UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

DROSOPHILA INCOMPTA, UMA ESPÉCIE DE ECOLOGIA RESTRITA, COMO MODELO DE ESTUDOS EM EVOLUÇÃO LIGADA A RESTRIÇÕES DE NICHO

PEDRO MESQUITA FONSECA

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de **Doutor em Genética e Biologia Molecular.**

Orientador: Élgion Lúcio da Silva Loreto

Co-orientadora: Lizandra Jaqueline Robe

Porto Alegre, Dezembro de 2020.

“Al fin y al cabo, somos lo que hacemos
para cambiar lo que somos.”

Eduardo Galeano

INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi desenvolvido no Laboratório de Genética e Biologia Molecular (LabDros) na Universidade Federal de Santa Maria. O autor desta tese contou com apoio de uma bolsa de doutorado do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). A parte experimental deste trabalho foi financiada por recursos do LabDros/UFSM e do Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM) da UFRGS.

À Ana Luisa.

AGRADECIMENTOS

Além de dedicar todo o meu trabalho à minha filha, é a ela que agradeço primeiramente. Agradeço por renovar a cada dia minha vontade de ser melhor e de fazer melhor por e para ela.

Poderia dedicar páginas e páginas só para agradecer às mulheres presentes em minha vida e às quais devo tudo que tenho, tudo que sou e tudo que fiz. Às minhas mães de coração Nana e Dulce, que sem querer me ensinaram o verdadeiro sentido de altruísmo, de fazer tudo pelos que se ama sem pensar em nada em troca. À minha irmã Mariana, meu porto seguro, minha edificação e inspiração. À minha esposa Larissa, pelo companheirismo, cumplicidade e divisão de sonhos, batalhas e conquistas. À minha avó Joanna e minha Dinda Branca, que mesmo de longe geograficamente estão sempre presentes em minha jornada. À minha prima Alice que desde cedo de certa forma me preparou para ser pai. À minha sogra Roselaine e minha cunhada Émile pela confiança e apoio em todos os sentidos. Às minhas amigas e comadres Gabriela, Camilla, Francine e Michelle pela amizade e carinho, mesmo com a distância e isolamento atuais. À minha Co-orientadora e mentora Lizandra, que além de todos os ensinamentos acadêmicos me inspira para a vida. Por fim, à memória de minha mãe Claudia, que sempre me inspirou pelo carinho da lembrança e legado de afeto. À todas vocês, muito obrigado.

Agradeço também ao meu orientador Élgion, pela confiança desde a graduação e por me fazer um amante da ciência, da genética e evolução.

Agradeço aos meus amigos/compadres/irmãos de sempre Thiago, Vinícius, Rodrigo, Filipe, Fernando e ao meu cunhado Elton por dividirem histórias de vida e planos futuros, de companheirismo e amizade.

Agradeço aos meus colegas de LabDros ‘das antiga’ Paloma, Valéria, Sinara, João, Gabriel, Mauro, Stela, Marcos, José, Geovani, Nader, Raquel, Karen, Tailini, Carla e Ronaldo, pelos ensinamentos e companhia no laboratório, em congressos, coletas e no cafofo, por dividirem o gosto pela biologia e por nossa amizade.

Agradeço também aos meus atuais colegas de Labdros Tiago, Daniel, Mônica, Taís, Mariana, Larissa, Dani, Tuane, Estéfani e Kátia e a todo o pessoal do Labdros de Porto Alegre, por todo o apoio e parcerias.

Sou muito orgulhoso e agradecido de fazer parte de um programa de referência e excelência nacional, portanto agradeço a todos os professores e em especial às professoras Vera e Maríndia, pelo acolhimento e inspiração.

Agradeço muito ao Elmo, nosso anjo da guarda de nosso PPG, pela atenção e cuidado para com os alunos, bem como por toda ajuda.

Agradeço aos membros titulares de minha banca, professora Karen, professora Mônica e professor Daniel por dedicarem um pouco de seu tempo e expertise para avaliar nosso trabalho. É uma honra tê-los como meus avaliadores.

Por fim, sou grato a todos que de certa forma tornaram meu caminho até aqui mais fácil, mais prazeroso e me fizeram ter mais vontade e orgulho de trabalhar até aqui. Obrigado.

SUMÁRIO

LISTA DE ABREVIATURAS E UNIDADES

Abreviaturas

GR: Gustatory Receptor (receptor gustatório)

OR: Odorant Receptor (receptor olfativo)

OBP: Odorant Binding Protein (proteína carregadora de odores)

ORCO: OR Coreceptor (correceptor de OR)

TE: Transposable Element (elemento transponível)

mRNA: Messenger RNA (RNA mensageiro)

siRNA: Small Interfering RNA (pequeno RNA de interferência)

piRNA: Piwi Interacting RNA (RNA associado à Piwi)

cDNA: Complementary DNA (DNA complementar)

Rag1/Rag2: Recombination Activating Gene 1/2 (Gene de ativação de recombinação)

HTT: Horizontal Transposon Transfer (Transferência horizontal de elemento transponível)

Unidades

pb/bp: Pares de base

ml: mililitros

g: gramas

aa: Amino ácidos

cm: Centímetros

RESUMO

Esta tese apresenta dois artigos associados a avaliação de padrões de evolução genômica que utilizam como modelo de estudo *Drosophila incompta*, uma espécie de drosofilídeos da ordem Diptera dos insetos, com registros bem distribuídos pelos Neotrópicos, a qual se notabiliza pela especificidade no uso de plantas do gênero *Cestrum* da família Solanaceae como sítios de corte, oviposição e alimentação de imaturos e adultos. No primeiro artigo fizemos a anotação e análise evolutiva do repertório de duas famílias gênicas de quimiorreceptores: receptores gustatórios (GR) e receptores olfativos (OR), além de outros dois grupos gênicos: *Yellow* e *nAChR*. Quimiorreceptores são a principal interface de um organismo com o ambiente em que este está inserido, sendo um bom foco de estudos evolutivos, especialmente para se investigar as pressões evolutivas que uma ecologia de nicho restrito impõe sobre espécies especialistas. Além do repertório gênico de *D. incompta*, o de outras 29 espécies do gênero *Drosophila* também foram contemplados a título de comparação, visto que este gênero é notabilizado por conter espécies com uma ampla gama de diferentes nichos, incluindo espécies especialistas e generalistas. Conseguimos verificar tendências de perdas gênicas ancestrais relacionadas à filogenia das espécies de *Drosophila*, pois conseguimos mostrar que boa parte dos genes presentes em espécies do subgênero *Sophophora* não estão presentes nas espécies não pertencentes a esse subgênero. Identificamos também sinais de seleção positiva em genes de *D. incompta*, indicando possível relação desses sinais evolutivos com a ecologia restrita desta espécie. Ainda, fizemos um experimento de preferência olfativa no qual testamos a preferência de *D. incompta* por odores relacionados às inflorescências de *Cestrum* contra odores já conhecidos como preferidos por outras espécies de *Drosophila*, e os resultados reforçam a clara preferência pelos odores da planta a qual *D. incompta* possui estrita relação. No segundo artigo caracterizamos e analisamos o repertório de elementos transponíveis (TEs) de *D. incompta*: seu mobiloma. Embora várias espécies de

Drosophila já contem com seus genomas sequenciados, poucas têm anotações criteriosas de seus mobilomas, os quais são considerados excelentes ferramentas para estudo de evolução molecular dessas espécies por parecerem responder a pressões evolutivas diversas, inclusive as relacionadas com a ecologia das espécies hospedeiras. Caracterizamos um total de 277 TEs dos quais 164 novos. Também analisamos a paisagem de TEs de *D. incompta* e de outras 31 espécies de *Drosophila*, onde não confirmamos a hipótese de que a paisagem de TEs é influenciada pela amplitude de nicho. Contudo achamos indicativos de diferenças entre as paisagens de espécies do gênero *Sophophora* e espécies não pertencentes a este gênero, indicando relação da disposição das paisagens com a filogenia das espécies. Ainda nesse artigo conseguimos verificar sinais de transferência horizontal de TEs (HTT) em elementos putativamente autônomos de *D. incompta*. Com esses estudos, foi possível, pois, corroborar a importância das espécies de ecologia restrita nas investigações de sinais de evolução molecular e evidenciar a presença de sinais evolutivos contrastantes em diferentes subgêneros de *Drosophila*.

Palavras-chave: quimiorreceptores; OR; GR; ecologia restrita; TE; mobiloma; HTT.

ABSTRACT

This thesis presents two scientific articles associated with an evaluation of patterns of genomic evolution that use *Drosophila incompta* as a study model, a Diptera: Drosophilidae species of insects, with wide distributed records along the Neotropics, which is notable for the specificity in the use of plants of the *Cestrum* genus (Solanaceae) as courtship, oviposition and larvae and adults feeding site. In the first article, we made the annotation and evolutionary analysis of the repertoire of two chemoreceptors gene families: gustatory receptors (GR) and odorant receptors (OR), in addition to two other gene groups: *Yellow* and *nAChR*. Chemoreceptors are the main interface of an organism with the environment in which it is inserted, being a focus of evolutionary studies, especially to investigate the evolutionary forces that a restricted niche ecology imposes on specialist species. In addition to the repertoire of *D. incompta* chemoreceptors, the repertoire of 29 other species of the genus *Drosophila* were also considered for comparison, since this genus is notable for containing species with a wide range of different niches, including specialist and generalist species. We were able to verify signals of ancestral gene losses related to the phylogeny of *Drosophila*, as we were able to show that a considerable amount of the genes present in species of the subgenus *Sophophora* are not present in non-*Sophophora* species. We also identified signals of positive selection in chemoreceptors of *D. incompta*, indicating a possible relation between these evolutionary clues and the restricted ecology of this species. We also did an olfactory choice experiment in which we tested the preference of *D. incompta* for odorants related to *Cestrum* flowers against odorants already known as preferred by other *Drosophila* species, and the results reinforce the clear preference for odorants from the plant which *D. incompta* has a strict relationship. In the second article, we characterize and analyze the repertoire of transposable elements (TEs) of *D. incompta*: its mobilome. Although several *Drosophila* species already have their genomes sequenced, a few of them have carefully annotated mobilomes, which are considered excellent tools for studying the molecular evolution of these species

because they seem to respond to different evolutionary pressures, including those related to the ecology and niche breadth of host species. We characterize 277 TEs, of which 164 brand new ones. We also analyzed the TE landscape of *D. incompta* and 31 other species of *Drosophila*, and we cannot confirm the hypothesis that the TE landscape is influenced by niche breadth. However, we found evidence of differences between the landscapes of *Sophophora* and non-*Sophophora* species, indicating a relation between the landscapes curves and the phylogenetic relation of *Drosophila* species. Still in this article, we were able to verify evidences of horizontal transposon transfer (HTT) in some putatively autonomous elements of *D*. *incompta*. With these studies, it was possible, therefore, to corroborate the value of species of restricted ecology in the investigation of signals of molecular evolution and to evince the presence of contrasting evolutionary signals in different subgenera of *Drosophila*.

Keywords: Chemoreceptors; OR; GR; restrict ecology; TE; mobilome; HTT.

CAPÍTULO I

INTRODUÇÃO GERAL

1. FAMÍLIA DROSOPHILIDAE

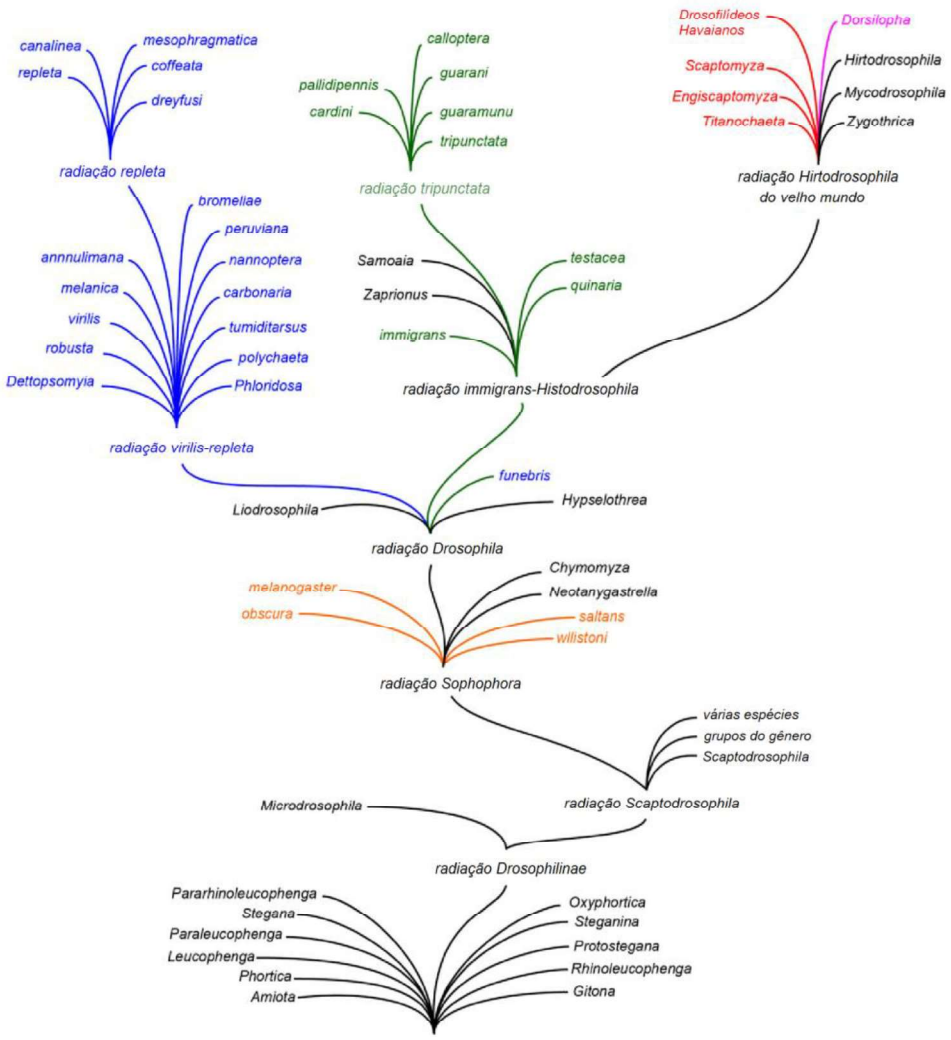
Dentro da ordem Diptera de Insecta, a família Drosophilidae está entre as mais diversas, contando com 4.547 espécies distribuídas em 77 gêneros (BÄCHLI, 2020). Por sua vez, os gêneros são agrupados em duas subfamílias: Drosophilinae e Steganinae. Segundo dados biogeográficos e evidências fósseis (THROCKMORTON, 1975), a origem dessa família remete a regiões tropicais do Velho Mundo e datam do Eoceno. Esse mesmo autor ainda sugere que a evolução desse grupo se deu a partir de ancestrais que utilizavam fungos crescidos no folhiço como principal fonte de alimento. Com a diversificação de substratos, aliada a susceptibilidade à evolução de nicho, oportunidades de adaptações teriam surgido (THROCKMORTON 1975; ROBE et al., 2010). O sucesso evolutivo da família torna-se evidente perante sua diversidade de espécies e nichos, bem como pela sua ampla distribuição.

O gênero *Drosophila* compreende o maior gênero da família, contando atualmente com 1.644 espécies descritas (BÄCHLI, 2020). Este gênero é amplamente estudado em diversas áreas da ciência, em especial na biologia evolutiva (PRUD'HOMME; GOMPEL, 2010, 2011; ROBE et al., 2013; KUNTZ; EISEN, 2014; WANGLER et al., 2015; UGUR et al., 2016; O'GRADY; DESALLE, 2018). *Drosophila* é um gênero parafilético com relação a diferentes gêneros de Drosophilidae, como *Lordiphosa, Samoaia, Scaptomyza* e *Zaprionus* (THROCKMORTON 1975; DESALLE; GRIMALDI, 1991; Robe et al. 2005; REMSEN; O'GRADY, 2002; O'GRADY et al., 2011). Este gênero encontra-se subdividido em nove subgêneros (O'GRADY; DESALLE, 2018) (Figura 1), dentre os quais o subgênero *Sophophora* é o mais estudado, enquanto o subgênero *Drosophila* é o mais diverso (O'GRADY; DESALLE, 2018; BÄCHLI, 2020).

Drosophila também é o gênero de Drosophilidae mais abundante no Brasil, com mais de 181 registros (GOTTSCHALK et al., 2008; TIDON et al., 2015). No entanto, considera-se que esse número seja uma subestimativa (DE RÉ, 2016), dado

o viés de capturas relacionadas ao método mais utilizado para a coleta de espécimes (TIDON; SENE, 1988), que contempla, em sua maioria, drosofilídeos atraídos por frutas fermentadas e/ou em decomposição em detrimento das espécies que utilizam outros recursos. Apesar disso, mais recentemente, outros métodos que contemplam grupos de espécies com outros hábitos também vêm sendo empregados (ROBE et al., 2014; MACHADO et al., 2016; DE RÉ et al., 2017; BARRIOS-LEAL et al., 2018; SCHMITZ; VALENTE, 2019; GAUTÉRIO et al., 2020).

Neste sentido, sabemos atualmente que o tipo preferencial de recursos, bem como o nível de amplitude de nicho pode ser bastante diverso entre diferentes espécies de drosofilídeos. Dentre os hábitos ecológicos mais comumente observados na família, destacam-se a frugivoria, a fitofagia, a micofagia e a antofilia (MARKOW; O'GRADY, 2008). Dentro de cada um destes nichos, encontramos desde espécies generalistas, que utilizam vários grupos taxonômicos como recursos, até espécies bastante especialistas (GERLACH, 2009; CORDEIRO et al., 2020; GAUTÉRIO et al., 2020). Neste âmbito, é interessante observar que mesmo *Drosophila melanogaster,* comumente apontada como generalista (DAVID; CAPY, 1988; LACHAISE; SILVAIN, 2004; MARKOW, 2015), parece ter apresentado, em sua origem, ecologia restrita a frutas de marula (*Sclerocarya birrea*) (Anacardiaceae), hábito este que ainda se manifesta sazonalmente em algumas regiões (MANSOURIAN et al., 2018). Entre as espécies especialistas, temos exemplos bem documentados, como *D. sechellia,* a qual depende intimamente dos frutos de *Morinda citrifolia* (Rubiaceae) para alimentação (JONES, 2005; MCBRIDE, 2007), e *D. erecta*, com sua associação a frutos de *Pandanus spp.* (Pandanaceae) (LACHAISE; TSACAS, 1974; RIO et al., 1983; LINZ et al., 2013). Além dessas, são conhecidas, dentro de *Drosophila*, espécies fitófagas com ecologia restrita a cactos. Este estado é bastante comum dentro do grupo *repleta*, sendo encontrado, por exemplo, em *D. mojavensis*, que utiliza cactos como local de acasalamento e fonte de alimento (MATZKIN, 2014) e *D. buzzatii*, que utiliza cactos em decomposição como fonte de alimento das larvas (PEREIRA et al., 1983; MANFRIN; SENE, 2006).



Throckmorton (1975)	Yassin (2013)
Drosofilídeos Havaianos	Idiomyia (gênero)
radiação virilis-repleta	Drosophila (Siphlodora)
radiação immigrans-tripunctata	Drosophila (Drosophila)
Drosophila (Sophophora)	Drosophila (Sophophora)
Drosophila (Dorsilopha)	Drosophila (Dorsilopha)

Figura 1 – Filogenia proposta por Throckmorton (1975) para a família Drosophilidae com as correspondências indicadas pelas cores e a nova nomenclatura proposta por Yassin (2013). Adaptado de O'Grady e DeSalle (2018).

1.1 GRUPO *flavopilosa* DE *Drosophila*

O grupo *flavopilosa* foi proposto no início da década de 60 por Wheeler et al. (1962). Este grupo pertence à radiação *virilis-repleta* do subgênero *Drosophila* (MALOGOLOWKIN, 1953; THROCKMORTON, 1975), atualmente considerada como um subgênero a parte, o subgênero *Siphlodora* (O'GRADY; DESALLE, 2018). No entanto, ainda existem controvérsias acerca do posicionamento filogenético do grupo *flavopilosa*, que parece constituir o grupo-irmão do grupo *repleta* ou do grupo *virilis* (DE RÉ et al., 2017). Dezoito espécies são consideradas, até agora, como pertencentes ao grupo *flavopilosa*, sendo elas: *D. acroria, D. crossoptera, D. lauta, D. cordeiroi, D. cestri, D. flavopilosa* e *D. hollisae*, pertencentes ao subgrupo *flavopilosa;* e *D. incompta, D. melina, D. nesiota, D. gentica* e *D. mariahelenae,* pertencendo ao subgrupo *nesiota,* além de outras seis espécies não alocadas a nenhum subgrupo (GUILLÍN; RAFAEL, 2018; BÄCHLI, 2020).

Registros de espécimes do grupo *flavopilosa* são encontradas em vários pontos dos Neotrópicos, desde o sul do Chile até a região central do México (ROBE et al., 2013; BÄCHLI, 2020). No Brasil, os estudos com esse grupo começaram com Brncic (1978), com coletas no estado do Rio Grande do Sul. Atualmente, seis espécies do grupo apresentam registros para o Brasil: *D. cestri*, *D. cordeiroi*, *D. flavopilosa*, *D. hollisae*, *D. incompta*, e *D. mariahelenae*. No entanto, esse número possivelmente está subestimado, devido ao baixo esforço amostral, que ainda não contempla por igual todas as regiões de ocorrência dos recursos preferenciais do grupo (ROBE et al., 2013). A monofilia recíproca dos subgrupos *flavopilosa* e *nesiota* foi confirmada recentemente para cinco das seis espécies do grupo que ocorrem no Brasil (MOREIRA, 2020).

Drosofilídeos pertencentes a esse grupo apresentam a particularidade de restrição de nicho relacionada a flores de algumas espécies do gênero *Cestrum* (Solanaceae) (Figura 2), além de *Sessea brasiliensis* (Solanaceae) (BRNCIC, 1966; HOFMANN, 1984 SANTOS; VILELA, 2005; ROBE et al., 2013). Essa ecologia restrita se apresenta na forma de oviposição exclusiva nas plantas de *Cestrum*, podendo ser em inflorescências desabrochadas ou ainda fechadas, que são ainda utilizadas como

sítio de alimentação de imaturos e adultos (WHEELER et al., 1962; BRNCIC, 1983; LUDWIG et al., 2002). Diferentes espécies do grupo *flavopilosa* apresentam uma série de adaptações morfológicas à exploração de *Cestrum*, entre as quais destaca-se a coloração, críptica às flores de *Cestrum*, e o formato especializado do ovopositor das fêmeas, que é robusto e apresenta e espinhos que facilitam a escarificação das superfícies onde os ovos serão postos (ROBE et al., 2013). Esta última característica também é encontrada em outros drosofilídeos antófilos monófagos (PIPKIN et al., 1966), refletindo uma potencial adaptação à especialização.

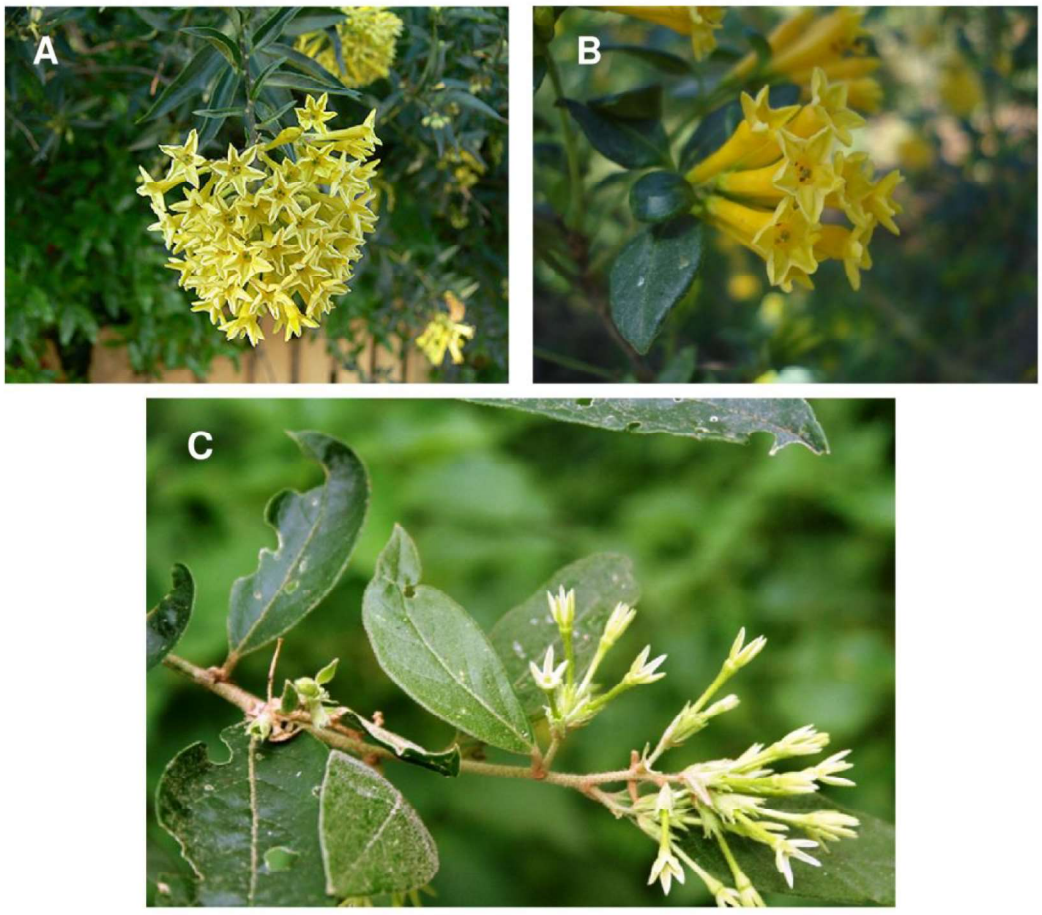


Figura 2 – Fotografias de inflorescências de três espécies do gênero *Cestrum* que foram registradas no Rio Grande do Sul. **A:** *C. parquii*; **B:** *C. euanthes*; **C:** *C. strigilatum.* Fonte arquivo pessoal.

Por estarem intimamente ligadas às plantas hospedeiras, espécies do grupo *flavopilosa* têm a sua distribuição afetada por variáveis ligadas à distribuição e à floração das espécies de *Cestrum* (ROBE et al., 2013). No entanto, devido à natureza efêmera e inconstante da floração de *Cestrum* ao longo do ano, Sepel et al. (2000) sugeriram que populações de *Drosophila* do grupo *flavopilosa* podem sofrer consecutivos *bottlenecks* seguidos de expansões populacionais. Além disso, esses mesmos autores sugerem que a taxa de ocupação das flores de *Cestrum* é diferenciada entre as espécies, mesmo no período de abundância de recursos, o que também poderia contribuir para os eventos de redução no tamanho populacional. Devido ao potencial efeito destes agentes 'estressores', Napp e Brncic (1978) sugeriram que, as populações dessas espécies deveriam apresentar baixos níveis de variabilidade genética. Isso foi posteriormente refutado por estudos com isozimas, inversões cromossômicas e até análises de polimorfismo em nível de genes individuais ou mesmo a nível de genoma mitocondrial (ROBE et al., 2013; DE RÉ et al., 2014).

1.2 *Drosophila incompta*

Drosophila incompta (Figura 3) pertence ao subgrupo *nesiota* do grupo *flavopilosa* de *Drosophila*, sendo encontrada desde a Argentina até o México (ROBE et al., 2013; BÄCHLI, 2020). Na região sul do Brasil, em particular, essa espécie é frequentemente encontrada em simpatria e até mesmo em sintopia com outras três espécies do subgrupo *flavopilosa*: *D. flavopilosa*, *D. cordeiroi* e *D. cestri* (ROBE et al., 2013). No entanto, parece existir uma segregação temporal no padrão de oviposição destas espécies: enquanto *D. incompta* parece preferir flores em estágios mais avançados de floração, outras espécies do grupo *flavopilosa*, como *D.* cestri, costumam ovipor em flores ainda fechadas ou em botão (NAPP; BRNCIC, 1978; DE RÉ, 2016). Além disso, frequentemente, flores de *Cestrum* apresentam registros simultâneos para *D. incompta* e diferentes espécies de *Zygothrica*, em especial àquelas pertencentes ao grupo *vittimaculosa* (VILELA, 1983; SANTOS; VILELA,

2005; FONSECA et al., 2017). Neste último caso, a eclosão tardia dos espécimes de *Zygothrica* também suportam a existência de um padrão de segregação temporal no uso de recursos (observação pessoal).

Figura 3 – Fotografia de um espécime macho de *Drosophila incompta*. Adaptado de De Ré et al., 2014.

Até o momento, *D. incompta* já foi coletada em associação a nove espécies de *Cestrum*: *C. amictum, C. corymbosum, C. euanthes, C. intermedium, C. nocturnum, C. parqui, C. schlechtendalii, C. sendtnerianum* e *C. strigilatum*, além de *Sessea brasiliensis* (SANTOS; VILELA, 2005; FONSECA et al., 2015). Esta espécie também se mostra mais tolerante a variações de precipitação e temperatura em comparação às outras espécies do grupo que co-ocorrem no Brasil (ROBE et al., 2013). Possivelmente, essas propriedades contribuem não apenas para um aumento na

área de distribuição potencial e real da espécie, como também para a grande abundância com que a mesma é frequentemente encontrada em coletas de *Cestrum* realizadas na região sul do Brasil.

2 EVOLUÇÃO MOLECULAR E GENÔMICA NO GÊNERO *Drosophila*

Os avanços recentes no campo de sequenciamento de larga escala ou sequenciamento de nova geração tem facilitado o acesso a genomas e transcriptomas completos, o que por sua vez aumenta a gama de possibilidades de estudos evolutivos, especialmente em organismos modelos. Espécies do gênero *Drosophila* sempre foram conhecidas como ótimos modelos na investigação de diversos aspectos da evolução, como em genética de populações, evolução ecológica, mecanismos de especiação, filogenia, evolução molecular e biologia do desenvolvimento (POWELL, 1997). Um dos fatores que torna esse grupo tão especial para diferentes enfoques é o fato dele conter espécies com diferentes propriedades ecológicas, contemplando desde espécies generalistas até espécies muito especialistas, como é o caso de *D. incompta.* Por outro lado, a grande disponibilidade de genomas previamente caracterizados para diferentes espécies do gênero certamente tem auxiliado na consolidação deste táxon em diferentes estudos voltados para a evolução genômica.

2.1 EVOLUÇÃO DE RECEPTORES SENSORIAIS

A principal forma de interação dos animais com o habitat onde estão inseridos é através de sistemas sensoriais, seja para localizar uma fonte de alimento, encontrar um parceiro reprodutivo ou até mesmo afastar potenciais predadores. Mais uma vez, *Drosophila* tem sido um excelente modelo para a investigação das formas de comunicação dos espécimes com o seu habitat, ou mesmo da evolução dos mecanismos sensoriais e dos genes a eles associados. Neste sentido, estudos sobre a evolução molecular associada a quimiorrecepção em *Drosophila* (MCBRIDE, 2007; ALMEIDA et al., 2014; MISSBACH et al., 2014) tem mostrado resultados

interessantes, tanto do ponto de vista qualitativo, quanto no âmbito quantitativo. Estudos semelhantes também têm sido realizados com genes associados à pigmentação (WITTKOPP et al., 2009; TAKAHASHI, 2013; REBEIZ; WILLIAMS, 2017), de detoxifixação (MATZKIN et al., 2006; GLOSS et al., 2014; ETGES, 2019).

Dentre os quimiosensores, sugere-se que os receptores de odores ou receptores olfativos (OR) (Figura 4) e aqueles responsáveis pela percepção dos sabores, ou receptores gustativos (GR), estejam mais predispostos a experimentarem alterações associadas a mudanças de nicho (MCBRIDE, 2007). Como referência, *D. melanogaster* possui 121 genes codificadores desses quimiorreceptores, 60 para cada família gênica e mais um correceptor (ROBERTSON et al., 2003). Contudo, esse número é variável tanto internamente, no gênero *Drosophila,* quanto em outros insetos. Em *Bombyx mori*, por exemplo, foram encontrados 48 ORs (WANNER et al., 2007) e 65 GRs (WANNER; ROBERTSON, 2008), enquanto na formiga *Linepithema humile* foram identificados, apenas para ORs, 367 genes (SMITH et al., 2011a).

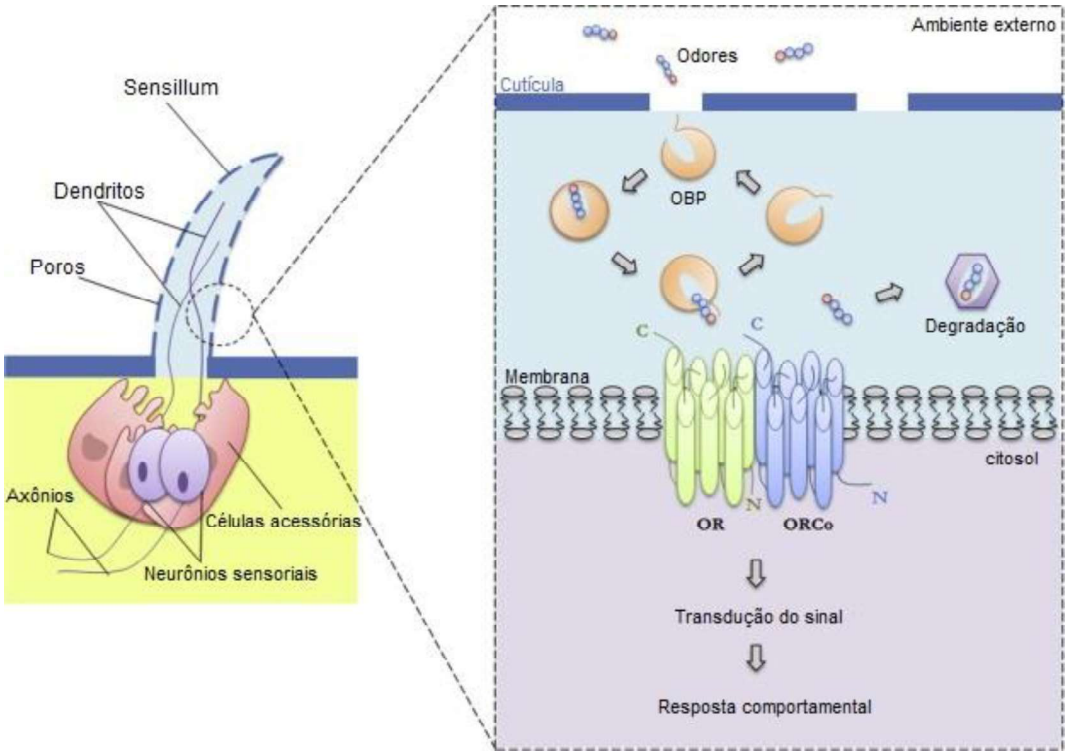


Figura 4 – Esquema genérico do processo de percepção de odores em *Drosophila*. **OBP:** Proteína carregadora de odores (do inglês *Odorant Binding Protein*); **OR:** Receptor olfativo genérico; **ORCO:** Cofator dos receptores olfativos. Figura modificada de Brito et al., 2016.

Estudos como os de Gardiner et al. (2008) apontam que, embora GRs e ORs tenham relações quanto sua origem, os mesmos possuem propriedades moleculares diferentes e também respondem a processos evolutivos de formas diferentes. Trabalhos com espécies especialistas de *Drosophila* como *D. sechellia* (MCBRIDE, 2007) e *D. erecta* (MCBRIDE; ARGUELLO, 2007), espécies especialistas nas frutas de *Morinda citrifolia* e em espécies do gênero *Pandanus*, respectivamente, tem mostrado tendência de perda gênica desses receptores relacionadas a restrição de nicho.

Neste âmbito, devido a sua ecologia restrita, *D. incompta* constitui um bom modelo para investigação das mudanças evolutivas potencialmente associadas à especialização ecológica. Além disso, a disponibilidade do genoma de pelo menos 30 outras espécies do gênero fornece uma oportunidade única de comparação dos padrões encontrados nessa espécie com espécies próximas, ecologicamente similares ou não, e sua potencial associação com diferentes pressões seletivas.

2.2 ELEMENTOS TRANSPONÍVEIS

Bárbara McClintock, no final da década de 40, estudando padrões de coloração de sementes de milho, foi a descobridora desse tipo único de sequências de DNA, que têm como principal característica a capacidade de mudarem suas posições nos genomas (MCCLINTOCK, 1948, 1956). A comunidade científica demorou para perceber a importância desse novo campo de pesquisa, e somente em 1983 McClintock teve o reconhecimento que merecia, sendo indicada ao prêmio Nobel de Fisiologia e Medicina por sua descoberta dos elementos transponíveis (TEs). Por muito tempo chamados de 'DNA-lixo' (*junk-DNA)* (OHNO, 1972), os TEs já não são mais considerados apenas genes saltadores, mas importantes componentes

dos genomas e fundamentais para a evolução dos mesmos (BIÉMONT, 2010; KIM et al., 2012; SCHRADER, SCHMITZ, 2019; GILBERT et al., 2020).

Presentes em praticamente todos os organismos vivos (WICKER et al., 2007) e até no genoma de alguns vírus de DNA (SUN et al., 2015), os TEs geralmente variam de 100 a 10.000 pb (WELLS; FESCHOTTE, 2020), mas podem ser muito maiores (ARKHIPOVA; YUSHENOVA, 2019). Sua contribuição para a composição do genoma do hospedeiro pode variar de 1%, em alguns insetos (KELLEY et al., 2014), até aproximadamente 90%, no caso de variedades de milho (*Zea mays*) (SANMIGUEL et al., 1996; SCHNABLE et al., 2009;). Em humanos, praticamente metade do genoma é composto por TEs (KONING et al., 2011). Contudo, a porcentagem de TEs em um genoma e consequentemente o tamanho desse genoma não estão ligados à complexidade do organismo, visto que desde organismos menos complexos, como o unicelular *Trichomona vaginalis,* até répteis, como salamandras, possuem frações similares de TEs (WELLS; FESCHOTTE, 2020).

Devido a sua grande capacidade de mobilização e/ou multiplicação nos genomas hospedeiros, os TEs são considerados importantes criadores de variabilidade genética, moldando não só a plasticidade dos genomas como também o tamanho dos mesmos (BIÉMONT, 2010).

Elementos transponíveis também são considerados cruciais para a evolução, pois além da capacidade de mudança de posição dentro do genomas em que se encontram, também são capazes de invadir novos genomas, e quando a aquisição de um novo TE provém de forma que não por herança parental chamamos de transferência horizontal de TEs (HTT) (SCHAACK et al., 2010; PECCOUD et al., 2017). Dessa forma, TEs tornam-se fontes inesgotáveis de inovação para os genomas, os quais ao longo desta corrida armamentista desenvolveram formas de tirar proveito das inúmeras possibilidades que as mobilizações podem gerar.

2.2.1 CLASSIFICAÇÃO DE TEs

A principal classificação dos TEs, proposta por Fineggan (1989) é amplamente aceita e usada até hoje. Essa classificação divide os TEs em duas principais classes

de acordo com os mecanismos de transposição dos elementos: Classe I ou retrotransposons e Classe II ou transposons.

A transposição de retroelementos da Classe I começa com a síntese de uma molécula de RNA mensageiro (mRNA) através de enzimas do hospedeiro, que por sua vez é utilizada como molde para a produção de uma molécula de DNA, através da enzima Transcriptase Reversa que é codificada pelo retroelemento. Essa nova molécula de DNA é integrada em alvos do DNA hospedeiro. Esse mecanismo de transposição é conhecido como 'copia-e-cola' (do inglês, *copy-and-paste*) ou transposição replicativa, e resulta no aumento do número de cópias do TE a cada evento de transposição (WICKER et al., 2007; BOURQUE et al.,

2018).

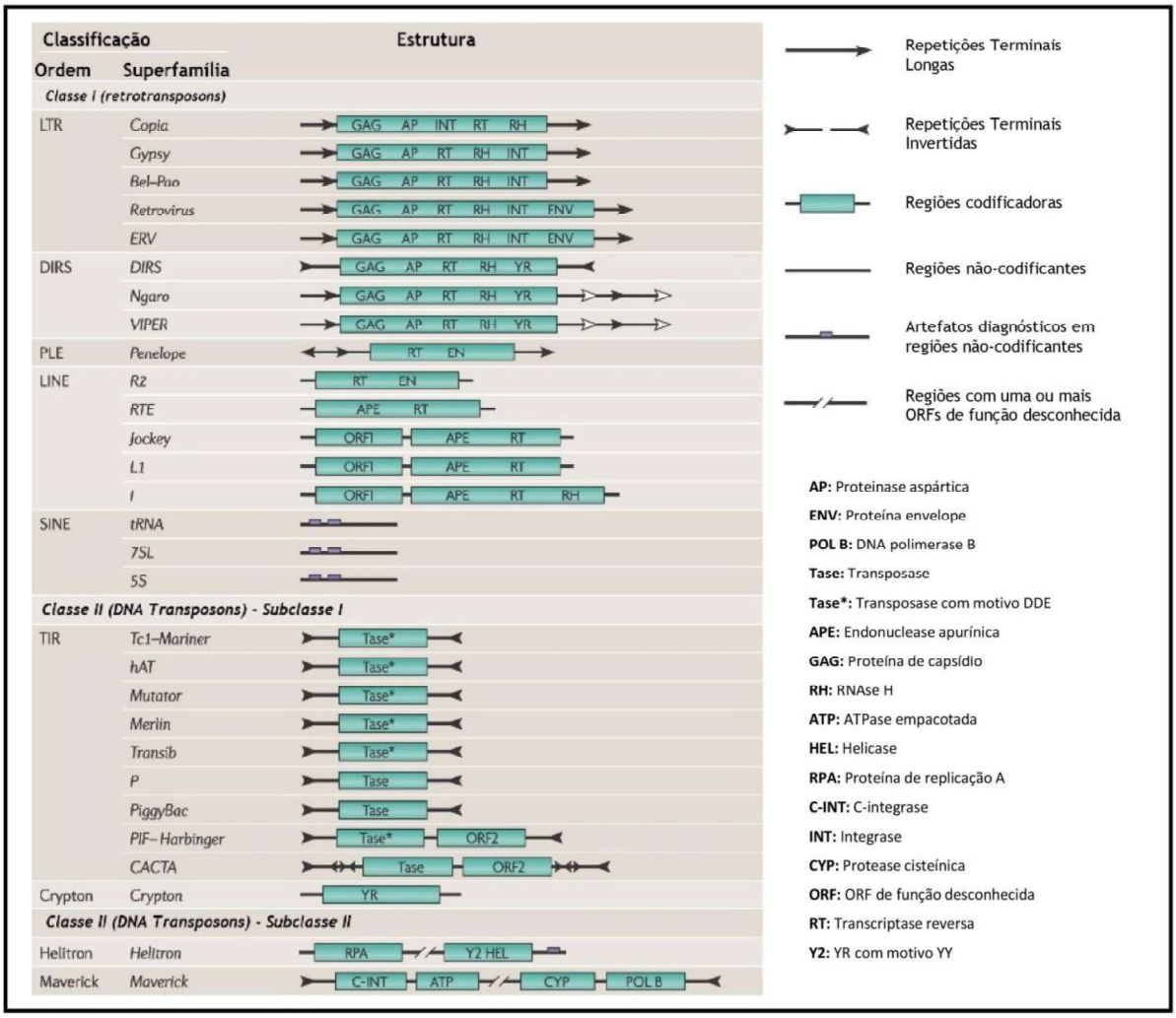


Figura 5 – Classificação dos TEs proposta por Wicker et al., 2007 levando em conta principalmente a estrutura dos elementos de transposição. Adaptado de Wicker et al., 2007.

Os transposons de Classe II se transpõem por enzimas transposases que são codificadas na sequência do próprio elemento, que apresenta autonomia para sua mobilização. As transposases apresentam domínios conservados que excisam o elemento de um sítio doador, pelo reconhecimento de sequências específicas das extremidades do elemento, e o inserem em um sítio receptor característico. Esse tipo de transposição é conhecido como transposição conservativa, ou 'corta-e-cola' (do inglês, *cut-and-paste*) ou ainda, no caso específico da subclasse dos Helitron,

‘descasca-e-cola’ (do inglês *peel-and-paste*), mecanismo esse que utiliza um intermediário de DNA circular (cDNA) para a transposição (WICKER et al., 2007; GRABUNDZIJA et al., 2016; BOURQUE et al., 2018).

Elementos das classes I e II apresentam elementos autônomos e não autônomos. Elementos autônomos possuem em sua estrutura sequências codificadoras das enzimas envolvidas com a sua transposição, enquanto os elementos não autônomos não apresentam essas sequências. Dentro das Classes os TEs são classificados em subclasses de acordo com sua estrutura e especificidades de inserção (WICKER et al., 2007; BOURQUE et al., 2018) (Figura 5). As subclasses ainda podem ser subdivididas em ordens, superfamílias ou famílias de acordo com as semelhanças estruturais e através de relações filogenéticas (WICKER et al.,2007; BOURQUE et al., 2018).

2.2.2 CICLO EVOLUTIVO DOS TEs E EVOLUÇÃO DOS HOSPEDEIROS

A relação dos TEs com seus hospedeiros pode ser comparada com várias relações comuns a diferentes seres vivos. Após a mobilização e inserção em um novo local (início do ‘ciclo de vida’ de um TE), interações comparáveis a competição, ao mutualismo e principalmente ao parasitismo ocorrem em conjunto com o genoma hospedeiro (ROBILLARD et al., 2016). Dependendo do local onde ocorre a inserção, a nova posição do TE pode ser deletéria para o hospedeiro. Isso ocorre, por exemplo, no caso de inserções em meio a sequências codificadoras ou regulatórias (KAZAZIAN, 2004). Outras modificações causadas pela mobilização de TEs incluem recombinações, deleções, duplicações, inversões e translocações (VOLFF, 2006; BURNS; BOEKE, 2012). Com isso, TEs vêm sendo considerados como importantes fontes de mutações e causadores de várias doenças genéticas em diversos organismos, com vários estudos desses casos em humanos (VORECHOVSKY, 2010; HANCKS; KAZAZIAN, 2012; CHÉNAIS, 2015; BURNS, 2017; ANWAR et al., 2017). Contrapondo isso, os genomas dos hospedeiros desenvolveram suas próprias defesas para essa corrida armamentista contra os TEs. Mecanismo de reparo e

produção de RNAs de interferência, como siRNAs e piRNAs (AGREN; WRIGHT, 2011; YAMASHIRO; SIOMI, 2017) são as primeiras defesas do hospedeiro. Uma vez que o TE seja silenciado, o mesmo perde sua capacidade de transposição/proliferação, tendo seu ciclo de vida interrompido, acumulando mutações e padecendo através de forças evolutivas como a seleção natural (LINCH; WALSH, 2007).

Porém, nem todas as inserções de TEs são deletérias, podendo assim estabelecer-se outras relações 'ecológicas' entre o TE e seu hospedeiro. Este cenário pode levar a domesticação, cooptação e exaptação, onde o hospedeiro utiliza-se da sequência do TE para produzir transcritos que interessam ao hospedeiro (VOLFF, 2006; REBOLLO et al., 2012; JANGAM et al., 2017; SCHRADER; SCHMITZ, 2019).

O exemplo mais estudado de domesticação de TEs é o da proteína Rag1 em vertebrados (com exceção de Agnatha). Essa proteína, juntamente com a Rag2, é responsável pela geração de diversidade de receptores de imunoglobulinas e células-T, aumentando assim a capacidade do sistema imunológico de reconhecer uma gama maior de antígenos invasores (VOLFF, 2006). A origem da Rag1 é atribuída a domesticação de uma transposase pertencente a um transposon Transib há aproximadamente 500 milhões de anos (ZHOU et al., 2004; KAPITONOV; JURKA, 2005).

2.3 MOBILOMAS E PAISAGENS DE TEs

Todo o conjunto de Elementos Transponíveis de um genoma juntamente com os seus vírus endógenos é chamado de mobiloma (SIEFERT, 2009). Estudos de mobilomas têm provido diferentes visões sobre a composição dos genomas e as forças evolutivas que influenciam na transposição de TEs e consequente modificações nos genomas, bem como comparações entre mobilomas podem ser uma alternativa às análises mais frequentes, servindo como ferramenta importante em diversas áreas (SIEFFERT, 2009; ELLIOTT; GREGORY, 2015; QUADRANA et al., 2016).

Paisagens de elementos transponíveis (do inglês Transposon Landscape) são representações gráficas da proporção de cada superfamília de TEs em relação à distância genética entre as cópias dos TEs no mobiloma, permitindo vislumbrar a história e a idade dessas superfamílias. Esses gráficos permitem termos ideia da diversidade de TEs em um mobiloma e a dinâmica dos elementos ao longo do tempo.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Análise de genes potencialmente relacionados a ecologia restrita a flores e efeito da ecologia restrita no mobiloma de *Drosophila incompta*.

3.2 OBJETIVOS ESPECÍFICOS

1) Caracterizar e analisar o repertório de receptores olfativos (ORs) e receptores gustatórios (GRs) em *Drosophila incompta*;

2) Comparar o repertório de GRs e ORs entre *D. incompta* e outras espécies também especialistas e espécies generalistas do gênero *Drosophila*;

3) Investigar sinais de seleção positiva nos genes dos quimiorreceptores de *D. incompta* e em sítios dos mesmos e verificar a relação dessas forças evolutivas com a especificidade de nicho dessa espécie;

4) Caracterizar e analisar o Mobiloma de *D. incompta*, comparando-o com os de outras espécies do gênero *Drosophila*;

5) Vislumbrar a paisagem de elementos transponíveis (TE Landscape) de *D. incompta* comparando-a com a de outras espécies especialistas e generalistas de *Drosophila*.

CAPÍTULO II

ARTIGO 1 - COMPARISONS INVOLVING THE REPERTOIRE OF CHEMORECEPTORS BETWEEN A HIGHLY SPECIALIZED FLY AND OTHER DROSOPHILA SPECIES

Fonseca, P. M., Robe, L.J., Loreto, E. L. S.

Manuscrito em preparação.

ABSTRACT

To explore the vast diversity of scenarios in nature, animals have developed some tools to interact with different conditions imposed by environments. Chemoreceptors encompass an important component of this interface, and are subdivided in odorant receptors (ORs), used to find food, detect partners and avoid predators, and gustatory receptors (GRs), used to distinguish healthy resources from dangerous ones. *Drosophila* has been widely employed as a model organism in many fields of science, partly because of diversity of species and niches. In this way, the contrast between generalist and specialist *Drosophila* species constitute an important model for the study of chemoreception evolution. Here we compare the repertoire of chemoreceptors of *Drosophila incompta*, a very specialized species which is ecologically restricted to flowers of *Cestrum* (Solanaceae), with that presented by other 30 *Drosophila* species, after associating the preferences of *D. incompta* to the odor of *Cestrum* flowers under olfactory choice tests. We found a clear preference of *D. incompta* for the extract of *Cestrum* flowers and presented evidence that the repertoire of GR and OR of this species is consistently smaller than that presented by species of the *Sophophora* subgenus. Nevertheless, a similar pattern was also detected for other non-*Sophophora* species, suggesting the presence of subjacent phylogenetic trends. Even so, we found some autapomorphic gene losses and many genes that seem to be under positive selection in *D. incompta*, indicating that the very specific lifestyle of these flies may be shaping the evolution of at least some of the members of these two gene families.

KEYWORDS

OR; GR; Olfaction; Taste; Odorant Receptor; Gustatory Receptor; *Yellow* gene family.

INTRODUCTION

Nature is vast of environments with different conditions and particularities. Animals living in each environment are in constant contact with many chemical compounds. Some of these compounds come from other living forms, and are received and interpreted with many different purposes, as meeting sexual partners (Wyatt 2003; Kurtovic et al. 2007; Davis 2007), identifying predators or parasites (Joseph & Carlson 2015; Ebrahim et al. 2015) or even finding food and feeling their taste (Dahanukar et al. 2005; Hallem et al. 2006). Despite the kind of chemical signals, animals usually interpret them through chemoreceptors present in their cells. As chemoreceptors are directly involved in this first contact with nature, they are subject to different selection pressures during animal evolution (Whiteman & Pierce 2008). In this sense, chemoreceptor genes generally evolve rapidly, producing a great diversification and birth and death of gene family members (McBride, 2007, McBride & Arguello, 2007; Vieira & Rozas, 2011, Cande et al., 2013). These patterns generally lead to a strict relationship between odorant recognition and individual ecological needs (Carey et al., 2010).

Since the early of 20th century, many species of *Drosophila* have been used as model organisms in different fields of science, among which we can highlight *Drosophila melanogaster* (Markow 2015). Nevertheless, the genus comprises 1,644 described species with many distinct characteristics, niches and distribution (O'Grady & DeSalle 2018; Bächli 2020). Many of these species have provided excellent models for evolutionary genomics, once there is wealthy phylogenetic information covering hundreds of species and at least thirty well-annotated genomes (Drosophila 12 Genomes Consortium et al. 2007; Chen et al. 2014; O'Grady & DeSalle 2018). *Drosophila* has also several advantages to the study of adaptations related to different ecological habits, since it encompasses species with contrasting behaviors and niche breaths. Among these, we can detach some generalist species, as *D. melanogaster*, *D. ananassae* and *D. simulans*, and other specialists, like *D. grimshawi*, *D. erecta* and *D. sechellia*. Generalist habits provide the advantage of

exploiting different resources for feeding and laying eggs, allowing for resilience under conditions of limited availability of resources or competition (Anholt 2020). Conversely, specialists take advantage of low levels of competition for food or oviposition sites once they can explore toxic or extreme niches (Anholt 2020).

In *Drosophila*, there are two well-studied families of chemoreceptors: gustatory receptors (GR) and olfactory receptors (OR), which are responsible for the detection of tastings or odorants, respectively (Clyne et al. 1999, Gao & Chess, 1999, Scott et al. 2001, Vosshall & Stocker 2007, Martin et al. 2013). In *D. melanogaster*, each of these two gene families presents 60 genes (Robertson et al. 2003). These genes are commonly involved in adaptations to different ecological contexts (Depetris-Chauvin et al. 2015; Diaz et al. 2018). Such a pattern has been demonstrated, for example, for the pestiferous species *D. suzukii,* which differs from other species of the *D. melanogaster* group due to its characteristic of exploring fresh fruit rather than rotting ones, which is reflected in the OR and GR repertoire (Hickner et al. 2016). Moreover, specialist *Drosophila* species can present a fivefold higher dn/ds ratio for OR and GR genes (McBride 2007), and may also present faster rates of gene loss for these two gene families (McBride & Arguello, 2007). Nevertheless, according to Gardiner et al. (2008), endemism rather than niche specialization may account for much of this straightforward chemosensory gene loss.

Within *Drosophila,* the *D. flavopilosa* species group draws attention for its close relationship with flowers of *Cestrum* (Solanaceae), since it uses this substratum as unique resource for oviposition, larval development, and adult feeding (Brncic 1966; Robe et al. 2013). This group consists of 18 species (Bächli 2020) divided into two subgroups and is considered to belong to the *virilis-repleta* radiation (Throckmorton 1975), currently considered the *Siphlodora* subgenus (Yassin, 2013). Nevertheless, the exact phylogenetic position of the group within the subgenus remains controversial, as this group can be considered either a sister-group of *D. annulimana* species group (Robe et al. 2010) or a sister-group of the *D. virilis* or *D. repleta* species groups (de Ré et al. 2017). Flies of the *D. flavopilosa* group share some strict morphological and physiological adaptations to the use of *Cestrum*

flowers, such as the light *Yellow* color of the body, that is cryptic with many *Cestrum* flowers; the medium to small size of the flies; the egg laying in advanced stages of embryonic development, pairing with the ephemeral nature of *Cestrum* flowers; and the robust and spine full characteristic of the ovipositors, that are used to scarify the flowers surface (Wheeler et al. 1962; Pipkin et al. 1966; Brncic 1983; Ludwig et al. 2002; Robe et al. 2013). Although the occurrence of this group covers a great part of the Neotropical region, most species are strictly endemic to small patches (Robe et al., 2013). The most widespread species of the group is *Drosophila incompta*, which occurs from Argentina to Mexico (Bächli 2020).

Due to the vast number of chemoreceptors in *Drosophila*, many different resources can be explored, establishing different niches. In order to associate the preference of a species with its specific resource, choice tests are usually performed comparing different substrates. In these tests, *Drosophila* usually prefers citric substrates for laying eggs (Dweck et al. 2013), but this choice can vary according to the niche of each species. Mansourian et al. (2018) demonstrated that specimens of *D. melanogaster* are seasonally specialist of marula fruits (*Sclerocarya birrea*), since they prefer marula volatile odors over orange ones, despite orange being preferred when tested against other fruit odors like banana. The host shift among populations can also be evaluated through the comparison of preferences to different odors. In fact, Date et al. (2013) found population specific preferences for *Drosophila mojavensis* (Date et al. 2013), a cactophilic species that can feed and breed on different species of cactus.

Thereby, this study aims to contribute to the knowledge involving the patterns of molecular evolution associated with the repertoire of chemoreceptors available in different ecological contexts. To this task, we first investigated the olfactory preferences of *D. incompta*, to test previous claims of antophilic specialization. After, we characterized OR and GR genes that are part of the repertoire of chemoreceptors of this species, and compared them with orthologous sequences in other *Drosophila* species, in order to get a glimpse about the forces that may shape the evolution of chemoreceptor genes in specialist species.

MATERIAL AND METHODS

Sampling

Specimens of the *Drosophila flavopilosa* group were obtained from flowers of *Cestrum strigilatum* (Solanaceae) collected in the city of Santa Maria, located at the south of Brazil (latitude -34.95303 and longitude -120.43572). The sampled flowers were maintained in the laboratory until the emergence of adult flies. Flies were identified by their external morphology and male genitalia patterns following Wheeler et al. (1962).

Olfactory choice tests

In order to test the flies' preference for different odors, an olfactory choice test was performed with adult flies of *D. incompta* and *D. melanogaster*. Flies of *D. incompta* were obtained from the field, as described above, whereas specimens of *D. melanogaster* were from the Oregon-R strain maintained in our lab stock. Flies were tested two days after emergence. The elected odors were: 1) extract of *Cestrum strigilatum* L. flowers (1ml of a solution of five entire non-rotting flowers – approximately 5g – macerated in 5ml of distilled water); 2) extract of *Brunfelsia uniflora* (pohl) d.don (Solanaceae) (1ml of a solution of two entire non-rotting flowers – approximately 5g – macerated in 5ml of distilled water); 3) extract of banana fruit (1ml of a solution of 5g of a mature banana macerated in 5ml of distilled water); and 4) extract of orange fruit (1ml of a solution of a piece of 5g of a mature orange with peel macerated in 5ml of distilled water). The odors were chosen according to the strict adaptation of *D. incompta* to *Cestrum spp.* (Hofmann & Napp 1984; Santos & Vilela 2005), and the general profile of exploring different fermenting fruits presented by frugivorous *Drosophila* species (Sevenster & Van Alphen 1993; Markow 2015). Conversely, *Brunfelsia* flowers were elected because they blossom concomitantly with *Cestrum* and have records of interaction with other Drosophilids (Frota-Pessoa 1952; Malogolowkin 1953; Sepel et al. 2000; Schmitz & Valente 2019).

In each case, the extracts were combined in pairs and dripped into an unused piece of cotton (Fig. 1). Ten flies of the same sex were put into the center of the olfactometer (adapted from Fuyama 1975), with gates closed for two minutes to acclimatization. After opening the gates, flies were left to choose a side. After five minutes, flies at each side of the artifact were counted. The preference tests were performed on a darkroom with only red light available. Each test was replicated at least six times, and involved both males and females, since both sexes showed similar patterns. ANOVA tests were performed to evaluate the significance of preferences.

Molecular Biology Procedures

Total DNA was isolated from 20 males of *D. incompta* using NucleoSpin Tissue XS kit (Macherey-Nagel, Düren, Germany) following the manufacture's protocol. Two approaches were used to obtain whole-genome shotgun sequences. First, a single-end approach with a read-size of approximately 100 bp was performed by the Fasteris DNA Sequencing Service, with a Solexa-Illumina HiSeq 2000 Next Generation Sequencing (NGS) device. The second NGS approach was performed by Macrogen Sequencing Service (Korea) using a Solexa-Illumina HiSeq 2000 device (Illumina Inc., San Diego, USA) under a paired-end approach, with read size of 100 bp.

Other 20 males of *D. incompta* were used to obtain total adult RNA using the TRIzol Reagent (ThermoFisher Scientific) according to the protocol of Rio et al. (2010). mRNA was isolated using mRNA Dynabeads C kit (Life Technologies) and the construction of libraries was performed with Seq RNA Total Ion V2 kit (ThermoFisher Scientific). Sequencing was performed in the Ion S5 sequencer (ThermoFisher Scientific).

Filtering and trimming were performed for both DNA and RNA sequences using Trim Galore software (Krueger 2015). Three different assembling methods were employed: the softwares ABySS 2.0 (Jackman et al. 2017), SPAdes 3.0 (Bankevich et al. 2012) and SOAPdenovo (Xie et al. 2014). Performance of the different *de novo*

assemblies was assessed by Quast software (Gurevich et al. 2013). The transcriptome *de novo* assembly was performed on Trinity (Grabherr et al. 2011), under default parameters, on Galaxy platform (Afgan et al. 2018).

Recovery of OR, GR, nAChR and Yellow gene sequences on D. incompta

Once the draft genome and the transcriptome assembly of *D. incompta* were available, three methods were used to retrieve sequences related to OR and GR chemoreceptors:

A) First, the repertory of gustatory and olfactory receptors genes of *D. melanogaster* (Robertson et al. 2003) and *D. virilis* (*Drosophila* 12 Genomes Consortium 2008) were downloaded from FlyBase (Thurmond et al. 2019), totaling 121 and 103 different genes sequences, respectively. These datasets were then employed as query under BLASTx searches (NCBI website) in order to recover the respective amino acid sequences. *Drosophila melanogaster* was chosen because it is a model organism that has a well-annotated and well-studied repertoire of chemoreceptors. In its turn, *D. virilis* entered as a query in our searches because it is closely related to *D. incompta* (Robe et al. 2010; De Ré et al. 2017). These sequences were then arranged in four separate matrices, subdivided by species and chemoreceptor gene family. Each of these sequence groups were then used as seed against the reads of *D. incompta* genome and transcriptome, through the use of aTRAM 2.0 software (Allen et al. 2018) with default parameters. This software simultaneously searches and assembles orthologous gene sequences.

B) The second procedure was the functional annotation of the entire genome with eggNOG-mapper (Huerta-Cepas et al. 2017) using default parameters, to get a general view of the entire genome and a glimpse into gene functions.

C) Finally, the repertoire of OR and GR gene sequences of *D. melanogaster* and *D. virilis* were used as query in BLASTx and BLASTn searches (NCBI website) against the assembled genome and transcriptome of *D. incompta*. In this sense, whereas the first strategy is directed against the amino acid sequences, the second aims to recover sequences potentially neglected in the BLASTx search because of putative stop-codons or due to introns.

BLASTn was also used to search for two other gene groups, *nAChR* and *Yellow*, with seven and 14 genes (the entire family), respectively, in order to compare with chemoreceptor gene families. The former gene family was chosen because it is straightforwardly distributed on insects, although its number of genes hardly varies (Dupuis et al. 2012). Nevertheless, three genes of this family (*nAChRα5, nAChRα6* and *nAChRβ3*) were withdrawn from analyses because we could not solve the relationships among them and other related genes and/or due to poor quality and/or small length of sequences on databases. On the other hand, as the Yellow gene family seems to be involved on flies' body pigmentation (Fergusson et al. 2011; Massey & Wittkopp 2016), and *D. incompta* specimens are totally light-*Yellow* colored, we elected these genes as potential targets to selection. *Drosophila melanogaster* repertoires stored on FlyBase database were used as seed on this approach.

Repertoire of OR, GR, nAChR and Yellow gene sequences on other Drosophila species

In order to recover *OR*, *GR*, *nAChR* and *Yellow* gene sequences from other *Drosophila* species, we relied on the repertoires found by Gardiner et al. (2008) for the first 12 *Drosophila* genomes (*D. ananassae, D. erecta, D. grimshawi, D. mojavensis, D. melanogaster, D. persimilis, D. pseudoobscura, D. sechellia, D. simulans, D. virilis, D. willistoni* and *D. yakuba*). To complement the matrices and provide a broader glimpse about species with different levels of niche specialization, we also used genes of other 19 *Drosophila* species (*D. arizonae, D. biarmipes, D. bipectinata, D. busckii, D. buzzatii, D. elegans, D. eugracilis, D. ficusphila, D. hydei, D. kikkawai, D. mauritiana, D. miranda, D. navojoa, D. novamexicana, D. obscura, D. rhopaloa, D. serrate, D. suzukii* and *D. takahashii)*, in addition to species used as outgroups in our evolutionary analyzes. These complementary datasets were recovered through BLASTn searches on NCBI database using genes of the four gene families previously recovered for both *D. virilis* and *D. melanogaster* as queries. In each case, only the main copy of each gene was considered, and duplicated copies

were not counted. Conversely, pseudogenes were eventually included in further analyses if they were the only available copies of the respective genes. We considered this approach a strategy to get a wider glimpse about the general evolutionary pattern of the four different gene families, since it is not strongly affected by recent duplications or gene losses in other *Drosophila* species. Moreover, this approach is more robust to problems related to sequencing or assembling strategies.

Evolutionary Tests

Matrices were constructed for each individual gene including all orthologous sequences found across the different *Drosophila* species and outgroups (Supplementary Table S1). *Scaptodrosophila lebanonensis* was generally employed as outgroup, except in cases when orthologous sequences were not found for this species. In these cases, another species closely related to Drosophilidae presenting orthologous sequences available on NCBI was employed as outgroup: *Calliphora stygia, Ctenopseustis obliquana* or *Rhopalosiphum maidis* for ORs and *Bactrocera latifrons, Musca domestica* or *Ceratits capitata* for GRs. After removing non-coding sequences, codon-based alignments were performed on MEGA X (Kumar et al 2018) for each gene of GR, OR, *nAChR* and Yellow gene families for which orthologous sequences of at least 135 bp were found for *D. incompta*.

Two different tests were then employed to investigate the presence of negative or positive selection in OR, GR, *nAChR* and Yellow genes of *D. incompta*: A) The Adaptive Branch-Site REL test for episodic diversification - aBSREL (Smith et al. 2015) was performed to test whether the branch leading to *D. incompta* has evolved under positive selection; and B) The Fixed Effects Likelihood test - FEL (Pond & Frost 2005) was used to infer the non-synonymous substitution (α) and synonymous substitution (β) rates while testing whether alpha is significantly higher than beta at each site across the sequence of *D. incompta*. The later test was only applied to matrices presenting a signal of positive selection for *D. incompta* genes on the aBSREL test. The two tests were performed in Datamonkey web platform (Weaver et al. 2018).

RESULTS

Olfactory Preference Tests

The smells perception tests showed a clear preference of *D. incompta* for *Cestrum* flowers compared to fermented bananas extract (p = 0). *Drosophila incompta* also prefers *Cestrum* flowers in comparison to extract of fermented orange and flowers of *Brunfelsia uniflora* (p = 0). For *D. melanogaster,* the test of preference for *Cestrum* versus fermented banana extract showed no statistically significant difference. On the other hand, our results corroborated the results found by Mansourian et al. (2018) about the significant preference for orange over banana extracts (Fig. 2, Supplementary Table S2).

Genome wide analyses

The best *de novo* assembly for the genome of *D. incompta* was obtained using SPAdes, with default parameters. This yielded 444,060 contigs, obtained from 62,895,766 paired-end raw reads (47,631,683 reads after quality filtering, trimming and merging) and 29,829,750 single-end raw reads that were not trimmed and filtered due to high sequencing quality. The assembly statistics can be assessed in Supplementary Table S3.

The total RNA sequences resulted in 12,346,389 raw single-end reads that were pruned to 8,816,476 reads after trimming and filtering. The assembly resulted in 4,818 contigs, with an L50 of 1,642 (Supplementary Table S3).

Repertoire of genes related to OR, GR, Yellow and nAChR gene families

We recovered a total of 78 sequences related to chemoreceptors genes in *D. incompta*, of which 36 and 42 are related to GRs and ORs, respectively (Table 1,

Supplementary Table S1). Of the 36 recovered GRs, 24 and 26 were found through homology searches using genes of *D. melanogaster* and *D. virilis* as queries, respectively, 25 through Egg-NOG Mapper tool and 12 through aTRAM assembly. For ORs, 31 and 36 were recovered through homology searches using genes of *D. melanogaster* and *D. virilis* as queries, respectively, 26 through Egg-NOG Mapper approach and 14 through aTRAM recovery. Only aTRAM did not recover any exclusive result, and 17 genes (eight GR and nine OR) were recovered in all searching strategies (Supplementary Table S4). Among the other gene families, all genes analyzed of the *nAChR* family and 11 of the 14 genes of the *Yellow* family were encountered for *D. incompta* (Table 1 and Supplementary Table S2).

Regarding the other *Drosophila* species, the number of chemoreceptor genes related to GRs varied from minimum values of 33 (as for *D. busckii* and *D. novamexicana*) to maximum values of 60 (as for *D. melanogaster*, *D. sechellia* and *D. simulans*). ORs genes varied from 37, for *D. busckii*, to 60, for *D. melanogaster*. Conversely, the lower and higher number of sequences related to the *Yellow* gene family were found for *D. incompta* (11) and the *melanogaster, suzukii* and *takahashii* subgroups of the *melanogaster* group (14). For the *nAChR* gene family, most species presented all seven genes, although only four copies were found for *D. rhopaloa* (Table 1 and Supplementary Table S2).

Regarding the distribution of losses or gains, a clear phylogenetic signal could be found for both, ORs and GRs (Figs. 3 and 4). In this context, it was interesting to see that the number of sequences retrieved for both, GR and OR gene families, was generally lower for non-*Sophophora* species. Even so, specific gene losses that are either autapomorphic of homoplasic for *D. incompta* were found for the ORs *Or1a*, *Or71a*, *Or85f* and *Or94b* genes, and for the GRs *Gr22e*, *Gr93c* and *Gr98b* genes. Among these, the two last losses were only found in *D. incompta* (Fig. 4).

For the *Yellow* gene family, two of the three genes that were not found in *D. incompta* were also absent in all non-*Sophophora* species analyzed here (*Yellow-e2* and *Yellow-f*). Among these, *Yellow-f* was only recovered for the *melanogaster* species group of the *Sophophora* subgenus, with the exception of the *ananassae*

species group (*D. ananassae* and *D. bipectinata*). Conversely, *Yellow-f2* was only absent in *D. incompta.*

Selection Tests

Of the 78 chemoreceptors genes found for *D. incompta*, 67 attained at least 135 bp of coding sequence (45 aa), and were included in further evolutionary analysis. Of these, 33 and 34 were GRs and ORs, respectively. For the *Yellow* and *nAChR* gene families, all of the 18 genes recovered for *D. incompta* were evaluated under the two sets of selection tests.

Of the 67 chemoreceptors matrices used for selection tests, 12 presented signals of positive selection on aBSREL approach: *gr21a*, *gr32a*, *gr57a*, *gr59f*, *gr68a*, *gr77a* and *gr89a* for GRs, and *or33c*, *or35a*, *or63a*, *or85d* and *or94a* for ORs ($p < 0.05$). These 12 matrices were also tested on FEL analysis and presented signals of positive selection ranging from one site for gr32a ($p = 0.0029$) to 17 sites on *or94a* ($p = 0.0033$). Also on three sites for gr21a ($p = 0.0003$), on four sites for gr68a ($p = 0.0212$) and gr89a ($p = 0.0116$), on eight sites for gr77a ($p = 0.0003$), on nine sites for gr59f ($p = 0.0026$) and or35a ($p = 0$), on 10 sites for or33c ($p = 0.0143$), on 15 sites for gr57a ($p = 0$) and on 16 sites for or63a ($p = 0$) and or85d ($p = 0$) (Table 2). For the *Yellow* and *nAChR* gene families none gene showed signals of positive selection for *D. incompta* on *aBSREL* test.

DISCUSSION

Drosophila has been widely used as a model to study the forces shaping the evolution of chemoreceptor gene families in different ecological contexts. Nevertheless, the role played by selection and genetic drift in the evolution of these genes on specialist species is still a matter of debate (McBride 2007; McBride and Arguello, 2007; Gardiner et al. 2008). This study adds to the comprehension about the patterns of molecular evolution presented by OR and GR gene families on a strictly adapted species of the *D. flavopilosa* group, *D. incompta*, and provides several

glimpses to the comprehension of general patterns presented by the entire *Drosophila* genus.

Molecular evolution in D. incompta

Although it has been previously demonstrated that *D. incompta* can only explore *Cestrum* flowers as feeding and breeding resources (Sepel et al. 2000; Santos & Vilela 2005; Robe et al. 2013), the ethological or physiological properties associated with this restricted ecology were not yet recognized. Our findings about the olfactory preferences of *D. incompta* suggests a chemosensory basis for this condition, as demonstrated by its clear preference for extracts of *Cestrum* flowers in comparison with other scents. It is generally expected that at least part of this specialization may be related to chemosensory modifications that are ultimately associated with genetic alterations. This hypothesis is supported by some findings presented here for the OR and GR gene families in *D. incompta*, such as some autapomorphic or homoplasic gene losses suggested for some genes, contrasted by the signals of positive selection detected for others. This is the case, for example, for the putative loss of *Or19a* in the genome of *D. incompta*. In fact, since the product of this gene seems to be the main *OR* responsible by the interaction with the valencene, one of the majors components of the citric aroma (Münch & Galizia 2016), its loss explains the poor attraction of *D. incompta* for citric scents observed in the olfactory preference tests. Conversely, for *D. melanogaster*, the results of these tests suggest a clear preference for the citric fruit in comparison to banana, which agrees with the findings about oviposition preferences provided by Dweck et al. (2013) and Mansourian et al. (2018), and can be associated with the presence of an intact copy of *Or19a*. Moreover, *D. melanogaster* was the only species with a copy of *Or19b*, probably originated by a duplication of *Or19a*, but here considered as a canonical gene due to its presence on FlyBase database.

It is known that both *OR* and *GR* gene families commonly evolve through gene duplication, gene loss and pseudogenization (Gardiner et al. 2008; Benton 2015).

More specifically, it has been suggested that a general trend of loss of some chemosensory genes, with action of positive selection on others could apply during host specialization (McBride 2007; McBride & Arguello 2007). We showed here that most missing chemoreceptor genes detected for *D. incompta* under different approaches seem to have been inherited from ancestral populations (see below), although there are some putative autapomorhic or homoplasic gene losses. In fact, from the 120 chemoreceptor genes that were evaluated in this study, only 78 retrieved orthologous sequences in *D. incompta*. Of the 42 missing genes, only seven presented evidence supporting an autapomorphic of homoplasic gene loss for the target species. This pattern generally fits the birth-and-death evolution model presented for chemoreceptor genes, showing the existence of a dynamic evolution related to both OR and GR gene families. The autapomorphic or homoplasic losses do also partially agree with the hypothesis suggesting a general trend toward gene losses in specialist species (McBride & Arguello, 2007; Croset et al. 2010). Nevertheless, it is important to consider that at least some of these putative gene losses may be an outcome of methodological sources of errors related to genome assembly or gene searches. The iterative approach wherein the results of two genomic and one transcriptome sequencing strategies were assembled and evaluated under at least three methods to retrieve sequences related to GR and OR chemoreceptors certainly reduced these sources of systematic errors, but possibly did not completely eliminate them. Such a reasoning can be further supported in face of the findings for the *nAChR* gene family, for which the number of genes found for *D. incompta* generally agrees with those found for other species of the *Drosophila* genus and especially for the *Siphlodora* subgenus.

Nevertheless, for the *Yellow* gene family, *D. incompta* detached as the species with the lower number of genes, which can be either related to such methodological artifacts of may reflect a genetic modification putatively associated with the differential colors presented by the target species. In fact, although *Yellow-e2* and *Yellow-f* were absent in all non-*Sophophora* species, *Yellow-f2* was only absent in *D. incompta.* It is known that *Yellow-f* and *Yellow-f2* code enzymes that participate in the melanization

process of *D. melanogaster*, with the last gene playing an important role in the pigmentation of the adult flies (Han et al. 2002). The absence of these two genes in *D. incompta* may contribute to the light-*Yellow* color of the adult flies of this species (Brncic 1983), which does not seem to present any polymorphism regarding body pigmentation, with rare records of other color related with feeding (Ludwig et al. 2002).

The effect of differential selection pressures was also detected on the repertoire of chemoreceptor genes that were encountered for *D. incompta*, since seven and five of its *GR* and *OR* related sequences, respectively, presented signals of positive selection. Five of the seven *GRs* and three of the five *OR* genes presenting signals of positive selection for *D. incompta* were also recovered in all other evaluated species, suggesting their products may perform some important functions that are somehow conserved across *Drosophila* phylogeny. In agreement with this pattern, among the GR genes presenting signals of positive selection in *D. incompta*, we can highlight *gr21a,* which is known to play roles of odorant and gustative receptor and is considered basal for its gene family, being acknowledged as one of the most conserved chemoreceptor in insects (Robertson et al. 2003; Robertson & Kent 2009; Hickner et al. 2016). Although the effect of the specific alterations found for *D. incompta* on the function of this and several other receptors remains yet to be assessed, it is quite interesting to find evidences of positive selection is such conserved pathways.

The association of the signals of positive selection with the restricted ecology of *D. incompta* can be further evoked when the specific roles of some *OR* genes under positive selection in this species are evaluated in more detail. *or33c*, for example, was previously found to be under positive selection in some species of the *D. melanogaster* subgroup (Tunstall et al., 2007). This gene encodes a receptor for cyclohexanone ligands (Münch & Galizia 2016), whose derivatives have been found in some essential oils of plants (Zhu 2011; Alvarez 2019) and may explain phytotoxic and cytotoxic activity of some essential oils (Berger 2007). Since many authors consider *Cestrum* toxic in many ways (McLennan & Kelly 1984; Al-Heza et al. 2009;

Bussmann et al. 2011b), the presence of cyclohexanone derivatives in the essential oils of species of this genus is an interesting topic to be addressed. Conversely, *or35a* is considered an indicator of food viability (Yao et al. 2005; Mansourian & Stensmyr 2015), being known to interact with many chemical classes, such as aldehydes, alcohols and esters (Hallem & Carlson 2006; Münch & Galizia 2016). Among the potential ligands of the product of this gene is the 1-hexanol, which can cause behavioral aversion in adult flies (Stensmyr et al. 2003 Hallem & Carlson 2006; Nemeth et al. 2018). Finally, *or86d* is a receptor to esters, specially ethyl-pentanoate (Nemeth et al. 2018), which provides some pleasant fruity aroma (Xiao et al. 2014), whereas *or94a* is considered a detector of ethylphenols, that are derived from hydroxycinnamic acids. This last component, in particular, can act as a potent dietary antioxidant, which may attract adult flies and larvae (Dweck et al. 2015).

Evolution of OR and GR genes in Drosophila

Increasing the number of evaluated species we could increase the comprehension about the patterns of evolution of OR and GR gene families across the phylogeny of *Drosophila*. In this sense, we found a great variation in the number of chemoreceptor genes among species, with the higher and lower number of both OR and GR genes being generally found for species of the *Sophophora* subgenus, in particular for species of the *melanogaster* subgroup, and for *D. busckii*, respectively. As *D. melanogaster* was one of the species employed as query or seed in our searches, the higher number of chemoreceptor genes found for this species, and other closely allied taxa could be a methodological artifact. The lower number of orthologous sequences found for *D. busckii* could also be a biased result, because this species is not closely related to any of the two species used as query or seed in our searches (Yassin 2013; O'Grady & DeSalle 2018). Moreover, no specific attention was yet paid to the annotation of chemoreceptors for this species. Nevertheless, such a pattern does not explain the results found for *D. virilis*, which does not only present a well-annotated genome but was also used as query or seed in our searches. Even so, this species retrieved a number of orthologous sequences close to the lower

range for both OR and GR genes (49 and 40, respectively). Another point to consider is the fact that we are just considering one copy of each gene, which will focus the depicted scenario to more ancient events of gene duplication. In this sense, the number of genes for some species may be somewhat underestimated. This is the case, for example, for *D. grimshawi*, which was already shown to possess several copies of some loci, related to recent gene duplications (Guo & Kim 2007; Nozawa & Nei 2007; Gardiner et al. 2008).

Despite these putative sources of bias, the general picture that emerges suggests that the distribution of OR and GR genes along the *Drosophila* phylogenetic tree possibly reflects a phylogenetic signal related to gene gains or gene losses on ancestral populations, that were stably inherited along millions of years. Such inferences are further supported when the presence/absence of each of the 24 *GR* and 18 *OR* genes of *D. melanogaster* that were not found in *D. incompta* are evaluated in a phylogenetic context. For the *ORs* that were not found for *D. incompta*, seven were also not found for all non-*Sophophora* species (*or7a, or33a, or33b, or43b, or45a, or59c* and *or65a*), and three were just recovered for one species of the *Siphlodora* subgenus (*or47a* in *D. navojoa, or 65c* in *D. arizonae* and *or85a* in *D. mojavensis*). These absences may indicate ancestral gene losses or may point a high divergence in relation to *D. melanogaster* sequences, reaching the limit of BLAST sensibility. Nevertheless, *Or22b* genes was just recovered for five of the evaluated species, among which the *melanogaster* subgroup and *D. incompta*, suggesting that neutral patterns and BLAST sensibility do not explain the complete picture. This also applies to the GR gene family, for which most genes that were not found in *D. incompta* seem to be widely restricted to the *Sophophora* subgenus. In fact, 17 GR genes (*gr10b, gr22a, gr22b, gr22c, gr22d, gr22f, gr36a, gr36b, gr36c, gr59a, gr59b, gr59c, gr92a, gr93b, gr93d, gr98c* and *gr98d*) were not recovered for none non-*Sophophora* species and two of them (*gr22c* and *gr93d*) were present only on five and four *Sophophora* species respectively. Even so, losses of *Gr93c* and *Gr98b* were autapomorhic to *D. incompta*. Comparing the mean number of chemoreceptor genes between *Sophophora* (56.27 for OR and 53.59 for GR genes) and non-*Sophophora*

species (43.67 for OR and 37 for GR genes) further agrees with the scenario of putative phylogenetic trends (Figure 5).

Nevertheless, as species of the *melanogaster* group frequently present the larger number of both OR and GR genes (60 for both in *D. melanogaster*), it is tempting to speculate that at least in some cases the number of chemoreceptor genes is also defining evolutionary potentials related not only with niche breadth, but also with the distribution range of individual species within each of the major lineages of *Drosophila*. In fact, several *Sophophora* species are recognized as generalist and cosmopolitan species, like *D. ananassae* (Singh 1996), *D. kikkawai* (Ramniwas 2012), *D. mauritiana* (Tsacas & David 1974), *D. melanogaster* (David & Capy 1988; Lachaise & Silvain 2004; Markow 2015), *D. obscura* (Begon 1975; Markow & O'Gardy 2008), *D. persimilis* (Wogaman & Seiger 1983), *D. pseudoobscura* (Carson 1971), *D. serrata* (Kelemen & Moritz 1999; Schiffer & McEvey 2006), *D. simulans* (Lemeunier et al. 1986), *D. suzukii* (Asplen et al. 2015; Biondi et al. 2016), *D. willistoni* (Spassky et al., 1971, Cordeiro & Winge, 1995, Powell, 1997), a pattern that seems to be less ubiquitous for non-*Sophophora* species. Nevertheless, this hypothesis also presents some important outliers, as is the case for *D. sechellia*, that presents 59 and 60 OR and GR genes, and is endemic to the Seychelles islands in the Indian Ocean and specialized on a single resource, the fruit *of Morinda citrifolia* (Tsacas & Bächli 1981; Louis & David 1986; Gerlach 2009); and *D. erecta*, that presents 58 and 56 OR and GR genes, and is endemic to west-central Africa and is specialized on ripe fruits of *Pandanus spp.* (Lachaise & Tsacas 1974; Rio et al. 1983; Linz et al. 2013). Thus, although such circumstances may be related to ecological potentials that were inherited from ancestral populations, the patterns and processes behind the evolution of OR and GR gene families are certainly quite complex and need to be evaluated under different perspectives.

CONCLUSIONS

The evolution of specialization towards the use of specific resources is an interesting topic of debate that is not yet fully understood (Etges 2019; Markow 2019;

Auer et al. 2020). Here we demonstrated the clear preference of *D. incompta* to the scent of macerated *Cestrum* flowers in comparison with other resources, and characterized some genetic alterations on OR and GR genes that are putatively related to this behavior. In this sense, during the characterization of the repertoire of chemoreceptors of *D. incompta*, we not only detected some autapomorphic or homoplasic gene losses, but also evidenced several signals of positive selection on different components of the two major gene families that provide the fundamental tools of interaction with nature (Benton, 2015). Even so, most gene losses presented by this species seem to have been inherited from ancestors distributed on different nodes of the *Drosophila* phylogenetic tree. This pattern supports the dynamic birth-and-death model of evolution of these genes first proposed by Vieira et al. (2007). Nevertheless, the exact relationship between these alterations and the ecological potentials presented by each species still remains widely allusive, and it is not possible now to disentangle the putative causes and consequences of ecological specialization.

REFERENCES

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic acids research. 2018 Jul 2;46(W1):W537-44.

Al-Reza SM, Rahman A, Kang SC. Chemical composition and inhibitory effect of essential oil and organic extracts of *Cestrum nocturnum* L. on food-borne pathogens. International journal of food science & technology. 2009 Jun;44(6):1176-82.

Allen JM, LaFrance R, Folk RA, Johnson KP, Guralnick RP. aTRAM 2.0: An improved, flexible locus assembler for NGS data. Evolutionary Bioinformatics. 2018 May 6;14:1176934318774546.

Alvarez MV, Ortega-Ramirez LA, Silva-Espinoza BA, Gonzalez-Aguilar GA, Ayala-Zavala JF. Antimicrobial, antioxidant, and sensorial impacts of oregano and rosemary essential oils over broccoli florets. Journal of Food Processing and Preservation. 2019 Mar;43(3):e13889.

Anholt RR. Chemosensation and Evolution of *Drosophila* Host Plant Selection. I science. 2020 Jan 24;23(1):100799.

Asplen MK, Anfora G, Biondi A, Choi DS, Chu D, Daane KM, Gibert P, Gutierrez AP, Hoelmer KA, Hutchison WD, Isaacs R. Invasion biology of spotted wing *Drosophila* (*Drosophila suzukii*): a global perspective and future priorities. Journal of Pest Science. 2015 Sep 1;88(3):469-94.

Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, Arguello JR, Hansson BS, Jefferis GS, Caron SJ, Knaden M. Olfactory receptor and circuit evolution promote host specialization. Nature. 2020 Mar;579(7799):402-8.

Bächli G (2020) The database on taxonomy of Drosophilidae. https://www.taxodros.uzh.ch/

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology. 2012 May 1;19(5):455-77.

Begon M. The relationships of *Drosophila obscura* Fallen and *D. subobscura* Collin to naturally-occurring fruits. Oecologia. 1975 Sep 1;20(3):255-77.

Benton R. Multigene family evolution: perspectives from insect chemoreceptors. Trends in ecology & evolution. 2015 Oct 1;30(10):590-600.

Berger RG, editor. Flavours and fragrances: chemistry, bioprocessing and sustainability. Springer Science & Business Media; 2007 Mar 6.

Biondi A, Traugott M, Desneux N. Special issue on *Drosophila suzukii*: from global invasion to sustainable control.

Brncic D. Ecological and cytogenetic studies of *Drosophila flavopilosa*, a neotropical species living in *Cestrum* flowers. Evolution. 1966 Mar 1:16-29.

Brncic D. The *flavopilosa* group of species as an example of flower-breeding species. The genetics and biology of *Drosophila*. 1983;6:360-77.

Bussmann RW, Malca G, Glenn A, Sharon D, Nilsen B, Parris B, Dubose D, Ruiz D, Saleda J, Martinez M, Carillo L. Toxicity of medicinal plants used in traditional medicine in Northern Peru. Journal of ethnopharmacology. 2011 Sep 1;137(1):121-40.

Cande J, Prud'homme B, Gompel N. Smells like evolution: the role of chemoreceptor evolution in behavioral change. Current opinion in neurobiology. 2013 Feb 1;23(1):152-8.

Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. Odorant reception in the malaria mosquito *Anopheles gambiae*. Nature. 2010 Mar;464(7285):66-71.

Carson HL. Speciation and the founder principle. Stadler Genetics Symposium, 1971 3, 51–70

Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki DC, FitzGerald PC, Artieri CG. Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. Genome research. 2014 Jul 1;24(7):1209-23.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron. 1999 Feb 1;22(2):327-38.

Cordeiro AR, Winge H. Levels of evolutionary divergence of *Drosophila willistoni* sibling species. Genetics of natural populations: The continuing importance of Theodosius Dobzhansky. 1995;262:280.

Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 2010 Aug 19;6(8):e1001064.

Dahanukar A, Hallem EA, Carlson JR. Insect chemoreception. Current opinion in neurobiology. 2005 Aug 1;15(4):423-30.

Date P, Dweck HK, Stensmyr MC, Shann J, Hansson BS, Rollmann SM. Divergence in olfactory host plant preference in D. mojavensis in response to cactus host use. PLOS one. 2013 Jul 25;8(7):e70027.

David JR, Capy P. Genetic variation of *Drosophila melanogaster* natural populations. Trends in Genetics. 1988 Apr 1;4(4):106-11.

Davis RL. The scent of *Drosophila* sex. Neuron. 2007 Apr 5;54(1):14-6.

De Ré FC, Robe LJ, Wallau GL, Loreto EL. Inferring the phylogenetic position of the *Drosophila flavopilosa* group: Incongruence within and between mitochondrial and nuclear multilocus datasets. Journal of Zoological Systematics and Evolutionary Research. 2017 Aug;55(3):208-21

Depetris-Chauvin A, Galagovsky D, Grosjean Y. Chemicals and chemoreceptors: ecologically relevant signals driving behavior in *Drosophila*. Frontiers in Ecology and Evolution. 2015 Apr 24;3:41.

Diaz F, Allan CW, Matzkin LM. Positive selection at sites of chemosensory genes is associated with the recent divergence and local ecological adaptation in cactophilic *Drosophila*. BMC evolutionary biology. 2018 Dec 1;18(1):144.

Drosophila 12 Genomes Consortium., Project Leaders., Clark, A. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450, 203–218 2007. https://doi.org/10.1038/nature06341

Dupuis J, Louis T, Gauthier M, Raymond V. Insights from honeybee (*Apis mellifera*) and fly (*Drosophila melanogaster*) nicotinic acetylcholine receptors: from genes to behavioral functions. Neuroscience & Biobehavioral Reviews. 2012 Jul 1;36(6):1553-64.

Dweck HK, Ebrahim SA, Kromann S, Bown D, Hillbur Y, Sachse S, Hansson BS, Stensmyr MC. Olfactory preference for egg laying on citrus substrates in *Drosophila*. Current Biology. 2013 Dec 16;23(24):2472-80.

Dweck HK, Ebrahim SA, Farhan A, Hansson BS, Stensmyr MC. Olfactory proxy detection of dietary antioxidants in *Drosophila*. Current Biology. 2015 Feb 16;25(4):455-66.

Ebrahim SA, Dweck HK, Stökl J, Hofferberth JE, Trona F, Weniger K, Rybak J, Seki Y, Stensmyr MC, Sachse S, Hansson BS. *Drosophila* avoids parasitoids by sensing their semiochemicals via a dedicated olfactory circuit. PLoS biology. 2015 Dec 16;13(12):e1002318.

Etges WJ. Evolutionary genomics of host plant adaptation: insights from *Drosophila*. Current opinion in insect science. 2019 Dec 1;36:96-102.

Ferguson LC, Green J, Surridge A, Jiggins CD. Evolution of the insect *Yellow* gene family. Molecular biology and evolution. 2011 Jan 1;28(1):257-72.

Frota-Pessoa O. Flower-feeding drosophilidae. *Drosophila* Information Service. 1952;26:101-2.

Fuyama Y. Behavior genetics of olfactory responses in *Drosophila*. I. Olfactometry and strain differences in *Drosophila melanogaster*. Behavior genetics. 1976 Oct 1;6(4):407-20.

Gao Q, Chess A. Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. Genomics. 1999 Aug 15;60(1):31-9.

Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. Molecular Ecology. 2008 Apr;17(7):1648-57.

Gerlach J. Conservation of the Seychelles sheath-tailed bat Coleura seychellensis on Silhouette Island, Seychelles. Endangered Species Research. 2009 Jul 9;8(1-2):5-13.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology. 2011 Jul;29(7):644.

Guo S, Kim J. Molecular evolution of *Drosophila* odorant receptor genes. Molecular biology and evolution. 2007 May 1;24(5):1198-207.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5.

Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. Cell. 2006 Apr 7;125(1):143-60.

Hallem EA, Dahanukar A, Carlson JR. Insect odor and taste receptors. Annu. Rev. Entomol.. 2006 Jan 7;51:113-35.

Han Q, Fang J, Ding H, Johnson JK, Christensen BM, Li J. Identification of *Drosophila melanogaster Yellow*-f and *Yellow*-f2 proteins as dopachrome-conversion enzymes. Biochemical Journal. 2002 Nov 15;368(1):333-40.

Hickner PV, Rivaldi CL, Johnson CM, Siddappaji M, Raster GJ, Syed Z. The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. BMC genomics. 2016 Dec 1;17(1):648.

Hjelmen CE, Johnston JS. The mode and tempo of genome size evolution in the subgenus Sophophora. PloS one. 2017 Mar 7;12(3):e0173505.

Hofmann PR, Napp M. Genetic-environmental relationships in *Drosophila incompta*, a species of restricted ecology. Rev Bras Genet. 1984 Jan 1;7:21-39.

Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Molecular biology and evolution. 2017 Aug 1;34(8):2115-22.

Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome research. 2017 May 1;27(5):768-77.

Joseph RM, Carlson JR. *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. Trends in Genetics. 2015 Dec 1;31(12):683-95.

Kelemen L, Moritz C. Comparative phylogeography of a sibling pair of rainforest *Drosophila* species (*Drosophila serrata* and *D. birchii*). Evolution. 1999 Aug;53(4):1306-11.

Krueger F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. 2015;516:517.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. 2018 Jun 1;35(6):1547-9.

Kurtovic A, Widmer A, Dickson BJ. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. Nature. 2007 Mar;446(7135):542-6.

Lachaise D, Silvain JF. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. In *Drosophila melanogaster*, *Drosophila simulans*: so similar, so different 2004 (pp. 17-39). Springer, Dordrecht.

Lemeunier F, Tsacas L, David J, Ashburner M. The *melanogaster* species group. In Ashburner, M., Thompson, J. R. Jr. and Carson, H. L. (eds) In The Genetics and Biology of *Drosophila*, Academic Press, London and New York, 1986 3, 147–256.

Linz J, Baschwitz A, Strutz A, Dweck HK, Sachse S, Hansson BS, Stensmyr MC. Host plant-driven sensory specialization in *Drosophila erecta*. Proceedings of the Royal Society B: Biological Sciences. 2013 Jun 7;280(1760):20130626.

Louis J, David JR. Ecological specialization in the *Drosophila melanogaster* species subgroup: a case study of D. sechellia. Acta oecologica. Oecologia generalis. 1986;7(3):215-29.

Ludwig A, Vidal NM, Loreto EL, Sepel LM. *Drosophila incompta* development without flowers. *Drosophila* Information Service. 2002;85:40-1.

Malogolowkin C. Sobre a genitália dos drosofilídeos. IV. A genitália masculina no subgênero *Drosophila* (Diptera, Drosophilidae). Rev Bras Biol. 1953;13:245-64.

Mansourian S, Stensmyr MC. The chemical ecology of the fly. Current opinion in neurobiology. 2015 Oct 1;34:95-102.

Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE, Stensmyr MC. Wild African *Drosophila melanogaster* are seasonal specialists on marula fruit. Current Biology. 2018 Dec 17;28(24):3960-8.

Markow TA. The natural history of model organisms: the secret lives of *Drosophila* flies. Elife. 2015 Jun 4;4:e06793.

Markow TA. Host use and host shifts in *Drosophila*. Current opinion in insect science. 2019 Feb 1;31:139-45.

Martin F, Boto T, Gomez-Diaz C, Alcorta E. Elements of olfactory reception in adult *Drosophila melanogaster*. The Anatomical Record. 2013 Sep;296(9):1477-88.

Massey JH, Wittkopp PJ. The genetic basis of pigmentation differences within and between *Drosophila* species. InCurrent topics in developmental biology 2016 Jan 1 (Vol. 119, pp. 27-61). Academic Press.

McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proceedings of the National Academy of Sciences. 2007 Mar 20;104(12):4996-5001.

McBride CS, Arguello JR. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics. 2007 Nov 1;177(3):1395-416.

McLennan MW, Kelly WR. *Cestrum parqui* (green cestrum) poisoning in cattle. Australian Veterinary Journal. 1984 Sep;61(9):289-91.

Münch D, Galizia CG. DoOR 2.0-comprehensive mapping of *Drosophila melanogaster* odorant responses. Scientific reports. 2016 Feb 25;6:21841.

Nemeth DC, Ammagarahalli B, Layne JE, Rollmann SM. Evolution of coeloconic sensilla in the peripheral olfactory system of *Drosophila mojavensis*. Journal of insect physiology. 2018 Oct 1;110:13-22.

Nozawa M, Nei M. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. Proceedings of the National Academy of Sciences. 2007 Apr 24;104(17):7122-7.

O'Grady PM, DeSalle R. Phylogeny of the genus *Drosophila*. Genetics. 2018 May 1;209(1):1-25

Pipkin SB, Rodriguez RL, Leon J. Plant host specificity among flower-feeding neotropical *Drosophila* (Diptera: Drosophilidae). The American Naturalist. 1966 Mar 1;100(911):135-56.

Pond SLK, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular biology and evolution. 2005 May 1;22(5):1208-22.

Powell JR. Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press; 1997 Sep 4.

Ramniwas S, Kajla B. Divergent strategy for adaptation to drought stress in two sibling species of montium species subgroup: *Drosophila kikkawai* and *Drosophila leontia*. Journal of insect physiology. 2012 Dec 1;58(12):1525-33.

Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harbor Protocols. 2010 Jun 1;2010(6):pdb-rot5439.

Robe LJ, Loreto EL, Valente VL. Radiation of the ,,*Drosophila*“ subgenus (Drosophilidae, Diptera) in the Neotropics. Journal of Zoological Systematics and Evolutionary Research. 2010 Nov;48(4):310-21.

Robe LJ, De Ré FC, Ludwig A, Loreto EL. The *Drosophila flavopilosa* species group (Diptera, Drosophilidae) An array of exciting questions. Fly. 2013 Apr 21;7(2):59-69.

Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences. 2003 Nov 25;100(suppl 2):14537-42.

Robertson HM, Kent LB. Evolution of the gene lineage encoding the carbon dioxide receptor in insects. Journal of Insect Science. 2009 Jan 1;9(1).

Santos RD, Vilela CR. Breeding sites of Neotropical Drosophilidae (Diptera): IV. Living and fallen flowers of *Sessea brasiliensis* and *Cestrum spp*.(Solanaceae). Revista Brasileira de Entomologia. 2005 Dec;49(4):544-51.

Schiffer M, McEvey SF. *Drosophila bunnanda*: a new species from northern Australia with notes on other Australian members of the montium subgroup (Diptera: Drosophilidae). Zootaxa. 2006 Oct 12;1333(1):1-23.

Schmitz HJ, Valente VL. The flower flies and the unknown diversity of Drosophilidae (Diptera): a biodiversity inventory in the Brazilian fauna. Papéis Avulsos de Zoologia. 2019;59.

Scott K, Brady Jr R, Cravchik A, Morozov P, Rzhetsky A, Zuker C, Axel R. A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. Cell. 2001 Mar 9;104(5):661-73.

Sepel LM, Golombieski RM, Napp M, Loreto EL. Seasonal fluctuations of D. cestri and D. *incompta*, two species of the *flavopilosa* group. *Drosophila* Information Service. 2000;83:122-6

Sevenster JG, Van Alphen JJ. A life history trade-off in *Drosophila* species and community structure in variable environments. Journal of Animal Ecology. 1993 Oct 1:720-36.

Singh BN. Population and behaviour genetics of *Drosophila ananassae*. Genetica. 1996 May 1;97(3):321-9.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular biology and evolution. 2015 May 1;32(5):1342-53.

Spassky B, Richmond RC, Perez-Salas S, Pavlovsky O, Mourao CA, Hunter AS, Hoenigsberg H, Dobzhansky T, Ayala FJ. Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. Evolution. 1971 Mar 1:129-43.

Stensmyr MC, Giordano E, Balloi A, Angioy AM, Hansson BS. Novel natural ligands for *Drosophila* olfactory receptor neurones. Journal of Experimental Biology. 2003 Feb 15;206(4):715-24.

Throckmorton LH. The phylogeny, ecology and geography of *Drosophila*. Invertebrates of genetic interest. 1975:421-69.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, Kaufman TC. FlyBase 2.0: the next generation. Nucleic acids research. 2019 Jan 8;47(D1):D759-65.

Tsacas L, David J. *Drosophila mauritiana* n. sp. du groupe *melanogaster* de l'Ile Maurice [Dipt. Drosophilidae]. Bulletin de la Societe entomologique de France. 1974;79(1):42-6.

Tsacas L, Bächli G. *Drosophila sechellia*, n. sp., huitieme espece du sous-groupe *melanogaster* des iles Sechelles (Diptera, Drosophilidae). REV. FR. ENTOMOL.; FRA; DA. 1981; VOL. 3; NO 4; PP. 146-150

Tunstall NE, Sirey T, Newcomb RD, Warr CG. Selective pressures on *Drosophila* chemosensory receptor genes. Journal of molecular evolution. 2007 Jun 1;64(6):628-36.

Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biology and Evolution. 2011 Jan 1;3:476-90.

Vieira FG, Sánchez-Gracia A, Rozas J. Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. Genome biology. 2007 Nov 1;8(11):R235.

Vosshall LB, Stocker RF. Molecular architecture of smell and taste in *Drosophila*. Annual review of neuroscience. 2007 Jun 28;30.

Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Molecular biology and evolution. 2018 Mar 1;35(3):773-7.

Wheeler MR, Takada H, Brncic D. The *flavopilosa* species group of *Drosophila*. Univ Texas Publ Genetics. 1962;6205:395-414.

Whiteman NK, Pierce NE. Delicious poison: genetics of *Drosophila* host plant preference. Trends in ecology & evolution. 2008 Sep 1;23(9):473-8.

Wogaman DJ, Seiger MB. Light intensity as a factor in the choice of an oviposition site by *Drosophila pseudoobscura* and *Drosophila persimilis*. Canadian journal of genetics and cytology. 1983 Aug 1;25(4):370-7.

Wyatt TD. Pheromones and animal behaviour: communication by smell and taste. Cambridge university press; 2003 Feb 27.

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014 Jun 15;30(12):1660-6.

Xiao Z, Yu D, Niu Y, Chen F, Song S, Zhu J, Zhu G. Characterization of aroma compounds of Chinese famous liquors by gas chromatography–mass spectrometry and flash GC electronic-nose. Journal of Chromatography B. 2014 Jan 15;945:92-100.

Yang Y, Hou ZC, Qian YH, Kang H, Zeng QT. Increasing the data size to accurately reconstruct the phylogenetic relationships between nine subgroups of the *Drosophila melanogaster* species group (Drosophilidae, Diptera). Molecular Phylogenetics and Evolution. 2012 Jan 1;62(1):214-23.

Yao CA, Ignell R, Carlson JR. Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. Journal of Neuroscience. 2005 Sep 14;25(37):8359-67.

Yassin A. Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the postgenomic era. Systematic Entomology. 2013 Apr;38(2):349-64.

Zhu H. Identification of essential oil from leaves of Solanum photeinocarpum (Solanaceae). Southwest China Journal of Agricultural Sciences. 2011;24(1):94-100

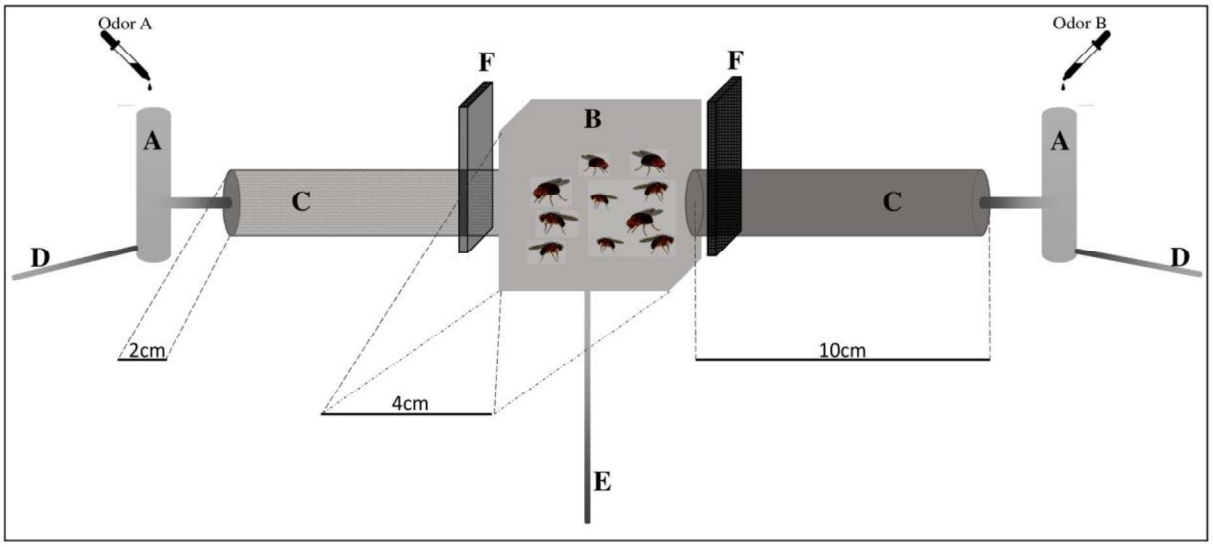


Figure 1 – Olfactometer model used to perform the olfactory preference tests. Legend: **A:** Extracts scent vessels. **B:** Flies waiting box. **C:** Preference tubes. **D:** Air inlet pipes. **E:** Air suction pipe. **F:** Gates separating the flies from scents.

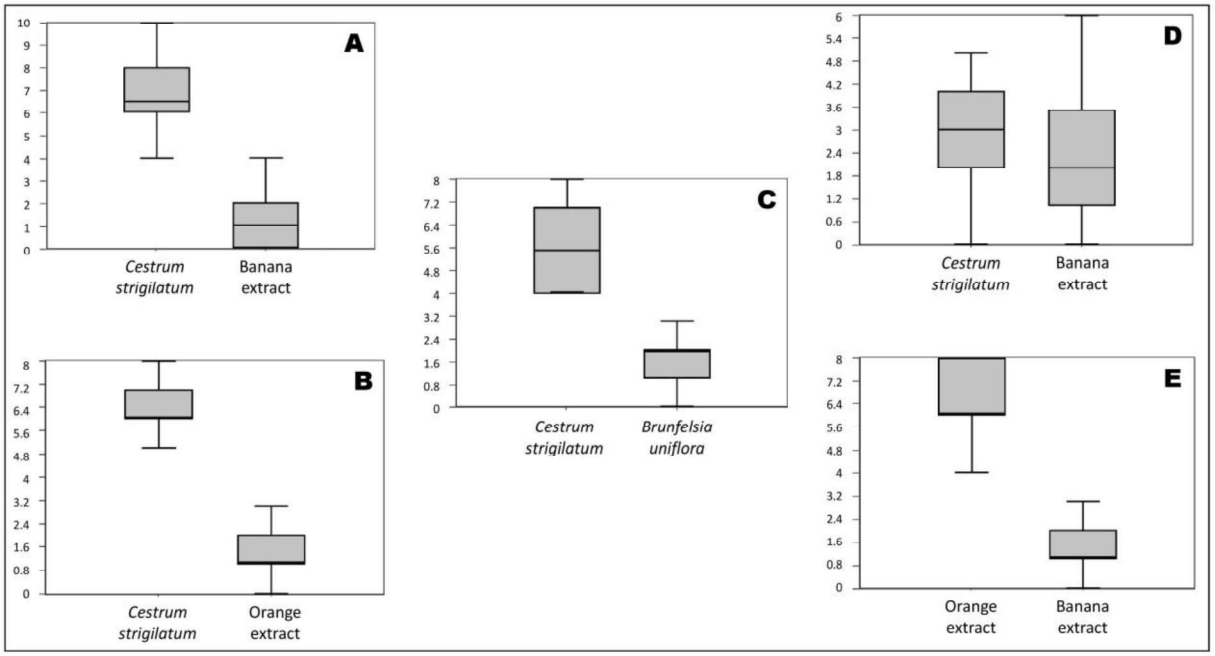


Figure 2 - Results of ANOVA performed to test the preference of *D. incompta* **(A, B and C)** and *D. melanogaster* **(D and E)** for each pair of extract scents. **A and D:** *Cestrum strigilatum* against banana. *F-Stat* = 267.0286, *p* = 0 and *F-Stat* = 0.03866, *p* = 0.84555, respectively; **B:** *C. strigilatum* against orange. *F-Stat* = 80.2002, *p* = 0; **C:** *C. strigilatum* against *Brunfelsia uniflora*. *F-Stat* = 25.71407, *p* = 0; **E:** Orange against banana. *F-Stat* = 74.88571, *p* = 0.

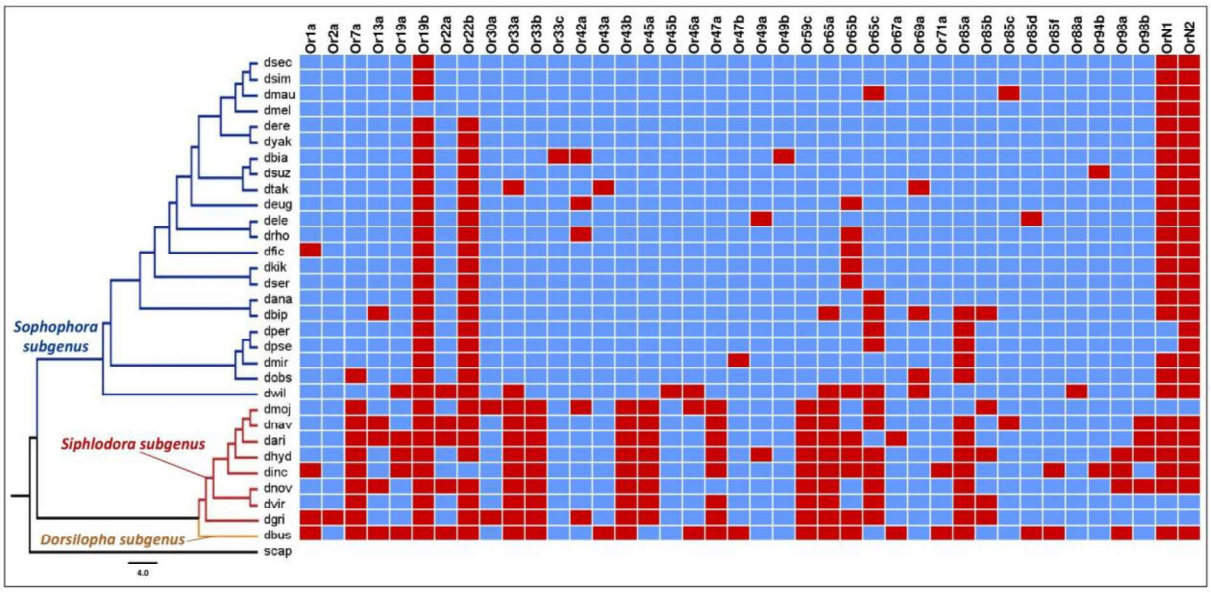


Figure 3 – Presence (blue boxes) and absence (red boxes) of 38 olfactory genes (***OR***) analyzed across *Drosophila* phylogeny, as modified from Robe et al. (2010), Yang et al. (2012), Yassin (2013), Hjelmen & Johnston (2017) and O'Grady & DeSalle (2018). Orthologues that were found in all analyzed species were not depicted on this graph.

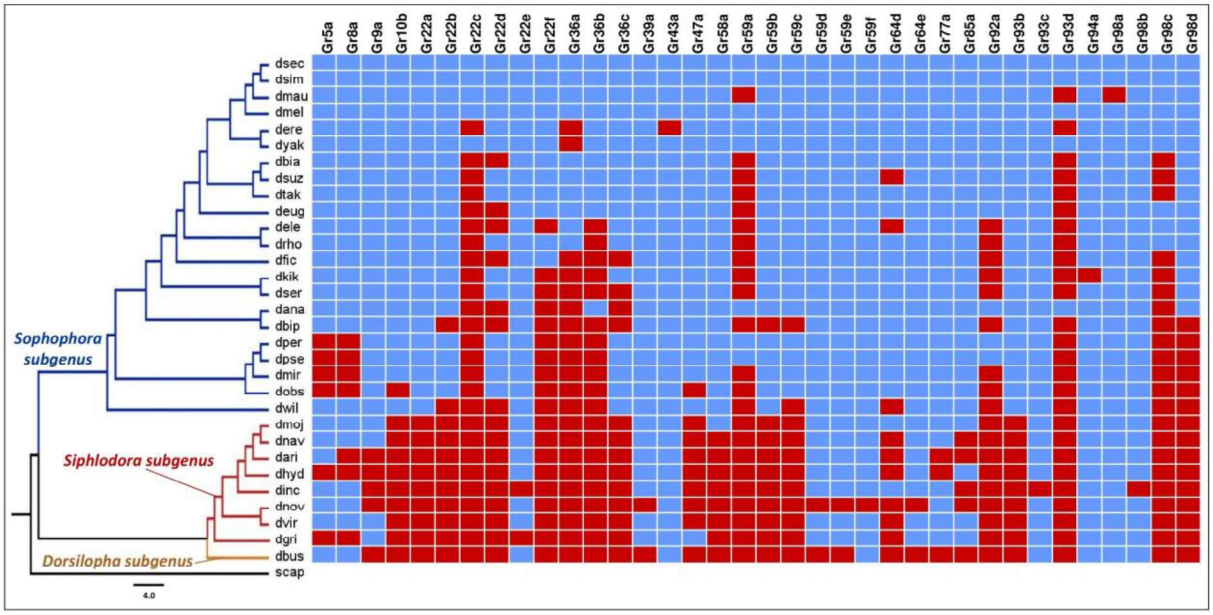


Figure 4 – Presence (blue boxes) and absence (red boxes) of 36 gustatory genes (***GR***) analyzed across *Drosophila* phylogeny, as modified from Robe et al. (2010), Yang et al. (2012), Yassin (2013), Hjelmen & Johnston (2017) and O'Grady & DeSalle (2018). Orthologs that were found in all analyzed species were not depicted on this graph.

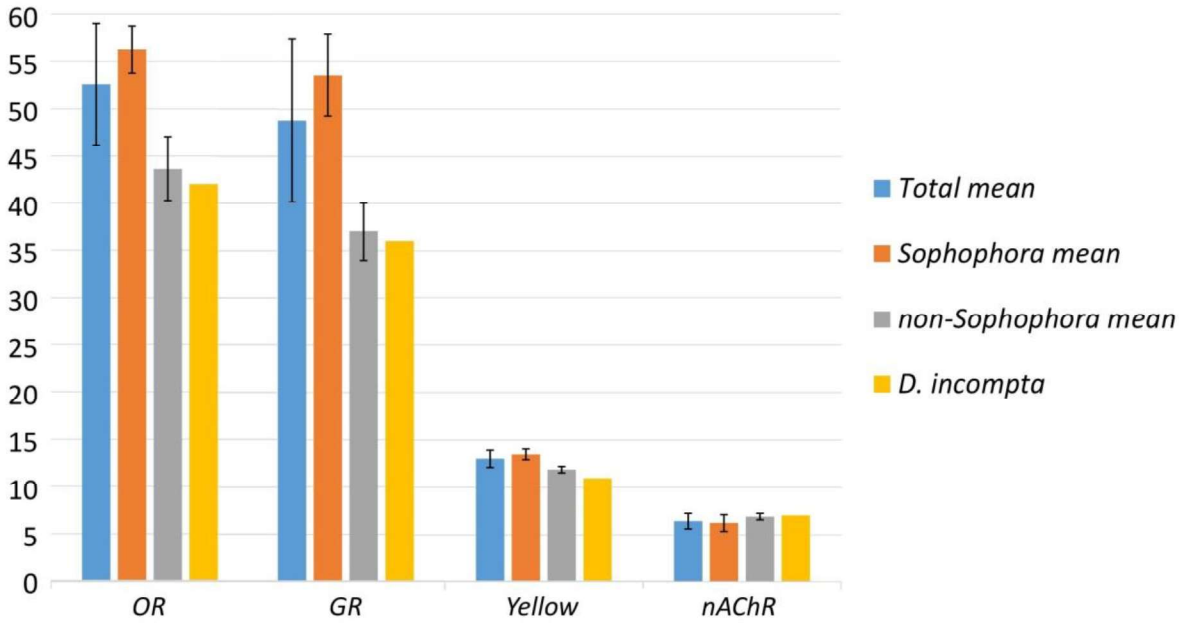


Figure 5 – Graph comparing the mean of gene numbers of all four analyzed genes with the total *D. incompta* founding genes.

Table 1 - Number of olfactory receptors (OR), gustatory receptors (GR), *Yellow* and *nAChR* related genes retrieved for *D. incompta* and each of the other 30 *Drosophila* species.

Species	Number of OR	Number of GR	Number of YELLOW	Number of nAChR
D. sechellia	59	60	14	7
D. simulans	59	60	14	7
D. mauritiana	57	57	14	7
D. melanogaster	60	60	14	7
D. erecta	58	56	14	7
D. yakuba	58	59	14	6
D. biarmipes	55	55	14	6
D. suzukii	57	55	14	5
D. takahashii	55	56	14	6
D. eugracilis	56	56	12	5
D. elegans	56	52	14	6
D. rhopaloa	56	55	14	4
D. ficusphila	56	51	14	6
D. kikkawaii	57	51	13	6
D. serrata	57	51	13	6
D. ananassea	57	54	13	7
D. bipectinata	52	46	13	6
D. persimilis	57	51	13	7
D. pseudoobscura	57	51	13	7
D. miranda	56	49	13	7
D. obscura	55	47	13	5
D. willistoni	48	47	13	7
D. mojavensis	47	42	12	7
D. navojoa	44	39	12	7
D. arizonae	43	36	12	6
D. hydei	42	36	12	7
D. incompta	**42**	**36**	**11**	**7**
D. novamexicana	45	33	12	7
D. virilis	49	40	12	7
D. grimshawi	44	38	12	7
D. busckii	37	33	12	7

Note - For each species, the total number of orthologous sequences was considered, not including any duplication but considering pseudogenes. *Drosophila melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta, D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. mojavensis*, *D. virilis* and *D. grimshawi* had their total numbers of *OR* and *GR* based on previous works (Guo & Kim 2007; Nozawa & Nei 2007; Gardiner et al. 2008).

Table 2 - Repertoire of olfactory receptors (OR) and gustatory receptors (GR) found in *D. incompta*, with their respective CDS fragment size and evidences of positive selection.

OR	CDS Fragment size recovered for *D. incompta* (bp)	Signals of positive selection for *D. incompta* according to aBSREL ($p \leq 0.05$)	Number of sites with evidence of positive selection according to FEL ($p < 0.1$)	*GR*	CDS Fragment size recovered for *D. incompta* (bp)	Signals of positive selection for *D. incompta* according to aBSREL ($p \leq 0.05$)	Number of sites with evidence of positive selection according to FEL ($p < 0.1$)
Or2a	296	-	-	*Gr2a*	1074	-	-
Or9a	1067	-	-	*Gr5a*	588	-	-
Or10a	178	-	-	*Gr8a*	143	-	-
Or13a	-	-	-	*Gr10a*	843	-	-
Or22a	-	-	-	*Gr21a*	1113	($p = 0.0003$)	3
Or22b	-	-	-	*Gr23a*	477	-	-
Or22c	304	-	-	*Gr28a*	603	-	-
Or23a	834	-	-	*Gr28b*	564	-	-
Or24a	1057	-	-	*Gr32a*	637	($p = 0.0029$)	1
Or30a	513	-	-	*Gr33a*	1442	-	-
Or33c	402	($p = 0.0143$)	10	*Gr39a*	609	-	-
Or35a	339	($p = 0$)	8	*Gr39b*	933	-	-
Or42a	811	-	-	*Gr43a*	441	-	-
Or42b	1194	-	-	*Gr47b*	570	-	-
Or43a	810	-	-	*Gr57a*	1245	($p = 0$)	15
Or45b	461	-	-	*Gr58b*		-	-
Or46a	738	-	-	*Gr58c*	963	-	-
Or47b	945	-	-	*Gr59d*	-	-	-
Or49a	-	-	-	*Gr59e*	-	-	-
Or49b	690	-	-	*Gr59f*	459	($p = 0.0026$)	9
Or56a	1101	-	-	*Gr61a*	335	-	-
Or59a	-	-	-	*Gr63a*	798	-	-
Or59b	1194	-	-	*Gr64a*	531	-	-
Or63a	1266	($p = 0$)	16	*Gr64b*	276	-	-
Or67a	326	-	-	*Gr64c*	999	-	-
Or67b	1032	-	-	*Gr64d*	581	-	-
Or67c	1011	-	-	*Gr64e*	432	-	-
Or67d	355	-	-	*Gr64f*	365	-	-
Or69a	990	-	-	*Gr66a*	753	-	-
Or74a	1215	-	-	*Gr68a*	703	($p = 0.0212$)	4
Or82a	201	-	-	*Gr77a*	812	($p = 0.0003$)	8
Or83a	1326	-	-	*Gr89a*	816	($p = 0.0116$)	4
Or83c	927	-	-	*Gr93a*	795	-	-
Or85b	496	-	-	*Gr94a*	1113	-	-
Or85c	-	-	-	*Gr97a*	787	-	-
Or85d	571	($p = 0$)	16	*Gr98a*	855	-	-
Or85e	1317	-	-				
Or88a	798	-	-				
Or92a	807	-	-				
Or94a	1158	($p = 0.0033$)	17				
Or98b	-	-	-				
Orco	-	-	-				

Note - Blank values on the CDS fragment size columns correspond to genes for which we could not reach the minimum size (45 aa) of translated protein to align in their respective matrices. Genes with no significant signal of positive selection present blank values on the selection tests columns.

Supplementary Tables

Supplementary Table S1 – List of presence (X on the table) or absence (- on the table) of *OR* (sheet 1), GR (sheet 2) and non-chemoreceptor genes (sheet 3) orthologues for each of the analyzed *Drosophila* species plus the outgroups. **Abbreviations**: Dinc (*Drosophila incompta*), Dana (*D. ananassae*), Dari (*D. arizonae*), Dbia (*D. biarmipes*), Dbus (*D. bisckii*), Dele (*D. elegans*), Dere (*D. erecta*), Deug (*D. eugracilis*), Dfic (*D. ficusphila*). Dgri (*D. grimshawi*), Dhyd (*D. hydei*), Dkik (*D. kikkawai*), Dmau (*D. mauritiana*), Dmel (*D. melanogaster*), Dmir (*D. miranda*), Dmoj (*D. nojavensis*), Dnav (*D. navojoa*), Dnov (*D. novamexicana*), Dobs (*D. obscura*), Dper (*D. persimilis*), Dpse (*D. pseudoobscura*), Drho (*D. rhopaloa*), Dsec (*D. sechellia*), Dsim (*D. simulans*), Dsuz (*D. suzukii*), Dtak (*D. takahashii*), Dvir (*D. virilis*), Dwil (*D. willistoni*), Dyak (*D. yakuba*), Blat (*Bactrocera. latifrons*), Mdom (*Musca domestica*), Ccap (*Ceratitis capitata*), Sleb (*Scaptodrosophila lebanonensis*), Csty (*Calliphora stygia*), Cobl (*Ctenopseustis obliquana*), Rmai (*Rhopalosiphum maidis*).

Supplementary Table S2 - Sample design of the five olfactory choice tests performed with *Drosophila incompta* (A, B and C) and *D. melanogaster* (D and E), comparing odors of *Cestrum* x Banana (sheets A and D), *Cestrum* x orange (sheet B) and *Cestrum* x *Brunfelsia uniflora* (sheet C) and Banana x orange (sheet E).

Supplementary Table S3 - Metrics and statistics of *Drosophila incompta* genome and transcriptome draft assemblies.

	Sequencing Platform	Runs	Mean Read Lenght	Total Raw Reads			Total Final Reads (after quality trimming and filtering)	Best *de novo* Assembler	Total Assembled Contigs	Largest Assembled Contig	N50	L50
				Single-end	Paired-end	Total						
DNA Data	Solexa-Illumina Hi-Seq 2000	1x Single-end Reads 1x Paired-end Reads	100 bp	29,829,750	62,895,766	92,725,516	77,461,433	SPAdes 3.0	444,060	36,473 bp	1,396	47,752
RNA Data	IonTorresnt 5S	2x Single-ens Reads	115 bp	12,346,389	-	12,346,389	8,816,476	Trinity	4,818	14,738	1,295	1,642

Supplementary Table S4 – Olfactory (OR **–** sheet A) and gustatory receptors (GR **–** sheet B) orthologues recovered for *Drosophila incompta* by each tool.

CAPÍTULO III

ARTIGO 2 - THE MOBILOME OF *DROSOPHILA INCOMPTA*, A FLOWER-BREEDING SPECIES: COMPARISON OF TRANSPOSABLE ELEMENT LANDSCAPES AMONG GENERALIST AND SPECIALIST FLIES

Fonseca, P. M., Moura, R. D., Wallau, G. L., Loreto, E. L. S.

Submetido: 9 de Março de 2019
Aceito: 22 de Abril de 2019
Publicado: 22 de Maio de 2019

Chromosome Res (2019) 27:203-219
Doi: https://doi.org/10.1007/s10577-019-09609-x

Chromosome Res (2019) 27:203–219
https://doi.org/10.1007/s10577-019-09609-x

ORIGINAL ARTICLE

The mobilome of *Drosophila incompta*, a flower-breeding species: comparison of transposable element landscapes among generalist and specialist flies

Pedro M. Fonseca · Rafael D. Moura · Gabriel L. Wallau · Elgion L. S. Loreto

Received: 9 March 2019 / Revised: 19 April 2019 / Accepted: 22 April 2019 / Published online: 22 May 2019

Abstract The *Drosophila* genus is one of the main model organisms in evolutionary studies, including those investigating the role of transposable elements (TE) in genomic evolution both at the nucleotide and chromosome levels. *D. incompta* is a species with restricted ecology, using *Cestrum* (Solanaceae) flowers as unique sources for oviposition, feeding and development. In the present study, we deeply characterise the *D. incompta* mobilome and generate a curated dataset. A total of 277 elements were identified, corresponding to approximately 14% of the genome, and 164 of these elements are new, of which 32.62% are putatively autonomous and 8.9% are transcriptionally active in adult flies. The restricted ecology does not seem to influence the dynamics of TE in this fly, since the proportion and diversity of TEs in its genome are similar to that of other *Drosophila* species. This result is reinforced by the absence of a clear pattern when comparing the TE landscape between generalist and specialist flies. Using 32 available *Drosophila* genomes—24 ecologically generalist species and 8 specialist species—no difference was found between their TE landscape patterns. However, differences were found between species of the *Sophophora* and *Drosophila* subgenus, indicating there are lineage-specific factors shaping TE landscapes.

Responsible Editor:Rachel O'Neill

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10577-019-09609-x) contains supplementary material, which is available to authorized users.

P. M. Fonseca · E. L. S. Loreto
Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

R. D. Moura · E. L. S. Loreto (✉)
Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Santa Maria (UFSM), Av. Roraima, 1000, Santa Maria, RS 97105-900, Brazil
e-mail: elgion@base.ufsm.br

G. L. Wallau
Departamento de Entomologia, Instituto Aggeu Magalhães (IAM), Fundação Oswaldo Cruz (FIOCRUZ/PE), Recife, PE, Brazil

Keywords Transposable Elements · Niche Amplitude · Horizontal Transfer · transposons · Genome

Abbreviations

TE	Transposable elements
bp	Base pairs
PA	Putative autonomous
PNA	Putative non-autonomous
DG	Degenerated
TIRs	Terminal inverted repeats
LTRs	Long terminal repeats
TSD	Target site duplication
ORF	Open reading frame
K2P	Kimura two parameters

Introduction

Transposable elements (TEs) are DNA fragments that are able to change their position along and between chromosomes through transposition substantial proportion of eukaryotic genomes from varied taxa. The complete set of TEs from a genome, along with endogenous viruses, is called the mobilome (Siefert 2009).

TEs are responsible for mutations and chromosomal rearrangements which are normally deleterious to the host organism. However, they are also an important source of genetic diversity and can be co-opted to new advantageous host functions. Currently, there is abundant evidence that the mobilome and its dynamics in the genome are a powerful driver of evolution (Hua-Van et al. 2011; McCullers and Steiniger 2017). The proportion of the genome occupied by TEs is highly variable, from approximately 1% in the Antarctic midge (Kelley et al. 2014) to up to 90% in the maize genome (SanMiguel et al. 1996). Although the entire TE content of a given species is usually characterised under a single term, the mobilome, these elements have different origins and normally show a huge diversity in their transposition processes, genes and the structural features they carry (Wicker et al. 2007). TEs can be autonomous, able to produce the enzymes required for their transposition, or non-autonomous elements, which require enzymes produced by autonomous ones. In addition, many degenerated elements, which are no longer able to mobilise, can also be found, which might present variable levels of degeneration that might hinder the identification of TEs (Platt et al. 2016). A number of hierarchical systems have been proposed to classify TEs based on classes, subclasses, orders and families (Wicker et al. 2007; Piégu et al. 2015).

Several bioinformatics methods have been developed for TE characterisation at the genomic scale, although there are as yet no standard benchmarks for it (Hoen et al. 2015). The tools used for TE identification and annotation are based on two methodologies: (i) homology, which identify TEs similar to a database of known TEs. These programmes are very sensitive but might miss several TEs not yet described and/or that do not have enough similarity to known TEs; (ii) via de novo, which uses structural features of TEs to identify and classify potential candidates. These methods can generate several false positive elements (Hoen et al. 2015; Platt et al. 2016), and a curated TE annotation only is possible through manual validation. There is a consensus in the scientific field that the most accurate assessment of TE landscapes uses a combination of de novo and homology-based methods followed by an additional manual curation step (Flutre et al. 2012; Platt et al. 2016).

TE landscapes are graphical representations of the proportion of each TE superfamily/family plot against the genetic distance observed between the TE copies and the consensus (ancestral element) derived from these copies, as a proxy for the TE superfamily/family age. These graphics allow an easy perception of the overall TE diversity and dynamic found in a given genome. For example, the graphic's shape can point out if TEs in a given genome experienced a recent transposition burst or if only one or a few superfamilies/families suffered this process and if all TEs have amplified constantly since some time in the past or if their copy number has decreased through time at a constant transposition rate (Barrón et al. 2014; Petersen et al. 2019). Petersen et al. (2019) analysed the TE landscapes of several arthropods with available genomes and found that even closely related species can show different TE landscapes. TEs and host genomes are subject to a continuous arms race, and understanding the evolutionary forces responsible for shaping the TE landscapes can reveal several facets of the intricate co-evolutionary process governing TE and host interactions.

The host species colonising process has been proposed to be involved in shaping the TE landscape, and invasive species are hypothesised to have a more recently active mobilome due to the reactivation of dormant TEs or due to the invasion of new TEs by horizontal transfer (Vieira et al. 1999, 2002). Stress factors associated with niche amplification, which normally follow the colonisation process, have been revealed as inducers of TE reactivation (Vieira et al. 1999, 2002; Picot et al. 2008). Thus, if niche amplitude is an important factor for shaping the TE landscape, it is expected that species with a wide range of realised niches would have different TE landscapes. More specifically, species with more limited realised niches would have a more ancient and degraded TE landscape than species with a wide range of realised niches, which would have more young and active TE copies. However, other factors, such as the effective population size, are known to influence the overall TE landscape and should be taken into consideration. For instance, in a given species with a low effective population size (Ne), natural selection might not be sufficient to remove detrimental TE insertions and instead genetic drift is considered the most

important factor shaping the TE landscape (Lynch and Conery 2003; Sessegolo et al. 2016).

For more than a century, different species from the genus *Drosophila* have been used as model organisms in many fields of science, being excellent models for evolutionary genomics (Markow 2015; O'Grady and DeSalle 2018). *Drosophila incompta* (Wheeler et al. 1962), together with 18 other species, belong to the *flavopilosa* group of *Drosophila* (Robe et al. 2013; Bächli 2018; Guillín and Rafael 2018). These species have a restricted realised niche depending on *Cestrum* spp. (Solanaceae) flowers for oviposition sites, to feed larvae, and for their development (Sepel et al. 2000; Santos and Vilela 2005; Robe et al. 2013). *D. incompta* has been sampled along with its *Cestrum* counterpart at several points along the Neotropics, since it is found from Mexico to Argentina (Bächli 2018). The flowering of *Cestrum* plants is irregular during the year, and as a consequence, the *D. incompta* populations have frequent and periodic bottlenecks. Napp and Brncic (1978) hypothesised that species from the *flavopilosa* group were likely to have low genetic variability due to their restricted ecology and frequent bottlenecks (Napp and Brncic 1978). However, studies analysing the polymorphism of chromosome inversions, isozymes and the mitogenome have shown that these flies have a substantial level of polymorphism (Robe et al. 2013; De Ré et al. 2014). The mobilome of *D. incompta* was first described by Ortiz et al. (2014), using homology-based methods. In the present study, we took complementary approaches to better characterise the *D. incompta* mobilome. A manual curation of the mobilome was implemented, resulting in a curated dataset for this mobilome. In addition, based on the particular ecological characteristic of *D. incompta*, we evaluated whether its mobilome has been shaped by its restricted realised niche and frequent bottlenecks. Finally, using 32 available *Drosophila* genomes, of which 24 species are ecologically generalist, using multiple resources and interacting with a wide range of other species, and 8 are specialists species using cactus, flowers, or specific fruits as feeding resources (see Supplementary Material 1 for a list of these species and references), we also sought to evaluate the impact of niche range on the TE landscape. Therefore, the main objective of this study was to provide an accurate description of the *D. incompta* mobilome and to test the hypothesis that the niche amplitude is a factor shaping the mobilome landscapes in *Drosophila*.

Materials and methods

Fly samples, sequencing and assembling

Flies of the *flavopilosa* group were obtained in *Cestrum* flowers (Solanaceae) collected in Santa Maria city, south of Brazil, at latitude 34.95303 and longitude − 120.43572. These flowers were maintained in the laboratory until the emergence of the adult flies. Specimens of *D. incompta* were identified by their external morphology and male genitalia, according to the pictures provided by Wheeler et al. (1962).

Total DNA was isolated from 20 *D. incompta* males using NucleoSpin Tissue XS kit (Macheray-Nagel, Düren, Germany) following the manufacture's protocol. Next-generation sequencing (NGS) was performed by Macrogen Sequencing Service (Korea) using a Solexa-Illumina HiSeq 2000 device (Illumina Inc., San Diego, USA). A paired-end approach, with a read size of 100 bp, was employed.

Total RNA was obtained from other 20 males of *D. incompta* using TRIzol (Thermo Fischer Scientific, Waltham, Massachusetts, EUA) according to the protocol describe by Rio et al. (2010). mRNA was isolated using mRNA Dynabeads C kit (Thermo Fischer), and libraries were constructed using the Seq RNA Total Ion V2 kit (Thermo Fisher Scientific). The sequencing was performed on an Ion S5 sequencer (Thermo Fisher Scientific).

Filtering and trimming processing was performed for both DNA and RNA reads using Trim Galore software (Krueger 2015). Three different assembling methods were used, and according to the Quast software (Gurevich et al. 2013), the best de novo assembly for our DNA data was obtained by using SPAdes 3.0 software (Bankevich et al. 2012) using default parameters. The transcriptome de novo assembly was implemented using Trinity (Grabherr et al. 2011) on Galaxy platform (Afgan et al. 2016) with default parameters.

Mobilome characterisation

Three complementary approaches were used to characterise the *D. incompta* mobilome. Two of them used the raw Illumina reads directly (after trimming and filtering) (Fig. 1). The de novo method uses structural features to discriminate between TEs and non-TE repeated sequences. For de novo searches, we used the DNAPipeTE pipeline (Goubert et al. 2015) with default parameters except for *-RM_t 0.5*. The *clustering method* was implemented using

the RepeatExplorer tool (Novak et al. 2013) on the Galaxy platform (Afgan et al. 2016). This approach uses the clustering of highly repetitive similar reads. After the clustering, the contigs obtained from RepeatExplorer were reassembled using the CAP3 program (Huang and Madan 1998) with the following parameters -a 20 -b 20 -c 12 -d 200 -e 30 -f 20 -g 6 -m 2 -n -5 -p 80 -r 1 -s 900 -t 300 -u 3 -v 2 -o 40. RepeatExplorer annotated the reads of each assembled cluster using RepeatMasker (http://www.repeatmasker.org) (Smit et al. 2013-2016). A third approach, based on a homology search, was implemented using the CENSOR tool on the RepBase website (https://www.girinst.org/censor/index.php) (Jurka et al. 2005). The assembled draft genome of *D. incompta* was used for this homology search.

A final TE database was obtained by clustering the three TE libraries obtained from the approaches described above using the CD-HIT-EST tool of the CD-HIT package (Weizhong and Godzik 2006; Fu et al. 2012) with default parameters. This final TE database was then submitted to a final homology search using rmBLASTn software (Altschul et al. 1990) with default parameters using the RepBase database (Version 24.10).

Manual curation of *D. incompta* mobilome

The TE dataset obtained using the approaches described previously was manually checked. All contigs that were outputted from CD-HIT-EST were individually annotated if they met a given threshold (score ≥ 160 in at least 130 bp in length) using rmBLASTn against the RepBase database. For the TE classification of superfamilies, we looked for the most related TE already annotated on RepBase. TE features, such as terminal repeats (terminal inverted repeats—TIRs and long terminal repeats—LTRs), open reading frames—ORFs,

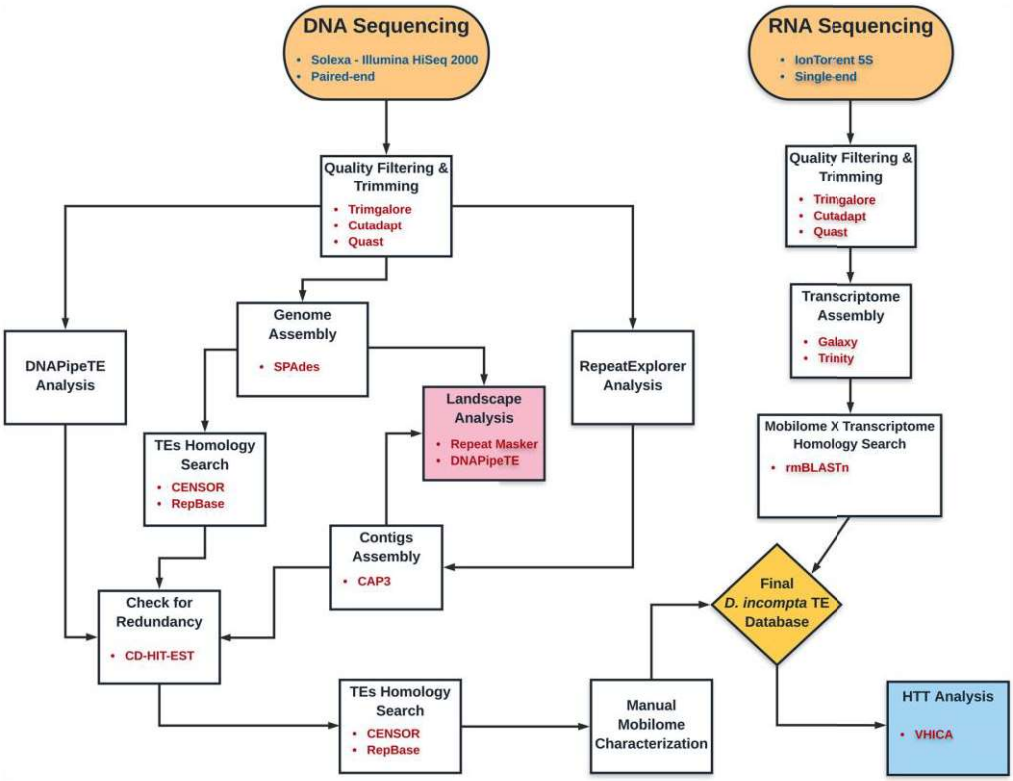


Fig. 1 Pipeline of bioinformatics analysis. Red text shows the tools used in the different analyses. Coloured boxes mark the start (orange) and the final results (blue, red and yellow)

target site duplication—TSD and total length, were manually annotated using UGENE software (Okonechnikov et al. 2012). In addition, TEs were classified as putative autonomous (PA), putative non-autonomous (PNA) or degenerated (DG) based on the integrity of their ORFs and the presence of TIRs or LTRs. LTR class I superfamilies had a further activity classification (PA*) due to the presence of a transposition viable structure but the absence of one LTR (or both). To identify and annotate new TEs, we used a threshold ≥ 80% of similarity with the corresponding TE in the database. Those TEs identified as new were given new names according to Wicker et al. (2007).

TE landscapes

TE landscapes were built using the RepeatMasker online tool (http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html). To compare the shape of TE landscapes observed between specialist and generalist flies, two approaches were used. In the first, the mode value observed in the landscapes of generalists and specialists was compared (see Supplementary Table S1). The statistical analysis was performed using a Welch's *t* test using R. In the second approach, the landscapes were assigned to one of three classes: (i) "L" when the mode was less than 5 at the Kimura two parameters (K2P) level; (ii) "bell curve," when the mode was higher than 5 at the K2P level; and (iii) "bimodal curve," when there were two modes and the second had at least half the value of the major mode. These distributions were statistically evaluated using a χ^2 test and Pearson correlation coefficient.

HTT investigation

To investigate the possible cases of horizontal transposon transfer (HTT) into the *D. incompta* genome, we used the Vertical and Horizontal Inheritance Consistence Analysis (VHICA) tool (Wallau et al. 2016a) with the same genes used in Bayesian analysis and using a Bonferroni multiple test correction parameter. *D. incompta* elements with a similarity higher than 90% to an annotated TE of *Drosophila* deposited in RepBase and with an ORF which codes for at least 100 amino acids were analysed. These selected ORFs (as well as the ORFs of related TEs of other Drosophilids) were used as query in a search for homologous sequences in the GenBank database (Clark et al. 2015). *Drosophila incompta* sequences with high similarity (above 90%) to TEs of distant organisms were not used in this analysis. However, these cases are described as putative HTT. The alignments were checked using the MEGA7 software (Kumar et al. 2018), and the dN/dS analysis was performed in DnaSP v.6.12 (Rozas et al. 2017).

Search for transcriptionally active TEs

In order to find out which TEs have their sequences transcribed in adult flies, we used the rmBLASTn tool (Altschul et al. 1990) with default parameters on our final dataset of annotated TEs from the *D. incompta* genome against the draft assembled transcriptome. After the search, we manually curated the output table using the following threshold values: score = 75; percentage of identity = 80%; alignment length = 100 bp; number of mismatches = 10; and number of gaps = 5.

Results

An overview of the *D. incompta* mobilome

The sequencing of the total *D. incompta* genome resulted in 31.447.883 paired-end raw reads. After quality filtering and trimming, the paired-end reads were merged resulting in a 47.631.683 reads file (4Gb) that was used for the ensuing analyses. The estimated sequencing coverage is 24,5 X, using the average of genome size for *Drosophila* belonging to virilis-repleta radiation.

The total RNA sequencing resulted in 12.346.389 raw single-end reads that after trimming and filtering became 8.816.476 reads and was used as the input for the transcriptome assembly and afterwards for a homology search of the mobilome against the transcriptome. Both assembly datasets and statistics can be assessed in Supplementary Table 3S.

The analysis and annotation performed by the DNAPipeTE software gave us an overview of the mobilome of *D. incompta*, indicating that 14.62% of its genome is comprised of TEs (Fig. 2a). This estimation deserves caution, since it was calculated by subtracting the reads used to assemble the TEs from the total reads obtained from Illumina sequencing. The genome size of *D. incompta* was not determined by flow cytometry or another independent procedure; hence, we

were not able to estimate precisely the real TE proportion of this genome. According to the DNApipeTE analysis, the most representative TE groups were *Helitron* (4.22%), LTR retrotransposons (3.19%), LINEs (2.8%) and other DNA transposons (1.75%) (Fig. 2a).

The analysis performed by RepeatExplorer software agrees with those of DNAPipeTE, showing that TEs represent 13.18% of the *D. incompta* genome, with class I comprising 7.14% and class II 6.04%. Of these, the two most abundant TEs were *Helitron* which accounted for 4.84% and LTR elements which accounted for 3.28% (Fig. 2b). Single and low copy DNAs complete the 23.8% of all repetitive sequences in the *D. incompta* genome. The *D. incompta* TE landscapes generated by the DNApipeTE analysis showed that many TEs are represented by young copies, indicating a low nucleotide divergence from consensus sequences (Fig. 2d).

Analysis of the curated dataset

Transposable elements identified by our pipeline were manually curated and their features, such as terminal repeats (TIRs, LTRs), ORFs and TSDs, were annotated. These annotations and sequences are available in Spreadsheets (Supplementary Tables 1S and 4S), summarised in Table 1 and described below.

Our general annotation characterised the mobilome of *D. incompta* as having 279 different elements of which 63% was LTR retrotransposons, 12% *LINEs*, 7% *Helitron*

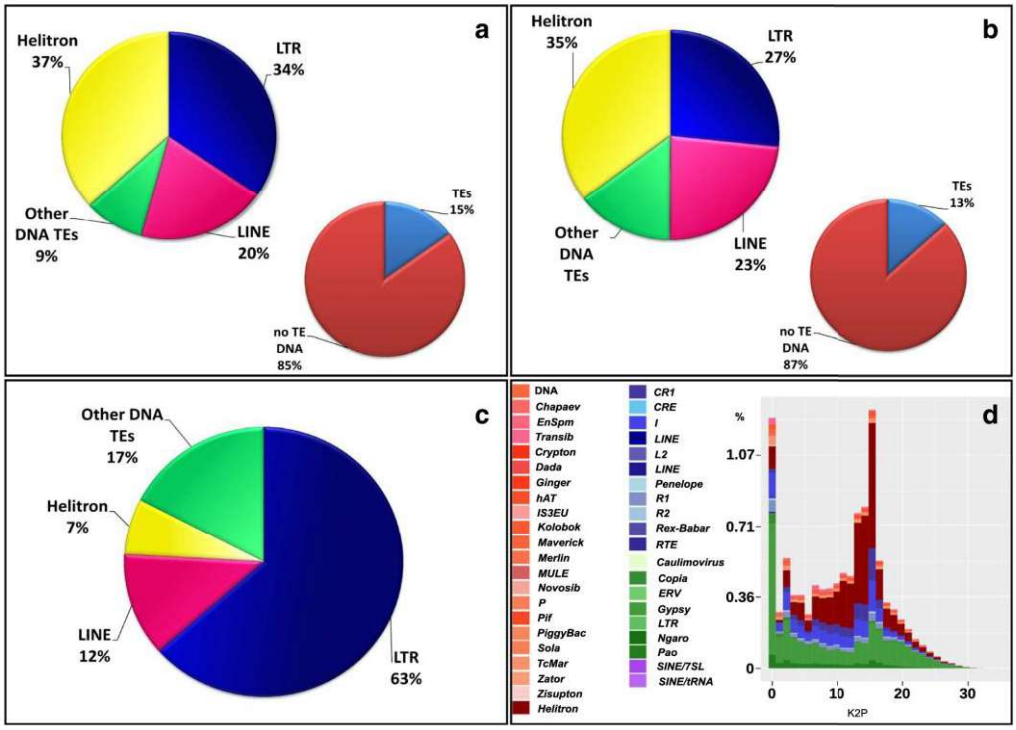


Fig. 2 **a** Percentage of TEs in the total assembled genome and the percentage of the major groups of TEs according to the DNAPipeTE analysis. **b** Percentage of TEs in the total assembled genome and the percentage of the major groups of TEs according to the RepeatExplorer analysis. **c** Percentage of each major group of TEs after manual curation. **d** Mobilome landscape of *D. incompta* generated by DNAPipeTE (X-axis = kimura substitution level; Y-axis = % of genome)

Table 1 Summary of manual curation of *D. incompta* TEs

Summary of TE annotation

Class	Superfamily	Annotated contigs	TEs	Families	New annotated TEs	PA TEs	PA* TEs	PNA TEs	DG TEs	Transcribed TEs
Class I	Gypsy	195	97	6	59	9	34	41	111	10
	Copia	40	20	1	16	9	5	5	21	2
	BEL	29	9	1	16	0	10	2	17	2
	Jockey	28	15	3	10	1	–	27	0	1
	R1	25	12	2	5	5	–	4	16	1
	CR1	9	1	1	1	2	–	0	7	1
	R2	7	1	1	0	0	–	2	5	1
	PENELOPE	3	1	1	1	1	–	1	1	0
	RTE	5	4	2	0	1	–	–	4	0
	Micropia	1	1	1	1	1	–	0	0	0
	Total class I	*342*	*161*	*19*	*109*	*29*	*49*	*82*	*182*	*18*
Class II	Helitron	101	20	2	8	3	–	9	89	5
	Mariner	68	36	7	18	9	–	11	48	0
	hAT	36	28	3	14	0	–	18	18	2
	Transib	30	21	1	7	0	–	8	22	0
	Harbinger	7	2	1	1	1	–	4	2	0
	PiggyBac	6	5	2	4	0	–	4	2	0
	Tetris	2	1	1	0	0	–	0	2	0
	Polinton	7	3	1	3	0	–	7	0	0
	Total class II	*257*	*116*	*18*	*55*	*13*	–	*61*	*183*	*7*
	Total	*509*	*277*	*37*	*164*	*42*	*49*	*143*	*365*	*25*

PA putative autonomous, *PA** LTR class I TEs putative autonomous however without one or both terminal repeats, *PNA* putative non-autonomous, *DG* degenerate

and 17% other DNA transposons (Fig. 2c). A total of 8.9% of these TEs were transcriptionally actives in adult flies.

Class I

Class I elements correspond to approximately 75.87% of all annotated TEs of *D. incompta*, of which, approximately 47.85% were classified as putative autonomous (PA) due to their respective total length, ORFs integrity and the presence of LTRs (Table 1). Some TEs were annotated as PA* on account of their absence of terminal repeats, despite the presence of conserved ORFs and protein domains.

Gypsy superfamily

This LTR superfamily is characterised by the presence of two main ORFs in their sequences (GAG and POL) and is found in most kingdoms of life (Wicker et al. 2007). They correspond to 60.35% of all class I TEs in *D. incompta*. Gypsy was by far the most abundant superfamily of retrotransposons, corresponding to 45.79% of the TEs annotated in this work. A total of 195 different clusters were assigned as belonging to this superfamily. Of these clusters, 79 did not show similarities to the TEs on the RepBase database and were designated news TEs. The other 116 elements were annotated as 38 already catalogued TEs from RepBase. Gypsy was the superfamily with the most putative autonomous TEs, corresponding to 51.32% of all Gypsy copies. Nine elements showed full length, functional ORFs and both LTRs, and these, in addition to the other 34 Gypsy elements, were classified as PA*. A total of ten Gypsy elements were transcriptionally active, of which six are PA*, two PNA and two corresponded to degenerated copies (DG).

Copia superfamily

Copia is the second most abundant superfamily in the *D. incompta* mobilome (9.52%) and also among the elements of class I (12.55%). A total of 40 clusters was found, of which 16 were designated as new elements. The other 24 clusters had high similarities to elements already deposited in databases, corresponding to five distinct TEs. Nine elements showed PA reconstructed sequences and five PA*. Of those classified as PA, seven clusters had 75 bp LTRs, which are shorter than that observed in all *Copia* elements described in RepBase. Only two elements of this superfamily were transcriptionally active in adults: a degenerated element and a PA element.

BEL superfamily

Also called the *Bel-Pao* superfamily, these LTR retrotransposons, similarly to the two superfamilies described before, have the GAG and POL ORFs (Wicker et al. 2007). *BEL* contains the third highest percentage of total TEs (8.14%) and of class I (10.73%). Of all the annotated TEs, this superfamily has the highest percentage of putative autonomous elements (72.47%). Of the 29 clusters obtained, 16 were annotated as new elements. Only the element Bel-8_Dmo-I was transcribed in adults. This element was classified as PA*.

Jockey superfamily

This non-LTR superfamily accounts for 4.27% of the total number of TEs and 5.63% of class I. It is composed of 28 clusters corresponding to 15 different elements, only one of which was marked as being putative autonomous. This superfamily contains the lowest number of PA/PA* among all class I TEs. Only one *Jockey* element was transcribed in adults and was classified as PNA.

R1 superfamily

This superfamily corresponds to 4.54% of mobilome and 5.98% of class I TEs. A total of 12 different TEs were annotated, five of which are new. Only one R1 element was transcriptionally active in adults, which was classified as PA.

Less represented class I superfamilies

Five other superfamilies had at last one representative TE annotated in the *D. incompta* genome. They were the following: *CR1*, *R2*, *Penelope*, *RTE* and *Micropia*. The sum of all these superfamilies represented 4.75% of the class I TEs and 3.61% of all annotated TEs. The *CR1* superfamily had nine clusters, all of the same element. This single TE did not present high similarity to any known TE and was classified as a new one. Two of their clusters were classified as PA and were active in the transcriptome data. The *R2* superfamily also had only one TE annotated with seven clusters. In this case, the annotated TE was already in the RepBase database (R2B_DM), and no element was classified as PA. *R2* also had correspondents in the transcriptome. The superfamilies *RTE* had five clusters corresponding to four different elements, three classified as DG and one as PA, none new TE and none were transcribed, and *Micropia*, which had only one copy classified as a new TE and also PA but were not transcribed.

Class II

DNA transposons of *D. incompta* represented approximately 24.13% of all the transposable elements, of which 11.20% were classified as PA. Eight superfamilies of class II TEs were annotated and are discriminated hereafter.

Helitron superfamily

This superfamily differs from the other DNA transposons by its singular transposition method via a rolling circle–like replication mechanism through a single-stranded DNA intermediate (Feschotte and Wessler 2001; Kapitonov and Jurka 2001). Another discriminating feature is the Rep/Hel domain in their ORFs, encoding a replication initiator and an helicase, respectively (Kapitonov and Jurka 2007). *Helitron* is the most abundant class II superfamily in the *D. incompta* mobilome, corresponding to 27.91% of all transposons and 6.73% of all TEs. This superfamily also had the highest number of annotated clusters (101) belonging to 20 different TEs. Only three of these different elements had clusters classified as PA. Eight new elements were found with one cluster each, and 12 TEs had a high similarity to TEs already catalogued. Five elements were transcriptionally active in adults, one of which

was designated as PA and the others were designated as DG or PNA elements.

Tc1/mariner superfamily

These TEs are ubiquitous in eukaryotes and are characterised by the presence of 5–27 bp TIRs as well as an ORF that encodes a transposase with a DDE domain (Hartl 2001; Wicker et al. 2007). Representing 24.6% of class II elements and 5.93% of all TEs, *mariner* was the most diverse superfamily of *D. incompta*. A total of 18 *mariner* elements were annotated as new transposons. The other 49 clusters corresponded to elements already present in the database and were classified as 18 different transposons. Nine elements were designated as PA, corresponding to 27.36% of the *mariner* elements. Three different Miniature Inverted-repeat Transposable Elements (MITEs) were found for *mariner* with very short length but conserved TIRs and TSDs. None of the elements of this superfamily was transcriptionally active in adults.

hAT superfamily

The *hAT* superfamily members accounted for 17.06% of class II elements and 4.11% of all the TEs in *D. incompta* mobilome. A total of 36 different clusters were found belonging to three families and 28 different TEs, of which 15 were new. No putative autonomous *hATs* were found. One MITE cluster was annotated. Two *hAT* elements were present in the transcriptome analysis.

Transib superfamily

This superfamily accounted for 17.06% of DNA transposons and 3.19% of all TEs. *Transib* elements are characterised by the presence of 9–60 bp TIRs, TSDs with 5 bp and an ORF with a DDE motif (Wicker et al. 2007). *Transib* had PA transposons neither detected nor transcriptionally active. Thirty sequences were annotated as *Transib*, nine of which were very similar and corresponded to one TE and the other 21 were distinct TEs, seven of which were new. Two MITEs were characterised in the *Transib* superfamily.

Less represented class II superfamilies

Harbinger, *PiggyBac*, *Tetris* and *Polinton* were characterised as less represented in the *D. incompta* mobilome. The sum of these superfamilies represented 4.15% of all the *D. incompta* TEs and 17.22% of those of class II. *Harbinger* had two TEs annotated in seven clusters; six of them were not new. The only one marked as new had just one cluster and was the only putative autonomous TE out of all the less represented class II superfamilies. The *PiggyBac* superfamily was represented by six sequences annotated as five TEs. Only one degenerated cluster was not classified as new; the other five copies represented four new TEs, all of which were also degenerated or truncated (PNA). Two degenerated *Tetris* members were annotated, and both had similarities to some TEs already annotated in the RepBase. The *Polinton* superfamily is known to represent the largest transposons (10–15+ kilobases) with many ORFs and high complexity (Kapitonov and Jurka 2005). We found three new *Polinton* elements in seven clusters, all designated as PNA. No transposons of these less represented superfamilies were transcribed in adults in our assay.

TE landscapes and niche amplitude

In order to test the hypothesis that the niche amplitude could be a significant factor in shaping the TE landscape, we annotated the TE landscape of 32 sequenced *Drosophila* genomes, of which 24 species were ecological generalists and 8 were specialists (see Supplementary Table 1). The TE landscapes depicted the relative abundance of the TE classes in the genome versus the Kimura two parameters divergence from the consensus. To compare the shape of TE landscapes observed between specialist and generalist flies, two approaches were used. In the first, the mode value observed in the landscapes of generalists and specialists was compared. This approach, although limited to a single parameter of the landscape shape, provided, as a quantitative value, an important descriptor related to the landscapes dynamics. In the second, the landscapes were assigned to one of three classes: L, bell curve and bimodal curve (Fig. 3).

The TE landscape patterns for 32 available *Drosophila* genomes were compared between species which are ecologically generalist and those that are specialists (Fig. 4). No significant differences were found between these groups. When the landscape modes were compared by Welch's t test, the calculated value was $t = 0.76347$ ($df = 10.311$, p value $= 0.4623$). When landscapes were classified in three classes, the calculated

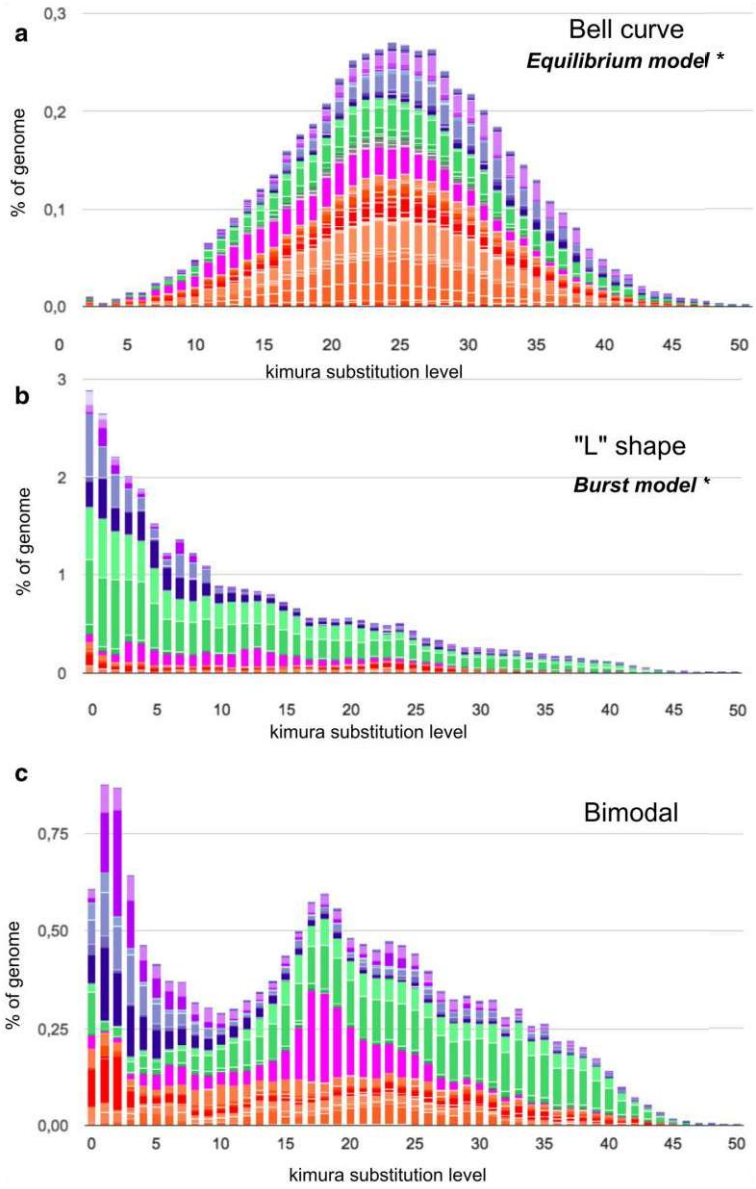


Fig. 3 Examples of TE landscapes and dynamics of transposable elements. **a** The landscape is a "bell curve" shape indicating the TE dynamics predicted by the equilibrium model (asterisk), which proposes that there is an equilibrium between transposition and excision over evolutionary time (*D. buzzatii*). **b** The landscape is an "L" type, where the mode value of the K2P distance is 0 (*D. ananassae*). This landscape is suggestive of TEs that are active in bursts and have a high turnover (the burst model (asterisk)). Many TE copies are very similar, indicating that most of them are active. **c** Bimodal landscape shape (*D. kikkawai*). (Asterisk, see discussion)

$\chi^2 = 1.9817$ ($p > 0.05$) and the Pearson correlation coefficient was $r = 0.1158$ (also not significant). However, visual inspection of Fig. 4 suggests that more species of the *Drosophila* subgenus had a TE landscape with the bell shaped curve than species of the *Sophophora* subgenus. Due to this, we decided to compare the landscapes between these groups, and this analysis showed they are statistically different. When the landscapes modes were compared by Welch's *t* test, the value was $t = 3.2259$ (df = 13.657, p value = 0.006269). When landscapes were classified in three classes, the calculated $\chi^2 = 11.565$ ($p < 0.01$) and the Pearson correlation coefficient was $r = 0.5841$ (also significant).

HTT

Three elements fit our criteria for HTT analysis using the VHICA software, which were 90% of similarity and an ORF encoding a minimum of 100 amino acids. These elements were *Mariner-9_Dan*, *Transib-5_Dbp* and *minos*. *Mariner-9_Dan* is an interesting case, since there is a strong HTT signal among elements found in *D. incompta* (subgenus *Drosophila*) and species of the *Sophophora* subgenus (*D. melanogaster*, *D. ananassae*, *D. kikkawai* and *D. willistoni*), but a weak HTT signal between *D. incompta* and the closer species of the subgenus *Drosophila* (*D. americana* and *D. virilis*) (Fig. 5a). *Transib-5_Dbp* is shared by South American flies of different subgenera (*D. incompta* and *D. willistoni*) and an Asian species (*D. bipectinata*) (Fig. 5b). The *minos* element appears to have been involved in several HTT events (Fig. 5c). Following the VHICA analysis, elements showing *dS* and codon bias incompatible with vertical transmission were found in *D. willistoni* (willistoni group); *D. erecta*, *D. yakuba* and *D. melanogaster* (melanogaster group); *D. mojavensis* and *D. hydei* (repleta group); and *D. incompta*.

Two suggestive HTT events involved *Helitron* elements and very distant species. The element *Helitron-1_Rpr* is shared with the Hemiptera *Rhodnius prolixus* with 99% nucleotide similarity. Furthermore, the element *Helitron-1_BM* is shared with the Lepidoptera *Bombyx mori* with 97% nucleotide similarity. The *mariner* elements *Smar25* and *Smar18* are found also in the planarian *Schmidtea mediterranea*, with a nucleotide similarity of 99% and 95%, respectively. One element found in *D. incompta* has a nucleotide similarity of more than 90% with the *mariner-1_Amel* element described

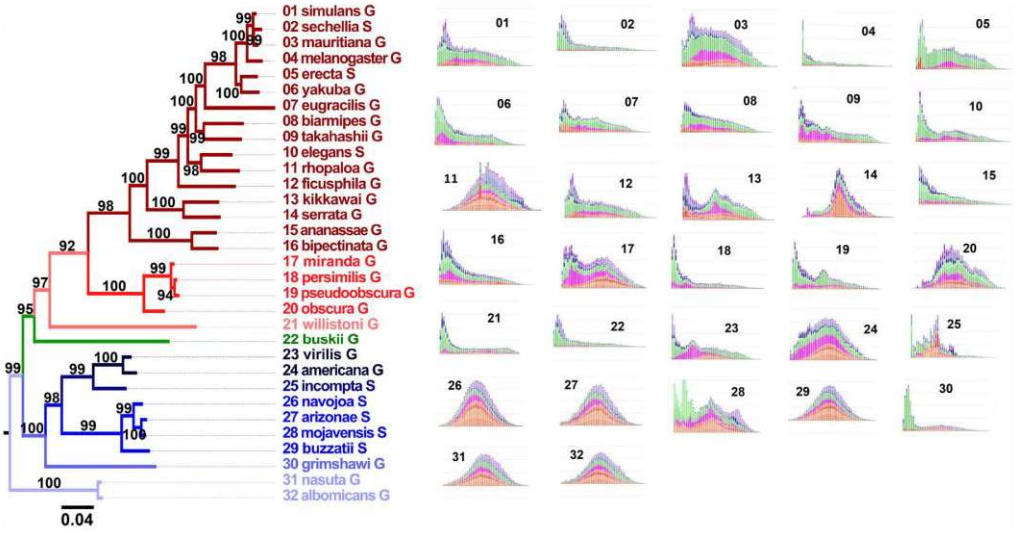


Fig. 4 Bayesian phylogenetic tree of 32 species of *Drosophila* and their mobilome landscapes. Shades of blue indicate the subgenus *Drosophila*. Shades of red indicate the subgenus *Sophophora*. The green branch indicates the subgenus *Dorsilopha*. G: generalist species; S: specialist species. Numbers before species specific epithet are relative to the landscape numbers. The phylogenetic relationship among the *Drosophila* species used in TE landscape comparison was reconstructed using Bayesian phylogenetic analyses (Supplementary Table 2S) of all 32 species and the mosquito *Aedes aegypti* as the outgroup

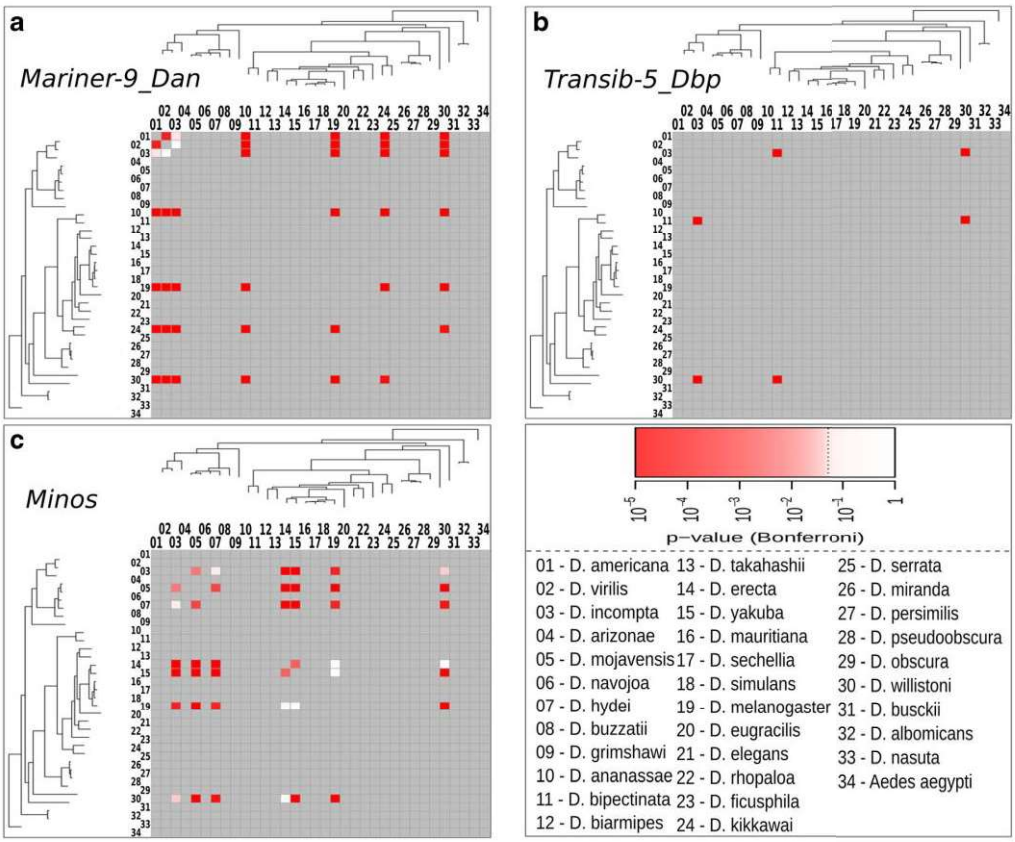


Fig. 5 VHICA result matrixes of possible cases of HTT involving *D. incompta* annotated TEs. **a** *Mariner-9_Dan* Rplot. **b** *Transib-5_Dbp* Rplot. **c** *minos* Rplot. Each square is coloured according to the colour bar of the *p* value calculated for the null hypothesis of vertical transmission

in *Apis mellifera*. These HTT events were not analysed using VHICA tools since the orthologous host genes are difficult to align with confidence.

Discussion

The TE content in eukaryotic genomes is highly variable, ranging from 3 to 5% in some yeasts to more than 90% in maize (Chénais et al. 2012). In *Drosophila*, the proportion of mobile elements varies from 4.65% in *D. busckii* to 30.80% in *D. suzukii* (Sessegolo et al. 2016). Our results estimate that the percentage of TEs in the *D. incompta* mobilome ranged from 13.18% (RepeatExplorer analyses) to 14.62% (DNAPipeTE analyses). These results are similar to those found for *D. virilis* (17.2%) (Sessegolo et al. 2016) which, among the analysed species, is the most phylogenetically related species to *D. incompta* (Robe et al. 2010).

The results obtained using automatic TE annotation tools were similar, highlighting *Helitron* as the most abundant group, followed by LTR elements and LINEs. From the *Drosophila* genomes annotated so far, it was only the most abundant TE in *D. buzzatii* (Rius et al. 2016), whereas LTR elements were always the most abundant TEs in the other *Drosophila* genomes, followed by LINEs (Clark et al. 2007; Sessegolo et al. 2016). In the manually curated annotation of the *D. incompta* mobilome, the LTR elements were the most abundant

group, followed by LINEs. The *Helitron* was the third most abundant group. This reduction is probably due to the threshold values in the final rmBLASTn analysis, in which very short sequences were removed, but which were not excluded in the automated analyses. To confirm this, we ran the rmBLASTn analysis with 30 bp as the threshold value for sequence length, and the percentage of *Helitron* increased significantly (data not shown).

Although several tools have been developed for mobilome annotation and characterisation, it is still a challenging task due to the inherent complexity of repetitive DNA. Hence, manual curation remains the best choice for an accurate and thorough TE characterisation including the identification of putative autonomous and active elements (Flutre et al. 2012; Hoen et al. 2015; Platt et al. 2016). The manual curation of the *D. incompta* mobilome allowed us to identify 279 elements, of which 164 were new. Importantly, 32.62% of the identified TEs were putatively autonomous (PA), of which 47.85% were class I elements and 11.20% were class II elements. However, few mobilomes have been characterised regarding the proportion of active elements. The *Drosophila melanogaster* genome is one of the best studied eukaryotic genomes and 30% of its elements are potentially autonomous (full length) and likely active (Kaminker et al. 2002; Barrón et al. 2014). A comparative study of the mobilome of *D. melanogaster* and other species of the *melanogaster* group (*D. simulans*, *D sechellia* and *D. yakuba*) found similar proportions of potentially autonomous TEs (Lerat et al. 2011). Thus, an important contribution of this study was the characterisation of the proportion of putative active elements in a species outside the *melanogaster* group and having a restricted ecology.

Two general models have been proposed to address the dynamics of TEs in genomes (Le Rouzic and Capy 2005; Barrón et al. 2014; Petersen et al. 2019). The first, the *equilibrium model*, assumes that each TE is transposed at a given rate and is subsequently removed by excision and purifying selection. The existence of an equilibrium between transposition and excision suggests, given that over time some TEs copies will be inactivated by mutations, the average divergence among copies will be associated with the average evolutionary time that TE copies are maintained in the genome. The TE landscapes are a very useful tool to characterise TE dynamics. Here we propose that following the equilibrium model, the expected TE landscape shape is represented by a typical bell curve (Fig. 3a). The second model of TE dynamics in genomes is the *burst model.* This model assumes that the transposition rate was not constant, given that TE transpositions can occur in bursts. If transposition occurs in bursts, the deleterious insertion needs be equipoised by fast TE removal through excision and purifying selection. In this scenario, the TE turnover is high and, in general, the TE copies will be very similar. Hence, the K2P divergence among copies is expected to be lower. The expected TE landscapes will have an L shape (Fig. 3b). Both patterns of TE landscapes have been found in the genomes analysed so far (Petersen et al. 2019). However, other shapes have also been found in the landscapes. In some genomes, the landscapes are bimodal or even multimodal. In others, the mode shows at a distance on the *Y* axis, producing an asymmetric curve. These asymmetric and bimodal landscapes might indicate that, in some genomes, the equilibrium of transposition/excision is intermediated by burst events.

The TE landscape of *D. incompta* shows the bimodal shape, with a significant number of very similar TE copies (a low K2P divergence) and a second mode at K2P 0.15. This shape reflects very well the TE annotation, which is composed by 32% of putative autonomous elements with the remaining copies showing a variable degree of degeneration.

Several studies have suggested that a low genetic variability and an increase in population genetic structure are expected in specialist species, since the effective population size of these organisms is limited by the limited resources they rely on and by environmental fragmentation (Kelley et al. 2000; Packer et al. 2005; Habel et al. 2009). In addition, these species likely face bottleneck events more frequently (Nossil 2002). However, generalist species have a larger availability of resources, and the populations are less genetically structured and have more genetic variation and gene flow (Kelley et al. 2000; Brouat et al. 2003; Zayed et al. 2005).

The comparative analyses of landscape shapes between specialist and generalist flies suggest that niche amplitude is not a major factor in shaping the TE dynamics in genomes, while other factors can be involved in TEs dynamics as, for example, the epigenetic mechanisms for TEs silencing and the chromatin features in that TEs are located. In the present study, no differences were found among TE

landscape between specialist and generalist flies. However, we found significant differences between the TE landscapes of flies of *Drosophila* and *Sophophora* subgenera. Petersen et al. (2019) found that lineage-specific TE landscapes occurred during the evolution of several insect species, probably reflecting differences in the evolutionary processes of each lineage. The observed TE landscape shape in each subgenus, which is preponderantly a bell curve in the *Drosophila* subgenus and an L shape in the *Sophophora* subgenus, suggests that in the first, the TE dynamics are under the equilibrium model of evolution, and in the second, they are under the burst model (Barrón et al. 2014).

Horizontal transposon transfer (HTT) is an important factor shaping host species' genomic evolution (Schaack et al. 2010; Peccoud et al. 2017; Wallau et al. 2018). *Drosophila incompta*, as a species of restrict ecology, should be less prone to HTT events, due its reduced ecological contacts (Venner et al. 2017). However, this hypothesis is not supported by the evidence, since several HTT cases have been reported involving this species (Ortiz et al. 2014). In the present study, more cases are added. *Mariner-9_Dan* is an interesting case, since there is a strong HTT signal among the elements found in *D. incompta* (subgenus *Drosophila*) and species of the *Sophophora* subgenus (*D. melanogaster*, *D. ananassae*, *D. kikkawai* and *D. willistoni*), but only a weak HTT signal between *D. incompta* and the more closely related species of the subgenus *Drosophila* (*D. americana* and *D. virilis*). Members of the *melanogaster* group are invasive species that have colonised South America and are now in geographic contact with *D. incompta* and *D. willistoni*. The *minos* element also involves *Drosophila* species of the *melanogaster* group (*D. melanogaster*, *D. yakuba*), *D. willistoni*, and two species of the *repleta* group (*D. mojavensis* and *D. hydei*). We point out that for this last TE, the results of the VHICA analysis might be underestimating HTT cases among *Drosophila* species since the copies of *minos* we identified were classified as degenerated or putative non-autonomous, and the one that was used by VHICA had an ORF of less than 140 aa, which, according to Wallau et al. (2016a) could bias the analysis. Interestingly, Wallau et al. (2016b) described a horizontal transfer of *Wolbachia* between *D. melanogaster* and *D. incompta*. *Wolbachia* has been suggested as a possible vector for HTT (Loreto et al. 2008).

The descriptions of HTT events involving distantly related species have increased recently, suggesting that these events are more frequent than was originally thought (Dotto et al. 2018; Venner et al. 2017). In the present study, we found *Helitron* elements with 97 to 99% similarity to elements described in Hemiptera and Lepidoptera. Heringer et al. (2017) showed that a *Helitron* element is shared by the distantly related insects *Drosophila* (Diptera), *Bombyx mori* (Lepidoptera) and *Cotesia* (Hymenoptera). The authors highlighted that this *Helitron* is also found in a polydnavirus which is transmitted to insects by Braconid wasps. *D. incompta* is parasitised by Braconid wasps, and although the *Helitron* elements described here were not found in the sequenced genome of this parasitoid wasps (Ortiz et al. 2014; da Silva et al. 2018), this wasp could be a vector for a polydnavirus carrying *Helitron* elements from distantly related species.

In this study, we provide an accurate annotation of mobilome of *D. incompta*, which, as a flower breeder, is a fly with a very restricted ecology. We identified 279 elements, 164 of which were new. Of the identified TEs, 32.62% were putatively autonomous and 8.9% were transcriptionally active. The restricted ecology of *D. incompta* does not seem to be the major determinant for dynamics of TEs in this fly, since the proportion and diversity of TE in its genome is similar to that of other *Drosophila* species. This conclusion is reinforced by the comparison of the landscapes of generalist and specialists flies, which do not show a difference. Restricted ecology, therefore, does not appear to be a barrier to horizontal transposon transfer, since eight cases have been identified in *D. incompta*.

Acknowledgements We are grateful to Dr. Lizandra Robe and two anonymous reviewers for suggestions.

Author contribution statement PMF and EL conceived and designed research. PMF, RMD and GLW conducted experiments and analysed data. PMF, GLW and EL wrote the manuscript. All authors read and approved the manuscript.

Funding information This study was supported by research grants and fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Pronex-Fapergs (16/2551—0000 499-4).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3–W10. https://doi.org/10.1093/nar/gkw343

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bächli G (2018) TaxoDros: the database on taxonomy of Drosophilidae. https://www.taxodros.uzh.ch/ Accessed 26 November 2018

Barrón MG, Fiston-Lavier AS, Petrov DA, González J (2014) Population genomics of transposable elements in Drosophila. Annu Rev Genet 48:561–581. https://doi.org/10.1146/annurev-genet-120213-092359

Bankevich A, Nurk S, Antipov D, Gurevich AA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Brouat C, Sennedot F, Audiot P, Leblois R, Rasplus JY (2003) Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. Mol Ecol 12:1731–1745. https://doi.org/10.1046/j.1365-294X.2003.01861.x

Chénais B, Caruso A, Hiard S, Casse N (2012) The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene 509:7–15. https://doi.org/10.1016/j.gene.2012.07.042

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2015) GenBank. Nucleic Acids Res 44:D67–D72. https://doi.org/10.1093/nar/gkv1276

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman T, Kellis M, Gelbart W et al (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203–218. https://doi.org/10.1038/nature06341

da Silva AF, Dezordi FZ, Loreto ELS, Wallau GL (2018) Drosophila parasitoid wasps bears a distinct DNA transposon profile. Mob DNA 9:23. https://doi.org/10.1186/s13100-018-0127-2

de Ré FC, Wallau GL, Robe LJ, Loreto ELS (2014) Characterization of the complete mitochondrial genome of flower-breeding Drosophila incompta (Diptera, Drosophilidae). Genetica 142:525–535. https://doi.org/10.1007/s10709-014-9799-9

Dotto BR, Carvalho EL, Freitas A, et al (2018) HTT-DB: new features and updates. Database 2018:bax102. https://doi.org/10.1093/database/bax102, HTT-DB: new features and updates

Feschotte C, Wessler SR (2001) Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. PNAS 98:8923–8924. https://doi.org/10.1073/pnas.171326198

Flutre T, Permal E, Quesneville H (2012) Transposable element annotation in completely sequenced eukaryote genomes. In: Plant Ttranspos able Eelements (eds Grandbastien M-A & Casacuberta J). Topics in current genetics, vol. 24, Springer, Heidelberg, pp. 17–40

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M (2015) De novo assembly and annotation of the Asian Tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). Gen Biol and Evol 7:1192–1205. https://doi.org/10.1093/gbe/evv050

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644–652 https://www.nature.com/articles/nbt.1883

Guillín ER, Rafael V (2018) Two new species in the Drosophila flavopilosa and Drosophila morelia species groups (Diptera: Drosophilidae) in the eastern Andes of Ecuador. Rev Peru Biol 25:69–74. https://doi.org/10.15381/rpb.v25i2.14684

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Habel JC, Augenstein B, Meyer M, Nève G, Rödder D, Assmann T (2009) Population genetics and ecological niche modelling reveal high fragmentation and potential future extinction of the endangered relict butterfly *Lycaena helle*. In: Habel JC, Assmann T (eds) Relict species. Springer, Berlin, Heidelberg

Hartl DL (2001) Discovery of the transposable element mariner. Genetics 157:471–475

Heringer P, Dias GB, GCS K (2017) A horizontally transferred autonomous Helitron became a full polydnavirus segment in *Cotesia vestalis*. G3 7:3925–3935. https://doi.org/10.1534/g3.117.300280

Hoen DR, Hickey G, Bourque G, Casacuberta J, Cordaux R, Feschotte C, Fiston-Lavier AS, Hua-van A, Hubley R, Kapusta A, Lerat E, Maumus F, Pollock DD, Quesneville H, Smit A, Wheeler TJ, Bureau TE, Blanchette M (2015) A call for benchmarking transposable element annotation methods. Mob DNA 6:13. https://doi.org/10.1186/s13100-015-0044-6

Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P (2011) The struggle for life of the genome's selfish architects. Biol Direct 6:19. https://doi.org/10.1186/1745-6150-6-19

Huang X, Madan A (1998) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467. https://doi.org/10.1159/000084979

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M, Celniker SE (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics

perspective. Genome Biol 3:research0084.1. https://doi.org/10.1186/gb-2002-3-12-research0084

Kapitonov VV, Jurka J (2001) Rolling-circle transposons in eukaryotes. PNAS 98:8714–8719. https://doi.org/10.1073/pnas.151269298

Kapitonov VV, Jurka J (2005) Self-synthesizing DNA transposons in eukaryotes. PNAS 103:4540–4545. https://doi.org/10.1073/pnas.0600833103

Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23:521–529. https://doi.org/10.1016/j.tig.2007.08.004

Kelley ST, Farrell BD, Mitton JB (2000) Effects of specialization on genetic differentiation in sister species of bark beetles. Heredity 84:218–227. https://doi.org/10.1046/j.1365-2540.2000.00662.x

Kelley JL, Peyton JT, Fiston-Lavier AS, Teets NM, Yee MC, Johnston JS, Bustamante CD, Lee RE, Denlinger DL (2014) Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nat Commum 5:4611. https://doi.org/10.1038/ncomms5611

Krueger F (2015) Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/]

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

Le Rouzic A, Capy P (2005) The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 169:1033–1043. https://doi.org/10.1534/genetics.104.031211

Lerat E, Burlet N, Biémont C, Vieira C (2011) Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene 473:100–109. https://doi.org/10.1016/j.gene.2010.11.009

Loreto ELS, Carareto CMA, Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity 100:545–554. https://doi.org/10.1038/sj.hdy.6801094

Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404

Markow TA (2015) The secret lives of Drosophila flies. eLife 4. https://doi.org/10.7554/eLife.06793

McCullers TJ, Steiniger M (2017) Transposable elements in *Drosophila*. Mob Genet Elem 7:1–18. https://doi.org/10.1080/2159256X.2017.1318201

Napp M, Brncic D (1978) Eletrophoretic variability in two closely related Brazilian species of the flavopilosa species group of *Drosophila*. Braz J Genet 1:1–10

Nossil P (2002) Transition rates between specialization and generalization in phytophagous insects. Evolution 56:1701–1706. https://doi.org/10.1111/j.0014-3820.2002.tb01482.x

Novak P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics. 29:792–793. https://doi.org/10.1093/bioinformatics/btt054

O'Grady PM, DeSalle R (2018) Phylogeny of the genus Drosophila. Genetics 209:1–25. https://doi.org/10.1534/genetics.117.300583

Okonechnikov K, Golosova O, Fursov M the UGENE team (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166–1167. https://doi.org/10.1093/bioinformatics/bts091

Ortiz MF, Wallau GL, Graichen DÂS, Loreto ELS (2014) An evaluation of the ecological relationship between *Drosophila* species and their parasitoid wasps as an opportunity for horizontal transposon transfer. Mol Gen Genomics 290:67–78. https://doi.org/10.1007/s00438-014-0900-y

Packer L, Zayed A, Grixti JC et al (2005) Conservation genetics of potentially endangered mutualism: reduced levels of genetic variation in specialist vs. generalist bees. Conserv Biol 19:195–202. https://doi.org/10.1111/j.1523-1739.2005.00601.x

Peccoud J, Loiseau V, Cordaux R, Gilbert C (2017) Massive horizontal transfer of transposable elements in insects. PNAS 114:4721–4726. https://doi.org/10.1073/pnas.1621178114

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof B (2019) Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol 19:11. https://doi.org/10.1186/s12862-018-1324-9

Picot S, Wallau GL, Loreto EL, Heredia FO, Hua-Van A, Capy P (2008) The mariner transposable element in natural populations of *Drosophila simulans*. Heredity (Edinb) 101(1):53–59. https://doi.org/10.1038/hdy.2008.27

Piégu B, Bire S, Arensburger P, Bigot Y (2015) A survey of transposable element classification systems–a call for a fundamental update to meet the challenge of their diversity and complexity. Mol Phylogenet Evol 86:90–109. https://doi.org/10.1016/j.ympev.2015.03.009

Platt RN, Blanco-Berdugo L, Ray DA (2016) Accurate transposable element annotation is vital when analyzing new genome assemblies. Genome Biol Evol 8:403–410. https://doi.org/10.1093/gbe/evw009

Rio DC, Ares M, Hannon GJ, Nilsen TW (2010) Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc 2010:pdb–prot5439

Rius N, Guillén Y, Delprat A, Kapusta A, Feschotte C, Ruiz A (2016) Exploration of the *Drosophila buzzatii* transposable element content suggests underestimation of repeats in *Drosophila* genomes. BMC Genomics 17:344. https://doi.org/10.1186/s12864-016-2648-8

Robe LJ, Loreto ELS, Valente VLS (2010) Radiation of the "*Drosophila*" subgenus (Drosophilidae, Diptera) in the Neotropics. J Zool Syst Evol Res 48:310–321. https://doi.org/10.1111/j.1439-0469.2009.00563.x

Robe LJ, De Ré FC, Ludwig A, Loreto ELS (2013) The *Drosophila flavopilosa* species group (Diptera, Drosophilidae): an array of exciting questions. Fly 7:59–69. https://doi.org/10.4161/fly.23923

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Santos R, Vilela C (2005) Breeding sites of Neotropical Drosophilidae (Diptera): IV. Living and fallen flowers of

Sessea brasiliensis and *Cestrum* spp. (Solanaceae). Revista Brasileira de Entomologia 49:544–551

Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537–547. https://doi.org/10.1016/j.tree.2010.06.001

Sepel LMN, Golombiesky RM, Napp M, Loresto ELS (2000) Seasonal fluctuation os *Drosophila cestri* and *Drosophila incompta*, two species of flavopilosa group. Drosophila Information Service 83:122–126

Sessegolo C, Burlet N, Haudry A (2016) Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. Biol Lett 12:20160407. https://doi.org/10.1098/rsbl.2016.0407

Siefert JL (2009) Defining the mobilome. Methods Mol Biol 532: 13–27. https://doi.org/10.1007/978-1-60327-853-9_2

Smit AFA, Hubley R, Green P (2013-2016) RepeatMasker Open-4.0, http://www.repeatmasker.org. Accessed 18 December 2018

Venner S, Miele V, Terzian C, Biémont C, Daubin V, Feschotte C, Pontier D (2017) Ecological networks to unravel the routes to horizontal transposon transfers. PLoS Biol 15:e2001536. https://doi.org/10.1371/journal.pbio.2001536

Vieira C, Lepetit D, Dumont S, Biémont C (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. Mol Biol Evol 16:1251–1255

Vieira C, Nardon C, Arpin C, Lepetit D, Biemont C (2002) Evolution of genome size in Drosophila. Is the invader's genome being invaded by transposable elements? Mol Biol Evol 19:1154–1161

Wallau GL, Capy P, Loreto ELS, Le Rouzic A, Hua-Van A (2016a) VHICA, a new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within Drosophila. Mol Biol Evol 33: 1094–1109. https://doi.org/10.1093/molbev/msv3

Wallau GL, da Rosa MT, De Ré FC, Loreto ELS (2016b) *Wolbachia* from *Drosophila incompta*: just a hitchhiker shared by *Drosophila* in the New and Old World? Insect Mol Biol 25(4):487–499. https://doi.org/10.1111/imb.1223741

Wallau GL, Vieira C, Loreto ELS (2018) Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. Mob DNA 9(6). https://doi.org/10.1186/s13100-018-0112-9

Weizhong L, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Wheeler M, Takada H, Brncic D (1962) The flavopilosa species group of *Drosophila*. The University of Texas Publication pp 395–413

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

Zayed A, Packer L, Grixti JC, Ruz L, Owen RE, Toro H (2005) Increased genetic differentiation in a specialist versus a generalist bee: implications for conservation. Conserv Genet 6: 1017–1026. https://doi.org/10.1007/s10592-005-9094-5

CAPÍTULO IV

CONCLUSÕES GERAIS E PERSPECTIVAS

Espécies de ecologia restrita têm sido uma interessante fonte de investigações de questões evolutivas devido às suas características especializadas de interação com seus nichos específicos. Lang et al. (2012), por exemplo, demonstram como pequenas mutações em um único gene são capazes de mudar a ecologia de uma espécie, tornando-a especialista. Nesse sentido, *Drosophila incompta* apresenta-se como um modelo ideal para estudos evolutivos, visto que apresenta ecologia restrita a plantas do gênero *Cestrum*, não explorando outro recurso. Neste trabalho conseguimos demonstrar mais um fator que reforça a especificidade de exploração de nicho de *D. incompta* com o experimento de preferência olfativa, no qual mostramos expressiva preferência dessa espécie pelos odores provenientes de flores de *Cestrum* em comparação com outros odores: de banana fermentada, de extrato de laranja fermentada, o qual também vimos pelo mesmo experimento ser a preferência de *D. melanogaster* em relação aos odores de banana fermentada, e de extrato de flores de *Brunfelsia uniflora*.

Uma das formas mais exploradas para estudar a evolução das espécies de ecologia restrita e as relações dessas espécies com seus nichos restritos é através de análises de quimiorreceptores, genes que codificam proteínas as quais interagem com compostos químicos ambientais e por isso estariam mais propensos às forças evolutivas. Ao caracterizarmos e analisarmos o repertório dos receptores gustatórios (GRs) e olfativos (ORs) de *D. incompta*, conseguimos mostrar padrões de evolução molecular dessas famílias gênicas dentro do gênero *Drosophila*, pois detectamos possíveis casos de perdas gênicas ancestrais, visto que vários genes não encontrados em *D. incompta* também se mostraram ausentes para as espécies não pertencentes ao agrupamento não-*Sophophora*. Neste âmbito, mostramos que a média geral do número de *loci* entre espécies pertencentes e não pertencentes ao subgênero *Sophophora* é significantemente menor no segundo grupo. Além dessa discrepância no número total de loci entre os grupos foi possível também observar sinais de seleção positiva em 12 genes quimiorreceptores de *D. incompta*, e em alguns desses genes ainda percebemos relações de interação com químicos

ambientais presentes nos substratos explorados por *D. incompta.* Das duas outras famílias gênicas investigadas nesse trabalho, verificamos que, apesar da discrepância se mostrar menor, espécies não pertencentes ao subgênero *Sophophora* também possuem menor número de *loci* de genes da família *Yellow*, sendo *D. incompta* a que registrou o menor número geral, podendo este fator indicar relação com sua pigmentação corporal, a qual é caracterizada pela total ausência de ornamentos e marcas escuras.

Padrões de evolução genômica associados a ecologia restrita de *D. incompta* também foram avaliados pela caracterização comparativa dos repertórios de Elementos transponíveis (TEs) de diferentes espécies. Esses elementos vêm tendo cada vez mais destaque em diversos campos da genética e genômica, dado a sua grande capacidade de mobilização e, consequentemente, de geração de variabilidade genética a seus hospedeiros. Nesse sentido, o mobiloma de uma espécie pode prover informações importantes sobre a composição e evolução do seu genoma, assim como a paisagem dos TEs de uma espécie é capaz de contar a história dos elementos naquela espécie e indicar, além de dados quantitativos e qualitativos, informações sobre atividade e viabilidade dos TEs. Mais de 30 espécies de *Drosophila* contam hoje com o genoma sequenciado e disponível, contudo poucas possuem o seu mobiloma anotado minuciosamente. Aqui fizemos uma anotação criteriosa e cuidadosa do mobiloma de *D. incompta*, onde anotamos 277 TEs, dos quais 164 novas descrições para a espécie alvo. Além disso, identificarmos 42 elementos putativamente autônomos e portanto potencialmente ativos, bem como 25 fragmentos de RNA que correspondem a TEs sendo transcritos, mais uma vez indicando potencial atividade.

Embora a transferência horizontal de elementos transponíveis (HTT) ainda careça de estudos mais amplos e aprofundados, ela é considerada como uma das mais importantes formas através das quais a dinâmica dos genomas pode gerar variabilidade genética. Além disso, a mesma é capaz de renovar o ciclo de vida de alguns TEs, fazendo com que os mesmos invadam genomas onde serão um componente novo e terão menor resistência, podendo evoluir mais facilmente através

de duplicações e se diversificar. Com as análises do mobiloma de *D. incompta* e comparações com bancos de dados de TEs, conseguimos inferir HTTs recentes para TEs que achamos em *D. incompta*, inclusive entre espécies distantes, tanto evolutiva quanto geograficamente.

Achados em ambos os artigos indicam distância evolutiva separando espécies do subgênero *Sophophora* e espécies não pertencentes a esse gênero: a diferença no número de quimiorreceptores entre os grupos e a diferença na análise modal das paisagens de TEs indicam uma relação de ancestralidade em tais eventos evolutivos, concordando com as relações mais aceitas atualmente (compiladas em O'Grady e DeSalle 2018) onde o subgênero *Sophophora* seria o clado ancestral, e tais diferenças evolutivas encontradas seriam então sinapomorfias dos outros subgêneros não-*Sophophora*.

Por fim, essa tese contribuiu de diferentes maneiras para o entendimento da evolução em organismos especialistas e com ecologia restrita, como é o caso de *D. incompta*. Estudos futuros onde mais genes com interação com o ambiente, como por exemplo genes de detoxificação e até genes relacionados ao ciclo circadiano de espécies de ecologia e/ou distribuição restrita e/ou específicas, certamente contribuirão para um maior entendimento dessa parcela das espécies que têm essas particularidades com seus nichos e vivenciam a evolução de forma particular e diferente em relação às espécies generalistas e/ou cosmopolitas. Para a questão de aprofundar-se no entendimento sobre *D. incompta*, experimentos que relacionam a particularidade de espera pelas flores de *Cestrum*, que não florescem o ano todo, com a genética e genômica dessa espécie podem contribuir para o completo entendimento do ciclo de vida e da exploração dos recursos desta espécie tão particular.

REFERÊNCIAS BIBLIOGRÁFICAS

Ågren JA, Wright SI. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? Chromosome research. 2011 Aug 1;19(6):777.

Almeida FC, Sanchez-Gracia A, Campos JL, Rozas J. Family size evolution in Drosophila chemosensory gene families: a comparative analysis with a critical appraisal of methods. Genome biology and evolution. 2014 Jul 1;6(7):1669-82.

Anwar SL, Wulaningsih W, Lehmann U. Transposable elements in human cancer: causes and consequences of deregulation. International journal of molecular sciences. 2017 May;18(5):974.

Arkhipova IR, Yushenova IA. Giant transposons in eukaryotes: Is bigger better? Genome biology and evolution. 2019 Mar;11(3):906-18.

Bächli G (2020) The database on taxonomy of Drosophilidae. https://www.taxodros.uzh.ch/

Barrios-Leal DY, Neves-da-Rocha J, Manfrin MH. Genetics and distribution modeling: the demographic history of the cactophilic *Drosophila buzzatii* species cluster in open areas of South America. Journal of Heredity. 2019 Jan 7;110(1):22-33.

Biémont C. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics. 2010 Dec 1;186(4):1085-93.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL. Ten things you should know about transposable elements. Genome biology. 2018 Dec;19(1):1-2.

Brito NF, Moreira MF, Melo AC. A look inside odorant-binding proteins in insect chemoreception. Journal of Insect Physiology. 2016 Dec 1;95:51-65.

Brncic D. Ecological and cytogenetic studies of Drosophila flavopilosa, a neotropical species living in Cestrum flowers. Evolution. 1966 Mar 1:16-29.

Brncic D. note on the *flavopilosa* group of species of *Drosophila* in Rio Grande do Sul, Brazil, with the description of two new species (Diptera, Drosophilidae). Revista Brasileira de Biologia. 1978.

Brncic D. The flavopilosa group of species as an example of flower-breeding species. The genetics and biology of *Drosophila*. 1983;6:360-77.

Burns KH. Transposable elements in cancer. Nature Reviews Cancer. 2017 Jul;17(7):415-24.

Burns KH, Boeke JD. Human transposon tectonics. Cell. 2012 May 11;149(4):740-52.

Chénais B. Transposable elements in cancer and other human diseases. Current cancer drug targets. 2015 Jan 1;15(3):227-42.

Cordeiro J, de Oliveira JH, Schmitz HJ, Vizentin-Bugoni J. High niche partitioning promotes highly specialized, modular and non-nested florivore–plant networks across spatial scales and reveals drivers of specialization. Oikos. 2020 May;129(5):619-29.

David JR, Capy P. Genetic variation of *Drosophila melanogaster* natural populations. Trends in Genetics. 1988 Apr 1;4(4):106-11.

De Ré FC, Wallau GL, Robe LJ, Loreto EL. Characterization of the complete mitochondrial genome of flower-breeding *Drosophila incompta* (Diptera, Drosophilidae). Genetica. 2014 Dec 1;142(6):525-35.

De Ré FC. *Evolução molecular e padrões macro e micro evolutivos em Drosophila incompta (Diptera, Drosophilidae)* (Tese de doutorado, Universidade Federal de Santa Maria). 2016.

De Ré FC, Robe LJ, Wallau GL, Loreto EL. Inferring the phylogenetic position of the *Drosophila flavopilosa* group: Incongruence within and between mitochondrial and nuclear multilocus datasets. Journal of Zoological Systematics and Evolutionary Research. 2017 Aug;55(3):208-21.

DeSalle R, Grimaldi DA. Morphological and molecular systematics of the Drosophilidae. Annual review of ecology and systematics. 1991 Nov;22(1):447-75.

Elliott TA, Gregory TR. Do larger genomes contain more diverse transposable elements? BMC evolutionary biology. 2015 Dec;15(1):1-0.

Etges WJ. Evolutionary genomics of host plant adaptation: insights from *Drosophila*. Current opinion in insect science. 2019 Dec 1;36:96-102.

Fonseca PM. Padrões evolutivos de *Zygothrica vittimaculosa* burla, 1956 (Diptera: Drosophilidae) no Brasil, com identificação e descrição de uma espécie-irmã. (Dissertação de mestrado, Universidade Federal de Santa Maria). 2015.

Fonseca PM, Loreto EL, Gottschalk MS, Robe LJ. Cryptic diversity and speciation in the *Zygothrica* genus group (Diptera, Drosophilidae): the case of *Z. vittimaculosa* Wiedemann. Insect Systematics & Evolution. 2017 Jun 13;48(3):285-313.

Gardiner MM, Landis DA, Gratton C, Schmidt N, O'Neal M, Mueller E, Chacon J, Heimpel GE, DiFonzo CD. Landscape composition influences patterns of native and exotic lady beetle abundance. Diversity and Distributions. 2009 Jul;15(4):554-64.

Gautério TB, Machado S, da Silva Loreto EL, Gottschalk MS, Robe LJ. Phylogenetic relationships between fungus-associated Neotropical species of the genera *Hirtodrosophila*, *Mycodrosophila* and *Zygothrica* (Diptera, Drosophilidae), with insights into the evolution of breeding sites usage. Molecular Phylogenetics and Evolution. 2020 Apr 1;145:106733.

Gerlach, J., editor (2009). The Diptera of the Seychelles. Pensoft, Sofia, Bulgaria.

Gilbert C, Peccoud J, Cordaux R. Transposable Elements and the Evolution of Insects. Annual Review of Entomology. 2020 Sep 15;66.

Gloss AD, Vassao DG, Hailey AL, Nelson Dittrich AC, Schramm K, Reichelt M, Rast TJ, Weichsel A, Cravens MG, Gershenzon J, Montfort WR. Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the Drosophilidae. Molecular Biology and Evolution. 2014 Sep 1;31(9):2441-56.

Gottschalk M, Hofmann P, Valente V. Diptera, Drosophilidae: historical occurrence in Brazil. Check list. 2008 Jan 11;4:485.

Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, Gogol-Döring A, Kapitonov V, Diem T, Dalda A, Jurka J. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nature communications. 2016 Mar 2;7(1):1-2.

Guillín ER, Rafael V. Two new species in the *Drosophila flavopilosa* and *Drosophila morelia* species groups (Diptera: Drosophilidae) in the eastern Andes of Ecuador. Revista peruana de biología. 2018 May 1;25(2):69-74.

Hancks DC, Kazazian Jr HH. Active human retrotransposons: variation and disease. Current opinion in genetics & development. 2012 Jun 1;22(3):191-203.

Hofmann PR. Variabilidade genética em espécies de nível ecológico restrito. Cienc Cult. 1984;37:579-81.

Jangam D, Feschotte C, Betrán E. Transposable element domestication as an adaptation to evolutionary conflicts. Trends in Genetics. 2017 Nov 1;33(11):817-31.

Jones CD. The genetics of adaptation in *Drosophila sechellia*. In Genetics of Adaptation 2005 (pp. 137-145). Springer, Dordrecht.

Kapitonov VV, Jurka J. RAG1 core and V (D) J recombination signal sequences were derived from Transib transposons. PLoS Biol. 2005 May 24;3(6):e181.

Kazazian HH. Mobile elements: drivers of genome evolution. Science. 2004 Mar 12;303(5664):1626-32.

Kelley JL, Peyton JT, Fiston-Lavier AS, Teets NM, Yee MC, Johnston JS, Bustamante CD, Lee RE, Denlinger DL. Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nature communications. 2014 Aug 12;5(1):1-8.

Kim YJ, Lee J, Han K. Transposable elements: no more 'Junk DNA'. Genomics & informatics. 2012 Dec;10(4):226.

de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 2011 Dec 1;7(12):e1002384.

Kuntz SG, Eisen MB. *Drosophila* embryogenesis scales uniformly across temperature in developmentally diverse species. PLoS Genet. 2014 Apr 24;10(4):e1004293.

Lachaise D, Silvain JF. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster - D. simulans* palaeogeographic riddle. In *Drosophila melanogaster*, *Drosophila simulans*: so similar, so different 2004 (pp. 17-39). Springer, Dordrecht.

Lachaise D, Tsacas L. Les Drosophilidae des savanes préforestières de la région tropicale de Lamto (Côte-d'Ivoire). II. Le peuplement des fruits de *Pandanus candelabrum* (Pandanacées). Ann. Univ. Abidjan Ser. E Ecol. 1974 7, 153-92.

Lang M, Murat S, Clark AG, Gouppil G, Blais C, Matzkin LM, Guittard É, Yoshiyama-Yanagawa T, Kataoka H, Niwa R, Lafont R. Mutations in the neverland gene turned *Drosophila pachea* into an obligate specialist species. Science. 2012 Sep 28;337(6102):1658-61.

Lynch M, Walsh B. The origins of genome architecture. Sunderland, MA: Sinauer Associates; 2007.

Linz J, Baschwitz A, Strutz A, Dweck HK, Sachse S, Hansson BS, Stensmyr MC. Host plant-driven sensory specialization in *Drosophila erecta*. Proceedings of the Royal Society B: Biological Sciences. 2013 Jun 7;280(1760):20130626.

Ludwig A, Vidal NM, Loreto EL, Sepel LM. *Drosophila incompta* development without flowers. *Drosophila* Information Service. 2002;85:40-1.

Machado S, Gottschalk MS, Robe LJ. Historical patterns of niche dynamics in Neotropical species of the *Drosophila* subgenus (Drosophilidae, Diptera). Evolutionary ecology. 2016 Feb 1;30(1):47-67.

Malogolowkin C. Sobre a genitália dos drosofilídeos. IV. A genitália masculina no subgênero *Drosophila* (Diptera, Drosophilidae). Rev Bras Biol. 1953;13:245-64.

Manfrin MH, Sene FM. Cactophilic *Drosophila* in South America: a model for evolutionary studies. Genetica. 2006 Jan 1;126(1-2):57-75.

Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE, Stensmyr MC. Wild African *Drosophila melanogaster* are seasonal specialists on marula fruit. Current Biology. 2018 Dec 17;28(24):3960-8.

Markow TA. The natural history of model organisms: the secret lives of *Drosophila* flies. Elife. 2015 Jun 4;4:e06793.

Markow TA, O'Grady P. Reproductive ecology of *Drosophila*. Functional Ecology. 2008 Oct 1:747-59.

Matzkin LM, Watts TD, Bitler BG, Machado CA, Markow TA. Functional genomics of cactus host shifts in *Drosophila mojavensis*. Molecular Ecology. 2006 Dec;15(14):4635-43.

Matzkin LM. Ecological genomics of host shifts in *Drosophila mojavensis*. In Ecological genomics 2014 (pp. 233-247). Springer, Dordrecht.

McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proceedings of the National Academy of Sciences. 2007 Mar 20;104(12):4996-5001.

McBride CS, Arguello JR. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics. 2007 Nov 1;177(3):1395-416.

McClintock, B. Mutable loci in maize. Carnegie Institution of Washington Yearbook. 1948 47, 155– 169.

McClintock, B. Controlling elements and the gene. Cold Spring Harbor Symposia on Quantitative Biology. 1956 21, 197– 216

Missbach C, Dweck HK, Vogel H, Vilcinskas A, Stensmyr MC, Hansson BS, Grosse-Wilde E. Evolution of insect olfactory receptors. Elife. 2014 Mar 26;3:e02115.

Moreira HD. História evolutiva de espécies do grupo *flavopilosa* de *Drosophila* (DIPTERA, DROSOPHILIDAE): uma abordagem filogenética, filogeográfica e taxonômica. (Dissertação de Mestrado, UFRGS) 2020.

Napp M, Brncic D. Eletrophoretic variability in two closely related Brazilian species of the *flavopilosa* species group of *Drosophila*. Braz J Genet. 1978;1:1-0.

O'Grady PM, Lapoint RT, Bonacum J, Lasola J, Owen E, Wu Y, DeSalle R. Phylogenetic and ecological relationships of the Hawaiian *Drosophila* inferred by

mitochondrial DNA analysis. Molecular Phylogenetics and Evolution. 2011 Feb 1;58(2):244-56.

O'Grady PM, DeSalle R. Phylogeny of the genus *Drosophila*. Genetics. 2018 May 1;209(1):1-25.

Ohno S. So much 'junk' DNA in our genome. In Evolution of Genetic Systems, Brookhaven Symp. Biol. 1972 (pp. 366-370).

Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. Proceedings of the National Academy of Sciences. 2017 May 2;114(18):4721-6.

Pereira MR, Vilela CR, Sene FM. Notes on breeding and feeding sites of some species of the repleta group of the genus *Drosophila* (Diptera, Drosophilidae). Ciência e Cultura. 1983;35(9):1313-9.

Pipkin SB, Rodriguez RL, Leon J. Plant host specificity among flower-feeding neotropical *Drosophila* (Diptera: Drosophilidae). The American Naturalist. 1966 Mar 1;100(911):135-56.

Powell JR. Progress and prospects in evolutionary biology: the *Drosophila* model. Oxford University Press; 1997 Sep 4.

Prud'Homme B, Gompel N. Genomic hourglass. Nature. 2010 Dec;468(7325):768-9.

Prud'Homme B, Gompel N. Fruity aphrodisiacs. Nature. 2011 Oct;478(7368):190-1.

Quadrana L, Silveira AB, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. The *Arabidopsis thaliana* mobilome and its impact at the species level. elife. 2016 Jun 3;5:e15716.

Rebeiz M, Williams TM. Using *Drosophila* pigmentation traits to study the mechanisms of cis-regulatory evolution. Current opinion in insect science. 2017 Feb 1;19:1-7.

Rebollo R, Farivar S, Mager DL. C-GATE-catalogue of genes affected by transposable elements. Mobile Dna. 2012 Dec 1;3(1):9.

Remsen J, O'Grady P. Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Molecular Phylogenetics and Evolution. 2002 Aug 1;24(2):249-64.

Rio B. Evolution d'une spécialisation saisonniére chez *Drosophila erecta* (Dipt., Drosophilidae). ANN. SOC. ENTOMOL. FR.; FRA; DA. 1983; VOL. 19; NO 2; PP. 235-248; ABS. ENG; BIBL. 30

Robe LJ, Valente VL, Budnik M, Loreto ÉL. Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: a nuclear versus mitochondrial gene approach. Molecular Phylogenetics and Evolution. 2005 Sep 1;36(3):623-40.

Robe LJ, Loreto EL, Valente VL. Radiation of the *,,Drosophila“* subgenus (Drosophilidae, Diptera) in the Neotropics. Journal of Zoological Systematics and Evolutionary Research. 2010 Nov;48(4):310-21.

Robe LJ, De Ré FC, Ludwig A, Loreto EL. The *Drosophila flavopilosa* species group (Diptera, Drosophilidae) An array of exciting questions. Fly. 2013 Apr 21;7(2):59-69.

Robe LJ, Machado S, Bolzan AR, dos Santos JP, Valer FB, dos Santos AP, Blauth ML, Gottschalk MS. Comparative ecological niche modeling and evolutionary ecology of Neotropical mycophagous Drosophilidae (Diptera) species. Studies on Neotropical Fauna and Environment. 2014 May 4;49(2):79-94.

Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences. 2003 Nov 25;100(suppl 2):14537-42.

Robillard É, Le Rouzic A, Zhang Z, Capy P, Hua-Van A. Experimental evolution reveals hyperparasitic interactions among transposable elements. Proceedings of the National Academy of Sciences. 2016 Dec 20;113(51):14763-8.

Sanmiguel P, Bennetzen JL. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Annals of Botany. 1998 Dec 1;82:37-44.

Santos RD, Vilela CR. Breeding sites of Neotropical Drosophilidae (Diptera): IV. Living and fallen flowers of *Sessea brasiliensis* and *Cestrum spp*.(Solanaceae). Revista Brasileira de Entomologia. 2005 Dec;49(4):544-51.

Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends in ecology & evolution. 2010 Sep 1;25(9):537-46.

Schmitz HJ, Valente VL. The flower flies and the unknown diversity of Drosophilidae (Diptera): a biodiversity inventory in the Brazilian fauna. Papéis Avulsos de Zoologia. 201859.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P. The B73 maize genome: complexity, diversity, and dynamics. science. 2009 Nov 20;326(5956):1112-5.

Schrader L, Schmitz J. The impact of transposable elements in adaptive evolution. Molecular ecology. 2019 Mar;28(6):1537-49.

Sepel LM, Golombieski RM, Napp M, Loreto EL. Seasonal fluctuations of *D. cestri* and *D. incompta*, two species of the flavopilosa group. *Drosophila* Information Service. 2000;83:122-6.

Siefert JL. Defining the mobilome. In Horizontal gene transfer 2009 (pp. 13-27). Humana Press.

Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Fave MJ. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proceedings of the National Academy of Sciences. 2011 Apr 5;108(14):5673-8.

Soares EL, Vignoli-Silva M, Mentz LA. O gênero *Cestrum* (Solanaceae) no Rio Grande do Sul, Brasil. Pesquisas. Botânica. 2007;58:263-82.

Sun C, Feschotte C, Wu Z, Mueller RL. DNA transposons have colonized the genome of the giant virus *Pandoravirus salinus*. BMC biology. 2015 Dec;13(1):1-2.

Takahashi A. Pigmentation and behavior: potential association through pleiotropic genes in *Drosophila*. Genes & genetic systems. 2013;88(3):165-74.

Throckmorton LH. The phylogeny, ecology and geography of *Drosophila*. Invertebrates of genetic interest. 1975:421-69.

Tidon R, Sene FM. A trap that retains and keeps *Drosophila* alive. *Drosophila* Information Service. 1988;672:89.

TIDON R, SCHMITZ HJ, MARTINS MB, GOTTSCHALK MS. Catálogo Taxonômico da Fauna do Brasil: Drosophilidae 2015 (Catálogo Eletrônico).

Ugur B, Chen K, Bellen HJ. *Drosophila* tools and assays for the study of human diseases. Disease models & mechanisms. 2016 Mar 1;9(3):235-44.

Vilela CR. Occurrence of the *Drosophila flavopilosa* species group (Diptera, Drosophilidae) in the state of SŃo Paulo (Brazil) with description of one new species. Revista Brasileira de Zoologia. 1983;2(2):63-9.

Volff JN. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays. 2006 Sep;28(9):913-22.

Vorechovsky I. Transposable elements in disease-associated cryptic exons. Human genetics. 2010 Feb 1;127(2):135-54.

Wangler MF, Yamamoto S, Bellen HJ. Fruit flies in biomedical research. Genetics. 2015 Mar 1;199(3):639-53.

Wanner KW, Anderson AR, Trowell SC, Theilmann DA, Robertson HM, Newcomb RD. Female-biased expression of odourant receptor genes in the adult antennae of the silkworm, *Bombyx mori*. Insect molecular biology. 2007 Feb;16(1):107-19.

Wanner KW, Robertson HM. The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. Insect molecular biology. 2008 Dec;17(6):621-9.

Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. Annual Review of Genetics. 2020 Sep 21;54.

Wheeler MR, Takada H, Brncic D. The flavopilosa species group of *Drosophila*. Univ Texas Publ Genetics. 1962;6205:395-414.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E. A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics. 2007 Dec;8(12):973-82.

Wittkopp PJ, Stewart EE, Arnold LL, Neidert AH, Haerum BK, Thompson EM, Akhras S, Smith-Winberry G, Shefner L. Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. Science. 2009 Oct 23;326(5952):540-4.

Yamashiro H, Siomi MC. PIWI-interacting RNA in *Drosophila*: biogenesis, transposon regulation, and beyond. Chemical reviews. 2017 Dec 27;118(8):4404-21.

Zhou CX, Zhang YL, Xiao L, Zheng M, Leung KM, Chan MY, Lo PS, Tsang LL, Wong HY, Ho LS, Chung YW. An epididymis-specific β-defensin is important for the initiation of sperm maturation. Nature cell biology. 2004 May;6(5):458-64.