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

GENOMIC SEARCH OF TRANSPOSABLE ELEMENTS AND THEIR IMPLICATIONS FOR THE VARIABILITY OF PEST SPECIES OF FRUIT CULTURE

Paula Dandara da Silva Berrutti

Master Dissertation submitted to the Graduate Program in Genetics and Molecular Biology of UFRGS as a partial requirement to obtain a Master's Degree in Biological Sciences.

Advisor: Porfª Drª Vera L. S. Valente Gaiesky

Co-advisor: Profª Drª Maríndia Deprá

Porto Alegre, 2018

This work was performed in the Drosophila laboratory of the Department of Genetics, Federal University of Rio Grande do Sul, with the support of CNPq resources.

Acknowledgments

À minha mãe, Andréia Luciana da Silva, que dedicou sua vida para que eu tivesse base para ir atrás dos meus sonhos.

À minha orientadora “Profi” Vera, por me ensinar os valores de ética e de humildade, e pela confiança depositada em mim.

À minha co-orientadora Maríndia, pela eterna paciência em acompanhar todos os ritmos deste trabalho e por tantas “luzes” que me deu.

À minha ex co-orientadora Carol, por me mostrar que o caminho da ciência exige dedicação, comprometimento e pragmatismo.

Aos colegas do lab, Natasha, Gilberto, Rebeca, Pâmela, Heleninha, Berê e Dani, por me auxiliarem em dúvidas mesmo que triviais sobre a ciência e a vida, e por estarem sempre prontos para um café e uma comemoração.

À Brenda, por ter me convidado para conhecer “o tal laboratório das moscas” em 2012.

Aos colegas do programa da Pós-Graduação em Genética e Biologia Molecular, em especial à Lílian e ao Elmo, por todo auxílio e atenção que sempre me deram.

Aos Professores da Pós-Graduação em Genética e Biologia Molecular pelo conhecimento transmitido.

Ao CNPq pelo auxílio financeiro, e pela bolsa de Mestrado.

Aos meus eternos amigos David, Clara, Aline, Alice, Sunamita, Thiago, Joseane, Valquíria, Bruna, Thalita, e Anselmo por me ouvirem e estarem do meu lado nos melhores e nos piores momentos durante essa jornada.

A todos vocês eu agradeço por fazerem parte do meu caminho profissional e pessoal, muito obrigada, eu amo vocês!

Eppur si muove

Galileo Galilei

SUMMARY

Abbreviation, symbols and units

bp base pair
CYP cytochrome P450 gene
DNA deoxyribonucleic acid
ORF open reading frame
PCR polymerase chain reaction
RNA ribonucleic acid
TE transposable element
TF transcription factor
TFBS transcription factor-binding site
SWD Spotted Wing *Drosophila*
UTR untranslated region

CHAPTER 1

INTRODUCTION

THE *DROSOPHILA* GENUS AND *DROSOPHILA SUZUKII*

The Drosophilidae family (suborder Brachycera, order Diptera, class Insecta) had its origin in the tropical regions, about 50 million years ago (Throckmorton, 1975). The most studied genus of this family is *Drosophila*, with more than 4,000 species already described (Bächli, 2012).

Drosophilids are considered the most successful and widely distributed Diptera. Several species of *Drosophila* have been studied extensively in both Genetics and Developmental Biology; which has contributed to understand several cellular and biological processes with applications, such as agriculture and medical sciences (Jennings 2011; Birney 2007), and has been the source of crucial insights in many biological processes. Therefore, because it is an organism relatively easy to collect in nature, and easy to maintain in the laboratory, besides being low cost, the *Drosophila* genus has served as a fundamental model in Genetics for over a century (Morgan 1910; Sturtevant 1913), being *Drosophila melanogaster* Meigen its representative most famous. *Accordi*ng to Pavan (1959), no other animal beyond man has been the target of so many studies as this fly.

One of the many advantages of using *Drosophila* as a model organism has been its harmless relationship to our own species, allowing the establishment of transgenic lineages and stocks that can be widely used without the risk of compromising human efforts or the natural environment (Ashburner *et al*., 2004). The conflict with humans was highlighted, however, after some of Drosophilidae species has been emerged as pest species. Most of the knowledge derived from studies with *Drosophila* was not transferred to applied entomological problems, since it reproduces and feeds on decaying fruits, it has rarely been considered an economically important pest. In a few decades, however, for the first time,

agricultural invasion pests are so closely related to members of the Drosophilidae family.

Studies with invasive species allow us to understand how the invaders react to new biotic and abiotic conditions, and how native species react to invasion (Silva *et al*., 2005). Bioinvasions are characterized by the intentional introduction or not of exotic species. *Accord*ing to Lincoln *et al*. (1998), exotic species are non-native organisms that were introduced within an area. Regardless of the process by which invasions take place, several consequences are possible; (Lodge 1993) and even the extinction of native species (Fritts & Rodda 1998). In addition, introduced species can bring public health risks (Ruiz *et al*., 2000) and damage to agriculture (Pimentel *et al*., 2001). The number of species transported, even unintentionally, by human action, breaking geographic barriers for example, is enormous. However, only a fraction of these species are able to establish themselves in a new territory and, among them, generally 1% has the potential to become a pest (Silva *et al*., 2005).

Usually associated with the popular name "fruit fly", members of the Drosophilidae family, however, do not feed on the fruits, but on the yeasts that grow in decaying organic matter (Carson, 1971). They present a wide diversity of ecological niches, as well as variation in the pattern of geographical distribution. In general, they are primary consumers of microorganisms, yeasts and bacteria, associated with the early stages of plant decomposition. For this reason they are not considered pest species. However, some species have already demonstrated their invasive potential as *Drosophila melanogaster* (David & Capy 1988), *Drosophila suboobscura* (Ayala et al., 1989), *Drosophila simulans* (Hamblin & Veuille 1999), *Drosophila malerkotliana* (Vogl *et al*., 2003), *Drosophila ananassae* (Val & Sene 1980), and *Zaprionus indianus* (Vilela*, 1999*). Several studies have proposed that the dispersal process of Drosophilidae species is directly related to anthropophilic actions (Tidon *et al*. 2003; Galego & Carareto 2007; Garcia *et al*. 2008; Yassin *et al*. 2008; Galego & Carareto 2010; Garcia *et al*. 2012).

Recently, a species of the Drosophilidae family, called *Drosophila suzukii* Matsumura, has emerged as a pest in several countries where it occurs. It is able to develop in a very wide range of soft-skinned fruits, both of cultivation and in wild

fruits of many native host plants in the invaded areas. This species was described in 1931, but the earliest records date from 1916, so little is known of its origin, whether it is native to Japan or if was introduced in the country (Hauser 2011).

Undeniably, *D. suzukii* has a high dispersion potential: it has expanded widely in Asia, and from there to Europe (Cini *et al*., 2012; Rota-Stabelli *et al*., 2013), North America (Kaneshiro 1983; Leblanc et al. 2009), and Central America (Walsh *et al*. 2011; Asplen *et al*. 2015; Lee *et al*. 2015). The first occurrence of this species in South America was verified by our research group in 2013 (Deprá *et al*., 2014) in southern Brazil, where it caused significant economic losses in orchards, especially in red fruits that seem to be their "preference" such as blackberry, cherry, raspberry, blueberry and strawberry (Goodhue *et al*. 2011; Bellamy *et al*. 2013; Santos 2014; De Ros *et al*. 2015; Ioriatti *et al*. 2015; Lee *et al*. 2015). Furtherly, the species was detected in several sites, dispersing to other regions of the country and even neighboring countries (Vilela & Mori 2014; Paula *et al.* 2014; Bitner-Mathé *et al.* 2014; Gonzaléz *et al*., 2015; Schlesener *et al*. 2015).

The ability of *D. suzukii* ovopositing its eggs into healthy fruits can lead to direct loss of yields with reductions up to 80% in some countries (Dreves *et al*. 2009; Walsh *et al*. 2010; Hauser 2011), and 100% of the ecologically grown cherries (Escudero *et al*., 2012). Its ability to grow in tomato under laboratory conditions has also been demonstrated (Cini *et al*., 2012). It has also recently been reported that *D. suzukii* has caused economic damage and significant losses in strawberry crops in southern Brazil (Santos 2014). Ecological differences in relation to most species of *Drosophila* reflect adaptations that allow their wide dispersion and can justify their success in the invasion of new habitats.

The *suzukii* subgroup, the same of the *Drosophila melanogaster* group, frequently exhibits sexual dimorphism in the color of the wings. This characteristic in males lead *D. suzukii* popularly be called Spotted Wing *Drosophila* (SWD) (Figure 1-A). Fruit damage is caused by females that have a serrated ovipositor with the ability to lay eggs within mature and healthy fruit (Walsh *et al.* 2011; Cini *et al*. 2012; Lee *et al*. 2015) (Figure 1-B). Injury caused by external piercing and/or oviposition allow pathogens to penetrate, increasing economic losses (Dreves *et al*. 2009; Bolda *et al*. 2010), as well as promoting the release of volatile products

(Abraham *et al*. 2015) which attract other pest species such as *Zaprionus indianus* (Timmeren & Isaacs 2013; Joshi *et al*. 2014; Lasa & Tadeo 2015). This ability of *Z. indianus* females to oviposite in healthy mature strawberries, to breed offspring and to benefit from injuries caused by *D. suzukii* or mechanical lesions may be associated with the attraction of these species to the odors released by ripe fruits of the "berries", as observed in adults *D. suzukii* (Ramniwas *et al*., 2012). Thus, this possible association of the mode of action of these two species of invasive Drosophilidae can contribute significantly to the increased incidence of *Z. indianus* in strawberry commercial fields (Bernardi *et al*., 2016) in grape orchards in the United States (Timmeren & Isaacs 2013), in sweet orange and guava crops in India (Fartyal *et al*., 2014) and in Mexico (Lasa & Tadeo 2015), and araçá, pitanga and guava in southern Brazil (Andreazza *et al*., 2015).

According to Lee et al. (2011) fruits may become susceptible to *D. suzukii* when they begin to change color. After the establishment of the fly, eradication is very difficult and the cost production increases permanently due to the need for monitoring, management, increased use of chemical products and secondary selection of fruits. As a result, some projects are underway: in the United States, a consortium of universities and institutions funded by the US Department of Agriculture has been in place since 2010 to monitor and control the spread of the fly. In Europe, several institutions are monitoring the species and there are proposals for monitoring and studying *D. suzukii* at the continental level (Rota-Stabelli *et al*. 2013). The interest in this species stems precisely from the fact that *D. suzukii* is one of the main pests associated with small-fruit farming in worldwide (Walsh *et al*. 2011; Cini *et al*. 2012; Santos 2014; Asplen *et al*. 2015), causing many losses to the fruit growers.

All efforts in the attempt to get to know *D. suzukii* come from the fact that this species is also a potential threat previously described for the biodiversity and ecology of the invaded areas (Dreves 2011; Cini *et al*. 2012; Deprá *et al*. 2014; Poppe *et al*. 2015, dos Santos *et al*., 2017; Fraimout A, *et al*. 2017). This behavior is attributed to its high polyphagia (Dreves *et al*., 2009), rapid population growth (Tochen *et al*., 2014) and dispersion capacity (Walsh *et al*., 2011; Cini *et al*., 2012).

Sequenced genomes have served as a powerful tool for gaining new insights into genetic, developmental, regulatory, and evolutionary processes; as well as helping the biologist to develop, validate and establish several evolutionary models (Ohler *et al*. 2002; Vogl *et al*. 2003; Duque *et al*. 2014). The availability of complete genomic sequences for the 12 species (*Drosophila* 12 Genomes Consortium, 2007) and many species of *Drosophila* sequenced until now (more than 24), allows now to examine the evolutionary diversification of genes in Drosophilidae. *D. suzukii* had its genome sequenced in 2013 (Chiu *et al*. 2013), and preliminary analysis comparison to other species of the *Drosophila melanogaster* group showed some peculiarities of pest species. The expansion of some gene families such as those encoding proteins involved in gustatory and olfactory perception - involved in the detection of stimuli, sensory transduction, endopeptidase inhibitors; metabolic processes of cellular regulation of proteins and glycerol, for example. On the other hand, other families of genes, such as those involved in defense mechanisms and detoxification of substances (esterases and cytochrome P450) appear to have decreased when compared to other species of the *melanogaster* group (Chiu *et al*. 2013).

Due to the high variability in ecological and behavioral strategies present in *Drosophila*, it has been seen that the characterization of genetic factors linked to genes associated with environmental responses, external stimuli (xenobiotic metabolism), with immunological functions and involved in the response to stress are less conserved, contributing to the plasticity of the genome (Chen & Li 2007; Van de Lagemaat *et al*., 2003). However, genes that encode hormone biosynthesis enzymes, transcription factors, and other factors involved in regulation of the development are essential for organism survival and tend to be highly conserved, since mutations would cause lethal effects (Simons *et al*., 2006).

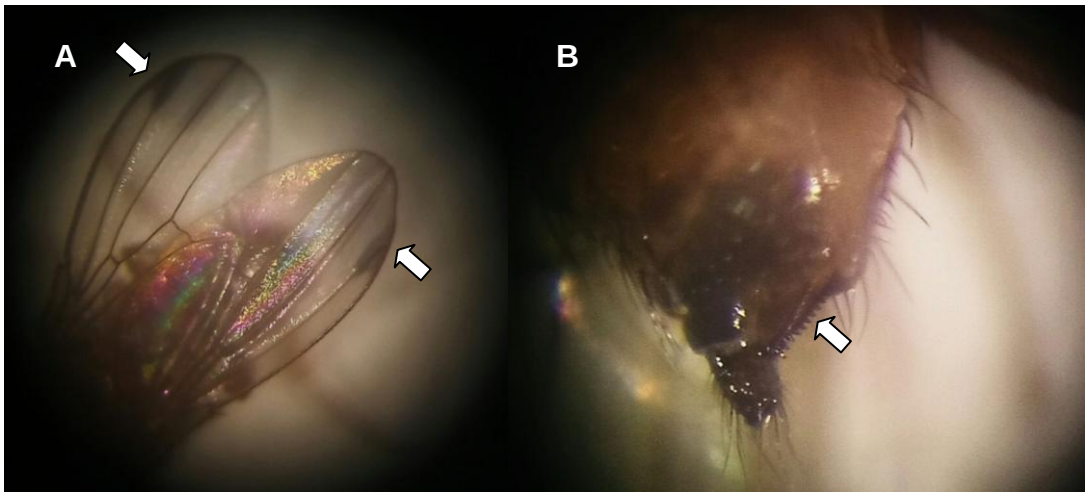

Figure 1. *Drosophila suzukii* (A) dark spot characteristic of male wings, (B) detail of saw-like ovipositor of the female.

THE *CYP* GENES SUPERFAMILY

Cytochromes P450 (CYPs) compose a superfamily of heme-thiolate proteins responsible for the metabolization of a large number of endogenous substrates (endobiotics: steroids, bile acids, fatty acids, prostaglandins, leukotrienes, retinoids and others) and exogenous (xenobiotics: environmental chemicals and natural plant products and others) (Nelson *et al*., 1996). Since its origin, more than 3.5 billion years ago, the CYP family of genes has diversified to modulate the metabolism of a growing number of environmental toxins, dietary compounds and drugs (Nelson *et al*., 1996).

When entering the organism, the xenobiotics undergo changes being the products of its metabolism less toxic, in other words, occurs a bioactivation of the initial chemical compost. Most xenobiotics are lipophilic compounds, and in order to be more easily excreted from the organism these compounds must undergo an enzymatic transformation into metabolites with more hydrophilic characteristics, in a process called biotransformation (Di Giulio *et al*., 1995). The metabolites resulting from this process are usually less active than the initial compound. However, the metabolism of xenobiotics can produce more reactive and toxic products that may be responsible for several forms of toxicity, and the accumulation of these more toxic metabolites promotes damage to the cellular

components, and to the DNA and RNA molecules (Liebler & Guengerich, 2005; Josephy *et al*., 2008), including the beginning and progression of tumors (Nelson *et al*., 1996). In this way, biotransformation usually results in an increased rate of excretion of xenobiotics, which reduces the risk of accumulation of these substances to toxic levels in the body, thus biotransformation of xenobiotics is the main mechanism for maintaining homeostasis during exposure to strange molecules from the body (Klaassen & Watkins, 1999).

The detoxification system is usually the first enzymatic defense against strange compounds. Os toxic compounds are typically non-reactive compounds, and as such do not contain reactive sites that can bind the water soluble groups. Thus, xenobiotics are primarily subjected to activation reactions, where oxidation, reduction or hydrolysis reactions introduce a functional group (-OH, -NH2, -SH or -COOH) transforming them into active substances ready for the conjugation process.

Bioactivations can be catalyzed by various enzymes, such as cytochrome P450 monooxygenases, flavin-containing monooxygenases, hydrolases, lipoxygenases, peroxidases, oxidases and reductases (Nebert, 1991; Klaassen & Watkins, 1999). One of the most important enzymatic systems in the bioactivation consists of the cytochromes P450 (CYP) and its redox partner NADPH oxidoreductase, both in terms of the high number of detoxifying xenobiotics and the catalytic versatility they present (Nebert, 1991, Nelson *et al*., 1996). Most of the metabolism mediated by the CYPs is based on an oxidation-reduction reaction, in which one oxygen atom (derived from O_2) is incorporated into the substrate, and the other atom is reduced to water with the reducing equivalents of NADPH (Klaassen & Watkins, 1999; Guengerich, 2007), as shown in the following reaction:

$$\textbf{Substrate (RH)} + \textbf{O}_2 + \textbf{NADPH} + \textbf{H}^+ \rightarrow \textbf{Product (ROH)} + \textbf{H}_2\textbf{O} + \textbf{NADP}^+$$

After, these molecules are conjugated by the addition of a water-soluble group to the reactive site. The reactions are mediated by several enzymes that may belong to superfamilies of distinct genes, including sulfotransferases, transaminases, acetyltransferases, methyltransferases, acyltransferases,

al*DOC*etoreductases, carboxylesterases, glycosylases, glucuronyltransferases and various hydrolases and esterases (Nebert, 1991). In conjugated metabolites there is normally an increase in hydrophilicity and as such these compounds are rapidly excreted (Meyer, 1996).

The ability of cells to oxidize hydrophobic exogenous compounds (detoxification) was already appreciated since the end of 19th century, although the enzymes responsible for this reaction were not known. In 1955, Williams and Klingenberg identified in mouse liver microsomes a pigment with specific spectrophotometric characteristics (review in Nebert & Gonzalez, 1987). Omura and Sato (1964) characterized this pigment as a hemoprotein that presented in its differential spectrum a Soret peak at 450 nm when complexed with carbon monoxide and designated cytochrome P450 (CYP). But it was in the studies conducted by Cooper et al. (1965) that they demonstrated the enzymatic function of CYPs and their importance in the metabolism of xenobiotics.

In this superfamily of enzymes, at least 21,000 named P450 sequences are known (Nebert 2005), which are distributed between plants (7,446 sequences), animals (6,313 sequences), fungi (5,729 sequences), bacteria (1,254 sequences), protozoa (247 sequences), archaea (48 sequences), and viruses (2 sequences). The wide variety of isoforms of these proteins needed the development of a universal nomenclature for the CYP superfamily, based on the comparison of the amino acid sequences and the evolutionary relationships of the corresponding genes based on a divergent evolution of this superfamily. Thus, to designate a cytochrome P450 gene is first included the acronym "CYP". Thereafter, the CYP enzymes within the same family are designated by a number and share more than 40% identity in the amino acid sequence. The families are then divided into subfamilies, the enzymes within the same subfamily being designated by the same letter. Genes within the same subfamily share more than 55% identity in their amino acid sequence. Finally, a number after the letter denotes each individual isoenzyme, differing by about 3% (Nebert *et al*., 1987; Nelson *et al*., 1996; review in Hemingway & Ranson, 2000).

Since CYPs are considered unique in the metabolic system of insects, and can also mediate resistance to all classes of insecticides (Feyereisen 2005, Li *et*

al., 2007), It has been observed more than 25 *CYP* genes of the families *CYP3, CYP4, CYP6, CYP9,* and *CYP12* related to insect resistance to insecticides (Tijet *et al*., 2001; Ranson *et al*., 2002). In all reported cases, it was observed overexpression of these enzymes in resistant insects (Li *et al*., 2007). In agriculture, to limit damage caused by pests such as *D. suzukii* populations are mostly suppressed with the use of pesticides. However, this can cause environmental and health problems because there are a high risk of chemical residues remaining in fruit, since the treatments are performed near harvest. In addition, several studies conducted with *Drosophila* species have associated insecticide resistance to overexpression of CYPs genes as a result of the insertion of transposable element (TE) fragments into their regulatory regions or even within the genes. These sequences of TEs may affect the expression of adjacent genes by introducing regulatory binding sites in flanking regions of the gene (Conte *et al.* 2002; Jordan *et al.* 2003; Kunarso *et al.* 2010; Molineris *et al.* 2011; Thornburg *et al.* 2006; Wang *et al.* 2007 e 2009).

Daborn (2002) reported that insecticidal resistance of the DDT-R locus in *Drosophila melanogaster* is due to overexpression of the *CYPGg1* gene. This overexpression is characterized by the insertion of the *ACCORD* retrotransposon fragment upstream of the gene (Catania *et al*., 2004). Chung *et al*. (2007) observed that this TE carries regulatory sequences, altering the spatial expression of the gene. Schlenke & Begun (2004) reported an association between TE insertion and resistance in *Drosophila simulans*. In this species, the *DOC* element inserted in the flanking region of the ortholog *CYPGg1* promotes its overexpression. Marsano *et al*. (2005) and Bogwitz *et al*. (2005), in turn, suggest that the presence of *Bari-1* transposon at the end of the 3' region of *CYP12a4* in *Drosophila melanogaster* increases the gene expression.

Chen & Li (2007) analyzed TEs in 13 *CYP*s in the *Drosophila melanogaster* species, eight of them associated with resistance and five involved in ecdysone biosynthesis and development regulation. Seven of eight resistance-associated CYP genes contained TEs inserted and none of these genes were associated with development. The authors hypothesize that TEs can be selectively enriched near genes in response to the environmental, but excluded from essential genes

(housekeeping), resulting in a great genomic plasticity. These results reveal an array of genomic events that may be associated with ecological adaptations of the species.

THE TRANSPOSABLE ELEMENTS AND THE HOST GENOME

Until the first half of the 20th century, science had the genome as a static entity, changing only on an evolutionary scale. The revolutionary idea that genomes possess DNA mobile sequences was conceptualized for the first time by Barbara McClintock before the discovery of the structure of DNA (McClintock, 1957). Today we know about the existence of previously unimaginable factors that are capable of generating genetic variability from one generation to another. Transposable Elements (TEs) were discovered in maize (*Zea mays*) by McClintock in the 1940s, and were initially described as duplicate segments, chromosomal modifications, chromosomal aberrations, transposition events, until they were called, in 1956, by transposable elements. TEs comprise a group of repetitive DNA sequences that have the intrinsic ability, or not, to change their location within the genomes. With the development of molecular biology techniques, in the 1980s TEs were rediscovered mobilizing in the genome of *Escherichia coli*, associated with mutations in *D. melanogaster* and maize (review in Varani *et al*., 2015). Since then, TEs have been found in all branches in the tree of life, from simpler organisms as bacteria and fungi, to more complex organisms such as invertebrates, plants, and vertebrates (Wicker *et al*., 2007; Pritham, 2009). However, some exceptions were found, restricted to unicellular species studied, as in the genome of red algae *Cyanidioschyzon merolae*, six species of *Apicomplexa*, and one species of *Unikont*, *Encephalitozoon cuniculi* (Pritham, 2009).

Due to mobilization and parasitic characteristics, TEs became the most abundant and ubiquitous sequences in nature (Aziz *et al*., 2010). Their high prevalence and distribution suggest that these genomic parasites can directly influence the evolution of host organisms that they parasitize, for example in the development of their immune systems (Kapitonov *et al*., 2005) and in the

dynamics of the chromosomes (Langdon *et al*. 2000). Some of these modifications are associated with events of molecular domestication, where copies of TEs play important roles in the genome of the organism. However, due to their mechanisms of replication and transposition, they can trigger modifications in the host organism as mutations, deletions, insertions, duplication, chromosomal rearrangements, a probable reproductive isolation and horizontal transmission system of genetic information between species (Kidwell & Lisch 2001), producing positive, negative or neutral effects in the host organism (Capy *et al*., 1998).

These elements are divided into groups that share common aspects of structure and transposition mechanisms. In the classification suggested by Wicker *et al*. (2007), hierarchically, the classification levels are: class, subclass, order, superfamily, family, and subfamily. The class level divides the TEs by the presence (class I) or absence (class II) of an RNA transposition intermediate. Class I elements (retrotransposons) transpose itself via an RNA intermediate, this copy is reversely transcribed to DNA by a reverse transcriptase encoded by the element. In this way, each replication cycle produces a new copy of the element. Class II, DNA transposons properly, has two subclasses that are distinguished by the number of DNA strands that are cleaved during the transposition process. Subclass 1 comprises the "cut-and-past" TEs, characterized by their terminal inverted repeats (TIRs). The transposition is mediated by the enzyme transposase which recognizes the TIRs and cleaves both strands. Subclass 2 comprises the “copy-and-paste” TEs, where the transposition process cleaves only one of the DNA strands. Both classes are subdivided into superfamilies based on structural characteristics, internal organization, size of duplication of the target site generated at the insertion, and sequence similarities at the DNA and protein level.

The TEs are also classified as autonomous and non-autonomous. The autonomous elements are those that encode all the sequences that enable their transposition, such as the transposase. Non-autonomous elements are structurally deficient in some aspects and depending on proteins produced by other elements in the genome to move. Many non-autonomous elements are derived from autonomous elements that have undergone deletions of some parts of their structures (Kidwell, 2005).

Repetitive sequences of mobile elements are particularly dynamic components of eukaryotic genomes. The transposition mechanism used by TEs is a recombination reaction that mediates the movement of these DNA segments between non-homologous sites. Thus, once they are mobile elements, TEs have the ability to change host genetic information by changing the structure of the chromosomes or the organization of the genes (Craig *et al*., 2002). In general, TEs can influence the evolutionary trajectory of their hosts in three different ways: (1) altering the function of a gene through its insertion, (2) through chromosomal rearrangements, and (3) as a source of coding or non-coding material that allows for the emergence of genetic novelties such as new genes and regulatory sequences (Feschotte & Pritham, 2007).

There is cumulative evidence suggesting that mechanisms of mutation played an important role in reformulating the cis-regulatory content of animal genomes (Maeso & Tena 2015), it was estimated that about 50-80% of *Drosophila* mutations result from the insertion of TEs (Biémont & Vieira 2006). However, the acquisition of TE in the regulatory region, can be advantageous, since it creates a new regulatory pattern by adding regulatory sequences of the TE introduced (Jordan *et al.* 2003; Pereira *et al.* 2009; Pooma *et al.* 2002; Erwin & Davidson 2009). In addition, the expression of TE depends on cis-regulatory and trans-acting elements in the host genome and consequently, changes in cis-regulatory elements are important for the determination of phenotypic differences (Bourque *et al*., 2008), such as polyadenylation, promoters, enhancers and silencers (Thornburg *et al*., 2006).

TEs are transcribed in sense and antisense orientation, and are involved in the regulation of transcription through interfering RNA (Brennecke *et al.* 2007; Girard & Hannon 2008). Deprá *et al*. (2009) described transcription expression pattern for the transposable elements canonical *hobo* and *hobo*^{VAHS} that were similar to that of developmental genes in the first larval stage, being in the later stage expressed in the central nervous system. This pattern suggests that TEs may have cis-regulatory sequences that are recognized by transcription factors of developmental genes.

With the development of nucleic acid sequencing techniques and the sequencing of the first prokaryotic and eukaryotic genomes, it was possible to observe that these elements may constitute a large part of the genome of some organisms, representing a certain 77% of the genus of the frog *Rana esculenta*, 0,3% of the *Escherichia coli* bacterium, 85% of the maize genome, and reaching 50% of the genome of the primates (Lander *et al*. 2001; Mikkelsen *et al*. 2005; Biémont & Vieira 2006; Rhesus Macaque Genome Sequencing and Analysis Consortium 2007; Schnable *et al*. 2009). In *Drosophila* the amount of transposable elements is variable, representing about 2.7% of the genomes of *D. simulans* and *Drosophila grimshawi* up to 25% in the genome of *Drosophila ananassae* (*Drosophila* 12 Genomes Consortium, 2007).

D. grimshawi, for example, has lower repetitive content/transposable elements (~ 2.7%) and this is possibly related to its ecological status: endemic to an island; which may minimize the chance for horizontal transfer of TEs families (*Drosophila* 12 Genomes Consortium 2007). Regarding the genome of *D. suzukii*, a lower content of TEs were observed - 4.9% of the total genome size (Chiu *et al*., 2013). This is intriguing and of great scientific interest to understand this low amount of TEs described for this pest species, since genetic diversity is often associated with adaptability to different niches.

Beyond the deleterious mutations, there are also cases where the insertion of transposable elements near the *CYP* genes in *D. melanogaster* and *D. simulans* led to resistence phenotype, reinforcing the idea that, while TEs in coding regions can have deleterious effects and are removed by purifying selection (Lipatov *et al.* 2005; Sela *et al.* 2007; Yang & Barbash 2008), in regulatory sequences TEs are better tolerated and may be playing an important role in the adaptation of *D. suzukii* in so many continents and substrates.

AIMS AND OBJECTIVES

It will be of academic interest and of applied importance to examine the consequences of the insertions of TEs in relation to the pest species *D. suzukii*, as

our research group since the 1990s has been carrying out several studies to understand the transposons. Besides that, based on the highly invasive nature of this species of fly and its economic importance, genomic studies can provide information for the identification of the genes responsible for adaptation to different ecological and climatic conditions.

In this work, two strategies were used to investigate this issue. The first was to investigate *in silico*, the occurrence of preferential insertions of TEs in *CYP*s genes and, thus, to infer their possible relation with the origin of resistance to insecticides in species of *Drosophila melanogaster* and *Drosophila suzukii*. The second was to examine the genomic content of transposable elements and to evaluate the possible consequences of TEs insertions in the pest species *Drosophila suzukii*.

REFERENCES

Abraham J, Zhang A, Angeli S, Abubeker S, Michel C, Feng Y and Rodriguez-Saona C (2015) Behavioral and antennal responses of *Drosophila suzukii* (Diptera: Drosophilidae) to volatiles from fruit extracts. Environ Entomol 44:356-367.

Andreazza F, Bernardi D, Botton M and Nava DE (2015) Índice de infestação natural de *Drosophila suzukii* e *Zaprionus indianus* (Diptera: Drosophilidae) em frutíferas nativas no município de Pelotas. In: XXIV Congresso de Iniciação Científica e XVII Encontro da Pós- Graduação, Pelotas. Anais Pelotas: Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul, Brasil, 4p.

Ashburner M, Golic K and Hawley RS (2004) *Drosophila*: a laboratory handbook. 2nd ed. New York, NY: Cold Spring Harbor Laboratory Press.

Asplen MK, Anfora G, Biondi A, Choi D-S, Chu D, Daane KM, Gibert P, Gutierrez AP, Hoelmer KA, Hutchison WD, *et al.* (2015) Invasion biology of spotted

wing *Drosophila* (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sc (2004)i 88:469-494.

Aziz RK, Breitbart M, Edwards RA (2010) Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Research 38(13):4207-4217.

Ayala FR, Serra LL and Prevosti A (1989) A grand experiment in evolution: the *Drosophila subobscura* colonization of the America. Genome 31:246-255.

Bächli G (2012) Taxodros: The database on Taxonomy of Drosophilidae. http://www.taxodros.uzh.ch.

Bellamy DE, Sisterson MS and Walse SS (2013) Quantifying host potentials: indexing postharvest fresh fruits for spotted wing *Drosophila*, *Drosophila suzukii*. PLoS One 8:e61227.

Bernardi D, Andreazza F, Botton M, Baronio CA and Nava DE (2016) Susceptibility and Interactions of *Drosophila suzukii* and *Zaprionus indianus* (Diptera: Drosophilidae) in Damaging Strawberry. Neotrop Entomol 58:1-7.

Biémont C and Vieira C (2006) Junk DNA as an evolutionary force. Nat Genet 443:521-524.

Birney E (2007) Evolutionary genomics: come fly with us. Nature 450(8):184-185. DOI: 10.1038/450184a.

Bitner-Mathé BC, Victorino J and Faria FS (2014) *Drosophila suzukii* has been found in tropical Atlantic rainforest in southeastern Brazil. Drosoph Inf Serv 97:136-137.

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 USA 102:12807-12812.

Bolda MP, Goodhue RE and Zalom FG (2010) Spotted wing *Drosophila*: potential economic impact of newly established pest. Agricutural Resource Economics. Update, University of California Gianni Foundation of Agricultural Economics 13:5-8.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng HH, *et al.* (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18:1752-1762.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R and Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell 128:1089-1103.

Capy P, Bazin C, Higuet D, Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience Austin Texas 197pp.

Carson HL (1971) The ecology of *Drosophila* breeding sites. Honolulu: Harold L. Lyon Arboretum Lecture, 2, 1-27.

Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH and Schlotterer C (2004) World-wide survey of an *ACCORD* insertion and its association with DDT resistance in *Drosophila melanogaster*. Mol Ecol 13:2491-2504.

Cini A, Ioriatti C and Anfora G (2012) A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. Bulletin of Insectology 65:149-160.

Chen S and Li X (2007) Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome *P450* genes. BMC Evol Biol 7:46.

Chiu JC, Jiang X, Zhao L, Hamm CA, Cridland JM, Saelao P, Hamby KA, Lee EK, Kwok RS, Zhang G, Zalom FG, Walton VM, Begun DJ (2013) Genome of *Drosophila suzukii*, the spotted wing *Drosophila*. G3 (Bethesda) 3(12):2257-71.

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P and Daborn PJ (2007) Cis-regulatory elements in the *ACCORD* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. Genetics 175:1071-1077.

Conte C, Dastugue B and Vaury C (2002) Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes. Mol Cell Biol 22:1767-1777.

Cooper CW, Yates CW, Talmage RV (1965). Some endogenous parathyroid hormone effects manifested by bone *in vitro*. Proc Sot Ezptl Biol Med 119:81-88.

Craig NL, Graigie R, Gellert M, Lambowitz AM (2002) Mobile DNA II. American Society for Microbiology Press, Whashington, DC, 1204pp.

David JR and Capy P. (1988) Genetic variation of *Drososphila melanogaster* natural populations. Trends Genet 4:106-111.

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, *et al.* (2002) A single P450 allele associated with insecticide resistance in *Drosophila*. Science 297:2253-2256.

Deprá M, Valente VLS, Margis R and Loreto ELS (2009) The *hobo* transposon and *hobo*-related elements are expressed as developmental genes in *Drosophila*. Gene 448:57-63.

Deprá M, Poppe JL, Schmitz HJ, De Toni DC and Valente VLS (2014) The first records of the invasive pest *Drosophila suzukii* in the South American continent. Journal of Pest Science 87(3):379-383.

De Ros G, Conci S, Pantezzi T and Savini G (2015) The economic impact of invasive pest *Drosophila suzukii* on berry production in the Province of Trento, Italy. J Berry Res 5:89–96.

Di Giulio RT, Benson WH, Sanders BM, Van Veld PA (1995) Biochemical Mechanisms: Metabolism, Adaptation, and Toxicity, in Fundamentals of Aquatic Toxicology – Effects, Environmental Fate, and Risk Assessment (ed. G. Pand). London: Taylor & Francis.

Dos Santos LA, Mendes MF, Krüger AP, Blauth ML, Gottschalk MS, Garcia FRM (2017) Global potential distribuition of Drosophila suzukii (Dipter, Drosophilidae). Plos One doi:10.1371/journal.pone.0174318.

Dreves AJ (2011) IPM program development for an invasive pest: coordination, outreach and evaluation. Pest Manag Sci 67(11):1403-10.

Dreves AJ, Walton V, Fisher G (2009) A new pest attacking healthy ripening fruit in Oregon. OSU Extension Service. http://ir.library.oregonstate.edu/jspui/bitstream/1957/13090/1/em8991.pdf.

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203-218.

Duque T, Samee MAH, Kazemian M, Pham HN, Brodsky MH and Sinha S (2014) Simulations of enhancer evolution provide mechanistic insights into gene regulation. Mol Biol and Evol 31(1):184-200.

Escudero LA, Bosch D and Batllori L (2012) *Drosophila suzukii*, una nueva plaga de los frutales. Vida Rural Junio 18-22.

Erwin DH & Davidson EH (2009) The evolution of hierarchical gene regulatory networks. Nat Rev Genet 10:141-148.

Fartyal RS, Sarswat M, Lhamo N, Sati PC and Asha L (2014) Records of *Zaprionus indianus* and *Drosophila suzukii indicus* as invasive fruit pests from mid valley region of Garhwal Uttarakhand, India. Drosoph Inf Serv 97:119-123.

Feschotte C and Pritham EJ (2007) DNA Transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331-368.

Feyereisen R (2005) Insect cytochrome, p. 450 in Comprehensive Molecular Insect Science Vol. 4, edited by L. I. Gilbert, K. Latrou, and S. S. Gill. Elsevier, Oxford.

Fraimout A, *et al*. (2017) Deciphering the Routes od invasion of *Drosophila suzikii* by Means of ABC Random Forest. Mol Biol Evol 34(4):980-996. doi:10.1093/molbev/msx050.

Fritts TH and Rodda GH (1998) The role of introduced species in the degradation of island ecosystems: a case history of Guam. Annu Rev Ecol Syst 29:113-140.

Galego LG, Ceron CR and Carareto CM (2007) Analysis of the drosophilid *Zaprionus indianus* introduction in Brazil: contribution of esterase loci polymorphisms. Dros Inf Serv 90:79-84.

Galego LG and Carareto CM (2010) Scenario for the spreading of the invasive species *Zaprionus indianus* Gupta 1970 (Diptera: Drosophilidae) throughout Brazil. Genet Mol Biol 33:767-773.

Garcia ACL, Valiati VH, Gottschalk MS, Rohde C and Valente VLS (2008) Two decades of colonization of the urban environment of Porto Alegre, southern Brazil, by *Drosophila paulistorum* (Diptera, Drosophilidae). Sér Zool 983: 329-338.

Garcia CF, Hochmüller CJC, Valente VLS and Schmitz HJ (2012) Drosophilid Assemblages at Different Urbanization Levels in the City of Porto Alegre, State of Rio Grande do Sul, Southern Brazil. Neotrop Entomol 41:32-41.

Girard A and Hannon GJ (2008) Conserved themes in small-RNA mediated transposon control. Trends Cell Biol 18:136-148.

Gonzaléz G, Mary AL and Goñi B (2015) *Drosophila suzukii* (Matsumura) found in Uruguay. Drosoph Inf Serv 98:103-107.

Goodhue RE, Bolda M, Farnsworth D, Williams JC and Zalom FG (2011) Spotted-wing *Drosophila* infestation of California strawberries and raspberries:

economic analysis of potential revenue losses and control costs. Pest Manag Sci 67:1396-1402.

Guengerich FP (2007) Mechanisms of cytochrome P450 substrate oxidation: MiniReview. J Biochem Mol Toxicol 21(4):163-8.

Hamblin MT and Veuille M. (1999) Population structure among African and derived populations of *Drosophila simulans:* evidence for ancient subdivision and recent admixture. Genetics 153(1):305-17.

Hauser M (2011) A historic account of the invasion of *Drosophila suzukii* (Matsumura) (Diptera: Drosophilidae) in the continental United States, with remarks on their identification. Pest Manag Sci 67(11):1352-7.

Hemingway J, Ranson H (2000) Review Insecticide Resistance in Insect Vectors of Human Disease. Annu Rev Entomol 45:371-391.

Ioriatti C, Walton V, Dalton D, Anfora G, Grassi A, Maistri S and Mazzoni V (2015) *Drosophila suzukii* (Diptera: Drosophilidae) and its potential impact to wine grapes during harvest in two cool climate wine grape production regions. J Econ Entomol 4:1-8.

Jennings BH (2011) *Drosophila* – a versatile model in biology & medicine. Mater Today 14(5):190-195. DOI: 10.1016/S1369-7021(11)70113-4

Jordan IK, Rogozin IB, Glazko GV and Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19:68-72.

Josephy PD, Guengerich FP, Miners JO (2008) "Phase I" and "Phase II" drug metabolism: therminology that we should phase out? Drug Metab Rev 37:575-580. DOI: 10.1080/03602530500251220.

Joshi KN, Biddinger JD, Demchak K and Deppen A (2014) First Report of *Zaprionus indianus* (Diptera: Drosophilidae) in Commercial Fruits and Vegetables in Pennsylvania. J Insect Sci 14(259).

Kaneshiro KY (1983) *Drosophila* (Sophophora) *suzukii* (Matsumura). Notes and exhibitions Proc Hawaiian Entomol Soc 24:179.

Kapitonov V and Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. PLoS Biology 3:998-1011.

Kidwell MG (2005) Transposable Elements. In: Gregory, T.R (org.). The Evolution of the Genome, New York: Elsevier academic Press, 2005. Cap. 3, p. 165-221.

Kidwell MG and Lisch DR (2001) Perspective: transposable elements, parasitic DNA and genome evolution. Evolution 55:1-24.

Klaassen CD, Watkins JB (1999) Biotransformation of xenobiotics in Casarett & Doull's Toxicology – The basic science of poisons. USA: McGraw-Hill Companion Handbook.

Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, Ng H-H and Bourque G (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 42:631-634.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al*. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921.

Lasa R and Tadeo E (2015) Invasive drosophilid pests *Drosophila suzukii* and *Zaprionus indianus* (Diptera: Drosophilidae) in Veracruz, Mexico. Fla Entomol 98:987-988.

Langdon T, Seago C, Mende M, Leggett M, Thomas H, Forster JW, Jones RN and Jenkins G (2000) Retrotransposon evolution in diverse plant genomes. Genetics 156:313-325.

Leblanc L, O'Grady PM, Rubinoff D and Montgomery SL (2009) New immigrant Drosophilidae in Hawaii, and a checklist of the established immigrant species. Proc Hawaiian Entomol Soc 41:121-127.

Lee JC, Bruck DJ, Dreves AJ, Ioriatti C, Vogt H and Baufeld P (2011) In Focus: Spotted wing *Drosophila*, *Drosophila suzukii*, across perspectives. Pest Manag Sci 67(11):1349-51.

Lee JC, Dreves AJ, CaveAM, Kawai S, Isaacs R, Miller JC, Van Timmeren S and Bruck DJ (2015) Infestation of wild and ornamental non crop fruits by *Drosophila suzukii* (Diptera: Drosophilidae). Ann Entomol Soc Am 3:1-13.

Li X, Schuler MA and. Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu Rev Entomol 52: 231-253.

Liebler DC and Guengerich FP (2005) Elucidating mechanisms of drug-induced toxicity. Nat Rev Drug Discov, 4(5):410-20.

Lincoln R, Boxshall G and Clark P (1998) A dictionary of ecology, evolution and systematics. 2nd ed Cambridge: Cambridge University Press p.361.

Lipatov M, Lenkov K, Petrov DA and Bergman CM (2005) Paucity of chimeric gene-transposable elements transcripts in the *Drosophila melanogaster* genome. BMC Biol *3*:*24.*

Lodge DM (1993) Biological invasions: lessons for ecology. Trends in ecology and evolution, Oxford 8(3) p 133-137.

Maeso I and Tena JJ (2015) Favorable genomic environments for cis-regulatory evolution: A novel theoretical framework. Semin Cell Dev Biol 57:2-10.

Marsano RM, Caizzi R, Moschetti R and Junakovic N (2005) Evidence for a functional interaction between the *Bari1* transposable element and the cytochrome P450 *CYP12a4* gene in *Drosophila melanogaster*. Gene 357:122-128.

Meyer UA (1996) Overview of enzymes of drug metabolism. J Pharmacokinet Biopharm 24:449- 459.

McClintock B (1956) Controlling elements and the gene. Cold Spring Harb Symp Quant Biol 21:197-216.

Mikkelsen TS, Hillier LW, Eichler EE, Zody MC, Jaffe DB, Yang S, Wolfgang E, Ines H, Kerstin L-T, Tasha KA, e*t al*. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87.

Molineris I, Grassi E, Ala U, Di Cunto F and Provero P (2011) Evolution of promoter affinity for transcription factors in the human lineage. Mol Biol Evol 28:2173-2183.

Morgan TH (1910) Sex limited inheritance in *Drosophila*. Science 32:120-122.

Nebert DW (2005) Cytochrome *P450* (*CYP*) Gene superfamily. Enciclopidia of live Sciences 1-9.

Nebert DW (1991) Proposed Role of Drug-Metabolizing Enzymes: Regulation of Steady State Levels of the Ligands that Effect Growth, Homeostasis, Differentiation, and Neuroen*DOC*rine Functions. Molecular En*DOC*rinology 5:1203–1214.

Nebert DW, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, Levin W (1987) The *P450* gene superfamily: recommended nomenclature. DNA 6(1):1-11.

Nebert DW, Gonzalez FJ (1987) *P450* Genes: Structure, Evolution, and Regulation. Ann Rev Biochem 56:945 – 993.

Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics 6(1):1-42.

Ohler U, Liao GC, Niemann H and Rubin GM (2002) Computational analysis of core promoters in the *Drosophila* genome. Genome Biol 3(12):1-12.

Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. J Biol Chem 239:2370 – 2378 and 2379 – 2385.

Paula MA, Lopes PHS and Tidon R (2014) First record of *Drosophila suzukii* in the Brazilian Savanna. Drosoph Inf Serv 97:113-115.

Pavan C (1959) Relações entre populações de *Drosophila* e o meio ambiente. Bol. fac. Filos. Cienc. e Let. USP, 221, Biol. Geral, 11: 1-81.

Pereira V, Enard D, Eyre-Walker A (2009) The effect of transposable element insertions on gene expression evolution in rodents. PLoS ONE 4:e4321.

Pimentel D, McNair S, Janeka J, Wightman J, Simmonds C, O´Connell C, Wong E, Russel L, Zern J, Aquino T, *et al.* (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. Agriculture Ecosystem environmentals 84:1-20.

Pooma W, Gersos C, Grotewold E (2002) Transposon insertion in the promoter of the *Zea mays* a1 gene differentially affect transcription by the Myb factors P and C1. Genetics 161:793-801.

Poppe JL, Schmitz HJ and Valente VLS (2015) The diversity of Drosophilidae in the South American pampas: update of the species records in an environment historically neglected. Drosoph Inf Serv 98:47-51.

Pritham EJ (2009) Transposable elements and factors influencing their success in eukaryotes. Journal of Heredity 100(5):648-655.

Ramniwas S, Kajla B and Parkash R (2012) Extreme physiological tolerance leads the wide distribution of *Zaprionus indianus* (Diptera: Drosophilidae) in temperate world. Acta Entomol Sin 55:1295-1305.

Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R (2002) Evolution of supergene families associated with insecticide resistance. Science 298:179-181.

Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionary and biomedical insights from the Rhesus macaque genome. Science 316:222-234.

Rota-Stabelli O, Blaxter M, Anfora G (2013) *Drosophila suzukii*. Curr Biol 23(1):R8-9.

Ruiz GM, Rawlings TK, Dobbs FC, Drake LKA, Mulladdy T, Hugh A and Colwrll RR (2000) Global spread of microorganisms by ships. Nature, London, v408 (6808) p.49-50. DOI: 10.1038/35040695

Santos RSS (2014) Ocorrência de *Drosophila suzukii* (Matsumura, 1931) (Diptera: Drosophilidae) atacando frutos de morango no Brasil. Embrapa Uva e Vinho, Bento Gonçalves 4p (Comunicado Técnico 159).

Silva NM, Fantinel CC, Valente VLS and Valiati VH (2005) Ecology of colonizing populations of the figfly *Zaprionus idnianus* (Diptera, Drosophilidade) in Porto Alegre, Southern Brazil. Iheringia Sér Zool 95:233-240.

Schlenke TA and Begun DJ (2004) Strong selective sweep associated with transposon insertion in *D. simulans*. Proc Natl Acad Sci USA 101:1626-1631.

Schlesener DCH, Wollmann J, Nunes AM, Cordeiro J, Gottschalk MS and Garcia FRM (2015) *Drosophila suzukii*: nova praga para a fruticultura brasileira. Biológico, São Paulo, v.77, n.1, p.45-51.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, *et al*. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115.

Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G (2007) Comparative analysis of transposed element insertion within human and mouse genomes reveals *Alu*'s unique role in shaping the human transcriptome. Genome Biol 8:R127.

Simons C, Pheasant M, Makunin IV, Mattick JS (2006) Transposon-free regions in mammalian genomes. Genome Res 16:164-172.

Sturtevant AH (1913) A third group of linked genes in *Drosophila ampelophila*. Science 37:990-992.

Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In: King, R.C., (Ed). Handbook of Genetics. Plenum Press, Nova York , 3, 421-469.

Thornburg BG, Gotea V, Makayowski W (2006) Transposable elements as a significant source of transcription regulating signals. Gene 365:104-110.

Tidon R, Leite DF, Leão BFD (2003) Impact of the colonization of *Zaprionus indianus* (Diptera: Drosophilidae) in different ecosystems of the neotropical region: 2 years after the invasion. Biological Conservation 112:299-305.

Tijet N, Helvig C, Feyereisen R (2001) The cytochrome *p450* gene superfamily in *Drosophila melanogaster*: annotation, intro-exon organization and phylogeny. Gene, Oxford 262:189-198.

Timmeren SV and Isaacs R (2013) Control of spotted wind *Drosophila, Drosophila suzukii*, by specific insecticides and by conventional and organic crop protection systems. Crop Prot 54:126-133.

Tochen S, Dalton DT, Wiman N, Hamm C, Shearer PW and Walton V (2014) Temperature-related development and population parameters for *Drosophila suzukii* (Diptera: Drosophilidae) on cherry and blueberry. Environ Entomol 43:501-510.

Val FC and Sene FM (1980) A newly introduced *Drosophila* species in Brazil (Diptera: Drosophilidae). Papéis avulsos do Departamento de Zoologia da Sociedade Agrícola 33:293-298.

Van de Lagemaat LM, Landry JR, Mager DL and Medstrand P (2003) Transposable elements in mammals promote regulatory variation and

diversification of genes with specialized functions. Trends Genet 19:530-536.

Varani A, Carvalho L, Zerillo MM, Monteiro-Vitorello CB (2015) Elementos de Transposição: Classificação e Mecanismos de Mobilização in: Elementos de Transposição, Diversidade, Evolução, Aplicações e Impacto nos genomas dos seres vivos. FIOCRUZ:11-43.

Vilela CR (1999) Is *Zaprionus indianus* Gupta, 1970 (Diptera: Drosophilidae) currently colonizing the Neotropical region? Drosoph Inf Serv 82:37-39.

Vilela CR and Mori L (2014) The invasive spotted-wing *Drosophila* (Diptera, Drosophilidae) has been found in the city of São Paulo (Brazil). Revista Brasileira de Entomologia 58(4):371-375.

Vogl C, Das AS, Beaumont M, Mohanty S and Stephan T W (2003) Population subdivision and molecular sequences variation: theory and analyses of *Drosophila ananassae* data. Genetics 165:1385-1395.

Walsh DB, Bolda MP, Goodhue RE, Dreves AJ, Lee JC, Bruckd J, Walton VM, O'Neal SD, Zalom FG (2011) *Drosophila suzukii* (Diptera: Drosophilidae): Invasive pest of ripening soft fruit expanding its geographic range and damage potential. J Pest Manag 106:289–295.

Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D (2007) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc. Natl Acad Sci U S A104:18613–18618.

Wang J, Bowen NJ, Mariño-Ramírez L and Jordan IK (2009) A c-Myc regulatory subnetwork from human transposable element sequences. Mol Biosyst 5:1831-1839.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al*. (2007) A unified classification system for eukaryotic transposable elements, Nat Rev Genet 8(12):973-82.

Yang HP and Barbash DA (2008) Abundant and species-specific *DINE-1* transposable elements in 12 *Drosophila* genomes. Genome Biol 9:R39.

Yassin A, Capy P, Madi-Ravazzi L, Ogereau D and David JR (2008) DNA barcode discovers two cryptic species and two geographical radiations in the invasive drosophilid *Zaprionus indianus*. Mol Ecol Notes 8:491-501.

CHAPTER 2

Increased *CYP* gene lengths are associated with increased transposable element content in *Drosophila suzukii*

Paula D. S. Berrutti[1*], Sidia M Callegari-Jacques[1,2], Maríndia Deprá[1] and Vera L. S. Valente[1]

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Laboratório de *Drosophila*, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFGRS), Porto Alegre, Rio Grande do Sul, Brazil

[2]Departamento de Estatística, Instituto de Matemática e Estatística, UFRGS.

Manuscript sent to **Gene** journal in 11/02/2018.
Reviewed manuscript resubmitted in 30/05/2018.

ABSTRACT

Background: An in silico analysis was performed to evaluate a possible connection between *CYP* genes and transposable elements (TEs) in a non-pest species (*Drosophila melanogaster*) and a pest species (*D. suzukii*).

Results: *Helitron* fragments have accumulated in introns of *CYP* genes. *Helitrons* are known as “exon-shuffling machines”, class II DNA transposons. Their transposition may result in capture of the flanking sequence, with consequent transduplication of the gene by transposition events. We found putative transcription-factor binding sites in all TE sequences, which reinforces the suggestion that TEs may influence gene regulation. In addition, our analysis indicated that the *D. suzukii* genome carries more TEs than the genome of *D. melanogaster.*

Conclusions: We hypothesize that the ten longer *CYP* genes occurring only in *D. suzukii* are enriched in TE fragments, which possibly resulted from *Helitron* transposition events. Selection of higher TE content within environmental-response genes could result in greater genomic plasticity of *D. suzukii*.

Highlights:

- Ten *CYP* genes show longer genes in *D. suzukii*.
- Longer *CYP* genes possibly resulted from *Helitron* transposition events.
- Putative TFBS were found in TE sequences nearby CYP genes.
- The TE content of *D. suzukii* genome is twice than *D. melanogaster*.
- *Helitron* is the most prevalent DNA transposon in the genome of *D. suzukii*.

Keywords: Cytochrome P450 monooxygenases; *Helitron*; transcription-factor binding site, genome TE content.

BACKGROUND

Cytochrome P450 monooxygenases (*CYP*s) are enzymes that play a role in

metabolic resistance in the detoxification of xenobiotics and endobiotics such as arthropod pests, natural plant allelochemicals and synthetic insecticides (Li et al., 2007). They are part of an ancient gene family that occurs in virtually all organisms (Feyereisen, 2005) and are a broad group of isoenzymes that vary in protein abundance and substrate specificity but use oxidant systems (Scott, 1999).

Resistance to insecticides is a widely used model for studying evolutionary phenomena, since the agent is known (pesticides) and the response to selection (evolution of resistance) is usually rapid (Mckenzie and Batterham, 1994). For example, resistance was studied by overexpression of the *CYP6p3* gene in *Anopheles gambiae* (Müller et al., 2008) and of *CYP6bq9* when expressed in the brain of *Tribolium castaneum* (Zhu et al., 2010). Resistance in the aphid *Myzus persicae* and the fruit fly *Drosophila melanogaster* occurs by duplication of the *CYP6cy3* and *CYP6g1* genes, respectively (Puinean et., 2010; Harrop et al., 2015), as well as by overexpression of the latter in *D. melanogaster* (Daborn et al., 2002).

Genes associated with environmental responses tend to be less well conserved, and the evolutionary response (adaptability to environment) is usually rapid. In vertebrates, it is suggested that evolutionary conserved *CYP*s function in endogenous pathways, while in the most divergent species, these *CYP*s function as evolutionary responses to different xenobiotics, contributing to genome plasticity (van de Lagemaat et al., 2003; Chen and Li, 2007; Thomas, 2007). In insects, *CYP*s have been identified as the only mechanism of resistance (Li et al., 2007; Feyereisen, 2005; Scott, 1999).

The role of transposable elements (TEs) inserted in *Drosophila CYP*s has long been predicted (Daborn et al., 2002; Catania et al. 2004; Schlenke and Begun, 2004; Bogwitz et al., 2005; Marsano et al., 2005; Chung et al., 2007; Carareto et al., 2013). TEs are repetitive DNA sequences that can change their location within and between genomes, except for copy paste elements, which cannot change their location once inserted. Transposons are also able to amplify the size of their host genome. TE insertions are also associated with phenotypic changes in insecticide resistance through changes in gene expression. For example, the overexpression of *CYP6g1* in *D. melanogaster* is characterized by

the insertion of the *Accord* retroelement upstream of the gene (Daborn et al., 2002). The ortholog of this gene in *D. simulans* is also overexpressed through the insertion of the *DOC* element in its flanking region (Schlenke and Begun, 2004). Moreover, the presence of the *Bari-1* element at the 3' end of *CYP12a4* in *D. melanogaster* increases the expression of this gene (Bogwitz et al., 2005; Marsano et al., 2005). Carareto et al. (2013) observed several putative insertions of TEs in the flanking regions of *CYP*s in *D. melanogaster* and *D. simulans* being *DNAREP1*, which belongs to the *Helitron* superfamily (review in Thomas and Pritham 2015), the most recurrent element observed. *Helitron* is a new class of repeats associated with gene capture, exon shuffling, genome rearrangement, and consequent transduplication (Kapitonov and Jurka, 2007).

Recently, *Drosophila suzukii* Matsumura (1931) has been more widely studied because it is one of the main pests associated with fruit growing in the world (Walsh et al., 2011; Asplen et al., 2015). First recorded from Japan, *D. suzukii* spread to Asia, and from there to North America (Walsh et al., 2011; Asplen et al., 2015; Cini et al., 2012; Rota-Stabelli et al., 2013), Europe (Kaneshiro, 1983; Leblanc et al., 2009) and South America, where our group first reported its occurrence (Deprá et al., 2014; Vilela and Mori, 2014; Paula et al., 2014; Bitner-Mathé et al., 2014). *D. suzukii* shows sexual dimorphism in the coloration of the wings and is popularly called Spotted Wing *Drosophila*. Females injure fruits when they deposit eggs with the serrated ovipositor, mainly in healthy fruits rather than in fallen fruits (Walsh et al., 2011; Lee et al., 2015). Drilling injuries allow pathogens to penetrate, increasing economic losses by as much as 80% (Dreves et al., 2009; Hauser, 2011; Escudero et al., 2012) and causing the release of volatile products (Abraham et al., 2015) that attract other drosophilid species (Timmeren and Isaacs, 2013; Joshi et al., 2014; Lasa and Tadeo, 2015). Once the fly becomes established, it is very difficult to eradicate, and production costs increase permanently due to the need for monitoring, management, increased use of chemicals and secondary selection of fruits.

The recently sequenced genome of *D. suzukii* (Chiu et al., 2013) has 69 annotated *CYP* genes (SpottedWingFlyBase). *D. suzukii* has fewer *CYP*s than most *Drosophila* species (Chiu et al., 2013). The genome of *D. melanogaster*

(FlyBase) contains 99 genes belonging to the *CYP* family. Because *CYP*s are considered the only metabolic system in insects that can mediate resistance to all classes of insecticides, examination of TEs associated with these genes and their possible consequences could provide interesting insights at the genetic and molecular levels for understanding of the insecticide-resistance phenotype.

Therefore, we have compared the *CYP* gene repertoire between *D. suzukii* and *D. melanogaster* to characterize putative TE sequences in them or in their flanking regions as well as regulatory elements from these TEs. The most frequent insertion in *D. suzukii CYP* genes stems from the *Helitron* superfamily, and transposition of these may result in the capture of a flanking sequence (Kapitonov and Jurka, 2007). We also characterized the genomic TE content in these species through NGS reads combined with graph-based clustering estimations of repeats. We detected a higher proportion of TEs in the *D. suzukii* genome, as well as in the proportion of the *Helitron* superfamily, than in *D. melanogaster.* The association of *Helitron* fragments and the differences in *CYP* genes of *D. suzukii* with respect to *D. melanogaster* may have resulted from a transduplication event, which suggests the existence of adaptive structural changes in the genome of this species. Here, we describe the possible association between these longer genes, TE insertions carrying putative TFBSs, and the larger genomic TE content. Our findings reinforce the hypothesis that TEs could be selectively enriched among environmental-response genes, resulting in greater genomic plasticity of *D. suzukii*.

METHODS

CYP genes *in silico* analysis

All of the *CYP* genes from *D. suzukii* and *D. melanogaster* were obtained and extracted from their Gbrowser in the websites spottedwingflybase.org (Chiu et al., 2013) and flybase.org, respectively (SpottedWingFlyBase; FlyBase), which provided genomic coordinates for all genes. The genes analyzed are described in Additional Tables S1-S4, for *D. suzukii* and *D. melanogaster*. For each gene, we also extracted 3 Kb upstream and downstream from the annotated transcription start and end coordinates. TEs inserted in the flanking regions could be altering

gene expression by contributing novel transcription regulatory signals. Gene sequences obtained were visually inspected on Gbrowser and were manually analyzed to compare their genomic information.

Visual display of longer genes was performed in R using the genoPlotR library (Guy et al., 2010), and all graphics were edited in Inkscape v0.92.1 (2017). For this comparison, the phylogeny generated by Chiu et al. (2013) was used. We broadened the analysis by adding the orthologous genes of two sister species of *D. suzukii*: *Drosophila biarmipes* and *Drosophila takahashii* (NCBI). To follow with a robust comparison among the species, we also searched for transposons in the orthologous genes of these sister species.

Transposons *in silico* analysis

In order to identify the presence or absence of TEs, the *CYP* gene sequences and their 3 Kb flanking regions were submitted to RepeatMasker web server (http://www.repeatmasker.org) using the database of *Drosophila* reference TEs stored in Repbase (Jurka et al., 2005). The search was applied using the parameters: crossmatch, fruit fly, and matrix based on a GC level query. The sequences were assigned to a given element on the basis of the best match obtained.

Transposon sequences inserted in *CYP* gene regions were analyzed to find putative transcription-factor binding sites. They were predicted using the web server ConSite (http://consite.genereg.net/), which accesses the JASPAR CORE Insecta (Bryne et al., 2008) of the *D. melanogaster* database with a 90% transcription factor cutoff score, following the study by Carareto et al. (2013).

Additionally, to confirm if the relationship between the number of TEs inserted in the *CYP* genes and the composition of TEs in the genomes of the species studied here is proportional, Illumina reads were downloaded from the SRA (Sequence Read Archive): *D. suzukii* – SRR942805, North-American sample sequenced by Chiu et al. (2013); *D. melanogaster* – SRR1738161. Graph-based clustering of NGS reads was performed with RepeatExplorer (Novák et al., 2013) using the latest Galaxy-based web server implementation, and also following the pipeline by Silva et al. (Silva et al., 2016).

Set of random genes

One set of 500 random genes from *D. suzukii* was created running BEDTools software v2.27.0 (Quinla 2014). All 500 random genes were manually inspected and the orthologous genes in *D. melanogaster* were selected. The previous methodology for transposon *in silico* analysis was applied.

Statistical tests

Due to asymmetry in the distribution of gene size, the Wilcoxon-Mann-Whitney test (WMW) was employed to compare the size of 10 selected *CYP* genes and that set of 500 random orthologous genes in each species.

As the total genome sizes were different between species, the sizes of 10 selected CYP genes were normalized to the median size of the 500 genes randomly selected from the entire species genome, for a fair comparison between species. The genes used to obtain the median in *D. suzukii* were the same as those chosen to calculate the median gene size for *D. melanogaster*. The normalized gene size was the size of a gene (in base pairs) divided by the median size of the 500 randomly selected genes. The Wilcoxon nonparametric rank test for paired data was used to compare species for the normalized CYP sizes.

The differences between *D. suzukii* and *D. melanogaster* regarding the TE enrichment within and near genes, TE in *CYP* or in the remaining genome genes, and *Helitron* enrichment in *CYP* or in the overall gene background were tested by Chi-square with continuity correction. For these comparisons, all genes and intergenic regions annotated for both species were considered.

The statistical analyses were done using SPSS® v.18. A P value equal to 0.05 was used as a threshold for statistical significance.

RESULTS

CYP genes harbor transposable-element fragments

Among 76 *CYP* genes annotated for *D. suzukii*, 36 genes have putative transposon sequences (Additional file 1: Table S1 and Additional file 2: Table S2), and in *D. melanogaster*, 34 of the 91 genes analyzed have transposons (Additional file 3: Table S3 and Additional file 4: Table S4). Despite the smaller number of *CYP* genes, a larger number of TEs fragments (103) were observed in *D. suzukii* compared to *D. melanogaster* (87). This difference is due largely to *Helitron* elements (Table 1), a DNA transposon present in 31 *CYP* genes (Figure 1 and Additional file 1: Table S1). In *D. melanogaster*, *Helitron* is also distributed in a larger number of *CYP*s (Figure 1), but the element with the most frequently observed found was *Gypsy*, an LTR retrotransposon, with 38 fragment insertions (Table 1).

We found TEs inserted in 5' and 3' flanking regions of the *CYP* genes (3 Kb up- and downstream) and in the intron region for both species (Figure 2). *D. suzukii* has a larger number of TE insertions in the 5' flanking region, where most promoter sequences are located. In contrast, *D. melanogaster* has a larger number of TE insertions in intron regions that is mostly due to a single gene, *CYP307a2*, which carries 31 TE fragments in the intron region (Additional file 3: Table S3). The ortholog *CYP307a2* in *D. suzukii* shows nine insertions of TEs in its intron (Additional file 1: Table S1), and in both species, the retroelements are the most numerous.

Helitron **elements shaping gene length**

In general, the organization of genes is well conserved among species belonging to the same order, and therefore, data on intron conservation and exon structure are well correlated with the phylogenetic position of the species (Rewitz et al., 2007). Exon and intron structures annotations are supported at the transcript level in *D. suzukii* (Chiu *et al*., 2013), in *D. melanogaster* (Graveley *et al*., 2010), in *D. biarmipes* (NCBI *Drosophila biarmipes* Annotation Release 101), and in *D. takahashii* (NCBI *Drosophila takahashii* Annotation Release 101). When we inspected the genes that had TE insertions in *D. suzukii* and *D. melanogaster*, we noted that some *CYP* genes of *D. suzukii* were longer compared to *D. melanogaster* (Figure 3). Altogether, ten of 36 genes with TE insertions in *D.*

suzukii have more exons and introns, as repetitive conserved blocks, than in *D. melanogaster* (Figure 3).

All these ten longer *CYP* genes have *Helitron* fragments inserted, with a total contribution of 5,577 pb in *D. suzukii* (Additional file 1: Table S1) and 653 pb in *D. melanogaster* (Additional file 3: Table S3). However, for each gene, the *Helitron* fragments per se represent the minor portion of the length, between 48 pb and 567 bp. The presence of *Helitron* repeats suggests that this TE could be a vehicle for generating these increased gene lengths due to their transposition and recombination activity.

Analysis with the genoPlotR returned similarities among the exons when comparing each ortholog among *D. suzukii*, *D. biarmipes*, *D. takahashii* and *D. melanogaster*. The genes *CYP12a4*, *CYP12e1*, *CYP6a18*, *CYP6a20*, *CYP6a21*, *CYP6a23*, *CYP6d5* and *CYP4e2* of *D. suzukii* (Figure 3.A-H) have at least one *Helitron* fragment in the intron region. Different from what was observed for *D. suzukii*, the *CYP12a4* and *CYP12e1* orthologous to *D. biarmipes* and *D. takahashii* do not have transposon insertions (Figure 3.A-B). However, in *D. melanogaster*, the *CYP12a4* ortholog has the *BARI* element in the 3' flanking region (Figure 3.A), as previously annotated (Bogwitz et al., 2005). The *D. suzukii CYP4e2* gene is increased in relation to the *D. melanogaster* ortholog, but with one fewer exon compared to the ortholog in *D. biarmipes* (Figure 3.H). Interestingly, even with one fewer exon, *D. suzukii CYP4e2* is a larger gene than the ortholog in *D. biarmipes*. Moreover, there is an *Helitron* fragment between exon six and seven that is not present in the *D. biarmipes* gene. The *CYP4c3* gene has two fragments of *Helitron* in the 3' flanking region and no intron fragment (Figure 3.I), and its sister species *D. takahashii* has the same fragment between exons five and six. These results agree with the literature about evolution mediated by *Helitrons* (Morgante et al., 2005; Kapitonov and Jurka et al., 2007; Lal et al., 2009; Barbaglia et al., 2012; Grabundzija et al., 2016). In these publications, authors show that the structure of the gene may have resulted from recombination or gene capture between *Helitron* insertions in the ancestral species, leading to a transduplication of this gene in *Drosophila suzukii*. Thus, we provide a hypothetical example in Figure 4 of *Helitron*-mediated gene capture that probably occurred in these 10 *D. suzukii CYP*

genes. This may occur when the end hairpin signal in *Helitron* is bypassed, and strand displacement continues through nearby gene regions until a new termination signal is reached (Kapitonov and Jurka, 2007; Grabundzijaet al., 2016). In *CYP12a4* (Figure 3-A) and *CYP6a20* (Figure 3-D), for example, the arrangement of *Helitron* in introns, the orientation and the high similarity of the exons suggest the hypothesis that the host gene was probably captured during its transposition (Figure 4-A).

The annotated biological processes for these genes, *Cyp12a4* and *Cyp6a20*, are described as responses to insecticide and aggressive behavior, respectively. Both characteristics are related to the successful adaptation of invasive species. Another interesting point is the absence of one exon in *D. suzukii CYP4e2* (Figure 3-H) in relation to its sister species *D. biarmipes*. We hypothesized that the reason for this loss may be the insertion of *Helitron*, which formerly was the seventh exon in the ancestral species. In *CYP4c3* (Figure 3-I), we hypothesized that the mechanism that led to the change in the structure of this gene is related to a possible recombination between the inserted *Helitron*s, since there is only one copy in the basal *D. takahashii* and two copies in *D. suzukii*. When Gilbert (1987) reported the shuffling exon, he observed that repetitive elements in intron regions can create hotspots for recombination, which leads to the shuffle of exons (Figure 4-B).

With the genoPlotR analysis, little or no similarity was observed only for the *CYP6w1* gene annotated in scaffold 2 (Figure 3-J). However, the same gene annotated for scaffold 8 showed high similarity to orthologs of other species (Figure 3-J). It is possible that there is an inaccurate annotation of this sequence, where this is likely another gene. An analysis of this sequence, using BLAST on the NCBI, revealed high identity with the *CYP6d2* genes of the sister species *D. biarmipes* (89%) and *D. takahashii* (87%). The *D. suzukii CYP6d2* gene is absent from the Gbrowser (http://spottedwingflybase.org) but is predicted by the NCBI genome browser. The same does not occur for the *CYP6d5* gene, which is annotated in two scaffolds 99 and 1273 (Figure 3-G), where both paralogs show high similarity to each other and to the ortholog of *D. biarmipes*, suggesting that it was probably an expanded event for this gene in *D. suzukii*; perhaps it could be

led by TE insertion and transposition.

From 500 random genes randomly selected from the entire genome, we visually inspected 124 genes that were longer in *D. suzukii* compared to orthologous genes from *D. melanogaster*. From these longer genes in *D. suzukii*, 45 genes carried 249 Helitron copies, whereas in *D. melanogaster*, 41 genes carried 110 Helitron copies (Additional file 8: Table S8).

We compared the size of *CYP*s with that of 500 random genes in both species. In *D. melanogaster, CYPs* are smaller (median, md = 2117) than the random genes (md = 5603) (WMW; P value = 0.025), but in *D. suzukii, CYPs* do not differ in size (md = 8032) from random genes (md = 6325) (WMW; P value = 0.526). Thus, relative to the global gene size, *D. suzukii CYP*s are statistically larger than the *D. melanogaster CYP*s.

As the size of CYP genes could be due to a larger genome size in *D. suzukii*, thus not being the result of arrangements of TEs but of a normal difference between major and minor genomes, we normalized the size of *CYP* genes to the median of the 500 random selected genes of the species for a fair comparison between species. In *D. melanogaster,* the relative size of the *CYPs* are smaller (md = 0.38) than those of *D. suzukii* (md = 1.27) (Wilcoxon test; P = 0.002). Thus, *D. melanogaster CYP*s sizes amount to 38% of the average overall genome genes, whereas *D. suzukii CYP*s are in general 27% larger than the overall genome genes.

Transposons are enriched in putative TFBS

Assuming that TEs carry transcription-factor binding sites (TFBS), since these sequences are preferentially retained in the genes because they harbor regulatory signals (Jordan et al., 2003; Feschotte, 2008), we searched for putative TFBS in all sequences of TEs found in the *CYP* genes of *D. suzukii* and *D. melanogaster* (Figure 5, Supplementary Table S7). Interestingly, with a differential retention of the TE classes in the genes (Table 1), it is also expected that different TFBS contents will be found (Thornburg et al., 2006). However, we observed little difference in the TFBS content among the different TEs (Figure 5 and Additional file 7: Table S7). Also, although *D. suzukii* has higher TE coverage (Additional file

1: Table S1), the highest number of TFBS is found in *CYP* genes TE fragments of *D. melanogaster* (Additional file 7: Table S7). *D. melanogaster* has higher TE base pair coverage in *CYP* genes (total = 38178 bp) (Additional file 3: Table S3) than *D. suzukii* (total = 21021 bp) (Additional file 1: Table S1). This difference in numbers could explain why fewer TFBS are observed in the TEs of *D. suzukii*. The disparity of TFBS number could also be explained by the quality of the TE fragments found in the *CYP* genes, LTRs, or TE 5' regions with more regulatory signals than internal TE sequences.

For both species, the putative TFBS *Hunchback* and *CF2-II* (Chorion factor 2) are over-represented (Figure 5). These proteins belong to the class of *Zinc Finger* transcription factors C2H2; *Hunchback* is strongly expressed early in development (Nüsslein-Volhard and Wieschaus, 1980; Lehmann, 1988), and *CF2-II* is expressed late in the embryonic stage (Shea et al., 1990). It is therefore likely that these TEs intrinsically carry TFBS as regulatory sequences, which may confer tissue-specific expression (Chung et al., 2007). Because TFBS are short, they occur randomly in both DNA and TEs (Thornburg et al., 2006). However, the presence of small fragments of TEs inserted in the flanking regions of *CYP*s could be affecting the gene expression, since they harbor putative TFBSs, and they are indications that TEs could be extremely important in adaptation to different environments for both species (Jordan et al., 2003; Feschotte, 2008; Shea et al., 1990; Thornburg et al., 2006).

TE content in *Drosophila* genomes

We analyzed the TE content in the genomes of the two *Drosophila* species, since TEs are recognized as the main contributors in the evolution of the genomes. Approximately 36% of the assembled *D. suzukii* genome contains TE sequences; the proportion is approximately 16% in the genome of *D. melanogaster* (Table 2).

*There are two available sequences of D. suzukii genome: one obtained by Chiu et al. (2013) from North American samples (*SRA096061*), and another by Ometto et al. (2013) from European samples (*ERP001893*). Both studies used paired-end sequencing on the Illumina Hiseq2000 platform.* Chiu et al. (2013) used

only one genome to run an automated homology comparison along with 6003 TEs from *D. melanogaster*. *Ometto et al. (2013*) *analyzed the D. suzukii and D. melanogaster genomes by* using the homology-based RepeatMasker and the Repbase Insect library*. They found ~11% of TE content in D. suzukii and ~17% in D. melanogaster.* A study by Rius et al. (2016) estimated the TE content of *D. suzukii* and *D. melanogaster* using the same genomic sequences used in this present study *(*SRA096061). However, they found that TEs represent 18.7% and 21.67% of *D. suzukii* and *D. melanogaster* genomes, respectively. The authors also annotated the TE content for *D. melanogaster* and *D. suzukii running RepeatMasker and Repbase*. Thus, a direct comparison cannot be made, although, as already noted by Rius et al. (2016), we suggest that the differences in TE content from the previous studies (Chiu, et al., 2013; Ometto et al., 2013; Rius et al., 2016) could be related to the applied methodologies, even when *all the analyzed genomes were sequenced* by Illumina Hiseq.

RepeatExplorer runs two broad strategies that are combined for the annotation of TE content: (1) Homology-based searches, which access Repbase library, and (2) de novo strategies, which scan the genome looking for structure and repetitive pattern of TEs. Together, these two strategies achieve better results. Thus, we believe that the results of previous studies using only the homology-based strategy for TE content annotation in D. suzukii and D. melanogaster were underestimated. A recent study from Sessegolo et al. (2016), running a de novo strategy with dnaPipeTE (Goubert et al., 2016), estimated the TE content in D. suzukii and D. melanogaster. The main difference between the software in terms of application and use is that for dnaPipeTE, it is necessary to compile different packages, while RepeatExplorer is available online. Despite this distinction, there is no methodological difference between RepeatExplorer and dnaPipeTE. In the D. suzukii genome, Sessegolo et al. (2016) found approximately 31% of TE content but only near 12% of TE content in D. melanogaster. These results are more similar to our findings (~36% and ~16%, respectively). Thus, we truly believe that the TE content for both species found in the present study is close to the reality found in nature due to the previously noted methodology differences.

In both species, most of these sequences are retrotransposons, in accordance with previous findings that class I elements predominate in *Drosophila* genomes (*Drosophila* Consortium 12 Genomes, 2007). Among DNA transposons, *Helitron* was the most important for both species and the second-largest element in the genome of *D. suzukii*. The percentages of the genomic TE content may still be higher, since the RepeatExplorer has a bias for a medium to high number of copies and more recent elements in the genome (copies more similar to each other). Older elements will have more divergent sequences and may not pass through similarity filters. Nevertheless, since it is a pipeline that uses clustering of reads by similarity, we concluded that the methodology of RepeatExplorer (Novák et al., 2013) is fast and easy to implement as an initial stage after the sequencing by Illumina.

As for the difference between species regarding TE distribution in genes and intergenic regions, the frequency of TEs in genes of *D. suzukii* is 8.6%, whereas in *D. melanogaster*, the distribution of TEs in genes is 41.6% (chi-square, $P < 0.001$). Moreover, 1.0% (103) of the total TE copies and 1.0% (68 copies) of the all *Helitrons* observed in *D. suzukii* are in *CYP* genes, while in *D. melanogaster*, these percentages are 0.2% (87 copies) of the total copies of TEs and 0.2% (19 copies) of the total amount of Helitrons are in *CYP* genes. These differences between species regarding the percentage of TEs and *Helitrons* inserted in *CYP* genes are statistically significant (chi-square, $P < 0.001$).

In *D. suzukii*, *Helitrons* are more abundant in the intergenic region (95.6%), whereas in *D. melanogaster*, more *Helitrons* are found in gene regions (85.9%). It is important to emphasize that the methodology does not allow separating complete *Helitrons* from fragmented copies, which could explain the difference between the species. This may be a particularity of the species and does not invalidate the fact that *D. suzukii* may have more *Helitrons* in intergenic regions.

DISCUSSION

Given the opportunistic nature and the ability of TEs to generate mutations, it is suggested that TEs are important engineers for evolution. Barbara McClintock

(1982) was the first to propose that the activation of TEs in response to stress induces mutations may help the body to adapt to new environmental conditions. Metabolic resistance based on cytochrome P450 is an important adaptation for a variety of insect species, including dipterans (Scott, 1999), and is a common mechanism by which insects develop resistance to pesticides (Feyereisen, 1999). TEs have often been found within or in proximity to resistance genes, providing indirect evidence that transposons are involved in the generation of adaptive genome-related changes in resistance (Catania et al., 2004; Chen and Li, 2007; Chung et al., 2013; Carareto et al., 2013; Casacuberta and González, 2013). In this study, we focused on the search for TEs associated with *CYP*s and on the genome of the successful invasive species *D. suzukii*. In this species, we documented *CYP*s with different TE contents, with TEs carrying putative TFBS and an exon-shuffling pattern probably caused by elements of the rolling-circle type, the *Helitron*s. We also found that the genome of *D. suzukii* has double the TE content of the genome of *D. melanogaster* and that *Helitron* is the most important of the class II DNA transposons.

Considering all TEs in the *CYP* genes studied here, all of these insertions are in flanking regions and introns, reinforcing the view that they are tolerated in non-coding regions. Another possible explanation is that when TEs are inserted close to genes, they can produce new regulatory networks (Feschotte, 2008), and changes in a gene-regulation network are thought to be very important during adaptive evolution (Casacuberta and González, 2013). As stressed above, the *CYP* genes of *D. suzukii* have TE insertions mostly in the 5' flanking region. Some studies have established that TE insertions in the 5'-UTR regions confer resistance to insecticides, especially in the case of the *Drosophila CYP6g1* gene (Daborn et al., 2002; Schmidt et al., 2010). The insertion of the *ACCORD* element in the *CYP6g1* gene has specific transcription enhancers (Chung et al., 2007); *CYP*s of *D. melanogaster* and *D. simulans* accumulate a large number of TE insertions, most of them belonging to the *Helitron* superfamily, which also carries putative TFBS (Carareto et al., 2013). These and other studies have added support to the idea that these elements are gradually co-opted for the regulation of host genes (Chung et al., 2007; Feschotte, 2008).

The possibility of acquiring changes in *cis*-regulatory elements implies that these create an opportunity to respond to new and different environmental factors (Casacuberta and González, 2013). It has been found that several LTR retrotransposons that contain *cis*-regulatory elements are more highly expressed in response to a particular stimulus (Kumar and Bennetzen, 1999). These regulatory sequences are similar to well-characterized motifs necessary for the activation of stress-response genes (Grandbastien et al., 2005). Our study showed that TE fragments carry putative TFBS that could play a role in fly development, such as *Hunchback* (involved in embryo development) and *CF2-II* (involved in cell differentiation). Such a pattern suggests that *CYP*s are permissive to TEs insertions, since these sequences may be donors of transcriptional regulatory signals that may be altering the host gene expression in early and late development. In the literature, there are few in silico incidences for *Hunchback*, *CF2-II* (Carareto et al., 2013), and Zinc finger domain (Thornburg et al., 2006, Babu et al., 2006) binding sites in TE sequences. TEs carrying putative TFBS support the hypothesis that these fragments could influence gene regulation, playing a key role in the adaptation of *Drosophila* species (Feschotte, 2008).

Several other processes that are directly or indirectly related to the presence of TEs in the genomes may also be affecting the coding regions, such as insertions, excisions, retrotranspositions, and exon shuffling. These processes may result in exonization and intronization of TE sequences in the genome and, ultimately, exaptation. If they provide some adaptive advantage, these insertions can even be maintained in the host genome. Feyereisen (1999) suggested two possible mechanisms of resistance to pesticides by *CYP* genes: structural changes in specific *CYP*s, such as gain or loss of exons, and increased expression of *CYP*s. One of the possible ways for an exon to emerge is by exon shuffling. Gilbert (1987), in his theory of exons, proposed that the greater protein diversity found in eukaryotes is the result of exon shuffling. Here, a total of ten *CYP* genes (Figure 3) were observed in structural change, as conserved blocks of exon gain with at least one insertion of the *Helitron* element. The retrotransposition mechanism is one of the factors that result in gene duplications (Holland, 1999). Mobilization of an element carrying gene sequences into a host gene (transduction

or transduplication) may give rise to new exons. The inserted sequence, if it carries splicing sites, can be processed to form alternative transcripts.

Transposons from the superfamilies *Helitron*, *CACTA* and *MULE* have been related to the transduplication of several gene segments in different organisms. *Helitron*s, included in subclass II of DNA transposons, constitute a particularly interesting superfamily, which is known to be involved in exon shuffling, transduplication, and the introduction of novel regulatory elements (Morgante et al., 2005; Pritham and Feschotte, 2007; Thomas et al., 2014). Elements of this unique subclass are mobilized by a different mechanism from the other transposons, the rolling-circle, from the displacement of the single strand of DNA in a loop shape, with subsequent cleavage and reintegration into the genome. These elements show great ability to capture and duplicate as transduplicated gene segments and constitute important genetic modelers in plants (Lopes et al., 2008). In maize, most copies of *Helitron*s have incorporated gene segments, suggesting that they have captured, amplified, and moved hundreds of these genes to various locations in the genome (Yang and Bennetzen, 2009). A recent study of insecticide resistance and *Helitron* in Palmer amaranth (*Amaranthus palmeri*) (Molin et al., 2017) found that in this species, the target-site gene amplification of the EPSPS cassette is an adaptive structural mechanism, which was led by *Helitron*, conferring resistance to glyphosate treatment. Our findings agree with the pattern of gene capture by *Helitron*s.

Molecular time trees estimated the divergence of *D. suzukii* from *D. biarmipes* in 7.3 Ma, *D. takahashii* in a period between 15 and 10 Ma, and latest divergent *D. melanogaster* was 2.5 Ma ago (Ometto *et al*., 2013). Thus, because *D. suzukii* occupy distinct habitats among Drosophilidae, feeding on a diversity of fruits, it is supposed that some classes of gene families in this pest species have evolved differently.

Genomic alterations leading to overexpression of the *CYP* gene were found in only some of the *CYP* genes implicated in insecticide resistance (Li et al., 2007). Previous studies showed that the number of the *CYP* gene family varies among genomes (Thomas, 2007; Chung et al, 2009) and that contraction may have occurred in *D. suzukii* (Chiu et al., 2013). Although *D. suzukii* has a smaller family

of CYP genes than does *D. melanogaster*, previous study has shown that longer-length genes are more important in the production of genomic novelties than are gene families with larger numbers of genes (Grishkevich and Yanai, 2017). This is because a longer gene has more splice variants, and the number of splice variants is inversely proportional to the size of the gene family (Kopelman et al., 2005). Grishkevich and Yanai (2017) suggest that gene length increases due, in part, to transposable elements.

Therefore, analyzing the expression of these longer *CYP* genes in *D. suzukii* and relating them to the inserted TE fragments will be useful in the ongoing search for resistance-management strategies. *Helitron*s are abundant in plant genomes and have been identified in many other eukaryotic genomes (Kapitonov and Jurka, 2007; Kapitonov and Jurka J, 2001). The events of capturing host genes are prominent in maize and have contributed to the evolution of the maize genome but are not well characterized in *Drosophila* genomes (Lal et al., 2009; Barbaglia et al., 2012). TEs are particularly important sequences for exaptation, since they have several regulatory motifs that can be used by the host genome, and they provide material that can evolve and generate evolutionary novelties. Moreover, studying the influence of the *Helitron* superfamily in the genomic context is important to understand the adaptive structural mechanism of this species that may have led to the evolution of this pest. Further studies that discuss the age of *Helitron* insertions, as well as the age of gene branches divergence, will be of great importance to continue the study of the dynamics of this element in the genome of *D. suzukii*, as we had not found any clear evidence for gene capture by Helitrons. Finally, progress in this research may help to elucidate the factors responsible for the successful colonization of the pest species *D. suzukii* and for its insecticide resistance. Studies in this area may assist in the theoretical understanding of the mobility of transposable elements, the evolution of genome size, as well as comparative analyses among genomes of native populations and invasive populations of pest species, with practical applications such as pest management.

Additional files

Additional file 1: Table S1. *CYP* genes with TE insertions, and their flanking regions, in *D. suzukii*. (XLSX 17 Kb)

Additional file 2: Table S2. *CYP* genes without TE insertion, and their flanking regions, in *D. suzukii*. (XLSX 11 Kb)

Additional file 3: Table S3. *CYP* genes with TE insertions, and their flanking regions, in *D. melanogaster.* (XLSX 16 Kb)

Additional file 4: Table S4. *CYP* genes without TE insertion, and their flanking regions, in *D. melanogaster*. (XLSX 9 Kb)

Additional file 5: Table S5. *CYP* genes with TE insertions, and their flanking regions, in *D. biarmipes.* (XLSX 10 Kb)

Additional file 6: Table S6. *CYP* genes with TE insertions, and their flanking regions, in *D. takahashii.* (XLSX 9 Kb)

Additional file 7: Table S7. TFBS identified in fragments of TEs in *D. suzukii* and *D. melanogaster*. (XLSX 17 Kb)

Additional file 8: Table S8. Random genes longer in *Drosophila suzukii* compared to *Drosophila melanogaster* and *Helitron* copies

DECLARATIONS

Ethics approval and consent to participate

Experiments in this research comply with applicable state laws. All institutional and national guidelines for the care and use of laboratory animals were followed.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article (and its additional files). The genome data that support the findings of this study are available in “NCBI-Sequence Read Archive (SRA)” with the identifiers SRR942805 (*D. suzukii*) and SRR1738161 (*D. melanogaster*). The *CYP* gene data that support the findings of this study are available on SpottedWingFlyBase http://spottedwingflybase.org/ (*D. suzukii*) and on FlyBase http://flybase.org/ (*D. melanogaster*).

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

PB performed the study. PB, MD, and VV analyzed the data. PB and SJ performed the statistical tests. PB, SJ, MD, and VV prepared the manuscript. All authors read and approved the manuscript.

Acknowledgements

The authors are grateful to Dr. Elgion Loreto for his inspiring comments, to Msc. Gilberto Cavalheiro Vieira and Dr. Petr Novak for their technical help. This study was supported by research grants and fellowships from CNPq (process n° 455101/2014-0), CAPES, PROPESQ-UFRGS, and FAPERGS.

REFERENCES

Abraham J, Zhang A, Angeli S, Abubeker S, Michel C, Feng Y, et al. Behavioral and Antennal Responses of *Drosophila suzukii* (Diptera: Drosophilidae) to Volatiles From Fruit Extracts. Chem. Ecol. 2015;44:356–67.

Asplen MK, Anfora G, Biondi A, Choi D, Chu D, Daane KM, et al. Invasion biology of spotted wing *Drosophila* (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci. 2015;88:469–94.

Babu MM, Iyer LM, Balaji S, Aravind L. The natural history of the WRKY-GCM1 zinc fingers and the relations between transcription factors and transposons. Nucleic Acids Res. 2006;34(22):6505-6520. doi: 10.1093/nar/gkl888

Barbaglia AM, Klusman KM, Higgins J, Shaw JR, Hannah LC, Lal SK. Gene Capture by *Helitron* Transposons Reshuffles the Transcriptome of Maize. Investigation. 2012;190:965–75.

Bitner-Mathé BC, Victorino J, Faria FS. *Drosophila suzukii* has been found in tropical Atlantic Rainforest in southeastern Brazil. Dros Inf Serv. 2014;97:136–7.

Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, Mckenzie JA, et al. *Cyp12a4* confers lufenuron resistance in a natural population of *Drosophila melanogaster*. PNAS 2005;102:12807–12.

Bryne JC, Valen E, Tang ME, Marstrand T, Winther O, Krogh A, et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res. 2008;36:102–6.

Carareto CMA, Hernandez EH, Vieira C. Genomic regions harboring insecticide resistance-associated *Cyp* genes are enriched by transposable element

fragments carrying putative transcription factor binding sites in two sibling *Drosophila* species. Gene Elsevier B.V.2013;1–7.

Casacuberta E, González J. The impact of transposable elements in environmental adaptation. Mol. Ecol. 2013;22:1503–17.

Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C. World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. Mol. Ecol. 2004;13:2491–504.

Chen S, Li X. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome *P450* genes. BMC Evol. Biol. 2007;7:1–13.

Chiu JC, Jiang X, Zhao L, Hamm CA, Cridland JM, Saelao P, et al. Genome of *Drosophila suzukii* , the Spotted Wing *Drosophila*. G3. 2013;3:2257–71.

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, et al. Cis-Regulatory Elements in the *Accord* Retrotransposon Result in Tissue-Specific Expression of the *Drosophila melanogaster* Insecticide Resistance Gene *Cyp6g1*. Genetics 2007;175:1071–7.

Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P, Daborn PJ. Characterization of *Drosophila melanogaster* cytochrome *P450* genes. PNAS. 2009;106:5731–6.

Cini A, Ioriatti C, Anfora G. A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. Bull.

Insectol. 2012;65:149–60.

ConSite webserver. http://consite.genereg.net/. 2004.

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, et al. A Single *P450* Allele Associated with Insecticide Resistance in *Drosophila*. Science 2002;297:2253–6.

Deprá M, Poppe JL, Toni DC De, Schmitz HJ, Valente VLS. The first records of the invasive pest *Drosophila suzukii* in the South American continent. J Pest Sci. 2014;87:379–83.

Dreves A, Walton V, Fisher G. New Pest Attacking Healthy Ripening Fruit in Oregon. OSU Extension Catalog, EM8991 2009,1-6.

Drosophila Consortium 12 Genomes. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature. 2007;450:203–18.

Escudero LA, Voschi D, Batllor L. *Drosophila suzukii*, una nueva plaga de los frutales. Vida Rural, ISSN 1133-8938 2012;347:18-22.

Feschotte C. The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet. 2008;9:397–405.

Feyereisen R. Insect cytochrome P450. Compr. Mol. Insect Sci. Oxford Elsevier. 2005;4:177.

Feyereisen R. Insect p450 enzymes. Annu. Rev. Entomol. 1999;44:507–33.

FlyBase. http://flybase.org. 2017.

Gilbert W. The Exon Theory of Genes. Cold Spring Harb. Lab. Press. 1987;901–5.

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

Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. A *Helitron* transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat. Commun. 2016;7:1–12.

Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa AP, et al. Stress activation and genomic impact of *Tnt1* retrotransposons in Solanaceae. Cytogenet Genome Res. 2005;110:229–41.

Graveley BR, Brooks AN, Carlson JW, Cherbas L, Choi J, et al. The *D. melanogaster* transcriptome: modENCODE RNA-Seq. 2010.

Grishkevich V, Yanai V. Gene length and expression level shape genomic novelties. Cold Spring Harb. Lab. Press Introd. 2017;1–22.

Guy L, Kultima JR, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010;26:2334–5.

Harrop TWR, Sztal T, Lumb C, Good RT, Daborn PJ, Batterham P, et al. Evolutionary Changes in Gene Expression, Coding Sequence and Copy-Number at the *Cyp6g1* Locus Contribute to Resistance to Multiple

Insecticides in *Drosophila*. PLoS ONE 2014;9(1):e84879.

Hauser M. A historic account of the invasion of *Drosophila suzukii* (Matsumura) (Diptera: Drosophilidae) in the continental United States, with remarks on their identification. Pest Manag Sci. 2011;67:1352–7.

Hemingway J, Hawkes NJ, Mccarroll L, Ranson H. The molecular basis of insecticide resistance in mosquitoes. Insect Biochem. Mol. Biol. 2004;34:653–65.

Holland PWH. Gene duplication: Past, present and future. Cell Dev. Biol. 1999;10:541–7.

Inkscape v0.92.1. http://gitlab.com/inkscape/inkscape. 2017.

Jordan IK, Rogozin IB, Glazko G V, Koonin E V. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 2003;19:68–72.

Joshi NK, Biddinger DJ, Demchak K, Deppen A. First Report of *Zaprionus indianus* (Diptera: Drosophilidae) in Commercial Fruits and Vegetables in Pennsylvania. J. Insect Sci. 2014;14:1–4.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 2005;110:462–7.

Kaneshiro KY. *Drosophila* (*Sophophora*) *suzukii* (Matsumura). Notes exhibitions. Proc. Hawaiian Entomol. Soc. 1983;24:179.

Kapitonov VV, Jurka J. *Helitrons* on a roll: eukaryotic rolling-circle transposons. Trends Genet. 2007;23:521–9.

Kapitonov VV, Jurka J. Rolling-circle transposons in eukaryotes. PNAS. 2001;98:8714–9.

Kopelman NM, Lancet D, Yanai I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat. Genet. 2005;37:588–9.

Kumar A, Bennetzen JL. Plant retrotransposons. Annu. Rev. Genet. 1999;33:479–532.

Lal S, Oetjens M, Hannah LC. *Helitrons*: Enigmatic abductors and mobilizers of host genome sequences. Plant Sci. 2009;176:181–6.

Lasa R, Tadeo E. Invasive Drosophilid Pests *Drosophila suzukii* and *Zaprionus indianus* (Diptera: Drosophilidae) in Veracruz, Mexico. Florida Entomol. 2015;98:987–8.

Leblanc L, Rubinoff D, Montgomery SL. New Immigrant Drosophilidae in Hawaii, and a Checklist of the Established Immigrant Species. Proc. Hawaiian Entomol. Soc. 2009;41:121–7.

Lee JC, Dreves MJ, Cave AM, Kawai S, Isaacs R, Miller JC, et al. Infestation of Wild and Ornamental Noncrop Fruits by *Drosophila suzukii* (Diptera: Drosophilidae). Arthropod Biol. 2015;108:117–29.

Lehmann R. Phenotypic comparison between maternal and zygotic genes

controlling the segmental pattern of the *Drosophila* embryo. 1988;17–27.

Li X, Schuler MA, Berenbaum MR. Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. Annu. Rev. Entomol. 2007;52:231–55.

Lopes FR, Carazzolle MF, Pereira GAG, Colombo CA, Carareto CMA. Transposable elements in *Coffea* (Gentianales: Rubiacea) transcripts and their role in the origin of protein diversity in flowering plants. Mol. Genet. Genomics. 2008;279:385–401.

Marsano RM, Caizzi R, Moschetti R, Junakovic N. Evidence for a functional interaction between the *Bari1* transposable element and the cytochrome *P450Cyp12a4* gene in *Drosophila melanogaster*. Gene 2005;357:122–8.

Mcclintock B. The Significance of Responses of the Genome to Challenge. Science (80)1982;226:792–801.

Mckenzie JA, Batterham P. The genetic, molecular and phenotypic consequences of selection for insecticide resistance. TRENDS Genet. 1994;9.

Molin WT, Wright AA, Lawton-rauh A, Saski CA. The unique genomic landscape surrounding the *EPSPS* gene in glyphosate resistant *Amaranthus palmeri*: a repetitive path to resistance. BMC Genomics [Internet]. BMC Genomics; 2017;18:1–16. Available from: http://dx.doi.org/10.1186/s12864-016-3336-4

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. Gene duplication and exon shuffling by *helitron-like* transposons generate intraspecies diversity in maize. Nat. Genet. 2005;37:997–1002.

Müller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, et al. Field-Caught Permethrin-Resistant *Anopheles gambiae* Overexpress *CYP6P3*, a *P450* That Metabolises Pyrethroids. PLoS Genet. 2008;4(11):e1000286.

National Center for Biotechnology Information (NCBI). https://www.ncbi.nlm.nih.gov/. 1988.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013;29:792–3.

Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. Nature. 1980;287:795–801.

Ometto L, Cestaro A, Ramasamy S, Grassi A, revadi S, et al. Linking Genomics and Ecology to Investigate the Complex Evolution of an Invasive *Drosophila* Pest. Genome Biol Evol 2013;5(4):745-757.

Paula, MA, Lopes PHS, Tidon R. First record of *Drosophila suzukii* in the Brazilian Savanna. Dros. Inf. Serv. 2014;97:113–5.

Pritham EJ, Feschotte C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. PNAS. 2007;104:1895–900.

Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a Cytochrome *P450* Gene Is Associated with Resistance to Neonicotinoid Insecticides in the Aphid *Myzus persicae*. PLoS Genet. 2010;6(6):e1000999.

Quinlan AR (2014) BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics 47:11.12.1-11.12.34. doi: 10.1002/0471250953.bi1112s47

RepeatMasker Open-4.0. http://www.repeatmasker.org. 2014.

Rewitz KF, O'Connor MB, Gilbert LI. Molecular evolution of the insect *Halloween* family of cytochrome *P450*s: Phylogeny, gene organization and functional conservation. Insect Biochem. Mol. Biol. 2007;37:741–53.

Rius N, Guillén Y, Delprat A, Kapusta A, Feschotte C, Ruiz A (2016) Exploration of the *Drosophila buzzatii* transposable element content suggests underestimation of repeats in *Drosophila* genomes. BMC Genomics 17(1):344.

Rota-Stabelli O, Blaxter M, Anfora G. *Drosophila suzukii*. Curr. Biol. Elsevier; 2013;23:R8–9.

Schlenke TA, Begun DJ. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. PNAS 2004;101:1626–31.

Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, et al. Copy Number Variation and Transposable Elements Feature in Recent, Ongoing Adaptation at the Cyp6g1 Locus. Plos Genet. 2010;6:1–11.

Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem. Mol. Biol. 1999;29:757–77.

Sessegolo C, Burlet N, Haudry A. Strong phylogenetic inertia on genome size and transposable element content among 29 species of flies. Biol Lett 2016,12:20160407. DOI: 10.1098/rsbl.2016.0407.

Shea MJ, King DL, Conboy MJ, Mariani BD, Kafatos FCK. Proteins that bind to *Drosophila* chorion cis-regulatory elements: A new C2H2 zinc finger protein and a C2C2 steroid receptor-like component. Genes Dev. 1990;4:1128–40.

Silva AF, Dezordi FZ, Wallau GL. Manual para caracterização genômica e análise evolutiva de elementos transponíveis utilizando diretamente reads de sequenciadores de alto desempenho. SBG Ribeir. 2016.

SpottedWingFlyBase. http://spottedwingflybase.org. 2013.

Thomas J and Pritham EJ. Helitrons, the eukaryotic rolling-circle transposable elements. Microbiol Spectr. 2015;3:MDNA3-0049-2014. doi: 10.1128/ microbiolspec.MDNA3-0049-2014

Thomas J, Phillips CD, Baker RJ, Pritham EJ. Rolling-Circle Transposons Catalyze Genomic Innovation in a Mammalian Lineage. GBE. 2014;6:2595–610.

Thomas JH. Rapid Birth–Death Evolution Specific to Xenobiotic Cytochrome P450 Genes in Vertebrates. PLoS Genet. 2007;3(5):e67.

Thornburg BG, Gotea V, Makałowski W. Transposable elements as a significant

source of transcription regulating signals. Gene. 2006;365:104–10.

Timmeren S Van, Isaacs R. Control of spotted wing *Drosophila*, *Drosophila suzukii*, by specific insecticides and by conventional and organic crop protection programs. Crop Prot. Elsevier Ltd; 2013;54:126–33.

van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. TRENDS Genet. 2003;19:530–6.

Vilela CR, Mori L. The invasive spotted-wing *Drosophila* (Diptera, Drosophilidae) has been found in the city of São Paulo (Brazil). Rev. Bras. Entomol. 2014;58:371–5.

Walsh DB, Bolda MP, Goodhue RE, Dreves AJ, Lee J, Bruck DJ, et al. *Drosophila suzukii* (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding its Geographic Range and Damage Potential. J. Integr. Pest Manag. 2011;2:1–7.

Yang L, Bennetzen JL. Distribution, diversity, evolution, and survival of *Helitrons* in the maize genome. PNAS. 2009;106:19922–7.

Zhu F, Parthasarathy R, Bai H, Woithe K, Kaussmann M, Nauen R, et al. A brain-specific cytochrome P450 responsible for the majority of deltamethrin resistance in the QTC279 strain of *Tribolium castaneum*. PNAS. 2010;107:8557–62.

Figure captions

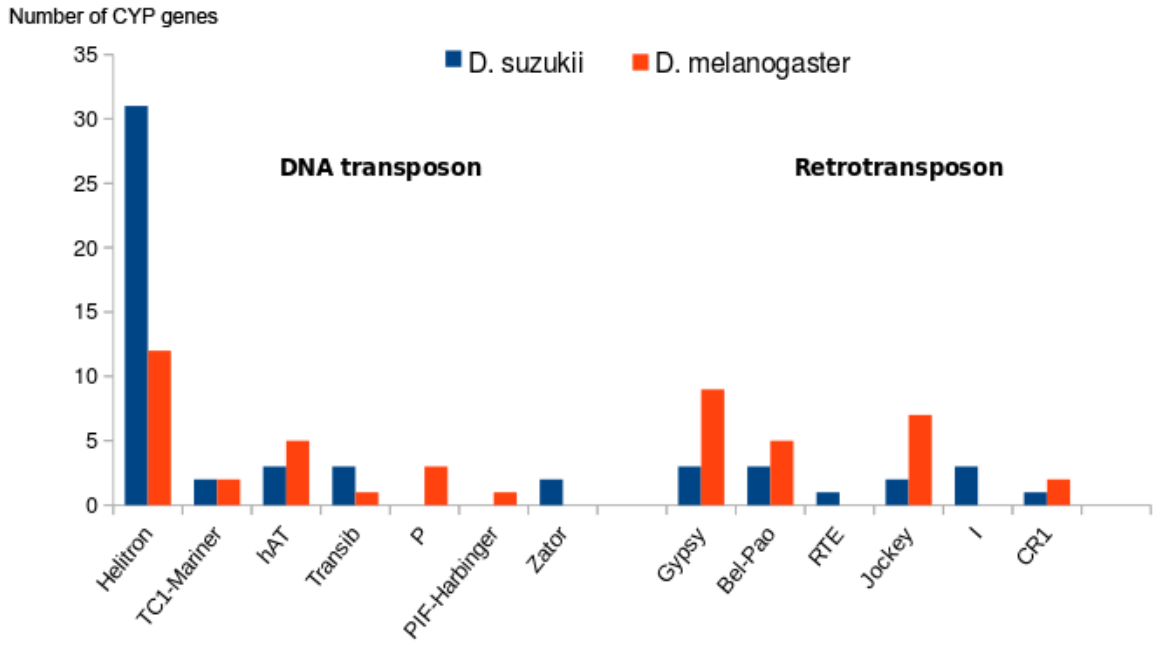

Fig. 1 Number of *CYP* genes within transposable element insertions in *Drosophila suzukii* and *Drosophila melanogaster*.

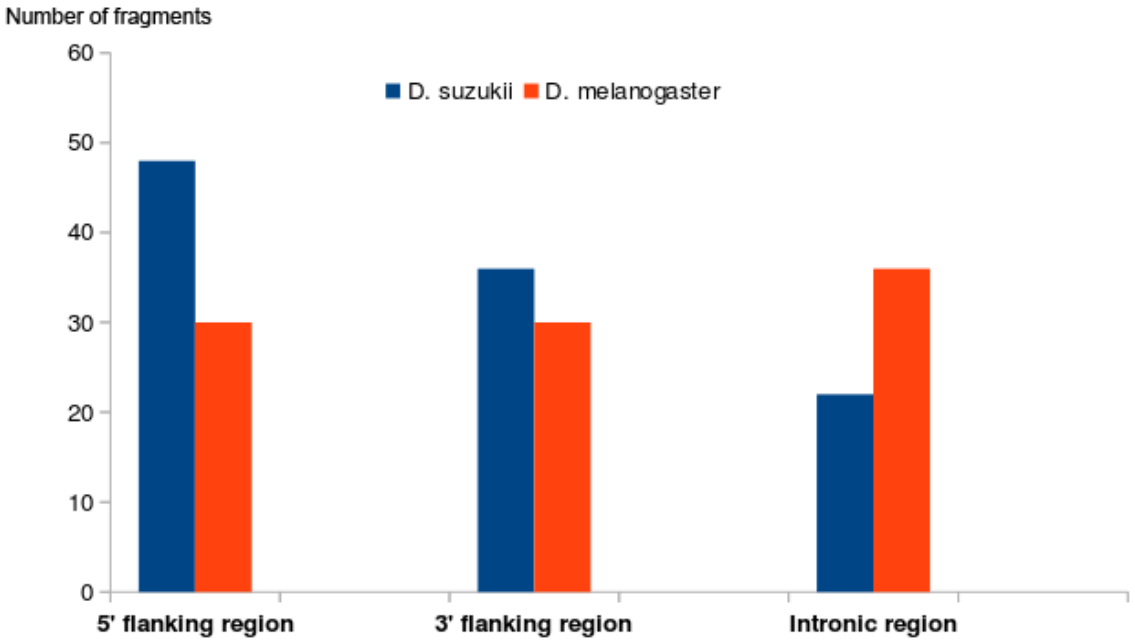

Fig. 2 Insertion position (5'- and 3'-flanking region, and intron region) of transposable elements in *Drosophila suzukii* and *Drosophila melanogaster*.

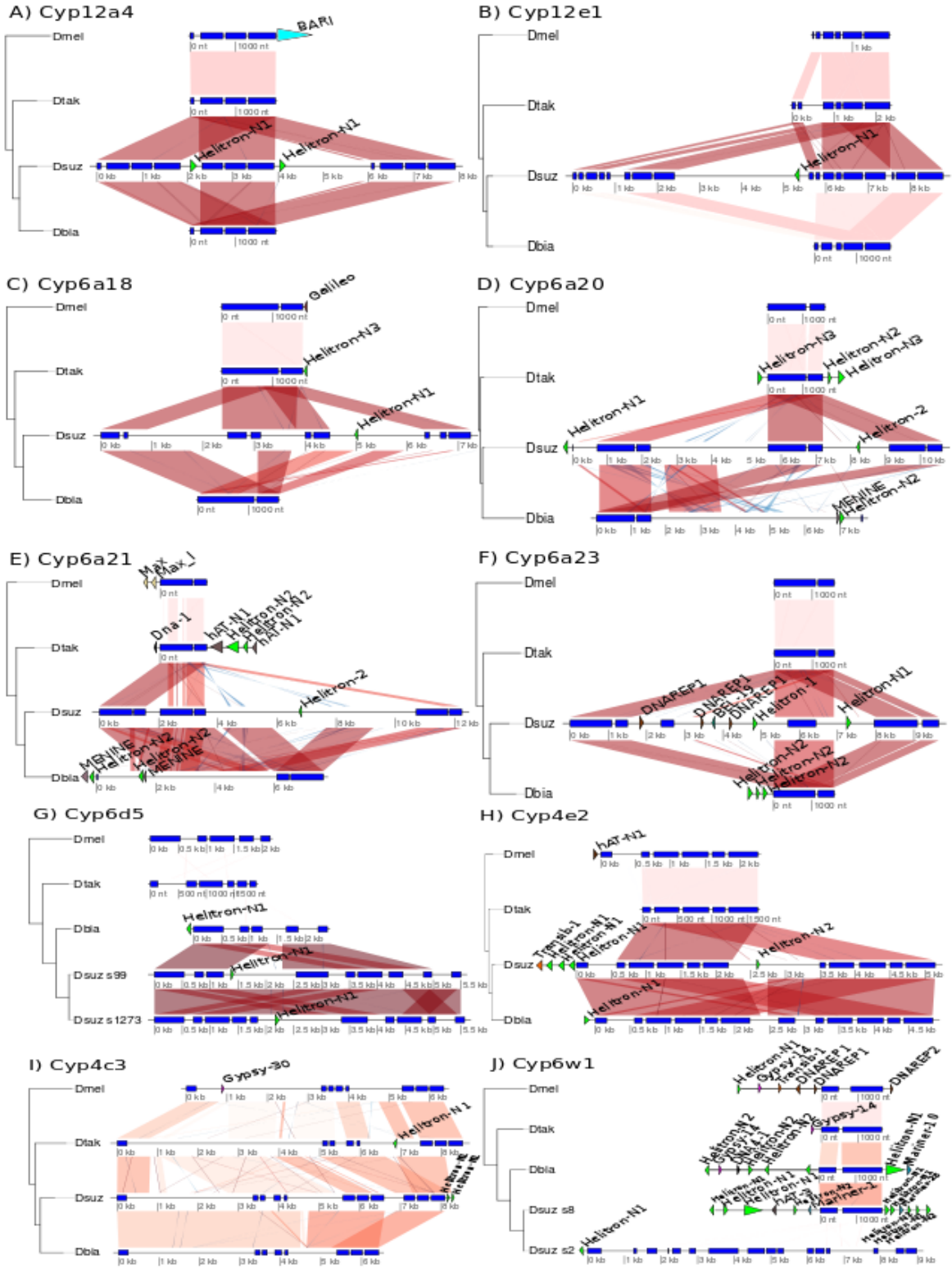

Fig. 3 Comparative analysis showing similarity between *CYP* genes in *Drosophila* species. The intensity of red boxes between genes highlights the

closest sequence above for which genes are denoted. The multiple transposable element insertions and their orientation are represented by triangles. The phylogeny on the left was inferred by maximum-likelihood methodology [40]. Genes are scaled to real length, except for flanking regions. Dmel, *Drosophila melanogaster*; Dtak, *Drosophila takahashii*; Dsuz, *Drosophila suzukii*; Dbia, *Drosophila biarmipes*; s99, scaffold 99; s1273, scaffold 1273; s8, scaffold; s2, scaffold 2.

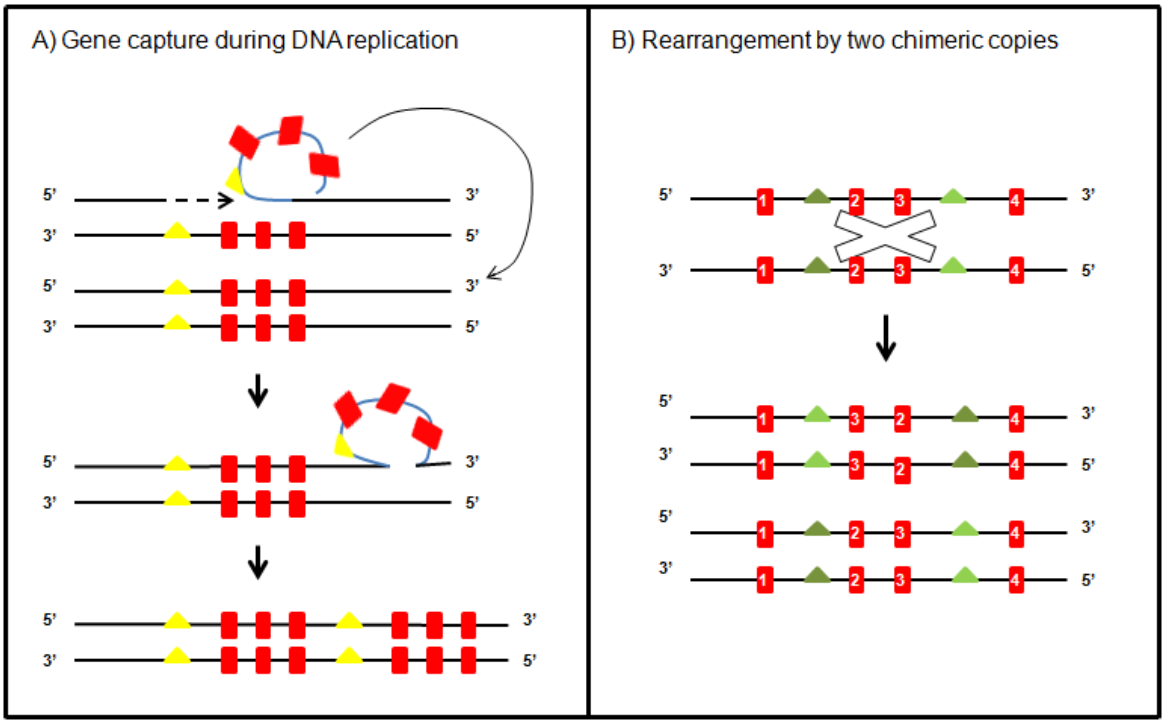

Fig. 4 Hypothetical exon shuffling by rolling-circle transposon: A) a longer gene formed by *Helitron* during its transposition; B) the hole of two *Helitron* copies rearranging due to the similarity in the sequences.

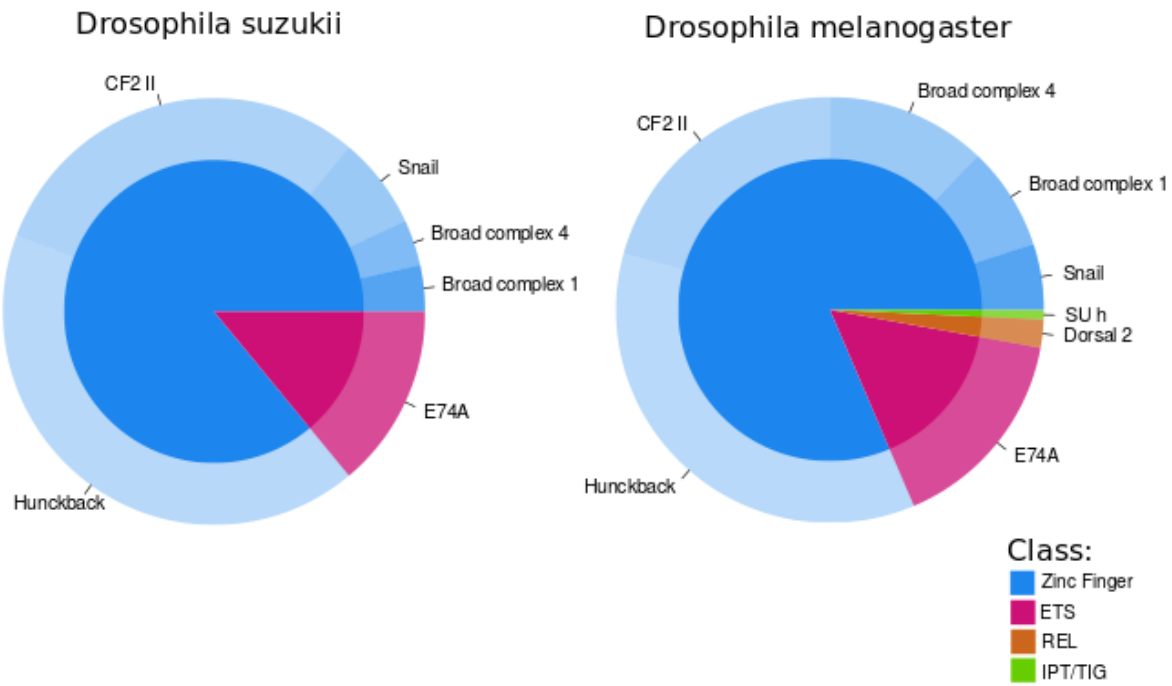

Fig. 5 Putative TFBS predicted for transposable elements inserted in *CYP* genes of *Drosophila suzukii* and *Drosophila melanogaster*.

Table 1. Transposable element fragments belonging to subclasses and orders in *CYP* genes and flanking regions.

		D. suzukii	*D. melanogaster*
Class I (retrotransposon)	LTR	12 (11.6%)	40 (46%)
	NON-LTR	12 (11.6%)	15 (17.2)
Class II (DNA transposon)	Subclass 1	11 (10.6%)	13 (15%)
	Subclass 2	68 (66%)	19 (21.8%)
	TOTAL	103 (100%)	87 (100%)

Long Terminal Repeat (LTR) = *Gypsy* and *Bel-Pao* superfamilies

Non-LTR = *RTE*, *I*, *Jockey* and *CR1* superfamilies

Subclass 1 = *TC1-Mariner*, *hAT*, *Transib*, *P*, *PIF-Harbinger* and *Zator* superfamilies

Subclass 2 = *Helitron* superfamily

Table 2. Genomic TE content in *Drosophila suzukii* and *Drosophila melanogaster.*

		D. suzukii	*D. melanogaster*
Class I (retrotransposon)	*Copia*	0.05%	0.37%
	Bel-Pao	4.67%	2.67%
	Gypsy	9.85%	5.44%
	LINE	7.00%	4.92%
	Kiri	0.02%	0.00%
	Outcast	0.02%	0.00%
Class II (DNA transposon)	*Tc1-mariner*	0.83%	0.30%
	hAT	0.76%	0.07%
	Transib	0.49%	0.16%
	PiggyBac	0.27%	0.00%
	CACTA	0.22%	0.00%
	PIF-Harbinger	0.05%	0.00%
	P	0.00%	0.37%
	Helitron	7.27%	0.45%
	Maverick	4.26%	0.00%
	Unknown	0.19%	1.21%
	TOTAL	35.94%	15.96%

CHAPTER 3

Final considerations

As data continue to accumulate over the next several years, the present study should be in a better position to evaluate definitively the role played by *Helitron* insertions in *CYP* gene family, as well as the role of transposable elements in shaping the genome and evolution of *D. suzukii*. Nevertheless, based on presently available evidence, it seems clear that the once popular notion that TEs are merely junk DNA and without evolutionary consequence is no longer tenable. On the contrary, these repetitive sequences are critically important to the emergence of phenotypic novelties over evolutionary time.

Abstract

In silico analyses were performed to evaluate a possible connection between *CYP* genes family, genome, and transposable elements of a non-pest species (*D. melanogaster*) and a pest species (*D. suzukii*). I found *Helitron* fragments accumulated in flanking regions of *CYP*s, and their transposition may have resulted in the capture of the flanking sequence, with consequent transduplication of the gene. *D. suzukii* genome carries more TEs than the genome of *D. melanogaster*, as well as the Helitron superfamily, is over-represented in the genome of the first species. I also found putative transcription-factor binding sites in TE fragments, which reinforces the idea that TEs may influence gene regulation.

Resumo

Foram realizadas análises *in silico* para avaliar uma possível conexão entre a família de genes *CYP*, o genoma e os elementos transponíveis de uma espécie não praga (*D. melanogaster*) e uma espécie praga (*D. suzukii*). Eu encontrei fragmentos de *Helitron* acumulados em regiões flanqueadoras de *CYP*s, e sua transposição pode ter resultado na captura da sequência flanqueadora, com consequente rearranjo do gene. O genoma de *D. suzukii* carrega mais TEs do que o genoma de *D. melanogaster*, bem como a superfamília de Helitron está representada em grande parte no genoma da primeira espécie. Eu também encontrei putativos sítios de ligação para fatores de transcrição nos fragmentos de transposons, o que reforça a ideia de que TEs podem influenciar na regulação dos genes.