JÉSSICA COUTINHO SILVA

CHROMOSOMAL DISTRIBUTION OF THE REPETITIVE SEQUENCES IN *Zea mays* AND *Psidium guajava*, AND ITS IMPLICATIONS IN KARYOTYPE

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

Orientador: Wellington Ronildo Clarindo

VIÇOSA - MINAS GERAIS

2020

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

T
S586c
2020

Silva, Jéssica Coutinho, 1990-
Chromosomal distribution of the repetitive sequences in *Zea mays* and *Psidium guajava*, and its implications in karyotype / Jéssica Coutinho Silva. – Viçosa, MG, 2020.
60 f. : il. (algumas color.) ; 29 cm.

Orientador: Wellington Ronildo Clarindo.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Citogenética. 2. Goiaba. 3. Cariótipos. 4. Milho. 5. Genomas. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 572.8

JÉSSICA COUTINHO SILVA

CHROMOSOMAL DISTRIBUTION OF THE REPETITIVE SEQUENCES IN *Zea mays* AND *Psidium guajava*, AND ITS IMPLICATIONS IN KARYOTYPE

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

APROVADA: 29 de setembro de 2020.

Assentimento:

Jéssica Coutinho Silva

Autora

Wellington Ronildo Clarindo

Orientador

Aos meus pais, Marluce e Agenário

DEDICO

AGRADECIMENTOS

À Universidade Federal de Viçosa por fornecer um ensino público de qualidade, sem o qual dificilmente teria acesso ao curso de pós-graduação.

Ao Programa de Pós-Graduação em Genética e Melhoramento pela oportunidade.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pelo suporte financeiro.

Ao meu orientador Wellington Ronildo Clarindo, pelos grandes ensinamentos e por ser um exemplo de profissional ético e dedicado.

A secretária do Programa de Pós-Graduação em Genética e Melhoramento nas pessoas do Marco Túlio e Odilon pelo auxílio e dedicação em todos os momentos.

Aos amigos do Laboratório de Citogenética e Citometria, pela torcida e pelos muitos momentos de descontração. Em especial a Mariana pela ajuda nos momentos difíceis. Não tenho palavras para expressar a minha gratidão por tamanha contribuição.

Aos meus familiares pelo companheirismo, amizade, carinho e amor.

Aqueles que são a base desta conquista: meus pais Marluce e Agenário pelo amor e apoio incondicional em todos os momentos, e por sonhar meus sonhos.

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

Aos professores da banca avaliadora, pela disponibilidade e pelas considerações que contribuirão para a melhoria deste trabalho.

A Deus por ter me concedido o dom da sabedoria e por me dar forças e coragem para lutar a cada dia.

***"Se cheguei até aqui foi porque me apoiei no ombro dos gigantes"*.**

(Isaac Newton)

RESUMO

SILVA, Jéssica Coutinho, D.Sc., Universidade Federal de Viçosa, setembro de 2020. **Distribuição cromossômica de sequências repetitivas em *Zea mays* e *Psidium guajava*, e suas implicações no cariótipo.** Orientador: Wellington Ronildo Clarindo.

Os genomas das plantas são ricos em sequências repetitivas, inicialmente reconhecidas como "DNA lixo". Porém, atualmente tem sido demonstrado que elas desempenham papéis importantes em diferentes processos, como rearranjos cromossômicos, expressão e função gênica, além de influenciar nas diferenças de conteúdo de DNA nuclear entre as espécies. Neste cenário, a identificação, comparação e mapeamento de sequências repetitivas geram evidencias acerca da organização cariotípica das plantas. O presente estudo comparou cariótipos de três acessos de *Zea mays*, por meio de citogenética molecular, buscando responder se a variação do tamanho do genoma nuclear entre eles é promovida por quantidades diferenciais de sequências repetitivas e/ou por rearranjos cromossômicos estruturais. Além disso, identificamos e caracterizamos a distribuição de duas sequências de retrotransposons-LTR no cariótipo de *Psidium guajava*. No primeiro estudo, o conteúdo de DNA nuclear dos acessos de *Z. mays* revelou uma variação intraespecífica notável de 2C = 2,00 pg a 2C = 6,10 pg. A partir da análise comparativa dos cariótipos, uma deleção terminal heterozigótica no cromossomo 3 foi apontada como a causa do menor valor 2C, assim como foi observada uma translocação no braço curto do cromossomo 1. O maior valor 2C foi associado à distribuição mais abundante de retrotransposons-LTR da família *Grande* no cariótipo. Além disso, heteromorfismos foram encontrados em relação ao número e à posição da sequência de repetição em tandem de 180 pb. No segundo trabalho, identificamos e caracterizamos dois retrotransposons *Copia* LTR do ponto de vista genômico e cromossômico em *P. guajava*, uma árvore frutífera economicamente importante. A análise citogenômica evidenciou que essas duas sequências de *Copia*, identificadas como *RLC_Pg_1* e *RLC_Pg_2*, estão relacionadas às linhagens *RLC_egBianca_1* e *RLC_egAle_2* de *Eucalyptus grandis*, respectivamente. Nosso trabalho fornece novos insights para esclarecer o papel dos elementos móveis na diversificação do cariótipo de *P. guajava* e *Z. mays*.

Palavras-chave: Citogenética. Goiaba. Cariótipo. Milho. Genoma de planta. Sequências repetitivas.

ABSTRACT

SILVA, Jéssica Coutinho, D.Sc., Universidade Federal de Viçosa, September, 2020. **Chromosomal distribution of the repetitive sequences in *Zea mays* and *Psidium guajava*, and its implications in karyotype.** Adviser: Wellington Ronildo Clarindo.

Plant genomes are rich in repetitive sequences, which were initially recognized as "junk DNA". However, currently it has been demonstrated that they play important roles different processes, such as chromosomal rearrangements, gene expression and function, in addition to influencing in differences of nuclear DNA content between species. In this scenario, the identification, comparison and mapping of repetitive sequences generate evidence about the karyotype organization of plants. The present study compared karyotypes of three *Zea mays* accessions, by molecular cytogenetic seeking to answer if the nuclear genome size variation among them is promoted by differential amounts of repetitive sequences and/or by structural chromosomal rearrangements. Furthermore, we identify and characterize the distribution of two LTR-retrotransposons sequence in *Psidium guajava* karyotype. In the first study, the measurement of the nuclear DNA content of *Z. mays* accessions revealed a remarkable intraspecific variation from 2C = 2.00 pg to 2C = 6.10 pg. From comparative karyotypes analysis, a heterozygous terminal deletion in chromosome 3 was pointed out as a cause of lower 2C value, as well as a translocation in the short arm of chromosome 1 was observed. Differently, higher 2C value was associated with the more abundant distribution of LTR-retrotransposons from the family *Grande* in the karyotype. Moreover, heteromorphisms were found regarding the number and position of the 180-bp tandem repeat sequence. In the second paper, we identified and characterized two *Copia* LTR-retrotransposons from a genomic and chromosomal point of view in *P. guajava*, an economically important fruit tree. The cytogenomic analysis evidenced that these two *Copia* sequences, identified as *RLC_Pg_1* and *RLC_Pg_2*, are related to the lineages *RLC_egBianca_1* and *RLC_egAle_2* from *Eucalyptus grandis*, respectively. Our work provides new insights to enlighten the role of mobile elements in diversification the *P. guajava* and *Z. mays* karyotype.

Keywords: Cytogenetics. Guava. Karyotype. Maize. Plant genome. Repetitive sequences.

SUMÁRIO

1 GENERAL INTRODUCTION

Repetitive sequences are recognized to play important roles the dynamism of genome structure and evolution in plants (Baucom et al. 2009). In addition, depending on their insertion sites, repetitive DNA may cause deleterious effects or gain of genes, chromosome rearrangements, and even can serve as key components to chromosome and chromatid segregation, maintaining chromosome stability (Zhong et al. 2002; Bennetzen and Wang 2014). In this latter case, differences in nuclear DNA content between species with the same chromosomal number are generally attributed to fluctuation in the abundance of repetitive sequences, as observed in *Zea mays* L. (Jian et al. 2017; Silva et al. 2020). In this scenario, the identification, comparison and mapping of repetitive sequences will allow to understand of plants karyotype organization.

Among plant genomes, TEs (transposable elements) represent the vast majority repetitive sequences (Wicker et al. 2007). They are mobile DNA sequences within the genome classified as retrotransposons (Class 1 elements), which move via synthesis of an intermediate RNA using a “copy and paste” mechanism, or transposons (Class 2 elements), in which the transposition occurs via a DNA sequence employing a “cut and paste” mechanism (Wicker et al. 2007). There is a considerable diversity of TEs throughout plant genomes, but the LTR retrotransposons (RTs) are the most abundant (see Table 1).

Table 1. Proportion (%) of transposable elements present in the total genome of plants.

Species	LTR-RTs (Class 1)	DNA transposons (Class 2)	Genome size (Megabases)	Reference
Zea mays L.	75%	8.6%	2,300	Schnable et al. (2009)
Solanum lycopersicum L.	59.24%	5%	900	Jouffroy et al. (2016)
Sorghum bicolor L.	54.43%	7.46%	730	Paterson et al. (2009)
Saccharum spp. (sugarcane)	40.86%	7.93%	10,000	Setta et al. (2014)
Eucalyptus grandis W. Hill ex Maiden	21.90%	5.60%	605	Myburg et al. (2014)
Oryza sativa L.	14.75%	12.96%	389	Sasaki (2005)
Metrosideros polymorpha Gaud.	5.60%	2.43%	304	Izuno et al. (2019)

This abundance in the genome is explained by their ability to replicate, as a *de novo* DNA sequence, and insert into the genome after a transposition event (Kejnovsky et al. 2012; Bennetzen and Wang 2014). Regarding their structure, the key feature of LTR-RTs is the presence of long terminal repeat (LTRs) flanking their internal coding region. Moreover, LTR-RTs possess two genes: *gag* and *polyprotein* (*pol)*. The *gag* gene encodes a protein with a domain that is similar to the viral capsid. The expression of the *pol* gene encodes four enzymatic domains: aspartic proteinase, integrase, reverse transcriptase and RNase H. The location of the integrase in the *pol* gene has been used to classify the LTR-RTs in two superfamilies in plants: *Copia* and *Gypsy* (Wicker et al. 2007; Neumann et al. 2019).

Another conspicuous component of plant genome are tandem repeats (satellites, minisatellites, and microsatellites) that consist of a large number of repeat units (Mehrotra and Goyal 2014). Tandem repeats accumulate particularly at

heterochromatic regions, such as *Z. mays* knobs, that are regions identified in pachytene and mitotic prometaphase and metaphase chromosomes by means of differential staining techniques (Mondin et al. 2014). They are composed of two tandem repeat sequences, of 180 base pairs (bp) (Peacock et al. 1981) and 350 bp (TR-1), or a mixture of both, besides harboring several LTR-RTs (Ananiev et al. 1998; Realini et al. 2018).

Considering that repetitive sequences can occupies the largest fraction of plants genomes and has a pivotal role in shaping the architecture of chromosomes, this study compared karyotypes of three *Z. mays* accessions, by molecular cytogenetic seeking to identify if the nuclear genome size variation among them is promoted by differential amounts of repetitive sequences and/or by structural chromosomal rearrangements. Furthermore, we identify and characterize the distribution of two LTR-RTs sequence in *Psidium guajava* L. karyotype.

2 REFERENCES

Ananiev EV, Phillips RL, Rines HW (1998) Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. Genetics 149:2025–2037

Baucom RS, Estill JC, Chaparro C, et al (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet 5:e1000732. doi: 10.1371/journal.pgen.1000732

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530. doi: 10.1146/annurev-arplant-050213-035811

Izuno A, Wicker T, Hatakeyama M, et al (2019) Updated genome assembly and annotation for *Metrosideros polymorpha*, an emerging model tree species of ecological divergence. G3 Genes|Genomes|Genetics 9:3513 LP – 3520. doi: 10.1534/g3.119.400643

Jian Y, Xu C, Guo Z, et al (2017) Maize (*Zea mays* L.) genome size indicated by 180-bp knob abundance is associated with flowering time. Sci Rep 7:5954. doi: 10.1038/s41598-017-06153-8

Jouffroy O, Saha S, Mueller L, et al (2016) Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. BMC Genomics 17:624. doi: 10.1186/s12864-016-2980-z

Kejnovsky E, Hawkins JS, Feschotte C (2012) Plant Genome Diversity Volume 1. Springer Vienna, Vienna

Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genomics Proteomics Bioinformatics 12:164–171. doi: 10.1016/j.gpb.2014.07.003

Mondin M, Santos-Serejo JA, Bertão MR, et al (2014) Karyotype variability in tropical maize sister inbred lines and hybrids compared with KYS standard line. Front Plant Sci 5:1–12. doi: 10.3389/fpls.2014.00544

Myburg AA, Grattapaglia D, Tuskan GA, et al (2014) The genome of *Eucalyptus grandis*. Nature 510:356–362. doi: 10.1038/nature13308

Neumann P, Novák P, Hoštáková N, Macas J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA 10:1. doi: 10.1186/s13100-018-0144-1

Paterson AH, Bowers JE, Bruggmann R, et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. doi:

10.1038/nature07723

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ (1981) Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci 78:4490–4494. doi: 10.1073/pnas.78.7.4490

Realini MF, Poggio L, Cámara Hernández J, González GE (2018) Exploring karyotype diversity of Argentinian guaraní maize landraces: relationship among South American maize. PLoS One 13:e0198398. doi: 10.1371/journal.pone.0198398

Sasaki T (2005) The map-based sequence of the rice genome. Nature 436:793–800. doi: 10.1038/nature03895

Schnable PS, Ware D, Fulton RS, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. doi: 10.1126/science.1178534

Setta N, Monteiro-Vitorello CB, Metcalfe CJ, et al (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. BMC Genomics 15:540. doi: 10.1186/1471-2164-15-540

Silva JC, Soares FAF, Sattler MC, Clarindo WR (2020) Repetitive sequences and structural chromosome alterations promote intraspecific variations in *Zea mays* L. karyotype. Sci Rep 10:8866. doi: 10.1038/s41598-020-65779-3

Wicker T, Sabot F, Hua-Van A, et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. doi: 10.1038/nrg2165

Zhong CX, Marshall JB, Topp C, et al (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836. doi: 10.1105/tpc.006106

3 RESEARCH PAPER 1: Repetitive sequences and structural chromosome alterations promote intraspecific variations in *Zea mays* L. karyotype

Scientific Reports, 10, 8866 (2020). https://doi.org/10.1038/s41598-020-65779-3

Jéssica Coutinho Silva[1*], Fernanda Aparecida Ferrari Soares[1], Mariana Cansian Sattler[1] and Wellington Ronildo Clarindo[1]

[1]Laboratório de Citogenética e Citometria, Departamento de Biologia Geral, Centro de Ciências Biológicas e da Saúde, Universidade Federal de Viçosa. ZIP 36570-900 Viçosa – MG, Brazil.

3.1 Abstract

LTR-retrotransposons, knobs and structural chromosome alterations contribute to shape the structure and organization of the *Zea mays* karyotype. Our initial nuclear DNA content data of *Z. mays* accessions revealed an intraspecific variation (2C = 2.00 pg to 2C = 6.10 pg), suggesting differences in their karyotypes. We aimed to compare the karyotypes of three *Z. mays* accessions in search of the differences and similarities among them. Karyotype divergences were demonstrated among the accessions, despite their common chromosome number (2n = 20) and ancestral origin. Cytogenomic analyses showed that repetitive sequences and structural chromosome alterations play a significant role in promoting intraspecific nuclear DNA content variation. In addition, heterozygous terminal deletion in chromosome 3 was pointed out as a cause of lower nuclear 2C value. Besides this, translocation was also observed in the short arm of chromosome 1. Differently, higher 2C value was associated with the more abundant distribution of LTR-retrotransposons from the family *Grande* in the karyotype. Moreover, heteromorphism involving the number and position of the 180-bp knob sequence was found among the accessions. Taken together, we provide insights on the pivotal role played by repetitive sequences and structural chromosome alterations in shaping the karyotype of *Z. mays*.

3.2 Introduction

The genus *Zea* is a group of annual and perennial grasses native to a region extending from Mexico to Central America[1]. Based on morphological traits, geographical distribution and genomic data, five *Zea* species are currently recognized: the closely related perennial species *Zea diploperennis* Iltis, Doebley & Guzman (2n = 2x = 20) and *Zea perennis* (Hitchcock) Reeves & Mangelsdorf (2n = 4x = 40); the annual species *Zea luxurians* (Durieu & Ascherson) Bird; *Zea nicaraguensis* Iltis & Benz; and *Zea mays* L. (2n = 2x = 20). *Z. mays* is divided into four subspecies: (i) ssp. *parviglumis* Iltis & Doebley, which represents the direct progenitor of *Z. mays* ssp. *mays*[1–3]; (ii) ssp. *mays*, a well-studied cereal crop with extensive genetic diversity, commonly known as maize or corn[4]; (iii) ssp. *huehuetenangensis* (Iltis & Doebley) Doebley; and (iv) ssp. *mexicana* (Schrader) Iltis. Although controversies still exist regarding the origin of maize, allozyme genetic analyses[5,6] and simple sequence repeat (SSR) markers[7] have provided strong evidence to support *Z. mays* ssp. *parviglumis* as the progenitor of *Z. mays* ssp. *mays*.

Transposable elements (TEs) contribute to the dynamics of the nuclear genome, either through polymorphic insertions and deletions or by mediating ectopic recombination events that can drive structural variation in the genome[8]. TEs constitute over 85% of the maize reference (B73) genome[4]; of these, ~70% belong to Class-I long terminal repeat (LTR) retrotransposons, which replicate through an RNA intermediate, as in the superfamilies *Gypsy* and *Copia*[8]. The families *Huck, Cinful, Tekay/Prem-1* and *Grande* belong to the superfamily *Gypsy*, while *Prem-2/Ji* and *Opie* are included in *Copia*[8,9]. These families constitute a large fraction of the *Z. mays* genome and are distributed throughout its ten chromosomes[10], but predominantly mapped in heterochromatic regions[9]. The cytogenetic determination of the genomic distribution of retrotransposons achieved to date in *Z. mays*[9,10] has evinced their abundance and physical location in the chromosomes.

The genomic dynamism of the LTR-retrotransposon families, which has been characterized by amplification and loss, is responsible for the variation in genome size among *Z. mays* accessions[11,12] that ranges from 2C = 4.50 to 7.11 pg[13,14]. DNA content divergences have also been shown at chromosome level employing image cytometry. The DNA content has been measured for the long and short arms of each chromosome

and the satellite of chromosome 6 of *Z. mays* 'AL Bandeirante'[15]. The DNA content of chromosome 9 (2C = 0.56 pg) was higher than that of chromosome 8 (2C = 0.53 pg)[15], a fact that can be related to accumulation of repetitive sequences in the knobs[16]. Hence, intraspecific variation in nuclear or chromosomal DNA content hints at karyotype differences, emphasizing the need to understand the causes of these variations in distinct *Z. mays* accessions.

In addition to TEs, knobs are also responsible for the intraspecific variation in *Z. mays* nuclear genome size[17,18]. Knobs are heterochromatic regions identified in pachytene and mitotic prometaphase and metaphase chromosomes by means of differential staining techniques[19]. They are composed of two tandem repeat sequences, of 180 base pairs (bp)[20] and 350 bp (TR-1), besides harboring several LTR–retrotransposons[17,21]. The positions, number and size of the knobs are variable among both the accessions and the chromosomes of the same karyotype[9,22]. In some cases, the knobs might serve as chromosomal markers that provide physical evidence of crossing-over events between non-homologous chromosomes[23]. Despite the recognized role of TEs and knobs on the dynamism of *Z. mays* genome size variation, little is known about the important of these sequences on plant fitness. However, Bilinski *et al.*[18] have found evidences that this variation may indeed be adaptive and that heterochromatic knob sequences are likely under the effect of natural selection. Therefore, considering that knob heteromorphism correlates with nuclear genome size, it is fundamental to map these portions in *Z. mays* chromosomes in order to verify their involvement in DNA content divergence.

In addition to LTR–retrotransposons and knobs, the chromosome structure and morphology can also be altered by structural rearrangements: duplication, deletion, translocations and/or inversions[24]. These rearrangements have been revealed in the genera *Solanum*[25], *Brachypodium*[26] and *Z. mays*[27] by comparative cytogenetics via chromosome painting, thus assisting the elucidation of their evolutionary histories. During the process of double-strand break repair, several rearrangements may occur as a result of illegitimate recombination or through recombination of homologous ectopic sequences[28]. Thus, repetitive sequences such as TEs can provide a template for repairing the double-strand breaks. For this reason, heterochromatic regions rich in similar repetitive sequences are considered hotspots for double-strand breaks[29,30].

Beyond the karyotype diversity and dynamism of *Z. mays* ssp. *mays*, this taxon also occupies a wide range of habitats and presents a diversity of morphological traits[1,31], being a crop with several agricultural varieties specific for different uses[31]. For example, the popcorn is characterized by small, hard kernels that explode when heated, forming large flakes (*popping expansion*), the major feature that separates popcorn from other types of maize[31,32]. Sturtevant[33] considered popcorn as a distinct species, *Zea everta*, which was posteriorly reduced to a subspecies[34] and then considered as a mutant of flint maize[35]. However, archeological evidence and the quantitative trait of popping ability rendered improbable the hypothesis of a mutant origin from flint maize[32]. Currently, taxonomists consider that popcorn belongs to the taxon *Z. mays* ssp. *mays* (https://www.itis.gov/about_itis.html; http://www.plantsoftheworldonline.org). The origin and evolutionary relationship of popcorn with other types of maize remains unknown[31]. Therefore, the genomic in situ hybridization (GISH) and the comparative chromosome painting via chromosome-specific probes might to discriminate the homologous chromosome regions, contributing to understand the evolutionary relationships of popcorn.

Considering the remarkable karyotype dynamism, intraspecific variation in nuclear genome size and chromosomal DNA[15,36] within *Z. mays*, the aim of this study was to perform a comparative analysis of the karyotypes of different *Zea* accessions, seeking to identify if the nuclear genome size variation among them is promoted by differential amounts of repetitive sequences and/or by structural chromosomal rearrangements.

3.3 Results

G_0/G_1 peaks of *Z. diploperennis* and 'Milho Pipoca Americano RS 20' overlapped with the internal standard *Z. mays* 'CE-777', even with coefficients of variation between 2.91% and 4.77%. Because these overlapped G_0/G_1 peaks compromise the reliability, we also utilized the internal standard *S. lycopersicum*, thus avoiding this flow cytometry linearity problem. *Z. diploperennis* showed 2C = 5.76 ± 0.06 pg, corresponding to 2C = 5.63 x 10^9 bp, whereas 'Milho Pipoca Americano RS 20' showed mean 2C = 5.55 ± 0.6 pg, corresponding to 2C = 5.43 x 10^9 bp. For the

popcorn ‘15-1149-1’ and ‘AL Bandeirante’, the nuclear 2C value was measured with the internal standard ‘CE-777’ and with *S. lycopersicum*. Popcorn ‘15-1149-1’ presented 2C = 2.00 ± 0.17 pg, corresponding to 2C = 1.96 x 10^9 bp, and ‘AL Bandeirante’ showed 2C = 6.10 pg, corresponding to 2C = 5.97 x 10^9 bp (Supplementary Figs. 1). Thus**,** the mean 2C value of *Z. diploperennis* was 2C = 3.76 pg higher than ‘15-1149-1’, 2C = 0.21 pg higher than ‘Milho Pipoca Americano RS 20’, and 2C = 0.34 pg lower than ‘AL Bandeirante’. ‘Milho Pipoca Americano RS 20’ presented 2C = 3.55 pg higher than ‘15-1149-1’, and 2C = 0.55 pg lower than ‘AL Bandeirante’. In a similar trend, ‘15-1149-1’ exhibited 2C = 4.10 pg lower than ‘AL Bandeirante’.

Given these nuclear genome size differences, we explored the karyotypes in order to understand the causes of these divergences among *Zea* accessions. For this, a metaphasic index of 60% was obtained from the cell cycle arrest, achieved by treatment involving hydroxyurea followed by amiprophos-methyl. Besides, to ensure morphologically preserved chromosomes with well-defined telomeres and primary constrictions, enzymatic maceration of root meristems and air-drying technique were used for slide preparation.

Once the differences in nuclear genome size were verified, the genomic homology among *Z. mays* accessions was confirmed. Even with GISH stringency at 85–90%, all chromosomes of at least ten ‘Milho Pipoca Americano RS 20’ and ‘AL Bandeirante’ metaphases were fully hybridized by the genomic probe of *Z. diploperennis*. The same result was observed for the genomic probe of ‘Milho Pipoca Americano RS 20’ applied to the ‘AL Bandeirante’ karyotype (Fig. 1). From hybridization of our previously constructed probe of *Z. mays* ‘AL Bandeirante’ chromosome 1, the homology between *Z. mays* accessions was also evidenced by specific painting of the chromosome 1 of ‘Milho Pipoca Americano RS 20’ (Supplementary Figs. 2). Therefore, these results provide substantial evidence that popcorn belongs to the subspecies *mays*.

Nevertheless, karyotype differences were demonstrated among the *Z. mays* ssp. *mays* accessions. Structural chromosome changes were identified in all karyotypes of ‘15-1149-1’, which were specifically stained by Feulgen reaction, DAPI and labeled by the 180-bp probe. Two alterations were identified: terminal deletion in

the long arm of chromosome 3 and translocation in the short arm of chromosome 1 (Fig. 2a, Supplementary Figs. 3). The translocation was identified as one small detectable chromosome fragment, but it was not classified as non-reciprocal or reciprocal. Differently, no structural chromosomal aberration was observed in 'Milho Pipoca Americano RS 20' and 'AL Bandeirante' karyotypes (Fig. 2b, c).

Apart from these structural chromosome aberrations, karyotype variations were also found regarding the number and position of the 180-bp knob sequence, including heteromorphism within the same karyotype between the chromosome pair (Fig. 3). The 180-bp sequence was mapped in ten different chromosome portions (1L, 2L, 3L, 4S, 4L, 5S, 5L, 6L or 8L) in the karyotypes of *Z. mays* '15-1149-1' and 'AL Bandeirante'. *Z. mays* '15-1149-1' exhibited positive signals in the chromosome portions 3L, 4S, 5L, 6L and 8L (Fig. 3a, b), whereas 'AL Bandeirante' displayed signals in 1L, 2L, 4L, 5S, 5L and 8L, also presenting heterozygosity for the presence/absence of signals in the chromosome pairs 1 and 5 (Fig. 3c, d). The karyotypes of the analyzed *Z. mays* accessions also differed in relation to *Grande* LTR–retrotransposon mapping. Uniform hybridization signals from this probe were obtained in ten metaphases of 'Milho Pipoca Americano RS 20' (Fig. 4a). The 'AL Bandeirante' karyotype exhibited stronger hybridization signals throughout the chromosome length in 15 metaphases (Fig. 4b), indicating that 'AL Bandeirante' possesses more copies of this LTR–retrotransposon.

3.4 Discussion

Comparing the nuclear DNA content of *Z. diploperennis* (2C = 5.76 pg), 'Milho Pipoca Americano RS 20' (2C = 5.55), '15-1149-1' (2C = 2.00 pg) and 'AL Bandeirante' (2C = 6.10 pg), a difference of up to 4.10 pg was found. Considering the chromosomal DNA content of 'AL Bandeirante', and according to mean values reported by Silva *et al.*[15], the 4.10 pg corresponds to approximately five times the chromosome 1 (2C = 0.80 pg = 4.00 pg) or ten times the chromosome 10 (2C = 0.38 pg = 3.80 pg). These data reinforce the intraspecific variation in nuclear DNA content observed in *Z. mays*, which has been reported to range from 2C = 4.50 pg to 7.11 pg[13,14]. Given this variation, karyotype differences and similarities were sought inside each *Z. mays* accession studied here.

Based on divergences regarding nuclear genome size, the first step was to verify the evolutionary relationship of the *Z. mays* spp. *mays* accessions. Genomic probes of *Z. diploperennis* provided hybridization signals, from telomere to telomere, in the chromosomes of 'AL Bandeirante' and 'Milho Pipoca Americano RS 20', confirming the genomic affinity of these accessions to the basal species *Z. diploperennis*[1]. Hybridization signals were also observed over all *Z. mays* ssp. *mays* chromosomes with the genomic probe of *Z. diploperennis*, but they were weaker than those detected with genomic probes of *Z. mays* spp. *mexicana* and *Z. mays* spp. *parviglumis*[37], which are more phylogenetically close to *Z. mays* spp. *mays*[1]. Reinforcing the evolutionary relationship, the F_1 hybrids of *Z. diploperennis* x *Z. mays* ssp. *mays* presented regular meiotic chromosome pairing and high pollen viability[38,39].

The genome homology between 'AL Bandeirante' and 'Milho Pipoca Americano RS 20' was also endorsed by chromosome painting of the chromosome 1. Using the application proposed by Soares *et al.*[40], Chromosome painting between 'AL Bandeirante' and 'Milho Pipoca Americano RS 20' was informative in comparative karyotype analysis, reflecting the common evolutionary origin of these accessions. This approach has been successfully used for the resolution of phylogenetic questions in *Brachypodium* genus[26]. The GISH and chromosome painting data obtained here confirmed that popcorn belongs to *Z. mays* ssp. *mays*.

After confirming the evolutionary origin, some karyotype divergences were evidenced among the *Z. mays* accessions, as well as differences between homologue chromosome pairs in the same karyotype, providing cytogenetic data to understand the intraspecific variation in nuclear DNA content. Translocation and terminal deletion were distinguished in '15-1149-1', which resulted in a morphological change in the chromosomes 1 and 3, respectively. Thus, cytogenetic preparations applying cell dissociation combined with air-drying technique were considered essential for the correct interpretation of these karyotype changes, since this methodology replaces the squashing step that can promote chromosome breakages[41]. The terminal deletion in the long arm of chromosome 3 occurred around the knob, which is considered a hotspot of chromosome structure alterations. The knobs present a complex organization, in which blocks of 180-bp sequence are interrupted by LTR–retrotransposons[21]. Structural chromosomal rearrangements occur within regions composed of repetitive DNA sequences[24,30]. In addition, Lysák and Schubert[42] related

that repetitive sequences scattered throughout the genome, especially TEs, are involved in various chromosomal rearrangements, such as deletion and translocation, because ectopic homologous sequences provide a template for recombination during the repair of double-strand breaks, a phenomenon denominated *ectopic recombination*.

The translocation and terminal deletion found in '15-1149-1' represent one of the causes associated to the relatively lower nuclear DNA content (2C = 2.00 pg) of this accession in relation to the others. In *S. lycopersicum*, a difference of 2C = 0.09 pg between the wild type and the mutant 'BHG 160' was found to be due to a heterozygous terminal deletion in the short arm of the chromosome 1[43]. The translocation is another karyotype aberration that includes a broken chromosome, resulting in a chromatid fragment. However, differently from a deletion or inversion, the chromatid fragment moves and joins the homologue pair or another chromosome[44]. Therefore, the translocation in the short arm of chromosome 1 also evidences that chromosome breakage occurred in the '15-1149-1' karyotype.

These structural chromosome aberrations were highlighted in '15-1149-1' and associated to low nuclear DNA content, yet other karyotype divergences were found from the mapping of the 180-bp sequence and *Grande* LTR-retrotransposon. Intraspecific variation in nuclear DNA content was also an outcome of the number and heterozygosity of the 180-bp sequence in the karyotype, as well as of the *Grande* LTR-retrotransposon signals found among the *Z. mays* accessions. This result shows that repetitive sequences typical of heterochromatin portions, LTR-retrotransposons and 180-bp knobs, are also responsible for genome size variation within *Z. mays*. The increase in nuclear genome size in this species has also been correlated with 180-bp knob abundance[45]. Furthermore, Bilinski *et al.*[18] demonstrated that the variations in nuclear genome size are driven by natural selection, causing changes in the abundance of repeat sequences across the genome of *Z. mays*, as significant reductions in heterochromatic knobs. Knobs, which are constituted by 180-bp and 350-bp sequences, are polymorphic in relation to their number, size and distribution across the ten *Z. mays* chromosomes[17,46], affecting 5–20% of the length of the chromosome arm[47]. Besides, the heterozygosity observed in the chromosome portions 1L and 5S only in 'AL Bandeirante' is also a karyotype evidence of the differential accumulation of the 180-bp sequence, and consequent change in the chromosomal DNA content

among accessions. A heterozygous condition has been appointed as a cause of crossing-over suppression[48]. Although the origin of this polymorphism is still uncertain, it has been presumed that knobs can move around according to the "complex megatransposons" hypothesis[49]. This hypothesis proposes that TR-1 tandem repeat sequences are capable of forming fold-back DNA segments, driving the knobs in the *Z. mays* genome[49].

Regarding the distribution of the *Grande* LTR-retrotransposon, which belongs to the *Gypsy* superfamily, 'AL Bandeirante' stood out with stronger hybridization signals than 'Milho Pipoca Americano RS 20'. In addition to this mapping, represented by karyograms, the influence of the *Grande* LTR-retrotransposon sequence on nuclear DNA content among *Z. mays* accessions was also demonstrated. *Grande* LTR-retrotransposon distribution produced a uniform hybridization pattern along the extension of metaphasic *Z. mays* chromosomes, differently from was reported by Mroczek and Dawe[10] and Lamb *et al.*[9] that found a speckled hybridization pattern along the chromosome extension. Distribution of the *Gypsy* LTR-retrotransposon in different chromosome regions has also been reported for other species. In *Arabidopsis thaliana* (L.) Heynh[50] and *Asparagus officinalis* L.[24], this LTR is mainly distributed in the centromeres. Differently, in *Silene latifolia* Poir., the signals for this LTR were observed in subtelomeric heterochromatin regions of the chromosomes[51].

Many mechanisms have shaped the karyotype organization in plants, such as LTR-retrotransposons[8]. The higher 2C value and more *Grande* LTR-retrotransposon signals in 'AL Bandeirante' than in 'Milho Pipoca Americano RS 20' reflect the consequences of the LTR-retrotransposon dynamism – amplification and/or loss. Increase in nuclear genome size is promoted by a *de novo* DNA sequence of the retrotransposon that is inserted into the genome after an RNA intermediate to be converted into a cDNA molecule by the reverse transcriptase[4,52]. On the other hand, unequal and illegitimate recombination are associated with a high frequency of genomic DNA loss, and may counterbalance the amplification of LTR-retrotransposons[53], as reported for *Arabidopsis*[54] and *Oryza sativa* L.[55]

3.5 Conclusions

Cytogenomic analysis showed that the intraspecific nuclear genome size variation in *Z. mays* spp. *mays* accessions was promoted by structural chromosome alterations and repetitive sequences. Considering the genomic relationship between the accessions, the dynamic genome of *Z. mays* was newly demonstrated by occurrence of terminal deletions, translocations, 180-bp knob sequence and *Grande* LTR-retrotransposon. Therefore, multiple karyotype factors are related to the changes in *Z. mays* and should be explored in other plant species.

3.6 Material and methods

3.6.1 Plant material

Seeds of *Z. diploperennis* and *Z. mays* '15-1149-1' (popcorn) were provided by the Maize Germplasm Bank of the Embrapa Maize and Sorghum (Sete Lagoas, Minas Gerais – Brazil) and Dr. Marcelo Soriano Viana (Universidade Federal de Viçosa, Minas Gerais – Brazil), respectively. Commercial seeds of *Zea mays* 'AL Bandeirante' and 'Milho Pipoca Americano RS 20' were also used. Seeds of the flow cytometry standards *Solanum lycopersicum* L. 'Stupické' and *Zea mays* 'CE-777' were provided by Dr. Jaroslav Doležel (Experimental Institute of Botany – Czech Republic). According to the *Zea* phylogeny proposed by Hufford *et al.*[1] based on data from ~1000 SNPs by Fang *et al.*[3], *Z. diploperennis* is basal in relation to *Z. mays* spp. *mays*. Therefore, this species was used to compare the nuclear genome size and to construct genomic probes in order to verify the ancestral relationship and homology among *Z. mays* spp. *mays* accessions.

3.6.2 Nuclear genome size

In order to avoid G_0/G_1 peak overlapping in flow cytometry histograms due to close nuclear DNA content, the nuclear genome sizes of *Z. diploperennis*, 'AL Bandeirante', 'Milho Pipoca Americano RS 20' and '15-1149-1' (samples) were measured using the reference standards *S. lycopersicum* or *Z. mays* (2C = 2.00 pg and 2C = 5.55 pg, respectively; Praça-Fontes *et al.*[56]. Leaf fragments from each sample and each internal standard (*S. lycopersicum* or *Z. mays*) were co-chopped[57], and the nuclei were isolated and stained using Otto buffers[58], following the procedure proposed by Praça-Fontes *et al.*[56]. The nuclei suspensions were stained with propidium iodide and analyzed in a BD Accuri C6 flow cytometer (Accuri cytometers, Belgium) equipped with a laser source to detect emissions at FL3 (> 670 nm). The histograms were analyzed using the BD CSampler software. Four technical replicates were performed for each sample with each standard, analyzing over 10,000 nuclei each time. The mean 2C nuclear genome size was measured for each *Zea* sample by dividing the mean channel of the fluorescence peak corresponding to the standard's G_0/G_1 nuclei by that of each sample.

Due to the intraspecific variation in mean 2C value among the *Z. mays* accessions, the karyotypes were characterized with the aim of identifying possible differences and similarities among them. For this, the 180-bp knob sequence and *Grande* LTR–retrotransposon were mapped via fluorescence in situ hybridization (FISH). In addition, GISH using the genomic DNA of the wild related species *Z. diploperennis* was performed to confirm the evolutionary origin of popcorn. Owing to constraints in seed availability and low germination rate, '15-1149-1' was replaced by 'Milho Pipoca Americano RS 20'.

3.6.3 Chromosome preparation

Roots of all accessions showing 1 cm in length were incubated for 18 h in 0.20 g L^{-1} MS salts (Sigma) and 1.75 mM hydroxyurea (inhibitor of ribonucleotide reductase, Sigma) at 30ºC. The roots were washed in dH_2O for four times of 15 min, and then treated with 3 µM amiprophos-methyl (inhibitor of microtubule polymerization, Sigma)

for 4 h at 30ºC. Later, the roots were fixed in 3:1 methanol : acetic acid solution, with three changes of 10 min each, and stored at -20ºC. Afterwards, the roots were again washed for three times in dH_2O, then macerated for 2 h at 36ºC in enzymatic pool (4% cellulase Sigma, 0.4% hemicellulase Sigma, 1% macerozyme Onozuka R10 Yakult, 100% pectinase Sigma) diluted in dH_2O in the proportion 1:8 (enzyme pool : dH_2O). After the maceration procedure, the roots were washed in dH_2O, fixed in 3:1 methanol : acetic acid solution and stored at -20ºC[15]. From the macerated root meristems, slides were prepared by cellular dissociation and air-drying techniques[41]. The slides were chosen for FISH and GISH based on their number of metaphases with morphologically preserved chromosomes, exhibiting well-defined telomere and primary constriction.

Slides showing karyotypes with possible structural chromosome alterations were subjected to Feulgen reaction, a method employed to stoichiometrically and specifically stain the DNA. Thereby, as done by Silva *et al.*[15], selected slides were immediately placed in a fixative solution of methanol : 37% formaldehyde : acetic acid (17:5:1) for 24 h at 25°C. After fixation, the slides were washed in dH_2O, air-dried, hydrolyzed in 5 M HCl for 18 min at 25°C, and stained with Schiff's reagent (Merck) for 16 h at 4°C.

3.6.4 Genomic homology among *Z. mays* accessions

Genomic DNA of 'Milho Pipoca Americano RS 20' and *Z. diploperennis* were obtained according to Doyle and Doyle[59]. DNA concentration and purity were determined by spectrophotometry using NanoDrop (Invitrogen), and DNA integrity was further verified by 1.5% agarose gel electrophoresis. The genomic DNA was amplified and labeled by degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR). The amplification reaction mix consisted of 4 µM degenerated oligonucleotide primer (DOP 5'-CCGACTCGAGNNNNNNATGTGG-3'), 200 ng of genomic DNA of 'Milho Pipoca Americano RS 20' or *Z. diploperennis*, 200 µM of each dNTP (Promega), 1X polymerase buffer, and 2.5 U AccuTaq LA DNA Polymerase (Sigma). The labeling reaction consisted of 4 µM DOP, 200 ng of the genomic DNA, 200 µM each of dATP, dCTP and dGTP, 150 µM dTTP, 50 µM ChromaTide Alexa Fluor 488-5-dUTP (Life Technologies), 1X enzyme reaction buffer (Sigma), and 2.5 U AccuTaq LA DNA Polymerase (Sigma). The PCR conditions were: 96°C for 3 min; 30 cycles of

denaturation at 91°C for 1 min; 56ºC for 1 min; increments of 0.1ºC/s until 68ºC; and 68ºC for 5 min. The labeled genomic probes were quantified in a NanoDrop spectrophotometer (Invitrogen) and evaluated by electrophoresis in 1.5% agarose gel. The amplified fragments ranged from 200 to 700 bp. The genomic DNA probe of *Z. diploperennis* was hybridized in 'AL Bandeirante' and 'Milho Pipoca Americano RS 20', and the genomic DNA probe of 'Milho Pipoca Americano RS 20' was hybridized in 'AL Bandeirante'.

Recently, our research group constructed a chromosome-specific probe for the chromosome 1 of *Z. mays* ssp. *mays* 'AL Bandeirante'[40], which was used for chromosome painting in *Z. mays*. DNA from the chromosome 1, previously amplified by DOP-PCR as reported in Soares *et al.*[40], was used in a new labeling reaction. The PCR program and labeling of the amplified fragment were performed as described above. The probe obtained from chromosome 1 was evaluated by electrophoresis in 1.5% agarose gel, showing fragments ranging from 100 to 900 bp.

3.6.5 Mapping of the 180-bp knob sequence and *Grande* LTR–retrotransposon

The *Grande* LTR–retrotransposon probe was generated by PCR using the primers *F*: 5′-TGCGAGGATAAGTCGGCGAAG-3′ and *R*: 5′-GGTGTTTTTAGGAGTAGGACGGTG-3′[10]. This family was selected for its wide distribution in the *Z. mays* genome. The probe of the 180-bp knob sequence was amplified from the primers *F*: 5′-ATAGCCATGAACGACCATTT-3′ and *R*: 5′-ACCCCACATATGTTTCCTTG-3′[14]. The reaction mixture consisted of: 0.5 µM of each primer, 200 ng of genomic DNA, 200 µM of each dNTP (Promega), 1X reaction buffer (Invitrogen), 2 mM $MgSO_4$ (Invitrogen), and 2 U Platinum Taq DNA Polymerase High Fidelity (Invitrogen). The PCR conditions for the *Grande* LTR–retrotransposon were as follows: initial denaturation at 95°C for 5 min; 30 cycles of denaturation at 95°C for 1 min; annealing at 66ºC for 1 min; extension at 68°C for 1 min and 30 sec; and final extension at 68°C for 5 min. For the 180-bp sequence, amplification conditions were the following: initial denaturation at 95°C for 5 min; 30 cycles of denaturation at 95°C for 1 min; annealing at 47ºC for 1 min; extension at 68°C for 1 min and 30 sec; and final extension at 68°C for 5 min. The labeling reaction consisted of 0.5 µM of each primer, 200 ng of the amplified DNA, 200 µM each of dATP, dCTP and dGTP, 150 µM

dTTP, 40 µM Tetramethylrhodamine 5-dUTP (Roche), 1X of the enzyme reaction buffer (Invitrogen), and 2.5 U AccuTaq LA DNA Polymerase. The PCR conditions were the same as described above for each sequence.

3.6.6 In situ hybridization

The procedures were performed as described by Soares *et al.*[40] and Schwarzacher and Heslop-Harrison[60], with modifications. Briefly, the slides were washed in 1X PBS buffer for 5 min, fixed with 4% formalin for 15 min, washed again in 1X PBS for 5 min, and dehydrated in cold ethanol series (70%, 85% and 100%) for 5 min each. Chromosome denaturation was carried out in 70% formamide / 2X saline-sodium citrate (SSC) buffer for 3 min, at 68°C for 'Milho Pipoca Americano RS' and '15-1149-1' and 70ºC for 'AL Bandeirante'. The difference in temperature is due to over-denaturation when popcorn chromosomes were submitted to 70ºC. Subsequently, the slides were dehydrated in cold ethanol series (70%, 85% and 100%). The hybridization mixture consisted of 50% formamide (Sigma) + 2X SSC (Sigma), 35 µg competitor DNA (Herring Sperm DNA, Promega) and 200 ng of the probe, with denaturation at 85°C for 5 min followed by immediate transfer to ice. Slides were incubated with 35 µL hybridization mixture, covered by plastic coverslip HybriSlip (Sigma) and sealed with Rubber Cement (Elmer's). The hybridization procedure was conducted in a ThermoBrite system (ThermoFisher) at 37°C for 24 h. After this period, stringency washes were performed in three solutions of 50% formamide / 2X SSC and one of 2X SSC, for 5 min each, at 42°C for 'Milho Pipoca Americano RS' and '15-1149-1' and 45ºC for 'AL Bandeirante'. Metaphases were counterstained with 40% glycerol/PBS + 6-diamidino-2-phenylindole (DAPI). The same slides used for FISH mapping of the 180-bp knob sequence were also used to evaluate the differential DAPI banding pattern.

The images were captured with a digital video camera 12-bit CCD (Olympus) coupled to a photomicroscope Olympus BX-60 equipped with epifluorescence and immersion objective of 100X, numeric aperture of 1.4. The frame was digitized using the Image-Pro Plus 6.1 software (Media Cybernetics).

3.7 Acknowledgments

The authors would like to thank the Conselho Nacional de Pesquisa (CNPq, Brazil), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) – Finance Code 001, and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, Brazil) for providing financial support to this study. We are also grateful to the Maize Germplasm Bank of Embrapa Maize and Sorghum (Minas Gerais, Brazil) and Dr. Marcelo Soriano Viana (Universidade Federal de Viçosa, Minas Gerais – Brazil) for kindly provided the *Zea diploperennis* and '15-1149-1' (popcorn) seeds, respectively.

3.8 References

1. Hufford, M. B., Bilinski, P., Pyhäjärvi, T. & Ross-Ibarra, J. Teosinte as a model system for population and ecological genomics. *Trends Genet.* **28**, 606–615 (2012).

2. Iltis, H. H. & Doebley, J. F. Taxonomy of *Zea* (Gramineae). II. subspecific categories in the *Zea mays* complex and a generic synopsis. *Am. J. Bot.* **67**, 994–1004 (1980).

3. Fang, Z. *et al.* Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883–894 (2012).

4. Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).

5. Doebley, J., Goodman, M. M. & Stuber, C. W. Patterns of isozyme variation between maize and Mexican annual teosinte. *Econ. Bot.* **41**, 234–246 (1987).

6. Doebley, J. Molecular evidence and the evolution of maize. *Econ. Bot.* **44**, 6–27 (1990).

7. Matsuoka, Y. *et al.* A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci.* **99**, 6080–6084 (2002).

8. Baucom, R. S. *et al.* Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet.* **5**, e1000732 (2009).

9. Lamb, J. C. *et al.* Distinct chromosomal distributions of highly repetitive sequences in maize. *Chromosom. Res.* **15**, 33–49 (2007).

10. Mroczek, R. J. & Dawe, R. K. Distribution of retroelements in centromeres and neocentromeres of maize. *Genetics* **165**, 809–819 (2003).

11. Sanmiguel, P. & Bennetzen, J. L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **82**, 37–44 (1998).

12. Roessler, K. *et al.* The genome-wide dynamics of purging during selfing in maize. *Nat. Plants* **5**, 980–990 (2019).

13. Díez, C. M. *et al.* Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytol.* **199**, 264–276 (2013).

14. Fourastié, M. F., Gottlieb, A. M., Poggio, L. & González, G. E. Are cytological parameters of maize landraces (*Zea mays* ssp. *mays*) adapted along an altitudinal cline? *J. Plant Res.* **131**, 285–296 (2018).

15. Silva, J. C., Carvalho, C. R. & Clarindo, W. R. Updating the maize karyotype by chromosome DNA sizing. *PLoS One* **13**, e0190428 (2018).

16. Poggio, L., Rosato, M., Chiavarino, A. M. & Naranjo, C. A. Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). *Ann. Bot.* **82**, 107–115 (1998).

17. Realini, M. F., Poggio, L., Cámara Hernández, J. & González, G. E. Exploring karyotype diversity of Argentinian guaraní maize landraces: relationship among South American maize. *PLoS One* **13**, e0198398 (2018).

18. Bilinski, P. *et al.* Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genet.* **14**, e1007162 (2018).

19. Mondin, M. *et al.* Karyotype variability in tropical maize sister inbred lines and hybrids compared with KYS standard line. *Front. Plant Sci.* **5**, 1–12 (2014).

20. Peacock, W. J., Dennis, E. S., Rhoades, M. M. & Pryor, A. J. Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proc. Natl. Acad. Sci.* **78**, 4490–4494 (1981).

21. Ananiev, E. V, Phillips, R. L. & Rines, H. W. Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. *Genetics* **149**, 2025–2037 (1998).

22. Albert, P. S., Gao, Z., Danilova, T. V. & Birchler, J. A. Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenet. Genome Res.* **129**, 6–16 (2010).

23. McClintock, B. A cytological demonstration of the location of an interchange between two non-homologous chromosomes of *Zea mays. Proc. Natl. Acad. Sci.* **16**, 791–796 (1930).

24. Li, S. F. *et al.* Chromosome evolution in connection with repetitive sequences and epigenetics in plants. *Genes (Basel)* **8**, 290 (2017).

25. Lou, Q., Iovene, M., Spooner, D. M., Buell, C. R. & Jiang, J. Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. *Chromosoma* **119**, 435–442 (2010).

26. Betekhtin, A., Jenkins, G. & Hasterok, R. Reconstructing the evolution of *Brachypodium* genomes using comparative chromosome painting. *PLoS One* **9**, e115108 (2014).

27. Albert, P. S. *et al.* Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. *Proc. Natl. Acad. Sci.* **116**, 1679–1685 (2019).

28. Schubert, I. & Lysak, M. A. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* **27**, 207–216 (2011).

29. Lysak, M. A. *et al.* Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci.* **103**, 5224–5229 (2006).

30. Schubert, I., Rieger, R., Fuchs, J. & Pich, U. Sequence organization and the mechanism of interstitial deletion clustering in a plant genome *Vicia faba*). *Mutat. Res. Lett.* **325**, 1–5 (1994).

31. Ziegler, K. Popcorn in *Specialty Corns* (ed. Hallauer, A. R.) 205–240 (CRC Press, 2001).

32. Zinsly, J. R. & Machado, J. A. Milho pipoca in *Melhoramento e produção do milho no Brasil* (ed. Paterniani, E.) 339–348 (Fundação Cargil, 1978).

33. Sturtevant, E. L. Notes on maize. *J. Torrey. Bot. Soc.* **21**, 319–343 (1894).

34. Bailey, L. H. *Manual of cultivated plants.* (Macmillan & Co. Ltd, 1924).

35. Erwin, A. T. The origin and history of pop corn, *Zea mays* L. var. indurata (Sturt.) Bailey mut. everta (Sturt.) Erwin. *Agron. J.* **41**, 53–56 (1949).

36. Rosado, T. B., Clarindo, W. R. & Carvalho, C. R. An integrated cytogenetic, flow and image cytometry procedure used to measure the DNA content of *Zea mays* A and B chromosomes. *Plant Sci.* **176**, 154–158 (2009).

37. González, G. E. & Poggio, L. Genomic affinities revealed by GISH suggests intergenomic restructuring between parental genomes of the paleopolyploid genus *Zea*. *Genome* **58**, 433–439 (2015).

38. Naranjo, C. A., Molina, M. C. & Poggio, L. Evidencias de un número básico X=5 en el género *Zea* y su importancia en estudios del origen del maíz. *Acad. Nac. Ciencias Exactas, Físicas y Nat. Buenos Aires* **5**, 43–53 (1990).

39. Poggio, L., González, G., Confalonieri, V., Comas, C. & Naranjo, C. A. The genome organization and diversification of maize and its allied species revisited: evidences from classical and FISH-GISH cytogenetic analysis. *Cytogenet. Genome Res.* **109**, 259–267 (2005).

40. Soares, F. A. F. *et al.* Plant chromosome-specific probes by microdissection of a single chromosome: is that a reality? *Front. Plant Sci.* **11**, 1–9 (2020).

41. Carvalho, C. R. & Saraiva, L. S. An air drying technique for maize chromosomes without enzymatic maceration. *Biotech. Histochem.* **68**, 142–145 (1993).

42. Lysák, M. A. & Schubert, I. Mechanisms of chromosome rearrangements in *Plant Genome Diversity Volume 2* (eds. Greilhuber, J., Dolezel, J. & Wendel, J. F.) 137–147 (Springer Vienna, 2013). doi:10.1007/978-3-7091-1160-4_9

43. Karsburg, I. V., Carvalho, C. R. & Clarindo, W. R. Identification of chromosomal deficiency by flow cytometry and cytogenetics in mutant tomato (*Solanum lycopersicum*, Solanaceae) plants. *Aust. J. Bot.* **57**, 444 (2009).

44. Burnham, C. R. Chromosomal interchanges in plants. *Bot. Rev.* **22**, 419–552 (1956).

45. Jian, Y. *et al.* Maize (*Zea mays* L.) genome size indicated by 180-bp knob abundance is associated with flowering time. *Sci. Rep.* **7**, 5954 (2017).

46. Xiong, Z. *et al.* Heterozygosity of knob-associated tandem repeats and knob instability in mitotic chromosomes of *Zea* (*Zea mays* L. and *Z. diploperennis* Iltis Doebley). *J. Integr. Plant Biol.* **47**, 1345–1351 (2005).

47. Kato, A., Lamb, J. C. & Birchler, J. A. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci.* **101**, 13554–13559 (2004).

48. Ghaffari, R., Cannon, E. K. S., Kanizay, L. B., Lawrence, C. J. & Dawe, R. K. Maize chromosomal knobs are located in gene-dense areas and suppress local recombination. *Chromosoma* **122**, 67–75 (2013).

49. Ananiev, E. V, Phillips, R. L. & Rines, H. W. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? *Proc. Natl. Acad. Sci.* **95**, 10785–10790 (1998).

50. Initiative, T. A. G. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).

51. Kejnovsky, E. *et al.* Retand: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat. *Mol. Genet. Genomics* **276**, 254–263 (2006).

52. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).

53. Tenaillon, M. I., Hollister, J. D. & Gaut, B. S. A triptych of the evolution of plant transposable elements. *Trends Plant Sci.* **15**, 471–478 (2010).

54. Devos, K., Brown, J. & Bennetzen, J. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12**, 1075–1079 (2002).

55. Vitte, C., Panaud, O. & Quesneville, H. LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. *BMC*

Genomics **8**, 218 (2007).

56. Praça-Fontes, M. M., Carvalho, C. R. & Clarindo, W. R. C-value reassessment of plant standards: an image cytometry approach. *Plant Cell Rep.* **30**, 2303–2312 (2011).

57. Galbraith, D. W. *et al.* Rapid flow cytometric nalysis of the cell cycle in intact plant tissues. *Science* **220**, 1049–1051 (1983).

58. Otto, F. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA in *Methods in Cell Biology* (eds. Darzynkiewiez, Z. & Crissman, H. A.) 105–110 (Academic Press, 1990).

59. Doyle, J. & Doyle, J. Isolation of plant DNA from fresh tissue. *Focus (Madison)* **12**, 13–15. (1990).

60. Schwarzacher, T. & Heslop-Harrison, P. *Practical in situ hybridization*. (BIOS Scientific Publishers Ltd, 2000).

3.9. Figures

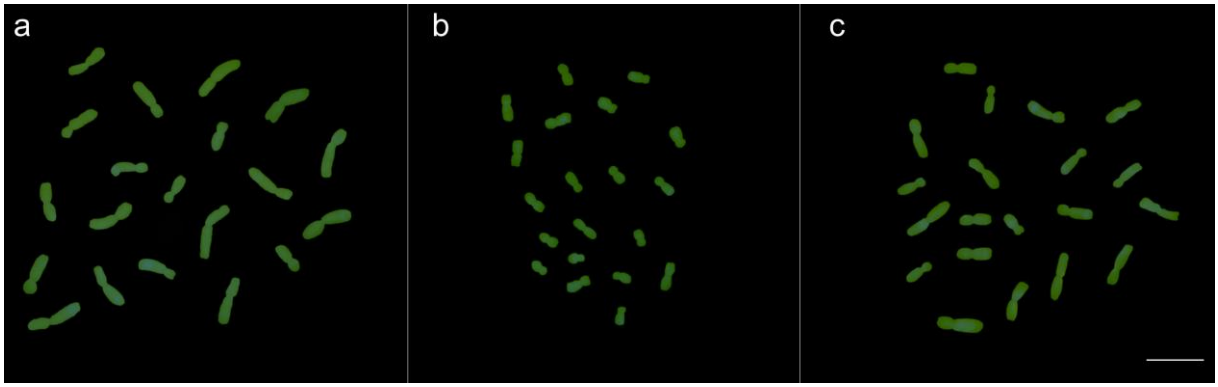


Fig. 1 – GISH in metaphase chromosomes of *Z. mays* labelled with ChromaTide-488-5-dUTP (green). **(a)** 'AL Bandeirante' chromosomes fully labelled by the genomic probe from 'Milho Pipoca Americano RS 20'. **(b) '**AL Bandeirante' and **(c)** 'Milho Pipoca Americano RS 20' chromosomes fully labeled by the genomic probe of *Z. diploperennis*. Bar = 10 µm. Images were digitized using the Image-Pro Plus software version 6.1 (https://www.mediacy.com/imageproplus).

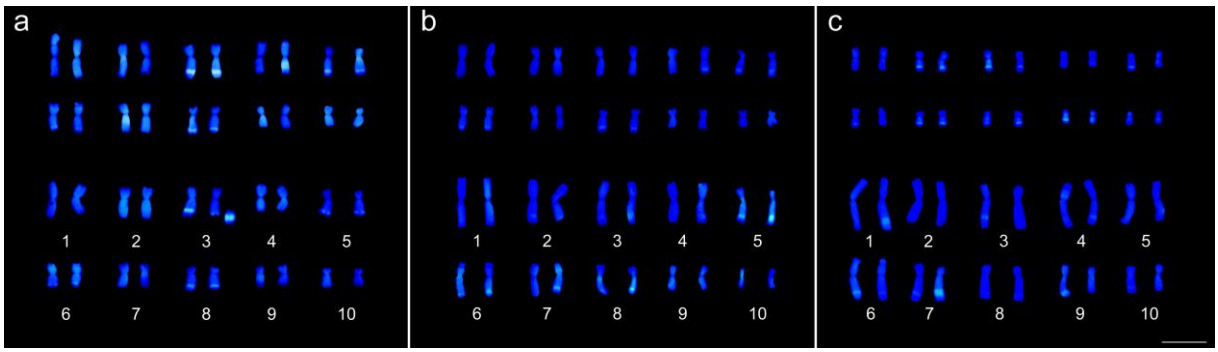


Fig. 2 – Differential DAPI staining (blue) in two metaphases of **(a)** '15-1149-1', **(b)** 'Milho Pipoca Americano RS 20' and **(c)** 'AL Bandeirante'. **(a)** Structural chromosome alterations in '15-1149-1': translocation in the short arm of chromosome 1 and heterozygous terminal deletion in the long arm of chromosome 3. **(b, c)** In 'Milho Pipoca Americano RS 20' and 'AL Bandeirante', no structural chromosomal aberration was observed. The DAPI-banding pattern in *Z. mays* chromosomes was promoted by the preferential binding of this fluorochrome to A-T rich sequences, allowing to evidence the knob portions in a cyan blue color. Bar = 10 µm. Images were digitized using the Image-Pro Plus software version 6.1 (https://www.mediacy.com/imageproplus).

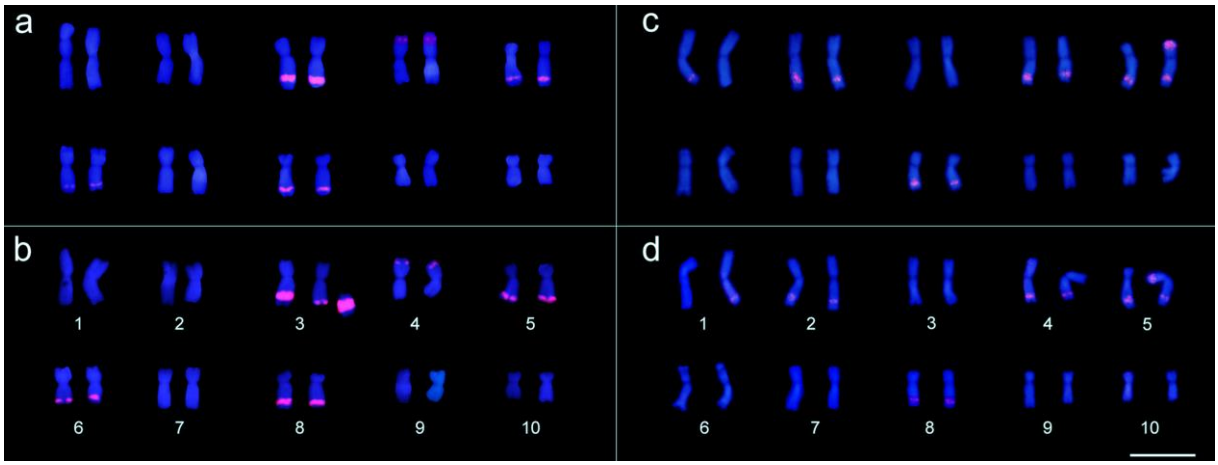


Fig. 3 – 180-bp sequence site mapping from probe labelled with Tetramethylrhodamine 5-dUTP (red) in metaphases of **(a, b)** '15-1149-1' and **(c, d)** 'AL Bandeirante'. **(a, b)** In '15-1149-1', 180-bp was mapped in the interstitial portion of the long arm in chromosomes 3, 5, 6 and 8, and in the terminal portion of the short arm on chromosome 4. **(a)** Note the translocation in the short arm of the chromosome 1, and **(b)** the heterozygous terminal deletion in the chromosome 3 long arm around the large block of 180-pb sequence. **(c, d)** In 'AL Bandeirante', 180-bp sequence was mapped on in the interstitial portion of the long arm in chromosomes 1, 2, 4, 5 and 8, and in the terminal portion of the short arm in chromosome 5. Note the heterozygosity of the 180-bp sequence in the chromosomes 1 and 5. Bar = 10 µm. Images were digitized using the Image-Pro Plus software version 6.1 (https://www.mediacy.com/imageproplus).

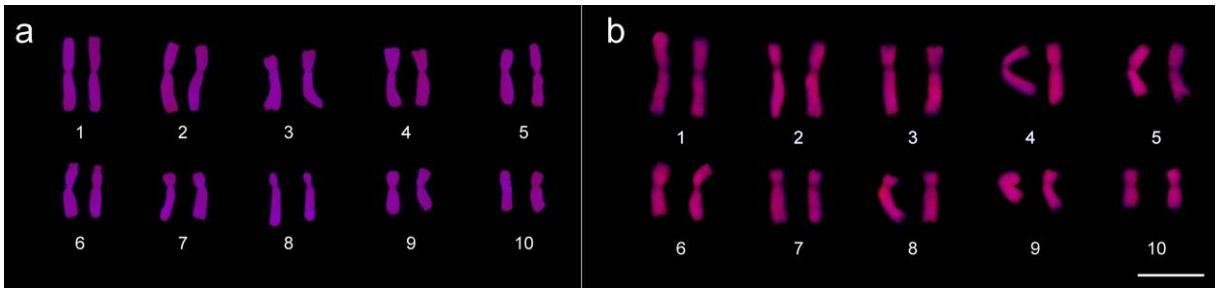
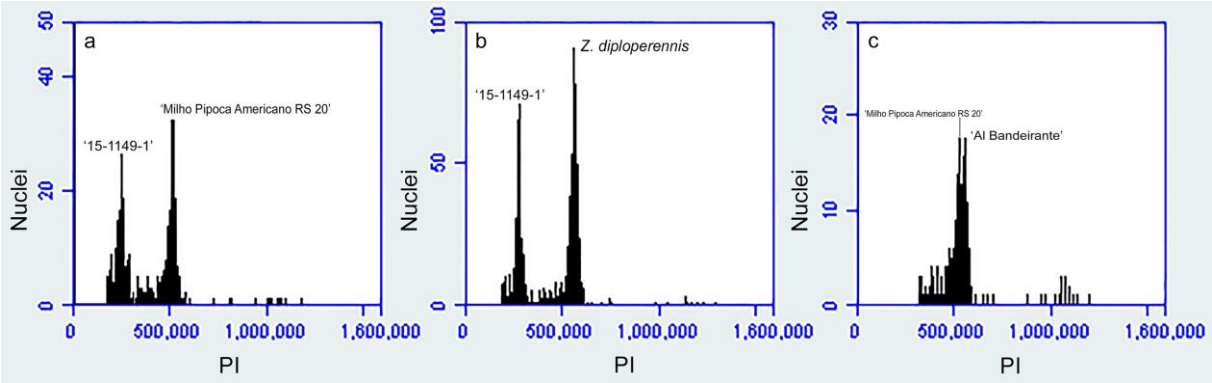
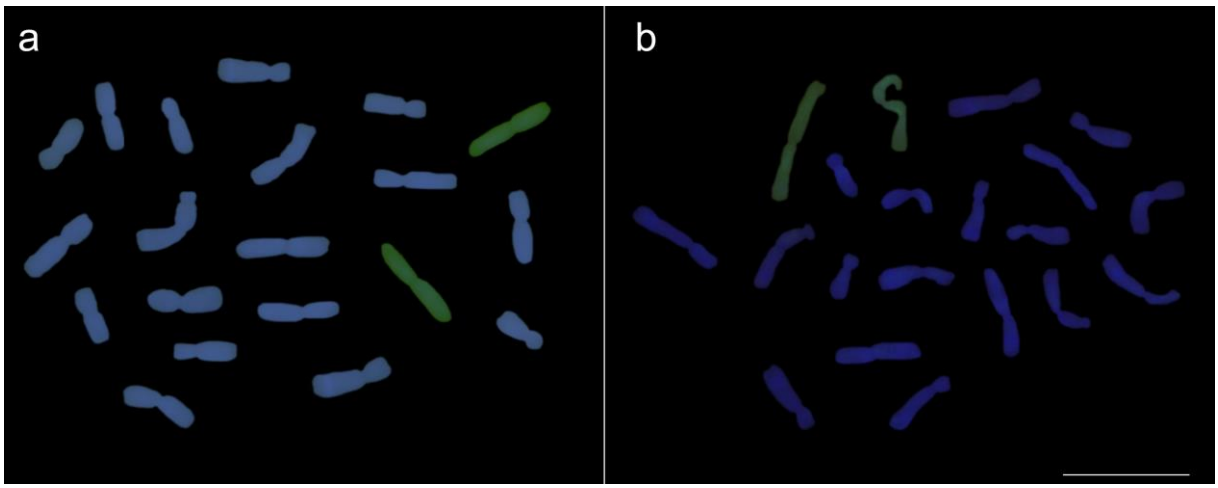


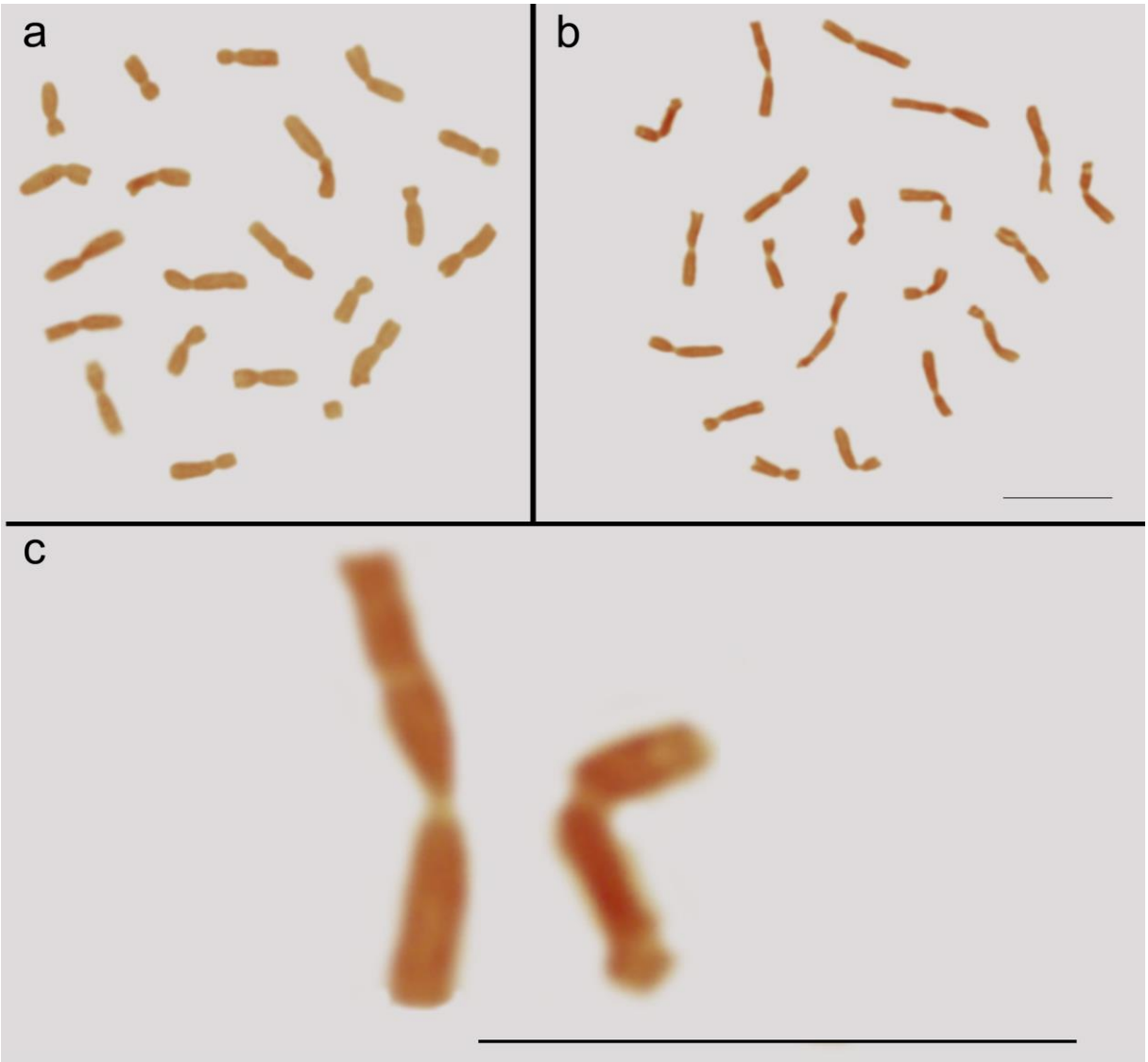
Fig. 4 – Mapping of the *Grande* LTR-retrotransposon sequence, from probes labelled with Tetramethylrhodamine 5-dUTP (red), in chromosomes of **(a)** 'Milho Pipoca Americano RS 20' and **(b)** 'AL Bandeirante'. **(a)** Hybridization signals were fully labeled in the ten chromosomes, telomere to telomere, in all metaphases of 'Milho Pipoca Americano RS 20'. **(b)** In 'AL Bandeirante' metaphases, the chromosomes displayed strong hybridization signals. Bar = 10 µm Images were digitized using the Image-Pro Plus software version 6.1 (https://www.mediacy.com/imageproplus).



Supplementary Figure 1 – Nuclear DNA content of the *Z. mays* spp. *mays* accessions and *Z. diploperennis*. **(a)** '15-1149-1' G_0/G_1 nuclei peak 2C = 2.00 pg and 'Milho Pipoca Americano RS 20' G_0/G_1 nuclei peak 2C = 5.55 pg. **(b)** '15-1149-1' G_0/G_1 nuclei peak 2C = 2.00 pg and *Z. diploperennis* G_0/G_1 nuclei peak 2C = 5.76 pg. **(c)** 'Milho Pipoca Americano RS 20' G_0/G_1 nuclei peak 2C = 5.55 pg and 'AL Bandeirante' G_0/G_1 nuclei peak 2C = 6.10 pg. Histograms were generated using BD Csampler software version Accuri C6.



Supplementary Figure 2 – Chromosome painting in *Z. mays* **(a)** 'Milho Pipoca Americano RS 20' and **(b)** 'AL Bandeirante' using chromosome-specific probe constructed for Chromosome 1 of *Z. mays* labelled with ChromaTide-488-5-dUTP (green). Bar = 10 µm. Images were digitized using the Image-Pro Plus software version 6.1 (https://www.mediacy.com/imageproplus).



Supplementary Figure 3 – Metaphases of '15-1149-1' stained by the Feulgen reaction. Metaphase evincing **(a)** a chromosome fragment resulted from a deletion and **(b)** translocations. **(c)** Enlarged view of chromosomes that exhibited structural alterations in **b**. Bar = 10 µm. Images were digitized using the Image-Pro Plus software version 6.1 (https://www.mediacy.com/imageproplus).

4 RESEARCH PAPER 2: Chromosomal distribution of *Copia* LTR-retrotransposons in *Psidium guajava* L.

Authors: Jéssica Coutinho Silva[1], Mariana Cansian Sattler[1], Miquéias Fernandes[2], Marcia Flores da Silva Ferreira[2], Wellington Ronildo Clarindo[1]

[1]Laboratório de Citogenética e Citometria, Departamento de Biologia Geral, Centro de Ciências Biológicas e da Saúde, Universidade Federal de Viçosa. ZIP: 36.570-900 Viçosa – MG, Brazil.
[2]Laboratório de Genética e Melhoramento Vegetal, Departamento de Agronomia, Centro de Ciências Agrárias e Engenharias, Universidade Federal do Espírito Santo, Alegre, ES, 29500-000, Brazil

✉Corresponding author: e-mail: coutinho.silva530@gmail.com
ORCID of Jéssica Coutinho Silva: https://orcid.org/0000-0002-9009-9942
Tel.: +55 31 3612-5028

4.1 Abstract

LTR-retrotransposons (LTR-RTs) are ubiquitous in plant genomes and variations in their abundance and distribution can give insights on karyotype evolution. Despite *Psidium guajava* L. being an important fruit and medicinal species of the Myrtaceae family, until now, cytogenetic studies in *P. guajava* are restricted to chromosome counting, morphometric characterization and nuclear genome size measurement. In this study, we used nine families of LTR-RTs from *Eucalyptus grandis* to identify and characterize the distribution of *Copia* LTR-RTs in *P. guajava* karyotype. The full-length sequences *RLC_Pg_1* and *RLC_Pg_2* of *P. guajava* showed 86.2% and 71.7% identity in relation to *RLC_egBianca_*1 and *RLC_egAle_2* of *E. grandis*, respectively. The identity between these sequences was also evidenced by their clustering in the phylogenetic analysis. Probes obtained from the identified *Copia* LTR-RTs sequences were hybridized in *P. guajava* metaphase chromosomes; the hybridization signals were telomere to telomere. For the first time, we identified and characterized two *Copia* LTR-RTs from a genomic and chromosomal point of view in a *Psidium* species. Considering the relevance of LTR-RTs characterization and classification in the current scenario of comparative sequencing and genomics, these results may lay the groundwork to understand the dynamics and mechanisms that shape the karyotype of *P. guajava* and other Myrtaceae species.

Keywords: Molecular cytogenetics, Guava, Myrtaceae, Cytogenetics, Plant genome, Mobile elements.

4.2 Introduction

Psidium L. (Myrtaceae) is a monophyletic genus comprising around 100 species (WCSP 2020). One of the most important species of the genus is *Psidium guajava* L., which is cultivated mainly for fruit production and/or source of secondary metabolites explored by the pharmaceutical industry (Gutiérrez et al. 2008). Providing insights about genomic features and changes related to agronomic traits, cytogenomic studies has been performed for different species, including the transposable elements (TEs) mapping (Huang et al. 2017; Deng et al. 2019; Assis et al. 2020). Until now, cytogenetic studies in *Psidium* are restricted to 2*n* chromosome number and nuclear genome size. *P. guajava* is the most-studied species regarding karyotype characterization, showing 2*n* = 2x = 22 chromosomes and a nuclear genome size of 2C = 0.95 pg (1C = 465,5 Mbp, Coser et al. 2012; Marques et al. 2016). The chromosomes of this species are relatively small, ranging from 2.03 µm (chromosome 1) to 0.85 µm (chromosome 11), and are classified as metacentric (3, 4, 8, 9, 10) and submetacentric (1, 2, 5, 6, 7, 11) (Coser et al. 2012; Marques et al. 2016).

TEs are genomic DNA sequences either categorized as retrotransposons (Class 1 elements), which move via an intermediate RNA using a “copy and paste” mechanism, or transposons (Class 2 elements), in which the transposition occurs via a DNA sequence employing a “cut and paste” mechanism (Wicker et al. 2007). There is a considerable diversity of mobile elements throughout plant genomes, but the long terminal repeat (LTR) retrotransposons (RTs) are the most abundant, representing for example 75% of the nuclear genome of *Zea mays* L. (Poaceae) (Schnable et al. 2009), 62% of *Solanum lycopersicum* L. (Solanaceae) (Paz et al. 2017), 21.90% of *Eucalyptus grandis* W. Hill ex Maiden (Myrtaceae) (Myburg et al. 2014), and 5.60% of *Metrosideros polymorpha* Gaud. (Myrtaceae) (Izuno et al. 2019). In plants, LTR-RTs are classified into the superfamilies *Copia* or *Gypsy* according to the position of the integrase domain along the encoded *polyprotein* gene. Both superfamilies are subdivided into lineages and families, considering encoding region similarities and overall structures (Neumann et al. 2019).

The LTR-RTs that present autonomous activity in the genome may display variability in their occurrence, abundance and chromosome distribution. Due to this,

the physical localization of TEs in chromosomes provides insights into genome structure and evolution (Bennetzen and Wang 2014; Assis et al. 2020). DNA sequence mapping by fluorescent in situ hybridization (FISH) has successfully been employed for a detailed investigation of TEs location and dynamics (Assis et al. 2020). FISH performed in plant metaphase chromosomes demonstrated that *Copia* retrotransposons are distributed throughout the genome, presenting a scattered or clustered distribution (Lamb et al. 2007). In *Vicia faba* L. and *Pinus elliottii* Engelm., these elements occur uniformly distributed over the chromosome length, but with reduced hybridization signal in the NOR, centromere and telomere (Kamm et al. 1996; Pearce et al. 1996). Differently, *Copia* retrotransposons were preferentially located in the centromeric and pericentromeric regions of *Arabidopsis thaliana* L. chromosomes (Heslop-Harrison 2000), as well as in the subterminal and telomeric regions of all chromosomes of *Erianthus arundinaceus* Hainan (Huang et al. 2017).

LTR-RTs are also recognized to play an important role in plant genome evolution. The integrated view of cytogenetic and genomic data accumulated until now leads to the knowledge that genome size variations observed in *Z. mays* involves mainly differences in the distribution of LTR-RTs (Estep et al. 2013; Silva et al. 2020). Furthermore, repetitive DNA-based FISH has been successfully applied to investigate taxonomic relationships comparing the physical localization and accumulation of retrotransposons (Huang et al. 2017; Deng et al. 2019; Assis et al. 2020).

Considering the LTR-RTs pivotal role in shaping the architecture of plant chromosomes, we aimed to identify and characterize the distribution of *Copia* retrotransposon elements in the *P. guajava* karyotype.

4.3 Material and methods

4.3.1 Plant material

Seeds of *Psidium guajava* ‘Paluma’ were provided from the active germplasm collection of the Universidade Federal do Espírito Santo (Alegre, Espírito Santo, Brazil). As few *P. guajava* roots were obtained in Petri dishes due to low seed germination and slow root growth and oxidation, we used in vitro tissue culture as an

alternative to increase the seed germination rates under controlled chemical and physical conditions. For this, *P. guajava* seeds were subjected to chemical scarification in 10% HCl for 24 h, washed in dH_2O, disinfected in a laminar flow chamber upon immersion in 70% ethanol for 1 min, then 100 mL 3.5% NaClO supplemented with 2 drops of Tween 20 (Sigma®) for 20 min, and washed four times in sterile dH_2O. *P. guajava* seeds were inoculated in germination medium supplemented with 2.15 g L^{-1} MS basal salts (Sigma®), 10 mL^{-1} MS vitamins, 0.2 g L^{-1} myo-inositol (Sigma®), 0.1 g L^{-1} L-cysteine (Sigma®), 0.1 g L^{-1} ascorbic acid (Sigma®), 0.1 g L^{-1} L-arginine (Sigma®), 0.1 g L^{-1} L-glutamine (Sigma®), 30 g L^{-1} sucrose (Dinâmica®) and 9.0 g L^{-1} Agar-agar (Dinâmica®). The pH was adjusted to 5.8, and the medium was autoclaved for 20 min at 121°C and 121 kPa. Cultures were maintained for 60 days in photoperiod of 16/8 h light/dark at 25 ± 2°C, with 36 µmol m^{-2} s^{-1} light irradiation.

4.3.2 Cytogenetic preparation

P. guajava roots with a length of 1.0 – 2.0 cm were treated with 3, 4, 5 or 6 µM amiprophos-methyl (Sigma®) solution supplemented with 0.3% dimethyl sulfoxide (Sigma®) for a period of 3, 4, 10, 12 or 15 h at 30°C. Subsequently, the roots were fixed in 3:1 methanol:acetic acid (Sigma®) solution, with three changes of 10 min each, and stored at -20ºC for 24 h. The roots were washed three times in dH_2O for 10 min each, and then the root meristems were excised and macerated for 2 h at 36ºC in enzymatic pool containing 4% cellulase Sigma®, 0.4% hemicellulase Sigma®, 1% macerozyme Onozuka R10 Yakult and 100% pectinase Sigma®. The four proportions of enzyme pool:dH_2O tested for *P. guajava* were 1:2, 1:4, 1:6 and 1:8. Next, the roots were washed for 10 min in dH_2O, fixed once more, and stored at -20°C. From the macerated root meristems, slides were prepared from *P. guajava* roots by cellular dissociation and air-drying techniques (Marques et al. 2016). Using phase contrast microscopy BX-40 (Olympus™), the slides were chosen for FISH based on their number of metaphases with morphologically preserved chromosomes, exhibiting well-defined primary constriction and telomeres.

4.3.3 Identification and phylogenetic analysis of the LTR-RTs sequences

As *E. grandis* also belongs to the Myrtaceae family and has the sequenced genome available, this genome was used as reference to identify the LTR-RTs in *P. guajava*. Transcriptionally active LTR-RTs of the superfamilies *Copia* and *Gypsy* of the species *E. grandis* (Marcon et al. 2015) were used in a BLASTN (e-value 1e−50) against the *P. guajava* TE set (Fernandes 2020). The size and sequences of *E. grandis* used for the BLASTN were: *RLC_egAle_1* (5509 bp), *RLC_egAle_2* (5395 bp), *RLC_egMax_1* (9670 bp), *RLC_egBianca_1* (5008 bp), *RLC_egAngela_1* (11280 bp), *RLC_egAngela_2* (7473 bp), *RLC_egIvana_1* (4440 bp), *RLG_egTat_1* (18300 bp), and *RLG_egTekay_1* (12159 bp). The hits with an identity <70% were discarded, and only sequences with a greater coverage (>58%) were retrieved for further analyses. *P. guajava* putative LTR-RTs were named as *RLC_Pg_1* and *RLC_Pg_2*, which corresponded to *RLC_egBianca_1* and *RLC_egAle_2* of *E. grandis*, respectively. The sequences of these two elements were deposited at GenBank under accessions MT732385 and MT732386. To confirm the classification of *RLC_Pg_1* and *RLC_Pg_2* as *Copia* elements, the sequences were analyzed using the online tool *LTRclassifier* (Monat et al. 2016). The tool BLASTX 2.10.1+ (Altschul et al. 1997) was also used to align the *RLC_Pg_1* and *RLC_Pg_2* sequences against conserved domain databases (Marchler-Bauer et al. 2017) and to search for conserved protein domains within the LTR-RTs.

The nine LTR-RTs sequences from *E. grandis* and the two sequences identified for *P. guajava* were used to estimate a phylogenetic tree. *Ty1-copia* RT from *Abies alba* Mill. (KJ000231.1) was used as the outgroup. DNA sequences were aligned using the online version of the multiple sequence alignment software MATFFT (Katoh et al. 2019). Phylogenetic analysis of the aligned sequences was performed in the MEGA X software (Kumar et al. 2018) applying the Maximum Likelihood method and a General Time Reversible model with 1,000 bootstrap replicates.

4.3.4 Distribution of the *Copia* LTR-RTs

From the *RLC_Pg_1* and *RLC_Pg_2* sequences of *P. guajava*, *F* and *R* primers were designed to amplify these sequences and constructing the probes for genome distribution analysis through FISH. Genomic DNA of *P. guajava* was extracted using the GenElute Plant Genomic DNA Miniprep (Sigma®) according to manufacturer's instruction. DNA concentration and purity were estimated using NanoDrop® (Invitrogen) and DNA integrity was further verified by 1.5% agarose gel electrophoresis. The probes of the *Copia* LTR-RTs were generated by conventional PCR reaction using specific primers, which were designed from PrimerQuest Tool (https://www.idtdna.com). The lengths of the amplified products and the primers used for the amplification are: *RLC_Pg_1*, 685 bp (F: 5′-CTACAATGAGAGCCCACCTAAC-3′ and R: 5′-GGCAGCTACGTGAGAAAGAA-3′) and of *RLC_Pg_2*, 301 bp (F: 5′-CAAGGTACACTCTCCACTCATC-3′ and R: 5′-CTAGCCGAGCCTTTAACCTATC-3′). The reaction mix consisted of: 0.5 µM of each primer, 50 ng of genomic DNA, 200 µM of each dNTP (Promega), 1X reaction buffer (Promega), 1.8 mM $MgCl_2$ (Promega) and 1.25 U of GoTaq® DNA polymerase (Promega). The PCR amplification was: initial denaturation at 95°C for 2 min; 30 cycles of denaturation at 95°C for 1 min; annealing at 51ºC for 45 s; extension at 72°C for 1 min; and final extension at 72°C for 2 min. The probes were labeled, using the product of the first PCR as template, in a new reaction containing 200 µM each of dATP, dCTP and dGTP, 150 µM dTTP, and of 40 µM ChromaTide® Alexa Fluor® 488-5-dUTP (Life Technologies®) for *RLC_Pg_1* or of 40 µM Tetramethylrhodamine 5-dUTP (Roche®) for *RLC_Pg_2*. Reactions were checked with 1.5% agarose gel electrophoresis.

For FISH, the selected *P. guajava* slides were aged for at least two days at 30°C, washed in 1X PBS buffer for 5 min, fixed with 4% formalin for 12 min, washed again in 1X PBS for 5 min, and dehydrated in cold ethanol series (70%, 85% and 100%) for 5 min each (Silva et al. 2020). To avoid unspecific signals, 10 µg competitor DNA was used (Pedrosa et al. 2002; Silva et al. 2020). Additionally, we carried out the post-hybridization washes with about 70% stringency (Schwarzacher and Heslop-Harrison 2000). Thereby, a mix was prepared with 10 µg competitor DNA (Herring Sperm DNA, Promega), 50% formamide (Sigma®), 2X SSC (Sigma®), and 200 ng of each probe (*RLC_Pg_1* and *RLC_Pg_2*). This mix was denatured at 85°C for 5 min,

transferred to ice and immediately applied to the slides. Chromosome and probe were denatured at 68ºC for 4 min and hybridized at 37ºC for 24 h in a ThermoBrite™ system (ThermoFisher®). After this period, post-hybridization washes were carried out in 2X SSC at 58°C for 20 min. The metaphases were counterstained with 40% glycerol/PBS + DAPI (Silva et al. 2020).

Chromosome images were captured with a digital video camera 12-bit CCD (Olympus®) coupled to a fluorescence microscope (BX-60, Olympus™) equipped with a 100× objective lens, and WB (*RLC_Pg_1* probe – green fluorescence), WG (*RLC_Pg_2* probe – red florescence) and WU (DAPI staining) filters. Chromosome images were captured using an exposure time of 1/1.8 sec. The frame was digitized using the Image ProPlus 6.1 software (Media Cybernetics®). Images were edited to the same brightness and contrast in Adobe Photoshop CC. The morphometric data was also characterized according to the arm's length (long and short in µm), as described by Guerra (1986).

4.4 Results

The best metaphasic index of 45% was obtained from cytogenetic preparations with 4 µM amiprophos-methyl and 0.3% dimethyl sulfoxide at 30°C during 15 h. The enzymatic maceration with 1:2 enzyme pool:dH_2O, in combination with the cell dissociation and air-drying techniques, generated slides containing morphologically well-preserved chromosomes with well-defined primary constrictions and telomeres. In addition, these metaphases did not show cytoplasmic background noise (Fig. 1). The chromosome number of *P. guajava* was confirmed here as $2n = 2x = 22$ chromosomes and no intraspecific variation was found. From the morphometric analysis, each homologue pair was identified and the karyogram was assembled (Fig. 1).

The full-length sequences *RLC_Pg_1* and *RLC_Pg_2* have 86.2% and 71.7% identity in relation to 5' and 3' LTRs of *RLC_egBianca*_1 and *RLC_egAle_2* of *E. grandis*, respectively (classified within the Bianca and Ale lineages; Marcon et al. 2015). The identity between these sequences is also evidenced by their clustering in the phylogenetic analysis (Fig. 2). The analysis performed with *LTRclassifier* confirmed the placement of *RLC_Pg_1* and *RLC_Pg_2* within the *Copia* superfamily and provided

the location of conserved domains that are 85% significant with a minimum score of 1.0e-50. As *LTRclassifier* conserved domain identification is based solely on the Pfam database, we also used the BLASTX tool. In addition to the domains provided by *LTRclassifier*, the BLASTX analysis also evidenced the location of *Ty1/Copia* family of the RNase H domain in *RLC_Pg_1* (e-value = 1.52e-35) and *RLC_Pg_2* (e-value = 2.80e-44), which were identified by comparing the LTR_RT sequences against the conserved domain database curated by NCBI (Table 1; Fig. 3).

Table 1. Location of the conserved domains of *RLC_Pg_1* and *RLC_Pg_2* RTs.

Conserved domain	Location within the LTR-RT sequence	
	RLC_Pg_1	*RLC_Pg_2*
Primer binding site (PBS)	3119 - 3133	438 - 451
Gag-polypeptide of LTR *Copia*-type	3475 - 3729	544 - 942
GAG-pre-integrase	4345 - 4542	-
Integrase core	4591 - 4932	2218 - 2556
Reverse transcriptase	5659 - 6396	3409 - 3888
RNase H	6666 - 7079	4393 - 4806
Poly-Purine Tract (PPT)	7221 - 7236	-
Long terminal repeat (LTR)	-	-
LTR-RT total size (bp)	9247	5311

The *Copia* LTR-RTs sequences in *P. guajava* were mapped in at least 15 metaphases, and uniform hybridization signals of these LTR-RTs were obtained along the 11 chromosomes, without any preference by chromosomal region (Figs. 1).

4.5 Discussion

Using nine families of transcriptionally active LTR-RTs from *E. grandis*, we identified and characterized two sequences of LTR-RTs in *P. guajava* that

corresponded to *RLC_egBianca_1* and *RLC_egAle_2* of *E. grandis*. Considering that *P. guajava* and *E. grandis* have shared a common ancestor at ~68 Mya (Thornhill et al. 2015) and considering the phylogenetic tree (Fig. 2), *RLC_Pg_1* and *RLC_Pg_2* repetitive elements might belong to the Bianca and Ale lineages, respectively. These RTs sequences of *P. guajava* exhibited the integrase domain upstream of the reverse transcriptase. Therefore, these RTs were classified as belonging to the *Copia* superfamily. However, the long terminal repeat (LTR) regions were not identified. LTR sequences are non-coding regions, but contain signals to start and stop the transcription, besides enhancers and polyadenylation signals that are crucial to the retrotransposition process (Gao et al. 2012; Orozco-Arias et al. 2019). Thus, the absence of LTR sequences can impair the ability of the RT to self-replicate (Orozco-Arias et al. 2019; Li et al. 2019).

The loss of LTR sequences in *RLC_Pg_1* and *RLC_Pg_2* have been associated to unequal crossing-over and/or nonhomologous recombination. These mechanisms have the potential to shape the structure of such elements within a few million years (SanMiguel and Vitte 2009; Estep et al. 2013). The unequal crossing-over occurs between homologous LTRs from the same retrotransposon class and produce solo-LTRs, which are single LTR sequences surrounded by target site duplications (Devos et al. 2002). However, the main mechanism related to the presence defective LTR-RTs is nonhomologous recombination. This event produces small deletions, which create degenerated fragments of LTR-RTs that might become unrecognizable or be eliminated (SanMiguel and Vitte 2009; Estep et al. 2013). In addition, the process of DNA sequence loss from LTR-RTs is variable among different families of RTs and species. In the genome of *E. grandis*, for instance, *RLC_egBianca_1* and *RLC_egAle_2* families showed more copies of internal domains than LTRs, suggesting a predominant loss of LTRs by recombination processes and conservation of the internal region. A different profile was observed for the related species *Eucalyptus urophylla* S. T. Blake, in which the LTR-RT families exhibiting predominant internal domain conservation were *RLC_egAngela_1* and *RLG_egTekay_1* (Marcon et al. 2015).

Copia LTR-RTs exhibited a uniform distribution over the length of 11 *P. guajava* chromosomes. Because the LTR-RTs may present autonomous activity in genomes, they may display variability in their occurrence, abundance and chromosomal

distribution (Assis et al. 2020). FISH in metaphase chromosomes of *Spinacia oleracea* L., for instance, showed that *Copia* were dispersed along all chromosomes (Li et al. 2019). Differently, *Copia* were preferentially located in the centromeric and pericentromeric regions of the *A. thaliana* chromosomes (Heslop-Harrison 2000), as well as in the subterminal and telomeric regions of the all chromosomes of *E. arundinaceus* (Huang et al. 2017).

4.6 Conclusions

In conclusion, we identified and characterized two *Copia* LTR-RTs from a genomic and chromosomal point of view in *P. guajava*. The two *Copia* LTR-RTs were evenly distributed along the 11 chromosomes of *P. guajava*. These results may lay the groundwork to understand the dynamics and mechanisms that shape the karyotype of *P. guajava*.

4.7 Acknowledgments

The authors would like to thank the Conselho Nacional de Pesquisa (CNPq, Brazil), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) – Finance Code 001, for providing financial support to this study. We are also grateful to the active germplasm collection of the Universidade Federal do Espírito Santo (Alegre, Espírito Santo, Brazil) for kindly providing the *Pisidium guajava* seeds.

4.8 Author contribution

JCS and WRC designed the study. MF and MFSF carried out the bioinformatics analysis. MCS carried out the phylogenetic analysis. JCS carried out the classical and molecular cytogenetic experiments with the help of WRC. JCS, MCS and WRC wrote the manuscript. All authors read and approved the final manuscript.

4.9 Data Availability

The *RLC_Pg_1* and *RLC_Pg_2* sequences have been deposited at the GenBank database under accessions MT732385 and MT732386. These two accession numbers are referenced in the “Material and methods” section.

4.10 References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389

Assis R, Baba VY, Cintra LA, Gonçalves LSA, Rodrigues R, Vanzela ALL (2020) Genome relationships and LTR-retrotransposon diversity in three cultivated *Capsicum* L. (Solanaceae) species. BMC Genomics 21:1–14. https://doi.org/10.1186/s12864-020-6618-9

Bennetzen JL, Wang H (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol 65:505–530. https://doi.org/10.1146/annurev-arplant-050213-035811

Coser SM, Ferreira MFS, Ferreira A, Mitre LK, Carvalho CR, Clarindo WR (2012) Assessment of genetic diversity in *Psidium guajava* L. using different approaches. Sci Hortic 148:223–229. https://doi.org/10.1016/j.scienta.2012.09.030

Deng H, Xiang S, Guo Q, Jin W, Cai Z, Liang G (2019) Molecular cytogenetic analysis of genome-specific repetitive elements in *Citrus clementina* Hort. Ex Tan. and its taxonomic implications. BMC Plant Biol 19:77. https://doi.org/10.1186/s12870-019-1676-3

Devos K, Brown J, Bennetzen J (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079. https://doi.org/10.1101/gr.132102

Estep MC, DeBarry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity 110:194–204. https://doi.org/10.1038/hdy.2012.99

Fernandes M (2020) GuavaDB: o banco de dados da genômica de *Psidium guajava* L. Dissertation, Universidade Federal do Espírito Santo

Gao D, Jimenez-Lopez JC, Iwata A, Gill N, Jackson SA (2012) Functional and structural divergence of an unusual LTR retrotransposon family in plants. PLoS One 7:e48595. https://doi.org/10.1371/journal.pone.0048595

Guerra MS (1986) Reviewing the chromosome nomenclature of Levan *et al*. Rev. Bras. Rev Bras Genet 9:741–743

Gutiérrez RMP, Mitchell S, Solis RV (2008) *Psidium guajava*: a review of its traditional uses, phytochemistry and pharmacology. J Ethnopharmacol 117:1–27. https://doi.org/10.1016/j.jep.2008.01.025

Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell 12:617–635. https://doi.org/10.1105/tpc.12.5.617

Huang Y, Luo L, Hu X et al (2017) Characterization, genomic organization, abundance, and chromosomal distribution of Ty1-copia retrotransposons in *Erianthus arundinaceus*. Front Plant Sci 8:1–11. https://doi.org/10.3389/fpls.2017.00924

Izuno A, Wicker T, Hatakeyama M, et al (2019) Updated genome assembly and annotation for *Metrosideros polymorpha*, an emerging model tree species of ecological divergence. G3 Genes|Genomes|Genetics 9:3513 LP – 3520. doi: 10.1534/g3.119.400643

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of Ty1-copia-like sequences as a component of large genomes in *Pinus elliottii* var. elliottii and other gymnosperms. Proc Natl Acad Sci U S A 93:2708–2713. doi: 10.1073/pnas.93.7.2708

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Lamb JC, Meyer JM, Corcoran B, Kato A, Han F, Birchler JA (2007) Distinct chromosomal distributions of highly repetitive sequences in maize. Chromosom Res 15:33–49. https://doi.org/10.1007/s10577-006-1102-1

Li S, Guo Y, Li J et al (2019) The landscape of transposable elements and satellite DNAs in the genome of a dioecious plant spinach (*Spinacia oleracea* L.). Mob DNA 10:3. https://doi.org/10.1186/s13100-019-0147-6

Marchler-Bauer A, Bo Y, Han L et al (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129

Marcon HS, Domingues DS, Silva JC, Borges RJ, Matioli FF, Fontes MRM, Marino CL (2015) Transcriptionally active LTR retrotransposons in *Eucalyptus* genus are differentially expressed and insertionally polymorphic. BMC Plant Biol 15:198. https://doi.org/10.1186/s12870-015-0550-1

Marques AM, Tuler AC, Carvalho CR, Carrijo TT, Ferreira MFS, Clarindo WR (2016) Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). Comp Cytogenet 10:117–128. https://doi.org/10.3897/CompCytogen.v10i1.6462

Monat C, Tando N, Tranchant-Dubreuil C, Sabot F (2016) LTRclassifier: a website for fast structural LTR retrotransposons classification in plants. Mob Genet Elements 6:e1241050. https://doi.org/10.1080/2159256X.2016.1241050

Myburg AA, Grattapaglia D, Tuskan GA et al (2014) The genome of *Eucalyptus grandis*. Nature 510:356–362. https://doi.org/10.1038/nature13308

Neumann P, Novák P, Hoštáková N, Macas J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA 10:1. doi: 10.1186/s13100-018-0144-1

Orozco-Arias S, Isaza G, Guyot R (2019) Retrotransposons in plant genomes: structure, identification, and classification through bioinformatics and machine learning. Int J Mol Sci 20:3837. https://doi.org/10.3390/ijms20153837

Paz RC, Kozaczek, ME, Rosli HG, Andino NP, Sanchez-Puerta MV (2017) Diversity, distribution and dynamics of full-length Copia and Gypsy LTR retroelements in *Solanum lycopersicum*. Genetica 145: 417-430

Pearce SR, Li D, Flavell AJ, et al (1996) TheTy1-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315. doi: 10.1007/BF02174388

Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. Genetics 161:1661–1672

SanMiguel P, Vitte C (2009) The LTR-Retrotransposons of maize. In: Bennetzen JL, Hake S (eds) Handbook of maize. Springer, New York, pp 307–327

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. https://doi.org/10.1126/science.1178534

Schwarzacher T, Heslop-Harrison P (2000) Practical in situ hybridization. BIOS Scientific Publishers Ltd, Oxford

Silva JC, Soares FAF, Sattler MC, Clarindo WR (2020) Repetitive sequences and structural chromosome alterations promote intraspecific variations in *Zea mays* L. karyotype. Sci Rep 10:8866. https://doi.org/10.1038/s41598-020-65779-3

Thornhill AH, Ho SYW, Külheim C, Crisp MD (2015) Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Mol Phylogenet Evol 93:29–43. https://doi.org/10.1016/j.ympev.2015.07.007

WCSP (2020) World checklist of selected plant families. http://apps.kew.org/wcsp/. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet. Accessed 20 July 2020

Wicker T, Sabot F, Hua-Van A et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165

4.11 Figures

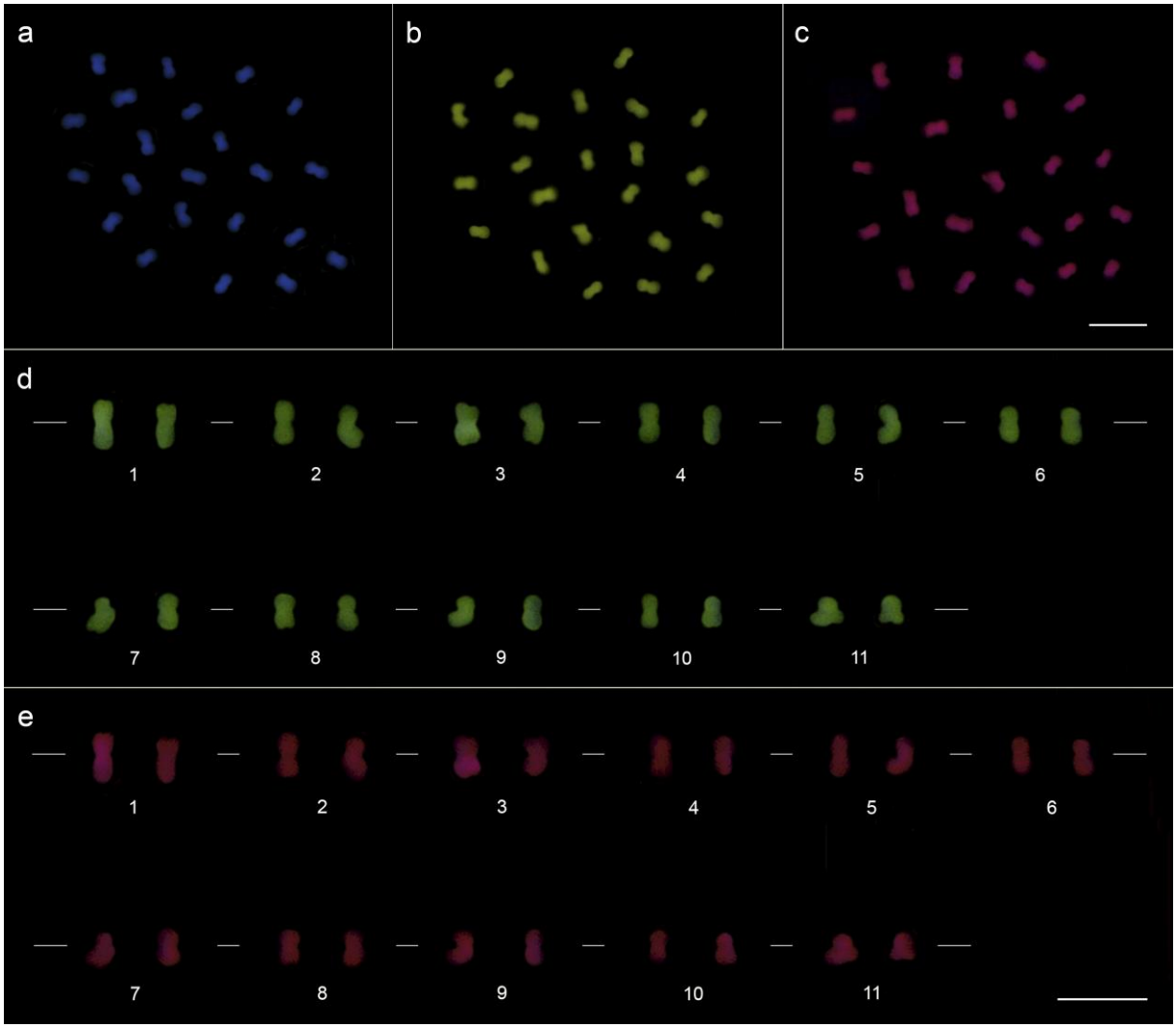


Fig. 1 – Distribution of the *Copia* superfamily LTR-retrotransposon in chromosomes of *P. guajava*. Metaphases stained with DAPI (blue) (a), showing hybridization signals of *RLC_Pg_1* (green) (b, d), and *RLC_Pg_2* (red) (c, e) probes. Note that all chromosomes were fully labeled. Bar = 5 µm.

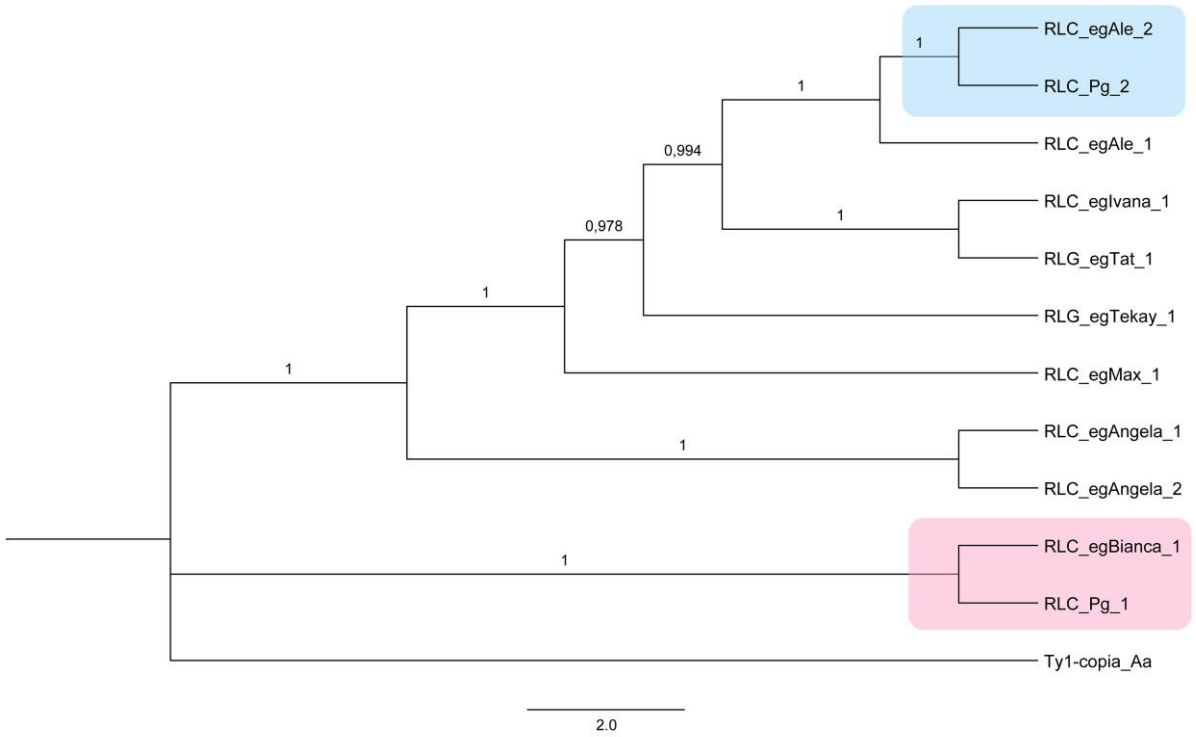


Fig 2. Phylogenetic tree with the highest log likelihood (-109327.62) obtained by using Maximum Likelihood method and General Time Reversible model. The numbers showed next to the branches indicates the proportion of trees in which the LTR-RT sequences clustered together. *Abies alba Ty1-copia* RT sequence was used as the external group.

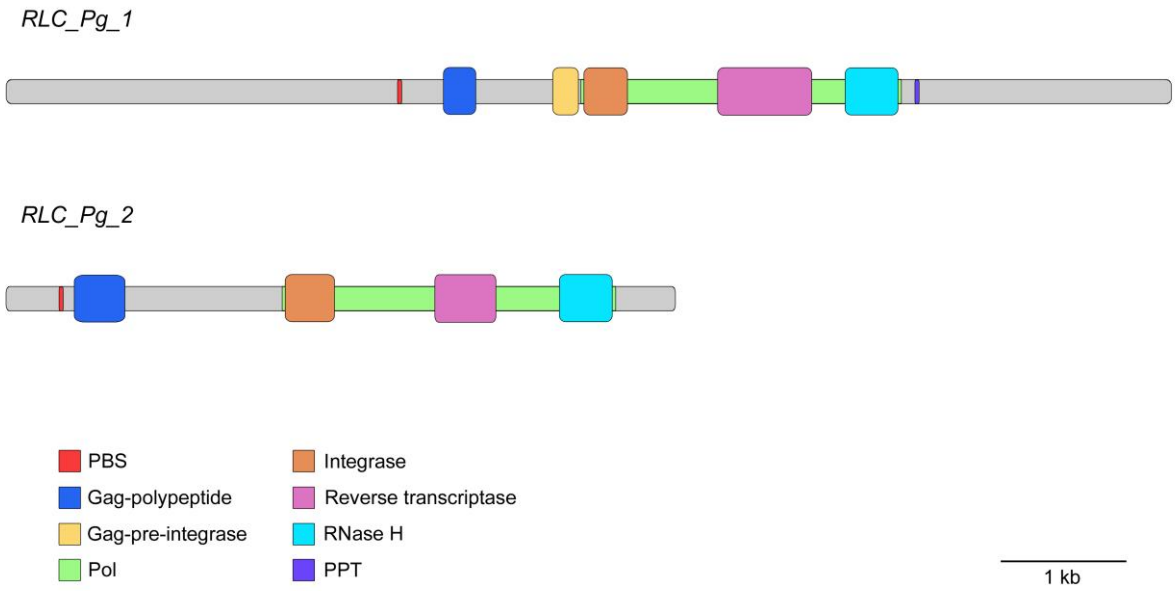


Fig 3. Structure of *RLC_Pg_1* and *RLC_Pg_2*. PBS = Primer binding site; PPT = Poly-Purine Tract.

5 OVERALL CONCLUSIONS

Cytogenomic analyses showed that repetitive sequences and chromosomal alterations are triggers of the striking divergence in nuclear DNA content observed among the studied accessions of *Z. mays*. Furthermore, we identified, characterized and classified two *P. guajava* LTR-RTs, *RLC_Pg_1* and *RLC_Pg_2*. These elements were classified within the *Copia* superfamily and were evenly distributed along the *P. guajava* chromosomes, each sequence exhibited a characteristic pattern of genomic abundance.

Considering the relevance of LTR-RTs annotation in the current scenario of comparative sequencing and genomics, we hope that these new data can lay the groundwork for understanding the role of mobile elements in the dynamics and evolution of the *P. guajava* and *Z. mays* genome.