



LUDMILA CRISTINA OLIVEIRA

**DNA REPETITIVO CENTROMÉRICO E
DISPERSO EM ESPÉCIES DE *Solanum***

**LAVRAS – MG
2015**

LUDMILA CRISTINA OLIVEIRA

**DNA REPETITIVO CENTROMÉRICO E DISPERSO EM ESPÉCIES DE
*Solanum***

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Doutor.

Orientadora
Dra. Giovana Augusta Torres

**LAVRAS – MG
2015**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Oliveira, Ludmila Cristina.

DNA repetitivo centromérico e disperso em espécies de
Solanum/ Ludmila Cristina Oliveira. – Lavras: UFLA, 2015.
113 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2015.
Orientador (a): Giovana Augusta Torres.
Bibliografia.

1. DNA repetitivo. 2. Centrômero. 3. DNA satélite. 4.
Retrotransposons centroméricos. 5. *Solanum*. I. Universidade
Federal de Lavras. II. Título.

LUDMILA CRISTINA OLIVEIRA

**DNA REPETITIVO CENTROMÉRICO E DISPERSO EM ESPÉCIES DE
*Solanum***

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética e Melhoramento de Plantas, para a obtenção do título de Doutor.

APROVADA em 26 de fevereiro de 2015.

Dra. Andrea Pedrosa-Harand	UFPE
Dr. Mateus Mondin	ESALQ
Dra. Lisete Chamma Davide	UFLA
Dra. Vânia Helena Techio	UFLA

Dra. Giovana Augusta Torres
Orientadora

**LAVRAS – MG
2015**

Aos meus formadores, principalmente àqueles que instigaram em mim o espírito científico e me permitiram fazer ciência.

DEDICO

À minha família e aos meus amigos, pelo apoio incondicional.

OFEREÇO

AGRADECIMENTOS

A Deus, força independente de tudo, coautor do que sou.

À Dra. Giovana, a quem nunca saberei agradecer apropriadamente, tamanha sua contribuição para este trabalho, para a minha formação como pesquisadora e acima de tudo como pessoa. Seu exemplo profissional é minha maior inspiração.

Ao Dr. Jiming, por ter me acolhido em seu laboratório na University of Wisconsin, por todo seu esforço para que eu tivesse êxito em minha pesquisa e pelos seus conselhos peculiares e valiosos.

Às professoras Dra. Lisete e Dra. Vânia, por todo o auxílio e atenção no decorrer do curso, e aos demais professores componentes da banca examinadora Dra. Andrea e Dr. Mateus, por todas as contribuições.

A todos os amigos do laboratório de Citogenética da UFLA, especialmente à Thaís, que dividiu comigo momentos bons e ruins, no Brasil e nos Estados Unidos, e foi muito importante para que eu conseguisse chegar até aqui. À Cristina e à Gisele, pelo convívio tão próximo e por terem participado com tanto cuidado principalmente da reta final dessa conquista. Agradeço também à Fernanda, Dayane e ao Guilherme pela amizade e pelo acolhimento, especialmente quando retornei do exterior.

A todos os amigos do laboratório do Dr. Jiming, pelos ensinamentos, pela paciência e pelo convívio tão harmonioso durante o ano que passei por lá.

Agradeço de forma especial ao Kai, à Pam e à Haiqin pela boa vontade em me ensinar tudo, mesmo que fosse a mais básica técnica molecular.

Aos meus pais e à minha irmã, que confiaram nas minhas escolhas e sempre me apoiaram incondicionalmente. Sem a nossa família como base, eu não teria sequer condições de sonhar este sonho, muito menos de realizá-lo.

À minha família, de forma especial ao Flávio e à tia Doralice, por estarem sempre tão próximos e presentes.

Aos companheiros de república, especialmente ao Wesley e à Fabiana, por terem se tornado minha família em Lavras durante tantos anos.

A todos os amigos de Madison, que contribuíram muito para que o ano que eu passei por lá fosse um dos mais felizes que já vivi.

Aos demais amigos (aos da família Lavras, aos da minha turma de graduação, do GEN, à Jeanne, aos Caios - Salgado e Quintiliano - ao Saulo, ao Carlos, ao Pascoal, à Mariana e à Quêlen) por compreenderem minhas ausências e me apoiarem.

Ao professor Ivan, por todas as oportunidades que me deu para o aprendizado da língua inglesa.

Ao Sérgio, pelo auxílio nos experimentos.

À UFLA, pela oportunidade de realização do doutorado, e ao seu corpo docente, pela contribuição para o meu crescimento profissional.

À University of Wisconsin, pela oportunidade de realização do doutorado sanduíche e de concretização de um sonho.

Aos funcionários do Departamento de Biologia, pelo pronto atendimento e colaboração durante a realização do curso.

À Capes e ao CNPq, pela concessão da bolsa de estudos.

**“Com grandes poderes
vêm grandes responsabilidades”**
(Stan Lee / Tio Ben)

RESUMO

O DNA repetitivo pode compor até 90% dos genomas de muitas espécies animais e vegetais e por vezes tem relação com a funcionalidade de estruturas cromossômicas. Existem dois grupos de DNA repetitivo, em tandem (DNA satélite), normalmente encontrado em regiões cromossômicas específicas como nos telômeros e centrômeros, e disperso, para o qual na maioria dos casos não há localização preferencial, embora ele seja frequentemente encontrado na região centromérica. A completa compreensão da complexidade e do funcionamento de estruturas como os centrômeros, e dos genomas de forma geral, sem informações acerca de sua fração repetitiva é improvável, uma vez que ela é seu mais abundante componente. O objetivo deste trabalho foi identificar e caracterizar o padrão de organização do DNA repetitivo presente nos genomas de espécies cultivadas e selvagens de *Solanum*, com ênfase no DNA centromérico, e fazer um estudo evolutivo de todas as sequências identificadas em espécies portadoras de diferentes genomas descritos para o gênero (A, B, E, P e T). Análises dos dados de sequenciamento do genoma da batata permitiram identificar três retrotransposons LTR da superfamília *Ty-3 gypsy*, localizados por FISH nos cromossomos da batata (*S. tuberosum*). St73 apresentou localização centromérica e é até então o único componente genético presente em todos os centrômeros da batata. St6 não mostrou localização preferencial, estando ausente apenas nas regiões terminais dos cromossomos, enquanto que St1 localizou-se na região centromérica. St1 e St6 estão presentes em todos os genomas estudados e St73 nos genomas A, P e E. St73 e St1 podem ser importantes para o funcionamento dos centrômeros do gênero. A aplicação da ChIP utilizando anticorpos contra CenH3 específica para o centrômero de batata permitiu isolar e analisar o DNA centromérico de *S. chomatophilum*, uma espécie selvagem portadora do genoma P. Pelo menos três, dos doze centrômeros do complemento, são compostos por DNA repetitivo em tandem e os demais provavelmente por DNA de cópia única ou por retrotransposons. Foi detectada uma grande variabilidade das sequências repetitivas centroméricas, que vai além da que já era conhecida (entre cromossomos), podendo variar entre centrômeros homólogos, entre acessos de uma mesma espécie e entre plantas de um mesmo acesso. As sequências centroméricas isoladas do genoma P apresentam alguma conservação nas espécies com genoma A, mas estão pouco presentes nos demais genomas. Os dados apontam para uma organização complexa e diferenciada dos centrômeros em *Solanum* e para a rápida evolução do DNA centromérico.

Palavras-chave: DNA repetitivo. Centrômero. DNA satélite. Retrotransposons Centroméricos. *Solanum*.

ABSTRACT

The repetitive DNA can compose up to 90% of many animal and plant genomes and sometimes is related to the functionality of chromosome structures. There are two groups of repetitive DNA, tandem DNA (satellite DNA), usually found at specific chromosomal regions as in the telomeres and centromeres, and dispersed DNA, which in most cases show no preferential location, though it is often found at the centromeric region. The broad comprehension of complexity and function of some chromosome structures, as centromeres, and of the genomes in general, is unlikely without information about its most abundant component, the repetitive fraction. This research aimed to identify and characterize the organization pattern of repetitive DNA in the genome of cultivated and wild *Solanum* species, with emphasis on centromeric DNA. Evolutionary aspects of all the sequences were also analyzed based on their presence and distribution in species representing different *Solanum* genomes (A, B, E, P, and T). Analysis of potato genome sequencing data identified three *Ty3* gypsy LTR retrotransposons, located by FISH in potato chromosomes (*S. tuberosum*). St73 mapped to the centromeres and it is the only genetic component present in all potato centromeres so far. St6 showed no preferential location, being absent only in chromosomes terminal regions, while St1 is located in the centromeric region. St1 and St6 are present in all studied genomes, while St73 is only in the A, P and E genomes. St73 and St1 may be important for the functioning of the *Solanum* centromeres. The ChIP technique using CenH3 antibodies, specific for the potato centromeric histone, allowed to isolating and analyzing the centromeric DNA of *S. chomatophilum*, a wild potato species carrying the P genome. At least three of the twelve centromeres are composed of tandem repetitive DNA and probably the others have single copy DNA or retrotransposons. We have detected a large variability of centromeric repeat sequences that goes beyond the variation among chromosomes already described for potato. We observed differences between homologous centromeres, accessions of the same species and plants of the same accession. The isolated centromeric sequences of the P genome are partially conserved in species with the A genome, but they are not found in other genomes often. The data indicate a complex and differentiated organization of *Solanum* centromeres and a rapid evolution of centromeric DNA.

Keywords: Repetitive DNA. Centromere. Satellite DNA. Centromeric retrotransposon. *Solanum*.

LISTA DE ILUSTRAÇÕES

PRIMEIRA PARTE

Figure 1	Sistema de classificação proposto para elementos.....	21
----------	---	----

SEGUNDA PARTE – ARTIGO 1

Figura 1	PCR utilizando primers desenhados para as repetições centroméricas de <i>S. chomatophilum</i> : Sc83, Sc92, Sc108 e Sc111....	71
Figura 2	Mapeamento por FISH das repetições centroméricas em <i>S. chomatophilum</i> , 2n=2x=24, acesso A. (A, C, E e G) Cromossomos de <i>S. chomatophilum</i> corados com DAPI (B) Sinais de FISH gerados pela sonda Sc83 (D) Sinais de FISH gerados pela sonda Sc92 (F) Sinais de FISH gerados pela sonda Sc108 (H) Sinais de FISH gerados pela sonda Sc111. Barras = 5 µm.....	72
Figura 3	Mapeamento por FISH das repetições centroméricas em <i>S. chomatophilum</i>	Erro! Indicador não definido.
Figura 4	Análise das repetições satélite centroméricas em <i>S. chomatophilum</i> por meio de fiber-FISH.	Erro! Indicador não definido.
Figura 5	Esquema ilustrando a distribuição das repetições centroméricas Sc111, Sc92/108, Sc92/108/111 e Sc83 nos cromossomos dos dois acessos de <i>Solanum</i>	Erro! Indicador não definido.
Figura 6	Variação no número de sinais centroméricos Sc83 em plantas do acesso B.....	Erro! Indicador não definido.
Figura 7	Mapeamento por FISH das repetições centroméricas isoladas de <i>S. chomatophilum</i> em diferentes espécies de <i>Solanum</i> .	Erro! Indicador não definido.
Tabela 1	Resumo das análises de FISH das repetições centroméricas isoladas de <i>S. chomatophilum</i> em espécies de <i>Solanum</i> . ..	Erro! Indicador não definido.

SEGUNDA PARTE – ARTIGO 2

- Figura 1 Distribuição de St73 por meio de FISH com marcação vermelha nos cromossomos de espécies diploides ($2n=2x=24$) de *Solanum*.**Erro! Indicador não definido.**
- Figura 2 Distribuição de St1 por meio de FISH com marcação verde nos cromossomos de espécies diploides ($2n=2x=24$) de *Solanum*.**Erro! Indicador não definido.**
- Figura 3 Alguns cromossomos de DM1-3 selecionados como exemplo para melhor visualização dos grandes sinais provenientes da sonda St1 na região pericentromérica e presença reduzida ou falta desses sinais na região do centrômero. Barra: 2.5 μm .**Erro! Indicador não definido.**
- Figura 4 Distribuição de St6 por meio de FISH com marcação verde nos cromossomos de espécies diploides ($2n=2x=24$) de *Solanum*.**Erro! Indicador não definido.**

LISTA DE ABREVIATURAS

cDNA	DNA complementar
Cen	Centrômero
DIRS	Do inglês <i>Dictyostelium Intermediate Repeat Sequence</i>
dNTP	Desoxirribonucleotídeo trifosfato
EDTA	Ácido etileno diamônio tetracético
ERVs	Do inglês <i>Endogenous Retroviruses</i>
FISH:	Hibridização fluorescente <i>in situ</i> (do inglês <i>Fluorescent in Situ Hybridization</i>)
Kb	1000 bases
LINEs	Do inglês <i>Long Interspersed Nuclear Elements</i>
LTRs	Longas terminações repetidas (do inglês <i>Long terminal repeats</i>)
Mb:	Milhões de bases
RNA_m	RNA mensageiro
ORF	Quadro aberto de leitura (do inglês <i>Open Reading Frame</i>)
pb	Pares de bases
PCR	Reação em cadeia de polimerase (do inglês <i>Polymerase Chain Reaction</i>)
PLE	Do inglês <i>Penelope-Like Elements</i>
rDNA	DNA ribossômico
rpm	Rotação por minuto
SINEs	Do inglês <i>Short Interspersed Nuclear Elements</i>
SMRT	Do inglês <i>Single-Molecule Real Time Sequencing</i>
TIR	Repetições terminais invertidas (do inglês <i>Terminal Inverted Repeats</i>)
TSD	Duplicação do sítio alvo (do inglês <i>Target Site Duplication</i>)

SUMÁRIO

	PRIMEIRA PARTE	
1	INTRODUÇÃO GERAL	15
2	REFERENCIAL TEÓRICO	18
2.1	Sequências de DNA repetitivo	18
2.1.1	Sequências repetitivas dispersas – Elementos transponíveis	18
2.1.1.1	Elementos transponíveis de classe I – RNA como intermediário de transposição	20
2.1.1.2	Elementos transponíveis de classe II – DNA como intermediário de transposição	26
2.1.1.3	O impacto dos elementos transponíveis na estrutura e funcionamento dos genomas	28
2.1.2	Sequências repetitivas em tandem – DNA satélite	30
2.2	DNA centromérico e sua organização em <i>Solanum</i>	33
2.3	Organização do genoma em Solanaceae	37
	REFERÊNCIAS	46
	SEGUNDA PARTE - ARTIGOS	
	ARTIGO 1	
	Composição e estrutura do DNA centromérico em <i>S. chomatophilum</i> e sua evolução no gênero <i>Solanum</i>	63
1	INTRODUÇÃO	66
2	MATERIAL E MÉTODOS	68
3	RESULTADOS	71
4	DISCUSSÃO	81
	REFERÊNCIAS	86
	DADOS SUPLEMENTARES	89
	ARTIGO 2	
	Retrotransposons centroméricos e pericentroméricos em <i>Solanum L.</i>	91
1	INTRODUÇÃO	94
2	MATERIAL E MÉTODOS	96
3	RESULTADOS	99
4	DISCUSSÃO	104
	REFERÊNCIAS	108

PRIMEIRA PARTE

1 INTRODUÇÃO GERAL

O gênero *Solanum* L. é o maior e mais complexo da família Solanaceae, com mais de 1.500 espécies (AGRA, 1999), um dos motivos pelos quais ele é tido como um dos mais interessantes gêneros do ponto de vista evolutivo (KNAPP et al., 2004). Uma de suas espécies, a batata (*Solanum tuberosum* L.), está entre os alimentos mais importantes para a humanidade, sendo cultivada em mais de 125 países e consumida por mais de um bilhão de pessoas. A batata ocupa a quarta posição entre as principais culturas produzidas mundialmente, superada apenas pelo trigo, milho e arroz (SOLANACEAE GENOMICS RESOURCE, 2012).

Diante de sua grande importância econômica, grande esforço é dispendido para o melhoramento da batata em vários países. Uma ferramenta que está disponível nos últimos anos para os programas é a incorporação de informações genômicas, que têm potencial de conferir mais flexibilidade a estes programas e, de forma gradativa e cumulativa, melhorar a eficiência de seleção (LEDUR et al., 2011). No caso da batata, um avanço significativo nesse sentido foi o sequenciamento do genoma de um genótipo monoploide ($2n=2x=24$) que estimou a existência de cerca de 39.000 genes e de pelo menos 62% de DNA repetitivo (POTATO GENOME SEQUENCING CONSORTIUM - PGSC, 2011).

Sequências repetitivas são fragmentos de DNA que apresentam múltiplas cópias no genoma eucariótico, podendo constituir até 90% do genoma de algumas espécies vegetais (MEHROTRA; GOYAL, 2014). Existem dois grupos de sequências repetitivas. No primeiro grupo estão as sequências intercaladas, que normalmente se encontram dispersas na maioria dos cromossomos. Grande parte delas são elementos transponíveis, principalmente aqueles pertencentes à classe dos retroelementos (JURKA et al., 2007). Eles

podem apresentar localização aleatória no genoma, mas é bastante comum que tenham localização preferencial, como nas regiões centroméricas e pericentroméricas, podendo desempenhar importantes funções em seus sítios de inserção (CHENG et al., 2002; COPENHAVER et al., 1999; PIMPINELLI et al. 1995; SCHUELER et al., 2001). Os retroelementos são uma importante fonte de diversidade, atuando continuamente na geração de mutações como consequência de sua amplificação e mobilidade. Existem também evidências de que eles contribuam para regulação gênica e que participem de respostas a estresse celular (DERAGON; CAPY, 2000).

O segundo grupo de sequências repetitivas é composto pelas sequências em tandem, que possuem unidades de repetição (monômeros) imediatamente adjacentes umas às outras, construindo grandes arranjos que podem ter até dezenas de megabases de comprimento (JURKA et al., 2007). As repetições em tandem são geralmente encontradas em regiões específicas e que desempenham um papel funcional nos cromossomos, como nos telômeros e subtelômeros, nos genes para DNA ribossomal e nos centrômeros (SHARMA; RAINA, 2005).

O centrômero é o sítio cromossômico responsável por manter unidas as cromátides irmãs e é onde o cinetócoro se organiza para que as fibras do fuso a ele se conectem para promover a segregação dos cromossomos durante a mitose e a meiose (JIANG et al., 2012). Repetições satélites têm sido descritas como componentes majoritários dos centrômeros de muitas espécies de plantas, como *Arabidopsis thaliana* (L.) Heynh. (MARTINEZ-ZAPATER et al., 1996), arroz (CHENG et al., 2002; DONG et al., 1998; NONOMURA; KURATA, 1999), milho (KAMM et al., 1994), sorgo (JIANG et al., 1996; MILLER et al., 1998), trigo (KISHII et al., 2001), milho (ALFENITO; BIRCHLER, 1993; ANANIEV et al., 1998; NAGAKI et al., 2003a) e batata (GONG et al., 2012; ZHANG et al., 2014).

Em *S. tuberosum* foi descrita uma estrutura bastante incomum dos

centrômeros, que podem ser constituídos tanto por DNA repetitivo quanto por DNA de cópia única, com diferenças marcantes entre os centrômeros de cada um dos cromossomos do complemento (GONG et al., 2012). Resultados semelhantes foram encontrados por Zhang et al. (2014) para os centrômeros de *Solanum verrucosum* Schltdl., uma espécie de batata selvagem proximamente relacionada à batata cultivada. Em ambos os casos os autores encontraram grande similaridade entre a maioria das repetições satélite centroméricas e retrotransposons do tipo Ty3-gypsy e acreditam que as repetições foram amplificadas a partir desses retrotransposons.

O conhecimento da organização genômica, localização cromossômica e origem evolutiva das sequências de DNA repetitivo, disperso e em tandem, é necessário para elucidação do potencial funcional destas sequências e para melhor compreensão de seu envolvimento no processo evolutivo dos genomas eucarióticos (PLOHL, 2010). Considerando a capacidade de atuação do DNA repetitivo na determinação da estrutura cromossômica, a plasticidade conferida aos genomas pelos retroelementos e a forte relação do DNA repetitivo com a estrutura e funcionalidade do centrômero, o presente trabalho visa elucidar a distribuição e organização dos retroelementos mais abundantes no genoma de *Solanum tuberosum*, caracterizar o DNA centromérico na batata selvagem *Solanum chomatophilum* Bitter, bem como verificar sua organização e distribuição e por fim investigar o padrão evolutivo de ambos (retrotransposons e sequências centroméricas) em outras espécies cultivadas e selvagens do gênero *Solanum*.

2 REFERENCIAL TEÓRICO

2.1 Sequências de DNA repetitivo

O termo “sequências de DNA repetitivo” refere-se a fragmentos de DNA que estão presentes em múltiplas cópias nos genomas. Considera-se que esses elementos sejam os responsáveis pela grande variação em tamanho e sequência dos genomas, influenciando as distâncias evolucionárias entre as espécies (MEHROTRA; GOYAL, 2014).

A compreensão dos mecanismos evolutivos na formação de genomas eucarióticos complexos é impossível sem a completa investigação das sequências de DNA repetitivo (SHAPIRO; VON STERNBERG, 2005). Essa necessidade fica ainda mais acentuada em plantas superiores, uma vez que a porção dessas sequências chega a compreender até 90% do DNA nuclear (MEHROTRA; GOYAL, 2014).

Em contraste com as famílias multigênicas, que são geralmente agrupadas com base no papel biológico que desempenham, as famílias de DNA repetitivo são definidas levando-se em consideração seus antepassados ativos e seus mecanismos de geração, sendo que ao longo do tempo os elementos individuais de famílias repetitivas podem adquirir diversas funções biológicas (JURKA et al., 2007).

Dessa forma, o DNA repetitivo é classificado primariamente em duas classes principais que, por sua vez, são consequências de seus modos de amplificação. São elas as sequências repetitivas dispersas e as sequências repetitivas em tandem (SLAMOVITS; ROSSI, 2002).

2.1.1 Sequências repetitivas dispersas – Elementos transponíveis

A caracterização genética dos elementos transponíveis iniciou-se na década de 40, com os trabalhos de citogenética em milho desenvolvidos por

Bárbara McClintock, nos quais foi demonstrada a existência de fatores capazes de se moverem e influenciarem a expressão de genes. Este conceito questionava o pensamento da comunidade científica da época, a qual considerava o genoma uma entidade estática e associava sua instabilidade a mutações deletérias ao organismo. Depois de sua descoberta original e caracterização em milho, os elementos transponíveis têm sido encontrados em todos os organismos, incluindo todas as espécies de plantas investigadas (BENNETZEN, 2000).

Com o advento do sequenciamento de DNA, tornou-se evidente que, longe de ser um componente raro de alguns genomas, como se pensava antes, os elementos transponíveis são os maiores componentes individuais do material genético da maioria dos eucariotos (FESCHOTTE; ZHANG; WESSLE, 2002), chegando a constituir 80% ou mais do DNA genômico total (WICKER et al., 2007).

Embora os principais grupos de elementos transponíveis sejam antigos e estejam presentes em todos os reinos, é observada uma diversidade extrema, existindo até dezenas de milhares de famílias diferentes (MORGANTE, 2005; SANMIGUEL; BENNETZEN, 1998). Dada a abundância e evidente diversidade dos elementos transponíveis, e mediante a dificuldade de nomeá-los, Wicker et al. (2007) propuseram um sistema unificado de nomenclatura, na intenção de integrar conceitos anteriormente adotados, fornecendo um consenso entre as várias classificações conflitantes e sistemas de nomeação existentes. Esta classificação divide os elementos transponíveis hierarquicamente em classe, subclasse, ordem, superfamília, família e subfamília. O nível mais alto (classe) separa-os de acordo com a presença ou ausência de um RNA intermediário no processo da transposição. As subclasses dividem elementos que copiam a eles próprios daqueles que deixam o local de

origem para se reintegrar em outro. A subdivisão em ordens marca as principais diferenças no mecanismo de inserção e, conseqüentemente, na organização geral e enzimologia dos elementos transponíveis. As superfamílias, dentro de uma ordem, compartilham da estratégia de replicação, mas distinguem-se por características generalizadas, tais como estrutura de proteínas ou domínios não codificadores e são diferenciadas também pela presença e pelo tamanho das TSDs (do inglês Target Site Duplication), a duplicação de uma sequência genômica curta no sítio de inserção do elemento transponível. As famílias são definidas pela conservação das sequências de DNA e as subfamílias com base em dados filogenéticos e podem, em casos específicos, distinguir elementos autônomos de não autônomos. Detalhes desta classificação e da estrutura dos elementos transponíveis podem ser observados na figura 1.

2.1.1.1 Elementos transponíveis de classe I – RNA como intermediário de transposição

Nestes elementos, conhecidos como retroelementos ou retrotransposons, é realizada, via transcriptase reversa, a síntese de DNA a partir de uma cópia de RNA transcrita do elemento transponível, e este DNA é, então, capaz de se inserir em um sítio alvo (WICKER et al., 2007). Cada ciclo completo de replicação produz, portanto, uma nova cópia do elemento, fazendo dos retrotransposons os principais contribuintes para a fração repetitiva dos grandes genomas (KUMAR; BENNETZEN, 1999).

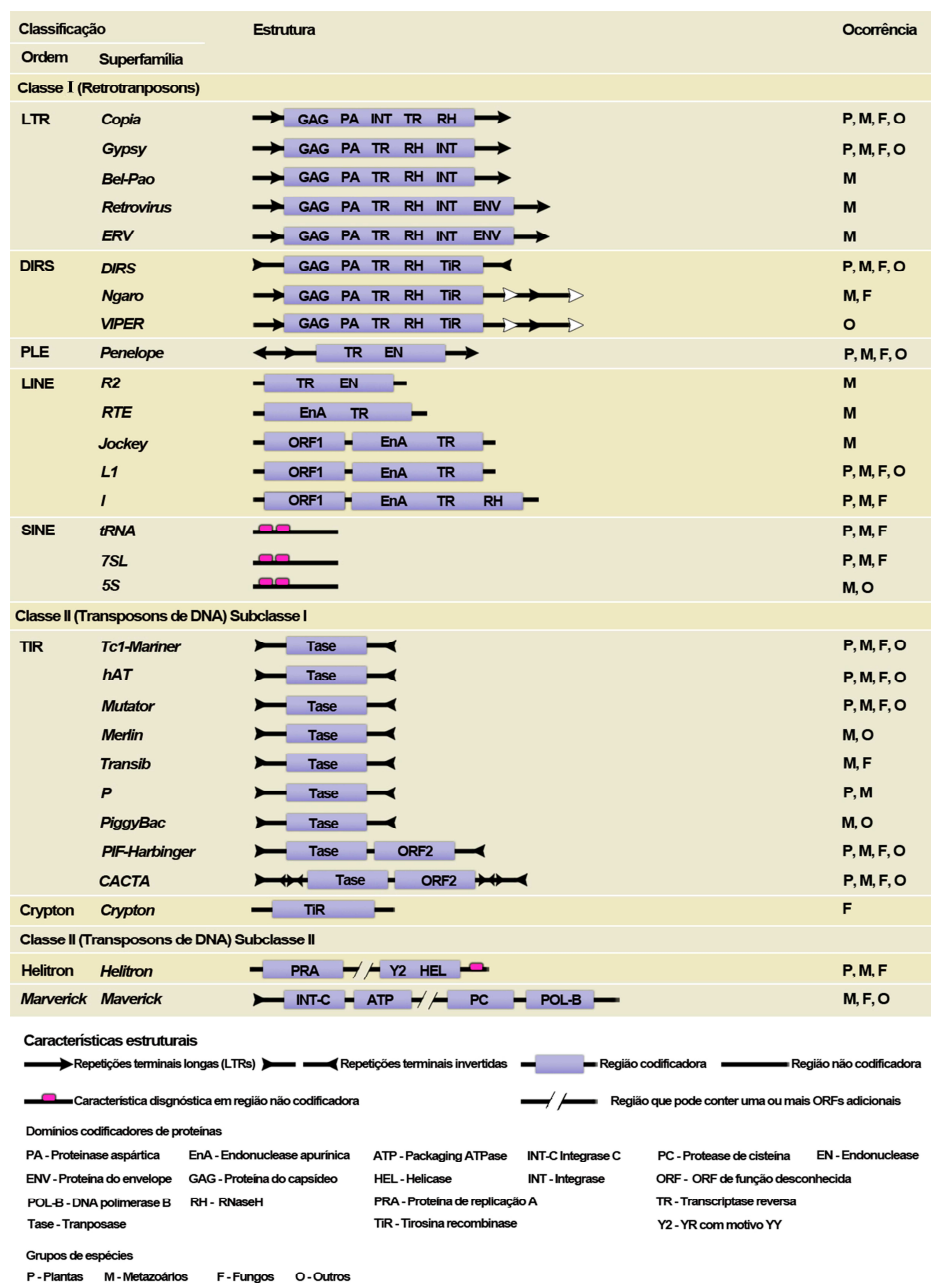


Figure 1 Sistema de classificação proposto para elementos
 Fonte: Adaptado de Wicker et al. (2007)

Dentre os elementos transponíveis existem as variantes autônomas, que codificam um conjunto completo de enzimas características de sua família e são autossuficientes em termos de transposição, e as não autônomas, que se transpõem por meio de empréstimos da maquinaria proteica codificada por seus parentes autônomos (JURKA et al., 2007).

Como no caso dos retrotransposons nenhum elemento é cortado ou transferido para outros sítios, não houve necessidade de subdividir a classe em subclasses, tendo sido, portanto, subdividida em cinco ordens: LTR (Long Terminal Repeat), DIRS (Dictyostelium Intermediate Repeat Sequence), PLE (*Penelope*-Like Elements), LINEs (Long Interspersed Nuclear Elements) e SINEs (Short Interspersed Nuclear Elements).

• LTRs - Long terminal repeats

Os retrotransposons LTR, assim chamados por possuírem longas repetições terminais (LTRs - Long Terminal Repeats), são menos abundantes em animais, mas constituem a ordem predominante em plantas e apresentam de poucas centenas de pares de bases até, excepcionalmente, 25 kb (NEUMANN; POZARKOVA; MACAS, 2003).

Após a transposição, os elementos LTRs produzem a duplicação de uma sequência genômica curta no sítio de inserção, a TSD (Target Site Duplication – Duplicação do sítio alvo). Estes retrotransposons contêm pelo menos dois genes: *gag* e *pol*. *Gag* codifica uma poliproteína estrutural para partículas semelhantes às de vírus e *pol* codifica proteínas catalíticas necessárias para a replicação. São elas: transcriptase reversa, RNase, integrase e proteinase aspártica (WICKER et al., 2007).

A ordem de retrotransposons LTR é subdividida em cinco superfamílias: *Copia*, *Gypsy*, *Bel-Pao*, *Retrovirus* e *ERVs* (do inglês: endogenous retroviruses). As superfamílias *Copia* e *Gypsy*, que se distinguem uma da outra pela ordem em que os genes aparecem, são as mais importantes dos genomas eucarióticos devido à sua

frequência e diversificação (FLAVELL et al., 1992; WICKER et al., 2007) e são amplamente distribuídas no reino vegetal (MARCO; MARÍN, 2005).

Embora em um genoma possam ser encontradas milhares de famílias de retrotransposons LTR, o que geralmente ocorre é que relativamente poucas famílias, muito mais abundantes que as demais, compõem a maior parte da fração repetitiva, como é o caso da família *Angela* em trigo (WICKER et al., 2001), *BARE* em cevada (VICIENT et al., 1999), *Opie* em milho (SANMIGUEL et al., 1998) e *Retrosor6* em sorgo (PETERSON et al., 2002).

• DIRS - Dictyostelium Intermediate Repeat Sequence

Uma das grandes diferenças entre os retroelementos DIRS e as outras ordens de retroelementos é a ausência de um gene para a integrase que é comum aos outros retroelementos, e a presença de outro gene que codifica para uma proteína diferente, porém também pertencente à família das integrases, chamada tirosina recombinase (GOODWIN; POULTER, 2011), a qual é necessária para o mecanismo de integração no genoma. Outra diferença notável é a de que os componentes desta ordem não geram sítios de duplicação no local de inserção, como é comum em alguns tipos de retrotransposons (WICKER et al., 2007) e apresentam uma estrutura incomum nas terminações, o que levou à classificação dos elementos DIRS como uma ordem separada na classe (GOODWIN; POULTER, 2004).

Retroelementos DIRS são delimitados por TIRs (do inglês Terminal Inverted Repeats – Repetições Terminais Invertidas) e possuem uma estrutura única composta por três ORFs: as duas primeiras codificando para *gag* e *pol*, como nos retrotransposons LTR e a terceira que codifica para tirosina recombinase (PIEDNOËL et al., 2011). Esta ordem é subdivida em 3 superfamílias: *DIRS*, *Ngaro* e *VIPER* (WICKER et al., 2007).

Inicialmente, estes retrotransposons foram relatados em apenas algumas espécies eucarióticas, tendo sido comumente encontrados em decápodes e

raramente em plantas e pouco se sabia a respeito de sua dinâmica e distribuição. Essa situação foi revertida quando Piednoël et al. (2011) analisaram, por meio de uma nova ferramenta computacional, 274 genomas que haviam sido sequenciados, o que revelou mais de 4.000 DIRS distribuídos entre 30 espécies diferentes e que podem ser agrupados em cerca de 300 famílias.

Uma vez que os retrotransposons DIRS são altamente diversificados e amplamente distribuídos em diferentes reinos eucarióticos, existe a hipótese de que eles sejam tão antigos quantos os retrotransposons LTR (JURKA et al., 2007) e, de acordo com Biémont e Vieira (2006) eles parecem também estar sobre forte pressão seletiva, sugerindo que eles têm alguma função que vem sendo conservada durante a evolução.

• PLE - *Penelope*-Like Elements

A ordem dos retrotransposons *Penelope*, que abriga apenas a superfamília *Penelope*, é composta por elementos que codificam uma única ORF composta por domínios de transcriptase reversa e endonuclease (JURKA et al., 2007). Eles compõem uma classe relativamente pouco estudada dentre os retroelementos eucarióticos e distinguem-se pela capacidade de reter íntrons em alguns representantes, pela semelhança de sua transcriptase reversa com a telomerase e também pela presença do domínio endonuclease semelhante à GIY-YIG, uma família de endonucleases não previamente encontrada em retroelementos e que induz a quebra da dupla fita de DNA (ARKHIPOVA, 2006). Retrotransposons *Penelope* são ainda caracterizados, assim como os DIRS, por longas repetições terminais incomuns ou repetições terminais invertidas não típicas dos retrotransposons sem LTR. (JURKA et al., 2007). Eles geram TSDs com 10-15 pares de bases (EVGEN'EV; ARKHIPOVA, 2005).

Uma análise filogenética abrangente dos retrotransposons *Penelope* foi realizada por Arkhipova (2006), com base em alinhamentos de sequências

estendidas e um conjunto ampliado de dados, incluindo muitos dos genomas sequenciados antes de 2005. A análise revelou um padrão de evolução semelhante ao de retrotransposons sem LTRs, formando clados profundamente ramificados e que datam da era pré-cambriana. Foram encontrados retrotransposons *Penelope* em representantes da maioria dos reinos eucarióticos, inclusive em plantas, embora eles sejam mais frequentes em genomas animais.

• **LINEs - Long Interspersed Nuclear Elements**

Os retroelementos do tipo LINE são autônomos, não possuem LTRs, podem chegar a vários kilobases de comprimento e são encontrados em todos os reinos eucarióticos (WICKER et al., 2007). Eles foram separados em cinco superfamílias (*R2*, *L1*, *RTE*, *I* e *Jockey*) e cada uma delas foi subdividida em muitas famílias.

Os LINEs possuem geralmente duas ORFs que codificam pelo menos uma transcriptase reversa e uma nuclease para sua transposição, no entanto, em algumas superfamílias, podem ser encontradas também ORFs para RNaseH, endonucleases apurínicas ou até mesmo ORFs com funções desconhecidas (WICKER et al., 2007).

De acordo com Kumar e Bennetzen (1999), os LINEs se inserem preferencialmente em regiões teloméricas ou próximas a outros elementos LINE. Não se sabe exatamente se esta preferência ocorre por uma real afinidade destes elementos pelos sítios em questão ou se devido à intensidade da pressão de seleção na remoção destes elementos nestes locais.

Embora existam algumas exceções, em plantas os LINEs parecem raros em comparação com os retrotransposons LTR (LEETON; SMYTH, 1993), sendo que os mais conhecidos são componentes das superfamílias *L1* e *RTE* (EICKBUSH; MALIK, 2002; ZUPUNSKI; GUBENSEK; KORDIS, 2001). No entanto, de acordo com Wicker et al. (2007), dada a falta de investigação sistemática em plantas, é provável que exista uma diversidade muito maior do que esta já conhecida.

- **SINEs - Short interspersed nuclear elements**

Os SINEs são elementos semelhantes aos LINEs, no entanto são mais curtos, mais simples e quase sempre dependentes da endonuclease e transcriptase reversa dos LINEs para sua retrotransposição (WEINER, 2002). A ordem que os compreende é dividida em 3 superfamílias: *tRNA*, *7SL* e *5S*, sendo que somente esta última não foi detectada em plantas.

Este é um grupo muito heterogêneo, no qual os elementos variam no comprimento de 90 a 300 pb e, tal como os LINEs, terminam em uma sequência simples repetitiva, geralmente uma cauda poli-A. Eles são considerados elementos defectivos que parasitam a maquinaria de transcrição dos LINEs (WESSLER, 2006).

2.1.1.2 Elementos transponíveis de classe II – DNA como intermediário de transposição

Assim como os retrotransposons, os elementos da classe II são evolutivamente antigos e encontrados em quase todos os eucariotos e também em procariotos (WICKER et al., 2007). Nestes últimos eles estão presentes em formas simples, chamadas sequências de inserção ou como parte de estruturas mais complexas (CHANDLER; MAHILLON, 2002).

De acordo com Feschotte, Zhang e Wessle (2002) este grupo se distingue pela transposição através de um intermediário de DNA, pela presença de TIRs e também por TSDs de comprimento conservado dentro das superfamílias. Cada superfamília inclui diversas famílias compostas por elementos autônomos e não autônomos (JURKA et al., 2007) e, usualmente, os membros de famílias não autônomas são derivados de família autônomas por deleções internas (FESCHOTTE; ZHANG; WESSLE, 2002). Na classificação de Wicker et al. (2007) estes elementos foram separados em duas subclasses.

➤ Subclasse I

A subclasse I compreende os elementos transponíveis clássicos do tipo “recorta-e-cola” da ordem TIR, caracterizados por suas repetições terminais invertidas de comprimento variável. Esta ordem é subdividida em nove superfamílias (*Tc1–Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger* e *CACTA*) que distinguem-se pelo tamanho das repetições terminais invertidas e das duplicações de sítio alvo. Sete delas apresentam apenas uma ORF para transposase e duas apresentam, além desta, uma outra ORF com função desconhecida (WICKER et al., 2007).

Durante a transposição dos elementos desta classe, o transposon é cortado de sua localização genômica original e colado em um novo local. Ambas as reações são catalisadas pela transposase, que reconhece as repetições terminais invertidas e corta ambas as fitas de DNA em cada extremidade, ligando-as posteriormente ao seu sítio alvo (JURKA et al., 2007).

Estes elementos transponíveis podem aumentar seu número pela transposição durante a replicação dos cromossomos, partindo de uma região que já tenha sido replicada para outra onde a forquilha de replicação ainda não tenha passado (GREENBLATT; BRINK, 1962).

Wicker et al. (2007) incluíram também nesta subclasse a ordem *Crypton*, a qual tem sido encontrada apenas em fungos e que abriga somente a superfamília *Crypton*. Os elementos pertencentes a esta ordem contêm uma tirosina recombinase, mas não têm um domínio para transcriptase reversa, o que sugere que eles se transpõem através de um intermediário de DNA.

➤ Subclasse II

A subclasse II comporta elementos que envolvem, em seu processo de transposição, replicação sem a quebra da dupla fita de DNA, sendo por isso

acentuadamente diferentes dos elementos inclusos na subclasse I. Helitron e Maverick são as ordens que a compõem e ambas abrigam apenas uma superfamília, nomeadas *Helitron* e *Maverick*, respectivamente (WICKER et al., 2007).

A transposição de elementos da ordem Helitron, descrita por Kapitonov e Jurka (2001), começa a partir de uma quebra num sítio iniciador de replicação específico chamado Rep (do inglês - replication initiator) codificado no transposon. Em seguida a extremidade livre 3-OH, recém produzida pela quebra, funciona como um primer para a síntese da fita líder, facilitada pela helicase (a qual é codificada em uma ORF no transposon) e por algumas proteínas de replicação do hospedeiro, incluindo a DNA polimerase. Quando finda-se a replicação do transposon, o sítio iniciador de replicação catalisa uma reação de transferência de cadeia, que resulta em um DNA de fita dupla, composto por uma fita “velha” e por aquela recém-sintetizada, e um DNA de cadeia simples, proveniente de uma das fitas “velhas”, que é liberado e que então reintegra-se ao genoma hospedeiro.

Os elementos transponíveis da ordem Maverick, também conhecidos como Polintrons, são cercados por TIRs e podem codificar até onze proteínas. A grande maioria deles codifica a DNA polimerase B e uma integrase; no entanto, eles não contêm codificação para transcriptase reversa, sugerindo que eles passem por transposição replicativa sem intermediários de RNA. É proposto que a transposição ocorra através da excisão do transposon de uma única fita de DNA, seguida por replicação extracromossômica e integração em um novo sítio (KAPITONOV; JURKA, 2006).

2.1.1.3 O impacto dos elementos transponíveis na estrutura e funcionamento dos genomas

Durante muito tempo, mesmo após a descoberta dos elementos transponíveis por McClintock, acreditou-se que eles não desempenhavam

funções importantes na evolução dos genomas, sendo classificados, juntamente com todo o DNA repetitivo, como DNA-lixo e descartados de grande parte dos trabalhos científicos. No entanto, com a caracterização do primeiro elemento transponível em bactéria (JORDAN et al., 1967) e posterior reconhecimento desta classe como um componente antigo dos genomas eucarióticos e procarióticos, surgiram inúmeras especulações a respeito da função cumprida por eles e sua relação com a evolução.

Biémont e Vieira (2006) relatam resultados de trabalhos que também contribuíram para o advento dos estudos com elementos transponíveis, mostrando que a expressão destes é tecido-específica. Alguns são altamente expressos durante determinadas fases da vida do organismo hospedeiro e outros são até mesmo expressos de formas diferentes nas células reprodutivas masculinas e femininas. Como tais padrões específicos de expressão não eram esperados para “DNA não funcional”, esta porção do genoma passou a ser tratada de forma diferente pela comunidade científica interessada na compreensão do genoma.

Evidências circunstanciais de que genomas de plantas são muito dinâmicos, em grande parte, devido à atividade dos elementos transponíveis estão se acumulando rapidamente (FESCHOTTE; ZHANG; WESSLE, 2002). Segundo Biémont e Vieira (2006) o progresso na compreensão de como esses elementos são regulados tem trazido uma apreciação de como o ambiente de um indivíduo pode afetar a expressão de seu complemento genético na determinação do fenótipo. Os mesmos autores afirmam ainda ser uma opinião emergente a de que estes elementos têm tido uma influência significativa na evolução dos genomas, particularmente controlando a atividade dos genes.

Algumas das consequências da atividade dos elementos transponíveis em relação aos genes aos quais se associam são alterações funcionais: quando inseridos dentro da região codificante de um determinado gene podem inativá-lo ou alterar

seu produto e quando inseridos dentro ou próximos a regiões promotoras podem alterar os padrões de expressão do gene (KASHKUSH; FELDMAN; LEVY, 2003). Os elementos transponíveis podem ainda causar quebras cromossômicas (DERAGON; CAPY, 2000), mobilização de outras porções do DNA durante a transposição, recombinação ilegítima e, quando são replicativos, geram como consequência do seu modo de transposição o aumento do tamanho do genoma hospedeiro (FERDOROFF, 2000; KIDWELL; LISCH, 1997).

Além da evidente relação dos elementos transponíveis com fatores genéticos, eles estão também associados a fatores epigenéticos. De acordo com Jurka et al. (2007) as sequências repetitivas dispersas são, em sua maioria, inativas e estudos realizados por SanMiguel (1996) e Meyers, Tingey e Morgante (2001) em milho relacionam esta inatividade às metilações. Nesta espécie, os elementos geneticamente inativos são hipermetilados e seus homólogos ativos são hipometilados.

Considerando-se toda a plasticidade conferida aos genomas pelos elementos transponíveis, seria impossível, de acordo com Bennetzen (2000) ter qualquer concepção significativa da estrutura do genoma e evolução das plantas sem compreender as contribuições e complicações trazidas por eles aos seus hospedeiros.

2.1.2 Sequências repetitivas em tandem – DNA satélite

As sequências repetitivas em tandem são frequentemente referidas como DNA satélite por terem sido inicialmente isoladas de bandas satélite sobre centrifugação por gradiente de densidade (TEK et al., 2005). A principal característica deste tipo de DNA é o arranjo sequencial de suas unidades de repetição, também conhecidas como monômeros, as quais se organizam uma após a outra construindo grandes arranjos que podem ter até dezenas de megabases de comprimento (CHARLESWORTH; SNIEGOWSKI; STEPHANW, 1994; PLOHL et al., 2008).

De acordo com Slamovits e Rossi (2002) são reconhecidas três classes distintas de repetições em tandem, com diferentes comprimentos, propriedades e distribuições genômicas. Repetições de 1 a 5 pares de bases caracterizam os microsatélites, e de até 100 pares de bases os minissatélites. Em ambas as classes as repetições ocorrem entre 10 e 100 vezes. Já os satélites, a terceira classe, possuem unidades de repetição que variam de centenas a milhares de pares de bases que se repetem milhares de vezes no genoma.

Os micro e minissatélites são os maiores componentes da fração repetitiva dos genomas de muitas espécies (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998), são encontrados com abundância e espalhados por todas as regiões genômicas (VOGT, 1990), mas raramente se localizam em regiões gênicas (RAO et al., 2010). Eles são os componentes dos telômeros que por sua vez são estruturas responsáveis por manter a estabilidade e proteger os cromossomos (ZAKIAN, 1995). Micro e minissatélites são também os componentes das regiões hipervariáveis, também conhecidas como VNTRs (variable number of tandem repeats) que contêm diferentes números de cópias dessas classes de DNA repetitivo (RAO et al. 2010). Esta variação no número de cópias dos micro e minissatélites em determinadas regiões pode ser uma informação importante na diferenciação de subespécies e cultivares (SCHMIDT et al., 1993). Alguns motivos dessas sequências de DNA repetitivo podem ser usados como impressões digitais, razão pela qual eles são considerados valiosos marcadores moleculares (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998).

Os satélites são considerados agentes importantes para estabilização e manutenção das estruturas cromossômicas (CSINK; HENIKOFF, 1998; IRICK, 1994). O comprimento das unidades monoméricas das famílias de DNA repetido em tandem é variável, mas os tamanhos preferenciais são de 150-180 pb e de 320-360 pb, tamanhos similares ao comprimento da sequência de DNA necessária para contornar um ou dois nucleossomos

(HEMLEBEN, 1990; TRAUT, 1991). Muitos exemplos de DNA satélite com monômeros maiores foram encontrados e são na maioria provenientes de eventos de duplicação, inserção ou de rearranjos complexos de unidades menores (KUBIS; SCHMIDT; HESLOP-HARRISON, 1998).

Mais de 160 famílias de repetições satélite têm sido descritas em plantas superiores (MACAS; MESZAROS; NOUZOVA, 2002). A sequência de DNA ou distribuição genômica de cada uma delas pode ser específica para um determinado grupo de espécies, podendo ainda ser observada especificidade para genomas ou até mesmo para cromossomos (PREIZNER et al., 1994; WANG et al., 1995; ZHAO et al., 1989).

Em um único genoma é possível e provável que se encontre inúmeras famílias de DNA satélite, e é comum que elas compartilhem entre si apenas duas características: o arranjo em tandem de seus monômeros e sua localização heterocromática. Enquanto isso parece não haver relação para outras características, tais como sequência de nucleotídeos, tamanho dos monômeros, complexidade e história evolutiva. Dentro de cada família satélite, a análise das sequências revelou uma alta homogeneidade, geralmente maior que 95% (PLOHL, 2010).

É bem estabelecido que as repetições satélite estão maciçamente presentes nas regiões centroméricas e subteloméricas dos cromossomos e elas têm sido muitas vezes associadas à estrutura da heterocromatina, havendo indícios de que, além do papel estrutural, muitas dessas sequências desempenhem importantes funções nos cromossomos eucarióticos (TEK et al., 2005). No caso das repetições satélites centroméricas, acredita-se que elas interajam com o cinetócoro e que essas interações sejam mediadas por processos de reconhecimento DNA-proteína (JIANG et al., 2003).

Segundo Jurka et al. (2007), ao longo do tempo os elementos individuais de famílias repetitivas podem adquirir diferentes papéis biológicos. Por outro

lado, alguns tipos de repetições apresentam sequências conservadas entre espécies e muitas vezes essas sequências possuem uma função conhecida e bem determinada, como é o caso dos DNAs ribossômicos (HESLOP-HARRISON, 2000). Apesar dos muitos indícios e suposições a respeito das funções desempenhadas pelo DNA satélite, sabe-se muito pouco a respeito de quais seriam elas e como elas são de fato desempenhadas por esta fração do genoma.

A natureza repetitiva do DNA satélite dificulta a obtenção de informações a respeito da riqueza destas sequências nos genomas e sua organização pós-sequenciamento, motivo pelo qual esta porção do DNA, embora muito abundante, é na maioria, pobremente representada nos dados de sequenciamento (NAVAJAS-PÉREZ; PATERSON, 2009). No entanto, o desenvolvimento de novos algoritmos e de programas de análise mais adequados (MACAS; NEUMANN; NAVRÁTILOVÁ, 2007) em conjunto com a aplicação da hibridização *in situ* fluorescente (GONG et al., 2012; TORRES et al., 2011; ZHANG et al., 2014) tem permitido cada vez mais a obtenção de estudos detalhados da composição e organização física deste DNA repetitivo.

2.2 DNA centromérico e sua organização em *Solanum*

Os centrômeros funcionais dos cromossomos eucariontes são responsáveis pela coesão de cromátides irmãs e pela segregação cromossômica regular durante a mitose e meiose, as quais são essenciais para o desenvolvimento e proliferação celular em todos os organismos. Eles são o loco para organização do cinetócoro, uma estrutura proteica onde se ligam as fibras do fuso para promover a movimentação cromossômica (CLEVELAND; MAO; SULLIVAN, 2003).

A identidade e a herança do centrômero funcional têm sido associadas à organização da cromatina, que difere dos outros domínios cromossômicos pela presença de uma variante da histona H3, chamada de CenH3 e também conhecida por CENP-A. A grande diferença entre a histona H3 canônica e a

CenH3 é a presença, nesta última, de uma cauda N-terminal altamente divergente, e que de forma geral varia consideravelmente em comprimento e composição mesmo entre organismos intimamente relacionados, podendo ser espécie-específica. Localizada exclusivamente no centrômero de todos os eucariotos estudados, a CenH3 constitui o único marcador de funcionalidade dessa região (JIANG et al., 2003; WANG; ZHANG; JIN, 2009).

A forte relação da CenH3 com a funcionalidade centromérica inicialmente desfoçou a importância do DNA constituinte dos centrômeros. No entanto, a obtenção de sucesso na construção de cromossomos artificiais humanos obtida somente após utilização de DNA repetitivo na composição dos centrômeros (HARRINGTON et al., 1997; HENNING et al., 1999; IKENO et al., 1998) sugeriu que o DNA também seja importante para a função centromérica.

Os centrômeros, que desempenham uma função conservada nos eucariotos superiores, são compostos, em sua maioria, por longos arranjos de DNA satélite (JIANG et al., 2003) que são, no entanto, pouco conservados nos organismos modelo e em grande parte das demais espécies já estudadas até o momento. Os centrômeros de *Arabidopsis thaliana* (L.) Heynh., por exemplo, apresentam longos arranjos de DNA satélite compostos por monômeros de 178 pb, encontrados como componente principal de todos os centrômeros do complemento cromossômico da espécie (NAGAKI et al., 2003b). Em humanos, monômeros de tamanho similar (171 pb), conhecidos como α -satélite, são da mesma forma encontrados em todos os centrômeros (WILLARD, 1998), mas sua composição em pares de bases é totalmente diferente do DNA repetitivo presente nos centrômeros de *A. thaliana*. As sequências de DNA satélite que compõem os centrômeros podem ser altamente divergentes até mesmo entre espécies estreitamente relacionadas, como acontece no arroz. Em *Oryza sativa* L., principal espécie de arroz cultivada, são encontradas no centrômero repetições de DNA satélite de 155 pb que não partilham homologia alguma com as repetições de 154 pb encontradas em *Oryza brachyantha* A. Chev. &

Roehr., uma espécie de arroz selvagem próxima a *O. sativa* (LEE et al., 2005).

Ainda que as sequências de DNA sejam por vezes divergentes mesmo entre espécies próximas, as repetições satélite centroméricas são frequentemente homogeneizadas no genoma inteiro. Assim, na maioria dos eucariotos superiores que foram analisados, um único tipo de repetição satélite domina todos os centrômeros, como é o caso do DNA α -satélite em humanos (WILLARD, 1998). Sob este aspecto, a batata (*Solanum tuberosum* L.) e espécies a ela relacionadas constituem um caso bastante peculiar. Um estudo realizado por Gong et al. (2012) descreveu centrômeros que não apresentam como componente principal o DNA repetitivo. Os autores mostraram que dentre os 12 cromossomos da batata, cinco são compostos majoritariamente por DNA de cópia única ou com baixo número de cópias e um apresenta tanto DNA repetitivo quanto DNA não repetitivo. Os seis centrômeros restantes, por outro lado, são compostos por arranjos de repetições satélite, organizados, no entanto, de forma rara em comparação com outros eucariotos. Cada um dos seis centrômeros apresentou diferentes unidades de repetição, variando tanto em comprimento (de 979 pb a 5.4 kb) quanto na sequência de DNA. Ao descrever tais arranjos o trabalho mostrou que o nível de divergência das sequências de DNA satélite vai além do que se pensava, uma vez que elas variam não somente entre espécies relacionadas, mas também entre os diferentes centrômeros de um mesmo complemento cromossômico.

Outro trabalho que contribuiu para o melhor entendimento da complexidade centromérica na batata foi realizado por Zhang et al. (2014) com *Solanum verrucosum* Schltdl., uma espécie selvagem proximamente relacionada a *S. tuberosum* e já proposta como seu ancestral. Assim como na batata, os autores descreveram para *S. verrucosum* tanto centrômeros compostos por DNA repetitivo como por DNA de cópia única ou com baixo número de cópias. Além da descrição, foi feita uma comparação minuciosa entre os centrômeros de *S. verrucosum* e da batata, permitindo aos autores lançar luz sobre a evolução das sequências

centroméricas. Em *S. verrucosum* os centrômeros dos cromossomos 2, 4, 7 e 10 apresentam essencialmente DNA satélite. Os centrômeros dos mesmos cromossomos em *S. tuberosum* são compostos tanto por DNA repetitivo (2 e 7) quanto por DNA não repetitivo (4 e 10) e as sequências que os compõem não apresentam homologia com aquelas encontradas em *S. verrucosum*. Os autores consideram a hipótese de que as repetições presentes nos centrômeros 4 e 10 de *S. verrucosum* tenham invadido o genoma da espécie após sua divergência de *S. tuberosum*, que teria por sua vez mantido a composição do DNA centromérico com cópias únicas. Cinco diferentes repetições satélites estão presentes nos genomas das duas espécies e sequências homeólogas de cópia única foram encontradas em comum apenas no centrômero do cromossomo 9, revelando sua rápida evolução entre espécies proximamente relacionadas.

A existência de repetições cromossomo-específicas pode ser explicada pela evolução concertada, que tende a criar famílias de DNA satélite com alta homogeneidade interna. A homogeneização das sequências de DNA se deve a diversos mecanismos moleculares estocásticos de transferência não-recíproca que atuam de forma mais eficiente dentro de subconjuntos de sequências adjacentes. A eficiência cai progressivamente quando as mutações são homogeneizadas entre arranjos localizados no mesmo cromossomo, em cromossomos homólogos e finalmente em cromossomos heterólogos, podendo levar à formação de variantes de monômeros que são cromossomo-específicos. O favorecimento de uma variante em detrimento de outra pode indicar, do ponto de vista evolutivo, restrições impostas a algumas sequências satélite que estão relacionadas a papéis funcionais, como interações específicas que complementam os processos epigenéticos (PLOHL et al, 2008). Outra possibilidade é de que algumas repetições sejam favorecidas por apresentarem um comprimento ideal que possa favorecer, por melhor conformação estrutural, a estabilização dos nucleossomos centroméricos (ZHANG et al., 2014).

Esta hipótese foi comprovada por Zhang et al. (2013) em centrômeros de arroz, compostos pelas repetições satélite CentO de 155 pb e também por sequências de cópia única. Os autores observaram que 90-100 pb da repetição eram sempre protegidos pelos nucleossomos CenH3 da digestão com nuclease micrococcal, uma enzima que digere preferencialmente os *linkers*, mantendo o DNA que envolve os nucleossomos intacto. Os autores verificaram que havia, portanto, uma periodicidade na distribuição das unidades de repetição em comparação com a distribuição nucleossômica e que o comprimento de 155 pb dos monômeros é bastante apropriado para dar uma volta em torno do nucleossomo e compor o *linker*. Essa relação foi descrita até o momento somente para as repetições satélite presentes nos centrômeros do arroz e em humanos (HASSON et al., 2013).

Além dessa função estrutural, acredita-se que as repetições satélite interajam com a cromatina e participem da funcionalidade centromérica. Um forte indício disso é a presença de um motivo de 17pb denominado CENP-B box, encontrado nos monômeros alfa-satélite humanos, que se liga à proteína CEN-B, provavelmente como um evento importante para a formação do cinetócoro (IKENO; MASUMOTO; OKAZAKI, 1994; MASUMOTO; NAKANO; OHZEKI, 2004; SCHUELER et al., 2005). Motivos semelhantes ao CENP-B box foram observados em diversas famílias de DNA satélite, mas não se tem conhecimento ainda a respeito de seu verdadeiro significado funcional (CANAPA et al., 2000; LORITE et al., 2004).

2.3 Organização do genoma em Solanaceae

De acordo com Wu et al. (2006), ao contrário de muitas famílias vegetais, a organização do genoma em Solanaceae foi influenciada por taxas relativamente pequenas de poliploidização, quando comparadas às taxas observadas em outras famílias como Poaceae e Brassicaceae. Em *Solanum* L., um dos principais gêneros da família (KNAPP et al., 2004), a

maioria das espécies é diploide. A seção Petota desse gênero é composta por 228 espécies selvagens e sete espécies cultivadas (HAWKES, 1994), dentre elas a batata (*S. tuberosum*), e é subdividida em 18 séries, sendo elas: Juglandifolia, Etuberosa, Morelliformia, Bulbocastana, Pinnatisecta, Commersoniana, Circaeifolia, Conicibaccata, Piurana, Acaulia, Demissa, Longipedicellata, Polyadenia, Cuneolata, Megistracoloba, Ingaefolia, Olmosiana e Tuberosa (HAWKES, 1956, 1958, 1963, 1978).

Todas as espécies componentes dessa seção tem número básico de cromossomos $x=12$ (GAVRILENKO, 2007), morfologia cromossômica bastante similar (GAVRILENKO, 2007) e compõem uma série poliploide (RYBIN, 1929, 1933). Dentre elas, cerca de 73% são diploides com $2n=2x=24$ cromossomos, 4% triploides ($2n=3x=36$), 15% tetraploides ($2n=4x=48$), 2% pentaploides ($2n=5x=60$) e 6% hexaploides ($2n=6x=72$) (HAWKES, 1990).

Diferentes genomas foram reconhecidos na seção Petota após análises de informações como taxas de cruzamento em combinações interespecíficas, grau de homologia cromossômica entre as espécies envolvidas nas hibridações e viabilidade e fertilidade dos híbridos gerados (HAWKES, 1958; IRIKURA, 1976, MARKS, 1955; MATSUBAYASHI, 1961, 1981; RAMANNA; HERMSEN, 1979). Devido a diferenças na nomenclatura e a alguns conflitos existentes nas descrições desses genomas, Matsubayashi (1991) fez uma reavaliação das relações genômicas entre as espécies de *Solanum* e procurou unificar as descrições já existentes, propondo a existência de cinco genomas diferentes: A, B, C, D e P, todos encontrados em espécies que tuberizam.

De acordo com Matsubayashi (1991) várias espécies sul-americanas, incluindo a espécie mexicana *Solanum verrucosum*, partilham um mesmo genoma, tendo sido ele designado genoma A e considerado o mais básico dentre os existentes. Os outros genomas receberam tais letras como identificação devido às iniciais dos nomes das séries nas quais foram primeiramente

identificados: genoma B - série *Longipedicellata* Bukasov, genoma C – série *Conicibaccata* Bitter, genoma D – série *Demissa* Bukasov e genoma P – série *Piurana* Hawkes.

Os genomas B, C, D e P foram inicialmente considerados por Matsubayashi (1991) como componentes genômicos de espécies aloploidoides e não foram identificados em nenhuma espécie diploide por meio de análises genômicas clássicas. A identificação em espécies diploides apenas foi possível após utilização da técnica de Hibridização genômica *in situ* (PENDINEN et al., 2008, 2012) e a disponibilidade e análise bioinformática de dados de sequenciamento do gene para a nitrato redutase (RODRÍGUEZ; SPOONER, 2009) e do gene nuclear Waxy (GBSSI - Granule-bound starch synthase), responsável pela síntese de grânulos de amido (SPOONER et al., 2008). As análises revelaram espécies diploides tais como *S. cardiophyllum* Lindl., *S. ehrenbergii* Rydb. e *S. jamesii* Torr. portadoras do genoma B e *S. chomatophilum* Bitter, *S. andreanum* Baker e *S. piurae* Bitter portadoras do genoma P. Os genomas C e D não foram identificados em espécies diploides (SPOONER et al., 2008).

Um genoma adicional, E, foi reconhecido na seção *Etuberosum*, que é proximamente relacionada à seção *Petota* e é composta por espécies que não tuberizam, tais como *S. etuberosum* Lindl. e *S. palustre* Poepp. ex Schltdl. (RAMANNA; HERMSEN, 1981). Por vezes é encontrada na literatura (GONG et al., 2012; TORRES et al., 2011; ZHANG et al., 2014) a citação de um outro genoma no gênero *Solanum*, o genoma T, pertencente ao tomate (*Solanum lycopersicum* L.).

A batata cultivada *Solanum tuberosum* é uma espécie que apresenta genoma A. De acordo com a árvore filogenética utilizada por Gong et al. (2012), construída com base nos trabalhos de análise de DNA plastidial de Castillo e Spooner (1997), Spooner e Castillo (1997) e Spooner, Anderson e Jansen

(1993), o genoma mais semelhante ao genoma A de *S. tuberosum* é o genoma P, seguido pelos genomas B, T e E, nessa ordem.

A batata cultivada, *S. tuberosum* é uma espécie tetraploide, com genoma AAAA e é considerada autotetraploide, principalmente com base no pareamento cromossômico quadrivalente que é observado durante a meiose da espécie (SPOONER et al., 2008). Hawkes et al. (1990) propuseram *Solanum verrucosum* como sua progenitora, mas de acordo com Spooner et al. (2008) a identidade deste genoma A ainda tem sido objeto de investigação, debate e incerteza.

Os cromossomos metafásicos de *S. tuberosum* são bem pequenos, medindo de 1.0 a 3.5 μm e sua identificação é bastante complicada dada a semelhança morfológica entre eles, dificuldade diminuída pela criação de uma biblioteca de BACs para identificação de cada um dos cromossomos do complemento (DONG et al., 2000).

Em 2011, o Consórcio para o Sequenciamento do Genoma da Batata (PGSC) publicou o sequenciamento de um genótipo monoploide, com estimativa de cerca de 39.000 genes codificadores de proteínas e pelo menos 62% de sequências repetidas (PGSC, 2011). A partir destes dados Torres et al. (2011) identificaram e localizaram, por meio de FISH, duas sequências repetitivas subteloméricas (CL14 e CL34), as quais estavam presentes em quase todas as extremidades dos 24 cromossomos. A repetição CL14 foi detectada nas regiões subteloméricas entre espécies do gênero *Solanum* altamente divergentes, incluindo o tomate. Em contraste, CL34 foi encontrada apenas na batata e em espécies intimamente relacionadas. O estudo revelou que as sequências estão arrançadas independentemente e que têm dinâmicas evolutivas distintas no gênero, sendo CL14 mais antiga e CL34 mais recente e em rápida evolução e forneceu novas evidências de que os domínios subteloméricos estão entre as mais dinâmicas regiões nos genomas eucarióticos.

Outro trabalho realizado no sentido de estudar o DNA repetitivo na batata foi realizado por Tek et al. (2005) que relataram, em uma espécie de batata diplóide (*Solanum bulbocastanum* Dunal), uma repetição satélite nomeada Sobo, não detectada em outras espécies de *Solanum*. A repetição Sobo foi mapeada para uma localização única na região pericentromérica do cromossomo 7. A análise da sequência revelou que a maior parte dos monômeros de Sobo tem similaridade significativa com LTRs de retrotransposons presentes no genoma da espécie.

Outro elemento repetitivo caracterizado no genoma da batata foi o elemento de Repetição Telomérica Intercalar pSbTC1 (ITR - Intercalary Telomeric Repeat), isolado da espécie diploide *Solanum bulbocastanum* Dunal. O elemento tem cerca de 2.8 kb, apresenta organização em tandem, localização principal nas regiões centroméricas dos cromossomos da batata e é amplamente distribuído em diversas espécies de *Solanum* selvagens. Os autores relatam que uma única cópia deste elemento pode ser amplificada ao ponto de ocupar vários megabases e por isso sugerem que ele não seja simplesmente um resquício de fusão cromossômica. Eles demonstram ainda que estas sequências são altamente metiladas e associadas com a condensação da heterocromatina centromérica (TEK; JIANG, 2004). He et al. (2013) também encontraram ITRs nas regiões centroméricas de alguns cromossomos de seis espécies de *Solanum* e relatam que existe uma amplificação massiva desses elementos principalmente em espécies portadoras dos genomas B e P. Análises de cromossomos paquitênicos de *Solanum pinnatisectum* Dunal, genoma B, realizadas pelos autores mostraram que pelo menos nove dos 12 cromossomos apresentam ITRs nas regiões centroméricas, sendo que em três deles os sinais de FISH estão associados inequivocamente aos centrômeros.

As repetições em tandem também foram estudadas na batata por Stupar et al. (2002), que relataram uma classe única de heterocromatina pericentromérica, nomeada 2D8, composta por sequências repetitivas de DNA dispostas em tandem e

altamente homólogas aos espaçadores intergênicos dos genes para RNA ribossomal 18S e 25S. De acordo com os autores há uma grande possibilidade de que os elementos 2D8 tenham derivado dos espaçadores intergênicos, uma vez que vários estudos prévios encontraram evidências de sequências relacionadas ao DNA ribossômico dispersas por todo o genoma eucariótico.

Zhu et al. (2008) analisaram 90 Mb do genoma da batata, o que representa 10,2% do genoma total, e fizeram comparações com dados referentes ao genoma do tomate. Os autores observaram a conservação da ordem e estrutura gênica entre as duas espécies. A sintenia observada, embora alta, não foi absoluta, indicando a ocorrência de inserções, deleções e microinversões desde a divergência entre o tomate e a batata. Apesar da aparente conservação gênica, houveram divergências no teor e composição das sequências de DNA repetitivo (46,29% no tomate e 34,18% na batata), as quais causaram, de acordo com os autores, impactos na arquitetura do genoma. Em relação à natureza destas sequências os autores identificaram, em ambas espécies, cinco principais classes: retrotransposons, transposons, DNA repetitivo relacionado aos telômeros, DNA ribossomal e sequências repetitivas não classificadas (mais da metade das sequências identificadas). Dentre as sequências de DNA repetitivo classificadas, os retrotransposons foram as mais abundantes, totalizando 11,77% na batata e 14,54% no tomate. Foram detectados mais retrotransposons do tipo *Ty3-gypsy* (4,60%) do que *Ty1-copia* (1,82%) no conjunto de dados de batata, e esta diferença no tomate foi muito menor, com 4,61% de *Ty3-gypsy* e 3,75% de *Ty1-copia*.

Um grande progresso no conhecimento genético do tomate (*Solanum lycopersicum*) foi obtido em 2012 com o sequenciamento de seu genoma (TOMATO GENOME CONSORTIUM, 2012). O tamanho genômico estimado foi de 900 Mb, consistente com estimativas previstas. Uma comparação com os dados do sequenciamento do genoma da batata (PGSC, 2011) mostrou 8,7% de divergência em relação aos nucleotídeos das regiões eucromáticas e mais de

30% de divergência nas regiões intergênicas e heterocromáticas, estas últimas ricas em DNA repetitivo. Foram identificados 18.320 genes ortólogos. As análises apontaram ainda a ocorrência de inversões cromossômicas, nove grandes e várias pequenas, envolvidas na diferenciação dos dois genomas.

Os estudos com DNA repetitivo em tomate, que tal como a batata é uma das peças centrais na investigação genética e molecular das solanáceas (VAN DER HOEVEN et al., 2002), tiveram os primeiros resultados relevantes em 1985, quando foi descoberta a primeira sequência repetitiva dispersa na espécie: THG2 (ZABEL et al., 1985). Posteriormente outros trabalhos foram realizados como o de Van der Hoeven et al. (2002) que analisaram uma coleção de 120.892 ESTs (Expressed Sequence Tag), derivadas de 26 bibliotecas de cDNA diferentes, trouxe importantes contribuições para o entendimento da organização do genoma desta espécie e dos processos evolutivos que lhe deram origem, uma vez que os dados foram comparados a dados do genoma de *Arabidopsis* (DC.) Heynh. Foram identificadas 27.274 sequências consenso únicas (unigenes), das quais 70% têm homólogos identificáveis no genoma de *Arabidopsis*. Os genes correspondentes ao metabolismo mantiveram-se mais conservados entre estes dois genomas, enquanto genes que codificam fatores de transcrição estão entre os que evoluem mais rapidamente. Grande parte das 10 maiores famílias multigênicas conservadas partilham números de cópias semelhantes em tomate e *Arabidopsis*, sugerindo que a multiplicidade dessas famílias pode ter ocorrido antes da divergência das duas espécies. Finalmente, seis clones BAC de diferentes partes do genoma do tomate foram isolados, geneticamente mapeados, sequenciados e anotados. A análise combinada da base de dados de ESTs e estes seis BACs sequenciados levou à previsão de que o genoma do tomate codifica 35.000 genes, presentes em grande parte nas regiões eucromáticas e correspondentes a menos de um quarto do DNA total.

Outro trabalho para caracterização do DNA repetitivo no tomate foi realizado por Chang et al. (2008) utilizando Cot-DNA, uma técnica que ajuda a mensurar quanto de DNA repetitivo existe em determinada amostra. Os autores utilizaram o DNA genômico total e identificaram pools representativos da fração repetitiva do tomate: Cot-1, Cot-10 e Cot-100. Os resultados com FISH mostraram que Cot-1 localizou-se mais intensamente nas extremidades dos cromossomos e nos organizadores nucleolares do cromossomo 2, com alguns sinais fracos em regiões distais e pericentroméricas. Os autores acreditam que esta fração seja, portanto, composta principalmente por rDNA 45S e repetições teloméricas. A FISH com a fração Cot-10 também mostrou sinais na maioria das extremidades dos cromossomos e em regiões pericentroméricas, enquanto que Cot-100 está presente em todas as regiões heterocromáticas. Os resultados permitiram concluir que a maioria do DNA repetitivo no tomate está confinado a regiões heterocromáticas nos telômeros, nas regiões organizadoras dos nucléolos e nos pericentrômeros.

Os telômeros do tomate são compostos pela repetição TT(T/A)AGGG arranjada em tandem (GANAL; BROUN; TANKSLEY, 1992). Uma repetição satélite de 162 pb, conhecida como TGR1, está presente nos subtelômeros de 20 dos 24 cromossomos da espécie (LAPITAN; GANAL; TANKSLEY, 1989; SCHWEIZER et al., 1988). Zhong et al. (1998) utilizaram a técnica de fiber-FISH para analisar com mais detalhamento a organização dos arranjos de ambas as repetições. Eles verificaram que os sinais gerados pelas repetições teloméricas eram sempre separados daqueles gerados pelas repetições subteloméricas por um espaço sem fluorescência, uma sequência espaçadora. Medições realizadas nas fibras revelaram uma grande variação (223 – 1.330 kb) no comprimento tanto dos blocos constituídos pelas repetições teloméricas, quanto naqueles formados por TGR1, sugerindo que cada cromossomo tem seu padrão único de organização destas sequências (ZHONG et al., 1998).

Um mapa molecular de ligação genética baseado no cDNA, DNA genômico e em marcadores EST de tomate foi construído para berinjela, *Solanum melongena* L., por Doganlar et al. (2002). O mapa composto por 12 grupos de ligação, abrangendo 1480 cM e com 233 marcadores revelou a conservação de grandes extensões de marcadores colineares, uma característica comum da evolução do genoma em Solanaceae e outras famílias de plantas. No geral, berinjela e tomate foram diferenciados por 28 rearranjos, o que poderia ser explicado por 23 inversões paracêntricas e cinco translocações na evolução a partir do último ancestral comum das espécies, sugerindo que a inversão paracêntrica tem sido o principal mecanismo de evolução cromossômica na família Solanaceae.

REFERÊNCIAS

- AGRA, M. F. New Species of *Solanum* subgenus e *Leptostemonum* (Solanaceae) from Chapada da Diamantina, Bahia, Brazil. **Novon**, Saint Louis, v. 9, p. 292-295, 1999.
- ALFENITO, M. R.; BIRCHLER, J. A. Molecular characterization of a maize B chromosome centric sequence. **Genetics**, Austin, v. 135, p. 589–597, 1993.
- ANANIEV, E. V. et al. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. **Proceedings of the national academy of sciences of the United States of America**, Washington, v. 95, p. 13073–13078, 1998.
- ARKHIPOVA, I. R. Distribution and Phylogeny of *Penelope-Like* Elements in Eukaryotes. **Systematic Biology**, Oxford, v. 55, n. 6, p. 875-885, 2006.
- BENNETZEN, J. L. Transposable element contributions to plant gene and genome evolution. **Plant Molecular Biology**, Dordrecht, v. 42, p. 251-269, 2000.
- BIÉMONT, C.; VIEIRA, C. Genetics: junk DNA as an evolutionary force. **Nature**, London, v. 443, n. 521-524, 2006.
- CANAPA, A. et al. A satellite DNA containing CENP-B box-like motifs is present in the antarctic scallop *Adamussium colbecki*. **Gene**, Amsterdam, v. 247, p. 175–180, 2000.
- CASTILLO, R. O.; SPOONER, D. M. Phylogenetic relationships of wild potatoes, *Solanum* series *Conicibaccata* (sect. *Petota*). **Systematic Botany**, Kent, v. 22, p. 45–83, 1997.

CHANDLER, M.; MAHILLON, J. Insertion sequences revisited. In: CRAIG, N. et al. (Ed.). **Mobile DNA II**. Herndon: ASM, 2002. p. 305–366.

CHANG, S. et al. "FISH mapping and molecular organization of the major repetitive sequences of tomato." **Chromosome Research**, Oxford, v. 16, n. 7, p. 919-933, 2008.

CHARLESWORTH, B.; SNIEGOWSKI, P. E.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, London, v. 37, p. 215–220, 1994.

CHENG, Z. K. et al. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. **Plant Cell**, Rockville, v. 14, p. 1691–1704, 2002.

CLEVELAND, D. W. et al. "Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling." **Cell Research**, Beijing, v. 112, n. 4, p. 407-421, 2003.

COPENHAVER, G. P. et al. Genetic definition and sequence analysis of Arabidopsis centromeres. **Science**, Washington, v. 286, p. 2468–2474, 1999.

CSINK, A. K.; HENIKOFF, S. Something from nothing: the evolution and utility of satellite repeats. **Trends in Genetics**, Cambridge, v. 14, p. 200-204, 1998.

DERAGON, J. M.; CAPY, P. Impact of transposable elements on the human genome. **Annals of Medicine**, Helsinki, v. 32, p. 264–273, 2000.

DOGANLAR, S. et al. A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. **Genetics**, Austin, v. 161, p. 1697–1711, Aug. 2002.

DONG, F. et al. Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. **Theoretical and applied genetics**, Berlin, v. 101, p. 1001–1007, 2000.

DONG, F. et al. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 95, p. 8135–8140, 1998.

EICKBUSH, T. H.; MALIK, H. S. **Mobile DNA II**. Herndon: ASM, 2002. p. 1111–1146

EVGEN'EV, M. B.; ARKHIPOVA, I. R. Penelope-like elements: a new class of retroelements: distribution, function and possible evolutionary significance. **Cytogenetic and Genome Research**, New York, v. 110, p. 510–521, 2005.

FERDOROFF, N. Transposons and genome evolution in plants. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 20, n. 13, p. 7002–700, 2000.

FESCHOTTE, C.; ZHANG, X.; WESSLER, S. Miniature inverted repeat transposable elements (MITEs) and their relationship with established DNA transposons. In: CRAIG, N. L. et al. (Ed.). **Mobile DNA II**. Washington: American Society for Microbiology, 2002. p. 1147–1158.

FLAVELL, A. J. et al. Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. **Nucleic Acids Research**, London, v. 20, p. 3639–3644, 1992.

GANAL, M. W.; BROUN, P.; TANKSLEY, S. D. Genetic mapping of tandemly repeated telomeric DNA sequences in tomato (*Lycopersicon esculentum*). **Genomics**, San Diego, v. 14, p. 444–448, 1992.

GAVRILENKO, T. "Potato cytogenetics." **Potato biology and biotechnology: Advances and Perspectives**, Wageningen, v. 10, p. 203-216, 2007.

GONG, Z. et al. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. **Plant Cell**, Rockville, v. 24, n. 9, p. 3559-3574, 2012.

GOODWIN, T. J.; POULTER, R. T. A new group of tyrosine recombinase-encoding retrotransposons. **Molecular Biology and Evolution**, Chicago, v. 21, p. 746-759, 2004.

GOODWIN, T. J.; POULTER, R. T. The DIRS1 group of retrotransposons. **Molecular Biology and Evolution**, Chicago, v. 18, p. 2067-2082, 2011.

GREENBLATT, I. M.; BRINK, R. A. Twin mutations in medium variegated pericarp maize. **Genetics**, Austin, v. 47, p. 489-501, 1962.

HARRINGTON, J. J. et al. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. **Nature Genetics**, New York, v. 15, p. 345-355, 1997.

HASSON, D. et al. The octamer is the major form of CENP-A nucleosomes at human centromeres. **Nature Structural & Molecular Biology**, New York, v. 20, n. 6, p. 687-695, 2013.

HAWKES, J. G. **A revision of the tuber-bearing Solanums**. Midlothian: Scottish Plant Breeding Station, 1963. p. 76-181.

HAWKES, J. G. "Biosystematics of the potato". In: _____. **The potato crop**. Dordrecht: Springer, 1978. p. 15-69.

HAWKES, J. G. Kartoffel: Taxonomy, cytology and crossability. In: KAPPERT, H.; RUDORF, W. (Ed.). **Handbuch Pflanzenzüchtung**. Berlin: P. Parey, 1958. p. 1–43.

HAWKES, J. G. Origins of cultivated potatoes and species relations hips. In: BRADSHAW, J. E.; MACKAY, G. R. (Ed.). **Potato genetics**. Wallingford: CAB International, 1994. p. 3-42.

HAWKES, J. G. "Taxonomic studies on the tuber-bearing solanums. 1: Solanum Tuberosum and the tetraploid species complex." **Proceedings of the Linnean Society of London**, Malden. v. 166, n. 1/2, p. 97-144, 1956.

HAWKES, J. G. **The Potato**: evolution, biodiversity, and genetic resources London: Belhaven, 1990. 259 p.

HE, L. et al. "Interstitial telomeric repeats are enriched in the centromeres of chromosomes in Solanum species." **Chromosome Research**, Oxford, v. 21, n. 1, p. 5-13, 2013.

HEMLEBEN, V. **Molekularbiologie der Pflanzen**. Stuttgart: G. Fischer, 1990.

HENNING, K. A. et al. Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 96, p. 592-597, 1999.

HESLOP-HARRISON, J. S. Comparative genome organization in plants: from sequence and markers to Chromatin and Chromosomes. **The Plant Cell**, Rockville, v. 12, p. 617-635, 2000.

IKENO, M. et al. Construction of YAC-based mammalian artificial chromosomes. **Nature Biotechnology**, New York, v. 16, p. 431-439, 1998.

IKENO, M.; MASUMOTO, H.; OKAZAKI, T. Distribution of CENP-B boxes reflected in CREST centromere antigenic sites on long-range alpha-satellite DNA arrays of human chromosome 21. **Human Molecular Genetics**, Oxford, v. 3, p. 1245–1257, 1994.

IRICK, H. A new function for heterochromatin. **Chromosoma**, Berlin, v. 103, p. 1-3, 1994.

IRIKURA, Y. Cytogenetic studies on the haploid plants of tuberbearing Solanum species. **Research Bulletin of the Hokkaido National Agricultural Research Station**, Hamilton, v. 115, p. 1–80, 1976.

JIANG, J. et al. A conserved repetitive DNA element located in the centromeres of cereal chromosomes. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, p. 14210–14213, 1996.

JIANG, J. et al. A molecular view of plant centromeres. **Trends in Plant Science**, Oxford, v. 8, n. 12, p. 570-575, 2003.

JORDAN, M. A.; SAEDLER, H.; STARLINGER, P. Strong-polar mutations in the transferase gene of the galactose operon E. coli. **Molecular Genetics and Genomics**, Berlin, v. 100, p. 296-306, 1967.

JURKA, J. et al. Repetitive sequences in complex genomes: structure and evolution. **Annual Review of Genomics and Human Genetics**, Baltimore, v. 8, p. 241-259, 2007.

KAMM, A. et al. Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. **Molecular Genetics and Genomics**, Berlin, v. 244, p. 420–425, 1994.

KAPITONOV, V. V.; JURKA, J. Rolling-circle transposons in eukaryotes. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 98, p. 8714–8719, 2001.

KAPITONOV, V.; JURKA, J. Self-synthesizing DNA transposons in eukaryotes. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 103, p. 4540–4545, 2006.

KASHKUSH, K.; FELDMAN, M.; LEVY, A. A. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. **Nature Genetics**, New York, v. 33, p. 102–106, 2003.

KIDWELL, M. G.; LISCH, D. R. Transposable elements as sources of variation on animals and plants. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 94, n. 15, p. 7704–7711, 1997.

KISHII, M. et al. A tandem repetitive sequence located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes. **Chromosome Research**, Oxford, v. 9, p. 417–428, 2001.

KNAPP, S. Solanaceae: a model for linking genomics with biodiversity. **Comparative and Functional Genomics**, Chichester, v. 5, p. 285–291, 2004.

KUBIS, S. E.; SCHMIDT, T.; HESLOP-HARRISON, J. S. Repetitive DNA elements as a major component of plant genomes. **Annals of Botany**, Oxford, v. 82, p. 45–55, 1998.

KUMAR, A.; BENNETZEN, J. Plant retrotransposons. **Annual Review of Genetics**, Palo Alto, v. 33, p. 479–532, 1999.

LAPITAN, N. L. V.; GANAL, M. W.; TANKSLEY, S. D. Somatic chromosome karyotype of tomato based on in situ hybridization of the TGRI satellite repeat. **Genome**, Austin, v. 32, p. 992–998, 1989.

LEDUR et al. **Novos rumos da genética do desenvolvimento avícola**. 2011. Disponível em: <<http://www.infoteca.cnptia.embrapa.br/bitstream/doc/902668/1/aviculturamonica0001.pdf>>. Acesso em: 6 jul. 2012.

LEE, H. R. et al. Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 102, p. 11793–11798, 2005.

LEETON, P. R.; SMYTH, D. R. An abundant LINE-like element amplified in the genome of *Lilium speciosum*. **Molecular Genetics and Genomics**, Berlin, v. 237, p. 97–104, 1993.

LORITE, P. et al. Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). **Gene**, Amsterdam, v. 332, p. 159–168, 2004.

MACAS, J.; MESZAROS, T.; NOUZOVA, M. PlantSat: a specialized database for plant satellite repeats. **Bioinformatics**, Oxford, v. 18, p. 28–35, 2002.

MACAS, J.; NEUMANN, P.; NAVRÁTILOVÁ, A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. **BMC Genomics**, San Diego, v. 8, p. 427, 2007.

MARCO, A.; MARÍN, I. Retrovirus-like elements in plants. **Recent Research Development in Plant Science**, Kerala, v. 3, p. 1-10, 2005.

MARKS, G. E. Cytogenetic studies in tuberous Solanum species. I. Genomic differentiation in the group Demissa. **Journal of Genetics**, Bangalore, v. 53, n. 2, p. 262–269, 1955.

MARTINEZ-ZAPATER, J. M. et al. A high repeated DNA sequence in Arabidopsis thaliana. **Molecular Genetics and Genomics**, Berlin, v. 204, p. 417–423, 1996.

MASUMOTO, H.; NAKANO, M.; OHZEKI, J. The role of CENP-B and alphsatellite DNA: de novo assembly and epigenetic maintenance of human centromeres. **Chromosome Research**, Oxford, v. 12, p. 543–556, 2004.

MATSUBAYASHI, M. "Cytogenetic studies in Solanum, Section **Tuberosum**, with special reference to the interspecific relationships." 1961. 121 p. Thesis (Ph.D) – Kyoto University, Kyoto, 1961.

MATSUBAYASHI, M. Phylogenetic relationships in the potato and its related species. In: TSUCHIYA, T.; GUPTA, P. K. (Ed.). **Chromosome engineering in plants: genetics, breeding, evolution**. Amsterdam: Elsevier Science, 1991. p. 93-118.

MATSUBAYASHI, M. "Species differentiation in tuberous Solanum and the origin of cultivated potatoes." **Recent Advances in Breeding**, Tokyo, v. 22, p. 86-106, 1981.

MEHROTRA, S, GOYAL, V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. **Genomics, Proteomics & Bioinformatics**, Beijing, v. 12, p. 164–171, 2014.

MEYERS, B. C.; TINGEY, S. V.; MORGANTE, M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. **Genome Research**, New York, v. 11, p. 1660–1676, 2001.

MILLER, J.T. et al. Cloning and characterization of a centromerespecific repetitive DNA element from *Sorghum bicolor*. **Theoretical and Applied Genetics**, Berlin, v. 96, p. 832–839, 1998.

MORGANTE, M. Plant genome organisation and diversity: the year of the junk! **Current Opinion in Biotechnology**, London, v. 17, p. 168–173, 2005.

NAGAKI, K. et al. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. **Genetics**, Austin, v. 163, p. 1221–1225, 2003b.

NAGAKI, K. et al. Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. **Genetics**, Austin, v. 163, p. 759–770, 2003a.

NAVAJAS-PÉREZ, R.; PATERSON, A. H. Patterns of tandem repetition in plant whole genome assemblies. **Molecular Genetics and Genomics**, Berlin, v. 281, p. 579–590, 2009.

NEUMANN, P.; POZARKOVA, D.; MACAS, J. Highly abundant pea LTR retrotransposon *Ogre* is constitutively transcribed and partially spliced. **Plant Molecular Biology**, Dordrecht, v. 53, p. 399–410, 2003.

NONOMURA, K. I.; KURATA, N. Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. **Molecular Genetics and Genomics**, Berlin, v. 261, p. 1–10, 1999.

PENDINEN, G. et al. Allopolyploid speciation of the Mexican tetraploid potato species *Solanum stoloniferum* and *S. hjertingii* revealed by genomic in situ hybridization. **Genome**, Ottawa, v. 51, n. 9, p. 714–720, 2008.

PENDINEN, G. et al. "Genomic in situ hybridization reveals both auto-and allopolyploid origins of different North and Central American hexaploid potato (*Solanum* sect. *Petota*) species." **Genome**, Ottawa, v. 55, n. 6, p. 407–415, 2012.

PETERSON, D. et al. Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene Discovery **Genome Research**, New York, v. 12, p. 795–807, 2002.

PIEDNOËL, M. et al. Eukaryote DIRS1-like retrotransposons: an overview. **BMC Genomics**, San Diego, v. 12, p. 621, 2011.

PIMPINELLI, S. et al. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 92, p. 3804–3808, 1995.

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

PLOHL, M. Those mysterious sequences of satellite DNAs. **Periodicum Biologorum**, Zagreb, v. 112, p. 403–410, 2010.

POTATO GENOME SEQUENCING CONSORTIUM. Genome sequence and analysis of the tuber crop potato. **Nature**, London, v. 475, p. 189–194, 2011.

PREIZNER, J. et al. Organization of a *Solanum brevidens* repetitive sequence related to the TGRI subtelomeric repeats of *Lycopersicon esculentum*. **Theoretical and Applied Genetics**, Berlin, v. 89, p. 1-8, 1994.

RAMANNA, M. S.; HERMSEN, J. G. T. Structural hybridity in the series *Etuberosa* of the genus *Solanum* and its bearing on crossability. **Euphytica**, Wageningen, v. 30, p. 15-31, 1981.

RAMANNA, M. S.; HERMSEN, J. G. T. "Unique meiotic behaviour in F1 plants from a cross between a non-tuberosus and a tuberosus *Solanum* species in section *Petota*." **Euphytica**, Wageningen, v. 28, n. 1, p. 9-15, 1979.

RAO, S. R. et al. DNA repetitive sequences-types, distribution and function: a review. *Journal of Molecular Cell Biology*, Oxford, v. 7, n. 2, p. 1-11, 2010.

RODRÍGUEZ, F.; SPOONER, D. M. Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. **Systematic Botany**, Kent, v. 34, p. 207-219, 2009.

RYBIN, V. A. "Cytological investigation of the South American cultivated and wild potatoes, and its significance for plant breeding." **Trudy po Prikladnoj Botanike Genetike i Selekcii**, Leningrad, v. 2, p. 3-100, 1933.

RYBIN, V. A. "Karyological investigation on some wild growing and indigenous cultivated potatoes of America." **Trudy po Prikladnoj Botanike Genetike i Selekcii**, Leningrad, v. 20, p. 655-720, 1929.

SANMIGUEL, P.; BENNETZEN, J. L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. **Annals of Botany**, London, v. 82, p. 37-44, 1998.

SANMIGUEL, P. et al. Nested retrotransposons in the intergenic regions of the maize genome. **Science**, Washington, v. 274, p. 765–768, 1996.

SANMIGUEL, P. et al. The paleontology of intergene retrotransposons in maize. **Nature Genetics**, New York, v. 20, p. 43–45, 1998.

SCHMIDT, T.; HESLOP-HARRISON, J. S. Variability and evolution of highly repeated DNA sequences in the genus Beta. **Genome**, Ottawa, v. 36, p. 1074–1079, 1993.

SCHUELER, M. G. Genomic and genetic definition of a functional human centromere. **Science**, Washington, v. 294, p. 109–115, 2001.

SCHUELER, M. G. et al. Progressive proximal expansion of the primate X chromosome centromere. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 102, p. 10563–10568, 2005.

SCHWEIZER, G. et al. Species-specific DNA sequences for the identification of somatic hybrids between *Lycopersicon esculentum* and *Solanum acaule*. **Theoretical and Applied Genetics**, Berlin, v. 75, p. 679–684, 1988.

SHAPIRO, J. A.; VON STERNBERG, R. Why repetitive DNA is essential to genome function. **Biological Reviews**, Cambridge, v. 80, p. 227–250, 2005.

SHARMA, S.; RAINA, S. N. Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. **Cytogenetic and Genome Research**, New York, v. 109, n. 15, p. 15–26, 2005.

SLAMOVITS, C. H.; ROSSI, M. S. Satellite DNA: agent of chromosomal evolution in mammals: a review. **Mastozoología Neotropical**, Mendoza, v. 9, n. 2, p. 297–308, 2002.

SOLANACEAE GENOMICS RESOURCE. Disponível em: <<http://solanaceae.plantbiology.msu.edu/>>. Acesso em: 6 jul. 2012.

SPOONER, D. M.; ANDERSON, G. J.; JANSEN, R. K. Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). **American Journal of Botany**, Baltimore, v. 80, p. 676–688, 1993.

SPOONER, D. M.; CASTILLO, R. Reexamination of series relationships of South American wild potatoes (Solanaceae: Solanum sect. Petota): evidence from chloroplast DNA restriction site variation. **American Journal of Botany**, Baltimore, v. 84, p. 671–685, 1997.

SPOONER, D. M. et al. "Genomic origins of potato polyploids: GBSSI gene sequencing data." **Crop Science**, Madison, v. 48, p. 27, 2008. Suppl. 1.

STUPAR, R. M. et al. Highly condensed Potato pericentromeric heterochromatin contains rDNA-Related tandem repeats. **Genetics**, Austin, v. 162, p. 1435–1444, 2002.

TEK, A. L. et al. Sobo, a recently amplified satellite repeat of Potato, and its implications for the origin of tandemly repeated sequences. **Genetics**, Austin, v. 170, p. 1231–1238, 2005.

TEK, A. L.; JIANG, J. The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. **Chromosoma**, Berlin, v. 113, p. 77–83, 2004.

TOMATO GENOME CONSORTIUM. "The tomato genome sequence provides insights into fleshy fruit evolution." **Nature**, London, v. 485, n. 7400, p. 635–641, 2012.

TORRES, G. A. et al. Organization and evolution of subtelomeric satellite repeats in the Potato genome. **G3**, Bethesda, v. 1, n. 2, p. 85-92, July 2011.

TRAUT, W. **Chromosomen. Klassische und molekulare Zytogenetik**. Heidelberg: Springer-Verlag, 1991.

VAN DER HOEVEN, R. et al. Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of large expressed sequence tag collection and selective genomic sequencing. **The Plant Cell**, Rockville, v. 14, p. 1441-1456, 2002.

VICIENT, C. M. et al. Structure, functionality, and evolution of the BARE - 1 retrotransposon of barley. **Genetica**, Dordrecht, v. 107, p. 53–63, 1999.

VOGT, P. Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved “chromatin folding code”. **Human Genetics**, Berlin, v. 84, p. 301–336, 1990.

WANG, G.; ZHANG, X.; JIN, W. An overview of plant centromeres. **Journal of Genetics and Genomics**, Beijing, v. 36, p. 529-537, 2009.

WANG, Z. X. et al. A chromosome 5-specific repetitive DNA-sequence in rice (*Oryza sativa* L.) **Theoretical and Applied Genetics**, Berlin, v. 90, p. 907-913, 1995.

WEINER, A. M. SINEs and LINEs: the art of biting the hand that feeds you. **Current Opinion in Cell Biology**, London, v. 14, p. 343–350, 2002.

WESSLER, S. R. **Eukaryotic transposable elements**: teaching old genomes new tricks. Oxford: Oxford University, 2006. P. 138-165.

WICKER, T. et al. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum*) reveals multiple mechanisms of genome evolution. **Plant Journal**, Oxford, v. 26, p. 307–316, 2001.

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

WILLARD, H. F. Centromeres: the missing link in the development of human artificial chromosomes. **Current Opinion in Genetics and Development**, London, v. 8, p. 219–225, 1998.

WU, F. N. et al. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. **Genetics**, Austin, v. 174, p. 1407–1420, 2006.

ZABEL, P. et al. Towards the construction of artificial chromosomes for tomato. In: VAN VLOTEN-DOTING, L.; GROOT, G. S. P.; HALL, T. C. (Ed.). **Molecular form and function of the plant genome**. New York: Plenum, 1985. p. 609–624.

ZAKIAN, V. A. Telomeres: beginning to understand the end. **Science**, Washington, v. 270, p. 1601–1607, 1995.

ZHANG, H. et al. Boom-Bust Turnovers of Megabase-Sized centromeric DNA in Solanum Species: Rapid Evolution of DNA Sequences Associated with Centromeres. **Plant Cell**, Rockville, v. 26, n. 4, p. 1436–1447, 2014.

ZHANG, T. et al. The Cento satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 110, p. 4875–4883, 2013.

ZHAO, X. et al. Genome-specific repetitive sequences in the genus *Oryza*. **Theoretical and Applied Genetics**, Berlin, v. 78, p. 201-209, 1989.

ZHONG, X. et al. "FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato." **The Plant Journal**, Oxford, v. 13, n. 4, p. 507-517, 1998.

ZHU, W. et al. Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition. **BMC Genomics**, London, v. 9, p. 286, 2008.

ZUPUNSKI, V.; GUBENSEK, F.; KORDIS, D. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. **Molecular Biology and Evolution**, Chicago, v. 18, p. 1849–1863, 2001.

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

Composição e estrutura do DNA centromérico em *S. chomatophilum* Bitter e sua evolução no gênero *Solanum* L.

Oliveira, LC; Wang, K; Macas, J; Torres, GA; Jiang, J

RESUMO

O centrômero é o loco onde se organiza o cinetócoro, placa proteica à qual se ligam as fibras do fuso para promover a orientação e a segregação cromossômica. Uma histona H3 modificada (CenH3), encontrada nos centrômeros de todos os organismos já estudados, é até então a mais proeminente marca associada à função centromérica. No entanto, existem indícios de que o DNA desempenhe também um importante papel na determinação desta função. Os centrômeros são estruturas intrigantes, uma vez que sua função é altamente conservada, mas as sequências de DNA que os compõem não o são. A sequência de DNA centromérico pode variar entre espécies proximamente relacionadas, e até mesmo entre centrômeros de um mesmo complemento, como é o caso da batata (*S. tuberosum* – genoma A). Os objetivos do trabalho foram identificar as sequências centroméricas e caracterizar sua organização nos cromossomos de *S. chomatophilum* (genoma P, o mais próximo ao genoma A), bem como analisar a dinâmica evolutiva dessas sequências em representantes dos demais genomas de *Solanum* (B, E e T). Após realização das técnicas de ChIP, ChIP-seq e análises bioinformáticas, foram identificadas quatro repetições centroméricas: Sc83, Sc111, Sc92/108 e Sc92/108/111. Seus comprimentos variam de aproximadamente 600 pb a 3 kb, tamanho incomum para repetições localizadas em centrômeros. Análises de FISH mostraram que as repetições compõem juntas de 6-7 centrômeros, dos 24 existentes no complemento cromossômico. Foi verificada variação na distribuição e na abundância para todas essas repetições quando elas foram mapeadas para dois acessos diferentes da espécie. Para Sc83 existe ainda variação entre plantas, para o número de sinais de FISH observados. O centrômero 2 pode ser composto por Sc111, Sc92/108 ou Sc92/108/111, existindo portanto três diferentes haplótipos associados a ele. Todas as repetições centroméricas isoladas de *S. chomatophilum* foram encontradas também no genoma A, embora com localização diferenciada. Todas elas estão ausentes nos demais genomas, com exceção de Sc83 que foi também identificada no genoma E. Estes dados apontam para a rápida divergência dessas sequências mesmo entre espécies proximamente relacionadas. Neste trabalho foram descritos aspectos da variação centromérica não relatados na literatura, que ampliam o conceito de complexidade dos centrômeros pela demonstração de variações entre acessos, plantas e cromossomos homólogos.

Palavras-chave: Centrômero. DNA centromérico. Evolução centromérica. DNA satélite. *Solanum*.

ABSTRACT

The centromere is the locus where assembles the kinetochore, a protein plate to which the spindle fibers bind to promote chromosome orientation and segregation. A modified histone H3 (CenH3), found in the centromeres of all organisms studied, is so far the most prominent mark associated with centromeric function. However, there are indications that the DNA also plays an important role in determining this function. Centromeres are intriguing structures, since its function is highly conserved, while their DNA sequences are diverse. Centromeric DNA can differ among closely related species and even between centromeres of the same complement, as reported in potato (*S. tuberosum* – genome A). The aims of this paper were to identify the centromeric sequences and characterize their organization in the chromosomes of *S. chomatophilum* (P genome, the closest to A genome). Evolutionary dynamics of these sequences was also analysed in species representing different *Solanum* genomes (B, T e E). After performing the ChIP, ChIP-seq and bioinformatic analysis, we identified four centromeric repeats: Sc83, Sc111, SC92/SC108 and Sc92/108/111. Their length ranges from about 600 bp to 3 kb, unusual sizes for repeats located in the centromeres. FISH analyses revealed that the repeats locates at 6-7 centromeres out of the 24 in the chromosome complement. Variation in the distribution and abundance for the repeats was observed when comparing two accessions of *S. chomatophilum*. For Sc83 there is variation on number of FISH signals between plants. The centromere 2 has three different haplotypes since it may be composed by Sc111, SC92/SC108 or Sc92/108/111. All centromeric repeats isolated from *S. chomatophilum* were also found in the A genome, although in different locations. All of them are absent in other genomes, except Sc83 which was also identified in the E genome. The data point to the rapid divergence of these sequences even among closely related species. This paper brings up new aspects of the centromeric variation, which extend the concept of centromere complexity by showing variations among accessions, plants and homologous chromosomes.

Keywords: Centromere. Centromeric DNA. Centromere evolution. Satellite DNA. *Solanum*.

1 INTRODUÇÃO

O centrômero é um elemento funcional dos cromossomos eucarióticos que, por ser o loco para organização do cinetócoro, é essencial para a regularidade dos processos de divisão celular (BIRCHLER; HAN, 2009). Acredita-se que a funcionalidade centromérica seja conferida por mecanismos genéticos e epigenéticos, embora não se conheça bem a relação existente entre as sequências de DNA, os componentes proteicos e as marcas epigenéticas centroméricas (PLOHL; MEŠTROVIĆ; MRAVINAC, 2014). A presença de uma variante da histona H3 (CenH3) específica dos centrômeros de todos os eucariotos estudados (BLACK; BASSETT, 2008; MALIK; HENIKOFF, 2009) é atualmente considerada como a mais proeminente marca da função centromérica (HENIKOFF; AHMAD; MALIK, 2001).

O DNA presente nos centrômeros tem, quase sempre, natureza repetitiva, com frequente organização em grandes blocos de DNA satélite (HENIKOFF; AHMAD; MALIK, 2001; JIANG et al., 2003). Em *Arabidopsis thaliana* (L.) Heynh., por exemplo, monômeros de 178 pb se repetem milhares de vezes e formam arranjos da ordem de megabases de DNA satélite que constituem todos os centrômeros da espécie (NAGAKI et al., 2003b). Em humanos, monômeros de tamanho similar (171 pb) são também encontrados no DNA satélite de todos os centrômeros do complemento cromossômico (WILLARD, 1998).

Embora a função dos centrômeros seja altamente conservada em todos os eucariotos, as sequências de DNA que os compõem geralmente não o são (BENSASSON, 2011). Dessa forma, ainda que tenham tamanhos parecidos, não existe similaridade de sequência entre os monômeros centroméricos descritos em *Arabidopsis* e em humanos. Mediante a alta conservação da função dos centrômeros, essa falta de conservação das sequências de DNA centroméricas caracteriza um paradoxo (HENIKOFF; AHMAD; MALIK, 2001).

Um aspecto que pode ser considerado comum para a maioria das espécies já estudadas é a alta similaridade das sequências de DNA repetitivo entre os centrômeros de um mesmo complemento cromossômico, como acontece em *Oryza sativa* L. (CHENG et al., 2002a), *Oryza brachyantha* A. Chev. & Roehr. (LEE et al., 2005), *Arabidopsis thaliana* (NAGAKI et al., 2003b), humanos (WILLARD, 1998), dentre outros. Esse não é o caso de espécies de *Solanum* L. Em *S. tuberosum* L. (batata cultivada) cada um dos doze centrômeros tem uma composição genética diferente dos demais. Seis são compostos por DNA repetitivo, cinco por DNA de cópia única e o último centrômero apresenta os dois tipos de DNA em sua composição (GONG et al., 2012). Zhang et al. (2014) também descreveram para *Solanum verrucosum* Schltdl., uma espécie selvagem já proposta como ancestral da batata, tanto centrômeros constituídos por DNA de cópia única quanto por DNA satélite. Foram encontradas repetições satélite diferentes para os dois genomas e repetições em comum, porém com localização diferente, apontando para uma rápida evolução das sequências centroméricas no gênero.

Tanto *S. tuberosum* quanto *S. verrucosum* são espécies portadoras do genoma A, um dos cinco genomas (A, B, P, T e E) reconhecidos em espécies diploides do gênero *Solanum* (MATSUBAYASHI, 1991; HAMANNA; HERMSEN, 1981). A dinâmica evolutiva das sequências centroméricas dessas duas espécies foi avaliada em espécies selvagens de *Solanum* portadoras dos outros quatro genomas (GONG et al., 2012; ZHANG et al., 2014). Os resultados obtidos em ambos os trabalhos apontam para grandes diferenças na composição do DNA centromérico mesmo entre espécies proximamente relacionadas.

Diante da complexidade e rápida evolução das sequências centroméricas do genoma A de *Solanum*, os objetivos do trabalho foram identificar as sequências centroméricas e caracterizar sua organização nos cromossomos do genoma P, o mais próximo ao genoma A, bem como analisar a dinâmica evolutiva dessas sequências em representantes dos demais genomas de *Solanum*.

2 MATERIAL E MÉTODOS

Material Vegetal

Solanum chomatophilum Bitter ($2n=2x=24$, genoma P) foi utilizada para as análises de ChIP e nos estudos citogenéticos. Dois acessos foram estudados: PI365339 e PI310943, doravante chamados, respectivamente, de acessos A e B. Para os estudos evolutivos foram utilizados um clone monoploide homozigótico duplicado, DM1-3 516R44 (DM1-3, $2n=2x=24$, genoma A), como representante da batata cultivada (*S. tuberosum*), além das espécies selvagens *S. verrucosum* ($2n=2x=24$, genoma A, PI275260) e *S. jamesii* Torr. ($2n=2x=24$, genoma B, PI458424). O Microtomate (*S. lycopersicum* Lam., $2n=2x=24$, genoma T) e *S. palustre* Poepp. ex Schltdl. ($2n=2x=24$, genoma E, PI473401) foram utilizadas como representantes de espécies distantes da batata. Sementes de todas as espécies selvagens de *Solanum* foram obtidas da USDA/ARS Potato Introduction Station, Sturgeon Bay, WI.

ChIP e ChIP-seq

Um anticorpo contra a proteína CenH3 da batata (GONG et al., 2012) foi utilizado nos experimentos de imunoprecipitação para o isolamento do DNA centromérico em *S. chomatophilum*. Aproximadamente 10 gramas de tecido foliar jovem foram coletados de plantas do acesso A de *S. chomatophilum* crescidas em casa de vegetação. As folhas foram trituradas em nitrogênio líquido até tornarem-se um pó fino, que foi ressuscitado em 10 mL de NBI (Nuclei Isolation Buffer) contendo 0,1% de celulase e 0,05% de pectinase. Os núcleos foram incubados por 30 min a 37°C, filtrados e centrifugados por 10 min a 4°C. O pellet foi lavado e ressuscitado em tampão específico para a digestão pela enzima nuclease micrococcal (Sigma-Aldrich) para redução do DNA em fragmentos do tamanho de

um nucleossomo. Após duas rodadas de centrifugação a 13.000 rpm por 10 min a 4°C, a cromatina digerida foi utilizada para os experimentos da ChIP, realizada de acordo com Nagaki et al. (2003b) com pequenas modificações. Aproximadamente 30 ng do DNA proveniente da ChIP foram utilizados para o preparo da biblioteca para o sequenciamento, utilizando o protocolo de ChIP-seq fornecido pela Illumina. Este DNA foi então sequenciado utilizando o sequenciador Illumina Genome Analyzer II que gerou reads de 36 pb.

Identificação, caracterização e clonagem das repetições

Utilizando o conjunto de ferramentas de software RepeatExplorer (NOVÁK et al., 2013) foi realizado um agrupamento baseado em similaridade que permitiu a identificação das repetições centroméricas. Para a análise dos gráficos de agrupamento foi utilizado o programa SeqGrapheR (NOVAK et al., 2013). Após a identificação das repetições, primers para PCR foram desenhados para cada uma delas de acordo com as sequências consenso. Elas foram então amplificadas por PCR a partir do DNA genômico de *S. chomatophilum* extraído de tecido foliar com o Kit QIAGEN (Valencia, CA). Na reação de PCR foram realizados 34 ciclos de 30 seg. a 95°C, 30 seg. a 58° e 1 min a 72°C. Os ciclos foram precedidos por uma desnaturação inicial (5 min a 95°C) e seguidos por um período para a de extensão final (5 min a 72°). Os produtos da reação foram visualizados em eletroforese com gel de agarose 1% para confirmação do tamanho dos fragmentos e verificação do padrão de amplificação. As bandas com os fragmentos de tamanho desejado foram recortadas do gel e o DNA foi purificado com a utilização do kit Gel extraction (QIAGEN). A enzima Klenow Fragment (BioLabs) foi utilizada para adição de uma adenina ao final dos fragmentos, que foram posteriormente inseridos no vetor pCR 4-TOPO e clonados com a utilização do kit TOPO TA cloning (Invitrogen). As bactérias (*E-coli*) transformadas foram cultivadas em meio LB, os DNAs plasmidiais

foram extraídos e seu tamanho foi verificado por eletroforese em gel de agarose 1%. As amostras foram preparadas para a reação de sequenciamento utilizando Clean bead seq (Beckman coulter). No mínimo cinco clones de cada fragmento foram sequenciados na plataforma 454.

FISH e Fiber-FISH

O preparo dos cromossomos mitóticos e as análises de FISH e fiber-FISH foram realizados seguindo protocolos publicados (JACKSON et al., 1998; DONG et al., 2000; LOU et al., 2010). A obtenção das sondas para cada repetição satélite se deu pela extração de DNA plasmidial a partir do cultivo dos clones selecionados. As sondas foram então marcadas com biotina-16-UTP ou digoxigenina-11-dUTP por meio de reação de nick translation padrão. Para a identificação dos cromossomos utilizou-se rDNA 5S para cromossomo 1 e rDNA 45S para o cromossomo 2. Para os demais foram utilizados BACs desenvolvidos para os cromossomos da batata (DONG et al., 2000). Os cromossomos foram contracorados com DAPI (4',6-diamidino-2-phenilindol) em solução de Vectashield. As imagens de FISH foram obtidas no microscópio Olympus BX60, pelo software AxioVision e contrastadas no programa Adobe Photoshop CS5. As medidas citológicas da fiber-FISH foram convertidas em kilobases utilizando a razão de conversão de 3.21 kb/ μ m de Cheng et al. (2002b).

3 RESULTADOS

Identificação das repetições e confirmação citológica de sua localização centromérica

Quatro clusters de repetições foram identificados como prováveis componentes dos centrômeros em *S. chomatophilum*: Sc83, Sc92, Sc108 e Sc111, com tamanhos previstos de 1.900, 1.223, 842 e 615 pares de bases, respectivamente. Bandas com tamanhos similares aos previstos computacionalmente foram obtidas para todas as repetições selecionadas (Figura 1). A sequência de DNA das repetições pode ser verificada na Tabela Suplementar 1.

Os resultados da FISH mostraram que todas as repetições identificadas nas análises bioinformáticas (Sc83, Sc92, Sc108 e Sc111) formam grandes blocos de localização centromérica, cada uma delas presente em um, dois ou três centrômeros do complemento cromossômico de *S. chomatophilum* (Figura 2).

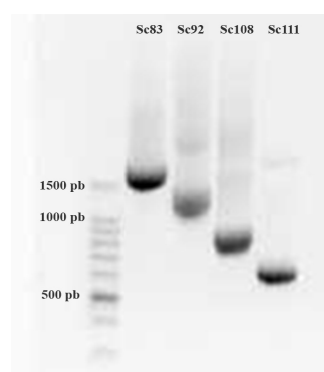


Figura 1 PCR utilizando primers desenhados para as repetições centroméricas de *S. chomatophilum*: Sc83, Sc92, Sc108 e Sc111.

Primers utilizados nas amplificações:

Sc 83F: CAATATATCACCAAATAGGCAACC

Sc83R:TCATGTTGAGCATTTGAGATGTG

Sc92F:GAAACTTCTAAACATAGTCCTCGAATC

Sc92R:ACTCCTTCGTTAGTTGAGATGTGAG

Sc108F: AATCAAAGAGCCAAATCAAAGTG

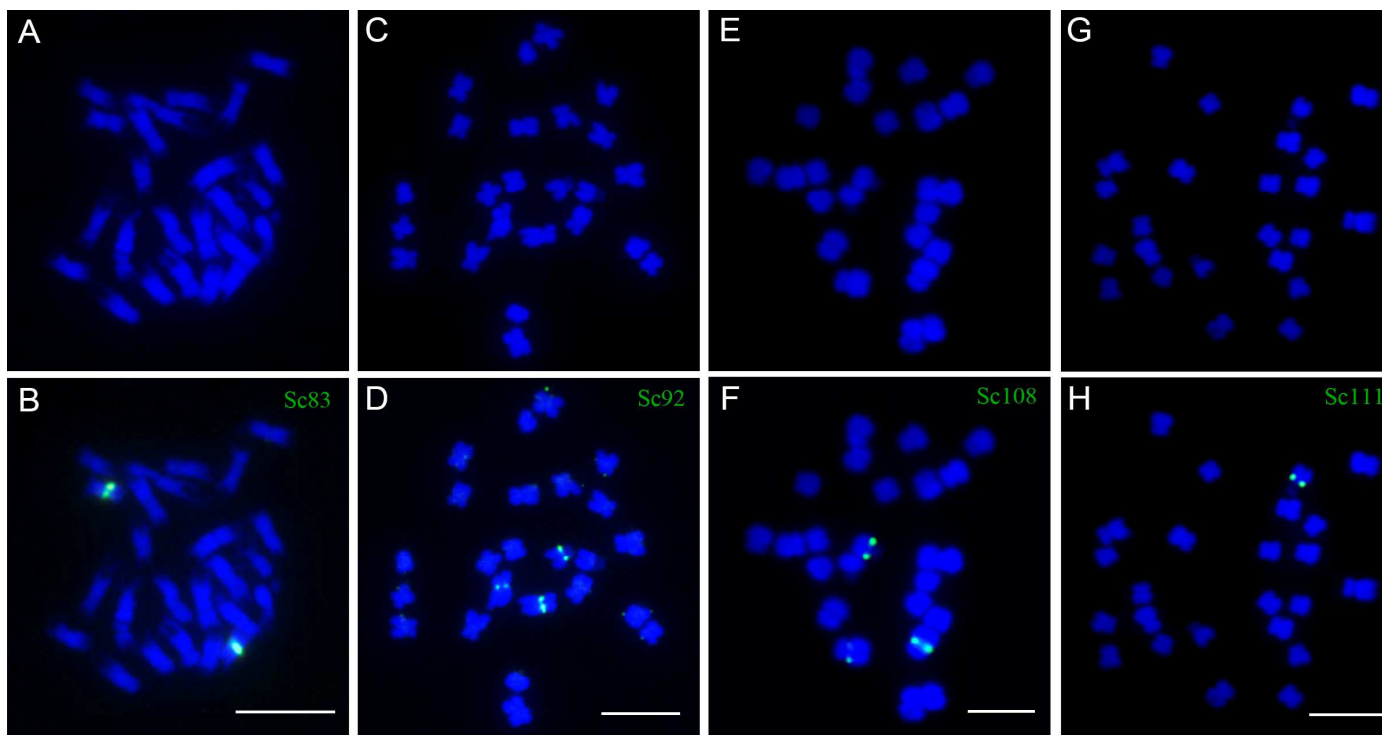


Figura 2 Mapeamento por FISH das repetições centroméricas em *S. chomatophilum*, $2n=2x=24$, acesso A. **(A, C, E e G)** Cromossomos de *S. chomatophilum* corados com DAPI **(B)** Sinais de FISH gerados pela sonda Sc83 **(D)** Sinais de FISH gerados pela sonda Sc92 **(F)** Sinais de FISH gerados pela sonda Sc108 **(H)** Sinais de FISH gerados pela sonda Sc111. Barras = 5 μ m.

Sc92 e Sc108: partes de um mesmo monômero

As sondas Sc92 e Sc108 se colocaram em todas as análises de FISH, (Figura 3, A, B, F e G). As análises iniciais mostraram que a combinação das duas sondas gerou três sinais no acesso A, sendo dois sinais maiores nos centrômeros de cada uma das cópias do cromossomo 1 - Cen1 (Figura 3A), identificadas pela sonda de rDNA 5S, e um sinal, um pouco menor, no centrômero de apenas uma das cópias do cromossomo 2 - Cen2 (Figura 3B), identificado pela sonda de rDNA 45S. Para o acesso B a configuração foi oposta. Foram observados também três sinais, porém apenas um é localizado no Cen1 (Figura 3F) e dois menores no Cen2 (Figura 3G). A colocação dos sinais da FISH derivados de Sc92 e Sc108 indica que essas repetições são na verdade partes de um mesmo monômero que foram separadas em diferentes clusters durante as análises computacionais.

Para melhor entendimento da relação existente entre essas duas sondas foi conduzida uma análise de fiber-FISH que mostrou sinais alternados de Sc92 e Sc108, sendo estes sinais sempre associados com a mesma fibra de cromatina (Figura 4 A e B). Mais de 50 fibras foram analisadas, não tendo sido observadas fibras contendo exclusivamente Sc92 ou Sc108, o que confirma que ambas não são distintas, mas sim partes de um mesmo monômero de repetição satélite, que será tratado adiante como Sc92/108. As análises de fiber-FISH permitiram também confirmar o arranjo em tandem das repetições e estimar o comprimento em pb do Cen1 do acesso A, que mediu 1.008 ± 171 kb. O comprimento do Cen2 não pôde ser determinado, uma vez que as fibras menores encontradas nas análises poderiam tanto ser provenientes do próprio Cen2 quanto de fibras quebradas do Cen1, não sendo possível sua distinção.

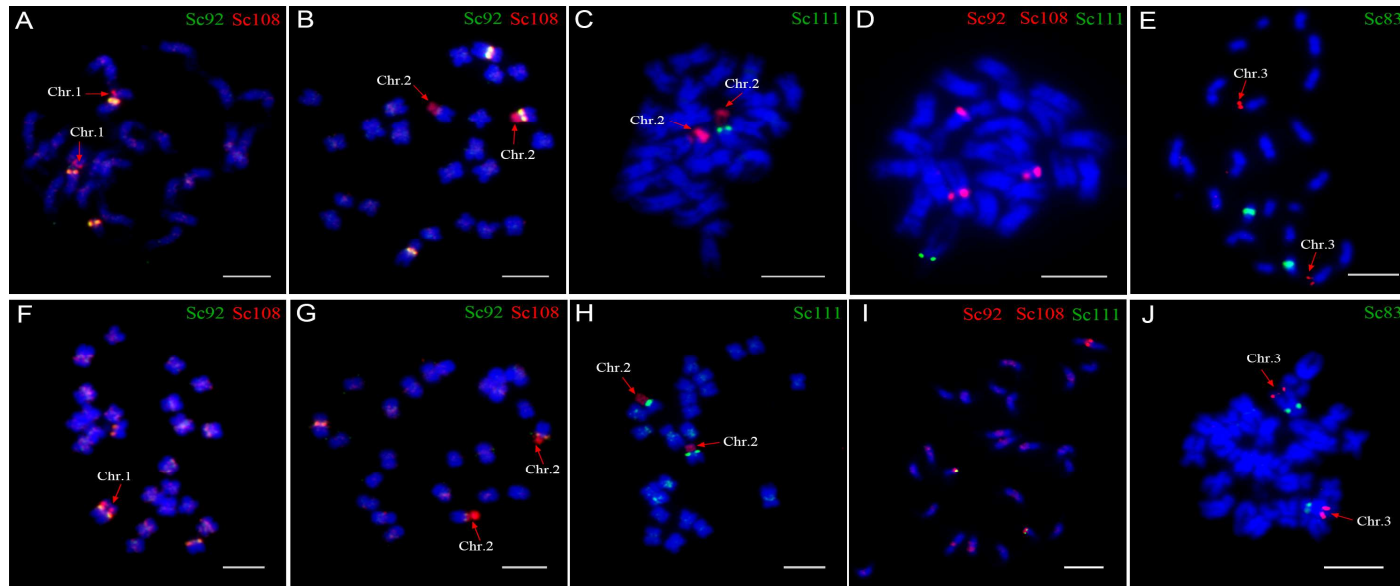


Figura 3 Mapeamento por FISH das repetições centroméricas em *S. chomatophilum*. A-E: acesso A (PI365339). F-J: acesso B (PI310943) **(A)** Repetições Sc92 e Sc108 mapeadas para o Cen 1 (identificado pela sonda rDNA 5S - vermelha) **(B)** Repetições Sc92 e Sc108 mapeadas para uma das cópias do Cen 2 (identificado pela sonda rDNA 45S - vermelha) **(C)** Repetição Sc111 mapeada para uma das cópias do Cen 2 (identificado pela sonda rDNA 45S - vermelha) **(D)** Repetições Sc92, Sc108 e Sc111. **(E)** Repetição Sc83 mapeada para cromossomo desconhecido junto com o BAC RH079E02 - vermelho, específico para o cromossomo 3 **(F)** Repetições Sc92 e Sc108 mapeadas para uma das cópias do Cen 1 (identificado pela sonda rDNA 5S - vermelha) **(G)** Repetições Sc92 e Sc108 mapeadas para o Cen 2 (identificado pela sonda rDNA 45S - vermelha) **(H)** Repetição Sc111 mapeada para o Cen 2 (identificado pela sonda rDNA 45S - vermelha) **(I)** Repetições Sc92, Sc108 e Sc111. **(J)** Repetição Sc83 mapeada para o Cen3 (identificado pela sonda com o BAC RH079E02 - vermelho). Barras = 5 μ m.

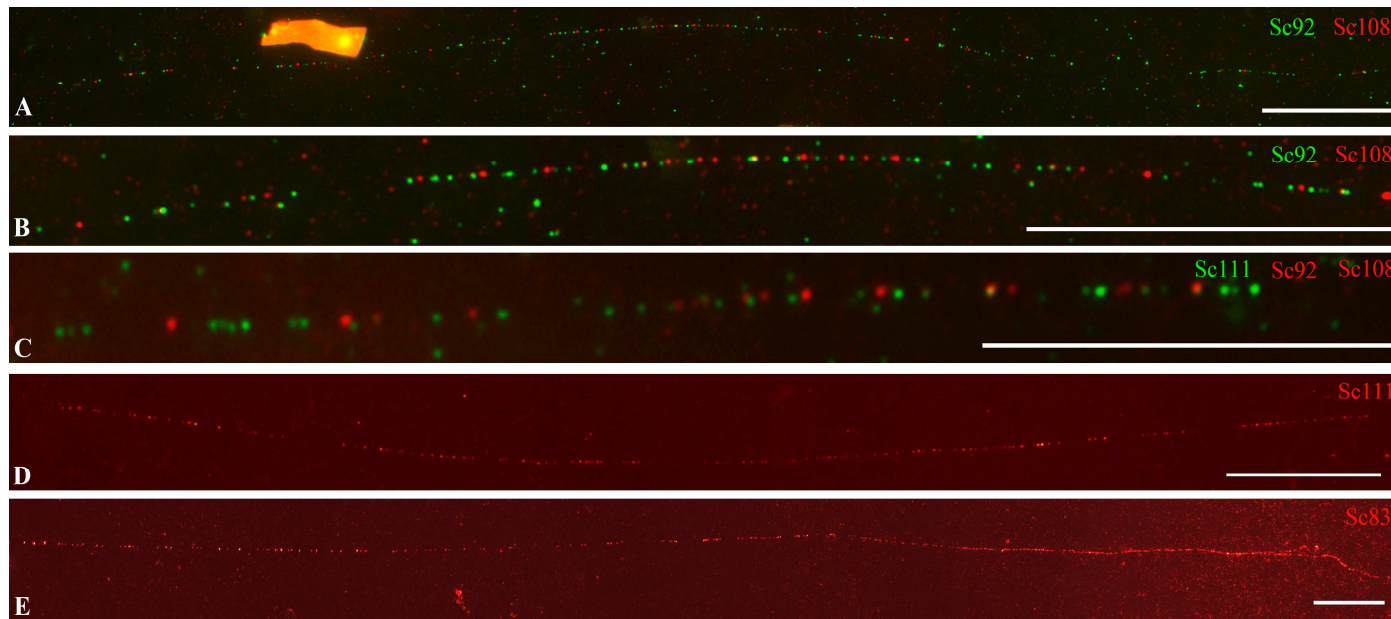


Figura 4 Análise das repetições satélite centroméricas em *S. chomatophilum* por meio de fiber-FISH. **(A)** Sinal completo de fiber-FISH derivado das sondas Sc92 e Sc108 no cromossomo 1 do acesso A. **(B)** Sinal parcial de fiber-FISH derivado das sondas Sc92 e Sc108 no cromossomo 1 do acesso A. **(C)** Sinal parcial de fiber-FISH derivado das sondas Sc92, Sc108 e Sc111 no cromossomo 2 do acesso B. **(D)** Sinal completo de fiber-FISH derivado da sonda Sc111 no cromossomo 2 do acesso A. **(E)** Sinal completo de fiber-FISH derivado da sonda Sc83 em cromossomo desconhecido do acesso A. Barras: 25 μ m

Complexa organização do Cen2

Além da repetição Sc92/108, a sonda com a repetição Sc111 também hibridizou no Cen2, identificado pela sonda de rDNA 45S, em ambos os acessos (Figura 3, C e H). Análises simultâneas das repetições Sc92/108 e Sc111 nos dois acessos mostraram uma organização complexa do Cen2 em *S. chomatophilum*, não relatada na literatura.

No acesso A, apenas um sinal de Sc111 foi detectado e ele não colocalizou com Sc92/108, sendo uma das cópias do Cen2 composta exclusivamente por Sc111, enquanto a outra é composta somente por Sc92/108 (Figura 3 B-D, Figura 5). Análises de fiber-FISH mostraram que o Cen2, composto por Sc111, mede 683 ± 77 kb (Figura 4D). Já no acesso B, Sc111 hibridizou em ambos Cen2 (Figura 3H). Os sinais gerados por essa sonda colocalizaram com os sinais de Sc92/108 (Figura 3I). As duas sondas, quando aplicadas em análises de fiber-FISH, foram encontradas associadas às mesmas fibras de cromatina, estando os sinais de Sc111 entremeados aos sinais gerados por Sc92/108 (Figura 4C, Figura 5). O resultado mostra que neste acesso (B) Sc111 compõe o mesmo monômero que Sc92/108, sendo este complexo monômero chamado adiante de Sc92/108/111.

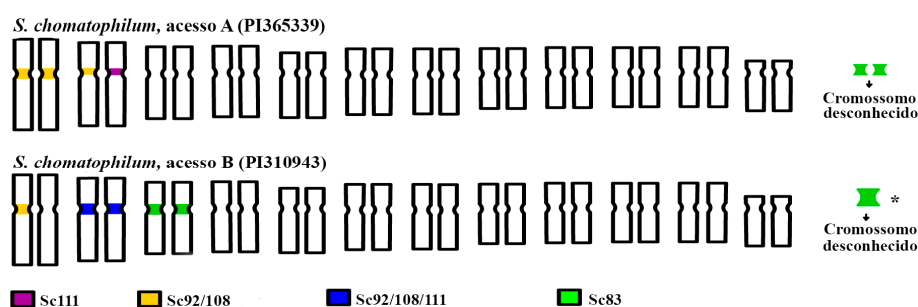


Figura 5 Esquema ilustrando a distribuição das repetições centroméricas Sc111, Sc92/108, Sc92/108/111 e Sc83 nos cromossomos dos dois acessos de *Solanum*. A morfologia dos cromossomos utilizada foi inspirada em um idiograma dos cromossomos da batata (TORRES et al., 2011) e não retrata portanto a real morfologia da espécie em questão. * Sinal detectado em 25 das 80 plantas analisadas para este acesso.

Variação na abundância e localização de Sc83

As análises iniciais de FISH para Sc83 revelaram sua localização em dois centrômeros em ambos os acessos (Figura 3 E e J). No acesso B a repetição está associada aos centrômeros de ambos os cromossomos 3 - Cen3 (Figura 3J), identificados pela sonda com o BAC RH079E02. O Cen3 para este acesso mediu de acordo com as análises de fiber-FISH 1.485 ± 172 kb (Figura 4E). Já no acesso A, a hibridização simultânea de Sc83 e do BAC RH079E02 mostrou que Sc83 não está localizada no Cen3.

Análises posteriores revelaram ainda a existência de variação no número de sinais de Sc83 para o acesso B (Figura 6). Além da configuração já relatada, com dois sinais centroméricos de intensidade média no Cen3 (Figura 3J e 6 A e B), foram encontradas plantas contendo, além destes, um sinal grande e muito intenso no centrômero de um cromossomo não identificado (Figura 6 C e D). A variabilidade de número de centrômeros contendo Sc83 foi verificada com a análise de 80 plantas individuais, das quais 55 apresentaram dois sinais (como visto nas figuras 6 A e B) e 25 apresentaram três sinais (como visto nas figuras 6 C e D).

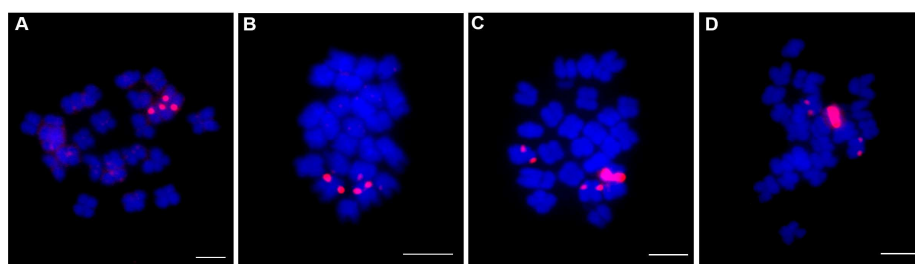


Figura 6 Variação no número de sinais centroméricos Sc83 em plantas do acesso B. **(A)** Metáfase da planta 06 com dois sinais de tamanho médio **(B)** Metáfase da planta 16 com dois sinais de tamanho médio **(C)** Metáfase da planta 10 com um sinal grande e dois sinais médios **(D)** Metáfase da planta 12 com um sinal grande e dois sinais médios. Barras: 5 μ m.

Distribuição das repetições centroméricas identificadas em *S. chomatophilum* no genoma de outras espécies de *Solanum*

As repetições Sc92/108 e Sc111, presentes nos Cen1 e Cen2 de *S. chomatophilum*, foram ambas encontradas nos Cen2 (Figura 7 A e B) e Cen10 (Figura 7 C e D) de *S. verrucosum*, o que indica que o monômero Sc92/108/111 é a repetição que compõe estes dois centrômeros.

Em DM1-3, a distribuição de Sc92/108 foi independente de Sc111. Esta última foi encontrada nos Cen3 (Figura 7E) e Cen9 (Figura 7F), enquanto que Sc92/108 apresentou distribuição mais ampla e dispersa nas regiões centroméricas e/ou pericentroméricas de todos os cromossomos (Figura 7 G e H).

Sc83 foi encontrada nos centrômeros de aproximadamente oito cromossomos não identificados de *S. verrucosum* (Figura 7 I e J) e em aproximadamente seis cromossomos também não identificados de *S. palustre* (Figura 7 K e L), mas os sinais gerados pela sonda foram menos intensos que em *S. chomatophilum*.

Nas demais espécies estudadas (*S. jamesii* e tomate) não foram observados sinais de FISH para todas as repetições analisadas. Para a certificação de que as repetições eram de fato ausentes nos cromossomos dessas espécies a FISH foi sempre realizada junto com uma sonda de rDNA 45S ou 5S, marcada com uma segunda cor, como controle positivo.

Um esquema da distribuição de todas as sondas identificadas em *S. chomatophilum* nos dois acessos da espécie e nos demais materiais onde as repetições foram encontradas e localizadas nos cromossomos pode ser visto na tabela 1.

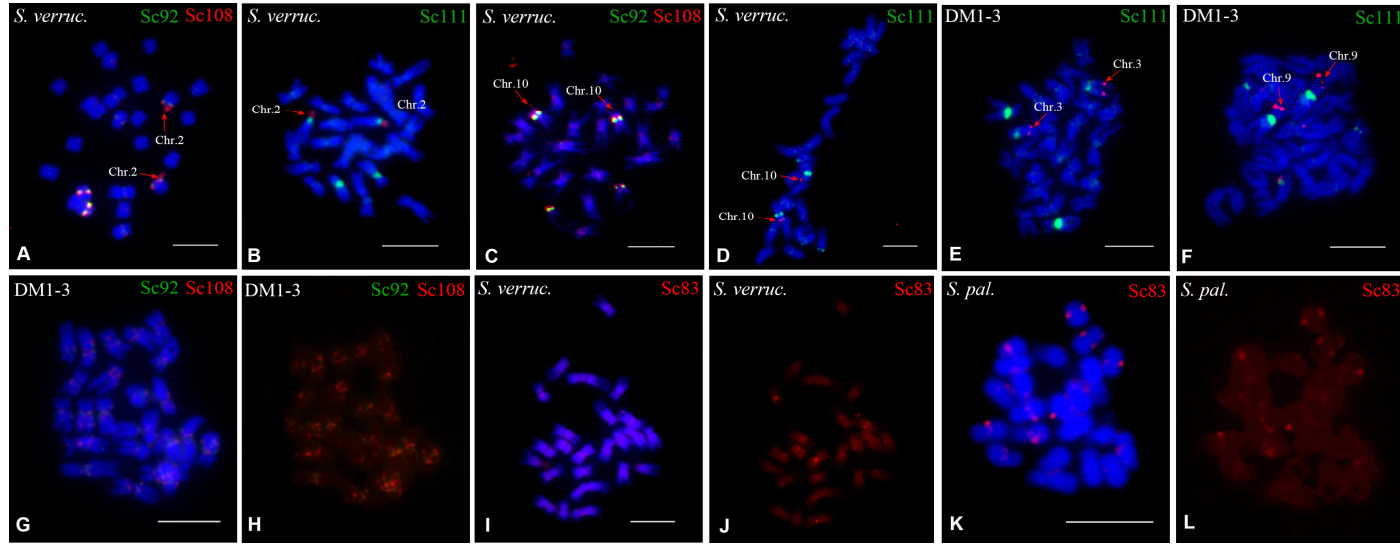


Figura 7 Mapeamento por FISH das repetições centroméricas isoladas de *S. chomatophilum* em diferentes espécies de *Solanum*. **(A)** Repetições Sc92 e Sc108 mapeadas em *S. verrucosum* para o Cen2 (identificado pela sonda rDNA 45S) **(B)** Repetição Sc111 mapeada em *S. verrucosum* para o Cen2 (identificado pela sonda rDNA 45S) **(C)** Repetições Sc92 e Sc108 mapeadas em *S. verrucosum* para o Cen10 (identificado pela sonda com o BAC RH178K07) **(D)** Repetição Sc111 mapeada em *S. verrucosum* para o Cen10 (identificado pela sonda com o BAC RH178K07) **(E)** Repetição Sc111 mapeada em DM1-3 para o Cen3 (identificado pela sonda com o BAC RH079E02) **(F)** Repetição Sc111 mapeada em DM1-3 para o Cen9 (identificado pela sonda com o BAC RH135I22) **(G)** Repetições Sc92 e Sc108 nos cromossomos de DM1-3 **(H)** Sinais da FISH das repetições Sc92 e Sc108 separados digitalmente da figura G. **(I)** Repetição Sc83 nos cromossomos de *S. verrucosum* **(J)** Sinais da FISH da repetição Sc83 separados digitalmente da figura I **(K)** Repetição Sc83 nos cromossomos de *S. palustre* **(L)** Sinais da FISH da repetição Sc83 separados digitalmente da figura K. Barras: 5 μm.

Tabela 1 Resumo das análises de FISH das repetições centroméricas isoladas de *S. chomatophilum* em espécies de *Solanum*.
* Sinal detectado em 25 das 80 plantas analisadas para este acesso.

	<i>S. chomat.</i> Acesso A Genoma P	<i>S. chomat.</i> Acesso B Genoma P	DM1-3 Genoma A	<i>S. verrucosum</i> Genoma A	<i>S. jamesii</i> Genome B	<i>S. palustre</i> Genome E	Tomate Genome T
Sc111	Específico para uma das cópias do Cen2	Ausente	Específico para Cen3 e Cen9	Ausente	Ausente	Ausente	Ausente
Sc92/108	Específico para o Cen1 e para uma das cópias do Cen2	Específico para uma das cópias do Cen1	Sinais dispersos em todos os cromossomos (mais abundantes nos centrômeros)	Ausente	Ausente	Ausente	Ausente
Sc92/108/111	Ausente	Específico para o Cen2	Ausente	Específico para Cen2 e Cen10	Ausente	Ausente	Ausente
Sc83	Específico para um centrômero não identificado	Específico para o Cen3 e para uma das cópias* de um centrômero não identificado	Ausente	Sinais fracos e dispersos em todos os cromossomos (não específicos para os centrômeros)	Ausente	Sinais centroméricos (fracos) específicos para ~6 centrômeros	Ausente

4 DISCUSSÃO

DNA de cópia única como provável componente centromérico em *S. chomatophilum*

As análises de ChIP, seguidas por ChIP-seq e FISH permitiram descrever a constituição de 6 centrômeros do acesso A e de 6-7 centrômeros do acesso B, dentre os 24 existentes no complemento cromossômico dos dois acessos (Figura 5). A não detecção de repetições em tandem nos demais centrômeros é um indicativo de que eles sejam compostos por retrotransposons, que se acumulam com bastante frequência na região centromérica (PIMPINELLI et al. 1995; COPENHAVER et al. 1999; SCHUELER et al. 2001; CHENG et al. 2002) ou por DNA de cópia única. Centrômeros contendo DNA de cópia única já foram relatados em cinco pares de cromossomos da batata (GONG et al., 2012) e em pelo menos um par em *S. verrucosum* (ZHANG et al., 2014).

Acredita-se que os centrômeros compostos por DNA de cópia única sejam centrômeros de origem mais recente. Eles são conhecidos como centrômeros evolucionariamente novos, centrômeros reposicionados ou apenas como novos centrômeros; são totalmente funcionais e surgem em locos cromossômicos não previamente associados com proteínas do cinetócoro (PLOHL, MEŠTROVIĆ; MRAVINAC, 2014). Por outro lado, centrômeros maduros são tipicamente compostos por longos arranjos de repetições satélite (WANG et al., 2014).

Diferentes haplótipos associados aos mesmos centrômeros

De acordo com Wang et al. (2014) um novo centrômero, composto por DNA de cópia única, adquire longos arranjos de DNA satélite durante sua evolução. A frequência deste novo haplótipo, agora composto por DNA

repetitivo, irá eventualmente aumentar na população até substituir o haplótipo original. Durante esse processo de substituição, seria possível encontrar diferentes haplótipos associados com o mesmo centrômero nas populações, ou seja, haplótipos contendo principalmente DNA de cópia única e outros com diferentes blocos de repetições satélite.

Neste trabalho, das 80 plantas avaliadas para Sc83 no acesso B, além dos dois sinais no Cen3, 25 plantas apresentaram um terceiro sinal ainda maior. O centrômero portador do terceiro sinal em algumas plantas é provavelmente composto por DNA de cópia única ou por retrotransposons naquelas plantas com apenas dois sinais. Desta forma, são reconhecidos dois haplótipos distintos (DNA satélite e provavelmente cópia única ou retrotransposons) para o cromossomo ainda desconhecido. De acordo com Gong et al. (2012) os retrotransposons são aparentemente a maior fonte para a origem de novas repetições satélite na batata. Quatro repetições originadas de sequências relacionadas a retrotransposons foram descritas pelos autores para *S. tuberosum*. A ausência de repetições em tandem no cromossomo desconhecido de *S. chomatophilum* indica que este centrômero esteja em franco processo de evolução, pela aquisição de DNA satélite e migração de uma configuração composta por DNA de cópia única e/ou retrotransposons para outra composta por longos arranjos de DNA repetitivo.

Outro dado interessante apresentado no presente trabalho é a existência de três haplótipos diferentes associados ao Cen2, uma vez que este centrômero pode ser composto pelos monômeros Sc111, Sc92/108 ou Sc92/108/111. Uma hipótese antiga, de que algumas repetições sejam favorecidas por apresentarem um comprimento ideal que possa favorecer, por melhor conformação estrutural, a estabilização dos nucleossomos centroméricos, foi recentemente provada por Zhang et al. (2014) em arroz e por Hasson et al. (2013) em humanos. Esta seria a justificativa para que o tamanho mais comum dos monômeros centroméricos já

descritos seja de 150 a 180 pb, tamanho ideal para que o monômero dê uma volta completa em torno de um único nucleossomo. As repetições aqui descritas não têm o tamanho ideal para conformação nucleossômica, uma vez que a menor delas, Sc111, tem mais de 800 pb e a maior, Sc 92/108/111, mais de 3 kb. Casos semelhantes foram descritos para repetições das outras duas espécies de batata já estudadas: *S. tuberosum*, com monômeros de 979 bp a 5 kb (GONG et al., 2012) e *S. verrucosum*, com monômeros maiores que 390 bp (ZHANG et al., 2014).

Zhang et al. (2014) consideram que diferentes repetições podem inicialmente surgir no complemento de uma espécie e sofrer ciclos de expansão e retração e que ao longo deste processo de evolução uma repetição estruturalmente favorável à uma melhor conformação nos nucleossomos CenH3 acabará por surgir. Tal repetição será seletivamente mantida em seu centrômero de origem e eventualmente, por mecanismos não muito bem conhecidos, invadir outros centrômeros até se fixar na população.

Sob este ponto de vista surge a hipótese de que os diferentes monômeros associados ao Cen2 estejam passando por este processo de seleção. Além de presente no Cen2, o monômero Sc92/108 está também presente no Cen1, enquanto que Sc111 e Sc92/108/111 encontram-se associados apenas ao Cen2. A presença de Sc92/108 no Cen1 é um indicativo que este monômero possa oferecer alguma vantagem aos centrômeros de *S. chomatophilum* e por isso esteja presente em mais de um centrômero do complemento cromossômico. Por esta razão Sc92/108 é também um monômero candidato a gerar grandes blocos de DNA satélite em todos os demais centrômeros do complemento. Outro fato, que também dá apoio a esta hipótese, é a existência de Sc92/108, de forma um pouco dispersa, em quase todos os centrômeros e/ou pericentrômeros de ambos os acessos de *S. chomatophilum* (Figura 3 A, B, F e G) e da mesma forma em todos os cromossomos de DM1-3 (Figura 7, G e H).

Aspectos evolutivos do DNA centromérico em *Solanum*

O mapeamento das repetições centroméricas de *S. chomatophilum* (genoma P) foi realizado em cromossomos metafásicos de diversas espécies de *Solanum*. Dentre os materiais investigados neste trabalho, de acordo com a árvore filogenética utilizada por Gong et al. (2012), construída com base nos trabalhos de análise de DNA plastidial de Spooner et al. (1993), Spooner e Castillo (1997) e Castillo e Spooner (1997), os mais próximos de *S. chomatophilum* são aqueles portadores do genoma A (*S. verrucosum* e DM1-3), seguidos da espécie com genoma B (*S. jamesii*). A espécie mais distante de *S. chomatophilum* é *S. palustre* (genoma E), seguida por *S. lycopersicum*, o tomate (genoma T).

A repetição Sc111, associada a uma das cópias do Cen2 no acesso A, está presente nas duas cópias do Cen3 e Cen9 em DM1-3 e não foi encontrada nas demais espécies analisadas. Aparentemente Sc111 surgiu no genoma de *S. chomatophilum* e em DM1-3 de forma independente, uma vez que essas repetições hibridizaram em diferentes centrômeros nas duas espécies. Outra possibilidade é de que rearranjos cromossômicos em uma ou nas duas espécies em questão tenham resultado no reposicionamento do próprio centrômero ou dos BACs utilizados para a identificação dos cromossomos.

Casos semelhantes a este foram relatados para as repetições St18 e St24, encontradas, respectivamente, no Cen1 e Cen9 de DM1-3, mas em diferentes centrômeros em *S. verrucosum* (Zhang et al., 2014). Sc111 é ainda uma repetição que em DM1-3 pode ter apresentado vantagens adaptativas e por esta razão teve sua distribuição ampliada para quatro centrômeros do complemento, enquanto que em *S. chomatophilum* ocupa apenas um.

Esta última hipótese é também aplicável à repetição Sc92/108/111, presente nas duas cópias do Cen2 no acesso B e que em *S. verrucosum* teve sua distribuição ampliada para além do Cen2, ocupando também o Cen10. Outro caso semelhante é o da repetição Sc83 que devido à variação apresentada está presente em dois ou três centrômeros de *S.*

chomatophilum e foi observada em cerca de oito centrômeros em *S. verrucosum*. A presença de ambas as repetições em um número maior de cromossomos em *S. verrucosum* é um bom indicativo de que nesta espécie elas tenham apresentado vantagens sobre as demais repetições e por isso tenham se espalhado.

Sc83 foi também encontrada em cerca de 6 centrômeros de *S. palustre*, que dentre as espécies estudadas é a mais distante de *S. chomatophilum*. Uma vez que esta repetição não foi encontrada em espécies menos distantes, como *S. jamesii* e *S. lycopersicum*, é provável que sua origem em *S. palustre* tenha ocorrido independentemente da origem em *S. chomatophilum*.

A repetição Sc92/108, encontrada em ambos Cen1 e em um Cen2 no acesso A e também presente em uma cópia do Cen1 do acesso B, foi encontrada apenas em DM1-3, onde se localiza nas regiões centromérica e/ou pericentromérica de todos os cromossomos do complemento.

De forma geral, as repetições isoladas de *S. chomatophilum* (genoma P), foram encontradas apenas nas espécies portadoras do genoma A, seu genoma mais próximo, indicando que elas foram recentemente amplificadas no gênero.

O presente trabalho mostra que a complexidade das sequências centroméricas em *Solanum* é ainda maior do que já demonstrado em comparação com outros gêneros, uma vez que a variação não se dá apenas em centrômeros de um mesmo complemento, mas se estende a grandes diferenças também entre centrômeros homólogos, entre acessos de uma mesma espécie e entre plantas de um mesmo acesso. Nossos dados corroboram várias hipóteses previamente propostas, como a da rápida evolução do DNA centromérico, do surgimento de centrômeros compostos por DNA repetitivo a partir daqueles compostos por DNA de cópia única e de que uma repetição satélite pode surgir inicialmente em um único centrômero e então invadir os centrômeros de todo o complemento. Além disso, os dados aqui relatados lançam luz sobre a origem e organização dos centrômeros de forma geral, especialmente em relação ao DNA repetitivo que os compõem.

REFERÊNCIAS

BENSASSON, D. (2011). Evidence for a high mutation rate at rapidly evolving yeast centromeres. *BMC Evol. Biol.* 11: 211.

BIRCHLER, J.A., HAN, F., 2009. Maize centromeres: structure, function, epigenetics. *Annu. Rev. Genet.* 21, 287e303.

BLACK BE, BASSETT EA (2008) The histone variant CENP-A and centromere specification. *Curr Opin Cell Biol* 20:91–100

CHENG , Z.K. , DONG , F.G. , LANGDON , T. , OUYANG , S. , BUELL , C.B. , GU , M.H. ,BLATTNER , F.R. , AND JIANG , J.M. (2002a). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14, 1691 – 1704.

CHENG, Z.K., BUELL, C.R., WING, R.A., and JIANG, J. (2002b). Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Res.* 10: 379–387.

COPENHAVER GP, NICKEL K, KUROMORI T, BENITO MI, KAUL S, LIN X, BEVAN M, MURPHY G, HARRIS B, PARNELL LD, MCCOMBIE WR, MARTIENSSEN RA, MARRA M, PREUSS D (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286:2468–2474.

DONG, F., SONG, J., NAESS, S.K., HELGESON, J.P., GEBHARDT, C., and JIANG, J. (2000). Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor. Appl. Genet.* 101: 1001–1007.

GONG, Z.Y., Wu, Y.F., KOBLÍZKOVÁ, A., TORRES, G.A., WANG, K., IOVENE, M., NEUMANN, P., ZHANG, W.L., NOVÁK, P., BUELL, C.R., MACAS, J., and JIANG, J.M. (2012). Repeatless and repeat-based centromeres in

potato: implications for centromere evolution. *Plant Cell* 24: 3559–3574.

HASSON D, et al. (2013) The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat Struct Mol Biol* 20(6):687–695.

HENIKOFF, S., AHMAD, K., and MALIK, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293: 1098–1102.

JACKSON, S.A., WANG, M.L., GOODMAN, H.M., and JIANG, J.M. (1998). Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. *Genome* 41: 566–572.

JIANG, J.M., BIRCHLER, J.A., PARROTT, W.A., and DAWE, R.K. (2003). A molecular view of plant centromeres. *Trends Plant Sci.* 8: 570–575.

LEE , H.R. , ZHANG , W.L. , LANGDON , T. , JIN , W.W. , YAN , H.H. , CHENG , Z.K. , JIANG, J.M. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc Nat Acad Sci U S A* 102, 11793 – 11798.

LOU, Q.F., IOVENE, M., SPOONER, D.M., BUELL, C.R., JIANG, J.M. (2010). Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. *Chromosoma* 119: 435–442.

MALIK HS, HENIKOFF S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138:1067–1082

MATSUBAYASHI, M. 1991. Phylogenetic relationships in the potato and its related species. p. 93–118. In T. Tsuchiya and P.K. Gupta (ed.) *Chromosome engineering in plants: Genetics, breeding, evolution*, part B. Elsevier Science BV, Amsterdam.

NOVÁK, P., NEUMANN, P., PECH, J., STEINHAIŠL, J., and MACAS, J. (2013). RepeatExplorer: a Galaxy-based web server for genomewide characterization of eukaryotic repetitive elements from nextgeneration sequence

reads. **Bioinformatics**, Oxford, v. 29, p. 792–793, 2013.

PIMPINELLI S, BERLOCO M, FANTI L, DIMITRI P, BONACCORSI S, MARCHETTI E, CAIZZI R, CAGGESE C, GATTI M (1995) Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proc Natl Acad Sci U S A* 92:3804–3808

PLOHL, M. MEŠTROVIĆ, N. MRAVINAC, B. Centromere identity from the DNA point of view. *Chromosoma*, 123 (2014), pp. 313–325.

RAMANNA, M. S.; HERMSEN, J. G. T. Structural hybridity in the series *Etuberosa* of the genus *Solanum* and its bearing on crossability. **Euphytica**, Wageningen, v. 30, p. 15-31, 1981.

SCHUELER MG, HIGGINS AW, RUDD MK, GUSTASHAW K, WILLARD HF (2001) Genomic and genetic definition of a functional human centromere. *Science* 294:109–115

WANG, L.S., ZENG, Z.X., ZHANG, W.L., and JIANG, J.M. (2014). Three potato centromeres are associated with distinct haplotypes with or without megabase-sized satellite repeat arrays. *Genetics* 196: 397–401.

WILLARD, H.F. (1998). Centromeres: the missing link in the development of human artificial chromosomes. *Curr Opin Genet Dev* 8, 219 – 225.

ZHANG, H.; KOBLÍZKOVÁ, A.; WANG, K.; GONG, Z.; OLIVEIRA, L.; TORRES, G.A.; Wu, Y.; Zhang, W.; Novák, P.; Buell, R.; Macas, J.; Jiang, J. (2014) Boom-Bust Turnovers of Megabase-Sized centromeric DNA in *Solanum* Species: Rapid Evolution of DNA Sequences Associated with Centromeres. *Plant Cell*. 26(4):1436-1447.

DADOS SUPLEMENTARES

Tabela Suplementar 1 Sequência de nucleotídeos das repetições centroméricas de *Solanum chomatophilum* obtida a partir de sequenciamento na plataforma 454.

Repetição	Sequência de nucleotídeos
Sc83	aataccctcactaaggtagctcctgcagggtttaacgaattcgccctcaatatcaccaaaatagg caaccacaaattccctatgaagggttcatcacatgattcatgagacgcataaaagtaccaggagcatta atggcatcatcaccaaccattcatatagaccaaacttggtcttgaatgtgatttccattcatcaccggggtt catccgaatttgatgataaccactcctaagatcaatttcgaaaacacacaagaaccattcaactcatcaa gcatatcatctaaccctaggaataggatgatgatactttaccgtaattttgtagggcttggcagtcagcac acatgcgccatgttccatcttcttggaaaccaatagaaccggcacggcacatggactcatgctctcttc acgcaaccttggtaggagttcctcaacttatctttagctccttgggtgtccccggattgcttataaagc cggcttgggcaattgtgatcctggcacaaaatcaattgatgctcaattcactgaagttggggcgatttc actggaagctccttgggaagacatcctcataatcattactgcaaaagaaaggaatagaattgagaag agaagaaatgagacttggtagttttagcacatgatccctatgtgtgataaggatcaaggaaacttccggac tcaatttctccttatctttagaattagccaacaacacttgggcctttccttcaacttctcgcctcctcct cttcttggctcctggcctccacgtgttctcttcacacttcttcttccctcgattcactcatccttatcattga ctttgagagggggtcatagggatgaagaacattctatgatcattcaacacaagtgaatacttgggtcct cccatcatgttgggtgaacgatcatgttccaagggtctcccgagtagcacatgacaagcttgattggga ccacatcatagaacctcatcttgatacttccgatcttgaactgatcattactgtcgatgcacctcaa ttctccacactcaccaaccattgaagttgtaagggttagcatgcttagtgtaggaagtttaagaattcc actaatgcagtactagcaacatttgcaacttctctatcaatgattaaggaacacacattatccttgatga agaagcacttggagtgaataaactctcccttgactcggatcatttaaggccttactaaccattgctcgtc tcaccacatatagaggagcgacactctctccttcttgacgcaaaggctgggtcttctccatcttcaaac atcctcttcttgaggatcttctcacaaccttcttctgctccttcttcaactaagatatattcccctccttgaa gtagggaatcagtcgattgggacactcacgcattatataggatcccctacattgtgacattgaattttt gtcttgcattaacacggcttccatgctacttttactagcatcacatgccactcaaatattttatcaaaatc aggtaattgcaacagaggagcggaactaagcatagattttaagcctcaaaagcttggcttctcctc cccacttaaaaaggtttatctttaggattacctcagtcagggggctgctatggtactaaaaccttgaca aatcgtctataaaaactagccaacccatgaaagcttcaacatctttaggatttcggaattggccaattc acccctcttcaactcacaacaaacctaagaagacaacttcatcaacataaatgaacacttaacaagatt agcataaagttgtcctcctaagcacatcaaacacacatctcaaatgctcaacatgaaagggcgaattc gcggccgctaattcatcgcc

Sc92	gcgatacctcctagggactagctctgcagggttaaacgaattcgcccttgaacttctaacatagtcctc gaatcttgaggacataaccaaaacatggagagtgagcaatccaagcaactccaaagcacaatctcc tcctagttaccacaacctacactggtagtaaaatagaagaaaggggtagtacaacacagctactaagta tgaggatatgcaaaacatgctaaaaaggacatttatttgaagctatgctttcatgccatttgaataaaatc tcataaacatacaataataggcaccttacaagtcacaaccaacgatacaaaatcatatgcttcacaatac ctcataaacatacaataccatttacctttaccttctgcttggctctccaccctaagtaacccttaagcaat acctttgtgcaatgtatgggcaacgtccataccatcacccacactaaggcaccaccttaaggctacaca aagtgcacttaccatgcatacctcaaccttcaccaattaatcacataagacataaagcacataagccaac aagtcacttcataaaactagaacactcatcaaatgtcatccaagactcacctaagtaagacatgggtgag gcagccatactactctccataccacaccaagtgatccttaaggctacacaagacaagctccatcca catctcaacataaaatcgtcattcaacttaatgcatttaaacacttcattcattaatgacataagatcatcac atacattaattcaacataagcattacattcaattcattatgactcatttcacattagctatgctcaaggacc aacactttcattaattagccaataagactcattcattcattactactaagacacaccaaggatctccctt aaatgtctacttgtgcaatgtctagatgatgtcccattccaccacctatcctaagtagtacacccttaagag accctagttcattcattcattaatgggggggagagccttaaccgacatagaccatggagcttgacatgg aatcccgtaataacccttaccggatgagggatcccacttgcctaaggtagggccattcttttagccttta cgtgtagagaccactagctatgtctatgtgggcacatagttaagggatagaaggattgcttctaagtac ccacatctcaactaacgacggggcacccatcttaagggttgctactaaatactcacatctcaactaacga aggagtaaggcggaattcgcgcc
Sc108	aagggactagtcctgcagggttaaacgaattcgcccttcattcattaaggagctctggattcaagatg agaaaacctttcatctaggaaccattcattaaaggagtacttatgagactatccttacaatagaccaac tttcattcattagcattaactccattctttcattcattcatacattcttcattcattctcactaggtgtgagag taggtaaacatagctctatctaacctcaccttcattaaagtcattcaagaagacaacaaccaagcataac acataaagtctctacacaagggtcacctttcatgctcttagaccataacaaattacatcatatcacattacctt cacataatcaacattactcaaaacaacaatcatccaactttgcctacaaggccatgatcacacatacttat aaagtaaacacctccccactttcaagggatcaaaccttacaattcaacatctacaacataatataaaat actagaatcaataatctattcacaactcatcatcataaaccttaggtcacaattgcccttcaaggcca acatacaatgtaacacaattgaccatcaattcctcataattcaatagaaacataacataatcaatagccc ataacaatctacacattaatatcacattattagctcataatcaatagaccactttaggcccaaaaattggat ttcaagatatgcaagggttcatggaagttttattatcttaaaacctcaatattaacataaatcattatttgaa atactttcatgcaagaaccaagattagaatgaaatttgggttttttagtgttttgaaaaacattgaaatcac tttgatttggctctttgattaaggcggaattcgcgccgctaaattcaattcgcccta
Sc111	Ctgtgggaaggcgatgcgggccttcgctattcagctggcgaaagggggatgtgctgcaaggcgataa gttgggtaacgccagggtttccagtcgcaggttgtaaaacgacggccagtggaattgtaacgactca ctatagggcgaattgaatttagcggccgcaattcgcccttaattgtgtaatgagcgtgatatgctatgga gaagtaagttcctcaaagaaatgagaaatggtaaaagtttgcattgagtggttagatgttattcctact tccttgaattgcatgaaagtcttggtaatgcttttatgtggagttgatgttctccttgatgtgctgtagt agtgatattacatgcttgttgggcttctagtcttagggcttactctaaaatgtttgaagtgtcattcgagg acgaatgttccaagtgggggatattgtaacaccctcaaaatgaaataggataaagtagagcctcacat gggttttgcattaataactgtttaaataatgaagaatagccttttagtcagtgttatgggttatgctcac caacttgaggagacgaagccttaggaagatgaatagagatatgaagagggttaggcctgatgaacaaa accaatctagggtctaagaagaggttctataaccaagattctcaatgggtcaacaatgatagggtgtctaac cctaaacctcaaagagggaatggaggtggttctacttttgagagacctaggtgtgctactgttgaaagc aacattaaggcggaattcggttaaacctgcaggactagtccttttagtgag

ARTIGO 2

Retrotransposons centroméricos e pericentroméricos em *Solanum* L.

Oliveira, LC; Macas, J; Jiang, J; Torres, GA

RESUMO

Retrotransposons LTR são componentes abundantes que contribuem para o alto dinamismo dos genomas devido à capacidade de se moverem, modificando a organização genômica. O sequenciamento do genoma da batata (*Solanum tuberosum*) revelou que pelo menos 62,2% de seu genoma é constituído por DNA repetitivo, dos quais 29,4% são retrotransposons LTR. Utilizando estes dados de sequenciamento, foram identificados três retrotransposons LTR (St1, St6 e St73) da superfamília Ty-3 Gypsy. A distribuição genômica dos três elementos foi verificada por meio de hibridização *in situ* fluorescente em cromossomos mitóticos da batata (*S. tuberosum*), do tomate (*S. lycopersicum*) e de outras quatro espécies de batatas selvagens (*S. verrucosum*, *S. chomatophilum*, *S. cardiophyllum* e *S. palustre*). St1 está concentrado principalmente nos pericentrômeros de todas as espécies envolvidas no trabalho e é o elemento mais abundante no genoma da batata. St6, encontrado de forma bastante dispersa em todos os genomas analisados, não apresentou localização preferencial. St73 localizou-se quase que exclusivamente nos centrômeros de todos os cromossomos da batata e de alguns cromossomos de *S. verrucosum*, *S. chomatophilum* e *S. palustre*. St1 e St6 são retroelementos mais antigos enquanto que St73 parece ter surgido mais recentemente no gênero. Parecem existir mecanismos de controle para inserção e excisão dos retroelementos aqui descritos e existem indícios de que St1 e principalmente St73 sejam importantes para os centrômeros de *S. tuberosum* e de outras espécies de *Solanum*.

Palavras-chave: Retrotransposons LTR. Retrotransposons centroméricos. Centrômero. Batata. *Solanum*.

ABSTRACT

LTR retrotransposons are abundant components that contribute to the high dynamics of the genomes due to its ability to move and change the genomic organization. The potato (*Solanum tuberosum*) genome sequencing revealed that at least 62.2% of its genome is repetitive DNA. The major fraction (29.4%) is made of LTR retrotransposon. Using these sequencing data, three Ty-3 Gypsy LTR retrotransposons were identified: St1, St6 and St73. Their genomic distribution was verified by fluorescence *in situ* hybridization in mitotic chromosomes of potato (*S. tuberosum*), tomato (*S. lycopersicum*) and four other wild potato species (*S. verrucosum*, *S. chomatophilum*, *S. cardiophyllum* and *S. palustre*). St1 is mainly concentrated in the pericentromeres of all species involved in this research and is the most abundant element of potato genome. St6, found dispersed throughout the analyzed genomes, showed no preferential location. St73 was located almost exclusively in the centromeres of all potato chromosomes and in some chromosomes of *S. verrucosum*, *S. chomatophilum* and *S. palustre*. St1 and St6 retrotransposons are older while St73 seems to have emerged more recently in the genus. The data suggest that insertion and excision of these retrotransposons are controlled and that St1 and especially St73 are important to *S. tuberosum* and other *Solanum* centromeres.

Keywords: LTR Retrotransposons. Centromeric retrotransposons. Centromere. Potato. *Solanum*.

1 INTRODUÇÃO

Genomas vegetais são extraordinariamente dinâmicos e de acordo com Ammiraju et al. (2007) uma força determinante de todo esse dinamismo é o fluxo de retrotransposons, que pode resultar em alterações dramáticas no tamanho e na organização genômica.

Retrotransposons são elementos genéticos capazes de se moverem dentro do genoma. Para que a transposição ocorra, é realizada a transcrição reversa do RNA transcrito do retroelemento, produzindo uma molécula de dupla fita de DNA que irá se inserir em um sítio alvo (WICKER et al., 2007). Dessa forma, cada ciclo completo de replicação produz uma nova cópia do elemento, fazendo dos retrotransposons os principais contribuintes para a fração repetitiva dos grandes genomas (KUMAR; BENNETZEN, 1999).

Estes retrotransposons geralmente contêm dois genes: *gag*, que codifica uma poliproteína estrutural semelhante às de vírus e *pol*, que codifica proteínas necessárias para a replicação, como a transcriptase reversa, RNase, integrase e proteinase aspártica (WICKER et al., 2007). Dentre as ordens definidas para esses elementos, a ordem LTR (do inglês Long Terminal Repeat) é considerada como a mais importante por ser a mais abundante nos genomas vegetais, como no caso do milho onde representa mais de 50% do genoma (SANMIGUEL et al., 1996, 1998).

Retrotransposons LTR são posteriormente classificados em cinco superfamílias: *TY-1 Copia*, *TY-3 Gypsy*, *Bel-Pao*, *Retrovirus* e *ERVs*, cada uma delas subdivididas em famílias, sendo que espécies com genomas grandes podem conter milhares delas (WICKER et al., 2007). As famílias de retrotransposons vegetais diferem consideravelmente umas das outras não somente por sua sequência e estrutura, mas também devido à sua distribuição cromossômica. Enquanto que algumas são dispersas, sem aparente localização preferencial, outras se concentram em regiões cromossômicas bastante

específicas (KEJNOVSKÝ et al., 2006; BALINT-KURTI et al., 2000).

Retrotransposons LTR se acumulam com bastante frequência nos centrômeros e pericentrômeros, tanto de plantas quanto de animais (PIMPINELLI et al. 1995; COPENHAVER et al. 1999; SCHUELER et al. 2001; CHENG et al. 2002). Existem evidências de que a atividade transposicional dos retroelementos LTR centroméricos influenciem na dinâmica evolutiva dos centrômeros, bem como em sua estrutura e função, pela geração de novas inserções que posteriormente podem passar por processos de recombinação homóloga ilegítima ou desigual (MA et al., 2007; WU et al., 2009). Segundo Plohl, Meštrović e Mravinac (2014), a compreensão dos mecanismos de acumulação, diversificação, capacidade de interação com proteínas e distribuição dos retroelementos é essencial para o completo entendimento de seu impacto nos genomas de forma geral e também nestas estruturas, como os centrômeros, onde eles preferencialmente se acumulam.

O sequenciamento completo do genoma da batata (*Solanum tuberosum* L.) revelou que pelo menos 62.2% do genoma é constituído de DNA repetitivo, dos quais 29.4% são retrotransposons LTR. Essas porcentagens devem ser ainda maiores visto que 117 Mb (14% do tamanho esperado de 844 Mb) não foram incluídos no genoma final e provavelmente são compostos por DNA repetitivo, previamente estimados em cerca de 75% do genoma por meio de sequências BAC e cosmídeos (PGSC, 2011).

Diante da reconhecida importância dos retroelementos LTR na dinâmica e evolução dos genomas e de estruturas cromossômicas, o objetivo deste trabalho foi verificar a existência e distribuição destes retroelementos no genoma da batata (*S. tuberosum*) e analisar seu padrão evolutivo em espécies cultivadas e selvagens do gênero *Solanum* L.

2 MATERIAL E MÉTODOS

Identificação dos retroelementos a partir de dados de sequenciamento 454

Cerca de 1.240.000 reads, com comprimento médio de 361 nucleotídeos, totalizando 447 Mb (53% do genoma haploide da batata - 840 Mb) foram disponibilizados pelo Consórcio de Sequenciamento (PGSC). Estes reads foram inicialmente avaliados para exclusão de sequências indesejáveis. As sequências remanescentes foram usadas para reconstrução das sequências repetitivas de acordo com a metodologia descrita por Macas et al., (2007).

Material genético

DM1-3 516R44 (DM1-3, $2n=2x=24$, genoma A), o clone monoploide homozigoto duplicado usado para montar o genoma da batata (PGSC, 2011), foi utilizado como representante da batata cultivada (*S. tuberosum*), sendo usado para a caracterização das sequências identificadas na etapa anterior.

A investigação do padrão evolutivo dessas sequências foi realizada em espécies selvagens de batata, incluindo representantes dos diferentes genomas básicos identificados no gênero *Solanum*: *S. verrucosum* Schltdl. ($2n=2x=24$, genoma A); *S. chomatophilum* Bitter ($2n=2x=24$, genoma P) e *S. cardiophyllum* Lindl. ($2n=2x=24$, genoma B). O tomate (*S. lycopersicum* L., $2n=2x=24$, genoma T) e *S. palustre* Poepp. ex Schltdl. ($2n=2x=24$, genoma E) foram utilizadas como representantes de espécies distantes da batata.

As sementes foram obtidas do USDA/ARS, Potato Introduction Station, Sturgeon Bay, Wisconsin, USA. Todos os genótipos foram cultivados em casa de vegetação para obtenção de radículas e tecido foliar.

Experimentos com FISH

- Preparações cromossômicas para análise de metáfase mitótica

Pontas de raízes foram coletadas de plantas jovens em casa de vegetação e pré-tratadas com 8-hidroxiquinoleína 3mM por 3 h a temperatura ambiente, lavadas por 30 min em água, fixadas em solução fixadora 3 etanol: 1 ácido acético e armazenadas a -20°C. As pontas de raiz fixadas foram lavadas três vezes em água destilada por cinco minutos cada e a região meristemática foi transferida para solução de enzimas contendo celulase 4% e pectinase 2% por cerca de 1 h a 37°C. A solução enzimática foi então substituída por água destilada gelada e após 30 minutos as pontas de raiz foram transferidas para solução fixadora. Para o preparo das lâminas utilizou-se o método de secagem à chama.

- Amplificação e marcação das sondas

Primers com 20 nucleotídeos foram desenhados por meio do programa Primer 3 v.0.4.0 (ROZEN AND SKALETISKY, 2000) para amplificação via PCR (Reação em Cadeia da Polimerase) de uma parte de cada uma das sequências identificadas a partir dos dados de sequenciamento 454, todas com cerca de 1kb. As reações de amplificação foram feitas usando DNA genômico isolado de folhas jovens do genótipo DM1-3 usando o Kit para extração de DNA genômico QIAGEN (Valencia, CA). Os produtos da reação foram visualizados em eletroforese com gel de agarose 1% para confirmação do tamanho do(s) fragmento(s) e verificação do padrão de amplificação.

Os produtos de PCR foram marcados com biotina-16-UTP ou digoxigenina-11-dUTP (Roche Diagnostic, Indianapolis) por meio de reação “nick translation” padrão e o tamanho dos fragmentos marcados foi confirmado em eletroforese com gel de agarose.

- Hibridização e detecção das sondas

Os cromossomos mitóticos foram desnaturados com formamida 70% em SSC 2x, por cerca de 1,5 min a 85° C e desidratados em série alcóolica: etanol 70% (-20 °C); 90% e 100% por 5 min cada. A mistura de hibridização (Formamida 50%, SSC 2x, sulfato de dextran 10%, 50-100 ng de sonda marcada) foi desnaturada por 8 min a 95° C e resfriada em gelo por pelo menos 5 minutos. A mistura de hibridização foi aplicada sobre a preparação cromossômica, coberta com lamínula 22 X 40 mm, vedada com adesivo de borracha e colocada em câmara úmida por no mínimo 16 horas.

A detecção das sondas foi feita com o uso dos anticorpos streptavidina conjugada com alexafluor 488 e antidigoxigenina conjugada com rodamina (Roche Diagnostics, Indianapolis, IN) após lavagens estringentes com tampão SSC 2x e TNT. Os cromossomos foram contracorados com DAPI (diamidínefenilindole) em meio de montagem “antifade” VectorShield (Vector Laboratories, Burlingame, CA). As imagens foram capturadas usando uma camera QImaging Retiga EXi CCD acoplada a um microscópio de fluorescência Olympus BX 60 e o contraste final feito com o software Photoshop CS5.

3 RESULTADOS

Análises bioinformáticas permitiram identificar três retroelementos: St1, St6 e St73, com comprimentos de cerca de 6.6, 8.1 e 6.6 kb respectivamente. Eles foram classificados como pertencentes à ordem LTR e superfamília *Ty-3 Gypsy*. Na classificação dentro da superfamília, St1 apresentou semelhanças a retrotransposons da linhagem *Athila* enquanto que St73 e St6 são cromovírus, linhagem caracterizada pela presença de um cromodomínio (domínio modificador da organização da cromatina). Dentre os cinco clados existentes na subdivisão dos cromovírus, St73 pertence ao CRM enquanto que St6 pertence ao *Tekay*.

Os experimentos com a hibridização *in situ* fluorescente (FISH) revelaram ampla distribuição destes retroelementos no genoma de DM1-3, bem como nos genomas de outras espécies do gênero *Solanum* analisadas. O padrão de distribuição genômica foi diferenciado para cada um dos retrotransposons.

St73 tem localização quase que exclusiva nos centrômeros de DM1-3 (Figura1 A-C) e está presente em todos os cromossomos do complemento. O número de cópias do retroelemento por cromossomo é variável, sendo mais abundante em alguns centrômeros do que em outros. St73 foi também identificado, com menor abundância, nos centrômeros de *S. verrucosum* (Figura1 D-F), *S. chomatophilum* (Figura1 G-I) e *S. palustre* (Figura1 J-L). Ele está presente em aproximadamente 16 cromossomos da primeira espécie, oito da segunda e seis da terceira. Essa repetição está ausente em *S. cardiophyllum* e em *S. lycopersicum*.

St1 é, dentre as sequências aqui estudadas, a que apresenta maior abundância no genoma de DM1-3 (Figura2 A-C). O retroelemento foi encontrado em todas as espécies avaliadas, porém em menor quantidade em *S. cardiophyllum*, *S. lycopersicum* e *S. palustre*, e é localizado principalmente em regiões pericentroméricas. Ele está também presente nas regiões intersticiais dos cromossomos, mas sua abundância diminui à medida que aumenta a proximidade dos telômeros, o que pode também ser observado principalmente em *S. verrucosum* (Figura1 D-F) e *S. chomatophilum* (Figura 1 G-I).

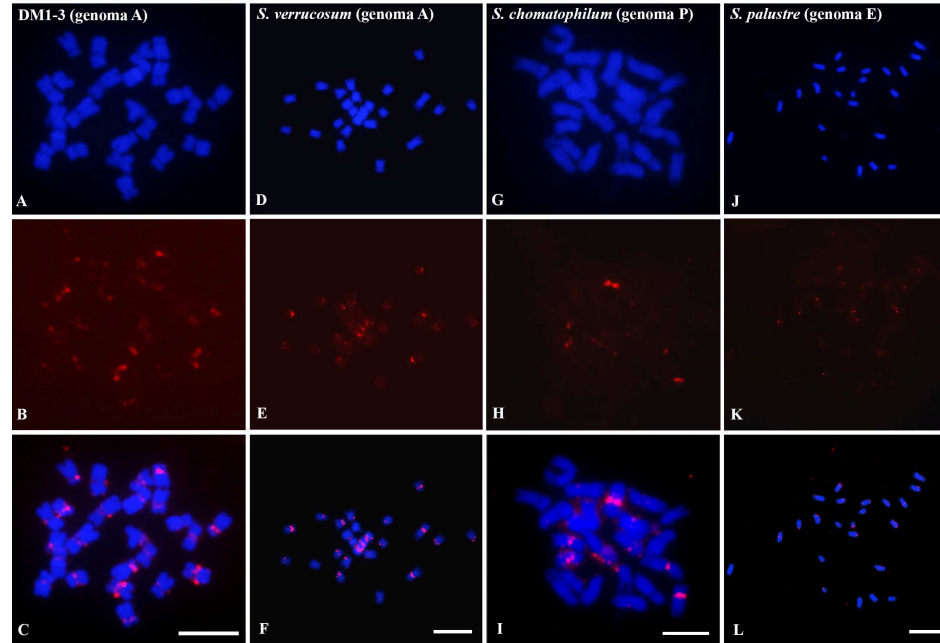


Figura 1 Distribuição de St73 por meio de FISH com marcação vermelha nos cromossomos de espécies diploides ($2n=2x=24$) de *Solanum*. **(A)** Cromossomos de DM1 corados com DAPI **(B)** Sinais de FISH gerados em todos os cromossomos de DM1-3 **(C)** Localização da sequência nos cromossomos de DM1-3 **(D)** Cromossomos de *S. verrucosum* corados com DAPI **(E)** Sinais de FISH gerados em cerca de 16 centrômeros de *S. verrucosum* **(F)** Localização da sequência nos cromossomos de *S. verrucosum* **(G)** Cromossomos de *S. chomatophilum* corados com DAPI **(H)** Sinais de FISH gerados em cerca de 8 centrômeros de *S. chomatophilum* **(I)** Localização da sequência nos cromossomos de *S. chomatophilum*. **(J)** Cromossomos de *S. palustre* corados com DAPI **(K)** Sinais de FISH gerados em cerca de 6 centrômeros de *S. palustre* **(L)** Localização da sequência nos cromossomos de *S. palustre*. Barras: 5µm.

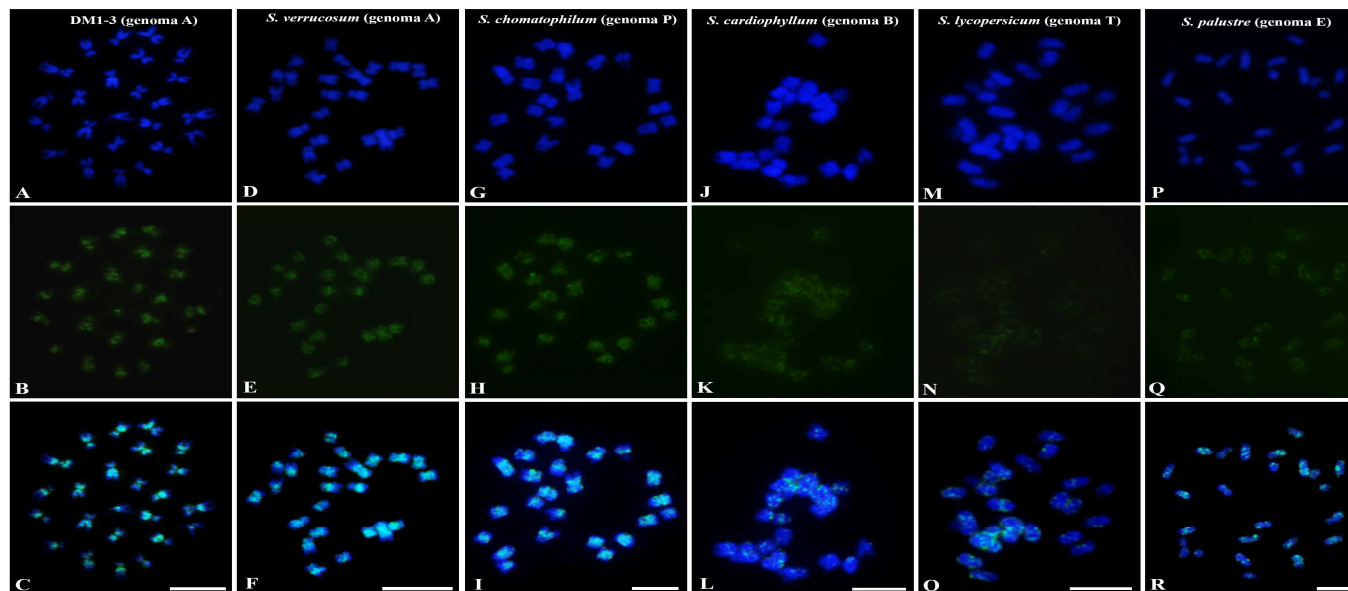


Figura 2 Distribuição de St1 por meio de FISH com marcação verde nos cromossomos de espécies diploides ($2n=2x=24$) de *Solanum*. **(A)** Cromossomos de DM1-3 corados com DAPI **(B)** Sinais de FISH gerados em DM1-3 **(C)** Localização da sequência nos cromossomos de DM1-3 **(D)** Cromossomos de *S. verrucosum* corados com DAPI **(E)** Sinais de FISH gerados em *S. verrucosum* **(F)** Localização da sequência nos cromossomos de *S. verrucosum* **(G)** Cromossomos de *S. chomatophilum* corados com DAPI **(H)** Sinais de FISH gerados em *S. chomatophilum* **(I)** Localização da sequência nos cromossomos de *S. chomatophilum* **(J)** Cromossomos de *S. cardiophyllum* corados com DAPI **(K)** Sinais de FISH gerados em *S. cardiophyllum* **(L)** Localização da sequência nos cromossomos de *S. cardiophyllum* **(M)** Cromossomos de *S. lycopersicum* corados com DAPI **(N)** Sinais de FISH gerados em *S. lycopersicum* **(O)** Localização da sequência nos cromossomos de *S. lycopersicum* **(P)** Cromossomos de *S. palustre* corados com DAPI **(Q)** Sinais de FISH gerados em *S. palustre* **(R)** Localização da sequência nos cromossomos de *S. palustre*. Barras: 5µm.

Em DM1-3, embora St1 seja muito abundante nas regiões pericentroméricas, é possível observar na maioria dos cromossomos uma falha do sinal exatamente na região centromérica, revelando sua presença pouco abundante ou até mesmo sua ausência no centrômero (Figura 1C e Figura 3).

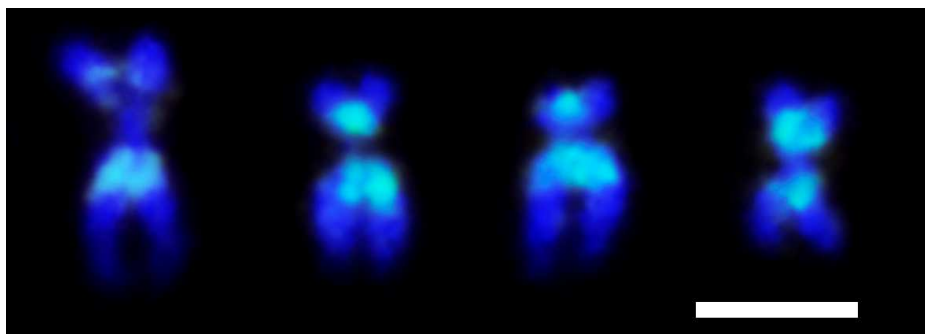


Figura 3 Alguns cromossomos de DM1-3 selecionados como exemplo para melhor visualização dos grandes sinais provenientes da sonda St1 na região pericentromérica e presença reduzida ou falta desses sinais na região do centrômero. Barra: 2.5 μ m.

O retroelemento CL6 tem distribuição ampla e uniforme no genoma de DM1-3, aparentemente sem localização preferencial (Figura4 A-C). Ele está presente de forma bastante dispersa por todos os cromossomos e ausente apenas nas regiões teloméricas e subteloméricas. O retroelemento foi encontrado em todas as demais espécies analisadas (Figura 4), com distribuição semelhante à observada em DM1-3, porém com menor abundância em *S. cardiophyllum* (Figura4 J-L), *S. lycopersicum* (Figura 4, M-O) e *S. palustre* (Figura 4 P-R).

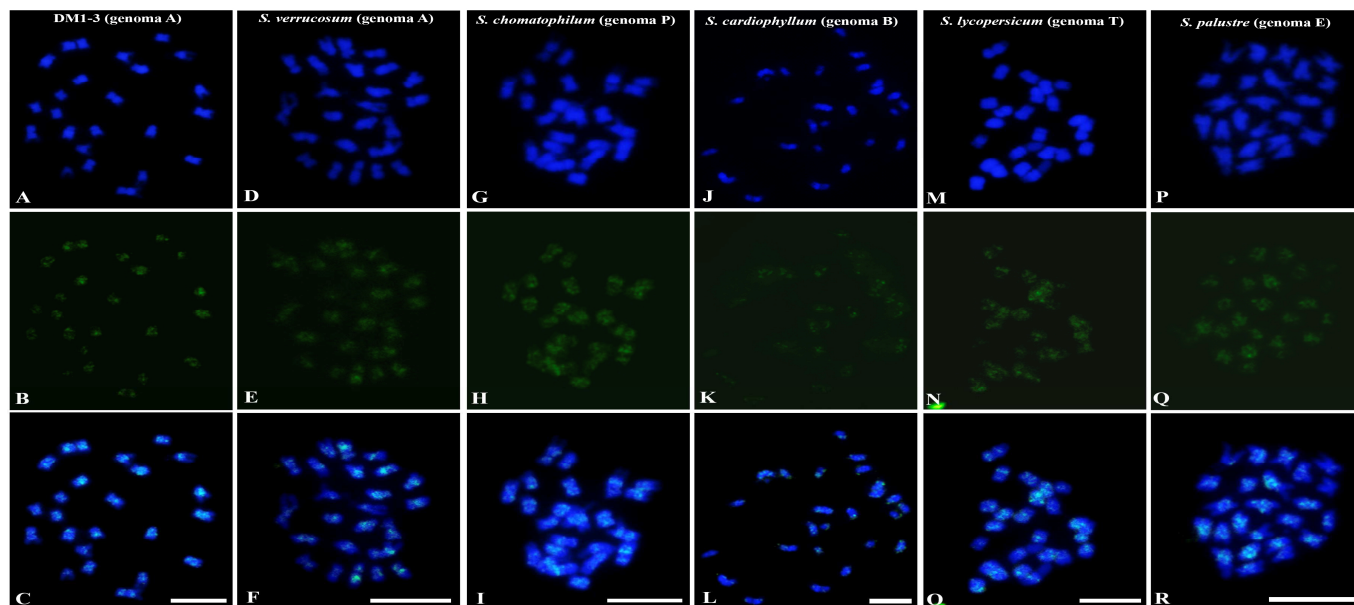


Figura 4 Distribuição de St6 por meio de FISH com marcação verde nos cromossomos de espécies diploides ($2n=2x=24$) de *Solanum*. **(A)** Cromossomos de DM1-3 corados com DAPI **(B)** Sinais de FISH gerados em DM1-3 **(C)** Localização da sequência nos cromossomos de DM1-3 **(D)** Cromossomos de *S. verrucosum* corados com DAPI **(E)** Sinais de FISH gerados em *S. verrucosum* **(F)** Localização da sequência nos cromossomos de *S. verrucosum* **(G)** Cromossomos de *S. chomatophilum* corados com DAPI **(H)** Sinais de FISH gerados em *S. chomatophilum* **(I)** Localização da sequência nos cromossomos de *S. chomatophilum* **(J)** Cromossomos de *S. cardiophyllum* corados com DAPI **(K)** Sinais de FISH gerados em *S. cardiophyllum* **(L)** Localização da sequência nos cromossomos de *S. cardiophyllum* **(M)** Cromossomos de *S. lycopersicum* corados com DAPI **(N)** Sinais de FISH gerados em *S. lycopersicum* **(O)** Localização da sequência nos cromossomos de *S. lycopersicum* **(P)** Cromossomos de *S. palustre* corados com DAPI **(Q)** Sinais de FISH gerados em *S. palustre* **(R)** Localização da sequência nos cromossomos de *S. palustre*. Barras: 5 μ m.

4 DISCUSSÃO

St73 é, dentre as sequências aqui estudadas, um retroelemento de destacada importância devido à sua localização centromérica. Ele divide o clado dos retrovírus CRM com inúmeros outros retrotransposons de localização também centromérica (GORINSEK; GUBENSEK; KORDIS, 2004; GORINSEK; GUBENSEK; KORDIS, 2005; KORDIS, 2005), que já foram caracterizados, dentre outras espécies, no trigo (LIU et al., 2008), milho (NAGAKI et al. 2003a; ZHONG et al. 2002), arroz (BAO et al., 2006), cana de açúcar (NAGAKI et al., 2005) e cevada (HUDAKOVA et al., 2001; PRESTING et al., 1998).

Os centrômeros dos eucariotos são, de forma geral, compostos por duas classes de DNA repetitivo: o DNA satélite e os retrotransposons (PLOHL; MEŠTROVIĆ; MRAVINAC, 2014). O DNA satélite centromérico, componente majoritário de inúmeros centrômeros, é geralmente conservado dentro da espécie, ou seja, uma única repetição satélite ou poucas delas compõem todos os centrômeros dentro de um complemento cromossômico; como ocorre em *Arabidopsis* (NAGAKI et al., 2003b) e humanos (WILLARD, 1998). Porém quando se faz uma comparação das repetições satélite centroméricas entre espécies, o nível de homologia geralmente é baixo, havendo grandes diferenças até mesmo entre espécies proximamente relacionadas (LEE et al., 2005).

Algumas poucas espécies apresentam configurações peculiares do DNA centromérico e uma delas é a batata (*S. tuberosum*). Alguns centrômeros da espécie são compostos por DNA de cópia única, o que é incomum para DNA centromérico, e outros por DNA satélite. O mais interessante é que cada um dos doze centrômeros da espécie apresenta uma configuração única do DNA, não tendo sido encontrada qualquer sequência conservada que estivesse presente em todos eles (GONG et al., 2012). Dessa forma, St73 é até então o único componente genético bem conservado e presente em todos os centrômeros da batata, um indicativo de que ele desempenhe ali um importante papel. Como foi

também encontrado em *S. verrucosum*, *S. chomatophilum* e *S. palustre*, St73 apresenta ainda relativa conservação entre as espécies do gênero *Solanum*, o que está de acordo com a afirmação feita por Nagaki et al. (2009), de que enquanto o DNA satélite centromérico evolui rapidamente, os retrotransposons centroméricos parecem evoluir de forma mais lenta.

A grande maioria dos retrotransposons centroméricos já estudados é ativamente transcrita (WEBER; SCHMIDT, 2009; TOPP et al., 2004; NEUMANN et al., 2007). Jiang et al. (2003) atribuem a esta transcrição a base da substituição das histonas H3 por histonas CenH3 (variantes da histona H3 centrômero-específicas essenciais para a montagem do cinetócoro e para o estabelecimento e a manutenção da função centromérica). A hipótese de Jiang et al. (2003) surgiu a partir de uma sugestão feita por Henikoff et al. (2001) de que a substituição das histonas se dá durante a transcrição, um momento que seria propício para a troca devido à desestabilização ou desmontagem dos nucleossomos. Uma vez que, em termos de composição genética, os principais componentes centroméricos são o DNA satélite, que geralmente não é transcrito, e os retrotransposons (PLOHL, MEŠTROVIĆ; MRAVINAC, 2014), a transcrição destes últimos seria elemento chave na determinação da função centromérica por propiciar a incorporação da CenH3. Acredita-se também que os promotores de retrotransposons centroméricos sejam importantes não somente para sua própria transcrição, mas também para a transcrição de sequências adjacentes a eles (JIANG et al., 2003).

Outro indicativo de que a transcrição dos retrotransposons seja de fato importante para os centrômeros foi a observação de que os transcritos do retrotransposon centromérico CRM do milho permanecem diretamente ligados à cromatina centromérica, o que sugere que eles desempenham um papel estabilizante em sua estrutura (TOPP et al., 2004).

Em alguns dos centrômeros de todas as espécies que apresentaram St73 é possível observar a existência de grandes blocos constituídos por este retroelemento, o que é um indício de que haja ali uma organização em tandem. Esta estrutura em tandem de retrotransposons, embora incomum, foi também descrita para retrotransposons não LTR em *Tetraodon nigroviridis*, o peixe baiacu (BOUNEAU et al., 2003)

Embora tanto St73 quanto St6 sejam cromovírus, a distribuição genômica verificada para os dois retroelementos foi bastante diferente. Os cromovírus são caracterizados pela presença de um cromodomínio, que é um domínio envolvido no remodelamento da cromatina e na regulação da expressão gênica nos eucariotos durante seu desenvolvimento (NOVIKOVA, 2009). Embora ele esteja presente em todos os cromovírus, segundo Kordis (2005) os componentes do clado CRM, ao qual pertence St73, apresentam cromodomínio diferenciado, e acredita-se que é este sítio diferenciado que garante a correta ligação dos retroelementos à região centromérica. St6, pertencente ao clado Tekay, que não possui tal sítio, apresenta por consequência distribuição mais ampla e desempenha provavelmente função estrutural nos cromossomos.

St1, que apresentou localização preferencial pericentromérica, pode também desempenhar uma relevante função no genoma da batata. De acordo com diversos autores, retroelementos nas regiões pericentroméricas desempenham um importante papel no recrutamento de enzimas modificadoras de histonas e também na manutenção da heterocromatina pericentromérica, fundamental para o controle de recombinações envolvendo os centrômeros (BENDER, 2004; HALL et al. 2002; LIPPMAN et al. 2004; VOLPE et al. 2002; ZHONG et al. 2002). Em *Saccharomyces pombe*, por exemplo, foi mostrado que repetições centroméricas presentes nos pericentrômeros são necessárias para a formação de heterocromatina mediada por RNAi (FOLCO et al., 2008; KAGANSKY et al., 2009).

Tratando-se, por fim, dos aspectos evolutivos, St1 e St6, uma vez encontrados até mesmo nas espécies mais distantemente relacionadas à *S. tuberosum*, são retroelementos bem conservados e provavelmente surgiram no gênero antes da divergência das espécies analisadas. St1 aparentemente enfrentou pressões seletivas distintas entre as espécies, uma vez que sua abundância e distribuição é bastante semelhante entre DM1-3, *S. verrucosum* e *S. chomatophilum*, mas diferente do que se observa em *S. cardiophyllum*, *S. lycopersicum* e *S. palustre*. St73, presente nas espécies com genoma A, P e E, parece ter surgido mais recentemente na história evolutiva do gênero. Uma vez que esta repetição não foi encontrada em *S. cardiophyllum* e *S. lycopersicum*, que são espécies mais próximas das espécies com genoma A e P, é possível que sua origem em *S. palustre* (genoma E) tenha ocorrido independentemente das demais espécies.

Nossos resultados sugerem a existência de sítios preferenciais e mecanismos de controle para a inserção dos diferentes retrotransposons e trazem indícios de que St1 e principalmente St73 sejam retroelementos importantes para os centrômeros de *S. tuberosum* e de outras espécies de *Solanum*.

REFERÊNCIAS

AMMIRAJU, J. S. S. et al. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. *Plant J.* 52, 342–351 (2007).

BALINT-KURTI PJ, CLENDENNEN SK, DOLEŽELOVÁ M, VALÁRIK M, DOLEŽEL J, BEETHAM PR, MAY GD: Identification and chromosomal localization of the monkey retrotransposon in *Musa* sp. *Mol Gen Genet* 2000, 263:908-915.

BAO W, ZHANG W, YANG Q, ZHANG Y, HAN B, GU M, XUE Y, CHENG Z: Diversity of centromeric repeats in two closely related wild rice species, *Oryza officinalis* and *Oryza rhizomatis*. *Mol Genet Genomics* 2006, 275:421-430.

BENDER J (2004) Chromatin-based silencing mechanisms. *Curr Opin lant Biol* 7:521–526.

BOUNEAU, LAURENCE, et al. "An active non-LTR retrotransposon with tandem structure in the compact genome of the pufferfish *Tetraodon nigroviridis*." *Genome research* 13.7 (2003): 1686-1695.

CHENG Z, DONG F, LANGDON T, OUYANG S, BUELL CR, GU M, BLATTNER FR, JIANG J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14: 1691–1704

COPENHAVER GP, NICKEL K, KUROMORI T, BENITO MI, KAUL S, LIN X, BEVAN M, MURPHY G, HARRIS B, PARNELL LD, MCCOMBIE WR, MARTIENSSEN RA, MARRA M, PREUSS D (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286:2468–2474.

FOLCO HD, PIDOUX AL, URANO T, ALLSHIRE RC: Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. *Science* 2008, 319:94-97.

GONG, Z.Y., Wu, Y.F., KOBLÍZKOVÁ, A., TORRES, G.A., WANG, K., IOVENE, M., NEUMANN, P., ZHANG, W.L., NOVÁK, P., BUELL, C.R., MACAS, J., and JIANG, J.M. (2012). Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* 24: 3559–3574.

GORINSEK B, GUBENSEK F, KORDIS D: Evolutionary genomics of chromoviruses in eukaryotes. *Mol Biol Evol* 2004, 21:781-798.

GORINSEK B, GUBENSEK F, KORDIS D: Phylogenomic analysis of chromoviruses. *Cytogenet Genome Res* 2005, 110:543-552.

HALL IM, SHANKARANARAYANA GD, NOMA KI, AYOUB N, COHEN A, GREWAL SI (2002) Establishment and maintenance of a heterochromatin domain. *Science* 297:2232–2237

HENIKOFF S, AHMAD K, MALIK HS: The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 2001, 293:1098-1102.

HUDAKOVA S, MICHALEK W, PRESTING GG, TEN HOOPEN R, DOS SANTOS K, JASENCAKOVA Z, SCHUBERT I: Sequence organization of barley centromeres. *Nucleic Acids Res* 2001, 29:5029-5035.

JIANG, J; BIRCHLER J. A.; PARROTT, W.A.; Dawe, R.K (2003) A molecular view of plant centromeres. *Trends in Plant Science*, 8 (12): 570-575.

KAGANSKY A, FOLCO HD, ALMEIDA R, PIDOUX AL, BOUKABA A, SIMMER F, URANO T, HAMILTON GL, ALLSHIRE RC: Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. *Science* 2009, 324:1716-1719.

KEJNOVSKÝ E, KUBÁT Z, MACAS J, HOBZA R, MRÁČEK J, VYSKOT B: RETAND: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat. *Mol Genet Genomics* 2006, 276:254-263.

KORDIS D: A genomic perspective on the chromodomain-containing retrotransposons: chromoviruses. *Gene* 2005, 347:161-173.

KUMAR, A. & BENNETZEN, J. Plant retrotransposons. *Annu. Rev. Genet.* 33, 479–532 (1999).

LEE , H.R. , ZHANG , W.L. , LANGDON , T. , JIN , W.W. , YAN , H.H. , CHENG , Z.K. , and JIANG , J.M. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. **Proc Nat Acad Sci U S A** 102, 11793 – 11798.

LIPPMAN Z, GENDREL AV, BLACK M, VAUGHN MW, DEDHIA N, MCCOMBIE WR, LAVINE K, MITTAL V, MAY B, KASSCHAU KD, CARRINGTON JC, DOERGE RW, COLOT V, MARTIENSSEN R (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476

LIU Z, YUE W, LI D, WANG R, KONG X, LU K, WANG G, DONG Y, JIN W, ZHANG X: Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. *Chromosoma* 2008, 117:445-456.

MA J, WING RA, BENNETZEN JL, JACKSON SA: Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet* 2007, 23:134-139.

MACAS J, NEUMANN, P. NAVRÁTILOVÁ, A Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 2007, 8:427, 2007.

NAGAKI , SONG J, STUPAR RM, PAROKONNY AS, YUAN Q, OUYANG S, LIU J, HSIAO J, JONES KM, DAWE RK, BUELL CR, JIANG J: Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. *Genetics* 2003a, 163:759-770.

NAGAKI , K. , TALBERT , P.B. , ZHONG , C.X. , DAWE , R.K. , HENIKOFF , S. , and JIANG , J.M. (2003b). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. **Genetics** 163, 1221 – 1225.

NAGAKI K, MURATA M: Characterization of CENH3 and centromere-associated DNA sequences in sugarcane. *Chromosome Res* 2005, 13:195-203.

NAGAKI K, WALLING J, HIRSCH C, JIANG J, MURATA M: Structure and evolution of plant centromeres. *Prog Mol Subcell Biol* 2009, 48:153-179.

NEUMANN P, YAN H, JIANG J: The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. *Genetics* 2007, 176:749-761.

NOVIKOVA O: Chromodomains and LTR retrotransposons in plants. *Comm & Integr Biol* 2009, 2:158-162.

PGSC (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–194

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Centromere identity from the DNA point of view. *Chromosoma*, 123 (2014), pp. 313–325.

PIMPINELLI S, BERLOCO M, FANTI L, DIMITRI P, BONACCORSI S, MARCHETTI E, CAIZZI R, CAGGESE C, GATTI M (1995) Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. *Proc Natl Acad Sci U S A* 92:3804–3808

PRESTING GG, MALYSHEVA L, FUCHS J, SCHUBERT I: A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J* 1998, 16:721-728.

ROZEN S, SKALESTSKY HJ. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Krawetz S, Misener S, editors. Humana Press: Totowa, NJ; 2000. p. 365-386.

SANMIGUEL, P., B. S. GAUT, A. TIKHONOV, Y. NAKAJIMA, AND J. L. BENNETZEN. 1998. The paleontology of intergene retrotransposons of maize. *Nature Genet.* 20:43–45.

SANMIGUEL, P., A. TIKHONOV, Y. K. JIN ET AL (11 co-authors). 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274:765–768.

SCHUELER MG, HIGGINS AW, RUDD MK, GUSTASHAW K, WILLARD HF (2001) Genomic and genetic definition of a functional human centromere. *Science* 294:109–115

TOPP CN, ZHONG CX, DAWE RK: Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci USA* 2004, 101:15986-15991.

VOLPE TA, KIDNER C, HALL IM, TENG G, GREWAL SIS, MARTIENSSSEN RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833–1837

WEBER B, SCHMIDT T: Nested Ty3-gypsy retrotransposons of a single *Beta procumbens* centromere contain a putative chromodomain. *Chromosome Res* 2009, 17:379-396.

WICKER, T., F. SABOT, A. HUA-VAN, J.L. BENNETZEN, P. CAPY, B. CHALOUB, A. FLAVELL, P. LEROY, M. MORGANTE, O. PANAUD, E. PAUX, P. SANMIGUEL, and A. SCHULMAN. 2007. A unified classification system for eukaryotic transposable elements. *Nature Rev. Genet.* 8: 973-982.

WU J, FUJISAWA M, TIAN Z, YAMAGATA H, KAMIYA K, SHIBATA M, HOSOKAWA S, ITO Y, HAMADA M, KATAGIRI S, KURITA K, YAMAMOTO M, KIKUTA A, MACHITA K, KARASAWA W, KANAMORI H, NAMIKI N, MIZUNO H, MA J, SASAKI T, MATSUMOTO T: Comparative analysis of complete orthologous centromeres from two subspecies of rice reveals rapid variation of centromere organization and structure. *Plant J* 2009, 60:805-819.

WILLARD, H.F. (1998). Centromeres: the missing link in the development of human artificial chromosomes. *Curr Opin Genet Dev* 8, 219 – 225.

ZHONG CX, MARSHALL JB, TOPP C, MROCZEK R, KATO A, NAGAKI K, BIRCHLER JA, JIANG JM, DAWE RK: Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 2002, 14:2825-2836.