



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

RAFAEL DE ASSIS

**RELAÇÕES CITOGENÔMICAS ENTRE ESPÉCIES DO  
GÊNERO *Capsicum* L. (SOLANACEAE)**

---

Londrina  
2019

RAFAEL DE ASSIS

**RELAÇÕES CITOGENÔMICAS ENTRE ESPÉCIES DO  
GÊNERO *Capsicum* L. (SOLANACEAE)**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

Orientador: Dr. André Luís Laforga Vanzela

Londrina  
2019

RAFAEL DE ASSIS

**RELAÇÕES CITOGENÔMICAS ENTRE ESPÉCIES DO GÊNERO**  
*Capsicum* L. (SOLANACEAE)

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

BANCA EXAMINADORA

---

Orientador: Prof. Dr. André Luís Laforga Vanzela  
Universidade Estadual de Londrina - UEL

---

Profa. Dra. Ana Lucia Dias  
Universidade Estadual de Londrina - UEL

---

Prof. Dr. Luiz Gustavo Rodrigues Souza  
Universidade Federal de Pernambuco - UFPE

Londrina, 25 de fevereiro de 2019.

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

de Assis, Rafael.

Relações citogenômicas entre espécies do gênero *Capsicum* L. (SOLANACEAE) / Rafael de Assis. - Londrina, 2019.  
81 f. : il.

Orientador: André Luís Laforga Vanzela.

Dissertação (Mestrado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, , 2019. Inclui bibliografia.

1. DNA repetitivo - Tese. 2. elementos de transposição - Tese. 3. FISH - Tese. I. Laforga Vanzela, André Luís . II. Universidade Estadual de Londrina. Centro de Ciências Biológicas. . III. Título.



PRO-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

UNIVERSIDADE ESTADUAL DE LONDRINA Centro de Ciências Biológicas

PRO-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

ATA DE DEFESA DE MESTRADO

Aos vinte e cinco dias do mês de fevereiro do ano de 2019, na sala 04 do LABESC, desta Universidade, às 14:00 horas, reuniu-se a Banca Examinadora homologada pelo Programa de Pós-Graduação, composta por André Luis Laforga Vanzela (orientador como presidente da banca), Ana Lúcia Dias e Luiz Gustavo Rodrigues Souza, na condição de membro remoto. A reunião teve por objetivo julgar o trabalho do estudante Rafael de Assis, sob o título Relações citogenômicas entre espécies do gênero Capsicum L. (Solanaceae). Os trabalhos foram abertos pelo Prof. Andre Luis Laforga Vanzela. A seguir foi dada a palavra ao estudante para apresentação do trabalho. Cada examinador arguiu o Mestrando, com tempos iguais de arguição e resposta. Terminadas as arguições, procedeu-se o julgamento do trabalho, sendo que a Dr. Luiz Gustavo Rodrigues Souza (membro remoto) enviou simultaneamente o formulário de avaliação o qual foi impresso e anexado à presente ata. A Banca Examinadora concluiu pela APROVAÇÃO do trabalho. Nada mais havendo a tratar, foi lavrada a presente ata, que vai assinada pelos demais membros da Banca Examinadora.

O(a) estudante deverá reformular seu trabalho no prazo de ..... dias: ( ) SIM ( X ) NÃO

Se houver alteração no título do trabalho, informar o novo título abaixo:

\_\_\_\_\_

OBS.: Este documento não deve conter rasuras ou corretivo e deve ser preenchido de forma legível)

Londrina, 25 de fevereiro de 2019

PRESIDENTE

Dr. André L. L. Vanzela
Vice - Coordenador
PPG - Genética Molecular
Universidade Estadual de Londrina

Instituição UEL, Londrina, PR, Brasil

TITULARES

Dr. Luiz Gustavo Rodrigues Souza

Instituição UFPE, Recife, PE, Brasil

Dra. Ana Lúcia Dias

Instituição UEL, Londrina, PR, Brasil



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
UNIVERSIDADE ANEXO II A  
ESTADUAL DE LONDRINA

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
DECLARAÇÃO DE PARTICIPAÇÃO A DISTÂNCIA EM BANCA EXAMINADORA  
DE DEFESA/QUALIFICAÇÃO DE MESTRADO OU DOUTORADO UNIVERSIDADE ESTADUAL DE  
LONDRINA, PARANÁ, BRASIL.

Declaro que no dia 25/02/2019, às 14:00 horas participei de forma síncrona remota com os demais membros que assinam a Ata física deste ato público, da banca de Defesa de Mestrado com o título **Relações citogenômicas entre espécies do gênero *Capsicum* L. (Solanaceae)** do candidato **Rafael de Assis**, do Programa de Pós-Graduação em Genética e Biologia Molecular.

Considerando o trabalho avaliado e as respostas do candidato aos questionamentos dos membros da banca examinadora, ao preencher e assinar este documento declaro minha decisão de que o candidato pode ser considerado(a): **APROVADO**

[(1) Aprovado(a); ou (2) Aprovado(a) com reformulação ; ou (3) Reprovado(a)].

Atenciosamente,

Prof. Dr. Luiz Gustavo Rodrigues Souza  
Departamento de Botânica, Universidade Federal de Pernambuco (UFPE)  
Recife, Pernambuco, Brasil

Dr. André L. L. Vanzela  
Vice - Coordenador  
PPG - Genética e Biologia Molecular  
Universidade Estadual de Londrina

## AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

A Universidade Estadual de Londrina, ao Programa de Genética e Biologia Molecular, a CAPES, CNPq, FINEP e Fundação Araucária pelo apoio financeiro.

Agradeço ao meu orientador, Prof. Dr. André Luís Laforga Vanzela, não só pela orientação neste trabalho, mas por ser exemplo de profissional, pelas conversas, pelos conselhos e pela paciência, que certamente mudaram minha maneira de pensar e agir, meu muito obrigado!

Aos colegas de laboratório, por todos os momentos que passamos e por toda a ajuda para a elaboração deste trabalho, desde os mais simples detalhes nas rotinas laboratoriais a explicações para qualquer dúvida que surgia. Muito de meu crescimento tem a ver com a convivência com vocês, muito obrigado.

Aos meus amigos de Guarapuava, por entenderem que muitas vezes não pude estar presente e mesmo com a distância estiveram sempre me apoiando e aguentando meus momentos de reclamação. Aos meus amigos de Londrina, que me acolheram e fizeram deste período um momento mais leve na vida.

Por último, mas não menos importante, agradeço aos meus pais. Minha mãe, Rosa, por ser sempre minha base e acreditar em meus sonhos junto comigo. Ao meu pai, Sergio, por estar sempre disposto a me ajudar em qualquer situação. Todo apoio e incentivo foi essencial para chegar até aqui.

1 ASSIS, Rafael de. **Relações citogenômicas entre espécies do gênero *Capsicum* L.**  
2 **(Solanaceae).** 2019. 81 f. Dissertação (Mestrado em Genética e Biologia Molecular) –  
3 Universidade Estadual de Londrina, Londrina, 2019.

## 6 RESUMO

9 Os genomas vegetais são compostos principalmente por sequências repetitivas. Elementos  
10 transponíveis (ETs), que formam a parte móvel dos genomas, são divididos nas Classes I e II,  
11 de acordo com seu mecanismo de transposição. Elementos pertencentes a Classe I, ou  
12 retrotransposons, são sintetizados e transpostos usando um RNA intermediário e podem se  
13 acumular diferencialmente dependendo do grupo vegetal. A fração repetitiva não codificante,  
14 representada pelas sequências satélite tende a se acumular em blocos ao longo dos  
15 cromossomos. As pimentas, pertencentes à família Solanaceae, possuem grande valor  
16 comercial, sendo comercializadas *in natura*, sob a forma de especiarias e são utilizadas  
17 também como ornamentação e de forma medicinal. Mesmo já possuindo seu sequenciamento  
18 publicado, trabalhos que focam na diversidade e distribuição de elementos repetitivos nos  
19 genomas das espécies de *Capsicum* ainda são superficiais. Diante disso, os objetivos deste  
20 trabalho foram compreender a organização, a distribuição e as relações genômicas e  
21 cariotípicas da fração repetitiva nos genomas de *Capsicum annuum*, *C. chinense* e *C.*  
22 *baccatum*. Para isso, foram utilizadas ferramentas de bioinformática para quantificar essa  
23 fração e métodos de citogenética molecular para a localização física dessas sequências.  
24 Sequenciamentos genômicos de alta cobertura foram contrastados com banco de dados de  
25 sequências conservadas de elementos transponíveis, elementos virais e DNA ribossômico. Os  
26 dados mostraram similaridades entre as sequências repetitivas nos genomas de *C. annuum* e  
27 *C. chinense*, em relação à *C. baccatum*. Os elementos da superfamília *Gypsy* foram mais  
28 abundantes que os *Copia*, especialmente os elementos Del, que foram mais representativos  
29 em *C. annuum* e *C. chinense*, enquanto que em *C. baccatum* houve maior acúmulo de  
30 elementos Athila. O bandeamento C-CMA/DAPI revelou diversidade de bandas entre as três  
31 espécies, contudo sempre com ocorrência de bandas terminais mais intensas CMA<sup>+</sup>, DAPI<sup>+</sup> ou  
32 CMA<sup>+</sup>/DAPI<sup>+</sup> colocalizadas. O resultado da FISH mostrou uma distribuição  
33 predominantemente dispersa dos retrotransposons, com exceção do elemento CRM que  
34 predominou na região pericentromérica, colocalizado com bandas de heterocromatina. Nossos  
35 resultados confirmaram a relação mais próxima entre *Capsicum annuum* e *C. chinense*, em  
36 relação à *C. baccatum*, concordando com a filogenia do gênero.

37 **Palavras-chave:** *Capsicu*. DNA repetitivo. elementos de transposição. FISH. Genomas.

1 ASSIS, Rafael de. **Cytogenomics relationships among species of the genus *Capsicum* L.**  
2 **(Solanaceae).** 2019. 81 p. Dissertation (Master's in Genetics and Molecular Biology) –  
3 Universidade Estadual de Londrina, Londrina. 2019.  
4  
5

## 6 ABSTRACT 7 8

9 Plant genomes have large quantities of repetitive sequences. Transposable elements (ETs),  
10 which form the moving part of the genomes, are divided into Classes I and II, according to  
11 their transposition mechanisms. Elements of the Class I, or retrotransposons, are synthesized  
12 and transposed by an intermediate RNA, and they may have differential accumulation  
13 depending on the plant group. The non-coding repetitive fraction, represented by the satellite  
14 sequences, tends to accumulate in blocks along the chromosomes. *Capsicum* species belong to  
15 the family Solanaceae, and they have great commercial value as spice and for fresh  
16 consumption. Although already having its genomes sequenced, works that focus on the  
17 diversity and distribution of repetitive sequences in the genomes are still superficial. The  
18 objectives of this work were to understand the organization, distribution and  
19 genomic/karyotypic relations among *Capsicum annuum*, *C. chinense* and *C. baccatum*, based  
20 on the repetitive DNA fraction. For this, bioinformatics tools were used to quantify this  
21 fraction, and the molecular cytogenetic methods were employed for their physical location.  
22 High-coverage genomic sequencing was contrasted against a database of conserved sequences  
23 of transposable elements, viral elements and ribosomal DNA. The data showed similarities  
24 between the repetitive sequences in the genomes of *C. annuum* and *C. chinense*, in relation to  
25 *C. baccatum*. The elements of the *Gypsy* superfamily were more abundant than the *Copia*,  
26 especially elements of Del family, which were more representative in *C. annuum* and *C.*  
27 *chinense*, whereas in *C. baccatum* there was a greater accumulation of Athila family elements.  
28 The technique of C-banding revealed band diversity among the three species, however, with  
29 the occurrence of more intense CMA<sup>+</sup>, DAPI<sup>+</sup> or CMA<sup>+</sup>/DAPI<sup>+</sup> terminal bands colocalized.  
30 The FISH results showed a predominantly dispersed distribution of retrotransposons, with the  
31 exception of the CRM family elements that predominated in the proximal region, colocalized  
32 with heterochromatin bands. Our results confirmed the closer relationship between *C. annuum*  
33 and *C. chinense*, in relation to *C. baccatum*, agreeing with the phylogeny.  
34

35 **Keywords:** *Capsicum*. FISH. Genomes. Repetitive DNA. Transposable elements.

## LISTA DE ILUSTRAÇÕES

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

- Figura 1** - Representantes dos elementos de transposição de Classe I e Classe II. GAG – Proteína capsidial. PR – Protease. INT – Integrase. RH – RNase H. CHR – Cromodomínio. RT – Transcriptase reversa. RPA – replication protein A. MP – movement protein. HEL – helicase. Note a diferença na posição da integrase na cadeia poligênica de Copia e Gypsy. ERVs não possuem LTR, contudo, são relacionados filogeneticamente com os elementos com LTR. .... 18
- Figura 2** - Distribuição relativa (%) dos elementos repetitivos nos três *datasets* de *Capsicum*. **(A)** Distribuição relativa (%) das classes de elementos repetitivos. Note que os elementos de Classe I possuem maior proporção em relação aos demais. **(B)** Distribuição relativa (%) das superfamílias de elementos transponíveis, microsatélites, DNAr 5S e DNAr 35S. Note, maior acúmulo de elementos da superfamília *Gypsy* em relação aos demais. **(C)** Distribuição relativa (%) dos retrotransposons com LTR. Observe maior acúmulo de elementos da família Del nos genomas de *C. annuum* e *C. chinense* e maior acúmulo do clado Athila/Tat em *C. baccatum*..... 50
- Figura 3** - Dotter representando sequências com LTRs em dois retrotransposons do clado Athila/Tat, extraídas do *dataset* de *C. annuum*. Note a diferença entre os tamanhos das LTRs (ponta de seta) nas duas sequências. **(A)** Athila\_annuum\_1 **(B)** Athila\_annuum\_2..... 56
- Figura 4** - Alinhamento das sequências de retrotransposons com a ferramenta Mauve. São representadas as sequências recuperadas de Del a partir do *dataset* de genomas de *Capsicum*. **(A)** Del\_annuum\_2. **(B)** Del\_baccatum\_1. **(C)** Del\_baccatum\_2. **(D)** Del\_baccatum\_3. **(E)** Del\_baccatum\_5. **(F)** Del\_annuum\_1. **(G)** Del\_chinense\_1. **(H)** Del\_annuum\_3. **(I)** Del\_baccatum\_4. **(J)** Del\_baccatum\_6 ..... 57

|    |   |    |
|----|---|----|
| 1  | <b>Figura 5</b> - Alinhamento das sequências de retrotransposons com a ferramenta Mauve. São representadas as sequências recuperadas de CRM a partir do <i>dataset</i> de genomas de <i>Capsicum</i> . (A) CRM_annuum_1. (B) CRM_annuu_2. (C) CRM_annuum_4. (D) CRM_chinense_2. (E) CRM_baccatum_2. (F) CRM_annuum_3. (G) CRM_chinense_1. (H) CRM_baccatum_1.....   | 58 |
| 7  | <b>Figura 6</b> - Alinhamento das sequências de retrotransposons com a ferramenta Mauve. São representadas as sequências recuperadas de Athila a partir do <i>dataset</i> de genomas de <i>Capsicum</i> . (A) Athila_annuum_1. (B) Athila_annuum_2.....   | 59 |
| 11 | <b>Figura 7</b> - Representação dos grupos de elementos de transposição. LTR – <i>long terminal repeat</i> , GAG – proteína capsidial, RT – <i>reverse transcriptase</i> , RH – <i>RNAse H</i> , INT – <i>integrase</i> , ASP - aspartase. Note a diferença na extensão em pares de base dos grupos formados, tamanho das LTRs e da posição das proteínas ao longo dos elementos evidenciando o agrupamento dentro da superfamília <i>Gypsy</i> . Tamanho das integrases faz referência ao tamanho do alinhamento obtido, tal variação não remonta ao tamanho real do gene, assim como nos demais domínios. ....  | 59 |
| 19 | <b>Figura 8</b> - Bandeamento C – CMA/DAPI em metáfases de espécies do gênero <i>Capsicum</i> . (A, B) <i>C. annuum</i> . (A) Pontas de seta indicam bancas terminais DAPI <sup>+</sup> . (B) Pontas de seta indicam bandas terminais intensas CMA <sup>+</sup> , setas indicam bandas intersticiais CMA <sup>+</sup> . (C, D) <i>C. chinense</i> (C) Pontas de setas indicam bandas terminais intensas DAPI <sup>+</sup> . (D) Pontas de setas indicam bandas terminais intensas CMA <sup>+</sup> . (E, F) <i>C. baccatum</i> . (E) Pontas de setas indicam bandas terminais e setas indicam bandas intersticiais DAPI <sup>+</sup> . (F) Pontas de seta indicam bandas terminais intensas e setas indicam bandas pericentroméricas..... | 61 |
| 28 | <b>Figura 9</b> - Hibridização <i>in situ</i> fluorescente com metáfases coradas com DAPI (azul), sonda de Del detectada com avidina-FITC (verde), em três espécies de <i>Capsicum</i> . (A-D) <i>C. annuum</i> com sinais dispersos apresentando maior concentração de sinais em relação às demais espécies. (E-G) <i>C. chinense</i> com sinais dispersos (H-J) <i>C. baccatum</i> apresentando sinais menos intensos. ....   | 63 |

|    |  |  |
|----|--|--|
| 1  | <b>Figura 10</b> - Hibridização <i>in situ</i> fluorescente com metáfases coradas com DAPI           |  |
| 2  | (azul), sonda de CRM detectada com avidina-FITC (verde), em três                                     |  |
| 3  | espécies de <i>Capsicum</i> . <b>(A e B)</b> <i>C. annuum</i> com sinais concentrados na             |  |
| 4  | região pericentromérica. <b>(C)</b> Detalhe do sinal da sonda de CRM                                 |  |
| 5  | evidenciando a localização pericentromérica do elemento de   |  |
| 6  | transposição. <b>(D-F)</b> <i>C. chinense</i> . <b>(G-I)</b> <i>C. baccatum</i> . ..... 65           |  |
| 7  | <b>Figura 11</b> - Hibridização <i>in situ</i> fluorescente com metáfases e prometáfases coradas     |  |
| 8  | com DAPI (azul), sonda de Oryco detectada com Cy3-dUTP   |  |
| 9  | (vermelho), em três espécies de <i>Capsicum</i> . <b>(A)</b> <i>C. annuum</i> . <b>(B)</b> <i>C.</i> |  |
| 10 | <i>chinense</i> . <b>(C)</b> <i>C. baccatum</i> . ..... 66   |  |
| 11 |  |  |

## LISTA DE TABELAS

1  
2  
3  
4  
5  
6  
7  
8

|  |    |
|--|----|
| <b>Tabela 1</b> - Descrição geral dos arquivos com os <i>scaffolds</i> das espécies de <i>Capsicum</i> . ....  | 37 |
| <b>Tabela 2</b> - Frequência e valores relativos dos elementos repetitivos nos três genomas<br>de <i>Capsicum</i> . ....   | 45 |
| <b>Tabela 3</b> - Descrição geral dos <i>primers</i> desenhados para regiões conservadas da<br>transcriptase reversa de retrotransposons de <i>Capsicum</i> . .... | 53 |

## SUMÁRIO

|          |   |    |
|----------|---|----|
| <b>1</b> | <b>INTRODUÇÃO</b> .....   | 13 |
| <b>2</b> | <b>FUNDAMENTAÇÃO TEÓRICA</b> .....                                    | 15 |
| 2.1      | <i>Valor C de DNA e elementos móveis</i> .....                        | 15 |
| 2.2      | <i>Sequências repetitivas não codificantes</i> .....                  | 19 |
| 2.3      | <i>Diversidade genômica e cariotípica na família Solanaceae</i> ..... | 20 |
| 2.4      | <i>Bioinformática e avanços na genômica de Capsicum</i> .....         | 21 |
| <b>3</b> | <b>OBJETIVOS</b> .....  | 23 |
| 3.1      | <i>Objetivo geral</i> .....   | 23 |
| 3.2      | <i>Objetivos específicos</i> .....                                    | 23 |
| <b>4</b> | <b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....                               | 24 |
| <b>5</b> | <b>ARTIGO</b> .....   | 32 |
|          | <b>Resumo</b> .....   | 32 |
|          | <b>Introdução</b> .....   | 33 |
|          | <b>Materiais e métodos</b> .....                                      | 37 |
|          | <b>Resultados</b> .....   | 43 |
|          | <b>Discussão</b> .....  | 67 |
|          | <b>Agradecimentos</b> .....   | 74 |
|          | <b>Referências</b> .....  | 74 |
| <b>6</b> | <b>CONSIDERAÇÕES FINAIS</b> .....                                     | 81 |

## 1. INTRODUÇÃO

Os genomas vegetais são compostos por uma fração de DNA não repetitiva (uma ou poucas cópias) e outra repetitiva (centenas ou até milhares de cópias). Essa última pode ser organizada em grupos variados de acordo com a natureza e o modo de repetição das sequências (HESLOP-HARRISON E SCHWARZACHER, 2011; BENNETZEN E WANG, 2014). A fração repetitiva pode alcançar até 90% nos genomas de algumas angiospermas (AMBROZOVA *et al.*, 2011), em *Capsicum* essa fração representa de 76% a 79% dos genomas (KIM *et al.*, 2014). De maneira geral, os repetitivos podem ser encontrados dispersos (elementos de transposição, TEs = *transposable elements*) ou organizados em *tandem* (DNA ribossômico, microsatélites, minissatélites, satélites, sequências teloméricas e centroméricas e TEs), e podem ser classificados de acordo com o tamanho dos motivos repetidos, composição de bases e localização nos cromossomos.

O gênero *Capsicum* (Solanaceae) exibe grande variedade de formas, tamanhos e cores de frutos (QIN *et al.*, 2014), além da pungência característica devida ao acúmulo dos alcalóides capsaicinóides, que garantem grande importância econômica para o grupo (AZA-GONZÁLEZ *et al.*, 2011; KIM *et al.*, 2014). O gênero possui 43 espécies (THE PLANT LIST, 2013), das quais cinco são consideradas domesticadas (*C. annuum* L., *C. baccatum* L., *C. chinense* Jacq., *C. frutescens* L. e *C. pubescens* R E P) e as demais silvestres (MOSCONE *et al.*, 2006; CARRIZO- GARCÍA *et al.*, 2013). Os estudos filogenéticos sugerem que *Capsicum annuum*, *C. chinense* e *C. frutescens* pertencem ao clado Annuum, *C. baccatum* ao clado Baccatum e *C. pubescens* ao clado Pubescens (CARRIZO-GARCÍA *et al.*, 2016). No gênero *Capsicum*, predominam os números haploides  $n= 12$  e  $13$ , com cariótipos simétricos e cromossomos meta- e submetacêntricos (MOSCONE *et al.*, 1993; 2007; SCALDAFERRO *et al.*, 2016).

Até o momento, foram sequenciados os genomas completos de três espécies (*C. annuum*, *C. chinense* e *C. baccatum*), todos tendo como referência o genoma de *C. annuum* (KIM *et al.*, 2014; QIN *et al.*, 2014). Tendo em vista que as descrições relativas à fração repetitiva desses genomas são muito superficiais e da livre disponibilidade dos genomas completamente sequenciados e montados, o foco desse estudo foi identificar, classificar e realizar uma ampla comparação citogenômica da fração repetitiva de DNA nessas três espécies, procurando compreender se a diversidade e a distribuição dos retrotransposons repetitivos seguem um padrão de distribuição cromossômica e acompanham a organização

35 dos clados propostos na filogenia do gênero, em relação à *C. annuum*, *C. chinense* e *C.*  
36 *baccatum*.

## 37 2. FUNDAMENTAÇÃO TEÓRICA

38

### 39 2.1. Valor C de DNA e elementos móveis

40 Os genomas vegetais variam em número e tamanho cromossômico, assim  
41 como em conteúdo de DNA nuclear. Em angiospermas, variação nos valores C de DNA  
42 alcançam cerca de 2.000 vezes, com os menores genomas tendo em torno de 60 Mb em  
43 espécies carnívoras do gênero *Genlisea* Lentibulariaceae, e os maiores com cerca de 150.000  
44 Mb em espécies de *Paris*, Melanthiaceae (HESLOP-HARRISON E SCHMIDT, 2007;  
45 PELLICER, FAY E LEITCH, 2010). Entre os mecanismos que levam ao aumento no  
46 tamanho dos genomas, pode-se destacar, por exemplo, a autopoliploidia (referente à  
47 duplicação do próprio genoma), a aloploidia (referente à duplicação de genomas híbridos)  
48 e os rearranjos estruturais, como deleções e ampliações de famílias repetitivas de DNA  
49 (BENNETZEN, 2002, HESLOP-HARRISON E SCHMIDT, 2007; BENNETZEN E WANG  
50 2014).

51 O DNA nuclear nos vegetais é composto por uma fração não repetitiva (uma  
52 ou poucas cópias de sequências codificantes, íntrons, promotores e sequências de DNA  
53 regulatórias) e outra fração repetitiva, com centenas ou até milhares de cópias. Essa última  
54 fração pode ser organizada em grupos variados de acordo com a natureza e o modo de  
55 repetição das sequências (HESLOP-HARRISON E SCHMIDT, 1998; HESLOP-HARRISON  
56 E SCHWARZACHER, 2011; BENNETZEN E WANG, 2014) e chegam a representar, por  
57 exemplo, 50 a 60% dos genomas de *Solanum tuberosum* e *S. lycopersicum*, respectivamente  
58 (TOMATO GENOME CONSORTIUM, 2012; MEHRA, GANGWAR, SHANKAR, 2015).  
59 Em um contexto geral, a fração repetitiva pode ser pouco representativa em alguns genomas,  
60 como 3% em *Utricularia gibba*, Lentibulariaceae (IBARRA-LACLETTE *et al.*, 2013), ou  
61 bem mais representativa, como 76,4% em *Capsicum annuum* e 79,6% em *C. chinense* (KIM  
62 *et al.*, 2014), ou até uma fração ainda maior, como 90% do genoma de algumas espécies do  
63 gênero *Fritillaria* (AMBROZOVA *et al.*, 2011).

64 A fração repetitiva pode ser encontrada dispersa nos genomas (elementos de  
65 transposição, TEs = *transposable elements*) ou organizadas em *tandem* (DNA ribossômico,  
66 microssatélites, minissatélites, satélites, sequências teloméricas e centroméricas e TEs). A  
67 fração repetitiva codificante engloba, por exemplo, elementos de transposição, sequências  
68 medianamente repetidas de DNAr 35S e 5S, genes que codificam para histonas e proteínas do  
69 citoesqueleto (KUBIS, SCHMIDT E HESLOP-HARRISON, 1998; CONTENTO HESLOP-  
70 HARRISON E SCHWARZACHER, 2005). A fração repetitiva não codificante pode ser

71 encontrada em *tandem*, representando sequências menores como os microssatélites com 1-6  
72 pb, minissatélites com 10-30 pb e DNAsat com mais de 30 pb (GARRIDO-RAMOS, 2017).

73 Os elementos transponíveis são divididos em Classes I e II (Figura 1), sendo  
74 os de Classe I (também chamados retrotransposons) os mais comuns nos vegetais. A  
75 transposição ocorre pelo mecanismo de “cópia e cola”, onde o RNAm transcrito é convertido  
76 em DNA complementar, por meio da enzima transcriptase reversa, e integrado em outra  
77 região do genoma por meio da enzima integrase (KAZAZIAN, 2004; WICKER *et al.*, 2007;  
78 LISCH, 2013). Os retrotransposons podem ser classificados de acordo com a presença ou  
79 ausência de LTR (repetições terminais longas), e subdivididos em quatro superfamílias, sendo  
80 *Copia* e *Gypsy* as mais comuns em plantas. Essa classificação é baseada nas diferenças  
81 observadas na ordem dos genes da cadeia poligênica, principalmente em relação à posição da  
82 integrase (Figura 1) (XIONG E EICKEBUSH, 1990; BENNETZEN, 2000). Retrotransposons  
83 sem LTR incluem os elementos LINE, SINE e DIRS. Elementos SINE podem variar de 80-  
84 500pb (WICKER *et al.*, 2007), os quais representam cerca de 0,3% do genoma de *Solanum*  
85 *tuberosum* (SEIBT *et al.*, 2016). Elementos LINE possuem grande diversidade entre os  
86 eucariotos e são predominantes em muitos animais, em comparação aos elementos LTR  
87 (WICKER *et al.*, 2007; PLATT, VANDEWEGE E RAY, 2018)

88 Das famílias de vírus que ocorrem em eucariotos, cinco são as que replicam  
89 o genoma via transcriptase reversa (PRINGLE, 1998). Entre elas destaca-se a família  
90 Caulimoviridae, por ser a única com genoma de DNA a infectar plantas  
91 (<https://talk.ictvonline.org/>). Essas famílias são similares a outros retrovírus e  
92 retrotransposons (ROTHNIE CHAPDELAIN E HOHN, 1994; HAAS *et al.*, 2002), estando  
93 relacionadas evolutivamente com os elementos com LTR. Os vírus da família Caulimoviridae  
94 são encontrados geralmente fragmentados (não autônomos) nos genomas vegetais  
95 (BOUSALEM, DOUZERY E SEAL, 2008; DIOP, 2018; GONG E HAN, 2018). Os ERVs  
96 (retrovírus endógenos) referem-se à retroelementos que no passado evolutivo infectaram  
97 células de diferentes linhagens e, desse modo, passaram a ser herdados como provírus por  
98 meio de herança mendeliana (FAUQUET *et al.*, 2005). A integração desses vírus endógenos e  
99 sua evolução junto ao genoma hospedeiro ocorreram provavelmente na forma de  
100 pseudogenes, acumulando mutações neutras (STOYE, 2001). Contudo, há evidências que  
101 algumas sequências são capazes de iniciar uma infecção frente a certas condições, como  
102 observados com os vírus *Banana streak OL* (BSOLV) (NDOWORA *et al.*, 1999), *Banana*  
103 *streak GF* (BSGFV) (GAYRAL *et al.*, 2008), *Petunia vein clearing* (PVCV) (RICHERT-  
104 PÖGGELER *et al.*, 2003) e *Tobacco vein clearing* (TVCV) (LOCKHART *et al.*, 2000). O

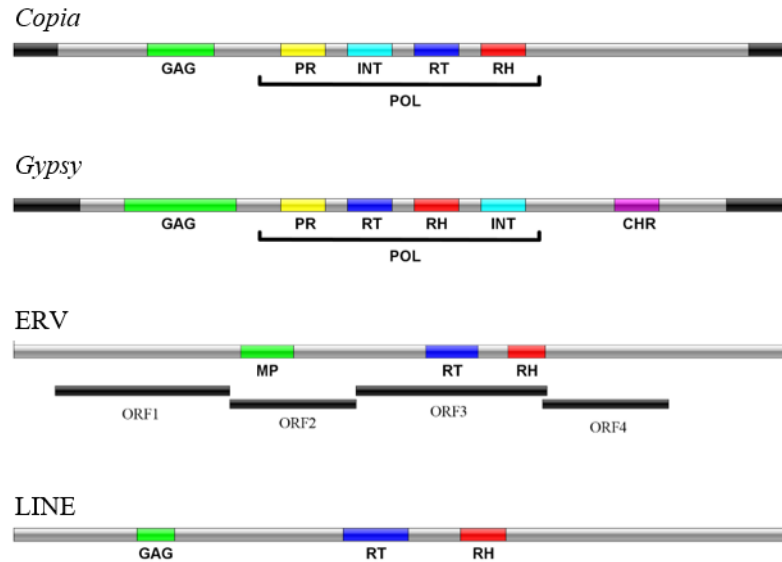
105 Comitê Internacional de Taxonomia de Vírus (sigla em inglês ICTV) classifica a família  
106 Caulimoviridae em oito gêneros, *Badnavirus*, *Caulimovirus*, *Cavemovirus*, *Petuvirus*,  
107 *Rosadnavirus*, *Solendovirus*, *Soymovirus* e *Tungrovirus* (ICTV, 2017).

108 Elementos de Classe II (Transposons) são transpostos por meio de um  
109 mecanismo de “corta e cola”, no qual o elemento é excisado e reintegrado em outra região do  
110 genoma utilizando a enzima transposase. Muitos transposons de DNA são flanqueados por  
111 sequências terminais repetidas invertidas (TIR – *terminal inverted repeat*), a transposase  
112 reconhece essas TIRs, permitindo assim a excisão do transposon e sua reintegração em outra  
113 região do genoma (MUÑOZ-LÓPEZ E GARCÍA-PÉREZ, 2010; LEVIN E MORAN, 2011).  
114 O baixo acúmulo de elementos de classe II, quando comparados aos LTR-RTs, se deve ao  
115 fato que o elemento é excisado de uma região e reintegrado em outra, enquanto os LTR-RTs  
116 são copiados e sua cópia é reintegrada.

117 Há também dois outros grupos que utilizam mecanismos diferentes de  
118 movimentação e acumulação como os Helitrons, que se espalham pelo genoma por “círculo  
119 rolante” via helicase, e os Polintons, que se transpõem por auto síntese, mediado pela  
120 polimerase B (KAPITONOV E JURKA, 2008; LISCH, 2009; LISCH, 2013). O sucesso da  
121 integração é alcançado em função do mecanismo de reparo na dupla fita de DNA, após a  
122 excisão do fragmento. Elementos pertencentes a essa classe diferenciam-se dos da Classe I  
123 pela ausência de um RNA intermediário (WICKER *et al.*, 2007; LISCH, 2013). Em geral, os  
124 elementos de Classe II se acumulam pouco nos genomas vegetais, contudo, podem ser  
125 responsáveis por diversas mutações (BENNETZEN E WANG, 2014).

126

## Class I



## Class II



127

128 **Figura 1:** Representantes dos elementos de transposição de Classe I e Classe II. GAG – Proteína  
 129 capsidial. PR – Protease. INT – Integrase. RH – RNase H. CHR – Cromodomínio. RT – Transcriptase  
 130 reversa. RPA – *replication protein A*. MP – *moviment protein*. HEL – helicase. Note a diferença na  
 131 posição da integrase na cadeia poligênica de *Copia* e *Gypsy*. ERVs não possuem LTR, contudo, são  
 132 relacionados filogeneticamente com os elementos com LTR.

133

## 2.2. Sequências repetitivas não codificantes

A fração repetitiva dos genomas vegetais, que ocorre organizada em *tandem*, pode ser classificada de acordo com o tamanho dos motivos que são repetidos, composição de bases e localização nos cromossomos. O DNA satélite corresponde à fração repetitiva não codificante com mais de 30pb e são considerados os principais componentes da heterocromatina, predominantes nas regiões pericentroméricas, centroméricas e subteloméricas dos cromossomos (GARRIDO-RAMOS, 2015, 2017). Os DNAsat foram descritos inicialmente como bandas adicionais de DNA após centrifugação do DNA em gradiente de densidade (HEMLEBEN, 1990). Desde então, *repeats* em *tandem* de diferentes composições, tamanhos e abundâncias vêm sendo identificados (GARRIDO-RAMOS, 2015).

Eventos de *crossing-over* desigual são importantes para a formação de satélites e para o aumento na variabilidade dos monômeros, contudo, o *crossing-over* por si só não é o único mecanismo que leva a formação e expansão dos DNA satélites. Amplificações gênicas ou de pedaços de DNA repetitivo codificante, seguido de duplicações e deleções também desempenham papel importante para o surgimento e evolução dessas sequências (SMITH, 1976; MEHROTRA E GOYAL, 2014). A região centromérica, por exemplo, é extremamente importante para a migração dos cromossomos durante as divisões celulares e pode conter TEs e DNAsat, além disso, muitos desses *repeats* podem ser cromossomo ou genoma específicos (PLOHL, MEŠTROVIĆ E MRAVINAC, 2012), como observado em cromossomos humanos (WILLARD E WAYE, 1987; RUDD E WILLARD, 2004). Em plantas, por exemplo, a variedade de TEs centroméricos foi reportada em diversas espécies e classificada em três grupos distintos quando à estrutura desses elementos (NEUMANN *et al.*, 2011). Mais recentemente em espécies diploides e tetraploides de *Coffea*, foram descritos dez grupos de retrotransposons centroméricos, associados ou não à heterocromatina (DE CASTRO NUNES *et al.*, 2018), evidenciando que esses elementos residentes nos genomas são extremamente variáveis, mesmo em espécies filogeneticamente próximas.

Diferentemente dos DNA satélite, os microssatélites e minissatélites são classificados de acordo com o número mínimo de bases ou número mínimo de *repeats*, contudo, tal diferença pode não ser tão clara, porque nem sempre o tamanho do motivo é suficiente para classificá-los (GARRIDO-RAMOS, 2017). Os principais tipos de microssatélites em plantas ocorrem na forma de mono- di- tri ou até heptanucleotídeos (KUBIS, SCHMIDT E HESLOP-HARRISON, 1998; ELLEGREN, 2004, GARRIDO-RAMOS, 2017). O mecanismo de mutação mais aceito para o aumento e diversificação de microssatélites e minissatélites é o *slippage*, que ocorre principalmente durante a replicação,

168 decorrente de um “escorrego” da polimerase em função da homologia nos *repeats*. Nesse  
169 caso, fitas de DNA podem se anelar de maneira irregular, produzindo uma alça levando ao  
170 reparo seguido do aumento ou diminuição dessas sequências (ELLEGRÉN, 2004).

171

### 172 2.3. Diversidade genômica e cariotípica na família Solanaceae

173 Os centros de origem propostos para *Capsicum* são as Américas Central e  
174 do Sul (DJIAN-CAPORALINO *et al.*, 2007). O gênero *Capsicum* abrange os pimentões e as  
175 pimentas, amplamente cultivados nas regiões tropicais e subtropicais, apresentando grande  
176 variedade de formas, tamanhos e cores de frutos (QIN *et al.*, 2014). A pungência  
177 característica é devida ao acúmulo de um grupo de alcalóides, os capsaicinóides, o que  
178 confere importância econômica ao grupo (AZA-GONZÁLEZ, NÚÑEZ-PALENIUS E  
179 OCHOA-ALEJO., 2011; KIM *et al.*, 2014). O gênero também é muito utilizado na produção  
180 de compostos medicinais, na ornamentação e para consumo, devido ao grande acúmulo de  
181 nutrientes que os frutos apresentam (MOSCONE *et al.*, 2003; 2007). Das 43 espécies descritas  
182 para o gênero (THE PLANT LIST. 2013), cinco são consideradas domesticadas (*C. annuum*  
183 L., *C. baccatum* L., *C. chinense* Jacq., *C. frutescens* L. e *C. pubescens* R E P) e as demais  
184 silvestres (MOSCONE *et al.*, 2007; CARRIZO GARCÍA *et al.*, 2013). *Capsicum annuum*, *C.*  
185 *chinense* e *C. frutescens* pertencem ao clado Annum, *C. baccatum* ao clado Baccatum e *C.*  
186 *pubescens* ao clado Pubescens (CARRIZO GARCÍA *et al.*, 2016).

187 As descrições citogenéticas presentes na literatura acerca da família  
188 Solanaceae sugerem a predominância de cariótipos com  $2n = 24$ , e variações quanto a esse  
189 valor podem estar relacionadas com eventos de poliploidia, aneuploidia e disploidia (BADR  
190 *et al.*, 1997; MOSCONE *et al.*, 2007; PADILHA *et al.*, 2016). Algumas espécies pertencentes  
191 ao gênero *Solanum* e *Lycianthes*, por exemplo, apresentam  $2n = 24$  e cariótipos simétricos,  
192 onde há predominância de cromossomos meta- e submetacêntricos (ACOSTA *et al.*, 2005;  
193 ACOSTA, GUERRA E MOSCONE, 2012). O gênero *Solanum* possui ainda espécies  
194 poliploides que podem apresentar até  $2n = 8x = 96$  (HUNZIKER, 2001). No gênero  
195 *Capsicum*, predominam os números haploides 12 e 13 e cariótipos simétricos com  
196 cromossomos meta- e submetacêntricos e, de modo geral, cariótipos com  $n=12$  são compostos  
197 por 11 pares metacêntrico e 1 par submetacêntrico, sendo essa característica mais observada  
198 nas espécies domesticadas (MOSCONE *et al.* 1993; 2007; SCALDAFERRO, GABRIELE E  
199 MOSCONE, 2013; SCALDAFERRO *et al.*, 2016).

200 Com relação às características citogenômicas, o gênero *Capsicum* apresenta  
201 variação quanto a quantidade de DNA nuclear entre as espécies. Os valores C de DNA foram

202 estimados por MOSCONE *et al.* (2003) em 25 amostras de nove espécies diploides de  
203 *Capsicum*, os quais variaram de 3,32 pg em *C. annuum* até 5,77 pg em *C. parvifolium*. Do  
204 mesmo modo que são observadas variações no valor C de DNA entre as espécies, os perfis de  
205 bandas de heterocromatina também são variáveis. Os primeiros trabalhos de bandeamento  
206 envolvendo espécies do gênero *Capsicum*, demonstraram diversidade quanto a presença de  
207 bandas terminais (bandas mais intensas ou tênues) e de bandas intersticiais e proximais  
208 (MOSCONE *et al.*, 1993; 1996), assim como em trabalhos mais recentes (SCALDAFERRO  
209 *et al.*, 2013; MARTINS *et al.*, 2018). A quantidade e localização de sítios de DNAr também é  
210 uma característica que apresenta variação nas espécies do gênero *Capsicum*,  
211 SCALDAFERRO *et al.* (2016) evidenciaram a posição de DNAr por meio de hibridizações  
212 com sondas de 35S e 5S. Em seu trabalho, puderam mostrar que *C. baccatum* possui maior  
213 quantidade de sítios de hibridização de 35S, enquanto *C. annuum* e *C. chinense* a quantidade  
214 foi menor. A literatura também evidencia a presença de poucos sítios, muitas vezes únicos, de  
215 DNAr 5S em espécies de *Capsicum* (PARK *et al.*, 1999a; PARK *et al.*, 1999b;  
216 SCALDAFERRO, GABRIELE E MOSCONE, 2013; AGUILERA, DEBAT E GRABIELE,  
217 2017).

218

#### 219 2.4. Bioinformática e avanços na genômica de *Capsicum*

220 O sequenciamento genômico baseado na técnica de Sanger, em 1977,  
221 marcou o início de uma nova era na genética. Avanços, principalmente na automatização de  
222 equipamentos de sequenciamento e nas análises de segmentos de DNA, permitiram a  
223 obtenção de um maior volume de dados, alcançando diversos organismos, modelos ou não.  
224 Posteriormente ao método de Sanger, outras metodologias foram desenvolvidas, como a  
225 plataforma Illumina (que gera *reads* de até 600 bases) e o PacBio com tamanho médio dos  
226 *reads* passando os 10 Kb de extensão (RHOADS E AU, 2015), que melhoraram a velocidade  
227 de geração de informações a custos mais baixos. Contudo, a integração de diferentes  
228 estratégias de sequenciamento ajudou a melhorar a cobertura e a qualidade no conjunto de  
229 dados, assim como a fidelidade das sequências montadas (SCHATZ, WITKIWSKI E  
230 MCCOMBIE, 2012; SIMS *et al.*, 2014). O primeiro sequenciamento genômico em plantas foi  
231 o da *Arabidopsis thaliana*, publicado em 2000, realizado por um consorcio multinacional  
232 (THE ARABIDOPSIS GENOME INITIATIVE, 2000). O consórcio *Sol Genomics*  
233 (FERNANDEZ-POZO *et al.*, 2014), que abriga os genomas de *Capsicum*, compreende uma  
234 plataforma *online* na qual podem ser depositados, removidos e alterados dados de diferentes  
235 fontes, fenotípicos, genotípicos e genômicos, garantindo que as informações depositadas

236 sejam as mais recentes. Essa estratégia de integração de dados aumentou também as  
237 possibilidades de análises, além do desenvolvimento de novas ferramentas computacionais.  
238 *Capsicum annuum*, que é a principal espécie do gênero, teve o genoma sequenciado e  
239 publicado em 2014 por dois grupos de pesquisa distintos. Kim *et al.* (2014) utilizaram os  
240 acessos CM334 para *C. annuum* e PI159236 para *C. chinense*, enquanto Qin *et al.* (2014)  
241 utilizaram a cultivar Zunla-1 e *C. annuum* var. *glabriusculum*. Com base no genoma  
242 referência de *C. annuum* foi montado o genoma de *C. baccatum* (KIM *et al.*, 2017).

243                 Dados de sequenciamento genômico são uma grande fonte de informações  
244 para o entendimento da distribuição e composição dos genomas com base na fração repetitiva,  
245 em especial os elementos de transposição (GORINSEK, GUBENSEK E KORDIS, 2004).  
246 Kim *et al.* (2008), com base no transcriptoma de *Capsicum annuum*, disponibilizaram um  
247 banco de dados de EST (*expressed sequences tags*) que facilitou os estudos relacionados aos  
248 padrões de expressão de genes de abrangência populacional. Tendo em vista a publicação e  
249 disponibilização dos genomas completamente sequenciados nas três principais espécies  
250 cultivadas de *Capsicum*, foi criada uma enorme possibilidade para explorar e compreender  
251 sobre a composição dos genomas, baseada em análises comparativas, com foco geral e em  
252 famílias de DNA específicas incluindo as expressas. KIM *et al.* (2017) demonstraram que  
253 cerca de 81% do genoma de *C. annuum* e *C. chinense* é composto por elementos  
254 transponíveis, para *C. baccatum* esse valor chega a 82%.

255                 O presente trabalho foi delineado com base nos seguintes questionamentos:  
256 i) Os mesmos elementos repetitivos se encontram qualitativa e quantitativamente homogêneos  
257 nesses genomas? ii) Tendo em vista o agrupamento filogenético entre as espécies, é possível  
258 estabelecer paralelos quanto ao acúmulo de repetitivos? iii) Há flutuações na ocorrência de  
259 famílias de TEs nesses três genomas? iv) É possível estabelecer uma relação física entre LTR-  
260 RTs e heterocromatina?

261 **3. OBJETIVOS**

262

263 *3.1. Objetivo geral*

264 Compreender como os genomas e os cromossomos de *Capsicum annuum*,  
265 *C. chinense* e *C. baccatum* se organizam e diferenciam com base na fração repetitiva de DNA.

266

267 *3.2. Objetivos específicos*

268

269 a) Buscar e comparar a frequência e a localização das famílias de elementos de  
270 transposição mais comuns nos genomas de *Capsicum* (Solanaceae);

271 b) Quantificar por meio de ferramentas de bioinformática a ocorrência dos LTR-RTs,  
272 ERVs, transposons e DNAsat;

273 c) Buscar, organizar e separar em grupos os elementos repetitivos mais representativos  
274 nesses três genomas;

275 d) Localizar os principais LTR-RTs presente nesses genomas e relacioná-los com as  
276 regiões de heterocromatina, bem como determinar o padrão de distribuição ao longo  
277 dos cromossomos;

278 e) Compreender como os genomas se organizam, e gerar informações e marcas úteis para  
279 estudos de diferenciação genômica e cariotípica nas três espécies mais importantes de  
280 pimentas e pimentões.

281 **4. REFERÊNCIAS BIBLIOGRÁFICAS**

- 282 ACOSTA, C. M. *et al.* Karyotype analysis in several South American species of *Solanum* and  
283 *Lycianthes rantonnei* (Solanaceae). **Taxon**, v. 54, n. 3, p. 713-723, 2005.
- 284 ACOSTA, C. M.; GUERRA, M.; MOSCONE, E. A. Karyological relationships among some  
285 South American species of *Solanum* (Solanaceae) based on fluorochrome banding and nuclear  
286 DNA amount. **Plant systematics and evolution**, v. 298, n. 8, p. 1547-1556, 2012.
- 287 AGUILERA, P. M.; DEBAT, H.; GRABIELE, M. An Integrated Physical Map of the  
288 Cultivated Hot Chili Pepper, *Capsicum baccatum* var. *Pendulum*. **International Journal of**  
289 **Agriculture and Biology**, v. 19, n. 3, 2017.
- 290 AMBROŽOVÁ, K. *et al.* Diverse retrotransposon families and an AT-rich satellite DNA  
291 revealed in giant genomes of *Fritillaria* lilies. **Annals of Botany**, v. 107, n. 2, p. 255-268,  
292 2010.
- 293 ARABIDOPSIS GENOME INITIATIVE *et al.* Analysis of the genome sequence of the  
294 flowering plant *Arabidopsis thaliana*. **nature**, v. 408, n. 6814, p. 796, 2000.
- 295 AZA-GONZÁLEZ, C.; NÚÑEZ-PALENIUS, H. G.; OCHOA-ALEJO, Neftalí. Molecular  
296 biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). **Plant cell reports**, v.  
297 30, n. 5, p. 695-706, 2011.
- 298 BADR, A.; SF, Khalifa; AI, Aboel-Atta. Chromosomal criteria and taxonomic relationships in  
299 the Solanaceae. **Cytologia**, v. 62, n. 2, p. 103-113, 1997.
- 300 BENNETT, M.D., LEITCH, I. J. **Plant DNA C-values database** (release 6.0, Dec. 2012)  
301 <http://www.kew.org/cvalues/>. 2012
- 302 BENNETZEN, J. L.; WANG, H. The contributions of transposable elements to the structure,  
303 function, and evolution of plant genomes. **Annual review of plant biology**, v. 65, p. 505-530,  
304 2014.
- 305 BENNETZEN, J. L. Transposable element contributions to plant gene and genome  
306 evolution. **Plant molecular biology**, v. 42, n. 1, p. 251-269, 2000.
- 307 BENNETZEN, J. L. Mechanisms and rates of genome expansion and contraction in flowering  
308 plants. **Genetica**, v. 115, n. 1, p. 29-36, 2002.
- 309 BENNETZEN, J. L., MA, J. e DEVOS, K.M. Mechanisms of recent genome size variation in  
310 flowering plants. **Annals of Botany**. 95, 127–132 .2005.
- 311 BERGMAN, C. M.; QUESNEVILLE, H. Discovering and detecting transposable elements in  
312 genome sequences. **Briefings in bioinformatics**, v. 8, n. 6, p. 382-392, 2007.
- 313 BOLGER, A. *et al.* The genome of the stress-tolerant wild tomato species *Solanum*  
314 *pennellii*. **Nature genetics**, v. 46, n. 9, p. 1034, 2014.
- 315 BOUSALEM, M.; DOUZERY, E. J. P; SEAL, S. E. Taxonomy, molecular phylogeny and  
316 evolution of plant reverse transcribing viruses (family Caulimoviridae) inferred from full-

- 317 length genome and reverse transcriptase sequences. **Archives of virology**, v. 153, n. 6, p.  
318 1085, 2008.
- 319 CARRIZO GARCÍA, C. *et al.* Wild Capsicums: identification and in situ analysis of Brazilian  
320 species. **Breakthroughs in the genetics and breeding of Capsicum and eggplant**. Edited  
321 by S. Lanteri, and GL Rotino, p. 205-213, 2013.
- 322 CARRIZO GARCÍA, C. *et al.* Phylogenetic relationships, diversification and expansion of  
323 chili peppers (*Capsicum*, Solanaceae). **Annals of botany**, v. 118, n. 1, p. 35-51, 2016.
- 324 CHEEMA, S. K.; PANT, M. R. Karyotype analysis of seven cultivated varieties of *Capsicum*  
325 *annuum* L. **Caryologia**, v. 66, n. 1, p. 70-75, 2013.
- 326 CONTENTO, A.; HESLOP-HARRISON, J. S.; SCHWARZACHER, T. Diversity of a major  
327 repetitive DNA sequence in diploid and polyploid Triticeae. **Cytogenetic and Genome**  
328 **Research**, v. 109, n. 1-3, p. 34-42, 2005.
- 329 DE CASTRO NUNES, R. *et al.* Structure and distribution of centromeric retrotransposons at  
330 diploid and allotetraploid *Coffea* centromeric and pericentromeric regions. **Frontiers in Plant**  
331 **Science**, v. 9, p. 175, 2018.
- 332 DI FILIPPO, M. *et al.* Euchromatic and heterochromatic compositional properties emerging  
333 from the analysis of *Solanum lycopersicum* BAC sequences. **Gene**, v. 499, n. 1, p. 176-181,  
334 2012.
- 335 DIOP, S. I. *et al.* Tracheophyte genomes keep track of the deep evolution of the  
336 *Caulimoviridae*. **Scientific Reports**, v. 8, n. 1, p. 572, 2018.
- 337 DJIAN-CAPORALINO, C. *et al.* Root-knot nematode (*Meloidogyne* spp.) Me resistance  
338 genes in pepper (*Capsicum annum* L.) are clustered on the P9 chromosome. **Theoretical**  
339 **and Applied Genetics**, v. 114, n. 3, p. 473-486, 2007.
- 340 DOLEŽEL, J. *et al.* Plant genome size estimation by flow cytometry: inter-laboratory  
341 comparison. **Annals of Botany**, v. 82, n. suppl\_1, p. 17-26, 1998.
- 342 ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature reviews**  
343 **genetics**, v. 5, n. 6, p. 435, 2004.
- 344 FAUQUET, C. M. *et al.* (Ed.). **Virus taxonomy: VIIIth report of the International**  
345 **Committee on Taxonomy of Viruses**. Academic Press, 2005.
- 346 FERNANDEZ-POZO, N. *et al.* The Sol Genomics Network (SGN)—from genotype to  
347 phenotype to breeding. **Nucleic acids research**, v. 43, n. D1, p. D1036-D1041, 2014.
- 348 GALBRAITH, D. W. *et al.* Rapid flow cytometric analysis of the cell cycle in intact plant  
349 tissues. **Science**, v. 220, n. 4601, p. 1049-1051, 1983.
- 350 GARRIDO-RAMOS, M. A. Satellite DNA in plants: more than just rubbish. **Cytogenetic**  
351 **and genome research**, v. 146, n. 2, p. 153-170, 2015.
- 352 GARRIDO-RAMOS, M. Satellite DNA: An evolving topic. **Genes**, v. 8, n. 9, p. 230, 2017.

- 353 GAYRAL, P. *et al.* A single Banana streak virus integration event in the banana genome as  
354 the origin of infectious endogenous pararetrovirus. **Journal of virology**, v. 82, n. 13, p. 6697-  
355 6710, 2008.
- 356 GEERING A. D. W., HULL, R. Family Caulimoviridae. In: King, A. M. Q.; ADAMS, M. J.;  
357 CARSTENS, E. B.; LEFKOWITZ, E. J. (eds) **Virus Taxonomy, Classification and**  
358 **Nomenclature of Viruses; Ninth Report of the International Committee on Taxonomy of**  
359 **Viruses**. London, UK, Elsevier Academic Press, pp 429–443. 2012.
- 360 GRABIELE, M. *et al.* Highly GC-rich heterochromatin in chili peppers (Capsicum-  
361 Solanaceae): A cytogenetic and molecular characterization. **Scientia Horticulturae**, v. 238, p.  
362 391-399, 2018.
- 363 GONG, Z.; HAN, G. Euphyllophyte paleoviruses illuminate hidden diversity and  
364 macroevolutionary mode of Caulimoviridae. **Journal of virology**, p. JVI. 02043-17, 2018.
- 365 GORINŠEK, B.; GUBENŠEK, F.; KORDIŠ, D. Evolutionary genomics of chromoviruses in  
366 eukaryotes. **Molecular biology and evolution**, v. 21, n. 5, p. 781-798, 2004.
- 367 HAAS, M. *et al.* Cauliflower mosaic virus: still in the news. **Molecular plant pathology**, v.  
368 3, n. 6, p. 419-429, 2002.
- 369 HARPER, G. *et al.* Viral sequences integrated into plant genomes. **Annual Review of**  
370 **Phytopathology**, v. 40, n. 1, p. 119-136, 2002.
- 371 HARPER, G. *et al.* Detection of petunia vein-clearing virus: model for the detection of DNA  
372 viruses in plants with homologous endogenous pararetrovirus sequences. **Journal of**  
373 **virological methods**, v. 107, n. 2, p. 177-184, 2003.
- 374 HAWKINS, J. S. *et al.* Rapid DNA loss as a counterbalance to genome expansion through  
375 retrotransposon proliferation in plants. **Proceedings of the National Academy of Sciences**, v.  
376 106, n. 42, p. 17811-17816, 2009.
- 377 HEMLEBEN, V. **Molekularbiologie der Pflanzen: mit 28 Tabellen**. Fischer, 1990.
- 378 HESLOP-HARRISON, J. S.; SCHMIDT, T. Genomes, genes and junk: the large-scale  
379 organization of plant chromosomes. **Trends in Plant Science**, v. 3, n. 5, p. 195-199, 1998.
- 380 HESLOP-HARRISON, J. S. SCHMIDT, T. Plant Nuclear Genome Composition.  
381 **Encyclopedia of Life Sciences**, p. 1–8, 2007.
- 382 HESLOP- HARRISON, J. S.; SCHWARZACHER, T. Organisation of the plant genome in  
383 chromosomes. **The Plant Journal**, v. 66, n. 1, p. 18-33, 2011.
- 384 HUNZIKER, A. T. **Genera Solanacearum: the genera of Solanaceae illustrated, arranged**  
385 **according to a new system**. ARG Gantner, 2001.
- 386 IBARRA-LACLETTE, E. *et al.* Architecture and evolution of a minute plant  
387 genome. **Nature**, v. 498, n. 7452, p. 94, 2013.
- 388 ICTV - **International Committee on Taxonomy of Viruses**. < <https://talk.ictvonline.org/>>.  
389 Acesso em 23 de setembro de 2017

- 390 JOUFFROY, O. *et al.* Comprehensive repeatome annotation reveals strong potential impact  
391 of repetitive elements on tomato ripening. **BMC genomics**, v. 17, n. 1, p. 624, 2016.
- 392 KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable  
393 elements implemented in Repbase. **Nature Reviews Genetics**, v. 9, n. 5, p. 411, 2008.
- 394 KAZAZIAN, H. H. Mobile elements: drivers of genome evolution. **science**, v. 303, n. 5664,  
395 p. 1626-1632, 2004.
- 396 KIM, H. *et al.* Pepper EST database: comprehensive in silico tool for analyzing the chili  
397 pepper (*Capsicum annuum*) transcriptome. **BMC Plant Biology**, v. 8, n. 1, p. 101, 2008.
- 398 KIM, S. *et al.* Genome sequence of the hot pepper provides insights into the evolution of  
399 pungency in *Capsicum* species. **Nature genetics**, v. 46, n. 3, p. 270, 2014.
- 400 KIM, S. *et al.*, New reference genome sequences of hot pepper reveal the massive evolution  
401 of plant disease-resistance genes by retroduplication. **Genome biology**, v. 18, n. 1, p. 210,  
402 2017.
- 403 KUBIS, S.; SCHMIDT, T.; HESLOP-HARRISON, J. Repetitive DNA elements as a major  
404 component of plant genomes. **Annals of Botany**, v. 82, n. suppl\_1, p. 45-55, 1998.
- 405 KUMAR, A.; BENNETZEN, J. L. Plant retrotransposons. **Annual review of genetics**, v. 33,  
406 n. 1, p. 479-532, 1999.
- 407 LEVIN, H. L.; MORAN, J. V. Dynamic interactions between transposable elements and their  
408 hosts. **Nature Reviews Genetics**, v. 12, n. 9, p. 615, 2011.
- 409 LISCH, D. Epigenetic regulation of transposable elements in plants. **Annual review of plant**  
410 **biology**, v. 60, p. 43-66, 2009.
- 411 LISCH, D. How important are transposons for plant evolution? **Nature Reviews Genetics**, v.  
412 14, n. 1, p. 49, 2013.
- 413 LIU, W. *et al.* IBS: an illustrator for the presentation and visualization of biological  
414 sequences. **Bioinformatics**, v. 31, n. 20, p. 3359-3361, 2015.
- 415 LLORENS, C. *et al.* The *Gypsy* Database (GyDB) of mobile genetic elements: release  
416 2.0. **Nucleic acids research**, v. 39, n. suppl\_1, p. D70-D74, 2010.
- 417 LOCKHART, B. E. *et al.* Characterization and genomic analysis of tobacco vein clearing  
418 virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in  
419 the host genome. **Journal of General Virology**, v. 81, n. 6, p. 1579-1585, 2000.
- 420 MARTINS, L. V. *et al.* Heterochromatin distribution and histone modification patterns of  
421 H4K5 acetylation and H3S10 phosphorylation in *Capsicum* L. **Crop Breeding and Applied**  
422 **Biotechnology**, v. 18, n. 2, p. 161-168, 2018.
- 423 MEHRA, M.; GANGWAR, I.; SHANKAR, R. A deluge of complex *repeats*: the *Solanum*  
424 genome. **PloS one**, v. 10, n. 8, p. e0133962, 2015.

- 425 MEHROTRA, Shweta; GOYAL, Vinod. Repetitive sequences in plant nuclear DNA: types,  
426 distribution, evolution and function. **Genomics, proteomics E bioinformatics**, v. 12, n. 4, p.  
427 164-171, 2014.
- 428 MELAYAH, D. *et al.* Distribution of the Tnt1 retrotransposon family in the amphidiploid  
429 tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives. **Biological Journal of the**  
430 **Linnean Society**, v. 82, n. 4, p. 639-649, 2004.
- 431 MELTERS, D. P. *et al.*, Comparative analysis of tandem repeats from hundreds of species  
432 reveals unique insights into centromere evolution. **Genome biology**, v. 14, n. 1, p. R10, 2013.
- 433 MOSCONE, E. A. *et al.* Giemsa C-banded karyotypes in *Capsicum* (Solanaceae). **Plant**  
434 **Systematics and Evolution**, v. 186, n. 3-4, p. 213-229, 1993.
- 435 MOSCONE, E. A. *et al.* Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow  
436 cytometry and Feulgen densitometry. **Annals of Botany**, v. 92, n. 1, p. 21-29, 2003.
- 437 MOSCONE, E. A. *et al.* The evolution of chili peppers (*Capsicum*-Solanaceae): a cytogenetic  
438 perspective. In: **VI International Solanaceae Conference: Genomics Meets Biodiversity**  
439 **745**. 2006. p. 137-170.
- 440 MUÑOZ-LÓPEZ, M.; GARCÍA-PÉREZ, J. L. DNA transposons: nature and applications in  
441 genomics. **Current genomics**, v. 11, n. 2, p. 115-128, 2010.
- 442 NAGAKI, K. *et al.* Coexistence of NtCENH3 and two retrotransposons in tobacco  
443 centromeres. **Chromosome Research**, v. 19, n. 5, p. 591-605, 2011.
- 444 NDOWORA, T. *et al.* Evidence that badnavirus infection in musaceae originate from integrated  
445 pararetroviral sequences. **Virology**, v. 255, n. 2, p. 214-220, 1999.
- 446 NEUMANN, P. *et al.* Plant centromeric retrotransposons: a structural and cytogenetic  
447 perspective. **Mobile DNA**, v. 2, n. 1, p. 4, 2011.
- 448 PADILHA, H. K. M. *et al.* Agronomic evaluation and morphological characterization of chili  
449 peppers (*Capsicum annuum*, Solanaceae) from Brazil. **Embrapa Clima Temperado-Artigo**  
450 **em periódico indexado (ALICE)**, 2016.
- 451 PARK, Y. *et al.* Karyotyping of the chromosomes and physical mapping of the 5S rRNA and  
452 18S-26S rRNA gene families in five different species in *Capsicum*. **Genes E genetic systems**,  
453 v. 74, n. 4, p. 149-157, 1999a.
- 454 PARK, Y. *et al.*, Chromosomal localization and sequence variation of 5S rRNA gene in five  
455 *Capsicum* species. **Molecules and cells**, v. 10, n. 1, p. 18-24, 1999b.
- 456 PARK, M. *et al.* Comparative analysis of pepper and tomato reveals euchromatin expansion  
457 of pepper genome caused by differential accumulation of Ty3/Gypsy-like elements. **BMC**  
458 **genomics**, v. 12, n. 1, p. 85, 2011.
- 459 PARK, M. *et al.*, Evolution of the large genome in *Capsicum annuum* occurred through  
460 accumulation of single- type long terminal repeat retrotransposons and their derivatives. **The**  
461 **Plant Journal**, v. 69, n. 6, p. 1018-1029, 2012.

- 462 PAZ, R. C. *et al.* Diversity, distribution and dynamics of full-length Copia and Gypsy LTR  
463 retroelements in *Solanum lycopersicum*. **Genetica**, v. 145, n. 4-5, p. 417-430, 2017.
- 464 PELLICER, J.; FAY, M. F.; LEITCH, I. J. The largest eukaryotic genome of them  
465 all?. **Botanical Journal of the Linnean Society**, v. 164, n. 1, p. 10-15, 2010.
- 466 PEZER, Ž. *et al.* Satellite DNA-mediated effects on genome regulation. In: **Repetitive DNA**.  
467 Karger Publishers, 2012. p. 153-169.
- 468 PLATT, R. N.; VANDEWEGE, M. W.; RAY, D. A. Mammalian transposable elements and  
469 their impacts on genome evolution. **Chromosome Research**, p. 1-19, 2018.
- 470 PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA evolution. In: **Repetitive**  
471 **DNA**. Karger Publishers, 2012. p. 126-152.
- 472 PRINGLE, C. R. The universal system of virus taxonomy of the International Committee on  
473 Virus Taxonomy (ICTV), including new proposals ratified since publication of the Sixth  
474 ICTV Report in 1995. **Archives of virology**, v. 143, n. 1, p. 203-210, 1998.
- 475 QIN, C. *et al.* Whole-genome sequencing of cultivated and wild peppers provides insights  
476 into *Capsicum* domestication and specialization. **Proceedings of the National Academy of**  
477 **Sciences**, v. 111, n. 14, p. 5135-5140, 2014.
- 478 RHOADS, A.; AU, K. F. PacBio sequencing and its applications. **Genomics, proteomics &**  
479 **bioinformatics**, v. 13, n. 5, p. 278-289, 2015.
- 480 RICHERT- PÖGGELER, K. R. *et al.* Induction of infectious petunia vein clearing (pararetro)  
481 virus from endogenous provirus in petunia. **The EMBO journal**, v. 22, n. 18, p. 4836-4845,  
482 2003.
- 483 ROMANO, E.; BRASILEIRO, A. C. M. Extração de DNA de plantas. **Biotecnologia,**  
484 **Ciência e Desenvolvimento**, Brasília, v. 2, n. 9, p. 40-43, 1999.
- 485 ROTHNIE, H. M.; CHAPDELAIN, Y.; HOHN, T. Pararetroviruses and retroviruses: a  
486 comparative review of viral structure and gene expression strategies. In: **Advances in virus**  
487 **research**. Academic Press, 1994. p. 1-67.
- 488 RUDD, M. K.; WILLARD, H. F. Analysis of the centromeric regions of the human genome  
489 assembly. **Trends in genetics**, v. 20, n. 11, p. 529-533, 2004.
- 490 SANGER, F. *et al.* Nucleotide sequence of bacteriophage  $\phi$ X174 DNA. **nature**, v. 265, n.  
491 5596, p. 687, 1977.
- 492 SANTOS, F. C. *et al.* Chromosomal distribution and evolution of abundant retrotransposons  
493 in plants: gypsy elements in diploid and polyploid *Brachiaria* forage grasses. **Chromosome**  
494 **research**, v. 23, n. 3, p. 571-582, 2015.
- 495 SCALDAFERRO, M. A.; GRABIELE, M.; MOSCONE, E. A. Heterochromatin type, amount  
496 and distribution in wild species of chili peppers (*Capsicum*, Solanaceae). **Genetic resources**  
497 **and crop evolution**, v. 60, n. 2, p. 693-709, 2013.

- 498 SCALDAFERRO, M. A. *et al.* FISH and AgNor mapping of the 45S and 5S rRNA genes in  
499 wild and cultivated species of *Capsicum* (Solanaceae). **Genome**, v. 59, n. 2, p. 95-113, 2016.
- 500 SCHATZ, M. C.; WITKIWSKI J.; MCCOMBIE, W.R. Current challenges in de novo plant  
501 genome sequencing and assembly. **Genome biology**, v. 13, n. 4, p. 1-7, 2012.
- 502 SCHWEIZER, D. Reverse fluorescent chromosome banding with chromomycin and  
503 DAPI. **Chromosoma**, v. 58, n. 4, p. 307-324, 1976.
- 504 SEIBT, K. M. *et al.* Short interspersed nuclear elements (SINEs) are abundant in Solanaceae  
505 and have a family- specific impact on gene structure and genome organization. **The Plant**  
506 **Journal**, v. 86, n. 3, p. 268-285, 2016.
- 507 SIMS, D. *et al.* Sequencing depth and coverage: key considerations in genomic analyses.  
508 **Nature Reviews**, v. 15, n. 2, fev. 2014.
- 509 SMITH, George P. Evolution of repeated DNA sequences by unequal crossover. **Science**, v.  
510 191, n. 4227, p. 528-535, 1976.
- 511 SONNHAMMER, E. L. L.; DURBIN, R. A dot-matrix program with dynamic threshold  
512 control suited for genomic DNA and protein sequence analysis. **Gene**, v. 167, n. 1-2, p. GC1-  
513 GC10, 1995.
- 514 STATON, S. E. *et al.* The sunflower (*Helianthus annuus* L.) genome reflects a recent history  
515 of biased accumulation of transposable elements. **The Plant Journal**, v. 72, n. 1, p. 142-153,  
516 2012.
- 517 STAVOLONE, L. *et al.* Cestrum yellow leaf curling virus (CmYLCV) promoter: a new  
518 strong constitutive promoter for heterologous gene expression in a wide variety of  
519 crops. **Plant molecular biology**, v. 53, n. 5, p. 703-713, 2003.
- 520 STOYE, J. P. Endogenous retroviruses: still active after all these years? **Current biology**, v.  
521 11, n. 22, p. R914-R916, 2001.
- 522 TEMNYKH, S. *et al.* Computational and experimental analysis of microsatellites in rice  
523 (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker  
524 potential. **Genome research**, v. 11, n. 8, p. 1441-1452, 2001.
- 525 THE PLANT LIST. **Version 1.1. Published on the Internet**; <http://www.theplantlist.org/>  
526 (acesso em 14 de janeiro de 2019). 2013.
- 527 TOMATO GENOME CONSORTIUM *et al.* The tomato genome sequence provides insights  
528 into fleshy fruit evolution. **Nature**, v. 485, n. 7400, p. 635, 2012.
- 529 VERNHETTES, S.; GRANDBASTIEN, M.; CASACUBERTA, J. M. The evolutionary  
530 analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its  
531 regulatory sequences. **Molecular biology and evolution**, v. 15, n. 7, p. 827-836, 1998.
- 532 WICKER, T. *et al.* A unified classification system for eukaryotic transposable  
533 elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973, 2007.

- 534 WILLARD, H. F.; WAYE, J. S. Hierarchical order in chromosome-specific human alpha  
535 satellite DNA. **Trends in Genetics**, v. 3, p. 192-198, 1987.
- 536 XIONG, Y.; EICKBUSH, T. H. Origin and evolution of retroelements based upon their  
537 reverse transcriptase sequences. **The EMBO journal**, v. 9, n. 10, p. 3353-3362, 1990.
- 538 XU, S. *et al.* Wild tobacco genomes reveal the evolution of nicotine  
539 biosynthesis. **Proceedings of the National Academy of Sciences**, p. 201700073, 2017.
- 540

541 **5. ARTIGO**542 Relações citogenômicas entre espécies do gênero *Capsicum* L. (SOLANACEAE)

543

544 Rafael de Assis<sup>1</sup>, André Luís Laforga Vanzela<sup>1\*</sup>

545

546 1 Laboratório de Citogenética e Diversidade Vegetal, Departamento de Biologia  
547 Geral, Centro de Ciências Biológicas, Universidade Estadual de Londrina, Londrina,  
548 Paraná, Brasil.

549

550 \*E-mail: andrevanzela@uel.br

551 **Resumo**

552 Os genomas vegetais são compostos principalmente por sequências repetitivas.  
553 Elementos transponíveis (ETs), que formam a parte móvel dos genomas, são  
554 divididos nas Classes I e II, elementos pertencentes a Classe I, ou retrotransposons,  
555 são sintetizados e transpostos usando um RNA intermediário e podem acumular  
556 diferencialmente dependendo do grupo vegetal. A fração repetitiva não codificante,  
557 representada pelas sequências satélite, tende a se acumular em blocos ao longo  
558 dos cromossomos. As pimentas, pertencentes à família Solanaceae, possuem  
559 grande valor comercial, sendo comercializadas *in natura* ou especiarias, sendo  
560 também utilizadas como ornamentação e de forma medicinal. Mesmo já possuindo  
561 seu sequenciamento publicado, trabalhos focando na diversidade e distribuição de  
562 elementos repetitivos nos genomas das espécies de *Capsicum* ainda são  
563 superficiais. Diante disso, os objetivos deste trabalho foram compreender a  
564 organização, a distribuição e as relações genômicas e cariotípicas da fração

565 repetitiva nos genomas de *Capsicum annuum*, *C. chinense* e *C. baccatum*. Para  
566 isso, foram utilizadas ferramentas de bioinformática para quantificar essa fração e  
567 métodos de citogenética molecular para a localização física dessas sequências.  
568 Sequenciamentos genômicos de alta cobertura foram contrastados contra um banco  
569 de dados de sequências conservadas de elementos repetitivos. Foram observadas  
570 similaridades entre as sequências repetitivas nos genomas de *C. annuum* e *C.*  
571 *chinense*, em relação à *C. baccatum*. A superfamília *Gypsy* foi mais abundante que  
572 *Copia*, especialmente os elementos Del, que foram mais representativos em *C.*  
573 *annuum* e *C. chinense*, enquanto que em *C. baccatum* houve maior acúmulo de  
574 elementos Athila. O resultado da FISH mostrou uma distribuição predominantemente  
575 dispersa dos retrotransposons, com exceção da família CRM que predominou na  
576 região pericentromérica, próximas a bandas de heterocromatina. Nossos resultados  
577 confirmaram a relação mais próxima entre *Capsicum annuum* e *C. chinense*, em  
578 relação à *C. baccatum*, concordando com a filogenia do gênero.

579

## 580 **Introdução**

581 O DNA nuclear nos vegetais é composto por uma fração não repetitiva (uma ou  
582 poucas cópias de sequências codificantes, íntrons, promotores e sequências de  
583 DNA regulatórias) e outra fração repetitiva, com centenas ou até milhares de cópias.  
584 Essa última fração pode ser organizada em grupos variados de acordo com a  
585 natureza e o modo de repetição das sequências [1–3]. A fração repetitiva pode ser  
586 encontrada dispersa nos genomas (elementos de transposição, TEs = *transposable*  
587 *elements*) ou organizadas em *tandem* (DNA ribossômico, microssatélites,  
588 minissatélites, satélites, sequências teloméricas e centroméricas e TEs). Os  
589 elementos transponíveis são divididos em classe I e II, os de Classe I, também

590 chamados retrotransposons, são os mais comuns nos vegetais. A transposição  
591 ocorre pelo mecanismo de “cópia e cola”, onde o RNAm transcrito é convertido em  
592 DNA complementar, por meio da enzima transcriptase reversa, e integrado em outra  
593 região do genoma por meio da enzima integrase [4–6]. Os retrotransposons podem  
594 ser classificados de acordo com a presença ou ausência de LTR (repetições  
595 terminais longas). Os retrotransposons com LTR são subdivididos em quatro  
596 superfamílias, sendo *Copia* e *Gypsy* as mais comuns em plantas. Essa classificação  
597 é baseada nas diferenças observadas na ordem dos genes da cadeia poligênica,  
598 principalmente em relação à posição da integrase [7,8]. Elementos de Classe II  
599 (Transposons) são transpostos por meio um mecanismo de “corta e cola”, no qual o  
600 elemento é excisado e reintegrado em outra região do genoma utilizando a enzima  
601 transposase. Muitos transposons de DNA são flanqueados por sequências terminais  
602 repetidas invertidas (TIR – *terminal inverted repeat*), a transposase reconhece as  
603 TIRs, permitindo assim a excisão do transposon e sua reintegração em outra região  
604 do genoma [9,10].

605 O DNA satélite corresponde à fração repetitiva não codificante com mais de 30pb e  
606 são considerados os principais componentes da heterocromatina, predominantes  
607 nas regiões pericentroméricas, centroméricas e subteloméricas dos cromossomos  
608 [11,12]. Diferentemente dos DNA satélite, os microssatélites e minissatélites são  
609 classificados de acordo com o número mínimo de bases ou número mínimo de  
610 *repeats*, contudo, tal diferença pode não ser tão clara, porque nem sempre o  
611 tamanho do motivo é suficiente para classificá-los [12]. Os principais tipos de  
612 microssatélites em plantas ocorrem na forma de mono- di- tri ou até  
613 heptanucleotídeos [12–14]. O mecanismo de mutação mais aceito para o aumento e  
614 diversificação de microssatélites e minissatélites é o *slippage*, que ocorre

615 principalmente durante a replicação, decorrente de um “escorrego” da polimerase  
616 em função da homologia nos *repeats*. Nesse caso, fitas de DNA podem se anelar de  
617 maneira irregular, produzindo uma alça levando ao reparo seguido do aumento ou  
618 diminuição dessas sequências [14].

619 O gênero *Capsicum* abrange as pimentas e pimentões, amplamente cultivados nas  
620 regiões tropicais e subtropicais, apresentando grande variedade de formas,  
621 tamanhos e cores de frutos [15]. A pungência característica é devida ao acúmulo de  
622 um grupo de alcalóides, os capsaicinóides [16,17]. O gênero também é utilizado na  
623 produção de compostos medicinais, na ornamentação e para consumo, devido ao  
624 grande acúmulo de nutrientes que os frutos apresentam [18,19]. Das 43 espécies  
625 descritas para o gênero [20], cinco são consideradas domesticadas (*C. annuum* L.,  
626 *C. baccatum* L., *C. chinense* Jacq., *C. frutescens* L. e *C. pubescens* R E P) e as  
627 demais silvestres [19,21]. *Capsicum annuum*, *C. chinense* e *C. frutescens* pertencem  
628 ao clado Annum, *C. baccatum* ao clado Baccatum e *C. pubescens* ao clado  
629 Pubescens [22]. As descrições citogenéticas acerca da família Solanaceae sugerem  
630 a predominância de cariótipos com  $2n = 24$ , e variações quanto a esse valor podem  
631 estar relacionadas com eventos de poliploidia, aneuploidia e disploidia [19,23,24]. No  
632 gênero *Capsicum*, predominam os números haploides  $n = 12$  e  $13$  e cariótipos  
633 simétricos com cromossomos meta- e submetacêntricos, de modo geral, cariótipos  
634 com  $n=12$  são compostos por 11 pares metacêntrico e 1 par submetacêntrico, sendo  
635 essa característica mais observada nas espécies domesticadas. Tais características  
636 foram descritas por MOSCONE *et al.* [22,25], SCALDAFERRO, GABRIELE E  
637 MOSCONE [26] e SCALDAFERRO *et al.* [27]. Os valores C de DNA foram  
638 estimados por MOSCONE *et al.* [18] em 25 amostras de nove espécies diploides de  
639 *Capsicum*, os quais variaram de 3,32 pg em *C. annuum* até 5,77 pg em *C.*

640 *parvifolium*. Do mesmo modo que são observadas variações no valor C de DNA  
641 entre as espécies, os perfis de bandas de heterocromatina também são variáveis.  
642 Os primeiros trabalhos de bandeamento envolvendo espécies do gênero *Capsicum*,  
643 demonstraram diversidade quanto a presença de bandas terminais (bandas mais  
644 intensas ou tênues) e quanto a presença de bandas intersticiais e proximais [25,28],  
645 assim como em trabalhos mais recentes [26,29].

646 Dados de sequenciamento genômico são uma grande fonte de informações para o  
647 entendimento da distribuição e composição dos genomas com base na fração  
648 repetitiva, em especial os elementos de transposição [30]. Até o momento, foram  
649 sequenciados os genomas completos de três espécies (*C. annuum*, *C. chinense* e *C.*  
650 *baccatum*), todos tendo como referência o genoma de *C. annuum* [15,17]. Tendo em  
651 vista que as descrições relativas à fração repetitiva desses genomas são muito  
652 superficiais e da livre disponibilidade dos genomas completamente sequenciados e  
653 montados, o foco desse estudo foi identificar, classificar e realizar uma ampla  
654 comparação citogenômica da fração repetitiva de DNA nessas três espécies,  
655 procurando compreender se a diversidade e a distribuição dos retrotransposons  
656 repetitivos seguem um padrão de distribuição cromossômica e acompanham a  
657 organização dos clados propostos na filogenia do gênero, em relação a *C. annuum*,  
658 *C. chinense* e *C. baccatum*.

659

660

## 661 **Materiais e métodos**

### 662 **Banco de dados genômicos em *Capsicum***

663 Os sequenciamentos das três espécies de *Capsicum* (*C. annuum*, *C. chinense* e *C.*  
664 *baccatum*) utilizadas neste trabalho encontram-se disponíveis na plataforma *Pepper*  
665 *Genome Platform* (<http://peppergenome.snu.ac.kr/>), cujas montagens foram  
666 realizadas com base no genoma referência de *C. annuum*. Os tamanhos e números  
667 de *scaffolds* estão descritos na Tabela 1.

668

669 **Tabela 1**: Descrição geral dos arquivos com os *scaffolds* das espécies de *Capsicum*.

|                     | <i>C. annuum</i> | <i>C. baccatum</i> | <i>C. chinense</i> |
|---------------------|------------------|--------------------|--------------------|
| % GC                | 35               | 35,4               | 34,9               |
| % AT                | 65               | 64,6               | 65,1               |
| Number of sequences | 37.989           | 25.349             | 51.917             |
| Max length          | 18.549.843       | 16.429.935         | 20.331.050         |
| Min length          | 264              | 267                | 274                |

670 Características dos bancos de dados genômicos das três espécies de *Capsicum*  
671 estudadas.

672 Para a comparação genômica, foram organizados dois bancos de dados locais com  
673 sequências obtidas do *GypsyDB* (<http://gydb.org/index.php/MainPage>), *RepBase*  
674 (<http://www.girinst.org/censor/>) e *NCBI* (<http://www.ncbi.nlm.nih.gov/>). Esses bancos  
675 foram utilizados para o levantamento e análise da fração repetitiva, separadamente  
676 em cada *dataset* de *scaffolds* genômicos, em *Blast* local. A fração repetitiva

677 analisada compreendeu: i) elementos de transposição (ETs) e elementos virais  
678 endógenos (ERVs), ii) DNA ribossômico (35S e 5S). Todas as análises foram  
679 realizadas em plataforma Linux, utilizando *scripts* de uso livre, assim como outros  
680 elaborados e editados pela equipe do LCDV.

681

## 682 **Avaliação da fração repetitiva codificante**

683 A avaliação da fração repetitiva foi baseada principalmente por meio das  
684 ferramentas BLASTx e *n* (*Basic Alignment Search Tool*), contrastando os *scaffolds* a  
685 bancos de dados locais. A contagem geral (ocorrência de elementos inteiros ou de  
686 suas frações) foi feita por filtragem baseada em *E-value* 10e-4 e identidade superior  
687 70%. Os arquivos de saída foram gerados de maneira tabular (-outfmt 6), sem  
688 alteração na ordem padrão da rotina BLAST. Para a fração repetitiva codificante  
689 (ETs, ERVs e DNAr), foram utilizadas sequências conservadas obtidas de diferentes  
690 espécies de plantas e animais, para aumentar diversidade na busca e as  
691 possibilidades de acerto.

692 A partir dos resultados dos Blastx e Blastn, as sequências similares à fração  
693 repetitiva foram plotadas em uma Tabela quali e quantitativa, organizadas de acordo  
694 com a filogenia proposta por LLORENS *et al.* (2009) [31], acrescentado à essa  
695 organização as frações de DNAr 35S e 5S. As linhagens de retrotransposons foram  
696 agrupadas dentro das superfamílias *Copia* (famílias Oryco, Retrofit, Sire e Tork) e  
697 *Gypsy* (famílias CRM, Del, Reina, Galadriel, Athila e Tat). Todos os dados  
698 quantitativos e qualitativos estimados na planilha SpreadSheets WPS, para  
699 elaboração de gráficos de comparação.

700 Os elementos transponíveis e os elementos virais endógenos mais representativos  
701 foram anotados utilizando o programa Artemis, com a identificação dos domínios  
702 realizada por meio de Blastx *online* no site do NCBI. Os esquemas representativos  
703 foram elaborados com o programa IBS, *Illustrator for Biological Sequences* [32].

704

## 705 **Desenho de *primers***

706 As sequências mais representativas de cada família repetitiva foram separadas para  
707 o desenho de *primers* com o auxílio da ferramenta *Custom Primers – OligoPerfect™*  
708 *Designer* da Thermo Fisher Scientific (<http://tools.thermofisher.com>). A viabilidade  
709 dos *primers* foi checada com a ferramenta *PCR primer stat*  
710 (<http://www.bioinformatics.org>). Os *primers* desenhados estão dispostos na Tabela 3.

711

## 712 **Reconstrução de retrotransposons íntegros**

713 Com base nos resultados das rodadas de Blast, sequências maiores que 3500pb  
714 foram analisadas quanto a presença das LTRs com o *script* Dotter l/b [33] em  
715 ambiente Linux. As sequências portadoras de LTR foram então submetidas ao  
716 Blastx *online* para a localização dos domínios proteicos e verificação da ordem das  
717 proteínas na cadeia poligênica. As sequências completas e com a ordem correta das  
718 proteínas foram alinhadas com programa MAUVE  
719 (<http://darlinglab.org/mauve/mauve.html>). A partir do alinhamento gerado, foram  
720 formados os grupos com base na similaridade gráfica obtida pelo programa. Estes  
721 grupos foram então validados por meio do *script* ClustalW em ambiente Linux. De

722 cada grupo gerado, um representante foi anotado e elaborado de maneira gráfica  
723 utilizando o programa IBS (<http://ibs.biocuckoo.org/>).

724

## 725 **Material vegetal**

726 Sementes de *C. annuum* (Criollo de Morelos) *C. chinense* (GBUEL27) e *C. baccatum*  
727 (GBUEL118), provenientes do banco de germoplasma da Universidade Estadual de  
728 Londrina – UEL, foram germinadas em bandeja com substrato padrão por  
729 aproximadamente 15 dias, replantadas em vasos contendo terra e fertilizante  
730 mineral misto NPK (N04/P14/K08) e mantidas na casa de vegetação do Laboratório  
731 de Citogenética e Diversidade Vegetal, CCB.

732

## 733 **Extração de DNA total e PCR**

734 Folhas jovens foram coletadas, maceradas em nitrogênio líquido e tratadas com  
735 tampão de extração CTAB [34]. As amostras foram purificadas com fenol:clorofórmio  
736 (1:1, v:v) e clorofórmio:álcool isoamílico (24:1 v:v), centrifugadas e o DNA precipitado  
737 em etanol absoluto. Após a eluição em Tris-HCl pH 8.0, a concentração do DNA foi  
738 estimada em NanoDrop 2000 (Thermo).

739 Fragmentos das regiões proteicas conservadas foram amplificados utilizando  
740 *primers* específicos. Para tal, foi utilizada uma PCR padrão, sendo o *mix* composto  
741 por: 0,5 µL de *Taq* Polimerase a 5 u/µL (Fermentas); 2,5 µL tampão 10 ×; 1,5 µL de  
742 MgCl<sub>2</sub> a 50 mM; 1 µL de dNTP a 10 mM, 2 µL de cada *primer* a 5 mM e H<sub>2</sub>O, para  
743 um volume final de 25 µL. As PCRs foram feitas em um termociclador PTC 100  
744 BioRad, na seguinte condição: 94 °C por 1 min, seguido de 30 ciclos de 94 °C por 30

745 s, 58 °C por 40 s, 72 °C por 50 s, finalizando com 72 °C por 10 min, sendo que as  
746 temperaturas de anelamento variaram para cada jogo de *primers*. As reações foram  
747 coradas com brometo de etídio e checadas por eletroforese em gel de agarose a 1%  
748 (3 V/cm).

749

## 750 **Bandeamento C- CMA/DAPI**

751 As raízes foram coletadas e pré-tratadas com uma solução de 0,5% de colchicina  
752 por 1 hora e 30 minutos e fixadas em solução de ácido acético:etanol PA (1:3, v:v)  
753 sob agitação constante. Para a confecção das lâminas, as raízes foram digeridas em  
754 solução composta por celulase e pectinase (2:20, m:v) por 2h30min a 37 °C, lavadas  
755 e maceradas em uma gota de ácido acético 60%. As lamínulas foram removidas  
756 após o congelamento em nitrogênio líquido e secas Após o envelhecimento das  
757 lâminas por 3 dias, as mesmas foram tratadas por 10 minutos em ácido acético 45%  
758 à 60°C, hidróxido de bário saturado por 8 min e 2x SSC à 60°C por 1 hora e 30  
759 minutos. Ao término da série de tratamentos as lâminas foram acomodadas em uma  
760 estufa à 37°C por 1 hora, e em seguida coradas com CMA<sub>3</sub> (cromomicina A3) por 30  
761 minutos e DAPI (4'- 6- diamidino- 2- fenilindole) por 30 minutos, segundo  
762 SCHWEIZER [35] com algumas adaptações. As lâminas foram montadas em meio  
763 DABCO (descrito anteriormente) e armazenadas no escuro por pelo menos três dias.

764

## 765 **Hibridização *in situ***

766 O pré-tratamento das raízes, bem com a confecção das lâminas seguiram os  
767 mesmos passos já citados. Após a confecção das lâminas, as mesmas foram

768 tratadas com 50 µL de uma solução de RNase a 10 mg/mL, a 37 °C por 1h e  
769 pepsina 10% (10 a 30 min em banho maria a 37 °C). Em seguida, essas foram  
770 desidratadas em série alcoólica e desnaturadas em formamida 70% a 70 °C por 10  
771 minutos.

772 Para a FISH, sondas foram preparadas por PCR, utilizando sequências de DNA  
773 relativas às regiões proteicas conservadas dos elementos de transposição. Os  
774 produtos das PCRs foram marcados também por PCR, utilizando um *mix* contendo  
775 um dos nucleotídeos conjugados com biotina-11-dUTP, digoxigenina-11-dUTP ou  
776 Cy3-dUTP. As sondas foram utilizadas em um mix (34 µL) composto da seguinte  
777 maneira: Formamida deionizada 100% (15 µL), Polietilenoglicol a 50% (6 µL), 20 ×  
778 SSC (3 µL), 100 ng de DNA de timo de bezerro (1 µL), 10% de SDS (1 µL) e 100 ng  
779 de sonda marcada (4 µL) e água ultrapura para um volume final de 34 µL.

780 O conjunto de lâminas e *mix* foi acondicionado em um termociclador (95 °C por 10  
781 min, 50 °C por 10 min e 38 °C por 10 min) para a desnaturação e anelamento. A  
782 hibridização foi feita em câmara úmida em uma estufa a 37 °C por 24 horas. Os  
783 banhos pós-hibridização foram realizados em tampão SSC a uma estringência de 70  
784 %. Na detecção foram utilizadas avidina-FITC e anti-digoxigenina-rodamina em BSA  
785 5% preparado em 4× SSC/0,2% Tween 20. Após as lavagens pós-deteção as  
786 lâminas foram montadas em um meio contendo 21 µL de antifade (DABCO), 4 µL de  
787 DAPI 4 µg/mL e 2 µL de cloreto de magnésio (25mM).

788

## 789 **Microscopia de fluorescência e Aquisição de imagens**

790 Os cromossomos foram observados e fotografados com microscópio de  
791 fluorescência DM 4500B, com a câmera DFC 300FX acoplada. As imagens

792 capturadas foram ajustadas para contraste e brilho, sobrepostas e editados no  
793 Photoshop CS6 (Adobe, USA).

## 794 **Resultados**

795

### 796 **Análise dos datasets com base na fração repetitiva**

797

798 *Scaffolds* de alta cobertura dos genomas de *C. annuum* (3,07Gb), *C. chinense*  
799 (3,22Gb) e *C. baccatum* (3,01Gb) foram baixados do *Pepper Genome Platform*.  
800 Esses conjuntos de dados foram utilizados para buscar trechos conservados das  
801 proteínas da fração repetitiva codificante, utilizando como referência um banco de  
802 dados compilado do *GypsyDB*, *RepBase* e *NCBI*. A análise mostrou que a fração  
803 repetitiva representa 2,32 %, 2,14 %, 2,08 % de cada *dataset*, respectivamente  
804 (Tabela 2). Quando comparados os elementos de Classe I com os de Classe II,  
805 notou-se que os de Classe I foram mais abundantes dentro dos genomas de  
806 *Capsicum*, com decréscimo nas porcentagens, já que em *C. annuum* essa fração  
807 correspondeu a 90,51 %, em *C. chinense* 88,05 %, enquanto que para *C. baccatum*  
808 essa fração representou cerca de 65 % (Tabela 2). Trechos conservados de  
809 elementos de Classe II apresentaram baixo acúmulo em *C. annuum*, *C. chinense* e  
810 *C. baccatum*, com valores inferiores a 1% para as três espécies, 0,21%, 0,16% e  
811 0,1%, respectivamente (Tabela 2). Elementos virais endógenos (ERVs) também  
812 foram pouco acumulados nos *datasets*, com 0,13% em *C. annuum*, 0,03% em *C.*  
813 *chinense* e 0,1% em *C. baccatum*, assim como os elementos de Classe II. A fração  
814 correspondente aos microssatélites apresentou valores próximos a 10%, sendo  
815 7,49%, 10,5% e 12,51%, respectivamente. A análise comparativa da fração

816 correspondente aos DNA ribossômicos 35S e 5S revelou um contraste maior, com  
817 acúmulo de sequências de DNAr 35S em *C. baccatum* (22,21%), enquanto que em  
818 *C. annuum* e *C. chinense* os valores foram bem menores, 1,54 e 0,99%,  
819 respectivamente (Tabela 2). Os valores para DNAr 5S não variaram muito entre as  
820 três espécies, com 0,05% em *C. baccatum*, 0,13% em *C. annuum* e 0,27% *C.*  
821 *chinense* (Tabela 2).

822 **Tabela 2:** Frequência e valores relativos dos elementos repetitivos nos três genomas de *Capsicum*.

| Species/Lineages |           | <i>C. annuum</i> (GC 35% - AT 65%) |        |          |          | <i>C. chinense</i> (GC 34.9% - 65.1%) |        |          |          | <i>C. baccatum</i> (GC 35.4% - AT 64.6%) |        |          |          |
|------------------|-----------|------------------------------------|--------|----------|----------|---------------------------------------|--------|----------|----------|--|--------|----------|----------|
|                  |           | num seq                            | NR (%) | size bp  | size CR% | num seq                               | NR (%) | size bp  | size CR% | num seq                                  | NR (%) | size bp  | size CR% |
| Copia            | Oryco     | 132                                | 0,02   | 26275    | 0,04     | 135                                   | 0,02   | 31442    | 0,05     | 152                                      | 0,02   | 34190    | 0,05     |
|                  | Sire      | 4447                               | 0,62   | 1067429  | 1,51     | 4174                                  | 0,56   | 1053086  | 1,63     | 1479                                     | 0,21   | 354593   | 0,53     |
|                  | Retrofit  | 672                                | 0,09   | 150518   | 0,21     | 441                                   | 0,06   | 95757    | 0,15     | 554                                      | 0,08   | 120451   | 0,18     |
|                  | Tork      | 2175                               | 0,30   | 463608   | 0,66     | 1529                                  | 0,21   | 335219   | 0,52     | 2048                                     | 0,29   | 514129   | 0,77     |
| Total Copia      |           | 7426                               | 1,03   | 1707830  | 2,41     | 6279                                  | 0,85   | 1515504  | 2,35     | 4233                                     | 0,61   | 1023363  | 1,53     |
| Gypsy            | Del       | 284583                             | 39,66  | 54492294 | 77,05    | 243622                                | 32,81  | 46980644 | 72,86    | 57195                                    | 8,21   | 11962770 | 17,89    |
|                  | Reina     | 732                                | 0,10   | 145860   | 0,21     | 521                                   | 0,07   | 108614   | 0,17     | 571                                      | 0,08   | 122132   | 0,18     |
|                  | CRM       | 3208                               | 0,45   | 626430   | 0,89     | 3064                                  | 0,41   | 606528   | 0,94     | 11411                                    | 1,64   | 2548213  | 3,81     |
|                  | Galadriel | 1415                               | 0,20   | 337975   | 0,48     | 3107                                  | 0,42   | 668432   | 1,04     | 1459                                     | 0,21   | 419793   | 0,63     |
|                  | Tat       | 17177                              | 2,39   | 4102662  | 5,80     | 16429                                 | 2,21   | 4041584  | 6,27     | 18729                                    | 2,69   | 5022947  | 7,51     |
|                  | Athila    | 11193                              | 1,56   | 2095490  | 2,96     | 13530                                 | 1,82   | 2497755  | 3,87     | 81772                                    | 11,74  | 22056512 | 32,98    |
| Total Gypsy      |           | 318308                             | 44,36  | 61800711 | 87,38    | 280273                                | 37,74  | 54903557 | 85,14    | 171137                                   | 24,56  | 42132367 | 62,99    |
| Bel              |           | 0                                  | 0,00   | 0        | 0,00     | 65                                    | 0,01   | 9239     | 0,01     | 0  | 0,00   | 0        | 0,00     |
| DIRS             |           | 0                                  | 0,00   | 0        | 0,00     | 0                                     | 0,00   | 0        | 0,00     | 0  | 0,00   | 0        | 0,00     |
| Total LTR-RTs    |           | 325734                             | 45,40  | 63508541 | 89,80    | 286617                                | 38,60  | 56428300 | 87,51    | 175370                                   | 25,17  | 43155730 | 64,52    |

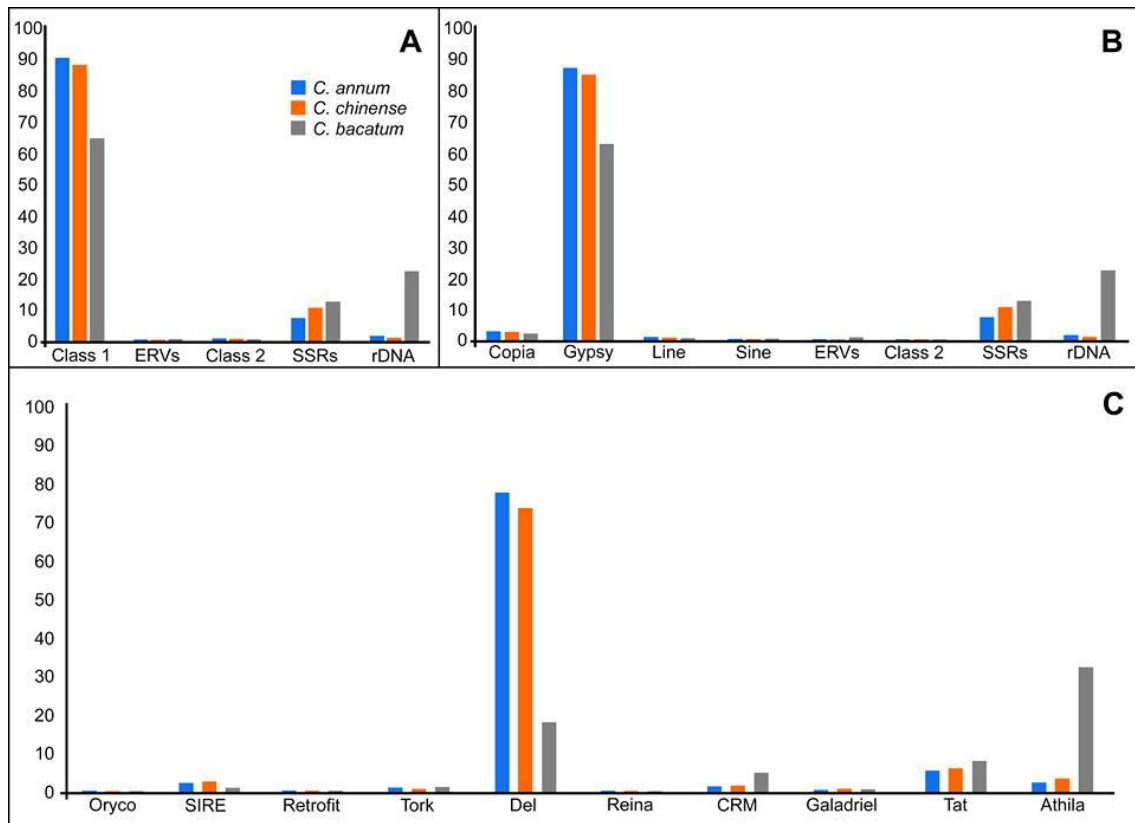
| Species/Lineages |                | <i>C. annuum</i> (GC 35% - AT 65%) |        |          |          | <i>C. chinense</i> (GC 34.9% - 65.1%) |        |          |          | <i>C. baccatum</i> (GC 35.4% - AT 64.6%) |        |          |          |
|------------------|----------------|------------------------------------|--------|----------|----------|---------------------------------------|--------|----------|----------|--|--------|----------|----------|
|                  |                | num seq                            | NR (%) | size bp  | size CR% | num seq                               | NR (%) | size bp  | size CR% | num seq                                  | NR (%) | size bp  | size CR% |
| LINE             | L1             | 1517                               | 0,21   | 140820   | 0,20     | 658                                   | 0,09   | 223482   | 0,35     | 254                                      | 0,04   | 102118   | 0,15     |
|                  | RTE            | 1034                               | 0,14   | 352148   | 0,50     | 820                                   | 0,11   | 122990   | 0,19     | 602                                      | 0,09   | 143627   | 0,21     |
| SINE             |                | 150                                | 0,02   | 9466     | 0,01     | 70                                    | 0,01   | 4214     | 0,01     | 88                                       | 0,01   | 5893     | 0,01     |
| Total            | Non-LTR-RTs    | 2701                               | 0,38   | 502434   | 0,71     | 1548                                  | 0,21   | 350686   | 0,54     | 944                                      | 0,14   | 251638   | 0,38     |
| Total Class 1    |                | 328435                             | 45,77  | 64010975 | 90,51    | 288165                                | 38,80  | 56778986 | 88,05    | 176314                                   | 25,30  | 43407368 | 64,90    |
| ERVs             | Caulimovirus   | 255                                | 0,04   | 28831    | 0,04     | 71                                    | 0,01   | 8652     | 0,01     | 154                                      | 0,02   | 18539    | 0,03     |
|                  | Soymovirus     | 14                                 | 0,00   | 891      | 0,00     | 0                                     | 0,00   | 0        | 0,00     | 6  | 0,00   | 449      | 0,00     |
|                  | Cavemovirus    | 21                                 | 0,00   | 2560     | 0,00     | 3                                     | 0,00   | 333      | 0,00     | 7  | 0,00   | 565      | 0,00     |
|                  | Tungrovirus    | 0                                  | 0,00   | 0        | 0,00     | 0                                     | 0,00   | 0        | 0,00     | 0  | 0,00   | 0        | 0,00     |
|                  | Badnavirus     | 151                                | 0,02   | 19580    | 0,03     | 27                                    | 0,00   | 3195     | 0,00     | 57                                       | 0,01   | 7113     | 0,01     |
|                  | Pararetrovirus | 181                                | 0,03   | 41158    | 0,06     | 33                                    | 0,00   | 6684     | 0,01     | 460                                      | 0,07   | 132639   | 0,20     |
| Total ERVs       |                | 622                                | 0,09   | 93020    | 0,13     | 134                                   | 0,02   | 18864    | 0,03     | 684                                      | 0,10   | 159305   | 0,24     |
| Transposons      | CACTA          | 136                                | 0,02   | 18564    | 0,03     | 114                                   | 0,02   | 15572    | 0,02     | 89                                       | 0,01   | 13619    | 0,02     |
|                  | hAT            | 681                                | 0,09   | 84556    | 0,12     | 345                                   | 0,05   | 55017    | 0,09     | 182                                      | 0,03   | 28114    | 0,04     |
|                  | PIF-Harbinger  | 12                                 | 0,00   | 824      | 0,00     | 9                                     | 0,00   | 1214     | 0,00     | 8  | 0,00   | 566      | 0,00     |



| Species/Lineages | <i>C. annuum</i> (GC 35% - AT 65%) |        |         |          | <i>C. chinense</i> (GC 34.9% - 65.1%) |        |         |          | <i>C. baccatum</i> (GC 35.4% - AT 64.6%) |        |         |          |
|------------------|------------------------------------|--------|---------|----------|---------------------------------------|--------|---------|----------|--|--------|---------|----------|
|                  | num seq                            | NR (%) | size bp | size CR% | num seq                               | NR (%) | size bp | size CR% | num seq                                  | NR (%) | size bp | size CR% |
| % mapped         | 0,00                               |        |         |          | 0,00                                  |        |         |          | 0,00                                     |        |         |          |

823 Comparação da fração repetitiva nos genomas de *Capsicum*. num seq – número de sequências encontradas após a rodada de  
824 Blast. NR (%) – valor relativo. Size pb – tamanho total que representa a classe repetitiva. Size CR (%) – porcentagem que a classe  
825 representa nos *scaffolds*.

1 Os elementos autônomos e não autônomos podem alcançar tamanhos longos em  
2 pares de bases (de ~1.000 a ~10.000). Considerando que a busca foi realizada por  
3 cores proteicas curtos e hiperconservados, em relação ao tamanho total que os  
4 retrotransposons podem alcançar, as porcentagens apresentadas aqui podem  
5 subestimar o tamanho relativo total dos elementos de Classe I presentes no três  
6 *datasets*. Mesmo assim, as regiões conservadas dos elementos de Classe I foram  
7 predominantes em relação às demais. *Capsicum annuum* teve 90,51% de elementos  
8 de Classe I, *C. chinense* apresentou valor próximo, com 88,05% enquanto *C.*  
9 *baccatum* teve 64,90% do *dataset* composto por elementos dessa classe (Tabela 2,  
10 Fig. 2A). Elementos pertencentes à Classe I foram subdivididos em sequências com  
11 LTR, sem LTR e ERVs (relacionados evolutivamente aos elementos com LTR,  
12 porém, sem as LTRs). Os elementos com LTR apresentaram valores próximos para  
13 as espécies filogeneticamente mais próximas, *C. annuum* e *C. chinense*  
14 apresentaram 89,90 e 87,51%, respectivamente, enquanto *C. baccatum* teve 64,52  
15 %. Quanto os elementos sem LTR, não houve uma clara relação entre valores  
16 obtidos e proximidade filogenética, 0,71 e 0,54% para *C. annuum* e *C. chinense*,  
17 respectivamente, e para *C. baccatum* 0,38. O mesmo foi observado para os ERVs,  
18 onde *C. annuum* teve 0,13% do *dataset* composto por essa fração, *C. chinense*  
19 0,03% e *C. baccatum* 0,24%. Considerando os dados dos três genomas analisados,  
20 as sequências de elementos LTR foram as mais abundantes, seguidas por aquelas  
21 sem LTR e dos ERVs (Tabela 2).



1  
2 **Figura 2:** Distribuição relativa (%) dos elementos repetitivos nos três *datasets* de  
3 *Capsicum*. **(A)** Distribuição relativa (%) das classes de elementos repetitivos. Note  
4 que os elementos de Classe I possuem maior proporção em relação aos demais. **(B)**  
5 Distribuição relativa (%) das superfamílias de elementos transponíveis,  
6 microsatélites, DNAr 5S e DNAr 35S. Note, maior acúmulo de elementos da  
7 superfamília *Gypsy* em relação aos demais. **(C)** Distribuição relativa (%) dos  
8 retrotransposons com LTR. Observe maior acúmulo de elementos da família Del nos  
9 genomas de *C. annuum* e *C. chinense* e maior acúmulo do clado Athila/Tat em *C.*  
10 *baccatum*.

1 Os elementos da superfamília *Gypsy* foram os mais acumulados nos três conjuntos  
2 de dados. *Capsicum annuum* e *C. chinense*, que são espécies filogeneticamente  
3 próximas, exibiram 87,49% e 85,37%, respectivamente, enquanto que *C. baccatum*  
4 exibiu 63,02% de sequências similares aos *Gypsy*. Retrotransposons da  
5 superfamília *Copia* apareceram menos acumulados nos *datasets*, sendo 2,41% (*C.*  
6 *annuum*), 2,35% (*C. chinense*) e 1,53% em *C. baccatum* (Tabela 2, Fig. 2B). Dentro  
7 da superfamília *Gypsy*, a família *Del* foi a predominante, com 77,15% e 73,05% em  
8 *C. annuum* e *C. chinense*, respectivamente, enquanto que em *C. baccatum* os  
9 elementos *Del* representaram apenas 17,89%. A família *Athila* foi mais  
10 representativa em *C. baccatum* (32,99%), enquanto que *C. annuum* (2,97%) e *C.*  
11 *chinense* (3,88%) os valores foram menores e mais próximos entre si (Fig. 2C).  
12 Elementos das famílias *CRM* e *Galadriel* apresentaram uma representatividade  
13 menor (entre 0,45% a 3,81%) quando comparadas com *Del*, sempre com valores  
14 mais próximos entre *C. annuum* e *C. chinense* em comparação a *C. baccatum*  
15 (Tabela 2, Fig. 2C).

16 Em relação aos outros elementos de origem retroviral, os retrotransposons sem  
17 LTRs foram pouco acumulados, com valores abaixo de 1% para as três espécies.  
18 *Capsicum annuum* apresentou 0,71%, *C. chinense* 0,54% enquanto *C. baccatum*  
19 0,38%. Desses, os elementos *LINE* foram os mais representativos, enquanto os  
20 *SINE* representaram 0,01% nos três genomas (Tabela 2). Os ERVs, diferentemente  
21 dos retrotransposons, foram estimados com base em uma compilação dos genomas  
22 virais completos da família *Caulimoviridae*, obtidas nos repositórios *online*  
23 (*Descriptions of Plant Viruses – DPVWEB*, *GiriRepbase*, *Gypsy Database* e *National*  
24 *Center of Biotechnology Information*). Os *datasets* dos genomas de *C. annuum*, *C.*  
25 *chinense* e *C. baccatum* exibiram sequências similares aos elementos do gênero

1 *Pararetrovirus*, sendo 0,06% em *C. annuum*, enquanto que em *C. chinense* e *C.*  
2 *baccatum* foram encontradas apenas 0,01% e 0,2%, respectivamente (Tabela 2). A  
3 fração mais representativa nos três genomas foi a que corresponde aos elementos  
4 endógenos conhecidos como *Tobacco vein clearing vírus*.

5 De maneira geral, os elementos de Classe II foram pouco acumulados nos *datasets*  
6 analisados, sem grande diferença entre as três espécies. *C. annuum* e *C. chinense*  
7 apresentam porcentagens próximas de transposons do clado hAT (0,13% e 0,09%,  
8 respectivamente) e para *C. baccatum*, o valor observado foi mais baixo, 0,04%. Os  
9 três *datasets* analisados apresentaram valores parecidos para os elementos  
10 CACTA, onde *C. annuum* obteve 0,03%, enquanto *C. chinense* e *C. baccatum*  
11 0,02%. Os elementos MuDR apresentaram valores próximos aos CACTA, 0,04%,  
12 0,02% e 0,03%, para *C. annuum*, *C. chinense* e *C. baccatum*, respectivamente.  
13 Helitrons também foram pouco acumulados em *C. annuum* e *C. chinense*  
14 apresentando o mesmo valor, 0,02%, contudo, esses elementos não foram  
15 amostrados em *C. baccatum* (Tabela 2).

16

## 17 **Seleção de sequências da transcriptase reversa e** 18 **produção de *primers***

19 A partir dos elementos previamente anotados, 91 sequências da transcriptase  
20 reversa foram selecionadas com base nos *cores* conservados do *GypsyDB* e  
21 alinhadas no programa Geneious. Dessas, nove sequências consenso foram  
22 utilizadas para o desenho de *primers* (Tabela 3). Em todos casos foram obtidas  
23 bandas com o tamanho previsto, correspondendo aos retrotransposons das

1 superfamílias *Gypsy* (CRM, Galadriel, Reina, Del e Athila) e *Copia* (Sire, Oryco,  
2 Retrofit e Tork).

3

4 **Tabela 3:** Descrição geral dos *primers* desenhados para regiões conservadas da  
5 transcriptase reversa de retrotransposons de *Capsicum*.

| <b>Superfamília</b> | <b>Linhagens</b> | <b>Amplicon (pb)</b> | <b>Primers</b>                                   |
|---------------------|------------------|----------------------|--|
| <i>Gypsy</i>        | CRM              | 192                  | F: CCACCAACAAGATAACGG<br>R: CCATCCATTCATAGAGACC  |
|                     | Galadriel        | 332                  | F: CCGTGCTCTCAACAAAGTCA<br>R: GCTCCTCCATTGAGCTAC |
|                     | Reina            | 286                  | F: CTTGAAGAATTGCATGGC<br>R: CCAGATGATCCTTCCAACTC |
|                     | Del              | 153                  | F: GTTCAGGGTGCCAAGTGT<br>R: GGGCGTTAGTCAACCTGAAG |
|                     | Athila           | 236                  | F: GGGTGGTATTGCTTCTTGGA<br>R: GAATCACCTACCACAGAG |
| <i>Copia</i>        | Sire             | 120                  | F: GTCGAACTTGCTGAGGA<br>R: CTCGGAGATTGCTTGAGG    |

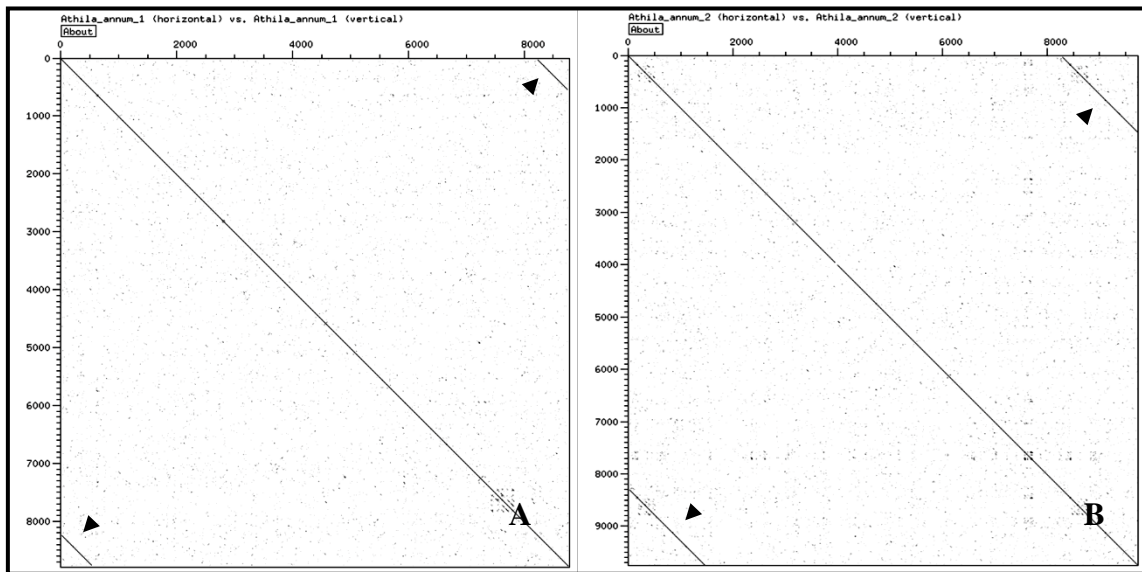
|          |     |   |
|----------|-----|---|
| Oryco    | 254 | F: GGTTCAAGGATCGGTTGATAG<br>R: GTTGAGCTTGCACACCATGT |
| Retrofit | 266 | F: GGTTACTGTCAGGAGTGTG<br>R: GAGTGAATCCAGCTTCAAG    |
| Tork     | 526 | F: CACAATGACCTGGTGTCGAT<br>R: CTCAATTCGCAAACCAAGGC  |

- 
- 1 Tamanho em pares de base do *amplicon* e sequências *forward* e *reverse* dos
  - 2 *primers* utilizados para amplificação dos elementos transponíveis das superfamílias
  - 3 *Gypsy* e *Copia*.
  - 4

## 1 Retrotransposons íntegros

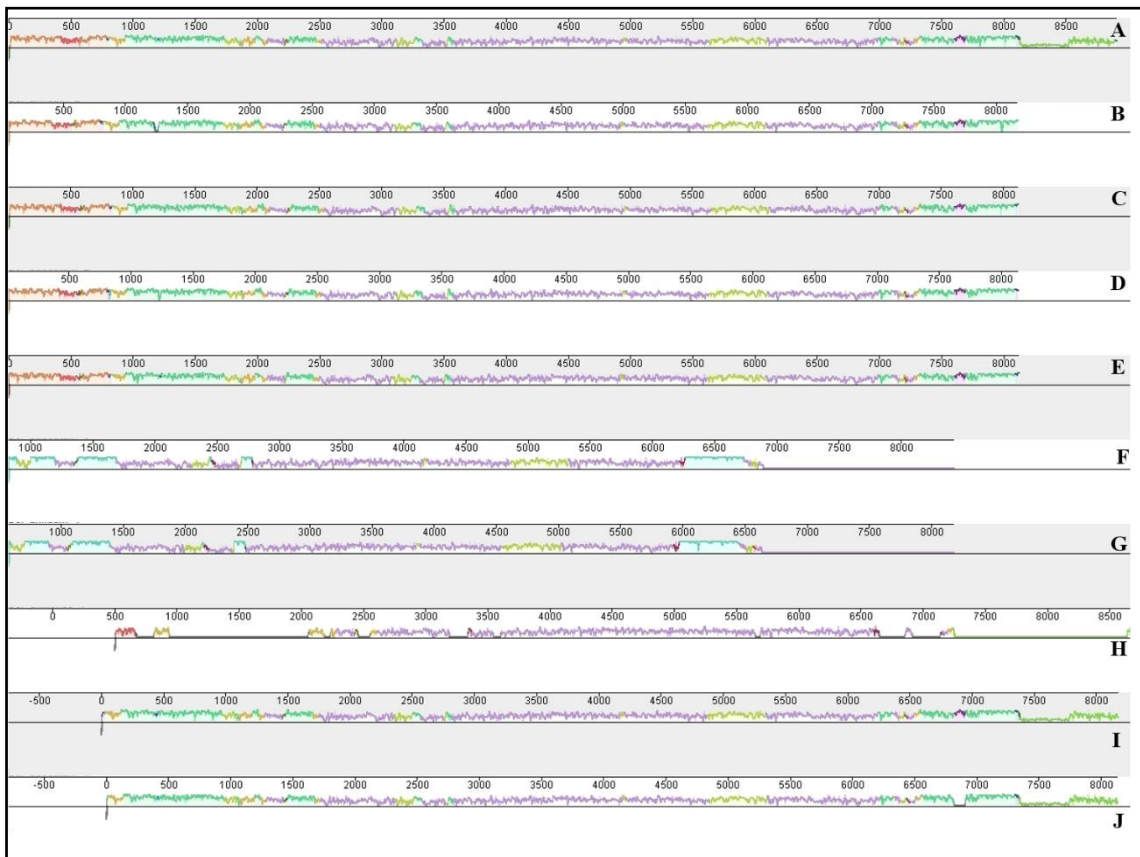
2 A busca por retrotransposons íntegros presentes nos *scaffolds* foi direcionada  
3 apenas para membros da superfamília *Gypsy*, famílias *Del*, *CRM* e *Athila/Tat*, pois  
4 foram as mais representativas nos três *datasets*. Para tal, foi utilizada a ferramenta  
5 *Dotter l/b* para buscar apenas sequências contendo ambas LTRs (Figura 4), seguida  
6 da localização do fragmento GAG e dos trechos das proteínas da cadeia poligênica,  
7 considerando a ordem dos genes aceitas para cada família de retrotransposon. Os  
8 três bancos de dados produziram 182 pseudomoléculas com mais de 3.500 pares de  
9 bases, porém, a maioria dessas foi identificada como não autônoma, já que faltavam  
10 ou LTR, ou GAG ou trechos da cadeia poligênica. Ao final foram selecionadas 8  
11 sequências pertencentes a família *CRM*, 10 de *Del* e 2 do clado *Athila/Tat*. Após  
12 duas rodadas de alinhamento com os programas *ClustalW* e *Mauve*, as sequências  
13 eleitas foram organizadas em grupos dentro de cada família . As sequências  
14 referentes à família *Del* foram organizadas em 5 grupos (Figura 5), variando de  
15 8105pb até 8902pb. Destes, o grupo 3 agrupou sequências de *C. annuum* e *C.*  
16 *chinense*, enquanto que grupos 1 e 4 foram exclusivos para *C. annuum* e os grupos  
17 2 e 5 foram exclusivos para *C. baccatum*. Os retrotransposons *CRM* foram  
18 separados em 3 grupos distintos (Figura 6), variando de 5259pb até 7328pb. Entre  
19 esses, apenas *C. annuum* apresentou um grupo exclusivo, os demais foram  
20 compartilhados entre as três espécies. As duas sequências pertencentes ao clado  
21 *Athila/Tat* foram distintas (Figura 7), formando cada uma um grupo. Essas  
22 sequências foram encontradas na montagem do genoma de *C. annuum*, sendo  
23 então exclusivas desta espécie. De cada grupo formado, uma sequência foi  
24 selecionada para a representação gráfica (Figura 8).

1



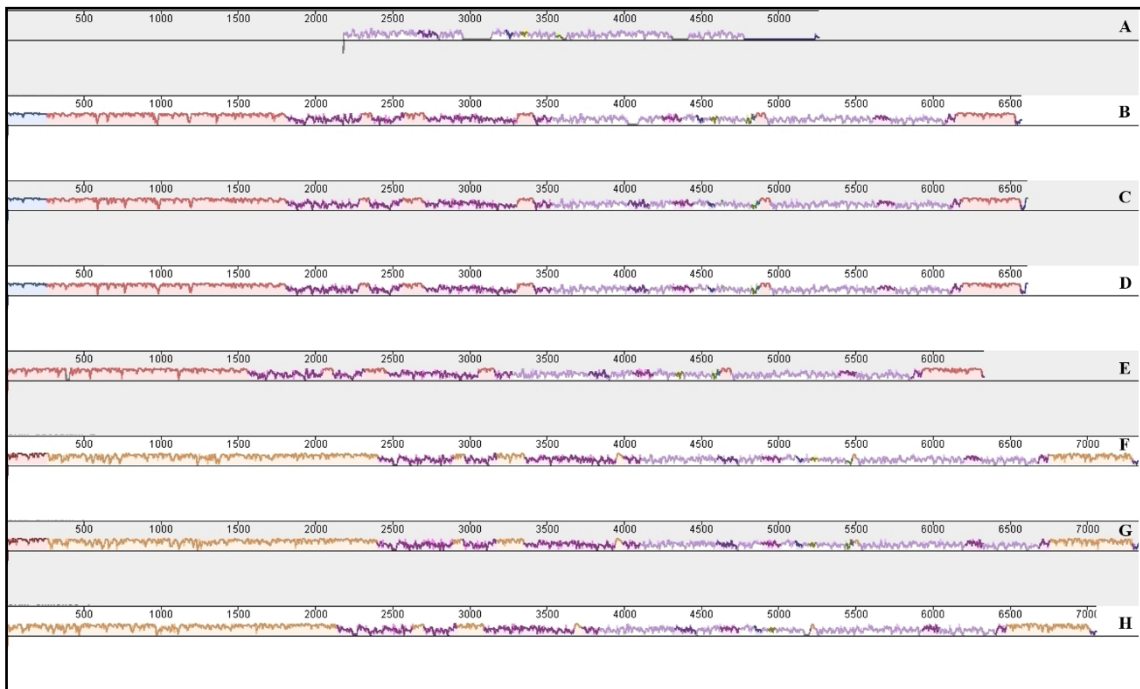
2

3 **Figura 3:** Dotter representando sequências com LTRs em dois retrotransposons do  
4 clado Athila/Tat, extraídas do *dataset* de *C. annum*. Note a diferença entre os  
5 tamanhos das LTRs (ponta de seta) nas duas sequências. **(A)** Athila\_annuum\_1 **(B)**  
6 Athila\_annuum\_2



1

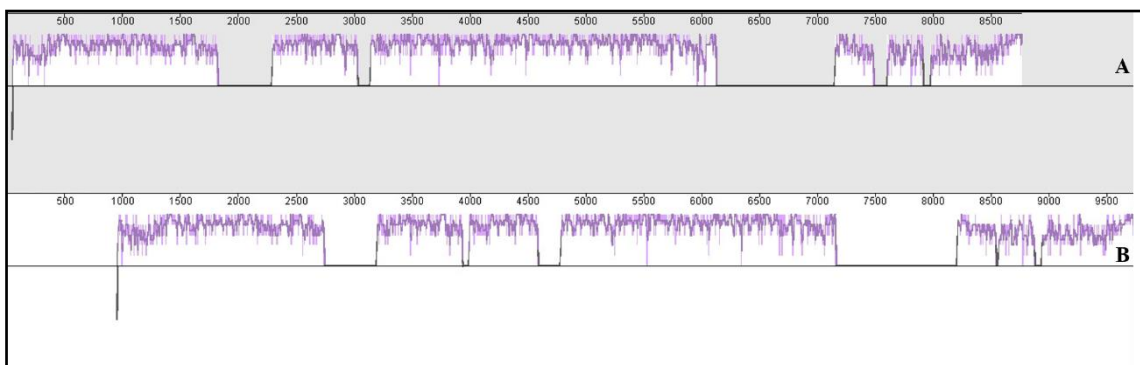
- 2 **Figura 4:** Alinhamento das sequências de retrotransposons com a ferramenta  
3 Mauve. São representadas as sequências recuperadas de Del a partir do *dataset* de  
4 genomas de *Capsicum*. **(A)** Del\_annuum\_2. **(B)** Del\_bacatum\_1. **(C)**  
5 Del\_bacatum\_2. **(D)** Del\_bacatum\_3. **(E)** Del\_bacatum\_5. **(F)** Del\_annuum\_1. **(G)**  
6 Del\_chinense\_1. **(H)** Del\_annuum\_3. **(I)** Del\_bacatum\_4. **(J)** Del\_bacatum\_6



1

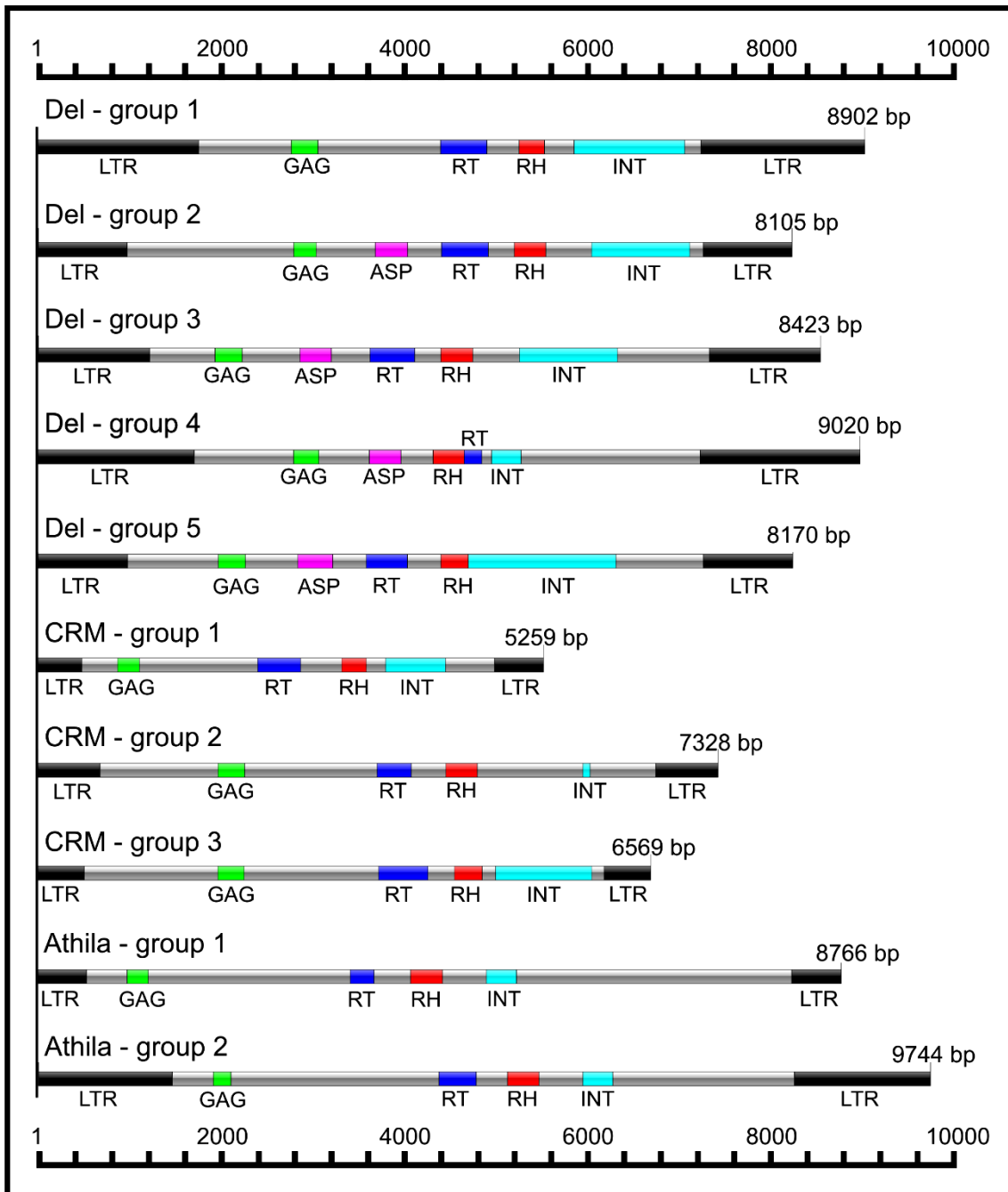
2 **Figura 5:** Alinhamento das seqüências de retrotransposons com a ferramenta  
 3 Mauve. São representadas as seqüências recuperadas de CRM a partir do *dataset*  
 4 de genomas de *Capsicum*. **(A)** CRM\_annuum\_1. **(B)** CRM\_annuu\_2. **(C)**  
 5 CRM\_annuum\_4. **(D)** CRM\_chinense\_2. **(E)** CRM\_baccatum\_2. **(F)**  
 6 CRM\_annuum\_3. **(G)** CRM\_chinense\_1. **(H)** CRM\_baccatum\_1.

7



8

- 1 **Figura 6:** Alinhamento das sequências de retrotransposons com a ferramenta  
 2 Mauve. São representadas as sequências recuperadas de Athila a partir do *dataset*  
 3 de genomas de *Capsicum*. **(A)** Athila\_annuum\_1. **(B)** Athila\_annuum\_2.



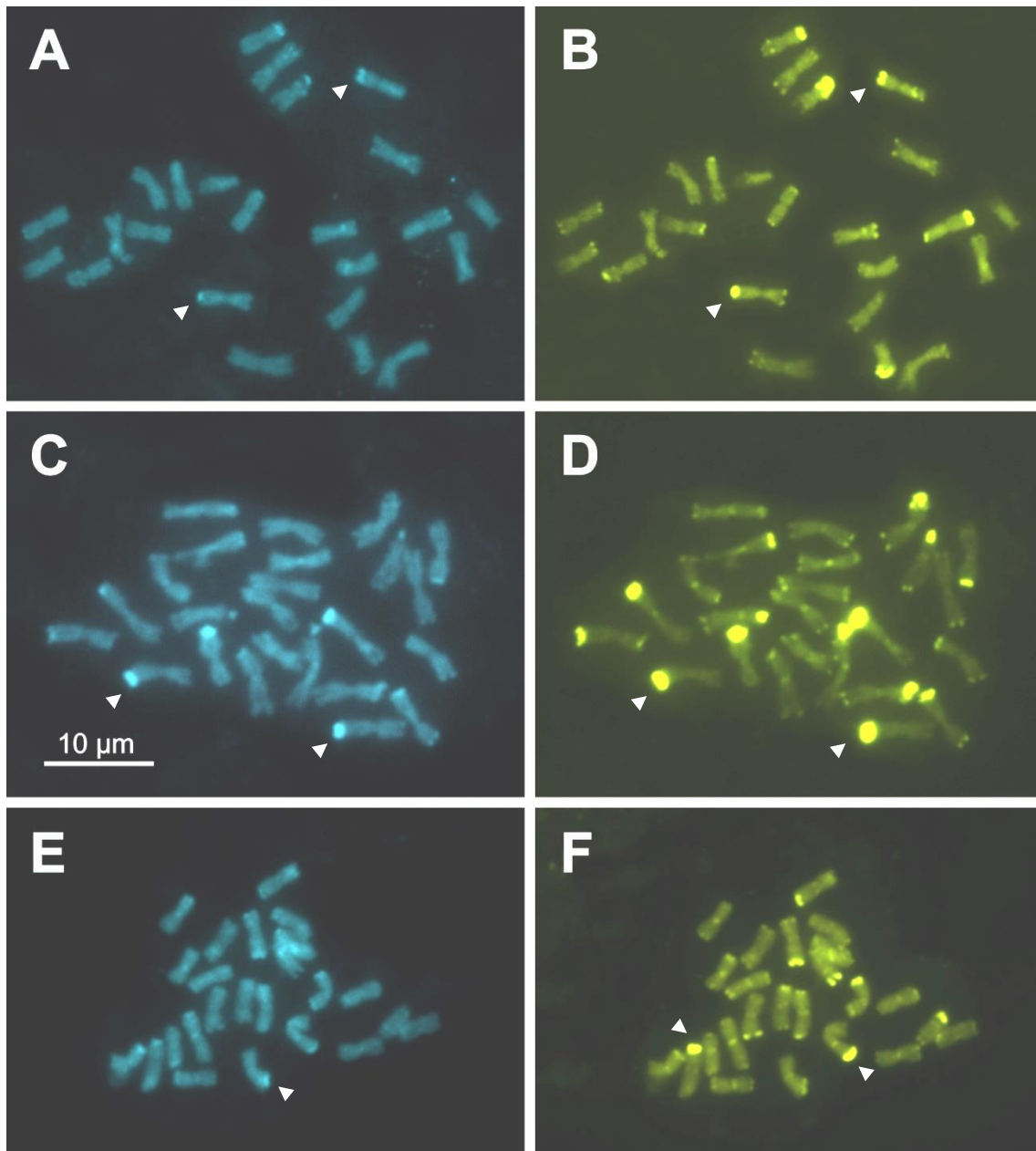
- 4
- 5 **Figura 7:** Representação dos grupos de elementos de transposição. LTR – *long*  
 6 *terminal repeat*, GAG – proteína capsidial, RT – *reverse transcriptase*, RH – *RNAse*  
 7 H, INT – *integrase*, ASP - aspartase. Note a diferença na extensão em pares de

1 base dos grupos formados, tamanho das LTRs e da posição das proteínas ao longo  
2 dos elementos evidenciando o agrupamento dentro da superfamília *Gypsy*.  
3 Tamanho das integrases faz referência ao tamanho do alinhamento obtido, tal  
4 variação não remonta ao tamanho real do gene, assim como nos demais domínios.

5

## 6 **Citogenética comparativa**

7 Os resultados do bandejamento C-CMA/DAPI mostraram que em *C. annuum*  
8 ocorrem três pares com bandas CMA<sup>+</sup> mais intensas adjacentes às bandas DAPI<sup>+</sup>  
9 de menor tamanho. Bandas terminais CMA<sup>+</sup> menores foram vistas em 11 pares  
10 cromossomos. Bandas intersticiais CMA<sup>+</sup> foram observadas em três pares, o mesmo  
11 número que as DAPI<sup>+</sup>, mas em apenas um par desses a banda foi CMA<sup>+</sup>/DAPI<sup>+</sup>  
12 (Figura 13, A-B).



1

2 **Figura 8:** Bandeamento C – CMA/DAPI em metáfases de espécies do gênero  
 3 *Capsicum*. **(A, B)** *C. annuum*. **(A)** Pontas de seta indicam bandas terminais DAPI<sup>+</sup>.  
 4 **(B)** Pontas de seta indicam bandas terminais intensas CMA<sup>+</sup>, setas indicam bandas  
 5 intersticiais CMA<sup>+</sup>. **(C, D)** *C. chinense* **(C)** Pontas de setas indicam bandas terminais  
 6 intensas DAPI<sup>+</sup>. **(D)** Pontas de setas indicam bandas terminais intensas CMA<sup>+</sup>. **(E, F)**  
 7 *C. baccatum*. **(E)** Pontas de setas indicam bandas terminais e setas indicam bandas

1 intersticiais DAPI<sup>+</sup>. **(F)** Pontas de seta indicam bandas terminais intensas e setas  
2 indicam bandas pericentroméricas.

3

4 *Capsicum chinense* exibiu um número bem maior de pares com bandas CMA<sup>+</sup> mais  
5 intensas (7 pares) e desses, 2 pares tiveram bandas DAPI<sup>+</sup> intensas adjacentes.  
6 Bandas terminais CMA<sup>+</sup> menores foram vistas em 11 pares cromossomos. Bandas  
7 proximais na forma de *dots* centroméricos foram vistas em 9 pares, muitos dos quais  
8 também foram DAPI<sup>+</sup>. Bandas intersticiais tênues foram marcadas com DAPI, mas  
9 não com CMA (Figura 13, C-D).

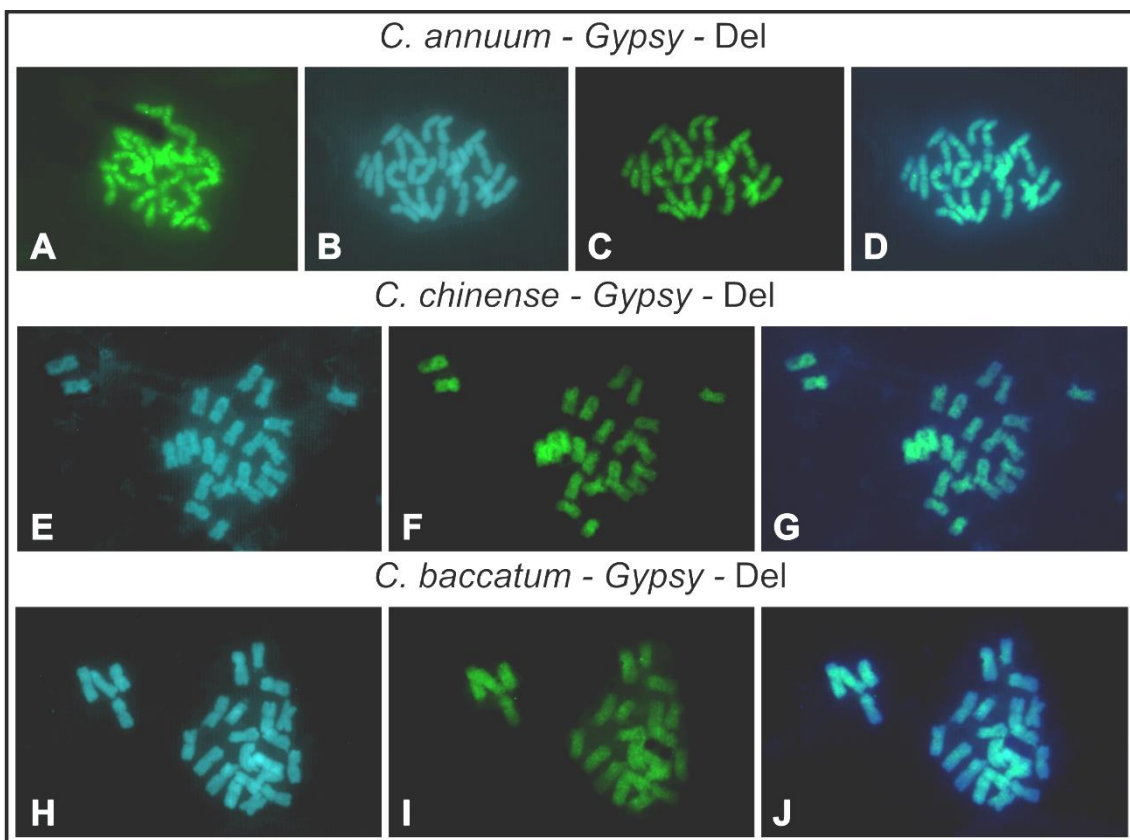
10 Assim como em *C. annuum* e *C. chinense*, *C. baccatum* também apresentou bandas  
11 terminais mais intensas (Figura 13, E-F), em três pares essas bandas foram  
12 CMA<sup>+</sup>/DAPI<sup>+</sup>, enquanto em um foi apenas CMA<sup>+</sup>. Bandas terminais mais tênues  
13 CMA<sup>+</sup> foram observadas em todos os cromossomos em pelo menos um dos braços.  
14 Bandas proximais na forma de *dots* centroméricos puderam ser observadas em 11  
15 pares, das quais apenas 3 também foram DAPI<sup>+</sup> (Figura 13, E- F).

16

## 17 **Hibridização *in situ***

18 As sondas das linhagens de *Gypsy* foram marcadas com biotina e detectadas com  
19 avidina-FITC (verde), e de *Copia* foram marcadas diretamente com Cy3-dUTP  
20 (vermelho). Para a superfamília *Gypsy* foram utilizadas sondas das famílias CRM e  
21 Del. A partir das hibridizações realizadas utilizando as sondas produzidas para a  
22 família Del, de modo geral pode-se observar um perfil de sinais predominantemente  
23 disperso nos três genomas, contudo, sinais mais intensos foram observados em *C.*

1 *annuum* (Figura 9, A-D) e *C. chinense* (Figura 9, E-G), em comparação a *C.*  
 2 *baccatum* (Figura 9, H-J). Em *C. annum* quatro pares cromossômicos  
 3 apresentaram sinais mais intensos em um dos braços (Figura 9D), *C. chinense*  
 4 também apresentou quatro pares com sinais mais intensos, nos demais  
 5 cromossomos dessas duas espécies houveram sinais sem presença de *dots*. Em *C.*  
 6 *baccatum*, assim como em *C. annum* e *C. chinense* foi possível observar maior  
 7 acúmulo de sinais em quatro pares, contudo nos demais pares do conjunto  
 8 cromossômico os sinais foram bem menos intensos. Estes resultados corroboram  
 9 com os resultados observados nas análises de bioinformática, que demonstraram  
 10 maior acúmulo de sequências da família Del em *C. annum* e *C. chinense*, enquanto  
 11 que em *C. baccatum* esse acúmulo não foi tão expressivo.



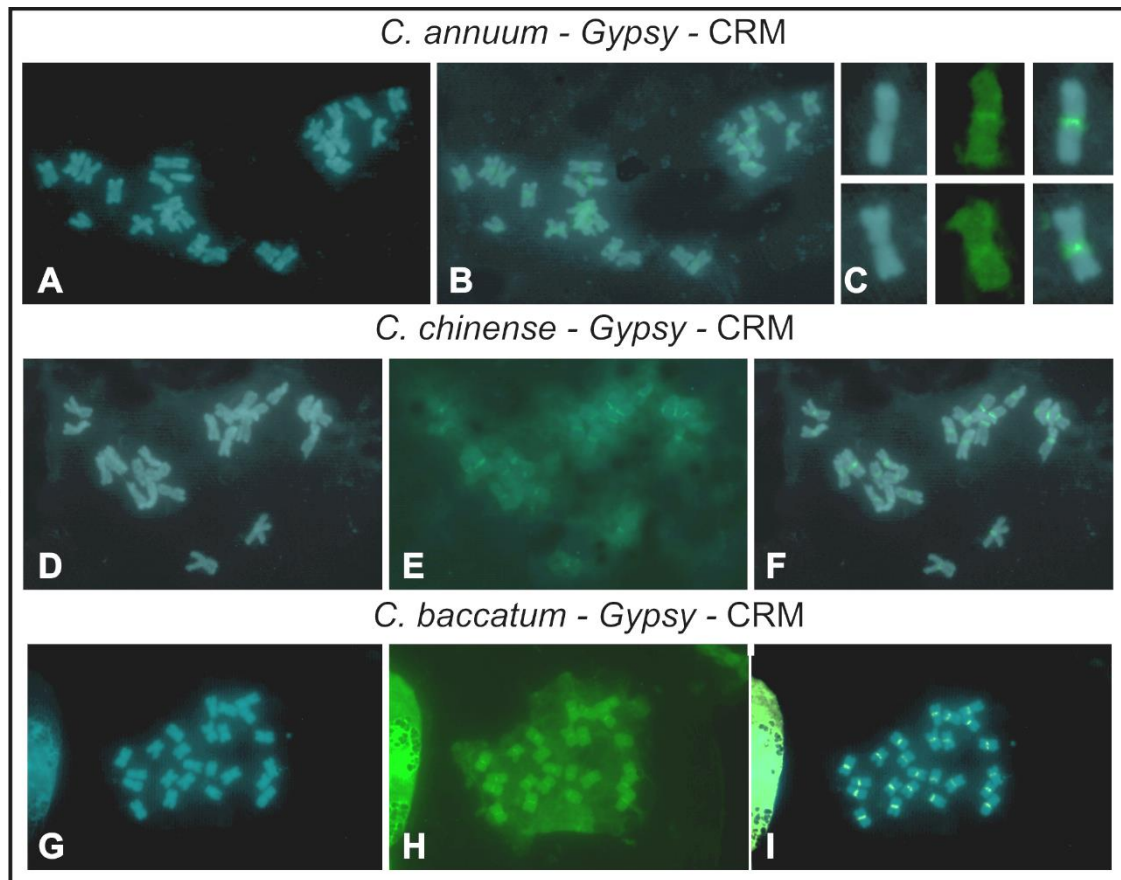
12

13 **Figura 9:** Hibridização *in situ* fluorescente com metáfases coradas com DAPI (azul),  
 14 sonda de Del detectada com avidina-FITC (verde), em três espécies de *Capsicum*.

1 **(A-D)** *C. annuum* com sinais dispersos apresentando maior concentração de sinais  
2 em relação às demais espécies. **(E-G)** *C. chinense* com sinais dispersos **(H-J)** *C.*  
3 *baccatum* apresentando sinais menos intensos.

4

5 As hibridizações realizadas com as sondas produzidas para a família CRM  
6 mostraram acúmulo de maneira muito similar entre as três espécies, em regiões  
7 proximais de todos os cromossomos (Figura 10). Contudo, os sinais observados em  
8 *C. baccatum* (Figura 10, G-I) foram mais intensos, quando comparados com *C.*  
9 *annuum* (Fig. 7, A e B) e *C. chinense* (Figura 10, D-F). *C. annuum* foi a única  
10 espécie onde um dos pares não apresentou sinais de hibridização com a sonda de  
11 CRM, nos demais pares os sinais foram observados foram similares quanto a  
12 intensidade. Similar a *C. annuum*, *C. chinense* apresentou sinais em todos os  
13 cromossomos, com variação quanto a intensidade em alguns. Entre as três  
14 espécies, *C. baccatum* foi a que apresentou maior acúmulo de sinais, dos quais em  
15 apenas dois pares foram menos intensos que os demais do mesmo conjunto  
16 cromossômico. Tal variação quanto ao acúmulo de sinais e intensidade, indica clara  
17 diferença no acúmulo de LTR-RTs da família CRM nos cromossomos das três  
18 espécies analisadas.

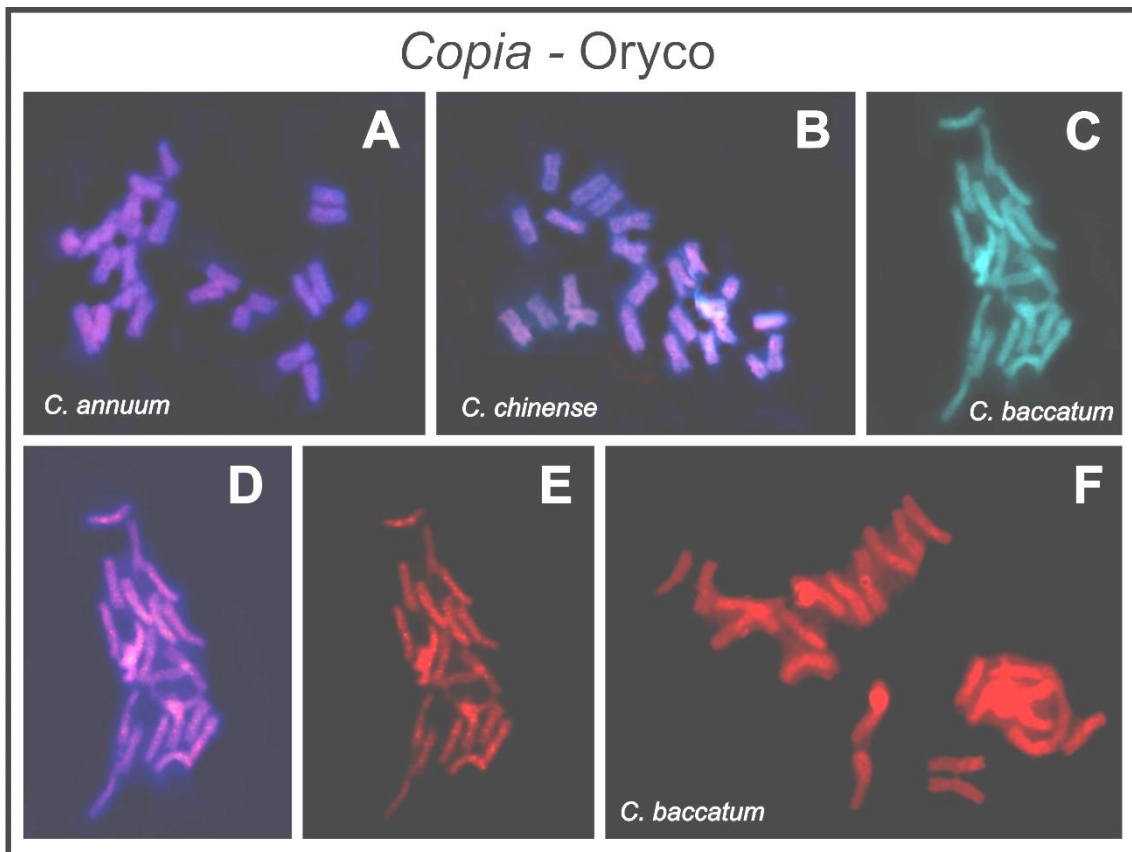


1

2 **Figura 10:** Hibridização *in situ* fluorescente com metáfases coradas com DAPI  
 3 (azul), sonda de CRM detectada com avidina-FITC (verde), em três espécies de  
 4 *Capsicum*. **(A e B)** *C. annuum* com sinais concentrados na região pericentromérica.  
 5 **(C)** Detalhe do sinal da sonda de CRM evidenciando a localização pericentromérica  
 6 do elemento de transposição. **(D-F)** *C. chinense*. **(G-I)** *C. baccatum*.

7

8 Para a superfamília *Copia*, as hibridizações realizadas com a família *Oryco*, foi  
 9 possível observar que não ocorreu, de maneira geral, formação de blocos tampouco  
 10 presença de sinais intensos, sugerindo baixa atividade desses elementos nos  
 11 genomas de *Capsicum* (Figura 11).



1

2 **Figura 11:** Hibridização in situ fluorescente com metáfases e prometáfases coradas  
 3 com DAPI (azul), sonda de Oryco detectada com Cy3-dUTP (vermelho), em três  
 4 espécies de Capsicum. **(A)** *C. annuum*. **(B)** *C. chinense*. **(C)** *C. baccatum*.

5

# 1 Discussão

2

## 3 **Fração repetitiva nos genomas de *Capsicum***

4 Nossas análises comparativas mostraram que a fração de elementos repetitivos  
5 pode impactar na composição dos genomas em *Capsicum*, especialmente porque  
6 retrotransposons são mais dinâmicos, por terem ocupado mais de 80% dos  
7 genomas, como mostra a predominância de *Gypsy*. Em tomate (*Solanum*  
8 *lycopersicum*), por exemplo, os elementos da superfamília *Gypsy* foram mais  
9 frequentes que os *Copia*, embora tal diferença seja pequena (JOUFFROY *et al.*,  
10 2016). Porém, quando PAZ *et al.* (2017) avaliaram os elementos transposição  
11 completos das superfamílias *Copia* e *Gypsy* em *Solanum lycopersicum*, notaram que  
12 os elementos das famílias Del de *Gypsy* e Tork de *Copia* foram predominantes.  
13 Quando comparamos os dados obtidos por PAZ *et al.* (2017) [36] com os nossos  
14 resultados em *Capsicum*, fica evidente que os elementos de *Copia* foram pouco  
15 representativos nos *datasets* de *Capsicum* e que possivelmente houve uma  
16 expansão de elementos da superfamília *Gypsy* em relação às outras famílias  
17 repetitivas. Considerando que os genomas de *Capsicum* são cerca de três vezes  
18 maiores que os de *Solanum* (PARK *et al.*, 2011; PARK *et al.*, 2012), ambos com  
19 predominância do número somático 24, sugerimos que essa variação não foi  
20 decorrente da poliploidia, mas do acúmulo ou perda diferencial de famílias  
21 repetitivas. Em *Nicotiana attenuata*, por exemplo, elementos com LTR também foram  
22 mais abundantes, representando 81% do genoma. Contudo, há contrastes dentro do  
23 mesmo gênero, já que em *N. obtusifolia* apenas 64% de elementos foram com LTR

1 Xu *et al.* (2017). Esse fato também foi observado entre as espécies aqui utilizadas,  
2 onde *C. annuum* e *C. chinense* apresentaram cerca de 90% e 88% de elementos  
3 LTR, respectivamente, enquanto que *C. baccatum* apresentou cerca de 65% do  
4 genoma com esses elementos. Essas características acompanham o agrupamento  
5 filogenético das espécies de *Capsicum*, onde *C. annuum* e *C. chinense* fazem parte  
6 do mesmo clado (Annuum), enquanto que *C. baccatum* pertence ao clado Baccatum  
7 (CARRIZO-GARCIA *et al.*, 2016). Contudo, é importante ressaltar que em *C.*  
8 *baccatum* houve um acúmulo de sequências de DNAr 35S levando a uma diferença  
9 de 20% em relação aos genomas de *C. annuum* e *C. chinense*. Esse evento poderia  
10 também explicar os contrastes nas porcentagens de elementos de Classe I entre  
11 esses três genomas, quando todos os repetitivos são analisados de maneira  
12 conjunta.

13 O clado Annuum apresentou valor aproximado de retrotransposons das famílias Del,  
14 CRM, Athila, Tat e Sire, em relação ao clado Baccatum. Contudo, quando  
15 considerados os elementos sem LTR, ERVs e transposons, essa diferença não tão  
16 foi evidente. Isso sugere que os retrotransposons foram mais ativos no processo de  
17 diferenciação genômica entre os cladros, do que os elementos ERVs e de Classe II.  
18 Atividade diferencial de retrotransposons entre genomas próximos também foi  
19 reportada para os gêneros *Helianthus* (STATON *et al.*, 2012) [37] e *Solanum* (DI  
20 FILIPPO *et al.*, 2012; BOLGER *et al.*, 2014) [38,39], concordando com o resultado  
21 em *Capsicum*. Nossos dados também mostraram pequenas diferenças entre *C.*  
22 *annuum* e *C. chinense*, especialmente no acúmulo de Del, ERVs e Line-RTE,  
23 sugerindo que outros elementos além dos LTR-RTs possuem algum papel do  
24 processo de diferenciação genômica. Além disso, nossos dados suportam a ideia de  
25 que os elementos repetitivos, principalmente os de transposição, funcionam de

1 maneira independente nos genomas vegetais. Em *Nicotiana*, por exemplo,  
2 retrotransposons Tnt1 evoluíram e se acumularam de maneira independente em  
3 diferentes espécies desse gênero (VERNHETTES, GRANDBASTIEN E  
4 CASACUBERTA, 1998; MELAYAH *et al.*, 2004) [40,41].

5 Considerando apenas os retrotransposons com LTR, nossos dados mostraram que  
6 apenas 0,002% do *dataset* de *C. annuum*, 0,001% de *C. chinense* e 0,004% de *C.*  
7 *baccatum* corresponderam à elementos possivelmente autônomos, sendo o restante  
8 fragmentos de elementos não autônomos. É importante ressaltar que os três bancos  
9 de dados foram gerados por sequenciamento de altíssima cobertura (KIM *et al.*,  
10 2014; QIN *et al.*, 2014) [15,17], o que favoreceu a montagem de  
11 pseudocromossomos, assim como de elementos completos. Segundo LISCH (2013)  
12 [6], a maior porção das sequências repetitivas codantes são relacionadas à  
13 fragmentos de elementos não autônomos, que podem ser amplificados em função  
14 da atividade dos autônomos. Isso explicaria a alta porcentagem de fragmentos de  
15 retrotransposons nos três *datasets*, em relação aos potencialmente autônomos. Dos  
16 cinco grupos de retrotransposons autônomos de Del, dois foram exclusivos do clado  
17 *Baccatum* e três do Clado *Annuum*. Nossos resultados sugerem que o acúmulo de  
18 retrotransposons da família Del aconteceu de modo independente em relação aos  
19 membros de outras famílias de LTR-RTs, além da possibilidade desses elementos  
20 Del terem sido mais ativos do que os demais, quando comparados aos de CRM e  
21 *Athila/Tat*.

22 Quando comparamos os valores quantitativos dos retrotransposons centroméricos,  
23 CR (família CRM), *C. baccatum* foi o que apresentou maior valor, quase quatro  
24 vezes mais do que nas outras duas espécies. Esses dados reforçam a diferença

1 entre os clados Annuum e Baccatum, assim como observado com os elementos Del.  
2 Dos três grupos de CRM, apenas um foi exclusivo de *C. annuum* e dois restantes  
3 foram compartilhados pelos três genomas. Exceto pelo grupo exclusivo de *C.*  
4 *annuum*, o compartilhamento pode ter alguma relação com o fato dos  
5 retrotransposons centroméricos serem residentes preferencialmente na região  
6 proximal bem como da importância desses elementos para a composição estrutural  
7 dos centrômeros. Neumann *et al.* (2011) classificou os CR em três grupos, aqueles  
8 com CR *motif*, os sem PTD (*putative targeting domain*) e os com cromodomínios. Ao  
9 que parece, os três grupos de CRM encontrados em *Capsicum* fazem referência ao  
10 grupo sem PTD já descritos na literatura. Os elementos sem PTD teoricamente não  
11 possuem associação preferencial com os centrômeros, contudo, esses elementos  
12 acumularam preferencialmente nas regiões centroméricas em *Musa* (NEUMANN *et*  
13 *al.*, 2011), assim como nos cromossomos de *Capsicum* detectados pela FISH.

14 O caso mais interessante foi o do clado Athila/Tat, pois foram os mais acumulados  
15 no *dataset* de *C. baccatum*, porém, sem sequências potencialmente autônomas  
16 nesse genoma. Como as sequências íntegras foram recuperadas exclusivamente no  
17 *dataset* de *C. annuum*, as explicações mais plausíveis podem ser a baixa frequência  
18 em *C. baccatum* decorrente da montagem dos genomas, o acúmulo diferencial de  
19 membros desse clado, ou menos provável degeneração desses retrotransposons  
20 em *C. baccatum*. Em *Gossypium*, Hawkins *et al.* (2009) demonstraram que linhagens  
21 de LTR-RTs diferentes apresentaram momentos de atividade diferentes ao longo da  
22 evolução dos genomas, o que promoveu a diferença de três vezes no conteúdo de  
23 DNA entre espécies diploides desse gênero. Em *Coffea*, o acúmulo de cópias de  
24 CRM em espécies diploides foi maior que no híbrido tetraploide (DE CASTRO  
25 NUNES *et al.* 2018), e isso levou os autores a sugerirem que mudanças na

1 quantidade de sequências de CRM aconteceram durante o processo de hibridização  
2 natural. Esse fato reforça a ideia do acúmulo diferencial, bem como da atividade  
3 independente de cada família de retrotransposons entre genomas próximos. Para  
4 que um elemento possa ser considerado autônomo ele necessita possuir todos os  
5 trechos da cadeia poligênica, apresentar ambas as LTRs e possuir mais que 3.500  
6 pares (KUMAR E BENNETZEN, 1999), na falta de alguma região da cadeia  
7 poligênica este é então considerado não autônomo. A diferença de 87,27% entre o  
8 número de sequências potencialmente íntegras (tamanho em pares de base e  
9 presença das LTRs) em *Capsicum* e as consideradas íntegras, sugere que os  
10 retrotransposons passam por diferentes eventos de degeneração nesses três  
11 genomas, como por exemplo transcrição incompleta, recombinação entre TEs e  
12 reparo na dupla-fita (BERGMAN EQUESNEVILLE, 2007) [42].

13

## 14 **LRT-RTs e localização física nos cromossomos**

15 Dados presentes na literatura mostram que elementos da família Del tendem a ser  
16 mais acumulados em espécies da família Solanaceae, como por exemplo em tomate  
17 e nas pimentas (PARK *et al*, 2011) [43]. Tais diferenças foram compatíveis com as  
18 hibridizações realizadas com a sonda da transcriptase reversa de Del em *C. annuum*  
19 e *C. chinense*, quando comparados com os sinais nos cromossomos de *C.*  
20 *baccatum*. Com base nas hibridizações realizadas nos cromossomos número 2 de  
21 tomate e pimenta, Park *et al.* (2011) [43] demonstraram que *Capsicum annuum*  
22 possui perfil disperso de Del, com sinais mais intensos ao longo dos cromossomos,  
23 enquanto que em *Solanum lycopersicum* a quantidade de sinais foi em menor  
24 quantidade, embora os sinais também tenham sido dispersos. Diferentemente dos

1 resultados observados em *Capsicum*, hibridizações realizadas por Santos *et al.*  
2 (2015) com sonda de Del mostraram sinais em forma de *dots* ao longo dos  
3 cromossomos de *Brachiaria brizantha*, mostrando que em grupos distintos o padrão  
4 de dispersão desses elementos pode variar.

5 A hibridização *in situ* com a sonda de CRM mostrou que *C. baccatum* teve sinais  
6 mais intensos na região cromossômica proximal em comparação com *C. annuum* e  
7 *C. chinense*, esses sinais foram compatíveis com as análises de bioinformática que  
8 mostraram maior acúmulo de CRM no genoma de *C. baccatum*. As três espécies de  
9 *Capsicum* mostraram acúmulo preferencial de CRM na região próxima ao  
10 centrômero com sinais muito fracos nas regiões intersticiais, contudo isso não é uma  
11 regra para espécies vegetais. Santos *et al.* (2015) observaram a presença de alguns  
12 *dots* de CRM nas regiões intersticiais dos cromossomos de *Brachiaria brizantha*,  
13 bem como a presença de sinais proximais em outras espécies desse gênero, fato  
14 também reportado por Neumann *et al.* (2011) em cromossomos de Espruce-da-  
15 Noruega. A preferência de acúmulo de CRM em regiões centroméricas pode ser  
16 explicada pela presença do domínio *zinc finger* (HHCC), o qual possui afinidade com  
17 a proteína CENH3 presente nos centrômeros. A presença desse domínio é  
18 característica de elementos do clado dos cromovírus, no qual os elementos da  
19 família CRM estão inclusos (NEUMANN *et al.*, 2011) [44]. Os poucos  
20 retrotransposons centroméricos íntegros recuperados em *Capsicum* foram no grupo  
21 sem PTD, não exclusivos da região centromérica. Como os sinais da FISH  
22 predominaram nas regiões proximais, podemos sugerir que essas regiões podem  
23 conter fragmentos da transcriptase reversa referentes à elementos não autônomos  
24 de diferentes grupos de CRM. Com base em sequências isoladas de parentais  
25 diploides de *Nicotiana tabacum*, NAGAKI *et al.* (2011) [45] demonstraram que

1 sequências de CR podem ser espécie específicas ou ortólogas, isso suporta nossos  
2 resultados de FISH, já que a mesma sonda da transcriptase reversa de *C. annuum*  
3 hibridizou também em *C. chinense* e *C. baccatum*.

4 Os resultados obtidos pelas análises de bioinformática mostraram baixo acúmulo de  
5 elementos da família Oryco (*Copia*), esse resultado foi também observado pela  
6 baixa quantidade de sinais de hibridização nos cromossomos de *Capsicum*. De  
7 modo geral, os sinais foram dispersos e fração, sugerindo assim a baixa atividade  
8 desses elementos nos genomas de *Capsicum*.

9 Hibridizações realizadas por SCALDAFERRO *et al.* (2016) com sondas de 35S e 5S  
10 em cromossomos de espécies de *Capsicum*, mostraram que *C. baccatum* possui  
11 maior quantidade de sítios de DNAr 35S em relação às demais espécies, com sinais  
12 nas pontas em quase todos os cromossomos, enquanto *C. annuum* apresentou dois  
13 pares e *C. chinense* quatro pares. Esses dados são compatíveis com nossa análise  
14 de bioinformática. Outros autores reportaram a presença de poucos sítios terminais,  
15 muitas vezes únicos, de DNAr 5S em várias espécies de *Capsicum* (PARK *et al.*,  
16 1999a; PARK *et al.*, 1999b; SCALDAFERRO, GABRIELE E MOSCONE, 2013;  
17 AGUILERA, DEBAT E GABRIELE, 2017). Nossos resultados de FISH com as  
18 sondas de retrotransposons não mostraram forte acumulação de sinais terminais,  
19 sugerindo que não há uma clara co-localização entre os LTR-RTs os DNAr em  
20 *Capsicum*. Essa mesma tendência foi observada quando comparamos os resultados  
21 das FISHs com o bandeamento CMA-DAPI, pois não houve co-localização de sinais  
22 fortes de FISH nas regiões terminais ricas em GC e AT. Os trabalhos de MOSCONE  
23 *et al.* (1993; 1996; SCALDAFERRO *et al.* 2013; MARTINS *et al.*, 2018) reportaram  
24 uma ampla variabilidade quanto a presença de bandas terminais (bandas mais

1 intensas ou tênues) e quanto a presença de bandas intersticiais e proximais. Esses  
2 dados corroboram com a diversidade de bandas CMA-DAPI observada em nosso  
3 trabalho com *C. annuum*, *C. chinense* e *C. baccatum*. O procedimento técnico  
4 empregado nesse trabalho (bandeamento C-CMA-DAPI) foi eficaz ao revelar bandas  
5 CMA+ em quase todos os centrômeros nas três espécies, e esses sinais foram  
6 colocalizados com os sinais de FISH com a sonda de CRM. A presença  
7 predominante de TEs em regiões hipermetiladas e com baixos níveis de  
8 recombinação, é sugerido por Bennetzen E Wang (2014), por serem regiões onde a  
9 inserção de retrotransposons pode ter um efeito mínimo. Isso explicaria a  
10 colocalização de elementos CRM associados a região centromérica.

## 11 **Agradecimentos**

12

## 13 **Referências**

14

- 15 1. Schmidt T, Heslop-Harrison JS. Genomes, genes and junk: The large-scale  
16 organization of plant chromosomes. *Trends in Plant Science* 1998;3:195–9.
- 17 2. Heslop-Harrison JSP, Schwarzacher T. Organisation of the plant genome in  
18 chromosomes. *Plant Journal* 2011;66:18–33.
- 19 3. Bennetzen JL, Wang H. The Contributions of Transposable Elements to the  
20 Structure, Function, and Evolution of Plant Genomes. *Annual Review of Plant*  
21 *Biology* 2014;65:505–30.
- 22 4. Jr HHK. Mobile Elements : Drivers of Genome Evolution. *Science* [Internet]  
23 2004;303:1626–32. Available from:  
24 <http://www.sciencemag.org/content/303/5664/1626.full>
- 25 5. Wicker T, Sabot F, Hua-van A, Bennetzen JL, Capy P, Chalhoub B, et al. Our

- 1 knowledge of the structure and com- position of genomes has progressed in  
2 pace with their sequencing. As was expected early on. Nature Reviews  
3 [Internet] 2007;8:973–82. Available from:  
4 <https://www.nature.com/articles/nrg2165.pdf>
- 5 6. Lisch D. How important are transposons for plant evolution? Nature Reviews  
6 Genetics [Internet] 2013;14:49–61. Available from:  
7 <http://dx.doi.org/10.1038/nrg3374>
- 8 7. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their  
9 reverse transcriptase sequences. The EMBO Journal 2018;9:3353–62.
- 10 8. Bennetzen JL. Transposable element contributions to plant gene and genome  
11 evolution. Plant Molecular Biology 2000;42:251–69.
- 12 9. Munoz-Lopez M, Garcia-Perez J. DNA Transposons: Nature and Applications  
13 in Genomics. Current Genomics 2010;11:115–28.
- 14 10. Levin HL, Moran J V. Dynamic interactions between transposable elements  
15 and their hosts. Nature Reviews Genetics [Internet] 2011;12:615–27. Available  
16 from: <http://dx.doi.org/10.1038/nrg3030>
- 17 11. Garrido-Ramos MA. Satellite DNA in plants: More than just rubbish.  
18 Cytogenetic and Genome Research 2015;146:153–70.
- 19 12. Garrido-Ramos MA. Satellite DNA: An evolving topic. Genes 2017;8.
- 20 13. Kubis S, Schmidt T, Heslop-harrison JS, Kubis S, Schmidtj T, Pat JS. Annals of  
21 Botany Repetitive DNA Elements as a Major Component of Plant Genomes.  
22 2008;82:45–55.
- 23 14. Ellegren H. Microsatellites: Simple sequences with complex evolution. Nature  
24 Reviews Genetics 2004;5:435–45.
- 25 15. Ling-Hu C, Wu H, Luo X, Cai H, Wu Z, Li S, et al. Whole-genome sequencing

- 1 of cultivated and wild peppers provides insights into *Capsicum* domestication  
2 and specialization. *Proceedings of the National Academy of Sciences*  
3 2014;111:5135–40.
- 4 16. Aza-González C, Núñez-Paleniús HG, Ochoa-Alejo N. Molecular biology of  
5 capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). *Plant Cell Reports*  
6 2011;30:695–706.
- 7 17. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, et al. Genome sequence of  
8 the hot pepper provides insights into the evolution of pungency in *Capsicum*  
9 species. *Nature Genetics* [Internet] 2014;46:270–8. Available from:  
10 <http://dx.doi.org/10.1038/ng.2877>
- 11 18. Moscone EA, Baranyi M, Ebert I, Greilhuber J, Ehrendorfer F, Hunziker AT.  
12 Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow cytometry  
13 and feulgen densitometry. *Annals of Botany* 2003;92:21–9.
- 14 19. Moscone EA, Scaldaferrò MA, Grabielle M, Jarret R, Ducasse DA, Gloria E, et  
15 al. The Evolution of Chili Peppers (. 2007;137–70.
- 16 20. BENNETT M. Nuclear DNA Amounts in Angiosperms—583 New Estimates.  
17 *Annals of Botany* [Internet] 1997;80:169–96. Available from:  
18 <https://academic.oup.com/aob/article-lookup/doi/10.1006/anbo.1997.0415>
- 19 21. Carrizo García C, Sterpetti M, Volpi P, Ummarino M, Saccardo F. Wild  
20 *Capsicums*: identification and in situ analysis of Brazilian species.  
21 *Breakthroughs in the genetics and breeding of Capsicum and eggplant*. Edited  
22 by S. Lanteri, and GL Rotino 2013;205–13.
- 23 22. Carrizo García C, Barfuss MHJ, Sehr EM, Barboza GE, Samuel R, Moscone  
24 EA, et al. Phylogenetic relationships, diversification and expansion of chili  
25 peppers (*Capsicum*, Solanaceae). *Annals of Botany* 2016;118:35–51.

- 1 23. Badr A, SF K, Al A-A, MM A-E-E. Chromosomal criteria and taxonomic  
2 relationships in the Solanaceae. *Cytologia* 1997;62:103–13.
- 3 24. Padilha HKM, Vasconcelos CS, Villela J, Valgas RA, Barbieri RL. Agronomic  
4 evaluation and morphological characterization of chili peppers ( *Capsicum*  
5 *annuum* , Solanaceae ) from Brazil. *Australian Journal of Basic and Applied*  
6 *Sciences* 2016;10:63–70.
- 7 25. Moscone EA, Lambrou M, Hunziker AT, Ehrendorfer F. Giemsa C-banded  
8 karyotypes in *Capsicum* (Solanaceae). *Plant Systematics and Evolution*  
9 [Internet] 1993;186:213–29. Available from:  
10 <http://link.springer.com/10.1007/BF00940799>
- 11 26. Scaldaferrero MA, Grabielle M, Moscone EA. Heterochromatin type, amount and  
12 distribution in wild species of chili peppers (*Capsicum*, Solanaceae). *Genetic*  
13 *Resources and Crop Evolution* 2013;60:693–709.
- 14 27. Scaldaferrero MA, da Cruz MVR, Cecchini NM, Moscone EA. FISH and AgNor  
15 mapping of the 45S and 5S rRNA genes in wild and cultivated species of *C*  
16 *apsicum* (Solanaceae) . *Genome* 2015;59:95–113.
- 17 28. Moscone EA, Lambrou M, Ehrendorfer F. Fluorescent chromosome banding in  
18 the cultivated species of *Capsicum* (Solanaceae). *Plant Systematics and*  
19 *Evolution* 1996;202:37–63.
- 20 29. Martins LDV, Peron AP, Lopes ÂC de A, Gomes RLF, Carvalho R de, Feitoza  
21 LDL. Heterochromatin distribution and histone modification patterns of H4K5  
22 acetylation and H3S10 phosphorylation in *Capsicum* L. *Crop Breeding and*  
23 *Applied Biotechnology* [Internet] 2018;18:161–8. Available from:  
24 [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1984-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-70332018000200161&lng=en&tlng=en)  
25 [70332018000200161&lng=en&tlng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-70332018000200161&lng=en&tlng=en)

- 1 30. Gorinšek B, Gubenšek F, Kordiš D. Evolutionary Genomics of Chromoviruses  
2 in Eukaryotes. *Molecular Biology and Evolution* [Internet] 2004;21:781–98.  
3 Available from: [https://academic.oup.com/mbe/article-](https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msh057)  
4 [lookup/doi/10.1093/molbev/msh057](https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msh057)
- 5 31. Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, et al.  
6 The Gypsy Database (GyDB) of Mobile Genetic Elements: Release 2.0.  
7 *Nucleic Acids Research* 2011;39:70–4.
- 8 32. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, et al. IBS: An illustrator for the  
9 presentation and visualization of biological sequences. *Bioinformatics*  
10 2015;31:3359–61.
- 11 33. Sonnhammer ELL, Durbin R. A dot-matrix program with dynamic threshold  
12 control suited for genomic DNA and protein sequence analysis. *Gene*  
13 1995;167:GC1-GC10.
- 14 34. Romano E, Brasileiro ACM. Extração de DNA de plantas. *Biotecnologia*  
15 1999;2:40–3.
- 16 35. Schweizer D. Reverse fluorescent chromosome banding with chromomycin  
17 and DAPI. *Chromosoma* 1976;58:307–24.
- 18 36. Paz RC, Kozaczek ME, Rosli HG, Andino NP, Sanchez-Puerta MV. Diversity,  
19 distribution and dynamics of full-length Copia and Gypsy LTR retroelements in  
20 *Solanum lycopersicum*. *Genetica* [Internet] 2017;145:417–30. Available from:  
21 <http://link.springer.com/10.1007/s10709-017-9977-7>
- 22 37. Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, et al.  
23 The sunflower (*Helianthus annuus* L.) genome reflects a recent history of  
24 biased accumulation of transposable elements. *Plant Journal* 2012;72:142–53.
- 25 38. Di Filippo M, Traini A, D'Agostino N, Frusciante L, Chiusano ML. Euchromatic

- 1 and heterochromatic compositional properties emerging from the analysis of  
2 *Solanum lycopersicum* BAC sequences. *Gene* [Internet] 2012;499:176–81.  
3 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378111912002466>
- 4 39. Bolger A, Scossa F, Bolger ME, Lanz C, Maumus F, Tohge T, et al. The  
5 genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature*  
6 *Genetics* [Internet] 2014;46:1034–8. Available from:  
7 <http://dx.doi.org/10.1038/ng.3046>
- 8 40. Vernhettes S, Grandbastien MA, Casacuberta JM. The evolutionary analysis of  
9 the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its  
10 regulatory sequences. *Molecular Biology and Evolution* [Internet] 1998;15:827–  
11 36. Available from: [https://academic.oup.com/mbe/article-](https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a025988)  
12 [lookup/doi/10.1093/oxfordjournals.molbev.a025988](https://academic.oup.com/mbe/article-lookup/doi/10.1093/oxfordjournals.molbev.a025988)
- 13 41. MELAYAH D, LIM KY, BONNIVARD E, CHALHOUB B, DORLHAC DE BORNE  
14 F, MHIRI C, et al. Distribution of the Tnt1 retrotransposon family in the  
15 amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives.  
16 *Biological Journal of the Linnean Society* [Internet] 2004;82:639–49. Available  
17 from: [https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/j.1095-](https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/j.1095-8312.2004.00348.x)  
18 [8312.2004.00348.x](https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/j.1095-8312.2004.00348.x)
- 19 42. Bergman CM, Quesneville H. Discovering and detecting transposable elements  
20 in genome sequences. *Briefings in Bioinformatics* [Internet] 2007;8:382–92.  
21 Available from: [https://academic.oup.com/bib/article-](https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbm048)  
22 [lookup/doi/10.1093/bib/bbm048](https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbm048)
- 23 43. Park M, Jo S, Kwon J-K, Park J, Ahn JH, Kim S, et al. Comparative analysis of  
24 pepper and tomato reveals euchromatin expansion of pepper genome caused  
25 by differential accumulation of Ty3/Gypsy-like elements. *BMC Genomics*

- 1 [Internet] 2011;12:85. Available from:  
2 <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-85>
- 3 44. Neumann P, Navrátilová A, Koblížková A, Kejnovský E, Hřibová E, Hobza R, et  
4 al. Plant centromeric retrotransposons: a structural and cytogenetic  
5 perspective. *Mobile DNA* [Internet] 2011;2:4. Available from:  
6 <http://mobilednajournal.biomedcentral.com/articles/10.1186/1759-8753-2-4>
- 7 45. Nagaki K, Shibata F, Suzuki G, Kanatani A, Ozaki S, Hironaka A, et al.  
8 Coexistence of NtCENH3 and two retrotransposons in tobacco centromeres.  
9 *Chromosome Research* [Internet] 2011;19:591–605. Available from:  
10 <http://link.springer.com/10.1007/s10577-011-9219-2>  
11  
12

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

## 6. CONSIDERAÇÕES FINAIS

- 1- A variação dos elementos de repetitivos pode impactar na composição dos genomas de *Capsicum*, especialmente porque os retrotransposons possuem um caráter mais ativo e ocuparam mais de 80% dos genomas, com predominância da superfamília *Gypsy*.
- 2- As espécies pertencentes ao mesmo clado, ou seja, mais próximas filogeneticamente apresentaram proporções similares no acúmulo de retrotransposons. *Capsicum annuum* e *C. chinense* tiveram maior acúmulo de TEs da família Del, enquanto *C. baccatum* apresentou maior acúmulo de Athila.
- 3- Entre as espécies analisadas, *C. baccatum* apresentou maior acúmulo de sequências de DNAr 35S e menor acúmulo de retrotransposons que as demais espécies. Quando analisadas as sequências repetitivas de maneira conjunta, a baixo frequência de elementos de Classe I parece ser compensado pelos maiores de DNAr 35S em *C. baccatum*.
- 4- O acúmulo diferencial de retrotransposons sugere que esses foram mais ativos no processo de diferenciação genômica entre os clado, do que os ERVs e os elementos de Classe II, suportando a ideia de que elementos repetitivos possuem atividade de maneira independente nos genomas vegetais, mesmo entre espécies filogeneticamente próximas.
- 5- O acúmulo de retrotransposons da família Del, em ambos os cladros, não demonstrou preferência por regiões de heterocromatina, uma vez que possuíram perfil disperso ao longo dos cromossomos.
- 6- Regiões de heterocromatina centromérica mostraram ser colocalizados com as regiões ricas em fragmentos da transcriptase reversa da família CRM, tal fato pode ser explicado por essa região ser hipermetilada e inserções de retrotransposons poderia ter um efeito mínimo.