Universidade de São Paulo Escola Superior de Agricultura "Luiz de Queiroz"

Análise comparativa de regiões microssintênicas entre *Passiflora edulis* e as Malpighiales *Populus trichocarpa* e *Manihot esculenta*

Alina del Carmen Egoávil Reátegui

Dissertação apresentada para obtenção do título de Mestra em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba 2019 Alina del Carmen Egoávil Reátegui Licenciada em Ciências Biológicas

Análise comparativa de regiões microssintênicas entre *Passiflora edulis* e as Malpighiales *Populus trichocarpa* e *Manihot esculenta*

Orientadora: Prof. Dra. MARIA LUCIA CARNEIRO VIEIRA

Dissertação para obtenção do título de Mestra em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba 2019

Dados Internacionais de Catalogação na Publicação DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Egoávil Reátegui, Alina del Carmen

Análise comparativa de regiões microssintênicas entre *Passiflora edulis* e as Malpighiales *Populus trichocarpa* e *Manihot esculenta /* Alina del Carmen Egoávil Reátegui. - Piracicaba, 2019.

67 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Genômica comparativa 2. Microssintenia 3. Malpighiales 4. *Passiflora edulis* 5. *Populus trichocarpa* 6. *Manihot esculenta* I. Título

DEDICATÓRIA

A mis padres (pilares fundamentales) y a mi hermano que siempre me apoyaron y motivaron para alcanzar mis anhelos.

AGRADECIMENTOS

A Deus, por ter me dado a força para seguir em frente.

À Escola Superior de Agricultura, "Luiz de Queiroz" (ESALQ/USP) e ao Departamento de Genética, pelo suporte acadêmico.

À Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Estudos Agrários Luiz de Queiroz (FEALQ) pelas bolsas concedidas, e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro para a realização do projeto.

À Prof. Dra. Maria Lucia Carneiro Vieira, pela orientação, paciência e oportunidade concedida. Meus sinceros agradecimentos por ter me aceitado em seu grupo de pesquisa.

Aos professores do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, pelos conhecimentos compartilhados.

À Dra. Carla de Freitas Munhoz, pelos ensinamentos nas análises de regiões mirossintênicas.

À Dra. Zirlane Portugal da Costa, pelo auxílio nas análises dos dados de elementos transponíveis, pela disponibilidade em esclarecer minhas dúvidas e pelos conhecimentos compartilhados.

Ao biotecnólogo Luiz Augusto Cauz dos Santos, pelos conhecimentos compartilhados.

Ao técnico de laboratório Carlos Alberto de Oliveira, pela amizade e pelo auxílio técnico durante a relização do trabalho.

Aos meus colegas do Laboratório de Genética Molecular de Plantas Cultivadas, pelo companherismo e bom convívio. Foi um enorme prazer conhecer todos vocês.

À minha querida familia, pelo amor incondicional e porque acreditaram em mim desde o primeiro instante.

Ao Angel, pelo carinho e valioso apoio em todos os momentos em que se tornou necessário.

Finalmente, minha gratidão a todos os que, de alguma forma, contribuíram com minha formação profissional e com o desenvolvimento deste trabalho.

SUMÁRIO

RESUMO	6
ABSTRACT	7
LISTA DE FIGURAS	8
LISTA DE TABELAS	10
1. INTRODUÇÃO	11
2. MATERIAL E MÉTODOS	15
2.1. Seleção das regiões genômicas de <i>P. edulis</i> a partir do conjunt sequenciados	to de insertos 15
2.2. Análise de microssintenia	15
2.3. Identificação dos elementos transponíveis	16
3. RESULTADOS	17
3.1. Seleção das regiões genômicas de P. edulis	17
3.2. Análise de microssintenia	17
3.3. Identificação dos elementos transponíveis	
4. DISCUSSÃO	51
5. CONCLUSÕES	55
REFERÊNCIAS	
APÊNDICE	65

RESUMO

Análise comparativa de regiões microssintênicas entre *Passiflora edulis* e as Malpighiales *Populus trichocarpa* e *Manihot esculenta*

A família Passifloraceae pertence à ordem Malpighiales com ~ 700 espécies e 16 gêneros. O gênero Passiflora, compreende a espécie comercial Passiflora edulis (2n = 18), o maracujá-azedo, que ocupa cerca de 95% dos pomares cultivados no Brasil. A produção é destinada à elaboração de sucos e ao consumo da fruta in natura. Embora a espécie tenha relevância econômica para o país, há carência de estudos genéticos e genômicos. Por esse motivo, e com o desenvolviemto de dados genômicos pelo nosso grupo de pesquisa, análises de microsintenia foram aqui conduzidas a partir de sequências ricas em genes de P. edulis comparativamente às sequências das Malpighiales, Populus trichocarpa (2n = 38) e Manihot esculenta (2n = 36). Neste estudo, foram selecionados 35 insertos BACs de P. edulis, ricos em genes e realizados os alinhamentos (BLASTN) das sequências, considerando um e-value de 10⁻¹⁰ e uma cobertura maior do 50%. Por comparações par a par, regiões com sintenia, colinearidade e com quatro ou mais pares de genes ortólogos foram tratadas como regiões microssintênicas. Para a identificação dos elementos transponíveis nos espaçamentos intergênicos, foi utilizado o pipeline REPET. No total, 32 insertos BACs de P. edulis mostraram microssintenia, 28 deles com P. trichocarpa e 25 com M. esculenta. Em ambas as comparações, a ordem dos genes foi conservada, e poucos rearranjos foram observados (inversões e translocações), sendo as inversões mais comuns. Maximamente, foram identificadas nove e cinco cópias de genes duplicados nas comparações com P. trichocarpa e M. esculenta, respectivamente. O maior grau de sintenia foi encontrado entre o inserto Pe173B16 de P. edulis e o cromossomo 9 de P. trichocarpa (29 ortólogos) e entre o inserto Pe185D11 de P. edulis e o cromossomo 12 de M. esculenta (27 ortólogos). Interessantemente, regiões de P. edulis mostraram microssintenia com segmentos distintos no mesmo cromossomo de P. trichocarpa ou M. esculenta. Além disso, regiões microssintênicas de P. edulis mostraram orientação em direção oposta. A busca de elementos transponíveis resultou em um total de 517 elementos, dentre os quais 248 em M. esculenta, 244 em P. trichocarpa e apenas 25 em P. edulis distribuídos em 17 regiões microssintênicas. Os retrotransposons (classe I) foram a classe predominante nas três espécies, destacando-se a ordem LTR. A ordem TIR de transposons de DNA (classe II) foi mais frequente em P. trichocarpa (52), seguindo-se M. esculenta (42) e P. edulis (2). Em síntese, tem-se um maior grau de microssintenia entre P. edulis e P. trichocarpa do que entre P. edulis e M. esculenta, corroborando a filogenia estabelecida. Comparativamente, a compactação dos genes é maior em P. edulis devido ao maior número de elementos repetitivos nos espacamentos intergênicos em P. trichocarpa e M. esculenta, embora ambas apresentem genomas menores. Sugere-se que o maior tamanho do genoma de P. edulis esteja associado à abundância de sequências repetitivas em regiões pobres em genes. Destaca-se que o método aqui usado visando à identificação de regiões microssintênicas representa uma importante ferramenta em análises comparativas, principalmente quando não se dispõe de genomas sequenciados, como é o caso de P. edulis.

Palavras-chave: Genômica comparativa; Microssintenia; Malpighiales; Passiflora edulis; Populus trichocarpa; Manihot esculenta

ABSTRACT

Comparative analysis of microsyntenic regions between *Passiflora edulis* and the Malpighiales *P. trichocarpa* and *M. esculenta*

The Passifloraceae family belongs to the order Malpighiales with ~ 700 species and 16 genera. The genus *Passiflora* includes the commercial species *P. edulis* (2n = 18), the sour passion fruit, which occupies approximately 95% of all orchards grown in Brazil. The production is destined for preparing juices and in natura fruit consumption. Although the species has economic relevance for the country, there is a lack of genetic and genomic studies. For this reason, and with the genomic data developed by our research group, microsyntenic analyzes were conducted herein using gene-rich sequences of P. edulis in comparison to those of the Malpighiales, P. trichocarpa (2n = 38) and M. esculenta (2n = 36). In this study, 35 gene-rich BACs fom P. edulis were selected and the alignments (BLASTN) of the sequences were performed, considering an e-value of 10^{-10} and coverage higher than 50%. By pairwise comparisons, regions with synteny, collinearity and four or more pairs of orthologous genes were treated as microsyntenic regions. For the identification of the transposable elements in the intergenic spaces, the REPET pipeline was used. In total, 32 BAC inserts showed microsynteny, 28 of them with P. trichocarpa and 25 with M. esculenta. In both comparisons, the gene order was conserved and few rearrangements were observed (inversions and translocations), being inversions more common. Maximally, nine and five copies of duplicate genes were found in the comparisons with P. trichocarpa and M. esculenta, respectively. The highest degree of sinteny was found between P. edulis insert Pe173B16 and P. trichocarpa chromosome 9 (29 orthologs) and between P. edulis insert Pe185D11 and M. esculenta chromosome 12 (27 orthologs). Interestingly, regions of P. edulis showed microsynteny with distinct segments on the same chromosome of P. trichocarpa or M. esculenta. In addition, microsyntenic regions of P. edulis showed orientation into opposite direction. The search for transposable elements resulted in a total of 517 elements, out of them 248 in M. esculenta, 244 in P. trichocarapa and only 25 in P. edulis distributed in 17 microsyntenic regions. The retrotransposons (class I) were the predominant class in the three species, specifically from LTR order. The TIR order of DNA transposons (class II) was more frequent in P. trichocarpa (52), followed by M. esculenta (42) and P. edulis (2). In synthesis, the degree of microsynteny between P. edulis and P. trichocarpa is higher than that between P. edulis and M. esculenta, corroborating the established phylogeny. Comparatively, the gene compaction is denser in P. edulis due to the greater number of repetitive elements in the intergenic spaces in P. trichocarpa and M. esculenta, although both have smaller genomes. It is suggested that the larger size of P. edulis genome is associated with the abundance of repetitive sequences in gene poor regions. Remarkable, the method used here to identify microsyntenic regions represents an important tool in comparative analyzes, mainly when sequenced genomes are not available, as is the case of *P. edulis*.

Keywords: Comparative genomics; Microsynteny; Malpighiales; Passiflora edulis; Populus trichocarpa; Manihot esculenta

LISTA DE FIGURAS

Figura 1. Fluxograma de identificação de regiões microssintênicas entre <i>P. edulis</i> , <i>P. trichocarpa</i> e <i>M. esculta</i> . 16
Figura 2. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 14) e M.esculenta (cromossomos 1 e 5).21
Figura 3. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 2) e M.esculenta (cromossomos 12 e 13).22
Figura 4. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 4 e 9) eM. esculenta cromossomos (15 e 17).23
Figura 5. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomos 4 e 9). 24
Figura 6. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 4 e 17)e M. esculenta (cromossomos 1 e 2).25
Figura 7. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 1) e M.esculenta (cromossomos 6 e 14).26
Figura 8. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 2 e 14)e M. esculenta (cromossomos 1 e 5).28
Figura 9. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 1) e M.esculenta (cromossomos 15 e 17).29
Figura 10. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 14) eM. esculenta (cromossomos 1 e 5).30
Figura 11. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 2 e 5)e M. esculenta (cromossomos 2 e 18).31
Figura 12. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 12 e15) e M. esculenta (cromossomos 6 e 14).32
Figura 13. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomo 7). 33
Figura 14. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 1) e M.esculenta (cromossomo 15).33
Figura 15. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomos 6 e 18).
Figura 16. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 6 e 18)e M. esculenta (cromossomos 3 e 15).35
Figura 17. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 4 e 9)e M. esculenta (cromossomo 4).36
Figura 18. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 12) eM. esculenta (cromossomo 1)37

Figura 19. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 2 e 5)e M. esculenta (cromossomo 18).38
Figura 20. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomo 5). 38
Figura 21. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>M. esculenta</i> (cromossomo 17). 39
Figura 22. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomos 4 e 9) e <i>M. esculenta</i> (cromossomo 4). 40
Figura 23. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomos 8 e 10) 41
Figura 24. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>M. esculenta</i> (cromossomos 6 e 14) 41
Figura 25. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomo 2). 42
Figura 26. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>M. esculenta</i> (cromossomo 13). 42
Figura 27. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 1 e 11)e M. esculenta (cromossomos 15 e 16).43
Figura 28. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>P. trichocarpa</i> (cromossomo 14) e <i>M. esculenta</i> (cromossomo 13). 44
Figura 29. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 4 e 9)e M. esculenta (cromossomo 4).45
Figura 30. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 1 e 3)e M. esculenta (cromossomos 4 e 11).46
Figura 31. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 8) e M.esculenta (cromossomos 1 e 5).47
Figura 32. Regiões microssintênicas colineares identificadas em <i>P. edulis</i> e <i>M. esculenta</i> (cromossomo 16). 48

Figura 33. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 12 e 15).

LISTA DE TABELAS

Tabela 1. Mapeamento genômico comparativo entre 28 sequências de *P. edulis* e o genoma de *P. trichocarpa*. 17
Tabela 2. Mapeamento genômico comparativo entre 25 sequências de *P. edulis* e o genoma de *M. esculenta*. 19
Tabela 3. Classificação dos elementos transponíveis identificados nos espaçamentos intergênicos de uma fração rica em genes de *P. edulis* comparativamente às regiões de *P. trichocarpa* e *M. esculenta*. 50

1. INTRODUÇÃO

A família Passifloraceae pertence à ordem Malpighiales e é membro do clado das Rosideas [1]. As passifloras ocorrem principalmente nas florestas tropicais e subtropicais dos Neotrópicos [2] e se caracterizam por serem trepadeiras ou lianas, além de arbustos e árvores [3]. A família possui cerca de 700 espécies e 16 gêneros [4], sendo o gênero *Passiflora* o mais rico, com cerca de 520 espécies descritas [2]. Dentre estas, 145 são de ocorrência no Brasil, sendo 83 endêmicas, o que coloca ao país em uma posição privilegiada em relação aos recursos genéticos do gênero [3].

Entre as passifloras, *Passiflora edulis*, comumente conhecida como maracujá azedo ou amarelo, é uma fruteira tropical, diploide (2n = 18), autoincompatível e, portanto, dependente da polinização cruzada, com flores muito atrativas a os polinizadores. O Brasil é considerado o maior produtor de maracujás [5, 6], com ~95% dos pomares cultivados com *P. edulis* [veja 7]. Basicamente, a produção brasileira tem razoável demanda no mercado nacional e está destinada, principalmente, para a elaboração de sucos e o consumo da fruta *in natura*. Por outro lado, uma parte reduzida da produção é direcionada para a exportação, sob a forma de suco concentrado, sendo os países europeus os principais importadores [veja 8]. No ano de 2017, a produção de maracujá azedo alcançou 554.598 toneladas, distinguindo-se a região nordeste como uma das principais produtoras [9].

Mesmo sendo uma planta usada para diversos fins industriais, produção de suco, fitoterápicos e cosméticos, é pouco estudada do ponto de vista genético e genômico, aspectos que podem auxiliar em pesquisas cujos objetivos são aumentar o rendimento e a qualidade dos frutos [10]. Em visto disso, nosso grupo tem gerado diversas informações científicas acerca do maracujá azedo [11, 12]. Os primeiros mapas genéticos de *P. edulis* foram estabelecidos com base em marcadores RAPD [13] e AFLP [14], o qual serviu para localizar locos quantitativos relacionados à resistência a *Xanthomonas axonopodis (Xap)*. Posteriormente, Oliveira e colaboradores obtiveram um mapa integrado, mais informativo, utilizando marcadores AFLP e SSR [15]. Além disso, foram identificados genes expressos em resposta à inoculação por *Xap*, destacando-se o gene da *lipoxygenase 2* que atua na defesa da planta [16]. Mais recentemente, Costa e colaboradores exploraram um conjunto de transcritos da biblioteca gerada por Munhoz et al. [16] e desenvolveram marcadores funcionais (SSRs e SNPs), os quais podem ser utilizados em estudos futuros [17].

Por outro lado, com o surgimento da Genômica como ciência, as tecnologias de sequenciamento do DNA tornaram-se ferramentas indispensáveis e, ainda, devido à necessidade de se obter uma maior cobertura genômica e em tempos mais curtos, certas estratégias para otimizar estas tecnologias foram desenvolvidas [18, 19]. Assim, ao longo dos anos, as plataformas de sequenciamento tornaram-se cada vez mais eficientes e accessíveis aos pesquisadores, por conta da melhoria na qualidade dos dados, aumento na velocidade de obtenção, e redução dos custos [20, 21]. No caso das plantas, até meados do 2019, foram depositadas no banco Phytozome (v12.1.6) sequências de dados genômicas de 82 espécies vegetais (https://phytozome.jgi.doe.gov/pz/portal.html). A disponibilidade de dados genômicos possibilitou o surgimento da Genômica comparativa que visa comparar dois ou mais genomas e estudar as semelhanças e diferenças relacionadas ao tamanho, presença de elementos transponíveis, número e organização dos genes, entre outros aspectos [22–25].

Estudos genômicos comparativos constituem uma abordagem importante para analisar informações biológicas relacionadas à organização, função e evolução dos genomas. Permitem também localizar regiões codificadores de proteínas e identificar segmentos conservados, possibilitando o estudo dos padrões evolutivos de espécies relacionadas [22, 26–29]. Pesquisas em genômica comparativa foram reportadas em diversas espécies vegetais, por exemplo, as realizadas entre *Brassica oleracea* e *Arabidopsis thaliana* [30], *Zea mays e A. thaliana* [31], *Fragaria* e outras espécies do clado das Rosideas [32], *Solanum lycopersicum* e *S. tuberosum* [33], *Oryza sativa* e *Ananas comosus* [34], *Glycine soja* e outras 9 espécies de *Glicine* [35], *Saccharum officinarum* e *S. spontaneum* [36], entre outras.

Um aspecto importante da genômica comparativa é a identificação de genes homólogos, que compartilham uma sequência de nucleotídeos ancestral comum. Esses são categorizados como ortólogos e parálogos [37–39]. Os genes ortólogos são separados por eventos de especiação [40]. Entretanto, a condição para serem considerados ortólogos é que as espécies comparadas compartilhem um gene ancestral. Além disso, uma particularidade dos ortólogos é que desempenham funções semelhantes nos genomas comparados, embora possam surgir variações na função do gene [41].

Genes parálogos são originados por eventos de duplicação no mesmo genoma, sendo que algumas delas são originadas nos primeiros estágios da evolução. A cópia do gene duplicado frequentemente adquire uma nova função [40].

A identificação de ortólogos e parálogos é essencial em análises evolutivas. A comparação de sequências possibilita conhecer a história evolutiva de uma família de genes ou do genoma de uma espécie [41, 42]. De igual forma, os termos sintenia e colinearidade são pontos chaves nas comparações de genomas aparentados. O conceito de sintenia refere-se à presença de genes ortólogos em cromossomos correspondentes nas espécies analisadas, enquanto a colinearidade refere-se à ordem conservada dos genes em uma determinada região cromossômica, sendo que ambos conceitos permitem uma melhor compreensão dos padrões evolutivos nos genomas analisados [43–45].

Outra estratégia usada pela genômica comparativa é o mapeamento genômico [46, 47] e consiste na identificação de genes homólogos e de variações estruturais, as quais são frequentes em plantas e constituem achados relevantes para elucidar a evolução dos genomas comparados [25]. Estas variações estruturais abrangem inversões, translocações, duplicações, inserções e deleções [25, 45, 48]. Estudos desenvolvidos em espécies vegetais reportaram que variações genômicas estruturais podem ser decorrentes de domesticação, processos metabólicos e catabólicos, tempo de floração, resistência a doenças, resistência ao estresse biótico e abiótico e tolerância a certos compostos [49].

Nas análises comparativas, um ponto frequentemente explorado é o estudo das sequências repetitivas, muito frequentes nos genomas dos eucariotos [22, 50]. Nas plantas, os elementos transponíveis (TEs) representam uma porção significativa do genoma [51–53]. Uma série de estudos indica que há uma relação entre o tamanho do genoma e a quantidade de TEs [52]. Por exemplo, no milho, os TEs ocupam mais de 85% do total do genoma [54].

Baseado no mecanismo de transposição, os TEs são divididos em classe I ou retrotransposons e classe II ou transposons de DNA [52, 53, 55, 56]. Segundo a classificação estabelecida por Wicker et al. (2007), a classe I está dividida em cinco ordens LTR (*Long Terminal Repeat*), DIRS (*Dictyostelium Intermediate Repeat Requence*), PLE (*Penelope-like Elements*), LINE (*Long Interspersed Nuclear element*) e SINE (*Short Interspersed Nuclear Element*) e a classe II está constituída por duas subclasses, a subclasse 1, composta pelas ordens TIR (*Terminal Inverted Repeat*) e *Cripton* e a subclasse 2 que é formada pelas ordens *Helitron* e *Maverick* [55].

Os retrotransposons são os TEs mais frequentes nos genomas de plantas, destacando-se a ordem LTR como a fração mais abundante [55, 57, 58]. Já os transposons de DNA abrangem pequenas frações nos genomas vegetais [55, 59]. Nos estudos comparativos as análises da dinâmica dos TEs podem ajudar no entendimento das mudanças evolutivas nos genomas de plantas [52].

No contexto das pesquisas genômicas, as bibliotecas inseridas em BACs (*Bacterial Artificial Chromosomes*) se tornaram amplamente utilizadas devido a sua eficiência em carregar grandes insertos de DNA (100 a 300 Kb) de forma estável, com baixo nível de recombinação, além de permitir a fácil manipulação e seleção dos clones [60–63]. Dessa forma, com o propósito de gerar informações sobre o genoma de *P. edulis*, nosso grupo de pesquisa construiu uma biblioteca genômica em BACs denominada Ped-B-Flav, a qual possui 82.944 clones (~108 kb/clone), com uma cobertura de ~6 vezes o genoma da espécie (~1.230 Mb) [64]. Esta biblioteca se encontra depositada no Centro Nacional de Recursos Genômicos Vegetais (CNRGV) do INRA (Toulouse, França) [11].

Santos et al. (2014) realizaram o primeiro estudo a partir dos dados da biblioteca (Ped-B-Flav) de *P. edulis*. Baseada na metodologia *BAC-end sequences* (BES), foram sequenciados 5.058 insertos nas duas direções (*forward* e *reverse*) e 916 insertos apenas em uma direção, gerando um total de 11.032 insertos sequenciados, e desses, foram exploradas ~10.000 BES de alta qualidade. Em 19,6% das BES foram identificados elementos repetitivos, além de repetições de sequências simples, sequências proteicas (atribuindo-lhes termos ontológicos) e regiões sintênicas com os genomas de *Arabidopsis thaliana, Populus trichocarpa* e *Vitis vinifera*. A primeira espécie foi escolhida por ser o genoma modelo em plantas e as duas últimas por possuírem o maior número de *top hits* após efetuarem-se buscas em banco de dados. Os resultados de mapeamento comparativo, revelaram que 106, 206, 367 regiões de potencial sintenia foram encontradas entre as BES de *P. edulis* e regiões do genoma de *A. thaliana, V. vinifera* e *P. trichocarpa*, respectivamente [11].

Posteriormente, a partir da plataforma de sequenciamento *PacBio* (que produz *reads* longas e de alta qualidade), foram sequenciados mais de 100 insertos dessa biblioteca [12]. A seleção desses insertos baseou-se inicialmente nos resultados do mapeamento comparativo realizado por Santos e colaboradores [11]. O segundo grupo de insertos foi selecionado a partir de *screenings* da biblioteca com sondas homológas aos transcritos de *P. edulis* descritos por Munhoz e colaboradores [16]. Assim, no total, foram selecionados para o sequenciamento 112 insertos de BACs. Em seguida, no momento da montagem destas sequências, foram descobertos dois casos de sobreposição, que resultaram em segmentos únicos. Além disso, três insertos de BACs correspondiam a DNA cloroplastidial, sendo que este achado fez com que Cauz-Santos e colaboradores [65] obtivessem o genoma cloroplastidial de *P. edulis*. Sendo assim, esses cinco insertos não foram incluídos, resultando em um conjunto de 107 insertos, que representam 10.4 Mb de dados (~1% do genoma). Com base nesse conjunto de dados,

identificaram-se ~1.900 genes, elementos repetitivos e regiões microssintênicas entre *P. edulis* e duas espécies de Malpighiales [12].

Assim, a partir dos dados gerados por Munhoz et al. [12] e com o intuito de dar continuidade aos estudos genômicos comparativos, um conjunto de 35 insertos BACs de *P. edulis*, rico em genes, foram comparados com as sequências de duas Malpighiales, filogeneticamente próximas, e cujos genomas estão disponíveis e bem anotados: *Populus trichocarpa* (2n=38) e *Manihot esculenta* (2n=36). Dessa forma, neste trabalho, pretende-se esclarecer a questão de qual das duas espécies comparadas possui maior grau de sintenia com *P. edulis*, além de estabelecer uma metodologia para a identificação de regiões microssintênicas que inclui a localização de genes ortólogos e dos elementos transponíveis nos espaçamentos intergênicos. Finalmente, através das informações geradas, pretende-se abrir o caminho para que novos estudos sejam realizados e facilitem o entendimento da evolução de *P. edulis* e sua relação com outras espécies filogeneticamente próximas.

2. MATERIAL E MÉTODOS

2.1. Seleção das regiões genômicas de *P. edulis* a partir do conjunto de insertos sequenciados

A partir da caracterização de 107 insertos de BACs provenientes da Biblioteca Ped-B-Flav, aproximadamente 1.900 genes foram anotados [12]. No presente estudo, foram selecionados 35 insertos de BACs que apresentaram o maior número de genes (mínimo 21 e máximo 36 genes) para serem comparados com as respectivas regiões dos genomas de *Populus trichocarpa* e *Manihot esculenta*, com o objetivo de identificar regiões microssintênicas entre esses genomas. Ambas são Malpighiales próximas de *P. edulis* e cujos genomas estão depositados no Phytozome [66].

2.2. Análise de microssintenia

Com a finalidade de identificar regiões microssintênicas, foram efetuados alinhamentos usando a ferramenta BLASTN das sequências gênicas de *P. edulis* (sequência query) contra as sequências genômicas *de P. trichocarpa* e *M. esculenta* (sequências subject) utilizando o banco de dados do Phytozome [66]. Apenas os alinhamentos com e-value de 10^{-10} e uma cobertura maior do 50% foram considerados. As localizações exatas do início e do final das possíveis regiões microssintênicas foram registradas em uma planilha excel.

Assim, com base nessas informações, utilizando a ferramenta JBrowse, de acesso livre na plataforma Phytozome v12.1.6 [66], se procedeu à identificação dos genes presentes em cada uma dessas regiões. Por comparações par a par, foram considerados microssintênicos [67] os blocos cromossômicos que apresentaram sintenia e colinearidade, além de incluírem pelo menos quatro genes ortólogos. Para tal, foram usadas tabelas comparativas e foi realizada a curadoria manual. Cada uma das regiões microssintênicas identificadas foram visualizadas usando o software Genome View [68]. Os gráficos comparativos foram construídos utilizando o CorelDRAW® Graphics Suite 2018 (Copyrigth© Corel Corporation, 2018). As etapas que compõem o processo de identificação de regiões microssintênicas estão resumidas na figura 1.



Figura 1. Fluxograma de identificação de regiões microssintênicas entre P. edulis, P. trichocarpa e M. esculenta.

2.3. Identificação dos elementos transponíveis

Os elementos móveis identificados nos espaçamentos intergênicos das três espécies foram analisados com base nos TEs (*Transposable Elements*) de *P. edulis*, previamente identificados [69], bem como o pipeline REPET [70, 71], cujas etapas são: auto alinhamento do genoma, agrupamento de sequências, criação de um consenso e anotação. Em linhas gerais, cada genoma é comparado com ele mesmo através de um alinhamento que identifica as regiões repetidas. Posteriormente, essas sequências repetidas são agrupadas em famílias. Na etapa seguinte, é criado um consenso para cada família e, por fim, é realizada a anotação dos TEs pela comparação com o banco de dados.

De *P. trichocarpa* e *M. esculenta*, foram utilizados os genomas completos. Após a identificação, as sequências dos TEs foram mapeadas contra as sequências das regiões microssintênicas previamente identificadas. Para o mapeamento foi utilizado o programa BWA-MEN [72]. As posições dos TEs foram confirmadas pela visualização no Genome View [68]. As posições dos TEs também foram anotadas e incluídas nos gráficos.

3. RESULTADOS

3.1. Seleção das regiões genômicas de P. edulis

Os 35 insertos selecionados para as análises microssintênicas, com os respectivos números de genes entre parêntesis, foram os seguintes: Pe101K14 + 141H13 (36), Pe185D11 (36), 164B18 (29), Pe214H11 (29), Pe164D9 (28), Pe186E19 (28), Pe43L2 (27), Pe164K17 (26), 215I8 (26), Pe84I14 (25), Pe84M23 (25), Pe93M2 (25), Pe171P13 (25), Pe207D11 (25), Pe93N7 (24), Pe108C16 (24), Pe173B16 (24), Pe185J16 (24), Pe198H23 (24), Pe212I1 (24), Pe93J9 (23), Pe135J12 (23), Pe195F4 (23), Pe74I6 (22), Pe84M18 (22), Pe101O4 (22), Pe141J23 (22), Pe201C11 (22), Pe69G18 (21), Pe69H24 (21), Pe93K19 (21), Pe125I23 (21), Pe164A12 (21), Pe168B17 (21) e Pe214A18 (21) (Apêncide A).

3.2. Análise de microssintenia

Dos 35 insertos selecionados de *P. edulis*, 32 apresentaram regiões microssintênicas com os genomas comparados. Os insertos Pe43L2, Pe198H23 e Pe168B17 não mostraram regiões microssintênicas com *P. trichocarpa* e *M. esculenta*.

Os 32 insertos de *P. edulis* com regiões microssintênicas representam um total de 3.518.244 pb (3.5 Mb). Deles, 28 insertos contêm regiões microssintênicas com *P. trichocarpa* abrangendo 2.573.530 pb (2.6 Mb) com 581 genes. As regiões microssintênicas localizadas em *P. trichocarpa* abrangem 10.496.396 pb (10.5 Mb) havendo um total de 1.405 genes, dos quais 747 genes são ortólogos. O tamanho médio dos genes em *P. edulis* foi 2.796 pb, relativamente inferior ao tamanho médio dos genes encontrados em *P. trichocarpa* com 3.310 pb. Os dados relacionados ao espaçamento intergênico revelaram que *P. edulis* apresenta em média um tamanho de 2.079 pb, em comparação com *P. trichocarpa* que apresenta 7.483 pb (Tabela 1). Além disso, 16 regiões cromossômicas de *P. edulis* apresentam genes ortólogos duplicados em *P. trichocarpa*.

Tabela 1. Mapeamento genômico comparativo entre 28 sequências de P. edulis e o genoma de P. trichocarpa.

(Continua)

		P. ed	ulis			P. trichocarpa					
Código do BAC	Tamanho do inserto (pb)	Tamanho da região sintênica (pb)	N° de genes na região sintênica	Tamanho médio dos genes (pb)	Tamanho médio do espaço intergênico (pb)	Chr	Tamanho da região sintênica (pb)	N° de genes na região sintênica (pb)	Tamanho médio dos genes (pb)	Tamanho médio do espaço intergênico (pb)	N° de genes ortólogos
Pe101K14- 141H13	172.337	159.949	32	2.923	2.157	14	213.942	31	4.055	2.962	12
Pe185D11	119.061	110.316	33	1.538	1.862	2	253.596	34	2.691	5.280	22
Do164D19	104 102	103.945	29	2.313	1.317	4	369.800	45	2.996	6.058	20
Pe104D18	104.102	103.945	29	2.313	1.317	9	189.230	42	2.345	2.214	18
Pe214H11	142 456	64.482	13	3.675	1.392	4	248.247	32	2,234	5.702	14
	142.456 -	62.181	13	3.950	904	4	222.504	37	2.630	3.745	11

		79.416	16	3.781	1.361	9	221.003	36	2.571	3.834	17
		60.720	12	3.950	968	9	202.191	38	3.039	2.650	13
		80.789	24	1.955	1.559	4	430.901	38	2.442	9.139	27
Pe164D9	93.527	85.112	26	1.951	1.451	17	209.253	38	2.320	3.855	26
		92.977	20	2.853	1.939	1	268.117	16	5.753	12.808	8
Pe186E19	115.218	17.442	7	3.554	893	1	27.583	5	2.625	3.615	5
	110 50 1	110.607	25	3.103	1.782	2	307.065	38	3.510	5.183	16
Pe164K17	113.504	113.313	26	3.037	1.762	14	332.637	46	3.834	3.584	23
Pe215I8	129.737	79.415	15	2.774	2.950	1	166.694	18	3.190	6.428	12
Pe84I14	97.848	93.065	23	3.114	976	14	141.647	20	4.958	6.435	13
D 041402	02.017	92.795	25	2.777	975	2	171.100	23	3.093	4.545	15
Pe84M23	93.217	89.339	24	2.820	943	5	206.947	19	3.013	8.693	12
D.02142	100.426	98.828	25	2.245	5.735	12	199.350	38	2.536	9.887	17
Pe93M2	100.436	88.334	23	2.269	5.982	15	207.961	33	3.192	10.369	18
Pe171P13	111.123	85.810	19	2.606	2.365	7	346.104	47	3.169	34.428	12
Pe207D11	111.690	31.090	9	3.019	723	1	122.497	12	3.600	13.387	8
D-02N7	100.000	105.007	24	3.121	1.423	6	340.655	46	3.745	11.582	23
Pe93N/	100.908	99.896	23	3.035	1.490	18	337.287	33	3.358	20.496	16
D 100016	06 752	68.880	18	2.280	1.640	6	137.749	23	3.747	5.188	16
Pe108C16	96.753	65.309	17	2.286	1.760	18	130.229	21	2.689	7.942	13
D-172D1(100 201	105,875	24	3.079	1.498	4	409.775	54	3.384	13.166	28
Pe1/3B10 109.801	109.801	105.875	24	3.079	1.498	9	166.729	37	2.809	3.315	29
Pe185J16	103.095	47.587	11	2.693	2.149	12	231.419	14	3.736	21.564	10
D-21211	101 204	85.114	18	3.138	1.882	2	162.212	25	4.547	11.759	14
F621211	121.364	85.114	18	3.138	1.882	5	169.126	27	4.056	9.702	13
Pe93J9	107.024	98.303	21	2.858	1.916	5	325.071	60	2.573	3.163	18
Pe105E/	113 //3	110.840	23	2.844	2.210	4	177.213	20	3.651	5.486	19
1019514	115.445	110.840	23	2.844	2.210	9	180.144	30	3.085	3.241	22
Do7/16	112.038	110.879	21	3.531	1.838	8	229.123	29	3.337	4.729	18
1 67410	112.058	110.879	21	3.531	1.838	10	421.891	52	3.016	6.153	26
Pe 101O4	117.581	93.089	17	2.420	3.249	2	216.603	15	5.112	11.000	7
Pe201C11	140 216	89.677	16	2.240	3.591	1	90.883	11	3.210	7.128	7
10201011	140.210	89.677	16	2.240	3.591	11	79.079	12	2.599	4.356	6
Pe69G18	91.665	80.375	17	3.219	1.605	14	236.747	28	3.243	6.132	13
Pe60H2/	10/ 185	101.965	20	2.815	2.405	4	200.595	26	3.289	4.984	18
10071124	104.105	103.909	21	2.753	2.307	9	163.737	25	3.468	3.212	20
Pe93K19	99 992	99.519	21	3.524	1.424	1	261.602	27	3.962	6.573	16
10/51(1)	<i>)).))2</i>	99.519	21	3.524	1.424	3	341.099	37	3.305	6.267	22
Pe125I23	96 568	42.041	9	2.648	2.278	8	75.424	10	3.576	6.151	9
1 0120120	20.200	39.779	10	2.219	2.611	8	53.334	7	3.362	4.968	6
Pe214418	106 977	106.404	21	1.945	3.773	12	139.481	22	3.398	3.179	10
10217/110	100.777	106.404	21	1.945	3.773	15	161.360	28	2.837	2.938	9
Total	3.131.946	2.573.530*	581*	2.796*	2.079*		10.496.936	1.405	3.310	7.483	747
	Chan Char	*	. Come no	dundânaia							

•Chr: Chromossomo; *: Sem redundância

No caso das comparações com *M. esculenta*, os 25 insertos de *P. edulis* com microssintenia abrangem 2.313.012 pb (2.3 Mb) com 546 genes. As regiões microssintênicas correspondentes a *M. esculenta* abrangem um total 7.846.432 pb (7.9 Mb) com 931 genes, dos quais 543 genes são ortólogos. No que se refere, ao tamanho médio dos genes, os resultados mostram o mesmo padrão que aquele encontrado nas comparações com *P. trichocarpa*. Por tanto, o tamanho médio dos genes em *P. edulis* é também inferior (2.677 pb) relativamente ao de *M. esculenta* (3.739 pb). O tamanho médio do espaçamento intergênico de *M. esculenta* (6.534 pb) é aproximadamente três vezes maior que o encontrado em *P. edulis* (1.974 pb) (Tabela 2). Além disso, 15 das regiões microssintênicas de *P. edulis* estão duplicadas nos cromossomos de *M. esculenta*.

Tabela 2. Mapeamento genômico comparativo entre 25 sequências de P. edulis e o genoma de M. esculenta.

										(Continua)	
		P. edi	ılis					N	1. esculenta		
Código do BAC	Tamanho do inserto (pb)	Tamanho da região sintênica (pb)	N° de genes na região sintênica	Tamanho médio dos genes (pb)	Tamanho médio do espaço intergênico (pb)	Chr	Tamanho da região sintênica (pb)	N° de genes na região sintênica (pb)	Tamanho médio dos genes (pb)	Tamanho médio do espaço intergênico (pb)	N° de genes ortólogos
Pe101K14-	170 227	170.391	36	2.721	2.072	1	183.133	29	3.466	3.081	16
141H13	172.337	164.887	35	2.673	2.100	5	259.161	33	4.516	3.884	14
De195D11	110.061	110.279	33	1.613	1.785	12	317.886	43	2.362	5.330	27
PerosDII	119.001	112.597	34	1.602	1.764	13	254.296	29	2.584	6.216	18
Do164D19	104 102	103.945	29	2.314	1.318	15	345.243	35	4.197	5.835	12
Pe104B18	104.102	103.945	29	2.314	1.318	17	182.720	38	2.430	2.525	17
Do164D0	02 527	85.112	26	1.951	1.452	1	118.187	21	2.636	3.144	15
Pe164D9 93.52	95.527	93.489	28	1.869	1.601	2	206.242	38	2,389	3.122	25
Pe186E19 115	115 019	50.679	16	2.224	1.110	6	101.361	12	4.021	4.828	8
	115.218	50.679	16	2.224	1.110	14	304.339	14	5.566	17.417	9
	112 504	101.996	22	3.149	1.923	1	189.788	23	3.879	7.638	11
Pe104K1/	115.504	110.607	25	3.104	1.782	5	393.258	38	4.822	6.328	17
D-21519	120 727	124.698	26	3.008	2.169	15	193,725	24	4,418	3,815	14
Pe21518	129.737	118.786	22	326	2.382	17	162.363	19	3.787	5.963	14
D 04114	07.040	96.433	24	3.092	969	1	135.657	19	5.005	6.468	12
Pe84114	97.848	94.441	24	3.033	944	5	211.686	26	4.890	9.873	14
D 041422	02 017	53.520	15	2.678	956	2	137.511	19	2.670	5.214	8
Pe84M23	93.217	81.630	17	2.865	1.126	18	148.682	18	3.680	5.096	13
D.02142	100.426	98.828	25	2.245	5.735	6	126.299	22	2.782	5.921	17
Pe93M2	100.436	78.587	19	2.256	2.198	14	151.939	26	3.883	7.350	12
Pe207D11	111.690	28.902	7	3.712	1.450	15	48.780	11	4.100	4.838	6
D 100016	06 750	68.880	18	2.280	1.640	3	76.043	14	3.672	4.372	10
Pe108C16	96./53	63.474	16	2.379	1,696	16	88.458	15	3.432	6.018	10
Pe173B16	109.801	92.992	23	3.094	1.087	4	235.649	23	4.748	9.462	20
Pe185J16	103.095	88.563	21	2.351	2.040	1	308.705	27	6.736	20.718	12

Pe212I1 121.384 85.114 19 3.138 1.882 18 172.143 23 4.370	11.733 14	
Pe135J12 103.564 93.675 22 2.451 1.895 17 236.805 21 3.010	9.033 20	
Pe195F4 113.443 106.357 21 3.000 2.333 4 321.722 18 3.789	14.966 14	
99.442 21 2.533 2.444 6 449.687 34 5.061	8.678 13	
92.535 19 2.652 2.491 14 125.428 17 3.616	4.016 9	
Pe141J23 95.795 95.715 22 2.568 1.870 13 221.517 20 4.322	7.765 10	
B-201C11 140.216 89.677 16 2.240 3.591 15 44.656 10 2.509	2.540 5	
Pe201C11 140.210 89.677 16 2.240 3.591 16 56.878 9 2.312	4.510 4	
Pe69G18 91.665 46.056 12 2.820 1.112 13 84.501 12 4.424	2.858 10	
Pe69H24 104.185 96.413 21 2.754 2.177 4 288.050 31 3.717	5.763 8	
99.519 21 3.524 1.424 4 213.242 29 3.227	4.276 14	
99.992 95.571 20 3.661 1.321 11 192.485 23 4.927	3.776 20	
95.477 21 2.527 2.407 1 187.380 29 3.111	4.072 18	
Pe125I23 96.568 39.779 10 2.219 2.611 5 61.847 11 1.921	4.074 5	
37.944 8 2.548 2.510 5 76.788 7 2.870	9.452 7	
Pe164A12 82.998 82.933 21 2.355 1.676 16 232.192 27 3.439	5.321 18	
Total 2.714.077 2.313.012* 546* 2.677* 1.974* 7.846.432 931 3.739	6.534 543	3

•Chr: Chromossomo; *: Sem redundância

A seguir, as figuras apresentadas obedecem a uma ordem a partir do número máximo até o mínimo de genes por inserto BAC de *P. edulis* e as respectivas microssintenias identificadas nas comparações com *P. trichocarpa* e *M. esculenta*.

Dois insertos de *P. edulis* apresentaram 36 genes. O inserto Pe101K14-141H13 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 14) e *M. esculenta* (cromossomos 1 e 5). O tamanho da região de *P. edulis* foi ligeiramente menor e não apresenta elementos móveis nas comparações com ambas as espécies. Na comparação com *P. trichocarpa*, a região do cromossomo 14 contêm 12 genes ortólogos além de dois TEs, pertencentes às ordens LTR (1) e TIR (1). Nas comparações com *M. esculenta*, há 11 genes ortólogos de *P. edulis* que estão duplicados e há três genes, localizados no final, que se encontram rearranjados. A região microssintênica entre *P. edulis* e a o cromossomo 1 de *M. esculenta* contêm 16 genes ortólogos e há dois elementos móveis das ordens LTR (1) e LINE (1). Esta região apresenta orientação oposta relativamente à de *P. edulis*. No caso da região microssintênica entre *P. edulis* e o cromossomo 5 foram identificados 14 ortólogos e seis elementos móveis, LTR (3) e TIR (3), no cromossomo 5 (Figura 2).



Figura 2. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 14) e *M. esculenta* (cromossomos 1 e 5).

O inserto Pe185D11 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 2) e *M. esculenta* (cromossomos 12 e 13). Em ambas as comparações, a região de *P. edulis* possui orientação oposta e um elemento móvel da ordem LTR. Na comparação com *P. trichocarpa*, a região do cromossomo 2 é aproximadamente 2x maior que a de *P. edulis*. Há 22 genes ortólogos e foi detectado um elemento móvel (Helitron) em *P. trichocarpa*. Já nas comparações com *M. esculenta*, 14 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 12 é cerca de 3x maior que a de *P. edulis*. Há 27 genes ortólogos, sendo esta a região que contém o maior número de ortólogos nas comparações com *M. esculenta*. Além disso, no cromossomo 12, foram identificados oito elementos móveis das ordens LTR (6), LINE (1) e SINE (1). No segundo caso, a região do cromossomo 13 é aproximadamente 2x maior. Foram identificados 18 genes ortólogos, 17 elementos móveis das ordens LTR (13), DIRS (1), LINE (1) e TIR (2) no cromossomo 13. Observe que o gene 18 de *P. edulis* possui duas cópias no cromossomo 12 e três no 13, havendo um total de cinco cópias em *M. esculenta* (Figura 3). Considerando todas as comparações, as duplicações gênicas encontradas em *M. esculenta* abrangem um mínimo de duas até o máximo de cinco cópias.



22



Figura 3. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 2) e *M. esculenta* (cromossomos 12 e 13).

Dois insertos de *P. edulis* contem 29 genes. O inserto Pe164B18 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 4 e 9) e *M. esculenta* cromossomos (15 e 17). A região de *P. edulis* possui genes ortólogos duplicados e dois elementos móveis das ordens LTR (1) e LINE (1) nas comparações com ambas as espécies. Nas comparações com *P. trichocarpa*, 14 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 4 é aproximadamente 3x maior que a de *P. edulis*. Há 20 genes ortólogos e nove elementos móveis das ordens LTR (3), TIR (5) e Helitron (1) na região do cromossomo 4. A região do cromossomo 9 apresenta um tamanho ligeiramente maior que o da região de *P. edulis*. Além disso, no cromossomo 9 foram identificados 18 genes ortólogos e um elemento móvel da ordem TIR. Já nas comparações com *M. eculenta*, sete genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 15 é cerca de 3x maior que a de *P. edulis*, sendo identificados 12 ortólogos e 13 elementos móveis das ordens LTR (9), DIRS (1), LINE (1) e TIR (2). O tamanho da região do cromossomo 17 foi ligeiramente maior que o da região de *P. edulis*.

ortólogos e há cinco elementos móveis das ordens LTR (4) e LINE (1) identificados no cromossomo 17. Distinguem-se quatro cópias em tandem na região de *P. edulis* do gene que codifica a proteína de ligação ao ácido salicílico tipo 2 (genes 9-12) com uma cópia no cromossomo 4 e três no cromossomo 9 de *P. trichocarpa* e uma cópia no cromossomo 17 de *M. esculenta*. Além disso, foram encontradas três cópias em *P. edulis* do gene que codifica a proteína CYP82D47 semelhante à do citocromo P450 (genes 17, 19 e 20) com duas cópias em *P. trichocarpa* (cromossomos 4 e 9) e *M. esculenta* (cromossomo 17) (Figura 4).



Figura 4. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 4 e 9) e *M. esculenta* cromossomos (15 e 17).

O inserto Pe214H11 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 4 e 9). A região de *P. edulis* possui orientação oposta, 21 genes ortólogos duplicados, não apresenta elementos móveis e interessantemente mostra microssintenia com dois segmentos separados por ~1,4 Mb no cromossomo 4

e ~1,2 Mb no cromossomo 9. Os segmentos do cromossomo 4 são aproximadamente 4x maiores que as respectivas regiões microssintênicas em *P. edulis*. A região microssintênica que envolve o 1° segmento do cromossomo 4 contêm 14 genes ortólogos e há cinco elementos móveis das ordens LTR (4) e SINE (1). Comparando-se o 2° e a região de *P. edulis*, foram identificados 11 genes ortólogos e há 12 elementos móveis das ordens LTR (9), DIRS (2) e LINE (1). No caso das comparações com o cromossomo 9, cada segmento é quase 3x maior que as respectivas regiões microssintênicas em *P. edulis*. A região microssintênica que envolve o 1° segmento do cromossomo 9 contêm 17 genes ortólogos e há seis elementos móveis das ordens LTR (1), DIRS (4) e SINE (1). Comparando-se o 2° segmento e a região de *P. edulis*, foram identificados 11 genes ortólogos e há três elementos móveis da ordem LTR. Observe que o gene 9 de *P. edulis* possui duas cópias no cromossomo 4 (2° segmento) e quatro cópias no cromossomo 9 (2° segmento). Além disso, os genes 15 e 23 de *P. edulis* têm duas e quatro cópias, respectivamente, no cromossomo 9 (1° segmento) (Figura 5).



Figura 5. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 4 e 9).

Dois insertos de *P. edulis* apresentaram 28 genes. O inserto Pe164D9 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 4 e 17) e *M. esculenta* (cromossomos 1 e 2). Nas comparações com *P. trichocarpa*, 16 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 4 é cerca de 5x maior que a de *P. edulis* e apresenta orientação oposta. Há 27 genes ortólogos e 29 elementos móveis das ordens LTR (16), DIRS (1) e TIR (13) no cromossomo 4. A região do cromossomo 17 é 2x maior que a de *P. edulis*. Há 26 genes ortólogos e sete elementos móveis das ordens LTR (1), LINE (3) e TIR (3) no cromossomo 17. Já nas comparações com *M. esculenta*, 12 genes ortólogos de *P. edulis* estão duplicados. Há 15 genes ortólogos e apenas um elemento móvel da ordem LTR no cromossomo 1. O tamanho da região do

cromossomo 2 é aproximadamente 2x maior que o da região de *P. edulis*. Há 25 genes ortólogos e sete elementos móveis das ordens LTR (5), TIR (1) e Helitron (1) no cromossomo 2. Observe que há oito cópias do gene 12 que codifica uma proteína induzida por estresse em *P. edulis* no cromossomo 4 e três cópias no cromossomo 17 de *P. trichocarpa*, entretanto em *M. esculenta* apenas foi encontrado um ortólogos no cromossomo 2 (Figura 6).



Figura 6. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 4 e 17) e *M. esculenta* (cromossomos 1 e 2).

O inserto Pe186E19 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 1) e *M. esculenta* (cromossomos 6 e 14). Na comparação com *P. trichocarpa*, há um segmento com cinco genes que se encontra translocado comparativamente a sua posição em *P. edulis* no cromossomo 1 de *P. trichocarpa* que é cerca de 3x maior, sendo identificados oito genes ortólogos e três elementos móveis das ordens LTR (2) e

TIR (1). O segmento translocado tem um tamanho ligeiramente maior o que de *P. edulis* havendo aí cinco genes ortólogos. Nas comparações com *M. esculenta*, cinco genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 6 possui orientação oposta e é 2x maior que a de *P. edulis*. Há oito ortólogos e quatro elementos móveis da ordem LTR no cromossomo 6. O tamanho do cromossomo 14 foi aproximadamente 6x maior que a de *P. edulis*. Foram identificados nove genes ortólogos, 14 elementos móveis da ordem LTR no cromossomo 14 foi aproximadamente 6x maior que a de *P. edulis*. Foram identificados nove genes ortólogos, 14 elementos móveis da ordem LTR no cromossomo 14 foi aproximadamente 6x maior que a de *P. edulis*. Foram identificados nove genes ortólogos, 14 elementos móveis da ordem LTR no cromossomo 14 (Figura 7).



Figura 7. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 1) e *M. esculenta* (cromossomos 6 e 14).

Dois insertos de *P. edulis* apresentaram 26 genes. O inserto Pe164K17 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 2 e 14) e *M. esculenta* (cromossomos 1 e 5). Em ambas as comparações, a região de *P. edulis* não apresenta elementos móveis. Nas comparações com *P. trichocarpa*, 14 genes de *P. edulis* estão duplicados. As regiões cromossômicas 2 e 4 são 3x maiores que a de *P. edulis* estão duplicados.

edulis. A primeira região contêm 16 genes ortólogos e cinco TEs da ordem LTR. Além disso, observe que os genes 8 e 9 de *P. edulis* encontram-se invertidos em relação aos seus ortólogos. A segunda região contêm 23 genes ortólogos e três TEs da ordem LTR. Observe que o gene 25 que codifica a proteína citocromo P450 em *P. edulis* possui duas cópias no cromossomo 14. Já nas comparações com *M. esculenta*, a região de *P. edulis* possui orientação oposta e há oito genes ortólogos duplicados. Há 11 genes ortólogos e sete TEs das ordens LTR (5), DIRS (1) e LINE (1) na região do cromossomo 1. Duas cópias do gene que codifica a proteína LOB (genes 12 e 13) foram encontradas em *P. edulis*, com um ortólogo no cromossomo 1. A região do cromossomo 5 é cerca de 3x maior que a de *P. edulis*, onde foram identificados 17 genes ortólogos e 11 TEs das ordens LTR (7) e TIR (4) (Figura 8).



Figura 8. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 2 e 14) e *M. esculenta* (cromossomos 1 e 5).

O inserto Pe215I8 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 1) e *M. esculenta* (cromossomos 15 e 17). A região de *P. edulis* apresenta um TE da ordem LTR. Na comparação com *P. trichocarpa*, a região do cromossomo 1 é aproximadamente 2x maior que a de *P. edulis*. Há 12 genes ortólogos e dois TEs das ordens LTR (1) e DIRS (1) no cromossomo 1. Duas cópias do gene que codifica a proteína 4-cumarato ligase (genes 6 e 7) em *P. edulis* possuem um ortólogo no cromossomo 1; além disso, o gene 13 que codifica a proteína transportadora de histidina e lisina possui duas cópias no cromossomo 1. Nas comparações com *M. esculenta*, 11 genes ortólogos de *P. edulis* estão duplicados. Os tamanhos das regiões cromossômicas 15 e 17 foram ligeiramente maiores que os respectivos tamanhos em *P. edulis*. A região do cromossomo 15 possui orientação oposta, foram identificados 14 genes ortólogos e seis TEs das ordens LTR (4) e TIR (2). Observe que o gene 14 de *P. edulis* possui duas cópias no cromossomo 15 e 17 (Figura 9).



Figura 9. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 1) e *M. esculenta* (cromossomos 15 e 17).

Cinco insertos de *P. edulis* apresentaram 25 genes. O inserto Pe84I14 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 14) e *M. esculenta* (cromossomos 1 e 5). Em *P. edulis* há um TE da ordem LTR. Na comparação com *P. trichocarpa*, a região do cromossomo 14 é um pouco maior que a de *P. edulis*. Há 13 genes ortólogos e dois TEs das ordens LTR (1) e TIR (1) no cromossomo 14. Nas comparações com *M. esculenta*, um ortólogo de *P. edulis* está duplicado. A região do cromossomo 1 possui orientação oposta e é ligeiramente maior que a de *P. edulis* e nela foram identificados 12 genes ortólogos e dois TEs das ordens LTR (1) e LINE (1). A região do cromossomo 5 é aproximadamente 2x maior que a região de *P. edulis*. Há 14 genes ortólogos e quatro TEs das ordens LTR (3) e TIR (1) no cromossomo 5. Observe que o gene 19 de *P. edulis* que codifica uma proteína cloroplastidial possui três cópias no cromossomo 1 e apenas uma no 5 (Figura 10).



Figura 10. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 14) e *M. esculenta* (cromossomos 1 e 5).

O inserto Pe84M23 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 2 e 5) e *M. esculenta* (cromossomos 2 e 18). Em ambas as comparações, a região de *P. edulis* apresenta orientação oposta, não possui TEs e segmentos rearranjados são observados. Nas comparações com *P. trichocarpa*, oito genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 2 é cerca de 2x maior que a de *P. edulis*. Há 15 genes ortólogos e dois TEs das ordens LTR (1) e Helitron (1) no cromossomo 2. Note que há uma inversão envolvendo dois genes de *P. edulis* nas comparações com o cromossomo 2 tanto de *P. trichocarpa* como de *M. esculenta*. A região do cromossomo 5 é aproximadamente 2x maior que a de *P. edulis* e nela foram identificados 12 genes ortólogos e seis TEs das ordens LTR (2), DIRS (1) e Helitron (3). Também foi identificada uma inversão envolvendo dois genes de *P. edulis* na comparação com o cromossomo 5. Observe que o gene 23 que codifica uma proteína transportadora de açúcares possui apenas um ortólogo no cromossomo 5 e três cópias no 2, e o gene 10 que codifica uma poligalacturonase em *P. edulis* tem duas cópias no cromossomo 5.

Já nas comparações com *M. esculenta*, cinco genes ortólogos de *P. edulis* estão duplicados. As regiões cromossômicas 2 e 18 são maiores (2x) que a de *P. edulis*. No cromossomo 2 há oito ortólogos e sete TEs das

ordens LTR (5), LINE (1) e TIR (1). No caso do cromossomo 18, foram identificados 13 ortólogos e um TE da ordem TIR. Note que o gene 13 que codifica a proteina álcool desidrogenase em *P. edulis* tem quatro cópias no18 (Figura 11).



Figura 11. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 2 e 5) e *M. esculenta* (cromossomos 2 e 18).

O inserto Pe93M2 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 12 e 15) e *M. esculenta* (cromossomos 6 e 14). Em *P. edulis* há genes ortólogos duplicados, dois TEs das ordens LTR (1) e TIR (1) e rearranjos em ambas as comparações. Nas comparações com *P. trichocarpa*, nove genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 12 é cerca de 2x maior que a de *P. edulis* e nela há 17 genes ortólogos e cinco TEs das ordens LTR (3) e LINE (2). A região do cromossomo 15 é aproximadamente 2x maior que a de *P. edulis* e nela foram identificados 18 genes ortólogos e sete TEs das ordens LTR (6) e TIR (1). Observe que os genes 8 e 9 de *P. edulis* estão invertidos nas comparações com *P.*

trichocarpa; além disso, o gene 8 que codifica a proteína da ciclina B em *P. edulis* possui duas cópias no cromossomo 12 e três no 15.

Nas comparações com *M. esculenta*, sete genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 6 é cerca de 2x maior que a região de *P. edulis* e nela foram identificados 17 genes ortólogos e três TEs das ordens LTR (1) e TIR (2). A região do cromossomo 14 possui orientação oposta e o seu tamanho é aproximadamente 2x maior que a de *P. edulis* e nela há 12 genes ortólogos. Observe a inversão dos genes 8 e 9 de *P. edulis* relativamente à posição no cromossomo 6 de *M. esculenta* (Figura 12).



Figura 12. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 12 e 15) e *M. esculenta* (cromossomos 6 e 14).

O inserto Pe171P13 apresenta sintenia colinear apenas com *P. trichocarpa* (cromossomo 7). Observe que um segmento cromossômico de *P. trichocarpa* de 186.404 pb está ausente em *P. edulis*, devido a rearranjos, possivelmente relacionados a eventos de deleção em *P. edulis*. Há dois TEs da ordem LTR em *P. edulis* cuja região é aproximadamente 4x menor que a do cromossomo 7. Foram identificados 12 genes ortólogos e há nove TEs pertencentes às ordens LTR (6), LINE (1), TIR (1) e Helitron (1) no cromossomo 7 (Figura 13).



Figura 13. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 7).

O inserto Pe207D11 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 1) e *M. esculenta* (cromossomo 15). A região de *P. edulis* não apresenta elementos móveis. Na comparação com *P. trichocarpa*, a região do cromossomo 1 possui orientação oposta e é 4x maior que a de *P. edulis*; além disso, foram identificados oito genes ortólogos e um elemento móvel da ordem Helitron no cromossomo 1. Na comparação com *M. esculenta*, a região do cromossomo 15 é aproximadamente 2x maior e seis genes ortólogos foram identificados (Figura 14).



Figura 14. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 1) e *M. esculenta* (cromossomo 15).

Cinco insertos de *P. edulis* apresentaram 24 genes. O inserto Pe93N7 apresenta microssintenia colinear apenas com *P. trichocarpa* (cromossomos 6 e 18). Na região de *P. edulis*, 15 genes ortólogos estão duplicados e um elemento móvel da ordem LTR foi identificado. As regiões cromossômicas (6 e 18) de *P. trichocarpa* são aproximadamente 3x maiores que a de *P. edulis*. Há 23 genes ortólogos e cinco elementos móveis das ordens LTR (4) e LINE (1) no cromossomo 6. A região do cromossomo 18 possui orientação oposta e foram identificados 16 genes ortólogos e seis elementos móveis das ordens LTR (2), DIRS (1) e TIR (3). Este caso é um bom exemplo de microssintenia com uma ordem muito conservada dos genes ortólogos (Figura 15).



Figura 15. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 6 e 18).

O inserto Pe108C16 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 6 e 18) e *M. esculenta* cromossomos (3 e 15). A região de *P. edulis* apresenta um elemento móvel da ordem LTR. Nas comparações com *P. trichocarpa*, 11 genes ortólogos de *P. edulis* estão duplicados. As regiões cromossômicas 6 e 18 são 2x maiores que a de *P. edulis*. Na região do cromossomo 6, há 16 genes ortólogos e dois elementos móveis das ordens LTR (1) e TIR (1). A região do cromossomo 18 possui orientação oposta, há 13 genes ortólogos e um elemento móvel da ordem Helitron. Observe que o gene 2 de *P. edulis* possui duas cópias nos cromossomos 6 e 18. Já nas comparações com *M. esculenta*, cinco genes ortólogos de *P. edulis* estão duplicados e esta região é ligeiramente comparativamente àquela dos cromossomos 3 (com orientação oposta) e 16. Em ambos os casos, foram identificados 10 genes ortólogos é não há elementos móveis (Figura 16).



Figura 16. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 6 e 18) e *M. esculenta* (cromossomos 3 e 15).

O inserto Pe173B16 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 4 e 9) e *M. esculenta* (cromossomo 4). Nas comparações com *P. trichocarpa*, a região de *P. edulis*, com orientação oposta, possui 17 genes ortólogos duplicados e um TE da ordem TIR. Observe que o cromossomo 4 tem um segmento cromossômico de 200.883 pb que está ausente em *P. edulis*; além disso, é cerca de 4x maior que a de *P. edulis*. Foram identificados 28 genes ortólogos e 24 TEs das ordens LTR (11), LINE (1), TIR (7) e Helitron (5) no cromossomo 4. A região do cromossomo 9 é ligeiramente maior e há 29 genes ortólogos, sendo esta a região com o maior número de ortólogos nas comparações com *P. trichocarpa*. Foram identificados três TEs das ordens TIR (1) e Helitron (2) no cromossomo 9. Observe que em uma das pontas da região de *P. edulis*, o gene que codifica a proteína induzida por auxina A15 possui três cópias, com um gene ortólogo em cada

cromossomo (4 e 9). Também foi observado que o gene 3, que codifica uma amidase, possui seis cópias no cromossomo 4 e três 9. Além disso, o gene 19 que codifica o conjunto de aquaporinas TIP1-2 presente em *P*. *edulis* possui cinco cópias no cromossomo 9.

Na comparação com *M. esculenta*, *P. edulis* não apresenta TEs. A região do cromossomo 4 é aproximadamente 2x maior que a de *P. edulis*. Há 20 genes ortólogos e 12 TEs das ordens LTR (6), DIR (2) e TIR (4) no cromossomo 4. Note que três genes de *P. edulis*, localizados na extremidade final, possuem cópias duplicadas no cromossomo 4 de *M. esculenta* (Figura 17).



Figura 17. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 4 e 9) e *M. esculenta* (cromossomo 4).

O inserto Pe185J16 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 12) e *M. esculenta* (cromossomo 1). A região de *P. edulis* não apresenta elementos móveis em ambas as comparações. Na comparação com *P. trichcocarpa*, a região do cromossomo 12 é cerca de 5x maior a de *P. edulis*. Há 10 genes ortólogos e seis elementos móveis das ordens LTR (3), DIRS (1) e TIR (2) no cromossomo 12. Na comparação com *M. esculenta*, a região do cromossomo 1 é cerca de 4x maior que a de *P. edulis* e nela foram identificados 12 genes ortólogos e cinco elementos móveis das ordens LTR (4) e TIR (1) (Figura 18).



Figura 18. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 12) e *M. esculenta* (cromossomo 1).

O inserto Pe212I1 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 2 e 5) e *M. esculenta* (cromossomo 18). Em ambas as comparações, a região de *P. edulis* possui dois TEs da ordem LTR. Nas comparações com *P. trichocarpa*, 12 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 2 possui orientação oposta e é 2x maior que a região de *P. edulis*. Foram identificados 14 genes ortólogos e 10 TEs das ordens LTR (8), TIR (1) e Helitron (1) no cromossomo 2. A região do cromossomo 5 é 2x maior a região de *P. edulis* e nela há 13 genes ortólogos e sete TEs da ordem LTR. Já na comparação com *M. esculenta*, a região do cromossomo 18 apresenta orientação oposta e é aproximadamente 2x maior que a de *P. edulis*. Foram identificados 14 genes ortólogos e três TEs das ordens LTR (2) e TIR (1) no cromossomo 18. Observe em *P. edulis*, três cópias do gene que codifica uma proteína hipotética transmembranar (genes 2-4) que possuem ortologia nas comparações com *P trichocarpa* (cromossomo 2) e *M. esculenta* (cromossomo 18) (Figura 19).



Figura 19. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 2 e 5) e *M. esculenta* (cromossomo 18).

Três insertos de *P. edulis* apresentaram 23 genes. O inserto Pe93J9 mostra microssintenia colinear apenas com *P. trichocarpa* (cromossomo 5), cujo tamanho é 3x maior que o da região de *P. edulis* que, por sua vez, apresenta um TE da ordem LTR. Há 18 genes ortólogos e um TE da ordem LTR identificado no cromossomo 5 (Figura 20).



Figura 20. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 5).

O inserto Pe135J12 apresenta microssintenia colinear somente com *M. esculenta* (cromossomo 17), cujo tamanho é aproximadamente 3x maior que o da região de *P. edulis* que, por sua vez, possui três elementos móveis da ordem LTR. Foram identificados 20 genes ortólogos e 14 elementos móveis das ordens LTR (13) e DIRS (1) no cromossomo 17. Observa-se que o gene que codifica a proteína glicosiltransferase mostra duas cópias (genes 3 e 4) em *P. edulis* com ortólogos duplicados em *M. esculenta*. Além disso, um conjunto de sete cópias gênicas (genes 13-19) que codificam uma família de proteínas transportadoras dos tipos nodulina 21 e WAT1 foram encontrados em *P. edulis*. Note que três dessas cópias gênicas (genes 16, 17 e 19) apresentam também ortologia com outros dois genes de *M. esculenta* (Figura 21).



Figura 21. Regiões microssintênicas colineares identificadas em P. edulis e M. esculenta (cromossomo 17).

O inserto Pe195F4 apresenta regiões microssintênicas colineares com *P. trichocarpa* cromossomos (4 e 9) e *M. esculenta* (cromossomo 4). Em *P. edulis* não há TEs. Nas comparações com *P. trichocarpa*, 18 genes ortólogos de *P. edulis* estão duplicados. As regiões cromossômicas 4 e 9 foram ligeiramente maiores que a região de *P. edulis*. Foram identificados 19 ortólogos no cromossomo 4. No cromossomo 9, há 22 ortólogos e dois TEs das ordens TIR (1) e Helitron (1), além de segmentos rearranjados. Na comparação com *M. esculenta*, a região microssintênica é aproximadamente 3x maior que a de *P. edulis*. Há 15 genes ortólogos e 14 TEs das ordens LTR (8), DIR (1) e TIR (5) no cromossomo 4 (Figura 22).



Figura 22. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* cromossomos (4 e 9) e *M. esculenta* (cromossomo 4).

Cinco insertos de *P. edulis* contem 22 genes. O inserto Pe74I6 apresenta microssintenia colinear apenas com *P. trichocarpa* (cromossomos 8 e 10). A região de *P. edulis* possui 16 genes ortólogos duplicados. A região do cromossomo 8 é aproximadamente 2x maior que a de *P. edulis*. Há 18 genes ortólogos e 10 TEs das ordens LTR (8), TIR (1) e Helitron (1) no cromossomo 8. A região do cromossomo 10 possui orientação oposta e é aproximadamente 4x maior que a de *P. edulis*. Foram identificados 26 genes ortólogos é 11 TEs das ordens LTR (8), SINE (1), TIR (1) e Maverick (1) no cromossomo 10. Observa-se que o gene de *P. edulis* (17) que codifica a proteína glutation S-tranferase mostra nove cópias no cromossomo 10 de *P. trichocarpa* (Figura 23). Considerando todas as comparações, as duplicações gênicas encontradas em *P. trichocarpa* abrangem um mínimo de duas até o máximo de nove cópias.



Figura 23. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 8 e 10).

O inserto Pe84M18 apresenta microssintenia colinear apenas com *M. esculenta* (cromossomos 6 e 14). Note que a região de *P. edulis* possui oito genes ortólogos duplicados. A região do cromossomo 6 apresenta orientação oposta e é aproximadamente 5x maior que a de *P. edulis*. Há 13 genes ortólogos e 30 TEs das ordens LTR (27), DIRS (2) e TIR (1) no cromossomo 6. Foram identificados três TEs das ordens LTR (1) e TIR (2) no cromossomo 14 (Figura 24).



Figura 24. Regiões microssintênicas colineares identificadas em P. edulis e M. esculenta (cromossomos 6 e 14).

O inserto Pe101O4 mostra microssintenia colinear apenas com *P. trichocarpa* (cromossomo 2). Em *P. edulis* há dois elementos móveis da ordem LTR. Observe que um segmento do cromossomo 2 (155.162 pb) não

está presente em *P. edulis*, assim como foi descrito em casos anteriores (Figuras 13 e 17). Além disso, a região do cromossomo 2 é 2x maior e contem sete genes ortólogos e oito TEs das ordens LTR (6), DIRS (1) e TIR (1). Finalmente, há dois elementos móveis da ordem LTR em *P. edulis* (Figura 25).



Figura 25. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomo 2).

O inserto Pe141J23 apresenta microssintênica colinear apenas com *M. esculenta* (cromossomo13). A região de *P. edulis* apresenta orientação oposta. O tamanho do cromossomo 13 é 2x maior que o da região de *P. edulis*. Foram identificados 10 genes ortólogos e 10 elementos móveis das ordens LTR (7), DIRS (1), LINE (1) e TIR (1) no cromossomo 13. Observe que ambos os genes 9 e 22 de *P. edulis* possuem duas cópias ortólogas em *M. esculenta* (Figura 26).



Figura 26. Regiões microssintênicas colineares identificadas em P. edulis e M. esculenta (cromossomo 13).

O inserto Pe201C11 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 1 e 11) e *M. esculenta* (cromossomos 15 e 16). A região de *P. edulis* apresenta genes ortólogos duplicados e um TE da ordem DIRS em ambas as comparações. Nas comparações com *P. trichocarpa*, a região de *P. edulis* apresenta orientação oposta e há sete genes ortólogos duplicados. No cromossomo 1 há sete ortólogos e três TEs da ordem LTR. Foram identificados seis ortólogos e um TE da ordem TIR no cromossomo 11. O tamanho de ambas regiões é similar em *P. edulis*.

Já nas comparações com *M. esculenta*, três ortólogos de *P. edulis* estão duplicados. As regiões dos cromossomos 15 e 16 são menores que a de *P. edulis*. No cromossomo 15 há cinco genes ortólogos enquanto no cromossomo 16 (com orientação oposta) há quatro ortólogos e três TEs da ordem LTR (Figura 27).



Figura 27. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 1 e 11) e *M. esculenta* (cromossomos 15 e 16).

Seis insertos de *P. edulis* apresentaram 21 genes. O inserto Pe69G18 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 14) e *M. esculenta* (cromossomo 13). Na comparação com *P. trichocarpa*, a região do cromossomo 14 é 3x maior que a região de *P. edulis*. Há 13 genes ortólogos e dois TEs das ordens LTR (1) e TIR (1) no cromossomo 14. Na comparação com *M. esculenta*, a região de *P. edulis* apresenta orientação oposta. Foram identificados 10 genes ortólogos e um TE da ordem LTR no cromossomo 13 (Figura 28).



Figura 28. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 14) e *M. esculenta* (cromossomo 13).

O inserto Pe69H24 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 4 e 9) e *M. esculenta* (cromossomo 4). Na região de *P. edulis* há inversões e poucas duplicações gênicas. Nas comparações com *P. trichocarpa*, 15 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 4 é 2x maior que a de *P. edulis*. Foram identificados 18 genes ortólogos e quatro TEs das ordens LTR (3) e Helitron (1) no cromossomo 4. A região do cromossomo 9 é 2x maior que a de *P. edulis* e contém 20 genes ortólogos. Na comparação com *M. esculenta*, a região de *P. edulis* apresenta orientação oposta. A região do cromossomo 4 é 3x maior que a de *P. edulis* e nela foram identificados oito genes ortólogos e 10 TEs das ordens LTR (6), DIRS (1) LINE (2) e TIR (1) (Figura 29).



Figura 29. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 4 e 9) *e M. esculenta* (cromossomo 4).

O inserto Pe93K19 apresenta regiões microssintênicas colineares com *P. trichocarpa* (cromossomos 1 e 3) e *M. esculenta* (cromossomos 4 e 11). Em ambas as comparações, a região de *P. edulis* possui genes ortólogos duplicados e um TE da ordem LTR. Nas comparações com *P. trichocarpa*, 15 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 1 apresenta orientação oposta e é aproximadamente 3x maior que a de *P. edulis*. Há 16 genes ortólogos e seis TEs das ordens LTR (2), TIR (1) e Helitron (3) no cromossomo 1. A região do cromossomo 3 é cerca de 4x maior que a de *P. edulis* e nela foram identificados 22 genes ortólogos e sete TEs das ordens LTR (4), DIRS (1) e TIR (2). Note que os genes 13 e 14 de *P. edulis* possuem duas e três cópias respectivamente no cromossomo 3.

Nas comparações com *M. esculenta*, 12 genes ortólogos de *P. edulis* estão duplicados. A região do cromossomo 4 é aproximadamente 2x maior que a de *P. edulis*. Há 14 ortólogos e três TEs das ordens LTR (2) e TIR (1) no cromossomo 4. A região do cromossomo 11 apresenta orientação oposta e é 2x maior que a de *P. edulis* e aí foram identificados 20 ortólogos e cinco TEs das ordens LTR (4) e TIR (1). Observe que há duas cópias do gene 14 de *P. edulis* no cromossomo 11 (Figura 30).



Figura 30. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomos 1 e 3) e *M. esculenta* (cromossomos 4 e 11).

O inserto Pe125I23 de *P. edulis* cotem regiões microssintênicas colineares com *P. trichocarpa* (cromossomo 8) e *M. esculenta* (cromossomos 1 e 5). Em ambas as comparações, a região de *P. edulis* apresenta microssintenia com dois segmentos separados no cromossomo 8 de *P. trichocarpa* e no cromossomo 5 de *M. esculenta* por ~ 3,5 Mb e ~ 171,3 Kb, respectivamente, assim como pode ser observado na figura 5. Nas comparações com *P. trichocarpa*, a região de *P. edulis* e o 2° segmento do cromossomo 8 apresentam orientação opostas. O 1° segmento do cromossomo 8 contem nove ortólogos e nenhum TE. Entre o 2° segmento do cromossomo 8 e a região de *P. edulis* foram identificados seis ortólogos e um elemento móvel da ordem LTR. Já nas comparações de *M. esculenta*, 11 genes ortólogos de *P. edulis* estão duplicados. O tamanho do cromossomo 1 é cerca de 2x maior que o da região de *P. edulis*. Há 18 genes ortólogos e um elemento móvel da ordem LINE (1) e SINE (1) no cromossomo 1. Foram identificados cinco ortólogos e um elemento móvel da ordem LINE no 1° segmento do cromossomo 5, enquanto o 2° apresenta orientação oposta e é 2x maior que a região de

P. edulis. Foram identificados sete ortólogos e quatro elementos móveis das ordens LTR (2) e TIR (2) no 2° segmento. Observe que em *P. edulis* há duas cópias do gene que codifica a proteina formina (genes 2 e 3) com um ortólogo em *P. trichocarpa* (2° segmento do cromossomo 8) e *M. esculenta* (cromossomo 1) (Figura 31).



Figura 31. Regiões microssintênicas colineares identificadas em *P. edulis* e *P. trichocarpa* (cromossomo 8) e *M. esculenta* (cromossomos 1 e 5).

O inserto Pe164A12 apresenta microssintenia colinear apenas com *M. esculenta* (cromossomo 16). A região de *P. edulis* tem orientação oposta e um TE da ordem LTR. A região do cromossomo 16 é 3x maior que a de *P. edulis* e nela há 18 genes ortólogos e 10 TEs das ordens LTR (5), LINE (2) e TIR (3). Também, observe a presença de segmentos rearranjados (Figura 32).



Figura 32. Regiões microssintênicas colineares identificadas em P. edulis e M. esculenta (cromossomo 16).

Finalmente, o inserto Pe214A18 apresenta microssintenia colinear apenas com *P. trichocarpa* (cromossomos 12 e 15). A região de *P. edulis* contem dois elementos móveis das ordens LTR. Há 10 genes ortólogos e um elemento móvel da ordem LTR no cromossomo 12 e nove genes ortólogos e dois elementos móveis das ordens TIR (1) e Helitron (1) no cromossomo 15 (Figura 33).



Figura 33. Regiões microssintênicas colineares identificadas em P. edulis e P. trichocarpa (cromossomos 12 e 15).

3.3. Identificação dos elementos transponíveis

A busca por elementos transponíveis nos espaços intergênicos das regiões microssintênicas nas três espécies resultou em um total de 517 elementos. Nas espécies *M. esculenta* e *P. trichocarapa* houve o maior número de elementos, 248 e 244 TEs, respectivamente, em comparação com *P. edulis*, onde foram encontrados apenas 25 elementos (Tabela 3). Importante mencionar que das 32 regiões microssintênicas de *P. edulis* apenas 17 tinham TEs nos espaços intergênicos (Figuras 3, 4, 9, 10, 12, 13, 15, 16, 17, 19, 20, 21, 25, 27, 30, 32 e 33).

Em termos gerais, os retrotransposons foram a classe predominante nas regiões analisadas, destacando-se a ordem LTR. Em *P. edulis*, foram encontrados 84% elementos desta ordem que abrangem 166.121 pb. Já em *P. trichocarpa* há 58,20% de LTRs abrangendo 404.906 pb, enquanto em *M. esculenta* há 71,37%, correspondendo a 557.865 pb (Tabela 3).

O maior número de elementos da ordem LTR em *P. edulis* (3) foi identificado no inserto Pe135J12 (Figura 21). Em *P. trichocarpa*, o maior número de LTRs foi reconhecido em determinadas regiões cromossômicas. No primeiro caso, corresponde aos cromossomos 4 (4 e 9 LTRs no 1° e 2° segmento, respectivamente) e 9 (1 e 3 LTRs no 1° e 2° segmento, respectivamete), quando comparados com o inserto Pe214H11 (Figura 5) e o segundo caso corresponde aos cromossomos 4 (16 LTRs) e 17 (1 LTR), quando comparadas com o inserto Pe164D9 de *P. edulis* (Figura 6). Em cada caso, foram reconhecidos 17 LTRs.

Já em *M. esculenta*, na comparação entre o inserto Pe84M18 e os cromossomos 6 e 14, há um total de 28 LTRs; desses, um elemento no cromossomo 14 e os 27 restantes nos espaçamentos intergênicos do cromossomo 6 (Figura 24).

As outras ordens da classe 1, como DIRS e LINE, foram representadas por poucos elementos em *P. trichocarpa* e *M. esculenta*. Em *P. edulis*, foi identificado apenas um elemento de cada ordem. No caso da ordem SINE, um reduzido número de elementos foi identificado tanto em *P. trichocarpa* (1,23%, 3/244) (Figuras 5 e 23) como em *M. esculenta* (0,81%, 2/248) (Figuras 3 e 31). Por último, não se encontraram elementos da ordem PLE nas três espécies (Tabela 3).

Em relação aos transposons de DNA, (subclasse 1), a ordem TIR foi a mais abundante. Em *P. trichocarpa* foram identificados 52 elementos (21,31%, 52/244), o maior número no cromossomo 4 (13 elementos) e 17 (3 elementos) perfazendo um total de 16 elementos (Figura 6). Em *M. esculenta* foram detectados 42 elementos TIR (16,94%, 42/248), o maior número (5) no cromossomo 4 (Figura 22). Em *P. edulis* apenas dois elementos foram identificados (8%, 2/25) (Figuras 12 e 17). Por outro lado, não foram encontrados elementos da ordem Crypton nas três espécies (Tabela 3).

Além disso, com relação à subclasse 2, evidenciaram-se elementos da ordem Helitron, principalmente em *P. trichocarpa*, o maior número nos cromossomos 4 (5) e 9 (2) totalizando sete elementos (Figura 17). Em *M. esculenta*, há apenas um elemento dessa ordem (Figura 6) e nenhum em *P. edulis*. Finalmente, foi identificado apenas um elemento da ordem Maverick no cromossomo 10 de *P. trichocarpa* (Figura 23 e Tabela 3).

		I	P. edulis	P. trichocarpa				M. esculenta				
	Número de elementos	Tamanho (pb)	Porcentagem do elemento	Porcentagem do tamanho	Número de elementos	Tamanho (pb)	Porcentagem do elemento	Porcentagem do tamanho	Número de elementos	Tamanho (pb)	Porcentagem do elemento	Porcentagem do tamanho
CLASSE I (Retrotransposons)	23	183.315	92,00	89,45	167	475.524	68,44	69,88	205	631.437	82,66	85,29
LTR	21	166.121	84,00	81,06	142	404.906	58,20	59,50	177	557.865	71,37	75,35
DIRS	1	11.919	4,00	5,82	13	53.236	5,33	7,82	11	45.589	4,44	6,16
PLE	0	0	0	0	0	0	0	0	0	0	0	0
LINE	1	5.275	4,00	2,57	9	16.309	3,69	2,40	15	27.181	6,05	3,67
SINE	0	0	0	0	3	1.073	1,23	0,16	2	802	0,81	0,11
CLASSE II (Transposons subclasse 1)	2	21.618	8,00	13,78	52	167.081	21,31	24,55	42	105.200	16,94	14,21
TIR	2	21.618	8,00	10,55	52	167.081	21,31	24,55	42	105.200	16,94	14,21
Crypton	0	0	0	0	0	0	0	0	0	0	0	0
CLASSE II (Transposons subclasse 2)	0	0	0	0	25	37.896	10,25	5,57	1	3.729	0,40	0,50
Helitron	0	0	0	0	24	36.918	9,84	5,43	1	3.729	0,40	0,50
Maverick	0	0	0	0	1	978	0,41	0,14	0	0	0	0
Total	25	204.933	100	100	244	680.501	100	100	248	740.366	100	100

Tabela 3. Classificação dos elementos transponíveis identificados nos espaçamentos intergênicos de uma fração rica em genes de *P. edulis* comparativamente às regiões de *P. trichocarpa* e *M. esculenta*.

4. DISCUSSÃO

Abordagens de genômica comparativa tem sido utilizadas para analisar sequências genômicas e encontrar similaridades e diferenças [22], contribuindo para o entendimento das relações evolutivas entre os organismos comparados. No caso das espécies pouco estudadas ou que não tem o genoma totalmente sequenciado, como é o caso de *P. edulis*, com apenas 1% do genoma sequenciado [12], as bibliotecas genômicas de grandes insertos são alternativas para se dar início às análises genômicas comparativas, sobretudo microssintênicas [73]. Nesses casos, sequências com uma boa anotação e de alta qualidade são consideradas indispensáveis [74], justificando aqui as análises aqui conduzidas.

Estudos de diversificação em angiospermas revelaram a diversificação da ordem Malpighiales aconteceu entre 100 a 90 milhões de anos atrás [75]. Apesar desse longo período de divergência, as análises de microssintenia entre uma região rica em genes de *P. edulis* (Passifloreaceae) e as regiões cromossômicas das Malpighiales *P. trichcocarpa* (Salicaceae) e *M. esculenta* (Euphorbiaceae), no geral, revelaram alta colinearidade. A conservação na ordem dos genes é justificada pela ancestralidade compartilhada dos genomas aqui analisados [76]. No entanto, por pertencerem a famílias distintas, em ambas comparações, foram observados rearranjos, sobretudo inversões. Raskina e colaboradores afirmaram que esses rearranjos se devem aos mecanismos de especiação de cada grupo, sendo mais frequentes as inversões [77]. Similarmente ao previamente encontrado nas comparações de um conjunto de 110 ortólogos de berinjela e tomate (Solanaceae), nossos resultados também apontam para a ocorrência de poucos segmentos translocados [78]. Além disso, nossas ánalises revelaram casos de regiões cromossômicas com orientação oposta, o que coincide com os resultados do estudo comparativo entre *Brassica oleracea* e *Arabidopsis* (Brassicaceae). Apesar de haver alta colinearidade dos genes nas regiões sintênicas, foram detectadas duas regiões em *Arabidospsis* com orientação oposta, neste caso, a presença desses segmentos está relacionado à ocorrência de rearranjos na linhagem ancestral de *Arabidopsis* [30].

A poliploidia representa um ponto chave na evolução dos genomas vegetais [12, 79–81], especialmente em angiospermas [82]. No geral, é comum que espécies com um histórico de poliploidia acumulem variações genômicas, como inversões, translocações, duplicações e deleções em seus genomas [83]. Eventos de duplicação do genoma de *P. trichocarpa* ocorreram 65 milhões de anos atrás, identificando-se cerca de 8.000 pares de genes duplicados, altamente conservados, e muitas duplicações em tandem [84]. No genoma de *M. esculenta*, também houve eventos de paleo-duplicação, provavelmente 47 milhões de anos atrás [85], sendo que muitos genes inicialmente duplicados foram perdidos e outros tem mais de duas cópias [86]. Em *Passiflora*, não se exclui que eventos de duplicação tenham ocorrido. Possivelmente isto se deu durante a diversificação do gênero a partir do ancestral, cerca de 38 milhões de anos atrás [87]. Há registros de espécies de *Passiflora* com um número cromossômico diploide 2n = 12, 18, 20 e 22 [88–90], tetraploide (2n = 24), hexaplóide (2n = 36) ou octaploide (2n = 72), considerando x = 6 como número básico primário [88, 91]. No entanto, números secundários foram propostos (x = 9, 10, 12) [88, 92, 93]. Hansen et al. [93] com base em análises de sequências indicaram x = 12 como o número básico mais provável. Há sugestão que a evolução do grupo teve como base os eventos de disploidia e poliploidia. *P. edulis*, por exemplo, seria resultado de disploidia ascendente $(6 \rightarrow 9)$ [94], dando suporte às duplicações gênicas aqui identificadas. Estudos recentes indicam que pelo menos dois eventos de duplicação genômica (*whole genome duplication*) acorreram em *P. edulis* (dados não publicados).

Sequências duplicadas são encontradas com frequência em estudos de genômica comparativa [95]. Aqui, maximamente, sete cópias foram descobertas em *P. edulis*, nove em *P. trichocarpa* e cinco em *M. eculenta*. Interessantemente, os resultados de sintenia mostraram que as sequências duplicadas são também ortólogas. Nesse contexto, é importante mencionar que a duplicação se dá em uma taxa de 0.01 gene por milhão de anos. Geralmente, genes duplicados são silenciados e aqueles que permanecem são submetidos a períodos de seleção purificadora [96]. A análise comparativa entre milho, arroz e sorgo revelou que as duplicações gênicas em tandem aconteceram após a divergência dos grupos [97]. Nas comparações com *P. trichocarpa*, foram identificados três casos de deleção em *P. edulis*. Segmentos deletados podem surgir durante a síntese de DNA por reparo de alguma ruptura da dupla fita [98]. Além disso, a recombinação também está envolvida no processo de perda de DNA [83].

Por outro lado, as análises filogenéticas a partir do genoma cloropastidial de *P. edulis* comparativamente ao de outras famílias da ordem Malpighiales evidenciaram que *P. edulis* (Passifloraceae) está mais próximo de *P. trichocarpa* (Salicaeace) do que de *M. esculenta* (Euphorbiaceae) [65]. Essas informações são corroboradas com os nossos resultados que mostraram um maior grau de microssintenia entre *P. edulis* e *P. trichocarpa*, do que entre *P. edulis* e *M. esculenta*. Do mesmo modo, estudos comparativos entre *Fragaria* (Rosidea do grupo I) e outras espécies do mesmo clado como *Populus*, *Medicago* (Rosideas do grupo I), *Vitis* (Rosidea basal) e *Arabidopsis* (Rosidea do grupo II) evidenciaram um maior grau de sintenia entre genomas filogeneticamente mais próximos [99], sendo identificados um menor número de segmentos sintênicos entre *Fragaria* e *Arabidopsis* [32]. Assim, demonstra-se que a genômica comparativa constitui um aspecto importante para esclarecer ou complementar inferências filogenéticas [12, 24].

Os resultados relativos a elementos transponíveis e, espaçamentos intergênicos revelaram que a ordem LTR dos retrotransposons é a mais abundante, com 21, 142 e 177 elementos identificados em *P. edulis, P. trichocarpa* e *M. esculenta*, respectivamente. TEs podem ocupar entre 80 a 90% do tamanho de genomas de plantas [100, 101], sendo os retrotransposons, especificamente da ordem LTR, mais abundantes [102–104]. Munhoz et al. [12] identificaram TEs (= 17,5% do total de sequências) em uma fração rica em genes de *P. edulis* (107 insertos de BACs), dos quais 181 correspondiam a LTRs. Parte dessa fração foi aqui analisada (32 insertos). Em *P. trichocarpa* e *M. esculenta*, espécies com os genomas totalmente sequenciados, foram identificados cerca de 42% [84] e 25,7% [105] TEs, respectivamente, a maioria retrotransposons da ordem LRT. Embora não sejam dados diretamente comparáveis, é importante ressaltar a presença e LTRs em regiões gênicas, nos três casos.

O tempo de inserção dos LTRs nos genomas de *M. esculenta* e *P. trichocarpa* foi estimado em 4 a 8 milhões de anos, respectivamente [106]. Também, recentemente, foi sugerido que a inserção de LTRs no genoma de *P. edulis* aconteceu há menos de 2 milhões de anos e foi evidenciada atividade transcripcional desses elementos [69]. Em tomate, foi revelado que os retrotransposons (LTRs) mais antigos habitam as regiões de heterocromatina; no entanto, os LTR jovens estão nas regiões eucromáticas, ricas em genes, mas tendem a ser eliminados [107].

No que concerne às outras ordens (DIRS, PLE, LINE, SINE), em geral, a sua frequência é muito menor comparativamente à de LTRs nos vegetais [108]. Retrotransposons da ordem DIRS ocorrem com maior constância em protistas [109] e em algumas algas verdes [110]. As ordens LINE e SINE são comuns em animais [111]; no entanto, também se encontram em plantas.

No que concerne aos transposons, a ocorrência deles é limitada em vegetais [103]; a ordem TIR foi a melhor representada, sobretudo em *P. trichocarpa* (52) e *M. esculenta* (42), e Helitron em *P. trichocarpa* (24). Os Helitrons são mais diversos em algumas espécies vegetais, embora de difícil identificação pela ausência de uma estrutura terminal [111].

Um aspecto particular relevante é que as regiões microssintênicas de *P. edulis* são menores e bem compactadas comparativamente às regiões de *P. trichocarpa* e *M. esculenta*, embora o genoma de *P. edulis* (1,23 Gb) seja três e duas vezes maior que o de *P. trichocarpa* (485 Mb) e *M. esculenta* (742 Mb), respectivamente. Essa aparente contradição nos faz levantar a seguinte questão: como se explica o tamanho do genoma de *P. edulis*, se (ao menos na região analisada) o conteúdo de TEs é baixo? E, tal como indicam vários estudos, o tamanho dos genomas vegetais está associado à proliferação de TEs [112, 113].

A resposta está no fato de terem sido analisadas regiões de *P. edulis* ricas em genes as quais mostraram, espaçamentos intergênicos com poucos TEs. Essa resposta está apoiada nas afirmações de Cossu et al. [114] ao indicarem que regiões ricas em genes apresentam principalmente TEs dispersos, ao contrário daquelas pobres em genes que mostram TEs agrupados. Os eventos de poliploidia também influenciam nessa variação, ou seja, presença de TEs e tamanho dos espaçamentos intergênicos [115]. Interessantemente, tanto *P. trichocarpa* como *M. esculenta* são espécies poliploides, cujos genomas são menores que o de *P. edulis*, e que apresentam espaçamentos intergênicos maiores. Resalta-se que regiões ricas em genes tendem a eliminar TEs dado o seu efeito deletério.

Após a poliploidização do genoma, a proliferação de TEs é ativada por meio de mecanismos epigenéticos [82, 116]. Em vista disso, é presumível que tanto a poliploidia como a proliferação de TEs venham a elucidar ganho em tamanho do genoma de *P. edulis* [69]. A multiplicação de TEs não é constante, há períodos em que a abundância de TEs aumenta rapidamente, levando ao aumento no tamanho dos genomas, enquanto outros períodos são quiescentes durante o tempo evolutivo [117]. Pesquisas futuras são indispensáveis para explorar a distribuição dos TEs e de outras sequências de DNA repetitivo (repetições em tandem, sequências teloméricas e centroméricas) no genoma de *P. edulis*. Há evidências que as frações repetidas estão associadas à variação no tamanho dos genomas embora sua contribuição seja menor que a dos TEs [118]. Em algumas espécies vegetais podem ocupar entre seis até 21% do tamanho dos genomas [118, 119].

Em *P. edulis* existem evidências citogenéticas da distribuição das sequências repetidas principalmente nas regiões pericentroméricas ou subcentroméricas, com sinais em um ou ambos telômeros, dependendo do cromossomo, embora, os retrotransposons apresentam um padrão repetitivo disperso [94]. Porém, somente com o genoma completo disponível e a descrição da distribuição e abundância do DNA repetitivo [120] em *P. edulis* contribuirá para o entendimento do papel dessas sequências na história evolutiva da espécie.

5. CONCLUSÕES

O método aqui utilizado para a identificação de regiões de microssintênicas representa um recurso importante é muito útil em análises comparativas nos casos em que não se dispõe do genoma sequenciado de uma das espécies analisadas.

Há um maior grau de microssintenia entre *Passiflora edulis* e *Populus trichocarpa* do que entre *P. edulis* e *Manihot esculenta*, havendo um maior número de genes ortólogos na 1ª comparação.

Há colinearidade dos genes ortólogos nos segmentos cromossômicos das três espécies e poucos rearranjos, sendo a inversão, o tipo de variação estrutural mais frequente.

Os retrotransposons da ordem LTR são os mais abundantes nas três espécies, ressalvando que este resultado refere-se às frações analisadas, ricas em genes, já que elementos transponíveis tendem a ser eliminados dessas regiões.

A compactação dos genes é maior em *P. edulis* devido à maior quantidade de elementos repetitivos em *P. trichocarpa* e *M. esculenta*, embora ambas apresentem genomas menores. Sugere-se que o maior tamanho do genoma de *P. edulis* esteja associado à abundância de sequências repetitivas em regiões pobres em genes. Claramente, futuras pesquisas devem ser realizadas para elucidar a distribuição e abundância de DNA repetitivo no genoma de *P. edulis*.

REFERÊNCIAS

1. APG. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. J Linn Soc. 2009;161:105–21.

2. Ulmer T, MacDougal JM. Passiflora : Passionflowers of the world. Timber Press. 2004;430.

3. Bernacci LC, Cervi AC, Milward-de-Azevedo MA, Nunes TS, Imig DC, Mezzonato AC. Passifloraceae in lista de espécies da flora do Brasil. Jardim Botânico do Rio de Janeiro. 2015.

4. Feuillet C. Passifloraceae (Passion flower family). In: Flowering plants of the neotropics (eds) Smith N, Mori SA, Herdenson A, Stevenson, DW, Stevenson-Heald, S.D. Princeton University Press, Oxford. 2004;286-287.

5. Lima AA, Cardoso CEL, Souza JS, Pires MM. Comercialização do maracujazeiro. 1ª edição. Cruz das Almas-BA; 2006.

6. Poll H, Benno AZ V, Kist B, Santos C, Carvalho C, Reetz ER, et al. Anuário brasileiro da fruticultura. Editora Gazeta Santa Cruz. Santa Cruz do Sul. 2011;128.

7. Meletti, LMM. e Bruckner CH. Melhoramento genético. In: Maracujá: tecnologia de produção, pós-colheita, agroindústria, mercado (eds) Bruckner CH; Picanço MC. Cinco Continentes. Porto Alegre. 2001; 345-385.

8. Meletti LMM. Avanços na cultura do maracujá no Brasil. Rev Bras Frutic. 2011;33:83-91.

9. IBGE. Produção Agrícola Municipal. 2017.

10. Santos VA, Ramos JD, Laredo RR, Silva FOR, Chagas EA, Pasqual M. Produção e qualidade de frutos de maracujazeiro-amarelo provenientes do cultivo com mudas em diferentes idades. Rev Ciências Agroveterinárias. 2017;16:33–40.

11. Santos AA, Penha HA, Bellec A, Munhoz C de F, Pedrosa-Harand A, Bergès H, et al. Begin at the beginning: A BAC-end view of the passion fruit (*Passiflora*) genome. BMC Genomics. 2014;15:816.

12. Munhoz CF, Costa ZP, Cauz-Santos LA, Reátegui ACE, Rodde N, Cauet S, et al. A gene-rich fraction analysis of the *Passiflora edulis* genome reveals highly conserved microsyntenic regions with two related Malpighiales species. Sci Rep. 2018;8:13024.

13. Carneiro MS, Camargo LEA, Coelho ASG, Vencovsky R, Júnior RPL, Stenzel NMC, et al. RAPD-based genetic linkage maps of yellow passion fruit (*Passiflora edulis* Sims. f. flavicarpa Deg.). Genome. 2002;45:670–8.

14. Lopes R, Lopes MTG, Carneiro MS, de Pina Matta F, Camargo LEA, Vieira MLC. Linkage and mapping of resistance genes to *Xanthomonas axonopodis* pv. *passiflorae* in yellow passion fruit. Genome. 2006;49:17–29.

15. Oliveira EJ, Vieira MLC, Garcia AAF, Munhoz CF, Margarido GRA, Consoli L, et al. An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. J Am Soc Hortic Sci. 2008;133:35–41.

16. Munhoz CF, Santos AA, Arenhart RA, Santini L, Monteiro-Vitorello CB, Vieira MLCC. Analysis of plant gene expression during passion fruit-*Xanthomonas axonopodis* interaction implicates *lipoxygenase 2* in host defence. Ann Appl Biol. 2015;167:135–55.

17. Costa ZP, Munhoz CDF, Vieira MLC. Report on the development of putative functional SSR and SNP markers in passion fruits. BMC Res Notes. 2017;10.

18. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011;52:413–35.

19. Gužvić M. J. The history of DNA sequencing. J Med Biochem. 2013;32:301-12.

20. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.

21. Yu J, Blom J, Glaeser SP, Jaenicke S, Juhre T, Rupp O, et al. A review of bioinformatics platforms for comparative genomics. Recent developments of the EDGAR 2.0 platform and its utility for taxonomic and phylogenetic studies. J Biotechnol. 2017;261:2–9.

22. Wei L, Liu Y, Dubchak I, Shon J, Park J. Comparative genomics approaches to study organism similarities and differences. J Biomed Inform. 2002;35:142–50.

23. Pan X, Stein L, Brendel V. SynBrowse: A synteny browser for comparative sequence analysis. Bioinformatics. 2005;21:3461–8.

24. Caicedo AL, Purugganan MD. Comparative plant genomics. Frontiers and prospects. Plant Physiol. 2005;138:545–7.

25. Chaney L, Sharp AR, Evans CR, Udall JA. Genome mapping in plant comparative genomics. Trends Plant Sci. 2016;21:770–80.

26. Andrew H. Paterson, John E. Bowers, Mark D. Burow, Xavier Draye, Christine G. Elsik, Chun-Xiao Jiang, et al. Comparative genomics of plant chromosomes. Plant Cell. 2000;12:1523–39.

27. Liu H, Sachidanandam R, Stein L. Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order. Genome Res. 2001;11:2020–6.

28. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, et al. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. PLoS Biol. 2003;1:e45–e45.

29. Touchman J. Comparative Genomics. Nat Educ Knowl. 2010;3:13.

30. Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, et al. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell Online. 2006;18:1348–59.

31. Brendel V, Kurtz S, Walbot V. Comparative genomics of *Arabidopsis* and maize: prospects and limitations. Genome Biol. 2002;3:REVIEWS1005.

32. Jung S, Cho I, Sosinski B, Abbott A, Main D. Comparative genomic sequence analysis of strawberry and other rosids reveals significant microsynteny. BMC Res Notes. 2010;3:168.

33. Lall R, Thomas G, Singh S, Singh A, Wadhwa G. Comparative genome analysis of *Solanum lycopersicum* and *Solanum tuberosum*. Bioinformation. 2013;9:923–8.

34. Wang J, Yu JJ, Sun P, Li Y, Xia R, Liu Y, et al. Comparative genomics analysis of rice and pineapple contributes to understand the chromosome number reduction and genomic changes in grasses. Front Genet. 2016;7:174.

35. Asaf S, Khan AL, Aaqil Khan M, Muhammad Imran Q, Kang S-M, Al-Hosni K, et al. Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. PLoS One. 2017;12:e0182281–e0182281.

36. Sharma A, Song J, Lin Q, Singh R, Ramos N, Wang K, et al. Comparative analysis of homologous sequences of *Saccharum officinarum* and *Saccharum spontaneum* reveals independent polyploidization events. Front Plant Sci. 2018;9 :1414.

37. Gogarten JP, Olendzenski L. Orthologs, paralogs and genome comparisons. Curr Opin Genet Dev. 1999;9:630-6.

38. Fang G, Bhardwaj N, Robilotto R, Gerstein MB. Getting started in gene orthology and functional analysis. PLoS Comput Biol. 2010;6:e1000703–e1000703.

39. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49–e49.

40. Studer RA, Robinson-Rechavi M. How confident can we be that orthologs are similar, but paralogs differ? Trends Genet. 2009;25:210–6.

41. Koonin E V. Orthologs, Paralogs, and Evolutionary Genomics. Annu Rev Genet. 2005;39:309-38.

42. Gabaldón T, Koonin E V. Functional and evolutionary implications of gene orthology. Nat Rev Genet. 2013;14:360–6.

43. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant Genomes. Science. 2008;320:486–8.

44. Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, et al. PLAZA: A comparative genomics resource to study gene and genome evolution in plants. Plant Cell. 2009;21:3718–31.

45. Xu Y, Bi C, Wu G, Wei S, Dai X, Yin T, et al. VGSC: A web-based vector graph toolkit of genome synteny and collinearity. Biomed Res Int. 2016;2016:7823429.

46. Fulton TM. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell Online. 2002;14:1457–67.

47. Huo N, Vogel JP, Lazo GR, You FM, Ma Y, McMahon S, et al. Structural characterization of *Brachypodium* genome and its syntenic relationship with rice and wheat. Plant Mol Biol. 2009;70:47–61.

48. Marroni F, Pinosio S, Morgante M. Structural variation and genome complexity: Is dispensable really dispensable? Curr Opin Plant Biol. 2014;18:31–6.

49. Saxena RK, Edwards D, Varshney RK. Structural variations in plant genomes. Brief Funct Genomics. 2014;13:296–307.

50. SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. Science. 1996;274:765–8.

51. Fedoroff N. Transposons and genome evolution in plants. Proc Natl Acad Sci U S A. 2000;97:7002-7.

52. Tenaillon MI, Hollister JD, Gaut BS. A triptych of the evolution of plant transposable elements. Trends Plant Sci. 2010;15:471–8.

53. Lisch D. How important are transposons for plant evolution? Nat Rev Genet. 2013;14:49-61.

54. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 Maize genome: complexity, diversity, and dynamics. Science. 2009;326:1112–5.

55. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.

56. Pantzartzi CN, Pergner J, Kozmik Z. The role of transposable elements in functional evolution of amphioxus genome: the case of opsin gene family. Sci Rep. 2018;8:2506.

57. Bennetzen JL. Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev. 2005;15:621–7.

58. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon J-MM, et al. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet. 2009;5:e1000732–e1000732.

59. Charles M, Belcram H, Just J, Huneau C, Viollet A, Couloux A, et al. Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics. 2008;180:1071–86.

60. Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, et al. A Bacterial artificial chromosome library for sequencing the complete human genome. Genome Res. 2001;11:483–96.

61. Gutman W, Pawełkowicz M, Woycicki R, Piszczek E, Przybecki Z. The construction and characteristics of a BAC library for Cucumis sativus L. "B10". Cell Mol Biol Lett. 2008;13:74–91.

62. Yamaguchi S, Niwa R, Kazuki Y, Ohbayashi T. Application of a bacterial artificial chromosome modification system for a human artificial chromosome vector. Yonago Acta Med. 2011;54:21–31.

63. Yu K. Bacterial artificial chromosome libraries of pulse crops: Characteristics and applications. J Biomed Biotechnol. 2012;2012:493186.

64. Yotoko KSCC, Dornelas MC, Togni PD, Fonseca TC, Salzano FM, Bonatto SL, et al. Does variation in genome sizes reflect adaptive or neutral processes? New clues from *Passiflora*. PLoS One. 2011;6:e18212–e18212.

65. Cauz-Santos LA, Munhoz CF, Rodde N, Cauet S, Santos AA, Penha HA, et al. The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in Malpighiales. Front Plant Sci. 2017;8:334.

66. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:D1178–86.

67. Cao Y, Han Y, Meng D, Li D, Jin Q, Lin Y, et al. Genome-wide analysis suggests high level of microsynteny and purifying selection affect the evolution of EIN3/EIL family in Rosaceae . PeerJ. 2017;5:e3400.

68. Abeel T, Van Parys T, Saeys Y, Galagan J, Van de Peer Y. GenomeView: a next-generation genome browser. Nucleic Acids Res. 2012;40:e12–e12.

69. Costa ZP. Genomic studies in *Passiflora edulis* (Passifloraceae). Tese de doutorado. ESALQ-USP. Piracicaba. 2018;94.

70. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol. 2005;1:166–75.

71. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. PLoS One. 2011;6:e16526–e16526.

72. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr arXiv13033997. 2013;00:1–3.

73. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499–510.

74. Grueber CE. Comparative genomics for biodiversity conservation. Comput Struct Biotechnol J. 2015;13:370–5.

75. Magallón S, Castillo A. Angiosperm diversification through time. Am J Bot. 2009;96:349-65.

76. Oh D-H, Dassanayake M. Network of orthologs and evolutionary context search unveil the landscape of gene transposition-duplication within the Brassicaceae family. bioRxiv. 2017;:236299.

77. Raskina O, Barber JC, Nevo E, Belyayev A. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. Cytogenet Genome Res. 2008;120:351–7.

78. Wu F, Eannetta NT, Xu Y, Tanksley SD. A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet. 2009;118:927–35.

79. Wendel JF, Cronn RC, Johnston JS, Price HJ. Feast and famine in plant genomes. Genetica. 2002;115:37-47.

80. Melo NF, Guerra M. Variability of the 5S and 45S rDNA sites in *Passiflora* L. Species with distinct base chromosome numbers. Ann Bot. 2003;92:309–16.

81. Leitch IJ, Bennett MD. Genome downsizing in polyploid plants. Biol J Linn Soc. 2004;82:651-63.

82. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. Genome Biol. 2016;17:37.

83. Gaeta RT, Chris Pires J. Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol. 2010;186:18–28.

84. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of Black Cottonwood, *Populus trichocarpa* (Torr. & amp; Gray). Science. 2006;313:1596–604.

85. Bredeson J V, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nat Biotechnol. 2016;34:562–70.

86. Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, et al. The Cassava Genome: Current Progress, Future Directions. Trop Plant Biol. 2012;5:88–94.

87. Muschner VC, Zamberlan PM, Bonatto SL, Freitas LB. Phylogeny, biogeography and divergence times in *Passiflora* (Passifloraceae). Genet Mol Biol. 2012;35 4 suppl 1:1036–43.

88. Melo NF, Cervi AC, Guerra M. Karyology and cytotaxonomy of the genus *Passiflora* L. (Passifloraceae). Plant Syst Evol. 2001;226:69–84.

89. Silvia MC, Vieira MLC, Mondin M, Aguiar-Perecin MLR. Comparative karyotype analysis of three *Passiflora* L. species and cytogenetic characterization of somatic hybrids. Caryologia. 2005;58:220–8.

90. Souza MM, Pereira TNS, Vieira MLC. Cytogenetic studies in some species of *Passiflora* L. (Passifloraceae): a review emphasizing Brazilian species. Brazilian Arch Biol Technol. 2008;51:247–58.

91. Snow N, MacDougal JM. New Chromosome Reports in Passiflora (Passifloraceae). Syst Bot. 1993;18:261.

92. Storey WB. Chromosomes numbers of some species of *Passiflora* occurring in Hawaii. Pac Sci. 1950;4:37–42.

93. Hansen AK, Gilbert LE, Simpson BB, Downie SR, Cervi AC, Jansen RK. Phylogenetic relationships and chromosome number evolution in *Passiflora*. Syst Bot. 2006;31:138–50.

94. Sader M. Mapeamento cromossômico do maracujá-azedo (*Passiflora edulis* Sims, Passifloraceae). Dissertação de mestrado.UFPE. Recife. 2016;48.

95. Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, et al. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. Nat Commun. 2013;4:1595.

96. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290:1151–5.
97. Wei F, Stein JC, Liang C, Zhang J, Fulton RS, Baucom RS, et al. Detailed analysis of a contiguous 22-Mb region of the maize genome. PLoS Genet. 2009;5:e1000728.

98. Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang T-H, et al. Architecture and evolution of a minute plant genome. Nature. 2013;498:94–8.

99. Fresnedo-Ramírez J, Martínez-García PJ, Parfitt DE, Crisosto CH, Gradziel TM. Heterogeneity in the entire genome for three genotypes of peach [*Prunus persica* (L.) Batsch] as distinguished from sequence analysis of genomic variants. BMC Genomics. 2013;14:750.

100. Bacci Jr. M, Soares RBS, Tajara E, Ambar G, Fischer CN, Guilherme IR, et al. Identification and frequency of transposable elements in *Eucalyptus*. Genet Mol Biol. 2005;28:634–9.

101. Yi F, Jia Z, Xiao Y, Ma W, Wang J. SPTEdb: a database for transposable elements in salicaceous plants. Database. 2018;2018:1–8.

102. Cavallini A, Natali L, Zuccolo A, Giordani T, Jurman I, Ferrillo V, et al. Analysis of transposons and repeat composition of the sunflower (*Helianthus annuus* L.) genome. Theor Appl Genet. 2010;120:491–508.

103. Oliver KR, McComb JA, Greene WK. Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol Evol. 2013;5:1886–901.

104. Zhang Q-J, Gao L-Z. Rapid and eccent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* Species. G3;Genes|Genomes|Genetics. 2017;7:1875–85.

105. Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, et al. Cassava genome from a wild ancestor to cultivated varieties. Nat Commun. 2014;5:5110.

106. Lyu H, He Z, Wu C-I, Shi S. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. New Phytol. 2018;217:428–38.

107. Xu Y, Du J. Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (*Solanum lycopersicum*) plants. Plant J. 2014;80:582–91.

108. Schulman AH, Wicker T. A field guide to transposable elements. In: Transposons and Genome Dynamics in Evolution (eds) Fedoroff NV. Plant. Oxford, UK: Wiley-Blackwell; 2013.15–40.

109. Poulter RTM, Butler MI. Tyrosine recombinase retrotransposons and Transposons. American Society of Microbiology. 2015;1271–91.

110. Piednoël M, Gonçalves IR, Higuet D, Bonnivard E. Eukaryote DIRS1-like retrotransposons: An overview. BMC Genomics. 2011;12:621.

111. Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. Genomics Inform. 2014;12:87–97.

112. Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. Genome Biol Evol. 2011;3:219–29.

113. Michael TP. Plant genome size variation: bloating and purging DNA. Brief Funct Genomics. 2014;13:308– 17. 114. Cossu RM, Buti M, Giordani T, Natali L, Cavallini A. A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. Tree Genet Genomes. 2012;8:61–75.

115. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, et al. Polyploidy and angiosperm diversification. Am J Bot. 2009;96:336–48.

116. Vicient CM, Casacuberta JM. Impact of transposable elements on polyploid plant genomes. Ann Bot. 2017;120:195–207.

117. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci U S A. 2014;111:5135–40.

118. Macas J, Novák P, Pellicer J, Čížková J, Koblížková A, Neumann P, et al. In Depth Characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe fabeae. PLoS One. 2015;10:e0143424.

119. Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novák P, Neumann P, et al. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol. 2015;208:596–607.

120. Plohl M. Those mysterious sequences of satellite DNA. Periodicum Biologorum. 2010;4:403-410.

APÊNDICE

APÊNDICE A. Conteúdo gênico de uma fração rica em genes do genoma de *P. edulis* e anotação estrutural [12]. Os insertos BACS com o maior número de genes foram usados para análises de microssintenia.

									Continua
BAC code	No. of predicte de genes	Intronless genes	Exons per gene	Gene length variation (bp)	Average gene length (bp)	Intergenic spacer length variation (bp)	Average intergenic spacer length (bp)	CDS length (bp)	Average CDS length (bp)
Pe101K14	36	17	2 - 17	264 - 11,778	2,720	33-6,312	2,070	264 - 6,576	1,187
Pe185D11	36	12	2 - 12	201 - 4,778	1,548	16-9,730	1,802	201 - 1,725	689
Pe164B18	29	9	2 - 19	243 - 16,279	2,313	42-7,449	1,316	243 - 11,409	1,393
Pe214H11	29	4	2 - 39	799 - 13,956	3,857	194 - 5,728	1,134	174 – 4,572	1,636
Pe164D9	28	13	2 - 11	198 - 5,817	1,868	114 - 5,844	1,600	156 - 2,202	1,066
Pe186E19	28	4	2 - 14	770 - 7,450	2,651	11 - 13,501	1,559	210-2,307	1,098
Pe43L2	27	3	2 - 18	339 - 10,097	2,718	162 - 2,768	973	279 - 3,123	1,145
Pe86F9	27	13	2 - 5	201 - 20,501	1,622	147 - 12,507	2,776	201 - 1,740	607
Pe164K17	26	4	2 - 13	436 - 9,502	3,037	11 – 7,775	1,761	204 - 5,334	1,310
Pe215I8	26	5	2 - 18	312 - 8,238	3,007	230 - 13,338	2,168	180 - 3,501	1,253
Pe75K15	25	14	2 - 5	186 - 4,193	857	10-11,721	2,951	186 - 2,100	591
Pe84I14	25	6	2 - 12	345 - 8,118	3,014	69 - 4,352	936	198 – 4,275	1,295
Pe84M23	25	5	2 - 13	305 - 8,652	2,753	52 - 5,197	998	177 – 3,018	1,168
Pe93M2	25	5	2 - 16	399 - 7,069	2,274	135 - 11,933	2,170	192 – 2,961	1,109
Pe171P13	25	8	2 - 20	461 - 9,727	2,759	158 - 15,960	2,392	330 - 4,035	1,193
Pe207D11	25	12	2 - 17	213 - 6,756	1,896	5 - 20,551	2,838	213 - 2,730	897
Pe93N7	24	5	2 - 11	921 - 8,889	3,120	18 – 7,588	1,421	387 - 5,085	1,486
Pe108C16	24	8	2 - 14	234 - 6,553	1,892	34 - 9,113	2,209	234 - 3,252	974
Pe173B16	24	6	2 - 32	475 - 15,390	3,079	151 - 15,127	2,134	279 - 6,375	1,523
Pe185J16	24	4	2 - 21	447 - 8,773	2,432	201 - 6,924	2,083	237 – 2,367	1,035
Pe198H23	24	8	2 - 6	180 - 5,279	1,943	1 - 11,008	2,681	180 - 3,510	1,143
Pe212I1	24	5	2 - 35	234 - 12,694	2,715	53 - 15,133	2,607	234 - 3,567	1,080
Pe93J9	23	3	2 - 16	615 - 6,131	2,907	3-9,066	1,824	201 - 3,321	1,295
Pe135J12	23	6	2 - 15	162 - 9,543	2,714	81 - 8,758	1,868	162 - 4,433	1,260
Pe195F4	23	2	2 - 20	261 - 8,364	2,843	9 - 11,133	2,208	177 – 5,442	1,192
Pe74I6	22	9	2 - 39	204 - 17,655	3,407	146 - 6,191	1,764	204 - 4,374	1,164
Pe84M18	22	6	2 - 10	321 - 8,124	2,563	22 - 15,224	2,364	321 - 4,356	1,160
Pe101O4	22	6	2 - 19	624 - 9,702	2,678	315 - 10,499	2,782	300 - 2,235	884
Pe141J23	22	6	2 - 15	189 – 9,258	2,567	608 - 12,079	2,407	189 - 2,550	870
Pe201C11	22	11	2 - 17	195 - 5,452	1,865	288 - 17,891	4,128	195 – 2,634	822
Pe69G18	21	3	2 - 22	228 - 8,658	2,958	61 - 19,104	2,304	210 - 3,582	1,192
Pe69H24	21	2	2 - 14	335 - 6,461	2,752	445 - 5705	2,306	234 - 2,559	1,142
Pe93K19	21	3	2 - 12	792 – 10,373	3,523	196 - 6,322	1,422	387 – 4,629	1,593
Pe125I23	21	5	2 - 14	414 - 7,993	2,526	51 - 8,659	2,406	414 - 1,776	1,106

BAC code	No. of predicte de genes	Intronless genes	Exons per gene	Gene length variation (bp)	Average gene length (bp)	Intergenic spacer length variation (bp)	Average intergenic spacer length (bp)	CDS length (bp)	Average CDS length (bp)
Pe164A12	21	7	2 - 11	384 - 7,964	2,354	26-7,406	1,675	228-4,503	1,050
Pe168B17	21	3	2 - 11	321 - 6,861	2,509	47 – 16,932	4,619	174 - 4,140	1,234
Pe214A18	21	7	2 - 11	243 - 6,314	1,944	237 - 27,586	3,916	243 - 2,184	924
Pe7M15	20	11	2 - 15	213 - 9,031	2,388	12 - 17,420	3,676	213 - 3,495	1,046
Pe28D11	20	16	2-4	189 – 2,430	780	22 - 28,073	5,567	189 – 1,410	558
Pe60G10	20	6	2 - 24	351 - 9,925	2,513	91 - 10,947	2,767	261 - 3,378	1,291
Pe65F7	20	8	2 - 14	306 - 7,081	1,973	12 - 25,539	2,702	213 - 3,252	844
Pe175N8	20	8	2 - 27	219 - 14,245	2,941	15 - 11,237	2,495	219 - 3,663	1,299
Pe214N19	20	9	2 - 13	234 - 5,913	1,594	37 – 15,598	3,485	189 – 2,470	788
Pe43D2	19	3	2 - 8	447 - 7,338	2,601	271 - 19,633	2,158	222-4,872	1,120
Pe51C2	19	5	2 - 16	357 - 8,889	3,603	493 - 6,756	2,110	357 - 5,088	1,520
Pe85B19	19	7	2 - 18	372 - 10,115	2,851	42-8,103	2,368	183 - 3,228	1,157
Pe101P7	19	3	2 - 20	234 - 8,484	3,742	16-2,340	963	234 - 2,712	1,247
Pe134H15	19	8	2 - 11	295 - 7,290	2,527	208 - 5,953	2,351	219 - 1,899	844
Pe216F3	19	2	2 - 37	393 - 14,151	3,198	241 - 3,160	914	393 - 8,943	1,626
Pe216F9	19	5	2 - 13	207 - 9,274	3,547	420 - 5,573	2,107	207 - 3,417	1,180
Pe20N3 + Pe64C12	18	5	2 - 12	441 - 6,941	2,557	266 - 10,519	2,009	276 – 2,364	1,223
Pe24G19	18	12	2 - 6	165 - 3,803	1,054	184 - 22,176	3,639	165 - 1,593	598
Pe69C7	18	7	2 - 22	210 - 8,505	3,745	132 - 18,029	2,165	210-4,164	1,450
Pe69O16	18	4	4 - 19	590 - 17,670	4,339	86 - 1,976	767	177 – 14,583	2,292
Pe212D7	18	7	2 - 36	171 - 21,131	2,654	415 - 20,035	4,436	171 - 9,330	1,229
Pe27H17	17	13	2 - 3	177 – 2,134	620	197 – 13,511	4,390	177 – 1,071	464
Pe85I9	17	5	2 - 12	207-8,578	2,908	334 - 20,210	2,892	207 - 1,956	1,107
Pe89E10	17	10	2 - 13	183 - 4,327	974	342 - 18,584	5,178	174 – 1,794	509
Pe101P13	17	4	2 - 21	666 - 13,552	4,437	90-4,941	1,072	210-2,307	1,261
Pe209G15	17	3	2 - 14	219 - 8,353	3,108	118 - 17,105	2,754	219-3,084	1,416
Pe21O15	16	7	2 - 13	189 - 4,570	1,512	106 - 14,572	3,633	156 - 1,902	595
Pe63J18	16	10	2 - 5	441 - 6,941	2,750	266 - 10,519	2,054	213 - 3,429	970
Pe84K8	16	3	2 - 18	162 - 12,356	3,570	178 – 4,867	1,891	162 - 2,295	1,072
Pe93M4	16	10	2 - 7	216 - 3,063	972	15 - 37,508	4,704	216 - 1,998	640
Pe117C17	16	11	2 - 12	153 - 6,852	979	7 – 18,168	5,302	153 - 1,188	414
Pe138G17	16	10	2 - 10	178 - 6,113	1,395	40 - 13,394	4,513	178 – 2,934	731
Pe141K8	16	4	2 - 24	1,053 - 11,592	4,060	283 - 5,091	2,179	387 – 3,975	1,653
Pe216B22	16	1	2 - 15	1013 - 8,815	3,931	47 – 19,862	3,119	795 – 3,768	1,575
Pe216I5	16	6	4 - 16	201 - 5,929	3,296	462 - 4,563	1,373	201 - 2,862	1,458
Pe61E2	15	4	3 - 12	231 - 8,598	3,100	223 - 18,187	3,244	231 - 2,103	973
Pe99P16	15	9	2 - 33	249 - 15,022	2,441	501 - 9,387	2,582	216-4,605	908
Pe123N8	15	5	2 - 22	163 - 10,051	2,938	70 - 13,306	39,979	163 - 2,397	1,028

BAC code	No. of predicte de genes	Intronless genes	Exons per gene	Gene length variation (bp)	Average gene length (bp)	Intergenic spacer length variation (bp)	Average intergenic spacer length (bp)	CDS length (bp)	Average CDS length (bp)
Pe3F10	14	4	2 - 14	652 - 6,552	2,471	90 - 4,389	1,557	285 - 3,252	1,080
Pe28E22	14	1	2 - 12	379 - 11,107	3,661	13 - 16,073	2,221	261 - 2,718	1,247
Pe34M7	14	6	2-4	225 - 1,298	652	82 - 39,701	6,611	192 – 1,026	459
Pe75F20	14	6	2 - 13	198 - 6,418	1,859	182 - 21,979	5,567	198 - 1,842	541
Pe85H4	14	1	2 - 51	489 - 22,481	3,938	178 – 17,578	2,764	300 - 5,706	1,546
Pe85J23	14	2	2 - 15	760 - 9,631	3,222	362 - 9,609	2,597	492 - 3,066	1,087
Pe101H15	14	10	2 - 5	225 - 24,687	2,257	122 - 15,195	6,521	255 - 1,008	524
Pe69F22	13	0	2 - 14	438 - 6,597	3,680	196 - 26,118	4,433	207 - 1,710	1,029
Pe75A21	13	8	3 - 10	162 - 5,730	1,569	10 - 15,569	4,038	162 – 2,076	630
Pe84M6	13	8	2 - 13	185 - 3,026	1,059	262 - 16,455	4,686	185 – 1,578	792
Pe86H7	13	7	2 - 3	213 - 4,497	1,429	31 - 28,575	6,964	213 - 3,459	875
Pe34H9	12	3	2 - 14	258-6,285	1,961	49 - 44,532	6,154	258 - 1,623	748
Pe213C9	12	8	2 - 5	327 - 3,599	1,246	213 - 31,653	7,880	234 - 2,016	749
Pe71E3	11	2	3 - 9	207 - 3,727	2,185	362 - 31,489	6,138	207 - 1,698	1,047
Pe93A7	11	8	2 - 4	162 - 1,374	582	18 – 25,472	7,604	162 - 759	373
Pe93F5	11	2	2 - 8	192 - 11,041	2,745	5-24,167	7,152	192 - 1722	707
Pe93O18	11	3	2 - 11	387 - 7,643	2,714	596 - 49,482	9,337	387 - 1,632	1,080
Pe101F21	11	7	2 - 7	243 - 4,835	947	58-27,172	8,438	198 – 1,806	534
Pe141B12	11	4	2 - 15	288 - 6,769	2,412	251 - 24,611	5,214	282 - 3,417	1,142
Pe75D12	10	6	2 - 5	219 - 3,255	778	109 - 39,945	8,052	216 - 1,224	456
Pe75N15	10	8	2	204 - 714	444	78 - 32,243	7,353	204 - 714	402
Pe9E4	9	4	2 - 14	342-6,100	2,099	654 - 13,925	6,177	342 - 2,898	1,171
Pe15E1	9	4	2 - 13	270 - 2,896	1,153	700 - 33,021	9,014	270 - 1,578	714
Pe20E10	9	4	2 – 2	159 – 1,578	605	278 - 35,112	9,958	159 – 1,578	496
Pe212M5	9	4	2 - 6	267 - 3,170	1,020	851 - 10,468	4,056	267 - 1,566	727
Pe103M2	8	2	2 - 17	222 - 12,656	3,122	418 - 32,453	6,547	222 - 2,010	807
Pe28I20	7	5	2 – 2	237 - 881	467	67 – 30,516	11,363	237 - 762	437
Pe75F13	7	4	2-3	180 - 1,636	654	16,743 – 92,497	58,535	180 - 1,245	519
Pe85O9	7	1	2 - 8	441 - 3,324	2,079	515-6,447	1,784	441 - 1,329	765
Pe1M17	6	1	2-4	312 - 2,473	1,099	256 - 10,848	5,311	312 - 1,404	784
Pe212J12	6	2	2 - 15	405 - 4,357	1,377	81 - 12,708	3,133	381 - 1,644	692
Pe216B2	6	1	2 - 24	218 - 15,969	5,097	830-4,575	2,306	218 - 3,819	1,605
Pe113A7	5	3	2	156 - 2254	1,206	3472 -26,026	13,464	156 - 681	503
Pe1K19	3	0	2 - 9	958 - 4,737	3,111	287 - 37,487	18,877	840 - 897	869
Pe33M2	3	2	3	210-2,037	824	4,001 - 69,199	36,600	210 - 697	377