

**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**

## 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 desenvolvimento 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 espaçamentos 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 from *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 synteny 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. trichocarpa* 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*

## 1. INTRODUÇÃO

A família Passifloraceae pertence à ordem Malpighiales e é membro do clado das Rosidae [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, 147 são de ocorrência no Brasil, sendo 85 exclusivas, 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 *lipxygenase* 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 acessíveis 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 de dados Phytozome (v12.1.6) sequências genômicas de 93 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 codificadoras 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 Rosídeas [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 indicam 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 homólogas 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, com evidências filogenéticas relativamente 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. CONCLUSÕES

O método aqui utilizado para a identificação de regiões de microssintênicas representa um recurso importante e 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 IV. *J Linn Soc.* 2016;181:1–20.
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. 14

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.