Universitat Autònoma de Barcelona

Transposable element misregulation in Drosophila buzzatii–Drosophila koepferae interspecific hybrids

Valèria Romero Soriano

Transposable element misregulation in *Drosophila buzzatii–Drosophila koepferae* interspecific hybrids

Valèria Romero Soriano

Departament de Genètica i Microbiologia

Universitat Autònoma de Barcelona

A thesis submitted for the degree of

Doctor in Genetics (PhD)

Director: *Dr. Maria Pilar García Guerreiro*

June 2016

Memòria presentada per la Llicenciada en Biotecnologia Valèria Romero Soriano per a optar al títol de Doctora en Genètica de la Universitat Autònoma de Barcelona.

Valèria Romero Soriano

Bellaterra, a 15 de juny de 2016

La Doctora María del Pilar García Guerreiro, Professora agregada del Departament de Genètica i Microbiologia de la Facultat de Biociències de la Universitat Autònoma de Barcelona,

Certifica que Valèria Romero Soriano ha dut a terme sota la seva direcció el treball de recerca realitzat al Departament de Genètica i Microbiologia de la Facultat de Biociències de la Universitat Autònoma de Barcelona, que ha portat a l'elaboració d'aquesta Tesi Doctoral, titulada «**Transposable element misregulation in *Drosophila buzzatii–Drosophila koepferae* interspecific hybrids**».

Per a que consti als efectes oportuns, signa el present certificat a Bellaterra, a 15 de juny del 2016.

Maria del Pilar García Guerreiro

A tots aquells que no hi han pogut ser.

Al somriures ja llunyans d'un Manel, i de l'altre (que de la foto era el més alt).

A l'inoblidable Mercè, i a la meravellosa Pepita.

Ens quedarem aquí baix. A tot estirar, cregui'm, ens espera l'improbable paradís del registre fòssil.

Jesús Moncada (*Calaveres atònites*, 1999)

Contents

Index of Figures

Index of Tables

Abstract

Transposable elements (TEs) are mobile genetic units present in almost all the eukaryotic sequenced genomes. Their mobilizing capacity, together with their repetitive nature, makes them powerful endogenous mutators able to create novel genetic variants, which will be then subject to selection. However, their mutagenic potential can also endanger their host's fitness, which has led to the development of several regulatory strategies against TE mobilization in eukaryotic organisms. These are especially important in the germline, where mutations can be transmitted to the offspring. In *Drosophila* ovaries, TEs are mainly regulated by a small RNA-mediated silencing mechanism, the piRNA (Piwi-interacting RNA) pathway, which affects transcriptional and post-transcriptional TE silencing. This strong regulation can be relaxed under several stress conditions, including interspecific hybridization, a genomic stressor that promotes TE mobilization. Several cases of transposition events have been described in hybrids of different species, including both animals and plants. In the case that concerns us, *D. buzzatii–D. koepferae* hybrids, a previous survey in our group detected mobilization of at least 28 TEs. However, the molecular mechanisms underlying this TE release remain elusive, although recent studies on hybrid TE expression seem to point to a transcriptional deregulation. Furthermore, little is known about the effects this phenomenon can have in the genome of the hybrid progeny. In this work, we first assess the impact that hybridization-induced TE proliferation has on the genome size of *D.buzzatii–D. koepferae* hybrids, throughout four generations of hybridization (an interspecific cross followed by three backcrosses). We demonstrate the existence of a sex-specific genome expansion, that affects only females at the first backcross. These results provide the first evidence of genome size increase in interspecific hybrids of animal species. We hypothesize that a TE deregulation at a transcriptional level occurs in F1 females, leading to new TE insertions that result in a genome size increase in the following generation. In order to test this hypothesis, we address two TE expression studies in the same hybrids, using two different approaches. First, we perform an in-depth analysis of the expression of one of the mobilized transposons, *Helena*, in both sexes and different tissues. We show that *Helena* expression in somatic tissues is not altered after hybridization, whereas in gonads sex-biased effects are observed. Indeed, *Helena* is repressed in F1 testes, in concordance with the unaltered genome size in males. In ovaries, an early *Helena* overexpression seems to occur in young flies, being then controlled in older ones. We subsequently performed a global analysis using a transcriptomic approach, in order to evaluate if the results for *Helena* could be extended to other TEs. To

disentangle the molecular mechanisms involved in TE deregulatiom, we analysed the piRNA populations of parental species and hybrids. We show that F1 testes indeed tend to present a TE expression lower than *D. buzzatii*, which is coupled with a global increase of piRNA amounts. In ovaries, TE overexpression is the more common effect, and seems to be mainly due to differences in piRNA production strategies between parental species. Actually, the piRNA pathway proteins are divergent between parental species and could be at the origin of the hybrid instability. Moreover, differences in piRNA amounts between *D. buzzatii* and *D. koepferae* cytoplasms could also account for some cases of deregulation, as occurs in hybrid dysgenesis syndrome. Finally, other explanations are needed to account for the whole pattern of deregulation, such as the failure of histone modification's deposition or of other TE silencing pathways.

Resum

Els elements transposables (ETs) són unitats genètiques mòbils presents en pràcticament tots els organismes eucariotes seqüenciats. La seva capacitat de moure's, juntament amb el seu caràcter repetitiu, els converteix en importants mutàgens amb l'habilitat de crear noves variants genètiques susceptibles a la selecció. Donat que el seu potencial mutagènic pot posar en perill la *fitness* de l'hoste, els organismes eucariotes han desenvolupat diferents estratègies de regulació per controlar la mobilització d'ETs. Cal destacar la importància d'aquestes estratègies en línia germinal, on les mutacions poden ser transmeses d'una generació a l'altra. En ovaris de *Drosophila*, el principal mecanisme de regulació d'ETs és la via dels piRNAs, que contribueix al seu silenciament transcripcional i post-transcripcional. La forta regulació a la que els ETs estan sotmesos es pot veure relaxada sota diferents condicions d'estrès, com és el cas de la hibridació interespecífica. Diversos estudis han descrit noves insercions d'ETs en híbrids interespecífics, tant d'animals com de plantes. En el cas que ens ocupa, els híbrids de *Drosophila buzzatii* i *Drosophila koepferae*, investigacions prèvies del nostre grup van detectar la mobilització d'almenys 28 ETs. No obstant, els mecanismes responsables d'aquesta activació són encara desconeguts, tot i que els estudis més recents del camp semblen apuntar a una desregulació a nivell d'expressió. També es desconeixen els efectes que la proliferació d'ETs pot tenir sobre el genoma dels híbrids. En aquest treball, comencem avaluant l'impacte de la hibridació sobre la mida del genoma dels híbrids de *Drosophila buzzatii* i *Drosophila koepferae* al llarg de quatre generacions d'encreuaments híbrids (un primer d'interespecífic seguit de quatre retroencreuaments). Demostrem l'existència d'una expansió genòmica sexe-específica, que afecta només les femelles del primer retroencreuament. Aquests resultats representen la primera evidència d'un augment de la mida del genoma en híbrids interespecífics d'espècies animals. La nostra hipòtesi és que una desregulació a nivell transcripcional té lloc a les femelles de la F1, donant lloc a noves insercions que es detecten a la següent generació. Per tal de testar aquesta hipòtesi, hem realitzat dos estudis d'expressió d'ETs, emprant dues aproximacions diferents. Primer, duem a terme una anàlisi en profunditat de l'expressió del retrotransposó *Helena* (un dels ETs que transposen en els nostres híbrids) en ambdós sexes i diferents teixits. Demostrem que l'expressió d'*Helena* en teixit somàtic no és alterada degut a la hibridació, mentre que en gònades s'observen efectes sexe-específics. En testicles de la F1, observem una repressió *d'Helena*, concordant amb l'absència de canvi en la mida del genoma dels mascles. En ovaris, sembla que *Helena* es desregula en mosques joves, però els nivells d'expressió

baixen en mosques de major edat. Posteriorment, descrivim una anàlisi a nivell transcriptòmic, on s'avalua si els resultats *d'Helena* són extrapolables a l'expressió global dels ETs. Per esbrinar quins mecanismes estan involucrats en la desregulació d'ETs, analitzem també les poblacions de piRNAs d'espècies parentals i híbrids. Els nostres resultats demostren que els testicles de la F1 tendeixen a presentar nivells d'expressió més baixos que *D. buzzatii*, probablement degut a un augment dels nivells de piRNAs. En ovaris, l'efecte més comú és la sobreexpressió d'ETs, que podria ser explicada per incompatibilitats en la via dels piRNA entre les dues espècies parentals. De fet, les proteïnes d'aquesta via es troben entre les més divergents entre les dues espècies. D'altra banda, alguns casos de desregulació poden ser explicats per diferències entre els nivells de piRNAs entre els citoplasmes de *D. buzzatii* i *D. koepferae*, com en el cas de la disgènesi híbrida. Finalment, cal destacar que són necessàries altres explicacions per explicar el patró global de desregulació, com ara un funcionament anormal d'altres vies de regulació d'ETs o de la modificació d'histones.

1 Introduction

In the present section, I will explain the conceptual basis underlying the study of transposable element expression and regulation in hybrids between *Drosophila* species. First, I will introduce natural hybridization and its importance in the evolution of eukaryotes, focusing on the mechanisms that govern the occurrence of hybrids in nature. Chief among the hybridization effects, we find the genetic instability triggered by transposable element activation. I will then present the model species used in this study, *Drosophila buzzatii* and *Drosophila koepferae*, a pair of cactophilic sibling species that live in sympatry in vast areas of South America; whose hybrids present an increase of transposition rates. Finally, I will discuss the biology of transposable elements, emphasizing their potential as an evolutionary force as well as the molecular mechanisms governing their expression and propagation in their host genomes.

1.1 Interspecific hybridization

'We used to make fun of Edgar Anderson by saying that he was finding hybrids under every bush. Then we realized that even the bushes were hybrids.'

Warren H. Wagner

1.1.1 Natural hybridization and evolution

Natural hybridization (or reticulation) refers to successful matings in nature between individuals from two populations that can be distinguished on the basis of at least one heritable character (Arnold 1997). In the case of interspecific hybridization, the mentioned individuals belong to two different taxa that have been defined as distinct species. This definition depends on the species concept employed, the most popular being the biological species concept (Dobzhansky 1937; Mayr 1942), which states that "species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups". According to this concept, the complete achievement of speciation is based on the development of solid barriers to reproduction. Thus, the occurrence of hybrids in nature is considered merely a mistake related to incomplete speciation, uncommon and without evolutionary effects due to a strong negative selection against introgressed phenotypes (Mayr 1963). However, we know nowadays that 10-30% of multicellular plant and animal species are involved in ongoing hybridization events, suggesting that the potential of natural hybridization as an evolutionary force has long been underestimated (Mallet 2005; Abbott et al. 2013).

Historically, investigations on hybridization were mainly used either to infer evolutionary relationships between taxa or to disentangle mechanisms that limit gene flow in order to understand speciation processes (Arnold 1997). A third approach, initially proposed by E. Anderson and G. L. Stebbins, conceded natural hybridization to be of evolutionary significance by itself, focusing on its ability to produce novel genotypes that are subject to selection (Anderson and Stebbins 1954). This last viewpoint was mainly supported by botanists, who emphasized the potential of hybrid genotypes to result in adaptive evolution and originate new evolutionary lineages. Indeed, natural hybridization in plants is considered to be widespread, especially in angiosperms, that are frequently allopolyploid (more than 50% of species would be of hybrid origin, Arnold 1997). Although studies of the fossil record point out the influence of hybridization in plants along

extended evolutionary periods, the distribution of reticulation among plant taxonomic groups is heterogeneous, and seems to depend on biological and geographical features. For instance, sympatry affects the propensity of species to hybridize by facilitating the occurrence of heterospecific matings.

A very well-studied case of plant hybridization concerns the sunflower genus *Helianthus*, in which several examples of homoploid hybrid speciation have occurred. At least three species of hybrid origin have arisen independently from the cross of *H. annuus* and *H. petiolaris*: *H. paradoxus* (Rieseberg et al. 1990), *H. anomalus* and *H. deserticola* (Rieseberg 1991). Interestingly, while parental species coexist in arid zones of central and western United States (with different soil preference), the three hybrid species are found in more restricted and extreme environments (Rieseberg et al. 1990; Rieseberg et al. 2003). For instance, *H. paradoxus* occurs only in brackish marshes in Texas and New Mexico, presenting several adaptive traits that attenuate the toxic effects of sodium (Rieseberg et al. 2003). These traits, such as leaf succulence and mineral ion uptake, are strongly selected in *H. annuus–H. petiolaris* synthetic hybrids when transplanted in *H. paradoxus* habitat (Rieseberg et al. 2003). This example highlights the creative role of natural hybridization in evolution, which can lead to the invasion of novel habitats (Seehausen 2004). Furthermore, the genomic composition of experimental hybrids between *H. annuus* and *H. petiolaris* was shown to be concordant with *H. anomalus* one, suggesting that selection plays indeed an important role in hybrid formation and speciation (Rieseberg et al. 1996). Finally, evidence of introgression in other *Helianthus* species (such as *H. bolanderi*, *H. exilis* and *H. debilis*) confirms that the evolution of this genus has been shaped, and still is, by reticulation events (Rieseberg 1991; Rieseberg et al. 2007).

Contrary to botanists, zoologists used to plead that natural hybridization was maladaptive and lacked evolutionary importance because heterospecific crosses' progeny is scarce and generally sterile. Actually, we must note that the impact of rare, incidental events on the pattern of organismal evolution cannot be dismissed. Furthermore, as already mentioned, the frequency of hybridization in animal taxa has been proven to be higher than previously thought (Mallet 2005). Since sterile hybrids can indeed be considered evolutionary dead ends, I will focus on examples of animal crosses producing viable F1 progeny with some degree of fertility. As in plants, the occurrence of hybridization is taxonomically widespread among animals, but unequally distributed. For instance, fishes have traditionally been the focus of hybridization surveys (Hubbs 1955), because they frequently hybridize in both freshwater and marine habitats (Gardner 1997; Scribner et al. 2000; Montanari et al. 2016). Another extensively studied taxonomic group are birds, such as Darwin's

finches (*Geospiza*), whose evolutionary history is highly influenced by introgression (Grant and Grant 2002; Grant and Grant 2008), and whose hybrids sometimes present higher fitness than parental species (Grant and Grant 1992). Although less frequently, hybridization has also been reported in different mammals like macropods (O'Neill et al. 1998; Metcalfe et al. 2007), dolphins (Amaral et al. 2014), bats (Larsen et al. 2010) and wolves (Anderson et al. 2009).

In *Drosophila*, crossability between closely related species is widespread. Successful interspecific matings in the laboratory have been described for all groups of both *Drosophila* and *Sophophora* subgenera (Bock 1984), but reported cases of natural hybridization are scarce. The most conspicuous examples concern different species pairs of Hawaiian *Drosophila*, such as *D. heteroneura* and *D. silvestris*, whose F1 and backcrossed hybrids have been collected in all the island localities where they live in sympatry (Kaneshiro 1990). Another evidence of natural introgression between *Drosophila* species was ascertained by the analysis of mitochondrial DNA of *D. simulans* and *D. mauritania* (Aubert and Solignac 1990; Ballard 2000), which showed that natural hybridization had also occurred between those species. Therefore, as in other animals, the evolutionary history of *Drosophila* species has been affected by reticulation events.

1.1.2 Reproductive isolation

Hybridization episodes are influenced by environmental factors, but also by the relative difficulty in producing hybrids at each generation (Arnold 1997). In fact, despite the important incidence of hybridization in nature, several isolating mechanisms are set up to prevent gene flow between species. These have to be overcome to produce (somewhat fertile) hybrids, but they are afterwards useful for the stabilization of hybrid lineages (Arnold 1997). Isolating barriers can act before and after fertilization, and only their combined accumulative action leads to complete reproductive isolation and speciation (Coyne and Orr 2004).

1.1.2.1 Prezygotic barriers

Among pre-fertilization barriers, we can classify those concerning habitat, pollinator and temporal isolation as ecological barriers (Coyne and Orr 2004). Habitat isolation reduces the probability of reproductive encounters between heterospecific individuals through spatial separation. It is based on genetic differences related to adaptation and does not necessarily involve geographic isolation. Indeed, allopatric species are considered ecologically isolated in a *macrospatial* form; whereas species that coexist in the same general area but have different ecological preferences are isolated in a *microspatial* form. An example of microspatial habitat isolation is the case of *Bombina* toad

hybrid zones in Croatia, where there is an important association between genotype and habitat: *B. bombina* alleles are more frequent in ponds (semipermanent water), whereas *B. variegata* ones are more common in temporary puddles (MacCallum et al. 1998).

In angiosperms, interspecific gene flow is also limited by the use of different pollinators (Coyne and Orr 2004). Isolation can be based on differential visitation by pollinators (either genetic or learned), which is called *ethological* pollinator isolation (Grant 1994). Otherwise, mechanical pollinator isolation is caused by morphological differences between species of flowers or pollinators, hindering cross-pollination (Grant 1994). A well-documented case of sympatric pollinator isolation concerns the monkeyflower genus *Mimulus*, where *M. cardinalis* (red, tubular flowers with high nectar volume) is pollinated almost exclusively by hummingbirds; while the closely related *M. lewisii* (broad, pink flowers with low nectar volume) uses mostly bees as pollinators (Schemske and Bradshaw 1999). The third and last ecological prezygotic barrier is temporal isolation (or allochrony). In plants, allochrony impedes interspecific gene flow due to differential flowering or pollen shedding periods between species; whereas in animals, it concerns mainly mating season and spawning time (Coyne and Orr 2004). For instance, a species of tropical Atlantic coral, *Montastraea franksi*, which lives in sympatry with its sister species *M. annularis*, spawns only two hours earlier –a sufficient delay for *M. franksi* gametes to lose viability (Levitan et al. 2004). Temporal isolation can be partly genetic but it can also be due environmental factors, the magnitude of time differences depending on the species under study (Coyne and Orr 2004).

On the other hand, nonecological prezygotic barriers include behavioural, mechanical and gametic isolation mechanisms that act before fertilization (Coyne and Orr 2004). Behavioural isolation is restricted to animals and is due to the reduction of sexual attraction between individuals of distinct species, reducing their probability of mating. This barrier demands the interaction between traits of different sexes: typically, males produce a signal that is preferentially recognized by conspecific females. For instance, differences in male wing patterns between the butterflies *Pieris occidentalis* and *P. protodice* (the former with darker forewings) act as a reproductive barrier that can be bypassed by artificially increasing wing melanization in *P. occidentalis* males (Wiernasz and Kingsolver 1992). Studies in *Drosophila* show the existence of a genetic basis of behavioural isolation, involving distinct sensory signals such as male courtship song, pheromones and particular morphologies (Coyne and Orr 2004; Laturney and Moehring 2012; Fan et al. 2013). Among the genes involved in these traits, the most well-known is the period (*per*) gene, which affects species-specificity in courtship song's rhythm (Kyriacou and Hall 1980; Wheeler et al. 1991).

Mechanical isolation consists in the inhibition of fertilization through incompatibilities between interspecific reproductive structures (Coyne and Orr 2004). This occurs in both plants and animals and can be structural or mediated by contact (Masly 2012). For instance, matings between *Drosophila simulans* and *D. mauritiana* are of abnormally short duration, probably due to interspecific differences in male genitalia shape that females are able to recognize, leading to sperm transfer interruption (Coyne 1993). Finally, gametic isolation barriers are those acting after gamete release (pollination, spawning or copulation) and before fertilization (Coyne and Orr 2004). For example, in *Drosophila*, male ejaculation produces an insemination reaction in females that leads to a swelling of their vagina (Alonso-Pimentel et al. 1994). The mass produced in this reaction disappears after a few hours in conspecific matings, but it lasts longer in interspecific ones, preventing egg formation (Asada and Kitagawa 1988) and causing ultimately female sterility or death (Marin et al. 1993).

1.1.2.2 Postzygotic barriers

Reproductive isolation mechanisms that act after fertilization, such as hybrid inviability and sterility, are called postzygotic barriers or hybrid incompatibilities (Maheshwari and Barbash 2011). Postzygotic isolation is considered to be a by-product of species divergence, caused by negative interactions between the two parental genomes that lead to deleterious phenotypes (Johnson 2010). Indeed, the hybrid genetic background is far from being an additive combination of the two parental species genomes. Sequence divergence, differences in heterochromatin content, unpredictable epistatic interactions and changes in gene expression may be at the origin of incompatibilities in the hybrid offspring (Maheshwari and Barbash 2011). Moreover, failure of epigenetic mechanisms and other uniparentally inherited factors also seem to play a role in hybrid instability (Michalak 2009), causing asymmetric effects between reciprocal interspecific crosses (Turelli and Moyle 2006; Fontdevila 2016).

Among the different hybrid phenotypes with reduced fitness, the most well-known are sterility and lethality, which have been long and extensively studied in *Drosophila* since the 1920s (Barbash 2010; Fontdevila 2016). In almost a century of research, less than ten genes involved in those hybrid incompatibilities have been described (reviewed in Maheshwari and Barbash 2011; Fontdevila 2016). It is important to note that these speciation genes are difficult to characterize, since their phenotype and function in hybrids are often different from those of parental species (Maheshwari and Barbash 2011). For instance, the gene *Odysseus*, involved in *D. simulans–D. mauritiana* hybrid male sterility, has only a mild function enhancing sperm production in parental

species (Sun et al. 2004). Another pitfall hindering the study of hybrid incompatibilities is the difficulty, or even impossibility, to perform serial hybrid crosses (Barbash 2010). Overall, many questions in this field remain unanswered. For example, it is still not clear whether sterility and lethality are mainly caused by a few genes of large effect (Masly et al. 2006; Phadnis and Orr 2009), or by the combined action of many minor factors (Fontdevila 2016).

In hybrids between *D. buzzatii* and *D. koepferae*, the latter explanation seems to be the most suitable. Studies in their backcrossed hybrids show that a minimum total size of autosomal introgressed fragments, corresponding approximately to 30% of the autosomal length, is required to produce male sterility (Naveira and Fontdevila 1986; Naveira and Fondevila 1991; Morán and Fontdevila 2014). These loci are dispersed in the genome and exchangeable, and only the accumulative effect of a sufficient number of them results in sterility (Fontdevila 2016). Interestingly, in the same hybrids, introgression of short fragments of the X chromosome can lead to hybrid sterility, or even inviability, without reaching any threshold (Naveira and Fontdevila 1986; Naveira and Fondevila 1991). The larger effect of this chromosome has also been described in *D. sechellia–D. mauritiana* hybrids and has been attributed to its higher density of incompatibility factors compared to autosomes (Masly and Presgraves 2007). Although the importance of some major effect genes cannot be dismissed, current literature in several species agrees that incompatibility phenotypes are based on the intricate epistatic interaction of multiple genes, ranging from a ten to hundreds (Phadnis 2011; Dzur-Gejdosova et al. 2012; Turner et al. 2014; Turner and Harr 2014; Phadnis et al. 2015).

Hybrid incompatibilities can lead to (or be caused by) genetic instability, a common feature of hybrid organisms (Fontdevila 2005). Actually, genetic instability is a potential source of diversity, but it can easily produce negative effects on hybrid fitness, hence contributing to reproductive isolation. At a chromosomal level, a high occurrence of reorganizations have been observed in hybrids of both animals (Naveira and Fontdevila 1985; O'Neill et al. 1998) and plants (Rieseberg et al. 1996; Wang et al. 2005). For instance, *Drosophila* hybrids present mostly inversions and duplications, yet deletions and translocations have also been described (Naveira and Fontdevila 1985). Another prominent phenomenon is polyploidization, which is frequent in hybrids of flowering plants and also produces aberrant karyotypes (Wendel 2000; Hegarty and Hiscock 2005). It is important to mention that genetic architecture can directly contribute to hybrid sterility and lethality, since karyotypic differences are likely to cause meiotic defects (Brown and O'Neill 2010).

Transposable elements (TEs) have been proposed as major drivers of the genomic instability produced during interspecific hybridization (Fontdevila 2005). Transposition events have been

reported in hybrids of different taxa, which concurs with Barbara McClintock's hypothesis that TEs are activated due to *genomic shocks* (McClintock 1984). For example, introgression of wild rice (*Zizania latifolia*) DNA in domesticated rice lines (*Oryza sativa*) leads to the activation of at least five transposons: *Tos17* and *RCS1* (Liu and Wendel 2000), *mPing* and *Pong* (Shan et al. 2005), and *Dart* (N. Wang et al. 2010); which is coupled with important changes in DNA methylation and transcription (Liu et al. 2004). Likewise, *Helianthus anomalus*, *H. deserticola* and *H. paradoxus*; the three mentioned sunflower species of hybrid origin (see **section 1.1.1**), have also experienced retrotransposon proliferation (Ungerer et al. 2006). In mammals, TE amplification was reported in centromeres of different macropodid hybrids (O'Neill et al. 1998; Metcalfe et al. 2007) and associated with genome-wide undermethylation (O'Neill et al. 1998). Finally, the first record of hybrid TE mobilization in *Drosophila* was the detection of a new *pDv111* insertion in hybrids between *D. virilis* and *D. littoralis* by *in situ* hybridization (Evgen'ev et al. 1982). A transposition increase of the retrotransposon *Osvaldo* was later described in *D. buzzatii–D. koepferae* hybrids (Labrador et al. 1999); where a more recent survey, performed at a genome-wide level, detected mobilization of 28 differents TEs (Vela et al. 2014). At the expression level, high TE transcription rates have been detected in several species hybrids, including sunflowers (Renaut et al. 2014), fishes (Dion-Côté et al. 2014) and *Drosophila* (Kelleher et al. 2012; Carnelossi et al. 2014; García Guerreiro 2015); suggesting that hybrid TE mobilization could be due to a silencing breakdown.

Interestingly, all signs of hybrid genetic instability that have hitherto been described have the ability of increasing genome size. Genome size (or C-value) presents a wide range of values among eukaryotes, reaching differences higher than 600,000-fold (Gregory 2005a), and is considered an important feature in the study of genome evolution and species diversification (Kraaijeveld 2010). Along with polyploidization, TE mobilization is one of the main forces that can contribute to genome expansion (Kidwell 2002). A striking example is the maize genome, which doubled its size in the last few million years due to several waves of transposition (SanMiguel et al. 1996). Therefore, hybridization-induced transposition bursts can potentially lead to hybrid genome size expansion. This has already occurred in sunflowers: the three above-mentioned species of hybrid origin are known to have genomes 50% larger than their parents (Baack et al. 2005). However, analyses of synthetic hybrids between the same *Helianthus* species, along with studies in other plants, reveal that this change in genome size does not always take place after hybridization (Baack et al. 2005; Mahelka et al. 2005; Zhou et al. 2010; Camillo et al. 2014).

1.2 *D. buzzatii* and *D. koepferae*

The genus *Drosophila* has been extensively studied over the course of the past century, becoming an important model system for the understanding of biological processes (Markow and O'Grady 2007). Overall, over 2000 species have been described and are usually divided in two subgenera: *Drosophila* and *Sophophora*, although the taxonomy of this genus is still discussed (Markow and O'Grady 2009). The most well-known and best characterized species, *D. melanogaster*, constitutes one of the most powerful genetic models among eukaryotes. The accumulated knowledge on *Drosophila* biology has yielded a high amount of experimental tools and resources that has greatly increased with the beginning of the genomic era (Matthews et al. 2005). Since the first release of the *D. melanogaster* genome (Adams et al. 2000), the sequencing of at least 25 *Drosophila* species genomes has been achieved (Richards et al. 2005; Drosophila 12 Genomes Consortium 2007; Zhou et al. 2012; Zhou and Bachtrog 2012; Chiu et al. 2013; Fonseca et al. 2013; Hu et al. 2013; Ometto et al. 2013; Chen et al. 2014; Guillén et al. 2015), and other projects are currently in progress. However, most of the sequenced species are members of the subgenus *Sophophora*, to which *D. melanogaster* belongs. In total, only six of them are part of the *Drosophila* subgenus: *D. virilis*, *D. mojavensis*, *D. grimshawi*, *D. albomicans*, *D. americana* and *D. buzzatii*.

The species pair studied in this work, *D. buzzatii* and *D. koepferae*, belong to the *repleta* group, *mulleri* subgroup, *buzzatii* complex and *buzzatii* cluster within the subgenus *Drosophila* (Fontdevila et al. 1988; Ruiz and Wasserman 1993). Both species are native to arid lands of South America and coexist in vast areas of northwestern Argentina and southern Bolivia. However, while *D. koepferae* is endemic to South America, *D. buzzatii* has been introduced to other continents, colonizing the Mediterranean region, the Canary Islands, Madeira, Equatorial Africa and Australia (Fontdevila et al. 1981; Sokal et al. 1987) and reaching a subcosmopolitan distribution (Manfrin and Sene 2006). As many of the species of the *repleta* group, they use necrotic tissues of different cacti as breeding and feeding sites, but they have different host preference. In nature, *D. buzzatii* is mainly associated to *Opuntia* cacti (prickly pears or tunas) and *D. koepferae* to columnar cacti (*Cereus* or *Trichocereus*), with a certain degree of niche overlap (Hasson et al. 1992; Fanara et al. 1999). Therefore, sympatric populations of *D. buzzatii* and *D. koepferae* are partially isolated by niche specificity (corresponding to microspatial habitat isolation, see **section 1.1.2.1**). Under experimental conditions, both species mimicked their behaviour in nature: *D. buzzatii* had a higher viability and preferred to oviposit on *Opuntia*, whereas *D. koepferae* laid more eggs and was more viable when

using *Trichocereus*; suggesting that host specificity could be the result of environmental adaptation (Fanara et al. 1999; Fanara and Hasson 2001). In addition, males exhibited greater mating success when flies developed in their preferred host (Hurtado et al. 2012). Actually, tunas and columnar cacti can be distinguished by several ecological features, like chemical compounds or microbiotic composition associated to the decaying process. For example, *Opuntia* is a relatively toxic-free habitat, whereas *Trichocereus* and *Cereus* contain alkaloid secondary compounds (Soto et al. 2014). Furthermore, while tunas are more abundant than columnars, they are also considered a more ephemeral breeding substrate; which has been linked to differences in life history traits between this species pair (Fanara et al. 1999; Soto et al. 2008). In particular, *D. koepferae* larger body size is thought to be a consequence of natural selection for greater dispersal ability (useful due to the greater distance between breeding sites) and *D. buzzatii* faster development could be an adaptation to more transient hosts. Recently, *D. buzzatii* has been registered for the first time emerging from a non-cactus host, *Cucumis melo* (Fanara et al. 2016).

Divergence between these closely related species is estimated to have occurred approximately 4-5 million years ago (Mya) (Gomez and Hasson 2003; Laayouni et al. 2003; Oliveira et al. 2012). They exhibit similar karyotypes, with five pairs of autosomes (including a dot) and one pair of sexual chromosomes (Ruiz and Wasserman 1993); and are morphologically undistinguishable except for striking differences in male genitalia. Besides the aforementioned spatial isolation, other prezygotic barriers may contribute to reproductive isolation between these species. For instance, male courtship song in *D. buzzatii* exhibits longer bursts than in *D. koepferae*, which could reduce the probability of interspecific matings (Oliveira et al. 2013); and differences in aedeagus size and shape could lead to failed insemination attempts (Soto et al. 2007; Soto 2012; Soto et al. 2013), as occurs in *D. mojavensis* species cluster (Richmond 2014). Finally, sperm competition could also play a role in reproductive isolation, particularly in *D. buzzatii* females, that take longer to deplete sperm reserves and are more prone to remate than females of *D. koepferae* (Fanara et al. 1999; Hurtado et al. 2013; Hurtado and Hasson 2013); as well as lethal insemination reaction, that is known to occur in interspecific crosses between species of the same group (Marin et al. 1993).

Hybrids have never been found in nature, but the observation of extensive sharing of polymorphic variants in several nuclear loci (Gomez and Hasson 2003; Piccinali et al. 2004; Franco et al. 2010) indicates that interspecific gene flow may have played a significant role in the evolution of this species pair. In the laboratory, successful crosses can be performed between *D. buzzatii* males and *D. koepferae* females, yielding sterile males and fertile females that can be backcrossed with *D. buzzatii* males (Naveira and Fontdevila 1986; Marín and Fontdevila 1998). However, the reciprocal

cross does not produce offspring –only once, a single larva (and no adults) was observed in a study involving 48 massal crosses (Marin et al. 1993). Asymmetry between reciprocal crosses is very common and often results from incompatibilities involving uniparentally inherited genetic factors (Turelli and Moyle 2006). Accordingly, *D. buzzatii* is the species that hybridizes less easily within the *buzzatii* cluster, especially when *D. buzzatii* females are involved in the heterospecific cross (Marin et al. 1993). This hints at maternally inherited factors (such as proteins, mRNAs and/or small RNAs) as the origin of inviability (Turelli and Moyle 2006). Male sterility in the successful cross direction results from the cooperative effect of incompatibilities at several genetic loci, following a polygenic threshold model that has already been discussed (see **section 1.1.2.2**).

Even in the successful cross, egg viability is extremely low, whereas larval viability and developmental time in hybrids were shown to be, at worst, not significantly different from parental species –and sometimes greater and faster, respectively (Soto et al. 2008). Therefore, *D. buzzatii–D. koepferae* viable hybrids do not seem to present a highly reduced fitness and could even perform better than their parents in some cases (Soto et al. 2008), which could explain the detection of reticulation in natural populations. Interestingly, studies on two morphological traits (wing length and aedeagus morphology) showed that hybrids usually present intermediate phenotypes, but some individuals are characterized by extreme phenotypes that differ from both parental species (Soto et al. 2007; Soto et al. 2008). It is also worth noting that genetic instability has been detected in hybrids between *D. buzzatii* and *D. koepferae*. For example, a high occurrence of inversions, duplications and other chromosome rearrangements was reported in their backcrossed hybrids (Naveira and Fontdevila 1985), which concurred with transpositional bursts of the retrotransposon *Osvaldo* (Labrador et al. 1999). More recently, a study at a genome-wide level has demonstrated transposition events can account for a high percentage of the genetic instability detected in these hybrids (Vela et al. 2014). Using AFLP markers and transposon display techniques, mobilization of a total of 28 different TE families was detected, including LTR, non-LTR and DNA transposons (Vela et al. 2011; García Guerreiro 2014). Expression studies of one of the mobilized elements, the retrotransposon *Osvaldo*, revealed that an increase of transcript abundance, especially in testes, could precede mobilization events, pointing to a failure of TE silencing mechanisms (García Guerreiro 2015). However, little is known about the mechanisms triggering this hybrid genomic instability and TE release, and we also ignore whether these results are TE-specific or can be generalized.

1.3 Transposable elements

Transposable elements (TEs) are DNA segments with the ability to mobilize to new genomic locations, often producing duplicate copies of themselves in the process. TEs that encode all the proteins needed for their transposition are called autonomous elements. However, not all of them actually transpose: some autonomous elements are inactivated epigenetically and remain silent in the genome. On the other hand, non autonomous elements are those that cannot transpose by themselves, but can use the proteins encoded by autonomous elements to mobilize. Non autonomous elements are often mutated relics of autonomous copies (Slotkin and Martienssen 2007).

1.3.1 History and classification

1.3.1.1 Historical perspective

The discovery of TEs took place in the late 1940s and is owed to Barbara McClintock, who was by then studying the mechanisms of chromosome breakage and fusion in maize (Ravindran 2012). McClintock observed the appearance of several kernel colour mutant phenotypes after the spontaneous insertion of a particular genetic unit, the *Ds* locus, close to the genes responsible for aleurone and endosperm pigmentation. Those mutations were reversible, showing that the *Ds* locus was able to mobilize from one chromosome to another, and were coupled with the occurrence of several chromosome rearrangements. She simultaneously discovered a second mobile element, the *As* locus, which was required for *Ds* transposition (McClintock 1950). Her findings, rigorously supported by experimental evidence, challenged the then established view of a static genome and were received with hostility and incredulity (Hua-Van et al. 2011; Fedoroff 2012). However, the gradual description of mobile sequences in other organisms, such as viruses, bacteria and *Drosophila* (Taylor 1963; Shapiro 1969; Engels and Preston 1981), led to the recognition of TEs as taxonomically widespread. McClintock was finally awarded the Nobel Prize in 1983 (McClintock 1983), more than three decades after her finding's publication (McClintock 1950).

The advent of large-scale sequencing technologies has revealed that TEs are present in almost all eukaryotic sequenced genomes. To date, only some parasitic Apicomplexa protozoans (Carlton et al. 2002; Gardner et al. 2002; Abrahamsen et al. 2004; Xu et al. 2004; Gardner et al. 2005; Brayton et al. 2007), the extremophile red alga *Cyanidioschyzon merolae* (Misumi et al. 2005) and the parasitic

Microsporidia fungi *Encephalitozoon cuniculi* (Katinka et al. 2001) are known to be devoid of mobile elements. The lack of TEs in those small eukaryotic genomes may be explained by a tendency towards genome size reduction (Hua-Van et al. 2011), a general feature of intracellular pathogens (Pritham 2009). Accordingly, the comparison between three trypanosomatid genomes showed that eukaryotic organisms are able to eradicate active TEs: *Trypanosoma brucei* and *T. cruzi* genomes both contain active retrotransposons that have been lost in the closely related species *Leishmania major* (Bringaud et al. 2006). Indeed, only remnants of degenerated TE copies can be detected in the latter genome, which was first considered to be depauperate in TEs (Ivens et al. 2005).

In addition to their ubiquity among eukaryotes, TEs often constitute an important fraction of their host genomes (Hua-Van et al. 2011; Lopez-Flores and Garrido-Ramos 2012). In animals, they represent 3% of the *Caenhorabditis elegans* genome, around 15% of *Drosophila melanogaster*, 37% of *Mus musculus* and more than 50% of the human genome (Dowsett and Young 1982; The C. elegans Sequencing Consortium 1998; Lander et al. 2001; Mouse Genome Sequencing Consortium 2002; Drosophila 12 Genomes Consortium 2007; de Koning et al. 2011). The genome fraction they occupy is even more variable in plants: from 10% in *Arabidopsis thaliana* genome to almost 90% in maize (The Arabidopsis Genome Initiative 2000; Schnable et al. 2009). The finding that repetitive sequences, and not genes, are major components of most eukaryotic genomes was crucial to explain the lack of correlation between genome size and gene number (or organismal complexity), also known as the “C-value paradox”. On the contrary, genome size and TE abundance seem to be positively correlated (Kidwell 2002). For instance, TE content accounts for genome size variation between different species and populations of *Drosophila* (Vieira et al. 2002; Boulesteix et al. 2006).

1.3.1.2 Classification of TEs

The gradual identification of new TEs in different species shed light on their intrinsic extraordinary diversity. As a result, D.J. Finnegan proposed a first classification system in 1989, which divided TEs in two classes according to their transposition mechanism. Class I included all elements that transpose via an RNA intermediate, whereas class II comprised elements that do not require this intermediate step (Finnegan 1989). The original system has been maintained, but it has since then been updated and adapted to facilitate the annotation of high amounts of emerging sequencing data (Jurka et al. 2005; Wicker et al. 2007). The most recent proposal classifies TEs in six hierarchical levels (class, subclass, order, superfamily, family and subfamily) according to their insertion

mechanism and structural features (Wicker et al. 2007). This classification, that does not necessarily rely on phylogenetic relationships, is summarized hereafter (**Figure 1**).

Class I elements

Class I elements, also named retrotransposons, are generally the most abundant in eukaryotic genomes. Their transposition starts with the transcription of a genomic TE copy into an RNA intermediate that is then reverse transcribed by a TE-encoded retrotranscriptase. This produces a new DNA copy of the element that is afterwards inserted in a new genomic location. This transposition mechanism, called *copy-and-paste*, is intrinsically replicative, which makes class I elements prone to increase their copy number in the genome (Lopez-Flores and Garrido-Ramos 2012). Retrotransposons have been divided in five different orders: LTR, LINE, SINE, DIRS-like and PLE-like (Wicker et al. 2007).

LTR (Long Terminal Repeat) retrotransposons are the predominant order in plants, as well as in a few animals, such as *Drosophila melanogaster*. Their main characteristic are their flanking LTRs that can recombine between them producing solo-LTR copies. They typically contain two open reading frames (ORFs): *gag*, that encodes a structural protein; and *pol*, that encodes various enzymatic domains (proteinase, reverse transcriptase, RNAse and integrase). LTR retrotransposons are closely related to retroviruses, which are included as a superfamily in the LTR order.

Elements from the LINE order (Long Interspersed Nuclear Element) are less common in plants but predominate in most animals, especially in birds and mammals. They lack LTRs and were used to be called non-LTR retrotransposons. All LINEs contain the *pol* ORF, with at least a reverse transcriptase and a nuclease domain; and some of them also carry a *gag*-like ORF upstream from *pol*, whose function remains unclear. They mobilize by target-site-primed reverse transcription, in which integration and reverse transcription steps are coupled. This transposition mechanism produces 5'-truncated copies due to the premature termination of reverse transcription. The retrotransposon *Helena*, which is the main focus of **chapter 3.2**, is a LINE-like element from the *Jockey* superfamily (Rebollo et al. 2008).

SINEs (Short Interspersed Nuclear Elements) are non-autonomous elements that are functionally related to LINEs. They are not deletion derivatives of any autonomous retrotransposon, but instead originate from the accidental retrotransposition of RNA polymerase III transcripts –tRNAs, 7SL RNAs and 5S rRNAs. Although they do not contain any ORF, they possess an internal promoter of RNApol III that allows them to be transcribed. Once expressed, they use LINEs' enzymatic

machinery to transpose. In humans, the SINE *Alu* family alone has more than one million copies representing 11% of the genome (Lander et al. 2001).

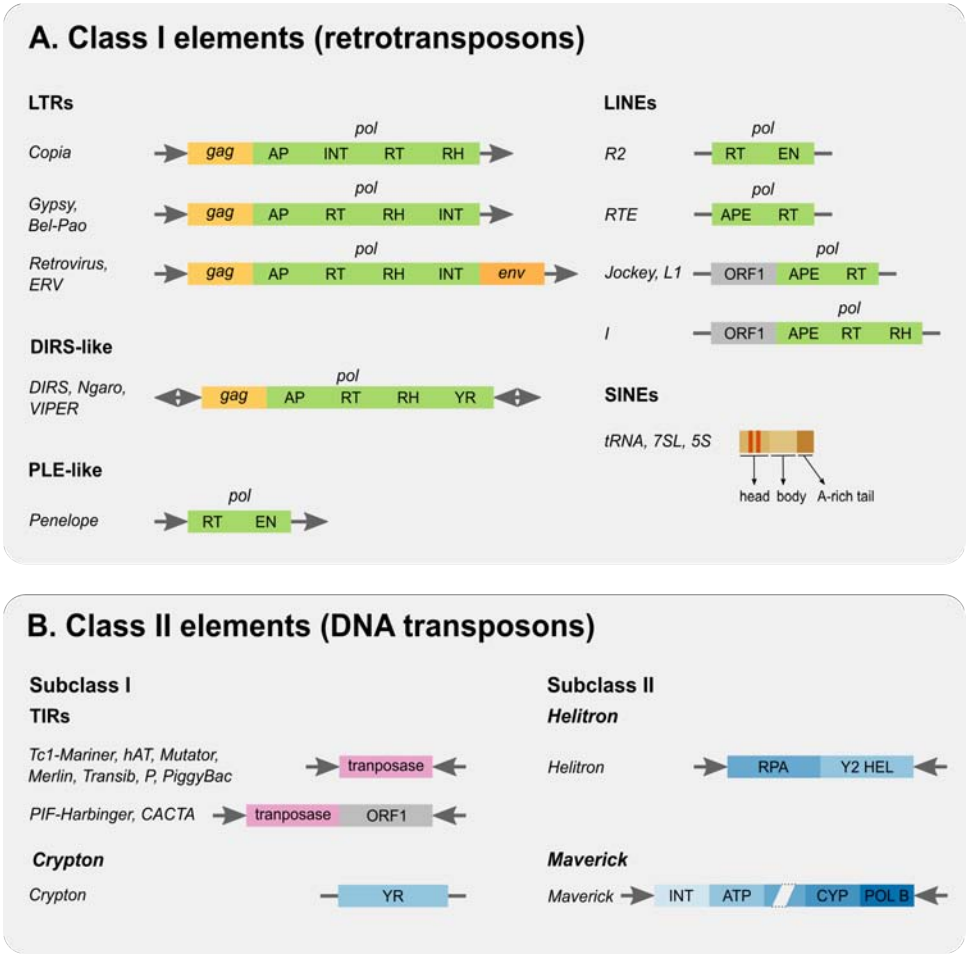


Figure 1: Schematic representation of the described TE classification system. Adapted from Wicker et al. (2007). (**A)** Retrotransposons are divided in five orders (in bold), each one composed of several superfamilies (in italics). **(B)** DNA transposons are classified into two subclasses, each divided in two orders (in bold) formed by one or more superfamilies (in italics). In alphabetical order: AP: aspartic proteinase, APE: apurinic endonuclease, ATP: packaging ATPase, C-INT: C-integrase, CYP: cysteine protease, EN: endonuclease, *env*: envelope protein, *gag*: capsid protein, HEL: helicase, INT: integrase, ORF1/2: open reading frame of unknown function, *pol*: polyprotein encoding various enzymatic domains, POL B: DNA polymerase B, RH: RNase H, RPA: replication protein A, RT: reverse transcriptase, YR: tyrosine recombinase, Y2: YR with YY motif.

In addition to the three classical retrotransposon types described, two orders –DIRS-like and PLE-like– have more recently been reported. DIRS-like (Dictyostelium intermediate repeat sequence) elements are similar to LTRs, but their *pol* gene contains a tyrosine recombinase instead of an integrase, which reveals differences in their mechanism of integration. On the other hand, PLE-like

elements (*Penelope*-like) usually encode a single ORF with a reverse transcriptase and an endonuclease domain, both phylogenetically distant from the other retrotransposons' ones. Interestingly, some of them contain an intron that is not lost during the transposition process.

Class II elements

Class II elements are named DNA transposons because they do not need an RNA intermediate to transpose: their DNA copies move directly from one chromosomal location to another. Although they are also found in almost all eukaryotic genomes, they are usually less abundant than retrotransposons. However, there are some notable exceptions, including *C. elegans*, *Hydra* and *Xenopus* genomes, where DNA transposons prevail (The C. elegans Sequencing Consortium 1998; Chapman et al. 2010; Hellsten et al. 2010). Class II elements are divided in two subclasses, according to the number of DNA strands cut during their transposition.

Elements of subclass 1 require the cleavage of both DNA strands for their transposition. Two orders have been defined within this subclass: TIR and *Crypton*. TIR (Terminal Inverted Repeat) elements, also known as *cut-and-paste* transposons, are characterized by their flanking TIRs. Their transposition, mediated by a self-encoded transposase, is not replicative. However, they can increase their copy number by transposing during chromosome replication from a newly-replicated position to an unreplicated site. On the other hand, the less well-known *Crypton* elements were first discovered in fungi and later described in several other animals (Kojima and Jurka 2011). They encode a tyrosine recombinase and seem to transpose via recombination and integration.

DNA transposons of subclass 2 cleave a single DNA strain during their transposition and are classified in two orders: *Helitron* and *Maverick*. *Helitron* elements seem to replicate via a rolling-circle mechanism thanks to a Y2 tyrosine recombinase, which allows them to proliferate within the genome. *Mavericks* also seem to undergo a replicative transposition via the excision of a single-strand DNA fragment that is replicated extrachromosomically and then integrated to a new genome site. The latter order comprises large elements with flanking TIRs that can encode up to 11 proteins, including a DNA polymerase and an integrase.

1.3.2 Interaction with host genome

'Given a sufficient lack of comprehension, anything (and that includes a quartet of Mozart) can be declared to be junk.'

Emile Zuckerkandl and Wolfgang Hennig (1995)

For a long time, transposable elements were considered molecular parasites with little or no phenotypic contribution on their hosts, and were consequently labelled with terms as *selfish* or *junk* DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). Their unique relevant function was thought to be their own proliferation, and hence their high frequency and subsistence in the genomes was attributed to their replicative advantage over other sequences (Kidwell and Lisch 2001). This viewpoint, focused solely on the ability of TEs to propagate, failed to capture a far more complex relationship between TEs and their hosts.

1.3.2.1 TE life cycle

Sequencing data has demonstrated that most TE-derived sequences in eukaryotic genomes are not functional, even though the fraction of active copies is variable between species. For instance, 17% of the euchromatic TE insertions are full-length in *Drosophila melanogaster* (Bartolomé et al. 2002); whereas only 1% of human *L1* elements are estimated to be complete (Sassaman et al. 1997; Prak and Kazazian 2000). Furthermore, a same element can be present at different stages in distinct host species, as well as within a same host genome. For example, the retrotransposon *Helena* has been found in 8 out of the 12 *Drosophila* sequenced genomes (Drosophila 12 Genomes Consortium 2007), but only *D. simulans* and *D. mojavensis* harbour full-length copies –along with incomplete ones (Rebollo et al. 2008; Granzotto et al. 2009). Interestingly, even between the latter two species the transcriptional state of *Helena* differs: it is silenced in most strains of *D. simulans* but generally active in *D. mojavensis*. Several TE life cycle models have been proposed throughout the last decade in order to explain the apparently stable presence of so many non functional insertions and the general dynamics of TE-host genome interaction. However, it is worth noting that most of them were proffered before the discovery piRNAs (a kind of small interference RNAs that silence TEs in the *Drosophila* germline, see **section 1.3.2.3**) and other mechanisms involved in TE control.

The life cycle of a TE in its host genome was initially thought to be divided in three main phases: invasion, maturity and senescence. A new TE would invade the genome and start a rapid proliferation, leading to a copy number increase along with sporadic mutations that would render some of them inactive. At some point, the frequency of copy loss by mutational processes would counterbalance the appearance of new insertions, reaching the maturity stage. Finally, when none of its copies would be active anymore, the TE would enter the senescence phase, which could last for millions of years. At this stage, non autonomous copies would be gradually lost, deleted, or otherwise would accumulate mutations until their remnants could not be identified as TEs (Kidwell and Lisch 2001). However, this model disregarded some crucial TE features that influence the

relationship between TEs and their hosts. For instance, it did neither take into account the variability of insertion effects (that range from deleterious to neutral or adaptive), nor the differences between copies. In fact, in a more recent and improved model (Le Rouzic et al. 2007), the transposition-deletion equilibrium previously described was shown to be reached only in unrealistic conditions (*i.e.* very low mutation rates and null probability of adaptive insertion).

According to this more accurate model (**Figure 2**), when a new TE arrives in a genome, it is likely rapidly lost due to genetic drift or natural selection, unless it is able to immediately and efficiently transpose and start the invasion (Le Rouzic and Capy 2005). The persistence of TE activity during the invasion phase depends mainly on newly inserted copies, because older ones have accumulated mutations and are more likely to have lost their ability of transposition. However, new insertions are often deleterious and are frequently eliminated by purifying selection. Thus, after an active transposition stage, the transposition rate tends to decrease, freezing the TE copy number. At this step, re-invasions can occur and prevent the complete loss of TE activity, leading to several invasion-regression cycles. After that, the remaining copies can be either domesticated (if they have an adaptive advantage) or lost by deletion or genetic drift (Le Rouzic et al. 2007). In terms of the fitness of the host, there is a decline during the active transposition stage (as a consequence of TE proliferation), but in the end it is recovered and exceeds the initial level.

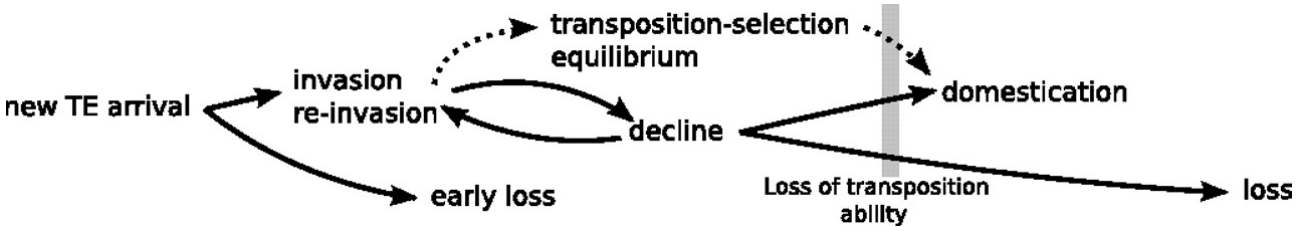


Figure 2: Summary of TE-host genome dynamics according to the model described by Le Rouzic, Boutin and Capy (2007). Excerpted from Le Rouzic et al. (2007).

It is noteworthy that this model does not take into account the existence of piRNA-mediated TE silencing (see **section 1.3.2.3**), which is important to neutralize active TEs in humans and flies (Lukic and Chen 2011; Kelleher and Barbash 2013). In particular, piRNAs have a major role at the onset of a TE invasion in *Drosophila* (Kelleher and Barbash 2013), probably reducing the duration of the invasion phase.

On the other hand, none of these models specifies the mechanisms by which a new TE arrives in the host genome. Actually, this can occur either by the reactivation of a quiescent copy, or by horizontal

transfer between species (Le Rouzic et al. 2007). The unexpected discovery that the DNA transposon *P* recently and rapidly invaded the *Drosophila melanogaster* genome after being horizontally transferred from *D. willistoni* (Anxolabéhère et al. 1988; Daniels et al. 1990) revealed that horizontal transmission might be an essential mechanism for TE subsistence. Since then, several studies describing TE horizontal transfer in eukaryotes have been reported, especially in *Drosophila* (Jordan et al. 1999; Sánchez-Gracia et al. 2005; Bartolomé et al. 2009; Gilbert et al. 2010; Lerat et al. 2011; Modolo et al. 2014; Kofler et al. 2015). However, the mechanisms by which TEs are transported between organisms remain elusive. A multitude of pathogens and parasites infecting eukaryotes have been proposed as potential vehicles for horizontal transfer. For instance, the suspected vectors in the case of the *P* element are mites (Houck et al. 1991), yet bacteria and viruses remain generally the most discussed (Gilbert et al. 2010; Schaack et al. 2010).

Otherwise, inactive TEs can sometimes be *awaken* and start a new invasion phase. For instance, different cases of transposition bursts have been described under different biotic and abiotic stress conditions, such as starvation, drought, extreme temperatures, chemical agents, radiation or viral infection (reviewed in García Guerreiro 2012; Casacuberta and González 2013). This concurs with Barbara McClintock's early idea that transposons can be activated after a *genome shock*, which may allow the host organism to respond and adapt rapidly to changes in environmental conditions (McClintock 1984). Remarkably, different TE families are activated under different environmental stresses, indicating that stress-induced TE activation depends on the regulatory sequences associated to a particular TE (or even insertion). On the other hand, those shocks can also be due to genomic stresses, such as hybridization between differentiated populations or species. For example, crosses between strains of the same *Drosophila* species that differ in the presence of a particular transposon can give rise to sterile offspring when females devoid of the TE are mated with males containing it (Picard 1976; Kidwell et al. 1977; Kidwell and Novy 1979; Evgen'ev et al. 1997). This is named the hybrid dysgenesis syndrome and is caused by the absence of piRNAs associated to the paternally-inherited TE in the maternal cytoplasm (Brennecke et al. 2008). Moreover, as explained in **section 1.1.2.2**; the merge of two different genomes during interspecific hybridization events can induce TE derepression both at transcriptional and transpositional levels, which has been reported in both plants (Liu and Wendel 2000; Shan et al. 2005; Ungerer et al. 2006; H.-Y. Wang et al. 2010; N. Wang et al. 2010; Renaut et al. 2014; Wu et al. 2014) and animals (Evgen'ev et al. 1982; O'Neill et al. 1998; Labrador et al. 1999; Metcalfe et al. 2007; Kelleher et al. 2012; Carnelossi et al. 2014; Dion-Côté et al. 2014; Vela et al. 2014; García Guerreiro 2015).

1.3.2.2 Harmful *vs.* evolutionary potential

Transposons, being able to induce a spectrum of mutations broader than any other known mutation mechanism, can be considered powerful endogenous mutators (Kidwell and Lisch 2001). For instance, the insertion and excision of TE copies during transposition can alter gene expression, especially when TEs move within coding regions or regulatory sequences. Furthermore, the presence of TEs can modulate the state of chromatin and the degree of methylation, which also affects transcription rates. On the other hand, the repetitive nature of TEs can be at the origin of ectopic recombination between copies (Hughes and Coffin 2005), which can cause deletions (Lagemaat et al. 2005), inversions (Sniegowski and Charlesworth 1994; Cáceres et al. 1999), duplications (Mishra 2008) and other chromosomal rearrangements (McClintock 1950). In addition, transposition bursts can cause drastic increases in genome size, as happened in the already mentioned example of the maize genome (SanMiguel et al. 1996). Finally, the occasional occurrence of aberrant transposition events can promote mobilization of host's sequences as well as the creation of pseudogenes (Kidwell and Lisch 2001).

All these described mutations can have deleterious phenotypic effects and endanger the viability of the host. In humans, TE-induced mutations have been well-studied and linked to cancer and several other diseases (Biémont and Vieira 2006). For example, recombinations and deletions involving mostly *Alu* and *LINE1* elements are known to contribute to cases of leukaemia, sarcoma, hepatoma, gastric and breast cancer; as well as to thalassemia, haemophilia and other genetic disorders (Chen et al. 2005; Callinan and Batzer 2006; Chénais 2015). Moreover, TEs can also be involved in human disease through the creation of new polyadenylation sites (encoded by the element itself) and the modification of alternative splicing (by exonization, exon skipping or intron retention); producing new transcript isoforms that can be harmful to the host (Chénais 2015). It is noteworthy that mutations leading to genetic diseases occur either in the germline (affecting the next generation) or in the first stages of the development, whereas those leading to some cancers can also be somatic (Lee et al. 2012). This shows that both germinal and somatic transposition can have negative effects on the fitness of the host, and hence that all TE insertions are subject to purifying selection.

On the other hand, the mutagenic potential of TEs is also a source of variability that is considered an evolutionary force (Biémont and Vieira 2006). Indeed, the perception of the selfish nature of TEs has considerably evolved with the rising number of studies demonstrating the beneficial impact they can have on host biology (Miller et al. 1999). For instance, a single insertion of the *P* element in the

methuselah (*mth*) gene of *Drosophila melanogaster* resulted in an increase of life span and enhanced resistance to different stresses –including starvation, high temperature and herbicides (Lin et al. 1998). The *mth* mutant line was created artificially in the laboratory, but other studies have revealed that natural insertions in *Drosophila* can also confer adaptive advantages (*e.g.* pesticide resistance) (Aminetzach et al. 2005; González et al. 2008; Guio et al. 2014; Mateo et al. 2014); as also occurs in other species of animals and plants (reviewed in Casacuberta and González 2013). This is not surprising given that transposons encode an extensive enzymatic machinery and a rich repertoire of regulatory elements with the potential to supply novel functional abilities to their hosts (Feschotte 2008; Sinzelle et al. 2009). Actually, co-optation (or exaptation) of TE sequences by the host genome to serve cellular or regulatory functions has occurred repeatedly in eukaryotes via the evolutionary process of molecular domestication (Miller et al. 1999; Sinzelle et al. 2009; Alzohairy et al. 2013).

One of the most striking examples of molecular domestication is V(D)J recombination, one of the crucial functions of the immune system of jawed vertebrates (Biémont and Vieira 2006). During lymphocyte maturation, recombination between V, D and J gene segments is mediated by the recombinases RAG1 and RAG2. All these segments are flanked by recombination signal sequences (RSSs) that are specifically recognized by RAG1 and RAG2 and resemble terminal inverted repeats. In each B or T cell, RAG proteins bind a particular combination of two RSSs and excise the internal region, leading to a unique cell-specific recombination between these segments. Altogether, a wide variety of rearrangements is created, which is at the origin of the enormous diversity of antibodies and T cell receptors needed to recognize an extensive assortment of antigens (Miller et al. 1999). This system emerged approximately 500 million years ago, at the onset of jawed vertebrate evolution, from transposon-encoded proteins. Indeed, in vitro assays showed that RAG1 and RAG2 together formed a transposase capable of excising and inserting DNA fragments containing RSSs (Jones and Gellert 2004). In particular, RAG1, RAG2 and the RSSs are derived from a *Transib* transposon (class II, subclass I, TIR) that shares sequence and structure similarity with TEs of the purple and green sea urchins genomes (Kapitonov and Jurka 2005; Kapitonov and Koonin 2015).

Another example of TE molecular domestication is the case of the syncytin genes in humans (Mi et al. 2000; Villesen et al. 2004), as well as in other primates (Blaise et al. 2003; Esnault et al. 2013) and mice (Dupressoir et al. 2005). These derive from *env* genes of endogenous retroviruses (*e.g. HER-W* in humans) and are involved in placental development (Dupressoir et al. 2009). Finally, it is worth mentioning the telomere maintenance mechanism in *Drosophila*, which relies on

the targeted transposition of three non-LTR transposons (*HeT-A*, *TART* and *TAHRE*) instead of using a telomerase (Silva-Sousa et al. 2012).

In conclusion, we must note that TEs are not exempt from the influence of evolutionary forces, notably genetic drift and selection. Mutations associated to TEs can modify the fitness of their hosts, either by providing new selective advantages that might lead to environmental adaptation; or by producing detrimental effects, such as lethality or disease. Their long-term maintenance in most eukaryotic genomes indicates that an accurate concept to describe TE-host dynamics would be coevolution, rather than parasitism. Furthermore, repeated TE recruitments to fulfill essential functions for the host cells reveals that they have been crucial actors of eukaryotic genome evolution. Indeed, it is thought that their capacity to facilitate rapid karyotypic evolution and promote the creation of new regulatory gene networks could ultimately be at the origin of speciation events (Fontdevila 2005; Rebollo et al. 2010). Accordingly, important changes in TE content, particularly transposition bursts, have been shown to be concomitant with several radiation episodes (Pascale et al. 1990; Verneau et al. 1998; Dobigny et al. 2005; de Boer et al. 2007; Lorenzi et al. 2008; Ray et al. 2008).

1.3.2.3 Regulation of TEs

Despite its role in creating genetic variability, TE mobilization is also a potential source of mutations that can cause damage to the host genome. In order to confront this threat, host organisms have evolved many diverse regulation strategies that limit TE activity. The presence of such controlling mechanisms is particularly important in the germline, because new insertions there are transmitted to future generations (Castel and Martienssen 2013). For instance, TE insertions in the genome are usually associated to DNA methylation (in mammals and plants) as well as to post-translational histone modifications related to repressive chromatin. Furthermore, small interfering RNAs (siRNAs) defend eukaryotic cells against TEs, using different pathways that differ in their relative importance between organisms. In particular, two kinds of siRNAs are known to be involved in TE silencing: endogenous siRNAs (endo-siRNAs) and Piwi-interacting small RNAs (piRNAs); that can act at the transcriptional or post-transcriptional levels (Levin and Moran 2011; Rigal and Mathieu 2011).

TE silencing mechanisms in plants

In plants, three interacting epigenetic strategies are set up to minimize the deleterious effects of transposition: endo-siRNAs, DNA methylation and histone modifications (Lisch 2009). Plant

siRNAs associated to transposons are transcribed from genomic TE insertions. First, single-strand precursor transcripts are produced by the DNA-dependent RNA polymerase IV (RNApol IV), then they are converted to double-strand by an RNA-dependent RNA polymerase (RDR). Finally, a Dicer-like RNAse (DCL) processes these precursors to generate siRNAs of 21-26 nucleotides (nt) in length. These are loaded into RNA-induced silencing complexes (RISCs) containing an Argonaute protein (Ago) to perform their regulatory function (Ito 2013). The existence of multiple paralogs of RDR, DCL and Ago genes in higher plants has diversified the production of small RNA populations that are involved in different silencing pathways (Xie et al. 2004). For instance, siRNAs of 21-22 nt in length participate in TE post-transcriptional silencing (PTS) via the RDR6/DCL4/Ago1 pathway; and others of 24 nt are involved in TE transcriptional silencing (TS) through the RDR2/DCL3/Ago4 pathway. Remarkably, these two pathways are known to interact, especially when the organism is under stress conditions, to ensure an efficient TE silencing (Ito 2013; Cui and Cao 2014).

In the RDR6/DCL4/Ago1 PTS pathway, siRNAs associated to Ago1 are guided to target transposon transcripts in the cytoplasm, which are then cleaved by the RISC complex. On the contrary, the RDR2/DCL3/Ago4 pathway produces Ago4-associated siRNAs that have their role in the nucleus, where the RISC complex scans for base-pair matching transposon transcripts. Upon target engagement, Ago4 interacts with the DNA-dependent RNA polymerase V (RNApol V) to ultimately recruit the DNA methyltransferase DRM2. This leads to RNA-directed de novo DNA methylation (RdDM), a major regulator of transposon activity in plants that is known to affect nearby genes (Wierzbicki et al. 2008; Zhong et al. 2014). Besides its own action as a transcription repressor, DNA methylation can trigger histone modifications via deposition of H3K9me3 marks. This step is performed by the histone methyltransferase Kryptonite (KYP) and leads to heterochromatin formation. Moreover, H3K9me3 marks can in turn recruit a DNA methyltransferase, CMT3, hence producing a self-reinforcing feedback loop that ensures a tight repression of TE transcription. Actually, several methyltransferases (such as Met1, CMT3, DRM2 and CMT2) and the chromatin remodeling factor DDM1 are known to control transposition of different subsets of TEs.

A conspicuous example of interaction between plant TE silencing pathways is the case of *Arabidopsis* DDM1 (and Met1) mutants, where a drastic loss of DNA methylation and H3K9me3 marks in TE-rich regions was reported, causing a transcriptional reactivation of several TEs, including the retrotransposon *Athila* (Slotkin et al. 2009; Creasey et al. 2014). While this was associated to a decrease in 24-nt siRNAs, a surprising increase of 21-nt siRNAs was also observed,

travelling from the vegetative nucleus to the sperm to reinforce TE silencing. This shows that TE control is particularly strong in the germline, hence allowing the maintenance of epigenetic TE control across generations. The biogenesis of these 21-nt *epigenetically activated* siRNAs (easiRNAs) was shown to depend on miRNA production, as well as on RDR6 and DCL4, the two factors usually involved in transposon PTS. Therefore, the RDR6/DCL4/Ago1 pathway can be activated in case of failure of the RDR2/DCL3/Ago4 pathway, which provides an alternative strategy to repress TEs that have evaded TS (Creasey et al. 2014). Similar interactions between plant transposon silencing pathways have been reported in case of *de novo* TE invasion (Marí-Ordóñez et al. 2013) and exposure to heat stress (Ito et al. 2011). Therefore, it is important to note that complementarity between different mechanisms of TE control may allow plants to face changes both in the genetic background and the environment.

TE silencing mechanisms in animals

Contrary to plants, animals possess two small RNA pathways associated to TE regulation: the endo-siRNA pathway and the piRNA pathway. Both can perform PTS and TS, triggering changes in chromatin conformation via DNA methylation and/or histone modification. Between them, the piRNA pathway has been, by far, the most extensively studied. Actually, endo-siRNAs are particularly important for TE silencing in plants (see **above**), nematodes and mammals, but its effects are subtler in the *Drosophila* germline (Chung et al. 2008; Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008). Given that this work is based on two *Drosophila* species, I will briefly introduce the endo-siRNA pathway but I will focus and expand primarily on the piRNA pathway.

Endo-siRNAs were first considered to be absent in *Drosophila* and vertebrates because these organisms do not encode any RNA-dependent RNA polymerase (RDR), an enzyme required for endo-siRNA production in yeast, plants (see **above**), and some animals (e.g. *Caenorhabditis elegans*). Therefore, their discovery in mice and *Drosophila* added one more layer of complexity to RNA silencing, while revealing an epigenetic strategy for TE silencing in somatic tissues (Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Okamura et al. 2008; Tam et al. 2008; Watanabe et al. 2008). Actually, no RDR polymerase is needed for their production: double-stranded RNAs (dsRNAs) are produced via natural antisense and hairpin transcripts or convergent transcription. In *Drosophila*, endo-siRNAs are 21 nt in length, bind specifically to Ago2 and their biogenesis is Dicer2-dependant; exactly as exogenous siRNAs, responsible for virus silencing (Kawamura et al. 2008; Xie et al. 2013). Two other factors, the dsRNA-binding proteins R2D2 and

Loquacious (isoform Loqs-PD), are required for endo-siRNA biogenesis (Zhou et al. 2009; Marques et al. 2010). Remarkably, endo-siRNAs are found in the soma, where they play a major role in TE silencing; but also in the germline, where they share their function with piRNAs (Lau et al. 2009). In *Drosophila* ovaries, when Dicer-2 or Ago2 gene functions are lost, mutant flies are viable and fertile and only a small subset of TEs are derepressed (Chung et al. 2008; Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008). This suggests that the endo-siRNA pathway is partially redundant with the piRNA-pathway.

Biogenesis and function of piRNAs

piRNAs are a class of germline-specific small RNAs of 23-30 nt that were first discovered in *Drosophila*, predominantly matching TEs and other repeats (Aravin et al. 2003). They were then shown to be associated to three specific Argonaute proteins: Argonaute3 (Ago3), Aubergine (Aub) and Piwi, all belonging to the Piwi subgroup (Saito et al. 2006; Vagin et al. 2006; Brennecke et al. 2007; Gunawardane et al. 2007). Populations of small RNAs associated to the Piwi subfamily proteins were soon described in several other species, such as mouse (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Watanabe et al. 2006), rat (Lau et al. 2006) and zebrafish (Houwing et al. 2007), revealing that the piRNA pathway is highly conserved in animals.

Unlike other small RNAs, piRNAs are specifically found in reproductive tissues and derive from single-stranded RNA precursors (rather than from double-stranded). Indeed, they are encoded by specific genomic loci called piRNA clusters, considered *transposon cemeteries* because they are mostly composed by remnants and nested fragments of inactive transposons (Brennecke et al. 2007). In mice, piRNA clusters are found in euchromatin, whereas in *Drosophila* they reside in heterochromatic regions, which are usually transcriptionally silent (Aravin et al. 2006; Girard et al. 2006; Brennecke et al. 2007). Actually, their transcription in flies seems to require chromatin repression marks (H3K9me3), without which there is a decrease of precursor piRNA amounts and hence a collapse of piRNA production (Rangan et al. 2011; Molla-Herman et al. 2015). In *Drosophila*, piRNA clusters can produce piRNAs either from a single or from both genomic strands, and are consequently called uni-strand or dual-strand clusters. The former resemble protein coding genes, with H3K4me3 marks at their promoters, and produce customary RNA polymerase II transcripts with 5' cap, polyadenilation sites, and liable to undergo alternative splicing (Li et al. 2013; Goriaux et al. 2014). On the contrary, dual-strand clusters do not have promoters and use noncanonical read-through transcription from their nearby genes, producing uncapped piRNA

precursors that use the Rhino-Deadlock-Cutoff protein complex (Rhi-Del-Cuff) to avoid degradation (Mohn et al. 2014; Zhang et al. 2014).

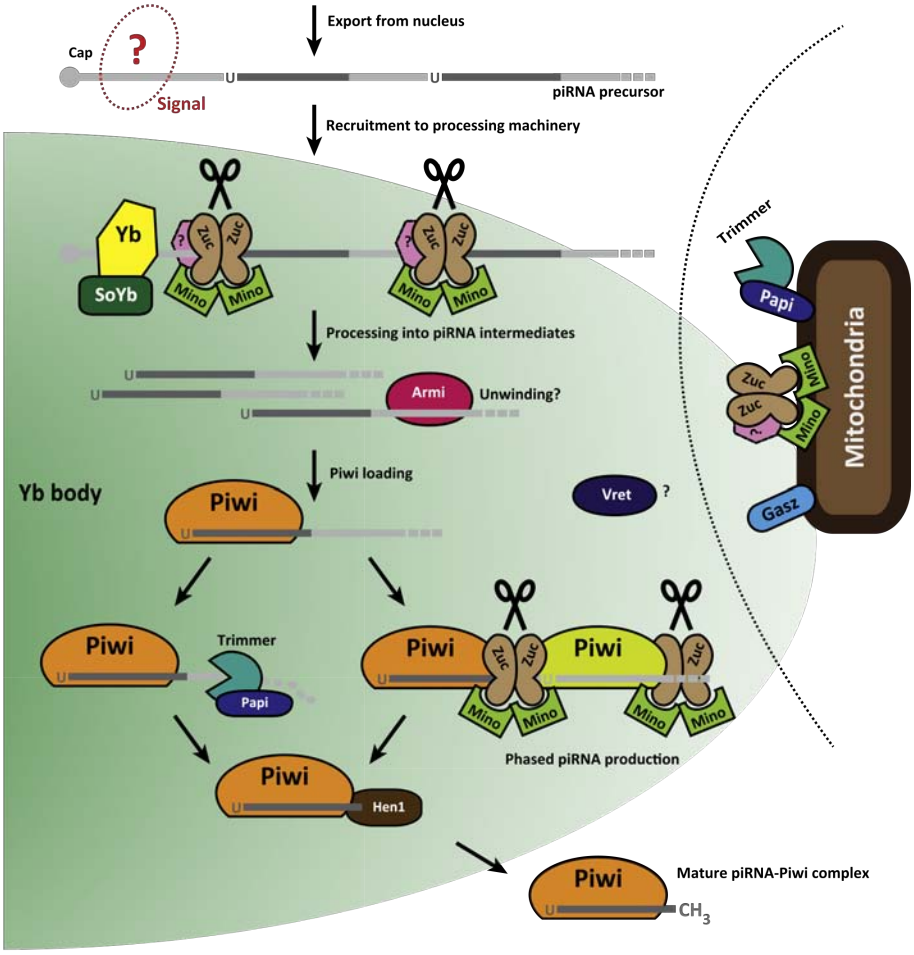


Figure 3: Biogenesis of piRNAs in somatic (follicle) cells of *Drosophila* ovaries. Excerpted from Czech and Hannon (2016).

piRNA precursors have to be transported from the nucleus to the cytoplasm in order to start the piRNA processing that triggers the biogenesis of primary piRNAs (**Figure 3**). Primary piRNA biogenesis takes place in both germ and follicle (somatic) cells of *Drosophila* ovaries, yet with several differences (Malone et al. 2009; Théron et al. 2014). For instance, in germ cells, piRNA processing occurs in a perinuclear structure called *nuage* and requires the helicase Hel25E (Zhang et al. 2012); while in follicle cells it happens at the Yb bodies (Olivieri et al. 2010; Saito et al. 2010). At this step, the nuclease Zucchini (Zuc) produces shorter piRNA intermediates (Ipsaro et al. 2012; Nishimasu et al. 2012), during a process that is still not fully understood but that requires Vreteno (Vret), Minotaur (Mino) and Gasz in both germ and follicle cells (Handler et al. 2011;

Zamparini et al. 2011; Czech et al. 2013; Handler et al. 2013; Vagin et al. 2013). The RNA helicase Armitage (Armi) is thought to act then resolving secondary structures of piRNA precursors (Vourekas et al. 2015). In follicle cells, the protein Yb binds piRNA intermediates and is also needed for piRNA processing (Olivieri et al. 2010; Saito et al. 2010); a function that would be performed by Sister of Yb and Brother of Yb (SoYb and BoYb) in germ cells (Handler et al. 2011). Afterwards, piRNA intermediates are loaded into Piwi-clade proteins –always Piwi in the case of follicle cells, where Aub and Ago3 are not expressed– using the heat shock protein Hsp90 and the co-chaperone Shutdown (Olivieri et al. 2012; Preall et al. 2012). To form the mature piRNA 3' ends, intermediates are trimmed either by an unknown endonuclease (*Trimmer* in **Figure 3**) and its co-factor Papi (Kawaoka et al. 2011; Saxe et al. 2013); or by Zuc, producing phased piRNAs (Han et al. 2015; Mohn et al. 2015; Senti et al. 2015). Finally, mature primary piRNAs are methylated by Hen1 with a protective purpose (Horwich et al. 2007; Saito et al. 2007).

A second path of piRNA biogenesis, called the ping-pong cycle, gives rise to secondary piRNAs in *Drosophila* germ cells (**Figure 4**). The observation of different orientation and nucleotide biases between piRNA populations associated to Piwi, Aub and Ago3 led to the description of a piRNA amplification loop in *Drosophila*; later found to be similar in many other animals, such as silkworm, fish and mouse (Czech and Hannon 2016). In particular, Piwi and Aub mainly bind piRNAs in antisense orientation to transposons that tend to have a uridine (U) in their first position; whereas Ago3 is associated to sense piRNAs with an adenine (A) in their 10th position. Furthermore, sequence complementarity in the first 10 nucleotides between Aub and Ago3-associated piRNAs was observed (Brennecke et al. 2007; Gunawardane et al. 2007).

Remarkably, the ping-pong cycle couples piRNA production with target post-transcriptional silencing (PTS). First, primary or maternal piRNAs (mostly antisense) associated to Aub detect active transposon transcripts by sequence complementarity (step *E* in **Figure 4**). These transcripts are cleaved by Aub slicer activity (hence performing PTS), which produces the 5' end of a new sense piRNA. The resulting fragment is loaded into Ago3 (step *F* in **Figure 4**) and cut either by an unknown nuclease or by Zuc, to form the 3' end of this sense piRNA (step *G*). After methylation by Hen1 (step *H* in **Figure 4**), the piRNA completes a maturation process where the piRNA-Ago3 complex suffers symmetric dymethyl-arginine (sDMA) modifications, added by the methyltransferase Capsuleen (Csul) and its cofactor Valois (Vls) (Kirino et al. 2009; Nishida et al. 2009). Mature piRNAs can then recognize and cleave complementary transcripts from piRNA clusters (step *I* in **Figure 4**), generating a new antisense piRNA that in turn associates to Aub (step *A*). This piRNA complex again undergoes maturation, involving first its trimming by an unknown

endonuclease or by Zuc (step *B* in **Figure 4**). After methylation and sDMA modifications (step *C* and *D* in **Figure 4**), the mature Aub-piRNA complex recognizes new transposon transcripts and cuts them, and so forth, in a sense-antisense complementary piRNA amplification cycle. It is important to note that Zuc-mediated 3' end formation of piRNAs associated to Aub (but not to Ago3) also results in the production of phased antisense piRNAs from the downstream transcript (step *J* in **Figure 4**) (Mohn et al. 2015; Wang et al. 2015). These phased piRNAs are mainly loaded into Piwi and provide sequence diversification in piRNA production, which increases the efficiency of TE targeting (Han et al. 2015; Senti et al. 2015).

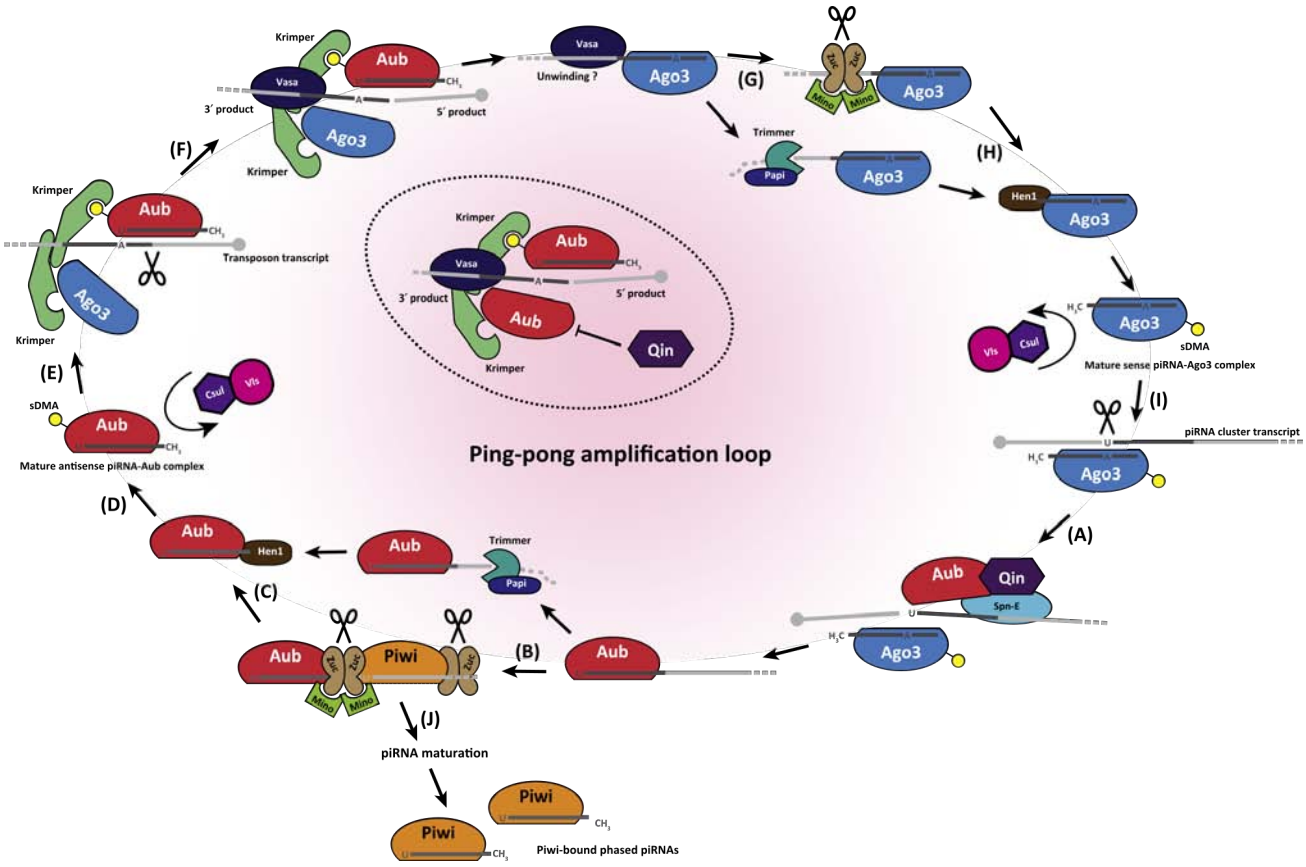


Figure 4: Biogenesis of secondary piRNAs (via ping-pong amplification) and phased piRNAs in germ cells of *Drosophila* ovaries. Excerpted from Czech and Hannon (2016).

Several proteins conforming the nuage are known to participate in the ping-pong cycle, some of them being essential for a successful secondary piRNA production (Lim and Kai 2007). For instance, the RNA helicase Vasa and the Tudor-domain proteins Spindle-E (SpnE) and Qin are known to interact with Aub in *Bombyx mori* germ cells (Nishida et al. 2015). Vasa is thought to prevent target RNAs degradation by transferring the Aub-sliced transcripts to Ago3; while Spindle-

E (SpnE) and Qin are responsible for the correct association of Aub with specific piRNAs (Xiol et al. 2014; Nishida et al. 2015). Remarkably, Qin prevents Aub cleavage products from associating to other Aub proteins, inhibiting homotypic Aub:Aub ping-pong amplification (inset in centre of **Figure 4**) (Zhang et al. 2011). In addition, Krimper (Krimp, also a Tudor-domain protein) interacts with loaded Aub via sDMA binding, and acts sequestering empty Ago3 proteins to prevent its illegitimate association to other RNAs (Sato et al. 2015; Webster et al. 2015). Other factors, such as the Tudor-domain proteins Tudor (Tud), Tejas (Tej) and Tapas are also involved in ping-pong amplification, probably executing scaffolding functions (Nishida et al. 2009; Patil and Kai 2010; Patil et al. 2014).

Besides its role in PTS, the piRNA pathway also represses transposons at a transcriptional level. Indeed, while Aub and Ago3 are localized in germ cells cytoplasm and act in secondary piRNA biogenesis (consequently contributing to PTS), Piwi plays its major role in the nucleus of both germ and follicle cells, where it is essential for transposon TS (Sienski et al. 2012; Rozhkov et al. 2013; Le Thomas et al. 2013; Klenov et al. 2014). Once loaded with mature piRNAs, Piwi proteins can be transported from the cytoplasm to the nucleus. There, they interact with a crucial player of transposon TS, the zinc finger protein Asterix (Arx), forming a complex that scans for nascent transposon transcripts (Dönertas et al. 2013; Muerdter et al. 2013; Ohtani et al. 2013). Panoramix (Panx) associates to Piwi-Arx complex upon target identification, triggering the deposition of H3K9me3 marks (Czech et al. 2013; Handler et al. 2013; Sienski et al. 2015; Yu et al. 2015). These histone modifications are set down by the methyltransferase Eggless (Egg) and its cofactor Windei (Wde), leading to HP1 recruitment and heterochromatin formation (Brower-Toland et al. 2007; Wang and Elgin 2011).

The precise molecular mechanisms involved in transposon TS are not completely understood, and ongoing research is still discovering new factors, interactions and functions that will help to disentangle the intricate details of this pathway. For example, Maelstrom is required for TS and is thought to block H3K9me3 spread downstream of TE sequences, although its presence in the cytoplasm suggests that this may not be its only function (Findley et al. 2003; Sienski et al. 2012). In addition, the lysine-specific demethylase 1 (Lsd1) contributes to TS by removing H3K4me2 marks from TE promoters (Czech et al. 2013; Yu et al. 2015). Finally, it is worth mentioning that TS also involves de novo CpG DNA methylation in the case of mammals (Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008).

2 Objectives

It is now acknowledged that the genomic stress induced by interspecific hybridization promotes TE mobilization in the hybrid offspring of animals and plants. In the case of hybrids between the species *Drosophila buzzatii* and *D. koepferae*, transposition of a total of 28 elements has been detected. However, little is known about the repercussion of this TE proliferation on the hybrid genome, nor about the molecular mechanisms unleashing this TE release. The **global aim of this thesis is to cast light on both the causes and the consequences of TE activation in hybrids**, first by assessing the impact of hybridization on genome size, then by examining TE deregulation at the transcription level, a step that precedes (and is sometimes required for) mobilization.

While genome expansions are frequent in hybrid plants, via polyploidization or transposon proliferation, the effects of interspecific hybridization on genome size have never been assessed in any animal species. Genome size is an important feature in eukaryotic species evolution and may play a role in speciation events, since differences between species' karyotypes can drive incompatibilities in meiosis, hence fostering reproductive isolation.

A failure of the piRNA pathway in *Drosophila* could enhance TE expression and be at the origin of mobilization events in hybrids. Two main explanatory hypotheses are proposed: the maternal cytotype hypothesis, by which the maternal cytoplasm (*D. koepferae*) might be unable to silence paternally inherited TEs; and the global piRNA pathway failure hypothesis, that states that the accumulated divergence of some of the piRNA pathway's proteins between parental species could cause incompatibilities and malfunction of this silencing pathway in hybrids.

The specific objectives of this thesis are described thereafter:

- To assess the male and female genome size of *D. buzzatii*, *D. koepferae* and their F1 and backcrossed hybrids (**chapter 3.1**).
- To determine whether interspecific hybridization has an effect on genome size of males and females (**chapter 3.1**).
- To perform a molecular characterization of one of the elements mobilized in *D. buzzatii–D. koepferae* hybrids, the retrotransposon *Helena*, in both parental species (**chapter 3.2**).
- To describe the impact of interspecific hybridization and introgression on *Helena* transcription and regulation, through three different strategies:
  - By quantifying the *Helena* transcript abundance in males and females of hybrids and parental species, separating gonads from somatic tissues, using quantitative PCR (**chapter 3.2**).
  - By localizing *Helena* transcripts in testes and ovaries of hybrids and parental species using fluorescent *in situ* hybridization (**chapter 3.2**).
  - By analyzing the piRNA populations associated to *Helena* in gonads of parental species and hybrids, using small RNA sequencing (**chapter 3.2**).
- To analyze the effects of hybridization and introgression on global gene and TE expression in gonads using a transcriptomic approach, focusing in:
  - Evaluating the extent of TE deregulation in hybrids compared to gene deregulation (**chapter 3.3**).
  - Assessing the sequence identity and differences in expression of the piRNA pathway proteins and genes (respectively) between our parental species (**chapter 3.3**).
  - Characterizing the piRNA populations of both hybrids and parents (**chapter 3.3**).
- To estimate the divergence time and compare the TE landscapes between the two parental genomes, *D. buzzatii* and *D. koepferae* (**chapter 3.3**).

3 Results

This section, divided in three chapters, encompasses a description of the results obtained in three different studies I carried out during the last three years.

Chapter 3.1 deals with the effects of interspecific hybridization on genome size, assessed in our parental species, *D. buzzatii* and *D. koepferae*, and four sequential generations of hybrids.

Chapter 3.2 depicts the expression dynamics of the retrotransposon *Helena*, an example of TE mobilized in *D. buzzatii-D. koepferae* hybrids, in parental species and four hybrid generations.

Chapter 3.3 reports a transcriptomic analysis that aims to characterize the expression of whole genome TEs in parental species and hybrids, as well as to unveil some molecular mechanisms responsible for the results observed.

3.1 Genome size evaluation in *D. buzzatii*, *D. koepferae* and their hybrids

This chapter consists of the article entitled "*Drosophila* females undergo genome expansion after interspecific hybridization" published in *Genome Biology and Evolution* journal in February 2016 (8(3):556-561).

3.1.1 *Drosophila* females undergo genome expansion after interspecific hybridization

Supplementary material of this article can be found in **Annex 8.1** and is available at *Genome Biology and Evolution* online (http://gbe.oxfordjournals.org/content/8/3/556/suppl/DC1).

Genome Biology and Evolution

Drosophila Females Undergo Genome Expansion after Interspecific Hybridization

Valèria Romero-Soriano[1], Nelly Burlet[2], Doris Vela[3], Antonio Fontdevila[1], Cristina Vieira[2], and María Pilar García Guerreiro[1,*]

[1]Departament De Genètica I Microbiologia (Edifici C), Grup De Genòmica, Bioinformàtica I Biologia Evolutiva. Universitat Autònoma De Barcelona, Spain

[2]Laboratoire De Biométrie Et Biologie Evolutive, UMR5558, Université Lyon 1, Université Lyon, Villeurbanne, France

[3]Laboratorio De Genética Evolutiva, Pontificia Universidad Católica Del Ecuador, Quito, Ecuador

*Corresponding author: E-mail: mariapilar.garcia.guerreiro@uab.cat.

Associate editor: Josefa Gonzalez

Accepted: February 8, 2016

Abstract

Genome size (or C-value) can present a wide range of values among eukaryotes. This variation has been attributed to differences in the amplification and deletion of different noncoding repetitive sequences, particularly transposable elements (TEs). TEs can be activated under different stress conditions such as interspecific hybridization events, as described for several species of animals and plants. These massive transposition episodes can lead to considerable genome expansions that could ultimately be involved in hybrid speciation processes. Here, we describe the effects of hybridization and introgression on genome size of *Drosophila* hybrids. We measured the genome size of two close *Drosophila* species, *Drosophila buzzatii* and *Drosophila koepferae*, their F_1 offspring and the offspring from three generations of backcrossed hybrids; where mobilization of up to 28 different TEs was previously detected. We show that hybrid females indeed present a genome expansion, especially in the first backcross, which could likely be explained by transposition events. Hybrid males, which exhibit more variable C-values among individuals of the same generation, do not present an increased genome size. Thus, we demonstrate that the impact of hybridization on genome size can be detected through flow cytometry and is sex-dependent.

Key words: genome size, flow cytometry, hybrids, *Drosophila*, transposable elements, AFLP markers.

Introduction

Genome size, also known as C-value, is the measure of DNA mass per haploid nucleus (Gregory 2005b) and represents a crucial feature for the understanding of genome evolution and speciation (Kraaijeveld 2010). Although this value is constant within individuals, eukaryotic species present a wide variation in genome size, reaching differences higher than 600,000-fold (Gregory 2005a). The lack of correlation between organisms' genome size and their number of genes or their complexity was called the "C-value paradox," an issue that was cleared up by the finding that genes are not the only (nor the major) components of genomes. It is now known that a large fraction of the genome of most eukaryotic organisms is noncoding repetitive DNA, including transposable elements (TEs), pseudogenes, introns, and satellites (Gregory 2005a). Together with polyploidization, transposition is considered to be one of the major forces of eukaryotic genome expansion (Kidwell 2002): for instance, the maize genome doubled its size during the last few million years after a series of transposition bursts (SanMiguel et al. 1996). In the *Drosophila* genus, some studies have demonstrated that TE amount can account for genome size variation between species (Boulesteix et al. 2006), as well as between populations of the same species (Vieira et al. 2002).

Although TE mobilization rates are usually low, spontaneous transposition bursts have been reported, often linked to different stressful conditions (reviewed in García Guerreiro 2012). Interestingly, some of these bursts seem to share timing with species radiation episodes (Rebollo et al. 2010).

The merge of two different genomes during interspecific hybridization events can be considered a genomic stress condition, which has been shown to lead to transposition bursts in several species. For example, different macropodid hybrids present amplified centromeres due to the presence of TE-related sequences (O'Neill et al. 1998; Metcalfe et al. 2007), and retrotransposon proliferation has also been described in three sunflower species of hybrid origin (Ungerer et al. 2006).

In *Drosophila*, the first evidence of hybrid TE mobilization was the detection of a new insertion of the *pDv111* element in *Drosophila virilis–Drosophila littoralis* hybrids by *in situ* hybridization (Evgen'ev et al. 1982). In the same way, an increase of transposition of the retrotransposon *Osvaldo* was reported in hybrids between *Drosophila buzzatii* and *Drosophila koepferae* (Labrador et al. 1999). More recently, a genome-wide study using AFLP markers in these same hybrids demonstrated that not only *Osvaldo* but at least 28 different TEs were mobilized (Vela et al. 2014), suggesting that transposition in *D. buzzatii–D. koepferae* hybrids is a widespread phenomenon. Other studies at a transcription level support the hypothesis of a TE derepression in hybrids between *Drosophila* species (Kelleher et al. 2012; Carnelossi et al. 2014; García Guerreiro 2015), as well as in hybrid lake whitefishes (Dion-Côté et al. 2014) and sunflowers (Renaut et al. 2014).

Massive bursts of transposition can cause drastic changes in genome size and composition. For instance, three hybrid-derived sunflower species present genome sizes 50% larger than parental species (Baack et al. 2005). This study shows that interspecific hybridization is a source of evolutionary novelties that may be at the origin of new species by the means of TE activation (reviewed in Fontdevila 2005; Rebollo et al. 2010). However, synthetic F_1 and F_6 hybrids between the same sunflower parental species do not present a genome increase, and neither do plants from hybrid-zone populations (Baack et al. 2005; Kawakami et al. 2011). These last results show that genome expansion is not a shared feature of all interspecific hybrids, which concurs with studies in other plants, such as oil palm, sea buckthorns, and grasses, where hybrids presented intermediate genome sizes between parental species (Mahelka et al. 2005; Zhou et al. 2010; Camillo et al. 2014).

In animals, despite the few studies describing TE activation in hybrids (O'Neill et al. 1998; Labrador et al. 1999; Metcalfe et al. 2007; Vela et al. 2014), information about the effect of hybridization on hybrid genome size is scarce. *D. buzzatii* and *D. koepferae* are two cactophilic species that only produce hybrid offspring when crossing *D. buzzatii* males with *D. koepferae* females—the reciprocal cross does not produce adult offspring (Marin et al. 1993). As previously mentioned, mobilization of different TEs in hybrids between these species has been reported by *in situ* hybridization, AFLPs and transposon display techniques (Labrador et al. 1999; Vela et al. 2011, 2014). We have estimated the genome size of these two parental species and their F_1 hybrids, as well as three subsequent generations of backcrossed hybrids (fig. 1). Thus, the present work aims to analyze the impact of interspecific hybridization, at different stages of genomic introgression, on genome size of male and female *Drosophila* hybrids.

Materials and Methods

Drosophila Stocks and Crosses

Six interspecific crosses were performed between ten *D. buzzatii* males (Bu28 strain) and ten *D. koepferae* females (Ko2 strain). Both strains are inbred lines originated by natural populations collected, respectively, in Bolivia and Argentina (Morán and Fontdevila 2014). Each cross was followed by three generations of backcrossing of ten hybrid females with ten *D. buzzatii* males. All stocks and crosses were reared at 25 °C in a standard *Drosophila* medium.

Genome Size Estimation

Genome size of *D. buzzatii*, *D. koepferae*, and their hybrids was estimated for males and females separately using flow cytometry technique. Nuclei were extracted from three heads of exactly 4 days-old flies, using *D. virilis* as internal control standard. Heads were homogenized in Galbraith buffer (30 mM trisodium citrate, 10^{-4} triton X-100, 2 µg/ml RNAse A, 20 mM MOPS, 21.3 mM $MgCl_2$) with 0.1 mg/ml propidium iodide (pH 7.2). After two filtering steps through 140 and 30-micron nylon meshes, samples were analyzed on a FACSCanto II flow cytometer fitted with an argon laser (488 nm wavelength). The relative fluorescence intensity between our flies and *D. virilis*, whose genome size estimate is 0.34 pg (Gregory and Johnston 2008), was determined. We performed 5–6 biological replicates for parental samples and 8–10 for hybrids (supplementary table S3, Supplementary Material online).

Statistical Analyses

Comparisons between parental species genome sizes were performed using the nonparametric Wilcoxon rank sum test (Mann and Whitney 1947), while hybrid genome size estimates were compared to a single theoretical mean (specific to each generation) with the Wilcoxon signed-rank test (Wilcoxon 1945). The single theoretical value specific to each generation was calculated for males and females separately, as follows:

$$\overline{GS}_{th} = \overline{f}_{ko} \times \overline{GS}_{ko} + \overline{f}_{bu} \times \overline{GS}_{bu}$$

where $\overline{f}$ is the *D. buzzatii* ($\overline{f}_{bu}$) or *D. koepferae* ($\overline{f}_{ko}$) mean genome fraction of each generation (for example, for BC1, $\overline{f}_{bu} = 0.75$ and $\overline{f}_{ko} = 0.25$) and $\overline{GS}$ is the mean genome size of *D. buzzatii* ($\overline{GS}_{bu}$) or *D. koepferae* ($\overline{GS}_{ko}$).

AFLP Genotyping

AFLP technique was suitable for our study because it did not require prior information on our species sequences (*D.*

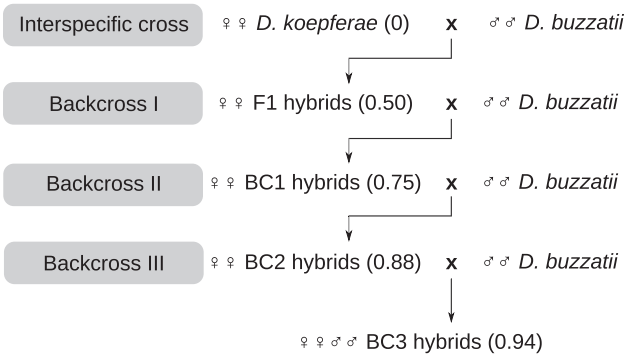


Fig. 1.—Diagram of crosses. A first interspecific massal cross of ten *D. koepferae* females with ten *D. buzzatii* males was followed by three subsequent backcrosses of ten hybrid females with ten *D. buzzatii* males. The *D. buzzatii* expected mean genome fraction of each generation is presented in parentheses.

koepferae available sequences are scarce) and had previously been used in our species and their hybrids (Morán and Fontdevila 2014; Vela et al. 2014). Markers were obtained following the protocol described in Vela et al. 2011, from six hybrid crosses used in a former study (Vela et al. 2014). Contrary to the previous study, where instability markers were checked, we here identified *D. koepferae*-specific markers for ten primer combinations (supplementary table S2, Supplementary Material online). The presence of these markers was then assessed in F_1 and BC1 hybrids, as detailed in supplementary fig. S1, Supplementary Material online. Finally, we determined the mean number of markers found per individual per family as explained in supplementary table S2, Supplementary Material online.

Results and discussion

As a first goal, we have determined both *D. buzzatii* and *D. koepferae* parental genome sizes to assess differences between them, and also between males and females. Our results show that *D. buzzatii* presents a mean C-value of 176.28 Mb for females and 169.07 Mb for males (fig. 2A). These values are significantly higher ($P = 0.022$, supplementary table S1*A*, Supplementary Material online) than the 146–153 Mb previously reported by other authors (Guillén et al. 2015). It is known that differences in the estimated values can depend on the technique used (flow cytometry *vs.* densitometry), the analyzed tissues (heads *vs.* testes) or even on fly rearing conditions (Nardon et al. 2003). Furthermore, it is important to note that the genome size reference used in the former study (Guillén et al. 2015) was the *Drosophila mojavensis* genome assembly size, which could likely suppose an *a priori* underestimation due to assembling issues. Indeed, most of the repeated sequences are not assembled and we know they can contribute to genome size variation. On the other hand, intraspecific variation in *Drosophila* genome size among different strains or populations has been reported in several studies (Vieira et al. 2002; Bosco et al. 2007; Gregory and Johnston 2008; Ellis et al. 2014). These differences have been attributed to changes in TE (Vieira et al. 2002) and satellite DNA amounts (Bosco et al. 2007), and seem to be correlated with several life history traits and metabolism genes expression (Ellis et al. 2014).

In this study, we globally observe that parental females have significantly larger genomes than males ($P = 1.48E-06$, supplementary table S1*A* Supplementary Material online), with differences of approximately 7 Mb for both species (fig. 2A). Similar results have been described in *Drosophila mauritania* or *Drosophila hydei* (Girard and Hannon 2008), but *Drosophila melanogaster* presents equivalent genome sizes for both sexes (Vieira et al. 2002) and *Drosophila simulans* males exhibit larger genomes than females (Vieira et al. 2002). However, different results have been found in other strains of the latter two species (Gregory and Johnston 2008), indicating that genome size differences between males and females are strain-specific and likely depend on specific increases of repetitive DNA in the Y chromosome heterochromatin (Vieira et al. 2002). In our species, we expect females to have a higher genome size than males because X chromosome is known to be longer than Y (Wasserman 1962; Fontdevila et al. 1988). Interestingly, the standard errors observed within replicates are ≈2-fold higher in males than in females, showing that males present greater genome size variability (supplementary table S1, Supplementary Material online). The dynamic gene content of the Y chromosome, which also contains a high amount of repetitive sequences, might account for this diversity (Bernardo Carvalho et al. 2009).

Differences between species are significant for females ($P = 0.015$, supplementary table S1*A*, Supplementary Material online), with *D. koepferae* genome about 3 Mb larger than *D. buzzatii* (fig. 2A). No significant difference was observed in males ($P = 0.126$, supplementary table S1*A*, Supplementary Material online), which is probably due to the lower genome size and the higher variability found in male samples.

According to our null hypothesis, the genome size of hybrids (F_1 and backcrosses) would present intermediate values between parental species and would be proportional to the *D. buzzatii*/*D. koepferae* genome fractions at each generation (fig. 1). Thus, we have compared the C-values of each hybrid generation to a theoretical weighted mean, reflecting the expected mean *D. buzzatii* introgression percentage in the hybrid genomes, assuming independent assortment of chromosomes during meiosis (see Materials and Methods). The accuracy of this assumption has been tested through AFLP genotyping of hybrids and parents: we have used 70 AFLP markers specific to *D. koepferae* and assessed which proportion of these markers is transmitted to hybrid progeny (see below).

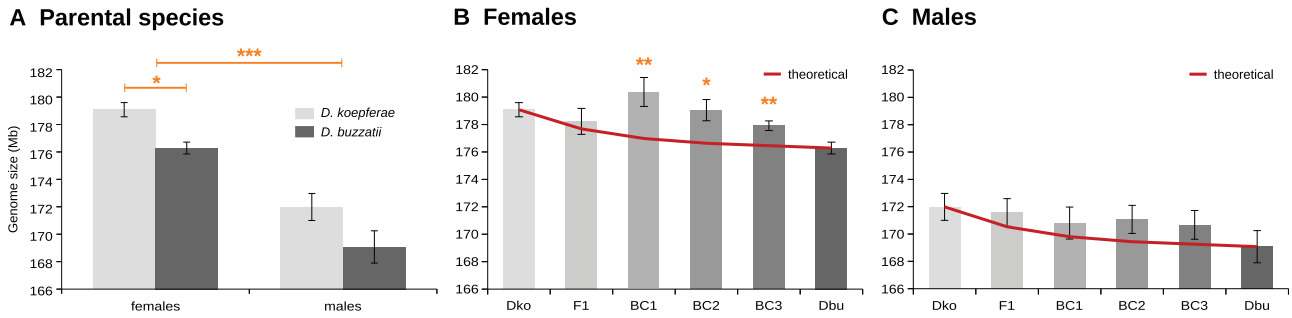


Fig. 2.—(*A*) Parental species mean genome size. *: *P* value < 0.05; **: *P* value < 0.01; ***: *P* value < 0.001 (Wilcoxon rank sum test W significant differences between species and sexes). (*B* and *C*) Mean genome size for parental species and all hybrid generations (gray bars) compared with theoretical mean values (red line) for female (*B*) and male (*C*) samples. Dbu: *D. buzzatii*; Dko: *D. koepferae*. Error bars represent standard error. *: *P* value < 0.05; **: *P* value < 0.01 Not useful (Wilcoxon signed-rank test V significant differences comparing experimental measures with the theoretical value).

In females, we show that the mean genome size of the four hybrid generations is higher than the theoretical value (fig. 2*B*), with statistically significant differences for the three backcrosses (supplementary table S1*B*, Supplementary Material online). The most striking results occur in the first backcross (BC1): the mean C-value (180.37 Mb) increases compared with the F_1 generation (178.23 Mb), and is also higher than in both parental species (176.28 and 179.08 Mb). These results are concordant with the transposition-related instability observed previously in our hybrids, where new insertions of 28 different TEs, including retrotransposons and DNA transposons, were detected in the three backcrosses (Vela et al. 2014). In the case of F_1, the vast majority of the detected instability markers were not transmitted to BC1, showing that the putative transposition events of F_1 take place after meiosis (Vela 2012), which is also coherent with our results: somatic transpositions are not expected to cause a genome size increase. TE activation in hybrids seems to be caused by the failure of epigenetic repression mechanisms (Michalak 2009), such as histone methylation or small RNA biogenesis. In *Drosophila* ovaries, TEs are mainly regulated by piRNAs, a kind of small RNAs associated to Piwi proteins. Differences in piRNA pools between parental species, or incompatibilities between their piRNA pathway effector proteins, might lead to a TE silencing failure in hybrids. If a TE derepression took place in F_1 ovaries at a transcriptional level, as shown for *D. simulans–D. melanogaster* hybrids (Kelleher et al. 2012), we would expect to detect new insertions in the following generations. Thus, new TE insertions could likely be responsible for the genome size increase observed after F_1.

It is worth noting that other phenomena could also account for the observed genome expansion, such as the amplification of satellites or other noncoding repetitive sequences (Bosco et al. 2007), which are responsible for the large *Drosophila orena* genome (Boulesteix et al. 2006). On the contrary, polyploidization can be discarded, since early studies of *D. buzzatii–D. koepferae* hybrids, based in *in situ* hybridization (Labrador et al. 1999), never reported a case of hybrid abnormal karyotype due to genome duplication.

Finally, a transmission bias favoring the larger parental genome, *D. koepferae*, could also be consistent with our results (e.g., due to reduced recombination or differential gamete viability). In order to test this hypothesis, we have determined the inheritance of 70 *D. koepferae*-specific AFLP markers in F_1 and BC1 hybrids from six different crosses. Our results, summarized in supplementary table S2, Supplementary Material online, show that almost 100% (92.9–97.1%) of the studied *D. koepferae*-specific markers are found in F_1, as expected: all F_1 individuals have an entire haploid copy of the *D. koepferae* genome. In the BC1, between 11.8% and 72.9% of the markers are found per individual (supplementary table S2, Supplementary Material online). This variability was also predictable, because inheritance of *D. koepferae* markers depends on the chromosomal assortment and recombination events occurring in each F_1 gamete. Thus, it is not surprising that BC1 and BC2 hybrids present higher standard errors on genome size measurements than parental species (supplementary table S1*B*, Supplementary Material online and fig 2*B*). The average proportion of *D. koepferae* markers found in BC1, 32.4% (95% confidence interval: 11.6–53.2%), is lower than the expected mean of 50%, which suggests that either the transmission of the smallest parental genome (*D. buzzatii*) is favored in BC1 hybrids, or there is not any transmission bias. It is also worth noting than even considering the most extremely biased case ($P = (½)^5 = 0.03$), in which 1) there is no recombination between *D. buzzatii* and *D. koepferae* chromosomes and 2) all individuals inherit all five *D. koepferae* chromosomes from their hybrid mothers; the *D. koepferae* genome fraction in backcrossed hybrids would be of 50% (as in F_1). Assuming these improbable particulars, genome size estimates remain significantly higher than the expected for BC1 (Wilcoxon signed-rank test V = 49, $P = 0.027$).

Despite the genome size is higher than expected in all backcrosses (fig. 2*B*), its value actually decreases through

generations after BC1 (fig. 2B). In rice (*Oryza sativa*), an important increase of *Tos17* and *RCS1* retrotransposons copy number was observed after introgression with *Zizania latifolia*, but no additional insertions were detected after a few generations (Liu and Wendel 2000), meaning that TE mobilization was by then controlled. Thereby, we can suppose that after a few generations of introgression, the preponderance of one of the parental genomes mitigates incompatibilities and palliates the hybridization effects. In this way, we can hypothesize that a greater transposition control in our hybrids would take place after BC1, which according to previous studies is true for the transposon *Galileo*, but not for *Helena* (high transposition rates observed also in BC2) and *Osvaldo* (higher transposition rates in BC3) (Vela et al. 2014). However, these elements represent only a small subset of these species' TEs and may not be representative of the whole set behavior.

The simple backcrossing with *D. buzzatii* (species with smallest genome) could by itself lead to the observed genome size decrease after BC1, but active mechanisms involving genome reduction might also be involved, especially those implicated in TE control. For instance, it is known that internal and complete deletions of TE copies can act as a prevention mechanism against genome invasions (Petrov and Hartl 1998; Liu and Wendel 2000; Senerchia et al. 2015), the latter being guided by the presence of multiple TE copies through recombination events.

The observed genome increase in hybrid females could also be a technical artefact due to changes in chromatin topology. In hybrids, the failure to maintain chromatin integrity could improve the accessibility of DNA to fluorochromes (Nardon et al. 2003), resulting in an increase of genome size estimates. However, this hypothesis can be discarded because the lowest levels of chromatin compaction are expected in F_1, whereas the highest genome size measures belong to BC1.

Regarding our hybrid males, it is worth mentioning that they are all sterile until BC3, when fertility is recovered for some individuals (Morán and Fontdevila 2014). Here, we show that all hybrid generations present intermediate genome size values between *D. buzzatii* and *D. koepferae* (fig. 2C and supplementary table S1*B*, Supplementary Material online). Although the mean C-value of each generation is higher than the theoretical, differences are not significant (supplementary table S1*B*, Supplementary Material online), meaning that the impact of hybridization and introgression on genome size is negligible in males. This seems contradictory with the fact that new TE insertions in our hybrids were also detected in males (Vela et al. 2014), where *Osvaldo* transcription rates were higher than in parents (García Guerreiro 2015). However, these male transposition events were thought to be partly somatic (Vela et al. 2014), and thus would not necessarily lead to a genome expansion. Furthermore, other transposons, such as *Helena*, seem to be repressed in hybrid males (Romero-Soriano and García Guerreiro 2016). This shows that TE regulation patterns differ between sexes and depend on the studied TEs, as proposed in a recent study (Senti et al. 2015). Indeed, the biogenesis of piRNAs has been shown to differ between males and females (Nagao et al. 2010; Siomi et al. 2010). Although we cannot rule out the involvement of particular TEs in the hybrid male sterile phenotype, our results suggest that, unlike hybrid females, males do not present a massive TE amplification.

Conclusions

We have shown that the increased transpositional activity previously reported in *D. buzzatii–D. koepferae* hybrids has an impact on hybrid female genome size. For the first time, an actual genome size increase due to interspecific hybridization has been described in animals. This allows us to validate flow cytometry as a technique to detect changes in C-value of *Drosophila* hybrids, probably due to transposition events. In males, the effects of hybridization are not significant, but we must note that changes in their genome size would lack direct evolutionary consequences, since they are all sterile until some individuals recover their fertility in BC3.

Supplementary Material

Supplementary figure S1 and tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors want to thank M. Santos and two anonymous reviewers for helpful discussions. This work was supported by research grant CGL2013-42432-P from the Ministerio de Economía y Competitividad (Spain) and grant 2014 SGR 1346 from Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (GGBE). V.R.-S. was supported by a PIF PhD fellowship from the Universitat Autònoma de Barcelona (Spain).

Literature Cited

Baack EJ, Whitney KD, Rieseberg LH. 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. New Phytol. 167:623–630.

Bernardo Carvalho A, Koerich LB, Clark AG. 2009. Origin and evolution of Y chromosomes: *Drosophila* tales. Trends Genet. 25:270–277.

Bosco G, Campbell P, Leiva-Neto JT, Markow TA. 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics 177:1277–1290.

Boulesteix M, Weiss M, Biémont C. 2006. Differences in genome size between closely related species: the *Drosophila melanogaster* species subgroup. Mol Biol Evol. 23:162–167.

Camillo J, et al. 2014. Reassessment of the genome size in *Elaeis guineensis* and *Elaeis oleifera*, and its interspecific hybrid. Genomics Insights 7:13–22.

Carnelossi EAG, et al. 2014. Specific activation of an *I*-like element in *Drosophila* interspecific hybrids. Genome Biol Evol. 6:1806–1817.

Dion-Côté AM, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. Mol Biol Evol. 31:1188–1199.
Ellis LL, et al. 2014. Intrapopulation genome size variation in *D. melanogaster* reflects life history variation and plasticity. PLoS Genet. 10:e1004522
Evgen'ev MB, Yenikolopov GN, Peunova NI, Ilyin YV. 1982. Transposition of mobile genetic elements in interspecific hybrids of *Drosophila*. Chromosoma 85:375–386.
Fontdevila A. 2005. Hybrid genome evolution by transposition. Cytogenet Genome Res. 110:49–55.
Fontdevila A, et al. 1988. *Drosophila koepferae*: a new member of the *Drosophila serido* (Diptera: Drosophilidae) superspecies taxon. Ann Entomol Soc Am. 81:380–385.
García Guerreiro M. 2012. What makes transposable elements move in the *Drosophila* genome? Heredity 108:461–468.
García Guerreiro MP. 2015. Changes of *Osvaldo* expression patterns in germline of male hybrids between the species *Drosophila buzzatii* and *Drosophila koepferae*. Mol Genet Genomics. 290:1471–1483.
Girard A, Hannon GJ. 2008. Conserved themes in small-RNA-mediated transposon control. Trends Cell Biol. 18:136–148.
Gregory TR. 2005a. Genome size evolution in animals. In: The Evolution of the Genome, edited by Gregory T.R. San Diego: Elsevier. pp. 3–87.
Gregory TR. 2005b. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. Ann Bot. 95:133–146.
Gregory TR, Johnston JS. 2008. Genome size diversity in the family Drosophilidae. Heredity 101:228–238.
Guillén Y, et al. 2015. Genomics of ecological adaptation in cactophilic *Drosophila*. Genome Biol Evol. 7:349–366.
Kawakami T, Dhakal P, Katterhenry AN, Heatherington CA, Ungerer MC. 2011. Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons. Genome Biol Evol.. 3:156–167.
Kelleher ES, Edelman NB, Barbash DA. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol. 10:e1001428
Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49–63.
Kraaijeveld K. 2010. Genome size and species diversification. Evol Biol. 37:227–233.
Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of *Osvaldo*. Mol Biol Evol. 16:931–937.
Liu B, Wendel JF. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874–880.
Mahelka V, Suda J, Jarolímová V, Trávnicek P, Krahulec F. 2005. Genome size discriminates between closely related taxa *Elytrigia repens* and *E. intermedia* (Poaceae: Triticeae) and their hybrid. Folia Geobot. 40:367–384.
Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 18:50–60.
Marin I, Ruiz A, Pla C, Fontdevila A. 1993. Reproductive relationships among ten species of the *Drosophila repleta* group from South America and the West Indies. Evolution 47:1616–1624.
Metcalfe CJ, et al. 2007. Genomic instability within centromeres of interspecific marsupial hybrids. Genetics 177:2507–2517.
Michalak P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. Heredity 102:45–50.
Morán T, Fontdevila A. 2014. Genome-wide dissection of hybrid sterility in *Drosophila* confirms a polygenic threshold architecture. J Hered. 105:381–396.
Nagao A, Mituyama T, Huang H, Chen D, Siomi MC. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. RNA 16:2503–2515.
Nardon C, Weiss M, Vieira C, Biémont C. 2003. Variation of the genome size estimate with environmental conditions in *Drosophila melanogaster*. Cytometry A 55A:43–49.
O'Neill RJW, O'Neill MJ, Marshall Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393:68–73.
Petrov DA, Hartl DL. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol. 15:293–302.
Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: towards new species. Gene 454:1–7.
Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. Philos Trans R Soc Lond B Biol Sci. 369:20130345
Romero-Soriano V, García Guerreiro MP. 2016. Expression of the retrotransposon Helena reveals a complex pattern of TE Deregulation in *Drosophila Hybrids*. PLoS One. doi: 10.1371/journal.pone.0147903.
SanMiguel P, et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768.
Senerchia N, Parisod C, Parisod C. 2015. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proc R Soc B Biol Sci. 282:20142874
Senti K-A, Jurczak D, Sachidanandam R, Brennecke J. 2015. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. Genes Dev. 29:1747–1762.
Siomi MC, Miyoshi T, Siomi H. 2010. piRNA-mediated silencing in *Drosophila* germlines. Semin Cell Dev Biol. 21:754–759.
Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol. 16:R872–R873.
Vela Peralta DJ. 2012. Estudio de la inestabilidad genómica inducida por transposición en los híbridos interespecíficos de *Drosophila buzzatii* y *Drosophila koepferae*. Universitat Autònoma de Barcelona.
Vela D, Fontdevila A, Vieira C, García Guerreiro MP. 2014. A genome-wide survey of genetic instability by transposition in *Drosophila* hybrids. PLoS One 9:e88992
Vela D, García Guerreiro MP, Fontdevila A. 2011. Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. Biotechniques 50:247–250.
Vieira C, Nardon C, Arpin C, Lepetit D, Biémont C. 2002. Evolution of genome size in *Drosophila* is the invader's genome being invaded by transposable elements? Mol Biol Evol. 19:1154–1161.
Wasserman M. 1962. Cytological studies of the *repleta* group of the genus *Drosophila*. V. The *mulleri* subgroup. Univ Texas Publ. 6205:85–118.
Wilcoxon F. 1945. Individual comparisons of grouped data by ranking methods. Biometrics Bull. 1:80–83.
Zhou X, et al. 2010. Genome size of the diploid hybrid species *Hippophae goniocarpa* and its parental species, *H. rhamnoides* ssp. sinensis and *H. neurocarpa* ssp. neurocarpa (Elaeagnaceae). Acta Biol Cracoviensia Ser Bot. 52:12–16.

3.2 Expression of *Helena* in *D. buzzatii*, *D. koepferae* and their hybrids

This chapter is composed by the paper entitled “Expression of the retrotransposon *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids” published in *PLoS ONE* journal in January 2016 (11(1):e0147903).

3.2.1 Expression of the retrotransposon *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids

Supplementary material of this article can be found in **Annex 8.2** and is available online at the PLoS ONE website (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147903#sec022).

PLOS ONE

RESEARCH ARTICLE

Expression of the Retrotransposon *Helena* Reveals a Complex Pattern of TE Deregulation in *Drosophila* Hybrids

Valèria Romero-Soriano, Maria Pilar Garcia Guerreiro*

Grup de Genòmica, Bioinformàtica i Biologia Evolutiva, Departament de Genètica i Microbiologia (Edifici C), Universitat Autònoma de Barcelona, 08193, Bellaterra, Barcelona, Spain

* mariapilar.garcia.guerreiro@uab.cat

Citation: Romero-Soriano V, Garcia Guerreiro MP (2016) Expression of the Retrotransposon *Helena* Reveals a Complex Pattern of TE Deregulation in *Drosophila* Hybrids. PLoS ONE 11(1): e0147903. doi:10.1371/journal.pone.0147903

Editor: Pawel Michalak, Virginia Tech Virginia, UNITED STATES

Received: September 1, 2015

Accepted: January 11, 2016

Published: January 26, 2016

Data Availability Statement: Sequence data from this article have been deposited in GenBank repository with the accession numbers KF280391, KP115213, KP115214 and KP115215.

Funding: This work was supported by research grant CGL2013-42432-P from the Ministerio de Economía y Competitividad (Spain) and grant 2014 SGR 1346 from Generalitat de Catalunya to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva (GGBE). VRS was supported by a PIF PhD fellowship from the Universitat Autònoma de Barcelona (Spain). The funders had no role in study

Abstract

Transposable elements (TEs), repeated mobile sequences, are ubiquitous in the eukaryotic kingdom. Their mobilizing capacity confers on them a high mutagenic potential, which must be strongly regulated to guarantee genome stability. In the *Drosophila* germline, a small RNA-mediated silencing system, the piRNA (Piwi-interacting RNA) pathway, is the main responsible TE regulating mechanism, but some stressful conditions can destabilize it. For instance, during interspecific hybridization, genomic stress caused by the shock of two different genomes can lead, in both animals and plants, to higher transposition rates. A recent study in *D. buzatii—D. koepferae* hybrids detected mobilization of 28 TEs, yet little is known about the molecular mechanisms explaining this transposition release. We have characterized one of the mobilized TEs, the retrotransposon *Helena*, and used quantitative expression to assess whether its high transposition rates in hybrids are preceded by increased expression. We have also localized *Helena* expression in the gonads to see if cellular expression patterns have changed in the hybrids. To give more insight into changes in TE regulation in hybrids, we analysed *Helena*-specific piRNA populations of hybrids and parental species. *Helena* expression is not globally altered in somatic tissues, but male and female gonads have different patterns of deregulation. In testes, *Helena* is repressed in F1, increasing then its expression up to parental values. This is linked with a mislocation of *Helena* transcripts along with an increase of their specific piRNA levels. Ovaries have additive levels of *Helena* expression, but the ping-pong cycle efficiency seems to be reduced in F1 hybrids. This could be at the origin of new *Helena* insertions in hybrids, which would be transmitted to F1 hybrid female progeny.

Introduction

Hybridization between species is well-known to cause genomic stress that leads to genetic instability in offspring. Hybrids show several features, including polyploidy (common in plants), high rates of chromosomal rearrangements, increased mutation rates, and high transpositional

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

activity [1]. These genome reorganizations are often considered *dysfunctions*, but several cases of hybrid speciation show their evolutionary potential (reviewed in [1]). Furthermore, interspecific hybridization, previously considered a very rare phenomenon in nature, is now estimated to occur in 25% of plants and 10% of animal species, suggesting that its potential has been largely underestimated [2].

Transposable elements (TEs) are dispersed repeated sequences found in the vast majority of the genomes of sequenced species. They have been proposed as major drivers of the genome reorganization occurring during hybridization. Both in animals [3,4] and plants [5–7], examples of TE mobilization due to interspecific crosses have been reported. In *Drosophila*, the pDv111 element transposes in *D. virilis—D. littoralis* hybrids [8], as does the retrotransposon *Osvaldo* in *D. buzzatii—D. koepferae* hybrids [9]. A whole-genome study of the latter hybrids using AFLP markers [10] found 70% of the hybrid instability to be caused by transposition events [11]. Increase of *Osvaldo* expression also occurred in hybrid testes [12]. In the same way, a widespread derepression of TEs at the expression level was noted in hybrids of *D. melanogaster* and *D. simulans* [13].

TEs can be divided in two classes, according to their transposition mechanism [14]. DNA transposons (or class II transposons) are TEs that do not require an RNA intermediate to mobilize: they are excised from their insertion site and inserted in a new position in the genome. Elements of class I, called retrotransposons (or RNA transposons), need a retrotranscription step to transpose: the TE is transcribed, its mRNA reverse transcribed and the resulting cDNA integrated in a new site. A recent classification divides each class into orders and superfamilies according to a more detailed description of their replication strategy and structural features [15]. Long Insterspersed Nuclear Elements (LINEs) are present throughout the eukaryotic kingdom, constituting one of the five distinct orders of class I elements included in this classification, characterized by the production of 5'-end truncated copies.

Helena is a LINE-like retrotransposon first described in *D. virilis*, as being responsible for one of the isolated mutations in the offspring of hybrid dysgenic crosses [16]. More recently, it has been found at different stages of its life cycle across the *Drosophila* genus [17]: absent or present, autonomous or not, and expressed or silenced. Although it is present in 8 out of 12 *Drosophila* sequenced genomes, its expression has only been detected in *D. mojavensis* [17] and some strains of *D. simulans* [18]. Our latest results showed that this element is also expressed in *D. koepferae* and *D. buzzatii*, with an increase in its transposition rate occurring in their hybrids (10^{-2}) compared to parental species ($0\text{-}10^{-3}$) [11].

On the other hand, TE expression in *Drosophila* is regulated through two small RNA-mediated silencing pathways. In somatic cells, the endogenous small interference RNA (endo-siRNA) pathway is the main TE silencing mechanism [19–22]. In gonads, the Piwi-interacting RNA (piRNA) pathway is important in TE repression at both transcriptional and post-transcriptional level [23]. The primary piRNA biogenesis involves the cleavage of long piRNA precursors, transcribed from specific genomic piRNA clusters [24]. These are loaded into a piRNA amplification loop, called the ping-pong cycle, that gives rise to secondary piRNAs [24].

Our aim has been to disentangle the molecular mechanisms responsible for transposition bursts during hybridization events. Initially we molecularly characterized the retrotransposon *Helena* in our target species, *D. buzzatii* and *D. koepferae*. Subsequently, analyses were run at three levels:

1. Quantification of *Helena* expression by quantitative real time PCR (qRT-PCR) in the offspring of crosses between *D. koepferae* females and *D. buzzatii* males (the reciprocal cross being unsuccessful), as well as in three subsequent generations of backcrossing hybrid females with *D. buzzatii* males (Fig 1). The effects of *D. buzzatii* introgression in our hybrid

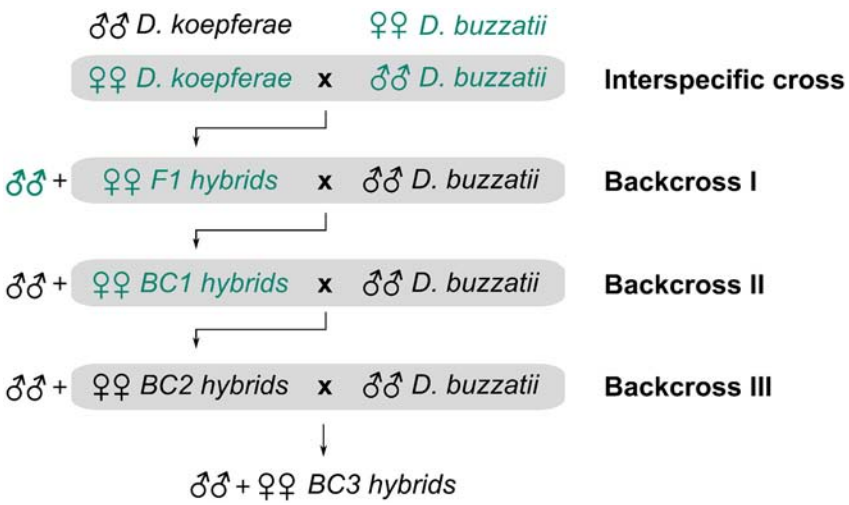


Fig 1. Crosses diagram. A first interspecific cross of 10 *D. koepferae* females with 10 *D. buzzatii* males was followed by three successive backcrosses of hybrid females with *D. buzzatii* males. Samples whose piRNA populations have been analysed are marked in green.

doi:10.1371/journal.pone.0147903.g001

genomes are particularly interesting since F1 males are sterile and the genetic variability created through interspecific hybridization can only be maintained by mating F1 females with parental males. Furthermore, increase of *Helena* transposition seen previously took place mainly in BC2 [11].

2. Localization of *Helena* transcripts in testes and ovaries by fluorescent *in situ* hybridization (FISH), to see if qualitative changes in *Helena* cellular expression patterns occurred after interspecific hybridization.

3. Analysis of *Helena* piRNA populations in germinal tissues of parents and hybrids, because breakdown of the TE silencing mechanisms could be responsible for their derepression in hybrids.

Unexpectedly, *Helena* expression tended to decrease in F1 hybrid testes compared to parental species. This repression might be explained by the high levels of *Helena*-specific piRNAs in hybrid testes, which seem to be mainly produced by the primary piRNA biogenesis pathway. The abundance of *Helena* transcripts in ovaries was significantly different between *D. koepferae* and *D. buzzatii*, but all hybrids have intermediate values. However, the ping-pong signature decreased, especially in F1 hybrid ovaries. Thus, a partial failure of the ping-pong amplification loop seems to be responsible for derepression of *Helena* in hybrid ovaries, which may sometimes be at the origin of transposition events. However, this activation seems to be compensated in some way by the production of *Helena*-specific primary piRNAs.

Results

Helena characterization in *D. buzzatii* and *D. koepferae* species

To characterize *Helena* and analyse its expression, the preliminary goal was to determine the sequence of this TE, which has not previously been done in our target species. From the *Helena* sequence of *D. mojavensis* [17], the most closely related sequenced species, we amplified a fragment of the TE in *D. buzzatii* (one copy) and *D. koepferae* (three copies: Table 1). For

Table 1. Characterization of *Helena* sequenced copies in *D. buzzatii* and *D. koepferae*.

Species	Length (bp)	Alignments vs. *D mojavensis* consensus [17]			Conserved domains			ORFs	
		% coverage	% identity	E-value	PRE_C2HC	EEP	RT	ORF1	ORF2
D. buzzatii	3840	85	88	0	+	+	+	+	+
***D. koepferae* 28**	2806	62	87	0	+	+	+	+	stop
***D. koepferae* 35–1**	3222	71	88	0	+	+	+	stop	stop
***D. koepferae* 35–2**	3247	72	86	0	+	+	-	+	stop
D. mojavensis	4502	-	-	-	+	+	+	+	+

PRE_C2HC: upstream to Cys-Cys-His-Cys (Zn finger motif) domain, EEP: Endonuclease, Exonuclease, Phosphatase; RT: Reverse Transcriptase. For conserved domain analysis [25]: "+" indicates domain presence and "–" indicates domain absence; for ORF analysis: "+" indicates untruncated gene and "stop" indicates that the ORF is interrupted by a stop codon.

doi:10.1371/journal.pone.0147903.t001

D. buzzatii, a 3840 bp sequence was obtained that covers 85% of the *D. mojavensis* consensus copy with 88% of identity, including almost the entire coding region of *Helena*. Even if the sequence is not a complete copy, two overlapping ORFs were identified; the first (*gag-like* protein) harbours a conserved PRE_C2HC domain (upstream of Cys-Cys-His-Cys Zn finger domain), and the second (*pol-like* protein) contains an exonuclease-endonuclease-phosphatase (EEP) as well as a reverse transcriptase domain. No premature stop codons were present, suggesting that the cloned amplicon could be an active *Helena* copy, although the complete sequence of this insertion is needed for confirmation. For *D. koepferae*, three different copies were sequenced, two of them (called *35–1* and *35–2*, with 3222 and 3247 bp, respectively) using a long template PCR system, and the other one (called *28*, with 2806 bp) using a different pair of primers. These sequences cover 62–72% of the *D. mojavensis* consensus copy with an identity of 86–88%. ORF1 (*gag*) seemed to be complete in two of the copies (*35–2* and *28*), but ORF2 (*pol*) carried deletions and was interrupted by premature stop codons in all three copies. However, all the described conserved domains could be identified and the two ORFs also overlapped in the three sequences.

Alignments of the *Helena* sequenced copies showed a high degree of sequence identity between *D. buzzatii* and *D. koepferae*, from 89 to 98% (S1 Table). Interestingly, the closest match was the unique copy of *D. buzzatii* and *D. koepferae*-28, being *D.koepferae*-35-1 the most divergent sequence. *D. koepferae*-35-1 and 35–2 share several internal deletions (two short deletions of 12 and 17 bp and a long one of 557 bp) compared to *D. buzzatii*. *D. koepferae*-35-1 also carries another 43 bp deletion and two short insertions of 9 and 6 bp. Although *D. koepferae*-28 does not seem to have any deletion compared to *D. buzzatii*, it is noteworthy that a different reverse primer was used to amplify this copy of *Helena*, and the presence of mutations after the primer region cannot be discarded.

A phylogenetic analysis of the *Helena* consensus copy identified in all *Drosophila* sequenced genomes [17,18] and other *Helena* characterized sequences [16,26] was made, together with our four sequences. The phylogenetic tree (Fig 2) divides the sequences in two clades that correspond to the *Drosophila* and *Sophophora* subgenera. The *Sophophora* clade is in concordance with its species phylogeny, except for *D. erecta*, which is actually grouped with *D. yakuba*. Within the *Drosophila* clade, *D. buzzatii* and *D. koepferae* form a monophyletic cluster, which is a sister group to *D. mojavensis*, expected in accordance with a vertical transmission scenario. According to this analysis, *D. buzzatii* and *D-koepferae*-28 have the closest related sequences, whereas *D. koepferae*-35-1, as previously seen in the alignments, is the most divergent copy.

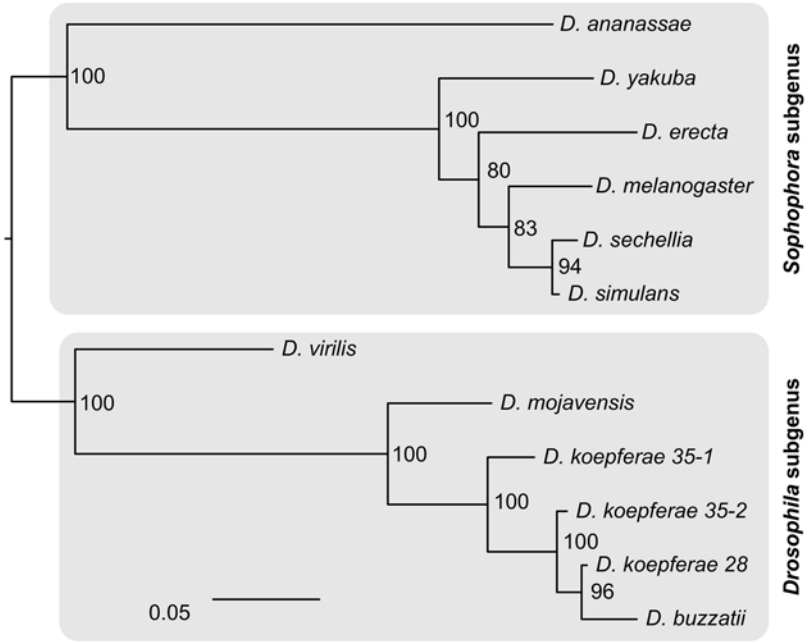


Fig 2. Maximum likelihood phylogenetic tree of *Helena* in the *Drosophila* genus, rooted using the midpoint-root option. Sequences are identified by the host species name. Numbers indicate nodal support, calculated using RAxML with 100 bootstrap replicates.

doi:10.1371/journal.pone.0147903.g002

The *Helena* copy number in the parental species was estimated by Southern blot (S1 Fig), a technique that allows the detection of both euchromatic and heterochromatic insertions. *D. buzzatii* has a higher *Helena* copy number (12–15 copies) than *D. koepferae* (6–12 copies). These results are in agreement with previous studies by FISH on polytene chromosomes [11], where 12 *Helena* euchromatic copies were detected in *D. koepferae* and 5 in *D. buzzatii*.

Overview of *Helena* expression in parental species

Helena expression has first been quantified in parental species by quantitative reverse transcription PCR (qRT-PCR), to evaluate differences in expression rates between *D. koepferae* (female parental species, Fig 1) and *D. buzzatii* (male parental species). Expression rates (ERs) were estimated using the comparative C_T method [27]. No introns have been described in the *Helena* sequence of any of the species in which this TE has been characterized [16–18,26,28]. Furthermore, the amplified fragment had the same length in both parental species, whether we used DNA or cDNA as a template [11], showing that our analyses concerned the only *Helena* splicing variant.

Germinal (testes or ovaries) and somatic tissues (*i.e.* male or female carcasses lacking testes or ovaries) were investigated separately. The results (summarized in S1 File) showed that ERs in somatic tissues are similar between sexes and species ($ER \approx 10^{-4}$, Fig A in S1 File). Indeed, neither the Wilcoxon rank sum test (which compares pairs of samples) nor the Kruskal-Wallis test (which compares multiple samples at once, $\chi^2 = 2.7139$; p-value = 0.4379) show significant differences between somatic tissue samples (Fig C in S1 File). For gonadal samples, the ERs were higher in the testes than the ovaries (Fig B in S1 File); in this case differences between sexes were statistically significant (Wilcoxon's W = 5, p-value = 4.9×10^{-7}, considering all parental samples). Concerning differences between species, *Helena* ERs in ovaries were significantly higher for *D.buzzatii* than *D. koepferae* (Fig D in S1 File, p-value = 6.66×10^{-4}). However,

expression in the testes of the parental species was not significantly different. Contrary to the results in somatic tissue, the Kruskal-Wallis test indicated significant differences in gonads ($\chi^2 = 22.7049$; p-value = 4.653×10^{-5}). It is worth noting the presence of some outlier replicates with particularly high ER values, which occurs mostly in *D. buzzatii*, the parental species with the highest transposition rate (8.2×10^{-3} vs. 0 in *D. koepferae* [11]). However, variances between parental species ERs were only statistically different in the ovaries (Levene's test, S2 Table).

Helena expression in hybrid somatic tissue

The ERs of *Helena* retrotransposon were investigated across four sequential generations of *D. buzzatii-D.koepferae* hybrids, a first interspecific cross and three subsequent backcrosses of hybrid females with *D. buzzatii* males (Fig 1). Our aim was to compare the ERs of each hybrid generation (F1, BC1, BC2 and BC3) with both parental species values. It is noteworthy that expression values of *D. koepferae* female samples (female parental species) and *D. buzzatii* male samples (male parental species) were those of the individuals involved in crosses to obtain F1 hybrids (each replicate belonging to a different cross). However, because *D. buzzatii* females and *D. koepferae* males are not involved in hybrid crosses, several individuals from the same laboratory stocks used in hybrid crosses were analyzed (Fig 3, in red).

In the case of males (Fig 3A), there were no obvious differences between parental and hybrid values for most of the replicates (average median of generations: ER = 2×10^{-4}), although BC2 samples seem to have slightly higher expression rates compared to the other generations. However, neither the Wilcoxon rank sum test (Table 2) nor the Kruskal-Wallis test ($\chi^2 = 1.931$; p-value = 0.8586) showed significant differences between hybrids and parental species. There were a few outranged values (considering ER$\geq10^{-3}$, represented by triangles in Fig 3), which might have been due to occasional transcription bursts taking place in F1, BC1 and BC2 (ER = 4.5×10^{-2}, 1.7×10^{-3} and 4.3×10^{-3}, respectively). Indeed, Levene's test for equality of variances shows that there were significant differences between generations (taking into account all samples: W = 1.45, p-value = 6.73E-06) and, in particular, F1 males had increased variance compared to *D. buzzatii* (S2 Table).

For females (Fig 3B), similar *Helena* ERs between generations were detected (average median of generations: ER = 10^{-4}). The highest expression rates belonged to one BC3 and some F1 biological replicates, but none of them reached the ER = 10^{-3} threshold. At a statistical level, there were no significant differences between parents and hybrids (Table 2), the groups not being distinguishable (Kruskal-Wallis test, $\chi^2 = 2.6058$; p-value = 0.7605), nor were their variances statistically different (Levene's test, W = 1.56, p-value = 0.1893 and S2 Table).In conclusion, *Helena* expression rates in somatic tissue do not change significantly after interspecific hybridization. However, in males, a few exceptional crosses gave outranged ER values responsible for the increase of variance between F1 hybrids and parents.

Helena expression in hybrid germinal tissue

Analogous experiments and analyses to those on somatic tissues were carried out on gonads of both males and females. ERs in testes (Fig 3C) were globally higher than in somatic tissues (average median between generations: ER = 4×10^{-4}). Comparing ERs between different generations, F1 testes seem to have the lowest transcript levels of the retrotransposon *Helena*; in fact, the Wilcoxon rank sum test indicated significant differences between F1 and both parental species expression rates (Table 2). No statistically significant differences were found between the other hybrid generations and parental species, except for BC3 and *D. koepferae* (the parental species showing the highest expression rates), these being at the boundaries of significance (p-value = 0.049, Table 2). The Kruskal-Wallis test also showed significant differences ($\chi^2 = 11.2107$;

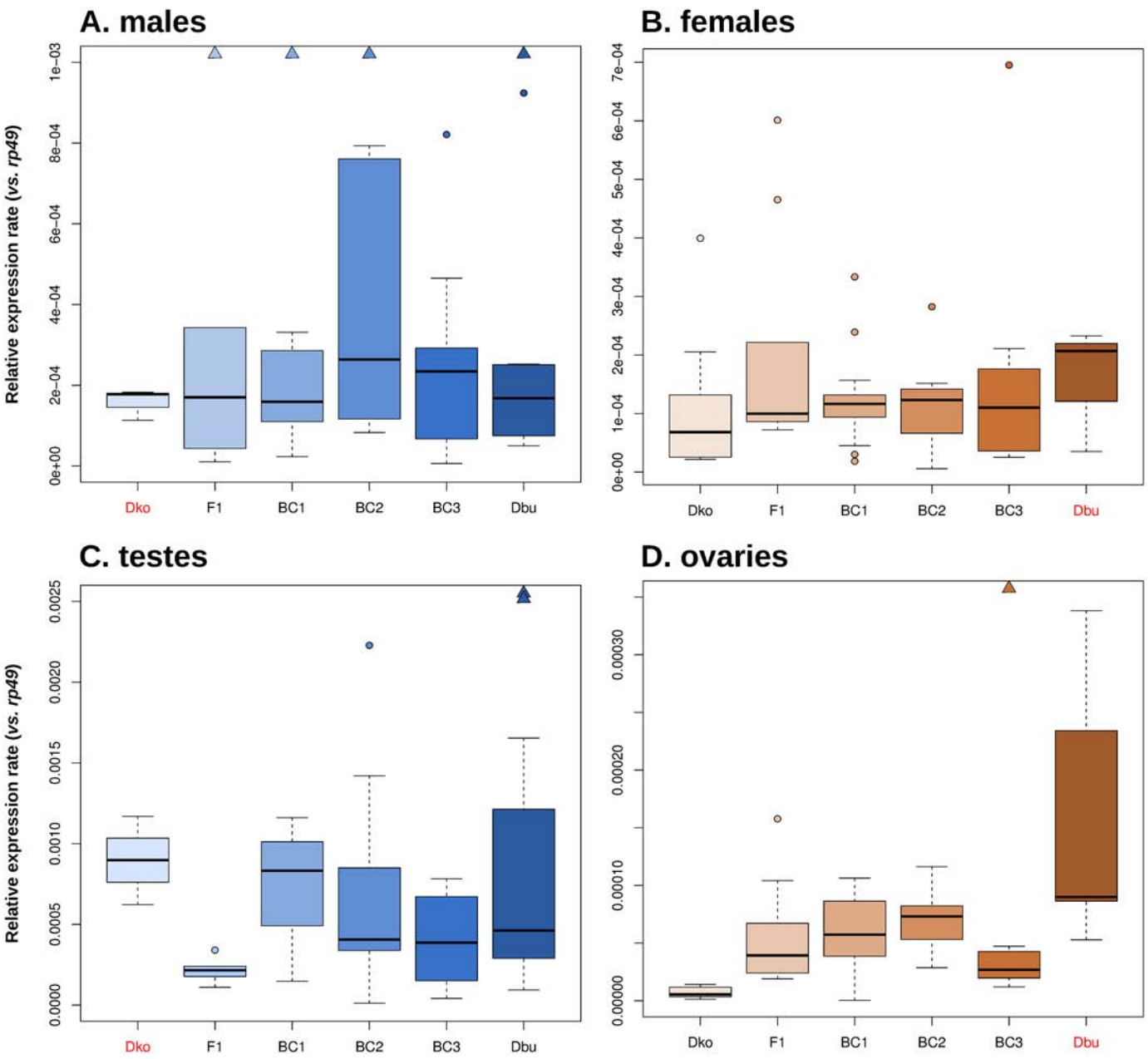


Fig 3. *Helena* expression rates relative to *rp49* housekeeping gene in parental species (*Dko* and *Dbu*) and hybrids. Boxes are determined by the first and third quartile values, with an intermediate deep line corresponding to the median value. Circles correspond to outliers (above or below 1.5-fold the interquartile range), and triangles represent those outliers whose ERs are extremely outranged and cannot be represented in the same scale. Male samples are represented in blue and female samples are represented in brown: the darker the colour, the higher the *D. buzzatii* genome fraction. Parental species which are not part of the interspecific crosses (*i.e.*, *Dko* for male tissues and *Dbu* for female tissues) are marked in red. **A)** results of male somatic tissues (outranged values represented by triangles are: ER = 4.5×10^{-2} for F1, ER = 1.7×10^{-3} for BC1, ER = 4.3×10^{-3} for BC2. ER = 2.9×10^{-3} for *Dbu*), **B)** results of female somatic tissues, **C)** results of testes (*Dbu* outranged values represented by triangles are: ER = 6.2×10^{-3} and ER = 3.6×10^{-3}), **D)** results of ovaries (BC3 outranged value represented by a triangle: ER = 8.5×10^{-3}).

doi:10.1371/journal.pone.0147903.g003

p-value = 0.04736), and in this case, the single outlier value seen in hybrids (BC2) proved to be lower than those in *D. buzzatii*. Levene's test indicated that variances were significantly different between generations (including all samples: W = 5.58, p-value = 3.934E-04), mainly due to the higher variance of *D. buzzatii* samples compared to hybrids (S2 Table). Thus, *Helena* expression

Table 2. Comparisons of *Helena* expression rates between each hybrid generation and both parental species (*D. buzzatii* and *D. koepferae*).

		N	median	SD	vs.*D. buzzatii*		vs. *D. koepferae*	
					W	p-value	W	p-value
males	**F1**	6	1.70E-04	1.81E-02	35	8.84E-01	9	1.00E+00
	BC1	11	1.59E-04	4.76E-04	59	9.49E-01	15	8.85E-01
	BC2	10	2.64E-04	1.29E-03	39	2.82E-01	8	2.87E-01
	BC3	9	2,34E-04	2.54E-04	48	9.41E-01	11	7.27E-01
females	**F1**	9	9.98E-05	1.92E-04	13	1.00E+00	23	1.36E-01
	BC1	13	1.16E-04	8.48E-05	24	6.11E-01	47.5	4.83E-01
	BC2	12	1.23E-04	7.31E-05	24	4.48E-01	47	6.51E-01
	BC3	10	1.10E-04	2.00E-04	19	5.54E-01	37	5.40E-01
testes	**F1**	5	2.16E-04	8.49E-05	47	2.75E-02*	15	3.57E-02*
	BC1	11	8.33E-04	3.52E-04	53	6.52E-01	21	5.55E-01
	BC2	14	4.06E-04	5.95E-04	83.5	7.43E-01	32	1.86E-01
	BC3	10	3.86E-04	2.56E-04	67	4.26E-01	27	4.90E-02*
ovaries	**F1**	12	3.93E-05	4.13E-05	52	1.94E-02*	0	3.09E-06***,[a]
	BC1	13	5.73E-05	3.23E-05	54.5	3.40E-02*	10	2.43E-04***,[a]
	BC2	14	7.31E-05	2.65E-05	55	9.45E-02	0	1.75E-06***,[a]
	BC3	12	2.68E-05	2.44E-03	0	6.14E-03**,[a]	2	1.24E-05***,[a]

N = number of replicates analyzed, SD = standard deviation, W = Wilcoxon rank sum test statistic, p-value = probability.

*: p-value < 0.05,

**: p-value < 0.01,

***: p-value < 0.001,

[a]: p-values that are significant after Bonferroni correction (p-value<0.0125). Each kind of sample (males, females, testes, ovaries) has been compared to the same tissue of both parental species (see S1 File for N, median and SD values of parental species groups).

doi:10.1371/journal.pone.0147903.t002

in hybrid testes tends to be similar or lower than in parental species, with lower variances compared to *D. buzzatii*.

Ovaries, where TEs are strongly regulated [29], had the lowest *Helena* expression rates (average median between generations of 5×10^{-5}). *Helena* expression in *D. buzzatii* was significantly higher than in *D. koepferae*, and the vast majority of hybrid replicates had intermediate values (Fig 3D). Expression differences between *D. koepferae* and each hybrid generation are highly significant (Table 2): expression was higher in hybrids for almost all the replicates of every generation. Furthermore, differences in *Helena* expression rates between *D. buzzatii* and hybrids were also significant for F1, BC1 and BC3 (Table 2). *Helena* expression gradually increases across generations F1 to BC2, and then unexpectedly decreases in BC3. Yet, the highest ER (ER = 8.5×10^{-3}, in red in Fig 3D) belonged to a BC3 replicate, which might be due to a sporadic transcription burst. However, variance was unchanged in BC3 compared to any of the parental species (S2 Table), although Levene's test show that there were significant differences between generations (taking into account all samples: W = 4.05, p-value = 3.138E-03). In particular, *D. buzzatii* had a more variable *Helena* expression than hybrids, whereas *D. koepferae* had the lowest variance (in both cases, results are significant for F1, BC1 and BC2 – S2 Table). Hence, *Helena* expression in ovaries can be considered to be additive between parental species, but the hybrids had higher *Helena* ERs and variances than the maternal species (*D. koepferae*).

Tissue expression patterns in hybrids and parental species

To see whether the quantitative differences in *Helena* expression between hybrids and parents involved changes in patterns of expression in tissues, FISH was used on male and female gonads (see [30] and [31] for annotated schemes of these tissues).

Hybridized testes had *Helena* expression signals in both parental and hybrid germinal tissue (Fig 4), with different patterns between generations. In *D. buzzatii* (male parental species), as well as in *D. koepferae* (female parental species), *Helena* transcripts were specifically localized in the mitotic spermatogonia region (Fig 4A and 4B). Hybrid expression presented a high variability between generations, as well as differences among individuals from the same generation. For example, in F1, transcripts were mostly detected in the elongating spermatids region (Fig 4C); but some testes had additional expression in the primary spermatocytes (mitotic spermatogonia, as in the parents: Figs A-C in S2 File) or a generalized expression from the apical zone to the elongation area (Figs D and F in S2 File), and only in one case, no detectable expression (Fig E in S2 File). In BC1, no transcript signals were detected in most cases (S3 File), with a few exceptions that had signals at the end of the ejaculatory duct (Fig 4D and Figs B and D in S3 File), where individualized mature sperm can be found [30]. Expression in BC2 testes has also been detected in the basal zone near to the ejaculatory duct (Fig 4E and Figs B, D and E in S4 File), and also in the apical end (Figs A-D and F in S4 File), including in some cases the stem cell area (Figs A and B in S4 File). In BC3, two different patterns were seen (Fig 4F); transcripts were localized in primary spermatocytes (as in parents), or there were no evident hybridization signals.

In ovaries, *Helena* expression was also detected in all hybrid and parental samples (Fig 5). In both parental species, transcripts were specifically detected in the nurse cell nucleus (Fig 5A and 5B). In F1 hybrid females, there was widespread expression not only in the nucleus but also in the nurse cell cytoplasm; and interestingly, transcripts are also found in follicle cells, which are somatic cells of the germinal tissue (Fig 5C). In the three backcrosses (Fig 5D–5F), *Helena* expression was restricted to nurse cells, and only in BC1 was expression also seen in the cytoplasm of these cells (Fig 5D).

Comparative analysis of *Helena* piRNA populations in interspecific hybrids

For greater insight of *Helena* regulation by the piRNA pathway, we sequenced the gonadal small RNA populations of some of the samples analysed by qRT-PCR (in green, Fig 1). We analysed the alignment of 23–32 nucleotides reads (corresponding to piRNAs length) to all the *Helena* copies described in the *D. buzzatii* genome [32], with two main objectives: (i) quantifying the amount of *Helena*-specific piRNAs (Fig 6A), and (ii) detecting the ping-pong signature levels for each sample (Fig 6B).

In the testes, differences in *Helena*-specific piRNA abundance between *D. buzzatii* and F1 hybrids were striking; F1 hybrids had a 3.75 fold higher expression of *Helena* piRNAs than their parents (Fig 6A). However, the amounts of ping-pong signature detected in *Helena* piRNAs (Fig 6B) were similar in both samples. Consequently, we hypothesized that an activation of the primary piRNA biogenesis pathway—which acts independently of the ping-pong cycle—might be occurring in hybrid testes. This activation could be at the origin of the repression of the retrotransposon *Helena* found by qRT-PCR (Fig 3C).

We also analysed the piRNA populations of ovaries of both parental species, as well as F1 and BC1 hybrids. *Helena* piRNAs were slightly more abundant in *D. buzzatii* than in *D. koepferae* (×1.44, Fig 6A), and the ping-pong signal was also higher in *D. buzzatii* (10 nt-overlap probability of 45.2 vs. 36.4%, Fig 6B). In hybrid ovaries, *Helena*-specific piRNA amounts were

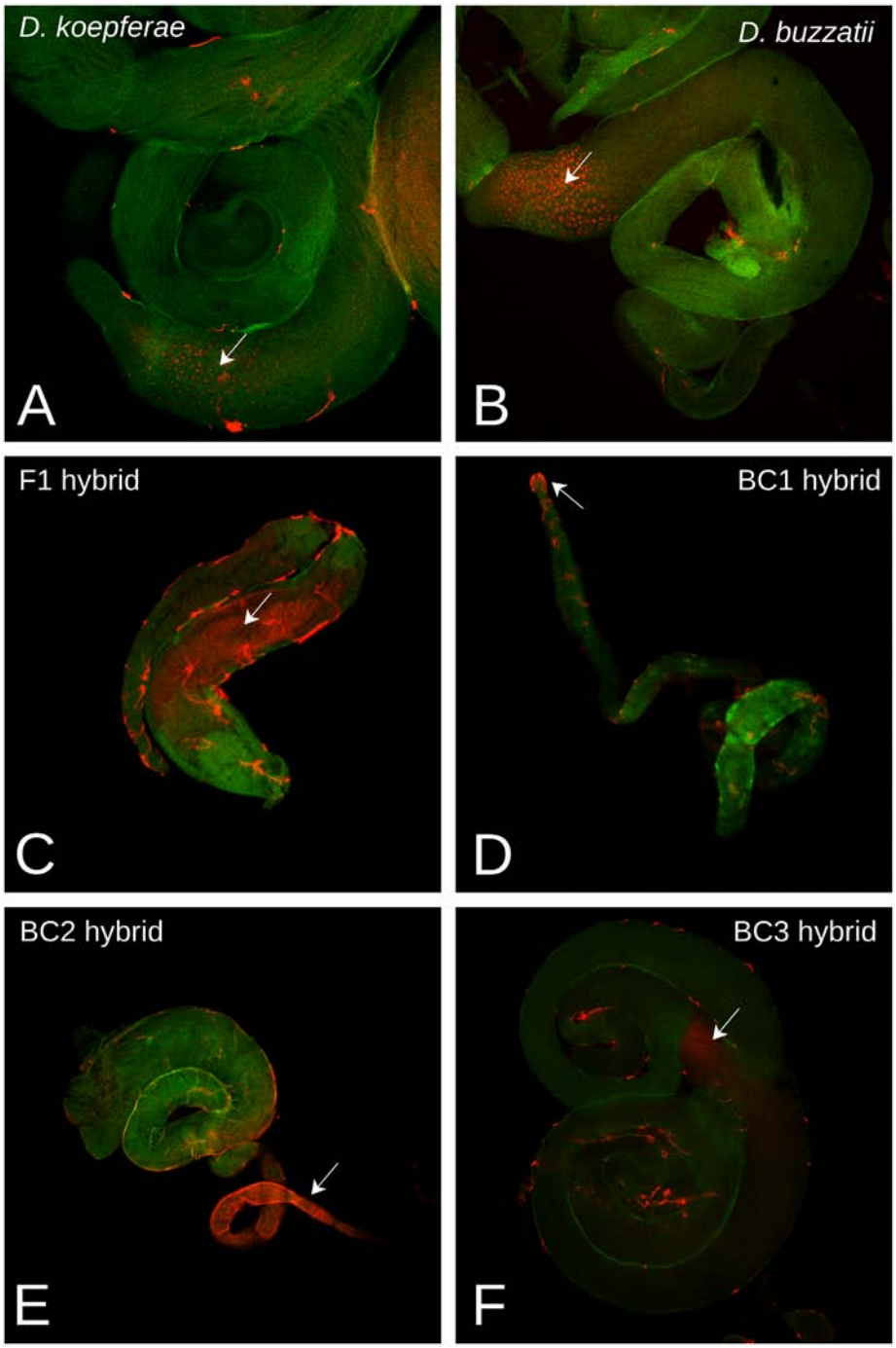


Fig 4. FISH of *Helena* RNA expression in testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts. **A)** *D. koepferae*, **B)** *D. buzzatii*, **C)** F1 hybrid, **D)** BC1 hybrid, **E)** BC2 hybrid, **F)** BC3 hybrid.

doi:10.1371/journal.pone.0147903.g004

at intermediate values between parental species (Fig 6A). F1 amounts were similar to *D. buzzatii*, but decreased after a generation of backcrossing. Curiously, F1 ovaries had a lower ping-pong signal than both parental species (×1.7–2.1, Fig 6B), while the BC1 signal was higher and very similar to *D. koepferae*. Thus, the results seem to indicate less efficient ping-pong cycle in

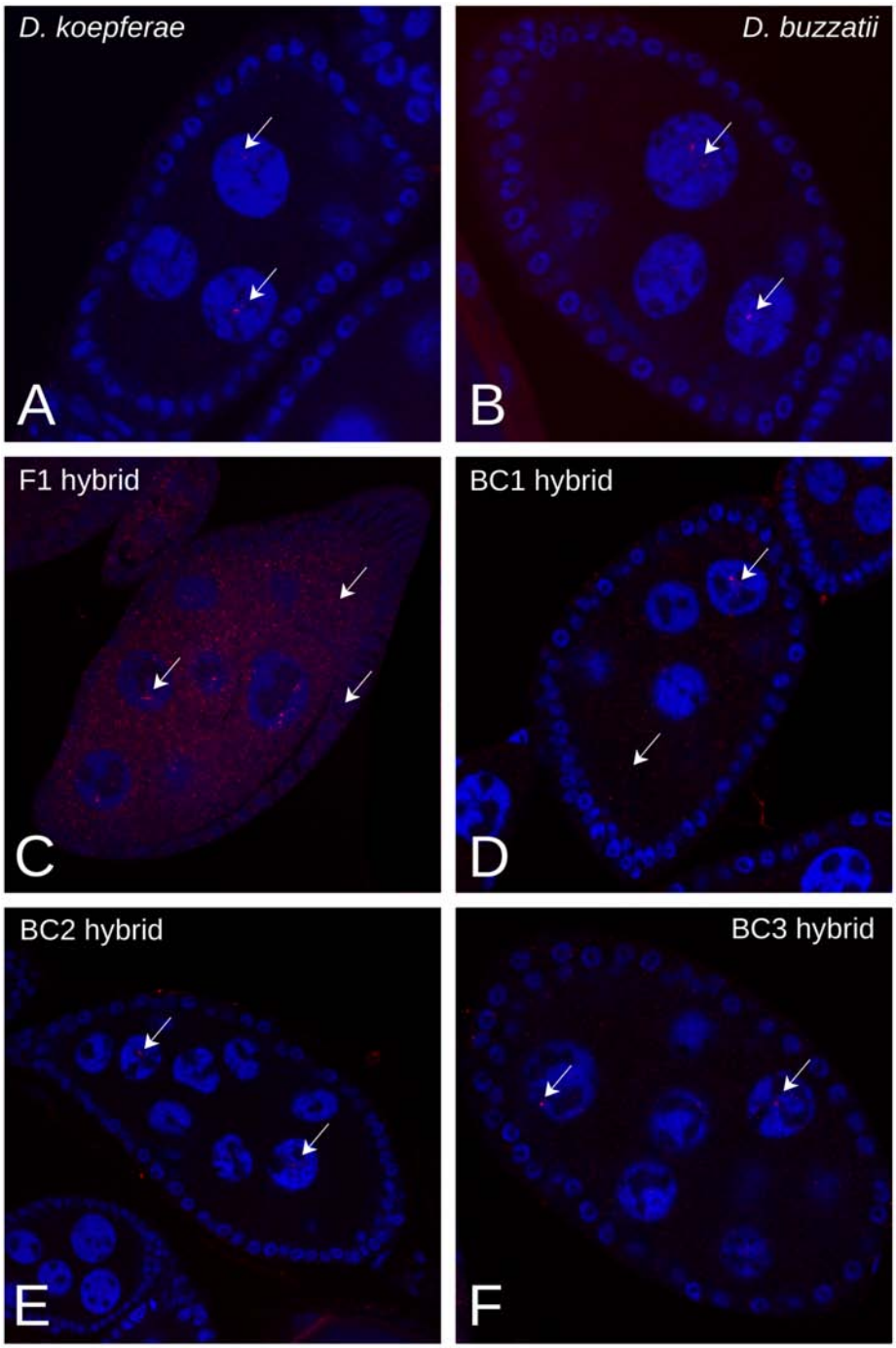


Fig 5. FISH of *Helena* RNA expression in ovaries. Red staining are *Helena* transcripts, blue staining is DAPI (cells nuclei). Arrows mark the presence of *Helena* transcripts. **A)** *D. koepferae*, **B)** *D. buzzatii*, **C)** F1 hybrid, **D)** BC1 hybrid, **E)** BC2 hybrid, **F)** BC3 hybrid.

doi:10.1371/journal.pone.0147903.g005

hybrids than in parental species. However, this hypothetical partial failure of the ping-pong amplification loop does not seem to alter substantially global piRNA production. As suggested for the testes, the primary pathway of piRNA production might also be responsible for maintaining the levels of *Helena* piRNAs.

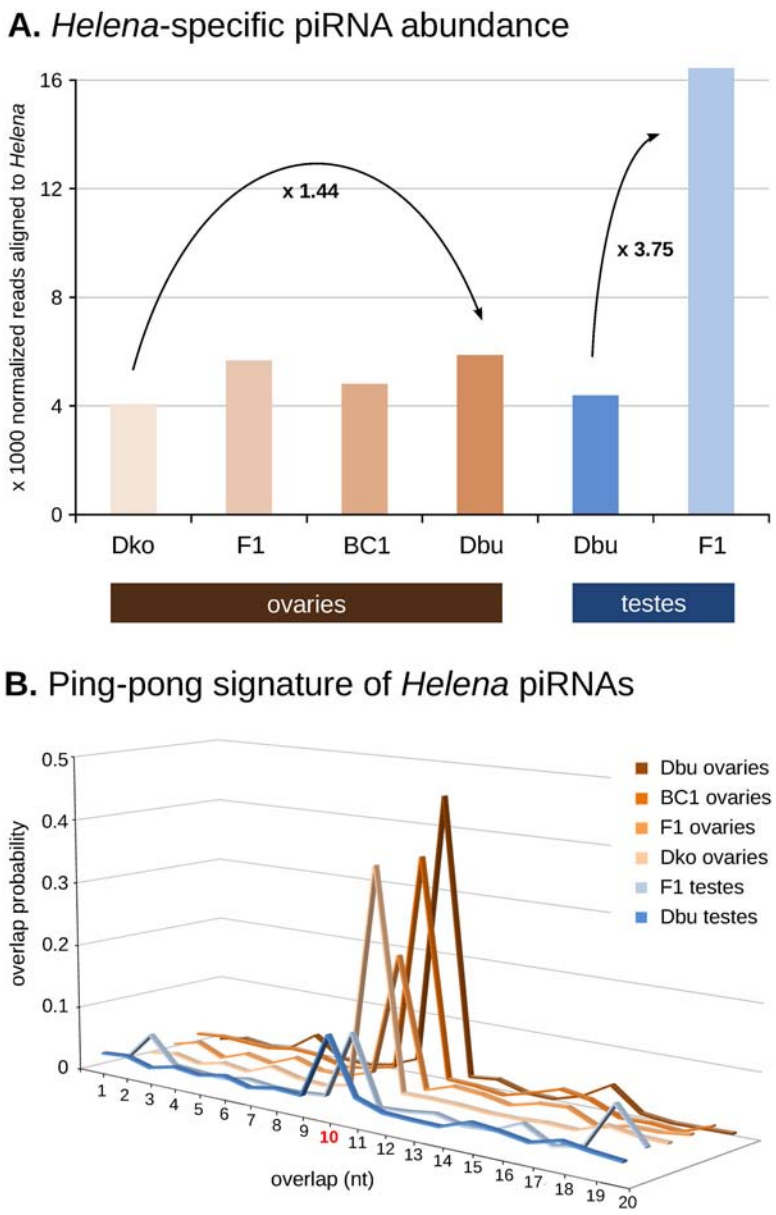


Fig 6. piRNA-mediated regulation of the retrotransposon *Helena*. A) Quantification of *Helena* piRNA populations: normalized read count of *Helena*-specific piRNAs in all sequenced samples, **B)** Ping-pong signature of *Helena*-specific piRNAs samples: probability of finding sense-antisense read pairs aligned to *Helena* sequences overlapping by 1 to 20 nucleotides; 10 nt overlap corresponds to ping-pong signal.

doi:10.1371/journal.pone.0147903.g006

Discussion

Helena is a well-conserved element in *D. buzzatii* and *D. koepferae* genomes

Our four *Helena* sequenced copies (*Dbu*, *Dko-28*, *Dko-35-1* and *Dko-35-2*) show high intraspecific and interspecific sequence identity levels (93–98%). This degree of conservation is remarkably high for a couple of species whose divergence time has been estimated at ~5 Myr [33], with a dS mode estimated in 0.85 (Romero-Soriano *et al.*, in prep). This could mean that

Helena is under purifiying selection in our model species, but the fact that LINE-like elements transposition mechanism produces non-functional 5'-truncated new insertions, which most probably display neutral evolution patterns [26], makes this hypothesis unlikely. Knowing that *D. mojavensis*, the closest sequenced species from *D. buzzatii* and *D. koepferae*, is one of the few species where a potentially active copy and expression of *Helena* has been detected [17], it is possible that our sequenced copies come from a recent invasion, escaping the genomic control, as in *D. mojavensis*. Internal deletions in non-LTR retrotransposons might act as a prevention mechanism against genome invasions by TEs [26,28]. In effect, the three *D. koepferae* sequenced copies present truncated *pol* ORFs, rendering them inefficient for transposition. Granzotto *et al.* [17] suggested that the case of *Helena* might offer a unique opportunity to real-time track a TE *life cycle* in *Drosophila*: its study in the few species where it remains active may lead us to understand the molecular mechanisms involved in TE neutralization, from internal deletions to expression repression.

Our phylogenetic analysis shows that the most closely related *Helena* sequences are, unexpectedly, *Dbu* and *Dko-28*, which might be explained by interspecific gene flow between *D. buzzatii* and *D. koepferae*, since they are sibling species sharing the same habitat in some arid areas of Argentina and Bolivia [34]. Although they are reproductively isolated by hybrid male sterility, introgression is possible through backcrosses of parental males with hybrid females. In fact, interspecific hybridization, eased in nature by sympatry, has been proposed between these species [33,35,36]. Another explanation could be horizontal transfer of *Helena* between these species, given that many examples of genetic horizontal transmission in eukaryotes involving transposable elements have been reported (reviewed in [37]).

Helena somatic expression remains globally unchanged after interspecific hybridization

Quantitative expression results show that the abundance of *Helena* transcripts in somatic tissues is not significantly different from hybrids and parental species (Table 2). However, a few replicates have extremely high ERs, especially in males (Fig 3), which might be due to exceptional transcription bursts. In F1, the presence of outranged values leads to a significant increase in variance (S2 Table), which could also be the consequence of experimental errors. However, we believe our careful controls and technical replicates were sufficiently accurate to rule out this hypothesis.

A small RNA-mediated silencing pathway, the endogenous siRNA (endo-siRNA) pathway, is responsible for TE silencing in *Drosophila* soma [19–22]. Failure of this post-transcriptional silencing system would result in a higher expression of *Helena* (and other TEs), as occasionally noted in our hybrid males. This punctual deregulation can be explained in two ways: (i) by the unique genetic background of each backcrossed hybrid, determined by the introgressed fragments of *D. buzzatii* genome. Different genetic backgrounds can result in significant differences in somatic transposition rates [38], which is also concordant with the different transposition rates found between hybrid individuals [11]. Furthermore, the endo-siRNA pathway components have quantitative [38] and overlapping [39] effects, which can lead to a wide range of consequences in case of partial failure. And (ii) because somatic transposition events are cell-specific and can take place at different stages of development [38], leading to different kinds of insertion mosaicisms. The earlier the silencing failure, the more *Helena* expression is expected both in F1 and backcrossed hybrids. So far, we ignore the ultimate cause of the occasional somatic deregulation occurring in a few samples, but it could be caused by either the divergence of the endo-siRNA pathway effector proteins between parental species or by the

absence/presence of *Helena*-specific endo-siRNAs in the genomic clusters responsible for their production.

Helena expression is repressed and mislocalized in F1 testes

In the male germline, *Helena* expression is first repressed in F1 hybrids, then restored to approximately the parental levels in subsequent generations (Fig 3C). These results are in contrast with those obtained for another mobilized retrotransposon, *Osvaldo*, in our hybrids in which significantly enhanced expression occurs in hybrid testes [12]. However, these differences are not rare because different classes of TEs can undergo differential regulation [24,40]. In the case of *Helena*, regulation studies provided us with a compelling molecular explanation for *Helena* low transcript abundances in F1; indeed, F1 testes have almost four times more *Helena*-specific piRNAs than the parental species, *D. buzzatii*, which might make its silencing more efficient (Fig 6A). It is noteworthy that the ping-pong cycle signature is maintained between F1 hybrids and *D. buzzatii* (Fig 6B), showing that the greater abundance of *Helena*-specific piRNA populations is not due to increased efficiency of secondary piRNA biogenesis. Therefore, we can hypothesize that the primary piRNA biogenesis pathway is enhanced in F1 interspecific hybrid testes in order to counterpart a putative TE activation. This was proposed in a recent study on wheat [41], where TE repression mechanisms were activated in F1 hybrids.

On the other hand, changes have also been detected in *Helena* cellular expression patterns after interspecific hybridization. In parental species, *Helena* transcripts have been detected in mitotic spermatogonia, a stage characterized by a general high level of gene expression [42], where other TEs such as *copia* [43] and *412* [44] are expressed. In hybrids, however, we have detected *Helena* expression in later stages of spermatogenesis, including elongating spermatids (in F1) and mature sperm (in BC1 and BC2) cells, which are considered far less active transcriptionally. These results are in agreement with previous studies in hybrids of the same species, where *Osvaldo* expression was also found in the basal region of the testes [12]. This could suggest that transcriptional misregulation of *Helena* and other TEs occurs in hybrids, a phenomenon that has been described for some genes, and linked to hybrid sterility and other hybrid incompatibilities [45]. Concordantly, fertility recovery in hybrid males of *D. buzzatii* and *D. koepferae* takes place in some individuals from BC3 [46], where tissue expression patterns are very similar to the parents. Thus, incorrect localization of *Helena* expression might be involved in hybrid male sterility, since we know that sterility in our hybrids is caused by the additive effect of several minor loci [46]. However, our FISH results can only be interpreted qualitatively; *Helena* is mislocalized in hybrid testes, but its expression at a quantitative level decreases (in F1) or is maintained (BC1, BC2 and BC3) after interspecific hybridization (Fig 3C).

Ovaries have additive values of *Helena* expression after interspecific hybridization

Ovaries are the only tissue where parental species expression differs significantly. *D. buzzatii* has higher *Helena* transcript levels than *D. koepferae* (Fig D in S1 File). The *Helena* copy number detected by Southern blot (S1 Fig) is higher in *D. buzzatii* (12–15) than in *D. koepferae* (6–12), but only 5 of the *D. buzzatii* copies are localized in the chromosome arms [11], less than in *D. koepferae* (12 copies). However, differences between their ERs are >10-fold (Fig 3D), which can only be explained if many *D. koepferae* copies are inactive. This hypothesis is in concordance with our *Helena* sequencing data, where the three sequenced *Helena* copies have truncated *pol* ORFs (Table 1).

In hybrid ovaries, *Helena* expression is at intermediate levels between *D. koepferae* and *D. buzzatii* for all generations (Fig 3D), but the amounts are always significantly higher than in *D. koepferae*. Since TE silencing in ovaries is crucial to maintain the genome integrity [29], organisms develop different strategies to efficiently control TE invasions. Thus, *Helena* ongoing neutralization might have followed different ways and reached different stages between our parental species. To explain *Helena* expression values, we focused on its regulation by piRNAs, which is the most important TE silencing mechanism in *Drosophila*.

We found that *D. buzzatii* has a larger *Helena*-specific population of piRNAs, whose ping-pong signature is higher than in *D. koepferae* (Fig 6). In F1 and BC1, the *Helena*-specific piRNA amounts are intermediate between those in the parental species (Fig 6A). However, secondary piRNA biogenesis seems to be less efficient in hybrid ovaries than in the parental species (Fig 6B), especially in F1, where there is a lower ping-pong signal in comparison to both *D. buzzatii* and *D. koepferae*. This reduced efficiency of ping-pong amplification may be due to a certain hybrid incompatibility in this pathway; indeed, even if our parental species are closely related, piRNA-mediated silencing effectors seem to be codified by rapidly evolving genes with positive selection marks [47], and whose expression varies widely between different populations of the same *Drosophila* species [48]. As suggested by results from testes, this malfunction might be compensated by the activation of the primary biogenesis pathway in order to maintain piRNA levels and preserve germline integrity.

At a cellular level, expression has been detected in nurse cells in all samples, but some generations of hybrid females (F1 and BC1) have a more widespread pattern of expression that also affects F1 follicular cells. Absence of a TE-specific piRNA in the mother cytoplasm can cause a transcriptional burst of this transposon in germ cells because maternally inherited piRNAs are responsible for TE silencing initiation at the first stages of development [49]. However, this putative increase of *Helena* expression in F1 hybrids compared with both parental species is not evident in our qRT-PCR results, which could be explained by the age of females, i.e. 3-day old for FISH and 10 for qRT-PCR. *Helena* expression rates might vary within a fly's lifetime, as noted in P-M dysgenic crosses where fertility recovery in old females has been attributed to the regulation of P elements by paternally inherited piRNA clusters [50]. In our case, although *D. koepferae* cytoplasm contains *Helena*-specific piRNAs, their levels are lower than in *D. buzzatii* (Fig 6A). Thus, the maternal cytoplasm could indeed be less efficient in silencing *Helena* expression and might cause the extensive presence of *Helena* transcripts noted in F1 and BC1 ovaries of young females (Fig 5) [51]. Interestingly, there was an atypical *Helena* expression pattern including ovary follicular cells in F1 hybrids. A similar phenomenon has been described in *D. simulans*, where the failure of the maternal cytotype to repress the transposon *tirant* also involved its unusual expression in follicular cells [52], which could explain the presence of *Helena* transcripts in F1 follicle cells (Fig 5C). On the other hand, the lower efficiency of the ping-pong cycle in F1 hybrid ovaries (Fig 6B) could also be at the origin of widespread *Helena* expression. However, it is important to emphasise that FISH results are only qualitative, and that the generalized localization of *Helena* transcripts might not be linked to a real increase of expression.

We propose the following landscape in ovaries: a first stage of *Helena* enhanced expression would occur in young flies, because the ping-pong cycle seems to be less productive (especially in F1) than in parental species (Fig 6B), and also probably because *D. koepferae* cytoplasm is unable to efficiently silence *Helena* expression. Eventually, new *Helena* transposition events could take place at this time. This derepression step would be followed by the activation of other TE regulation mechanisms, such as primary piRNA biogenesis. This would allow *Helena*-specific piRNA levels to be maintained and would consequently decrease *Helena*

expression. Therefore, our FISH experiments may be detecting transcripts that will later be post-transcriptionally silenced.

In conclusion, we have shown that interspecific hybridization modifies the expression of *Helena* retrotransposon in gonads. However, we demonstrate that *Helena* can be either transcriptionally repressed (as in F1 testes) or enhanced (as in F1 and BC1 young ovaries) in hybrids. Therefore, our study underlines the complexity of TE deregulation in hybrids, which not only differs between sexes but also presents different patterns between transposons [12]. The molecular understanding of the intricate mechanisms involved in TE silencing in hybrids might be crucial to cast light on the evolutionary role of TEs in phenomena such as hybrid sterility and speciation.

Material and Methods

Drosophila stocks and crosses

The strains used for interspecific crosses were (i) the Bu28 strain of *D. buzzatii*, an inbred line originated by the union of different populations (LN13, 19, 31 and 33) collected in 1982 in Los Negros (Bolivia); and (ii) the Ko2 strain of *D. koepferae*, an inbred line originated from a population collected in 1979 in San Luis (Argentina). Both lines were maintained by brother-sister mating for over a decade and are now kept by mass culturing.

Because hybrid egg viability between *D. buzzatii* and *D. koepferae* is low [53] and the hybrid offspring scarce, we performed 14 different crosses for qPCR experiments, denoted as families A through N. For each family, 10–15 Ko2 virgin females were crossed with 10–15 Bu28 males of the same age. This cross was followed by three generations of backcrossing of 10–15 hybrid females (whenever possible) with the same number of *D. buzzatii* males. Five additional crosses (as just described) were used for FISH analyses. All stocks and crosses were reared at 25°C in a standard *Drosophila* medium supplemented with yeast.

Helena molecular characterization

PCR reactions were carried out in a final volume of 50 μl, including 1× High Yield Reaction Buffer with Mg^{2+} (Kapa Biosystems), 0.2 mM of each dNTP (Roche), 0.4 μM of each primer (Sigma-Aldrich), template DNA (≈10–20 ng) and 0.04 U/μl of Taq polymerase (KapaTaq from Kapa Biosystems). A MJ Research Inc. thermocycler was used, with the following program: 5 min at 94°C (preliminary denaturation); 30 cycles of 45 s at 94°C (denaturation), 45 s at specific PCR annealing temperatures and 1 min 30 s at 72°C (extension); and 10 min at 72°C (final extension). The two longest copies of *Helena* from *D. koepferae* were amplified using Roche's Expand Long Template PCR system. Amplified samples were stored at 4°C, gel purified with the Nucleospin Gel and PCR Clean-Up kit (Macherey-Nagel), and cloned with the pGEM-T Easy Vector System I (Promega).

Primers were designed from the longest copy of *Helena* from the *D. mojavensis* genome [17], the closest sequenced species to *D. buzzatii* and *D. koepferae*: HelMojF2A (5′ -AGCAGCCCAGAAAATGCTTA-3′) and HelMojR2B (5′ -TCTCAGCGGTAAGGTGCTCT-3′). For the *Helena* shortest copy from *D. koepferae*, HelMojR1A (5′ -GTCCACAACCACAACCACAG-3′) was used instead of HelMojR2B.

Helena isolated clones were sequenced using capillary sequencing technique (Macrogen Inc); they were analyzed using different NCBI tools and databases, such as nucleotide BLAST [54] (megablast algorithm, for highly similar sequences), ORF Finder and Conserved Domain Search [25]. Sequence data from this article have been deposited in GenBank repository (accession numbers KF280391, KP115213, KP115214 and KP115215).

For the phylogenetic analysis, we used some of the *Helena* consensus sequences identified in the 12 *Drosophila* genomes: from *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. mojavensis* [17] and *D. simulans* [18], as well as other *Helena* characterized sequences from *D. virilis* (Repbase ID = HELENA) [16] and *D. melanogaster* (accession number AF012030) [28]. The retrieved sequences, together with *D. buzzatii* and *D. koepferae* obtained copies, were aligned using MAFFT 7.123b E-INS-i algorithm, optimized for sequences with multiple conserved domains and long gaps [55]. The alignment was automatically cleaned using Gblocks [56] web server, allowing smaller final blocks, gap positions within the final blocks and less strict flanking positions (S1 Text for the fasta alignment). A graphical representation of the final alignment using ESPript [57] (http://espript.ibcp.fr) is also available (S2 Text). Maximum likelihood (ML) phylogenetic trees were estimated by RaxML 7.2.8 [58], using the GTRCAT model. Statistical support for bipartitions was estimated from 100-bootstrap replicates with RaxML (same model).

Southern blot

Genomic DNA from *D. buzzatii* and *D. koepferae* (Bu28 and Ko2 strains, respectively) was extracted from a pool of 30 flies according to [59]. The DNA was digested with *AatII* (Roche), which has no restriction sites within the *Helena* sequence, allowing us to estimate the number of complete copies. After agarose gel electrophoresis and denaturing steps, the DNA was transferred to a positively charged nylon membrane (Roche). Pre-hybridization and hybridization steps were carried out at 42°C in a solution containing 5x SSC, 50% of formamide, 0.1% of *N*-laurylsarcosine, 0.02% of SDS and 5% of blocking reagent (Roche). The membrane was hybridized with a dig-labelled DNA probe of ≈3.8 kb, corresponding to the longest sequenced fragment of *Helena* in *D. buzzatii*.

Quantification of *Helena* transcripts by RT-PCR

Four types of samples were analyzed for each generation: ovaries, testes, female somatic carcasses and male somatic carcasses. Flies were dissected in PBT (1× phosphate-buffered saline [PBS], 0.2% Tween 20) 10 days after their birth, to ensure a sufficient number of offspring for analysis. Total RNA was purified from >10 ovaries, 14 testes or 10 carcasses per sample with the RNeasy kit (Qiagen) and then treated with DNaseI (Ambion). cDNA synthesis was carried out with anchored-oligo$(dT)_{18}$ primers using Roche's Transcriptor First Strand cDNA Synthesis Kit. Transcript abundance was estimated by fluorescence intensity using Biorad's iQ SYBR Green Supermix on a CFX96 BioRad Real-Time lightcycler with primers specific to the *Helena* endonuclease region. Relative quantification was performed using the ribosomal *rp49* housekeeping gene, which is equally expressed in *D.buzzatii* and *D. koepferae*, as an endogenous control for the standard curve method. Two technical replicates were run for each sample. ΔC_T values for all samples are summarized in S3 Text and have been used to calculate expression rates as in [27].

Primers used to amplify *Helena* were designed from a fragment of *Helena* characterized earlier from the same hybrids [11]. The qPCR fragment corresponds to 200 bp of the endonuclease region amplified with the following primers: HelenaF1 (5′ -CGACATACTCGCTTCCTGTG-3′) and HelenaR1 (5′ -TCACACTCCCTCTTGCATTG-3′). For *rp49*, the published primers [60] designed from *D. mojavensis* genome were used, that give a qPCR amplicon of 196 bp. The primer efficiencies were 96.6 and 99% for *Helena* and *rp49* respectively.

Fluorescent *in situ* hybridization in ovaries and testes

We dissected the ovaries and testes of 3-days old flies in PBT, which is the ideal age for optimal visualization of the different cells from ovaries. We followed the protocol described in [61]. The *Helena* antisense RNA probe was a 984-pb fragment corresponding to the *pol-like* gene

(primers HelenaF1 and HelMojR1A), which included T7 and SP6 promoter sites. It was labelled by *in vitro* transcription of SP6/T7 using DIG RNA Labelling Kit (Roche). Labelled probes were detected using anti-DIG POD antibody (Roche) and fluorescence amplification (TSA PLUS Cyanine3 kit, PerkinElmer), visualized with a TCS-SP5 Leica confocal scanning laser microscope.

piRNA analyses: small RNA extraction, library preparation, sequencing and alignment

We dissected 5 to 6-days old flies as described above. Small RNA was purified from ovaries (n = 70 pairs for all samples) and testes (n = 96 pairs for *D. buzzatii* and n = 333 pairs for F1 sterile males), following the manual small RNA purifying protocol described of Grentzinger *et al.* [62]. After small RNA isolation, samples were gel-purified and precipitated. A single Illumina library was prepared for each sample and sequenced on an Illumina Hiseq 2500 platform by FASTERIS SA (Switzerland). Reads of 23–32 nucleotides were selected as piRNAs and trimmed using UrQt [63] to remove low-quality nucleotides. The trimmed reads were aligned to the *D. buzzatii* genome TE library [32] using Bowtie1 v1.1.1 [64] (the most sensitive option and keeping a single alignment for reads mapping to multiple positions). The read count step (built in TE tools: https://github.com/l-modolo/TEtools) was computed per TE family by adding all reads mapped on copies from the same family. Finally, read counts were normalized using the R Bioconductor package DESeq2 [65]. Only the results for *Helena* retrotransposon were used for this study.

Ping-pong signature was analyzed by checking the presence of sense-antisense read pairs overlapping by 10 nucleotides, using Antoniewski's signature.py pipeline [66]. For this analysis, we used the raw 23–32 nucleotide reads since a trimming step would bias the real small RNA length aligned to the *Helena* sequences of the same TE library (as described above).

Statistical methods

R software was used for statistical analyses. Because the assumptions of Gaussian distribution and equal variances are not valid in qRT-PCR experiments with small sample sizes, the most suitable test is the robust non-parametric Wilcoxon rank sum test (also called the Mann-Whitney test [67]), which was used to compare expression rates of hybrids and parental species at each generation. Kruskal-Wallis test [68] was used to determine whether differences between all groups (including all parents and hybrids) were significant. Finally, Levene's test for equality of variances was used to assess changes in variance between groups.

Supporting Information

S1 Fig. Southern blot analysis of *Helena* in parental species, *D. buzzatii* (left) and *D. koepferae* (right). No restriction sites for *AatII* are present in *Helena*'s probe sequence. Thus, digestions with this enzyme allow us to distinguish different *Helena* copies. Arrows in red indicate strong-signaled bands; arrows in black indicate faint bands.
(TIFF)

S1 File. *Helena* expression results in parental species. (Fig A and B) *Helena* expression rates relative to *rp49* housekeeping gene in *D. koepferae* (Dko) and *D. buzzatii* (Dbu) somatic tissues **(A)** and gonads **(B)**. Male samples are represented in blue and female samples are represented in brown. Boxes are determined by the first and third quartile values, with an intermediate deep line corresponding to the median value. Circles correspond to outliers (above or below 1.5-fold the interquartile range), and triangles represent those outliers whose ERs are extremely outranged and cannot be represented in the same scale (triangle in A: ER = 2.9×10^{-3}, in B: ER = 3.6×10^{-3}

and 6.2×10^{-3}). **(Fig C and D)** Comparison of *Helena* expression rates between all different parental samples for somatic tissues **(C)** and gonads **(D)**. N = number of replicates analyzed, SD = standard deviation, W = Wilcoxon rank sum test statistic, p-value = probability. *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001. In red, p-values that are significant after Bonferoni correction (p-value<0.008).
(PDF)

S2 File. FISH of *Helena* RNA expression in different F1 hybrid testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts.
(TIFF)

S3 File. FISH of *Helena* RNA expression in different BC1 hybrid testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts.
(TIFF)

S4 File. FISH of *Helena* RNA expression in different BC2 hybrid testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts.
(TIFF)

S1 Table. Summary of BLAST alignment results between *Helena* sequenced copies. *Dbu* = *D. buzzatii*, *Dko28* = *D. koepferae*-28, *Dko35-1* = *D. koepferae*-35-1, *Dko35-2* = *D. koepferae*-35-2.
(PDF)

S2 Table. Variance comparisons of *Helena* expression rates between each hybrid generation and parental species. W = Levene's test for equality of variances satistic, p-value = probability. *: p-value<0.05, **: p-value<0.01, ***: p-value<0.001. In red, p-values that are significant after Bonferoni correction (p-value<0.01). Each kind of sample (males, females, testes, ovaries) has been compared to the same tissue of both parental species.
(PDF)

S1 Text. Alignment of *Helena* sequences (in fasta format) obtained with MAFFT E-INS-i algorithm and cleaned using Gblocks. This alignment was used to construct the phylogenetic tree on Fig 2.
(PDF)

S2 Text. Graphical representation of the *Helena* alignment obtained with MAFFT E-INS-i algorithm and cleaned using Gblocks. Highly conserved residues (similarity score per position > 0.5) are framed in blue and used to build the consensus sequence. Each nitrogenous base in a conserved position is represented in a different colour.
(PDF)

S3 Text. Summary of ΔC_T values for all studied replicates (from different crosses) of each kind of sample for all generations.
(PDF)

Acknowledgments

We thank M. Peiró for her technical assistance; X. Grau-Bové for his valuable help with phylogenetic methods; N. Rius and A. Ruiz for providing an advanced access to the *D. buzzatii* genome and their TE copies list; and B. Mugat and S. Chambeyron for the small RNA extraction.

Author Contributions

Conceived and designed the experiments: MPGG. Performed the experiments: VRS. Analyzed the data: VRS. Wrote the paper: VRS MPGG.

References

1. Fontdevila A (2005) Hybrid genome evolution by transposition. Cytogenet Genome Res 110: 49–55. PMID: 16093657
2. Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20: 229–237. PMID: 16701374
3. O'Neill RJW, O'Neill MJ, Marshall Graves JA (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393: 68–73. PMID: 9590690
4. Metcalfe CJ, Bulazel KV, Ferreri GC, Schroeder-Reiter E, Wanner G, Rens W, et al. (2007) Genomic instability within centromeres of interspecific marsupial hybrids. Genetics 177: 2507–2517. PMID: 18073443
5. Liu B, Wendel JF (2000) Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43: 874–880. PMID: 11081978
6. Ungerer MC, Strakosh SC, Zhen Y (2006) Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16: R872–R873. PMID: 17055967
7. Wang N, Wang H, Wang H, Zhang D, Wu Y, Ou X, et al. (2010) Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with Zizania latifolia. BMC Plant Biol 10: 190. doi: 10.1186/1471-2229-10-190 PMID: 20796287
8. Evgen'ev MB, Yenikolopov GN, Peunova NI, Ilyin YV (1982) Transposition of Mobile Genetic Elements in Interspecific Hybrids of Drosophila. Chromosoma 85: 375–386. PMID: 6126320
9. Labrador M, Farré M, Utzet F, Fontdevila A (1999) Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol 16: 931–937. PMID: 10406110
10. Vela D, García Guerreiro MP, Fontdevila A (2011) Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. Biotechniques 50: 247–250. doi: 10.2144/000113655 PMID: 21548908
11. Vela D, Fontdevila A, Vieira C, García Guerreiro MP (2014) A genome-wide survey of genetic instability by transposition in Drosophila hybrids. PLoS One 9: e88992. doi: 10.1371/journal.pone.0088992 PMID: 24586475
12. García Guerreiro MP (2015) Changes of Osvaldo expression patterns in germline of male hybrids between the species Drosophila buzzatii and Drosophila koepferae. Mol Genet Genomics 290: 1471–1483. doi: 10.1007/s00438-015-1012-z PMID: 25711309
13. Kelleher ES, Edelman NB, Barbash DA (2012) Drosophila Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. PLoS Biol 10: e1001428. doi: 10.1371/journal.pbio.1001428 PMID: 23189033
14. Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5: 103–107. PMID: 2543105
15. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 10: 276.
16. Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis. Proc Natl Acad Sci U S A.
17. Granzotto A, Lopes FR, Lerat E, Vieira C, Carareto CMA (2009) The evolutionary dynamics of the Helena retrotransposon revealed by sequenced Drosophila genomes. BMC Evol Biol 9: 174. doi: 10.1186/1471-2148-9-174 PMID: 19624823
18. Rebollo R, Lerat E, Kleine LL, Biémont C, Vieira C (2008) Losing helena: the extinction of a drosophila line-like element. BMC Genomics 9: 149. doi: 10.1186/1471-2164-9-149 PMID: 18377637
19. Siomi MC, Saito K, Siomi H (2008) How selfish retrotransposons are silenced in Drosophila germline and somatic cells. FEBS Lett 582: 2473–2478. doi: 10.1016/j.febslet.2008.06.018 PMID: 18572018
20. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, et al. (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science (80-) 320: 1077–1081.
21. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, et al. (2008) An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798–802. doi: 10.1038/nature07007 PMID: 18463631

22. Okamura K, Chung W-J, Ruby JG, Guo H, Bartel DP, Lai EC (2008) The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803–806. doi: 10.1038/nature07015 PMID: 18463630

23. Rozhkov NV, Hammell M, Hannon GJ (2013) Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev.

24. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Kellis M, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128: 1089–1103. PMID: 17346786

25. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 41: D348–D352. doi: 10.1093/nar/gks1243 PMID: 23197659

26. Petrov DA, Lozovskaya ER, Hartl DL (1996) High intrinsic rate of DNA loss in Drosophila. Nature 384: 346–349. PMID: 8934517

27. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3: 1101–1108. PMID: 18546601

28. Petrov DA, Hartl DL (1998) High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. Mol Biol Evol 15: 293–302. PMID: 9501496

29. O'Donnell KA, Boeke JD (2007) Mighty Piwis defend the germline against genome intruders. Cell 129: 37–44. PMID: 17418784

30. Hirst J, Carmichael J (2011) A potential role for the clathrin adaptor GGA in Drosophila spermatogenesis. BMC Cell Biol 12: 22. doi: 10.1186/1471-2121-12-22 PMID: 21599933

31. Olovnikov IA, Kalmykova AI (2013) piRNA Clusters as a Main Source of Small RNAs in the Animal Germline. Biochem 78: 572–584.

32. Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, et al. (2015) Genomics of ecological adaptation in cactophilic Drosophila. Genome Biol Evol 7: 349–366.

33. Gomez GA, Hasson E (2003) Transpecific Polymorphisms in an Inversion Linked Esterase Locus in Drosophila buzzatii. Mol Biol Evol 20: 410–423. PMID: 12644562

34. Marín I, Fontdevila A (1998) Stable Drosophila buzzatii—Drosophila koepferae Hybrids. J Hered 89: 336–339. PMID: 9703688

35. Piccinali R, Aguadé M, Hasson E (2004) Comparative molecular population genetics of the Xdh locus in the cactophilic sibling species Drosophila buzzatii and D. koepferae. Mol Biol Evol 21: 141–152. PMID: 14595098

36. Franco FF, Silva-Bernardi ECC, Sene FM, Hasson ER, Manfrin MH (2010) Intra- and interspecific divergence in the nuclear sequences of the clock gene period in species of the Drosophila buzzatii cluster. J Zool Syst Evol Res 48: 322–331.

37. Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25: 537–546. doi: 10.1016/j.tree.2010.06.001 PMID: 20591532

38. Xie W, Donohue RC, Birchler JA (2013) Quantitatively increased somatic transposition of transposable elements in Drosophila strains compromised for RNAi. PLoS One 8: e72163. doi: 10.1371/journal.pone.0072163 PMID: 23940807

39. Mirkovic-Hösle M, Förstemann K (2014) Transposon defense by endo-siRNAs, piRNAs and somatic piRNAs in Drosophila: contributions of Loqs-PD and R2D2. PLoS One 9: e84994. doi: 10.1371/journal.pone.0084994 PMID: 24454776

40. Sienski G, Dönertas D, Brennecke J (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151: 964–980. doi: 10.1016/j.cell.2012.10.040 PMID: 23159368

41. Senerchia N, Parisod C, Parisod C (2015) Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proc R Soc B Biol Sci 282: 20142874.

42. White-Cooper H (2012) Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. Spermatogenesis 2: 11–22. PMID: 22553486

43. Filatov DA, Morozova TV, Pasyukova EG (1998) Age dependence of the copia transposition rate is positively associated with copia transcript abundance in a Drosophila melanogaster isogenic line. Mol Gen Genet MGG 258: 646–654. PMID: 9671033

44. Borie N, Maisonhaute C, Sarrazin S, Loevenbruck C, Biémont C (2002) Tissue-specificity of 412 retrotransposon expression in Drosophila simulans and D. melanogaster. Heredity (Edinb) 89: 247–252.

45. Moehring AJ, Teeter KC, Noor MAF (2007) Genome-wide patterns of expression in Drosophila pure species and hybrid males. II. Examination of multiple-species hybridizations, platforms, and life cycle stages. Mol Biol Evol 24: 137–145. PMID: 17032727

46. Morán T, Fontdevila A (2014) Genome-Wide Dissection of Hybrid Sterility in Drosophila Confirms a Polygenic Threshold Architecture. J Hered 105: 381–396. doi: 10.1093/jhered/esu003 PMID: 24489077
47. Simkin A, Wong A, Poh Y-P, Theurkauf WE, Jensen JD (2013) Recurrent and Recent Selective Sweeps in the piRNA Pathway. Evolution (N Y) 67: 1081–1090.
48. Fablet M, Akkouche A, Braman V, Vieira C (2014) Variable expression levels detected in the Drosophila effectors of piRNA biogenesis. Gene 537: 149–153. doi: 10.1016/j.gene.2013.11.095 PMID: 24361206
49. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ (2008) An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. Science (80-) 322: 1387–1392.
50. Levine MT, Malik HS (2011) Learning to Protect Your Genome on the Fly. Cell 147: 1440–1441. doi: 10.1016/j.cell.2011.12.001 PMID: 22196722
51. Théron E, Dennis C, Brasset E, Vaury C (2014) Distinct features of the piRNA pathway in somatic and germ cells : from piRNA cluster transcription to piRNA processing and amplification. Mob DNA 5: 1–11.
52. Akkouche A, Grentzinger T, Fablet M, Armenise C, Burlet N, Braman V, et al. (2013) Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. EMBO Rep 14: 1–7.
53. Soto EM, Soto IM, Carreira VP, Fanara JJ, Hasson E (2008) Host-related life history traits in interspecific hybrids of cactophilic Drosophila. Entomol Exp Appl: 18–27.
54. McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32: W20–W25. PMID: 15215342
55. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772–780. doi: 10.1093/molbev/mst010 PMID: 23329690
56. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56: 564–577. PMID: 17654362
57. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42: W320–W324. doi: 10.1093/nar/gku316 PMID: 24753421
58. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690. PMID: 16928733
59. Piñol J, Francino O, Fontdevila A, Cabré O (1988) Rapid isolation of Drosophila high molecular weight DNA to obtain genomic libraries. Nucleic Acids Res 16: 2763.
60. Granzotto A, Lopes FR, Vieira C, Carareto CMA (2011) Vertical inheritance and bursts of transposition have shaped the evolution of the BS non-LTR retrotransposon in Drosophila. Mol Genet Genomics 286: 57–66. doi: 10.1007/s00438-011-0629-9 PMID: 21618036
61. Akkouche A, Rebollo R, Burlet N, Esnault C, Martinez S, Viginier B, et al. (2012) tirant, a newly discovered active endogenous retrovirus in Drosophila simulans. J Virol 86: 3675–3681. doi: 10.1128/JVI.07146-11 PMID: 22278247
62. Grentzinger T, Armenise C, Pelisson A, Brun C, Mugat B, Serrano V, et al. (2013) A user-friendly chromatographic method to purify small regulatory RNAs. Methods 67: 91–101. doi: 10.1016/j.ymeth.2013.05.011 PMID: 23727218
63. Modolo L, Lerat E (2015) UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. BMC Bioinformatics 16: 137. doi: 10.1186/s12859-015-0546-8 PMID: 25924884
64. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25. doi: 10.1186/gb-2009-10-3-r25 PMID: 19261174
65. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 1–21.
66. Antoniewski C (2014) Computing siRNA and piRNA Overlap Signature. In: Werner A, editor. Animal Endo-SiRNAs: Methods and Protocols. Springer New York. pp. 135–146.
67. Mann HB, Whitney DR (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann Math Stat 18: 50–60.
68. Kruskal WH, Wallis WA (1952) Use of Ranks in One-Criterion Variance Analysis. J Am Stat Assoc 47: 583–621.

3.3 Transcriptomic analysis of TE deregulation in *D. buzzatii-D. koepferae* hybrids

This chapter consists of the manuscript entitled “Divergence in piRNA pathway effector proteins partially explains *Drosophila buzzatii-D. koepferae* hybrid instability” that has been recently submitted and is currently under review. Supplementary material of this article can be found in **Annex 8.3**.

3.3.1 Divergence in piRNA pathway effector proteins partially explains *Drosophila buzzatii-D. koepferae* hybrid instability

Authors: Valèria Romero-Soriano[1], Laurent Modolo[2], Hélène Lopez-Maestre[2], Bruno Mugat[3], Eugénie Pessia[2], Séverine Chambeyron[3], Cristina Vieira[2], Maria Pilar Garcia Guerreiro[1].

[1] Grup de Biologia Evolutiva. Departament de Genètica i Microbiologia (Edifici C). Universitat Autònoma de Barcelona. 08193 Bellaterra, Barcelona, Spain.

[2] Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Université Lyon 1, Villeurbanne, France.

[3] Institut de Génétique Humaine, CNRS, UPR1142, 34396 Montpellier Cedex 5, France.

Keywords: tranposable elements, piRNAs, interspecific hybridization, RNA-seq, *Drosophila*.

3.3.1.1 Abstract

Hybridization between species is a genomic stress condition that can lead to the activation of transposable elements (TEs) in both animal and plant species. Previous studies in *Drosophila buzzatii-Drosophila koepferae* hybrids showed mobilization of 28 TE families, as well as abnormal expression of *Osvaldo* and *Helena* retrotransposons in gonads. However, we ignore the precise molecular mechanisms involved in this TE release. To give insight on the causes of TE deregulation, we have performed a transcriptomic analysis of TEs in ovaries (notorious for playing a major role in TE silencing) of both parental species, as well as of F1 and backcrossed hybrids (BC). We find that 15.2% of the expressed TEs are deregulated in F1 ovaries, a proportion that decreases to 10.6% in BC1; with a bias towards overexpression in both cases. Sequencing of piRNA populations shows that differences between parental piRNAs cannot entirely explain these results. Instead, we find that piRNA pathway proteins are differentially expressed and have divergent sequences between parental species. Thus, a functional divergence of the piRNA pathway between *D. buzzatii* and *D. koepferae* may cause incompatibilities in hybrids and be at the origin of TE deregulation. However, other lines of evidence are required to understand the whole set of alterations. These analyses have been complemented with the study of F1 testes, which surprisingly exhibit a tendency towards TE underexpression. Compared to *D. buzzatii*, piRNA production seems to be enhanced in hybrid testes, showing that TE expression and regulation is sex-biased.

3.3.1.2 Introduction

Transposable elements (TEs) are mobile DNA fragments that are dispersed throughout the genome of the vast majority of both prokaryotic and eukaryotic organisms. Their capacity to mobilize, together with their repetitive nature, confers them a high mutagenic potential. TE insertions can be responsible for the disruption of genes or regulatory sequences, and can also cause chromosomal rearrangements, representing a threat to their host genome integrity (Hedges and Deininger 2007). To mitigate these deleterious effects, mechanisms of TE control are especially important in the germline, where novel insertions (as well as other mutations) can be transmitted to the progeny (Iwasaki et al. 2015; Czech and Hannon 2016).

Animal genomes have developed a TE silencing system, the piRNA (Piwi-interacting RNA) pathway (Klattenhoff and Theurkauf 2008; Brennecke and Senti 2010), that acts in the germline at both post-transcriptional and transcriptional levels (Rozhkov et al. 2013). piRNA templates form specific genomic clusters, whose transcription produces long piRNA precursors that are cleaved to

produce primary piRNAs (Brennecke et al. 2007). The resulting piRNAs can initiate an amplification loop called the ping-pong cycle, giving rise to secondary piRNAs (Brennecke et al. 2007; Gunawardane et al. 2007). A third kind of piRNAs are produced by phased cleavage of piRNA cluster transcript remnants that have first been processed during secondary piRNA biogenesis (Han et al. 2015; Mohn et al. 2015). In the soma, another small-RNA mediated silencing system, the endo-siRNA (endogenous small interference RNA) pathway, has been shown to be involved in post-transcriptional silencing of TEs (Ghildiyal et al. 2008).

These strong mechanisms of TE regulation can be relaxed under different stress conditions, leading to unexpected TE mobilization events (García Guerreiro 2012). Hybridization between species is a genomic stress that can lead to several genome reorganizations that seem to be driven by TEs (Fontdevila 2005; Michalak 2009; García Guerreiro 2014; Romero-Soriano et al. 2016). In the literature, several cases of TE proliferation in interspecific hybrids have been reported for a wide range of species, including plants (Liu and Wendel 2000; Ungerer et al. 2006; N. Wang et al. 2010) as well as animals (Evgen'ev et al. 1982; O'Neill et al. 1998; Metcalfe et al. 2007). Studies describing an enhanced TE expression in hybrids suggest that this may be caused by a TE silencing breakdown (Kelleher et al. 2012; Carnelossi et al. 2014; Dion-Côté et al. 2014; Renaut et al. 2014; García Guerreiro 2015). In this work, we propose two possible explanatory hypotheses –not mutually exclusive– to understand this breakdown, since the molecular mechanisms allowing TE release in hybrids remain unknown.

The first hypothesis, that we call the maternal cytotype failure, recalls the hybrid dysgenesis phenomenon (Picard 1976; Kidwell et al. 1977), where an increase of TE activity is observed. This occurs when *Drosophila* females whose genome is devoid of a particular TE are mated with males containing it, and is associated with the absence of specific piRNAs in the maternal cytoplasm (Brennecke et al. 2008), which are crucial to initiate an efficient TE silencing response in the progeny (Grentzinger et al. 2012). In the same logic, differences between parental species piRNA pools could lead to a transcriptional activation of some paternally-inherited TEs in interspecific hybrids. Under this hypothesis, only a subset of TE families, specific to the male species, would be deregulated after hybridization.

The second hypothesis claims that a global failure of the piRNA pathway is responsible for the observed TE activation in hybrids. It has been shown that piRNA pathway effector proteins show adaptive evolution marks (Obbard et al. 2009; Simkin et al. 2013) and their expression levels can significantly differ between different populations of the same *Drosophila* species (Fablet et al. 2014). Thus, genetic incompatibilities involving this pathway could arise even between closely

related species. The accumulated functional divergence of these proteins would cause a widespread transcriptional TE derepression, as suggested in *D. melanogaster–D. simulans* artificial (*Hmr*-rescued) hybrids (Kelleher et al. 2012).

In order to test these hypotheses and provide new insight into the mechanisms underlying TE activation in hybrids, we have performed a whole-genome study of TE expression and regulation using the species *D. buzzatii* and *D. koepferae* (*buzzatii* complex, *repleta* group). We chose this species pair as a model because hybridization between them can occur in nature (Gomez and Hasson 2003; Piccinali et al. 2004; Franco et al. 2010), providing a source of genetic variability that makes them particularly interesting for natural hybridization and speciation studies. Contrarily to *D. melanogaster* and *D. simulans*, our species allow backcrosses to be performed (Marín and Fontdevila 1998; Barbash 2010), even if their divergence time appears to be higher: 4.0-5.0 Mya for *D. buzzatii–D. koepferae* (Gomez and Hasson 2003; Laayouni et al. 2003; Oliveira et al. 2012) compared to 1.0-3.0 for *D. melanogaster–D. simulans* (Russo et al. 1995; Lachaise and Silvain 2004; Cutter 2008). Furthermore, several TE mobilization events have previously been detected in our hybrids by *in situ* hybridization (Labrador et al. 1999), amplified fragment length polymorphism (AFLP) markers (Vela et al. 2011) and/or transposon display (Vela et al. 2014). Finally, at least two of the mobilized elements, the retrotransposons *Osvaldo* and *Helena*, present abnormal patterns of expression in hybrids (García Guerreiro 2015; Romero-Soriano and García Guerreiro 2016), pointing to a failure of TE silencing.

We demonstrate that 15.2% of the expressed TE families are deregulated in F1 hybrid ovaries, in most cases overexpressed. This proportion decreases to 10.6% after a generation of backcrossing. However, even if differences between parental piRNA pools can be linked to the misexpression of some TE families, they do not explain the whole pattern of deregulation. Accordingly, our analyses of genomic TE content show that parental TE landscapes are very similar, and hence big differences in their piRNA populations are not expected. On the other hand, we demonstrate that the piRNA pathway proteins are particularly divergent between *D. buzzatii* and *D. koepferae* translated transcriptomes, which seems to lead to dissimilarities in their piRNA production strategies. Interestingly, a high proportion of the overexpressed TEs do not have associated piRNA populations in parents (nor in hybrids), pointing out a complex TE deregulation network where a failure of the piRNA pathway together with other TE silencing mechanisms would take place. Finally, we show that the effects of hybridization are sex-biased, because in testes (contrarily to ovaries) TE deregulation is globally biased towards underexpression, which can be explained by a higher production of piRNAs in hybrid males.

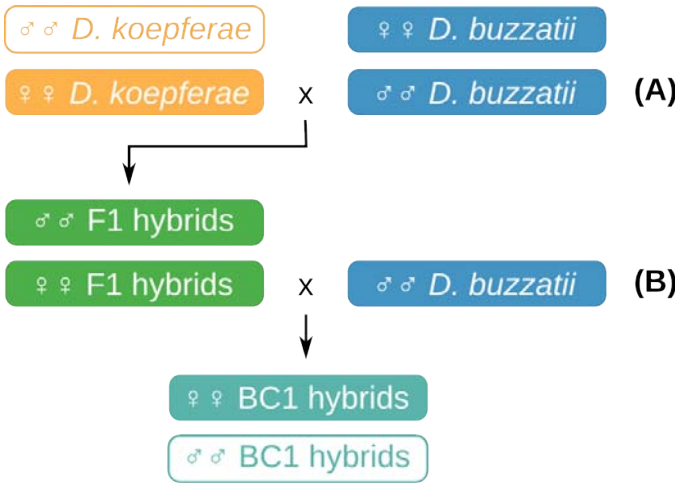


Figure 5: Crosses diagram. (A) is the first interspecific cross between *D. koepferae* (yellow) females and *D. buzzatii* (blue) males, and **(B)** is the backcross between F1 hybrids (green) females and *D. buzzatii* (blue) males, that gives rise to BC1 (turquoise). Colours have been assigned according to the *D. buzzatii/D.koepferae* genome content: yellow for *D. koepferae*, blue for *D. buzzatii*, green for F1 hybrids and turquoise for BC1 hybrids. Samples marked with a white background rectangle have not been sequenced.

3.3.1.3 Results

Qualitative changes in TE expression after interspecific hybridization

We sequenced the ovarian transcriptomes of both parental species and two hybrid generations, the F1 and a first backcross BC1 (**Figure 5**), and examined their TE expression. We also sequenced and analysed the testicular transcriptomes of *D. buzzatii* (male parental species) and F1 hybrids. Globally, we have detected expression of 415 out of 658 candidate TE families (see **Methods** and File S1 in **Annex 8.3.2**). We show that ovaries present significantly higher TE global alignment rates than testes (**Figure 6A**; Student's t=4.09, p=0.0035) whereas the global TE alignment rate between hybrids and parental species is not significantly different (Student's t=-1.10, p=0.30). However, at a qualitative level, we observe notable differences between parents and hybrids: LTR proportion is increased in both hybrid testes (from 14.2 to 31.4%) and ovaries (from 7.7-8.3 to 14.4-13.8%), as well as are RC elements (or *Helitrons*) in F1 testes (from 4.3 to 8.1%, **Figure 6B**). TE expression profiles are very similar between ovaries of *D. buzzatii* and *D. koepferae*, but parental testes (*D. buzzatii*) present a considerably lower LINE proportion (**Figure 6B**). In all cases, TE expression is mainly represented by retrotransposons (LINEs are the most expressed category

followed by LTRs). Therefore, even if the global amounts of TE expression remain unchanged after interspecific hybridization, we observe differences at the TE family expression level.

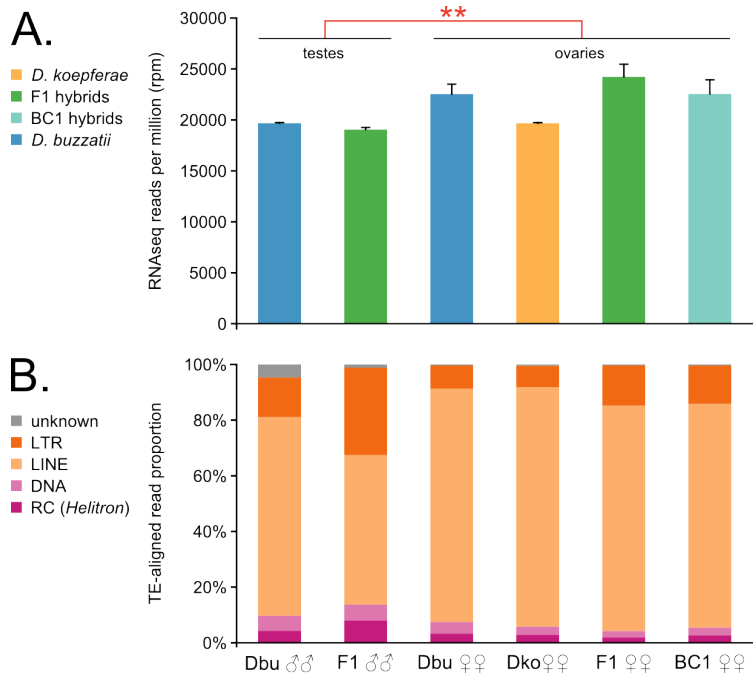


Figure 6: TE expression summary. Dbu= *D. buzzatii*; Dko= *D. koepferae*; ♂♂= testes; ♀♀= ovaries. **(A)** Mean proportion of reads aligning to the TE library. Bars represent standard deviation between replicates. ** Student's t=4.09, p=0.0035. **(B)** TE expression profiles following Repbase classification (Jurka et al. 2005): LTR and LINE (class I), DNA and RC/*Helitron* (class II), Unknown (unclassified). LTR= elements with Long Terminal Repeats; LINE= Long Interspersed Nuclear Element; RC= Rolling Circle elements (or *Helitrons*).

TE deregulation in hybrid ovaries is biased towards overexpression

Compared to *D. buzzatii* and *D. koepferae* separately, F1 ovaries present a similar number of differentially expressed TE families (221 and 234, respectively), while in BC1 expression is closer to *D. buzzatii* (149 and 254, **Figure 7A**). In both cases, hybrid ovaries present a bias towards TE overexpression compared to parental species (**Figure 7A**), with 55% of the deregulated families (on average) more expressed in hybrids (Table S1 in **Annex 8.3.9**).

When compared to both parental species, 37 TE families are significantly overexpressed in F1 and only 27 in BC1 (most of them are shared between generations, **Table 1**). Among them, 77% are retrotransposons, and *Gypsy* elements exhibit the highest fold change (FC) values. Surprisingly, we also observe 26 underexpressed families in F1 and 17 in BC1 (**Table 2**). Underexpressed TE

families are also mainly retrotransposons (71%) and their FC values tend to be lower than those of overexpressed families (**Table 1** and **2**).

Therefore, after a generation of backcrossing, the global amount of TE deregulation decreases from 15.2 to 10.6% of the 415 expressed families. In the same way, we observe that FC values are often lower in BC1 than in F1 (**Table 1** and **2**). All the deregulated TE families are transcriptionally active in both parental species (**Figure 7B**), but only 21% of them exhibit differences of expression higher than 2-fold between parental species (a total of 16 families; 14 overexpressed and 2 underexpressed, see names in **Figure 7B**).

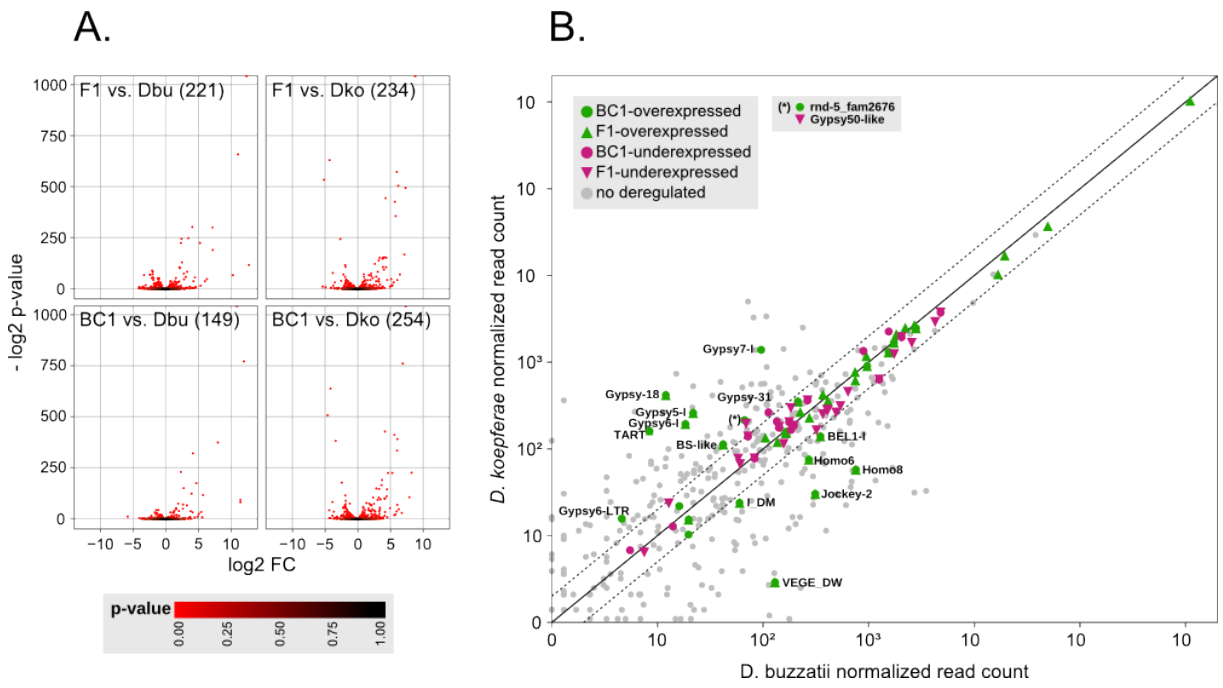


Figure 7: TE differential expression analyses in ovaries. (A) Differentially expressed TE families in hybrids compared separately to *D. buzzatii* (Dbu) and *D. koepferae* (Dko). The total number of differentially expressed TE families of each comparison is written in parenthesis. FC= fold change (hybrid vs. parent). **(B)** Expression of TE families in *D. koepferae* vs. *D. buzzatii*. In colour, deregulated TE families in hybrids (compared to both parental species). Dot lines represent 2-fold changes between parental expression and the solid line represents the same amount of expression between Dbu and Dko. Names of those TE families with differences of expression higher than 2-fold between parental species are indicated.

Table 1: Overexpressed TE families in hybrid ovaries. Dbu= *D. buzzatii*; Dko= *D. koepferae*; FC= fold change; BH= Benjamini–Hochberg correction; [a] overexpressed only in BC1; [b] FC increases after BC.

TE family	Order	Superfamily	F1 ovaries				BC1 ovaries			
			log2(FC) vs.		BH adjusted p-value		log2(FC) vs.		BH adjusted p-value	
			Dbu	Dko	Dbu	Dko	Dbu	Dko	Dbu	Dko
Homo6	DNA	hAT	2.46	4.32	5.47E-75	7.81E-135	2.38	4.25	2.26E-70	5.04E-130
Homo8	DNA	hAT	2.55	6.26	3.35E-40	5.01E-153	1.97	5.68	8.03E-24	1.77E-125
R=81	DNA	hAT	0.68	0.79	1.23E-03	1.44E-04	0.62	0.73	5.92E-03	4.50E-04
rnd-5_family-1117	DNA	hAT	0.63	0.37	1.44E-03	7.44E-02	-	-	-	-
VEGE_DW[b]	DNA	hAT	1.26	6.53	3.28E-04	2.02E-22	2.69	7.96	1.64E-16	3.04E-33
Rehavkus-2_Nvi	DNA	MULE-MuDR	0.77	0.46	8.12E-08	2.00E-03	-	-	-	-
rnd-5_family-4211	DNA	MULE-MuDR	0.37	0.56	7.16E-02	3.61E-03	-	-	-	-
DNA8-7_CQ	DNA	OtherDNA	0.61	0.65	9.85E-06	1.51E-06	0.38	0.43	1.49E-02	2.51E-03
rnd-4_family-786	DNA	Transib	0.41	0.67	5.59E-02	9.17E-04	-	-	-	-
rnd-5_family-1551	DNA	Transib	0.69	0.48	4.49E-04	1.76E-02	-	-	-	-
CR1-1_CQ	LINE	CR1	1.16	0.80	2.25E-04	1.31E-02	-	-	-	-
CR1-2_CQ	LINE	CR1	0.52	0.53	2.94E-02	2.24E-02	-	-	-	-
I_DM	LINE	I	1.28	2.58	1.07E-02	2.61E-07	1.27	2.57	1.82E-02	2.27E-07
rnd-5_family-156	LINE	I	1.68	0.96	1.65E-08	1.81E-03	1.36	0.64	1.28E-05	4.89E-02
BS-like	LINE	Jockey	5.33	3.90	5.91E-69	1.82E-45	4.73	3.31	4.52E-54	1.02E-32
Jockey-2_Dya	LINE	Jockey	2.39	5.77	5.28E-69	1.98E-129	0.32	3.70	9.10E-02	2.50E-51
rnd-3_family-39	LINE	Jockey	0.39	0.58	4.60E-03	7.14E-06	-	-	-	-
TART_B1[a]	LINE	Jockey	-	-	-	-	1.46	2.30	3.53E-02	3.45E-04
TART	LINE	Jockey	7.24	3.14	1.13E-58	2.60E-26	5.74	1.64	1.43E-36	1.11E-07

rnd-4_family-338	LINE	L2	0.57	0.40	4.36E-04	1.83E-02	-	-	-	-
rnd-5_family-2046	LINE	L2	0.71	0.65	1.84E-04	6.54E-04	-	-	-	-
Bilbo	LINE	LOA	0.83	1.02	8.33E-13	8.82E-19	0.78	0.97	4.22E-11	4.64E-17
R1_Dps	LINE	R1	0.56	0.81	3.23E-05	5.52E-10	0.53	0.78	1.57E-04	1.91E-09
rnd-5_family-1630	LINE	R1	0.53	0.63	1.03E-04	2.48E-06	0.30	0.40	7.15E-02	4.93E-03
RT2	LINE	R1	0.74	0.53	1.21E-08	5.45E-05	-	-	-	-
RTAg3	LINE	R1	0.93	1.02	3.33E-05	5.48E-06	0.54	0.63	4.22E-02	7.98E-03
RTAg4	LINE	R1	0.51	0.60	2.20E-04	6.74E-06	-	-	-	-
BEL1-I_Dmoj	LTR	BelPao	2.81	4.13	5.42E-24	1.03E-47	1.02	2.34	1.33E-03	1.15E-15
BEL1-LTR	LTR	BelPao	1.53	1.92	3.80E-03	3.25E-04	1.05	1.45	9.10E-02	9.24E-03
Gypsy-14_Dwil-I[a]	LTR	Gypsy	-	-	-	-	3.94	3.91	7.45E-02	4.72E-02
Gypsy-151_AA-I	LTR	Gypsy	0.43	0.71	4.33E-03	8.58E-07	-	-	-	-
Gypsy16-I_Dpse	LTR	Gypsy	12.76	7.39	2.88E-36	5.41E-150	11.47	6.09	2.94E-29	5.80E-102
Gypsy-172_AA-I	LTR	Gypsy	0.64	0.81	4.66E-02	7.87E-03	-	-	-	-
Gypsy-18_Dwil-I[b]	LTR	Gypsy	11.10	6.04	1.49E-199	8.22E-174	12.01	6.95	8.02E-234	2.40E-230
Gypsy-18_Dwil-LTR[b]	LTR	Gypsy	10.35	7.19	2.00E-21	9.12E-52	11.48	8.32	5.49E-26	2.18E-69
Gypsy5-I_Dya	LTR	Gypsy	12.40	8.88	0.00E+00	0.00E+00	10.94	7.41	0.00E+00	0.00E+00
Gypsy61-I_AG	LTR	Gypsy	0.31	1.00	5.90E-02	7.47E-13	-	-	-	-
Gypsy6-I_Dya[b]	LTR	Gypsy	7.21	3.87	1.15E-91	6.99E-47	8.03	4.69	5.22E-114	3.81E-69
Gypsy6-LTR_Dya[a]	LTR	Gypsy	-	-	-	-	4.17	2.48	5.89E-11	5.30E-07
Gypsy7-I_Dmoj[a]	LTR	Gypsy	-	-	-	-	4.23	0.38	5.37E-98	5.37E-02
Gypsy8-I_Dpse	LTR	Gypsy	0.42	0.84	2.23E-03	3.08E-11	-	-	-	-
R=961[a]	LTR	Gypsy	-	-	-	-	1.71	1.28	6.75E-03	3.08E-02
rnd-5_family-2676[a]	LTR	Gypsy	-	-	-	-	2.72	1.04	1.74E-22	8.93E-05
	mean		2.48		6.16E-03		3.34		1.22E-02	

Table 2: Underexpressed TE families in hybrid ovaries. Dbu= *D. buzzatii*; Dko= *D. koepferae*; FC= fold change; BH= Benjamini–Hochberg correction; [a] underexpressed only in BC1; [b] FC increases after BC.

TE family	Order	Superfamily	F1 ovaries				BC1 ovaries			
			log2(FC) vs.		BH adjusted p-value		log2(FC) vs.		BH adjusted p-value	
			Dbu	Dko	Dbu	Dko	Dbu	Dko	Dbu	Dko
Howilli1 [a]	DNA	hAT	-	-	-	-	-1.70	-1.59	8.09E-02	7.33E-02
MINOS	DNA	Tc1Mariner	-1.32	-0.53	8.12E-08	6.02E-02	-	-	-	-
rnd-5_family-1477 [a]	DNA	Tc1Mariner	-	-	-	-	-0.59	-1.13	1.21E-06	6.24E-24
rnd-5_family-3658 [a]	DNA	Tc1Mariner	-	-	-	-	-0.66	-0.97	2.23E-02	8.48E-05
Transib1_DP [b]	DNA	Transib	-0.57	-0.90	8.58E-02	2.44E-03	-0.64	-0.97	6.76E-02	8.76E-04
Transib3_DP	DNA	Transib	-2.01	-2.86	9.45E-02	8.46E-03	-	-	-	-
HELITRON1_DM	RC	Helitron	-3.37	-3.11	1.34E-02	2.37E-02	-	-	-	-
Helitron-1_Dvir	RC	Helitron	-0.81	-0.32	4.66E-08	5.73E-02	-	-	-	-
rnd-3_family-48	RC	Helitron	-0.95	-0.59	1.29E-16	7.62E-07	-0.60	-0.23	6.44E-07	7.37E-02
rnd-4_family-133	RC	Helitron	-1.08	-0.53	1.50E-06	3.50E-02	-	-	-	-
DMCR1A-like	LINE	CR1	-1.21	-0.65	8.95E-11	1.27E-03	-	-	-	-
DPSEMINIME-like	LINE	CR1	-0.76	-0.26	2.38E-08	9.53E-02	-	-	-	-
DMRER1DM-like	LINE	R1	-1.55	-1.08	4.39E-09	1.08E-04	-	-	-	-
BEL-11_Dta-I	LTR	BelPao	-1.91	-1.29	7.37E-18	1.24E-08	-	-	-	-
BEL-20_AA-I [a]	LTR	BelPao	-	-	-	-	-0.67	-0.52	2.23E-02	6.39E-02
BEL-3_Dta-I	LTR	BelPao	-0.70	-0.61	8.23E-03	2.24E-02	-0.57	-0.48	5.13E-02	7.61E-02
BEL-6_Dwil-I	LTR	BelPao	-1.08	-1.47	1.10E-02	2.05E-04	-	-	-	-
BEL-8_Dwil-I	LTR	BelPao	-2.08	-1.10	5.93E-17	3.88E-05	-	-	-	-
Nobel_I [b]	LTR	BelPao	-0.81	-0.73	9.17E-06	6.08E-05	-0.82	-0.74	9.24E-06	3.64E-05

rnd-4_family-529 [b]	LTR	BelPao	-0.45	-0.91	9.41E-02	1.06E-04	-0.70	-1.16	8.53E-03	4.98E-07
rnd-5_family-1078	LTR	BelPao	-1.00	-0.44	2.92E-12	3.79E-03	-	-	-	-
rnd-5_family-2670	LTR	BelPao	-2.02	-1.11	2.35E-28	1.50E-08	-	-	-	-
Copia-3-like [a]	LTR	Copia	-	-	-	-	-0.45	-1.04	6.63E-02	8.92E-08
rnd-5_family-4686	LTR	Copia	-0.92	-1.08	1.24E-02	2.22E-03	-	-	-	-
Beagle-like	LTR	Gypsy	-0.59	-1.27	1.58E-02	5.00E-09	-	-	-	-
Gypsy1-I_Dmoj	LTR	Gypsy	-0.85	-1.05	8.73E-04	2.01E-05	-0.53	-0.73	6.52E-02	2.80E-03
Gypsy-22_Dya-I [b]	LTR	Gypsy	-1.74	-1.63	1.23E-04	3.51E-04	-2.13	-2.02	5.53E-06	9.98E-06
Gypsy2-I_DM	LTR	Gypsy	-1.17	-0.65	3.86E-10	1.20E-03	-	-	-	-
Gypsy-31_Dwil-I [a]	LTR	Gypsy	-	-	-	-	-1.11	-2.33	5.27E-02	5.69E-07
Gypsy4-I_Dpse	LTR	Gypsy	-1.90	-0.90	1.40E-26	1.62E-06	-1.37	-0.38	8.49E-15	6.15E-02
Gypsy50-like	LTR	Gypsy	-0.98	-2.47	1.34E-02	4.85E-13	-	-	-	-
QUASIMODO-like [a]	LTR	Gypsy	-	-	-	-	-0.58	-1.20	1.62E-02	1.38E-09
rnd-5_family-1084	LTR	Gypsy	-0.91	-1.85	8.70E-03	1.66E-09	-0.67	-1.61	7.57E-02	2.96E-08
TABOR_DA-LTR [a]	LTR	Gypsy	-	-	-	-	-3.27	-3.46	5.43E-02	2.13E-02
	mean		-1.19		1.29E-02		-1.11		2.81E-02	

Divergence time between parental species and TE landscapes influence deregulation

In a previous study, *D. simulans–D. melanogaster* artificial hybrid (*Hmr*-rescued) ovaries displayed a proportion of deregulated TE families of 12.1% (similar to *D. buzzatii–D. koepferae* 15.2% in F1) which was considered to be widespread compared to the 0.7% found for protein-coding genes (Kelleher et al. 2012). To evaluate the extent of gene deregulation in our hybrids, we produced a *de novo* transcriptome assembly for each parental species. Parental transcriptomes were annotated using BLAT alignments against gene models of *D. buzzatii* (Guillén et al. 2015) and *D. mojavensis* (Drosophila 12 Genomes Consortium 2007) genomes (see **Methods**).

Table 3: Summary of assemblies and annotation. NA= not annotated; [a] clustering with CD-HIT.

		D. buzzatii	***D. koepferae***	**addition**
Trinity assemblies	contig number	49,474	26,105	75,579
Final assemblies (splitted chimers)	contig number	51,772	27,212	78,984
	% GC	44.93	45.29	45.04
	N50	3,021	2,400	2,798
	median length	697	664	686
	mean length	1,482.9	1,285.1	1414.7
	assembled bases	76,771,744	34,970,044	111,741,788
Annotation	*D. buzzatii* genome	30,386	16,897	47,283
	D. mojavensis genome	1,942	898	2,840
	total	32,328	17,795	50,123
	% annotated contigs	62.4%	65.4%	63.5%
NA contigs	contig number	19,444	9,417	28,861
	after clustering*	-	-	20,525
Final transcriptome	contig number	-	-	**70,648**
	annotated	-	-	**70.9%**

We annotated 70.9% of the final transcriptome contigs (**Table 3**) as 11,190 different protein-coding genes. Among these, 657 are overexpressed and 821 underexpressed in F1 ovaries (File S2 in **Annex 8.3.3**), reaching a proportion of deregulation of 13.2%. In BC1, it decreases to 12.3%, with 711 overexpressed and 662 underexpressed genes (File S2 in **Annex 8.3.3**). Thus, both TE and gene expression are affected at similar levels (~10-15%) in ovaries of *D. buzzatii–D. koepferae* hybrids, but they follow distinct patterns (only TEs are biased towards overexpression). It is noteworthy that

F1 and BC1-overexpressed genes have in common three enriched Gene Ontology (GO) terms: response to methotrexate, GABA receptor activity and cation-aminoacid symporter activity (**Table 4**). More interestingly, in the case of underexpressed genes, several enriched GO terms related to aminoacid metabolism, ion transport and oogenesis are shared between F1 and BC1 (**Table 4**), which may be related to the hybrid loss of fertility.

Alteration of gene expression is remarkably higher in our hybrids than in *D. simulans–D. melanogaster* ones, which might be due to differences in divergence times between these species pairs. We have calculated the most common rate of substitution per synonymous site between our parental species (dS=0.139; File S3 in **Annex 8.3.4**) and estimated their divergence time at 4.96 Mya using the mutation rate estimate of Keightley et al. (2014). This result concurs with the few available estimations of divergence between this species pair, that range between 4.02-4.63 Mya (Gomez and Hasson 2003; Laayouni et al. 2003; Oliveira et al. 2012). Using the same formula, *D. melanogaster* and *D. simulans* (with dS=0.068, Cutter 2008) would have diverged 2.43 Mya, which is in concordance with the most commonly used estimation (2-3 Mya, Lachaise and Silvain 2004) and confirms that the latter species pair are more closely related.

In spite of being closely related, *D. melanogaster* and *D. simulans* have radically different TE contents: while mostly recent and active TE copies that account for 15% of the genome are found in *D. melanogaster*; *D. simulans* carries mainly old and deteriorated copies, representing 6.9% of the genome (Modolo et al. 2014). We have examined the repeatomes of our parental species using dnaPipeTE (Goubert et al. 2015), which revealed that both their TE landscapes and abundance are very similar (**Figure 8** and File S4 in **Annex 8.3.5**). Both species seem to share similar kinds and proportions of recent and active TEs, suggesting that species divergence, rather than differences in TE content, would cause TE deregulation in our hybrids, which recalls the piRNA pathway failure hypothesis.

Differences in parental piRNA pools cannot fully explain hybrid TE expression

Differences in piRNA pools between parental species ovaries can be at the origin of TE silencing impairment (Brennecke et al. 2008), especially when piRNA levels of a particular TE are lower in the maternal species, *D. koepferae*. To test the maternal cytotype failure hypothesis, we sequenced and analysed the piRNA populations of the samples presented in **Figure 5**. Globally, antisense regulatory piRNA populations (23-30nt) were detected for 392 out of 658 candidate TE families (File S5 in **Annex 8.3.6**), mostly retrotransposons. In this case, we performed the differential expression analyses using FC values (see **Methods**).

Table 4: Gene Ontology (GO) terms with significant enrichment in overexpressed and underexpressed genes of hybrid ovaries. Only GO terms common in F1 and BC1 are shown.

Gene set	GO term	Ontology	Description	Ancestor term
Overexpressed	GO:0031427	Biological process	response to methotrexate	response to stimulus
	GO:0016917	Molecular function	GABA receptor activity	signal transduction
	GO:0005416	Molecular function	cation:aminoacid symporter activity	ion/aminoacid transport
Underexpressed	GO:0030707	Biological process	ovarian follicle cell development	oogenesis
	GO:0007304	Biological process	chorion-containing eggshell formation	oogenesis
	GO:0030703	Biological process	eggshell formation	oogenesis
	GO:0008258	Biological process	head involution	embryo development
	GO:1902221	Biological process	erythrose 4-phosphate/phosphoenolpyruvate family amino acid metabolic process	aminoacid metabolic process
	GO:1902222	Biological process	erythrose 4-phosphate/phosphoenolpyruvate family amino acid catabolic process	aminoacid metabolic process
	GO:0006558	Biological process	L-phenylalanine metabolic process	aminoacid metabolic process
	GO:0006559	Biological process	L-phenylalanine catabolic process	aminoacid metabolic process
	GO:0009074	Biological process	aromatic amino acid family catabolic process	aminoacid metabolic process
	GO:0015695	Biological process	organic cation transport	ion transport
	GO:0015101	Molecular function	organic cation transmembrane transporter activity	ion transport
	GO:0005213	Molecular function	structural constituent of chorion	-
	GO:0030312	Cellular component	external encapsulating structure	-
	GO:0042600	Cellular component	chorion	-

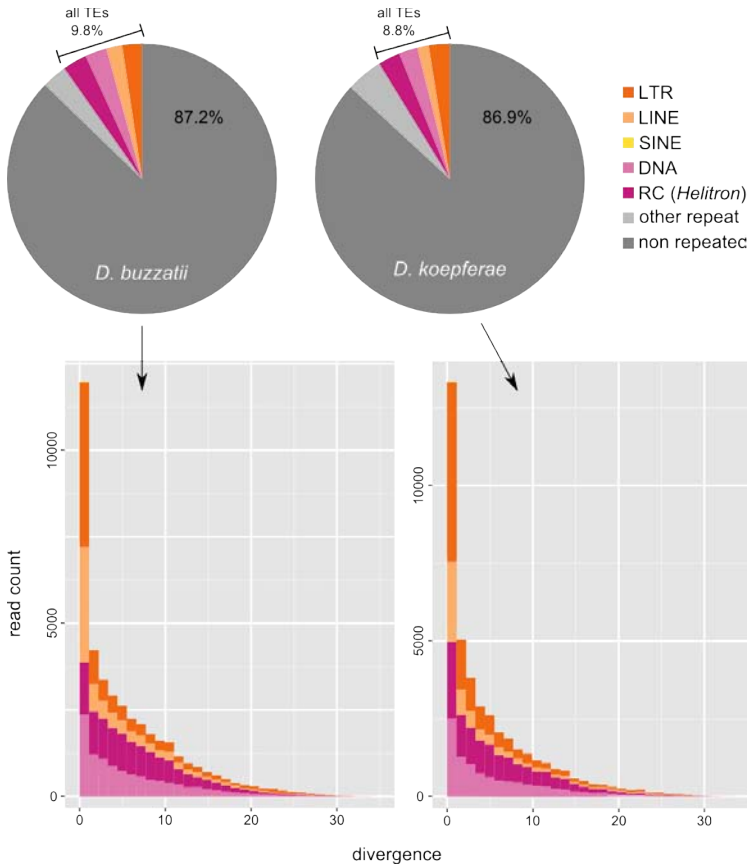


Figure 8: *D. buzzatii* and *D. koepferae* present highly similar repeatomes. (A) TE abundance in parental species genome. **(B)** TE landscapes of our parental species: genomic reads are classified according to their identity against the TE contig assembled with dnaPipeTE.

Comparisons between *D. buzzatii* and *D. koepferae* ovaries reveal that 196 TE families present differences higher than 2-fold in their antisense piRNA populations (**Figure 9A**). Families having lower levels of piRNAs in the maternal species are not always overexpressed. Indeed, among the 98 TE families that exhibit reduced abundance of piRNAs in *D. koepferae*, only 8 are overexpressed in hybrids (either in F1 or BC1, **Figure 9B-i**). Reciprocally, families having higher levels of piRNAs in the maternal species are not more commonly underexpressed: only 12 out of 98 families with higher piRNA abundance in *D. koepferae* are classified as underexpressed (**Figure 9B-iii**). Actually, some deregulated TE families present the opposite pattern (e.g. *Gypsy6-I* or *Howili1*, **Figure 9A**). However, this does not mean that differences between piRNA pools cannot account for some specific cases of TE deregulation (e.g. *TART_B1* or *MINOS*, **Figure 9A**).

Interestingly, 12 of the overexpressed families are among those without associated piRNA populations (**Figure 9B-iv**), indicating that other TE regulation mechanisms (if any) could be responsible for their regulation in the ovaries.

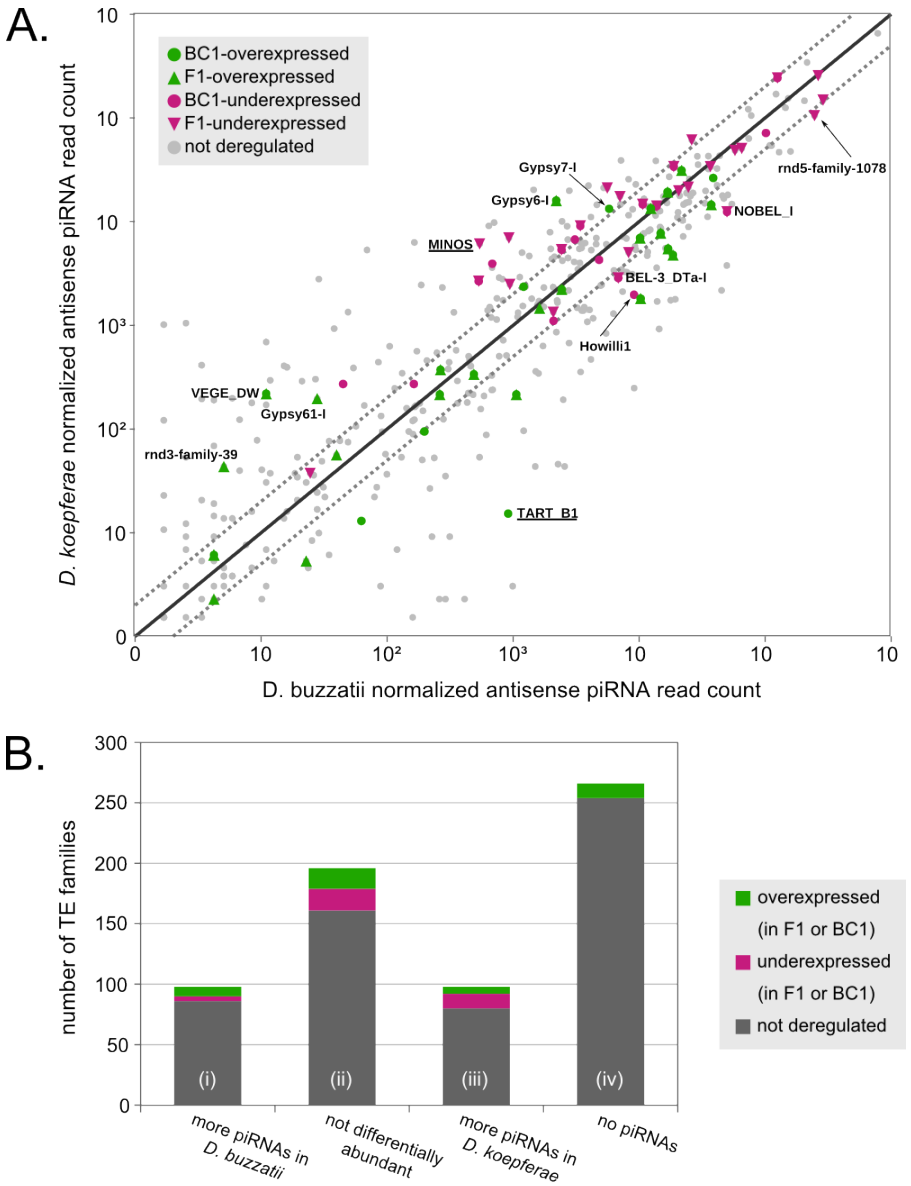


Figure 9: Parental piRNA populations and TE deregulation in ovaries. (A) Expression of TE-associated piRNA populations in *D. koepferae* (Dko) *vs. D. buzzatii* (Dbu). Dot lines represent 2-fold changes between parental piRNA amounts and the solid line represents the same piRNA levels between Dbu and Dko. Underlined TE names are examples of families that may be deregulated due to the maternal cytotype hypothesis (underexpressed with more piRNAs in *D. koepferae*, overexpressed with more piRNAs in *D. buzzatii*). Names of deregulated TE families with unexpected differences in piRNA amounts (underexpressed with more piRNAs in *D. buzzatii*, overexpressed with more piRNAs in *D. koepferae*) are also indicated, with an arrow in some cases. **(B)** Proportion of deregulated TE families of different categories, classified according to differences (of at least 2-fold) between parental piRNA populations: **(i)** more piRNAs in *D. buzzatii*, **(ii)** not differentially abundant between parental species, **(iii)** more piRNAs in *D. koepferae*, **(iv)** absence of piRNAs in both species.

piRNA production strategies differ between parental species

Artificial hybrids between *D. simulans* and *D. melanogaster* present deficient piRNA production, which displaces the size distribution of ovarian piRNAs (23-30nt) towards miRNAs and siRNAs

(18-22 nt) (Kelleher et al. 2012). However, our hybrids present an overall size distribution pattern similar to *D. koepferae* (**Figure 10A**) and similar (to higher) levels of piRNAs than parental species (File S5 in **Annex 8.3.6**). Thus, our results show that piRNAs are produced in *D. buzzatii*-*D. koepferae* hybrids.

Interestingly, we note that size distribution of small RNA populations differs between our parental species (**Figure 10A**): *D. koepferae* exhibits abundant piRNAs and lower levels of miRNAs and siRNAs, whereas the opposite is observed in *D. buzzatii*. These differential amounts of piRNAs between our parental species might be due to a functional divergence in their piRNA biogenesis pathways. To get greater insight into piRNA production strategies, we have assessed the functionality of the secondary biogenesis pathway in our samples. In the germline, mature piRNAs (either maternal or primary) can initiate an amplification loop called the ping-pong cycle, yielding sense and antisense secondary piRNAs (Brennecke et al. 2007; Gunawardane et al. 2007). In this loop, piRNAs are cleaved 10 bp after the 5' end of their template, a feature that is specific to this pathway and can be used to recognize secondary piRNAs. We have determined the ping-pong signature in our sequenced piRNA populations (Antoniewski 2014) and revealed that *D. buzzatii* ping-pong fraction is higher than *D. koepferae* (**Figure 10B**), which is in agreement with the idea of divergence in piRNA biogenesis between them.

In hybrids, ping-pong signature levels in F1 and BC1 ovaries are intermediate between parental species (F1 is more similar to *D. koepferae* and BC1 to *D. buzzatii*, **Figure 10B**), whereas in *D. simulans*-*D. melanogaster* artificial hybrids, a reduced ping-pong fraction was observed (Kelleher et al. 2012). Therefore, our hybrids differ from *D. melanogaster–D. simulans* model in that they are not characterized by a widespread decrease of piRNA production: although a few TE families present lower levels of piRNAs than both parental species (File S6 in **Annex 8.3.7**), they do not always coincide with the upregulated ones.

Interestingly, half of the overexpressed TE families (a total of 20, including the 12 without associated piRNA populations described in **Figure 9B-iv**) do not present traces of ping-pong amplification (Figure S1 in **Annex 8.3.1**). Eleven of them are LINE retrotransposons, of which five belong to the *R1* clade, whose members have a high target-specificity for 28S rRNA genes in arthropods (Eickbush et al. 1997; Kojima and Fujiwara 2003). The eight families with associated piRNA populations but without ping-pong signal could possibly be somatic elements, expressed in follicle cells of the ovaries, where secondary piRNA biogenesis does not take place.

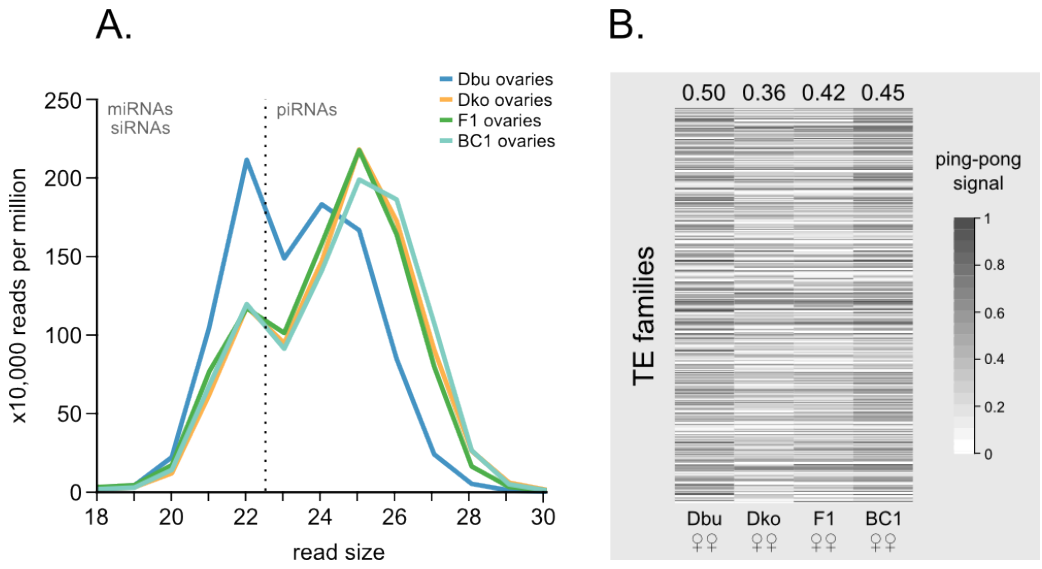


Figure 10: Characterization of piRNA populations in parental and hybrid ovaries. Dbu= *D. buzzatii*; Dko= *D. koepferae*; ♀♀= ovaries. **(A)** Read length distribution of ovarian small RNAs. The vertical dot line separates miRNAs and siRNAs (left) from piRNAs (right). **(B)** piRNA ping-pong fraction for each TE family (grey lines) and for the whole piRNA population (upper number). Only families with detectable ping-pong signal (>0) for at least one ovarian sample are represented.

piRNA pathway proteins have rapidly evolved

Although the piRNA pathway is highly conserved across the metazoan lineage, some of its effector proteins are encoded by genes bearing marks of positive selection (Simkin et al. 2013). The accumulated divergence between these proteins has been proposed to account for the TE silencing failure in *Hmr*-rescued interspecific hybrids (Kelleher et al. 2012). To elucidate the global failure hypothesis, we have aligned *D. buzzatii* and *D. koepferae* translated transcriptomes (see **Methods**) against each other and assessed their identity percentage distribution, with a resulting median identity of 97.2% (**Figure 11**).

We have then identified in *D. buzzatii* and *D. koepferae* translated transcriptomes a total of 30 protein-coding genes known to be involved in TE regulation (Yang and Pillai 2014) as reciprocal best BLAST hits of their *D. melanogaster* putative orthologs (their names and symbols are listed in **Table 5**). Alignments of all these genes between our parental species exhibit identity percentages lower than the median –their own median equals 92.5%– with the exception of the helicase Hel25E, whose sequence is identical in *D. buzzatii* and *D. koepferae* (**Figure 11**). Among the 10 most divergent proteins (identity ≤90%), we find factors involved in both piRNA biogenesis (*e.g.* zucchini, tejas) and TE silencing (*e.g.* Panoramix, maelstrom, Hen1 and qin). Thus, protein

divergence between our studied species could cause hybrid incompatibilities in both biogenesis and function of piRNAs.

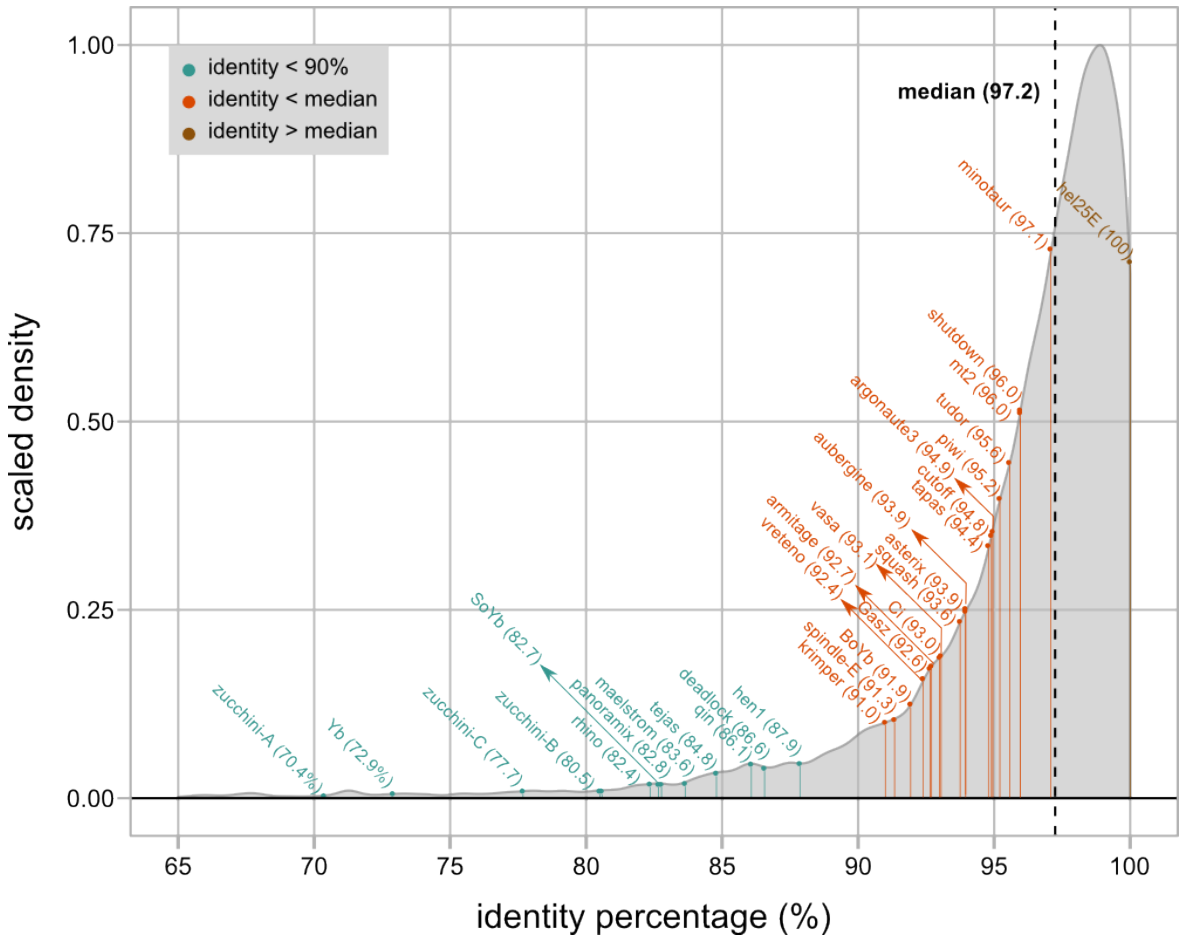


Figure 11: Distribution of identity percentages between *D. buzzatii* and *D. koepferae* proteomes (see **Methods**). A total of 30 proteins involved in the piRNA pathway were identified as reciprocal best BLAST hits of their *D. melanogaster* orthologs (represented by vertical bars, their identity in parenthesis). For Zucchini, four sequences were recognized as putative paralogs and named zucchini-A, B, C and D (only zucchini-A, B and C are shown because zucchini-D was only identified in *D. buzzatii*). At least in two other species of the genus *Drosophila*, *D. melanogaster* and *D. grimshawi*, paralogs of Zucchini have been identified (Drosophila 12 Genomes Consortium 2007).

We have also examined the expression of these 30 protein-coding genes and revealed significant differences between our parental species for all of them, with the exception of *Hen1*, *Panoramix* (*Panx*) and *tejas* (*tej*, **Table 5**). The highest FC (log2FC=5.0) is attributed to *krimper* (*krimp*, more expressed in *D. buzzatii*), known to participate in the ping-pong amplification process (Sato et al. 2015; Webster et al. 2015). Moreover, the two main genes involved in secondary piRNA biogenesis, *Aubergine* (*Aub*) and *Argonaute3* (*Ago3*), are also more expressed in *D. buzzatii* (**Table 5**). Altogether, these results are consistent with the higher ping-pong fraction reported in this species

(**Figure 10B**). Therefore, divergence in piRNA production between our parental species can be explained by the accumulated divergence in their piRNA pathway effector proteins as well as by the important differences in their expression levels.

When comparing hybrids to both parental species (**Table 5**), we observe significant underexpression of Hen1 (involved in primary and secondary piRNA biogenesis) and Sister of Yb (SoYb, involved in primary piRNA biogenesis) in both F1 and BC1. On the other hand, significant overexpression of Panx (involved in transcriptional silencing) also occurs in both hybrid generations. Those three genes are among the most divergent between parental species (identity≤90%, **Figure 11**) and their altered expression could also partially account for TE deregulation.

Interspecific hybridization has sex-biased effects on TE deregulation

An enhanced piRNA production may cause a bias to TE underexpression in hybrid testes

F1 testes present 256 differentially expressed TE families compared to *D. buzzatii* (more than any hybrid-parent comparison in ovaries, **Figure 12A**), and, as in ovaries, most of them are retrotransposons (File S7 in **Annex 8.3.8**). Although we cannot compare hybrids to both parental species, we observe that TE underexpression in hybrid testes prevails over their overexpression (Table S1 in **Annex 8.3.9**), showing that TE deregulation exhibits sex-biased patterns.

Regarding piRNA populations, the global piRNA production seems to be enhanced in F1 hybrids compared to *D. buzzatii* (**Figure 12B**), and the ping-pong fraction is also increased (**Figure 12C**). Besides, there is a bias towards piRNA overexpression of TE families in hybrids: 130 TE families exhibit more piRNAs in hybrids than in *D. buzzatii*, whereas 87 families have lower piRNA levels in hybrids (considering ≥2-fold differences, File S7 in **Annex 8.3.8**). Therefore, in the case of males, the bias towards TE underexpression seems to be explained by a higher production of piRNAs.

TE expression and piRNA production are sex-biased

The described sex-biased TE deregulation patterns are consistent with the remarkable differences in TE expression observed between testes and ovaries. Our results show that opposite sex samples always present more differences than samples of the same sex (Table S1 in **Annex 8.3.9**). In particular, testes tend to present higher TE expression than ovaries (Table S1 in **Annex 8.3.9**): for instance, 303 TE families present differential expression between ovaries and testes of *D. buzzatii*, of which 164 are more expressed in males than in females (**Figure 12A**). piRNA production also differs between sexes in *D. buzzatii*: testes exhibit lower global piRNA amounts (**Figure 12B**) and

lower ping-pong signature levels than ovaries (**Figure 12C**). Accordingly, alignment rates of piRNAs to TEs are significantly higher in ovaries than in testes (File S5 in **Annex 8.3.6**, Student's t=-9.26, p=0.01586). Therefore, males tend to present higher TE expression and lower amounts of piRNAs than females.

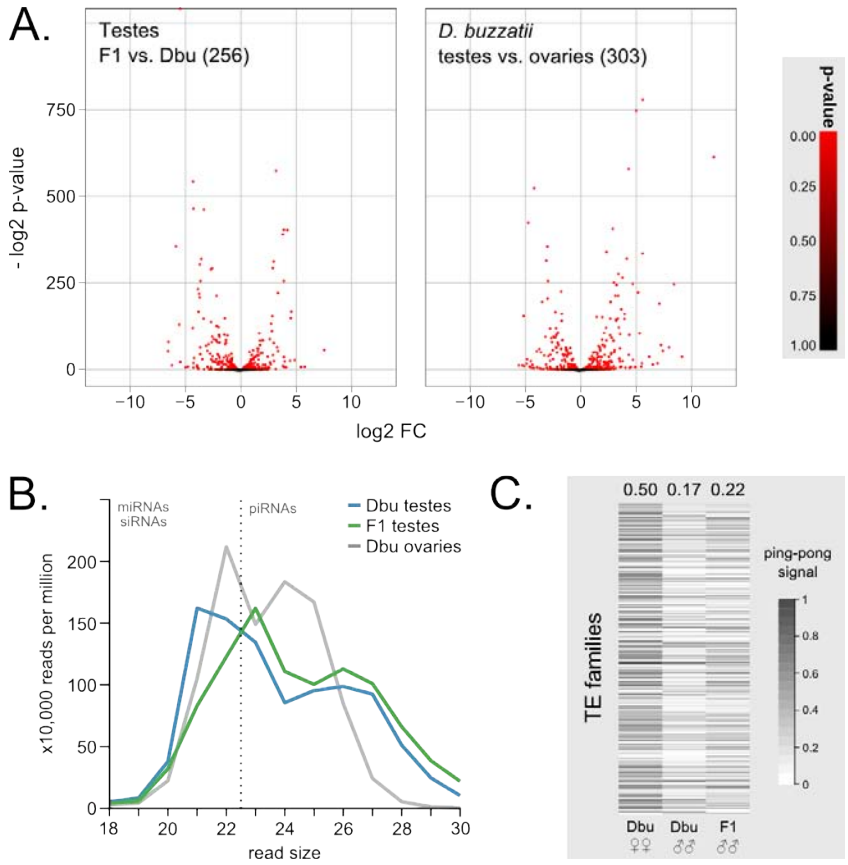


Figure 12: Differential expression analyses in testes. Dbu= *D. buzzatii*; ♂♂= testes; ♀♀= ovaries. **(A)** Differentially expressed TE families between F1 testes and Dbu (left) and between sexes of *D. buzzatii* (right). The total number of significant differences of each comparison is written in parenthesis. FC= fold change. **(B)** Read length distribution of *D. buzzatii* (testes and ovaries) and F1 testes small RNAs. The vertical dot line separates miRNAs and siRNAs (left) from piRNAs (right). **(C)** piRNA ping-pong fraction for each TE family (grey lines) and for the whole piRNA population (upper number). Only families with detectable ping-pong signal (>0) for at least one sample are represented.

Table 5: Summary of differential expression analyses of piRNA pathway genes: comparisons between parental species and between parents and hybrids. Dbu= *D. buzzatii*; Dko= *D. koepferae*; FC= fold change; BH= Benjamini–Hochberg correction; * significant p-value.

Gene name	Gene symbol	*D. buzzatii* vs. *D. koepferae*			F1 vs. parental species				BC1 vs. parental species			
		% id	log2(FC)	BH p-value	log2(FC)		BH adjusted p-value		log2(FC)		BH p-value	
					Dbu	Dko	Dbu	Dko	Dbu	Dko	Dbu	Dko
Argonaute3	**Ago3**	94.90	0.80	3.60E-29*	-0.77	0.02	1.69E-27*	7.68E-01	-0.76	0.04	6.66E-26*	6.41E-01
Armitage	**armi**	92.70	-0.59	1.51E-18*	0.43	-0.16	2.77E-10*	2.24E-02*	0.27	-0.33	1.86E-04*	1.43E-06*
asterix	**arx**	93.89	1.73	4.67E-65*	-0.30	1.43	3.21E-03*	2.45E-44*	-0.02	1.71	8.72E-01	2.43E-63*
aubergine	**aub**	93.92	2.62	3.45E-183*	-0.98	1.64	1.24E-26*	1.56E-72*	-0.46	2.16	1.08E-06*	1.40E-124*
Brother of Yb	**BoYb**	91.93	-0.42	9.63E-09*	0.52	0.10	1.79E-12*	1.83E-01	0.49	0.07	7.25E-11*	3.39E-01
cubitus interruptus	**Ci_tf**	92.97	-1.52	2.73E-18*	0.34	-1.18	6.40E-02*	1.66E-11*	0.24	-1.28	2.55E-01	2.78E-13*
cutoff	**cuff**	94.79	1.85	1.62E-78*	-0.64	1.22	2.77E-10*	2.16E-34*	-0.07	1.78	5.77E-01	1.68E-72*
deadlock	**del**	86.56	-0.88	7.51E-14*	0.32	-0.57	8.98E-03*	2.57E-06*	-0.03	-0.91	8.72E-01	1.82E-14*
GASZ ortholog	**Gasz**	92.64	0.65	1.00E-21*	0.07	0.72	3.05E-01	3.98E-26*	0.37	1.02	1.01E-07*	8.22E-52*
helicase at 25E	**Hel25E**	100	-0.41	1.36E-17*	0.25	-0.16	2.97E-07*	1.29E-03*	0.07	-0.34	2.51E-01	1.40E-12*
Hen1	**Hen1**	87.86	-0.02	9.13E-01	-0.44	-0.46	2.50E-06*	1.87E-06*	-0.50	-0.51	2.48E-07*	7.01E-08*
krimper	**krimp**	91.00	5.04	0.00E+00*	-0.62	4.41	3.02E-32*	0.00E+00*	-0.07	4.97	2.59E-01	0.00E+00*
maelstrom	**mael**	83.64	-1.20	8.48E-66*	0.77	-0.43	1.69E-27*	8.37E-10*	0.39	-0.81	1.13E-07*	6.11E-31*
minotaur	**mino**	97.08	-0.30	1.11E-04*	0.31	0.01	9.79E-05*	9.17E-01	0.03	-0.27	7.79E-01	5.30E-04*
Methyltransferase2	**Mt2**	95.95	0.74	9.90E-18*	-0.07	0.67	3.65E-01	2.95E-14*	-0.06	0.68	5.77E-01	6.58E-15*
Panoramix	**Panx**	95.95	0.01	9.20E-01	0.48	0.50	3.89E-09*	1.81E-09*	0.32	0.33	1.86E-04*	5.27E-05*
piwi	**piwi**	95.21	0.13	4.58E-02*	-0.23	-0.11	2.51E-04*	1.03E-01	-0.20	-0.07	2.49E-03*	2.63E-01
qin	**qin**	86.07	-1.30	9.28E-14*	0.47	-0.83	8.98E-03*	2.85E-06*	0.02	-1.29	9.23E-01	2.94E-13*
rhino	**rhi**	82.35	-1.03	7.85E-27*	0.34	-0.69	6.93E-04*	5.76E-13*	-0.06	-1.09	6.61E-01	1.13E-29*

shutdown	**shu**	95.97	2.26	0.00E+00*	-0.64	1.63	1.09E-53*	4.43E-302*	-0.17	2.10	1.37E-04*	0.00E+00*
Sister of Yb	**SoYb**	82.65	-0.32	4.11E-02*	-1.30	-1.62	1.43E-16*	4.20E-25*	-0.50	-0.82	2.11E-03*	9.76E-08*
spindle E	**spn-E**	91.34	-0.85	3.11E-17*	0.52	-0.33	5.13E-07*	1.29E-03*	0.23	-0.62	3.73E-02*	1.27E-09*
squash	**squ**	93.55	1.34	8.63E-23*	-0.72	0.62	1.10E-07*	9.35E-06*	-0.73	0.61	1.45E-07*	1.09E-05*
tapas	**tapas**	94.42	-0.94	3.03E-19*	0.63	-0.31	3.74E-09*	3.97E-03*	0.17	-0.77	1.67E-01	3.09E-13*
tejas	**tej**	84.79	0.01	9.62E-01	0.15	0.15	1.95E-01	1.83E-01	0.02	0.02	8.90E-01	8.52E-01
tudor	**tud**	95.56	-0.50	7.43E-04*	0.32	-0.19	3.89E-02*	2.26E-01	0.14	-0.37	4.80E-01	1.50E-02*
vasa	**vas**	93.05	0.67	1.41E-43*	-0.16	0.51	1.56E-03*	5.27E-26*	-0.11	0.56	4.90E-02*	3.57E-31*
vret	**vreteno**	92.39	0.68	7.64E-21*	-0.29	0.39	9.79E-05*	1.09E-07*	-0.26	0.42	7.92E-04*	6.71E-09*
Yb	**Yb**	72.89	1.05	4.22E-43*	-0.09	0.96	2.23E-01	1.11E-35*	-0.37	0.68	5.50E-07*	2.91E-18*
zucchini (A)	**zucA**	70.37	-1.55	4.19E-62*	1.21	-0.34	8.74E-38*	3.07E-04*	0.87	-0.67	5.21E-20*	4.03E-13*
zucchini (B)	**zucB**	80.50	-2.17	2.02E-04*	1.02	-1.15	1.10E-01	2.24E-02*	0.71	-1.45	3.57E-01	4.31E-03*
zucchini (C)	**zucC**	77.68	1.16	8.18E-53*	-0.28	0.88	1.65E-04*	2.05E-30*	-0.22	0.95	5.11E-03*	4.67E-35*
zucchini (D)	**zucD**	-	-0.43	6.87E-01	0.04	-0.39	9.62E-01	7.01E-01	0.48	0.05	6.61E-01	9.55E-01

3.3.1.4 Discussion

TE overexpression prevails over underexpression in *D. buzzatii–D. koepferae* hybrid ovaries (**Table 1**,**2** and S1 in **Annex 8.3.9**). This concurs with several studies focused on a single or few TEs, where higher transcription levels in hybrids than in parents were observed (Kawakami et al. 2011; Carnelossi et al. 2014; García Guerreiro 2015). At a whole-genome level, a few surveys also report cases of TE families underexpressed in hybrids, but these results are generally out of the main attention focus and consequently poorly discussed. For instance, in lake whitefish hybrids, approximately 38% of differentially expressed TEs are underexpressed in hybrids (Dion-Côté et al. 2014), a similar result to what we find in ovaries. Another well-studied case is that of hybrid sunflowers, where F1 hybrids present lower expression of the majority of TEs compared to parental species (Renaut et al. 2014). The presence of both overexpressed and underexpressed TEs suggests that hybrid TE deregulation is more complex than previously expected and may depend on the TE family.

Functional divergence between parental piRNA pathways can lead to hybrid incompatibilities

We demonstrate that TE families with important differences in their piRNA amounts between *D. buzzatii* and *D. koepferae* are not more commonly deregulated than families with similar levels (**Figure 9**). This shows that the maternal cytotype failure hypothesis cannot completely account for the entire pattern of TE deregulation observed, which is consistent with the similiarity of TE landscapes between our parental species (**Figure 8**). Thus, this explanation might be valid only for some particular TE families (**Figure 9**).

Sequence divergence between maternal piRNAs and paternal TE transcripts (and the reciprocal) could also lead to a decrease of silencing efficacy in hybrids, as suggested by piRNA alignment results on our TE library (File S5 in **Annex 8.3.6**). A genome-wide comparison of sequences within a TE family between parental species cannot be performed because sequenced TEs in *D. koepferae* are scarce and its genome has not been sequenced yet. However, some TE families, such as *Helena*, have been shown to be highly conserved between these species (Romero-Soriano and García Guerreiro 2016). The presence of underexpressed TE families in hybrids also seems to rule out this explanation.

Therefore, our results point rather to the piRNA pathway global failure hypothesis, which states that accumulated divergence of piRNA pathway effector proteins is responsible for hybrid TE

deregulation. In this way, we show that proteins involved in piRNA biogenesis and function are more divergent than expected between *D. buzzatii* and *D. koepferae* (**Figure 11**). Consistent with this observation, previous studies in other *Drosophila* species have demonstrated that some of these proteins are encoded by rapidly evolving genes with marks of adaptive selection (Obbard et al. 2009; Simkin et al. 2013). Furthermore, we find that almost all piRNA pathway genes present significant differences in expression between *D. buzzatii* and *D. koepferae* (**Table 5**). Such level of variability was also observed between different populations of a same species, *D. simulans* (Fablet et al. 2014).

D. koepferae seems to produce higher amounts of piRNAs compared to *D. buzzatii*, that exhibits higher levels of ping-pong signature (**Figure 10**). Those differences in global piRNA production strategies between parental species could be linked to the divergence and variability in expression between piRNA pathway genes. Indeed, the two main effectors of ping-pong amplification, *Aub* and *Ago3*, are more expressed in *D. buzzatii* than in *D. koepferae* (log2FC=2.62 and 0.80, respectively – **Table 5**), which is consistent with the important ping-pong fraction detected in this species. Furthermore, an excess of *Aub* expression relative to *Piwi* could lead to a decrease of piRNA production due to a less efficient phased piRNA biogenesis. After the cleavage of a piRNA cluster transcript by Ago3 in the ping pong cycle, the remnants of this transcript are loaded into Aub and processed to form the 3' end of an antisense Aub-bound piRNA (Czech and Hannon 2016). The excised fragment of the piRNA cluster transcript is usually loaded into Piwi (and to a lesser extent, into Aub) and cut by Zucchini (Zuc) every 27-29 nucleotides, producing phased antisense piRNAs that allow sequence diversification (Han et al. 2015; Mohn et al. 2015). We can hypothesize that an excess of *Aub* expression leads to a more frequent loading of this protein for phased piRNA production; impairing the efficiency of phasing in *D. buzzatii*. This would lead to lower levels of piRNAs in *D. buzzatii*, that would mostly be produced by ping-pong amplification.

Contrary to *Aub*, *qin* is more expressed in *D. koepferae* than in *D. buzzatii* (log2FC=-1.30, **Table 5**), which can be at the origin of the observed lower amounts of antisense piRNAs in *D. buzzatii* (File S5 in **Annex 8.3.6**). Qin is known to enforce heterotypic ping-pong between Aub and Ago3 by preventing futile homotypic Aub:Aub cycles, which mainly produce sense piRNAs (Zhang et al. 2011). A recent study has demonstrated that homotypic Aub:Aub ping-pong also generates lower Piwi-bound antisense phased piRNAs, because qin ensures the correct loading of Piwi with antisense sequences (Wang et al. 2015). Therefore, a lower expression of *qin* (coupled with an excess of Aub) could lead to a less efficient production of antisense piRNAs (both secondary and phased) in *D. buzzatii* compared to *D. koepferae*. However, we must note that the remarkably higher

expression levels of *krimper* in *D. buzzatii* (log2FC=5.0, **Table 5**) may diminish these effects, because krimper contributes to heterotypic ping-pong cycle formation by sequestering unloaded Ago3 proteins to prevent illegitimate access of other RNA sequences into them (Sato et al. 2015; Webster et al. 2015).

D. buzzatii and D. koepferae seem to present a functional divergence of the piRNA pathway, which could likely be at the origin of TE misregulation in hybrids. However, contrary to the observed in *D. melanogaster–D. simulans* artificial hybrids, our hybrids do not exhibit deficient piRNA production (Kelleher et al. 2012). Indeed, global piRNA amounts in hybrids are higher than in *D. buzzatii* and resemble the amounts observed in *D. koepferae* (**Figure 10A** and File S5 in **Annex 8.3.6**); and hybrid secondary piRNA biogenesis presents intermediate levels between parental species (**Figure 10B**). Thus, incompatibilities in our hybrids may entail piRNA-mediated silencing effectors rather than proteins involved in piRNA biogenesis, even though both kinds of protein are among those with the lowest identity percentages (**Figure 11**).

Misexpression of *SoYb, Hen1* and *Panoramix* can influence hybrid TE expression

Two of the piRNA pathway genes, *SoYb* and *Hen1*, are underexpressed in hybrids (**Table 5**). Hen1 is known to methylate piRNAs at their 3' ends in both follicle and germ cells (Horwich et al. 2007; Saito et al. 2007), but the impact of its mutation on TE expression may depend on the TE family. For instance, overexpression of *HeT-A* retrotransposon was observed in *Hen1* mutants due to a higher instability of piRNAs (Horwich et al. 2007), but other mutants exhibited an unchanged expression of retrotransposons (Saito et al. 2007). *SoYb* seems to be involved in primary piRNA biogenesis and has a partially redundant function with its paralog *BoYb* (Handler et al. 2011). Thus, even a complete gene loss of *SoYb* could be compensated by *BoYb* and would not lead to a widespread TE overexpression. Curiously, *BoYb* was underexpressed in *D. simulans–D. melanogaster* artificial hybrids (Kelleher et al. 2012). Although downregulation of *Hen1* and *SoYb* cannot explain the whole pattern of TE deregulation, we cannot dismiss it as a possible contributor to TE overexpression in some cases.

On the other hand, overexpression of *Panoramix*, known to be essential for TE transcriptional silencing (Czech et al. 2013; Handler et al. 2013; Sienski et al. 2015; Yu et al. 2015) may compensate silencing deficiencies (especially at a post-transcriptional level) and be at the origin of TE underexpression.

TE deregulation may involve other mechanisms

We have shown that TE deregulation in hybrid ovaries may be related to the piRNA pathway in terms of i) incompatibilities due to its divergence between parental species, ii) misregulation of some genes involved in TE silencing and iii) differences between parental piRNA pools (for a few TE families). However, changes in this pathway may not explain the whole set of alterations of TE expression observed in hybrids. Actually, an important fraction of overexpressed TE families does not present any associated piRNA (**Figure 9B**).

For instance, the endo-siRNA pathway is known to silence TEs in somatic and germinal tissues, with a partially redundant function with the piRNA pathway in gonads (Saito and Siomi 2010). Although our hybrids do not present lower global levels of 21 nucleotide reads than parental species (**Figure 10A**), we cannot completely reject the involvement of a putative endo-siRNA pathway dysfunction in TE deregulation, particularly for somatic elements. With our data, we cannot distinguish between somatic and germinal elements, and related bibliography in our species model is virtually nonexistent. However, the presence of *gypsy* elements among deregulated families (**Table 1** and **2**) could indicate that some of them are indeed expressed in follicle somatic cells.

In wild wheat hybrids, two TE defence mechanisms have been proposed to be activated: deletion and methylation (Senerchia et al. 2015). In *Drosophila*, DNA methylation is not common, but internal or complete deletions of TE copies have been suggested to act as a TE prevention mechanism against genome invasions (Petrov and Hartl 1998; Lerat et al. 2011; Romero-Soriano and García Guerreiro 2016). In that case, suppression of active insertions could reduce the RNA amounts of some TE families, contributing to their underexpression. Furthermore, recombination between copies is known to control R1 elements expansion in *Drosophila*. These elements are specifically inserted in 28S rRNA genes and their copies are often deleted by recombination events (Eickbush and Eickbush 2014).

Finally, histone methylation marks linked with permissive or repressive chromatin states have frequently been associated with TE sequences and their surroundings (Klenov et al. 2007; Yasuhara and Wakimoto 2008; Riddle et al. 2011; Yin et al. 2011). We must note that this has been shown to be tightly connected with the piRNA pathway. For instance, expression of piRNA clusters depends (directly or indirectly) on methylation marks (Rangan et al. 2011; Goriaux et al. 2014; Mohn et al. 2014; Molla-Herman et al. 2015), and piRNA-mediated transcriptional silencing triggers the deposition of repressive H3K9me3 marks. However, other mechanisms (including endo-siRNAs)

are also able to recruit this silencing machinery leading to heterochromatin formation. Failure in the deposition of histone modifications could hence result in abnormal TE expression.

TE deregulation across generations of hybridization

Interspecific gene flow between *D. buzzatii* and *D. koepferae* is a natural source of genetic diversity that can only be maintained through introgression of a parental genome in F1 females (F1 males are all sterile, Marin et al. 1993). Therefore, the study of backcrossed hybrids delves into the understanding of the real impact of hybridization in nature. We show that differences in ovarian TE expression between hybrids and parents are concordant with the expected *D.buzzatii*/*D.koepferae* genome fraction at each generation: F1 is equally distant from both parental species, whereas BC1 drifts apart from D. koepferae (**Figure 7A**). Furthermore, the total amount of deregulated TE families is lower in BC1 (10.6% of the expressed TEs) than in F1 (15.2%): a generation of backcrossing seems to be sufficient to restore the regulatory mechanisms of some families, but not of the totality. A similar result was reported in inbred lines of *Oryza sativa* introgressed with genetic material from the wild species *Zizania latifolia*, where *copia* and *gypsy* retrotransposons were activated and then rapidly repressed within a few selfed generations (Liu and Wendel 2000). F1 and BC1 ovaries exhibit the lowest number of differentially expressed TEs within one-to-one sample comparisons (Table S1 in **Annex 8.3.9**) and present similar TE expression profiles (**Figure 6B**). This points to the hypothesis that more generations would be necessary to restore TE expression to the parental levels. Indeed, if TE activation in hybrids is caused by the failure of different epigenetic mechanisms (Michalak 2009), these are expected to be mitigated after several backcrosses thanks to the dominance of one of the parental genomes. In agreement to this hypothesis, we showed in a recent study that TE activation causes a genome expansion in *D. buzzatii–D. koepferae* hybrid females, but the C-value decreases after the first backcross (Romero-Soriano et al. 2016).

Tendency to TE repression in hybrid testes demonstrates that TE regulation is sex-biased

We show that TE expression presents different patterns between ovaries and testes, both at the quantitative and qualitative levels (**Figure 6**). Other studies have reported tissue-specific expression of transposons between male and female gonads. For instance, in *D. simulans* and *D. melanogaster*, transcripts of *412* are only found in testes (Borie et al. 2002), *I-like* elements are more expressed in testes than in ovaries of *D. mojavensis* and *D. arizonae* (Carnelossi et al. 2014), as well as are *Osvaldo* and *Helena in D. buzzatii and D. koepferae* (García Guerreiro 2015; Romero-Soriano and García Guerreiro 2016). All these studies show higher transcript abundances in male gonads, which

is consistent with the bias we observe towards testes overexpression compared to ovaries (Table S1 in **Annex 8.3.9**).

These findings point out a differential TE regulation between male and female gonads, which was previously suggested by studies in *Drosophila* testes demonstrating that male piRNA biogenesis is not always performed by the same mechanisms as in ovaries (Nagao et al. 2010; Siomi et al. 2010). Concordantly, we observe that testes have lower piRNA amounts and a less efficient ping-pong cycle than ovaries (**Figure 12**). It has indeed been shown that piRNAs in testes are not only involved in TE repression but also in gene silencing, particularly of *Stellate* and *vasa* (Nishida et al. 2007).

Our results on TE deregulation in hybrids fully support the idea of sex-specificity in TE silencing. Contrarily to ovaries, hybrid testes exhibit a bias towards TE underexpression compared to *D. buzzatii* (Table S1 in **Annex 8.3.9**). Accordingly, the retrotransposon *Helena* was shown to exhibit lower transcript abundances in F1 testes than in *D. buzzatii* and *D. koepferae* (Romero-Soriano and García Guerreiro 2016), as was the case for most TE families in a transcriptomic study in F1 sunflower hybrids (Renaut et al. 2014). Although two other studies in *Drosophila* hybrids, focused on individual TEs, displayed the opposite effect (Carnelossi et al. 2014; García Guerreiro 2015), we consider that disparity between specific studies fits in our global results.

TE underexpression prevalence in our hybrid testes can be explained by an increase of piRNA production and ping-pong signal in F1 testes (**Figure 12B** and **C**). Thus, activation of piRNA biogenesis, especially through the ping-pong cycle, seems to be responsible for TE repression in testes. Consistent with this tight repression of TE activity in males, the genome size increase observed in *D. buzzatii–D. koepferae* hybrids occurs only in females, whereas the hybridization impact in male genome size is undetectable (Romero-Soriano et al. 2016).

3.3.1.5 Conclusion and perspectives

We suggest that TE deregulation in ovaries of *D. buzzatii–D. koepferae* hybrids might be the result of several interacting phenomena. First, a partial failure of the piRNA pathway due to a functional divergence between parental species, especially in silencing effector proteins. Second, misexpression of some piRNA pathway genes, that act at different levels. Finally, different amounts of TE-specific piRNAs in maternal cytoplasm, which could be responsible for the deregulation of some TE families.

Furthermore, we cannot discard that other TE repression mechanisms might partially account for the observed set of deregulations. For instance, the endo-siRNA pahway function could also be affected, deletions could play a role in TE underexpession, and histone post-translational modifications may alter the chromatin state pattern of the hybrid genome and cause either overexpression or underexpression (depending on the TE insertion). The study of these mechanisms would be an interesting focus for future investigations, as it could shed light on other causes of hybrid TE deregulation.

On the other hand, comparison of ovaries and testes show that TE regulation is sex-biased. Surprisingly, piRNA biogenesis is enhanced in hybrid testes, which underlines that hybridization is a genomic stress that can activate response pathways to counteract TE deregulation. Further work in testes needs to be performed to elucidate the observed differences in TE silencing, which could be crucial to understand the molecular basis of hybrid breakdown and sterility.

3.3.1.6 Methods

Drosophila stocks and crosses

Interspecific crosses were performed between males of *D. buzzatii* Bu28 strain, an inbred line originated by the union of different populations (LN13, 19, 31 and 33) collected in 1982 in Los Negros (Bolivia); and females of *D. koepferae* Ko2 strain, an inbred line originated from a population collected in 1979 in San Luis (Argentina). Both lines were maintained by brother-sister mating for more than a decade and are now kept by mass culturing.

We performed 45 different interspecific crosses of 10 *D. buzzatii* males with 10 *D. koepferae* virgin females (in order to obtain F1 individuals), then 30 backcrosses of 10 *D. buzzatii* males with 10 hybrid F1 females (which gave rise to BC1 females). All stocks and crosses were reared at 25ºC in a standard *Drosophila* medium supplemented with yeast.

RNA extraction, library preparation and sequencing

Flies were dissected in PBT (1× phosphate-buffered saline [PBS], 0.2% Tween 20), 5-6 days after their birth. Total RNA was purified from testes (n=30 pairs per sample for *D. buzzatii* and n=45 pairs per sample for F1 hybrids) or ovaries (n=20 pairs per sample) with the Nucleospin RNA purification kit (Macherey-Nagel). RNA quality and concentration was evaluated using Experion Automated Electrophoresis System (Bio-rad), in order to keep only high quality samples. Two Illumina libraries of 250-300bp fragments were prepared for each kind of sample (*D. buzzatii*, *D. koepferae*, F1 and BC1 ovaries; and *D. buzzatii* and F1 testes), using 2µg of purified RNA.

Duplicate libraries correspond to biological replicates (ovaries from different crosses and separate RNA extractions). Sequencing was performed using the Illumina mRNA-seq paired-end protocol on a HiSeq2000 platform, at the INRA–UMR AGAP (Montpellier, France). We obtained 53.5 to 59.1 million paired-end reads for each sample (divided in two replicates) resulting in a total of 332.7 million paired-end reads.

Assembly and annotation

A *de novo* reference transcriptome was constructed for each of our target species using Trinity r2013_08_14 (Grabherr et al. 2011) with options *–group_pairs_distance 500* and *–min_kmer_cov 2*. All contigs were aligned to *D. buzzatii* genome (Guillén et al. 2015) using BLAT v.35x1 (Kent 2002), with parameters *–minIdentity=80* and *–maxIntron=75000*, in order to identify chimers. Contigs that aligned partially (≤60%) on up to 3 genomic locations with a total alignment coverage of ≥80% were considered chimeric and split consequently.

Finally, to annotate protein-coding genes, all contigs of both transcriptomes were aligned against the *D. buzzatii* predicted gene models and the *D. buzzatii* genome (Guillén et al. 2015) using BLAT v.35x1 (same parameters as before). This approach allows us to identify untranslated regions and double check the genomic position associated to a contig. Only contigs with alignment coverages ≥70% and whose best hit genomic coordinates overlapped in both alignments were annotated. The same approach was applied to the remaining non annotated contigs with *D. mojavensis*' gene models. The remaining contigs were clustered using CD-HIT v4.5.4 (Fu et al. 2012) with options *-c 0.8*, -T 0, *-aS 0.8*, *-A 80*, *-p 1*, *-g 1*, *-d 50*; and annotated with the name of the longest sequence of each cluster. **Table 3** depicts a summary of annotation statistics.

TE library construction

Our library is mainly constituted by the list of all TE copies masked in the *D. buzzatii* genome (Guillén et al. 2015; Rius et al. 2016), because *D. koepferae* has not until now been sequenced. In order to have a better representation of *D. koepferae* TE landscape and increase specificity in further analyses, we annotated TE transcripts from our *de novo* assemblies by aligning them to a consensus TE library (the same used to mask *D. buzzatii* genome) using BLAT v.35x1. Contigs whose alignments covered ≥80% of their sequences with a minimum 80% identity and ≥80 bp long (“the three 80 criteria”) were kept as TE transcripts and included in our TE library. To improve our coverage and sensitivity to detect poorly expressed TEs, a third *de novo* assembly, using all the reads from all sequenced samples (from both parents and hybrids) was performed and annotated as described above.

This resulted in 65,772 final TE copies belonging to 699 TE families, which were assigned to only 658 families after two steps of clustering. Clustering was performed using the three 80 criteria; manually through BLAT alignments, and automatically using CD-HIT v4.5.4 (same parameters as in gene annotation). These 658 families were divided in 5 categories, following Repbase classification (Jurka et al. 2005): LTR and LINE (class I), DNA and RC (class II) and Unknown (unclassified).

Small RNA extraction, library preparation and sequencing

Small RNA was purified from ovaries (n=70 pairs for all samples) and testes (n=96 pairs for *D. buzzatii* and n=333 pairs for F1 sterile males), following the manual small RNA purifying protocol described by Grentzinger et al. (2013), which significantly reduces endogenous contamination and degradation products abundance. After small RNA isolation, samples were gel-purified and precipitated. A single Illumina library was prepared for each sample and sequencing was performed on an Illumina Hiseq 2500 platform by FASTERIS SA (Switzerland). We obtained a total of 401.1 million reads (21.4 to 58.7 million reads per sample). Reads of 23-30 nucleotides were kept as piRNAs.

TE analyses: read mapping and differential expression

All our sequencing data was trimmed using UrQt (Modolo and Lerat 2015), in order to remove polyA tails (for RNA-seq) and low-quality nucleotides (for both RNA-seq and piRNA-seq). The resulting trimmed reads were aligned to our TE library using Bowtie2 v2.2.4 for RNA-seq (Langmead and Salzberg 2012) and Bowtie1 v1.1.1 for piRNAs (Langmead et al. 2009), with the default options implemented in TEtools pipeline (the most sensitive option and keeping a single alignment for reads mapping to multiple positions, *--very-sensitive* for Bowtie2 and *-S* for Bowtie). The read count step (built in TE tools: https://github.com/l-modolo/TEtools) was computed per TE family (adding all reads mapped on copies of the same family). Finally, we performed the differential expression analyses between TE families using the R Bioconductor package DESeq2 (Love et al. 2014) on the raw read counts, using the Benjamini-Hochberg multiple test correction (FDR level of 0.1). Statistical summaries of these analyses are available in File S1 and S5 (see **Annexes 8.3.2 and 8.3.6**), including both raw and normalized read count tables. TE families with ≤10 aligned reads per sample are considered to be unexpressed in the text. For piRNA analyses, no significant differences could be detected at the TE family level due to the lack of replicates, leading us to perform the analyses using FC values.

Gene analyses: read mapping, differential expression and GO enrichment

Gene expression analyses were performed following the same approach used for TEs. RNA-seq reads were aligned against the addition of *D. buzzatii* and *D. koepferae* transcriptomes, and read count was computed per annotated gene (by adding all reads mapped on contigs with the same annotation).

Trinity's tool TransDecoder (Haas et al. 2013) was employed to predict ORFs within *D. buzzatii* and *D. koepferae* transcriptomes, using Pfam-A database v.29 (Punta et al. 2012). Then, we performed a functional annotation of the resulting proteomes using GO terms (The Gene Ontology Consortium 2000). For that, we used eggnog-mapper tool (https://github.com/jhcepas/eggnog-mapper): we first mapped our sequences to eggNOG orthologous groups from eukaryotic, bacterial and archaeal databases (Huerta-Cepas et al. 2016) using an e-value of 0.001. Then, we transferred the GO terms of the best orthologous group hit for each gene. GO enrichments for deregulated genes in hybrids were analysed using the Topology-Weighted method built in Ontologizer (Bauer et al. 2008), with a p-value threshold of 0.01.

Divergence time and TE landscapes of parental species

In order to identify contig pairs between *D. buzzatii* and *D. koepferae*, all sequences ≥2000 bp of the *D. buzzatii de novo* transcriptome were aligned against *D. koepferae* using BLAST (McGinnis and Madden 2004). We kept only the best hit for each query and subject, resulting in a total of 2,656 contig pairs, which were translated using EMBOSS getorf (Rice et al. 2000). We used the most likely protein sequences of each contig pair (*i.e.* the longer) to perform codon alignments with MUSCLE (Edgar 2004). Finally, the dS rate of each pair was calculated using the codeml program in PAML version 4 (Yang 2007). Divergence time was estimated as in Keightley et al. (2014) using the obtained dS mode.

We examined the repeatomes of *D. buzzatii* and *D. koepferae* using dnaPipeTE pipeline (Goubert et al. 2015), which assembles repeats from low coverage genomic NGS data and annotates them with RepeatMasker Open-4.0 (Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996–2010, http://www.repeatmasker.org, last accessed February 24, 2016) and Tandem repeats finder (Benson 1999). We employed Repbase library version 2014-01-31 (Jurka et al. 2005). For both species, two iterations were performed using a read sample size corresponding to a genome coverage of 0.25X (Guillén et al. 2015; De Panis and Hasson, unpublished), according to genome size estimates in Romero-Soriano et al. (2016). Because mitochondrial DNA is usually assembled, all dnaPipeTE contigs were aligned to BLAST nucleotide collection (McGinnis and Madden 2004) to distinguish

nuclear from mitochondrial sequences. Reads mapping to mitochondrial contigs were identified using Bowtie2 with default parameters (Langmead and Salzberg 2012) and filtered out. DnaPipeTE was then run without mithocondrial reads (same parameters).

Ping-pong signature identification

The ping-pong cycle is mediated by Aub and Ago3 proteins, which cleave the piRNA precursor (or TE transcript) preferentially 10 bp after its 5' end. Thus, sense and antisense reads overlapped by 10 nucleotides are produced during secondary piRNA biogenesis (Klattenhoff and Theurkauf 2008). We aligned our piRNA raw reads (23-30nt, without any trimming step in order to maintain their real size) against the whole TE library using Bowtie1 (*-S* option, Langmead et al. 2009) and checked for the presence of 10nt-overlapping sense-antisense read pairs using signature.py pipeline (Antoniewski 2014). The same analysis was carried out separately for each of the TE families of the library.

piRNA pathway proteins ortholog search

Proteomes of *D. buzzatii* and *D. koepferae* (see **above**) were aligned against each other using BLAST. Identity percentages of each protein best hit were kept and used to calculate the median identity percentage between *D. buzzatii* and *D. koepferae*.

We identified the orthologs of 30 proteins involved in piRNA biogenesis (Yang and Pillai 2014) in *D. buzzatii* and *D. koepferae* proteomes by reciprocal best blast hit analysis, using their *D. melanogaster* counterparts as seeds (EnsemblMetazoa 27 release, Cunningham et al. 2015), with and e-value cutoff of 1e-05. *D. buzzatii* proteins were aligned against their *D. koepferae* ortholog using BLAST, in order to evaluate their identity percentage.

3.3.1.7 Acknowledgements

The authors wish to thank Nuria Rius and Alfredo Ruiz for providing advanced access to the *D. buzzatii* genome and its TE list; Xavier Grau-Bové for his useful help with orthology methods; Esteban Hasson and Diego Nicolás de Panis for sharing sequencing data of *D. koepferae* genome; Clément Goubert for his valuable help with dnaPipeTE; and INRA – UMR AGAP (France) and Fasteris SA (Switzerland) for RNA sequencing services.

This work was supported by the Ministerio de Economía y Competitividad (Spain, grant CGL2013-42432-P); Generalitat de Catalunya (grant 2014 SGR 1346 to the Grup de Genòmica, Bioinformàtica i Biologia Evolutiva); Centre National de la Recherche Scientifique (France,

UPR1142); Fondation pour la Recherche Médicale (France, FRM:DEP20131128518); Agence Nationale de la Recherce (France, grant Exhyb ANR-14-CE19-0016-01 to CV) and Institut Universitaire de France (grant to CV). VRS was supported by a PIF PhD fellowship from the Universitat Autònoma de Barcelona (Catalunya, Spain) and by a CMIRA Accueil Doc travel grant from Région Rhône-Alpes (France).

Sequence data from this article will be submitted to Sequence Read Archive (SRA) before acceptation.

4 Discussion

This thesis intends to shed light onto the evolutionary consequences and causal mechanisms of TE activation in *D. buzzatii-D. koepferae* hybrids. In **chapter 3.1**, “Drosophila females undergo genome expansion after interspecific hybridization”, we investigate the impact of interspecific hybridization on genome size, a feature that is likely affected by TE proliferation. In **chapters 3.2** and **3.3**, we assess the hypothesis that transposition release is due to a TE transcription deregulation by using two different approaches.

In the research article “Expression of the retrotransposon Helena reveals a complex pattern of TE deregulation in Drosophila hybrids”, we perform an exhaustive analysis of *Helena* expression dynamics through four hybrid generations. The focus on a single transposon allows an in-depth dissection of its expression in different tissues and through different techniques, providing detailed information about *Helena* transcripts’ localization and abundance.

Another strategy is employed in the third chapter, “Divergence in piRNA pathway effector proteins partially explains Drosophila buzzatii-D. koepferae hybrid instability”, where we perform a transcriptomic analysis of genome-wide TEs. This survey supplies a broadened vision of global TE expression and regulation in hybrids, but the technical approaches used here have an ineluctably lower sensitivity than the former.

In each of these three chapters (**3.1, 3.2** and **3.3**), the obtained results are thoroughly discussed. The purpose of the present section is to review and unify the three independent discussions, in order to mold and consolidate a global message comprehending the scientific contributions of this thesis.

TE proliferation is known to occur in hybrids of a wide range of species, including plants (Liu and Wendel 2000; Shan et al. 2005; Ungerer et al. 2006; Wu et al. 2014) and animals (Evgen'ev et al. 1982; O'Neill et al. 1998; Labrador et al. 1999; Metcalfe et al. 2007). A recent study in our lab detected mobilization of 28 different TE families in hybrids between the sibling species *D. buzzatii* and *D. koepferae* (Vela et al. 2014). This species pair is particularly interesting from an evolutionary viewpoint, since reticulation events between *D. buzzatii* and *D. koepferae* seem to have occurred in nature (Gomez and Hasson 2003; Piccinali et al. 2004; Franco et al. 2010). Thus, our model allows the investigation of *real* hybrids which may have sculpted the evolutionary history of their parental species.

4.1 On hybrid genome size increase

4.1.1 An evolutionary role for TEs in shaping hybrid genome size and structure

In **chapter 3.1**, we describe a genome expansion in our hybrid females, associated with the interspecific hybridization process (**section 3.1.1** Figure 2). This genome size increase is not the only feature that characterizes *D. buzzatii–D. koepferae* hybrids, since an increase of inversions and duplications (among other chromosomal reorganizations) was described three decades ago (Naveira and Fontdevila 1985). Interestingly, chromosomal rearrangements and other karyotypic alterations can contribute to the origin of new species (Brown and O'Neill 2010; Nevo 2012), a phenomenon that has been described in both plants (Ramsey and Schemske 1998) and animals (Nevo 2012) and is called chromosomal speciation. Chromosomal rearrangements in heterozygosis can cause meiotic defects, producing unbalanced gametes that are often unviable. Those gametes present low recombination rates due to karyotypic differences between parents, which lead to the creation of linkage groups allowing the maintenance of *speciation* genes (Noor et al. 2001; Rieseberg 2001). Therefore, changes in genome architecture may act as reproductive isolation barriers, ultimately contributing to hybrid speciation events.

Interestingly, all the aforementioned genetic instabilities of *D. buzzatii–D. koepferae* hybrids can be triggered by transposons. In this regard, several chromosomal inversions in *D. buzzatii* are known to be caused by TEs (Cáceres et al. 1999; Casals et al. 2003; Delprat et al. 2009). Moreover, TE mobilization can produce (often intrinsically) duplication events, especially in the case of retrotransposons (see **section 1.3.1.2**). Finally, proliferation of TEs is one of the most important

mechanisms leading to genome size increase (Gregory 2005b). Indeed, waves of transpositional activity can rapidly cause significant genome expansions, such as the doubling of the maize genome size, that occurred in only a few million years (SanMiguel et al. 1996). In conclusion, TEs may be responsible for the rapid karyotypic evolution of hybrids between *D. buzzatii* and *D. koepferae*, emphasizing their role as an evolutionary force (see **section 1.3.2.2**).

4.1.2 Transposition events account for a sex-biased genome expansion

Eukaryotic genome size is strongly related to TE abundance (Kidwell 2002). Hence, the TE release described in *D. buzzatii–D. koepferae* hybrids (Labrador et al. 1999; Vela et al. 2014) could likely be at the origin of the female genome expansion described in this thesis. In **chapter 3.1**, we contemplate other explanations that could also account for this phenomenon. In particular, we assess a possible transmission bias towards the larger parental genome, *D. koepferae*, using AFLP markers. However, our results rule out this hypothesis, since we show that a putative transmission bias (if any) would favor the genome of *D. buzzatii*. We also reject a possible role of polyploidization in genome size increase, because not a single case of polyploid karyotype has been reported in the many cytological studies concerning our species' hybrids (Naveira and Fontdevila 1985; Marín and Fontdevila 1998; Labrador et al. 1999). Finally, we recognize proliferation of satellite DNA (and other repetitive sequences) as a possible causative agent of genomic expansion, as shown in *D. orena* genome (Boulesteix et al. 2006). This last conjecture cannot be dismissed but lacks of experimental verification in our hybrids.

Genome size has been assessed throughout four hybrid generations, including the F1 and three sequential backcrosses of hybrid females with *D. buzzatii* males (**section 3.1.1** Figure 1). In the first hybrid generation, female genome size estimates are not significantly different from the theoretical value (*i.e.* a weighted parental mean) (**section 3.1.1** Figure 2). A clear genome expansion occurs in BC1, where not only C-value estimates are significantly greater than the theoretical value, but their mean also increases compared to the previous generation. After BC1, genome size estimates remain higher than the theoretical mean, but their value actually decreases all over generations. This decrease is probably due to the introgression of the smaller parental genome, *D. buzzatii;* along with a better control of TE mobilization. Curiously, no genome size increase is observed in any generation of hybrid males (**section 3.1.1** Figure 2), meaning that TE release in our hybrids is sex-specific and has no impact on male genome size.

4.1.3 TE mobilization occurs in F1 female gametes

According to the pattern of genome size variation described in section 3.1.1 (Figure 2), novel TE insertions in females are detectable only in backcrossed hybrids, but not in F1. This scenario implies that transposition events have to take place before BC1; that is, in gametes of F1. Thus, a failure of TE silencing mechanisms in F1 ovaries would be at the origin of the transposition release (and genome expansion) observed in our hybrids. Concordantly, an enhanced TE expression was described in F1 ovaries of *D. melanogaster–D. simulans* hybrids, as well as in hybrids of lake whitefishes and sunflowers (Kelleher et al. 2012; Dion-Côté et al. 2014; Renaut et al. 2014). In *Drosophila* ovaries, TEs are mainly silenced by piRNAs at both transcriptional and post-transcriptional levels (see **section 1.3.2.3**). Our genome size survey suggests that an impairment of piRNA pathway's function in F1 would cause an increase in TE copy number in subsequent generations, as observed in *D. buzzatii–D. koepferae* backcrossed hybrids (Vela et al. 2014). Studies on TE transcription and its regulation may be the key to understand the causal mechanisms involved in hybrid TE release.

These findings have guided us to the analysis of TE expression in hybrids, a step that precedes TE mobilization and that is essential for retrotransposon transposition. To start, we have studied one of the mobilized elements in hybrids between *D. buzzatii* and *D. koepferae*, the LINE retrotransposon *Helena* (Evgen'ev et al. 1997), which presented higher transcription rates in hybrids than in both parental species (Vela et al. 2014). An analogous survey reporting the expression patterns in hybrids and parental species of another mobilized retrotransposon, *Osvaldo*, was previously carried out (García Guerreiro 2015).

4.2 On *Helena* retrotransposon expression

In **chapter 3.2**, the expression of *Helena* is assessed in carcasses (somatic tissues) and gonads of *D. buzzatii*, *D. koepferae* and their hybrids. Given that the effects of interspecific hybridization on TE activation and genome expansion are sex-biased (demonstrated in **chapter 3.1**), males and females are studied separately.

4.2.1 Hybridization effects on somatic expression are specific to each individual and depend on the studied TE

Results in carcasses show that *Helena* expression remains globally unchanged in somatic tissues of both sexes after specific hybridization. However, a few occasional transcription bursts seem to occur in male samples (see **section 3.2.1** Figure 3A), leading sometimes to a variance increase in hybrid expression rates. In the case of *Osvaldo*, expression in male carcasses is globally higher in hybrids than in parents, while in females it presents additive values (García Guerreiro 2015). The high transcript abundances of *Helena* in some samples may be the consequence of a failure in the endo-siRNA pathway affecting only some hybrid flies. This emphasizes the uniqueness of the genomic background of backcrossed hybrids, which depends on the introgressed fragments of each individual. However, the presence of one of these transcription bursts in F1 is surprising, since hybrids are expected to be very similar at this generation (given that the parental strains used are highly inbred). Although I cannot provide a compelling molecular explanation for the variability observed in F1, it is noteworthy that *Osvaldo* expression in F1 also differs between replicates (García Guerreiro 2015). Furthermore, in a study measuring larval viability and developmental time in the same hybrids, different hybrid phenotypic categories were also observed (Soto et al. 2008).

4.2.2 Hybridization effects on gonads are sex-biased and TE-dependent

Helena expression in testes decreases significantly in F1 compared parental species, and is restored to parental levels after the first backcross (see **section 3.2.1** Figure 3C). Besides, *Helena* transcripts are mislocalized in hybrids (see **section 3.2.1** Figure 4). Interestingly, the lower *Helena* levels found in F1 hybrids can be explained by a striking increase of *Helena*-associated piRNA populations (see

section 3.2.1 Figure 6). Since the fraction of secondary piRNAs (ping-pong signature level) is maintained after hybridization, we hypothesize that the higher piRNA levels are not mainly due to an intensified activity of the ping-pong cycle, but to a more efficient phasing or primary piRNA biogenesis (see **section 1.3.2.3**). The activation of piRNA production in male gonads is probably a defense mechanisms that counteracts TE activation. A tighter control of TE activity could explain the negligible effects of interspecific hybridization on male genome size (see **section 3.1.1**), but results of *Osvaldo* expression reveal that reality is far more complex. While *Helena* transcription in hybrids is repressed or similar to parental levels; *Osvaldo* has a trend towards overexpression (especially in BC1 and BC2, García Guerreiro 2015). In conclusion, we can state that *Helena* is down-regulated in F1 hybrids, but we must note that the impact of interspecific hybridization on testicular TE expression depends on the studied TE.

In ovaries, *Helena* expression significantly differs between parental species, and hybrids globally present intermediate values across all generations (see **section 3.2.1** Figure 3D). However, our FISH results suggest that an increase of TE expression may occur at first in young flies, eventually leading to TE mobilization (see **section 3.2.1** Figure 5C and D). Afterwards, an implementation of TE silencing mechanisms could take place in older flies, reducing *Helena* transcript abundances (see **section 3.2.1** Figure 3D). This outline is concordant with our piRNA population analyses, where we observe that piRNA amounts are roughly maintained between hybrids and parents, although the fraction of secondary piRNAs decreases in F1 (see **section 3.2.1** Figure 6). Hence, a failure of the ping-pong cycle would cause an early TE derepression, subsequently hindered by the activation of other piRNA production strategies, like phasing and primary piRNA biogenesis. *Osvaldo* expression results, which lack of regulation analyses, do not show any transcription alteration in hybrid ovaries compared to parental species (García Guerreiro 2015). Therefore, interspecific hybridization affects ovarian TE expression in a TE-dependent way that differs from the described in testes. This demonstrates that alteration of TE expression is sex-biased, as predicted by our genome size results (see **section 3.1.1**).

The striking differences in expression dynamics between two of the mobilized transposons in *D. buzzatii–D. koepferae* hybrids reveal the existence of a complex TE misregulation pattern, involving both over and underexpression, as well as sex-specificity. Thus, the study described in **chapter 3.2** raises several questions on the nature of hybrid TE activation, particularly regarding the

extent and the direction of TE deregulation. Since performing equivalent studies on other TEs requires in most cases their molecular characterization, this strategy would be considerably time-consuming and still would not guarantee a comprehensive advance in the understanding of the molecular mechanisms underlying hybrid TE deregulation. Thus, we chose to perform a transcriptomic analysis of global TE expression. Although less meticulous, the scope of this survey grants a widened insight on hybrid TE deregulation.

4.3 On global TE deregulation

The last Results chapter (**3.3**) describes a genome-wide analysis of TE expression and regulation using high-throughput sequencing technologies. We carry out this study in gonads, since the most striking results of *Helena* and *Osvaldo* expression surveys occur in ovaries and testes. In the case of ovaries, two hybrid generations are analysed (F1 and BC1), whereas only the F1 is studied in testes (**Figure 5**).

4.3.1 TE deregulation is sex-biased and diminish across generations

Global TE expression patterns differ between ovaries and testes, both quantitatively (higher expression in ovaries) and qualitatively (*e.g.* lower LINE expression in testes, **Figure 6**). At the TE family level, the majority of the families are more expressed in testes than in ovaries (Table S1 in **Annex 8.3.9**). This tissue-specificity indicates that TE regulation is not as efficient in testes as in ovaries. Accordingly, we observe lower piRNA amounts in testes than in ovaries (**Figure 12**). These results concur with previous results showing that piRNAs in testes are not only involved in TE silencing, but also in gene silencing (Nishida et al. 2007). Assuming hence that TE regulation is sex-biased, it is not surprising that hybrid TE deregulation tendencies differ between males and females. Hybrid ovaries exhibit a bias towards TE overexpression, whereas underexpression prevails in hybrid testes (Table S1 in **Annex 8.3.9**). In the same way, a widespread overexpression of TEs was reported in ovaries of *Hmr*-rescued *D. melanogaster-D. simulans* hybrids (Kelleher et al. 2012). On the other hand, the repression of *Helena* expression in hybrid testes described in **chapter 3.2** is in concordance with the tendency observed in F1 testes. It is worth noting that, despite their respective biases, both tissues present cases of TE over and underexpression, as observed for *Osvaldo* (upregulated) and *Helena* (downregulated) in hybrid testes (García Guerreiro 2015).

In F1 ovaries, 15.2% of the expressed TEs are deregulated, a proportion that decreases to 10.6% in BC1. After a generation of backcrossing, incompatibilities in TE silencing seem to be partially mitigated and TE control is recovered in almost a third of the F1-deregulated TEs. Although we cannot estimate the number of generations necessary to restore parental TE silencing levels, our genome size results suggest that the TE copy number does not increase after BC1, when the mean

genome size starts to decrease (see **section 3.1.1** Figure 2). It is also noteworthy fertility recovery in males occurs for some individuals in the third backcross (BC3) (Morán and Fontdevila 2014).

4.3.2 Several mechanisms are at the origin of TE deregulation

In **chapter 3.3**, we show that the tendency towards TE underexpression in hybrid testes could likely be explained by an increase of piRNA production. F1 testes present larger piRNA populations with stronger ping-pong signal than *D. buzzatii*, suggesting that an enhancement of the secondary piRNA biogenesis is at the origin of TE repression (**Figure 12**). However, in the particular case of *Helena* (**chapter 3.2**), an increased activity of phased and primary piRNA biogenesis pathways seem to be responsible for this retrotransposon repression (see **section 3.2.1**). Therefore, the intensification of piRNA production may involve primary, secondary and phased piRNA production. Again, this concurs with the imperceptible effects of hybridization in male genome size (see **section 3.1.1**).

In ovaries, our analyses are more complete and allow a more in-depth description of hybrid TE expression and deregulation. Two hypotheses are tested, the maternal cytotype hypothesis and the global piRNA pathway breakdown hypothesis (see **section 3.3.1.2**). We show that the first hypothesis can only account for a small number of deregulation cases, in which either lower amounts of piRNAs in the cytoplasm of the maternal species (*D. koepferae*) compared to the paternal (*D. buzzatii*) lead to overexpression, or higher amounts of piRNAs in *D. koepferae* compared to *D. buzzatii* lead to underexpression (**Figure 9**). On the other hand, we demonstrate that piRNA pathway proteins tend to be more divergent than the average between our parental species (**Figure 11**), which could lead to a dysfunction of this pathway in hybrids, concordant with the second hypothesis scenario. Furthermore, high differences in expression of several piRNA pathway proteins between parental species (**Table 5**) could account for a functional divergence in piRNA production strategies. Actually, *D. buzzatii* has a higher proportion of secondary piRNAs but lower global piRNA levels than *D. koepferae* (**Figure 10**). Apparently, phased piRNA production in *D. buzzatii* is less efficient than in *D. koepferae;* and a higher frequency of homotypic Aub:Aub ping-pong cycles in *D. buzzatii* could lessen the levels of antisense piRNAs in this species.

We hypothesize that the accumulated divergence in piRNA pathway proteins leads to a dysfunctional silencing in hybrids and is the main cause of TE deregulation. However, it is noteworthy that, contrary to *D. simulans-D. melanogaster Hmr*-rescued hybrids (Kelleher et al. 2012), a deficiency in piRNA production is not observed in our hybrids (**Figure 10**). Otherwise, our hybrids present abnormal expression of *SoYb*, *Hen1* and *Panoramix*, three of the proteins involved in piRNA-mediated silencing (**Table 5**). Mutants of *SoYb* and *Hen1*, both underexpressed in our

hybrids (**Table 5**), do not exhibit extreme phenotypes (regarding neither TE expression nor fertility) in *D. melanogaster* (Horwich et al. 2007; Saito et al. 2007; Handler et al. 2011). Therefore, their decrease in expression can only partially account for TE deregulation. Overexpression of *Panoramix*, involved in TE transcriptional silencing (Czech et al. 2013; Handler et al. 2013; Sienski et al. 2015; Yu et al. 2015) can contribute to the underexpression observed in some cases.

Finally, we must note that other mechanisms could be involved in TE deregulation in hybrids between in *D. buzzatii* and *D. koepferae*, since some of the overexpressed TE families are devoid of piRNA populations in parental species and hybrids (**Figure 9B-iv**). For example, an impairment of endo-siRNA-mediated silencing could lead to the deregulation of some TE families, particularly those of somatic expression (Nilsen 2008). Internal deletions and recombination-mediated complete deletions have been proposed as a defence mechanism against TE proliferation (Petrov and Hartl 1998; Lerat et al. 2011) and could also play a role in TE underexpression. Last but not least, an abnormal recruitment of permissive/repressive hystone modification marks at the TE genomic insertions could also be at the origin of aberrant TE expression.

5 Conclusions

(1) In both parental species, *D. buzzatii* and *D. koepferae*, females present significantly larger genomes than males (differences of 7Mb). Furthermore, variability in genome size is higher in males than in females.

(2) In hybrid females, the genome size estimates are higher than the theoretical value (*i.e.* a weighted mean between parentals species) for the three generations of backcrosses. However, there is no evidence of any impact on hybrid males genome size due to interspecific hybridization. Therefore, the effects of hybridization on genome size are sex-specific.

(3) A putative transmission bias in backcrossed hybrids would favour *D. buzzatii* genome (smaller than *D. koepferae*).

(4) A genome expansion in hybrid females takes place between F1 and BC1, most probably due to transposable element mobilization.

(5) All the *Helena* insertions characterized in parental species present two overlapping ORFs and three conserved domains (PRE_C2HC in ORF1; EEP, and RT in ORF2).

(6) *D. buzzatii* has a higher *Helena* copy number than *D. koepferae*. However, only the *D. buzzatii* sequenced copy is putatively active.

(7) Internal deletions may act as a regulatory mechanism preventing *Helena* proliferation.

(8) In ovaries of parental species ovaries, *Helena* expression is higher in *D. buzzatii* than in *D. koepferae*.

(9) *Helena* expression in somatic tissues remains globally unchanged after interspecific hybridization, with the exception of a few transcription bursts in males.

(10) F1 testes present lower *Helena* transcript abundance than both parental species, likely due to an enhanced production of its associated piRNAs. However, hybrid ovaries present intermediate *Helena* expression levels, although a hypothetical transcription burst might occur in young females. Hence, the effects of hybridization on *Helena* quantitative expression are sex-specific.

(11) Localization of *Helena* transcripts is altered in both ovaries and testes after interspecific hybridization.

(12) In gonads of hybrids and parental species, global TE expression is mainly represented by retrotransposons, especially LINEs.

(13) TE landscapes are extremely similar between *D. buzzatii* and *D. koepferae* genomes.

(14) Global TE expression, piRNA production and TE deregulation patterns are sex-biased, because:

(a) TE expression is higher in testes than in ovaries for the majority of the families.

(b) Global piRNA levels are lower in testes than in ovaries.

(c) After hybridization, TE overexpression is more common in ovaries and TE underexpression in testes.

(15) In ovaries, TE and gene expression are affected at similar levels by interspecific hybridization. After a generation of backcrossing, ovarian TE deregulation decreases from 15.2 to 10.6% of the expressed TEs. In the cases of genes, it decreases from 13.2 to 12.3% of the annotated genes in our *de novo* assembly.

(16) Differences in piRNA pools between *D. buzzatii* and *D. koepferae* account for a small portion of TE deregulation.

(17) piRNA production differs between *D. buzzatii* and *D. koepferae* ovaries: global piRNA amounts, ping-pong fraction and expression of piRNA pathway genes are strikingly different between these species.

(18) Most of the piRNA pathway proteins are among the most divergent between *D. buzzatii* and *D. koepferae* proteomes.

(19) The functional divergence in the piRNA pathway between parental species may be one of the main causes of TE deregulation.

(20) Other explanations are needed to explain the whole ovarian TE deregulation pattern, such as a failure of the endo-siRNA pathway or of the deposition of histone modifications.

(21) In hybrid testes, TE repression can be explained by an enhanced piRNA production compared to *D. buzzatii*.

6 Bibliography

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, et al. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, et al. 2004. Complete Genome Sequence of the Apicomplexan, Cryptosporidium parvum. Science 304:441–445.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The Genome Sequence of Drosophila melanogaster. Science 287:2185–2195.

Alonso-Pimentel H, Tolbert LP, Heed WB. 1994. Ultrastructural examination of the insemination reaction in Drosophila. Cell Tissue Res. 275:467–479.

Alzohairy AM, Gyulai G, Jansen RK, Bahieldin A. 2013. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid 69:1–15.

Amaral AR, Gretchen L, Maria M C, George A, Howard C R. 2014. Hybrid speciation in a marine mammal: The clymene dolphin (Stenella clymene). PLoS One 9:1–8.

Aminetzach YT, Macpherson JM, Petrov DA. 2005. Pesticide Resistance via Transposition-Mediated Adaptive Gene Truncation in Drosophila. Science 309:764–767.

Anderson E, Stebbins GL. 1954. Hybridization as an Evolutionary Stimulus. Evolution (N. Y). 8:378–388.

Anderson TM, VonHoldt BM, Candille SI, Musiani M, Greco C, Stahler DR, Smith DW, Padhukasahasram B, Randi E, Leonard JA, et al. 2009. Molecular and Evolutionary History of Melanism in North American Gray Wolves. Science 323:1339–1344.

Antoniewski C. 2014. Computing siRNA and piRNA Overlap Signature. In: Werner A, editor. Animal Endo-SiRNAs: Methods and Protocols. Springer New York. p. 135–146.

Anxolabéhère D, Kidwell MG, Periquet G. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. Mol. Biol. Evol. 5:252–269.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442:203–207.

Aravin AA, Hannon GJ, Brennecke J. 2008. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. Science 761.

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T. 2003. The Small RNA Profile during Drosophila melanogaster Development. Dev. Cell 5:337–350.

Arnold ML. 1997. Natural hybridization and evolution. New York: Oxford University Press

Asada N, Kitagawa O. 1988. Insemination reaction in the Drosophila nasuta subgroup. Japanese J. Genet. 63:137–148.

Aubert J, Solignac M. 1990. Experimental evidence for mitochondrial DNA introgression between Drosophila species. Evolution (N. Y). 44:1272–1282.

Baack EJ, Whitney KD, Rieseberg LH. 2005. Hybridization and genome size evolution: Timing and magnitude of nuclear DNA content increases in Helianthus homoploid hybrid species. New Phytol. 167:623–630.

Ballard JWO. 2000. When One Is Not Enough: Introgression of Mitochondrial DNA in Drosophila. Mol. Biol. Evol. 17:1126–1130.

Barbash DA. 2010. Ninety years of Drosophila melanogaster hybrids. Genetics 186:1–8.

Bartolomé C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. Genome Biol. 10:R22.

Bartolomé C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of Drosophila melanogaster. Mol. Biol. Evol. 19:926–937.

Bauer S, Grossmann S, Vingron M, Robinson PN. 2008. Ontologizer 2.0 - A multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics 24:1650–1651.

Benson G. 1999. Tandem Repeats Finder: a program to analyse DNA sequences. Nucleic Acids Res. 27:573–578.

Biémont C, Vieira C. 2006. Junk DNA as an evolutionary force. Nat. Genet. 443:521–524.

Blaise S, de Parseval N, Benit L, Heidmann T. 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc. Natl. Acad. Sci. U. S. A. 100:13013–13018.

Bock IR. 1984. Interspecific hybridization in the genus Drosophila. Evol. Biol. 18:41–70.

de Boer JG, Yazawa R, Davidson WS, Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 8:422.

Borie N, Maisonhaute C, Sarrazin S, Loevenbruck C, Biémont C. 2002. Tissue-specificity of 412 retrotransposon expression in Drosophila simulans and D. melanogaster. Heredity (Edinb). 89:247–252.

Boulesteix M, Weiss M, Biémont C. 2006. Differences in genome size between closely related species: The Drosophila melanogaster species subgroup. Mol. Biol. Evol. 23:162–167.

Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, Bidwell SL, Brown WC, Crabtree J, Fadrosh D, et al. 2007. Genome sequence of Babesia bovis and comparative analysis of apicomplexan hemoprotozoa. PLoS Pathog. 3:1401–1413.

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

Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. Science 322:1387–1392.

Brennecke J, Senti K-A. 2010. The piRNA pathway : a fly's perspective on the guardian of the genome. Trends Genet. 26:499–509.

Bringaud F, Ghedin E, Blandin G, Bartholomeu DC, Caler E, Levin MJ, Baltz T, El-Sayed NM. 2006. Evolution of non-LTR retrotransposons in the trypanosomatid genomes: Leishmania major has lost the active elements. Mol. Biochem. Parasitol. 145:158–170.

Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SCR, Lin H. 2007. Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev. 21:2300–2311.

Brown JD, O'Neill RJ. 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. Annu. Rev. Genomics Hum. Genet. 11:291–316.

Cáceres M, Ranz JM, Barbadilla A, Long M, Ruiz A. 1999. Generation of a Widespread Drosophila Inversion by a Transposable Element. Science 285:415–418.

Callinan PA, Batzer MA. 2006. Retrotransposable elements and human disease. Genome Dyn. 1:104–115.

Camillo J, Leão AP, Alves AA, Formighieri EF, Azevedo ALS, Nunes JD, de Capdeville G, de A. Mattos JK, Souza Jr MT. 2014. Reassessment of the Genome Size in Elaeis guineensis and Elaeis oleifera, and Its Interspecific Hybrid. Genomics Insights 7:13–22.

Carlton JM, Angiuoli S V, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, et al. 2002. Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature 419:512–519.

Carnelossi EAG, Lerat E, Henri H, Martinez S, Carareto CMA, Vieira C. 2014. Specific activation of an I-like element in Drosophila interspecific hybrids. Genome Biol. Evol. 6:1806–1817.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. Mol. Ecol.:1503–1517.

Casals F, Cáceres M, Ruiz A. 2003. The Foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of Drosophila buzzatii. Mol. Biol. Evol. 20:674–685.

Castel SE, Martienssen RA. 2013. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat. Rev. Genet. 14:100–112.

Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D, et al. 2010. The dynamic genome of Hydra. Nature 464:592–596.

Chen JM, Stenson PD, Cooper DN, Férec C. 2005. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum. Genet. 117:411–427.

Chen Z, Sturgill D, Qu J, Jiang H. 2014. Comparative validation of the D. melanogaster modENCODE transcriptome annotation. Genome Res. 24:1209–1223.

Chénais B. 2015. Transposable elements in cancer and other human diseases. Curr. Cancer Drug Targets 15:227–242.

Chiu JC, Jiang X, Zhao L, Hamm CA, Cridland JM, Saelao P, Hamby KA, Lee EK, Kwok RS, Zhang G, et al. 2013. Genome of Drosophila suzukii, the spotted wing drosophila. G3 3:2257–2271.

Chung WJ, Okamura K, Martin R, Lai EC. 2008. Endogenous RNA Interference Provides a Somatic Defense against Drosophila Transposons. Curr. Biol. 18:795–802.

Coyne JA, Orr HA. 2004. Speciation. Massachussets: Sinauer Associates

Coyne JA. 1993. The Genetics of an Isolating Mechanism between Two Sibling Species of Drosophila. Evolution (N. Y). 47:778–788.

Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature 508:411–415.

Cui X, Cao X. 2014. Epigenetic regulation and functional exaptation of transposable elements in higher plants. Curr. Opin. Plant Biol. 21:83–88.

Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2015. Ensembl 2015. Nucleic Acids Res. 43:D662–D669.

Cutter AD. 2008. Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. Mol. Biol. Evol. 25:778–786.

Czech B, Hannon GJ. 2016. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem. Sci. 41:324–337.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in Drosophila. Nature 453:798–802.

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol. Cell 50:749–761.

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for Horizontal Transmission of the P Transposable Element Between Drosophila Species. Genetics 124:339–355.

Delprat A, Negre B, Puig M, Ruiz A. 2009. The transposon Galileo generates natural chromosomal inversions in Drosophila by ectopic recombination. PLoS One 4:e7883.

Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq Reveals Transcriptomic Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species. Mol. Biol. Evol. 31:1188–1199.

Dobigny G, Ozouf-Costaz C, Waters PD, Bonillo C, Coutanceau JP, Volobouev V. 2005. LINE-1 amplification accompanies explosive genome repatterning in rodents. Chromosom. Res. 12:787–793.

Dobzhansky T. 1937. Genetics and the Origin of Species. New York: Columbia University Press

Dönertas D, Sienski G, Brennecke J. 2013. Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev. 27:1693–1705.

Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284.

Dowsett AP, Young MW. 1982. Differing levels of dispersed repetitive DNA among closely related species of Drosophila. Proc. Natl. Acad. Sci. U. S. A. 79:4570–4574.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203–218.

Dupressoir A, Marceau G, Vernochet C, Bénit L, Kanellopoulos C, Sapin V, Heidmann T. 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc. Natl. Acad. Sci. U. S. A. 102:725–730.

Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, Heidmann T. 2009. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. Proc. Natl. Acad. Sci. U. S. A. 106:12127–12132.

Dzur-Gejdosova M, Simecek P, Gregorova S, Bhattacharyya T, Forejt J. 2012. Dissecting the Genetic Architecture of F1 Hybrid Sterility in House Mice. Evolution (N. Y). 66:3321–3335.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eickbush TH, Burke WD, Eickbush DG, Lathe WC. 1997. Evolution of R1 and R2 in the rDNA units of the genus Drosophila. Genetica 100:49–61.

Eickbush TH, Eickbush DG. 2014. Integration, Regulation, and Long-Term Stability of R2 Retrotransposons. :1–20.

Engels WR, Preston CR. 1981. Identifying P factors in Drosophila by means of chromosome breakage hotspots. Cell 26:421–428.

Esnault C, Cornelis G, Heidmann O, Heidmann T. 2013. Differential Evolutionary Fate of an Ancestral Primate Endogenous Retrovirus Envelope Gene, the EnvV Syncytin, Captured for a Function in Placentation. PLoS Genet. 9:e1003400.

Evgen'ev MB, Yenikolopov GN, Peunova NI, Ilyin Y V. 1982. Transposition of Mobile Genetic Elements in Interspecific Hybrids of Drosophila. Chromosoma 85:375–386.

Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG. 1997. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis. Proc. Natl. Acad. Sci. U. S. A. 94:196–201.

Fablet M, Akkouche A, Braman V, Vieira C. 2014. Variable expression levels detected in the Drosophila effectors of piRNA biogenesis. Gene 537:149–153.

Fan P, Manoli DS, Ahmed OM, Chen Y, Agarwal N, Kwong S, Cai AG, Neitz J, Renslo A, Baker BS, et al. 2013. Genetic and Neural Mechanisms that Inhibit Drosophila from Mating with Other Species. Cell:1–14.

Fanara JJ, Fontdevila A, Hasson E. 1999. Oviposition preference, viability, developmental time and body size in the cactophilic sibling species Drosophila koepferae and D. buzzatii in association to their natural hosts. Evol. Ecol. 13:173–190.

Fanara JJ, Hasson E. 2001. Oviposition Acceptance and Fecundity Schedule in the Cactophilic Sibling Species Drosophila buzzatii and D. koepferae on Their Natural Hosts. Evolution (N. Y). 55:2615–2619.

Fanara JJ, Soto IM, Lipko P, Hasson E. 2016. First Record of Drosophila buzzatii (Patterson & Wheeler) (Diptera: Drosophilidae) Emerging from a Non-Cactus Host. Neotrop. Entomol.

Fedoroff N V. 2012. McClintock's challenge in the 21st century. Proc. Natl. Acad. Sci. U. S. A. 109:20200–20203.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9:397–405.

Findley SD, Tamanaha M, Clegg NJ, Ruohola-Baker H. 2003. Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development 130:859–871.

Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. Trends Genet. 5:103–107.

Fonseca NA, Morales-Hojas R, Reis M, Rocha H, Vieira CP, Nolte V, Schlötterer C, Vieira J. 2013. Drosophila americana as a model species for comparative studies on the molecular basis of phenotypic variation. Genome Biol. Evol. 5:661–679.

Fontdevila A, Pla C, Hasson E, Wasserman M, Sanchez A, Naveira H, Ruiz A. 1988. Drosophila koepferae: a new member of the Drosophila serido (Diptera: Drosophilidae) superspecies taxon. Ann. Entomol. Soc. Am. 81:380–385.

Fontdevila A, Ruiz A, Alonso G, Ocaña J. 1981. Evolutionary History of Drosophila buzzatii. I. Natural Chromosomal Polymorphism in Colonized Populations of the Old World. Evolution (N. Y). 35:148–157.

Fontdevila A. 2005. Hybrid genome evolution by transposition. Cytogenet. Genome Res. 110:49–55.

Fontdevila A. 2016. Hybrid Incompatibility in Drosophila: An Updated Genetic and Evolutionary Analysis. eLS:1–16.

Franco FF, Silva-Bernardi ECC, Sene FM, Hasson ER, Manfrin MH. 2010. Intra- and interspecific divergence in the nuclear sequences of the clock gene period in species of the Drosophila buzzatii cluster. J. Zool. Syst. Evol. Res. 48:322–331.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.

García Guerreiro M. 2012. What makes transposable elements move in the Drosophila genome? Heredity (Edinb). 108:461–468.

García Guerreiro MP. 2014. Interspecific hybridization as a genomic stressor inducing mobilization of transposable elements in Drosophila. Mob. Genet. Elements 4:e34394.

García Guerreiro MP. 2015. Changes of Osvaldo expression patterns in germline of male hybrids between the species Drosophila buzzatii and Drosophila koepferae. Mol. Genet. Genomics 290:1471–1483.

Gardner JPA. 1997. Hybridization in the sea. Adv. Mar. Biol. 3:1–78.

Gardner MJ, Bishop R, Shah T, de Villiers EP, Carlton JM, Hall N, Ren Q, Paulsen IT, Pain A, Berriman M, et al. 2005. Genome Sequence of Theileria parva, a Bovine Pathogen That Transforms Lymphocytes. Science 309:134–138.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, et al. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW, Zapp ML, Weng Z, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science 320:1077–1081.

Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C. 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1350.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442:199–202.

Gomez GA, Hasson E. 2003. Transpecific Polymorphisms in an Inversion Linked Esterase Locus in Drosophila buzzatii. Mol. Biol. Evol. 20:410–423.

González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. 2008. High rate of recent transposable element-induced adaptation in Drosophila melanogaster. PLoS Biol. 6:2109–2129.

Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. 2014. Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep. 15:411–418.

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

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Grant BR, Grant PR. 2008. Fission and fusion of Darwin's finches populations. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363:2821–2829.

Grant PR, Grant BR. 1992. Hybridization of Bird Species. Science 256:193–197.

Grant PR, Grant BR. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707–711.

Grant V. 1994. Modes and origins of mechanical and ethological isolation in angiosperms. Proc. Natl. Acad. Sci. U. S. A. 91:3–10.

Granzotto A, Lopes FR, Lerat E, Vieira C, Carareto CMA. 2009. The evolutionary dynamics of the Helena retrotransposon revealed by sequenced Drosophila genomes. BMC Evol. Biol. 9:174.

Gregory TR. 2005a. Genome Size Evolution in Animals. In: The Evolution of the Genome. p. 3–87.

Gregory TR. 2005b. The C-value enigma in plants and animals: A review of parallels and an appeal for partnership. Ann. Bot. 95:133–146.

Grentzinger T, Armenise C, Brun C, Serrano V, Pelisson A, Chambeyron S. 2012. piRNA-mediated transgenerational inheritance of an acquired trait piRNA-mediated transgenerational inheritance of an acquired trait. Genome Res. 22:1877–1888.

Grentzinger T, Armenise C, Pelisson A, Brun C, Mugat B, Chambeyron S. 2013. A user-friendly chromatographic method to purify small regulatory RNAs. Methods 67:91–101.

Grivna ST, Beyret E, Wang Z, Kim VN, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev.:1709–1714.

Guillén Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, Casillas S, Ramia M, Egea R, Negre B, et al. 2015. Genomics of Ecological Adaptation in Cactophilic Drosophila. Genome Biol. Evol. 7:349–366.

Guio L, Barrón MG, González J. 2014. The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Mol. Ecol. 23:2020–2030.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5’ end formation in Drosophila. Science 315:1587–1590.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Han BW, Wang W, Li C, Weng Z, Zamore PD. 2015. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science 348:817–821.

Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, Brennecke J. 2013. The genetic makeup of the Drosophila piRNA pathway. Mol. Cell 50:762–777.

Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, Stark A, Sachidanandam R, Brennecke J. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J. 30:3977–3993.

Hasson E, Naveira H, Fontdevila A. 1992. The breeding sites of Argentinian cactophilic species of the Drosophila mulleri complex (subgenus Drosophila-repleta group). Rev. Chil. Hist. Nat. 65:319–326.

Hedges DJ, Deininger PL. 2007. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. - Fundam. Mol. Mech. Mutagen. 616:46–59.

Hegarty MJ, Hiscock SJ. 2005. Hybrid speciation in plants: new insights from molecular studies. New Phytol. 165:411–423.

Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, et al. 2010. The genome of the Western clawed frog Xenopus tropicalis. Science 328:633–636.

Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD. 2007. The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. Curr. Biol. 17:1265–1272.

Houck MA, Clark JB, Peterson KR, Kidwell MG. 1991. Possible horizontal transfer of Drosophila genes by the mite Proctolaelaps regalis. Science 253:1125–1128.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov D V., Blaser H, Raz E, Moens CB, et al. 2007. A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. Cell 129:69–82.

Hu TT, Eisen MB, Thornton KR, Andolfatto P. 2013. A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome Res. 23:89–98.

Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. 2011. The struggle for life of the genome's selfish architects. Biol. Direct 6:1–29.

Hubbs CL. 1955. Hybridization between Fish Species in Nature. Syst. Zool. 4:1–20.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293.

Hughes JF, Coffin JM. 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. Genetics 171:1183–1194.

Hurtado J, Hasson E. 2013. Inter and intraspecific variation in female remating propensity in the cactophilic sibling species Drosophila buzzatii and D. koepferae. J. Insect Physiol. 59:569–576.

Hurtado J, Iglesias PP, Lipko P, Hasson E. 2013. Multiple paternity and sperm competition in the sibling species Drosophila buzzatii and Drosophila koepferae. Mol. Ecol. 22:5016–5026.

Hurtado J, Soto EM, Orellana L, Hasson E. 2012. Mating success depends on rearing substrate in cactophilic Drosophila. Evol. Ecol. 26:733–743.

Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. 2012. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature 491:279–283.

Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472:115–119.

Ito H. 2013. Small RNAs and regulation of transposons in plants. Genes Genet. Syst. 88:3–7.

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, et al. 2005. The genome of the kinetoplastid parasite, Leishmania major. Science 309:436–442.

Iwasaki YW, Siomi MC, Siomi H. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. Annu. Rev. Biochem. 84:405–433.

Johnson NA. 2010. Hybrid incompatibility genes: Remnants of a genomic battlefield? Trends Genet. 26:317–325.

Jones JM, Gellert M. 2004. The taming of a transposon: V(D)J recombination and the immune system. Immunol. Rev. 200:233–248.

Jordan IK, Matyunina L V, McDonald JF. 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. Proc. Natl. Acad. Sci. U. S. A. 96:12621–12625.

Jurka J, Kapitonov V V., Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110:462–467.

Kaneshiro KY. 1990. Natural hybridization in Drosophila, with special reference to species from Hawaii. Can. J. Zool. 68:1800–1805.

Kapitonov V V, Koonin E V. 2015. Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. Biol. Direct 10:20.

Kapitonov V V., Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 3:0998–1011.

Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, et al. 2001. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature 414:450–453.

Kawakami T, Dhakal P, Katterhenry AN, Heatherington CA, Ungerer MC. 2011. Transposable element proliferation and genome expansion are rare in contemporary sunflower hybrid populations despite widespread transcriptional activity of LTR retrotransposons. Genome Biol. Evol. 3:156–167.

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature 453:793–797.

Kawaoka S, Izumi N, Katsuma S, Tomari Y. 2011. 3’ End Formation of PIWI-Interacting RNAs In Vitro. Mol. Cell 43:1015–1022.

Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics 196:313–320.

Kelleher ES, Barbash DA. 2013. Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. Mol. Biol. Evol. 30:1816–1829.

Kelleher ES, Edelman NB, Barbash DA. 2012. Drosophila Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. PLoS Biol. 10:e1001428.

Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. Genome Res. 12:656–664.

Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86:813–833.

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

Kidwell MG, Novy JB. 1979. Hybrid dysgenesis in Drosophila melanogaster: Sterility resulting from gonadal dysgenesis in the P-M system. Genetics 92:1127–1140.

Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49–63.

Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat. Cell Biol. 11:652–658.

Klattenhoff C, Theurkauf W. 2008. Biogenesis and germline functions of piRNAs. Development 135:3–9.

Klenov MS, Lavrov SA, Korbut AP, Stolyarenko AD, Yakushev EY, Reuter M, Pillai RS, Gvozdev VA. 2014. Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res. 42:6208–6218.

Klenov MS, Lavrov SA, Stolyarenko AD, Ryazansky SS, Aravin AA, Tuschl T, Gvozdev VA. 2007. Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res. 35:5430–5438.

Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. 2015. The recent invasion of natural Drosophila simulans populations by the P-element. Proc. Natl. Acad. Sci. 112:6659–6663.

Kojima KK, Fujiwara H. 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol. Biol. Evol. 20:351–361.

Kojima KK, Jurka J. 2011. Crypton transposons: identification of new diverse families and ancient domestication events. Mob. DNA 2:12.

de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over Two-Thirds of the human genome. PLoS Genet. 7.

Kraaijeveld K. 2010. Genome Size and Species Diversification. Evol. Biol. 37:227–233.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 22:908–917.

Kyriacou CP, Hall JC. 1980. Circadian rhythm mutations in Drosophila melanogaster affect short-term fluctuations in the male's courtship song. Proc. Natl. Acad. Sci. U. S. A. 77:6729–6733.

Laayouni H, Hasson E, Santos M, Fontdevila A. 2003. The evolutionary history of Drosophila buzzatii. XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. Mol. Biol. Evol. 20:931–944.

Labrador M, Farré M, Utzet F, Fontdevila A. 1999. Interspecific hybridization increases transposition rates of Osvaldo. Mol. Biol. Evol. 16:931–937.

Lachaise D, Silvain J-F. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: The Drosophila melanogaster-D. simulans palaeogeographic riddle. Genetica 120:17–39.

Lagemaat LN Van De, Gagnier L, Medstrand P, Mager DL. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res.:1243–1249.

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.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Larsen PA, Marchán-Rivadeneira MR, Baker RJ. 2010. Natural hybridization generates mammalian lineage with species characteristics. Proc. Natl. Acad. Sci. U. S. A. 107:11447–11452.

Laturney M, Moehring AJ. 2012. The genetic basis of female mate preference and species isolation in Drosophila. Int. J. Evol. Biol. 2012:1–13.

Lau NC, Robine N, Martin R, Chung W-J, Niki Y, Berezikov E, Lai EC. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. Genome Res. 19:1776–1785.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA Complex from Rat Testes. Science 313:363–367.

Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette III LJ, Lohr JG, Harris CC, Ding L, Wilson RK, et al. 2012. Landscape of Somatic Retrotransposition in Human Cancers. Science 337:967–971.

Lerat E, Burlet N, Biémont C, Vieira C. 2011. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene 473:100–109.

Levin HL, Moran J V. 2011. Dynamic interactions between transposable elements and their hosts. Nat. Rev. Genet. 12:615–627.

Levitan DR, Fukami H, Jara J, Kline D, McGovern T, McGhee KE, Swanson CA, Knowlton N. 2004. Mechanisms of Reproductive Isolation Among Sympatric Broadcast- Spawning Corals of the Montastraea Annularis Species Complex. Evolution (N. Y). 58:308–323.

Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW, Xu J, Moore MJ, Schimenti JC, Weng Z, et al. 2013. An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. Mol. Cell 50:67–81.

Lim AK, Kai T. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc. Natl. Acad. Sci. 104:6714–6719.

Lin Y-J, Seroude L, Benzer S. 1998. Extended Life-Span and Stress Resistance in the Drosophila Mutant methuselah. Science 282:943–947.

Lisch D. 2009. Epigenetic regulation of transposable elements in plants. Annu. Rev. Plant Biol. 60:43–66.

Liu B, Wendel JF. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874–880.

Liu Z, Wang Y, Shen Y, Guo W, Hao S, Liu B. 2004. Extensive alterations in DNA methylation and transcription in rice caused by introgression from Zizania latifolia. Plant Mol. Biol. 54:571–582.

Lopez-Flores I, Garrido-Ramos MA. 2012. The repetitive DNA content of eukaryotic genomes. Genome Dyn. 7:1–28.

Lorenzi H, Thiagarajan M, Haas B, Wortman J, Hall N, Caler E. 2008. Genome wide survey, discovery and evolution of repetitive elements in three Entamoeba species. BMC Genomics 9:595.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:1–21.

Lukic S, Chen K. 2011. Human piRNAs are under selection in Africans and repress transposable elements. Mol. Biol. Evol. 28:3061–3067.

MacCallum CJ, Nürnberger B, Barton NH, Szymura JM. 1998. Habitat Preference in the Bombina Hybrid Zone in Croatia. Evolution (N. Y). 52:227–239.

Mahelka V, Suda J, Jarolímová V, Trávnicek P, Krahulec F. 2005. Genome size discriminates between closely related taxa Elytrigia repens and E. intermedia (Poaceae: Triticeae) and their hybrid. Folia Geobot. 40:367–384.

Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. Annu. Rev. Genet. 45:331–355.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–237.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell 137:522–535.

Manfrin MH, Sene FM. 2006. Cactophilic Drosophila in South America: a model for evolutionary studies. Genetica 126:57–75.

Marí-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. 2013. Reconstructing de novo silencing of an active plant retrotransposon. Nat. Genet. 45:1029–1039.

Marín I, Fontdevila A. 1998. Stable Drosophila buzzatii - Drosophila koepferae Hybrids. J. Hered. 89:336–339.

Marin I, Ruiz A, Pla C, Fontdevila A. 1993. Reproductive Relationships among Ten Species of the Drosophila repleta Group from South America and the West Indies. Evolution (N. Y). 47:1616–1624.

Markow TA, O'Grady P. 2009. Phylogenetic taxonomy in Drosophila - Problems and prospects. Fly (Austin).:10–14.

Markow TA, O'Grady PM. 2007. Drosophila biology in the genomic age. Genetics 177:1269–1276.

Marques JT, Kim K, Wu PH, Alleyne TM, Jafari N, Carthew RW. 2010. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. Nat. Struct. Mol. Biol. 17:24–30.

Masly JP, Jones CD, Noor MAF, Locke J, Orr HA. 2006. Gene transposition as a cause of hybrid sterility in Drosophila. Science 313:1448–1450.

Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. PLoS Biol. 5:1890–1898.

Masly JP. 2012. 170 Years of "Lock-and-Key": Genital Morphology and Reproductive Isolation. Int. J. Evol. Biol. 2012:1–10.

Mateo L, Ullastres A, González J. 2014. A transposable element insertion confers xenobiotic resistance in Drosophila. PLoS Genet. 10:e1004560.

Matthews KA, Kaufman TC, Gelbart WM. 2005. Research resources for Drosophila: the expanding universe. Nat. Rev. Genet. 6:179–193.

Mayr E. 1942. Systematics and the origin of the species. New York: Columbia University Press

Mayr E. 1963. Animal Species and Evolution. Boston: Harvard University Press

McClintock B. 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. 36:344–355.

McClintock B. 1983. The Significance of Responses of the Genome to Challenge. Nobel Lect.:180–199.

McClintock B. 1984. The Significance of Responses of the Genome to Challenge. Science 226:792–801.

McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 32:W20–W25.

Metcalfe CJ, Bulazel K V, Ferreri GC, Schroeder-Reiter E, Wanner G, Rens W, Obergfell C, Eldridge MDB, O'Neill RJ. 2007. Genomic instability within centromeres of interspecific marsupial hybrids. Genetics 177:2507–2517.

Mi S, Lee X, Li X-P, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang X-Y, Edouard P, Howes S, et al. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403:785–789.

Michalak P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. Heredity (Edinb). 102:45–50.

Miller WJ, McDonald JF, Nouaud D, Anxolabéhère D. 1999. Molecular domestication - more than a sporadic episode in evolution. Genetica 107:197–207.

Mishra B. 2008. Transposable Element-driven Duplications during Hominoid Genome Evolution. Encycl. Life Sci.

Misumi O, Matsuzaki M, Nozaki H, Miyagishima S, Mori T, Nishida K, Yagisawa F, Yoshida Y, Kuroiwa H, Kuroiwa T. 2005. Cyanidioschyzon merolae genome. A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. Plant Physiol. 137:567–585.

Modolo L, Lerat E. 2015. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. BMC Bioinformatics 16:137.

Modolo L, Picard F, Lerat E. 2014. A new genome-wide method to track horizontally transferred sequences: Application to Drosophila. Genome Biol. Evol. 6:416–432.

Mohn F, Handler D, Brennecke J. 2015. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science 348:812–817.

Mohn F, Sienski G, Handler D, Brennecke J. 2014. The Rhino-Deadlock-Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell 157:1364–1379.

Molla-Herman A, Vallés AM, Ganem-Elbaz C, Antoniewski C, Huynh J-R. 2015. tRNA processing defects induce replication stress and Chk2-dependent disruption of piRNA transcription. EMBO J. 34:3009–3027.

Montanari SR, Hobbs J-PA, Pratchett MS, van Herwerden L. 2016. The importance of ecological and behavioural data in studies of hybridisation among marine fishes. Rev. Fish Biol. Fish.

Morán T, Fontdevila A. 2014. Genome-Wide Dissection of Hybrid Sterility in Drosophila Confirms a Polygenic Threshold Architecture. J. Hered. 105:381–396.

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ. 2013. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. Mol. Cell 50:736–748.

Nagao A, Mituyama T, Huang H, Chen D, Siomi MC. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. RNA 16:2503–2515.

Naveira H, Fondevila A. 1991. The evolutionary history of Drosophila buzzatii. XXI. Cumulative action of multiple sterility factors on spermatogenesis in hybrids of D. buzzatii and D. koepferae. Heredity (Edinb). 67:57–72.

Naveira H, Fontdevila A. 1985. The evolutionary history of Drosophila buzzatii. IX. High frequencies of new chromosome rearrangements induced by introgressive hybridization. Chromosoma 91:87–94.

Naveira H, Fontdevila A. 1986. The Evolutionary History of Drosophila buzzatii. XII. The Genetic Basis of Sterility in Hybrids between D. buzzatii and Its Sibling D. serido from Argentina. Genetics 114:841–857.

Nevo E. 2012. Speciation: Chromosomal Mechanisms. eLS:1–10.

Nilsen TW. 2008. Endo-siRNAs: yet another layer of complexity in RNA silencing. Nat. Struct. Mol. Biol. 97:341–344.

Nishida KM, Iwasaki YW, Murota Y, Nagao A, Mannen T, Kato Y, Siomi H, Siomi MC. 2015. Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in bombyx germ cells. Cell Rep. 10:193–203.

Nishida KM, Okada TN, Kawamura T, Mituyama T, Kawamura Y, Inagaki S, Huang H, Chen D, Kodama T, Siomi H, et al. 2009. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. EMBO J. 28:3820–3831.

Nishida KM, Saito K, Mori T. 2007. Gene silencing mechanisms mediated by Aubergine − piRNA complexes in Drosophila male gonad. RNA 13:1911–1922.

Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, et al. 2012. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature 491:284–287.

Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. U. S. A. 98:12084–12088.

O'Neill RJW, O'Neill MJ, Marshall Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393:68–73.

Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364:99–115.

Ohtani H, Iwasaki YW, Shibuya A, Siomi H, Siomi MC, Saito K. 2013. DmGTSF1 is necessary for Piwi–piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. Genes Dev. 27:1656–1661.

Okamura K, Chung W-JJ, Ruby JG, Guo H, Bartel DP, Lai EC. 2008. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453:803–806.

Oliveira CC, Manfrin MH, Sene FDM, Etges WJ. 2013. Evolution of male courtship songs in the Drosophila buzzatii species cluster. In: Speciation: Natural Processes, Genetics and Biodiversity. New York: Nova Science. p. 137–164.

Oliveira DCSG, Almeida FC, O'Grady PM, Armella M a, Desalle R, Etges WJ. 2012. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the Drosophila repleta species group. Mol. Phylogenet. Evol.

Olivieri D, Senti K-A, Subramanian S, Sachidanandam R, Brennecke J. 2012. The Cochaperone Shutdown Defines a Group of Biogenesis Factors Essential for All piRNA Populations in Drosophila. Mol. Cell 47:954–969.

Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J. 29:3301–3317.

Ometto L, Cestaro A, Ramasamy S, Grassi A, Revadi S, Siozios S, Moretto M, Fontana P, Varotto C, Pisani D, et al. 2013. Linking genomics and ecology to investigate the complex evolution of an invasive Drosophila pest. Genome Biol. Evol. 5:745–757.

Orgel LE, Crick FH. 1980. Selfish DNA: the ultimate parasite. Nature 284:604–607.

Pascale E, Valle E, Furano A V. 1990. Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. Proc. Natl. Acad. Sci. 87:9481–9485.

Patil VS, Anand A, Chakrabarti A, Kai T. 2014. The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. BMC Biol. 12:1–15.

Patil VS, Kai T. 2010. Repression of Retroelements in Drosophila Germline via piRNA Pathway by the Tudor Domain Protein Tejas. Curr. Biol. 20:724–730.

Petrov DA, Hartl DL. 1998. High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. Mol. Biol. Evol. 15:293–302.

Phadnis N, Baker EP, Cooper JC, Frizzell KA, Hsieh E, de la Cruz AFA, Shendure J, Kitzman JO, Malik HS. 2015. An essential cell cycle regulation gene causes hybrid inviability in Drosophila. Science 350:1552–1555.

Phadnis N, Orr HA. 2009. A Single Gene Causes Both Male Sterility and Segregation Distortion in \emph{{D}rosophila} Hybrids. Science 323:376–379.

Phadnis N. 2011. Genetic architecture of male sterility and segregation distortion in drosophila pseudoobscura bogota-USA hybrids. Genetics 189:1001–1009.

Picard G. 1976. Non mendelian female sterility in Drosophila melanogaster: hereditary transmission of I factor. Genetics 83:107–123.

Piccinali R, Aguadé M, Hasson E. 2004. Comparative molecular population genetics of the Xdh locus in the cactophilic sibling species Drosophila buzzatii and D. koepferae. Mol. Biol. Evol. 21:141–152.

Prak ET, Kazazian HH. 2000. Mobile elements and the human genome. Nat. Rev. Genet. 1:134–144.

Preall JB, Czech B, Guzzardo PM, Muerdter F, Hannon GJ. 2012. shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA 18:1446–1457.

Pritham EJ. 2009. Transposable elements and factors influencing their success in eukaryotes. J. Hered. 100:648–655.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2012. The Pfam protein families databases. Nucleic Acids Res. 40:D290–D301.

Ramsey J, Schemske DW. 1998. Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. Annu. Rev. Ecol. Syst. 29:467–501.

Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, Hannon GJ, Lehmann R. 2011. PiRNA production requires heterochromatin formation in Drosophila. Curr. Biol. 21:1373–1379.

Ravindran S. 2012. Barbara McClintock and the discovery of jumping genes. Proc. Natl. Acad. Sci. 109:20198–20199.

Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL. 2008. Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Res. 18:717–728.

Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: Towards new species. Gene 454:1–7.

Rebollo R, Lerat E, Kleine LL, Biémont C, Vieira C. 2008. Losing helena: the extinction of a drosophila line-like element. BMC Genomics 9:149.

Renaut S, Rowe HC, Ungerer MC, Rieseberg LH. 2014. Genomics of homoploid hybrid speciation: diversity and transcriptional activity of long terminal repeat retrotransposons in hybrid sunflowers. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 369:20130345.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, et al. 2005. Comparative genome sequencing of Drosophila pseudoobscura: Chromosomal, gene, and cis-element evolution. Genome Res. 15:1–18.

Richmond MP. 2014. The role of aedeagus size and shape in failed mating interactions among recently diverged taxa in the Drosophila mojavensis species cluster. BMC Evol. Biol. 14:1–9.

Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D, et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res. 21:147–163.

Rieseberg LH, Carter R, Zona S. 1990. Molecular Tests of the Hypothesized Hybrid Origin of Two Diploid Helianthus Species (Asteraceae). Evolution (N. Y). 44:1498–1511.

Rieseberg LH, Kim SC, Randell RA, Whitney KD, Gross BL, Lexer C, Clay K. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. Genetica 129:149–165.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major Ecological Transitions in Wild Sunflowers Facilitated by Hybridization. Science 301:1211–1216.

Rieseberg LH, Sinervo B, Randal Linder C, Ungerer MC, Arias DM. 1996. Role of Gene Interactions in Hybrid Speciation: Evidence from Ancient and Experimental Hybrids. Science 272:741–745.

Rieseberg LH. 1991. Homoploid Reticulate Evolution in Helianthus (Asteraceae): Evidence from Ribosomal Genes. Am. J. Bot. 78:1218–1237.

Rieseberg LH. 2001. Chromosomal Rearrangements and Speciation. Trends Ecol. Evol. 16:351–358.

Rigal M, Mathieu O. 2011. A “mille-feuille” of silencing: Epigenetic control of transposable elements. Biochim. Biophys. Acta 1809:452–458.

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

Romero-Soriano V, Burlet N, Vela D, Fontdevila A, Vieira C, Garcia Guerreiro MP. 2016. Drosophila females undergo genome expansion after interspecific hybridization. Genome Biol. Evol. 8:556–561.

Romero-Soriano V, García Guerreiro MP. 2016. Expression of the Retrotransposon Helena Reveals a Complex Pattern of TE Deregulation in Drosophila Hybrids. PLoS One 11:e0147903.

Le Rouzic A, Boutin TS, Capy P. 2007. Long-term evolution of transposable elements. Proc. Natl. Acad. Sci. U. S. A. 104:19375–19380.

Le Rouzic A, Capy P. 2005. The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift. Genetics 169:1033–1043.

Rozhkov N V, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev.

Ruiz A, Wasserman M. 1993. Evolutionary cytogenetics of the Drosophila buzzatii species complex. Heredity (Edinb). 70 (Pt 6):582–596.

Russo CAM, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol. Biol. Evol. 12:391–404.

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. 24:2493–2498.

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev. 20:2214–2222.

Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2’-O-methylation of Piwi-interacting RNAs at their 3' ends. Genes Dev. 21:1603–1608.

Saito K, Siomi MC. 2010. Small RNA-mediated quiescence of transposable elements in animals. Dev. Cell 19:687–697.

Sánchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in Drosophila. Trends Genet. 21:200–203.

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

Sassaman DM, Dombroski BA, Moran J V, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH. 1997. Many human L1 elements are capable of retrotransposition. Nat. Genet. 16:37–43.

Sato K, Iwasaki YW, Shibuya A, Carninci P, Tsuchizawa Y, Ishizu H, Siomi MC, Siomi H. 2015. Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Mol. Cell 59:553–563.

Saxe JP, Chen M, Zhao H, Lin H. 2013. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. EMBO J. 32:1869–1885.

Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol. Evol. 25:537–546.

Schemske DW, Bradshaw HD. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (Mimulus). Proc. Natl. Acad. Sci. U. S. A. 96:11910–11915.

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.

Scribner KT, Page KS, Bartron ML. 2000. Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. Rev. Fish Biol. Fish. 10:293–323.

Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Senerchia N, Parisod C, Parisod C. 2015. Genome reorganization in F1 hybrids uncovers the role of retrotransposons in reproductive isolation. Proc. R. Soc. B Biol. Sci. 282:20142874.

Senti K-A, Jurczak D, Sachidanandam R, Brennecke J. 2015. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. Genes Dev. 29:1747–1762.

Shan X, Liu Z, Dong Z, Wang Y, Chen Y, Lin X, Long L, Han F, Dong Y, Liu B. 2005. Mobilization of the active MITE transposons mPing and Pong in rice by introgression from wild rice (Zizania latifolia Griseb.). Mol. Biol. Evol. 22:976–990.

Shapiro JA. 1969. Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. J. Mol. Biol. 40:93–105.

Sienski G, Batki J, Senti K-A, Dönertas D, Tirian L, Meixner K, Brennecke J. 2015. Silencio / CG9754 connects the Piwi – piRNA complex to the cellular heterochromatin machinery. Genes Dev. 29:1–14.

Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151:964–980.

Silva-Sousa R, López-Panadès E, Casacuberta E. 2012. Drosophila telomeres: An example of co-evolution with transposable elements. Genome Dyn. 7:46–67.

Simkin A, Wong A, Poh Y-P, Theurkauf WE, Jensen JD. 2013. Recurrent and Recent Selective Sweeps in the piRNA Pathway. Evolution (N. Y). 67:1081–1090.

Sinzelle L, Izsvák Z, Ivics Z. 2009. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. Cell. Mol. Life Sci. 66:1073–1093.

Siomi MC, Miyoshi T, Siomi H. 2010. piRNA-mediated silencing in Drosophila germlines. Semin. Cell Dev. Biol. 21:754–759.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8:272–285.

Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA. 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell 136:461–472.

Sniegowski PD, Charlesworth B. 1994. Transposable Element Numbers in Cosmopolitan Inversions From a Natural Population of Drosophila melanogaster. Genetics.

Sokal RR, Oden NL, Barker JSF. 1987. Spatial Structure in Drosophila buzzatii Populations: Simple and Directional Spatial Autocorrelation. Am. Nat. 129:122–142.

Soto EM, Soto IM, Carreira VP, Fanara JJ, Hasson E. 2008. Host-related life history traits in interspecific hybrids of cactophilic Drosophila. Entomol. Exp. Appl. 126:18–27.

Soto IM, Carreira VP, Corio C, Padró J, Soto EM, Hasson E. 2014. Differences in tolerance to host cactus alkaloids in Drosophila koepferae and D. buzzatii. PLoS One 9:1–9.

Soto IM, Carreira VP, Fanara JJ, Hasson E. 2007. Evolution of male genitalia: environmental and genetic factors affect genital morphology in two Drosophila sibling species and their hybrids. BMC Evol. Biol. 7:77.

Soto IM, Carreira VP, Soto EM, Márquez F, Lipko P, Hasson E. 2013. Rapid Divergent Evolution of Male Genitalia Among Populations of Drosophila buzzatii. Evol. Biol. 40:395–407.

Soto IM. 2012. Aedeagal Divergence in Sympatric Populations of Two Sibling Species of Cactophilic Drosophila (Diptera: Drosophilidae): Evidence of Character Displacement? Neotrop. Entomol. 41:207–213.

Sun S, Ting C-T, Wu C-I. 2004. The Normal Function of a Speciation Gene, Odysseus, and Its Hybrid Sterility Effect. Science 305:81–83.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453:534–538.

Taylor AL. 1963. Bacteriophage-induced mutation in Escherichia coli. Proc. Natl. Acad. Sci. 50:1043–1051.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815.

The C. elegans Sequencing Consortium. 1998. Genome Sequence of the Nematode C. elegans: A Platform For Investigating Biology. Science 282:2012–2018.

The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25:25–29.

Théron E, Dennis C, Brasset E, Vaury C. 2014. Distinct features of the piRNA pathway in somatic and germ cells: from piRNA cluster transcription to piRNA processing and amplification. Mob. DNA 5:1–11.

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin A a, Tóth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev.

Turelli M, Moyle LC. 2006. Asymmetric Postmating Isolation: Darwin’s Corollary to Haldane's Rule. Genetics 176:1059–1088.

Turner LM, Harr B. 2014. Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. Elife 3:e02504.

Turner LM, White MA, Tautz D, Payseur BA. 2014. Genomic Networks of Hybrid Sterility. PLoS Genet. 10:18–22.

Ungerer MC, Strakosh SC, Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. Curr. Biol. 16:R872–R873.

Vagin V V, Sigova A, Li C, Gvozdev V, Zamore PD. 2006. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science 313:320–324.

Vagin V V, Yu Y, Jankowska A, Luo Y, Wasik KA, Malone CD, Harrison E, Rosebrock A, Wakimoto BT, Fagegaltier D, et al. 2013. Minotaur is critical for primary piRNA biogenesis. RNA 19:1064–1077.

Vela D, Fontdevila A, Vieira C, García Guerreiro MP. 2014. A genome-wide survey of genetic instability by transposition in Drosophila hybrids. PLoS One 9:e88992.

Vela D, García Guerreiro MP, Fontdevila A. 2011. Adaptation of the AFLP technique as a new tool to detect genetic instability and transposition in interspecific hybrids. Biotechniques 50:247–250.

Verneau O, Catzeflis F, Furano A V. 1998. Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. Proc. Natl. Acad. Sci. U. S. A. 95:11284–11289.

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

Villesen P, Aagaard L, Wiuf C, Pedersen FS. 2004. Identification of endogenous retroviral reading frames in the human genome. Retrovirology 1:1–13.

Vourekas A, Zheng K, Fu Q, Maragkakis M, Alexiou P, Ma J, Pillai RS, Mourelatos Z, Jeremy Wang P. 2015. The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev. 29:617–629.

Wang H-Y, Tian Q, Ma Y-Q, Wu Y, Miao G-J, Ma Y, Cao D-H, Wang X-L, Lin C, Pang J, et al. 2010. Transpositional reactivation of two LTR retrotransposons in rice-Zizania recombinant inbred lines (RILs). Hereditas 147:264–277.

Wang N, Wang H, Wang H, Zhang D, Wu Y, Ou X, Liu S, Dong Z, Liu B. 2010. Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with Zizania latifolia. BMC Plant Biol. 10:190.

Wang SH, Elgin SCR. 2011. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc. Natl. Acad. Sci. 108:21164–21169.

Wang W, Han BW, Tipping C, Ge DT, Zhang Z, Weng Z, Zamore PD. 2015. Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol. Cell 59:819–830.

Wang YM, Dong ZY, Zhang ZJ, Lin XY, Shen Y, Zhou D, Liu B. 2005. Extensive de novo genomic variation in rice induced by introgression from wild rice (Zizania latifolia Griseb.). Genetics 170:1945–1956.

Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev. 20:1732–1743.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453:539–543.

Webster A, Li S, Hur JK, Wachsmuth M, Bois JS, Perkins EM, Patel DJ, Aravin AA. 2015. Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. Mol. Cell 59:564–575.

Wendel JF. 2000. Genome evolution in polyploids. Plant Mol. Evol. 42:225–249.

Wheeler DA, Kyriacou CP, Greenacre ML, Yu Q, Rutila JE, Rosbash M, Hall JC. 1991. Molecular transfer of a Species-Specific behavior from Drosophila simulans to Drosophila melanogaster. Science 251:1082–1085.

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. 10:276.

Wiernasz DC, Kingsolver JG. 1992. Wing melanin pattern mediates species recognition in Pieris occidentalis. Anim. Behav. 43:89–94.

Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135:635–648.

Wu Y, Jiang T, Sun Y, Wang Z, Guo G, Sun S, Wang J, Li N, Wang Z, Zhang D, et al. 2014. Mobilization of Diverse Transposable Elements in Rice Induced by Alien Pollination Without Entailing Genetic Introgression. Plant Mol. Biol. Report. 33:1181–1191.

Xie W, Donohue RC, Birchler JA. 2013. Quantitatively increased somatic transposition of transposable elements in Drosophila strains compromised for RNAi. PLoS One 8:e72163.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004. Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. 2:642–652.

Xiol J, Spinelli P, Laussmann MA, Homolka D, Yang Z, Cora E, Couté Y, Conn S, Kadlec J, Sachidanandam R, et al. 2014. RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell 157:1698–1711.

Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, et al. 2004. The genome of Cryptosporidium hominis. Nature 431:1107–1112.

Yang Z, Pillai RS. 2014. Fly piRNA biogenesis: tap dancing with Tej. BMC Biol. 12:77.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Yasuhara JC, Wakimoto BT. 2008. Molecular landscape of modified histones in Drosophila heterochromatic genes and euchromatin-heterochromatin transition zones. PLoS Genet. 4:0159–0172.

Yin H, Sweeney S, Raha D, Snyder M, Lin H. 2011. A High-Resolution Whole-Genome map of key chromatin modifications in the adult Drosophila melanogaster. PLoS Genet. 7.

Yu Y, Gu J, Jin Y, Luo Y, Preall JB, Ma J, Czech B, Hannon GJ. 2015. Panoramix enforces piRNA-dependent cotranscriptional silencing. Science 350:339–342.

Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, Hannon GJ, Lehmann R. 2011. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development 138:4039–4050.

Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD, et al. 2012. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell 151:871–884.

Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, Theurkauf WE. 2014. The HP1 homolog Rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell 157:1353–1363.

Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, Weng Z, Theurkauf WE, Zamore PD. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol. Cell 44:572–584.

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE. 2014. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell 157:1050–1060.

Zhou Q, Bachtrog D. 2012. Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in Drosophila. Science 337:341–345.

Zhou Q, Zhu H, Huang Q, Zhao L, Zhang G, Roy SW, Vicoso B, Xuan Z, Ruan J, Zhang Y, et al. 2012. Deciphering neo-sex and B chromosome evolution by the draft genome of Drosophila albomicans. BMC Genomics 13:109.

Zhou R, Czech B, Brennecke J, Sachidanandam R, Wohlschlegel JA, Perrimon N, Hannon GJ. 2009. Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. RNA 15:1886–1895.

Zhou X, Ma J, Wang W, Gong N, Zhang Y, Liu J. 2010. Genome size of the diploid hybrid species Hippophae goniocarpa and its parental species, h. rhamnoides ssp. sinensis and H. neurocarpa ssp. neurocarpa (Elaeagnaceae). Acta Biol. Cracoviensia Ser. Bot. 52:12–16.

Zuckerkandl E, Hennig W. 1995. Tracking heterochromatin. Chromosoma 104:75–83.

7 Acknowledgements

Mirant enrere em sembla molt llunyà l'estiu en què, tot just havent acabat la carrera, vaig saber que m'havien concedit la oportunitat de començar una tesi doctoral. Estava de colònies a la vall de Pineta en el que va ser el meu últim any com a monitora. Tant la cobertura com l'accés a internet eren limitats, però vaig trobar un moment per utilitzar l'únic ordinador i consultar la resolució de la convocatòria de les beques PIF que concedia la Universitat Autònoma de Barcelona. Recordo perfectament (a aquestes alçades no serviria de res mentir) no tenir gens clar si la investigació m'agradaria, però em semblava impensable rebutjar una oportunitat així. Quatre anys després, no trobo cap motiu per penedir-me de la meva decisió. Arribats aquí, m'agradaria agrair la confiança que la meva directora de tesi, la Dra. Maria Pilar Garcia Guerreiro, va dipositar en mi en aquell moment. Muchísimas gracias por darme la oportunidad de conocer el mundo de la investigación, por todo el esfuerzo que has vertido en mi trabajo durante estos años, y también por tu disponibilidad y característica eficiencia.

Després de cinc anys, comença a ser difícil recapitular tota la gent que ha anat passant pel laboratori. Vull concedir-li un lloc especial a la Montse Peiró, perquè no m'imagino els meus principis al grup sense ella, per la paciència que va tenir i encara té, i perquè em posa de bon humor sentir *«bon dia, lady!»* tots els matins quan ens creuem als passadissos. També és important que esmenti la Carmen, en qui vaig trobar una amiga i aliada política els meus dos primers anys aquí. Gracias a las doctorandas *sénior* que me sirvieron de modelo y ejemplo al principio de mi tesis, las ya doctoras Doris y Luz. También a Luis, a quien admiro profundamente como científico y persona (además de tener la familia más adorable del mundo). El dia a dia al laboratori no hagués sigut (ni remotament) el mateix sense la Marta. Gràcies per aguantar el meu mal humor, ha estat tota una experiència descobrir la ciència al teu costat. Gracias también a la indomable Eila, por dejarse domesticar y a la vez domesticarme a mí. Y a Víctor, el entrañable descubrimiento tardío. Finalment, no puc oblidar-me de tots aquells que formen o han format part, més o menys temps, del Grup de Biologia Evolutiva. Gràcies a l'Alba, en Sergi, la Hayley, la Blanca, el Joan, la Lluna i l'Aline; la Rosa, el Francisco, en Mauro, i en especial al Dr. Fontdevila, pel seu interès.

Han estat imprescindibles al llarg d'aquests anys la Montse Sales i la Raquel, dos dels meus principals motius per somriure sense voler quan baixo al primer pis (en oposició, clar està, a la col·lecció de cucs de seda amb qui conviviu). Formeu un dels tàndems més divertits que he conegut mai. Gràcies d'altra banda a l'Elena, la Maite i la Mariajosep. És espectacular com treballar amb vosaltres facilita la (meva) feina. Ha estat un privilegi poder gaudir de la vostra professionalitat.

De plus, je souhaite remercier Cristina d'avoir bien voulu m'héberger en France. Mes trois séjours à Lyon m'ont donné l'opportunité de connaître une nouvelle façon de percevoir et comprendre la

science. Merci à toute l'équipe *Élements Transposables, Évolution, Populations* de votre accueil : Laurent, Hélène, Clément, Bianca, Emanuel, Sébastien, Lain, Nelly, Marie, Emmanuelle, Annabelle, Matthieu et (encore une fois) Cristina. Merci aussi à Eugénie et Robin. J'ai l'énorme chance d'avoir à présent de vrais amis à Lyon (et Villeurbanne).

No vull imaginar-me en quin estat hagués acabat aquesta tesi si al llarg d'aquests anys no hagués seguit gaudint del volei. Potser, qui sap, no hagués ni aconseguit acabar-la. Sempre dic que el volei m'és necessari per tocar de peus a terra. Amb l'especialització professional, les persones que vas coneixent s'assemblen més i més a tu, pel que fa a interessos i curiositat. Passar gran part de la teva vida amb gent semblant a un mateix és agradable i sobretot còmode, però no puc evitar pensar que tanca portes a visions diferents de la realitat. En paraules (del traductor) de Stefan Zweig, *«plus un esprit se limite, plus il touche par ailleurs à l'infini»*. En el meu cas, és gràcies al volei que no sóc una *«monomaníaca»*.

Per consistència cronològica, vull començar donant les gràcies al Club Voleibol Esplugues per la seva acollida. En aquest club, on vaig jugar al principi del meu doctorat, vaig tenir la sort de conèixer persones increïbles, moltes de les quals encara puc considerar amigues. Gràcies a tothom que va compartir amb mi aquella etapa, amb una menció especial a la Virgínia, la Núria, la Júlia, la Zoe, la Quimey, la Katia, la Íngrid, la Sofi i la Cris. Va ser un plaer jugar amb vosaltres.

Siguiendo la cronología, me toca darle las gracias a Rafa por darme la oportunidad de volver a jugar en Sant Cugat, mi club de toda la vida. Gràcies també a totes les jugadores que em van rebre de tornada, tant a les que coneixia com a les que vaig descobrir llavors: Andrea Zuco, Estefania, Pilar, Andrea Garmendia, Maria, Sara, Cris; i en especial a la Neus i a la Marina, que han passat a ser pilars de la meva estabilitat dins i fora del camp. Les segueixen les petites Anna, Ainhoa, Carol i Sesé, juntament amb l'Esteve – gràcies a tots vosaltres, que veu formar part d'una de les temporades de les que guardo millor record. De l'etapa més recent, em queda dir gràcies a l'Iris, l'Alicia, la Laura Alcalde, el Lluc, i a la capitana eterna, la Mònica. Gràcies a tots per la vostra empatia durant l'últim any (però senyalo a la Neus i senyalo a la Mònica). Finalment, gràcies a les bebès juvenils, que han superat tots els estàndards de cuquisme de l'univers (us confesso que m'és totalment impossible d'establir un rànquing).

No cal ni dir que qui diu que no es poden combinar estudis i esport no sap del que parla. Estic convençuda de que l'esport m'ha donat els pitjors moments de la meva vida, però d'alguna manera

sempre queden sobradament compensats pels bons. És curiós com aquesta descripció s'assembla a la que faria actualment de la ciència.

Tot i que com sempre m'allargo massa, no puc acabar aquesta secció sense donar les gràcies a tots els meus amics, que d'una manera o altra han contribuït a minimitzar l'horror en els moments més crítics, així com a enriquir els moments de tranquil·litat. De la meva infància i adolescència, vull mencionar la Claire, la Marta Baró, la Jordina, la Carolina i la Íngrid. Em considero afortunada de mantenir la vostra amistat. De la universitat, vull començar donant les gràcies a la Sandra Motas, la Blanca, la Sílva i la Núria. També a la Karen, l'Ángel, a la Neus, al Jordi, a l'Alba i a l'Anna. Tant els que heu emigrat com els que esteu a prop us heu tornat difícils (com jo) de veure, però la part positiva és que sabem que és culpa d'un mateix amant, molt absorbent, que compartim. A tots aquells que esteu acabant la tesi, us desitjo molta sort! La llum al final del túnel potser us sembla tènue però us hi esteu apropant cada cop més. Finalment, moltes gràcies a l'únic que em queda de les colònies, una etapa que vaig tancar al començar la tesi. Albert: pots ser groc, blau, verd, lila, negre, canviar de color i desaparèixer, però torna sempre, si us plau.

Gràcies a la meva família per la confiança, el suport incondicional i els ànims. A la meva germana Adriana, a qui el món li sembla tan fàcil d'entendre que molt sovint aconsegueix que sigui així. A la meva mare, que m'ha ensenyat la increïble força que tenen la perseverança i el romanticisme. Al meu pare, que com a bon físic m'ha fet entendre la relativitat i a acceptar que ni el bé es bé ni el mal és mal. A la Minchin, per existir *per se*.

Agraeixo d'altra banda a la Carme, la Teresa, el Xavi i el Josep haver-me fet veure que l'estabilitat no només existeix sinó que es pot viure amb plenitud. Alguns d'aquests noms, per suposat, inclouen més d'un individu.

També donaré gràcies per l'oxigen que li dec a molts personatges. Will, Dwight, Michael, Leslie, Ron, Malcolm, Ruzena. Entre centenars d'altres.

Finalment, no sé com agrair amb paraules la paciència i el suport del Xavi. Gràcies infinites per caminar al meu costat, per la teva contribució (moral, estètica i científica) a la meva tesi, per ensenyar-me que és possible viure sense fantasmes. Aviat et tocarà a tu. Much more than six seasons and a movie (and worth it).

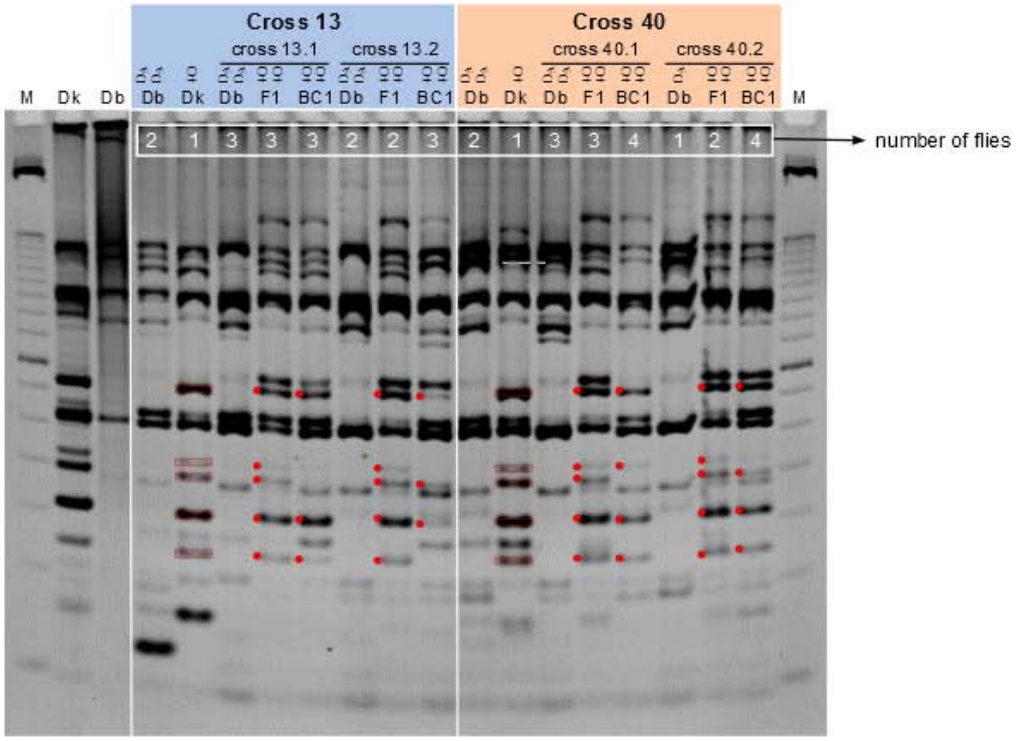
8 Annexes

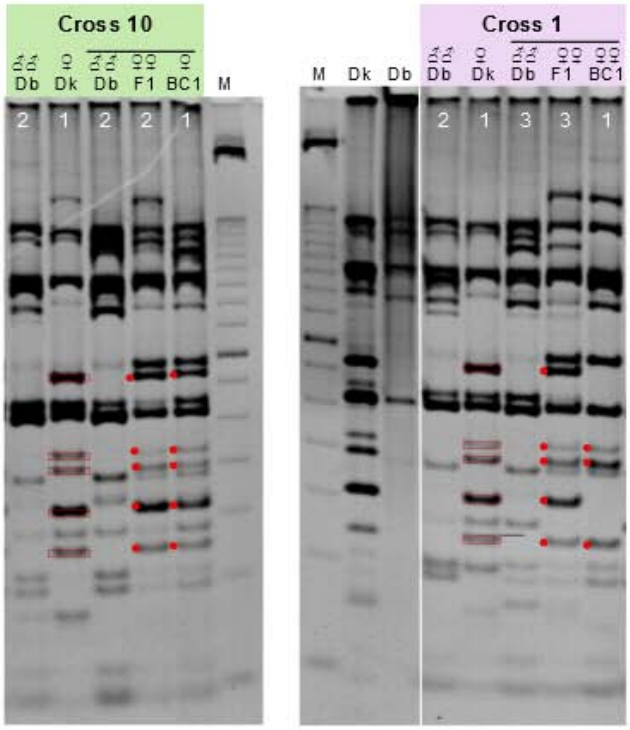
8.1 Supplementary data of "*Drosophila* females undergo genome expansion after interspecific hybridization" (chapter 3.1)

8.1.1 Figure S1 – AFLP genotyping example

Figure S1. Example of AFLP genotyping for the six analyzed families for one out of 10 analyzed primer combinations (GG-CGG). M= molecular weight marker, Db= *D. buzzatii*, Dk= *D. koepferae*. Red rectangles indicate *D. koepferae*-specific markers identified in the six studied families. Red dots indicate the presence of those markers in F1 and BC1 hybrids. The table summarizes the counting results for each family for this primer combination (Dk-sp= *D. koepferae*-specific).

Primer combination: GG-CGG

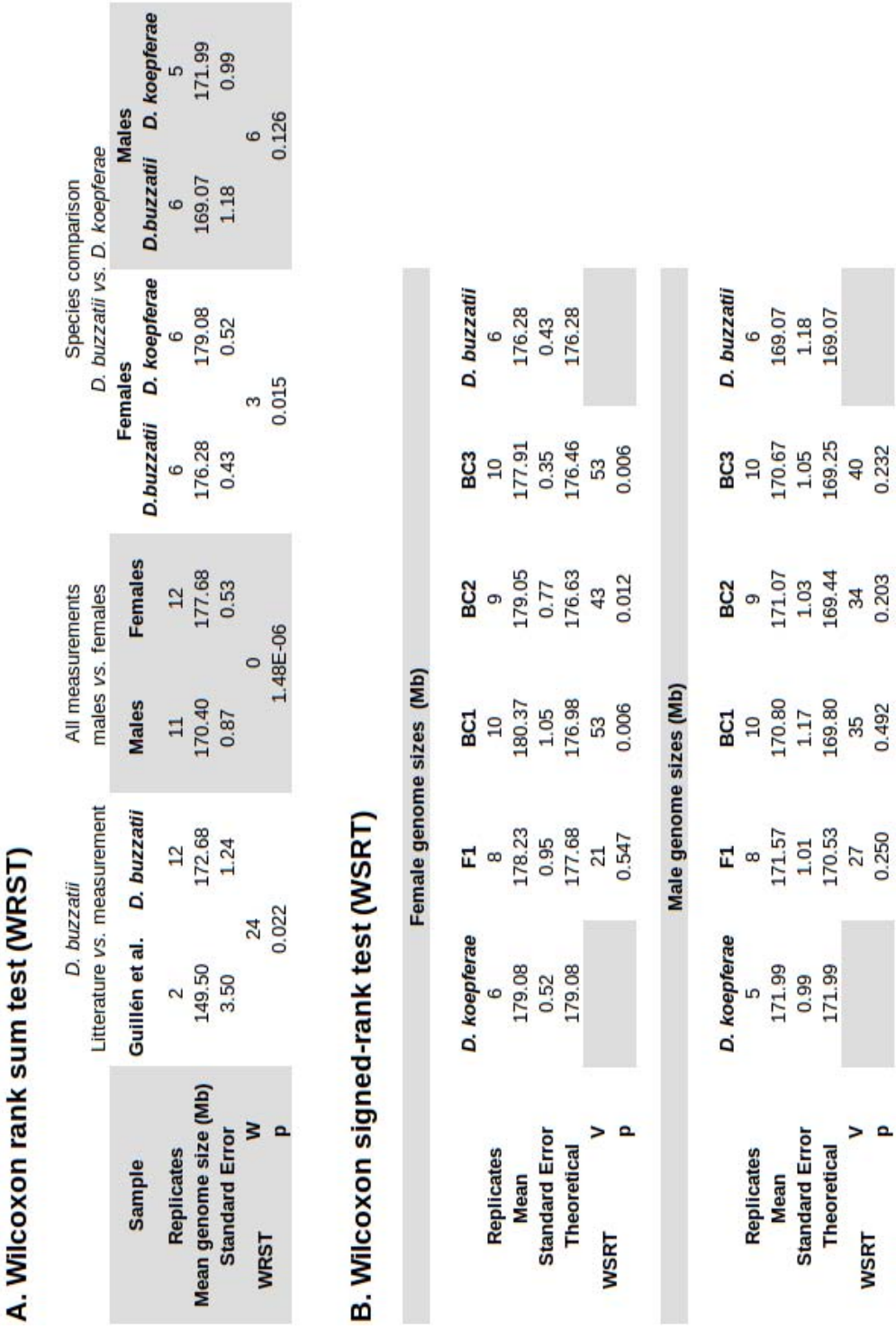




cross	number of markers	
all	Dk-sp	5
13.1	F1	5
	BC1	4
13.2	F1	5
	BC1	3
40.1	F1	5
	BC1	4
40.2	F1	5
	BC1	5
10	F1	5
	BC1	5
1	F1	5
	BC1	3
mean	F1	5
	BC1	4

8.1.2 Table S1 – Genome size comparisons statistics

Table S1. A. Parental samples genome size comparisons. *W:* Wilcoxon rank sum test, *p:* probability. **B. Comparisons of hybrid genome sizes to the theoretical mean value.** *V:* Wilcoxon signed-rank test, *p:* probability.



A. Wilcoxon rank sum test (WRST)

	D. buzzatii Litterature *vs.* measurement		All measurements males *vs.* females		Species comparison *D. buzzatii vs. D. koepferae*			
					Females		Males	
Sample	Guillén et al.	*D. buzzatii*	Males	Females	*D.buzzatii*	*D. koepferae*	*D.buzzatii*	*D. koepferae*
Replicates	2	12	11	12	6	6	6	5
Mean genome size (Mb)	149.50	172.68	170.40	177.68	176.28	179.08	169.07	171.99
Standard Error	3.50	1.24	0.87	0.53	0.43	0.52	1.18	0.99
W	24		0		3		6	
WRST p	0.022		1.48E-06		0.015		0.126	

B. Wilcoxon signed-rank test (WSRT)

Female genome sizes (Mb)							
	D. koepferae	F1	BC1	BC2	BC3	*D. buzzatii*	
Replicates	6	8	10	9	10	6	
Mean	179.08	178.23	180.37	179.05	177.91	176.28	
Standard Error	0.52	0.95	1.05	0.77	0.35	0.43	
Theoretical	179.08	177.68	176.98	176.63	176.46	176.28	
V		21	53	43	53		
WSRT p		0.547	0.006	0.012	0.006		

Male genome sizes (Mb)							
	D. koepferae	F1	BC1	BC2	BC3	*D. buzzatii*	
Replicates	5	8	10	9	10	6	
Mean	171.99	171.57	170.80	171.07	170.67	169.07	
Standard Error	0.99	1.01	1.17	1.03	1.05	1.18	
Theoretical	171.99	170.53	169.80	169.44	169.25	169.07	
V		27	35	34	40		
WSRT p		0.250	0.492	0.203	0.232		

8.1.3 Table S2 – AFLP genotyping summary

Table S2. Number of *D. koepferae*-specific (Dk-sp) AFLP markers found in F1 and BC1 hybrids for each primer combination and cross.

cross	generation	primer combination										total	observed proportion	pooled flies in BC1	BC1 mean proportion per individual *
		TG-GGG	TG-GCA	CG-CAT	GG-CGG	TG-GCG	GG-CAT	GT-CTC	CG-GCA	GT-CGG	TG-GCC				
13.1	Dk-sp	7	9	7	5	5	7	10	6	7	7	70	100.0%		
	F1	7	7	5	5	5	7	10	5	7	7	65	92.9%	3	17.7%
	BC1	3	3	1	4	1	3	3	4	5	4	31	44.3%		
13.2	Dk-sp	7	9	7	5	5	7	10	6	7	7	70	100.0%		
	F1	7	9	5	5	5	7	10	6	7	7	68	97.1%	3	11.8%
	BC1	3	1	2	3	1	2	3	2	3	2	22	31.4%		
40.1	Dk-sp	7	9	7	5	5	7	10	6	7	7	70	100.0%		
	F1	7	8	7	5	5	6	10	6	7	7	68	97.1%	4	13.6%
	BC1	2	0	3	4	1	4	6	4	4	3	31	44.3%		
40.2	Dk-sp	7	9	7	5	5	7	10	6	7	7	70	100.0%		
	F1	7	7	7	5	5	6	10	6	6	7	66	94.3%	4	21.2%
	BC1	4	2	7	5	1	5	6	5	4	4	43	61.4%		
10	Dk-sp	7	9	7	5	5	7	10	6	7	7	70	100.0%		
	F1	7	9	6	5	4	6	10	6	6	7	66	94.3%	1	57.1%
	BC1	4	4	4	5	3	3	7	4	2	4	40	57.1%		
1	Dk-sp	7	9	7	5	5	7	10	6	7	7	70	100.0%		
	F1	7	9	5	5	5	7	10	6	6	6	66	94.3%	1	72.9%
	BC1	6	5	4	3	4	6	8	4	4	7	51	72.9%		

***** The proportion of Dk-sp markers found in BC1 depends on the number of pooled flies: the more flies in a pool, the more markers observed, following the principle of inclusion-exclusion. In order to make samples with different number of pooled individuals comparable, we estimated the mean proportion of markers per individual and cross, using the following formula:

$$observed\ proportion = \sum_{i=1}^{n} p_i - \sum_{1 \le i \le j \le n}^{n} p_i \times p_j + \sum_{1 \le i \le j \le k \le n}^{n} p_i \times p_j \times p_k - \ ... + (-1)^{n+1}(p_1 \times ... \times p_n)$$

where the *observed proportion* is the proportion of Dk-sp markers found in BC1 for a pool of *n* flies (see table) and *p* is the mean proportion of Dk-sp markers for a single fly. We consider that $p_1 = p_2 = \cdots = p_n = p$ (as individual flies are not distinguishable in the pool) and that results are independent for each individual ($P(A \cap B) = P(A) \times P(B)$).

For example, for **cross 13.1**, a total of 3 flies were pooled for the analyses:

$$observed\ proportion = 3p - 3p^2 + p^3 \quad \xrightarrow{yields} \quad 0.443 = 3p - 3p^2 + p^3 \quad \xrightarrow{yields} \quad p = 0.1772$$

In the same way, for **cross 40.2**, a total of 4 flies were pooled for the analyses:

$$observed\ proportion = 4p - 6p^2 + 4p^3 - p^4 \quad \xrightarrow{yields} \quad 0.614 = 4p - 6p^2 + 4p^3 - p^4 \quad \xrightarrow{yields} \quad p = 0.2118$$

8.1.4 Table S3 – Genome size measurements

Table S3. Summary of all measurements and genome size data. IP=propidium iodure fluorescence, GS=genome size. GS units have been converted as follows: 1 pg= 978 Mb.

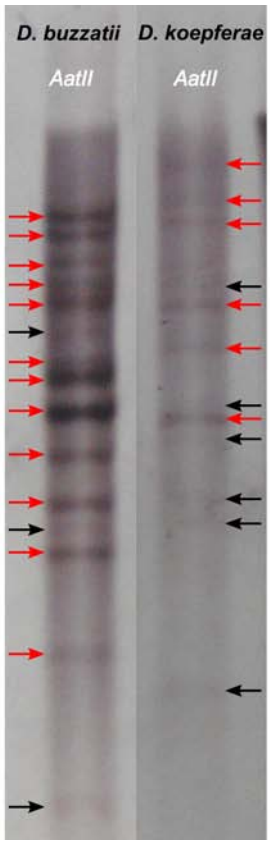
Species	Sample	IP D. virilis	GS D. Virilis (pg)	IP sample	GS sample (pg)	GS sample (Mb)
D. buzzatii ♀♀	Bu 1 F	133.045	0.340	70.853	0.181	177.083
	Bu 2 F	132.104	0.340	70.583	0.182	177.665
	Bu 3 F	135.256	0.340	71.578	0.180	175.971
	Bu 4 F	160.524	0.340	84.324	0.179	174.674
	Bu 5 F	161.391	0.340	85.297	0.180	175.741
	Bu 6 F	162.962	0.340	86.534	0.181	176.571
D. buzzatii ♂♂	Bu 1 M	128.720	0.340	66.232	0.175	171.096
	Bu 2 M	133.192	0.340	66.732	0.170	166.600
	Bu 3 M	130.850	0.340	67.612	0.176	171.818
	Bu 4 M	154.829	0.340	79.803	0.175	171.390
	Bu 5 M	149.365	0.340	75.798	0.173	168.743
	Bu 6 M	153.108	0.340	75.877	0.168	164.790
D. koepferae ♀♀	Ko 1 F	130.945	0.340	70.880	0.184	179.992
	Ko 2 F	131.129	0.340	70.913	0.184	179.823
	Ko 3 F	133.457	0.340	72.029	0.184	179.467
	Ko 4 F	155.830	0.340	84.101	0.183	179.460
	Ko 5 F	153.570	0.340	81.546	0.181	176.569
	Ko 6 F	155.759	0.340	83.928	0.183	179.173
D. koepferae ♂♂	Ko 1 M	128.845	0.340	67.664	0.179	174.626
	Ko 2 M	126.760	0.340	66.013	0.177	173.167
	Ko 3 M	128.205	0.340	66.600	0.177	172.738
	Ko 4 M	147.013	0.340	74.975	0.173	169.582
	Ko 5 M	146.993	0.340	75.069	0.174	169.817
F1 hybrids ♀♀	F1 1 F	132.807	0.340	71.540	0.183	179.121
	F1 2 F	133.805	0.340	72.048	0.183	179.047
	F1 3 F	134.248	0.340	73.051	0.185	180.941
	F1 4 F	134.370	0.340	72.005	0.182	178.188
	F1 5 F	133.282	0.340	72.686	0.185	181.341
	F1 6 F	157.280	0.340	81.980	0.177	173.321
	F1 7 F	153.659	0.340	82.449	0.182	178.421
	F1 8 F	155.303	0.340	81.952	0.179	175.468
F1 hybrids ♂♂	F1 1 M	130.113	0.340	68.597	0.179	175.308
	F1 2 M	129.379	0.340	67.482	0.177	173.437
	F1 3 M	128.037	0.340	66.736	0.177	173.318
	F1 4 M	133.715	0.340	67.484	0.172	167.818
	F1 5 M	130.456	0.340	67.203	0.175	171.294
	F1 6 M	152.688	0.340	77.378	0.172	168.512
	F1 7 M	149.116	0.340	75.725	0.173	168.862
	F1 8 M	146.375	0.340	76.614	0.178	174.044
BC1 hybrids ♀♀	BC1 1 F	134.253	0.340	73.414	0.186	181.833
	BC1 2 F	133.709	0.340	72.835	0.185	181.133
	BC1 3 F	133.164	0.340	73.323	0.187	183.093
	BC1 4 F	134.788	0.340	72.062	0.182	177.776
	BC1 5 F	132.760	0.340	70.348	0.180	176.199
	BC1 6 F	132.285	0.340	70.564	0.181	177.374
	BC1 7 F	133.167	0.340	72.078	0.184	179.980
	BC1 8 F	153.776	0.340	82.349	0.182	178.069
	BC1 9 F	157.053	0.340	88.573	0.192	187.531
	BC1 10 F	154.970	0.340	84.240	0.185	180.754

BC1 hybrids ♂♂	BC1 1 M	127.373	0.340	66.348	0.177	173.208
	BC1 2 M	137.038	0.340	67.810	0.168	164.540
	BC1 3 M	132.798	0.340	66.946	0.171	167.630
	BC1 4 M	135.707	0.340	67.578	0.169	165.585
	BC1 5 M	130.827	0.340	67.587	0.176	171.784
	BC1 6 M	129.181	0.340	67.822	0.179	174.578
	BC1 7 M	129.440	0.340	67.739	0.178	174.016
	BC1 8 M	148.540	0.340	75.947	0.174	170.014
	BC1 9 M	147.763	0.340	77.426	0.178	174.236
	BC1 10 M	146.689	0.340	76.069	0.176	172.436
BC2 hybrids ♀♀	BC2 1 F	135.184	0.340	72.862	0.183	179.223
	BC2 2 F	133.039	0.340	71.240	0.182	178.059
	BC2 3 F	133.284	0.340	70.424	0.180	175.695
	BC2 4 F	133.384	0.340	71.238	0.182	177.593
	BC2 5 F	134.545	0.340	72.226	0.183	178.502
	BC2 6 F	153.022	0.340	81.403	0.181	176.890
	BC2 7 F	151.491	0.340	82.524	0.185	181.139
	BC2 8 F	155.179	0.340	84.992	0.186	182.122
	BC2 9 F	158.963	0.340	87.111	0.186	182.219
BC2 hybrids ♂♂	BC2 1 M	130.098	0.340	67.298	0.176	172.008
	BC2 2 M	131.314	0.340	66.615	0.172	168.686
	BC2 3 M	131.751	0.340	67.963	0.175	171.529
	BC2 4 M	130.323	0.340	67.500	0.176	172.227
	BC2 5 M	129.373	0.340	67.325	0.177	173.042
	BC2 6 M	145.807	0.340	73.106	0.170	166.722
	BC2 7 M	156.506	0.340	78.330	0.170	166.424
	BC2 8 M	148.702	0.340	78.052	0.178	174.536
	BC2 9 M	146.668	0.340	76.954	0.178	174.467
BC3 hybrids ♀♀	BC3 1 F	132.031	0.340	70.645	0.182	177.919
	BC3 2 F	132.090	0.340	70.591	0.182	177.704
	BC3 3 F	133.098	0.340	70.716	0.181	176.670
	BC3 4 F	132.630	0.340	70.076	0.180	175.689
	BC3 5 F	133.516	0.340	71.332	0.182	177.651
	BC3 6 F	132.401	0.340	71.421	0.183	179.371
	BC3 7 F	150.681	0.340	81.238	0.183	179.274
	BC3 8 F	150.674	0.340	80.717	0.182	178.133
	BC3 9 F	152.120	0.340	81.693	0.183	178.573
	BC3 10 F	152.704	0.340	81.815	0.182	178.156
BC3 hybrids ♂♂	BC3 1 M	128.366	0.340	65.793	0.174	170.431
	BC3 2 M	127.564	0.340	65.798	0.175	171.515
	BC3 3 M	133.442	0.340	66.306	0.169	165.226
	BC3 4 M	132.028	0.340	67.020	0.173	168.794
	BC3 5 M	128.643	0.340	66.133	0.175	170.942
	BC3 6 M	128.255	0.340	65.938	0.175	170.954
	BC3 7 M	128.357	0.340	66.769	0.177	172.971
	BC3 8 M	149.016	0.340	78.457	0.179	175.072
	BC3 9 M	147.279	0.340	77.476	0.179	174.922
	BC3 10 M	148.952	0.340	74.299	0.170	165.865

8.2 Supplementary data of "Expression of the retrotransposon *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids" (chapter 3.2)

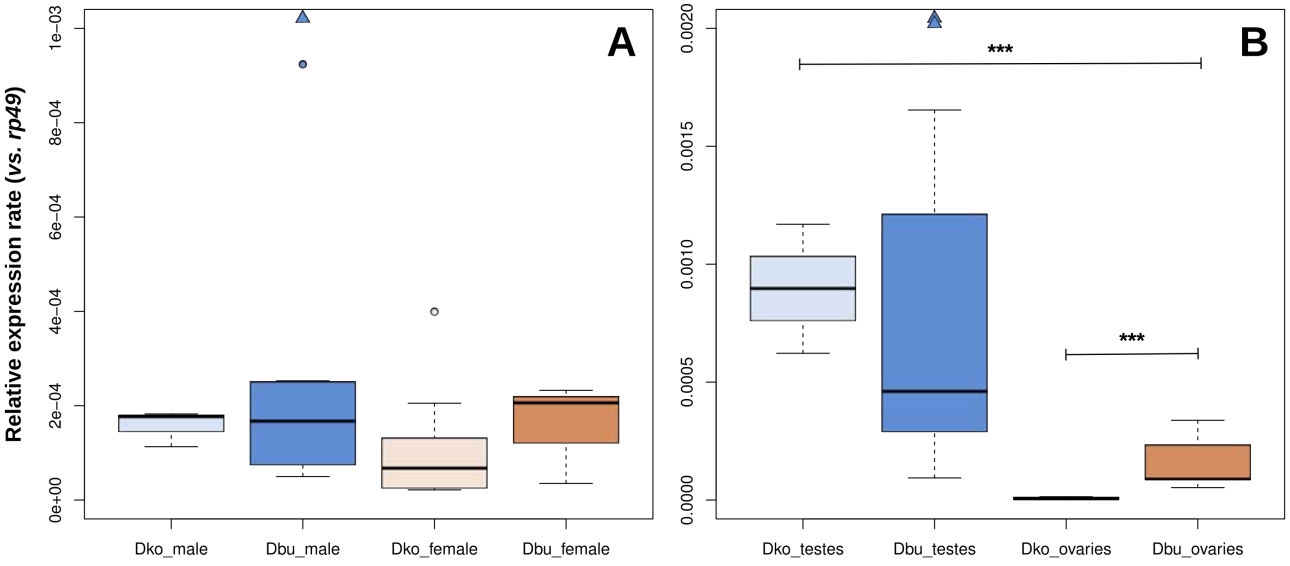
8.2.1 S1 Fig – *Helena* Southern blot

S1 Fig. Southern blot analysis of *Helena* in parental species, *D. buzzatii* (left) and *D. koepferae* (right). No restriction sites for *AatII* are present in *Helena*'s probe sequence. Thus, digestions with this enzyme allow us to distinguish different *Helena* copies. Arrows in red indicate strong-signaled bands; arrows in black indicate faint bands.



8.2.2 S1 File – Expression in parental species

S1 File. *Helena* expression results in parental species. (Fig A and B) *Helena* expression rates relative to *rp49* housekeeping gene in *D. koepferae* (Dko) and *D. buzzatii* (Dbu) somatic tissues **(A)** and gonads **(B)**. Male samples are represented in blue and female samples are represented in brown. Boxes are determined by the first and third quartile values, with an intermediate deep line corresponding to the median value. Circles correspond to outliers (above or below 1.5-fold the interquartile range), and triangles represent those outliers whose ERs are extremely outranged and cannot be represented in the same scale (triangle in A: ER = 2.9×10^{-3}, in B: ER = 3.6×10^{-3} and 6.2×10^{-3}). **(Fig C and D)** Comparison of *Helena* expression rates between all different parental samples for somatic tissues **(C)** and gonads **(D)**. N = number of replicates analyzed, SD = standard deviation, W = Wilcoxon rank sum test statistic, p-value = probability. *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001. In red, p-values that are significant after Bonferoni correction (p-value<0.008).



C

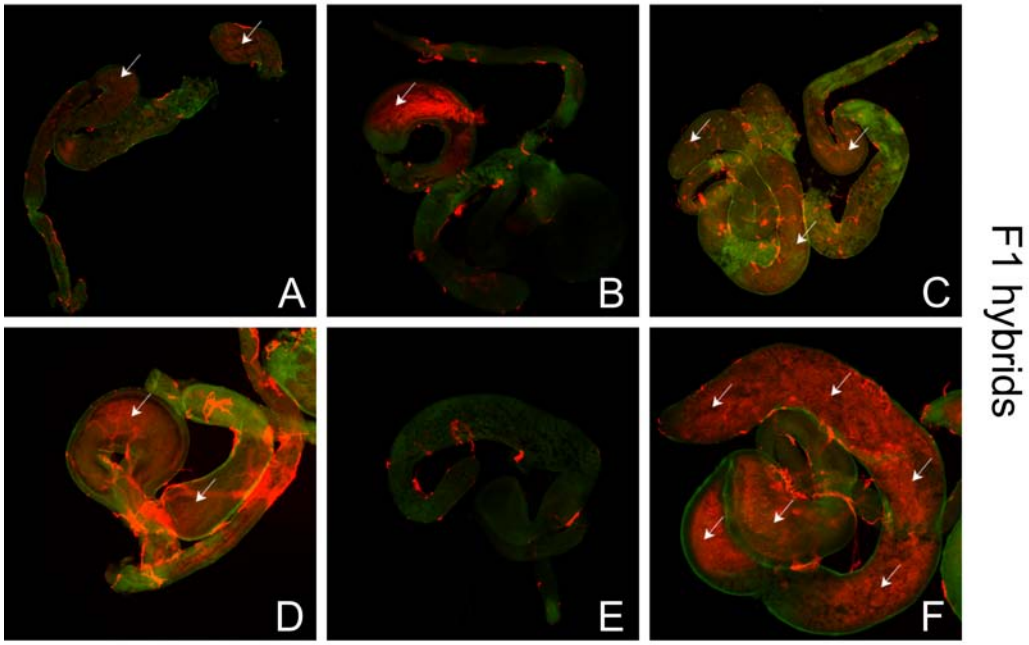
					males				females			
					D. buzzatii		*D. koepferae*		*D. buzzatii*		*D. koepferae*	
		N	median	SD	W	p-value	W	p-value	W	p-value	W	p-value
males	*D. buzzatii*	11	1.68E-04	8.40E-04			18	8.85E-01	14	7.69E-01	31	1.75E-01
	D. koepferae	3	1.77E-04	3.87E-05					6	7.00E-01	7	2.82E-01
females	*D. buzzatii*	3	2.07E-04	1.07E-04							8	3.73E-01
	D. koepferae	9	6.82E-05	1.22E-04								

D

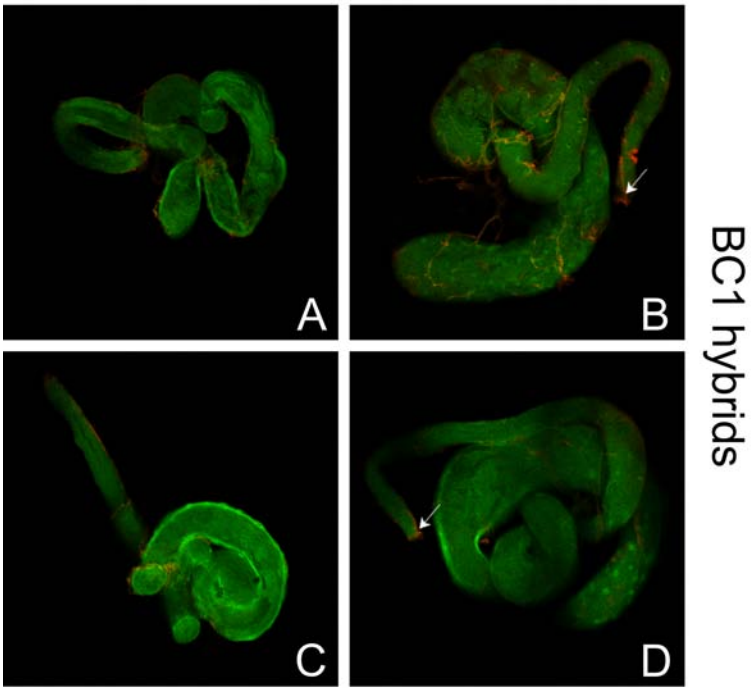
					testes				ovaries			
					D. buzzatii		*D. koepferae*		*D. buzzatii*		*D. koepferae*	
		N	median	SD	W	p-value	W	p-value	W	p-value	W	p-value
testes	*D. buzzatii*	11	4.62E-04	1.91E-03			23	3.68E-01	5	8.70E-03**	0	5.67e-06***
	D. koepferae	3	8.99E-04	2.74E-04					0	3.57E-02*	0	6.99E-03**
ovaries	*D. buzzatii*	5	9.00E-05	1.22E-04							0	6.66E-04***
	D. koepferae	10	5.50E-06	4.80E-06								

8.2.3 S2 File – *Helena* localization in F1 testes

S2 File. FISH of *Helena* RNA expression in different F1 hybrid testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts.



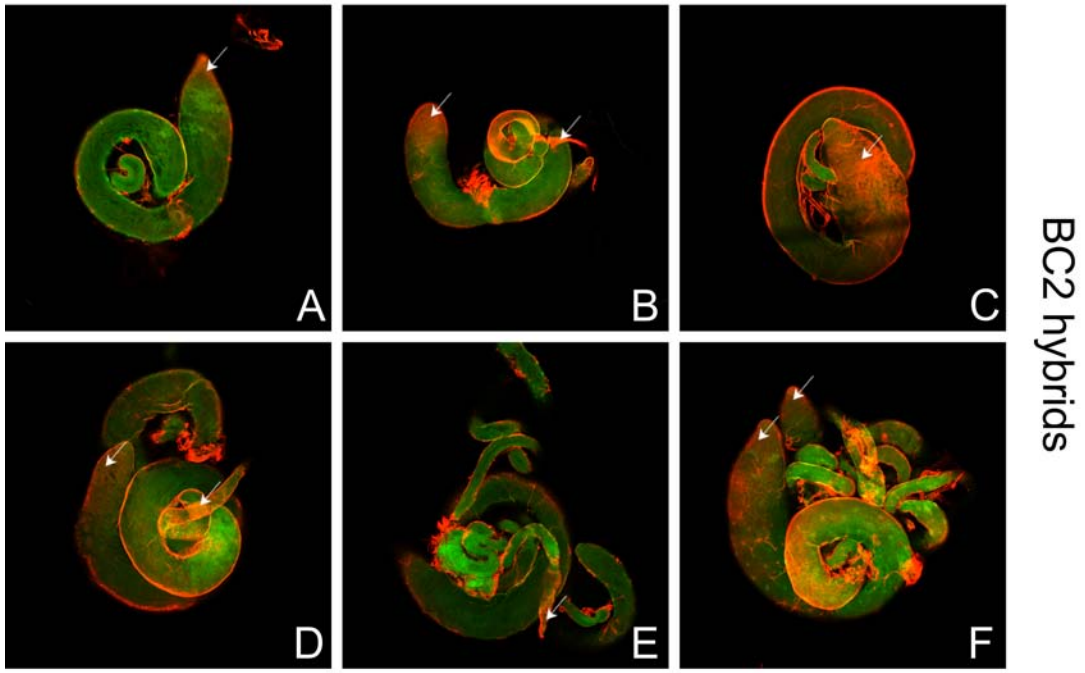
8.2.4 S3 File – *Helena* localization in BC1 testes



S3 File. FISH of *Helena* RNA expression in different BC1 hybrid testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts.

8.2.5 S4 File – *Helena* localization in BC2 testes

S4 File. FISH of *Helena* RNA expression in different BC2 hybrid testes. Red staining are *Helena* transcripts, green staining is tissue autofluorescence. Arrows mark the presence of *Helena* transcripts.



8.2.6 S1 Table – BLAST alignments

S1 Table. Summary of BLAST alignment results between *Helena* sequenced copies. *Dbu = D. buzzatii*, *Dko28 = D. koepferae*-28, *Dko35-1 = D. koepferae*-35-1, *Dko35-2 = D. koepferae*-35-2.

Species	Sequence Identity	E-value	Max Score
Dbu* vs. *Dko28	98%	0.0	4823
Dbu* vs. *Dko35-1	89%	0.0	3472
Dbu* vs. *Dko35-2	97%	0.0	4275
Dko28* vs. *Dko35-1	90%	0.0	3517
Dko28* vs. *Dko35-2	98%	0.0	4405
Dko35-1* vs. *Dko35-2	93%	0.0	4848

8.2.7 S2 Table – Variance comparison statistics

S2 Table. Variance comparisons of *Helena* expression rates between each hybrid generation and parental species. W = Levene's test for equality of variances satistic, p-value = probability. *: p-value<0.05, **: p-value<0.01, ***: p-value<0.001. In red, p-values that are significant after Bonferoni correction (p-value<0.01). Each kind of sample (males, females, testes, ovaries) has been compared to the same tissue of both parental species.

		variance	vs.*D. buzzatii*		vs.*D. koepferae*	
			W	p-value	W	p-value
males	*D. koepferae*	1.50E-09	1.72	2.14E-01	-	-
	D. buzzatii	7.06E-07	-	-	1.72	2.14E-01
	F1	3.28E-04	11.06	**4.61E-03****	2.90	1.32E-01
	BC1	2.26E-07	1.41	2.50E-01	0.97	3.44E-01
	BC2	1.66E-06	0.27	6.06E-01	1.24	2.89E-01
	BC3	6.47E-08	2.52	1.30E-01	2.12	1.76E-01
females	*D. koepferae*	1.49E-08	1.56	1.89E-01	-	-
	D. buzzatii	1.15E-08	-	-	1.56	1.89E-01
	F1	3.70E-08	0.80	3.91E-01	1.53	2.33E-01
	BC1	7.20E-09	0.47	5.05E-01	0.99	3.32E-01
	BC2	5.34E-09	0.98	3.41E-01	1.51	2.33E-01
	BC3	4.01E-08	0.19	6.73E-01	0.38	5.44E-01
testes	*D. koepferae*	7.48E-08	2.43	1.45E-01	-	-
	D. buzzatii	3.64E-06	-	-	2.43	1.45E-01
	F1	7.21E-09	5.06	**4.11E-02***	2.85	1.42E-01
	BC1	1.24E-07	7.64	**1.20E-02***	0.99	3.40E-01
	BC2	3.54E-07	6.71	**1.63E-02***	1.17	2.96E-01
	BC3	6.53E-08	8.13	**1.02E-02***	0.08	7.89E-01
ovaries	*D. koepferae*	2.30E-11	48.20	**1.02E-05*****	-	-
	D. buzzatii	1.48E-08	-	-	48.20	**1.02E-05*****
	F1	1.70E-09	23.45	**2.26E-06*****	9.83	**5.21E-03****
	BC1	1.04E-09	27.03	**8.79E-05*****	17.18	**4.60E-04*****
	BC2	7.02E-10	33.03	**3.00E-05*****	11.52	**2.73E-03****
	BC3	5.97E-06	1.65	2.18E-01	3.98	6.00E-02

8.2.8 S1 Text – *Helena* alignment (fasta)

S1 Text. Alignment of *Helena* sequences (in fasta format) obtained with MAFFT E-INS-i algorithm and cleaned using Gblocks. This alignment was used to construct the phylogenetic tree on Fig 2 (see **section 3.2.1**).

This file is available online at http://journals.plos.org/plosone/article/asset?unique&id=info:doi/10.1371/journal.pone.0147903.s008 or at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.2**, file **S1_Text.pdf**).

8.2.9 S2 Text – *Helena* alignment (graphical)

S2 Text. Graphical representation of the *Helena* alignment obtained with MAFFT E-INS-i algorithm and cleaned using Gblocks. Highly conserved residues (similarity score per position > 0.5) are framed in blue and used to build the consensus sequence. Each nitrogenous base in a conserved position is represented in a different colour.

This file is available online at http://journals.plos.org/plosone/article/asset?unique&id=info:doi/10.1371/journal.pone.0147903.s009 or at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.2**, file **S2_Text.pdf**).

8.2.10 S3 Text – Expression measurements (ΔC_T values)

S3 Text. Summary of ΔC_T values for all studied replicates (from different crosses) of each kind of sample for all generations.

sample	generation	cross	Δct
males	buzzatii	A	14,07
males	buzzatii	B	11,95
males	buzzatii	C	14,05
males	buzzatii	E	12,95
males	buzzatii	F	14,29
males	buzzatii	G	12,47
males	buzzatii	H	12,54
males	buzzatii	I	11,97
males	buzzatii	J	13,43
males	buzzatii	L	8,44
males	buzzatii	M	10,08
males	F1	A	11,51
males	F1	B	14,49
males	F1	C	11,96
males	F1	D	4,49
males	F1	EFJMN	16,55
males	F1	GHI	13,45
males	BC1	A	11,68
males	BC1	B	11,56
males	BC1	C	13,24
males	BC1	D	12,82
males	BC1	E	9,19
males	BC1	H	15,39
males	BC1	I	11,87
males	BC1	J	12,16
males	BC1	K	12,62
males	BC1	M	14,62
males	BC1	N	13,06
males	BC2	A	11,76
males	BC2	B	10,3
males	BC2	C	10,36
males	BC2	D	10,41
males	BC2	E	13,07
males	BC2	H	13,56
males	BC2	I	12,03
males	BC2	J	13,52
males	BC2	K	7,85
males	BC2	M	12,25
males	BC3	A	11,74
males	BC3	B	11,88
males	BC3	C	11,07
males	BC3	D	13,86
males	BC3	E	12,06
males	BC3	F	14,3
males	BC3	G	12,74
males	BC3	H	17,32
males	BC3	M	10,25
males	koepferae		12,42
males	koepferae		12,46
males	koepferae		13,11

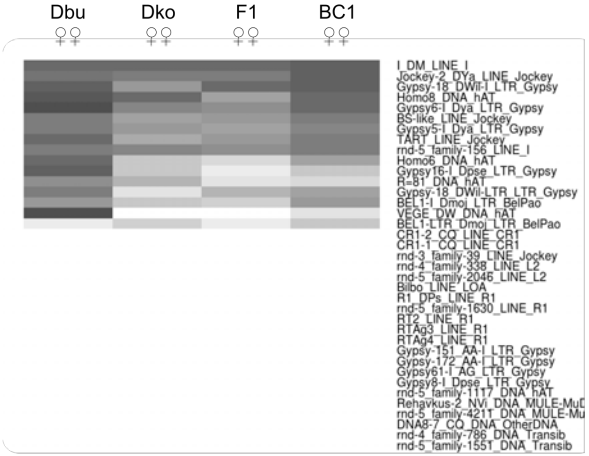
sample	generation	cross	Δct
females	koepferae	A	15,3
females	koepferae	B	15,5
females	koepferae	F	15,26
females	koepferae	H	12,89
females	koepferae	J	13,84
females	koepferae	K	13,98
females	koepferae	L	13,24
females	koepferae	M	12,25
females	koepferae	N	11,29
females	F1	A	13,47
females	F1	B	13,76
females	F1	C	13,29
females	F1	D	13,5
females	F1	E	12,34
females	F1	G	12,14
females	F1	K	10,7
females	F1	M	13,58
females	F1	N	11,07
females	BC1	B	12,94
females	BC1	C	13,38
females	BC1	D	13,07
females	BC1	E	12,89
females	BC1	F	14,44
females	BC1	G	15,02
females	BC1	H	13,26
females	BC1	I	12,9
females	BC1	J	11,55
females	BC1	K	12,03
females	BC1	L	13,22
females	BC1	M	12,64
females	BC1	N	15,73
females	BC2	A	13,08
females	BC2	B	12,73
females	BC2	C	12,69
females	BC2	D	11,79
females	BC2	E	13,64
females	BC2	FG	12,84
females	BC2	I	12,87
females	BC2	J	13,26
females	BC2	K	12,9
females	BC2	L	16,12
females	BC2	M	14,18
females	BC2	N	17,41
females	BC3	A	12,47
females	BC3	B	12,47
females	BC3	C	12,63
females	BC3	D	14,76
females	BC3	E	14,21
females	BC3	F	15,28
females	BC3	G	13,97
females	BC3	H	15,02
females	BC3	J	12,21
females	BC3	L	10,49
females	buzzatii		14,8
females	buzzatii		12,07
females	buzzatii		12,24

8.3 Supplementary data of “Divergence in piRNA pathway effector proteins partially explains *Drosophila buzzatii–D. koepferae* hybrid instability” (chapter 3.3)

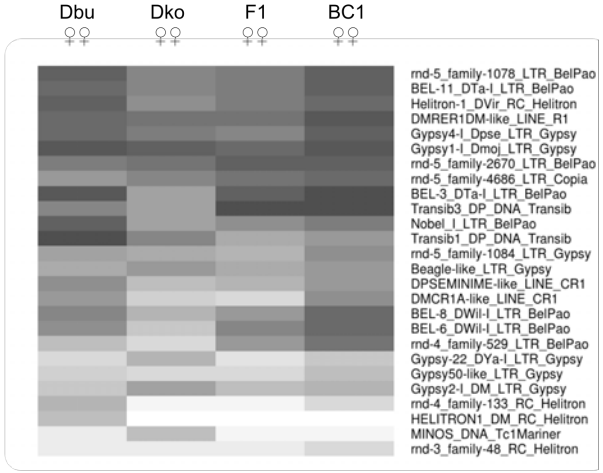
8.3.1 Figure S1 – Ping-pong fraction of deregulated TEs

Figure S1. Ping-pong fraction of ovarian piRNA populations associated to deregulated TE families. **(A)** Overexpressed in F1. **(B)** Underexpressed genes in F1. **(C)** Overexpressed genes in BC1. **(D)** Underexpressed genes in BC1.

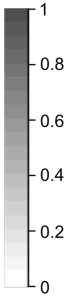
A. F1-overexpressed TEs



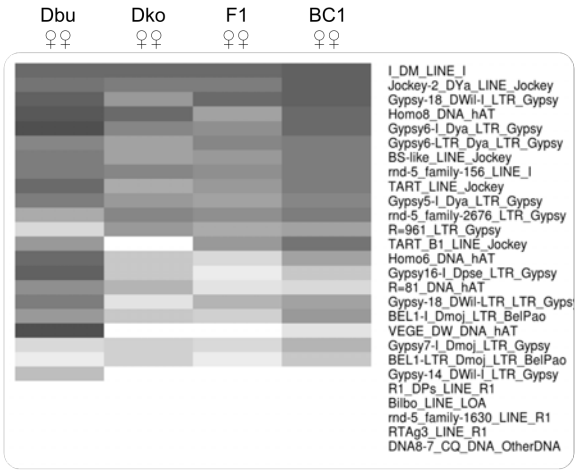
B. F1-underexpressed TEs



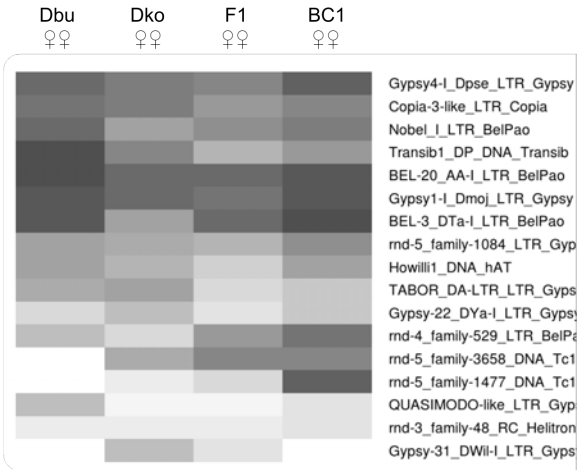
ping-pong signal



C. BC1-overexpressed TEs



D. BC1-underexpressed TEs



8.3.2 File S1 – RNA-seq statistics

Supplementary file 1: RNA-seq statistics summary. (A) Number of reads at each analysis step. **(B)** Raw read count per TE family after alignment to the TE library. **(C)** Read count per TE family after normalization by DESeq2.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file1-4.xls**).

8.3.3 File S2 – Gene expression analyses

Supplementary file 2: Deregulated genes in ovaries. FC= Fold Change; BH= Bonferroni-Hochberg**. (A)** Overexpressed genes in F1. **(B)** Overexpressed genes in BC1. **(C)** Underexpressed genes in F1. **(D)** Underexpressed genes in BC1.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file1-4.xls**).

8.3.4 File S3 – Divergence time analyses

Supplementary file 3: Summary of codeml results. Rate of substitution per non-synonymous site (dN) and per synonymous site (dS) for each *D. buzzati-D. koepferaee* contig pair.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file1-4.xls**).

8.3.5 File S4 – TE landscapes analyses

Supplementary file 4: Summary of dnaPipeTE results. Read count and proportion (%) of each class of repetitive sequences for *D. buzzatii* and *D. koepferae* genomic reads.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file1-4.xls**).

8.3.6 File S5 – small RNA populations statistics

Supplementary file 5: small RNA population sequencing statistics summary. (A) Number of reads at each analysis step. **(B)** Raw piRNA read count per TE family after alignment to the TE library. **(C)** piRNA read count per TE family after normalization by DESeq2.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file5-6.xls**).

8.3.7 File S6 – piRNA populations analyses

Supplementary file 6: TE families with notable differences (≥2-fold) in their piRNA populations in hybrid ovaries (F1 or BC1) compared to both parental species. FC= Fold Change. **(A)** Lower piRNA levels in parents. **(B)** Lower piRNA levels in hybrids.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file5-6.xls**).

8.3.8 File S7 – TE expression in testes

Supplementary file 7: Differential expression of TEs in F1 testes compared to *D. buzzatii*. FC= Fold Change; BH= Bonferroni-Hochberg**. (A)** Overexpressed TE families in F1. **(B)** Underexpressed TE families in F1. **(C)** TE families with lower piRNA abundance in F1. **(D)** TE families with higher piRNA abundance in F1.

This file is available online at https://www.dropbox.com/sh/1bp777lptc6ty9k/AAAc69-IOvFXdLLF4GdodJdna?dl=0 (Folder **Annex 8.3**, file **supp_file7.xls**).

8.3.9 Table S1 – Differential expression analyses

Supplementary table 1: Differential expression summary. Dbu= *D. buzzatii*, Dko= *D. koepferae*. Above the main diagonal (grey), number of TE families with significant differential expression for each comparison. In parenthesis, fraction (%) of differentially expressed TE families of *column* sample showing overexpression (green) or underexpression (red) compared to the sample in *row*. Below the main diagonal, fraction of the differentially expressed families which present 1.5 fold or higher differences.

	Dbu testes	F1 testes	Dbu ovaries	Dko ovaries	F1 ovaries	BC1 ovaries
Dbu testes		256 (54.3/45.7)	303 (54.1/45.9)	325 (56.0/44.0)	325 (52.3/47.7)	302 (50.3/49.7)
F1 testes	89.8%		304 (52.6/47.4)	284 (57.0/43.0)	316 (51.6/48.4)	297 (50.8/49.2)
Dbu ovaries	92.1%	93.8%		284 (51.1/48.9)	221 (45.3/54.8)	149 (46.3/53.7)
Dko ovaries	96.0%	90.1%	90.5%		234 (44.9/55.1)	254 (44.1/55.9)
F1 ovaries	93.5%	97.5%	88.2%	87.6%		92 (46.7/53.3)
BC1 ovaries	91.4%	96.3%	79.9%	89.4%	75%	