University of Thessaly

School of Health Sciences

Department of Biochemistry and Biotechnology

MSc. Advanced Experimental and Computational Biosciences

Genomic analysis of the Y chromosome from the major olive fruit insect pest, *Bactrocera oleae.*

Rallis Dimitris

Larissa

2022

Τίτλος:
Γονιδιωματική ανάλυση του Υ χρωμοσώματος από το παράσιτο του καρπού της ελιάς, *Bactrocera oleae.*

Thesis advisory committee:

Mathiopoulos Kostas: *Professor, Department of Biochemistry and Biotechnology, University of Thessaly*

Giakountis Antonios: *Assistant Professor, Department of Biochemistry and Biotechnology, University of Thessaly*

Papanicolaou Alexie: *Associate Professor, Hawkesbury Institute for the Environment, Western Sydney University*

Abstract

The presence of two discrete sexes is a convergent characteristic of higher organisms, tightly linked to the evolution of sex chromosomes. Those elements have evolved independently multiple times and the study of their sequence, content and structure is important for unravelling the processes that drive sex chromosome formation. The order of Diptera have been proved great models in this effort, as a multitude of sex determination mechanisms and sex chromosome lineages exist among its members. Moreover, the economical and epidemiological impact that many Diptera have on human activity poses an exceptional interest in studying those elements, especially the sex-limited Y chromosome, as a necessity in the implementation of Integrated Pest Management (IPM) approaches. In this study, two novel methods for the characterization of Y-linked sequences were developed, benchmarked and employed on the novel assembly of the olive fruit pest *Bactrocera oleae* (Diptera: Tephritidae). Thorough analysis of Y regions revealed an RNA-mediated amplification of the karyopherin member *importin-4,* consistent to similar instances of nuclear transport machinery gene duplications observed in Drosophilidae. Those Y copies have acquired testes specific expression and possibly exert their function through the production of non-coding transcripts. The fertility role attributed to homologues of this gene in both *D. melanogaster* and rats implied a similar function in *B. oleae*, which is of great interest for IPM applications. Towards this, we attempted the characterization of a putative fertility role for *importin-4* gene in *B. oleae*, through the systemic knock-down of this gene's function by RNAi pathway activation.

Περίληψη

Η ύπαρξη δύο διακριτών φύλων είναι ένα κοινό χαρακτηριστικό των ανώτερων ευκαρυωτικών οργανισμών, το οποίο βρίσκεται σε άμεση συσχέτιση με την εξέλιξη των φυλετικών χρωμοσωμάτων. Τέτοια χρωμοσώματα έχουν εξελιχθεί ανεξάρτητα, πολλές φορές, ενώ η μελέτη της αλληλουχίας, του περιεχομένου, και της δομής τους δύναται να αποκαλύψει τις διαδικασίες που οδηγούν στον σχηματισμό τους. Η τάξη των Διπτέρων έχει αποδειχτεί ως ένα εξαιρετικό μοντέλο για την μελέτη των συγκεκριμένων γενετικών στοιχείων, καθώς εμπεριέχει μεταξύ των μελών της πλείστες διαφορετικές γενεαλογίες φυλετικών χρωμοσωμάτων και συστημάτων φυλοκαθορισμού. Επιπλέον, η δράση ορισμένων Διπτέρων επιφέρει οικονομικές και επιδημιολογικές επιπτώσεις στην ανθρώπινη δραστηριότητα, καθιστώντας την μελέτη των φυλετικών χρωμοσωμάτων, και συγκεκριμένα του αρρενοειδικού Υ, ως αντικείμενο εξέχουσας σημασίας για την εφαρμογή μεθόδων Ολοκληρωμένης Διαχείρισης Παρασίτων (ΟΔΠ). Στη συγκεκριμένη μελέτη αναπτύξαμε και αξιολογήσαμε δύο νέες μεθόδους για τον εντοπισμό Υ-ειδικών αλληλουχιών. Στη συνέχεια οι παραπάνω μέθοδοι εφαρμόστηκαν στο γονιδίωμα αναφοράς του εντόμου *Bactrocera oleae* (Diptera: Tephritidae), γνωστό και ως Δάκος, που αποτελεί παράσιτο του καρπού της ελιάς. Η εκτενής ανάλυση του Υ χρωμοσώματος αποκάλυψε την ύπαρξη πολλαπλών αντιγράφων ενός γονιδίου καρυοφερίνης, της *ιμπορτίνης-4*, η οποία διπλασιάστηκε στο Υ μέσω ενός RNA ενδιάμεσου μορίου. Παρόμοιες περιπτώσεις διπλασιασμού γονιδίων που εμπλέκονται στο σύστημα πυρηνικής μεταφοράς, και οι οποίες διαμεσολαβούνται από μόρια RNA, έχουν πρόσφατα χαρακτηριστεί στην οικογένεια των Δροσοφιλίδων. Τα συγκεκριμένα αντίγραφα που εντοπίζονται στο Υ χρωμόσωμα έχουν αποκτήσει αρρενοειδική έκφραση, συγκεκριμένα στους όρχεις του εντόμου, και πιθανώς δρουν μέσω της παραγωγής μη-κωδικοποιητικών RNA. Τα ομόλογα γονίδια *ιμπορτίνης-4* που εντοπίζονται στην *Drosophila melanogaster* και στα τρωκτικά εμπλέκονται στην γονιμότητα των συγκεκριμένων οργανισμών, ενώ ο χαρακτηρισμός μιας παρόμοιας λειτουργίας στο δάκο είναι εξέχουσας σημασίας για την εφαρμογή προσεγγίσεων ΟΔΠ. Προς την κατεύθυνση αυτή, επιχειρήσαμε την μελέτη του ρόλου της *ιμπορτίνης-4* στην γονιμότητα του Δάκου μέσω συστημικής μετα-μεταγραφικής σίγησης του συγκεκριμένου γονιδίου με αξιοποίηση του συστήματος RNAi.

[…] There is a war between the man and the woman.

There is a war between the ones who say there is a war,

and the ones who say there isn't […]

L. Cohen

Table of Contents

Introduction

Diptera is an order of arthropods, towards which extensive studies are focused. Many of its members possess great economical and epidemiological importance, as agronomic pests and disease vectors respectively. Therefore, studies are targeted to the containment of the destructive action through population reduction below an economic injury level in the former case, or minimization of infectious bites in the latter. The use of chemical insecticides for insect population control has been the golden standard for decades. However, their extensive application has raised major concerns regarding health and ecological impact. Additionally, instances of resistance to insecticides posed a threat to the efficiency of these core pest control tools, highlighting the need for more sustainable approaches that will make prudent use of chemical methods.

Integrated Pest Management (IPM) includes the careful consideration of all available pest control techniques and subsequent integration of appropriate measures that discourage the development of pest populations, while minimizing the risks to human health and the environment (Food and Agriculture Organization (FAO)). It consists of three interconnected areas: prevention, monitoring and intervention. IPM intervention targets the reduction of the field population through the employment of chemical, biological and genetic control methods. In this case, the use of chemical insecticides is judicious and acts auxiliary to natural enemies of the insect (biological method) or more advanced genetic control methods that target the reproductive potential of the population. The Sterile Insect Technique (SIT) is an alternative, eco-friendly and species-specific genetic control method introduced in 1995 by Knipling. Since then, it has been efficiently used for both population reduction and elimination purposes (De Longo 2000, Orankanok 2006, 2007). SIT is based on the rearing, sterilization and release of male-only insects. The principles of SIT have been the base for the development of more sophisticated genetic control methods like Release of Insects carrying a Dominant Lethal (RIDL) (Phuc 2007), Incompatible Insect Technique (IIT) (Pagendam 2020) and more recently the gene-drive based precision guided SIT (pgSIT) (Kandul 2019). Current advances in genomic technologies and molecular tools assist in the thorough genome analysis of insect pest species, towards a wide and efficient application of IPM programs for the containment of their destructive action.

This study initiates from the importance of Y chromosomes for sex determination and reproduction, two key elements of IPM intervention methods, and explores the sequence, content and functional role of Y chromosomes using the pest species *Bactrocera oleae* as an insect model.

Evolution of sex determination

The presence of two separate sexes is a common characteristic found throughout the tree of life and is particularly widespread in metazoans (Furman 2020). Sexual reproduction involves the mixing of two genomes through the fusion of haploid gametes that are the result of meiotic division of cells. When present, this reproductive mechanism confers the advantage of mixing genomic elements from two individuals, revealing an excessive genetic variation that will be otherwise hidden by asexual reproduction (Otto 2009). Sexual reproduction has been proven, through mathematical modelling, to inevitably lead to the production of gametes by higher multicellular organisms that differ in size between males and females, a process known as anisogamy (Bulmer & Parker 2002). Therefore, a distinct developmental program is required for each of the two sexes to produce the primary sexual characteristics onto which the future reproductive potential of the individual relies. Sex determination cascades are a convergent characteristic of higher organisms and are responsible for tightly controlling this process.

Although the outcome of sex determination is strongly conserved across species, the way that male and female developmental programs are initiated and regulated differs across taxa. Sex determination has evolved independently numerous times among plants and animals, which suggests that hermaphroditism poses an evolutionary burden to organisms. Focusing on Diptera, those are characterized by a plethora of different sex determination mechanisms, while the absence of hermaphroditism in them indicates that these different systems evolved from a common gonochoric ancestor (Blackmon 2016).

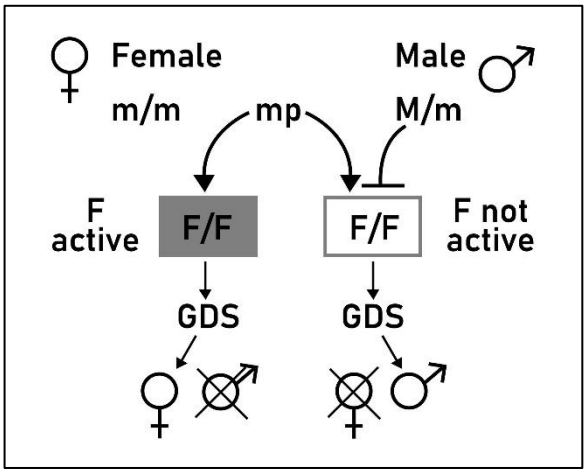

Figure 1 Classical four-element model of sex determination in dipteran insects. (Shearman 2002)

The classical model of sex determination in Diptera can be described in terms of four key elements (Figure 1): the primary signal (M), maternal products (mp), the key gene (F) and a genetic double switch (GDS). In this scheme the primary signal (M), together with the maternal products (mp), regulates the activity state of the key gene (F), which in turn sets the GDS to either male or female mode (Shearman 2002). Despite the conservation of this mode, differences exist in the genetic composition of sex determination cascades, which imply the fast evolution of those mechanisms even in the same clade. Intriguingly, the acquisition of a different role in the cascade for the key *transformer (tra)* gene is observed between *Drosophilidae* and *Muscidae* (Figure 2). In the former case, *tra* is alternatively spliced based on the isoform of the upstream *sex lethal* gene (*sxl*), while in the latter, it has evolved as the master switch gene in several members of *Musca domestica* (Hediger 2010). At the same time, *sxl* is present in studied members of Tephritidae insects like *C. capitata*, but with a paralog character (Saccone 1998). The final gene in *D. melanogaster* cascade is the *doublesex (dsx)* that acts as a GDS and is alternatively spliced depending on the activation or not of upstream elements of the cascade. Both sex specific isoforms of *dsx* are functional and differ in their C-terminal protein domain. This functional dimorphism leads to the respective downstream activation of genes responsible for shaping the primary sexual characteristics of the individual (Pomiankowski 2004). The role of *dsx* as the base of the sex determination cascade is highly conserved across several Diptera taxa and genes with structural similarity to *dsx* are found even outside arthropods, in distantly related organisms like *Caenorhabditis elegans* (Lints & Emmons 2002). This observation supports the notion that the evolution of sex determination cascades is bottom-up, and highlights the constant need for tight regulation of sex determination. An eager plurality of sex determination cascades exists across taxa, and the more non-model species are studied the more exciting mechanisms will be revealed.

In sharp contrast to the highly conserved role of downstream sex determination cascade regulatory genes, like the *doublesex-mab3* (DM) family (Haag & Doty 2005, Kopp 2012), the master switch genes are highly diverse across species. There are many instances where the mode of activation for the sex determination cascade depends on external stimuli, namely environmental or social cues (Environmental Sex Determination, ESD). Such examples include temperature-dependent sex determination in reptiles (Merchant 2013), photoperiod in marine amphipods (Guler 2012) and social factors in limpets (Warner 1996). However, the well-studied model species *D. melanogaster, C. elegans* and also mammals, all determine sex through specific genetic elements (Genotypic Sex Determination, GSD) (Bachtrog 2014). In heterogametic GSD systems two important distinctions are made: 1) whether it is the male or female that produces different gametes (male vs. female heterogamety), 2) whether the molecular mechanism of sex determination involves a dominant sex determining gene (dominant Y/W system), or if sex is determined by the number or ratio of X or Z chromosomes to autosomes (Blackmon 2017). It is therefore clear that although not universal, sex and sex-determining genes are often associated with the distinct genomic elements known as ***sex chromosomes***.

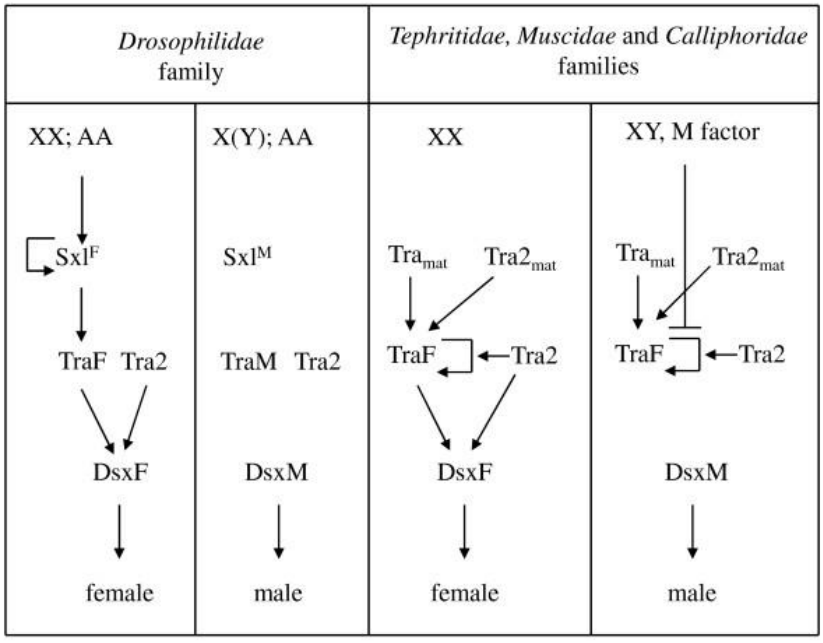

Figure 2 Comparison of the sex determination gene cascades between Drosophilidae (Drosophila), and Tephritidae (Ceratitis Anastrepha and Bactrocera), Muscidae (Musca) and Calliphoridae (Lucilia) families of Diptera. (Sarno 2010)

Evolution of sex chromosomes

Sex chromosomes are commonly found in heterogametic species and have evolved independently numerous times with frequent turnovers from one system to the other (Bachtrog 2014). Two main types of sex chromosome systems are observed (Figure 3). The most familiar is the XX/XY system, where males are the heterogametic species possessing one X copy (hemizygous) and the sex-limited Y chromosome, while females possess two X copies (homozygous). The initial discovery of sex chromosomes by Nettie Stevens in 1905 involved an XY system in mealworms. Similarly, XY systems can be found across heterogametic species, with mammals and the majority of Tephritidae being two examples of interest (Bachtrog 2014). The second sex chromosome system, the ZZ/ZW, is based on the exact converse motif, where the sex-limited W chromosome is present only in females, in contrast to ZZ homozygous males. Examples of the ZW systems are found within vertebrates (in all birds and snakes) and in Lepidoptera, Trichoptera and some Diptera (Tree of sex Consortium 2014, Sahara 2012, Frías 1992). The respective system and state of sex chromosomes should not be taken as eternally stable. Although transitions to female heterogamety from male heterogamety are considered difficult, as this will result in YY offspring, such transitions have been observed in Diptera and specifically within the Tephritidae family, highlighting their dynamic state (Bush 1966).

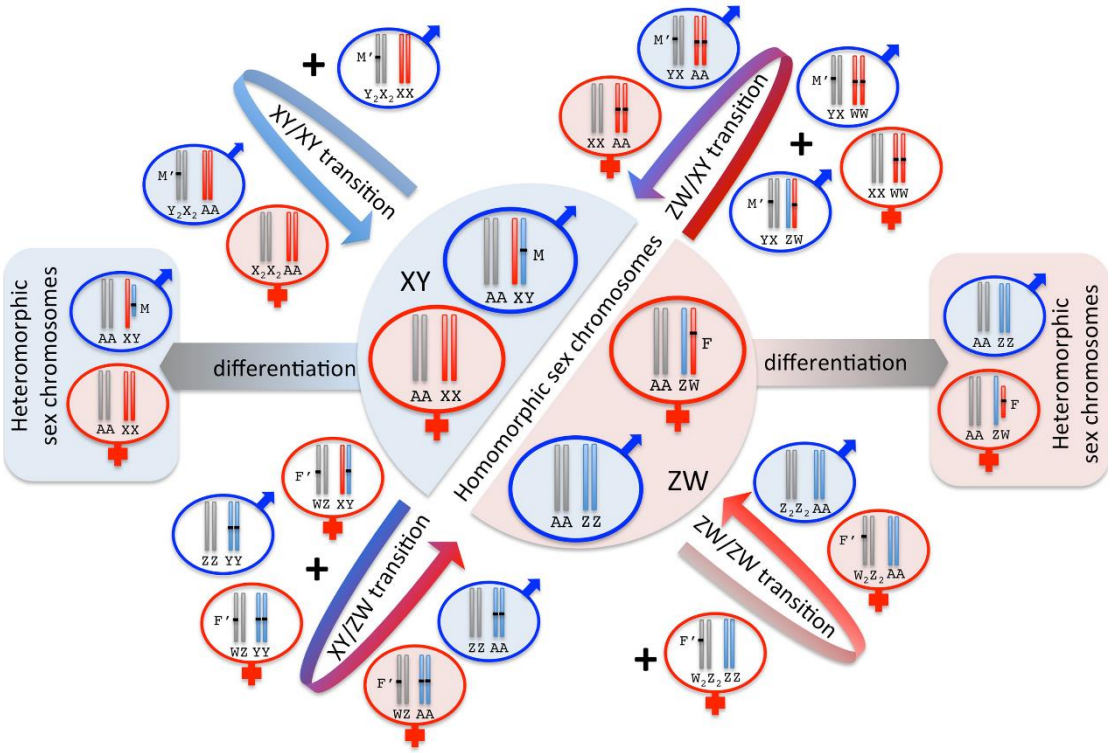

Figure 3 Transitions of sex chromosomes and the respective effect on sex determination switch genes. (Bachtrog 2014)

The classic model for sex chromosome evolution suggests that those distinct genetic elements arise from a pair of autosomal chromosomes upon acquisition of a sex determining role (Figure 4) (Bull 1983). There are multiple ways by which such function can arise, ranging from point mutations that knock-out a gene's function, to gene duplication and neofunctionalization of the duplicate and possibly other ways yet undiscovered (Kamiya 2012, Yoshimoto 2008). Thereafter, the genetic variant that takes control over the sex determination cascade can act as a dominant switch gene or in a dosage dependent manner and might induce the evolution of two distinct sex chromosomes. However, there is increasing evidence that this mechanism is not universal and sex chromosomes exist that do not share a common ancestry between them. Characteristically, supernumerary B chromosomes, small non-essential chromosomes that are often inherited

in a selfish manner, are found to segregate as Y chromosomes in independent lineages (Nokkala 2003, Clark & Kocher 2019).

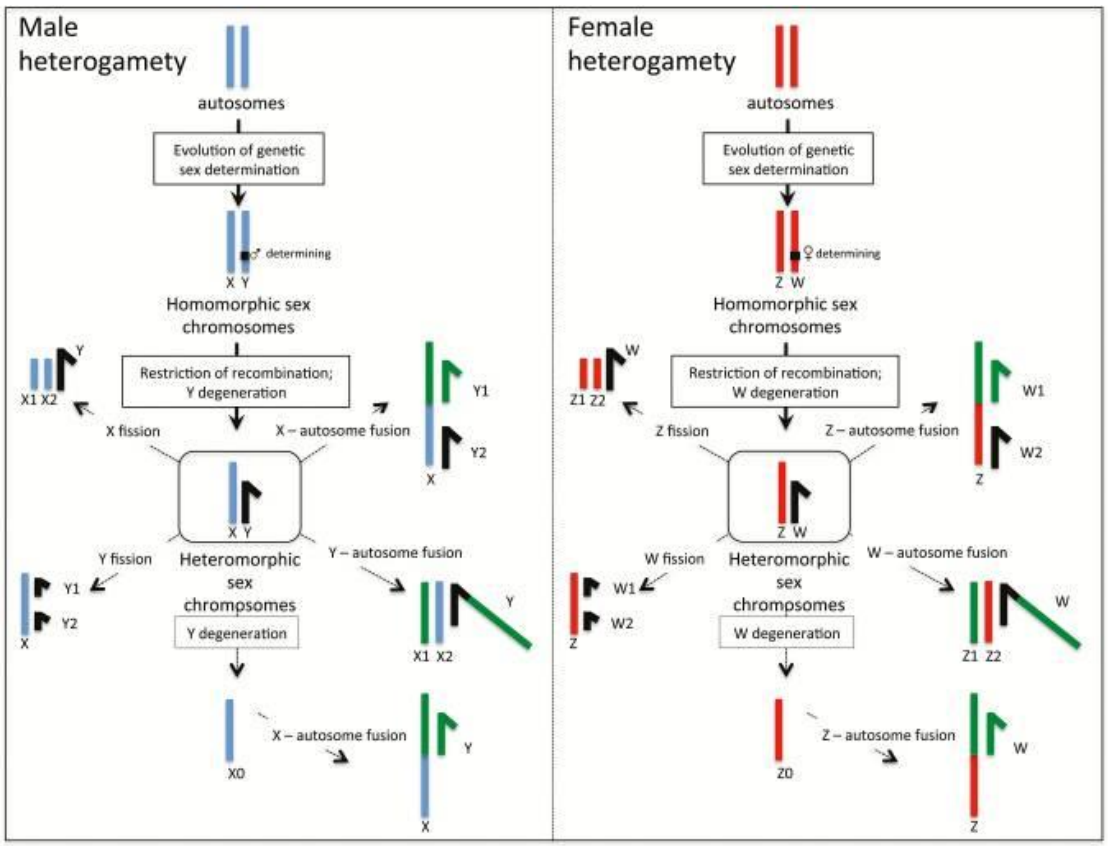

Figure 4 Differentiation of XY and ZW sex chromosomes and alternate pathways towards origination of complex sex chromosomes. (Blackmon 2016)

Upon generation of a sex chromosome pair, two types of morphology can be observed: homomorphic and heteromorphic. Heteromorphic sex chromosomes are the most common type, where the sex-limited Y or W chromosome usually degenerates and obtains a distinct structure and content compared to its X or Z chromosomal pair (Figure 4). However, it's not rare for sex chromosomes to be morphologically similar, in which case are known as homomorphic. Characteristically, mosquitoes of the genus *Aedes* have homomorphic sex chromosomes, in sharp contrast to the heteromorphic *Anopheles* species (Presgraves 1998, Zdobnov 2002). In both cases, the chromosomal region that contains the sex determining function stops recombining between different sex chromosomes. It is thought that ***recombination suppression*** is the driver force for sex chromosome differentiation. Non-recombining sex chromosomes are not the result of a single event (e.g. inversion) but rather evolve through a more progressive mechanism. The linkage of sexually antagonistic alleles (alleles that are preferentially advantageous to one sex but not to the other) close to a sex determining locus creates selective pressure favouring recombination suppression between sex chromosomes, thus ensuring that the sexually antagonistic allele is preferentially transmitted by the sex that it benefits (Rice 1987). Male beneficial genes that are under sexual conflict are favoured over linkage to non-recombining regions (e.g. Y chromosome) since in this way they are no longer present in females and therefore cannot be selected against.

Several mechanisms have been proposed to cause recombination suppression, possibly in a synergistic way. The observed outcome suggests two models for the expansion of non-recombining regions: discrete and progressive, which differ in the presence or not of distinct gene blocks with similar evolutionary divergence form the homologous X copies (Figure 5) (Furman 2020). Chromosomal inversions are a dominant mechanism which can halt recombination in newly formed sex chromosomes. A chromosomal inversion spanning the sex determining region and linked sex-biased alleles is often assumed to promote divergence between sex chromosomes. Once an inversion has been established in the sex determining region, further inversion events can expand the non-recombining regions of the newly formed sex chromosomes. This process results in the generation of distinct strata, regions where genomic characteristics cluster into distinct groups (Figure 5.A), which represent independent inversion events with age-dependent degeneration (Lahn 1999, Wright 2014, Vicoso 2013). Recent findings coming from the planarian species *Schmidtea mediterranea* suggested a dominant role for chromosomal inversions in the recombination suppression of neo-sex chromosomes (Guo 2022). Specifically, the researchers identified three population-specific inversions of >20Mb in size that halt recombination on a chromosome that possesses much of the master regulators related to the reproductive system. Additionally, the same chromosome shows allele-specific expression of sex-related genes, and collectively provides ideal foundation for the evolution as a primer sex chromosome. Furthermore, the accumulation of transposable elements (TE) near the sex determining locus is also assumed to supress recombination through creating sequence divergence between sex chromosomes. Transposable elements comprise a category of selfish genetic elements that are capable of replicating and inserting themselves throughout the genome. TEs are thought to accumulate after sex chromosomes stop to recombine, as a result of reduced positive selection. However, the insertion of TE sequences near the sex determining locus would force the silencing of these genetic elements, which might result in a reduced recombination of the region adjacent to the TE insertion (Kent 2017). Additionally, TE sequences can promote the recombination between ectopic regions, causing genomic rearrangements that would amplify recombination suppression (Bonchev & Willi 2018).

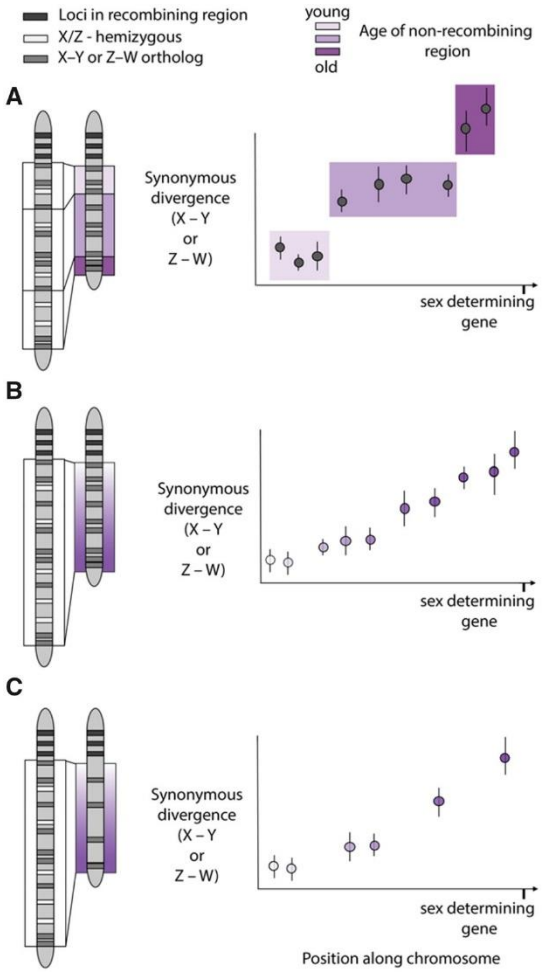

Figure 5 Expected patterns of sex chromosome divergence following recombination suppression. (A) Stepwise progression results in distinct strata. (B) Progressive expansion (e.g., TE accumulation, methylation changes) results in a linear relationship between ortholog divergence across the range of the sex chromosome. (C) A potential problem of inadequate gene sampling is that progressive expansion can be misinterpreted as stepwise. (Furman 2020).

Morphological differentiation of sex chromosomes

In heteromorphic sex chromosomes, the X and Z can exist in a homozygous state (XX or ZZ individuals) and therefore continue to recombine. The persistence of recombination on X, together with the reduction of their effective population size due to hemizygous individuals, has led to the fast X effect hypothesis which implies a faster evolution rate for genes on X, especially for strong male-beneficial recessive alleles (Charlesworth 1987). Therefore, these sex chromosomes tend to maintain most of their former autosomal content. In sharp contrast, it is common for the sex-limited chromosome (Y or W) to morphologically differentiate from its pair through a degeneration process. Two major changes are observed during this process that reshape the sequence, structure and content of the sex-limited chromosome: 1) the heterochromatinization and repeat acquisition, and 2) the loss of its former autosomal gene content. These two are considered interconnected and are both thought to be driven by the evolutionary forces acting post recombination suppression.

Although the outcome of Y masculinization can be easily observed in heteromorphic species, it cannot be monitored in real time. The dynamic interplay between evolutionary processes and genomic rearrangements leads in the formation of the known heteromorphic-Y morphology, being hard though to infer a cause-and-effect relationship for instances acting post recombination suppression like heterochromatinization and repeat acquisition. It is a common instance for repeats on the Y to consist of repetitive satellite DNA and transposable elements (TEs). The random insertion of the latter is known for intervening with functional regions of the genome and can cause disturbance to the Open Reading Frame (ORF) of coding genes or interfere with regulatory elements (Bonaccorsi and Lohe 1991). Since TEs might be detrimental for genome's integrity, several defence mechanisms exist that regulate them from spreading in the genome. Those include RNA-mediated silencing through RNAi or piRNAs that can cause extensive DNA methylation of TEs (Śliwińska 2016, Allshire 2018). Therefore, one assumption is that the accumulation of such sequences on the Y might force the organism to suppress them through initiating an extensive heterochromatinization. These large-scale changes in the chromatin structure might cause the wide silencing of genes residing in that region, known as position-effect variegation (Dorer 1994, Gatti 1992, Schotta 2003, Lippman 2004). Evidence pointing in the direction of TE-induced heterochromatinization arise from Drosophilae Y chromosome, where the age-dependent loss of Y heterochromatin and the expression of TE regions significantly reduce the lifespan of male insects (Brown 2020, Nguyen 2021). On the other hand, the above observations cannot exclude the possibility that TEs might accumulate on the Y post heterochromatinization. Additionally to TEs, large ampliconic regions of genes belonging to different gene families are commonly found to form the repetitive sequences on Y chromosomes, and usually reside in palindromes (Skaletsky 2003). The functionality of such palindromes is thought to be intra-chromosomal recombination and gene conversion between them. This confers the advantage of correcting mutated copies, increasing the number and possibly expression of advantageous genes and counteracting degenerative forces on Y chromosomes (Lange 2009, Marais 2010, Rozen 2003). Collectively, the above result in the characteristic repetitive and heterochromatic nature of Y (or W) heteromorphic sex chromosomes.

Masculinization process also includes the translocation of the former autosomal gene content from the degenerating Y to the autosomes, leading to the shrinkage of the Y (Potrzebowski 2008, *Mořkovský* 2010). This is suggested from empirical observations, where the Y chromosomes are extremely gene poor compared to the respective X pair. Emerging data from the neo-Y of *Drosophila miranda* provide valuable insights into the process of Y masculinization. In this species, the ancestral degenerate Y chromosome of the clade fused with former autosome 3, generating a pair of neo-sex chromosomes. This turnover took place ~1.5 million years ago and the pair of neo-X and neo-Y are currently in the process of differentiating from one another. These neo-sex chromosomes still share a high homology between them (~98% shared sequence), that can be attributed to their former autosomal origin and therefore are still able to recombine throughout most of their

length (Bachtrog 2019). However, the expected former autosomal gene loss from the neo-Y is profound, with 5% of the ancestral autosomal genes being lost from the non-recombining region of the neo-Y. The respective percentage of gene loss from the neo-X is ~0.5%, which complies with the expected rate for autosomal chromosomes (Bachtrog 2019). Although these results are consistent with the theoretical and empirical expectations, a radical observation was that the neo-Y chromosome dramatically increased in size by threefold compared to the neo-X. The size increase is attributed to the massive amplification of genes, which in some cases were found to be co-amplified also on the former X and neo-X chromosomes. This gene amplification was observed to take place in distinct gene clusters that were amplified in a palindrome orientation, as discussed previously. Additionally, TEs start to accumulate on the non-recombining region, which are often integrated into intronic sequences of coding genes leading to the dramatic increase of their size (Koerich 2008).

The degeneration of Y is a convergent process observed in cases of heteromorphic sex chromosomes. However, it should not lead to the misconception that all sex-chromosome pairs are intended to become heteromorphic. There are numerous cases where ancient homomorphic sex chromosomes exist, which recombine and are undifferentiated in most of their length (Bachtrog 2014). Even in the case where heteromorphic sex chromosomes exist, the level of divergence from their ancestral state is not linearly related to their age (Mank 2012). The traditional theory for the evolution of Y chromosomes implied that these genetic elements will be led to extinction based on the assumption of a continuous degeneration process; cases of minor sex chromosome loss exist to support this notion (Steinemann & Steinemann 2005, Graves 2006). Sex chromosome systems that completely lack a Y are termed X0/XX, in which case Y loss can be tolerated if required male-specific Y-linked genes are translocated to autosomes and an alternative sex determination scheme is established (Bachtrog 2013). That is, determination of sex based on the number of X chromosomes or the X/A ratio. Such systems exist in nematodes and are widespread among insects, having been proposed as the ancestral state in many lineages (Nakamura 2009, Blackman 1995). Apparently, *D. melanogaster's* XY/XX system was suggested to have evolved from an X0/XX when a supernumerary B chromosome started segregating as the Y (Carvalho 2010, Carvalho 2002, Hackstein 1996). Furthermore, interesting examples are found in tsetse fly (Glossina, Wiedemann) populations; consistent to what was mentioned before, in this family of insects, sex is not determined by a dominant locus on the Y chromosome, thus irregular X/Y segregation instances are observed with the co-existence of X0, XY or XYY males in natural populations (Maudlin 1979). Despite the exceptions of X0/XX systems, the current opinion on the degeneration of Y chromosomes is that this process does not proceed linearly, but rather degeneration and gene loss slows down and even seizes at a certain point where an equilibrium gene content has been reached. Conversely, evolutionary old gene-poor Y chromosomes, like *D. melanogaster*'s, show a net rate of increased gene gain rather than a continuous gene loss (Koerich 2008). The tight linkage of sex chromosomes with sex determination might generate evolutionary traps that favour the preservation of a degenerated Y with an explicit role in sexual reproduction, instead of transitioning to a different sex chromosome system (Bachtrog 2014).

Genomic composition of sex-limited chromosomes

There is a convergent pattern for genes persisting on degenerate Y chromosomes across different species. Usually, ancient Ys possess few coding genes, compared to their X pair, which are enhanced for male-specific functions and appear to have translocated on the Y from other autosomal locations (Mahajan & Bachtrog 2017). Additionally to the coding content of Y chromosomes, there is a manifold of regulatory roles and cryptic variation attributed to those sex chromosomes (Lemos 2008, Maier 2021). To understand the processes that form the gene content of sex-limited chromosomes we must consider the unique selective forces acting on these genetic elements throughout their evolution /degeneration.

Random mutagenesis is a process naturally occurring in populations and can lead to the generation of deleterious (common) or beneficial mutations (rare). Natural selection is responsible for favouring beneficial mutations and eliminating deleterious ones. Upon recombination suppression, the efficacy of natural selection on the sex-limited chromosome is dramatically reduced and leads to the accumulation of deleterious mutations. In recombining chromosomes, natural selection can operate on each mutation independently. However, non-recombining regions are locked into a complete linkage disequilibrium, with the net effect of the linkage group being selected as a whole, rather than each locus independently (Mank 2012). Therefore, when sex chromosomes completely stop recombining, natural selection can only act at a chromosome-wide level on the sex-limited chromosome. Different evolutionary models have been proposed to account for the selection under non-recombining conditions. These are background selection, selective sweeps and Muller's ratchet. Background selection suggests that strongly deleterious mutations are eliminated from a population in a deterministic fashion, therefore only mutation-free non-recombining chromosomes can survive in the long run (Charlesworth 1996). On the other hand, the selective sweep scenario accounts for the hitchhiking process that follows the fixation of a highly advantageous allele in a non-recombining region. In this case, deleterious mutations that are linked to the beneficial allele can persist in the population given that the fitness advantage of the beneficial gene outweighs their adverse effect (Bachtrog 2004, Rice 1987). Finally, Muller's ratchet suggests that in a population the Ys vary in the number of mutations they carry, and the least-loaded mutational class can be lost in a stochastic way, since recombination cannot restore its composition (Charlesworth 1978). It is thought that the above forces, together with the father-to-son inheritance pattern of Y, shape the genetic content of degenerate Ys. The relative contribution of each one is still debatable, and it is even thought that different forces might have stronger effects depending on the degeneration stage of the Y. For example, on a young Y chromosome background selection might have a stronger effect due to the presence of many functional genes, while in later stages of degeneration selective sweeps are considered to be more important (Bachtrog 2008).

Under the scope of sexual conflict, the fact that heteromorphic Ys are present only in males makes them an ideal location for male-beneficial genes. This notion can easily explain the conserved patterns of genetic content across different Y chromosomes, by assuming that male-beneficial genes become Y-linked to avoid negative selection by females. However, this hypothesis is difficult to test, and newly emerged sex chromosomes are again valuable models into the processes that form the gene content of Y. Emerging data from the massive gene amplification on *D. miranda*' s neo-Y indicate that some of the co-amplified neo-X and -Y genes are dosage sensitive, which might indicate a putative dosage compensatory role of Y-linked copies (Bachtrog 2019). On the other hand, neo-Y amplified genes tend to acquire male-specific expression patterns, with the majority of them being expressed in testis and male accessory glands. Additionally, accelerated protein evolution rates are observed for neo-Y genes that are classified as male beneficial, which might indicate a trend for neo-Y genes to acquire or improve male-functions (Zhou & Bachtrog 2012). Most of the X/Y amplified genes of *D. miranda* originated from the autosome to which the ancestral Y linked to, therefore the gene amplification might act as a mechanism for the evolution of sex-specific functions in young sex chromosomes. Contrary to this, the gene content observed in ancient Y chromosomes is still consistent to a

male-specific role, but appears to have translocated to the Y from other autosomal regions. Such is the case for Y-linked genes of both humans and *D. melanogaster,* where the majority of fertility related genes have translocated to the Y from other autosomes (Skaletsky 2003, Carvalho 2002). However, it should not be misconceived that there is a deterministic way of linking genes on the Y. Rather, the selective forces acting on Y chromosomes favour random events to an extent where they confer an advantage to the non-recombining region of Y or resolve cases of sexual conflict.

It would be naive to assume that the genetic content of the Y has only a protein coding role. The non-coding content of Y chromosomes has been found to be implicated in metabolic regulation (Molina 2017), and male disease in humans, through transcription of long, small and the newly discovered yet promising class of circular RNA (Maier 2021). Similarly, studies of the regulatory dynamics between different Y chromosomes of *D. melanogaster* revealed a manifold of cryptic variation among them (Lemos 2008). Under the same genetic background, different Y chromosomes were able to differentially regulate the expression of autosomal and X located genes, with a preference in male-biased genes. The mechanism of this gene regulatory role was not resolved. However, variation in heterochromatin blocks that contain TEs and repeats was hypothesized as being responsible, through altering the availability of limited transcription factors or chromatin regulators (Liu 2007).

Meiotic gene drives

Another approach for the elements that reside on the Y and their functional role suggests that intragenomic genetic conflict is a major contributor to their evolution. Mendelian inheritance predicts an equal probability for transmission of each chromosomal copy to the progeny of a diploid organism. However, selfish genetic elements exist that increase their transmission to the next generation through manipulating meiosis or gametogenesis (Sandler 1957). Such mechanisms, known as ***meiotic gene drives***, consist of a *distorter* gene and a *responder* locus with variable sensitivity. The *distorter* gene must be able, upon presence of a sensitive *responder* allele, to trigger or halt a molecular cascade that will result in a bias inclusion of its allele into functional gametes. Sex chromosomes can be highly differentiated and are considered relatively easy targets for segregation distorters (Bachtrog 2020). Although such selfish mechanisms have been evolved multiple times in different organisms (Courret 2019, Cocquet 2012), they are usually forced to suppression due to the burden caused by population sex-ratio distortion. The process of X/Y gene co-amplification has been proposed as being such a meiotic gene drive repressor mechanism that arises during the early stages of neo-X and -Y differentiation, during which sex chromosomes are vulnerable (Bachtrog 2020). In fact, a Gene Ontology analysis performed on co-amplified X/Y genes revealed an over-representation of biological processes related to meiosis and chromosome segregation which might indicate an ongoing X-Y interchromosomal conflict over segregation (Bachtrog 2019). Additionally, components of the RNAi machinery were also found amplified, which together with the occurrence of both orientation transcripts from co-amplified genes, implies a putative dsRNA mediated gene regulation.

An interesting example of a meiotic gene drive system is the Segregation Distorter (SD) system, which can be found at low frequency in almost all wild populations of *D. melanogaster.* In a male *SD/SD*$^{+}$ heterozygote, *SD*$^{+}$ bearing spermatids fail to complete histone-to-protamine-transition, a process needed for the hypercondensation of DNA in mature sperm cells. Therefore, only SD-bearing spermatids can mature and elongate properly, thus selfishly favouring the transmission of the SD allele. In contrast, female *SD/SD*$^{+}$ individuals pass the *SD* and *SD*$^{+}$ alleles equally to their progeny. Many years of research for the understanding of the drive mechanism led to the detailed characterization of the *distorter* and *responder* (*Rsd*) locus. However, the complete mechanisms of the distorter's mode of action is yet to be found.

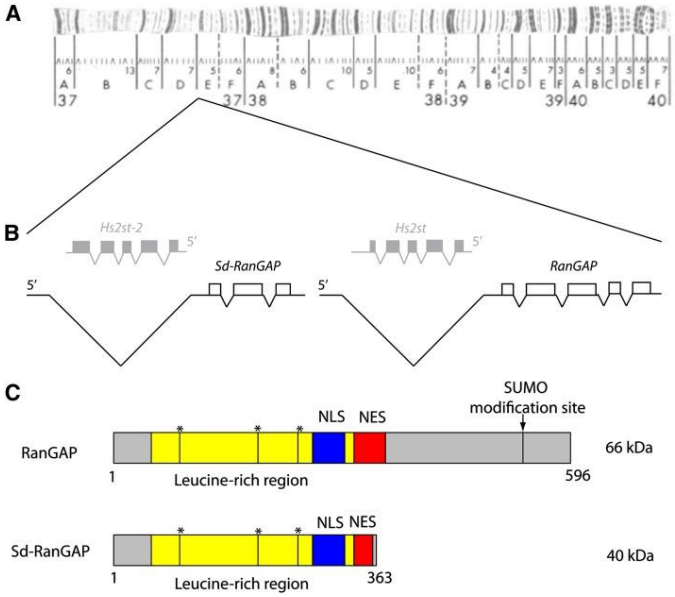

Figure 6 Structure of the SD region. (A) A polytene map showing the location of the Sd locus at band 37D2-6. (B) Gene arrangement on the Sd locus. (C) A schematic representation of RanGAP and Sd–RanGAP proteins. (Larracuente & Presgraves 2012)

The components of the meiotic gene drive are mapped on chromosome 2 of *D. melanogaster.* All SD flies share a ~5kb tandem duplication characterized as the *Sd* locus (Powers & Ganetzky 1991). Two nested genes reside inside the duplication, which encode for a *heparan sulfate 2-O-sulfotransferase* (*Hs2st*) and a *Ran GTPase activating protein* (*RanGAP*) (Figure 6). The distorter role has been attributed to the *RanGAP* gene duplication, and specifically to one of the two duplicates that appears to be structurally altered compared to the WT allele (Merrill 1999). In the distal part of the duplication, the *RanGAP* copy is cleaved at the duplication junction, while an introduced early stop codon leads in the production of a truncated protein that lacks 234 amino acids of the COOH-terminus termed Sd-RanGAP (Figure 6.C). Two important amino acid motifs are absent from the truncated product: a SUMOylation site that is responsible for anchoring the *RanGAP* on

the nuclear envelope, and one of the two Nuclear Export Signals (NES) (Mahajan 1998, Kusano 2003). The *Sd*-RanGAP protein has altered subcellular distribution, being diffused in the nucleus and cytoplasm consistent to exclusion of SUMOylation and NES signals, while its normal localization is associated with the outer periphery of the nuclear envelope (Figure 7). The *Responder* locus of the SD system was more challenging to identify since it corresponds to a class of DNA elements for which little is known, satellite DNA. *Rsp* was mapped near the centromere of chromosome 2, in a highly heterochromatic region (Brittnacher and Ganetzky 1989). It was found that a repeated array of a dimeric 120bp AT-rich repeat correlates with the SD phenotype, and specifically the degree of sensitivity is proportional to the copy number of the repeats: Rsp^{ss} alleles include ~ 2500 copies, Rsp^{s} alleles include ~700 copies, Rsp^{i} alleles associated with SD^{+} chromosomes (no *Sd* duplication) include 100-200 copies, and Rsp^{i} alleles associated with *SD* chromosomes include <20 copies of the repeat (Wu 1988, Pimpinelli 1989).

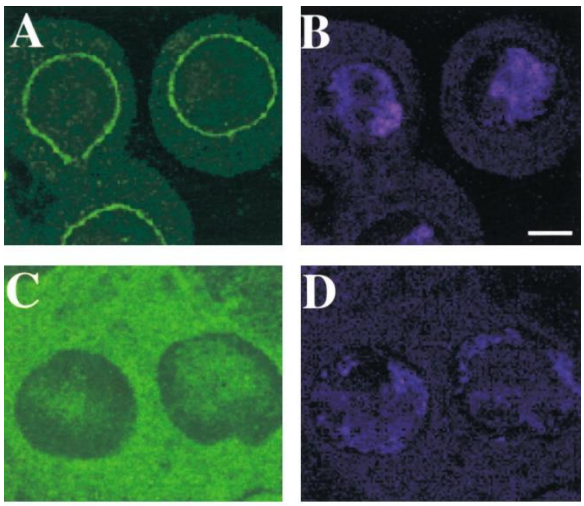

Figure 7 Nuclear mislocalization of Sd-RanGAP in primary spermatocytes. Confocal images show Sd-RanGAP-HA or wild type Sd-RanGAP anti-body staining (green) and propidium iodine staining of chromatin (blue). A & B: primary spermatocytes from wild-type males, C & D: primary spermatocytes from distorting males expressing a Sd-RanGAP-HA transgene (Kusano 2001).

Several modes of action have been proposed for the SD system, however the exact mechanism of segregation distortion is unknown. Ran protein and its cofactor are associated with a number of different cellular processes, including mitotic spindle assembly, nuclear envelope assembly, chromosome segregation and other cellular processes that take place during anaphase and cytokinesis (Quimby & Dasso 2003, Renso & Wilde 2011). In addition to that, Ran protein is mostly known for its essential role in the nuclear transport process. Ran belongs to the Ras protein family and is a small GTPase, meaning that it hydrolyses GTP and exploits the released energy to change its conformation. Ran is found in two forms, GTP- and GDP-bound, and it cycles between them through the help of its co-factors RanGAP and RanGEF (Ran guanin nucleotide exchange factor). These proteins catalyze the nucleotide hydrolysis and nucleotide exchange respectively and their differential localization (RanGAP: nuclear envelope, RanGEF: chromatin associated) generates a concentration gradient of Ran-GTP and Ran-GDP across the nuclear envelope (Figure 8) (Kalab 2002). An essential role for the directionality and trafficking of components between the nucleus and the cytoplasm has been attributed to the differential distribution of Ran-GTP and Ran-GDP (Gorlich & Kutay 1999). Import complexes that carry cargo proteins into the nucleus, enter through the nuclear pore and are dissociated through the help of nuclear Ran-GTP, thus releasing their cargo. Similarly, proteins and RNAs that are intended to exit the nucleus form export complexes that require Ran-GTP for their stability, which is hydrolyzed to Ran-GDP by the co-factor RanGAP upon entering the cytosol leading to complex dissociation (Figure 8). *Sd*-RanGAP is thought to perturb the concentration gradient of GTP- and GDP-bound Ran through its aberrant subcellular distribution.

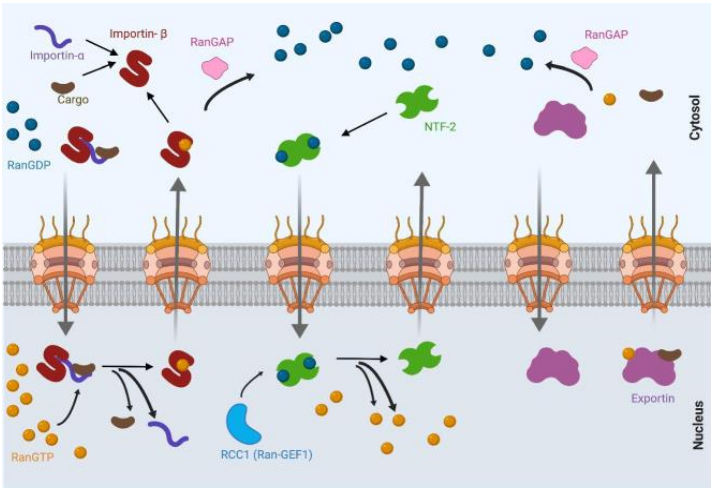

Figure 8 Schematic view of conventional nuclear transport mechanism using RanGTP/RanGDP gradient. (Mirsalehi 2021)

Karyopherins in gametogenesis

The compartmentalization of DNA and the shuttling of proteins and RNA to and from the nucleus is a fundamental aspect of eukaryotic systems. Molecules smaller than ~40kDa are capable of translocating between cytoplasm and nucleus through diffusion. However, larger macromolecules require a carrier to move them through a basket-like structure called nuclear pore. This movement is mediated by karyopherins, a family of proteins that use the Ran-GTP/-GDP gradient described before to transfer molecular cargo (Figure 8) (Kimura 2014). Three distinct types of karyopherins are found in eukaryotes: alpha-importins (IMPα), beta-importins (IMPβ) and exportins. As suggested by their names, the two former types cooperate for the import of cargo in the nucleus while the latter type is responsible for the export process. For the purposes of this thesis, I will focus on importins as it appears to have an intriguing role in gametogenesis. IMPα are characterized by an importin-β binding domain on the N-terminus (IBB) and armadillo (ARM) repeats on the C-terminal domain, while IMPβ consist mostly of HEAT motif repeats on the C-terminus and a Ras binding domain on the N-terminus (Bednenko 2003). The ARM and HEAT protein domains share common ancestry, with IMPβs being the progenitors of IMPas (Malik 1997).

Proteins that are intended for nuclear localization often carry respective signals (NLS: Nuclear Localization Signal) which are recognized by importins. A classical NLS includes stretches of basic amino acids that are recognized by the ARM repeats of IMPα. The NLS-ARM interaction causes a conformational change on IMPα that allows the interaction with IMPβ1 through the IBB domain; the protein complex can now enter the nucleus through the nuclear pore. Upon translocation, a Ran-GTP molecule interacts with the Ran binding domain at the N-terminal of IMPβ and causes the dissociation of the import complex. On the other hand, members of the IMPβ family have the ability to independently mediate the translocation of proteins to the nucleus without the need for prior IMPα interaction while the NLSs recognized by IMPβs contain much greater variation (Lee 2003, Lee 2006). The developmental transitions and differentiation of eukaryotic cells seems to be simultaneously controlled by regulatory elements, like transcription factors, and importins that mediate the translocation of TFs in the nucleus at the proper time (Loveland 2015). Therefore, there is an expected plurality over the number of IMPβ members, the type of NLSs each IMPβ member recognizes, as in its spatiotemporal expression (Major 2011). The members of IMPβs found across species range from 14 in *Saccharomyces cerevisiae* to 20 in humans (Figure 9), highlighting the parallel transport pathways that exist (Chook 2011). In *D. melanogaster*, mutations in IMPβ genes that disrupt the transport pathways are found to be detrimental for specific developmental and morphogenesis processes (Kimura & Imamoto 2014). Of high interest for the purposes of this thesis is the role of IMPβ members in spermiogenesis and testis physiology functions.

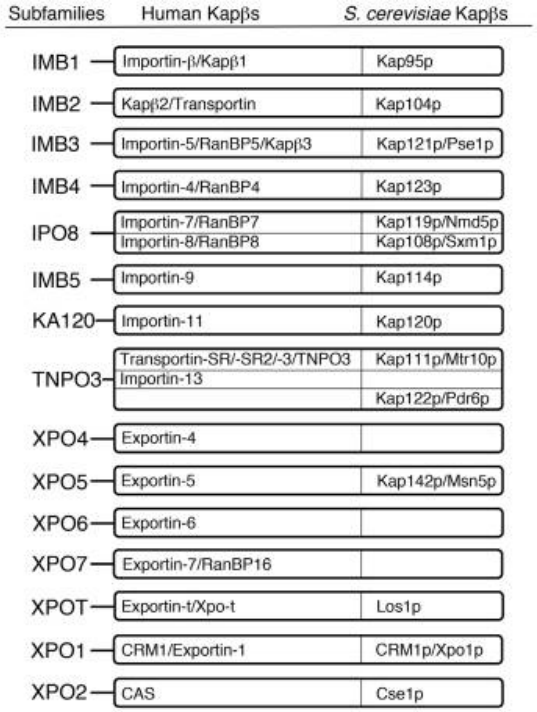

Figure 9 The importin beta family of nuclear transport factors. Schematic of all known human and S. cerevisiae IMPβs divided into evolutionary-derived subfamilies (Chook 2011).

Spermiogenesis is a process that involves drastic rearrangements in the cell, in a way that the unique morphology of sperm can be accomplished. Therefore, shuttling regulation of specific transcripts and proteins to and from the nucleus is vital for the differentiation to be successful. Major (2011) described a pattern of differential cell-type expression and subcellular localization of karyopherins throughout spermiogenesis in rodents. During this process, the expression profiles of two IMPβ members, *importin-4* and *importin-5,* were found to be associated with specific stages of spermiogenesis (Figure 10). A convergent characteristic of mature sperm cells is the tight condensation of chromatin in the nucleus of mature spermatids. To achieve this, canonical histones, the proteins that form nucleosomes, are replaced by protamines in a process called histone-to-protamine transition (Miller 2010). In mammals, this process is initiated by intermediate mediators, transition protein 1 (TP1), TP2 and TP4, that replace ~90% of the somatic and testis specific histones. Consistent with the overexpression of *importin-4* during that stage, studies in rodents revealed the involvement of imp-4 in the translocation of TP2 during mammalian spermiogenesis process (Pradeepa 2008).

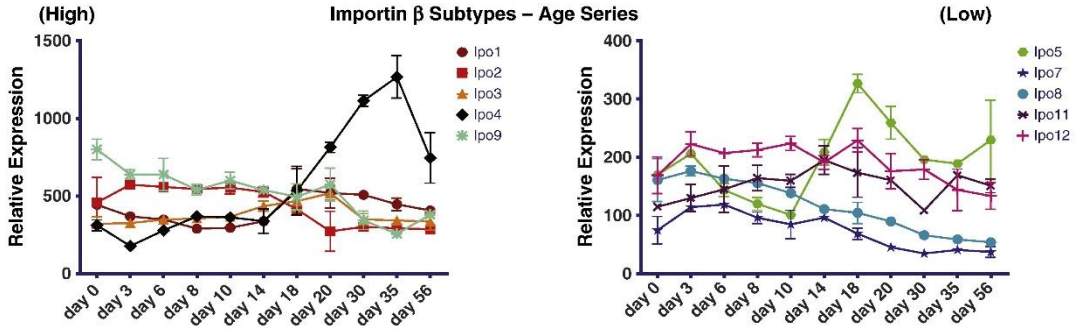

Figure 10 Expression of importin βs throughout mouse spermatogenesis. Data are separated into lower/higher detection levels for visualisation purposes (Major 2011).

The essential role of imp4 in gametogenesis appears to be also conserved in the Diptera clade. A tandem duplication of *imp4* homologue in *D. melanogaster* generated two genes termed *Apollo (Apl)* and *Artemis (Arts),* which evolved to resolve intragenomic sexual conflict (Vankuren 2018). Based on the mutation acquisition of the two copies, *Apl* seems to have derived from *Arts* ~208.000 years ago. Each copy is characterized by a tissue enriched expression pattern, with *Apl* being predominantly expressed in testis while *Arts* has acquired higher expression in ovaries. Respective knockouts of each allele indicate that each copy has acquired a vital sex-specific role, with $Apl^{-/-}$ induced sterility in males and $Arts^{-/-}$ induced sterility in females. The sterility phenotype of *Apl* and *Arts* null mutants appears to result from a disruption in the actin/myosin network responsible for gamete elongation, that leads in round spermatid nuclei and eggs. Interestingly, this no-elongation phenotype is also apparent in dysfunctional spermatids of the *SD* system, in which the nuclear translocation system is perturbed by the aberrant Ran-GTP/GDP gradient (Tokuyasu 1977). Sexual conflict is thought to be the selective force that drove the duplication of the ancestral *imp4* gene, since its ancestral pleiotropic state would not allow for the optimization of its male- or female-specific function (Vankuren 2018). Therefore, the duplication and subsequent rapid evolution of sex-enriched role offered a resolution to the intragenomic conflict. It is worth mentioning that the ancestral *imp4* gene was also found duplicated in three additional drosophila lineages through RNA-mediated duplication, and acquired the similar tissue-enriched expression as *Apl/Arts* pair.

In fact, *imp4,* or the precursor of *Apl* and *Arts,* is not the only member of the nuclear transport pathway that was duplicated in the Drosophila lineage. A broad comparative analysis of 131 nuclear transport genes from 22 *Drosophila* species identified eight classes of duplicated genes, with the highest number of individual duplication events being observed for *Ran, importin-α* and *Ntf-2* (Figure 11) (Mirsalehi 2021). *Importin-α2* was found as having the most independent duplication events, and intriguingly its function is specialized in *Drosophila* lineage for gametogenesis (Mason 2009, Mason 2002, Phadnis 2012). In sharp contrast to the *Apl/Arts* example, the majority of duplicated genes (~78%) were found to be RNA mediated retroposed copies, as the coding exons are conserved between copies while introns are either differentiated or absent. A general pattern was observed for nuclear transport gene copies, where the ancestral copy retained a ubiquitous expression while the retrocopy obtained tissue-enriched expression in testis. Furthermore, in several lineages the retrocopies are under positive selection, indicating a functional role for the duplication of these genes, possibly male-related. The duplication events were found to predominantly exist in the *Drosophila* lineage, although the genomes of several other Diptera were studied. However, it should be noted that no members of the Tephritidae family were included in the analysis.

Gene family	Ntf-2	Ran	importin-α	Nup93	Nxt-1	e(y)2b	Tnpo	Importin-6
D. melanogaster	Ntf-2r(1)	Ran-like(2)	αKap4	Nup93-like	-	e(y)2	CG8219	Apl / Arts
D. simulans	Ntf-2r(1)	Ran-like(2) / Ran-like(3)	αKap4	Nup93-like	-	e(y)2	CG8219	-
D. sechellia	Ntf-2r(1)	Ran-like(2) / Ran-like(3)	αKap4	Nup93-like	-	e(y)2	-	-
D. yakuba	-	-	αKap4	Nup93-like	-	e(y)2	CG8219	-
D. erecta	-	Ran-like(2) / Ran-like(3)	αKap4	Nup93-like	-	e(y)2	CG8219	-
D. takahashi	-	Ran-like(4)	αKap5/ αKap6	Nup93-like	-	e(y)2	-	-
D. biarmipes	Ntf-2r(2)	Ran-like(5)	αKap5/ αKap6	Nup93-like	-	e(y)2	CG8219	-
D. rhopaloa	Ntf-2r(2)	Ran-like(6)/ Ran-like(7)/ Ran-like(8)	αKap4/αKap6/ αKap7	Nup93-like	Nxt1-like(1)	e(y)2	-	-
D. elegans	-	-	αKap6/αKap7	Nup93-like	-	e(y)2	CG8219	-
D. eugracilis	-	-	αKap2D/ αKap6	Nup93-like	-	e(y)2	CG8219	-
D. ficusphila	Ntf-2r(3)	Ran-like(9)/Ran-like(10)	αKap6	Nup93-like	-	e(y)2	-	-
D. kikkawai	-	-	-	Nup93-like/ Nup93-like(2)	-	e(y)2	-	-
D. ananassae	Ntf-2r(2)/ Ntf-2r(4)	Ran-like(11)	αKap2B/ αKap5	Nup93-like	-	e(y)2	-	-
D. bipectinata	Ntf-2r(2)	Ran-like(12)	-	Nup93-like	-	e(y)2	-	-
D. pseudoobscura	Ntf-2r(S)	Ran-like(1)	αKap5/ αKap5B	Nup93-like	-	e(y)2	-	Apl / Arts
D. persimilis	Ntf-2r(S)	Ran-like(1)	αKap5	Nup93-like	-	e(y)2	-	Apl / Arts
D. miranda	Ntf-2r(S)	Ran-like(1)	αKap5	Nup93-like	-	e(y)2	-	-
D. willistoni	-	-	αKap5/αKap2E/ αKap2F/αKap2G	Nup93-like	-	e(y)2	-	Apl / Arts
D. mojavensis	-	Ran-like(13)	-	Nup93-like	-	e(y)2	-	-
D. virilis	-	Ran-like(13)	αKap5/αKap2C	Nup93-like	Nxt1-like(2)	e(y)2	-	-
D. albomicans	-	-	αKap5	Nup93-like/ Nup93-like(3)	Nxt1-like(3)	e(y)2	-	-
D. grimshawi	Ntf-2r(S)	Ran-like(14)	αKap2A/αKap2H	Nup93-like	-	e(y)2	-	-

Figure 11 Summary of nuclear transport genes duplication events across 22 Drosophila species is shown in the phylogeny. Each rectangle represents a duplication event. DNA-mediated duplications are shown in bold in the table and with striped boxes in the phylogeny (Mirsalehi 2021).

The olive fly *B. oleae* as a member of the Tephritidae family

Bactrocera oleae (Rossi), also known as the olive fruit fly, is a member of the Tephritidae family of diptera, belonging to the Daculus subgenus. The Tephritidae, or true fruit flies, is a family of insects known for its worldwide economic importance. Out of an approximate of 4.000 members, more than 250 Tephritidae species are known to attack a wide range of fruits and vegetables, making them one of the most important family of pests (Hoy 2018). Their destructive action results from the developmental cycle of the insects, and specifically from the process called holometabolism. During the development of diptera, an individual undergoes complete metamorphosis meaning that the early developmental stages are morphologically different from the adult insect. Therefore, four characteristic life stages are observed: egg, larva, pupae and adult. In an effort to protect the embryos until pupal formation and to provide direct access to nutrients just after egg emergence, adult female pest species oviposit their eggs inside crops from which the larvae feed upon (Figure 12). As a result, the crop is destroyed and made inappropriate for consumption.

Figure 12 Female adult B. oleae ovipositing inside an olive crop (above). The larvae hatched from the eggs feed upon the olive's flesh (bellow).

The majority of Tephritidae pests are polyphagous, and their larvae can feed upon a wide range of fruits and vegetables. Characteristic examples of such pests are *Ceratitis capitata* (Wiedemann), *Bactrocera dorsalis* (Hendel) and *Bactrocera tryoni* (Froggatt). However, the larvae of the olive fly *B. oleae* are strictly monophagous and can survive through feeding exclusively from crops of *olea* species, mainly *Olea europaea* (Daane & Johnson 2010). Due to the dietary association of *B. oleae* to olive crops, the distribution of the insect is tightly linked to olive cultivating areas. Olive fly is widely distributed in the Mediterranean basin where the vast majority of olive trees are cultivated (~98%), and it was accidentally introduced in California in 1998, spreading to N. America and Mexico by 2004 (Burrack & Zalom 2008, Zalom 2008). An endemic species of this pest was also reported in Pakistan, *B. oleae var asiatica* (Nardi 2006, 2010), while Australia is one of the few areas where olive trees can be cultivated for commercial purposes and the olive fly is not established yet (Figure 13) (Malheiro 2015).

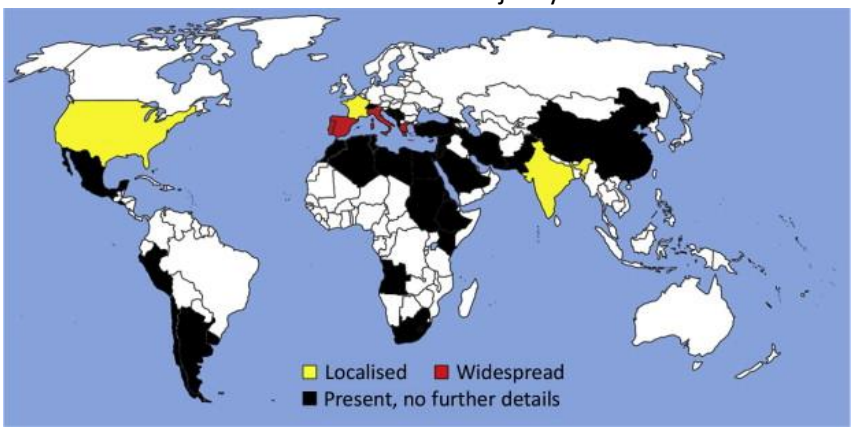

Figure 13 World distribution of Bactrocera oleae. (Malheiro 2015)

The population control measurements taken for the containment of the destructive action caused by *B. oleae* reside mainly on conventional methods. Traditionally, chemical insecticides are used against the olive fly, including organophosphates (Skouras 2007), pyrethroids (Margaritopoulos 2008) and the biopesticide compound Spinosad (Thomas 2005). However, the absence of species-specificity for those compounds make them detrimental to human health and the environment. The potentially serious health effects of insecticides in humans are well reported (Yano 2002, Mansour 2008a, 2008b), resulting in a total of 1 million word-wide deaths and chronic diseases per year (Environews Forum, 1999). Additionally, the extensive use of insecticides underlies the risk of resistance development in wild populations, with such instances being reported in

laboratory conditions (Skouras 2007, Vontas 2011, Kakani 2010). From an ecological point of view, the lack of species-specificity poses a serious threat to biodiversity through affecting advantageous insects, birds, aquatic organisms and animals (Denholm and Rowland 1992).

Eco-friendly alternatives to the use of insecticides for population control purposes have been suggested, with Sterile Insect Technique (SIT) being the most widely approved method (Knipling 1955). SIT is based on the mass rearing and release on the field of sterilized, preferably male-only insects (Figure 14). The sterile males are intended to mate with wild type (WT) females, thus reducing the reproductive potential of the wild population, leading to its suppression. In contrast to chemical control methods, SIT is a species-specific and non-disruptive method that allows for the suppression or even elimination of a local pest population (Dyck 2005). Although SIT is efficiently applied in IPM programs for the containment of other major Tephritidae pest species like *C. capitata* and *B. dorsalis* (De Longo 2000, Orankanok 2006, 2007), its application on *B. oleae* is hindered by the inability to separate male-only insect at early developmental stages. Such separation requires Genetic Sexing Strains (GSS) that are based on the translocation of genetic phenotypic markers on the male-limited Y chromosome. The mutation of those markers on the autosomal loci and the subsequent phenotypic rescue on the Y confer a distinct phenotype on females (eg. white pupae pigmentation, embryonic thermal sensitivity) while the males possessing the WT phenotype are selected over. Efforts for the genetic translocation of phenotypic markers on olive fly's Y chromosome were unsuccessful, as were both-sex release SIT trials (Economopoulos 1972, Economopoulos 1977, Economopoulos & Zervas 1982).

Figure 14 Principles of Sterile Insect Technique (SIT) method (New scientist©)

The principle of male-only releases in SIT highlights the importance of sex chromosomes, and specifically of the sex-limited Y, for the efficient implementation of IPM programs. In many higher Diptera, the chromosomal composition consists of 5 large rod chromosomes, described as A-E Muller elements in *Drosophila melanogaster*, and a smaller heterochromatic dot-like chromosome (Muller 1940). Consistent to this is the typical karyotype for Tephritidae, which is composed of six chromosomal pairs, including a pair of heteromorphic sex chromosomes (XY), with the males being the heterogametic sex (Figure 15) (Garcia & Martinez 2009, Kounatidis 2008, Mavragani & Tsipidou 1992, Zhao 1998). In contrast to *D. melanogaster* where the sex is determined by the A:X chromosome ratio (Cline 1993), the Y chromosome of Tephritidae was long known to house the male determining factor. The *Maleness-On-the-Y* (*MoY*) gene was recently characterized on the Y chromosome of major Tephritidae pest species, including *B. oleae*, as the necessary and sufficient factor for male development (Meccariello 2019). Despite the high conservation of MoY's sequence and localization on the Y, there is a high variability in Tephritidae's Y chromosome size, in both inter- and intra-specific level. In the karyotype of *B. oleae*, the Y appears as a dot-like heterochromatic chromosome of ~4 Mb in size (Figure 15.B). Similarly, *B. dorsalis*, *B. zonata* and *Rhagoletis completa* also possess dot-like Y chromosomes while the Y of *Anastrepha ludens*, *B. tryoni* and *B. curcubitae* are larger than dot but still short in size. An exception to this morphology is the Y chromosome of *C. capitata* which is large and characterized by distinct chromosomal arms (Figure 15.A).

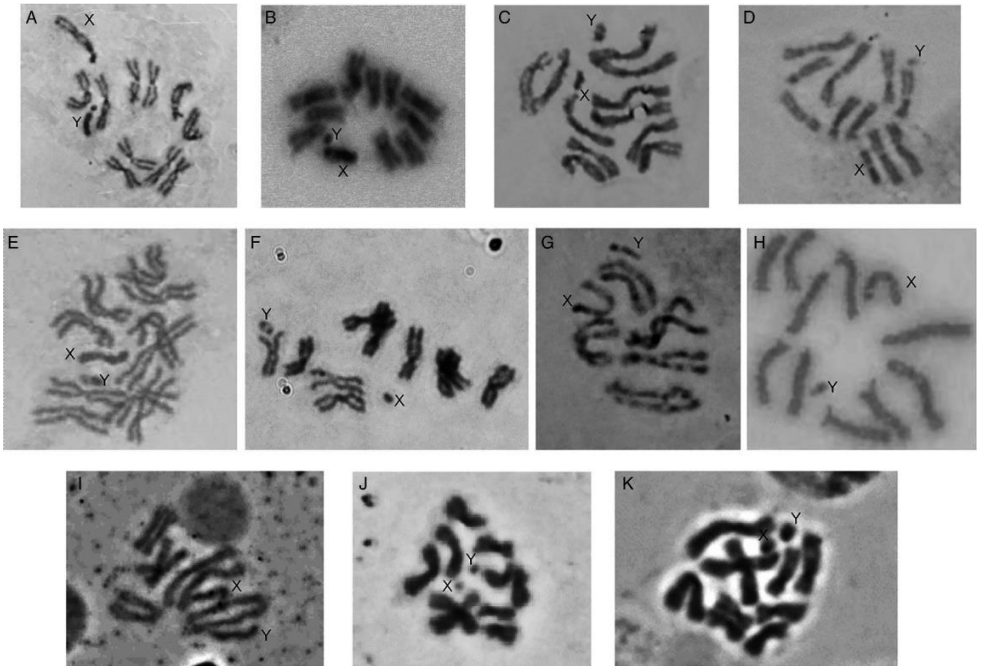

Figure 15 Male mitotic karyotype of Tephritidae species. X and Y chromosomes are indicated. (A) Ceratitis capitata, (B) Bactrocera oleae, (C) B. tryoni, (D) B. dorsalis, (E) Zeugodacus cucurbitae, (F) Dacus ciliatus, (G) Anastrepha fraterculus, (H) A. ludens, (I) Rhagoletis cerasi, (J) R. completa, and (K) R. cingulate. (Zacharopoulou 2017)

Studying the Y chromosome of *B. oleae*

Considering the importance of sex chromosomes and sex-related genes for the efficient application and expansion of IPM approaches, the thorough study of these elements becomes inevitable. Moreover, model species are extremely useful for studying the biology of different lineages in the tree of life. However, studying the sex chromosomes, sex determination and sexual conflict only in a model species could lead to misconceptions about the universality of characteristics. Notably, those mechanisms appear to have independently and convergently evolved multiple times even in organisms of the same clade. Therefore, new data emerging from sex chromosomes of non-model organisms provide interesting exceptions and empirical insight into the evolutionary processes and dynamics of sex chromosomes and sex-related characteristics.

The study of sex chromosomes, and especially of differentiated Ys, was traditionally challenging. Due to their heterochromatic and highly repetitive nature, Y chromosomes are notoriously difficult to sequence and assemble with standard bioinformatic approaches. Usually, these parts of the genome are underrepresented in the final assembly, with Y regions appearing highly fragmented and mis-assembled due to unresolved repeats (Hall 2013, Carvalho 2013). Characteristically, out of the ~40Mb heterochromatic Y chromosome of *D. melanogaster*, only ~4Mb were included in the final reference assembly while the rest required custom curation in order to be resolved (Chang & Larracuente 2021). Additionally, the difficulty in assembling the fuzzy Y has led to female-only sequencing efforts, resulting in Y chromosomes falling behind in genomic studies (Jiang 2022). Even when males are included in the sequencing, the identification of the regions in the assembly that represent sequences of the Y remains a big challenge. Traditionally, the Chromosome Quotient (CQ) approach is used for this purpose (Hall 2013). CQ can identify both X and Y assembly regions in the presence of independent male and female sequencing data of high depth. The sequencing reads are aligned on the reference assembly and different chromosomal regions are characterized by the differences in depth of coverage. Single copy Y is expected to be covered only by male datasets, while the ratio of depth of coverage for X chromosome is expected to be 1/2 based on the presence of single copy X on males. This method was proven efficient in the characterization of human and mosquito sex chromosomes, however the unique evolutionary history and sequence divergence of different Y chromosomes, in addition to variations in sequencing technology and depth utilized across species, sets the global efficacy of the method at a risk.

The use of hybrid sequencing and assembly methods has provided remarkable results in *de novo* assembly efforts of non-model organisms. Consistently, the use of hybrid technologies in the recent whole genome sequencing of *B. oleae* resulted in one of the most contiguous assemblies in Tephritidae and closed many gaps in the composition of the genome (Bayega 2020). Moreover, it provided exceptional insights into the sequence of the dot Y chromosome of this insect, which hosts the maleness factor. The Y chromosome sequences were characterized through the initial CQ method (Hall 2013), which predicted ~3.9Mb in the assembly as being Y-linked, the ~1.7Mb of which were PCR validated. Surprisingly, the size of predicted Y sequences is consistent with the estimated ~4Mb size for *B. oleae*'s dot Y. However, the reference assembly still contains major errors on Y-linked regions, which are apparent from the mis-assemblies observed on the maleness gene locus. The *MoY* gene was recently characterized in many Tephritidae, including *B. oleae*, as the necessary and sufficient factor for male development (Meccariello 2019). Notably, its computational identification on the Y of *B. oleae* was hindered by assembly errors on the coding region of *MoY*. Currently, *MoY* gene is the only functionally characterized gene on the Y, but considering the evolutionary implications and empirical observations on sex-limited chromosomes, it is expected that additional male fitness and fertility roles should be attributed to it.

Thesis goal and objectives

The goal of this thesis is to provide the appropriate methods for the exploitation of publicly available genomic and transcriptomic data from non-model insect species towards a thorough characterization of the sequence, content, and role of Y chromosomes. We use the high-quality assembly of the olive fruit pest *Bactrocera oleae* as a case study. Initiating from the identification of the Y-linked genomic content we extend the analysis to the point of functionally characterizing a Y-linked gene, acknowledging the importance of Y chromosome's content for the monitoring and implementation of IPM approaches. Specifically, the objectives of this thesis are:

i. Develop novel Y identification method and benchmark them against existing ones
ii. Computationally characterize predicted coding genes on Y regions
iii. Validate Y-linkage of selected assembly regions through PCR
iv. Experimentally characterize the function of a Y-linked gene with IPM-related role

Materials and methods

Genomic resources

The analysis was based solely on publicly available data deposited on NCBI SRA dataset. The recent *de novo* assembly of *B. oleae* was used as a reference (GCA_001188975.4) and genomic sequencing data from male (*SRR8788979, SRR8767399*) and female (*SRR8788622*) insects were used for the Y-identification analysis. The transcriptomic dataset used included data from different developmental stages (*SRR8820069-SRR8820076, SRR8819878 - SRR8819887, SRR5559327, SRR5559328, SRR826667, SRR826668, SRR8800824, SRR8800826, SRR8800833*) and adult tissues (*SRR826665, SRR826666, SRR8800821- SRR8800823, SRR8800825, SRR8800827- SRR8800832*) of male and female insects. For benchmarking the novel algorithms, *D. melanogaster* 's chromosome-level reference assembly (*GCF_000001215.4*) was used together with male (*ERR2163721, ERR2163723*) and female (*ERR2163720, ERR2163722*) DNA-sequencing datasets. Both DNA and RNA reads were aligned on the reference genome using GSNAP algorithm. The gene annotations and Apollo genome browser were provided by A/Professor Alexie Papanicolaou.

DNA extraction and Y contig validation

DNA was extracted from adult *B. oleae* flies from Demokritos laboratory strain using CTAB buffer in-house protocol. The insect samples for DNA isolation were collected from the general insect laboratory population and included whole body of adult males and head-thorax of adult females. Abdomen of females was discarded to avoid male DNA contamination from sperm maintained in spermatheca post mating. Male and female genomic DNA was used for PCR validation of putative Y contigs. One primer pair per contig was designed using Geneious Primer3 plug-in and each primer was validated for its uniqueness through BLASTn of its sequence against the reference genome of *B. oleae.* PCR reactions were conducted with 30ng/ul of genomic DNA using KapaTaq polymerase (Sigma-Aldrich), following the suggested PCR conditions. The sequence and respective T_m of each primer pair can be found in Table 10**.**

Insect rearing conditions and injections

The Demokritos laboratory strain of *B. oleae* was used for the experimental procedures. Insects were reared under controlled conditions of maintained 60% humidity, a temperature of 25°C and 16:8 day:night photoperiod. Artificial larval and adult diets were provided to the insects as previously described and paraffin cones were used for egg laying and collection (Tzanakakis 1989). Prior to injection, emerging adults were sexed and kept at separate cages, to maintain virgin populations. At the first day after emergence, individuals were removed from the sexed populations and injected with 96nl of dsRNA (1,5μg/μl). Injections were performed with Nanoject II using glass needles, at the metathoracic segment as suggested by Tsoumani (2020), and injected individuals were kept in sexed populations. Quadruplicates of injected flies were sampled from day 0 to day 6 post injection, and kept at extrazol reagent until RNA extraction.

dsRNA preparation

Synthesis of IMP4 and GFP dsRNA for the injections was done through *in vitro* transcription using T7 RNA polymerase (Thermo Scientific). Specific primers were designed that amplified a 400bp region on the 4^{rth} exon of autosomal *importin-4* gene. Both primers contained a T7 promoter overhang that was incorporated into the final amplicon. DNA *in vitro* transcription template was amplified from female genomic DNA using KapaTaq polymerase, ran on agarose gel, size selected and extracted using NucleoSpin gel and PCR clean-up kit (Macherey-Nagel). Similarly, a 500bp DNA template for *in vitro* transcription of dsGFP was amplified from a plasmid containing the GFP gene using primers with T7 overhangs and isolated

as described previously. In both cases dsRNA was diluted and administered in water for injection (wfi). Primer sequences can be found in Table 10.

RNA extraction cDNA synthesis and RNAi evaluation

Total RNA was extracted from the whole body of injected insects using Extrazol reagent (BLIRT) based on manufacturer's instructions, with an additional DNase-I treatment step to remove DNA contaminants. Synthesis of cDNA was done with MMLV-RT (Invitrogen), using random hexamer primers to initiate first strand reverse transcription. The total quantity of RNA introduced to the RT reaction was 1 μg, while the synthesized cDNA was diluted with wfi to a final volume of 400μl. Subsequently, the levels RNAi on *imp4* transcript were evaluated using qPCR. A primer pair was designed for qPCR that amplified a 104bp region on exon 3 of autosomal *imp4* gene, standard primers for the housekeeping *GAPDH* and *14-3-3zeta* transcripts were used for normalization of expression and reactions were carried out at conditions suggested by Sagri (2017). Primer sequences can be found in Table 10**.**

Importin-4 sequence alignments

Reference sequences for *importin-4* gene were obtained from the official gene set of each organism on NCBI. These included *B. oleae's importin-4* (LOC106622256), *D. melanogaster's Apollo* (Dmel_CG32165) and *Artemis* (Dmel_CG32164) and *D. simulans importin-4* (LOC6738355). Pairwise alignments of genes were performed using MAFT alignment plugin of Geneious. Multiple alignments of *B. oleae's importin-4* Y copies were done with Geneious mapper using autosomal *importin-4* as the reference sequence.

Results

Development and benchmarking of novel methods

The identification of genomic elements on the Y chromosome requires the thorough characterization of assembly regions that reflect Y-linked sequences. Bayega (2020) supported the characterization of the majority (3.9 Mb) of sequences linked to the ~4 Mb dot Y of *B. oleae* through the use of a published CQ method (Hall 2013). In brief, CQ employees the male/female ratio of read coverage to predict Y-linked regions, based on the low or none support of those regions by female reads. In an attempt to expand the knowledge over the regions of Y, and to test the efficiency of Y-identification methods across different organisms and/or assembly approaches, we developed two additional Y-prediction methods ("Thessaly CQ", "Kmer") and evaluated them against methods used by friendly labs ("Israel", "CQ repeats"). The CQ repeats method is an improved version of the traditional CQ used by Bayega (2020), as it masks repetitive elements identified on contigs. The two novel methods used different approaches to identify Y regions, these included: i) a CQ approach and ii) a kmer-based approach.

Traditional CQ approaches require separate male and female high-depth (>50X) sequencing data sets for the valid assessment of male/female read coverage. The implementation of our own CQ approach uses male and female read coverage files as input and calculates a median read-coverage ratio, in sequence windows of predetermined size. It accomplishes this by using an autosomal contig to calibrate the depth of each sequencing dataset, it normalizes the depth of the least covered dataset (usually females) and calculates the male/female read-coverage for each contig while also accounting for repetitive sequences. Each contig is partitioned in smaller sequences (100 bp default window size) and the male/female mean value of read-coverage is estimated for each window. Subsequently, the ratio of male/female read-coverage is calculated for each window, while the contig's ratio is derived from the calculation of the median ratio from all respective windows.

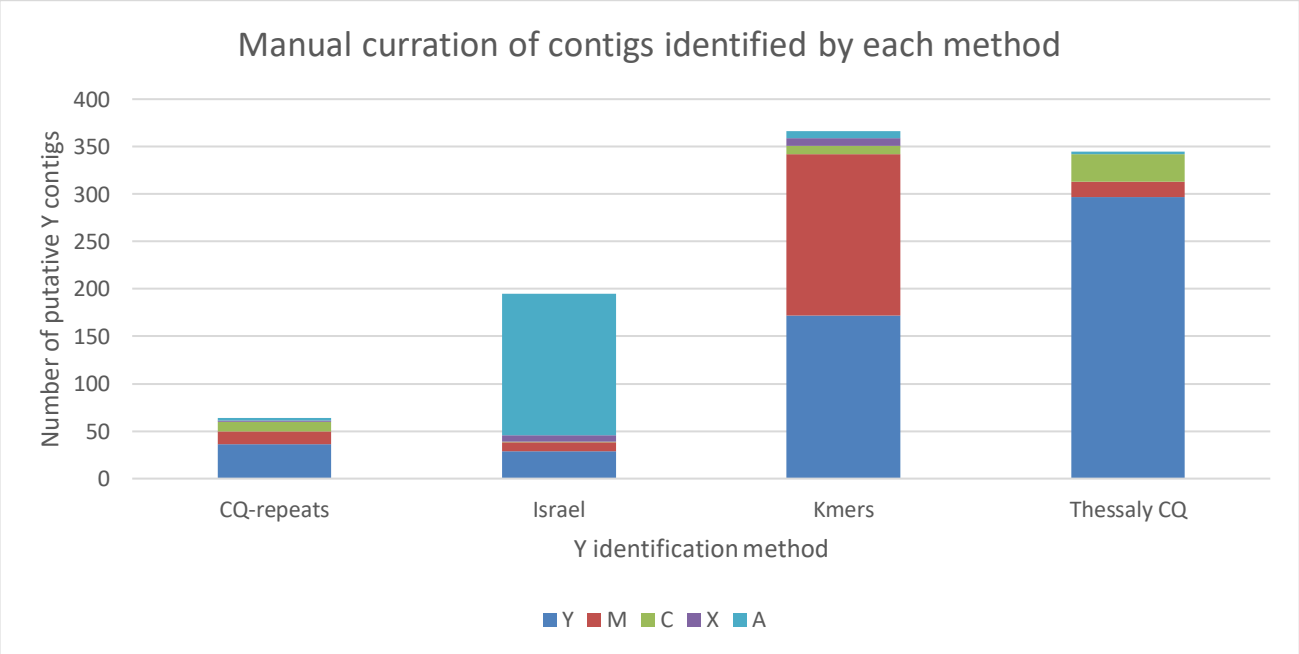

Figure 16 Categorization of putative Y contigs predicted by each of the methods assessed, to one of the five contigs categories: Y, X, autosomal (A), misassembly (M) and contamination (C). Bar plots indicate the number of predicted contigs by each respective method and colour categories indicate the result of manual curation.

On the other hand, Kmer method uses an alignment-free approach to predict Y-linked contigs. The estimation of read coverage through CQ principles requires the prior alignment of whole-genome-sequencing reads on the reference sequence, a process that is time consuming and uses major computational resources. Additionally, the high sequencing depth required for a credible estimation of the male/female ratio increases the cost and limits the applicability of the method. In an effort to account for these drawbacks, a novel method was developed that predicts Y contigs by accounting for male-specific regions across the genome. Specifically, the method still requires separate male and female read datasets and uses them to generate sex-specific kmer databases. Continuously, the percentage of sequence supported by kmers of the male and female-specific database is determined and male-specific contigs can be predicted through a simple division of the male-female (or male-common) percentage of coverage. We expect male-specific regions (Y chromosome) to have higher coverage by the male-specific kmer database, while common genomic regions (X and Autosomes) are expected to possess identical kmer coverage by male- and female-specific databases. This method has the advantage of smaller computational time, since it bypasses the need for read alignments, however it falls behind in the identification of X-linked contigs since the X sequence is present in both male and female datasets.

The two novel methods, as well as those developed by collaborating labs, were employed on the reference assembly of *B. oleae,* which outputted four lists of putative Y contigs. In addition to this, a list of PCR-validated contigs was available from Bayega (2020). The number and IDs of predicted Y contigs differed between the methods used. Specifically, Israel method predicted 195 contigs, CQ-repeats outputted 64 contigs and the novel Thessaly CQ and kmer methods resulted in 350 and 370 putative Y contigs respectively. It should be noted that the predictions of our methods are based on a user defined threshold, above which the contigs are considered putatively Y. For this analysis, the threshold for Thessaly CQ method was set at lnMedian > 2 male/female ratio while for the kmer method it was set at >5% differential coverage by male and female kmer databases (Figure 22, Figure 21). The thresholds used were relaxed and based on empirical observations, while no size filter was applied in the analysis, therefore small contigs of ~2kb were also included in the final lists based on the observation that Y chromosomes appears highly fragmented in the assembly (Tomaszkiewicz 2017). Indeed, Figure 21 suggests an inverse relationship between contig size and number of putative Ys, with the final size category completely lacking any of them. Overall, 13 putative Y predictions overlapped between the four assessed methods (Figure 17, Table 7), resulting in a total of ~1.75 Mb of resolved Y sequence. Additionally, the largest number of shared putative Ys was observed between the two novel methods which highlights their increased sensitivity compared to the others.

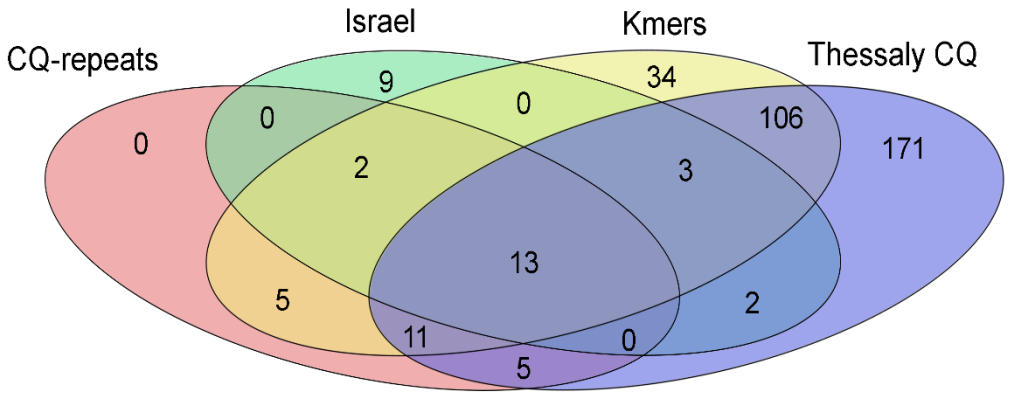

Figure 17 Venn diagram with the number of shared curated Y and Y/M contigs identified among the four assessed methods.

In order to evaluate the efficiency of each method, the total 768 putative Y contigs predicted by all methods were manually assessed for their chromosomal origin. Male and female reads were aligned on the reference genome and the curation of the chromosomal origin was done in Apollo genome browser, through observation of the depth and consistency of read coverage across each contig's sequence. In this way, the contigs were categorized in: X, Y, A (Autosomal) or as being erroneous assemblies (misassemblies) of unknown origin (M) or misassemblies of two chromosomes (e.g. X/Y/M). An additional category was included for contigs that were considered contaminations (C) of non-fly DNA and mainly reflected bacterial sequences of the insect's microbiome (Figure 16). Those contaminants were supposedly removed from the final assembly during quality control (Bayega 2020), however we were still able to identify bacterial genes and sequences in the published assembly, the read-coverage of which did not support a hypothesis of exogenous element integration in the fly's genome. The above manually curated dataset was used to estimate the efficiency of each method in recalling Y contigs as well as the amount of false positives that each one produces (Table 1, Table 2). Since several contigs were identified by more than one method, it was appropriate to define the number of curated Y contigs that are uniquely identified by each method as this reflects its superiority and sensitivity over the others (Table 3, Table 4). Apparently, the newly developed Thessaly CQ method appears to be the most sensitive of all, with a recall percentage of ~93% for curated Y contigs and a false positive rate of ~15% (53/350 predicted Y contigs) (Table 1, Table 2). Kmer method was benchmarked as the second-best method in identifying curated Y contigs, with a recall rate of 43.3 % and a false positive rate of ~54% (198/370 predicted Y contigs). The high false positives derive from the abundance of misassembled contigs predicted as putative Ys. In fact, contig regions that were not supported by adequate male or female short reads were categorized as M, since their origin could not be confidently interpreted. As Kmer method accounts only for horizontal read coverage and not depth, few male reads aligning on misassembled regions led to their characterization as putative Ys by the algorithm. We cannot conclude whether this is an algorithmic drawback or a higher sensitivity, PCR validation of those regions could provide an answer to this.

Although the results from *B. oleae* suggested that the two novel methods overperformed the previously used ones, we wanted to test their efficiency over the reference genome of *D. melanogaster.* We used the model species of Diptera for this analysis since its scaffold characterization is at chromosome level, therefore we could avoid the manual curation for this organism. In fact, the Kmer method had the highest efficiency, achieving an 81.1 % recall rate for validated Y contigs and predicting 52.6 % of the Unknown contigs as Y-linked (Table 5). Similarly, Thessaly CQ had a recall rate of 71 % for validated Y contigs and predicted 65.9 % of the Unknown as being Y-linked (Table 6). It is possible that the different sequencing and assembly methods used in *D. melanogaster* and the unique sequence characteristics of its Y chromosome compared to *B. oleae,* might be responsible for the variance in performance. Overall, the two methods performed efficiently in characterizing >70 % of validated Y contigs and also predicting uncharacterized contigs as Y-linked.

Gene predictions on putatively Y contigs

In an attempt to identify Y-linked genes, the reference assembly of *B. oleae* was automatically annotated using the JAMg pipeline. Continuously, the manually curated Y contigs were inspected for the presence of automatic gene annotations, which resulted in a total of 80 computationally predicted genes. It should be noted that the *MoY* gene was not included in the computational predictions, probably due to its genomic region being erroneously assembled in the reference assembly. Apparently, *MoY* was absent from both NCBI's official gene set and JAMg gene predictions, while the genomic region contained significant read coverage inconsistencies (Figure 19), suggesting that errors in the assembly of that region could impeded the prediction of the Open Reading Frame (ORF) from the maleness factor. The translated ORF of each predicted gene was subjected to a BLASTp search and the results are presented in Table 7. Although the majority of predicted genes had at least partial similarities with gene annotations from other organisms, their expression

was not always supported by *B. oleae* RNA-seq datasets used in the analysis. This could be interpreted either as false predictions of the annotation algorithm or as low expression regions of the genome. Additionally, it is expected for Y regions to regulate the expression of autosomal genes, possibly through production of small RNAs, which could be missed in the RNA seq dataset due to the library preparation strategy used (poly-A enrichment).

One gene was found to be present in multiple copies in the curated Y contig dataset, the karyopherin family member *importin-4.* Specifically, 13 copies of *imp4* were found to be distributed in 5 curated Y contigs (Table 8), while the expected number of actual Y linked copies is expected to be higher. This is inferred from the high DNA-seq read coverage of Y *imp4* regions, that exceed the expected 150X depth of Y-linked regions. Such coverage inconsistencies are expected in highly similar repetitive regions, since the short-read assembly approaches used in the reference assembly fail to resolve long repeats that exceed in length the size of the library. In Y contigs that contained more than one *imp4* copies, those were arranged either in a palindrome orientation (LGAM02015762.1, LGAM02021661.1, LGAM02007390.1) or in tandem (LGAM02019339.1). An assessment of the Y *imp4* expression, based on available RNAseq data revealed an exclusive expression of the Y copies in male tissues, namely testes (SRR826665, SRR8800830) and sex organs of male insects (SRR8800827). In many cases, the expression data were inconsistent to the automated gene annotations, partially supporting the expression of the predicted *imp4* genes. Furthermore, the transcription orientation of Y *imp4* genes was determined through identifying the strand to which the first sequencing reads of the paired RNAseq supportive data align to, and comparing it to that of the ausosomal *imp4* gene. Similar to what is observed for Y amplified genes in *Drosophila miranda,* some *imp4* Y copies were transcribed in the sense orientation while at least one copy was transcribed in the antisense.

A group of 20 contigs from the computationally predicted putative Ys were chosen to further assess their Y-linkage through PCR validation. The prioritization of the putative Y contigs was done based on the following criteria: i) Number of predicted genes, ii) Size, iii) Presence of *imp4* copies, iv) Transcriptome data, v) Identification method. The design of the PCR validation included one primer pair per contig, which was preferentially designed in regions of the predicted genes. Higher genetic variability is expected for Y regions since the absence of recombination allows the fixation of neutral variance. Therefore, all PCR reactions were performed using as template genomic DNA from Demokritos laboratory strain, that was used for the generation of the reference assembly, so to avoid primer mismatches due to genetic variability in Y regions. The results suggested the Y-linkage of 7/20 putative Y contigs that included a total of 10 computationally predicted genes, 3 of which were predicted *imp4* copies (Table 9). The rest of the copies were mapped on contigs that were PCR validated as Y-linked by Bayega (2020). Although these results suggest a ~35% efficiency of our Y curation method, both sex amplification in regions of Y curated contigs could be a false positive result of the one-contig one-primer pair approach.

The genomic regions of Y *imp4* gene copies were isolated, and their sequence was studied in comparison to the autosomal copy. Structurally, the Y copies' annotations indicate the presence of small introns between the CDS, with the exceptions of LGAM02019339.1 (1), LGAM02015762.1 (1) and LGAM02021661.1 (2) that possess long introns and LGAM02021661.1 (3), LGAM02015762.1 (3), LGAM02007390.1 (1) and LGAM02019339.1 (2) that are intronless. Mapping their sequences onto the reference *imp4* reveals that Y copies consist fragments of the autosomal *imp4* gene, which collectively cover large part of autosomal *imp4* coding sequence (CDS). The gene region that is shared among Y copies and the autosomal gene includes parts of the CDS, while there is a total absence of intron conservation. In fact, the small annotated introns of Y copies are erroneous algorithmic predictions, as their sequence corresponds to autosomal's CDS. As for the larger introns, those could not be mapped anywhere on the autosomal gene, therefore might consist secondary sequence acquisition. Collectively, those observations indicate an RNA mediated mechanisms of *imp4* gene amplification on the Y chromosome of *B. oleae* (Figure 18). The sequence

conservation appears to be higher between Y copies, which share common SNPs, small insertions and deletions, rather than among Y copies and the autosomal. This might imply a common origin of Y copies from an ancestral RNA mediated duplication of *imp4*, which could be further amplified on the Y through non-allelic homologous recombination induced by Y-located repeats.

The recent duplication of the *imp4* homologue in *D.* melanogaster gave rise to two copies termed *Apollo* and *Artemis,* which rapidly evolved sex-specific gametogenesis essential role, possibly resolving sexual conflict over the original gene (Vankuren 2018). Considering the fertility role attributed to the *imp4* gene duplications, we compared their sequence to that of the homologous autosomal *imp4* of *B. oleae.* Although the two *D. melanogaster* copies have sex biased roles, they share an impressive amount of sequence similarity at nucleotide level. The comparison of the two genomic loci indicates a 92.72 % identity in their sequence, that corresponds to a 96.2 % similarity in the predicted amino acid sequences. In fact, the most profound difference is an additional intron in *Apl* sequence that removes 20 amino acids from exon-3, near the N-terminal domain (Figure 31). *Arts* was suggested to be the initial copy by which *Apl* emerged, therefore the additional intron might be the result of a secondary acquisition related to the sex-specific function. This intron is also absent from *B. oleae imp4* homologue, although the two organisms share a 46.8 % similarity in the *imp4* amino acid sequence. Additionally, *B. oleae* is characterized by two *imp4* isoforms, which differ in the presence of an intervening exon 4(2) between exons 4 and 5. *Apl* and *Arts* of *D. melanogaster* miss this small exon in their gene annotation, however the closely related species *Drosophila simulans* includes one between exons 4 and 5, in one of the two predicted isoform annotations. The size of the exons is almost identical between the two species (*B. oleae:* 108 bp, *D. simulans:* 111 bp), in contrast to the flanking introns that differ in size, therefore positioning the intervening exons differently in the linear gene alignment (Figure 32). However, an alignment of the exons' sequence reveals a ~48.2 % nucleotide sequence conservation, similar to the total CDS similarity shared by the two species (~48.3 %). Overall, *B. oleae* and *D. simulans* produce two isoforms of autosomal *imp4* gene that differ in the inclusion of an additional exon, whereas *D. melanogaster* produces a single isoform from each of its *imp4* duplicates (*Apl* and *Arts*) that differ in the exclusion or not of an exonic region. Taking into consideration the above, it could be hypothesized that there is a sex specific role in each isoform, which in the case of *D. melanogaster* was transposed from the intervening exon to the structural variation found in *Apl* gene duplicate.

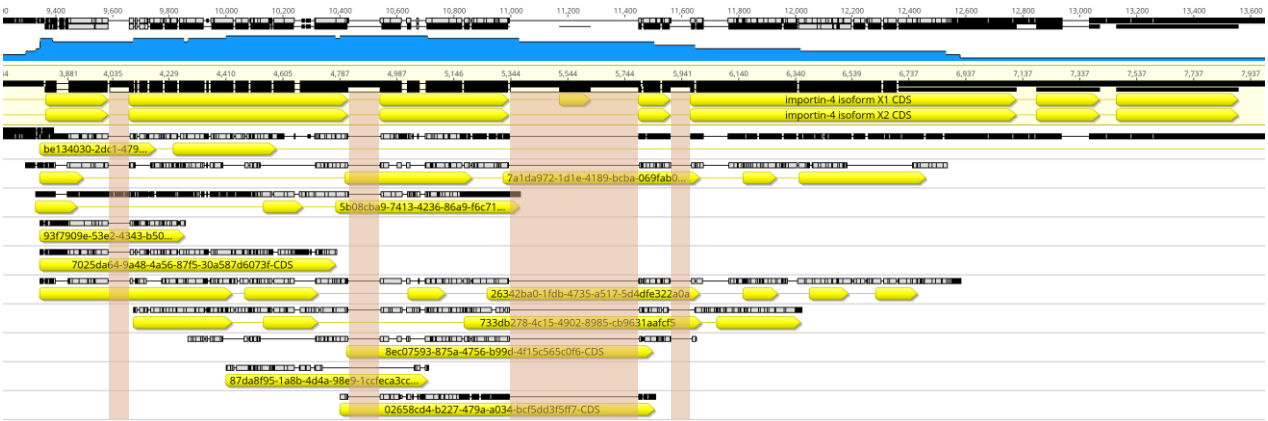

Figure 18 Graphic representation of Y importin-4 gene copies' alignment on autosomal importin-4. Grey regions on genes indicate sequence similarity with autosomal, black regions indicate dissimilar sequence. Orange boxes highlight autosomal intronic regions that lack homologues in Y copies. Aligned copies are (top to bottom): LGAM02021661.1 (4), LGAM02021661.1 (1), LGAM02015762.1 (2), LGAM02021661.1 (3), LGAM02015762.1 (3), LGAM02019339.1 (3), LGAM02007390.1 (2), LGAM02006872.1 (1), LGAM02007390.1 (1) and LGAM02019339.1 (2). Copies LGAM02021661.1 (2), LGAM02019339.1 (1) and LGAM02015762.1 (1) were not mapped to reference due to the presence of large introns, their sequence similarity to imp4 was validated by aligning CDSs only.

Concerning the genomic composition of the Y chromosome of *B. oleae*, the only functionally annotated gene is that of the maleness factor (*MoY*). As noted before, the computational prediction of this gene was impeded in *B. oleae*, possibly due to erroneous assembly of the respective genomic region. A manual inspection of the annotated *MoY* gene revealed major inconsistencies in read coverage (Figure 19). The expected 150X coverage depth was exceeded in the genomic region of *MoY* on contig LGAM02002022.1, with the respective reads reaching in some regions a depth of ~280X. Additionally, the linear coverage depth was inconsistent, with sudden drops of coverage from ~280X to the expected ~150X and back. A BLAT search of the annotated *MoY* region on the whole genome of *B. oleae* revealed the presence of a second contig that was supported mainly by long sequencing reads and few short reads. The comparison of a ~3kb region that included the *MoY* gene in those two contigs indicated a ~91.2 % similarity between them, while a whole contig alignment was not possible due to high divergence in the rest of their sequence. It could be hypothesized that the presence of two *MoY* genomic regions was responsible for the initial error, since the short read assembly that was initially used in the hybrid approach pipeline could not distinguish between highly similar reads deriving from those two independent loci. The presence of multiple *MoY* copies was also validated through qPCR through the ΔΔCq relative quantification method. Two single copy housekeeping genes, *14-3-3 zeta* and *α-tubulin* were used as reference, the amplification of which was based on primers designed and evaluated by Sagri (2016). The results of relative DNA quantification indicated the presence of 3.4 copies of *MoY* gene, which is peculiar, considering that we were expecting a relative quantity of ~1 in the case of duplication (same ΔCq as the single copy housekeeping genes). The higher relative quantity of *MoY* could not be attributed to an increased efficiency of the primers used for its amplification, since this was calculated to be within optimal rates (E: 94.1 %, R^2: 0.997). Possibly, the presence of repetitive *MoY* regions on the Y, that however do not hit as *MoY* in the BLAT search, are responsible for the aberrant calculated copy number. In fact, a manual inspection of the expressed loci on Y contigs revealed the presence of at least 5 assembly regions that contained transcriptomic data and share partial sequence similarity with *MoY* gene, suggesting that fragments of the maleness factor might also be amplified on the Y.

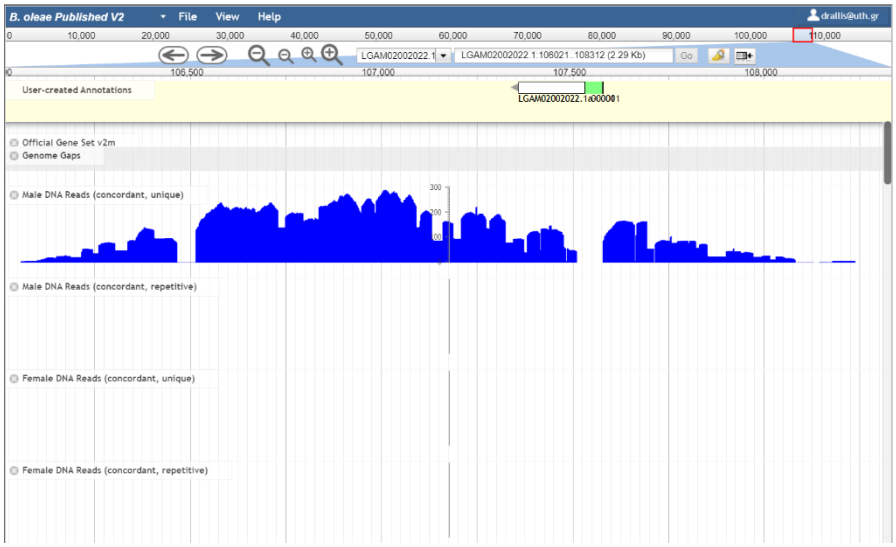

Figure 19 Genomic locus of MoY gene on the reference assembly of B. oleae. Coverage inconsistencies are observed in regions supported by genomic data across the locus, with sudden drops peaks of coverage, indicating the erroneous assembly of the region.

Functional analysis of importin-4 gene

The presence of amplified *imp4* copies on the Y chromosome was intriguing, considering the suggested fertility role of its homologues in other organisms. We proceeded to the functional characterization of autosomal *imp4* gene in *B. oleae*, as a first step in inferring a putative role for the Y copies. Towards that, we

induced RNAi silencing of the *imp4* gene in virgin adult populations by the administration through microinjection of double stranded RNA (dsRNA). The design of the injections was based on similar functional assessments done on the same organism by Tsoumani (2020). Specifically, a 400 bp region of autosomal *imp4* exon-4 was amplified by female genomic DNA, with primers containing T7 promoter overhangs. Male genomic DNA was avoided due to the amplification of Y by-products (Figure 33). Followingly, the amplicon containing the minimal T7 promoter was transcribed *in vitro* from both orientations, generating complementary single strand RNA molecules that formed the desired dsRNA. The double strand state of transcribed RNA molecules was validated through electrophoresis in agarose gel. A final dsRNA concentration of 1.5 μg/μl was obtained, that was 50 % higher than that suggested by Tsoumani (2020), resulting in the administration of 144 ng of dsRNA per injection. Similarly, a dsRNA including a 500 bp region of the non-endogenous GFP gene was amplified from a plasmid and used as treatment control to account for changes in expression caused by the microinjection treatment and RNAi mechanism activation. Since our hypothesis for the fertility role was mainly based on the duplication of *imp4* in *D. melanogaster,* we could not predict the sex towards which the autosomal *imp4* gene is beneficial. Therefore, microinjections were performed in groups of both male and female adult *B. oleae* insects. The microinjections were performed the first day post adult emergence at the metathoracic segment of the insects, prior to sexual maturation. This design was intended as to avoid false positive phenotypic results caused by a pre-injection fertile state.

Prior to the phenotyping process, it was appropriate to optimize the microinjection conditions as to obtain a credible and repetitive RNAi induced silencing. The levels of *imp4* expression were assessed using qPCR reactions, for which a specific pair of primers was designed that amplified a region of *imp4* exon-3. Care was taken as not to design qPCR primers that overlapped with the dsRNA region, so to avoid possible false negative results caused by the reverse transcription and amplification of dsRNA (Figure 34). Relative expression of *imp4* transcripts was determined in quadruplicates at days 0-6 post injection by the ΔΔCq method, using respective dsGFP treated insects as control samples. Two housekeeping genes were used for normalization, *GAPDH* and *14-3-3 zeta*, as suggested by Sagri (2016). The quantification results indicated that by using the above approach, there was no conclusive silencing phenotype. Based on Tsoumani (2020), a reduction in the relative quantity of the transcript is expected at day 3 post injection. For male insects, our results suggest an ~50 % reduction in the transcript abundance of *imp4* at day 3 post injection which is followed by an increase in relative expression at day 4 and finally a second gradual decrease at days 5 and 6 post injection (Figure 20, Figure 23). Among these changes in relative expression, the ~50% decrease at day 3 post injection is the only statistically significant one ($p < 0,05$, paired Student's t-test). In sharp contrast, females do not show a reduction of *imp4* transcript abundance at day 3 post injection, rather the respective expression levels appear to increase without any statistical significance (Figure 20, Figure 24). A slight increase is observed from day 2, similar to that of males, and continues until day 5, where a reduction at control's levels is reached. The final timepoint of day 6 post injection indicates a second increase in abundance of almost ~45 %. It is worth noting that a significant biological variability was observed among control samples (Figure 27, Figure 29), which is reflected in the ΔCq standard deviation values. Although outliers were removed from the analysis, an increase in standard deviation was noted from day 3 and onwards, specifically for female individuals (Figure 23, Figure 24). Interestingly, a differential level of *imp4* expression between day 0 and day 6 post injection was also observed (Figure 25, Figure 26), assuming that dsGFP treated controls reflect the normal variance in expression. By assessing the ΔCq values of *imp4* amplification, the normalized expression levels seem to increase by time, having a final difference of ~0.93 cycles for males and ~1.67 cycles for females, which based on the ΔΔCq method indicate a ~1.9 fold increase in expression for males and ~3.2 fold increase for females. The biological importance of this increase might be reflected in the fertility role of this gene, since *B. oleae* insects reach complete sexual maturity at ~6 post emergence.

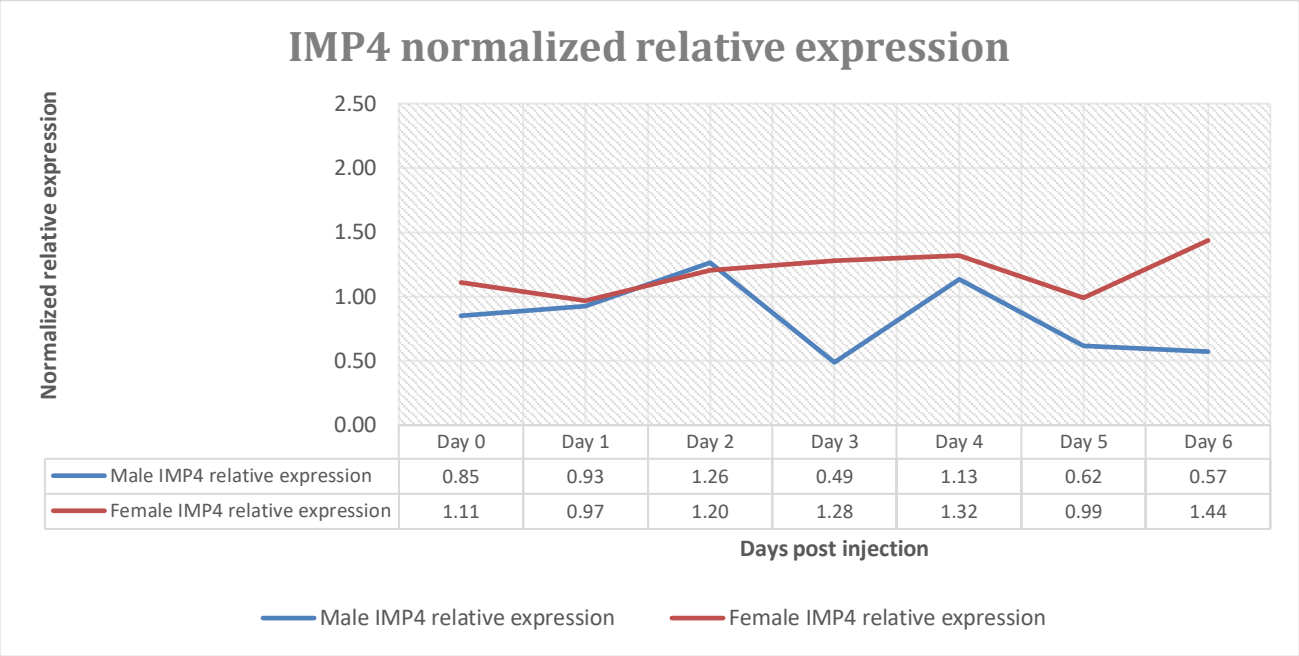

Figure 20 Normalized relative expression of importin-4 gene at males and females from day 0 to 6 post injection.

Discussion

In this study, we focused on the computational identification of Y chromosome regions in sequence assemblies, inspired by the complex process of sex chromosome evolution and the important role these elements have in the development and application of IPM methods. For this purpose, the novel hybrid assembly of the major olive fruit insect pest *Bactrocera oleae* was used as a model, which contained high quality and high depth sequencing data, that included large part of the 4 Mb dot Y chromosome. We developed two novel computational methods, kmer and Thessaly CQ, for the identification of Y-linked regions on the assembly and compared them against other methods used by friendly labs. Our analysis of *B. oleae* Y regions suggested the amplification of one specific gene, the karyopherin *importin-4 (imp4),* on the male specific chromosome. In fact, the amplification of this gene on the Y of *B. oleae* was already known, even prior to the first genome assembly of the insect, as Y-linked *imp4* gene fragments were identified in a Representational Difference Analysis (Gabrieli 2011). However, possible off-targets of the FISH approach performed in the above work, mapped the imp4 fragments also in the pericentromeric region of chromosomes and led scientist to the false conclusion of it being part of the cell's heterochromatin. The high-quality sequencing data that are currently available and our PCR validation (Figure 33), limit the amplified *imp4* copies only in male datasets, revealing their organization on the genome and their amplification origin. At the same time, transcriptomic data support the testis expression of those copies thus refuting the previous claims. Considering the fertility role that has been attributed to homologues of this gene in *D. melanogaster* and rats, we intended the functional characterization of *imp4* in *B. oleae.*

The Kmer and Thessaly CQ method, as well as the other methods used by our collaborators, were benchmarked for their efficiency using a dataset of contigs that were manually curated for their chromosomal origin. Relying on the manual characterization as a metric, our methods had the highest Y recall efficiency, ~93 % for Thessaly CQ and ~43 % for Kmer, ranking as the best two of the methods assessed (Table 2). The percentages do not correspond to actual Y chromosome coverage, since the manually curated contig dataset does not include the total contigs of the assembly but rather those that were predicted as Y-linked by any of the methods assessed. Therefore, although Thessaly CQ appears to be the most sensitive method, its efficiency might drop in case an even more sensitive method was to be introduced in the analysis. It is also worth noting that none of the methods was able to achieve a 100 % recall efficiency and 0 % of false positives, indicating that every available approach suffers from weaknesses. Some Y predicted contigs of high interest were subjected to PCR validation, which suggested that the manual curations does not always reflect real Y sequences, as 7/20 of the tested contigs were validated as Y linked, further reducing the efficiency of the methods. These false positive results question the initial characterization of 3.9 Mb from the dot Y of *B. oleae* (Bayega 2020). However, this assumption does not override the fact that a large part of the Y has been sequenced. In many of the Y contigs assessed, local abnormal increases in the read coverage were observed which suggest the presence of unresolved repeats on the Y regions of the reference assembly. Such inconsistencies could arise from the assembly strategy that was used, which included a first round of short read assembly from 10X linked-reads that were sequenced from high molecular weight DNA libraries, and then proceeded to scaffolding with long read sequencing datasets. In this way, highly similar repeats that appear in tandem or share flanking regions could have been assembled as one region if the size of the repeat exceeded the size of the high molecular weight DNA library. It is expected for the Y chromosome to be highly repetitive, and although standard hybrid assembly strategies can be efficiently employed for the assembly of other genomic regions, resolving the Y might be in need of custom pipelines that rely primarily on long reads. Overall, the above results highlight the fact that computational approaches comprise useful tools for reconstructing the genomic characteristics of organisms, however a blind faith on such methods might be misleading, especially in the study of quirky genomic elements like the Y chromosomes. Rather, the manual assessment of computational outputs might reveal the weaknesses of current respective methods and adjust the results to the peculiarities of each genome, sequencing, or assembly method. Characteristically, the Thessaly CQ method had the second-best recall rate over PCR-validated Y contigs of *D. melanogaster*, although it was ranked as the most efficient in Y-recall on *B. oleae* (Table 5*,* Table 6)*.*

The employment of our novel methods on the Y chromosome of *B. oleae* and the subsequent manual curation of contigs revealed the presence of 80 computationally predicted genes, 69 of which contained at least few or partial coverage by RNA-seq data (Table 7). As discussed previously, since not all of the selected manually curated contigs were identified as Y-linked by PCR-validation, these predictions should also be PCR-tested for Y-linkage and male expression to exclude false positives. It is expected for many of these genes to be false computational predictions, as in many instances no gene similarity was found using BLASTp algorithm while it was rare for gene predictions to be consistent and supported by adequate transcriptomic data. In fact, empirical observations from the Y chromosomes of other fruit-flies, consistent to the evolutionary process of Y-degeneration, set the number of Y-linked genes to only 15-20, ergo the 80 predicted genes far exceed these expectations. Conversely to this, the instance of *MoY* highlights a weakness of gene prediction algorithms in efficiently identifying Y-linked genes, related to local misassemblies observed in the fuzzy Y. Therefore, it would not be of a surprise if more Y-linked genes were missed from the predictions. During manual curation, confidently Y-linked contigs were assessed for the presence of highly transcribed regions that were not predicted as coding genes. Several candidates were found that contained expression data from male tissues or male embryos, which need to be further assessed through qPCR for their male-only expression and their coding or not nature. Additionally, non-predicted genes should exist that were missed due to the fragmented character of the Y on the reference assembly. Studies on *D. melanogaster* indicate a trend for Y-linked genes

to gain large introns due to the acquisition of repetitive elements in them (Koerich 2008). It is therefore expected for the exons of a single Y-linked gene to reside in different contigs of the highly fragmented Y, thus hindering its computational prediction.

The identification and characterization of *MoY* region had been long-awaited by the Tephritidae studying community, due to the prospects that this gene offers in the development of novel population suppression methods. As discussed, the automatic prediction of this gene was hindered possibly by the erroneous assembly of its genomic region that disturbed the ORF on the reference sequence. It was therefore important to understand the reason why assembly algorithms failed in resolving its sequence, as to improve the methods for Y sequence reconstruction and exclude the possibility that other Y-linked genes are missed due to similar errors. The read coverage on *MoY* region suggested that the reference sequence was a misassembly of at least two independent loci, a hypothesis that was validated by the presence of two *MoY* containing contigs and the relative quantification of the copy number of *MoY* through qPCR. This observation could provide valuable insights toward the characterization of the sex determination mechanism is Tephritidae. It could be assumed that each *MoY* copy produces opposite orientation transcripts that lead in the activation of the RNAi pathway, which might cause the downstream regulation of transcripts. Upon RNAi pathway activation, long dsRNA is digested to 20-22nt fragments, the sequence of which determines the target transcripts. Therefore, no linear similarity is required for *MoY* to act against other transcripts and determine sex through post transcriptional regulation of female specific genes. Although exciting, such mode of action is unlike since the injection of *MoY* dsRNA on *B. oleae* embryos causes feminization, suggesting that targeting the *MoY* transcripts emits their functional role. Another hypothesis for the existence of multiple copies could be the dose-dependent action of *MoY* gene. Since transcriptomic data suggest the expression of both *MoY* containing contigs and an ORF is predicted on those transcripts, the existence of a second gene copy might ensure the presence of adequate MoY proteins during early embryogenesis. Considering the highly heterochromatic nature of *B. oleae's* Y chromosome and the fact that no other Y region is found to be expressed during embryonic stages, multiple *MoY* copies might be needed to reach required expression levels from an otherwise transcriptionally idle Y.

Among the predicted genes on Y contigs, one was found to be amplified on the male-specific chromosome. Thirteen partial copies of the karyopherin member *importin-4* were located on PCR-validated Y contigs, the majority of which were partially supported by expression data exclusively from testes datasets. The genomic analysis of those copies indicated an RNA-mediated mechanism of amplification, which was suggested from the total absence of autosomal introns on Y copies (Figure 18). However, this does not imply that all 13 copies were transposed on the Y through RNA intermediates. If this were the case, we would expect for each single *imp4* Y copy to have followed an individual divergence path, being characterized by unique mutations. However, shared polymorphism is found among Y copies, which indicates that the amplification on Y was not the result of individual translocations but rather the expansion of some initial RNA-mediated *imp4* integrations, possibly through non-allelic homologous recombination mediated by intra-Y repeats. Recent results from the study of neo-Y chromosome from *D. miranda* suggest that the amplification of gene families on the Y and even co-amplification between X and Y is a common instance, with hypothesized sex related or dosage compensating function (Bachtrog 2019). Moreover, the Y amplified copies of *D. miranda* are found to produce both sense and anti-sense transcripts that are almost exclusively expressed in testes, supporting the male-specific role. This amplification is observed at the initial stages of neo-Y differentiation and involves genes that resided on the autosome from which the neo-Y derived. Although *imp4* Y copies of *B. oleae* match the expression pattern of amplified *D. miranda's* Y genes, the RNA-mediated integration comes in contrast to the amplification scheme on the neo-Y of this insect, probably suggesting a secondary translocation and fixation of an autosomal gene on the ancient *B. oleae's* Y that should provide an advantage to males. We can make several hypotheses about the origin and mode-of-action of Y copies. *Imp4* copies might infer their function through the production of coding transcripts. The fact that the copies are fragmented could suggest

the production of truncated proteins that act by interfering through their binding domain with targets of the autosomal *imp4,* thus reducing their translocation to the nucleus. However, the production of anti-sense transcripts, together with the sparse and partial coverage by RNA data suggests that this is not the dominant mode-of-action. Rather, the production of both orientation transcripts could serve a regulatory role through non-coding *imp4* dsRNA formation. This could also explain the sparse RNA data, since the method that was used for RNA library preparation targeted the poly-A tails of mature transcripts, which non-coding Y copies could be deprived of. This scenario also serves the fact that the copies were RNA-mediated. The translocation of the CDS only, avoids the carryover of any regulatory elements from the parental gene, in which case the random integration on the Y could confer to the copies the transcription orientation of the insertion region. Based on parsimony, this scenario would easily fix anti-sense producing copies if, in this way, a direct advantage was conferred for males. Finally, the testis only expression might not actually serve a functional role, but it could rather be a side effect of the spermatid maturation. During this process, chromatin undergoes drastic changes with the exchange of somatic histones to protamines, a process known as histone-to-protamine transition. The relaxation of chromatin prior to hypercondensation by protamines might cause leaked expression of *imp4* Y copies. However, such trend of random testis expression on Y regions was not observed during the manual curation of the Y contigs, thus excluding this possibility.

The changes observed on genomic content during the masculinization of heteromorphic Y chromosomes suggest that a male-specific genomic character is acquired on the non-recombining region of those elements. This includes the translocation of non-sex-related genes to the autosomes, the neo-functionalization of genes toward a male-beneficial role and the secondary acquisition of genes characterized by a male-beneficial function (e.g. male fertility genes). The reason behind the fixation of these changes has been attributed to sexual conflict, which is avoided through the linkage of sexually antagonistic genes that confer an advantage to males, to the Y chromosome that has only male-to-male transmission. Therefore, it is expected for the Y-amplification of *imp4* gene to also serve a male beneficial role, since copies are found solely on the Y and not in any other chromosome. Imp4 belongs to karyopherin family of proteins which are part of the nuclear transportation machinery, responsible for shuttling cargo to the nucleus through the nuclear pore. Several members of karyopherins are found to mediate nuclear transportation in different cell types and/or developmental stages, with notions suggesting that nuclear gene activation and expression is spatio-temporarily related to the karyopherin proteins available. Importins of the beta category like *imp4* mainly identify non-canonical NLSs thus further specifying their interactome. This could possibly be an advantage for proteins that are intended to enter the nucleus at specific cell-cycle timepoints, as by sharing a common non-canonical NLS they can ensure synchronous nuclear translocation upon presence of the respective importin beta member on the cytosol. One such process is spermatogenesis, which requires the tight regulation of meiosis stages while drastic rearrangements take place in the morphology of the nucleus through chromatin hypercondensation. In humans, the *imp4* homologue has been found to be implicated in this process, being expressed for a specific time window during spermatogenesis (Figure 10). The critical role of nuclear transportation for the successful outcome of gamete maturation is further highlighted by examples of meiotic gene drive systems, like the SD system in Drosophila, where dysfunctional spermatids are produced upon perturbation of nuclear transportation. Meiotic gene drives tend to arise on genomic regions that lack recombination, like sex chromosomes or centromeric regions, and usually originate from duplication events that cause imbalance in expression (Courret 2019). On the other hand, suppressors of meiotic gene drives have also been reported to originate from gene duplication events suggesting a major role for gene copy number in the battle over fair genomic transmission (Lin 2018, Zhang 2015). Interestingly, *imp4* copies of *B. oleae* are not the only karyopherins that are amplified in the Diptera clade. A comparative analysis on members of the nuclear transport machinery from Diptera revealed that the duplication of karyopherin genes has happened independently multiple times in the Drosophila lineage (Mirsalehi 2021). However, no Tephritidae members were introduced in the comparative analysis. The researchers of this study suggest that

piRNAs, a class of small RNAs expressed in germ cells, are involved in maintaining integrity of the genome through silencing transposons and other repetitive elements during gametogenesis; nuclear translocation genes are essential to mediate the insertion of piRNA-induced silencing complexes (piRISCs). Our results suggest that the karyopherin duplications in Drosophilidae are consistent to what is observed for *B. oleae imp4*, with the events being mainly RNA-mediated and the copies usually acquiring testis-specific expression. Although an exception to the above trend, the tandem duplication of the *imp4* homologue from *D. melanogaster* has recently given rise to two independent genes that rapidly acquired sex-specific roles. Each of the two copies, termed *Apollo* and *Artemis*, serve a fertility role for one of the two sexes, through mediating the correct elongation of the respective gametes. This is assumed to be the result of sexual conflict over the function of the ancestral *imp4* gene, which was resolved by the duplication and acquisition of sex-biased expression and function. The high conservation of *imp4's* role in gametogenesis, from mammals to Diptera, suggests that *B. oleae's imp4* should also be under sexual conflict and the amplification events on Y might be part of a compensatory mechanism. To validate these assumptions, we initiated the functional characterization of the autosomal *imp4* gene with a focus in assessing its fertility role in males and females. This characterization was based on the silencing of the autosomal gene copy through RNAi, which was activated by the administration through microinjection of a dsRNA molecule possessing a 400 bp part of the autosomal *imp4* gene. The silencing validation results indicated that the levels of *imp4* transcript's abundance were temporarily reduced in males with variable efficiency in each individual, while females were highly opposed to the silencing (Figure 20). Additionally, the assessment of *imp4* expression levels on control samples suggested an increase in transcript's abundance as insects were approaching sexual maturation (Figure 29, Figure 27). The relative increase between day 0 and 6 was higher for females, which might indicate a stronger dependency on *imp4's* function compared to males. It is also worth noting that the injections were not targeted in the germline of the insects, neither the expression of autosomal *imp4* is limited in gametic cells. This suggests a pleiotropic role for *imp4* on different tissues, which can explain the higher female expression under the scope of a non-gametogenesis female function. Furthermore, during the injection process females were empirically observed to be larger in size compared to males, thus possibly higher doses of dsRNA might be needed for an efficient silencing. Overall, since the silencing was not succeeded in both sexes and the reduction in males was not persistent, further injection rounds should be performed with higher dosages of *imp4* dsRNA in order to confidently move to the assessment of *imp4's* role in male and female fertility.

Apparently, we cannot conclude to a possible role for Y *imp4* copies without having in context the role of the autosomal gene. However, the collected evidence suggest that a sex-related function should be attributed to the amplification on the Y that is served through the production of non-coding partial sense and anti-sense transcripts. As discussed previously, the coding role for *imp4* copies is excluded by transcriptomic evidence, which suggests that it is most likely for the anti-sense transcripts to be functionally responsible. The activation of the RNAi pathway in testes through *imp4* anti-sense transcription could be a defense mechanism against a meiotic gene drive. If any factor that affects the proper segregation of the Y chromosome in male gametes exists and is dependent on *imp4* for nuclear translocation, then the reduction in the abundance of *imp4* transcripts during specific stages of spermatogenesis could impede the function of this system, acting as a silencer for the gene drive. However, this approach lacks an explanation for the role of sense transcribed copies. A hypothesis that can account for this could be based on the observation that deleterious alleles can hijack the Y chromosome in case they are linked to an advantageous allele, since no recombination takes place to extrude them. *B. oleae's* Y chromosome is known to possess copies of an active transposable element (Tsoumani 2015), the activation of which can have detrimental effect in the integrity of the genome. The integration of an *imp4* copy inside an active Y transposon could interrupt its sequence therefore inactivating it. This advantageous effect could fix the duplication on the Y, even if the production of anti-sense *imp4* transcripts compromises the fitness of the individual. In such case, further expansion of *imp4* on the Y could fix sense transcribed copies that dynamically counterbalance the effect through "sponging" the anti-sense

transcripts. Alternatively, the non-coding *imp4* transcripts could act through gene regulation mechanisms other than RNAi. One such hypothesis is the production of transcripts that induce RNA activation (RNAa) pathway, which causes epigenetic activation of target genes by a yet uncharacterized mechanism (Tan 2021). Interestingly, the autosomal *imp4* gene resides near the *transformer* (*tra*) gene of the sex determination cascade. Sexual conflict could arise over *imp4* if it serves a male-beneficial role but is located in a region that gets deactivated during spermatogenesis due to female-specific function. The conflict could be possibly resolved through the generation of endogenous dsRNA by sense and anti-sense transcripts produced from Y copies. This could ensure the downstream RNAa-mediated maintenance of an active transcriptional state for autosomal *imp4* during spermatid maturation, serving a trans-acting fertility role. The fact that amplification of *imp4* on Y does not suggest a straightforward role, but rather leaves room for different hypotheses, further increases the interest in unraveling the underlying mechanism and proves that Y chromosomes and their evolutionary dynamics are diamonds in the rough.

In order to experimentally check those speculations, an adequate and persistent post-transcriptional gene silencing should be reached (~80 %) in both male and female individuals, which will require an increase in the quantity of administered dsRNA. Furthermore, insect dissections should be performed in injected flies to validate that the systemic RNAi affects the sex specific tissues (testes, ovaries). Thereafter, pairwise crossing male/female injected and WT individuals followed by assessment of their fertility rate would provide insights into the role of *imp4* expression in each sex. A complete knock-out (KO) of the autosomal *imp4* gene using CRISPR/Cas9 system would better assist in the phenotyping process, but poses the risk of a detrimental fitness cost to individuals due to the pleiotropic effect of the gene. Our transcriptomic data indicate a somatic expression for the autosomal *imp4*, which should be counterbalanced by another karyopherin member upon KO, otherwise lethality is expected. An alternative approach would include the insertion of a construct that constantly produces shRNA targeting *imp4,* under the control of a testis specific promoter. However, such promoters are not characterized in *B. oleae,* and although beta-2 tubulin expression is known to be testes-limited, the characterization of its promoter would be laborious. The functional characterization of the autosomal copy will reveal the role and the sex towards which *imp4* expression is more favorable while limiting the hypotheses over the amplification on Y. Continuously, and after a fertility role has been attributed to autosomal *imp4,* assessing the expression levels of that gene in tissues of interest (testes, ovaries) under intersex conditions (Y^- males, Y^+ females) would offer to provide a plausible explanation for Y amplification. Intersexes could be achieved through knock-down of *MoY* (Y^+ females) or *tra* (Y^- males) genes during embryogenesis. In this way, the effect of Y-specific *imp4* transcripts on the expression and function of autosomal *imp4* gene would be revealed, assisting in determining the mode-of-action, role and origin of Y amplified copies.

Appendix

Table 1 Number of contigs falling in each manual curation category (rows), for each of the methods assessed (columns).

Category	*CQ-repeats*	*Israel*	*Kmers*	*Thessaly CQ*	***Category's total***
A	1	88	1	2	**91**
A/M	2	61	3	0	**65**
A/X	0	0	2	0	**2**
A/X/M	0	1	0	0	**1**
A/Y	0	0	1	1	**1**
C	10	1	13	34	**48**
M	14	8	168	15	**194**
X	0	4	5	0	**9**
X/?/M	0	0	1	0	**1**
X/M	1	3	3	0	**7**
Y	**8**	**6**	**97**	**208**	**224**
Y/M	**28**	**23**	**75**	**89**	**124**
Y/X/M	0	0	1	1	**1**
Total	**64**	**195**	**370**	**350**	**768**

Table 2 Percentage of contigs identified in each category (rows) by each different method (column) over the total of contigs identified in that category by all methods (Category's total).

Category	*CQ-repeats*	*Israel*	*Kmers*	*Thessaly CQ*	***Category's total***
A	1.10%	96.70%	1.10%	2.20%	**91**
A/M	3.08%	93.85%	4.62%	0.00%	**65**
A/X	0.00%	0.00%	100.00%	0.00%	**2**
A/X/M	0.00%	100.00%	0.00%	0.00%	**1**
A/Y	0.00%	0.00%	100.00%	100.00%	**1**
C	20.83%	2.08%	27.08%	70.83%	**48**
M	7.22%	4.12%	86.60%	7.73%	**194**
X	0.00%	44.44%	55.56%	0.00%	**9**
X/?/M	0.00%	0.00%	100.00%	0.00%	**1**
X/M	14.29%	42.86%	42.86%	0.00%	**7**
Y	**3.57%**	**2.68%**	**43.30%**	**92.86%**	**224**
Y/M	**22.58%**	**18.55%**	**60.48%**	**71.77%**	**124**
Y/X/M	0.00%	0.00%	100.00%	100.00%	**1**
Total	**1.10%**	**96.70%**	**1.10%**	**2.20%**	**768**

Table 3 Number of contigs falling in each manual curation category (rows) that were uniquely identified by each of the different methods assessed (columns).

Category	*CQ-repeats*	*Israel*	*Kmers*	*Thessaly CQ*	***Category's total***
A	0	86	1	2	**91**
A/M	1	61	2	0	**65**
A/X	0	0	2	0	**2**
A/X/M	0	1	0	0	**1**
A/Y	0	0	0	0	**1**
C	7	0	6	24	**48**
M	7	8	157	10	**194**
X	0	4	5	0	**9**
X/?/M	0	0	1	0	**1**
X/M	1	3	3	0	**7**
Y	**0**	**1**	**14**	**123**	**224**
Y/M	**0**	**8**	**20**	**36**	**124**
Y/X/M	0	0	0	0	**1**
Total	**16**	**172**	**211**	**195**	**768**

Table 4 Percentage of contigs uniquely identified in each category (rows) by each different method (column) over the total of contigs identified in that category by all methods (Category's total).

Category	*CQ-repeats*	*Israel*	*Kmers*	*Thessaly CQ*	***Category's total***
A	0.00%	94.51%	1.10%	2.20%	**91**
A/M	1.54%	93.85%	3.08%	0.00%	**65**
A/X	0.00%	0.00%	100.00%	0.00%	**2**
A/X/M	0.00%	100.00%	0.00%	0.00%	**1**
A/Y	0.00%	0.00%	0.00%	0.00%	**1**
C	14.58%	0.00%	12.50%	50.00%	**48**
M	3.61%	4.12%	80.93%	5.15%	**194**
X	0.00%	44.44%	55.56%	0.00%	**9**
X/?/M	0.00%	0.00%	100.00%	0.00%	**1**
X/M	14.29%	42.86%	42.86%	0.00%	**7**
Y	**0.00%**	**0.45%**	**6.25%**	**54.91%**	**224**
Y/M	**0.00%**	**6.45%**	**16.13%**	**29.03%**	**124**
Y/X/M	0.00%	0.00%	0.00%	0.00%	**1**
Total	**16**	**172**	**211**	**194**	**768**

Table 5 Benchmarking of Kmer method using contig data from D. melanogaster. Descr [sub]/ total [sub] indicates the percentage of D. melanogaster contigs corresponding to each chromosomal category that were predicted as putative Y by Kmer method (Descr [subtracted]), over the total contigs predicted as putative Ys (total [sub]). Descr [sub]/Descr [total] indicates the percentage of D. melanogaster contigs corresponding to each chromosomal category that were predicted as putative Y by Kmer method (Descr [subtracted]), over the total of contigs that are characterized in this category in the dataset (Descr [total]).

	Descr [subtracted]	*Descr [total]*	*Descr [sub]/ total [sub] %*	*Descr [sub]/Descr [total] %*
Y	369	455	41.46 %	81.10 %
XY	14	69	1.57 %	20.29 %
Mitochondrial	0	3	0	0.00 %
X	0	477	0	0.00 %
Autosome	38	270	4.2 %	14 %
Unknown	469	1099	52.6 %	42.67 %
Total	890	2374		

Table 6 Benchmarking of Thessaly-CQ method using contig data from D. melanogaster. Descr [sub]/ total [sub] indicates the percentage of D. melanogaster contigs corresponding to each chromosomal category that were predicted as putative Y by Thessaly-CQ method (Descr [subtracted]), over the total contigs predicted as putative Ys (total [sub]). Descr [sub]/Descr [total] indicates the percentage of D. melanogaster contigs corresponding to each chromosomal category that were predicted as putative Y by Thessaly-CQ method (Descr [subtracted]), over the total of contigs that are characterized in this category in the dataset (Descr [total]).

	Descr (subtracted)	*Descr (total)*	*Descr (sub)/ total (sub) %*	*Descr (sub)/Descr (total) %*
Y	142	200	25 %	71 %
XY	9	67	1.6 %	13 %
Mitochondrial	0	1	0	0.00 %
X	1	447	0	0.00 %
Autosome	12	177	2.1 %	6.7 %
Unknown	369	978	65.9 %	37.7 %
Total	560	1870		

Table 7 Similarity of predicted genes on the manually curated Y contigs. IDs on bold indicate contigs that were predicted as putative Ys by all four assessed methods. The predicted amino acid sequence was used as query on BLASTp algorithm. The table includes the contig over which each gene was predicted, the gene with the highest BLASTp similarity, the respective organism, the e-value of the hit, the size of the gene, the presence of large introns and the existence of expression data.

Contig ID	*Gene similarity*	*Organism*	*e-value*	*Size (bp)*	*Introns*	*Expression*
LGAM02012741.1	Hypothetical	*Bactrocera tryoni*	7e-51	1,515	No	No
	Uncharacterized	*Bactrocera oleae*	1e-111	864	No	No
	Uncharacterized	*Bactrocera oleae*	1e-111	1,267	No	Low
	ATP dependent DNA helicase	*Teleopsis dalmanni*	7e-70	658	No	Low

	Uncharacterized	*Bactrocera dorsalis*	2e-159	3,295	Small	Low
LGAM02017130.1	Uncharacterized	*Nylanderia fulva*	0.0	100,748	Large	Yes
	Uncharacterized	*Rhagoletis zephyria*	0.0	24,896	Large	Yes
	Bacterial RNase	*Candidatus Erwinia dacicola*	5e-122	133,085	Large	Yes
	Uncharacterized	*Rhagoletis pomonella*	3e-126	5,195	Large	Low
	hypothetical	*Bactrocera tryoni*	0.0	2,595	Small	Low
LGAM02003750.1	Hypothetical	*Eumeta japonica*	3e-118	1,586	Small	Low
	Hypothetical	*Bactrocera tryoni*	4e-84	2,164	Small	No
	Uncharacterized	*Rhagoletis pomonella*	3e-165	39,255	Large	Yes
	Hypothetical	*Arctia plantaginis*	1e-144	68,001	Large	Few
LGAM02021661.1	Importin-4	*Bactrocera oleae*	1e-78	1,565	Small	Yes
	Importin-4	*Bactrocera oleae*	6e-93	9,198	Large	Yes
	Importin-4	*Bactrocera dorsalis*	2e-66	417	No	Yes
	Importin-4	*Zeugodacus cucurbitae*	7e-104	5,402	Large	Yes
LGAM02015762.1	Importin-4	*Bactrocera latifrons*	0.0	8,541	Large	Yes
	Importin-4	*Bactrocera oleae*	3e-53	1,453	Small	Yes
	Importin-4	*Bactrocera tryoni*	5e-139	906	No	Yes
LGAM02019339.1	Uncharacterized	*Rhagoletis zephyria*	2e-15	157	No	No
	Importin-4	*Bactrocera oleae*	2e-30	483	No	Few
	Importin-4	*Bactrocera tryoni*	6e-100	2,095	Small	Yes
LGAM02023136.1	Dynein regulatory complex subunit 2	*Bactrocera oleae*	2e-138	884	Small	Few
	N66 matrix protein-like	*Bactrocera oleae*	0.0	1,061	Small	Yes
	Dynein regulatory complex subunit 2	*Bactrocera oleae*	0.0	2,192	Large	Few
LGAM02025169.1	Histone-Lysin N-methyltransferase SETMAR-like	*Ctenocephalides felis*	2e-135	10,891	Large	Few
	ATP-dependent DNA helicase	*Aphis craccivora*	9e-149	57,361	Large	Few
	Uncharacterized	*Bactrocera dorsalis*	2e-116	1,895	Small	Few
LGAM02030549.1	Sentrin-specific protease 7-like	*Bactrocera oleae*	3e-81	1,279	No	Few

	Uncharacterized	*Rhagoletis zephyria*	1e-60	1,695	No	Yes
	Putative gag-pol protein	*Drosophila ananassae*	0.0	16,325	Large	Few
LGAM02034668.1	Uncharacterized	*Zeugodacus cucurbitae*	0.0	45,395	Large	Yes
	Uncharacterized	*Bactrocera oleae*	2e-114	2,609	Small	Few
	Ubiquitin-like specific protease 2	*Bactrocera oleae*	4e-152	2,401	Small	Few
LGAM02037669.1	Alpha-(1,6)- fucosyltransferase	*Eumeta japonica*	3e-82	1,493	No	No
	Zinc-finger BED domain-containing protein 5-like	*Manduca sexta*	0.0	2,459	No	Yes
	No significant similarity	-	-	45,989	Large	Few
LGAM02002186.1	Helicase POLQ-like	*Bactrocera oleae*	0.0	4,024	Large	Yes
	General transcriptional corepressor trfA	*Bactrocera oleae*	0.0	8,987	Large	Yes
LGAM02007390.1	Importin-4	*Bactrocera dorsalis*	3e-49	327	No	Yes
	Importin-4	*Zeugodacus cucurbitae*	9e-74	1,573	Small	No
LGAM02009529.1	Unnamed protein product	*Ceratitis capitata*	8e-38	1,132	No	Yes
	Zinc finger protein chimno	*Bactrocera dorsalis*	2e-34	16,816	Large	Few
LGAM02010183.1	Hypothetical	*Eumeta japonica*	3e-106	39,308	Large	Few
	Uncharacterized	*Lucilia sericata*	0.0	29,937	Large	Yes
LGAM02012394.1	Antigen 5 like allergen Cul n 1-like	*Bactrocera oleae*	3e-126	2,467	Large	Yes
LGAM02016159.1	Putative DD34D transposase	*Bactrocera tryoni*	1e-18	714	No	Few
LGAM02022200.1	Uncharacterized	*Bactrocera oleae*	1e-09	4,805	Large	No
	Uncharacterized	*Bactrocera oleae*	5e-07	9,922	Large	Few
LGAM02026604.1	Unnamed protein	*Ceratitis capitata*	1e-46	974	No	No
	Histone-lysin N methyltransferase SETMAR	*Harpegnathos saltator*	5e-60	1,398	No	Few
LGAM02030084.1	No significant similarity	-	-	895	No	Few
	No significant similarity	-	-	1,948	No	Few
LGAM02036318.1	G protein activated inward rectifier potassium channel 3	*Bactrocera oleae*	1e-56	282	No	Yes
	Uncharacterized	*Bactrocera oleae*	0.0	1,844	No	Few
LGAM02002022.1	Serine tRNA ligase	*Aphis craccivora*	3e-36	5,857	Large	Few
LGAM02006383.1	Bumetadine-sensitive sodium chloride cotransporter	*Bactrocera oleae*	0.0	4,635	Large	Yes
LGAM02006501.1	Sodium dependent dopamine transporter	*Bactrocera oleae*	0.0	1,067	Small	Yes
LGAM02006872.1	Importin-4	*Bactrocera oleae*	2e-47	745	No	No

LGAM02008989.1	Uncharacterized	*Bactrocera oleae*	1e-92	430	No	No
LGAM02014419.1	No significant similarity	-	-	1,583	Small	Yes
LGAM02016277.1	Uncharacterized	*Bactrocera oleae*	0.0	3,045	Small	Yes
LGAM02016868.1	Uncharacterized	*Bactrocera tryoni*	5e-25	37,207	Large	No
LGAM02017966.1	Uncharacterized	*Drosophila sechellia*	0.003	270	Small	Yes
LGAM02020433.1	Gamma-butyrobetain dioxygenase	*Bactrocera oleae*	0.0	1,200	Small	Yes
LGAM02022011.1	Separin	*Bactrocera oleae*	0.0	4,235	Large	Yes
LGAM02022521.1	Glutamate receptor-like	*Bactrocera oleae*	4e-31	524	Small	No
LGAM02023287.1	Protein furry-like	*Bactrocera oleae*	3e-105	5,119	Large	Yes
LGAM02024152.1	TLRP translocation defect protein 14-like	*Bactrocera oleae*	2e-53	3,899	Large	Yes
LGAM02024329.1	Cytoplasmic dynein 2 light intermediate chain 1-like	*Bactrocera oleae*	1e-98	426	No	Yes
LGAM02025742.1	Glutathione S-transferase theta-1-like	*Bactrocera oleae*	5e-140	1,254	Small	Yes
LGAM02026204.1	Mycin-17	*Bactrocera oleae*	2e-89	2,436	Large	Yes
LGAM02027773.1	No significant hits	-	-	1,003	Small	Few
LGAM02029489.1	Uncharacterized	*Acyrthosiphon pisum*	0.0	45,370	Large	Yes
LGAM02030391.1	Carbonic anhydrase 6-like	*Bactrocera oleae*	4e-170	1,921	Small	Few
LGAM02032337.1	Transient receptor potential cation channel trpm	*Drosophila simulans*	2e-37	8,004	Large	Yes
LGAM02034698.1	Trypsin beta	*Bactrocera oleae*	0.0	1,865	Small	Yes
LGAM02036569.1	Uncharacterized	*Bactrocera oleae*	3e-157	5,077	Small	Few

Table 8 Number and IDs of PCR validated Y-linked contigs containing importin-4 copies. Presence and origin of RNA data are also included.

ID	Size (Kb)	PCR verified	Number of genes	Number of predicted importins	Transcriptomic data	RNA-seq ID	RNA-seq tissue
LGAM02015762.1	76.72	Yes	3	3	Yes	SRR826665	Testes
						SRR8800827	Sex organs of adult male insects
						SRR8800830	Testes
LGAM02021661.1	78.27	Yes	4	4	Yes	SRR826665	Testes
						SRR8800827	Sex organs of adult male insects
						SRR8800830	Testes
LGAM02019339.1	71.16	Yes	3	3	Yes	SRR826665	Testes
						SRR8800827	Sex organs of adult male insects
						SRR8800830	Testes
LGAM02007390.1	14.49	Yes	2	2	Yes	SRR826665	Testes
						SRR8800827	Sex organs of adult male insects
						SRR8800830	Testes
LGAM02006872.1	2.12	Yes	1	1	No		

Table 9 Group of contigs that were validated for Y-linkage through PCR.

ID	Contig size (Kb)	Gene similarity (BLASTp)	Transcriptomic data	Validation primers on gene region
LGAM02002186.1	16	Helicase POLQ-like	Yes	Yes
		Transcriptional corepressor trA	Yes	No
LGAM02006872.1	2	Importin-4	No	Yes
LGAM02007390.1	13.7	Importin-4	Yes	No
		Importin-4	No	Yes
LGAM02026204.1	5	Mucin	Yes	Yes
LGAM02029489.1	56	ATP dependent DNA helicase	No	Yes (in intron)
LGAM02030549.1	90	Protease	Yes	No
		No-hits	Yes	No
		Viral	Yes	No
LGAM02033308.1	10.6	-	-	-

Table 10 Primer names and sequences

	Primer name (F)	Forward primer sequence	Primer name (R)	Reverse primer sequence
Putative Y validation	Bo_2186-F	TCGGTGTGCTTTTTCTTTGCC	Bo_2186-R	GCTTCAGCTTCCTCTTCCAG
	Bo_2951-F	AAAACCACGCTTGAGCACAC	Bo_2951-R	TAGCGCGCTTGTTGTTGTTG
	Bo_6872-F	AAGAGCCGAACAAACGTTGC	Bo_6872-R	TTGTCCCAGCGCATAGAAGG
	Bo_7390-F	ACGTCATGGTATCAGTGGCC	Bo_7390-R	AATGAATTTCGAAGCGGCGG
	Bo_9529-F	ACAAACGGAGCAAAGCATCC	Bo_9529-R	CCGCCAATACGTCCGATTTG
	Bo_1797-F	TGAGTTTGTTGTTGCGGCAG	Bo_1797-R	GGTGCAGTAGCGTTGTGTTG
	Bo_4419-F	AAAAGTTTGCCCGAACTCCAC	Bo_4419-R	CATACCCACAAAGTGCCGTG
	Bo_6277-F	GCTTACGGGTGTCATTGCAC	Bo_6277-R	ACCATGAAGCACGAAAAGGC
	Bo_2011-F	GAATAACGGGTTTGTGGGCG	Bo_2011-R	GTATTGGCTACCGCATTGGC
	Bo_3136-F	ATTTCGACGACAACAGGCAC	Bo_3136-R	TCAACACCGACGACTATCCG
	Bo_3287-F	AGCAACACACTACTGTCGGG	Bo_3287-R	TGCTCTTCAGGTGAGTTCCAC
	Bo_4152-F	TCACGCAGACTCGATTGTCC	Bo_4152-R	TCGCATGTTAAGTACTTTCTGGATG
	Bo_5742-F	CCTGCATTATCGACGATGGC	Bo_5742-R	TGCAAAAATATCTGCCGCCG
	Bo_6204-F	TACTGGGGCATCATGGCAAG	Bo_6204-R	ACAAGCGCATATGCCTCAATC
	Bo_7299-F	AGGCGCTTTGAAAATTTTGGTG	Bo_7299-R	TTCTTTGTATTGTATTGCTGTGTAATAC
	Bo_9489-F	TCGAGAGCAAGGTAATTGTTGG	Bo_9489-R	CAATCCCAAGACCTACCTCACC
	Bo_0391-F	CACTTAACTATGTGCCAGTATGTCAG	Bo_0391-R	TCGGGGTAAAACTTTGCGTG
	Bo_0549-F	TATGGTCGCCCTCAACTAGG	Bo_0549-R	TTTCTTTGCGTTTGGGAGCG
	Bo_2337-F	TCAGGTCGATCTCACAGCTG	Bo_2337-R	TTAGTTTTGTTGGCGGTGGG
	Bo_3308-F	CACCGATGATTGTGGCTTCG	Bo_3308-R	AACCCAACGGCTTAAAAGGG
Imp4 silencing	Bo_imp4_ex 4_T7-F	TAATACGACTCACTATAGGGTTTTTGGGT GAAGACAACTCG	Bo_imp4_ex 4_T7-R	TAATACGACTCACTATAGGGGCAATTCA TTGCAGAGCTGA
	T7-GFP-F	TAATACGACTCACTATAGGGCCGCCAGT GTGCTGGAA	T7-GFP-R	TAATACGACTCACTATAGGGGATATCTG CAGAATTCGCC
	Aut_Imp4-F	ACAAACGTTGCAGAAATCCATACC	Aut_Imp4-R	TCTCGAGAATATCAAAAGCGTGAA
	MoY-OLFLY_F	AAATGAGATCAGTGTGGATAAT	MoY-OLFLY_R1	TGTGGTTTCGTTTCCTTTG

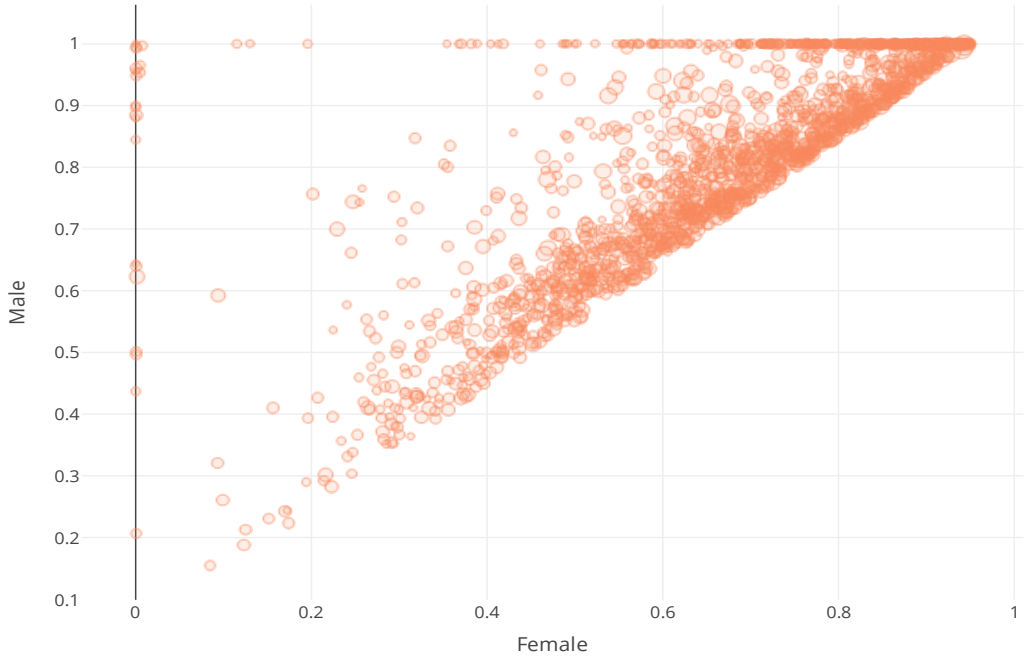

Figure 22 Scatter plot representing the percentage of coverage by male and female kmer databases for each B. oleae contig. Contigs with less than 5% differential coverage (diagonal) were excluded from plotting as were considered autosomal or X-linked. The size of each dot is determined by the common logarithm of each contig's size.

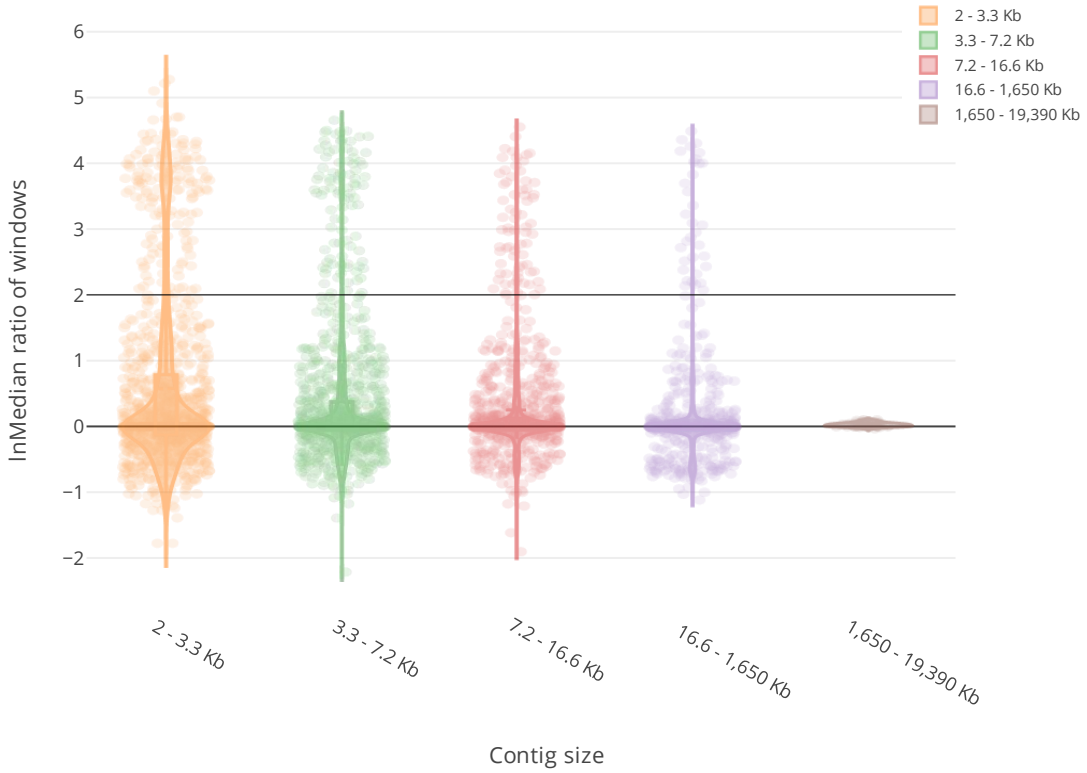

Figure 21 Violin plots of the different size categories of B. oleae's contigs over the lnMedian value of the male to female ratio for each contig. All contigs with lnMedian > 2 were considered putative Y and were subjected to manual curation.

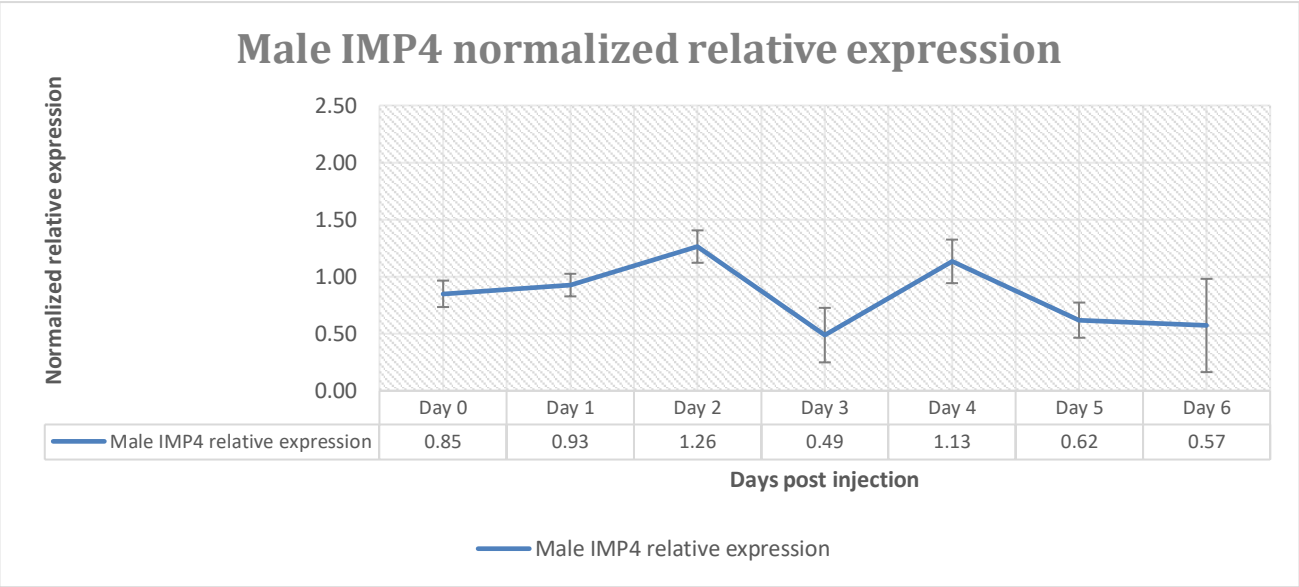

Figure 23 Normalized relative expression of importin-4 gene on males from day 0 to 6 post injection. Error is calculated as $\sqrt{((s1)^2/n1)} + \sqrt{((s2)^2/n2)}$*, where S1 and S2 are the standard deviations of dsGFP treated and dsIMP4 treated respectively while n1 and n2 are the group sizes.*

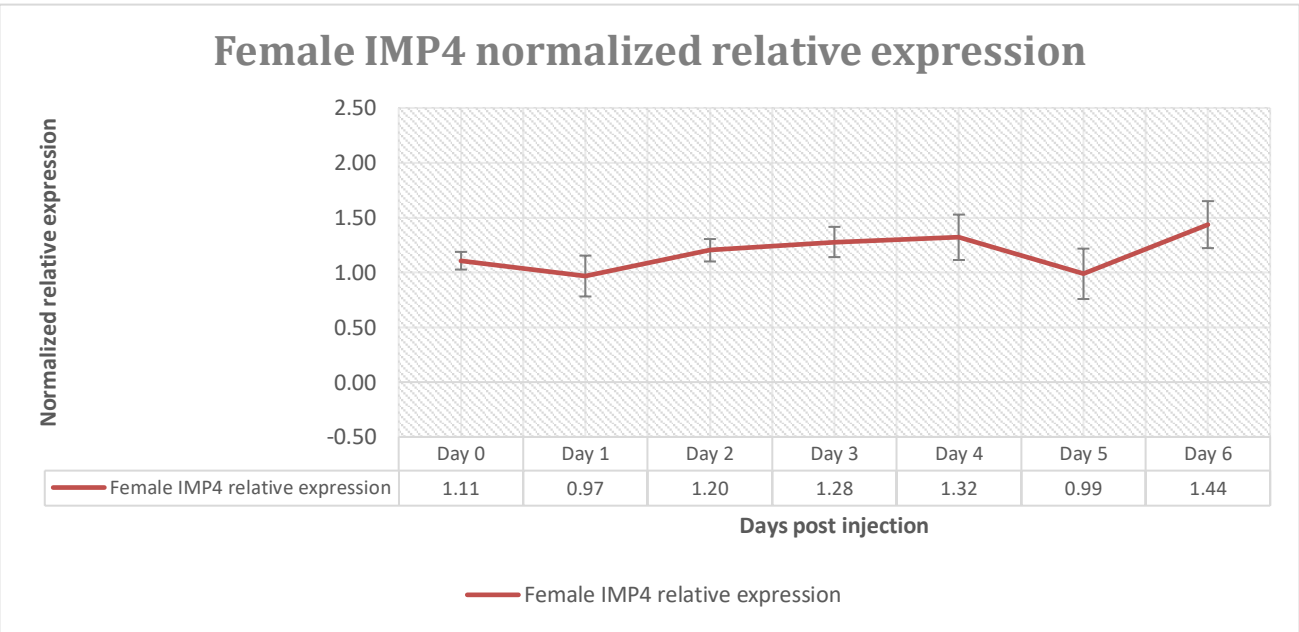

Figure 24 Normalized relative expression of importin-4 gene on females from day 0 to 6 post injection. Error is calculated as $\sqrt{((s1)^2/n1)} + \sqrt{((s2)^2/n2)}$*, where S1 and S2 are the standard deviations of dsGFP treated and dsIMP4 treated respectively while n1 and n2 are the group sizes.*

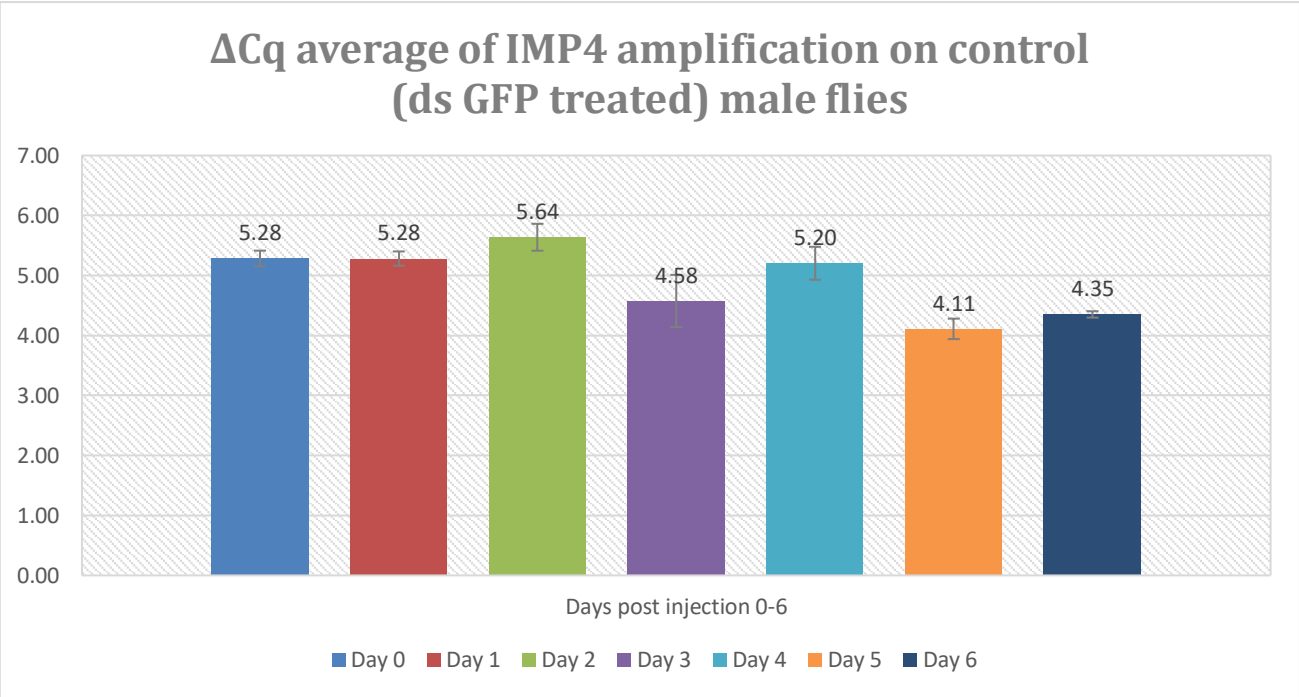

Figure 25 Average normalized ΔCq values from impotin 4 amplification on male control dsGFP treated flies at day 0 to 6 post injection. Error bars indicate the SD for each sample group.

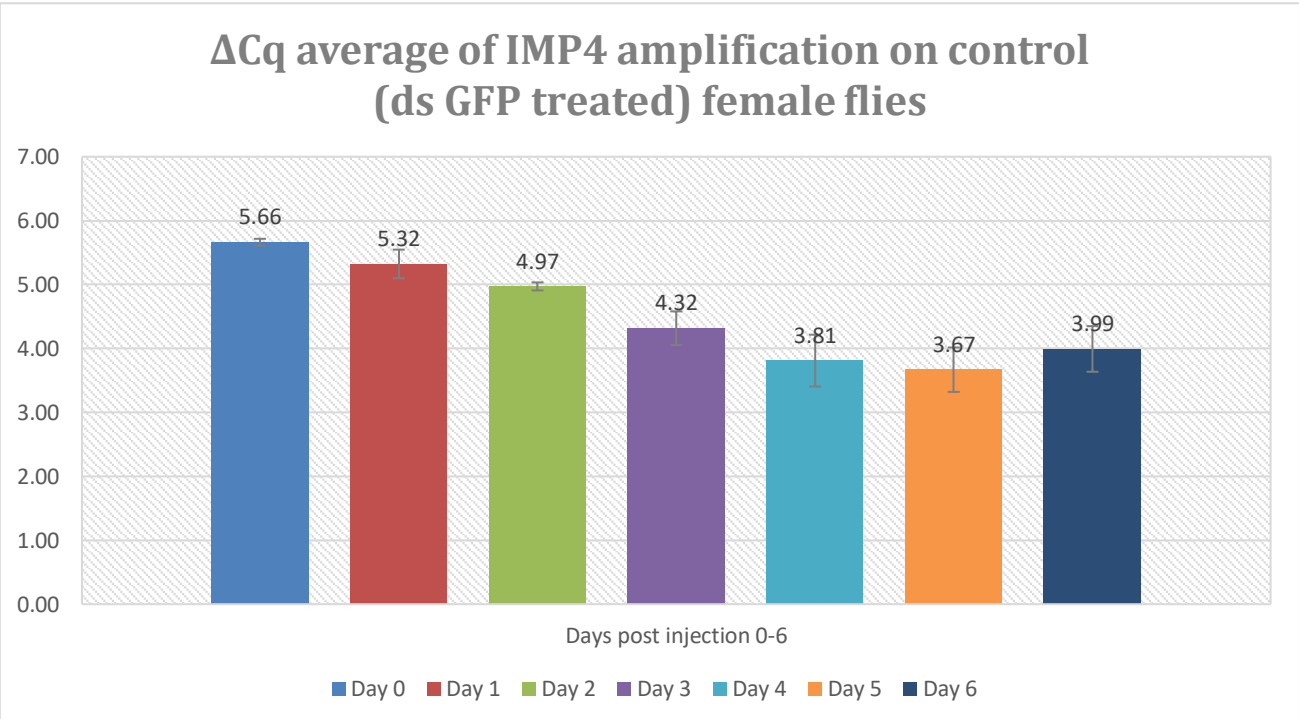

Figure 26 Average normalized ΔCq values from impotin 4 amplification on female control dsGFP treated flies at day 0 to 6 post injection. Error bars indicate the SD for each sample group

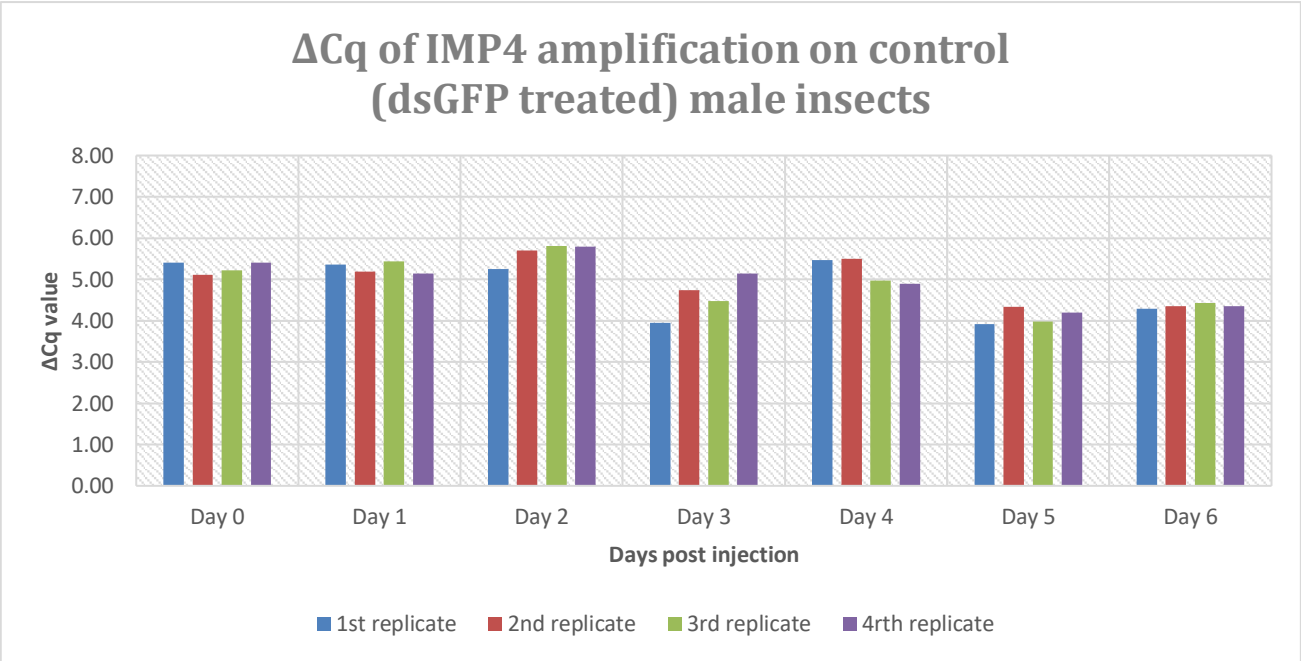

Figure 27 Normalized ΔCq values of importin 4 amplification on quadruplicates from dsGFP treated male insects sampled from day0 to 6 post injection.

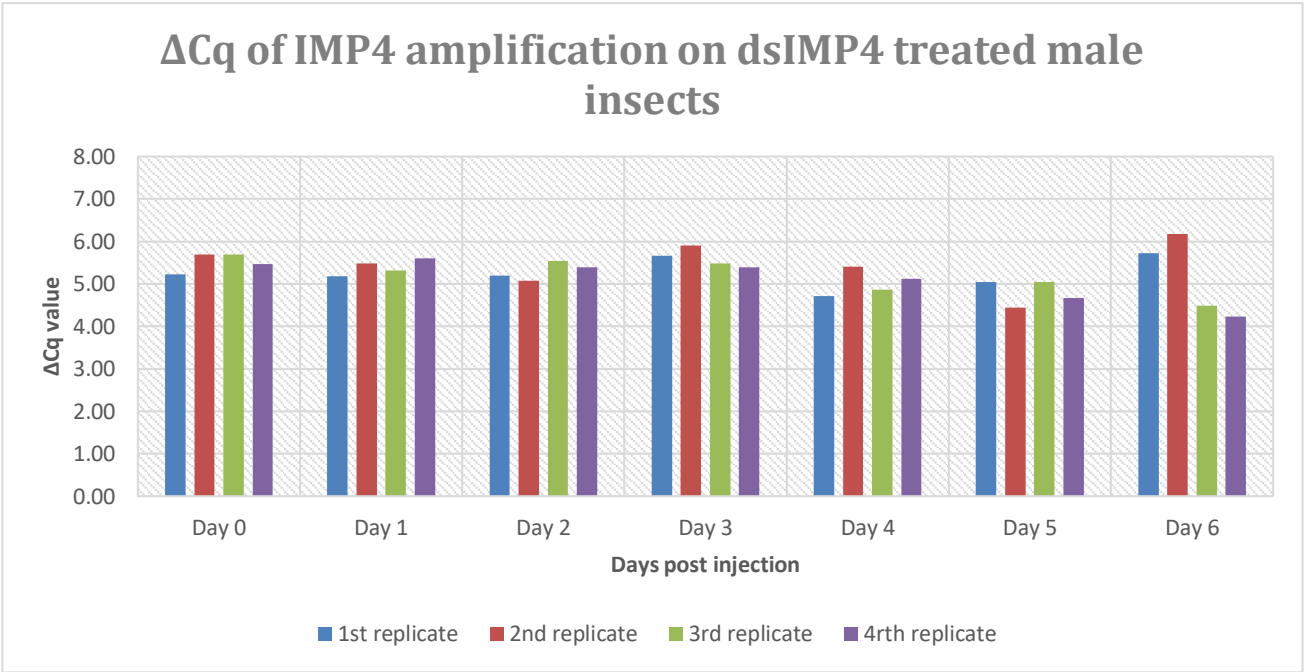

Figure 28 Normalized ΔCq values of importin 4 amplification on quadruplicates from dsIMP4 treated male insects sampled from day0 to 6 post injection.

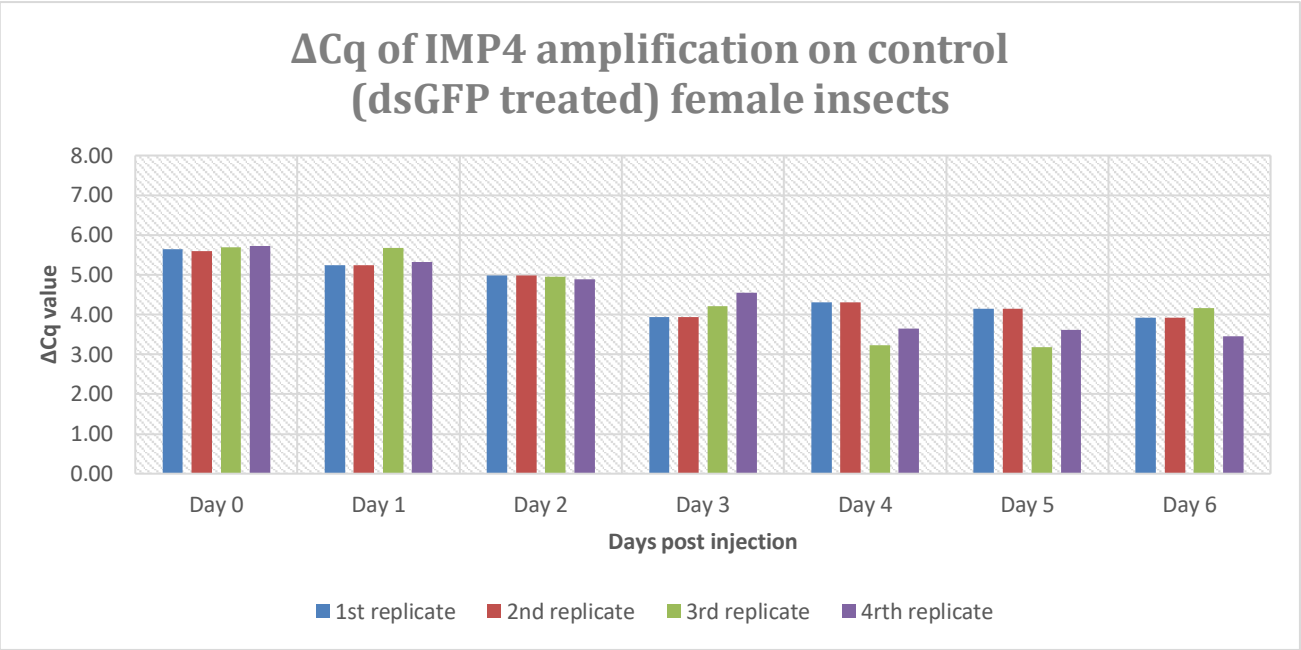

Figure 29 Normalized ΔCq values of importin 4 amplification on quadruplicates from dsGFP treated female insects sampled from day0 to 6 post injection.

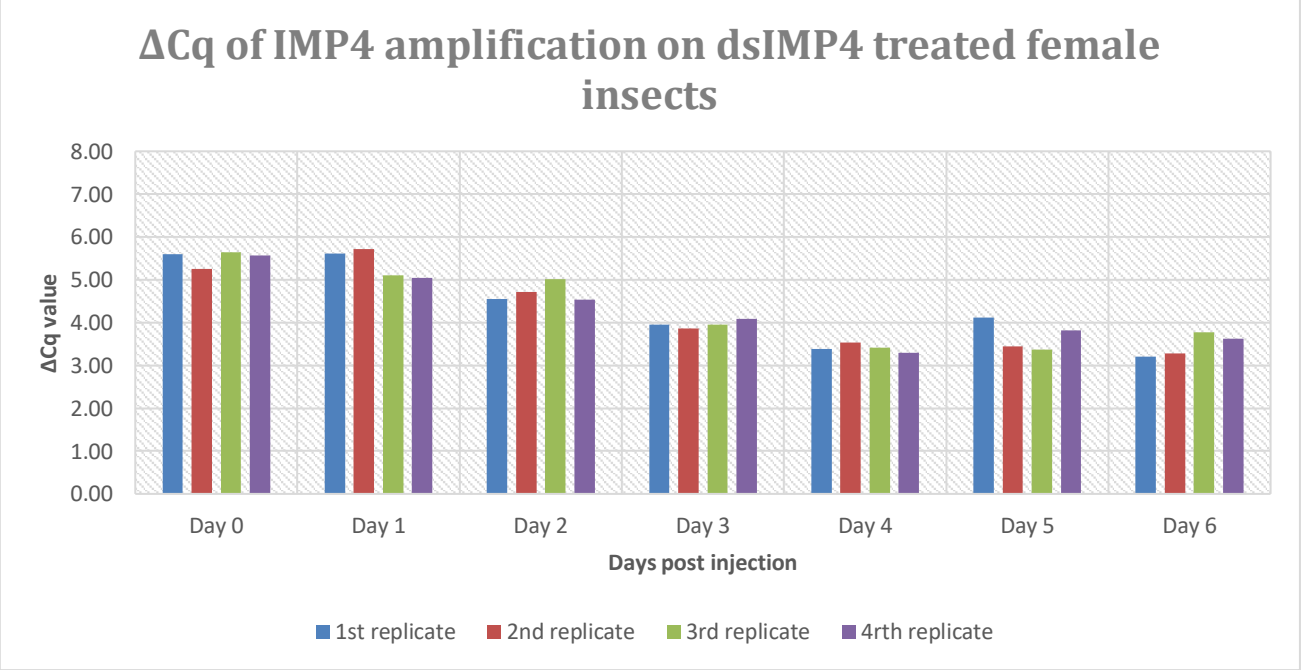

Figure 30 Normalized ΔCq values of importin 4 amplification on quadruplicates from dsIMP4 treated female insects sampled from day0 to 6 post injection.

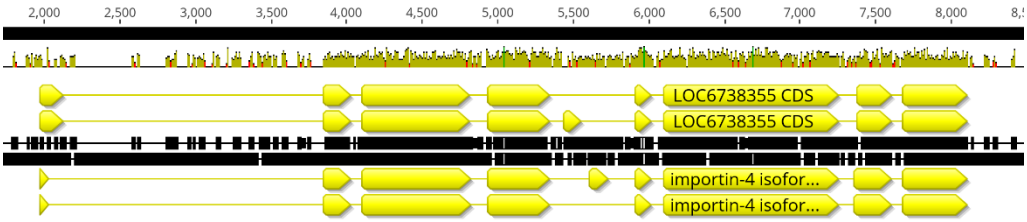

Figure 32 Graphic representation of nucleotide alignment between D. simulans (above) and B. oleae (bellow) importin 4 homologues. The intervening exon is located dissimilarly in the linear alignment. Consensus identity bar plot indicates level of sequence conservation in colour scale (red < green).

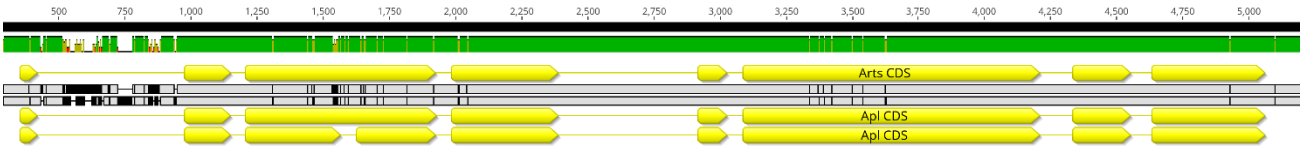

Figure 31 Graphic representation of nucleotide alignment between Artemis (above) and Apollo (bellow) copies. The additional intron of Apollo is visible. Consensus identity bar plot indicates level of sequence conservation in colour scale (red < green).

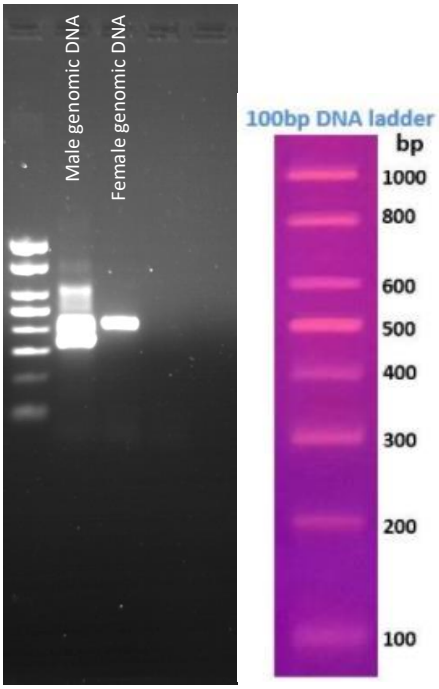

Figure 33 Amplification of template for in vitro transcription of importin 4. The use of male genomic DNA as template results in the amplification of major by-products due to amplification of imp4 on Y.

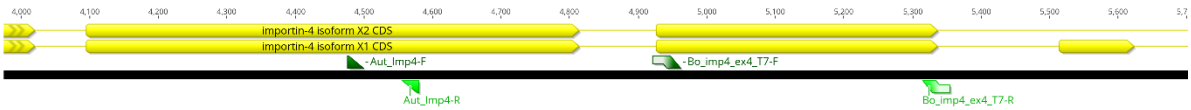

Figure 34 Exon 3 and 4 of the B. oleae autosomal importin 4 gene. T7 overhand containing primers (white-green) were used for amplification of the dsRNA transcription template. Aut_Imp4 primer pair (green) was used for the validation of silencing through qPCR.

References

- *Allshire RC, Madhani HD. (2018) Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol.; 19:229–244. 10.1038/nrm.2017.119*
- *B.J. Lee, A.E. Cansizoglu, K.E. Suel, T.H. Louis, Z. Zhang, Y.M. Chook, (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2, Cell 126 543–558.*
- *Bachtrog D, et al. (2014). Sex determination: why so many ways of doing it? PLoS Biol. 12(7):e1001899.*
- *Bachtrog D, Gordo I. (2004) Adaptive evolution of asexual populations under Muller's ratchet. Evol Int J Org Evol.; 58:1403–13.*
- *Bachtrog D. (2008) The temporal dynamics of processes underlying Y chromosome degeneration. Genetics.; 179:1513–25.*
- *Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. Nature Reviews Genetics, 14(2), 113–124.*
- *Bachtrog, D. (2020). The Y chromosome as a battleground for intragenomic conflict. Trends in Genetics, 36(7), 510–522.*
- *Bachtrog, D., Mahajan, S., & Bracewell, R. (2019). Massive gene amplification on a recently formed Drosophila Y chromosome. Nature Ecology and Evolution, 3(11), 1587–1597.*
- *Bayega, A., Djambazian, H., Tsoumani, K. T., Gregoriou, M. E., Sagri, E., Drosopoulou, E., Mavragani-Tsipidou, P., Giorda, K., Tsiamis, G., Bourtzis, K., Oikonomopoulos, S., Dewar, K., Church, D. M., Papanicolaou, A., Mathiopoulos, K. D., & Ragoussis, J. (2020). De novo assembly of the olive fruit fly (Bactrocera oleae) genome with linked-reads and long-read technologies minimizes gaps and provides exceptional y chromosome assembly. BMC Genomics, 21(1), 1–21.*
- *Bednenko, J., Cingolani, G., and Gerace, L. (2003) Importin contains a COOH-terminal nucleoporin binding region important for nuclear transport. J. Cell Biol. 162, 391– 401*
- *Blackman, RL. (1995) Sex determination in insects. In: Leather, SR.; Hardie, J., editors. Insect reproduction. CRC press;*
- *Blackmon, H., Ross, L., & Bachtrog, D. (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. Journal of Heredity, 108(1), 78–93.*
- *Bonaccorsi S, Lohe A. (1991). Fine mapping of satellite DNA sequences along the Y chromosome of Drosophila melanogaster: relationships between satellite sequences and fertility factors. Genetics 129:177–189.*
- *Bonchev G, Willi Y. (2018). Accumulation of transposable elements in selfing populations of Arabidopsis lyrata supports the ectopic recombination model of transposon evolution. New Phytol. 219(2):767–778.*
- *Brittnacher, J. G., and B. Ganetzky, (1989) On the components of segregation distortion in Drosophila melanogaster. IV. Construction and analysis of free duplications for the Responder locus. Genetics 121: 739–750*
- *Brown EJ, Nguyen AH, Bachtrog D. (2020) The Y chromosome may contribute to sex-specific ageing in Drosophila. Nat Ecol Evol. Jun;4(6):853-862. doi: 10.1038/s41559-020-1179-5. Epub 2020 Apr 20. PMID: 32313175; PMCID: PMC7274899.*
- *Bull JJ. (1983). The evolution of sex determining mechanisms. Menlo Park (CA): Benjamin/Cummings Publishing Company*
- *Bulmer M. G. and Parker G. A. (2002) The evolution of anisogamy: a game-theoretic approach Proc. R. Soc. Lond. B.2692381–2388*
- *Burrack, H. J., & Zalom, F. G. (2008). Olive fruit fly (Diptera: Tephritidae) ovipositional preference and larval performance in several commercial important olive varieties in California. Journal of Economic Entomology, 101, 750e758.*
- *Bush GL (1966) Female Heterogamety in the Family Tephritidae. Am Nat 100: 119–126.*
- *Carvalho AB (2002) Origin and evolution of the Drosophila Y chromosome. Curr. Opin. Genet. Dev 12, 664–668*
- *Carvalho AB. Origin and evolution of the Drosophila Y chromosome (2002). Current Opinion in Genetics & Development ;12:664–668.*
- *Carvalho, A. B., & Clark, A. G. (2013). Efficient identification of Y chromosome sequences in the human and drosophila genomes. Genome Research, 23(11), 1894–1907.*
- *Charlesworth B, Coyne JA, Barton NH. (1987) The relative rates of evolution of sex chromosomes and autosomes. Am Nat.; 130:113–46.*
- *Charlesworth B. (1978) Model for evolution of Y chromosomes and dosage compensation. Proc Natl Acad Sci USA.; 75:5618–22.*
- *Charlesworth B. (1996) The evolution of chromosomal sex determination and dosage compensation. Curr Biol.; 6:149–62.*
- *Chook, Y. M., & Süel, K. E. (2011). Nuclear import by karyopherin-βs: Recognition and inhibition. Biochimica et Biophysica Acta - Molecular Cell Research, 1813(9), 1593–1606.*
- *Clark FE, Kocher TD. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. Sci Rep. 9(1):20213*
- *Cline, T. W. (1993). The Drosophila sex determination signal: how do flies count to two? Trends Genet. 9, 385–390*
- *Cocquet J. et al. (2012) A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. PLoS Genet. 8, e1002900*
- *Courret Cécile, Chang Ching-Ho, Wei Kevin H.-C., Montchamp-Moreau Catherine and Larracuente Amanda M. (2019) Meiotic drive mechanisms: lessons from Drosophila Proc. Biol. Sci 286, 20191430*
- *Daane, K. M., & Johnson, M. W. (2010). Olive fruit fly: managing an ancient pest in modern times. Annual Review of Entomology, 55, 151e169.*

- *De Longo, O., Colombo, A., Gomez-Riera, P., Bartolucci, A., & Tan, K. H. (2000). The use of massive SIT for the control of the Medfly, Ceratitis capitata (Wied.), Strain SEIB 6-96, in Mendoza, Argentina. Area-wide control of fruit flies and other insect pests, 782.*
- *Denholm I and Rowland MW (1992) Tactics for managing pesticide resistance in arthropods: theory and practice. Annu Rev Entomol 37: 91-112*
- *Dorer DR, Henikoff S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell.; 77:993–1002.*
- *Dyck VA, Hendrichs J, Robinson AS. (2005). Sterile insect technique: principles and practice in area-wide integrated pest management. Dordrecht: Springer*
- *Economopoulos A. (1972) Sexual competitiveness of gamma-ray sterilized males of Dacus oleae. Mating frequency of artificially reared and wild females. Env Entomol. 490-497.*
- *Economopoulos AP and Zervas GA (1982) Sterile insect technique and radiation in insect control. IAEA-SM-255/39 357–368.*
- *Economopoulos AP, Avtzis N, Zervas G, Tsitsipis J, Haniotakis G, Tsiropoulos G, Manoukas A (1977) Experiments on control of olive Xy, Dacus oleae (Gmelin), by combined eVect of insecticides and releases of gamma-ray sterilized insects. J Appl Entomol 83:201–215.*
- *EnviroNews Forum (1999). Killer environment. Environmental Health Perspectives 107(2): A62.*
- *Foster JW, Graves JAM (1994). An Sry-Related Sequence on the Marsupial X-Chromosome - Implications for the Evolution of the Mammalian Testis determining Gene. Proc Natl Acad Sci U S A 91: 1927–1931.*
- *Frías D. (1992). Aspectos de la biología evolutiva de especies de Tephritidae (Diptera) de distribución chilena. Acta Entomol Chil. 17:69–79.*
- *Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., Shu, J. J., Mank, J. E., & Fraser, B. (2020). Sex Chromosome Evolution: So Many Exceptions to the Rules. Genome Biology and Evolution, 12(6), 750–763.*
- *Gabrieli, P., Gomulski, L. M., Bonomi, A., Siciliano, P., Scolari, F., Franz, G., Jessup, A., Malacrida, A. R., & Gasperi, G. (2011). Interchromosomal duplications on the Bactrocera oleae Y chromosome imply a distinct evolutionary origin of the sex chromosomes compared to Drosophila. PLoS ONE, 6(3).*
- *Garcia-Martinez V, Hernandez-Ortiz E, Zepeta-Cisneros CS, Robinson AS, Zacharopoulou A, et al. (2009) Mitotic and polytene chromosome analysis in the Mexican fruit fly, Anastrepha ludens (Loew) (Diptera: Tephritidae). Genome 52: 20–30. 5.*
- *Gatti M, Pimpinelli S. (1992) Functional elements in Drosophila melanogaster heterochromatin. Annu Rev Genet.; 26:239–75.*
- *Gorlich, D., and U. Kutay, (1999) Transport between the cell nucleus and the cytoplasm. Annu. Rev. Cell Dev. Biol. 15: 607– 660.*
- *Graves JAM. (2006) Sex chromosome specialization and degeneration in mammals. Cell.; 124:901–914.*
- *Guler Y, Short S, Kile P, Ford AT (2012) Integrating field and laboratory evidence for environmental sex determination in the amphipod, Echinogammarus marinus. Mar Biol 159: 2885–2890.*
- *Guo, L., Bloom, J. S., Dols-Serrate, D., Boocock, J., Ben-David, E., Schubert, O. T., Kozuma, K., Ho, K., Warda, E., Chui, C., Wei, Y., Leighton, D., Lemus Vergara, T., Riutort, M., Sánchez Alvarado, A., & Kruglyak, L. (2022). Island-specific evolution of a sex-primed autosome in a sexual planarian. Nature, 606(7913), 329–334.*
- *Haag ES, Doty AV (2005) Sex determination across evolution: connecting the dots. PLoS Biol 3: e21.*
- *Hackstein JH, et al. (1996) Is the Y chromosome of Drosophila an evolved supernumerary chromosome? Bioessays; 18:317–323.*
- *Hall, A.B., Qi, Y., Timoshevskiy, V. et al. (2013) Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. BMC Genomics 14, 273*
- *Hediger M, Henggeler C, Meier N, Perez R, Saccone G, et al. (2010). Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex determining pathway in the housefly. Genetics 184: 155–170.*
- *Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG, et al. . (2002). Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. Genome Biol. 3(12):research0085.1.*
- *Hoy, M. A. (2018). Insect Molecular Genetics. Elsevier Gezondheidszorg.*
- *Jiang, F., Liang, L., Wang, J. et al. (2022) Chromosome-level genome assembly of Bactrocera dorsalis reveals its adaptation and invasion mechanisms. Commun Biol 5, 25*
- *Kakani, E. G., Zygouridis, N. E., Tsoumani, K. T., Seraphides, N., Zalom, F. G., & Mathiopoulos, K. D. (2010). Spinosad resistance development in wild olive fruit fly Bactrocera oleae (Diptera: Tephritidae) populations in California. Pest Management Science, 66(4), 447–453.*
- *Kalab P, Weis K, Heald R. (2002) Visualization of a Ran-GTP gradient in interphase and mitotic Xenopus egg extracts. Science; 295:2452– 2456.*
- *Kamiya T, et al. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (Fugu). PLoS Genet. 8(7): e1002798.*
- *Kandul, N.P., Liu, J., Sanchez C., H.M. et al. (2019). Transforming insect population control with precision guided sterile males with demonstration in flies. Nat Commun 10, 84*
- *Kent TV, Uzunovic J, Wright SI. (2017). Coevolution between transposable elements and recombination. Philos Trans RSocB.372(1736):20160458.*

- *Kimura, N. Imamoto (2014). Biological significance of the importin-β familydependent nucleocytoplasmic transport pathways, Traffic 15 (7) 727–748.*
- *Knipling E. (1955). Possibilities of insect control or eradication through the use of sexually sterile malesJ. Econ. Entomol., 48, p. 459*
- *Koerich LB, Wang X, Clark AG, Carvalho AB (2008) Low conservation of gene content in the Drosophila Y chromosome. Nature 456: 949–951.*
- *Kopp A (2012) Dmrt genes in the development and evolution of sexual dimorphism. Trends Genet 28: 175–184.*
- *Kounatidis I, Papadopoulos N, Bourtzis K, Mavragani-Tsipidou P (2008) Genetic and cytogenetic analysis of the fruit fly Rhagoletis cerasi (Diptera: Tephritidae). Genome 51: 479–491. 6.*
- *Kusano, A., Staber, C., Chan, H. Y. E., & Ganetzky, B. (2003). Closing the (Ran)GAP on segregation distortion in Drosophila. BioEssays, 25(2), 108–115.*
- *Lahn BT, Page DC. (1999). Four evolutionary strata on the human X chromosome. Science 286(5441):964–967.*
- *Lange J, Skaletsky H, van Daalen SKM, Embry SL, Korver CM, Brown LG, Oates RD, Silber S, Repping S, Page DC. (2009) Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell.; 138:855–869.*
- *Larracuente, A. M., & Presgraves, D. C. (2012). The selfish Segregation Distorter gene complex of Drosophila melanogaster. Genetics, 192(1), 33–53.*
- *Lemos, B., Araripe, L. O., & Hartl, D. L. (2008). Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science, 319(5859), 91–93.*
- *Lin CJ, Hu F, Dubruille R, Vedanayagam J, Wen J, Smibert P, Loppin B, Lai EC. (2018) The hpRNA/ RNAi pathway is essential to resolve intragenomic conflict in the Drosophila male germline. Dev. Cell 46, 316–326.*
- *Lints, Robyn, and Scott W. Emmons. (2002) Regulation of Sex-Specific Differentiation and Mating Behavior in C. Elegans by a New Member of the DM Domain Transcription Factor Family, 15 Sept. 2002, genesdev.cshlp.org/content/16/18/2390.short.*
- *Lippman Z, Gendrel A, Black M, Vaughn M, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature.; 430:471–76.*
- *Liu, X., Wu, B., Szary, J., Kofoed, E. M., & Schaufele, F. (2007). Functional sequestration of transcription factor activity by repetitive DNA. Journal of Biological Chemistry, 282(29), 20868–20876.*
- *Loveland, K. L., Major, A. T., Butler, R., Young, J. C., Jans, D. A., & Miyamoto, Y. (2015). Putting things in place for fertilization: Discovering roles for importin proteins in cell fate and spermatogenesis. Asian Journal of Andrology, 17(4), 537–544.*
- *Mahajan R, Gerace L, Melchior F. (1998) Molecular characterization of the SUMO-1 modification of RanGAP1 and its role in nuclear envelope association. J Cell Biol;140:259–270*
- *Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of y chromosome gene content in flies. Nature Communications, 8(1).*
- *Maier, M. C., McInerney, M. R. A., Graves, J. A. M., & Charchar, F. J. (2021). Noncoding Genes on Sex Chromosomes and Their Function in Sex Determination, Dosage Compensation, Male Traits, and Diseases. Sexual Development, 15(5–6), 432–440.*
- *Major, A. T., Whiley, P. A. F., & Loveland, K. L. (2011). Expression of nucleocytoplasmic transport machinery: Clues to regulation of spermatogenic development. Biochimica et Biophysica Acta - Molecular Cell Research, 1813(9), 1668–1688.*
- *Malik HS, Eickbush TH, Goldfarb DS. (1997) Evolutionary specialization of the nuclear targeting apparatus. Proc Natl Acad Sci U S A.;94(25):13738–42*
- *Mank, J. E. (2012). Small but mighty: The evolutionary dynamics of W and y sex chromosomes. Chromosome Research, 20(1), 21–33.*
- *Mansour, S. A., A. H. Mossa, and T. M. Heikal. 2008a. Cytogenetic and hormonal alteration in rats exposed to recommended "safe doses" of spinosad and malathion insecticides. Int. J. Agric. Biol. 10:9–14.*
- *Mansour, S. A., T. M. Heikal, and A. H. Mossa. 2008b. Biochemical and histopathological effects of formulations containing malathion and spinosad in rats. Toxicol. Int. 15:71–78.*
- *Marais GAB, Campos PRA, Gordo I. (2010) Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. Genome Biol Evol.; 2:347–357.*
- *Margaritopoulos JT, Skavdis G, Kalogiannis N, Nikou D, Morou E, Skouras PJ, Tsitsipis JA, Vontas J (2008). Efficacy of the pyrethroid alpha-cypermethrin against Bactrocera oleae populations from Greece and improved diagnostic foran iAChE mutation, Pest Manag Sci. 64 900–908.*
- *Mason DA, Fleming RJ, Goldfarb DS. (2002) Drosophila melanogaster importin alpha1 and alpha3 can replace importin alpha2 during spermatogenesis but not oogenesis. Genetics.;161(1):157–70.*
- *Mason DA, Stage DE, Goldfarb DS. (2009) Evolution of the metazoan-specifc importin alpha gene family. J Mol Evol.;68(4):351–65*
- *Maudlin I. (1979) Chromosome polymorphism and sex determination in a wild population of tsetse. Nature; 277:300–301.*
- *Mavragani-Tsipidou P, Karamanlidou G, Zacharopoulou A, Koliais S, Kastritisis C (1992) Mitotic and polytene chromosome analysis in Dacus oleae (Diptera: Tephritidae). Genome 35: 373–378. 7.*

- *Meccariello, A., Salvemini, M., Primo, P., Hall, B., Koskinioti, P., Dalíková, M., Gravina, A., Gucciardino, M. A., Forlenza, F., Gregoriou, M. E., Ippolito, D., Monti, S. M., Petrella, V., Perrotta, M. M., Schmeing, S., Ruggiero, A., Scolari, F., Giordano, E., Tsoumani, K. T., … Saccone, G. (2019). Maleness-on-the-Y (MoY) orchestrates male sex determination in major agricultural fruit fly pests. Science, 365(6460), 1457–1460.*
- *Merchant-Larios H, Diaz-Hernandez V (2013) Environmental sex determination mechanisms in reptiles. Sex Dev 7: 95–103.*
- *Merrill, C., L. Bayraktaroglu, A. Kusano, and B. Ganetzky, (1999) Truncated RanGAP encoded by the Segregation Distorter locus of Drosophila. Science 283: 1742–1745.*
- *Miller, D., M. Brinkworth, and D. Iles, (2010) Paternal DNA packaging in spermatozoa: More than the sum of its parts?: DNA, histones, protamines and epigenetics. Reproduction 139: 287– 301.*
- *Mirsalehi, A., Markova, D. N., Eslamieh, M., & Betrán, E. (2021). Nuclear transport genes recurrently duplicate by means of RNA intermediates in Drosophila but not in other insects. BMC Genomics, 22(1), 1–13.*
- *Molina, E., Chew, G.S., Myers, S.A. et al. (2017). A Novel Y-Specific Long Non-Coding RNA Associated with Cellular Lipid Accumulation in HepG2 cells and Atherosclerosis-related Genes. Sci Rep **7**, 16710*
- *Mořkovský, L., Storchová, R., Plachý, J. et al. (2010). The Chicken Z Chromosome Is Enriched for Genes with Preferential Expression in Ovarian Somatic Cells. J Mol Evol 70, 129–136*
- *Muller, H. J. (1940). The New Systematics (ed. Huxley, J.) 185–268 (Clarendon Press,).*
- *Nakamura, M. (2009) Sex determination in amphibians. Semin. Cell Dev. Biol. 20, 271–282*
- *Nardi, F., Carapelli, A., Boore, J. L., Roderick, G. K., Dallai, R., & Frati, F. (2010). Domestication of olive fly through a multi-regional host shift to cultivated olives: comparative dating using complete mitochondrial genomes. Molecular Phylogenetics and Evolution, 57, 678e686.*
- *Nardi, F., Carapelli, A., Vontas, J. G., Dallai, R., Roderick, G. K., & Frati, F. (2006). Geographical distribution and evolutionary history of organophosphate-resistant Ace alleles in the olive fly (Bactrocera oleae). Insect Biochemistry and Molecular Biology, 36, 593e602.*
- *Nguyen AH, Bachtrog D. (2021) Toxic Y chromosome: Increased repeat expression and age-associated heterochromatin loss in male Drosophila with a young Y chromosome. PLoS Genet. Apr 22;17(4):e1009438.*
- *Nokkala S, Grozeva S, Kuznetsova V, Maryanska-Nadachowska A. (2003). The origin of the achiasmatic XY sex chromosome system in Cacopsylla peregrina (Frst.) (Psylloidea, Homoptera). Genetica 119(3):327–332*
- *Orankanok W, Chinvinijkul S, Sittilob P, Thanaphum S, Somboonprasert P, Enkerlin WR (2006). Integrating the sterile insect technique (SIT) for fruit fly control in Thailand. In: Proceeding 7th economic fruit fly symposium, 10–15 September 2006, Salvador, Bahia*
- *Orankanok W, Chinvinijkul S, Thanaphum S, Sittilob P, Enkerlin WR (2007). Area-wide integrated control of oriental fruit fly Bactrocera dorsalis and guava fruit fly Bactrocera correcta in Thailand. In: Vreysen MJB, Hendrichs J, Robinson AS (eds) Area-wide control of insect pests: from research to field implementation. Springer, Dordrecht, pp 517–526*
- *Otto, Sarah P. (2009). "The evolutionary enigma of sex." The American naturalist vol. 174 Suppl 1: S1-S14. doi:10.1086/599084*
- *Pagendam, D.E., Trewin, B.J., Snoad, N. et al. (2020) Modelling the Wolbachia incompatible insect technique: strategies for effective mosquito population elimination. BMC Biol 18, 161*
- *Phadnis N, Hsieh E, Malik HS. (2012) Birth, death, and replacement of karyopherins in Drosophila. Mol Biol Evol.;29(5):1429–40.*
- *Phuc, H.K., Andreasen, M.H., Burton, R.S. et al. (2007) Late-acting dominant lethal genetic systems and mosquito control . BMC Biol 5, 11*
- *Pimpinelli S, Dimitri P. (1989) Cytogenetic analysis of Segregation Distortion in Drosophila melanogaster: the cytological organization of the Responder (Rsp) locus. Genetics;121:765–772.*
- *Pomiankowski A, Nothiger R, Wilkins A. (2004). The evolution of the Drosophila sex-determination pathway. Genetics.; 166:1761–73.*
- *Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jegou B, Kaessmann H. (2008) Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biology.; 6:709–716.*
- *Powers, P. A., and B. Ganetzky, (1991) On the components of segregation distortion in Drosophila melanogaster. V. Molecular analysis of the Sd locus. Genetics 129: 133–144.*
- *Pradeepa, M. M., Manjunatha, S., Sathish, V., Agrawal, S., & Rao, M. R. S. (2008). Involvement of Importin-4 in the Transport of Transition Protein 2 into the Spermatid Nucleus. Molecular and Cellular Biology, 28(13), 4331–4341.*
- *Presgraves D, Orr HA. (1998). Haldane's rule in taxa lacking a hemizygous X. Science; 282:889– 91.*
- *Quimby, B. B., and M. Dasso, (2003) The small GTPase Ran: interpreting the signs. Curr. Opin. Cell Biol. 15: 338–344.*
- *Renshaw, M. J., and A. Wilde, (2011) The Role of RanGTPase in Mitotic Spindle Assembly, eLS.*
- *Rice WR. (1987) Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics.; 116:161–67.*
- *Rice WR. (1987). The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. Evolution 41(4):911–914*

- *Rozen, S. et al. (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423, 873–876.*
- *S.J. Lee, T. Sekimoto, E. Yamashita, E. Nagoshi, A. Nakagawa, N. Imamoto, M. Yoshimura, H. Sakai, K.T. Chong, T. Tsukihara, Y. Yoneda, (2003) The structure of importin-beta bound to SREBP-2: nuclear import of a transcription factor, Science 302 1571–1575.*
- *Saccone, G; Peluso, I; Artiaco, D; Giordano, E; Bopp, D; Polito, L C (1998). The Ceratitis capitata homologue of the Drosophila sex-determining gene sex-lethal is structurally conserved, but not sex-specifically regulated. Development, 125(8):1495-1500.*
- *Sagri, E., Koskinioti, P., Gregoriou, ME. et al. (2017). Housekeeping in Tephritid insects: the best gene choice for expression analyses in the medfly and the olive fly. Sci Rep 7, 45634*
- *Sahara K, Yoshido A, Traut W. (2012). Sex chromosome evolution in moths and butterflies. Chromosome Res. 20:83–94.*
- *Sandler L and Novitski E (1957) Meiotic Drive as an Evolutionary Force. The American Naturalist 91, 105–110*
- *Sarno, Francesca & Ruiz, Maria Fernanda & Eirin-Lopez, Jose & Perondini, A.L.P. & Selivon, Denise & Sánchez, Lucas. (2010). The gene transformer-2 of Anastrepha fruit flies (Diptera, Tephritidae) and its evolution in insects. BMC evolutionary biology. 10. 140. 10.1186/1471-2148-10-140.*
- *Schotta, G., Ebert, A., Dorn, R., & Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. Seminars in Cell & Developmental Biology, 14(1), 67–75.*
- *Shearman, D. (2002) The Evolution of Sex Determination Systems in Dipteran Insects Other than Drosophila. Genetica 116, 25–43*
- *Skaletsky H. et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825–837*
- *Skaletsky, H. et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825–837*
- *Skouras PJ, Margaritopoulos JT, Seraphides NA, Ioannides IM, Kakani EG, Mathiopoulos KD and Tsitsipis JA (2007). Organophosphate resistance in olive fruit fly, Bactrocera oleae, populations in Greece and Cyprus. Society of Chemical Industry. Pest Manag Sci 63(1):42-8*
- *Skouras PJ, Margaritopoulos JT, Seraphides NA, Ioannides IM, Kakani EG, Mathiopoulos KD and Tsitsipis JA (2007) Organophosphate resistance in olive fruit fly, Bactrocera oleae, populations in Greece and Cyprus. Society of Chemical Industry. Pest Manag Sci 63(1):42-8*
- *Śliwińska EB, Martyka R, Tryjanowski P. (2016) Evolutionary interaction between W/Y chromosome and transposable elements. Genetica. Jun;144(3):267-78.*
- *Steinemann S, Steinemann M. (2005) Y chromosomes: born to be destroyed. Bioessays.; 27:1076–1083.*
- *Tan CP, Sinigaglia L, Gomez V, Nicholls J, Habib NA. (2021) RNA Activation-A Novel Approach to Therapeutically Upregulate Gene Transcription. Molecules. Oct 28;26(21):6530.*
- *Thomas DB, Mangan RL (2005). Non target impact of spinosad GF-120 bait sprays for control of the Mexican fruit fly (Diptera: Tephritidae) in Texas citrus. J. Econ. Entomol. 98:1950–56*
- *Tokuyasu, K. T., W. J. Peacock, and R. W. Hardy, (1977) Dynamics of spermiogenesis in Drosophila melanogaster. VII. Effects of segregation distorter (SD) chromosome. J. Ultrastruct. Res. 58: 96–10*
- *Tomaszkiewicz M, Medvedev P, Makova KD. (2017) Y and W Chromosome Assemblies: Approaches and Discoveries. Trends Genet.;33:266–82.*
- *Tree of Sex Consortium. (2014). Tree of Sex: a database of sexual systems. Sci Data. 1:140015.*
- *Tsoumani KT, Drosopoulou E, Bourtzis K, Gariou-Papalexiou A, Mavragani-Tsipidou P, et al. (2015) Achilles, a New Family of Transcriptionally Active Retrotransposons from the Olive Fruit Fly, with Y Chromosome Preferential Distribution. PLOS ONE 10(9): e0137050.*
- *Tsoumani, K.T., Belavilas-Trovas, A., Gregoriou, ME. et al. (2020). Anosmic flies: what Orco silencing does to olive fruit flies. BMC Genet 21 (Suppl 2), 140*
- *Tzanakakis M. (1989) Small scale rearing: Dacus oleae. Fruit Flies.;3:105 –18.*
- *Vankuren, N. W., & Long, M. (2018). Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. Nature Ecology and Evolution, 2(4), 705–712.*
- *Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. (2013). Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11(8):e1001643.*
- *Vontas J, Hernandez-Crespo P, Margaritopoulos JT, Ortego F, Feng HT, Mathiopoulos KD, Hsu JS (2011) Insecticide resistance in Tephritid flies. Pesticide Biochemistry and Physiology 100: 199-205.*
- *Warner RR, Fitch DL, Standish JD (1996) Social control of sex change in the shelf limpet, Crepidula norrisiarum: size-specific responses to local group composition. J Exp Mar Bio Ecol 204: 155–167.*
- *Wright AE, Harrison PW, Montgomery SH, Pointer MA, Mank JE. (2014). Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. Evolution 68(11):3281–3295.*

- *Wu CI, Lyttle TW, Wu ML, Lin GF. (1988) ssociation between DNA satellite sequences and the Responder of Segregation Distorter in D. melanogaster. Cell;54:179–189.*
- *Yano, B. L., D. M. Bond, M. N. Novilla, L. G. McFadden, and M. J. Reasor. 2002. Spinosad insecticide: Subchronic and chronic toxicity and lack of carcinogenicity in Fischer 344 rats. Toxicol. Sci. 65:288–98. doi:10.1093/toxsci/65.2.288.*
- *Yoshimoto S, et al. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in Xenopus laevis. Proc Natl Acad Sci U S A. 105(7):2469–2474.*
- *Zalom, F. G., Burrack, H. J., Bingham, R., Price, R., & Ferguson, L. (2008). Olive fruit fly (Bactrocera oleae) introduction and establishment in California. Acta Horticulturae, 791, 619e628.*
- *Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, et al. (2002). Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science.; 298:149– 59.*
- *Zhang L, Sun T, Woldesellassie F, Xiao H, Tao Y. (2015) Sex ratio meiotic drive as a plausible evolutionary mechanism for hybrid male sterility. PLoS Genet. 11, e1005073.*
- *Zhao JT, Frommer M, Sved JA, Zacharopoulou A (1998) Mitotic and polytene chromosome analyses in the Queensland fruit fly, Bactrocera tryoni (Diptera: Tephritidae). Genome 41: 510–526.*
- *Zhou Q and Bachtrog D (2012) Sex-specific adaptation drives early sex chromosome evolution in Drosophila. Science 337, 341–345*