UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ESTUDO DA HISTÓRIA EVOLUTIVA DE *Drosophila lutzii* (DIPTERA: DROSOPHILIDAE), UMA ESPÉCIE ANTOFÍLICA, UTILIZANDO SEQUENCIAMENTO DE NOVA GERAÇÃO

Carolina Prediger

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Mestre em Genética e Biologia Molecular

Orientadora: Dra. Maríndia Deprá
Coorientadora: Dra. Juliana Cordeiro

Porto Alegre, Fevereiro 2020

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

TRACING THE EVOLUTIONARY HISTORY OF THE ANTOPHILIC SPECIES *Drosophila lutzii* (DIPTERA: DROSOPHILIDAE) USING NEXT GENERATION SEQUENCING APPROACH

Carolina Prediger

Dissertation submitted to the Programa de Pós Graduação em Genética e Biologia Molecular at UFRGS as a partial requirement to obtain a Master's Degree in Biological Sciences.

Advisor: Maríndia Deprá, Ph.D.
Co-advisor: Juliana Cordeiro, Ph.D.

Porto Alegre, February 2020

CIP - Catalogação na Publicação

Prediger, Carolina
ESTUDO DA HISTÓRIA EVOLUTIVA DE Drosophila lutzii (DIPTERA: DROSOPHILIDAE), UMA ESPÉCIE ANTOFÍLICA, UTILIZANDO SEQUENCIAMENTO DE NOVA GERAÇÃO / Carolina Prediger. -- 2020.
72 f.
Orientadora: Maríndia Deprá.

Coorientadora: Juliana Cordeiro.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Programa de Pós-Graduação em Genética e Biologia Molecular, Porto Alegre, BR-RS, 2020.

1. Sequenciamento de genoma completo. 2. Drosofilídeos de flores. 3. P450. 4. Cyp. 5. filogenética. I. Deprá, Maríndia, orient. II. Cordeiro, Juliana, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

The present dissertation was carried out on Laboratório de *Drosophila* at Universidade Federal do Rio Grande do Sul, and funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul).

Agradecimentos

Primeiramente gostaria de agradecer as pessoas que contribuíram diretamente com os capítulos apresentados aqui, à professora Dra. Vera Lucia da Silva Valente Gaiesky, por todo o suporte. À M.Sc. Tuane Letícia Carvalho pelo DNA de indivíduos corretamente identificado. Ao Dr. Zacchary Cohen e ao Center for High Throughput Computing Team da University of Wisconsin-Madison. Dr. Charles Robin da University of Melbourne pelo envio dos dados utilizados em sua pesquisa. Ao CNPq a CAPES e a FAPERGS pelo financiamento.

Também Gostaria de agradecer aos colegas de laboratório, em especial aqueles que me acompanharam desde o início do mestrado, dando aquele incentivo para arrancar o tapete da sala de bolsistas. Aos colegas das disciplinas, em especial a turma das disciplinas de biodiversidade.Ao time do professor Dr. Elgion Lucio Da Silva Loreto e à Dra. Evelise Bach pelas dicas de como montar o genoma. A equipe do Centro Nacional de Supercomputação (CESUP) pela ajuda com os scripts no comecinho de tudo. Aos "*sinners*" que me ouviram falando da vontade de jogar o computador pela janela e os choramingos quando faltava luz no vale. Ao 601F nossa 'família tradicional brasileira' é a melhor. A toda a minha bolha, por me dar esperanças.

Finalmente gostaria de agradecer à minha família, pelo apoio e compreensão, sei que minhas escolhas são um pouco "diferentes". Às minhas orientadoras, por terem me dado a oportunidade de trabalhar nesse projeto, tento o privilégio de poder fazer o que eu gosto, e principalmente por acreditarem que eu e o computador nos daríamos bem, mesmo quando eu usava o *windows*. Um muito obrigada especial a Maríndia, que apoiou minha escolha e me enviou uma ótima carta de recomendação.

Concluindo, foi um prazer trabalhar tendo o suporte de vocês. Muito obrigada!

Table of content

List of Abbreviation

12S – 12SrRNA

16S – 16SrRNA

28S – 28SrRNA

AA – Amino acids

Adh – Alcohol dehydrogenase gene

Amd – α methyl dopa hypersensitive protein gene

BI – Bayesian Inference of phylogeny

BS – Bootstrap

bp – Base pair

CDS - Coding region of the genes

COI – Cytochrome c oxidase subunit I mitochondrial gene

COII – Cytochrome c oxidase subunit II mitochondrial gene

Cyps – Cytochrome P450 monooxygenases

CytB – Cytochrome b mitochondrial gene

Ddc – Dopa decarboxylase gene

DNA – Deoxyribonucleic acid

FBD – Flower-Breeding Drosophilids

Gpdh – Glycerol-3-phosphate dehydrogenase 1 gene

Gr – Gustatory Receptors

GSTs – Glutathione-S-transferases

IRs – Ionotropic receptors

Mdr65 – Multidrug resistance 65

ML – Maximum Likelihood Tree

NOMP C – No mechanoreceptor potential C

NVD – Neverland gene

Od – Odorant binding

Or – Odorant receptors

P450 – Cytochrome P450 monooxygenase

Pkc98E – Proteinkinase C98E

PP – Posterior Probability

TMC – Transmembrane channel-like gene

WGS – Whole genome sequencing

Resumo

Apesar dos drosofilídeos serem muito conhecidos por utilizarem frutos em decomposição como fonte de alimentação e oviposição, muitos recursos podem ser utilizados por eles como flores, fungos, folhas, entre outros. Essas espécies com nichos peculiares foram pouco estudadas por muito tempo e atualmente há um grande esforço para aumentar o conhecimento de sua ecologia, genética e evolução. Com relação ao grupo de espécies da *Drosophila lutzii,* um grupo de espécies com ecologia restrita a flores, que pode inclusive utilizar flores tóxicas, não há muito conhecimento acumulado. O mais recente trabalho filogenético que amostrou o grupo o posicionou como um grupo de espécies pertencente ao subgênero *Drosophila*, contudo a comunidade científica apresenta ressalvas quanto a sua relação evolutiva e a nomenclatura sugerida. Apesar do grupo de espécies utilizarem flores tóxicas, os mecanismos fisiológicos por trás do uso de nicho ainda não foram estudados para o grupo de espécies. Aqui, foi realizado o primeiro sequenciamento do genoma de *D. lutzii*. A partir disso, as relações evolutivas da espécie foram inferidas por meio de onze marcados moleculares. Adicionalmente, foi caracterizada a família gênica citocromo P450 de *D. lutzii*, família muito associada à detoxificação de compostos. Os resultados mostram que *D. lutzii* se posiciona dentro do subgênero *Drosophila* e apresenta relação estreita com os grupos de espécies *D. pallidipennis* e *D. tripunctata*. Com relação à família gênica citocromo P450 verificou-se a presença de 80 genes sobre seleção purificadora, um pseudogene e oito potenciais pseudogenes. Verificou-se muitos eventos de duplicação de genes associados à detoxificação, contudo outras analises são necessárias para verificar os efeitos dessas duplicações.

Palavras chave: Sequenciamento de genoma completo, Drosofilídeos de flores, P450, Cyp, filogenética.

Abstract

The use of decaying fruits as feeding and breeding by drosophilids is widely known, even that they can use a many other resources, such as, fungi, flowers, leaves, and so on. The species with peculiar niches were neglected for a while, and nowadays there is a great effort to increase the knowledge of their ecology, genetics and evolution. Regarding *D. lutzii* species group, a flower-breeding *Drosophila* that is capable of use toxic flowers as feeding and breeding site very little is known. The last phylogenetic study that sampled the group placed it within the *Drosophila* subgenus; however the scientific community has reservations regarding the proposed evolutionary relationships and nomenclature. In spite of the group of species use toxic flowers, the physiological mechanisms correlated to niche use were not study for this species group. Here, the first sequencing of the *D. lutzii* genome was carried out. From that, the evolutionary relationships were reconstructed by phylogenetic approach using eleven molecular markers. Also, the gene family cytochrome P450 (P450), which is correlated to detoxification, was characterized for the *Drosophila lutzii*. Regarding the phylogenetic, *D. lutzii* species group fall within the *Drosophila* subgenus, and are close related to the *D. pallidipennis* and *D. tripunctata* species groups. Considering de detoxification genes, 80 P450 genes were found within the *D. lutzii* genome, all of them under purifying selection. One pseudo-gene and 8 potential pseudo-genes were seen. Duplication process was noted to many P450 genes related to detoxification, however further analysis are needed to verify the effect of this duplications.

Key words: Whole genome sequencing, Flower-breeding drosophilids, P450, Cyp, phylogenetics

Chapter 1 – Why is the *Drosophila lutzii* species group an interesting model for evolutionary studies?

1. Introduction

1.1. Flower-Breeding Drosophilids (FBD) and the *Drosophila lutzii* group of species

Drosophila melanogaster has been intensively studied in medicine, genetics, development and evolutionary fields. This is partially due to the use of inexpensive and convenient banana-bait traps to capture this fly and also inexpensive culture medium to keep it at the lab. Furthermore, lineage maintenance is possible for many other Drosophilidae species. These flies also have a short lifetime and well-defined development stages, therefore, the species of this family became an excellent model organism for several studies (Yamaguchi and Yoshida, 2018).

Regarding diversity, the Drosophilidae family encompasses approximately 4,200 species (Bächli, 2020). This high diversity is also seen when feeding and breeding sites are taken into account. There are records of drosophilids breeding and feeding at a wide range of resources such as fruit, sap, mushrooms, cacti, leaves, flowers, living crabs, predating Hemiptera and mosquitoes' larvae, and there are even reports of species being attracted by baits of decaying animal (Carson 1971; Ashburner 1981; Lachaise and Tsacas 1983; Ferraz 2014). However, until recently, most of the diversity studies focus on *Drosophila* species that feed on yeast found on decaying fruits, once these species are easily collected by banana-bait traps. This bias was observed by many research groups, including scientists specialized in Neotropical diversity, and a large effort was taken in the last decade to understand the ecology, biology, and evolution of species of different guilds. Significant advances have been made regarding mycophagous (Gottschalk *et al.,* 2009; Valer *et al*., 2016; Machado *et al*. 2017; Santa-Brígida *et al*., 2019) and anthophilous drosophilids (Brncic, 1966; 1983; Schmitz and Hofmann, 2005; Schmitz *et al.*, 2007; Robe *et al.*, 2013; De Ré *et al.*, 2014; 2017; Grimaldi *et al*., 2016; Fonseca *et al*., 2019; Schmitz and Valente, 2019; Cordeiro *et al*., 2020).

Regarding the Neotropical anthophilous drosophilids, a remarkable study of Schmitz and Valente (2019) highlights the unknown diversity of this guild, once approximately 50% of its collected species were yet not described. Since the primordial work of Brncic (1983), who has performed the first review on the anthophilous drosophilids worldwide, several new species in the Neotropic region were described (see Cordeiro *et al.*, 2020 for the list of these species). Anthophily in Drosophilidae species is shown in different forms: some species use flowers as a rendezvous site to mate with their partner; some species use only as a feeding site, and; some species use it as rendezvous, feeding and breeding site, where flowers are a necessary resource to develop the larvae. The species that show this latter developmental strategy is also referred as flower-breeding drosophilid (FBD) (Brncic, 1983). In this way, several Neotropical lineages of the *Drosophila* genus show restricted ecology to flowers (Markow and O'Grady, 2008), that is, these species require the use of flowers in some step of their life cycle. Considering these Neotropical lineages of the Drosophilidae family with restricted ecology to flowers, three of them are of particular interest: *D. bromeliae* species group, *D. flavopilosa* species group and *D. lutzii* species group [sensu Yassin, 2013] once that all species within these species group are FBD (Brncic, 1983; Robe *et al*. 2010a; 2013; Yassin, 2013; De Ré *et al*., 2017; Schmitz and Valente, 2019).

Recent phylogenetic studies showed that *D. flavopilosa* and *D. bromeliae* species group are independent lineages belonging to the *Siphlodora* subgenus (sensu Yassin, 2013 – the previous *virilis-repleta* radiation, according to Throckmorton 1975), as well as the last Drosophilidae phylogenetic hypothesis places *D. lutzii* species group [sensu Yassin, 2013] into the *Drosophila* subgenus within the *Drosophila* genus (the previous *immigrans-tripunctata* radiation, according to Throckmorton 1975) (Robe *et al.,* 2010a; 2013; Yassin, 2013; De Ré *et al.,* 2017). Before Yassin's work (Yassin, 2013), *D. lutzii* species group [sensu Yassin, 2013] were placed as a separate subgenus within *Drosophila* genus, the *Drosophila (Phloridosa)* subgenus, and the classical evolutionary studies of Throckmorton (1975) and Grimaldi (1990) were incongruent in their conclusions about this group of species. On one hand, Throckmorton's study, based on ecological traits and geographical distribution, suggests that this lineage is related

to the "*virilis-repleta* radiation" [subgenus *Siphlodora*, according to Yassin (2013)] (Throckmorton 1975). On the other hand, Grimaldi's study, based on external morphology, consider that this lineage diverged earlier than the divergence of the *Drosophila* and *Sophophora* subgenera, placing these species in the *Drosophila* (*Phloridosa*) subgenus (Grimaldi 1990). Later on, the Yassin's review on the Drosophilidae family, using morphological and molecular markers, placed the *D. lutzii* lineage within the *Drosophila* (*Drosophila*) subgenus, categorizing this lineage as another group within *Drosophila* subgenus instead of a distinct subgenus (Yassin, 2013) (Figure 1).

Figure 1. Dendrograms representing the phylogenetic relationships reported for *D. lutzii* species group [sensu Yassin, 2013]. In red it is shown the clade where the *D. lutzii* species group [sensu Yassin, 2013] had been reported by the authors.

The *D. lutzii* species group [sensu Yassin, 2013] encompasses eight species (*D. alei* Brncic, 1962, *D. alfari* Sturtevant, 1921, *D. cuzcoica* Duda, 1927, *D. denieri* Blanchard, 1938, *D. lutzii* Sturtevant, 1916 (Figure 2), *D. merzi* Vilela and Bachli, 2002, *D. monsterae* Vilela and Prieto, 2018 and *D. tristani* Sturtevant, 1921). This group of species oviposits in flowers belonging to families Araceae, Boraginaceae, Convolvulaceae, Cucurbitaceae, Malvaceae, Passifloraceae and Solanaceae to complete their life cycle (Schmitz and Valente, 2019; Cordeiro *et al.*, 2020). The species *D. lutzii* is frequently found competing with *D. denieri* and *D. bromelioides* in *Brugmansia suaveolens* and *Ipomoea* sp. flowers. *Drosophila lutzii* seems to be more abundant than its competitors in the *Ipomoea* species, although the effects of intra and interspecific competition for host's flowers still need further investigation.

Figure 2. Male (upper) and female (lower) *D. lutzii* specimens. Source: João Henrique Figueredo de Oliveira`s personal archive.

Regarding the geographical range (Figure 3), the species of the *D. lutzii* group [sensu Yassin, 2013] occurs in Brazil, Mexico, Puerto Rico, Costa Rica, Colombia, Ecuador, Argentina, Uruguay, Peru, Chile, United States of America, Hawaiian Islands and Caribbean Islands (Bächli, 2020). Within the species of the group, the largest geographical range comes from *D. lutzii* which has been reported from Florida to the south of Brazil, including many Caribbean Islands and also being reported in the Hawaiian archipelago (Bächli, 2020).

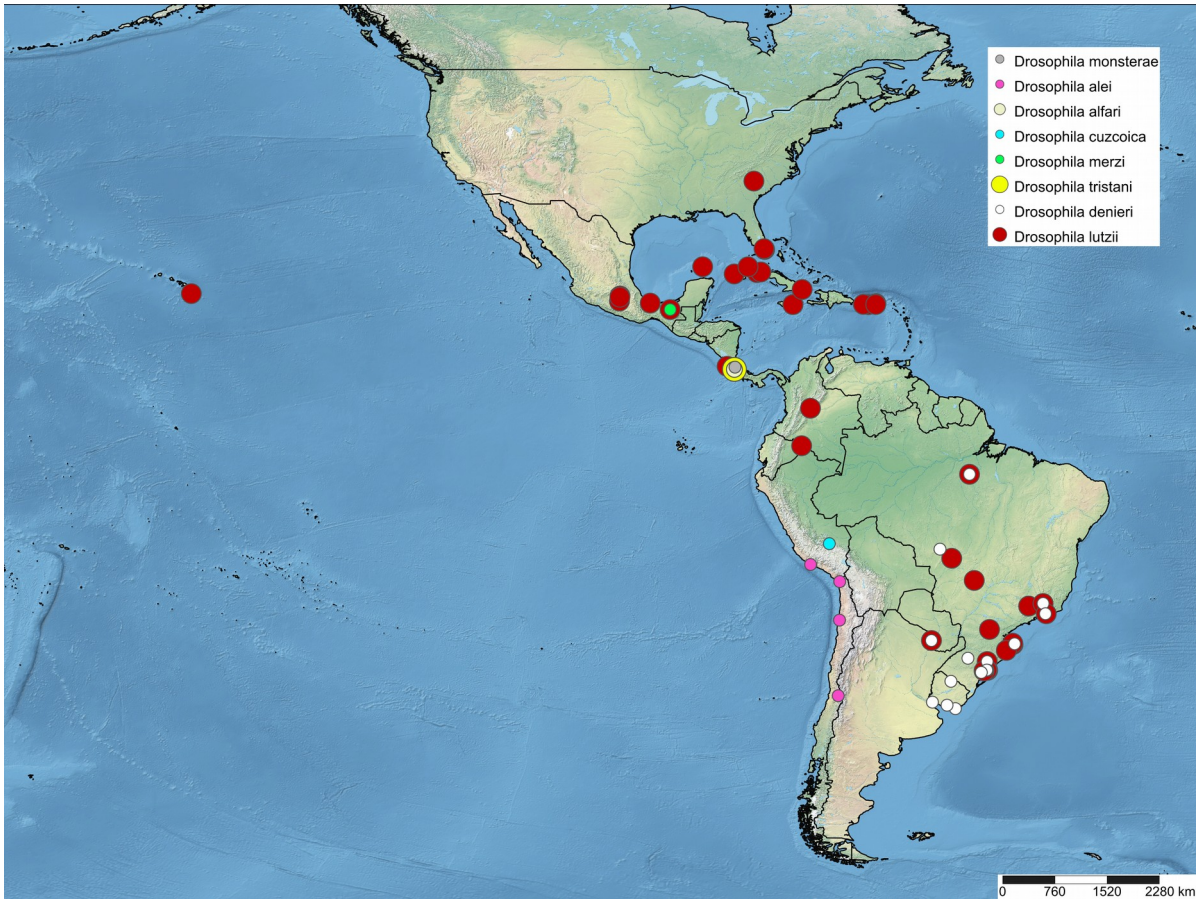


Figure 1. Reported occurrence of species belonging to the *D. lutzii* group [sensu Yassin, 2013]. Source: TaxoDros coordinates database (BÄCHLI, 2020) and Vilela and Prieto (2018).

1.2. Ecological specialization

Understanding why species feed and breed in a determined resource and not the other is a triggering question, and a precise answer still needs to be acquired. A recent study of the interaction between FBD and their host plants shows highly modular, non nested network, formed by highly specialized drosophilids, i.e. one clusters of drosophilids interacts with one cluster of plants and does not tend to interact with other cluster of plants, furthermore interaction between specialist-to-specialist are relatively frequented. The authors propose that physiological barriers and spatio-temporal co-occurrence of host and FBD possibly are responsible for the interaction network pattern found (Cordeiro *et al.*, 2020).

Considering the genetic and physiological mechanism behind the resource

choice, Markow (2019) and Etges (2019) reviewed and systematically organized the requirements that make the species able to find, accept and ultimately use the resource and also listed the genes that are related to ecological specialization. Both authors point out some gene families associated with feeding and breeding choices, such as the chemosensory related genes [Gustatory Receptors (Gr), Odorant Receptors(Or), Odorant binding (Ob), Ionotropic receptors (IRs), No mechanoreceptor potential C (NOMP C), transmembrane channel-like gene (TMC)], and the detoxification related genes [Glutathione-S-tranferases (GSTs) and cytochrome P450 monooxygenase (P450 or Cyp)]. Also, they point out that genes related to metabolic pathways may also be related to resource choice (Etges, 2019; Markow, 2019).

Regarding chemosensory related gene families, some studies that compared different levels of specialization showed interesting findings. For example, *Drosophila sechellia* is an endemic species to the Seychelles Archipelago and is specialized in *Morinda citrifolia*'s fruits which are highly toxic to other closed related drosophilids, and *D. simulans*, a closely related species is cosmopolitan and generalist. In one hand, McBride and Arguello (2007) reported Or and Gr's gene lost in *D. sechellia* comparing with *D. simulans* and suggested that it is a consequence of host specialization. On the other hand, Gardiner *et al.* (2008) conclude that the gene lost were due to endemism once the statistical signal to specialist selection was lost by adding other specialized species in the analysis (*D. erecta*, a 'seasonal specialist' with *Pandanus*, and *D. mojavensis* cactophilic species). Another example regarding chemoreceptors genes comes from a study of the drosophilid species *Scaptomyzaflava*, a leaf-mining specialist on Brassicaceae plant family. In this case, the genomic evolution events included gene loss and pseudogenization of Or genes related to the perception of compounds typically produced by yeasts, and gene duplication events of at least one Or gene related to perception of green leaf volatiles (Goldman-Huertas *et al.,* 2015). Regarding FBD, De Ré (2016) evaluated the Or and Gr families in *D. incompta*, a species belonging to the *D. flavopilosa* species group and strongly associated with flowers of *Cestrum* genus (Solanaceae). The results pointed out that gene loss was a very common event for Or and Gr gene families for this species. It is concluded that the gene loss

events were due to the species' specialization.

Regarding detoxification, one example came from *Drosophila* mycophagous specialist. Many of these species are tolerant to α-amanitin produced by some mushrooms, a substance with affinity to RNA polymerase II. Some lineages of *D. melanogaster* are tolerant to α-amanitin, and it seems that different pathways led to the resistance. The genes found to be related to this resistance are: Multidrug resistance 65 (*Mdr65*), Protein kinase *C98E* (*Pkc98E*), tequila (*teq*), megalin (*mgl*), widerborst (*wdb*) (these last three genes are related to the Target of Rapamycin pathway, repressor of autophagy) and three P450 genes *Cyp6a2*, *Cyp12d1-d*, and *Cyp12d1-p* (these P450 genes are reported to be responsible for detoxification of a large spectrum of non-related chemical substances, Chialvo and Werner, 2018).

Another interesting example comes from *Drosophila mettleri*, a cactophic species. This species is able to oviposit in soil with much higher cacti's alkaloids concentration than fresh tissue. A transcriptome study that compared different food sources used by this species showed differential expression related to the gene families: P450 (genes *Cyp6t*, *Cyp4e1*, *Cyp4e2*, *Cyp4e3*, *Cyp307a1*, *Cyp307a2*, *Cyp6d5*, *Cyp6d5*, *Cyp313a4*, *Cyp304a1, Cyp4p3, Cyp4p1, Cyp4p2, Cyp28a5, Cyp312a1*, Disembodied, Phantom), carboxylesterases, GST and UGT-glycosyltransferases. Interestingly, the authors also reported changes on sensory perception of the flies (Hoang *et al*., 2015).

Another interesting case of apparently single gene mutation related to host specialization comes from *D. panchea*, a cactofilic drosophilid with restricted use to *Lophocereus schottii*. According to Lang *et al.* (2012) mutations on the Neverland gene (*NVD*) made the biochemical reaction of converting cholesterol to 7-dehydrocholesterol (7DHC) impossible. The 7DHC is an ecdysone precursor. Instead, the mutated NVD protein now uses lathosterol as the precursor to produce 7DHC. The lathosterol is only synthesized by the cactus host, and therefore the drosophilid species became unable to use other cactus host species.

Considering these, study these genes families in species specialized to a particular resource became interesting to understand the mechanisms that are related to ecological specialization. *Drosophila lutzii* have potential particularities that make it an interesting model to study host specialization. One particularity is

that some of its host plant shows a high level of toxicity, especially *B. suaveolens*. This plant species has tropane alkaloids on its leaves, and scopolamine is the most abundant tropane alkaloid, but also hyoscyamine and atropine are found. These chemicals are likely to be also found in the plant flowers (Arab and Trigo, 2011). The scopolamine appears to reduce herbivory in Lepidoptera (Arab *et al.*, 2012). In this way, the analysis of the P450 gene family in *D. lutzii* could help to understand the genomic evolution required to resource specialization.

2. Objectives

2.1. Main Objective

The main objective of this study is assembling the first genome of *D. lutzii* to generate insights on the evolutionary history of *D. lutzii* species group [sensu Yassin, 2013].

2.2. Specific objectives

1) Sequence and assemble the entire genome of *D. lutzii* (Chapter 02 and 03);

2) Analyze the evolutionary relationships of *D. lutzii* [sensu Yassin, 2013] species group within the Drosophilidae family (Chapter 02);

3) Characterize the P450 gene family with an emphasis on the evolution of detoxification related genes (Chapter 03).

Chapter 2 – What is the phylogenetic position of *Drosophila lutzii* species group within the Drosophilidae's phylogeny?

Carolina Prediger[1], Juliana Cordeiro[2], Maríndia Deprá[1]

1 – Programa de Pós Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Brazil.

2 – Departamento de Zoologia, Ecologia e Genética, Instituto de Biologia, Universidade Federal de Pelotas, Brazil.

Manuscript in preparation to be submitted to the Journal of Genetics and Molecular Biology

Abstract

The comprehension of the evolutionary relationships among species is a requirement to many study fields. Therefore, clarifying the phylogenetic relationships of species group is important to understanding the organism's biology. *Drosophila lutzii* species group is a Neotropical group of flower-breeding flies previously known as the unique species belonging to subgenus *Phloridosa* within genus *Drosophila*. This group is now placed in the *immigrans-tripunctata* radiation within *Drosophila* subgenus. However, the phylogenetic relationship of this group within Drosophilidae remains controversial. This study aims to reconstruct the evolutionary relationship of *D. lutzii* group species based on genomic data. Genomic DNA of *D. lutzii* males was submitted to whole genome sequencing. Genome assembly was performed in SPAdes software and filter with Redundans pipeline. BLAST+ was used to identify 11 genes (*Ddc, Adh, Amd, Gpdh, 28S, COI, COII, CytB, 12S* and *16S*) used for phylogenetic reconstruction. These genes were recovered from 63 Drosophilidae species. The phylogenetic relationships were inferred using MrBayes and PhyML reconstructions. Our data shows that *D. lutzii* is a member of the subgenus *Drosophila* within the genus *Drosophila*, and the group of species is closely related to the *tripunctata* lineage, especially the *D. pallidipennis*

and *D. tripunctata* species group.

Key words: Flower-breeding Drosophila, anthophilous drosophilid, phylogenetics

1. Introduction

The knowledge of the evolutionary relationships of species is one of the required traits in several applied research such as ecology, genetics, conservation biology, biogeography, comparative physiology and epidemiology (Graham *et al.,* 2018; O'Grady and DeSalle, 2018). However, even for drosophilids, taxa that have been intensively studied for over a hundred years, with many phylogenetic studies reported in literature (Markow and O'Grady, 2006; Robe *et al*., 2010a; Robe *et al*., 2010b; Russo *et al*., 2013; Yassin, 2013; O'Grady and DeSalle, 2018) there are uncertainties regarding phylogenetic position for many groups and species. The reasons are: i) intrinsic characteristics of the species' evolution process such as rapidly speciation leading to phylogenetic incongruencies and difficult interpretation of the evolutionary relationships (the Darwinian shortfall; Diniz-Filho *et al.,* 2013); or ii) poorly knowledge for many taxa (the Linnean shortfall; Brown and Lomolino, 1998).

The flower-breeding drosophilids (FBD) are anthophilous drosophilid species usually neglected in the phylogenetic studies. Due to their ecological restricted requirements, since female seems to be dependent on flowers to oviposit, the long term maintenance of these species in laboratory conditions is laborious (Brncic, 1983; Ludwig *et al*., 2002; Cordeiro *et al*., 2020). Even though, efforts were done to include these species in the Neotropical phylogenetic studies of Drosophilidae (Pélalandakis and Solignac, 1993; Tatarenkov and Ayala, 2001; Remsen and O'Grady, 2002; Da Lage *et al.,* 2007; van der Linde and Houle, 2008; Robe *et al*., 2010a; 2013; Yassin, 2013; De Ré *et al*., 2017). The Neotropical *D. flavopilosa* and *D. bromeliae* species groups belong to the *Siphlodora* subgenus of the *Drosophila* genus (Yassin, 2013). These species groups are independent evolutionary lineages, showing that anthophyly arise several times in the Drosophilidae species (Markow and O'Grady, 2008). In spite of the phylogenetic studies scarcity including

the Neotropical *D. lutzii* species group [sensu Yassin, 2013], the species of this group seems to belong to the *Drosophila* subgenus within the *Drosophila* genus (Yassin, 2013). Previous studies provided incongruent phylogenetic data (Throchmorton, 1975; Grimaldi, 1990) and since then the eight species belonging to the *D. lutzii* species group were placed in the *Phloridosa* subgenus within *Drosophila* genus (Grimaldi 1990). However, the recent *D. lutzii* systematic reclassification was based on information provided by only one gene (*Amyrel*) and morphological data (Yassin, 2013), and the relationships of this group of species within the *Drosophila* subgenus are still unknown.

In this study, we analyzed the *D. lutzii* phylogenetic relationship among 63 other Drosophillidae species. Our results point out that the *D. lutziii* species group falls within the *Drosophila* subgenera. This group of species is closely related to the 'lineage' *tripunctata*, especially the *D. tripunctata* and *D. pallidipennis* species groups.

2. Material and Methods

2.1. Biological sample

Since we had no success maintaining an isoline of *D. lutzii* at the laboratory, the following strategy was performed to avoid heterozygosity increase in the *D. lutzii* genome assembly: i) only three flowers from the same branch of *Ipomea purpurea* (Convolvulaceae) were collected. These flowers were kept in sterile vial at 25ºC and high humidity condition until *Drosophila* adults' emergency. Males were identified through genitalia morphology, following appropriate literature, and DNA barcode. Then ii) the DNA was extracted from 10 males of *D. lutzii* using the NucleoSpin DNA Insect kit (Macherey Nagel, Düren, Germany). The de novo whole genome sequencing was performed at MacrogenInc..

2.2. de novo Whole Genome Sequencing strategy

The library construction was performed using Truseq DNA Nano 350bp and the sequence platform was Illumina Novaseq 150 bp, pair-end, with final coverage

approximately 200x. The quality of the generated reads were checked on fastQC (Andrews, 2010). The best k-mer were identified by kmergenie (Chikhi and Medvedev, 2014). The *D. lutzii* genome was assembled using *denovo* approach on SPAdes (Bankevich*et al*., 2012). To avoid heterozygosity, the Redundans pipeline were applied generating the scaffolds for downstream analysis (Pryszcz and Gabaldón, 2016). The assembly quality and posterior filtering were check in QUAST (Gurevich*et al*., 2013). All analyses were performed in the Center for High Throughput Computing at the University of Wiscosin-Madison (chtc.cs.wisc.edu/).

2.3. Molecular markers and phylogenetic analyses

To reconstruct the evolutionary relationships of species 11 molecular markers were chosen based on previous work (O'Grady and DeSalle, 2018): the nuclear genes Amyrel, Dopa Decarboxylase (*Ddc*), alcohol dehydrogenase (*Adh*), α methyl dopa-resistant (*Amd*), Glycerol-3-phosphate dehydrogenase 1 (*Gpdh*), 28SrRNA (*28S*), and the mitochondrial genes Cytochrome c oxidase subunit I (*COI*), Cytochrome c oxidase subunit II (*COII*), Cytochrome B (*CytB*), 12SrRNA (*12S*) and 16SrRNA (*16S*). To recover these genes from *D. lutzii* genome, *D. melanogaster* sequences were downloaded from FlyBase (Thurmond *et al*., 2019) and then used as local BLAST+ query through BLASTn searches. Genes that were not recovered using this strategy were then retrieved by BLASTn using a consensus sequence for several *Drosophila* species as query, generated in the UGENE software (Okonechnikov*et al*., 2012).

In order to sample all major clades of *Drosophila* genus, 59 species of *Drosophila* genus, two species of *Zaprionus* and two species of *Scaptodrosophila* (used as out group) had all the genes above mentioned retrieved from GenBank. For species with whole genome available a BLASTn strategy was used to retrieve the genes. However, each genome was analyzed individually in order to balance sequence length and accuracy, since sanger method is more accurate than WGS (Wang *et al*., 2012). For 28S marker, sequences from the available species were downloaded from the Eckbush lab webpage (http://blogs.rochester.edu/EickbushLab/) (Supplementary Table S1). For each

maker, the alignments were conducted with the method G-INS-i implemented in MAFFT (Katoh, 2002). The alignments were manually checked and the 5' and 3' end were cut. Nuclear genes had their introns removed. After checking the alignment, three matrices were built: nuclear data, mitochondrial data and nuclear+mitochondrial data (Supplementary data). PartitionFinder 2.1.1. (Lanfear*et al*., 2017) were used to find the best partition within the data and the evolutionary models for phylogenetic analysis using Bayesian Inference approach. Phylogenetic reconstruction were performed for each group of data (nuclear; mitochondrial; nuclear+mitochondrial) using Bayesian Inference (BI) through MrBayes 3.2.6 (Ronquist *et al*., 2012) implemented at CIPRES platform (Miller *et al*., 2010). For BI analysis 50,000,000 generations were set and burn-in of 25%. The nuclear+mitochondrial matrix was analyzed through Maximum Likelihood (ML) reconstruction. In this case, the evolutionary model for the concatenated alignment was chosen using KAKUSAM software (Tanabe, 2011). ML analyses were performed in PhyML (Guindon and Gascuel, 2003) through 500 bootstrap (BS). All phylogenetic reconstruction obtained were visualized and edited in iTOL v. 4 (Letunic and Bork, 2019).

3. Results

3.1. *denovo* whole genome sequencing and assembly quality

The FastQC report showed that 177,410,711 forward sequences and 175,914,276 reverse sequences were generated. Overall the parameters analyzed indicated high quality sequencing (e.g.: Phred quality score higher than Q30, Supplementary Figure S1), and the contigs quality was also improved using the Redundants pipeline as shown in Table 1 (for full report see Supplementary Table S2).

Table 1. Quality parameters reported by QUAST after genome assembly with Spades and implementation of Redundants' pipeline. All statistics are based on contigs of size ≥ 500 bp.

	Spades' Contigs	Redundants' scaffolds
Number of contigs	87,033	63,818
Largest contig	250,057	250,057
Total length	226,483,272	202,420,868
GC (%)	40.04	40.64
N50	6,571	8,400
N75	1,613	2,412
L50	7,112	5,491
L75	25,607	16,989
# N's per 100 kbp	0.00	0.00

3.2. Phylogenetics analysis

The final alignment matrices showed 5,669 basepairs (bp) for the nuclear data, 3,556 bp for the mitochondrial data and 9,216 bp for nuclear+mitochondrial data (Table 2). The phylogenetic reconstructions with all matrices nuclear+mitochondrial, and nuclear largely recovered the major *Drosophila* genus clades following Yassin (2013) (*Drosophila*, *Siphlodora*, *Dorsilopha* and *Sophophora* subgenera), and also recovered the relationship within their species groups (Figure 1, Figure 2 and Supplementary Figure S2). The mitochondrial data matrix fails to recover the relationship between the major clades and does not recover the monophyly of *Sophophora* (Supplementary Figure S3). The nuclear+mitochondrial data BI phylogenetic tree (Figure 1) showed higher support than the nuclear+mitochondrial data ML and will be used for comparison among clades. Major differences of the ML tree will be pointed out (Figure 2). Regarding the general topology of the *Drosophila* genus, *Sophophora* subgenus appears as the sister clade of the clade encompassing *Dorsilopha*, *Siphlodora*, *Drosophila* subgenera and the genus *Zaprionus* (PP=1, BS=0.99). The subgenera *Dorsilopha* is the next clade branching off (PP=1, BS=1). The subgenus *Siphlodora* appears as the sister clade *Zaprionus* genus and the *Drosophila* subgenus clade (Figure 1 and *Figure 2*) (PP=1, BS=0.5). This close relation between *Zaprionus* genus and *Drosophila* subgenus has been reported previously by van der Linde and Houle (2008), Robe *et al.* (2010a, 2010b) and Yassin (2013) and represent one of the

many genera of Drosophilidae that are placed by molecular phylogeny within the *Drosophila* genus.

Table 2. Size (bp) of the molecular markers used in the reconstruction of the phylogenetic relationships

COI	COII	CytB	12S	16S	28s	Adh	Amd	Ddc	Gpdh	Amyrel
765	684	1035	444	624	705	729	996	1104	699	1431
nuclear+mitochonrial data matrix										
1 - 765	766 - 1450	1451 - 2486	2487 - 2931	2932- 3556	3557 - 4262	4263 - 4992	4993 - 5989	5990 - 7094	7095 - 7794	7795 - 9226
mitochondrial data matrix					**nuclear nuclear matrix**					
1 - 765	766 - 1450	1451 - 2486	2487 - 2931	2932 - 3556	1 - 705	706 - 1435	1436 - 2432	2433 - 3537	3538 - 4237	4238 - 5669

Regarding the *Sophophora* subgenus, the close relationship between *D. willistoni* and *D. saltans* species groups was recovered by the nuclear+mitochondrial and nuclear matrices, also the relationships between the *D. melanogaster* and *D. obscura* species groups. Incongruences was seen when comparing this with the tree generated by the mitochondrial data which failed to recover the monophyly of the subgenera, the main clades were recovered showing the Neotropical *D. willistoni+D. saltans* species groups closely related to the *Dorsilopha* subgenus and *Zaprionus* genus, while the *D. melanogaster+D. obscura* groups as closely related to the *D. immigrans* (Supplementary Figure S3). The subgenera *Siphlodora* showed some disagreement between the BI and ML reconstructions, BI grouped *D. virilis* + *D. robusta*) as sister of *D. flavopilosa* species group (Figure 1), not recovered in the ML (Figure 2). Regarding other clades within the *Siphlodora* subgenus, the *D. repleta* and *D. mesophragmatica* groups were recovered as sister clades and close related with *D. canalinea* group, showing the same topology as reported by Robe *et al.*, (2010a) and Russo *et al*., (2013).

Regarding the *Drosophila* subgenus, three main lineages were recovered in our analysis resembling the phylogenetic relationships recovered by Robe *et al.* (2010) and Yassin (2013) (Figure 3). Here, the lineages will be named after the first species described belonging to that lineage (Robe *et al*., 2010). The *mediostriata* lineage (Figure 3, PP=1, BS=51) encompasses the *D. calloptera* and *D. guaramunu*

groups and some species belonging to *D. tripunctata* group. The *cardini* lineage (PP=0.76, BS=75) comprehends the two sister groups of species: *D. cardini* and *D. guarani*. The *tripunctata* lineage (PP=1;BS=96) holds the *D. pallidipennis* group and some species of the *D*. *tripunctata* group. Regarding *D. lutzii*, the BI reconstruction places *D. lutzii* group as sister clade of *D. pallidipennis* group (Figure 1; PP=1) within the *tripunctata* lineage, and this clade is recovered as sister clade to *D. paraguayensis*, belonging to subgroup II of *D*. *tripunctata* group (Figures 1; PP=0.92). The ML phylogenetic reconstruction places *D. lutzii* group as sister clade of *D. pallidipenis+D. paraguayensis* (Figure 2; BS=0.54), however with low branch support (BS=0.65) (Figure 2).

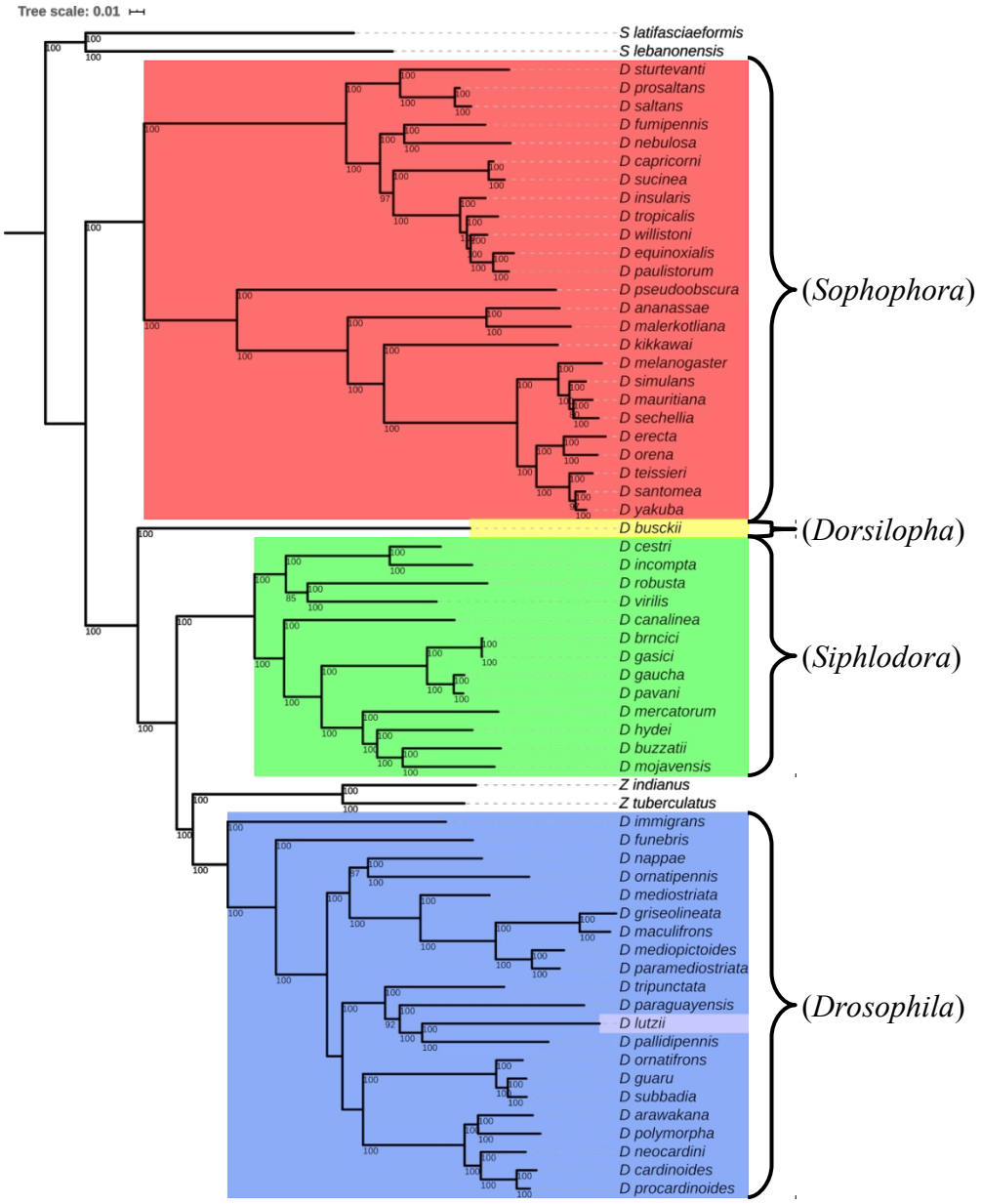


Figure 1. Bayesian Inference phylogenetic relationship reconstruction based on nuclear+mitocondrial data. Posterior probability (PP) values are shown under the branches, and values lower than 80 were omitted. The subgenera *Drosophila*, *Siphlodora*, *Dorsilopha* and *Sophophora* are highlighted in blue, green, yellow and red, respectively. Light blue highlight the *D. lutzii*'s clade.

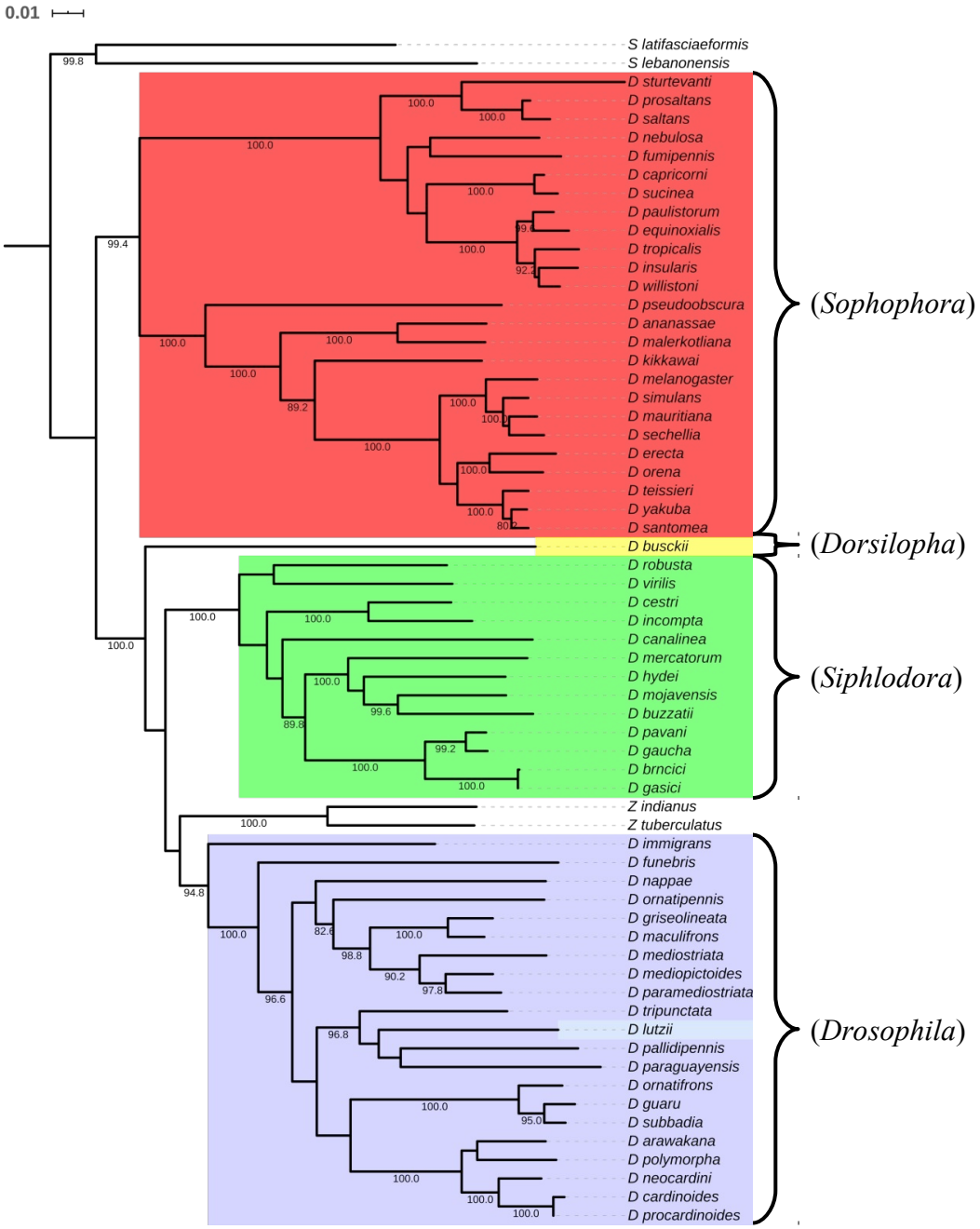


Figure 2. Maximum Likelihood phylogenetic relationship reconstruction based on nuclear+mitochondrial data. Bootstrap (BS) values are shown under the branches and values lower than 80 were omitted. The subgenera *Drosophila, Siphlodora, Dorsilopha* and *Sophophora* are highlighted in blue, green, yellow and red, respectively. Light blue highlights *D. lutzii*'s clade.

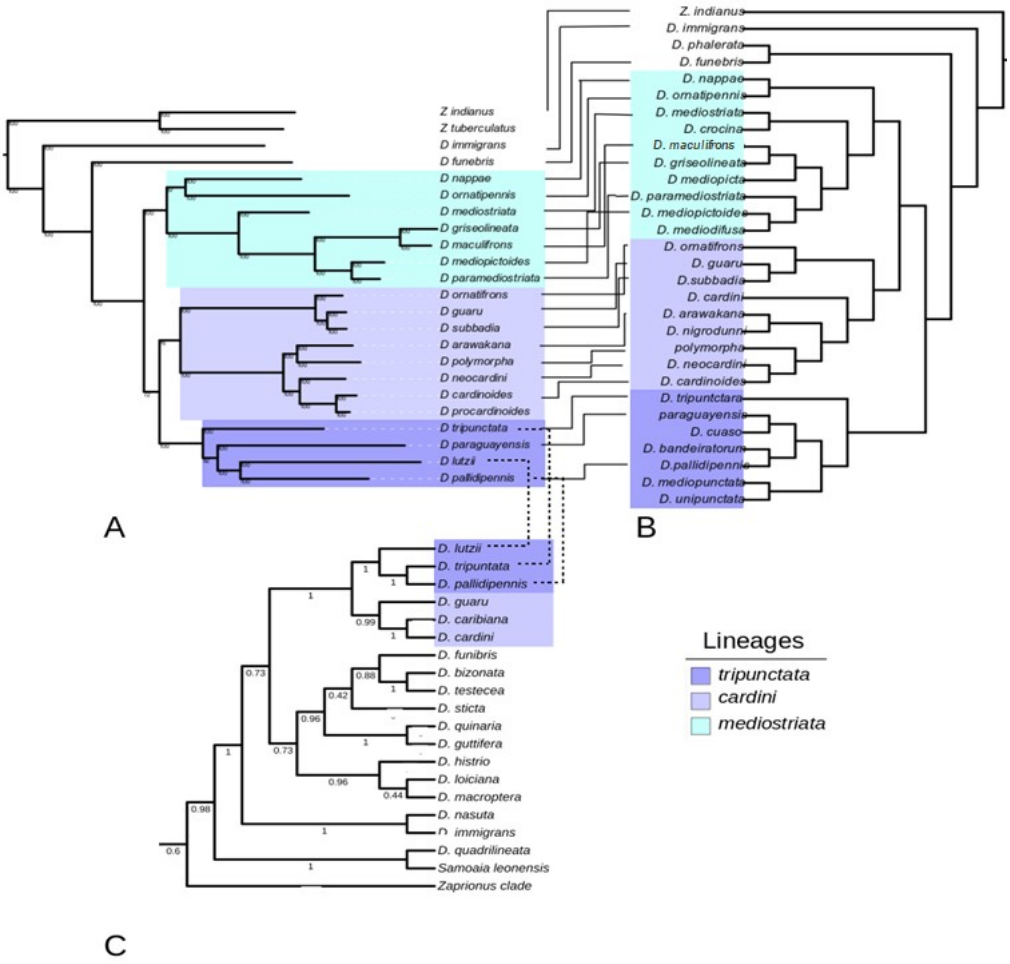


Figure 3.Phylogenetic relationships hypothesis based on Bayesian Inferences emphasizing the subgenus *Drosophila* (*Drosophila*) [sensu Yassin, 2013]. (A) Phylogenetic relationship recovered by this study, using 11 molecular markers, (B) Robe *et al.* (2010) phylogenetic relationship using four molecular markers, and (C) Yassin (2013) phylogenetic relationships using seven molecular markers. Lineage names follows Robe *et al.* (2010).

4. Discussion

The present study represents important advances for FBD studies, not only by the robustness of the phylogenetic reconstruction presented here but also because of the first the whole genome of *D. lutzii* were sequenced, enabling future researches that requires genome sequencing and clarity regarding evolutionary relationship. Even that we used a traditional phylogenetic method, the high number of molecular markers provided an unprecedented molecular reconstruction of the

evolutionary relationship of *D. lutzii* once that no many genomes of *Drosophila* subgenus are available for phylogenomic approaches. Regarding the general topology of the genus *Drosophila* our results recovered the previous data (Markow and O'Grady, 2006; Robe *et al*., 2010a; 2010b; Russo *et al*., 2013; Yassin, 2013; O'Grady and DeSalle, 2018). Regarding the *Sophophora* subgenus, the close relationship between the *D.willistoni* and *D. saltans* species groups and between the *D. melanogaster* and *D. obscura* species groups was recovered (Markow and O'Grady, 2006; Russo *et al*., 2013; Yassin, 2013; O'Grady and DeSalle, 2018). In fact, this subgenus seems to be paraphyletic, since the *Lordiphosa* genus appears to be closely related to the clade *willistoni*+*saltans* species group (O'Grady and DeSalle, 2018). Concerning the relationships within the *Siphlodora* subgenus, the major pattern shown by Robe *et al*. (2010a) was also recovered here. The exception is the phylogenetic position of *D. flavopilosa* species group. In our analysis, the BI and ML methods resulted in incongruencies. This may be a consequence of the absence of *D. annulimana* species group in our data, since it seems to be the *D. flavopilosa* sister clade (Robe *et al.,* 2010a).

Analyzing the phylogeny of the subgenus *Drosophila*, three main lineages are seen by Robe *et al*. (2010b): the *mediostriata*, *cardini* and *tripunctata* lineages. The topology seen is similar to Robe *et al*. (2010), regarding the lineage *cardini*, increasing the molecular markers do not solve the low support previously reported (0.76 by this study and 0.67 by Robe *et al.*, 2010b). However, different than Robe *et al*. (2010), our phylogenetic reconstruction places *cardini* group species as sister clade to *tripunctata* group species. The *D. lutzii* species group is placed within the *tripunctata* lineage, in agreement to Yassin (2013). However, the inner clades do not agree, our BI reconstruction places *D. lutzii* as sister clade to *D. pallidipennis*, and *D. lutzii*+*D. pallidipennis* clade as sister of *D. paraguayensis*. A distinct pattern was shown by ML analysis, were *D. lutzii* is the sister clade of *D. paraguayensis*+*D. pallidipennis* clade, however showing lower bootstrap values (BS=0.62).

Considering this and comparing the morphology of male terminalia of *D. lutzii* species group with the *Drosophila* and *Siphlodora* subgenus it is possible to highlight the similarities in the morphology between *D. lutzii* species group and the *tripunctata* radiation (Yassin, 2013). Following Yassin (2013), different from

Siphlodora species, *D. lutzii* as well as the *Drosophila* subgenus species their cercus are not merged to the epandrium. Many species of *Siphlodora* subgenus shows numerous and longs bristles at the inner part of the epandrium, not seen in the *D. lutzii* species group. Another interesting morphological trait shared by *D. lutzii* and the *Drosophila* subgenus species are the inner paraphyles fused to the aedeagus and aedeagal apodeme and the salient dorsal arch (Yassin, 2013).

Regarding the classical studies, based on ecological, morphological and biogeographical traits, Throckmorton (1975) placed the *D. lutzii* species group [sensu Yassin, 2013] within the *Siphlodora* subgenus [sensu Yassin, 2013]. At that time, the male terminalia was overlooked due to the shortage of specimens with described terminalia, even though the analysis of this structure is highly appropriate for insect cladistics (Song and Bucheli,2010). Regarding Grimaldi's studies, the apomorphy related to the resource use (flowers in the case of *D. lutzii*) have promoted a phylogenetic noise in the analysis, leading to the basal placement of this species group in the Drosophilidae phylogeny. This is also observed for the *D. flavopilosa* species group (also FBD species) (Grimaldi 1990). In this way, our results agree with modern data (Yassin, 2013), in which *D. lutzii* species is placed within the *Drosophila* subgenus, phylogenetically close to the *D. tripunctata* and *D. pallidipennis* species groups.

Acknowledgements

The authors thank Dra. Vera Lucia da Silva Valente Gaiesky for all the support as head of the Laboratório de Drosophila. M.Sc. Tuane Letícia Carvalho for help with DNA extraction; and Zacchary Cohen and the Center for High Throughput Computing Team at the University of Wisconsin-Madison for the bioinformatic support. Funding CNPq (grant: Universal/CNPq 402447/2016-6), CAPES and FAPERGS.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 12 Oct 2019

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham SON, Prjibelski AD et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477. doi: 10.1089/cmb.2012.0021

Brncic D (1966) Ecological and Cytogenetic Studies of *Drosophila flavopilosa*, a Neotropical Species Living in *Cestrum* Flowers. Evolution (N Y) 20:16–29. doi: 10.2307/2406146

Brncic D (1983) Ecology of flower-breeding *Drosophila*. In: Ashburner M, Carson HL and Thompson Jr. T (eds) The Genetics and Biology of *Drosophila* (vol 3d). pp 333–382

Brown JH and Lomolino M V. (1998) Biogeography. Doi: 10.4324/9781315841236

Chikhi R and Medvedev P (2014) Informed and automated k-mer size selection for genome assembly. Bioinformatics 30:31–37. doi: 10.1093/bioinformatics/btt310

Cordeiro J, de Oliveira JHF, Schmitz HJ and Vizentin-Bugoni J (2020) High niche partitioning promotes highly specialized, modular and non-nested florivore–plant networks across spatial scales and reveals drivers of specialization. Oikos. doi: 10.1111/oik.06866

Da Lage JL, Kergoat GJ, Maczkowiak F, Silvain JF, Cariou ML and Lachaise D (2007) A phylogeny of Drosophilidae using the *Amyrel* gene: Questioning the *Drosophila melanogaster* species group boundaries. J Zool Syst Evol Res 45:47–63. doi: 10.1111/j.1439-0469.2006.00389.x

De Ré FC, Robe LJ, Wallau GL and Loreto ELS (2017) Inferring the phylogenetic position of the *Drosophila flavopilosa* group: Incongruence within and between mitochondrial and nuclear multilocus datasets. J Zool Syst Evol Res 55:208–221. doi: 10.1111/jzs.12170

Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO and Bini LM (2013) Darwinian shortfalls in biodiversity conservation. Trends Ecol Evol 28:689–695. doi: 10.1016/j.tree.2013.09.003

Gao H, Yang Q, Yan X, Wang Z, Feng J, Yang X, Gao S, Feng L, Cheng X, Jia C et al. (2012) Exploring antibiotic resistant mechanism by microcalorimetry. 321–324. doi: 10.1007/s10973-011-1362-7

Graham CH, Storch D and Machac A (2018) Phylogenetic scale in ecology and evolution. Glob Ecol Biogeogr 27:175–187. doi: 10.1111/geb.12686

Grimaldi DA (1990) A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). Bull Am MUSEUM Nat Hist 197:1–139.

Guindon S and Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol 52:696–704. doi: 10.1080/10635150390235520

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

Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. doi: 10.1093/nar/gkf436

Lanfear R, Frandsen PB, Wright AM, Senfeld T and Calcott B (2017) Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 34:772–773. doi: 10.1093/molbev/msw260

Letunic I and Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. doi: 10.1093/nar/gkz239

Ludwig A, Vidal NM, Loreto ELS and Sepel LM. (2002) *Drosophila incompta* development without flowers. Drosoph Inf Serv 85:19–20.

Markow TA and O'Grady P (2008) Reproductive ecology of *Drosophila*. Funct Ecol 22:747–759. doi: 10.1111/j.1365-2435.2008.01457.x

Markow TA and O'Grady PM (2006) Phylogenetic relationships of Drosophilidae. *Drosophila*. doi: 10.1016/b978-012473052-6/50001-9

Miller MA, Pfeiffer W and Schwartz T (2010) The CIPRES science gateway: A community resource for phylogenetic analyses. Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA, pp 1–8

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

Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, German Grehov OG, Kandrov D, Rasputin K, Syabro M et al. (2012) Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28:1166–1167. doi: 10.1093/bioinformatics/bts091

Pélandakis M and Solignac M (1993) Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J Mol Evol 37:525–543. doi: 10.1007/BF00160433

Pryszcz LP and Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113. doi: 10.1093/nar/gkw294

Remsen J and O'Grady P (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol Phylogenet Evol 24:249–264. doi: 10.1016/S1055-7903(02)00226-9

Robe LJ, De Ré FC, Ludwig A and Loreto ELS (2013) The *Drosophila flavopilosa* species group (Diptera, Drosophilidae) An array of exciting questions. Fly (Austin) 7:59–69. doi: 10.4161/fly.239523

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

Robe LJ, Valente VLS and Loreto ELS (2010b) Phylogenetic relationships and macro-evolutionary patterns within the *Drosophila tripunctata* "radiation" (Diptera: Drosophilidae). Genetica 138:725–735. doi: 10.1007/s10709-010-9453-0

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. doi: 10.1093/sysbio/sys029

Russo CAM, Mello B, Frazão A and Voloch CM (2013) Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). Zool J Linn Soc 169:765–775. doi: 10.1111/zoj.12062

Song H and Bucheli SR (2010) Comparison of phylogenetic signal between male genitalia and non-genital characters in insect systematics. Cladistics 26:23–35. doi: 10.1111/j.1096-0031.2009.00273.x

Tanabe AS (2011) Kakusan4 and Aminosan: Two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. Mol Ecol Resour 11:914–921. doi: 10.1111/j.1755-0998.2011.03021.x

Tatarenkov A and Ayala FJ (2001) Phylogenetic relationships among species groups of the *virilis-repleta* radiation of *Drosophila*. Mol Phylogenet Evol 21:327–331. doi: 10.1006/mpev.2001.1002

Throckmorton LH (1975) The phylogeny, ecology, and geography of *Drosophila*. Handb Genet Vol 3 421–469.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V et al. (2019) FlyBase 2.0: The next generation. Nucleic Acids Res 47:D759–D765. doi: 10.1093/nar/gky1003

van der Linde K and Houle D (2008) A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). Insect Syst Evol 39:241–267. doi: 10.1163/187631208788784237

Yassin A (2013) Phylogenetic classification of the Drosophilidae Rondani (Diptera): The role of morphology in the postgenomic era. Syst Entomol 38:349–364. doi: 10.1111/j.1365-3113.2012.00665.x

Supplementary material

Table S1. Accession number for the referenced database for the sequences used

species	COII	COI	CYTB	12S	16S	ADH	Amyrel	GPDH	AMD	DDC	28S
D. busckii	AF478416.1	MF882465.1	-	-	KP730763.1	NC_030802.1	JXOY01000603.1	CP012526.1	AF293707.1	AF293733.1	JF735900.1
D. ornatipennis	EU493704.1	EU493573.1	EU494089.1	-	EU494307.1	AY081443.1	-	-	EU444585.1	-	AY081386.1
D. arawakana	HM006889.1	HM006881.1	-	-	AF246516.1	AY695384.1	AF491630.1	AB932683.1	EU444555.1	EU446060.1	AB932809.1
D. cardinoides	AY162975.1	HM006875.1	-	-	AF246518.1	AB932651.1	-	AB932687.1	EU444559.1	EU446064.1	AB932813.1
D. neocardini	AY847770.1	HM006876.1	-	-	AF246505.1	-	-	-	EU444581.1	EU446089.1	-
D. polymorpha	AB932792.1	HM006879.1	-	-	AF246507.1	AB932668.1	AY736495.1	AB932703.1	EU444591.1	EU446098.1	AB932837.1
D. procardinoides	HM006888.1	HM006880.1	-	-	AF246508.1	-	-	-	HM006864.1	-	-
D. funebris	EU390744.1	MG078611.1	EU494095.1	EU494432.1	EU494313.1	Y13252.1	AF335557.1	AB932692.1	-	AF293734.1	AB932819.1
D. griseolineata	EU493711.1	EU493581.1	EU494097.1	EU494434.1	EU494315.1	-	-	-	EU444565.1	EU446071.1	-
D. guaru	AY847763.1	EF569997.1	-	-	-	-	AF491631.1	-	EU444566.1	EU446072.1	-
D. maculifrons	KC571599.1	EF569998.1	-	-	-	-	-	-	EU444570.1	EU446077.1	-
D. ornatifrons	AY162977.1	EU493582.1	EU494098.1	EU494435.1	EU494316.1	AB932657.1	-	AB932693.1	EU444583.1	EU446091.1	AB932820.1
D. subbadia	AY847772.1	-	-	-	-	-	-	-	EU444594.1	EU446099.1	-
D. immigrans	EU493716.1	HQ981790.1	EU494102.1	EU494439.1	EU494320.1	M97638.1	AF491632.1	AB261142.1	EU444568.1	EU446074.1	AB932824.1

D. brncici	AY847757.1	-	-	-	-	0_EF560573.1	-	-	EF560566.1	EF559365.1	-	
D. gasici	AY847762.1	-	-	-	-	0_EF560576.1	-	-	EF560569.1	EF559368.1	-	
D. gaucha	EU390745.1	EU390733.1	-	-	-	0_EF560579.1	-	-	AF324955.1	AF324971.1	X71253.1	
D. pavani	EU390756.1	EU390732.1	EU494099.1	EU494436.1	EU494317.1	0_EF560580.1	AY736490.1	-	EF560570.1	EF559370.1	-	
D. pallidipennis	AY162982.1	EU493600.1	EU494115.1	EU494453.1	EU494334.1	-	AY736487.1	-	EU444586.1	EU446092.1	X71269.1	
D. mediopictoides	EU493746.1	EU493617.1	EU494131.1	EU494469.1	EU494349.1	-	AY733055.1	-	EU444576.1	EU446080.1	X71265.1	
D. mediostriata	EU493747.1	KX052967.1	EU494132.1	EU494470.1	EU494350.1	-	-	-	EU444577.1	EU446085.1	-	
D. nappae	AY162983.1	EF570005.1	-	-	-	-	-	-	EU444579.1	EU446088.1	-	
D. paraguayensis	AY162987.1	EF570012.1	-	-	-	-	-	-	EU444588.1	EU446094.1	-	
D. paramediostriata	AY162996.1	EF570013.1	-	-	-	-	-	-	EU444590.1	EU446096.1	-	
D. tripunctata	EU493748.1	MF882507.1	EU494133.1	EU494471.1	EU494351.1	AB932679.1	-	AB932714.1	AF293728.1	AF324964.1	AB932849.1	
D. canalinea	JF736114.1	KX275234.1	EU494091.1	JF736039.1	EU494309.1	-	KF632678.1	-	AF324952.1	AF324968.1	AF184011.1	
D. cestri	AY847758.1	JX993112.1	-	-	-	-	-	-	EU444560.1	EU446065.1	-	
D. incompta	AY847764.1	NC_025936.1	NC_025936.1	KM275233.1	KM275233.1	-	-	-	EU444569.1	EU446075.1	-	
D. buzzatii	DQ202011.1	KX275224.1	lcl|	DQ201971.1	KP730772.1	U65746.1	lcl|scaffold17	scaffold73*	lcl|scaffold9	AF324980.1	AF184008.1	

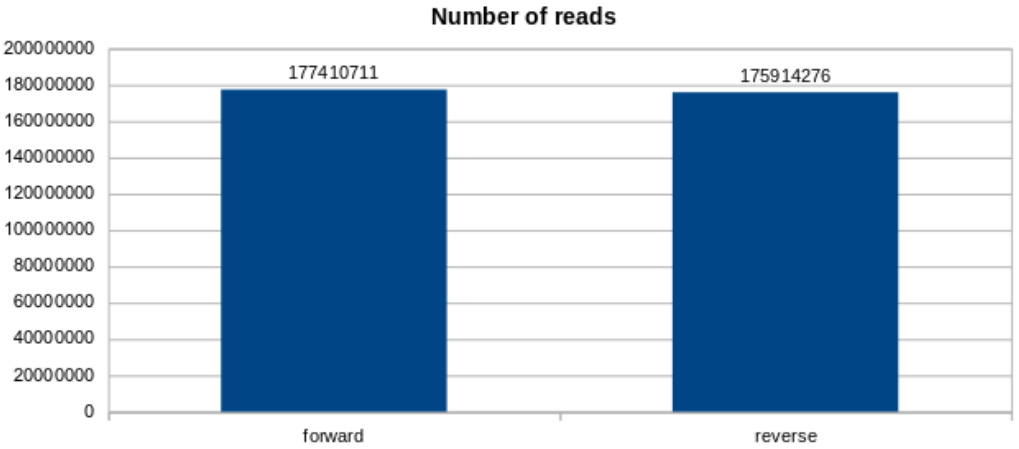
			scaffold342								
D. hydei	EU390746.1	MG086288.1	EU494230.1:177	EU494459.1	-	X58694.1	AY733042.1	L41650.1	QMEQ02000036.1	AF293737.1	-
D. mercatorum	DQ202028.1	KX275228.1	EU494121.1	EU494460.1	EU494340.1	DQ471664.1	-	-	AF324957.1	AF324973.1	AF184010.1
D. mojavensis	EU493738.1	DQ383703.1	EU494122.1	BK006339.1	BK006339.1	AAPU01010615.1	AAPU01010236.1	NW_001979113.1	AAPU01010519.1	NW_001979113.1	‡Eickbush
D. robusta	EU390758.1	KR384693.1	EU494128.1	HQ849826.1	HQ849826.1	AY750138.1	SCDW01004201.1	-	AF293724.1	AF293747.1	GU597381.1
D. virilis	EU493751.1	MH423345.1	EU494135.1	HQ849831.1	HQ849831.1	KU559568.1	AF136603.1	XR_001450130.1	XM_002057599.2	AF293749.1	‡Eickbush
D. ananassae	AF474077.1	MG557557.1	BK006336.1	BK006336.1	BK006336.1	AB194426.1	AF024691.3	FJ795593.1	AAPP01015682.1	HQ631615.1	‡Eickbush
D. erecta	GQ244453.1	KX771108.1	BK006335.1	BK006335.1	BK006335.1	X54116.1	AF039562.2	DQ167751.1	QMER02000014.1	-	‡Eickbush
D. malerkotliana	EU493756.1	JQ679118.1	EU494140.1	EU494479.1	MK106019.1	AB194422.1	AY733054.1	HQ631718.1	-	HQ631632.1	HQ631448.1
D. mauritiana	AF474081.1	HM630860.1	NC_005779.1	AF200830.1	AF200830.1	Z00033.1	JF815750.1	NIGA01000001.1	NIGA01000001.1	-	JF735893.1
D. melanogaster	KT174472.1	KT174474.1	NC_024511.2	KY310613.1	KT174474.1	X78384.1	AF022713.2	NP_476565.1	AE014134.6	NP_724164.1	‡Eickbush
D. orena	VCKV01000136.1	AY757281.1	VCKV010000-70.1		AF164584.1	Z00032.1	U96158.2	DQ167752.1	VCKV01000314.1	VCKV01000314.1	JF735909.1
D. santomea	DQ382822.1	JQ679120.1	KF824871.1	KF824871.1	KF824871.1	AY804554.1	AY736503.1	-	-	-	-
D. sechellia	GQ244458.1	KJ426007.1	CM016413.1	AF200832.1	AF200832.1	X04672.1	AF039558.1	NW_001999694.1	NIFZ01000001.1	NIFZ01000001.1	‡Eickbush

D. simulans	AF474082.1	KJ767247.1	CM016414.1	AF200839.1	AF200839.1	X00607.1	NIFY01000002.1	L41647.1	NIFY01000001.1	AAST01028456.1	‡Eickbush
D. teissieri	DQ382774.1	KX771111.1	AF164586.1	-	AF164586.1	X54118.1	AF039557.2	U47809.1	AF293727.1	-	JF735895.1
D. yakuba	X03240.1	KX771113.1	KF824874.1	KF824874.1	KF824874.1	X57368.1	CM000158.2	DQ167753.1	AAEU02000279.1	AAEU02000279	‡Eickbush
D. kikkawai	AB669790.1	KY973965.1	AF164583.1	-	AF164583.1	NW_016067405.1	U96156.3	HQ631711.1	AFFH02007724.1	HQ631625.1	X71185.1
D. pseudoobscura	EU493762.1	EU493633.1	NC_018348.1	FJ899745.1	FJ899745.1	Y00602.1	NC_009006.2	SDMN01000002.1	AY754405.1	AY754449.1	‡Eickbush
D. prosaltans	HQ110561.1	AF045103.1	-	-	-	HQ110515.1	-	-	-	-	HQ110538.1
D. saltans	AF050741.1	GU597450.1	-	-	-	AY335198.1	-	AY335216.1	-	-	HQ110540.1
D. sturtevanti	HQ110562.1	MG010104.1	MG010122.1	-	-	AB026535.1	AY736506.1	AY335217.1	-	MG010071.1	HQ110542.1
D. capricorni	EU532079.1	MG010100.1	MG010118.1	EU494488.1	EU494367.1	AY335196.1	-	AY335214.1	-	MG010068.1	-
D. equinoxialis	MG010105.1	MG010090.1	MG010108.1	EU494489.1	EU494368.1	U95268.1	-	-	FJ664506.1	MG010058.1	-
D. fumipennis	EU532081.1	MG010101.1	MG010119.1	EU494490.1	EU494369.1	EU532133.1	-	-	FJ664509.1	-	-
D. insularis	MG010106.1	MG010091.1	MG010109.1	-	-	U95273.1	-	-	FJ664507.1	MG010059.1	-
D. nebulosa	EU532083.1	MG010102.1	MG010120.1	EU494491.1	EU494370.1	U95275.1	AY733060.1	L41250.1	AF293717.1	MG010069.1	JF735890.1
D. paulistorum	EU532093.1	MG010099.1	MG010117.1	EU494492.1	EU494371.1	U95270.1	-	L41648.1	FJ664503.1	MG010067.1	HQ110536.1
D. sucinea	EU532094.1	MG010103.1	MG010121.1	EU494493.1	EU494493.1	AY335197.1	-	AY335215.1	FJ664510.1	MG010070.1	-
D. tropicalis	EU532096.1	MG010092.1	MG010110.1	-	-	U95274.1	AF251140.1	-	FJ664505.1	MG010060.1	KJ746538.1
D. willistoni	MG010107.1	MG010093.1	BK006338.1	BK006338.1	BK006338.1	L08648.1	AF039560.1	L41248.1	AF293730.1	MG010061.1	‡Eickbush
Z. indianus	FJ393919.1	KJ463786.1	LWKS01000065.1	MK216814.1	LWKS01000065	KX384733.1	EF458322.1	LWKS01001500.1	EU444597.1	EU446103.1	GU597395.1

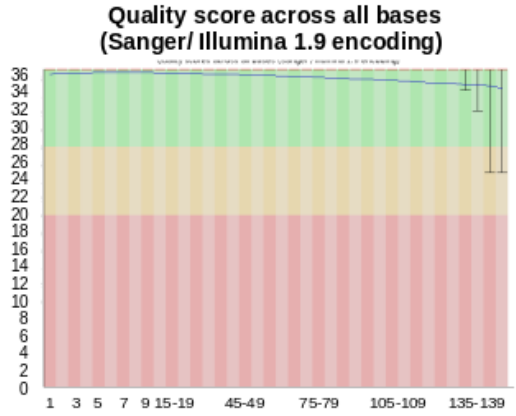
Z. tuberculatus	EU595373.1	EU493691.1	EU494193.1	EU494538.1	M93998.1	X63955.1	AY736524.1	L37039.1	AF293731.1	AF293751.1	KJ746539.1
S. latifasciaeformis	EU493813.1	EU493684.1	EU444596.1	EU446102.1	EU494410.1	-	EU494186.1	GU597377.1	GQ352255.1	EU494532.1	-
S. lebanonensis	EU493815.1	EU493686.1	QMEN02000033.1	AF091329.1	EU494411.1	JXPJ01023272.1	EU494188.1	HQ110555.1	QMEN02000001.1	EU494534.1	JXPJ01025478.1

Table S2. Full QUAST report to the genome assembly using SPAdes and Redundants pipeline. All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted

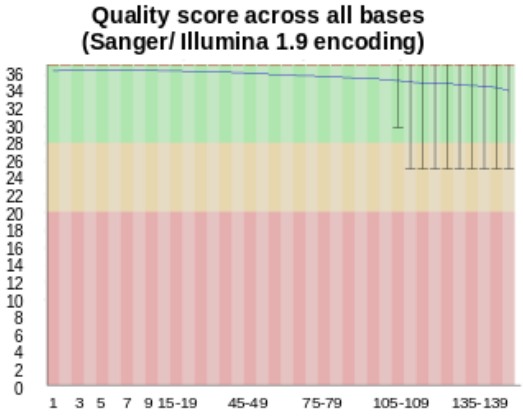
	SPAdes's Contigs	Redundants's scaffolds
# contigs (>= 0 bp)	615,741	129,541
# contigs (>= 1000 bp)	48,656	39,311
# contigs (>= 5000 bp)	9,308	9,194
# contigs (>= 10000 bp)	4,506	4,503
# contigs (>= 25000 bp)	994	994
# contigs (>= 50000 bp)	149	149
Total length (>= 0 bp)	315,034,500	223,085,922
Total length (>= 1,000 bp)	198,240,342	184,621,117
Total length (>= 5,000 bp)	125,829,728	125,085,072
Total length (>= 10,000 bp)	922,151,75	92,171,341
Total length (>= 25,000 bp)	38,527,990	38,527,990
Total length (>= 50,000 bp)	10,372,062	10,372,062
# contigs	87,033	63,818
Largest contig	250,057	250,057
Total length	226,483,272	202,420,868
GC (%)	40.04	40.64
N50	6,571	8,400
N75	1,613	2,412
L50	7,112	5,491
L75	25,607	16,989
# N's per 100 kbp	0.00	0.00



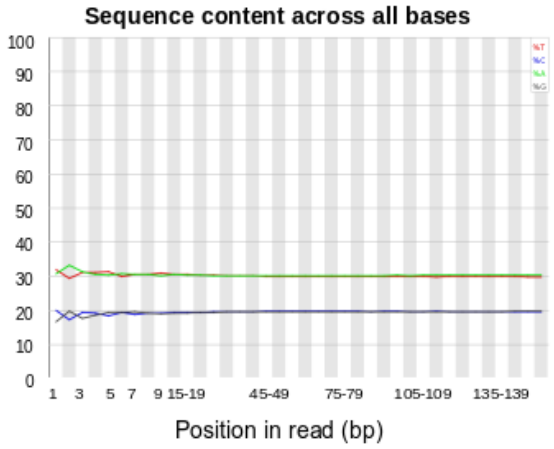
1



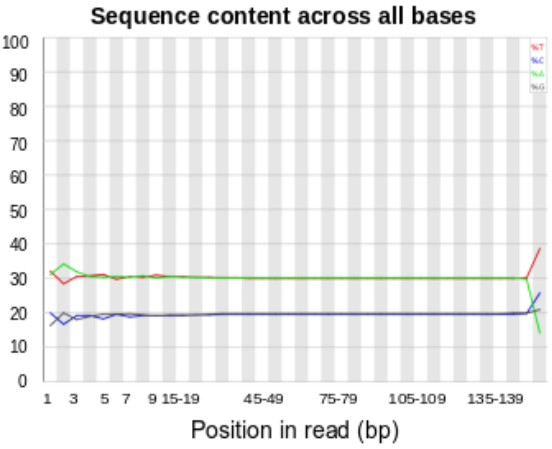
2A



2B



3A



3B

To be continue

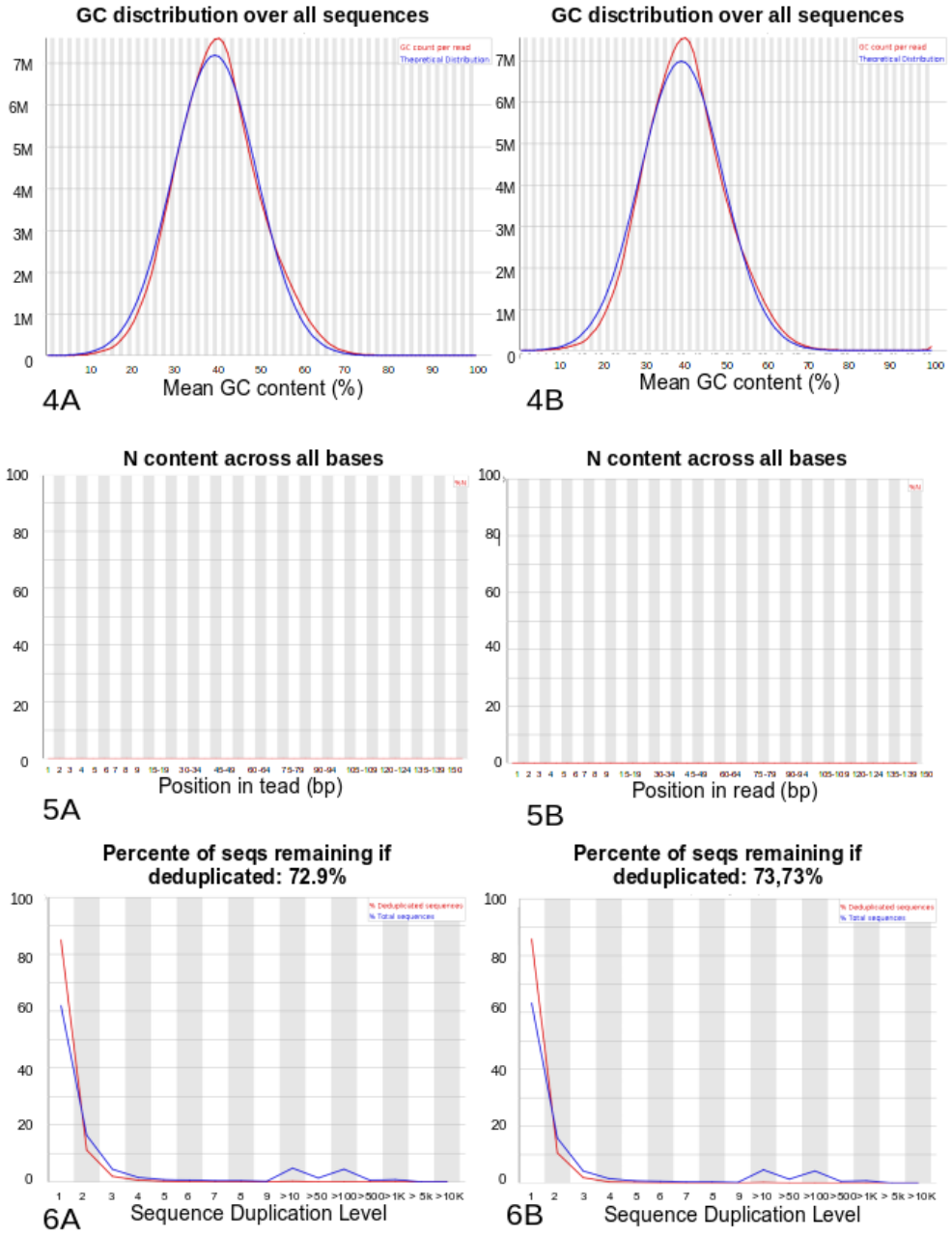


Figure S1. Sequencing quality report (fastQC) 1 - Number of reads generated by foward and reverse sequencing. 2 -Phred quality report. 3 - content of GC throw read length. 4 - GC content overall sequenced reads. 5 - number of ambiguous nucleotide throw reads length. 6 - level of sequence duplication.

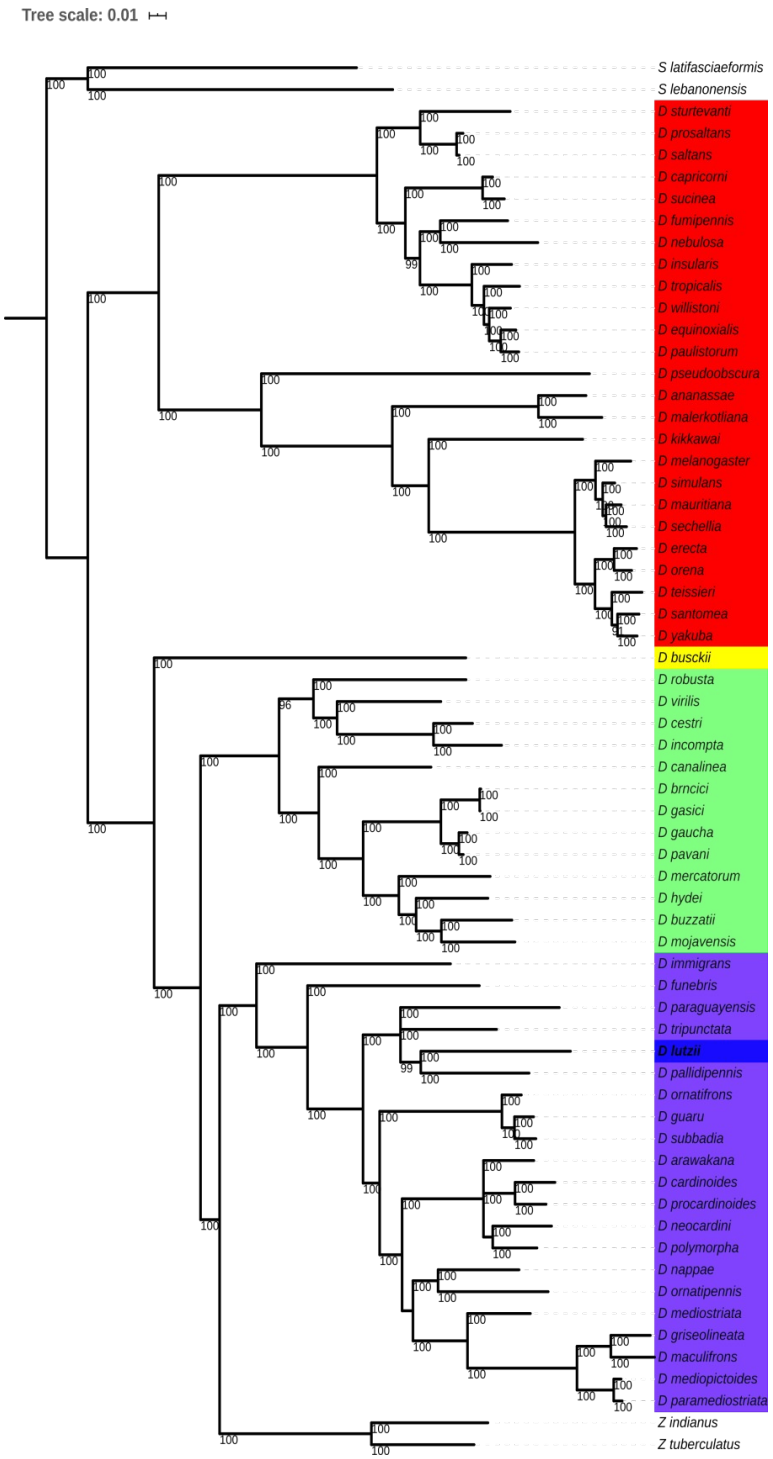


Figure S2. Nuclear data BI reconstruction, *Sophophora*, *Dorsilopha*, *Siphlodora* and *Drosophila* are represented in red, Yellow, green and blue, respectively.

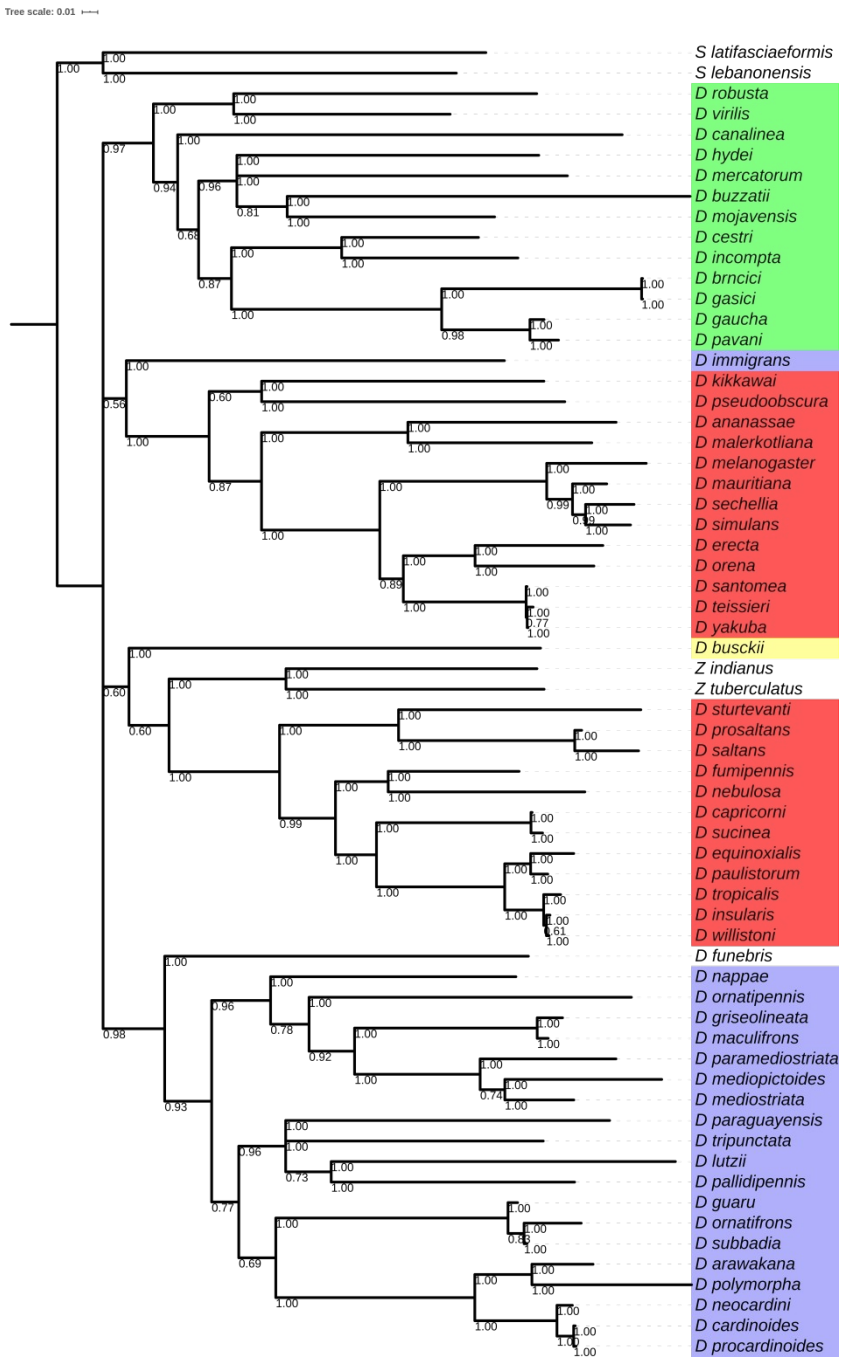


Figure S3. Mitochondrial data BI reconstruction, *Sophophora*, *Dorsilopha*, *Siphlodora* and *Drosophila* are represented in red, Yellow, green and blue, respectively.

Chapter 3 – Cytochrome P450 genes repertoire in *Drosophila lutzii* (Diptera: Drosophilidae) genome: insights from an anthophilous species

Carolina Prediger[1], Juliana Cordeiro[2], Maríndia Deprá[1]

1 – Programa de Pós Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Brazil.

2 – Departamento de Zoologia, Ecologia e Genética, Instituto de Biologia, Universidade Federal de Pelotas, Brazil.

Manuscript under preparation to be submitted as short communication to Journal of Genetics and Molecular Biology

Abstract

The study of niche specialization, niche shift and the physiological aspects related to them reveal interesting questions such as “How the use of different resources can affect the evolution of genes or gene families?” The answer to this question requires the study of different systems, and maybe the most specialized ones, to be achieved. *Drosophila* species that are strictly related to flowers are a good model to study niche specialization. *Drosophila lutzii* is a flower-breeding drosophilid able to use toxic flowers as breeding and feeding resource, for example the plant species belonging to the Solanaceae and Convolvulaceae families. Such resource use could change the selective pressure of detoxification gene families. The cytochrome P450 (P450) super-family of genes related to detoxification was characterized in *Drosophila lutzii*. The *D. lutzii*’s genome was sequenced through Illumina NovaSeq 150-bp paired-end and assembled using SPAdes algorithm. *Drosophila melanogaster*’s P450 genes were used as query in the genomic BLAST searches and GeneWise algorithm were used to annotate the *D. lutzii´*s P450 genes. Also, P450 genes of twelve *Drosophila* species were used in the *D. lutzii* genes comparison. Synonymous and non-synonymous analysis and

gene duplication events were carried out. The *D. lutzii* results show a repertoire of 80 P450 genes. It was identified one pseudo-gene and eight sequences were tagged as potential pseudo-genes. Interestingly, several gene duplication events occurred in genes associated with detoxification. Finally, all P450 genes retrieved seem to be under purifying selection. Further analysis will show how the use of different resources can affect the evolution of Cytochrome P450 genes repertoire in *D. lutzii* genome.
Key words: Cyp genes, flower-breeding *Drosophila*, anthophilous drosophilids, niche specialization, detoxification

Several species traits are associated with the ability to use a specific resource as feeding and/or breeding site. Especially for toxic resources, a trade-off between nutrient acquisition and toxic compounds processing should determine the species niche breadth. Clearly, the close relationship between the species and its feeding and breeding resources results in selection pressure that shapes the evolution history of some gene families, such as those related to chemical perception and detoxification (McBride and Arguello, 2007; Goldman-Huertas *et al*., 2015; Hoang *et al*., 2015; De Ré, 2016; Chialvo and Werner, 2018; Etges, 2019; Markow, 2019). Flower-breeding drosophilids (FBD) are anthophilous Drosophilidae species that use flowers as feeding and breeding sites, and are unable to complete their life cycle without this resource (Brncic, 1983). FBD species are an interesting model to study the effects of niche specialization on genes and genomes, in part due to the large number of available genomes of species with different ecological requirements allowing comparative genomics among species with different levels of host specialization (so far, 74 Drosophilidae genomes are available at National Center of Biotechnology Information website).

Drosophila lutzii is a generalist FBD species. Adults emerge from flowers of plant species belonging to Convolvulaceae, Cucurbitaceae, Passifloraceae and Solanaceae families (Cordeiro *et al*., 2020). Some plant species of these families produce several alkaloids involved with herbivory avoidance. For example, the alkaloid scopolamine produced by several Solanaceae species reduces herbivory

by Leptoptera caterpillars (Arab and Trigo, 2011; Arab *et al*., 2012).Therefore, the ability to use such plant species as feeding and breeding sites must require a drosophilid physiological adaptation to detoxify these chemicals.

The enzymes produced by the Cytochrome P450 genes (P450) are one of those responsible for detoxification processes, in fact this super-family is able to catalyze a large spectrum of endogenous and exogenous substrates. These genes have been largely studied in insects specially because of the association between P450 with resistance to insecticides (Giraud *et al*., 2010). It is expected that somehow the close relationship between the insects and their plant hosts shape the evolution of some P450 genes, in structure of the gene and/or expression patterns (Etges, 2019; Markow, 2019). Indeed, several evidences shows the association between P450 genes with feeding and breeding resource use (Frank and Fogleman, 1992; Danielson *et al.*, 1998; Bono *et al*., 2008; Guillén *et al*., 2014; Soto *et al.,* 2014; Chialvo and Werner, 2018). However, except for Rane *et al*. (2019) study, the evolution of P450 genes in FBD species is unknown. In this way, the analysis of the P450 genes in *D. lutzii* will improve the comprehension of the genomic evolution associated with ecological specialization. In this study, we aimed to characterize the P450 genes repertoire in the *D. lutzii* genome and compare with P450 genes within the 12 genomes previously analyzed by Good *et al.* (2014).

To acquire *D. lutzii*, flowers of *Ipomea purpurea* were collected and kept in sterile vials until *Drosophila*'s adults emergency. Males were identified by genitalia morphology and DNA barcode. The NucleoSpin DNA Insect kit were used to DNA extraction of 10 males. The *denovo* whole genome sequencing was performed at Macrogen Inc. using Truseq DNA Nano 350-bp and the sequence platform was Illumina Novaseq 150-bp, pair-end. Genome assemble approach used was *denovo* on SPAdes (Bankevich *et al.*, 2012) followed by cleaning with Redundans pipeline (Pryszcz and Gabaldón, 2016). The P450 gene coding region (CDS) and the amino acids (AA) sequences from *D. melanogaster* genome were downloaded from Flybase website (flybase.org; Thurmond *et al*., 2019).These sequences were used as query for a local tBLASTn searches against the *D. lutzii* genome using

BLAST+ (Camacho *et al*., 2009) The matches and a contingency of 5-kb up and downstream were retrieved using SAMtools (Li *et al*., 2009). In order to predict the gene structure, the GeneWise algorithm implemented in the wise2 package (Birney *et al*., 2004) were used combining all *D. melanogaster* P450 protein sequences with all potential *D. lutzii*'s P450 nucleotide sequences. Early stop codons were searched and identified for the *D. lutzii* sequences, and sequences with early stop codon were tagged as pseudo-genes. CDS and AA sequences from the 12 *Drosophila* species (*D. melanogaster, D. simulans, D. sechelia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, persimilis, D. willistoni, D. mojavensis, D. virilis* and *D. grimshawi*) were obtained from Good *et al.* (2014). The AA sequences of the 13 species were aligned using the G-ins-i algorithm implemented on MAFFT (Katoh, 2002). To identify the genes of each retrieved *D. lutzii* sequence, a UPGMA dendrogram were constructed using MAFFT (Katoh, 2002). In the resulted UPGMA clade, the *D. lutzii* sequences were named after the orthologous gene clade where the sequences were located in. Sequences that do not group with any orthologous were called Cyp-like. Process of gene duplication and gene loss were checked for *D. lutzii* P450 genes.

In the UPGMA tree, for each clade containing one or more *D. lutzii* sequence genes, the DNA sequences were aligned by codon using the algorithm MUSCLE (Edgar, 2004) implemented in the MEGAX (Kumar *et al*. 2018). Each *D. lutzii* sequence had its status checked into one of the following categories: i) complete gene, sequence retrieved from start to stop codon (C); ii) partial gene, missing the upstream gene sequence (Pu); iii) partial gene missing the downstream gene sequence (Pd); and, iv) partial gene missing the upstream and downstream gene sequence (Pud). For each *D. lutzii* sequence gene, the number of synonymous substitutions per synonymous site (dS) and non-synonymous substitutions per non-synonymous site (dN) was calculated. It was also calculated the dS/dN ratio for pairwise comparison with the 12 species genes using SNAP program (Synonymous Non-synonymous Analysis Program), as available on the HIV database website (www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) (Korber, 2000).

After the searches it was retrieved 89 potential P450 genes from the *D. lutzii* genome. Of these, we recovered 34 complete gene sequence with length ranging from 1,461-bp to 1,914-bp; 22 partial genes missing the upstream gene sequence with length ranging from 525-pb to 1,638-pb; 12 partial genes missing the downstream gene sequence with length ranging from 189-bp to 1,622-bp; and 12 partial genes missing the upstream and downstream gene sequences with length ranging from 126-bp to 1,200-bp. The analysis of dS and dN for these 80 *D. lutzii* Cyp sequences suggest the action of purifying selection in these sequences (Table 1). A pseudogene sequence was found with 542-bp and multiples early stop codons. Eight retrieved sequences were identified as Cyp-like genes due to the ungrouped arrangement in the UPGMA dendrogram of orthologous sequences (Figure 1, also available at here). Although these sequences do not show stop codons, their length size (from 312-bp to 1,551-bp) and the ungrouped status in the UPGMA tree might indicate that these sequences are pseudogenes.

Table 1. Repertoire of retrieved P450 genes within *D. lutzii* genome. Information about recovered fragment size, mean values of dS, dN and dS/dN ratios are shown. Gene status regarding the recovered sequence gene and number of genes that they were compared to are also shown. C: complete gene; Pd: partial gene missing downstream gene sequence; Pu: partial gene missing upstream gene sequence; Pud: partial gene missing upstream and downstream gene sequence, U – unknown. ♦ Sequence above saturation, unable to calculate neither dS nor dN. ◊ Non calculated due to the ungrouped arrangement in the UPGMA dendrogram of orthologous sequences.

genes	dS	dN	dS/dN	Sequence length	Status	N. Seq. Compared
Cyp12a4/5	2.13	0.27	7.90	558	Pd	33
Cyp12a4/5	2.00	0.20	9.82	1044	Pu	33
Cyp12a4/5	2.91	0.29	9.71	1638	Pu	33
Cyp12b2	1.98	0.16	12.24	1506	Pu	11
Cyp12c1	2.13	0.32	6.59	1581	C	8
Cyp12d1/2	1.50	0.21	7.00	1449	Pu	16
Cyp12e1	2.69	0.16	15.81	1299	Pu	15
Cyp12g1	1.81	0.20	9.09	1464	Pu	2

genes	dS	dN	dS/dN	Sequence length	Status	N. Seq. Compared
Cyp18A1	1.96	0.66	2.98	417	Pud	12
Cyp28a5	2.12	0.26	8.08	1524	C	16
Cyp28a5	3.19	0.21	15.02	1584	C	16
Cyp28c1	1.67	0.33	7.37	189	Pd	14
Cyp28c1	2.00	0.40	7.89	525	Pu	14
Cyp28c1	1.72	0.21	7.30	762	Pu	14
Cyp28d1/2	2.01	0.55	5.32	1509	C	20
Cyp301a1	1.55	0.09	17.11	1683	C	12
Cyp302a1	1.64	0.24	8.60	564	Pud	13
Cyp302a1	1.42	0.21	8.97	606	Pud	13
Cyp303a1	1.62	0.08	20.65	1425	Pu	12
Cyp304a1	2.23	0.17	13.23	1542	C	14
Cyp305a1	1.34	0.13	13.52	1437	Pu	12
Cyp306a1	3.11	0.42	7.46	360	Pd	12
Cyp307a2	3.28	0.19	17.52	753	Pud	25
Cyp307a2	4.10	0.17	24.41	1161	Pud	25
Cyp309a1	2.54	0.42	5.99	561	Pd	13
Cyp309a1	2.27	0.28	8.92	1371	Pu	13
Cyp309a2	1.65	0.18	9.27	974	Pu	12
Cyp310a	2.66	0.21	11.98	1530	C	11
Cyp311a1	1.82	0.25	7.12	1464	C	12
Cyp312a1	1.58	0.22	7.44	1542	C	12
Cyp313a4	2.44	0.42	5.76	1482	C	20
Cyp313a4	3.65	0.22	16.77	1251	Pu	20
Cyp313b1	1.41	0.11	13.42	1398	Pu	12
Cyp314a1	1.86	0.05	37.20	564	Pu	13
Cyp314a1	1.00	0.05	19.48	597	Pud	13
Cyp317a	1.49	0.15	9.67	1557	C	12
Cyp318a1	1.48	0.21	10.42	969	Pd	13
Cyp318a1	1.92	0.29	10.57	660	Pu	13
Cyp49a1	1.57	0.11	14.13	1914	C	12

genes	dS	dN	dS/dN	Sequence length	Status	N. Seq. Compared
Cyp4aa1	0.90	0.10	9.01	1521	Pu	12
Cyp4ac	3.31	0.20	16.28	1483	Pd	23
Cyp4ad1	1.89	0.15	12.36	1443	Pu	12
Cyp4ae1	2.09	0.34	6.03	204	Pud	12
Cyp4c3	1.65	0.15	21.07	471	Pd	12
Cyp4c3	2.39	0.75	3.27	621	Pu	12
Cyp4d1	1.62	0.18	10.53	1551	C	22
Cyp4d1	1.82	0.46	4.05	234	Pud	22
Cyp4d14	2.30	0.21	11.04	1545	C	12
Cyp4d20	1.72	0.18	10.09	1524	C	13
Cyp4d8	1.94	0.20	9.40	1425	Pu	12
Cyp4g1	1.41	0.07	19.68	1659	C	12
Cyp4g15	1.01	0.05	24.40	1622	Pd	13
Cyp4p1/2/3	2.44	0.19	13.05	1617	C	13
Cyp4s3	1.25	0.11	12.06	605	Pd	12
Cyp4s3	1.43	0.18	7.85	903	Pu	12
Cyp6a120	1.96	0.17	11.18	1500	C	6
Cyp6a13	3.08	0.17	17.75	1542	C	11
Cyp6a17/23	3.08	0.17	17.75	1497	C	24
Cyp6a2	2.13	0.15	14.73	1521	C	21
Cyp6a2	2.20	0.14	16.78	1521	C	21
Cyp6a22	2.02	0.14	14.63	1491	C	12
Cyp6a9	2.23	0.16	14.56	652	Pd	20
Cyp6d2	1.88	0.19	9.91	1128	Pd	11
Cyp6d2	1.38	0.11	12.11	315	Pud	11
Cyp6d4	1.62	0.18	9.65	690	Pd	13
Cyp6d4	1.30	0.20	6.86	882	Pu	13
Cyp6d5	2.39	0.26	8.94	1542	C	15
Cyp6g1	1.71	0.18	9.40	1566	C	14
Cyp6g2	1.13	0.16	7.00	1551	C	12
Cyp6t1	2.38	0.16	14.85	1584	C	11

genes	dS	dN	dS/dN	Sequence length	Status	N. Seq. Compared
Cyp6t3	1.87	0.24	7.66	1503	C	11
Cyp6u1	2.03	0.23	8.72	1461	C	12
Cyp6v1	1.29	0.08	17.47	1587	C	12
Cyp6w1	2.11	0.25	8.41	1518	C	11
Cyp9b1/2	1.74	0.16	10.97	1515	C	16
Cyp9c1	1.50	0.21	7.00	1584	C	12
Cyp9f2	1.35	0.16	8.52	1200	Pud	13
Cyp9h	2.17	0.34	6.81	1545	C	15
Cyp4d2	♦	♦	♦	274	Pd	11
Cyp4d2	♦	♦	♦	126	Pd	11
Pseudo-gene	◊	◊	◊	542	U	◊
Cyp-like	◊	◊	◊	455	U	◊
Cyp-like	◊	◊	◊	408	U	◊
Cyp-like	◊	◊	◊	546	U	◊
Cyp-like	◊	◊	◊	312	U	◊
Cyp-like	◊	◊	◊	486	U	◊
Cyp-like	◊	◊	◊	534	U	◊
Cyp-like	◊	◊	◊	1551	U	◊

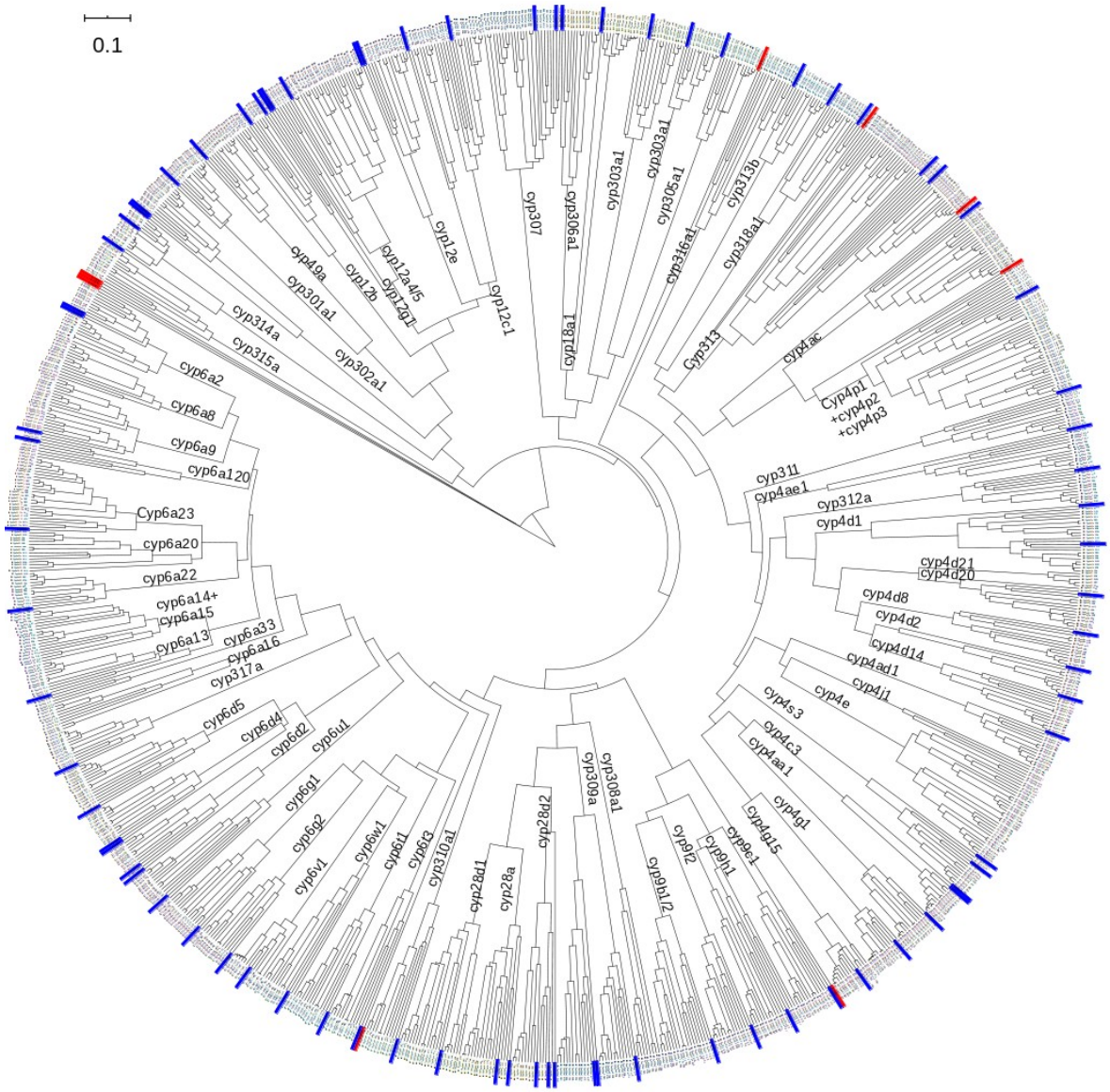


Figure 1. UPGMA dendrogram for P450 gene family sequences of 13 *Drosophila* species. The blue bar indicates the *D. lutzii* P450 genes and the red bar represent Cyp-like sequences. Cyp gene names are identified in each dendrogram branch.

Gene duplication and gene deletion events were identified in the *D. lutzii* P450 genes. In the UPGMA dendrogram the duplication events are observed for 15 genes (*Cyp12a4/5, Cyp28a5, Cyp28c1, Cyp302a1, Cyp307a2, Cyp309a1, Cyp313a4, Cyp314a1, Cyp318a1, Cyp4C3, Cyp4d1, Cyp4s3, Cyp6a2, Cyp6d2* and *Cyp6d4*) (Figure 1). Some of these genes are involved in the following processes: ecdysone synthesis pathway (Gilbert, 2004; Namiki *et al*., 2005; Chung *et al*., 2009); sex pheromones detection (Maïbèche-Coisne*et al.*, 2002); detoxification insecticides (Araine*t al*., 2018; Xu *et al*., 2018); detoxification of lithium chloride (Kasuya*et al.*, 2009); detoxification of α-amanitin (Chialvo and Werner, 2018); and detoxification of camptothecin (Thomas *et al.,* 2013) (Table 2). Some duplicated genes (*Cyp28a5,Cyp6a2, Cyp302a1,Cyp309a2*) were also found up regulated in the genome of other *Drosophila* species and they have been associated to detoxification of alkaloids related to the use of cactus and fungus resources (Dworkin and Jones, 2009; Hoang *et al*., 2015; Chialvo and Werner, 2018). Furthermore, it was detected gene loss events in *D. lutzii* genome (Table 2) (*Cyp307a1, Cyp308a1, Cypyp313a1/2/3/5, Cyp315a1, Cyp316a1, Cyp4d21, Cyp4e1/2, Cyp4e3, Cyp6a14/15, Cyp6a16, Cyp6a19/20* and *Cyp6a8/18*).

Table 2: Or gene lost or duplicated in *D. lutzii* and their biological function.

OR gene (*D. lutzii*)	Related to the following biological process	Source
Cyp307a1 (deleted)	Ecdysone pathway	Namiki *et al*. (2005); Gilbert (2004)
Cyp315a1 (deleted)	Ecdysone pathway	Namiki *et al*. (2005); Gilbert (2004)
Cyp4d21 (deleted)	Circadian cycle	Erion *et al.* (2016)
Cyp4e3 (deleted)	Regulation of H_2O_2 levels	Terhzaz *et al.*, (2015)
Cyp6a8/18 (deleted)	insecticides resistance (DDT)	Le Goff *et al.* (2003)
Cyp6a19/*20* (deleted)	Associated with males Aggressive behavior	(Dierick and Greenspan, 2006)
Cyp307a2 (2 copies)	Ecdysone pathway	Namiki *et al*. (2005);Gilbert (2004)
Cyp302a1 (2 copies)	Ecdysone pathway	Namiki *et al*. (2005);Gilbert (2004)
Cyp314a1	Ecdysone pathway	Namiki *et al*.

(2 copies)		(2005);Gilbert (2004)
Cyp4s3 (2 copies)	Sex pheromones detection	Maïbèche-Coisne *et al.* (2002)
Cyp309a1 (2 copies)	Resistance to lithium chloride	Kasuya *et al.* (2009)
Cyp6d2 (2 copies)	Tolerance to α-amanitin, camptothecin	Chialvo and Werner (2018); Thomas *et al.* (2013)
Cyp12a4/5 (3 copies)	insecticide resistance (lufenuron)	Bogwitz *et al*. (2005)
Cyp28a5 (2 copies)	Upregulated in *D. mettleri* on cactus soil media	Hoang *et al.* (2015)
Cyp28c1 (3 copies)	Expressed on salivary glands	Chung *et al*. (2009)
Cyp318a1 (2 copies)	For *D. melanogaster,* reducing its expression reduces survival rates	Chung *et al*. (2009)
Cyp4C3 (2 copies)	For *D. melanogaster,* reducing its expression reduces survival rates	Chung *et al*. (2009)
Cyp4d1 (2 copies)	Downregulated by high ecdysone levels	Davies *et al.* (2006)
Cyp6a2 (2 copies)	Tolerance to caffeine and DDT,	Brun *et al.* (1996)

We did not find the biological processes to the followed deleted genes *Cyp313a1/2/3/5, Cyp316a1, Cyp4e1/2, Cyp6a14/15, Cyp6a16, Cyp308a1* and the flowed duplicated genes *Cyp6d4* and *Cyp313a4*.

The dynamic of gene duplication and loss within the *Drosophila* genus phylogeny (Figure 2) shows a large variation on P450 gene number even within close related species. Intriguingly, most of the *D. lutzii* P450 duplicated genes are related to exogenous chemicals processing, suggesting that the number of P450 gene copies may be a natural selection target to feeding niche shift (Good *et al.*, 2014).

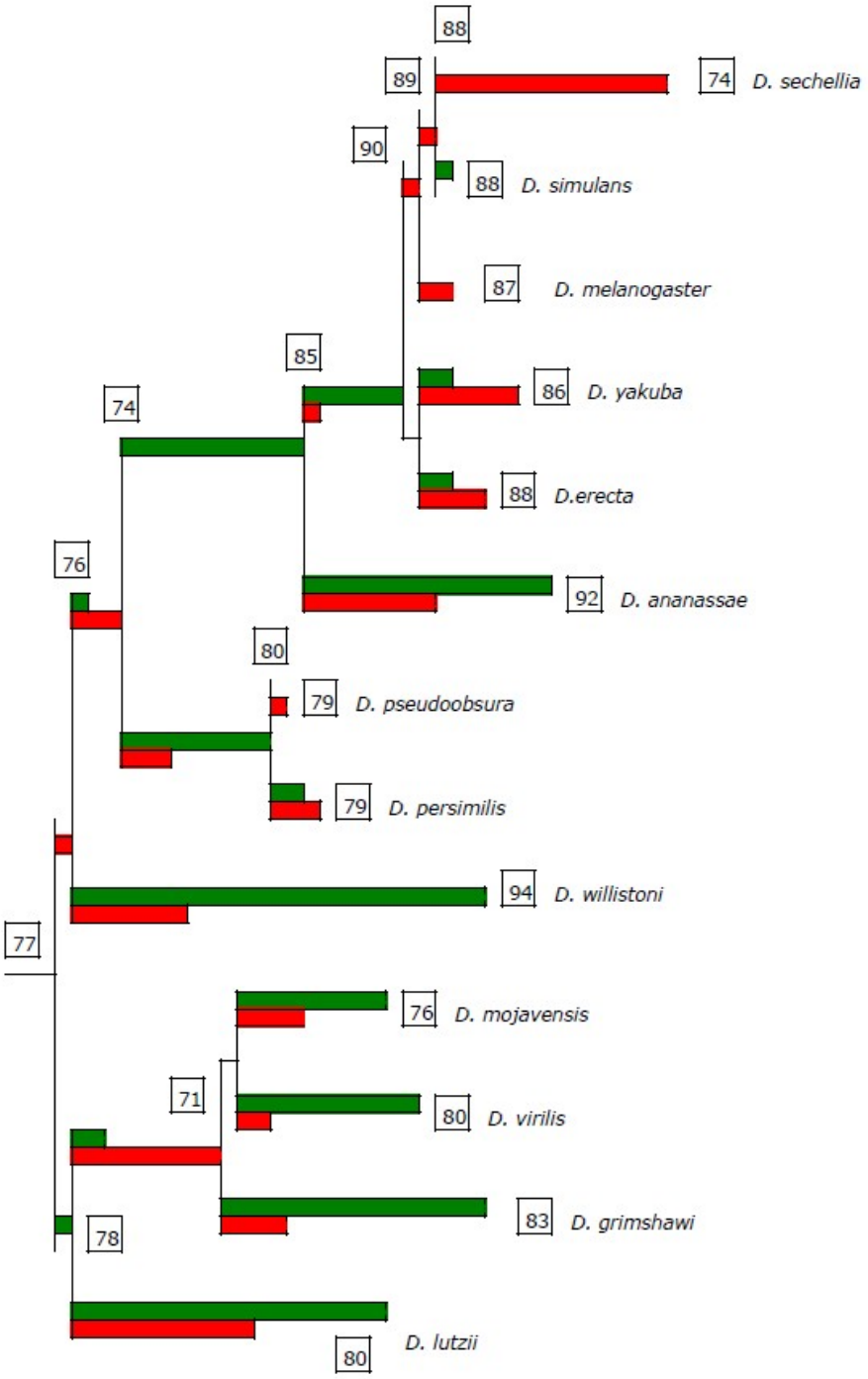


Figure 2. P450 gene duplication and lost across the 13 *Drosophila* species phylogeny. The number of P450 genes found in each genome are shown inside the squares. Green bars represent the gene duplication events and red bars the gene loss events. Modified from Good *et al.* (2014).

Taken together, our data shows that *D. lutzii* genome had a great increase

in P450 genes due to duplications and also some gene loss. This pattern is typically associated to ecological specialization for other species (Goldman-Huertas *et al*., 2015; Yassin *et al*., 2016). Analysis including other species genome associated with the identification of sites under positive selection and of transcription-factor biding sites within the *D. lutzii* P450 genes will improve our results and help to elucidate the P450 genes duplication/loss events in this species.

Acknowledgements

The authors thank Dra. Vera Lucia da Silva Valente Gaiesky for all the support. M.Sc. TuaneLetícia Carvalho for help with DNA extraction; and Zacchary Cohen and the Center for High Throughput Computing Team at the University of Wisconsin-Madison for the bioinformatic support. Funding: CNPq (grant Universal/ CNPq 402447/2016-6), CAPES and FAPERGS.

References

Arab A, Alves MN, Sartoratto A, Ogasawara DC and Trigo JR (2012) Methyl Jasmonate Increases the Tropane Alkaloid Scopolamine and Reduces Natural Herbivory in *Brugmansia suaveolens*: Is Scopolamine Responsible for Plant Resistance? Neotrop Entomol 41:2–8. doi: 10.1007/s13744-011-0001-0

Arab A and Trigo JR (2011) Host Plant Invests in Growth Rather than Chemical Defense When Attacked by a Specialist Herbivore. J Chem Ecol 37:492–495. doi: 10.1007/s10886-011-9955-y

Arain MS, Shakeel M, Elzaki MEA, Farooq M, Hafeez M, Shahid MR, Shah SAH, Khan FZA, Shakeel Q, Salim AMA et al. (2018) Association of detoxification enzymes with butene-fipronil in larvae and adults of *Drosophila melanogaster*. Environ Sci Pollut Res 25:10006–10013. doi: 10.1007/s11356-018-1202-4

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham SON, Prjibelski AD et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477. doi: 10.1089/cmb.2012.0021

Bette Korber (2000) Computational Analysis of HIV Molecular Sequences. In: Learn AGR and GH (ed) HIV Signature and Sequence Variation Analysis. Kluwer Academic Publishers., Dordrecht, Netherlands, pp 55–72

Birney E, Clamp M and Durbin R (2004) GeneWise and Genomewise. Genome Res 14:988–995. doi: 10.1101/gr.1865504

Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, O'Keefe M, McKenzie JA, Batterham P and Daborn PJ (2005) *Cyp12a4* confers lufenuron resistance in a natural population of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 102:12807–12812. doi: 10.1073/pnas.0503709102

Bono JM, Matzkin LM, Castrezana S and Markow TA (2008) Molecular evolution and population genetics of two *Drosophila mettleri* Cytochrome P450 genes involved in host plant utilization. Mol Ecol 17:3211–3221. doi: 10.1111/j.1365-294X.2008.03823.x

Brncic D (1983) Ecology of flower-breeding *Drosophila*. In: Ashburner M, Carson HL and Thompson Jr. T (eds) The Genetics and Biology of *Drosophila* (vol 3d). pp 333–382

Brun A, Cuany A, Le Mouel T, Berge J and Amichot M (1996) Inducibility of the *Drosophila melanogaster* Cytochrome P450 gene, *Cyp6A2*, by phenobarbital in insecticide susceptible or resistant strains. Insect Biochem Mol Biol 26:697–703. doi: 10.1016/S0965-1748(96)00036-7

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9. doi: 10.1186/1471-2105-10-421

Chialvo CHS and Werner T (2018) *Drosophila*, destroying angels, and deathcaps! Oh my! A review of mycotoxin tolerance in the genus *Drosophila*. Front Biol (Beijing) 13:91–102. doi: 10.1007/s11515-018-1487-1

Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P and Daborn PJ (2009) Characterization of *Drosophila melanogaster* Cytochrome P450 genes. Proc Natl Acad Sci U S A 106:5731–5736. doi: 10.1073/pnas.0812141106

Cordeiro J, de Oliveira JHF, Schmitz HJ, Vizentin–Bugoni J, Oliveira JHF de, Schmitz HJ and Bugoni JV- (2020) High niche partitioning promotes highly specialized, modular and non-nested florivore–plant networks across spatial scales and reveals drivers of specialization. Oikos. doi: 10.1111/oik.06866

Danielson PB, Foster JLM, McMahill MM, Smith MK and Fogleman JC (1998) Induction by alkaloids and phenobarbital of family 4 Cytochrome P450s in *Drosophila*: Evidence for involvement in host plant utilization. Mol Gen Genet 259:54–59. doi: 10.1007/s004380050788

Davies L, Williams DR, Aguiar-Santana IA, Pedersen J, Turner PC and Rees HH (2006) Expression and down-regulation of Cytochrome P450 genes of the *Cyp4* family by ecdysteroid agonists in *Spodoptera littoralis* and *Drosophila*

melanogaster. Insect Biochem Mol Biol 36:801–807. doi: 10.1016/j.ibmb.2006.08.001

De Panis DN, Padró J, Furió-Tarí P, Tarazona S, Milla Carmona PS, Soto IM, Dopazo H, Conesa A and Hasson E (2016) Transcriptome modulation during host shift is driven by secondary metabolites in desert *Drosophila*. Mol Ecol 25:4534–4550. doi: 10.1111/mec.13785

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

Dierick HA and Greenspan RJ (2006) Molecular analysis of flies selected for aggressive behavior. Nat Genet 38:1023–1031. doi: 10.1038/ng1864

Dworkin I and Jones CD (2009) Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. Genetics 181:721–736. doi: 10.1534/genetics.108.093419

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

Erion R, King AN, Wu G, Hogenesch JB and Sehgal A (2016) Neural clocks and neuropeptide F/Y regulate circadian gene expression in a peripheral metabolic tissue. Elife 5:1–21. doi: 10.7554/eLife.13552

Etges WJ (2019) Evolutionary genomics of host plant adaptation: insights from *Drosophila*. Curr Opin Insect Sci 36:96–102. doi: 10.1016/j.cois.2019.08.011

Frank MR, Danielson PB and Fogleman JC (1997) Comparison of *Drosophila* Cytochrome P450 metabolism of natural and model substrates. J Insect Physiol 43:953–957. doi: 10.1016/S0022-1910(97)00038-3

Frank MR and Fogleman JC (1992) Involvement of Cytochrome P450 in host-plant utilization by Sonoran Desert *Drosophila*. Proc Natl Acad Sci U S A 89:11998–12002. doi: 10.1073/pnas.89.24.11998

Gilbert LI (2004) Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in *Drosophila melanogaster*. Mol Cell Endocrinol 215:1–10. doi: 10.1016/j.mce.2003.11.003

Giraud M, Unnithan GC, Goff G Le and Feyereisen R (2010) Regulation of Cytochrome P450 expression in *Drosophila*: Genomic insights. Pestic Biochem Physiol 97:115–122. doi: 10.1038/jid.2014.371

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG and Whiteman NK (2015) Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. Proc Natl

Acad Sci U S A 112:3026–3031. doi: 10.1073/pnas.1424656112

Good RT, Gramzow L, Battlay P, Sztal T, Batterham P and Robin C (2014) The molecular evolution of Cytochrome P450 genes within and between *Drosophila* species. Genome Biol Evol 6:1118–1134. doi: 10.1093/gbe/evu083

Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, Casillas S, Ràmia M, Egea R, Negre B et al. (2014) Genomics of ecological adaptation in cactophilic *Drosophila*. Genome Biol Evol 7:349–366. doi: 10.1093/gbe/evu291

Hoang K, Matzkin LM and Bono JM (2015) Transcriptional variation associated with cactus host plant adaptation in *Drosophila mettleri* populations. Mol Ecol 24:5186–5199. doi: 10.1111/mec.13388

Kasuya J, Kaas G and Kitamoto T (2009) Effects of lithium chloride on the gene expression profiles in *Drosophila* heads. Neurosci Res 64:413–420. doi: 10.1016/j.neures.2009.04.015

Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. doi: 10.1093/nar/gkf436

Kumar S, Stecher G, Li M, Knyaz C and Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. doi: 10.1093/molbev/msy096

Le Goff G, Boundy S, Daborn PJ, Yen JL, Sofer L, Lind R, Sabourault C, Madi-Ravazzi L and Ffrench-Constant RH (2003) Microarray analysis of Cytochrome P450 mediated insecticide resistance in *Drosophila*. Insect Biochem Mol Biol 33:701–708. doi: 10.1016/S0965-1748(03)00064-X

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Maïbèche-Coisne M, Jacquin-Joly E, François MC and Nagnan-Le Meillour P (2002) cDNA cloning of biotransformation enzymes belonging to the Cytochrome P450 family in the antennae of the noctuid moth *Mamestra brassicae*. Insect Mol Biol 11:273–281. doi: 10.1046/j.1365-2583.2002.00335.x

Markow TA (2019) Host use and host shifts in *Drosophila*. Curr Opin Insect Sci 31:139–145. doi: 10.1016/j.cois.2019.01.006

McBride CS and Arguello JR (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics 177:1395–1416. doi: 10.1534/genetics.107.078683

Namiki T, Niwa R, Sakudoh T, Shirai KI, Takeuchi H and Kataoka H (2005)

Cytochrome P450 *Cyp307A1*/*Spook*: A regulator for ecdysone synthesis in insects. Biochem Biophys Res Commun 337:367–374. doi: 10.1016/j.bbrc.2005.09.043

Pryszcz LP and Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113. doi: 10.1093/nar/gkw294

Rane R V., Clarke DF, Pearce SL, Zhang G, Hoffmann AA and Oakeshott JG (2019) Detoxification Genes Differ between Cactus-, Fruit-, and Flower-Feeding *Drosophila*. J Hered 110:80–91. doi: 10.1093/jhered/esy058

Soto IM, Carreira VP, Corio C, Padró J, Soto EM and Hasson E (2014) Differences in tolerance to host cactus alkaloids in *Drosophila koepferae* and *D. buzzatii*. PLoS One 9:1–9. doi: 10.1371/journal.pone.0088370

Terhzaz S, Cabrero P, Brinzer RA, Halberg KA, Dow JAT and Davies SA (2015) A novel role of *Drosophila* Cytochrome *P4504e3* in permethrin insecticide tolerance. Insect Biochem Mol Biol 67:38–46. doi: 10.1016/j.ibmb.2015.06.002

Thomas AM, Hui C, South A and McVey M (2013) Common variants of *Drosophila melanogaster Cyp6d2* cause camptothecin sensitivity and synergize with loss of *Brca2*. G3 Genes, Genomes, Genet 3:91–99. doi: 10.1534/g3.112.003996

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V et al. (2019) FlyBase 2.0: The next generation. Nucleic Acids Res 47:D759–D765. doi: 10.1093/nar/gky1003

Xu J, Su X, Bonizzoni M, Zhong D, Li Y, Zhou G, Nguyen H, Tong S, Yan G and Chen XG (2018) Comparative transcriptome analysis and RNA interference reveal *Cyp6A8* and SNPs related to pyrethroid resistance in *Aedes albopictus*. PLoS Negl Trop Dis 12:1–17. doi: 10.1371/journal.pntd.0006828

Yassin A, Debat V, Bastide H, Gidaszewski N, David JR and Pool JE (2016) Recurrent specialization on a toxic fruit in an island *Drosophila* population. Proc Natl Acad Sci 113:4771 LP – 4776. doi: 10.1073/pnas.1522559113

Chapter 4 –Final considerations

1. General Discussion

In the last decade, researchers specialized on Neotropical drosophilids diversity reacted to the lack of knowledge regarding mycophagous and anthophilous *Drosophila* species, and a large effort to acquire information about biological diversity, ecology, genetics have been done since then (Brncic, 1966; 1983; Schmitz and Hofmann, 2005; Schmitz *et al*., 2007; Gottschalk *et al.*, 2009; Robe *et al*., 2013; De Ré*et al*., 2014; 2017; Valer *et al.*, 2016; Grimaldi *et al*., 2016; Machado *et al.*, 2017; Santa-Brígida *et al*., 2019; Fonseca *et al*., 2019; Schmitz and Valente, 2019; Cordeiro *et al*., 2020). However, the scientific community still have reservation regarding the evolutionary relationship and the nomenclature used for this species group, even with a recent phylogenetic hypothesis that places the *D. lutzii* species group [sensu Yassin, 2013] within the *Drosophila* subgenus (Yassin, 2013). Regarding this matter, this study clarifies the phylogenetic relationships of *D. lutzii*, confirming its position within the subgenus *Drosophila* (Yassin, 2013) phylogenetically near *D. pallidipennis* species group and *D. tripunctata* species group, with great branch support.

The physiological constrains that are responsible for feeding and breeding choices, and the niche restriction have been most studied in cactophilic drosophilids (Frank *et al*., 1997; Bono *et al*., 2008; Guillén *et al*., 2014; Soto *et al*., 2014; Hoang *et al.*, 2015; De Panis *et al.*, 2016), and some studies in mycophagous (Chialvo and Werner, 2018) and anthophilous species (De Ré, 2016; Rane *et al*., 2019). Even though, lots of information still are missing to high-quality overview of physiological mechanisms driving resources use. Considering this, we initiated the analysis of P450 genes and were able to identify 80 Cyp genes, one pseudogene and eight potential pseudogenes. From the 80 genes, several duplicated genes are potentially involved in detoxification process of plant-produced chemicals. However, due to the absence of some important P450 genes, the methodology used here to retrieve these genes from the *D. lutzii*

genome might have to be improved.

2. Prospects

After the analysis of the obtained data by this study it is possible to visualize some future research strategies:

- To perform a phylogenomic approach within the *Drosophila* subgenus;
- To perform the analysis of site selection and branch selection for the P450 gene family, focusing on the *D. lutzii* genome;
- To characterize the odorant receptor gene family in *D. lutzii;*
- To characterize the gustatory receptor gene family in *D. lutzii;* and
- Since we have the *D. lutzii* genome with a great sequencing coverage, the mobilome characterization would give us a great insight about the drivers of ecological specialization.

3. Conclusion

The main conclusions of this dissertation are:

- *Drosophila lutzii* species groups was placed within the subgenus *Drosophila* (chapter 2);
- *Drosophila lutzii* species group is closely related to the '*tripunctata* radiation' (chapter 2);
- *Drosophila lutzii* has 80 P450 genes and they are under purifying selection (chapter 3) .

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed 12 Oct 2019

Arab A, Alves MN, Sartoratto A, Ogasawara DC and Trigo JR (2012) Methyl Jasmonate Increases the Tropane Alkaloid Scopolamine and Reduces Natural Herbivory in *Brugmansia suaveolens*: Is Scopolamine Responsible for Plant Resistance? Neotrop Entomol 41:2–8. doi: 10.1007/s13744-011-0001-0

Arab A and Trigo JR (2011) Host Plant Invests in Growth Rather than Chemical Defense When Attacked by a Specialist Herbivore. J Chem Ecol 37:492–495. doi: 10.1007/s10886-011-9955-y

Arain MS, Shakeel M, Elzaki MEA, Farooq M, Hafeez M, Shahid MR, Shah SAH, Khan FZA, Shakeel Q, Salim AMA et al. (2018) Association of detoxification enzymes with butene-fipronil in larvae and adults of *Drosophila melanogaster*. Environ Sci Pollut Res 25:10006–10013. doi: 10.1007/s11356-018-1202-4

Ashburner M (1981) Entomophagous and other Bizarre Drosophilidae. In: Ashburner M, Carson HLN and Thompson Jr. J (eds) The Genetics and Biology of *Drosophila*, 3th ed. Academic Press, London, pp 395–429

Bächli G TaxoDros. https://www.taxodros.uzh.ch/. Accessed 1 Jan 2020

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham SON, Prjibelski AD et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19:455–477. doi: 10.1089/cmb.2012.0021

Bette Korber (2000) Computational Analysis of HIV Molecular Sequences. In: Learn AGR and GH (ed) HIV Signature and Sequence Variation Analysis. Kluwer Academic Publishers., Dordrecht, Netherlands, pp 55–72

Birney E, Clamp M and Durbin R (2004) GeneWise and Genomewise. Genome Res 14:988–995. doi: 10.1101/gr.1865504

Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, O'Keefe M, McKenzie JA, Batterham P and Daborn PJ (2005) *Cyp12a4* confers lufenuron resistance in a natural population of *Drosophila melanogaster*. Proc Natl Acad Sci U S A 102:12807–12812. doi: 10.1073/pnas.0503709102

Bono JM, Matzkin LM, Castrezana S and Markow TA (2008) Molecular evolution and population genetics of two *Drosophila mettleri* Cytochrome P450 genes involved in host plant utilization. Mol Ecol 17:3211–3221. doi: 10.1111/j.1365-294X.2008.03823.x

Brncic D (1983) Ecology of flower-breeding *Drosophila*. In: Ashburner M, Carson HL and Thompson Jr. T (eds) The Genetics and Biology of *Drosophila* (vol 3d). pp 333–382

Brncic D (1966) Ecological and Cytogenetic Studies of *Drosophila flavopilosa*, a

Neotropical Species Living in *Cestrum* Flowers. Evolution (N Y) 20:16–29. doi: 10.2307/2406146

Brown JH and Lomolino M V. (1998) Biogeography. doi: 10.4324/9781315841236

Brun A, Cuany A, Le Mouel T, Berge J and Amichot M (1996) Inducibility of the *Drosophila melanogaster* Cytochrome P450 gene, *Cyp6A2*, by phenobarbital in insecticide susceptible or resistant strains. Insect Biochem Mol Biol 26:697–703. doi: 10.1016/S0965-1748(96)00036-7

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9. doi: 10.1186/1471-2105-10-421

Carson HL (1971) The ecology of *Drosophila* breeding sites. Harold Lyon Arboretum Lecture Number Two edition, University of Hawaii

Chialvo CHS and Werner T (2018) *Drosophila*, destroying angels, and deathcaps! Oh my! A review of mycotoxin tolerance in the genus *Drosophila*. Front Biol (Beijing) 13:91–102. doi: 10.1007/s11515-018-1487-1

Chikhi R and Medvedev P (2014) Informed and automated k-mer size selection for genome assembly. Bioinformatics 30:31–37. doi: 10.1093/bioinformatics/btt310

Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P and Daborn PJ (2009) Characterization of *Drosophila melanogaster* Cytochrome P450 genes. Proc Natl Acad Sci U S A 106:5731–5736. doi: 10.1073/pnas.0812141106

Cordeiro J, de Oliveira JHF, Schmitz HJ, Vizentin–Bugoni J, Oliveira JHF de, Schmitz HJ and Bugoni JV- (2020) High niche partitioning promotes highly specialized, modular and non-nested florivore–plant networks across spatial scales and reveals drivers of specialization. Oikos. doi: 10.1111/oik.06866

Da Lage JL, Kergoat GJ, Maczkowiak F, Silvain JF, Cariou ML and Lachaise D (2007) A phylogeny of Drosophilidae using the *Amyrel* gene: Questioning the *Drosophila melanogaster* species group boundaries. J Zool Syst Evol Res 45:47–63. doi: 10.1111/j.1439-0469.2006.00389.x

Danielson PB, Foster JLM, McMahill MM, Smith MK and Fogleman JC (1998) Induction by alkaloids and phenobarbital of family 4 Cytochrome P450s in *Drosophila*: Evidence for involvement in host plant utilization. Mol Gen Genet 259:54–59. doi: 10.1007/s004380050788

Davies L, Williams DR, Aguiar-Santana IA, Pedersen J, Turner PC and Rees HH (2006) Expression and down-regulation of Cytochrome P450 genes of the *Cyp4* family by ecdysteroid agonists in *Spodoptera littoralis* and *Drosophila melanogaster*. Insect Biochem Mol Biol 36:801–807. doi: 10.1016/j.ibmb.2006.08.001

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

De Ré FC, Robe LJ, Wallau GL and Loreto ELS (2017) Inferring the phylogenetic

position of the *Drosophila flavopilosa* group: Incongruence within and between mitochondrial and nuclear multilocus datasets. J Zool Syst Evol Res 55:208–221. doi: 10.1111/jzs.12170

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

Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO and Bini LM (2013) Darwinian shortfalls in biodiversity conservation. Trends Ecol Evol 28:689–695. doi: 10.1016/j.tree.2013.09.003

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. doi: 10.1093/nar/gkh340

Etges WJ (2019) Evolutionary genomics of host plant adaptation: insights from *Drosophila*. Curr Opin Insect Sci 36:96–102. doi: 10.1016/j.cois.2019.08.011

Ferraz danilo riberio (2014) Atratividade de iscas de origem animal para dípteros muscóides em área de cerrado do sudeste brasileiro, com ênfase na família Calliphoridae. Universidade Estadual de Campinas

Fonseca PM, Moura RD, Wallau GL and Loreto ELS (2019) The mobilome of *Drosophila incompta*, a flower-breeding species: comparison of transposable element landscapes among generalist and specialist flies. Chromosom Res 27:203–219. doi: 10.1007/s10577-019-09609-x

Frank MR, Danielson PB and Fogleman JC (1997) Comparison of *Drosophila* Cytochrome P450 metabolism of natural and model substrates. J Insect Physiol 43:953–957. doi: 10.1016/S0022-1910(97)00038-3

Frank MR and Fogleman JC (1992) Involvement of Cytochrome P450 in host-plant utilization by Sonoran Desert *Drosophila*. Proc Natl Acad Sci U S A 89:11998–12002. doi: 10.1073/pnas.89.24.11998

Gardiner A, Barker D, Butlin RK, Jordan WC and Ritchie MG (2008) *Drosophila* chemoreceptor gene evolution: Selection, specialization and genome size. Mol Ecol 17:1648–1657. doi: 10.1111/j.1365-294X.2008.03713.x

Gilbert LI (2004) Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in *Drosophila melanogaster*. Mol Cell Endocrinol 215:1–10. doi: 10.1016/j.mce.2003.11.003

Giraud M, Unnithan GC, Goff G Le and Feyereisen R (2010) Regulation of Cytochrome P450 expression in *Drosophila*: Genomic insights. Pestic Biochem Physiol 97:115–122. doi: 10.1038/jid.2014.371

Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG and Whiteman NK (2015) Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. Proc Natl Acad Sci U S A 112:3026–3031. doi: 10.1073/pnas.1424656112

Good RT, Gramzow L, Battlay P, Sztal T, Batterham P and Robin C (2014) The molecular evolution of Cytochrome P450 genes within and between *Drosophila*

species. Genome Biol Evol 6:1118–1134. doi: 10.1093/gbe/evu083

Gottschalk MS, Bizzo L, Döge JS, Profes MS, Hofmann PRP and Valente VLS (2009) Drosophilidae (Diptera) associated to fungi: differential use of resources in anthropic and Atlantic Rain Forest areas. Iheringia Série Zool 99:442–448. doi: 10.1590/s0073-47212009000400016

Graham CH, Storch D and Machac A (2018) Phylogenetic scale in ecology and evolution. Glob Ecol Biogeogr 27:175–187. doi: 10.1111/geb.12686

Grimaldi DA (2016) Revision of the *Drosophila bromeliae* Species Group (Diptera: Drosophilidae): Central American, Caribbean, and Andean Species. Am Museum Novit 3859:1–55. doi: 10.1206/3859.1

Grimaldi DA (1990) A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). Bull Am MUSEUM Nat Hist 197:1–139.

Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, Casillas S, Ràmia M, Egea R, Negre B et al. (2014) Genomics of ecological adaptation in cactophilic *Drosophila*. Genome Biol Evol 7:349–366. doi: 10.1093/gbe/evu291

Guindon S and Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol 52:696–704. doi: 10.1080/10635150390235520

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

Hoang K, Matzkin LM and Bono JM (2015) Transcriptional variation associated with cactus host plant adaptation in *Drosophila mettleri* populations. Mol Ecol 24:5186–5199. doi: 10.1111/mec.13388

Kasuya J, Kaas G and Kitamoto T (2009) Effects of lithium chloride on the gene expression profiles in *Drosophila* heads. Neurosci Res 64:413–420. doi: 10.1016/j.neures.2009.04.015

Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. doi: 10.1093/nar/gkf436

Kumar S, Stecher G, Li M, Knyaz C and Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. doi: 10.1093/molbev/msy096

Lachaise D and Tsacas L (1983) Breeding-sites in tropical African Drosophilids. In: Ashburner M, Carson HL and Jr. JNT (eds) The Genetics and Biology of *Drosophila*, 3rd ed. Academic Press, London, pp 221–332

Lanfear R, Frandsen PB, Wright AM, Senfeld T and Calcott B (2017) Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 34:772–773. doi: 10.1093/molbev/msw260

Lang M, Murat S, Clark AG, Gouppil G, Blais C, Matzkin LM, Guittard É, Yoshiyama-Yanagawa T, Kataoka H, Niwa R et al. (2012) Mutations in the neverland Gene Turned *Drosophila pachea* into an Obligate Specialist Species. Science (80) 337:1658–1661.

Letunic I and Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. doi: 10.1093/nar/gkz239

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Ludwig A, Vidal NM, Loreto ELS and Sepel LM. (2002) *Drosophila incompta* development without flowers. Drosoph Inf Serv 85:19–20.

Machado S, Junges dos Santos JP, Fonseca PM, Bolzan AR, David J, Loreto EL da S, Gottschalk MS and Robe LJ (2017) Neotropical mycophagous drosophilids (Diptera: Drosophilidae): DNA barcoding as a way of overcoming the taxonomic impediment. Insect Conserv Divers 10:271–281. doi: 10.1111/icad.12223

Maïbèche-Coisne M, Jacquin-Joly E, François MC and Nagnan-Le Meillour P (2002) cDNA cloning of biotransformation enzymes belonging to the cytochrome P450 family in the antennae of the noctuid moth *Mamestra brassicae*. Insect Mol Biol 11:273–281. doi: 10.1046/j.1365-2583.2002.00335.x

Markow TA (2019) Host use and host shifts in *Drosophila*. Curr Opin Insect Sci 31:139–145. doi: 10.1016/j.cois.2019.01.006

Markow TA and O'Grady P (2008) Reproductive ecology of *Drosophila*. Funct Ecol 22:747–759. doi: 10.1111/j.1365-2435.2008.01457.x

Markow TA and O'Grady PM (2006) Phylogenetic relationships of Drosophilidae. *Drosophila*. doi: 10.1016/b978-012473052-6/50001-9

McBride CS and Arguello JR (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. Genetics 177:1395–1416. doi: 10.1534/genetics.107.078683

Miller MA, Pfeiffer W and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew Comput Environ Work GCE 2010. doi: 10.1109/GCE.2010.5676129

Namiki T, Niwa R, Sakudoh T, Shirai KI, Takeuchi H and Kataoka H (2005) Cytochrome P450 *Cyp307A1*/*Spook*: A regulator for ecdysone synthesis in insects. Biochem Biophys Res Commun 337:367–374. doi: 10.1016/j.bbrc.2005.09.043

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

Okonechnikov K, Golosova O, Fursov M, Varlamov A, Vaskin Y, Efremov I, German Grehov OG, Kandrov D, Rasputin K, Syabro M et al. (2012) Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28:1166–1167. doi: 10.1093/bioinformatics/bts091

Pélandakis M and Solignac M (1993) Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J Mol Evol 37:525–543. doi: 10.1007/BF00160433

Pryszcz LP and Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 44:e113. doi: 10.1093/nar/gkw294

Rane R V., Clarke DF, Pearce SL, Zhang G, Hoffmann AA and Oakeshott JG (2019) Detoxification Genes Differ between Cactus-, Fruit-, and Flower-Feeding *Drosophila*. J Hered 110:80–91. doi: 10.1093/jhered/esy058

Remsen J and O'Grady P (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol Phylogenet Evol 24:249–264. doi: 10.1016/S1055-7903(02)00226-9

Robe LJ, De Ré FC, Ludwig A and Loreto ELS (2013) The *Drosophila flavopilosa* species group (Diptera, Drosophilidae) An array of exciting questions. Fly (Austin) 7:59–69. doi: 10.4161/fly.239523

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

Robe LJ, Valente VLS and Loreto ELS (2010b) Phylogenetic relationships and macro-evolutionary patterns within the *Drosophila tripunctata* "radiation" (Diptera: Drosophilidae). Genetica 138:725–735. doi: 10.1007/s10709-010-9453-0

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. doi: 10.1093/sysbio/sys029

Russo CAM, Mello B, Frazão A and Voloch CM (2013) Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). Zool J Linn Soc 169:765–775. doi: 10.1111/zoj.12062

Santa-Brígida R, Wartchow F, Medeiros PS, Gottschalk MS, Martins MB and de Carvalho CJB (2019) Mycophagous Drosophilidae (Diptera) guild and their hosts in the Brazilian amazon. Pap Avulsos Zool 59:0–4. doi: 10.11606/1807-0205/2019.59.20

Schmitz HJ and Hofmann PRP (2005) First record of subgenus *Ploridosa* of *Drosophila* in southern Brazil, with notes on breeding sites. Drosoph Inf Serv 88:97–101.

Schmitz HJ and Valente VL da S (2019) The flower flies and the unknown diversity of Drosophilidae (Diptera): A biodiversity inventory in the Brazilian fauna. Pap Avulsos Zool 59:0–4. doi: 10.11606/1807-0205/2019.59.45

Schmitz HJ, Valente VLS and Hofmann PRP (2007) Taxonomic survey of Drosophildae (Diptera) from mangrove forests of Santa Catarina Island, Southern Brazil. Neotrop Entomol 36:53–64. doi: 10.1590/S1519-566X2007000100007

Song H and Bucheli SR (2010) Comparison of phylogenetic signal between male genitalia and non-genital characters in insect systematics. Cladistics 26:23–35.

doi: 10.1111/j.1096-0031.2009.00273.x

Soto IM, Carreira VP, Corio C, Padró J, Soto EM and Hasson E (2014) Differences in tolerance to host cactus alkaloids in *Drosophila koepferae* and *D. buzzatii*. PLoS One 9:1–9. doi: 10.1371/journal.pone.0088370

Tanabe AS (2011) Kakusan4 and Aminosan: Two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. Mol Ecol Resour 11:914–921. doi: 10.1111/j.1755-0998.2011.03021.x

Tatarenkov A and Ayala FJ (2001) Phylogenetic relationships among species groups of the *virilis-repleta* radiation of *Drosophila*. Mol Phylogenet Evol 21:327–331. doi: 10.1006/mpev.2001.1002

Thomas AM, Hui C, South A and McVey M (2013) Common variants of *Drosophila melanogaster* Cyp6d2 cause camptothecin sensitivity and synergize with loss of Brca2. G3 Genes, Genomes, Genet 3:91–99. doi: 10.1534/g3.112.003996

Throckmorton LH (1975) The phylogeny, ecology, and geography of *Drosophila*. Handb Genet Vol 3 421–469.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V et al. (2019) FlyBase 2.0: The next generation. Nucleic Acids Res 47:D759–D765. doi: 10.1093/nar/gky1003

Valer FB, Bernardi E, Mendes MF, Blauth ML and Gottschalk MS (2016) Diversity and associations between Drosophilidae (Diptera) species and basidiomycetes in a neotropical forest. An Acad Bras Cienc 88:705–718. doi: 10.1590/0001-3765201620150366

Van Der Linde K and Houle D (2008) A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). Insect Syst Evol 39:241–267. doi: 10.1163/187631208788784237

Vilela CR and Prieto D (2018) A new Costa Rican species of *Drosophila* visiting inflorescences of the hemi-epiphytic climber *Monstera lentii* (Araceae). Rev Bras Entomol 62:225–231. doi: 10.1016/j.rbe.2018.06.002

Wang X victory, Blades N, Ding J, Sultana R and Parmigiani G (2012) Estimation of sequencing error rates in short reads. BMC Bioinformatics 13:1–12. doi: 10.1186/1471-2105-13-185

Xu J, Su X, Bonizzoni M, Zhong D, Li Y, Zhou G, Nguyen H, Tong S, Yan G and Chen XG (2018) Comparative transcriptome analysis and RNA interference reveal *Cyp6A8* and SNPs related to pyrethroid resistance in *Aedes albopictus*. PLoS Negl Trop Dis 12:1–17. doi: 10.1371/journal.pntd.0006828

Yamaguchi M and Yoshida H (2018) *Drosophila* as a Model Organism. In: Yamaguchi M (ed) *Drosophila* models for human diseases, 1s ed. Springer, Singapore, pp 1–11

Yassin A (2013) Phylogenetic classification of the Drosophilidae Rondani (Diptera): The role of morphology in the postgenomic era. Syst Entomol 38:349–

364. doi: 10.1111/j.1365-3113.2012.00665.x

Yassin A, Debat V, Bastide H, Gidaszewski N, David JR and Pool JE (2016) Recurrent specialization on a toxic fruit in an island *Drosophila* population. Proc Natl Acad Sci 113:4771 LP – 4776. doi: 10.1073/pnas.1522559113