**JAQUELINE AMORIM PEREIRA**

# EVOLUÇÃO DA *C*-HETEROCROMATINA BASEADA NO ESTUDO DE SEQUÊNCIAS DE DNA REPETITIVO EM ABELHAS DOS GÊNEROS *MELIPONA* E *TRIGONA* (APIDAE: MELIPONINI)

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Biologia Celular e Estrutural, para obtenção do título de *Doctor Scientiae*.

Orientadora: Denilce Meneses Lopes

Coorientadora: Tânia Maria Fernandes Salomão

**VIÇOSA - MINAS GERAIS**
**2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa**

T
P436e
2022

Pereira, Jaqueline Amorim, 1993-
Evolução da *c*-heterocromatina baseada no estudo de sequências de DNA repetitivo em abelhas dos gêneros *Melipona* e *Trigona* (Apidae: Meliponini) / Jaqueline Amorim Pereira. – Viçosa, MG, 2022.
1 dissertação eletrônica (84 f.): il. (algumas color.).

Orientador: Denilce Meneses Lopes.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Biologia Geral, 2022.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2022.698
Modo de acesso: World Wide Web.

1. Abelhas sem ferrão - Evolução. 2. Abelhas sem ferrão - Citogenética. I. Lopes, Denilce Meneses, 1978-. II. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Biologia Celular e Estrutural. III. Título.

CDD 22. ed. 595.799

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

**JAQUELINE AMORIM PEREIRA**

# EVOLUÇÃO DA *C*-HETEROCROMATINA BASEADA NO ESTUDO DE SEQUÊNCIAS DE DNA REPETITIVO EM ABELHAS DOS GÊNEROS *MELIPONA* E *TRIGONA* (APIDAE: MELIPONINI)

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Biologia Celular e Estrutural, para obtenção do título de *Doctor Scientiae*.

APROVADA: 27 de outubro de 2022.

Assentimento:

Jaqueline Amorim Pereira
Autora

Denilce Meneses Lopes
Orientadora

## AGRADECIMENTOS

A Deus por conceder força, perseverança e saúde.

Aos meus pais, Rogério e Helena, por todo incentivo, apoio e amor! À minha irmã, Gabriela, pelo incentivo e amizade.

Ao Guilherme, meu companheiro, pela paciência, incentivo, companheirismo, amizade e amor.

À minha orientadora, professora Denilce Menezes Lopes por todos os ensinamentos, confiança, incentivo e amizade durante todos esses anos.

À minha coorientadora, professora Tânia, pelos ensinamentos nesses anos.

Ao professor Diogo Cavalcante Cabral de Mello por ter me acolhido em seu laboratório, por todos ensinamentos e confiança.

Ao professor Lucio Campos pelos ensinamentos e amizade que vem desde a minha graduação.

Aos professores Maria José e Marcelo, por me acolherem em sua casa em Rio Claro durante meus experimentos, pela amizade e ensinamentos.

As meninas do Laboratório de Citogenética de Insetos, Gisele, Marina, Natália e Camila obrigada pela amizade, ensinamentos, toda ajuda e companheirismos nesses anos.

Ao Insect Cytogenomics Group pelo acolhimento, ensinamento e amizade. Ana Eliza por ter me acolhido em sua casa e ajudado nos experimentos. Ana Beatriz, Ana Eliza, Diogo Milani e Vanessa obrigada por ter contribuído com os experimentos e pela prontidão em ajudar.

À minha professora de inglês, Fernanda, pela disponibilidade em ajudar.

À todos os amigos e pessoas especiais que fizeram parte dessa jornada.

Ao Instituto de Biociências da UNESP – Rio Claro, onde desenvolvi parte desse trabalho.

Ao programa de Pós-graduação em Biologia Celular e Estrutural da Universidade Federal de Viçosa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

PEREIRA, Jaqueline Amorim, D.Sc., Universidade Federal de Viçosa, outubro de 2022. **Evolução da *c*-heterocromatina baseada no estudo de sequências de DNA repetitivo em abelhas do gênero *Melipona* e *Trigona* (Apidae: Meliponini).** Orientadora: Denilce Meneses Lopes. Coorientadora: Tânia Maria Fernandes Salomão.

A heterocromatina constitutiva (*c*-heterocromatina) está relacionada à estabilidade dos cromossomos, formação de centrômeros e telômeros. Sua composição é basicamente de sequência de DNA repetitivo, como DNA satélite (DNAsat) e elementos transponíveis (TE). Essas sequências estão sujeitas a rápidas modificações, as quais podem influenciar na evolução da *c*-heterocromatina e do genoma das espécies. Em Meliponini, abelhas sem ferrão, normalmente as espécies apresentam um dos braços cromossômicos com grandes blocos de *c*-heterocromatina, como ocorre no gênero *Trigona*. Entretanto, o gênero *Melipona* é uma exceção a esse padrão, com espécies em que a *c*-heterocromatina ocupa quase toda a extensão dos cromossomos (alto conteúdo de *c*-heterocromatina) e espécies com pouca quantidade de *c*-heterocromatina (baixo conteúdo de *c*-heterocromatina), sendo nesse caso restrita à região centromérica e pericentromérica dos cromossomos. Devido a essas características cromossômicas, as espécies dos gêneros *Melipona* e *Trigona* são um modelo pertinente para estudar a evolução da *c*-heterocromatina, a partir da abordagem de sua composição molecular. Desse modo, no capítulo I, caracterizamos as sequências de DNA repetitivo dos cromossomos de duas espécies de *Melipona*, *Melipona (Michmelia) scutellaris*, com alto conteúdo de *c*-heterocromatina e *Melipona (Melipona) quadrifasciata*, baixo conteúdo de *c*-heterocromatina, por meio do sequenciamento genômico, métodos de bioinformática e citogenética. Nossos dados revelaram que a composição da *c*-heterocromatina é diferente entre as duas espécies analisadas e à amplificação dessa região nos cromossomos de *M. scutellaris* é decorrente da intensa amplificação de famílias especificas de DNA repetitivo. No capítulo II estendemos as análises anteriores para representantes de diferentes subgêneros de *Melipona*: *Melipona (Michmelia) mondury, Melipona (Melikerria) interrupta, Melipona (Melikerria) fasciculata, Melipona (Eomelipona) bicolor,* com alto e baixo conteúdo de *c*-heterocromatina. A composição da *c*-heterocromatina mostrou diferente entre os quatro subgêneros e isso foi decorrente de mudanças quantitativas na biblioteca de DNAsat das espécies. Os subgêneros com baixa proporção de *c*-heterocromatina (*Eomelipona* e *Melipona*) apresentaram baixa abundância de DNAsat, enquanto o oposto é observado em espécies com

alto conteúdo (*Michmelia* e *Melikerria*). Notamos, que a amplificação da *c*-heterocromatina ocorreu de forma independente em *Michmelia* e *Melikerria*, e esse processo foi decorrente de duas famílias de DNAsat que forma amplificadas de forma diferente em cada subgênero. No capítulo III usamos diferentes técnicas, enzimas de restrição, sequenciamento de genoma, análises de bioinformática e hibridização fluorescente *in situ*, com intuito de identificar as sequências que estão contribuindo para evolução da *c*-heterocromatina em *Trigona.* Identificamos uma família de DNAsat altamente abundante do genoma de *Trigona hyalinata* como um dos principais constituintes da *c*-heterocromatina dessa espécie, bem como de espécies pertencentes ao clado B. Essa família não foi observada em nenhuma espécie analisada do clado A. Desse modo, a *c*-heterocromatina em *Trigona* está evoluindo de forma diferente entre clado A e B, e isso é decorrente da modificação da biblioteca de DNAsat ao longo do tempo.

**Palavras-Chave:** Genoma. DNA repetitivo. RepeatExplorer. Citogenética.

## ABSTRACT

PEREIRA, Jaqueline Amorim, M.Sc., Universidade Federal de Viçosa, October, 2022. **Evolution of c-heterochromatin based on the study of repetitive DNA sequences in bees of the genus *Melipona* and *Trigona* (Apidae: Meliponini).** Advisor: Denilce Meneses Lopes. Co-advisor: Tânia Maria Fernandes Salomão.

Constitutive heterochromatin (*c*-heterochromatin) is related to chromosome stability, centromeres and telomeres formation. Its composition is basically a repetitive DNA sequence, such as satellite DNA (DNAsat) and transposable elements (TE). These sequences are subject to rapid modifications, which can influence the evolution of *c*-heterochromatin and the genome of the species. In Meliponini, stingless bees, the species normally have one of the chromosome arms with large blocks of *c*-heterochromatin, as in the genus *Trigona*. However, the genus *Melipona* represents an exception to this pattern, with species in which c-heterochromatin occupies almost the entire length of the chromosomes (high content of *c*-heterochromatin) and species with little amount of *c*-heterochromatin (low content of *c*-heterochromatin). Due to these chromosomal characteristics, the species of the genera *Melipona* and *Trigona* are relevant model to study the evolution of c-heterochromatin based on its molecular composition. Thus, in chapter I, we characterized the repetitive DNA sequences of the chromosomes of two species of *Melipona*, *Melipona (Michmelia) scutellaris*, with high *c*-heterochromatin content and *Melipona (Melipona) quadrifasciata*, low *c*-heterochromatin content, through of genomic sequencing, bioinformatics and cytogenetics methods. Our data revealed that the composition of *c*-heterochromatin is different between the two analyzed species and the amplification of this region in the chromosomes of *M. scutellaris* is due to the intense amplification of specific families of repetitive DNA. In chapter II we extend the previous analyzes to representatives of different subgenera of *Melipona: Melipona (Michmelia) mondury*, *Melipona (Melikerria) interrupta*, *Melipona (Melikerria) fasciculata*, *Melipona (Eomelipona) bicolor*, with high and low *c*-heterochromatin content. The composition of *c*-heterochromatin showed differences between the four subgenera and this was due to quantitative changes in the satDNA library of the species. The subgenera with a low proportion of *c*-heterochromatin (*Eomelipona* and *Melipona*) showed a low abundance of satDNA, while the opposite is observed in species with a high content (*Michmelia* and *Melikerria*). We noticed that the amplification of *c*-heterochromatin occurred independently in *Michmelia* and *Melikerria*, and this process was due to two families of satDNA that were

amplified differently in each subgenus. In chapter III we used different techniques, restriction enzymes, genome sequencing, bioinformatics analysis and fluorescent *in situ* hybridization, in order to identify the sequences that are contributing to the evolution of *c*-heterochromatin in *Trigona*. We identified a family of highly abundant satDNAs from the genome of *Trigona hyalinata* as one of the main constituents of *c*-heterochromatin of this species, as well as of species belonging to the clade B. This family was not observed in any analyzed species from clade A. Thus, *c*-heterochromatin in *Trigona* is evolving differently between clade A and B, and this is due to the modification of the satDNA library over time.

**Keywords:** Genome. Repetitive DNA. RepeatExplorer. Cytogenetics.

**SUMÁRIO**

# 1. INTRODUÇÃO

## 1.1 Heterocromatina

A heterocromatina foi identificada pela primeira vez por Heitz, ao analisar o comportamento dos cromossomos durante a divisão celular (HEITZ, 1928). Em suas análises foi notada a presença de duas regiões distintas nos cromossomos: a eucromatina, com diferentes graus de condensação, e a heterocromatina, apresentando-se condensada ao longo do ciclo celular. A heterocromatina, por sua vez, pode ser classificada em heterocromatina facultativa e heterocromatina constitutiva (LIU; ALI; ZHOU, 2020; PIMPINELLI; GATTI; DE MARCO, 1975). A heterocromatina facultativa constitui uma região dos cromossomos que contém genes, de modo que a sua condensação varia ao longo do desenvolvimento do organismo. Por outro lado, a heterocromatina constitutiva (*c*-heterocromatina) é caracterizada por ser pobre em genes e apresentar maior grau de condensação que a eucromatina. A *c*-heterocromatina apresenta importante papel na manutenção da estabilidade dos cromossomos, formação de centrômeros, pareamento e segregação de cromossomos durante a divisão celular (ALLSHIRE; MADHANI, 2018; LIU; ALI; ZHOU, 2020).

O estudo da *c*-heterocromatina tem demonstrado que essa região é constituída preferencialmente por sequências de DNA repetitivo, como por exemplo, as repetições *in tandem*, que incluem o DNA satélite; e as sequências dispersas, como elementos transponíveis (BOSTOCK, 1980; CHARLESWORTH; SNIEGOWSKI; STEPHAN, 1994; RICHARD; KERREST; DUJON, 2008). Devido à sua constituição, a *c*-heterocromatina pode exibir rápida modificação ao longo do tempo, o que pode refletir na evolução do genoma e das espécies (FERREE; PRASAD, 2012; GARRIDO-RAMOS, 2017; HUGHES; HAWLEY, 2009).

## 1.2 Elementos Transponíveis

Os elementos transponíveis (TEs), foram identificados por Barbara McClintock, quando conduzia experimentos com milho e observou mudanças no padrão de coloração dos grãos após os cruzamentos (MCCLINTOCK, 1950, 1951, 1956). Diante desses resultados ela concluiu que existiam “elementos controladores” que se moviam pelo genoma e podiam influenciar a atividade dos genes e consequentemente nas características fenotípicas (MCCLINTOCK, 1950, 1951, 1956). Dentre as características dos TEs estão a localização dispersa e a capacidade de se mover pelo genoma, replicar e multiplicar (WICKER et al., 2007). Como consequência disso, essas sequências contribuem para modulação da arquitetura genômica e evolução cromossômica, já que podem facilitar a ocorrência de mutações e rearranjos cromossômicos

(GRABUNDZIJA et al., 2016; GRAY, 2000; KLEIN; O'NEILL, 2018). Os TEs podem estar envolvidos na regulação gênica, alterando os padrões de transcrição quando inseridos nas regiões ocupadas por genes (CARBONE et al., 2014; STUART et al., 2016). Além disso, essas sequências podem exibir proporção variada no genoma de diferentes espécies e isso pode estar relacionado com o tamanho genoma (GREGORY, 2005; ZUCCOLO et al., 2007).

Usualmente TEs são divididos em duas classes de acordo com os mecanismos de transposição (WICKER et al., 2007). Os elementos da classe I, são chamados de retrotransposons e realizam transposição através de um RNA intermediário e da ação de uma transcriptase reversa. Os retrotransposons podem ser divididos em dois grupos, devido à presença ou não de repetições terminais longas (LTR) nas extremidades. Os elementos da classe II, são conhecidos como transposons, e usam DNA para transposição. Elementos da classe II podem codificar o gene da transposase, que normalmente é flanqueado por repetições terminais invertidas (TIR) (WICKER et al., 2007).

Os TEs estão sujeitos a modificações epigenéticas a fim de minimizar sua atividade de transposição, como exemplo o mecanismo de metilação na lisina 9 da histona H3, o que resulta em uma região condensada e inativa (LIPPMAN; MARTIENSSEN, 2004; SLOTKIN; MARTIENSSEN, 2007). Desse modo, essas sequências contribuem para formação da *c*-heterocromatina, região na qual essas sequências predominam (LIPPMAN; MARTIENSSEN, 2004; MARSANO; DIMITRI, 2022; SLOTKIN; MARTIENSSEN, 2007). Alguns estudos vêm demonstrando uma relação entre TEs e DNA satélite, com a tendência de TEs invadirem regiões de DNA satélite (MEŠTROVIĆ et al., 2015). Nos centrômeros, por exemplo, foi relatado que os TEs podem participar da sua formação e funcionalidade quando se misturam com o DNA satélite presente nessa região (LI et al., 2013). Os TEs também podem ser modelo para formação de DNA satélite, apresentando homologias entre eles e contribuindo para a evolução dessas sequências (MEŠTROVIĆ et al., 2015).

### 1.3 DNA satélite

Os DNAs satélites (DNAsat) caracterizam como arranjos de sequências repetidas *in tandem* e são classificados de acordo com o tamanho da unidade de repetição em microssatélite (1 – 6pb), minissatélite (7 – 100pb) e satélite (> 100pb) (CHARLESWORTH; SNIEGOWSKI; STEPHAN, 1994; RICHARD; KERREST; DUJON, 2008). Essas sequências foram primeiramente identificadas por meio de experimentos de ultracentrifugação em gradiente de cloreto de césio, na qual identificou-se duas bandas com diferentes densidades, sendo a banda menos densa, cerca de 10% do genoma, caracterizada como DNAsat (KIT, 1961; SUEOKA,

1961). Por meio do uso da técnica de cinética de reassociação ($C_0t$-DNA), observou-se que essa região apresentava taxa de re-naturação maior que o restante do DNA genômico devido à sua natureza repetitiva (BRITTEN; DAVIDSON, 1971; WARING; BRITTEN, 1966). O uso dessa técnica para estudo de sequências repetitivas predominou por anos, no entanto por volta de 1980, outra metodologia foi introduzida, o uso de enzimas de restrição, permitindo isolar uma sequência repetitiva específica (HORZ; ZACHAU, 1977; SINGER, 1982). Posteriormente essas metodologias foram atreladas à Hibridização Fluorescente *in situ* (FISH) e à clonagem, possibilitando identificar sua localização nos cromossomos e a constituição das sequências repetitivas (RATAN et al., 2017).

Ao longo de anos e com avanços dos estudos, o papel do DNAsat no genoma começou a ser desvendado (GARRIDO-RAMOS, 2017). Essas sequências podem ser encontradas em quantidade elevada no genoma de algumas espécies, e em geral têm sido observadas na *c*-heterocromatina e em alguns casos também na eucromatina, ocupando regiões centroméricas, pericentroméricas, subteloméricas e algumas regiões intersticiais (GREWAL; JIA, 2007; RUIZ-RUANO et al., 2016). Mesmo prevalecendo em regiões com baixa taxa transcricional, algumas sequências de DNA satélite podem ser processadas em pequenos RNAs, como piwiRNAs e siRNAs que atuam na regulação epigenética do genoma, na formação e manutenção da *c*-heterocromatina (BISCOTTI et al., 2015). Além disso, também podem estar envolvidas no alongamento de telômeros, formação de centrômeros, montagem do cinetócoro, pareamento de cromossomos e segregação cromossômica (PLOHL; MEŠTROVIĆ; MRAVINAC, 2012, 2014).

Foi postulado que a introdução de uma família de DNA satélite no genoma, começa com a duplicação *de novo* de qualquer sequência de dois ou mais pares de bases, seja por "derrapagem" da polimerase durante a replicação ou por reinserção de cópias replicadas por círculo rolante de intermediários de DNA circulares extracromossômicos (COHEN; SEGAL, 2009; RUIZ-RUANO et al., 2016). Posteriormente, esse novo segmento pode ser disseminado pelo genoma (RUIZ-RUANO et al., 2016). Após o estabelecimento dessas sequências no genoma, as repetições que compõem uma família de DNAsat podem mudar rapidamente ao longo do tempo devido à amplificação ou deleção de suas repetições e ao acúmulo de mutações (GARRIDO-RAMOS, 2017; PLOHL; MEŠTROVIĆ; MRAVINAC, 2012). Isso pode ocasionar modificação no genoma e consequentemente favorecer o isolamento reprodutivo, contribuindo para a especiação (FERREE; PRASAD, 2012; HUGHES; HAWLEY, 2009; JAGANNATHAN; YAMASHITA, 2021).

Com a ampliação do uso do sequenciamento de nova geração (NGS), novas informações sobre genomas têm sido fornecidas. No entanto, sempre foi um desafio a caracterização de sequências repetitivas de genomas sequenciados, muitas vezes devido às ferramentas de bioinformática disponíveis exigir a montagem completa do genoma (ALTEMOSE et al., 2014; GARRIDO-RAMOS, 2017). Atualmente o desenvolvimento de novas ferramentas de bioinformática tem auxiliado na identificação dessas sequências, como o pipeline RepeatExplorer (NOVÁK; NEUMANN; MACAS, 2020). Esse método se baseia na identificação e clusterização de *reads* semelhantes, provenientes de sequenciamento de nova geração que foram cortadas aleatoriamente (NOVÁK; NEUMANN; MACAS, 2020). Com base nas semelhanças entre as *reads,* um gráfico é construído e sua topologia analisada a fim de identificar sequências de DNA repetitivo (NOVÁK; NEUMANN; MACAS, 2020). Com isso o NGS juntamente com ferramentas de bioinformática tem permitido a análise detalhada e abrangente do conteúdo repetitivo do genoma, sendo possível um estudo minucioso do DNAsat e TEs, enquanto outras técnicas (uso de enzimas de restrição) permitiam estudo de apenas poucas sequências. O uso dessas técnicas está sendo empregado em vários organismos, incluindo plantas (PAMPONÉT et al., 2019), insetos (CAMACHO et al., 2022; LIMA; RUIZ-RUANO, 2022; RUIZ-RUANO et al., 2016), peixes (SERRANO-FREITAS et al., 2020), e juntamente com a citogenética tem revelando informações sobre a evolução dessas sequências no genoma, que de modo geral, contribui para o entendimento da *c*-heterocromatina e dos cromossomos.

### 1.4 *Melipona*

*Melipona* Illiger, 1806, compreende um gênero da tribo Meliponini (abelhas sem ferrão), com distribuição exclusivamente neotropical, rico em espécies, sendo 73 espécies descritas e 43 com ocorrência no Brasil (CAMARGO; PEDRO, 2013; MICHENER, 2007). Espécies desse gênero, assim como outros Meliponídeos, constituem importantes polinizadores, já que coletam recursos alimentares como pólen e néctar das flores, contribuindo para a reprodução das plantas (GIANNINI et al., 2015).

Há décadas, esse grupo tem sido alvo de estudos morfológicos, filogenéticos e citogenéticos, contribuindo para ampliar o conhecimento dessas espécies (CAMARGO; PEDRO, 2013; DA CUNHA et al., 2018; RAMÍREZ et al., 2010). Com base em estudos morfológicos, as espécies desse gênero são distribuídas em quatro subgêneros: *Eomelipona* Moure, 1992, *Melikerria* Moure, 1992, *Michmelia* Moure, 1975 e *Melipona* Illiger, 1806 (CAMARGO; PEDRO, 2013). Análises moleculares de genes mitocondriais e nucleares

corroboram essa classificação, e mostram que possivelmente o ancestral mais recente do grupo existiu a 14-17~ milhões de anos (RASMUSSEN; CAMERON, 2010).

Estudos citogenéticos, demonstraram que o número cariotípico predominante em *Melipona* é 2n = 18, com variações apenas em *Melipona seminigra* Illiger, 1806 com 2n = 22, *Melipona quinquefasciata* Lepeletier, 1836 e *Melipona rufiventris* Lepeletier, 1836, sendo que as duas últimas diferem devido à presença de cromossomos B (CUNHA et al., 2021; SILVA et al., 2018). Dados obtidos por bandamento C revelaram uma característica peculiar desse grupo em relação à *c*-heterocromatina, que é distinta dos demais constituintes da tribo (CUNHA et al., 2021). Na maioria das espécies de Meliponini a *c*-heterocromatina é restrita a um dos braços cromossômicos (COSTA; BRITO; MIYAZAWA, 2004; GODOY; FERREIRA; LOPES, 2013; MIRANDA; FERNANDES; LOPES, 2013; ROCHA; POMPOLO; CAMPOS, 2003), já em *Melipona* essa região pode variar quanto a localização e quantidade, o que permite a separação desse gênero em dois grupos (LOPES et al., 2008, 2011; ROCHA et al., 2002; ROCHA; POMPOLO, 1998; ROCHA; POMPOLO; CAMPOS, 2003). O primeiro grupo é constituído de espécies com baixa quantidade de *c*-heterocromatina, menor que 50%, localizada na região pericentromérica dos cromossomos. Neste grupo encontramos espécies pertencentes aos subgêneros *Eomelipona*, *Melikerria* e *Melipona* sensu stricto. No segundo grupo, as espécies possuem alta quantidade de *c*-heterocromatina, maior que 50%, distribuída em quase toda a extensão dos cromossomos. No segundo grupo encontramos representantes dos subgêneros *Melikerria* e *Michmelia*. TAVARES et al (2010) corroboram com essa classificação citogenética quando analisaram o conteúdo de DNA de 15 espécies de *Melipona* e constataram que esse grupo possuía variação quanto ao tamanho do genoma, que estava relacionado com a quantidade de *c*-heterocromatina.

Visto as características desse gênero, alguns estudos vêm sendo desenvolvidos visando a compreensão da evolução da *c*-heterocromática nesse grupo de abelhas (CUNHA et al., 2018; PEREIRA et al., 2021; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018; TRAVENZOLI et al., 2019). Esses estudos, por sua vez revelam que possivelmente o ancestral putativo desse gênero possuía baixo conteúdo de *c*-heterocromatina e o alto conteúdo surgiu pelo menos duas vezes ao longo da evolução desse gênero, nas espécies de *Michmelia* e *Melikerria* (CUNHA; CAMPOS; LOPES, 2020; CUNHA et al., 2018; PEREIRA et al., 2021; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018). Além disso a composição da *c*-heterocromatina mostra-se divergente entre espécies de *Melipona* (PEREIRA et al., 2021; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018), assim possivelmente as sequências repetitivas que compõem essa região estão evoluindo de forma diferente. Isso é evidenciado

quando hibridizamos uma sequência repetitiva isolada da *c*-heterocromatina de uma espécie de *Michmelia* com espécies dos outros subgêneros e não observamos marcações (PEREIRA et al., 2021; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018). Além disso, observou-se a predominância em quase toda a *c*-heterocromatina de uma sequência de DNA repetitivo, quando analisadas espécies de alto conteúdo pertencentes ao subgênero *Michmelia* (LOPES et al., 2014; PEREIRA et al., 2021). No entanto, mesmo que os estudos mencionados tenham contribuído para ampliar o conhecimento da *c*-heterocromatina em *Melipona*, a investigação da composição e evolução dessa região de forma detalhada é necessária para entender os processos que levaram à amplificação maciça dessa região.

### 1.5 *Trigona*

O gênero *Trigona* Jurine, 1807, assim como *Melipona*, também constitui importantes polinizadores pertencentes à tribo Meliponini, a qual possui 32 espécies descritas, com distribuição na Região Neártica e Região Neotropical (CAMARGO; PEDRO, 2013). No Brasil, podemos encontrar representantes desse gênero ao longo de todo o território (CAMARGO; PEDRO, 2013). De acordo com dados moleculares e considerando as características dos ninhos, *Trigona* pode ser dividido em dois clados: A e B, cujo ancestral comum data cerca de 19 milhões de anos atrás (RASMUSSEN; CAMARGO, 2008; RASMUSSEN; CAMERON, 2010). Esses dois clados englobam 9 grupos: "*pallens"*, *"fulviventris"*, *"cilipes"*, *"amalthea", "spinipes", "fuscipennis", "crassipes"* e *"recursa"* (RASMUSSEN; CAMARGO, 2008).

De acordo com os dados citogenéticos, as espécies de *Trigona* apresentam cariótipo com número de cromossomos conservado, sendo 2n = 34, a única exceção é *Trigona braueri* Friese, 1900 (como *Trigona fulviventris*) com 2n = 32 (COSTA; BRITO; MIYAZAWA, 2004; TRINDADE DOMINGUES et al., 2005). Diferente de *Melipona,* a distribuição da *c*-heterocromatina é mais semelhante a de outros representantes da tribo, nos quais essa região se concentra em um dos braços cromossômicos (BARBOZA; COSTA, 2021; BARTH et al., 2011; COSTA; BRITO; MIYAZAWA, 2004; SANTOS et al., 2018; FERNANDES; BARTH; SAMPAIO, 2013; GODOY; FERREIRA; LOPES, 2013; TAVARES et al., 2021; DOMINGUES et al., 2005). O mapeamento de sonda microssatélites $(GA)_{15}$ e sonda de rDNA mostrou variações entre populações de *Trigona spinipes* (BARBOZA; COSTA, 2021; TAVARES et al., 2021). Esses microssatélites se localizaram principalmente em eucromatina e ambas as sondas mostraram variações quanto à localização e ao número de cromossomos que abrigam essas sequências (BARBOZA; COSTA, 2021; TAVARES et al., 2021). A escassez de trabalhos mostra o quão pouco se sabe sobre a constituição e evolução da *c*-heterocromatina neste grupo. Desse modo, o estudo com sequências repetitivas pode ser uma oportunidade para

o entendimento da evolução dessa região nos dois grupos, *Melipona* e *Trigona*, os quais possuem grande quantidade de *c*-heterocromatina distribuída em um padrão distinto nos cromossomos. Assim, no capítulo I e II realizamos uma análise citogenômica de seis espécies de *Melipona,* pertencentes aos quatro subgêneros. Nosso objetivo foi investigar a composição da *c*-heterocromatina, bem como as sequências e mecanismos envolvidos na amplificação e evolução dessa região.

No capítulo III objetivamos avaliar a evolução da *c*-heterocromatina em espécies do gênero *Trigona*, cujas características em relação a distribuição dessa região se assemelham a outros gêneros de Meliponini. Nesse estudo aplicamos diferentes técnicas, incluindo enzimas de restrição, sequenciamento de nova geração e citogenética.

## 2. REFERÊNCIAS BIBLIOGRÁFICAS

ALLSHIRE, R. C.; MADHANI, H. D. Ten principles of heterochromatin formation and function. **Nature Reviews Molecular Cell Biology**, 2018.

ALTEMOSE, N. et al. Genomic Characterization of Large Heterochromatic Gaps in the Human Genome Assembly. **PLoS Computational Biology**, v. 10, n. 5, p. e1003628, 15 maio 2014.

BARBOZA, V. P.; COSTA, M. A. Cytogenetic Analysis in *Trigona spinipes* Fabricius (Hymenoptera, Meliponina) Reveals Intraspecific Variation. **Neotropical Entomology**, v. 50, n. 5, p. 846–849, 1 out. 2021.

BARTH, A. et al. Occurrence of b chromosomes in *Tetragonisca* latreille, 1811 (hymenoptera, apidae, meliponini): A new contribution to the cytotaxonomy of the genus. **Genetics and Molecular Biology**, v. 34, n. 1, p. 77–79, 2011.

BISCOTTI, M. A. et al. Transcription of tandemly repetitive DNA: functional roles. **Chromosome Research**, v. 23, n. 3, p. 463–477, 25 set. 2015.

BOSTOCK, C. A function for satellite DNA? **Trends in Biochemical Sciences**, v. 5, n. 5, p. 117–119, maio 1980.

BRITTEN, R. J.; DAVIDSON, E. H. Repetitive and Non-Repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty. **The Quarterly Review of Biology**, v. 46, n. 2, p. 111–138, jun. 1971.

CAMACHO, J. P. M. et al. Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. **BMC Biology**, v. 20, n. 1, p. 36, 7 dez. 2022.

CAMARGO JMF; PEDRO SRM. Meliponini Lepeletier, 1836. Em Moure, JS, Urbano, D. & Melo, GAR (Orgs). **Catálogo de Abelhas (Hymenoptera, Apoidea) na Região Neotropical - versão online.** Disponível online em: http://www.moure.cria.org.br/catalogue. Acesso em: 20 de junho de 2022.

CARBONE, L. et al. Gibbon genome and the fast karyotype evolution of small apes. **Nature**, v. 513, n. 7517, p. 195–201, 11 set. 2014.

CHARLESWORTH, B.; SNIEGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v. 371, n. 6494, p. 215–220, set. 1994.

COHEN, S.; SEGAL, D. Extrachromosomal Circular DNA in Eukaryotes: Possible Involvement in the Plasticity of Tandem Repeats. **Cytogenetic and Genome Research**, v. 124, n. 3–4, p. 327–338, 2009.

COSTA, K. F.; BRITO, R. M.; MIYAZAWA, C. S. Karyotypic description of four species of *Trigona* (Jurine, 1807) (Hymenoptera, Apidae, Meliponini) from the State of Mato Grosso, Brazil. **Genetics and Molecular Biology**, v. 27, n. 2, p. 187–190, 2004.

CUNHA, M. S. et al. The Bee Chromosome database (Hymenoptera: Apidae). **Apidologie**, v. 52, n. 2, p. 493–502, 25 abr. 2021.

CUNHA, M. S.; CAMPOS, L. A. O.; LOPES, D. M. Insights into the heterochromatin evolution in the genus *Melipona* (Apidae: Meliponini). **Insectes Sociaux**, v. 67, n. 3, p. 391–398, 27 ago. 2020.

CUNHA, M. S. et al. Comparative cytogenetics in three *Melipona* species (Hymenoptera: Apidae) with two divergent heterochromatic patterns. **Genetics and Molecular Biology**, v. 41, n. 4, p. 806–813, 2018.

FERNANDES, A.; BARTH, A.; SAMPAIO, W. S. Caracterização citogenética da espécie *Trigona chanchamayoensis* (Hymenoptera, Apidae, Meliponini) encontrada no cerrado brasileiro. **Evolução e Conservação da Biodiversidade**, v. 4, n. 2, p. 63, 1 dez. 2013.

FERREE, P. M.; PRASAD, S. How Can Satellite DNA Divergence Cause Reproductive Isolation? Let Us Count the Chromosomal Ways. **Genetics Research International**, v. 2012, p. 1–11, 29 jan. 2012.

GARRIDO-RAMOS, M. Satellite DNA: An Evolving Topic. **Genes**, v. 8, n. 9, p. 230, 18 set. 2017.

GIANNINI, T. C. et al. Crop pollinators in Brazil: a review of reported interactions. **Apidologie**, v. 46, n. 2, p. 209–223, 18 mar. 2015.

GODOY, D. C.; FERREIRA, R. P.; LOPES, D. M. Chromosomal Variation and Cytogenetics of *Plebeia lucii* and *P. phrynostoma* (Hymenoptera: Apidae). **Florida Entomologist**, v. 96, n. 4, p. 1559–1566, dez. 2013.

GRABUNDZIJA, I. et al. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. **Nature Communications**, v. 7, n. 1, p. 10716, 2 abr. 2016.

GRAY, Y. H. M. It takes two transposons to tango:transposable-element-mediated chromosomal rearrangements. **Trends in Genetics**, v. 16, n. 10, p. 461–468, out. 2000.

GREGORY, T. R. Synergy between sequence and size in Large-scale genomics. **Nature Reviews Genetics**, v. 6, n. 9, p. 699–708, set. 2005.

GREWAL, S. I. S.; JIA, S. Heterochromatin revisited. **Nature Reviews Genetics**, v. 8, n. 1, p. 35–46, jan. 2007.

HEITZ, E. Das Heterochromatin der Moose. **Jahrbücher für Wissenschaftliche Botanik**, v. 69, p. 762–818, 1928.

HORZ, W.; ZACHAU, H. G. Characterization of Distinct Segments in Mouse Satellite DNA by Restriction Nucleases. **European Journal of Biochemistry**, v. 73, n. 2, p. 383–392, mar. 1977.

HUGHES, S. E.; HAWLEY, R. S. Heterochromatin: A Rapidly Evolving Species Barrier. **PLoS Biology**, v. 7, n. 10, p. e1000233, 27 out. 2009.

JAGANNATHAN, M.; YAMASHITA, Y. M. Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid Incompatibility in *Drosophila*. **Molecular Biology and Evolution**, v. 38, n. 11, p. 4977–4986, 27 out. 2021.

KIT, S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. **Journal of Molecular Biology**, v. 3, n. 6, p. 711-IN2, dez. 1961.

KLEIN, S. J.; O'NEILL, R. J. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. **Chromosome Research**, v. 26, n. 1–2, p. 5–23, 13 mar. 2018.

LI, B. et al. Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. **The Plant Journal**, v. 73, n. 6, p. 952–965, mar. 2013.

LIMA, L. G.; RUIZ-RUANO, F. J. In-Depth Satellitome Analyses of 37 *Drosophila* Species Illuminate Repetitive DNA Evolution in the *Drosophila* Genus. **Genome Biology and Evolution**, v. 14, n. 5, 3 maio 2022.

LIPPMAN, Z.; MARTIENSSEN, R. The role of RNA interference in heterochromatic silencing. **Nature**, v. 431, n. 7006, p. 364–370, 15 set. 2004.

LIU, J.; ALI, M.; ZHOU, Q. Establishment and evolution of heterochromatin. **Annals of the New York Academy of Sciences**, 2020.

LOPES, D. M. et al. Cytogenetic characterization of *Melipona rufiventris* Lepeletier 1836 and *Melipona mondury* Smith 1863 (Hymenoptera, Apidae) by C banding and fluorochromes staining. **Genetics and Molecular Biology**, v. 31, n. 1, p. 49–52, 2008.

LOPES, D. M. et al. Cytogenetics of three *Melipona* species (Hymenoptera, Apidae, Meliponini). **Sociobiology**, v. 58, n. 1, p. 185–194, 2011.

LOPES, D. M. et al. Similarity of heterochromatic regions in the stingless bees (Hymenoptera: Meliponini) revealed by chromosome painting. **Caryologia**, v. 67, n. 3, p. 222–226, 3 jul. 2014.

MARSANO, R. M.; DIMITRI, P. Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements. **Cells**, v. 11, n. 5, p. 761, 22 fev. 2022.

MCCLINTOCK, B. The origin and behavior of mutable loci in maize. **Proceedings of the National Academy of Sciences**, v. 36, n. 6, p. 344–355, 30 jun. 1950.

MCCLINTOCK, B. Chromosome organization and genic expression. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 16, n. 0, p. 13–47, 1 jan. 1951.

MCCLINTOCK, B. Controlling Elements and the Gene. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 21, n. 0, p. 197–216, 1 jan. 1956.

MEŠTROVIĆ, N. et al. Structural and functional liaisons between transposable elements and satellite DNAs. **Chromosome Research**, v. 23, n. 3, p. 583–596, 21 set. 2015.

MICHENER CD. **The Bees of the World**. 2nd Ed. ed. University Press, Baltimore: [s.n.].

MIRANDA, R. V.; FERNANDES, A.; LOPES, D. M. Karyotype description of *Cephalotrigona femorata* Smith (Hymenoptera: Apidae) and the C-banding pattern as a specific marker for Cephalotrigona. **Sociobiology**, v. 60, n. 1, 27 mar. 2013.

NOVÁK, P.; NEUMANN, P.; MACAS, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. **Nature Protocols**, v. 15, n. 11, p. 3745–3776, 23 nov. 2020.

PAMPONÉT, V. C. C. et al. Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: citogenomic characterization of transposable elements and satellite DNA. **BMC Genomics**, v. 20, n. 1, p. 262, 2 dez. 2019.

PEREIRA, J. A. et al. Molecular cytogenetics in the study of repetitive sequences helping to understand the evolution of heterochromatin in *Melipona* (Hymenoptera, Meliponini). **Genetica**, v. 149, n. 1, 2021.

PICCOLI, M. C. A.; BARDELLA, V. B.; CABRAL-DE-MELLO, D. C. Repetitive DNAs in *Melipona scutellaris* (Hymenoptera: Apidae: Meliponidae): chromosomal distribution and test of multiple heterochromatin amplification in the genus. **Apidologie**, v. 49, n. 4, p. 497–504, 2018.

PIMPINELLI, S.; GATTI, M.; DE MARCO, A. Evidence for heterogeneity in heterochromatin of *Drosophila melanogaster*. **Nature**, v. 256, n. 5515, p. 335–337, jul. 1975.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA Evolution. In: [s.l: s.n.]. p. 126–152.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Centromere identity from the DNA point of view. **Chromosoma**, v. 123, n. 4, p. 313–325, 25 ago. 2014.

RAMÍREZ, S. R. et al. A molecular phylogeny of the stingless bee genus *Melipona* (Hymenoptera: Apidae). **Molecular Phylogenetics and Evolution**, v. 56, n. 2, p. 519–525, ago. 2010.

RASMUSSEN, C.; CAMARGO, J. M. F. A molecular phylogeny and the evolution of nest architecture and behavior in *Trigona s.s.* (Hymenoptera: Apidae: Meliponini). **Apidologie**, v. 39, n. 1, p. 102–118, 25 jan. 2008.

RASMUSSEN, C.; CAMERON, S. A. Global stingless bee phylogeny supports ancient divergence, vicariance, and long distance dispersal. **Biological Journal of the Linnean Society**, v. 99, n. 1, p. 206–232, 2010.

RATAN, Z. A. et al. Application of Fluorescence In Situ Hybridization (FISH) Technique for the Detection of Genetic Aberration in Medical Science. **Cureus**, 9 jun. 2017.

RICHARD, G.-F.; KERREST, A.; DUJON, B. Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. **Microbiology and Molecular Biology Reviews**, v. 72, n. 4, p. 686–727, dez. 2008.

ROCHA, M. P. et al. DNA characterization and karyotypic evolution in the bee genus *Melipona* (Hymenoptera, Meliponini). **Hereditas**, v. 136, n. 1, p. 19–27, abr. 2002.

ROCHA, M. P.; POMPOLO, S. DAS G. Karyotypes and heterochromatin variation (C-bands) in *Melipona* species (Hymenoptera, Apidae, Meliponinae). **Genetics and Molecular Biology**, v. 21, n. 1, p. 41–45, mar. 1998.

ROCHA, M.; POMPOLO, S.; CAMPOS, L. Citogenética da tribo Meliponini (Hymenoptera, Apidae). **UNESC, Criciúma**, p. 311–320, 2003.

RUIZ-RUANO, F. J. et al. High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports**, v. 6, n. 1, p. 28333, 7 set. 2016.

SANTOS, J. M. et al. Heterochromatin Distribution and Chromosomal Mapping of Microsatellite Repeats in the Genome of *Frieseomelitta* Stingless Bees (Hymenoptera: Apidae: Meliponini). **Florida Entomologist**, v. 101, n. 1, p. 33–39, 2018.

SERRANO-FREITAS, É. A. et al. Satellite DNA content of B chromosomes in the characid fish *Characidium gomesi* supports their origin from sex chromosomes. **Molecular Genetics and Genomics**, v. 295, n. 1, p. 195–207, 17 jan. 2020.

SILVA, A. A. et al. Karyotypic description of the stingless bee *Melipona quinquefasciata* Lepeletier, 1836 (Hymenoptera, Meliponini) with emphasis on the presence of B chromosomes. **Comparative Cytogenetics**, v. 12, n. 4, p. 471–482, 9 nov. 2018.

SINGER, M. F. Highly Repeated Sequences in Mammalian Genomes. In: [s.l: s.n.]. p. 67–112.

SLOTKIN, R. K.; MARTIENSSEN, R. Transposable elements and the epigenetic regulation of the genome. **Nature Reviews Genetics**, v. 8, n. 4, p. 272–285, abr. 2007.

STUART, T. et al. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. **eLife**, v. 5, 2 dez. 2016.

SUEOKA, N. Variation and heterogeneity of base composition of deoxyribonucleic acids: A compilation of old and new data. **Journal of Molecular Biology**, v. 3, n. 1, p. 31-IN15, fev. 1961.

TAVARES, M. G. et al. Karyotypic variation in the stingless bee *Trigona spinipes* (Hymenoptera: Apidae: Meliponini) from different geographical regions of Brazil. **Apidologie**, v. 52, n. 6, p. 1358–1367, 26 dez. 2021.

TAVARES, M. G.; CARVALHO, C. R.; SOARES, F. A. F. Genome size variation in *Melipona* species (Hymenoptera: Apidae) and sub-grouping by their DNA content. **Apidologie**, v. 41, n. 6, p. 636–642, 2 jun. 2010.

TRAVENZOLI, N. M. et al. Cytogenetic Analysis and Chromosomal Mapping of Repetitive DNA in *Melipona* Species (Hymenoptera, Meliponini). **Cytogenetic and Genome Research**, v. 158, n. 4, p. 213–224, 2019.

TRINDADE DOMINGUES, A. M. et al. Karyotype characterization of *Trigona fulviventris* Guérin, 1835 (Hymenoptera, Meliponini) by C banding and fluorochrome staining: Report of a new chromosome number in the genus. **Genetics and Molecular Biology**, v. 28, n. 3, p. 390–393, 2005.

WARING, M.; BRITTEN, R. J. Nucleotide Sequence Repetition: A Rapidly Reassociating Fraction of Mouse DNA. **Science**, v. 154, n. 3750, p. 791–794, 11 nov. 1966.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973–982, dez. 2007.

ZUCCOLO, A. et al. Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. **BMC Evolutionary Biology**, v. 7, n. 1, p. 152, 29 dez. 2007.

**CAPÍTULO I**

# The extensive amplification of heterochromatin in *Melipona* bees revealed by high throughput genomic and chromosomal analysis

Artigo publicado na revista *Chromosoma*

Chromosoma (2021) 130:251–262
https://doi.org/10.1007/s00412-021-00764-x

ORIGINAL ARTICLE

# The extensive amplification of heterochromatin in *Melipona* bees revealed by high throughput genomic and chromosomal analysis

Jaqueline A. Pereira[1] · Diogo Milani[2] · Ana Beatriz S. M. Ferretti[2] · Vanessa B. Bardella[2] · Diogo C. Cabral-de-Mello[2] · Denilce M. Lopes[1]

Received: 1 July 2021 / Revised: 18 October 2021 / Accepted: 29 October 2021 / Published online: 26 November 2021


## Abstract

Satellite DNAs (satDNAs) and transposable elements (TEs) are among the main components of constitutive heterochromatin (*c*-heterochromatin) and are related to their functionality, dynamics, and evolution. A peculiar case regarding the quantity and distribution of *c*-heterochromatin is observed in the genus of bees, *Melipona*, with species having a low amount of heterochromatin and species with high amount occupying almost all chromosomes. By combining low-pass genome sequencing and chromosomal analysis, we characterized the satDNAs and TEs of *Melipona quadrifasciata* (low *c*-heterochromatin) and *Melipona scutellaris* (high low *c*-heterochromatin) to understand *c*-heterochromatin composition and evolution. We identified 15 satDNA families and 20 TEs for both species. Significant variations in the repeat landscapes were observed between the species. In *M. quadrifasciata*, the repetitive fraction corresponded to only 3.78% of the genome library studied, whereas in *M. scutellaris*, it represented 54.95%. Massive quantitative and qualitative changes contributed to the differential amplification of *c*-heterochromatin, mainly due to the amplification of exclusive repetitions in *M. scutellaris*, as the satDNA MscuSat01-195 and the TE LTR/Gypsy_1 that represent 38.20 and 14.4% of its genome, respectively. The amplification of these two repeats is evident at the chromosomal level, with observation of their occurrence on most *c*-heterochromatin. Moreover, we detected repeats shared between species, revealing that they experienced mainly quantitative variations and varied in the organization on chromosomes and evolutionary patterns. Together, our data allow the discussion of patterns of evolution of repetitive DNAs and *c*-heterochromatin that occurred in a short period of time, after separation of the *Michmelia* and *Melipona* subgenera.

**Keywords** Satellite DNAs · Transposable elements · Repeatome · Stingless bees · Fluorescence in situ hybridization

## Introduction

Constitutive heterochromatin (*c*-heterochromatin) domains correspond to regions of DNA that are fundamental for the organization of eukaryotic genomes. They are characterized by dense condensation, paucity of genes, enrichment of multiple classes of repetitive DNA, low transcription levels, and late replication compared to euchromatin. Moreover, they ensure stability, centromere assembly, chromosome pairing, and segregation during cell division (Allshire and Madhani 2018). Due to the enrichment of repetitive DNA, which usually presents rapid turnover in nucleotide sequence and abundance, *c*-heterochromatin is highly dynamic (Plohl et al. 2012; Garrido-Ramos 2017; Liu et al. 2020).

Among the main components of *c*-heterochromatin are transposable elements (TEs) and satellite DNA (satDNA) (López-Flores and Garrido-Ramos 2012; Liu et al. 2020). TEs are usually scattered and can multiply in genomes; moreover, they are mobile and can change position on chromosomes (Bourque et al. 2018; Klein and O'Neill, 2018). The satDNAs are characterized as tandem repeats and generally tend to occupy specific chromosomal positions, mainly

✉ Diogo C. Cabral-de-Mello
cabral.mello@unesp.br

✉ Denilce M. Lopes
denilce.lopes@ufv.br

[1] Laboratório de Citogenética de Insetos, Departamento de Biologia Geral, Universidade Federal de Viçosa, Campus Viçosa, Av. P.H. Rolfs s/n, Viçosa, Minas Gerais CEP 36570-900, Brazil

[2] Departamento de Biologia Geral E Aplicada, Instituto de Biociências/IB, UNESP–Universidade Estadual Paulista, Rio Claro, São Paulo, Brazil

enriched with *c*-heterochromatin (Garrido-Ramos 2017). Molecular mechanisms, such as unequal crossing-over, gene conversion, intrastrand homologous recombination, transposition, slippage rolling circle replication, and re-insertion are involved in the evolution of satDNAs. In this way, satDNA members tend to occur with greater intraspecific similarity and greater interspecific divergence (Dover, 1982; Charlesworth et al. 1994; Garrido-Ramos 2017). In combination, the satDNAs and TEs are subject to amplifications, losses, and nucleotide sequence variation, resulting in qualitative and quantitative changes in genomes, which can favor chromosomal evolution by facilitating chromosomal rearrangements and differentiation between species (Plohl et al. 2012; Skinner and Griffin 2012; Bourque et al. 2018; Christmas et al. 2019).

*Melipona* stingless bees are significant pollinators, with over 76 species exclusively distributed in the Neotropical region (Michener 2007; Camargo and Pedro 2013). The genus is divided into four subgenera, *Melipona*, *Eomelipona*, *Melikerria*, and *Michmelia* whose diversification occurred ~ 14–17 million years (My) ago (Ramírez et al. 2010). Chromosomal studies in *Melipona* revealed a modal diploid number of 2n = 18, with slight variations to 2n = 22, and the occurrence of B chromosomes (Rocha and Pompolo 1998; Lopes et al. 2008; Silva et al. 2018; Cunha et al. 2020). An intriguing characteristic of *Melipona* chromosomes is the vast difference in the *c*-heterochromatin distribution revealed by classical cytogenetic techniques such as C-banding. According to the amount of *c*-heterochromatin, measured by the relative size of *c*-heterochromatin blocks with respect to the whole chromosome length, *Melipona* is artificially divided into two groups. Group I consists of species with low amounts of *c*-heterochromatin, less than 50% of the karyotype, located in the pericentromeric region of the chromosomes, and group II consists of species with high amounts of *c*-heterochromatin, more than 50% of the karyotype, distributed along almost the entire extension of the chromosomes. Group I are species belonging to the subgenera *Eomelipona*, *Melikerria*, and *Melipona s. str.*, while group II represents *Melikerria* and *Michmelia* (Rocha and Pompolo 1998; Rocha et al. 2002, 2003; Lopes et al. 2008, 2011). Putatively, as a consequence of variable amount of *c*-heterochromatin, the genome size of *Melipona* is also variable. Species with high amounts of *c*-heterochromatin have greater genome sizes ranging from 0.7 to 1.38 pg, while species with low *c*-heterochromatin have genome sizes ranging from 0.27 to 0.35 pg (Tavares et al. 2010).

The chromosomal studies, along with the phylogenetic hypothesis for *Melipona*, suggested that the putative ancestor had low heterochromatin content, and the high content appeared at least twice during the evolution of this group in species of the subgenera *Michmelia* and *Melikerria*. This was proposed because *Michmelia* species with high heterochromatin content is more phylogenetically related to species from *Eomelipona* with low heterochromatin content than with representatives of *Melikerria* that also present high heterochromatin content (Ramírez et al. 2010; Lopes et al. 2011; Piccoli et al. 2018; Travenzoli et al. 2019; Cunha et al. 2018, 2020; Pereira et al. 2021). The studies show that the general composition of *c*-heterochromatin is divergent between *Melipona* subgenera, but this information is limited to the chromosomal mapping of an anonymous repetitive DNA fraction ($C_0t$-DNA), to a unique repetitive DNA sequence obtained by enzymatic restriction, and to some microsatellites (Piccoli et al. 2018; Travenzoli et al. 2019; Cunha et al. 2018, 2020; Pereira et al. 2021).

This extensive difference in the amount of *c*-heterochromatin in *Melipona* provides a unique opportunity to study *c*-heterochromatin evolution from a molecular viewpoint, elucidating the mechanisms and specific sequences involved in the amplification of this genomic fraction. In this way, we aimed to advance our understanding of the composition and evolution of the intriguing *c*-heterochromatin in *Melipona* species. Thus, we performed a detailed analysis of satDNAs and main TEs, in *M. (Melipona) quadrifasciata* of group I and *M.* (*Michmelia*) *scutellaris* of group II. We combined high-throughput low-pass genome sequencing obtained from the Illumina platform, RepeatExplorer (Novák et al. 2013), satMiner (Ruiz-Ruano et al. 2016), and other bioinformatics protocols for repetitive DNA identification and characterization, followed by mapping of repeats on chromosomes. Our analysis focused on repetitive DNAs putatively related to the differential amplification of *c*-heterochromatin, revealing intense amplification of specific repetitive families between species.

## Materials and methods

### Species sampling, chromosome preparation, and DNA extraction

We collected larvae and adult bees of *M.* (*Melipona*) *quadrifasciata* (MQUA) and *M.* (*Michmelia*) *scutellaris* (MSCU) at two localities, Viçosa/Minas Gerais, Brazil (20° 45′ 17″ S, 42° 52′ 57″ W) and Encruzilhada/Bahia, Brazil (15° 31′ 53″ S, 40° 54′ 34″ W), respectively. To obtain mitotic chromosomes, we dissected the brain ganglia of post-defecating larvae according to Imai et al. (1988). Whole adults were stored in absolute ethanol for DNA extraction. Genomic DNA of the two species was obtained using the DNeasy kit (Qiagen Inc., CA, USA), according to the manufacturer's protocol.

## Genome sequencing and identification of repetitive sequences

We sequenced the genome of MSCU using Illumina Hiseq 2000 (Inc., San Diego, CA, USA), obtaining a paired-end library (2 × 101 bp) using the service of Macrogen Inc. (Korea). Sequencing of the MSCU genome provided 20,720,210 raw reads of 2 × 101 nt, corresponding to ~ 4 × of genome coverage. The genomic reads of this species were deposited in the Sequence Read Archive (SRA) under accession number PRJNA759201. The genome of MQUA is available in the SRA database (accession number SRX972105) and corresponds to the ~ 35 × of genome size. Similar to the MSCU genome, the genome of MQUA was sequenced using Illumina Hiseq 2000, obtaining paired-end raw reads (2 × 101 bp).

To identify the repetitive DNA sequences (satDNAs and TEs) putatively involved in the differential amplification of heterochromatin between MQUA and MSCU, we performed qualitative and quantitative comparative analyses between the two sequenced genomes. To this end, we used RepeatExplorer2 (Galaxy Version 1.3.1) to perform graph-based clustering analysis (Novák et al. 2013), available on the online public platform https://repeatexplorer-elixir.cerit-sc.cz/galaxy/ and the satMiner protocol (Ruiz-Ruano et al. 2016), available on GitHub (https://github.com/fjruizruano/satminer). The reads were pre-processed using "quality trimming tool," "FASTQ interlace" on paired-end reads, and "FASTQ to FASTA" converter. After pre-processing, to run the satMiner protocol (Ruiz-Ruano et al. 2016), we selected 2 × 200,000 reads of each genome for the first round of clustering. These steps were repeated until no satDNAs were identified; for this, two more clustering rounds were necessary, adding 2 × 300,000 reads and 2 × 400,000 reads for each species. For RepeatExplorer2, we randomly selected 500,000 reads from each genome (MQUA and MSCU). Subsequently, the clusters were analyzed, and we selected those with linear structure and Repbase v20.10 (Bao et al. 2015) annotation as putative TEs. For satDNAs, we selected clusters with spherical or ring-like shapes and high graphic densities. The contigs of each cluster were analyzed by dot plot graphs using Geneious software v4.8 (Drummond et al. 2009). Sequences that did not have a tandem structure were considered putative TEs. The TEs were confirmed by searching for similarity with families of TEs published in databases such as Repbase v20.10 (Bao et al. 2015). Sequences showing tandem structures were characterized as satDNAs. The satDNA and TEs were named according to the order of abundance in the genomes of each species, according to Ruiz-Ruano et al. (2016).

## Computational analysis of satDNAs and TEs

After identification of putative satDNAs and TEs, we compared all satDNA sequences to identify homology between them and classify as superfamilies, families, or variants, according to Ruiz-Ruano et al. (2016), following the same protocol with RepeatMasker v4.0.5 using a cross-match search engine. In addition, we also checked the similarity between sequences with the Muscle (Edgar 2004) alignment tool in Geneious software v4.8 (Drummond et al. 2009). To verify the relationship between variants of families for superfamily 1 (SF1; see results), we generated a minimum spanning tree with Arlequin v3.5 (Excoffier and Lischer 2010). Each indel position was considered a single change and represented the relative abundance for MQUA and MSCU.

To infer sequence abundance and divergence, we used 2 × 6 million paired pre-processed reads for both genomes by quality trimming with the same criteria used for the previous analyses. The abundance and divergence of each satDNA and TE sequence in the genome of each species were estimated using RepeatMasker (Smit et al. 2017). We inferred the average sequence divergence of the sequences by applying the Kimura 2-parameter model with the script calcDivergenceFromAlign.pl within the RepeatMasker suite. The abundance of each repetitive family was estimated as the proportion of nucleotides aligned with the reference consensus sequence divided by the size of the library. Finally, repeat landscapes were generated by comparing the sequence divergence versus the abundance of each satDNA and TE sequence among MQUA and MSCU to check the temporal amplification patterns.

## Primers design and PCR

For chromosomal analysis, we selected satDNAs and TEs that showed different abundances between species and those that were species-specific. Thus, we designed primers to amplify sequences by PCR and synthesized probes for fluorescence in situ hybridization (FISH) mapping. The consensus sequence of each family was used for the manual design of primers for satDNAs and TEs (Supplementary Table S1). PCR reactions were conducted using 10 × PCR Rxn Buffer, 0.2 mM $MgCl_2$, 0.16 mM dNTPs, 2 mM of each primer, 1 U *Taq* Platinum DNA Polymerase (Invitrogen, CA, USA), and 50–100 ng/µL template DNA. The amplification of the sequences by PCR used the following conditions: initial denaturation at 94 °C for 5 min, followed by 30 cycles at 94 °C (30 s), 55 °C (30 s), and 72 °C (80 s), and one final extension at 72 °C for 5 min. PCR products were checked on a 1% agarose gel. The monomeric bands were isolated and used as a source of reamplification for further analysis.

Sequences were deposited in GenBank under accession numbers MZ983437-MZ983471.

### C-banding and FISH chromosomal mapping of repeats

We followed for C-banding the protocol proposed by Rocha and Pompolo (1998) allowing *c*-heterochromatin detection. The fluorescence in situ hybridization (FISH) was performed according to Cabral-de-Mello and Marec (2021). The selected satDNA and TE sequences were labeled by the indirect method through nick translation or by PCR with digoxigenin-11-dUTP (Roche, Mannheim, Germany) or biotin-14-dATP (Invitrogen). To detect the signal, we used anti-digoxigenin-rhodamine (Roche Applied Science) and streptavidin Alexa Fluor 488 (Invitrogen). The chromosomes were contrasted with 2′,4-diamidino-2-phenylindole, dihydrochloride (DAPI), and slides were mounted with VECTASHIELD (Vector, Burlingame, CA, USA). Metaphases were analyzed using an Olympus BX61 microscope equipped with a fluorescence lamp and appropriate filters. The images were captured using a DP70-cooled digital camera in grayscale. The FISH signals were pseudo-colored in red, and images of chromosomes and signals were combined and optimized for brightness and contrast using Adobe Photoshop CS2 software.

## Results

### Extensive difference between repeatome of two Melipona species with high and low c-heterochromatin content

We identified the heterochromatin of MQUA restricted to the pericentromeric region of three metacentric/submetacentric chromosomes and the short arm of the acrocentric chromosome (Fig. 1a). In contrast, heterochromatin occupied almost the entire extension of the chromosomes of MSCU (Fig. 1a), as previously reported (Rocha and Pompolo 1998). Furthermore, we observed a highly divergent amount of satDNAs and TEs, between MQUA and MSCU. The repetitive fraction characterized here comprised only 3.78% of the MQUA genome library, corresponding to about 0.00998 Gb of its 0.26 Gb total genome (Tavares et al. 2010). Meanwhile, in MSCU, the repetitive DNA comprised 54.95% of the genome library, corresponding to about 0.58 Gb of its 1.056 Gb total genome (Tavares et al. 2010) (Table 1, Fig. 1b,c). In general, the accumulation (calculated by abundance) and divergence (measured by Kimura-2 parameter) among the repeats differed greatly between species, highlighting the differentiation of repetitive DNA profiles (Fig. 1d).

Among all repetitive DNAs, we identified 15 families of satDNAs, of which 11 were shared between MQUA and MSCU (Table 1). The sequence monomer length ranged from 20 to 586 bp, and the A + T sequence content ranged from 38.3 to 70%. These satDNAs comprised 2.84% of the MQUA genome and 38.41% of the MSCU genome (Table 1, Fig. 1b). The most abundant satDNA in MQUA (MquaSat01-586) corresponded to 0.949% of its genome, representing 25.11% of the repetitive DNA genome fraction characterized. In contrast, the most abundant satDNA in MSCU (MscuSat01-195) presented an abundance of 38.20%, corresponding to 69.53% of the characterized repetitive DNA sequences, or 0.403 Gb of the genome of the species (Fig. 1c). The less abundant satDNA in MQUA (MquaSat07-20) represented 0.5741% of the genome, and in MSCU the lowest abundance was 0.00021% for MquaSat08_MscuSat13-306 (Table 1).

For the 11 satDNAs shared between the two species, we noticed that the abundance was quite variable, as follows: seven satDNA families, specifically MquaSat06_MscuSat02-145, MquaSat04_MscuSat03-195, MquaSat09_MscuSat09-43, MquaSat02_MscuSat10-196, MquaSat05_MscuSat11-225, MquaSat03_MscuSat12-146, and MquaSat08_MscuSat13-306, were more represented in the MQUA genome, whereas four, specifically MquaSat11_MscuSat05-147, MquaSat13_MscuSat06-146, MquaSat12_MscuSat07-145, and MquaSat10_MscuSat08-145, were more represented in MSCU genome. The differences of abundances for the satDNAs identified in the two species were from approximately 1.8 times more (MquaSat06_MscuSat02-145) to more than one thousand times (MquaSat03_MscuSat12-146), both more represented in the MQUA genome. Two of the four satDNA families identified in only one species were exclusive for MQUA (MquaSat01-586 and MquaSat07-20) and two for MSCU (MscuSat01-195 and MscuSat04-115) (Table 1).

The average sequence divergence of the 13 satDNA families in MQUA was 11.19%, ranging from 1.65 to 25.80% (standard deviation of 9%). In MSCU, the average divergence for 13 satDNA families was 14.48%, varying from 2.11 to 30.3% (standard deviation 7.66%) (Table 1). By inspecting satDNA landscape plots (abundance versus divergence) for each satDNA family, we observed distinct profiles of amplification, homogeneous (new variants) or differentiated (old variants), depending on the repetitive DNA element (Fig. 2, Supplementary Fig. 1). For the two satDNA families exclusive of the MSCU genome, we observed two peaks of sequence abundance for MscuSat01-195, one at low divergence and the second with higher divergence, which was more abundant, and for MscuSat04-115, a single sequence abundance peak at low divergence was observed (Fig. 2a). Interestingly, between the two peaks of MscuSat01-195, a medium degree of sequence abundance was

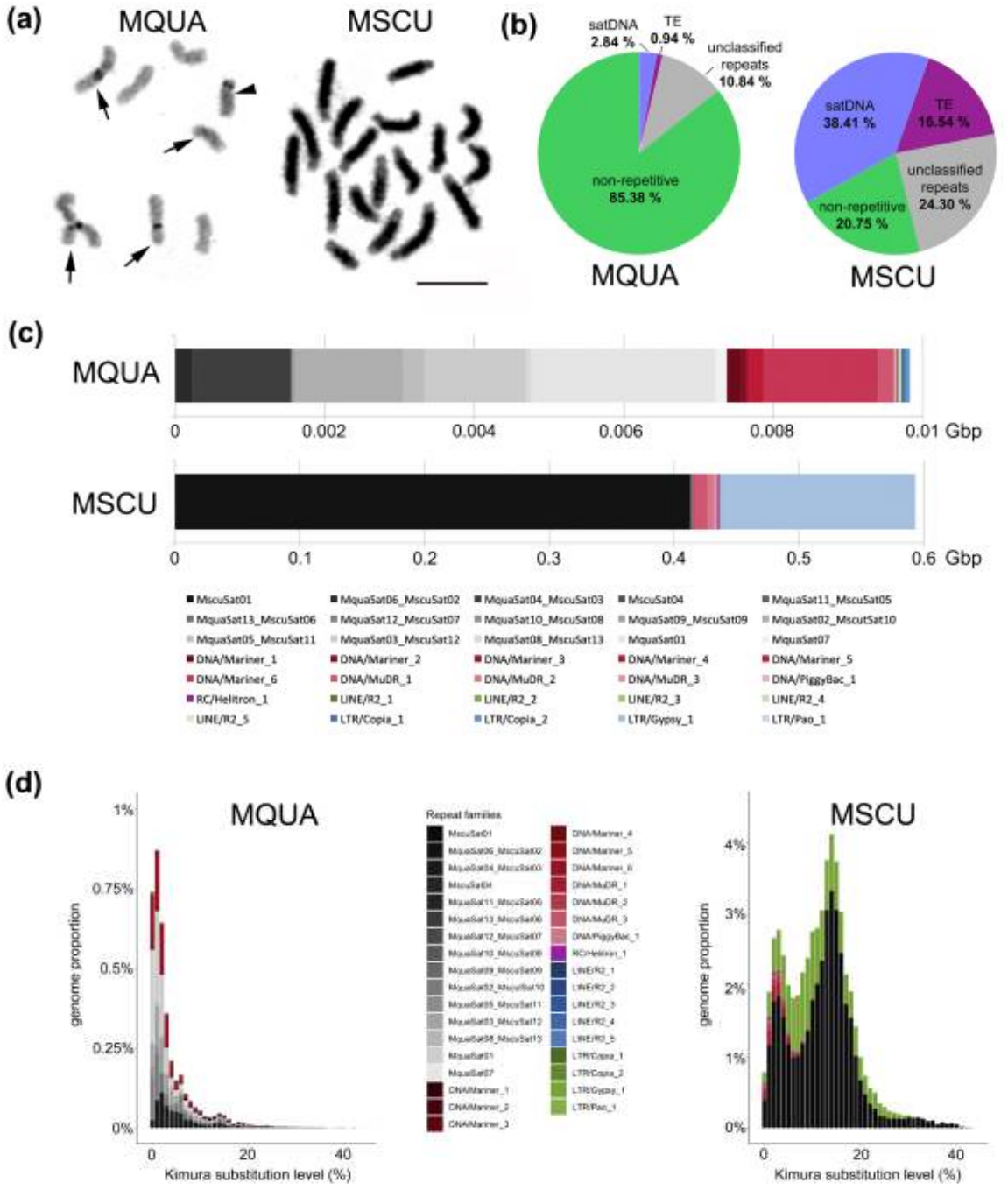



**Fig. 1** Summary of repetitive DNAs identified in the genomes of two *Melipona* species, *M. quadrifasciata* (MQUA) and *M. scutellaris* (MSCU). **a** *c*-Heterochromatin distribution revealed by C-banding technique; **b** genome composition, i.e. repetitive fraction and non-repetitive DNAs; **c** proportion of the distinct repetitive DNAs compared to the genomic size of each species; **d** landscape (abundance versus divergence) for all repeats identified. Observe the high difference in *c*-heterochromatin content in (**a**) and the specific repetitive DNAs in (b-d). In (**a**), the arrows point the chromosomes with pericentromeric *c*-heterochromatin block, and the arrowhead shows a chromosome with *c*-heterochromatin on one chromosomal arm. Scale bar = 5 µm

**Table 1** Characteristics of repetitive DNAs (satDNA and TEs) identified in the genome of two *Melipona* species, *M. quadrifasciata* (MQUA) and *M. scutellaris* (MSCU). SF: superfamily, ML: monomer length, A + T%: A + T sequence content

| Repetitive elemento | SF | ML | A + T% | Abundance % | | Divergence % | | Abundance (coefficient) |
|---|---|---|---|---|---|---|---|---|
| | | | | MQUA | MSCU | MQUA | MSCU | MSCU/MQUA |
| Satellite DNA | | | | | | | | |
| MscuSat01-195 | | 195 | 61.5 | 0.00000 | 38.2062 | 0.00 | 12.58 | |
| MquaSat06_MscuSat02-145 | 1 | 145 | 52.4 | 0.08749 | 0.04904 | 14.38 | 16.33 | 0.56052 |
| MquaSat04_MscuSat03-195 | | 195 | 65.1 | 0.50782 | 0.04779 | 4.97 | 30.30 | 0.09411 |
| MscuSat04-115 | | 115 | 38.3 | 0.00000 | 0.04234 | 0.00 | 5.92 | |
| MquaSat11_MscuSat05-147 | 1 | 147 | 54.4 | 0.00167 | 0.01771 | 20.13 | 10.76 | 10.60479 |
| MquaSat13_MscuSat06-146 | 1 | 146 | 59.6 | 0.00049 | 0.01306 | 22.11 | 13.22 | 26.65306 |
| MquaSat12_MscuSat07-145 | 1 | 145 | 51.7 | 0.00141 | 0.01076 | 23.58 | 7.88 | 7.63121 |
| MquaSat10_MscuSat08-145 | 1 | 145 | 53.8 | 0.00199 | 0.01040 | 25.80 | 11.08 | 5.22613 |
| MquaSat09_MscuSat09-43 | | 43 | 41.9 | 0.01838 | 0.00687 | 11.58 | 2.11 | 0.37378 |
| MquaSat02_MscuSat10-196 | | 196 | 62.2 | 0.55339 | 0.00372 | 3.83 | 21.09 | 0.00672 |
| MquaSat05_MscuSat11-225 | | 225 | 41.3 | 0.10950 | 0.00232 | 3.27 | 21.14 | 0.02119 |
| MquaSat03_MscuSat12-146 | 1 | 146 | 59.6 | 0.52112 | 0.00046 | 3.71 | 22.37 | 0.00088 |
| MquaSat08_MscuSat13-306 | | 306 | 67.3 | 0.02697 | 0.00021 | 1.90 | 13.52 | 0.00779 |
| MquaSat01-586 | | 586 | 57 | 0.94929 | 0.00000 | 1.65 | 0.00 | |
| MquaSat07-20 | | 20 | 70 | 0.05741 | 0.00000 | 8.60 | 0.00 | |
| Total | | | | **2.83692** | **38.41095** | | | |
| Transposable elements | | | | | | | | |
| DNA/Mariner_1 | | | | 0.07450 | 0.02904 | 13.51 | 8.78 | 0.38979 |
| DNA/Mariner_2 | | | | 0.02382 | 0.00687 | 4.64 | 10.49 | 0.28841 |
| DNA/Mariner_3 | | | | 0.00054 | 0.01334 | 8.28 | 0.61 | 24.7037 |
| DNA/Mariner_4 | | | | 0.01587 | 0.00844 | 12.88 | 5.63 | 0.53182 |
| DNA/Mariner_5 | | | | 0.07366 | 0.00428 | 3.07 | 7.70 | 0.58104 |
| DNA/Mariner_6 | | | | 0.58672 | 0.00050 | 1.96 | 28.48 | 0.00085 |
| DNA/MuDR_1 | | | | 0.08187 | 1.04175 | 7.15 | 2.92 | 12.7244 |
| DNA/MuDR_2 | | | | 0.00551 | 0.51516 | 12.73 | 3.26 | 93.4954 |
| DNA/MuDR_3 | | | | 0.00505 | 0.22556 | 12.29 | 2.42 | 44.6653 |
| DNA/PiggyBac_1 | | | | 0.00357 | 0.01313 | 22.72 | 2.58 | 3.67787 |
| LINE/R2_1 | | | | 0.00237 | 0.00488 | 14.88 | 6.03 | 2.05907 |
| LINE/R2_2 | | | | 0.00220 | 0.00568 | 13.49 | 5.33 | 2.58181 |
| LINE/R2_3 | | | | 0.00685 | 0.01540 | 13.02 | 4.94 | 2.24817 |
| LINE/R2_4 | | | | 0.00440 | 0.00880 | 13.87 | 7.38 | 2 |
| LINE/R2_5 | | | | 0.00413 | 0.00940 | 9.55 | 3.96 | 2.27602 |
| LTR/Copia_1 | | | | 0.02164 | 0.00235 | 4.13 | 11.23 | 0.10859 |
| LTR/Copia_2 | | | | 0.02144 | 0.00356 | 4.89 | 11.17 | 0.16604 |
| LTR/Gypsy_1 | | | | 0.00000 | 14.47401 | 0.00 | 11.72 | |
| LTR/Pao_1 | | | | 0.00428 | 0.02023 | 18.89 | 3.81 | 4.72663 |
| RC/Helitron_1 | | | | 0.00486 | 0.13519 | 10.89 | 2.43 | 27.8168 |
| Total | | | | 0.94329 | 16.53756 | | | |
| All known repeats | | | | 3.78021 | 54.94851 | | | |
| Non characterized repeats | | | | 10.83959 | 24.30222 | | | |
| All repeats total | | | | 14.61980 | 79.25073 | | | |
| Non repeats total | | | | 85.38020 | 20.74927 | | | |
| Genome size | | | | 0.26 Gb | 1.056 Gb | | | |

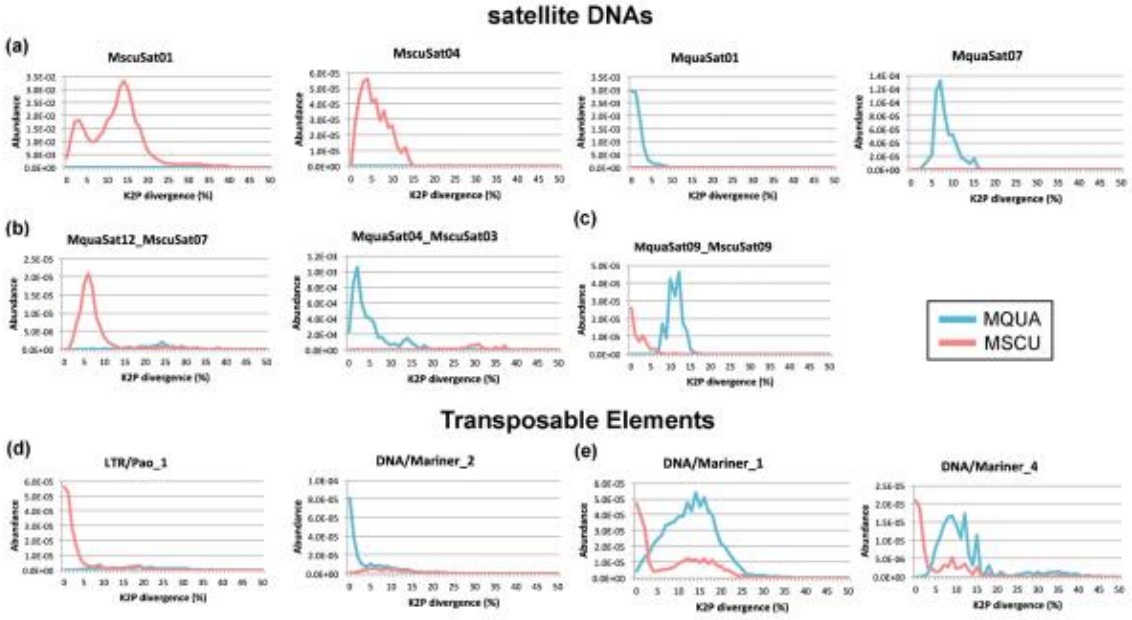



**Fig. 2** Comparative landscapes (abundance versus divergence) for the satDNAs (**a**–**c**) and TEs (**d**,**e**) identified in the genomes of the two *Melipona* species, *M. quadrifasciata* (MQUA, blue) and *M. scutellaris* (MSCU, red). For landscape plots, the repeat element divergence in K2P model distance to consensus is showed on the *x*-axis and the satDNA abundance on the *y*-axis. (**a**) satDNAs exclusively observed in the genome of only one species, showing distinct patterns of abundances in distinct divergences; (**b**) satDNAs shared between the genomes of both species, revealing expected accumulation of repeats with lower divergence in the species with more abundance of copies; (**c**) satDNAs shared between the genomes of both species with an accumulation of repeats with lower divergence in the species with less copy abundance; (**d**) TEs shared between species with an accumulation of lower divergent copies in the species with higher abundance for the repeat; (**e**) TEs shared between the two species with lower divergence in the species with less abundance for the repeat. In (**b**–**e**), selected examples for the distinct patterns, and the landscapes for the other repeats can be seen in Supplementary Figs. 1 (satDNAs) and 2 (TEs)

also observed. For MquaSat01-586 and MquaSat07-20, which occurred exclusively in the MQUA genome, we observed a single peak (Fig. 2a). The peak was in low divergence for MquaSat01-586 and a higher divergence for MquaSat07-20 (Fig. 2a). Concerning the shared satDNAs, as expected, in the species in which they were more abundant, they presented a more evident peak with low divergence, demonstrating the process of amplification (Fig. 2b). The exception was MquaSat09_MscuSat09-43, that although slightly less abundant in the MSCU genome, presenting a peak with lower divergence in comparison to the peak in MQUA (Fig. 2c).

Our analysis of sequence similarity identified one superfamily (SF) comprising six satDNA families. All satDNA families were shared between the two species, with variations in the sequence abundance of each family (Table 1). The six families showed slight variation in the length of the monomer (145–147 bp), so the differences between them were mainly due to the substitution of nucleotides that varied from 35 to 43 bp. For the family MquaSat06_MscuSat02-145, we noticed three sequence variants (named A, B, and C), while for the other five satDNA families, only one variant occurred. Therefore, we constructed a minimum spanning three (MST) for the six satDNA families belonging to SF1, that included the three variants of MquaSat06_MscuSat02-145, and found that one variant of MquaSat06_MscuSat02-145 (variant C) was the probable ancestor for the other satDNA families belonging to SF1 (Fig. 3).

Compared to satDNAs, the TEs that could be character ized were less abundant in the genome of both *Melipona*. We identified a total of 20 elements that represented 0.9432% of the MQUA and 16.64% of the MSCU genomes (Table 1, Fig. 1b). All TEs were shared between the two species, except for LTR/Gypsy_1, which was exclusively present in the MSCU genome, accounting for 14.4% of the total abundance (Table 1, Fig. 1c,d). Among the 19 shared TEs, seven were more abundant in the MQUA genome and 12 in the MSCU genome. Generally, we observed a minor difference in sequence abundance between the species, ranging from approximately 1.9 times more (DNA/Mariner_4 in MQUA) to approximately 93.5 times more (DNA/MuDR_2 in MSCU). The only exception was DNA/Mariner_6, which

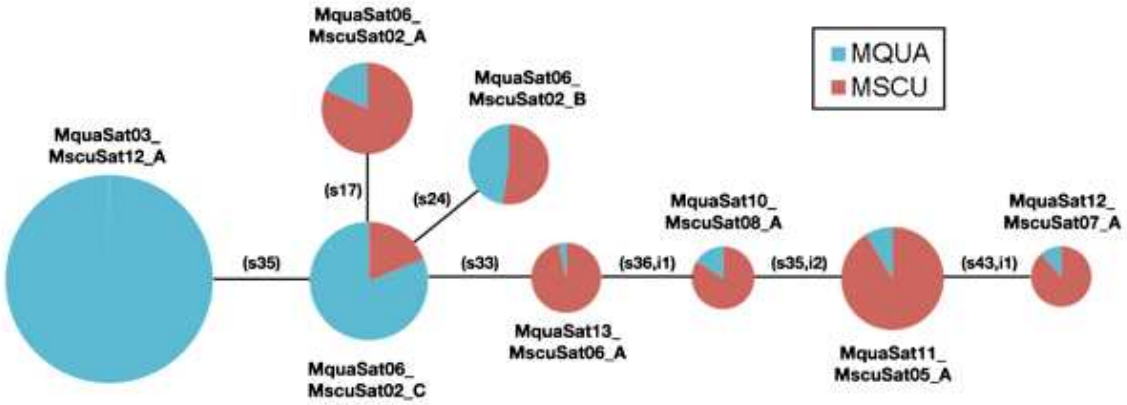



**Fig. 3** Minimum spanning tree (MST) representing the relationships for the six satDNA families (and their variants) belonging to SF1 superfamily in both species of *Melipona*, *M. quadrifasciata* (MQUA, blue) and *M. scutellaris* (MSCU, red). The sizes of the circles are proportional to their abundance. In parenthesis, the number of substitutions (s) or indels (i) between the distinct sequences are shown. Note: MquaSat06-MscuSat02-145 is the putative ancestor for the other satDNA families

had a low abundance (0.58672%) in the MQUA genome; it was more than one thousand times more abundant than MSCU (Table 1). The landscape plots for the TEs were consistent with the putative patterns of more recent sequence amplification with peaks with lower K2P in species that present more sequence abundance (Fig. 2d, Supplementary Fig. 2). Exceptions were observed for DNA/Mariner_1 and DNA/Mariner_4, which, although more abundant in the MQUA genome, presented a peak with lower K2P sequence divergence for MSCU copies (Fig. 2e).

Our similarity analysis between TEs and satDNAs did not reveal any sequence similarity, suggesting no origin relationships among them. The analysis of satDNA with TEs databases (Repbase) also did not retrieve any significant similarity. For MscuSat01-195 and LTR/Gypsy_1 that presented major abundances and similar chromosomal distribution (see next section), we also compared in detail each cluster/contig generated by RepeatExplorer and no similarity was noticed between them.

## Chromosomal mapping of repetitive DNAs reveals highly differentiated heterochromatin in Melipona

We studied the chromosomal distribution of mostly differentially enriched repetitive DNAs between the MQUA and MSCU genomes that comprised nine satDNAs and three TEs. Among the nine satDNAs, six were successfully mapped in the chromosomes of MQUA and five in MSCU; for TEs, we observed signals for one in MQUA and two in MSCU chromosomes. These repetitive sequences were located primarily in the heterochromatic region of the chromosomes in both species (Fig. 4). Some repeats that we identified by computational analysis of the genomes of the species did not reveal signals in chromosomes by FISH, suggesting non-clustered organization or small signals below FISH resolution (results not shown). FISH data is summarized on Supplementary Table S2.

In MQUA, the most representative satDNA (MquaSat01-586), which was also identified only in this species, was observed in the pericentromeric region of a single chromosome pair (Fig. 4a). In contrast, the more representative satDNA (MscuSat01-195) in MSCU occupied almost the entire heterochromatin extension of the chromosomes, showing spread dots (Fig. 4b). The other satDNA exclusive to the MSCU genome (MscuSat04-115) was located terminally in one pair of chromosomes, occupying the euchromatic region (Fig. 4c).

Among the six satDNAs shared between the two *Melipona* genomes, only two revealed signals in both species, MquaSat04_MscuSat03-195 (Fig. 4d,f) and MquaSat11_MscuSat05-147 (Fig. 4e,g). In MQUA, MquaSat04_MscuSat03-195 revealed signals in the pericentromeric region in two chromosome pairs (Fig. 4d), whereas in MSCU, this repeat was distributed in all chromosomes along heterochromatin (Fig. 4f). MquaSat11_MscuSat05-147 was present in one chromosome pair in the two species. However, in MQUA, it was located in the interstitial region (Fig. 4e), and in MSCU, it was the termini of the chromosomes (Fig. 4g). Among the four other satDNAs identified in the genomes of both *Melipona*, three revealed signals only on the chromosomes of MQUA (MquaSat02_MscuSat10-196, MquaSat03_MscuSat12-246, and MquaSat08_MscuSat13-306) and one in the chromosomes of MSCU (MquaSat13_MscuSat06-146) as follows: MquaSat02_MscuSat10-196 was located on the pericentromeric region of seven chromosomal pairs, (Fig. 4h); MquaSat03_MscuSat12-146 (Fig. 4i) and

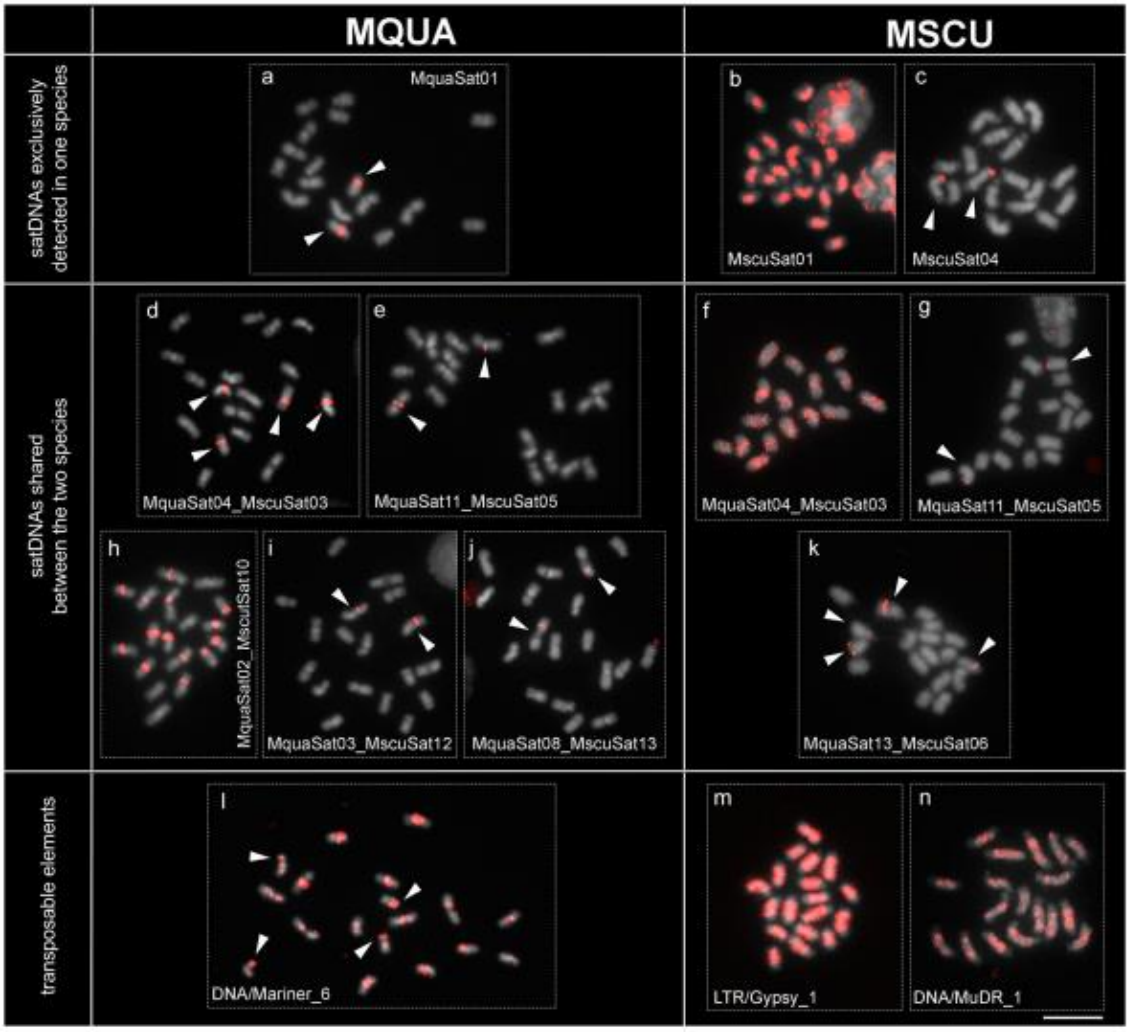



**Fig. 4** Chromosomal organization of repetitive DNAs revealed through FISH mapping in *Melipona* species. *M. quadrifasciata* (MQUA) and *M. scutellaris* (MSCU). (**a**–**c**) satDNAs with exclusive occurrence in one of the species; (**d**–**k**) satDNAs shared between species; (**i**–**n**) TEs. Gray corresponds to chromosomes stained with DAPI and red corresponds to FISH signals (probes). Arrowheads indicate chromosomes harboring signals for repetitive DNAs located in one or few chromosomes. Note the difference of hybridization patterns between the species with the occurrence of some repeats dispersed along almost the entire chromosomal extension in MSCU and a discrete location in MQUA. Scale bar = 10 µm

MquaSat08_MscuSat13-306 (Fig. 4j) were located in the interstitial region of different chromosome pairs of MQUA; MquaSat13_MscuSat06-146 showed clusters in the termini of one chromosomal pair and the interstitial region of different chromosome pair (Fig. 4k) in MSCU.

Concerning the three TEs, we observed FISH signals only for DNA/Mariner_6 in MQUA, which presented signals along the pericentromeric region of virtually all chromosomes and one chromosome arm of two chromosome pairs (Fig. 4l). The other two TEs, LTR/Gypsy_1 and DNA/MUDR_1, were observed only in MSCU chromosomes with the occurrence of spread signals along the *c*-heterochromatin extension (Fig. 4m,n). The signals of LTR/Gypsy_1 (Fig. 4m) were more intense than those of DNA/MUDR_1 (Fig. 4n), in accordance with the difference in sequence abundance between these two families.

## Discussion

The molecular composition and the causes of unusual amplification of the *c*-heterochromatin of *Melipona* species discovered more than 20 years ago (Rocha and Pompolo 1998), remained completely unknown until the recent years (Piccoli et al. 2018; Cunha et al. 2020; Pereira et al. 2021). Here, we advanced in the understanding of composition and amplification of *c*-heterochromatin in *Melipona* by characterizing the repeatome in MQUA and MSCU. Although the difference in genome size, as observed in *Melipona* (Tavares et al. 2010), also occurs in other species even at the intraspecies level (Alvarez-Fuster et al. 1991; Díez et al. 2013; Stelzer et al. 2019), the intense amplification of *c*-heterochromatin in congeneric species, colonizing almost entire chromosomes, is unusual and renders *Melipona* a special case. Genome size variation and enlargement can occur through diverse mechanisms, including polyploidization, supernumerary DNAs (B chromosomes), or repetitive DNA amplification (revised by Blommaert 2020). The last mechanism is the most parsimonious for enlargement of the MSCU genome compared to MQUA, as the two species have the same diploid number and no B chromosomes. Moreover, as other Apidae bees generally have genomes smaller than 0.5 Gb (Gregory 2021), it is a more parsimonious gain of repeats than repeated elimination. Supporting this hypothesis, we noticed significant differences between the proportion of repetitive DNA composition of the MQUA (3.78%) and MSCU (54.94%) genomes libraries, besides the observed difference in amount of *c*-heterochromatin as well as the difference genome size between the species (Rocha et al. 2002; Tavares et al. 2010). Most of these differences are attributed only to one satDNA (MscuSat01-195) and one TE (LTR/Gypsy_1), which are highly abundant in the MSCU genome and distributed along the most heterochromatic extension of MSCU, suggesting that these two elements are the leading players in *c*-heterochromatin amplification of this species.

Recent studies have shown that at least part of the repetitive DNA pool is shared between *Michmelia* species, occupying a large portion of *c*-heterochromatin (Piccoli et al. 2018; Cunha et al. 2020; Pereira et al. 2021). MSCU belongs to *Michmelia*; thus, the dissemination and amplification of MscuSat01-195 and LTR/Gypsy_1 in the genome can occur during the diversification of this subgenus. Our results support this idea, considering that in the MSCU genome, there are differentiated copies (old variants) for both sequences. These old variants could have originated in the common ancestor of the subgenus and it deserves investigation. Moreover, the occurrence of homogeneous sequences (new variants) for MscuSat01-195 and LTR/Gypsy_1, revealing more recent amplification of repeats is evident also. Although much less abundant compared to MscuSat01, another satDNA identified only in MSCU, MscuSat04-115, also contributed to heterochromatin amplification, presenting patterns of recent amplification, i.e. sequence homogeneity. In MQUA, despite the low heterochromatin content, some repeats also experienced some degree of amplification. This is the case for MquaSat01-586, which is exclusive to the genome of the species, corresponding to the most abundant repetitive DNA. This element is located in one chromosome pair and is homogeneous, indicating that this family may have experienced recent local amplification. The action of molecular mechanisms, such as unequal crossing-over, replication slippage, rolling circle amplification, and transposition, are among the main factors responsible for the spreading and amplifying repeats (Charlesworth et al. 1994) and consequently interfering with the content and structure of heterochromatin.

Although less representative in abundance, our investigation showed 25 repeats, shared between MQUA and MSCU, as part of the library present in the ancestor of *Melipona s. str* and *Michmelia* subgenus. Despite the conservation of the sequences, we observed variations in the number of copies, mainly for satDNA families. This is expected and in accordance with the Library Hypothesis, which predicts that phylogenetically related species share an ancestral collection of satDNA families that experience quantitative changes after speciation events, generating species specific or population specific profiles (Salser et al. 1976; Fry and Salser 1977; Landais et al. 2000; Ferree and Barbash 2009; Ricci et al. 2013; Wei et al. 2014; 2018; Feliciello et al. 2015; Palacios-Gimenez et al. 2020). For TEs, we noticed considerable differences in abundance for DNA/MuDR_1 and DNA/Mariner_6, and both were more abundant in MSCU, putatively due to independent recent activity, with the occurrence of homogeneous repeats that occupy the *c*-heterochromatin.

Among the shared satDNA families, we identified a relationship between six of them, comprising the SF1, all shared between the two species. This suggests that these families emerged in the common ancestor and experienced mainly quantitative changes and few nucleotide substitutions among them and between the two *Melipona* species. The accumulation of mutations can cause the emergence of new variants in a family of satDNA, which can be homogenized and fixed in the genome (Plohl et al. 2012). Thus, a set of satDNA families can originate from an ancestral family (Meštrović et al. 2006; Garrido-Ramos 2017) in the case of SF1 from variant C of MquaSat06_MscuSat02. Interestingly, FISH demonstrated that some members of SF1 occur in distinct positions on the chromosomes, as is the case for MquaSat11_MscuSat05 (long chromosomal arm of the second pair) and MquaSat03_MscuSat12 (short chromosomal arm of the second pair) in MQUA,

and MquaSat11_MscuSat05 (terminal region of one chromosome pair) and MquaSat13_MscuSat06 (interstitial regions of two chromosome pairs). This could be a consequence of local amplification of these repeats from ancestral short repeat arrays, as proposed in the grasshopper *Locusta migratoria* (Ruiz-Ruano et al. 2016).

Based on the location of 18S rDNA in *Melipona* species, it has been suggested that amplification of *c*-heterochromatin occurs from the interstitial region of the chromosomes (Cunha et al. 2018, 2020; Pereira et al. 2021). In species with low *c*-heterochromatin content, the 18S rDNA cluster is located in the interstitial region of one chromosome, whereas in species with high *c*-heterochromatin content, it is located in the terminal region of a chromosome pair. The satDNA MquaSat11_MscuSat05-147 support this hypothesis about starting point of heterochromatin amplification as our data for FISH mapping revealed that signal for this repeat displaced from the interstitial location in the MQUA to the terminal position in MSCU.

Taken together, our detailed characterization of satDNAs and TEs in the two *Melipona* expanded our knowledge about the evolutionary patterns of repetitive DNAs in the genus and species with highly differentiated amounts of *c*-heterochromatin. We report that the amplification and emergence of specific repetitive DNAs were the leading causes of heterochromatin expansion and genome diversification in short evolutionary periods since *Michmelia* subgenus diversification. On the contrary, we identified that despite intense amplification and differentiation of a few repeats, some repeats remained conserved, experiencing primarily quantitative changes. Finally, the study of other species belonging to the genus *Melipona* that also show differences in heterochromatin amount could serve as a model to help the broad understanding of rapid changes in heterochromatin from a phylogenetic perspective.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00412-021-00764-x.

**Acknowledgements** We acknowledge the anonymous reviewers for the suggestions and manuscript improvement.

**Funding** The authors would like to thank the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), process number 2019/19069–7. DCC-d-M is a recipient of a research productivity fellowship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (process number 308290/2020–8). VBB and ABSMF are supported by FAPESP, processes numbers 2018/21772–5 and 2020/06188–5, respectively.

## References

Alvarez-Fuster A, Juan C, Petitpierre E (1991) Genome size in *Tribolium* flour beetles: inter-and intraspecific variation. Genetics Res 58:1–5. https://doi.org/10.1017/S0016672300029542

Allshire RC, Madhani HD (2018) Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol 19:229. https://doi.org/10.1038/nrm.2017.119

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:1–6. https://doi.org/10.1186/s13100-015-0041-9

Blommaert J (2020) Genome size evolution: towards new model systems for old questions. Proc R Soc B 287:20201441. https://doi.org/10.1098/rspb.2020.1441

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan ST, Mager LD, Feschotte C (2018) Ten things you should know about transposable elements. Genome Biol 19:1–12. https://doi.org/10.1186/s13059-018-1577-z

Cabral-de-Mello DC, Marec F (2021) Protocolo de hibridização in situ de fluorescência universal (FISH) para mapear DNAs repetitivos em insetos e outros artrópodes. Mol Genet Genomics 296:513–526. https://doi.org/10.1007/s00438-021-01765-2

Camargo JMF, Pedro SRM (2013) Meliponini Lepeletier, 1836. In Moure, J. S., Urban, D. & Melo, G. A. R. (Orgs). Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region - online version. Available at http://www.moure.cria.org.br/catalogue. Accessed 10 Aug 2020.

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220. https://doi.org/10.1038/371215a0

Christmas MJ, Wallberg A, Bunikis I, Olsson A, Wallerman O, Webster MT (2019) Chromosomal inversions associated with environmental adaptation in honey-bees. Mol Ecol 28:1358–1374. https://doi.org/10.1111/mec.14941

Cunha MSD, Travenzoli NM, Ferreira RDP, Cassinela EK, Silva HBD, Oliveira FPM, Salomão TMF, Lopes DM (2018) Comparative cytogenetics in three *Melipona* species (Hymenoptera: Apidae) with two divergent heterochromatic patterns. Genet Mol Biol 41:806–813. https://doi.org/10.1590/1678-4685-GMB-2017-0330

Cunha MS, Campos LAO, Lopes DM (2020) Insights into the heterochromatin evolution in the genus *Melipona* (Apidae: Meliponini). Insect Soc 67:391–398. https://doi.org/10.1007/s00040-020-00773-6

Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, Eguiarte LE, Tenaillon MI (2013) Genome size variation in wild and cultivated maize along altitudinal gradients. New Phytol 199:264–276. https://doi.org/10.1111/nph.12247

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 199:111–117. https://doi.org/10.1016/0168-9525(86)90211-8

Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson S. (2009) Geneious v4.8. http://geneious.com. Accessed 19 Aug 2019

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: uma nova série de programas para realizar análises genéticas de populações em Linux e Windows. Mol Ecol Resour 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Feliciello I, Akrap I, Brajković J, Zlatar I, Ugarković Đ (2015) Satellite DNA as a driver of population divergence in the red flour beetle *Tribolium castaneum*. Genome Biol Evol 7:228–239. https://doi.org/10.1093/gbe/evu280

Ferree PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. PLoS Biol 7:e1000234. https://doi.org/10.1371/journal.pbio.1000234

Fry K, Salser W (1977) Nucleotide sequence of HS-a satellite DNA from kangaroo rat *Dipodomys ordii* and characterisation of similar sequences in other rodents. Cell 12:1069–1084. https://doi.org/10.1016/0092-8674(77)90170-2

Garrido-Ramos MA (2017) Satellite DNA: an Evolving Topic Genes 8:230. https://doi.org/10.3390/genes8090230

Gregory TR (2021) Animal genome size database. http://www.genomesize.com. Accessed 20 Apr 2021

Imai HT, Taylor RW, Crosland MW, Crozier RH (1988) Modes of spontaneous chromosomal mutation and karyotype evolution in ants with reference to the minimum interaction hypothesis. Jpn J Appl Entomol Z 63:159–185. https://doi.org/10.1266/jjg.63.159

Klein SJ, O'Neill RJ (2018) Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosom Res 26:5–23. https://doi.org/10.1007/s10577-017-9569-5

Liu J, Ali M, Zhou Q (2020) Establishment and evolution of heterochromatin. Ann N Y Acad Sci 1476:59–77. https://doi.org/10.1111/nyas.14303

Landais I, Chavigny P, Castagnone C, Pizzol J, Abad P, Vanlerberghe-Masutti F (2000) Characterization of a highly conserved satellite DNA from the parasitoid wasp *Trichogramma brassicae*. Gene 255:65–73. https://doi.org/10.1016/S0378-1119(00)00318-8

Lopes DM, Pompolo SDG, Campos LADO, Tavares MG (2008) Cytogenetic characterization of *Melipona rufiventris* Lepeletier 1836 and *Melipona mondury* Smith 1863 (Hymenoptera, Apidae) by C banding and fluorochromes staining. Genet Mol Biol 31:49–52. https://doi.org/10.1590/S1415-47572008000100010

Lopes DM, Fernandes A, Praça-Fontes MM, Werneck HDA, Resende HC, Campos LADO (2011) Cytogenetics of three *Melipona* species (Hymenoptera, Apidae, Meliponini). Sociobiology 58:185–194

López-Flores I, Garrido-Ramos MM (2012) The repetitive DNA content of eukaryotic genomes. Cytogenet Genome Res 7:1–28. https://doi.org/10.1159/000337118

Meštrović N, Castagnone-Sereno P, Plohl M (2006) High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. Gene 376:260–267. https://doi.org/10.1016/j.gene.2006.04.008

Michener CD (2007) The Bees of the World. The John Hopkins, 2nd Ed. University Press, Baltimore. pp953

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genomewide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics 29:792–793. https://doi.org/10.1093/bioinformatics/btt054

Palacios-Gimenez OM, Milani D, Song H et al (2020) Eight million years of satellite DNA evolution in grasshoppers of the genus Schistocerca illuminate the ins and outs of the library hypothesis. Genome Biol Evol 12:88–102. https://doi.org/10.1093/gbe/evaa018

Pereira JA, Travenzoli NM, de Oliveira MP, de Azevedo WH, Salomão TMF, Lopes DM (2021) Molecular cytogenetics in the study of repetitive sequences helping to understand the evolution of heterochromatin in *Melipona* (Hymenoptera, Meliponini). Genetica 149:55–62. https://doi.org/10.1007/s10709-020-00111-5

Piccoli MCA, Bardella VB, Cabral-de-Mello DC (2018) Repetitive DNAs in *Melipona scutellaris* (Hymenoptera: Apidae: Meliponidae): chromosomal distribution and test of multiple heterochromatin amplification in the genus. Apidologie 49:497–504. https://doi.org/10.1007/s13592-018-0577-z

Plohl M, Meštrović N, Mravinac B (2012) Satellite DNA evolution. Genome 7:126–152. https://doi.org/10.1159/000337122

Ramírez SR, Nieh JC, Quental TB, Roubik DW, Imperatriz-Fonseca VL, Pierce NE (2010) A molecular phylogeny of the stingless bee genus Melipona (Hymenoptera: Apidae). Mol Phylogenet Evol 56:519–525. https://doi.org/10.1016/j.ympev.2010.04.026

Ricci M, Luchetti A, Bonandin L, Mantovani B (2013) Random DNA libraries from three species of the stick insect genus Bacillus (Insecta: Phasmida): repetitive DNA characterization and first observation of polyneopteran MITEs. Genome 56:729–735. https://doi.org/10.1139/gen-2013-0107

Rocha MP, Pompolo SG (1998) Karyotypes and heterochromatin variation (C-bands) in *Melipona* species (Hymenoptera, Apidae, Meliponinae). Genet Mol Biol 21:41–45. https://doi.org/10.1590/S1415-47571998000100008

Rocha MP, Pompolo SG, Dergam JA, Fernandes A, Campos LAO (2002) DNA characterization and karyotypic evolution in the bee genus *Melipona* (Hymenoptera Meliponini). Hereditas 136:19–27. https://doi.org/10.1034/j.1601-5223.2002.1360104.x

Rocha MP, Cruz MP, Fernandes A, Waldschmidt AM, Silva JC, Pompolo SG (2003) Longitudinal differentiation in *Melipona mandacaia* (Hymenoptera, Meliponini) cromossomes. Hereditas 138:133–137. https://doi.org/10.1034/j.1601-5223.2003.01699.x

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016) High-throughput analysis of the satellitome illuminates satellite DNA evolution. Sci Rep 6:28333. https://doi.org/10.1038/srep28333

Salser W, Bowen S, Browne D, El-Adli F, Fedoroff N, Fry K, Heidell H, Paddock G, Wallace B, Whitcome P (1976) Investigation of the organization of mammalian chromosomes at the DNA sequences level. Fed Proc 35:23–35

Silva AA, Rocha MP, Pompolo SG, Campos LAO, Tavares MG (2018) Karyotypic description of the stingless bee Melipona quinquefasciata Lepeletier, 1836 (Hymenoptera, Meliponini) with emphasis on the presence of B chromosomes. Comp Cytogenet 12:471–482. https://doi.org/10.3897/CompCytogen.v12i4.29165

Skinner BM, Griffin DK (2012) Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity 108:37–41. https://doi.org/10.1038/hdy.2011.99

Smit AFA, Hubley R, Green P (2017) RepeatMasker Open-4.0. http://www.repeatmasker.org

Stelzer CP, Pichler M, Stadler P, Hatheuer A, Riss S (2019) Within-population genome size variation is mediated by multiple genomic elements that segregate independently during meiosis. Genome Biol Evol 11:3424–3435. https://doi.org/10.1093/gbe/evz253

Tavares MG, Carvalho CR, Soares FAF (2010) Genome size variation in *Melipona* species (Hymenoptera: Apidae) and sub-grouping by their DNA content. Apidologie 41:636–642. https://doi.org/10.1051/apido/20010023

Travenzoli NM, Lima BA, Cardoso DC, Dergam JA, Fernandes-Salomão TM, Lopes DM (2019) Cytogenetic analysis and chromosomal mapping of repetitive DNA in Melipona species (Hymenoptera, Meliponini). Cytogenet Genome Res 158:213–224. https://doi.org/10.1159/000501754

Wei KHC, Grenier JK, Barbash DA, Clark AG (2014) Correlated variation and population differentiation in satellite DNA abundance among lines of Drosophila melanogaster. Proc Natl Acad Sci 111:18793–18798. https://doi.org/10.1073/pnas.1421951112

Wei KHC, Lower SE, Caldas IV, Sless TJ, Barbash DA, Clark AG (2018) Variable rates of simple satellite gains across the *Drosophila* phylogeny. Mol Biol Evol 35:925–941. https://doi.org/10.1093/molbev/msy005

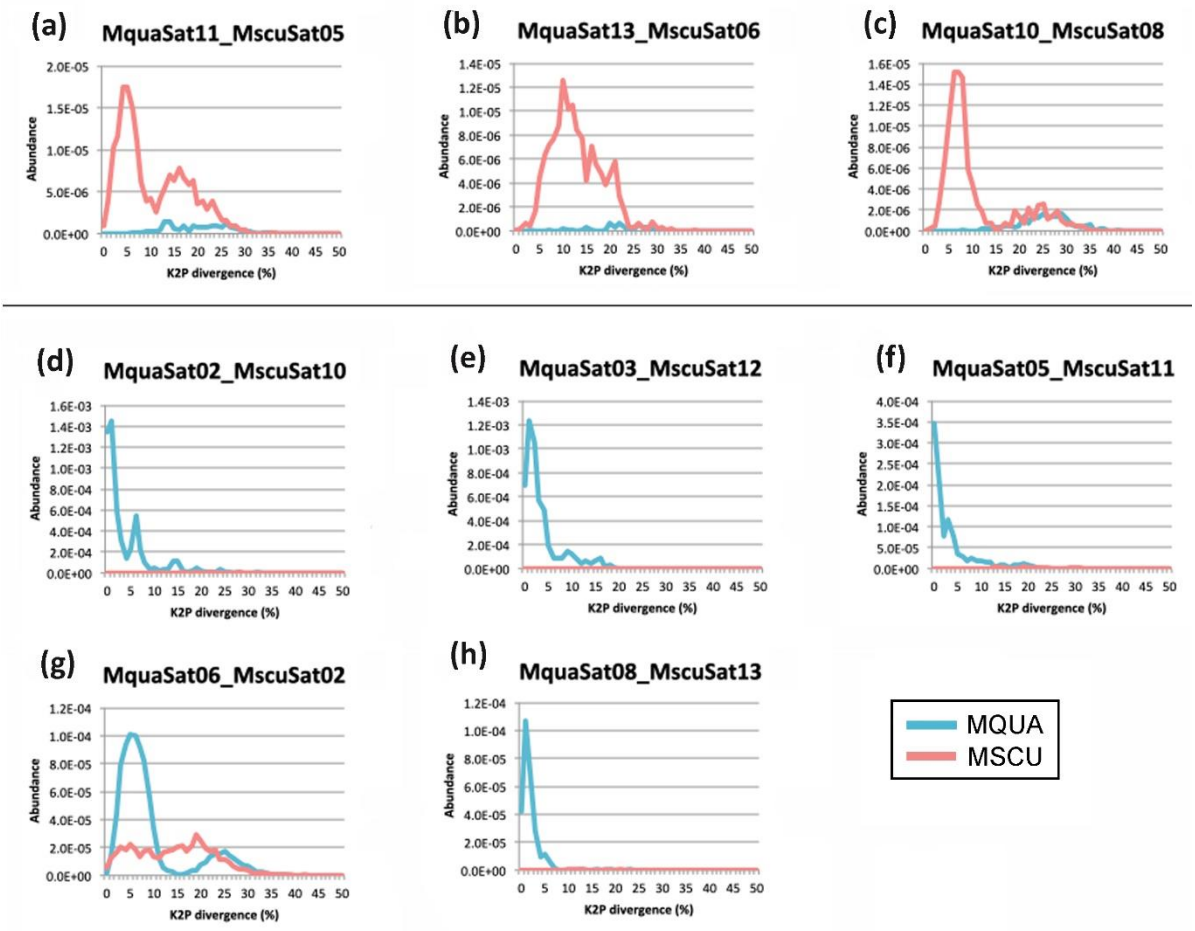
**Supporting information** - The extensive amplification of heterochromatin in *Melipona* bees revealed by high throughput genomic and chromosomal analysis

| Sequence names | Forward primers | Reverse primers |
|---|---|---|
| **Satellite DNA** | | |
| MscuSat01 | TCATTTGATGTTCTACAGCATG | CCGAGGAACGAGAAGAATCT |
| MquaSat04_MscuSat03 | GATTCAGTACAAACACGATTC | CCTTAGCAGGGATTGATAAAT |
| MscuSat04 | ACTTCGGCCGGCAGTGTC | CGTCCTTGGGTCTGCAACT |
| MquaSat11_MscuSat05 | TATCAGGAGAACCGCGCAT | CACAGGCTTTTCGAAAAATCGT |
| MquaSat13_MscuSat06 | TAGAAATTTCAGTTAGACAAAC | CTTTTGTATCAGGAGCATATG |
| MquaSat02_MscutSat10 | GGATTTCGACCAAACATGAC | GTGACTTCTGACAGATTTTTC |
| MquaSat03_MscuSat12 | GTTCTCCTGATACAGGCTTC | CGTTTATCCCTGGACAATGTA |
| MquaSat08_MscuSat13 | TTCAGACATAATGCCTCATAAA | ACTTACATTCTCACGCGAATTG |
| MquaSat01 | TGAACATGAATGTTATGCCTGA | GGTACACCGTATCACAAGTG |
| **Transposable Elements** | | |
| LTR/Gypsy_1 | ATTTTGATTTGTGTCGTTCTAT | GCGATGAGGTGCAGAATGA |
| DNA/MuDR_1 | TGGATATTGATACACGAGCATA | AATATCGTCGTGGTATTCCTAT |
| DNA/Mariner_6 | GACTCTAAGTGGCTGTGTTTA | TTTCATGTCGCTGGAGAAGC |

**Supplementary Table S1** - List of primers used for amplification through PCR of satDNA and TEs from *Melipona* genomic DNAs

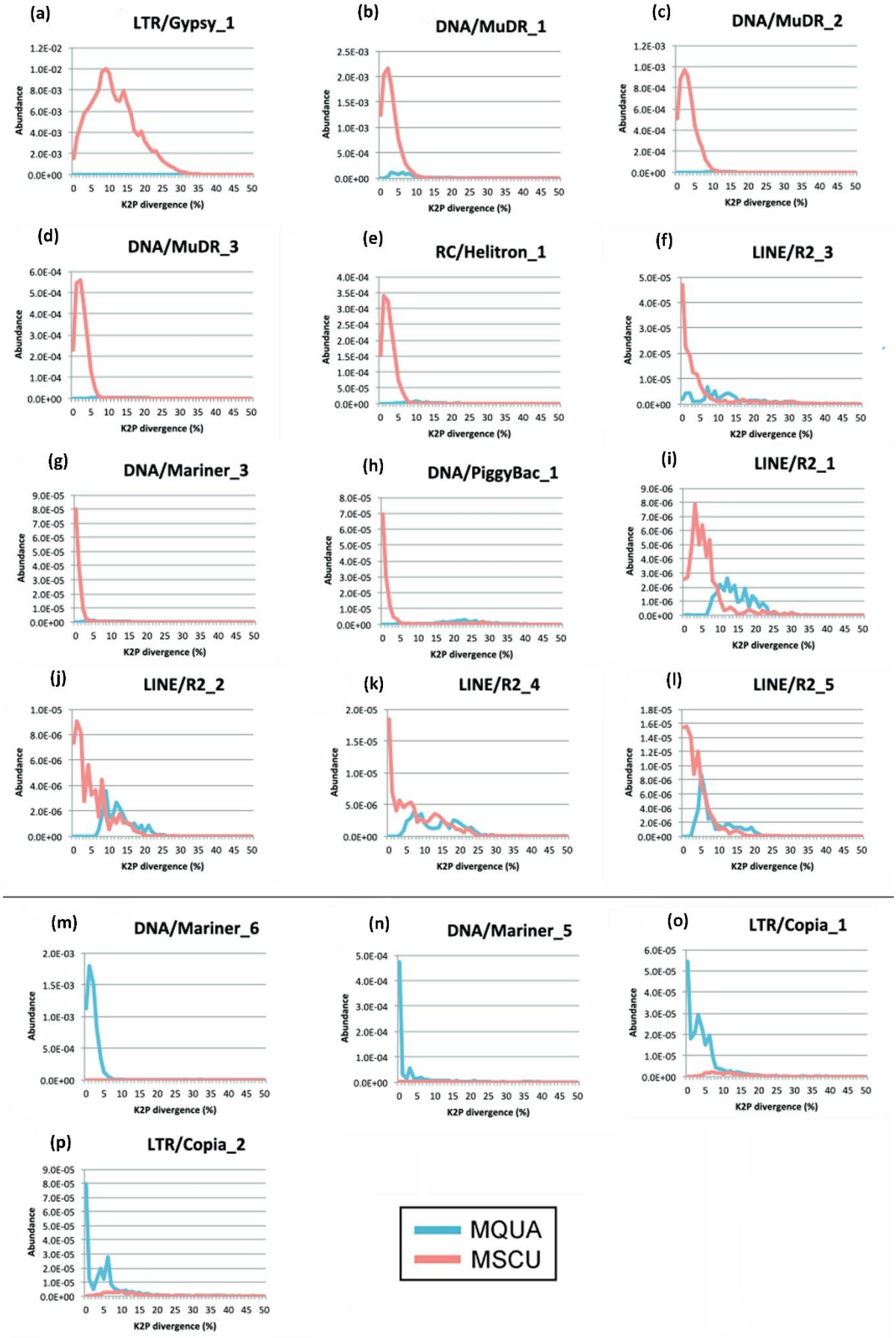
| Repetitive element | FISH signal | | Number of chromosome pairs labeled | | Chromosome location | |
|---|---|---|---|---|---|---|
| | MQUA | MSCU | MQUA | MSCU | MQUA | MSCU |
| **satellite DNA** | | | | | | |
| MscuSat01 | No | Yes | | 9 | | D (het) |
| MquaSat04_MscuSat03 | Yes | Yes | 2 | 9 | P | D (het) |
| MscuSat04 | No | Yes | | 1 | | T (eu) |
| MquaSat11_MscuSat05 | Yes | Yes | 1 | 1 | I | T (het) |
| MquaSat13_MscuSat06 | No | Yes | | 2 | | T (het), I (eu) |
| MquaSat02_MscutSat10 | Yes | No | 6 | | P | |
| MquaSat03_MscuSat12 | Yes | No | 1 | | I | |
| MquaSat08_MscuSat13 | Yes | No | 1 | | I | |
| MquaSat01 | Yes | No | 1 | | P | |
| **transposable elements** | | | | | | |
| LTR_Gypsy_1 | No | Yes | | 9 | | D (het) |
| DNA_MuDR_1 | No | Yes | | 9 | | D (het) |
| DNA_Mariner_6 | Yes | No | 9 | | P, A | |

**Supplementary Table S2** List of satDNA and TEs used for fluorescence *in situ* hybridization (FISH) mapping on chromosomes of MQUA and MSCU. Chromosome location: location of signals on chromosomes, P – pericentromeric, I – interstitial, D – dispersed, T – terminal, A – one chromosome arm, eu – euchromatin, het – heterochromatin.




**Supplementary Figure 1** Comparative landscapes (abundance *versus* divergence) for the satDNAs shared between the genome of both species. Repeat element divergence in the K2P

model distance to consensus is shown on the *x*-axis and the satDNA abundance on the *y*-axis. (a-c) satDNAs with more abundance of copies in the MSCU genome, and (d-h) more abundance of copies in the MQUA genome


(a)
LTR/Gypsy_1
(b)
DNA/MuDR_1
(c)
DNA/MuDR_2
(d)
DNA/MuDR_3
(e)
RC/Helitron_1
(f)
LINE/R2_3
(g)
DNA/Mariner_3
(h)
DNA/PiggyBac_1
(i)
LINE/R2_1
(j)
LINE/R2_2
(k)
LINE/R2_4
(l)
LINE/R2_5
(m)
DNA/Mariner_6
(n)
DNA/Mariner_5
(o)
LTR/Copia_1
(p)
LTR/Copia_2
Abundance
K2P divergence (%)
MQUA
MSCU

**Supplementary Figure 2** Comparative landscapes (abundance *versus* divergence) for the TEs. Repeat element divergence in the K2P model distance to consensus is shown on the *x*-axis and the satDNA abundance on the *y*-axis. (a) TE exclusively observed in the genome of MSCU; (b-n) TEs shared between both species with more abundance in the MSCU genome; and (m-p) more abundance in the MQUA genome

## CAPÍTULO II

### DNA satellite amplification drives the differentiation of *c*-heterochromatin into *Melipona* (Apidae: Meliponini).

**DNA satellite amplification drives the differentiation of *c*-heterochromatin into *Melipona* (Apidae: Meliponini).**

Jaqueline Amorim Pereira[1]; Diogo Milani[2]; Diogo C. Cabral-de-Mello[2,3]; Denilce Meneses Lopes[1]

[1] Laboratório de Citogenética de Insetos, Departamento de Biologia Geral, Universidade Federal de Viçosa, Campus Viçosa, Av. P.H. Rolfs s/n, CEP 36570-900, Viçosa, Minas Gerais, Brazil.

[2] Departamento de Biologia Geral e Aplicada, UNESP – Universidade Estadual Paulista, Instituto de Biociências/IB, Rio Claro, São Paulo, Brazil.

[3] Department of Experimental Biology, Genetics Area, University of Jaén, Paraje las Lagunillas s/n, 23071 Jaén, Spain

* Author to whom correspondence should be addressed: Denilce Meneses Lopes,

Departamento de Biologia Geral, Universidade Federal de Viçosa, Av. P. H. Rolfs, s/n,

Centro, Viçosa, 36570-900, Minas Gerais, Brazil. E-mail: denilce.lopes@ufv.br

**Abstract**

Satellite DNA (satDNA) is one of the main constituents of *c*-heterochromatin (constitutive heterochromatin). The dynamics of these sequences relate to changes in *c*-heterochromatin over time, as well as genome and species evolution. In *Melipona* bees, *c*-heterochromatin has been investigated due to the large variation that this region can occupy in chromosomes of some of these species. However, this characteristic does not extend to all representatives of the genus, which makes this an unusual case in relation to the distribution of *c*-heterochromatin. In order to investigate the composition and processes that led to the differentiation of *c*-heterochromatin in *Melipona,* we combined low-pass genome sequencing and chromosomal analysis for satellite DNA analysis in representatives of four subgenera of this group: *Michmelia*, *Melikerria*, *Melipona* e *Eomelipona*. We identified a total of 11 new satDNA families. Most families are shared, but the abundance varies widely between species of the four subgenera. In *Michmelia* and *Melikerria* the high proportion of *c*-heterochromatin was related to the high abundance of satDNA in the genome. MeliponaSat01-395 was highly abundant in *Michmelia*, with 22.64% and 24.03% in *Melipona* (*Michmelia*) *scutellaris* and *Melipona* (*Michmelia*) *mondury*, respectively. On the other hand, in *Melikerria* the MeliponaSat02-196 with 20.80% in *Melipona* (*Melikerria*) *interrupta* and 5.35% in *Melipona* (*Melikerria*) *fasciculata*. In species with low amounts of *c*-heterochromatin, the proportion of all satDNA families was 1.24% in *Melipona* (*Eomelipona*) *bicolor* and 3.05% in *Melipona* (*Melipona*) *quadrifasciata*. Thus, we showed that the amplification of *c*-heterochromatin occurred independently between *Michmelia* and *Melikerria* and this process was due to the massive amplification of specific satDNA families. Furthermore, c-heterochromatin is also evolving differently between the *Melipona* and *Eomelipona* subgenera.

**Key-words:** bees, RepeatExplorer, genome, chromosome.

## Introduction

Since the term heterochromatin was coined for the first time (HEITZ, 1928) this region has become the target of studies aimed at understanding it. Heterochromatin can be categorized into facultative heterochromatin and constitutive heterochromatin. Facultative heterochromatin harbors genes that are kept silent (ALLSHIRE; MADHANI, 2018; LIU; ALI; ZHOU, 2020). On the other hand, constitutive heterochromatin (*c*-heterochromatin) is gene-poor, highly condensed and has low transcriptional activity (ALLSHIRE; MADHANI, 2018; LIU; ALI; ZHOU, 2020). In addition, *c*-heterochromatin represents an important structural component and contributes to the maintenance of genome integrity (LIU; ALI; ZHOU, 2020). *C*-heterochromatin is composed of repetitive DNA sequences, especially satellite DNA (satDNA) (GARRIDO-RAMOS, 2017; PLOHL; MEŠTROVIĆ; MRAVINAC, 2012).

SatDNA are noncoding nuclear repetitive sequences tandemly organized, which normally constitute large portions of the eukaryotic genome (GARRIDO-RAMOS, 2017; PLOHL; MEŠTROVIĆ; MRAVINAC, 2012). Usually, satDNA exhibits a concerted pattern of evolution, obtained by homogenization and fixation of intraspecific sequences (DOVER, 1982, 1986; GARRIDO-RAMOS, 2017). This implies more homogeneous variants within the species than in comparison between two species. The form of evolution and absence of functional selective constraints contribute to variations in satDNA sequences to occur quickly, so that quantitative and qualitative changes can be easily noticed between closely related species (PALACIOS-GIMENEZ et al., 2020; RUIZ-RUANO et al., 2016). Genome size and architecture can be impacted by contraction/amplification events of different satDNA families (LIMA; RUIZ-RUANO, 2022; FERREE; PRASAD, 2012). Thus, repetitive DNA sequences can drive changes in *c*-heterochromatin (FERREE; PRASAD, 2012; GARRIDO-RAMOS, 2017). Therefore, evaluating their composition and behavior over time can reveal valuable information about the modification of the species' genome.

*Melipona* corresponds to a genus of stingless bees belonging to the Meliponi tribe. In *Melipona*, the distribution of *c*-heterochromatin in the chromosomes is different from the pattern found in other representatives of this tribe. In most Meliponini, *c*-heterochromatin is restricted to one of the chromosomal arms (BARTH et al., 2011; SANTOS et al., 2018; GODOY; FERREIRA; LOPES, 2013; ROCHA; POMPOLO; CAMPOS, 2003). On the other hand, in *Melipona*, some species present *c*-heterochromatin along the entire length of the chromosomes, corresponding to more than 50% of the karyotype, while in others, this region is present only in the centromeric region and short arm of the chromosomes and corresponds to

less of 50% of the karyotype (LOPES et al., 2008, 2011; ROCHA et al., 2002; ROCHA; POMPOLO, 1998; ROCHA; POMPOLO; CAMPOS, 2003; TRAVENZOLI et al., 2019a). This variation in the distribution and amount of *c*-heterochromatin observed among the species of this group is possibly due to two amplification events, which occurred after the diversification of the group approximately ~ 14–17 million years ago (My) (CUNHA; CAMPOS; LOPES, 2020b; DA CUNHA et al., 2018; PEREIRA et al., 2021a; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018; RASMUSSEN; CAMERON, 2010; TRAVENZOLI et al., 2019b). Thus, considering the four subgenera of *Melipona*: *Melipona, Eomelipona, Melikerria* and *Michmelia*, the large proportion of *c*-heterochromatin is observed in species of *Melikerria* and *Michmelia*, while in *Melipona* and *Eomelipona* only small portions of *c*-heterochromatin are observed (CAMARGO; PEDRO, 2013; CUNHA; CAMPOS; LOPES, 2020b; DA CUNHA et al., 2018; PEREIRA et al., 2021a; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018; TRAVENZOLI et al., 2019b). Therefore, possibly the first amplification event would have occurred in *Melikerria* (high *c*-heterochromatin), and the second in *Michmelia* (high *c*-heterochromatin) which is phylogenetically closer to *Eomelipona* (low *c*-heterochromatin) (CUNHA; CAMPOS; LOPES, 2020a; PEREIRA et al., 2021a; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018; TRAVENZOLI et al., 2019b).

This characteristic of *Melipona* makes this group a special case, in which studies with the *c*-heterochromatin of its representatives is essential to advance knowledge about this chromosomal region. Recently, using methods of high-throughput sequencing methods and bioinformatics tools Repeatome, we were able to characterize two species of this genus: *Melipona scutellaris* (high *c*-heterochromatin) and *Melipona quadrifasciata* (low *c*-heterochromatin) (PEREIRA et al., 2021b). A large variation in the proportion of repetitive DNA sequences that compose the genomic library of these species was identified. This large variation is related to the proportion of *c*-heterochromatin of each species. Thus, it was noted that a transposable element (LTR/Gypsy_1) and a satDNA (MscuSat01-195) highly abundant in *M. scutellaris* are responsible for the amplification of *c*-heterochromatin in this species. These data show that satellite DNA can perform a crucial role in the changes that occur in the *c*-heterochromatin of this group.

Here we will expand the work carried out previously with *Melipona* (PEREIRA et al., 2021b) in order to elucidate the sequences and mechanisms that contributed to the amplification of *c*-heterochromatin in *Melikerria* and *Michmelia.* Thus, we will analyze the Satellitome of four species of *Melipona*: *Melipona* (*Michmelia*) *mondury*, *Melipona* (*Eomelipona*) *bicolor*, *Melipona* (*Melikerria*) *faciculata*, *Melipona* (*Melikerria*) *interrupta*. With this, we identified

that the expansion of heterochromatin in the chromosomes is due to specific families of satDNA and the evolution of this region is occurring differently among the four subgenera.

## Materials and methods

### *Species sampling and chromosome preparation*

Adult and larvae *Melipona* were collected from the following locations: *Melipona (Michemelia) mondury* (MMON), *Melipona (Melipona) quadrifasciata* (MQUA) and *Melipona (Eomelipona) bicolor* (MBIC) (Viçosa, Minas Gerais, Brazil); *Melipona (Michmelia) scutellaris* (MSCU) (Encruzilhada/Bahia, Brazil); *Melipona (Melikerria) interrupta* (MINT) (Iranduba, Amazonas, Brazil); *Melipona (Melikerria) fasciculata* (MFAS) (São Luiz, Maranhão, Brazil). The adults were stored in absolute ethanol. For all these species collected, we obtain genomic DNA using the DNeasy kit (Qiagen Inc., CA, USA), according to the manufacturer's protocol. Mitotic metaphase chromosomes were obtained from brain ganglia of post-defecating larvae of the MMON, MBIC, MINT and MFAS, according to IMAI et al. (1988), and used for fluorescent *in situ* hybridization.

### *Genome and sequencing*

Genome sequencing was performed for four species: MMON, MBIC, MINT, MFAS, using the Illumina Hiseq 2000 (Inc., San Diego, CA, USA) with the service of Novogene (China), obtaining 2 × 150 bp paired reads. The genomic reads from these four species are deposited in Sequence Read Archive (SRA), XXXXX. To MSCU (accession number PRJNA759201) and MQUA (accession number SRX972105), we used the satDNA library previously obtained by PEREIRA et al. (2021b), in addition the genomes of these species were also used for abundance and divergence analyses performed later.

To identify the main satDNA families that make up the MMON, MBIC, MINT, MFAS genome, we adopted the graph-based clustering analysis using the RepeatExplorer2 (NOVAK et al., 2013), available on the public platform https://repeatexplorer-elixir.cerit-sc.cz/galaxy. That way, each genome was analyzed separately using the suggestions by NOVÁK et al. (2020), with the "repeat identification in a single species" protocol, that includes the tools "read pre-processing" and "RepeatExplorer2 clustering". We identified the supposed satDNA with the use of tandem repeat analyzer (TAREAN) (NOVÁK et al., 2017). All identified satDNA families were compared against previously characterized sequences deposited in Repbase v20.10 (BAO; KOJIMA; KOHANY, 2015) and in the GenBank/ NCBI DNA databases with

the BASTn service (ALTSCHUL et al., 1990) in order to identify possible similarities with each other.

The abundance and divergence of each satDNA family that we identified and for the satDNA families previously characterized for MSCU and MQUA (PEREIRA et al., 2021b) were estimated for the six *Melipona* genomes by means of RepeatMasker (SMIT; HUBLEY; GREEN, 2017) using the Cross_match option. The divergence of the sequences was determined by the Kimura 2-parameter (K2P) using the calcDivergenceFromAlign.pl script in software RepeatMasker (SMIT; HUBLEY; GREEN, 2017). The abundance of each satDNA was estimated as a proportion of the genome, constituting the sum of mapped nucleotides belonging to it divided by the total number of nucleotides in the library. Based on abundance, satDNA families were named in order of increasing abundance according to RUIZ-RUANO et al, (2016). Finally, we verified the temporal amplification patterns of satDNA families, for that we generated a repetitive landscape for each genome relating sequence divergence versus abundance.

To identify possible homologies between the satDNA families, we followed the protocol with RepeatMasker v4.0.5 using a cross-match search engine. With this it was possible to classify as superfamilies, families, or variants, according to RUIZ-RUANO et al (2016).

***PCR amplification and chromosomal mapping of satDNAs***

We selected the most representative satDNAs in each of the species for chromosomal analyses. For one of these selected satDNAs, we used the primer previously published by PEREIRA et al. (2021b), for others we designed primers manually (Table 1). The polymerase chain reaction (PCR) was used to amplify the satDNAs in each of the species, as applied by PEREIRA et al. (2021b). Monomeric bands for each satDNA were isolated and used for PCR reamplification for next analysis.

We performed fluorescence *in situ* hybridization (FISH) following the protocol proposed by CABRAL-DE-MELLO AND MAREC (2021). The monomers of each selected satDNA were used as probe. The sequences were labeled by nick translation or by PCR with digoxigenin-11-dUTP (Roche, Mannheim, Germany) or biotin-14-dATP (Invitrogen), and for detection anti-digoxigenin-rhodamine (Roche Applied Science) and Alexa Fluor 488 streptavidin (Invitrogen) were used. The chromosomes were contrasted with 2′,4-diamidino-2-phenylindole, dihydrochloride (DAPI), and slides were mounted with VECTASHIELD (Vector, Burlingame, CA, USA). The analysis of the results and image capture was performed

using an Olympus microscope BX61 equipped with fluorescence lamp and appropriate filters and a DP70-cooled digital camera.

## Results

### SatDNA library is variable in *Melipona*

A total of 11 satDNA families were recovered for the MMON, MBIC, MINT, MFAS genomes. These new families together with the satDNA library had previously been described for MSCU and MQUA genome, totalizing 24 satDNA families for the six *Melipona* species (Table 2). For the MeliponaSat01-395 and MeliponaSat15-315 families, we identified β variants, in addition to the α variant that had been identified previously (PEREIRA et al., 2021b), in which there are differences in monomer length. For the analysis we use the β variants that were most prevalent in the genomes. The range of variation for the monomer length was 20-4454 pb, and the A + T content of consensus sequences varied between a range of 38,3% to 72,5%, for the 24 satDNA families (Table 2).

Comparison between the species revealed that five satDNA families were not shared among the genome of all species, as follows: MeliponaSat01-395, MeliponaSat02-195, MquaSat01-586, MeliponaSat06-195, MquaSat07-20, MscuSat04-115. Furthermore, a great variation in the abundance of some satDNAs can be noticed between species, which is related to the proportion of *c*-heterochromatin and the subgenus to which each species belongs. In this way, we can highlight the main results that show these variations in three parts.

Firstly, in *Michmelia*, MSCU and MMON (high *c*- heterochromatin) the abundance of satDNA was 22.97 % and 24,37 %, respectively (Table 2). The most abundant satDNA for these species, MeliponaSat01-395, corresponded to 22.64 % and 24.037 %, respectively (Table 2). Second, in *Melikerria*, the satDNA represented 23.54 % and 6.50 % of the MINT and MFAS genome (high *c*-heterochromatin), respectively. MeliponaSat02-196 was more representative in these genomes with 20.80 % (MINT) and 5.35 % (MFAS), corresponding to an abundance difference between the two species of approximately 3.8 times more (Table 2). In these species MeliponaSat03-145 was also abundant with 1.486% in MINT and 0.526% in MFAS. In the end, we noticed *Melipona s. str.*, MQUA and *Eomelipona*, MBIC, (low *c*-heterochromatin) whose satDNA fraction corresponded only 3.05 % and 1.24 %, respectively. In MQUA the most abundant satDNA was MquaSat01-586 corresponding to 0.94% of the genome, whereas in MBIC the MeliponaSat15-314 and MquaSat06-MscutSat02-145 comprised 0.246 % and 0.238% of the genome (Table 2).

The divergence (measured by Kimura-2 parameter) between satDNA families in *Michmelia* was 16,15% (MSCU) and 19,19% (MMON), in *Melikerria* was 21,86% (MINT) and 21,35% (MFAS), in *Eomelipona* was 16,84% (MBIC) and in *Melipona* was 15,86% (MQUA) (Table 2). Based on satDNA landscape plots, we identified the possible changes in relation to the abundance and divergence of each satDNA family over time. The increase in divergence is mainly dependent on point mutations, leading to different variants (old variants). On the other hand, the lower divergence is a consequence of amplification and homogenization (new variants). The MeliponaSat01-395 shows a peak with low divergence in the MSCU and MMON species, evidencing the recent process of amplification and homogenization of this family in the subgenus *Michmelia*. In the two most abundant families in MINT and MFAS of the subgenus *Melikerria*, we can observe a recent peak of amplification for MeliponaSat2-196 and a smaller peak with greater divergence for MeliponaSat03-145 (Figure 1). For the MquaSat06-MscuSat02-145 and MeliponaSat15-314 families, we noticed an amplification peak with greater divergence and another less divergent, respectively, in MBIC (Figure 1).

The comparison between the 24 satDNA families revealed similarity between some of them. In this way, we can include two new families, MeliponaSat03-145 and MeliponaSat24-145, in the superfamily (SF1) that had been previously defined (PEREIRA et al. 2021b). Thus, the superfamily (SF1) now has eight satDNA families: MeliponaSat03-145, MquaSat03_MscuSat12-146, MquaSat06_MscuSat02-145, MquaSat11_MscuSat05-147, MquaSat10_MscuSat08-145, MquaSat12_MscuSat07-145, MquaSat13_MscuSat06-146, MeliponaSat24-145. The monomer size between these families showed small variations from 145pb to 147pb. A second superfamily was also defined (SF2), composed of five satDNA families: MeliponaSat02-196, MeliponaSat06-195, MquaSat04_MscuSat03-195, MquaSat02_MscutSat10-196, MeliponaSat14-196. In SF2 we also observed little variation between the size of the monomers of each satDNA family (195pb-196pb). Interestingly, we noted that the most abundant families of MINT and MFAS are components of different superfamilies, with MeliponaSat02-196 being a member of SF2 and MeliponaSat03-145 of SF1 (Table 2).

***Chromosomal mapping reveals that more abundant satDNA are the main constituents of c-heterochromatin***

We mapped the satDNA families that were most abundant in the species, in order to identify the distribution of these sequences in the chromosomes. The families with the highest abundance and that could be related to the amplification of *c*-heterochromatin, MeliponaSat01-

395 (β variant), MscuSat01-195 (α variant) and MeliponaSat02-196, were analyzed in MMON, MINT, MFAS and MBIC. Already MeliponaSat03-145 were analyzed only in MINT and MFAS, and MquaSat06_MscuSat02-145 and MeliponaSat15-315 in MBIC. In some species we did not identify signals in the chromosomes by FISH, possibly this is related to absence or the low proportion of these sequences in the genome, which can generate small signals below the detection of FISH.

For MeliponaSat01-395 we performed analyzes with the α and β variants, however we were only successful with the α variant in which it exhibited hybridization signals only in MMON, in which this family is more representative, like what was observed for MSCU (see data published in PEREIRA et al., 2021b). MscuSat01-195 was in almost the entire length of the *c*-heterochromatin of MMON chromosomes, in some chromosomes we can see more evident blocks and in other signals more scattered (Figure 2 E).

The MeliponaSat02-196 and MeliponaSat03-145 families were successfully mapped only in MINT and MFAS, in these analyzes we performed two-color FISH. In both MINT and MFAS, MeliponaSat02-196 showed to localize in most of the *c*-heterochromatin of the chromosomes of these species (Figure 2 A,C). Similarly, we also noticed MeliponaSat03-145 localizing part of the *c*-heterochromatin in these species (Figure 2 B, D). Two-color FISH shows that in some chromosomes there is a co-localization of MeliponaSat02-196 and MeliponaSat03, while in others there was a predominance of one of the satDNAs. It is possible to notice that MeliponaSat02-196 signals are more intense and occupy a greater extent of *c*-heterochromatin than MeliponaSat03-145, which is consistent with the difference in abundance between these families.

In MBIC the two mapped satDNA revealed signals in the chromosomes of this species, being MquaSat06_MscuSat02-145 located in the pericentromeric region in two chromosome pairs (Figure 2 F). It is possible to observe differences in signal strength between the two chromosome pairs. MeliponaSat15-314 was also identified in the pericentromeric region however in only one chromosome pair (Figure 2 G).

**Discussion**

In this study, we analyzed the satellite of six species of *Melipona* which provided clear evidence on the mechanisms behind the differentiation of *c*-heterochromatin in this group. Here we show that the composition of *c*-heterochromatin is divergent among the four subgenera of *Melipona*, and this is due to the evolution of the satDNA library present in the ancestor of this group. We note that during about ~14–17 My (RASMUSSEN; CAMERON, 2010), different

satDNA families underwent amplification, which resulted in high variation in satDNA abundances between subgenera, stemming from different amounts of *c*-heterochromatin.

Quantitative changes in *c*-heterochromatin distribution led to genome size expansion and may be related to the satDNA proportion (CRADDOCK; GALL; JONAS, 2016). PEREIRA et al (2021b) clearly demonstrated that the amplification of *c*-heterochromatin in MSCU was driven by the addition of satDNA in the genome of this species, which also contributes to genomic size increase. Our results reveal that this relationship also extends to another species of *Michmelia*, as well as to species of high *c*-heterochromatin content of *Melikerria*. Although most satDNA families are shared between *Michmelia* and *Melikerria* species, we identified that the composition of *c*-heterochromatin is different due to the variable abundance of determined families. Thus, these results are consistent with the hypothesis of *c*-heterochromatin independent amplification (CUNHA; CAMPOS; LOPES, 2020b; DA CUNHA et al., 2018; PEREIRA et al., 2021a; PICCOLI; BARDELLA; CABRAL-DE-MELLO, 2018; TRAVENZOLI et al., 2019b) and show that quantitative changes were the main responsible for the amplification of this region in the subgenera *Michmelia* and *Melikerria*.

In *Michmelia* the satDNA family MeliponaSat01-395 is highly abundant in MSCU and MMON, presenting as one of the main components of *c*-heterochromatin. This finding reinforces previous data showing that among *Michmelia* species the composition of *c*-heterochromatin is similar (CUNHA; CAMPOS; LOPES, 2020b; PEREIRA et al., 2021a). Despite this, MeliponaSat01-395 is not exclusive to *Michmelia*, being found in low abundance in the genome of *Melikerria* and *Eomelipona* species, but absent in MQUA (*Melipona*), in which probably was degenerated and lost over time in this species, or even in a closest ancestor of the subgenus it belongs to. This hypothesis is plausible, as the subgenera *Melikerria* and *Melipona* share a common ancestor (RAMÍREZ et al., 2010). So that the accumulation of mutations over time leads to sequence degeneration, on the other hand, amplification implies the massive duplication of satDNA arrays throughout the genome (CAMACHO et al., 2022). In this sense, we can assume that MeliponaSat01-395 was part of the satDNA library present in the ancestral species of the genus *Melipona*. The MeliponaSat01-395 family experienced massive amplification even in the ancestor of the *Michmelia* species, at about 5 My (RASMUSSEN; CAMERON, 2010), causing the spread of satDNA arrays and the homogenization of the family. This can be shown by the low divergence for this family. In addition, this satDNA may have undergone other modifications, which led to the emergence of variants of different sizes. The increase in tandem structure length and complexity appears to be an evolutionary trend for satDNA, so mechanisms such as unequal crossover and replication

slippage might be involved in this process (GARRIDO-RAMOS, 2017; RUIZ-RUANO et al., 2018).

In *Melikerria*, the composition of *c*-heterochromatin is similar between the two species analyzed (MINT and MFAS), with a predominant satDNA family, MeliponaSat02-196. This family is part of a superfamilies, which indicates that this satellite has homologies with other satDNA families. One of the main factors responsible for the increase in divergence between the set of repeats in a satDNA family are mutations (PLOHL; MEŠTROVIĆ; MRAVINAC, 2012). In fact, in this superfamily the monomer size did not vary greatly, indicating that mutations over time may have led to the formation of new families from an ancestral family. As in *Michmelia*, MeliponaSat02-196 was probably present in the ancestor of the genus *Melipona*, as it is present in all subgenera in low proportion, and was highly amplified in *Melikerria*.

In MBIC and MQUA we observed a lower abundance of satDNA when compared to species with high content. Despite the similarity in the proportion of *c*-heterochromatin between the *Melipona* and *Eomelipona* subgenera, their composition does not show similar. The divergence in *c*-heterochromatin composition between these two species is due to the amplification/degeneration of specific families over time. This indicates that although there was no amplification of *c*-heterochromatin in these subgenera, the satDNA library experienced changes that are leading to a modification in the composition of this region.

Taking these data together, we can draw some conclusions about the evolution of *c*-heterochromatin in *Melipona* genus. First, we noted that interspecific differences were predominantly quantitative, with most satDNA families conserved throughout *Melipona* evolution. Thus, our results relate to the patterns proposed by the “library hypothesis”, which indicates that related species share a common collection of satDNA families, each of which can be independently amplified in each species (FRY; SALSER, 1977). Thus, mechanisms such as unequal crossover, replication of extrachromosomal circles of tandem repeats by the rolling-circle mechanism and reinsertion of replicated arrays, transposition and mutations can drive changes in tandem arrays, leading to amplification or degeneration (CAMACHO et al., 2022; SATOVIĆ et al., 2016; WALSH, 1987).

Next we observed that each subgenus presents particularities in relation to its satDNA library reflecting changes in *c*-heterochromatin composition. In *Melipona* it is possible that these changes are related to the diversification of the subgenus. Some studies show that repetitive sequence divergence located in *c*-heterochromatin can be directly linked to reproductive isolation between related species (FERREE; PRASAD, 2012; JAGANNATHAN;

YAMASHITA, 2021). This is because variations in these sequences can alter chromosomal dynamics and chromatin structure. In *Drosophila*, satDNA has been shown to be related to chromocenter formation, which involves the organization of heterologous chromosomes within the nucleus. Thus, the divergence of satDNA between phylogenetically close species of *Drosophila* can compromise this organization resulting in hybrid incompatibility and reproductive isolation (JAGANNATHAN; YAMASHITA, 2021).

In the end, we noted two major global bursts of satDNA amplification that led to the increase in *c*-heterochromatin in *Michmelia* and *Melikerria*. SatDNA is highly dynamic, so rapid changes in the number of arrays in the genome are predictive for these sequences (GARRIDO-RAMOS, 2017; PLOHL et al., 2008; PLOHL; MEŠTROVIĆ; MRAVINAC, 2012). Furthermore, it has been noted that massive amplification of certain satDNA appears to be a trend in Meliponini (PEREIRA et al., 2021; PEREIRA; SALOMÃO; LOPES, 2020). However, what would have contributed to the occurrence of these events only in these two subgenera (*Michmelia* and *Melikerria)*, or being a recent process that could be extended to the other subgenera (*Melipona* and *Eomelipona*) still awaits further investigation. Some authors suggest that the most abundant satDNA in the genome may be involved in centromeric function (MELTERS et al., 2013; PLOHL; MEŠTROVIĆ; MRAVINAC, 2014). Moreover, the sequences that make up this region can change rapidly and experience amplifications, which contribute to the expansion of centromeres (HUANG et al., 2016; PLOHL; MEŠTROVIĆ; MRAVINAC, 2014). In Hymenoptera, such events have been observed in ants, and show that centromere elongation can be advantageous in these species (HUANG et al., 2016; TEIXEIRA et al., 2022). In *Melipona* species with high c-heterochromatin content, the high compaction of this region makes it impossible to identify the location of the centromeres, so it is not possible to establish any relationship between the amplification of satDNA and the expansion of centromeres. Studies involving centromeric proteins in this species will throw much light on this subject.

## References

ALLSHIRE, R. C.; MADHANI, H. D. Ten principles of heterochromatin formation and function. **Nature Reviews Molecular Cell Biology**, v. 19, n. 4, p. 229–244, 13 abr. 2018.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.

BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, n. 1, p. 11, 2 dez. 2015.

BARTH, A. et al. Occurrence of b chromosomes in *Tetragonisca* latreille, 1811 (hymenoptera, apidae, meliponini): A new contribution to the cytotaxonomy of the genus. **Genetics and Molecular Biology**, v. 34, n. 1, p. 77–79, 2011.

CABRAL-DE-MELLO, D. C.; MAREC, F. Universal fluorescence *in situ* hybridization (FISH) protocol for mapping repetitive DNAs in insects and other arthropods. **Molecular Genetics and Genomics**, v. 296, n. 3, p. 513–526, 2021.

CAMACHO, J. P. M. et al. Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. **BMC Biology**, v. 20, n. 1, p. 36, 2022.

CAMARGO, J.M.F.; PEDRO, S.R.M. *Meliponini* Lepeletier, 1836. Em Moure, JS, Urbano, D. & Melo, GAR (Orgs). **Catálogo de Abelhas (Hymenoptera, Apoidea) na Região Neotropical - versão online.**

CRADDOCK, E. M.; GALL, J. G.; JONAS, M. Hawaiian *Drosophila* genomes: size variation and evolutionary expansions. **Genetica**, v. 144, n. 1, p. 107–124, 20 fev. 2016.

CUNHA, M. S.; CAMPOS, L. A. O.; LOPES, D. M. Insights into the heterochromatin evolution in the genus *Melipona* (Apidae: Meliponini). **Insectes Sociaux**, v. 67, n. 3, p. 391–398, 2020a.

CUNHA, M. S.; CAMPOS, L. A. O.; LOPES, D. M. Insights into the heterochromatin evolution in the genus *Melipona* (Apidae: Meliponini). **Insectes Sociaux**, v. 67, n. 3, p. 391–398, 27 ago. 2020b.

CUNHA, M. S. et al. Comparative cytogenetics in three *Melipona* species (Hymenoptera: Apidae) with two divergent heterochromatic patterns. **Genetics and Molecular Biology**, v. 41, n. 4, p. 806–813, 2018.

DOVER, G. Molecular drive: A cohesive mode of species evolution. **Nature**, v. 299, n. 5879, p. 111–117, 1982.

DOVER, G. A. Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. **Trends in Genetics**, v. 2, p. 159–165, 1986.

FERREE, P. M.; PRASAD, S. How Can Satellite DNA Divergence Cause Reproductive Isolation? Let Us Count the Chromosomal Ways. **Genetics Research International**, v. 2012, p. 1–11, 29 jan. 2012.

FRY, K.; SALSER, W. Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. **Cell**, v. 12, n. 4, p. 1069–1084, 1977.

GARRIDO-RAMOS, M. Satellite DNA: An Evolving Topic. **Genes**, v. 8, n. 9, p. 230, 18 set. 2017.

GODOY, D. C.; FERREIRA, R. P.; LOPES, D. M. Chromosomal variation and cytogenetics of *Plebeia lucii* and *P. phrynostoma* (Hymenoptera: Apidae). **Florida Entomologist**, v. 96, n. 4, p. 1559–1566, 2013.

HEITZ, E. Das Heterochromatin der Moose. **Jahrbücher für Wissenschaftliche Botanik**, v. 69, p. 762–818, 1928.

HIROTAMI, T. I. et al. Modes of spontaneous chromosomal mutation and karyotype evolution in ants with reference to the minimum interaction hypothesis. **the japanese journal of genetics**, v. 63, n. 2, p. 159–185, 1988.

HUANG, Y.-C. et al. Evolution of long centromeres in fire ants. **BMC Evolutionary Biology**, v. 16, n. 1, p. 189, 2016.

JAGANNATHAN, M.; YAMASHITA, Y. M. Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid Incompatibility in *Drosophila*. **Molecular Biology and Evolution**, v. 38, n. 11, p. 4977–4986, 2021.

LIMA, L. G.; RUIZ-RUANO, F. J. In-Depth Satellitome Analyses of 37 *Drosophila* Species Illuminate Repetitive DNA Evolution in the *Drosophila* Genus. **Genome Biology and Evolution**, v. 14, n. 5, 2022.

LIU, J.; ALI, M.; ZHOU, Q. Establishment and evolution of heterochromatin. **Annals of the New York Academy of Sciences**, 2020.

LOPES, D. M. et al. Cytogenetic characterization of *Melipona rufiventris* Lepeletier 1836 and *Melipona mondury* Smith 1863 (Hymenoptera, Apidae) by C banding and fluorochromes staining. **Genetics and Molecular Biology**, v. 31, n. 1, p. 49–52, 2008.

LOPES, D. M. et al. Cytogenetics of three *Melipona* species (Hymenoptera, Apidae, Meliponini). **Sociobiology**, v. 58, n. 1, p. 185–194, 2011.

MELTERS, D. P. et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. **Genome Biology**, v. 14, n. 1, p. R10, 2013.

NOVAK, P. et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792–793, 2013.

NOVÁK, P. et al. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. **Nucleic Acids Research**, v. 45, n. 12, p. e111–e111, 2017.

NOVÁK, P.; NEUMANN, P.; MACAS, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. **Nature Protocols**, v. 15, n. 11, p. 3745–3776, 2020.

PALACIOS-GIMENEZ, O. M. et al. Eight Million Years of Satellite DNA Evolution in Grasshoppers of the Genus *Schistocerca* Illuminate the Ins and Outs of the Library Hypothesis. **Genome Biology and Evolution**, v. 12, n. 3, p. 88–102, 2020.

PEREIRA, J. A. et al. Molecular cytogenetics in the study of repetitive sequences helping to understand the evolution of heterochromatin in *Melipona* (Hymenoptera, Meliponini). **Genetica**, v. 149, n. 1, p. 55–62, 15 fev. 2021a.

PEREIRA, J. A. et al. The extensive amplification of heterochromatin in *Melipona* bees revealed by high throughput genomic and chromosomal analysis. **Chromosoma**, v. 130, n. 4, p. 251–262, 26 dez. 2021b.

PICCOLI, M. C. A.; BARDELLA, V. B.; CABRAL-DE-MELLO, D. C. Repetitive DNAs in *Melipona scutellaris* (Hymenoptera: Apidae: Meliponidae): chromosomal distribution and test of multiple heterochromatin amplification in the genus. **Apidologie**, v. 49, n. 4, p. 497–504, 2018.

PLOHL, M. et al. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. **Gene**, v. 409, n. 1–2, p. 72–82, fev. 2008.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA Evolution. p. 126–152, 2012.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Centromere identity from the DNA point of view. **Chromosoma**, v. 123, n. 4, p. 313–325, 25 ago. 2014.

RAMÍREZ, S. R. et al. A molecular phylogeny of the stingless bee genus *Melipona* (Hymenoptera: Apidae). **Molecular Phylogenetics and Evolution**, v. 56, n. 2, p. 519–525, ago. 2010.

RASMUSSEN, C.; CAMERON, S. A. Global stingless bee phylogeny supports ancient divergence, vicariance, and long distance dispersal. **Biological Journal of the Linnean Society**, v. 99, n. 1, p. 206–232, 2010.

ROCHA, M. P. et al. DNA characterization and karyotypic evolution in the bee genus *Melipona* (Hymenoptera, Meliponini). **Hereditas**, v. 136, n. 1, p. 19–27, abr. 2002.

ROCHA, M. P.; POMPOLO, S. DAS G. Karyotypes and heterochromatin variation (C-bands) in *Melipona* species (Hymenoptera, Apidae, Meliponinae). **Genetics and Molecular Biology**, v. 21, n. 1, p. 41–45, mar. 1998.

ROCHA, M.; POMPOLO, S.; CAMPOS, L. Citogenética da tribo Meliponini (Hymenoptera, Apidae). **UNESC, Criciúma**, p. 311–320, 2003.

RUIZ-RUANO, F. J. et al. High-throughput analysis of the satellitome illuminates satellite DNA evolution. **Scientific Reports**, v. 6, n. 1, p. 28333, 7 set. 2016.

RUIZ-RUANO, F. J. et al. High-throughput analysis of satellite DNA in the grasshopper *Pyrgomorpha conica* reveals abundance of homologous and heterologous higher-order repeats. **Chromosoma**, v. 127, n. 3, p. 323–340, 2018.

SANTOS, J. M. et al. Heterochromatin Distribution and Chromosomal Mapping of Microsatellite Repeats in the Genome of *Frieseomelitta* Stingless Bees (Hymenoptera: Apidae: Meliponini). **Florida Entomologist**, v. 101, n. 1, p. 33–39, 2018.

SATOVIĆ, E. et al. Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. **BMC Genomics**, v. 17, n. 1, p. 997, 6 dez. 2016.

SMIT, A.; HUBLEY, R.; GREEN, P. RepeatMasker Open-4.0. 2013-2015. 2017.

TEIXEIRA, G. A. et al. Multiple heterochromatin diversification events in the genome of fungus-farming ants: insights from repetitive sequences. **Chromosoma**, v. 131, n. 1–2, p. 59–75, 2022.

TRAVENZOLI, N. M. et al. Karyotypic description and repetitive DNA chromosome mapping of *Melipona interrupta* latreille, 1811 (Hymenoptera: Meliponini). **Caryologia**, v. 72, n. 2, p. 91–95, 2019a.

TRAVENZOLI, N. M. et al. Cytogenetic Analysis and Chromosomal Mapping of Repetitive DNA in *Melipona* Species (Hymenoptera, Meliponini). **Cytogenetic and Genome Research**, v. 158, n. 4, p. 213–224, 2019b.

WALSH, J. B. Persistence of Tandem Arrays: Implications for Satellite and Simple-Sequence DNAs. **Genetics**, v. 115, n. 3, p. 553–567, 1 mar. 1987.

| Sequence names | Forward primers | Reverse primers |
|---|---|---|
| **Satellite DNA** | | |
| MeliponaSat01-395 β | ACAGCATGCCGTGCAATTGT | AGAACATCAAATGTCCGACGA |
| MscuSat01-195 α | TCATTTGATGTTCTACAGCATG | CCGAGGAACGAGAAGAATCT |
| MeliponaSat02-196 | TTCGATGATAGTTTCAAATGCG | GTCCTCAATTGCTAGCGATAT |
| MeliponaSat03-145 | GAGAACTGCTTGTCGGGCT | CTGTTACAGGCGTCTTTCCT |
| MquaSat06_MscuSat02-145 | ATCCAGACGATGTACATCTCT | CTCGAGATATTCCCGATTGAA |
| MeliponaSat15-314 | AGAGCATGGACTCAAAGGGA | AGGTCGGCAAGAGTGTAATG |

**Table 1**. List of primers used for PCR amplification of satDNA including the new primers and the primers used in PEREIRA et al. 2021b.

| Satellite | SF | ML | A+T (%) | *M. bicolor* K2P (%) | *M. bicolor* Abundance | *M. quadrifasciata* K2P (%) | *M. quadrifasciata* Abundance | *M. scutellaris* K2P (%) | *M. scutellaris* Abundance | *M. mondury* K2P (%) | *M. mondury* Abundance | *M. interrupta* K2P (%) | *M. interrupta* Abundance | *M. fasciculata* K2P (%) | *M. fasciculata* Abundance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MeliponaSat01-395 | | 395 | 60,3 | 7,33 | 0,00202% | ----- | ------ | 10,00 | 22,64499% | 7,30 | 24,03782% | 22,01 | 0,00004% | 6,07 | 0,00100% |
| MeliponaSat02-196 | 2 | 196 | 62,2 | 6,58 | 0,00194% | 24,61 | 0,00036% | ------ | ------ | 33,09 | 0,00001% | 6,31 | 20,80750% | 6,15 | 5,35920% |
| MeliponaSat03-145 | 1 | 145 | 55,2 | 21,07 | 0,00108% | 26,95 | 0,00077% | 24,57 | 0,00022% | 28,26 | 0,00013% | 16,72 | 1,48631% | 16,37 | 0,52681% |
| MquaSat01-586 | | 586 | 57 | 38,41 | 0,00032% | 1,65 | 0,94930% | ------ | ----- | ------ | ----- | 9,48 | 0,00001% | ------ | ------ |
| MeliponaSat05-4454 | | 4454 | 63,2 | 7,92 | 0,16209% | 15,85 | 0,03321% | 13,90 | 0,02001% | 14,58 | 0,05581% | 24,85 | 0,30139% | 24,35 | 0,19410% |
| MeliponaSat06-195 | 2 | 195 | 67,7 | ------ | ------ | 28,92 | 0,00002% | ----- | ------ | ------ | ----- | 25,39 | 0,28785% | 25,52 | 0,07564% |
| MeliponaSat07-3390 | | 3590 | 53,8 | 6,78 | 0,17592% | 15,28 | 0,03815% | 16,46 | 0,08439% | 17,31 | 0,12611% | 15,78 | 0,20122% | 16,44 | 0,09428% |
| MquaSat04-MscuSat03-195 | 2 | 195 | 65,1 | 33,61 | 0,00067% | 0 | 0,50781% | 30,30 | 0,04780% | 30,74 | 0,00030% | 27,83 | 0,00137% | 29,40 | 0,00052% |
| MquaSat02-MscuSat10-196 | 2 | 196 | 62,2 | 37,39 | 0,00001% | 3,83 | 0,55339% | 21,09 | 0,00373% | 26,67 | 0,00003% | 36,60 | 0,00059% | 36,52 | 0,00013% |
| MquaSat03-MscuSat12-146 | 1 | 146 | 59,6 | 27,30 | 0,00029% | 3,71 | 0,52111% | 22,37 | 0,00046% | 27,23 | 0,00010% | 27,12 | 0,00228% | 27,21 | 0,00111% |
| MquaSat05-MscuSat11-225 | | 225 | 41,3 | 19,03 | 0,03791% | 3,27 | 0,10950% | 21,14 | 0,00232% | 20,56 | 0,02268% | 22,70 | 0,21950% | 22,72 | 0,10704% |
| MquaSat06-MscuSat02-145 | 1 | 145 | 52,4 | 17,27 | 0,23871% | 14,38 | 0,08749% | 16,33 | 0,04904% | 18,44 | 0,02998% | 22,97 | 0,04170% | 21,16 | 0,03091% |
| MeliponaSat13-2660 | | 2660 | 62,7 | 6,02 | 0,13793% | 16,1 | 0,12580% | 17,09 | 0,01290% | 17,52 | 0,02054% | 21,45 | 0,05890% | 19,15 | 0,04529% |
| MeliponaSat14-196 | 2 | 196 | 59,7 | 8,40 | 0,13256% | 26,45 | 0,03263% | 25,45 | 0,00900% | 27,96 | 0,02222% | 30,76 | 0,10771% | 30,26 | 0,04375% |
| MeliponaSat15-314 | | 314 | 68,5 | 2,25 | 0,24615% | 17,59 | 0,00163% | 11,45 | 0,00007% | 15,55 | 0,00008% | 16,83 | 0,00004% | 16,78 | 0,00028% |
| MquaSat07-20 | | 20 | 70 | ------ | ------ | 8,6 | 0,05740% | ----- | ----- | ----- | ----- | ----- | ----- | 21,88 | 0,00005% |
| MscuSat04-115 | | 115 | 38,3 | 24,53 | 0,00163% | ------ | ------ | 5,92 | 0,04235% | 13,82 | 0,02048% | 24,18 | 0,00004% | 22,27 | 0,00000% |
| MquaSat11-MscuSat05-147 | 1 | 147 | 54,4 | 14,58 | 0,03413% | 20,13 | 0,00167% | 10,76 | 0,01769% | 13,50 | 0,00790% | 24,02 | 0,00119% | 23,41 | 0,00331% |
| MquaSat09-MscuSat09-43 | | 43 | 41,9 | 8,46 | 0,00489% | 11,58 | 0,01838% | 2,11 | 0,00687% | 7,24 | 0,00589% | 13,60 | 0,01410% | 13,77 | 0,00466% |
| MquaSat10-MscuSat08-145 | 1 | 145 | 53,8 | 17,90 | 0,02214% | 25,8 | 0,00199% | 11,08 | 0,01040% | 12,21 | 0,01171% | 24,21 | 0,00216% | 23,55 | 0,00466% |
| MeliponaSat21 | | 131 | 72,5 | 11,53 | 0,01864% | 15,07 | 0,00464% | 16,59 | 0,00064% | 15,49 | 0,00111% | 15,32 | 0,00269% | 14,27 | 0,00343% |
| MquaSat12-MscuSat07-145 | 1 | 145 | 51,7 | 18,62 | 0,01228% | 23,58 | 0,00141% | 7,88 | 0,01075% | 12,94 | 0,00228% | 29,25 | 0,00080% | 27,02 | 0,00103% |
| MquaSat13-MscuSat06-146 | 1 | 146 | 59,6 | 20,84 | 0,00263% | 22,11 | 0,00049% | 13,22 | 0,01307% | 17,11 | 0,00881% | 23,82 | 0,00084% | 25,27 | 0,00147% |
| MeliponaSat24 | 1 | 145 | 53,8 | 14,70 | 0,01412% | 23,59 | 0,00286% | 25,23 | 0,00040% | 25,49 | 0,00099% | 21,52 | 0,00328% | 21,60 | 0,00563% |
| | | | | | **1,24805%** | | **3,05000%** | | **22,97711%** | | **24,37499%** | | **23,54150%** | | **6,50429%** |

**Table 2.** Characteristics of the satDNAs of the *Melipona* species including families, superfamilies (SF), monomer length (ML), A+T sequence content (A+T%), abundance and divergence (K2P)

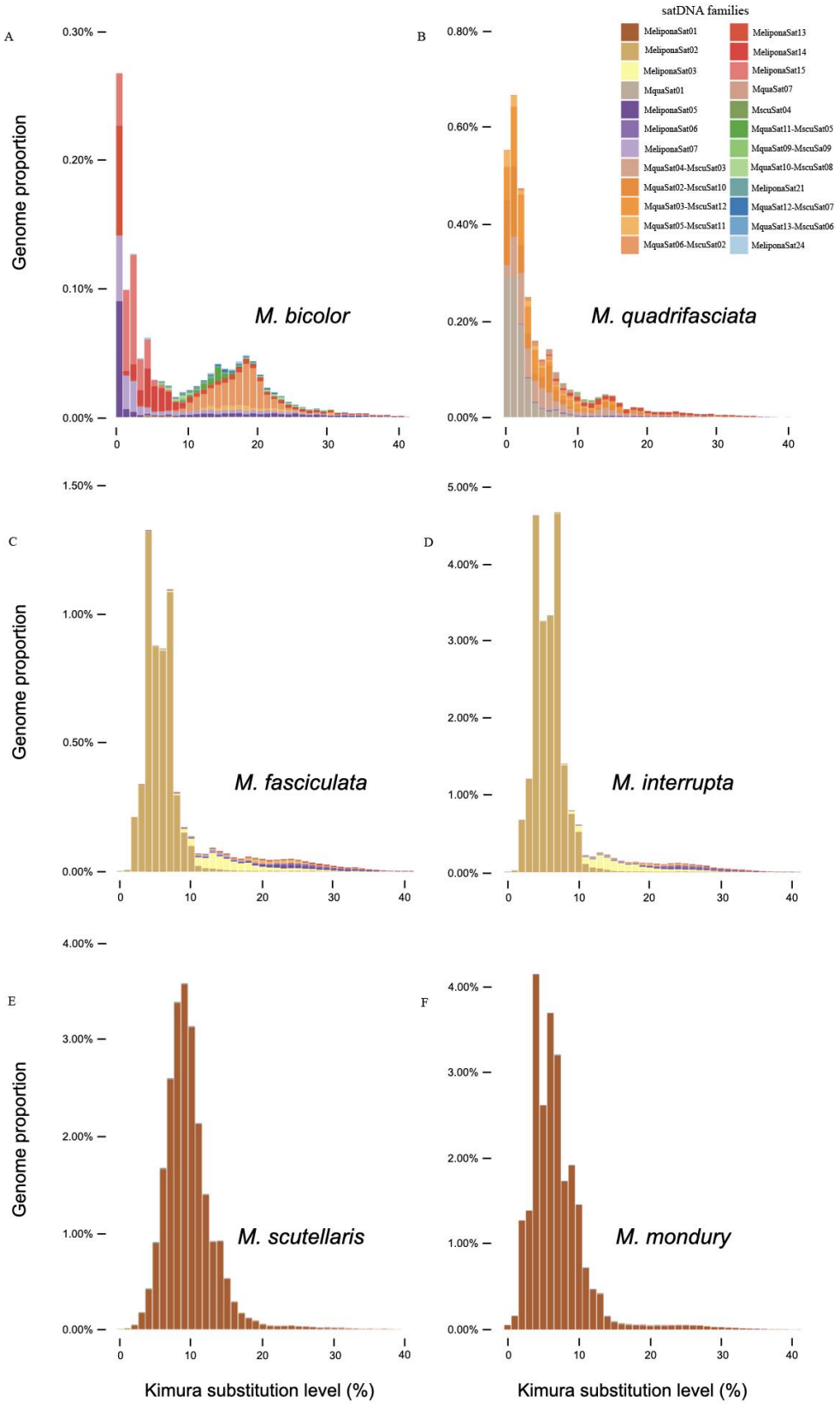


**Figure 1.** Landscape (abundance versus divergence) for the satDNAs of *Melipona*: (A) *M. bicolor;* (B); *M. quadrifasciata;* (C) *M. fasciculata*; (D) *M. interrupta*; (E) *M. scutellaris*; (F) *M. mondury.*

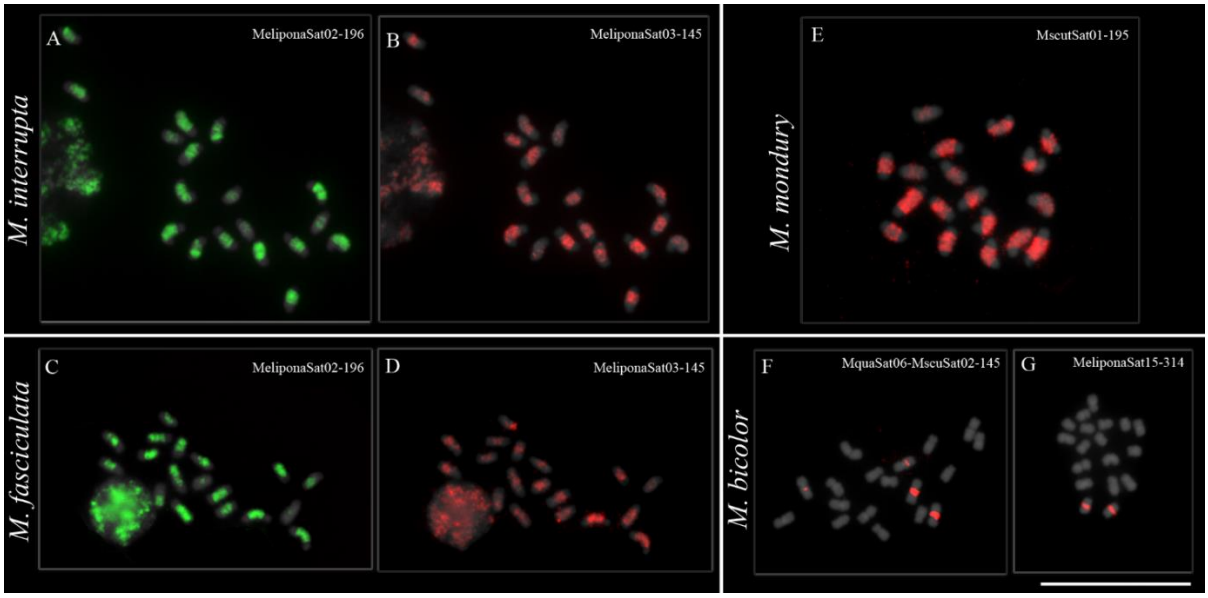



**Figure 2.** Distribution and chromosomal location of the most abundant satDNA families in *Melipona* species: *M. interrupta* (A) MeliponaSat02-196, (B) MeliponaSat03-145; *M. fasciculata* (C) MeliponaSat02-196, (D) MeliponaSat03-145; *M. mondury* (E) MscuSat01-195; *M. bicolor* (F) MquaSat06-MscuSat02-145, (G) MquaSat15-314.

## CAPÍTULO III

### The satellite DNAs populating the genome of *Trigona hyalinata* and the sharing of a highly abundant satDNA in *Trigona* genus

**The satellite DNAs populating the genome of *Trigona hyalinata* and the sharing of a highly abundant satDNA in *Trigona* genus**

Jaqueline Amorim Pereira[1]; Diogo C. Cabral-de-Mello[2,3]; Denilce Meneses Lopes[1,]

[1] Laboratório de Citogenética de Insetos, Departamento de Biologia Geral, Universidade Federal de Viçosa, Campus Viçosa, Av. P.H. Rolfs s/n, CEP 36570-900, Viçosa, Minas Gerais, Brazil.

[2] Departamento de Biologia Geral e Aplicada, UNESP – Universidade Estadual Paulista, Instituto de Biociências/IB, Rio Claro, São Paulo, Brazil.

[3] Department of Experimental Biology, Genetics Area, University of Jaén, Paraje las Lagunillas s/n, 23071 Jaén, Spain

* Author to whom correspondence should be addressed: Denilce Meneses Lopes, Departamento de Biologia Geral, Universidade Federal de Viçosa, Av. P. H. Rolfs, s/n, Centro, Viçosa, 36570-900, Minas Gerais, Brazil. E-mail: denilce.lopes@ufv.br

**Abstract**

Among Meliponini species, *c*-heterochromatin can occupy large portions of chromosomes. Although this characteristic could be useful for understanding evolutional patterns of satellite DNAs (satDNAs), few sequences have been characterized in these bees. In *Trigona*, phylogenetically represented by clades A and B, the *c*-heterochromatin is mostly located in one chromosome arm. Here we used different techniques, including restriction endonucleases and genome sequencing followed by chromosomal analysis, to identify satDNAs that may be contributing to the evolution of *c*-heterochromatin in *Trigona*. Our results revealed a highly abundant ThyaSat01-301 satDNA family, corresponding to about 13.77% of the *Trigona hyalinata* genome. Another seven satDNAs were identified, one corresponding to 2.24%, and other six 0.545% of the genome. The satDNA ThyaSat01-301 family was shown to be one of the main constituents of the *c*-heterochromatin of this species, as well as of other species belonging to clade B of *Trigona*. However, this satDNA was not observed on the chromosomes of species from clade A, demonstrating that the *c*-heterochromatin is evolving divergently between species of clade A and B, as a consequence of the evolution of repetitive DNA sequences. Finally, our data suggest molecular diversification of the karyotypes, despite a conservated macro-chromosomal structure on the genus.

**Key-words:** *c*-heterocromatin, genome, RepeatExplorer, Meliponini, bees.

## 1. Introduction

Eukaryotic genomes are enriched with multiple classes of repetitive DNAs, including satellite DNA (satDNA) and Transposable Elements (TEs) [1]. The satDNAs can represent large portions of the genomes and are characterized as tandem repeats which have a repeat unit length frequently larger than 100 bp [1,2]. Over the years, the use of different techniques to study repetitive DNA, such as $C_0$t-DNA renaturation kinetics, restriction enzymes, chromosomal analysis, and more recently sequencing technologies combined with computational approaches, has allowed to understand the structure and evolution the satDNAs in genomes [3,4]. This class of repetitive DNA is found primarily on constitutive heterochromatin (*c*-heterochromatin) domains, and are predominant on chromosomal regions, such as telomeres and pericentromeres [3,5]. SatDNA sequences are highly dynamic, and some molecular mechanisms (unequal crossing-over, transposition, rolling circle replication and re-insertion) contribute to variations in the number of copies, nucleotide base pair composition, and ultimately to their evolution [3,4]. Usually, groups of related species share families of satDNA from a common ancestor that can evolve independently within each species, according to the library hypothesis model [6].

In stingless bees (Meliponini), studies about evolution of *c*-heterochromatin and repetitive DNA sequences have been limited [7–11]. Since Kerr et al. [12] first described the karyotype of Meliponini species, the study of chromosomes was mainly limited to the use of techniques such as C-banding and DAPI/$CMA_3$ fluorochromes staining. This approach gives only superficial information about *c*-heterochromatin [13–15], but revealed variable patterns of chromosomal organization of *c*-heterochromatin in Meliponini. In most species the *c*-heterochromatin is located in one of the chromosomal arms, frequently occupying large extension of the chromosomes, and specially in *Melipona* representatives, the *c*-heterochromatin occupies almost the entire length of the chromosomes or is restricted to the pericentromeric region [15–19]. Recently, some studies using $C_0$t-DNA and restriction endonucleases to isolate repetitive DNA sequence, show that the composition of *c*-heterochromatin is divergent among some subgenera and genera of Meliponini [7–10]. Furthermore, the molecular composition of *c*-heterochromatin was presented in details for two species *Melipona quadrifasciata* and *Melipona scutellaris*, revealing that quantitative and qualitative changes in the repetitive DNA library contributed to the evolution of the highly divergent *c*-heterochromatin of the genus *Melipona* [11].

*Trigona* (Meliponini) bees has about 32 species and based on morphological, biological characteristics, and geographical distribution it is divided into 9 groups of species [20].

According to the molecular phylogeny proposed by Rasmussem and Camargo [20], *Trigona* can be divided into two major clades, A and B, with the last shared common ancestor dating from approximately 19 million years, Mya [21]. The study of the karyotypes of *Trigona* species showed that the chromosome number of 2n = 34 is conserved, except for *Trigona braueri* (as *Trigona fulviventris*) with 2n = 32 [17,22]. Similar to most species of Meliponini high amount of *c*-heterochromatin is observed in *Trigona*, occupying one of the chromosomal arms [15,17–19,22–26]. The *c*-heterochromatin in *Trigona* was characterized only by fluorochrome staining, revealing enrichment of G+C base pairs in few [22,24] or multiple chromosomes [25,26].

Since most Meliponini species have a frequent occurrence of large proportions of *c*-heterochromatin on chromosomes, putatively as result of massive amplification, studies on this group could contribute to broadening the knowledge of repetitive DNAs and their role in genome evolution. Here, our aim was to investigate repetitive DNA sequences and their relationship with the evolution of *c*-heterochromatin in *Trigona* species at a phylogenetic context. For this, we use different approaches, including restriction endonucleases and low-coverage genome sequencing followed by computational and chromosomal analysis. Our data showed a highly abundant satDNA family as the main repetitive DNA that constitute the *c*-heterochromatin of some of the species of this genus belonging to the clade B, but not to the clade A. This reveal conservation of heterochromatin in close related species, but also some degree of molecular differentiation of the chromosomes in the genus, despite conservation of macro-chromosomal structure.

## 2. Materials and methods

### *2.1 Genomic DNA isolation and digestion by restriction enzymes*

Total genomic DNA from adults of *Trigona hyalinata* was extracted according to Waldschmidt et al [27]. In order to obtain repetitive DNAs, we digested *T. hyalinata* genomic DNA with the restriction enzymes *Hin*fI, *Eco*RI, *Eco*571, *Vsp*I, *Bcn*I (Thermo Fisher Scientific). For the reactions, we used 2.5 µL DNA (100 ng/µL), 0.5 µL enzyme (10 U/µL), and 2.0 µL Buffer (10×) incubated in a water bath at 37 °C for 4 h. The restriction products were separated on a 1.2% agarose gel, and the result of DNA restriction that showed a band pattern typical for tandem repetitive sequences, was isolated and purified from the gel using the NucleoSpin® Gel and PCR kit (Neumann-Neander-Str. 6-8 · 52355 Düren · Germany) following the manufacturer's instructions.

### *2.2 Cloning, sequencing and sequence analysis*

The isolated and purified bands were modified using the A-tailing procedure and inserted in pGEM-T Easy Vector (Promega) and used for transformation of competent *Escherichia coli* cells (DH5α). We obtained a total of eight positive clones by PCR screening using the primers M13 Reverse (-48) and T7 Promoter in which the plasmids were sequenced in both directions by the Sanger sequencing method, carried out by Myleus Facility (Brazil). The nucleotide sequences obtained from the sequencing were analyzed using the software Geneious v4.8 [28]. To verify if the obtained sequences had similarities with other sequences previously described, we made a search on Repbase v20.10 [29] and GenBank/NCBI DNA databases using the BLAST tool [30].

After characterizing the repetitive sequence isolated by enzymatic digestion, we designed primers manually: Forward 5'-AAGTGATCAGAGGAGAACGAT-3' and Reverse 5'-ATATAGCAATGTGGCGGCCA-3'. PCR reactions were performed using 10×PCR Rxn Buffer, 0,2 mM of $MgCl_2$, 0,16 mM of dNTPs, 2mM of each primer, 1 U de Taq Platinum DNA Polymerase (Invitrogen) and 100 ng/µl of genomic DNA. PCR conditions included an initial denaturation at 94 ° C for 5 min and 30 cycles at 94 ° C (30 s), 55 ° C (30 s) and 72 ° C (80 s), and final extension at 72 ° C for 5 minutes. The PCR products were checked on a 1% agarose gel. The monomeric bands were isolated and used for reamplification for further protocols.

*2.3 Genome Sequencing and satellite DNA sequence analysis*

For the genome sequencing, the genomic DNA of the *T. hyalinata* was obtained using the DNeasy kit (Qiagen Inc., CA, USA), according to the manufacturer's protocol. The sequencing of the genome was performed using Illumina Hiseq 2000 (Inc., San Diego, CA, USA), obtaining paired-end reads (2 × 150 bp) using the service of Novogene (China). The genomic reads were deposited in GenBank under accession number PRJNA885409. To identify the abundant satDNAs populating the genome of *T. hyalinata*, we used RepeatExplorer2 (Galaxy Version 2.3.7), available on the public platform https://repeatexplorer-elixir.cerit-sc.cz/galaxy. These analyzes followed the protocol suggested by Novák et al. [31]. After grouping by RepeatExplorer2, satDNAs were identified using the TAREAN tool [32]. All sequences were analyzed to identify possible similarities with satDNA previously identified by comparison with sequences deposited on Repbase v20.10 [29] and in the GenBank/NCBI DNA databases with the BASTn tool [30]. The abundance and divergence of each satDNA family was estimated by means of RepeatMasker [33] using the Cross_match. The abundance of each satDNA family was estimated based on its proportion in the genome, so the sum of the mapped nucleotides belonging to one specific satDNa family was divided by the total number of

nucleotides in the library. The divergence of the satDNA family was determined by the Kimura 2-parameter (K2P) using the calcDivergenceFromAlign.pl script in software RepeatMasker [33]. Finally, we identified possible homology between satDNA families using the software Geneious v4.8 [28]. Finally, the sequences were named according to Ruiz-Ruano et al. [34], considering the abundance.

*2.4 Fluorescence in situ hybridization (FISH)*

Mitotic chromosomes were obtained according to Imail et al. [35] from brain ganglia of post defecating larvae of different species of *Trigona* from clades A and B (Table 1). FISH followed the protocol described by Pinkel et al. [36] with modifications. First, the slides were treated with RNase A (100 µg/ml) and kept at 37ºC for 1 hour. Then, a wash was performed in 2 × SSC for 5 min and the slides were incubated in a solution containing 5 µg/ml of pepsin in 0.01 N HCl for 10 min. After, the slides were washed in 2 × SSC for 5 min and dehydrated in alcoholic series at the following concentrations 50%, 70% and 100% for 2 min each. For denaturation of metaphase chromosomes, we use 70% formamide/2 × SSC for 5 min. The repetitive sequence isolated from the *T. hyalinata* genome was labeled by indirect method through PCR with digoxigenin-11-dUTP (Roche, Mannheim, Germany). For probe hybridization in the chromosomes, the slides were kept overnight at 37°C. The slides were washed in 2 × SSC and for signal detection we used anti-digoxigenin-rodamine (Roche Applied Science) for 1h at 37ºC. Finally, slides were washed in 4 × SSC/Tween and dehydrated in an Ethanol series of 50%, 70% e 100%. Chromosomes were counterstained with 4',6-Diamidino-2-Phenylindole, Dihydrochloride (Sigma). The metaphases were captured in an Olympus BX53F microscope equipped with an Olympus MX10 camera, using the CellSens imaging software.

## 3. Results

*3.1 Extensive amplification of a satDNA superfamily on the genome of* Trigona hyalinata

Through enzymatic digestion of genomic DNA with *Hin*fI allowed us to identify one repetitive DNA sequence from the genome of *T. hyalinata*, this was the only restriction enzyme that give a ladder-like pattern, which represents one of the characteristics of a putative authentic tandem repeat (Figure 1A). Two fragments of approximately 300 bp and 600 bp were isolated. The sequence analysis from cloned fragments revealed a sequence with 302 bp, percentage of A+T = 63.2%, with repetitive pattern, since the 600 bp fragment corresponds to two monomers. PCR assays for the sequence revealed a ladder pattern, characteristic for in tandem repeats (Figure 1B). As we did not identify any correspondence of these fragments with other sequences of

repetitive DNA previously characterized (TEs for example) and because they present repetitively in tandem by the genome, we classify it as a satDNA.

Through RepeatExplorer clustering analysis we identified eight satDNA families (deposited on GenBank under accession numbers OP559492 - OP559499) which corresponded together to 16.565% of the *T. hyalinata* genome considering the characterized repetitive fraction (Table 2). These satDNAs monomer sequences ranged in length from 145 bp to 542 bp, and the AT content of the sequences ranged from 55.9% to 66.1%. The homology analysis among the eight *T. hyalinata* satDNA families revealed similarity between some of them, so that it was possible to group into two superfamilies. The superfamily (SF1) being represented by ThyaSat01-301 and ThyaSat02-300, and the second superfamily (SF2) by ThyaSa04-145, ThyaSat06-145, ThyaSat07-145 and ThyaSat08-145. In SF1 we noticed difference of one base pair between the length of the monomer of consensus sequences, while in SF2 the monomers have the same length. This indicates that the divergence between them is mainly due to nucleotide substitution, sush as point mutation events (Table 2).

ThyaSat01-301 is the most abundant satDNA family corresponding to 13.774% of the genome, that is about 83% of the characterized satDNAs (Table 2, Figure 1C). Furthermore, the ThyaSat01-301 corresponds to the satDNA isolated with the restriction enzyme *Hin*fI (Figure 1D). Although much less abundant than ThyaSat01-301, another satDNA family also showed relevant abundance in *T. hyalinata* genome, the ThyaSat02-300, corresponding to 2.246% (14% of the satDNA content) of the genome. These two satDNA families are member of SF1 and together represent 97% of the satDNA content of the species genome (Table 2, Figure 1C). The other six satDNAs were very low abundant, representing together only 0.545% of the genome. The average of nucleotide K2P divergence of the eight satDNA families was 13.39, ranging from 5.5 to 20.59 (Table 2).

The analysis of landscapes (abundance *versus* divergence) evidences more recent amplification of the members of SF1 in comparison to other satDNAs, as it has high abundance with lower divergence (K2P) (Figure 2A,B). Low K2P was also observed for ThyaSat05-291, but with no amplification pattern (Figure 2C). For the ThyaSat03-542 variable K2P was noticed (Figure 2D). Finally, for the members of SF2 the distinct peaks of abundance in variable K2P for ThyaSat04-145 and the other members, i.e. ThyaSat06-145, ThyaSat07-145, ThyaSat08-145 suggested that these families are under distinct patterns of homogenization/amplification (Figure 2E).

*3.2 A phylogenetically conserved satDNA on heterochromatin of* Trigona *species*

We analyzed the chromosomal distribution of the main satDNA isolated from the *T. hyalinata* genome (ThyaSat01-301) in other six species of *Trigona* belonging to the clades A and B, and we observed positive FISH signals on four of them, all these from clade B (Figure 3,4). According to previously published data, the *c*-heterochromatin for these species is predominant in one of the chromosomal arms [17,26]. The *c*-heterochromatin is easily observed by DAPI staining as more brightness regions after formamide FISH treatment (Figure 3A,D,G,J,M,P). Thus, we observed that the ThyaSat01-301 satDNA is exclusively enriched on the *c*-heterochromatin region of the chromosomes (Figure 3).

In *T. hyalinata*, the signals formed large blocks occupying all the *c*-heterochromatin of almost all chromosomes, except the *c*-heterochromatin of one chromosome pair (Figure 3C). Similarly, in *T. spinipes* (Figure 3F), *T.* aff. *fuscipennis* (Figure 3I) and *T. recursa* (Figure 3L) the signals also coincided primarily with *c*-heterochromatic regions. However, we observed differences in the intensity of the signals among these species, with scattered signals for *T.* aff. *fuscipennis* (Figure 3I) and *T. recursa* (Figure 3L). In *T. spinipes* (Figure 3F) and *T. recursa* (Figure 3L) two and four chromosomes lacked signals, respectively. All of these species with positive FISH signal belong to clade B and no signal has been identified for *T. williana* (Figure 3N) and *T. pallens* (Figure 3Q), which belong to clade A.

## 4. Discussion

One of the most striking characteristics of bees from the Meliponini tribe is the high abundance of *c*-heterochromatin [16,17,19], as noticed on the *Trigona* species selected for this study. Most species share the same chromosome number and heterochromatin distribution on one arm of most chromosomes is the commonly observed pattern. Recently, methods for repetitive DNAs study, as the used here, have been allowing to reveal differences on the karyotypes of Meliponini representatives, mainly on the *c*-heterochromatin. Even among congeneric species, like *Trigona*, relevant differences have been noticed, suggesting that the repetitive DNAs in this group have undergone intense differentiation processes and can be an important tool for evolutionary genomic and karyotypic studies [11].

Here we were able to identify and characterize eight satDNAs in the genome of *T. hyalinata.* The most abundant, ThyaSat01-301, is A+T enriched, similarly to observed on most satDNAs characterized on insects so far [37]. The ThyaSat01-301 form large blocks virtually covering the heterochromatin of most chromosomes of the species and is shared with among some other *Trigona* representatives from clade B, suggesting that this is the main constituting

of the *c*-heterochromatin in these species. This idea is supported by the high abundance of ThyaSat01-301 on the genome of *T. hyalinata*, corresponding to 13.774% of its genome.

The prevalence of one specific family of repetitive DNA occupying a large part of *c*-heterochromatin also seems to be recurrent for some genera of Meliponini, as already reported for *Tetragonisca* [9] and *Melipona* [10,11]. In some insects the presence of one or more highly abundant families in the genome can also be observed in *Triatoma infestans* [38]*, Hippodamia variegata* [39] and *Holhymenia histrio* [40]. However, these studies are based on the analysis of isolated species, so it is not possible to propose whether this characteristic also extends to other species belonging to the same subgenus or genus. Furthermore, for most insects the most abundant satDNA may not represent more than 1% of the genome [34,41–44]. With this, we can infer that the massive amplification of a certain family seems to be a trend in Meliponini and occurs independently in each genus or subgenus. For ThyaSat01-301 and the other member of SF1, i.e. ThyaSat02-300 the high abundance and low divergence suggest that this amplification was more recent in comparison to other satDNAs families in *T. hyalinata* genome.

According to the phylogeny proposed by Rasmussen et al. [20], *Trigona* species are divided into two clades A and B. Our FISH results revealed signals ThySat01-301 exclusively on chromosomes of representatives of Clade B species, *T. hyalinata*, *T. spinipes*, *T. recursa* and *T.* aff. *fuscipennis*. The sharing of the same satDNA family between these species suggests an ancient origin that was possibly present in the common ancestor of this clade, approximately 11 Mya [21]. Although, we can not completely rule out that this satDNA is also present in species of clade A (*T. williana* e *T. pallens*) in very low proportion not detected by FISH analysis. In any case, our data clearly evidence that the *c*-heterochromatin is evolving in a divergent manner between species of Clade A and B, achieving different molecular composition. This suggest that the amplification of heterochromatin in the genus, that has similar distribution on species of clade A and B, occupying one entire arm of most chromosomes, occurred independently, or after ancestral amplification it was completely differentiated by mutational processes. This is a similar case to observed in the genus *Melipona*, in which the analysis of repetitive DNA revealed that the composition of *c*-heterochromatin is divergent between subgenera, suggesting more than one round of amplification of *c*-heterochromatin [7,8,10].

Among insects, analyzes using high-throughput sequencing data and molecular cytogenetics clearly illustrate the evolutionary dynamics of satDNA over time [45–47]. Most satDNAs are highly dynamic, and follow the principles of the Library Hypothesis that predicts that sister species shares a set of satDNA from the common ancestor and certain members of

the library may be differentially amplified, appearing as the main component of satDNAs, while others occur at low abundance, leading to distinct profiles of satDNA between species [6,45]. In *Trigona* the difference in *c*-heterochromatin composition of between species of clade A and B, illustrate very well the predictions of the evolution of satDNA library that differentiated, approximately 19 Mya [21]. Also, in accordance with phylogenetic history of the *Trigona* genus, our FISH results reveal variations in the intensity of hybridization signals between species of clade B, which may be related to quantitative changes in ThySat01-301 over time, as predicted in satDNA library hypothesis. In *T. hyalinata* and *T. spinipes*, ThySat01-301 shows a similar distribution on chromosomes, covering almost all *c*-heterochromatic extensions. These two species are closely related and belong to the *"spinipes"* group, and they may not have had sufficient time for the accumulation of large variations between the variants of ThySat-301. On the other hand, in *T. recursa* and *T.* aff. *fuscipennis*, the ThySat01-301 signals were more scattered along the heterochromatin of most chromosomes. These two species are more distantly related to *T. hyalinata* and belong to different groups, *"recursa"* and *"fuscipennis"* [20]. Thus, other variants of the ThySat01-301 family may be present in the genome of these species in comparison to variants from *T. hyalinata*, as well as there may have been both amplifications and contractions for this satDNA family over time between these species' groups.

In *Trigona*, both intraspecific and interspecific variations were demonstrated in the number and location of 18 rDNA sites [25,26,48]. Interspecific variations occur mainly between species of Clade A and B. Thus, in Clade A the ribosomal sites are located in the interstitial region of the chromosomes, and in Clade B they are mainly placed on the terminal region of the chromosomes [48]. According to these previous works, the genomic dispersion of rDNA sites can be facilitated by the interaction with repetitive sequences located in *c*-heterochromatin. Furthermore, the molecular mechanisms acting on spreading of satDNA monomers, as transposition and replication of extrachromosomal circles may be related to the dispersion of rDNA sites [3,26]. Although less variable, chromosomal distribution of other repetitive DNAs are also variable on *Trigona* species, like the microsatellite $(GA)_{15}$ [25,26]. Thus, these cytogenetic data along with the *c*-heterochromatin divergence between Clade A and B observed here highlight the genomic differentiation on *Trigona* species, despite the conservation of the diploid number (2n=34).

**Author Contributions:**

J.A.P and D.M.L. conceived the study and designed the experiments. J.A.P, D.M.L and D.C.C.M performed the experiments, analyzed the data and wrote the manuscript. All authors approved the final manuscript.

**Funding:** The authors would like to thank the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado de São Paulo.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We acknowledge the anonymous reviewers for the suggestions and manuscript improvement.

**Conflicts of Interest:** The authors declare no conflict of interest.

## 5. References

1. Charlesworth, B.; Sniegowski, P.; Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **1994**, *371*, 215–220, doi:10.1038/371215a0.

2. López-Flores, I.; Garrido-Ramos, M.A. The Repetitive DNA Content of Eukaryotic Genomes. *Genome Dyn.* **2012**, *7*, doi:10.1159/000337118.

3. Garrido-Ramos, M. Satellite DNA: An Evolving Topic. *Genes (Basel).* **2017**, *8*, 230, doi:10.3390/genes8090230.

4. Lower, S.S.; McGurk, M.P.; Clark, A.G.; Barbash, D.A. Satellite DNA evolution: old ideas, new approaches. *Curr. Opin. Genet. Dev.* **2018**, *49*, 70–78, doi:10.1016/j.gde.2018.03.003.

5. Plohl, M.; Meštrović, N.; Mravinac, B. Satellite DNA evolution. In *Genome Dynamics*; 2012; Vol. 7, pp. 126–152 ISBN 9783318021493.

6. Fry, K.; Salser, W. Nucleotide Sequences of HS-α Satellite DNA from Kangaroo Rat Dipodomys Ordii and Characterization of Similar Sequences in Other Rodents. *Cell* **1977**, *12*, 1069–1084, doi:10.1016/0092-8674(77)90170-2.

7. Piccoli, M.C.A.; Bardella, V.B.; Cabral-de-Mello, D.C. Repetitive DNAs in *Melipona scutellaris* (Hymenoptera: Apidae: Meliponidae): chromosomal distribution and test of multiple heterochromatin amplification in the genus. *Apidologie* **2018**, *49*, 497–504, doi:10.1007/s13592-018-0577-z.

8. Cunha, M.S.; Campos, L.A.O.; Lopes, D.M. Insights into the heterochromatin evolution in the genus *Melipona* (Apidae: Meliponini). *Insectes Soc.* **2020**, *67*, 391–398, doi:10.1007/s00040-020-00773-6.

9. Pereira, J.A.; Salomão, T.M.F.; Lopes, D.M. Different repetitive DNA sequences make up heterochromatin in Meliponini. *Apidologie* **2020**, *51*, 855–860, doi:10.1007/s13592-020-00766-1.

10. Pereira, J.A.; Travenzoli, N.M.; de Oliveira, M.P.; de Azevedo Werneck, H.; Salomão, T.M.F.; Lopes, D.M. Molecular cytogenetics in the study of repetitive sequences helping to understand the evolution of heterochromatin in *Melipona* (Hymenoptera, Meliponini). *Genetica* **2021**, *149*, doi:10.1007/s10709-020-00111-5.

11. Pereira, J.A.; Milani, D.; Ferretti, A.B.S.M.; Bardella, V.B.; Cabral-de-Mello, D.C.; Lopes, D.M. The extensive amplification of heterochromatin in *Melipona* bees revealed by high throughput genomic and chromosomal analysis. *Chromosoma* **2021**, *130*, 251–262, doi:10.1007/s00412-021-00764-x.

12. Kerr, W.E. Estudos sobre o gênero *Melipona*. *An. da Esc. Super. Agric. Luiz Queiroz* **1948**, *5*, 181–276, doi:10.1590/S0071-12761948000100005.

13. Brito, R.M.; Caixeiro, A.P. de A.; Pompolo, S. das G.; Azevedo, G.G. Cytogenetic data of *Partamona peckolti* (Hymenoptera, Apidae, Meliponini) by C banding and fluorochrome staining with DA/CMA3 and DA/DAPI. *Genet. Mol. Biol.* **2003**, *26*, 53–57, doi:10.1590/S1415-47572003000100009.

14. Lopes, D.M.; Pompolo, S. das G.; Campos, L.A. de O.; Tavares, M.G. Cytogenetic characterization of *Melipona rufiventris* Lepeletier 1836 and *Melipona mondury* Smith 1863 (Hymenoptera, Apidae) by C banding and fluorochromes staining. *Genet. Mol. Biol.* **2008**, *31*, 49–52, doi:10.1590/S1415-47572008000100010.

15. Barth, A.; Fernandes, A.; das Pompolo, S.G.; Costa, M.A. Occurrence of b chromosomes in *Tetragonisca* Latreille, 1811 (Hymenoptera, Apidae, Meliponini): a new contribution to the cytotaxonomy of the genus. *Genet. Mol. Biol.* **2011**, *34*, 77–79, doi:10.1590/S1415-47572010005000100.

16. Rocha, M.P.; Pompolo, S. das G. Karyotypes and heterochromatin variation (C-bands) in *Melipona* species (Hymenoptera, Apidae, Meliponinae). *Genet. Mol. Biol.* **1998**, *21*, 41–45, doi:10.1590/S1415-47571998000100008.

17. Costa, K.F.; Brito, R.M.; Miyazawa, C.S. Karyotypic description of four species of *Trigona* (Jurine, 1807) (Hymenoptera, Apidae, Meliponini) from the State of Mato Grosso, Brazil. *Genet. Mol. Biol.* **2004**, *27*, 187–190, doi:10.1109/MMSP.2004.1436524.

18. Godoy, D.C.; Ferreira, R.P.; Lopes, D.M. Chromosomal Variation and Cytogenetics of *Plebeia lucii* and *P. phrynostoma* (Hymenoptera: Apidae). *Florida Entomol.* **2013**, *96*, 1559–1566, doi:10.1653/024.096.0439.

19. Santos, J.M.; Diniz, D.; Rodrigues, T.A.S.; Cioffi, M.D.B.; Waldschmidt, A.M. Heterochromatin Distribution and Chromosomal Mapping of Microsatellite Repeats in the Genome of *Frieseomelitta* Stingless Bees (Hymenoptera: Apidae: Meliponini). *Florida Entomol.* **2018**, *101*, 33–39, doi:10.1653/024.101.0107.

20. Rasmussen, C.; Camargo, J.M.F. A molecular phylogeny and the evolution of nest architecture and behavior in *Trigona s.s.* (Hymenoptera: Apidae: Meliponini). *Apidologie* **2008**, *39*, 102–118, doi:10.1051/apido:2007051.

21. Rasmussen, C.; Cameron, S.A. Global stingless bee phylogeny supports ancient divergence, vicariance, and long distance dispersal. *Biol. J. Linn. Soc.* **2010**, *99*, 206–232, doi:10.1111/j.1095-8312.2009.01341.x.

22. Trindade Domingues, A.M.; Waldschmidt, A.M.; Andrade, S.E.; Andrade-Souza, V.; de Oliveira Alves, R.M.; da Silva, J.C.; Costa, M.A. Karyotype characterization of *Trigona fulviventris* Guérin, 1835 (Hymenoptera, Meliponini) by C banding and fluorochrome staining: report of a new chromosome number in the genus. *Genet. Mol. Biol.* **2005**, *28*, 390–393, doi:10.1590/s1415-47572005000300009.

23. Rocha, M.; Pompolo, S.; Campos, L. Citogenética da Tribo Meliponini (Hymenoptera, Apidae). *UNESC, Criciúma* **2003**, 311–320.

24. Fernandes, A.; Barth, A.; Sampaio, W.S. Caracterização citogenética da espécie *Trigona chanchamayoensis* (Hymenoptera, Apidae, Meliponini) encontrada no cerrado brasileiro. *Evolução e Conserv. da Biodiversidade* **2013**, *4*, 63, doi:10.7902/4issecbvol2.2013n55.

25. Barboza, V.P.; Costa, M.A. Cytogenetic Analysis in *Trigona spinipes* Fabricius (Hymenoptera, Meliponina) Reveals Intraspecific Variation. *Neotrop. Entomol.* **2021**, *50*, 846–849, doi:10.1007/s13744-021-00853-7.

26. Tavares, M.G.; Ferreira, R. de P.; Travenzoli, N.M.; Lopes, D.M. Karyotypic variation in the stingless bee *Trigona spinipes* (Hymenoptera: Apidae: Meliponini) from different geographical regions of Brazil. *Apidologie* **2021**, *52*, 1358–1367, doi:10.1007/s13592-021-00906-1.

27. Waldschmidt, A.M.; Fernandes Salomão, T.M.; De Barros, E.G.; De Antônio Oliveira Campos, L. Extraction of genomic DNA from *Melipona quadrifasciata* (Hymenoptera: Apidae, Meliponinae). *Brazilian J. Genet.* **1997**, *20*, doi:10.1590/S0100-84551997000300011.

28. Drummond, A.; Ashton, B.; Cheung, M.; Helid, J.; Kearse, M.; Moir, R.; Stones-Havas, S.; Thierer, T.; Wilson, S. Geneious v4.8. **2009**.

29. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes. *Mob. DNA* **2015**, *6*, 4–9, doi:10.1186/s13100-015-0041-9.

30. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.

31. Novák, P.; Neumann, P.; Macas, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* **2020**, *15*, 3745–3776, doi:10.1038/s41596-020-0400-y.

32. Novák, P.; Ávila Robledillo, L.; Koblížková, A.; Vrbová, I.; Neumann, P.; Macas, J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **2017**, *45*, e111–e111, doi:10.1093/nar/gkx257.

33. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013-2015. **2017**.

34. Ruiz-Ruano, F.J.; López-León, M.D.; Cabrero, J.; Camacho, J.P.M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **2016**, *6*, 28333, doi:10.1038/srep28333.

35. Hirotami, T.I.; Robert, W.T.; Michael, W.J.C.; Ross, H.C. Modes of spontaneous chromosomal mutation and karyotype evolution in ants with reference to the minimum interaction hypothesis. *Japanese J. Genet.* **1988**, *63*, 159–185, doi:10.1266/jjg.63.159.

36. Pinkel, D.; Straume, T.; Gray, J.W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 2934–2938, doi:10.1073/pnas.83.9.2934.

37. Palomeque, T.; Lorite, P. Satellite DNA in insects: a review. *Heredity (Edinb).* **2008**, *100*, 564–573, doi:10.1038/hdy.2008.24.

38. Pita, S.; Panzera, F.; Mora, P.; Vela, J.; Cuadrado, Á.; Sánchez, A.; Palomeque, T.; Lorite, P. Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. *PLoS One* **2017**, *12*, e0181635, doi:10.1371/journal.pone.0181635.

39. Mora, P.; Vela, J.; Ruiz-Ruano, F.J.; Ruiz-Mena, A.; Montiel, E.E.; Palomeque, T.; Lorite, P. Satellitome Analysis in the Ladybird Beetle *Hippodamia variegata* (Coleoptera, Coccinellidae). *Genes (Basel).* **2020**, *11*, 783, doi:10.3390/genes11070783.

40. Bardella, V.B.; Milani, D.; Cabral-de-Mello, D.C. Analysis of *Holhymenia histrio* genome provides insight into the satDNA evolution in an insect with holocentric chromosomes. *Chromosom. Res.* **2020**, *28*, 369–380, doi:10.1007/s10577-020-09642-1.

41. Pavlek, M.; Gelfand, Y.; Plohl, M.; Meštrović, N. Genome-wide analysis of tandem repeats in *Tribolium castaneum* genome reveals abundant and highly dynamic tandem repeat families with satellite DNA features in euchromatic chromosomal arms. *DNA Res.* **2015**, *22*, 387–401, doi:10.1093/dnares/dsv021.

42. Palacios-Gimenez, O.M.; Dias, G.B.; de Lima, L.G.; Kuhn, G.C. e S.; Ramos, É.; Martins, C.; Cabral-de-Mello, D.C. High-throughput analysis of the satellitome revealed enormous diversity of satellite DNAs in the neo-Y chromosome of the cricket *Eneoptera surinamensis*. *Sci. Rep.* **2017**, *7*, 6422, doi:10.1038/s41598-017-06822-8.

43. Palacios-Gimenez, O.M.; Bardella, V.B.; Lemos, B.; Cabral-de-Mello, D.C. Satellite DNAs are conserved and differentially transcribed among *Gryllus* cricket species. *DNA Res.* **2018**, *25*, 137–147, doi:10.1093/dnares/dsx044.

44. Camacho, J.P.M.; Cabrero, J.; López-León, M.D.; Martín-Peciña, M.; Perfectti, F.; Garrido-Ramos, M.A.; Ruiz-Ruano, F.J. Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. *BMC Biol.* **2022**, *20*, 36, doi:10.1186/s12915-021-01216-9.

45. Palacios-Gimenez, O.M.; Milani, D.; Song, H.; Marti, D.A.; López-León, M.D.; Ruiz-Ruano, F.J.; Camacho, J.P.M.; Cabral-De-Mello, D.C.; O'Neill, R. Eight Million Years of Satellite DNA Evolution in Grasshoppers of the Genus *Schistocerca* Illuminate the Ins and Outs of the Library Hypothesis. *Genome Biol. Evol.* **2020**, *12*, 88–102, doi:10.1093/gbe/evaa018.

46. Cabral-de-Mello, D.C.; Zrzavá, M.; Kubíčková, S.; Rendón, P.; Marec, F. The Role of Satellite DNAs in Genome Architecture and Sex Chromosome Evolution in Crambidae Moths. *Front. Genet.* **2021**, *12*, doi:10.3389/fgene.2021.661417.

47. de Lima, L.G.; Ruiz-Ruano, F.J. In-Depth Satellitome Analyses of 37 *Drosophila* Species Illuminate Repetitive DNA Evolution in the *Drosophila* Genus. *Genome Biol. Evol.* **2022**, *14*, doi:10.1093/gbe/evac064.

48. Teixeira, G.; Ferreira, R.; Lopes, D. Comparative cytogenetic analysis reveals chromosomal variability in five stingless bees of the genus *Trigona* (Apidae, Apinae, Meliponini) (*manuscript in preparation*).

| Species | Species group* | Clades* | Sample locality |
|---|---|---|---|
| *Trigona williana* | *"fulviventris"* | A | Altamira, Pará |
| *Trigona pallens* | *"pallens"* | A | Altamira, Pará |
| *Trigona hyalinata* | *"spinipes"* | B | Viçosa, Minas Gerais |
| *Trigona spinipes* | *"spinipes"* | B | Ribeirão Preto, São Paulo |
| *Trigona aff. fuscipennis* | *" fuscipennis"* | B | Florestal, Minas Gerais |
| *Trigona recursa* | *"recursa"* | B | Januária, Minas Gerais |

*Segundo Rasmussem e Camaron (2008)
**Table 1.** *Trigona* species and respective collection sites.

| **Satellite DNA family** | **SF** | **ML** | **AT %** | **Divergence** | **Abundance** | **Genome Size** |
|---|---|---|---|---|---|---|
| ThyaSat01-301 | 1 | 301 | 63.5 | 5.74 | 13.774% | 0.53pg |
| ThyaSat02-300 | 1 | 300 | 62.7 | 5.50 | 2.246% | |
| ThyaSat03-542 | | 542 | 66.1 | 8.99 | 0.370% | |
| ThyaSat04-145 | 2 | 145 | 55.9 | 20.05 | 0.110% | |
| ThyaSat05-291 | | 291 | 58.8 | 9.26 | 0.016% | |
| ThyaSat06-145 | 2 | 145 | 61.4 | 20.13 | 0.015% | |
| ThyaSat07-145 | 2 | 145 | 58.6 | 16.87 | 0.017% | |
| ThyaSat08-145 | 2 | 145 | 58.6 | 20.59 | 0.017% | |
| **Total** | | | | | **16.565%** | |

**Table 2.** Main attributes of satDNAs characterized from *Trigona hyalinata* including families, superfamilies (SF), monomer length (ML), A+T sequence content (A+T%), abundance and divergence (K2P).

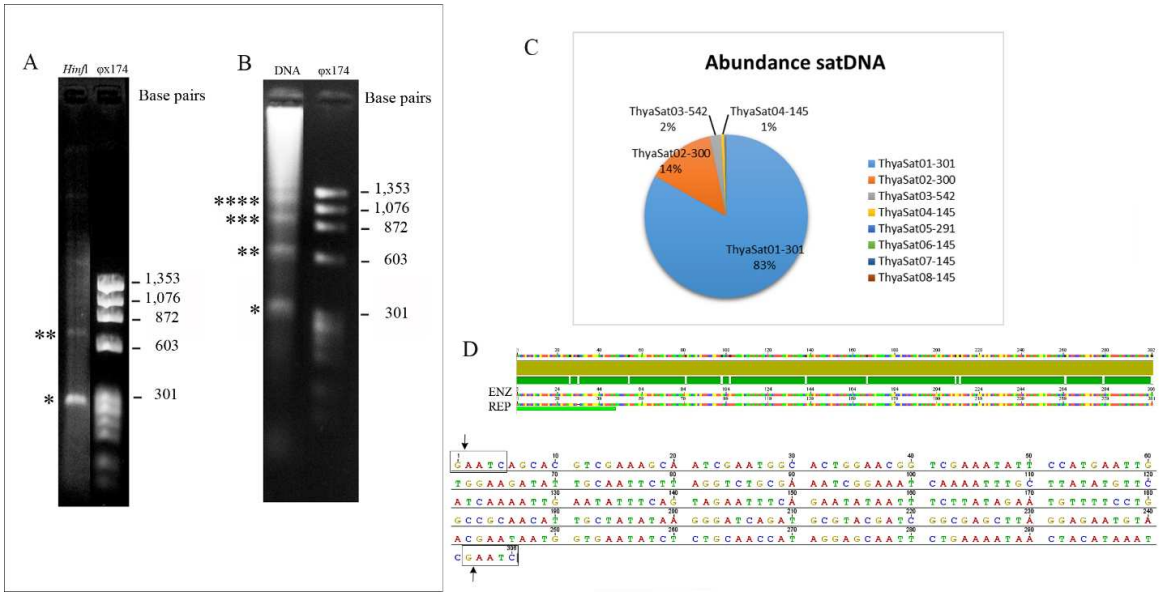



**Figure 1.** Satellite DNAs populating the genome of *Trigona hyalinata*. (**A**) agarose gel showing the enzymatic digestion with *Hinf*I that revealed occurrence of ladder pattern; one asterisk (monomer) and two asterisks (dimer). (**B**) agarose gel after PCR amplification of ThyaSat01-301 showing the ladder pattern; quantity of asterisks corresponds to monomer multiples. (**C**) relative abundance of each satDNA family identified by RepeatExplorer. (**D**) upper panel shows the alignment of consensus sequences obtained by restriction enzyme (ENZ) and by RepeatExplorer (REP) for ThyaSat01-301, note that the sequences are highly similar. The lower panel shows the consensus sequence from cloned fragments and arrows indicate the restriction sites.

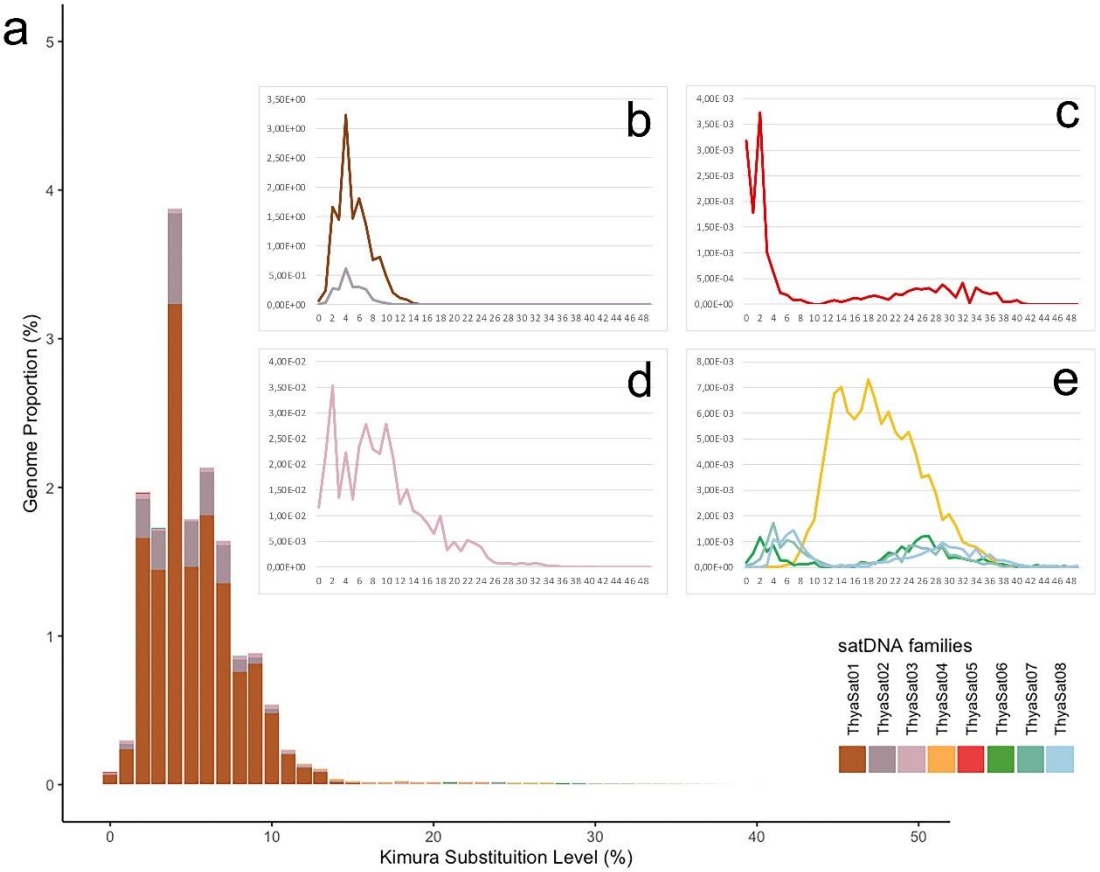



**Figure 2.** Landscape (abundance *versus* divergence) for the satDNAs identified on the genome of *Trigona hyalinata* through analysis of RepeatExplorer. The abundance and divergence were estimated by RepeatMasker. (**A**) general landscape for all satDNAs. In (**B**-**E**) are showed landscapes in more details for specific satDNA families; (**B**) members of SF1, ThyaSat01-301 and ThyaSat02-300; (**C**) ThyaSat05-291; (**D**) ThyaSat03-542; (**E**) members of SF2, ThyaSat04-145, ThyaSat06-145, ThyaSat07-145, and ThyaSat08-145.

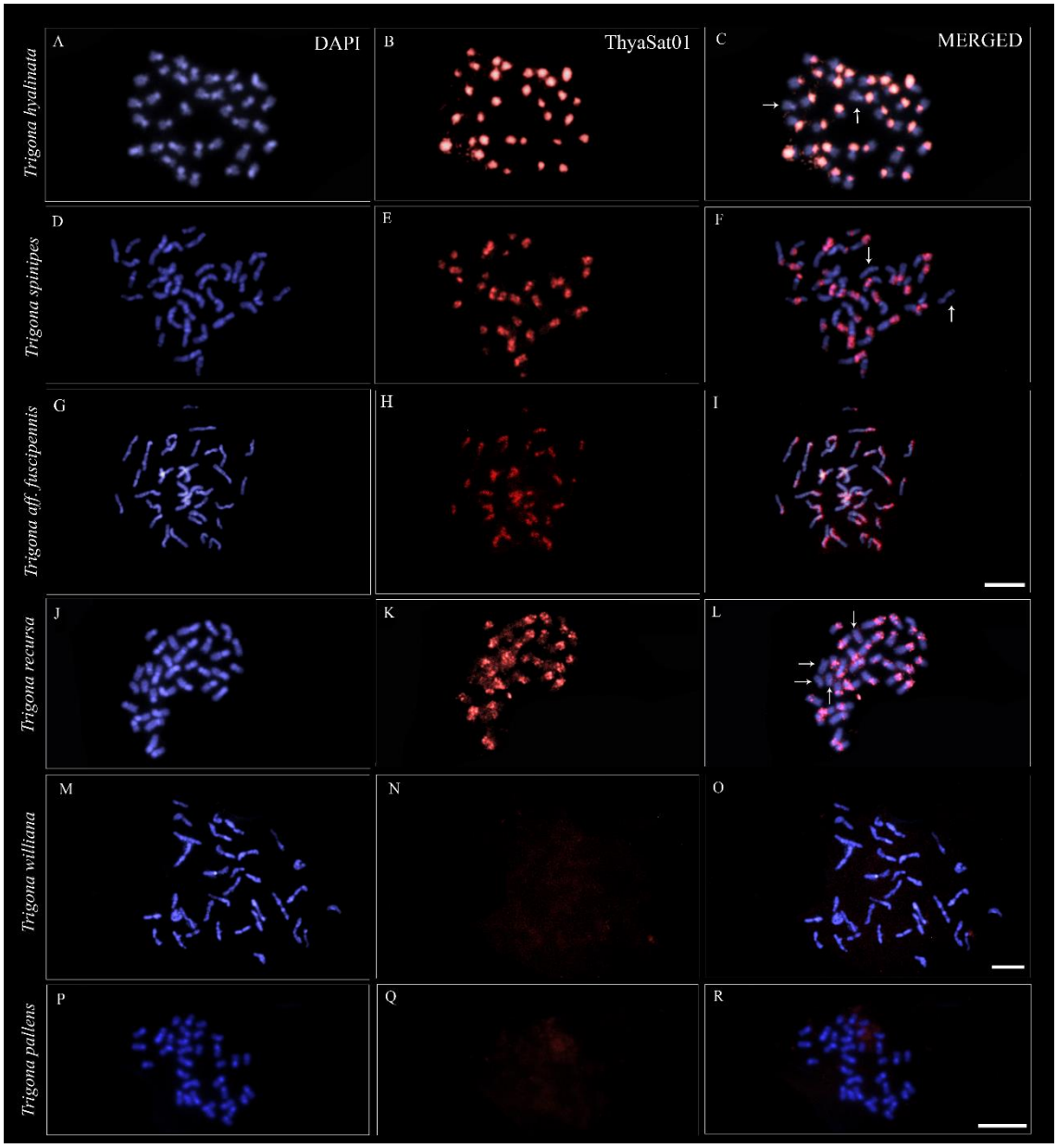



**Figure 3.** Chromosomal distribution of ThyaSat01-301 detected by FISH in *Trigona.* (**A**,**B**,**C**) *Trigona hyalinata*; (**D**,**E**,**F**) *Trigona spinipes;* (**G**,**H**,**I**) *Trigona aff. fuscipennis;* (**J**,**K**,**L**) *Trigona recursa* (**M**,**N**,**O**) *Trigona williana;* (**P**,**Q**,**R**) *Trigona pallens.* (**A**-**L**) Clade B representatives, (**M**-**R**) Clade A representatives. Blue corresponds to chromosomes stained with DAPI and red corresponds to FISH signals. The arrows correspond to the chromosomes where no FISH signals were observed. Note differences in the intensity of hybridization signals between species, with occurrence of scattered signals for *Trigona fuscipennis* and *Trigona recursa*.

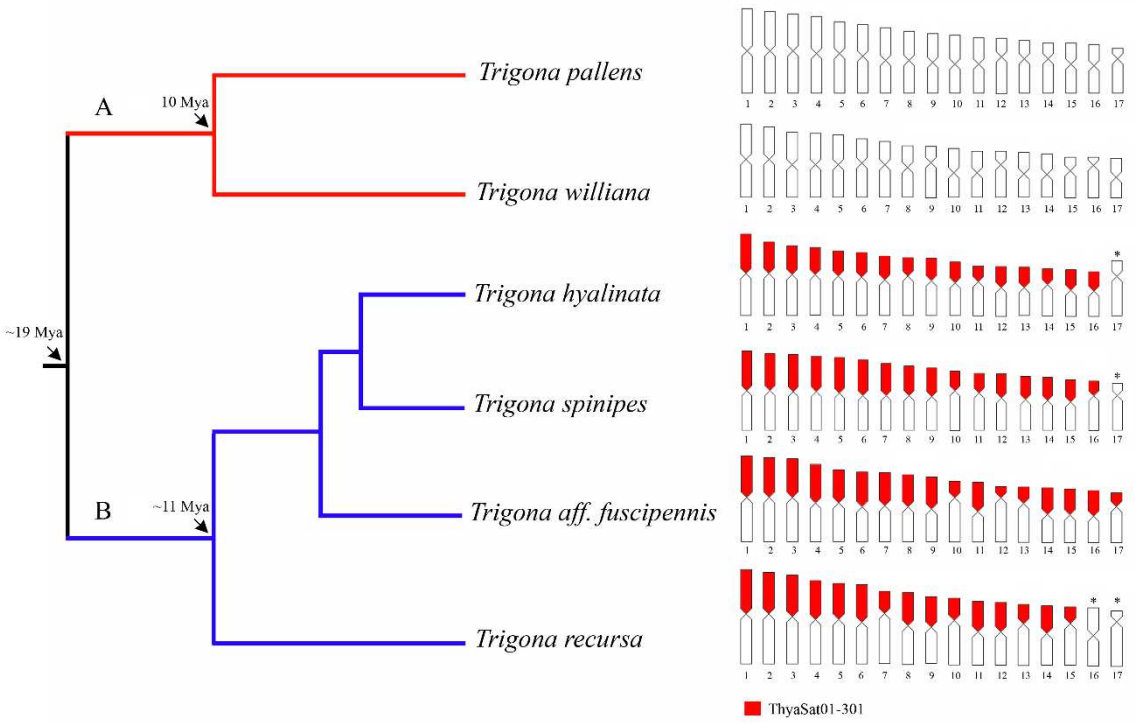



**Figure 4.** Simplified phylogeny for *Trigona* species based on Rasmussem and Camargo [20], Rasmussen and Cameron [21]. The ideograms on the right show the distribution of ThyaSat01-301 on the *Trigona* karyotypes. (*) Chromosomes in which no signs of ThyaSat01-301 were observed. Note that ThyaSat01-301 is located on the heterochromatic arm of chromosomes in Clade B species (blue), but it is not present on Clade A species (red).

## 6.CONCLUSÕES GERAIS

- A análise citogenômica com *M. scutellaris* e *M. quadrifasciata* revelou que entre essas duas espécies existem diferenças quantitativas e qualitativas em relação a porção repetitiva do genoma. Vimos que a grande quantidade de *c*-heterocromatina nos cromossomos de *M. scutellaris* está relacionada à amplificação de uma família de satDNA e um elemento transponível.

- A maioria das famílias de DNAsat identificadas nas diferentes espécies de *Melipona* são compartilhadas, de modo que a diferenciação da *c*-heterocromatina entre os subgêneros se deve principalmente a eventos de amplificação de famílias específicas.

- Em *Trigona* a composição da *c*-heterocromatina é diferente entre clado A e B. ThyaSat01-301 mostrou abundante no genoma de *T. hyalinata* e o principal constituinte da *c*-heterocromatina das espécies do clado B.

- Apesar de *Melipona* e *Trigona* apresentarem diferenças em relação a distribuição da *c*-heterocromatina, nos dois gêneros notamos que a evolução da *c*-heterocromatina está relacionada a modificação da biblioteca de DNAsat ao longo do tempo. Além disso em abelhas parece ser comum uma família de DNAsat altamente abundante como o principal componente da *c*-heterocromatina.