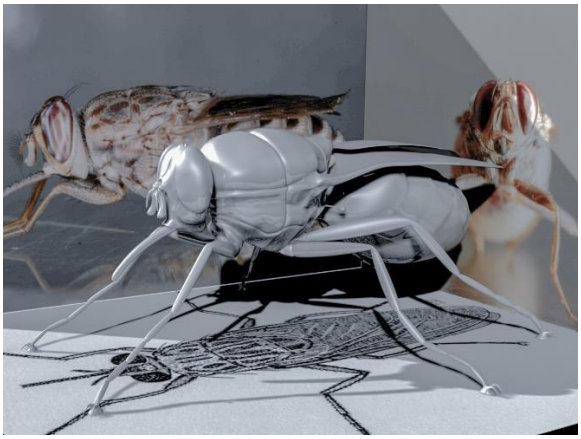
UNIVERSITA' DEGLI STUDI DI PAVIA

Dipartimento di Biologia e Biotecnologie

Genomic and molecular evolutionary analyses of insects of economic and public health relevance



Grazia Savini

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
XXIX Ciclo – A.A. 2013-2016

UNIVERSITA' DEGLI STUDI DI PAVIA
Dipartimento di Biologia e Biotecnologie

Genomic and molecular evolutionary analyses of insects of economic and public health relevance

Grazia Savini

Supervised by Prof. Anna R. Malacrida

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
XXIX Ciclo – A.A. 2013-2016

[Cover image courtesy of Geoffrey Attardo (Yale School of Public Health)]

Acknowledgements

I would like to express my special appreciation and thanks to Professors Anna R. Malacrida and Giuliano Gasperi, you have been inspiring mentors for me. I would like to thank you for encouraging my research and for being so passionate about science.

I am very grateful to Professor Serap Aksoy and to all members of her lab. Thank you for giving me such an extraordinary opportunity. I loved everyone in the lab and I really enjoyed the time I spent in New Haven.

A special thanks to Francesca Scolari, Geoffrey Attardo and Lino Ometto. Each of you taught me most of the things I learned during these years. Thank you Francesca for your invaluable support, Thanks Geoff for your impressive pictures and thank you Lino for your private lessons on molecular evolution!

Thanks to Prof. Ludvik Gomulski and Omar Rota-Stabelli for providing valuable suggestions and support for my work.

Thanks to my colleagues of the insects genetics and genomics laboratory, each of you gave me special memories.

Finally, thanks to mum and dad. Without your support I would never be able to achieve this goal.

Abstract

Human African Trypanosomiasis, or sleeping sickness, causes 10,000 deaths per year. Despite the extensive knowledge of trypanosome biology acquired over the last 25 years, the toolbox for disease control is still limited, with neither vaccines nor effective and affordable drugs available in the near future (Aksoy and IGGI, 2010). Eradication of trypanosomiasis via tsetse populations control has a great potential and novel genomic resources are needed to improve currently available tools and develop new effective strategies.

In 2015, the release of genomic resources of five *Glossina* species has facilitated comparative genomics analyses with the genome of *G. m. morsitans* (Attardo *et al.*, 2014) and other Diptera genomes already available, providing additional information on vector competence, haematophagy, viviparity, as well as on species-specific genes involved in host seeking and vectorial capacity.

In this framework, I firstly reconstructed the molecular phylogeny of the six species (*G. morsitans, G. pallidipes, G. austeni, G. fuscipes, G. palpalis, G. brevipalpis*) using a genome-wide set of orthologous genes. Our genomic phylogenetic analysis suggests that Morsitans, Palpalis and Fusca groups all show a monophyletic origin, with *G. brevipalpis* (Fusca group) being the most differentiated species. It is noteworthy that the phylogenetic relationships established by our analysis reflects the biological diversities exhibited by each of the *Glossina* group, such as the geographical distribution, the eco/ethological constrains, the degrees of vectorial capacity and the mating behavior.

Based on this phylogeny, I used the molecular clock approach to derive the time of splitting among the *Glossina* groups and the included species. Orthologs genes derived from *Musca domestica* allowed us to date also the divergence between *Musca* and the *Glossina* genus at 209 Mya, in the Triassic period. Within the *Glossina* genus, the divergence time has been estimated between the most distantly related *G. brevipalpis* and the Palpalis and Morsitans groups at~ 54 Mya, i.e. Eocene period, while the divergence between the Morsitans and Palpalis groups has been dated at~21 Mya, i.e. Miocene period.

The availability of genome sequences from six *Glossina* species provides unprecedented opportunities to study the evolution of reproductive traits relevant for interpreting the species differentiation. Indeed, reproduction is one of the fundamental processes influencing several aspects of insect's life and it is often the target of control methods aiming at reducing the insect population size, particularly effective in reproductive systems similar to the one displayed by tsetse flies (Vreysen *et al.*, 2000). In this framework, a key role is played by male reproductive genes that encode for proteins involved in fertilization, such as seminal fluid proteins (SFPs). These proteins are fundamental in influencing female post-mating physiological and behavioral changes as well as other processes involved in sexual competition. I identified gene candidates for positive selection among those expressed in the male reproductive tissues, male

accessory glands (MAGs) and testes. Our analysis was based on 2,563 genes selected from *G. m. morsitans* MAG- and testes-specific transcriptomes, because of their higher expression in MAG compared to testes, and vice versa. The genes were tested for the occurrence of positive selection using codon substitution models in PAML. The analysis led to the identification of a number of genes candidate for positive selection. Interestingly, most of the identified genes remained uncharacterized. These "novel" genes putatively involved in species- or group-specific processes are particularly interesting for further evolutionary and applicative analyses.

Considering the applicative potential of fast evolving and species-specific MAG sequences, I analyzed the regulatory elements responsible for the transcription of the 24 most highly transcribed accessory glands genes in *G. m. morsitans*. MAG-specific transcriptomic data revealed the presence of a transcription factor known as Paired. *In silico* analysis confirmed the transcription binding in the promoter region of most of the genes under study. The Paired transcription factor expression profile and protein localization have been demonstrated to be MAG-specific, thus suggesting a role in regulating the expression of MAG genes and, in turn, male reproduction. Its important role in reproduction is also confirmed by its sequence conservation among the six *Glossina* species analyzed. From the applicative point of view, it can be an important target for the manipulation of male reproduction in tsetse flies. Under this perspective, I have started performing RNAi experiments in order to evaluate the silencing effects of Paired on ejaculate composition, fertility and female post-mating behaviors.

During my PhD studies, I have also been involved in the genome projects of two insect species: the Asian tiger mosquito *Ae. albopictus* and the medfly *C. capitata*. In the framework of the *Ae. albopictus* genome project (Italian strain), I have been involved in the determination of the genome size of the Fellini strain and, for both genomes, part of my research was dedicated to the manual curation and annotation of Odorant Binding Proteins (OBPs) and Odorant Receptor (OR) genes.

Abbreviations and acronyms

AAT	Animal African Trypanosomiasis
AW-IPM	Area Wide Integrated Pest Management
BBB	Blood-Brain-Barrier
CDS	Coding Sequence
CSP	Chemosensory Protein
dN	Nonsynonymous substitution
DP	Differential Product
DPE	Downstream core Promoter Element
dS	Synonymous substitution
FDR	False Discovery rate
Gaut	Glossina austeni
Gbre	Glossina brevipalpis
Gfui	Glossina fuscipes
Gmor	Glossina morsitans
GO	Gene Ontology
Gpai	Glossina pallidipes
Gppi	Glossina palpalis
GR	Gustatory Receptor
HAT	Human African Trypanosomiasis
IGGI	International *Glossina* Genome Initiative
Inr	Initiator motif
IR	Ionotropic Receptor
MAGs	Male Accessory Glands
Mya	Million Years Ago
NTDs	Neglected Tropical Diseases
OBP	Odorant-Binding Proteins
ODE	Odorant degrading enzyme
OR	Odorant Receptor
OR	Olfactory Receptor
Orco	Odorant Receptor co-receptor
OSN	Olfactory Receptor neuron
prd	Paired transcription factor
RDA	Representational Difference Analysis
SFPs	Seminal Fluid Proteins
SGHV	Salivary Gland Hypertrophy Virus
SIT	Sterile Insect Technique

Table of Contents

Chapter 1: General introduction

1.1 Insects disease vectors and agricultural pests

The past two decades have witnessed a growing interest toward several aspects of insect's biology. The reasons are diverse and complex:

- the growing insect pest issues related to the intensification of crop-livestock production systems;
- the increasingly frequent invasions of exotic disease vectors and agricultural pests due to trade globalization and climate change;
- the increased disease outbreaks (including Zika fever, dengue, chikungunya) worldwide;
- the spread of insecticide resistance and, consequently, demand for alternative/more efficient strategies for environmentally-friendly insect pest management.

Vector-borne diseases account for more than 17% of all infectious diseases, causing more than 1 million deaths annually. Most of the organisms responsible for such diseases are bloodsucking insects, such as mosquitoes, ticks, tsetse flies, sandflies and fleas that can spread the disease through humans and animals during blood meals (http://www.who.int/mediacentre/factsheets/fs387/en/). Malaria alone is responsible for causing more than 400.000 deaths every year globally. While malaria and dengue are two of the most relevant insect-borne diseases, several other particularly affect marginalized populations living in tropical and sub-tropical regions of the world. Many of these are regarded as Neglected Tropical Diseases (NTDs).

Insects are also major constraints to agriculture, forestry and fishery production causing, in Europe, losses for at least 12 billion EUR per year (Scalera *et al.*, 2012; Szyniszewska and Tatem, 2014). Some species are highly invasive and polyphagous, and, as such, they threaten the global horticultural production. Among these is the Mediterranean fruit fly (medfly) *Ceratitis capitata*, with an estimated damage amounting to US $192 million per year in countries of the Mediterranean Basin.

Despite considerable research, the toolbox for disease and pest control is limited, with neither vaccines nor effective and affordable drugs available in the near future. So far, the most effective measures against insect-borne diseases and agricultural losses have been those targeted at the insect itself.

1.2 Insect population control in the genomic era

High-throughput sequencing technologies has brought a revolution in the study of insect biology. The genome sequences of a huge number of non-model organisms have been released, allowing comparative analyses of gene families expansion and positive selection, and consequently inferences of adaptive evolution. These resources are also helping in revealing demography, local adaptation and speciation mechanisms.

Genomic and phylogenomic resources are particularly relevant for insect pests and disease vectors since they can lead to the identification of the molecular bases of biological traits responsible for high reproductive rate, ability to resist to infections and frequency of contact with the host (Hill *et al.*, 2005). Taking into account that even the same arthropod borne disease can be transmitted by more than one insect species, the development of successful vector control programs has always been severely impaired by poorly understood complex ecological traits and intricate population dynamics (Hill *et al.*, 2005). Similarly, the availability of genomic information derived from polyphagous, oligophagous and monophagous insect pests will allow to understand the molecular bases of feeding behavior, invasive capacity and adaptability to different environments and climates.

As resistance to insecticides is emerging, alternative approaches are required. One option is represented by the Sterile Insect Technique (SIT), an environmentally-friendly method of insect pest/vector that integrates into Area Wide Integrated Pest Management (AW-IPM). The SIT relies on the mass rearing, sterilization and release of male insects into the field. Mating of these released sterile males with wild females leads to a decreased females' reproductive potential, thus leading to a reduction in population size and, ultimately, to the local elimination or suppression of the population.

SIT programs have been proven to be successful in the elimination of the New World screwworm *Cochliomyia hominivorax* Coquerel from United States, Mexico, Central America, and Libya, the medfly and other tephritid fruit flies in United States, Central and South America, South Africa, Europe, and Asia, the pink bollworm *Pectinophora gossypiella* Saunders in United States and codling moth *Cydia pomonella* L. in Canada (Alphey *et al.*, 2010). However, this strategy is expensive and depends on large biofactories.

In the last decade, new generation sequencing technologies have contributed to provide an ever-expanding volume of high quality genomic and transcriptomic data. Insect genome projects, together with sophisticated computer-modeling approaches, insect transgenesis (Catteruccia *et al.*, 2000; Coates *et al.*, 1998; Jasinskiene *et al.*, 1998), arthropod functional genomics (Dimopoulos *et al.*, 2002) and intensified studies of vector populations provided novel insights into the physiology, behavior and evolution of disease vectors, devising new methods for their control.

In 2002, the publication of the *An. gambiae* genome represented a turning point in vector genomics. Afterwards, similar efforts have been made for *Culex pipiens* vector of West Nile virus, the yellow fever mosquito *Aedes aegypti*, the Asian tiger mosquito *Aedes albopictus*, the Lyme disease tick *Ixodes scapularis* and the tsetse fly *Glossina morsitans* vector of African trypanosomiasis (Yin *et al.*, 2016). Thanks to genome initiatives, over the past 10 years more than 100 insect genomes have been released.

During my PhD, I worked in the framework of three genome sequencing projects:

- i5k: the 5000 insect genome project, among which the genome of the medfly, *Ceratitis capitata*, was sequenced and analyzed (Consortium, 2013);

- INFRAVEC: whose objective was to provide high-throughput sequencing of different mosquitoes species, including the Italian *Fellini* strain of *Ae. albopictus* (http://www.infravec.eu);
- IGGI (International *Glossina* Genome Initiative): focused on the sequencing of six *Glossina* species.

I performed genome-based analyses to provide an annotated framework for functional genes involved in the reproductive biology and sexual development of these insects. For development and reproduction to occur, insects need to be able to respond to relevant chemical stimuli, at the right time. These include behaviours related to mate finding, mating and progeny production. I thus focused my work on unravelling the functional role and evolution of genes related to reproduction and chemosensory perception.

1.3 Insect reproduction

Together with behaviors associated with survival, much of an insect life is dedicated to reproduction (Klowden, 2008). In nature insect makes adaptation due to selective environmental factors acting in a population, the so-called *adaptive strategy* (Horn, 1976). This concept is theorized in an evolutionary ecology paradigm: the r- and k-selection theory. The theory was developed in the late 1960s by the ecologists Robert MacArthur and E. O. Wilson and relates to the selection, in an organism, of a combination of traits which stabilize the equilibrium between the quantity and quality of the offspring (Pianka, 1970). r-strategist species live in unstable environments which select traits that enhance population growth rate, including early maturity, small body size, high reproductive effort and high fecundity. On the contrary, k-strategists tend to maintain their populations at a constant level, close to the carrying capacity of environment (Atkins, 1980). They are selected for traits enhancing high survival ability, including large body size, high investment in individual maintenance at the cost of low reproductive effort, low fecundity and longer life span (Reznick *et al.*, 2002).

Locust, aphids, mosquitoes and fruit flies (i.e. the medfly *C. capitata*) are example for r-strategist insect, while tsetse flies and carpenter bees are example for k-selected insects (Matthews and Kitching, 1984). Knowledge about the reproductive strategies used by insects of medical and agricultural importance is fundamental for controlling population size in the field. In order to be effective, biological control and other control strategies need to account for the reproductive characteristics of the species under study. Indeed, for r-strategist species, like most mosquitoes and invasive pests, a very high proportion of individuals have to be killed to cause some impact on the target population (Malacrida *et al.*, 2007; Regis *et al.*, 2008).

Contrarily to r-strategist insects, k-strategist such as tsetse flies are more suitable for eradication through such control methods. Compared to mosquitoes and the medfly, they show a low reproduction rate and, moreover, being adapted to stable habitats with low

level of cross-breeding, they show a reduced genetic variability and a limited capacity to respond quickly to selective pressures imposed by various control interventions (Dujardin and Schofield, 2004).
The differences at the genomic level between k- and r-strategist species can now be approached and integrated with data relative to the physiological adaptations displayed by female insects. A very interesting case is represented by the tsetse fly: females in these species produce a single offspring at each reproductive cycle through adenotrophic viviparity (development and feeding of intrauterine offspring). To achieve this, tsetse females have undergone extreme adaptations, including ovarian follicle reduction, uterine expansion and development of accessory glands (milk glands) to accommodate and feed a developing larva that can increase over 100-fold in dry mass over a week (Benoit *et al.*, 2015).
Whether (and to which extent) this is reflected in different reproductive adaptations also from the male side is still unclear. The reproductive strategy of male insects shapes their fitness by means of the production and management of semen, ability to find partners for mating, capacity to outcompete rival males, female choice, and investment in progeny (Wedell *et al.*, 2002). Males normally invest little in individual sperm but, instead, increase reproductive fitness by maximizing both the number of spermatozoa produced, the composition of the seminal fluid transferred to the female together with sperm, and the number of mating performed (Parker, 1970). In addition, male insects have developed multiple strategies to ensure successful transfer of sperm, including transfer of free sperm, spermatophores and production of mating plugs, to regulate sperm storage and release in the female (Scolari *et al.*, 2016).
To begin disentangling this biological question at the molecular level, I focused on the analyses of genes encoding seminal fluid proteins, both in an example of r-strategist species (the medfly) and of a k-stretegist (tsetse flies). Insect seminal fluid proteins (SFPs) are indeed important for fertilization and have been shown to play a key role in manipulating post-mating physiological and behavioral changes in the female (Laflamme and Wolfner, 2013). Seminal fluid proteins are diverse and perform a number of different functions. Under natural selection and under female selection, SFPs genes are under both intra- and inter-sexual selection, due to male-male competition and female sexual selection, respectively (Andres *et al.*, 2006; Poiani, 2006).

1.4 Insect olfaction

All living organisms need to detect chemicals in their environment to survive. Many aspects of the life of an insect are influenced by olfaction, which is indeed crucial in nutrition, mating, courtship behavior, aggregation, alarming, aggression and prey-predator interactions (Hansson and Stensmyr, 2011). In mosquitoes, for example, it enables females to find suitable hosts for a blood meal. Chemicals involved in insect

communication are referred to as semiochemicals and odorants. Insects detect semiochemicals with specialized peripheral sensory structures, the olfactory sensilla, present on different chemosensory tissues such as antennae, maxillary palps, and proboscises (Leal, 2012). The signal transduction cascade which derive from the detection involves three main groups of molecules: i) odorant-binding proteins (OBPs), ii) chemosensory proteins (CSPs) , and iii) the chemoreceptor superfamily formed by the olfactory (OR), gustatory (GR) and ionotropic (IR) receptor families (Leal, 2013; Pelosi *et al.*, 2006; Vogt and Riddiford, 1981). OBPs transport the odor molecules from the aqueous lumen of a sensillum to transmembrane olfactory receptor proteins (ORs) (Leal, 2013; Vogt, 1988). The number of genes encoding putative OBPs is very variable across different insect orders (Gong *et al.*, 2009; Manoharan *et al.*, 2013). Two hypothesis have been proposed regarding OBPs mode of action:

- OBPs act as passive carriers and ORs are activated by the ligand per se, as observed in moths and mosquitoes (Damberger *et al.*, 2007);
- The formation of a specific OBP-ligand complex is necessary for receptor activation (Ronderos and Smith, 2010).

In the membrane, ORs form a heteromeric complex with a highly conserved non-conventional OR, the odorant receptor co-receptor (Orco), which is required for the functional integration of receptor proteins in the dendritic part of the olfactory receptor neurons (OSNs) within the sensillum shaft (Larsson *et al.*, 2004; Neuhaus *et al.*, 2005). Once the odor molecule is no longer required, it needs to be degraded in order to allow the olfactory system to remain responsive to other odors. The enzymes responsible for the degradation of the odors are odorant degrading enzymes (ODEs). Different classes of ODEs are specific to different volatile compounds and localize within the sensillum lumen and within the cells lining the base of the sensillum **(fig.1.1)**.

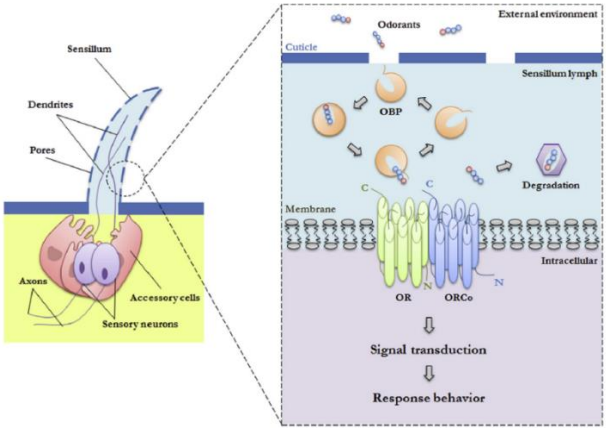


Fig 1.1. Schematic view of the odorant perception process in insects (Brito *et al.*, 2016).

OBPs have been recently reported to be expressed in male accessory glands and testes of diverse insects, such as *Drosophila*, *Aedes aegypti* and *Tribolium castaneum* (Chapman, 2008; Sirot *et al.*, 2008; South *et al.*, 2011). These proteins may be also involved in function other than olfaction, acting as carriers for hormones or ligands that are transferred from the male to the female during copulation (Arya *et al.*, 2010).
Therefore, the results of our studies are particularly important not only to enable comparative analyses among related species, but also to develop novel environmentally-friendly attractants and repellents for the control of wild populations.

Chapter 2: Comparative genomic analysis of *Glossina* species: insights into male contributions to reproduction

2.1 Background

2.1.1 Trypanosomiasis: the role of tsetse flies

Ranking as fifth on the world's deadliest animals list, tsetse flies are vectors of pathogenic trypanosomes responsible for sleeping sickness in humans (HAT, Human African Trypanosomiasis which causes 10,000 deaths per year) and nagana in livestock and wild animals (AAT, Animal African Trypanosomiasis). These parasites belong to the genus *Trypanosoma*. There are two forms of HAT. One, mostly defined by chronic infection, is caused by *Trypanosoma brucei gambiense* and accounts for more than 98% of reported cases, affecting 24 countries in West and Central Africa. The other form, characterized by acute infection, is caused by *Trypanosoma brucei rhodesiense*, it represents less than the 2% of reported cases and affects 13 countries in Eastern and Southern Africa (WHO, 2013).
Between 20,000 and 30,000 people are thought to be infected with sleeping sickness every year. Diagnosis and treatment are complex and most of the affected populations have no access to adequate therapy. Moreover, the clinical signs and symptoms are unspecified, frequently leading to misdiagnosis. In addition to public health issues, also the economy of the country suffers from the consequences of the AAT, with economic losses in cattle production estimated at US$4.75 billion per year (Budd, 1999).

2.1.2 The genus Glossina: overview

Tsetse flies are confined to sub-Saharan Africa **(fig 2.2).** They require a temperature of 16 to 38°C and 50-80% relative humidity for their survival and the lifespan is variable depending on the season: it ranges from one to two months in the dry seasons to 3-5 months in the rainy seasons. Both sexes are obligate hematophagous and they require a blood meal every 2-4 days, depending on host availability (WHO, 2013).
Tsetse species belong to the genus *Glossina*, which is the sole member of the family Glossinidae. The Glossinidae are placed within the superfamily Hippoboscoidea, together with louse flies and two families of bat flies, Hippoboscidae, Streblidae and Nycteribiidae, respectively (Petersen *et al.*, 2007). Nowadays, 22 tsetse fly species are present in Africa. The genus *Glossina* can be generally divided into three groups of species based on a combination of geographical distribution as well as behavioral, molecular and morphological features. The genus includes the Savannah flies (subgenus Morsitans), the forest flies (subgenus Fusca) and the riverine flies (subgenus Palpalis) (Krafsur, 2009). **Table 2.1** and **fig. 2.2** show key features of the six species target of my study.

Species of the Morsitans group are largely savanna and woodland inhabitants, although *G. pallidipes* may also be found in forests. The Morsitans group is adapted to drier habitats than the other two subgenera. Species of the Palpalis group tend to occur in riverine and lacustrine habitats. Fusca group flies largely inhabit moist forests of West Africa although one of these species, *G. brevipalpis*, occurs discontinuously in East Africa, Zaire, and Mozambique **(Fig.2.2)**.

The host-specificity varies among the three groups, with the Palpalis flies displaying strong antrophilicity and morsitans and fusca showing more zoophilic preferences (Aksoy and IGGI, 2010; Barrett *et al.*, 2003) **(Table 2.1, Fig. 2.2)**.

With the exception of *G. p. palpalis* in Bonon and Cote d'Ivoire, populations within each *Glossina* species complex mostly occupy non-overlapping geographic areas (allopatric population). This means that hybrid males are sterile and the hybrid females vary in their fertility (Krafsur, 2009).

Table 2.1. Biological and ecological traits of the six *Glossina* species analyzed. Species-specific information on habitat, host preference and vector status. (Clausen *et al.*, 1998; Motloang *et al.*, 2012; Munks *et al.*, 2005).

Species	Group	Habitat	Vector Status
G. fuscipes	*palpalis*	Lowland rainforest	Human African trypanosomiasis
G. palpalis	*palpalis*	Lowland rainforest	Human African trypanosomiasis
G. austeni	*morsitans*	Savanna woodlands	Nagana
G. pallidipes	*morsitans*	Savanna woodlands	Nagana
G. morsitans	*morsitans*	Savanna woodlands	Nagana
G. brevipalpis	*fusca*	Forest islands	*Trypanosoma brucei* resistant

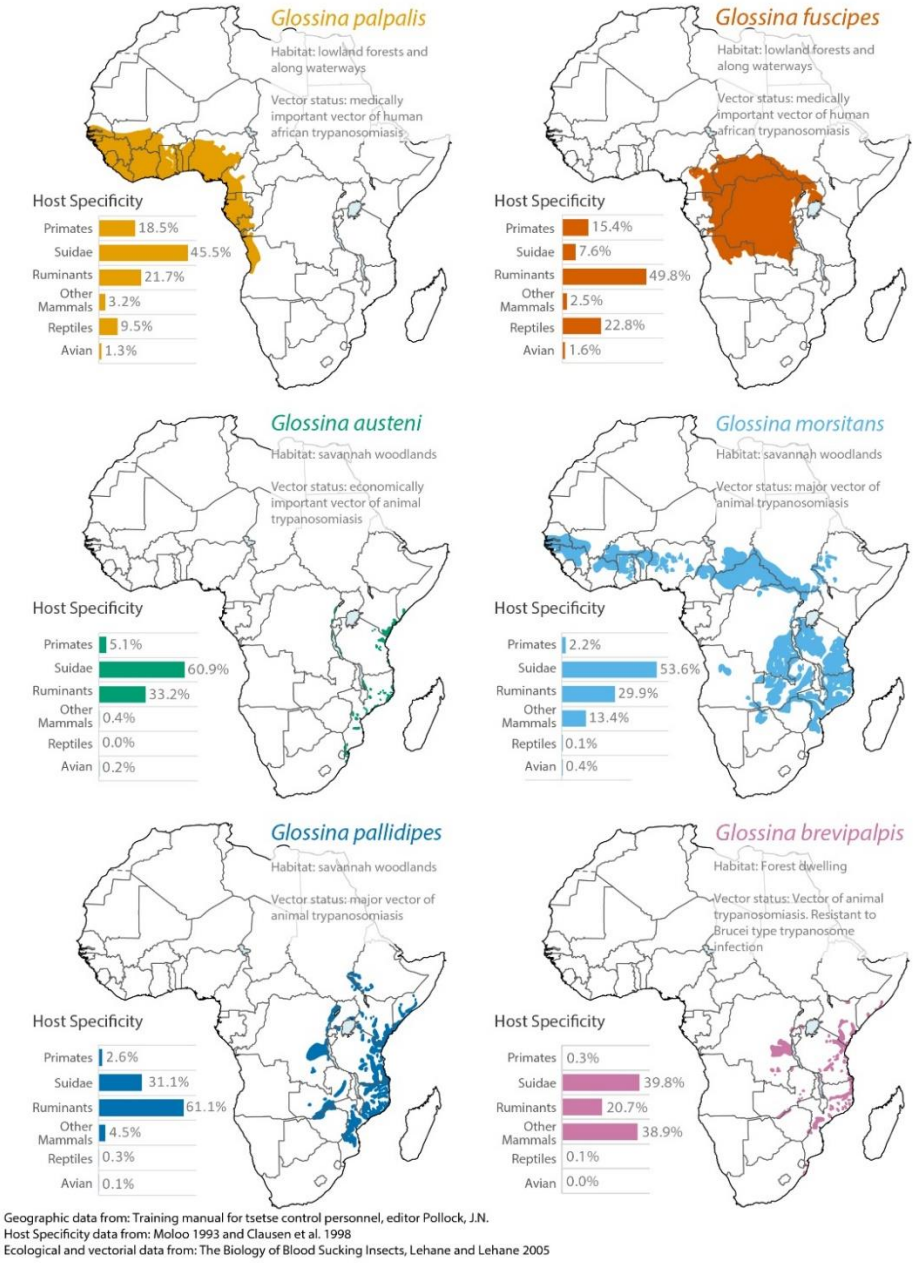


Fig. 2.2 *Glossina* geographic distribution, host preference and ecology (Attardo *et al*, Genome paper, in preparation).

2.1.3 The parasite and the disease: Trypanosoma brucei and sleeping sickness

Sleeping sickness clinically evolves in two stages: a hemo-lymphatic stage (early stage) in which the parasite dwell in the lymphatic system and bloodstream and a meningo-encephalitic stage (late stage), with the parasite crossing the blood-brain-barrier (BBB) and invading the central nervous system.
Intermittent fever, headache, dermatologic problems, cardiac disorders, endocrine disturbances are the main symptoms of the first stage, while neuropsychiatric signs are typical of the second stage, which gradually lead to coma, severe organ failure and death (Brun *et al.*, 2010).
The parasite responsible for the disease is a protozoan hemoflagellate belonging to the genus *Trypanosoma*. Two subspecies of *Trypansoma brucei* are responsible for the two form of sleeping sickness disease which occur in different African regions. The Rift Valley geographically separates the two forms of the parasites, with *T. b. Rhodesiense* present in the east of the Valley and *T. b. gambiense* found to the west (**fig 2.3**). Both sexes of adult *Glossina* flies feed on blood and can transmit the parasite to a mammalian host, which is necessary for the completion of the parasite's life cycle.

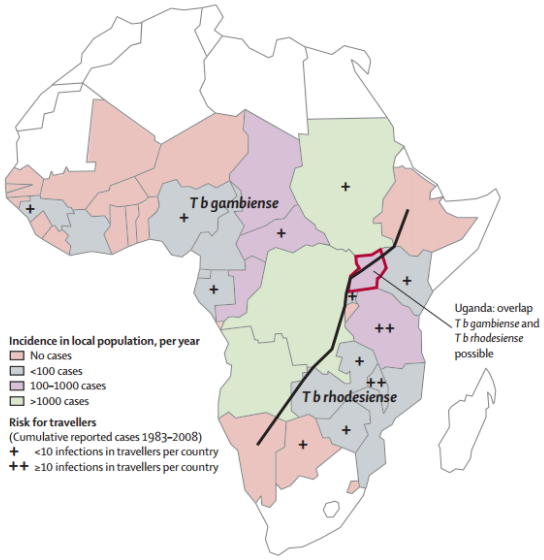


Fig. 2.3 Distribution of human African trypanosomiasis. The black line divides the areas of prevalence for *Trypanosoma brucei gambiense* (left) and for *Trypanosoma brucei rhodesiense* (right) (Brun *et al.*, 2010).

The parasite undergoes different biological stages both in the insect vector and in the mammalian reservoir:
a) bloodstream trypomastigote form: this is the form ingested by the fly; it moves to the midgut of tsetse, where the infection process starts. Some *Glossina* species are refractory to certain trypanosome species and this is the stage where the infection can still be blocked;
b) procyclic form: trypomastigote forms can replicate and differentiate into procyclic forms that cross the peritrophic membrane and reach the proventriculus, a small muscular organ where the blood meal flows forward to the crop, where the blood is stored, and backward from the crop to the midgut;
c) mesocyclic trypomastigote and epimastigote forms: these forms develop in the proventriculus and then migrate to the salivary gland, where they multiply and become infectious;
d) metacyclic form: this is the only infective stage to vertebrates. It is transmitted to the mammalian host during the blood meal and, once in the bloodstream, it develops into a trypomastigote that have access to different body fluids, including lymph, cerebrospinal fluid and the placenta.
The tsetse fly competence for trypanosomiasis is influenced by several factors: density of the tsetse population, tsetse fly longevity, vector's susceptibility to infection, host symbionts (Roditi and Lehane, 2008).
Less than 1% of flies is infective for *T. brucei* spp. and only 2-5% of infected flies present the infective metacyclic form in the salivary glands (Molyneux, 1980). This makes the tsetse fly a low competent vector of African Trypanosomiasis, however, it is also true that a low number of parasites in the blood meal are sufficient to infect a tsetse fly and a single infected bite is enough to transmit the parasite to a mammalian host. It is also common to observe tsetse flies co-infected by more than one trypanosome species (Lehane *et al.*, 2000). This finding explain the possibility of genetic recombination between different trypanosome strains within tsetse salivary glands (Gibson and Stevens, 1999), which is also supported by experimental evidences showing that tsetse salivary glands can host two trypanosome strains (Peacock *et al.*, 2007).

2.1.4 Tsetse microbiota: impact on host vector competence

Symbiotic microbes of many insect vectors play a fundamental role in the transmission of pathogens, influencing host vector competence, fecundity and nutrition. In *Glossina* laboratory strains, four vertically transmitted microbes have been found to establish a symbiotic association with the fly (Wang *et al.*, 2013) (**Fig. 2.4**):
- the obligate *Wigglesworthia glossinidia,*
- the commensal *Sodalis glossinidius,*
- the parasitic *Wolbachia,*

- the Salivary Gland Hypertrophy Virus (SGHV)

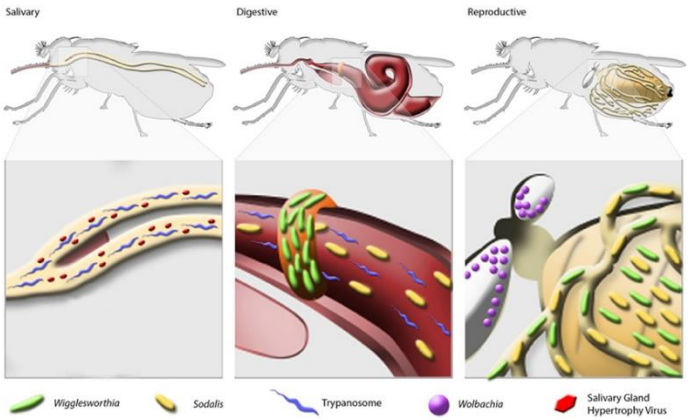


Fig. 2.4 Localization of symbionts and SGHV in tsetse (Wang *et al.*, 2013).

The obligate mutualistic *Wigglesworthia* is found both intracellularly, in midgut-associated bacteriome cells, and extracellularly, in the lumen of milk glands (i.e. a series of tubules that ramify throughout the female abdomen and are responsible for the transfer of nutrients and symbionts from the mother to the intrauterine larvae). Its presence influences several physiological mechanisms of tsetse fly. Firstly, it provides several nutrients such as key B-complex vitamins, thiamine monophosphate, lipoic acid, folate and many other metabolites that are absent, or present at low titers, in vertebrate blood and that are crucial for the development of the intrauterine larva. Indeed, several studies demonstrated that the absence of this bacterium causes impairment in the intrauterine larval development, leading to abortion (Nogge, 1978; Nogge and Gerresheim, 1982; Pais *et al.*, 2008; Schlein, 1977). The absence of *Wigglesworthia* affects also the fly's immune system, with *Wigglesworthia*-free flies presenting severely compromised immune system during adulthood (Weiss *et al.*, 2012; Weiss *et al.*, 2011; Weiss *et al.*, 2013).
The commensal *Sodalis* is a gram-negative organism closely related to free-living microbes within the Enterobacteriacaea. It is found in various tissues including midgut, fat body, milk glands, salivary glands and hemocoel. No functional roles have been defined yet for *Sodalis* within the tsetse host and it is present only in some natural tsetse populations. However, few studies reported its possible involvement in favoring the establishment of the trypanosome infection (Farikou *et al.*, 2010; Soumana *et al.*, 2013).
Wolbachia is a widespread alpha-proteobacteria endosymbiont. It infects approximately 70% of insect species (Hilgenboecker *et al.*, 2008) and it is also found in some tsetse

populations. It is widely studied in the frame of vector control strategies because of its ability in manipulating the reproductive biology of its host through cytoplasmic incompatibility (where zygotes formed from crosses between uninfected females and infected males do not develop), male killing, feminization and parthenogenesis (Werren *et al.*, 2008).
The SGH virus is also of particular interest because it causes hypertrophy of the salivary glands, gonadal lesions and reduced fecundity and lifespan (Abd-Alla *et al.*, 2011) in the flies. It occurs both in colony-reared and natural populations of tsetse and it can be transmitted either vertically, through the maternal milk gland secretions, or horizontally during the feeding (Abd-Alla *et al.*, 2011; Wang *et al.*, 2013).

2.1.5 Vector control strategies against Glossina species

Vector-borne infections have largely affected human activities. Despite decades of research, effective vaccines against African trypanosomes have yet to be developed. Current vector control interventions involve the use of traps or screens, insecticides and the sterile insect technique (SIT) (http://www.who.int/tdr/research/vectors/methods _strategies/en/). Although extensive insecticide ground spraying was used to control tsetse in many African countries, Zimbabwe and Nigeria above others, insecticides are rarely used today because of health and economic issues. Odor baited traps and screen impregnated with insecticide effectively suppressed tsetse population by 99% in many countries. However, effective baits have been developed for tsetse living in Savannah (Morsitans group) but not for riverine tsetse (Palpalis group), major vectors of HAT. Knowledge of the molecular mechanisms underlying olfaction may be crucial to develop more potent attractants and repellents.
The SIT is the other main approach used to reduce tsetse populations. The SIT consists in the massive release in the environment of sterilized adult males, which compete with wild males for mating with wild females. The technique has already been effectively used for the eradication of tsetse flies (*G. austeni*) from Unguja Island, Zanzibar (Vreysen *et al.*, 2000). However, in order to avoid re-invasion of the treated area, ecological islands are needed for the success of the technique. Moreover, it requires expensive procedures for the rearing of a large number of male flies, sterilization, and their massive release in the field.

Reproductive physiology of *Glossina* species

2.1.6 Female reproductive system: the adenotrophic viviparity

Knowledge of the reproductive physiology of *Glossina* species is essential for the improvement and development of population control strategies.
One of the most remarkable aspects of the tsetse biology is the adenotrophic viviparity, i.e. the ability to nourish intrauterine offspring from glandular secretions and to give birth to fully developed larvae. This peculiar reproductive strategy yields very few progeny and, as such, provide excellent target to develop population control strategies.

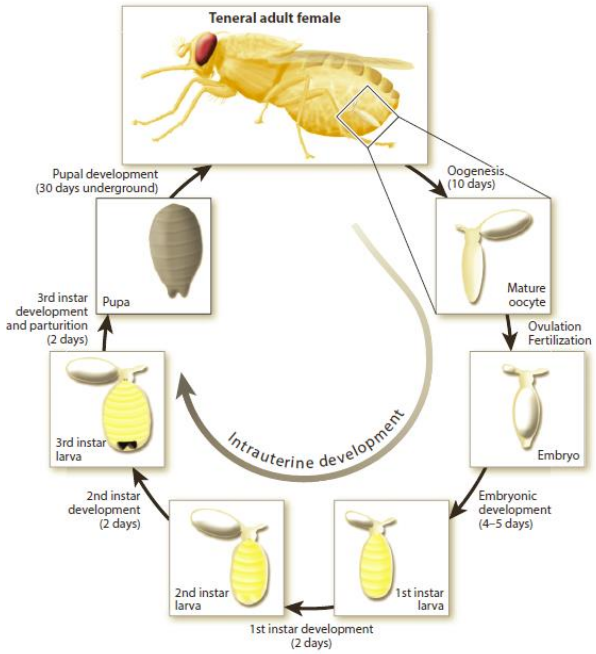


Fig. 2.5 Schematic of the first gonotrophic cycle of a *G. morsitans* female under optimal environmental and nutritional conditions. The different stages of oogenesis, embryogenesis, and larvigenesis within the *Glossina* reproductive tract (ovaries and uterus) are shown (Benoit *et al.*, 2015).

Adenotrophic viviparity is shared by all members of *Hippoboscoidea*, present also in *Mesembrinellinae* and in one species of the *Sarcophagidae* family (Benoit *et al.*, 2015). Mating usually occurs within 3-5 days of emergence of the female and, in the field, it probably occurs at the first blood meal (Tobe and Langley, 1978). After the mating, spermatozoids are stored in the female spermathecae, where they can survive for about 200 days. One mating is sufficient for female to remain fertile, nonetheless, remating is

not unusual. The number of times a female mates in the wild influences the effective population size and may constitute a critical factor in determining the success of control methods (Bonomi *et al.*, 2011). The first ovulation occurs ~10 days after adult emergence, followed by intrauterine embryogenesis and larvigenesis **(Fig. 2.5)**. Embryogenesis takes 3–4 days and is followed by 5–6 days of larval development (Saunders and Dodd, 1972; Tobe and Langley, 1978). The female produces a single offspring at regular intervals, giving birth to the first progeny ~20 days posteclosion. In order to provide all the nutrients to the larva, a tsetse female undergoes large changes in volume and weight during each pregnancy cycle. After parturition, the larva pupates within 1-2 hours, and approximately 30 days later, the adult fly emerges.
This reproductive system is supported by peculiar features of reproductive organs, such as a reduced number of ovaries (with two ovarioles per ovary), an expansion of the milk glands and a greatly expanded uterus (Pimley and Langley, 1982), together with two spermathecae (yellow/orange spherical bodies) which act as sperm-storage organs (**Fig. 2.5)**.

2.1.7 Male reproductive system: Testes, Accessory glands and Seminal Fluid Proteins (SFPs)

Tsetse male reproductive organs are similar to those of other *cyclorrhaphous* (Diptera), consisting in a pair of testes and accessory glands, which both empty into the ejaculatory duct (Itard, 1970). Males of the *Glossina* genus transfer sperm to females through the spermatophore (Pollock, 1970), which is a capsule-like structure assembled within the uterus of the female with Male Accessory Glands (MAGs) secretions during the later stages of copulation (Odhiambo, 1983; Scolari *et al.*, 2016). The spermatophore functions as a protective container for the ejaculate, ensuring spermatozoa to reach the female spermathecae (Scolari *et al.*, 2016) **(Fig. 2.6)**.
Insect seminal fluid proteins (SFPs) have been studied as a model for seminal protein function in many animals, including mammals (McGraw *et al.*, 2015). Sperm storage, female post-mating response and sperm competition are only some of the many roles that have been described for SFPs in female insects after mating (Avila *et al.*, 2011). The seminal fluid comprises both MAG products and testes-specific proteins, including sperm structural components and secreted proteins.
Components of the ejaculate are associated with sex-specific costs and benefits. Moreover, several studies demonstrated that testis-specific genes, SFPs and spermatogenesis genes show lineage-specific bursts of accelerated evolution and positive selection playing a role in shaping the plasticity of the species (Haerty *et al.*, 2007; Perry *et al.*, 2013).

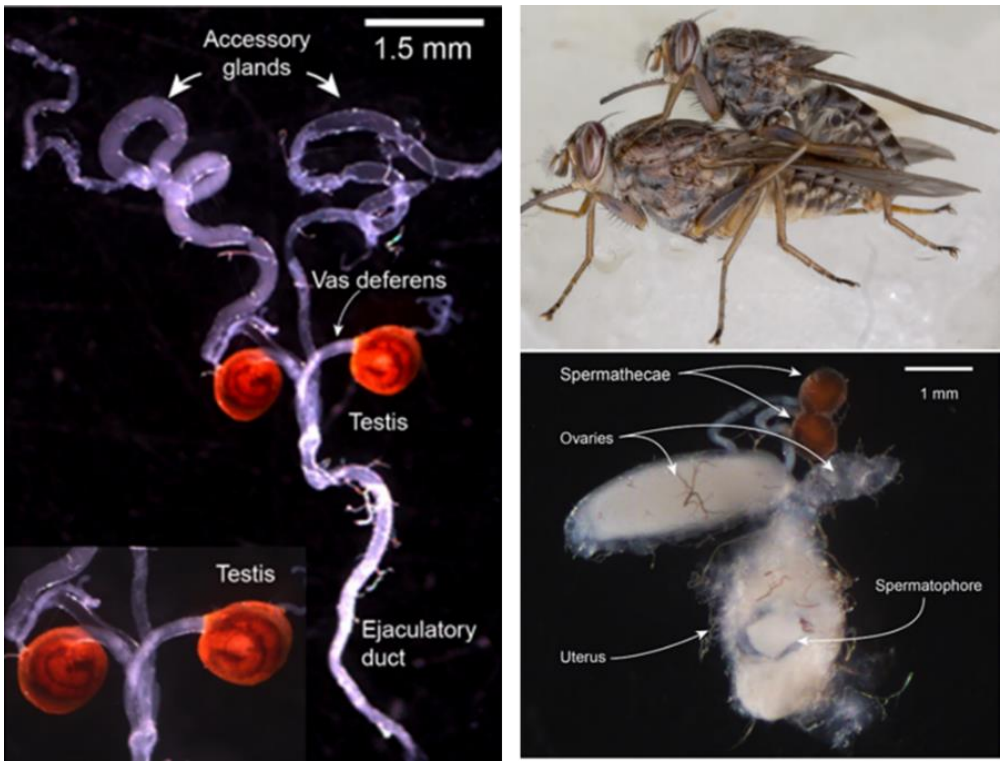


Fig. 2.6 From left to right: male reproductive tract; mating pair, and female reproductive tract dissected immediately after mating: a spermatophore is evident in the uterus. Scale bar = 1 mm (Scolari *et al.*, 2016).

2.2 Aim of the work

Despite the extensive understanding of trypanosome biology acquired over the last 25 years, the toolbox for disease control is still limited, with neither vaccines nor effective and affordable drugs available in the near future (Aksoy and IGGI, 2010). Eradication of trypanosomiasis via tsetse populations control has a great potential and novel genomic resources are needed to improve currently available tools and develop new effective strategies.

Started in 2010, the *Glossina* genome projects aimed at producing and releasing the genome sequence of five *Glossina* species and of two related Dipterans, the stable fly *Stomoxys calcitrans* and the house fly *Musca domestica*.

The completion of the release of the genomic resources in 2015 has facilitated comparative genomic analyses between *G. m. morsitans* (Attardo *et al.*, 2014) and the twelve *Drosophila* genomes already available, providing researchers with additional information on vector competence, haematophagy, viviparity, as well as on the identity of species-specific genes involved in host seeking and vectorial capacity.

Evolutionary and comparative genomics analyses provide powerful tools to study molecular changes among organisms, supporting the identification of both conserved and species-specific genes and the evaluation of selective pressures acting on them. On this background, investigating the evolution of reproduction-related genes can, in the long term, provide important information about the process of speciation and species divergence.

Studies focused at identifying the presence of adaptive evolution in proteins of the reproductive system have never been performed in tsetse species.

The aim of my project was thus characterizing the presence of selective pressure in proteins encoded by genes expressed in the reproductive tissues of tsetse species.

The following section will be divided in two parts:

a) Evolutionary and comparative genomics analyses of Glossina species

a1) Phylogeny of the *Glossina* genus and molecular clock

To uncover the genes involved in the development of species-specific traits, the genome of six species have been analysed: *G. m. morsitans, G. pallidipes, G. austeni, G. palpalis, G. fuscipes* and *G. brevipalpis.* We first reconstructed the molecular phylogeny of these six species using a genome-wide set of orthologous genes. Moreover, the same dataset has been used to infer the time of divergence among the species. The molecular clock was used to date the first evolutionary split among the six sequenced *Glossina* species at about 35 million years ago (Mya), consistent with the *Glossina* fossil records (Cockerell, 1907).

a2) Rate of molecular evolution of male reproductive genes across *Glossina* species

We selected 2,563 male-biased genes, expressed in male accessory glands (MAGs) and testes of *G. m. morsitans*. Orthology was assessed in the other 5 Glossina species and the sequences were tested for the occurrence of positive selection using codon substitution models in PAML. Accessory gland proteins are those with the highest d_N/d_S ratios, indicating relaxed selection and/or pervasive positive selection; in contrast, genes expressed in testes are subjected to evolutionary pressures similar to those acting on genes with no male-biased expression.

b) Identification and functional analysis of elements involved in the regulation of male reproductive genes in the reference species G. m. morsitans: the Paired transcription factor

After the identification of MAG and testes genes contributing to male reproduction in the six *Glossina* species, my interest was addressed in exploring the mechanisms involved in their regulation. In particular, the attention has been focused on the reference genome *G. m. morsitans* and on sequences highly expressed in accessory glands, which display faster evolutionary rates compared to testes genes.
In order to identify MAG-specific transcription factors, the promoter region of the 24 most highly expressed accessory glands genes has been analyzed. The analyses led to the identification of the transcription factor Paired, highly transcribed in the male accessory glands and whose transcription-factor binding site has been found in the regulatory region of most of the candidate genes. Tissue-specific expression profiles of the transcription factor Paired have been determined and RNAi experiments have been performed to derive information on its putative activity in regulating the expression of MAG genes.

2.3 Evolutionary and comparative genomics analyses of *Glossina* species

2.3.1 Materials and Methods

Molecular phylogeny analyses

Data from the *G. m. morsitans* genome were used to conduct a comprehensive multi-locus phylogenetic analysis in the context of genome data from 5 additional *Glossina* species (*G. pallidipes, G. austeni*, *G. fuscipes*, *G. palpalis, G. brevipalpis*). The genomes of 3 fruit fly species (*D. melanogaster, D. ananassae, D. grimshawi*), the house fly *Musca domestica*, the sand fly *Lutzomyia longipalpis* and the mosquito *Anopheles gambiae* were also included to complement and facilitate comparative analyses.
Bayesian and maximum likelihood (ML) phylogenetic analyses have been performed. Maximum likelihood tree was constructed from a Raxml analyses using LG+G+F model and a concatenated alignment of 117783 amino acids from 286 orthologs genes in all the species involved in the analysis (full dataset after Gblock). Posterior probabilities were calculated from a Bayesian analysis of the same dataset using heterogeneous model of evolution CAT+G and bootstrap supports from an Astral analysis based on all single genes trees analyzed at the nucleotide level under GTR+G model.

Molecular Clock Analyses

Divergence times were calculated for the genomic dataset previously described using PhyloBayes (Lartillot *et al.*, 2009), employing the optimal tree inferred for each data set under the best fitting substitution model (see above), and the best-fitting relaxed clock models, selected using Bayes Factor (calculated using thermodynamic integration) in Phylobayes. 3 node calibrations have been used, setting the time of divergence between *D. melanogaster* and *An. gambiae* at 295.4-238.5 Mya; the split between *D. melanogaster and G. morsitans* at 64 Mya (as minimal value); and the split between *G. morsitans* and *G. brevipalpis* at 33 Mya (as minimal value).

Identification and annotation of male reproductive genes

To identify male genes with biased expression in testes and male accessory glands (MAGs) respectively, we used *G. m. morsitans* RNA-Seq libraries specifically derived from these tissues (Scolari *et al.*, 2016). Briefly, total RNA was separately extracted from testes and MAGs of teneral, 3 days old virgin and 6-8 h post-mating adult flies, to cover multiple physiological and developmental states. For each of the three time points, RNA isolation, Illumina library preparation and sequencing were performed as previously described (Benoit *et al.*, 2014). The Sequence Read Archive numbers for the

individual libraries at NCBI are SAMN04054396, SAMN04054395, SAMN04054394, SAMN04054393, SAMN04054392 and SAMN04054391 (Bioproject ID: PRJNA295435). Transcriptional levels were analyzed using CLC Genomics Workbench (CLC bio, Cambridge, MA). As measure of relative gene transcription, Fragments Per Kilobase of Exon per Million reads mapped (FPKM) statistic was used. For both testes and MAGs, an average transcription value over the three libraries was separately derived for each transcript.

To identify genes with significantly higher transcription in testes and MAGs relatively to the other tissue, the proportion of read counts for each sequence relative to total read counts was determined to calculate the P-value differences in proportions by Z-tests following FDR correction. Fold changes were determined as the ratio of FPKM of testes vs. MAGs. In all comparisons, we considered one gene to be tissue-biased when fold change in that tissue was at least 5 fold higher than in the other, as previously described in (Scolari *et al.*, 2016).

Functional annotation of male-biased proteins

Functional classification of testes and MAG-biased genes was performed using the Blast2GO software v.2.8 (https://www.blast2go.com/b2ghome). The CDS of these genes were used to perform BLASTx against the NCBI non-redundant (nr) database (e-value < 10^{-10}). For Gene Ontology mapping (GO; http://www.geneontology.org) we used Blast2GO to extract GO terms associated with homologies identified by NCBI's BLAST. We retained annotations with e-value < 10^{-10}. We then performed InterPro and InterProScan (Quevillon *et al.*, 2005) searches remotely from Blast2GO via the InterPro EBI web server and merged InterProScan GOs with the original GO annotations.

Orthologous gene set identification and sequence alignment

The list of testes- and MAG-biased genes we identified in *G. m. morsitans* was used to identify orthologous sequences in other five species: *G. m. morsitans*, *G. pallidipes*, *G. austeni*, *G. fuscipes*, *G. palpalis gambiensis* and *G. brevipalpis*. The respective coding sequences (CDSs) were retrieved from the BioMart database (Durinck *et al.*, 2005) in Vectorbase (Lawson *et al.*, 2009) (http://biomart.vectorbase.org/biomart/martview). Gene loci that showed incomplete domains and/or had incomplete sequences were manually curated using Web Apollo and CLC Main Workbench sequence annotation tools.

To assess the presence of different selective pressures across tsetse species we used the codon models implemented in PAML (Yang, 2007). We considered only those *G. m. morsitans* genes whose putative ortholog sequence was identified in at least four of the five other species. All orthologous sequences were aligned using Prank (Loytynoja and Goldman, 2008), as implemented in TranslatorX (Abascal *et al.*, 2010), which aligns

protein-coding nucleotide sequences based on their corresponding amino acid translations. To minimize the possibility of spurious matches, orthologous sets with sequences shorter than 50 amino acids were removed. We also used a custom perl script to remove problematic alignment regions using an approach similar to that proposed by (Han *et al.*, 2009; Ramasamy *et al.*, 2016).

Analyses of the rate of DNA and Protein Evolution

Rates of molecular evolution were analyzed using PAML 4.7 (Yang, 2007). The rate of nonsynonymous substitution, d_N (leading to amino acid changes), and synonymous substitution, d_S (which should accumulate neutrally), have been estimated over all branches of the phylogenetic tree using the "free-ratio" model. Following the results of the phylogenetic analysis, the unrooted tree used in all analyses had the structure (((*G. morsitans,G. pallidipes),G. austeni),(G. fuscipes,G. palpalis),G. brevipalpis*). Different models of substitution rates across coding sites were then tested (Yang and Nielsen, 2000) with the aim of detecting genes that either evolved at a different rate or underwent positive selection along one of the *Glossina* lineages.

In the first test, we compared models that assumed one or more substitution rates across the phylogeny. The first of such models is the basic "one-ratio" branch model (M0), which assumes a constant omega across the phylogeny (model=0 and NSsites=0). This model was also used to estimate the branch lengths for each gene tree, which were then copied into the tree structure file to be used with the "branch-site" substitution models. The likelihood of the M0 model was compared with that of a branch model that assumed two omega values, one for each of the *Glossina* species (the so-called foreground branch, with the exception of *G. brevipalpis*, since by using an unrooted tree we cannot separate the processes that acted along its lineage from those that took place in the lineage subtending the clade containing the other five species) and one for the rest of the tree (the background branches; model=2 and NSsites=0).

Subsequently, the value of twice the difference between the two likelihoods ($2\delta\lambda$) was tested using a chi-squared test with 1 degree of freedom. The occurrence of positive selection was tested by the branch–site test. In this test (branch–site model A, test 2 (Yang *et al.*, 2005), omega can vary both among sites in the protein and across branches on the tree (model=2, NSsites=2). As for the branch model, we used tree structures with branch lengths estimated by model M0. The null model fixed omega2=1 (fix_omega=1, omega=1), whereas the positive selection model allowed omega2>1 (fix_omega=0, omega=1). The likelihood ratio test had 1 degree of freedom. The occurrence of positive selection was also tested by comparing (nearly) neutral models to models that allow for the occurrence of positive selection (site tests). In a first approach we compared the likelihood of a model (M1a; model = 0 and NSsites = 1) that assumes two sets of sites with neutral ($\omega = 1$) or nearly neutral evolution ($0 < \omega < 1$), to a model with an additional class of sites with $\omega > 1$ (M2a; model = 0 and NSsites = 2). In a second more realistic

approach, we compared the likelihood of a model where ten site classes have ω values drawn from a β distribution (M7; model = 0 and NSsites = 7) to a model that incorporates an additional class of sites under positive selection (M8; model = 0 and NSsites = 8). In these cases, each comparison was tested using a χ^2 test with 2 degrees of freedom. To account for multiple testing, we also estimated the false discovery rate (FDR) of each test using the q value approach (Storey, 2002) implemented in R (R Development Core Team, 2009). We note that the reciprocal best-hit approach is prone to miss genes with high sequence divergence, including those that underwent particularly intense divergent adaptive evolution. Thus, we could have missed targets of positive selection among our sequenced genes.

2.3.2 Results and Discussion

a1) Phylogeny of the Glossina genus and molecular clock

Molecular phylogeny of Glossina species

Physiological, ecological, and behavioural differences, including trypanosome vectorial capacity, are well-known among *Glossina* species. In order to reach a reliable phylogenetic tree, we considered for the analysis 12 insect species for which genomic resources were already available. This comprehensive multi-locus phylogenetic analysis was performed on the six *Glossina* species and on six additional species: three *Drosophila* species (*D. melanogaster, D. ananassae, D. grimshawi*), the house fly *Musca domestica*, the sand fly *Lutzomyia longipalpis* and the mosquito *Anopheles gambiae*.
Whole genome analysis of the 12 species allowed us to recognize the presence of 286 orthologous genes. We traced the evolutionary history of a concatenated alignment of 117,783 amino acids from the 286 genes, using both Bayesian and ML (maximum likelihood) phylogenetic frameworks **(Fig. 2.7)**. The resulting tree supports the monophyletic origin both of the Morsitans group (*G. morsitans, G. pallidipes* and *G. austeni*) and the Palpalis group (*G. fuscipes* and *G. palpalis*), as previously suggested (Dyer *et al.*, 2008). Moreover, the phylogenetic tree clarify the *G. austeni* relation with the Morsitans group. It is noteworthy that the position of *G. austeni* was previously controversial with respect both to the Palpalis and Morsitans groups (Dias J.A., 1987). Notably, the Fusca group (*G. brevipalpis*) can be considered as a sister group to other tsetse species (Hoppenheit *et al.*, 2013), forming the ancestral subgenera.

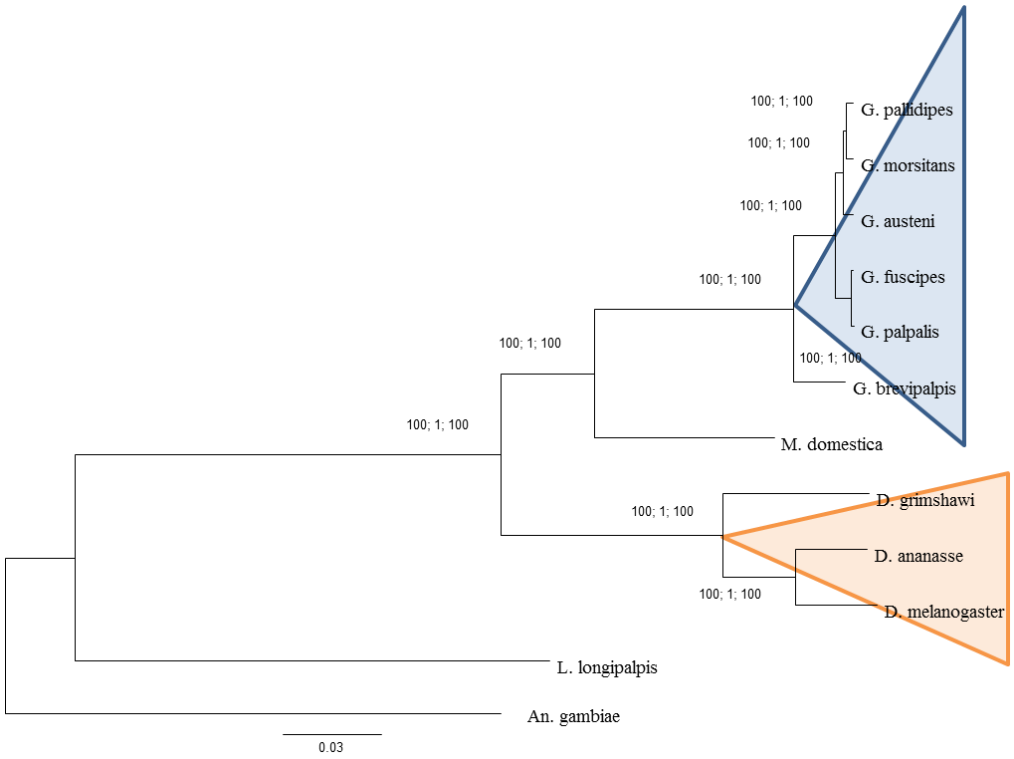


Fig. 2.7 Phylogenetic analyses based on 117,783 amino acid sequences from 286 orthologous genes detected in six *Glossina*, three *Drosophila*, *Musca domestica*, *Lutzomyia longipalpis* and *Anopheles gambiae*.

Glossina Molecular Clock

In order to deduce the time of divergence of the *Glossina* species, the phylogenetic tree was used to calibrate the molecular clock showed in **Fig. 2.8**.
The accuracy of a molecular clock completely rely on prior assumptions and, in particular, on calibrations (Battistuzzi *et al.*, 2010; Warnock *et al.*, 2012). Based on our assumptions, the divergence time for *G. brevipalpis* and the other *Glossina* species has been estimated at 54 Mya and at 21 Mya for the split between the Morsitans and Palpalis groups. The advantage of having genomic data coming from more than one species belonging to the same *Glossina* groups allows to clarify the controversial situation within the Morsitans group, dating the *G. austeni* species differentiation before the split between *G. morsitans* and *G. pallidipes* (*G. austeni*(*G. morsitans*, *G.pallidipes*)), at 12

Mya. According to our analysis, the more recent split is that between *G. palpalis* and *G. fuscipes*, estimated at 3.7 Mya (Palpalis group).
It is interesting to mention that these results are consistent with *Glossina* fossil records found in Colorado (USA), estimated as fossils of the Oligocene period (33.9 million to 23 million years before the present) (Cockerell, 1907; Junqueira *et al.*, 2016).
However, it is important to note that 22 *Glossina* species are living in Africa today and that our analysis is limited to those species for which genomic data are available.

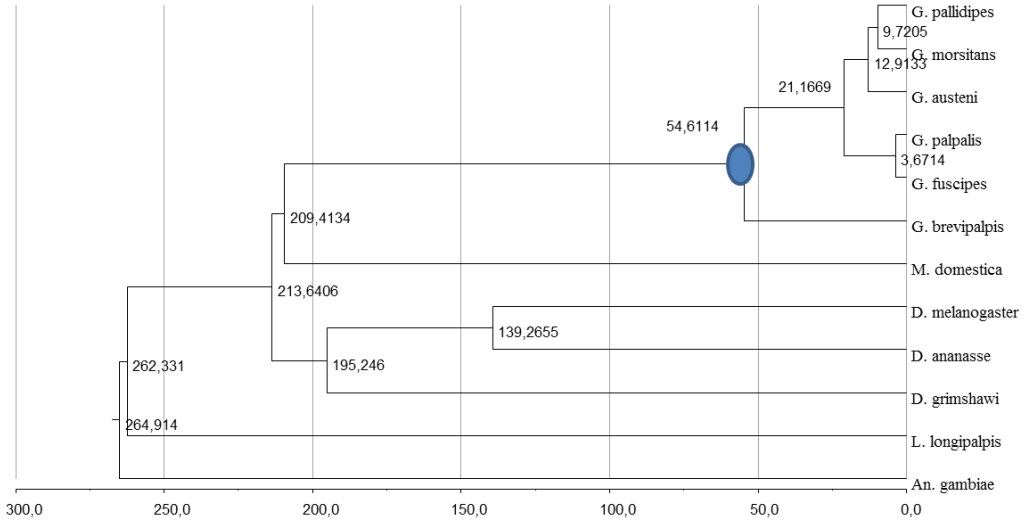


Fig. 2.8 Molecular clock for the species analyzed. On the x-axis from 300 million years ago (Mya) to present time. Time of divergence are shown at node points.

a2) Molecular evolution of male reproductive genes across *Glossina* species

Genes can be gained and lost over the course of evolution and have an important impact on the organism's fitness (Ranz and Parsch, 2012). Studies in *D. melanogaster* demonstrated that new genes can quickly become integrated into the genome and become essential for survival or fertility, increasing the genetic diversity (Chen *et al.*, 2010; Ding *et al.*, 2010). Specifically, numerous evidences suggest that traits associated to male reproduction may have evolved extremely rapidly (Coulthart and Singh, 1988; Eberhard, 1985; Wu *et al.*, 1996) as has been observed, for example, among male reproductive proteins between closely related *Drosophila* species (Nurminsky *et al.*, 1998; Ting *et al.*, 1998; Tsaur *et al.*, 1998). On this background, my interest has been directed toward the analysis of the molecular evolution of male reproductive genes in *Glossina* species for which the phylogeny has been clarified.

Functional annotation of male reproductive genes according to Gene Ontology

We focused our analysis on genes highly expressed in the two tissues responsible for the reproductive biology of the male, i.e. testes and MAGs. These tissues have been demonstrated, in *G. m. morsitans*, to participate to the ejaculate production transferred from the male to the female through the mean of a spermatophore (Scolari *et al.*, 2016). Using *G. m. morsitans* as reference genome, I identified in this species a total of 127 and 2436 genes which are highly expressed in MAGs and testes, respectively. We used a five-fold change as threshold to discriminate the expression rate of male reproductive genes in the two different tissues (i.e. when fold change in one tissue was at least five-fold higher than in the other). We considered for further comparative analyses with the other *Glossina* species only sequences whose orthology could be determined for at least five species. The comparative analysis resulted in the detection of 92 and 1924 orthologous sequences for MAGs and testes, respectively (**Table 2.9**).

Table 2.9 Number of RNA-seq reads and inferred testes- and MAGs sequences identified in *G. m. morsitans*. The number of orthologous genes identified in at least five species is also reported.

	G. m. morsitans			***6 Glossina***
Tissue	*Total RNA-seq reads*	*Reads after trimming*	*5-fold higher*	*orthology in at least 5 species*
MAG	128.565.696,67	126.765.703,33	127	92
Testes	131.973.344,00	129.097.917,67	2436	1924

The sharp difference in the number of genes differentially expressed in the two tissues can be explained by different functions associated to testes and MAGs.
To derive an initial picture of the potential functions of these tissues in the six *Glossina* species (*G. morsitans, G. pallidipes, G. austeni, G. fuscipes, G. palpalis, and G. brevipalpis*), we assigned different Gene Ontology (GO) functional classes to testes- and MAG-biased genes **(Fig. 2.10)**. Gene Ontology was assessed for both Molecular Function and Biological Processes. Binding and catalytic activity (GO Molecular Function), and metabolic and cellular processes (GO Biological Process) are the classes comprising the highest number of genes, in both testes and MAG datasets **(Fig. 2.10)**. The presence of such genes can be explained by the secretory activity and the rapid cell proliferation typical of these tissues.
A finer analysis of the members of each GO functional class revealed further interesting aspects. Indeed, among the testes genes, a deeper inspection of the GO term "reproduction" (GO Biological Process) allowed the identification of 11 sequences involved in different phases of the spermatogenesis process **(Table 2.11)**. For MAG genes, the GO analysis could not be as informative as for testes genes because of the smaller dataset and the limited availability of information. The lack of functional information about MAG genes is expected since genes expressed in the accessory glands display particularly rapid evolution and are thus very diverged from sequences of model species for which functional annotation has been verified. For example, in *D. melanogaster*, the evolutionary rate of many of these genes was so fast that they lack detectable orthologs even in other *Drosophila* species (Swanson *et al.*, 2001).

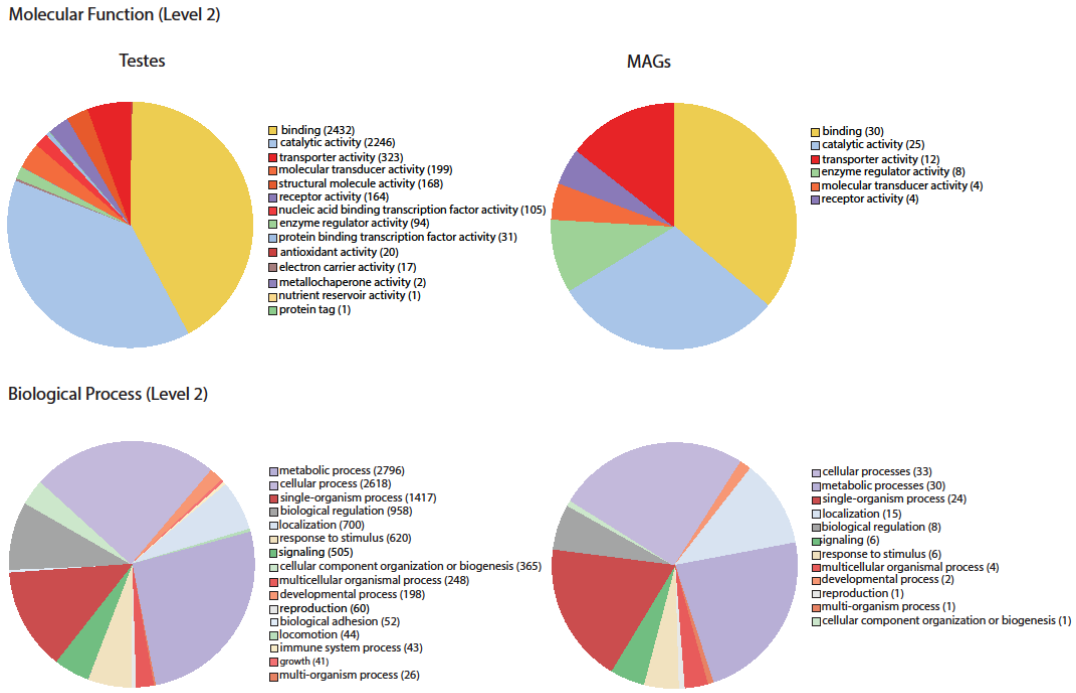


Fig. 2.10 Pie charts representing GO Molecular Function and Biological Process categories Level 3) for both testes and MAG genes.

Table 2.11 Eleven testes genes involved in different phases of the spermatogenesis process.

#Seqs	GO ID	Term	Sequences
10	GO:0007283	spermatogenesis	GMOY009988, GMOY004171, GMOY004199, GMOY007773, GMOY007085, GMOY009942, GMOY010205, GMOY003464, GMOY007703, GMOY007347
7	GO:0048515	spermatid differentiation	GMOY004199, GMOY007773, GMOY007085, GMOY009942, GMOY010205, GMOY003464, GMOY007347
6	GO:0007286	spermatid development	GMOY004199, GMOY007773, GMOY007085, GMOY009942, GMOY010205, GMOY003464
3	GO:0007291	sperm individualization	GMOY004199, GMOY007773, GMOY007085
2	GO:0048137	spermatocyte division	GMOY009988, GMOY004171
2	GO:0008215	spermine metabolic process	GMOY012095, GMOY012183
2	GO:0006597	spermine biosynthetic process	GMOY012095, GMOY012183
1	GO:0008295	spermidine biosynthetic process	GMOY012095
1	GO:0008216	spermidine metabolic process	GMOY012095

Selective pressure across testes and MAG genes

A gene can acquire new functions either because of changes in the protein sequence or through new patterns of expression (Ranz and Parsch, 2012). Because of the redundancy of the genetic code, point mutations which occur within a protein coding sequence can alter either only the nucleotide sequence or both the nucleotide and protein sequence, leading to synonymous or nonsynonymous substitutions (Ranz and Parsch, 2012).
Synonymous mutations do not affect the protein sequence and consequently the structure and function of the peptide. They are selectively neutral and accumulate over time in a clock-like fashion. By contrast, nonsynonymous mutations are subject to purifying and, at times, positive selection, and are a proxy for the rate of protein evolution. Based on this background, the ratio of nonsynonymous to synonymous divergence between sequences belonging to different species (d_N/d_S) can be used as a measure of level of selection experienced by that gene (Hurst, 2002; Ranz and Parsch, 2012). For instance, when in a gene d_N equals to d_S (i.e. $d_N/d_S=1$) we can hypothesize that neutral evolution is acting in the gene although we cannot exclude the combined action of positive and purifying selection in different portions of the sequence, as d_N/d_S is averaged along the gene. When $d_N/d_S<1$ negative (purifying) selection is the major determinant of the protein evolution, however, positive selection in few sites cannot be excluded. Finally we can hypothesize the action of positive (diversifying) selection when $d_N/d_S>1$.
Based on these assumptions, we started exploring the patterns of evolution of our set of orthologous MAG and testes-derived genes across the six *Glossina* species. In particular, we compared the rate of molecular evolution among a subset of the previously described set of orthologous male reproductive genes derived from testes (n=1,495) and MAG (n=59) sequences identified in all the six *Glossina* species. For such analyses, we took into account also the rate of evolution of 5,513 *G. m. morsitans* genes selected from the total genomic dataset, the orthologous of which have been identified in the other five species (*G. pallidipes, G. austeni, G. fuscipes, G. palpalis, G. brevipalpis)*.
As shown in **Fig. 2.12**, on average, the level of overall genomic selective pressure measured by the ratio d_N/d_S is significantly higher in "MAG genes" than for "testes genes" and "all genes". This suggests either more relaxed or positive selection in MAG genes. Moreover, the Bartlett test ((Bartlett test of homogeneity of variances, Bartlett's K-squared = 210.26, df = 1, p-value < 2.2e-16) suggests a high heterogeneity in the selective pressure across MAG genes, which is expected when some are under contrasting (i.e. positive, purifying and relaxed) selection.
Contrarily to MAG genes, testes genes show to be exposed to the same selective pressure acting along the entire genome, both categories being characterized by a slow rate of molecular evolution. Indeed, the estimated d_N/d_S ratio ranges from 0.10 to 0.15 for testes genes and all genes, and is significantly higher for accessory gland genes ranging from 0.25 to 0.30 ($P < 10\text{-}5$, Wilcoxon test).

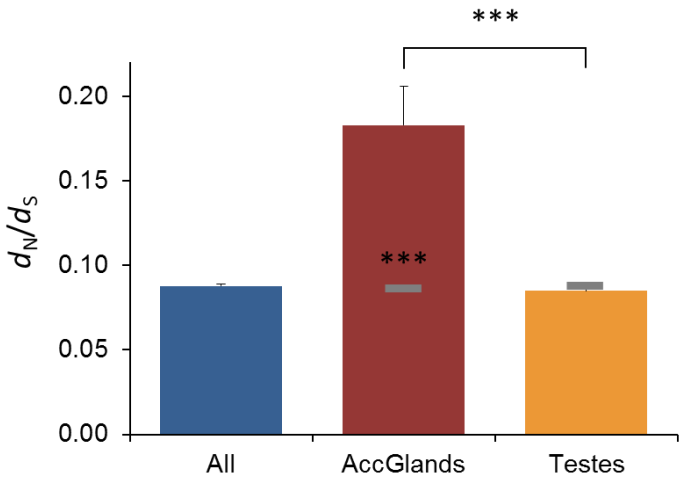


Fig. 2.12 Average selective pressure (Standard Error) in all genes (n = 5,513), genes overexpressed in accessory glands (n = 59) and in genes overexpressed in testes (n = 1,495). Grey bars are the mean of genes not overexpressed in either of the two tissues. *** = P < 10-5, Wilcoxon test.

Such result is even clearer when the rate of molecular evolution is estimated for testes and MAG genes present in each of the six *Glossina* species. For each of the three datasets analyzed, a consensus evolutionary tree both for the d_N and the d_S substitution rates has been generated **(Fig. 2.13)**. The selective pressure levels are more than doubled in all branches of the MAG genes phylogenetic tree.
These data are consistent with the observation that male reproductive genes, especially those highly expressed in MAGs undergo a rapid evolution. We can speculate that the rapid adaptive evolution of the male genes examined in this study could contribute to the *Glossina* male-hybrid sterility that is often suggested as a first step in reproductive isolation (Müller *et al.*, 2012).

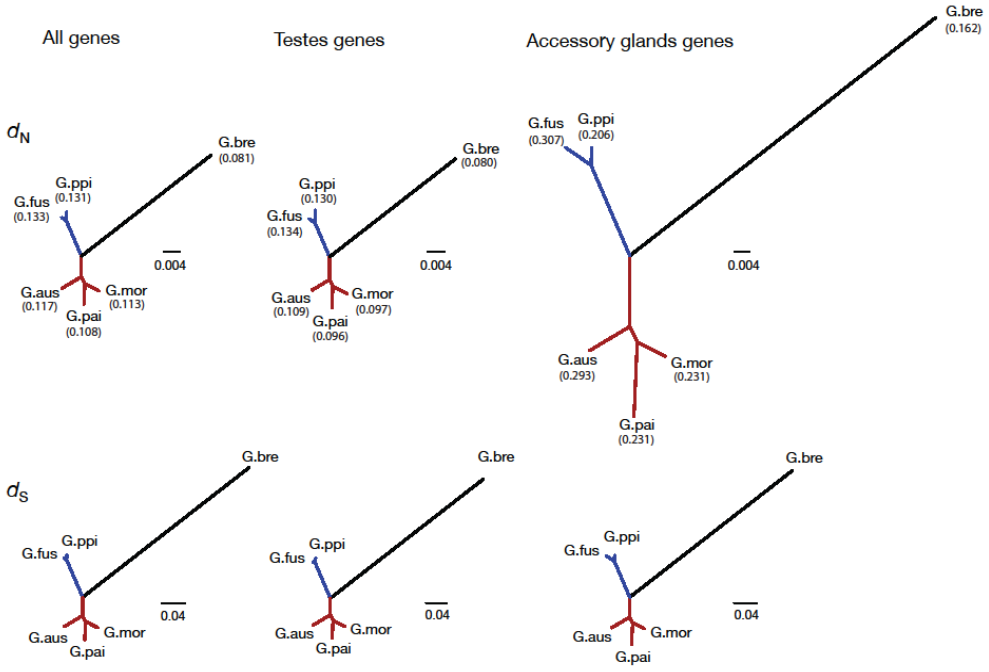


Fig. 2.13 Consensus evolutionary analysis of orthologous genes in the six *Glossina* species. Upper and lower are, respectively, the trees derived from analyses of nonsynonymous (d_N) and synonymous (d_S) substitutions. All genes, testes-biased and MAG-biased genes are represented (from left to right). In red, the *morsitans* group (savannah flies), in blue, the *palpalis* group (riverine flies), *G. brevipalpis* (forest flies) in black. The dN/dS for each species is given in parentheses.

Candidate male reproductive genes for positive selection in six Glossina species

Genomic and transcriptomic *G. m. morsitans* resources allowed us to study the pattern of evolution of male reproductive genes, with a focus on rapidly evolving sequences.
We took advantage of the release of the complete genomes of six *Glossina* species (*G. morsitans*, *G. pallidipes*, *G. austeni*, *G. fuscipes*, *G. palpalis* and *G. brevipalpis*) to performe comparative genomics analysis.
We focused our analysis on a set of 2,016 orthologous genes from six *Glossina* species, namely 1924 and 92 testes- and MAG-biased genes **(Tables 2.14, 2.15)**. Both site- and branch-site models have been tested. For branch tests, considering that male reproductive tissue-specific transcriptomic data are available only for *G. m. morsitans*,

we considered this species as the foreground species (i.e. the species of interest). Positive selection acting on particular sites across the entire phylogeny has been tested comparing codon substitution models M2a (positive selection), M1 (nearly neutral), M7 (beta) and M8 (beta+ω>1: continuous), both for testes- and MAG-biased genes (Yang *et al.*, 2000) **(Tables 2.14, 2.15)**.

Table 2.14 Number of candidate testes genes for lineage-specific selective pressure in *G. morsitans* (G.mor*), G. austeni* (G.aus), *G. fuscipes* (G.fus), *G. pallidipes* (G.pai) and *G. palpalis* (G.ppi).

	G.mor	G.aus	G.fus	G.pai	G.ppi	G [a]
Tot. genes [b]	1924	1909	1883	1877	1883	1924
Site test A [c]	--	--	--	--	--	57 (35)
Site test B [d]	--	--	--	--	--	141 (57)
Branch test [e]	136 (20)	199 (56)	124 (10)	162 (15)	117 (15)	586 (93)
Branch-site test [f]	55 (14)	64 (12)	36 (11)	50 (14)	34 (8)	216 (55)

[a] Candidate genes in at least one lineage.
[b] Total number of tested genes. Branch and branch-site tests were performed using each of the five species as foreground branch. Numbers may differ because alignments contained either six or five species (including the one used as foreground branch).
[c] Site test for positive selection using codon substitution models M2a and M1. Here and for the other tests, we report total number of candidates (at P<0.05) and, in brackets, those passing multiple-test correction (FDR < 0.20).
[d] Site test for positive selection using codon substitution models M8 and M7.
[e] Branch test for divergent selective pressure acting on the foreground species.
[f] Branch-site test for positive selection acting on some sites in the foreground species.

Table 2.15 Number of candidate accessory-gland genes for lineage-specific selective pressure in *G. morsitans* (G.mor*), G. austeni* (G.aus), *G. fuscipes* (G.fus), *G. pallidipes* (G.pai) and *G. palpalis* (G.ppi).

	G.mor	G.aus	G.fus	G.pai	G.ppi	G [a]
Tot. genes [b]	92	92	92	91	88	92
Site test A [c]	--	--	--	--	--	6 (3)
Site test B [d]	--	--	--	--	--	7 (3)
Branch test [e]	5 (1)	13 (2)	8 (0)	3 (0)	4 (1)	30 (4)
Branch-site test [f]	5 (2)	9 (2)	2 (2)	5 (1)	1 (0)	19 (6)

[a] Candidate genes in at least one lineage.
[b] Total number of tested genes. Branch and branch-site tests were performed using each of the five species as foreground branch. Numbers may differ because alignments contained either six or five species (including the one used as foreground branch).
[c] Site test for positive selection using codon substitution models M2a and M1. Here and for the other tests, we report total number of candidates (at $P<0.05$) and, in brackets, those passing multiple-test correction (FDR < 0.20).
[d] Site test for positive selection using codon substitution models M8 and M7.
[e] Branch test for divergent selective pressure acting on the foreground species.
[f] Branch-site test for positive selection acting on some sites in the foreground species.

Testes and MAG genes subjected to positive and neutral selection in the six Glossina species

Approaching this analysis, it must be taken into account that positive selection is difficult to be detected since it rarely occurs over all sites in a gene or in all species of a phylogeny (Jeffares *et al.*, 2015). To test each of these possibilities, we performed three tests. We used:

- site models (Site A, M1-M2 and B, M7-M8) to evaluate the possibility that positive selection acted in some sites in the gene;
- branch models (Br) to study its effect over different branches in the phylogeny;
- branch-site models (BrS) to consider the action of positive selection in particular sites only along a single branch of the phylogeny.

Site and branch-site models, where the d_N/d_S ratio varies only within specific sites of the genes in all species or both among sites and across the branches of the phylogeny, are the most realistic. If only few sites of the protein are fast evolving, adaptive evolution will not be detected by branch-models since most of the sites in the protein will remain under purifying selection (Jeffares *et al.*, 2015).

Tables 2.16 and **2.18** show testes (n=1924) and MAGs genes (n=92) tested for positive and neutral selection. The analysis, as evident from the tables, concerned only five species *G. morsitans, G. pallidipes, G. austeni, G. fuscipes and G. palpalis*. Indeed, because of the lack of knowledge about the ancestral root, we used an unrooted tree using *G. brevipalpis* as outgroup, with the ancestral root being necessarily between the outgroup and the rest of the taxa in the tree.

It appears clearly that in both tissues there is heterogeneity within and between the genes, which in the different considered species, display positive selection.

Among the listed genes, those showing a q-value (FDR, false discovery rate) < 0.20 were selected for further analysis. The results of such analysis based on site B (M7-M8) and branch-site models are shown in **tables 2.17 and 2.19** (*G. m. morsitans* foreground branch).

From these tables it appears clear that 58 out of 1924 **(2.8%)** testes genes are under positive selection after site B (M7-M8) test, while only thirteen **(0.7%)** after branch-site tests **(table 2.17)**. As far as it concern MAG genes, 3 out of 92 **(3.2%)** resulted under positive selection from the site B (M7-M8) model test and two **(2.17%)** from the branch-site model **(table 2.19)**. These data clearly demonstrate that the MAG tissue displays a higher percentage of genes that are under positive selection compared to testes.

Table 2.16 Testes genes (n= 1924) in *G. m. morsitans* (G.mor*)*, *G. austeni* (G.aus), *G. fuscipes* (G.fus), *G. pallidipes* (G.pai) and *G. palpalis* (G.ppi) tested for positive selection after site (A and B), branch (Br) and branch-site (BrS) models.

			Site test		G.mor		G.aus		G.fus		G.pai		G.ppi	
dataset	**gene**	**putative_function**	**A**	**B**	**Br**	**BrS**	**Br**	**BrS**	**Br**	**BrS**	**Br**	**BrS**	**Br**	**BrS**
all	GMOY000082	chaoptin					*							
all	GMOY000168	sorting nexin-25	**	**				*						
all	GMOY000485	NA						*						
all	GMOY000539	ataxin-2 homolog isoform X2					*							
all	GMOY000654	Palmitoyltransferase ZDHHC17			*									
all	GMOY000863	ubiquitin carboxyl-terminal hydrolase 34		*								**		
all	GMOY001139	intraflagellar transport 122 homolog					*							
all	GMOY001316	phosphatase PP2A 55 kDa regulatory subunit isoform X1									*			
all	GMOY001320	split ends										*		
all	GMOY001350	ethanolamine kinase	**	***										
all	GMOY001744	serine-rich adhesin for platelets							*					
all	GMOY001748	NA			*									
all	GMOY001860	isoform C												**
all	GMOY001887	leucine-rich repeats and immunoglobulin-like domains 3				**								
all	GMOY001913	ubiquitin- ligase E3C					*							

all	GMOY001941	sarcalumenin isoform X1											*	
all	GMOY001977	synaptotagmin 4 isoform X9					*							
all	GMOY002025	tyrosine- kinase Fer isoform X1			*									
all	GMOY002064	zinc finger 271-like		*										
all	GMOY002287	apoptotic chromatin condensation inducer in the nucleus	*	**										
all	GMOY002349	sorbitol dehydrogenase											**	***
all	GMOY002375	short spindle isoform A			**									
all	GMOY002564	serine-rich adhesin for platelets							*					
all	GMOY002800	furin-like protease 2 isoform X1		***										
all	GMOY002820	dmX 2					*							
all	GMOY002841	DNA repair RAD50											***	
all	GMOY002881	uridine diphosphate glucose pyrophosphatase-like		*										
all	GMOY003057	cyclin-dependent kinase 14 isoform X1	***	***										
all	GMOY003081	helix-loop-helix 11				*								
all	GMOY003155	von Willebrand factor D and EGF domain-containing					*							
all	GMOY003188	double-strand break repair MRE11											**	
all	GMOY003237	GTPase-activating r					**							
all	GMOY003370	Tyrosine- kinase shark	**	**									*	

all	GMOY003432	sprint isoform X1					*							
all	GMOY003582	NA					*							
all	GMOY003628	katanin p60 ATPase-containing subunit A-like 1	***	***										
all	GMOY003636	NA		**										
all	GMOY003638	nidogen-2					*							
all	GMOY003771	ras-related and estrogen-regulated growth inhibitor					*							
all	GMOY003823	PREDICTED: uncharacterized protein LOC106091657 isoform X1						*						
all	GMOY003915	cullin-3 isoform X2	***	***	***		***		***		***		***	
all	GMOY004004	ubiquitin-conjugating enzyme E2 Q2	***	***										
all	GMOY004064	DNA repair REV1								**				
all	GMOY004222	glutamate receptor kainate 2				**								
all	GMOY004238	NA		*										
all	GMOY004338	cytochrome P450 CYP12A2-like					*							
all	GMOY004441	isoform A	*	***										
all	GMOY004688	zinc finger Xfin			*									
all	GMOY004744	alpha- sarcomeric isoform X1	***	***										
all	GMOY004783	ATP-dependent 6-phosphofructokinase isoform X1					*				***			

all	GMOY004875	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase	*	**										
all	GMOY004932	rab GTPase-activating 1-like isoform X1		*										
all	GMOY005070	dynein heavy cytoplasmic isoform X1					*							
all	GMOY005075	NA					***							
all	GMOY005314	DNA fragmentation factor subunit beta	*	***										
all	GMOY005361	NA									*			
all	GMOY005379	low-density lipo receptor-related 2					***							
all	GMOY005409	NA											*	
all	GMOY005561	JHEH2					**							
all	GMOY005627	MON2 homolog	**	***	**		**							
all	GMOY005703	myosin heavy muscle isoform X11	***	***										
all	GMOY005863	NA	***	***									**	
all	GMOY005903	Netrin-B										*		
all	GMOY005954	NA	***	***		***								
all	GMOY006023	ring canal kelch					*							
all	GMOY006100	Laminin subunit alpha					*							
all	GMOY006119	LETM1 and EF-hand domain-containing anon- mitochondrial			*									

all	GMOY006132	MATH and LRR domain-containing PFE0570w		*										
all	GMOY006227	phospholipase A1 2	***	***										
all	GMOY006302	CD81 antigen							**	**				
all	GMOY006377	histone acetyltransferase p300 isoform X1					*							
all	GMOY006415	NA			*									
all	GMOY006503	chitooligosaccharidolytic beta-N-acetylglucosaminidase					**							
all	GMOY006579	RNA polymerase II large subunit		**			*							
all	GMOY007021	inorganic phosphate cotransporter		**										
all	GMOY007044	NADP-dependent malic enzyme				*								
all	GMOY007157	ubiquitin conjugation factor E4 B		*										
all	GMOY007253	ryanodine receptor 44F isoform X5	***	***							***	***		
all	GMOY007388	enhancer of mRNA-decapping 3					**							
all	GMOY007547	monocarboxylate transporter 8	***	***			**		***	***	***	**		
all	GMOY007750	gamma-tubulin complex component 5	*	*										
all	GMOY007797	NA					*							
all	GMOY007818	hypothetical protein FF38_05875				**								
all	GMOY007838	uncharacterized serine-rich		*										

all	GMOY007883	papilin isoform X2												*
all	GMOY008094	microtubule-associated futsch												**
all	GMOY008119	dedicator of cytokinesis 7 isoform X2										*		
all	GMOY008138	probable ATP-dependent RNA helicase kurz					*					*		
all	GMOY008388	disco- isoform A				*								
all	GMOY008433	NA						*						
all	GMOY008456	A disintegrin and metallo ase with thrombospondin motifs 9 isoform X1						*						
all	GMOY008545	NA		*			**							
all	GMOY008639	non-specific lipid-transfer	**	***					*					
all	GMOY008690	unconventional myosin-Va isoform X1						***						
all	GMOY008811	coiled-coil and C2 domain-containing 1-like isoform X2					*							
all	GMOY008814	manganese-transporting ATPase 13A1					**							
all	GMOY008876	Transcription initiation factor TFIID subunit 2		*										
all	GMOY008882	#NOME?					*							
all	GMOY009072	outer dense fiber 3-B					*							
all	GMOY009123	tudor domain-containing 7B					*							

all	GMOY009167	mediator of RNA polymerase II transcription subunit 23											***	
all	GMOY009235	histone acetyltransferase KAT6A												***
all	GMOY009251	over compensating isoform A										*		
all	GMOY009256	dynein heavy chain axonemal			**		*							
all	GMOY009618	NA		*			***	*						
all	GMOY009682	cell division cycle 7-related kinase					***							
all	GMOY009683	NA				***						**		
all	GMOY009744	apolipo D-like	*	*										
all	GMOY009833	elav isoform X2							**					
all	GMOY009862	baculoviral IAP repeat-containing 6		*		*	***							
all	GMOY009931	probable serine threonine-kinase MARK-A isoform X2						**						
all	GMOY009936	Serine threonine- kinase Warts						**						
all	GMOY009959	filaggrin isoform X2								*				
all	GMOY010088	NA					*							
all	GMOY010094	NA		*										
all	GMOY010114	Inactive rhomboid 1					*							
all	GMOY010152	voltage-dependent T-type calcium channel subunit alpha-1G				*								

all	GMOY010173	nesprin-1 isoform X10			**						**			
all	GMOY010213	DNA topoisomerase 2			*						***			
all	GMOY010214	tRNA-dihydrouridine(20a 20b) synthase [NAD(P)+]-like							*					
all	GMOY010234	alsin homolog					*							
all	GMOY010310	ethanolaminephosphotransferase 1 isoform X1						**						
all	GMOY010311	probable cytochrome P450 4d14		*										
all	GMOY010344	leucine-rich repeat-containing 20 isoform X1		*										
all	GMOY010384	flocculation FLO11-like								***				
all	GMOY010447	WD repeat and FYVE domain-containing 3					*							
all	GMOY010651	peanut											***	
all	GMOY010684	scalloped isoform X1	***	***										
all	GMOY010944	helicase domino isoform X1	**				***				**		**	
all	GMOY011064	NA			*		***				*			
all	GMOY011075	hippocampus abundant transcript 1 isoform X1										*		
all	GMOY011328	E3 ubiquitin- ligase TRIM33 isoform X1	***	***						*				
all	GMOY011499	tubulin polyglutamylase TTLL5								**				
all	GMOY011575	Homeotic empty spiracles						*						
all	GMOY011709	xmas-2		**										

all	GMOY011918	NA					**							
all	GMOY011994	discs overgrown kinase					*							
all	GMOY012004	Organic cation transporter			***	***	***			***				***
Noaus	GMOY009305	NA												*
NOb	GMOY000017	NA									**			
NOb	GMOY000301	structural maintenance of chromosomes 2				*								
NOb	GMOY000392	Transcription termination factor 2									***			
NOb	GMOY001122	chromatin-remodeling complex ATPase chain Iswi			*		*				***			
NOb	GMOY002664	NA	***	***						**				
NOb	GMOY002933	NA		*										
NOb	GMOY003749	NA	***	***										
NOb	GMOY003818	Src oncogene at isoform B				*								
NOb	GMOY005118	sialin					*							
NOb	GMOY005278	NA										*		
NOb	GMOY005825	titin isoform X1					*							
NOb	GMOY006700	lamin Dm0	*	*										
NOb	GMOY007135	pellino	***	***	***		***		***		***		***	
NOb	GMOY007250	dnaJ homolog subfamily B member 2 isoform X1					*							
NOb	GMOY007798	E3 ubiquitin- ligase sina	**	***										

NOb	GMOY007904	sevenless								**				
NOb	GMOY007927	chitooligosaccharidolytic beta-N-acetylglucosaminidase		*										
NOb	GMOY009121	ecdysone receptor isoform X2			*									
NOb	GMOY010705	NA	***	***										
NOb	GMOY010738	high affinity cAMP-specific						**						
NOb	GMOY010740	ubiquitin carboxyl-terminal hydrolase 35			*									
NOb	GMOY010796	NA		*										
NOb	GMOY011696	isoform G										*		
NOb	GMOY011858	titin								*				
NOf	GMOY000269	proton-coupled folate transporter	***	***			**					*		
NOf	GMOY010920	ubiquitin-like modifier-activating enzyme ATG7			***									
NOf	GMOY011611	NA											*	
NOpai	GMOY002453	NA					*							
NOpai	GMOY003306	proteoglycan 4-like isoform X1											**	
NOpai	GMOY004000	NA							***				**	*
NOpai	GMOY007507	transcription elongation factor B polypeptide 2				*								
NOpai	GMOY008607	nose resistant to fluoxetine 6	***	***										
NOppi	GMOY002073	NA	***	***								*		
NOppi	GMOY003159	eater			*									

NOppi	GMOY004902	3-hydroxyisobutyryl-mitochondrial isoform X1									***			
NOppi	GMOY006151	low-density lipo receptor-like isoform X9	***	***										
NOppi	GMOY008445	1-phosphatidylinositol 3-phosphate 5-kinase					*							

[a] Gene name is that of the *G. m. morsitans* ortholog.
[b] Site test for positive selection using codon substitution models M2a and M1.
[c] Site test for positive selection using codon substitution models M8 and M7.
[d] Branch test for divergent selective pressure acting on the foreground species.
[e] Branch-site test for positive selection acting on some sites in the foreground species.
* FDR < 0.20; ** FDR < 0.05; *** FDR < 0.005.

Table 2.17 Testes genes showing a q-value (FDR, false discovery rate) < 0.20 based on site B (M7-M8) and branch-site models results (*G. m. morsitans* foreground branch).

Testes* site B (M7-M8)_morsitans foreground branch, q-value < 0.2

dataset	gene	putative_function	P value	qvalue (FDR)
all	GMOY007547	monocarboxylate transporter 8	1.77E-43	1.44E-39
all	GMOY007253	ryanodine receptor 44F isoform X5	4.43E-25	8.97E-22
all	GMOY005954	NA	3.27E-22	5.29E-19
all	GMOY004744	alpha- sarcomeric isoform X1	4.65E-18	5.39E-15
all	GMOY011328	E3 ubiquitin- ligase TRIM33 isoform X1	7.18E-18	7.27E-15

NOpai	GMOY008607	nose resistant to fluoxetine 6	3.74E-17	3.36E-14
all	GMOY003628	katanin p60 ATPase-containing subunit A-like 1	1.05E-16	8.49E-14
all	GMOY005703	myosin heavy muscle isoform X11	1.65E-16	1.22E-13
all	GMOY002800	furin-like protease 2 isoform X1	3.15E-15	1.96E-12
NOb	GMOY007135	pellino	7.21E-13	3.44E-10
NOb	GMOY003749	NA	1.55E-12	6.96E-10
NOb	GMOY010705	NA	1.51E-10	5.83E-08
all	GMOY005863	NA	3.11E-10	1.10E-07
all	GMOY003915	cullin-3 isoform X2	5.94E-10	1.92E-07
all	GMOY004004	ubiquitin-conjugating enzyme E2 Q2	1.54E-09	4.79E-07
all	GMOY010684	scalloped isoform X1	2.54E-09	7.61E-07
NOppi	GMOY002073	NA	5.05E-08	0.000011
all	GMOY005627	MON2 homolog	1.46E-07	0.000027
all	GMOY006227	phospholipase A1 2	1.47E-07	0.000027
NOppi	GMOY006151	low-density lipo receptor-like isoform X9	5.02E-07	0.000086
all	GMOY003057	cyclin-dependent kinase 14 isoform X1	1.26E-06	0.000197
NOb	GMOY002664	NA	1.26E-06	0.000197
all	GMOY008639	non-specific lipid-transfer	6.28E-06	0.000795
all	GMOY004441	isoform A	7.34E-06	0.000914
all	GMOY005314	DNA fragmentation factor subunit beta	7.59E-06	0.000931
NOb	GMOY007798	E3 ubiquitin- ligase sina	0.000014	0.001607
NOf	GMOY000269	proton-coupled folate transporter	0.000014	0.001650

all	GMOY001350	ethanolamine kinase	0.000050	0.004913
all	GMOY000168	sorting nexin-25	0.000061	0.005712
all	GMOY006579	RNA polymerase II large subunit	0.000077	0.006847
all	GMOY003370	Tyrosine- kinase shark	0.000087	0.007614
all	GMOY003636	NA	0.000331	0.025807
all	GMOY004875	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II isoform X4	0.000340	0.026056
all	GMOY002287	apoptotic chromatin condensation inducer in the nucleus	0.000341	0.026056
all	GMOY007021	inorganic phosphate cotransporter	0.000520	0.036420
all	GMOY011709	xmas-2	0.000522	0.036420
all	GMOY002881	uridine diphosphate glucose pyrophosphatase-like	0.001058	0.068251
all	GMOY007750	gamma-tubulin complex component 5	0.001075	0.068251
NOb	GMOY006700	lamin Dm0	0.001066	0.068251
all	GMOY006132	MATH and LRR domain-containing PFE0570w	0.001537	0.090209
all	GMOY008545	NA	0.001768	0.101337
all	GMOY010094	NA	0.001832	0.103772
all	GMOY009744	apolipo D-like	0.001999	0.110928
all	GMOY002064	zinc finger 271-like	0.002229	0.121162
all	GMOY008876	Transcription initiation factor TFIID subunit 2	0.002582	0.129488
all	GMOY007157	ubiquitin conjugation factor E4 B	0.002788	0.134448
all	GMOY010311	probable cytochrome P450 4d14	0.003007	0.142833
NOb	GMOY007927	chitooligosaccharidolytic beta-N-acetylglucosaminidase	0.003015	0.142833

all	GMOY004238	NA	0.003433	0.158923
all	GMOY004932	rab GTPase-activating 1-like isoform X1	0.003551	0.162739
all	GMOY009862	baculoviral IAP repeat-containing 6	0.003710	0.167890
all	GMOY009618	NA	0.003835	0.169847
all	GMOY000863	ubiquitin carboxyl-terminal hydrolase 34	0.004080	0.178677
all	GMOY010344	leucine-rich repeat-containing 20 isoform X1	0.004196	0.180807
all	GMOY007838	uncharacterized serine-rich	0.004799	0.194388
NOb	GMOY002933	NA	0.004878	0.196612
NOb	GMOY010796	NA	0.004945	0.198322

Testes Branchsite_morsitans foreground branch, q-value < 0.2**

dataset	gene	putative_function	P value	qvalue (FDR)
all	GMOY012004	Organic cation transporter	4.25E-20	3.45E-16
all	GMOY005954	NA	1.57E-07	0.000211
all	GMOY009683	NA	5.26E-06	0.004750
all	GMOY001887	leucine-rich repeats and immunoglobulin-like domains 3	8.48E-06	0.006243
all	GMOY007818	hypothetical protein FF38_05875	0.000045	0.021439
all	GMOY004222	glutamate receptor kainate 2	0.000149	0.044724
all	GMOY008388	disco- isoform A	0.000233	0.060701
all	GMOY009862	baculoviral IAP repeat-containing 6	0.000247	0.060701
all	GMOY007044	NADP-dependent malic enzyme	0.000256	0.061105

Nob	GMOY000301	structural maintenance of chromosomes 2	0.000325	0.071201
Nopai	GMOY007507	transcription elongation factor B polypeptide 2	0.000457	0.084646
all	GMOY010152	voltage-dependent T-type calcium channel subunit alpha-1G	0.000717	0.118594
all	GMOY003081	helix-loop-helix 11	0.000767	0.119638
Nob	GMOY003818	Src oncogene at isoform B	0.001087	0.149306

Table 2.18. Number of MAG genes (=92) in *Glossina m. morsitans* (G.mor*), G. austeni* (G.aus), *G. fuscipes* (G.fus), *G. pallidipes* (G.pai) and *G. palpalis* (G.ppi) tested for positive seletion after site (A and B), branch (Br) and branch-site (BrS) models.

			Site test		**G.mor**		**G.aus**		**G.fus**		**G.pai**		**G.ppi**	
dataset	**gene**	**putative_function**	**A**	**B**	**Br**	**BrS**	**Br**	**BrS**	**Br**	**BrS**	**Br**	**BrS**	**Br**	**BrS**
all	GMOY002550	plancitoxin-1			**									
all	GMOY002583	NA										*		
all	GMOY005874	NA				*								
all	GMOY005914	transmembrane channel 1								*				
all	GMOY007314	NA	**	**										
NOb	GMOY002399	NA	***	***			*	*		**				
NOb	GMOY004505	NA				*								
NOb	GMOY007759	NA	***	***										

NOb	GMOY010622	synaptic vesicular amine transporter isoform X1						*					*	
NOppi	GMOY009723	angiotensin-converting enzyme-related					*							

[a] Gene name is that of the *G. m. morsitans* ortholog.
[b] Site test for positive selection using codon substitution models M2a and M1.
[c] Site test for positive selection using codon substitution models M8 and M7.
[d] Branch test for divergent selective pressure acting on the foreground species.
[e] Branch-site test for positive selection acting on some sites in the foreground species.
* FDR < 0.20; ** FDR < 0.05; *** FDR < 0.005.

Table 2.19 Candidate MAG genes showing a q-value (FDR, false discovery rate) < 0.20 based on site B (M7-M8) and branch-site models results (*G. m. morsitans* foreground branch).

MAG* Site B (M7-M8)_morsitans foreground branch, q-value < 0.2

dataset	gene	putative_function	P value	qvalue (FDR)
Nob	GMOY002399	NA	1.04E-08	2.55E-06
Nob	GMOY007759	NA	5.08E-06	0.000686
all	GMOY007314	NA	0.000056	0.005341

MAG Branchsite_morsitans foreground branch, q-value < 0.2**

dataset	gene	putative_function	P value	q value (FDR)
Nob	GMOY004505	NA	0.000897	0.134543
all	GMOY005874	NA	0.000929	0.136767

dN/dS ratio of MAG and testes genes selected for functional GO categories in six Glossina species

In order to go deeper into the analysis of the evolutionary rates of MAG and testes orthologous genes across the six *Glossina* species (*G. morsitans, G. pallidipes, G. austeni, G. fuscipes, G. palpalis, G. brevipalpis*), we estimated the d_N/d_S ratio associated to each sequence (Yang, 2007).
The relationships among gene function and gene evolutionary traits across the six species has been approached in both testes and MAGs by coupling GO functional classes with orthologous group evolutionary properties. Functional categories have been derived from curated lists associated with GO Molecular Function Level III. In the case of novel tsetse proteins, for which GO terms could not be assigned, but displaying high d_N/d_S values, potential functional role has been inferred, when possible, using InterPro.
Glossina orthologous genes were categorized into 17 functional classes in the MAGs **(Fig. 2.20)** and 42 in the testes **(Fig. 2.21)**. It is noteworthy that assignment of GO terms are made through homology/orthology to genes with annotated terms from *D. melanogaster*, thus assignments are usually biased towards slower evolving, well-conserved genes.

d_N/d_S ratio in MAG GO functional categories

The different d_N/d_S ratios estimated for the functional GO classes of MAG orthologous genes across the six *Glossina* species are represented in **Fig. 2.20**. It appears that the estimated d_N/d_S average values are very heterogeneous among the classes, ranging from 0.5 for odorant binding to 0.05 for carbohydrate derivative binding genes. Other interesting categories with relatively high d_N/d_S average ratios (0.25-0.44) are represented by novel genes ("none" in the figure), lyase activity, peptidase regulator activity and enzyme inhibitor activity. On the other hand, organic cyclic compound binding, oxidoreductase activity, small molecule binding, ion binding, transmembrane transporter activity, substrate specific transporter activity are those displaying the lowest d_N/d_S estimations. Considering their basic role in metabolic pathways, is not unexpected that such genes are evolving at a slow rate.
By contrast, OBPs associated to reproductive tissues display high d_N/d_S values. The role of OBPs in tissues other than those involved in chemoreception may be explained by their specific carrier activities. Indeed, these proteins can transport any physiologically active ligands, including those transferred from males to females during the mating (Arya *et al.*, 2010). Sex pheromones and other male products present in the seminal fluid and transferred to the females during copulation are poorly soluble compounds that, in aqueous media, need to be transported by binding proteins. The role of pheromone carrier has been described for both OBPs and CSPs, and the delivery of male products

to reproductive organs might suggests species- and sex-specific roles, thus justifying their fast evolution (Li *et al.*, 2008; Sirot *et al.*, 2008).

Special attention must be addressed to the relatively high d_N/d_S values estimated for "novel" genes. They shared no similarity to sequences present in the GenBank database and were thus classified as "novel". The proteins encoded by these genes have been found in the ejaculate of *G. m. morsitans*, which is transferred to the female during the mating. They have been suggested to be implicated in post-mating behaviors and as such, they may have implication in sexual selection and speciation. Moreover, in *Glossina* genomes, these putative novel genes revealed many duplication events, suggesting lineage-specific gene duplications, as already observed for other seminal fluid proteins (Almeida and Desalle, 2008). These considerations predict the fast evolutionary rate of this category, as indicated by the high d_N/d_S estimated average value.

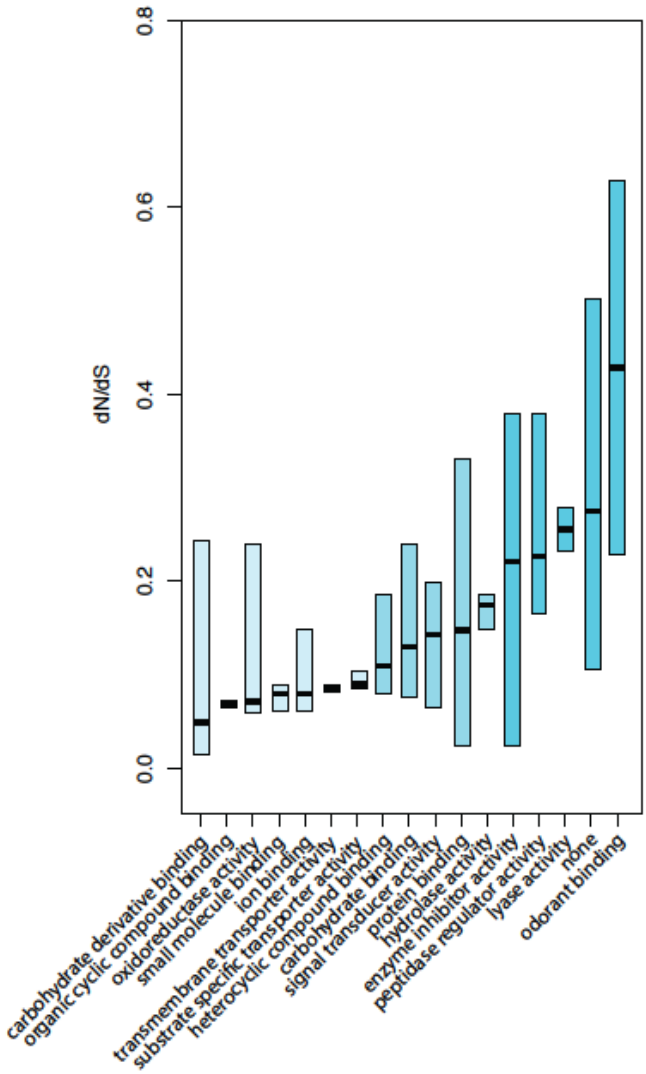


Fig. 2.20 d_N/d_S ratio of selected gene functional categories in MAGs. Notched box plots show medians and extend to the first and third quartiles.

dN/dS ratios in testes GO functional categories

Contrarily to MAGs, testes genes display a low variability in the distribution of d_N/d_S estimate across the functional categories. Forty-two GO functional classes have been

identified in testes genes **(Fig. 2.21)**. The mean for d_N/d_S values is lower than 0.3 for all the GO categories, meaning purifying selection. It is known that testes and MAGs participate in the maintenance of reproductive functions of insect males and, based on the information available from other insect taxa, it is likely that, in tsetse testes, key regulatory genes of spermatogenesis tend to be conserved to guarantee the male-specific processes required for sperm production (Bonilla and Xu, 2008). On the other hand, the MAG secretions may act as key factors in male reproductive success and, as such, the accessory gland protein-encoding genes are subject to rapid evolution as a result of sexual conflict and competition (Ram and Wolfner, 2007).

The differences in term of evolution between testes and MAGs is presented in **Fig. 2.22**. It is interesting to notice that the selective pressure is acting differently on categories of genes associated to different male reproductive tissues.

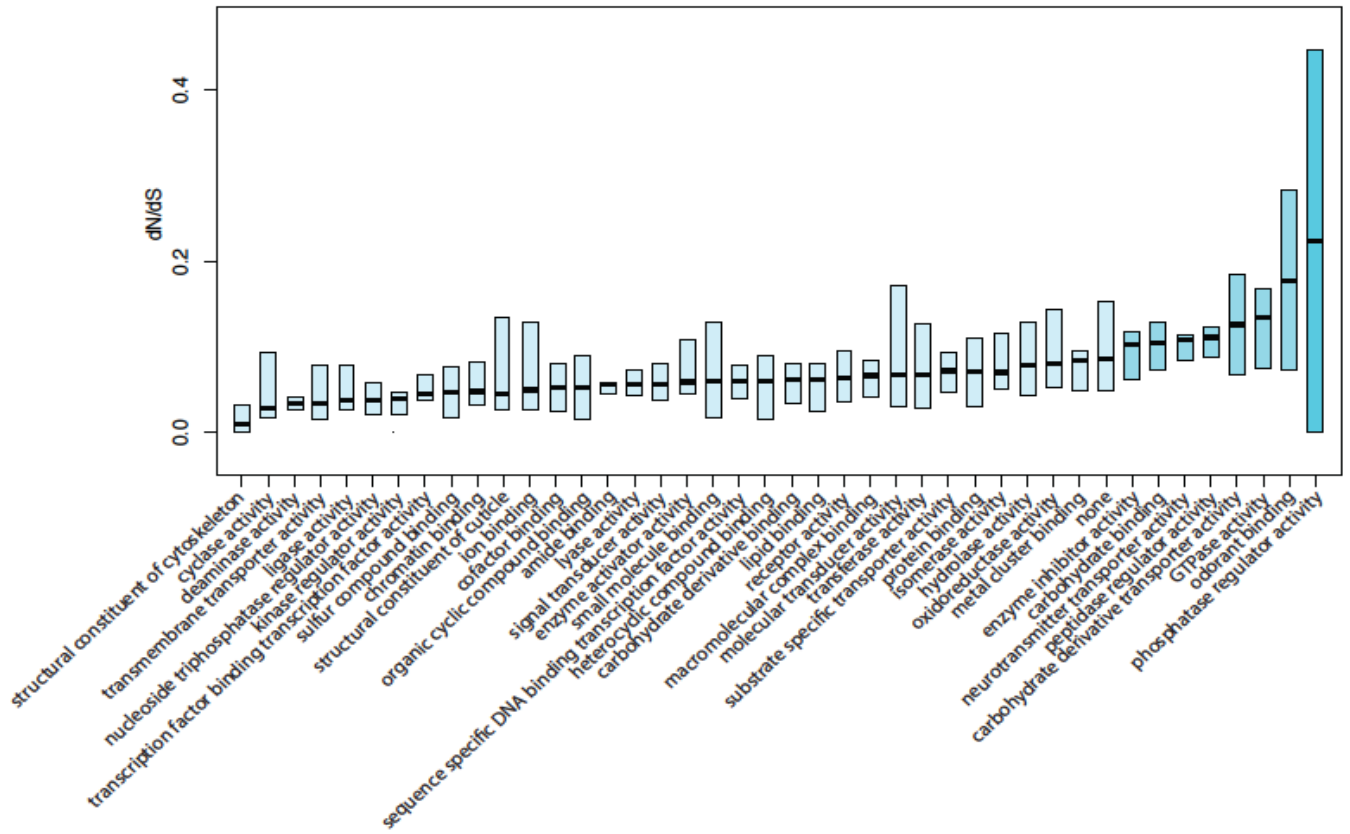


Fig. 2.21 Evolutionary properties of selected gene functional categories in testes. Notched box plots show medians and extend to the first and third quartiles.

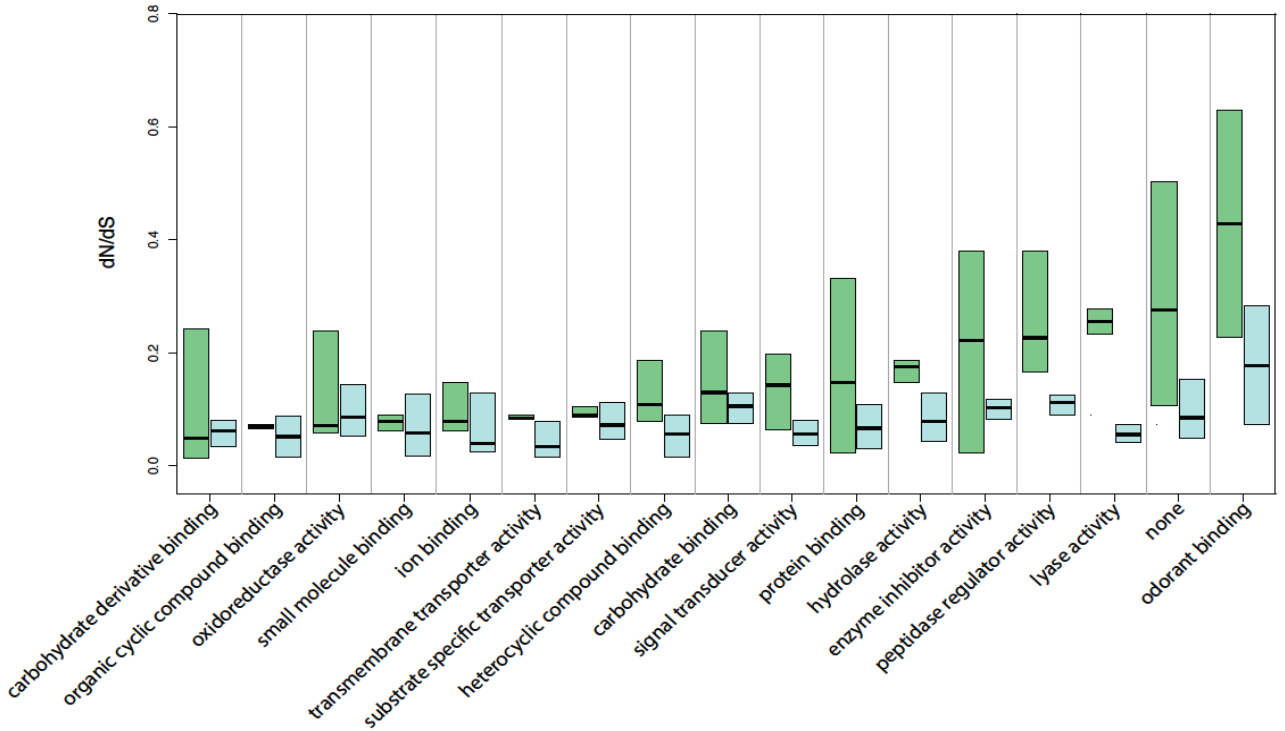


Fig. 2.22 Gene functional categories comparison between accessory gland and testes proteins. Testes = blue; AG = green. Notched box plots show medians and extend to the first and third quartiles.

MAGs and testes genes under positive and neutral selection

Genes within different functional categories displaying the highest d_N/d_S ratio values in MAGs and testes were considered for cross comparison with the outcome of previously performed site B (M7-M8) and branch-site model tests. Genes were selected using 0.6 d_N/d_S ratio as a threshold. As it appears from **table 2.23**, some of these genes with the highest d_N/d_S values have been also sorted by site-specific (*) and branch-site models (**) as candidates for positive selection **(Table 2.23)**.

Eight novel genes and two odorant binding genes where considered in MAGs. In testes, ten novel genes, one transferase activity and one carbohydrate derivative binding gene have been sorted. Six out of height novel MAG genes and the testes genes display orthology in only five species (*G. morsitans, G. pallidipes, G. austeni, G. fuscipes, G. palpalis).* All the other genes show orthology in all the six *Glossina* species.

For MAGs, one novel (GMOOY007759) and one OBP gene (GMOY007314) resulted under positive selection also by site-specific B (M7-M8) test (* in **Table 2.23)**, while only GMOY004505 resulted under positive selection also by branch-site model (** in **Table 2.23)**.

For testes, four novel genes resulted positive for selection also by site B (M7-M8) test. None of the testes genes resulted positive to selection according to the branch-site test.

The identification of genes from the male reproductive tissues which are clearly under positive selection it is of particular value since the majority of them are coding for proteins which are component of the male ejaculate. As such, they are transferred to the female during the mating. They are expected to reflect the life histories of the species to promote post-mating behaviors, optimal sperm delivery, survival and egg fertilization capacity.

Table 2.23 Twenty-two transcripts displaying male-biased expression and significant d_N/d_S ratio. Omegafix is defined as the selective pressure estimated for the whole phylogeny.
*site B (M7-M8) *G. m. morsitans* as foreground species.
**Branch-site *G. m. morsitans* as foreground species.

Accessory Glands			
Proposed functional class	Transcript name	# species	omegafix
Novel proteins (8)			
	GMOY007759*	5	1.22209
	GMOY009778	5	1.07741
	GMOY004506	5	0.85647
	GMOY006927	6	0.74206
	GMOY004505**	5	0.72393
	GMOY002398	5	0.71211
	GMOY004319	5	0.64849
	GMOY005771	6	0.60202
Odorant binding (2)			
	GMOY007314*	6	0.73572
	GMOY007757	6	0.62947
Testes			
Proposed functional class	Transcript name	# species	omegafix
Novel proteins (10)			
	GMOY006700*	5	0.91611
	GMOY007135*	5	0.8239
	GMOY010796*	5	0.79202
	GMOY006845	5	0.77696
	GMOY011410	5	0.68709
	GMOY000403	5	0.64065
	GMOY008545*	5	0.63115
	GMOY006788	5	0.61941
	GMOY012123	5	0.6159
	GMOY010304	5	0.60521
Transferase activity (1)			
	GMOY006539	5	0.75232
Carbohydrate derivative binding (1)			
	GMOY001312	5	0.69082

2.4 The Paired transcription factor in G. m. morsitans

After having identified the genes which, in the MAGs, codify for proteins that may have an important role in post-mating behaviors and fertilization, my interest was addressed to the analysis of the regulation of such genes.
Promoters play an essential role in controlling gene expression as they are the binding sites for RNA polymerase. Multiple functional sites are involved in the binding of the polymerase; these elements, such as the TATA box, the Initiator motif (Inr) and the downstream core promoter element (DPE) serve as binding sites for transcription factors, which are also crucial in gene regulation.
I studied the transcriptional regulation and promoter regions of the 24 highly expressed and MAG-specific genes because of the relevance they may have in the discovery of transcription factors that can be manipulated to interfere with their expression and, in turn, with tsetse flies fertility.

2.4.1 Materials and Methods

Promoter analysis

The promoter region of the 24 highly expressed and MAG-specific genes has been analyzed: GMOY000899, GMOY002399, GMOY004505, GMOY004506, GMOY005771, GMOY006927, GMOY006928, GMOY007759, GMOY007760, GMOY008628, GMOY009777, GMOY009778, GMOY002279, GMOY002442, GMOY004724, GMOY004725, GMOY010053, GMOY005874, GMOY005875, GMOY005876, GMOY007314, GMOY007757.
The TRANSFAC (TRANScription FACtor) database of eukaryotic transcription factors was used to predict potential transcription factor binding sites. Optimized position weight matrices in prediction of putative binding sites for transcription factors identified in *Drosophila melanogaster* have been used.

RT-PCR tissue expression

RNA isolation and RT-PCR from fresh tissues. Total RNA was isolated from MAG, testes, ejaculatory duct, fat body, gut and carcass according to the protocol supplied with Direct-zol RNA miniprep (Zymo Research). The concentration and purity of the RNA samples were determined using the NanoDrop 2000. Total RNA (2 μg) was reverse transcribed (RT) using iScript cDNA Synthesis kit (BioRad). The expression of Paired and β-tubulin cDNAs was determined using RT-PCR:
GMOY001037 (Paired) F 5'-CATCGGCAAGTTCATCATATTCG-3'
GMOY001037 (Paired) R 5'-GCGGCGGTAGTGTTGTTG-3'
Tubulin F 5'-ACGTATTCATTTCCCTTTGG-3'
Tubulin R 5'-AATGGCTGTGGTGTTGGACAAC-3'.
The final PCR products were electrophoresed on 1.5% agarose gels containing ethidium bromide along with 100bp DNA markers. The experiment was repeated three times.

Western blotting

Accessory glands, testes, ejaculatory duct, fat body, gut and carcass from 10 virgin males (5 days old) were homogenized in cracking buffer (8M urea, 2% SDS, 5% Beta-Mercaptoethanol, 125mM Tris-HCL pH 6.8). Freshly dissected tissues were added to cracking buffer and homogenization was performed in eppendorf tubes, using plastic pestles. Proteins were quantified using the RC-DC Protein Assay. 20 ug of protein mixture from each tissue were separated by SDS-PAGE and transferred to a nitrocellulose membrane using a transfer apparatus according to the manufacturer's protocol (BioRad). After incubation with 5% dry milk 0.05% PBS-T for 2 hours, the

membrane was washed once with PBS-T and incubated with rabbit polyclonal anti-segmentation protein paired antibody [0.5 ug/ml] at 4°C overnight. Membranes were washed three times for 10 minutes in PBST and incubated with a 1:20.000 dilution of goat-anti rabbit HRP conjugated antibody for 2 hours. Blots were washed with PBST three times for 5 min and developed with the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) according to the manufacturer's protocols. The membrane was washed with a stripping buffer (PBS 7ul/ml B-mercaptoetanol, 2%SDS) to remove the paired antibody. The membrane was washed twice with PBS and incubated with rabbit polyclonal tubulin antibody (1:5000) at 4°C overnight. Incubation with goat-anti rabbit HRP conjugated antibody and membrane development were repeated/performed as described above.

RNAi

Paired siRNA and negative control (GFP siRNA) were purchased from Integrated DNA Technologies. Two duplex sequences were used to interfere with the paired gene:
morsitans_paired _1: 5'-UCGCAAAUACCGUCCUUUAUCAGCUUU-3'
morsitans_paired_5: 5'-CUAUCGUUUGAUUUCCAUCUGCUUCAA-3'.
Paired and GFP siRNA were resupended in PBS (for a concentration of 1ug/ul) and injected in the abdomen of 14 male teneral flies (within 24 h from eclosion and before the first blood meal): 7 with paired siRNAs and 7 with GFP siRNAs. Every male was injected with 3ul of siRNA, paired or GFP (negative control). Injections were performed using a personalized injector. The flies received the first blood meal 24 hours after injection. Flies were dissected 72 hours after injection and the paired RNA silencing was confirmed by q-PCR.

q-PCR

The RNA was obtained from the abdomen of the injected flies and the paired expression was tested by q-PCR. Paired q-PCR primers: GAPDH expression was used as control.
Paired qPCR-F 5'- GGATATAGCAGATATAGAAGG-3'
Paired qPCR-R 5'- AAGTTGAAAGTCGTATTGA-3'
GAPDH qPCR-F 5'-CTGATTTCGTTGGTGATACT-3'
GAPDH qPCR-R 5'-CCAAATTCGTTGTCGTACCA3'.

2.4.2 Results and Discussion

Promoter analysis of highly expressed MAG genes: the GMOY005874 example

I focused on the transcriptional regulation and promoter regions of the 24 highly expressed and MAG-specific genes.
Here we show the promoter analysis results for the most abundant *G. m. morsitans* MAG gene: GMOY005874. The gene presents two paralogs, GMOY005875 and GMOY005876, and orthologs in *G. pallidipes*, *G. austeni*, *G. fuscipes*, *G. palpalis* and *G. brevipalpis* have been identified. The three paralogs showed a conserved gene structure in all six species, with two exons and a gene product that ranges from 131 to 135 amino acids in length. The predicted protein sequences resembles that of OBPs, presenting six α-helical domains and six highly conserved cysteine residue that have distinct disulphide bonds (Manoharan *et al.*, 2013). Protein similarity is about 50% between GMOY005874 and GMOY005875, 28% between GMOY005874 and GMOY005876 and 26% between GMOY005875 and GMOY005876.
The TRANSFAC (TRANScription FACtor) database of eukaryotic transcription factors has been used to predict potential transcription factor binding sites (Wingender *et al.*, 2000). A consistent conservation of transcription binding sites is present among the three paralogous genes in all the putative orthologous sequences. In particular, the analysis led to the identification of the paired transcription factor-binding site (**Fig. 2.24**).

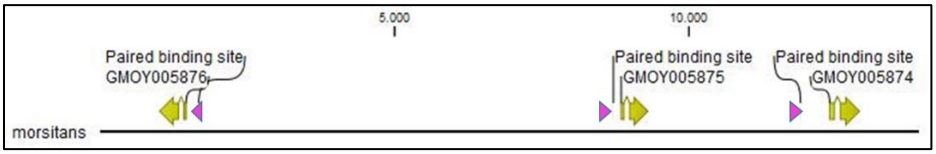


Fig. 2.24 Genomic organization of the three paralogs GMOY005874-5875-5876 in *G. m. morsitans*. Coding sequences (CDS) are represented in yellow and the paired binding site in pink. The paired consensus sequence is 5'-AAATTGC-3'. The illustration is not in scale.

Paired transcription factor: gene expression and tissue specificity

The *paired* gene (*prd*) was firstly described in the fruit fly, *Drosophila melanogaster*. *prd* belongs to the pair-rule gene family that participates in the determination of anterior–posterior axis during early embryogenesis (Baumgartner and Noll, 1990); it is necessary

for postembryonic viability and for male fertility through the regulation of accessory gland development (Xue *et al.*, 2001; Xue and Noll, 2000).
prd is the founding member of the Pax genes. They encode an evolutionarily conserved family of transcription factors with multiple DNA binding motifs and play key roles in animal development. All members of the Pax family are defined by the presence of a highly conserved 128-amino-acid paired domain (PD) (Noll, 1993).
The evolutionarily conserved domain consists of the N-terminal subdomain (PAI) and the C-terminal (RED) domains, each containing a helix– turn–helix motif capable of binding DNA (Xu *et al.*, 1999). In *Drosophila*, the RED domain has been recently found to be specifically required for the expression in the male accessory glands of Acp26Aa (an accessory gland specific peptide) and the sex peptide, and for the induction of female post-mating response (Li *et al.*, 2015) (**Fig. 2.25a**).
I investigated the tissue specificity of *paired* using previously developed MAG- and testes RNAseq datasets. RNAseq data are available from three male physiological states: teneral (i.e. within 24h post-eclosion), 3 days old virgin and 3 days old mated flies. In all the three groups, *paired* expression is significantly higher in accessory glands than in testes (**Fig. 2.25b**).

FlyAtlas Organ/Tissue Expression, larval vs. adult

Larval Expression Level	Tissue	Adult Expression Level
NA	Head	5.7
NA	Eye	2.025
NA	Brain	1.9
7.65	Central Nervous System	NA
NA	Thoracic-Abdominal Ganglion	3.5
NA	Crop	5.9
4.9	Midgut	3
4.8	Hindgut	3.6
4.6	Malpighian Tubules	5
no informative data	Fat Body	*no informative data*
9.9	Salivary Gland	9.8
NA	Heart	2.15
0.55	Trachea	NA
NA	VirginFemale Spermatheca	*no informative data*
NA	InseminatedFemale Spermatheca	*no informative data*
NA	Ovary	3.7
NA	Testis	25.2
NA	Male Accessory Gland	3050.9
1.525	Carcass	9.2

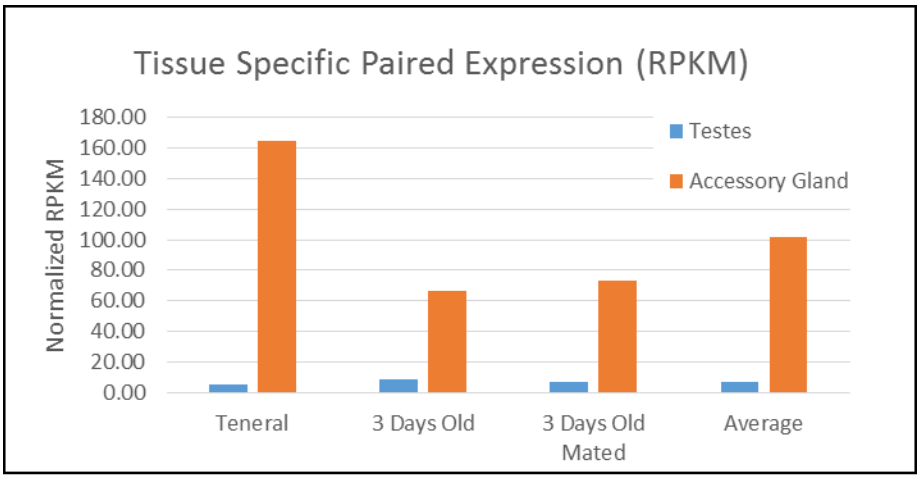


Fig. 2.25 a) *Paired* tissue expression in *D. melanogaster* (source: FlyBase). **b**) *Paired* tissue enrichment in *G. m. morsitans* male accessory glands; transcriptome data from male accessory glands and testes in teneral, 3 days old virgin, 3 days old mated flies respectively.

An RT-PCR analysis on RNA derived from five body compartments (male accessory glands, testes, fat body, gut and carcass) confirmed that paired is transcribed and enriched in the MAGs (**Fig. 2.26**). Samples have been obtained from 5 days old virgin *G. m. morsitans* males.

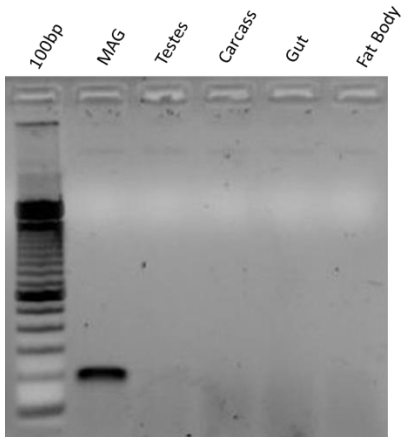


Fig. 2.26 Tissue distribution of *paired*. Paired (GMOY001037) RT-PCR primers: F: 5'CATCGGCAAGTTC-CATCATATTCG-3'; R: 5'-GCGGCGGTAGTGTTGTTG-3'.

To localize *paired* at the protein level, I performed a western blot analysis on paired protein. Taking advantage of commercially available rabbit polyclonal antibodies against the *Drosophila* paired protein, I probed six *G. m. morsitans* tissues for the presence of paired. The protein sequence recognized by the antibody corresponds to the N-terminal portion (1-50) of the *D. melanogaster* paired protein. As shown in the alignment between the two paired protein sequences (**Fig. 2.27**) this portion of protein is highly conserved between the two species.

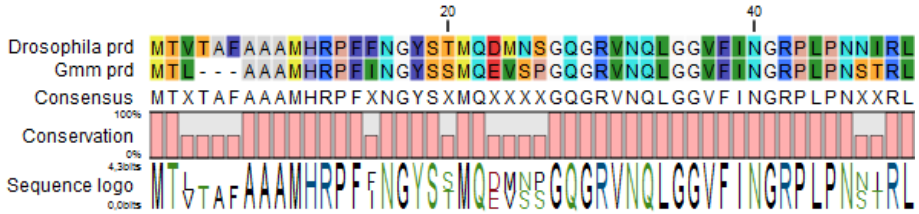


Fig. 2.27 Domain of interest of anti-paired antisera. Alignment between the region of interest in *D. melanogaster* and *G. m. morsitans*.

Six male tissues (accessory glands, testes, ejaculatory duct, fat body, gut and carcass) were dissected from 5 days old virgin males. Western blot analysis revealed protein only at one band on the membrane corresponding to the male accessory gland tissue. As expected by *in silico* prediction, the size of the detected band is of about 63 kDa, between 50 and 75 kDa marker bands (**Fig. 2.28**). The *Drosophila* paired gene is composed by two exons and the gene product is a protein made up of 613 amino acids with a molecular weight of 65 kDa.

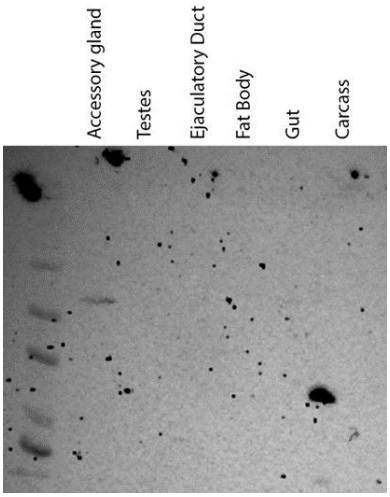


Fig. 2.28 Paired protein detection in male accessory glands through western blot analysis. Six tissues from 5 days old virgin males have been analyzed: MAG, testes, ejaculatory duct, fat body, gut and carcass.

Paired and GMOY005874 RNAi

dsRNA primers for *paired* and GMOY005874 have been designed (**Table 2.29**). Three days-old virgin males (fed the day before the injection) have been injected in the abdomen with 3 ug of paired dsRNAs. As a control, an equal number of flies have been injected with 3 ug of GFP dsRNAs. Preliminary results based on qPCR analyses performed on 7 paired-injected flies and 7 GFP injected flies are shown in **Fig. 2.30**.

Paired	F_taatacgactcactatagggGCCAACTCCGAATAGCTTGA R_taatacgactcactatagggCTTTGGAAGGTGAACCGAAA
GMOY005874	F_taatacgactcactatagggTGGCTGTAGCTGCAAACATC R_taatacgactcactatagggTGGATTACGGTCTCCTGCTT

Table 2.29 Paired and GMOY005874 primers for dsRNA synthesis. T7 tag peptide is included in all the sequences.

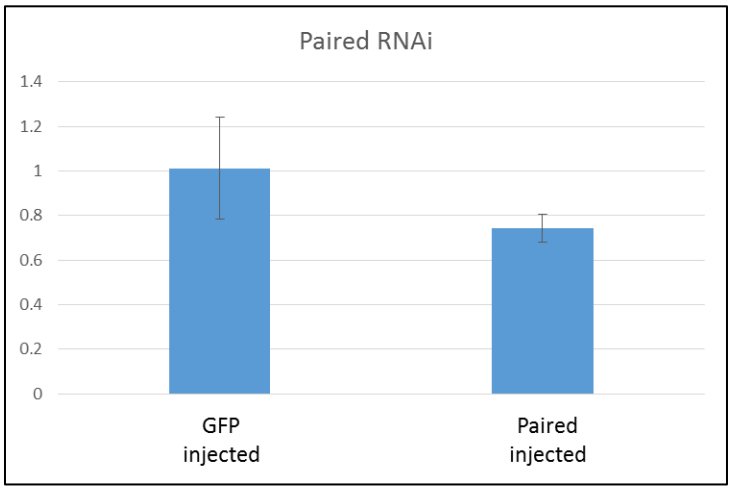


Fig. 2.30 RNAi experiment using paired dsRNA and GFP dsRNA as control. Preliminary results were obtained from seven paired injected flies and seven controls.

These results suggest the presence of a reduction in *paired* expression, but additional samples are needed, especially to minimize controls' variability. A reduction in gene GMOY005874 expression has been detected, indicating, at least for this gene, the potentiality of this transcription factor in controlling the expression of MAG genes.

2.5 Conclusions and Perspectives

The genetic diversity and the phylogenetic relationships among the *Glossina* species have been previously analyzed based on mitochondrial and nuclear ribosomal DNA sequences (Dyer *et al.*, 2008; Schneider *et al.*, 2013). However, species phylogenies derived from comparisons of single genes are rarely consistent with each other, often leading to the assessment of controversial phylogenetic relationships among the species. Completely sequenced genomes has been proven to allow the construction of a phylogeny that is less sensitive to inconsistencies such as horizontal gene transfer and unrecognized paralogy, and more representative of whole-genomes than are single-gene trees (Snel *et al.*, 1999).

For this reason, we took advantage of the release of six *Glossina* species (*G. morsitans, G. pallidipes, G. austeni, G. fuscipes, G. palpalis, G. brevipalpis*) and we used a genome-wide approach to disentangle their phylogenetic relationships. Moreover, this gave us the possibility to approach comparative genomic analyses among these species.

Comparative genomic analyses are particularly valuable to understand species-specific biology as well as recent evolutionary histories. Our genomic phylogenetic analysis suggests that Morsitans, Palpalis and Fusca groups all show a monophyletic origin, with *G. brevipalpis* (Fusca group) being the most differentiated species. It is noteworthy that the phylogenetic relationships estabilished by our analysis reflects the biological diversities, such as the geographical distribution, the eco/ethological constrains, the degrees of vectorial capacity and the mating behavior, characteristic of each of the *Glossina* group.

Based on this phylogeny, we used the molecular clock approach to derive the time of splitting among the *Glossina* groups and the included species. Orthologs genes derived from *Musca domestica* allowed us to date also the divergence between *Musca* and the *Glossina* genus at 209 Mya, in the Triassic period. Within the *Glossina* genus, we estimated the divergence time between the most distantly related *G. brevipalpis* and the Palpalis and Morsitans groups at ~ 54 Mya, i.e. Eocene period, while the divergence between the Morsitans and Palpalis groups has been dated at ~ 21 Mya, i.e. Miocene period. These data provide a good support to disentangling the evolutionary relationships within Diptera.

The availability of genome sequences from six *Glossina* species provides unprecedented opportunities to study the evolution of reproductive traits relevant for interpreting the species differentiation. Indeed, reproduction is one of the fundamental processes influencing several aspects of insect's life and it is often the target of control methods aiming at reducing the insect population size. Such control strategies are particularly effective in reproductive systems similar to the one displayed by tsetse flies (Vreysen *et al.*, 2000). In this frame, a key role is played by male reproductive genes that encode for

proteins involved in fertilization, such as seminal fluid proteins (SFPs). They are fundamental in influencing female post-mating physiological and behavioral changes as well as other processes involved in sexual competition and, thus, subjected to strong evolutionary pressures. Evolutionary and comparative genomics analyses gave us the opportunity to identify the presence of adaptive evolution in proteins of the reproductive system in *Glossina* species. We identified gene candidates for positive selection among those expressed in the male reproductive tissues, MAG and testes. Our analysis was based on 2,563 male genes, whose transcripts have been selected, in the *G. m. morsitans* transcriptome, on the basis of their higher expression in one tissue compared to the other. Orthology was assessed in the other five *Glossina* species and the genes were tested for the occurrence of positive selection using codon substitution models in PAML. d_N/d_S ratios, site-specific and branch-site models allowed us to identify potentially highly evolving sequences along the *Glossina* lineage, leading to the identification of a number of genes candidate for positive selection according to more than one test. It is noteworthy that positive selection is more likely to occur in some site of the gene rather than along all the sequence, and only along a single branch of the phylogeny. Interestingly, most of the identified genes in both tissues remained uncharacterized, since similarities with sequences described in reproductive tissues of different species are difficult to be detected, confirming their status of highly evolving sequences. Such considerations allow us to consider our "novel" genes putatively involved in species- or group-specific processes and, thus, of relevant interest for further evolutionary and applicative analyses. Considering the applicative potential of male reproductive genes and, in particular of fast evolving and species-specific male accessory gland sequences, we analyzed the regulatory elements responsible for the expression of the 24 most highly expressed accessory glands genes in *G. m. morsitans*.

Tissue-specific transcriptomic data revealed the presence of a transcription factor known as Paired. *In silico* analysis confirmed the transcription binding in the promoter region of most of the genes under study. The Paired transcription factor expression profile and protein localization have been demonstrated to be MAG-specific, thus suggesting a role in regulating the expression of MAG genes and, in turn, male reproduction. The important role in reproduction of this transcription factor is also confirmed by its sequence conservation among the *Glossina* species analyzed. From the applicative point of view, it can be an important target for the manipulation of male reproduction in tsetse flies. Under this perspective, we are performing RNAi experiments in order to evaluate the silencing effect of Paired on ejaculate composition, fertility and female post-mating behaviors.

Chapter 3: Identification and characterization of male-specific sequences in the Asian tiger mosquito *Aedes albopictus*

3.1 Background

3.1.1 Aedes albopictus and its impact on public health

The tiger mosquito *Aedes albopictus* (Skuse) (Diptera: Culicidae) is classified among the 100 most invasive species worldwide according to the Global Invasive Species Database (Lowe *et al.*, 2000). It is recognized as an important vector for many arboviruses such as those responsible for Dengue, Chikungunya, West Nile and Zika (Faraji *et al.*, 2014; Gratz, 2004), epizootic viruses (Lambrechts *et al.*, 2010; Paupy *et al.*, 2009) and nematodes (Giangaspero *et al.*, 2013; Pietrobelli, 2008). Because of the continuous evolutionary adaptation occurring between the host (the mosquito) and the virus, *Ae. albopictus* competence for various viruses is of particular concern. Moreover, this mosquito ability to feed both on humans (anthropophilic behavior) and animals (zoophilic behavior) exponentially increases the risk to propagate zoonotic pathogens from animal-to-animal and from animal-to-human (Paupy *et al.*, 2009).
Aedes albopictus impact on public health has strongly increased in the last 30-40 years, when the species spread from its native home-range area in East Asia and colonized parts of Europe, the Americas, the Caribbean, Africa and the Middle East. Following the spread of *Aedes aegypti* and *Culex pipiens*, the expansion of *Ae. albopictus* is the third example of diffusion favored by human activities. Nevertheless, biological traits such as physiological and ecological plasticity contributed to its rapid diffusion, increasing its potential to adapt and colonize new environments and forcing the coexistence with other vector species (Bonizzoni *et al.*, 2013).

3.1.2 Aedes albopictus biological traits

The life cycle of *Ae. albopictus* consists of four distinct stages: egg, larvae, pupa, and adult **(Fig. 3.1)**. Temperature and photoperiod both affect all the developmental stages of the mosquito, thus determining the number of generation per year, which ranges from 5-7, in temperate regions, to 17 in tropical areas (Hawley, 1988). An average of 40-90 eggs are produced by females after each blood meal, with about 300-350 eggs per lifetime (Hawley, 1988). Eggs can overwinter, through the so-called phenomenon of diapause. The aquatic phase can be completed in 5-6 days if the water temperature reaches 20-25°C, or up to three weeks at lower temperatures (14-18°C). During the larval stage, which include four instars (L1,L2,L3,L4), the size of the mosquito can increase from 2 to 10 mm in length. The L4 instar is followed by the pupal stage, when the mosquito stop feeding and the metamorphosis into the adult occurs.
Both males and females take nutrients from plant fluids, such as nectars and decaying fruits, and display a lifespan of about three weeks (Fang, 2010). Females also feed on

blood, which is necessary for eggs maturation. Although showing an anthropophilic behavior, *Ae. albopictus* can also feed on a variety of other hosts, including other mammals, birds and reptiles (Delatte *et al.*, 2010; Sullivan *et al.*, 1971). The preferential feeding on human blood is the main reason why *Ae. albopictus* is such an efficient disease vector and why it is so difficult to reduce its disease transmission potential. Females can be infected by pathogens taking the blood meal from an infected host: some pathogens are able to colonize the mosquito salivary glands, increasing their chances to be injected into a new host during the subsequent blood meal.

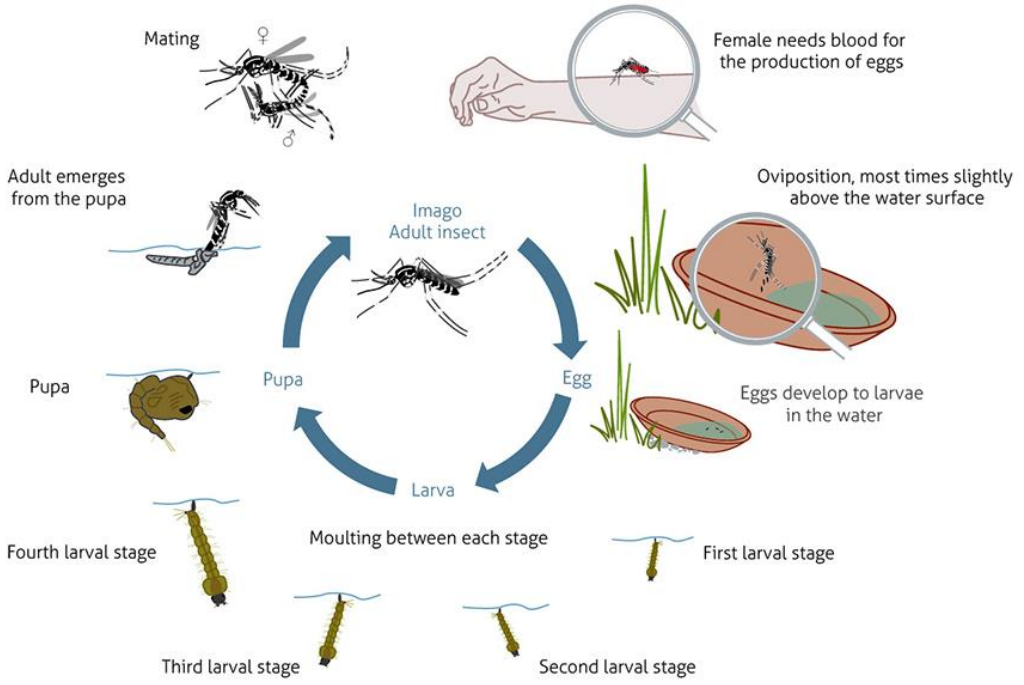


Fig. 3.1 The life cycle of the Asian tiger mosquito.

3.1.3 Aedes albopictus distribution

Aedes albopictus is native of the tropical forests of Southeast Asia. However, over the past 30 years, the increasing globalization of trade and human movements along with environmental changes facilitated the introduction and establishment of this invasive

mosquito in every continent except Antarctica (Enserink, 2008; Gasperi *et al.*, 2012; Lambrechts *et al.*, 2010; Medlock *et al.*, 2012).

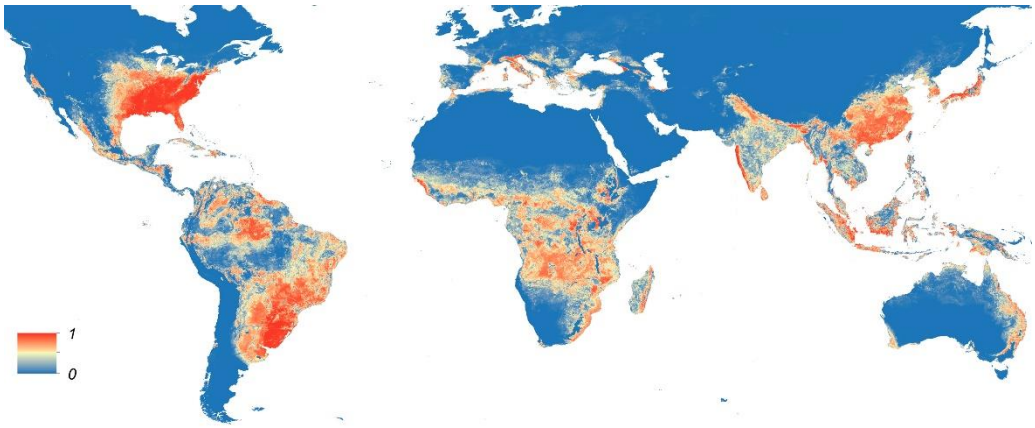


Fig. 3.2 Global map of the predicted distribution of *Ae. albopictus*. The map depicts the probability of occurrence (from 0 blue to 1 red) at a spatial resolution of 5 km × 5 km (Kraemer *et al.*, 2015).

Besides its ecological plasticity, the dispersal and spread of the Asian tiger mosquito is, indeed, a typical example of invasion mediated by human activities. The eggs of *Ae. albopictus* can withstand desiccation for many months and survive long transportation times, thus, being easily transported via the trade of used tyres and the importation of live plants (Bonizzoni *et al.*, 2013; Delatte *et al.*, 2011; Manni *et al.*, 2015). Used tires that collect rainwater provide suitable habitats for eggs and larvae survival, as well as suitable breeding environments for adult mosquitoes. Eggs resistance to desiccation, allow them to hatch, even after several months, in new geographic locations with the result of large numbers of mosquitoes arising in different places (Gasperi *et al.*, 2012).
The tiger mosquito colonization of Europe started in Albania in the late 1970s (Adhami and Reiter, 1998). The second invaded country was Italy in the 1990s (Sabatini *et al.*, 1990), and subsequently the species spread into the Mediterranean basin. Since its importation into Italy through Genoa (Enserink, 2008), *Ae. albopictus* has become established in most areas of the country, mostly in urban areas and at an altitude of less than 600m above sea level. In the Americas, the species was initially reported in Texas (USA) in 1985 and, since then, it spread northward and eastward, having now been reported in over 25 US states. In Latin America it was first reported in Brazil in 1986, and later in Mexico in 1988 (Eritja *et al.*, 2005). Reports of species importation were recorded also in New Zealand in 1994, in Bolivia, Cayman Islands, Costa Rica, Cuba,

Honduras, Dominican Republic, El Salvador, Guatemala, and Panama in 1995, in Columbia in 1997, and in Argentina in 1998 (Eritja *et al.*, 2005) **(Fig. 3.2)**.
Because of its rapid adaptation to new environments, once it is established in a region, the eradication process of the mosquito can be very difficult and constant surveillance and appropriate control strategies are required (Holder *et al.*, 2010).

3.1.4 Aedes albopictus control methods

Currently, there are neither effective vaccines nor treatments available against mosquito-borne diseases. The ongoing strategies rely mainly on vector population control programs. In order to impair the ability of a mosquito to serve as a disease vector, it is necessary to interfere with biological aspects relevant to its vectorial capacity, such as reproduction, olfaction, longevity and immunity.
Such vector control strategies include:

i) Environmental modifications, based on removing/covering temporary water containers (Fonseca *et al.*, 2013);
ii) Mechanical methods, such as odor baits that help reducing adult populations of mosquitoes targeting gravid or host-seeking females (Mackay *et al.*, 2013);
iii) Chemical compounds, like insect growth regulators used as chemical larvicides, and pyrethroids used as chemical adulticides, commonly applied because of their low toxicity towards mammals (Vontas *et al.*, 2012);
iv) Biological measures, for example using entomopathogenic fungi, copepods, or *Bacillus thuringiensis* and plants essential oils as microbial and botanical larvicides respectively. Another emerging strategy takes advantage of the *Wolbachia*-induced cytoplasmic incompatibility that results in embryonic mortality in matings between insects of the same species with different *Wolbachia* infection status (Bourtzis *et al.*, 1998; Minard *et al.*, 2013; Valiente Moro *et al.*, 2013; Zouache *et al.*, 2009);
v) Genetic-based strategies, such as the use of sterile mosquitoes (within Sterile Insect Technique - SIT - programmes) and related techniques. Preliminary studies confirmed the feasibility of using SIT to suppress natural populations of *Ae. albopictus* in Europe, supporting the development of novel genetic-based approaches (Bellini *et al.*, 2007; Bellini *et al.*, 2013; Madakacherry *et al.*, 2014) **(Fig. 3.3)**.

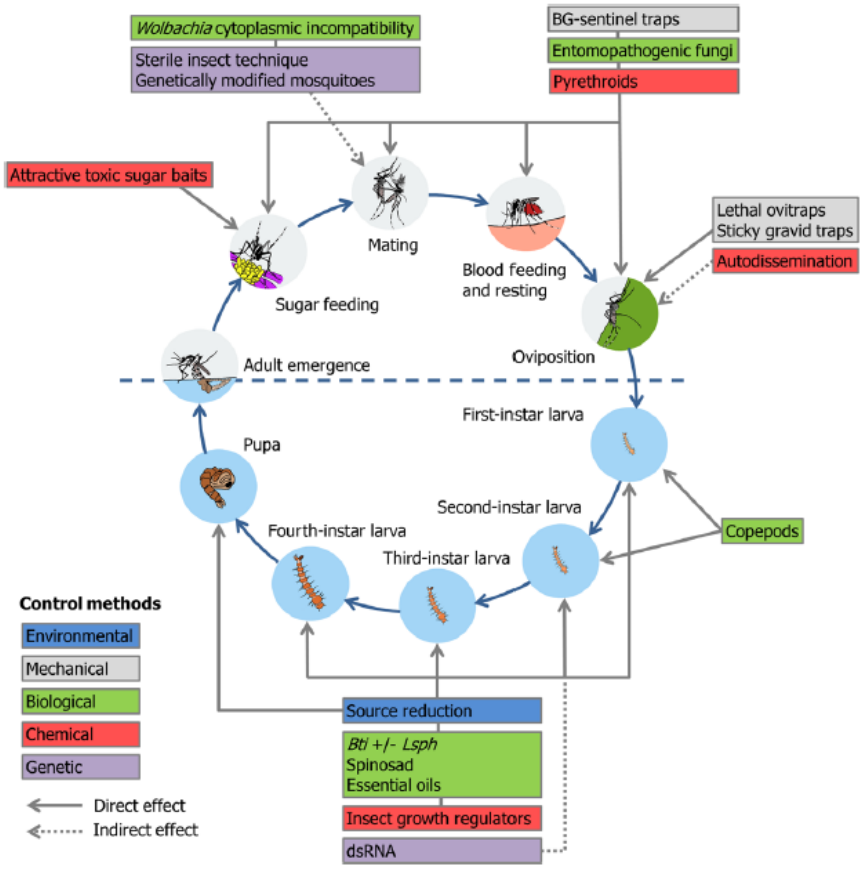


Fig. 3.3 Control methods available against *Aedes spp.* (Bti: *Bacillus thuringiensis var. Israelensis*; Lsph: *Lysinibacillus sphaericus*; dsRNA: double-stranded RNA) (Baldacchino *et al.*, 2015).

3.1.5 Features of the Aedes albopictus genome

The recent release of two genome assemblies for two strains of the Asian tiger mosquito, respectively, provided a considerable contribution to advance studies in this species. The two sequenced strains are the *Foshan* laboratory strain, originally sampled around 30 years ago from the native area of this mosquito (Chen *et al.*, 2015), and the *Fellini* laboratory strain originally sampled from an Italian region (Dritsou *et al.*, 2015).
Although both genome projects faced difficulties in obtaining a high quality assembly, they are providing valuable information about its genomic structure and its pattern of

evolution. Comparative analyses with other insect genomes is revealing several interesting features, one of these is the genome size. So far, the *Aedes* genomes are the largest among mosquitoes, ranging from 1,376 mega base pairs (Mb) of *Ae. aegypti* to 1,967 Mb and 940 Mb of *Ae. albopictus Foshan* and *Fellini* strain, respectively (Chen *et al.*, 2015; Dritsou *et al.*, 2015).

The huge difference in genome size in respect to other mosquito species, like *An. gambiae*, which is 264 Mb, is probably due to the large amount of transposable elements (TEs) and repetitive elements found in *Aedes* genomes (Holt *et al.*, 2002; Severson and Behura, 2012). *Aedes* mosquitoes also show differences in the karyotype structure when compared to *An. gambiae*. The mosquito family (Diptera: Culicidae) is indeed divided into two subfamilies: *Anophelinae* and *Culicinae* in heteromorphic and homomorphic sex chromosomes have evolved, respectively. In *Aedes* and *Culex* mosquitoes (*Culicinae*), male development is initiated by a dominant male-determining locus (M-locus) located on a homomorphic sex-determining chromosome (Newton *et al.*, 1974). On the other hand, *Anopheles* mosquitoes have fully differentiated heteromorphic sex chromosomes where the male-determining locus resides on the non-recombining Y chromosome (Marin and Baker, 1998).

The mitotic chromosome complement of *Ae. albopictus*, as well as *Ae. aegypti*, consists of three pairs of metacentric chromosomes (Rai, 1963), with the sex determination locus associated to the smallest homomorphic autosome 1 (McClelland, 1962).

Also in species with well-defined heteromorphic sex chromosomes such as *Drosophila* and *Anopheles*, Y chromosome sequences have proven resistant to traditional methods of sequence assembly because of its repetitive nature, thus being rarely annotated in published genomes of non-model organisms (Bernardo Carvalho *et al.*, 2009). This is the reason why, despite the publication of the *Ae. albopictus* genome, sequences from the M-locus are still uncharacterized.

3.2 Aim of the work

During my PhD studies, I have been involved in two projects targeting *Ae. albopictus*, with the aim of i) achieve a better understanding of the mosquito genomic features, and ii) provide the genetic basis for improving the vector control strategies against this important vector. The first one was in the framework of the *Ae. albopictus* genome project (Italian strain), in which I have been involved in the determination of the genome size of the *Fellini* strain and in the manual curation and annotation of Odorant Binding Proteins (OBPs) and Odorant Receptor (OR) genes.
The second project was focused on the identification of male-specific sequences suitable for the development of a genetic sexing strain to be used in SIT campaigns. Using cytogenetic and molecular approaches, I characterized the chromosome complement of the *Fellini* strain to shed light on the *Ae. albopictus* male determining factor.

3.3 Materials and Methods

Mosquito strain

This study was performed using mosquitoes of the *Fellini* strain provided by the Laboratory of Romeo Bellini (CAA, Crevalcore - Italy).

Genome size estimation

A series of ten-fold dilutions of a known concentration of a linearized plasmid containing the amplicon sequence corresponding to part of the *Ae. albopictus* G6PDH gene identified from RNA-seq data (Genbank/EMBL Accession Nr KT279821) was monitored in real-time using a MiniOpticon (Bio-Rad) machine. As the length of the amplicon was known, the concentrations of the dilutions could be calculated as copies per microlitre. Comparison between the calibration curve of the CT values derived from the standard dilution series and those derived from the same G6PDH amplicon in genomic *Ae. albopictus* DNA samples of known concentration, allowed the absolute number of target gene copies to be calculated, and hence the genome size (Dritsou *et al.*, 2015).

Gene annotation

TBLASTN searches were performed using as queries the aminoacid sequences of the 111 Odorant binding proteins (OBPs) and 112 Odorant receptors (ORs) of *Ae. aegypti* (Bohbot *et al.*, 2014; Manoharan *et al.*, 2013). The contigs and scaffolds that produced hits (cut-off e-value of $1E^{-10}$) were used to interrogate, using BLASTX, local protein databases of the *Ae. aegypti* OBPs and ORs. Multiple contigs and scaffolds with hits to an OBP or OR were assembled using CAP3. GeneWise was used to obtain gene model predictions based on homology with *Ae. aegypti* OBP and OR proteins (Dritsou *et al.*, 2015).

Chromosome preparations

Chromosome preparations were obtained from imaginal discs (IDs) of *Ae. albopictus* 4th instar larvae of as described in (Sharakhova *et al.*, 2011) for *Ae. aegypti*. Larvae were placed in ice for immobilization. Larval heads were removed in cold hypotonic solution (0.5% sodium citrate) and the thorax dissected in order to extract the IDs. After 10 min. in hypotonic solution, a drop of fixative solution (ethanol/acetic acid in 3∶1 ratio) was applied. 50% propionic acid was added for maceration, and IDs were squashed under a

cover slip (22×22 mm). Chromosomes were DAPI-stained and the slides were mounted using the VECTASHIELD mounting medium (Vector Laboratories). Chromosomes were screened under an epifluorescence Zeiss Axioplan microscope and images were captured using an Olympus DP70 digital camera.

Representational difference analysis (RDA)

The Representational Difference Analysis (RDA) (Chang, 2002) was applied to identify male-specific or male-enriched sequences in *Ae. albopictus*.
To this aim, I extracted genomic DNA from seven male and seven female adult mosquitoes (*Fellini* strain). The genomic DNA was then digested with two four-cutter endonucleases, DpnII (recognition site: 5'-↓GATC-3', 3'-CTAG↑-5') and MseI (recognition site: 5'-T↓TAA-3', 3'-AAT↑T-5'), resulting in the DpnII and MseI male and female Representations. To generate the "tester", R adaptors were replaced by J adaptors only in the male Representation.
As a second step, we performed the subtractive hybridization using a 1/100 tester/driver ratio. Then, male-specific DNA was re-amplified to generate the Differential Product 1 (DP1) using J adaptors as primers. Following this procedure, we performed a second and a third round of subtractive hybridizations: the DP2 was obtained using a new DNA tester generated by substituting the DP1 J adaptors for N adaptors and using two different tester/driver ratios 1/100 and 1/800, while the DP3 was obtained substituting the DP2 N adaptors for J adaptors and using three different tester/driver ratio 1/100, 1/400 and 1/800 (Fig.2).

3.4 Results and Discussion

The Aedes albopictus genome project: Genome size determination

The draft genome of the Italian strain was determined using a standard NGS-Illumina (next generation sequencing) approach and derived from L4 larvae of an isofemale line developed in lab rearing conditions.
The genome size, which describes the DNA content in picograms per haploid genome (C-value). Real-time PCR-based method relies on the absolute quantification of a single copy gene (Higuchi *et al.*, 1993). The C-value can then be easily calculated by dividing the mass of sample DNA by the copy number determined for single-copy genes, where the amount of sample used is determined by UV absorption spectrometry. The single copy gene utilized for *A. albopictus* was the G6PDH gene. A series of ten-fold dilutions of known concentration of the linearized plasmid containing the amplicon sequence corresponding to the gene was monitored in real-time PCR and, as the length of the amplicon was known, the concentrations of the dilutions has been calculated as copies per µl. Using genomic *Ae. albopictus* DNA samples of know concentration, the C_t values calibration curves derived from the standard dilution series was compared to those derived from the same G6PDH amplicon, allowing the absolute quantification of target gene copies, thus the genome size. For further details, please see the insert "*Aedes albopictus* genome paper".
I estimated the genome size of the Fellini strain at 0.94 Gbp. This value is relatively close to the estimate genome size of the assembly before scaffolding, measuring 1.12 Gb.
The size of the assembled genome is comparable to that of the other *Aedes* genome available, i.e. *Ae. aegypti*. Interestingly, the *Aedes* genomes (*Ae. albopictus* and *Ae. aegypti*) are the largest of any mosquito species sequenced to date **(Fig. 3.4)**, which vary from 278 Mb for *An. gambiae* to 540 Mb and 1376 Mb for *Culex quinquefasciatus* and *Ae. aegypti*, respectively (Arensburger *et al.*, 2010; Nene *et al.*, 2007).
Previous studies already emphasized the particularly large size of the *Ae. albopictus* genome compared to other insect species, with genome size variation observed also among different populations of the same species and that can be attributed to changes in the amounts and organization of repetitive DNA (Black and Rai, 1988) and that might explain the extraordinary adaptive capacity of this species (Rai and Black, 1999; Severson and Behura, 2012). A huge variation in the estimation of the genome size is observed also in the comparison between the *Fellini* strain and the *Foshan* strain (1967 Mb (Chen *et al.*, 2015)). However, a proportion of this variation might be attributed to the different strategies adopted for the genome size determination in the two different consortia.

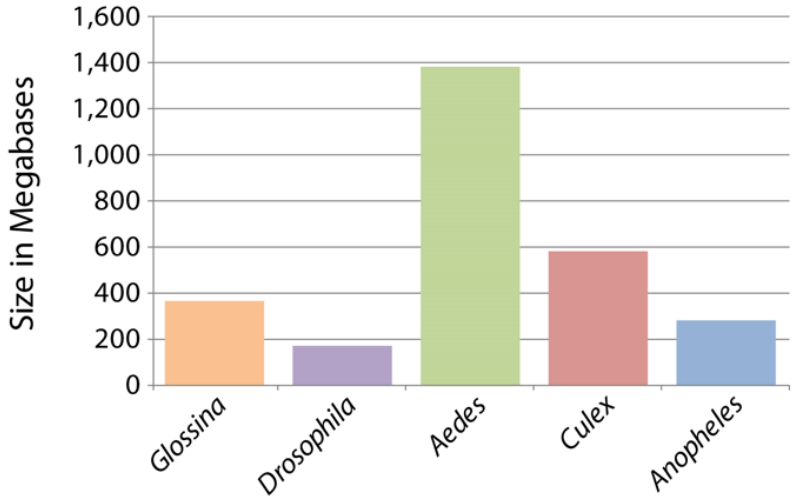


Fig. 3.4 Dipteran genome size. From left to right: *G. m. morsitans* 389 Mb, *D. melanogaster* 175 Mb, *Ae. aegypti* 1376 Mb, *C. quinquefasciatus* 540 Mb, *A. gambiae* 278 Mb (Aksoy and IGGI, 2010; Arensburger *et al.*, 2010; Ellis *et al.*, 2014; Holt *et al.*, 2002; Nene *et al.*, 2007) (Source: Geoffrey Attardo).

Annotation of olfactory genes

The genome revealed the expansion of gene families involved in immunity, olfaction, sex determination, diapause and insecticide resistance, contributing to the plasticity of its genome as well as abundant repetitive DNA classes (Chen *et al.*, 2015; Dritsou *et al.*, 2015).

Ae. albopictus, like other mosquito species, relies on olfactory cues for host-seeking, blood-feeding and the detection of oviposition sites. In the frame of the genome project, I identified and annotated OBP and OR genes of *Ae. albopictus*.

The protein sequences of 111 Odorant binding proteins (OBPs) and 112 Odorant receptors (ORs) of *Ae. aegypti* (Bohbot *et al.*, 2007; Bohbot *et al.*, 2014; Manoharan *et al.*, 2013) were used as queries for TBLASTN and BLASTX searches. Orthologues were identified for 110 and 98 *A. aegypti* OBPs and ORs, respectively.

Due to the extensive fragmentation of many of the genes, it was not possible to determine the actual number of OBPs and ORs present in the genome. Fifty-four of the *Ae. albopictus* OBP sequences were complete, although the sequence often spanned more than one contig. Partial sequences were obtained for 56 OBP orthologues, for 25 of which it was not possible to identify the region corresponding to the signal peptide,

typical features of OBPs, perhaps due to the higher variability of these regions compared to the rest of the gene.

Thirty-five of the OR sequences were complete (although two contained frame-shifts), again frequently spanning several different contigs; six of the complete ORs were identified in single contigs, whereas 12 were identified in single scaffolds. Twenty-one of the incomplete OR sequences lacked one or both termini and encoded conceptual polypeptides between 305 and 411 amino acids in length. Orthologues of 14 *Ae. aegypti* ORs (namely OR3, 5, 14, 16, 17, 28, 65, 96, 98, 103, 107, 114, 124 and 127) appear to have been lost in *Ae. albopictus*, however, a number of gene lineage expansions are clearly evident, for example, AaegOR60, AaegOR61 and AaegOR63 have at least two homologues in *Ae. albopictus*, whereas AaegOR70 has at least three. Further analyses will help completing the *Ae. albopictus* OBP and OR repertoires and will open the way for the functional characterization of these proteins with the aim of developing novel and effective attractants and repellents for monitoring and control purposes (extracted from Dritzou *et al*., 2015).

Building a genetic sexing strain for the tiger mosquito

3.4.1 Cytogenetic characterization of the mitotic chromosome complement of Aedes albopictus

For the SIT technique to be successful, several factors must be taken into account: i) the species should be easily colonized and mass-reared at a reasonable cost; ii) for species where females can transmit disease, it is necessary that they are removed from the release insects; iii) the released males need to be as competitive as wild males, and iv) the target population should be reduced in size (using other techniques) before the release of the sterile males (Yamada *et al.*, 2015).
In order to identify male-specific sequences suitable for the development of a genetic sexing strain for *Ae. albopictus*, we used cytogenetic and molecular approaches. The idea is to use molecular markers, obtained by molecular techniques such as the Representational Difference Analysis, that can be associated to the banding pattern observed in mitotic chromosomes prepared from male and female mosquitoes.
Successful physical mapping for any organism relies on a robust source of high-quality, easily obtainable chromosome preparations. Because of the abundance of repetitive elements in the genome, the polytene chromosomes of *Aedes* mosquitoes are unsuitable for cytogenetic studies. At the same time, the large size of the genome make mitotic chromosomes large and easily identifiable (Sharakhova *et al.*, 2011).
In this context, using the *Fellini* strain provided by Romeo Bellini Laboratory (CAA, Crevalcore), we characterized the mitotic karyotype of *Ae. albopictus*. Following a protocol established for *Ae. aegypti*, we used imaginal discs (IDs) of 4th instar larva as a source of mitotic chromosomes **(Fig. 3.5)**. In order to identify unique banding patterns that characterize the three different chromosomes, slides were stained with both DAPI and GIEMSA stains **(Fig. 3.6)**.
Ae. albopictus has three pairs of chromosomes as all the mosquito species (2n=6), with the exception of *Chagasia bathana* (2n=8) of the subfamily Anophelinae.
Unlike Anophelinae, sex chromosomes of *Aedes* and *Culex* are homomorphic. In accordance with *Ae. aegypti* linkage groups, chromosome 1 is the shortest, chromosome 2 the longest, and chromosome 3 the one of medium length (Sharakhova *et al.*, 2011). In *Ae. aegypti*, the M-locus has been mapped to chromosome 1, band 1q21; nevertheless, in other Culicinae mosquitoes, it has been found in non-canonical locations, indicating either translocation or turnover of the sex-determining gene (Ferdig *et al.*, 1998; Venkatesan *et al.*, 2009).

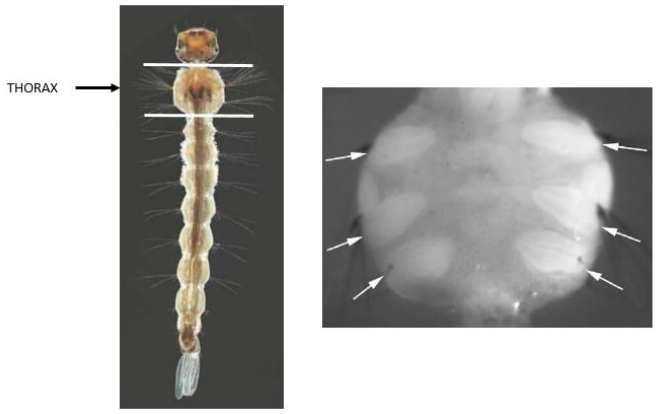


Fig. 3.5 On the left, an example of *Ae. aegypti* larva. On the right, IDs are shown. The location of the IDs under the cuticle in thoracic segments of 4th instar larvae are indicated by the arrows (Sharakhova *et al.*, 2011).

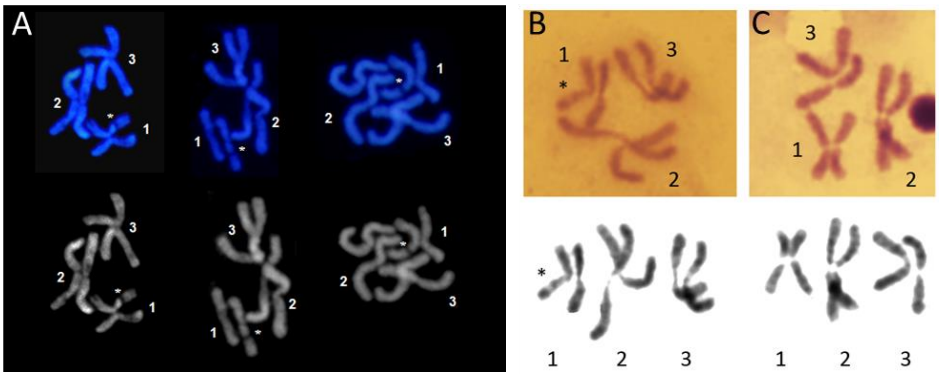


Fig. 3.6 Mitotic chromosome complement of individual *Ae. albopictus* larvae. From top to bottom: DAPI (A) and Giemsa stained chromosomes (B,C).

In order to identify cytological differences between male- and female-derived *Ae. albopictus* chromosomes, we prepared 40 mitotic chromosome slides from individual larvae. Our cytogenetic results, using DAPI and GIEMSA stains, show two different banding patterns in mitotic chromosome preparations **(Fig. 3.6)**. One of the two banding pattern observed comprise a GC-rich band on one chromatid on the long arm of chromosome 1 and not in the other (putative male pattern, **Fig. 3.6 A-B**). The same band

is instead not visible in neither of the two chromatids of mitotic chromosome 1, referring to the second pattern observed (putative female pattern, **Fig. 3.6 C**). Since *Aedes* larvae do not exhibit morphological dimorphism, it was not possible to know whether the slides refer to a female or a male karyotype. However, using only the thoracic part of the larva give us the possibility to further analyze the same sample, obtaining genomic DNA or RNA. The identification of male-specific DNA markers could be used to know *a priori* the sex of the larvae is one of the main objective of this research project.

3.4.2 Molecular characterization of Aedes albopictus sex chromosomes through the identification of male-linked sequences

Characterization of male-specific sequences may play a fundamental role in the elucidation of the molecular nature of the male determining factor and could shed light on the evolution of sex chromosomes within this dipteran family. In order to identify male-specific or male-enriched sequences, we used a Representational Difference Analysis approach (RDA), previously used in our laboratory to identify male-linked sequences in *Bactrocera oleae* (Gabrieli *et al.*, 2011), to be associated to the banding pattern detected by cytogenetic analysis.
The RDA is a reliable way to detect differences between two similar complex genomes. With this method, series of subtractive hybridizations are performed using two representations (amplicons) of the two genomes to be compared. Reduction of the complexity of the genome is essential for efficient subtractive hybridization, and excessive reduction of the complexity will result in the loss of target (Ushijima *et al.*, 1997). As a result, target restriction fragments initially amplified from one genome, the "tester", that are not amplified from the other genome, the "driver", are further amplified approximately 10^6-fold in excess of common fragments (Baldocchi and Flaherty, 1997). We were interested in male specific sequences so we used the female genomic DNA as "driver" and the male genomic DNA as "tester".
As a first step, the RDA protocol requires the generation of male and female Representations by PCR of adaptors-ligated genomic fragments. We used three pairs of adaptors: the R series, used in the preparation of the amplicon representation, and the J and N series used for hybridization-amplifications **(Table 3.7)**.

Primer	Sequence (5'-3')
R-DpnII 24	AGCACTCTCCAGCCTCTCACCGCA
R-DpnII 12	GATCTGCGGTGAGA
J-DpnII 24	ACCGACGTCGACTATCCATGAACA
J-DpnII 12	GATCTGTTCATGGA
N-DpnII 24	AGGCAACTGTGCTACTCGAGGGAA
N-DpnII 12	GATCTTCCCTCGAG
R-MseI 24	AGCACTCTCCAGCCTCTCACCGCAT
R-MseI 12	TAATGCGGTGAG
J-MseI 24	ACCGACGTCGACTATCCATGAACAT
J-MseI 12	TAATGTTCATGG
N-MseI 24	AGGCAACTGTGCTACTCGAGGGAAT
N-MseI 12	TAATTCCCTCGA

Table 3.7 Primers used for Representation Difference analysis (RDA). The R-series primers were used in the preparation of the amplicon representation; the J- and N-series primers were used for hybridization-amplifications.

We observed a putative male-specific band only in samples digested with the MseI enzyme (1/100 tester/driver ratio) and not in samples digested with the DpnII enzyme **(Fig. 3.8)**.

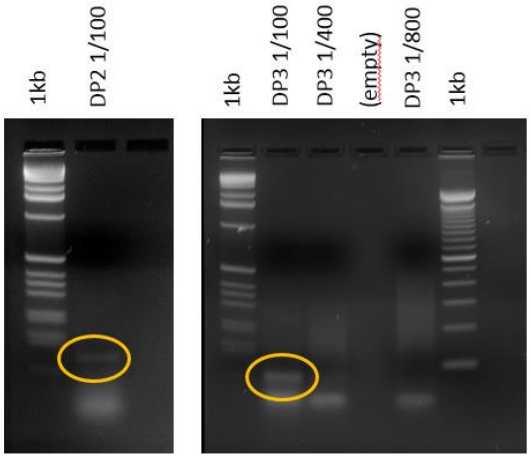


Fig. 3.8 Gel electrophoresis showing MseI-DP2 using a 1/100 tester/driver ratio and MseI-DP3 using 1/100, 1/400 and 1/800 ratios.

The putative male-specific band **(Fig. 3.8)** present in the DP3 and referring to the 1/100 tester/driver ratio was gel eluted and cloned, and ten positive clones were selected and sequenced **(Fig. 3.9)**.

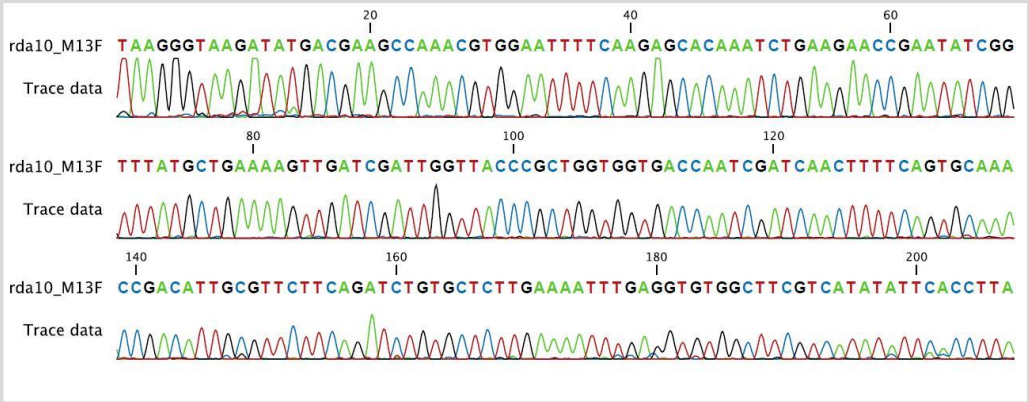


Fig. 3.9 Sequencing result of the rda10_M13F clone.

Sequences were subjected to BLAST analyses and best hits are reported in **table 3.10**. Sequencing results revealed six out of ten different sequences: rda1, rda2, rda3, rda7, rda9 and rda10. rda4 and rda5 sequences were not suitable because of a poor sequencing results and rda6 and rda8 sequences were identical to rda1 and rda10 sequences, respectively. Putative male-specific/-enriched sequences on genomic male/female DNA have been tested on both male and female genomic DNA but the results showed no male-specificity.
Thus, alternative approaches such as chromosome quotient (Hall *et al.*, 2013) can be performed in the future.

Clone	Clone length (bp)	BLASTn best hits	Accession number	e-value	Score	Maximum identity (%)
rda1	255	Solanum lycopersicum chromosome ch12, complete genome	HG975524.1	0.004	52	97
rda2	206	Salmo salar clone BAC S0142M23, complete sequence	EU816603.1	3.00E-28	134	84
rda3	275	Aedes albopictus microsatellite alb212 sequence	JQ886085.1	3.00E-11	79	82
rda7	207	Aedes aegypti clone p515, Pony-Aa-A14 MITE repeat region	AF208671.1	3.00E-28	134	76
rda9	256	Drosophila ananassae retrotransposon tom, 3' LTR	M37656.1	0.17	46	89
rda10	207	Aedes aegypti, clone XX-97O18, complete sequence	AC150257.12	6.00E-50	206	82

Table 3.10 BLAST analysis of the RDA clones identified.

3.5 Conclusions and Perspectives

Aedes albopictus is a worldwide invasive species. It represents a serious threat for public health, and therapeutic treatments as well as effective control strategies against vector-borne diseases are still missing.

The two genome projects independently designed by two different consortia are good examples of the efforts of the scientific community towards the study of this important mosquito species. Genomic data are providing novel insights into the biology of the species, allowing the identification of molecular targets to be further explored to enhance the vector control strategies. *Aedes albopictus* genome revealed the presence of abundant repetitive DNA classes together with the expansion of gene families involved in immunity, olfaction, sex determination, diapause and insecticide resistance. In particular, I found its genome to be 0.94 Gbp in size, and I begin characterizing the set of olfactory genes (OBPs and ORs), which play fundamental roles in host-seeking behavior. However, due to the extensive fragmentation of many of the genes, it was not possible to determine the actual number of OBPs and ORs present in the genome.

Together with the large genome and the abundance of repetitive elements, another relevant feature of this species is represented by its homomorphic sex chromosomes.

The identification of male-specific sequences aimed at improving/developing SIT-based control strategies are particularly challenging. Probably due to the repetitive nature of the *Ae. albopictus* genome, the RDA approach, which relies on subtractive hybridizations of male and female genomic DNA, failed to retrieve library-specific short sequences and it is not suitable in such cases. Recently, six Y chromosome genes have been identified in *Anopheles* species (Hall *et al.*, 2013) using the so-called CQ, Chromosome Quotient approach. This technique can be exploited for the characterization of the M-factor also in other insect species, where males and females are sequenced separately. Probably, it will be the best approach to identify male-specific sequences also in *Ae. albopictus*, as it has been demonstrated to work on species with highly fragmented genome assemblies. The characterization of male-specific DNA markers will provide unique markers for population and phylogenetic analysis, and opportunities for developing genetic sexing strains reliable for SIT-based applications.

Chapter 4: Genomic insights into the reproductive biology of the medfly, *Ceratitis capitata*

4.1 Background

4.1.1 The Mediterranean fruit fly, Ceratitis capitata: economic importance

The Mediterranean fruit fly (medfly) *Ceratitis capitata* is a highly invasive agricultural Dipteran pest belonging to the Tephritidae family (Maddison and Bartlett, 1989). From its sub-Saharan Africa home range, this species quickly spread to Asia, Europe, South and Central America, becoming a global invader (Malacrida *et al.*, 2007; Malacrida *et al.*, 1992).
The success of medfly invasion worldwide, is due to several biological features of this species, including:
i) its highly polyphagous behavior (i.e. the ability to infest a wide range of different plant species): the medfly can cause significant damages to more than 300 species of fruits, ornamental plants and vegetables, including citrus, subtropical fruits, grapes and pomegranates (Liquido *et al.*, 1991).
ii) its ability to tolerate cooler climates: medflies can overwinter as adults, as pupae in the ground, and also as eggs and larvae in the fruits; therefore the species can be easily transported across continents through international trades (Harris and Carey, 1989; Harris and Lee, 1989; Scolari *et al.*, 2012; Vera *et al.*, 2002). For these reasons, quarantine restrictions are in place in many countries, including USA and Japan, on fruits imported from infested areas. The economic loss due to medfly is estimated around US$192 million/year.
iii) its extremely high reproductive potential: females can produce 300-800 eggs during their two-months lifespan. In addition, in temperate climates this species can reach six/seven generations per year, and even more than twelve under tropical weather conditions (i.e. multivoltine life cycle) (Thomas *et al.*, 2001).

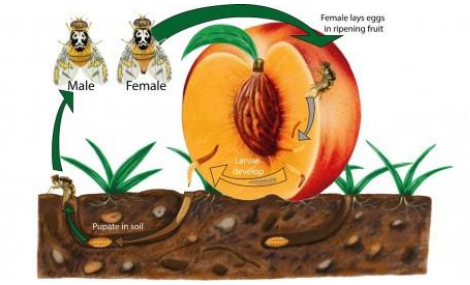


Fig. 4.1 Medfly life cycle (http://www.oxitec.com/agriculture/our-products/medfly/).

The damage to crops begins when a female lays her eggs beneath the skin of a host fruit (approximately 1mm). The hole produced by the female ovipositor allows easy access to many microorganisms, such as *Rhizopus stolonifera* and *Phytophthora spp* (Cayol *et al.*, 1994), responsible for premature rotting of the fruit. After two-three days, the eggs hatch and the larvae start to feed and develop in the fruit's pulp, making it inedible. After three larval stages lasting six to ten days, the larvae "jump" out from the fruit to pupate in the soil **(Fig. 4.1)**.

4.1.2 Ceratitis capitata distribution

In the last decades, population genetic studies clarified the medfly colonization processes. The most supported hypothesis set the origin of the medfly diffusion in the sub-Saharan East Africa (Hagen *et al.*, 1981), and its subsequent spread to adjoining African regions and towards the Mediterranean coasts. The genetic differentiation between these populations is significantly related with the geographic distance, suggesting a hierarchical migration structure through Spain and along the Mediterranean basin to the East (Gomulski *et al.*, 1998; Malacrida *et al.*, 1998). The strong adaptability showed by this species to the different Mediterranean ecological condition can easily explain why the medfly is considered a global invader (Malacrida *et al.*, 2007).
The species is now distributed in most tropical and temperate regions of the world and constantly threatens to invade or reinvade new areas (Carey, 1991) **(Fig. 4.2)**.
Indeed, besides Mediterranean Europe and the Middle East, the medfly is present in most African regions including Indian Ocean islands, South and Central America, western Australia and the Pacific region (Szyniszewska and Tatem, 2014). The species has been established for about a century in Hawaii and, although extensive eradication efforts have been made, it is repeatedly detected into Florida and California (Carey, 1996; Jang, 2007). The cost of each previous medfly incursions in California is estimated to range from US$300,000 to US$200,000 (APHIS, 1992). However, despite more than 250 emergency eradication projects have been directed against medfly in California and 100% success have been declared for each program, more than nine tephritid species are still established in the area (Papadopoulos *et al.*, 2013).

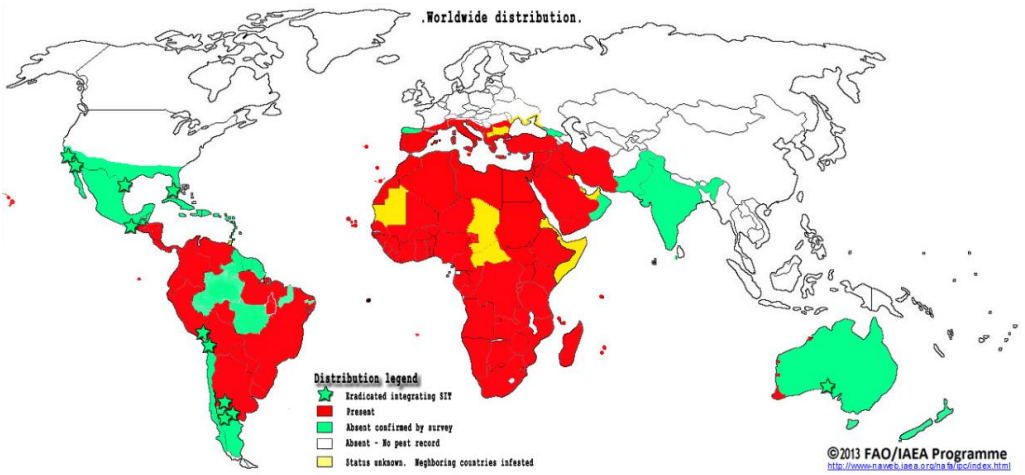


Fig. 4.2 Worldwide distribution of the medfly (source: IAEA, http://nucleus.iaea.org).

4.1.3 Ceratitis capitata control methods and the SIT

At present, the control of medfly populations in the wild mainly relies on the use of chemicals (i.e. insecticides or attractants) that have a broad spectrum of action and affect also non-target species **(Fig. 4.3)**. Moreover, insecticide resistance is a constantly emerging threat. The development of environmentally friendly and cheap methods to control this pest would be of great benefit (Kapongo, 2007).
Because of the highly invasiveness and the difficulties faced in the eradication process, medfly is the most thoroughly studied tephritid fruit fly species at the genetic and molecular levels and it has become a model for the analysis of other fruit flies invasions. *Ceratitis capitata* was the first target species for a large-scale fruit fly AW-IPM (Area-Wide Integrated Pest Management) program based on SIT (Hendrichs *et al.*, 2002; Klassen, 2005), a method that for more than 25 years kept Mexico, USA and half of Guatemala free of the pest. This technique consists in the release in the field of sterile males, which can compete with wild males to mate with females. Females that mate with a sterile male produce no offspring, thus reducing the next generation's population. Only sterile males contribute to population suppression and the success of SIT depends on the fitness and competitiveness of sterilized males. Generally, X-ray sterilized males show reduced mating competitiveness and low mating success compared to wild males (Lance *et al.*, 2000). Despite the fact that this technique has proved to be useful in eradicating medflies populations in certain areas, this strategy is expensive and relies on insect mass rearing in large biofactories. Moreover, females of *C. capitata* can remate, thus affecting SIT effectiveness. Currently, the SIT is the most promising alternative to pesticides and

several efforts have been made to improve the efficacy and decrease the costs of this technique.

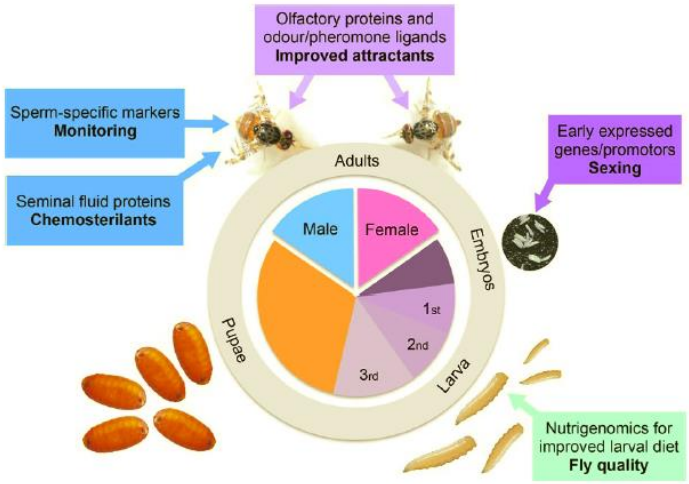


Fig. 4.3 Medfly functional genomics resources and their impact on the improvement of the SIT (Scolari *et al.*, 2014).

Current SIT applications are based on the use of a genetic sexing strain based on the temperature sensitive lethal (tsl) mutation. Resultant males are mass-reared in billions per week for sterilization and released in North, Central and South America, Australia, South Africa, and Mediterranean countries including Spain and Israel, to not only control existing populations but to also prevent new invasions/re-invasions (Dyck *et al.*, 2005). In the last 20 years, enormous progress has been made in understanding medfly biology, with the goal of developing/optimizing a wide range of molecular tools for the implementation of population control strategies (Scolari *et al.*, 2014)

However, currently, the SIT is the most promising alternative to pesticides, and a lot of research is currently ongoing to improve the efficacy and decrease the costs of this technique. To be most successful, this approach requires i) knowledge of the genetic basis of its complex mating behavior ii) knowledge of the genetic background of the released males and the genetic structure of the target population, iii) a genetic sexing strain for male-only production, iv) a sterilization system that inflicts the least possible fitness load, and v) effective procedures to monitor the efficiency of the programs. Since the mating in medfly is a composite process (Lance and McInnis, 2005), with males producing multiple sexual signals (visual, acoustic and olfactory) and females being highly selective based on male courtship performances (Whittier *et al.*, 1992), the ability

of sterilized males to attract and mate with wild females is crucially important for the success of the SIT. Our group is currently focusing on understanding the genetic/genomic bases of this complex mating behavior. In the following paragraph I introduce the biological features I studied in the framework of the medfly genome project with the aim of deepen our understanding of medfly biology to improve the SIT application.

4.1.4 Ceratitis capitata reproductive biology

The mating process in the medfly consists of four main phases: i) males' aggregation (also called "leks"), ii) sexual signaling from males (release of sex pheromones to attract females), iii) male courtship, and iv) sperm transfer.
The lek polygyny is the main mating strategy used by *C. capitata*: usually, two to 20 males aggregate under plant leaves to form small groups (called "leks") and, with a fanning movement of the wings, they start releasing pheromones and generating an acoustic "calling song" with the aim of attracting females (Eberhard, 2000). The "calling" behavior of males consists in the evagination of the rectal ampulla, which is a bubble-shaped structure from which the sex pheromone blend is emitted (Arita and Kaneshiro, 1989; Levinson *et al.*, 1987; Prokopy and Hendrichs, 1979; Whittier *et al.*, 1992). Chemicals produced by males attract females to the "lekking" leaves. The pheromone blend is made by three to five major chemicals known to be attractive to the females and by other 50 components with still an unknown function (Heath *et al.*, 1991; Siciliano *et al.*, 2014). Mating occurs on the leaves or leaf nodes and rarely on fruits of host trees and, and it can lasts from one to three hours. Male diet, and particularly the amount of carbohydrate and proteins, influences both the courting performance in leks, the copulatory frequency and success, and also the female receptivity (Eberhard, 2000).

a) *Ceratitis capitata chemoreception*

The molecular machinery of chemoreception plays a fundamental role in the medfly reproductive behavior, for example in behaviors associated with courtship. For these reasons, chemoreception is a good target for the development of innovative pest control strategies.
However, still little is known about the chemosensory machinery components involved in the detection of olfactory clues.
In order to characterize the medfly chemosensory repertoire, a recent study identified a number of OBP transcripts starting from EST libraries derived from different tissues (adult heads, male reproductive tracts and embryos) (Gomulski *et al.*, 2012). A total of seventeen putative *OBP* transcripts were identified, five of which share high sequence similarities and close phylogenetic affinities to *Drosophila melanogaster* pheromone

binding protein related proteins (PBPRPs). This finding paved the way to evolutionary and functional comparisons with homologous sequences from other tephritids of economical relevance such as those of the genera *Bactrocera* and *Rhagoletis* (Siciliano *et al.*, 2014).
One of these five genes (CcapOBP83a-2) is highly expressed in the antennae, with the highest expression in sexually mature individuals (Siciliano *et al.*, 2014). Phylogenetic analysis revealed that this medfly OBP clusters with *Drosophila* OBPs involved in the detection of pheromones. Additional studies revealed that this OBP displays high binding affinity for (E,E)-α-farnesene, one of the five major compounds in the medfly male pheromone mixture (Siciliano *et al.*, 2014), corroborating its role in chemoreception, and more specifically in the medfly reproductive behavior.
Odorant receptors are another important class of chemosensory proteins that are part of a large family of highly divergent seven-transmembrane domain proteins. In the membrane they form a heteromeric complex with a highly conserved non-conventional OR named as Odorant receptor co-receptor (Orco) (Vosshall and Hansson, 2011). Orco is highly conserved among all insect species analyzed to date and is involved in the localization of ORs to ORN dendrites (Larsson *et al.*, 2004).
The result section report the characterization of medfly OBPs and ORs, which has also been included in the recently published medfly genome paper.

b) *Ceratitis capitata seminal fluid*

Insect ejaculate transferred from the male to the female comprises not only the spermatozoa, but a number of different components, which have an important role in triggering a series of known physiological and behavioral changes in the mated female. The identification of the ejaculate components and their functions have been studied also in the medfly and resulted in a series of interesting findings (Eberhard, 2000; Jang, 2002; Jang *et al.*, 1998), which are briefly described in the following section. Such previous knowledge was the starting point for my work on the identification and annotation of the genes encoding proteins which have the potential to be transferred in the medfly seminal fluid.

Sperm transfer and use

In behavioral ecology, polyandry refers to a class of mating systems where females mate multiply, most often with different males. In medfly, polyandry is a frequent event in the field, and it is regulated by environmental, genetic and physiological factors (Bonizzoni *et al.*, 2007; Bonizzoni *et al.*, 2002; Gavriel *et al.*, 2009; Kraaijeveld *et al.*, 2005).
Females store spermatozoa in their reproductive tracts, which comprise two spermathecae (used for long-term storage) and a fertilization chamber (Marchini *et al.*, 2001; Twig and Yuval, 2005). On the other hand, males have an aedeagus with three

ejaculatory openings, one for each female storage compartment **(Fig. 4.4)**. Sperm are injected within the female sperm storage organs simultaneously (Marchini *et al.*, 2001). After insemination, sperms can survive for up to 18 days (Scolari *et al.*, 2008; Twig and Yuval, 2005).

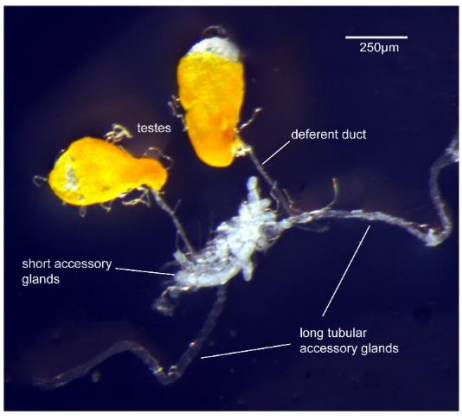


Fig. 4.4 Dissection of the male reproductive tract including the testes and accessory glands (Scolari *et al.*, 2012).

High fecundity and high genetic variability are crucial for the success of an invasive insect species, and polyandry is a good reproductive strategy to ensure egg fertilization even in presence of a genetic bottleneck (Chapman *et al.*, 1998; Gavriel *et al.*, 2009; Kraaijeveld *et al.*, 2005).
Noticeably, polyandry and sperm use negatively impact the SIT (see page xx), which, to be efficient, has to be monitored in terms of number of sterile males released, and their mating success. In 2008, Scolari *et al.*, described the first sperm-specific marking systems in *C. capitata* (Scolari *et al.*, 2008; Scolari *et al.*, 2014) **(Fig. 4.5)**.

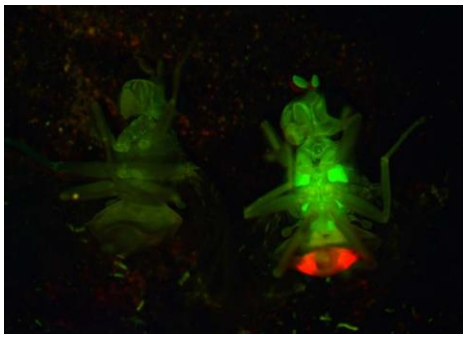


Fig. 4.5 On the right a transgenic Mediterranean fruit fly male carrying two different types of molecular markers (overall EGFP green and Dsred testes-specific colour). On the left a non-transgenic male (Source: Francesca Scolari).

The system was based on the spermatogenesis-specific β2-tubulin (Ccβ2t) promoter. Fluorescent sperm can be easily isolated from both testes and spermathecae, and the corresponding strains did not show major competitiveness disadvantages in laboratory conditions. This is of great importance for SIT. Moreover, such markers should facilitate studies related to the mating behavior of the medfly, unraveling mechanisms of sperm transfer, sperm storage, use, precedence and sperm competition (Scolari *et al.*, 2008).

Seminal fluid proteins (SFPs)

Testes and male accessory glands (MAGs) participate in the maintenance of complementary reproductive functions. In the testes, the key regulatory genes of spermatogenesis tend to be conserved to guarantee the male specific processes required for sperm production (White-Cooper and Bausek, 2010). By contrast, the accessory gland secretions act as key factors in male insect reproductive success and the genes expressed in the MAGs are subject to rapid evolution as a result of sexual conflict and competition (Ravi Ram and Wolfner, 2007).
Transcriptomic resources (Scolari *et al.*, 2012) developed from medfly testes and male accessory gland tissues were useful to identify the molecular components of the male reproductive system, among which transcripts encoding proteins involved in spermatogenesis, fertility, sperm-egg binding, as well as those enconding SFPs. As previously mentioned, SFPs are fast-evolving genes, and they are ideal target for the development of species-specific diagnostic markers or for novel and environmentally-friendly chemosterilants which will mimic the effects of MAG proteins, like impairing the sperm storage, or interfering with female remating (Le *et al.*, 2014).

4.1.5 Ceratitis capitata in the post-genomic era

Ceratitis capitata was the first non-drosophilidae species in which the germline was genetically transformed (Loukeris *et al.*, 1995), enabling studies on its biology in ways that were previously impossible (Ogaugwu *et al.*, 2013; Schetelig *et al.*, 2009; Schetelig *et al.*, 2009) **Fig.4.5**.

The medfly genome sequence has been recently published (Papanicolaou *et al.*, 2016), allowing a more detailed analysis of the complex biological traits responsible of its development, and its peculiar adaptation to the environment. This is one of 30 arthropod genome sequencing projects that have been initiated as a part of a pilot project of the i5K consortium, whose aim is to sequence the genome of 5000 insect species in the coming years (Consortium, 2013).

Thanks to this genome sequencing effort, new genes, promoters and regulatory sequences are being characterized and are now studied for i) developing and improving sexing strains, ii) monitoring the released males in the field, and iii) determining the mating status of wild females (Scolari *et al.*, 2014).

This new knowledge, along with genetic tools available for functional and applied studies, have the potential to significantly advance the means of controlling the size and invasive potential of medfly populations. The availability of the genome sequence of two closely related dipterans, i.e. *D. melanogaster* and *Musca domestica*, enabled comparative genomic analyses on gene families involved in adaptation and invasiveness. The medfly genome has been sequenced from DNA extracted from pools of adult flies from lines inbred for 20 generations and the genome size was estimated to be of 479 Mb (Papanicolaou *et al.*, 2016). In-depth curation of more than 1,800 mRNAs shows specific gene expansions that can be related to invasiveness and host adaptation, including gene families for chemoreception, xenobiotics metabolism, cuticle proteins, opsins and aquaporins among others. In addition, genes relevant to IPM control have been characterized, including those that could be target of genetic modifications to improve medfly strains used in SIT (Papanicolaou *et al.*, 2016).

4.2 Aim of the work

The medfly has been an established lab organism for several decades, and is notable as being the closest non-drosophilid relative to *Drosophila* subject to genetic analysis (Papanicolaou *et al.*, 2016). In order to improve the SIT, already used to control several tephritid species in the contest of integrated pest management approaches, several genetic studies have been performed on medfly in order to use genetic manipulation of the species in this direction.

It was the first to have its germ-line efficiently transformed by a transposon-based vector system and it is now a model system for genetic manipulation in non-drosophilids, including functional genomics analysis, new vector systems for transgene stabilization, genomic targeting, and transgenic strains created for population control (Papanicolaou *et al.*, 2016).

During my PhD studies, I participated to the medfly genome project. Part of my research was dedicated to the manual curation and annotation of genes involved on reproduction and chemoreception, specifically OBPs, ORs and putative seminal fluid proteins.

4.3 Materials and Methods

Chemoreceptor genes identification and annotation

OBP and OR gene families have been annotated performing tBLASTn searches on the genomic scaffold sequences, using *D. melanogaster* and *M. domestica* OBPs and ORs as queries. BLASTx, local protein databases of the *D. melanogaster* and *M. domestica* OBPs and ORs, was used for putative proteins encoded by the identified gene models that produced hits (<1e–10). Gene models were manually curated on WebApollo and named using a numerical system, with genes on the same scaffold numbered sequentially. Signal peptide sequences were excluded before alignment and the phylogenetic analyses of OBP protein sequences.
For each family the amino acid sequences were aligned using MAFFT v7 (Katoh and Standley, 2013) with the E-INS-i strategy, BLOSUM62 matrix, 1000 maxiterate and offset 0 or CLUSTALX v2.0 (Larkin *et al.*, 2007) using default settings with the relevant families of *D. melanogaster*. The most appropriate model of molecular evolution for each dataset was determined using MEGA 6.0.6 (Tamura *et al.*, 2013). Phylogenetic relationships were estimated using maximum likelihood with 1000 bootstrap replications using MEGA 6.06 retaining positions present in at least 75 % of the sequences. The resulting mid-point rooted tree was drawn using FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) and iDraw (www.indeeo.com). BLASTp searches were performed on the JAMg Consensus Gene Set v1, as well as high-confidence and low-confidence protein sets from NCBI. For further details, see pag. 127.

Seminal fluid proteins genes annotation

Putative medfly SFP genes have been annotated using the amino acid sequences of the 146 characterized *D. melanogaster* SFPs (Findlay *et al.*, 2009) (tBLASTn, e-value $<10^{-10}$). Sixty-four of the *Drosophila* SFPs gave no significant hits to the medfly genome, whereas the remaining sequences resulted in multiple hits.
The predicted amino acid sequences of the identified medfly gene models were considered for annotation if they gave significant reciprocal BLASTp hits (e-value $<10^{-10}$) in the NCBI nr database to sequences belonging to known SFP functional classes. In addition, we queried (tBLASTx, e-value $<10^{-10}$) the genome scaffolds using the ESTs previously derived from medfly male testes and male accessory glands (Scolari *et al.*, 2012). The predicted amino acid sequences of the identified gene models were considered for annotation if they gave significant reciprocal BLASTp (e-value $<10^{-10}$) hits in the NCBI nr database to sequences belonging to known SFP functional classes. For further details, see pag. 127.

4.4 Results and Discussion

Chemoreception genes: insights into adaptation and invasiveness

As previously described, medfly is a phytophagous species which strongly relies on its host plants. This is a crucial difference compared to the dietary habits of both *Drosophila* and *Musca*, which mainly feed on decaying organic matter and carcasses, respectively. The capacity of the medfly to exploit a wide range of host plants is also due to its sophisticated chemosensory system.
The functional characterization of molecules involved in chemoreception, such as those involved in courtship or fruit detection, not only provides insights of its extraordinary adaptation, but might be crucial for developing new synthetic ligands that can act as attractants or repellents. Such molecules can be integrated as part of Integrated Pest management (IPM) control strategies, and specifically used for disrupting oviposition and mating behaviors.
Forty-six OBP genes were identified in the medfly genome, compared to 52 in *D. melanogaster* (Vieira and Rozas, 2011) and 85 in *M. domestica* (Scott *et al.*, 2014). In addition, we could classify the identified sequences in the different subfamilies. The members of the thirty Classic OBP subfamily (containing six highly conserved cysteine residues that form three disulphide bonds) have been found in the *C. capitata* genome, together with ten members of the Minus-C OBP subfamily (four conserved cysteine residues and only two disulphide bonds), six Plus-C (contain additional conserved cysteine residues), and two Dimer or Double OBPs **(Fig. 4.6)**.

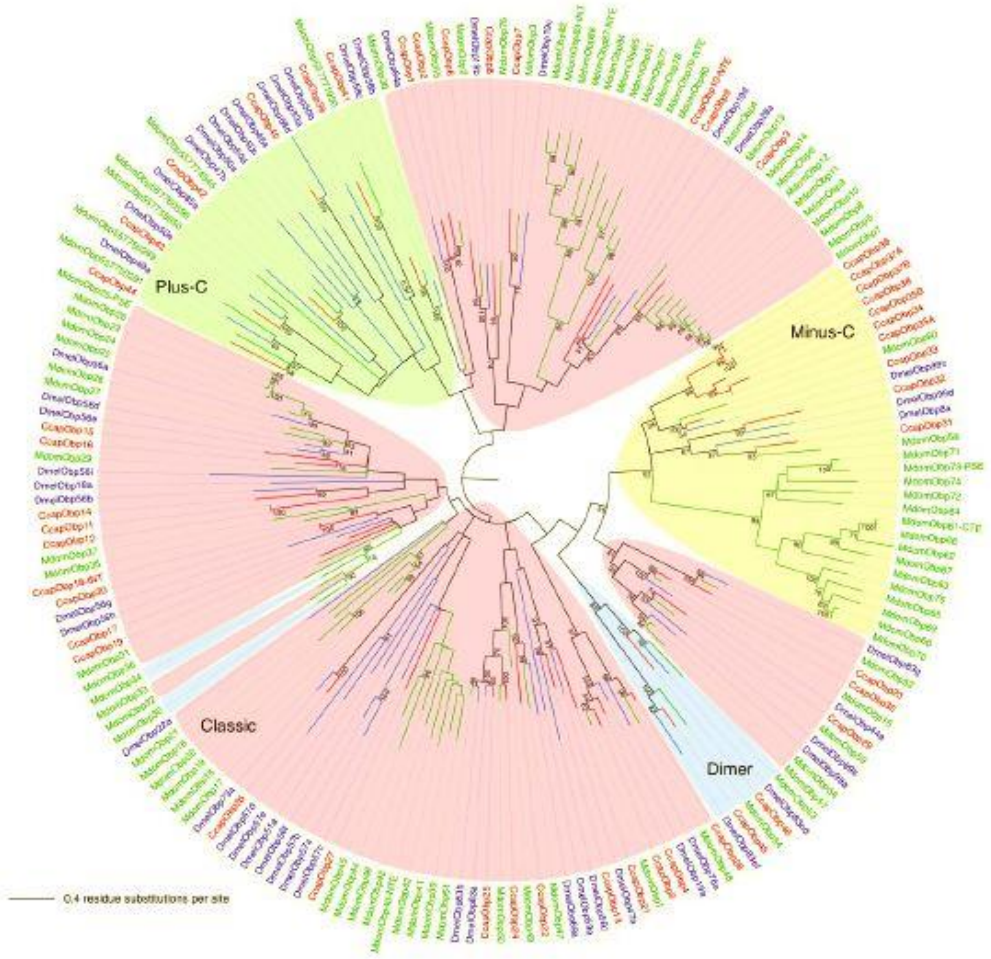


Fig. 4.6 Phylogenetic relationships of OBPs from *C. capitata*, *D. melanogaster* and *M. domestica*. The midpoint rooted maximum likelihood (log likelihood = -33834) tree was inferred using the Le & Gascuel model (Le and Gascuel, 2008) with a discrete Gamma distribution and some invariable sites. Bootstrap values greater than 50% (1000 replications) are shown. The different subfamilies of OBPs are highlighted by colored shading (Papanicolaou *et al.*, 2016).

Moreover, besides having several separate expansions of receptors implicated in fruit detection and courtship, the similarities observed between *C. capitata* and *Drosophila* are most likely due to similarities in their ecology (Papanicolaou *et al.*, 2016). Also 76 OR have been identified in the medfly genome.

For further discussion and results, see the insert "The whole genome sequence of the Mediterranean fruit fly, *C. capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species", pag. 127.

Ceratitis capitata seminal fluid protein (SFP) genes

As previously discussed, SFPs are known to induce numerous physiological and behavioral post-mating changes in females. These changes include the i) receptivity to re-mating; ii) sperm storage parameters; iii) egg production; iv) sperm competition and feeding behaviors (Avila *et al.*, 2011). SFPs play critical roles in sperm activation, gamete interaction and ovulation but only limited information are available on their functional roles, mostly because of their rapid evolution as demonstrated by (Begun *et al.*, 2006), therefore representing a large repertoire of unexplored potential targets for controlling reproduction, which can enhance population control strategies, such as the SIT.
We identified 459 putative SFP genes that have been further grouped into 17 functional classes previously defined for *Drosophila* SFPs (**Fig. 4.7**) (Findlay *et al.*, 2008).

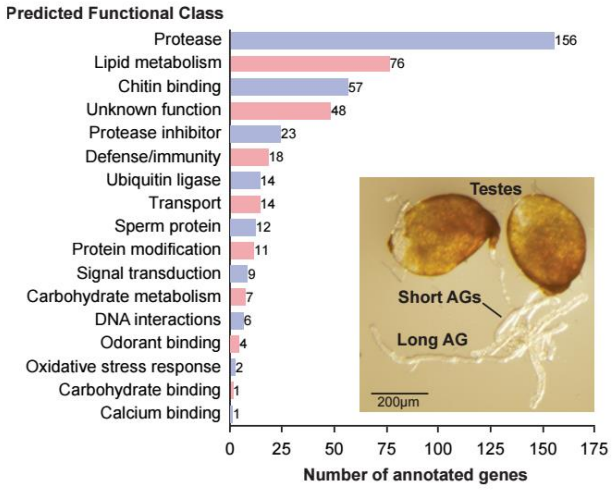


Fig. 4.7 Proposed functional classes of putative Seminal Fluid Protein (SFP) coding genes identified in the medfly genome. In the box, dissected male reproductive tract including the testes and the accessory glands (AGs) (Papanicolaou *et al.*, 2016).

Proteases account for the most abundant class of genes, followed by genes involved in lipid metabolism and chitin binding, and sequences with unknown function. Male and

female whole body RNA-Seq libraries (female ISPRA SRR836189 and male ISPRA SRR836190), as well as reproductive tissue datasets, showed that 37 of all annotated genes are male-biased, with 31 of them being predominantly transcribed in the male reproductive tissues. Although SFP-encoding genes do not necessarily display a male-biased expression profile, these features make them putative candidates for further analyses. For detailed discussion on the most abundant seminal fluid proteins identified in the medfly genome, please see the insert "The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species", pag. 127.

4.5 Conclusion and Perspectives

The resources I contributed to generate in the medfly genome project are already of key importance for the development of several research lines:

1) The annotated chemosensory proteins are being expressed in heterologous systems to perform binding assays with odorant molecules derived from male pheromone mixture and plants. This will allow to identify both the natural ligands but also understand the features of the binding pockets (in the case of OBPs), allowing the planning of binding assays to identify novel and powerful attractants/repellents.
2) The annotation of a wide dataset of seminal fluid proteins is the basis for proteomic-based investigations aimed at identifying the actual components of the seminal fluid. Given the amount of tools (including transgenesis) available for the study of gene functions in the medfly, it will be possible to rapidly proceed to: i) RNAi-based assays on selected candidates, to evaluate the individual effects of specific SFP on the post-mating physiology of the medfly females (i.e. manipulating its fertility or remating capacity); ii) comparative genomics analyses, as I performed in tsetse fly species (see chapter 2), to identify similarities and differences among the tephritidae family, as several genomes are being currently sequenced.

List of original manuscripts

Insect Biochemistry and Molecular Biology 79 (2016) 13–26

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Sperm-less males modulate female behaviour in *Ceratitis capitata* (Diptera: Tephritidae)

Paolo Gabrieli, Francesca Scolari, Alessandro Di Cosimo, Grazia Savini, Marco Fumagalli, Ludvik M. Gomulski, Anna R. Malacrida, Giuliano Gasperi*

Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, Pavia, Italy

ARTICLE INFO

Article history:
Received 29 June 2015
Received in revised form
22 September 2016
Accepted 4 October 2016
Available online 6 October 2016

Keywords:
Ceratitis capitata
Innexin
Seminal fluid
Sperm
RNA interference
Post-mating behaviour

ABSTRACT

In the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann)(Diptera: Tephritidae), mating has a strong impact on female biology, leading to a decrease in sexual receptivity and increased oviposition and fecundity. Previous studies suggest that sperm transfer may play a role in inducing these behavioural changes. Here we report the identification of a medfly *innexin* gene, *Cc-inx5*, whose expression is limited to the germ-line of both sexes. Through RNA interference of this gene, we generated males without testes and, consequently, sperm, but apparently retaining all the other reproductive organs intact. These sperm-less males were able to mate and, like their wild-type counterparts, to induce in their partners increased oviposition rates and refractoriness to remating. Interestingly, matings to sperm-less males results in oviposition rates higher than those induced by copulation with control males. In addition, the observed female post-mating behavioural changes were congruent with changes in transcript abundance of genes known to be regulated by mating in this species. Our results suggest that sperm transfer is not necessary to reduce female sexual receptivity and to increase oviposition and fecundity. These data pave the way to a better understanding of the role/s of seminal components in modulating female post-mating responses. In the long term, this knowledge will be the basis for the development of novel approaches for the manipulation of female fertility, and, consequently, innovative tools to be applied to medfly control strategies in the field.

1. Introduction

In insects, mating dramatically alters female behaviour and physiology: it induces ovulation and increases egg-laying rates, decreases propensity to remate, modulates flight and feeding behaviour, and affects lifespan (Avila et al., 2011; Kubli and Bopp, 2012). These effects are triggered by the male ejaculate transferred to the female upon mating (McGraw et al., 2004). Ejaculate components, i.e. sperm and seminal fluid, are produced in the reproductive tract, which comprises the testes, the male accessory glands (MAGs), the ejaculatory duct and bulb. These tissues secrete seminal fluid proteins (SFPs), including those synthetized in the MAGs, the accessory gland proteins (Acps) (Wolfner, 2007).

Although there is now a general consensus on the strong effects mating exerts on female insect behaviour (Avila et al., 2011), the identity of the main factor/s affecting female behaviour is not conserved between different organisms. At least three components have been shown to be implicated in this phenomenon: mechanical stimulation and transfer of seminal fluid and/or sperm, which appear to co-interact differently depending on the species. For instance, physical stimuli suppress female receptivity soon after mating in the Hessian fly *Mayetiola destructor* (Bergh et al., 1992) and in the tsetse fly *Glossina morsitans* (Gillot and Langley, 1981). In the malaria mosquito, *Anopheles gambiae*, the 20-hydroxyecdysone hormone secreted in the MAGs and transferred upon mating is sufficient to switch female behaviour from virgin to mated status, irrespective of the presence of sperm (Gabrieli et al., 2014; Thailayil et al., 2011). In *Drosophila melanogaster*, factor/s in the seminal fluid are responsible for the reduction of female receptivity (Kalb et al., 1993), but sperm is necessary to retain this activity ("sperm effect") (Manning, 1962, 1967; Swanson, 2003). In this last species, a 36 amino acid peptide synthetized in the MAGs, the Sex-Peptide

Abbreviations: MAGs, Male Accessory Glands; SFPs, Seminal Fluid Proteins; Acps, ACcessory gland Proteins; SP, Sex-Peptide; Medfly, Mediterranean fruit fly; *inx*, *innexin*.

* Corresponding author.
E-mail address: gasperi@unipv.it (G. Gasperi).

http://dx.doi.org/10.1016/j.ibmb.2016.10.002
0965-1748/

(SP), is the main trigger of this response in the female (Chapman et al., 2003; Liu and Kubli, 2003), through the interaction with its receptor (Sex-Peptide Receptor) (Haussmann et al., 2013; Yapici et al., 2008) at the level of reproductive tract and central nervous system. Sperm are required for the correct transfer of SP and its retention (Liu and Kubli, 2003; Peng et al., 2005).

Apart from being a question of biological relevance and diversification among Diptera, the role of sperm and SFPs in modulating female post-mating responses acquires particular importance with regard to the development of pest control strategies (Robinson et al., 2009). This is especially relevant in species with high reproductive capacities, such as the Mediterranean fruit fly (medfly), *Ceratitis capitata* (Wiedemann), an extremely invasive agricultural pest with an almost worldwide distribution (Gasperi et al., 2002; Malacrida et al., 2007). Over the past two decades, much progress has been made in understanding medfly biology both in terms of structure/physiology/function of the reproductive apparatus (Marchini et al., 2003; Williamson, 1989), mating/remating behaviour (Bertin et al., 2010; Bonizzoni et al., 2002, 2007; Eberhard, 2000; Kraaijeveld et al., 2005; Miyatake et al., 1999; Scolari et al., 2014; Yuval and Hendrichs, 2000), and functional genomics of reproduction (Gomulski et al., 2008, 2012; Scolari et al., 2012; Scolari et al., 2014; Siciliano et al., 2014a, 2014b).

In this fly, polyandry, associated with a variable degree of sperm precedence, is a frequent behaviour in the wild (Bonizzoni et al., 2002, 2006; Gavriel et al., 2009; Kraaijeveld et al., 2005).

It has also been suggested that sperm transfer itself has an impact on female post-mating behaviour since surgically castrated males were inefficient in reducing female remating receptivity two days after *copula* (Miyatake et al., 1999). Conversely, physiological assays aimed at investigating the role/s of MAG secretions revealed that their injection into virgin females induced a significant switch from attraction to male pheromone for copulation to host-seeking behaviour for oviposition (Jang, 1995; Jang et al., 1998). The composition of the seminal fluid in the medfly, and, consequently, the exact functional roles of its components are poorly understood. To address this biological question, molecular studies have been initiated aimed at the identification of the genes expressed in the male reproductive tract (Davies and Chapman, 2006; Scolari et al., 2012) and their transcriptional profiles in response to multiple matings, revealing potentially interesting candidates that may be involved in sperm transfer and female post-copulatory behaviour (Scolari et al., 2012).

Here we used a functional genomics approach to unravel the role of sperm transfer in modulating female behaviour. To do so, we generated males without sperm by knocking down the expression of a newly identified medfly germ-line specific *innexin* gene that, like its *D. melanogaster* (Tazuke et al., 2002) and *An. gambiae* (Magnusson et al., 2011; Thailayil et al., 2011) homologues, is essential for male and female germ-line development. We assessed the mating ability of the resulting sperm-less males, focusing on the behavioural and molecular consequences on their female partners. We discuss our results in the light of the reproductive biology of the species and their relevance for the control strategies currently used against the medfly.

2. Materials and methods

2.1. Medfly samples and experimental strategy

The medfly laboratory strain ISPRA, established in 1968 at the European Community Joint Research Centre (Ispra, Italy) with wild flies from Sicily and Greece, was used in this study. The strain has been maintained in the insectary at the Dept. of Biology and Biotechnology, University of Pavia (Italy), since 1979 under standard rearing conditions (Saul, 1982).

This work was organized in the following experiments: 1) collection of fertilized eggs from the medfly ISPRA strain, 2) embryo microinjection of double-stranded RNA (dsRNA) against the newly identified *innexin-5* gene to generate sperm-less males, 3) comparison of the mating performance between normal (control) and sperm-less males, 4) comparison of the physiological/behavioural response(s) between females mated to control and sperm-less males, 5) evaluation of the transcriptional activities of mating-responsive genes in females mated to control and sperm-less males.

2.2. Mating assays

Adult flies were sexed at emergence and used in mating and remating assays to compare the performance of control and sperm-less males.

For mating assays, four day-old females were placed in a 30 × 30 × 30 cm Plexiglas cage together with age-matched virgin sperm-less or control males (about 150 females and 200 males per cage); couples were isolated and checked for *copula* duration. To verify that the injected males used in mating assays were sperm-less, they were dissected following copulation to check for the absence of testes/sperm. Only females mated to sperm-less males were retained and transferred to a separate cage for subsequent tests.

For remating assays, pools of 25 mated females (either to control or sperm-less males, respectively) were placed in a 30 × 30 × 30 cm Plexiglas cage and exposed to 50 virgin males (either to sperm-less or control males) at different time points after the first mating (0 h, 3 h, 6 h, 1 d, 3 d, 6 d and 12 d, respectively). It is noteworthy that females laid eggs in the interval between the two copulations, thus using sperm received from the first mating for fertilization (in the case of matings with control males). Remating events were scored in a 4 h time-window, couples were isolated and *copula* duration was recorded. Three biological replicates were performed for each time point and data of the proportion of remated females were analysed using the Cochran-Mantel-Haenszel test after having tested the homology of odds ratios using the Woolf test. These analyses (and all those that follow) were performed using the R statistical software environment (R Core Team, 2015).

We counted the number of sperm in the spermathecae of mated females according to the protocol described by (Bertin et al., 2010). Sperm numbers were normalized using the square root function (following Kolmogorov-Smirnov test and Bartlett's test for equal variance) and analysed with one-way ANOVA followed by Tukey's post-hoc test.

We additionally scored the number of eggs laid by mated and remated females, placed individually in small cages. Seventy-five females were considered for each treatment. Eggs were counted every 24 h. Propensity of females to oviposit was compared using the Log-Rank Mantel Cox test and point by point comparison was performed using Fisher's test. The number of laid eggs was compared using a Stratified Analysis with Nonparametric Covariable Adjustment (in the case of the virgin vvs mated comparison, as the data were not normally distributed, as assessed by the D'Agostino test) (Kawaguchi and Koch, 2015) or a two-way ANOVA followed by Tukey's post-hoc test (in the case of the mated vvs remated comparison; the normality and homoscedasticity of the data was assessed using the D'Agostino test and the Brown-Forsythe Levene-type test).

2.3. Identification of the Cc-inx5 gene and phylogenetic analyses

In order to generate sperm-less males, we searched for the

orthologue/s of the *Drosophila* germ-line-specific gene *innexin-5* in the medfly. Putative orthologues of *Drosophila innexin* genes were identified in the medfly genome (http://arthropodgenomes.org/wiki/Ceratitis_capitata, GenBank ID: GCA_000347755.2) using DELTA-BLASTP searches on the NCBI server (Boratyn et al., 2012).

A phylogenetic analysis was performed including the five medfly innexin predicted amino acid sequences and 14 innexin and innexin-like sequences from *D. melanogaster* and *An. gambiae*. The amino acid sequences were aligned using MAFFT ver. 7 (Katoh and Standley, 2013) with the E-INS-i strategy, BLOSUM62 matrix, 1000 maxiterate and offset 0. The most appropriate model of molecular evolution for the dataset was determined using MEGA 5.2.2. (Tamura et al., 2011). Phylogenetic relationships were estimated using Maximum Likelihood with 1000 bootstrap replications using MEGA 5.2.2. The resulting mid-point rooted tree was drawn using FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.4. Cc-inx5 transcriptional profile

RNA was separately isolated from different body compartments of males and females, namely heads, thoraxes, and abdomens deprived of reproductive apparatus (from five, three and three individuals for each sex, respectively). RNA was additionally isolated from MAGs, testes and ovaries (from 60, 20, and 20 individuals for each tissue, respectively). RNA was extracted using TRIzol® Reagent (Invitrogen, Carlsbad, CA, USA) following manufacturer's instructions and resuspended in RNase-free water. Contaminating genomic DNA was digested using DNA-free (Ambion, Foster City, CA, USA). RNA concentrations were quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, USA). cDNA was generated using the cloned AMV first-strand synthesis kit (Invitrogen) according to manufacturer's instructions.

To determine *Cc-inx5* transcript abundance in the six above-mentioned tissues, RT-PCRs were performed using specific primers designed with *Primer3 v.0.4.0* (Rozen and Skaletsky, 2000) (Table S1). One microlitre of synthetized cDNA was used for the reactions (3 min at 94 °C; 25 cycles of 30 s at 94 °C, 30 s at 59 °C, 1 min at 72 °C; 5 min at 72 °C), performed using Taq DNA Polymerase (Invitrogen). The amplification products were electrophoresed on 1.0% agarose-gels (Agarose-LE Ultrapure USB).

2.5. dsRNA preparation and RNA interference assay

Cc-inx5 dsRNA was obtained and injected in medfly embryos according to the protocols described for *Ceratitis* and *Drosophila* (Kennerdell and Carthew, 1998; Pane et al., 2002). A *Cc-inx5* fragment from positions 619 to 863 was amplified from ovary cDNA with primers that incorporated a T7 promoter sequence at each end of the product (Table S1). *In vitro* transcriptions were performed with the Megascript Kit (Ambion), RNA was ethanol precipitated and resuspended in injection buffer (Rubin and Spradling, 1982). As a control, turbo Green Fluorescent Protein (tGFP) gene (Evrogen) fragments flanked by T7 promoter sequences were generated by amplifying tGFP using specific primers and dsRNA was obtained as previously described. Embryos were collected for 30 min, dechorionated and microinjected in the posterior pole with 2 μg/μl dsRNA solution (Gabrieli and Scolari, 2016). Injected eggs were covered with Halocarbon oil 800 (Sigma-Aldrich), placed at 25 °C in Parafilm-sealed Petri dishes with moist Whatman paper under the lid. Hatched larvae were reared on a carrot-based food (Gabrieli and Scolari, 2016). Freshly eclosed virgin male and female adults were dissected using fine forceps and their genitalia were observed under a Zeiss Axioplan microscope. Images were captured using an Olympus DP70 digital camera.

2.6. Tissue dissections and DAPI staining

Sexually mature adults (at least 4 days old) were cold-anaesthetized and dissected in 1x PBS under a dissecting stereomicroscope using fine forceps. Spermathecae of virgin females or females mated to sperm-less or control males were stained using DAPI as previously described (Bertin et al., 2010).

2.7. Validation of RNA interference by Real Time qPCR

To quantify the extent of dsRNA-mediated *Cc-inx5* knockdown, Real Time qPCRs were performed. Total RNA was extracted from whole abdomens of adult individuals and reverse-transcribed into double stranded cDNA using the iScript ™ cDNA Synthesis Kit (Biorad). RT-qPCR reactions were performed with an MJ Mini™ Personal Thermal Cycler (Biorad) using the SsoFast EvaGreen Supermix kit (Biorad) and MiniOpticon (Biorad). Primers were designed using *Primer3Plus* to obtain amplification products of about 100–250 bp (See primer list in Table S1). Real Time qPCR was performed in 15 μl reactions containing 7.5 μl of SsoFast EvaGreen Supermix kit, 1.5 μl of forward and reverse primers (300 nM final) and 4.5 μl of cDNA (diluted 1:10). Cycling parameters were: 3 min at 95 °C, 40 cycles of 10 s at 95 °C and 30 s at 55 °C. Fluorescence was detected at the end of each cycle. Five biological replicates were performed and the specificity of the amplification product was determined by melt-curve analysis. Relative quantification was performed using CFX Manager Software (Biorad). All data were normalized to the expression levels of *GAPDH2* and *G6PDH*. To calculate the relative fold changes compared to control individuals, the error propagation formula was used (Bustin et al., 2009) and one-sample *t*-test statistical analyses were performed using *GraphPad InStat 3* (GraphPad Software).

2.8. Tissue in situ hybridization

Cc-inx5 and *Cc-Vasa* (GeneBank Accession number: XP_004520983) were amplified using primers listed in Table S1. We chose to use *Vasa* because this gene is specifically expressed in the germ-lines of both sexes, and only in germ-line progenitor cells and developing spermatocytes, and not in supporting somatic cells (Papathanos et al., 2009). Probes for *in situ* were prepared using the Biotin High Prime kit (Roche). Dissected testes were fixed overnight in 4% paraformaldehyde at 4 °C. The tissues were then incubated with the probe for 24 h at 70 °C in hybridization buffer (50% formamide, 5x SSC, 0.1% Tween and 10 mM citric acid). The signals were detected using streptavidin-HRP and Alexa Fluor 594 Tyramide (Invitrogen) and visualized using a Zeiss Axioplan epifluorescence microscope. Images were captured using an Olympus DP70 digital camera.

2.9. Protein extraction and SDS-PAGE

Testes and MAGs from 30 males, either sperm-less or controls, were separately dissected in 1x PBS under a stereomicroscope and extracted in 20 μl extraction buffer (PBS, 20 mM EDTA, protease inhibitors cocktail – Sigma Aldrich). After quantification of protein content using the Bicinchroninic Acid Assay (BCA), about 20 μg of proteins were separated on a 12.5% polyacrylamide gel and stained with Coomassie.

2.10. Analyses of mating-induced transcriptional changes using Real Time qPCR

To determine the role of sperm transfer in regulating gene expression in medfly females, we studied the transcriptional

profiles of a subset of genes by Real Time qPCR assays.

Total RNA was extracted from pools of heads and abdomens of females that were either i) virgin, or ii) mated to control or iii) to sperm-less males. For each of the three conditions, RNA was extracted from four pools of three females at 6, 24 and 72 h after mating. The RNA was reverse-transcribed into cDNA and used as the template for Real Time qPCR reactions according to the protocol described above. In the analysis of the relative gene transcription fold-change, virgin females were taken as calibrators. The relative fold changes were calculated using the Pfaffl method (Bustin et al., 2009). Differential transcription between mated (either with sperm-less or control males) and virgin females was tested using one-sample *t*-test. Differential transcription between mated females was tested using the two-way ANOVA followed by Tukey's post-hoc test (after having assessed the normality and homoscedasticity of the data using the Lilliefors test and the Brown-Forsythe Levene-type test).

3. Results

In this study our goal was to determine the effects of sperm on the behaviour/physiology and gene transcriptional profiles of mated females. To do so, we first established a base-line of behavioural/physiological parameters related to mating and remating using wild-type flies.

3.1. Mating effects on female behaviour

The effects of mating on female behaviour were evaluated through analysis of three key behavioural/physiological parameters: i) the remating tendency at different time-points after the first mating, ii) the number of ovipositing females and iii) the number of eggs laid by each female after one or two matings.

3.1.1. Female receptivity

The receptivity of medfly females showed a significant reduction in the initial hours following the first mating (Fig. 1A, Suppl. Table S2; 0 h vvs 3 h: Mantel-Haenszel $\chi^2 = 5.789$, $P = 0.016$, Su). It remained low for at least three days post mating (3 h vvs 6 h: Mantel-Haenszel $\chi^2 = 0.435$, $P = 0.509$; 3 h vvs 1d: $\chi^2 = 1.292$, $P = 0.256$; 3 h vvs 3d: $\chi^2 = 2.773$, $P = 0.096$), after which it displayed a clear increase (3d vvs 6d: Mantel-Haenszel $\chi^2 = 6.755$, $P = 0.009$) and, 12 days post mating, it was back to the level of the first time-point (i.e. 0 h; 0 h vvs 12d: Mantel-Haenszel $\chi^2 = 0.031$, $P = 0.860$).

To assess whether the presence of sperm in the female sperm storage organs could have an impact on remating tendency, sperm cells were counted in the two spermathecae of 30 females 1, 3, 6 and 12 days after first mating, respectively. We did not consider early post-*copula* time-points to allow sperm transferred upon mating to move completely to the spermathecae (Bertin et al., 2010). Sperm numbers in the spermathecae were significantly lower at 6 and 12 days, compared to the counts at 1 and 3 days, after the first mating (One-way ANOVA $P < 0.0001$, Fig. 1B, Table S3).

Mating and remating duration did not significantly differ at all time-points analysed, with the exception of the 1-day time-point: here remating duration was shorter than that of the first mating (Unpaired *t*-test, $t_{631} = 2.136$, $P = 0.03$; Fig. 1C, Table S4).

3.1.2. Female oviposition rate and fecundity

Mating enhanced female oviposition rates (i.e. the percentage of females laying eggs) in the first 4 days after *copula*, as shown by the oviposition rate curves of virgin and mated females (Fig. 2A, Log-Rank Mantel Cox test $\chi^2 = 11.10$; $P = 0.001$), with the difference being significant from day 2 (Fig. 2A, Fisher's test 2d $P = 0.0169$; 3d $P = 0.0033$; 4d $P = 0.0022$). It is important to enotethat, while the portion of females laying eggs at day 1 and day 3 is significantly different in the mated females (Fisher's test, $P = 0.001$), this is not true in the case of virgin females (Fisher's test, $P = 0.084$; Table S5).

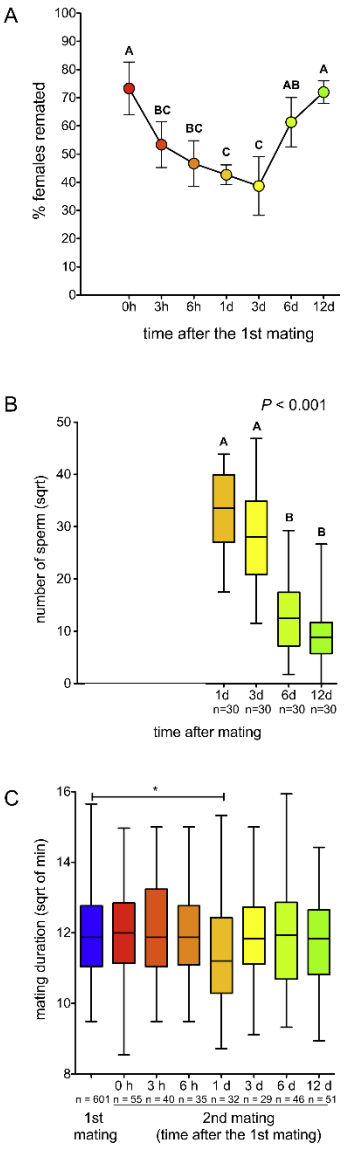
Mating also enhancedfemale fecundity, i.e. the number of laid eggs, as determined comparing the number of eggs deposited by mated and virgin females (Stratified Analysis with Nonparametric Covariable Adjustment $\chi^2 = 16.1$, $P < 0.001$; Fig. 2B, Table S6). This difference is significant at day 2 after mating (Mann-Whitney $U = 23.5$, $P = 0.032$), although the fecundity increased, compared to day 1, in both virgin (Mann-Whitney $U = 2$, $P = 0.015$) and mated (Mann-Whitney $U = 24$, $P = 0.003$) females.

Twice-mated females shared the same overall tendency to oviposit as once-mated females (Fig. 2C, Log-Rank Mantel Cox test, $\chi^2 = 0.122$, $P = 0.727$). Surprisingly, however, the immediate effect of mating and remating on oviposition was not the same (ANOVA interaction $P = 0.002$; Fig. 2D, Table S7), as the number of eggs laid on the day following remating (corresponding to the second day after the first mating) is lower in twice-mated females than in mated females (Tukey post-hoc test after ANOVA, $P = 0.004$). This difference was not evidentevident in the following days (Tukey post-hoc test after ANOVA, day 3 $P = 0.980$; day 4 $P = 0.818$).

3.2. Medfly Cc-inx5: a novel germ-line specific innexin gene

The release of the medfly genome (ISPRA strain) within the i5K initiative (http://arthropodgenomes.org/wiki/Ceratitis_capitata, GenBank ID: AOHK00000000.1) facilitated the analysis of the repertoire of medfly genes encoding Innexin proteins. DELTA-BLASTP searches using the NCBI server (Boratyn et al., 2012) were performed to identify *C. capitata* orthologues of *D. melanogaster* Innexin encoding genes (*ogre/inx1, inx2, inx3, zpg/inx4, inx5, inx6* and *inx7*). Five genes were identified (*Cc-inx1, Cc-inx2, Cc-inx3, Cc-inx5* and *Cc-inx7*) that share significant similarities to *D. melanogaster ogre/inx1, inx2, inx3, inx5* and *inx7*, respectively (Table 1). The *Drosophila zpg/inx4* and *inx6* genes do not appear to have clear orthologues in the medfly genome. The phylogenetic relationships among the predicted *innexin* proteins from *C. capitata* (n = 5), *D. melanogaster* (n = 7) and *An. gambiae* (n = 6) are shown in Fig. 3A. Cc-inx5 clusters with Innexin proteins that are expressed in the germ-line of *D. melanogaster* (zpg/inx4, inx5 and inx6) and *An. gambiae* (AGAP006241). Indeed, microarray and high-throughput expression data in *Drosophila* (Flyatlas; http://www.flyatlas.org) indicate that *inx5* and *inx6* are expressed only in the testes of adult males, whereas *zpg* expression is specific to the ovaries of adult females. In *An. gambiae*, the Innexin gene AGAP006241 is germ-line-specific but does not display sex-limited expression (Thailayil et al., 2011). To assess whether the medfly *inx5* orthologue (*Cc-inx5*) has germ-line and sex-specific transcriptional activity, its profile was analysed by RT-PCR in different adult body compartments/organs (Fig. 3B). The expected amplification product (1200 bp) was obtained exclusively in the testes and ovary samples, clearly indicating that the transcription of this medfly gene is restricted to the germ-line of both sexes. Notably, *Cc-inx5* is not transcribed in the male accessory glands. In addition, although Cc-inx5 shares higher identity to both *D. melanogaster* inx5 (2e-107, 56%) and inx6 (3e-114, 51%) than to zpg (3e-112, 48%), the genomic location of the *Cc-inx5* gene closely resembles that of *zpg* in *D. melanogaster* (Fig. S1).

A complete 1580 bp *Cc-inx5* mRNA sequence was amplified from cDNA derived from RNA extracted from the testes and the predicted CDS of 1134 bp (nucleotide positions 152 to 1285) putatively encodes a 377 aa protein. The predicted protein possesses all the features typical of the Innexin multiprotein family (Bauer et al., 2005).



3.3. *Cc-inx5 is essential for germ-line development*

A 243 bp fragment of the newly isolated *Cc-inx5* gene (nucleotide positions 640–882 in the mRNA sequence) was used as template to produce dsRNA (ds*Cc-inx5*) that was injected into the posterior pole of *C. capitata* embryos. To discriminate between potential lethal effects of *Cc-inx5* knockdown and mortality due to the microinjection procedure, a mock experiment was set up using a 486 bp fragment of the tGFP gene (ds*tGFP*). Survival rates were very similar between the ds*Cc-inx5* and ds*tGFP* injection experiments (Yates test, $\chi^2 = 1.773$; $P = 0.183$; Table 2).

Of the 1959 adults that emerged from ds*Cc-inx5*-injected embryos, 1042 (53%) were males; similarly, of the 992 adult flies that emerged from ds*tGFP*-injected embryos, 517 (52%) were males, indicating that *Cc-inx5* knockdown does not cause sex-specific lethality or sex-distortion effects.

To assess the effect of the injections on the morphology of the reproductive tract, the abdomens of the adults of both sexes were dissected. All the ds*tGFP*-injected individuals showed fully developed testes or ovaries; on the contrary, the majority of the ds*Cc-inx5*-treated flies showed abnormalities in the germ-line development (Fig. 4). In the interfered individuals, males lacked sperm and females lacked mature eggs, while the male and female accessory glands and the spermathecae appeared to be fully developed (Fig. S2).

Dissection of the spermathecae of WT females mated with sperm-less males further confirmed the absence of sperm transfer (Fig. S3), while protein SDS-PAGE uggestesuggested that the MAGs have very similar protein profiles in sperm-less and control males (Fig. S4).

The extent of *Cc-inx5* knockdown in both sexes was assessed by Real Time qPCR analyses using RNA extracted from the whole abdomens of 4 day-old adult flies. *Cc-inx5* transcription levels were separately evaluated for 5 males and 5 females injected with ds*Cc-inx5* and showing no germ-line (as determined by dissection before RNA extraction) and compared to the transcription levels of five ds*tGFP*-injected males or females (Fig. S5). The *Cc-inx5* transcription levels were significantly reduced (one sample *t*-test males: mean = 0.257; $t_4 = 7.1612$, $P = 0.002$; females: mean = 0.408; $t_4 = 5.6354$, $P = 0.0049$). The transcription levels of *Ccinx-2* and *Ccinx-3* were not statistically different between test and control flies, indicating that the dsRNA injected was specific for *Ccinx-5* (one sample *t*-test; *Ccinx-2* males: mean = 1.308; $t_4 = 1.053$, $P = 0.352$; females: mean = 2.319; $t_4 = 2.508$, $P = 0.066$. *Ccinx-3* males: mean = 1.636; $t_4 = 2.508$, $P = 0.051$; females: mean = 1.310; $t_4 = 0.846$, $P = 0.445$).

The strong reduction of *Ccinx5* transcription was further confirmed by whole mount *in situ* hybridization on dissected testes from control and sperm-less males (Fig. S6). This result showed also that in WT males the transcription of *Ccinx5* was limited to the cysts containing developing sperm. A similar pattern of transcription was detected also for the germ-line specific gene *CcVasa*.

Fig. 1. Mating with WT males reduces female remating tendency. A) Percentage of females that remate at different time-points after the first mating in a 4 h period. Mean and SD of three biological replicates are shown, using 25 females for each replicate at each time point analysed; letters indicate that the difference between groups of females is statistically significant. **B)** Number of sperm (square root-transformed) in the spermathecae (30 females) at different time points after first mating. Letters indicate that the difference between groups of females is statistically significant. **C)** Copula duration (square root-transformed) of the first mating (blue) and rematings at different time points after first *copula*. The asterisk indicates the difference in duration between the first mating and remating after 1d. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

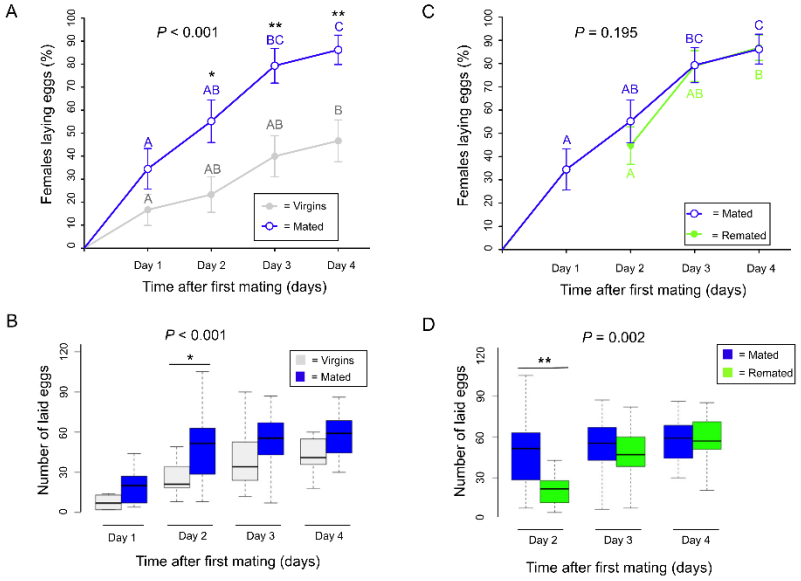


Fig. 2. Mating induces oviposition in medfly females. For all the experiments, results using 30 females for each group are presented. **A)** Log-rank analysis of the percentage of females laying eggs after mating (blue) or as virgins (grey). The *P* value of the test is shown. Letters indicate if the difference of females laying eggs from the same group is statistically significant between the time points analysed. Asterisks, instead, indicate hwhether a significant difference in the percentage of females laying eggs between groups on the same day exists (* – P < 0.05; ** – P < 0.01). **B)** Number of eggs laid by females, at the same time-points as above. Eggs were counted for each female every 24 h. The asterisk indicates that the numbers of eggs deposited at day 2 by virgin and mated females are significantly different. **C)** Log-rank analysis of the percentage of females laying eggs after one mating (blue) or after two matings (remating occurred 1 day after the first copula; green). Letters are as in panel A. **D)** Number of eggs laid by single or double mated females. Asterisks indicate that the number of eggs deposited at day 2 (day 1 after remating) is significantly different. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
C. capitata predicted protein sequences that share significant similarity to *D. melanogaster* Innexin proteins.

C. capitata innexins	Accession number	*D. melanogaster* best hit	Query coverage (%)	e-value	Identity (%)
Cc-inx1	XP_004521052.1	*ogre/inx1*	100	3e-144	90
Cc-inx2	XP_004521056.1	*inx2*	100	5e-173	92
Cc-inx3	XP_004530094.1	*inx3*	100	5e-155	79
Cc-inx5	XP_004525700.1	*inx5*	97	2e-107	56
Cc-inx7	XP_004521054.1	*inx7*	94	1e-123	61

3.4. *Sperm-less males are able to induce post-mating behavioural/physiological changes in females*

Sperm-less males were mated to WT females to assess their mating ability and sterility. The same experimental conditions were applied to control assays involving *dstGFP* injected males.

Sperm-less males were as able to mate with WT females as control males, but there was a significant difference in *copula* duration (Unpaired *t*-test, $t_{676} = 3.322$, $P = 0.0009$) which was, on average, 9 min longer for sperm-less males (166 ± 35 SD min) compared to control males (157 ± 32 SD min) (Fig. 5A).

Based on the behavioural/physiological baseline analysis on WT males described above, we assessed the ability of sperm-less males to elicit refractoriness to remating in females at the following three time-points: 6 h, 1 d and 3 d after the first mating. Data from three experimental replicates indicated that female refractoriness to remating is induced by sperm-less and control males in a similar manner (Fig. 5, Table S8).

Oviposition rates were calculated for females mated to sperm-less or control males (Fig. 6). Eggs were collected daily across an oviposition period of 6 d post mating. As expected, a higher percentage of females mated to either sperm-less or control males laid eggs with respect to virgin females, at all tested time-points (Log-Rank Mantel-Cox test, control vs. virgin $\chi^2 = 36.89$; $P < 0.0001$; sperm-less vs. virgin $\chi^2 = 52.80$; $P < 0.0001$). Interestingly, a higher percentage of females mated to sperm-less males laid eggs

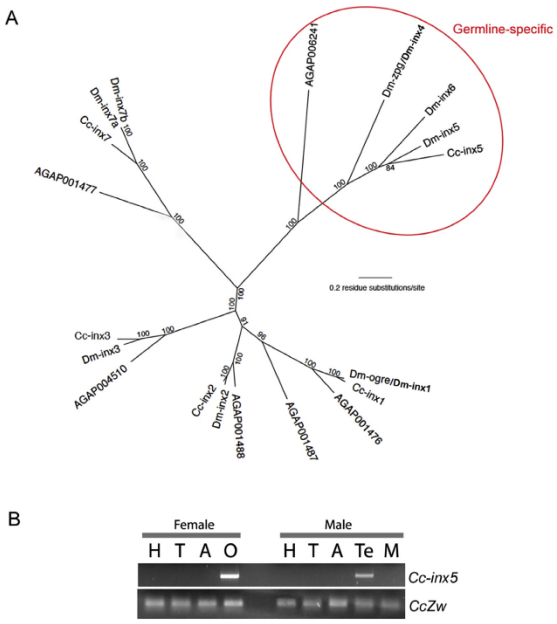


Fig. 3. Phylogenetic relationships and transcriptional profile of *Cc-inx5*. **A)** Maximum Likelihood tree of innexin proteins from *C. capitata*, *D. melanogaster* and *An. gambiae*. Bootstrap values (percentage of 1000 replications) are shown. The scale represents the number of amino acid substitutions per site. **B)** Transcriptional profiles of *Cc-inx5* in different body compartments of medfly adult males and females as determined by RT-PCR. *CcZw* (Glucose-6-phosphate dehydrogenase) was used as control (H – head, T – thorax, A – abdomen without reproductive tissues, O – ovaries, Te – testes, M – male accessory glands).

Table 2
Hatching, pupation and adult eclosion rates estimated from medfly embryo injections using ds*Cc-inx5*.

	Embryos injected	Hatching rate (%)	Pupation rate (%)	Adults (%)	Males	Females
dsCc-inx5	22,124	5089 (23)	2205 (10)	1959 (9)	1042	917
dstGFP	10,654	2696 (25)	1125 (10)	992 (9)	517	475

compared to females mated to control males (Fig. 6A, Log-Rank Mantel Cox test, $\chi^2 = 5.04$; $P = 0.025$), particularly at days 2, 3 and 4 after mating (Fisher's exact test 2 days $P = 0.028$; 3 days $P = 0.025$, 4 days $P = 0.049$). This difference is also evident comparing the trend of the two curves following the point by point comparison using Fisher's test (represented with letters in Fig. 6A): while the curve of females mated with control males can be easily divided in three steps (one including day 1, the second including day 2, 3 and 4 and the third with day 5 and 6), the curve of females mated with sperm-less males showed two steps only (one including day 1 and the second including all the remaining days).

Both sperm-less and control males induced the oviposition of a greater number of eggs compared to virgin females (Tukey post-hoc test after two-way ANOVA, $P < 0.001$, Table S9). However, females mated to sperm-less males laid a higher number of eggs than those mated to control males (Tukey post-hoc test after two-way ANOVA, $P < 0.001$, Fig. 6B, Table S10), particularly on days 2 and 4 after mating (Tukey post-hoc test after ANOVA; day 2: $P = 0.028$; day 4: $P = 0.004$ Fig. 6B, Table S10).

3.5. Effect of mating with sperm-less males on female transcriptional activity

We finally examined whether mating with sperm-less males elicited different responses in mated females at the transcriptional level. To do so, we determined the relative transcription levels of five genes in females mated to either sperm-less or control males at three time-points: 6 h, 1 d and 3 d after mating. We selected the genes to investigate based on available data from other insect species. Mating, indeed, induces transcriptional changes in females of many insect species, including *Anopheles*, *Drosophila* and *Ceratitis* (Gabrieli et al., 2014; Gomulski et al., 2012; McGraw et al., 2004,

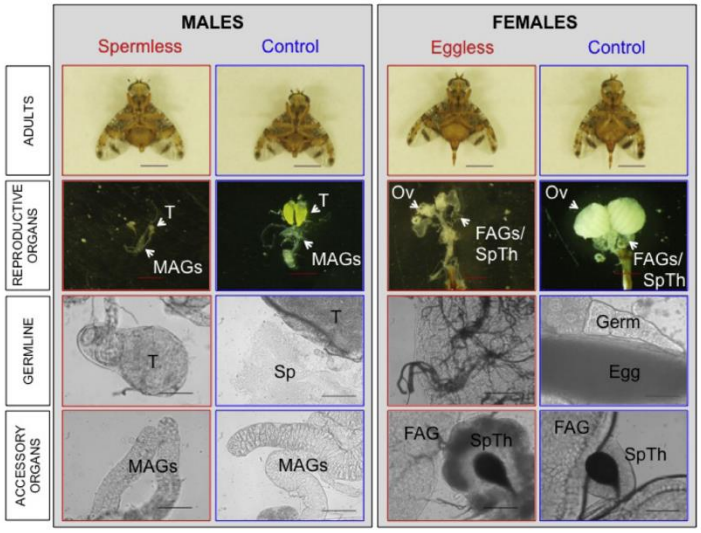


Fig. 4. Adult flies and the corresponding dissected reproductive tracts. Panels referred to *dsCc-inx5*-injected individuals are marked in red, those referred to controls in blue. First row: from the left, *dsCc-inx5*-injected and control (*dsGFP*-injected) medfly males and females, respectively. Scale bar: 2 mm. Second row: corresponding dissected reproductive tracts. Scale bar = 0.5 mm. Third row: Higher magnification of testes and ovaries. Scale bar: 0.1 mm. Fourth row: accessory reproductive tissues (glands and spermathecae). Scale bar: 0.1 mm. Abbreviations: T – Testis; Ov – Ovaries; Sp – Sperm; SpTh – Spermathecae; FAGs – Female Accessory Glands; MAGs – Male Accessory Glands; Germ – Germline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2008; Rogers et al., 2008). Different male contributions to mating (sperm, Acps, and other components) have distinct effects on *Drosophila* female gene expression (McGraw et al., 2004), while lack of sperm transfer does not alter gene expression changes in *Anopheles* females (Thailayil et al., 2011).

In both *D. melanogaster* and *An. gambiae* (Dalton et al., 2010; Gabrieli et al., 2014; McGraw et al., 2008; Rogers et al., 2008), mating changes the transcriptional activity of three genes in the female abdomen (*Vitellogenin, Vg; major royal jelly, mrj; flightin, fln*) and of two genes in the female head (*dopamine N-acetyl transferase, dat; larval serum protein 2, lsp2*).

As expected, mating of WT virgin females with control males induced changes in the transcription of these five genes at different time-points after mating (marked with blue asterisks in Fig. 7. One sample *t*-test; *Vg* 3d $t_3 = 3.570$, $P = 0.0376$; *fln* 1d $t_3 = 3.214$, $P = 0.0488$; *dat* 6 h $t_3 = 31.858$, $P < 0.0001$; *lsp2* 1d $t_3 = 10.003$, $P = 0.0021$, 3d $t_3 = 10.212$, $P = 0.0020$). Mating with sperm-less males also induce changes in the transcription of these genes compared to virgin levels (marked with red asterisks in Fig. 7. One sample *t*-test; *Vg* 6 h $t_3 = 3.288$, $P = 0.0462$; *mrj* 1d $t_3 = 9.405$, $P = 0.0025$; *fln* 6 h $t_3 = 4.966$, $P = 0.0157$; *dat* 6 h $t_3 = 3.197$, $P = 0.050$; *lsp2* 1d $t_3 = 7.007$, $P = 0.0060$). The expression of *Vg*, *mrj* and for *lsp2*, however, was different mated either with sperm-less or with control males and these differences were statistically significant (marked with black asterisks in Fig. 7, Table S11; Tukey post-hoc test after two-way ANOVA: *Vg* 6 h $P = 0.0119$; *mrj* 1d $P = 0.0001$; *lsp2* 6 h $P = 0.0090$, 1d $P < 0.0001$, 3d $P < 0.0001$).

4. Discussion

In this study, a germ-line specific medfly innexin gene, *Cc-inx5*, was identified that shares a high similarity with *D. melanogaster inx5*. The knockdown of *Cc-inx5* during embryonic development resulted in the generation of sperm-less males with fully developed male accessory glands. The availability of these sperm-less males enabled us to investigate the role of sperm in inducing behavioural/physiological changes in mated females.

4.1. Mating temporally inhibits female receptivity and induces oviposition and fecundity

In normal conditions (with WT control males), mating induces a temporal inhibition of female remating from 3 h after mating that lasts for at least 3 days in the well-established ISPRA laboratory strain. However, from day 6 females display no further inhibition and start to remate at a frequency similar to that of virgins. This phenomenon seemed to be correlated to the presence of sperm in the spermathecae, as their numbers are strongly reduced in females 6 days after the first mating, accordingly to previous studies (Twig and Yuval, 2005). If this was true, mating with sperm-less males should not fully induce re-mating inhibition.

The absence of remating inhibition at 6 days post-mating appears to be in contrast with previous results reporting a continuous and strong inhibition of remating in female medfly (Miyatake et al., 1999); however, in that study females were continuously exposed to males and mating events were recorded in a 30 min interval. In

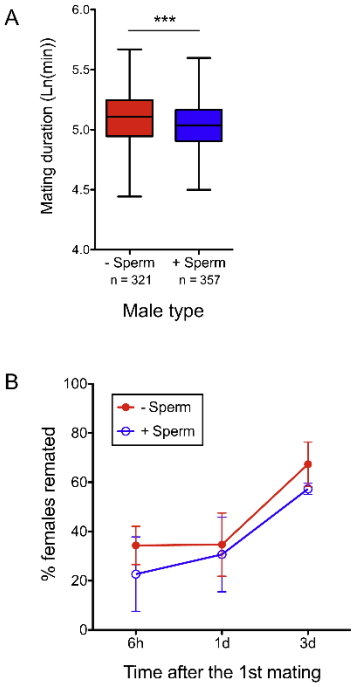


Fig. 5. Sperm-less males induce remating inhibition. A) Mating duration (logarithmic-transformed) of couples formed with sperm-less (red) or control (blue) males. Asterisks indicate that the difference is statistically significant. **B)** Proportion of twice-mated females mated first with sperm-less (red) or control (blue) males and exposed to virgin WT at 6 h, 1 and 3 days after the first mating. Results are from three biological replicates using 25 females for each group and each at time point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

our study, couples were instead removed from the cage and females were maintained without males until re-exposure for the remating assays. The experimental design we adopted has been already used for studies on medfly mating behaviour (Gavriel et al., 2009) and it closely resembles the receptivity assay used in *Drosophila* (Chapman et al., 2003) and in *An. gambiae* (Gabrieli et al., 2014; Rogers et al., 2009), allowing a tight control of female mating status at each time point analysed.

As expected, mating also induces oviposition and fecundity in females. However, remating does not result in an increase in these parameters, suggesting that the effect of mating on oviposition and fecundity is not cumulative. On the contrary, remating appears to temporally inhibit female fecundity the day after the second *copula*. The reason for this is currently unknownhowever. However, given that the entry of sperm in the storage organs requires time (Bertin et al., 2010), as reported also in *Drosophila* (Gilbert, 1981), ovulation might be temporally inhibited to avoid sperm loss.

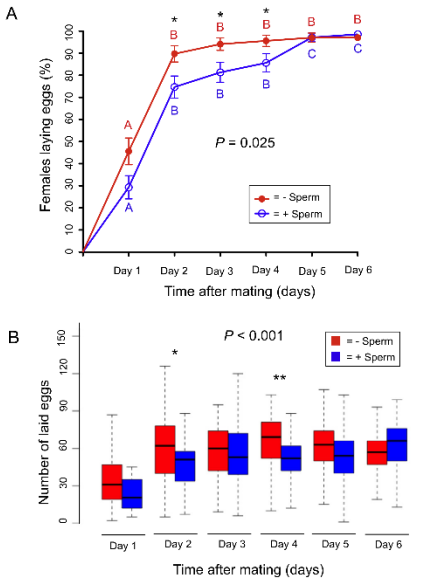


Fig. 6. Mating with sperm-less males induces higher oviposition levels in medfly females. For all the experiments, results using 75 females for each group are presented. **A)** Long-rank analysis of the percentage of females laying eggs mating with sperm-less males (red) or with control males (blue). Asterisks indicate significant difference in the percentage of females laying eggs between those mated to sperm-less and control males. Letters, instead, indicate if the difference of females laying eggs from the same group is statistically significant between the time points analysed. **B)** Number of eggs laid by females. Eggs were counted for each female every 24 h. Asterisks indicate significant difference in the numbers of eggs laid by females mated to sperm-less and control males. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Sperm-less males reduce female receptivity

Our data did not show any significant difference in terms of remating inhibition in females mated to sperm-less or control males, at the three time points analysed. It isevident that the temporal profile of female receptivitydiffers slightly from that reported in Fig. 1A using wild-type flies and from that previously reported (Gavriel et al., 2009). The different genetic background of the strain used in this and previous studies might explain this effect; in this case, although the trend of the female receptivity is similar, the timeframe is different. In the case of the differences observed in thepresent work, they might have been induced by the manipulation of the flies used in the experiment, as they wewere all microinjected at the embryonic stage, which might have influencedinfluenced male reproductive fitness. More importantly, the experimental flies were reared on a different larval diet (compared to the wild-type flies) to facilitate manipulation and assessmentassessment of the rearing conditions, and this might have altered male fitness and sexual performance, as previously reported in the medfly (Gavriel et al., 2009), in mosquitoes (Yahouedo et al., 2014), and in other fruit flies (Aluja et al., 2009).

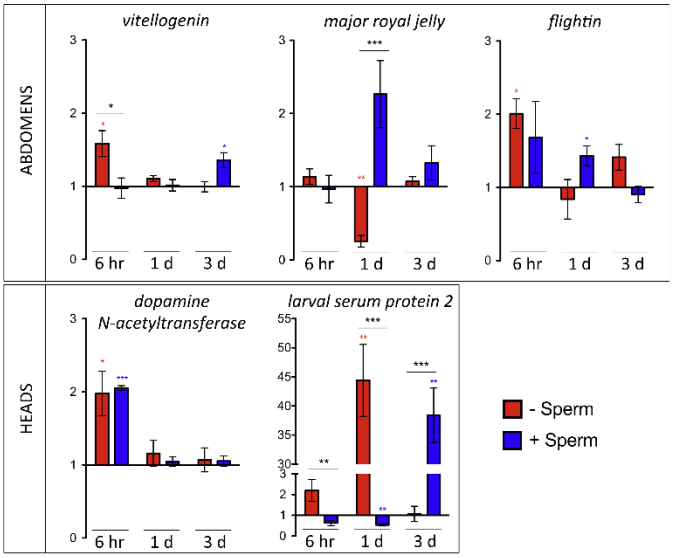


Fig. 7. Differential transcript levels of five genes in mated females at different time points. Transcript abundances were determined in the abdomens (upper row) and heads (lower row) of females mated to sperm-less (red bars) or to control males (blue bars). The analyses were performed on four biological replicates of pools of tissues from three females for each group at 6, 24 and 72 h after mating. Values are presented as relative to expression levels in virgins (transcriptional level = 1). Coloured asterisks indicate significant difference in transcript abundances between mated (either with sperm-less -red asterisks- or with control males -blue asterisks-) and virgin females. Black asterisks indicate significant difference between females mated with sperm-less or with control males. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As the absence of sperm in the spermathecae does not alter the remating tendency of the females mated with sperm-less males, we reject the hypothesis that sperm is necessary to inhibit remating. The initial observation that the percentage of remated females increases as number of sperm stored in the spermathecae diminishes may be an "artefact". It is, indeed, conceivable that the action of other male factors that inhibit female remating has evolved to last for the length of time required by females to utilize the majority of stored spermatozoa to fertilize her eggs.

In the literature it was previously reported that i) surgical castration (Miyatake et al., 1999); ii) low sperm numbers transferred to the female (Bloem et al., 1993; Gavriel et al., 2009; Katiyar and Ramirez, 1970; Kraaijeveld and Chapman, 2004; Nakagawa et al., 1971; Vera et al., 2003); and iii) irradiation, all result in inefficient female remating inhibition (Gavriel et al., 2009; Katiyar and Ramirez, 1970). Our data are only apparently in contrast with these findings that support a direct sperm effect on medfly female remating inhibition. Firstly, comparison with the results derived from surgically castrated males (Miyatake et al., 1999) is difficult due to the profound differences in the protocols used to test female receptivity, as previously discussed. However, one possible explanation is that in surgically castrated males the testes are completely removed, both in terms of germ-line and somatic lineage. Furthermore, the MAGs and testes are closeclosely connected and surgical castration might have damaged the MAGs.

Secondly, the amount of sperm transferred and stored during the first mating has been proposed to be a key factor associated with short-term (24 h after first mating) renewal of receptivity (Mossinson and Yuval, 2003). However, it is important to consider that, in *Drosophila* at least, the amount of ejaculate is not fixed and the male can adjust its seminal fluid components by increasing key proteins, for example in response to sperm competition (Wigby et al., 2009). We cannot exclude the opposite, i.e. that a reduced amount of sperm transferred by the medfly male may be accompanied by a reduced quantity of seminal fluid, thus explaining the higher female remating rates.

Finally, to date there is no information regarding the effect of irradiation on the quality and quantity of SFPs, thus the inefficiency of irradiated medfly males in inducing female remating inhibition (Bloem et al., 1993; Mossinson and Yuval, 2003; Seo et al., 1990) might be due not only to sperm damage, but also to other fitness-related factors. The quantity and/or quality of transferred SFPs, indeed, plays a key role in fertility (LaFlamme et al., 2012) and further research is needed to assess the impact of irradiation on medfly seminal fluid components other than sperm.

The results we report reveal a marked difference between *Drosophila* and *Ceratitis*. In *D. melanogaster* the mating effect on female behaviour lasts for at least 5 days (Manning, 1962), while sperm-less males induce a shorter effect, lasting 1–2 days (Kalb et al., 1993). A similar effect can be achieved by the injection of purified SP, which is produced in the MAGs (Chen et al., 1988; Hihara, 1981). Furthermore, mating with MAG-less males is not

able to induce the switch of female behaviour, but the prolonged effect of MAG secretions requires the transfer of sperm (Xue and Noll, 2000). Sperm is indeed necessary for the correct transfer and retention of SP (Liu and Kubli, 2003; Peng et al., 2005). In contrast, in *Ceratitis* the inhibition of female receptivity is similar in both females mated to sperm-less or to control males, at all the time points analysed, thus excluding the presence of a *Drosophila*-like "sperm effect". In this regard, the medfly is similar to *An. gambiae*, in which sperm-less males fully elicit the female post-mating response (Thailayil et al., 2011).

4.3. Sperm-less males induce oviposition and female fecundity

Mating with sperm-less males induced females to continuously oviposit for days after mating. Thisunderlines fundamental differences with *Drosophila*, in which sperm-less males are able to induce oviposition only transiently (Manning, 1962, 1967). In *Drosophila*, oogenesis and oviposition are indeed stimulated by the transfer of SP, Ovulin and DUP99B during mating (Chapman et al., 2003; Herndon and Wolfner, 1995; Liu and Kubli, 2003), but sperm-less males are not able to retain SP for a prolonged time (Liu and Kubli, 2003). Conversely, in *An. gambiae*,as in the medfly, sperm-less males are able to induce continuous oviposition (Thailayil et al., 2011). Both in *Drosophila* and in *Anopheles* mating induces the remodelling of the tissues of the reproductive organs, likely preparing the female for oviposition and fertilization (Kapelnikov et al., 2008; Rogers et al., 2008); whether this also occurs in *Ceratitis* andwhether it is correlated to the presence of spermatozoa or to the action of other male factors is still unknown.

In *An. gambiae* the male-transferred 20-hydroxyecdysone (20E) hormone stimulates oviposition (Gabrieli et al., 2014) and, once set in motion, oviposition occurs after every blood meal. Interestingly, the ecdysteroid levels in *Drosophila* females also increase after mating (Harshman et al., 1999) and ovarian ecdysteroid levels are induced by the action of SP, but independently from its known receptor (SPR) (Ameku and Niwa, 2016). These increased hormonal levels, in turn, stimulate gametogenesis by increasing the number of germline stem cells (GSCs) in the ovaries. Also in *Anopheles*, the direct transfer of 20E induces oogenesis through the action of a female Mating-Induced Stimulator of Oogenesis (MISO) protein (Baldini et al., 2013). In the medfly, oogenesis is also enhanced after mating and sperm transfer is not required for this effect, suggestingsuggesting that a male factor, most likely MAG-derived, is involved. We expect that this MAG-derived male factor enhances female gametogenesis possibly through the manipulation of hormonal signals, as in *Drosophila* and in *Anopheles*.

The effect of mating on female physiology, however, is not limited to the reproductive tissues and central nervous system, but it involves many other organs. Recently, it has been demonstrated that mating profoundly remodels the *Drosophila* female midgut (Reiff et al., 2015), preceding the increased nutrient intake required for offspring production. This tissue remodelling is under hormonal control, but in this case the juvenile hormone (JH), which is secreted by the *corpora allata* under the stimulation of SP, plays a fundamental role (Moshitzky et al., 1996). It is therefore likely that the mating effect in the medfly would be pleiotropic and under a sperm-independent male signal and that the female physiological switch we measured is the result of the complex and harmonic regulation of these effects. However, when elements of this fine regulation are manipulated, like, for example, the presence/absence of spermatozoa, the system might show irregularirregular responses. This is probably the case for the increased oviposition and fecundity we observed when sperm was not transferred. In this case, the absence of sperm has probably caused a misregulation of one or more pathways that, involved in the females, respond to thefemale's mating response, that are related to fecundity and oviposition.

4.4. Sperm-less males engage in longer matings

Our data also show that the lack of sperm transfer by sperm-less males results in a longer mating duration compared to matings with control males. The difference is minimal (9 min) compared to the total length of the *copula* (>2 h), but given that the process duringaffected mating phase (seminal fluid transfer, sperm transfer, etc …) that is affected, not allowing to is unknown, we cannot draw any conclusions about the biological significance of this difference. The relationship between sperm transfer to the spermathecae and *copula* duration, however, is not linear. We found differences in *copula* duration only when females remated 24 h after the first mating, and this is in accordance with our previous results (Bertin et al., 2010). The amount of sperm stored in the spermathecae cannot explain this difference; if that were the case, the remating duration should have been shorter than that of the first mating at least on the first three days after mating. Similarly, it was previously reported that no relationship exists between *copula* duration and the quantity of sperm transferred to the female (Costa et al., 2012; Taylor et al., 2000). Like *Drosophila* (Bairati, 1968; Lung and Wolfner, 2001), the transfer of seminal fluid in the medfly occurs before sperm transfer (which begins 90 min after *copula* initiation (Yuval et al., 1996)), possibly to create an optimal environment in the female sperm storage organs that will receive the spermatozoa (Marchini et al., 2003), like it occurs in *Drosophila* (Bairati, 1968; Lung and Wolfner, 2001). It is possible that the longer *copula* duration detected for the sperm-less males could be related to their attempt to transfer as much seminal fluid as possible and/or to increase the opportunities to influence the way in which their ejaculate is stored. One more option might be that the males can perceive the quantity of sperm transferred and given that they detect the passage of no sperm, they prolong copula duration in an attempt to increase the number of sperm transferred. However, in the Queensland fruit fly, *Bactrocera tryoni*, copula duration has been suggested to be under greater control from females than males (Collins et al., 2012) and this could be true in the medfly too. Furthermore, this observation might reflect a potential role of sperm in providing the female with a signal for *copula* termination.

4.5. Mating with sperm-less males alters gene transcription in females

Mating induces a profound changes in gene expression in *Drosophila* females (Lawniczak and Begun, 2004; Mack et al., 2006) and distinctdistinct subsets of genes are controlled by sperm, fluid from the MAGs or other male-derived factors (Gioti et al., 2012; McGraw et al., 2004, 2008). Similarly, also in *Anopheles* mating induces transcriptional changes (Gabrieli et al., 2014; Rogers et al., 2008), but sperm seems not to be necessary for this control (Thailayil et al., 2011). We detected mating-induced transcriptional changes in the *Ceratitis* female abdomen and head, and some of thesethese changes are altered when females do not receive sperm during the mating.

The changes in gene transcription between females mated with sperm-less or control males reflect the behavioural differences previously reported in this study.

No differences were detected in the transcription of genes that are related to the inhibition of remating. The *dopamine N-acetyl transferase* gene was up-regulated in mated female heads compared to virgin females, but no differences were detected between sperm-less or control male mated females. Dopamine is required for female receptivity to courtship (Neckameyer, 1998)

and dopamine N-acetyltransferase converts dopamine to N-acetyl dopamine, providing a mechanism to reduce dopamine activity. Similarly, *flightin* is up-regulated in the abdomen of mated females; this gene is involved in muscle contraction, and the changes in its expression may contribute to induce morphological changes in the female's reproductive tract, as described in *Drosophila* (Adams and Wolfner, 2007).

On the contrary, genes that might be related to oviposition showed significant differences in transcription in females mated with sperm-less or control males. *Vitellogenin* was highly transcribed soon after mating in the abdomen of females mated to sperm-less males, while its up-regulation occurred later in females mated to control males. Vitellogenin is the major component of the yolk of insect eggs (Gelti-Douka et al., 1974), and its transcriptional profile is congruent with the observation that sperm-less males induce higher oviposition rates sooner after mating compared to matings with control males. A similar pattern of transcription was detected in the female heads for the *larval serum protein 2* gene. *Lsp2* encodes nutrient reservoirs that function as fat body storage of amino acids (Liu et al., 2009).

Surprisingly, the *major royal jelly* gene, which is homologous to the *Drosophila yellow-g*, is down-regulated in female abdomens mated to sperm-less males, while it appears to be slightly (but just not significantly) up-regulated in females mated to control males. This gene is expressed in follicle cells and is required for eggshell formation during oogenesis; it is possible to speculate that the absence of sperm induces females to invest less in the proper formation of the eggs.

5. Conclusions

The data we present in this study suggest that sperm transfer is not necessary to induce female remating inhibition, fecundity and oviposition. However, the absence of sperm transfer probably results in the misregulation of mating-responsive pathways in females, such that females mated with sperm-less males increased oviposition.

Future studies are required to assess the molecular mechanisms by which inhibition and oviposition are induced in the medfly, which will likely involve the identification of the medfly SFPs. It will be interesting to determine whether medfly SFPs exist that are bound to the sperm tail and are thus retained in the female for a prolonged period, as is the case for the *Drosophila* SP (Peng et al., 2005).

These findings also represent the basis for future studies aimed at improving current population-control strategies, including the Sterile Insect Technique (SIT). We have shown, for example, the importance of further investigations on the effect of irradiation, currently used to produce sterile males, on the male accessory glands and the activity and integrity of SFPs.

By understanding the molecular mechanisms responsible for the inhibition of female remating and the behavioural switch to oviposition hosts, it may be eventually possible to enhance the fitness of sterile males used in the SIT.

Acknowledgements

This work was partially funded by the FAO/IAEA (Food and Agriculture Organization of the United Nations (FAO)/International Atomic Energy Agency (IAEA)) Technical Contract No. 16966 (GG), a PRIN grant from the Italian Ministry of Education, University and Research (MIUR) (20077 RCHRW) (LMG), and Cariplo-Regione Lombardia 'IMPROVE' (FS). The work was also carried out within the FAO/IAEA research CRP programme "Identification of molecular and transgenic tools for sexing strategies and development of strain and sperm marking systems for fruitfly pests" - "Development and evaluation of improved strains of insect pests for SIT" (GG). The authors wish to thank Ioannis Livadaras and Alfred M. Handler for technical advice and Flaminia Catteruccia and Andrea Crisanti for initial discussion and help on this project.

Appendix ASupplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ibmb.2016.10.002.

References

Adams, E.M., Wolfner, M.F., 2007. Seminal proteins but not sperm induce morphological changes in the *Drosophila melanogaster* female reproductive tract during sperm storage. J. Insect Physiol. 53, 319–331.
Aluja, M., Rull, J., Sivinski, J., Trujillo, G., Perez-Staples, D., 2009. Male and female condition influence mating performance and sexual receptivity in two tropical fruit flies (Diptera: Tephritidae) with contrasting life histories. J. Insect Physiol. 55, 1091–1098.
Ameku, T., Niwa, R., 2016. Mating-induced increase in germline stem cells via the neuroendocrine system in female Drosophila. PLoS Genet. 12, e1006123.
Avila, F.W., Sirot, L.K., LaFlamme, B.A., Rubinstein, C.D., Wolfner, M.F., 2011. Insect seminal fluid proteins: identification and function. Ann. Rev. Entomol. 56, 21–40.
Bairati, A., 1968. Structure and ultrastructure of the male reproductive system of *Drosophila melanogaster* Meig. 2. The genital duct and accessory glands. Mon. Zool. It 2, 105–182.
Baldini, F., Gabrieli, P., South, A., Valim, C., Mancini, F., Catteruccia, F., 2013. The interaction between a sexually transferred steroid hormone and a female protein regulates oogenesis in the malaria mosquito *Anopheles gambiae*. PLoS Biol. 11, e1001695.
Bauer, R., Loer, B., Ostrowski, K., Martini, J., Weimbs, A., Lechner, H., Hoch, M., 2005. Intercellular communication: the *Drosophila* innexin multiprotein family of gap junction proteins. Chem. Biol. 12, 515–526.
Bergh, J.C., Harris, M.O., Rose, S., 1992. Factors inducing mated behaviour in female Hessian flies (Diptera: cecidomyiidae). Ann. Entomol. Soc. Am. 85, 224–233.
Bertin, S., Scolari, F., Guglielmino, C.R., Bonizzoni, M., Bonomi, A., Marchini, D., Gomulski, L.M., Gasperi, G., Malacrida, A.R., Matessi, C., 2010. Sperm storage and use in polyandrous females of the globally invasive fruitfly, *Ceratitis capitata*. J. Insect Physiol. 56, 1542–1551.
Bloem, S., Bloem, K., Rizzo, N., Chambers, D., 1993. Female medfly refractory period: effect of first mating with sterile males of different sizes. In: Aluja, M., Liedo, P. (Eds.), Fruit Flies. Biology and Management. Springer-Velag, New York, NY, USA, pp. 191–192.
Bonizzoni, M., Gomulski, L.M., Bertin, S., Scolari, F., Guglielmino, C.R., Yuval, B., Gasperi, G., Malacrida, A., 2007. Unfaithful Mediterranean fruit fly Ceratitis capitata: impact on the SIT? In: Vreysen, M.J.B., Robinson, A.S., Hendrichs, J. (Eds.), Area-wide Control of Insect Pests. Springer, Dordrecht, pp. 175–182.
Bonizzoni, M., Gomulski, L.M., Mossinson, S., Guglielmino, C.R., Malacrida, A.R., Yuval, B., Gasperi, G., 2006. Is polyandry a common event among wild populations of the pest *Ceratitis capitata*? J. Econ. Entomol. 99, 1420–1429.
Bonizzoni, M., Katsoyannos, B.I., Marguerie, R., Guglielmino, C.R., Gasperi, G., Malacrida, A., Chapman, T., 2002. Microsatellite analysis reveals remating by wild Mediterranean fruit fly females, *Ceratitis capitata*. Mol. Ecol. 11, 1915–1921.
Boratyn, G.M., Schaffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., Madden, T.L., 2012. Domain enhanced lookup time accelerated BLAST. Biol. Direct 7, 12.
Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622.
Chapman, T., Bangham, J., Vinti, G., Seifried, B., Lung, O., Wolfner, M.F., Smith, H.K., Partridge, L., 2003. The sex peptide of *Drosophila melanogaster*: female post-mating responses analyzed by using RNA interference. Proc. Natl. Acad. Sci. U. S. A. 100, 9923–9928.
Chen, P.S., Stumm-zollinger, E., Aigaki, T., Balmer, J., Bienz, M., Bohlen, P., 1988. A male accessory-gland peptide that regulates reproductive-behavior of female *Drosophila Melanogaster*. Cell 54, 291–298.
Collins, S.R., Perez-Staples, D., Taylor, P.W., 2012. A role for copula duration in fertility of Queensland fruit fly females mated by irradiated and unirradiated males. J. Insect Physiol. 58, 1406–1412.
Costa, A.M., Anjos-Duarte, C.S., Roriz, A.K.P., Dias, V.S., Joachim-Bravo, S., 2012. Male diet and age influence to inhibit female remating in *Ceratitis capitata* (Diptera: Tephritidae). J. Appl. Entomol. 136, 456–463.
Dalton, J.E., Kacheria, T.S., Knott, S.R.V., Lebo, M.S., Nishitani, A., Sanders, L.E., Stirling, E.J., Winbush, A., Arbeitman, M.N., 2010. Dynamic, mating-induced gene expression changes in female head and brain tissues of *Drosophila melanogaster*. BMC Genomics 11, 541.
Davies, S.J., Chapman, T., 2006. Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (*Ceratitis capitata*). Insect biochem. Mol. Biol. 36, 846–856.

Eberhard, W., 2000. Sexual behavior and sexual selection in the Mediterranean fruit fly, Ceratitis capitata (Dacinae: ceratitidini). In: Aluja, M., Norrbom, A. (Eds.), Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior. CRC Press., Boca Raton, Florida, USA, pp. 457–489.

Gabrieli, P., Kakani, E.G., Mitchell, S.N., Mameli, E., Want, E.J., Mariezcurrena Anton, A., Serrao, A., Baldini, F., Catteruccia, F., 2014. Sexual transfer of the steroid hormone 20E induces the postmating switch in Anopheles gambiae. Proc. Natl. Acad. Sci. U. S. A. 111, 16353–16358.

Gabrieli, P., Scolari, F., 2016. Delivery of nucleic acids through embryo microinjection in the worldwide agricultural pest insect, Ceratitis capitata. JOVE 116, e54528.

Gasperi, G., Bonizzoni, M., Gomulski, L.M., Murelli, V., Torti, C., Malacrida, A.R., Guglielmino, C.R., 2002. Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*. Genetica 116, 125–135.

Gavriel, S., Gazit, Y., Yuval, B., 2009. Remating by female Mediterranean fruit flies (*Ceratitis capitata*, Diptera: Tephritidae): temporal patterns and modulation by male condition. J. Insect Physiol. 55, 637–642.

Gelti-Douka, H., Gingeras, T.R., Kambysellis, M.P., 1974. Yolk proteins in *Drosophila*: identification and site of synthesis. J. Exp. Zool. 187, 167–172.

Gilbert, D.G., 1981. Ejaculate esterase 6 and initial sperm use by female *Drosophila melanogaster*. J. Insect Physiol. 27, 641–650.

Gillot, C., Langley, P.A., 1981. The control of receptivity and ovulation in the tsetse fly, *Glossina morsitans*. Physiol. Entomol. 6, 269–281.

Gioti, A., Wigby, S., Wertheim, B., Schuster, E., Martinez, P., Pennington, C.J., Partridge, L., Chapman, T., 2012. Sex peptide of *Drosophila melanogaster* males is a global regulator of reproductive processes in females. Proc. R. Soc. B 279, 4423–4432.

Gomulski, L.M., Dimopoulos, G., Xi, Z., Scolari, F., Gabrieli, P., Siciliano, P., Clarke, A.R., Malacrida, A.R., Gasperi, G., 2012. Transcriptome profiling of sexual maturation and mating in the Mediterranean fruit fly, *Ceratitis capitata*. PLoS ONE 7, e30857.

Gomulski, L.M., Dimopoulos, G., Xi, Z., Soares, M.B., Bonaldo, M.F., Malacrida, A.R., Gasperi, G., 2008. Gene discovery in an invasive tephritid model pest species, the Mediterranean fruit fly, Ceratitis capitata. BMC Genomics 9, 243.

Harshman, L.G., Loeb, A.M., Johnson, B.A., 1999. Ecdysteroid titers in mated and unmated *Drosophila melanogaster* females. J. Insect Physiol. 45, 571–577.

Haussmann, I.U., Hemani, Y., Wijesekera, T., Dauwalder, B., Soller, M., 2013. Multiple pathways mediate the sex-peptide-regulated switch in female *Drosophila* reproductive behaviours. Proc. Proc. R. Soc. B 280, 20131938.

Herndon, L.A., Wolfner, M.F., 1995. A *Drosophila* seminal fluid protein, Acp26aa, stimulates egg-laying in females for 1 Day after mating. Proc. Natl. Acad. Sci. U. S. A. 92, 10114–10118.

Hihara, F., 1981. Effects of the male accessory gland secretion on oviposition and remating in females of *Drosophila melanogaster*. Zool. Mag. 90, 307–316.

Jang, E.B., 1995. Effects of mating and accessory-gland injections on olfactory-mediated behavior in the female mediterranean fruit-fly, *ceratitis-capitata*. J. Insect Physiol. 41, 705–710.

Jang, E.B., McInnis, D.O., Lance, D.R., Carvalho, L.A., 1998. Maring-induced changes in olfactory-mediated behavior of laboratory-reared normal, sterile and wild female Mediterranean fruit flies (Diptera: Tephritidae) mated to cospecific males. Ann. Entomol. Soc. Am. 91, 139–144.

Kalb, J.M., Dibenedetto, A.J., Wolfner, M.F., 1993. Probing the function of *Drosophila Melanogaster* accessory-glands by directed cell ablation. Proc. Natl. Acad. Sci. U. S. A. 90, 8093–8097.

Kapelnikov, A., Rivlin, P.K., Hoy, R.R., Heifetz, Y., 2008. Tissue remodeling: a mating-induced differentiation program for the *Drosophila* oviduct. BMC Dev. Biol. 8, 114.

Katiyar, K.P., Ramirez, E., 1970. Mating frequency and fertility of Mediterranean fruit fly females alternately mated with normal and irradiated males. J. Econ. Entomol. 63, 1247–1250.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kawaguchi, A., Koch, G.G., 2015. Sanon: an R package for stratified analysis with nonparametric covariable adjustment. J. Stat. Softw. 67, 1–37.

Kennerdell, J.R., Carthew, R.W., 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell 95, 1017–1026.

Kraaijeveld, K., Chapman, T., 2004. Effects of male sterility on female remating in the mediterranean fruitfly, Ceratitis Capitata. Proc. R. Soc. B 271 (Suppl. 4), S209–S211.

Kraaijeveld, K., Katsoyannos, B.I., Stavrinides, M., Kouloussis, N.A., Chapman, T., 2005. Remating in wild females of the Mediterranean fruit fly, *Ceratitis capitata*. Anim. Behav. 69, 771–776.

Kubli, E., Bopp, D., 2012. Sexual behavior: how Sex Peptide flips the postmating switch of female flies. Curr. Biol. 22, R520–R522.

LaFlamme, B.A., Ram, K.R., Wolfner, M.F., 2012. The *Drosophila melanogaster* seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet. 8, e1002435.

Lawniczak, M.K.N., Begun, D.J., 2004. A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females. Genome 47, 900–910.

Liu, H., Kubli, E., 2003. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 100, 9929–9933.

Liu, Y., Liu, H., Liu, S., Wang, S., Jiang, R.J., Li, S., 2009. Hormonal and nutritional regulation of insect fat body development and function. Arch. Insect Biochem. Physiol. 71, 16–30.

Lung, O., Wolfner, M.F., 2001. Identification and characterization of the major *Drosophila melanogaster* mating plug protein. Insect biochem. Mol. Biol. 31, 543–551.

Mack, P.D., Kapelnikov, A., Heifetz, Y., Bender, M., 2006. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 103, 10358–10363.

Magnusson, K., Mendes, A.M., Windbichler, N., Papathanos, P.A., Nolan, T., Dottorini, T., Rizzi, E., Christophides, G.K., Crisanti, A., 2011. Transcription regulation of sex-biased genes during ontogeny in the malaria vector *Anopheles gambiae*. PloS One 6, e21572.

Malacrida, A.R., Gomulski, L.M., Bonizzoni, M., Bertin, S., Gasperi, G., Guglielmino, C.R., 2007. Globalization and fruitfly invasion and expansion: the medfly paradigm. Genetica 131, 1–9.

Manning, A., 1962. A sperm factor affecting the receptivity of *Drosophila melanogaster* females. Nature 194, 252–253.

Manning, A., 1967. The control of sexual receptivity in female *Drosophila*. Anim. Behav. 15, 239–250.

Marchini, D., Del Bene, G., Cappelli, L., Dallai, R., 2003. Ultrastructure of the male reproductive accessory glands in the medfly *Ceratitis capitata* (Diptera: Tephritidae) and preliminary characterization of their secretions. Arthropod Struct. Dev. 31, 313–327.

McGraw, L.A., Clark, A.G., Wolfner, M.F., 2008. Post-mating gene expression profiles of female *Drosophila melanogaster* in response to time and to four male accessory gland proteins. Genetics 179, 1395–1408.

McGraw, L.A., Gibson, G., Clark, A.G., Wolfner, M.F., 2004. Genes regulated by mating, sperm, or seminal proteins in mated female Drosophila melanogaster. Curr. Biol. 14, 1509–1514.

Miyatake, T., Chapman, T., Partridge, L., 1999. Mating-induced inhibition of remating in female Mediterranean fruit flies Ceratitis capitata. J. Insect Physiol. 45, 1021–1028.

Moshitzky, P., Fleischmann, I., Chaimov, N., Saudan, P., Klauser, S., Kubli, E., Applebaum, S.W., 1996. Sex-peptide activates juvenile hormone biosynthesis in the *Drosophila melanogaster* corpus allatum. Arch. Insect Biochem. Physiol. 32, 363–374.

Mossinson, S., Yuval, B., 2003. Regulation of sexual receptivity of female Mediterranean fruit flies: old hypotheses revisited and a new synthesis proposed. J. Insect Physiol. 49, 561–567.

Nakagawa, S., Farias, G.J., Suda, D., Cunningham, R.T., Chambers, D.L., 1971. Reproduction of the Mediterranean fruit fly: frequency of mating in the laboratory. Ann. Entomol. Soc. Am. 64, 949–950.

Neckameyer, W.S., 1998. Dopamine modulates female sexual receptivity in *Drosophila melanogaster*. J. Neurogenet. 12, 101–114.

Pane, A., Salvemini, M., Delli Bovi, P., Polito, C., Saccone, G., 2002. The transformer gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. Development 129, 3715–3725.

Papathanos, P.A., Windbichler, N., Menichelli, M., Burt, A., Crisanti, A., 2009. The *vasa* regulatory region mediates germline expression and maternal transmission of proteins in the malaria mosquito Anopheles gambiae: a versatile tool for genetic control strategies. BMC Mol. Biol. 10, 65.

Peng, J., Zipperlen, P., Kubli, E., 2005. *Drosophila* sex-peptide stimulates female innate immune system after mating via the Toll and Imd pathways. Curr. Biol. 15, 1690–1694.

R Core Team, 2015. R: a language and environment for statistical computing. In: R Foundation for Statistical Computing, Austria, Vienna.

Reiff, T., Jacobson, J., Cognigni, P., Antonello, Z., Ballesta, E., Tan, K.J., Yew, J.Y., Dominguez, M., Miguel-Aliaga, I., 2015. Endocrine Remodelling of the Adult Intestine Sustains Reproduction in Drosophila. Elife 4, e06930.

Robinson, A.S., Knols, B.G., Voigt, G., Hendrichs, J., 2009. Conceptual framework and rationale. Malar. J. 8 (Suppl. 2), S1.

Rogers, D.W., Baldini, F., Battaglia, F., Panico, M., Dell, A., Morris, H.R., Catteruccia, F., 2009. Transglutaminase-mediated semen coagulation controls sperm storage in the malaria mosquito. PLoS Biol. 7, e1000272.

Rogers, D.W., Whitten, M.M.A., Thailayil, J., Soichot, J., Levashina, E.A., Catteruccia, F., 2008. Molecular and cellular components of the mating machinery in *Anopheles gambiae* females. Proc. Natl. Acad. Sci. U. S. A. 105, 19390–19395.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Method Mol. Biol. 132, 365–386.

Rubin, G.M., Spradling, A.C., 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science 218, 348–353.

Saul, S.H., 1982. Rearing methods for the medfly, *Ceratitis capitata*. Ann. Entomol. Soc. Am. 75, 480–483.

Scolari, F., Gomulski, L.M., Ribeiro, J.M., Siciliano, P., Meraldi, A., Falchetto, M., Bonomi, A., Manni, M., Gabrieli, P., Malovini, A., Bellazzi, R., Aksoy, S., Gasperi, G., Malacrida, A.R., 2012. Transcriptional profiles of mating-responsive genes from testes and male accessory glands of the Mediterranean fruit fly, *Ceratitis capitata*. PLoS One 7, e46812.

Scolari, F., Yuval, B., Gomulski, L.M., Schetelig, M.F., Gabrieli, P., Bassetti, F., Wimmer, E.A., Malacrida, A.R., Gasperi, G., 2014. Polyandry in the medfly - shifts in paternity mediated by sperm stratification and mixing. BMC Genet. 15 (Suppl. 2), S10.

Seo, S.T., Vargas, R.I., Gilmore, J.E., Kurashima, R.S., Fujimoto, M.S., 1990. Sperm transfer in normal and gamma-irradiated laboratory-reared Mediterranean fruit flies (Diptera: Tephritidae). J. Econ. Entomol. 85, 1949–1953.

Siciliano, P., He, X.L., Woodcock, C., Pickett, J.A., Field, L.M., Birkett, M.A., Kalinova, B., Gomulski, L.M., Scolari, F., Gasperi, G., Malacrida, A.R., Zhou, J.J., 2014a.

Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, *Ceratitis capitata*. Insect biochem. Mol. Biol. 48, 51–62.

Siciliano, P., Scolari, F., Gomulski, L.M., Falchetto, M., Manni, M., Gabrieli, P., Field, L.M., Zhou, J.J., Gasperi, G., Malacrida, A.R., 2014b. Sniffing out chemosensory genes from the Mediterranean fruit fly, *Ceratitis capitata*. PLoS One 9, e85523.

Swanson, W.J., 2003. Sex peptide and the sperm effect in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. 100, 9643–9644.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Taylor, P.W., Kaspi, R., Yuval, B., 2000. Copula duration and sperm storage in Mediterranean fruit flies from a wild population. Physiol. Entomol. 25, 94–99.

Tazuke, S.I., Schulz, C., Gilboa, L., Fogarty, M., Mahowald, A.P., Guichet, A., Ephrussi, A., Wood, C.G., Lehmann, R., Fuller, M.T., 2002. A germline-specific gap junction protein required for survival of differentiating early germ cells. Development 129, 2529–2539.

Thailayil, J., Magnusson, K., Godfray, H.C., Crisanti, A., Catteruccia, F., 2011. Spermless males elicit large-scale female responses to mating in the malaria mosquito *Anopheles gambiae*. Proc. Natl. Acad. Sci. U. S. A. 108, 13677–13681.

Twig, E., Yuval, B., 2005. Function of multiple sperm storage organs in female Mediterranean fruit flies (Ceratitis capitata, Diptera : Tephritidae). J. Insect Physiol. 51, 67–74.

Valdez, J.M., 2001. Ultrastructure of the testis of the Mexican fruit fly (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 94, 251–256.

Vera, M.T., Cladera, J.L., Calcagno, G., Vilardi, J.C., McInnis, D.O., Stolar, E., Group, F.W., 2003. Remating of wild *Ceratitis capitata* (Diptera: Tephritidae) females in field cages. Ann. Entomol. Soc. Am. 96, 563–570.

Wigby, S., Sirot, L.K., Linklater, J.R., Buehner, N., Calboli, F.C., Bretman, A., Wolfner, M.F., Chapman, T., 2009. Seminal fluid protein allocation and male reproductive success. Curr. Biol. 19, 751–757.

Williamson, D.L., 1989. Oogenesis and spermatogenesis. In: Robertson, A.S., Hooper, G. (Eds.), Fruit Flies. Their Biology, Natural Enemies and Control. Elsevier, Amsterdam, The Netherlands, pp. 141–151.

Wolfner, M.F., 2007. S.P.E.R.M."(seminal proteins (are) essential reproductive modulators): the view from Drosophila. Soc. Reprod. Fertil. (Suppl. 65), 183–199.

Xue, L., Noll, M., 2000. Drosophila female sexual behavior induced by sterile males showing copulation complementation. Proc. Natl. Acad. Sci. U. S. A. 97, 3272–3275.

Yahouedo, G.A., Djogbenou, L., Saizonou, J., Assogba, B.S., Makoutode, M., Gilles, J.R.L., Maiga, H., Mouline, K., Soukou, B.K., Simard, F., 2014. Effect of three larval diets on larval development and male sexual performance of *Anopheles gambiae* s.s. Acta Trop. 132, S96–S101.

Yapici, N., Kim, Y.J., Ribeiro, C., Dickson, B.J., 2008. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. Nature 451, 33–U31.

Yuval, B., Blay, S., Kaspi, R., 1996. Sperm transfer and storage in the Mediterranean fruit fly (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 89, 486–492.

Yuval, B., Hendrichs, J., 2000. Behavior of flies in the genus *ceratitis* (Dacinae: ceratitidini). In: Aluja, M., Norrbom, A.L. (Eds.), Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior. CRC Press, Boca Raton, Florida, USA, pp. 429–458.

Papanicolaou *et al. Genome Biology* (2016) 17:192
DOI 10.1186/s13059-016-1049-2

Genome Biology

RESEARCH Open Access

The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species

Alexie Papanicolaou[1], Marc F. Schetelig[2], Peter Arensburger[3], Peter W. Atkinson[4,5], Joshua B. Benoit[6], Kostas Bourtzis[7,8], Pedro Castañera[9], John P. Cavanaugh[6], Hsu Chao[10], Christopher Childers[11], Ingrid Curril[12], Huyen Dinh[10], HarshaVardhan Doddapaneni[10], Amanda Dolan[13], Shannon Dugan[10], Markus Friedrich[14], Giuliano Gasperi[15], Scott Geib[16], Georgios Georgakilas[17], Richard A. Gibbs[10], Sarah D. Giers[18], Ludvik M. Gomulski[15], Miguel González-Guzmán[9], Ana Guillem-Amat[9], Yi Han[10], Artemis G. Hatzigeorgiou[17], Pedro Hernández-Crespo[9], Daniel S. T. Hughes[10], Jeffery W. Jones[19], Dimitra Karagkouni[17], Panagiota Koskinioti[20], Sandra L. Lee[10], Anna R. Malacrida[15], Mosè Manni[15], Kostas Mathiopoulos[20], Angela Meccariello[21], Shwetha C. Murali[10], Terence D. Murphy[22], Donna M. Muzny[10], Georg Oberhofer[17], Félix Ortego[9], Maria D. Paraskevopoulou[17], Monica Poelchau[11], Jiaxin Qu[10], Martin Reczko[23], Hugh M. Robertson[18], Andrew J. Rosendale[6], Andrew E. Rosselot[6], Giuseppe Saccone[21], Marco Salvemini[21], Grazia Savini[15], Patrick Schreiner[5], Francesca Scolari[15], Paolo Siciliano[15], Sheina B. Sim[16], George Tsiamis[8], Enric Ureña[9], Ioannis S. Vlachos[17], John H. Werren[13], Ernst A. Wimmer[12], Kim C. Worley[10], Antigone Zacharopoulou[24], Stephen Richards[10] and Alfred M. Handler[25*]

Abstract

Background: The Mediterranean fruit fly (medfly), *Ceratitis capitata*, is a major destructive insect pest due to its broad host range, which includes hundreds of fruits and vegetables. It exhibits a unique ability to invade and adapt to ecological niches throughout tropical and subtropical regions of the world, though medfly infestations have been prevented and controlled by the sterile insect technique (SIT) as part of integrated pest management programs (IPMs). The genetic analysis and manipulation of medfly has been subject to intensive study in an effort to improve SIT efficacy and other aspects of IPM control.

Results: The 479 Mb medfly genome is sequenced from adult flies from lines inbred for 20 generations. A high-quality assembly is achieved having a contig N50 of 45.7 kb and scaffold N50 of 4.06 Mb. In-depth curation of more than 1800 messenger RNAs shows specific gene expansions that can be related to invasiveness and host adaptation, including gene families for chemoreception, toxin and insecticide metabolism, cuticle proteins, opsins, and aquaporins. We identify genes relevant to IPM control, including those required to improve SIT.
(Continued on next page)

* Correspondence: al.handler@ars.usda.gov
[25]USDA-ARS, Center for Medical, Agricultural, and Veterinary Entomology, 1700 S.W. 23rd Drive, Gainesville, FL 32608, USA
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: The medfly genome sequence provides critical insights into the biology of one of the most serious and widespread agricultural pests. This knowledge should significantly advance the means of controlling the size and invasive potential of medfly populations. Its close relationship to *Drosophila*, and other insect species important to agriculture and human health, will further comparative functional and structural studies of insect genomes that should broaden our understanding of gene family evolution.

Keywords: Medfly genome, Tephritid genomics, Insect orthology, Gene family evolution, Chromosomal synteny, Insect invasiveness, Insect adaptation, Medfly integrated pest management (IPM)

Background

The Mediterranean fruit fly (medfly, *Ceratitis capitata*, Diptera: Tephritidae) is one of the most destructive agricultural pests throughout the world due to its broad host plant range that includes more than 260 different fruits, vegetables, and nuts [1]. Host preferences vary in different regions of the world, which can be associated with its ability to invade and adapt to ecological niches throughout tropical and subtropical regions. While the species originated in sub-Saharan Africa [2, 3], it is currently endemic throughout Africa, the Middle East, European countries adjacent and proximal to the Mediterranean Sea, the Hawaiian Islands, the Caribbean, and Central and South America [4]. Thus the worldwide economic costs due to crop damage, export control due to quarantine restrictions, and control and prevention of medfly infestation reach many US$ billions each year [5] (for an overview of medfly biology, ecology, and invasiveness, see: http://www.cabi.org/isc/datasheet/12367).

Medfly has also been an established lab organism for several decades and is notable as being the closest non-drosophilid relative to *Drosophila* subject to intensive genetic analysis, with broad chromosomal syntenic relationships established. These studies have been largely driven by efforts to use genetic manipulation to improve the sterile insect technique (SIT), which is the primary biologically based method used to control medfly as a component of area-wide multi-tactical integrated pest management (IPM) approaches, which include the use of natural enemies and insecticide/bait formulations. Current SIT applications are based on the use of a classical genetic sexing strain that incorporates female-specific activity of an embryonic temperature-sensitive lethal (*tsl*) mutation. Resultant males are mass-reared in billions per week for sterilization and release in North, Central and South America, Australia, South Africa, and Mediterranean countries including Spain and Israel, to not only control existing populations but to also prevent new invasions [6]. As such, medfly has served as a model system for developing genetic analyses and manipulations that might improve these population control programs that are applicable to a large number of tephritid fruit fly species throughout the world, which range from similarly polyphagous species to ones that are more highly specialized.

Previous studies in medfly mapped ~30 cloned genes and ~40 microsatellite sequences by in situ hybridization to larval salivary gland polytene chromosomes [7, 8]. It was also the first non-drosophilid insect to have its germ-line efficiently transformed by a transposon-based vector system [9], an approach that has since been applied to several orders of non-drosophilid species. This has included functional genomics analysis, new vector systems for transgene stabilization, genomic targeting, and transgenic and *Wolbachia*-infected strains created for potential population control.

To further our understanding of this critical agricultural pest and its genomic organization in comparison to *Drosophila* and other dipteran/insect species, we now present the results of the medfly whole genome sequencing (WGS) project. This is one of 30 arthropod genome sequencing projects that have been initiated as a part of a pilot project for the i5K arthropod project [10] at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC). Notably, the quality of this analysis is unusually strong for an insect genome, comparable to the more compact genome of *Drosophila melanogaster*, where half the 479 Mb medfly genome sequence was assembled in 35 scaffolds larger than 4 Mb (NG50). A thorough automated structural annotation of the genome was conducted, aided by RNA sequencing (RNA-Seq) data, which allowed a curation community of 20 groups to make key sequence assignments related to genome structure, orthology, and genetic regulation, and to manually annotate key gene families related to invasiveness and adaptation, insecticide resistance and detoxification, and aspects of sex-determination, reproduction, and cell death.

This extensive resource is expected to provide a foundation for continued research on fundamental and comparative studies of insect genomes and gene family evolution and the high-quality reference genome assembly should have far-reaching practical applications in pest management research. It will be instrumental to the

development of methods for the identification of genome-wide polymorphisms that can be used for population genetic analysis and source determination of medflies identified in ports of entry. Furthermore, its extensive annotated gene set will facilitate identifying the molecular basis of mutations in strains used for SIT (e.g. *tsl* sexing strain) and the identification of novel targets that can be utilized to facilitate higher efficiency and efficacy of IPM programs.

Results and discussion

Genome sequence, structure, orthology, and function

Whole genome sequencing and assembly

The medfly WGS project reported here is a continuation of an initial project initiated at HGSC that is summarized in Additional file 1: Supplementary material A. Briefly, the initial 454 sequencing project used mixed-sex embryonic DNA from a long-term caged population of the ISPRA strain maintained at the University of Pavia, Italy. This approach yielded relatively low N50 values for both contigs (~3.1 kb) and scaffolds (~29.4 kb) that are presumed to be the result of high levels of polymorphism and repetitive DNA. Thus, the subsequent sequencing attempt reported here used DNA from 1–3 adults that arose from ISPRA lines inbred in single pairs for 12–20 generations. This DNA was used to create 180 bp to 6.4 kb insert-size libraries for Illumina HiSeq2000 sequencing followed by an ALLPATHS-LG assembly (Additional file 2: Table S1; see "Methods"). This yielded a highly improved assembly (GB assembly acc: GCA_000347755.1), though it was determined that 5.7 Mb comprised endosymbiotic bacterial sequences (Enterobacteriaceae and Comamonadaceae; see Additional file 1: Supplementary material C) localized to 18 scaffolds. The majority of the contaminant sequences represent the genome of *Pluralibacter gergoviae* that was recovered in two contigs (see Additional file 1: Supplementary material D and Additional file 2: Tables S2 and S3 for the *P. gergoviae* genome details and annotation). After removal of these bacterial sequences, the new assembly (GB assembly acc: GCA_000347755.2) revealed a final genome size of 479.1 Mb, corresponding to the initial estimated size of 484 Mb that included the bacterial sequences. The 479 Mb assembly size is slightly less than earlier estimates of 540 Mb and 591 Mb, derived from Feulgen stain [11] and qPCR [12] studies, respectively, due to the difficulty of assembling highly repetitive heterochromatic sequences. Re-estimation of the genome size by k-mer analysis, using Jellyfish [13], of the 500 bp insert library sequences obtained a value of 538.9 Mb, in agreement with the Feulgen stain study. Using this estimate, we presume the remaining 11 % of the genome is repetitive heterochromatic regions that could not be assembled with our short read procedure.

The revised assembly yielded 25,233 contigs with an N50 of 45,879 bp assembled into 1806 scaffolds with an N50 of 4.1 Mb (Table 1; see Table 2 for additional assembly features). Using BUSCO [14] on the final genome assembly, it was determined that the assembly correctly identified the full sequence of 2556 genes from a total of 2675 (95 %) found to be conserved across most arthropods. Furthermore, partial coverage of 91 (3.4 %) genes was identified, with only 28 (1.0 %) missing, and an additional 153 (5.7 %) being duplicated. For comparison, the same analysis run on the *D. melanogaster* genome sequence (v. 5.53) identified 98 % of the genes as complete, 0.7 % partial, 0.3 % missing, and 6.5 % duplicated (see Additional file 2: Table S4 for comparisons to *Drosophila* and tephritid species).

Curation and gene ontology

Automated annotations were performed using three approaches (see "Methods" and Additional file 1: Supplementary material B): (1) Maker 2.0 [15] at HGSC with the assembled genome and adult male and female RNA-Seq data used to improve gene models; (2) at NCBI using the Gnomon pipeline; and (3) our in-house Just_Annotate_My_genome (JAMg) annotation platform that makes use of RNA-Seq data and de novo predictions (http://jamg.sourceforge.net). Preliminary analysis showed that the NCBI and JAMg annotations broadly agreed and had fewer false positives than Maker 2.0. For manual annotations, curators were provided with the WebApollo manual curation tool [16, 17] hosted by the U.S. Department of Agriculture, National Agricultural Library (USDA-NAL), and data from the JAMg annotation pipeline (and associated tools) with NCBI Ref-Seq derived models. The annotation of 20 key gene sets has resulted, thus far, in curation of 1823 gene (messenger RNA [mRNA]) models, making medfly one of the most highly curated non-drosophilid insects. This has allowed in depth genomic analyses that have revealed divergent genes exhibiting rapid evolutionary rates. These data have been integrated into the dipteran phylogenetic framework by undertaking orthology and synteny comparisons, especially with the closely related species, *D. melanogaster*, and the housefly, *Musca domestica*.

Orthology to other arthropod genomes

To assess the conservation of protein-coding genes between *C. capitata* and other arthropods, complete proteomes from 14 additional arthropod species were used with *C. capitata* to determine orthology. The analysis of 254,384 protein sequences from 15 species identified 26,212 orthologous groups (defined as containing at least two peptide sequences), placing 202,278 genes into orthologous groups while failing to allocate 52,106 (unique) protein-coding genes into any group. The

Table 1 Medfly genome assembly metrics for NCBI Genome assembly accession GCA_000347755.2 that replaces assembly GCA_000347755.1 after removal of bacterial contaminant sequences

Genome assembly	Contigs (n)	25,233
	Contig N50	45,879 bp
	Scaffolds (n)	1806
	Scaffold N50	4,118,346 bp
	Size of final assembly	479,047,742
	Size of final assembly - without gaps	440,703,716 bp
	NCBI Genome Assembly Accession	GCA_000347755.2 http://www.ncbi.nlm.nih.gov/assembly/GCA_000347755.2

majority of the unique proteins were identified in *Acyrthosiphon pisum* and *Daphnia pulex*, while interestingly, *C. capitata* had the largest proportion (87 %) of proteins placed into an orthologous group (Fig. 1). This could have been influenced by the larger sampling of dipteran genomes relative to other taxa.

The distribution of proteins among orthologous groups is presented in Additional file 2: Table S5a. When examining conserved proteins, 1345 orthologous groups were found to contain a single-copy protein from all species, while an additional 1879 orthologous groups were found to contain multiple members in one or more species. Moreover, there are 5767 orthologous groups unique to Diptera, 224 of which are present in all dipterans. Within *C. capitata*, 1608 putative peptide sequences could not be placed into any orthologous group, thus identifying them as more recently evolved orphan genes (see Additional file 2: Tables S5b and S5c for orthologous groups and the number of groups for analyzed species, respectively). The distribution of these orphans across the genome is relatively uniform, with no clear pattern or clusters of genes. While these orphan genes are unique to *C. capitata* within this analysis, we would expect this to be less likely if more dipteran species were included within the analysis. Current work is ongoing to provide a more robust orthology of proteins within the family Tephritidae compared to related taxa in Diptera.

Table 2 Medfly genome NCBI annotation features for the assembly Ccap_1.0 (see http://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ceratitis_capitata/101/ for details and additional features)

Feature	Count	Mean length (bp)
Genes and pseudogenes	14,652	–
Protein-coding	14,162	–
Non-coding	385	–
Pseudogenes	105	–
Genes with variants	3527	–
Genes	14,547	16,014
All transcripts	24,125	2903
Messenger RNA	23,075	2979
Miscellaneous RNA	238	3506
Transfer RNA	416	74
Long non-coding RNA	396	1074
Single-exon transcripts	2833	1193
CDSs	23,075	2198
Exons	77,742	465
Introns	62,132	4117

Chromosomal assignment of scaffolds

A physical map of the genome that assigns scaffolds to chromosomal loci helps to refine and verify the genomic assembly and allows the analysis of syntenic relationships between species for evolutionary comparisons [18, 19]. It should also aid in the design and analysis of genetic manipulations (e.g. genomic targeting, creation of chromosome inversions and translocations) and especially gene-editing approaches. Similar to *Drosophila*, *C. capitata* is among the few species subject to genomic analysis for which a larval salivary gland polytene chromosome map is available that has been subjected to cytogenetic analysis by in situ hybridization of cloned genes and microsatellite sequences [7, 8]. This has allowed the initiation of a physical genome map by assigning 43 scaffolds, linked to these genes and sequences, to defined loci on five autosomal chromosomes (chromosomes 2 to 6) and a single scaffold to the X (Fig. 2, Additional file 2: Table S6). Four repetitive DNA clones proven to be Y-linked by in situ hybridization to mitotic chromosomes [20] were associated with three scaffolds, though neither the X nor Y have defined polytene mapped loci. Thus, *ceratatoxin* (*ctx*) genes, linked to a single 6.4 Mb scaffold, were also mapped to chromosome 1 (X chromosome) by mitotic chromosome hybridization [21]. Scaffolds with a combined length of 42.6 Mb were linked to chromosome 2, 7.2 Mb to chromosome 3, 60.9 Mb to chromosome 4, 49.1 Mb to chromosome 5, and 45.8 Mb to chromosome 6. The Y-linked sequences could be assigned to more than ten scaffolds, though all were relatively short with the four mapped sequences comprising 0.81 Mb. Thus, more than 212 Mb has been mapped, representing nearly 45 % of the genome, allowing a large proportion of the curated genes to be localized to chromosomal map positions.

This initial step in the development of a physical map already provides significant support for the assembly since localization of 14 (of 45) mapped scaffolds were supported by two or more mapped genes/sequences, and in no case was discontinuity by intervening scaffolds observed. This includes the 15.8 Mb scaffold 3 (NW_004523802.1) on 6L,

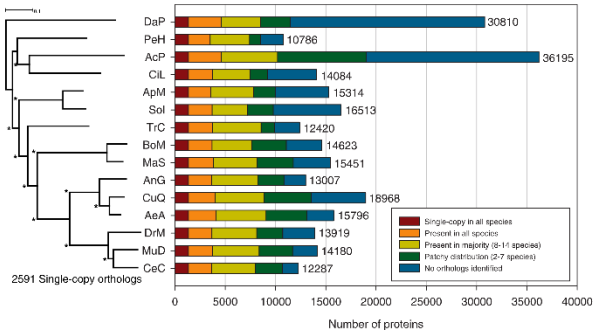


Fig. 1 Genome-wide phylogenomics and orthology. The phylogenetic relationship of *C. capitata* and 13 species in Arthropoda was estimated using a maximum likelihood analysis of a concatenation of 2591 single-copy orthologous protein sequences, 1000 bootstrap replicates, and rooted with *D. pulex*. The *scale bar* represents 0.1 amino acid substitution per site and the *asterisks* represent nodes with a bootstrap value of 100. *Horizontal bars* for each species show the absolute number of proteins that are: single-copy orthologs in all species, present in all species (not necessarily in single-copy), present in the majority of species in the analysis, present in a minority of the species (patchy distribution) in the analysis, and unique to the species. Species/strain designations are: *Acyrthosiphon pisum* (AcP), *Aedes aegypti* (strain Liverpool) (AeA), *Anopheles gambiae* (strain PEST) (AnG), *Apis mellifera* (ApM), *Bombyx mori* (BoM), *Ceratitis capitata* (CeC), *Cimex lectularius* (CiL), *Culex quinquefasciatus* (strain Johannesburg) (CuQ), *Daphnia pulex* (DaP), *Drosophila melanogaster* (DrM), *Manduca sexta* (MaS), *Musca domestica* (MuD), *Pediculus humanus* (PeH), *Solenopsis invicta* (Sol), and *Tribolium castaneum* (TrC)

to which *Medflymic80* at 85B and the *chorion S15-S19* cluster/*paramyosin* genes at 88B are linked. The extrapolated length spanning three map sections (~4.8 Mb/ map section) is also consistent with approximately 90 % of the scaffold size and the high quality of its linkage. The assignment of the 11 Mb scaffold 7 (NW_004524245.1) to 5L-62 to 65C by linkage to three in situ hybridized sequences (*Medflymic41*, *integrin-αPS2*, and *white*) is also consistent with scaffold length and integrity.

Continued scaffold assignment to linkage groups by genome-wide SNP analysis and continued chromosomal hybridizations of annotated genes should aid in the further expansion of the physical map and assembly confirmation. This may also be facilitated by insertion site sequencing of transformant vector integrations that have been localized by in situ hybridization, especially a series of *piggyBac* vector insertions [22] mapped to chromosome 5 of the D53 inversion strain used (to suppress recombination) in the *temperature sensitive lethal* genetic-sexing strain (Fig. 3). Insertion site sequencing for these transformant lines, among others using several transposon-based vectors, should allow further scaffold assignments to these loci.

Transposable elements

Mobile or transposable elements (TE) are major constituents of many metazoan genomes that play a significant role in the generation of spontaneous mutations, either by insertional mutagenesis or imprecise excisions that delete critical coding or regulatory sequences [23], making their presence and activity a major contributor to the evolutionary potential of most organisms. The ability of many functional autonomous elements to undergo lateral inter-species transfer, expansion in the new host genome and then accumulation of mutations rendering them inactive, also provides several means of defining phylogenetic relationships.

The presence and activity of TEs (especially Class II elements), in species such as *C. capitata*, has particular significance since medfly was the first non-drosophilid insect to be genetically transformed using a transposon-based vector system [9] and has been the subject of genetic manipulations for improved functional analysis and sexing and sterility strains for improved SIT [24–29]. In this regard, expansion of transformation-based manipulations may depend on the further identification of new TE vector systems, as well as ensuring the stability of previous and new transgenic strains by identifying the genomic presence of potential cross-mobilizing systems [28, 30, 31].

Approximately 18 % of the assembled *C. capitata* genome contains sequences of TE origin (see Additional file 2: Table S7). Of these sequences, 55.9 % are composed of retrotransposons (15.7 % long terminal repeats [LTRs] and 40.2 % non-LTRs) and 44.1 % are DNA transposons. This is higher than the proportion estimated for the

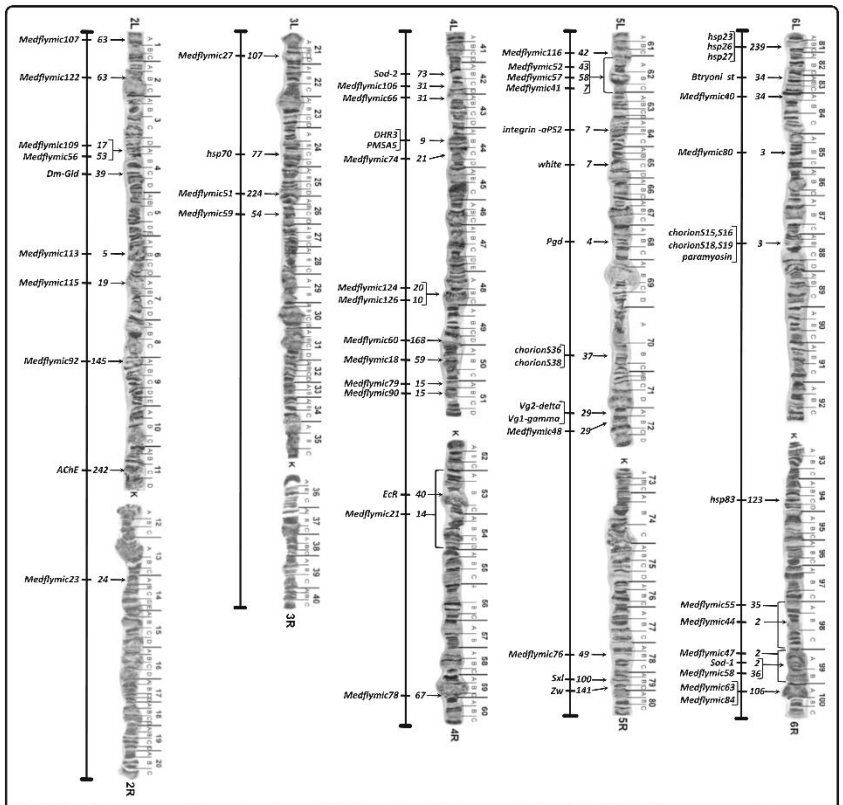


Fig. 2 *C. capitata* genome scaffold map based on scaffold linkage of annotated genes and microsatellite (*Medflymic*) sequences previously localized to map banding positions by in situ hybridization to autosomal polytene chromosomes (chromosomes 2 to 6). The larval salivary gland polytene chromosome map [193] presented includes left (*l*) and right (*R*) autosomal chromosome arms linked at a centromeric region (*K*). *Arrows* with adjacent scaffold numbers point to mapped loci positions of designated genes/microsatellites, with *bracketed positions* used for less precise mapping. See Additional file 2: Table S6 for sequence and scaffold accession numbers and sizes, in addition to map positions for sex-linked (chromosome 1; X and Y) genes/sequences mapped to undefined loci on mitotic non-polytenized chromosome spreads [20, 21]

euchromatic region of *D. melanogaster* (5.3 %), but not much above the combined estimate as a proportion of the total euchromatic and heterochromatic *D. melanogaster* genome (10–15 %) [32]. Among Class I elements, the *RTE* and *Jockey* non-LTR TE subclasses occupy 5.27 % and 1.47 % of the genome, respectively, while the *Gypsy* LTR TE subclass occupies 1.4 % of the genome. Among DNA transposons, all the major superfamilies are represented with sequences derived from the *Tc1/mariner* superfamily comprising a large majority (82.1 %) of the identified genomic sequences (approximately 6.8 % of the entire genome). This is consistent with the observation in *Drosophila* species in which *Tc1/mariner* elements are a large percentage of DNA transposons, and also with studies from other tephritid species [33, 34].

The large *Tc1/mariner* element representation is of interest since the *Tc1*-like *Minos* element from *D. hydei* [35] was used for the first germ-line transformation of

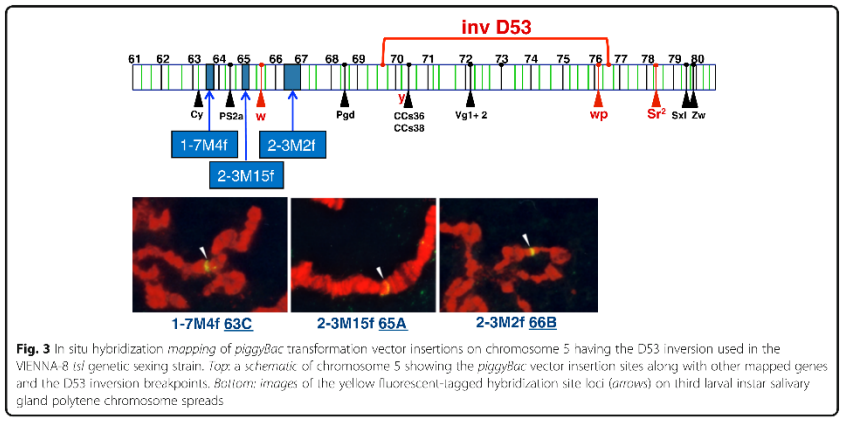


Fig. 3 In situ hybridization *mapping* of *piggyBac* transformation vector insertions on chromosome 5 having the D53 inversion used in the VIENNA-8 *tsf* genetic sexing strain. *Top*: a schematic of chromosome 5 showing the *piggyBac* vector insertion sites along with other mapped genes and the D53 inversion breakpoints. *Bottom*: images of the yellow fluorescent-tagged hybridization site loci (*arrows*) on third larval instar salivary gland polytene chromosome spreads

C. capitata [9] and the *mariner Mos1* element has a similar potential [11, 36]. BLASTp with the *Minos* transposase amino acid sequence provided no significant alignments, though tBLASTn provided several high similarity alignments including a 1040 nt translated sequence on scaffold 145 (NW_004523255.1) having 55 % identity ($1e^{-102}$).

Medfly has been most commonly transformed with *piggyBac*-based vectors [22] and many of these lines have been stably maintained under artificial marker selection for ten years or more. It is thus unlikely that *piggyBac*-related elements have cross-mobilizing activity, which is supported by relatively low similarities to *piggyBac* at the nucleotide or amino acid sequence level. This is in contrast to nearly identical *piggyBac* elements (>95 % nucleotide coding-region identity) that exist in a broad range of *Bactrocera* species, and throughout the *Bactrocera dorsalis* species complex, though none have been shown to be functional [37, 38]. Indeed, amino acid sequences in the medfly database that are most highly similar to *piggyBac* transposase by BLASTp include the human Pgdb4 isoform X2 ($1e^{-27}$; XP_012161257.1) and *piggyBac*-derived protein 4-like ($1e^{-25}$; XP_004524835.1) sequences, all of which belong to the transposase IS4 family having the DDE_Tnp_1_7 domain. Alignments to the *piggyBac* transposase by tBLASTn, however, yielded multiple sequences with e-values <$1e^{-10}$, with a translated 1281 nucleotide sequence on scaffold 297 (NW_004523799.1) yielding the most significant e-value of $7e^{-148}$.

The *hAT*-family *Hermes* element from *M. domestica* [39] has also been used to transform medfly [40], and similar to *Minos*, significant similarities in the medfly genome have yet to be found to the complete nucleotide sequence by BLASTn, nor the transposase amino acid sequence by BLASTp. However, tBLASTn provided alignments to two translated nucleotide sequences on scaffold 50 (NW_004524024.1) having significant identities to *Hermes* transposase (43 % identity, $4e^{-128}$; and 39 % identity, $1e^{-116}$). Notably, overlapping sequences on scaffold 50 also show relatively high similarity to the *D. melanogaster hobo* (HFL1) *hAT* element that, thus far, has not shown vector function beyond drosophilid species. Partial sequences of several *hobo*-like elements were previously identified in medfly [41, 42].

As expected for members of Class II transposon superfamilies, related elements (or sequences) exist in medfly, though the relatively low levels of identity to transposon vector systems functional in this species, especially in the transposase-coding regions, suggest that significant divergence has occurred during vertical inheritance. If functional, the potential for cross-mobilization would require empirical evidence given the relative stability of transformant lines created with the three vector systems thus far. However, it may be assumed that transposon vectors currently in use will remain stable within the medfly genome with respect to potential re-mobilization by a cross-mobilizing transposase. This is a critical consideration for evaluating strain stability and environmental risks for transgenic strains used in field release programs.

microRNAs and microRNA machinery

microRNAs (miRNAs) are small (~22 nts) non-coding RNAs that play a critical role in gene regulation by inducing mRNA degradation and/or translation suppression

for genes involved in reproduction, metamorphosis, aging, and social behavior, among numerous other aspects of insect biology [43]. This RNA interference (RNAi) mechanism for regulation of gene expression has facilitated the development of novel strategies for population control. Notably, the role of *Cctra* and *Cctra-2* in medfly sex determination has been clarified by RNAi studies where chromosomal XX females have undergone a sex reversal to phenotypic males [44, 45]. This result has also provided a unique approach to creating male-only populations for SIT release.

To identify precursor and mature miRNAs in *Ceratitis*, a step-wise annotation methodology was utilized. In brief, Hexapoda miRNA precursors and mature sequences were aligned against the assembled genome, while novel miRNA precursors were detected by a modified version of the Maple algorithm [46]. In total, 158 mature miRNAs were identified with high confidence along with 83 precursors. A total of 129 presented significant homology with known Hexapoda miRNAs (Additional file 2: Table S8). The number of identified miRNAs is considerably less than those annotated in *D. melanogaster* (n = 430), but comparable to *Aedes aegypti* (n = 125), and twice the number found in *Musca domestica* (n = 69). A set of 13 putative clusters comprising more than one pre-miRNA have been identified. The majority of clusters (10 out of 13) range in size between 260 and 8514 bp, while the remaining three have sizes of 14,436 bp, 18,467 bp, and 25,941 bp, respectively. Thirty-three of the 83 precursors (~39.7 %) belong to a cluster. The genomic localization of the identified pre-miRNAs (based on the NCBI Gnomon annotation) are putatively detected as six read-through pre-miRNAs (7.2 %; same strand orientation within 4 kb of a start/stop coding region), 30 intronic (36.2 %), three antisense (3.6 %), and 44 intergenic (53 %).

In addition to the mature and precursor miRNA sequences, genes associated with RNAi have been characterized, including *dicer1*, *drosha*, *dgcr8* (*pasha*), and *argonaute-2*, which play key roles in miRNA biogenesis and function [47, 48] (Additional file 2: Table S9). Other important identified miRNA-related genes include *snd1*, *gawky*, *dcp2*, and *ccr4-not*. *Snd1* is a member of the RISC complex [49, 50], while *gawky* is required for miRNA-mediated gene silencing, promoting mRNA deadenylation and decapping via *ccr4-cnot* and *dcp2* recruitment [51–53]. Moreover, genes encoding *exportin-5* and *ran* proteins, which are responsible for miRNA precursor nuclear export [54] were also characterized.

The identified miRNA-related proteins and the number of predicted miRNAs, exhibit a well-formed layer of post-transcriptional regulation through miRNA-induced translation suppression or mRNA degradation. The high number of conserved miRNAs indicates regulation networks conserved in Hexapoda. No *Sid-1* homologue, which is involved in systemic miRNA, was found in the *C. capitata* genome, similar to other dipteran species. Genes related to siRNA and piRNA biogenesis and function were also identified, including *prkra*, which is required for miRNA/siRNA production by *dicer1*, and *piwi*, *argonaute-3*, *aubergine*, *hen1*, *spindle-e*, *tejas*, *vasa*, *maelstrom*, *deadlock*, *cutoff*, and *tdrkh*, which are part of the piRNA pathway [55]. Interestingly, only *deadlock* and *cutoff* were identified, while a *rhino* homologue was not detected in the medfly genome. These three proteins form the proposed *rhino*, *deadlock*, *cutoff* (RDC) complex and loss of *rhino* leads to loss of piRNAs from dual-strand, but not from uni-strand clusters in *Drosophila* [56]. Furthermore, *Yb*-like genes were not identified and it is of interest to determine how the lack of highly homologous *Yb* and *rhino* genes affect piRNA germline transcription or if there are proteins presenting similar functions. Nevertheless, the numerous piRNA-related proteins signify a well-structured pathway.

Gene families associated with adaptation and invasiveness

Unlike *Drosophila* that inhabits and feeds on rotting and decaying organic matter, and *Musca* that feeds and develops in excreta, carcasses, and other septic matter, medfly is an opportunistic phytophagous species whose survival and dispersion is tightly dependent on its interactions with its different host plants [57]. Adult medfly reproductive behavior is very elaborate and involves the use of sexual pheromones [58, 59] and adults must seek out a rich diet based on carbohydrates and proteins to support a high reproductive rate [60], which involves chemoreception and vision to detect appropriate plant hosts and adaptation to aqueous larval environments. In addition, all invasive insects require adaptation to new and diverse microbial environments requiring immunity mechanisms that defend against multiple pathogens. These are especially important to medfly which oviposits its eggs in microbiologically rich environments. These different resource exploitation and survival strategies are reflected in adaptive differences in the chemoreception, water transport, and visual and immunity system pathways of these species. Indeed, the impressive biological success of medfly is supported by these and additional adaptive traits involving larval to adult life stages in which plasticity for these pathways play a fundamental role [61].

Chemoreception

Insect olfaction and gustation is the product of a signal transduction cascade that includes four major gene families [62]. These include the water-soluble odorant-binding proteins (OBPs) that bind to the membrane-

bound odorant receptors (ORs) [62, 63], the gustatory taste receptors (GRs) [64], and the ionotropic receptors (IRs) that evolved from the ionotropic glutamate receptor superfamily that respond to amines, acids, and other odorants not perceived by the ORs [65, 66].

The *C. capitata* OBP and chemoreceptor family repertoires (ORs, GRs, and IRs) were compared with those of *D. melanogaster* [63, 65, 67] and the housefly, *M. domestica* [68] (Fig. 4). The total of 46 OBP genes encoding 48 proteins augments the 17 OBP transcripts identified previously from a transcriptome study [58], which is less than the other flies; however, medfly has chemoreceptor repertoires of intermediate size between the lower and higher counts of *D. melanogaster* and *M. domestica*, respectively (see Additional file 3: Figure S1 and Additional file 2: Table S10). The 17 previously identified OBPs were characterized for their transcription activities in different body compartments of each sex and a subset showed transcriptional changes related to maturation, mating, and time of day [59, 69]. Further biochemical assays indicated that CcapObp24 (previously named CcapOBP83a-2) showed high affinity for (E,E)-a-farnesene, one of the five major components of the medfly male pheromone [59]. Detailed examination of the gene family relationships in *C. capitata*, *D. melanogaster*, and *M. domestica* revealed expected patterns of birth-and-death gene family evolution typical of environmentally relevant genes.

Medfly shares the highly conserved members of these families, such as the Orco protein that functions with each specific OR to make a functional olfactory receptor (Fig. 5; see Additional file 2: Table S11) and the equivalent Ir8a/25a proteins, along with the conserved suites of sugar and carbon dioxide GRs and some bitter taste GRs. In contrast to *Musca*, which exhibited expansions of several lineages of candidate bitter taste receptors in the GR and IR families, as well as expansion of the DmelOr45a lineage implicated in repulsion from aversive chemicals in larvae [70], *Ceratitis* is more similar to *Drosophila* in having few differential expansions of candidate bitter GR and IRs and a single ortholog of DmelOr45a (Fig. 6; see Additional file 2: Tables S12 and S13 and Additional file 3: Figure S2). It also has an expansion of the DmelGr32a and 39b lineages (single orthologs in *Musca*), as well as an ortholog for DmelGr68a (lost from *Musca*), all of which are implicated in pheromone perception [71]. *Ceratitis* also has an expansion (CcapOr58-62) equidistant from the DmelOr67d and DmelOr83c lineages that have, apparently, very different functions. DmelOr67d is involved in the perception of a male-produced pheromone [72] whereas DmelOr83c is specific for farnesol, a component of citrus peel [73], as well as a major component of the medfly pheromone [59]. An extensive expansion of the DmelOr67d lineage is present in *Musca* (MdomOr53-65) suggesting that the housefly, and perhaps medfly, may have more complex pheromone repertoires than *D. melanogaster*. *Ceratitis* also differs from the other flies in having four genes related to the DmelGr43a lineage, which is a fructose receptor [74] expressed in both mouthparts and the brain, and details of the expression patterns of these four genes may reveal sub-functions for this lineage. Finally, the DmelOr7a lineage that includes a receptor for fruit odors in *Drosophila* [75] shows a medfly-specific expansion to ten genes, while this lineage was lost in *Musca*, which presumably does not require fruit detection. *Ceratitis* has a similar expansion of seven GRs related to the DmelGr22a-f expansion, and two proteins related to DmelGr93b-d, both lineages lost from *Musca*, raising the possibility for the bitter taste receptors involvement in fruit perception.

In summary, *Ceratitis* is more similar to *Drosophila* than *Musca* in its repertoire of chemoreceptors, most likely due to a more similar ecology, despite being more basal phylogenetically [76]. *Ceratitis* is also distinctive in having

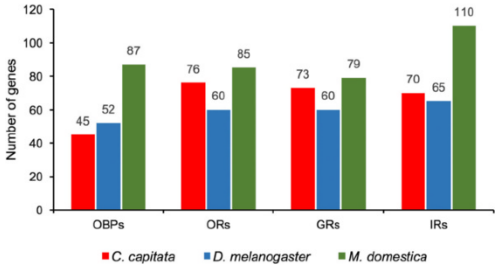


Fig. 4 Comparison of gene numbers for odorant-binding proteins (OBPs), odorant receptors (ORs), gustatory receptors (GRs), and ionotrophic receptors (IRs) in *C. capitata*, *D. melanogaster*, and *M. domestica*. Gene numbers provided above each *bar*

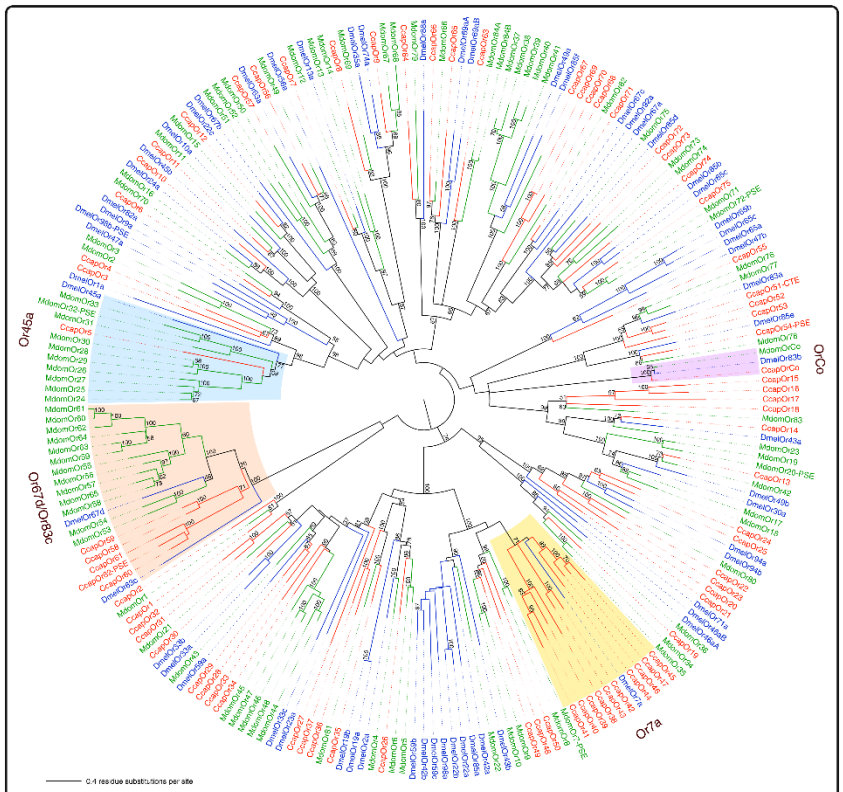


Fig. 5 Phylogenetic relationships of OR proteins from *C. capitata*, *D. melanogaster*, and *M. domestica*. The unrooted maximum likelihood (log likelihood = −140908) tree was inferred using the Le and Gascuel model [208] with a discrete Gamma distribution and some invariable sites. Bootstrap values greater than 50 % (1000 replications) are shown. Suffixes after the gene/protein names are: -CTE, C-terminus missing; -PSE, pseudogene

several separate expansions of receptors implicated in fruit detection and courtship in *Drosophila*. See Additional file 4 for further results, discussion, and protein sequences for the OBP/OR and GR/IR chemoreception gene families.

Opsin genes

In addition to odorants, field experiments have shown that the visual system plays a significant role in medfly host fruit detection, for which medflies are equipped with prominent, colorfully patterned compound eyes [77]. This has allowed the development of visual traps for both trapping and monitoring that are effective in both sexes, particularly using coloration in the yellow-orange wavelength range [78, 79]. The attractive effect of specific color hues, however, appears to vary across populations, which reflects genetic plasticity that fine-tunes color preference to the range of available host fruits [80]. In addition, shape recognition represents another vision-guided component in host fruit detection [81] and vision is further assumed to generally play a role in the courtship of tephritid flies [82]. Physiological measurements revealed sensitivity peaks in the UV range (365 nm) and in the yellow-green range (485–500 nm) [77].

These data are compatible with the apparent absence of a member of the blue sensitive opsin

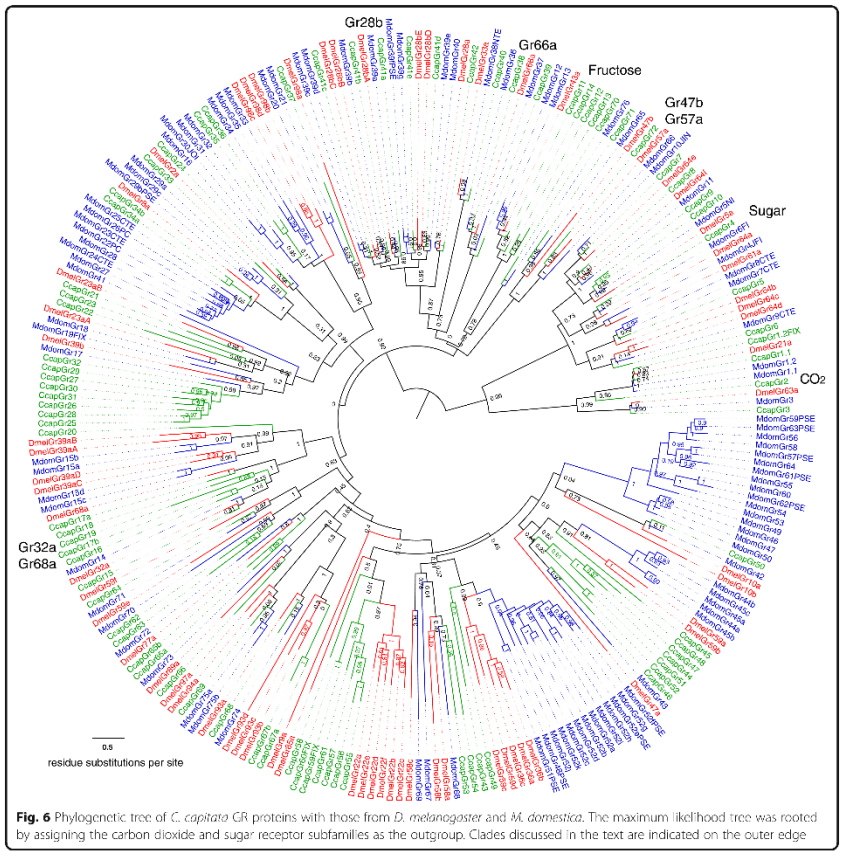


Fig. 6 Phylogenetic tree of *C. capitata* GR proteins with those from *D. melanogaster* and *M. domestica*. The maximum likelihood tree was rooted by assigning the carbon dioxide and sugar receptor subfamilies as the outgroup. Clades discussed in the text are indicated on the outer edge

subfamily in the *Ceratitis* genome, while the repertoire of long wavelength-sensitive (Rh1, Rh2, and Rh6) and UV-sensitive (Rh3 and Rh4) opsins is conserved between *Ceratitis* and *Drosophila* (Additional file 3: Figure S3). In *Drosophila*, the blue-sensitive *opsin* gene Rh5 is specifically expressed in ommatidia sensitive to shorter wavelengths ("pale" ommatidia), whereas the long wavelength-sensitive opsin Rh6 is present in ommatidia sensitive to green ("yellow" ommatidia) [83]. Given that the blue *opsin* gene subfamily is equally conserved in winged insects [84], as are the UV and green *opsin* gene subfamilies, the absence of a blue opsin from *Ceratitis* indicates unexpected regressive differences in the organization and genetic regulation of differential *opsin* expression compared to *Drosophila*, deserving further exploration.

Aquaporin genes

Twelve putative *aquaporin* (AQP) genes have been identified from the *C. capitata* genome and predicted gene sets (Additional file 2: Table S14), which is the highest number of AQP genes documented for any insect thus far (two more than those discovered in *Glossina morsitans* [85] and four more than those in *D. melanogaster* [86]). This increased gene number is the result of an expanded water

transporting Drip/Prip (Clade A) family, which has been demonstrated in other brachycerans such as *Glossina* and *Musca* [85] (Fig. 7). In addition, four putative entomoglyceroporins (Clade B), a recently described Prip-like channel capable of transporting glycerol and urea due to a single mutation of transmembrane domain 5 from a charged histidine to an uncharged amino acid [87], are present in the medfly genome, which is similar to *Drosophila*, *Glossina*, and *Musca* and twice the number found in mosquitoes (Fig. 7). As expected, at least one gene for each of the five previously identified insect AQP groups are present in the *C. capitata* genome [87, 88] (Additional file 2: Table S14). The identification of expanded Drip/Prip genes among most higher flies suggests increased or specialized water transport, but functional studies will be necessary to establish the role of these additional AQPs in relation to brachyceran physiology.

Immunity-related genes

The immunity response includes melanization, phagocytosis, encapsulation, clotting, and biosynthesis of antimicrobial peptides and proteins (AMP) by the fat body [89]. Four main signaling pathways are involved in pathogen recognition and defense response: the Toll, immune deficiency (IMD), JAK/STAT, and JNK pathways [90]. The recognition of bacteria in insects is achieved through two families of pattern recognition receptor (PRRs), peptidoglycan recognition proteins (PGRPs), and Gram-negative binding proteins (GNBPs), that bind microbial ligands and activate the immunity pathways [91–93]. Genes representative of all four pathways were identified in the *C. capitata* genome.

The *C. capitata* genome annotation initially identified 413 putative immunity genes by a search of the innate immunity databases for *D. melanogaster* and *An. gambiae*

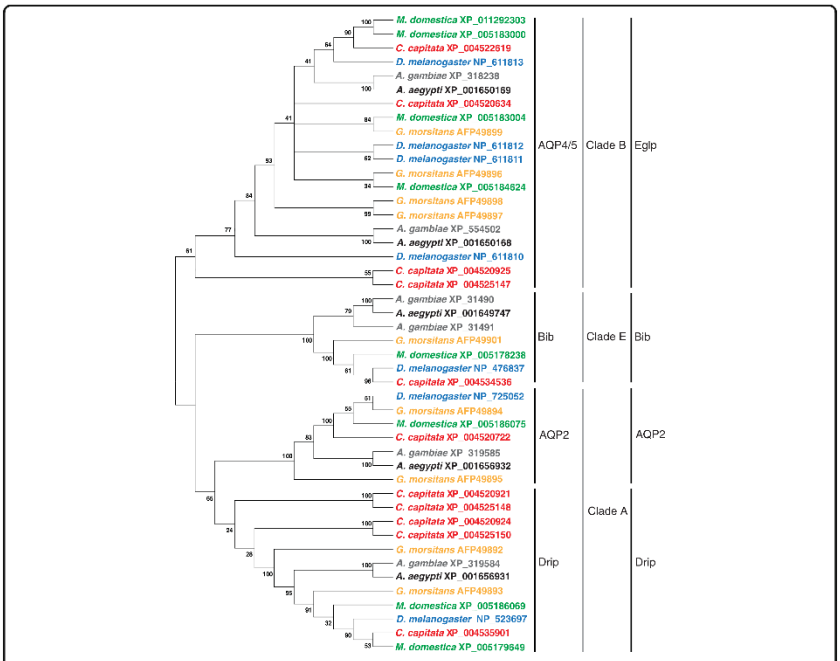


Fig. 7 Comparison of predicted AQP amino acid sequences from *C. capitata* and other indicated Diptera. The *neighbor-joining tree* was produced using MrGA6 using Dayhoff Model and pairwise matching; *branch values* indicate support following 3000 bootstraps with values below 20 % omitted. Classification is based upon Finn et al., Benoit et al., and Fabrick et al. [87, 88, 226]. *Drip Drosophila* integral protein, *Prip Pyrocoelia rufa* integral protein, *Eglp* entomoglyceroporin, *AQP* aquaporin. Unorthodox AQPs are not included in this analysis

(see http://bordensteinlab.vanderbilt.edu/IIID/test_immunity.php) [94]. Of these, 63 showed direct homology to both *D. melanogaster* and *Anopheles gambiae* putative immunity genes, 77 solely to *An. gambiae* genes, and 122 solely to *D. melanogaster* genes (Additional file 2: Table S15). The remaining 151 genes were identified, via a Hidden Markov model (HMM) analysis, by possessing all the aforementioned components of the insect immune system. The availability of additional insect genomes now allows for more comprehensive comparisons of the various insect immune systems. The total number of medfly immunity genes (n = 413) is somewhat higher than *D. melanogaster* (n = 379), but less so than the housefly (n = 771), which must cope with a pathogen-rich environment [68]. Nevertheless, the enormously diverse host-choice for medfly and its cosmopolitan nature has resulted in a robust immune system providing a defense for the diverse pathogens encountered in the various hosts and environmental conditions.

Specific gene families provide several interesting insights. The antibacterial ceratotoxin peptide family (seven genes), for example, is thus far unique to medfly and several other species in the *Ceratitis* genus, where they exist in the female accessory glands to protect the reproductive tract from bacterial infection during mating. They are also found on the surface of oviposited eggs where they may create a microbiologically controlled environment that favors early larval development [95–98]. These genes are clustered together on the X chromosome and apparently arose as a result of gene duplication [99]. The *spätzle* gene family is also highly expanded in medfly, where it is an immune response effector that activates the Toll signaling pathway [89, 100, 101] induced by Gram-positive bacteria and fungi [102]. Indeed, fungal infections commonly follow fruit punctures after oviposition, which may contribute to the expanded *spätzle* family in medfly. In this respect there is a significant expansion in the Toll receptor family, having 17 genes, relative to nine in *Drosophila* and *M. domestica*. The clip-domain serine protease gene family, required for *spätzle* activation [103, 104], is also expanded in medfly, having 50 genes relative to 45 in *D. melanogaster* and 28 in *M. domestica*. However, *Toll* is also involved in embryonic development [105–107] and, therefore, the observed expansion may involve other systems in addition to immune response.

Gene families associated with insecticide resistance and detoxification

The emergence of resistance to insecticides is recognized as a major challenge for IPM control of economically important tephritid flies such as *C. capitata* [108]. Thus, a high priority is the identification of genes associated with insecticide resistance, including the three major detoxification enzyme families (cytochrome P450, carboxylesterases, and glutathione S-transferases), known receptors/targets for the main groups of insecticides (cholinesterase, cys-loop ligand-gated ion channel, and voltage-gated sodium channels genes), and cuticle proteins. This analysis also provides significant knowledge relevant to the role of these genes and their gene families in biological processes fundamental to development and behavior.

Cytochrome P450 genes

The P450 enzymes, including mixed function oxidases and cytochrome P450 (CYP450) mono-oxygenases, have a highly diverse array of functions including synthesis of hormones critical to insect development and reproduction, as well as chemical metabolism that facilitates host plant adaptation and survival in toxic environments (e.g. insecticide detoxification). This array of functions is achieved, typically, by a large number of related, though structurally independent P450 proteins, whose total number and rate of expansion is influenced by species-specific physiology and environmental challenges.

The *C. capitata* cytochrome P450 family is composed of 103 genes and nine pseudogenes (Additional file 2: Table S16) having a greater level of expansion compared to *D. melanogaster* where 88 CYP450 genes and three pseudogenes have been identified (http://flybase.org/). This expansion is mainly found in CYP6 (clan 3) and CYP12 (mitochondrial clan) genes (Fig. 8), but is less expanded than the respective clans found in *M. domestica* [68]. The medfly CYP6 family is composed of 40 genes and four pseudogenes, almost doubling the 23 genes found in *D. melanogaster*, by the notable expansion of subfamilies CYP6A (14 genes), CYP6G (nine genes), and CYP6D (five genes). Interestingly, members of the three subfamilies have been previously associated with insecticide resistance in higher Diptera and clan 3 has been previously characterized as proliferating rapidly by gene cluster duplications [109]. Indeed, a cluster of 18 consecutive CYP genes (13 of which belong to the CYP6A subfamily) is found in the *Ceratitis* genome. Two related, but shorter clusters of two and nine consecutive CYP6 genes are found in the *D. melanogaster* genome separated by 6 Mb. Notably, orthologs for the flanking genes upstream (*mtt*, FBgn0050361) and downstream (*Kank*, FBgn0027596) of the first and the second cluster, respectively, in *D. melanogaster* are found flanking the CYP6 gene cluster in *Ceratitis* (Additional file 3: Figure S4A). The CYP6A51 gene (XP_004534861), whose overexpression had been previously associated with lambda-cyhalothrin resistance in *Ceratitis* [110], is located at one end of this cluster.

The two in tandem CYP6G genes in opposite orientation followed by one CYP6T gene found in *D. melanogaster* is also found in *Ceratitis*, having the same conserved gene order and orientation. However, the tandem array

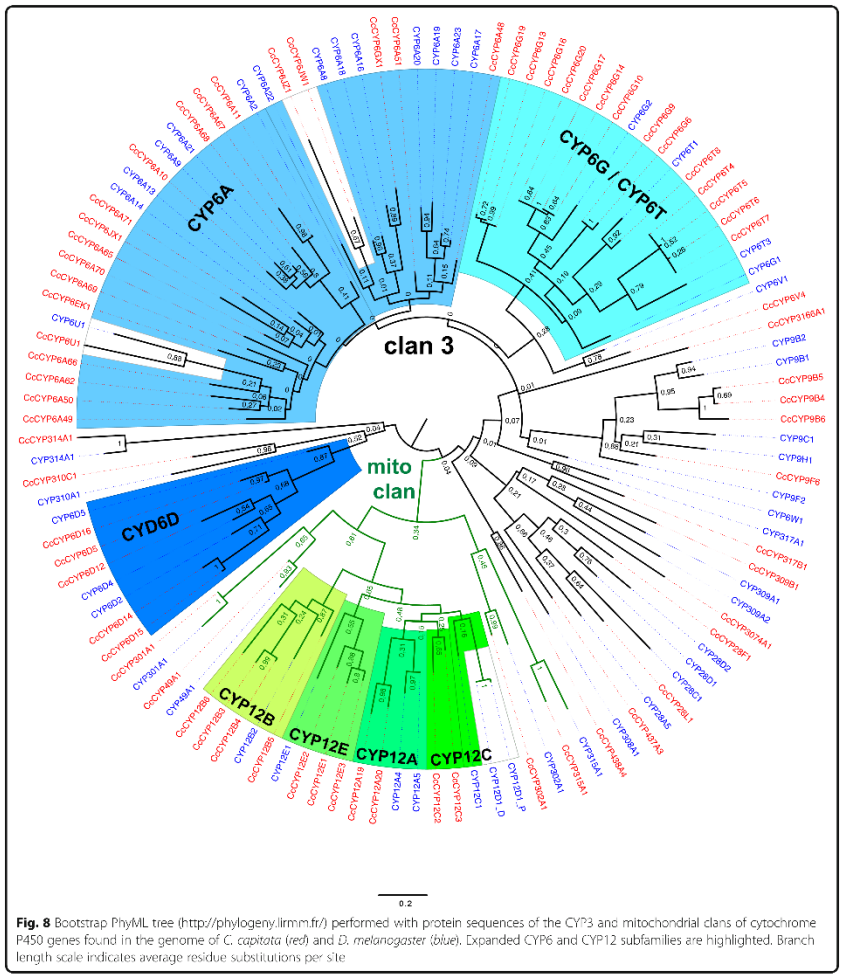


Fig. 8 Bootstrap PhyML tree (http://phylogeny.lirmm.fr/) performed with protein sequences of the CYP3 and mitochondrial clans of cytochrome P450 genes found in the genome of *C. capitata* (*red*) and *D. melanogaster* (*blue*). Expanded CYP6 and CYP12 subfamilies are highlighted. Branch length scale indicates average residue substitutions per site

in *Ceratitis* includes three CYP6G genes and one CYP6T, and is repeated at least three times resulting in nine CYP6G genes, four CYP6T genes, and four CYP6G pseudogenes (Additional file 3: Figure S4B). We have also found two more CYP6D genes in *C. capitata* than in *D. melanogaster*.

Finally, the CYP12 family is formed by 11 genes and one pseudogene in *Ceratitis*. Homologs for five of the seven genes of this family identified in *D. melanogaster*, except for Cyp12d1-d and Cyp12d2-p, are found in *Ceratitis*. Interestingly, the additional genes found in *Ceratitis* most likely resulted from duplication events, since they are

located in four clusters of two CYP12A (+1 pseudogene), four CYP12B, two CYP12C, and three CYP12E genes (+1 pseudogene). These duplications may be explained by their participation in environmental responses, suggesting that these genes should be of interest to studies on P450-mediated resistance in *Ceratitis*. Indeed, genes of the CYP12 family have been associated with insecticide resistance in both *Musca* and *Drosophila* [109].

Among cytochrome P450 genes there are also components of the ecdysone biosynthesis pathway. 20-hydroxyecdysone (20E) plays a critical role in both early development and female reproduction in most, if not all, insects and, thus, can be targeted by highly specific insect growth regulators (IGR) for population control. Of particular interest are orthologs of the four P450 Halloween genes that act in the final steps of ketodiol conversion to the active hormone [111] (Additional file 2: Table S16). Those found in *Ceratitis* include: *phantom* (*CcCYP306A1*), *disembodied* (*CcCYP302A1*), *shadow* (*CcCYP315A1*), and *shade* (*CcCYP314A1*). By contrast, one of the two genes that may participate as stage-specific components in 20E biosynthesis in *D. melanogaster* (*Cyp307a1*, *spook*, and *Cyp307a2*, *spookier*) is absent in medfly. The gene found in *Ceratitis* contains an intron and has been consequently named *CcCYP307A2*, considering that *D. melanogaster Cyp307a1* is an intronless mRNA-derived paralog of *Cyp307a2* [112] that is only found in the Sophophoran subgenus of *Drosophila* [113].

Carboxyl/cholinesterase genes

Insect carboxyl/cholinesterases have been classified within 14 clades (A to N), while two of the clades (C and G) have yet to be found in higher Diptera [114]. A total of 35 carboxyl/cholinesterase genes were identified in *D. melanogaster* while 43 were found in *C. capitata*. Orthologous genes for all members of the acetylcholinesterase and non-catalytic neurodevelopmental clades (I to N) are found in *Ceratitis* with no additional gene duplications or deletions. The differences in the number of carboxyl/cholinesterases genes in *Ceratitis* compared to *D. melanogaster* are found in clades A, D, E, F, and H (Fig. 9; see Additional file 2: Table S17). Three tandem copies of a *cricket-like* gene (clade A, FBgn0000326), putatively influencing male mating behavior in *D. melanogaster* [115], are found in *C. capitata*. The microsomal alpha-esterase gene cluster [116] (Clade B) involved in detoxification is also found in the *C. capitata* genome, and has two extra genes compared to *D. melanogaster*. Here we find two copies of the "*aliesterase*" gene (*alpha-7* or *E3*) that has been associated with organophosphate resistance in *M. domestica* and *Lucilia cuprina* [117]. Interestingly, an unknown mechanism of malathion resistance mediated by an alteration in aliesterase activity is also being studied in *C. capitata* [118]. The cluster of *Est6* and *Est7* genes (Clade E, beta-esterases) conserved in the *Drosophila* species group [116] is not found in the *Ceratitis* genome, which possesses only one of the two genes. Nonetheless, the total number of genes in Clade E is preserved in *C. capitata* due to the tandem duplication of another gene, similar to CG6414 (FBgn0029690). Finally, a single *juvenile hormone esterase* (*JHE*) gene and three tandem copies of the *juvenile hormone esterase duplication* (*JHEdup*) gene, which encode proteins lacking the GQSAG motif found in active JHE [119], are found in *Ceratitis*. A proliferation in the *JHEdup* gene by a duplication event in cactophilic *Drosophila* has been associated with ethanol degradation in rotting fruits [120], although the exact function of these genes remains uncertain. Nevertheless, *JHE* and related genes involved in JH metabolism may also provide important targets for IGRs.

Glutathione S-transferase genes

Glutathione S-transferases (GSTs) are a large family of proteins that are involved in metabolic detoxification. We have identified 28 cytosolic GST genes in the *C. capitata* genome and 1 pseudogene (Additional file 2: Table S18), which is fewer than the 36 GST genes found in *D. melanogaster* and equal to the number of genes in *M. domestica*. Phylogeny-based classification has allowed their grouping into six GST subclasses (Additional file 3: Figure S5): seven *Delta* GSTs, 14 *Epsilon* GSTs, one *Omega* GST, three *Theta* GSTs, two *Zeta* GSTs, and one *Sigma* GST. As in *Drosophila*, many of the insect-specific genes of the *Delta* and *Epsilon* subclasses, putatively involved in insect responses to environmental conditions [121, 122], are organized within clusters in the *C. capitata* genome. Genes of these two subclasses have been involved in insect responses to xenobiotics and in insecticide resistance [121, 122], such as OP-resistance in *M. domestica* mediated by gene amplification of *MdGSTD3* and DDT-resistance in *D. melanogaster* associated to overexpression of *DmGSTD1*. However, resistance to insecticides mediated by GSTs have not yet been reported in *Ceratitis*, which may be related to the few cases in which resistance mechanisms have been elucidated for this species when compared to *Musca* and *Drosophila*.

Cys-loop ligand-gated ion channel superfamily genes

Members of the cys-loop ligand-gated ion channel (cysLGIC) superfamily [123], including the highly conserved nicotinic acetylcholine receptor (nAChR) subunits and the GABA receptors, GluCls and HisCls, are targets for insecticides. In the medfly genome we find orthologous genes for most of the cysLGIC members described in insects (Additional file 2: Table S19). Interestingly, an additional divergent nAChR subunit gene,

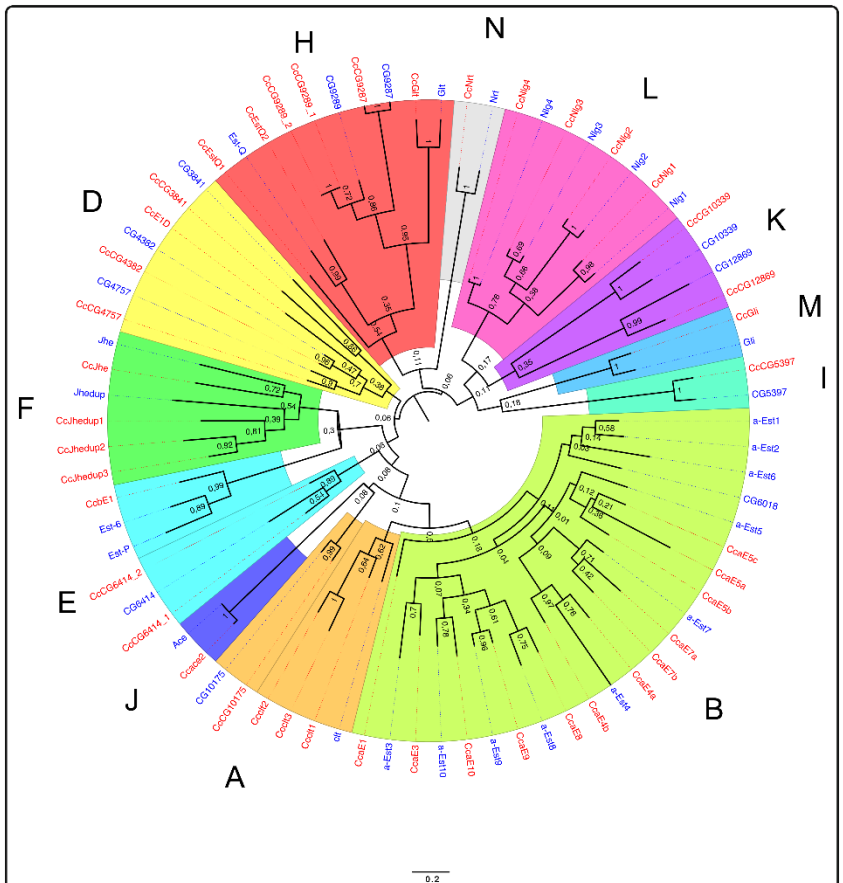


Fig. 9 Bootstrap PhyML tree of *C. capitata* (*red*) and *D. melanogaster* (*blue*) esterase protein sequences (http://phylogeny.lirmm.fr/). Clades are indicated by letters, A–N. *Branch length scale* indicates average residue substitutions per site

coding for an α subunit receptor (nAChRα8) that conserve the two adjacent cysteine residues involved in acetylcholine binding, is identified in *Ceratitis* (Additional file 3: Figure S6). Orthologous genes for this divergent subunit are only found in other tephritid flies such us *Bactrocera cucurbitae* (XP_011189556), *B. dorsalis* (XP_011213957), and *B. oleae* (XP_014090995). This represents a minor expansion of the *C. capitata* nAChR group, which has also been observed in *Apis mellifera* (nine α and two β subunits), *Tribolium castaneum* (11 α and one β subunits), and, more noticeably, in *Nasonia vitripennis* (12 α and four β subunits) [124] (Additional file 2: Table S19). Recently, resistance to spinosad, a major natural control compound for medfly, has been selected in a *C. capitata* laboratory strain [125]. Since spinosyns target nAChR subunits, the functional characterization of this group

should be important to elucidating the molecular mechanism involved in this resistance.

Voltage-gated sodium channel genes

The voltage-gated sodium channel, the target site for DDT and pyrethroid insecticides, is composed of a pore-forming subunit encoded by a unique gene with extensive alternative splicing and RNA editing, which generate a large collection of sodium channel isoform variants [126] interacting with auxiliary subunits that modulate their function [127]. In *Ceratitis*, orthologs are found for the *D. melanogaster DmNa$_v$* gene (formerly *para*) and the auxiliary subunits: *TipE* and four *TipE*-homologous genes (*Teh1*, *Teh2*, *Teh3*, *Teh4*) [127, 128]. The genomic arrangement of the TipE gene family members in *C. capitata* coincides with the one observed in *D. melanogaster* (*TipE*, *Teh2*, *Teh3*, and *Teh4* genes in a cluster separated from *Teh1*), which is believed to be conserved among the Insecta [129]. A well characterized mechanism of resistance to pyrethroids is target site insensitivity mediated by mutations in the voltage-gated sodium channel gene [126], often referred to as knockdown resistance or "kdr resistance." The information acquired after sequencing the genome would be highly valuable to study this complex target. Resistance to pyrethroids has been reported in both Spanish field populations and a laboratory selected strain of *Ceratitis* [110]; however, this resistance is suspected to be mediated by P450 detoxification as mentioned above.

Cuticle protein genes

Using sequence motifs that are characteristic for several families of cuticle proteins [130], 202 genes coding for putative cuticle proteins were identified (Additional file 2: Table S20). These genes were analyzed with CutProtFam-Pred, a cuticular protein family prediction tool [131], and 195 genes could be assigned to one of eight families (CPR, CPAP1, CPAP3, CPF, CPCFC, CPLCA, CPLCG, and TWDL) (Additional file 2: Table S21). The remaining seven proteins lack a defining conserved domain but possess characteristics commonly associated with cuticle proteins, including the repeated low-complexity sequences (AAP[A/V]/GGY). Many of the genes (~77 %) are arranged in clusters of 3–28 genes (Additional file 2: Table S22 and Additional file 3: Figure S7) that are primarily specific to the type of cuticular protein. However, in several cases multiple family types were found in a single cluster. The size and number of clusters was similar to that observed in other species [132, 133] and may be a common feature of cuticle proteins in arthropods. Clustering may be important for the coordinated expression of these genes during critical points in development [134] and it has been suggested that clustering could facilitate the development of insecticide resistance [132].

Similar to other insects, the CPR family, with the RR-1 (soft cuticle), RR-2 (hard cuticle), and unclassifiable types, constitute the largest group of cuticle protein genes in the *Ceratitis* genome. The 110 CPR genes identified in medfly are comparable to the number in *Tribolium*, but slightly less than the 137 genes found in *D. melanogaster* [134, 135]. The number of genes in the protein families CPAP1, CPAP3, CPF, CPCFC, and CPLCG are similar to the number in other insects [134, 136]; however, *Ceratitis* shows an expansion in the Tweedle (TWDL) and CPLCA families similar to that in *D. melanogaster* and *M. domestica* (Fig. 10; see Additional file 3: Figure S8). Most insects have only 2–5 members of the TWDL family, while dipterans have an expanded family with mosquitoes (Culicidae) possessing 6–12 Tweedle genes, and *Drosophila*, *Musca*, and *Ceratitis* exhibiting a greater expansion of ~30–40 Tweedle genes. Similarly, *Drosophila* and *Ceratitis* also show a greater number of CPLCA genes (13–25 genes) than that found in other species (1–9 genes). The notable exception to the expansion of these gene families within Schizophora is *Glossina* (Fig. 10). The expansion of cuticle protein families likely reflects adaptive evolution [133, 135] and the lack of expansion in *Glossina* may reflect the difference in developmental strategies among these dipterans, with larval development occurring in utero in *Glossina* females. However, the precise role of these protein families, and the functional implications of expanded gene families, requires further study.

Gene families associated with sex, reproduction, and population control

SIT is the major biologically-based tactic for the control of medfly populations, in addition to several other tephritid pest species. Genomic data related to sex-determination, sex-specific gene expression, reproduction, and programmed cell death have already proven essential to the development of transgenic strains that improve the efficiency of female-lethality for male-only populations, male sterilization, and genetic marking systems for sperm and trapped males. Continued development and improvement of these strains will depend on new strategies that result in: (1) the suppression of testis-specific genes to induce male sterilization (as an alternative to irradiation); (2) the manipulation of sex determination genes in chromosomal females resulting in their development as sterile phenotypic males; (3) the identification of lethal effector genes for tissue and sex-specific conditional lethality; and (4) defining the molecular effects of seminal fluid proteins on female physiology and behavior, which may provide mechanisms that interfere with medfly reproduction.

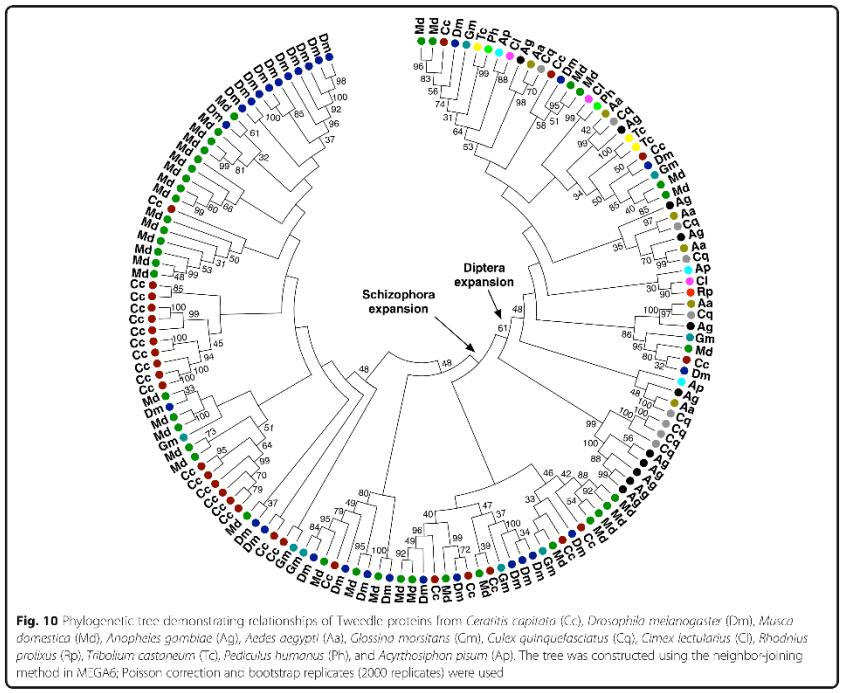


Fig. 10 Phylogenetic tree demonstrating relationships of Tweedle proteins from *Ceratitis capitata* (Cc), *Drosophila melanogaster* (Dm), *Musca domestica* (Md), *Anopheles gambiae* (Ag), *Aedes aegypti* (Aa), *Glossina morsitans* (Gm), *Culex quinquefasciatus* (Cq), *Cimex lectularius* (Cl), *Rhodnius prolixus* (Rp), *Tribolium castaneum* (Tc), *Pediculus humanus* (Ph), and *Acyrthosiphon pisum* (Ap). The tree was constructed using the neighbor-joining method in MEGA6; Poisson correction and bootstrap replicates (2000 replicates) were used

Sex-determination, sex-linked, and sex-specific genes

Sex determination is a fundamental developmental process that regulates male-specific and female-specific sexual differentiation, and thus, various sex-specific aspects of fertility, courtship behavior, and, in some species, dosage compensation. Given the importance of these functions to medfly IPM population control, the sex-determining genetic constituents in this species, and their interactions, have been studied for many years [44, 137]. This has revealed many commonalities with *Drosophila*, including most of the identified sex-determination genes, yet several important distinctions exist [138].

In *Drosophila*, the *Sex-lethal* (*Sxl*) > *transformer/transformer-2* (*tra/tra-2*) > *doublesex/fruitless* (*dsx/fru*) sex-determination gene pathway hierarchy initiates female-specific differentiation when *Sxl* > *tra* transcripts encode functional splicing factors resulting in *dsx*-female expression. In the default state, when *Sxl* > *tra* nonfunctional products result in *dsx*-male expression, male-specific differentiation ensues. Orthologues to all of these genes were previously identified and tested in *C. capitata*, resulting in a similar hierarchy of activity [44, 45, 139, 140], except that the *Sex-lethal* ortholog (*CcSxl*) does not act as the upstream regulator of *Cctra* in females. Indeed, *Sxl* does not have any apparent sex-determining function in *Ceratitis*, similar to other other non-drosophilid dipterans including *Musca* [141, 142]. Alternatively, it has been shown that *Cctra* activity is required in XX embryos to establish the female developmental pathway, initiating *Cctra* positive auto-regulation by maternal *Cc*TRA, while male differentiation is regulated by a male determining factor (M-factor) that prevents *Cctra* activation [44, 143].

To further elucidate and confirm the relationship between sexual differentiation in *Drosophila* and *C. capitata*, 35 cognates of *Drosophila* genes were identified in the medfly genome that are directly or indirectly involved in sex determination or sexual differentiation (25 genes including *Cctra*, *Ccdsx*, and *CcSxl*), six sex-specifically spliced genes, and four genes having somatic sex-specific functions such as dosage compensation

[144, 145] (see Additional file 2: Table S23). A tBLASTn analysis showed sequence conservation for all 35 orthologs, having amino acid sequence identities in the range of 35–98 %, with 20 genes expressed at early embryonic stages [142]. Novel sex-determining genes have evolved from gene duplications in other insects (e.g., *complementary sex-determiner/feminizer* in *Apis*, *Nix* in *Aedes*, and *Sxl/sister-of-Sex-lethal* [*ssx*] in *Drosophila*) [140, 146, 147], though paralogs of the *Ceratitis Sxl*, *tra*, and *tra-2* genes have yet to be identified in the medfly genome. While *Nix* is part of the male-determining M-locus in *Aedes*, its relationship to sex-determining genes in medfly (and other species) appears to be limited to the RNA recognition motif (RMM) most often found in *tra-2*. Thus, the molecular nature of the upstream splicing regulator(s) of *Cctra* and the putative Y-linked male determining factor have yet to be clarified, which remains a high priority [137, 142].

Known Y-linked genes are highly limited, and while none are known to encode the M factor, they do provide scaffold identification for sequences that are potentially related. These include four 1–6 kb highly repetitive Y-linked genes (GB acc: AF071418.1, AF154063.1, AF115330.1, and AF116531.1) first identified in a phage library and found to be male-specific and Y-linked by southern blot and mitotic chromosome in situ hybridizations, respectively [20] (see Additional file 2: Table S6). Y-linkage was later confirmed by a Bowtie mapping analysis of $>10^8$ male and female genomic reads against the four sequences [137].

For X-linked scaffold identification, the *ceratotoxin* (*ctx*) genes (GB acc: CtxA2, Y15373.1; CtxC1, Y15374.1, and CtxC2-CtxD, Y15375.1), previously mapped by in situ hybridization to the mitotic X chromosome [21], were found on the 6.4 Mb genomic scaffold 23 (NW_004523725), which otherwise provides very low gene content (14 transcribed regions), as expected for a highly heterochromatic chromosome. Within a 0.8 Mb flanking region of the *ctx* family, only the orthologs of *Drosophila carboxylesterase 4* and *tolloid* exist (*Cc*CG4757-like).

The identification of key medfly sex-determining genes has been important to novel sexing strategies for SIT population control, that have incorporated the *Cctra* sex-specific first intron splicing cassette into cell death genes to achieve female-specific lethality [24, 26, 27, 148]. Of particular interest has been the potential use of conditional knock-outs of *Cctra* or *Cctra-2* to transform chromosomal females to phenotypic XX males [44, 45] for high level production of male-only populations for SIT release programs.

Seminal fluid protein genes

Insect seminal fluid proteins (SFPs), transferred from males to females during mating along with sperm, are powerful modulators of multiple aspects of female reproductive physiology and behavior, including sperm storage and use, ovulation, oviposition, and receptivity to re-mating [149–153]. These proteins belong to functional classes that are rather conserved across different insect species, and include proteases and protease inhibitors, lipases, sperm-binding proteins, antioxidants, lectins, and prohormones [154, 155]. However, their identification based on sequence similarity searches is challenging, as many have been shown to undergo rapid evolution and gene expansion [156]. This can be explained by the critical roles they play in sperm activation, gamete interaction, and ovulation. Only limited information is currently available relevant to the molecular identity and functional roles of medfly SFPs [157–160]. Recent transcriptomic analyses on the testes and male accessory glands identified transcripts that exhibit mating-induced changes in abundance, most likely related to replenishment of their protein products after multiple matings [159] that are frequent in nature [161–164]. Patterns of sperm use in twice-mated females have also been investigated, revealing that sperm are stored in the female fertilization chamber in a stratified fashion, mostly likely to initially favor the fresher ejaculate from the second male [165]. Studies on the effects of SFPs on female physiology and fertilization dynamics may provide the key to understanding how sperm mobilization within the female reproductive tract is regulated.

A total of 459 genes were annotated in the medfly genome and grouped into 17 functional classes based on the categories defined for *Drosophila* SFPs [166] (see Additional file 2: Table S24 and Additional file 3: Figure S9). The most abundant class corresponds to predicted protease genes, genes involved in lipid metabolism and chitin binding, and sequences with yet unknown function, respectively. Comparison of transcriptional levels between male (ISPRA SRR836190) and female (ISPRA SRR836189) whole body RNA-Seq libraries, as well as reproductive tissue datasets revealed that 37 of all annotated genes are male-biased, with 31 of them being predominantly transcribed in the male reproductive tissues (see Additional file 2: Table S24). These features make them particularly interesting candidates for further functional analysis, although it is noteworthy that SFP-encoding genes do not necessarily display a male-biased expression profile [166, 167].

Proteases also represent a major class among *Drosophila* SFPs, which are thought to be involved in the regulation of female post-mating responses [168], including cleavage of inactive molecules into their active forms [169]. A previous analysis of medfly testes and male accessory glands expressed sequence tags (ESTs) found that one of the proteases, *trypsin alpha-3*, is a mating-responsive gene [159]. This gene, indeed,

displays a significant increase in transcript abundance immediately after male copulation, including after successive matings. This may indicate that the depletion of its protein products may trigger transcription to replenish the proteins to be transferred upon mating.

Lipid metabolism genes that may encode SFPs are also abundant in the medfly genome and include sequences that may be active in the breakdown of complex energy sources to be used by stored sperm, or in the remodeling of the sperm phospholipid membrane for capacitation [170]. The high number of genes encoding proteins with predicted chitin-binding activity may be related to antimicrobial roles, and indeed, chitin-binding abilities have been reported for several antifungal peptides [171]. Proteins with such chitin-binding activity have been previously identified not only in *Drosophila* [166], but also in *An. gambiae* [172]. We also identified genes putatively encoding proteolysis regulators, which are a highly represented protein class in the seminal fluid of multiple species [173]. This finding supports the notion that proteolysis-mediated sperm activation might have broad phylogenetic conservation and that proteolytic activity is essential for male reproductive success [174].

The identification of genes encoding proteins involved in odor perception is in agreement with several studies reporting the expression of such genes in the male accessory glands and testes of multiple species [175–182]. The identification of the medfly putative orthologs for Obp56e and Obp56g, which encode proteins found in the *Drosophila* seminal fluid, suggests that medfly OBPs may act as carriers for physiologically active ligands, such as hormones, that are transferred from the male to the female upon mating.

Approximately 10 % of the putative SFP genes annotated could not be associated to a specific functional class. Interestingly, several of these genes (n = 19) displayed a male-biased transcriptional profile and a particularly high abundance in male reproductive tissues. Among them is *CG5867-like*, for which previous ESTs analyses revealed a transcriptional profile possibly related to the replenishment of ejaculate components after mating [159]. Its *Drosophila* ortholog has a hemolymph juvenile hormone-binding domain that has been suggested to be involved in the regulation of hormone levels. Among the genes with unknown functions, of particular interest is the Uncharacterized protein LOC101454281. While lacking significant sequence similarity to known sequences, the presence of multiple glycosylation sites allows us to speculate on its potential mucin nature. In *Drosophila*, mucins have been shown to participate, together with other proteins and lipids, in the formation of mating plugs, often produced within the female reproductive tract during or shortly after mating [149, 183]. Medfly does not produce a plug, but mucins may have a role in protecting sperm and assisting their movement through the female tract, as occurs in mammals [184, 185].

These data lay the foundation for deeper proteomics-based investigations aimed at identifying and quantifying the peptides delivered to the female reproductive tract by medfly males. A deeper understanding of the identity and functional roles of the medfly SFPs will allow their exploitation for manipulating female reproductive physiology, behavior, and fertility. This could possibly lead to the development of novel environmentally acceptable species-specific chemosterilants [186, 187], capable of mimicking the behavior-modulating effects of an SFP by impeding correct sperm storage or interfering with female re-mating.

Programmed cell death genes

Pro-apoptotic proteins from the *reaper* (*rpr*), *hid*, *grim* (RHG) gene family, first described for insects in *Drosophila*, are primary regulators of programmed cell death by their negative control of the inhibitor of apoptosis (IAP) proteins, thereby allowing caspase activation resulting in cell death [188]. As such, they have critical roles in development, especially in the larval to adult cellular transitions during metamorphosis, and the removal of cells damaged by environmental stress. In *Drosophila*, their vital roles in development have been demonstrated by lethality resulting from *hid* and *rpr* null mutations or their ectopic misexpression from transgenes [189]. Ectopic misexpression of *hid*, in particular, has been used as an effective lethal effector for uni-sex and female-specific conditional lethality for improved SIT in several species, including medfly where the *Drosophila* cognate was found to be functional [28]. In the caribfly, *Anastrepha suspensa*, the native RHG genes were isolated and functionally validated using cell death assays [190], with the *A. ludens* (mexfly) *Alhid* cognate subsequently used for highly effective conditional lethality in *A. suspensa* [25].

Conservation between the *Drosophila* and medfly apoptotic cognates was first tested by performing the BLASTn algorithm on 95 *Drosophila* genes from four Gene Ontology (GO) groups against the medfly genome: programmed cell death (GO:12501), germ cell programmed cell death (GO:35234), negative regulation of apoptotic process (GO:43066), and apoptotic processes (GO:6915) (Additional file 2: Table S25). Eighty-one genes were highly conserved, of which 57 had e-values less than $1e^{-30}$, while 14 genes did not show significant similarity to medfly sequences. Notably, the medfly *reaper*, *hid*, and *grim* genes were identified by similarities to multiple regions of the orthologous protein, while *sickle* was only identified by homology of two conserved protein motifs after an additional tBLASTx algorithm

search—one being the N-terminal IAP motif and the second being the GH3-binding motif. Both motifs are essential for apoptotic function in *D. melanogaster* and *A. suspensa* [190–192] and their conservation suggests a conserved pathway for the two species.

Comparisons of the genomic structure of *Drosophila reaper*, *hid*, *grim*, and *sickle* to their medfly orthologs revealed conserved synteny and genomic organization of the respective regions. In *Drosophila*, all four genes are located within a 272 kb region on chromosome 3L, while in *Ceratitis*, the region is located on scaffold 2 (NW.004523691) that maps to chromosome 6R (see Fig. 2, Additional file 2: Table S6). These loci are syntenic based on the polytene map [193], and the region in medfly is nearly three-fold longer, consistent with relative genome sizes (Fig. 11). Nevertheless, the orientation and relative distances among the genes are conserved between the two species.

Conclusions

Here we report the whole genome sequence of the Mediterranean fruit fly, *C. capitata*, which is one of the most highly invasive and destructive plant pests throughout the world. Of particular interest are the comparative relationships of gene families between this species and two closely related dipterans, *D. melanogaster* and *M. domestica*, having implications for the adaptation and invasiveness of medfly and presenting specific methods and targets for control of its population size. Current research may also utilize the genome assembly as a foundation to characterize population structure of this pest insect throughout tropical regions of the world and the genomic context for genetic sexing strategies for SIT can now be explored utilizing this as a reference genome.

The final 479 Mb genome assembly size varies from our 538.9 Mb k-mer value and an earlier 540 Mb Feulgen stain estimate [11], and we presume this is due to an inability to assemble highly repetitive heterochromatic sequences that account for approximately 11 % of the genome. The high quality of the assembly, however, is reflected in a contig N50 of 45.7 kb and scaffold N50 of 4.06 Mb, and the integrity of 11–15.8 Mb scaffolds consistent with physical mapping. We conclude that this resulted from the use of genomic DNA from highly inbred single and small-pooled flies to minimize polymorphisms. This is in comparison to an initial sequencing attempt using DNA extracted from non-inbred flies from the same laboratory strain, which yielded low-quality assemblies. This protocol has now been established for all species in the i5K pilot project and should serve as a guide for future projects.

The adaptation of medfly to diverse fruits and vegetables, and its successful invasion of associated habitats, may be related to specific gene expansions relative to *Drosophila* and *Musca*. The more similar behavioral ecologies between *Drosophila* and medfly, for instance, seem to be reflected in more similar expansions of the IR and GR taste receptor gene families, as well as receptors for pheromone attractants, compared to the housefly disease vector, *M. domestica*, which is considered to be more closely related to *D. melanogaster* [76]. On the other hand, the larger number of cytochrome P450 genes and the common expansion of CYP6 subfamilies in medfly and *Musca*, relative to *Drosophila*, may reflect their cosmopolitan nature requiring an increased need for adaptation, as well as their more pestiferous behavior. This comparison is consistent with the higher number of immunity system genes relative to *Drosophila*, with the notable expansion of the *Toll* and *spätzle* families, and the unique existence and expansion of *ceratotoxins*, thus far specific to medfly, and both of which may enhance protection of eggs

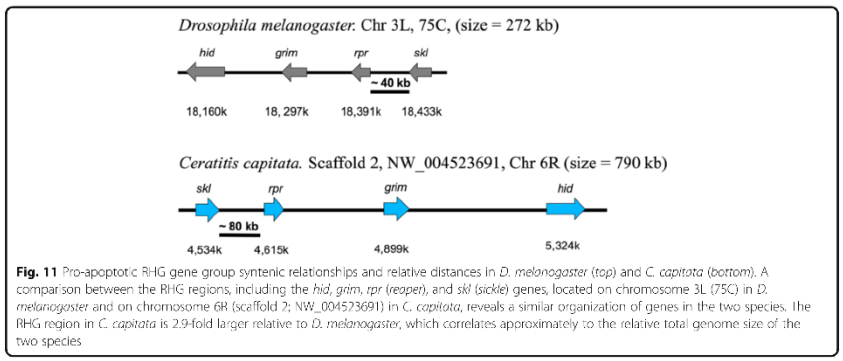


Fig. 11 Pro-apoptotic RHG gene group syntenic relationships and relative distances in *D. melanogaster* (*top*) and *C. capitata* (*bottom*). A comparison between the RHG regions, including the *hid*, *grim*, *rpr* (*reaper*), and *skl* (*sickle*) genes, located on chromosome 3L (75C) in *D. melanogaster* and on chromosome 6R (scaffold 7; NW_004523691) in *C. capitata*, reveals a similar organization of genes in the two species. The RHG region in *C. capitata* is 2.9-fold larger relative to *D. melanogaster*, which correlates approximately to the relative total genome size of the two species

oviposited into diverse microbial-rich environments. Adaptation and invasiveness for medfly may also be reflected in the expansion of the TWDL and CPLCA cuticle protein families, and the highest number of *aquaporin* genes reported thus far in insects.

Another critical role for the medfly genome analysis is defining potential gene targets and genetic reagents relevant to the IPM control of its behavior and population size. The functional characterization of chemoreception molecules, especially those implicated in courtship/pheromone or fruit detection, may permit the development of new synthetic ligands that act as attractants, repellents, or antagonists to disrupt oviposition or mating behaviors. These molecules may also be used as lures for trapping or pest population monitoring. Knowledge of the spectral sensitivity of opsins should also guide the use of optimal trap colors, and while there is conservation for most of the *opsin* genes with *Drosophila*, the absence of the typically conserved blue *opsin* subfamily in medfly is a notable distinction that should be explored. More generally, an improved understanding of the genetic systems involved in fundamental biological processes should aid in the development of specific and more ecologically sound insecticides. The genomic data generated will also facilitate the development of specific tools for the detection of incipient resistance and the implementation of appropriate resistance management strategies.

Genetically modified strains are currently being developed for control of medfly and related tephritid species, and further advances in the effectiveness and safety of these strains will depend on the identification and isolation of stage, tissue, and sex-specific genes and their promoters, and the lethal effectors they will control. The RHG pro-apoptotic cell death gene family was identified in medfly and found to have a genomic organization analogous to *Drosophila*. Functional conservation of these genes is also expected, raising the possibility of using the *reaper* and *grim* genes, along with *hid*, for redundant secondary lethality [194]. SFPs present a large repertoire of potential targets for controlling reproduction, including peptides that might be modified for improved suppression of female multiple matings to enhance SIT.

Central to current gene modification protocols is the use of transposon-mediated germline transformation, for which new and highly efficient TE systems are an ongoing need. Thus, the relatively high euchromatic representation of all the DNA transposon superfamilies is encouraging for the discovery of new elements that may be used as vectors. However, these elements may also have the potential for cross-mobilizing related elements, some of which are in current use as transformation vectors. The full genome sequence now allows the identification of potential mobilizing and cross-mobilizing systems, which is critical to evaluating potential instability of transposon vectors for genome modification and mitigation of associated risks. Therefore, the lack of apparent mobilizing systems in the medfly genome for current vectors diminishes the potential for transgene instability and possible lateral inter-species transfer. This alleviates a serious concern for an environmental risk related to the release of genetically modified medflies.

The relatively high quality of this genome sequence is also a prerequisite for highly specific gene editing and the identification of specific genomic sites that can be used for insertional targeting that avoids deleterious genomic effects. Routine use of such target sites could be made most efficacious by initially introducing small recombination sites by gene editing that could be subsequently used for repetitive recombinase-mediated megabase transgene insertions and deletions [29, 30]. This would allow for genomic modifications that avoid position effects on transgene expression and insertional mutations that debilitate the host strain, resulting in enhanced functional studies and modified strains for the most effective and ecologically sound means of population control.

Methods

Genome sequencing and assembly

The second attempt to improve genome sequence quality was initiated by consecutive single pair sib inbreeding of the *C. capitata* ISPRA strain for 20 generations, to achieve a high level of genome homozygosity thereby reducing sequence polymorphisms that were the likely cause of previous weak assemblies (Additional file 1: Supplementary material A). The more highly successful genome assembly reported here was thus, importantly, dependent on this level of inbreeding, and it was used as a strong recommendation to perform inbreeding for 10–12 generations for all i5K projects [10], or as many generations as possible. The assembly reported here used an enhanced Illumina-ALLPATHS-LG sequencing and assembly strategy that is being used for other species in the BCM-HGSC i5K pilot project, which enabled multiple species to be approached efficiently in parallel.

Details for the insert library preparation and sequencing are available in Additional file 1: Supplementary material B. Briefly, the sequencing read data were assembled using ALLPATHS-LG (v35218) [195] and further scaffolded and gap-filled using the in-house tools Atlas-Link (v.1.0) and Atlas gap-fill (v.2.2, https://www.hgsc.bcm.edu/software/). Alignments were conducted as part of the ALLPATHS-LG assembly process and the true insert size of mate pair libraries was estimated, with the 8 kb library adjusted to 6.4 kb. This assembly was 484.7 Mb in total length with a contig N50 of 45,711 bp and a scaffold N50 of 4.06 Mb;

however, initial annotations revealed the presence of significant bacterial sequence that was identified as 5.7 Mb of endosymbiotic bacterial sequences localized to 18 scaffolds (see Additional file 1: Supplementary material C for details). Analysis of the metagenomics content was conducted using Blast [196] and Kraken [197] and is described in detail within Additional file 1: Supplementary material C. Removal of the contaminant sequences (with the GenBank assembly accession iterated to GCA_000347755.2) revealed a final genome size of 479.1 Mb with a contig N50 of 45,879 bp and a scaffold N50 of 4.12 Mb, which has been deposited in the NCBI: BioProject PRJNA168120 (see Table 1 and Additional file 2: Table S1).

Genome annotation and downstream informatics analysis

To facilitate annotation, RNA-Seq data were generated from three samples, including mixed-sex embryos and whole body male and female adults, using RNA extracted with Trizol reagent (Life Technologies) followed by DNase treatment (DNAfree, Ambion). A total of 5.3 Gb sequence data were produced for the embryo, 4.9 Gb for the female, and 7.8 Gb for the male samples (see Additional file 2: Table S1 and NCBI BioSample: SAMN02055687 - SAMN02055689). These data were aggregated with data contributed by the community (see Additional file 2: Table S26; complete dataset including experimental procedures available at GEO, accession number GSE80605). The assembly was annotated by the consortium using three distinct approaches: (1) Maker 2.0 [15] at HGSC with the assembled genome and adult male and female RNA-Seq data used to improve gene models; (2) at NCBI using the Gnomon pipeline; and (3) our in-house JAMg, all publicly available with details for Maker 2.0 and Gnomon described in Additional file 1: Supplementary material B. Briefly, JAMg was used to produce automated annotations which made use of GSNAP, Trinity RNA-Seq de novo, Trinity RNA-Seq genome-guided [198], PASA [199], Augustus [200], and other tools before deriving a consensus gene set using EvidenceModeler [201]. Then PASA was used again to annotate UTRs and generate alternative splicing isoforms. Annotations from the three platforms were provided to the curation community using the WebApollo JBrowse tool as hosted by the USDA, National Agricultural Library (https://apollo.nal.usda.gov/cercap/sequences).

Orthology assignment

The final annotation set for *C. capitata* (CeC) was compared to other arthropod genomes to characterize orthology. First, the following annotation sets were extracted from genomic databases for the arthropod species: *Acyrthosiphon pisum* (AcP), aphidbase.com ACYPI OGS 2.1B; *Aedes aegypti* (strain Liverpool) (AeA), vectorbase.org, OGS AaegL3.3; *Anopheles gambiae* (strain PEST) (AnG), vectorbase.org, OGS AgamP4; *Apis mellifera* (ApM), hymenopteragenome.org/beebase, OGS Amel_4.5; *Bombyx mori* (BoM), Ensembl build 29, GCA_000151625.1.29; *Cimex lectularius* (CiL), i5k.nal.usda.gov, OGS v1.2; *Culex quinquefasciatus* (strain Johannesburg) (CuQ), vectorbase.org, OGS CpipJ2.2; *Daphnia pulex* (DaP), genome.jgi.doe.gov, OGS FrozenGeneCatalog20110204; *Drosophila melanogaster* (DrM), flybase.org, OGS r6.08; *Manduca sexta* (MaS), i5k.nal.usda.gov, OGS v2.0; *Musca domestica* (MuD), vectorbase.org, OGS MdomA1.1; *Pediculus humanus* (PeH) vectorbase.org, OGS PhumU2.1; *Solenopsis invicta* (Sol) hymenopteragenome.org/ant_genomes, OGS 2.2.3 w/ HGD-IDs; and *Tribolium castaneum* (TrC), NCBI WGS, OGS GCF_000002335.2_Tcas_3.0. For each gene set, the longest peptide sequence was selected for each gene model from all available isoforms, removing low-quality and short sequences. The final counts of proteins for each species is indicated in Fig. 1. The OrthoMCL pipeline (v2.0.9) was followed to define orthologous groups of proteins between these peptide sets [202]. Briefly, after formatting the peptide sequence file for each species, an all-by-all BLASTp search was performed between all proteins from all species. The resulting blast hits were loaded into the OrthoMCL schema within a MySQL database. Ortholog groups were calculated using the scripts provided with OrthoMCL and the MCL clustering algorithm [203]. This results in sets of orthologs, co-orthologs, and in-paralogs defined between all peptides from all species. From this, counts of shared proteins between species were calculated and summarized in Fig. 1. To place the species within a phylogenetic context, single copy orthologs were identified between all species using BUSCO [14, 204]. A total of 2591 single copy orthologs were used to generate a multigene alignment. Peptide sequences from each species for each orthologous group were aligned independently using MUSCLE [205], trimmed using trimAl with parameters "-w 3 -gt 0.95 -st 0.01" [206], and trimmed sequences were concatenated using ElConcatenero (https://github.com/ODiogoSilva/ElConcatenero). Phylogenetic analysis in RAxML was performed using the PROTGAMMA amino acid substitution model and 1000 bootstrap replicates [207]. This substitution model was selected due to its use of empirical base frequencies and the LG substitution model which is a general amino-acid replacement matrix that was demonstrated to produce a tree topology with a higher likelihood than trees produced using an alternative amino acid substitution model such as WAG or JTT [208]. This tree was rooted with *D. pulex* and was visualized using Dendroscope 3.2.10 [209]. The resulting tree was used to order the species in Fig. 1.

Manual annotation of specific genome characters and gene families

Transposable elements

DNA transposons The assembled *C. capitata* genome was analyzed for potential DNA transposon sequences using the program RepeatModeler and a custom library of DNA transposon sequences from available publications and databases (http://www.repeatmasker.org). The output of the RepeatModeler program was aligned to custom protein database of DNA transposon sequences using the fasty36 program [210] with e-value cutoff of 0.5 to further identify potential genuine TE sequences [211]. Duplicate entries were identified using the program BLASTclust [212].

LTR elements LTR annotation was both structurally and homology-based. First, a structurally based LTR search was performed by running the LTR_STRUC program on genomic scaffolds [213]. Second, a homology-based annotation of the repeat families, which were generated by running RepeatModeler on the scaffolds was compared to a database of known, RepBase Drosophila LTRs using tblastx searches via the CENSOR program [214].

Non-LTR retrotransposons A modified version of the homology-based TESeeker [215] was used to identify non-LTR retrotransposons. TESeeker was run with representative TEs included with it, as well as those identified by RepeatModeler. TEs were classified with an in-house classifier that uses reverse transcriptase conserved domains to classify based on the open reading frame of the TEs. tBLASTn searches were then performed using the classified TEs to help reconstruct a full-length element.

microRNAs

Mature miRNA sequences along with miRNA precursors were retrieved from miRBase 21 [216]. Precursors were subsequently aligned against the assembled genome using BLASTn, with positive hits presenting e-values $<1e^{-10}$. Mature miRNA sequences from Hexapoda were aligned against the assembly using Bowtie v1 [217], allowing up to one mismatch. The aligned mature miRNAs and precursors were selected and annotated, and for homologous loci, the secondary structure of the region was further investigated using RNAfold [218]. A modified version of Maple module from the ShortStack [46] algorithm was utilized in order to identify characteristic pre-miRNA features such as complementarity at 5' and 3', a 3' overhang, 3' and 5' bulges. Sequences having the most stable structure were selected for each locus.

To identify homologous transcripts in Hexapoda, mature miRNAs presenting identity in seed region and total sequence, allowing 1–2 mismatches in 3' or 5' (excepting seed region) were collapsed into clusters. Clusters with >3 Hexapoda members were marked as presenting high homology in the subphylum. Only miRNAs marked as having experimental support in miRBase were included in order to avoid sequences identified solely by homology studies, and to enhance the robustness of the analysis.

Chemoreceptor genes

For the annotation of the OBP and OR gene families, tBLASTn searches were performed on the genomic scaffold sequences, using *D. melanogaster* and *M. domestica* OBPs and ORs as queries. The putative proteins encoded by the identified gene models that produced hits ($<1e^{-10}$) were used to query, using BLASTx, local protein databases of the *D. melanogaster* and *M. domestica* OBPs and ORs. The gene models were modified, where necessary, in WebApollo. The medfly genes were named using a numerical system, with genes on the same scaffold numbered sequentially. However, for the OBPs, the sequential numbering system was modified to permit sequential naming of the different OBP subfamilies. The medfly genes and their encoded proteins are detailed in Additional file 2: Tables S10 and S11 and putative protein sequences are provided in Additional file 4. Pseudogenes (suffix PSE) were translated as well as possible so that they could be aligned with the other proteins for the phylogenetic analysis. In the case of the OBP protein sequences, the signal peptide sequences were excluded before alignment and the phylogenetic analyses. For each family the amino acid sequences were aligned using MAFFT v7 [219] with the E-INS-i strategy, BLOSUM62 matrix, 1000 maxiterate and offset 0. The most appropriate model of molecular evolution for each dataset was determined using MEGA 6.0.6 [220]. Phylogenetic relationships were estimated using maximum likelihood with 1000 bootstrap replications using MEGA 6.06 retaining positions present in at least 75 % of the sequences. The resulting mid-point rooted tree was drawn using FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) and iDraw (www.indeeo.com).

The GR and IR gene families were manually annotated and analyzed with the aid of maximum likelihood phylogenetic trees. BLASTp searches were performed on the JAMg Consensus Gene Set v1, as well as high-confidence and low-confidence protein sets from NCBI. tBLASTn searches were also performed using all *D. melanogaster* and *M. domestica* relatives as queries. If the gene models appeared to be intact or were easily repaired in the WebApollo tool employed for manual gene annotation, they were manipulated and named therein, but more difficult gene models were manually assembled in TextWrangler before being modified in WebApollo. All of the *Ceratitis* genes and encoded proteins are detailed in Additional file

2: Tables S12 and S13, with protein sequences provided in Additional file 4.

Only a few difficulties with the genome assembly were encountered in these two gene families, such as truncation of exons in gaps between contigs within scaffolds or off ends of scaffolds (suffix NTE in the figures, tables, and proteins). In most cases these gene models were corrected using raw reads (suffix FIX). One IR model was designed that spans scaffolds, with no support other than the agreement of the available exons on both scaffolds, and their appropriate relatedness to similar genes (suffix JOI). Pseudogenes were translated to an encoded protein as a best match for alignment with intact proteins for phylogenetic analysis (suffix PSE). Pseudogene translations included in the analysis were limited to those having at least half the average length of related proteins and several shorter fragments were not included. Protein families were aligned in CLUSTALX v2.0 [221] using default settings with the relevant families of *D. melanogaster*. Problematic gene models and pseudogenes were refined in light of these alignments. Less obvious pseudogenes (e.g. those with small in-frame deletions or insertions, crucial amino acids changes, or promoter defects) would not be recognized, so the provided functional protein totals might be high.

For phylogenetic analysis, the alignments were trimmed using TRIMAL v1.4 [206] retaining only positions present in more than 80 % of the sequences. Phylogenetic analysis was performed using maximum likelihood methods in PHYML v3.0 [222] using default settings, and trees were prepared in FIGTREE v1.4 (http://tree.bio.ed.ac.uk/software/figtree/) and Adobe Illustrator.

Immunity-related genes

D. melanogaster and *An. gambiae* immune-related genes were retrieved from the Insect Innate Immunity Database (sec: http://bordensteinlab.vanderbilt.edu/IIID/test_immunity.php) [94] and aligned against *C. capitata* gene models using Blastp [196]. Medfly genes that showed the best Blast hit against *D. melanogaster* or *An. gambiae* were assumed to be putatively involved in the medfly immune system and were annotated manually.

An additional HMM analysis was performed in order to enrich the medfly immunity gene repertoire. Thirty-four curated multiple sequence alignments of potential immune-related genes from *D. melanogaster*, *An. gambiae*, *Ae. Aegypti*, and *C. quinquefasciatus* were retrieved from ImmunoDB [223] and HMMs were built using HMMER software, version 3.1b1 [224]. These HMMs were used to calculate the likelihood of having any of the 34 domains for each of the *C. capitata* predicted proteins, with calls having an e-value $<10^{-2}$ annotated manually.

Seminal fluid proteins genes

To annotate putative medfly SFP genes, we queried (tBLASTn, e-value $<10^{-10}$) the genome scaffolds using the amino acid sequences of the 146 characterized *D. melanogaster* SFPs [166, 225]. Sixty-four of the *Drosophila* SFPs gave no significant hits to the medfly genome, whereas the remaining sequences resulted in multiple hits. The predicted amino acid sequences of the identified medfly gene models were considered for annotation if they gave significant reciprocal BLASTp (e-value $<10^{-10}$) hits in the NCBI nr database to sequences belonging to known SFP functional classes. In addition, we queried (tBLASTx, e-value $<10^{-10}$) the genome scaffolds using the ESTs previously derived from medfly male testes and male accessory glands [159]. The predicted amino acid sequences of the identified gene models were considered for annotation if they gave significant reciprocal BLASTp (e-value $<10^{-10}$) hits in the NCBI nr database to sequences belonging to known SFP functional classes.

Additional files

Additional file 1: Supplementary material A. *C. capitata* genome sequencing approaches, B. Automated annotations, C. Detection of bacterial sequence contamination and D. Screening the *C. capitata* genome sequence for potential horizontal gene transfer events. (DOCX 601 kb)

Additional file 2: Supplementary Tables S1–S26. **Table S1** *C. capitata* genome and RNA-seq source material and sequencing runs. **Table S2** *Pluralibacter gergoviae* genome metrics. **Table S3** *P. gergoviae* genes associated with general functional categories. **Table S4** BUSCO genome assembly comparisons between *C. capitata*, *D. melanogaster*, and *Bactrocera* species. **Table S5a** Orthology tables - Copy numbers. **Table S5b** Orthology tables - Orthologous groups. **Table S5c** Orthology tables - Counts by species. **Table S6** Chromosomal positions for mapped scaffolds. **Table S7** *C. capitata* transposable element sequences. **Table S8** *C. capitata* microRNA sequences. **Table S9** microRNA/siRNA/piRNA machinery in *C capitata*. **Table S10** *C. capitata* odorant-binding protein (OBP) genes. **Table S11** *C. capitata* odorant receptor (OR) genes. **Table S12** *C. capitata* gustatory receptor (GR) gene assignments. **Table S13** *C. capitata* ionotrophic receptor (IR) gene assignments. **Table S14** *C. capitata* aquaporin genes. **Table S15** Immunity-related gene comparisons for *C. capitata*, *D. melanogaster*, and *M. domestica*. **Table S16** P450 genes in the *C. capitata* genome. **Table S17** Carboxylesterase genes in the *C. capitata* genome. **Table S18** Glutathione S-transferase (GST) genes in the *C. capitata* genome. **Table S19** CysLGIC superfamily genes in *C. capitata* and other insect genomes. **Table S20** *C. capitata* cuticle protein genes. **Table S21** Putative cuticle proteins per family in the *C. capitata* genome. **Table S22** Cuticle protein gene clusters in the *C. capitata* genome. **Table S23** *C. capitata* sex-determination gene orthologs. **Table S24** Putative seminal fluid protein (SFP) genes in the *C. capitata* genome. **Table S25** *C. capitata* genes related to the apoptotic pathway of *D. melanogaster*. **Table S26** Community RNA-Seq data for the genome assembly (XLSX 6240 kb)

Additional file 3: Supplementary figures S1–S9. **Figure S1** Odorant binding protein (OBP) genes phylogenetic tree. **Figure S2** Ionotrophic receptor (IR) genes phylogenetic tree. **Figure S3** Opsin genes phylogenetic tree. **Figure S4** CYP gene clusters A and B. **Figure S5** Glutathione S-transferase genes phylogenetic tree. **Figure S6** cysLGIC genes superfamily phylogenetic tree. **Figure S7** Cuticle protein gene clusters. **Figure S8**

CPLCA cuticle protein genes phylogenetic tree. **Figure S9** Seminal fluid protein functional classes. (PDF 9425 kb)

Additional file 4: Supplementary material: *C. capitata* chemoreceptor genes. (DOCX 194 kb)

Abbreviations

20E: 20-hydroxyecdysone; AMP: Antimicrobial peptides; *ctx*: *Ceratotoxin*; CYP450: Cytochrome P450; cysLGIC: Cys loop ligand gated ion channel; *dsx*: *Doublesex*; EST: Expressed sequence tag; *fru*: *Fruitless*; GNBP: Gram-negative binding protein; GO: Gene Ontology; GR: Gustatory taste receptor; GST: Glutathione S-transferase; HMM: Hidden Markov Model; IAP: Inhibitor of apoptosis; IGR: Insect growth regulator; IPM: Integrated pest management; IR: Ionotropic receptor; JAMg: Just_Annotate_My_genome; JHE: Juvenile hormone esterase; LTR: Long terminal repeat; nAChR: Nicotinic acetylcholine receptor; OBP: Odorant-binding protein; OR: Odorant receptor; PGRP: Peptidoglycan recognition protein; PRR: Pattern recognition receptor; RDC: *Rhino, deadlock, cutoff*; RHG: *Reaper, hid, grim*; RMM: RNA recognition motif; *rpr*: *Reaper*; SFP: Seminal fluid protein; SIT: Sterile insect technique; *ssx*: *Sister-of-Sex-lethal*; *Sxl*: *Sex-lethal*; TE: Transposable element; *tra*: *Transformer*; *tra-2*: *Transformer-2*; *tsl*: *Temperature-sensitive lethal*; TWDL: Tweedle; WGS: Whole genome sequencing

Acknowledgements

Grateful appreciation is extended to Drs. Kenneth Vick and Kevin Hackett (USDA-ARS, National Program Staff) for their support of the initiation of this project, to Drs. A. Mintzas, D. Kritikou, A. Garou Papalexiou, P. Gourzi, E. Stratikopoulos, and A. Augustinos (University of Patras, Greece) for their contributions to the medfly cytogenetic map construction utilized for scaffold and transgene map localizations, and to the FAO/IAEA for support of this effort (to AZ).

Funding

Support of this project was provided by the U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), Animal and Plant Health Inspection Service (APHIS), and National Institute of Food and Agriculture (NIFA)-Biotechnology Risk Assessment Grants Program (grant no. 2011-39211-30769 to AMH) for funding the initial phase of this project, and to the National Institutes of Health (NIH)-National Human Genome Research Institute (NHGRI) for funcing the medfly genome sequencing, assembly, and Maker 2.0 automated annotation as part of the i5K 30 genome pilot project (grant no. U54 HG003273 to RAG). The NIH Intramural Research Program, National Library of Medicine funded the NCBI Gnomon annotation and the USDA-National Agricultural Library (NAL) provided support for the WebApollo curation website. Support was provided for: toxin metabolism and insecticide resistance gene studies from MINECO, Spain (AGL2013-42632-R to FO and PH-C); microRNAs, horizontal gene transfer, and bacterial contaminant studies from the European Social Fund and National Strategic Reference Framework-THALES (MIS375869 to KB, GT, AGH, and KM) and the U.S. National Science Foundation (DEB 1257053 to JHW); cuticle protein gene studies from USDA-NIFA (grant no. 2016-67012-24652 to AJR); sex-determination studies from LR Campania (grant 5/02, 2008 to GS); male reproduction and sexual differentiation studies from the FAO/IAEA (Technical Contract No.: 16966 to GGa) and Cariplo IMPROVE (to FS); and programmed cell death gene studies and genomic data analysis (to MFS) from the Emmy Noether program, DFG (SCHE 1833/1-1) and the LOEWE Center for Insect Biotechnology & Bioresources grant of the Hessen State Ministry of Higher Education, Research and the Arts (HMWK), Germany and from the USDA-NIFA-Biotechnology Risk Assessment Grants Program (grant no. 2015-33522-24094 to AMH).

Availability of data and materials

All genome sequence data are publicly available at the NCBI BioProject: PRJNA168120, and RNA-Seq transcriptome data at BioProject: PRJNA198743, with the genome assembly at NCBI accession number GCA_000347755.1 (see Table 1). Raw sequence data are available at the NCBI SRA site with accession numbers for each library (SRX275786–SRX275788 and SRX276046–SRX276048) and source material, as well as BioProject sites, listed in Additional file 2: Table S1. Annotation and gene model data including a WebApollo browser are available at the USDA-National Agricultural Library i5K Workspace (https://i5knal.usda.gov/Ceratitis_capitata). These and additional genomic resources can also be accessed at the BCM-HGSC sites: https://www.hgsc.bcm.edu/arthropods/mediterranean-fruit-fly-genome-project and ftp://ftp.hgsc.bcm.edu/I5K-pilot/Mediterranean_fruit_fly/.

Authors' contributions

AMH and SR conceived of and directed the project whose management was assisted by MFS, AP, and RAG. LMG reared the inbred ISPRA lines and extracted genomic DNA and RNA that was used for library construction, sequencing, assembly, and Maker annotations at HGSC by SR, SD, SLL, HC, HVD, HD, YH, JQ, SGM, DSTH, KCW, DMM, and RAC. Additional community RNA-Seq analysis was performed by GO, IC, and EAW; JAMg annotations were performed by AP and Gnomon (NCBI) annotations were directed by TDM that were provided for manual annotation at the USDA-NAL WebApollo site managed by MP and CC; AD, JHW, AP, TDM, GT, and KB identified and analyzed bacterial sequence contaminants; SC and SBS performed orthology and phylogeny analysis; AMH, AZ, and PH-C made scaffold map assignments; GT and KB performed symbiont horizontal gene transfer analysis; PS, PA, PWA, and AMH analyzed transposable elements; GGe, DK, MDP, ISV, and AGH manually annotated and analyzed microRNAs; LMG, FS, GS, PS, MM, ARM, and GCa manually annotated and analyzed odorant binding protein, odorant receptor, and seminal fluid protein genes; SDG and HMR manually annotated and analyzed gustatory and ionotrophic receptor genes; MF and JWJ manually annotated and analyzed opsin genes; AJR, AFR, JPC, and JBB manually annotated and analyzed aquaporin and cuticle protein genes; PK, MR, and KM manually annotated and analyzed immunity-related genes; PC, FO, PH-C, MG-G, EU, and AC-A manually annotated and analyzed toxin metabolism and insecticide resistance genes (P450s, carboxylesterases, GSTs, cysLGICs, and sodium channels); MS and GS manually annotated and analyzed sex determination genes; and MFS and AMH manually annotated and analyzed programmed cell death genes. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

[1]Hawkesbury Institute for the Environment, Western Sydney University, Sydney, Australia. [2]Justus-Liebig-University Giessen, Institute for Insect Biotechnology, 35394 Giessen, Germany. [3]Department of Biological Sciences, Cal Poly Pomona, Pomona, CA 91768, USA. [4]Department of Entomology and Center for Disease Vector Research, University of California Riverside, Riverside, CA 92521, USA. [5]Interdepartmental Graduate Program in Genetics, Genomics & Bioinformatics, University of California Riverside, Riverside, CA 92521, USA. [6]Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221, USA. [7]Insect Pest Control Laboratory, Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture, Seibersdorf, Vienna, Austria. [8]Department of Environmental and Natural Resources Management, University of Patras, Agrinio, Greece. [9]Department of Environmental Biology, Centro de Investigaciones Biológicas, CSIC, 28040 Madrid, Spain. [10]Human Genome Sequencing Center, Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [11]National Agricultural Library, USDA, Beltsville, MD 20705, USA. [12]Georg-August-Universität Göttingen, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, 37077 Göttingen, Germany. [13]Department of Biology, University of Rochester, Rochester, NY 14627, USA. [14]Department of Biological Sciences, Wayne State University, Detroit, MI 48202, USA. [15]Department of Biology and Biotechnology, University of Pavia, 27100 Pavia, Italy. [16]USDA-ARS, Pacific Basin Agricultural Research Center, Hilo, HI 96720, USA. [17]DIANA-Lab, Department of Electrical & Computer Engineering, University of Thessaly, 382 21 Volos, Greece and Hellenic Pasteur Institute, 11521 Athens, Greece. [18]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [19]Department of Biological Sciences, Oakland University, Rochester, MI 48309, USA. [20]Department of Biochemistry and Biotechnology, University of Thessaly, Larissa, Greece. [21]Department of Biology, University of Naples Federico II, 80126 Naples, Italy.

[22]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA. [23]Institute of Molecular Biology and Genetics, Biomedical Sciences Research Centre "Alexander Fleming", Vari, Greece. [24]Department of Biology, University of Patras, Patras, Greece. [25]USDA-ARS, Center for Medical, Agricultural, and Veterinary Entomology, 1700 S.W. 23rd Drive, Gainesville, FL 32608, USA.

Received: 11 April 2016 Accepted: 26 August 2016
Published online: 22 September 2016

References

1. Liquido NJ, Shinoda LA, Cunningham RT. Host plants of the Mediterranean fruit fly (Diptera, Tephritidae). An annotated world list. Ann Entomol Soc Am. 1991;77:1–57.
2. Gasperi G, Guglielmino CR, Malacrida AR, Milani R. Genetic variability and gene flow in geographical populations of Ceratitis capitata (Wied.) (medfly). Heredity (Edinb). 1991;67:347–56.
3. Malacrida AR, Guglielmino CR, Gasperi G, Baruffi L, Milani R. Spatial and temporal differentiation in colonizing populations of Ceratitis capitata. Heredity. 1992;69:107–11.
4. Szyniszewska AM, Tatem AJ. Global assessment of seasonal potential distribution of Mediterranean fruit fly, Ceratitis capitata (Diptera: Tephritidae). PLoS One. 2014;9:e111582.
5. Enkerlin WR. Impact of fruit fly control programmes using the sterile insect technique. In: Dyck VA, Hendrichs J, Robinson AS, editors. Sterile insect technique - principles and practice in area-wide integrated pest management. Dordrecht: Springer; 2005. p. 651–76.
6. Dyck VA, Hendrichs J, Robinson AS, editors. Sterile insect technique - principles and practice in area-wide integrated pest management. Dordrecht: Springer; 2005.
7. Zacharopoulou A, Frisardi M, Savakis C, Robinson AS, Tollas P, Konsolaki M, et al. The genome of the Mediterranean fruitfly Ceratitis capitata: localization of molecular markers by in situ hybridization to salivary gland polytene chromosomes. Chromosoma. 1992;101:448–55.
8. Stratikopoulos EE, Augustinos AA, Pavlopoulos ID, Economou K, Mintzas A, Mathiopoulos KD, et al. Isolation and characterization of microsatellite markers from the Mediterranean fruit fly, Ceratitis capitata: cross species amplification in other Tephritidae species reveals a varying degree of transferability. Mol Genet Genomics. 2003;282:283–306.
9. Loukeris TG, Livadaras I, Arca B, Zabalou S, Savakis C. Gene transfer into the medfly, Ceratitis capitata, with a Drosophila hydei transposable element. Science. 1995;270:2002–5.
10. i5K Consortium. The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. J Hered. 2013;104:595–600.
11. Gomulski LM, Torti C, Malacrida AR, Gasperi G, Comari, a full-length mariner element from the Mediterranean fruit fly, Ceratitis capitata. Insect Mol Biol. 1997;6:241–53.
12. Tsoumani KT, Mathiopoulos KD. Genome size estimation with quantitative real-time PCR in two Tephritidae species: Ceratitis capitata and Bactrocera oleae. J Appl Entomol. 2012;136:626–31.
13. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.
14. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
15. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18:188–96.
16. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. JBrowse: a next-generation genome browser. Genome Res. 2009;19:1630–8.
17. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, et al. Web Apollo: a web-based genomic annotation editing platform. Genome Biol. 2013;14:R93.
18. Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, et al. Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics. 2008;179:1601–55.
19. Hierst JL. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. Front Genet. 2015;6:220.
20. Anleitner JE, Haymer DS. Y enriched and Y specific DNA sequences from the genome of the Mediterranean fruit fly, Ceratitis capitata. Chromosoma. 1992;101:271–8.
21. Rosetto M, de Filippis T, Mandrioli M, Zacharopoulou A, Gourz P, Manetti AG, et al. Ceratotoxins: female-specific X-linked genes from the medfly, *Ceratitis capitata*. Genome. 2000;43:707–11.
22. Handler AM, McCombs SD, Fraser MJ, Saul SH. The lepidopteran transposon vector, piggyBac, mediates germ-line transformation in the Mediterranean fruit fly. Proc Natl Acad Sci U S A. 1998;95:7520–5.
23. Tordai H, Nagy A, Farkas K, Banyai L, Patthy L. Modules, multidomain proteins and organismic complexity. FEBS J. 2005;272:5064–78.
24. Fu G, Condon KC, Epton MJ, Gong P, Jin L, Condon GC, et al. Female-specific insect lethality engineered using alternative splicing. Nat Biotechnol. 2007;25:353–7.
25. Schetelig MF, Handler AM. A transgenic embryonic sexing system for Anastrepha suspensa (Diptera: Tephritidae). Insect Biochem Mol Biol. 2012;42:790–5.
26. Schetelig MF, Handler AM. Strategy for enhanced transgenic strain development for embryonic conditional lethality in Anastrepha suspensa. Proc Natl Acad Sci U S A. 2012;109:9348–53.
27. Ogaugwu CE, Schetelig MF, Wimmer EA. Transgenic sexing system for Ceratitis capitata (Diptera: Tephritidae) based on female specific embryonic lethality. Insect Biochem Mol Biol. 2013;43:1–8.
28. Schetelig MF, Caceres C, Zacharopoulou A, Franz G, Wimmer EA. Conditional embryonic lethality to improve the sterile insect technique in Ceratitis capitata (Diptera: Tephritidae). BMC Biol. 2009;7:4.
29. Schetelig MF, Scolari F, Handler AM, Kittelmann S, Gasperi G, Wimmer EA. Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly Ceratitis capitata. Proc Natl Acad Sci U S A. 2009;106:18171–6.
30. Horn C, Handler AM. Site-specific genomic targeting in Drosophila. Proc Natl Acad Sci U S A. 2005;102:12483–8.
31. Dafa'alla TH, Condon GC, Condon KC, Phillips CE, Morrison NI, Jin L, et al. Transposon-free insertions for insect genetic engineering. Nat Biotechnol. 2006;24:820–1.
32. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol. 2005;1:166–75.
33. Green CL, Frommer M. The genome of the Queensland fruit fly Bactrocera tryoni contains multiple representatives of the mariner family of transposable elements. Insect Mol Biol. 2001;10:371–86.
34. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7.
35. Franz G, Loukeris TG, Dialektaki G, Thompson CR, Savakis C. Mobile Minos elements from Drosophila hydei encode a two exon transposase with similarity to the paired DNA-binding domain. Proc Natl Acad Sci U S A. 1994;91:4746–50.
36. Robertson HM. The mariner transposable element is widespread in insects. Nature. 1993;362:241–5.
37. Handler AM, Zimowska GJ, Armstrong KF. Highly similar piggyBac elements in Bactrocera that share a common lineage with elements in noctuid moths. Insect Mol Biol. 2008;17:387–93.
38. Bonizzoni M, Gomulski LM, Malacrida AR, Capy P, Gasperi G. Highly similar piggyBac transposase-like sequences from various Bactrocera (Diptera, Tephritidae) species. Insect Mol Biol. 2007;16:645–50.
39. Warren WD, Atkinson PW, O'Brochta DA. The Hermes transposable element from the house fly, Musca domestica, is a short inverted repeat-type element of the hobo, Ac, and Tam3 (hAT) element family. Genet Res. 1994;64:87–97.
40. Michel K, Stamenova A, Pinkerton AC, Franz G, Robinson AS, Garcia-Papalexiou A, et al. Hermes-mediated germ-line transformation of the Mediterranean fruit fly Ceratitis capitata. Insect Mol Biol. 2001;10:155–62.
41. Handler AM, Gomez SP. The hobo transposable element excises and has related elements in tephritid species. Genetics. 1996;143:1339–47.
42. Torti C, Gomulski LM, Bonizzoni M, Murelli V, Moralli D, Guglielmino CR, et al. Cohobo, a hobo-related sequence in Ceratitis capitata. Genetica. 2005;123:313–25.
43. Lucas K, Raikhel AS. Insect microRNAs: biogenesis, expression profiling and biological functions. Insect Biochem Mol Biol. 2013;43:24–38.

44. Pane A, Salvemini M, Delli Bovi P, Polito C, Saccone G. The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. Development. 2002;129:3715–25.
45. Salvemini M, Robertson M, Aronson B, Atkinson P, Polito LC, Saccone G. Ceratitis capitata transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for female sex determination. Int J Dev Biol. 2009;53:109–20.
46. Axtell MJ. ShortStack: comprehensive annotation and quantification of small RNA genes. RNA. 2013;19:740–51.
47. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol. 2008;9:22–32.
48. Yeom K-H, Lee Y, Han J, Suh MR, Kim VN. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. Nucleic Acids Res. 2006;34:4622–9.
49. Caudy AA, Ketting RF, Hammond SM, Denli AM, Bathoorn AMP, Tops BBJ, et al. A micrococcal nuclease homologue in RNAi effector complexes. Nature. 2003;425:411–4.
50. Tsuchiya N, Nakagama H. MicroRNA, SND1, and alterations in translational regulation in colon carcinogenesis. Mutat Res. 2010;693:94–100.
51. Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev. 2006;20:1885–98.
52. Schneider MD, Najand N, Chaker S, Pare JM, Haskins J, Hughes SC, et al. Gawky is a component of cytoplasmic mRNA processing bodies required for early Drosophila development. J Cell Biol. 2006;174:349–58.
53. Temme C, Simonelig M, Wahle E. Deadenylation of mRNA by the CCR4–NOT complex in Drosophila: molecular and developmental aspects. Front Genet. 2014;5:143.
54. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003;17:3011–6.
55. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011;12:246–58.
56. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell. 2014;157:1364–79.
57. Drew R, Yuval B. The evolution of fruit fly feeding behavior. In: Aluja M, Norrbom A, editors. Fruit flies (Tephritidae): phylogeny and evolution of behavior. Boca Raton: CRC Press; 2000. p. 731–49.
58. Siciliano P, He X-L, Woodcock C, Pickett JA, Field LM, Birkett MA, et al. Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, Ceratitis capitata. Insect Biochem Mol Biol. 2014;48:51–62.
59. Siciliano P, Scolari F, Gomulski LM, Falchetto M, Manni M, Gabrieli P, et al. Sniffing out chemosensory genes from the Mediterranean fruit fly, Ceratitis capitata. PLoS One. 2014;9:e85523.
60. Malacrida AR, Gomulski LM, Bonizzoni M, Bertin S, Gasperi G, Guglielmino CR. Globalization and fruitfly invasion and expansion: the medfly paradigm. Genetica. 2007;131:1–9.
61. Yuval B, Hendrichs J. Behavior of flies in the Genus Ceratitis. In: Aluja M, Norrbom A, editors. Fruit flies (Tephritidae): phylogeny and evolution of behavior. Boca Raton: CRC Press; 2000. p. 429–57.
62. Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol. 2013;58:373–91.
63. Vieira FG, Rozas J. Comparative genomics of the odorant binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biol Evol. 2011;3:476–90.
64. Kwon JY, Dahanukar A, Weiss LA, Carlson JR. The molecular basis of CO2 reception in Drosophila. Proc Natl Acad Sci U S A. 2007;104:3574–8.
65. Rytz R, Croset V, Benton R. Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in Drosophila and beyond. Insect Biochem Mol Biol. 2013;43:888–97.
66. Koh TW, He Z, Gorur-Shandilya S, Menuz K, Larter NK, Stewart S, et al. The Drosophila IR20a clade of ionotropic receptors are candidate taste and pheromone receptors. Neuron. 2014;83:850–65.
67. Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2003;100 Suppl 2:14537–42.
68. Scott JC, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, et al. Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. Genome Biol. 2014;15:466.
69. Gomulski LM, Dimopoulos G, Xi Z, Scolari F, Gabrieli P, Siciliano P, et al. Transcriptome profiling of sexual maturation and mating in the Mediterranean fruit fly, Ceratitis capitata. PLoS One. 2012;7:e30857.
70. Bellmann D, Richardt A, Freyberger R, Nuwal N, Schwärzel M, Fiala A, et al. Optogenetically induced Olfactory Stimulation in *Drosophila* Larvae Reveals the Neuronal Basis of Odor-Aversion Behavior. Front Behav Neurosci. 2010;4:27.
71. Thorne N, Bray S, Amrein H. Function and expression of the Drosophila gr genes in the perception of sweet, bitter and pheromone compounds. Chem Senses. 2005;30 Suppl 1:270–2.
72. Kurtovic A, Widmer A, Dickson BJ. A single class of olfactory neurons mediates behavioural responses to a Drosophila sex pheromone. Nature. 2007;446:542–6.
73. Ronderos DS, Lin CC, Potter CJ, Smith DP. Farnesol-detecting olfactory neurons in Drosophila. J Neurosci. 2014;34:3959–68.
74. Miyamoto T, Slone J, Song X, Amrein H. A fructose receptor functions as a nutrient sensor in the Drosophila brain. Cell. 2012;151:1113–25.
75. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. Cell. 2006;125:143–60.
76. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, et al. Episodic radiations in the fly tree of life. Proc Nat Acad Sci U S A. 2011;108:5690–5.
77. Agee HR, Boller E, Remund U, Davis JC, Chambers DL. Spectral sensitivities and visual attractant studies on the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), olive fly, Dacus oleae (Gmelin), and the European cherry fruit fly, Rhagoletis cerasi (L.) (Diptera, Tephritidae)1. Zeitschrift für Angewandte Entomologie. 1982;93:403–12.
78. Levinson H, Levinson A, Osterried E. Orange-derived stimuli regulating oviposition in the Mediterranean fruit fly. J Appl Entomol. 2003;127:269–75.
79. Prokopy RJ, Economopoulos AP. Color responses of Ceratitis capitata flies. Zeitschrift für Angewandte Entomologie. 1976;80:434–7.
80. Katsoyannos BI, Panagiotidou K, Kechagia I. Effect of color properties on the selection of oviposition site by Ceratitis capitata. Entomol Exp Appl. 1986;42:187–93.
81. Nakagawa S, Prokopy RJ, Wong TTY, Ziegler JR, Mitchell SM, Urago T, et al. Visual orientation of Ceratitis capitata flies to fruit models. Entomol Exp Appl. 1978;24:193–8.
82. Benelli G, Daane KM, Canale A, Niu C-Y, Messing RH, Vargas RI. Sexual communication and related behaviours in Tephritidae: current knowledge and potential applications for Integrated Pest Management. J Pest Sci. 2014;87:385–405.
83. Johnston Jr RJ. Lessons about terminal differentiation from the specification of color-detecting photoreceptors in the Drosophila retina. Ann N Y Acad Sci. 2013;1293:33–44.
84. Henze MJ, Oakley TH. The dynamic evolutionary history of pancrustacean eyes and opsins. Integr Comp Biol. 2015;55:830–42.
85. International Glossina Genome Initiative. Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis. Science. 2014;344:380–6.
86. Benoit JB, Hansen IA, Attardo GM, Michalkova V, Mireji PO, Bargul JL, et al. Aquaporins are critical for provision of water during lactation and intrauterine progeny hydration to maintain tsetse fly reproductive success. PLoS Negl Trop Dis. 2014;8:e2717.
87. Finn RN, Chauvigné F, Stavang JA, Belles X, Cerdà J. Insect glycerol transporters evolved by functional co-option and gene replacement. Nat Commun. 2015;6:7814.
88. Benoit JB, Hansen IA, Szuter EM, Drake LL, Burnett DL, Attardo GM. Emerging roles of aquaporins in relation to the physiology of blood-feeding arthropods. J Comp Physiol B. 2014;184:811–25.
89. Lemaitre B, Hoffmann J. The host defense of Drosophila melanogaster. Annu Rev Immunol. 2007;25:697–743.
90. Boutros M, Agaisse H, Perrimon N. Sequential activation of signaling pathways during innate immune responses in Drosophila. Dev Cell. 2002;3:711–22.
91. Kang D, Liu G, Lundström A, Gelius E, Steiner H. A peptidoglycan recognition protein in innate immunity conserved from insects to humans. Proc Natl Acad Sci. 1998;95:10078–82.
92. Kim Y-S, Ryu J-H, Han S-J, Choi K-H, Nam K-B, Jang I-H, et al. Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and β-1,3-glucan that mediates the signaling for the induction of innate immune genes in Drosophila melanogaster Cells. J Biol Chem. 2000;275:32721–7.

93. Lee WJ, Lee JD, Kravchenko VV, Ulevitch RJ, Brey PT. Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, Bombyx mori. Proc Natl Acad Sci. 1996;93:7888–93.
94. Brucker RM, Funkhouser LJ, Setia S, Pauly R, Bordenstein SR. Insect Innate Immunity Database (IID): an annotation tool for identifying immune genes in insect genomes. PLoS One. 2012;7:e45125.
95. Marchini D, Bernini LF, Marri L, Giordano PC, Dallai R. The female reproductive accessory glands of the medfly Ceratitis capitata: Antibacterial activity of the secretion fluid. Insect Biochem. 1991;21:597–605.
96. Marchini D, Manetti AGO, Rosetto M, Bernini LF, Telford JL, Baldari CT, et al. cDNA sequence and expression of the ceratotoxin gene encoding an antibacterial sex-specific peptide from the medfly Ceratitis capitata (diptera). J Biol Chem. 1995;270:6199–204.
97. Marchini D, Marri L, Rosetto M, Manetti AGO, Dallai R. Presence of antibacterial peptides on the laid egg chorion of the medfly Ceratitis capitata. Biochem Biophys Res Commun. 1997;240:657–63.
98. Marchini D, Rosetto M, Dallai R, Marri L. Bacteria associated with the oesophageal bulb of the medfly Ceratitis capitata (Diptera: Tephritidae). Curr Microbiol. 2002;44:120–4.
99. Rosetto M, Marchini D, de Filippis I, Ciolfi S, Frati F, Quilici S, et al. The *ceratotoxin* gene family in the medfly *Ceratitis capitata* and the Natal fruit fly *Ceratitis rosa* (Diptera: Tephritidae). Heredity (Edinb). 2003;90:382–9.
100. Hoffmann JA. The immune response of Drosophila. Nature. 2003;426:33–8.
101. Valanne S, Wang J-H, Rämet M. The Drosophila Toll Signaling Pathway. J Immunol. 2011;186:649–56.
102. Lemaitre B, Reichhart J-M, Hoffmann JA. Drosophila host defense: Differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. Proc Natl Acad Sci. 1997;94:14614–9.
103. Jang I-H, Chosa N, Kim S-H, Nam H-J, Lemaitre B, Ochiai M, et al. A Spätzle-processing enzyme required for Toll signaling activation in Drosophila innate immunity. Dev Cell. 2006;10:45–55.
104. Kambris Z, Brun S, Jang I-H, Nam H-J, Romeo Y, Takahashi K, et al. Drosophila immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for Toll activation. Curr Biol. 2006;16:808–13.
105. Halfon MS, Hashimoto C, Keshishian H. The Drosophila Toll gene functions zygotically and is necessary for proper motoneuron and muscle development. Dev Biol. 1995;169:151–67.
106. Belvin MP, Anderson KV. A conserved signaling pathway: The Drosophila Toll-Dorsal Pathway. Annu Rev Cell Dev Biol. 1996;12:393–416.
107. Qiu P, Pan PC, Govind S. A role for the Drosophila Toll/Cactus pathway in larval hematopoiesis. Development. 1998;125:1909–20.
108. Vontas J, Hernández-Crespo P, Margaritopoulos JT, Ortego F, Feng H-T, Mathiopoulos KD, et al. Insecticide resistance in Tephritid flies. Pestic Biochem Physiol. 2011;100:199–205.
109. Feyereisen R. Insect CYP genes and P450 enzymes. In: Gilbert LI, editor. Insect Molecular Biology and Biochemistry. Oxford: Elsevier; 2012. p. 236–316.
110. Arouri R, Le Goff G, Hemden H, Navarro-Llopis V, M'Saad M, Castanera P, et al. Resistance to lambda-cyhalothrin in Spanish field populations of Ceratitis capitata and metabolic resistance mediated by P450 in a resistant strain. Pest Manag Sci. 2015;71:1281–91.
111. Ono H, Rewitz KF, Shinoda T, Itoyama K, Petryk A, Rybczynski R, et al. Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. Dev Biol. 2006;298:555–70.
112. Rewitz KF, O'Connor MB, Gilbert LI. Molecular evolution of the insect Halloween family of cytochrome P450s: Phylogeny, gene organization and functional conservation. Insect Biochem Mol Biol. 2007;37:741–53.
113. Sztal T, Chung H, Gramzow L, Daborn PJ, Batterham P, Robin C. Two independent duplications forming the Cyp307a genes in Drosophila. Insect Biochem Mol Biol. 2007;37:1044–53.
114. Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD, Russell RJ. Biochemical genetics and genomics of insect esterases. In: Gilbert LI, Gill SS, editors. Insect pharmacology: channels, receptors, toxins and enzymes. London: Elsevier; 2010. p. 229–301.
115. Ellis LL, Carney GE. Mating alters gene expression patterns in Drosophila melanogaster male heads. BMC Genomics. 2010;11:558.
116. Robin GCD, Claudianos C, Russell RJ, Oakeshott JG. Reconstructing the diversification of alpha-esterases: Comparing the gene clusters of Drosophila buzzatii and D-melanogaster. J Mol Evol. 2000;51:149–60.
117. Claudianos C, Russell RJ, Oakeshott JG. The same amino acid substitution in orthologous esterases confers organophosphosohate resistance on the house fly and a blowfly. Insect Biochem Mol Biol. 1999;29:675–86.
118. Magana C, Hernandez-Crespo P, Brun-Barale A, Couso-Ferrer F, Bride J-M, Castanera P, et al. Mechanisms of resistance to malathion in the medfly Ceratitis capitata. Insect Biochem Mol Biol. 2008;38:756–62.
119. Crone EJ, Sutherland TD, Campbell PM, Coppin CW, Russell RJ, Oakeshott JG. Only one esterase of Drosophila melanogaster is likely to degrade juvenile hormone in vivo. Insect Biochem Mol Biol. 2007;37:540–9.
120. Guillen Y, Rius N, Delprat A, Williford A, Muyas F, Puig M, et al. Genomics of ecological adaptation in Cactophilic Drosophila. Genome Biol Evol. 2015;7:349–66.
121. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Mol Biol. 2005;14:3–8.
122. Li X, Schuler MA, Berenbaum MR. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu Rev Entomol. 2007;52:231–53.
123. Thompson AJ, Lester HA, Lummis SCR. The structural basis of function in Cys-loop receptors. Q Rev Biophys. 2010;43:449–99.
124. Jones AK, Bera AN, Lees K, Sattelle DB. The cys-loop ligand-gated ion channel gene superfamily of the parasitoid wasp, Nasonia vitripennis. Heredity. 2010;104:247–59.
125. Couso Ferrer F. Bases moleculares de la resistencia a insecticidas en la mosca mediterránea de la fruta Ceratitis Capitata (Wiedemann). Madrid: E.T.S.I. Agrónomos (UPM), Biotecnología; 2012.
126. Dong K, Du Y, Rinkevich F, Nomura Y, Xu P, Wang L, et al. Molecular biology of insect sodium channels and pyrethroid resistance. Insect Biochem Mol Biol. 2014;50:1–17.
127. Derst C, Walther C, Veh RW, Wicher D, Heinemann SH. Four novel sequences in Drosophila melanogaster homologous to the auxiliary Para sodium channel subunit TipE. Biochem Biophys Res Commun. 2006;339:939–48.
128. Wang L, Nomura Y, Du Y, Dong K. Differential effects of TipE and a TipE-homologous protein on modulation of gating properties of sodium channels from Drosophila melanogaster. Plos One. 2013;8:e67551.
129. Li J, Waterhouse RM, Zdobnov EM. A remarkably stable TipE gene cluster: evolution of insect Para sodium channel auxiliary subunits. BMC Evol Biol. 2011;11:337.
130. Willis JH. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. Insect Biochem Mol Biol. 2010;40:189–204.
131. Ioannidou ZS, Theodoropoulou MC, Papandreou NC, Willis JH, Hamodrakas SJ. CuticleProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile hidden Markov models. Insect Biochem Mol Biol. 2014;52:51–9.
132. Benoit JB, Adelman ZN, Reinhardt K, Dolan A, Poelchau M, Jennings EC, et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat Commun. 2016;7:10165.
133. Comman RS. Molecular evolution of Drosophila cuticular protein genes. PLoS One. 2009;4:e8345.
134. Togawa T, Dunn WA, Emmons AC, Nagao J, Willis JH. Developmental expression patterns of cuticular protein genes with the R&R Consensus from *Anopheles gambiae*. Insect Biochem Mol Biol. 2008;38:508–19.
135. Comman RS, Togawa T, Dunn WA, He N, Emmons AC, Willis JH. Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae*. BMC Genomics. 2008;9:22.
136. Willis J, Papandreou N, Iconomidou VA, Hamodrakas SJ. Cuticular proteins. In: Insect Molecular Biology and Biochemistry. San Diego: Academic Press; 2012. p. 134–66.
137. Salvemini M, D'Amato R, Petrella V, Ippolito D, Ventre G, Zhang Y, et al. Subtractive and differential hybridization molecular analyses of Ceratitis capitata XX/XY versus XX embryos to search for male-specific early transcribed genes. BMC Genet. 2014;15 Suppl 2:S5.
138. Nagaraju J, Saccone G. How is sex determined in insects? An epilogue. J Genet. 2010;89:389–90.
139. Saccone G, Salvemini M, Pane A, Polito LC. Masculinization of XX Drosophila transgenic flies expressing the Ceratitis capitata DoublesexM isoform. Int J Dev Biol. 2008;52:1051–7.
140. Bopp D, Saccone G, Beye M. Sex determination in insects: variations on a common theme. Sex Dev. 2014;8:20–8.
141. Saccone G, Peluso I, Artiaco D, Giordano E, Bopp D, Polito LC. The Ceratitis capitata homologue of the Drosophila sex-determining gene sex-lethal is structurally conserved, but not sex-specifically regulated. Development. 1998;125:1495–500.

142. Saccone G, Louis C, Zhang H, Petrella V, Di Natale M, Perri M, Salvemini M. Male-specific phosphorylated SR proteins in adult flies of the Mediterranean fruitfly Ceratitis capitata. BMC Genet. 2014;15 Suppl 2:S6.
143. Graham P, Penn JK, Schedl P. Masters change, slaves remain. Bioessays. 2003;25:1–4.
144. Gomulski LM, Dimopoulos G, Xi Z, Soares MB, Bonaldo MF, Malacrida AR, et al. Gene discovery in an invasive tephritid model pest species, the Mediterranean fruit fly, Ceratitis capitata. BMC Genomics. 2008;9:243.
145. Gabrieli P, Falaguerra A, Siciliano P, Gomulski LM, Scolari F, Zacharopoulou A, et al. Sex and the single embryo: early development in the Mediterranean fruit fly, Ceratitis capitata. BMC Dev Biol. 2010;10:12.
146. Schmieder S, Colinet D, Poirié M. Tracing back the nascence of a new sex determination pathway to the ancestor of bees and ants. Nat Commun. 2012;3:895.
147. Hall AB, Basu S, Jiang X, Qi Y, Timoshevsky VA, Biedler JK, et al. Sex determination. A male-determining factor in the mosquito Aedes aegypti. Science. 2015;348:1268–70.
148. Handler AM. Prospects for using genetic transformation for improved SIT and new biocontrol methods. Genetica. 2002;116:137–49.
149. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. Insect seminal fluid proteins: identification and function. Annu Rev Entomol. 2011;56:21–40.
150. Wolfner MF. "S.P.E.R.M." (seminal proteins (are) essential reproductive modulators): the view from Drosophila. Soc Reprod Fertil Suppl. 2007;65:183–99.
151. Poiani A. Complexity of seminal fluid: a review. Behav Ecol Sociobiol. 2006;60:289–310.
152. Perry JC, Sirot L, Wigby S. The seminal symphony: how to compose an ejaculate. Trends Ecol Evol. 2013;28:414–22.
153. Dhole S, Servedio MR, Lloyd EA. Sperm competition and the evolution of seminal fluid composition. Evolution. 2014;68:3008–19.
154. Kelleher ES, Watts TD, LaFlamme BA, Haynes PA, Markow TA. Proteomic analysis of *Drosophila mojavensis* male accessory glands suggests novel classes of seminal fluid proteins. Insect Biochem Mol Biol. 2009;39:366–71.
155. Mueller JL, Ripoll DR, Aquadro CF, Wolfner MF. Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. Proc Natl Acad Sci U S A. 2004;101:13542–7.
156. Begun DJ, Lindfors HA, Thompson ME, Holloway AK. Recently evolved genes identified from Drosophila yakuba and D. erecta accessory gland expressed sequence tags. Genetics. 2006;172:1675–81.
157. Miyatake T, Chapman T, Partridge L. Mating-induced inhibition of remating in female Mediterranean fruit flies *Ceratitis capitata*. J Insect Physiol. 1999;45:1021–8.
158. Davies SJ, Chapman T. Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (*Ceratitis capitata*). Insect Biochem Mol Biol. 2006;36:846–56.
159. Scolari F, Gomulski LM, Ribeiro JM, Siciliano P, Meraldi A, Falchetto M, et al. Transcriptional profiles of mating-responsive genes from testes and male accessory glands of the Mediterranean fruit fly, Ceratitis capitata. PLoS One. 2012;7:e46812.
160. Jang E. Effects of mating and accessory-gland injections on olfactory-mediated behavior in the female Mediterranean Fruit-Fly, *Ceratitis-Capitata*. J Insect Physiol. 1995;41:705–10.
161. Bonizzoni M, Gomulski LM, Bertin S, Scolari F, Guglielmino CR, Yuval B, et al. Unfaithful Mediterranean fruit fly Ceratitis capitata females: Impact on the SIT? In: Vreysen MJB, Robinson AS, Hendrichs J, editors. Area Wide Control of Insect Pests. Netherlands: Springer; 2007. p. 175–82.
162. Bonizzoni M, Katsoyannos BI, Marguerie R, Guglielmino CR, Gasperi G, Malacrida A, et al. Microsatellite analysis reveals remating by wild Mediterranean fruit fly females, Ceratitis capitata. Mol Ecol. 2002;11:1915–21.
163. Gavriel S, Gazit Y, Yuval B. Remating by female Mediterranean fruit flies (Ceratitis capitata, Diptera: Tephritidae): temporal patterns and modulation by male condition. J Insect Physiol. 2009;55:637–42.
164. Kraaijeveld K, Katsoyannos BI, Stavrinides M, Kouloussis NA, Chapman T. Remating in wild females of the Mediterranean fruit fly, Ceratitis capitata. Anim Behav. 2005;69:771–6.
165. Scolari F, Yuval B, Gomulski LM, Schetelig MF, Gabrieli P, Bassetti F, et al. Polyandry in the medfly - shifts in paternity mediated by sperm stratification and mixing. BMC Genet. 2014;15 Suppl 2:S10.
166. Findlay GD, Yi X, Maccoss MJ, Swanson WJ. Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS Biol. 2008;6:e178.
167. Sirot LK, Hardstone MC, Helinski ME, Ribeiro JM, Kimura M, Deewatthanawong P, et al. Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. PLoS Negl Trop Dis. 2011;5:e989.
168. LaFlamme BA, Ram KR, Wolfner MF. The Drosophila melanogaster seminal fluid protease "seminase" regulates proteolytic and post-mating reproductive processes. PLoS Genet. 2012;8:e1002435.
169. Heifetz Y, Vandenberg LN, Cohn HI, Wolfner MF. Two cleavage products of the *Drosophila* accessory gland protein ovulin can independently induce ovulation. Proc Natl Acad Sci U S A. 2005;102:743–8.
170. Prokupek AM, Eyun SI, Ko L, Moriyama EN, Harshman LG. Molecular evolutionary analysis of seminal receptacle sperm storage organ genes of Drosophila melanogaster. J Evol Biol. 2010;23:1386–98.
171. Fehlbaum P, Bulet P, Chernysh S, Briand JP, Roussel JP, Letellier L, et al. Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides. Proc Natl Acad Sci U S A. 1996;93:1221–5.
172. Dottorini T, Persampieri T, Palladino P, Baker DA, Spaccapelo R, Senin N, et al. Regulation of Anopheles gambiae male accessory gland genes influences postmating response in female. FASEB J. 2013;27:86–97.
173. LaFlamme BA, Wolfner MF. Identification and function of proteolysis regulators in seminal fluid. Mol Reprod Dev. 2013;80:80–101.
174. Zhao Y, Sun W, Zhang P, Chi H, Zhang MJ, Song CQ, et al. Nematode sperm maturation triggered by protease involves sperm-secreted serine protease inhibitor (Serpin). Proc Natl Acad Sci U S A. 2012;109:1542–7.
175. Allen AK, Spradling AC. The Sf1-related nuclear hormone receptor Hr39 regulates *Drosophila* female reproductive tract development and function. Development. 2008;135:311–21.
176. Chapman T. The soup in my fly: evolution, form and function of seminal fluid proteins. PLoS Biol. 2008;6:e179.
177. Yamamoto MT, Takemori N. Proteome profiling reveals tissue-specific protein expression in the male reproductive system of *Drosophila melanogaster*. Fly (Austin). 2010;4:36–9.
178. Ban L, Napolitano E, Serra A, Zhou X, Iovinella I, Pelosi P. Identification of pheromone-like compounds in male reproductive organs of the oriental locust Locusta migratoria. Biochem Biophys Res Commun. 2013;437:620–4.
179. Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, Harrington LC. Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. Insect Biochem Mol Biol. 2008;38:168–80.
180. South A, Sirot LK, Lewis SM. Identification of predicted seminal fluid proteins in *Tribolium castaneum*. Insect Mol Biol. 2011;20:447–56.
181. Zhou S, Stone EA, Mackay TF, Anholt RR. Plasticity of the chemoreceptor repertoire in *Drosophila melanogaster*. PLoS Genet. 2009;5:e1000681.
182. Scolari F, Benoit JB, Michalkova V, Aksoy E, Takac P, Abd-Alla AM, et al. The spermatophore in Glossina morsitans morsitans: insights into male contributions to reproduction. Sci Rep. 2016;6:20334.
183. Avila FW, Wong A, Sitnik JL, Wolfner MF. Don't pull the plug! the Drosophila mating plug preserves fertility. Fly (Austin). 2015;9(2):62–7.
184. Russo CL, Spurr-Michaud S, Tisdale A, Pudney J, Anderson D, Gipson IK. Muc1 gene expression in human male urogenital tract epithelia. Hum Reprod. 2006;21:2783–93.
185. Eriksen GV, Carstedt I, Uldbjerg N, Ernst E. Cervical mucins affect the motility of human spermatozoa in vitro. Fertil Steril. 1998;70:350–4.
186. Baldini F, Gabrieli P, Rogers DW, Catteruccia F. Function and composition of male accessory gland secretions in Anopheles gambiae: a comparison with other insect vectors of infectious diseases. Pathog Glob Health. 2012;106:82–93.
187. Le BV, Klocx C, Schatz A, Nguyen JB, Kakani EG, Catteruccia F, et al. Dihydroisoxazole inhibitors of Anopheles gambiae seminal transglutaminase AgTG3. Malar J. 2014;13:270.
188. Lee CY, Bachrecke EH. Genetic regulation of programmed cell death in Drosophila. Cell Res. 2000;10:193–204.
189. Zhou L, Schnitzler A, Agapite J, Schwartz LM, Steller H, Nambu JR. Cooperative functions of the reaper and head involution defective genes in the programmed cell death of Drosophila central nervous system midline cells. Proc Natl Acad Sci U S A. 1997;94:5131–6.
190. Schetelig MF, Nirmala X, Handler AM. Pro-apoptotic cell death genes, hid and reaper, from the tephritid pest species, Anastrepha suspensa. Apoptosis. 2011;16:759–68.
191. Zhou L. The 'unique key' feature of the lap-binding motifs in RHG proteins. Cell Death Differ. 2006;12:1148–51.

192. Nirmala X, Schetelig MF, Zimowska GJ, Zhou L, Handler AM. Pro-apoptotic gene regulation and its activation by gamma-irradiation in the Caribbean fruit fly, *Anastrepha suspensa*. Apoptosis. 2015;20:1–9.
193. Gariou-Papalexiou A, Gouzi P, Delprat A, Kritikou D, Rapti K, Chrysanthakopoulou B, et al. Polytene chromosomes as tools in the genetic analysis of the Mediterranean fruit fly, Ceratitis capitata. Genetica. 2002;116:59–71.
194. Handler AM. Enhancing the stability and ecological safety of mass-reared transgenic strains for field release by redundant conditional lethality systems. Insect Sci. 2016;23:225–34.
195. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108:1513–8.
196. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
197. Zamani N, Sundstrom G, Meadows JR, Hoppner MP, Dainat J, Lantz H, et al. A universal genomic coordinate translator for comparative genomics. BMC Bioinformatics. 2014;15:227.
198. Haas B., Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et a.. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
199. Haas B., Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31:5654–66.
200. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34:W435–9.
201. Haas B., Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.
202. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.
203. Van Dongen S. Graph clustering by flow simulation. Utrecht: University of Utrecht; 2000.
204. Benoit JB, Adelman ZN, Reinhardt K, Dolan A, Poelchau M, Jennings EC, et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat Commun. 2016;7:10165. http://doi.org/10.1038/ncomms10165.
205. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:1–19.
206. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
207. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
208. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25:1307–20.
209. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012;61:1061–7.
210. Pearson W. Finding protein and nucleotide similarities with FASTA. Curr Protoc Bioinformatics. 2004;Chapter 3:Unit3 9.
211. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A. 1988;85:2444–8.
212. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
213. McCarthy EM, McDonald JF. LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics. 2003;19:362–7.
214. Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR--a program for identification and elimination of repetitive elements from DNA sequences. Comput Chem. 1996;20:119–21.
215. Kennedy RC, Unger MT, Christley S, Collins FH, Madey GR. An automated homology-based approach for identifying transposable elements. BMC Bioinformatics. 2011;12:130.
216. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39:D152–7.
217. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.
218. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008;36:W70–4.
219. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
220. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.
221. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.
222. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
223. Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science. 2007;316:1738–43.
224. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.
225. Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Res. 2009;19:886–96.
226. Fabrick JA, Pei J, Hull JJ, Yool AJ. Molecular and functional characterization of multiple aquaporin water channel proteins from the western tarnished plant bug, *Lygus hesperus*. Insect Biochem Mol Biol. 2014;45:125–40.

Original Article

A draft genome sequence of an invasive mosquito: an Italian *Aedes albopictus*

Vicky Dritsou*[1], Pantelis Topalis*[2], Nikolai Windbichler[3], Alekos Simoni[3], Ann Hall[3], Daniel Lawson[4], Malcolm Hinsley[4], Daniel Hughes[4], Valerio Napolioni[1], Francesca Crucianelli[1], Elena Deligianni[2], Giuliano Gasperi[5], Ludvik M. Gomulski[5], Grazia Savini[5], Mosè Manni[5], Francesca Scolari[5], Anna R. Malacrida[5], Bruno Arcà[6], José M. Ribeiro[7], Fabrizio Lombardo[6], Giuseppe Saccone[8], Marco Salvemini[8], Riccardo Moretti[9], Giuseppe Aprea[9], Maurizio Calvitti[9], Matteo Picciolini[1], Philippos Aris Papathanos[3,10], Roberta Spaccapelo[10], Guido Favia[11], Andrea Crisanti[1,3,10], Christos Louis[2,10]

[1]Polo d'Innovazione di Genomica, Genetica e Biologia (Polo GGB), Loc. S. Andrea delle Fratte, Perugia, Italy, [2]Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology Hellas, Heraklion, Crete, Greece, [3]Imperial College London, Department of Life Sciences, South Kensington Campus, UK, [4]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genomes Campus, Cambridge, Cambridgeshire, UK, [5]Department of Biology and Biotechnology, University of Pavia, Italy, [6]Department of Public Health and Infectious Diseases – Division of Parasitology, Sapienza University of Rome, Italy, [7]National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA, [8]Department of Biology, University of Naples Federico II, Italy, [9]ENEA – Italian National Agency for New Technologies, Energy and Sustainable Economic Development, CR Casaccia, Rome, Italy, [10]Department of Experimental Medicine, University of Perugia, Italy, [11]Scuola di Bioscienze e Medicina Veterinaria, University of Camerino, Italy

The draft genome sequence of Italian specimens of the Asian tiger mosquito *Aedes (Stegomyia) albopictus* (Diptera: Culicidae) was determined using a standard NGS (next generation sequencing) approach. The size of the assembled genome is comparable to that of *Aedes aegypti*, the two mosquitoes are also similar as far as the high content of repetitive DNA is concerned, most of which is made up of transposable elements. Although, based on BUSCO (Benchmarking Universal Single-Copy Orthologues) analysis, the genome assembly reported here contains more than 99% of protein-coding genes, several of those are expected to be represented in the assembly in a fragmented state. We also present here the annotation of several families of genes (tRNA genes, miRNA genes, the sialome, genes involved in chromatin condensation, sex determination genes, odorant binding proteins and odorant receptors). These analyses confirm that the assembly can be used for the study of the biology of this invasive vector of disease.

Keywords: NGS, WGS, BUSCO, Repetitive DNA, Transposable elements, Invasive species, Disease vector, Dengue fever, Chikungunya

Introduction

Aedes albopictus, the Asian tiger mosquito native in Southeast Asia, depicts a very aggressive host seeking and biting behaviour that causes a high degree of nuisance and, most significantly, it also acts as a vector for several important diseases such as arboviral (e.g. dengue virus, West Nile virus, chikungunya virus) and other parasitic (e.g. dirofilaria) infections.[1] Ever since its presence was recorded in Albania[2] and Italy[3] in the mid 1970s and late 1980s, respectively, there have been increased voices asking for intensified monitoring in order to prevent a potential re-emergence of dengue fever in Europe. Indeed, a major dengue fever epidemic in Greece occurred about 90 years ago: two back-to-back dengue fever epidemics, transmitted by *Aedes aegypti* (*Stegomyia fasciata*), resulted in almost one million cases and, officially, 1553 deaths mostly in Athens.[4–6] Since then, vectors of dengue fever were only occasionally reported in Europe. *A. aegypti* for example, which

Correspondence to: Christos Louis, IMBB-FORTH, N. Plastira 100, Vassilika Vouton, GR-700 13 Heraklion, Crete, Greece. Email: louis@imbb.forth.gr

* Should be considered as equal first authors

DOI 10.1179/2047773215Y.0000000031

was first recorded in Macedonia[7,8] and was then found throughout Greece,[9] was never again collected there, even in systematic searches, following the DDT-based malaria eradication campaign that took place in the 1950s.[10] In spite of the absence of dengue fever's classical vector, the recent continent-wide invasion of the Asian tiger mosquito[11] has prominently included this disease in the agenda of European control agencies.

The spread of *A. albopictus* in Europe, and in particular in Italy, seems to have originated from the Americas, whose invasion is assumed to have occurred less than a decade earlier.[12,13] Most, if not all movements of this species across continents are assumed to be the result of worldwide trade, especially, that of used car tyres.[14] By now, wild Asian tiger mosquito specimens have been isolated in a large number of countries in all continents with the exception of the Antarctic (see[15] for a map of its current global distribution).

Although European disease control agencies feared the imminent re-emergence of dengue fever, it was, first, a small epidemic caused by a different arbovirus, chikungunya virus (CHIKV) that was attributed to transmission by *A. albopictus*. In the summer of 2007, a few hundred people near the coast of Emilia-Romagna, Italy, were infected with the CHIKV virus (reviewed by Tomasello and Schlagenhauf[16]). This was followed, indeed, by several isolated cases of dengue fever in different locations in Europe, and an epidemic outbreak of the disease did then occur in Madeira.[17] It should be noted, of course, that although it belongs to Portugal, Madeira is located about 900 km southwest of continental Europe.

The present situation is certainly not dramatic from the public health point of view, at least not in terms of the actual cases of diseases transmitted by *A. albopictus*. It is clear, though, that to keep it from becoming worse, measures have to be taken, ideally in the sense of a potential eradication of *A. albopictus* from Europe the soonest possible. This implies a boosted investment in the research on the biology of the Asian tiger mosquito, especially in its European environment.[18] There has been an increased emphasis in the study of 'European' tiger mosquito 'strains', mostly referring to its population biology and ecology.[15] As members of the European Infravec Consortium,[19] we decided to shift gears and to determine the sequence of specimens of *A. albopictus* isolated from the wild in Italy. This would provide an additional advantage, namely the possibility to compare its 'post-invasion' genome to the genome of Asian specimens once this is publicly available (Xiaoguang Chen, Anthony A. James, *et al.*, personal communication) as well as the genome of a strain isolated in the South Indian Ocean island of La Réunion.[20]

Although the time frame of the establishment of the species in Italy is extremely short in an evolutionary scale (probably <30 years), it is a fact that the Asian tiger mosquitoes in this country have gone through one, and possibly two major bottlenecks. These bottlenecks occurred during their probable initial transfer from East Asia to the United States, followed a few years later by the second transfer to Italy. Furthermore, the mosquitoes may have been crossed to insects that arrived in Italy from other geographic locations such as, for example, Albania. This could have happened both in the field and, possibly, even in the laboratory, during the process of establishing the strain *Fellini*, which was utilised as the source of biological material in the study reported here.[21]

We report here the determination of the whole genome sequence of the *Fellini* strain of *A. albopictus*, as well as its preliminary characterisation. We discuss the assembly obtained, present several examples of annotated genetic features (gene families and repetitive elements) and also discuss the reasons for the difficulties in obtaining a higher quality assembly.

Materials and Methods

Biological material and preparation of genomic DNA

The strain Fellini (a.k.a. Rimini) was used for the determination of the genome sequence. This was derived from 500 eggs collected in 2004 with ovitraps in an urban environment from the northern Italian city of Rimini (44°03′24″ N 12°33′52″ E, Romeo Bellini, personal communication). At generation 40, 10 000 eggs were transferred to Imperial College. Multiple rounds of isofemale selection to reduce heterozygosity of the colony were carried out. At generation 73 about 3800 larvae (stage L4) were collected and total genomic DNA was extracted in batches of 15–25 using either the Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions, or by Phenol:Chloroform:Isoamyl alcohol (25:24:1, Sigma-Aldrich, St. Louis, MO, USA) and purified by EtOH. The extracted genomic DNA (8465 µg) was resuspended in 10 mM Tris-1 mM EDTA buffer.

Determination of genome size

The genome size of the Fellini strain was estimated at 0.94 Gb using a real-time PCR-based method.[22] A series of ten-fold dilutions of known concentration of a linearised plasmid containing the amplicon sequence corresponding to part of the *A. albopictus* G6PDH gene identified from RNA-seq data (Genbank/EMBL Accession Nr KT279821) was monitored in real-time using a MiniOpticon (Bio-Rad). As the length of the amplicon was known, the concentrations of the dilutions could be

calculated as copies per microlitre. Comparison of the calibration curve of the C_T values derived from the standard dilution series with those derived from the same G6PDH amplicon in genomic *A. albopictus* DNA samples of known concentration, permitted the absolute number of target gene copies to be calculated, and hence the genome size.

Library preparation and next generation sequencing

Mate-Pair (MP) and Paired-End (PE) libraries were prepared using Nextera Mate-Pair and Nextera DNA Sample Preparation kits (Illumina, San Diego, CA, USA), respectively, starting from one microgram of gDNA for MP and 50 ng of gDNA for the PE libraries; the manufacturer's instructions were used. A total of 14 libraries were generated (seven for each type). Mate-Pair and PE libraries were then separately pooled and clustered on eight different lanes of one HiSeq PE Flow Cell v3 (Illumina, San Diego, CA, USA). Library quality control was performed on a 2100 Bioanalyzer using a High Sensitivity DNA chip (Agilent Technologies). The libraries were quantified using Qubit dsDNA BR Assay (Life Technologies) and by qPCR using the Illumina library quantification kit (Kapa Biosystem). The sequencing was performed on Illumina's HiSeq1500 with a 2×93 cycling. The whole sequencing run generated 252 Gb (94% $\geq$Q30). The paired-end library had a $60\times$ coverage and an insert size equal to 400 bp.

To determine the average insert size of the MP library to be used for scaffolding, we mapped all 438 867 831 reads included in the contigs of the Mi and So assemblies (see *Assembly*) with a size of 10 Kb or larger (256 and 218 contigs, respectively) using Seqmap[23] allowing no mismatches. We then identified the respective mate pairs that could be mapped within the same contig and calculated the distance between the two reads.

Small RNA sequencing

The protocol described by Friedländer *et al.*[24] was followed. Essentially, total RNA was prepared from 30 whole 4- to 5-day-old mosquitoes using TRIzol (Invitrogen) according to manufacturer's protocol. Males, sugar-fed females and blood-fed females (1 day post feeding) were sampled. Three independent biological replicates for sugar fed and two for blood-fed mosquitoes were used for the small RNA library preparation. Strand-specific cDNA libraries with different barcodes (six base index) were generated using a TruSeq Small RNA kit (Illumina). Small RNA library was validated with 2100 Bioanalyser (Agilent) and sequenced with MiSeq (illumina). Adaptor and index sequences were extracted from the Miseq raw data.

Assembly

For the assembly of the reads, we used a 64-core cluster with a total of 512 Gb of RAM. After having tested a number of software/assemblers (e.g. Velvet,[25] ABySS[26]), we decided to proceed with SOAPdenovo,[27] which produced the most qualitative result. Due to hardware limitations, we divided the PE reads into two groups named Mi and So, each one containing half of the initial reads (i.e. each one containing ~625 million reads or a $30\times$ genome coverage each). We then assembled them individually and subsequently used the MP library for the scaffolding procedure. Following this, the two assemblies were merged, a procedure that included the removal of redundant contigs as well as all contigs that were shorter than 260 bp (see Results and Discussion section). The assembly was concluded by building scaffolds using, again, SOAPdenovo. Assemby 1.0 (see Results and Discussion) was submitted to the NCBI with the BioProject ID number PRJNA289460.

Given the relatively poor statistics of the assembly obtained, we assessed its overall inclusiveness/completeness by performing a BUSCO analysis.[28]

Annotation

The automatic annotation of the genome was performed using the MAKER pipeline.[29] In an attempt to provide more evidence to the pipeline for the annotation process, we have used, within MAKER, evidence from publicly available *A. aegypti* ESTs as well as from RNA-seq data of male and female antennae (Fellini strain, Gomulski, Gasperi *et al.*, manuscript in preparation), and female antennae and palps, whole females and male heads isolated from a strain of *A. albopictus* established in 2012 from wild mosquitoes collected in Rome (Arcà *et al.*, manuscript in preparation). In addition to the automatic annotation, a group of collaborators working on different research areas are currently working towards the manual annotation of the genome. A summary analysis of some of these data is reported here in order to provide an assessment of the quality of the genome assembly. Further details will be published in a series of manuscripts that will follow. For the manual annotation of the following gene families the specifics are as follows.

Manual annotation of specific genes

Masking and identification of repetitive sequence elements

To mask the repeats present in the assembly, we used the MAKER pipeline[29] due to several advantages that this offers: First, it consists of two different masking pieces of software, RepeatMasker and RepeatRunner that both accept and use external libraries of repeated elements. In our case, we used the version 20.0.1 of RepBase plus the default RepeatRunner library,[30]

in addition to the *A. aegypti* repeated elements reported in VectorBase [https://www.vectorbase.org/download-info/aedes-aegypti-liverpoolrepeatslib]. Second, the MAKER pipeline can take advantage of computers with multiple cores rendering the analysis more rapid and, third, it processes each scaffold/contig separately allowing the whole process to be resumed in case of involuntary interruptions during the run.

tRNA genes

To identify tRNA genes, we used the tRNA-SE[31] software on the previously masked sequences. Again, this software can be used from within the MAKER pipeline for the reasons stated above.

miRNA-encoding genes

For the detection of reads representing miRNAs, miRDeep2 was used using default parameters.[24] The miRDeep2 core algorithm requires as inputs two reference files: one of mature and one of precursors miRNAs of related organisms in addition to a reference file of mature miRNAs of similar organisms. For the related species, mature and precursor miRNAs sequences from the *A. albopictus* cell line C710 were used[31] while for the similar species, we used all the mature miRNA sequences of metazoan found in miRBase v.21 (http://www.mirbase.org)

Genes involved in chromatin condensation

Various genes involved in chromatin compaction[32] have been selected based on their differential expression associated to the presence of the endosymbiont *Wolbachia*[33] while trying to identify factors of paternal origin, which contribute to the Cytoplasmic Incompatibility (CI) phenomenon.[34] Specifically, three RNA-seq samples from *A. albopictus* testes belonging to three lines of *A. albopictus* characterised by a different *Wolbachia* infection pattern (A: wild-type infection, i.e. *w*AlbA and *w*AlbB *Wolbachia* strains; 0: no infection, obtained by antibiotic treatment; P: artificial infection with *w*Pip *Wolbachia* strain, obtained by transinfection)[35] have been used for a differential gene expression analysis. Reads were aligned to the *A. albopictus* transcriptome[36] using TopHat v2.0.13,[37 39] with default options except for the following: read mismatch set to 3. Cuffdiff v2.2.1[40,41] was then used to identify differentially expressed genes. Cufflinks was run, followed by Cuffmerge and Cuffdiff, in all cases with default options. Significance was scored at a cut-off false discovery rate (FDR) of 0.05. After that, genes known to be involved in chromatin condensation and showing a significantly different expression related to *Wolbachia* infection pattern were mapped to the *A. albopictus* genome with the TBLASTN tool v2.2.28+,[42,43] with the following non default option: max intron length set to 5000, estimated according to Nene *et al.*[44] In addition, a set of 10 key genes from *Drosophila melanogaster* (infected by *w*Mel *Wolbachia* strain) known to be implicated in chromatin regulation were also tentatively mapped to the *A. albopictus* genome to ascertain the presence of orthologues.

Odorant binding protein and Odorant receptor genes

TBLASTN searches were performed using as queries protein sequences of the 111 Odorant binding proteins (OBPs) and 112 Odorant receptors (ORs) of *A. aegypti*.[45 47] The contigs and scaffolds that produced hits ($e < 1E\text{-}10$) were used to interrogate, using BLASTX, local protein databases of the *Ae aegypti* OBPs and ORs. Multiple contigs and scaffolds with hits to an OBP or OR were assembled using CAP3.[48] GeneWise[49] was used to obtain gene model predictions based on homology with *A. aegypti* OBP and OR proteins.

Sex-determination genes

A TBLASTN search of the present assembly was performed using, as queries, the protein sequences of 14 putative sex determining genes of *A. aegypti*[44] and the *Ceratitis capitata* TRA protein sequence (GenBank acc. num.: AF434936.1[50]). For the *A. aegypti* genes, we utilised the longest *protein sequence, in case* of multiple isoforms for a given gene, except for the *dsx* and *fru* orthologues for which we utilised the full protein sequence of each sex-specific isoforms.

Salivary gland genes

Salivary proteins/cDNAs previously identified in *A. albopictus*[51] and *A. aegypti*[52] were mapped to the genome using the BLAT tool,[53] and compared to other databases using BLASTX or RPS BLAST.[54] Results are displayed in a hyperlinked spreadsheet as previously described.[55] Manual annotation of salivary genes was done with the tool Artemis.[56]

Results and Discussion

Assembly

We obtained sequencing reads on an Illumina's HiSeq1500 automatic sequencer to a final coverage of about 60-fold from a paired-end library and about 60-fold from a Mate Pair library. We faced a clear failure of most assembly software used to achieve an assembly that would have an 'acceptable' N50 value for contigs. The most obvious explanation for this is the high number of repetitive sequences present in the genome of the Asian tiger mosquito[57] combined with the short length of the NGS sequence reads. After repeated attempts, we decided to settle on an assembly that was based on the merger of two sub-assemblies (Mi and So), each performed with one half of the 'short' paired end reads obtained that yielded the best results, and then try to improve it by further 'cleaning', i.e. removal of duplicate contigs (see Table 1 for assembly statistics).

We proceeded with the elimination of singletons with a length <100 nt (~52% of the total). Then all remaining contigs were aligned to each other using the nucmer application of the MUMmer package.[58] The contigs of Mi and So were then divided into four categories: (i) Contigs identical between Mi and So, (ii) Contigs in So fully contained within Mi contigs (identity 99%), (iii) Contigs in So which contained a Mi contig (identity 99%) and finally (iv) contigs that showed no or only partial overlap. We then removed from the assembly all Mi contigs that were identical to So contigs as well all Mi contigs that were contained within So contigs, a total of 3 152 748. Similarly, we also removed from the assembly 2 445 970 So contigs contained in Mi contigs, making sure that we did not remove contigs identical to previously removed Mi contigs from the first two steps. We therefore verified this and re-added those to the assembly (1 169 778 contigs). At the end 5 326 714 contigs remained, which were then used as input to SOAPdeNovo for scaffolding.

For scaffolding, the average insert size determined for the MP library was used. We calculated the distance between all the mate pairs we identified within contigs longer than 10 Kb and we determined an average distance between them of 3150 bp with a median value 2930 bp. Based on that, we used an insert size of 3000 bp for the scaffolding of the contigs to produce the pre-final assembly that we called MiSo 1.0. After analysing the MiSo 1.0 assembly for the presence of repetitive DNA segments, we finally removed from the assembly all contigs that had a length between 100 and 260 bp (see the BUSCO analysis, below), and finally called this assembly MiSo 1.01, which is the assembly that we used for annotation purposes. The relevant overall statistics of the assembly are shown in Table 1.

At the time of writing this paper, we repeated the sizing of the MP library, this time using all contigs from the MiSo assembly that had a size of 3.5 Kb or larger (3120 contigs) and this led to an estimated average insert size of 2032 bp with a mean value of 2186 bp.

Evaluation of the fragmentation and completeness of the assembly

Given the computed low contig N50 number obtained for our assembly, we used an alternative approach for the evaluation of its completeness. This consisted of a BUSCO analysis, which is based on the determination of the presence, in a given assembled genome of a set of single copy orthologues.[28] In our case, we used BUSCOs from both *A. gambiae* and *D. melanogaster*, the closest species to *A. albopictus* for which such sets were available. The results are shown in Table 2.

The fruit fly based analysis yielded results that were very similar to those obtained with the mosquito orthologues (less than 5% divergence) verifying the solidity of the methodology. The overall analysis with 2685 *A. gambiae* BUSCOS showed that 2662 (99.14%) of them were found in our assembly while 23 were missing. Of these positive BUSCOs, 1747 (or 65.63% of all present in the assembly) were found to produce single hits, while 925 (the remaining 34.37%) yielded multiple hits. As BUSCOS are ideally expected to produce single hits, the multiple-hit BUSCOs can only represent genes that are all present in the assembly, though exhibiting varying degrees of fragmentation.

Concerning the BUSCOs that yield no hit, one explanation is that they are missing from the assembly, but another acceptable explanation is that a missing BUSCO does not represent *a priori*, a problem in the assembly. For example, one *A. gambiae* BUSCO is also missing from the complete, version 6 assembly of the *D. melanogaster* genome, while in the opposite analysis, 21 fruit fly BUSCOs (0.68%) are missing from the *A. gambiae* assembled genome (not shown) that is based on a much more complete assembly than the one reported here. In addition to a given BUSCO being absent from the assembly, other reasons for such missing BUSCOs may be an extreme fragmentation of the gene, a gene that has undergone substantial sequence changes in evolution such that, although present in the genome, it no longer fulfils the criteria set for a positive result in the analysis; finally, a gene can always be missing even from an evolutionarily close relative. This latter explanation can be best exemplified by the fact that although <1% of the *A. gambiae* BUSCO orthologues are absent from the assembly, the percentage of missing BUSCOs goes up to 2.6 when the fruit fly orthologues are analysed (not shown). Finally, as stated, around one-third of all BUSCOs found in the assembled Asian tiger mosquito genome are found in more than one contig/scaffold, as a result of fragmentation of the genes in question. This fragmentation is

Table 1 The table shows the summary statistics of the assemblies described in this paper

	Size with N, in bp	Size without N, in bp	No of scaffolds + contigs	No of scaffolds	No of singletons	Scaffolds N50	Contigs N50
MiSoClean (Ass. 1.0)	2 432 868 255	1 589 519 495	4 901 513	189 306	4 712 207	1105	341
MiSoClean with contigs >260 (Ass. 1.01)	1 965 518 921	1 122 114 752	2 141 557	189 141	1 952 416	3255	516

Table 2 The table shows the summary statistics of the BUSCO analysis described in this paper

	No hit DMELA	Single hits	Multiple hits	%	No hit AGAMB	Single hits	Multiple hits	%
Drosophila melanogaster	–				1	3	2681	99.96
A. Gambiae	21	2922	129	99.32	–			
All scaffs and contigs	80	1988	1004	97.4	23	1747	915	99.14
Only scaffolds	881	1867	324	71.32	653	1707	325	75.68
All contigs only	410	1703	959	86.66	232	1540	913	91.36
All contigs ≥ 2000 bp	2650	372	50	13.74	2299	347	39	14.38
260 ≤ contigs < 2000	525	1634	913	82.91	318	1486	881	88.16
Contigs ≥ 260	438	1711	923	85.74	246	1555	884	90.84

DMELA and AGAMB refer to the *D. melanogaster* and *A. gambiae* BUSCOS, respectively. The first two rows show the number of hits detected when using the two sets of BUSCOS on the current assemblies of the fruit fly and the *A. gambiae*, while the next rows show the respective analysis performed on the *A. albopictus* assembly.

certainly smaller for the BUSCOs that are present in larger contigs and scaffolds (10–15%).

We also performed the BUSCO analysis on subsets of the assembly, namely on scaffolds alone, on all contigs, and on all contigs longer that 2000 bp and those with a size ranging from 260 to 2000 bp. These data are also shown in Table 2. Here, two interesting observations are worth mentioning. First, the higher frequency at which BUSCOs are found in large contigs, not unexpectedly, is $\sim 10\times$ higher than the total average and, second, the rate of single to multiple hits is also highest for large contigs followed by BUSCOs found in scaffolds.

The results of the BUSCO analysis, taken together, show that the assembled *A. albopictus* genome can be used for the isolation of most protein-coding genes that are to be studied, although due to fragmentation, this may require some additional steps such as PCR-based isolation and additional sequencing steps.

Finally, the BUSCO analysis provides another significant piece of information. With only one exception, all *A. gambiae* BUSCOs present in the assembled genome are found in contigs and scaffolds that are longer than 260 bp, strongly suggesting that the high number of short contigs present in the assembly are comprised of sequences that do not code for proteins. Based on that and additional pieces of evidence such as the repeat masking and further computations (not shown) we decided, as briefly mentioned earlier, to include in the 'final' assembly only contigs that are longer than 260 bp as well as all scaffolds that were created.

We also performed a masking of repeats present in the assembly; the results are described in Table 3. We have separated the total assembly in four categories, namely scaffolds, contigs between 261 and 2000 bp in length and contigs longer than 2,000 bp (the longest contig in the assembly has a length of 39 933 bp) as well as, naturally, the total assembly.

Assembly 1.01 has a total length of ~ 1.12 Gb (before scaffolding) and ~ 1.97 after scaffolding using an average of 3,000 bp for the length of the insert of the MP library. The former number is relatively close to genome size determined by quantitative PCR (i.e. 0.94 Gb), while the latter is clearly much higher. Obviously, using the recalculated average distance of the Mate Pairs, the length would become approximately 0.2 Gb shorter. The masking resulted in $\sim 34.12\%$ of the total length being tagged as representing repetitive sequences. Although the percentage of masked sequences is only 3% higher when the shorter contigs are examined, we have found that the percentage increases dramatically for the shortest contigs with a length of <260 bp (not shown); this reaffirms our decision to exclude them from the assembly 1.01 (see above). We stress, of course, that the masking of the repetitive segments of the assembly is certainly not complete. The software used as well as the libraries used by the software can certainly not be comprehensive, thus several such repetitive elements may have remained undetected. This will naturally affect the size of the computed genome size that will also be affected by all contigs still present in the assembly that are not unique.

Annotation

Repetitive sequence segments

One of the problematic features of the project anticipated since its beginning was the large number of repeated sequences present in the genome of *A. albopictus.* The fact that older publications reported extremely varying sizes of the genome for wild specimens collected at different places and times made it probable that this was due to the massive expansion or loss of non-coding DNA, very possibly representing transposable elements.[57,59] Since the differences of DNA content could between several populations be the result of the invasion process, we decided to address the issue in more detail. We approached the matter using the data obtained by the repeat masking of the assembly. A total of 27 501 repetitive segments were identified in the assembly that were non-identical; of those 797 were not further recognised through the databases used for the masking. These numbers represent a minimum of elements since if two or more copies of a

Table 3 Detailed statistics performed of the *A. albopictus* assembly, performed on fractions or the total assembly

	Fellini scaffolds	Fellini small contigs (260–000 bps)	Fellini large contigs (> 2000 bps 2001–39 933)	Total
Number #	189 306	1 946 237	6030	2 141 573
Total bases	1 151 063 905	798 233 382	16 259 831	1 965 557 118
Scaffolding Ns	843 348 760	0	0	843 348 760
% Scaffolding Ns	73.27	0	0	42.90
Remaining bases	307 715 145	798 233 382	16 259 831	1 122 208 358
Masked bases	101 286 315	276 390 922	5 164 936	382 842 173
% Masked bases	32.92	34.63	31.77	34.12
Unique bases	206 428 830	521 842 460	11 094 895	739 366 185
% unique bases (Scaffolding Ns not included)	67.08	65.37	68.23	65.88
% Unique bases (Scaffolding Ns included)	17.93	65.37	68.23	37.61

The numbers indicate bp except for the cases in which percentages are shown.

given sequence are present in the assembly, these will be counted as 1. The remaining hits contained several sequences derived from transposable elements known from other species, but also of sequences representing satellite DNA (97× or 5.4% of the assembly) and simple repeats (12 496× or 1.12% of the assembly). Table 4 shows the percentage of the assembly taken up by different classes/families. Interestingly, 22 contigs in the assembly were found to contain rDNA sequences, a family of genes that is known to be extremely difficult to be assembled correctly in WGS (Whole Genome Sequence) projects.[60] The most abundant elements of the 'Other' class of Table 4 includes (but is not limited to) the fruit fly elements *17.6*, *S*, *Doc*, *Bari1*, *piwi*, *F*, *TART* and *Asterix* as well as the rice gypsy-type retrotransposon; these are listed in decreasing order of frequency identified, which, in no case, was higher that 0.2%.

A recent publication reported the assembly of the 'repeatome' of *A. albopictus*.[20] In this case, the strain sequenced was from the island of La Réunion in the Indian Ocean, where the Asian tiger mosquito has also recently been established. In addition to the high number of repeats, a direct comparison is rendered difficult by the fact that the two studies used different software utilising different algorithms and databases for the identification of repeated sequences. Still, in the 12 families of repeated sequences that can be directly compared between the two strains, one can only distinguish a few differences that could be considered to be significant (see Table 4). With the possible exception of the LINE/LOA family whose members are more frequently encountered in the assembly of the strain from La Réunion all other differences in the frequency of transposable elements are not significant enough to pinpoint to a noteworthy variance. Finally, although the frequency of the satellite sequences in the two assemblies is clearly different (6.5 times higher in the Italian strain), we believe that this only represents a dissimilarity that has 'technical' reasons (i.e. software used) and does not reflect real biological differences.

Automatic annotation pipeline

The *ab initio* analysis led to the 'identification of 719 636 ORFs as well as a number of "gene models"'. Given the relatively low N50 value for the assembly as well as the previously concluded BUSCO analysis, we were not able to determine a somewhat more accurate value for the number of 'genes' present in the *A. albopictus* genome. To remedy this, we checked all sequences longer than 50 bps reported in the gff output (6 154 839 out of a total 77 066 684) and subjected them to both BLASTN and TBLASTX analysis querying the set of transcripts of *A. aegypti* obtained from Vectorbase. The results are shown in Table 5. The nucleotide searches indicate that the annotated genome contains 238 898 unique high scoring pairs (HSP) that are similar to 14 581 different *A. aegypti* transcripts with an identity that is higher than 80%. This value represents ~77.34% of the total of *A. aegypti* genes. The value obtained from the TBLASTX analysis is ~99.87% (18 815 with 14 822 to produce high scoring pairs longer than 50 bps). The fact that we miss approximately 0.13% of the genes in our assembly, based on the automatic annotation, could be interpreted as being due to a variety of reasons. The most obvious one, with a potentially biological significance, would be a divergence of the sequences due to changes that occurred throughout evolution, but technical reasons such as extreme fragmentation could also be blamed. Nevertheless, the values obtained at this level of analysis also suggest that our assembly is useful in identifying genes of *A. albopictus*. Significantly, this result also points to the high conservation between the two Aedes species on the level of the primary structure of the proteins encoded by the two genomes.

Annotation of specific genes: a potpourri of examples

tRNA genes

The number of tRNA genes in a given genome varies tremendously from organism to organism. This is true not only when comparing species that are distantly related to each other but it is also the case

Table 4 Percentage of the assembly taken up by different classes/families of repeated sequences

Family	Average size of family members	% of assembly	'French' data
Satellite	306	5.40	0.83%
LINE/RTE-BovB	251	3.25	3.41%
LINE/R1	256	1.45	1.26%
Simple_repeat	53	1.12	
LTR/Gypsy	200	1.00	
LTR/Pao	262	0.94	
LINE/LOA	231	0.94	1.93%
DNA	136	0.86	
DNA/Sola	141	0.54	0.25%
LTR/Copia	199	0.49	
LINE/CR1	222	0.49	0.14%
LINE/I	160	0.40	0.57%
SINE/tRNA	141	0.38	0.23%
DNA/TcMar?	167	0.21	
DNA/CMC-Chapaev	121	0.20	0.33%
LINE/RTE	158	0.19	
DNA/CMC-EnSpm	117	0.19	0.41%
LINE/Penelope	130	0.18	
LINE/L2	213	0.17	
DNA/hAT-hATm	144	0.16	0.11%
SINE/tRNA-I	115	0.14	0.13%
DNA/CMC-Chapaev-3	144	0.14	
DNA/CMC-Transib	156	0.14	
LINE/Jockey	221	0.11	
DNA/Zator	104	0.10	
LINE/Dong-R4	280	0.10	
DNA/Kolobok-Hydra	124	0.09	
Unknown	116	0.08	
RC/Helitron	161	0.07	
Low_complexity	53	0.07	
DNA/TcMar-Tc1	103	0.07	
LINE/L1	171	0.07	
DNA/Crypton	131	0.06	
DNA/TcMar-Fot1	98	0.05	
DNA?/Crypton?	104	0.04	
Unspecified	122	0.04	
DNA/MULE-MuDR	131	0.04	
LINE/L1-Tx1	222	0.03	
DNA/hAT-Charlie	77	0.03	
tRNA	66	0.03	
DNA/hAT-Tip100	107	0.02	
DNA/hAT-Ac	102	0.01	
DNA/hAT	101	0.01	
DNA/PiggyBac	103	0.01	
DNA/hAT-hATx	88	0.01	
LTR/ERV1	81	0.01	
DNA/P	103	0.00	
DNA/Maverick	68	0.00	
DNA/PIF-Harbinger	106	0.00	
DNA/TcMar-Tigger	83	0.00	
DNA/Dada	119	0.00	
DNA/hAT-Blackjack	88	0.00	
Other	80	13.99	

for species that are part of the same genus or even represent sibling species as was shown in comparisons of the genomes of species in the genera *Drosophila*[61] and *Anopheles*.[57] In *A. Aegypti*, a total of 906 tRNA genes, including pseudogenes, have been reported for its annotated genome;[62] the corresponding number is indicated as 984 in VectorBase (see Table 6). In the present analysis, whose results are also shown in Table 6, we identified 5852 tRNA genes, including 4094 pseudogenes. This extremely high number of genes includes 422 genes that encode tRNA-Gly, 432 tRNA-Ala and 276 tRNA-Met genes. Significantly, in these three cases, 88.6, 76.9, and 73.9%, respectively, are found in small contigs with a length between 260 and 2000 bp. Since our overall data indicated that this part of the assembly contains a higher number of repetitive sequences and protein-coding genes are found at a lower frequency, we would like to speculate that these three classes of tRNA genes may be associated to specific families of repetitive sequences, which were not masked, for example transposable elements not present in the libraries used for masking. We should note that a disproportionately high number of tRNA-Ala genes are also found in the genome of *A. aegypti*.[62] Given the draft assembly, gene numbers

Table 5 Results of querying the *ab initio* gene predictions of the assembly using the set of transcripts of *A. aegypti* with BLASTN and TBLASTX

Identity% and alignments length (bps)	No. of genes similar to *A. aegypti* based on blastn	% (Blastn)	No. of genes similar to *A. aegypti* based on tlastx	% (tblastx)
90% and alignment length >90 bps	9005	47.80	9033	47.95
90% and alignment length >80 bps	9324	49.49	9930	52.71
90% and alignment length >70 bps	10075	53.48	10905	57.88
90% and alignment length >60 bps	10311	54.73	11874	63.03
90% and alignment length >50 bps	10496	55.71	12749	67.67
90% and alignment length >40 bps	10680	56.69	13456	71.42
90% and any alignment length	10913	57.92	18545	98.43
80% and alignment length >90 bps	13542	71.88	11091	58.87
80% and alignment length >80 bps	13754	73.00	12032	63.86
80% and alignment length >70 bps	14296	75.88	13022	69.12
80% and alignment length >60 bps	14394	76.40	13999	74.30
80% and alignment length >50 bps	14443	76.66	14822	78.67
80% and alignment length >40 bps	14497	76.95	15423	81.86
80% and any alignment length	14571	77.34	18815	99.87
70% and alignment length >90 bps	13844	73.48	12080	64.12
70% and alignment length >80 bps	14040	74.52	13072	69.38
70% and alignment length >70 bps	14574	77.36	14062	74.64
70% and alignment length >60 bps	14664	77.83	14998	79.61
70% and alignment length >50 bps	14705	78.05	15711	83.39
70% and alignment length >40 bps	14751	78.30	16249	86.25
70% and any alignment length	14821	78.67	18839	99.99
60% and alignment length >90 bps	13846	73.49	12699	67.40
60% and alignment length >80 bps	14042	74.53	13646	72.43
60% and alignment length >70 bps	14576	77.37	14582	77.40
60% and alignment length >60 bps	14666	77.85	15486	82.20
60% and alignment length >50 bps	14707	78.06	16132	85.63
60% and alignment length >40 bps	14753	78.31	16595	88.08
60% and any alignment length	14823	78.68	18840	100.00
50% and alignment length >90 bps	13846	73.49	13018	69.10
50% and alignment length >80 bps	14042	74.53	13938	73.98
50% and alignment length >70 bps	14576	77.37	14898	79.08
50% and alignment length >60 bps	14666	77.85	15733	83.51
50% and alignment length >50 bps	14707	78.06	16327	86.66
50% and alignment length >40 bps	14753	78.31	16768	89.00
50% and any alignment length	14823	78.68	18840	100.00

are likely overestimates and will change as the assembly is improved.

Odorant binding protein and Odorant receptor genes

A. albopictus, like most other mosquito species, relies on olfactory cues for host-seeking, blood-feeding and oviposition site detection. Insect olfaction is the result of a signal transduction cascade involving odorant binding proteins (OBPs) and odorant receptors (ORs), amongst others. The water-soluble OBPs bind odorant molecules that enter the pores of the sensilla and transport them through the lymph to activate the membrane-bound ORs.[63] In *A. aegypti*, 111 OBPs and 112 ORs have been identified.[45–47] Previous studies have identified a small number of OBPs[64 66] and only one OR[67] in *A. albopictus*.

Based on searches using the *A. aegypti* OBP and OR gene families, orthologues were identified for 110 and 98 *A. aegypti* OBPs and ORs, respectively (Tables S1 and S2). It was not possible to determine the actual number of OBPs and ORs present in the genome due to the extensive fragmentation of many of the genes. Fifty-four of the *A. albopictus* OBP sequences were complete, although the sequence often spanned more than one contig. Partial sequences were obtained for 56 OBP orthologues, for 25 of which, it was not possible to identify the region corresponding to the signal peptide, perhaps due to the higher variability of these regions compared to the rest of the gene.

Thirty-five of the OR sequences were complete (although two contained frame-shifts), again frequently spanning several different contigs; six of the complete ORs were identified in single contigs, whereas 12 were identified in single scaffolds. Twenty-one of the incomplete OR sequences lacked one or both termini and encoded conceptual polypeptides between 305 and 411 amino acids in length. Orthologues of 14 *A. aegypti* ORs (OR3, 5, 14, 16, 17, 28, 65, 96, 98, 103, 107, 114, 124 and 127) appear to have been lost in *A. albopictus*, however, a number of gene lineage expansions are clearly evident, for example, AaegOR60, AaegOR61 and AaegOR63 have at least two homologues in *A. albopictus*, whereas AaegOR70 has at least three.

Further analyses, currently in progress including transcriptome data, will help complete the *A. albopictus* OBP and OR repertoires and will open

Table 6 tRNA genes identified in the assembly

Genes	Scaffolds	Small contigs (260-2000 bps)	Large contigs (2001–9993)	Total	*A. aegypti* tRNA genes in VectorBase
Total tRNAs	1172	4616	64	5852	984
Pseudo tRNAs	856	3202	36	4094	0
Remaining tRNAs	316	1414	28	1758	984
Gly	78	640	4	722	47
Val	11	34	2	47	48
Ala	90	332	10	432	327
Leu	8	35	0	43	58
Ile	7	8	2	17	5
Cys	3	4	0	7	14
Met	66	204	6	276	35
Asp	5	20	4	29	54
Asn	1	2	0	3	23
Glu	6	8	0	14	49
Gln	5	12	0	17	25
His	0	2	0	2	37
Arg	1	10	0	11	42
Lys	0	4	0	4	54
Phe	1	0	0	1	19
Tyr	1	0	0	1	28
Trp	1	4	0	5	14
Ser	13	26	0	39	38
Thr	12	32	0	44	24
Pro	3	6	0	9	38
SeC	1	5	0	6	1
Undetermined	3	20	0	23	4
Sup-TTA	0	6	0	6	0
Sup-CTA	0	0	0	0	0

the way for the functional characterisation of these proteins with the aim of developing novel and effective attractants and repellents for monitoring and control.

Genes involved in chromatin condensation

Cytoplasmic Incompatibility is a phenomenon of conditional sterility observed in various insect species, which is caused by the symbiotic bacterium *Wolbachia*; CI is initiated in an embryo when male and female pronuclei originated from parents that exhibited a different *Wolbachia* infection pattern. Specifically, if females are uninfected or infected by a non-compatible *Wolbachia* strain, kariogamy fails and an embryo does not develop. CI can be exploited to produce functionally sterile males to be released against wild target populations as suppressive tool.[33] In addition, it has also been suggested as a method to spread genes of interest or useful physiological features into wild populations (as an example, resistance to pathogens).[68] Therefore, understanding the genetic basis of this phenomenon could have a profound impact on the exploitation of *Wolbachia* for genetic control of insect pests in both agriculture and medicine. *Wolbachia* is not present in mature sperm but it is known that male pronuclei are indirectly affected by its presence during spermatogenesis.[69] Although various studies have aimed at determining all factors implicated in CI,[70–72] no model that can fully describe the phenomenon has yet been developed. What is certainly known is that the embryonic death resulting from CI crosses is a consequence of improper paternal chromatin condensation leading to an abnormal migration of paternal chromosomes. This epigenetic modification has to be induced by *Wolbachia* before spermatid stage, when these bacteria are completely removed. Thus, we selected and analysed a series of *A. aegypti* genes known to be involved in chromatin condensation and whose orthologues were found to be differently expressed when *Wolbachia* is present or not in *A. albopictus* testes (Moretti *et al.*, unpublished data).

Based on a TBLASTN analysis, summarised in Table S3, *A. aegypti* orthologues mostly showed a significant sequence conservation in the genome assembly. The amino acid sequence identity referred to the single BLAST hits peaked in some cases at 100%. However, orthologues are generally scattered among different scaffolds/contigs (ranging 6–185). The best results regard two isoforms of the nucleosome assembly protein gene (*Nap1*) that are located in seven and six scaffolds/contigs, respectively, of the Fellini genome assembly. In the above cases, single scaffolds account for 63 and 99% of query coverage respectively with 32% of sequence identity in the former and 90% in the latter. In addition, regarding the first case, a contig was found to host a hit showing 95% similarity and representing 58% of the query coverage. Two chromatin accessibility complex (CHRAC) gene isoforms are also represented in a relatively low number of scaffolds/contigs of the genome. The analysis of the

first gave seven hits, among which the best scoring showed 62% of query coverage and 97% of max identity. Eight hits characterise the second CHRAC gene isoform with the best hit showing just 24% query coverage and 87% maximum identity.

The analysis of a second set of genes from *D. melanogaster*, known to be involved in chromatin regulation, also gave interesting results even though fewer hits were obtained generally with lower levels of query coverage and sequence identity (Table S4). The histone-binding protein Caf1 is significantly conserved in *A. albopictus*. A hit, out of 10, is located on a scaffold and showed 52% of query coverage and 97% of max sequence identity. An orthologue of the heterochromatin protein (HP1) was also found with 63% query coverage and 39% max identity. Again, the Nap1 protein mapped to the assembly, although with lower similarity compared to the two isoforms from *A. aegypti* (53% query coverage and 45% max identity referring to the best scored hit). Finally, three isoforms of a HIRA protein homologue were located in the assembly. The expression of this gene is already known to be down-regulated by *Wolbachia*, possibly determining some of the cytological effects characterising the CI phenomenon.[70,71] Based on the genes/proteins analysed in this work, future research will aim at identifying the main factors determining the sperm modifications responsible for the CI.

Sex-determination genes

As is the case in other culicine mosquitoes, *A. albopictus* lacks heteromorphic sex chromosomes.[73,74] In this species the primary signal of sex determination seems to be a dominant male determining factor linked to the chromosome 1, according to *A. aegypti–A. albopictus* linkage maps.[75,76] We initiated the search for the *A. albopictus* genes involved in sex determination using the fruit fly homologues as starting blocks.

The results of this study are shown in Table S5. A TBLASTN analysis revealed sequence conservation between *Drosophila* orthologues and *A. albopictus* putative genes in the genome assembly, amino acid sequence identity ranging from 44 to 98%. Only two genes (the homologues of *dpn* and *vir*) are located in a single scaffold/contig of the present genome assembly. The other orthologues are distributed among scattered scaffolds/contigs. However, for each gene we identified, although scattered, all the *A. aegypti* orthologous exons. The name, length and ORF strand of the scaffolds/contigs containing each of the *A. albopictus* orthologues, as well as links to *Drosophila* and *A. aegypti* orthologue annotations (Flybase/Ensembl) and scaffolds/contigs statistics, are reported in Table S5. Interestingly, we have found that the male-specific and the female-specific *dsx* isoforms, similarly to *A. aegypti*, are conserved in *A. albopictus* (80 and 90%, respectively); moreover, *A. aegypti* and *A. albopictus* possess FRU orthologues sharing 90% identity. As previously described for *A. aegypti* and other mosquitoes, a *tra* orthologue seems to be absent in the *A. albopictus* genome, leaving open the question on the molecular nature of the upstream splicing regulator/regulators of *dsx* and *fru* genes in this mosquito species. The *dsx* and *fru* genomic regions of *A. albopictus* will provide a basis for an in-depth comparative search of these and other potential *cis* elements eventually conserved in *Aedes* mosquitoes. Finally, a very recent report[77] showed that the *A. aegypti nix* gene encoding a putative splicing factor related to *Drosophila* tra-2 seems to correspond to a Male determining factor only expressed in males. Aedes nix controls the *dsx* and *fru* male-specific splicing and interestingly it seems to be conserved in *A. albopictus* as a male-specific gene.

It is obviously conceivable that sex-specifically expressed genes such as *dsx* and *fru*, could be utilised to devise and apply biotechnological approaches for vector population suppression, proposed for other mosquito species.[78–80] Their usage in approaches such as SIT and similar techniques in which sex separation is involved obviously jumps to mind.[81]

Salivary gland genes

Mosquito saliva plays an important physiological role in blood feeding through its anti-platelet, anti-clotting and vasodilatory activities, however it also affects pathogen transmission by virtue of its immuno-modulatory and anti-inflammatory properties.[82,83] Salivary genes of blood feeding arthropods, perhaps also as a consequence of the host immune pressure, exhibit an accelerated evolutionary rate and this explains, at least in part, both the elevated diversification observed when comparing the salivary repertoires of anopheline to culicine mosquitoes and the presence of a substantial number of genus-specific salivary genes and gene families.[84,85]

A total of 68 putative secreted products, mostly found in adult female saliva or expressed in the salivary glands of both sexes, were identified in a previous *A. albopictus* sialotranscriptome analysis.[51] We selected 58 putatively secreted salivary proteins and searched the *A. albopictus* genome to proceed to manual annotation and obtain corresponding gene models. Genomic sequences encoding 35 full-length salivary proteins, and including among others serine protease inhibitors, D7 and Antigen five family members, 34 kDa proteins, a salivary lysozyme and putative antimicrobials of the HHH family, were retrieved. For other 19 proteins only partial sequences could be obtained, whereas a short D7 family member (gi 56417443), two putative 13 kDa proteins (gi 56417419, gi 56417421) and a 7.6 kDa protein (gi 56417477) were found to be most likely alleles or alternative splicing products

(not shown). The availability of the complete *A. albopictus* genome also allowed searching for orthologues of *A. aegypti* salivary proteins[52] not previously identified in the tiger mosquito. This way 52 novel putative secreted salivary proteins were identified in *A. albopictus* and among these 20 represent full-length genes (Table S6). The availability of the complete salivary repertoire of *A. albopictus* will allow further functional studies and may help in developing reliable serological tools to evaluate human exposure to Aedes arboviral vectors, as previously done for Anopheles malaria vectors.[86,87]

miRNA-encoding genes

The class of microRNAs (miRNAs) contains small, 22–23 nt long molecules that have been found to play crucial roles in regulating gene expression both in plants and metazoans. Several papers have dealt with the characterisation of miRNAs in mosquitoes, including *A. albopictus*.[88,89] In the first referenced case a total of 104 miRNA species were identified while in the second referenced study 64 species were found; it should be stressed, though, that the latter study concerned the analysis of a cell line of the Asian tiger mosquito. In our analysis of the Assembly 1.01, we were able to discover a total of 93 miRNAs, of which 30 were characterised as novel. From those novel miRNAs, 26 have the same seed with a miRNA found in miRBase while four share no homology or similarity with any known miRNA stored in miRBase. The allocation of the miRNAs to three groups (male adult mosquitoes, sugar- and blood-fed adult females) was 84, 51 and 65, respectively. Naturally, some of the miRNAs were shared by the three groups; thus, 47 were common between male and sugar fed females, 47 between sugar fed and blood females while 57 between males and blood fed females. The miRNA genes identified are listed in Table S7. Whether our analysis has identified all miRNA genes present in the genome studied cannot be assessed, but due to the small size of the genes we are confident that, no genes will remain undetermined due to the fragmentation reported earlier.

Conclusions

We reported here the summary of the whole genome sequencing project of *A. albopictus*. This assembled genome does not represent an optimal result; indeed, as was the case for *A. aegypti*, the genome of *albopictus* is comprised of a large proportion of repetitive DNA segments, including many transposable elements, a fact that makes an optimal assembly extremely tedious, if not impossible. In addition to that, of course, in contrast to the published *A. aegypti*, genome the *A. albopictus* genome was based entirely on NGS. Using this approach, the assembly into contigs of short sequences that are often repeated hundreds and thousands of times is an extremely difficult undertaking. Given the fact that the assembly, as it stands today, can be used to isolate and study most protein-coding genes it can, thus, be considered a milestone to be used to 'find one's way'. This is especially true when one keeps in mind the cost-benefit of such a project.

The analysis of the genome did not provide any final clues as to the reasons of the dramatic differences seen in the size of the genomes of different strains of *A. albopictus*. It is more than probable that the high number of repetitive sequences (among which a large number of transposable elements) could be the cause for the differing genome size determined for this species. Whether the diversity only concerns, in this magnitude, specimens that have been collected in different parts of the globe, of whether it is also true for individuals that form part of the same population cannot be answered with the present data. We hope to be able to provide more precise answers when the three genomes that are currently near closure will be compared to each other. Finally, since our annotation had used the assembly 1.01, we reported these data in the present report, but we plan to release, in the near future, an updated assembly onto which the features presented here will be remapped, and which will be based on updated data such as, for example, the newly determined average insert size of the MP library.

Acknowledgements

We would like to dedicate this paper to our friend, colleague and mentor Fotis Kafatos on occasion of his 75th birthday. We want to thank Romeo Bellini for furnishing the initial stock of *A. albopictus* used for the determination of the genome sequence, Francesca Marini and Angiola Desiderio who helped with the transcriptomic analysis of *A. albopictus* testes aiming at identifying *Wolbachia*-induced sperm modifications.

Disclaimer statements

Contributors

VD and PT were responsible for the assembly and overall annotation of the genome, as well as for the annotation of the tRNA, miRNA genes and the repetitive sequences in the genome, NW, AS, AH, DL, MH, DH, VN, FC were responsible for the acquisition of the primary sequence data, ED assisted the acquisition of the short RNAseq data and was responsible for the annotation of genes encoding miRNAs, GG, LMG, GS, MM, FS, AM were responsible for the annotation of the odorant binding protein and odorant receptor genes, BA, JMR, FL were responsible for the annotation of the salivary gland genes, GS, MS were responsible for the annotation of the sex-determination genes, RM, GA, MC, were responsible for the annotation of the genes

involved in chromatin condensation, MP assisted with the bioinformatics analysis, PAP contributed to the interpretation of data, RS, GF, AC as well as all authors mentioned above helped draft and revise the article, AC coordinated the overall project, CL coordinated the bioinformatics part of the project and drafted the article.

Funding

This work was funded by the Infravec Consortium (FP7 programme of the European Union) and, partially, by funds from the Hellenic Secretariat General for research and Technology. VD and CL were supported, in part, by i-Move fellowships from the EU's Marie Curie programme. JMR was funded by the intramural programme of the NIAID. The sex determination studies were supported by a grant from Campania Region (LG 5/2001; 2008; Biotechnological approaches for the control of the human disease vector Aedes albopictus) to GS and by a grant from the University of Naples Federico II and Compagnia di San Paolo, Naples, ITALY, to MS in the frame of Programme STAR (STAR2013_25) from the University of Naples Federico II and Compagnia di San Paolo, Naples, ITALY. ARM was funded by a grant from the Italian Ministry of Health (RF-2010-2318965).

Conflicts of interest

None for any of the authors.

Ethics approval

Not applicable for this paper.

References

1 Gratz NG. Critical review of the vector status of *Aedes albopictus*. Med Vet Entomol. 2004;18:215–27.
2 Adhami J, Murati N. Prani e mushkonjës *Aedes albopictus* në shqipëri. Rev Mjekësore. 1987;1:13–16.
3 Sabatini A, Raineri V, Trovato G, Coluzzi M. *Aedes albopictus* in Italia e possibile diffusione della specie nell'area mediterranea. Parasitologia. 1990;32:301–04.
4 Halstead SB, Papaevangelou G. Transmission of dengue 1 and 2 viruses in Greece in 1928. Am J Trop Med Hyg. 1980;29(4):635–37.
5 Rosen L. Dengue in Greece in 1927 and 1928 and the pathogenesis of dengue hemorrhagic fever: new data and a different conclusion. Am J Trop Med Hyg. 1986;35(3):642–53.
6 Louis C. Daily newspaper view of dengue fever epidemic, Athens, Greece, 1927-1931. Emerg Infect Dis. 2012;18(1):78–82.
7 Waterston J. On the mosquitoes of Macedonia. Bull Entomol Res. 1918;9:1–12.
8 Joyeux C. Notes sur les Culicides de Macedoine. Bull Soc Pathol Exot. 1918;11:530–47.
9 Samanidou-Voyadjoglou A, Darsie RF Jr. New country records for mosquito species in Greece. J Am Mosq Control Assoc. 1993;9(4):465–66.
10 Livadas GA. Malaria eradication in Greece. Riv Malariol. 1958;37:173–91.
11 Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, *et al.* A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. Vector Borne Zoonotic Dis. 2012;12:435–47.
12 Forratini OP. Identificacao de *Aedes (Stegomyia) albopictus* do Brasil. Rev Saude Publica. 1986;20(244):245.
13 Hawley WA. The biology of *Aedes albopictus*. J Am Mosq Control Assoc. 1988;4(Suppl):1–40.
14 Knudsen AB. Global distribution and continuing spread of *Aedes albopictus*. Parassitologia. 1995;37(2-3):91–97.
15 Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species *Aedes albopictus*: current knowledge and future perspectives. Trends Parasitol. 2013;29(9):460–68.
16 Tomasello D, Schlagenhauf P. Chikungunya and dengue autochthonous cases in Europe, 2007-2012. Travel Med Infect Dis. 2013;11(5):274–84.
17 Schaffner F, Mathis A. Dengue and dengue vectors in the WHO European region: past, present, and scenarios for the future. Lancet Infect Dis. 2014;14(12):1271–80.
18 Medlock JM, Hansford KM, Versteirt V, Cull B, Kampen H, Fontenille D, *et al.* An entomological review of invasive mosquitoes in Europe. Bull Entomol Res. 2015;25:1–27.
19 Crisanti A, INFRAVEC: research capacity for the implementation of genetic control of mosquitoes. Pathog Glob Health. 2013;107:458–62.
20 Goubert C, Modolo L, Vieira C, Valiente-Moro C, Mavingui P, Boulesteix M. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). Genome Biol Evol. 2015;7(4):1192–205.
21 Bellini R, Calvitti M, Medici A, Carrieri M, Celli G, Maini S. Use of the sterile insect technique against *Aedes albopictus* in Italy: first results of a pilot trial. In: Vreysen MJB, Robinson AS, Hendrichs J, editors. Area-wide control of insect pests: from research to field implementation. Dordrecht, The Netherlands: Springer; 2007; p. 505–15.
22 Wilhelm J, Pingoud A, Hahn M. Real-time PCR-based method for the estimation of genome sizes. Nucleic Acids Res. 2003;31(10):e56.
23 Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics. 2008;24:2395–96.
24 Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40(1):37–52.
25 Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–29.
26 Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19(6):1117–23.
27 Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, *et al.* De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010;20:265–72.
28 Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva EV. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 2013; 41(Database issue):D358–65.
29 Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinformatics. 2014;48:4.11.1–4.11.39.
30 Smith CD, Edgar RC, Yandell MD, Smith DR, Celniker SE, Myers EW, *et al.* Improved repeat identification and masking in dipterans. Gene. 2007;389(1):1–9.
31 Lowe TM, Eddy S. tRNAScan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.
32 Molla-Herman A, Matias NR, Huynh J-R. Chromatin modifications regulate germ cell development and transgenerational information relay. Curr Opin Insect Sci. 2014;1:10–18.
33 Bourtzis K, Dobson SL, Xi Z, Rasgon JL, Calvitti M, Moreira LA, *et al.* Harnessing mosquito-*Wolbachia* symbiosis for vector and disease control. Acta Trop. 2014;132(Suppl):S150–63.
34 Landman F, Orsi GA, Loppin B, Sullivan W. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. PLoS Pathog. 2009;5(3):e1000343.
35 Calvitti M, Moretti R, Lampazzi E, Bellini R, Dobson SL. Characterization of a new *Aedes albopictus* (Diptera: Culicidae)-*Wolbachia pipientis* (Rickettsiales: Rickettsiaceae) symbiotic association generated by artificial transfer of the *w*Pip strain from *Culex pipiens* (Diptera: Culicidae). J Med Entomol. 2010;47:179–87.
36 Poelchau MF, Reynolds JA, Denlinger DL, Elsik CG, Armbruster PA. A de novo transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. BMC Genomics. 2011;12:619.
37 Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
38 Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

39 Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
40 Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–15.
41 Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013;31:46–53.
42 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
43 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al.* Gapped BLAST and PSI-BLAST – a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
44 Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, *et al.* Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science. 22 June 2007;316(5832):1718–23.
45 Manoharan M, Chong MNF, Vaïtinadapoulé A, Frumence E, Sowdhamini R, Offmann B. Comparative genomics of odorant binding proteins in *Anopheles gambiae*, *Aedes aegypti*, and *Culex quinquefasciatus*. Genome Biol Evol. 2013;5(1):163–80.
46 Bohbot J, Pitts RJ, Kwon H-W, Rützler M, Robertson HM, Zwiebel LJ. Molecular characterization of the *Aedes aegypti* odorant receptor gene family. Insect Mol Biol. 2007;16(5):525–37.
47 Bohbot JD, Sparks JT, Dickens JC. The maxillary palp of *Aedes aegypti*, a model of multisensory integration. Insect Biochem Mol Biol. 2014;48:29–39.
48 Huang X, Madan A. CAP3: a DNA sequence assembly program. Genome Res. 1999;9(9):868–77.
49 Birney E, Durbin R. Using Genewise in the *Drosophila* annotation experiment. Genome Res. 2000;10(4):547–48.
50 Pane A, Salvemini M, Delli Bovi P, Polito C, Saccone G. The transformer gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. Development. 2002;129(15):3715–25.
51 Arca B, Lombardo F, Francischetti IM, Pham VM, Mestres-Simon M, Andersen JF, *et al.* An insight into the sialome of the adult female mosquito *Aedes albopictus*. Insect Biochem Mol Biol. 2007;37:107–27.
52 Ribeiro JM, Arca B, Lombardo F, Calvo E, Phan VM, Chandra PK. An annotated catalogue of salivary gland transcripts in the adult female mosquito, *Aedes aegypti*. BMC Genomics. 2007;8:6.
53 Kent WJ. BLAT – the BLAST-like alignment tool. Genome Res. 2002;12:656–64.
54 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
55 Ribeiro JM, Topalis P, Louis C. AnoXcel: an *Anopheles gambiae* protein database. Insect Mol Biol. 2004;13:449–57.
56 Berriman M, Rutherford K. Viewing and annotating sequence data with Artemis. Brief Bioinform. 2003;4:124–32.
57 McLain DK, Rai KS, Fraser MJ. Intraspecific and interspecific variation in the sequence and abundance of highly repeated DNA among mosquitoes of the *Aedes albopictus* subgroup. Heredity (Edinb). 1987;58(Pt 3):373–81.
58 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, *et al.* Versatile and open software for comparing large genomes. Genome Biol. 2004;5:R12.
59 Rao PN, Rai K. Inter and intraspecific variation in nuclear DNA content in *Aedes* mosquitoes. Heredity (Edinb). 1987;59:253–58.
60 Dritsou V, Deligianni E, Dialynas E, Allen J, Poulakakis N, Louis C, *et al.* Non-coding RNA gene families in the genomes of anopheline mosquitoes. BMC Genomics. 2014;15:1038.
61 Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature. 2007;450:203–18.
62 Behura SK, Severson DW. Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in *Aedes aegypti* and *Anopheles gambiae*. Insect Mol Biol. 2011;20:177–87.
63 Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol. 2013;58:373–91.
64 Arnbruster P, White S, Dzundza J, Crawford J, Zhao X. Identification of genes encoding atypical odorant-binding proteins in *Aedes albopictus* (Diptera: Culicidae). J Med Ent. 2009;46(2):271–80.
65 Li C, Yan T, Dong Y, Zhao T. Identification and quantitative analysis of genes encoding odorant binding proteins in *Aedes albopictus* (Diptera: Culicidae). J Med Entomol. 2012;49(3):573–80.
66 Deng Y, Yan H, Gu J, Xu J, Wu K, Tu Z, *et al.* Molecular and functional characterization of odorant-binding protein genes in an invasive vector mosquito, *Aedes albopictus*. PLoS One. 2013;8(7):e68836.
67 Scialò F, Hansson BS, Giordano E, Polito CL, Digilio FA. Molecular and functional characterization of the odorant receptor2 (OR2) in the tiger mosquito *Aedes albopictus*. PLoS One. 2012;7(5):e36538.
68 Walker T, Johnson PH, Moreira LA, Iturbe-Ormaetxe I, Frentiu FD, McMeniman CJ, *et al.* The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. Nature. 2011;476:450–53.
69 Presgraves DC. A genetic test of the mechanism of *Wolbachia*-induced cytoplasmic incompatibility in *Drosophila*. Genetics. 2000;154:771–76.
70 Zheng Y, Ren P-P, Wang J-L, Wang Y-F. *Wolbachia*-induced cytoplasmic incompatibility is associated with decreased hira expression in male *Drosophila*. PLoS One. 2011;6(4):e19512.
71 Landmann F, Orsi GA, Loppin B, Sullivan W. *Wolbachia*-mediated cytoplasmic incompatibility is associated with impaired histone deposition in the male pronucleus. PLoS Pathog. 2009;5(3):e1000343.
72 Beckmann JF, Fallon AM. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. Insect Biochem Mol Biol. 2013;43:867–78.
73 McClelland G. Sex-linkage in *Aedes aegypti*. Trans R Soc Trop Med Hyg. 1962;56:4.
74 Craig GB, Hickey WA. Genetics of *Aedes aegypti*. In: Wright JW, Pal R, editors. Genetics of insect vectors of disease. New York: Elsevier; 1967; p. 67–131.
75 Severson DW, Mori A, Kassner VA, Christensen BM. Comparative linkage maps for the mosquitoes, *Aedes albopictus* and *A. aegypti*, based on common RFLP loci. Insect Mol Biol. 1995;4(1):41–45.
76 Sutherland IW, Mori A, Montgomery J, Fleming KL, Anderson JM, Valenzuela JG, *et al.* A linkage map of the Asian tiger mosquito (*Aedes albopictus*) based on cDNA markers. J Hered. 2011;102(1):102–12.
77 Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler J. A male-determining factor in the mosquito *Aedes aegypti*. Science. 2015;348(6240):1268–70.
78 Papathanos PA, Bossin HC, Benedict MQ, Catteruccia F, Malcolm CA, Alphey L, *et al.* Sex separation strategies: past experience and new approaches. Malar J. 2009;8(Suppl 2):S5.
79 Alphey L, McKemey A, Nimmo D, Neira Oviedo M, Lacroix R, Matzen K, *et al.* Genetic control of *Aedes* mosquitoes. Pathog Glob Health. 2013;107(4):170–79.
80 Marinotti O, Jasinskiene N, Fazekas A, Scaife S, Fu G, Mattingly ST, *et al.* Development of a population suppression strain of the human malaria vector mosquito, *Anopheles stephensi*. Malar J. 2013;12:142.
81 Koukidou M, Alphey L. Practical applications of insects' sexual development for pest control. Sex Dev. 2014;8(1–3):127–36.
82 Ribeiro JM, Mans BJ, Arca B. An insight into the sialome of blood-feeding Nematocera. Insect Biochem Mol Biol. 2010;40:767–84.
83 Ribeiro JMC, Arcà B. From sialomes to the sialoverse: an insight into salivary potion of blood-feeding insects. Adv Insect Physiol. 2009;37:59–118.
84 Arca B, Struchiner CJ, Pham VM, Sferra G, Lombardo F, Pombi M, *et al.* Positive selection drives accelerated evolution of mosquito salivary genes associated with blood-feeding. Insect Mol Biol. 2014;23:122–31.
85 Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. Science. 2015;347:1258522.
86 Rizzo C, Ronca R, Fiorentino G, Verra F, Mangano V, Poinsignon A. Humoral response to the *Anopheles gambiae* salivary protein gSG6: a serological indicator of exposure to Afrotropical malaria vectors. PLoS One. 2011;6:e17980.
87 Stone W, Bousema T, Jones S, Gesase S, Hashim R, Gosling R, *et al.* IgG responses to *Anopheles gambiae* salivary antigen gSG6 detect variation in exposure to malaria vectors and disease risk. PLoS One. 2012;7:e40170.
88 Skalsky RL, Vanlandingham DL, Scholle F, Higgs S, Cullen BR. Identification of microRNAs expressed in two mosquito vectors, *Aedes albopictus* and *Culex quinquefasciatus*. BMC Genomics. 2010;11:119.
89 Gu J, Hu W, Wu J, Zheng P, Chen M, James AA, *et al.* miRNA genes of an invasive vector mosquito, *Aedes albopictus*. PLoS One. 2013;8(7):e67638.

Scolari et al. BMC Genetics 2014, 15(Suppl 2):S11
http://www.biomedcentral.com/1471-2156/15/S2/S11

BMC Genetics

RESEARCH Open Access

How functional genomics will impact fruit fly pest control: the example of the Mediterranean fruit fly, *Ceratitis capitata*

Francesca Scolari, Ludvik M Gomulski, Paolo Gabrieli, Mosè Manni, Grazia Savini, Giuliano Gasperi, Anna R Malacrida*

Abstract

The highly invasive agricultural insect pest *Ceratitis capitata* (Diptera: Tephritidae) is the most thoroughly studied tephritid fruit fly at the genetic and molecular levels. It has become a model for the analysis of fruit fly invasions and for the development of area-wide integrated pest management (AW-IPM) programmes based on the environmentally-friendly Sterile Insect Technique (SIT). Extensive transcriptome resources and the recently released genome sequence are making it possible to unravel several aspects of the medfly reproductive biology and behaviour, opening new opportunities for comparative genomics and barcoding for species identification. New genes, promotors and regulatory sequences are becoming available for the development/improvement of highly competitive sexing strains, for the monitoring of sterile males released in the field and for determining the mating status of wild females. The tools developed in this species have been transferred to other tephritids that are also the subject of SIT programmes.

Background

The Mediterranean fruit fly (medfly), *Ceratitis capitata* Wiedemann, is one of the world's most destructive agricultural insect pests [1-3]. Due to its global distribution and history of rapid and devastating outbreaks [4-6], the medfly is the most thoroughly studied "true" fruit fly (Diptera: Tephritidae) [7] at the genetic and molecular levels. It has thus become a model species for the analysis of fruit fly invasions [8] and for the development of control strategies [9]. Medfly outbreaks have been successfully controlled through area-wide integrated pest management (AW-IPM) programmes based on the environmentally-friendly Sterile Insect Technique (SIT) [10]. In the SIT, the reduction of pest population size is achieved through mass release of reproductively sterile male insects into a wild-type population [11]. Males rendered sterile through ionizing radiation compete with wild-type males for matings and deplete female reproductive success. Preventative sterile male releases have been and are currently applied in areas where the climatic conditions and the availability of suitable hosts for oviposition are particularly favourable for medfly establishment, such as California, Southern Australia and Florida [12-16]. To be most successful, this approach requires i) knowledge of the genetic background of the released males and the genetic structure of the target population, ii) a sexing strain for male-only production, iii) a sterilization system that inflicts the least possible fitness load, and iv) effective procedures to monitor the efficiency of the programmes.

In the last 20 years, enormous progress has been made in understanding medfly biology, with the goal of developing and optimizing a wide range of molecular tools for the implementation of population control strategies (Figure 1). Population genetics provided useful approaches for reconstructing the routes of medfly invasion, highlighting the complexity of the process [4,5,17-25]. *Ceratitis capitata* was the first non-drosophilid species in which the germline was transformed [26], enabling studies on its biology in ways that were previously impossible [27-34].

The application of functional genomics tools, together with the recent release of the medfly genome sequence (http://arthropodgenomes.org/wiki/i5K;https://www.hgsc.bcm.edu/arthropods/medfly-genome-annotation-groups),

* Correspondence: malacrid@unipv.it
Department of Biology and Biotechnology, University of Pavia, Pavia, Italy

Scolari *et al. BMC Genetics* 2014, **15**(Suppl 2):S11
http://www.biomedcentral.com/1471-2156/15/S2/S11

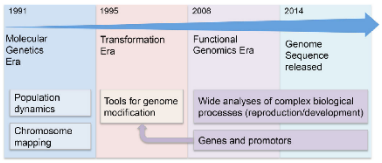


Figure 1 Molecular timeline of medfly research.

allows a more detailed analysis of the complex biological traits that underpin the adaptive potential of this fly at all developmental stages (Figure 2)[8,35]. Indeed, functional genomics provides powerful evolutionary tools to interpret how medfly (either wild or transgenic) develop and respond to the environment. Different aspects of development, behaviour, sexual maturation, and reproduction can now be examined both in terms of gene expression profiles and protein analyses [36-43]. New genes, promotors and regulatory sequences are consequently becoming available for i) the development/improvement of competitive sexing strains, ii) the monitoring of released males in the field, and iii) for determining the mating status of wild females.

Medfly embryogenesis

A reservoir of early male/female differentially expressed genes and sex regulatory sequences is now available for unravelling the first steps of medfly embryogenesis, i.e. when the maternal-to-zygotic transition (MTZ) occurs and when the sexual fate is established at the molecular level [36,38]. As a practical consequence, promotor and enhancer sequences that are active in early stages of development are becoming available as tools for the future generation and/or improvement of the existing conditional embryonic and female-specific lethality systems developed using conventional techniques. Female-specific lethality systems were developed based on alternative splicing of the *Cctransformer* gene (*Cctra*) [31]. Moreover, cellularisation-specific promotors/enhancers allowed the development of a transgenic embryo-specific lethality system [33]. More recently, the combination of the *Cctra*-based female-specific lethality [31] with the embryonic lethality system [33], yielded a female-specific embryonic lethality (FSEL) system in this species [44].

In this context, the medfly genes with vital functions in early embryonic development, such as those involved in sex determination and cellular blastoderm formation, are of direct use [38]. Their zygotic transcriptional activation follows two waves. The first wave starts within four hours after oviposition and includes the zygotic genes *Ccsisterless A* (*CcsisA*), *Ccdeadpan* (*Ccdpn*) and *Ccslow-as-molasses* (*Ccslam*). The second major burst of expression activation begins five hours after oviposition and includes the maternal-zygotic genes *Ccgroucho* (*Ccgro*), *CcSex-lethal* (*CcSxl*), *Cctransformer* (*Cctra*), *Ccfemale-lethal d* (*Ccfl(2)d*), *CcRho1* and *Ccserendipity-α* (*Ccsry-α*) [38]. During this transition, sexual identity is established at the molecular level, before cellularisation of the embryo occurs. Unlike *Drosophila* [46,47], *Cctra* is the key-gene of the sex-determination cascade: it generates mRNAs encoding full-length active proteins only in females and displays an autocatalytic function, that

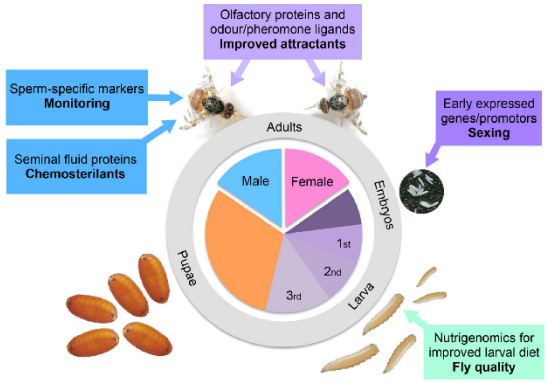


Figure 2 Medfly functional genomics resources and their impact on the improvement of the SIT.

Scolari et al. BMC Genetics 2014, 15(Suppl 2):S11
http://www.biomedcentral.com/1471-2156/15/S2/S11

guarantees the female-specific development of cell memory [47]. *Cctra*, in cooperation with *Cctra2*, determines the sex-specific splicing of *Ccdsx*, the transcription factor that is the regulator of the sex-differentiation processes. The mother supplies the embryos with *Cctra* and *Ccdsx* female-specific splicing variants. Subsequently, the maternal information for female-specific development is reset in embryos through the reprogramming of *Cctra* mRNA splicing and the degradation of the maternal *Ccdsx* mRNAs [38]. The precise timing of sex-specific splicing [38], as well as the proof of evidence that transgenic dsRNA for *tra* is effective in the conditional production of male-only progeny [48], can be exploited for the development of novel sexing strains.

Metabolic regulation of sexual maturation and mating

The production of highly competitive males is an essential requisite for effective SIT. Transcriptome and microarray-based functional analyses performed on whole flies and specific tissues are providing basic information on the pathways involved in primary metabolism, hormone synthesis, neurological-related processes, gametogenesis, signalling, and sensory perception [36,37,39,42]. The regulation of these specific pathways and biological processes can be affected by long-term artificial rearing, that may translate into reduced quality of individuals released for SIT [42]. In particular, down-regulation of signaling and neurological processes, especially those related to light and chemical stimuli, muscle development, muscle differentiation and locomotion, have been reported as a consequence of mass rearing in artificial conditions [42]. In this context, nutrigenomics can provide valuable information on how nutrition affects gene expression patterns, offering the means to measure male and female medfly responses to changes in the food stream, but also providing information on diet limitations [49]. This is a priority for operational SIT. Transcriptional baseline profiles of key biological pathways involved in sexual maturation of both sexes, and also in response to mating, are available for medflies reared on the standard diet used in mass rearing facilities [37]. Indeed, we know that medfly female maturation requires the activation of fatty acid metabolism as a reflection of the high energy requirements for female reproductive success, such as foraging, nutrient storage and egg development [50]. In addition, Gene Ontology (GO) enrichment analyses revealed that, in mature females, transcript categories related to memory/learning behaviours and visual and olfactory functions are significantly overrepresented [37]. By contrast, male sexual maturation requires the activation of carbohydrate and protein metabolism for energy production and muscle activities, memory formation, smell recognition and pheromone production [37]. All these activities suggest an investment required for lek formation and courtship [51]. Despite extensive post-mating transcriptional changes in the male, changes in the female were surprisingly modest. Indeed, in the male, mating does not down-regulate the transcriptional activities of genes implicated in lek formation/courtship, whereas it increases the activities of genes related to fitness (i.e. *double time* and *Basigin*) [37].

Some of these pathways are down-regulated by irradiation [42]. This is the case of processes related to visual and chemical responses, and those associated with muscle development and locomotion. These irradiation-related changes may have an impact on the competitiveness of mass reared flies.

Studies on improved diets or chemical manipulation of the adult environment offer promising options for the improvement of sterile male competitiveness. Approaches aimed at the improvement of the sexual performance of sterile males include i) altering the olfactory environment experienced by freshly eclosed individuals, providing high-quality post-teneral nutrition [52] and ii) inoculating males with probiotic bacteria [53,54].

Male reproduction

A better understanding of the reproductive biology of the medfly should permit the development of novel or improved approaches to impact male reproductive success and/or regulate female mating behaviour and fertility. In this respect, transcriptomics and proteomics of reproductive tissues will help to identify genes and promotors. Testes and male accessory glands (MAGs) participate in the maintenance of complementary reproductive functions. In the testes, the key regulatory genes of spermatogenesis tend to be conserved to guarantee the male-specific processes required for sperm production [55,56]. By contrast, the accessory gland secretions act as key factors in male insect reproductive success, and the genes expressed in the MAGs are subject to rapid evolution as a result of sexual conflict and competition [57]. A transcriptome-based analysis performed on medfly testes and male accessory gland tissues resulted in a database of 3344 unique sequences [39]. Transcripts related to spermatogenesis, fertility, sperm-egg binding, as well as those involved in the production of seminal fluid proteins (SFPs), were identified. Some of the SFP transcripts displayed a mating-responsive profile [39]. These will be ideal targets for the development of novel and more specific environmentally-friendly chemosterilants [58,59] that mimic the behaviour-modulating effects of MAG proteins, i.e. by impeding correct sperm storage, or interfering with female remating.

Over a third of the transcripts from these two tissues shared no significant similarities to known genes from other organisms. Considering that they may represent novel and/or fast-evolving sequences, they represent

Scolari *et al. BMC Genetics* 2014, **15**(Suppl 2):S11
http://www.biomedcentral.com/1471-2156/15/S2/S11

ideal targets for the development of species-specific diagnostic markers.

Improved SIT monitoring strategies

A major issue in the monitoring activities for evaluating SIT effectiveness is the difficulty in assessing the capacity of released sterile males to inseminate wild-type females [60-62]. The availability of the testes- and sperm-specific *Ccβ2-tubulin* gene has allowed the use of its promotor for fluorescent protein marking of the spermatozoa, and hence to detect females that have mated with released males [32]. Using this marking system, strains have been generated and evaluated for their ability to transfer green or red fluorescent sperm to the female spermathecae. It has been proven that these sperm remain viable and fluorescent for a long time within the spermathecae, also after female death [32] (Figure 3). The transgene previously inserted in one of these lines, namely 1260_F-3_m-1, was then efficiently modified by the use of the site-specific integration system from phage *phiC31* [34]. Post-integrational excision of one of the *piggyBac* inverted terminal repeats resulted in stably integrated transgene insertions that, being inert to the *piggyBac* transposase, could not be remobilized. This allowed the development of an optimized strain for pest control that minimizes environmental concerns (stab_1260_F-3_m-1)[34]. Once integrated into the medfly GSS Vienna-8 strain, this sperm marking system may offer valuable alternatives to the currently used fluorescent powders [63] that are detected in trapped flies using UV light. Moreover, this sperm marking system can also be integrated into strains carrying diverse transgenes in tandem, for example with conditional embryonic lethality [33] and sexing systems [31].

For monitoring activities, one of the priorities is the development of powerful species- and sex-specific attractants. In this context, it is essential to identify the components of the molecular machinery that recognizes and binds external ligands (odours and pheromone components) and translates this interaction into electrical signals to the central nervous system. Three main groups of molecules are involved: odorant-binding proteins (OBPs), chemosensory proteins (CSPs), and the chemoreceptor superfamily formed by the olfactory (OR), gustatory (GR) and ionotropic (IR) receptor families [64,65]. The chemosensory gene repertoire of the medfly is currently being characterized at the functional genomics and structural level [40,41]. So far, one antennal-enriched OBP appears to be particularly promising for practical applications. Indeed, it displayed highest binding specificity for (E,E)-α-farnesene, a major component of male pheromone blend, and also for Trimedlure, a strong synthetic medfly attractant [41]. The resolution of the three-dimensional structure of this medfly OBP will be the premise for the design of synthetic molecules able to act as antagonists of the natural ligand/s. Such optimized molecules need to be further evaluated and tested for side-effects before they can be used in AW-IPM approaches.

Conclusions

The extensive transcriptome resources now available for the medfly (Table 1) will greatly improve the on-going annotation of the genome. They will also facilitate the generation of genomic data from other tephritid species of agricultural importance [66-71], opening new ways for comparative genomics and barcoding for species identification. In addition, the structural and functional genomics (transcriptomics, proteomics, RNA interference etc) tools that are being developed in the medfly can be extended to other tephritids that are also the subject of

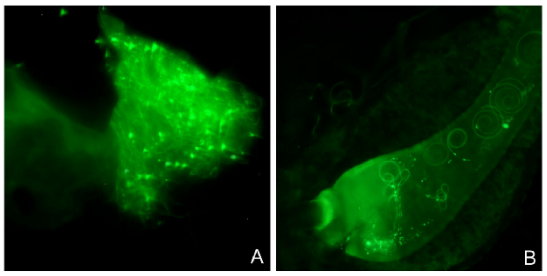


Figure 3 Transgenic sperm can be easily traced in the reproductive tract of laboratory wild-type females. Mechanically opened spermatheca isolated from a laboratory wild-type female mated with a transgenic male with green fluorescent sperm [32], three days after death (A). Spermathecal duct dissected from a laboratory wild type female mated to a transgenic male with green fluorescent sperm [32] 24 hours after mating (B). Images were captured using an epifluorescence Zeiss Axioplan microscope at 400x magnification with the Zeiss filters set 13.

Table 1 Transcriptome and microarray resources available for the medfly.

Database	Tissue/status	Strain	Type	Accession numbers	Ref.
NCBI dbEST database	Embryo (from 30 min to 36 hr after oviposition)	Ispra	Expressed sequence tags	FG068301-FG078567	[36]
NCBI dbEST database	Adult male and female heads (from 30 min to 8 days after emergence).	Ispra	Expressed sequence tags	FG078568-FG089553	[36]
NCBI dbEST database	Adult testes and male accessory glands.	Ispra	Expressed sequence tags	JK832450-JK838363	[39]
NCBI dbEST database	Adult male accessory glands.	Guatemala mass-rearing strain (Moscamed)	Expressed sequence tags	DQ406805-DQ406817	[72]
NCBI GEO Dataset	Adult female head. Immature versus mature.	Ispra	Microarray	GSE19571	[37]
NCBI GEO Dataset	Adult male head. Immature versus mature.	Ispra	Microarray	GSE19572	[37]
NCBI GEO Dataset	Adult female head. Mated versus Virgin.	Ispra	Microarray	GSE19573	[37]
NCBI GEO Dataset	Adult Male head. Mated versus Virgin.	Ispra	Microarray	GSE19608	[37]
NCBI Sequence Read Archive	Adult, whole body, Irradiated	Vienna 7	Illumina Hiseq 2000 sequencing	SRX312172-SRX312174	[42]
NCBI Sequence Read Archive	Adult, whole body, Non-irradiated	Vienna 7	Illumina Hiseq 2000 sequencing	SRX312183-SRX312185	[42]
NCBI Sequence Read Archive	Pupae, Irradiated	Vienna 7	Illumina Hiseq 2000 sequencing	SRX312176-SRX312180	[42]
NCBI Sequence Read Archive	Pupae, Non-irradiated	Vienna 7	Illumina Hiseq 2000 sequencing	SRX312186-SRX312188	[42]
NCBI Sequence Read Archive	Adult, whole body, Non-irradiated	Wild Hawaii	Illumina Hiseq 2000 sequencing	SRX312189-SRX312191	[42]
NCBI Sequence Read Archive	Pupae, Non-irradiated	Wild Hawaii	Illumina Hiseq 2000 sequencing	SRX312192-SRX312194	[42]

SIT programmes, such as *Anastrepha* and *Bactrocera* species (*A. ludens, A. suspensa, A. obliqua, A. fraterculus, B. cucurbitae, B. tryoni, B. dorsalis, B. correcta*)[10].

The increased knowledge of the biology of the medfly acquired through genomic approaches will also facilitate the further development of regulations for the transfer and potential field release of genetically modified fruit flies.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
The work was carried out within the FAO/IAEA research CRP programme "Development and evaluation of improved strains of insect pests for SIT" (GG) and the IAEA Technical Contract No:16966 (GG). The work was also partially funded by a PRIN grant from the Italian Ministry of Education, University and Research (MIUR) (20077 RCHRW) (LMG).
The authors would like to thank the two anonymous referees for their valuable comments which helped to improve the manuscript.
This article has been published as part of *BMC Genetics* Volume 15 Supplement 2, 2014: Development and evaluation of improved strains of insect pests for SIT. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcgenet/supplements/15/S2.
Publication of this supplement was funded by the International Atomic Energy Agency. The peer review process for articles published in this supplement was overseen by the Supplement Editors in accordance with BioMed Central's peer review guidelines for supplements. The Supplement Editors declare that they have no competing interests.

Published: 1 December 2014

References

1. Bateman MA: **The ecology of fruit flies.** *Annu Rev Entomol* 1972, **17**:493-518.
2. Aluja M, Mangan RL: **Fruit fly (Diptera: Tephritidae) host status determination: critical conceptual, methodological, and regulatory considerations.** *Annu Rev Entomol* 2008, **53**:473-502.
3. Papadopoulos NT, Plant RE, Carey JR: **From trickle to flood: the large-scale, cryptic invasion of California by tropical fruit flies.** *Proc Biol Sci* 2013, **280(1768)**:20131466.
4. Gasperi G, Bonizzoni M, Gomulski LM, Murelli V, Torti C, Malacrida AR, Guglielmino CR: **Genetic differentiation, gene flow and the origin of infestations of the medfly, *Ceratitis capitata*.** *Genetica* 2002, **116(1)**:123-135.
5. Malacrida AR, Gomulski LM, Bonizzoni M, Bertin S, Gasperi G, Guglielmino CR: **Globalization and fruitfly invasion and expansion: the medfly paradigm.** *Genetica* 2007, **131(1)**:1-9.
6. Carey JR: **The Mediterranean fruit fly (*Ceratitis capitata*) invasion of California deepens: response to an alternative explanation for recurring outbreaks.** *Am Entomol* 2010, **56**:158-163.
7. White IM, Elson-Harris MM: *Fruit flies of economic significance: their identification and bionomics* Wallingford, Oxon, UK: CAB International; 1992.
8. Diamantidis AD, Carey JR, Nakas CT, Papadopoulos NT: **Population-specific demography and invasion potential in medfly.** *Ecol Evol* 2011, **1(4)**:479-488.
9. Hendrichs J, Robinson AS, Cayol JP, Enkerlin W: **Medfly Area Wide Sterile Insect Technique Programmes for Prevention, Suppression or Eradication: the importance of Mating Behavior Studies.** *Fla Entomol* 2002, **85(1)**:1-13.
10. Dyck VA, Hendrichs J, Robinson AS: **Sterile Insect Technique: Principles and practice in Area-wide Integrated Pest Management.** Dordrecht The Netherlands: Springer; 2005.
11. Knipling EF: **Possibilities of insect control or eradication through the use of sexually sterile males.** *J Econ Entomol* 1955, **48**:459-462.
12. Smallridge CJ, Hoskins DC: **Preventative sterile fly release for the management of a Mediterranean fruit fly outbreak in South Australia.** In

Proceedings of the 6th International Symposium on Fruit Flies of Economic Importance. Johannesburg: Ultra Litho (Pty) Ltd;Brian BN 2004:213-215.
13. Dowell RV, Siodiqaw IA, Meyer F, Spaugy EL: **Mediterranean fruit fly preventative release programme in southern California.** In *International conference on area-wide control of insect pests and the fifth international symposium on fruit flies of economic importance; Penang, Malaysia*. Penerbit Universiti Sains Malaysia, Penang, Malaysia;Tan K-H 2000:369-377.
14. Hood-Nowotny R, Mayr L, Islam A, Robinson A, Caceres C: **Routine isotope marking for the Mediterranean fruit fly (Diptera: Tephritidae).** *J Econ Entomol* 2009, **102(3)**:941-947.
15. CDFA: **Mediterranean fruit fly preventive release program. March 2002.** CDFA, Sacramento, CA, USA; 2002.
16. Andress T, Walters I, del Toro M, Shelly T: **Release-Recapture of Sterile Male Mediterranean Fruit Flies (Diptera: Tephritidae) in Southern California.** *Proceedings of the Hawaiian Entomological Society* 2013, **45**:11-29.
17. Baruffi L, Damiani G, Guglielmino CR, Bandi C, Malacrida AR, Gasperi G: **Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data.** *Heredity (Edinb)* 1995, **74(Pt 4)**:425-437.
18. Gasparich GE, Silva JG, Han HY, McPheron BA, Steck GJ, Shepard WS: **Population genetic structure of Mediterranean fruit fly (Diptera: Tephritidae) and implications for worldwide colonization patterns.** *Annals of the Entomological Society of America* 1997, **90**:790-797.
19. Gomulski LM, Bourtzis K, Brogna S, Morandi PA, Bonvicini C, Sebastiani F, Torti C, Guglielmino CR, Savakis C, Gasperi G, *et al*: **Intron size polymorphism of the Adh1 gene parallels the worldwide colonization history of the Mediterranean fruit fly, *Ceratitis capitata*.** *Molecular Ecology* 1998, **7**:1729-1741.
20. Malacrida AR, Marinoni F, Torti C, Gomulski LM, Sebastiani F, Bonvicini C, Gasperi G, Guglielmino CR: **Genetic aspects of the worldwide colonization process of *Ceratitis capitata*.** *J Hered* 1998, **89(6)**:501-507.
21. Kouri A: **Estimates of heterozygosity and patterns of geographic differentiation in natural populations of the medfly (*Ceratitis capitata*).** *Hereditas* 2002, **137**:173-179.
22. Kouri A: **Estimates of gene flow from rare alleles in natural populations of medfly *Ceratitis capitata* (Diptera: Tephritidae).** *Bulletin of Entomological Research* 2004, **94**:449-456.
23. Kouri A: **Patterns of variation within and between Greek populations of *Ceratitis capitata* suggest extensive gene flow and latitudinal clines.** *Journal of Economic Entomology* 2004, **97**:1186-1190.
24. Bonizzoni M, Guglielmino CR, Smallridge CJ, Gomulski M, Malacrida AR, Gasperi G: **On the origins of medfly invasion and expansion in Australia.** *Mol Ecol* 2004, **13(12)**:3845-3855.
25. Barr NB: **Pathway analysis of *Ceratitis capitata* (Diptera: Tephritidae) using mitochondrial DNA.** *J Econ Entomol* 2009, **102(1)**:401-411.
26. Loukeris TG, Livadaras I, Arcà B, Zabalou S, Savakis C: **Gene transfer into the medfly, *Ceratitis capitata*, with a *Drosophila hydei* transposable element.** *Science* 1995, **270(5244)**:2002-2005.
27. Handler AM, McCombs SD, Fraser MJ, Saul SH: **The lepidopteran transposon vector, *piggyBac*, mediates germ-line transformation in the Mediterranean fruit fly.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:7520-7525.
28. Christophides GK, Livadaras I, Savakis C, Komitopoulou K: **Two medfly promoters that have originated by recent gene duplication drive distinct sex, tissue and temporal expression patterns.** *Genetics* 2000, **156(1)**:173-182.
29. Michel K, Stamenova A, Pinkerton AC, Franz G, Robinson AS, Cariou-Papaleixou A, Zacharopoulou A, O'Brochta DA, Atkinson PW: **Hermes-mediated germ-line transformation of the Mediterranean fruit fly *Ceratitis capitata*.** *Insect Mol Biol* 2001, **10(2)**:155-162.
30. Gong P, Epton MJ, Fu G, Scaife S, Hiscox A, Condon KC, Condon GC, Morrison NI, Kelly DW, Dafa'alla T, *et al*: **A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly.** *Nat Biotechnol* 2005, **23(4)**:453-456.
31. Fu G, Condon KC, Epton MJ, Gong P, Jin L, Condon GC, Morrison NI, Dafa'alla TH, Alphey L: **Female-specific insect lethality engineered using alternative splicing.** *Nat Biotechnol* 2007, **25(3)**:353-357.
32. Scolari F, Schetelig MF, Bertin S, Malacrida AR, Gasperi G, Wimmer EA: **Fluorescent sperm marking to improve the fight against the pest insect *Ceratitis capitata* (Wiedemann; Diptera: Tephritidae).** *N Biotechnol* 2008, **25(1)**:76-84.
33. Schetelig MF, Caceres C, Zacharopoulou A, Franz G, Wimmer EA: **Conditional embryonic lethality to improve the sterile insect technique in *Ceratitis capitata* (Diptera: Tephritidae).** *BMC Biol* 2009, **7**:4.
34. Schetelig MF, Scolari F, Handler AM, Kittelmann S, Gasperi G, Wimmer EA: **Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly *Ceratitis capitata*.** *Proc Natl Acad Sci USA* 2009, **106(43)**:18171-18176.
35. Yuval B, Hendrichs J: **Behavior of flies in the genus *Ceratitis* (Dacinae: Ceratitidini).** In *Fruit flies (Tephritidae): Phylogeny and Evolution of Behavior*. Boca Raton, Florica: CRP Press;Aluja M, Norrbom A 2000:429-457.
36. Gomulski LM, Dimopoulos G, Xi Z, Soares MB, Bonaldo MF, Malacrida AR, Gasperi G: **Gene discovery in an invasive tephritid model pest species, the Mediterranean fruit fly, *Ceratitis capitata*.** *BMC Genomics* 2008, **9**:243.
37. Gomulski LM, Dimopoulos G, Xi Z, Scolari F, Gabrieli P, Siciliano P, Clarke AR, Malacrida AR, Gasperi G: **Transcriptome profiling of sexual maturation and mating in the Mediterranean fruit fly, *Ceratitis capitata*.** *PLoS One* 2012, **7(1)**:e30857.
38. Gabrieli P, Falaguerra A, Siciliano P, Gomulski LM, Scolari F, Zacharopoulou A, Franz G, Malacrida AR, Gasperi G: **Sex and the single embryo: early development in the Mediterranean fruit fly, *Ceratitis capitata*.** *BMC Dev Biol* 2010, **10**:12.
39. Scolari F, Gomulski LM, Ribeiro JM, Siciliano P, Meraldi A, Falchetto M, Bonomi A, Manni M, Gabrieli P, Malovini A, *et al*: **Transcriptional profiles of mating-responsive genes from testes and male accessory glands of the Mediterranean fruit fly, *Ceratitis capitata*.** *PLoS One* 2012, **7(10)**:e46812.
40. Siciliano P, Scolari F, Gomulski LM, Falchetto M, Manni M, Gaoricli P, Field LM, Zhou JJ, Gasperi G, Malacrida AR: **Sniffing out chemosensory genes from the Mediterranean fruit fly, *Ceratitis capitata*.** *PLoS One* 2014, **9(1)**:e85523.
41. Siciliano P, He XL, Woodcock C, Pickett JA, Field LM, Birkett MA, Kalinova B, Gomulski LM, Scolari F, Gasperi G, *et al*: **Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, *Ceratitis capitata*.** *Insect Biochem Mol Biol* 2014, **48**:51-62.
42. Calla B, Hall B, Hou S, Geib SM: **A genomic perspective to assessing quality of mass-reared SIT flies used in Mediterranean fruit fly (*Ceratitis capitata*) eradication in California.** *BMC Genomics* 2014, **15**:98.
43. San Andrés V, Castañera P, Sabater-Muñoz B: **Transcriptome analysis in *Ceratitis capitata* to unveil genes involved in ageing-maturation process.** *Spanish Journal of Agricultural Research* 2013, **11(3)**:842-854.
44. Ogaugwu CE, Schetelig MF, Wimmer EA: **Transgenic sexing system for *Ceratitis capitata* (Diptera: Tephritidae) based on female-specific embryonic lethality.** *Insect Biochem Mol Biol* 2013, **43(1)**:1-8.
45. Salz HK, Erickson JW: **Sex determination in *Drosophila*: The view from the top.** *Fly (Austin)* 2010, **4(1)**:60-70.
46. Graham P, Penn JK, Schedl P: **Masters change, slaves remain.** *Bioessays* 2003, **25(1)**:1-4.
47. Pane A, Salvemini M, Delli Bovi P, Polito C, Saccone G: **The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate.** *Development* 2002, **129(15)**:3715-3725.
48. Saccone G, Pane A, de Simone A, Salvemini M, Milano A, Annunziata L, Mauro U, Polito L: **New sexing strains for Mediterranean fruit fly *Ceratitis capitata*: transforming females into males.** Dordrecht, the Netherlands: Springer;Edited by Vreysen MJB, Robinson AS, Hendrichs J 2007:95-102.
49. Yocum GD, Coudron TA, Brandt SL: **Differential gene expression in *Perillus bioculatus* nymphs fed a suboptimal artificial diet.** *J Insect Physiol* 2006, **52(6)**:586-592.
50. Drew RAI, Yuval B: **The evolution of fruit fly feeding behaviour.** In *Fruit flies (Tephritidae): phylogeny and evolution of behavior*. Boca Raton, Florida: CRC Press;Aluja M, Norrbom AL 2000:731-749.
51. Eberhard W: **Sexual behavior and sexual selection in the Mediterranean fruit fly, *Ceratitis capitata* (Dacinae: Ceratitidini).** In *Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior*. Boca Raton, Florida: CRC Press;Aluja M, Norrbom A 2000:457-489.
52. Yuval B, Maor M, Levy K, Kaspi R, Taylor P, Shelly T: **Breakfast of champions or kiss of death? Survival and sexual performance of protein-fed, sterile Mediterranean fruit flies (Diptera: Tephritidae).** *Florida Entomologist* 2007, **90**:115-122.
53. Pereira R, Yuval B, Liedo P, Teal PEA, Shelly TE, Malavasi DO, Hendrichs J: **Improving sterile male performance in support of programmes**

integrating the sterile insect technique against fruit flies. *Journal of Applied Entomology* 2013, **137(S1)**:178-190.
54. Ben Ami E, Yuval B, Jurkevitch E: **Manipulation of the microbiota of mass-reared Mediterranean fruit flies *Ceratitis capitata* (Diptera: Tephritidae) improves sterile male sexual performance.** *ISME J* 2010, **4(1)**:28-37.
55. Bonilla E, Xu EY: **Identification and characterization of novel mammalian spermatogenic genes conserved from fly to human.** *Mol Hum Reprod* 2008, **14(3)**:137-142.
56. White-Cooper H, Doggett K, Ellis R: **The evolution of spermatogenesis.** In *Sperm biology: an evolutionary perspective.* New York: Academic Press; Birkhead TR, Hosken DJ, Pitnick SS 2009:151-183.
57. Ravi Ram K, Wolfner MF: **Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction.** *Integr Comp Biol* 2007, **47(3)**:427-445.
58. Baldini F, Gabrieli P, Rogers DW, Catteruccia F: **Function and composition of male accessory gland secretions in *Anopheles gambiae*: a comparison with other insect vectors of infectious diseases.** *Parasit Glob Health* 2012, **106(2)**:82-93.
59. Le BV, Klöck C, Schatz A, Nguyen JB, Kakani EG, Catteruccia F, Khosla C, Baxter RH: **Dihydroisoxazole inhibitors of *Anopheles gambiae* seminal transglutaminase AgTG3.** *Malar J* 2014, **13(1)**:210.
60. McInnis DO: **Size differences between normal and irradiated sperm heads in mated female Mediterranean fruit flies (Diptera: Tephritidae).** *Annals of the Entomological Society of America* 1993, **86**:305-308.
61. McInnis DO, Tam S, Grace C, Miyashita D: **Population suppression and sterility rates induced by variable sex ratio, sterile insect releases of *Ceratitis capitata* (Diptera: Tephritidae) in Hawaii.** *Annals of the Entomological Society of America* 1994, **87**:231-240.
62. San Andrés V, Urbaneja A, Sabater-Muñoz B, Castañera P: **A novel molecular approach to assess mating success of sterile *Ceratitis capitata* (Diptera: Tephritidae) males in sterile insect technique programs.** *J Econ Entomol* 2007, **100(4)**:1444-1449.
63. FAO/IAEA/USDA - Food and Agriculture Organisation - IAEA, United States Department of Agriculture: **Manual for product quality control and shipping procedures for sterile mass-reared tephritid fruit flies. Version 5.0.** Vienna, Austria: International Atomic Energy Agency; 2003.
64. Pelosi P, Mastrogiacomo R, Iovinella I, Tuccori E, Persaud KC: **Structure and biotechnological applications of odorant-binding proteins.** *Appl Microbiol Biotechnol* 2014, **98(1)**:61-70.
65. Leal WS: **Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes.** *Annu Rev Entomol* 2013, **58**:373-391.
66. Nirmala X, Schetelig MF, Yu F, Handler AM: **An EST database of the Caribbean fruit fly, *Anastrepha suspensa* (Diptera: Tephritidae).** *Gene* 2013, **517(2)**:212-217.
67. Pavlidi N, Dermauw W, Rombauts S, Chrisargiris A, Van Leeuwen T, Vontas J: **Analysis of the Olive fruit fly *Bactrocera oleae* transcriptome and phylogenetic classification of the major detoxification gene families.** *PLoS One* 2013, **8(6)**:e66533.
68. Schwarz D, Robertson HM, Feder JL, Varala K, Hudson ME, Raglanc GJ, Hahn DA, Berlocher SH: **Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*.** *BMC Genomics* 2009, **10**:633.
69. Zheng W, Peng T, He W, Zhang H: **High-throughput sequencing to reveal genes involved in reproduction and development in *Bactrocera dorsalis* (Diptera: Tephritidae).** *PLoS One* 2012, **7(5)**:e36463.
70. Hsu JC, Chien TY, Hu CC, Chen MJ, Wu WJ, Feng HT, Haymer DS, Chen CY: **Discovery of genes related to insecticide resistance in *Bactrocera dorsalis* by functional genomic analysis of a de novo assembled transcriptome.** *PLoS One* 2012, **7(8)**:e40950.
71. Ramsdell KM, Lyons-Sobaski SA, Robertson HM, Walden KK, Feder JL, Wanner K, Berlocher SH: **Expressed sequence tags from cephalic chemosensory organs of the northern walnut husk fly, *Rhagoletis suavis*, including a putative canonical odorant receptor.** *J Insect Sci* 2010, **10**:51.
72. Davies SJ, Chapman T: **Identification of genes expressed in the accessory glands of male Mediterranean Fruit Flies (*Ceratitis capitata*).** *Insect Biochem Mol Biol* 2006, **36(11)**:846-856.

doi:10.1186/1471-2156-15-S2-S11
Cite this article as: Scolari *et al.*: How functional genomics will impact fruit fly pest control: the example of the Mediterranean fruit fly, *Ceratitis capitata*. *BMC Genetics* 2014 15(Suppl 2):S11.

References

Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic acids research. 2010;38:W7-13.

Abd-Alla AMM, Parker AG, Vreysen MJB, Bergoin M. Tsetse Salivary Gland Hypertrophy Virus: Hope or Hindrance for Tsetse Control? PLoS Neglected Tropical Diseases. 2011;5:e1220.

Adhami J, Reiter P. Introduction and establishment of Aedes (Stegomyia) albopictus skuse (Diptera: Culicidae) in Albania. J Am Mosq Control Assoc. 1998;14:340-343.

Aksoy S, IGGI. A proposal for tsetse fly (Glossina) genome projects (white paper). 2010.

Almeida FC, Desalle R. Evidence of adaptive evolution of accessory gland proteins in closely related species of the Drosophila repleta group. Mol Biol Evol. 2008;25:2043-2053.

Alphey L, Benedict M, Bellini R, Clark GG, Dame DA, Service MW, Dobson SL. Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis. Vector Borne and Zoonotic Diseases. 2010;10:295-311.

Andres JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. Molecular evolution of seminal proteins in field crickets. Mol Biol Evol. 2006;23:1574-1584.

APHIS. Risk assessment, mediterranean fruit fly. Washington D.C.: Planning and Risk Analysis Systems. Policy and Program Development. Animal and Plant Health Inspection Service USDA. 1992.

Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F, *et al.* Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics. Science (New York, NY). 2010;330:86-88.

Arita LH, Kaneshiro KY. Sexual selection and lek behavior in the Mediterranean fruit fly, Ceratitis capitata (Diptera: Tephritidae). Pacific Science. 1989;43:8.

Arya GH, Weber AL, Wang P, Magwire MM, Negron YL, Mackay TF, Anholt RR. Natural variation, functional pleiotropy and transcriptional contexts of odorant binding protein genes in Drosophila melanogaster. Genetics. 2010;186:1475-1485.

Atkins MD. Introduction to Insect Behavior. Mac Millan Publishing CoNew York. 1980:1.

Attardo GM, Abila PP, Auma JE, Baumann AA, Benoit JB, Brelsfoard CL, Ribeiro JMC, Cotton JA, Pham DQD, Darby AC, *et al.* Genome Sequence of the Tsetse Fly Glossina morsitans: Vector of African Trypanosomiasis. Science (New York, NY). 2014;344:380-386.

Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. Insect seminal fluid proteins: identification and function. Annu Rev Entomol. 2011;56.

Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. Insect Seminal Fluid Proteins: Identification and Function. Annual review of entomology. 2011;56:21-40.

Baldacchino F, Caputo B, Chandre F, Drago A, della Torre A, Montarsi F, Rizzoli A. Control methods against invasive Aedes mosquitoes in Europe: a review. Pest Manag Sci. 2015;71:1471-1485.

Baldocchi RA, Flaherty L. Isolation of genomic fragments from polymorphic regions by representational difference analysis. Methods. 1997;13:337-346.

Barrett MP, Burchmore RJ, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, Krishna S. The trypanosomiases. Lancet (London, England). 2003;362:1469-1480.

Battistuzzi FU, Filipski A, Hedges SB, Kumar S. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. Mol Biol Evol. 2010;27:1289-1300.

Baumgartner S, Noll M. Network of interactions among pair-rule genes regulating paired expression during primordial segmentation of Drosophila. Mech Dev. 1990;33:1-18.

Begun DJ, Lindfors HA, Thompson ME, Holloway AK. Recently evolved genes identified from Drosophila yakuba and D. erecta accessory gland expressed sequence tags. Genetics. 2006;172:1675-1681.

Bellini R, Calvitti M, Medici A, Carrieri M, Celli G, Maini S. Use of the Sterile Insect Technique Against Aedes albopictus in Italy: First Results of a Pilot Trial. In: Vreysen MJB, Robinson AS, Hendrichs J, editors. Area-Wide Control of Insect Pests: From Research to Field Implementation. Dordrecht: Springer Netherlands; 2007. p. 505-515.

Bellini R, Medici A, Puggioli A, Balestrino F, Carrieri M. Pilot field trials with Aedes albopictus irradiated sterile males in Italian urban areas. J Med Entomol. 2013;50:317-325.

Benoit JB, Attardo GM, Baumann AA, Michalkova V, Aksoy S. Adenotrophic viviparity in tsetse flies: potential for population control and as an insect model for lactation. Annual review of entomology. 2015;60:351-371.

Benoit JB, Attardo GM, Michalkova V, Krause TB, Bohova J, Zhang Q, Baumann AA, Mireji PO, Takac P, Denlinger DL, *et al.* A novel highly divergent protein family identified from a viviparous insect by RNA-seq analysis: a potential target for tsetse fly-specific abortifacients. PLoS genetics. 2014;10:e1003874.

Bernardo Carvalho A, Koerich LB, Clark AG. Origin and evolution of Y chromosomes: Drosophila tales. Trends in genetics : TIG. 2009;25:270-277.

Black WCt, Rai KS. Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. Genetical research. 1988;51:185-196.

Bohbot J, Pitts RJ, Kwon HW, Rutzler M, Robertson HM, Zwiebel LJ. Molecular characterization of the Aedes aegypti odorant receptor gene family. Insect Mol Biol. 2007;16:525-537.

Bohbot JD, Sparks JT, Dickens JC. The maxillary palp of Aedes aegypti, a model of multisensory integration. Insect biochemistry and molecular biology. 2014;48:29-39.

Bonilla E, Xu EY. Identification and characterization of novel mammalian spermatogenic genes conserved from fly to human. Molecular human reproduction. 2008;14:137-142.

Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species Aedes albopictus: current knowledge and future perspectives. Trends Parasitol. 2013;29:460-468.

Bonizzoni M, Gomulski LM, Bertin S, Scolari F, Guglielmino CR, Yuval B. Unfaithful Mediterranean fruit fly Ceratitis capitata females: Impact on the SIT? In: Vreysen MJB, Robinson AS, Hendrichs J, editors. Area-Wide Control of Insect Pests. Netherlands: Springer; 2007.

Bonizzoni M, Katsoyannos BI, Marguerie R, Guglielmino CR, Gasperi G, Malacrida A. Microsatellite analysis reveals remating by wild Mediterranean fruit fly females, Ceratitis capitata. Mol Ecol. 2002;11.

Bonomi A, Bassetti F, Gabrieli P, Beadell J, Falchetto M, Scolari F, Gomulski LM, Regazzini E, Ouma JO, Caccone A, *et al.* Polyandry is a common event in wild populations of the Tsetse fly Glossina fuscipes fuscipes and may impact population reduction measures. PLoS Negl Trop Dis. 2011;5:e1190.

Bourtzis K, Dobson SL, Braig HR, O'Neill SL. Rescuing Wolbachia have been overlooked. Nature. 1998;391:852-853.

Brito NF, Moreira MF, Melo ACA. A look inside odorant-binding proteins in insect chemoreception. Journal of insect physiology. 2016;95:51-65.

Brun R, Blum J, Chappuis F, Burri C. Human African trypanosomiasis. Lancet (London, England). 2010;375:148-159.

Budd LT. DFID-funded Tsetse and Trypanosome Research and Development Since 1980: Economic analysis: DFID; 1999.

Carey JR. Establishment of the Mediterranean fruit fly in California. Science (New York, NY). 1991;253:1369-1373.

Carey JR. The Incipient Mediterranean Fruit Fly Population in California: Implications for Invasion Biology. Ecology. 1996;77:1690-1697.

Catteruccia F, Nolan T, Loukeris TG, Blass C, Savakis C, Kafatos FC, Crisanti A. Stable germline transformation of the malaria mosquito Anopheles stephensi. Nature. 2000;405:959-962.

Cayol JP, Causse R, Louis C, Barthes J. Medfly Ceratitis capitata Wiedemann (Dipt., Trypetidae) as a rot vector in laboratory conditions. Journal of Applied Entomology. 1994;117:338-343.

Chang Y. Representational difference analysis. Current protocols in molecular biology. 2002;Chapter 25:Unit 25B.27.

Chapman T. The soup in my fly: evolution, form and function of seminal fluid proteins. PLoS Biol. 2008;6.

Chapman T, Miyatake T, Smith HK, Partridge L. Interactions of mating, egg production and death rates in females of the Mediterranean fruit fly, Ceratitis capitata. Proceedings of the Royal Society B: Biological Sciences. 1998;265:1879-1894.

Chen S, Zhang YE, Long M. New genes in Drosophila quickly become essential. Science (New York, NY). 2010;330:1682-1685.

Chen X-G, Jiang X, Gu J, Xu M, Wu Y, Deng Y, Zhang C, Bonizzoni M, Dermauw W, Vontas J, *et al.* Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. Proceedings of the National Academy of Sciences. 2015;112:E5907-E5915.

Clausen PH, Adeyemi I, Bauer B, Breloeer M, Salchow F, Staak C. Host preferences of tsetse (Diptera: Glossinidae) based on bloodmeal identifications. Med Vet Entomol. 1998;12:169-180.

Coates CJ, Jasinskiene N, Miyashiro L, James AA. Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti. Proceedings of the National Academy of Sciences of the United States of America. 1998;95:3748-3751.

Cockerell TDA. A Fossil Tsetse-fly in Colorado. Nature. 1907;76:1.

Consortium iK. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. Journal of Heredity. 2013;104:595-600.

Coulthart MB, Singh RS. Low genic variation in male-reproductive-tract proteins of Drosophila melanogaster and D. simulans. Mol Biol Evol. 1988;5:167-181.

Damberger FF, Ishida Y, Leal WS, Wüthrich K. Structural Basis of Ligand Binding and Release in Insect Pheromone-binding Proteins: NMR Structure of Antheraea polyphemus PBP1 at pH 4.5. Journal of Molecular Biology. 2007;373:811-819.

Delatte H, Bagny L, Brengue C, Bouetard A, Paupy C, Fontenille D. The invaders: phylogeography of dengue and chikungunya viruses Aedes vectors, on the South West islands of the Indian Ocean. Infect Genet Evol. 2011;11:1769-1781.

Delatte H, Desvars A, Bouetard A, Bord S, Gimonneau G, Vourc'h G, Fontenille D. Blood-feeding behavior of Aedes albopictus, a vector of Chikungunya on La Reunion. Vector Borne Zoonotic Dis. 2010;10:249-258.

Dias J.A. TS. Contribuição para o estudo da sistemática do género Glossina Wiedemann, 1830 (Insecta, Brachycera, Cyclorrhapha, Glossinidae) Proposta para a criação de um novo subgénero. Garcia Orta Sér Zool Lisboa. 1987;14:11.

Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, Kafatos FC. Genome expression analysis of Anopheles gambiae: responses to injury, bacterial challenge, and malaria infection. Proceedings of the National Academy of Sciences of the United States of America. 2002;99:8814-8819.

Ding Y, Zhao L, Yang S, Jiang Y, Chen Y, Zhao R, Zhang Y, Zhang G, Dong Y, Yu H, *et al.* A young Drosophila duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. PLoS genetics. 2010;6:e1001255.

Dritsou V, Topalis P, Windbichler N, Simoni A, Hall A, Lawson D, Hinsley M, Hughes D, Napolioni V, Crucianelli F, *et al.* A draft genome sequence of an invasive mosquito: an Italian Aedes albopictus. Pathogens and global health. 2015;109:207-220.

Dujardin JP, Schofield CJ. Triatominae: systematics, morphology and population biology. Maudlin, I; Holmes, P; MILES, MA, (eds) The Trypanosomiases CAB. 2004.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics (Oxford, England). 2005;21:3439-3440.

Dyck VA, Hendrichs J, Robinson AS. Sterile insect technique - principles and practice in area wide integrated pest management. Dordrecht, NL: Springer. 2005:787.

Dyer NA, Lawton SP, Ravel S, Choi KS, Lehane MJ, Robinson AS, Okedi LM, Hall MJ, Solano P, Donnelly MJ. Molecular phylogenetics of tsetse flies (Diptera: Glossinidae) based on mitochondrial (COI, 16S, ND2) and nuclear ribosomal DNA sequences, with an emphasis on the palpalis group. Molecular phylogenetics and evolution. 2008;49:227-239.

Eberhard W. Sexual behavior and sexual selection in the Mediterranean fruit fly, Ceratitis capitata (Dacinae: Ceratitidini). In Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior. Edited by Aluja M, Norrbom A Boca Raton, Florida: CRC Press. 2000:32.

Eberhard WG. Sexual selection and animal genitalia. Harvard Univ Press, Cambridge, MA, 1985. 1985.

Ellis LL, Huang W, Quinn AM, Ahuja A, Alfrejd B, Gomez FE, Hjelmen CE, Moore KL, Mackay TFC, Johnston JS, *et al.* Intrapopulation genome size variation in D.

melanogaster reflects life history variation and plasticity. PLoS genetics. 2014;10:e1004522.

Enserink M. Entomology. A mosquito goes global. Science (New York, NY). 2008;320:864-866.

Eritja R, Escosa R, Lucientes J, Marquès E, Roiz D, Ruiz S. Worldwide invasion of vector mosquitoes: present European distribution and challenges for Spain. Biological Invasions. 2005;7:87-97.

Fang J. Ecology: A world without mosquitoes. Nature. 2010;466:432-434.

Faraji A, Egizi A, Fonseca DM, Unlu I, Crepeau T, Healy SP, Gaugler R. Comparative host feeding patterns of the Asian tiger mosquito, Aedes albopictus, in urban and suburban Northeastern USA and implications for disease transmission. PLoS Negl Trop Dis. 2014;8:e3037.

Farikou O, Njiokou F, Mbida Mbida JA, Njitchouang GR, Djeunga HN, Asonganyi T, Simarro PP, Cuny G, Geiger A. Tripartite interactions between tsetse flies, Sodalis glossinidius and trypanosomes—An epidemiological approach in two historical human African trypanosomiasis foci in Cameroon. Infection, Genetics and Evolution. 2010;10:115-121.

Ferdig MT, Taft AS, Severson DW, Christensen BM. Development of a comparative genetic linkage map for Armigeres subalbatus using Aedes aegypti RFLP markers. Genome research. 1998;8:41-47.

Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome research. 2009;19:886-896.

Findlay GD, Yi X, Maccoss MJ, Swanson WJ. Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS biology. 2008;6:e178.

Fonseca DM, Unlu I, Crepeau T, Farajollahi A, Healy SP, Bartlett-Healy K, Strickman D, Gaugler R, Hamilton G, Kline D, *et al.* Area-wide management of Aedes albopictus. Part 2: gauging the efficacy of traditional integrated pest control measures against urban container mosquitoes. Pest Manag Sci. 2013;69:1351-1361.

Gabrieli P, Gomulski LM, Bonomi A, Siciliano P, Scolari F, Franz G, Jessup A, Malacrida AR, Gasperi G. Interchromosomal duplications on the Bactrocera oleae Y chromosome imply a distinct evolutionary origin of the sex chromosomes compared to Drosophila. PloS one. 2011;6:e17747.

Gasperi G, Bellini R, Malacrida AR, Crisanti A, Dottori M, Aksoy S. A new threat looming over the Mediterranean basin: emergence of viral diseases transmitted by Aedes albopictus mosquitoes. PLoS Negl Trop Dis. 2012;6:e1836.

Gavriel S, Gazit Y, Yuval B. Remating by female Mediterranean fruit flies (Ceratitis capitata, Diptera: Tephritidae): temporal patterns and modulation by male condition. J Insect Physiol. 2009;55.

Giangaspero A, Marangi M, Latrofa MS, Martinelli D, Traversa D, Otranto D, Genchi C. Evidences of increasing risk of dirofilarioses in southern Italy. Parasitol Res. 2013;112:1357-1361.

Gibson W, Stevens J. Genetic exchange in the trypanosomatidae. Advances in parasitology. 1999;43:1-46.

Gomulski LM, Bourtzis K, Brogna S, Morandi PA, Bonvicini C, Sebastiani F, Torti C, Guglielmino CR, Savakis C, Gasperi G, *et al.* Intron size polymorphism of the Adh1

gene parallels the worldwide colonization history of the Mediterranean fruit fly, Ceratitis capitata. Molecular Ecology. 1998;7:1729-1741.

Gomulski LM, Dimopoulos G, Xi Z, Scolari F, Gabrieli P, Siciliano P. Transcriptome profiling of sexual maturation and mating in the Mediterranean fruit fly, Ceratitis capitata. PLoS One. 2012;7.

Gong D-P, Zhang H-J, Zhao P, Xia Q-Y, Xiang Z-H. The Odorant Binding Protein Gene Family from the Genome of Silkworm, Bombyx mori. BMC Genomics. 2009;10:1-14.

Gratz NG. Critical review of the vector status of Aedes albopictus. Med Vet Entomol. 2004;18:215-227.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, *et al.* Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics. 2007;177:1321-1335.

Hagen KS, Allen WW, Tassan RL. Mediterranean fruit fly: The worst may be yet to come. Calif Agric. 1981;36:2.

Hall AB, Qi Y, Timoshevskiy V, Sharakhova MV, Sharakhov IV, Tu Z. Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. BMC Genomics. 2013;14:273.

Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. Adaptive evolution of young gene duplicates in mammals. Genome research. 2009;19:859-867.

Hansson Bill S, Stensmyr Marcus C. Evolution of Insect Olfaction. Neuron. 2011;72:698-711.

Harris EJ, Carey JR. Laboratory Studies of the Mediterranean Fruit Fly (Diptera: Tephritidae) in Coffee. Environmental Entomology. 1989;18:103-110.

Harris EJ, Lee CY. Development of Ceratitis capitata (Diptera: Tephritidae) in Coffee in Wet and Dry Habitats. Environmental Entomology. 1989;18:1042-1049.

Hawley WA. The biology of Aedes albopictus. Journal of the American Mosquito Control Association Supplement. 1988;1:1-39.

Heath RR, Landolt PJ, Tumlinson JH, Chambers DL, Murphy RE, Doolittle RE, Dueben BD, Sivinski J, Calkins CO. Analysis, synthesis, formulation, and field testing of three major components of male mediterranean fruit fly pheromone. Journal of chemical ecology. 1991;17:1925-1940.

Hendrichs J, Robinson AS, Cayol JP, Enkerlin W. MEDFLY AREAWIDE STERILE INSECT TECHNIQUE PROGRAMMES FOR PREVENTION, SUPPRESSION OR ERADICATION: THE IMPORTANCE OF MATING BEHAVIOR STUDIES. Florida Entomologist. 2002;85:1-13.

Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Bio/technology (Nature Publishing Company). 1993;11:1026-1030.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. How many species are infected with Wolbachia?--A statistical analysis of current data. FEMS microbiology letters. 2008;281:215-220.

Hill CA, Kafatos FC, Stansfield SK, Collins FH. Arthropod-borne diseases: vector control in the genomics era. Nature reviews Microbiology. 2005;3:262-268.

Holder P, George S, Disbury M, Singe M, Kean JM, Mcfadden A. A Biosecurity Response to Aedes albopictus (Diptera: Culicidae) in Auckland, New Zealand. Journal of Medical Entomology. 2010;47:600-609.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, *et al.* The genome sequence of the malaria mosquito Anopheles gambiae. Science (New York, NY). 2002;298:129-149.

Hoppenheit A, Murugaiyan J, Bauer B, Steuber S, Clausen P-H, Roesler U. Identification of Tsetse (<italic>Glossina</italic> spp.) Using Matrix-Assisted Laser Desorption/Ionisation Time of Flight Mass Spectrometry. PLoS Negl Trop Dis. 2013;7:e2305.

Horn DJ. Biology of insects. Philadelphia : Saunders. 1976:1.

Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends in genetics : TIG. 2002;18:486.

Itard J. L'appareil reproducteur mâle des Glossines (Diptera-Muscidae) Les étapes de sa formation chez la pupe. La spermatogénèse. Rev Élev Méd vét Pays trop 1970;23:24.

Jang EB. Fruit flies and their impact on agriculture in Hawaii. 2007;39:2.

Jang EB. Physiology of mating behavior in Mediterranean fruit fly (Diptera:Tephritidae): chemoreception and male accessory gland fluids in female post-mating behavior. Florida Entomologist. 2002;85:89-93.

Jang EB, McInnis DO, Lance DR, Carvalho LA. Mating-Induced Changes in Olfactory-Mediated Behavior of Laboratory-Reared Normal, Sterile, and Wild Female Mediterranean Fruit Flies (Diptera: Tephritidae) Mated to Conspecific Males. Annals of the Entomological Society of America. 1998;91:139-144.

Jasinskiene N, Coates CJ, Benedict MQ, Cornel AJ, Rafferty CS, James AA, Collins FH. Stable transformation of the yellow fever mosquito, Aedes aegypti, with the Hermes element from the housefly. Proceedings of the National Academy of Sciences of the United States of America. 1998;95:3743-3747.

Jeffares DC, Tomiczek B, Sojo V, dos Reis M. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. Methods in molecular biology (Clifton, NJ). 2015;1201:65-90.

Junqueira AC, Azeredo-Espin AM, Paulo DF, Marinho MA, Tomsho LP, Drautz-Moses DI, Purbojati RW, Ratan A, Schuster SC. Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. Sci Rep. 2016;6:21762.

Kapongo JPK, P.G; Giliomee, J.H. Control of Mediterranean Fruit Fly Ceratitis capitata (Diptera: Tephritidae) with the Parasitoid Muscidifurax raptor (Hymenoptera: Pteromalidae) in Vineyards. HortScience. 2007;42:1400-1404.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30.

Klassen W. Area-Wide Integrated Pest Management and the Sterile Insect Technique. In: Dyck VA, Hendrichs J, Robinson AS, editors. Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management. Dordrecht: Springer Netherlands; 2005. p. 39-68.

Klowden MJ. CHAPTER 4 - Reproductive Systems. Physiological Systems in Insects (Second Edition). San Diego: Academic Press; 2008. p. 181-238.

Kraaijeveld K, Katsoyannos BI, Stavrinides M, Kouloussis NA, Chapman T. Remating in wild females of the Mediterranean fruit fly, Ceratitis capitata. Anim Behav. 2005;69.

Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. eLife. 2015;4:e08347.

Krafsur ES. Tsetse flies: genetics, evolution, and role as vectors. Infect Genet Evol. 2009;9:124-141.

Laflamme BA, Wolfner MF. Identification and function of proteolysis regulators in seminal fluid. Mol Reprod Dev. 2013;80.

Lambrechts L, Scott TW, Gubler DJ. Consequences of the expanding global distribution of Aedes albopictus for dengue virus transmission. PLoS Negl Trop Dis. 2010;4:e646.

Lance DR, McInnis DO. Biological Basis of the Sterile Insect Technique. In: Dyck VA, Hendrichs J, Robinson AS, editors. Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management. Dordrecht: Springer Netherlands; 2005. p. 69-94.

Lance DR, McInnis DO, Rendon P, Jackson CG. Courtship Among Sterile and Wild Ceratitis capitata (Diptera: Tephritidae) in Field Cages in Hawaii and Guatemala. Annals of the Entomological Society of America. 2000;93:1179-1185.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H. Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England). 2007;23.

Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. Neuron. 2004;43:703-714.

Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics (Oxford, England). 2009;25:2286-2288.

Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, *et al.* VectorBase: a data resource for invertebrate vector genomics. Nucleic acids research. 2009;37:D583-D587.

Le BV, Klöck C, Schatz A, Nguyen JB, Kakani EG, Catteruccia F, Khosla C, Baxter RH. Dihydroisoxazole inhibitors of Anopheles gambiae seminal transglutaminase AgTG3. Malaria Journal. 2014;13:1-12.

Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25.

Leal WS. An acute olfactory system is essential for reproduction: the Raison d'etre for adult insects. Proceedings of the American Philosophical Society. 2012;156:295-302.

Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annual review of entomology. 2013;58:373-391.

Lehane MJ, Msangi AR, Whitaker CJ, Lehane SM. Grouping of trypanosome species in mixed infections in Glossina pallidipes. Parasitology. 2000;120 (Pt 6):583-592.

Levinson HZ, Levinson AR, Schäfer K. Pheromone biology of the Mediterranean fruit fly (Ceratitis capitata Wied.) with emphasis on the functional anatomy of the pheromone glands and antennae as well as mating behaviour. Journal of Applied Entomology. 1987;104:448-461.

Li L, Li P, Xue L. The RED domain of Paired is specifically required for Drosophila accessory gland maturation. Open Biology. 2015;5:140179.

Li S, Picimbon JF, Ji S, Kan Y, Chuanling Q, Zhou JJ, Pelosi P. Multiple functions of an odorant-binding protein in the mosquito Aedes aegypti. Biochemical and biophysical research communications. 2008;372:464-468.

Liquido NJ, Shinoda LA, Cunningham RT, America ESo. Host Plants of the Mediterranean Fruit Fly (Diptera: Tephritidae) : an Annotated World Review: Entomological Society of America; 1991.

Loukeris TG, Livadaras I, Arca B, Zabalou S, Savakis C. Gene transfer into the medfly, Ceratitis capitata, with a Drosophila hydei transposable element. Science. 1995;270.

Lowe S, Browne M, Boudjelas S, Poorter M. 100 of the world's worst invasive alien species: a selection from the Global Invasive Species Database. . IUCN/SSC Invasive Species Specialist Group, Auckland. 2000;12.

Loytynoja A, Goldman N. A model of evolution and structure for multiple sequence alignment. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2008;363:3913-3919.

Mackay AJ, Amador M, Barrera R. An improved autocidal gravid ovitrap for the control and surveillance of Aedes aegypti. Parasit Vectors. 2013;6:225.

Madakacherry O, Lees RS, Gilles JR. Aedes albopictus (Skuse) males in laboratory and semi-field cages: release ratios and mating competitiveness. Acta Trop. 2014;132 Suppl:S124-129.

Maddison P, Bartlett B. A contribution towards the zoogeography of the Tephritidae. World Crop Pests. 1989;3A:8.

Malacrida AR, Gomulski LM, Bonizzoni M, Bertin S, Gasperi G, Guglielmino CR. Globalization and fruitfly invasion and expansion: the medfly paradigm. Genetica. 2007;131:1-9.

Malacrida AR, Guglielmino CR, Gasperi G, Baruffi L, Milani R. Spatial and temporal differentiation in colonizing populations of Ceratitis capitata. Heredity. 1992;69:101-111.

Malacrida AR, Marinoni F, Torti C, Gomulski LM, Sebastiani F, Bonvicini C, Gasperi G, Guglielmino CR. Genetic aspects of the worldwide colonization process of Ceratitis capitata. The Journal of heredity. 1998;89:501-507.

Manni M, Gomulski LM, Aketarawong N, Tait G, Scolari F, Somboon P, Guglielmino CR, Malacrida AR, Gasperi G. Molecular markers for analyses of intraspecific genetic diversity in the Asian Tiger mosquito, Aedes albopictus. Parasit Vectors. 2015;8:188.

Manoharan M, Ng Fuk Chong M, Vaitinadapoule A, Frumence E, Sowdhamini R, Offmann B. Comparative genomics of odorant binding proteins in Anopheles gambiae, Aedes aegypti, and Culex quinquefasciatus. Genome Biol Evol. 2013;5:163-180.

Marchini D, Del Bene G, Falso LF, Dallai R. Structural organization of the copulation site in the medfly Ceratitis capitata (Diptera: Tephritidae) and observations on sperm transfer and storage. Arthropod structure & development. 2001;30:39-54.

Marin I, Baker BS. The evolutionary dynamics of sex determination. Science (New York, NY). 1998;281:1990-1994.

Matthews EG, Kitching RL. Insect Ecology. University Quesland PressAsutralia. 1984:1.

McClelland G. Sex-linkage in Aedes aegypti. Trans Roy Soc Trop Med Hyg. 1962;56:1.

McGraw LA, Suarez SS, Wolfner MF. On a matter of seminal importance: The emerging influence of seminal plasma components on fertility and future progeny. BioEssays : news and reviews in molecular, cellular and developmental biology. 2015;37:142-147.

Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, Van Bortel W. A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. Vector Borne Zoonotic Dis. 2012;12:435-447.

Minard G, Tran FH, Raharimalala FN, Hellard E, Ravelonandro P, Mavingui P, Valiente Moro C. Prevalence, genomic and metabolic profiles of Acinetobacter and Asaia associated with field-caught Aedes albopictus from Madagascar. FEMS Microbiol Ecol. 2013;83:63-73.

Molyneux D. Host-trypanosomes interactions in Glossina. Insect Science and its Application. 1980;1:7.

Motloang M, Masumu J, Mans B, Van den Bossche P, Latif A. Vector competence of Glossina austeni and Glossina brevipalpis for Trypanosoma congolense in KwaZulu-Natal, South Africa. The Onderstepoort journal of veterinary research. 2012;79:E1-6.

Müller L, Grath S, von Heckel K, Parsch J. Inter- and Intraspecific Variation in Drosophila Genes with Sex-Biased Expression. International Journal of Evolutionary Biology. 2012;2012:963976.

Munks RJ, Sant'Anna MR, Grail W, Gibson W, Igglesden T, Yoshiyama M, Lehane SM, Lehane MJ. Antioxidant gene expression in the blood-feeding fly Glossina morsitans morsitans. Insect Mol Biol. 2005;14:483-491.

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu Z, Loftus B, Xi Z, Megy K, Grabherr M, *et al.* Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science (New York, NY). 2007;316:1718-1723.

Neuhaus EM, Gisselmann G, Zhang W, Dooley R, Stortkuhl K, Hatt H. Odorant receptor heterodimerization in the olfactory system of Drosophila melanogaster. Nat Neurosci. 2005;8:15-17.

Newton ME, Southern DI, Wood RJ. X and Y chromosomes of Aedes aegypti (L.) distinguished by Giemsa C-banding. Chromosoma. 1974;49:41-49.

Nogge G. Aposymbiotic tsetse flies, Glossina morsitans morsitans obtained by feeding on rabbits immunized specifically with symbionts. Journal of insect physiology. 1978;24:299-304.

Nogge G, Gerresheim A. Experiments on the elimination of symbionts from the tsetse fly, Glossina morsitans morsitans (Diptera: Glossinidae), by antibiotics and lysozyme. Journal of Invertebrate Pathology. 1982;40:166-179.

Noll M. Evolution and role of Pax genes. Current opinion in genetics & development. 1993;3:595-605.

Nurminsky DI, Nurminskaya MV, Aguiar DD, Hartl DL. Selective sweep of a newly evolved sperm-specific gene in Drosophila. Nature. 1998;396:572-575.

Odhiambo TR, Kokwaro, E. D. & Sequeira, L. M. Histochemical and ultrastructural studies of the male accessory reproductive glands and spermatophore of the tsetse, Glossina morsitans morsitans Westwood. . Insect Sci Appl. 1983;4:9.

Ogaugwu CE, Schetelig MF, Wimmer EA. Transgenic sexing system for Ceratitis capitata (Diptera: Tephritidae) based on female-specific embryonic lethality. Insect biochemistry and molecular biology. 2013;43:1-8.

Pais R, Lohs C, Wu Y, Wang J, Aksoy S. The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly. Applied and environmental microbiology. 2008;74:5965-5974.

Papadopoulos NT, Plant RE, Carey JR. From trickle to flood: the large-scale, cryptic invasion of California by tropical fruit flies. Proceedings Biological sciences / The Royal Society. 2013;280:20131466.

Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K, Castañera P, Cavanaugh JP, Chao H, Childers C, *et al.* The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome Biology. 2016;17:1-31.

Parker GA. Sperm competition and its evolutionary consequences in the insects. Biological Reviews. 1970;45:525-567.

Paupy C, Delatte H, Bagny L, Corbel V, Fontenille D. Aedes albopictus, an arbovirus vector: from the darkness to the light. Microbes Infect. 2009;11:1177-1185.

Peacock L, Ferris V, Bailey M, Gibson W. Dynamics of infection and competition between two strains of Trypanosoma brucei brucei in the tsetse fly observed using fluorescent markers. Kinetoplastid biology and disease. 2007;6:4.

Pelosi P, Zhou JJ, Ban LP, Calvello M. Soluble proteins in insect chemical communication. Cellular and Molecular Life Sciences CMLS. 2006;63:1658-1676.

Perry JC, Sirot L, Wigby S. The seminal symphony: how to compose an ejaculate. Trends Ecol Evol. 2013;28.

Petersen FT, Meier R, Kutty SN, Wiegmann BM. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. Molecular phylogenetics and evolution. 2007;45:111-122.

Pianka ER. On r and K selection. Amer Naturalist. 1970;104:7.

Pietrobelli M. Importance of Aedes albopictus in veterinary medicine. Parassitologia. 2008;50:113-115.

Pimley RW, Langley PA. Hormone stimulated lipolysis and proline synthesis in the fat body of the adult tsetse fly, Glossina morsitans. Journal of insect physiology. 1982;28:781-789.

Poiani A. Complexity of seminal fluid: a review. Behav Ecol Sociobiol. 2006;60.

Pollock JN. Sperm Transfer by Spermatophores in Glossina austeni Newstead. Nature. 1970;225:1063-1064.

Prokopy RJ, Hendrichs J. Mating Behavior of Ceratitis capitata on a Field-Caged Host Tree. Annals of the Entomological Society of America. 1979;72:642-648.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic acids research. 2005;33:W116-120.

R Development Core Team. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. 2009.

Rai K. A comparative study of mosquito karyotypes. Ann Ent Soc Am. 1963;56:10.

Rai KS, Black WCt. Mosquito genomes: structure, organization, and evolution. Advances in genetics. 1999;41:1-33.

Ram KR, Wolfner MF. Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. PLoS genetics. 2007;3:e238.

Ramasamy S, Ometto L, Crava CM, Revadi S, Kaur R, Horner D, Pisani D, Dekker T, Anfora G, Rota-Stabelli O. The evolution of olfactory gene families in Drosophila and the genomic basis of chemical-ecological adaptation in Drosophila suzukii. Genome Biology and Evolution. 2016.

Ranz JM, Parsch J. Newly evolved genes: moving from comparative genomics to functional studies in model systems. How important is genetic novelty for species adaptation and diversification? BioEssays : news and reviews in molecular, cellular and developmental biology. 2012;34:477-483.

Ravi Ram K, Wolfner MF. Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. Integrative and comparative biology. 2007;47:427-445.

Regis L, Monteiro AM, Melo-Santos MA, SilveiraJr JC, Furtado AF, Acioli RV, Santos GM, Nakazawa M, Carvalho MS, Ribeiro Jr PJ, *et al.* Developing new approaches for detecting and preventing Aedes aegypti population outbreaks: basis for surveillance, alert and control system. Memorias do Instituto Oswaldo Cruz. 2008;103:50-59.

Reznick D, Bryant MJ, Bashey F. r- and k- selection revisited: the role of population regulation in life-history evolution. Ecology. 2002;83:1509-1520.

Roditi I, Lehane MJ. Interactions between trypanosomes and tsetse flies. Current opinion in microbiology. 2008;11:345-351.

Ronderos DS, Smith DP. Activation of the T1 neuronal circuit is necessary and sufficient to induce sexually dimorphic mating behavior in Drosophila melanogaster. J Neurosci. 2010;30:2595-2599.

Sabatini A, Raineri V, Trovato G, Coluzzi M. [Aedes albopictus in Italy and possible diffusion of the species into the Mediterranean area]. Parassitologia. 1990;32:301-304.

Saunders DS, Dodd CWH. Mating, insemination, and ovulation in the tsetse fly, Glossina morsitans. Journal of insect physiology. 1972;18:187-198.

Scalera R, Genovesi P, Essl F, Rabitsch W. The impacts of invasive alien species in Europe. . Technical Report Luxembourg: European Environmental Agency 2012.

Schetelig MF, Caceres C, Zacharopoulou A, Franz G, Wimmer EA. Conditional embryonic lethality to improve the sterile insect technique in Ceratitis capitata(Diptera: Tephritidae). BMC Biology. 2009;7:1-13.

Schetelig MF, Scolari F, Handler AM, Kittelmann S, Gasperi G, Wimmer EA. Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly Ceratitis capitata. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:18171-18176.

Schlein Y. Lethal effect of tetracycline on tsetse flies following damage to bacterioid symbionts. Experientia. 1977;33:450-451.

Schneider DI, Garschall KI, Parker AG, Abd-Alla AM, Miller WJ. Global Wolbachia prevalence, titer fluctuations and their potential of causing cytoplasmic incompatibilities in tsetse flies and hybrids of Glossina morsitans subgroup species. J Invertebr Pathol. 2013;112 Suppl:S104-115.

Scolari F, Benoit JB, Michalkova V, Aksoy E, Takac P, Abd-Alla AM. The spermatophore in Glossina morsitans morsitans: insights into male contributions to reproduction. Sci Rep. 2016;6.

Scolari F, Benoit JB, Michalkova V, Aksoy E, Takac P, Abd-Alla AM, Malacrida AR, Aksoy S, Attardo GM. The Spermatophore in Glossina morsitans morsitans: Insights into Male Contributions to Reproduction. Sci Rep. 2016;6:20334.

Scolari F, Gomulski LM, Gabrieli P, Manni M, Savini G, Gasperi G, Malacrida AR. How functional genomics will impact fruit fly pest control: the example of the Mediterranean fruit fly, Ceratitis capitata. BMC Genet. 2014;15 Suppl 2:S11.

Scolari F, Gomulski LM, Ribeiro JM, Siciliano P, Meraldi A, Falchetto M, Bonomi A, Manni M, Gabrieli P, Malovini A, *et al.* Transcriptional profiles of mating-responsive genes from testes and male accessory glands of the Mediterranean fruit fly, Ceratitis capitata. PloS one. 2012;7:e46812.

Scolari F, Schetelig MF, Bertin S, Malacrida AR, Gasperi G, Wimmer EA. Fluorescent sperm marking to improve the fight against the pest insect Ceratitis capitata (Wiedemann; Diptera: Tephritidae). N Biotechnol. 2008;25:76-84.

Scolari F, Yuval B, Gomulski LM, Schetelig MF, Gabrieli P, Bassetti F. Polyandry in the medfly - shifts in paternity mediated by sperm stratification and mixing. BMC Genet. 2014;15.

Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, *et al.* Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. Genome Biology. 2014;15:1-17.

Severson DW, Behura SK. Mosquito genomics: progress and challenges. Annual review of entomology. 2012;57:143-166.

Sharakhova MV, Timoshevskiy VA, Yang F, Demin S, Severson DW, Sharakhov IV. Imaginal discs-a new source of chromosomes for genome mapping of the yellow fever mosquito Aedes aegypti. PLoS Negl Trop Dis. 2011;5:e1335.

Sharakhova MV, Timoshevskiy VA, Yang F, Demin SI, Severson DW, Sharakhov IV. Imaginal Discs – A New Source of Chromosomes for Genome Mapping of the Yellow Fever Mosquito Aedes aegypti. PLoS Neglected Tropical Diseases. 2011;5:e1335.

Siciliano P, He XL, Woodcock C, Pickett JA, Field LM, Birkett MA. Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, Ceratitis capitata. Insect Biochem Mol Biol. 2014;48.

Siciliano P, Scolari F, Gomulski LM, Falchetto M, Manni M, Gabrieli P. Sniffing out chemosensory genes from the Mediterranean fruit fly, Ceratitis capitata. PLoS One. 2014;9.

Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, Harrington LC. Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, Aedes aegypti: potential tools for control of female feeding and reproduction. Insect biochemistry and molecular biology. 2008;38:176-189.

Snel B, Bork P, Huynen MA. Genome phylogeny based on gene content. Nature genetics. 1999;21:108-110.

Soumana IH, Simo G, Njiokou F, Tchicaya B, Abd-Alla AM, Cuny G, Geiger A. The bacterial flora of tsetse fly midgut and its effect on trypanosome transmission. J Invertebr Pathol. 2013;112 Suppl:S89-93.

South A, Sirot LK, Lewis SM. Identification of predicted seminal fluid proteins in Tribolium castaneum. Insect Mol Biol. 2011;20:447-456.

Storey JD. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64:479-498.

Sullivan MF, Gould DJ, Maneechai S. Observations on the host range and feeding preferences of Aedes albopictus (Skuse). J Med Entomol. 1971;8:713-716.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proceedings of the National Academy of Sciences of the United States of America. 2001;98:7375-7379.

Szyniszewska AM, Tatem AJ. Global assessment of seasonal potential distribution of Mediterranean fruit fly, Ceratitis capitata (Diptera: Tephritidae). PLoS One. 2014;9.

Szyniszewska AM, Tatem AJ. Global Assessment of Seasonal Potential Distribution of Mediterranean Fruit Fly, Ceratitis capitata (Diptera: Tephritidae). PloS one. 2014;9:e111582.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30.

Thomas MC, Heppner JB, Woodruff RE, Weems HV, Steck GJ, Fasulo TR. Mediterranean Fruit Fly, Ceratitis capitata (Wiedemann) (Insecta: Diptera: Tephritidae). Featured Creatures [Online Database] Florida Department of Agriculture and Consumer Services 2001.

Ting C-T, Tsaur S-C, Wu M-L, Wu C-I. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science (New York, NY). 1998;282:1501-1504.

Tobe SS, Langley PA. Reproductive Physiology of Glossina. Annual review of entomology. 1978;23:283-307.

Tsaur SC, Ting CT, Wu CI. Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of Drosophila: II. Divergence versus polymorphism. Mol Biol Evol. 1998;15:1040-1046.

Twig E, Yuval B. Function of multiple sperm storage organs in female Mediterranean fruit flies (Ceratitis capitata, Diptera: Tephritidae). Journal of insect physiology. 2005;51:67-74.

Ushijima T, Morimura K, Hosoya Y, Okonogi H, Tatematsu M, Sugimura T, Nagao M. Establishment of methylation-sensitive–representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. Proceedings of the National Academy of Sciences of the United States of America. 1997;94:2284-2289.

Valiente Moro C, Tran FH, Raharimalala FN, Ravelonandro P, Mavingui P. Diversity of culturable bacteria including Pantoea in wild mosquito Aedes albopictus. BMC Microbiol. 2013;13:70.

Venkatesan M, Broman KW, Sellers M, Rasgon JL. An initial linkage map of the West Nile Virus vector Culex tarsalis. Insect Mol Biol. 2009;18:453-463.

Vera MT, Rodriguez R, Segura DF, Cladera JL, Sutherst RW. Potential Geographical Distribution of the Mediterranean Fruit Fly, Ceratitis capitata (Diptera: Tephritidae), with Emphasis on Argentina and Australia. Environmental Entomology. 2002;31:1009-1022.

Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biol Evol. 2011;3.

Vogt RG. The Molecular-Basis of Pheromone Reception. Chemical Senses. 1988;13:1.

Vogt RG, Riddiford LM. Pheromone binding and inactivation by moth antennae. Nature. 1981;293:161-163.

Vontas J, Kioulos E, Pavlidi N, Morou E, della Torre A, Ranson H. Insecticide resistance in the major dengue vectors Aedes albopictus and Aedes aegypti. Pesticide Biochemistry and Physiology. 2012;104:126-131.

Vosshall LB, Hansson BS. A unified nomenclature system for the insect olfactory coreceptor. Chem Senses. 2011;36:497-498.

Vreysen MJ, Saleh KM, Ali MY, Abdulla AM, Zhu Z-R, Juma KG, Dyck VA, Msangi AR, Mkonyi PA, Feldmann HU. Glossina austeni (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. J Econ Entomol. 2000;93.

Wang J, Weiss BL, Aksoy S. Tsetse fly microbiota: form and function. Frontiers in cellular and infection microbiology. 2013;3:69.

Warnock RC, Yang Z, Donoghue PC. Exploring uncertainty in the calibration of the molecular clock. Biology letters. 2012;8:156-159.

Wedell N, Gage MJG, Parker GA. Sperm competition, male prudence and sperm-limited females. Trends in Ecology & Evolution. 2002;17:313-320.

Weiss BL, Maltz M, Aksoy S. Obligate symbionts activate immune system development in the tsetse fly. Journal of immunology (Baltimore, Md : 1950). 2012;188:3395-3403.

Weiss BL, Wang J, Aksoy S. Tsetse Immune System Maturation Requires the Presence of Obligate Symbionts in Larvae. PLoS biology. 2011;9:e1000619.

Weiss BL, Wang J, Maltz MA, Wu Y, Aksoy S. Trypanosome Infection Establishment in the Tsetse Fly Gut Is Influenced by Microbiome-Regulated Host Immune Barriers. PLoS Pathog. 2013;9:e1003318.

Werren JH, Baldo L, Clark ME. Wolbachia: master manipulators of invertebrate biology. Nat Rev Micro. 2008;6:741-751.

White-Cooper H, Bausek N. Evolution and spermatogenesis. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2010;365:1465-1480.

Whittier TS, Kaneshiro KY, Prescott LD. Mating Behavior of Mediterranean Fruit Flies (Diptera: Tephritidae) in a Natural Environment. Annals of the Entomological Society of America. 1992;85:214-218.

WHO. Control and surveillance of human African trypanosomiasis. World Health Organization technical report series. 2013:1-237.

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F. TRANSFAC: an integrated system for gene expression regulation. Nucleic acids research. 2000;28:316-319.

Wu C-I, Johnson NA, Palopoli MF. Haldane's rule and its legacy: Why are there so many sterile males? Trends in Ecology & Evolution. 1996;11:281-284.

Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO. Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. Genes & development. 1999;13:1263-1275.

Xue L, Li X, Noll M. Multiple protein functions of paired in Drosophila development and their conservation in the Gooseberry and Pax3 homologs. Development. 2001;128:395-405.

Xue L, Noll M. Drosophila female sexual behavior induced by sterile males showing copulation complementation. Proceedings of the National Academy of Sciences. 2000;97:3272-3275.

Yamada H, Vreysen MJB, Bourtzis K, Tschirk W, Chadee DD, Gilles JRL. The Anopheles arabiensis genetic sexing strain ANO IPCL1 and its application potential for the sterile insect technique in integrated vector management programmes. Acta Tropica. 2015;142:138-144.

Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution. 2007;24:1586-1591.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586-1591.

Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 2000;17:32-43.

Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 2000;155:431-449.

Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107-1118.

Yin C, Shen G, Guo D, Wang S, Ma X, Xiao H, Liu J, Zhang Z, Liu Y, Zhang Y, *et al.* InsectBase: a resource for insect genomes and transcriptomes. Nucleic acids research. 2016;44:D801-807.

Zouache K, Voronin D, Tran-Van V, Mousson L, Failloux AB, Mavingui P. Persistent Wolbachia and cultivable bacteria infection in the reproductive and somatic tissues of the mosquito vector Aedes albopictus. PloS one. 2009;4:e6388.

Curriculum Vitae

Personal Information

Name	Grazia
Surname	Savini
Address	Corso Garibaldi, 54 Pavia (PV), Italy
Date and Place of birth	May 28, 1988 – Atri (TE), Italy
Nationality	Italian
E-mail	grazia.savini01@universitadipavia.it
Phone (Cell)	+39 3299863023

Personal Profile:

Problem solver, open-minded and hard-working molecular biologist. Curious and passionate about science, sensitive to sustainability issues and fascinated by technological development aimed at improving lives. Creative thinker and team player with the ability to work alone without supervision.

Education & Qualifications

Nov. 2013 - present **Ph.D. in Genetics, Molecular and Cellular Biology** University of Pavia, Italy
Laboratory of Genetics and Genomics of Insects
Defense date: December, 2016
Supervisor: Prof. Anna R. Malacrida

Apr. – Dec. 2015 **Research Scholar** Yale School of Public Health, USA
Laboratory of Epidemiology of Microbial Diseases
Supervisor: Prof Serap Aksoy

Sept. 2011 – Jul. 2013 **Master Degree in Molecular Biology and Genetics** University of Pavia, Italy
Score: 110/110 with distinction
English as official language
Laboratory of Molecular Cytogenetics
Scientific supervisor: Prof. Elena Raimondi

Sept. 2007 - Jul. 2011 **Bachelor Degree in Biology** University of L'Aquila, Italy
Score: 108/110
Laboratory of Patology and Immunology

Supervisor: Prof. Adriano Angelucci

Sept. 2002 - Jun. 2007 **High School Degree**, Liceo Scientifico Luca da Penne – Penne (PE), Italy.
Score: 97/100

Relevant Experience

Technical competence/experience:

- Main molecular biology techniques: PCR, iPCR, RT-PCR, cloning, transformations, electrophoresis, nucleic acid purification, blotting (Southern, Northern and Western), FISH, PRINS, RDA (representational difference analysis), fluorescence microscopy;
- Real Time PCR, with knowledge of related software, such as MiniOpticon and CFX by BioRad;
- Cell culture (fibroblast cell lines);
- Chromosome isolation and staining;
- Analyses and ability to handle genomic data, genome manual annotation;
- Analyses of large dataset of transcriptomic sequences (RNA-seq libraries), sequence analyses, alignments;
- Use of bioinformatics and web-based software such as WebApollo, CLC Main Workbench and CLC Genomics;
- Evolutionary analyses and estimation of dN/dS ratio;
- Knowledge of good laboratory practices;
- Accurate, detailed data recording;
- Ability to lead simultaneous scientific research projects;
- Ability to lead and mentor undergraduate and master students;
- Results discussion via oral presentations;
- Teaching in general and evolutionary zoology, University of Pavia (A.A.2015/16);
- Scientific reviewer for “BMC Genomics” and “Biological Sciences” Journals, 2016-current.

Computing skills and competence:

- Operating systems: GNU/Linux, Mac OS X, MS Windows;
- Basic knowledge of Unix scripting;
- Ability to handle large data sets efficiently using scripts, databases, and other tools.

Language skills:

- Italian, native speaker;

- Excellent knowledge of spoken and written English, used to give oral presentations;
- Basic knowledge of French.

Teaching/supervisory experience

2013-present	Supervised several students in the Lab. of Genetics and Genomics of Insects at University of Pavia, Italy.
2013-present	Teaching Assistant of General and Evolutionary Zoology, University of Pavia, Italy.

Extracurricular Activities

Courses/Workshops:

Heidelberg,Germany 19-23rd June 2016	EMBO Practical Course: Computational biology: Genomes to systems.
Lyon, France 6-10th June 2016	Third RCM on Enhancing Vector Refractoriness to Trypanosome Infection.
Lyon, France 2-4th June 2016	Workshop on Tissue Localization and Analysis of Insect Endosymbionts by Fluorescence Microscopy: Theoretical Background and Practical Applications.
Lyon, France 30 May-1st June 2016	Workshop on Bioinformatics Approaches for Microbiota Profiling Based on Amplicon Sequencing Data.
Muguga, Kenya 15-21st March 2015	Multiple Glossina Genome Annotation Workshop.
Volos, Greece 1-5th September 2014	COST action, Training school “Managing cherry pests in space and time”.
Cortona, Italy 23-25th June 2014	AGI course: Microbiota come genotipo esteso.

Publications

2016

Gabrieli P, Scolari F, Di Cosimo A, **Savini G**, Fumagalli M, Gomulski LM, Malacrida AR, Gasperi G. **Sperm-Less Males Modulate Female Behaviour in Ceratitis capitata (Diptera: Tephritidae)**. Insect Biochemistry and Molecular Biology, 13-26. doi: 10.1016/j.ibmb.2016.10.002

Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K, Castañera P, Cavanaugh JP, Chao H, Childers C, Curril I, Dinh H, Doddapaneni H, Dolan A, Dugan S, Friedrich M, Gasperi G, Geib S, Georgakilas G, Gibbs RA, Giers SD, Gomulski LM , González-Guzmán M, Guillem-Amat A, Han Y, Hatzigeorgiou AG, Hernández-Crespo P, Hughes DST, Jones JW, Karagkouni D, Koskinioti P, Lee SL, Malacrida AR, Manni M, Mathiopoulos K, Meccariello A, Murali SC, Murphy TD, Muzny DM, Oberhofer G, Ortego F, Paraskevopoulou MD, Poelchau M, Qu J, Reczko M, Robertson HM, Rosendale AJ, Rosselot AE, Saccone G, Salvemini M, **Savini G**, Schreiner P, Scolari F, Siciliano P, Sim SB, Tsiamis G, Ureña E, Vlachos IS, Werren JH, Wimmer EA, Worley KC, Zacharopoulou A, Richards S, and Handler AM**. The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitate (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species.** *Genome Biology*, 17:192. doi: 10.1186/s13059-016-1049-2.

Submitted / In submission:

Manni M, Guglielmino C, Scolari F, Vega-Rùa A, Failloux A, Somboon P, Lisa A, **Savini G**, Bonizzoni M, Gomulski L, Malacrida A. **Global invasion of Aedes albopictus as a vector of arboviruses: what consequences for disease outbreaks?** PLoS Negl. Trop. Dis., 2016 (submitted)

2015

Dritsou V, Topalis P, Windbichler N, Simoni A, Hall A, Lawson D, Hinsley M, Hughes D, Napolioni V, Crucianelli F, Deligianni E, Gasperi G, Gomulski LM, **Savini G,** Manni M, Scolari F, Malacrida AR, Arcá B, Ribeiro JM, Lombardo F, Saccone G, Salvemini M, Moretti R, Aprea G, Calvitti M, Picciolini M, Papathanos PA, Spaccapelo R, Favia G, Crisanti A, Louis C (2015) **A draft genome sequence of an invasive mosquito: an Italian *Aedes albopictus*.** *Pathogens and Global Health*, 109: 207-220. doi: 10.1179/2047773215Y.0000000031.

2014

Scolari F, Gomulski LM, Gabrieli P, Manni M, **Savini G**, Gasperi G, Malacrida AR (2014) **How the 'omics' era will impact fruit fly pest control: the example of the Mediterranean fruit fly, *Ceratitis capitata*.** *BMC Genetics*, 15 (Suppl 2): S11. doi: 10.1186/1471-2156-15-S2-S11.

Abstracts/Posters

2016

ICE 2016 XXV International Congress of Entomology, Orlando, Florida USA, 25-30 September 2016

Scolari , **Savini G**, Malacrida AR, Aksoy S, Attardo G. Exploring the role of endosymbionts in male reproductive physiology: impact on ejaculate composition and function in the tsetse fly Glossina m. morsitans.

2015

EMBO Conference on "Molecular and population biology of mosquitoes and other disease vectors: current, resurgent and emerging results", Kolymbari, Crete, Greece, 24-29 July 2015

Gomulski LM, Manni M, Scolari F, **Savini G**, Tait G, Nolan T, Lowson D, Ribeiro JMC, Malacrida AR, Gasperi G. Olfaction as a window between *Aedes albopictus* and the environment: chemosensory gene repertoire of wild populations.

Keystone Symposia: The Arthropod Vector: the controller of transmission, Taos, New Mexico, USA, 12-17 May 2015

Gomulski LM, Manni M, Scolari F, **Savini G**, Tait G, Nolan T, Lawson D, Ribeiro JMC, Malacrida AR, Gasperi G. Differential transcript abundances and single nucleotide polymorphisms in the chemosensory gene repertoire of wild *Aedes albopictus* populations.

2nd FAO/IAEA RCM on "Exploring Genetic Molecular, Mechanical and Behavioural Methods of Sex Separation in Mosquitoes". Juazeiro, Brazil - 9 March - 13 March 2015

Scolari F, **Savini G**, Manni M, Gabrieli P, Gomulski LM, Bonizzoni M, Gasperi G, Malacrida M - Cytogenetic and molecular characterization of sex-specific markers in the tiger mosquito, *Aedes albopictus*.

2014

7th International Symposium on Molecular Insect Science, Amsterdam (NL), 3-16 July 2014

Manni M, Aketarawong N, Gomulski LM, Tait G, Scolari F, **Savini G**, Falchetto M, Malacrida AR, Gasperi G. Origin and migration routes of the Asian tiger mosquito, *Aedes albopictus*.

2013

AGI congress (Italian Geneticists Association), Cortona (Italy), 2013

Belloni E, Piras F, Mazzagatti A, Badiale C, Meinardi B, Castro A, **Savini G**, Cerutti F, Bensi M, Nergadze S, Raimondi E and Giulotto E. Analysis of the physical organization and of the CENP binding ability of horse centromeric satellite DNA families.

19th International Chromosome Conference, Bologna (Italy), 2013

Raimondi E, Belloni E, Piras F, Mazzagatti A, Badiale C, Meinardi B, Castro A, **Savini G**, Bensi M, Nergadze S, Giulotto E. Physical organization and CENP binding ability of horse centromeric satellite DNA families.

References

Prof. Anna R. MALACRIDA
PhD Supervisor
Dept. of Biology and Biotechnology
University of Pavia
Via Ferrata 9, Pavia 27100, Italy
Tel: +39 0382 98 6059
Email: malacrid@unipv.it

Prof. Serap AKSOY
Research Scholar Supervisor
Yale School of Public Health
60 College Street
New Haven, CT 06520-8034
Tel: 203.737.2180
Email: serap.aksoy@yale.edu

Prof. Elena RAIMONDI
Master thesis Supervisor
Dept. of Biology and Biotechnology
University of Pavia
Via Ferrata 9, Pavia 27100, Italy
Tel: +39 0382 98 5556
Email: elena.raimondi@unipv.it