyploidy in *P. cattleyanum*, especially when the ploidy level is not very high. Measuring the nuclear DNA content by flow cytometry is more accurate than measuring chromosomes and is a faster process when compared with classical cytogenetics techniques (Dolezel et al. 2007).

Cytogenetic data and insights into the origin of P. cattleyanum cytotypes

Considering the genome repetition of x = 11, indicated by the linear increase in the number of CMA^+ bands, rDNA sites and DNA content in relation to the increase of ploidy level, we suggest an autopolyploid origin hypothesis for *P. cattleyanum*. A possible autopolyploid origin had already been suggested by Vázques (2014) for Uruguayan populations. And our results support this hypothesis for Brazilian populations too. The 45S rDNA sites from one of the parental genomes prevail over the other in many allopolyploids (Chiarini et al. 2014).

Singhal et al. (1985) analyzed meiotic behavior in *P. coriaceum* (synonymous of *P. cattleyanum*) and detected in 2n = 77 cytotype a high frequency of univalent formation. A strong indicator of autopolyploid is the presence of multivalent (even-numbered) and univalent (odd-numbered) formation during the meiotic process (Gregory and Mable 2005; Tate et al. 2005). In addition, *P. cattleyanum* has low rates of pollen viability and fertility (Singhal et al. 1985; Raseira and Raseira 1996; Souza-Pérez and Speroni 2017). The linear increase in the number of CMA^+ bands, rDNA sites, DNA content, high rates of pollen infertility and formation of univalent in meiosis are suggestive of autopolyploid origin.

Autopolyploids are difficult to identify because sometimes are indistinguishable from their diploid progenitors but evolve as cryptic evolutionary units (Soltis et al. 2009; Baduel et al. 2019). Cytogenetic data providing insights into the origin of cytotypes was demonstrated in *L. alba* (Reis et al. 2014) and *Solanum elaeagnifolium* (Solanaceae) (Chiarini et al. 2014) for autopolyploid formation, and in *Linum suffruticosum* L. complex (Linaceae) (Afonso et al. 2020) for allopolyploid formation. Due to the difficult to distinguish between autopolyploid and allopolyploid formation, further evaluation of meiotic process of all ploidy levels, phylogenetic and phylogeographic studies, and niche modelling analyses are necessary to confirm this hypothesis of *P. cattleyanum* polyploid complex origin.

Conclusions

In this study, we confirmed the occurrence of polyploid series for *P. cattleyanum*, derived from basic number x = 11 and added new cytogenetic records. Investigating cytological aspects of *P. cattleyanum* may be a key to understand the origin and the evolution of *Psidium*. Moreover, our findings add the knowledge about the evolution of tropical polyploid plants. Our study is the first step towards the understanding of current distribution patterns and formation of *P. cattleyanum* polyploid complex. We pointed out genomic rearrangements using CMA/DAPI banding, physical mapping using FISH and genome size analyses, which shows that polyploidy is a dynamic process. We demonstrated that genome size assessment is a useful tool for identifying cytotypes of *P. cattleyanum*. In addition, the high variation in chromosome number found shows that *P. cattleyanum* is an interesting model to study mechanisms that are involved in the evolution of natural polyploidy complexes.

References

Afonso A, Loureiro J, Arroyo J, Olmedo-Vicente E, Castro S (2020) Cytogenetic diversity in the polyploid complex *Linum suffruticosum* sl (Linaceae). J Bot J Linn Soc. https://doi.org/10.1093/botlinnean/boaa060

Atchison E (1947) Chromosome numbers in the Myrtaceae. Am J Bot 34: 159-164

Baduel P, Bray S, Vallejo-Marin M, Kolár F, Yant L (2018) The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. Front Ecol Evol. https://doi.org/10.3389/fevo.2018.00117

Brown JD, O'Neill RJ (2010) Chromosomes, conflict, and epigenetics: Chromosomal speciation revisited. Annu Rev Genomics Hum Genet. https://doi.org/10.1146/annurev-genom-082509-141554

Chiarini FE (2014) Variation in rDNA loci of polyploid *Solanum elaeagnifolium* (Solanaceae). N Z J Bot. https://doi.org/10.1080/0028825X.2014.888087

Costa IR, Forni-Martins ER (2006) Chromosome studies in Brazilian species of *Campomanesia* Ruiz & Pávon and *Psidium* L. (Myrtaceae Juss.). Caryologia. https://doi.org/10.1080/00087114.2006.10797891

Costa IR (2009) Estudos evolutivos em Myrtaceae: aspectos citotaxonômicos e filogenéticos em Myrteae, enfatizando *Psidium* e gêneros relacionados. Thesis, Universidade Estadual de Campinas

Costa IR, Dornelas MC, Forni-Martins ER (2008) Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. Plant Syst Evol. https://doi.org/10.1007/s00606-008-0088-x

D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome. https://doi.org/10.1139/g98-023

Dodsworth S, Chase MW, Leitch AR (2016) Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? J Bot J Linn Soc. https://doi.org/10.1111/boj.12357

Dolezel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc. https://doi.org/10.1038/nprot.2007.310

Dolezel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, Obermayer R (1998) Plant genome size estimation by flow cytometry: inter-laboratory comparison. Ann Bot. https://doi.org/10.1093/oxfordjournals.aob.a010312

Ebihara A, et al (2005) Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the Vandenboschia radicans complex (Hymenophyllaceae) in Japan and adjacent areas. Am J Bot. https://doi.org/10.3732/ajb.92.9.1535

Éder-Silva E, Felix LP, Bruno RLA (2007) Citogenética de algumas espécies frutíferas nativas do nordeste do Brasil. Rev Bras Frutic. http://dx.doi.org/10.1590/S0100-29452007000100024

Franzon, RC, Campos LDO, Proença CEB, Sousa-Silva JC (2009) Araçás do gênero psidium: principais espécies, ocorrência, descrição e usos. Embrapa: Planaltina

Gao J (2019) Dominant plant speciation types. A commentary on: ‘Plant speciation in the age of climate change’. Ann Bot. https://doi.org/10.1093/aob/mcz108

Gomes SSL, et al (2018) Genome size and climate segregation suggest distinct colonization histories of an orchid species from Neotropical high-elevation rocky complexes. Biol J Linn Soc. https://doi.org/10.1093/biolinnean/bly065

Gregory TR, Mable BK (2005) Polyploidy in animals. In: Gregory TR (ed) The evolution of the genome, 1st edn. Academic Press, San Diego, pp 427-517

Guerra M (2000) Patterns of heterochromatin distribution in plant chromosomes. Genet Mol Biol. https://doi.org/10.1590/S1415-47572000000400049

Guerra M, Souza MD (2002) Como observar cromossomos: um guia de técnicas em citogenética vegetal, animal e humana. FUNPEC, Ribeirão Preto

Hodálová I, Mereďa P, Kučera J, Marhold K, Kempa M, Olšavská K, Slovák M (2015) Origin and systematic position of *Jacobaea vulgaris* (Asteraceae) octoploids: genetic and morphological evidence. Plant Syst Evol. https://doi.org/10.1007/s00606-014-1163-0

Hirano RT, Nakazone HY (1969) Chromosome numbers of ten species and clones in the genus *Psidium*. J Am Soc Hortic Sci 94: 83-86.

Husband BC, Baldwin SJ, Suda J (2013) The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: Leitch IJ, Greilhuber J, Dolezel J, Wendel JF (eds) Plant genome diversity, 2nd edn. Springer-Verlag, Wien, pp 255–276

Kharrat-Souissi A, Siljak-Yakovlev S, Pustahija F, Chaieb M (2012) Physical mapping of 5S and 18S-5.8 S-26S RNA gene families in polyploid series of *Cenchrus ciliaris* Linnaeus, 1771 (Poaceae). Comp Cytogenet. https://doi.org/10.3897/CompCytogen.v6i3.3380

Kitamura S, Inoue M, Shikazono N, Tanaka A (2001) Relationships among *Nicotiana species* revealed by the 5S rDNA spacer sequence and fluorescence *in situ* hybridization. Theor Appl Genet 103: 678-686

Kolár F, Čertner M, Suda J, Schönswetter P, Husband BC (2017) Mixed-ploidy species: progress and opportunities in polyploid research. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2017.09.011

Leitch IJ, Bennett MD (2004) Genome downsizing in polyploid plants. Biol J Linn Soc. https://doi.org/10.1111/j.1095-8312.2004.00349.x

Leitch IJ, Chase MW, Bennett MD (1998) Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. Ann Bot. https://doi.org/10.1006/anbo.1998.0783

Levin DA. 2019. Plant speciation in the age of climate change. Ann Bot. https://doi.org/10.1093/aob/mcz108

Machado RM, de Oliveira FA, de Matos Alves F, de Souza AP, Forni-Martins ER (2020) Population Genetics of Polyploid Complex *Psidium cattleyanum* Sabine (Myrtaceae): Preliminary Analyses Based on New Species-Specific Microsatellite Loci and Extension to Other Species of the Genus. Biochem Genet. https://doi.org/10.1007/s10528-020-10002-1

Mandáková T, Gloss AD, Whiteman NK, Lysak MA (2016) How diploidization turned a tetraploid into a pseudotriploid. Am J Bot. https://doi.org/10.3732/ajb.1500452

Marhold K, Kudoh H, Pak JH, Watanabe K, Španiel S, Lihová J (2010) Cytotype diversity and genome size variation in eastern Asian polyploid Cardamine (Brassicaceae) species. Ann Bot. https://doi.org/10.1093/aob/mcp282

Marques A, Tuler AC, Carvalho CR, Carrijo T, da Silva Ferreira MF, Clarindo W (2016) Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). Comp Cytogenet. https://doi.org/10.3897/CompCytogen.v10i1.6462

Mazzella C, Speroni G, Pritsch C, Souza-Pérez M, Bonifacino M, Vázquez S, Vaio M, Trujillo C, Cabrera D, Vignale B (2016) Arazá, Especie Frutal De Interés: Sistema Reproductivo, Distribución Natural Y Citotipos en Urugay. J Basic Appl Genet 27: 144

Normand F, Habib R (2001) Phenology of strawberry guava (Psidium cattleianum) in ReÂunion Island. J Hortic Sci Biotechnol. https://doi.org/10.1080/14620316.2001.11511407

Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. New phytol. https://doi.org/10.1111/j.1469-8137.2009.03142.x

Raseira MCB, Raseira A (1996) Contribuição ao estudo do araçazeiro. Embrapa, Pelotas

Reis AC, et al (2014) Lippia alba (Verbenaceae): A new tropical autopolyploid complex? Am J Bot. https://doi.org/10.3732/ajb.1400149

Rice A, et al (2015) The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. New Phytol. https://doi.org/10.1111/nph.13191

Rice A, et al (2019). The global biogeography of polyploid plants. Nat Ecol Evol. https://doi.org/10.1038/s41559-018-0787-9

Roa F, Guerra M (2012) Distribution of 45S rDNA sites in chromosomes of plants: Structural and evolutionary implications. BMC Evol Biol. https://doi.org/10.1186/1471-2148-12-225

Roa F, Guerra M (2015) Non-Random Distribution of 5S rDNA Sites and Its Association with 45S rDNA in Plant Chromosomes. Cytogenet Genome Res. https://doi.org/10.1159/000440930

Rye B (1979) Chromosome number variation in the Myrtaceae and its taxonomic implications. Aust J Bot 27: 547-573.

Schwarzacher TP, Heslop-Harrison P (2000) Pratical in situ hybridization. Bios Scientific Publishers, Oxford

Schweizer D (1976) Reverse fluorescent chromosome banding with chromomycin and DAPI. Chromosoma 58: 307-324

Silveira RM, Machado RM, Forni-Martins ER, Verola CF, Costa IR (2016) Environmental variations drive polyploid evolution in neotropical Eugenia species Myrtaceae. Genet Mol Res. http://dx.doi.org/10.4238/gmr15048842

Singhal VK, Gill BS, Bir SS (1985) Cytology of cultivated woody species (Polypetalae). Proc. Indian Sci. Congr. Assoc 71: 143-144

Sobral MEG, Sousa LP (2007) Morfotipos do Araçazeiro (*Psidium cattleianum* Sabine) no Estado do Paraná. In: Pedrosa-Macedo JH, Dalmolin A, Smith CW (eds) O Araçazeiro: Ecologia e Controle Biológico, 1st edn. FUPEF, Curitiba, pp 19-28

Soltis DE, et al (2014) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). New Phytol. https://doi.org/10.1111/nph.12756

Soltis DE, et al (2009) Polyploidy and angiosperm diversification. Am J Bot. https://doi.org/10.3732/ajb.0800079

Souza AG, Resende LV, Lima IP, Martins LSS, Techio VH (2014) Chromosome number and nuclear DNA amount in *Psidium* app. Resistant and susceptible to *Meloidogyne enterolobii* and its relation with compatibility between rootstocks and commercial varieties of guava tree. Plant Syst Evol. https://doi.org/10.1007/s00606-014-1068-y

Souza-Pérez M, Speroni G (2017) New apomictic pathway in Myrtaceae inferred from Psidium cattleyanum female gametophyte ontogeny. Flora. https://doi.org/10.1016/j.flora.2017.06.010

Srivastava HC (1977) Cytological studies in *Psidium friedrichstalianum* Niedenzu. Cytologia. https://doi.org/10.1508/cytologia.42.395

Tacuatiá LO, et al (2017) Physical mapping of 35S rRNA genes and genome size variation in polyploid series of Sisyrinchium micranthum and S. rosulatum (Iridaceae: Iridoideae). Pl Biosystems. https://doi.org/10.1080/11263504.2016.1179691

Tate JA, Soltis DE, Soltis PS (2005) Polyploidy in plants. In: Gregory TR (ed) The evolution of the genome, 1st edn. Academic Press, San Diego, pp 371-426

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2011) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot. https://doi.org/10.1093/aob/mcr277

Tuler AC, et al (2019) Diversification and geographical distribution of *Psidium* (Myrtaceae) species with distinct ploidy levels. Trees. https://doi.org/10.1007/s00468-019-01845-2

Tuler AC, Proença CEB, Costa IR (2020) *Psidium* in Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Avaible at: <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB10855>

Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. Nat Rev Genet. https://doi.org/10.1038/nrg.2017.26

Vázques SNM (2014) *Psidium cattleyanum* Sabine y *Acca sellowiana* (Berg.) Burret (Myrtaceae): caracterización cromossómica y cariotípica em poblaciones silvestre y genótipos selecionados em programas nacionales de mejoramiento. Monography, Universidad de la República Uruguay.

Vijayakumar N, Subramanian D (1985) Cytotaxonomical studies in South Indian Myrtaceae. Cytologia. https://doi.org/10.1508/cytologia.50.513

Wendel JF (2000) Genome evolution in polyploids. In: Doyle JJ, Gaut BS (eds) Plant Molecular Evolutio, 1st edn. Springer, Dordrecht, pp 225-249

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR Protocols: A Guide to Methods and Applications, , 1st edn. Academic Press, New York, pp 315-322

Wikler C (2007) Problemática do Araçazeiro como Planta Invasora. In: Pedrosa-Macedo JH, Dalmolin A, Smith CW (eds) O Araçazeiro: Ecologia e Controle Biológico, 1st edn. FUPEF, Curitiba, pp 07-18

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. PNAS. https://doi.org/10.1073/pnas.0811575106

Online Resource 1- Direct chromosome counts for 109 samples of *Psidium cattleyanum*

Sample	2n	Locality	Reference	Color of Fruit	Cytometry code
1	66	Porto Alegre-RS-Brazil	This Study	Yellow	-
2	88	Porto Alegre-RS-Brazil	This Study	Yellow	-
3	66	Porto Alegre-RS-Brazil	This Study	Yellow	-
4	88, 132	Porto Alegre-RS-Brazil	This Study	Red	-
5	55	Santa Maria-RS-Brazil	This Study	Yellow	-
6	77	Santa Maria-RS-Brazil	This Study	Yellow	-
7	132	Uruguaiana-RS-Brazil	This Study	Yellow	-
8	77	Caxias do Sul-RS-Brazil	This Study	Red	-
9	66	Estrada para Blumenau-SC-Brazil	This Study	Yellow	-
10	44	Imbituba-Praia Redmelha-SC-Brazil	This Study	Yellow	-
11	44	Imbituba-Praia Redmelha-SC-Brazil	This Study	Yellow	-
12	44	Imbituba-Praia Redmelha-SC-Brazil	This Study	Yellow	-
13	33, 66	Dr. Pedrinho-SC-Brazil	This Study	Yellow	-
14	66	Dr. Pedrinho-SC-Brazil	This Study	Yellow	-
15	66	Dr. Pedrinho-SC-Brazil	This Study	Yellow	-
16	44, 88	Dr. Pedrinho-SC-Brazil	This Study	Red	-
17	44, 77, 132	Curitiba-PR-Brazil	This Study	Red	Curi01
18	88	Curitiba-PR-Brazil	This Study	Red	Curi02
19	77, 132	Curitiba-PR-Brazil	This Study	Red	-
20	77, 88, 132	Curitiba-PR-Brazil	This Study	Red	-
21	77	Curitiba-PR-Brazil	This Study	Yellow	Curi03

22	77	Curitiba-PR-Brazil	This Study	Yellow	Curi04
23	77, 110	Curitiba-PR-Brazil	This Study	Yellow	Curi09
24	88	Curitiba-PR-Brazil	This Study	Red	-
25	44	Paranaguá- Ilha do Mel-PR-Brazil	This Study	Yellow	-
26	44	Paranaguá- Ilha do Mel-PR-Brazil	This Study	Yellow	Im17
27	44	Paranaguá- Ilha do Mel-PR-Brazil	This Study	Yellow	Im28
28	44	Pontal do sul-PR-Brazil	This Study	Yellow	Psul09
29	77	Ubatuba-Praia da Fazenda-SP-Brazil	This Study	Yellow	-
30	77	Ubatuba-Praia da Fazenda-SP-Brazil	This Study	Yellow	-
31	77	Ubatuba-Praia da Fazenda-SP-Brazil	This Study	Yellow	-
32	77	Ubatuba-Praia da Fazenda-SP-Brazil	This Study	Yellow	-
33	77	Ubatuba-Praia da Fazenda-SP-Brazil	This Study	Yellow	-
34	66	S. L. do Paraitinga- PESM-SP-Brazil	This Study	Yellow	-
35	66	S. L. do Paraitinga- PESM-SP-Brazil	This Study	Yellow	-
36	77	S. L. do Paraitinga- PESM-SP-Brazil	This Study	Red	-
37	99	Ubatuba-Vila de Picinguaba-SP-Brazil	This Study	Yellow	-
38	44	Bertioga-SP-Brazil	This Study	Yellow	Bert06
39	44	Bertioga-SP-Brazil	This Study	Yellow	Bert20
40	44	Bertioga-SP-Brazil	This Study	Yellow	Bert23
41	44	Bertioga-SP-Brazil	This Study	Yellow	Bert24
42	44	Bertioga-SP-Brazil	This Study	Yellow	Bert25
43	44	Bertioga-SP-Brazil	This Study	Yellow	Bert30
44	44	Ilha Comprida-SP-Brazil	This Study	Yellow	Ic18
45	44	Ilha do Cardoso-SP-Brazil	This Study	Yellow	
46	44	Ilha do Cardoso-SP-Brazil	This Study	Yellow	Icar06
47	44	Ilha do Cardoso-SP-Brazil	This Study	Yellow	Icar08
48	44	Ilha do Cardoso-SP-Brazil	This Study	Yellow	-
49	44	Ilha do Cardoso-SP-Brazil	This Study	Yellow	Icar 13
50	44	Ilha do Cardoso-SP-Brazil	This Study	Yellow	-
51	66	Paraty- Pedra da Macela RJ-Brazil	This Study	Yellow	Pm01
52	66, 77	Paraty- Pedra da Macela RJ-Brazil	This Study	Yellow	-

53	77	Paraty- Pedra da Macela RJ-Brazil	This Study	Yellow	Pm02
54	77	Paraty- Praia de Trindade-RJ-Brazil	This Study	Yellow	-
55	77	Paraty- Praia de Trindade-RJ-Brazil	This Study	Yellow	Trind01
57	88	Paraty- Praia do Taquari-RJ-Brazil	This Study	Yellow	-
58	132	Macaé- PARNAJurubatiba-RJ-Brazil	This Study	Yellow	Juru01
59	132	Macaé- PARNAJurubatiba-RJ-Brazil	This Study	Yellow	-
60	132	Macaé- PARNAJurubatiba-RJ-Brazil	This Study	Yellow	-
61	33,66	Itatiaia-RJ-Brazil	This Study	Yellow	-
62	66	Itatiaia-RJ-Brazil	This Study	Yellow	-
63	110	Guarapari-ES-Brazil	This Study	Yellow	-
64	132	Pontal do Ipiranga- Linhares-ES-Brazil	This Study	Yellow	Lin01
65	132	Barra Nova- São Mateus-ES-Brazil	This Study	Yellow	-
66	132	Barra Nova- São Mateus-ES-Brazil	This Study	Yellow	-
67	110	Itaúnas-ES-Brazil	This Study	Yellow	Itau01
68	66	Santa Teresa-ES-Brazil	This Study	Yellow	-
69	66	Santa Teresa-ES-Brazil	This Study	Yellow	-
70	66	Santa Teresa-ES-Brazil	This Study	Yellow	-
71	110	Caravelas-BA-Brazil	This Study	Yellow	Ba01
72	110	Caravelas-BA-Brazil	This Study	Yellow	Ba02
73	110	Caravelas-BA-Brazil	This Study	Yellow	Ba03
74	99	Trancoso-BA-Brazil	This Study	Yellow	-
75	110	Trancoso-BA-Brazil	This Study	Yellow	Ba04
76	88	-	Atchison 1947	Red	-
77	88	-	Atchison 1948	Yellow	-
78	88	-	Atchison 1949	-	-
79	77	Hawaii	Hirano & Nakasone 1969	Red	-
80	66	Hawaii	Hirano & Nakasone 1970	Yellow	-
81	77	Ínidia	Singhal 1985	-	-
82	66, 77	Sul do RS-RS-Brazil	Raseira & Raseira 1995	Red	-

83	88	Passo Fundo-RS-Brazil	Raseira & Raseira 1996	Red	-
84	66	Itatiaia-RJ-Brazil	Costa 2009	Red	-
85	66	Pelotas-RS-Brazil	Raseira & Raseira 1996	Yellow	-
86	66	Passo Fundo-RS-Brazil	Raseira & Raseira 1996	Yellow	-
87	44	Campinas-SP-Brazil	Costa 2006	Yellow	-
88	44	Cananeia-SP-Brazil	Costa 2006	Yellow	-
89	44	Sete Barras-SP-Brazil	Costa 2006	Yellow	-
90	44	Campinas-SP-Brazil	Costa 2009	Yellow	-
91	77	Uruguay	Medina 2014	Red	-
92	88	Uruguay	Medina 2014	Yellow	-
93	77	Uruguay	Medina 2014	Yellow	-
94	88	Curitiba-PR-Brazil	Machado et al. 2020	Red	-
95	77	Curitiba-PR-Brazil	Machado et al. 2020	Yellow	-
96	77	Curitiba-PR-Brazil	Machado et al. 2020	Yellow	-
97	77	Curitiba-PR-Brazil	Machado et al. 2020	Yellow	-
98	66	Cunha-PR-Brazil	Machado et al. 2020	Yellow	-
99	44	Ingaí-MG-Brazil	Souza et al. 2015	-	-
100	46	Recife-PE-Brazil	Souza et al. 2015	-	-
101	46	Recife-PE-Brazil	Souza et al. 2015	-	-
102	46	Itumirim-MG-Brazil	Souza et al. 2015	-	-
103	55	Itumirim-MG-Brazil	Souza et al. 2015	-	-
104	58	Recife-PE-Brazil	Souza et al. 2015	-	-
105	66	Lavras-MG-Brazil	Souza et al. 2015	-	-
106	66	Lavras-MG-Brazil	Souza et al. 2015	-	-

107	66	Lavras-MG-Brazil	Souza et al. 2015	-	-
108	82	Recife-PE-Brazil	Souza et al. 2015	-	-
109	66	-	Tuler et al. 2019	-	-

Online Resource 2- DNA content (2C values- pg) data of *Psidium cattleyanum.* – = data not available; *= individuals with ploidy confirmed by direct chromosome counts in literature data. Cytometry code is according table 01.

Sample	Cytometry code	Ploidy	Chromosome number	2C values – pg	CV	Locality	Reference
1*	-	4x	44	1.05	3.9	Campinas- SP- Brazil	Costa et al. 2008
2*	-	4x	44	1.99	0.96	Ingaí-MG- Brazil	Souza et al. 2015
3	Bert06	4x	44	2.16	5	Bertioga-SP-Brazil	This study
4	Bert20	4x	44	2.18	4.5	Bertioga-SP-Brazil	This study
5	Bert23	4x	44	2.19	4.6	Bertioga-SP-Brazil	This study
6	Bert24	4x	44	2.17	4.9	Bertioga-SP-Brazil	This study
7	Bert25	4x	44	2.13	4.8	Bertioga-SP-Brazil	This study
8	Bert30	4x	44	2.15	4.9	Bertioga-SP-Brazil	This study
9	Ic18	4x	44	2.26	4.2	Ilha Comprida-SP-Brazil	This study
10	Icar06	4x	44	2.24	4.4	Ilha do Cardoso-SP-Brazil	This study
11	Icar08	4x	44	2.24	5.0	Ilha do Cardoso-SP-Brazil	This study
12	Icar 13	4x	44	2.26	5.1	Ilha do Cardoso-SP-Brazil	This study
13	Im17	4x	44	2.21	4.9	Paranaguá- Ilha do Mel-PR-Brazil	This study
14	Im28	4x	44	2.23	4.2	Paranaguá- Ilha do Mel-PR-Brazil	This study
15	Psul09	4x	44	2.25	4.7	Pontal do sul-PR-Brazil	This study
16*	-	4x	46	2.20	0.68	Recife-PE-Brazil	Souza et al. 2015

17*	-	4x	46	2.54	0.9	Recife-PE-Brazil	Souza et al. 2016
18*	-	4x	46	2.70	0.93	Itumirim-MG-Brazil	Souza et al. 2017
19*	-	5x	55	2.88	0.64	Itumirim-MG-Brazil	Souza et al. 2018
20*	-	5x	58	3.00	0.67	Recife-PE-Brazil	Souza et al. 2019
21*	-	6x	66	2.91	-	Itatiaia-RJ-Brazil	Costa 2009
22*	-	6x	66	3.01	0.57	Lavras-MG-Brazil	Souza et al. 2019
23*	-	6x	66	3.11	0.47	Lavras-MG-Brazil	Souza et al. 2020
24*	-	6x	66	5.32	0.55	Lavras-MG-Brazil	Souza et al. 2021
25*	-	6x	66	3.57	-	-	Tuler et al. 2019
26	Curi 31	6x	66	3.39	4.4	Curitiba-PR-Brazil	This study
27							
28	Pm01	6x	66	3.43	5	Paraty- Pedra da Macela RJ-Brazil	This study
29*	-	7x	77	3.82	-	Uruguay	Medina 2014
30	Curi01	7x	77	3.96	3.9	Curitiba-PR-Brazil	This study
31	Curi03	7x	77	3.87	4.1	Curitiba-PR-Brazil	This study
32	Curi04	7x	77	3.88	4.8	Curitiba-PR-Brazil	This study
33	Curi06	7x	77	3.92	4.3	Curitiba-PR-Brazil	This study
34	Curi09	7x	77	4.05	4.5	Curitiba-PR-Brazil	This study
35	Curi12	7x	77	3.87	4.3	Curitiba-PR-Brazil	This study
36	Curi13	7x	77	3.95	4.1	Curitiba-PR-Brazil	This study
37	Curi31	7x	77	3.81	3.8	Curitiba-PR-Brazil	This study
38	SV01	7x	77	3.71	4.7	S. L. do Paraitinga-PESM-SP-Brazil	This study
39	Trind01	7x	77	3.73	5.0	Paraty- Praia de Trindade-RJ-Brazil	This study
40	P.Brav01	7x	77	3.86	5.0	Paraty- Praia Brava-RJ-Brazil	This study
41	P.Brav02	7x	77	3.92	5.0	Paraty- Praia Brava-RJ-Brazil	This study
42	P.Brav03	7x	77	3.88	5.0	Paraty- Praia Brava-RJ-Brazil	This study
43	P.Brav04	7x	77	3.87	5.0	Paraty- Praia Brava-RJ-Brazil	This study
44	Pm02	7x	77	4.01	5.0	Paraty- Pedra da Macela RJ-Brazil	This study
45	Pm03	7x	77	3.95	5.0	Paraty- Pedra da Macela RJ-Brazil	This study
46	Pm04	7x	77	3.82	5.0	Paraty- Pedra da Macela RJ-Brazil	This study

47*	-	8x	82	5.47	0.5	Recife-PE-Brazil	Souza 2014
48*	-	8x	88	4.25	-	Uruguay	Speroni et al. 2012
49	Curi02	8x	88	4.27	3.8	Curitiba-PR-Brazil	This study
50	Curi18	8x	88	4.42	4.4	Curitiba-PR-Brazil	This study
51	Ba01	10x	110	4.44	4.4	Caravelas-BA-Brazil	This study
52	Curi25	10x	110	5.77	3.7	Curitiba-PR-Brazil	This study
53	Ba02	10x	110	4.20	4.6	Caravelas-BA-Brazil	This study
54	Ba04	10x	110	4.85	4.9	Trancoso-BA-Brazil	This study
55	Itau01	10x	110	5.61	4.6	Itaúnas-ES-Brazil	This study
56	Itau02	10x	110	3.75	4.7	Itaúnas-ES-Brazil	This study
57	Juru01	12x	132	5.40	5.8	Macaé-PARNAJurubatiba-RJ-Brazil	This study
58	Lin01	12x	132	5.88	5.6	Pontal do Ipiranga-Linhares-ES-Brazil	This study

Capítulo 2: Population genetics of polyploid complex *Psidium cattleyanum* Sabine (Myrtaceae): preliminary analyses based on new species-specific microsatellite loci and extension to other species of the genus

Raquel Moura Machado[1,*], Fernanda Ancelmo de Oliveira[2], Fábio de Matos Alves[2], Anete Pereira de Souza[2] and Eliana Regina Forni-Martins[2]

[1]Programa de Pós-graduação em Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Rua Monteiro Lobato, 255, 13083-862, Campinas, SP, Brazil.
[2]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Rua Monteiro Lobato, 255, 13083-862, Campinas, SP, Brazil.
*Corresponding author: raquelmouramachado@gmail.com

Published in Biochemical Genetics. https://doi.org/10.1007/s10528-020-10002-1
The final publication is available at https://link.springer.com/article/10.1007%2Fs10528-020-10002-1

Authors' ORCID:
Raquel Moura Machado - 0000-0002-9603-5747
Fernanda Ancelmo de Oliveira - 0000-0002-9488-0867
Fábio de Matos Alves -0000-0002-7902-4915
Anete Pereira de Souza - 0000-0003-3831-9829
Eliana R. Forni-Martins - 0000-0002-8889-4955

Declarations

Funding- This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and FAEPEX/UNICAMP. FAO received a Post Doctoral fellowship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2018/18527-9).

Conflicts of interest- The authors declare no conflicts of interest.

Ethics approval- This article does not include studies about humans or animals performed by any of the authors.

Authors' contributions - All authors contributed equally to this manuscript.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and FAEPEX/UNICAMP for the financial support. FAO received a Post Doctoral fellowship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2018/18527-9). We thank Prof. Marcia Dr. Flores da Silva Ferreira, of Universidade Federal do Espírito Santo, for sending the different *Psidium* species plant material, and Dr. André Vito Scatigna, for his support in fieldwork and collecting plant material from the individual grown in Uruguaiana (RS). We also thank the staff of Laboratório de Análises Genéticas e Moleculares (LAGM/UNICAMP).

A**bstract** – Polyploidy is a phenomenon that alters the genetic diversity of populations and has been reported as one of the most important evolutionary forces for plant diversification. The *Psidium cattleyanum* complex comprises a group of wild populations with several ploidy levels reported in the literature. The multiple cytotypes, associated with its wide distribution area, make this species a potential key model for understanding evolutionary processes related to polyploidization. In this study, we isolated and characterized nuclear microsatellite markers of *P. cattleyanum* and tested their transferability to other nine species of the genus. We performed a preliminary analysis of genetic diversity and population structure in three populations of *P. cattleyanum*. The three populations analyzed had different chromosome numbers, being polyploid cytotypes (2n=6x=66, 2n=7x=77 and 2n=8x=88). We designed 46 primer pairs and successfully amplified 37 markers, from which the 10 best were selected for analysis. Considering both the PIC and DP values, most of markers were highly informative. The new SSR markers were used to assess the levels of genetic diversity of the populations and detected one population with predominance of sexual reproduction. DAPC analysis pointed the formation of three groups, which corresponded to the populations analyzed. The markers were successfully amplified in related species, with some species presenting 80% transferability. By producing this panel of polymorphic microsatellites, we contribute to the understanding evolution in groups of natural polyploids for future studies.

Keywords: Araçá, Cattley guava, cytotype, cross-amplification, genetic structure, polyploidy, SSR.

Introduction

Polyploidization has been reported as an evolutionary mechanism that promotes plant diversity (Soltis et al. 2009, 2010, 2014). The use of molecular biology techniques has revolutionized the study of polyploidy, revealing that a single species can frequently form polyploids (Soltis et al. 2014). Polyploidy arises from the multiplication of a single genome (autopolyploidy) or the combination of two or more different genomes, followed by the multiplication of this new hybrid genome (allopolyploidy) (Chen and Ni 2006; De Wet 1980). The existence of multiple origins of polyploidy and the possibility of gene flow between populations (diploids and polyploids) and between polyploids with different origins (auto and allopolyploids) demonstrate that polypolidization is a mechanism that causes genetic variation (Schifino-Wittmann 2004). Higher genetic diversity in polyploid organisms allows a species to expand its geographic distribution (Otto and Whitton 2000; Soltis and Soltis 2000), whereas the effects of polyploidy on plant growth, epigenetics and physiology is likely to lead to niche shift between the different ploidy levels (Baniaga et al. 2019; te Beest et al. 2011).

The neotropical genus *Psidium* L. (Myrtaceae) includes species of high economic importance, like *P. guajava* L. (common guava), and other species with great economic potential, with promising futures for their fruits, such as *P. cattleyanum* Sabine (Cattley guava) and *P. guineense* Sw. (Brazilian guava) (Franzon et al. 2009; Mani et al. 2011). In this genus, polyploidization events are recurrent, with records of diploid, triploid, tetraploid, hexaploid and octaploid species (Marques et al. 2016), and also species with two or more ploidy level (Costa and Forni-Martins 2006).

Psidium cattleyanum presents several ploidy levels, which characterizes it as a polyploid complex. The chromosomal numbers reported in the literature include 2n=33, 44, 46, 48, 55, 58, 66, 77, 82, 88, 99, 110 and 132 (Atchison 1947; Costa and Forni-Martins 2006; Costa 2009; de Souza et al. 2014; Hirano and Nakazone 1969; Machado 2016; Raseira and Raseira 1996; Singhal et al. 1984; Souza-Pérez and Speroni 2017). Diploid individuals of *P. cattleyanum* (2n=22) have not yet been observed and the origin of polyploid (either auto or allopolyploidy) in this taxon remains unknown. Two

hypotheses regarding the origin of polyploidy in *P. cattleyanum* have been discussed: based on *in situ* hybridization studies (*FISH*), Costa (2009) raised the hypothesis that the tetraploid *P. cattleyanum* could have originated by the hybridization of two species with 2n = 22 and differentiated genomes; meanwhile Medina (2014) and Machado (2016) suggested the hypothesis of an autopolyploid origin due to the repetition of a genome x = 11, based on observations of fluorochromes CMA/DAPI banding and mapping DNAr sites (using *FISH*).

Natural populations of *P. cattleyanum* are distributed in eastern Brazil, from Bahia to Rio Grande do Sul (Wikler 2007), and in Uruguay (Souza-Pérez and Speroni 2017). The species was introduced to other tropical regions, where it became an aggressive invader, suppressing native vegetation and causing ecological disturbance (Wikler 2007) and continues to invade new areas, being considered one of the top 100 invasive plants of the world according to the Global Database of Invasive Species (2020). This highly invasive potential can be related to its reproduction strategies, among other factors.

The occurrence of asexual reproduction (i.e. apomixis and clonal reproduction) is strongly associated with polyploidy (Kólar et al. 2017; Van Drunen and Husband 2019). Recently, Souza-Pérez and Speroni (2017) detected apomixis in two cytotypes of *P. cattleyanum*. Regarding sexual reproduction, studies showed that the species can be autogamous (Raseira and Raseira 1996), and allogamous as well, with this strategy varying from individual to individual (Normand and Habib 2001). The occurrence of root sprout regeneration is also possible for this species (Global Database of Invasive Species 2020; Machado, personal observation).

The several ploidy levels recorded in *P. cattleyanum* reveal its potential as a model for understanding evolutionary processes in groups with polyploidy. Despite this potential, there is no data about the genetic diversity of *P. cattleyanum* populations. Most genetic analyses of *Psidium* have focused on investigating the genetic variability among common guava cultivars (Coser et al. 2012; Kherwar et al. 2018; Noia et al. 2017), whereas only few attempts have been made to assess the genetic diversity of indigenous *Psidium* species, using dominant markers (Mani et al. 2011; Sharma et al. 2007) and SSR (Simple Sequence Repeats) markers transferred from *P. guajava* (Tuler et al. 2015; Tuler et al. 2019).

Due to their high polymorphism, molecular microsatellite (or SSR) markers are useful for evaluating the genetic similarity between closely related individuals or taxa (Hodel et al. 2016). However, the use of markers transferred from related species is

associated with selection of the most polymorphic loci, which may lead to a bias in genetic diversity parameters (Queirós et al. 2015). Hence, despite the large panel of SSR markers developed for *P. guajava* (Noia et al. 2017; Tuler et al. 2015), analyses using species-specific microsatellites could be more accurate for measuring genetic variability in *P. cattleyanum*.

Therefore, the purpose of this study was to develop and characterize genomic microsatellite markers for *P. cattleyanum* and test the cross-amplification of these markers in other species of *Psidium*. Additionally, we performed preliminary analysis of genetic diversity and population structure in three wild populations of *P. cattleyanum* from Brazil. We also used three cytotypes to develop the libraries and tested whether the SSR markers are transferred to individuals of *P. cattleyanum* with other ploidy levels.

Materials and methods

DNA extraction and ploidy level assessment - For DNA extraction of all individuals analyzed in this study (Tables 1, 5), we followed a modified method based Tel-Zur et al. (1999) using approximately 20 mg of leaves dehydrated in silica gel of each sample. We quantified the genomic DNA with a Nanodrop spectrophotometer.

Ploidy levels of the three individuals used in the SSR-enriched library development were previously assessed by Machado (2016), whereas those of the studied populations of *P. cattleyanum* (PR-MX, RJ-8X and SP-6X from Table 1) were determined here based on direct chromosomal counts, following the protocol proposed by Guerra (2002). For each population, we sampled all 15 individuals for PR-MX, 10 individuals for RJ-8X and nine individuals for SP-6X and counted at least five cells per individual (Online Resource 3). Vouchers of all populations were deposited in the Herbarium UEC of the University of Campinas (Table 1).

Construction of the SSR-enriched library - To construct the *P. cattleyanum* microsatellites library, individuals of three ploidy levels were selected: 2n=4x=44 (L1 = Bertioga, SP – Voucher: *R.M. Machado 2* (UEC)); 2n=8x=88 (L2 = Curitiba, PR – Voucher: *R.M. Machado 1* (UEC); and 2n=12x=132 (L3 = Uruguaiana, RS – Voucher: *A.V. Scatigna 670* (UEC), cultivated in a public park). We built the three microsatellite libraries (L1, L2 and L3) according to a modified protocol of Billotte et al. (1999). The total DNA was digested using the *Afa* I (10u/µL) restriction enzyme. After digestion, we

performed a connection reaction to the Rsa21 (10μM) (5'CTCTTGCTTACGCGTGGACTA3') and Rsa25 (10μM) (5'TAGTCCACGCGTAAGCAAGAGCACA3') adaptors to facilitate the enrichment step. In the enrichment step, we used probes (magnetic beads – Invitrogen-Dynal, Lillestrøm, Norway) to capture the DNA fragments containing microsatellite sequences, via hybridization with oligonucleotides Biotin-IIIII(CT)8 and Biotin-IIIII(GT)8.

The captured fragments were amplified by PCR and the resulting products were cloned in a pGEM-T vector (Promega, Madison, WI, USA). Afterwards, they were inserted into a competent *E. coli* cell (Agilent Technologies, Stratagene Products Division, La Jolla, CA, USA) via electroporation. We cultured the competent cells in petri dishes with agar and selected 94 positive colonies from each plate for transformation, then we extracted the vectors from them and sequenced a total of 282 recombinant clones.

Subsequently, we removed the adapters and performed trimming to eliminate contamination by the vectors using the Geneious v.9 platform (https://www.geneious.com). We also used this program to obtain the consensus sequences and searched for vectors using the VecScreen tool (https://www.ncbi.nlm.nih.gov/tools/vecscreen/). The consensus sequences were aligned using the NCBI nonredundant (Nr) database, through the BLASTN algorithm with an e-value cutoff of 1e-06 using the BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We identified the sequences containing microsatellites with the PHOBOS plug-in (http://www.rub.de/ecoevo/cm/cm_phobos.htm) and designed the primers using Prime3Plus on Geneious v.9. All the primer pairs were synthesized with the sequence m13- 5'CACGACGTTGTAAAACGA3' (Schuelke 2000).

Validation of the library and statistical analysis – To validate the microsatellite library, we used 45 individuals from three *P. cattleyanum* populations (15 individuals for population) (Table 1).

The PCR reaction used 4 µl of genomic DNA (20ng), 0.75 µl 10x Buffer (50 mM KCl, 20 mM Tris-HCl (pH 8.4)), 0.3 µl MgCl2 (50 mM), 0.4 µl dNTP (2.5 mM), 0.8 µl BSA (2.5 mg/ml), 0.1 µl of each primer (10 mM), 0.1 µl of fluorophore 700 or 800, 0.3 µl of Taq DNA Polymerase 5 U/µl (Roche®) and water to complete a final volume of 10 µl. The amplification conditions were based on those described by Tuler et al. (2015), with the specific annealing temperature of each primer pair (Table 2). After the optimization phase of the PCR reactions, the amplified products were checked on 6.5%

acrylamide gel using the Li-Cor device (4300 DNA Analyzer) to confirm if the fragments were polymorphic. We also tested whether the primers designed for each cytotype (2n=4x=44, 2n=8x=88 and 2n=12x=132) were transferred to ploidy levels that differed from those of the original library. After the tests, we performed the amplification and genotyping of the studied populations (Table 1).

In the present study, we treated the microsatellites as dominant markers. The data was visually scored based on the presence (1) and absence (0) matrix for the bands on the acrylamide gel. This effectively reduces the information of the markers content and precludes analyses that consider observed heterozygosity of individuals or allele frequency distributions. However, polyploid microsatellite genotyping is difficult due to stutter peaks and/or allele overlap caused by multiple copies of the same size.

The polymorphism content (PIC – Botstein et al. 1980) and the discriminating power (DP – Tessier et al. 1999) (Table 3) were used to evaluate and characterize each marker. We estimated the Shannon index (Shannon 1948), Simpson index (Simpson 1949), and Nei diversity index (Nei 1978) to evaluate the genetic diversity of individuals, using the *poppr* (Kamvar et al. 2014) and *polysat* (Clark and Jasieniuk 2011) packages in the R 3.4.0 platform.

Since apomixis can occur in *P. cattleyanum* (Souza-Pérez and Speroni 2017), we decided to estimate the number of multilocus genotypes (MLG) and the eMLG, which is the MLG value of the smallest sample. Both of which are indirect inferences of asexual reproduction. The *poppr* package (Kamvar et al. 2014) in the R 3.4.0 platform was also used for these analyses.

We used a Discriminant Analysis of Principal Components (DAPC; Jombart et al. 2010), which uses a nonparametric approach, free from Hardy-Weinberg constraints, to detect the population structure and grouping of the individuals studied. DAPC combines PCA, K-means clustering and discriminant analysis to detect and visualize population structure and can be applied to polyploid data sets (Meirmans et al. 2018). This analysis was implemented in the R package *adegenet*. Two approaches were conducted: (1) the first DAPC analysis was performed, providing the information of three groups (populations of *P. cattleyanum*) and (2) the number of clusters was assessed using the function *find.clusters*. We assumed 15 as the maximum number of clusters. The optimal number of clusters was estimated using the Bayesian information criterion (BIC), both DAPC results were presented as scatter plots and the second approach was presented as a barplot.

Cross-amplification of the *P. cattleyanum* library - We tested the transferability of the SSR-enriched library developed for *P. cattleyanum* to nine species of the genus (one sample for each species): *P. acutangulum* DC., *P. australe* Cambess, *P. brownianum* Mart. ex DC., *P. friedrichsthalianum* (O. Berg) Nied., *P. gaudichaudianum* Proença & Faria, *P. guajava* L., *P. myrtoides* O. Berg, *P. oblongatum* O.Berg and *P. rombeum* O.Berg. Dehydrated leaves of these nine species were provided by Prof. Marcia Flores da Silva Ferreira, who cultivates these plants at the Universidade Federal do Espírito Santo. PCR parameters were the same used herein for *P. cattleyanum*.

Results and Discussion

This is the first effort to assess genetic diversity and mating system in *P. cattleyanum* using species-specific SSR marker, which are also herein developed, and spanning different ploidy levels. Below, we present and discuss our results of the characterization of the microsatellite library, its cross-amplification to other nine species of *Psidium* and the preliminary analysis of population genetics in natural populations of *P. cattleyanum*.

The population from Paraná (PR-MX) presented two cytotype, 2n = 77 and 2n = 88, but the first was predominant (14/15 samples); the population from São Paulo (SP-6X) presented only hexaploid specimens (2n = 66); and the population of Rio de Janeiro (RJ-8X) presented only octaploid specimens (2n=88) (Table 1; Online Resource 3).

From the consensus sequences, we designed a total of 46 primer pairs (Online Resource 1) for the microsatellites' nuclear flanking regions, obtaining 18 primer pairs from the library of cytotype 2n=4x=44, 12 from the library of cytotype 2n=8x=88, and 16 from the library of cytotype 2n=12x=132. We found the annealing temperatures for 37 pairs of primers, which are listed in Online Resource 1. Nine were discarded from the analysis because they did not amplify fragments under the various conditions tested.

Even if the primer pair was designed from the sequence of a specific cytotype library, we observed no amplification problems across different ploidy levels. Thus, we resumed the analysis considering only one general library for the species. We identified six pairs of primers with monomorphic products and 31 with polymorphic microsatellites, but 21 primer pairs had amplification artifacts (i.e. overlapping fragments or non-specific amplification resulting in several bands), which made genotyping difficult. Therefore, we selected the 10 pairs that resulted in ideal genotyping conditions, clear fragments and easy interpretation on the gel (Table 2).

The PIC values obtained for the 10 markers ranged from 0.37 to 0.89, with an average of 0.73 (Table 3). According to the Botstein classification (Botstein et al. 1980), PIC values ≥ 0.5 are highly informative. Thus, nine loci were highly informative and only Pca-UNICAMP05 was moderately informative, with a PIC of 0.37. DP values ranged from 0.31 to 0.85, with an average of 0.74 (Table 3). Considering both the PIC and DP values found for each locus, four markers showed the highest levels of informativeness: Pca-UNICAMP04, Pca-UNICAMP10, Pca-UNICAMP08 and Pca-UNICAMP11 (Table 3).

For the 10 loci analyzed, the number of bands per locus ranged from 2 to 15 for *P. cattleyanum*, with an average of 8.2 (Table 3). Most of these bands were exclusive, i.e. bands observed in only one population; only two loci showed no exclusive bands (Pca-UNICAMP05 and Pca-UNICAMP07). The population with the highest number of exclusive bands was PR-MX, with a total of 33 bands. The exclusive bands represent a simple measurement of a population's genetic distinction (Oliveira et al. 2016).

The highest diversity levels (Table 4) were observed for SP-6X population (Shannon index: 2.52 and Simpson index: 0.91). Meanwhile, we found that the population with two cytotypes (PR-MX) showed the lowest levels of genetic diversity (Shannon index: 0.88 and Simpson index: 0.551), when compared to populations with one cytotype only.

The PR-MX population also presented a low number of multilocus genotypes (Table 4), only three were detected, suggesting that this population was formed by a small number of generating individuals (founder effect), and was probably propagated via apomictic seeds. This hypothesis explains the low genetic diversity of the population indicated by the Shannon and Simpson indices, reinforced by the higher Nei diversity index (Nei 1978), showing the highest heterozygosity in the PR-MX population (0.279). One consequence of apomixis is the preservation of heterozygosity of the mother plant (Cruz et al. 1998). Apomixis has already been detected for Uruguayan *P. cattleyanum* populations, in the heptaploid and octaploid cytotypes (Souza-Pérez and Speroni 2017). Once the levels of genetic diversity in different cytotypes depend on many factors, i.e. changing mating system, geographical distribution and environmental niche (Baduel et al. 2018; Baniaga et al. 2019; Meirmans et al. 2018), a more in-depth population analysis encompassing a larger number of individuals sampled can help to elucidate these results.

The results of the two DAPC analyses are presented in Figs. 1 and 2. The screen plots of eigenvalues of discriminant analysis and the amount of variation contained in the

different principal components are presented in Online Resource 2. In the first approach, DAPC showed the formation of three groups (Figure 1), which corresponded to the populations analyzed in this study: PR-MX= orange dots; SP-6X= gray dots; RJ-8X= dark blue dots. In the second analysis, we assumed no prior information about groupings of the individuals evaluated. Using the K-means algorithm and inspection of BIC values, six clusters were revealed to explain the variance in these groups of individuals. Therefore, the data was divided into six clusters (Fig 2), with no clear delimitation of groups within our dataset. The barplot (Fig 2.c) showed a subdivision (structuring) in the populations analyzed. Both results of the DAPC analyses showed that the microsatellite markers (SSRs) developed for *P. cattleyanum* can detect genetic structure of populations. The ability to detect population structure with a clustering approach is an important part of genetic studies (Meirmans et al. 2018).

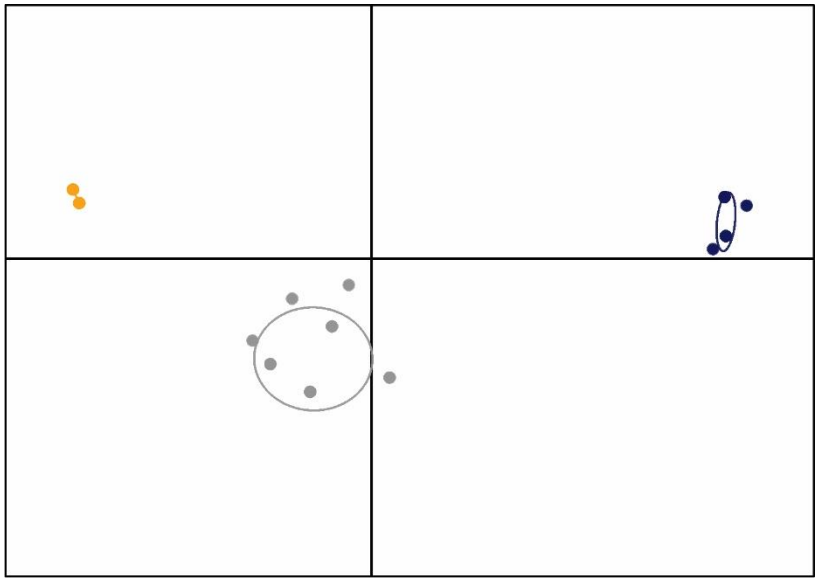


Fig 1 Scatterplot of DAPC using the three populations of *Psidium cattleyanum* as pre-determined groupings. Dots represent individuals and different colors represent the populations: PR-MX= orange dots; SP-6X= gray dots; RJ-8X= dark blue dots.

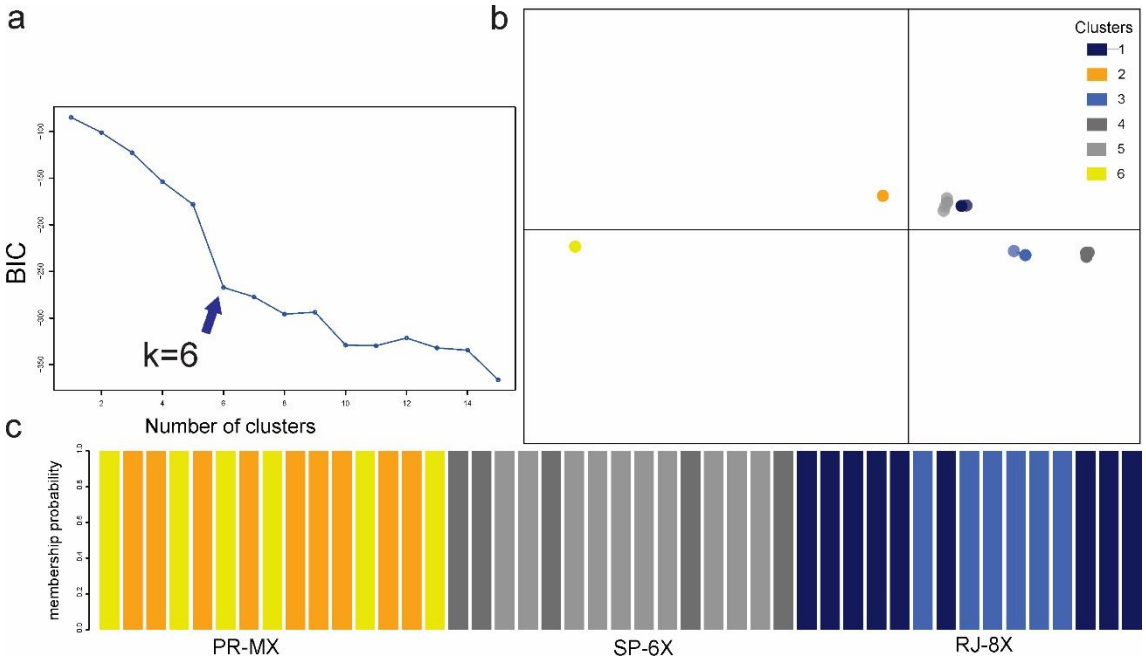


Fig 2 DAPC scatterplots based on the K-means algorithm; (A) k number is selected based on BIC value for clusters up to k = 15; (B) scatter plot shows genetic patterns of SSR data; (C) bar plot showing the probabilities of assignment of individuals to k = 6 genetic DAPC clusters. PR-MX= yellow and orange dots/bars; SP-6X= gray and dark gray dots/bars; RJ-8X= blue and dark blue dots/bars.

Of the nine species used in the transferability tests of the SSR library developed for *P. cattleyanum* (Table 5), seven showed amplification of at least two markers: *P. acutangulum* (30%), *P. australe* (10%), *P. brownianum* (70%), *P. friedrichsthalianum* (60%), *P. gaudichaudianum* (80%), *P. guajava* (20%) and *P. oblongatum* (70%). Only two species (*P. rhombeum* and *P. myrtoides*) presented no amplification of any marker. It is unclear whether there was no amplification due to the quality of extracted DNA or because the loci mutated or were lost in the differentiation of these species.

Many factors could be related to the low transferability rate found for some species, such as the genetic distance between groups and/or the difference in ploidy between *P. cattleyanum* and the other species tested in this study. Some studies have evaluated the relationships between *Psidium* species using the SSR library developed for *P. guajava* to compare and delimit them (Noia et al. 2017; Tuler et al. 2015). The transferability results demonstrate its potential for identifying groups in the genus based on the amplification patterns of the transferred library (de Oliveira Bernardes et al. 2018; Noia et al. 2017; Tuler et al. 2015). In both studies (Noia et al. 2017; Tuler et al. 2015),

the *P. cattleyanum* species was grouped with *P. myrtoides*; however, the SSR markers developed for *P. cattleyanum* were not amplified for this species. These conflicting results demonstrate the importance of developing a species-specific library for detailed studies of genetics and population structure, with the development of the *P. cattleyanum* library being an important step.

The loci characterized herein were successfully cross-amplified, suggesting that these markers can be used in genetic studies for other species of the genus. In addition, these newly developed SSR markers can be used to develop future breeding programs, since some species have fruits that are highly appreciated by local populations, i.e., *araçá* (*P. cattleyanum*), or global ones, i.e., guava (*P. guajava*). Additionally, markers can detect asexual reproduction, with apomixis being a huge advantage for commercial development, producing uniform *P. cattleyanum* offspring.

In the future, we aim to estimate the genetic diversity of additional natural *P. cattleyanum* populations by relating it to the amount of cytotypes and ploidy levels within each population. Furthermore, our markers could help with invasive species control programs that target *P. cattleyanum*. Finally, the SSR markers are an important contribution to research involving groups with natural polyploids.

References

Atchison E (1947) Chromosome numbers in the Myrtaceae. Am J Bot. https://doi.org/10.1002/j.1537-2197.1947.tb12970.x

Baduel P, Bray S, Vallejo-Marin M et al. (2018) The “polyploid hop”: shifting challenges and opportunities over the evolutionary lifespan of genome duplications. Front Ecol Evol. https://doi.org/10.3389/fevo.2018.00117

Baniaga AE, Marx HE, Arrigo N, Barker MS (2019) Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. Ecol Lett. https://doi.org/10.1111/ele.13402

Billotte N, Lagoda PJL, Risterucci AM, Baurens FC (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits 54:277-288.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314.

Chen ZJ, Ni Z (2006) Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays. https://doi.org/10.1002/bies.20374

Clark LV, Jasieniuk M (2011) POLYSAT: an R package for polyploid microsatellite analysis. Mol Ecol Resour. https://doi.org/10.1111/j.1755-0998.2011.02985.x

Coser SM, Ferreira MFS, Ferreira A, Mitre LK, Carvalho CR, Clarindo WR (2012) Assessment of genetic diversity in Psidium guajava L. using different approaches. Sci Hortic. https://doi.org/10.1016/j.scienta.2012.09.030.

Costa IR (2009) Estudos evolutivos em Myrtaceae: aspectos citotaxonômicos e filogenéticos em Myrteae, enfatizando *Psidium* e gêneros relacionados. Dissertação de Mestrado, Universidade Estadual de Campinas

Costa IR, Forni-Martins ER (2006) Chromosome studies in Brazilian species of Campomanesia Ruiz & Pávon and Psidium L. (Myrtaceae Juss.). Caryologia. https://doi.org/10.1080/00087114.2006.10797891

Cruz RP, Federizzi LC, Milach SCK (1998) A apomixia no melhoramento de plantas. Cienc Rural. http://dx.doi.org/10.1590/S0103-84781998000100028.

De Oliveira Bernardes C, Tuler AC, Ferreira A, Carvalho MS, Nogueira AM, da Silva Ferreira MF (2018) Transferability of Psidium microsatellite loci in Myrteae (Myrtaceae): a phylogenetic signal. Euphytica. https://doi.org/10.1007/s10681-018-2195-2

De Souza ADG, Resende LV, De Lima IP, Martins LSS, Techio VH (2015) Chromosome number and nuclear DNA amount in Psidium spp. resistant and susceptible to Meloidogyne enterolobii and its relation with compatibility between rootstocks and commercial varieties of guava tree. Plant Syst Evol. https://doi.org/10.1007/s00606-014-1068-y

De Wet JMJ (1980) Origins of polyploids. In: Lewis WH (ed). Polyploidy: biological relevance. Springer, Boston

Franzon RC, Campos LDO, Proença CEB, Sousa-Silva JC (2009) Araçás do gênero Psidium: principais espécies, ocorrência, descrição e usos. Embrapa, Planaltina https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/697560/1/doc266.pdf

Global Invasive Species Database (2020) Species profile: Psidium cattleianum. Downloaded from http://www.iucngisd.org/gisd/speciesname/Psidium+cattleianum on 28-04-2020.

Guerra M (2002) Como observar cromossomos: um guia de técnicas em citgenética vegetal, animal e humana. Funpec, Ribeirão Preto

Hirano RT, Nakazone HY (1969) Chromosome numbers of ten species and clones in the genus Psidium. J. Am Soc Hortic Sci 94:83-86.

Hodel RG, Segovia-Salcedo MC, Landis JB et al. (2016) The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. Appl Plant Sci. https://doi.org/10.3732/apps.1600025

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genet. https://doi.org/10.1186/1471-2156-11-94

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. Peer J. https://doi.org/10.7717/peerj.281

Kherwar D, Usha K, Mithra SA, Singh B (2018) Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (Psidium guajava L.). J Plant Biochem Biot. https://doi.org/10.1007/s13562-017-0438-2

Kolár F, Čertner M, Suda J, Schönswetter P, Husband BC (2017) Mixed-ploidy species: progress and opportunities in polyploid research. Trends Plant Sci. https://doi.org/10.1016/j.tplants.2017.09.011

Machado RM (2016) Distribuição geográfica e análise cariotípica de citótipos de Psidium cattleianum Sabine (Myrtaceae). Dissertação de Mestrado, Universidade Estadual de Campinas

Mani A, Mishra R, Thomas G (2011). Elucidation of diversity among Psidium species using morphological and SPAR methods. J Phytol. http://updatepublishing.com/journal/index.php/jp/article/view/2305.

Marques A, Tuler AC, Carvalho CR et al. (2016) Refinement of the karyological aspects of Psidium guineense (Swartz,1788): a comparison with Psidium guajava (Linnaeus, 1753). Comp Cytogenet. https://doi.org/10.3897/CompCytogen.v10i1.6462

Medina SNV (2014) Psidium cattleyanum Sabine y Acca sellowiana (Berg.) Burret (Myrtaceae): caracterización cromossómica y cariotípica em poblaciones silvestre y genótipos selecionados em programas nacionales de mejoramiento. Monografia, Universidade de la República

Meirmans PG, Liu S, Van Tienderen PH (2018). The analysis of polyploid genetic data. J Hered. https://doi.org/10.1093/jhered/esy006.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number individuals. Genetics 89:583-590.

Noia LR, Tuler AC, Ferreira A, Ferreira MFS (2017). Relationship between Psidium species (Myrtaceae) by resistance gene analog markers: focus on nematode resistance. Genet Mol Res. http://dx.doi.org/10.4238/gmr16019441

Normand F, Habib R (2001) Phenology of strawberry guava (Psidium cattleianum) in ReÂunion Island. J. Hortic Sci Biotech. https://doi.org/10.1080/14620316.2001.11511407

Oliveira FA, Cidade FW, Fávero AP, Vigna BB, Souza AP (2016) First microsatellite markers for Paspalum plicatulum (Poaceae) characterization and cross-amplification in different Paspalum species of the Plicatula group. BMC Res Notes. https://doi.org/10.1186/s13104-016-2312-z

Otto SP, Whitton J (2000) Polyploidy: incidence and evolution. Annu Rev Genet. https://doi.org/10.1146/annurev.genet.34.1.401

Queirós J, Godinho R, Lopes S et al. (2015) Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. Mol Ecol Resour. https://doi.org/10.1111/1755-0998.12349

Raseira MCB, Raseira A (1996) Contribuição ao estudo do araçazeiro. Embrapa, Pelotas Schifino-Wittmann MT (2004) Poliploidia e seu impacto na origem e evolução das plantas silvestres e cultivadas. Agric sci technol. http://dx.doi.org/10.18539/cast.v10i.2.904

Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. https://doi.org/10.1038/72708

Shannon CE (1948) A mathematical theory of communication. Mob Comput Commun. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sharma AS, Sehrawat SK, Singhrot RS, Boora KS (2005). Assessment of genetic diversity and relationship among Psidium spp. through RAPD analysis. In: I International Guava Symposium. Acta Hort. https://doi.org/10.17660/ActaHortic.2007.735.7

Simpson EH (1949) Measurement of diversity. Nature. https://doi.org/10.1038/163688a0

Singhal VK, Gill BS, Bir SS (1984) Cytology of cultivated woody species (Polypetalae). Proc Indian Sci Congr Assoc 71: 143–144.

Soltis DE, Albert VA, Leebens-Mack J et al. (2009) Polyploidy and angiosperm diversification. Am J Bot. https://doi.org/10.3732/ajb.0800079

Soltis DE, Buggs RJA, Doyle JJ, Soltis PS (2010) What we still don't know about polyploidy. Taxon 59:1387-1403.

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I et al. (2014) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al.* (2011). New Phytol. https://doi.org/10.1111/nph.12756

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.97.13.7051

Souza-Pérez M, Speroni G (2017) New apomictic pathway in Myrtaceae inferred from Psidium cattleyanum female gametophyte ontogeny. Flora. https://doi.org/10.1016/j.flora.2017.06.010

te Beest M, Le Roux JJ, Richardson DM et al. (2011) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot. https://doi.org/10.1093/aob/mcr277

Tel-Zur N, Abbo S, Myslabodski D et al. (1999) Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus (Cactaceae). Plant Mol Biol Rep. https://doi.org/10.1023/A:1007656315275

Tessier C, David J, This P et al. (1999) Optimization of the choice of molecular markers for varietal identification in Vitis vinifera L. Theor Appl Genet. https://doi.org/10.1007/s001220051054

Tuler AC, Carrijo TT, Nóia LR et al. (2015) SSR markers: a tool for species identification in Psidium (Myrtaceae). Mol Biol Rep. https://doi.org/10.1007/s11033-015-3927-1

Tuler AC, Carrijo TT, Peixoto AL et al. (2019) Diversification and geographical distribution of Psidium (Myrtaceae) species with distinct ploidy levels. Trees. https://doi.org/10.1007/s00468-019-01845-2

Van Drunen WE, Husband BC (2019) Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. New Phytol. https://doi.org/10.1111/nph.15999

Wikler C (2007) Problemática do Araçazeiro como Planta Invasora. In: Pedrosa-Macedo JH, DalMolin A, Smith CW (ed). O Araçazeiro: Ecologia e Controle Biológico. Fupef, Curitiba

Table 1 *Psidium cattleyanum* populations used to validate the microsatellite-enriched library (SSR)

Population	Number of individuals	Ploidy level	Proportion of ploidy	Collection site	Latitude	Longitude	Voucher UEC
PR-MX	15	Mixed, 7x and 8x	93.3% 2n=77 : 6.6% 2n= 88	Curitiba – PR	-25.445917	-49.233472	*R.M. Machado 1*
SP-6X	15	6x	100% 2n= 66	Cunha – SP	-23.138069	-44.901639	*R.M. Machado 3*
RJ-8X	15	8x	100% 2n= 88	Paraty – RJ	-23.064306	-44.678722	*R.M. Machado 4*

Table 2 Characteristics of the 10 primer pairs designed for the microsatellites of the *Psidium cattleyanum* genome. SZ: Size of the amplified fragment. AT°C: specific annealing temperature of the primer pairs.

Library	Name	GenBank		5´-3´ sequence	Motif	SZ	AT°C
L1	Pca-UNICAMP01	MN047794	F	GACTTGACAAGGGCAAAGTC	(CT)14(CT)8	206-148	55
			R	TAAAGGTGCATTTGTCTGCG			
L1	Pca-UNICAMP02	MN047795	F	AAGTTGGCAGGTCTAGTTCC	(TG)9	294-280	60
			R	TCAAGCTAGGTATGCTTCCC			
L1	Pca-UNICAMP04	MN047797	F	CCTTTTACACATTAGCTCTCTC	(CT)22	182-130	55
			R	GACCTGGGGTGTCATAACAA			
L1	Pca-UNICAMP05	MN047798	F	CAAAGTAGGTATGCTGCGTG	(CT)16(CA)10	100-94	63
			R	GCAAGTTAAACCGATCTGCA			
L2	Pca-UNICAMP06	MN047799	F	GACTTGACAAGGGCAAAGTC	(CT)16(CT)8	190-148	60
			R	CTGCGTGTGCTAGACCTTAA			
L2	Pca-UNICAMP07	MN047800	F	ACTAATGACGGTCCTTGAGAC	(GT)8(CT)20	214-210	51
			R	TTGTTGAGACTGCATGCATG			
L2	Pca-UNICAMP08	MN047801	F	GCACGTGCAAGAAAGAGAG	(AG)31(CT)26(TC)16(TG)9	234-200	60
			R	GTTCACACAGCACGCTAATT			
L2	Pca-UNICAMP09	MN047802	F	CATGAAAAATGAGTAGGCTCTC	(CT)15(AG)23	222-192	60
			R	CTCAGCTGGTTGTGCATAAC			
L2	Pca-UNICAMP10	MN047803	F	ACAACCCTTCTTTGCCCTAA	(TC)15(AC)11(CT)21	242-174	60
			R	ACAGATGTCATCAGAAGACACT			
L2	Pca-UNICAMP11	MN047804	F	CGTTATCTCCTTCCTCCGAG	(CT)11(TTC)8	290-176	60

R ATCGCCGATCAACTTCGAG

Table 3 Description of the SSR markers developed for *Psidium cattleyanum*. NM = number of bands, PIC = polymorphism content, DP = discriminatory power, variation in band size and exclusive bands for each population

				Exclusive bands		
Locus' name	NM	PIC	DP	PR-MX	SP-6X	RJ-8X
Pca-UNICAMP01	8	0.79	0.79	4	0	0
Pca-UNICAMP02	4	0.58	0.61	1	1	0
Pca-UNICAMP04	15	0.91	0.83	7	0	3
Pca-UNICAMP05	2	0.37	0.31	0	0	0
Pca-UNICAMP06	9	0.79	0.77	6	2	0
Pca-UNICAMP07	3	0.51	0.74	0	0	0
Pca-UNICAMP08	10	0.88	0.84	3	0	5
Pca-UNICAMP09	6	0.74	0.81	2	1	1
Pca-UNICAMP10	14	0.89	0.84	6	4	1
Pca-UNICAMP11	10	0.85	0.85	4	1	1
		Total exclusive bands		33	9	11

Table 4 Genetic diversity estimates for the *Psidium cattleyanum* populations (Pop) evaluated in this study. Population code according to Table 2. N = number of individuals per population, MLG = number of multilocus genotypes, eMLG = expected number of MGL, considering the smallest sample size, Shannon Index = Shannon index (Shannon, 1948), Simpson Index = Simpson diversity index (Simpson, 1949), Nei Index = Nei diversity index (Nei, 1978)

Pop	N	MLG	eMLG	Shannon Index	Simpson Index	Nei Index
PR-MX	15	3	3	0.8820	0.551	0.279
SP-6X	15	13	13	2.5230	0.916	0.119
RJ-8X	15	10	10	2.1760	0.871	0.117
Total	45	26	11.3	2.9590	0.926	0.326

Table 5 Transferability of the *Psidium cattleyanum* species' SSR library to other species of the *Psidium* genus

Species	Pca-UNICAMP01	Pca-UNICAMP02	Pca-UNICAMP04	Pca-UNICAMP05	Pca-UNICAMP06	Pca-UNICAMP07	Pca-UNICAMP08	Pca-UNICAMP09	Pca-UNICAMP10	Pca-UNICAMP11	% amplified loci
P. acutangulum DC.	-	+	-	-	+	-	+	-	-	-	30
P. australe Cambess	-	-	-	-	+	-	-	-	-	-	10
P. brownianum Mart. ex DC.	+	-	+	+	+	-	-	+	+	+	70
P. friedrichsthalianum (O. Berg) Nied.	+	-	+	+	-	+	-	+	-	+	60
P. gaudichaudianum Proença & Faria	+	+	-	+	+	+	+	+	-	+	80
P. guajava L.	-	+	-	-	+	-	-	-	-	-	20
P. myrtoides O. Berg	-	-	-	-	-	-	-	-	-	-	0
P. oblongatum O.Berg	+	-	+	+	+	+	-	+	+	-	70
P. rhombeum O.Berg.	-	-	-	-	-	-	-	-	-	-	0
Total	4	3	3	4	6	3	2	4	2	3	

Online Resource 1- Characteristics of the 46 primer pairs designed for the microsatellites of the *Psidium cattleyanum* genome. SZ: Size of the amplified fragment. AT°C: specific annealing temperature of the primer pairs. – annealing temperature not determined. The loci in italics were monomorphic.

Library	Name	GenBank		5´-3´ sequence	Motif	SZ	AT°C
L1	Pca-UNICAMP12	MN629941	F	GAGTCATGGGTCAAGAGGAC	(AG)20	222	55
			R	GAACTCCAAAATTCTGGCCG			
L1	Pca-UNICAMP01	MN047794	F	GACTTGACAAGGGCAAAGTC	(CT)14(CT)8	158	55
			R	TAAAGGTGCATTTGTCTGCG			
L1	Pca-UNICAMP02	MN047795	F	AAGTTGGCAGGTCTAGTTCC	(TG)9	274	60
			R	TCAAGCTAGGTATGCTTCCC			
L1	Pca-UNICAMP13	MN629942	F	ATCTCAAGTGCTCTTTGAGA	(GA)20	185	51
			R	TGTCAAAGTGAGAGTTTGGG			
L1	Pca-UNICAMP14	MN629943	F	CCAGTGCCATAGTTTCTGTG	(CT)19	161	60
			R	TTGCCTTCCCATTTTGGTTG			
L1	Pca-UNICAMP15	MN629944	F	GCTCCCACTAGATCTTGCTT	(TC)26	132	51
			R	TCACAAGTATGATGCTCCCC			
L1	Pca-UNICAMP03	MN047796	F	GACTTGACAAGGGCAAAGTC	(CT)14(CT)8	158	51
			R	TAAAGGTGCATTTGTCTGCG			
L1	Pca-UNICAMP16	MN629945	F	GGCTGGGAATTCCTAACTGA	(AG)24	155	-
			R	ACAGAAATCACAGCCCAATT			
L1	Pca-UNICAMP17	MN629946	F	TCCTTGACCCTCTCTCTTCA	(TC)22	188	-
			R	TGCCCATTCAAACCATTGTC			
L1	Pca-UNICAMP18	MN629947	F	GATTGGATCGAGGCTAGGAG	(GA)28	200	51
			R	GACAAGATCTCTTCTGGCCA			
L1	Pca-UNICAMP04	MN047797	F	CCTTTTACACATTAGCTCTCTC	(CT)22	138	55
			R	GACCTGGGGTGTCATAACAA			
L1	Pca-UNICAMP19	MN629948	F	AGGCATCATTGACCGGTAAT	(AC)7	150	55
			R	CCTATAACCACTTCGCATGC			
L1	Pca-UNICAMP20	MN629949	F	TCTGCAGCAATTTCCATTCT	(TC)24	142	55
			R	ACACTAGAACACAACTCTCT			

L1	Pca-UNICAMP05	MN047798	F	CAAAGTAGGTATGCTGCGTG	(CT)16(CA)10	115	63
			R	GCAAGTTAAACCGATCTGCA			
L1	Pca-UNICAMP21	MN629950	F	GCACAAGAACTAAGACACAT	(AC)7	143	60
			R	ACATATTCCTGGGGAAAGGG			
L1	Pca-UNICAMP22	MN629951	F	GGAGGGGTGGTAGATTCCAT	(AG)28	136	-
			R	TCTCTCCCTCTTGTCCACTA			
L1	Pca-UNICAMP23	MN629952	F	AGTATCCCTGTCGCGTATTG	(GA)13(TG)13(GT)19(GA)16	139	55
			R	CCAAAATATCAACTACTCTCGC			
L2	Pca-UNICAMP06	MN047799	F	GACTTGACAAGGGCAAAGTC	(CT)16(CT)8	143	60
			R	CTGCGTGTGCTAGACCTTAA			
L2	Pca-UNICAMP24	MN629953	F	CCCACCCCATGTTTTAATCG	(AC)22(CT)20	163	-
			R	CAAATGCCCAAACCCCAC			
L2	Pca-UNICAMP07	MN047800	F	ACTAATGACGGTCCTTGAGAC	(GT)8(CT)20	198	51
			R	TTGTTGAGACTGCATGCATG			
L2	Pca-UNICAMP25	MN629954	F	TGAAAACCAAGTAGCACCCT	(TG)25(GT)24	244	51
			R	TTTGCTGAGTTCACTCCCAT			
L2	Pca-UNICAMP26	MN629955	F	ATCGTCGTCTTCTGCTGAAT	(TC)24	285	-
			R	TAGTGTGGATCGAGTTCAGC			
L2	Pca-UNICAMP27	MN629956	F	GGGATAATTGCTGAGGTGGT	(GA)25	220	60
			R	AAGGACCACTTTAGCTGCTT			
L2	Pca-UNICAMP28	MN629957	F	CATTAATCCCTAGGTCCCCC	(TC)18(CA)14	210	60
			R	ACTGCAATCCAATTGACACG			
L2	Pca-UNICAMP08	MN047801	F	GCACGTGCAAGAAAGAGAG	(AG)31(CT)26(TC)16(TG)9	211	60
			R	GTTCACACAGCACGCTAATT			
L2	Pca-UNICAMP09	MN047802	F	CATGAAAAATGAGTAGGCTCTC	(CT)15(AG)23	134	60
			R	CTCAGCTGGTTGTGCATAAC			
L2	Pca-UNICAMP29	MN629958	F	CATTAATCCCTAGGTCCCCC	(TC)19(CA)9(TC)21(CT)16	106	55
			R	TGTATATTCACCACCGCCTC			
L2	*Pca-UNICAMP30*	*MN629959*	*F*	*AAGCCAAGTGGGAACATGAT*	*(TG)13(AG)20(CA)19(TC)16*	*138*	*55*
			R	*TACCCTTTCACACACACACA*			

L2	Pca-UNICAMP10	MN047803	F	ACAACCCTTCTTTGCCCTAA	(TC)15(AC)11(CT)21	187	60
			R	ACAGATGTCATCAGAAGACACT			
L2	Pca-UNICAMP11	MN047804	F	CGTTATCTCCTTCCTCCGAG	(CT)11(TTC)8	256	60
			R	ATCGCCGATCAACTTCGAG			
L3	Pca-UNICAMP31	MN629960	F	TGCGAAGTTACCACCCAAT	(TC)30(CT)28	231	60
			R	GAAAAATCACACCCTACGCC			
L3	Pca-UNICAMP32	MN629961	F	AAGCCGAAGAGCAGAAGAAA	(AG)31(AG)15(GA)8(AG)7(GA)12	180	55
			R	ACCCGACATCTCTTCGTAAA			
L3	Pca-UNICAMP33	MN629962	F	TCGAAAGTTGGTTGGACTCA	(CT)20(CT)16	114	-
			R	TCTCACCCCTCTTGACTCTT			
L3	Pca-UNICAMP34	MN629963	F	TGTGAAGTTGCCTCATAGCT	(TC)28	128	63
			R	CCCTAGAAGAACCCATTGCA			
L3	Pca-UNICAMP35	MN629964	F	TGTTTATGTAGCCGTGTTGG	(GA)24	123	-
			R	TTCTGCTGTAGCCGTAATGT			
L3	Pca-UNICAMP36	MN629965	F	CCGCAGCTTCTATCAAACTT	(TC)13	144	-
			R	TACGACTCACTATAGGGCGA			
L3	Pca-UNICAMP37	MN629966	F	CAGCCTAACAGTTTGTGCTG	(CT)31	174	-
			R	TCGCAATTGAGAAGATCCCA			
L3	*Pca-UNICAMP38*	*MN629967*	*F*	*ACTCTCGAGACGTCCGAGA*	*(GA)8*	*143*	*55*
			R	*ACCATGATTACGCCAAGCTA*			
L3	Pca-UNICAMP39	MN629968	F	AGGAACTTGTTTGTCCCTTG	(TG)23(AT)13	294	60
			R	CCCCCTCCTTTTTGGAAGAA			
L3	*Pca-UNICAMP40*	*MN629969*	*F*	*CCGCAGCTTCTATCAAACTT*	*(TC)13(TC)9(AG)22*	*144*	*60*
			R	*TACGACTCACTATAGGGCGA*			
L3	*Pca-UNICAMP41*	*MN629970*	*F*	*TTGTGGTTGTGTGTTTGTGA*	*(GT)15*	*111*	*55*
			R	*AGAAGATATCTGCCCTACACA*			
L3	*Pca-UNICAMP42*	*MN629971*	*F*	*AAAGTGTTCACTGTGGGCAA*	*(CT)25(GA)26*	*103*	*60*
			R	*ATGGCCGCGGGATTAGAGA*			
L3	*Pca-UNICAMP43*	*MN629972*	*F*	*GTGCTGTCAATGCAACACTA*	*(TC)28(AG)25(AG)31*	*117*	*60*

			R	*ATGTTGAAGGGAGTGGTTGT*			
L3	Pca-UNICAMP44	MN629973	F	TTCGCTGCACAAAGTAGAGA	(AG)23(GA)9	120	63
			R	TACTCAAACAACCACCTTGC			
L3	Pca-UNICAMP45	MN629974	F	CATTCACTACCAGTTACTGTCA	(CT)21	125	55
			R	GCGAAACAATGAAAGGAACG			
L3	Pca-UNICAMP46	MN629975	F	ACACTCAACTAATCTTACCGCT	(GA)19	109	55
			R	CAACTCTCTTGGACCCTCTC			

Online Resource 2- Scatterplots of DAPC using the individuals of *Psidium cattleyanum*.

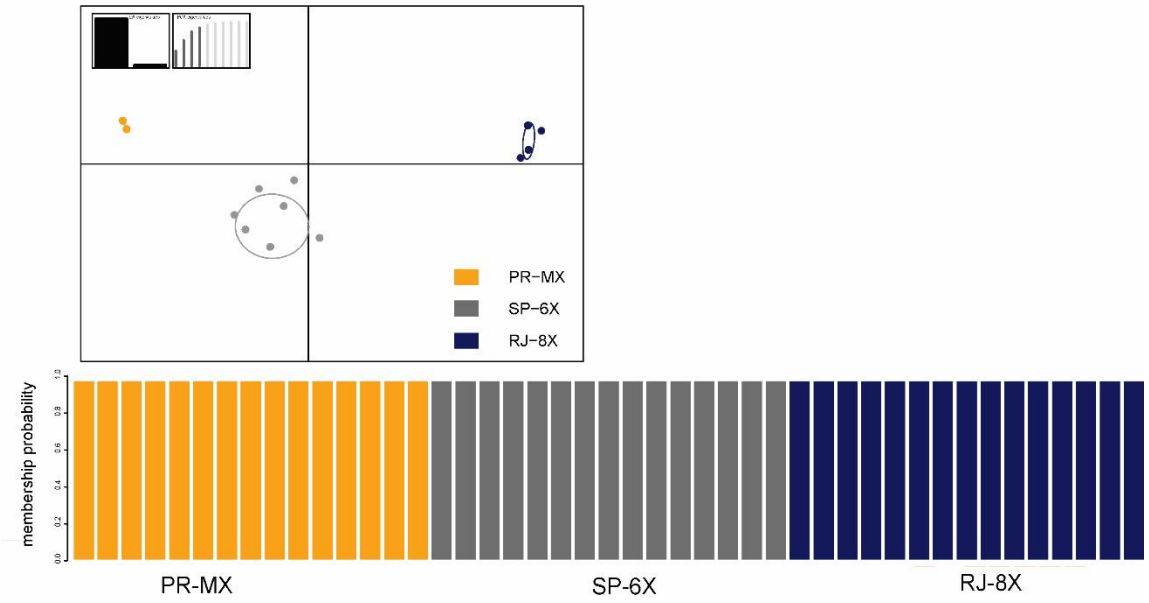


Fig 1.1 Scatterplot of DAPC using the three populations of *Psidium cattleyanum* as pre-determined groupings. Dots and bars represent individuals and different colors represent the populations: PR-MX= orange dots/bars; SP-6X= gray dots/bars; RJ-8X= dark blue dots/bars. Bar plot showing the probabilities of assignment of individuals. The screen plots of eigenvalues of discriminant analysis and the amount of variation contained in the different principal components are presented (insert).

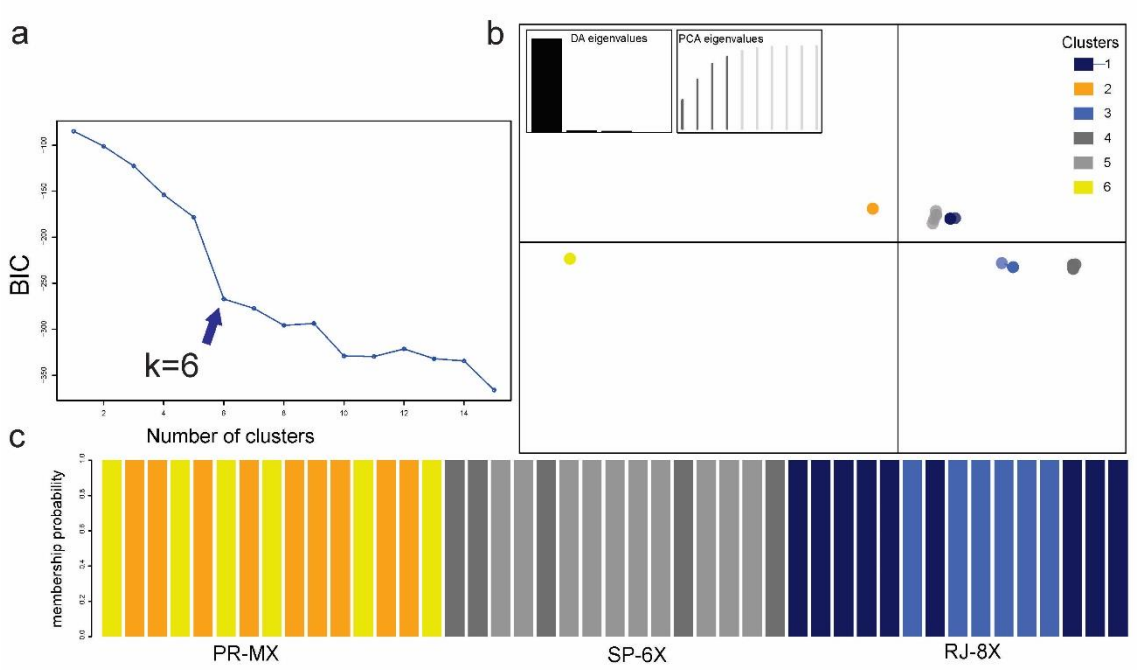


Fig 2.2 DAPC scatterplots based on the K-means algorithm; (A) k number is selected based on BIC value for clusters up to k = 15; (B) scatter plot shows genetic patterns of SSR data; (C) bar plot showing the probabilities of assignment of individuals to k = 6 genetic DAPC clusters. PR-MX= yellow and orange dots/bars; SP-6X= gray and dark gray dots/bars; RJ-8X= blue and dark blue dots/bars. The screen plots of eigenvalues of discriminant analysis and the amount of variation contained in the different principal components are presented (insert).

Online Resource 3- Mitotic metaphases of *Psidium cattleyanum*: (a) individual 2n= 66 from SP-6X (Cunha- SP), (b) individual 2n= 77 from PR-MX (Curitiba- PR) and (c) individual 2n= 88 from RJ-8X (Paraty- RJ).

Capítulo 3: Population structure and intraspecific ecological niche divergence of *Psidium cattleyanum* points to evidence of a future divergent lineage

Raquel Moura Machado[1,*], Fernanda Ancelmo de Oliveira[2], Ana Carolina Devides Castello[3], Fábio de Matos Alves[2], Anete Pereira de Souza[2] and Eliana Regina Forni-Martins[2]

[1]Programa de Pós-graduação em Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Rua Monteiro Lobato, 255, 13083-862, Campinas, SP, Brazil.
[2]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Rua Monteiro Lobato, 255, 13083-862, Campinas, SP, Brazil.
[3]Universidade do Estado de Minas Gerais, unidade de Ituiutaba, Rua. Ver. Geraldo Moisés da Silva, s/n, 38302-192, Ituiutaba, SP, Brazil.
*Corresponding author: raquelmouramachado@gmail.com

Abstract

Background and Aims: Polyploidy is defined as the presence of more than two complete chromosome sets in an organism and has frequently occurred throughout the history of Angiosperms. Polyploidization is a process that typically results in instant speciation. Using *Psidium cattleyanum*, a natural polyploid complex with several polyploid records, we aim to answer two questions regarding the polyploidy speciation: polyploidy promotes interruption of gene flow and intraspecific niche divergence.

Methods: We analyzed 12 natural populations of *P. cattleyanum* and, integrated population genetics data accessed by microsatellite markers and climatic niche analysis, using environmental niche modelling, we provided insights about polyploid speciation.

Key Results: Our data supports a strong genetic structure in populations and low environmental niche similarity between cytotypes. Genetic diversity declines with increasing ploidy level, probably associated with asexual reproduction. Thus, we can corroborate that polyploidy is both generating a reproductive barrier and it is associated with niche divergence.

Conclusions: Our data evidenced of a future divergent lineage between cytotypes of *P. cattleyanum,* and confirm the role of polyploidy as an evolutionary step in speciation. Additionally, this study provides new information for the discussion about how polyploidy affects genetic diversity of taxa and ecological niches.

Keywords: *Araçá*; Cattley guava; SSR; strawberry guava; ENM; Polyploidy;

INTRODUCTION

The speciation process is characterized by the emergence of barriers to gene flow between previously interbreeding populations (Rieseberg and Willis, 2007). Polyploidization is an important step in plant speciation (Grant, 1981) and the role of this events has been widely discussed (Baduel *et al.*, 2018; Jighly *et al.*, 2019; Gao, 2019; Levin, 2019; Baniaga *et al.*, 2020; Lavania, 2020). Polyploidization events are often associated with instant speciation (Gao, 2019) due to the rapid postzygotic reproductive barrier among polyploids and their progenitors (Stace, 1991; Rieseberg and Willis, 2007). Cytotypes are considered an intermediate step in the evolution of groups, favoring the speciation process (Briggs and Walters, 1997). The union of unreduced gametes are the major mechanism of polyploid formation and is greatly affected by environmental factors, particularly temperature variations (Ramsey and Schemske, 1998; Schinkel *et al.*, 2016). Although polyploidization is a ubiquitous event in plant evolution (Wendel *et al.*, 2005; Wood *et al.*, 2009), the knowledge about the evolutionary dynamics of polyploids is smaller when compared to diploid species (Jighly *et al.*, 2019).

Polyploidy is often associated with increased genetic diversity and empirical data suggests that established polyploids can accumulate adaptive allelic diversity (reviewed by Baduel *et al.*, 2018). The extra set of chromosomes in allopolyploids promotes a stable multiallelic state throughout generations, while in autopolyploids this higher genetic diversity comes from polysomic inheritance (Haldane 1930). The higher genetic diversity in polyploid organisms allows a species to expand its geographic and environmental distribution range (Otto and Whitton, 2000; Soltis and Soltis, 2000; Brochmann *et al.*, 2004). This is the case for invasive species, where polyploidy has already been shown to contribute to the success of biological invasions (Pandit *et al.*, 2011, te Beest *et al.*, 2011, Moura et al. 2021). Despite this, the evolutionary advantages of polyploidization events are still controverse (Baduel *et al.*, 2018).

Alterations caused by polyploidization include characters related to plant growth, epigenetics and physiology (te Beest *et al.*, 2011; Baniaga *et al.*, 2020), such as higher tolerance to stress, allowing cytotypes to occupy new ecological niches and expand their geographical ranges (Hijmans *et al.*, 2007; Pandit *et al.*, 2011; te Beest *et al.*, 2011; Moura *et al.* 2021). Ecological niche differentiation is an important step for polyploid speciation (Baniaga *et al.*, 2020). However, there is no unified theory about the influence of polyploidization events on ecological niches (Parisod and Broennimann, 2016). Both patterns, niche expansion and contraction, have been suggested for polyploid taxa, so this subject remains controversial (Parisod and Broennimann, 2016; López-Jurado *et al.*, 2019).

Due to the difficulty of obtaining direct empirical data, measuring the extent of polyploidy effects can be difficult (Baduel *et al.*, 2018), and knowledge about the influence of whole genome duplications (WGD) events in the diversification, speciation and ecology of species remains scarce (Tuler *at al.*, 2019), especially for tropical lineages. For example, autopolyploid organisms are sometimes indistinguishable from their diploid progenitors, but evolve as distinct lineages (Baduel *et al.*, 2018; Soltis *et al.*, 2007). Some autopolyploid cytotypes are actually new species, but they have been recognized as cytotypes because of the tradition of including cytotypes in a single named species and the convenience of adhering to a broad morphology-based species concept (Soltis et al. 2007).

The neotropical genus *Psidium* L. (Myrtaceae) is a good model for investigating the effects of polyploidy in evolutionary processes. In this genus, records of diploid, triploid, tetraploid, hexaploid, and octaploid species are common (Costa and Forni-Martins, 2006; Marques *et al.*, 2016; Tuler *et al*., 2015, 2019; Machado *et al.*, 2020;). Polyploidy has played an important role in the evolution and diversification of *Psidium* (Marques *et al.*, 2016, Tuler *et al.*, 2019, Machado *et al.*, 2020). Few studies have associated polyploidy with differential geographic distribution in Neotropical Myrtaceae, e.g. *Eugenia* (Silveira *et al.*, 2016).

For *Psidium*, Tuler *et al.* (2019), using SSR, evaluated the relationship between genetic diversity and geographic range. The authors found that the polyploid species of the genus have higher levels of genetic diversity and wider occurrence ranges, when compared to diploid species (Tuler *et al.*, 2019).

Psidium cattleyanum has characteristics that make it an excellent model for studying the influence of polyploidy in speciation (Machado 2020, Chapter 1). *Psidium cattleyanum* (Cattley guava) is a polyploid complex with several ploidy levels reported in the literature, and chromosome number ranging from 2n = 33 to 132 (Marques *et al.*, 2016; Souza-Pérez and Speroni, 2017; Tuler *et al*., 2019; Machado *et al.*, 2020; Machado 2021- Chapter 1). This species has a wide distribution range, and is naturally distributed in eastern Brazil, from Ceará to Rio Grande do Sul (Flora do Brasil 2020), and in Uruguay (Souza-Pérez and Speroni, 2017). In other tropical and subtropical areas of the world, *P. cattleyanum* is an aggressive invader, causing ecological disturbances (Wikler, 2007). Many characteristics could favor its highly invasive potential, such as several ploidy levels reported for species and reproductive strategies. The reproductive strategies of *P. cattleyanum* include autogamy (Raseira and Raseira, 1996) and allogamy (Normand and Habib, 2001), or asexual propagation by apomixis (Souza-Pérez and Speroni, 2017) and root sprout (Global Invasive Species Database, 2020; Wilker, 2007). Polyploidy is frequently associated with asexual reproduction (Kolár *et al.*, 2017; Van Drunen

and Husband, 2019), and could prevent pollinator dependence in colonization of new habitats (Schinkel *et al.*, 2016; Baduel *et al.*, 2018; Rezende *et al.*, 2020).

Mixed ploidy species (species with more than two cytotypes) are excellent models to understand the role of polyploidy in lineages divergence and diversification (Kolár et al., 2017). The wide distribution of *P. cattleyanum* forces its populations to confront different environmental conditions and landscapes. Spatially separated populations may experience isolation-by-distance and isolation-by-environment, both of which can reduce gene flow and increase population divergence (Wang *et al.,* 2013). The polyploid records of *P. cattleyanum* offer insights into the mechanisms of polyploid evolution. Identifying the factors that have influenced gene flow among populations is crucial to provide insights into the earliest stages of biological divergence. The combination of different approaches, such as population genetics, environmental niche modeling, and structural equation modeling, can help investigate and describe patterns related to the evolution of polyploids.

The genetic diversity of *P. cattleyanum* populations was recently accessed by Machado *et al.* (2020), and, even with a small sample, the preliminary analysis showed that polyploidy and reproductive traits greatly influence population genetics and structure of *P. cattleyanum.* In this study, we assessed the genetic diversity and genetic structure patterns throughout *P. cattleyanum's* distribution range, using microsatellites. Also, we used a climatic niche analysis under current climate to estimate the similarity and divergence of environmental niche for each cytotype. Additionally, we investigated the relative contribution of geographic and environmental variables shaping the organization of genetic diversity in *P. cattleyanum* populations. Using *P. cattleyanum* as a model, we tested two hypotheses related to polyploid speciation: (1) cytotypes act as independent evolutionary units, promoting divergence between populations and interruption of gene flow, with the expectation to observe high levels of differentiation and structuring among populations, (2) polyploidy promotes habitat divergence, thus, cytotypes will tend to occupy different ecological niches. We also asked the following questions: 1) Do higher ploidy levels have greater genetic diversity? 2) Is there evidence of gene-flow between populations? 3) Does isolation by distance or environment play a role in population differentiation?

MATERIALS AND METHODS

Population sampling

We sampled 12 natural populations (with 14 to 30 individuals per population) of *P. cattleyanum* (Table 1) by collecting plants located at least 10 meters apart. We sought to cover the entire natural distribution of *P. cattleyanum* in the eastern region Brazil: Paraná (PR1-MX, PR2-4X), São Paulo (SP1-4X, SP2-4X, SP3-4X, SP4-6X), Rio de Janeiro (RJ1-8X, RJ2-7X, RJ3-MX), Espírito Santo (ES1-12X, ES2-12X) and Bahia (BA-10X). The 15 individuals from the PR1-MX, SP4-6X and RJ2-7X populations were previously analyzed in a preliminary study conducted by Machado *et al.* (2020) to characterize the microsatellite library. Herein, we analyzed these same 15 individuals plus 15 new samples, increasing the sampling to 30 individuals per population. Voucher materials for each population were deposited in the herbarium of the University of Campinas (UEC; table 1). GPS coordinates were recorded for all individuals. Leaf samples from all individuals were collected and dried in silica gel until DNA extraction. The ploidy levels of *P. cattleyanum* populations used in this study were determined by Machado *et al.* (in prep, Cap. 1).

DNA Extraction and SSR Amplification

To extract the total DNA of all 328 individuals analyzed in this study, we used approximately 20 mg of dehydrated leaves and modified the method of Tel-Zur *et al.* (1999). The concentration and quality of the DNA in the samples were verified using a Nanodrop spectrophotometer (Thermo Scientific).

We amplified 10 microsatellite markers developed for *P. cattleyanum* and the PCR protocols described by Machado *et al.* (2020). The amplified products were separated by vertical electrophoresis on 6.5% acrylamide gel using the Li-Cor device (4300 DNA Analyzer). For microsatellites genotyping, we treated the microsatellites as dominant markers. We build a presence (1) and absence (0) matrix for the bands based on the visually inspection of the acrylamide gel.

Molecular analyses of microsatellite data

For each population, we estimated the number of bands (NB), number of private bands (PB), Shannon index (Shannon, 1948), Simpson index (Simpson, 1949), Nei diversity index (Nei, 1978) and number of multilocus genotypes (MLG), which is an indirect inference of asexual reproduction, using the *poppr* (Kamvar *et al.*, 2014) and *polysat* (Clark and Jasieniuk, 2011) packages in the R 3.4.0 platform (R Development Core Team, 2017). Linear regression models (Aalen, 1989) were used to show the relationship between ploidy level and estimated genetic population parameters (Shannon index, Simpson index, Nei index, and number of

bands). We removed an outlier (PR1-MX) to fit the linear model in two parameters (Shannon index and Simpson index) and met model assumptions.

To infer the population structure of the individuals studied, we used two approaches: STRUCTURE (Pritchard *et al.*, 2000) and Discriminant Analysis of Principal Components (DAPC; Jombart *et al.*, 2010). We used the software STRUCTURE to determine the approximate number of genetic clusters within the full data set and to assign individuals to the most appropriate cluster. STRUCTURE is well-suited for analyzing polyploids because it performs "soft" instead of "hard" clustering and individuals can be partly assigned to several clusters (Meirmans *et al.*, 2018). The STRUCTURE analysis was performed using an algorithm based on cluster models, following the Bayesian approach. For the model that allows mixing and ten interactions, we tested 250,000 replicas for burn-in and 1,000,000 MCMC iterations. The number of clusters (K) tested ranged from 1 to 15, and the final graphs were viewed on the CLUMPAK server (Kopelman *et al.*, 2015) and plotted on *strataG* v0.9.4 (Archer *et al.*, 2016) package in the R environment. The best k values as determined by Evanno's Delta K were also evaluated in CLUMPAK server.

We performed a DAPC using the R package *adegenet* (Jombart, 2008). DAPC uses a nonparametric approach, free from Hardy-Weinberg constraints, and can be applied to polyploid data sets with mixed ploidy populations (Meirmans *et al.*, 2018). We conducted two approaches: (1) the first DAPC analysis was performed providing information for 12 groups (populations of *P. cattleyanum*) and (2) the number of clusters was assessed using the function *find.clusters*. We assumed 25 as the maximum number of clusters. The optimal number of clusters was estimated using the Bayesian information criterion (BIC), and both DAPC results were presented as bar plots.

To detect population differentiation, analysis of molecular variance (AMOVA) was performed using the R package *poppr* (Kamvar *et al.*, 2014). We investigated the partitioning of genetic variance within and among populations (1) or within and among populations, as well as among groups pre-defined according to ploidy level (2). Before we performed the AMOVA analysis, we applied the *clonecorrect* function to our data matrix to remove potential bias caused by cloned genotypes. The estimates of *Gst global* (Nei and Chesser, 1983) and pairwise *Gst* for populations were calculated using the *polysat* package (Clark and Jasieniuk, 2011).

Climatic niche analysis

To test whether cytotypes occupy different environmental niches, we used the same occurrence points of the sampled cytotypes for genetic analysis. When a population had two or

more cytotypes, we used the same GPS point of population for each cytotype. A total of 20 climatic variables were defined in Worldclim 2.0 (Fick and Hijimans, 2017) in a spatial resolution of approximately 1 km², which were 19 bioclimatic variables and a monthly solar radiation. To eliminate the collinearity between these variables, we performed a principal component analysis (PCA) in the ENMTML package (Andrade *et al.*, 2020), implemented in R, where the components that represent 95% of the data variation were transformed into *rasters* used to generate models.

We generated ecological niche models using the PCA *rasters* and cytotype occurrence data (Fig. 6) using the ModleR package (Sánchez-Tapia *et al.*, 2018). We used the following parameters: partition cross validation with five partitions; buffer type - max; number of background points - 1000; algorithms - Maxent, Bioclim, Domain and Mahalanobis. The final models were obtained with the default parameters (consensus level: 0.5, majority consensus approach) to generate both the binary consensus models and the ensemble. From the binary ensemble models, we estimated the overlap of the ecological niche of the cytotypes using the Schoener D (Schoener, 1968) and Warren I (Warren *et al.*, 2008) indices. Both indices range from 0 (no niche overlap) to 1 (total niche overlap). This analysis was performed using the phyloclim package (Heibl and Calenge, 2018) also implemented in R platform. Values of D and I less than or equal to 0.5 indicate low niche overlap, and these results only mean that the niches are not identical, but do not indicate that they are different (from Luis *et al.*, 2018).

To test the ecological niche equivalence (Warren *et al.*, 2008), we cut the *rasters* generated by the PCA with the delimitation of the Atlantic Forest. This test is based on Warren *et al.* (2008) and estimates through random permutations whether the compared niches are equivalent. This test was performed with the *ecospat.niche.equivalency.test* function of the *ecospat* package (Broennimann *et al.*, 2018), also implemented in R, using the lower parameter, which tests the niche divergence, that is, that the overlap of cytotype niche is less equivalent than expected at random, performing 1000 repetitions.

Through random permutations of niches according to the environmental conditions available in the area, the niche similarity test estimates whether the niches compared are less similar than expected at random. This test compares the overlap observed between the cytotypes with the overlap expected at random if one or both choose the habitat at random within its geographic range. This test was performed with the *background.test* function of the *ecospat* package (Broennimann *et al.*, 2018) in R, using the *lower* parameter to test the niche divergence, that is, if the niche overlap of the cytotypes is less similar to that expected at random, performing 1000 repetitions.

Estimation of environmental differences

To verify differences in the environment of the cytotypes, we selected the least correlated variables of the model ($r^2 < 0.7$) using the *removeCollinearity* function of the *virtualspecies* package (Leroy *et al.*, 2015). We used the same occurrence points of genetic analysis to extract the values of the environmental variables, using the *extract* function of the *raster* package (Hijmans, 2015), to perform a Principal component analysis, as indicated in the package *FactoMineR* (Le *et al.*, 2008).

Associative test

We also evaluated if the environmental or geographical distance could explain the genetic data of the *P. cattleyanum* populations. First, we used the same PCA for the environmental variables from the modelling to extract the values of the population's coordinates using package *Raster*. We used the averages of the principal components to calculate Euclidean distances between populations. After, with the distance matrices of genetic and geography used in other analyses (see above) and the environmental distance matrices, we conducted a Multiple Matrix Regression with Randomization (MMRR), performed with the MMRR function in R (Wang *et al.*, 2013) using 10,000 permutations. The comparisons among genetic distance with both geography and environmental distance were tested.

RESULTS

Population diversity and structure

From the 10 microsatellite loci *amplified*, we detected a total of 715 SSR bands among the 12 *P. cattleyanum* populations (Table 2). The number of bands per population ranged from 25 to 106, and private bands ranged from zero to nine, with a total of 43 (Table 2). The genetic diversity parameters of Shannon index, Simpson index and Nei index estimated for each population ranged from 0.769 to 3.401, from 0.487 to 0.967, and from 0.005 to 0.130, respectively (Table 2). Analyses of genetic diversity reveled high genetic diversity in populations with cytotypes 2n = 44 (Shannon index = 3.401, Simpson index = 0.967, Nei index = 0.11 to 0.13). Linear regression models pointed to a reduction of genetic diversity parameters estimated with the increase of ploidy levels for all parameters analyzed (Fig.1A- $R^2 = 0.80$, $p < 0.001$; B- $R^2 = 0.60$, $p < 0.002$; C- $R^2 = 0.78$, $p < 0.001$; D- $R^2 = 0.73$, $p < 0.001$).

In four populations of *P. cattleyanum*, all individuals had distinct multilocus genotypes (PR2-4X, SP1-4X, SP2-4X and SP3-4X; Table 2). In the other five populations, some individuals shared the same MLG (SP4-6X, RJ1-8X, RJ2-7X, RJ3-MX and ES2-12X; Table

2). Three populations exhibited a low quantity of genotypes (PR1-MX, ES1-12X and BA-10X; Table 2).

STRUCTURE analysis identified K = 11 as the optimal number of genetic clusters (Fig. 2, K means is available at Supplementary Data Fig. S1). *P. cattleyanum* populations with cytotypes 2n = 44 (PR2-4X, SP1-4X, SP2-4X and SP3-4X) exhibited high levels of admixture. Some *P. cattleyanum* populations exhibited low levels of admixture PR1-MX, SP4-6X, RJ2-7X and BA-10X, except the populations ES1-12X/ES2-12X and RJ1-8X/RJ3-MX that shared the same cluster but diverged from other populations. The results of the two DAPC analyses are presented as bar plots in Fig. 3 and Supplementary Data Fig. S3. We decided to use the first approach to compare with STRUCTURE, since the allocation of individuals to clusters from DAPC was very similar to those achieved by STRUCTURE. In DAPC analysis, the delimitation of populations was clearer than STRUCTURE. In general, in both analyses, most populations were grouped as independent genetic clusters or consisted of either a single population or of two nearby populations, indicating low gene flow among populations.

The AMOVA showed that a large fraction of the variation was partitioned among populations in both analyses (ϕ_{ST} = 0.49 and 0.51, $p < 0.001$; Table 3), suggesting restricted gene flow among population and cytotypes groups. The analysis showed significantly higher levels of structure among the cytotypes groups (ϕ_{CT} = 0.22, $p < 0.001$; Table 3). The genetic divergence values of *P. cattleyanum* in this study were high: 0.20 for *Gst*. A heatmap of pairwise estimates of *Gst* to each population is presented in Supplementary Data Fig. S4. The most divergent populations were ES1-12X/RJ2-7X and least divergent populations were SP1-4X/ SP2-4X and SP1-4X/ PR2-4X.

Estimation of environmental differences

The PCA showed that environmental variables with most contributions were related to solar radiation (January and March), temperature (mean temperature annual, in the warmest quarter, and temperature annual range), and precipitation (annual, seasonality, of wettest quarter and of coldest quarter) (Fig. 4). The individual analysis of each variable [Supplementary Data Fig. S5] shows that the distribution of individuals of *P. cattleyanum* with higher ploidy levels (decaploid and dodecaploid) is influenced by extreme environmental conditions (higher index of solar radiation and temperature, and less precipitation) when compared to other cytotypes.

Associative test

The result of MMRR pointed to a combined effect of geographic and environmental distances influencing the genetic distances ($R^2 = 0.61$; $p < 0.001$). However, when compared separately, the linear regression showed that geographic distances contributed more to describe genetic distances ($R^2 = 0.59$; $p < 0.001$, Fig. 5-A), while environmental distances contributed weakly and not significantly ($R^2 = 0.024$; $p = 0.2$; Fig. 5-B).

Climatic niche analysis

The cytotypes with some level of niche overlap ($D > 0.5$) were octaploids with tetraploids and octaploids with hexaploids (Fig 6 C and Fig 6 D). The cytotypes with the highest levels of ploidy (deca and dodecaploid) showed no overlap between them or with the other cytotypes (Fig 6 C and Fig 6 D). The results of the equivalence tests indicate that none of the cytotype niches are equivalent, except for the Heptaploid and Octaploid niches [Supplementary Data Figs. S6 and S7 A-C]. When accounting for the surrounding environmental conditions, our test indicated that none of the cytotypes are similar [Supplementary Data Figs. S7 D-I and S8].

DISCUSSION

In this study, we integrated molecular data and climatic niche analysis to provide insights into the influence of polyploidy on divergences between lineages. Using cytotypes of *P. cattleyanum* as a model (4x, 6x, 7x, 8x, 10x and 12x), we tested two hypotheses on polyploid speciation: (1) cytotypes act as discrete evolutionary units, promoting divergence between populations and interruption of gene flow, and (2) polyploidy promotes habitat divergence, both of which are factors that drive speciation. Through these independent approaches, our results suggest divergence between populations, which present distinct ploidy levels. The data indicates strong populational genetic structure, suggesting that results should involve mechanisms that produce reproductively isolated lineages. Our results also pointed to low niche similarity between cytotypes, corroborating those obtained through molecular analyzes. Cytotypes with the highest ploidy (deca and dodecaploid) were the only cytotypes that did not overlap with any of the other cytotypes and were more genetically distant than the other cytotypes, with the lowest genetic diversity. Thus, we can corroborate both hypotheses about the effect of polyploidy on the early stages of genetic and environment niche divergence between lineages.

Our study is the first to include natural populations of *P. cattleyanum* along its natural distribution, using species-specific SSR markers developed by Machado *et al.* (2020) and considering the ploidy level of the populations. In this study, the genetic diversity patterns of

P. cattleyanum cytotypes were determined in an ecological context using environmental niche modeling (ENM).

Psidium cattleyanum genetic diversity decreased as the ploidy level increased

It is very challenging to study population genetics and polyploid evolution (Jighly *et al.*, 2019). Many aspects of polyploids organisms can be determinant in genetic diversity and population structure, e.g., changing mating system, multiple origins of cytotypes, the gene flow between polyploids with different origins and cytotypes within populations (Baduel *et al.*, 2018; Meirmans *et al.*, 2018; Baniaga *et al.*, 2020). In our study, the populations of *P. cattleyanum* with high ploidy levels showed fewer bands and private bands (and consequently mean bands), while in Tuler *et al.* (2019) showed that polyploid species in *Psidium* have higher mean alleles. The use of species-specific SSR markers could be more accurate for measuring genetic variability (Queirós *et al.*, 2015; Machado *et al.*, 2020), which could explain the divergent results between our study and Tuler *et al.* (2019). Previously, most of the efforts to quantify the genetic diversity in *Psidium* have been concentrated on *P. guajava* (Coser *et al*., 2012; Noia *et al.*, 2017; Kherwar *et al.*, 2018).

The reproductive strategies of *P. cattleyanum* can help explain the genetic patterns uncovered herein. Although the diploid cytotype (2x) has not yet been discovered (Machado *et al.*, 2020, Tuler *et al.*, 2019), our SSR data suggests that most of the *P. cattleyanum* populations analyzed here could be highly selfing and/or apomictic. We found populations with low levels of genetic diversity, low numbers of MGL and high structuration (PR1-MX, SP4-6X, RJ1-8X, RJ2-7X, ES1-12X and BA-10X). On the other hand, we also found populations composed by tetraploids (2n=4x=44) with high levels of genetic diversity, genotypes (MLG), and number of bands and private bands when compared to populations with the highest cytotypes, suggesting sexual reproduction. For *Potentilla puberula* Krašan (Rosaceae), the diploid and tetraploid cytotypes are sexual, and cytotypes with greater ploidy are apomictic (Dobes *et al.*, 2013, Van Drunen and Husband, 2019). The relationship between polyploidy and reproductive strategies has been investigated in other groups of plants (Kólar *et al.*, 2017; Van Drunen and Husband, 2019). In general, diploids and cytotypes with lower ploidies have sexual reproduction, while the highest ploidy levels could be apomictic (Dobes *et al.*, 2013).

Based on MLG analysis, we hypothesized that the populations PR1-MX, ES1-12X and BA-10X were probably formed by a small number of generating individuals (founder effect), and propagated via asexual reproduction. The populations of Uruguay showed apomixis in

ploidy levels 7x and 8x (Souza-Pérez and Speroni, 2017) and low genetic diversity (Souza-Pérez in personal communication). Polyploidy is frequently associated with the invasion of novel habitats and expansion of geographic distribution (Soltis and Soltis, 2000; Pandit *et al.*, 2011; te Beest *et al.*, 2011; Baduel *et al.*, 2018; Moura *et al.* 2021). The populations PR1-MX and BA-10X are the extreme southernmost and northernmost in the distribution range addressed in this study, but the ES1-12X/RJ2-7X ones are more genetically divergent than the other pairwise combinations, showing that polyploid promotes divergence between populations (Supplementary Data Fig. S4).

The colonization of new habitats typically proceeded by a small number of founders and tolerance to selfing is greatly important (Baduel *et al.*, 2018). Since apomixis and self-compatibility are closely associated with polyploidy (deWet, 1980), such characteristics could be an advantage that assures reproductive success for the dispersed invaders (Baduel *et al.*, 2018). The shift to apomixis could prevent the minor cytotype disadvantages (Levin, 1975), and also facilitate the genotypes of *P. cattleyanum* better adapted to local environmental conditions. The facultative apomixis influencing the colonization of higher elevations and range expansions of *Ranunculus kuepferi* cytotypes has been reported by Schinkel *et al.* (2016). In the long term, asexual reproduction reduces genetic differences among individuals within populations and increases the genetic differentiation among populations (Zhang *et al.*, 2019), decreasing gene flow. Structure and DAPC analyses showed strong genetic structure at population levels and pointed to an absence of gene flow between cytotypes. Many of the populations analyzed in this study are differentiated by polyploid cytotypes, which also contributes to the genetic differentiation of *P. cattleyanum*. Therefore, polyploidization affects the genetic structure of species, as well as the ecological processes (Baniaga *et al.*, 2020).

Ecological and geographical factors can also reduce gene flow leading to population divergence (Wang *et al.*, 2013). Our MMRR results showed a significant effect of combining environmental and geographic distances. The high influence of geographic distances in *P. cattleyanum* populations and the considerable genetic differentiation (Gst= 0.20; AMOVA-population $\phi_{ST} = 0.49$ and cytotypes $\phi_{ST} = 0.51$) observed in the present study confirm that gene flow is restricted among *P. cattleyanum* populations, and consequently between cytotypes. Thus, the populations of *P. cattleyanum* differentiated by cytotypes act as discrete evolutionary units, and this species provides insights about the earliest stages of the speciation process.

Although having higher levels of genetic and genomic diversity is a characteristic of polyploid organisms (Haldane, 1930; Soltis and Soltis, 2000; Schifino-Wittmann, 2004; Gao, 2019; Lavania, 2020), our results from the linear regression pointed to a significant decrease in

genetic diversity with increased ploidy for all diversity indexes (index Shannon, Simpson and Nei) and the number of bands per population. These results differ from those observed for other polyploid species (Brochmann *et al.*, 2004; Garcıa-Verdugo *et al.*, 2009). The same patterns of *P. cattleyanum* were obtained by Burnier *et al.* (2009) with AFLP markers. The polyploidy populations, which are probably apomictic, had a greater capability for colonization in previously glaciated areas, but showed reduced genetic diversity, probably due to the bottleneck effect during colonization. The highest ploidy levels of *Prunus lusitanica* L. (Rosaceae) having low levels of genetic variation across broad geographical areas has also been reported by Garcia-Verdugo *et al.* (2013). In *P. cattleyanum*, the lower rates of genetic diversity in high ploidy cytotypes may be related to three factors: the species' asexual reproduction characteristics, geographic isolation, and intraspecific divergence of environmental niche between cytotypes to expand the geographical range (see next topic). These results contribute to the discussion regarding the importance of polyploidy for the genetic diversity of species and, consequently, divergence between lineages.

The niche differentiation between cytotypes of P. cattleyanum

In the *P. cattleyanum* polyploid complex, we observed low similarity and a significant divergence of ecological niche between cytotypes, suggesting segregation of ploidy probably related to an ecological specialization of exploiting environmental resources. The niche overlap between cytotypes octaploids/tetraploids, and octaploids/hexaploids, suggests that octaploid cytotypes could have multiple origins as discussed by López-Jurado *et al.* (2019) for tetraploids of *Dianthus broteri* Boiss. & Reut. complex (Caryophyllaceae). Niche differentiation between cytotypes has already been described for other polyploid taxa (Marchant *et al.*, 2016; Visger *et al.*, 2016; Karunarathne *et al.*, 2018; López-Jurado *et al.*, 2019; Decanter *et al.*, 2020). Baniaga *et al.* (2020) found that polyploids are often more climatically differentiated from their diploid parents. Decanter *et al.* (2020) showed that high ploidy cytotypes of *Saxifraga rosacea* Moench (Saxifragaceae) were associated with a greater tolerance to extreme conditions. The diploid species of *Psidium* show a restricted geographical distribution in the Atlantic Forest and the polyploid species occur in several biomes in Brazil (Tuler *et al.*, 2019), we hypothesize that several ploidy levels of *P. cattleyanum* could facilitate the occupation of new ecological niches, enabling species to occupy different environmental conditions in Brazil.

Our data shows that extreme environmental conditions could predict higher polyploidy occurrence in *P. cattleyanum*. Polyploids tend to increase their frequency in habitats affected

by environmental disturbances (Baduel *et al.*, 2018). In this species, the high ploidy cytotypes (11x and 12x) are associated with tolerance to higher solar incidence and temperature and less precipitation and occupying a distinct ecological niche. The same pattern was observed for *Dianthus broteri* complex (2x, 4x, 6x and 12x) by López-Jurado et al. (2019): the higher ploidy levels were found in different and nonoverlapping niches, with each cytotype occupying a distinct ecological niche (López-Jurado *et al.*, 2019). The effects of polyploidization events on ecological niche remain largely unknown (Marchant *et al.*, 2016). Obtaining empirical data about similarity and divergence of ecological niche use in polyploid systems could make it easier to develop a unified theory about the ecological consequences of polyploidy (Parisod and Broennimann, 2016).

Diploids and tetraploids can respond differently to local environmental factors (Certner *et al.*, 2018). For wild potatoes (*Solanu*m sect. *Petota* Dumort., Solanaceae), diploids and triploids tend to occur in warmer and drier areas, whereas higher-level polyploids tend to occur in relatively cold areas (Hijmans *et al.*, 2007). Polyploidy has played an important role in environmental differentiation and range expansion (Otto and Whitton, 2000; Hijmans *et al.*, 2007). The environmental variables that contribute the most to the distribution of *P. cattleyanum* cytotypes (temperature, precipitation, and solar radiation) could increase the production of unreduced gametes. The formation of higher ploidy levels is frequently associated with the union of unreduced gametes (deWet, 1980).

Environmental abiotic stresses, such as temperature, water and plant nutrition, have influenced the formation of unreduced gametes (Ramsey and Schemske, 1998). In *Brachypodium distachyon* (L.) P. Beauv. (Poaceae) (Manzaneda *et al.*, 2012), aridity is an important predictor of polyploid occurrence. In *Ranunculus auricomus* L. (Ranunculaceae), the exposure to extended light periods changed the proportion of seed formation, with diploid cytotypes presenting mostly sexual seed formation, while polyploid cytotypes formed predominantly apomictic seeds (Ulum *et al.*, 2020). The influence of solar radiation on populations of *P. cattleyanum* with 11x and 12x ploidy may have formed and maintained these higher levels of ploidy, and consequently, resulted in a divergent ecological niche and greater tolerance to more extreme conditions for *P. cattleyanum*.

For *P. cattleyanum*, our data suggests that polyploidy is a process driving intraspecific diversification. The hypothesis about possible speciation processes has been discussed by Pedrosa-Macedo and Smith (2007), based on species' morphological variation. The absence of gene flow combined with the influence of geographic distance and intraspecific divergence of cytotypes niches, showed that cytotypes are a step towards intraspecific divergence. Also, the

several ploidy levels and reproductive traits of *P. cattleyanum* increase potential colonization, and these two characteristics were decisive for maintaining populations in different environmental conditions.

In this study, we detected low gene flow, and high rates of differentiation between populations/cytotypes, and niche intraspecific divergence. *Psidium cattleyanum* populations are considered a good model for demonstrating the influence of polyploidy as an intermediate stage in divergence between lineages. The confirmation of our hypothesis indicates that cytotypes act as discrete evolutionary units and points to a possible divergence among lineages in the future. Our results reveal how polyploidy affects the genetic diversity of taxa and promotes habitat divergence in the *P. cattleyanum* polyploid complex, providing new information for the discussion about the impact of polyploidy on ecological niches. Historical processes must be evaluated in the future to clarify the patterns found for *P. cattleyanum* cytotypes.

SUPPLEMENTARY DATA

Supplementary data are available below.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and FAEPEX/UNICAMP for the financial support. FAO received a Post Doctoral fellowship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2018/18527-9). We thank Dr. A. V. Scatigna for his support in fieldwork and collecting plant material from the individual in Uruguaiana (RS), B. Leal for being available to answer questions at any time, G. M. Serbin and N. S. Streher for critical comments on previous versions of this manuscript. We also thank the staff of Laboratório de Análises Genéticas e Moleculares (LAGM/UNICAMP), A. Alves-Pereira and Y. A. Moura for providing the R codes for genetic analyses and statistical support.

LITERATURE CITED

Aalen OO. 1989. A linear regression model for the analysis of life times. *Statistics in medicine* 8: 907-925.

Andrade AFA, Velazco SJE, Júnior PDM. 2020. ENMTML: An R package for a straightforward construction of complex ecological niche models. *Environmental Modelling & Software* 125: 104615. doi.org/10.1016/j.envsoft.2019.104615

Archer FI, Adams PE, Schneiders BB. 2017. strataG: An r package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* 17: 5– 11.

Baduel P, Bray S, Vallejo-Marin M, Kolár F, Yant L. 2018. The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Frontiers in Ecology and Evolution 6*: 117. doi: 10.3389/fevo.2018.00117

Baniaga AE, Marx HE, Arrigo N, Barker MS. 2020. Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecology letters* 23: 68-78.

Briggs D, Walters SM. 1997. Plant variation and evolution. Cambridge: Cambridge University Press.

Brochmann C, Brysting AK, Alsos IG, *et al*. 2004. Polyploidy in arctic plants. *Biological journal of the Linnean society* 82: 521-536.

Broennimann O, Di Cola V, Petitpierre B, *et al.* 2014. Package 'ecospat'. Available at: https://cran. r-project. org/web/packages/ecospat/index. html

Burnier J, Buerki S, Arrigo N, Küpfer P, Alvarez N. 2009. Genetic structure and evolution of Alpine polyploid complexes: Ranunculus kuepferi Ranunculaceae as a case study. *Molecular Ecology* 18: 3730-3744.

Čertner M, Sudová R, Weiser M, Suda J, Kolár F. 2019. Ploidy-altered phenotype interacts with local environment and may enhance polyploid establishment in Knautia serpentinicola Caprifoliaceae. *New Phytologist* 221: 1117-1127.

Clark LV, Jasieniuk M. 2011. POLYSAT: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources* 11: 562-566.

Coser SM, Ferreira MFS, Ferreira A, Mitre LK, Carvalho CR, Clarindo WR. 2012. Assessment of genetic diversity in Psidium guajava L. using different approaches. *Scientia Horticulturae* 148: 223-229.

Costa IR, Forni-Martins ER. 2006. Chromosome studies in Brazilian species of Campomanesia Ruiz & Pávon and Psidium L. Myrtaceae Juss. *Caryologia* 59: 7-13.

Decanter L, Colling G, Elvinger N, Heiðmarsson S, Matthies D. 2020. Ecological niche differences between two polyploid cytotypes of Saxifraga rosacea. *American Journal of Botany* 107: 423-435.

DeWet JMJ. 1980. Origins of polyploids. In: Lewis HW, ed. *Polyploidy*. Springer: Boston, 3-15.

Dobeš C, Milosevic A, Prohaska D, Scheffknecht S, Sharbel TF, Hülber K. 2013. Reproductive differentiation into sexual and apomictic polyploid cytotypes in Potentilla puberula Potentilleae, Rosaceae. *Annals of botany* 112: 1159-1168.

Hijmans RJ, Fick SE. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology* 37: 4302-4315.

Gao J. 2019. Dominant plant speciation types. A commentary on: 'Plant speciation in the age of climate change'. *Annals of botany* 124: iv-vi.

García-Verdugo C, Calleja JA, Vargas P, Silva L, Moreira O, Pulido F. 2013. Polyploidy and microsatellite variation in the relict tree P runus lusitanica L.: how effective are refugia in preserving genotypic diversity of clonal taxa? *Molecular Ecology* 22: 1546-1557.

García-Verdugo C, Fay MF, Granado-Yela C, *et al.* 2009. Genetic diversity and differentiation processes in the ploidy series of Olea europaea L.: a multiscale approach from subspecies to insular populations. *Molecular Ecology* 18: 454-467.

García-Verdugo C, Sajeva M, La Mantia T, Harrouni C, Msanda F, Caujapé-Castells J. 2015. Do island plant populations really have lower genetic variation than mainland populations?

Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology* 24: 726-741.

Global Invasive Species Database. 2021. Available at: http://www.iucngisd.org/gisd/search.php on 21-01-2021

Grant V. 1981. Plant speciation. Columbia: Columbia University Press.

Haldane JBS. 1930. Theoretical genetics of autopolyploids. *Journal of Genetics* 22: 359-372.

Heibl C, Calenge C, Heibl MC. 2018. Package 'phyloclim'. Available at: http://rsync.udc.es/CRAN/web/packages/phyloclim/phyloclim.pdf

Hijmans RJ, Gavrilenko T, Stephenson S, Bamberg J, Salas A, Spooner DM. 2007. Geographical and environmental range expansion through polyploidy in wild potatoes Solanum section Petota. *Global ecology and biogeography* 16: 485-495.

Hijmans RJ, Van Etten J, Cheng J, *et al.* 2015. Package 'raster'. Available at: https://cran.r-project.org/web/packages/raster/raster.pdf

Hodel RG, Segovia-Salcedo MC, Landis JB, *et al.* 2016. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Applications in plant sciences* 4: 1600025. doi.org/10.3732/apps.1600025

Jighly A, Abbott RJ, Daetwyler HD. 2019. Editorial: Polyploid Population Genetics and Evolution-From Theory to Practice. *Frontiers in Ecology and Evolution* 7: 460. doi: 10.3389/fevo.2019.00460

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403-1405. doi: 10.1093/bioinformatics/btn129

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* 11: 94.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer J* 2: e281. doi:10.7717/peerj.281

Karunarathne P, Schedler M, Martínez EJ, Honfi AI, Novichkova A, Hojsgaard D. 2018. Intraspecific ecological niche divergence and reproductive shifts foster cytotype displacement and provide ecological opportunity to polyploids. *Annals of botany* 121: 1183-1196.

Kherwar D, Usha K, Mithra SA, Singh B. 2018. Microsatellite SSR marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava Psidium guajava L. *Journal of Plant Biochemistry and Biotechnology* 27: 284-292

Kolár F, Čertner M, Suda J, Schönswetter P, Husband BC. 2017. Mixed-ploidy species: progress and opportunities in polyploid research. *Trends in Plant Science* 22: 1041-1055.

Lavania UC. 2020. Plant speciation and polyploidy: in habitat divergence and environmental perspective. *Nucleus* 63: 1-5.

Lê S, Josse J, Husson F. 2008. FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 25: 1-18.

Leroy B, Meynard CN, Bellard C, Courchamp F. 2016. virtualspecies, an R package to generate virtual species distributions. *Ecography* 39: 599-607.

Levin DA. 2019. Plant speciation in the age of climate change. *Annals of Botany* 124: 769-775.

Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24: 35-43.

López-Jurado J, Mateos-Naranjo E, Balao F. 2019. Niche divergence and limits to expansion in the high polyploid Dianthus broteri complex. *New Phytologist* 222: 1076-1087.

Machado RM, de Oliveira FA, de Matos Alves F, de Souza AP, Forni-Martins ER. 2020. Population Genetics of Polyploid Complex Psidium cattleyanum Sabine Myrtaceae: Preliminary Analyses Based on New Species-Specific Microsatellite Loci and Extension to

Other Species of the Genus. *Biochemical Genetics*, in press. doi.org/10.1007/s10528-020-10002-1

Mani A, Mishra R, Thomas G. 2011. Elucidation of diversity among Psidium species using morphological and SPAR methods. *Journal of Phytology* 3: 53-61.

Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. 2012. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in Brachypodium distachyon Poaceae. *New phytologist* 193:797-805.

Marchant DB, Soltis DE, Soltis PS. 2016. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytologist* 212: 708-718.

Marques A, Tuler AC, Carvalho CR, Carrijo T, da Silva Ferreira MF, Clarindo W. 2016. Refinement of the karyological aspects of *Psidium guineense* Swartz, 1788: a comparison with *Psidium guajava* Linnaeus, 1753. *Comparative Cytogenetics* 101: 117-128.

Meirmans PG, Liu S, Van Tienderen PH. 2018. The analysis of polyploid genetic data. *Journal of Heredity* 109: 283-296.

Moura RF, Queiroga D, Vilela E, Moraes AP. 2021. Polyploidy and high environmental tolerance increase the invasive success of plants. *Journal of Plant Research* 134: 105–114

Nei M. 1978. Estimation of average heterozygosity and genetic distance from small number individuals. *Genetics* 89: 583-590.

Nei M, Chesser RK. 1983. Estimation of fixation indices and gene diversities. *Annals of human genetics* 47: 253-259.

Noia LR, Tuler AC, Ferreira A, Ferreira MFS 2017. Relationship between Psidium species Myrtaceae by resistance gene analog markers: focus on nematode resistance. *Genetics and Molecular Research* 16: 1-13.

Normand F, Habib R. 2001. Phenology of strawberry guava Psidium cattleianum in ReÂunion Island. *The journal of Horticultural Science and Biotechnology* 76: 540-545.

Otto SP, Whitton J. 2000. Polyploidy: incidence and evolution. *Annual Review of Genetics* 34: 401–437.

Pandit MK, Pocock MJ, Kunin WE. 2011. Ploidy influences rarity and invasiveness in plants. *Journal of Ecology* 99: 1108-1115.

Parisod C, Broennimann O. 2016. Towards unified hypotheses of the impact of polyploidy on ecological niches. *New Phytologist*, 212: 540-542.

Pedrosa-Macedo JH, Smith CW. 2007. Histórico dos Estudos do Projeto Araçá. In: Pedrosa-Macedo JH, DalMolin A, Smith CW, eds. *O Araçazeiro: Ecologia e Controle Biológico*. Curitiba: FUPEF, 01-06.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Psidium in Flora do Brasil 2020 em construção. Jardim Botânico do Rio de Janeiro. Avaible at http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB10858

Queirós J, Godinho R, Lopes S, *et al.* 2015. Effect of microsatellite selection on individual and population genetic inferences: an empirical study using cross-specific and species-specific amplifications. *Molecular ecology resources* 15: 747-760.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* 29: 467-501.

Raseira MCB, Raseira A. 1996. Contribuição ao estudo do araçazeiro. Pelotas: Embrapa-Centro de Pesquisa Agropecuária de Clima Temperado.

Rezende L, Suzigan J, Amorim FW, Moraes AP. 2020. Can plant hybridization and polyploidy lead to pollinator shift? *Acta Botanica Brasilica* 34: 229-242.

Rieseberg LH, Willis JH. 2007. Plant speciation. S*cience* 317: 910-914.

Sánchez-Tapia A, de Siqueira MF, Lima RO, *et al.* 2017. Model-R: a framework for scalable and reproducible ecological niche modeling. In: Mocskos E, Nesmachnow S, eds *Latin American High Performance Computing Conference*. California: Springer, 218-232.

Schifino-Wittmann, MT. 2004. Poliploidia e seu impacto na origem e evolução das plantas silvestres e cultivadas. *Current Agricultural Science and Technology* 10: 151-157.

Schinkel CC, Kirchheimer B, Dellinger AS, *et al.* 2016. Correlations of polyploidy and apomixis with elevation and associated environmental gradients in an alpine plant. *AoB Plants* 8: plw064.

Schoener TW. 1968. Sizes of feeding territories among birds. *Ecology* 49: 123-141.

Shannon CE. 1948. A mathematical theory of communication. *Mobile Computing and Communications Review* 5: 3-55.

Silveira RM, Machado RM, Forni-Martins ER, Verola CF, Costa IR. 2016. Environmental variations drive polyploid evolution in neotropical Eugenia species Myrtaceae. *Genetics and Molecular Research* 15: gmr15048842

Simpson EH. 1949. Measurement of diversity. *Nature* 163: 688-688.

Soltis DE, Soltis PS, Schemske DW, *et al.* 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56: 13-30.

Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences* 97: 7051-7057.

Souza-Pérez M, Speroni G. 2017. New apomictic pathway in Myrtaceae inferred from Psidium cattleyanum female gametophyte ontogeny. *Flora* 234: 34-40.

Stace CA. 1991. Plant taxonomy and biosystematics. Cambridge: Cambridge University Press.

te Beest M, Le Roux JJ, Richardson DM, *et al*. 2011. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109: 19-45.

Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y. 1999. Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera Hylocereus and Selenicereus Cactaceae. *Plant Molecular Biology Reporter* 17: 249-254.

Tuler AC, Carrijo TT, Nóia LR, Ferreira A, Peixoto AL, da Silva Ferreira MF. 2015. SSR markers: a tool for species identification in Psidium Myrtaceae. *Molecular biology reports* 42: 1501-1513.

Tuler AC, Carrijo TT, Peixoto AL, *et al*. 2019. Diversification and geographical distribution of Psidium Myrtaceae species with distinct ploidy levels. *Trees* 33: 1101-1110.

Ulum FB, Costa Castro C, Hörandl E. 2020. Ploidy-dependent effects of light stress on the mode of reproduction in the Ranunculus auricomus complex Ranunculaceae. *Frontiers in Plant Science* 11: 104. doi.org/10.3389/fpls.2020.00104

Van Drunen WE, Husband BC. 2019. Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. *New Phytologist* 224: 1266-1277.

Visger CJ, Germain-Aubrey CC, Patel M, Sessa EB, Soltis PS, Soltis DE. 2016. Niche divergence between diploid and autotetraploid Tolmiea. *American Journal of Botany* 103: 1396-1406.

Wang IJ, Glor RE, Losos JB. 2013. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology letters* 16: 175-182.

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution: International Journal of Organic Evolution* 62: 2868-2883.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *American journal of botany* 102: 1753-1756.

Wikler C. 2007. Problemática do Araçazeiro como Planta Invasora. In: Pedrosa-Macedo JH, DalMolin A, Smith CW, eds. *O Araçazeiro: Ecologia e Controle Biológico*. Curitiba: FUPEF, 07-18.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the national Academy of sciences* 106: 13875-13879.

Zhang K, Wang X, Cheng F. 2019. Plant Polyploidy: Origin, Evolution, and Its Influence on Crop Domestication. *Horticultural Plant Journal* 5: 231-239.

Zhao S, Guo Y, Sheng Q, Shyr Y. 2014. Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC bioinformatics* 15: 1-2.

Table 1 *Psidium cattleyanum* populations used in this study. Pop= populations; N= Number of individuals collected; Lat= Latitude; Long= Longitude;

Pop	N	Ploidy level	% of Ploidy	Collection site	Lat	Long	Voucher
PR1-MX	30	Mixed	90% 7x- 7% 8x- 3% 10x	Curitiba- PR	-25.446	-49.233	*R.M. Machado 1*
PR2-4X	30	4x	100% 2n= 44	Ilha do Mel-Paranaguá- PR	-25.561	-48.304	*R.M. Machado 10*
SP1-4X	30	4x	100% 2n= 44	Ilha do Cardoso-Cananéia- SP	-25.066	-47.906	*R.M. Machado 15*
SP2-4X	30	4x	100% 2n= 44	Ilha do Comprida-SP	-24.853	-47.701	*R.M. Machado 14*
SP3-4X	30	4x	100% 2n= 44	Bertioga- SP	-23.778	-45.958	*R.M. Machado 2*

SP4-6X	30	6x	100% 6x	São Luiz do Paraitinga- SP	-23.336	-45.145	*R.M. Machado 13*
RJ1-8X	30	8x	100% 8x	Praia do Taquari- Paraty- RJ	-23.064	-44.679	*R.M. Machado 4*
RJ2-7X	30	7x	100% 7x	Praia de Trindade- Paraty- RJ	-23.338	-44.708	*R.M. Machado 17*
RJ3-MX	30	Mixed	93% 6x- 7% 7x	Serra da Bocaina- Paraty- RJ	-23.138	-44.901	*R.M. Machado 16*
ES1-12X	30	12x	100% 12x	Praia de Setiba- Guarapari- ES	-20.606	-40.421	*R.M. Machado 20*
ES2-12X	14	12x	100% 12x	Guarapari- ES	-20.606	-40.421	*R.M. Machado 20*
BA-10X	14	10x	100% 10x	Porto Seguro- BA	-16.47	-39.063	*R.M. Machado 26*
Total	328						

Table 2 Parameters of genetic diversity estimates for the *Psidium cattleyanum* populations (Pop) evaluated in this study. Population codes follow Table 1. N = number of individuals per population; NB= number of bands; PB= number of private bands; Shannon Index = Shannon diversity index (Shannon, 1948); Simpson Index = Simpson diversity index (Simpson, 1949); Nei Index = Nei diversity index (Nei, 1978); MLG = number of multilocus genotypes.

Population	N	NB	PB	Shannon Index	Simpson Index	Nei Index	MLG
PR1-MX	30	53	3	0.769	0.487	0.110	3
PR2-4X	30	85	4	3.401	0.967	0.110	30
SP1-4X	30	103	6	3.401	0.967	0.130	30
SP2-4X	30	106	9	3.401	0.967	0.118	30
SP3-4X	30	90	4	3.401	0.967	0.118	30
SP4-6X	30	61	2	2.488	0.871	0.073	18
RJ1-8X	30	39	0	3.089	0.949	0.056	24
RJ2-7X	30	50	3	2.913	0.931	0.059	22
RJ3-MX	30	36	5	3.216	0.958	0.059	26

ES1-12X	30	25	1	1.372	0.607	0.005	8
ES2-12X	14	33	1	2.168	0.867	0.041	10
BA-10X	14	34	5	1.332	0.673	0.038	5
Global	328	715	43	5.117	0.988	0.157	236

Table 3 Analysis of molecular variance (AMOVA) based on 10 microsatellite markers in *Psidium cattleyanum*. Two tests were performed: (1) the population being higher hierarchy and (2) cytotype groups being higher hierarchy. Plots of significance test are available in Suplem X.

Source of variation	d.f.	*Sum of squares*	*% variation*
Population			
Among *populations*	11	2115.38	38.68
Among *groups* within *populations*	3	34.71	10.20
Individuals within populations	313	2240.43	51.12
Total	327	4390.52	
Cytotypes groups			
Among *cytotypes*	5	1361.62	22.00
Among *groups* within *cytotypes*	9	788.46	28.57
Individuals within cytotypes	313	2240.43	49.43
Total	327	4390.52	

* ϕ population differentiation statistics:

Population: ϕ_{ST}= 0.49*, ϕ_{SC}= 0.17$^{\#}$, ϕ_{CT}= 0.38*

Cytotypes groups: ϕ_{ST}= 0.51*, ϕ_{SC}= 0.37*, ϕ_{CT}= 0.22*

*pvalue < 0.01; $^{\#}$ pvalue= 0.18

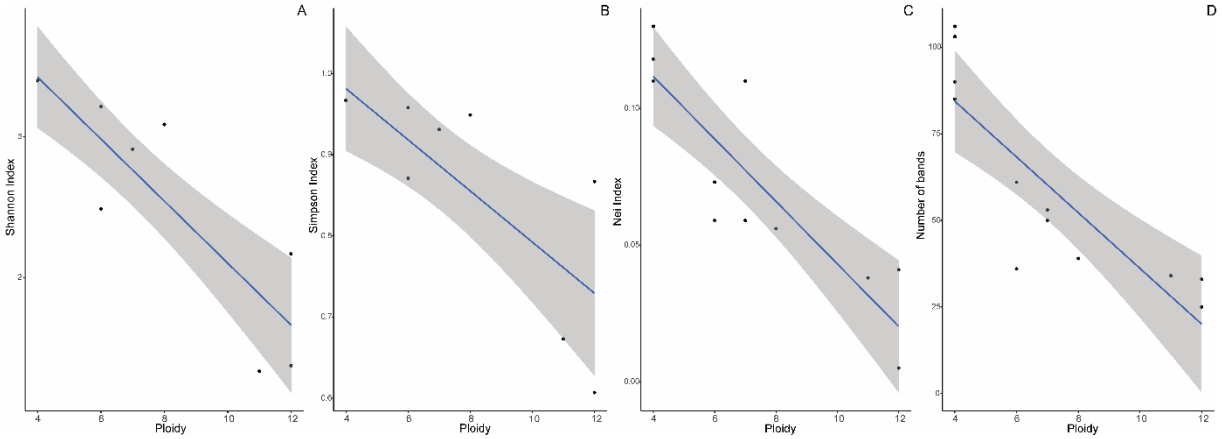


Fig 1. Linear regression models (blue line; standard deviation- gray shadow) for *Psidium cattleyanum* ploidy vs. diversity indexes: in A- Shannon index, B- Simpson index, C- Nei index, and D- Number of bands.

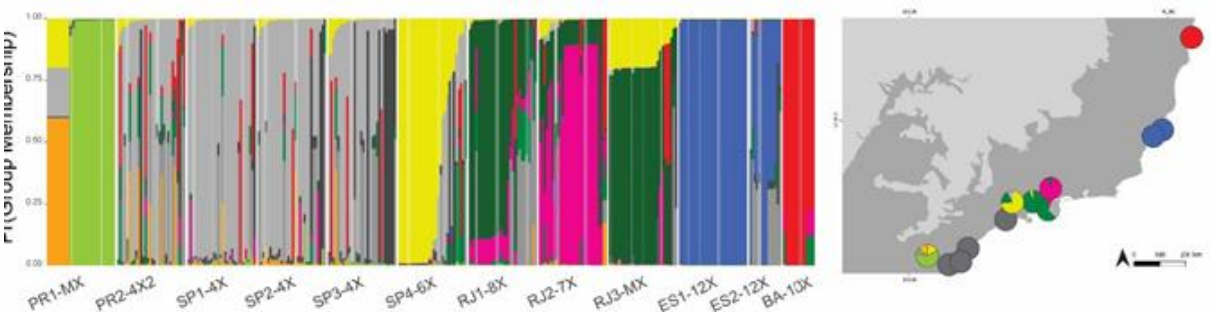


Fig 2. Bar plots of the estimated membership coefficient for each of the 328 individuals of *Psidium cattleyanum* from 12 populations based on ten microsatellite markers. The most likely value of K inferred by STRUCTURE was 11. A vertical bar represents each genotype. Codes below bars correspond to population codes from Table 1. Each color represents distinct clusters, and populations are colored in the map according to the most representative genetic cluster.

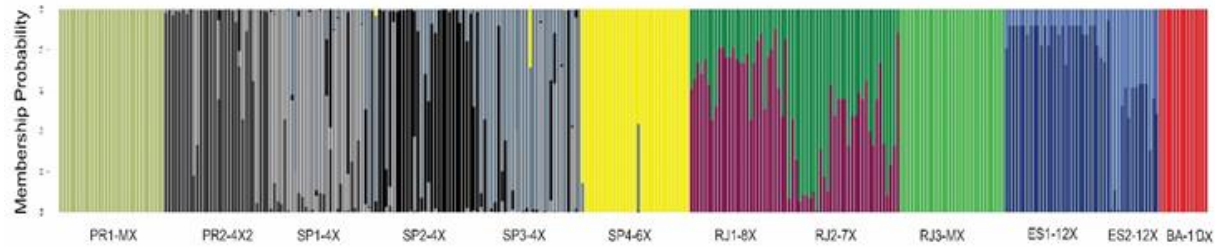


Fig 3. Discriminant analysis of principal components (DAPC) bar plots assigning 328 *Psidium cattleyanum* accessions from 12 populations obtained from 10 microsatellite loci. Colors in each

bar represent the probability a sampled individual belongs to a genetic cluster. Codes below bars correspond to population codes from Table 1.

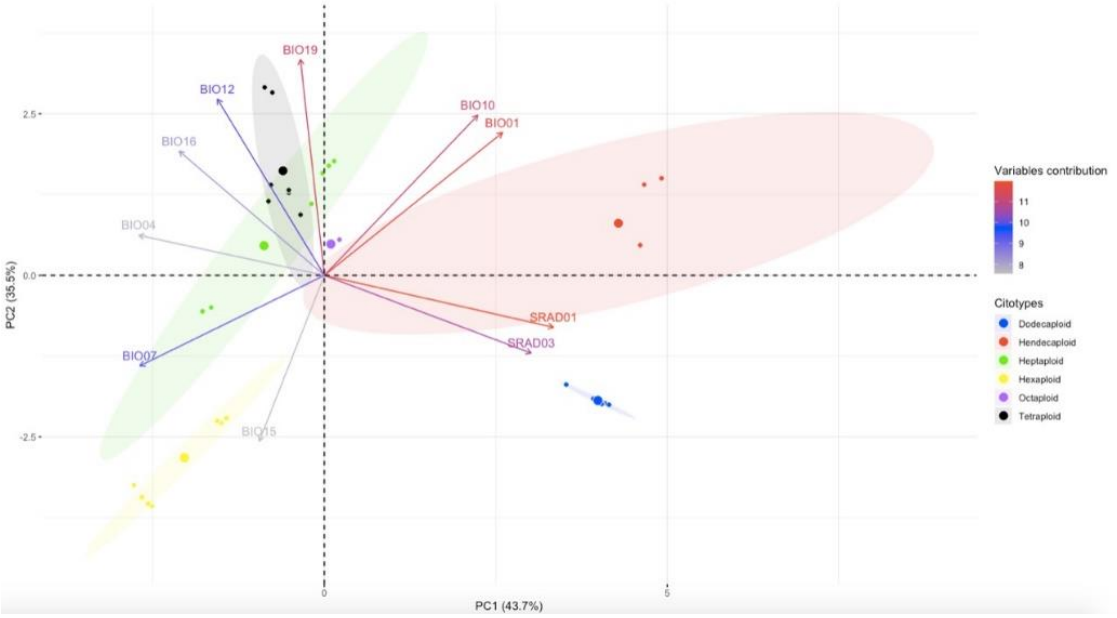


Fig 4. Principal component analysis of environmental variables for *Psidium cattleyanum* cytotypes. Variables codes are: SRAD01 - Solar radiation in January (kJ m-2 day-1); SRAD03 - Solar radiation in March (kJ m-2 day-1); BIO01 - Annual Mean Temperature; BIO4 - Temperature Seasonality (standard deviation ×100); BIO07 - Temperature Annual Range; BIO10 - Mean Temperature of Warmest Quarter; BIO 12 - Annual Precipitation; BIO 15 - Precipitation Seasonality (Coefficient of Variation); BIO 16 - Precipitation of Wettest Quarter; BIO 19 - Precipitation of Coldest Quarter.

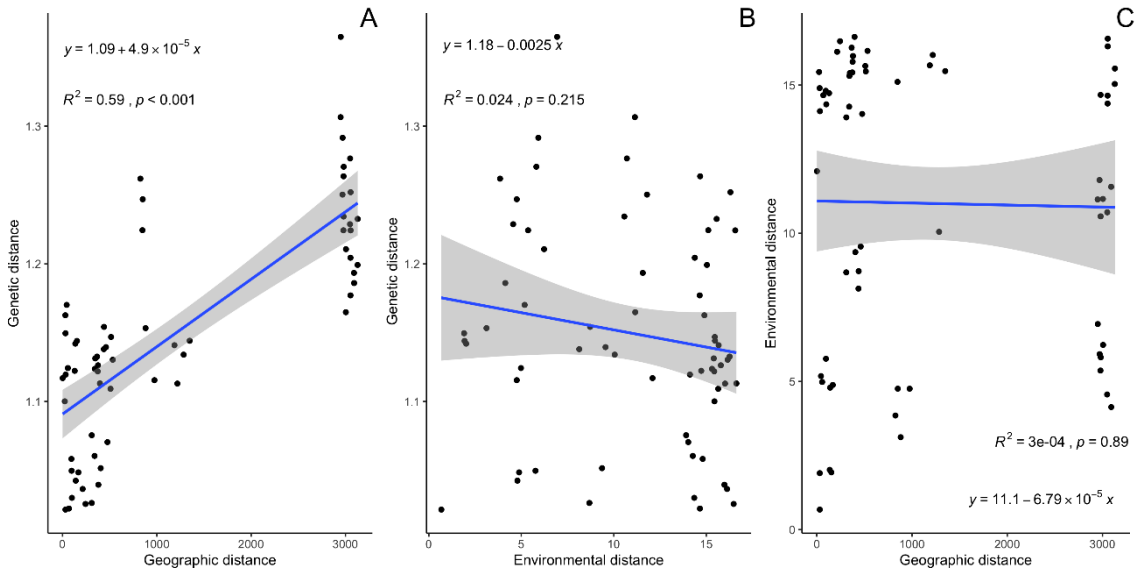
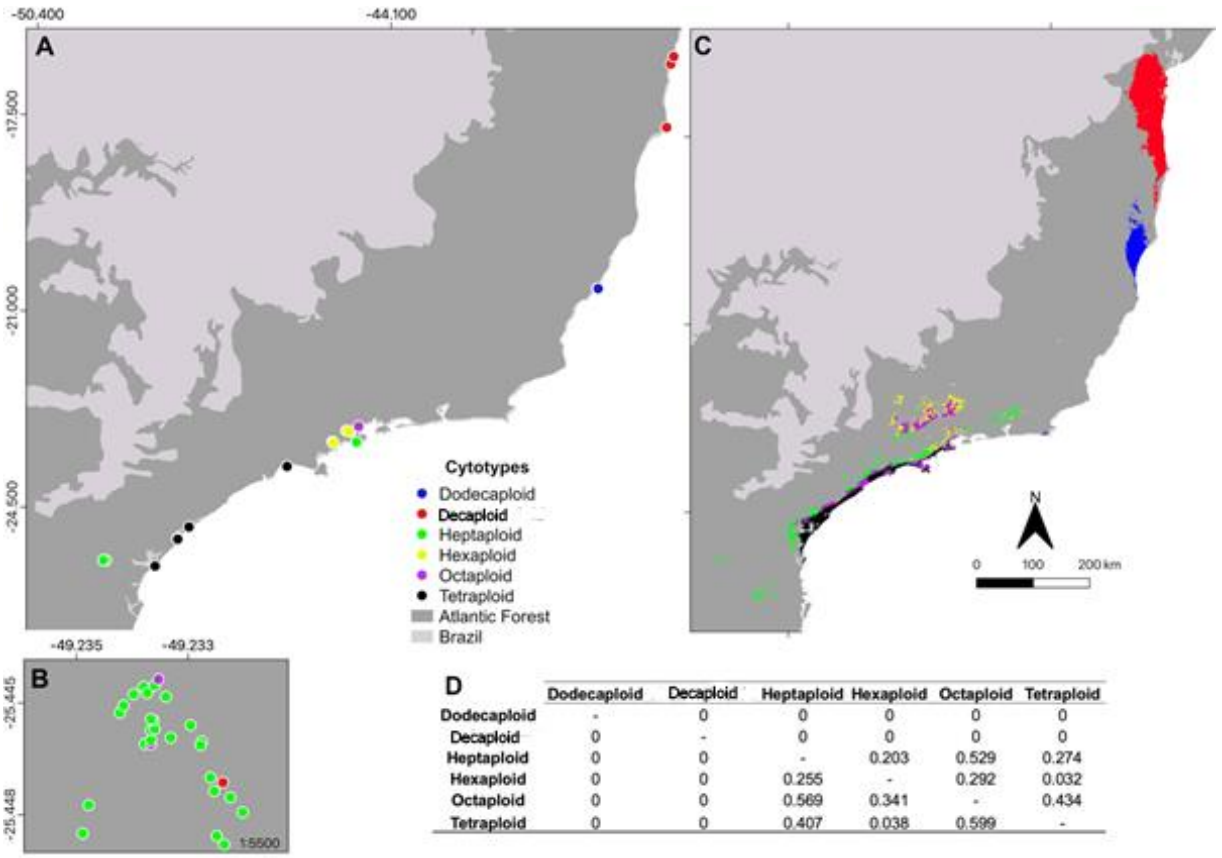


Fig 5. Linear regression model (blue line; standard deviation- gray shadow) of pairwise comparisons, A- genetic distance vs. geographic distance, B- genetic distance vs. environmental distance, and C- environmental distance vs. geographic distance.

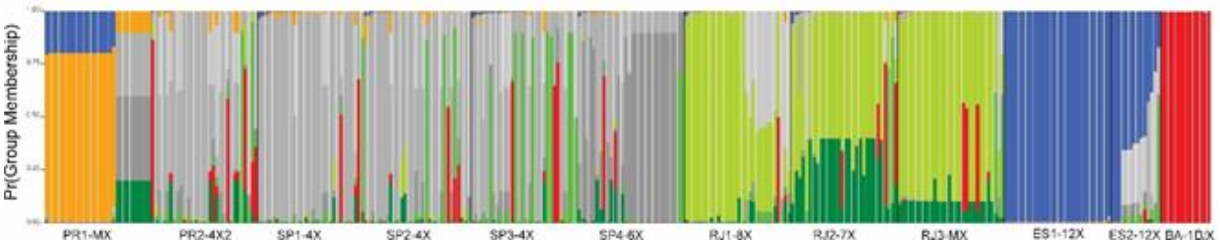


D

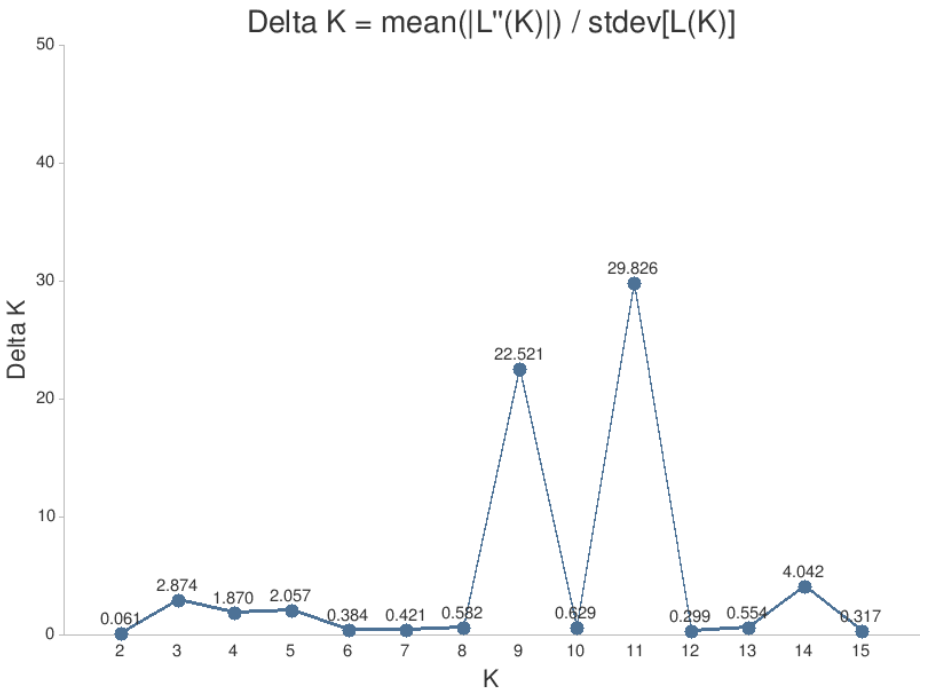
	Dodecaploid	Decaploid	Heptaploid	Hexaploid	Octaploid	Tetraploid
Dodecaploid	-	0	0	0	0	0
Decaploid	0	-	0	0	0	0
Heptaploid	0	0	-	0.203	0.529	0.274
Hexaploid	0	0	0.255	-	0.292	0.032
Octaploid	0	0	0.569	0.341	-	0.434
Tetraploid	0	0	0.407	0.038	0.599	-

Fig 6. Niche overlap results. A. Distribution of the sampled populations. B. Detail of the populations sampled in PR1-MX. C. Binary ensembles (majority consensus) of cytotypes. D. Niche overlap table between cytotypes, higher values correspond to Schoener's D (Schoener 1968) and lower values to Warren's I.

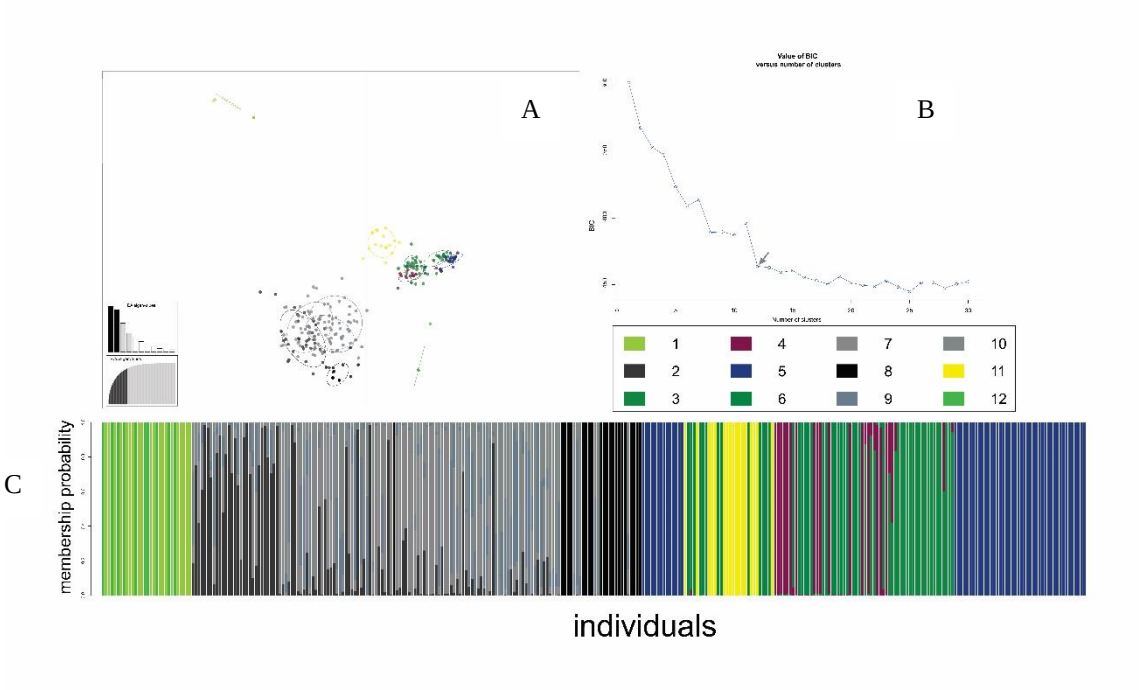
SUPPLEMENTARY DATA



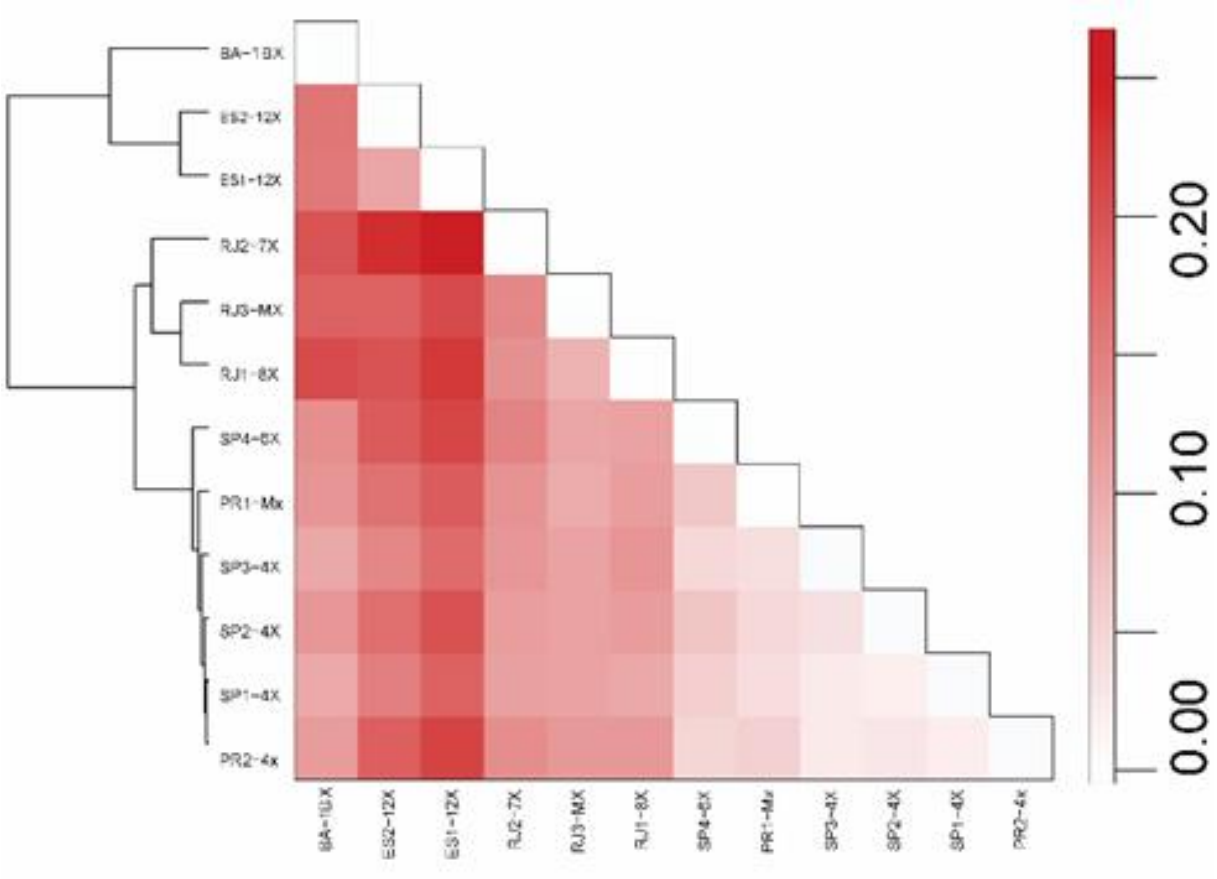
Supplementary Fig S1. Bar plots of the estimated membership coefficient for each of the 328 individuals of *Psidium cattleyanum*. The optimal K inferred by STRUCTURE was 9. A vertical bar represents each genotype. Codes below bars correspond to population codes according to Table 1.



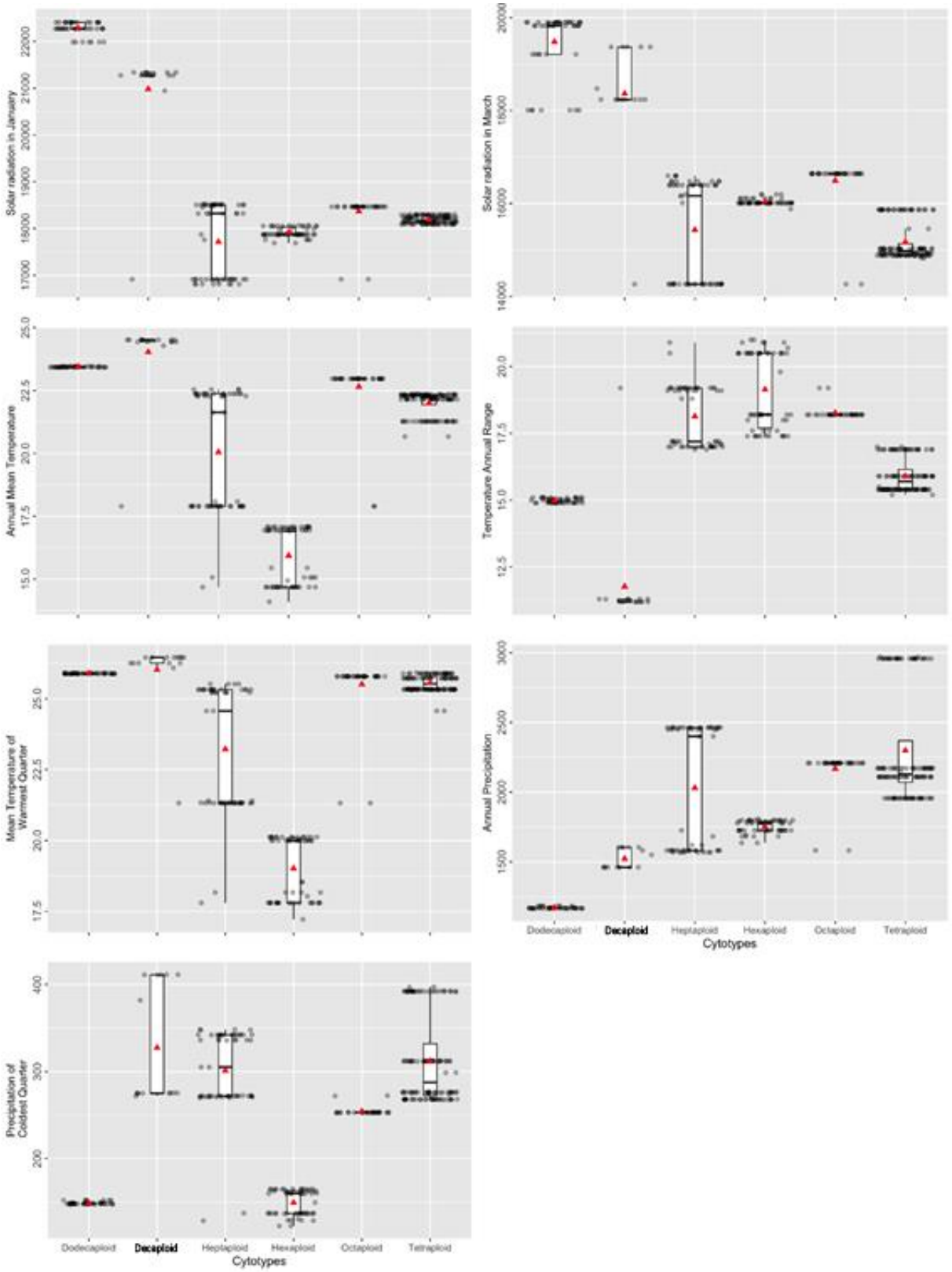
Supplementary Fig S2. The approximate numbers of genetic clusters (K) within the full data set of 328 individuals of *Psidium cattleyanum* based on results from the software package STRUCTURE.



Supplementary Fig S3. DAPC scatter plots based on the K-means algorithm for 328 individuals of *Psidium cattleyanum*; (A) Scatterplot shows genetic patterns of SSR data; The screen plots of eigenvalues of discriminant analysis and the amount of variation contained in the different principal components are presented (insert). (B) K number is selected based on BIC value for clusters up to k = 15 (arrow; (C) bar plot showing the probabilities of assignment of individuals to k = 12 genetic DAPC clusters.



Supplementary Fig S4. Heatmap of calculated pairwise estimates of Gst for populations of *Psidium cattleyanum*. Annotations on the left side of the heatmap show clustering of the samples.



Supplementary Fig S5. Boxplot of individual analysis of each environmental variable determinant for cytotypes occurrence.

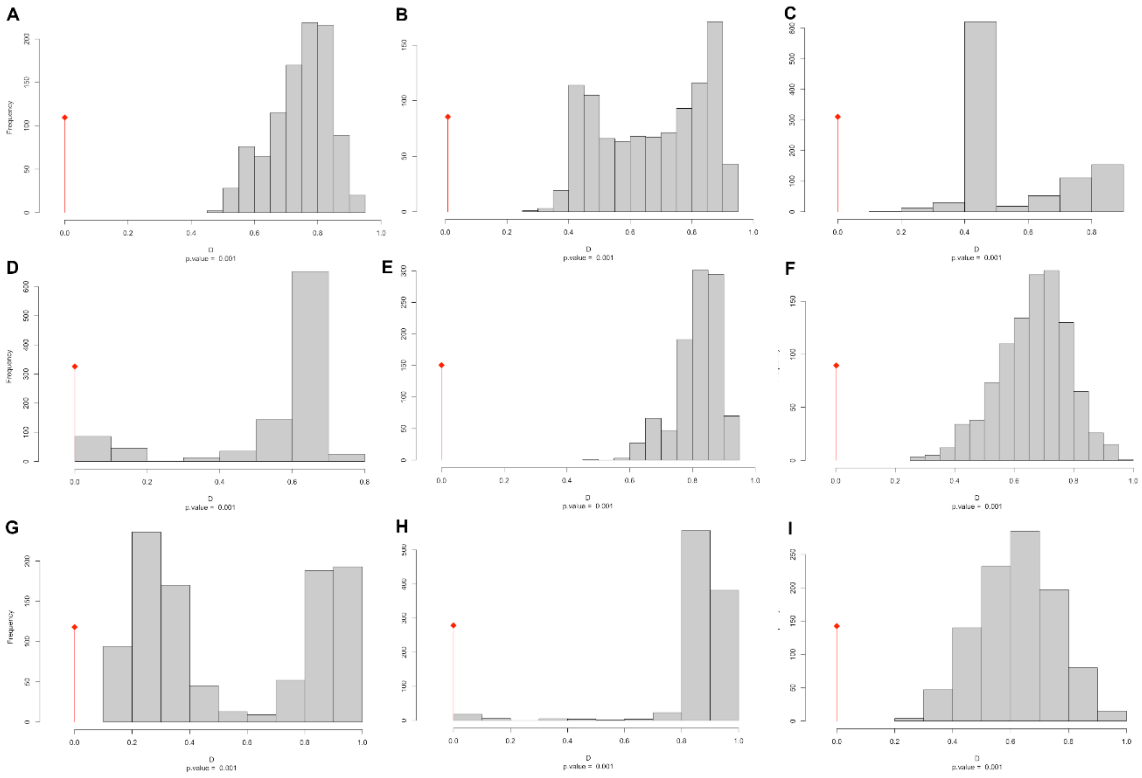
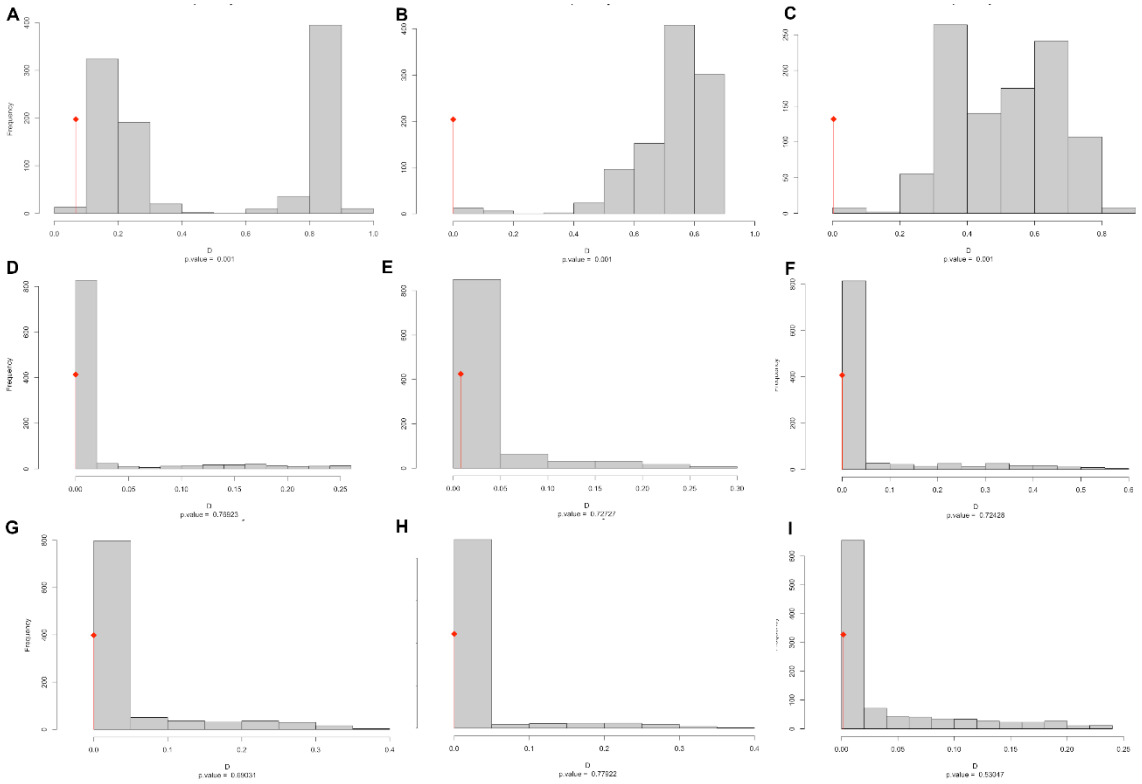
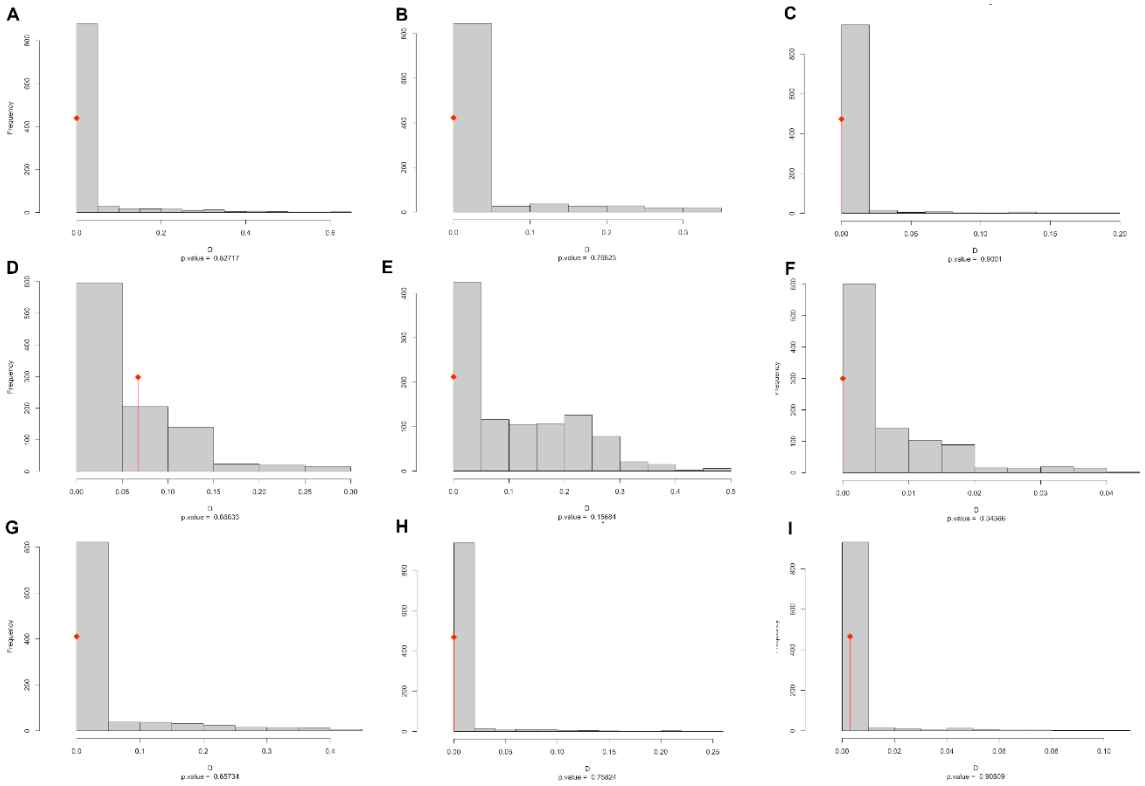


Fig 6. Scatterplot of the results of the equivalence tests.



Supplementary Fig S7. Scatterplot of the results of the equivalence tests A to C, and similarity tests D to I.



Supplementary Fig S8. Scatterplot of the results of similarity tests.

Considerações finais

A poliploidia é um tema que gera amplas discussões, assim o conhecimento sobre aspectos da citogenética, genética, ecologia de organismos poliploides faz-se extremamente necessário. Diversos esforços têm sido feitos para gerar dados experimentais sobre formação e estabelecimento de organismos e populações poliploides. Apresento, nesta tese, três manuscritos que contribuem para o avanço dos estudos sobre poliploidização.

No primeiro manuscrito, nós confirmamos a ocorrência de uma série poliploide em *P. cattleyanum*, derivada de x=11, também demostramos a ocorrência de rearranjos genômicos em dois níveis de ploidia mais altos, assim como mostramos que a estimativa do conteúdo de DNA pode ser útil para identificar os citótipos de *P. cattleyanum*.

No segundo manuscrito mostramos que marcadores microssatélites são úteis para estimar a diversidade genética e estrutura genética das populações naturais *P. cattleyanum*, também detectamos a ocorrência de reprodução assexuada, o que é importante para programas de melhoramento genético.

No terceiro manuscrito mostramos que os citótipos de *P. cattleyanum* podem ser considerados como um passo intermediário na divergência entre linhagens, adicionando novas informações sobre o impacto da poliploidia no nicho ambiental, diversidade genética e divergência de populações.

Com uma grande quantidade de citótipos relatados na literatura e uma ampla distribuição geográfica *P. cattleyanum* é uma espécie chave para testar diversas teorias sobre a influência da poliploidia, estudos futuros de biologia reprodutiva de origem da poliploidia e biogeografia do complexo poliploide, podem ajudar a esclarecer os padrões encontrados nessa tese.

Referências

Atchison E (1947) Chromosome numbers in the Myrtaceae. Am J Bot 34: 159-164

Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L (2018) The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. Front Ecol Evol. doi:10.3389/fevo.2018.00117

Baniaga AE, Marx HE, Arrigo N, Barker MS (2020) Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. Ecol Lett. doi:10.1111/ele.13402

Briggs D, Walters SM (1997) Plant variation and evolution. Cambridge University Press, Cambridge

Costa IR (2009) Estudos evolutivos em Myrtaceae: aspectos citotaxonômicos e filogenéticos em Myrteae, enfatizando *Psidium* e gêneros relacionados. Thesis, Universidade Estadual de Campinas, Campinas

Costa IR, Dornelas MC, Forni-Martins ER (2008) Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. Plant Syst Evol. doi:10.1007/s00606-008-0088-x

Costa IR, Forni-Martins ER (2006) Chromosome studies in Brazilian species of *Campomanesia* Ruiz & Pávon and *Psidium* L. (Myrtaceae Juss.). Caryologia. doi:10.1080/00087114.2006.10797891

Decanter L, Colling G, Elvinger N, Heiðmarsson S, Matthies D. 2020. Ecological niche differences between two polyploid cytotypes of *Saxifraga rosacea*. Am J Bot. doi:10.1002/ajb2.1431

Éder-Silva E, Felix LP, Bruno RLA (2007) Citogenética de algumas espécies frutíferas nativas do nordeste do Brasil. Rev Bras Frutic. doi:10.1590/S0100-29452007000100024

Fox DT, Soltis DE, Soltis PS, Ashman TL, Van de Peer Y (2020) Polyploidy: a biological force from cells to ecosystems. Trends Cell Biol. doi:10.1016/j.tcb.2020.06.006

Franzon RC, Campos LDO, Proença CEB, Sousa-Silva JC (2009) Araçás do gênero *Psidium*: principais espécies, ocorrência, descrição e usos. Embrapa, Planaltina

Gao J (2019) Dominant plant speciation types. A commentary on: 'Plant speciation in the age of climate change'. Ann Bot. doi:10.1093/aob/mcz108

Glennon KL, Ritchie ME, Segraves KA (2014) Evidence for shared broad-scale climatic niches of diploid and polyploid plants. Ecol Lett. doi:10.1111/ele.12259

Global Invasive Species Database (2021). Available at: http://www.iucngisd.org/gisd/search.php on 21-01-2021

Grant V (1981) Plant Speciation. Columbia University Press, New York

Haldane JBS (1930) Theoretical genetics of autopolyploids. J Genet 22: 359-372.

Hirano RT, Nakazone HY (1969) Chromosome numbers of ten species and clones in the genus *Psidium*. J Am Soc Hortic Sci 94: 83-86.

Jighly A, Abbott RJ, Daetwyler HD (2019) Editorial: Polyploid Population Genetics and Evolution-From Theory to Practice. Front Ecol Evol. doi:10.3389/fevo.2019.00460

Kolár F, Čertner M, Suda J, Schönswetter P, Husband BC (2017) Mixed-ploidy species: progress and opportunities in polyploid research. Trends Plant Sci. doi:10.1016/j.tplants.2017.09.011

Levin DA (1975) Minority cytotype exclusion in local plant populations. Taxon 24: 35-43.

Levin DA (2019) Plant speciation in the age of climate change. Ann Bot. doi:10.1093/aob/mcz108 124: 769-775

López-Jurado J, Mateos-Naranjo E, Balao F (2019) Niche divergence and limits to expansion in the high polyploid *Dianthus broteri* complex. New Phytol. doi:10.1111/nph.15663

Machado RM, de Oliveira FA, de Matos Alves F, de Souza AP, Forni-Martins ER (2020) Population Genetics of Polyploid Complex Psidium cattleyanum Sabine (Myrtaceae): Preliminary Analyses Based on New Species-Specific Microsatellite Loci and Extension to Other Species of the Genus. Biochem Genet. doi:10.1007/s10528-020-10002-1

Marchant DB, Soltis DE, Soltis PS (2016) Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. New Phytol. doi:10.1111/nph.14069

Marques A, Tuler AC, Carvalho CR, Carrijo T, da Silva Ferreira MF, Clarindo W (2016) Refinement of the karyological aspects of *Psidium guineense* Swartz, 1788: a comparison with Psidium guajava Linnaeus, 1753. Comp Cytogenet. doi:10.3897/CompCytogen.v10i1.6462

Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011) Recently formed polyploid plants diversify at lower rates. Science. doi: 10.1126/science.1207205

Moura RF, Queiroga D, Vilela E, Moraes AP (2021) Polyploidy and high environmental tolerance increase the invasive success of plants. J Plant Res. doi:10.1007/s10265-020-01236-6.

Normand F, Habib R (2001) Phenology of strawberry guava *Psidium cattleianum* in ReÂunion Island. J Hortic Sci Biotech. doi:10.1080/14620316.2001.11511407

Otto SP (2007) The evolutionary consequences of polyploidy. Cell. doi:10.1016/j.cell.2007.10.022

Otto SP, Whitton J (2000) Polyploidy: incidence and evolution. Annu Rev Genet 34: 401–437

Pandit MK, Pocock MJ, Kunin WE (2011) Ploidy influences rarity and invasiveness in plants. J Ecol. doi:10.1111/j.1365-2745.2011.01838.x

Parisod C, Broennimann O (2016) Towards unified hypotheses of the impact of polyploidy on ecological niches. New Phytol. doi:10.1111/nph.14133

Parisod C, Holderegger R, Brochmann C (2010) Evolutionary consequences of autopolyploidy. New Phytol. doi:10.1111/j.1469-8137.2009.03142.x

Pedrosa-Macedo JH, Smith CW. 2007. Histórico dos Estudos do Projeto Araçá. In: Pedrosa-Macedo JH, DalMolin A, Smith CW (eds) O Araçazeiro: Ecologia e Controle Biológico, 1st edn. FUPEF, Curitiba, pp 01-06

Psidium in Flora do Brasil 2020 em construção. Jardim Botânico do Rio de Janeiro. Avaible at http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB10858

Ramsey J, Schemske DW (1998) Pathways, mechanisms and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29: 467-501

Raseira MCB, Raseira A (1996) Contribuição ao estudo do araçazeiro. Embrapa-Centro de Pesquisa Agropecuária de Clima Temperado, Pelotas

Rezende L, Suzigan J, Amorim FW, Moraes AP (2020) Can plant hybridization and polyploidy lead to pollinator shift? Acta bot bras. https://doi.org/10.1590/0102-33062020abb0025

Rice A, Šmarda P, Novosolov M, et al (2019). The global biogeography of polyploid plants. Nat Ecol Evol. https://doi.org/10.1038/s41559-018-0787-9

Schifino-Wittmann MT (2004) Poliploidia e seu impacto na origem e evolução das plantas silvestres e cultivadas. Current Agricultural Science and Technology. doi:10.18539/cast.v10i2.904

Singhal VK, Gill BS, Bir SS (1985) Cytology of cultivated woody species (Polypetalae). Proc. Indian Sci. Congr. Assoc 71: 143-144

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. PNAS. doi:10.1073/pnas.97.13.7051

Soltis DE, Segovia-Salcedo MC, Jordon-Thaden I, et al. (2014). Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et l. (2011). New Phytol. doi:10.1111/nph.12756

Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS (2007) Autopolyploidy in angiosperms: have we grossly underestimated the number of species? Taxon. doi:10.2307/25065732

Souza AG, Resende LV, Lima IP, Martins LSS, Techio VH (2014) Chromosome number and nuclear DNA amount in *Psidium* app. resistant and susceptible to *Meloidogyne enterolobii* and its relation with compatibility between rootstocks and commercial varieties of guava tree. Plant Syst Evol. doi:10.1007/s00606-014-1068-y

Souza-Pérez M, Speroni G (2017) New apomictic pathway in Myrtaceae inferred from *Psidium cattleyanum* female gametophyte ontogeny. Flora. https://doi.org/10.1016/j.flora.2017.06.010

Stebbins GL (1971) Chromosomal Evolution in Higher Plants. Edward Arnold, London.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2011) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot. doi:10.1093/aob/mcr277

Tuler AC, Carrijo TT, Peixoto AL, et al. (2019) Diversification and geographical distribution of Psidium (Myrtaceae) species with distinct ploidy levels. Trees. doi:10.1007/s00468-019-01845-2

Tuler AC, Proença CEB, Costa IR (2020) Psidium in Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Avaible at: <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB10855>

Van de Peer Y, Ashman TL, Soltis PS, Soltis DE (2020) Polyploidy: an evolutionary and ecological force in stressful times. Plant Cell. doi:10.1093/plcell/koaa015

Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. Nat Rev Genet. doi:10.1038/nrg.2017.26

Van Drunen WE, Husband BC (2019) Evolutionary associations between polyploidy, clonal reproduction, and perenniality in the angiosperms. New Phytol. doi:10.1111/nph.15999

Vasconcelos TN, Proença CE, Ahmad B, et al. (2017) Myrteae phylogeny, calibration, biogeography and diversification patterns: increased understanding in the most species rich tribe of Myrtaceae. Mol Phylogenetics Evol. doi:10.1016/j.ympev.2017.01.002

Vázques SNM (2014) *Psidium cattleyanum* Sabine y *Acca sellowiana* (Berg.) Burret (Myrtaceae): caracterización cromossómica y cariotípica em poblaciones silvestre y genótipos selecionados em programas nacionales de mejoramiento. Monography, Universidad de la República, Uruguay.

Wendel JF (2015) The wondrous cycles of polyploidy in plants. doi:10.3732/ajb.1500320

Wikler C (2007) Problemática do Araçazeiro como Planta Invasora. In: Pedrosa-Macedo JH, Dalmolin A, Smith CW (eds) O Araçazeiro: Ecologia e Controle Biológico, 1st edn. FUPEF, Curitiba, pp 07-18

Wittmann MTS, Dallagnol M (2003) Indução de poliploidia no melhoramento de plantas. Pesq Agrop Gaúcha 9: 155-164

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. PNAS. doi:10.1073/pnas.0811575106

WCSP (2017). “World Checklist of Selected Plant Families”. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet at http://wcsp.science.kew.org/

ANEXOS

1. Relatório do SISGEN

Ministério do Meio Ambiente
CONSELHO DE GESTÃO DO PATRIMÔNIO GENÉTICO

SISTEMA NACIONAL DE GESTÃO DO PATRIMÔNIO GENÉTICO E DO CONHECIMENTO TRADICIONAL ASSOCIADO

Comprovante de Cadastro de Acesso

Cadastro nº A60388A

A atividade de acesso ao Patrimônio Genético, nos termos abaixo resumida, foi cadastrada no SisGen, em atendimento ao previsto na Lei nº 13.123/2015 e seus regulamentos.

Número do cadastro:	**A60388A**
Usuário:	**Raquel Moura Machado**
CPF/CNPJ:	**043.551.153-05**
Objeto do Acesso:	**Patrimônio Genético**
Finalidade do Acesso:	**Pesquisa**

Espécie

Psidium cattleyanum

Título da Atividade:	**Poliploidia em Psidium cattleyanum Sabine (Myrtaceae): implicações citogenéticas e evolutivas**

Equipe

Raquel Moura Machado	**UNICAMP**

Data do Cadastro:	**04/03/2021 09:52:15**
Situação do Cadastro:	**Concluído**

Conselho de Gestão do Patrimônio Genético
Situação cadastral conforme consulta ao SisGen em **9:52** de **04/03/2021**.

SISTEMA NACIONAL DE GESTÃO DO PATRIMÔNIO GENÉTICO E DO CONHECIMENTO TRADICIONAL ASSOCIADO - SISGEN

2. Declaração de direitos autorais

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas cientificas ou anais de congressos sujeitos a arbitragem, que constam da minha Tese de Doutorado, intitulada "POLIPLOIDIA EM *Psidium cattleyanum* Sabine (Myrtaceae): IMPLICAÇÕES CITOGENÉTICAS E EVOLUTIVAS", não infringem os dispositivos da Lei n." 9.610/98, nem o direito autoral de qualquer editora.

O capítulo 2 desta Tese de Doutorado foi publicado na Biochemical Genetics, aqui encontra-se a versão submetida para a revista, a publicação final está disponível em: https://link.springer.com/article/10.1007%2Fs10528-020-10002-1

Published in Biochemical Genetics. https://doi.org/10.1007/s10528-020-10002-1

The final publication is available at https://link.springer.com/article/10.1007%2Fs10528-020-10002-1

Campinas, 16 de junho de 2021

Assinatura: Raquel M. Machado

Nome do(a) autor(a): Raquel Moura Machado

RG n. 2006010330518

Assinatura: Eliana R Forni Martins

Nome do(a) orientador(a): Eliana Regina Forni Martins

RG n. 8580582-8