CRANFIELD UNIVERSITY

Emma Bailey

Generating beneficial predator genomes to provide comparative insights into insecticide resistance-related gene families

Environment and Agrifood

PhD

Academic Year: 2018 - 2022

Supervisor: Dr Fady Mohareb

Associate Supervisor: Dr Keywan Hassani-Pak & Dr Robert King

January 2022

CRANFIELD UNIVERSITY

Emma Bailey

Generating beneficial predator genomes to provide comparative insights into insecticide resistance-related gene families

Environment and Agrifood

PhD

Academic Year: 2018 - 2022

Supervisor: Dr Fady Mohareb

Associate Supervisor: Dr Keywan Hassani-Pak & Dr Robert King

January 2022

Contents

Acknowledgements

Firstly, I am extremely grateful to my supervisor at Cranfield university, Dr Fady Mohareb, for his continuous support and patience throughout not only my PhD, but also my Master's degree. He has always boosted my confidence in my abilities and I would undoubtedly not be where I am now in my academic journey without his encouragement.

I am also hugely appreciative of my project supervisors at Rothamsted Research, Dr Keywan Hassani-Pak and Dr Robert King for their ongoing advice and assistance throughout my PhD and for providing a friendly and caring environment to work in.

I would like to acknowledge my funding body, the Pest Genomics Initiative, a collaborative project between Rothamsted Research, Bayer Crop Science and Syngenta AG. Additionally I would like to thank all members of the Pest Genomics Initiative for devising this PhD project. Special mentions go to Dr Linda Field, Dr Christopher Rawlings, Dr David Hughes, Dr Benjamin Buer, Dr Ralf Nauen and Dr Eric Ganko for their kind support throughout and patience with reviewing my manuscripts. Additionally I am grateful to Martin Williamson for his technical support in the lab.

I am very thankful for my peer mentor, Dr Dana McGregor for her wonderful support and for our morale boosting chats over cake! I would also like to thank Donna Fellowes for giving life to the student community at Rothamsted, for endlessly organising events for us and for her boundless positivity. I am grateful to the student community at Rothamsted as a whole, for providing a strong and unique sense of unity as we faced the ups and downs of PhD life together. Special thanks to Caitlin Willis and Patry Ortega-Ramos for their help with collecting and rearing insects.

Finally, my appreciation goes out to my family and friends for supporting me through nearly a decade of being a student - the end is here! And of course, I am eternally grateful for my partner Liam who provided endless encouragement and forced me to take breaks from my computer occasionally - we can finally have our kitchen table back!

Abstract

With a rapidly growing population to feed, finding ways to increase crop yields has become more important than ever. Insect pests contribute hugely to yield losses every year and finding methods to effectively control pest levels is therefore crucial to reduce these losses. Insecticide use alone is no longer a viable solution, due to ever increasing levels of resistance developing amongst crop pests, as well as the environmental concerns associated with their overuse. Biological control – the use of natural predators to keep pest populations under control – has proven to be a highly effective method of pest control and generally has less severe environmental impacts than pesticides (although introducing non-native species can result in undesirable changes to local biodiversity). Biological control agents are therefore a key component of Integrated Pest Management (IPM) strategies, which aim to manage pest populations in a sustainable and economical manner. IPM programs prioritise selective insecticides which target the pest species and are harmless to the beneficial predators. However, the numbers of reported insecticide resistance cases are far lower in beneficial predators as opposed to crop pests. As a result, insecticide application often harms beneficial predator populations and reduces their biological control capabilities, which may allow resistant pest populations to surge after application.

Genomic information is readily available for a multitude of crop pest species, however, when this project began, there was minimal genomic data available for beneficial predators. By increasing the availability of genomic data for beneficial predators, we can perform comparative analyses between crop pests and their predators of insecticide target-sites and genes encoding metabolic enzymes potentially responsible for insecticide resistance. These analyses may help uncover whether there is any genomic basis for the reduced number of insecticide resistance cases in beneficial predators compared to crop pests.

The aim of this project was to firstly sequence and assemble the genomes of key beneficial predators for which no genomic information is currently available. This included *Orius*

laevigatus (minute pirate bug), *Sphaerophoria rueppellii* (European hoverfly) and *Microctonus brassicae* (parasitoid wasp). Next, these genomes were annotated and manual curation of resistance-associated detoxification genes was performed. The resultant detoxification gene sets were then used to perform comparative analyses between beneficial predators and crop pests.

The results from the comparative analysis suggest a greater degree of detoxification family gene expansion within crop pests compared to beneficial predators. This difference was particularly apparent in the families associated with detoxification of plant xenobiotics and suggests that the plant-based diet of crop pests provided increased selection pressure for resistance mechanisms prior to the introduction of insecticides. Once insecticides were introduced, crop pests may therefore have had an advantage over beneficial predators in terms of developing insecticide resistance. In addition, variation in the levels of resistance between different beneficial predators correlated to some extent with gene expansion, with several factors having likely had some influence on this, including diet, migration and length of commercial use.

The knowledge gained from this project could contribute to our understanding of insecticide resistance from a genomic perspective and aid in the development of successful IPM strategies.

Executive summary

This thesis project was performed as part of the Pest Genomics Initiative (PGI) - a collaboration between Rothamsted Research, Bayer Crop Science and Syngenta Crop Protection - which aims to sequence and annotate the genomes of key global insect pests and beneficial predators. As the world population grows, so does the demand for high-quality and sustainable food. There is therefore a need for new pest-control strategies which can reduce crop losses whilst having minimal impact on biodiversity and ecosystems. The high-quality annotated genomes generated by the PGI will be made publicly available to assist with the development of future pest-control strategies.

This project will focus on sequencing and annotating beneficial predator genomes which will subsequently be used for comparative analyses with pest genomes. These analyses will compare detoxification family genes which are potentially involved in insecticide resistance between pests and predators to assess if there is any genomic basis which could explain why beneficial predators have a lesser degree of insecticide resistance compared to pest species.

This thesis will begin with a literature review in Chapter 1, covering the current status of the crop pest problem; strategies which can be used to control pests; the genetics of insecticide resistance; and an overview of genome assembly and annotation methods which could be suitable for insect genomes.

The first species to be sequenced and annotated is *Orius laevigatus*, a minute pirate bug which is a widely-used commercially available beneficial predator. Chapter 2 will report a scaffold-level genome assembly for this species, and a comparative analysis of insecticide resistance-related gene families with hemipteran crop pests.

The second species to be sequenced and annotated is *Sphaerophoria rueppellii*, a European hoverfly. This species is a key pollinator and has recently been made commercially available due to its efficacy as a beneficial predator. Chapter 3 will report a near-chromosome level assembly for this species and a comparative analysis of insecticide resistance-related gene families with hemipteran crop pests and pollinators.

The final species to be sequenced and annotated is *Microctonus brassicae*, a parasitic wasp of *Psylliodes chrysocephala* (the cabbage-stem flea beetle). This species is not yet commercially available, but shows promise for use in future pest control strategies against the highly resistant *P. chrysocephala*. Chapter 4 will cover a scaffold-level genome assembly for this species and a comparative analysis of insecticide resistance-related gene families in hemipteran crop pests.

Chapter 5 will include a final comparative analysis of insecticide resistance-related gene families between several beneficial predator species and crop pests. The results will be used to establish if differences in these gene families could account for differing levels of insecticide resistance. The beneficial predator species will include those covered in Chapters 2 to 4, as well as *Chrysoperla carnea*, a green lacewing, and other publicly available species. Here the limitations of such comparative analyses will be covered, as well as the potential causes of the differences seen in these gene families.

Finally, Chapter 6 will conclude the thesis with some overall conclusions from Chapters 2 to 5 as well as some proposals for future work.

Chapter 1. Introduction

1.1 Feeding a growing population

With the human population expected to reach 9 billion by 2050, it has been estimated that food production needs to increase by 70-100% in order to feed this growing population. Increasing the amount of agricultural land is not a viable option due to other pressures such as urbanization and biodiversity protection. Therefore the focus is currently on increasing crop productivity in order to achieve higher overall yields. Options for this include the use of genetically modified crops, altering agronomic practices and Integrated Pest Management (IPM) [1].

1.2 Integrated Pest Management (IPM)

IPM is a hugely important aspect of increasing crop yields, with losses as a result of pests estimated at 50-80% in some major crops. Pests responsible for these losses include weeds, pathogens and animal pests, with animal pests suggested to account for ~18% of these losses [2]. This figure unlikely represents the true impact which insect pests have, as this does not include the damage they impart on crops through virus transmission, which can often have a far greater impact than the direct damage caused by insect pests [3]. In cases of a particularly severe outbreak, insect pests can even result in a complete loss of yield, as was the case in India where the diamondback moth, *Plutella xylostella*, destroyed 100% of a cauliflower crop [4].

IPM aims to control pests through a carefully selected combination of strategies. Efforts are made to prevent pests becoming an issue in the first place, through the destruction of pest habitats, using pest resistant crop varieties, altering planting dates and creating environments to encourage beneficial predators. Pest populations (and the populations of beneficial

predators) are also closely monitored to ensure pesticides are only used when necessary to minimise damage to the environment, human health and beneficial predators [5].

1.3 Pesticide drawbacks

Whilst pesticides may appear to be a fast-acting and thorough solution to insect pest control, their use in fact brings an abundance of issues to the environment. Additionally, the temporary nature of pesticides means they often need frequent reapplication in order to control the pest population – resulting in further damage.

Pesticides tend to be non-specific, and therefore whilst they may be effective at eradicating vast numbers of pests, they can also be toxic to other species, including: domestic animals, humans [6] and crucial pollinators including the honey bee [7]. This can occur as a result of direct exposure to pesticides (via feeding on contaminated seeds and plant matter, or physical contact with sprayed plants) in the case of pollinators [8], foraging birds [9] and wild mammals, as well as indirect exposure through bioaccumulation in the food chain for predatory mammals [10]. Pesticide application in agricultural fields near surface water may also result in runoff/leaching which can harm amphibian and aquatic species via both direct toxicity and sublethal effects such as alterations to their morphology, nervous systems and fecundity [11–14], and has also resulted in a direct increase in mortality in several invertebrate species [15]. Additionally, pesticide use is linked to the reduction of species diversity of plants - likely as a result of agricultural intensification - resulting in a loss of the habitats and food sources of a variety of mammals, birds and insects, including pollinators and natural enemies [16]. This could result in a reduction of the population levels of these species, and in terms of pest control, the loss of natural enemies would be a significant disadvantage.

Pesticides can limit any potential benefit of beneficial predators [16]. Pesticides will often reduce the predator population to a level where they are no longer able to provide sufficient

biological control against the pests, resulting in a resurgence of the pest population - despite a potential initial reduction in the pest population immediately after pesticide application. (It is often this initial reduction in pest population that prompts farmers to use subsequent applications of pesticides.) A key example of this is the brown planthopper (BPH), a pest of rice in Asia, which prior to pesticide use had been kept under control with natural enemies of the pest, including water bugs and mirid bugs. However, once pesticides became the primary method of pest control used by Asian rice farmers, multiple large-scale BPH outbreaks occurred. This resulted in policy changes which emphasised a focus on the use of biological control and encouraged farmers to avoid the use of insecticides [17]. Another example of pest resurgence occurred in citrus crops in California, where the vedalia ladybeetle was successfully being used as a natural enemy to keep the cottony-cushion scale under control. However, when DDT pesticides were introduced in the 1940s, their usage resulted in a virtual extinction of the vedalia population, and a massive surge of the pest population as a result. A swift return was made to the use of biological control with a reintroduction of the vedalia ladybeetle, and the pest population was under control once again [18].

Pesticide resistance is also an ever increasing problem, with ~600 insect species being reported as resistant to at least one pesticide in 2014 [19]. This reduced efficacy of many pesticides results in higher doses being used in an attempt to control pest populations, which will thereby increase the cost of their use [20]. It is especially concerning that of the species showing resistance to pesticides, 96% are pests/parasites, and only 4% are beneficial predators [21], although it should be noted that there will likely be fewer reports for beneficial predator species due to a higher interest in crop-damaging pest species.

Fundamentally, the use of pesticide-centric strategies for pest control is unsustainable. Pesticides result in considerable off-target damage to other species and the environment; they limit the efficacy of natural enemies and even if pesticides appear effective at first use, their efficacy is likely to decline as resistance increases within the target pest species.

IPM provides a promising alternative to such pesticide-centric strategies. With IPM, the aim is to avoid the use of broad-spectrum pesticides which are harmful to the environment, and use targeted pesticides, but only when absolutely necessary. The overall goal is to implement pesticide strategies which target only the damaging pest and leave other species, particularly natural enemies, unharmed [22]. Without knowing the impact of pesticides on beneficial predators, IPM programs could be much less effective than expected. Even naturally derived pesticides, such as the widely-used spinosyn biopesticide, need to be assessed for their impact on non-target organisms. Even if not immediately lethal to beneficial predators, biopesticides can still have lasting effects which result in a reduced population and a lesser ability to control pest populations [23].

1.4 Biological control

Pest management strategies utilizing biological control agents have proved highly effective all over the world, and in the most successful cases they have even provided permanent control of the target species with additional chemical control rendered unnecessary [18,24,25]. Biological control has several advantages over chemical control, including a lower cost [26] and significantly lesser damage to the environment and human health [27]. Resistance is also unlikely to present much of an issue with biological control methods, although there are a few examples of pests sequestering defensive compounds from plants as protection against natural enemies [28–30].

However, IPM strategies do have to be thoroughly researched, as certain aspects can conflict with each other. For example, species interactions amongst beneficial predators can result in a reduced level of pest control if not managed sufficiently. These predators may not just feed on their target pests, but could also have detrimental effects on other beneficial predators by feeding on them or consuming their eggs [31]. This has especially been found to be the case

when thrip and aphid predators are used in unison, and has resulted in current strategies often failing to control aphids [32].

It is also important to consider the impact of introducing non-native species for pest control, as the impacts to local ecology could be disastrous. For example the introduction of a tachinid fly parasitoid, *Bessa remota*, is thought to be responsible for the extinction of the coconut moth, *Levuana iridescens*, in Fiji [33] and also hugely reduced population numbers of a non-target moth species, *Heteropan dolens* [34]. Another example is the multicolored Asian lady beetle, *Harmonia axyridis*, which was released as a biological control agent in North America to control aphid populations [35]. However, it has since become something of a pest itself by feeding on fruits in orchards and vineyards, aggregating in people's homes, feeding on aphid parasitoids and immature monarch butterflies (*Danaus plexippus*) and displacing native coccinellids - potentially due to resource competition [36].

We can attempt to avoid such negative impacts by performing pre-release studies to assess any potential risks to non-target species and also to study their efficacy. It is crucial to avoid situations such as in North America, where over 60 beneficial predators were introduced to control the gypsy moth, *Lymantria dispar*, with little impact on the pest population and unknown impacts on non-target species [37,38]. Post-release monitoring is also key to continually assess efficacy and environmental impact, as effects may not be seen in the short-term. For example, it may not be until the target pest population is severely diminished that biological controls start to impact non-target species; and wider impacts, such as the ecological consequences of biodiversity loss and food chain disruption, may not be immediately apparent.

Arthropod predators, including *Orius laevigatus*, *Chrysoperla carnea* and Syrphidae predatory hoverflies as well as parasitoid wasps are well-established as effective biological control agents against insect pest populations in orchards, glasshouses and crop fields [39–51]. Many arthropod predators are now commercially available as biological control agents and are widely used for this purpose as part of IPM strategies targeting whiteflies, aphids and thrips,

which are well known to be some of the most damaging crop pests in the world [52].

1.5 Importance of genetics

The genetics of resistance has been well studied, with genes responsible for resistance mechanisms having been identified in certain insect species [53]. Insecticide resistance falls into three main categories: a) increasing levels of certain cuticular proteins to reduce cuticle permeability to insecticides [54,55]; b) altering levels of detoxification enzymes which can metabolize or sequester insecticides [56–58] and c) target site mutations which reduce sensitivity to the insecticide [59,60].

Supergene families are common for detoxification enzymes and have generally evolved as a result of gene duplications / amplification. These families include: cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione *S*-transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxyl/choline esterases (CCEs) [61–66]. Through the mapping of these gene families in certain species, potential resistance-associated loci have been identified, which have in turn resulted in the identification of further resistance-linked genes. Genes associated with resistance can also be used to infer orthologues in other species [67].

Predicting the resistance status of individuals can be as simple as identifying the presence of a resistance allele, especially in the case of target site resistance [68–70] Or it may involve looking for gene duplications which are responsible for increased levels of detoxification gene expression. For example, in aphids, gene duplications have been found which result in higher concentrations of insecticide-metabolising enzymes being produced [71]. Increased production of metabolic resistance enzymes can also result from mutations in regulatory and/or promoter regions which induce an upregulation of gene products. Such mutations are reportedly responsible for cases of insecticide resistance in *D. melanogaster* and *M. domestica* [59].

Searching for the presence of such causal genes in the genomes of insects could give an insight into the mechanism of pesticide resistance, thus allowing a targeted pest management strategy to be adopted. Identifying beneficial predators which possess genes/mutations for insecticide resistance allows for their selection before being used commercially for biological control [72]. Screening of pest populations could give an insight into the likelihood / timescale for resistance to certain pesticides developing, by studying the frequency of known resistant alleles within the population [19].

1.6 Comparative genomics & genomic data uses

When this project began, minimal genomic data was available for beneficial predators. The few published beneficial predator genomes included: *Metaseiulus/Galendromus occidentalis*, which served as the first reference genome for phytoseiid mites [73] and three *Nasonia* parasitoid wasp genomes [74]. Increasing the availability of predator genomes would provide a useful resource for research in multiple areas, including: isolating genes which specify the target prey (the parasitoid wasp genomes have permitted the identification of a genomic region responsible for host preference); identifying mechanisms responsible for locating prey and host plants (e.g. interactions with insect pheromones and plant semiochemicals) and also identifying certain dietary requirements, which could prove useful for efficient large-scale rearing of biological control agents. Interestingly, in certain parasitic species, the genome may also permit the identification of venoms which could be used as a stand-alone method of biological control [74].

In contrast to beneficial predators, many pest genomes are already publicly available, including: *Aphis glycines* (soybean aphid) [75], *Acyrthosiphon pisum* (pea aphid) [76], *Diuraphis noxia* (Russian wheat aphid) [77], *Bemisia tabaci* (whitefly) [78,79], *Spodoptera frugiperda* (fall armyworm) [80], *Bombyx mori* (silkworm) [81], *Tetranychus urticae* (two-spotted spider mite)

[82], *Pieris rapae* (cabbage white butterfly) [83] and *Tribolium castaneum* (red flour beetle) [84].

The availability of these pest genomes have permitted comparative genomics studies looking at expansions in insecticide resistance-related gene families in insect pests such as Triatomine bugs [85,86], aphids [87] and the red flour beetle [88]. However, no such studies exist for beneficial predators compared with pests. The good availability of these pest genomes makes this an ideal time to sequence beneficial predators because this will enable us to identify differences between their genomes through comparative analysis studies. This could reveal vital information regarding the relationship between these beneficial predators and the pests they feed on, such as how they have co-evolved alongside each other. In this project, the focus of these comparative analyses will be on the detoxification gene families which are associated with insecticide resistance. Expansion of these gene families can indicate how insects may have evolved insecticide resistance mechanisms, and the level of expansion may indicate the degree of resistance a species possesses. Previous studies have shown that these gene families are evolving at a rapid rate, with a diverse set of detoxification genes found across differing insect species [67,89]. Comparative analyses of these gene families may therefore give an indication as to why beneficial predators are generally more susceptible to insecticides than crop pests [21].

1.7 Sequencing

Insects possess several attributes which can hinder the production of high quality genome assemblies. Their small physical size means multiple individuals often need to be pooled to generate sufficient DNA for sequencing, however, pooling individuals results in a high level of polymorphism due to the presence of multiple genomes, and rearing inbred colonies to reduce polymorphism is often not possible due to time constraints and inbreeding depression [90]. Highly polymorphic samples can negatively impact assembly quality and contiguity [91]. In

addition, insect genomes often have significant repeat content, which can impact the quality of the assembly. Past arthropod genome assemblies have shown a repeat content as high as 70% (in *Rhipicephalus microplus*) [92]. Care therefore must be taken when developing sequencing strategies for insects, if a high quality assembly is to be achieved.

The low cost of short-read genome assembly using technologies such as Illumina and 454 sequencing, has prompted the sequencing of many species. Short-read technologies have low DNA input requirements, potentially avoiding the need for pooling individuals and reducing polymorphism in the sample. However, short-read technologies often struggle with high heterozygosity and vast repetitive regions, resulting in assemblies with low contiguity and large gaps that will ultimately have a poor annotation model and only ever achieve draft assembly status [91]. There are methods which can improve genome assemblies, such as the use of gap-closing software and postprocessing to correct mis-assemblies. Nevertheless, the best approach to overcome these issues is through the incorporation of long-reads, which span large enough regions that the repetitive regions can be accurately assembled, thus hugely reducing the proportion of gaps in the final assembly [93].

Long-read technologies, such as PacBio and Oxford Nanopore, cope better with high polymorphism and repeat content. However, generating sufficient high-molecular weight DNA material is a significant issue for long-read technologies, which have high input requirements due to stringent size selection steps. The recently developed low input PacBio protocol does claim to make long-read sequencing approaches possible with inputs as low as 100ng through the removal of size selection steps [94], but unfortunately limited commercial availability at the time of writing restricts its use for this project. The removal of size selection steps also requires the starting DNA to be relatively free of short fragments, which may require an improvement of current DNA extraction and storage protocols. In addition, potentially the most significant issue of long-read sequencing approaches is the high-error rate: 11-15% with PacBio [95] and 5-15% with Oxford Nanopore [96]. The high cost of long-read sequencing means it is often too costly to simply increase the coverage to a sufficient level to account for

the high error rate, and the high input requirements can also put a restriction on the achievable coverage levels. In contrast, short-read sequencing has a low cost and low error rate. The error rate can be as low as ~0.01% for methods such as Illumina [97].

To overcome the weaknesses of either sequencing method as a standalone approach, a hybrid approach can be used, which combines both long- and short-read sequencing technologies. Combining short-read data with long-read data reduces the coverage depth required for long-read sequencing, thereby lowering the cost of sequencing. The low-error rate short-reads can be used to perform error correction of the long-reads, which in turn will span much longer regions of DNA than the short-reads [95]. This results in an assembly with a low error rate whilst still maintaining high contiguity. Another benefit of the hybrid approach is the inclusion of repetitive regions within the assembly, regions which have often been poorly studied in other sequenced organisms. Their importance is becoming more widely known, with repetitive DNA now being recognised as playing a critical role in gene regulation and evolution. These repetitive regions have even been directly linked to certain diseases [98]. The hybrid sequencing approach has been tested on several arthropod species which previously only had short-read assemblies available. These updated assemblies consistently had higher contiguity and showed a greater level of completeness compared to the short-read assembly version with [92,99].

Whilst not available at the start of the project, the recently released PacBio HiFi sequencing method provides a solution to some of the aforementioned issues, providing long-read data with a significantly higher accuracy (99.5%) than standard PacBio data (85%). Using HiFi data for *de novo* assembly removes the need for error correction methods which can introduce mis-assemblies and has been shown to produce assemblies of a dramatically increased quality compared to the methods covered previously [100].

Alongside long- and short-read technologies, another addition to genome assembly pipelines involves the generation of chromosome conformation capture (3C) data such as Hi-C. This method produces a ‘map’ of chromatin interactions throughout the genome which can be used

to group and scaffold contigs based on their physical proximity within the genome [101]. A combination of the hybrid sequencing approach alongside Hi-C data significantly improved the contiguity and completeness of the *Apis mellifera* genome, with an N50 score 120-fold higher than the original short-read only assembly [102].

Based on the strong successes of the hybrid and 3C approaches for other arthropod assemblies, it seems sensible to apply these methods to the beneficial predator species of interest. However, due to the high cost of long-read and 3C data, this was not possible for all species. PacBio long-read and Illumina short-read data was obtained for both *Orius laevigatus* and *Sphaerophoria rueppellii*, and Hi-C data was obtained for *S. rueppellii* only. Had the PacBio HiFi method been available at the start of the project, it is likely HiFi data would have been obtained for all species. Instead it was used for only the final species - *Microctonus brassicae*.

1.8 Assembly

There are a variety of tools available for de novo genome assembly, as well as a further wide range of parameters for each of these tools, leading to an almost exponential number of ways that assemblies can be performed. Each of these tools has their own strengths and therefore decisions on which assemblers to use will be based on multiple factors, including: what genome size it is best suited for; how well it copes with highly heterozygous genomes; whether it is suited for PacBio or Illumina reads, or if it is a fully hybrid assembler which accepts both types of reads. Whilst papers exist which aim to evaluate the most successful assembler [103], their efficacy can vary hugely depending on the characteristics of the genome, in terms of ploidy level, size, proportion of repetitive elements, heterozygosity and so on. With no single widely accepted protocol for genome assembly, especially in the case of insect genomes, the process is often somewhat of a trial and error procedure where many assemblers are trialed and then the one(s) with the best resultant assembly quality metrics (covered in section 1.9) will be selected.

The assemblers which could be used for this process should be designed to cope with highly polymorphic genomes, as will be the case for non-inbred insect populations. Short-read assemblers notoriously struggle to cope with this type of genome assembly, whereas long-read assemblers tend to perform better. For example, Canu is a long-read assembler designed for repetitive and heterozygous genomes which can produce high contiguity assemblies [104]. However, when using a long-read only assembler such as Canu, subsequent error correction is required due to the high error rate of long-read PacBio data. Error correction can be performed using the raw reads once the assembly has been completed; for example the Pilon error correction tool [105] can identify differences between the draft assembly and raw reads, to fix mis-assemblies and resolve gaps.

An alternative is to use a long-read assembler which incorporates error polishing such as Flye which is capable of generating highly contiguous and accurate assemblies for repetitive genomes, using long-reads for error-correction [106]. Another alternative is FALCON/FALCON-Unzip, which has been designed with non-inbred heterozygous genomes in mind and incorporates a long-read error correction step using Arrow [107].

Read coverage needs to be high (>70x) if the long-reads are to be used to perform error correction against themselves [108]. However, the low error rate of short reads means they will generally provide better quality error correction, resulting in a more accurate final assembly. Using short-reads to polish a long-read assembly is a form of 'hybrid assembly'.

Hybrid assemblies incorporate both long- and short-read data, and they can be used to circumvent the shortcomings of long- or short-read assemblers. There are two general approaches to hybrid assembly. The "short-read first" approach is where contigs generated using Illumina short-reads are scaffolded using long-reads. This takes advantage of the low error rate of Illumina reads to produce highly accurate contigs, whilst the long–reads can be used to remove gaps and vastly improve the contiguity of the assembly [95]. However, using pre-assembled short-reads can introduce mis-assemblies due to the difficulty short-read assemblers face when attempting to resolve repetitive elements or highly polymorphic regions.

A better hybrid assembly option for outbred insect populations is the “long-read first” method, which involves using long-reads to produce the assembly [109], and then using the short-reads purely for error correction [95]. A “long-read first” hybrid assembler such as Platanus-allee [110] is well-suited to heterozygous genomes. Alternatively, any well-suited long-read assembler (such as Flye, Falcon or Canu) could be used and followed by error polishing with short-reads using a tool such as Pilon [105].

When using PacBio HiFi data there is less need for short-reads for error polishing, due to the low-error rate. A *de novo* assembler such as Hifiasm works well with heterozygous genomes and can produce high-quality assemblies with high contiguity using only long-read HiFi data [111].

Another issue with highly heterozygous genomes is the introduction of redundancy which can result in highly fragmented assemblies with a larger total size than expected. A tool such as the Redundans pipeline [112] can be used to remove redundant regions, i.e. alternative heterozygous contigs. In addition, Redundans can perform scaffolding and gap-closing, resulting in reduced fragmentation and a more accurate assembly. RNA-seq data can also be used to perform scaffolding with a tool such as Rascaf [113]. However, scaffolding using these methods may be inaccurate due to the high repetitive content of the genomes as well as mis-assemblies in the draft assembly. In order to perform truly chromosome-level quality scaffolding, Hi-C data is required.

Pipelines are available for processing and utilizing Hi-C data for super-scaffolding. This includes a platform called Juicer which can process raw Hi-C data into a list of contacts, generate contact matrices and annotate structural features of the genome [114]. The 3D-DNA pipeline can then be used to generate a candidate chromosome-length genome assembly using the output from Juicer and the draft assembly [115].

Endless time could be spent making slight improvements to a genome assembly to improve contiguity and completeness scores. With such a wide range of available assemblers and

protocols it can be difficult to ascertain which assembly is the 'best'. If certain aspects of each assembly are better quality than others produced using different assemblers (e.g. one assembly may have high completeness, whilst another assembly has high contiguity) then multiple assemblies can be merged to improve overall contiguity using a meta-assembler such as Quickmerge [116]. However, if the genome is at the level of completeness that it can be successfully used for the required analyses (i.e. comparative genomics), then it may be worth concentrating on trying to answer biological questions as a better use of time. A balance is needed between having the perfect assembly and having an assembly which can be used to answer biological questions, although where this balance lies is a fairly subjective decision.

1.9 Determining Assembly Completeness

In order to assess which assembly is the 'best' quality, we can assess the 'three C's' of genome quality: correctness (i.e. the concordance of an assembly to a good quality reference); completeness (i.e. whether it has all the genes we would expect it to have) and contiguity (i.e. how complete the assembled sequence is and how many separate fragments there are). Reference genomes are not available in the case of *de novo* assemblies, meaning a simple comparison between genomes to check for correctness will not be possible in these cases. Therefore the focus will be on completeness and contiguity.

BUSCO assesses genome completeness by searching the assembly for single-copy orthologs which are known to be present in over 90% of insect species. These genes/orthologs are included with the BUSCO software, however datasets can be manually curated to be used with BUSCO [117].

To assess contiguity, a tool such as QUAST can be used which generates metrics including N50 (the length of the shortest contig/scaffold at 50% of the total genome length) as well as the length of the longest contig and total contig number [118].

Values related to the contiguity of the genome do not take into account gene content and therefore do not necessarily reflect the quality of the genome [119]. BUSCO scores could arguably be considered the most important measure of assembly completeness, since it is the gene content of the assembly which tends to be most essential in order for biological questions to be answered. However, in cases where the non-coding or repetitive regions are of particular interest, N50 scores would perhaps give a better indication of overall completeness.

1.10 Annotation

Once the assemblies are complete, they must undergo annotation to find gene models. Annotating eukaryotic genomes is challenging due to vast non-coding and repetitive regions. A good partial solution to this is the use of RNA-seq data, which can be aligned to the genome to identify the correct exon-intron structure for each gene; however, this will only capture expressed genes [120]. In order to capture as many genes as possible, a method such as the MAKER annotation pipeline can be used [121]. This pipeline incorporates RNA-seq data, as well as alignments to databases of known proteins and novel gene predicting software to identify gene models. In addition, repetitive regions are masked using RepeatMasker [122] to ensure accurate gene annotation.

However there are still limitations to annotation using this approach: fragmented assemblies can result in genes being split across several contigs and these may be annotated as fragment genes; sequencing errors may also cause erroneous annotation, and finally, incorrectly annotated genes within databases used for alignment will introduce errors into the annotation [120].

There is currently little progress when it comes to improving the quality of automated genome annotation. However, manual curation can be performed for genes of interest using RNA-seq

data aligned to the genome and gene models from closely related species to ensure the intron-exon structure of predicted gene models is correct [120].

To check that the annotation is of good quality, BUSCO can be run in 'protein' mode to assess the completeness of the annotated gene set [117].

Functional annotation of the gene models is essential for further biological research and is generally done using homology searches. Blast2GO [123] incorporates BLAST searches against public sequence databases (such as NCBI nr) alongside other annotation resources, such as the InterPro database (a database of 'signatures' for protein domains, families and functional sites) [124].

1.11 Aims and Objectives

In summary, the aims of this project are as follows: to sequence three beneficial predator species using a range of different technologies; to develop best-suited assembly pipelines for each species based on the sequencing data available; to annotate the genomes and perform manual curation of resistance-linked detoxification genes and finally, to perform comparative analyses of beneficial predator detoxification gene families with publicly available pest genomes to assess if there is a genomic basis for the heightened insecticide susceptibility in predators compared to pests.

1.12 Literature cited

[1] Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. Food security: the challenge of feeding 9 billion people. Science 2010;327:812–8.
[2] Oerke E-C. Crop losses to pests. J Agric Sci 2006;144:31–43.
[3] Dedryver C-A, Le Ralec A, Fabre F. The conflicting relationships between aphids and men: a review of aphid damage and control strategies. C R Biol 2010;333:539–53.
[4] Ahmad T, Ansari MS, Ali H. Outbreak of diamondback moth, Plutella xylostella in Aligarh, India 2009.
[5] Flint ML. IPM in Practice, 2nd Edition: Principles and Methods of Integrated Pest Management. University of California Agriculture and Natural Resources; 2012.
[6] Mostafalou S, Abdollahi M. Pesticides: an update of human exposure and toxicity. Arch Toxicol 2017;91:549–99.
[7] Goulson D, Nicholls E, Botías C, Rotheray EL. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 2015;347:1255957.
[8] Fletcher, Barnett. Bee pesticide poisoning incidents in the United Kingdom. Bull Insectology 2003.
[9] Mineau P, Downes CM, Kirk DA, Bayne E, Csizy M. Patterns of bird species abundance in relation to granular insecticide use in the Canadian prairies. Écoscience 2005;12:267–78.
[10] Harris S, Morris P, Wray S, Yalden D. A review of British mammals: Population estimates and conservation status of British mammals other than Cetaceans, Joint Nature Conservation Committee, Peterborough, UK 1995. http://www.jncc.gov.uk/page-2759 (accessed June 9, 2022).
[11] Adams E, Leeb C, Brühl CA. Pesticide exposure affects reproductive capacity of common toads (Bufo bufo) in a viticultural landscape. Ecotoxicology 2021;30:213–23.
[12] Sparling DW, Fellers GM, McConnell LL. Pesticides and amphibian population declines in California, USA. Environ Toxicol Chem 2001;20:1591–5.
[13] Woodley SK, Mattes BM, Yates EK, Relyea RA. Exposure to sublethal concentrations of a pesticide or predator cues induces changes in brain architecture in larval amphibians. Oecologia 2015;179:655–65.
[14] Greulich K, Pflugmacher S. Differences in susceptibility of various life stages of amphibians to pesticide exposure. Aquat Toxicol 2003;65:329–36.
[15] Liess M, Von Der Ohe PC. Analyzing effects of pesticides on invertebrate communities in streams. Environ Toxicol Chem 2005;24:954–65.
[16] Geiger F, Bengtsson J, Berendse F, Weisser WW, Emmerson M, Morales MB, et al. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. Basic Appl Ecol 2010;11:97–105.
[17] Bottrell DG, Schoenly KG. Resurrecting the ghost of green revolutions past: The brown planthopper as a recurring threat to high-yielding rice production in tropical Asia. J Asia Pac Entomol 2012;15:122–40.
[18] DeBach P, Rosen D. Biological Control by Natural Enemies. CUP Archive; 1991.
[19] Sparks TC, Nauen R. IRAC: Mode of action classification and insecticide resistance management. Pestic Biochem Physiol 2015;121:122–8.
[20] Georghiou GP. The magnitude of the resistance problem. Pesticide Resistance: Strategies and Tactics for Management 1986.
[21] Georghiou GP, Lagunes-Tejeda A. The occurrence of resistance to pesticides in arthropods 1991.
[22] Croft BA. Developing a Philosophy and Program of Pesticide Resistance Management. In: Roush RT, Tabashnik BE, editors. Pesticide Resistance in Arthropods, Boston, MA: Springer US; 1990, p. 277–96.
[23] Biondi A, Desneux N, Siscaro G, Zappalà L. Using organic-certified rather than synthetic pesticides may not be safer for biological control agents: selectivity and side effects of 14 pesticides on the predator Orius laevigatus. Chemosphere 2012;87:803–12.
[24] Gurr G, Wratten SD. Biological control: Measures of success. 2000th ed. London, England: Chapman and Hall; 2000.
[25] Waage JK, Greathead DJ, Brown R, Paterson RRM, Haskell PT, Cook RJ, et al. Biological control: challenges and opportunities. Philos Trans R Soc Lond B Biol Sci 1988;318:111–28.
[26] Naranjo SE, Ellsworth PC, Frisvold GB. Economic value of biological control in integrated pest management of managed plant systems. Annu Rev Entomol 2015;60:621–45.
[27] Howarth FG. Environmental Impacts of Classical Biological Control. Annu Rev Entomol 1991;36:485–509.
[28] Ode PJ. Plant chemistry and natural enemy fitness: effects on herbivore and natural enemy interactions. Annu Rev Entomol 2006;51:163–85.
[29] Nishida R. Sequestration of defensive substances from plants by Lepidoptera. Annu Rev Entomol 2002;47:57–92.
[30] Robert CA, Zhang X, Machado RA, Schirmer S, Lori M, Mateo P, et al. Sequestration and activation of plant toxins protect the western corn rootworm from enemies at multiple trophic levels. Elife 2017;6. https://doi.org/ 10.7554/eLife.29307.
[31] Messelink GJ, Bloemhard CMJ, Sabelis MW, Janssen A. Biological control of aphids in the presence of thrips and their enemies. Biocontrol 2013;58:45–55.
[32] Bloemhard C, Ramakers P. Strategies for aphid control in organically grown sweet pepper in the Netherlands. IOBC WPRS BULLETIN 2008;32:25.
[33] Kuris AM. Did biological control cause extinction of the coconut moth, Levuana iridescens, in Fiji? In: Pederson J, editor. Marine Bioinvasions: Patterns, Processes and Perspectives, Dordrecht: Springer Netherlands; 2003, p. 133–41.

[34] Robinson GS. Macrolepidoptera of Fiji and Rotuma: a taxonomic and biogeographic study. Doctoral. Durham University, 1974.
[35] Gordon RD. Journal of the New York Entomological Society. The Coccinellidae (Coleoptera) of America North of Mexico 1985;93:1–912.
[36] Koch RL. The multicolored Asian lady beetle, Harmonia axyridis: a review of its biology, uses in biological control, and non-target impacts. J Insect Sci 2003;3:32.
[37] Beirne BP. Avoidable obstacles to colonization in classical biological control of insects. Can J Zool 1985;63:743–7.
[38] Pimentel D, Glenister C, Fast S, Gallahan D. Environmental Risks of Biological Pest Controls. Oikos 1984;42:283–90.
[39] Tenhumberg B, Poehling H-M. Syrphids as natural enemies of cereal aphids in Germany: Aspects of their biology and efficacy in different years and regions. Agric Ecosyst Environ 1995;52:39–43.
[40] Freier B, Triltsch H, Möwes M, Moll E. The potential of predators in natural control of aphids in wheat: Results of a ten-year field study in two German landscapes. Biocontrol 2007;52:775–88.
[41] Brewer MJ, Elliott NC. Biological control of cereal aphids in north america and mediating effects of host plant and habitat manipulations. Annu Rev Entomol 2004;49:219–42.
[42] Pineda A, Marcos-García MÁ. Seasonal Abundance of Aphidophagous Hoverflies (Diptera: Syrphidae) and Their Population Levels In and Outside Mediterranean Sweet Pepper Greenhouses. Ann Entomol Soc Am 2008;101:384–91.
[43] Symondson WOC, Sunderland KD, Greenstone MH. Can generalist predators be effective biocontrol agents? Annu Rev Entomol 2002;47:561–94.
[44] Frescata C, Mexia A. Biological Control of Thrips (Thysanoptera) byOrius laevigatus(Heteroptera: Anthocoridae) in Organically-Grown Strawberries. Biological Agriculture & Horticulture 1996;13:141–8. https://doi.org/ 10.1080/01448765.1996.9754773.
[45] Chambers RJ, Long S, Helyer NL. Effectiveness of Orius laevigatus (Hem.: Anthocoridae) for the control of Frankliniella occidentalis on cucumber and pepper in the UK. Biocontrol Sci Technol 1993;3:295–307.
[46] Hassan SA. Über die Massenzucht von Chrysopa carnea Steph. (Neuroptera, Chrysopidae). Z Angew Entomol 2009;79:310–5.
[47] T. R. NEW. The biology of Chrysopidae and Hemerobiidae (Neuroptera), with reference to their usage as biocontrol agents: a review. Trans R Entomol Soc Lond 1975;127:115–40.
[48] Stelzl M, Devetak D. Neuroptera in agricultural ecosystems. Agric Ecosyst Environ 1999;74:305–21.
[49] Canard M, Séméria Y, New TR. Biology of chrysopidae. vol. 27. Springer; 1984.
[50] NEW, T. R. Neuroptera. Aphids Their Biology, Natural Enemies and Control 1988:249–58.
[51] Wang Z-Z, Liu Y-Q, Shi M, Huang J-H, Chen X-X. Parasitoid wasps as effective biological control agents. J Integr Agric 2019;18:705–15.
[52] Malais M. The Biology of Glasshouse Pests and Their Natural Enemies: Knowing and Recognizing. Koppert.Biological systems; 1992.
[53] Roush RT, McKenzie JA. Ecological genetics of insecticide and acaricide resistance. Annu Rev Entomol 1987;32:361–80.
[54] Silva AX, Jander G, Samaniego H, Ramsey JS, Figueroa CC. Insecticide resistance mechanisms in the green peach aphid Myzus persicae (Hemiptera: Aphididae) I: A transcriptomic survey. PLoS One 2012;7:e36366.
[55] Wood O, Hanrahan S, Coetzee M, Koekemoer L, Brooke B. Cuticle thickening associated with pyrethroid resistance in the major malaria vector Anopheles funestus. Parasit Vectors 2010;3:67.
[56] Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. Insect Mol Biol 2005;14:3–8.
[57] Feyereisen R. Molecular biology of insecticide resistance. Toxicology Letters 1995;82-83:83–90. https://doi.org/10.1016/0378-4274(95)03470-6.
[58] Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. PLoS Genet 2010;6:e1000999.
[59] Li X, Schuler MA, Berenbaum MR. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu Rev Entomol 2007;52:231–53.
[60] Hemingway J, Hawkes NJ, McCarroll L, Ranson H. The molecular basis of insecticide resistance in mosquitoes. Insect Biochem Mol Biol 2004;34:653–65.
[61] Li X, Shi H, Gao X, Liang P. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in Plutella xylostella (L.). Pest Manag Sci 2018;74:695–704.
[62] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[63] Pavlidi N, Vontas J, Van Leeuwen T. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. Curr Opin Insect Sci 2018;27:97–102.
[64] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[65] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[66] Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Current Opinion in Insect Science 2019;31:131–8.
[67] Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. Evolution of supergene families associated with insecticide resistance. Science 2002;298:179–81.
[68] Liu Z, Williamson MS, Lansdell SJ, Denholm I, Han Z, Millar NS. A nicotinic acetylcholine receptor mutation conferring target-site resistance to imidacloprid in Nilaparvata lugens (brown planthopper). Proc Natl Acad Sci U S A 2005;102:8420–5.
[69] Crossthwaite AJ, Rendine S, Stenta M, Slater R. Target-site resistance to neonicotinoids. J Chem Biol 2014;7:125–8.
[70] Kwon DH, Im JS, Ahn JJ, Lee J-H, Marshall Clark J, Lee SH. Acetylcholinesterase point mutations putatively associated with

monocrotophos resistance in the two-spotted spider mite. Pestic Biochem Physiol 2010;96:36–42.
[71] Devonshire AL, Moores GD. A carboxylesterase with broad substrate specificity causes organophosphorus, carbamate and pyrethroid resistance in peach-potato aphids (Myzus persicae). Pestic Biochem Physiol 1982;18:235–46.
[72] Roderick GK, Navajas M. Genes in new environments: genetics and evolution in biological control. Nat Rev Genet 2003;4:889–99.
[73] Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. Genome Biol Evol 2016;8:1762–75.
[74] Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 2010;327:343–8.
[75] Wenger JA, Cassone BJ, Legeai F, Johnston JS, Bansal R, Yates AD, et al. Whole genome sequence of the soybean aphid, Aphis glycines. Insect Biochem Mol Biol 2020;123:102917.
[76] International Aphid Genomics Consortium. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol 2010;8:e1000313.
[77] Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, et al. The genome of Diuraphis noxia, a global aphid pest of small grains. BMC Genomics 2015;16:429.
[78] Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol 2016;14:110.
[79] Xie W, Chen C, Yang Z, Guo L, Yang X, Wang D, et al. Genome sequencing of the sweetpotato whitefly Bemisia tabaci MED/Q. Gigascience 2017;6:1–7.
[80] Kakumani PK, Malhotra P, Mukherjee SK, Bhatnagar RK. A draft genome assembly of the army worm, Spodoptera frugiperda. Genomics 2014;104:134–43.
[81] Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, et al. The genome sequence of silkworm, Bombyx mori. DNA Res 2004;11:27–35.
[82] Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 2011;479:487–92.
[83] Shen J, Cong Q, Kinch LN, Borek D, Otwinowski Z, Grishin NV. Complete genome of Pieris rapae, a resilient alien, a cabbage pest, and a source of anti-cancer proteins. F1000Res 2016;5:2631.
[84] Tribolium Genome Sequencing Consortium, Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, et al. The genome of the model beetle and pest Tribolium castaneum. Nature 2008;452:949–55.
[85] Traverso L, Lavore A, Sierra I, Palacio V, Martinez-Barnetche J, Latorre-Estivalis JM, et al. Comparative and functional triatomine genomics reveals reductions and expansions in insecticide resistance-related gene families. PLoS Negl Trop Dis 2017;11:e0005313.
[86] Schama R, Pedrini N, Juárez MP, Nelson DR, Torres AQ, Valle D, et al. Rhodnius prolixus supergene families of enzymes potentially associated with insecticide resistance. Insect Biochem Mol Biol 2016;69:91–104.
[87] Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol 2010;19 Suppl 2:155–64.
[88] Shi H, Pei L, Gu S, Zhu S, Wang Y, Zhang Y, et al. Glutathione S-transferase (GST) genes in the red flour beetle, Tribolium castaneum, and comparative analysis with five additional insects. Genomics 2012;100:327–35.
[89] Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, et al. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol Biol 2006;15:615–36.
[90] Richards S, Murali SC. Best Practices in Insect Genome Sequencing: What Works and What Doesn't. Curr Opin Insect Sci 2015;7:1–7.
[91] Richards S. Arthropod Genome Sequencing and Assembly Strategies. Methods Mol Biol 2019;1858:1–14.
[92] Barrero RA, Guerrero FD, Black M, McCooke J, Chapman B, Schilkey F, et al. Gene-enriched draft genome of the cattle tick Rhipicephalus microplus: assembly by the hybrid Pacific Biosciences/Illumina approach enabled analysis of the highly repetitive genome. Int J Parasitol 2017;47:569–83.
[93] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 2011;13:36–46.
[94] Kingan SB, Heaton H, Cudini J, Lambert CC, Baybayan P, Galvin BD, et al. A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing. Genes 2019;10. https://doi.org/10.3390/genes10010062.
[95] Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, et al. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 2012;30:693–700.
[96] Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol Detect Quantif 2015;3:1–8.
[97] Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of next generation sequencing platforms. Next Gener Seq Appl 2014;1. https://doi.org/10.4172/jngsa.1000106.
[98] Slotkin RK. The case for not masking away repetitive DNA. Mob DNA 2018;9:15.
[99] Kawamoto M, Jouraku A, Toyoda A, Yokoi K, Minakuchi Y, Katsuma S, et al. High-quality genome assembly of the silkworm, Bombyx mori. Insect Biochem Mol Biol 2019;107:53–62.
[100] Hon T, Mars K, Young G, Tsai Y-C, Karalius JW, Landolin JM, et al. Highly accurate long-read HiFi sequencing data for five complex genomes. Sci Data 2020;7:399.
[101] Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods 2015;72:65–75.
[102] Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, et al. A hybrid de novo genome assembly of the honeybee, Apis mellifera, with chromosome-length scaffolds. BMC Genomics 2019;20:275.

[103] Zhang W, Chen J, Yang Y, Tang Y, Shang J, Shen B. A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. PLoS One 2011;6:e17915.
[104] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 2017;27:722–36.
[105] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9:e112963.
[106] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 2019;37:540–6.
[107] Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 2016;13:1050–4.
[108] Salmela L, Walve R, Rivals E, Ukkonen E. Accurate self-correction of errors in long reads using de Bruijn graphs. Bioinformatics 2017;33:799–806.
[109] Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 2017;27:737–46.
[110] Kajitani R, Yoshimura D, Okuno M, Minakuchi Y, Kagoshima H, Fujiyama A, et al. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat Commun 2019;10:1702.
[111] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 2021;18:170–5.
[112] Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 2016;44:e113.
[113] Song L, Shankar DS, Florea L. Rascaf: Improving Genome Assembly with RNA Sequencing Data. Plant Genome 2016;9. https://doi.org/10.3835/plantgenome2016.03.0027.
[114] Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 2016;3:95–8.
[115] Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 2017;356:92–5.
[116] Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 2016;44:e147.
[117] Simão 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.
[118] Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 2013;29:1072–5.
[119] Thrash A, Hoffmann F, Perkins A. Toward a more holistic method of genome assembly assessment. BMC Bioinformatics 2020;21:249.
[120] Salzberg SL. Next-generation genome annotation: we still struggle to get it right. Genome Biol 2019;20:92.
[121] Cantarel BL, Korf I, Robb SMC, 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.
[122] Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 2004;Chapter 4:Unit 4.10.
[123] Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005;21:3674–6.
[124] Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. Nucleic Acids Res 2009;37:D211–5.

Chapter 2. *Orius laevigatus* genome assembly, functional annotation and comparative genomics

This chapter was submitted to BMC Genomics on 18th May 2021, and was accepted for publication on 2nd December 2021.

A scaffold-level genome assembly of a minute pirate bug, *Orius laevigatus* (Hemiptera: Anthocoridae), and a comparative analysis of insecticide resistance-related gene families with hemipteran crop pests

Emma Bailey*†‡, Linda Field*, Christopher Rawlings†, Rob King†, Fady Mohareb‡, Keywan-Hassani Pak†, David Hughes†, Martin Williamson*, Eric Ganko**, Benjamin Buer§, Ralf Nauen§

* Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, UK

† Department of Computational and Analytical Sciences, Rothamsted Research, Harpenden, UK

‡ The Bioinformatics Group, Cranfield Soil and Agrifood Institute, Cranfield University, Cranfield, UK

** Seeds Research, Syngenta Crop Protection, LLC, Research Triangle Park, NC, USA.

§ Bayer AG, Crop Science Division, R&D, Monheim, Germany

2.1 Abstract

Background: *Orius laevigatus*, a minute pirate bug, is a highly effective beneficial predator of crop pests including aphids, spider mites and thrips in integrated pest management (IPM) programmes. No genomic information is currently available for *O. laevigatus*, as is the case for the majority of beneficial predators which feed on crop pests. In contrast, genomic information for crop pests is far more readily available. The lack of publicly available genomes for beneficial predators to date has limited our ability to perform comparative analyses of genes encoding potential insecticide resistance mechanisms between crop pests and their predators. These mechanisms include several gene/protein families including cytochrome P450s (P450s), ATP binding cassette transporters (ABCs), glutathione S-transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxyl/cholinesterases (CCEs).

Methods and findings: In this study, a high-quality scaffold level *de novo* genome assembly for *O. laevigatus* has been generated using a hybrid approach with PacBio long-read and Illumina short-read data. The final assembly achieved a scaffold N50 of 125,649bp and a total genome size of 150.98Mb. The genome assembly achieved a level of completeness of 93.6% using a set of 1,658 core insect genes present as full-length genes. Genome annotation identified 15,102 protein-coding genes - 87% of which were assigned a putative function.

Comparative analyses revealed gene expansions of sigma class GSTs and CYP3 P450s. Conversely the UGT gene family showed limited expansion. Differences were seen in the distributions of resistance-associated gene families at the subfamily level between *O. laevigatus* and some of its targeted crop pests. A target site mutation in ryanodine receptors (I4790M, PxRyR) which has strong links to diamide resistance in crop pests and had previously only been identified in lepidopteran species was found to also be present in hemipteran species, including *O. laevigatus*.

Conclusion and significance: This assembly is the first published genome for the Anthocoridae family and will serve as a useful resource for further research into target-site selectivity issues and potential resistance mechanisms in beneficial predators. Furthermore, the expansion of gene families often linked to insecticide resistance may be an indicator of the capacity of this predator to detoxify selective insecticides. These findings could be exploited by targeted pesticide screens and functional studies to increase effectiveness of IPM strategies, which aim to increase crop yields by sustainably, environmentally-friendly and effectively control pests without impacting beneficial predator populations.

Keywords:

Orius laevigatus; pirate bug; PacBio; Illumina; whole genome sequencing; beneficial predator; insecticide resistance; comparative genomics; Hemiptera; crop pests

2.2 Introduction

Loss of crops to insect pests can account for ~10% of potential yield, as a result of both direct feeding damage and the transfer of viral plant diseases [1]. Thus, to maximise crop yields and sustain food production for a growing world population, pests need to be controlled. At present this control relies mainly on the use of synthetic pesticides, many of which are non-selective and are therefore toxic to both their target pest species and to beneficial predators and parasitoids. As a result there may be a reduction in the predator populations to a level where they are no longer able to contribute natural pest control. This, along with the development of insecticide resistance in pests, can lead to pest populations surging, sometimes to even higher levels than pre-pesticide application [2–4]. Beneficial predators, such as those in the genus *Orius,* have proven to be especially effective in the biological control of crop pests [5]. As generalist predators, *Orius* species target a wide variety of pest species including aphids, beet armyworm, leafhoppers, mites, thrips and whiteflies, many of which are the world's most damaging crop pests [6,7]. Some *Orius* species are commercially available as biological control agents and are widely used for this purpose as part of integrated pest management (IPM) strategies, especially in covered crops [8–10].

Whole genome sequences of insects are helping us to understand many aspects of their biology and behaviour, and this can be applied to potential insecticide resistance mechanisms in pest insects and their natural enemies. However, only a few genomes of beneficial predator species have been published to date, including a phytoseiid mite, *Galendromus occidentalis* [11]; three parasitoid wasps, *Nasonia giraulti, Nasonia longicornis* and *Nasonia vitripennis* [12] and two lady beetles, *Harmonia axyridis* and *Coccinella septempunctata* [13]. To date there are no published genomes for species of the Hemiptera: Anthocoridae (i.e. minute pirate bug) family of predators. In contrast, a growing number of genomes of crop pests are available [14–26]. This larger number of pest genomes, relative to beneficial predator genomes could be in part because up until recently, the genomes of the pests themselves have appeared more useful in

terms of developing targeted pesticides and investigating mechanisms of pesticide resistance. However, agriculture is now moving increasingly away from pesticide use – particularly with the Directive on Sustainable Use of Pesticides 2009/128/EC [27] - and towards IPM strategies, which includes the use of beneficial predators. Future studies of pesticide resistance mechanisms should therefore include beneficial predator genomes alongside pest genomes in order to help select targeted pesticides which do not harm beneficials and subsequently improve the efficacy of IPM strategies [28–32].

The aim of the work reported here was to develop a high-quality genome assembly for *O. laevigatus*, shown in figure 1, to serve as a resource for research into this species as well as the wider Anthocoridae family, which consists of 400-600 mostly predaceous minute pirate bug species - a potentially valuable source of biological control agents [33]. The *O. laevigatus* genome was then used for comparative analyses between beneficial predators and crop pests, focusing on genes encoding potential insecticide resistance mechanisms.

There are two main types of insecticide resistance mechanisms: increased expression of genes encoding protein families involved in metabolic resistance and point mutations in genes encoding insecticide target proteins [34]. Gene families involved in insecticide resistance in pest species are known to include cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione S-transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxyl/cholinesterases (CCEs) [35–40]. Comparisons of the genes/proteins which may be involved in insecticide resistance in crop pests with the corresponding genes in beneficial insects, could aid the development of insecticides which target crop pests but have limited impact on beneficial predator populations. This could prove key to developing successful IPM strategies which exploit differences in the ability of predators and crop pests to tolerate pesticides. Improving the availability of beneficial predator genomes could also help the selection of beneficial predators with genes/mutations for inherent insecticide resistance before being released in the field for biological control [41].

The results presented here provide a comprehensive foundation for further study of potential insecticide resistance mechanisms in beneficial predators and how they compare to crop pests.

Figure 1. ***Orius laevigatus*****.** Image from the Ukrainian Biodiversity Information Network, taken by Boris Loboda, 20.04.19 in Ukraine. Retrieved from: https://ukrbin.com/show_image.php?imageid=106989.

2.3 Methods

2.3.1 Sample preparation and sequencing

Orius laevigatus (commonly known as a minute pirate bug) were obtained from 'Bioline AgroSciences'. CO_2 was used for anaesthesia to allow the insects to be sorted from the substrate. Both adults and nymphs were then flash frozen with liquid N_2 and stored at -80°C. The whole process was done within 48 hours of arrival.

~1000 individuals were pooled for genomic DNA/RNA extractions, which were carried out in-house at Rothamsted Research. The commercial DNAzol reagent was used for the DNA extractions, and the Bioline Isolate II RNA Mini Kit was used for the RNA extractions. The DNA and RNA were sent for library preparation and sequencing by Genewiz (New Jersey, US).

The genome assembly was developed using a hybrid assembly strategy with both Illumina short reads and Pacific Biosciences (PacBio) long reads.

Short reads were sequenced using 2mg of DNA and a library with an insert size of 200bp. Sequencing was done using Illumina HiSeq 4000 with a 2x150bp paired-end configuration. 413,143,574 reads were obtained with a total length of 123 Gb (820x). Raw reads are available under SRA accession: ERR6994870. K-mer counting of the raw Illumina DNA data was done using Jellyfish 2.2.6[42]. Canonical (-C) 21-mers (-m 21) were counted and a histogram of k-mer frequencies outputted. GenomeScope 2.0 [43] was used to process this histogram with 'ploidy' set at 2 and 'maximum k-mer coverage cut-off' set at 10,000.

To obtain long read PacBio data, 3.7mg of DNA first underwent blue pippin size selection (>=10kb) to remove low molecular weight DNA. <500ng of DNA remained after size selection,

and so a low input protocol was used for library construction with an insert size of 20kb. Sequencing was done using the PacBio Sequel I platform and 537,651 reads were obtained with a total length of 6Gb (44x) and an N50 of 11,287bp. Raw reads are available under SRA accession: ERR6941611.

Transcriptome sequencing used 10mg of RNA and a library construction with an insert size of 150bp and PolyA selection for rRNA removal. Sequencing was done using Illumina HiSeq 4000 with a 2x150bp paired-end configuration. 413,137,378 reads were obtained. Raw reads are available under SRA accession: ERR7012629.

FastQC v.0.11.8. [44] was used for quality checks on the raw Illumina HiSeq DNA and RNA sequence data. Adapters were trimmed, low-quality bases (below a score of 3) were removed from the start and end of reads and any reads with a length less than 36 bases were also removed. Trimmomatic v.0.38. [45] was used for these trimming steps. Quality trimming of reads using Trimmomatic resulted in a 0.2% loss of reads for whole genome sequencing and a 5% loss of reads for transcriptome sequencing (table 1).

Table 1. Number of paired-end Illumina HiSeq DNA sequences present before and after trimming

	Illumina DNA Reads	Illumina RNA Reads
Total sequences before trimming	413,143,574	413,137,378
Total sequences after trimming	412,474,208	389,150,727
Sequences lost	669,366	23,986,651

2.3.2 Genome quality assessment

Basic metrics from the genome assembly were calculated using a script developed for the 'Assemblathon' [46]. These metrics include scaffold/contig N50, longest and shortest scaffold length, number of scaffolds exceeding a range of lengths and number of gaps/N's in the assembly.

The completeness of the genome assembly and annotation for *Orius laevigatus* was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) [47] of the insect gene set (insecta odb9). 'Genome' mode was used to assess the assembly, and 'protein' mode to assess the annotation. 'Fly' was used as the training species for Augustus gene prediction. BUSCO assessments were then run with default parameters.

2.3.3 *De novo* genome assembly

The overall assembly pipeline is shown in figure 2. The raw PacBio long reads were assembled into contigs with the Flye v2.5. *de novo* assembler [48,49]. Rascaf was then used to improve the Flye genome assembly with RNA-seq data [50]. Contigs were also produced with the raw PacBio long reads using Canu v1.8 [51] as well as with FALCON v1.3.0 and FALCON-Unzip, which is recommended for heterozygous/outbred organisms with diploid or higher ploidy (and also includes phased-polishing with Arrow) [52,53].

QuickMerge v0.3 [54] was used to merge the assemblies, with Flye as the reference assembly. BUSCO outputs were compared between the merged assembly and the standalone assemblies to identify genes which had been lost during the merging process. Full-length contigs containing these missing genes were extracted from the standalone assemblies and added to the merged assembly, based on the assumption that these contigs would also contain other

missed genes (i.e. those not included in BUSCO's list of 1,658 core insect genes). Multiple rounds of Pilon error polishing [55] were performed, using the Illumina short read data, until no further improvement in BUSCO score was seen.

Redundans [56] was used for scaffolding and redundant contig removal. Redundans is geared towards highly heterozygous genomes. Some redundant regions had to be removed manually, as Redundans does not detect redundancy when only part of the contig is duplicated. The nucmer tool from the MUMmer4 package [57] was used to detect these redundant regions through a whole genome self-alignment.

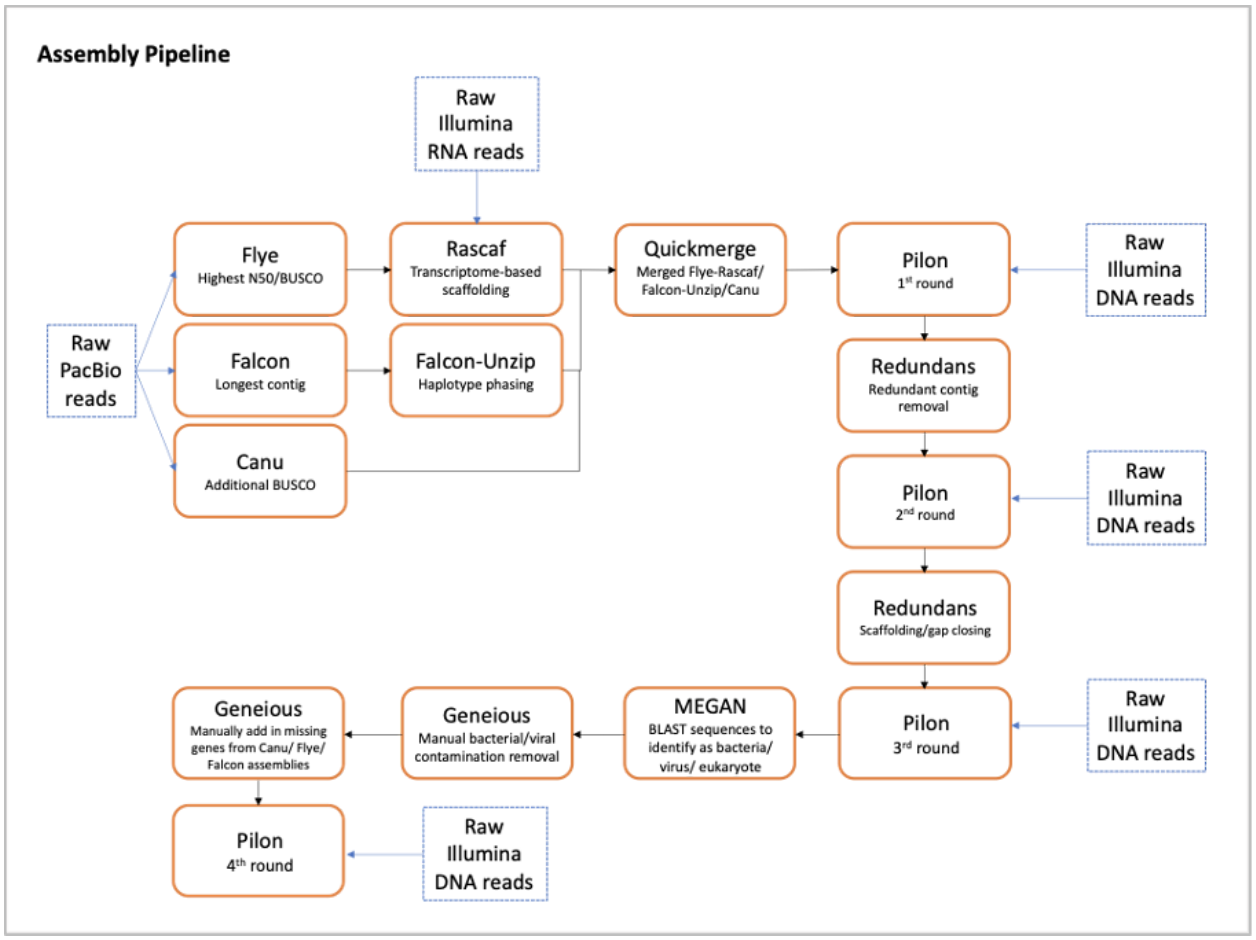


Figure 2. The assembly pipeline for the *Orius laevigatus* genome

A BLAST search against the NCBI Reference Sequence (Refseq) database release 93 [58], was performed using the Tera-BLAST algorithm on a TimeLogic DeCypher system (Active Motif Inc., Carlsbad, CA). The results were processed with Megan [59] to identify any bacterial or viral sequences which were then removed manually in Geneious v10.2.6.

The mitochondrial genome sequence was identified and extracted by running a BLAST search of the *O. laevigatus* genome against the *Orius sauteri* mitochondrial genome which is publicly available at NCBI, GenBank accession No. KJ671626 [60].

2.3.4 Genome annotation

Gene prediction was performed using the MAKER v2.31.8 pipeline [61] through the incorporation of both transcriptome evidence and *ab initio* gene prediction as well as a custom repeat library (see below). MAKER was run using Augustus v3.3.1 [62], GeneMark-ES v4.32 [63] and FGeneSH v8.0.0 [64] as well as EVidenceModeler v1.1.1 [65] with default masking options.

A *de novo* species specific repeat library was constructed using RepeatModeler v1.0.7 [66] to identify repeat models. These models were searched against the GenBank non-redundant (*nr*) protein database for Arthropoda (e value $<10^{-3}$) using Blastx to remove any potential protein-coding genes. This was combined with transposon data to create a custom library. Transposons were identified from the transcriptome assembly by running HMMER: hmmscan [67] against the Pfam database [68] and filtering the resultant Pfam descriptions for those containing “transposon”. A search for transposons was also done on transcripts produced from MAKER and these transposons were then added to the custom repeat library which was used for a second round of MAKER. RepeatMasker v4.0.7 [69] was used to mask repeats in the genome assembly using these repeat libraries, as well as to estimate the abundances of all predicted repeats.

RNA-seq reads were mapped to the genome with HISAT2 v2.0.5 [70] for assembly with StringTie v1.0.1 [71]. A *de novo* assembly was also done using Trinity v2.5.1 [72]. The best transcripts (classified by reasonable transcript size and homology to other species) were selected from the Trinity and StringTie assemblies using Evigene v19.jan01 [73].

Evidence from assembled transcripts was transferred to the genome assembly via MAKER. The output from this was then used to produce a high confidence level gene model training set - overlapping and redundant gene models were removed. Augustus and GeneMark were trained using this training set prior to being used for *ab initio* gene predictions. FGeneSH was run based on the *Drosophila melanogaster* genome.

The best transcripts from both the *ab initio* gene prediction annotation and the transcriptome-based annotation were selected using Evigene (classified by reasonable protein size and homology to other species) and combined to create the final annotation.

Orius laevigatus protein sequences were aligned using Blastp against the non-redundant (*nr*) NCBI protein database for Arthropoda. InterProscan searches were run against several databases (CDD, HAMAP, HMMPAnther, HMMPfam, HMMPIR, FPrintScan, BlastProDom, ProfileScan, HMMTigr) for functional annotation. BLAST2GO [74] was used to assign gene ontology (GO annotations). Infernal v1.1.2 [75] was used to predict and annotate non-coding RNAs.

The mitochondrial genome was annotated using MITOS2 [76] with reference database 'RefSeq 81 Metazoa' and genetic code '5 Invertebrate'.

2.3.5 Comparative Genomics and Phylogenetic Analysis

To produce the species tree, orthogroup gene trees were produced using Orthofinder [77] and the tree was inferred from these using the STAG method [78].

In order to identify genes potentially involved in insecticide resistance, the PFAM domains assigned to gene models during annotation (as described in the 'Genome Annotation' methods section) were used as follows: CCEs (PF00135/IPR002018), GSTs (IPR004045/PF02798), (IPR004046/PF00043), P450s (IPR001128/PF00067), ABCs (IPR003439/PF00005) and UGTs (IPR002213/PF00201). Proteins from UniProt for the classes of interest, from hemipteran species, were used for BLAST queries against *O. laevigatus* to identify any missed genes and to assist with subfamily assignment within these classes. Subfamily assignment for *O. laevigatus* gene families was finalised using phylogenetic trees produced using MAFFT alignments [79,80] and RaxML v8.2.11 [81]. The GAMMA LG protein model [82] was used and a bootstrap consensus tree was inferred from 100 replicates.

Manual checks and curation were performed for genes potentially involved in insecticide resistance. Increased copy numbers of these genes often led to adjacent tandem duplications being incorrectly annotated as one gene model, therefore curation was important to prevent incorrect gene numbers being reported in later analyses. The exon/intron boundaries and start/stop codons of the genes were confirmed through visualization in IGV [83] of RNAseq data mapped to the genome using HISAT2 v2.0.5 [70] and the gene models were edited in Geneious where necessary.

The P450s were classified and named by Dr David Nelson [84]. The UGTs were classified and named by Dr Michael Court [85]. Nomenclature of P450s and UGTs is based on the evolutionary relationships of the sequences. P450 and UGT sequences were BLAST searched against named insect sequences and were assigned to known families if they were >40% (for P450 families) or

>45% (for UGT families) identical. Other sequences were assigned to new families based on their clustering on trees and their percent identity to each other.

2.4 Results and Discussion

2.4.1 Sequencing

In order to produce enough DNA and RNA for sequencing, ~1,000 individuals of *O. laevigatus* were required. Because they were obtained commercially, the level of inbreeding of the culture was not known. However, all individuals were obtained from a single colony within the rearing facility. A high heterozygosity level was therefore a possibility and this was kept in mind when making decisions during the assembly process.

2.4.2 Genome metrics evaluation based on raw reads

The raw read k-mer analysis with GenomeScope 2.0 estimated a haploid genome size of ~141Mb (table 2), in line with the final assembly size of 151Mb. A genome size estimate using methods such as flow cytometry would have provided a more accurate estimate, however, such data was not available for the *Orius* genus. This could be considered a limitation to the study, as 141Mb was provided as a genome size estimate to Canu, Flye and FALCON-Unzip which may have affected the outputted assemblies. Genome repeat length was 20Mb, 16.5% of the total estimated genome size.

Table 2. Genome characteristics obtained from GenomeScope v2.0

Using a k-mer length of 21 and a maximum k-mer coverage of 10,000

	Minimum	Maximum
Heterozygosity, %	1.197	1.297
Genome Haploid Length (Mb)	140.7	142.2
Genome Repeat Length (Mb)	20.2	20.5
Genome Unique Length (Mb)	120.4	121.8
Read Error Rate, %	0.86	0.86

The heterozygosity rate ranged from 1.20% to 1.30%. This alongside the small 'shoulder' to the left of the main 'full-model' peak (figure 3), indicates a fairly high level of heterozygosity, which was taken into consideration in the assembly strategy.

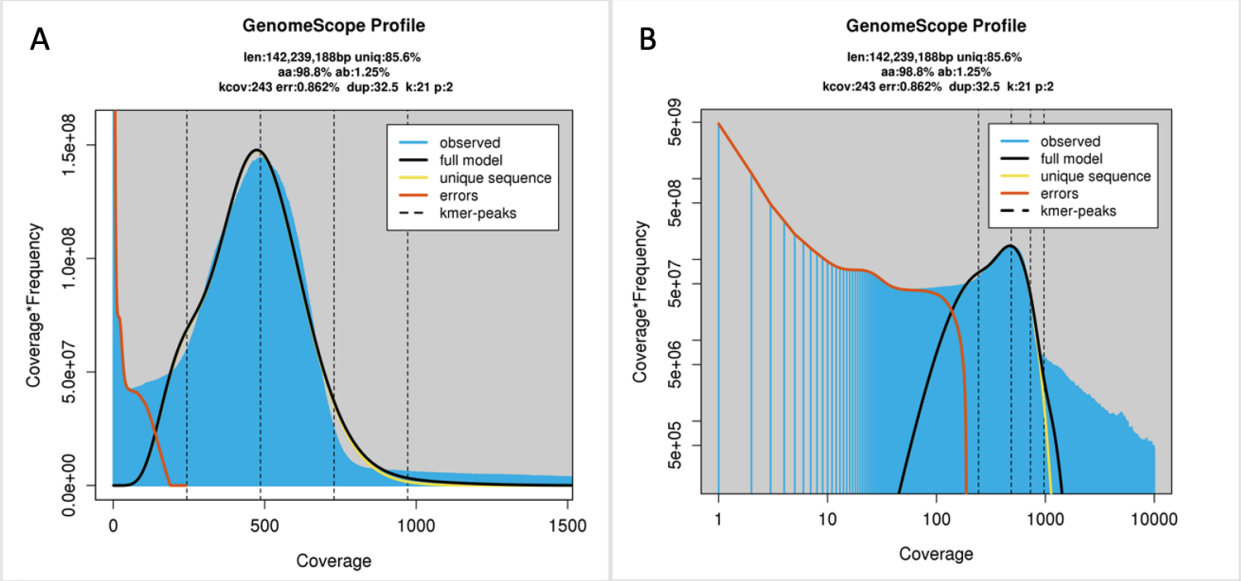


Figure 3. GenomeScope v2.0 profile plots of A: a transformed linear plot of k-mer frequency; and B: a transformed log plot of k-mer coverage at a k-mer length of 21 and a maximum k-mer coverage of 10,000.

2.4.3 Assembly

After trialing a variety of different assemblers, focusing on those suited to heterozygous genomes, Flye, FALCON and Canu achieved the best contiguity (N50) and completeness scores (BUSCO) and so were used to produce 3 separate genome assemblies. The statistics for these assemblies, as well as for subsequent versions of the assembly outlined in this section are shown in appendix i. Rascaf improved the contiguity of the Flye assembly through alignment of the RNA-seq data to the genome, likely because it is less affected by the use of multiple individuals versus genome assembly tools which include non-conserved sequences from a population of individuals. FALCON-Unzip improved the FALCON assembly contiguity with a 4.5-fold decrease in the total number of scaffolds (although this coincided with a ~9% loss of complete gene models found using BUSCO and suggests that FALCON-Unzip may have been too stringent for this genome - perhaps because it was designed with plant and fungal genomes in mind [52,53]).

Flye (both with and without Rascaf) had the best assembly statistics in terms of scaffold N50 and BUSCO score. However, FALCON-Unzip achieved the largest 'longest scaffold' of the three assemblers.

Quickmerge was used to merge the FALCON-Unzip assembly, Rascaf improved Flye assembly and the Canu assembly. The resultant merged assembly had better continuity than any of the stand-alone assemblies, however, the BUSCO completeness was slightly worse than the standalone Flye assembly (and worsened with the second round of Quickmerge). This was likely due to mis-assemblies in the component assemblies causing alignment issues, which resulted in sections of the misassembled contigs being discarded.

Pilon was used for error polishing and improved the BUSCO completeness score. Redundans (redundancy removal and scaffolding/gap-closing) improved the scaffold N50 and removed redundant scaffolds.

A comparison of the gene models (core insect genes from the insecta odb9 BUSCO gene set) found in the original Flye / FALCON-Unzip / Canu assemblies versus the merged assembly showed that some of the gene models were found in at least one of the original assemblies, but were missing in the merged assembly. Of the 154 missing or fragmented genes in the merged assembly (out of a total 1,658 core insect genes), 5 were found in the FALCON-Unzip assembly, 5 in the Flye assembly and 46 in the Canu assembly. Manual editing to bring the full-length contigs containing these missing genes into the merged assembly took the BUSCO completeness score up by 5%. A final round of Pilon improved this score by an additional 0.5% (further rounds of Pilon did not improve the score).

Table 3. Final assembly statistics for the *O. laevigatus* genome

Number of scaffolds	2,050
Total size of scaffolds	150,957,203 bp
Longest scaffold	2,051,674 bp
Shortest scaffold	1,007 bp
Number of scaffolds > 1K nt	2,050 (100.0%)
Number of scaffolds > 10K nt	1,832 (89.4%)
Number of scaffolds > 100K nt	386 (18.8%)
Number of scaffolds > 1M nt	4 (0.2%)
Number of scaffolds > 10M nt	0 (0.0%)
N50 scaffold length	125,649 bp
Number of N's	21, 965 *
Number of gaps	187 *

*(1 gap was 17,239 N's, and another gap was 1,243 N's. All other gaps were <100 N's.)

This brought the final assembly statistics to 93.6% BUSCO (insecta) complete, scaffold N50: 125,649bp, the longest scaffold: 2,051,674bp and 89.4% of scaffolds >10k in length (table 3). The final assembly is available under GenBank accession: GCA_018703685.1. Transcriptome sequences are available under accessions: HBWI01000001-HBWI01209903.

2.4.4 Annotation

Gene prediction with MAKER identified 15,102 protein-coding genes with the encoded proteins having a mean length of 464 amino acids. Of these, 12,949 (86%) had a match to NCBI's non-redundant (*nr*) database and 11,616 (77%) contained InterPro motifs, domains or signatures. In total, 13,112 (87%) were annotated with either blastp or InterPro and 10,192 were annotated with a GO ID. More information on the InterPro member database annotations is given in appendix ii. The longest protein found was an 'egf-like protein' at 14,628 amino acids. The resultant gene set was 84.5% BUSCO (insecta) complete.

From the Infernal tool inference of RNA alignments, a total of 791 non-coding RNA elements and 269 *cis*-regulatory elements were found in the genome (table 4).

Table 4. Number of ncRNAs predicted in the *Orius laevigatus* genome

ncRNA element	Number of elements
tRNA	503
rRNA	182
snRNA	53
miRNA	41
srpRNA	6
snoRNA	3
lncRNA	3

2.4.5 Repeat Annotation

Transposable and repetitive elements made up 27.07% of the assembled *O. laevigatus* genome (table 5) and the majority of these (20.4%) were unclassified repeats. This is close to the reported repeat content of other hemipteran species, for example: *Cimex lectularius* - 31.63% [86] and *Acyrthosiphon pisum* - 38% [15], an exception is *Rhodnius prolixus* which has an unusually low repeat content of 5.6% [87].

Table 5. Summary of transposable and repetitive elements in the *Orius laevigatus* genome

		Number of Elements	Length Occupied	Percentage of Sequence
SINES		705	59,683 bp	0.04%
LINES		3,309	1,556,653 bp	1.03%
	LINE1	0	0 bp	0.00%
	LINE2	496	257,681 bp	0.17%
	L3/CR1	2,310	890,133 bp	0.59%
LTR elements		959	501,171 bp	0.33%
DNA elements		5,490	1,715,984 bp	1.14%
	hAT-Charlie	784	222,164 bp	0.15%
	TcMar-Tigger	99	41,650 bp	0.03%
Unclassified		105,531	30,830,578 bp	20.42%
Total interspersed repeats		**NA**	**34,664,069 bp**	**22.96%**
Small RNA		127	35,035 bp	0.02%
Satellites		4,867	3,456,707 bp	2.29%
Simple repeats		30,022	2,273,603 bp	1.51%
Low complexity		7,742	444,236 bp	0.29%
Total:		**NA**	**42,285,278 bp**	27.07%

2.4.6 Mitochondrial Genome

A circularized mitochondrial genome of 16,246bp, assembled and annotated using MITOS2, consisted of 13 protein coding genes, 19 tRNA genes, 2 rRNA genes and an A+T rich region with a length of 1,460bp and an A+T content of 72.7% (figure 4). This closely matches the *Orius sauteri* mitochondrial genome, which is also 16,246bp and has 13 protein-coding genes, 22 tRNA genes, 2rRNA genes and an A+T rich region of 1,758 bp and an A+T content of 73.5% [60].

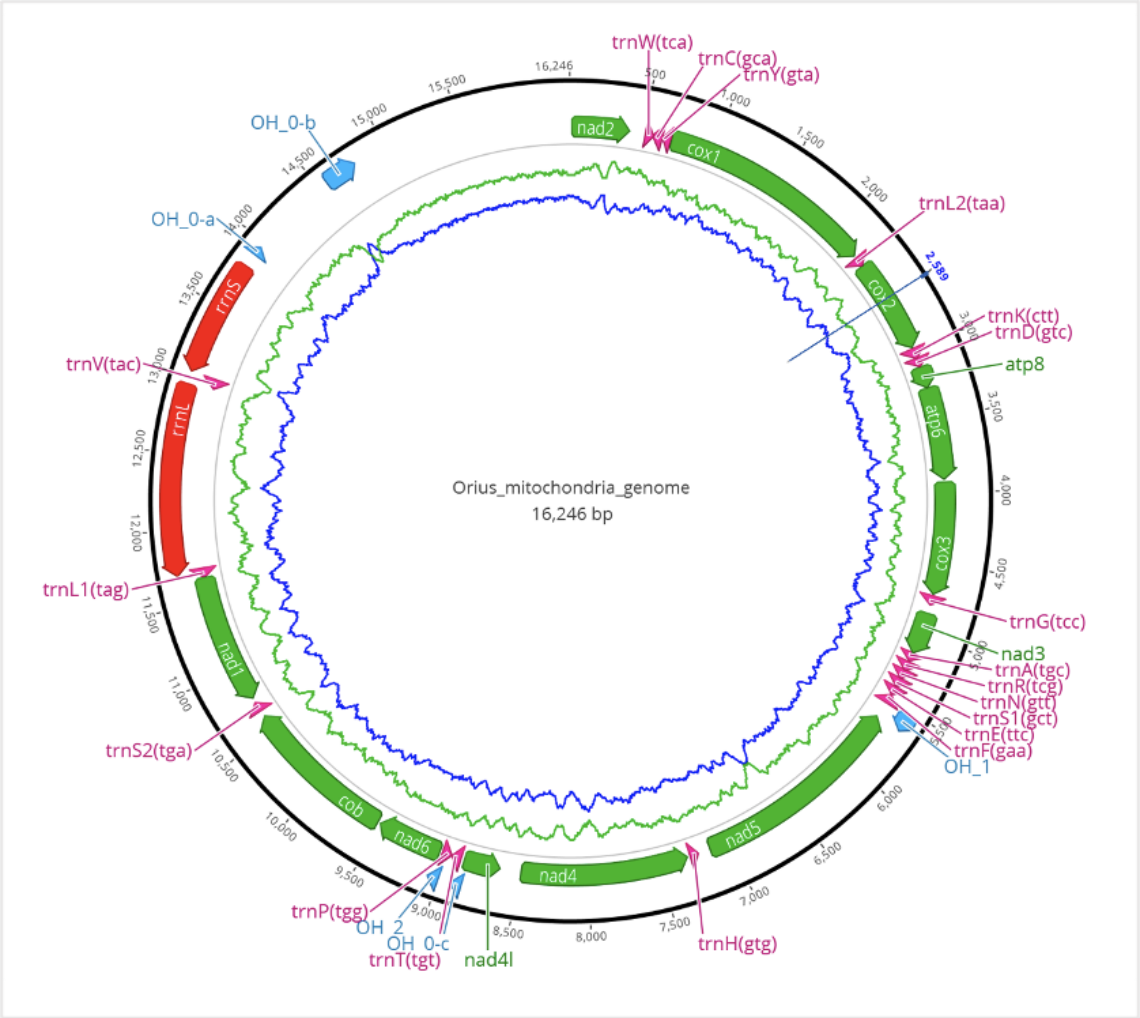


Figure 4. The mitochondrial genome for *Orius laevigatus,* visualised using Geneious and annotation track obtained using MITOS2. The innermost graphs represent AT content shown in green, and GC content shown in blue.

2.4.7 Phylogeny

OrthoFinder assigned 318,985 genes (88.8% of total) to 27,481 orthogroups. There were 1,621 orthogroups with all species present and 45 of these consisted entirely of single-copy genes. Phylogenetic analysis correctly clustered *O. laevigatus* within the hemipteran clade (figure 5) and identified *Cimex lectularius* as its closest relative.

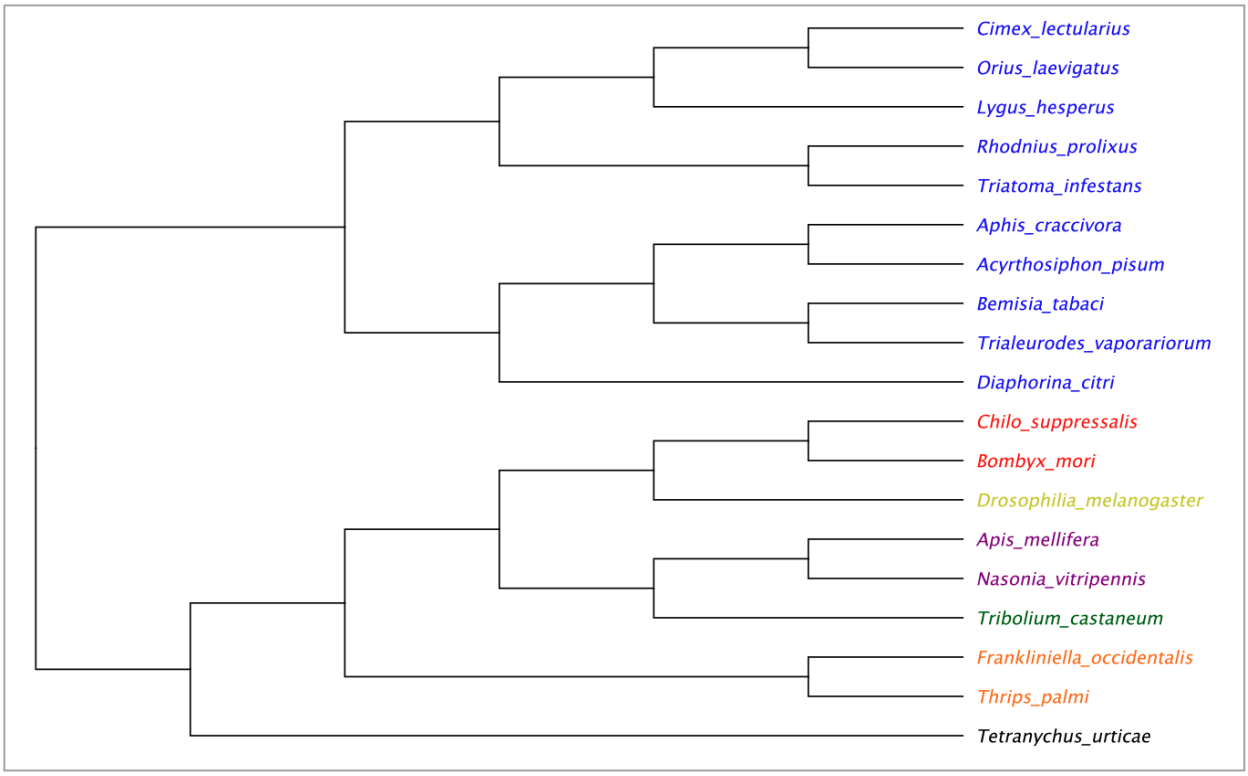


Figure 5. Phylogeny and divergence of Insecta. Nodes are coloured by order, blue=Hemiptera, red=Lepidoptera, yellow=Diptera, purple=Hymenoptera, green=Coleoptera, orange=Thysanoptera, black=Chelicerata. Produced using the STAG tree inference method and full proteomes of the following species: *C. lectularius:* PRJNA167477, *L. hesperus*: PRJNA284294, *R. prolixus:* PRJNA13648, *T. infestans*: PRJNA589079, *A. craccivora:* PRJNA558689, *A. pisum*: PRJNA13657, *B. tabaci*: PRJNA312470, *T. vaporariorum*: PRJNA553773, *D. citri*: PRJNA2944, *C. suppressalis*: PRJNA506136, *B. mori:* PRJNA205630, *D. melanogaster*: PRJNA13812, *A. mellifera*: PRJNA471592, *N. vitripennis*: PRJNA575073, *T. castaneum*: PRJNA12540, *F. occidentalis*: PRJNA203209, *T. palmi*: PRJNA607431, *T. urticae:* PRJNA315122.

2.4.8 Comparative Genomics

ABC Transporters

ATP-binding cassette transporters (ABCs), the largest known group of active transporters, can eliminate xenobiotic compounds - such as secondary metabolites produced by plants or insecticides - through translocation [36]. These transporters are subdivided into eight subfamilies: ABCA-H. ABCB, ABCC and ABCG are the subfamilies most associated with resistance to a variety of insecticides including pyrethroids, carbamates, organophosphates and neonicotinoids [88]. 41 of the 64 transporters in *O. laevigatus* belong to these 3 class-specific expansions (table 6) which could confer resistance to insecticides (a phylogenetic tree showing relationships of ABC transporters in *O. laevigatus* is included in appendix iii.

Table 6. Numbers of ABC transporter genes annotated in *Orius laevigatus* (this study), *Cimex lectularius* [89]**, *Lygus hesperus*** [90]**, *Frankliniella occidentalis*** [91]**, *Thrips palmi*** [92]**, *Aphis gossypii*** [93]**, *Trialeurodes vaporariorum*** [94]**, *Diuraphis noxia* and *Bemisia tabaci*** [95]

	O. laevigatus + close relatives			Crop pests					
	O. laevigatus	***C. lectularius***	***L. hesperus***	***F. occidentalis***	***T. palmi***	***D. noxia***	***A. gossypii***	***T. vaporariorum***	***B. tabaci***
ABCA	11	6	11	3	3	3	4	3	8
ABCB	9	7	6	5	4	6	5	9	3
ABCC	9	6	12	19	12	24	25	7	6
ABCD	1	2	2	2	2	3	2	4	2
ABCE	1	1	1	1	2	1	1	1	1
ABCF	5	4	3	3	3	3	4	3	3
ABCG	23	23	19	22	16	26	30	9	23
ABCH	2	2	11	13	7	11	0	9	9
Total	**64**	**51**	**65**	**70**	**49**	**77**	**71**	**45**	**55**

Table 6 shows a comparison of numbers of ABC transporter genes found in the current study with those reported for some pest species. The gene family expansions were generally seen in the ABCC and ABCG classes for all hemipteran species and slightly larger expansions were seen in some crop pests compared to *O. laevigatus* for the ABCC class, however, the expansions were of very similar size for both crop pests and *O. laevigatus* in the ABCG class. Overall, the total numbers of ABC transporter genes were similar across all the hemipteran species compared.

Glutathione S-Transferases

The glutathione S-transferases (GSTs) protein family is large and functionally diverse, and is known to confer resistance to all main insecticide classes. GST-mediated detoxification of insecticides takes place via several different mechanisms, including protecting against oxidative stress, binding and sequestration of the insecticide, and by catalysing the conjugation of glutathione to the insecticide to reduce their toxicity [37].

Table 7. Numbers of glutathione S-transferase genes annotated in *Orius laevigatus* (this study), *Cimex lectularius* [96]**, *Rhodnius prolixus, Triatoma Infestans*** [97]**, *Thrips palmi*** [92]**, *Myzus persicae, Acyrthosiphon pisum, Trialeurodes vaporariorum, Bemisia tabaci, Halyomorpha halys*** [98] **and *Murgantia histrionica*** [24]

	O. laevigatus + close relatives				Crop pests						
	O. laevigatus	**** C. lectularius***	***R. prolixus***	***T. infestans***	***T. palmi***	***M. persicae***	***A. pisum***	***T. vaporariorum***	***B. tabaci***	***H. halys***	***M. histrionica***
Delta	1	1	1	1	14	3	11	9	14	2	4
Epsilon	0	0	0	0	0	0	0	1	0	0	0
Omega	2	1	1	0	1	1	1	0	1	3	0
Sigma	16	5	7	9	6	12	5	3	6	19	25
Theta	1	2	3	2	1	1	2	0	0	3	2
Zeta	1	1	1	0	2	0	0	2	2	1	0
Microsomal	3	0	1	2	1	2	2	3	2	5	3
Total	**24**	**10**	**14**	**14**	**25**	**19**	**21**	**18**	**25**	**33**	**34**

* C. lectularius numbers may be an underestimate as sequencing coverage was low for this study.

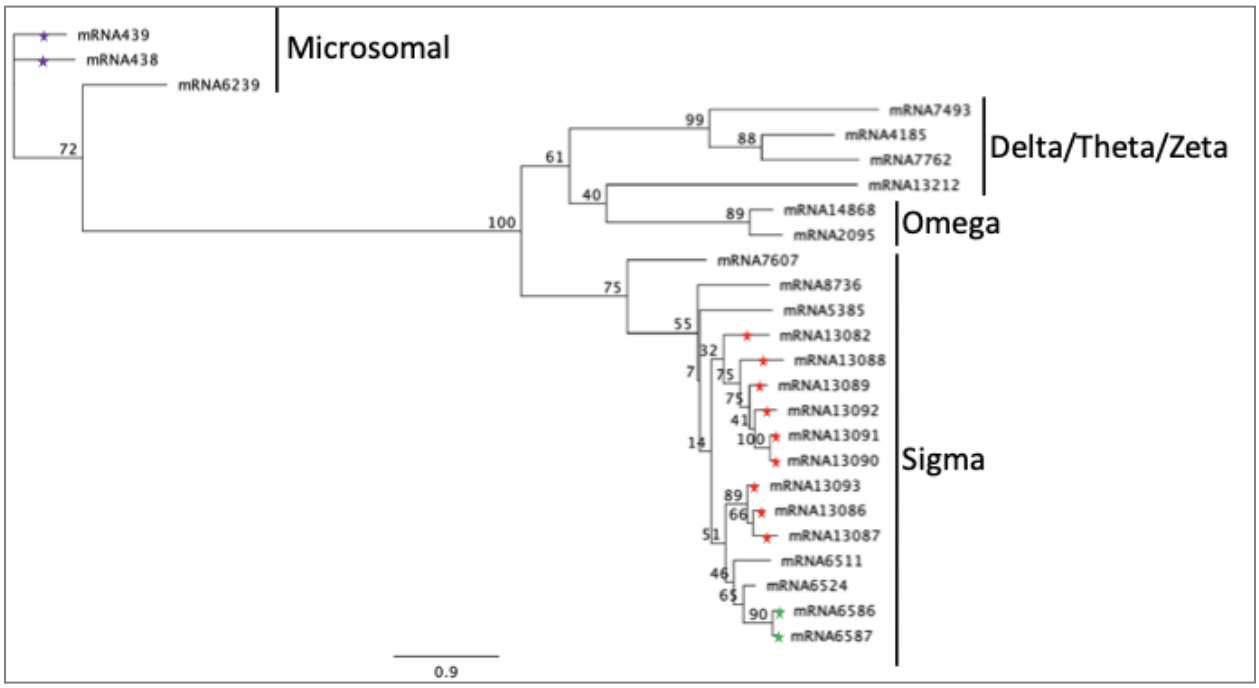


Figure 6. Phylogenetic tree of the *Orius laevigatus* glutathione S-transferases. Amino acid sequences were aligned using MAFFT and analysed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured stars on branches indicate tandem duplications, with each colour representing a different scaffold/set of tandem duplications in the *O. laevigatus* assembly.

The number of GST genes in *O. laevigatus* was fairly similar to other hemipteran close relatives, with the exception of the sigma class, which was notably larger (table 7). Of the 16 genes in the sigma class, 9 genes (mRNA13082 and mRNA13086-13093) were adjacent on the same scaffold, indicating a lineage specific expansion (fig 6). Expansions in this class have been reported in several hemipteran species including *Triatoma infestans, Myzus persicae, Halyomorpha halys* and *Murgantia histrionica* [23,24,97,99]. The sigma class has been found to play an important role in detoxification of organophosphorus insecticides in hemipteran species [100], therefore this expansion could potentially confer some tolerance to organophosphates in *O. laevigatus*. The delta and epsilon classes of GSTs are linked to insecticide resistance to pyrethroids [101,102]. The delta class is much larger in several crop pests compared to *O. laevigatus* and its close relatives which suggests these crop pests could

exhibit a higher level of delta class GST-mediated pyrethroid resistance. The epsilon class has previously been thought to be specific to Holometabola [103], and whilst *Trialeurodes vaporariorum* has a single member, the epsilon class is absent from all other Hemiptera species, suggesting potential epsilon class GST-mediated pyrethroid resistance is most likely absent in *O. laevigatus* and its close relatives, as well as most Hemiptera crop pests.

Carboxyl/cholinesterases

Many carboxyl/cholinesterases (CCEs) are linked to detoxification of organophosphorus, carbamate and pyrethroid insecticides and acetylcholinesterase (AChE) is the target for organophosphate and carbamate insecticides, with amino acid substitutions being linked to resistance [39]. 32 members of the CCE superfamily, including 1 AChE gene, were found in the *O. laevigatus* genome (table 8) which is a similar number to that reported for *Cimex lectularius*, which had 30 CCE genes and 2 AChE genes [89].

The dietary class of CCEs is involved in insecticide and xenobiotic detoxification [104]. *O. laevigatus* has no genes within this class, in line with *T. infestans* and *C. lectularius,* whereas the crop pest species (i.e. thrips, aphids and whiteflies) all have at least 5 members in this class (table 8). *R. prolixus* has 22 genes which have been classed as dietary; however this assignment was based heavily on a species-specific expansion which is characteristic of the dietary class. The real number of genes in the dietary class for *R. prolixus* may be 0, since this clade of 22 genes clusters with the hormone/semiochemical class in both the *R. prolixus* study [105] and this study (appendix iv). A lack of dietary esterases in *R. prolixus* would make sense, as *R. prolixus, C. lectularius* and *T. infestans* are all blood-sucking insects and do not require dietary esterases to process the secondary metabolites found in plants. This could also explain why *O. laevigatus,* a beneficial predator of crop pests in both nymph and adult life stages, does not require dietary esterases.

Table 8. Numbers of Carboxyl/cholinesterases annotated in Orius laevigatus (this study), *Cimex lectularius* [106]**, *Rhodnius prolixus*** [105]**, *Triatoma infestans*** [97]**, *Frankliniella occidentalis*** [91]**, *Myzus persicae*** [99]**, *Acyrthosiphon pisum, Bemisia tabaci*** [107] **and *Trialeurodes vaporariorum*** [108] **and their distribution across classes and clades.**

		O. laevigatus + close relatives				Crop pests				
		O. laevigatus	***C. lectularius***	***R. prolixus***	***T. infestans***	***F. occidentalis***	***M. persicae***	***A. pisum***	***T. vaporariorum***	***B. tabaci***
Dietary class		0	0*	22 (0)***	0	28	5	5	12	6
Hormone/semiochemical processing class		16	20*	9 (31)***	18	7	12	16	6	19
Neuro-developmental class	Glutactins	1	0*	2	0	2	0	0	1	1
	AChE	2	1*	2	1	2	3	2	2	4
	uncharacterised	1	1*	2	0	2	1	1	1	1
	gliotactin	3	0*	1	0	1	1	1	1	1
	neuroligin	8	0*	4	0	7	0	3	3	10
	neurotactin	1	0*	1	0	1	0	0	1	0
	Subtotal	16	2*	12	1	15	5	7	9	17
Total		**32**	**22* (30)****	**43**	**19**	**50**	**22**	**28**	**27**	**42**

* *C. lectularius* numbers may be an underestimate as sequencing coverage was low for this study, clade assignment was also uncertain as a result.

** A more recent study[89] found 30 CCE genes in *C. lectularius*, and is more likely to be a true representation, but they had not been assigned into classes/clades.

*** Numbers in brackets represent the possible true numbers of *R. prolixus* CCEs, based on a potential misassignment of 22 genes to the dietary class instead of the hormone/semiochemical processing class

The dietary class is involved in pyrethroid resistance [109]; however, *T. infestans* exhibits pyrethroid esterase activity despite having no dietary esterases [110]. *O. laevigatus* has also shown the ability to develop pyrethroid resistance - although the exact mechanism of this resistance is not yet known [111]. The hormone and semiochemical processing class is also involved in insecticide metabolism, due to the presence of β-esterases [112,113]. There may be

some redundancy in genes potentially involved in insecticide detoxification from the dietary and hormone/semiochemical processing classes. This might explain why only one of these classes shows an increased number of genes for each of these hemipteran species (table 8), as having increased numbers of both classes would be redundant, whilst very low numbers of both classes would be detrimental. The lack of the dietary class may therefore not impact the xenobiotic resistance abilities of *O. laevigatus,* as it has 16 genes within the hormone/semiochemical processing class.

The remaining CCEs in *O. laevigatus* belong to the neurodevelopmental class and include the neuroligins, gliotactins, glutactins and neurotactins, which are non-catalytic due to the lack of a critical serine residue. Acetylcholinesterase is the only protein in this class which has been linked to organophosphate resistance [114,115].

UDP-glycosyltransferases

UDP-glycosyltransferases (UGTs) are detoxification enzymes speculated to be involved in insecticide metabolism. Although the exact mechanisms of UGT-mediated resistance have not yet been identified, their upregulation has been shown in resistant strains of *P. xylostella* [35] and they have been linked to diamide resistance in *Chilo suppressalis* [116], neonicotinoid resistance in *Diaphorina citri* [117] and they also contribute to insecticide detoxification via the elimination of oxidative stress in *Apis cerana* [118].

The number of UGT genes in *O. laevigatus* was much lower than for other hemipteran species (table 9). The UGTs were submitted to Dr Michael Court for naming. Numbers of UGTs have been reported to be lower in non-phytophagous insects [92], which could explain the low numbers seen in *O. laevigatus* and *R. prolixus* compared to crop pests. This suggests that UGT-mediated detoxification may be lower in *O. laevigatus* than in crop pests.

Table 9. Numbers of UDP-glycosyltransferase genes found in *O. laevigatus* (this study), *Rhodnius prolixus, Tetranychus urticae, Nilaparvata lugens, Acyrthosiphon pisum, Bemisia tabaci* [19], *Myzus persicae* [119] and *Trialeurodes vaporariorum* [120].

	O. laevigatus + close relatives		Crop pests					
	O. laevigatus	***R. prolixus***	***T. urticae***	***N. lugens***	***M. persicae***	***A. pisum***	***T. vaporariorum***	***B. tabaci***
Total	10	16	81	20	101	72	55	76

Cytochrome P450s

Cytochrome P450s are a diverse superfamily capable of metabolizing a huge variety of endogenous and exogenous substrates. In insects they are associated with growth and development, metabolism of pesticides and plant toxins as well as the production and metabolism of insect hormones and pheromones. P450s are associated with resistance to insecticides from a variety of classes, including pyrethroids, carbamates and neonicotinoids. They are also linked to the activation of organophosphates and other pro-insecticides (a form of insecticide which is metabolized into an active form inside the host) [38]. Upregulation of P450s in insects has been shown to confer insecticide resistance [121–124], and conversely downregulation occurs in response to pro-insecticides [125,126].

A total of 58 full-length P450 genes were identified in the *O. laevigatus* genome, 11 P450 fragment genes were also found as well as 1 pseudogene. These sequences were named by Dr David Nelson using his in-house pipeline [84]. The majority of these genes (34) belonged to the diverse CYP3 class, which was a similar size to other hemipteran species (table 10).

Table 10. Total numbers of Cytochrome P450 genes annotated in *Orius laevigatus* (this study), *Cimex lectularius* [89], *Rhodnius prolixus, Triatoma infestans* [97], *Frankliniella occidentalis, Thrips palmi* [91], *Myzus persicae, Acyrthosiphon pisum* [99], *Trialeurodes vaporariorum* [94], *Bemisia tabaci* [127], *Halyomorpha halys* [23] and *Murgantia histrionica* [24].

	O. laevigatus + close relatives				Crop pests							
	O. laevigatus	***C. lectularius***	***R. prolixus***	***T. infestans***	***F. occidentalis***	***T. palmi***	***M. persicae***	***A. pisum***	***T. vaporariorum***	***B. tabaci***	***H. halys***	***M. histrionica***
CYP2	6	6	7	1	12	12	3	10	7	18	6	7
CYP3	34(41)*	36	55	65	22	26	63	33	41	76	84	43
CYP6	11	10**	8	15	18	-	-	29	34	47	-	-
CYP9	0	0**	0	0	0	-	-	0	0	0	-	-
Other	23	26**	47	50	4	-	-	4	7	-	-	-
CYP4	13(17)*	11	49	22	37	42	48	32	25	73	45	30
Mito	5	6	8	6	10	11	1	8	7	4	6	6
Total	**58**	**59**	**119**	**94**	**81**	**91**	**115**	**83**	**80**	**171**	**141**	**86**

*Values in brackets represent total gene numbers including partial and fragment genes. For other species partial and fragment p450 genes were excluded in cases where they were listed as such - some may remain in the counts if official naming and curation had not taken place.

**Values used are those from[89], but values differed by study[128] identified 5 CYP9s, 35 CYP6s and 5 others;[129] identified 0 CYP9s, 8 CYP6s and 15 others (these were also officially named by David Nelson)

The CYP3 clade is currently the P450 clade most associated with insecticide resistance - notably the CYP6 and CYP9 families [130]. Interestingly the CYP9 family was not present in *O. laevigatus*, as found for *T. infestans*, *R. prolixus, M. histrionica* and *H. halys* [23,24,97]. Further investigation into the assignment of classes within the CYP3 clade suggests the lack of the CYP9 class could be a common feature within Hemiptera (table 10).

Expansion of the CYP397 gene family was seen in *O. laevigatus*, (fig 7) with 7 full-length CYP397 genes and 1 fragment. CYP397B1, CYP397B2, CYP397B6 and CYP397C1 were directly adjacent on the same scaffold, indicating tandem duplications. Sequence similarity of the CYP397 genes to CYP397B1 ranged from 52% to 86%, which suggests a variation in ages of

these tandem duplications. *Cimex lectularius* also showed an increased copy number of CYP397 with 6 copies (A1-A6) [89]. CYP397A1 is significantly upregulated (>36 fold) in pyrethroid-resistant strains of *C. lectularius* [129], therefore the expansion of this gene family could potentially confer some tolerance to pyrethroids in *O. laevigatus*.

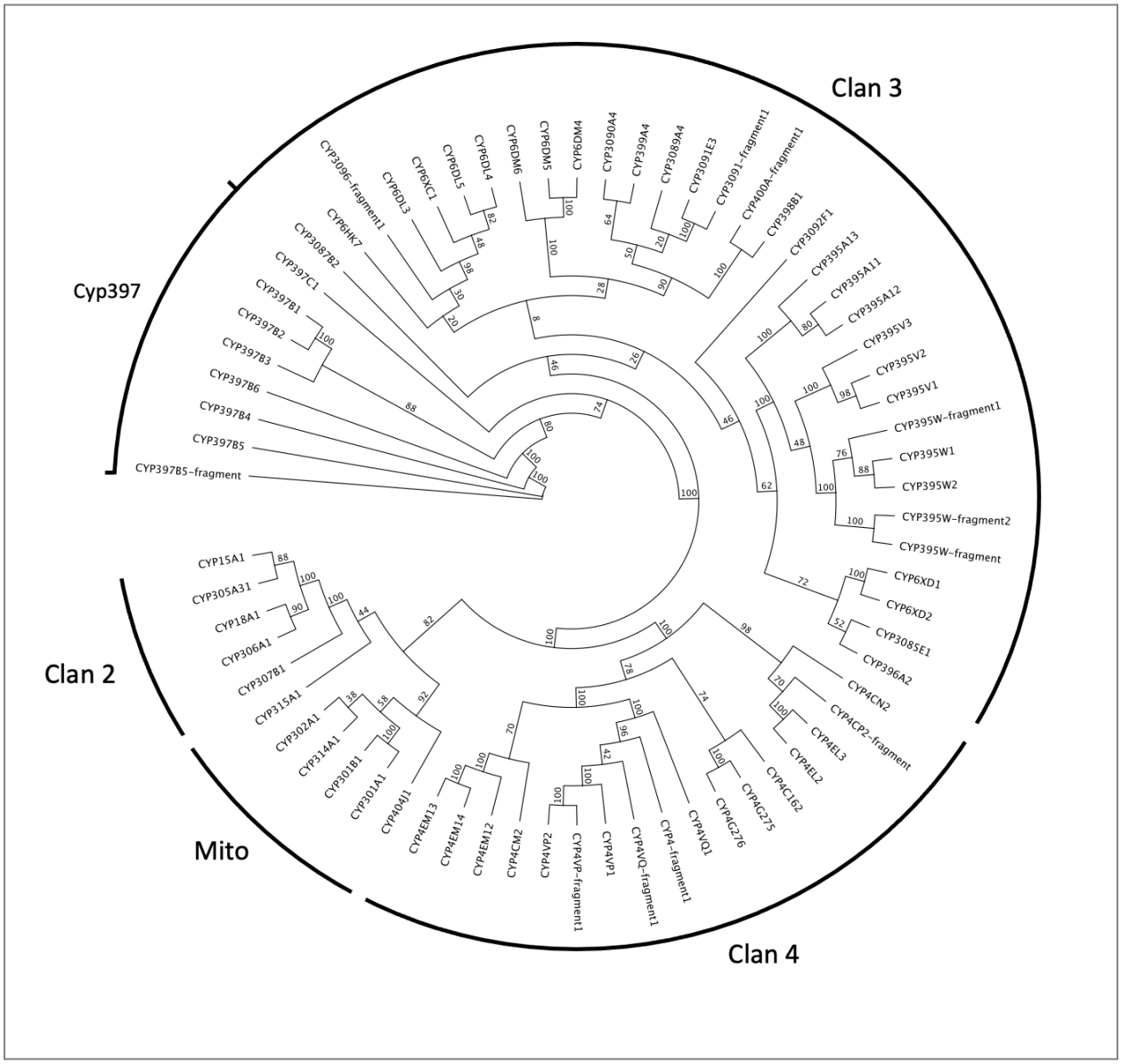


Figure 7. Phylogenetic tree of the *Orius laevigatus* cytochrome P450s. The Cyp397 gene family is a member of clan 3. Amino acid sequences were aligned using MAFFT and analysed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates.

A previous study [131] looked at the effect of insecticide synergists on *Orius tristicolor* (another minute pirate bug of the Anthocoridae family), and found that PBO (an inhibitor of P450s and esterases) significantly increased the mortality rate when combined with indoxacarb (an oxadiazine insecticide). Whereas inhibition of solely GSTs or esterases did not reduce mortality. Upregulation of P450s, esterases and GSTs have all been seen in response to oxadiazines [132], therefore the fact that only P450 inhibition had an impact on mortality rate suggests P450s may be the primary detoxification mechanism of *O. laevigatus*.

Target site mutations

Point mutations resulting in amino acid substitutions in the target proteins of insecticides have been characterised in many insecticide resistant insect species, including in the sodium channel gene *para* which confers resistance to pyrethroids [133]; the acetylcholinesterase-1 (ace-1) enzyme associated with organophosphate resistance [134] and the acetyl-coenzyme A Carboxylase (ACC) enzyme linked to keto-enol (spirotetramat) resistance [135]. Despite these mutations having been observed in a variety of hemipteran crop pests, none were observed in this *O. laevigatus* assembly. Although, it is important to note that the *O. laevigatus* assembly was a consensus of ~1000 individuals, therefore differences in target sites would likely only be apparent if they were present in the majority of the population. Overall, tolerance of insecticides by *O. laevigatus* resulting from target site differences seems unlikely compared to what is seen in crop pests, where there has been intensive selection pressure.

The ryanodine receptor (RyR) is the target of diamide insecticides, and two target site resistance mutations conferring amino acid substitutions (I4790M and G4946E - numbering according to *Plutella xylostella*, PxRyR) have been identified in lepidopteran pests [21,136]. Interestingly, *O. laevigatus* has the I4790M substitution which has been shown to confer varying levels of resistance to diamides. This point mutation was also present in other hemipteran species as shown in figure 8 (except for *Lygus hesperus* which had an I>L mutation).

I4790M has been detected in lepidopteran populations across the globe and is considered to be a 'selectivity switch' for diamides [137]. *O. tristicolor* showed high levels of resistance to chlorantraniliprole (a diamide insecticide) with <5% mortality [131] with the I4790M substitution being the main cause [138]. It is therefore possible that I4790M may confer some tolerance to diamides in *O. laevigatus,* and indeed, diamide resistance has been reported in *O. laevigatus* [139]. However, I4790M could potentially also confer diamide tolerance in crop pests - diamide resistance has already been shown in *F. occidentalis* [139]. Therefore this would likely not be an exploitable difference for IPM strategies.

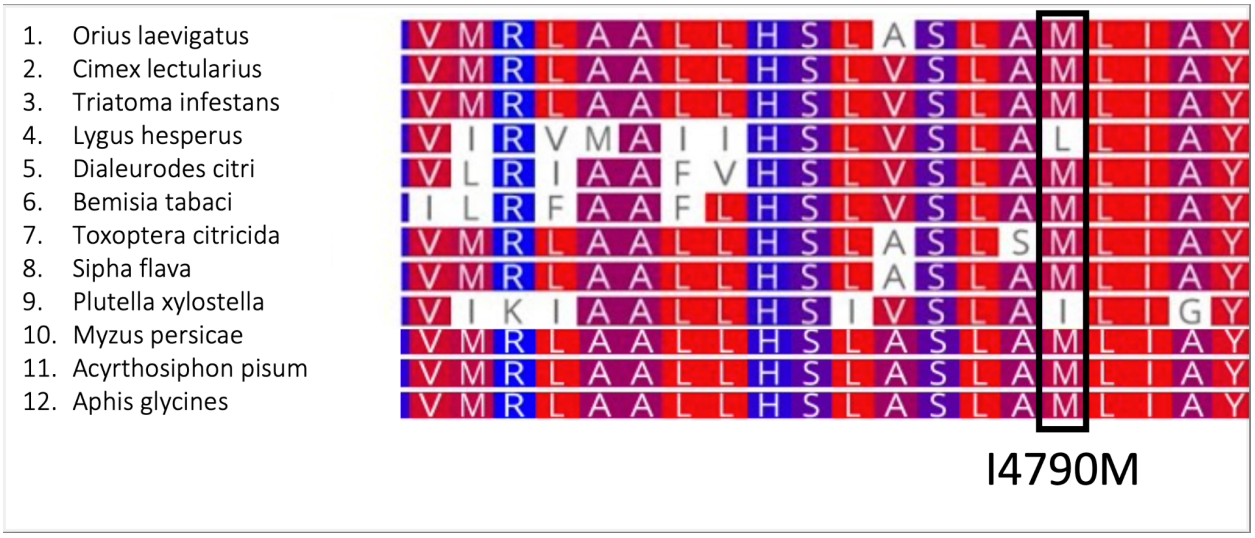


Figure 8. An alignment of amino acid sequences to compare transmembrane domain 3 of the conserved ryanodine receptor (RyR) from *Plutella xylostella* with hemipteran species. The box indicates the I4790M RyR point mutation linked to diamide resistance (numbered according to PxRyR). A strain of *P. xylostella* without the point mutation was used for the alignment. RyR sequences were obtained from the UniProt database, excluding *Orius laevigatus*. UniProt entry names are as follows: *Cimex lectularius*: A0A7E4RNZ4_CIMLE, *Triatoma infestans*: A0A023F678_TRIIF, *Lygus hesperus*: A0A146M8X0_LYGHE, *Dialeurodes citri*: A0A141BN13_DIACT, *Bemisia tabaci*: A0A1U9JHP1_BEMTA, *Toxoptera citricida*: A0A0H3XSN1_TOXCI, *Sipha flava:* A0A2S2QG40_9HEMI, *Plutella xylostella*: I3NWV8_PLUXY, *Myzus persicae:* A0A0A7RS32_MYZPE, *Acyrthosiphon pisum*: X1WXB1_ACYPI, *Aphis glycines:* A0A6G0U418_APHGL.

2.5 Conclusions

PacBio long-read technology combined with low error-rate short-read Illumina sequencing enabled the production of a high-quality genome and mitochondrial assembly for *O. laevigatus*. Whilst genome continuity may not be as good as an assembly generated from a single insect, the genome completeness is still of a sufficient quality to aid with comparative and functional genomics analyses and provides a useful first reference genome for the *Anthocoridae* family. An experimental estimate to confirm genome size and Hi-C based scaffolding would likely be the next best steps to significantly improve this genome in the future.

Comparative analyses of *O. laevigatus* with hemipteran crop pests showed evidence of possible differences in xenobiotic tolerance, including a potential increase in GST-mediated tolerance of organophosphates in *O. laevigatus*, whilst GST-mediated pyrethroid tolerance may be more prevalent in crop pests. There may also be less UGT-mediated tolerance to diamides and neonicotinoids in *O. laevigatus* compared to crop pests - although, the I4790M target site mutation may confer some degree of diamide insensitivity to *O. laevigatus*.

A recent study shows that there is significant variation in the susceptibility of *O. laevigatus* to pyrethroids when a variety of wild and commercial populations are assessed [111]. This suggests that beneficial predators such as *O. laevigatus* are certainly capable of developing insecticide resistance, but a combination of factors result in resistance developing slower than in pest species. This could be due to beneficial predators having smaller population sizes, longer life cycles, less exposure to pesticides and a lack of continuous selection pressure - beneficial predators often need to be re-released each season as populations migrate to new areas in search of food sources. These differences will have resulted in a lesser degree of selection for resistance mechanisms in *O. laevigatus* and therefore any observed differences in potential sensitivity would only be at low levels. Further comparisons looking at differences in

gene expansions, expression levels and key target site mutations between resistant and susceptible strains of *O. laevigatus* would provide more concrete evidence for the findings in this study.

In conclusion, this study indicates differences in potential mechanisms of resistance between crop pests and *O. laevigatus* which could be exploited when selecting targeted insecticides. An increase in the number of pesticides which are safe for beneficial predators such as *O. laevigatus* would be of significant impact to pest management, especially at a time when the list of pesticides effective against crop pests is growing ever shorter. The findings also suggest that *O. laevigatus* has the ability to develop resistance to a variety of insecticides which could be used to our advantage through the selective breeding and selection of resistant strains of *O. laevigatus* for use in IPM strategies.

2.6 Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The genome assembly generated in this study, as well as the raw PacBio and Illumina data used, are available under the BioProject accessions: PRJNA721944 and PRJEB38143.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Pest Genomics Initiative, a collaborative project between Rothamsted Research, Bayer Crop Science and Syngenta AG. Rothamsted Research gratefully acknowledges strategic funding from the UK Biotechnology and Biological and Sciences Research Council (UKRI-BBSRC). The content is solely the responsibility of the authors and does not necessarily reflect the official views of the funding bodies.

Authors' contributions

The Pest Genomics Initiative (BB, CR, EG, KH-P, LF, RK and RN) devised the original conceptual ideas. EB performed the DNA and RNA extractions with assistance from MW. EB assembled and annotated the genome with guidance from RK and DH. EB performed the comparative analyses with input from RN. FM, RK and KH-P supervised the project. EB wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Acknowledgements go to Dr David R. Nelson of the International Committee on the Nomenclature for Cytochrome P450 Enzymes for classification of *Orius laevigatus* P450s and to Dr Michael H. Court of the UGT Nomenclature Committee for classification of *Orius laevigatus* UGTs.

2.7 Literature Cited

[1] Oerke E-C. Crop losses to pests. J Agric Sci 2006;144:31–43.
[2] Geiger F, Bengtsson J, Berendse F, Weisser WW, Emmerson M, Morales MB, et al. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. Basic Appl Ecol 2010;11:97–105.
[3] Bottrell DG, Schoenly KG. Resurrecting the ghost of green revolutions past: The brown planthopper as a recurring threat to high-yielding rice production in tropical Asia. J Asia Pac Entomol 2012;15:122–40.
[4] Debach P, Rosen D. Biological control by natural enemies (second edition). J Trop Ecol 1992;8:216–216.
[5] Hernandez LM. A review of the economically important species of the genus Orius (Heteroptera: Anthocoridae) in East Africa. Journal of Natural History 1999;33:543–68. https://doi.org/10.1080/002229399300245.
[6] Kelton LA. The Anthocoridae of Canada and Alaska: Heteroptera, Anthocoridae. Agriculture Canada; 1978.
[7] Malais M. Knowing and Recognizing: The Biology of Glasshouse Pests and Their Natural Enemies. 1992.
[8] Chambers RJ, Long S, Helyer NL. Effectiveness of Orius laevigatus(Hem.: Anthocoridae) for the control of Frankliniella occidentalis on cucumber and pepper in the UK. Biocontrol Sci Technol 1993;3:295–307.
[9] Sanchez JA, Sanchez JA, Lacasa A. Modelling population dynamics of Orius laevigatus and O. albidipennis (Hemiptera: Anthocoridae) to optimize their use as biological control agents of Frankliniella occidentalis (Thysanoptera: Thripidae). Bull Entomol Res 2002;92:77–88.
[10] Tommasini MG, Maini S, Nicoli G. Advances in the integrated pest management in protected-eggplant crops by seasonal inoculative releases of Orius laevigatus. Adv Hortic Sci 1997;11:182–8.
[11] Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. Genome Biol Evol 2016;8:1762–75.
[12] Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. Science 2010;327:343–8.
[13] Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, et al. Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles. Nat Commun 2018;9:1–13.
[14] Tribolium Genome Sequencing Consortium, Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, et al. The genome of the model beetle and pest Tribolium castaneum. Nature 2008;452:949–55.
[15] International Aphid Genomics Consortium. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol 2010;8:e1000313.
[16] You M, Yue Z, He W, Yang X, Yang G, Xie M, et al. A heterozygous moth genome provides insights into herbivory and detoxification. Nat Genet 2013;45:220–5.
[17] Xue J, Zhou X, Zhang C-X, Yu L-L, Fan H-W, Wang Z, et al. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genome Biol 2014;15:521.
[18] Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, et al. The genome of Diuraphis noxia , a global aphid pest of small grains. BMC Genomics 2015;16:1–16.
[19] Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol 2016;14:1–15.
[20] Wenger JA, Cassone BJ, Legeai F, Johnston JS, Bansal R, Yates AD, et al. Whole genome sequence of the soybean aphid, Aphis glycines. Insect Biochem Mol Biol 2020;123:102917.
[21] Wang X, Cao X, Jiang D, Yang Y, Wu Y. CRISPR/Cas9 mediated ryanodine receptor I4790M knockin confers unequal resistance to diamides in Plutella xylostella. Insect Biochem Mol Biol 2020;125:103453.
[22] Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). Sci Rep 2018;8:1–18.
[23] Sparks ME, Bansal R, Benoit JB, Blackburn MB, Chao H, Chen M, et al. Brown marmorated stink bug, Halyomorpha halys (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. BMC Genomics 2020;21:227.
[24] Sparks ME, Rhoades JH, Nelson DR, Kuhar D, Lancaster J, Lehner B, et al. A Transcriptome Survey Spanning Life Stages and Sexes of the Harlequin Bug, Murgantia histrionica. Insects 2017;8. https://doi.org/ 10.3390/insects8020055.
[25] Cao C, Sun L, Wen R, Shang Q, Ma L, Wang Z. Characterization of the transcriptome of the Asian gypsy moth Lymantria dispar identifies numerous transcripts associated with insecticide resistance. Pestic Biochem Physiol 2015;119:54–61.
[26] Sparks ME, Nelson DR, Haber AI, Weber DC, Harrison RL. Transcriptome Sequencing of the Striped Cucumber Beetle, Acalymma vittatum (F.), Reveals Numerous Sex-Specific Transcripts and Xenobiotic Detoxification Genes. BioTech 2020;9:21.
[27] European Commission. Directive 2009/128/EC on the sustainable use of pesticides. Official Journal of the European Union; 2009. https://doi.org/10.2861/78.
[28] Cameron PJ, Walker GP, Hodson AJ, Kale AJ, Herman TJB. Trends in IPM and insecticide use in processing tomatoes in New Zealand. Crop Prot 2009;28:421–7.
[29] Kranthi KR, Russell DA. Changing Trends in Cotton Pest Management. In: Peshin R, Dhawan AK, editors. Integrated Pest Management: Innovation-Development Process: Volume 1, Dordrecht: Springer Netherlands; 2009, p. 499–541.

[30] Meissle M, Mouron P, Musa T, Bigler F, Pons X, Vasileiadis VP, et al. Pests, pesticide use and alternative options in European maize production: current status and future prospects. J Appl Entomol 2009;134:357–75.
[31] Hillocks RJ. Farming with fewer pesticides: EU pesticide review and resulting challenges for UK agriculture. Crop Prot 2012;31:85–93.
[32] Lechenet M, Dessaint F, Py G, Makowski D, Munier-Jolain N. Reducing pesticide use while preserving crop productivity and profitability on arable farms. Nature Plants 2017;3:1–6.
[33] Lattin JD. BIONOMICS OF THE ANTHOCORIDAE. Annu Rev Entomol 1999;44:207–31.
[34] Heckel DG. Insecticide Resistance After Silent Spring. Science 2012;337:1612–4.
[35] Li X, Shi H, Gao X, Liang P. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in Plutella xylostella (L.). Pest Manag Sci 2018;74:695–704.
[36] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[37] Pavlidi N, Vontas J, Van Leeuwen T. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. Curr Opin Insect Sci 2018;27:97–102.
[38] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[39] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[40] Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Current Opinion in Insect Science 2019;31:131–8.
[41] Roderick GK, Navajas M. Genes in new environments: genetics and evolution in biological control. Nat Rev Genet 2003;4:889–99.
[42] Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 2011;27:764–70.
[43] Rhyker Ranallo-Benavidez T, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 2020;11:1–10.
[44] Andrews S. FastQC n.d. https://github.com/s-andrews/FastQC (accessed April 20, 2021).
[45] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
[46] Bradnam K. A script to calculate a basic set of metrics from a genome assembly 2011. http://korflab.ucdavis.edu/Datasets/Assemblathon/Assemblathon2/Basic_metrics/assemblathon_stats.pl.
[47] Simão 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.
[48] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 2019;37:540–6.
[49] Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. Assembly of long error-prone reads using de Bruijn graphs. Proc Natl Acad Sci U S A 2016;113:E8396–405.
[50] Song L, Shankar DS, Florea L. Rascaf: Improving Genome Assembly with RNA Sequencing Data. Plant Genome 2016;9. https://doi.org/10.3835/plantgenome2016.03.0027.
[51] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 2017;27:722–36.
[52] Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 2013;10:563–9.
[53] Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 2016;13:1050–4.
[54] Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 2016;44:e147.
[55] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 2014;9:e112963.
[56] Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res 2016;44:e113–e113.
[57] Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol 2018;14:e1005944.
[58] O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 2016;44:D733–45.
[59] Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res 2007;17:377–86.
[60] Du B-Z, Niu F-F, Wei S-J. The complete mitochondrial genome of the predatory bug Orius sauteri (Poppius) (Hemiptera: Anthocoridae). Mitochondrial DNA Part A 2016;27:777–8.
[61] Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 2011;12:1–14.
[62] Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 2004;32:W309–12.
[63] Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 2005;33:6494–506.
[64] Solovyev V. Statistical approaches in eukaryotic gene prediction. In: D.J. Balding, M. Bishop, C. Cannings, editor. Handbook of Statistical Genetics, Chichester: John Wiley & Sons Ltd; 2001, p. 83–127.

[65] Haas BJ, 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;9:1–22.
[66] Smit AFA, Hubley R. RepeatModeler Open-1.0 2008-2015. http://www.repeatmasker.org.
[67] Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics 2013;29:2487–9.
[68] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42:D222–30.
[69] Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0 2013-2015. http://www.repeatmasker.org.
[70] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–60.
[71] Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33:290–5.
[72] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644–52.
[73] Gilbert D. EvidentialGene - Evidence Directed Gene Construction for Eukaryotes 2013. https://sourceforge.net/projects/evidentialgene/.
[74] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36:3420–35.
[75] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 2013;29:2933–5.
[76] Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 2013;69:313–9.
[77] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 2019;20:1–14.
[78] Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. bioRxiv 2018:267914. https://doi.org/ 10.1101/267914.
[79] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–80.
[80] Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059–66.
[81] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.
[82] Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol 2008;25:1307–20.
[83] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.
[84] Nelson DR. The Cytochrome P450 Homepage. Hum Genomics 2009;4:1–7.
[85] UGT Committee. UGT Committee Home. UGT Committee Home n.d. https://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage (accessed March 25, 2021).
[86] Rosenfeld JA, Reeves D, Brugler MR, Narechania A, Simon S, Durrett R, et al. Genome assembly and geospatial phylogenomics of the bed bug Cimex lectularius. Nat Commun 2016;7:1–10.
[87] Mesquita RD, Vionette-Amaral RJ, Lowenberger C, Rivera-Pomar R, Monteiro FA, Minx P, et al. Genome of Rhodnius prolixus, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc Natl Acad Sci U S A 2015;112:14936–41.
[88] Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 2014;45:89–110.
[89] 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;7:1–10.
[90] Joe Hull J, Chaney K, Geib SM, Fabrick JA, Brent CS, Walsh D, et al. Transcriptome-Based Identification of ABC Transporters in the Western Tarnished Plant Bug Lygus hesperus. PLoS One 2014;9:e113046.
[91] Rotenberg D, Baumann AA, Ben-Mahmoud S, Christiaens O, Dermauw W, Ioannidis P, et al. Genome-enabled insights into the biology of thrips as crop pests. BMC Biol 2020;18:142.
[92] Guo S-K, Cao L-J, Song W, Shi P, Gao Y-F, Gong Y-J, et al. Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance. Mol Ecol Resour 2020;20:1110–25.
[93] Pan Y, Zeng X, Wen S, Gao X, Liu X, Tian F, et al. Multiple ATP-binding cassette transporters genes are involved in thiamethoxam resistance in Aphis gossypii glover. Pestic Biochem Physiol 2020;167:104558.
[94] Pym A, Singh KS, Nordgren Å, Emyr Davies TG, Zimmer CT, Elias J, et al. Host plant adaptation in the polyphagous whitefly, Trialeurodes vaporariorum , is associated with transcriptional plasticity and altered sensitivity to insecticides. BMC Genomics 2019;20:1–19.
[95] Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, Bemisia tabaci. BMC Genomics 2017;18:1–16.
[96] Adelman ZN, Kilcullen KA, Koganemaru R, Anderson MAE, Anderson TD, Miller DM. Deep sequencing of pyrethroid-resistant bed bugs reveals multiple mechanisms of resistance within a single population. PLoS One 2011;6:e26228.
[97] Traverso L, Lavore A, Sierra I, Palacio V, Martinez-Barnetche J, Latorre-Estivalis JM, et al. Comparative and functional triatomine genomics reveals reductions and expansions in insecticide resistance-related gene families. PLoS Negl Trop Dis 2017;11:e0005313.
[98] Aidlin Harari O, Santos-Garcia D, Musseri M, Moshitzky P, Patel M, Visendi P, et al. Molecular Evolution of the Glutathione S-Transferase Family in the Bemisia tabaci Species Complex. Genome Biol Evol 2020;12:3857–72.

[99] Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol 2010;19 Suppl 2:155–64.
[100] Gawande ND, Subashini S, Murugan M, Subbarayalu M. Molecular screening of insecticides with sigma glutathione S-transferases (GST) in cotton aphid Aphis gossypii using docking. Bioinformation 2014;10:679–83.
[101] Lumjuan N, Rajatileka S, Changsom D, Wicheer J, Leelapat P, Prapanthadara L-A, et al. The role of the Aedes aegypti Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. Insect Biochem Mol Biol 2011;41:203–9.
[102] Vontas JG, Small GJ, Hemingway J. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens. Biochem J 2001;357:65–72.
[103] Friedman R. Genomic organization of the glutathione S-transferase family in insects. Mol Phylogenet Evol 2011;61:924–32.
[104] Oakeshott J, Claudianos C, Campbell PM, Newcomb RD, Russell RJ. Biochemical genetics and genomics of insect esterases. In: Gilbert LI, Gill SS, Iatrou K, editors. Insect Pharmacology: Channels, Receptors, Toxins and Enzymes, Elsevier; 2010, p. 229–81.
[105] Schama R, Pedrini N, Juárez MP, Nelson DR, Torres AQ, Valle D, et al. Rhodnius prolixus supergene families of enzymes potentially associated with insecticide resistance. Insect Biochem Mol Biol 2016;69:91–104.
[106] Adelman ZN, Kilcullen KA, Koganemaru R, Anderson MAE, Anderson TD, Miller DM. Deep sequencing of pyrethroid-resistant bed bugs reveals multiple mechanisms of resistance within a single population. PLoS One 2011;6:e26228.
[107] Xia J, Xu H, Yang Z, Pan H, Yang X, Guo Z, et al. Genome-Wide Analysis of Carboxylesterases (COEs) in the Whitefly, (Gennadius). Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20204973.
[108] Karatolos N. Molecular mechanisms of insecticide resistance in the greenhouse whitefly, Trialeurodes vaporariorum. PhD. University of Exeter, 2011. https://doi.org/ https://ore.exeter.ac.uk/repository/bitstream/handle/10036/3350/KaratolosN.pdf.
[109] Flores AE, Albeldaño-Vázquez W, Salas IF, Badii MH, Becerra HL, Garcia GP, et al. Elevated α-esterase levels associated with permethrin tolerance in Aedes aegypti (L.) from Baja California, Mexico. Pestic Biochem Physiol 2005;82:66–78.
[110] Orihuela PLS, Vassena CV, Zerba EN, Picollo MI. Relative Contribution of Monooxygenase and Esterase to Pyrethroid Resistance in Triatoma infestans (Hemiptera: Reduviidae) from Argentina and Bolivia. J Med Entomol 2014;45:298–306.
[111] Balanza V, Mendoza JE, Cifuentes D, Bielza P. Selection for resistance to pyrethroids in the predator Orius laevigatus. Pest Manag Sci 2021;77:2539–46.
[112] Ganesh KN, Vijayan VA, Urmila J, Gopalan N, Prakash S. Role of esterases and monooxygenase in the deltamethrin resistance in Anopheles stephensi Giles (1908), at Mysore. Indian J Exp Biol 2002;40:583–8.
[113] Prasad KM, Raghavendra K, Verma V, Velamuri PS, Pande V. Esterases are responsible for malathion resistance in *Anopheles stephensi*: A proof using biochemical and insecticide inhibition studies. J Vector Borne Dis 2017;54:226.
[114] Jyoti, Singh NK, Singh H, Singh NK, Rath SS. Multiple mutations in the acetylcholinesterase 3 gene associated with organophosphate resistance in Rhipicephalus (Boophilus) microplus ticks from Punjab, India. Vet Parasitol 2016;216:108–17.
[115] Zhang Y, Li S, Xu L, Guo HF, Zi J, Wang L, et al. Overexpression of carboxylesterase-1 and mutation (F439H) of acetylcholinesterase-1 are associated with chlorpyrifos resistance in Laodelphax striatellus. Pestic Biochem Physiol 2013;106:8–13.
[116] Zhao J, Xu L, Sun Y, Song P, Han Z. UDP-Glycosyltransferase Genes in the Striped Rice Stem Borer, (Walker), and Their Contribution to Chlorantraniliprole Resistance. Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20051064.
[117] Tian F, Wang Z, Li C, Liu J, Zeng X. UDP-Glycosyltransferases are involved in imidacloprid resistance in the Asian citrus psyllid, Diaphorina citri (Hemiptera: Lividae). Pestic Biochem Physiol 2019;154:23–31.
[118] Cui X, Wang C, Wang X, Li G, Liu Z, Wang H, et al. Molecular Mechanism of the UDP-Glucuronosyltransferase 2B20-like Gene (AccUGT2B20-like) in Pesticide Resistance of Apis cerana cerana. Front Genet 2020;11:592595.
[119] Pan Y, Xu P, Zeng X, Liu X, Shang Q. Characterization of UDP-Glucuronosyltransferases and the Potential Contribution to Nicotine Tolerance in Myzus persicae. Int J Mol Sci 2019;20:3637.
[120] Xie W, He C, Fei Z, Zhang Y. Chromosome-level genome assembly of the greenhouse whitefly (Trialeurodes vaporariorum Westwood). Mol Ecol Resour 2020;20:995–1006.
[121] Karunker I, Benting J, Lueke B, Ponge T, Nauen R, Roditakis E, et al. Over-expression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). Insect Biochem Mol Biol 2008;38:634–44.
[122] Liang X, Xiao D, He Y, Yao J, Zhu G, Zhu KY. Insecticide-mediated up-regulation of cytochrome P450 genes in the red flour beetle (Tribolium castaneum). Int J Mol Sci 2015;16:2078–98.
[123] Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. PLoS Genet 2010;6:e1000999.
[124] Yang T, Liu N. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, Culex quinquefasciatus. PLoS One 2011;6:e29418.
[125] Main BJ, Everitt A, Cornel AJ, Hormozdiari F, Lanzaro GC. Genetic variation associated with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii. Parasit Vectors 2018;11:1–9.
[126] Vlogiannitis S, Mavridis K, Dermauw W, Snoeck S, Katsavou E, Morou E, et al. Reduced proinsecticide activation by cytochrome P450 confers coumaphos resistance in the major bee parasite Varroa destructor. Proc Natl Acad Sci U S A 2021;118. https://doi.org/10.1073/pnas.2020380118.
[127] Ilias A, Lagnel J, Kapantaidaki DE, Roditakis E, Tsigenopoulos CS, Vontas J, et al. Transcription analysis of neonicotinoid resistance in Mediterranean (MED) populations of B. tabaci reveal novel cytochrome P450s, but no nAChR mutations associated with the phenotype. BMC Genomics 2015;16:1–23.
[128] Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O. Transcriptomics of the Bed Bug (Cimex lectularius). PLoS One

2011;6:e16336.
[129] Zhu F, Gujar H, Gordon JR, Haynes KF, Potter MF, Palli SR. Bed bugs evolved unique adaptive strategy to resist pyrethroid insecticides. Sci Rep 2013;3:1–8.
[130] Feyereisen R. Evolution of insect P450. Biochem Soc Trans 2006;34:1252–5.
[131] Pereira RR, Picanço MC, Santana PA Jr, Moreira SS, Guedes RNC, Corrêa AS. Insecticide toxicity and walking response of three pirate bug predators of the tomato leaf minerTuta absoluta. Agric For Entomol 2014;16:293–301.
[132] Shi L, Shi Y, Zhang Y, Liao X. A systemic study of indoxacarb resistance in Spodoptera litura revealed complex expression profiles and regulatory mechanism. Sci Rep 2019;9:1–13.
[133] Martinez-Torres D, Foster SP, Field LM, Devonshire AL, Williamson MS. A sodium channel point mutation is associated with resistance to DDT and pyrethroid insecticides in the peach-potato aphid, Myzus persicae (Sulzer) (Hemiptera: Aphididae). Insect Mol Biol 1999;8:339–46.
[134] Alon M, Alon F, Nauen R, Morin S. Organophosphates' resistance in the B-biotype of Bemisia tabaci (Hemiptera: Aleyrodidae) is associated with a point mutation in an ace1-type acetylcholinesterase and overexpression of carboxylesterase. Insect Biochem Mol Biol 2008;38:940–9.
[135] Lueke B, Douris V, Hopkinson JE, Maiwald F, Hertlein G, Papapostolou K-M, et al. Identification and functional characterization of a novel acetyl-CoA carboxylase mutation associated with ketoenol resistance in Bemisia tabaci. Pestic Biochem Physiol 2020;166:104583.
[136] Zuo Y-Y, Ma H-H, Lu W-J, Wang X-L, Wu S-W, Nauen R, et al. Identification of the ryanodine receptor mutation I4743M and its contribution to diamide insecticide resistance in Spodoptera exigua (Lepidoptera: Noctuidae). Insect Sci 2020;27:791–800.
[137] Richardson EB, Troczka BJ, Gutbrod O, Davies TGE, Nauen R. Diamide resistance: 10 years of lessons from lepidopteran pests. J Pest Sci 2020;93:911–28.
[138] Nauen R, Steinbach D. Resistance to Diamide Insecticides in Lepidopteran Pests. In: Horowitz AR, Ishaaya I, editors. Advances in Insect Control and Resistance Management, Cham: Springer International Publishing; 2016, p. 219–40.
[139] Dáder B, Colomer I, Adán Á, Medina P, Viñuela E. Compatibility of early natural enemy introductions in commercial pepper and tomato greenhouses with repeated pesticide applications. Insect Sci 2020;27:1111–24.

Chapter 3. *Sphaerophoria rueppellii* genome assembly, functional annotation and comparative genomics

This chapter was submitted to BMC Genomics on 4th November 2021, and has been accepted for review. It is currently in the revisions stage.

A near-chromosome level genome assembly of the European hoverfly, *Sphaerophoria rueppellii*, and a comparative analysis of insecticide resistance-related gene families in hemipteran crop pests and pollinators

Emma Bailey*†‡, Linda Field*, Christopher Rawlings†, Rob King†, Fady Mohareb‡, Keywan-Hassani Pak†, David Hughes†, Martin Williamson*, Eric Ganko**, Benjamin Buer§, Ralf Nauen§

* Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, UK

† Department of Computational and Analytical Sciences, Rothamsted Research, Harpenden, UK

‡ The Bioinformatics Group, Cranfield Soil and Agrifood Institute, Cranfield University, Cranfield, UK

** Seeds Research, Syngenta Crop Protection, LLC, Research Triangle Park, NC, USA.

§ Bayer AG, Crop Science Division, R&D, Monheim, Germany

3.1 Abstract

Background: *Sphaerophoria rueppellii*, a European species of hoverfly, is a highly effective beneficial predator of crop pests including aphids, thrips and coleopteran/lepidopteran larvae in integrated pest management (IPM) programmes. It is also a key pollinator of a wide variety of important agricultural crops. No genomic information is currently available for *S. rueppellii*. Without genomic information for such beneficial predator species, we are unable to perform comparative analyses of insecticide target-sites and genes encoding metabolic enzymes potentially responsible for insecticide resistance, between crop pests and their predators. These metabolic mechanisms include several gene families - cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione-*S*-transferases (GSTs), UDP-glucosyltransferases (UGTs) and carboxyl/choline esterases (CCEs).

Methods and findings: In this study, a high-quality near-chromosome level *de novo* genome assembly (as well as a mitochondrial genome assembly) for *S. rueppellii* has been generated using a hybrid approach with PacBio long-read and Illumina short-read data, followed by super scaffolding using Hi-C data. The final assembly achieved a scaffold N50 of 87Mb, a total genome size of 537.6Mb and a level of completeness of 96% using a set of 1,658 core insect genes present as full-length genes. The assembly was annotated with 14,249 protein-coding genes.

Comparative analysis revealed gene expansions of CYP6Zx P450s, epsilon-class GSTs, dietary CCEs and multiple UGT families (UGT37/302/308/430/431). Conversely, ABCs, delta-class GSTs and non-CYP6Zx P450s showed limited expansion. Differences were seen in the distributions of resistance-associated gene families at the subfamily levels between *S. rueppellii* and some crop pests.

Conclusion and significance: This assembly is the first published genome for a predatory member of the Syrphidae family and will serve as a useful resource for further research into selectivity and potential tolerance of insecticides by beneficial predators. Furthermore, the expansion of some gene families often linked to insecticide resistance and selectivity may be an indicator of the capacity of this predator to detoxify IPM selective insecticides. These findings could be exploited by targeted insecticide screens and functional studies to increase effectiveness of IPM strategies, which aim to increase crop yields by sustainably and effectively controlling pests without impacting beneficial predator populations.

Keywords: *Sphaerophoria rueppellii*, hoverfly, PacBio; Illumina; Hi-C; Whole genome sequencing; beneficial predator; insecticide resistance; comparative genomics; diptera; crop pests

3.2 Introduction

Loss of crops to insect pests can account for more than 10% of potential yield, as a result of both direct feeding damage and the transfer of plant viruses via insect feeding [1]. Methods of controlling insect pests are therefore critical to ensure that crop yields are maximised to sustain the growing world population. Insecticides play a key role in pest control strategies. Many modern insecticides are known to be selective for pests without harming beneficials. However, some insecticides such as pyrethroids tend to be non-specific, and as a result are often toxic to both their target pest species and beneficial predators. Applications of such non-specific insecticides can reduce predator populations so that they are unable to act as an effective natural control. This can lead to pest populations surging, with instances of higher populations than pre-pesticide application [2–4].

Beneficial predators, such as those in the Syrphid family, are effective in the biological control of crop pests. Syrphid adults typically feed on nectar and pollen, however, the larvae of roughly one-third of syrphid species feed on crop pests such as aphids, thrips and coleopteran and lepidopteran larvae [5–11]. Predatory syrphidae are able to feed on up to ~500 aphids during their larval stage, which is a higher daily feeding rate than other aphid predators [12]. For example, *S. rueppellii* were able to reduce aphid (*Myzus persicae)* populations by 84% in a field experiment [13]. Specialised adaptations present within adult female syrphidae allow them to detect aphid pheromones and increase their efficacy as biological control agents - for example, adult females often lay their eggs in close proximity to aphid colonies to ensure a plentiful food supply for emerging larvae [14]. Syrphid adults also avoid laying their eggs close to parasitised aphids [15] which reduces intraguild predation between parasitoids and hoverflies, and thus allows for them to be safely combined in IPM strategies. Such strategies can result in more effective pest control compared to using only one beneficial predator species, especially when attempting to control more than one species of pest [16]. Overall, it is unsurprising that

syrphidae are considered to be amongst the most important aphid predators and a key tool for biological control [17,18].

Alongside pest control, adult hoverflies play a key role in pollination [19] and are considered the second most important pollinator after the Apidae bee families [20]. Unlike bees, hoverflies are highly migratory and therefore capable of transporting pollen over long distances - which has benefits for both the plants and other non-migratory pollinators [21]. Pollination experiments have showed that hoverflies increase seed number in food crops such as strawberry, oilseed rape and sweet pepper (which also showed increased fruit abundance) [13,22,23].

This dual role as effective pollinators and biological control agents [11] makes hoverflies hugely attractive for commercial use and also highlights the need to develop IPM strategies which conserve their populations. The aim of this work was to produce a high-quality genome assembly for *S. rueppellii*, shown in figure 1, to serve as a resource for research into this species as well as the wider Syrphidae family, which consists of ~6000 species worldwide [19,24] and is therefore a potentially valuable source of biological control agents.

The number of beneficial predator genomes has been trailing behind crop pest genomes in recent years, although numbers are now on the rise, especially with the progress being made by the Darwin tree of Life (DToL) sequencing project. High quality genomes have already been released by DToL for some beneficial predators such as green lacewing (*Chrysoperla carnea*) and the seven spotted ladybird (*Coccinella septempunctata*) [25]. Other publicly available beneficial predator genomes include: a phytoseiid mite, parasitoid wasps, a minute pirate bug and lady beetles [26–29]. To date the only available genome for the Syrphid family is the non-predatory European hoverfly (*Eristalis pertinax*) released by DToL (but not yet annotated at the time of writing), so this *S. rueppellii genome* is the first available for a predatory member of the Syrphid family.

Figure 1. ***Sphaerophoria rueppellii.*** Image by Sandy Rae, licensed under CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=18497052

The EU Directive on Sustainable Use of Pesticides 2009/128/EC [30] means that IPM strategies, including the use of beneficial predators [31–35], are growing in necessity. These strategies can be supported by comparative analyses of the genomes of predators and pests, focusing on potential differences in insecticide tolerance mechanisms based on both target-site selectivity and metabolism.

There are two main types of insecticide resistance mechanisms: mutations in insecticide target genes that prevent the insecticide binding to the target [36] and duplication or increased

expression of genes encoding enzymes which can metabolise insecticides. Gene families associated with metabolic resistance include cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione S-transferases (GSTs), UDP-glucosyltransferases (UGTs) and carboxyl/choline esterases (CCEs) [37–42]. Comparisons of these mechanisms in beneficial predators and crop pests could help identify insecticides which will target crop pests but have limited impact on beneficial predator populations. This information could prove key to developing successful IPM strategies which exploit differences in insecticide selectivity between the predator and crop pests. Improving the availability of beneficial predator genomes could also aid the selection of beneficial predators with genes/mutations for insecticide resistance before being released in the field for biological control [43].

The genome assembly and comparative analysis presented here provide a comprehensive foundation for further study of insecticide tolerance and selectivity mechanisms in beneficial predators and how they compare to crop pests.

3.3 Materials and Methods

3.3.1 Sample preparation and sequencing

S. rueppellii larvae were obtained from 'biopestgroup.com'. CO_2 was used for anaesthesia to allow the insects to be sorted from the substrate. The larvae were then flash frozen with liquid N_2 and stored at -80°C. The whole process was completed within 48 hours of arrival.

For transcriptome sequencing, RNA extractions were carried out in-house at Rothamsted Research using the Bioline Isolate II RNA Mini Kit. 30µg of RNA was obtained from ~5

individuals. The library was constructed with an insert size of 150bp and PolyA selection for rRNA removal. Sequencing was performed by Genewiz (New Jersey, US) using Illumina HiSeq 4000 with a 2x150bp paired-end configuration.

For short-read genomic sequencing, DNA extractions were performed in-house at Rothamsted Research using the commercial DNAzol reagent. Short reads were sequenced using 1.1µg of DNA obtained from ~5 individuals and a library with an insert size of 200bp. Sequencing was performed by Genewiz (New Jersey, US) using Illumina HiSeq 4000 with a 2x150bp paired-end configuration. K-mer counting of the raw Illumina DNA data was performed using Jellyfish 2.2.6 [44]. Canonical (-C) 21-mers (-m 21) were counted and a histogram of k-mer frequencies outputted. GenomeScope 2.0 [45] was used to process this histogram with ploidy set to 2 and maximum k-mer coverage cut-off set to 10,000.

For long-read genomic sequencing, whole insects were sent directly to Georgia Genomics (University of Georgia, US) who performed the DNA extractions using ~15 individuals. To obtain long-read PacBio data, a 15-30Kb SMRTbell library was produced with an insert size of 24,000bp and a 15 hour sequencing run was carried out using PacBio Sequel II.

For Hi-C sequencing, whole insects were sent directly to Arima Genomics (San Diego, US) who carried out the DNA extractions using 10 individuals. Arima-QC and library preparation were also performed in-house. Sequencing was performed using Illumina HiSeq X with a 2 x 150bp paired-end configuration.

3.3.2 Genome quality assessment

Basic metrics from the genome assembly were calculated using a script developed for the 'Assemblathon' [46]. These metrics include scaffold/contig N50, longest and shortest scaffold

length, number of scaffolds exceeding a range of lengths and number of gaps/N's in the assembly.

The completeness of the genome assembly and annotation for *S. rueppellii* was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) [47] of the insect gene set (insecta odb9). 'Genome' mode was used to assess the assembly, and 'protein' mode to assess the annotation. 'Fly' was used as the training species for Augustus gene prediction. BUSCO assessments were then run with default parameters.

3.3.3 *De novo* genome assembly

FastQC v.0.11.8. [48] was used to perform quality checks on the raw Illumina HiSeq DNA and RNA sequence data. Adapters were trimmed, low-quality bases (below a score of 3) were removed from the start and end of reads and any reads with a length less than 36 bases were also removed. Trimmomatic v.0.38. [49] was used for these trimming steps.

The raw PacBio reads were subsetted using a 'SelectLongestReads' script from: https://github.com/yechengxi/AssemblyUtility to reduce coverage from 277x to 150x coverage prior to assembly. The subsetted PacBio long reads were then assembled into contigs with the Flye v2.5. *de novo* assembler [50,51] with the following parameters: '--genome-size 300m -i 3 --meta'. This subsetting was used to reduce duplication in the assembly outputted by Flye whilst maintaining the completeness of the genome.

The subsetted PacBio long reads and Illumina DNA short reads were also assembled into contigs using Platanus Allee v2.2.2 [52] with default parameters. This is a hybrid assembler designed for heterozygous data.

QuickMerge v0.3 [53] was used to merge the Flye and Platanus-Allee assemblies, with Flye as the reference assembly. BUSCO outputs were compared between the merged assembly and the standalone assemblies to identify core insect genes which had been lost during the merging process. Full-length contigs containing these missing genes were extracted from the standalone assemblies and added to the merged assembly, based on the assumption that these contigs would also contain other missing genes (i.e. those not included in BUSCO's list of 1,658 core insect genes).

Purge Haplotigs v1.0.0 [54] was next used to perform redundant contig removal from the merged assembly. Parameters '-l 5 -m 30 -h 190' were chosen from the coverage histogram outputted in the first step of the pipeline. The percent cutoff for identifying a contig as a haplotig was set to '-a 40', (the default value is 70, however a lower cutoff was chosen due to a very high level of duplication). This tool takes read depth coverage into consideration to reduce over-purging of repetitive regions and paralogous contigs, whilst still coping well with highly heterozygous assemblies.

The Hi-C data was processed using Juicer v1.5 [55] and used as input to the 3D-DNA de novo genome assembly pipeline (version 180922) [56] alongside the draft assembly to produce a candidate chromosome-length genome assembly. Contact matrices were generated by aligning the Hi-C dataset to the genome assembly after Hi-C scaffolding, and were then visualised using JuiceBox Assembly Tools v1.11.08 [57]. The parameters used were as follows: '--mode haploid --build-gapped-map --sort-output'. Additional finishing on the scaffolds was performed in JuiceBox to correct mis-joins.

Three rounds of Pilon [58] error polishing were performed, using the Illumina short read data, after which no further improvement in BUSCO score was seen. A final round of Purge Haplotigs was then performed to reduce duplication further. Parameters '-l 10 -m 50 -h 150' were chosen

from the coverage histogram outputted in the first step of the pipeline. The percent cutoff for identifying a contig as a haplotig was set to '-a 80'.

3.3.4 Mitochondrial genome assembly

The mitochondrial genome was found and extracted by running a BLAST search of the *S. rueppellii* genome against the *Syrphus ribesii* mitochondrial genome, which is publicly available at NCBI, GenBank accession number: MW091497.1.

3.3.5 Annotation

Gene prediction was performed using the MAKER v2.31.8 pipeline [59] through the incorporation of both transcriptome evidence and *ab initio* gene prediction as well as a custom repeat library (see below). MAKER was run using Augustus v3.3.1 [60], GeneMark-ES v4.32 [61] and FGeneSH v8.0.0 [62] as well as EVidenceModeler v1.1.1 [63] with default masking options.

A *de novo* species specific repeat library was constructed using RepeatModeller v1.0.7 [64] to identify repeat models. These models were searched against the GenBank non-redundant (*nr*) protein database for Arthropoda (e value $<10^{-3}$) using Blastx to remove any potential protein-coding genes. This was combined with transposon data to create a custom library. Transposons were identified from the transcriptome assembly by running HMMER: hmmscan [65] against the Pfam database [66] and filtering the resultant Pfam descriptions for those containing "transposon". A search for transposons was also performed on transcripts produced from MAKER and these transposons were then added to the custom repeat library which was used for a second round of MAKER. RepeatMasker v4.0.7 [67] was used to mask repeats in the

genome assembly using these repeat libraries, as well as to estimate the abundances of all predicted repeats.

RNA-seq reads were mapped to the genome with HISAT2 v2.0.5 [68] for assembly with StringTie v1.0.1 [69]. A de novo assembly was also done using Trinity v2.5.1 [70]. The best transcripts (classified by reasonable transcript size and homology to other species) were selected from the Trinity and StringTie assemblies using Evigene v19.jan01 [71].

Evidence from assembled transcripts was transferred to the genome assembly via MAKER. The output from this was then used to produce a high confidence level gene model training set. Overlapping and redundant gene models were removed. Augustus and GeneMark were trained using this training set prior to being used for *ab initio* gene predictions. FGeneSH was run based on the *Drosophila melanogaster* genome.

The best transcripts from both the *ab initio* gene prediction annotation and the transcriptome-based annotation were selected using Evigene and combined to create the final annotation.

S. rueppellii protein sequences were aligned using Blastp against the non-redundant (nr) NCBI protein database for Arthropoda. InterProscan searches were run against several databases (CDD, HAMAP, HMMPAnther, HMMPfam, HMMPIR, FPrintScan, BlastProDom, ProfileScan, HMMTigr) for functional annotation. BLAST2GO [72] was used to assign gene ontology (GO annotations). Infernal v1.1.2 [73] was used to predict and annotate non-coding RNAs.

The mitochondrial genome was annotated using MITOS2 [74] with reference database 'RefSeq 89 Metazoa' and genetic code '5 Invertebrate'.

3.3.6 Comparative genomics and phylogenetic analysis

To produce the species tree, orthogroup gene trees were produced using OrthoFinder [75] and the tree was inferred from these using the STAG method [76].

In order to identify genes potentially involved in insecticide resistance, the PFAM domains assigned to gene models during annotation (as described in the 'Genome Annotation' methods section) were used as follows: CCEs (PF00135/IPR002018), GSTs (IPR004045/PF02798), (IPR004046/PF00043), P450s (IPR001128/PF00067), ABCs (IPR003439/PF00005) and UGTs (IPR002213/PF00201). Proteins from UniProtKB for the classes of interest, from hemipteran species, were used for BLAST queries against *S. rueppellii* to identify any missed genes and to assist with subfamily assignment within these classes. Subfamily assignment for *S. rueppellii* gene families was finalised using phylogenetic trees which were produced using MAFFT alignments [77,78] and RaxML v8.2.11 [79]. The GAMMA LG protein model [80] was used and a bootstrap consensus tree was inferred from 100 replicates.

Manual checks and curation were performed for genes potentially involved in insecticide resistance. Increased copy numbers of genes linked to insecticide resistance often led to adjacent tandem duplications being incorrectly annotated by MAKER as one gene model; therefore curation was important to prevent incorrect gene numbers being reported in later analyses. The exon/intron boundaries and start/stop codons of the genes were confirmed through visualization in IGV [81] of RNAseq data mapped to the genome using HISAT2 v2.0.5 [68] and the gene models were edited in Geneious where necessary.

The P450s were classified and named by Dr David Nelson [82]. The UGTs were classified and named by Dr Michael Court [83]. Nomenclature of P450s and UGTs is based on the evolutionary relationships of the sequences. P450 and UGT sequences were BLAST searched against named insect sequences and were assigned to known families if they were >40% (for P450 families) or

>45% (for UGT families) identical. Other sequences were assigned to new families based on their clustering on trees and their percent identity to each other.

3.4 Results and Discussion

3.4.1 Sequencing

~30 individuals of *S. rueppellii* were required to produce sufficient DNA and RNA for sequencing. Since they were obtained commercially, the level of inbreeding of the culture was not known. However, all individuals were obtained from a single colony within the rearing facility. A high heterozygosity level was therefore a possibility and this was kept in mind during the assembly process.

3.4.2 Raw data

The DNA sequencing generated 6,748,327 PacBio reads with a total length of 83.2 Gbp (277x) and a polymerase read length N50 of 63,285bp. A total of 125.3Gb was produced from the Illumina HiSeq platform for whole genome sequencing, as well as 36.9 Gb for transcriptome sequencing. Quality trimming of Illumina reads using Trimmomatic to remove adapters and any reads <36bp resulted in a 2.9% loss of reads for whole genome sequencing and a 5.18% loss of reads for transcriptome sequencing (table 1). A total of 21.6Gb of sequencing data was produced from Arima-HiC. Analysis of proximal ligation gave a library QC metric of 30% (a high-quality Arima-HiC library is >15%).

Table 1. Number of paired-end Illumina HiSeq DNA/RNA sequences present before and after trimming using Trimmomatic. Adapters and reads <36bp were removed.

	Illumina DNA Reads	Illumina RNA Reads
Total sequences before trimming	417,662,063	123,298,454
Total sequences after trimming	405,634,072	116,917,664
Sequences lost	12,027,991	6,380,790

3.4.3 Genome metrics evaluation based on raw reads

The raw read k-mer analysis with GenomeScope 2.0 (see figure 2) estimated a haploid genome size of ~400Mb (table 2), which is an underestimate of the final assembly size of 537Mb. However, such discrepancies are often seen when using k-mer frequency to estimate genome size in genomes with high repeat content and high heterozygosity [84]. Genome repeat length was 170Mb, 42% of the total estimated genome size.

Table 2. Genome characteristics obtained from GenomeScope v2.0
Using a k-mer length of 21, ploidy set at 2 and a maximum k-mer coverage of 10,000

Genome characteristics	Minimum	Maximum
Heterozygosity, %	3.235	3.356
Genome Haploid Length (Mb)	397.7	403.1
Genome Repeat Length (Mb)	169.1	171.4
Genome Unique Length (Mb)	228.6	231.7
Read Error Rate, %	0.153	0.153

The heterozygosity rate ranged from 3.24% to 3.36%. This indicates a fairly high level of heterozygosity, which was taken into consideration in the assembly strategy.

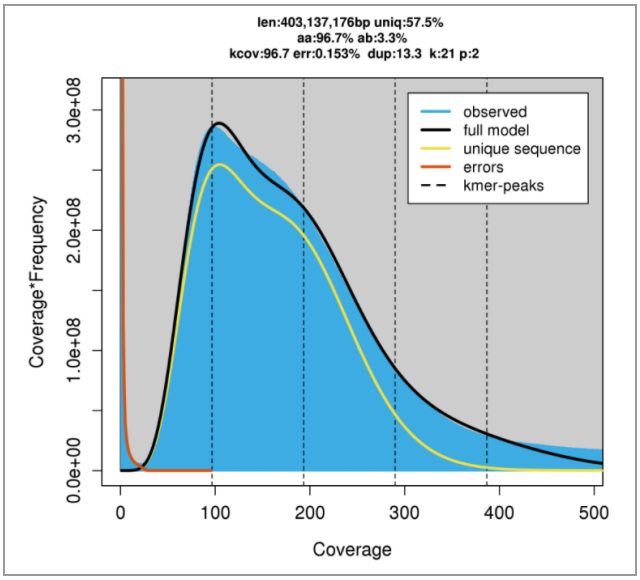


Figure 2. GenomeScope v2.0 k-mer profile plot for the *S. rueppellii* genome, based on 21-mers in Illumina reads. The observed k-mer frequency distribution is depicted in blue, whereas the GenomeScope fitmodel is shown as a black line. The unique and putative error k-mer distributions are plotted in yellow and red, respectively.

3.4.4 Assembly

After trialing a variety of different assemblers, focusing on those suited to heterozygous genomes, and assessing the resultant contiguity (N50) and completeness scores (BUSCO), Flye and Platanus-Allee were chosen to produce 2 separate assemblies. Flye had the best assembly statistics in terms of scaffold N50 (100,207bp with 18 scaffolds >1 million bp) and BUSCO completeness score (99.2%), however, duplication was very high (48.3%) for this assembly, even after subsetting the longest reads to get 150x coverage (duplication was 63.8% prior to subsetting). The total number of scaffolds was 50,164. Platanus-Allee had a lower scaffold N50 (42,845bp with 0 scaffolds >1 million bp) and a slightly lower BUSCO completeness score (97.6%), however, duplication was much lower (3.6%). The total number of scaffolds was 67,142.

The Flye and Platanus-Allee assemblies were merged using QuickMerge, and some manual curation was performed to bring back falsely removed contigs. This resulted in an assembly with a completeness score of 96.5%, duplication of 15.5%, a scaffold N50 of 67,653bp and a total of 59,284 scaffolds, 16 of which were >1 million bp. A subsequent round of Purge Haplotigs brought the duplication score down to 4.6% whilst still maintaining a completeness of 95.6%. Scaffold N50 increased to 126,450bp and the total number of scaffolds was reduced to 15,009.

This draft assembly was next used for scaffolding with Hi-C data using the 3D-DNA *de novo* genome assembly pipeline. This increased the scaffold N50 to 87,361,475 bp, with 5 scaffolds > 10 million bp. The total number of scaffolds was reduced to 11,549, with 6 chromosomal-level scaffolds, numbered by sequence length (figure 3). The BUSCO completeness score was reduced to 94.6%, however, a round of Pilon error polishing brought this back up to 96.4% (subsequent rounds of Pilon worsened the BUSCO score). A final run with Purge Haplotigs reduced duplication from 4% to 3%. Statistics of the final assembly are shown in table 3.

Table 3. Final assembly statistics for the *S. rueppellii* genome

Number of scaffolds	8,476
Total size of scaffolds	537,631,316 bp
Longest scaffold	125,413,692 bp
Shortest scaffold	957 bp
Number of scaffolds > 1K nt	8,412 (99.2%)
Number of scaffolds > 10K nt	2,095 (24.7%)
Number of scaffolds > 100K nt	70 (0.8%)
Number of scaffolds > 1M nt	9 (0.1%)
Number of scaffolds > 10M nt	5 (0.1%)
N50 scaffold length	87,097,991 bp
Number of N's	56,988,920
BUSCO	C:96.0%[S:93.0%,D:3.0%], F:1.2%,M:2.8%

The final assembly size of 537.6Mb was slightly larger than the assembled genome size for *E. pertinax* (482Mb) [85], but closely matches the genome size for *Episyrphus balteatus* (530Mb) from the Syrphidae family, which was calculated using flow cytometry and can therefore be considered a more accurate estimate [86].

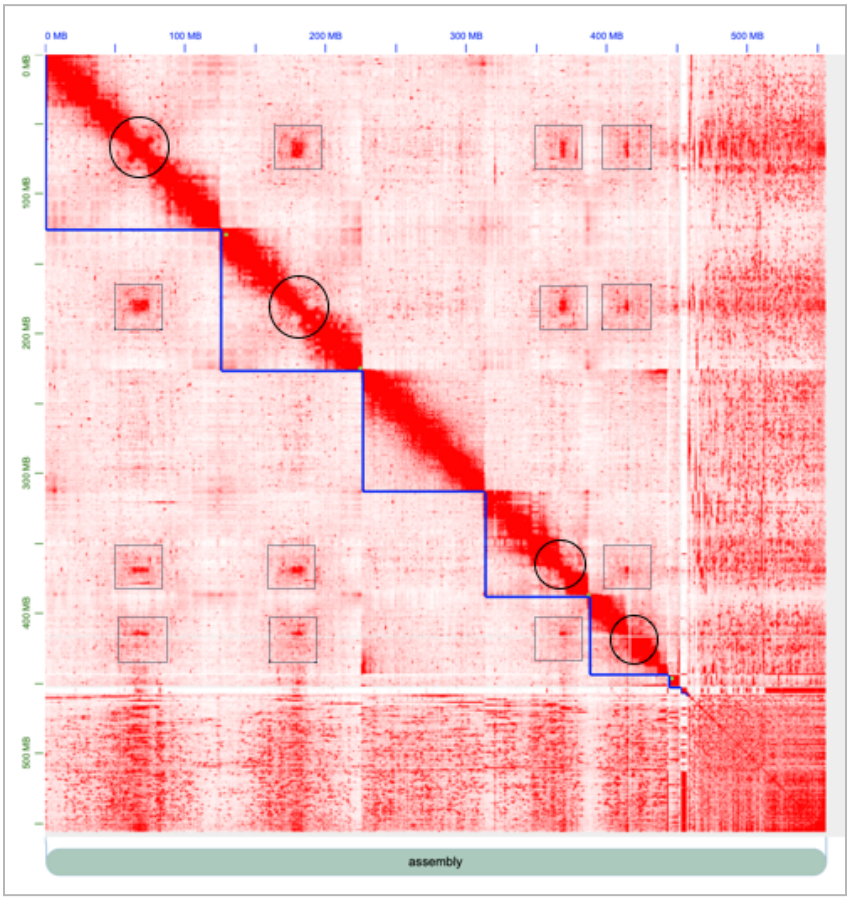


Figure 3. The *Sphaerophoria rueppellii* genome visualised in JuiceBox, with Hi-C contacts shown in red. Blue edges = superscaffolds/chromosomes. Black circles = likely centromeres. Grey boxes = centromere - centromere inter-chromosomal interactions. (Potential chromosome 3 had no obvious centromere, which may have been due to a mis-assembly. The 7th scaffold was mostly repeat regions - evidenced by the lack of interactions with the rest of the genome.)

3.4.5 Annotation

Gene prediction with MAKER identified 14,249 protein-coding genes with the proteins having a mean length of 465 amino acids. Of these, 10,789 (76%) had a match to NCBI's non-redundant (*nr*) database and 12,000 (84%) contained InterPro motifs, domains or signatures. The longest protein found was a 'nesprin-1 isoform' at 17,083aa. The final proteome had a BUSCO completeness score of 87.3% (with 4.9% duplication).

From the Infernal tool inference of RNA alignments, a total of 2,292 non-coding RNA elements were found in the genome (table 4).

Table 4. Number of ncRNAs predicted in the *S. rueppellii* genome

ncRNA element	Number of elements
tRNA	2,058
rRNA	37
snRNA	79
miRNA	81
srpRNA	7
snoRNA	28
lncRNA	2

3.4.6 Repeat annotation

Transposable and repetitive elements made up 30% of the *S. rueppellii* genome (table 5). This is consistent with the reported repeat content of genomes of Diptera species, which ranges widely from 7% (*Drosophila simulans*) to 55% (*Aedes aegypti*) [87]. 16.15% of the *S. rueppellii* genome (77,619,601bp) was masked for annotation - some repeats are annotated but not masked, such as those less than 10bp in length. The majority of these were LINES (9.97%) and interspersed repeats (14.35%). Details of transposable and repetitive elements are shown in table 5.

Table 5. Summary of transposable and repetitive elements in the *S. rueppellii* genome

Percentages do not include runs of X/Ns >=20

			Number of Elements	Length Occupied (bp)	Percentage of Sequence
Retroelements			83,508	57,419,755	11.94
	SINES		21	6045	0.00
	Penelope		114	59,699	0.01
	LINES		75,547	47,912,743	9.97
		L2/CR1/Rex	10,835	4,939,603	1.03
		R1/LOA/Jockey	659	459,946	0.10
		R2/R4/NeSL	37	26,828	0.01
		RTE/Bov-B	56,568	39,492,974	8.22
	LTR elements		7,490	9,500,967	1.98
		BEL/Pao	1,433	2,121,736	0.44
		Ty1/Copia	2,122	2,534,993	0.53
		Gypsy/DIRS1	3,935	4,844,238	1.01
DNA transposons			16,586	8,092,263	1.68
	hobo-Activator		88	35,361	0.01
	Tc1-IS630-Pogo		16,004	7,819,876	1.63
	PiggyBac		27	8,274	0.00
	Other (Mirage, P-element, Transib)		23	13,960	0.00
Rolling-circles			3,616	462,376	0.10
Unclassified			6,299	3,448,721	0.72
Total interspersed repeats			-	68,960,739	14.35
Small RNA			54	15,691	0.00
Satellites			11	817	0.00
Simple repeats			150,503	6,316,344	1.31
Low complexity			36,195	1,890,143	0.39
Total			**296,772**	**146,606,849**	**30.00**

3.4.7 Mitochondria

A circularized mitochondrial genome of 16,387bp, assembled and annotated using MITOS2, consisted of 13 protein coding genes, 22 tRNA genes, 2 rRNA genes and an A+T rich region with a length of 1,500bp (figure 4). This is very similar to the *Syrphus ribesii* mitochondrial genome which is 16,530bp in length and also has 13 protein coding genes, 22 tRNA genes, 2 rRNA genes and an A-T rich region [88].

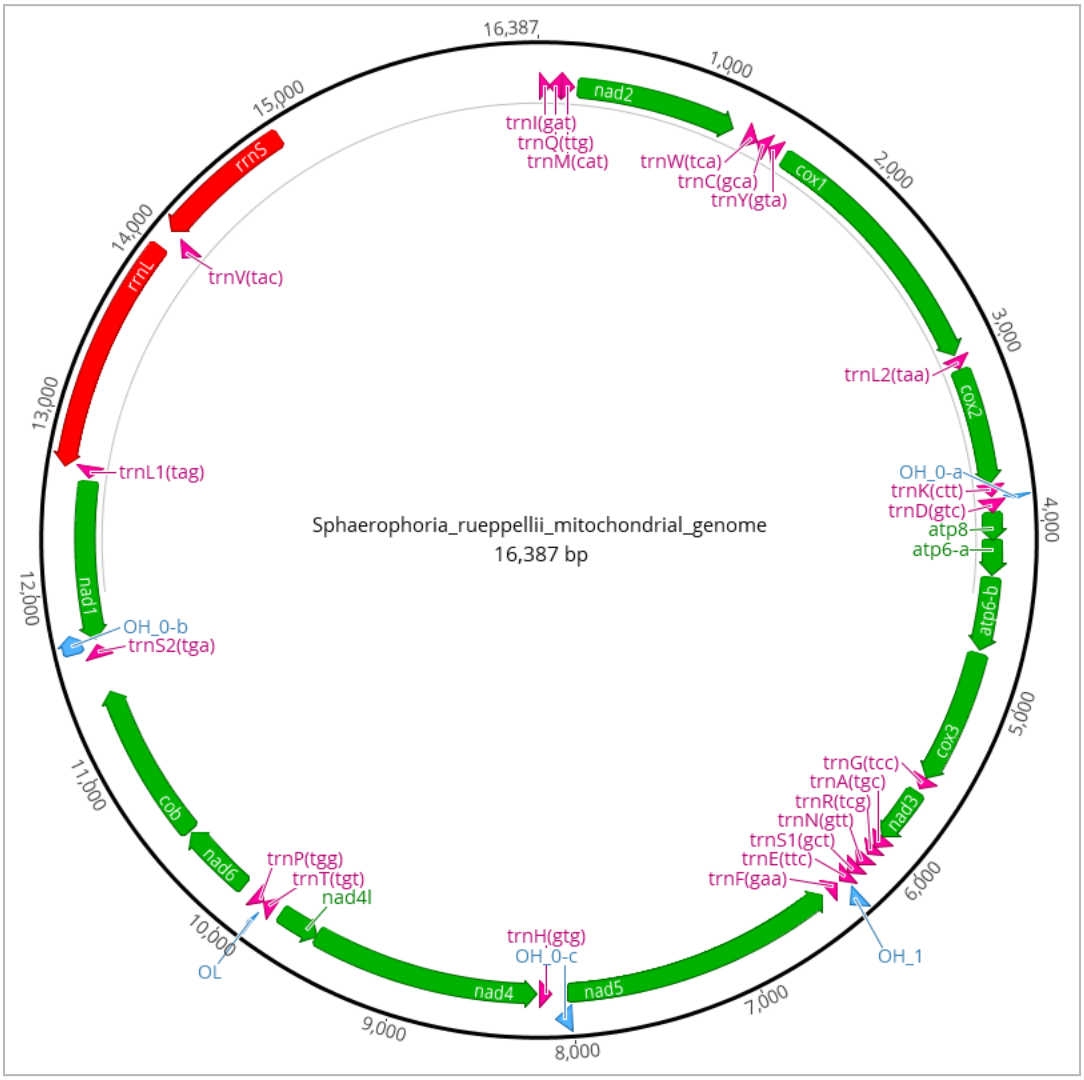


Figure 4. The mitochondrial genome for *Sphaerophoria rueppellii*, visualised using Geneious and annotation track obtained using MITOS2.

3.4.8 Phylogeny

OrthoFinder assigned 435,592 genes (93.6% of total) to 28,834 orthogroups. There were 1,805 orthogroups with all species present and one of these consisted entirely of single-copy genes. Phylogenetic analysis correctly clustered *S. rueppellii* within the dipteran clade, between the Phoridae and Drosophilidae families [89] (figure 5).

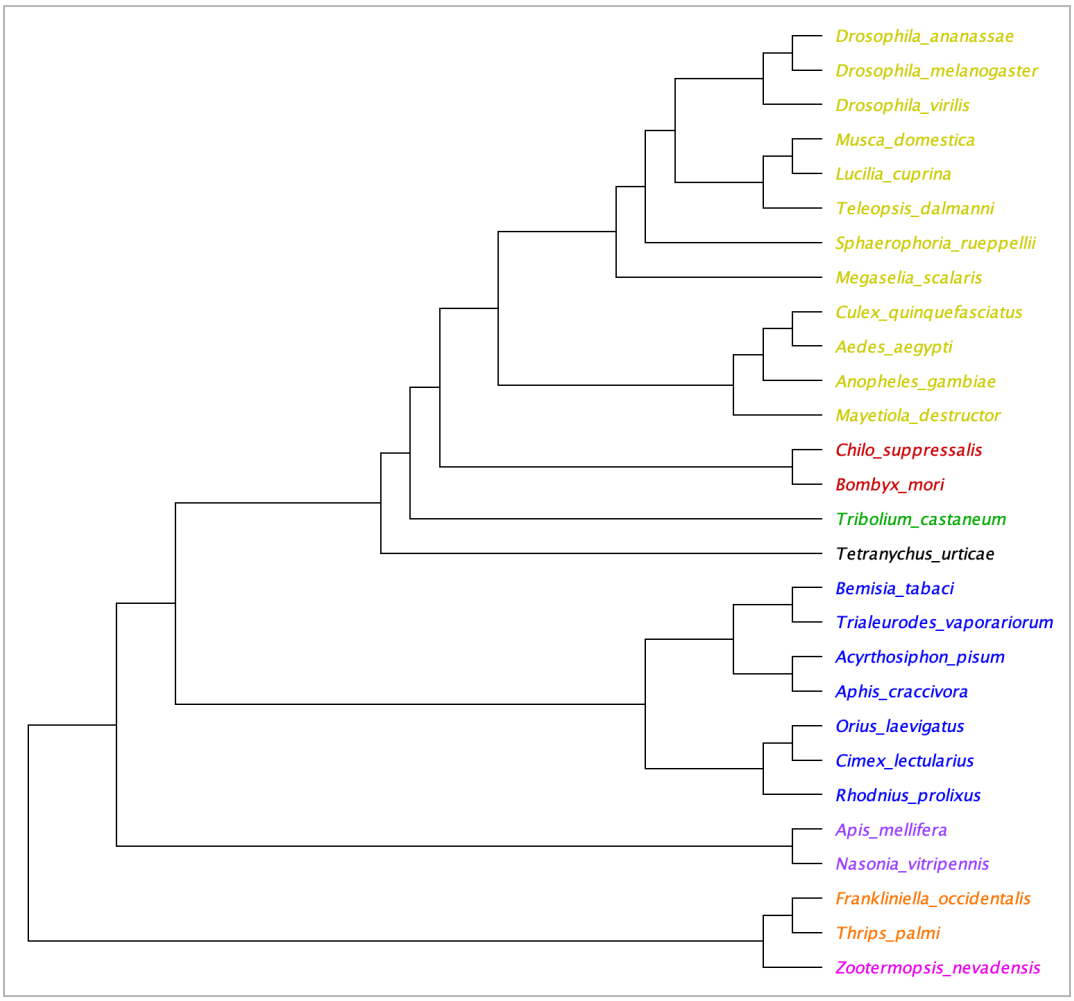


Figure 5. Phylogeny and divergence of Insecta, including crop pests and other beneficial predators

Nodes are coloured by order, yellow=Diptera, red=Lepidoptera, green=Coleoptera, black=Chelicerata, blue=Hemiptera, purple=Hymenoptera, orange=Thysanoptera, pink=Isoptera. Produced using the STAG tree inference method and full proteomes of the following species: *D. ananassae:* PRJNA12651, *D. melanogaster:* PRJNA13812, *D. virilis:* PRJNA12688, *M. domestica:* PRJNA176013, *L. cuprina:* PRJNA248412, *T. dalmanni:* PRJNA391339, *S. rueppellii:* (this study), *M. scalaris:* PRJEB1273, *C. quinquefasciatus:* PRJNA18751, *A. aegypti:* PRJNA318737, *A. gambiae:* PRJNA1438, *M. destructor:* PRJNA45867, *C. suppressalis*: PRJNA506136, *B. mori:* PRJNA205630, *T. castaneum*: PRJNA12540, *T. urticae:* PRJNA315122, *B. tabaci*: PRJNA312470, *T. vaporariorum*: PRJNA553773, *A. pisum*: PRJNA13657, *A.*

craccivora: PRJNA558689, *O. laevigatus:* PRJNA721944, *C. lectularius:* PRJNA167477, *R. prolixus:* PRJNA13648, *A. mellifera*: PRJNA471592, *N. vitripennis*: PRJNA575073, *F. occidentalis*: PRJNA203209, *T. palmi*: PRJNA607431, *Z. nevadensis:* PRJNA203242.

3.4.9 Comparative genomics

UDP-glucosyltransferases

UDP-glucosyltransferases (UGTs) are phase II detoxification enzymes which are involved in insecticide metabolism. The mechanisms of UGT-mediated resistance are for example based on the conjugation of P450-functionalized substrates, however, their upregulation has been shown in resistant strains of *P. xylostella* [37], and they have been linked to diamide resistance in *Chilo suppressalis* [90] and neonicotinoid resistance in *Diaphorina citri* [91] and they also contribute to insecticide detoxification via the elimination of oxidative stress in *Apis cerana* [92].

Table 6. Numbers of annotated UDP glucosyltransferase genes found in *Sphaerophoria rueppellii* (this study), *Drosophila melanogaster* [93]**, *Anopheles sinensis, Aedes aegypti, Anopheles gambiae*** [94]**, *Apis mellifera, Bombus impatiens, Bombus huntii*** [95]**, *Tetranychus urticae, Nilaparvata lugens, Acyrthosiphon pisum, Bemisia tabaci*** [96]**, *Myzus persicae*** [97]**, *Trialeurodes vaporariorum*** [98] **and *Thrips palmi*** [99]**.**

	***S. rueppellii* + close relatives**					**Pollinators**			**Crop pests**						
	Sr	***Dm***	***As***	***Aa***	***Ag***	***Am***	***Bi***	***Bh***	***Tu***	***Nl***	***Mp***	***Ap***	***Tv***	***Bt***	***Tp***
Total	46	35	30	32	23	2	8	2	81	20	101	72	55	76	17

There are 46 UGTs in the *S. rueppellii* genome (table 6), which are classified into 14 families as shown in figure 6 (UGT36, UGT37, UGT49, UGT50, UGT301, UGT302, UGT308, UGT314, UGT316, UGT430, UGT431, UGT432, UGT433, UGT435). Of these families, UGT430-435 are species specific to *S. rueppellii*, whilst all other families are present in at least one other

Diptera species [94]. The UGT genes are distributed across predicted chromosomes 1-5 (with the exception of 1 gene, which is located on a scaffold additional to the chromosome superscaffolds) and the majority (26) are on potential chromosome 2. 38 of the genes are located within clusters of 2-13 tandem UGT genes which generally consist of genes from the same UGT family. This indicates that a high degree of tandem duplication within the UGT gene family likely occurred in *S. rueppellii*.

39 out of 46 UGT genes belong to 7 of the UGT families (UGT308, UGT36, UGT49, UGT302, UGT430, UGT37 and UGT431), suggesting a significant lineage-specific expansion within these 7 families. There is a greater degree of UGT lineage specific expansion and subsequently a higher total number of UGTs within *S. rueppellii* compared to other dipteran species. For example, in the *Drosophila melanogaster* genome, expansion is only seen in 3 UGT families (UGT35, UGT303, UGT37); and in three mosquito species (*Anopheles sinensis, Anopheles gambiae, Aedes aegypti)* expansion is only seen in UGT308 [94]. *S. rueppellii* also has a much higher number of UGT genes compared to other pollinator species.

Hemiptera crop pest species tend to have higher numbers of UGT genes than Diptera, as shown in table 6. This tends to be the result of substantial gene expansion concentrated within a single UGT family. For example: a UGT352 expansion in *Bemisia tabaci* accounted for 36 of its 76 UGTs; the UGT344 family accounted for 35 of *Acyrthosiphon pisum's* 72 UGTs and the UGT201 family accounted for 33 of *Tetranychus urticae's* 81 UGTs. These lineage specific expansions have previously been linked to increased detoxification of plant secondary metabolites [100] and therefore the increased number of UGTs in Hemiptera compared to Diptera may be linked to differences in diet. Host plant adaptation alone has been shown to usually be insufficient to confer insecticide resistance, and therefore higher numbers of UGTs in Hemiptera cannot be assumed to correspond to increased insecticide tolerance/resistance [101]. However, upregulation of UGTs from 7 different UGT families, including 6 UGT344 members, has been associated with thiamethoxam resistance in *Aphis gossypii* [102]. It is

therefore possible that expansion in UGT families may be associated with both host plant adaptation and insecticide resistance. Further study into the role of individual UGTs would be needed to clarify whether differences in total numbers of UGTs are associated with differences in insecticide tolerance/resistance between Hemiptera and Diptera.

9 of the *S. rueppellii* UGT genes belong to the UGT302 and UGT308 families, which are suggested to be the ones most associated with resistance to pyrethroid insecticides [94]. This suggests that expansion of these families in *S. rueppellii* could be a response to pyrethroid exposure. Expansion of these gene families has been reported in *A. sinensis* - 14 of its 30 UGT genes belonged to the UGT302/308 families and 7 of these were considered strong candidates for pyrethroid resistance [94].

The most significant expansion for *S. rueppellii* is seen in the UGT431 family. This family is unique to *S. rueppellii*, but is closest in sequence similarity to the UGT37 and UGT430 families which also exhibited some expansion. The UGT37 family has been shown to be upregulated during organophosphorus pesticide exposure in *Caenorhabditis elegans [103]*. The UGT37 family exhibits lineage specific expansion in *D. melanogaster* and is its largest UGT gene family with members spread across five different genome locations [93]. This differs from the *S. rueppellii* genome, where the majority (12/14) of the UGT37 and UGT431 families are located in a cluster of adjacent genes on chromosome 2 within 0.17Mb of genomic space. This could suggest the UGT37 family may have expanded more recently in *S. rueppellii*, however, the percentage identity within this cluster ranges from 33% to 70%, which indicates that at least part of the cluster can be considered “old”. Since these genes have not been fully dispersed in the genome, there may be a selective advantage for preserving the cluster on chromosome 2 as a heritable unit, i.e. UGT37/431 members may be required for the same mechanism. Based on the links of UGT37 to pesticide resistance, the expansion of the UGT37/431 families and preservation of the gene cluster could be an adaptational response to pesticide exposure.

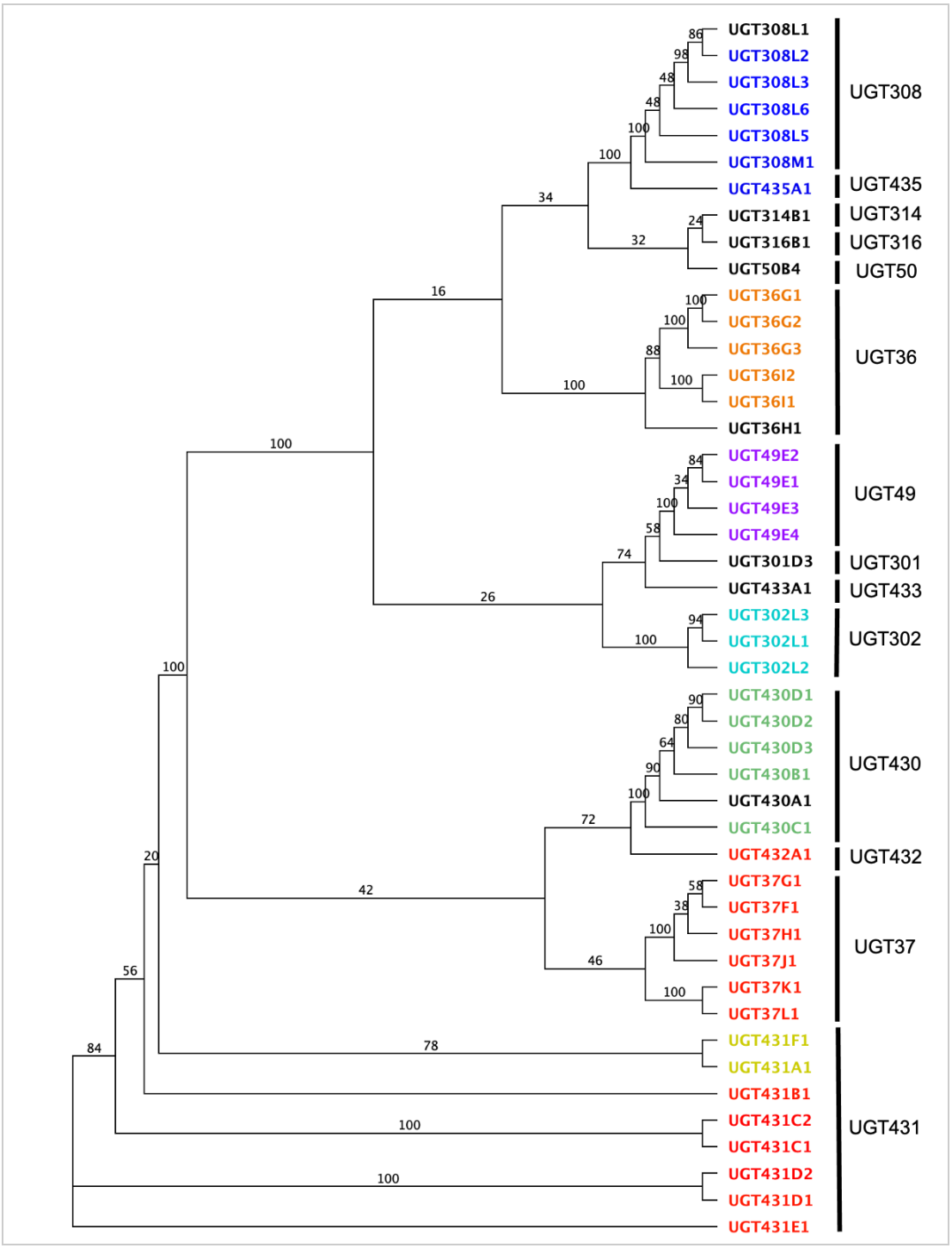


Figure 6. Phylogenetic tree of *S. rueppellii* UDP-glucosyltransferases. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on genes within the cluster having >70% similarity using Blosum45 with threshold 0, and being located adjacently in the genome.

Glutathione *S*-transferases

The glutathione *S*-transferases (GSTs) family is large and functionally diverse, and has been shown to confer resistance to all main insecticide classes. For example, the delta and epsilon classes have been linked to pyrethroid resistance in *A. aegypti* and *N. lugens* [104,105]. GST-mediated detoxification of insecticides takes place via several different mechanisms, including protecting against oxidative stress, binding and sequestration of the insecticide and by catalysing the conjugation of glutathione to insecticides and their metabolites to reduce their toxicity and facilitate excretion, respectively [39].

Table 7. Numbers of GST genes annotated in *Sphaerophoria rueppellii* (this study), *Drosophila melanogaster* [107]**, *Aedes aegypti*** [108]**, *Anopheles gambiae*** [109]**, *Culex pipiens quinquefasciatus*** [110]**, *Apis mellifera, Bombus impatiens, Bombus huntii*** [111]**, *Thrips palmi*** [99]**, *Myzus persicae, Acyrthosiphon pisum, Trialeurodes vaporariorum* and *Bemisia tabaci*** [112] **and their distribution across classes.**

	***S. rueppellii* + close relatives**					**Pollinator**			**Crop pests**				
	Sr	***Dm***	***Aa***	***Ag***	***Cp***	***Am***	***Bi***	***Bh***	***Tp***	***Mp***	***Ap***	***Tv***	***Bt***
Delta	4	9	8	12	14	1	-	-	14	3	11	9	14
Epsilon	11	14	8	8	10	0	-	0	0	0	0	1	0
Omega	3	4	1	1	1	1	-	-	1	1	1	0	1
Sigma	1	1	1	1	1	4	-	3	6	12	5	3	6
Theta	3	4	4	2	6	1	-	-	1	1	2	0	0
Zeta	1	2	1	1	0	1	-	-	2	0	0	2	2
Microsomal	0	3	3	3	3	2	-	-	1	2	2	3	2
Total	**23**	**37**	**26**	**28**	**35**	**10**	**15**	**11**	**25**	**19**	**21**	**18**	**25**

S. rueppellii has 23 GSTs (table 7), which are located on proposed chromosomes 1-3, with members of the same family located on the same chromosome. (Chr1: Theta and Omega, Chr2:

Epsilon, Chr3: Sigma, Delta and Zeta.) The total number of GSTs is slightly lower in *S. rueppellii* compared to other Diptera species, although higher than other pollinators. Phylogeny of these GSTs is shown in figure 7.

Sigma-GSTs are associated with detoxification of oxidants produced during pollen and nectar metabolism in bees [106]. However, *S. ruepplii* has a reduced number of sigma-GSTs compared to other pollinators. This suggests *S. rueppellii* may use different detoxification enzymes to cope with these oxidants, or perhaps a different pathway for pollen and nectar metabolism.

Within the Diptera species the majority of GSTs are present within the epsilon and delta class, however, for *S. rueppellii* whilst the numbers of epsilon GSTs are comparable to other Diptera species, the numbers of delta class GSTs are notably lower.

The epsilon class is the largest class in *S. rueppellii*, as a result of substantial class-specific expansion. 7 epsilon members are clustered within 31kb, with a percentage identity ranging from 35% to 81%, this indicates that whilst some members of the cluster are the result of recent tandem duplications, others are the result of far older duplications. Clusters of epsilon GSTs are common across Diptera species, with clusters of 8 epsilon genes seen in *A. aegypti* and *A. gambiae* and a cluster of 11 epsilon genes in *D. melanogaster*. The preservation of these clusters suggests that maintaining epsilon genes as a heritable cluster confers a selective advantage, likely in terms of conferring increased insecticide resistance. This cluster and class specific expansion may therefore imply an increased degree of GST delta-linked pyrethroid tolerance/resistance in *S. rueppellii* compared to Hemiptera crop pests, which have at most 1 epsilon gene.

In contrast to the epsilon class, *S. rueppelli's* delta class is far smaller, as a result of minimal recent class-specific expansion. Only 2 of the genes are directly adjacent, and were likely a recent tandem duplication based on their 88% sequence identity, whilst the other two

members are dispersed through the genome (across 7.8Mb of genomic space). This follows the pattern seen in some other Diptera species, which also have delta genes more widely dispersed than epsilon. For example, 3 separate clusters are seen in both *A. aegypti* and *A. gambiae*, (although in *D. melanogaster* a single cluster of 11 delta genes is present) [113]. This reduced number of delta GSTs could imply a reduced degree of GST delta-linked pyrethroid resistance in *S. rueppellii* compared to Hemiptera crop pests, although this may be counteracted by the significant expansion within the epsilon class. The lack of preservation of delta clusters may also suggest that they confer a less significant selective advantage than do the epsilon GSTs.

The sigma class of GSTs has been associated with the detoxification of organophosphorus insecticides [114]. All Diptera species included in analysis had only 1 sigma gene, and this was also the case for *S. rueppellii*. All crop pest species had larger sigma classes. This may imply a reduced level of GST sigma-linked organophosphorus resistance compared to Hemiptera crop pests.

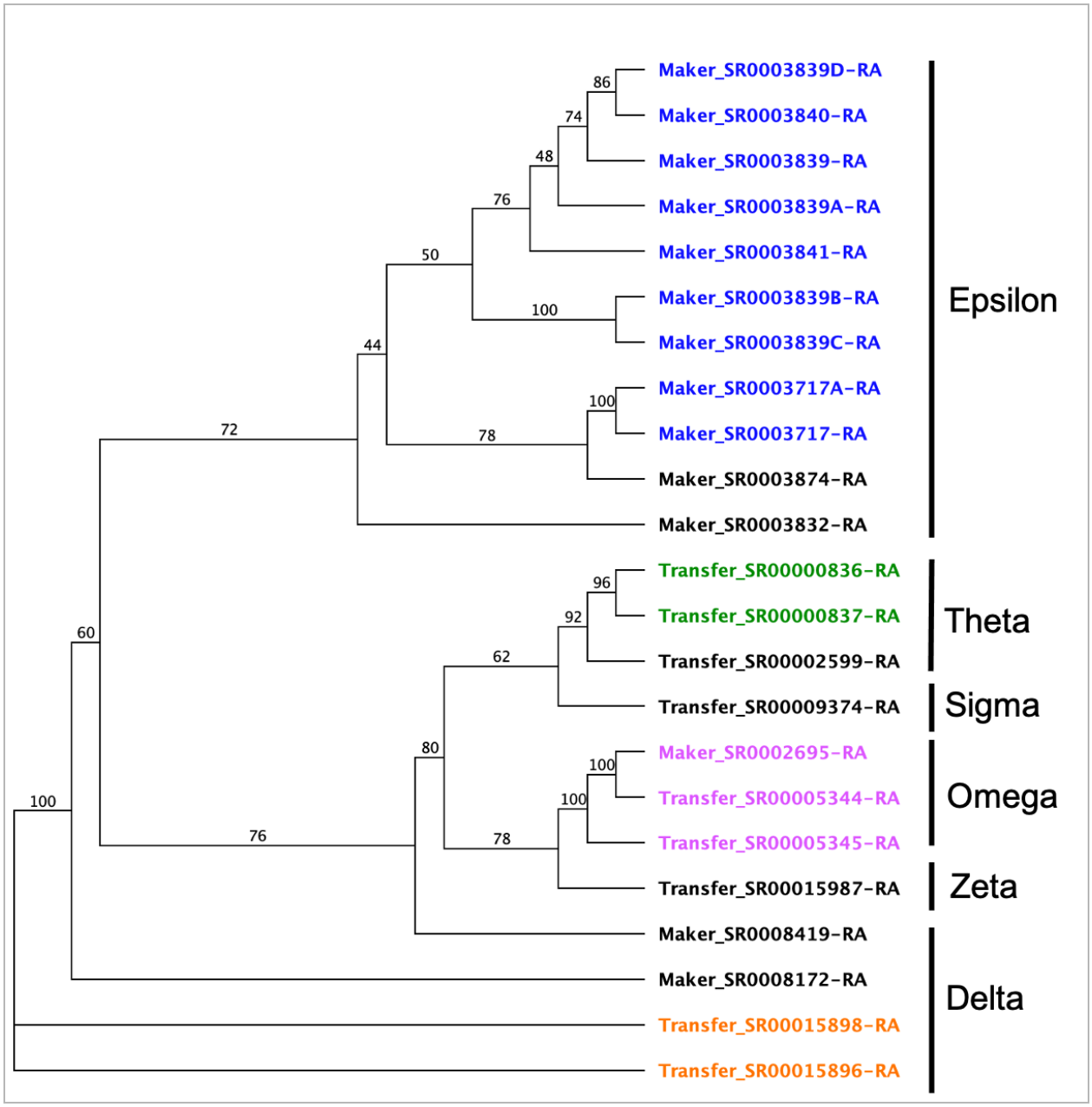


Figure 7. Phylogenetic tree of the *Sphaerophoria rueppellii* glutathione S-transferases. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates.

Coloured nodes indicate groups of likely recent tandem duplications, based on genes within the cluster having >70% similarity using Blosum45 with threshold 0, and being located adjacently in the genome.

Carboxyl/choline esterases

Carboxyl/choline esterases (CCEs) are associated with insecticide resistance, notably to organophosphates, and to a lesser degree carbamates and pyrethroids [115]. For example esterase-based organophosphate resistance has been reported in three *Culex* species [116] and synergist bioassays have shown that esterases are responsible for metabolic resistance to pyrethroids (deltamethrin) and organophosphates (temephos) in *A. aegypti* [117].

Table 8. Numbers of CCEs annotated in *Sphaerophoria rueppellii* (this study), *Drosophila melanogaster, Aedes aegypti, Anopheles gambiae* [118]**, *Culex pipiens quinquefasciatus*** [110]**, *Apis mellifera, Bombus impatiens, Bombus huntii*** [111]**, *Frankliniella occidentalis*** [119]**, *Myzus persicae*** [120]**, *Acyrthosiphon pisum*, *Bemisia tabaci*** [121] **and *Trialeurodes vaporariorum*** [122] **and their distribution across classes and clades.**

		***S. rueppellii* and close relatives**					**Pollinators**			**Crop pests**				
		Sr	***Dm***	***Cp***	***Aa***	***Ag***	***Am***	***Bi***	***Bh***	***Fo***	***Mp***	***Ap***	***Tv***	***Bt***
Dietary class		15	13	30	22	16	8	-	-	28	5	5	12	6
Hormone/semiochemical processing class		13	8	26	15	14	5	-	-	7	12	16	6	19
Neuro-developmental class	Glutactins	4	5	6	7	10	0	-	-	2	0	0	1	1
	AChE	1	1	1	2	2	2	-	-	2	3	2	2	4
	uncharacterised	-	1	2	1	1	3	-	-	2	1	1	1	1
	gliotactin	1	1	1	1	1	1	-	-	1	1	1	1	1
	neuroligin	5	4	3	5	5	5	-	-	7	0	3	3	10
	neurotactin	1	2	2	2	2	-	-	-	1	0	0	1	0
	Subtotal	12	14	15	18	21	11	-	-	15	5	7	9	17
Total		**40**	**35**	**71**	**55**	**51**	**24**	**22**	**23**	**50**	**22**	**28**	**27**	**42**

S. rueppellii has 40 full-length carboxylesterase genes (table 8) which are distributed across proposed chromosomes 1-5 with 19 of the genes arranged in 7 clusters of 2-4 genes (figure 8). The total number of CCEs for *S. rueppellii* and the distribution of genes across the 3 main classes is comparable to other Diptera species. The numbers and distribution of CCEs is also similar between Diptera and Hemiptera, with the only noticeable differences being a lower average number of 'dietary' esterases in Hemiptera species and a higher number of 'glutactins' in Diptera. Compared to other pollinators, *S. rueppellii* has a much higher number of CCE genes.

The so-called 'dietary' class of CCEs has been shown to be involved in insecticide and xenobiotic detoxification [118] and amplification of genes within this class, i.e. esterase E4/B1-like genes, has been linked to organophosphate resistance in hemipteran and dipteran species (*M persicae, N. lugens, S. graminum* and *Culex* mosquitoes) [116,123–127]. Within the *S. rueppellii* genome, multiple clusters of high similarity, adjacent esterase E4/B1 genes indicate recent tandem duplications, which could confer some tolerance/resistance to organophosphorus insecticides. In cases where the number of dietary genes in *S. rueppellii* is higher than Hemiptera crop pests there could be an increased degree of organophosphate resistance.

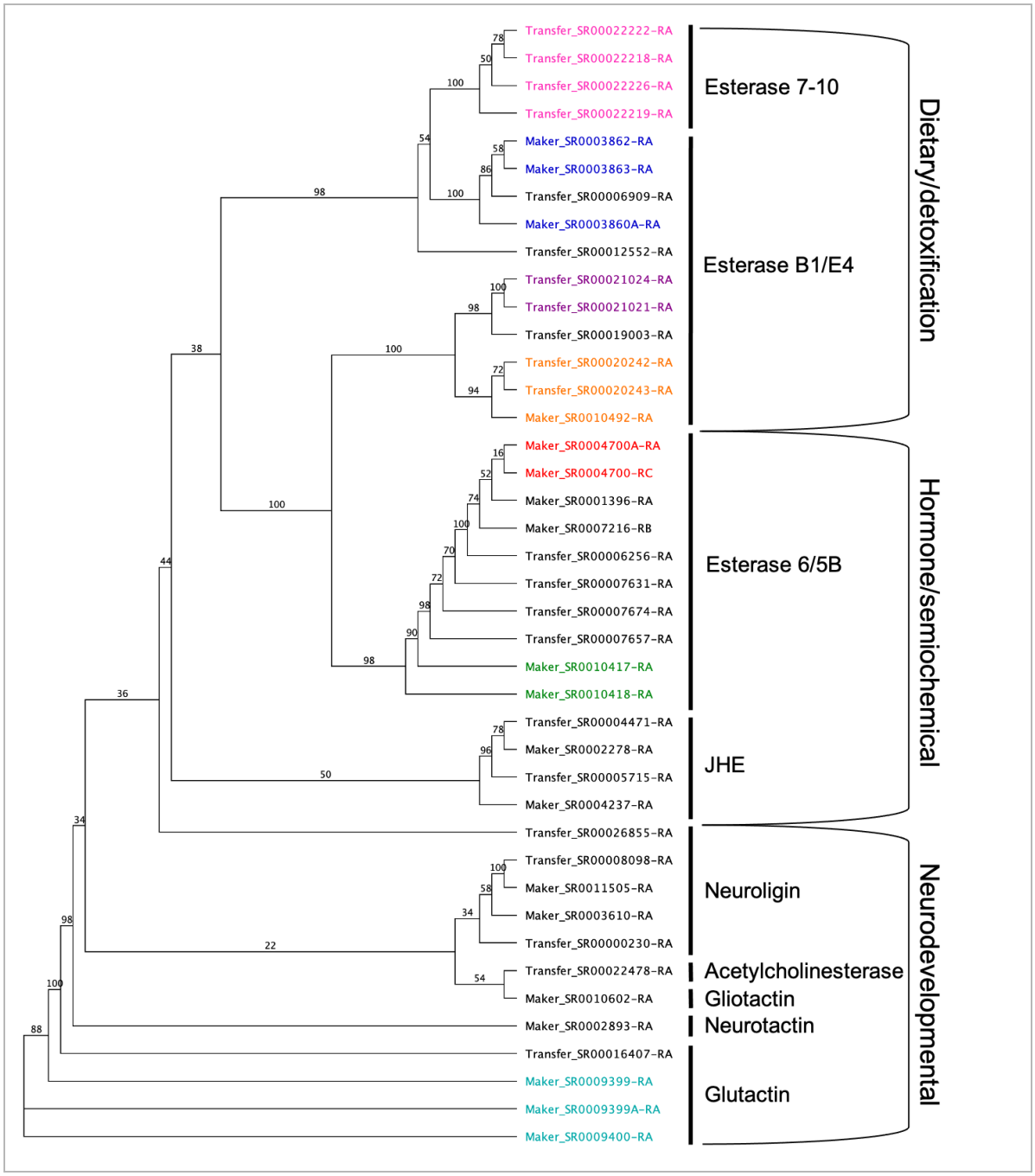


Figure 8. Phylogenetic tree of the *Sphaerophoria rueppellii* carboxyl/cholinesterases. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates.

Coloured nodes indicate groups of likely recent tandem duplications, based on genes within the cluster having >70% similarity using Blosum45 with threshold 0, and being located adjacently in the genome. Maker_SR0001396-RA was a gene fragment, and was not included in the final gene count or analysis; all others are full-length genes.

ABC Transporters

ATP-binding cassette transporters (ABCs) are the largest known group of active transporters and are able to eliminate by translocation xenobiotic compounds such as secondary metabolites produced by plants or insecticides [128]. The ABC transporters are subdivided into eight subfamilies (ABCA-H), of which ABCB, ABCC and ABCG are the most associated with resistance to a variety of insecticides including pyrethroids, carbamates, organophosphates and neonicotinoids [129].

Table 9. Numbers of ABC transporter genes annotated in *Sphaerophoria rueppellii* (this study), *Drosophila melanogaster* [129]**, *Bactrocera dorsalis*** [130]**, *Anopheles gambiae, Culex pipiens quinquefasciatus*** [131]**, *Apis mellifera*** [132]**, *Aedes aegypti*** [133]**, *Anopheles sinensis*** [134]**, *Frankliniella occidentalis*** [119]**, *Thrips palmi*** [99]**, *Aphis gossypii*** [135]**, *Trialeurodes vaporariorum*** [136] ***Diuraphis noxia*** **and *Bemisia tabaci*** [137] **and their distribution across subfamilies.**

	***S. rueppellii* + close relatives**							***Pollinators***	**Crop pests**					
	Sr	***Dm***	***Bd***	***Aga***	***Aa***	***As***	***Cp***	***Am***	***Fo***	***Tp***	***Dn***	***Ago***	***Tv***	***Bt***
ABCA	11 (12*)	10	7	8	10	10	9	3	3	3	3	4	3	8
ABCB	6 (7*)	8	7	5	5	5	5	5	5	4	6	5	9	3
ABCC	8	14	9	15	15	16	18	9	19	12	24	25	7	6
ABCD	3	2	2	2	2	2	2	2	2	2	3	2	4	2
ABCE	1	1	1	2	1	1	2	1	1	2	1	1	1	1
ABCF	3	3	3	3	3	3	3	3	3	3	3	4	3	3
ABCG	10	15	15	17	15	21	28	15	22	16	26	30	9	23
ABCH	3	3	3	3	4	3	3	3	13	7	11	0	9	9
Total	**45 (47*)**	**56**	**47**	**55**	**53 (62 with 9 in ABCJ)**	**61**	**70**	**41**	**70**	**49**	**77**	**71**	**45**	**55**

*including fragment genes >200bp

S. rueppellii has 47 ABC genes (table 9), which are distributed across proposed chromosomes 1-6, with 3 of the genes located on scaffolds external to the chromosome superscaffolds. 20 of the genes are located in 9 clusters of 2-3 (figure 9). The total number of ABC genes in *S. rueppellii* is at the lower end of that seen for other Diptera species as well as for Hemiptera crop pests, but slightly higher than another species of pollinator.

The distribution of *S. rueppelli*'s ABC genes across subfamilies is similar to other species, except for the ABCC and ABCG subfamilies, which are smaller in *S. rueppellii* than all other Diptera species and the majority of Hemiptera crop pests. These are two of the families most associated with insecticide resistance, and so their reduced size suggests that ABC-mediated tolerance/resistance to insecticides could be lower in *S. rueppellii* compared to these other species.

The ABCA subfamily is expanded in Diptera, whilst the ABCH subfamily is expanded in Hemiptera. However these subfamilies do not have strong links to insecticide resistance, and so these differences would likely not contribute to any variation in tolerance/resistance levels.

The percentage identity of ABC genes within *S. rueppellii* ranges from 0%-71%, with the exception of one pair of genes with an identity of 89%. This suggests that there has been little recent lineage specific expansion within the *S. rueppellii* ABC transporter family, and this is supported by the numbers of the genes in the ABC subfamilies, which are either similar to or lower than other Diptera species. Any lineage-specific expansion seen in *S. rueppellii* is minimal, demonstrated by the small size of gene clusters. Species-specific and lineage-specific ABC expansions on a much larger scale have been reported in a variety of arthropods such as *Tribolium castaneum* and *Tetranychus urticae*, although whether these expansions contribute directly to increased insecticide resistance is not yet known [129].

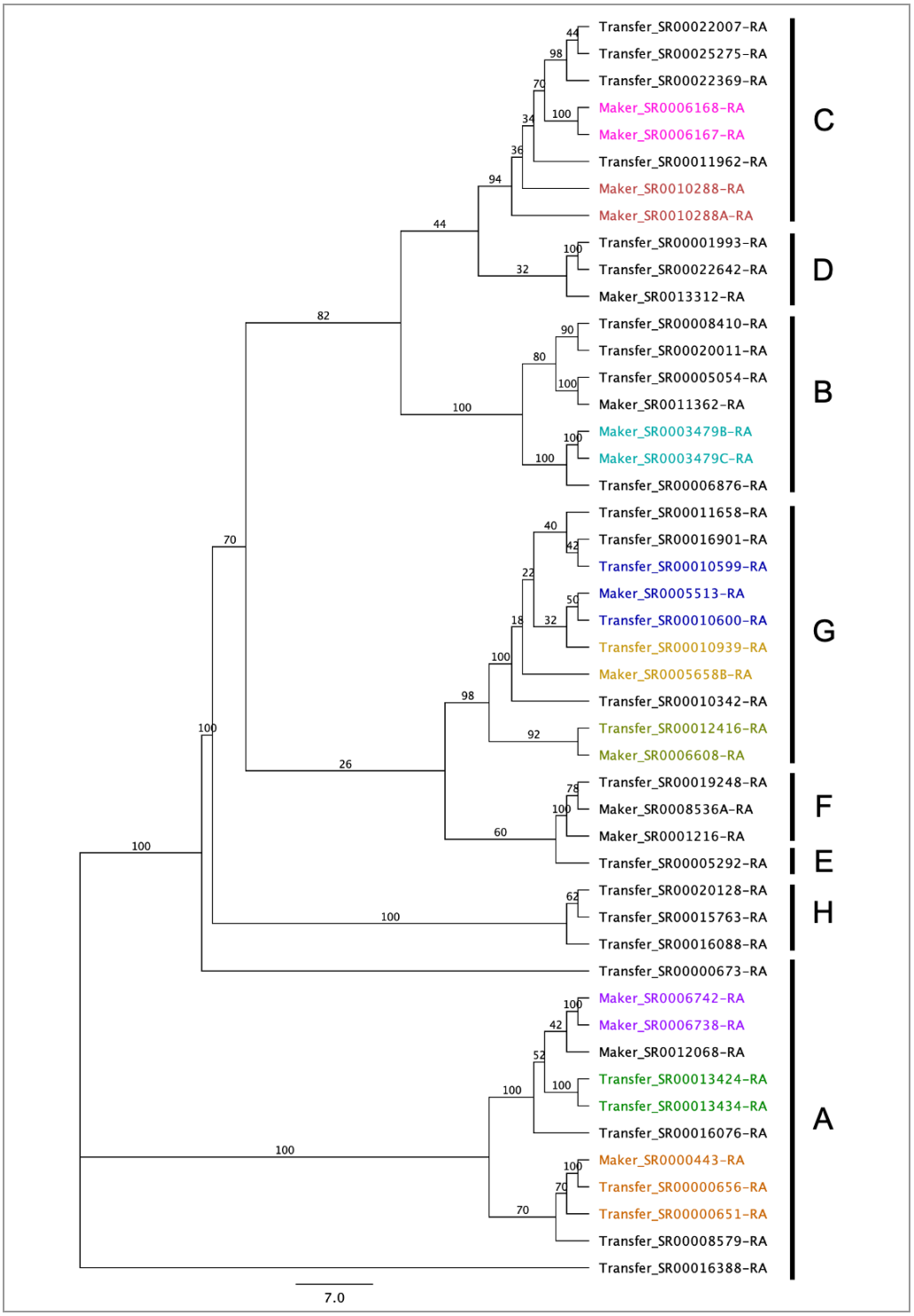


Figure 9. Phylogenetic tree of the *Sphaerophoria rueppellii* ABC transporters. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on genes within the cluster having >70% similarity using Blosum45 with threshold 0, and being located adjacently in the genome.

Cytochrome P450 monooxygenases

Cytochrome P450 monooxygenases (P450s) are a diverse superfamily capable of metabolizing a huge variety of endogenous and exogenous substrates. In insects they are associated with growth and development, metabolism of pesticides and plant toxins as well as the production and metabolism of insect hormones and pheromones [138,139]. P450s are associated with the resistance to insecticides from a variety of classes, including pyrethroids, carbamates and neonicotinoids and many examples of resistance are linked to upregulated P450s [140–143]. They are also linked to the activation of organophosphates and other pro-insecticides (a form of insecticide which is metabolized into an active form inside the host) [144] often as a result of downregulation [145,146].

Table 10. Total numbers of Cytochrome P450 genes annotated in *Sphaerophoria rueppellii* (this study), *Musca domestica, Drosophila melanogaster* [147]**, *Anopheles gambiae, Aedes aegypti*** [148]**, *Culex pipiens quinquefasciatus*** [110]**, *Apis mellifera*** [149]**, *Bombus impatiens, Bombus huntii*** [95]**, *Frankliniella occidentalis, Thrips palmi*** [119]**, *Myzus persicae, Acyrthosiphon pisum*** [120]**, *Trialeurodes vaporariorum*** [136] **and *Bemisia tabaci*** [150]**.**

	***S. rueppellii* + close relatives**						**Pollinator**			**Crop pests**					
	Sr	***Md***	***Dm***	***Ag***	***Aa***	***Cp***	***Am***	***Bi***	***Bh***	***Fo***	***Tp***	***Mp***	***Ap***	***Tv***	***Bt***
CYP2	6	8	7	10	11	14	8	-	-	12	12	3	10	7	18
CYP3	34(37)*	65	35	41	80	88	31	-	-	22	26	63	33	41	76
CYP6	22	46	22	-	-	-	-	-	-	18	-	-	29	34	47
CYP9	2	7	5	-	-	-	-	-	-	0	-	-	0	0	0
Other	10	12	8	-	-	-	-	-	-	4	-	-	4	7	-
CYP4	15(16)*	55	33	45	58	83	4	5	2	37	42	48	32	25	73
Mitochondrial	14	18	11	9	9	11	6	-	-	10	11	1	8	7	4
Total	**69(73)***	**146**	**86**	**105**	**158**	**196**	**49**	**49**	**44**	**81**	**91**	**115**	**83**	**80**	**171**

*Values in brackets represent total gene numbers including partial and fragment genes. For other species partial and fragment P450 genes were excluded in cases where they were listed as such - some may remain in the counts if official naming and curation had not taken place.

A total of 69 full-length P450 genes were identified in the *S. rueppellii* genome, as well as 4 P450 fragment genes (table 10). The phylogeny of these P450s is shown in figure 10. These genes were named by Dr David Nelson using his in-house pipeline [82]. The total number of P450s varies widely between insect species, ranging from 44 for *Bombus huntii* to 196 for *C. pipiens*. *S. rueppellii* falls at the lower end of this range, however when compared to other dipteran species, this is mainly due to the reduced size of the CYP4 clan.

34 of the P450 genes have 55-97% identity to another sequenced P450, 38 have 40-55% identity, and 1 gene has <40% identity. 9 genes (CYP18A1, CYP301-304A1, CYP307A2, CYP314A1, CYP315A1 and CYP49A1) were classified as orthologs to P450s from *Lucilia cuprina, Ceratitis capitata* and *Musca domestica*. These genes are involved in a conserved pathway, found in all insects, for the essential growth hormone 20-hydroxyecdysone [151]. Orthologs were not present for other genes, likely because other P450s are involved in detoxification, and therefore vary during evolution based on the organism's environment and adaptation.

The CYPome diversity value was 52%, based on the presence of 38 CYP subfamilies and 73 genes. The CYPome follows the pattern of other arthropods, with most CYP families having few genes, whilst only a few CYP families have many genes [149].

The majority of *S. rueppellii* P450s (34) belong to the CYP3 clan (table 10), which is the one most associated with insecticide resistance, notably the CYP6 and CYP9 families [139], both of which were present in *S. rueppellii*. CYP3 is also the largest clan in other pollinators and in several other diptera species and hemipteran crop pest species.

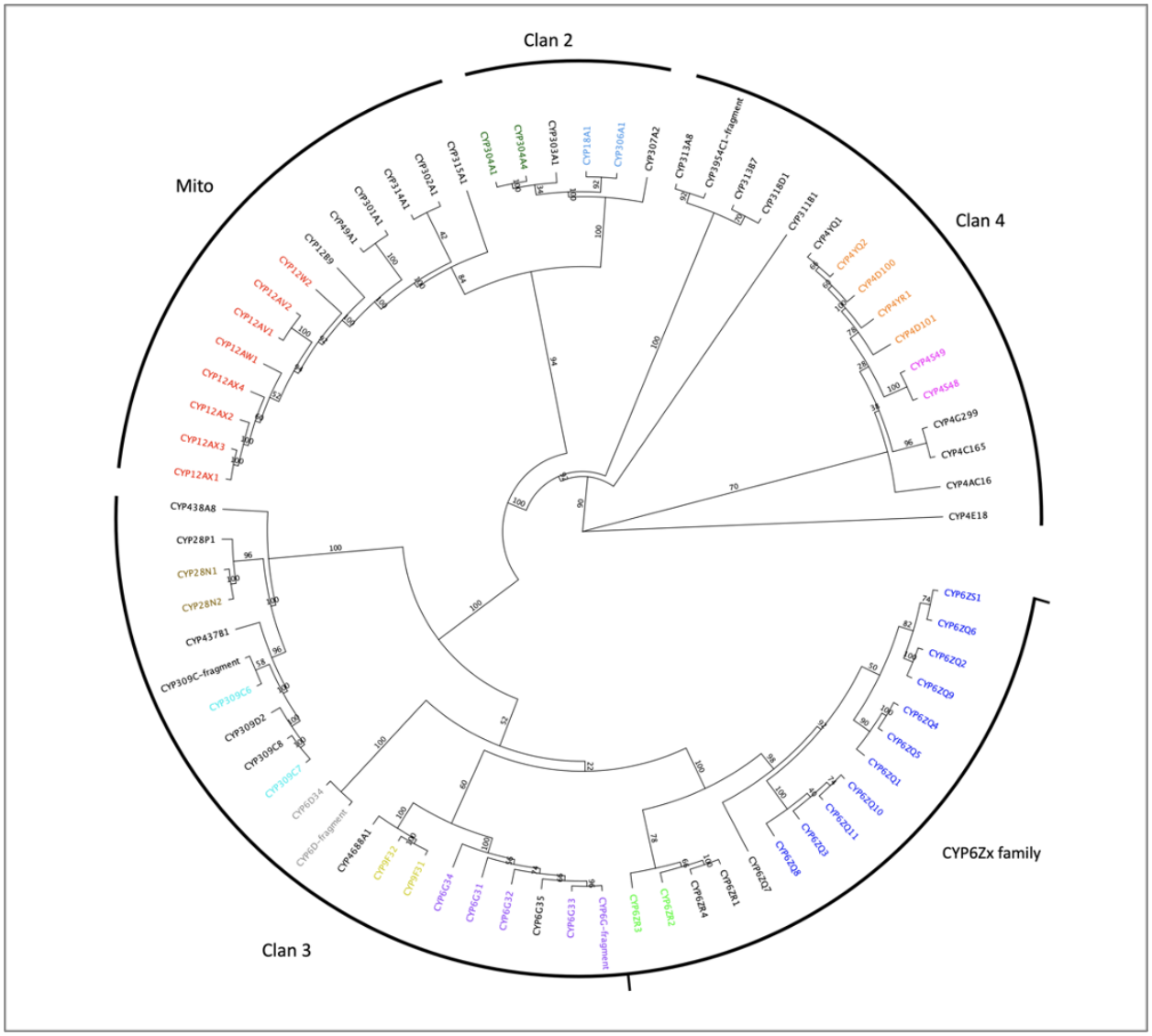


Figure 10. Phylogenetic tree of the *Sphaerophoria rueppellii* Cytochrome P450s. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on genes within the cluster having >70% similarity using Blosum45 with threshold 0, and being located adjacently in the genome. The CYP6Zx family is part of clan 3.

The largest sub-family specific expansion is in clan 3, within the CYP6Zx family, with 16 members: CYP6ZQ1-11, CYP6ZR1-4 and CYPZS1. CYP6ZQ1-11 (excluding Q7) are located contiguously within a 0.2Mb region of chromosome 3 (figure 11). Within this cluster there is no

consistent relationship or pattern between the proximity of the CYP6Zx genes or their gene structure with their percent identity, which ranged from 33-90%. The lower end of the percent identity within the cluster indicates that at least part of the cluster can be considered "old", and therefore, since these genes have not been fully dispersed in the genome, there may be a selective advantage for preserving the cluster on chromosome 3 as a heritable unit.

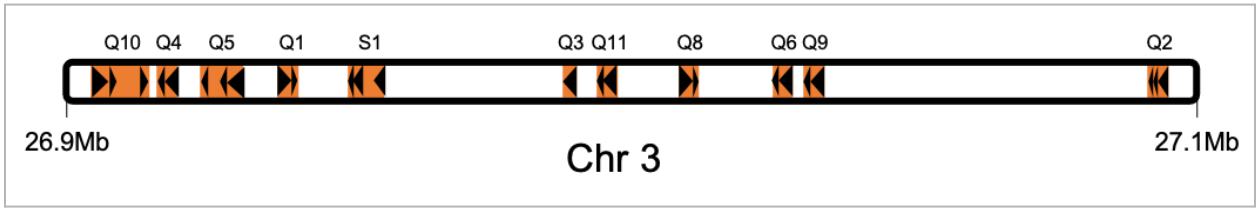


Figure 11. Arrangement of the CYP6Zx subfamily on chromosome 3. Orange boxes represent genes, black arrows represent exons as well as gene orientation.

Whether the large CYP6Zx expansion may confer an increased degree of tolerance to xenobiotics in *S. rueppellii* remains to be investigated. Overall, numbers of the resistance-associated CYP3 clan are similar or lower than Hemiptera crop pests, suggesting that P450-mediated insecticide tolerance/resistance mechanisms may not be as prevalent as for other species.

The CYP4 clan is vastly expanded in many arthropods [152], and whilst the CYP4 clan is not as strongly associated with insecticide resistance as CYP3, studies have shown upregulation of some CYP4 genes in response to insecticide exposure [141,153–155]. *S. rueppellii* has a lower number of CYP4 genes compared to many other dipteran species and crop pests, however, compared to other pollinators the CYP4 subfamily is relatively large. A reduced number of CYP4 genes is common within pollinators [95,156], but the reasons behind this are not yet known.

Pollinators use P450s for the detoxification of pollen flavonoids, notably the CYP6AS subfamily which is often expanded in honey bees; however, this subfamily is absent in *S. rueppellii* [157,158]. It is likely that another subfamily is responsible for flavonoid detoxification in *S. rueppellii* (possibly the expanded CYP6Zx subfamily) and future studies assessing P450 upregulation in response to flavonoids could help identify this.

Point Mutations

Point mutations in genes encoding insecticide targets which are known to confer insensitivity to insecticides were searched for in the *S. rueppellii* genes. This includes those in the sodium channel para gene, which can confer resistance to pyrethroids; the GABA-gated ion channel RDL which can lead to multiple insecticide resistance; the acetylcholinesterase (*ace-2*) enzyme which is associated with organophosphate and carbamate resistance; the Ryanodine receptor which is linked to diamide resistance and acetyl CoA carboxylase which is linked to keto-enol resistance. Despite mutations in these genes having been observed across Diptera species including house flies and mosquitoes as well as crop pests such as whiteflies, aphids and diamondback moths, none were found in this *S. rueppellii* genome [159–173].

Overall, target site mediated tolerance/resistance is not seen in *S. rueppellii*. Although it is important to note that the *S. rueppellii* genome assembly was a consensus of ~30 individuals, therefore mutations would likely only be apparent if they were present in the majority of the population.

3.5 Conclusions

Here we present the first high quality genome of *S. rueppellii* including the mitochondrial genome enabled by PacBio long-read technology combined with low error-rate short-read Illumina sequencing. Hi-C data permitted further scaffolding of this genome to a near-chromosome level assembly. The genome completeness is of excellent quality for comparative and functional genomics analyses and provides a useful first reference for predatory syrphidae.

Comparative analyses of *S. rueppellii* with crop pests showed evidence that *S. rueppelli* has a detoxification gene inventory comparable to selected crop pests, with a few notable differences: lineage-specific expansions were seen within detoxification gene families such as UGTs and P450, whereas the ABC transporter family lacks such expansions compared to some crop pests. No mutations were found in common insecticide target-sites, suggesting a lack of selectivity of insecticides at the protein/receptor binding level.

Comparative analyses of *S. rueppellii* with pollinators showed that *S. rueppellii* has an increased number of genes in all detoxification families, in particular: UGTs, non-sigma class GSTs and CYP4 P450s. This could be in part due to *S. rueppellii* needing more detoxification genes for its diet: hoverflies lack the eusocial behavioural mechanisms seen in bees, such as processing nectar into honey and converting pollen into 'beebread', which result in a dilution of toxins and hence reduce the need for detoxification enzymes in bees [156]. Additionally, the considerably longer migratory distance covered by hoverflies compared to bees [21] may have resulted in hoverflies being exposed to a wider variety of xenobiotics, and could perhaps have resulted in expansion of associated detoxification genes.

Despite the reduced number of detoxification genes in pollinators such as *A. mellifera*, they appear to be no more sensitive to insecticides than other insects [156,174]. Insects with a

pollen-based diet have been found to have an increased degree of insecticide tolerance, with many of the same genes being upregulated in response to both pollen and to certain insecticides [175]. This suggests that the unique set of detoxification genes required by pollinators for their diet, could perhaps impart an increased degree of insecticide tolerance without the need for the extent of gene expansion seen in other insect species. This may mean that despite *S. rueppellii* having fewer detoxification genes than some crop pests, this might not necessarily be indicative of reduced insecticide tolerance. However, this is not to say that insecticides are not a major problem for *S. rueppellii*, with clear evidence that the same neonicotinoids (imidacloprid and thiamethoxam) which are toxic to honey bees are also toxic to *S. rueppellii* [176,177].

This study provides a good basis for beginning to identify differences in genes encoding potential tolerance/resistance mechanisms between crop pests and *S. rueppellii* which could be exploited when selecting targeted insecticides for use in IPM strategies. Evidence of gene expansions in resistance-associated gene families implies that *S. rueppellii* is certainly capable of developing resistance to a variety of insecticides, which could be used to our advantage through the selective breeding and selection of resistant strains of *S. rueppellii* for use in IPM.

An interesting future comparison could be to look at the differences in olfactory genes between *S. rueppellii* and *E. pertinax* (the non-predatory European hoverfly), as this may give some indication of the genes involved in detecting aphid pheromones and avoidance of parasitised aphids.

3.6 Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The genome and transcriptome assemblies generated in this study (as well as the raw sequencing data used to produce them) are available under bioproject: PRJEB48036.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Pest Genomics Initiative, a collaborative project between Rothamsted Research, Bayer Crop Science and Syngenta AG. Rothamsted Research gratefully acknowledges strategic funding from the UK Biotechnology and Biological and Sciences Research Council (UKRI-BBSRC). The content is solely the responsibility of the authors and does not necessarily reflect the official views of the funding bodies.

Authors' contributions

The Pest Genomics Initiative (BB, CR, EG, KH-P, LF, RK and RN) devised the original conceptual ideas. EB performed the DNA and RNA extractions with help from MW. EB assembled and annotated the genome with guidance from RK and DH. EB performed the comparative analyses. FM, RK and KH-P supervised the project. EB wrote the manuscript. All authors read, edited and approved the final manuscript.

Acknowledgements

Acknowledgements go to Dr. David R. Nelson of the International Committee on the Nomenclature for Cytochrome P450 Enzymes for classification of *Sphaerophoria rueppellii* P450s and to Dr. Michael H. Court of the UGT Nomenclature Committee for classification of *Sphaerophoria rueppellii* UGTs.

3.7 Literature Cited

[1] Oerke E-C. Crop losses to pests. J Agric Sci 2006;144:31–43.
[2] Geiger F, Bengtsson J, Berendse F, Weisser WW, Emmerson M, Morales MB, et al. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. Basic Appl Ecol 2010;11:97–105.
[3] Bottrell DG, Schoenly KG. Resurrecting the ghost of green revolutions past: The brown planthopper as a recurring threat to high-yielding rice production in tropical Asia. J Asia Pac Entomol 2012;15:122–40.
[4] Debach P, Rosen D. Biological control by natural enemies (second edition). J Trop Ecol 1992;8:216–216.
[5] Rojo S, coaut GF, coaut M-GM, coaut NJ, coaut MM. A world review of predatory hoverflies (Diptera, Syrphidae: Syrphinae) and their prey. sidalc.net; 2003.
[6] Wotton KR, Gao B, Menz MHM, Morris RKA, Ball SG, Lim KS, et al. Mass Seasonal Migrations of Hoverflies Provide Extensive Pollination and Crop Protection Services. Curr Biol 2019;29:2167–73.e5.
[7] Ramsden M, Menendez R, Leather S, Wäckers F. Do natural enemies really make a difference? Field scale impacts of parasitoid wasps and hoverfly larvae on cereal aphid populations. Agric For Entomol 2017;19:139–45.
[8] Tenhumberg B, Poehling H-M. Syrphids as natural enemies of cereal aphids in Germany: Aspects of their biology and efficacy in different years and regions. Agric Ecosyst Environ 1995;52:39–43.
[9] Tenhumberg B. Predicting predation efficiency of biocontrol agents: linking behavior of individuals and population dynamics. International Congress on Environmental Modelling and Software, scholarsarchive.byu.edu; 2004.
[10] ROTHERAY, GE. Colour guide to hoverfly larvae (Diptera, Syrphidae). Dipter Dig 1993;9:1–155.
[11] Dunn L, Lequerica M, Reid CR, Latty T. Dual ecosystem services of syrphid flies (Diptera: Syrphidae): pollinators and biological control agents. Pest Manag Sci 2020;76:1973–9.
[12] Hopper JV, Nelson EH, Daane KM, Mills NJ. Growth, development and consumption by four syrphid species associated with the lettuce aphid, Nasonovia ribisnigri, in California. Biol Control 2011;58:271–6.
[13] Pekas A, De Craecker I, Boonen S, Wäckers FL, Moerkens R. One stone; two birds: concurrent pest control and pollination services provided by aphidophagous hoverflies. Biol Control 2020;149:104328.
[14] Mizuno M, Itioka T, Tatematsu Y, Itô Y. Food utilization of aphidophagous hoverfly larvae (Diptera: Syrphidae, Chamaemyiidae) on herbaceous plants in an urban habitat. Ecol Res 1997;12:239–48.
[15] Dib H, Simon S, Sauphanor B, Capowiez Y. The role of natural enemies on the population dynamics of the rosy apple aphid, Dysaphis plantaginea Passerini (Hemiptera: Aphididae) in organic apple orchards in south-eastern France. Biol Control 2010;55:97–109.
[16] Messelink GJ, Janssen A. Increased control of thrips and aphids in greenhouses with two species of generalist predatory bugs involved in intraguild predation. Biol Control 2014;79:1–7.
[17] Freier B, Triltsch H, Möwes M, Moll E. The potential of predators in natural control of aphids in wheat: Results of a ten-year field study in two German landscapes. Biocontrol 2007;52:775–88.
[18] Brewer MJ, Elliott NC. Biological control of cereal aphids in north america and mediating effects of host plant and habitat manipulations. Annu Rev Entomol 2004;49:219–42.
[19] Rotheray GE, Gilbert F. The natural history of hoverflies. Forrest text; 2011.
[20] Rader R, Cunningham SA, Howlett BG, Inouye DW. Non-Bee Insects as Visitors and Pollinators of Crops: Biology, Ecology, and Management. Annu Rev Entomol 2020;65:391–407.
[21] Doyle T, Hawkes WLS, Massy R, Powney GD, Menz MHM, Wotton KR. Pollination by hoverflies in the Anthropocene. Proc Biol Sci 2020;287:20200508.
[22] Hodgkiss D, Brown MJF, Fountain MT. The effect of within-crop floral resources on pollination, aphid control and fruit quality in commercial strawberry. Agric Ecosyst Environ 2019;275:112–22.
[23] Jauker F, Wolters V. Hover flies are efficient pollinators of oilseed rape. Oecologia 2008;156:819–23.
[24] Thompson FC, Rotheray GE, Zumbado MA, Brown BV, Borkent A, Cumming JM, et al. Manual of Central American Diptera 2010.
[25] Darwin Tree of Life – Reading the genomes of all life: a new platform for understanding our biodiversity n.d. https://www.darwintreeoflife.org/ (accessed July 22, 2021).
[26] Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. Genome Biol Evol 2016;8:1762–75.
[27] Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. Science 2010;327:343–8.
[28] Bailey E, Field L, Rawlings C, King R, Mohareb F, Pak K-H, et al. A scaffold-level genome assembly of the pirate bug, Orius laevigatus, and a comparative analysis of insecticide resistance-related gene families with hemipteran crop pests. Research Square 2021. https://doi.org/10.21203/rs.3.rs-537204/v1.
[29] Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, et al. Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles. Nat Commun 2018;9:1–13.
[30] European Commission. Directive 2009/128/EC on the sustainable use of pesticides. Official Journal of the European Union; 2009. https://doi.org/10.2861/78.

[31] Cameron PJ, Walker GP, Hodson AJ, Kale AJ, Herman TJB. Trends in IPM and insecticide use in processing tomatoes in New Zealand. Crop Prot 2009;28:421–7.
[32] Kranthi KR, Russell DA. Changing Trends in Cotton Pest Management. In: Peshin R, Dhawan AK, editors. Integrated Pest Management: Innovation-Development Process: Volume 1, Dordrecht: Springer Netherlands; 2009, p. 499–541.
[33] Meissle M, Mouron P, Musa T, Bigler F, Pons X, Vasileiadis VP, et al. Pests, pesticide use and alternative options in European maize production: current status and future prospects. J Appl Entomol 2009;134:357–75.
[34] Hillocks RJ. Farming with fewer pesticides: EU pesticide review and resulting challenges for UK agriculture. Crop Prot 2012;31:85–93.
[35] Lechenet M, Dessaint F, Py G, Makowski D, Munier-Jolain N. Reducing pesticide use while preserving crop productivity and profitability on arable farms. Nature Plants 2017;3:1–6.
[36] Heckel DG. Insecticide Resistance After Silent Spring. Science 2012;337:1612–4.
[37] Li X, Shi H, Gao X, Liang P. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in Plutella xylostella (L.). Pest Manag Sci 2018;74:695–704.
[38] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[39] Pavlidi N, Vontas J, Van Leeuwen T. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. Curr Opin Insect Sci 2018;27:97–102.
[40] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[41] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[42] Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Current Opinion in Insect Science 2019;31:131–8.
[43] Roderick GK, Navajas M. Genes in new environments: genetics and evolution in biological control. Nat Rev Genet 2003;4:889–99.
[44] Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 2011;27:764–70.
[45] Rhyker Ranallo-Benavidez T, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 2020;11:1–10.
[46] Bradnam K. A script to calculate a basic set of metrics from a genome assembly 2011. http://korflab.ucdavis.edu/Datasets/Assemblathon/Assemblathon2/Basic_metrics/assemblathon_stats.pl.
[47] Simão 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.
[48] Andrews S. FastQC. Github; n.d.
[49] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
[50] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 2019;37:540–6.
[51] Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. Assembly of long error-prone reads using de Bruijn graphs. Proc Natl Acad Sci U S A 2016;113:E8396–405.
[52] Kajitani R, Yoshimura D, Okuno M, Minakuchi Y, Kagoshima H, Fujiyama A, et al. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat Commun 2019;10:1702.
[53] Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res 2016;44:e147.
[54] Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 2018;19:460.
[55] Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 2016;3:95–8.
[56] Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 2017;356:92–5.
[57] Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst 2016;3:99–101.
[58] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One 2014;9:e112963.
[59] Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 2011;12:1–14.
[60] Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 2004;32:W309–12.
[61] Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 2005;33:6494–506.
[62] Solovyev V. Statistical approaches in eukaryotic gene prediction. In: DJ. Balding, M. Bishop, C. Cannings, editor. Handbook of Statistical Genetics, Chichester: John Wiley & Sons Ltd; 2001, p. 83–127.
[63] Haas BJ, 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;9:1–22.
[64] Smit AFA, Hubley R. RepeatModeler Open-1.0 2008-2015. http://www.repeatmasker.org.

[65] Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics 2013;29:2487–9.
[66] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42:D222–30.
[67] Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0 2013-2015. http://www.repeatmasker.org.
[68] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–60.
[69] Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33:290–5.
[70] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644–52.
[71] Gilbert D. EvidentialGene - Evidence Directed Gene Construction for Eukaryotes 2013. https://sourceforge.net/projects/evidentialgene/.
[72] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36:3420–35.
[73] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 2013;29:2933–5.
[74] Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 2013;69:313–9.
[75] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 2019;20:1–14.
[76] Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. bioRxiv 2018:267914. https://doi.org/ 10.1101/267914.
[77] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–80.
[78] Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059–66.
[79] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.
[80] Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol 2008;25:1307–20.
[81] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.
[82] Nelson DR. The Cytochrome P450 Homepage. Hum Genomics 2009;4:1–7.
[83] UGT Committee. UGT Committee Home. UGT Committee Home n.d. https://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage (accessed March 25, 2021).
[84] Pflug JM, Holmes VR, Burrus C, Spencer Johnston J, Maddison DR. Measuring genome sizes using read-depth, k-mers, and flow cytometry: methodological comparisons in beetles (Coleoptera). bioRxiv 2019:761304. https://doi.org/ 10.1101/761304.
[85] Hawkes W, Wotton K, University of Oxford and Wytham Woods Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, et al. The genome sequence of the tapered dronefly, Eristalis pertinax (Scopoli, 1763). Wellcome Open Res 2021;6:292.
[86] Hanrahan SJ, Johnston JS. New genome size estimates of 134 species of arthropods. Chromosome Res 2011;19:809–23.
[87] Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, et al. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol 2019;19:11.
[88] Chen M, Peng K, Su C, Wang Y, Hao J. The complete mitochondrial genome of Syrphus ribesii (Diptera: Syrphoidea: Syrphidae). Mitochondrial DNA B Resour 2021;6:519–21.
[89] Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, et al. Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A 2011;108:5690–5.
[90] Zhao J, Xu L, Sun Y, Song P, Han Z. UDP-Glycosyltransferase Genes in the Striped Rice Stem Borer, (Walker), and Their Contribution to Chlorantraniliprole Resistance. Int J Mol Sci 2019;20. https://doi.org/ 10.3390/ijms20051064.
[91] Tian F, Wang Z, Li C, Liu J, Zeng X. UDP-Glycosyltransferases are involved in imidacloprid resistance in the Asian citrus psyllid, Diaphorina citri (Hemiptera: Lividae). Pestic Biochem Physiol 2019;154:23–31.
[92] Cui X, Wang C, Wang X, Li G, Liu Z, Wang H, et al. Molecular Mechanism of the UDP-Glucuronosyltransferase 2B20-like Gene (AccUGT2B20-like) in Pesticide Resistance of Apis cerana cerana. Front Genet 2020;11:592595.
[93] Ahn S-J, Marygold SJ. The UDP-Glycosyltransferase Family in Drosophila melanogaster: Nomenclature Update, Gene Expression and Phylogenetic Analysis. Front Physiol 2021;12:648481.
[94] Zhou Y, Fu W-B, Si F-L, Yan Z-T, Zhang Y-J, He Q-Y, et al. UDP-glycosyltransferase genes and their association and mutations associated with pyrethroid resistance in Anopheles sinensis (Diptera: Culicidae). Malar J 2019;18:62.
[95] Xu J, Strange JP, Welker DL, James RR. Detoxification and stress response genes expressed in a western North American bumble bee, Bombus huntii (Hymenoptera: Apidae). BMC Genomics 2013;14:874.
[96] Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol 2016;14:1–15.
[97] Pan Y, Xu P, Zeng X, Liu X, Shang Q. Characterization of UDP-Glucuronosyltransferases and the Potential Contribution to Nicotine Tolerance in Myzus persicae. Int J Mol Sci 2019;20:3637.
[98] Xie W, He C, Fei Z, Zhang Y. Chromosome-level genome assembly of the greenhouse whitefly (Trialeurodes vaporariorum Westwood). Mol Ecol Resour 2020;20:995–1006.
[99] Guo S-K, Cao L-J, Song W, Shi P, Gao Y-F, Gong Y-J, et al. Chromosome-level assembly of the melon thrips genome yields insights

into evolution of a sap-sucking lifestyle and pesticide resistance. Mol Ecol Resour 2020;20:1110–25.
[100]Guo L, Xie W, Yang Z, Xu J, Zhang Y. Genome-Wide Identification and Expression Analysis of Udp-Glucuronosyltransferases in the Whitefly Bemisia Tabaci (Gennadius) (HemipterA: Aleyrodidae). Int J Mol Sci 2020;21. https://doi.org/ 10.3390/ijms21228492.
[101]Dermauw W, Pym A, Bass C, Van Leeuwen T, Feyereisen R. Does host plant adaptation lead to pesticide resistance in generalist herbivores? Curr Opin Insect Sci 2018;26:25–33.
[102]Pan Y, Tian F, Wei X, Wu Y, Gao X, Xi J, et al. Thiamethoxam Resistance in Aphis gossypii Glover Relies on Multiple UDP-Glucuronosyltransferases. Front Physiol 2018;9:322.
[103]Lewis JA, Szilagyi M, Gehman E, Dennis WE, Jackson DA. Distinct patterns of gene and protein expression elicited by organophosphorus pesticides in Caenorhabditis elegans. BMC Genomics 2009;10:202.
[104]Lumjuan N, Rajatileka S, Changsom D, Wicheer J, Leelapat P, Prapanthadara L-A, et al. The role of the Aedes aegypti Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. Insect Biochem Mol Biol 2011;41:203–9.
[105]Vontas JG, Small GJ, Hemingway J. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens. Biochem J 2001;357:65–72.
[106]Corona M, Robinson GE. Genes of the antioxidant system of the honey bee: annotation and phylogeny. Insect Mol Biol 2006;15:687–701.
[107]Low WY, Ng HL, Morton CJ, Parker MW, Batterham P, Robin C. Molecular evolution of glutathione S-transferases in the genus Drosophila. Genetics 2007;177:1363–75.
[108]Lumjuan N, Stevenson BJ, Prapanthadara L-A, Somboon P, Brophy PM, Loftus BJ, et al. The Aedes aegypti glutathione transferase family. Insect Biochem Mol Biol 2007;37:1026–35.
[109]Ding Y, Ortelli F, Rossiter LC, Hemingway J, Ranson H. The Anopheles gambiae glutathione transferase supergene family: annotation, phylogeny and expression profiles. BMC Genomics 2003;4:35.
[110]Yan L, Yang P, Jiang F, Cui N, Ma E, Qiao C, et al. Transcriptomic and phylogenetic analysis of Culex pipiens quinquefasciatus for three detoxification gene families. BMC Genomics 2012;13:609.
[111]Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, et al. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol Biol 2006;15:615–36.
[112]Aidlin Harari O, Santos-Garcia D, Musseri M, Moshitzky P, Patel M, Visendi P, et al. Molecular Evolution of the Glutathione S-Transferase Family in the Bemisia tabaci Species Complex. Genome Biol Evol 2020;12:3857–72.
[113]Friedman R. Genomic organization of the glutathione S-transferase family in insects. Mol Phylogenet Evol 2011;61:924–32.
[114]Gawande ND, Subashini S, Murugan M, Subbarayalu M. Molecular screening of insecticides with sigma glutathione S-transferases (GST) in cotton aphid Aphis gossypii using docking. Bioinformation 2014;10:679–83.
[115]Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[116]Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol 2000;45:371–91.
[117]Bisset JA, Marín R, Rodríguez MM, Severson DW, Ricardo Y, French L, et al. Insecticide resistance in two Aedes aegypti (Diptera: Culicidae) strains from Costa Rica. J Med Entomol 2013;50:352–61.
[118]Oakeshott J, Claudianos C, Campbell PM. Biochemical genetics and genomics of insect esterases. Molecular Insect … 2010.
[119]Rotenberg D, Baumann AA, Ben-Mahmoud S, Christiaens O, Dermauw W, Ioannidis P, et al. Genome-enabled insights into the biology of thrips as crop pests. BMC Biol 2020;18:142.
[120]Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol 2010;19 Suppl 2:155–64.
[121]Xia J, Xu H, Yang Z, Pan H, Yang X, Guo Z, et al. Genome-Wide Analysis of Carboxylesterases (COEs) in the Whitefly, (Gennadius). Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20204973.
[122]Karatolos N. Molecular mechanisms of insecticide resistance in the greenhouse whitefly, Trialeurodes vaporariorum. PhD. University of Exeter, 2011. https://doi.org/ https://ore.exeter.ac.uk/repository/bitstream/handle/10036/3350/KaratolosN.pdf.
[123]Bass C, Field LM. Gene amplification and insecticide resistance. Pest Manag Sci 2011;67:886–90.
[124]Ono M, Swanson JJ, M. Field L, Devonshire AL, D. Siegfried B. Amplification and methylation of an esterase gene associated with insecticide-resistance in greenbugs, Schizaphis graminum (Rondani) (Homoptera: Aphididae). Insect Biochem Mol Biol 1999;29:1065–73.
[125]Vontas JG, Small GJ, Hemingway J. Comparison of esterase gene amplification, gene expression and esterase activity in insecticide susceptible and resistant strains of the brown planthopper, Nilaparvata lugens (Stål). Insect Mol Biol 2000;9:655–60.
[126]Raymond M, Chevillon C, Guillemaud T, Lenormand T, Pasteur N. An overview of the evolution of overproduced esterases in the mosquito Culex pipiens. Philos Trans R Soc Lond B Biol Sci 1998;353:1707–11.
[127]Vaughan A, Rodriguez M, Hemingway J. The independent gene amplification of electrophoretically indistinguishable B esterases from the insecticide-resistant mosquito Culex quinquefasciatus. Biochem J 1995;305 (Pt 2):651–8.
[128]Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[129]Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 2014;45:89–110.
[130]Xiao L-F, Zhang W, Jing T-X, Zhang M-Y, Miao Z-Q, Wei D-D, et al. Genome-wide identification, phylogenetic analysis, and expression profiles of ATP-binding cassette transporter genes in the oriental fruit fly, Bactrocera dorsalis (Hendel) (Diptera: Tephritidae). Comp Biochem Physiol Part D Genomics Proteomics 2018;25:1–8.

[131]Lu H, Xu Y, Cui F. Phylogenetic analysis of the ATP-binding cassette transporter family in three mosquito species. Pestic Biochem Physiol 2016;132:118–24.
[132]Liu S, Zhou S, Tian L, Guo E, Luan Y, Zhang J, et al. Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, Bombyx mori. BMC Genomics 2011;12:491.
[133]Figueira-Mansur J, Schrago CG, Salles TS, Alvarenga ESL, Vasconcellos BM, Melo ACA, et al. Phylogenetic analysis of the ATP-binding cassette proteins suggests a new ABC protein subfamily J in Aedes aegypti (Diptera: Culicidae). BMC Genomics 2020;21:463.
[134]He Q, Yan Z, Si F, Zhou Y, Fu W, Chen B. ATP-Binding Cassette (ABC) Transporter Genes Involved in Pyrethroid Resistance in the Malaria Vector Anopheles sinensis: Genome-Wide Identification, Characteristics, Phylogenetics, and Expression Profile. Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20061409.
[135]Pan Y, Zeng X, Wen S, Gao X, Liu X, Tian F, et al. Multiple ATP-binding cassette transporters genes are involved in thiamethoxam resistance in Aphis gossypii glover. Pestic Biochem Physiol 2020;167:104558.
[136]Pym A, Singh KS, Nordgren Å, Emyr Davies TG, Zimmer CT, Elias J, et al. Host plant adaptation in the polyphagous whitefly, Trialeurodes vaporariorum , is associated with transcriptional plasticity and altered sensitivity to insecticides. BMC Genomics 2019;20:1–19.
[137]Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, Bemisia tabaci. BMC Genomics 2017;18:1–16.
[138]Feyereisen R. INSECT P450 ENZYMES. Annu Rev Entomol 1999;44:507–33.
[139]Feyereisen R. Evolution of insect P450. Biochem Soc Trans 2006;34:1252–5.
[140]Karunker I, Benting J, Lueke B, Ponge T, Nauen R, Roditakis E, et al. Over-expression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). Insect Biochem Mol Biol 2008;38:634–44.
[141]Liang X, Xiao D, He Y, Yao J, Zhu G, Zhu KY. Insecticide-mediated up-regulation of cytochrome P450 genes in the red flour beetle (Tribolium castaneum). Int J Mol Sci 2015;16:2078–98.
[142]Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. PLoS Genet 2010;6:e1000999.
[143]Yang T, Liu N. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, Culex quinquefasciatus. PLoS One 2011;6:e29418.
[144]Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[145]Main BJ, Everitt A, Cornel AJ, Hormozdiari F, Lanzaro GC. Genetic variation associated with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii. Parasit Vectors 2018;11:225.
[146]Vlogiannitis S, Mavridis K, Dermauw W, Snoeck S, Katsavou E, Morou E, et al. Reduced proinsecticide activation by cytochrome P450 confers coumaphos resistance in the major bee parasite Varroa destructor. Proc Natl Acad Sci U S A 2021;118. https://doi.org/10.1073/pnas.2020380118.
[147]Scott JG, 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.
[148]The genome of the model beetle and pest Tribolium castaneum. Nature 2008;452:949–55.
[149]Dermauw W, Van Leeuwen T, Feyereisen R. Diversity and evolution of the P450 family in arthropods. Insect Biochem Mol Biol 2020;127:103490.
[150]Ilias A, Lagnel J, Kapantaidaki DE, Roditakis E, Tsigenopoulos CS, Vontas J, et al. Transcription analysis of neonicotinoid resistance in Mediterranean (MED) populations of B. tabaci reveal novel cytochrome P450s, but no nAChR mutations associated with the phenotype. BMC Genomics 2015;16:1–23.
[151]Sztal T, Chung H, Berger S, Currie PD, Batterham P, Daborn PJ. A cytochrome p450 conserved in insects is involved in cuticle formation. PLoS One 2012;7:e36544.
[152]Zhang H, Zhao M, Liu Y, Zhou Z, Guo J. Identification of cytochrome P450 monooxygenase genes and their expression in response to high temperature in the alligatorweed flea beetle Agasicles hygrophila (Coleoptera: Chrysomelidae). Sci Rep 2018;8:17847.
[153]Scharf ME, Parimi S, Meinke LJ, Chandler LD, Siegfried BD. Expression and induction of three family 4 cytochrome P450 (CYP4)* genes identified from insecticide-resistant and susceptible western corn rootworms, Diabrotica virgifera virgifera. Insect Mol Biol 2001;10:139–46.
[154]Shi W, Sun J, Xu B, Li H. Molecular characterization and oxidative stress response of a cytochrome P450 gene (CYP4G11) from Apis cerana cerana. Z Naturforsch C 2013;68:509–21.
[155]Ingham VA, Jones CM, Pignatelli P, Balabanidou V, Vontas J, Wagstaff SC, et al. Dissecting the organ specificity of insecticide resistance candidate genes in Anopheles gambiae: known and novel candidate genes. BMC Genomics 2014;15:1018.
[156]Berenbaum MR, Johnson RM. Xenobiotic detoxification pathways in honey bees. Curr Opin Insect Sci 2015;10:51–8.
[157]Mao W, Rupasinghe SG, Johnson RM, Zangerl AR, Schuler MA, Berenbaum MR. Quercetin-metabolizing CYP6AS enzymes of the pollinator Apis mellifera (Hymenoptera: Apidae). Comp Biochem Physiol B Biochem Mol Biol 2009;154:427–34.
[158]Johnson RM, Mao W, Pollock HS, Niu G, Schuler MA, Berenbaum MR. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis mellifera. PLoS One 2012;7:e31051.
[159]Zhao M, Dong Y, Ran X, Guo X, Xing D, Zhang Y, et al. Sodium channel point mutations associated with pyrethroid resistance in Chinese strains of Culex pipiens quinquefasciatus (Diptera: Culicidae). Parasit Vectors 2014;7:369.
[160]Weill M, Malcolm C, Chandre F, Mogensen K, Berthomieu A, Marquine M, et al. The unique mutation in ace-1 giving high insecticide

resistance is easily detectable in mosquito vectors. Insect Mol Biol 2004;13:1–7.
[161]Singh OP, Dykes CL, Das MK, Pradhan S, Bhatt RM, Agrawal OP, et al. Presence of two alternative kdr-like mutations, L1014F and L1014S, and a novel mutation, V1010L, in the voltage gated Na+ channel of Anopheles culicifacies from Orissa, India. Malar J 2010;9:146.
[162]Tan WL, Li CX, Wang ZM, Liu MD, Dong YD, Feng XY, et al. First detection of multiple knockdown resistance (kdr)-like mutations in voltage-gated sodium channel using three new genotyping methods in Anopheles sinensis from Guangxi Province, China. J Med Entomol 2012;49:1012–20.
[163]Liu N, Pridgeon JW. Metabolic detoxication and the kdr mutation in pyrethroid resistant house flies, Musca domestica (L.). Pestic Biochem Physiol 2002;73:157–63.
[164]Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL. Identification of mutations in the houseflypara-type sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. Mol Gen Genet 1996;252:51–60.
[165]Vontas JG, Hejazi MJ, Hawkes NJ, Cosmidis N, Loukas M, Janes RW, et al. Resistance-associated point mutations of organophosphate insensitive acetylcholinesterase, in the olive fruit fly Bactrocera oleae. Insect Mol Biol 2002;11:329–36.
[166]Kozaki T, Shono T, Tomita T, Kono Y. Fenitroxon insensitive acetylcholinesterases of the housefly, Musca domestica associated with point mutations. Insect Biochem Mol Biol 2001;31:991–7.
[167]da Silva NM, de Carvalho RA, de Azeredo-Espin AML. Acetylcholinesterase cDNA sequencing and identification of mutations associated with organophosphate resistance in Cochliomyia hominivorax (Diptera: Calliphoridae). Vet Parasitol 2011;177:190–5.
[168]Ffrench-Constant RH, Rocheleau TA, Steichen JC, Chalmers AE. A point mutation in a Drosophila GABA receptor confers insecticide resistance. Nature 1993;363:449–51.
[169]Scott JG. Evolution of resistance to pyrethroid insecticides in Musca domestica. Pest Manag Sci 2017;73:716–22.
[170]Yang C, Huang Z, Li M, Feng X, Qiu X. RDL mutations predict multiple insecticide resistance in Anopheles sinensis in Guangxi, China. Malar J 2017;16:482.
[171]Troczka B, Zimmer CT, Elias J, Schorn C, Bass C, Davies TGE, et al. Resistance to diamide insecticides in diamondback moth, Plutella xylostella (Lepidoptera: Plutellidae) is associated with a mutation in the membrane-spanning domain of the ryanodine receptor. Insect Biochem Mol Biol 2012;42:873–80.
[172]Lueke B, Douris V, Hopkinson JE, Maiwald F, Hertlein G, Papapostolou K-M, et al. Identification and functional characterization of a novel acetyl-CoA carboxylase mutation associated with ketoenol resistance in Bemisia tabaci. Pestic Biochem Physiol 2020;166:104583.
[173]Pan Y, Zhu E, Gao X, Nauen R, Xi J, Peng T, et al. Novel mutations and expression changes of acetyl-coenzyme A carboxylase are associated with spirotetramat resistance in Aphis gossypii Glover. Insect Mol Biol 2017;26:383–91.
[174]Hardstone MC, Scott JG. Is Apis mellifera more sensitive to insecticides than other insects? Pest Manag Sci 2010;66:1171–80.
[175]Schmehl DR, Teal PEA, Frazier JL, Grozinger CM. Genomic analysis of the interaction between pesticide exposure and nutrition in honey bees (Apis mellifera). J Insect Physiol 2014;71:177–90.
[176]Manjon C, Troczka BJ, Zaworra M, Beadle K, Randall E, Hertlein G, et al. Unravelling the Molecular Determinants of Bee Sensitivity to Neonicotinoid Insecticides. Curr Biol 2018;28:1137–43.e5.
[177]Calvo-Agudo M, González-Cabrera J, Picó Y, Calatayud-Vernich P, Urbaneja A, Dicke M, et al. Neonicotinoids in excretion product of phloem-feeding insects kill beneficial insects. Proc Natl Acad Sci U S A 2019;116:16817–22.

Chapter 4. *Microctonus brassicae* genome assembly, functional annotation and comparative genomics

This chapter is ready for submission to a journal at a future date.

A scaffold-level genome assembly of the parasitic wasp, *Microctonus brassicae* (Hymenoptera: Braconidae), and a comparative analysis of insecticide resistance-related gene families in hemipteran crop pests

Emma Bailey*†‡, Linda Field*, Christopher Rawlings†, Rob King†, Fady Mohareb‡, Keywan-Hassani Pak†, David Hughes†, Martin Williamson*, Eric Ganko**, Benjamin Buer§, Ralf Nauen§

* Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, UK

† Department of Computational and Analytical Sciences, Rothamsted Research, Harpenden, UK

‡ The Bioinformatics Group, Cranfield Soil and Agrifood Institute, Cranfield University, Cranfield, UK

** Seeds Research, Syngenta Crop Protection, LLC, Research Triangle Park, NC, USA.

§ Bayer AG, Crop Science Division, R&D, Monheim, Germany

4.1 Abstract

Background: *Microctonus brassicae* (Hymenoptera: Braconidae) is a parasitoid wasp of *Psylliodes chrysocephala*, commonly known as the cabbage-stem flea beetle, a major pest of oilseed rape across Europe. The neonicotinoid seed treatment ban and subsequent widespread pyrethroid resistance in *P. chrysocephala*, has resulted in such severe pest problems that for many farmers it is no longer economically viable to grow oilseed rape. *M. brassicae* could therefore have a vital role to play in future integrated pest management (IPM) strategies for this pest. No genomic information is currently available for *M. brassicae*. Without genomic information for such beneficial predator species, we are unable to perform comparative analyses of insecticide target-sites and genes encoding metabolic enzymes potentially responsible for insecticide resistance, between crop pests and their predators. These metabolic mechanisms include several gene families - cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione-S- transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxyl/cholinesterases (CCEs).

Methods and findings: In this study, a high-quality scaffold-level *de novo* genome assembly for *M. brassicae* has been generated using high quality PacBio HiFi data. The final assembly consisted of 109 scaffolds, with an N50 of 7.1Mb, a total genome size of 139Mb and level of completeness of 98.5% using a set of 1,658 core insect genes present as full-length genes. The assembly was annotated with 11,873 protein-coding genes and the resultant gene set achieved a completeness of 95.9%.

Comparative analysis revealed differences in genes encoding potential insecticide tolerance/resistance mechanisms between *M. brassicae* and crop pests. This included lineage specific expansions of *M. brassicae* gene families associated with insecticide detoxification: UGTs, P450s and CCEs. Conversely, other insecticide-related gene families showed minimal

lineage specific expansion in *M. brassicae* whilst exhibiting increased numbers in crop pests, such as GSTs and ABCs.

Conclusion and significance: This assembly is the first published genome for a predatory member of the *Microctonus* genus and will serve as a useful resource for further research into selectivity and potential tolerance of insecticides by beneficial predators. Furthermore, the expansion of gene families often linked to insecticide resistance and selectivity may be an indicator of the capacity of this predator to detoxify IPM selective insecticides. These findings could be exploited by targeted insecticide screens and functional studies to increase effectiveness of IPM strategies, which aim to increase crop yields by sustainably and effectively controlling pests without impacting beneficial predator populations.

Keywords: *Microctonus brassicae*; *Psylliodes chrysocephala*; Parasitic wasp; Parasitoid; PacBio HiFi; Whole genome sequencing; beneficial predator; insecticide resistance; comparative genomics; Hymenoptera; crop pests

4.2 Introduction

Every year, insect pests are responsible for worldwide crop yield losses of 18-26% at a value of over £340 billion. These percentages are only increasing with the constant evolution of insecticide resistance (as well as a warming climate) [1–3]. With a growing world population to feed, it is therefore more important than ever to find alternatives to insecticides to maintain and increase crop yields. Integrated pest management (IPM) strategies aim to take a sustainable approach to managing pests - minimising damage to human health and the environment, as well as economic risk. They focus on monitoring pest levels to ensure they remain below the economic threshold and applying control methods when levels threaten to exceed this threshold. Such measures may include cultural prevention, e.g. site and cultivar selection; mechanical controls, e.g. traps and tilling; biological control, e.g. attracting or releasing beneficial insects which eat or parasitize target pests and chemical control i.e. applying selective pesticides only when necessary [4–7].

Microctonus brassicae (Hymenoptera: Braconidae), shown in figure 1, belongs to the Braconidae family, one of the largest hymenopteran families, with over 15,000 species of parasitic wasp associated with hosts from over 120 insect families, making it an excellent source of biological control agents [8,9]. Many species of parasitoid wasps have been successfully used as biological control agents against agricultural pests [10–15], including another *Microctonus* species which had such a high field parasitism level nine years after it was introduced, that insecticide use was considered unnecessary [16]. Parasitoid wasps have developed a wide range of physiological and behavioral mechanisms to effectively locate and parasitize their respective host [17–21]. Their host specificity makes them perfectly suited for IPM strategies, as they are unlikely to have potentially harmful impacts on non-target insect populations. Alongside this, there is minimal evidence of host resistance developing towards parasitoids [22].

Figure 1. Female *Microtonus brassicae* ovipositing an adult *Psylliodes chrysocephala* [23]**.**

M. brassicae parasitises *Psylliodes chrysocephala*, commonly known as the cabbage-stem flea beetle [23]. *P. chrysocephala* is a major pest of oilseed rape within the UK and across Europe [24–28] and was traditionally controlled using neonicotinoid seed treatments in combination with pyrethroid insecticides. However, following the 2013 EU neonicotinoid seed treatment ban, a shift to heavy pyrethroid use has resulted in widespread pyrethroid resistance within *P. chrysocephala* populations [29–32]. As a result, crop yield losses have soared dramatically over the past few years, with many growers in the UK ceasing future growth of oilseed rape [33–35]. The knock on effect of this is a loss of flowering crops in the spring, which could be detrimental to foraging pollinators - the very group which the neonicotinoid seed treatment ban was intended to protect. An alternative pest control strategy which does not rely solely on pyrethroid application is therefore of utmost importance, and biological control agents such as

M. brassicae should play a large role in such future strategies. Research has shown that *M. brassicae* can be successfully reared in captivity (although this is currently slow and costly) and the parasitism rate achieved in captivity suggests they could also achieve good parasitism rates in the field [23].

Pesticides can have a negative impact on parasitoid populations and their ability to act as successful biological control agents, either through acute toxicity leading to reduction in population size or sublethal effects which could impact their ability to reproduce and parasitize even if they initially survive pesticide exposure. This can occur as a result of many factors, not limited to: insecticide odours inhibiting parasitoid detection of herbivore-induced plant volatiles; reduced longevity; reduced oviposition capability; reduction in viable eggs; reduced emergence rates from the host; reduced pupation and metamorphosis success and reduced foraging ability [36–46]. The use of insecticides alongside parasitoid biological control agents should therefore be carefully assessed prior to use, with particular attention paid to the sublethal effects as opposed to focusing on mortality levels from direct insecticide exposure.

Genomes are currently available for several parasitoid wasp species, including three *Nasonia* wasp species [47], six *Cotesia* wasp species [48], *Microplitis demolitor* [49], *Fopius arisanus* [50], *Pteromalus puparum* [51], *Diachasma alloleum* [52] *and Habrobracon hebetor* [53]. This study presents the first genome for the *Microctonus* genus.

Increasing the availability of beneficial predator / parasitoid genomes permits for comparative analyses of predator and pest genomes, focusing on potential differences in insecticide tolerance mechanisms based on both target site selectivity and metabolism.

There are two main types of insecticide resistance mechanisms: mutations in insecticide target genes that prevent the insecticide binding to the target [54] duplication or increased expression of genes encoding enzymes which can metabolise insecticides Gene families

associated with metabolic resistance include cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione S-transferases (GSTs), UDP-glucosyltransferases (UGTs) and carboxyl/choline esterases (CCEs) [55–60]. Comparisons of these mechanisms in beneficial predators and crop pests could help identify insecticides which can target crop pests but have limited impact on beneficial predator populations. This information could prove key to developing successful IPM strategies which exploit differences in insecticide selectivity between the predator and crop pests. Improving the availability of beneficial predator genomes could also aid the selection of beneficial predators with genes/mutations for insecticide resistance before being released in the field for biological control [61].

The genome assembly and comparative analysis presented here provide a comprehensive foundation for further study of insecticide tolerance and selectivity mechanisms in beneficial predators and how they compare to crop pests.

4.3 Materials and Methods

4.3.1 Sample preparation and sequencing

In July/August 2019, live *Psylliodes chrysocephala* adults were collected from oilseed rape pods freshly harvested from the fields at Rothamsted Research, Harpenden, Hertfordshire, using a hand-held battery-powered pooter.

Long-read HiFi data was obtained through sequencing of a *P. chrysocephala* adult which was parasitised with *M. brassicae*. Coverage was calculated by mapping the raw HiFi reads to the *M. brassicae* genome using Minimap2 [62].

4.3.2 De novo genome assembly

HiFi data from the parasitised *P. chrysocephala* adult was assembled using Hifiasm v0.10 (r299) [63] with default settings. Hifiasm was chosen due to its ability to quickly generate high quality genomes and because it was designed specifically for working with HiFi data. Hifiasm also performed best compared to other HiFi assemblers in terms of genome contiguity and quality [64]. Purge Haplotigs v1.0.0 [65] was used to perform redundant contig removal from the assembly. Parameters '-l 2 -m 50 -h 80' were chosen from the coverage histogram outputted in the first step of the pipeline. This tool takes read depth coverage into consideration to reduce over-purging of repetitive regions and paralogous contigs.

The subsequent assembly was then scaffolded using Hi-C data from multiple *P. chrysocephala* individuals. The Hi-C data was processed using Juicer v1.5 [66] and used as input to the 3D-DNA de novo genome assembly pipeline (version 180922) [67] alongside the draft assembly to produce a candidate chromosome-length genome assembly. Contact matrices were generated by aligning the Hi-C dataset to the genome assembly after Hi-C scaffolding, and were then visualised using JuiceBox Assembly Tools v1.11.08 [68]. The parameters used were as follows: '--mode haploid --build-gapped-map --sort-output'.

139Mbp of sequence from the resultant assembly did not align with *P. chrysocephala* Illumina short-read sequence data, and also showed a visible lack of Hi-C contacts with the rest of the assembly. A BLAST search of this sequence data matched most closely to *M. demolitor*, confirming the presence of *M. brassicae* within the HiFi data. These scaffolds were therefore extracted to form the *M. brassicae* genome assembly.

4.3.3 Annotation

Gene prediction was performed using the MAKER v2.31.8 pipeline [69]. No transcriptome evidence was available, so gene models were inferred using protein homology evidence and *ab initio* gene prediction. Proteomes from closely related species were used: *Microplitis demolitor* (GCA_000572035.2), *Cotesia rubecula* (bipaa.genouest.org), *Cotesia vestalis* (bipaa.genouest.org), *Fopius arisanus* (GCA_000806365.1), *Diachasma alloeum* (GCA_001412515.3) and *Nasonia vitripennis* (GCA_009193385.2). Maker was run using Augustus v3.3.1 [70], GeneMark-ES v4.32 [71] and EVidenceModeler v1.1.1 [72] with default masking options.

A *de novo* species specific repeat library was constructed using RepeatModeller v1.0.7 [73] to identify repeat models. These models were searched against the GenBank non-redundant (*nr*) protein database for Arthropoda (e value $<10^{-3}$) using Blastx to remove any potential protein-coding genes. This was combined with transposon data to create a custom library. Transposons were identified from the transcriptome assembly by running HMMER: hmmscan [74] against the Pfam database [75] and filtering the resultant Pfam descriptions for those containing “transposon”. A search for transposons was also performed on transcripts produced from Maker and these transposons were then added to the custom repeat library which was used for a second round of Maker. RepeatMasker v4.0.7 [76] was used to mask repeats in the genome assembly using these repeat libraries, as well as to estimate the abundances of all predicted repeats.

M. brassicae protein sequences were aligned using Blastp against the non-redundant (nr) NCBI protein database for Arthropoda. InterProscan searches were run against several databases (CDD, HAMAP, HMMPAnther, HMMPfam, HMMPIR, FPrintScan, BlastProDom, ProfileScan, HMMTigr) for functional annotation. BLAST2GO [77] was used to assign gene ontology (GO annotations). Infernal v1.1.2 [78] was used to predict and annotate non-coding RNAs.

4.3.4 Genome quality assessment

Basic metrics from the genome assembly were calculated using a script developed for the 'Assemblathon' [79]. These metrics include scaffold/contig N50, longest and shortest scaffold length, number of scaffolds exceeding a range of lengths and number of gaps/N's in the assembly.

The completeness of the genome assembly and annotation for *M. brassicae* was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) [80] of the insect gene set (insecta odb 10). 'Genome' mode was used to assess the assembly, and 'protein' mode to assess the annotation. 'Fly' was used as the training species for Augustus gene prediction. BUSCO assessments were then run with default parameters.

4.3.5 Comparative genomics and phylogenetic analysis

To produce the species tree, orthogroup gene trees were produced using orthofinder [81] and the tree was inferred from these using the STAG method [82].

In order to identify genes potentially involved in insecticide resistance, the PFAM domains assigned to gene models during annotation (as described in the 'Genome Annotation' methods section) were used as follows: CCEs (PF00135/IPR002018), GSTs (IPR004045/PF02798), (IPR004046/PF00043), P450s (IPR001128/PF00067), ABCs (IPR003439/PF00005) and UGTs (IPR002213/PF00201). Proteins from UniProtKB for the classes of interest, from hymenopteran species, were used for BLAST queries against *M. brassicae* to identify any missed genes and to assist with subfamily assignment within these classes. Subfamily assignment for *M. brassicae* gene families was finalised using phylogenetic trees produced using MAFFT alignments [83,84]

and RaxML v8.2.11 [85]. The GAMMA LG protein model [86] was used and a bootstrap consensus tree was inferred from 100 replicates.

Manual checks and curation were performed for genes potentially involved in insecticide resistance. Increased copy numbers of genes linked to insecticide resistance often led to adjacent tandem duplications being incorrectly annotated as one gene model, therefore curation was important to prevent incorrect gene numbers being reported in later analyses. Alignment with gene models from closely related species was used to confirm the exon/intron boundaries and start/stop codons and the gene models were edited in Geneious where necessary.

The P450s were classified and named by Dr David Nelson [87]. The UGTs were classified and named by Dr Michael Court [88]. Nomenclature of P450s and UGTs is based on the evolutionary relationships of the sequences. P450 and UGT sequences were BLAST searched against named insect sequences and were assigned to known families if they were >40% (for P450 families) or >45% (for UGT families) identical. Other sequences were assigned to new families based on their clustering on trees and their percent identity to each other.

4.4 Results and Discussion

4.4.1 Sequencing

The *M. brassicae* long-read data was obtained from a parasitized *P. chrysocephala* individual. This parasitoid was concluded to be *M. brassicae*, as the adult *P. chrysocephala* was obtained from in-house colonies at Rothamsted Research, from which emerging parasitoids had already

been confirmed as being *M. brassicae*. It can be assumed that the assembly was generated from a single individual, as *M. brassicae* is a solitary koinobiont endoparasitoid, meaning that a single individual wasp develops from each host individual [23].

4.4.2 Raw data

DNA sequencing of the parasitized *P. chrysocephala* individual generated 2,074,955 PacBio HiFi reads with a total length of 29.9 Gbp, a read length N50 of 15,228bp and a mean read quality of 32.2. 34.73% of these reads mapped to the *M. brassicae* genome, giving an estimate coverage of 75x.

4.4.3 Assembly

HiFi data from the parasitised *P. chrysocephala* adult was assembled using Hifiasm and scaffolded using Hi-C data from multiple *P. chrysocephala* adults. *M. brassicae* contigs assembled independently from *P. chrysocephala*, and were extracted to produce the final assembly.

The resultant *M. brassicae* assembly consisted of 109 scaffolds, with a scaffold N50 of 7.1Mb, a total genome size of 139Mb and level of completeness of 98.5% using a set of 1,658 core insect genes present as full-length genes. Statistics of the final assembly are shown in table 1.

The genomes of other Braconid wasps vary in size from 127.9Mb (*Macrocentrus cingulum*: PRJNA361069*)* to 388.8Mb (*Diachasma alloeum*: PRJNA306876). At 139Mb, the *M. brassicae* genome is therefore at the lower end, but within the range of other currently sequenced Braconid wasps.

Table 1. Final assembly statistics for the *M. brassicae* genome

Number of scaffolds	109
Total size of scaffolds	139,242,784
Longest scaffold	17,203,980
Shortest scaffold	35,569
Number of scaffolds > 10K nt	109
Number of scaffolds > 100K nt	92
Number of scaffolds > 1M nt	19
Number of scaffolds > 10M nt	4
N50 scaffold length	7,123,232
Number of N's	0
Number of gaps	0
BUSCO	C:98.5%[S:95.4%,D:3.1%], F:0.2%, M:1.3%

The GC content of the assembly was 29.66%. This falls within the GC content range of other Braconid wasps which ranges from 23.8% (*Lysiphlebus fabarum*: PRJNA587428) to 39.4% (*Fopius arisanus*: PRJNA258104).

4.4.4 Annotation

Gene prediction with MAKER identified 11,873 protein-coding genes with the proteins having a mean length of 545 amino acids. Of these 72% had a match to NCBI's non-redundant (nr) database and 92% contained InterPro motifs, domains or signatures. The longest protein found was 'titin isoform X11' at 20,104 amino acids. The final proteome has a busco completeness score of 95.9% (with 18.5% duplication). This total number of protein-coding genes fits with the number identified in other Hymenoptera species, which ranged from 10,443 (Aphidius gifuensis) to 20,226 (Aphidius ervi) [89].

From the Infernal tool inference of RNA alignments, a total of 419 non-coding RNA elements were found in the genome (table 2).

Table 2. Number of ncRNAs predicted in the *M. brassicae* genome

ncRNA element	Number of elements
tRNA	140
rRNA	118
snRNA	24
miRNA	52
srpRNA	2
snoRNA	40
lncRNA	2
Other	41

4.4.5 Repeat annotation

Transposable and repetitive elements made up 9.92% of the *M. brassicae* genome (table 3). This is close to the reported repeat content of *Apis mellifera* (8%) [90], but lower than other Hymenoptera species, which ranged from 24% for *Cotesia vestalis* to 49% for *Diachasma alloeum* [89]. Repeat content is often positively correlated with increased genome size, and therefore the small size of the *M. brassicae* genome (139Mb) likely contributes to the low repeat content. 8.79% of the *M. brassicae* genome (12,233,744bp) was masked for annotation - some repeats are annotated but not masked, such as those less than 10bp in length. The majority of these were simple repeats (6.25%). Details of transposable and repetitive elements are shown in table 3.

Table 3. Summary of transposable and repetitive elements in the *Microctonus brassicae* genome

			Number of Elements	Length Occupied (bp)	Percentage of Sequence
Retroelements			1,218	1,401,809	1.01%
	SINES		0	0	0.00%
	LINES		394	184,416	0.13%
		L2/CR1/Rex	114	46,383	0.03%
		R1/LOA/Jockey	100	47,106	0.03%
		R2/R4/NeSL	5	6,387	0.00%
		RTE/Bov-B	22	8,935	0.01%
	LTR elements		824	1,217,393	0.87%
		BEL/Pao	138	288,842	0.21%
		Ty1/Copia	114	208,123	0.15%
		Gypsy/DIRS1	572	720,428	0.52%
DNA transposons			344	67,076	0.05%
	Tc1-IS630-Pogo		169	42,275	0.03%
	PiggyBac		3	1,644	0.00%
Rolling-circles			95	21,649	0.02%
Unclassified			215	88,103	0.06%
Total interspersed repeats			-	1,556,988	1.12%
Small RNA			76	54,298	0.04%
Satellites			1	634	0.00%
Simple repeats			182,326	8,696,104	6.25%
Low complexity			36,325	1,912,693	1.37%
Total			**220,600**	**13,799,354**	**9.92%**

4.4.6 Phylogeny

OrthoFinder assigned 637,452 genes (95.3% of total) to 31,964 orthogroups. There were 1,910 orthogroups with all species present. Phylogenetic analysis correctly clustered *M. brassicae* within the Hymenoptera: Braconidae family of parasitoids (figure 1).

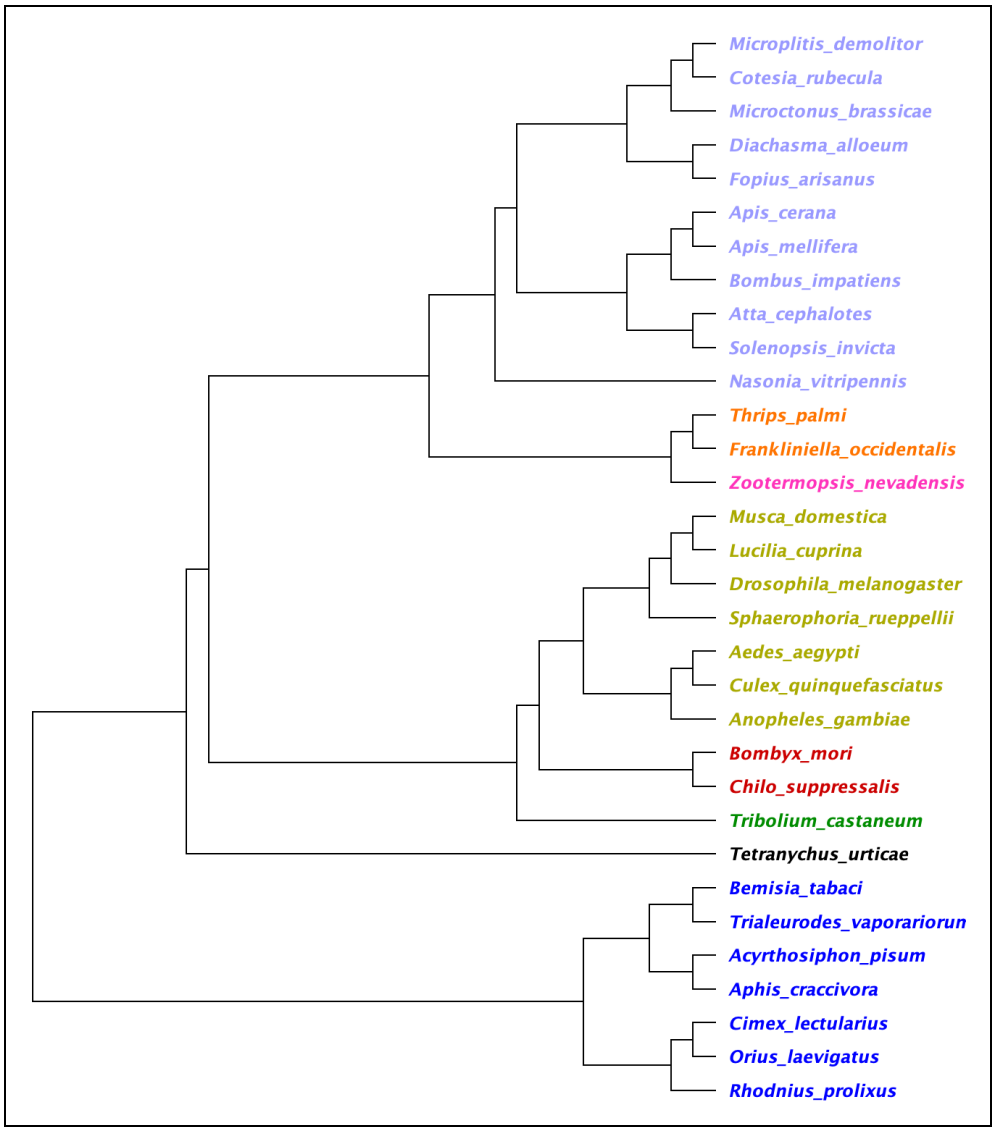


Figure 1. Phylogeny of Insecta

Nodes are coloured by order, yellow=Diptera, red=Lepidoptera, green=Coleoptera, black=Chelicerata, blue=Hemiptera, purple=Hymenoptera, orange=Thysanoptera, pink=Isoptera. Produced using the STAG tree inference method and full proteomes of the following species: *A. cephalotes: PRJNA48091, A. cerana: PRJNA235974, B. impatiens: PRJNA61101, S. invicta: PRJNA49629, D. alloeum: PRJNA284396, F. arisanus: PRJNA258104, M. demolitor: PRJNA195937, S. invicta: PRJNA49629, C. rubecula: https://bipaa.genouest.org/sp/cotesia/ (OGS1.0), D. melanogaster:* PRJNA13812, *M. domestica:* PRJNA176013, *L. cuprina:* PRJNA248412, *S. rueppellii:* (author's previous work), *C. quinquefasciatus:* PRJNA18751, *A. aegypti:* PRJNA318737, *A. gambiae:* PRJNA1438, *C. suppressalis*: PRJNA506136, *B. mori:* PRJNA205630, *T. castaneum*: PRJNA12540, *T. urticae:* PRJNA315122, *B. tabaci*: PRJNA312470, *T. vaporariorum*: PRJNA553773, *A. pisum*: PRJNA13657, *A. craccivora:* PRJNA558689, *O. laevigatus:* PRJNA721944, *C. lectularius:* PRJNA167477, *R. prolixus:* PRJNA13648, *A. mellifera*: PRJNA471592, *N. vitripennis*: PRJNA575073, *F. occidentalis*: PRJNA203209, *T. palmi*: PRJNA607431, *Z. nevadensis:* PRJNA203242.

4.4.7 Comparative genomics

ABC Transporters

ATP-binding cassette transporters (ABCs) are the largest known group of active transporters and are able to eliminate by translocation xenobiotic compounds such as secondary metabolites produced by plants or insecticides [91]. The ABC transporters are subdivided into eight subfamilies: ABCA-H, and of these, ABCB, ABCC and ABCG are the most associated with resistance to a variety of insecticides including pyrethroids, carbamates, organophosphates and neonicotinoids [92].

44 full-length ABC transporter genes were found for *M. brassicae* which was similar to other hymenopteran species (table 4). The distribution of genes across ABC transporter subfamilies was also very similar, with the majority of genes being present in the insecticide-resistance linked ABCC and ABCG subfamilies.

There is evidence of some lineage-specific gene expansion in the ABCC subfamily, with two high identity pairs of genes exhibiting recent tandem duplication (figure 2). Within the ABCG class, there were two clusters of genes which had a low percentage identity. As these genes had not been dispersed throughout the genome, this could indicate the conservation of these genes as heritable units may confer a selective advantage.

The distribution of ABC transporter genes across subfamilies was similar for Hymenoptera species compared to the crop pest species, although there is evidence of more significant gene expansion in the resistance-associated ABCC and ABCG subfamilies in some crop pests - *D. noxia*, *F. occidentalis* and *A. gossypii*. This could potentially confer an increased degree of insecticide resistance/tolerance in these pests compared to *M. brassicae*.

Table 4. Numbers of ABC transporter genes annotated in *Microctonus brassicae* and *Microplitis demolitor* (this study), *Aphidius gifuensis* [89], *Cotesia congregata, Cotesia rubecula, Cotesia glomerata, Cotesia vestalis, Cotesia flavipes, Cotesia sesamiae, Nasonia vitripennis, Apis mellifera* [48], *Drosophila melanogaster* [92], *Frankliniella occidentalis* [93], *Thrips palmi* [94], *Aphis gossypii* [95], *Trialeurodes vaporariorum* [96] *Diuraphis noxia* and *Bemisia tabaci* [97] and their distribution across subfamilies.

	Species	ABCA	ABCB	ABCC	ABCD	ABCE	ABCF	ABCG	ABCH	Total
M. brassicae & other Hymenoptera	*Microctonus brassicae*	4	6	13	2	1	3	15	0	**44**
	Microplitis demolitor	4	4	13	2	1	3	18	0	**45**
	Aphidius gifuensis	9	5	13	3	1	4	16	0	**50**
	Cotesia congregata	4	4	16	2	1	3	13	3	**46**
	Cotesia rubecula	4	4	17	2	2	4	13	3	**49**
	Cotesia glomerata	4	3	17	2	1	3	14	3	**47**
	Cotesia vestalis	4	3	14	2	2	4	14	3	**46**
	Cotesia flavipes	4	3	16	2	1	3	12	3	**44**
	Cotesia sesamiae	4	3	16	2	0	3	13	3	**44**
	Nasonia vitripennis	10	8	20	1	1	3	11	1	**55**
	Apis mellifera	3	5	9	2	1	3	15	3	**41**
	Drosophila melanogaster	10	8	14	2	1	3	15	3	**56**
Hemiptera crop pests	*Frankliniella occidentalis*	3	5	19	2	1	3	22	13	**70**
	Thrips palmi	3	4	12	2	2	3	16	7	**49**
	Diuraphis noxia	3	6	24	3	1	3	26	11	**77**
	Aphis gossypii	4	5	25	2	1	4	30	0	**71**
	Trialeurodes vaporariorum	3	9	7	4	1	3	9	9	**45**
	Bemisia tabaci	8	3	6	2	1	3	23	9	**55**

The ABCH subfamily is unique to arthropods and zebrafish [92], and had been thought to be present in all arthropod species. *Aphis gossypii* appears to be one of the first species identified without any ABCH members [95], and this was speculated to be the result of poor assembly quality. However, as more arthropod species are sequenced and annotated, it has become increasingly likely that this subfamily is not always present. In this study, no members of the ABCH subfamily were found in *M. brassicae or M. demolitor*, and no members were found in the

parasitoid wasp *Aphidius gifuensis* [89]. Lineage-specific losses of the ABCH subfamily may therefore be common across other arthropod species.

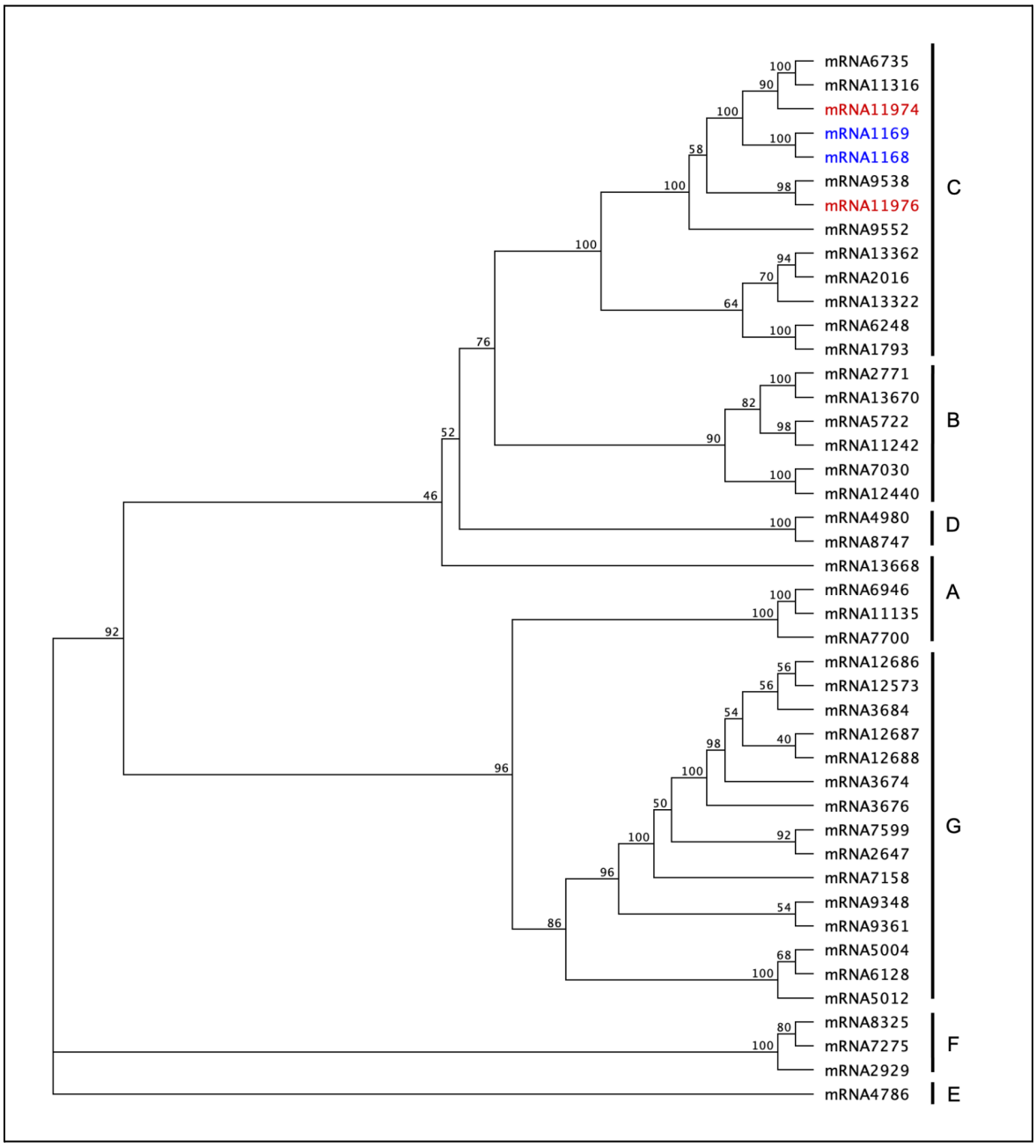


Figure 2. Phylogenetic tree of the *Microctonus brassicae* ABC transporters. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on >70% Similarity (using Blosum45 with threshold 0) and proximity of genes within the assembly.

Lineage specific expansion of ABCH genes has occurred in several crop pest species. The ABCH subfamily is not currently as strongly associated with pesticide resistance as ABCC and ABCG, however, upregulation of ABCH genes has been observed in *M. persicae* in response to carbamates [98] as well as for multi-pesticide resistant versus susceptible strains of *T. urticae* [99,100] and *P. xylostella* [101]. Therefore, the lack of this subfamily could potentially result in reduced pesticide resistance capabilities in *M. brassicae* compared to certain crop pests.

Cytochrome P450s monooxygenases

Cytochrome P450 monooxygenase (P450s) are a diverse superfamily capable of metabolizing a huge variety of endogenous and exogenous substrates. In insects they are associated with growth and development, metabolism of pesticides and plant toxins as well as the production and metabolism of insect hormones and pheromones [102,103]. P450s are associated with the resistance to insecticides from a variety of classes, including pyrethroids, carbamates and neonicotinoids. Many examples of resistance are linked to upregulated P450s [104–107]. They are also linked to the activation of organophosphates and other proinsecticides [108] often as a result of downregulation [109,110].

A total of 64 full-length P450 genes were identified in the M. brassicae genome (table 5). These genes were named by Dr David Nelson using his in-house pipeline [87].

46 of the P450 genes had 55-97% identity to another Hymenoptera p450, 16 had 40-55% identity and 2 genes had <40% identity. CYP343A1 - a member of the CYP2 clan - was the only gene classified as an ortholog to another species (*Cotesia*). The lack of orthologs for other genes is likely a result of P450 variation during evolution based on the organism's environment and adaptation. The CYPome diversity value for *M. brassicae* was 52%, based on the presence of

33 CYP subfamilies and 64 genes. The CYPome follows the pattern of other arthropods, with most CYP families having few genes, whilst only a few CYP families have many genes [111].

The total number of P450s for insect species ranges widely from 28 for *Meteorus pulchricornis* to 171 for *Bemisia tabaci*. This is mostly due to variation seen in the CYP3 and CYP4 clans, which tend to show a large degree of variation amongst insect species, as a result of their role in detoxification, whereas the CYP2 and mito clan exhibit much less variation between species, likely due to their roles in essential physiological functions [112].

Compared to other Hymenoptera species, the total number of *M. brassicae* P450s was fairly average. However, the total number was lower than all Hemiptera crop pests. Overall, the crop pests tended to have higher numbers of P450s than Hymenoptera species - most notably in the CYP3 and CYP4 clans. Distribution of *M. brassicae* P450 genes across clans was similar to other Hymenoptera species, with the majority of *M. brassicae* P450s belonging to the CYP3 (29) and CYP4 (21) clans.

The CYP3 clan is the most associated with insecticide resistance, notably the CYP6 and CYP9 subfamilies [103], both of which were present in *M. brassicae*. Recent expansion was seen in several clan 3 subfamilies (figure 3): CYP6SP7-9 were located within a 7kb region with identity ranging from 69-92%; CYP9HH1-3 were located within an 8kb region with identity ranging from 80-83%; CYP6AS204-8 were located within an 11kb region with identity ranging from 46-72% and CYP6AQ72-6 were located within an 11kb region with identity ranged from 67-91%. These clusters of adjacent genes with high identity indicate expansion likely occurred by tandem duplication. Expansions of the CYP6AS subfamily were also seen in *Apis mellifera [116]*, and members of this subfamily are reported to be upregulated in response to honey/pollen [117] and pyrethroids [118]. The CYP6AS subfamily expansion in *M. brassicae* could therefore be associated with both dietary metabolism and potential increased resistance/tolerance to pyrethroids.

Table 5. Total numbers of Cytochrome P450 genes annotated in *Microctonus brassicae* (this study), *Microplitis demolitor, Macrocentrus cingulum, Fopius arisanus, Diachasma alloeum, Trichomalopsis sarcophagae, Orussus abietinus* [113]**, *Aphidius gifuensis*** [89]**, *Cotesia congregata, Cotesia rubecula, Cotesia vestalis, Cotesia flavipes, Cotesia sesamiae, Nasonia vitripennis, Apis mellifera, Drosophila melanogaster*** [48]**, *Meteorus pulchricornis*** [114]**, *Pteromalus puparum, Trichogramma pretiosum, Copidosoma floridanum*** [51]**, *Frankliniella occidentalis, Thrips palmi*** [93]**, *Myzus persicae, Acyrthosiphon pisum*** [115]**, *Trialeurodes vaporariorum*** [96] **and *Bemisia tabaci*** [93]**.**

	Parasitoid Wasp Family	Species	CYP2	CYP3	CYP4	Mito	Total
M. brassicae & other Hymenoptera	Braconidae	*Microctonus brassicae*	7	29	21	7	**64**
	Braconidae	*Microplitis demolitor*	7	21	12	7	**47**
	Braconidae	*Aphidius gifuensis*	11	24	16	8	**59**
	Braconidae	*Cotesia congregata*	10	36	18	6	**70**
	Braconidae	*Cotesia flavipes*	9	25	9	6	**49**
	Braconidae	*Cotesia rubecula*	9	33	15	6	**63**
	Braconidae	*Cotesia sesamiae*	9	24	10	6	**49**
	Braconidae	*Cotesia vestalis*	9	31	14	6	**60**
	Braconidae	*Diachasma alloeum*	9	27	28	7	**71**
	Braconidae	*Fopius arisanus*	9	20	19	6	**54**
	Braconidae	*Macrocentrus cingulum*	11	15	10	5	**41**
	Braconidae	*Meteorus pulchricornis*	2	13	7	6	**28**
	Encyrtidae	*Copidosoma floridanum*	5	43	23	5	**76**
	Orussidae	*Orussus abietinus*	7	13	9	8	**37**
	Pteromalidae	*Nasonia vitripennis*	7	48	30	7	**92**
	Pteromalidae	*Pteromalus puparum*	8	64	38	7	**171**
	Pteromalidae	*Trichomalopsis sarcophagae*	7	40	21	7	**75**
	Trichogrammatidae	*Trichogramma pretiosum*	5	29	21	9	**64**
		Apis mellifera	8	28	4	6	**46**
		Drosophila melanogaster	7	36	32	11	**86**
Hemiptera crop pests		*Frankliniella occidentalis*	12	22	37	10	**81**
		Thrips palmi	12	26	42	11	**91**
		Myzus persicae	3	63	48	1	**115**
		Acyrthosiphon pisum	10	33	32	8	**83**
		Trialeurodes vaporariorum	7	41	25	7	**80**
		Bemisia tabaci	18	76	73	4	**171**

Partial and fragment p450 genes were excluded in cases where they were listed as such - some may remain in the counts if official naming and curation had not taken place.

In *M. pulchricornis*, CYP6SP6 expression was significantly increased after exposure to a variety of insecticides, including a 4-fold increase in response to pyrethroid (cypermethrin) and a 25.7-fold increase in response to a pyrrole class pro-insecticide (chlorfenapyr). The *MpulCYP6SP6* and *MpulCYP6SQ-1* genes were also speculated to contribute to neonicotinoid resistance [114]. *M. brassicae* had a gene cluster which included: CYP6SP7-9 and CYP6SQ22 (figure 3). CYP6SP9 had 79% identity to *MpulCYP6SP6*. It is therefore possible that these genes could confer potential increased resistance to a variety of insecticides.

The other CYP3 clan subfamilies which exhibited expansion have not been studied in response to xenobiotic compounds, however, as they are also members of the detoxification-associated CYP6 and CYP9 subfamilies it is also possible that these expansions could confer potential increased insecticide tolerance.

The CYP4 clan is vastly expanded in many arthropods [119], and whilst the CYP4 clan is not as strongly associated with insecticide resistance as CYP3, studies have shown upregulation of some CYP4 genes in response to insecticide exposure [105,120–122].

Recent expansion was seen in clan 4 subfamilies: CYP4CA16-20 were located within a 17kb region with identity ranging from 60-84% and CYP4AB81-85 were located within a 21kb region with identity ranging from 61-86%.

The largest P450 subfamily expansion was within the CYP4249 family with 7 members: CYP4249B1-5 and CYP4249C1-2. B1-5 were located contiguously within a 0.16Mb region. C1-2 were located within a 5kb region. Percentage identity ranged from 40-83%. A study looking at P450s in *M. pulchricornis* identified that CYP4249-1 was upregulated in response to a variety of insecticides and showed the highest upregulation of all tested P450 genes (4.7-fold) in response to organophosphate (phoxim) and the second-highest (5.7-fold) in response to pyrethroid (cypermethrin) [114]. The CYP4249 subfamily exhibited large expansion in *M.*

brassicae, with 7 members. It is therefore possible that this expansion could confer potential increased resistance to a variety of insecticides.

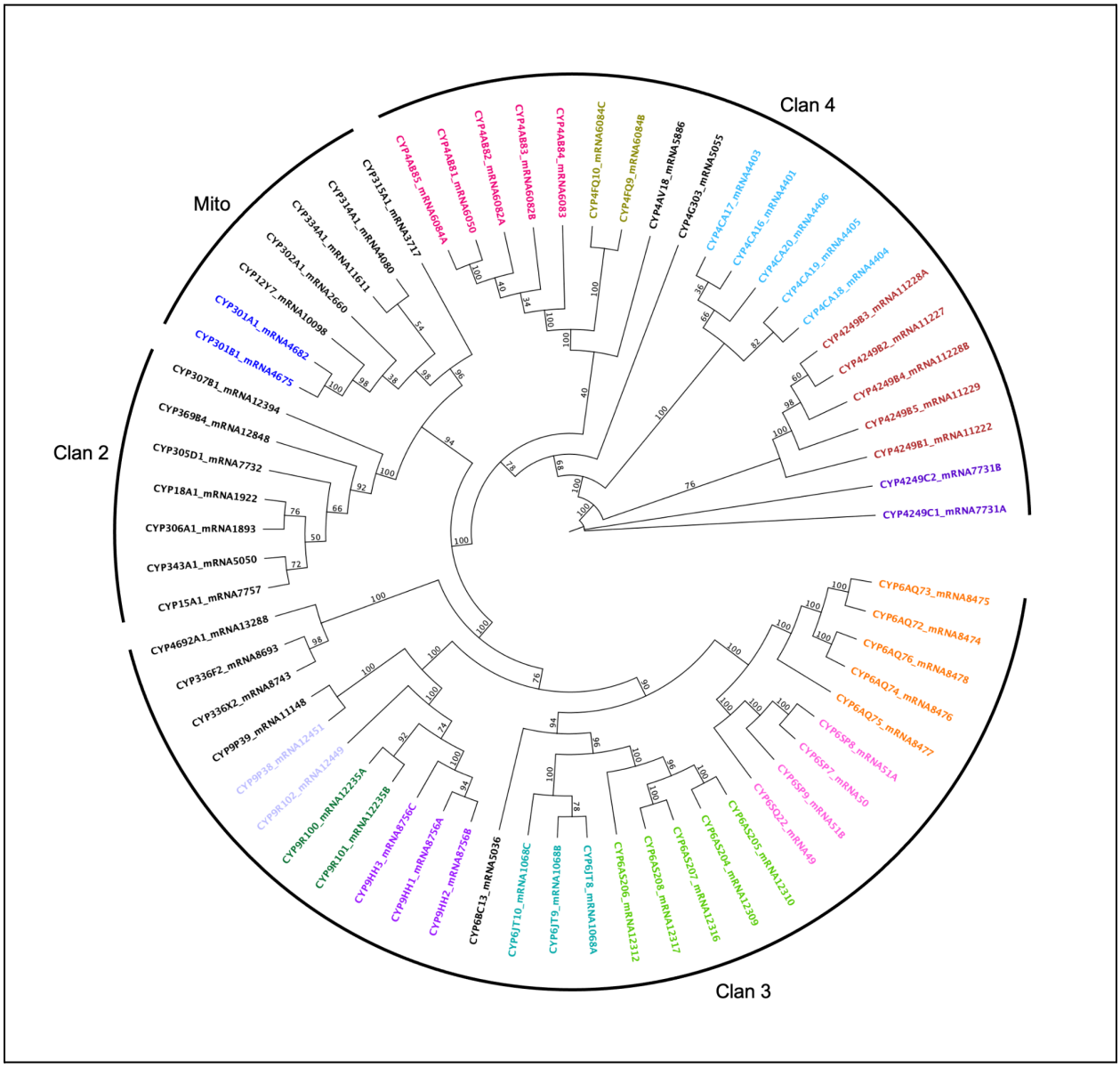


Figure 3. Phylogenetic tree of the *Microctonus brassicae* Cytochrome P450s. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on >70% Similarity (using Blosum45 with threshold 0) and proximity of genes within the assembly.

These clusters of high identity genes indicate recent tandem duplication was likely responsible for gene expansion within the P450 genes. However, despite these expansions, total numbers of CYP3 and CYP4 P450s were generally higher in hemipteran crop pests than in *M. brassicae*, which may suggest a greater degree of P450-mediated insecticide resistance in crop pests.

Carboxyl/choline esterases

Carboxyl/choline esterases (CCEs) are associated with insecticide resistance, notably to organophosphates, and to a lesser degree carbamates and pyrethroids [123]. For example, amplification of genes for esterases are associated with resistance to a variety of insecticides in *Myzus persicae* via degradation and sequestration [124–126]. Elevated esterases are proposed to confer organophosphate and carbamate resistance in *Nilaparvata lugens* and other planthopper species [127,128]. Esterase-based organophosphate resistance has been reported in three *Culex* species [129] and synergist bioassays have shown that esterases are responsible for metabolic resistance to pyrethroids (deltamethrin) and organophosphates (temephos) in *Aedes aegypti* [130].

M. brassicae has 26 full-length carboxylesterase genes (table 6). The total number of CCEs for *M. brassicae* and distribution across the 3 main classes is comparable to other Hymenoptera species (albeit at the lower end of the range). The numbers and distribution of CCEs is also similar between Hymenoptera and Hemiptera crop pests, with the most notable difference being a higher average number of 'dietary' esterases in hymenoptera.

Table 6. Numbers of CCE genes annotated in *Microctonus brassicae* and *Microplitis demolitor* (this study), *Cotesia congregata, Nasonia vitripennis, Apis mellifera, Drosophila melanogaster* *[48]*, *Frankliniella occidentalis* *[93]*, *Myzus persicae* [115], *Acyrthosiphon pisum, Bemisia tabaci* *[131]* and *Trialeurodes vaporariorum* *[132]* and their distribution across classes and clades.

	Species	Dietary class	Hormone/ semiochemical processing class	Neurodevelopmental						Total
				Glutactins	AChE	uncharacterised	gliotactin	neuroligin	neurotactin	
M. brassicae & other Hymenoptera	*Microctonus brassicae*	13	4	0	2	0	0	6	1	**26**
	Microplitis demolitor	20	7	0	3	0	0	6	1	**37**
	Cotesia congregata	19	7	-	-	-	-	-	-	**26**
	Nasonia vitripennis	13	17	-	-	-	-	-	-	**30**
	Apis mellifera	8	4	-	-	-	-	-	-	**12**
	Drosophila melanogaster	13	8	-	-	-	-	-	-	**21**
Hemiptera crop pests	*Frankliniella occidentalis*	28	7	2	2	2	1	7	1	**50**
	Myzus persicae	5	12	0	3	1	1	0	0	**22**
	Acyrthosiphon pisum	5	16	0	2	1	1	3	0	**28**
	Trialeurodes vaporariorum	12	6	1	2	1	1	3	1	**27**
	Bemisia tabaci	6	19	1	4	1	1	10	0	**42**

The so-called 'dietary' class of CCEs has been shown to be involved in insecticide and xenobiotic detoxification [133] and amplification of genes within this class, i.e. esterase E4/B1-like genes, has been linked to resistance to a variety of insecticides in hemipteran and dipteran species (*M persicae, N. lugens, S. graminum* and *Culex* mosquitoes) *[126,129,134–138]*. Within the *M. brassicae* genome, there were two clusters of similar esterase E4/B1 genes (figure 4). One cluster of 6 genes with percent identity ranging from 51-66% identity, and another cluster of 3 genes with percent identity ranging from 43-59%. These clusters indicate tandem duplications, which could confer some tolerance/resistance to insecticides. In cases where the number of dietary genes in *M. brassicae* is higher than hemipteran crop pests, there could be a potential increased degree of resistance.

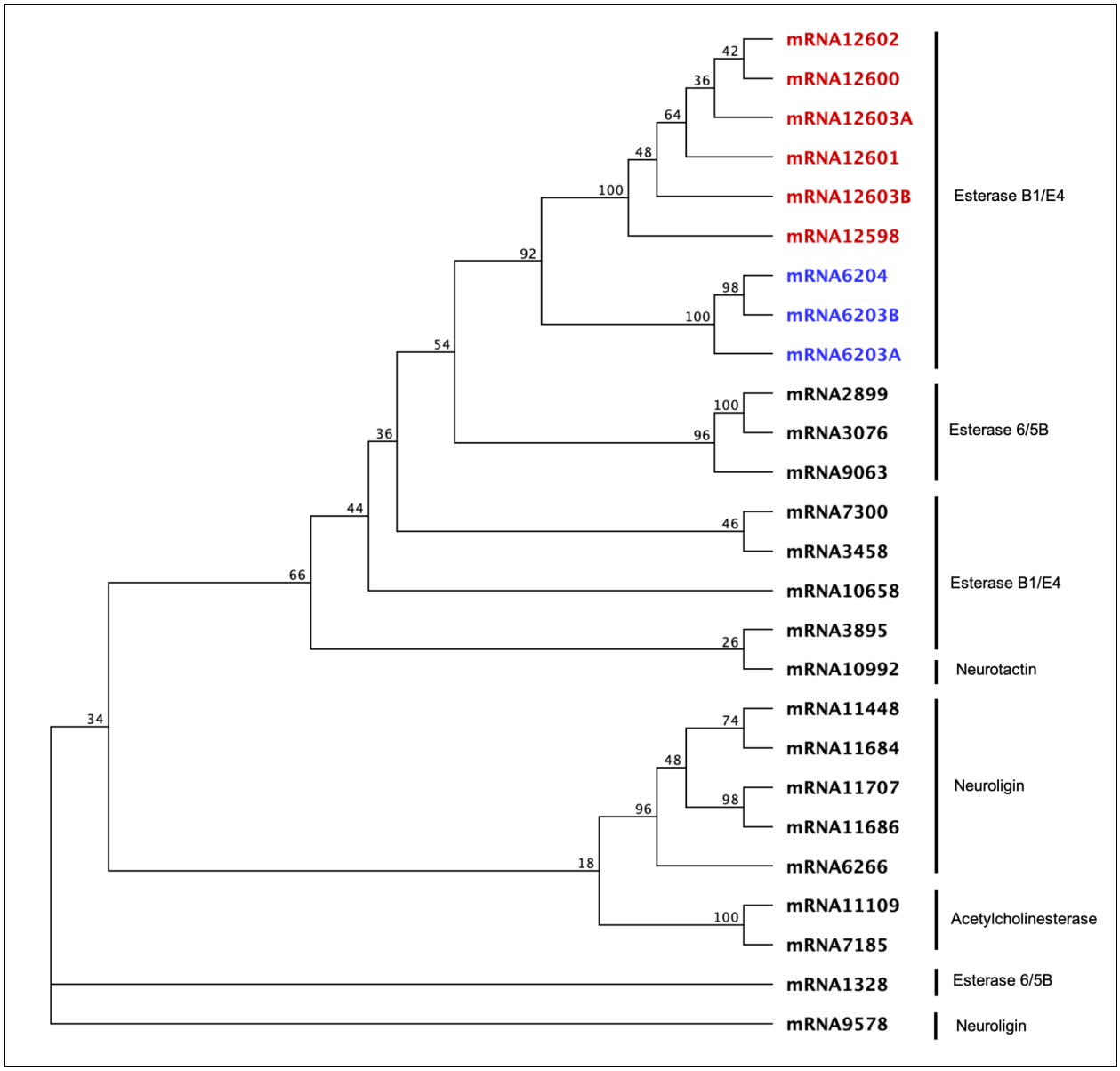


Figure 4. Phylogenetic tree of the *Microctonus brassicae* Carboxyl/cholinesterases. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on >70% Similarity (using Blosum45 with threshold 0) and proximity of genes within the assembly.

Glutathione-*S*-transferases

The glutathione-*S*-transferases (GSTs) family is large and functionally diverse, and has been shown to confer resistance to all main insecticide classes. GST-mediated detoxification of insecticides takes place via several different mechanisms, including protecting against oxidative stress, binding and sequestration of the insecticide and their metabolites and by catalysing the conjugation of glutathione to insecticides to reduce their toxicity and facilitate excretion respectively [57].

Table 7. Numbers of GST genes annotated in *Microctonus brassicae and Microplitis demolitor* (this study), *Meteorus pulchricornis* [139]**, *Cotesia congregata, Cotesia rubecula, Cotesia vestalis, Cotesia flavipes, Cotesia sesamiae, Nasonia vitripennis, Apis mellifera, Drosophila melanogaster*** [48]**, *Thrips palmi*** [94]**, *Myzus persicae, Acyrthosiphon pisum, Trialeurodes vaporariorum* and *Bemisia tabaci*** [140] **and their distribution across classes.**

	Species	Delta	Epsilon	Omega	Sigma	Theta	Zeta	Microsomal	Unclassified	Detoxification Total *
M. brassicae & other Hymenoptera	*Microctonus brassicae*	0	1	1	4	0	1	4	1	**8**
	Microplitis demolitor	3	2	1	4	0	1	-	1	**13**
	Meteorus pulchricornis	4	0	3	7	0	1	2	0	**15**
	Cotesia congregata	5	0	2	5	1	1	2	2	**16**
	Cotesia rubecula	7	0	2	5	1	1	16	2	**20**
	Cotesia vestalis	6	0	2	5	1	1	15	2	**17**
	Cotesia flavipes	6	0	2	5	1	1	15	2	**17**
	Cotesia sesamiae	6	0	2	5	1	1	15	2	**17**
	Nasonia vitripennis	5	0	2	8	3	1	0	0	**19**
	Apis mellifera	1	0	1	4	1	1	2	0	**8**
	Drosophila melanogaster	11	14	4	1	4	2	3	0	**36**
Hemiptera crop pests	*Thrips palmi*	14	0	1	6	1	2	1	-	**24**
	Myzus persicae	3	0	1	12	1	0	2	0	**17**
	Acyrthosiphon pisum	11	0	1	5	2	0	2	0	**19**
	Trialeurodes vaporariorum	9	1	0	3	0	2	3	0	**15**
	Bemisia tabaci	14	0	1	6	0	2	2	0	**23**

(*microsomal not included)

M. brassiccae has 7 cytosolic GSTs, 4 microsomal (non-detoxification) genes and 1 unclassified GST (table 7) which most closely matched glutathione S-transferase 1 (*Cephus cinctus*) with 79% identity when searching the non-redundant Insecta proteins database via NCBI BlastP. The total number of detoxifying GSTs was the same as *A. mellifera*, but less than that seen for other hymenopteran species. The distribution of GSTs across classes was similar to other Hymenoptera species, with the exception of an absence of delta class GSTs.

Members of the delta class in *N. lugens* [141] and epsilon class in *A. aegypti* [142] have been linked to pyrethroid resistance through reduction of lipid peroxidation and the epsilon class has also been associated with direct metabolism of pyrethroids [143]. Both classes are also able to metabolize DDT [144,145]. Despite all other hymenoptera species having at least 1 delta class GST, *M. brassicae* had no delta class GSTs. However, the epsilon class, which is generally missing in hymenopteran species, had 1 member. A lack of epsilon GSTs is common in hymenopteran species, and has been speculated to be due to parasitoids receiving protection via their host's detoxification capabilities [139]. It is therefore possible that the loss of delta class GSTs in *M. brassicae* is a result of the detoxification capabilities of *Psylliodes chrysocephala*. Indeed, in recent years *P. chrysocephala* has developed a high degree of pyrethroid resistance as a result of widespread pyrethroid use following the neonicotinoid seed treatment ban [32].

The theta class had no members, which was the same as for *M. demolitor* and *M. pulchricornis*, as well as for the hemipteran crop pests *T. vaporariorum* and *Bemisia tabaci*. The theta class is likely the origin of cytosolic GSTs [146] and was thought to be generally highly conserved amongst insects, with most insects having at least 1 theta GST [147], although it is now apparent that there are increasing numbers of insect species lacking the theta class. The theta class is not strongly associated with insecticide resistance, although it may play a role in protection against oxidative stress [148].

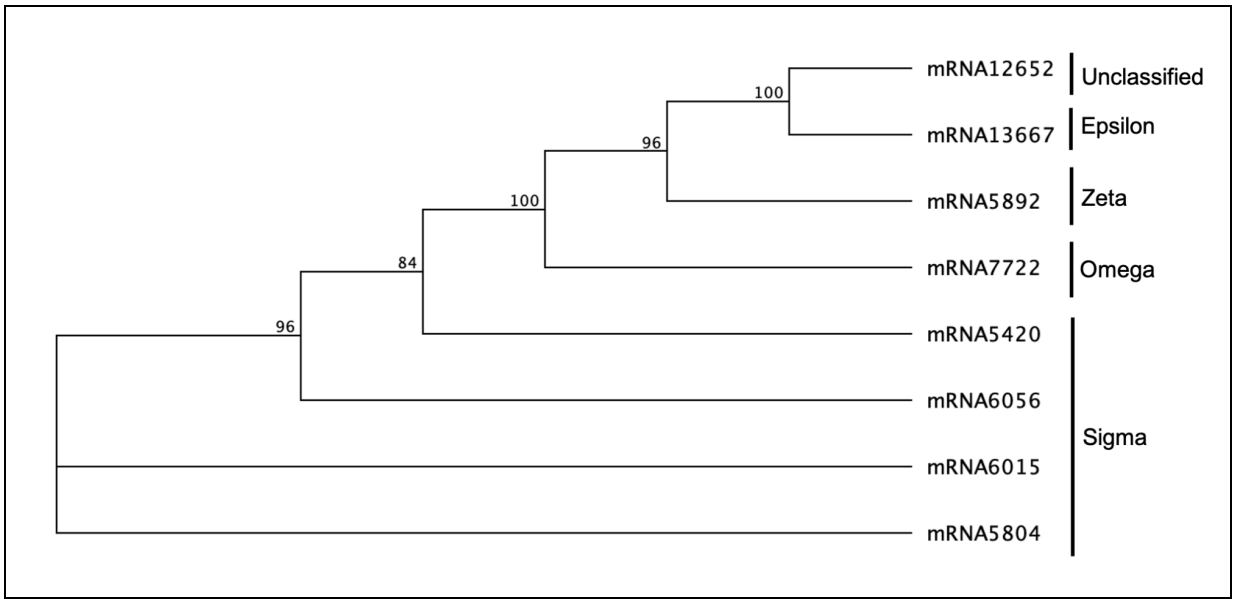


Figure 5. Phylogenetic tree of the *Microctonus brassicae* Glutathione S-Transferases. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates.

Numbers of omega, zeta, sigma and microsomal classes were similar to the numbers seen for other species. These classes are thought to be involved in various cellular processes, including defense against oxidative stress [148–150]. Organophosphates (phoxim) induced increased expression of GST omega 1 in *Bombyx mori* [151]. Pyrethroids and broad spectrum insecticides (chlorantraniliprole and beta-cypermethrin) induced significantly higher expression of omega and sigma GSTs in *Heortia vitessoides* [152]. Organophosphate and pyrethroid (phoxim and cypermethrin) exposure resulted in upregulated expression levels of sigma and microsomal GSTs in *M. pulchricornis* [139].

There is no evidence of any recent gene expansion within the GST gene family in *M. brassicae* (figure 5), suggesting that this may not be its most prevalent potential insecticide resistance mechanism. Compared to hemipteran crop pests, *M. brassicae* has a lower number of GSTs, notably delta class GSTs, and therefore may have reduced potential GST-mediated resistance.

UDP-glycosyltransferases

UDP-glycosyltransferases (UGTs) are phase II detoxification enzymes which are involved in insecticide resistance. The mechanisms of UGT-mediated resistance are for example based on the conjugation of P450-functionalized substrates. They have been linked to pyrethroid resistance in *M. pulchricornis* and *Anopheles gambiae* [153,154], carbamate resistance in *M. persicae* [155], diamide resistance in *Chilo suppressalis* [156] and neonicotinoid resistance in *Diaphorina citri* and *Aphis gossypii* [157,158] and they also contribute to insecticide detoxification via the elimination of oxidative stress in *Apis cerana* [159].

M. brassicae had 15 full-length UGT genes (table 8), distributed across 7 families: UGT334 (UGT334AA6-13), UGT338 (UGT338F2), UGT391 (UGT391B2), UGT403 (UGT403G2), UGT461 (UGT461A1-2), UGT462 (UGT462A1) and UGT50 (UGT50C8). The total number of genes is comparable to other hymenopteran species. However, total numbers in hymenopteran species are far lower than in hemipteran crop pests. This may be in part associated with differences in diet. Lineage specific expansions of UGTs have previously been linked to increased detoxification of plant secondary metabolites [160]. As hemipteran crop pests are plant-feeding insects, whilst hymenoptera feed on nectar, pollen and host insects, this would explain the increased number of UGTs in hemiptera. Host plant adaptation alone has been shown to usually be insufficient to confer insecticide resistance, and therefore higher numbers overall of UGTs in Hemiptera cannot be assumed to correspond to increased insecticide tolerance/resistance [160].

Table 8. Numbers of annotated UDP glycosyltransferase genes found in *Microctonus brassicae and Microplitis demolitor* (this study), *Meteorus pulchricornis* [151], *Pteromalus puparum, Trichogramma pretiosum, Copidosoma floridanum* [51], *Cotesia congregata, Cotesia rubecula, Cotesia vestalis, Cotesia flavipes, Cotesia sesamiae, Nasonia vitripennis, Apis mellifera, Drosophila melanogaster* [48], *Tetranychus urticae, Nilaparvata lugens, Acyrthosiphon pisum, Bemisia tabaci* [161], *Myzus persicae* [162], *Trialeurodes vaporariorum* [163] and *Thrips palmi* [94].

	Species	Total
M. brassicae & other Hymenoptera	*Microctonus brassicae*	**15**
	Microplitis demolitor	**7**
	Meteorus pulchricornis	**10**
	Pteromalus puparum	**27**
	Trichogramma pretiosum	**18**
	Copidosoma floridanum	**19**
	Cotesia congregata	**11**
	Cotesia rubecula	**10**
	Cotesia vestalis	**10**
	Cotesia flavipes	**10**
	Cotesia sesamiae	**10**
	Nasonia vitripennis	**22**
	Apis mellifera	**12**
	Drosophila melanogaster	**35**
Hemiptera crop pests	*Tetranychus urticae*	**81**
	Nilaparvata lugens	**20**
	Thrips palmi	**17**
	Myzus persicae	**101**
	Acyrthosiphon pisum	**72**
	Trialeurodes vaporariorum	**55**
	Bemisia tabaci	**76**

10 of the genes (UGT334AA6-13, UGT461A2 and UGT462A1) were located adjacently within a single cluster within a 50kb region. Percent identity within this cluster ranged from 37% to 87%. This suggests a high degree of tandem duplication is likely responsible for UGT gene expansion. The wide range of percentage identity indicates that these are a mix of 'old' and

'new' duplications. The fact that these genes have not been dispersed throughout the genome over time (with the exception of UGT461A1, which is located on a separate scaffold but has 65% identity to UGT461A2) suggests that there may be a selective advantage for maintaining this gene cluster as a heritable unit.

The UGTs present in *M. brassicae* were very similar to those found in *M. pulchricornis* (Hymenoptera: Braconidae). The largest family in both species was the UGT334 (UGT334AA) family, containing 5 members in *M. pulchricornis* and 8 members in *M. brassicae*. Both species also each had a single member of the UGT338 (UGT338F), UGT391 (UGT391B), UGT403 (UGT403G) and UGT50 (UGT50C) families [154].

Several of the UGTs shared by both species were upregulated in response to insecticides in *M. pulchricornis*: UGT334A1 was upregulated in response to a pyrethroid insecticide (cypermethrin); UGT334A1 and UGT403G1 were upregulated in response to a broad spectrum insecticide (chlorfenapyr) and UGT338F1, UGT403G1 and UGT50C7 were all highly upregulated in response to an organophosphate (phoxim). Knockdown of UGT403G1 and UGT334AA1 significantly increased the mortality of *M. pulchricornis* in response to phoxim and chlorfenapyr respectively [154].

The presence of these insecticide-associated UGT family genes in *M. brassicae* may confer some resistance/tolerance to a variety of insecticides, particularly the significantly expanded UGT334 family. Members of the *M. pulchricornis* UGT334AA subfamily were both upregulated and downregulated to varying degrees in response to several insecticides [154], which suggests there may be a variety of resistance-related functions within UGT334AA subfamily members.

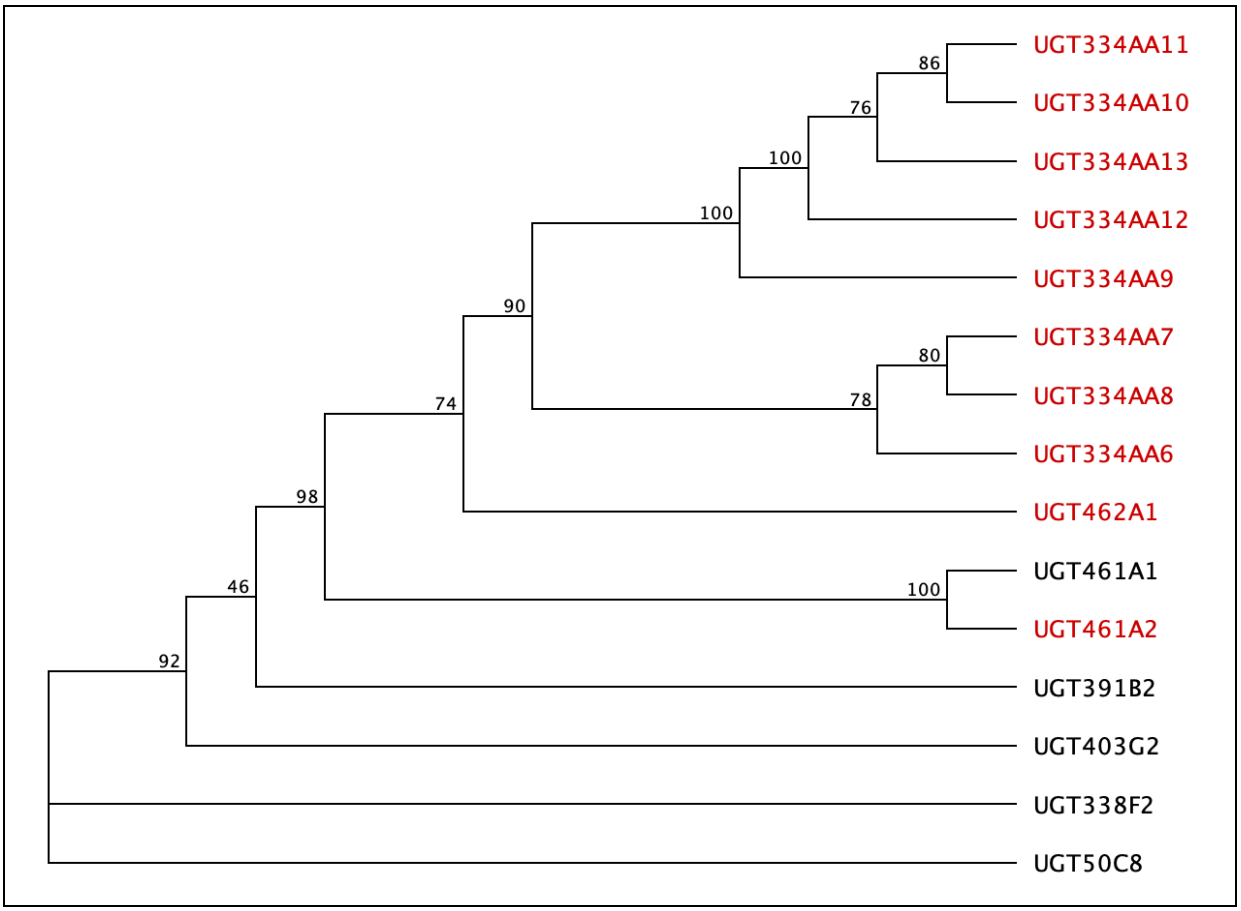


Figure 6. Phylogenetic tree of the *Microctonus brassicae* UDP-glucosyltransferases. Amino acid sequences were aligned using MAFFT and analyzed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. Coloured nodes indicate groups of likely recent tandem duplications, based on >70% Similarity (using Blosum45 with threshold 0) and proximity of genes within the assembly.

Point Mutations

Known pyrethroid resistance-associated point mutations in the voltage-gated sodium channel were searched for in *M. brassicae*. These included the L1014F point mutation which is widespread in *Psylliodes chrysocephala* populations across Europe and the L925I mutation which is less common, but still present in some UK populations [29,32,164,165]. As it is likely

that both *P. chrysocephala* and *M. brassicae* would be under very similar selection pressures, it was thought that at least one of these target-site mutations had a high likelihood of being present in *M. brassicae*. However, neither were present (although this does not exclude the possibility that the mutation could be present within other members of the population).

4.5 Conclusions

PacBio HiFi long read technology enabled the production of a high quality genome for *M. brassicae*. The genome completeness is of excellent quality for comparative and functional genomics analyses and provides a useful first reference for the *Microctonus* genus.

Comparative analyses of *M. brassicae* with Hemiptera crop pests showed evidence that *M. brassicae* has a detoxification gene inventory comparable to these selected crop pests. However, lineage-specific expansions within detoxification gene families such as UGTs, P450s and CCEs were seen, whereas the GST and ABC transporter gene families lack such expansions compared to some crop pests. No mutations were found in common insecticide target-sites, suggesting a lack of selectivity of insecticides at the protein/receptor binding level.

An interesting next step would be to compare the numbers of detoxification genes of *M. brassicae* with its host *P. chrysocephala*. Host-parasitoid relationships can influence the numbers of resistance genes found in the parasitoid. A higher level of insecticide resistance in the host has been suggested to increase the level of resistance seen in its corresponding parasitoid, with resistance genes potentially being selected for during larval development [166]. PBO inhibition of P450s in *P. chrysocephala* suppressed lambda-cyhalothrin resistance, suggesting that P450s could be its most prevalent resistance mechanism [32]. In *M. brassicae*, the P450 gene family showed the most lineage specific expansion compared with its other detoxification

gene families. It is therefore possible that this expansion could have potentially been promoted by its host. (Although, similar exposure to insecticides could also have easily played a role in this selection process.) On the other hand, host protection has also been speculated to decrease the level of certain resistance mechanisms, such as reduced epsilon GST numbers in hymenopteran species [139], and so perhaps host protection by *P. chrysocephala* could be responsible for the low number of GSTs in *M. brassicae*.

This study provides a good basis for beginning to identify differences in genes encoding potential tolerance/resistance mechanisms between crop pests and *M. brassicae* which could be exploited when selecting targeted insecticides for use in IPM strategies. Evidence of gene expansions in resistance-associated gene families implies that *M. brassicae* is certainly capable of developing resistance to a variety of insecticides, which could be used to our advantage through the selective breeding and selection of resistant strains of *M. brassicae* for use in IPM.

4.6 Declarations

Availability of data and materials

The genome assembly and annotation generated in this study is available in the GenBank repository, under bioproject: PRJEB48227. The raw PacBio HiFi reads are available from SRA under accession number: ERS8071542.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Pest Genomics Initiative, a collaborative project between Rothamsted Research, Bayer Crop Science and Syngenta AG. Rothamsted Research gratefully acknowledges strategic funding from the UK Biotechnology and Biological and Sciences Research Council (UKRI-BBSRC). The content is solely the responsibility of the authors and does not necessarily reflect the official views of the funding bodies.

Authors' contributions

The Pest Genomics Initiative (BB, CR, EG, KH-P, LF, RK and RN) devised the original conceptual ideas. EB assembled and annotated the genome with guidance from RK and DH. EB performed the comparative analyses. FM, RK and KH-P supervised the project. EB wrote the manuscript. All authors read, edited and approved the final manuscript.

Acknowledgements

Acknowledgements go to Dr. David R. Nelson of the International Committee on the Nomenclature for Cytochrome P450 Enzymes for classification of *Microctonus brassicae* P450s and to Dr. Michael H. Court of the UGT Nomenclature Committee for classification of *Microctonus brassicae* UGTs.

4.7 Literature Cited

[1] Oerke E-C. Crop losses to pests. J Agric Sci 2006;144:31–43.
[2] Culliney TW. Crop Losses to Arthropods. In: Pimentel D, Peshin R, editors. Integrated Pest Management: Pesticide Problems, Vol.3, Dordrecht: Springer Netherlands; 2014, p. 201–25.
[3] Deutsch CA, Tewksbury JJ, Tigchelaar M, Battisti DS, Merrill SC, Huey RB, et al. Increase in crop losses to insect pests in a warming climate. Science 2018;361:916–9.
[4] Barzman M, Bàrberi P, Birch ANE, Boonekamp P, Dachbrodt-Saaydeh S, Graf B, et al. Eight principles of integrated pest management. Agron Sustain Dev 2015;35:1199–215.
[5] Stenberg JA. A Conceptual Framework for Integrated Pest Management. Trends Plant Sci 2017;22:759–69.
[6] Prokopy R, Kogan M. Chapter 139 - Integrated Pest Management. In: Resh VH, Cardé RT, editors. Encyclopedia of Insects (Second Edition), San Diego: Academic Press; 2009, p. 523–8.
[7] Ehler LE. Integrated pest management (IPM): definition, historical development and implementation, and the other IPM. Pest Manag Sci 2006;62:787–9.
[8] Quicke DLJ, van Achterberg C. Phylogeny of the subfamilies of the family Braconidae (Hymenoptera: Ichneumonoidea) n.d. https://repository.naturalis.nl/document/148935 (accessed September 23, 2021).
[9] Wharton RA. Bionomics of the Braconidae. Annu Rev Entomol 1993;38:121–43.
[10] Wang Z-Z, Liu Y-Q, Shi M, Huang J-H, Chen X-X. Parasitoid wasps as effective biological control agents. J Integr Agric 2019;18:705–15.
[11] Ba MN, Baoua IB, N'Diaye M, Dabire-Binso C, Sanon A, Tamò M. Biological control of the millet head miner Heliocheilus albipunctella in the Sahelian region by augmentative releases of the parasitoid wasp Habrobracon hebetor: effectiveness and farmers' perceptions. Phytoparasitica 2013;41:569–76.
[12] Flanders SE, Others. Mass Production of Egg Parasites oí the Genus Trichogramma. Hilgardia 1930;4.
[13] Orr DB. Scelionid Wasps as Biological Control Agents: A Review. Fla Entomol 1988;71:506–28.
[14] Bale JS, van Lenteren JC, Bigler F. Biological control and sustainable food production. Philos Trans R Soc Lond B Biol Sci 2008;363:761–76.
[15] Herren HR, Neuenschwander P. Biological Control of Cassava Pests in Africa. Annu Rev Entomol 1991;36:257–83.
[16] Kean JM, Barlow ND. Long-term Assessment of the Biological Control of Sitona discoideus by Microctonus aethiopoides and Test of a Model. Biocontrol Sci Technol 2000;10:215–21.
[17] Pennacchio F, Strand MR. Evolution of developmental strategies in parasitic hymenoptera. Annu Rev Entomol 2006;51:233–58.
[18] Desneux N, Blahnik R, Delebecque CJ, Heimpel GE. Host phylogeny and specialisation in parasitoids. Ecol Lett 2012;15:453–60.
[19] Canale A, Benelli G. Impact of mass-rearing on the host seeking behaviour and parasitism by the fruit fly parasitoid Psyttalia concolor (Szépligeti) (Hymenoptera: Braconidae). J Pest Sci 2012;85:65–74.
[20] Pizzol J, Desneux N, Wajnberg E, Thiéry D. Parasitoid and host egg ages have independent impact on various biological traits in a Trichogramma species. Journal of Pest Science 2012;85:489–96. https://doi.org/10.1007/s10340-012-0434-1.
[21] Zappalà L, Campolo O, Grande SB, Saraceno F, Biondi A, Siscaro G, et al. Dispersal of Aphytis melinus (Hymenoptera: Aphelinidae) after augmentative releases in citrus orchards. Eur J Entomol 2012;109:561–8.
[22] Hufbauer RA, Roderick GK. Microevolution in biological control: Mechanisms, patterns, and processes. Biol Control 2005;35:227–39.
[23] Jordan A, Broad GR, Stigenberg J, Hughes J, Stone J, Bedford I, et al. The potential of the solitary parasitoid Microctonus brassicae for the biological control of the adult cabbage stem flea beetle, Psylliodes chrysocephala. Entomol Exp Appl 2020;168:360–70.
[24] Graham CW, Alford DV. The distribution and importance of cabbage stem flea beetle (Psylliodes chrysocephala (L.)) on winter oilseed rape in England. Plant Pathol 1981;30:141–5.
[25] Broman B. Diversities in oilseed rape growing within the Western Palearctic Regional Section [Brassica napus var. oleifera]. Bulletin SROP (France) 1990.
[26] Winfield AL, Others. Management of oilseed rape pests in Europe. Agricultural Zoology Reviews 1992;5:51–95.
[27] Alford DV, Nilsson C, Ulber B. Insect pests of oilseed rape crops. Biocontrol of Oilseed Rape Pests 2003;1.
[28] Williams IH. The Major Insect Pests of Oilseed Rape in Europe and Their Management: An Overview. In: Williams IH, editor. Biocontrol-Based Integrated Management of Oilseed Rape Pests, Dordrecht: Springer Netherlands; 2010, p. 1–43.
[29] Højland DH, Nauen R, Foster SP, Williamson MS, Kristensen M. Incidence, Spread and Mechanisms of Pyrethroid Resistance in European Populations of the Cabbage Stem Flea Beetle, Psylliodes chrysocephala L. (Coleoptera: Chrysomelidae). PLoS One 2015;10:e0146045.
[30] Robert C, Ruck L, Carpezat J, Lauvernay A, Siegwart M, Leflon M, et al. Pyrethroids resistance monitoring in French cabbage stem flea beetle (Psylliodes chrysocephala) and rape winter stem weevil (Ceutorhynchus picitarsis) populations in oilseed rape. Écologie Chimique: nouvelles contributions à la protection des cultures contre les ravageurs et 11e Conférence Internationale sur les Ravageurs et Auxiliaires en Agriculture, 24 au 26 octobre 2017, Montpellier, France, Association Française de Protection des Plantes (AFPP); 2017, p. 196–208.
[31] Stará J, Kocourek F. Cabbage stem flea beetle's (Psylliodes chrysocephala L.) susceptibility to pyrethroids and tolerance to thiacloprid

in the Czech Republic. PLoS One 2019;14:e0214702.
[32] Willis CE, Foster SP, Zimmer CT, Elias J, Chang X, Field LM, et al. Investigating the status of pyrethroid resistance in UK populations of the cabbage stem flea beetle (Psylliodes chrysocephala). Crop Prot 2020;138:105316.
[33] Wynn S, Ellis S, Alves L. Cabbage stem flea beetle snapshot assessment--incidence and severity at end September 2014. HGCA Project Report 2014;546.
[34] Dewar AM. The adverse impact of the neonicotinoid seed treatment ban on crop protection in oilseed rape in the United Kingdom. Pest Manag Sci 2017;73:1305–9.
[35] Kathage J, Castañera P, Alonso-Prados JL, Gómez-Barbero M, Rodríguez-Cerezo E. The impact of restrictions on neonicotinoid and fipronil insecticides on pest management in maize, oilseed rape and sunflower in eight European Union regions. Pest Manag Sci 2018;74:88–99.
[36] Desneux N, Pham-Delègue M-H, Kaiser L. Effects of sub-lethal and lethal doses of lambda-cyhalothrin on oviposition experience and host-searching behaviour of a parasitic wasp, Aphidius ervi. Pest Manag Sci 2004;60:381–9.
[37] Desneux N, Rafalimanana H, Kaiser L. Dose–response relationship in lethal and behavioural effects of different insecticides on the parasitic wasp Aphidius ervi. Chemosphere 2004;54:619–27.
[38] Desneux N, Denoyelle R, Kaiser L. A multi-step bioassay to assess the effect of the deltamethrin on the parasitic wasp Aphidius ervi. Chemosphere 2006;65:1697–706.
[39] Biondi A, Zappalà L, Stark JD, Desneux N. Do biopesticides affect the demographic traits of a parasitoid wasp and its biocontrol services through sublethal effects? PLoS One 2013;8:e76548.
[40] Haseeb M, Amano H. Effects of contact, oral and persistent toxicity of selected pesticides on Cotesia plutellae (Hym., Braconidae), a potential parasitoid of Plutella xylostella (Lep., Plutellidae). J Appl Entomol 2002;126:8–13.
[41] Shimoda T, Yara K, Kawazu K. The effects of eight insecticides on the foraging behavior of the parasitoid wasp Cotesia vestalis. J Plant Interact 2011;6:189–90.
[42] Xu R, Mortimer PE, Kuang RP, He J, Zhang WD, Yin F. Sublethal impact of paraquat on the life span and parasitic behavior of Diaeretiella rapae M'Intosh. J Environ Sci Health B 2013;48:651–7.
[43] Saber M, Hejazi MJ, Kamali K, Moharramipour S. Lethal and sublethal effects of fenitrothion and deltamethrin residues on the egg parasitoid Trissolcus grandis (Hymenoptera: Scelionidae). J Econ Entomol 2005;98:35–40.
[44] Stapel JO, Cortesero AM, Lewis WJ. Disruptive Sublethal Effects of Insecticides on Biological Control: Altered Foraging Ability and Life Span of a Parasitoid after Feeding on Extrafloral Nectar of Cotton Treated with Systemic Insecticides. Biological Control 2000;17:243–9. https://doi.org/10.1006/bcon.1999.0795.
[45] Schneider MI, Smagghe G, Gobbi A, Viñuela E. Toxicity and pharmacokinetics of insect growth regulators and other novel insecticides on pupae of Hyposoter didmator (Hymenoptera: Ichneumonidae), a parasitoid of early larval instars of lepidopteran pests. J Econ Entomol 2003;96:1054–65.
[46] Heneberg P, Bogusch P, Astapenková A, Řezáč M. Neonicotinoid insecticides hinder the pupation and metamorphosis into adults in a crabronid wasp. Sci Rep 2020;10:7077.
[47] Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 2010;327:343–8.
[48] Gauthier J, Boulain H, van Vugt JJFA, Baudry L, Persyn E, Aury J-M, et al. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. Commun Biol 2021;4:104.
[49] Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR. Whole Genome Sequence of the Parasitoid Wasp Microplitis demolitor That Harbors an Endogenous Virus Mutualist. G3 2018;8:2875–80.
[50] Geib SM, Liang GH, Murphy TD, Sim SB. Whole Genome Sequencing of the Braconid Parasitoid Wasp Fopius arisanus, an Important Biocontrol Agent of Pest Tepritid Fruit Flies. G3 2017;7:2407–11.
[51] Ye X, Yan Z, Yang Y, Xiao S, Chen L, Wang J, et al. A chromosome-level genome assembly of the parasitoid wasp Pteromalus puparum. Mol Ecol Resour 2020;20:1384–402.
[52] Tvedte ES, Walden KKO, McElroy KE, Werren JH, Forbes AA, Hood GR, et al. Genome of the Parasitoid Wasp Diachasma alloeum, an Emerging Model for Ecological Speciation and Transitions to Asexual Reproduction. Genome Biol Evol 2019;11:2767–73.
[53] Ye X, Yang Y, Tian Z, Xu L, Yu K, Xiao S, et al. A high-quality de novo genome assembly from a single parasitoid wasp. bioRxiv 2020:2020.07.13.200725. https://doi.org/10.1101/2020.07.13.200725.
[54] Heckel DG. Insecticide Resistance After Silent Spring. Science 2012;337:1612–4.
[55] Li X, Shi H, Gao X, Liang P. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in Plutella xylostella (L.). Pest Manag Sci 2018;74:695–704.
[56] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[57] Pavlidi N, Vontas J, Van Leeuwen T. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. Curr Opin Insect Sci 2018;27:97–102.
[58] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[59] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[60] Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Current Opinion in Insect Science 2019;31:131–8.
[61] Roderick GK, Navajas M. Genes in new environments: genetics and evolution in biological control. Nat Rev Genet 2003;4:889–99.

[62] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34:3094–100.
[63] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 2021;18:170–5.
[64] Rabanal FA, Gräff M, Lanz C, Fritschi K, Llaca V, Lang M, et al. Pushing the limits of HiFi assemblies reveals centromere diversity between two Arabidopsis thaliana genomes. bioRxiv 2022:2022.02.15.480579. https://doi.org/ 10.1101/2022.02.15.480579.
[65] Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 2018;19:460.
[66] Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 2016;3:95–8.
[67] Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 2017;356:92–5.
[68] Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst 2016;3:99–101.
[69] Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 2011;12:1–14.
[70] Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 2004;32:W309–12.
[71] Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 2005;33:6494–506.
[72] Haas BJ, 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;9:1–22.
[73] Smit AFA, Hubley R. RepeatModeler Open-1.0 2008-2015. http://www.repeatmasker.org.
[74] Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics 2013;29:2487–9.
[75] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42:D222–30.
[76] Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0 2013-2015. http://www.repeatmasker.org.
[77] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36:3420–35.
[78] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 2013;29:2933–5.
[79] Bradnam K. A script to calculate a basic set of metrics from a genome assembly 2011. http://korflab.ucdavis.edu/Datasets/Assemblathon/Assemblathon2/Basic_metrics/assemblathon_stats.pl.
[80] Simão 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.
[81] Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 2019;20:1–14.
[82] Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. bioRxiv 2018:267914. https://doi.org/ 10.1101/267914.
[83] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–80.
[84] Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059–66.
[85] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.
[86] Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol 2008;25:1307–20.
[87] Nelson DR. The Cytochrome P450 Homepage. Hum Genomics 2009;4:1–7.
[88] UGT Committee. UGT Committee Home. UGT Committee Home n.d. https://prime.vetmed.wsu.edu/resources/udp-glucuronsyltransferase-homepage (accessed March 25, 2021).
[89] Li B, Du Z, Tian L, Zhang L, Huang Z, Wei S, et al. Chromosome-level genome assembly of the aphid parasitoid Aphidius gifuensis using Oxford Nanopore sequencing and Hi-C technology. Mol Ecol Resour 2021;21:941–54.
[90] Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, et al. A hybrid de novo genome assembly of the honeybee, Apis mellifera, with chromosome-length scaffolds. BMC Genomics 2019;20:275.
[91] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[92] Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 2014;45:89–110.
[93] Rotenberg D, Baumann AA, Ben-Mahmoud S, Christiaens O, Dermauw W, Ioannidis P, et al. Genome-enabled insights into the biology of thrips as crop pests. BMC Biol 2020;18:142.
[94] Guo S-K, Cao L-J, Song W, Shi P, Gao Y-F, Gong Y-J, et al. Chromosome-level assembly of the melon thrips genome yields insights into evolution of a sap-sucking lifestyle and pesticide resistance. Mol Ecol Resour 2020;20:1110–25.
[95] Pan Y, Zeng X, Wen S, Gao X, Liu X, Tian F, et al. Multiple ATP-binding cassette transporters genes are involved in thiamethoxam resistance in Aphis gossypii glover. Pestic Biochem Physiol 2020;167:104558.
[96] Pym A, Singh KS, Nordgren Å, Emyr Davies TG, Zimmer CT, Elias J, et al. Host plant adaptation in the polyphagous whitefly, Trialeurodes vaporariorum , is associated with transcriptional plasticity and altered sensitivity to insecticides. BMC Genomics

2019;20:1–19.
[97] Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, Bemisia tabaci. BMC Genomics 2017;18:1–16.
[98] Silva AX, Jander G, Samaniego H, Ramsey JS, Figueroa CC. Insecticide resistance mechanisms in the green peach aphid Myzus persicae (Hemiptera: Aphididae) I: A transcriptomic survey. PLoS One 2012;7:e36366.
[99] Dermauw W, Wybouw N, Rombauts S, Menten B, Vontas J, Grbic M, et al. A link between host plant adaptation and pesticide resistance in the polyphagous spider mite Tetranychus urticae. Proc Natl Acad Sci U S A 2013;110:E113–22.
[100] Dermauw W, Osborne EJ, Clark RM, Grbić M, Tirry L, Van Leeuwen T. A burst of ABC genes in the genome of the polyphagous spider mite Tetranychus urticae. BMC Genomics 2013;14:317.
[101] You M, Yue Z, He W, Yang X, Yang G, Xie M, et al. A heterozygous moth genome provides insights into herbivory and detoxification. Nat Genet 2013;45:220–5.
[102] Feyereisen R. INSECT P450 ENZYMES. Annu Rev Entomol 1999;44:507–33.
[103] Feyereisen R. Evolution of insect P450. Biochem Soc Trans 2006;34:1252–5.
[104] Karunker I, Benting J, Lueke B, Ponge T, Nauen R, Roditakis E, et al. Over-expression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). Insect Biochem Mol Biol 2008;38:634–44.
[105] Liang X, Xiao D, He Y, Yao J, Zhu G, Zhu KY. Insecticide-mediated up-regulation of cytochrome P450 genes in the red flour beetle (Tribolium castaneum). Int J Mol Sci 2015;16:2078–98.
[106] Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. PLoS Genet 2010;6:e1000999.
[107] Yang T, Liu N. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, Culex quinquefasciatus. PLoS One 2011;6:e29418.
[108] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[109] Main BJ, Everitt A, Cornel AJ, Hormozdiari F, Lanzaro GC. Genetic variation associated with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii. Parasit Vectors 2018;11:225.
[110] Vlogiannitis S, Mavridis K, Dermauw W, Snoeck S, Katsavou E, Morou E, et al. Reduced proinsecticide activation by cytochrome P450 confers coumaphos resistance in the major bee parasite Varroa destructor. Proc Natl Acad Sci U S A 2021;118. https://doi.org/10.1073/pnas.2020380118.
[111] Dermauw W, Van Leeuwen T, Feyereisen R. Diversity and evolution of the P450 family in arthropods. Insect Biochem Mol Biol 2020;127:103490.
[112] Feyereisen R. 8 - Insect CYP Genes and P450 Enzymes. In: Gilbert LI, editor. Insect Molecular Biology and Biochemistry, San Diego: Academic Press; 2012, p. 236–316.
[113] Yin C, Ye X, Chen M, Mei Y, Xiao H, Li F, et al. Evolution analysis of cytochrome P450 gene family in parasitoid wasps. Zhongguo Sheng Wu Fang Zhi 2019;35:335–42.
[114] Xing X, Yan M, Pang H, Wu F 'an, Wang J, Sheng S. Cytochrome P450s Are Essential for Insecticide Tolerance in the Endoparasitoid Wasp Meteorus pulchricornis (Hymenoptera: Braconidae). Insects 2021;12. https://doi.org/ 10.3390/insects12070651.
[115] Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in Acyrthosiphon pisum and Myzus persicae. Insect Mol Biol 2010;19 Suppl 2:155–64.
[116] Berenbaum MR, Johnson RM. Xenobiotic detoxification pathways in honey bees. Curr Opin Insect Sci 2015;10:51–8.
[117] Johnson RM, Mao W, Pollock HS, Niu G, Schuler MA, Berenbaum MR. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis mellifera. PLoS One 2012;7:e31051.
[118] Schmehl DR, Teal PEA, Frazier JL, Grozinger CM. Genomic analysis of the interaction between pesticide exposure and nutrition in honey bees (Apis mellifera). J Insect Physiol 2014;71:177–90.
[119] Zhang H, Zhao M, Liu Y, Zhou Z, Guo J. Identification of cytochrome P450 monooxygenase genes and their expression in response to high temperature in the alligatorweed flea beetle Agasicles hygrophila (Coleoptera: Chrysomelidae). Sci Rep 2018;8:17847.
[120] Scharf ME, Parimi S, Meinke LJ, Chandler LD, Siegfried BD. Expression and induction of three family 4 cytochrome P450 (CYP4)* genes identified from insecticide-resistant and susceptible western corn rootworms, Diabrotica virgifera virgifera. Insect Mol Biol 2001;10:139–46.
[121] Shi W, Sun J, Xu B, Li H. Molecular characterization and oxidative stress response of a cytochrome P450 gene (CYP4G11) from Apis cerana cerana. Z Naturforsch C 2013;68:509–21.
[122] Ingham VA, Jones CM, Pignatelli P, Balabanidou V, Vontas J, Wagstaff SC, et al. Dissecting the organ specificity of insecticide resistance candidate genes in Anopheles gambiae: known and novel candidate genes. BMC Genomics 2014;15:1018.
[123] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[124] Minks, K. A, Harrewijn, P. Aphids : their biology, natural enemies, and control. Amsterdam; New York: Elsevier; 1987.
[125] Field LM, Williamson MS, Moores GD, Devonshire AL. Cloning and analysis of the esterase genes conferring insecticide resistance in the peach-potato aphid, Myzus persicae (Sulzer). Biochem J 1993;294 (Pt 2):569–74.
[126] Field LM, Blackman RL, Tyler-Smith C, Devonshire AL. Relationship between amount of esterase and gene copy number in insecticide-resistant Myzus persicae (Sulzer). Biochem J 1999;339 (Pt 3):737–42.
[127] Chen W-L, Sun C-N. Purification and characterization of carboxylesterases of a rice brown planthopper, Nilaparvata lugens Stål.

Insect Biochem Mol Biol 1994;24:347–55.
[128] Karunaratne SHPP, Small GJ, Hemingway J. Characterization of the elevated esterase-associated insecticide resistance mechanism in Nilaparvata lugens (Stal) and other planthopper species. Int J Pest Manage 1999;45:225–30.
[129] Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol 2000;45:371–91.
[130] Bisset JA, Marín R, Rodríguez MM, Severson DW, Ricardo Y, French L, et al. Insecticide resistance in two Aedes aegypti (Diptera: Culicidae) strains from Costa Rica. J Med Entomol 2013;50:352–61.
[131] Xia J, Xu H, Yang Z, Pan H, Yang X, Guo Z, et al. Genome-Wide Analysis of Carboxylesterases (COEs) in the Whitefly, (Gennadius). Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20204973.
[132] Karatolos N. Molecular mechanisms of insecticide resistance in the greenhouse whitefly, Trialeurodes vaporariorum. PhD. University of Exeter, 2011. https://doi.org/ https://ore.exeter.ac.uk/repository/bitstream/handle/10036/3350/KaratolosN.pdf.
[133] Oakeshott J, Claudianos C, Campbell PM. Biochemical genetics and genomics of insect esterases. Molecular Insect … 2010.
[134] Bass C, Field LM. Gene amplification and insecticide resistance. Pest Manag Sci 2011;67:886–90.
[135] Ono M, Swanson JJ, M. Field L, Devonshire AL, D. Siegfried B. Amplification and methylation of an esterase gene associated with insecticide-resistance in greenbugs, Schizaphis graminum (Rondani) (Homoptera: Aphididae). Insect Biochem Mol Biol 1999;29:1065–73.
[136] Vontas JG, Small GJ, Hemingway J. Comparison of esterase gene amplification, gene expression and esterase activity in insecticide susceptible and resistant strains of the brown planthopper, Nilaparvata lugens (Stål). Insect Mol Biol 2000;9:655–60.
[137] Raymond M, Chevillon C, Guillemaud T, Lenormand T, Pasteur N. An overview of the evolution of overproduced esterases in the mosquito Culex pipiens. Philos Trans R Soc Lond B Biol Sci 1998;353:1707–11.
[138] Vaughan A, Rodriguez M, Hemingway J. The independent gene amplification of electrophoretically indistinguishable B esterases from the insecticide-resistant mosquito Culex quinquefasciatus. Biochem J 1995;305 (Pt 2):651–8.
[139] Zhang X-R, Zhang J-Q, Shao Y-Y, Xing X-R, Wang J, Liu Z-X, et al. Identification of glutathione-S-transferase genes by transcriptome analysis in Meteorus pulchricornis (Hymenoptera: Braconidae) and their expression patterns under stress of phoxim and cypermethrin. Comp Biochem Physiol Part D Genomics Proteomics 2019;31:100607.
[140] Aidlin Harari O, Santos-Garcia D, Musseri M, Moshitzky P, Patel M, Visendi P, et al. Molecular Evolution of the Glutathione S-Transferase Family in the Bemisia tabaci Species Complex. Genome Biol Evol 2020;12:3857–72.
[141] Vontas JG, Small GJ, Hemingway J. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens. Biochem J 2001;357:65–72.
[142] Lumjuan N, Rajatileka S, Changsom D, Wicheer J, Leelapat P, Prapanthadara L-A, et al. The role of the Aedes aegypti Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. Insect Biochem Mol Biol 2011;41:203–9.
[143] Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. Genome Biol 2014;15:R27.
[144] Ranson H, Jensen B, Wang X, Prapanthadara L, Hemingway J, Collins FH. Genetic mapping of two loci affecting DDT resistance in the malaria vector Anopheles gambiae. Insect Mol Biol 2000;9:499–507.
[145] Ranson H, Prapanthadara L-A, Hemingway J. Cloning and characterization of two glutathione S-transferases from a DDT-resistant strain of Anopheles gambiae. Biochemical Journal 1997;324:97–102. https://doi.org/10.1042/bj3240097.
[146] Pemble SE, Taylor JB. An evolutionary perspective on glutathione transferases inferred from class-theta glutathione transferase cDNA sequences. Biochem J 1992;287 (Pt 3):957–63.
[147] Shi H, Pei L, Gu S, Zhu S, Wang Y, Zhang Y, et al. Glutathione S-transferase (GST) genes in the red flour beetle, Tribolium castaneum, and comparative analysis with five additional insects. Genomics 2012;100:327–35.
[148] Yamamoto K, Zhang P, Miake F, Kashige N, Aso Y, Banno Y, et al. Cloning, expression and characterization of theta-class glutathione S-transferase from the silkworm, Bombyx mori. Comp Biochem Physiol B Biochem Mol Biol 2005;141:340–6.
[149] Tu C-PD, Akgül B. Drosophila glutathione S-transferases. Methods Enzymol 2005;401:204–26.
[150] Hassan F, Singh KP, Ali V, Behera S, Shivam P, Das P, et al. Detection and functional characterization of sigma class GST in Phlebotomus argentipes and its role in stress tolerance and DDT resistance. Sci Rep 2019;9:19636.
[151] Wang YH, Gu ZY, Wang JM, Sun SS, Wang BB, Jin YQ, et al. Changes in the activity and the expression of detoxification enzymes in silkworms (Bombyx mori) after phoxim feeding. Pestic Biochem Physiol 2013;105:13–7.
[152] Cheng J, Wang C-Y, Lyu Z-H, Lin T. Multiple Glutathione S-Transferase Genes in Heortia vitessoides (Lepidoptera: Crambidae): Identification and Expression Patterns. J Insect Sci 2018;18. https://doi.org/10.1093/jisesa/iey064.
[153] Vontas J, Blass C, Koutsos AC, David J-P, Kafatos FC, Louis C, et al. Gene expression in insecticide resistant and susceptible Anopheles gambiae strains constitutively or after insecticide exposure. Insect Mol Biol 2005;14:509–21.
[154] Yan M-W, Xing X-R, Wu F-A, Wang J, Sheng S. UDP-glycosyltransferases contribute to the tolerance of parasitoid wasps towards insecticides. Pestic Biochem Physiol 2021:104967.
[155] Silva AX, Bacigalupe LD, Luna-Rudloff M, Figueroa CC. Insecticide resistance mechanisms in the green peach aphid Myzus persicae (Hemiptera: Aphididae) II: Costs and benefits. PLoS One 2012;7:e36810.
[156] Zhao J, Xu L, Sun Y, Song P, Han Z. UDP-Glycosyltransferase Genes in the Striped Rice Stem Borer, (Walker), and Their Contribution to Chlorantraniliprole Resistance. Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20051064.
[157] Tian F, Wang Z, Li C, Liu J, Zeng X. UDP-Glycosyltransferases are involved in imidacloprid resistance in the Asian citrus psyllid, Diaphorina citri (Hemiptera: Lividae). Pestic Biochem Physiol 2019;154:23–31.
[158] Chen X, Tang C, Ma K, Xia J, Song D, Gao X-W. Overexpression of UDP-glycosyltransferase potentially involved in insecticide resistance in Aphis gossypii Glover collected from Bt cotton fields in China. Pest Manag Sci 2020;76:1371–7.

[159] Cui X, Wang C, Wang X, Li G, Liu Z, Wang H, et al. Molecular Mechanism of the UDP-Glucuronosyltransferase 2B20-like Gene (AccUGT2B20-like) in Pesticide Resistance of Apis cerana cerana. Front Genet 2020;11:592595.
[160] Dermauw W, Pym A, Bass C, Van Leeuwen T, Feyereisen R. Does host plant adaptation lead to pesticide resistance in generalist herbivores? Curr Opin Insect Sci 2018;26:25–33.
[161] Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol 2016;14:1–15.
[162] Pan Y, Xu P, Zeng X, Liu X, Shang Q. Characterization of UDP-Glucuronosyltransferases and the Potential Contribution to Nicotine Tolerance in Myzus persicae. Int J Mol Sci 2019;20:3637.
[163] Xie W, He C, Fei Z, Zhang Y. Chromosome-level genome assembly of the greenhouse whitefly (Trialeurodes vaporariorum Westwood). Mol Ecol Resour 2020;20:995–1006.
[164] Zimmer CT, Müller A, Heimbach U, Nauen R. Target-site resistance to pyrethroid insecticides in German populations of the cabbage stem flea beetle, Psylliodes chrysocephala L. (Coleoptera: Chrysomelidae). Pestic Biochem Physiol 2014;108:1–7.
[165] Højland DH, Kristensen M. Target-site and metabolic resistance against λ-cyhalothrin in cabbage stem flea beetles in Denmark n.d. https://projects.au.dk/fileadmin/projects/norbarag/Insecticide_group/Documents_Website/Hoejland_and_Kristensen_2018.pdf (accessed September 22, 2021).
[166] Liu S-S, Li Y-X, Tang Z-H. Host resistance to an insecticide favors selection of resistance in the parasitoid, Cotesia plutellae (Hymenoptera: Braconidae). Biol Control 2003;28:137–43.

Chapter 5. Final Review

This chapter is ready for submission to a journal at a future date.

Pest vs Predator: a comparative analysis of the expansion of detoxification gene families in crop pests compared to beneficial predators

Emma Bailey*†‡, Linda Field*, Christopher Rawlings†, Rob King†, Fady Mohareb‡, Keywan-Hassani Pak†, David Hughes†, Martin Williamson*, Eric Ganko**, Benjamin Buer§, Ralf Nauen§

* Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, UK

† Department of Computational and Analytical Sciences, Rothamsted Research, Harpenden, UK

‡ The Bioinformatics Group, Cranfield Soil and Agrifood Institute, Cranfield University, Cranfield, UK

** Seeds Research, Syngenta Crop Protection, LLC, Research Triangle Park, NC, USA.

§ Bayer AG, Crop Science Division, R&D, Monheim, Germany

5.1 Abstract

Insect pests are responsible for a huge amount of crop yield losses each year, and the numbers of insecticide resistance cases in these pests are continually on the rise. In contrast, the beneficial predators which form an important part of Integrated Pest Management (IPM) strategies appear to develop fewer cases of insecticide resistance. This is challenging when developing pest control strategies - the number of insecticides safe for use with beneficials is already low and so as the range of insecticides suitable for crop pests also decreases, the options for pest control become limited. This means that pesticides may have to be used which are harmful to beneficials and this can result in secondary outbreaks of the pest if the populations of beneficials are heavily impacted.

Now that beneficial predator genomes have increased in availability, they can be compared to pest genomes to assess whether there is any genetic basis for these differing capacities for developing resistance to insecticides. These investigations include comparative analyses between crop pests and their predators of insecticide target-sites and genes encoding metabolic enzymes potentially responsible for insecticide resistance. The metabolic mechanisms include several gene families - cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione-S- transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxyl/cholinesterases (CCEs).

This review has amalgamated data on the detoxification families of a variety of beneficial predators and crop pests to assess how gene expansion varies between them. The findings suggest a definite increase in gene expansion in crop pests when compared to predators, most notably in the UGT and P450 gene families. This is likely related to the diet of insect pests, as these are the main detoxification mechanisms used by insect herbivores to metabolise plant toxins. Additionally, gene expansion and resistance varies amongst beneficial predators themselves, and several factors are discussed which may play a role in this such as: migration,

diet and extent of commercial use. Finally, we discuss the shortcomings of using gene expansion as an indicator of overall resistance and some alternative resistance mechanisms which could provide us with more knowledge of why resistance levels vary.

Keywords: Beneficial predators; *Chrysoperla carnea*; *Orius laevigatus*; *Sphaerophoria rueppellii*; *Microctonus brassicae*; crop pests; insecticide; resistance; detoxification; comparative analysis

5.2 Introduction

Every year, insect pests are responsible for worldwide crop yield losses of 18-26% at a value of over £340 billion [1,2]. A constant rise in the number of insecticide resistance cases can only exacerbate this problem, and so with a growing world population to feed, it is more important than ever to find ways to maintain and increase crop yields. Integrated pest management (IPM) strategies aim to take a sustainable approach to managing pests - minimising damage to human health and the environment, as well as economic risk. They focus on monitoring pest levels to ensure they remain below the economic threshold and applying control methods when levels threaten to exceed this threshold. The cornerstone of many IPM programmes is the use of biological control, i.e. attracting or releasing beneficial insects which will eat or parasitize target pests. Chemical control still forms an important part of IPM, but applied only when absolutely necessary [3–6].

Broad-spectrum insecticides - organophosphates, carbamates and pyrethroids - have been widely used to control pest populations since being introduced in the 1940s and 1950s. Whilst these insecticides are effective against a broad range of pest species, they are also harmful to non-target species, such as beneficial predators [7]. Additionally, cases of insecticide resistance have been seen in pest species as early as 2 years after their introduction [8,9], with extensive use of these products resulting in a continual rise in insecticide resistance cases ever since [10].

Modern insecticides aim to be more selective, with reduced toxicity towards beneficial predators and increased toxicity towards pest species [11]. These insecticides are well suited to IPM strategies, as they can be used in conjunction with biological control agents to provide effective control of pest populations [12,13]. However, despite aiming to be selective, these insecticides can still be harmful to beneficial predators, through either acute or sublethal effects - such as reduced fecundity, lifespan and foraging ability [14–17]. In order for IPM

strategies to achieve their highest potential in terms of pest control, insecticides must be chosen which are selective for the pest species and cause limited damage to beneficial predator populations.

Beneficial predators are known to exhibit a lesser degree of insecticide resistance compared to pest species [18]. It has been hypothesised that beneficial predators may have limited resistance mechanisms compared to pest species, perhaps as a result of pests requiring detoxification mechanisms due to a dietary exposure to xenobiotics in plants. The smaller populations and longer life cycles of beneficial predators compared to pests would also result in a slower evolution of resistance mechanisms in beneficials.

With beneficial predator genomes having recently increased in availability [19–25], and crop pest genomes already widely available [26–32], comparative analyses can now be performed to give some indication as to why resistance is generally lower in beneficial predators compared to crop pests. There are two main types of insecticide resistance mechanisms: mutations in insecticide target genes which prevent the insecticide from binding to the target [33] and duplication or increased expression of genes encoding enzymes which can metabolise insecticides. Gene families associated with metabolic resistance include cytochrome P450 monooxygenases (P450s), ATP binding cassette transporters (ABCs), glutathione S-transferases (GSTs), UDP-glycosyltransferases (UGTs) and carboxyl/choline esterases (CCEs) [34–39]. Previous comparative analyses have focused on the expansion of detoxification gene families, which have been shown to confer resistance to insecticides [40–44]. However, expansion is not the only resistance mechanism and therefore, results from such comparative analyses should be considered carefully. It is more difficult to incorporate target site resistance into comparative analyses as the causal mutations may not be observable within the genome unless they are present within the individual or within a significant proportion of the individuals used to produce the genome. Other factors may have a significant impact on resistance mechanisms, includings: epigenetic modifications, post transcriptional regulation or

unknown pleiotropic effects from other mechanisms. Additionally, to ensure reliable conclusions when comparing genomes, the methods used to generate the assemblies/annotations and their resultant quality should be taken into account.

This review aims to cover whether there are any clear patterns in gene expansion of detoxification families and resistance levels when comparing pests and beneficial predators; the potential causes of differing degrees of gene expansion; what the limitations are when performing comparative analyses and also the challenges faced when rearing beneficials for commercial use.

5.3 Overview of beneficial predators

5.3.1 *Chrysoperla carnea*

C. carnea (green lacewing) is considered to be the most important lacewing species with regards to its efficacy as a beneficial predator of insect crop pests. It has a wide host plant range and can be used to control pests in orchards, greenhouses and crop fields. It also has a broad prey range: the larvae can feed on aphids, leafhoppers, whiteflies, thrips, Lepidoptera eggs and larvae, scales, psylla, mealybugs and spider mites, whilst the adults feed on pollen, nectar and honeydew [45–49]. *C. carnea* has been mass reared and used widely as a biological control agent all over the world, and has in fact been recognised as an effective biological control agent for over 250 years [46,50,51].

C. carnea is widespread as a result of their migratory behaviours. Female adults undergo pre-ovipository migration flights, depositing eggs in habitats far from where they emerged [52] and adults also cover vast distances of up to 300km during seasonal migration [53].

In the late 1900s, *C. carnea* was already known to be highly tolerant of many pyrethroid and microbial insecticides, with cases of carbamate and organophosphate resistance being present in select populations. Variability in resistance levels between populations from different geographical locations was correlated to pesticide usage [54–60].

Metabolic resistance (P450s and CCEs) and target site insensitivity were identified as the main mechanisms of resistance to a variety of insecticides - pyrethroids, organophosphates and carbamates - in *C. carnea* field populations [60]. Additionally, a strain of *C. carnea* reared in the laboratory achieved high levels of neonicotinoid resistance over 5 generations of selection (and maintained this for a further 4 generations), with metabolic resistance (P450s and CCEs) being identified as the main resistance mechanism [61].

5.3.2 *Orius laevigatus*

O. laevigatus (a minute pirate bug) is a highly polyphagous species which is currently most commonly used for the greenhouse control of thrips, for which they show a strong preference, however they will also feed on whiteflies, aphids and lepidoptera larvae [62–66]. *O. laevigatus* is available commercially and is widely used in pest control strategies [67] most notably for the control of thrips, for which it has been used since at least 1996 [68]. *O. laevigatus* can be found on a number of crops including deciduous fruits, corn, cotton, soybean, alfalfa and grapes [69].

O. laevigatus populations are able to persist even as crop pest populations fluctuate by feeding on alternative food sources, such as pollen or nectar, when prey populations are low [70,71]. This is beneficial for IPM strategies, as *O. laevigatus* can be pre-applied to crops before pests emerge allowing their population to establish and they will also continue to control pest

populations at low density. However, the reproductive fitness of *O. laevigatus* significantly decreases when feeding exclusively on non-prey food [71,72]. Another benefit for IPM strategies is that migration and geographical spread of *O. laevigatus* appears to be limited, even amongst neighbouring crops, with populations tending to remain in the areas they were applied, therefore reducing the need for reapplications of the predator [73].

Laboratory trials have confirmed the susceptibility of *O. laevigatus* to most pesticides, showing that neonicotinoids (imidacloprid) were very toxic by contact (but only slightly when ingested); organophosphates and pyrethroids were generally detrimental and whilst carbamates had a low toxicity to adults, there was evidence of toxicity to larval stages [74–77]. Even in greenhouse populations, where *O. laevigatus* was used to control thrips and was therefore exposed to pesticides frequently, only one of the pesticides tested was compatible with *O. laevigatus* [78]. In fact, to date there are no examples of field-developed resistance in any heteropteran predators, even though they have been widely used in pest management [79].

O. laevigatus has been shown capable of developing resistance to neonicotinoids and pyrethroids, however, this has only been achieved through the exploitation of genetic variation in wild and commercial populations to produce artificially selected strains [80,81].

5.3.3 *Sphaerophoria rueppellii*

Syrphidae (hoverfly) larvae are mainly predators of aphids, but also feed on thrips and spider mites [82–84]. Adults feed on pollen and nectar and are important pollinators [85]. *S. rueppellii* (European hoverfly) has only been made commercially available as a biological control agent in recent years [86,87], although their use in aphid control precedes this, with a 2008 study recording *S. rueppellii* as the most abundant syrphid species in Mediterranean greenhouses where flowering plants being used to attract aphidophagous hoverflies [88]. Additionally, *S.*

rueppellii has been recorded as an abundant aphid predator in northern European crops since the late 1900s, [89,90].

S. rueppellii larvae adjust their feeding rate based on aphid population levels, with low prey availability not impacting mortality, but instead increasing their development time and reducing their body size/weight. However, there were some sublethal effects on *S. rueppellii* adults if aphid numbers were particularly low [86]. This is beneficial for IPM strategies, as *S. rueppellii* populations would likely fluctuate less in response to changes in aphid population levels.

Aphidophagous syrphid species are reported to migrate long distances. While part of the population will remain in summer habitats and overwinter, the rest of the population will fly from northern Europe to southern Europe or northern Africa in autumn and return in the spring [91–93]. This has not been studied for *S. rueppellii* specifically, but it is likely they could exhibit similar migratory behaviours. *S. rueppellii* females will travel to find oviposition sites based on the presence of green leaf volatiles and the aphid-alarm pheromone [94].

Currently there is very limited information on the insecticide resistance status of *S. rueppellii*, however, it has been shown to be susceptible to pyridines and neonicotinoids via the consumption of honeydew excreted by phloem-feeding insects [95,96].

5.3.4 *Microctonus brassicae*

Literature for *M. brassicae* (a parasitoid wasp) is limited, as it was only recently discovered. Currently there are only two known populations (Harpenden, UK and Norfolk, UK), however, these populations are geographically widespread, suggesting there are likely more populations

[97]. *M. brassicae* is not currently commercially available and has not yet been used in the field as a biological control agent, although its efficacy has been demonstrated in the laboratory [97].

M. brassicae parasitises *Psylliodes chrysocephala*, commonly known as the cabbage-stem flea beetle [98]. *P. chrysocephala* feed primarily on Brassicaceae plants [99]. This is currently the only known host of *M. brassicae*, although it is certainly possible that it has multiple hosts. The resistance status of *M. brassicae* is not currently known, however, its host has developed substantial resistance to pyrethroids [100–103]. Considering *M. brassicae* has also likely been heavily exposed to pyrethroids, it is possible it has developed increased resistance mechanisms for pyrethroid detoxification. Although, host-protection may have reduced the selection pressure for such mechanisms [104,105].

5.3.5 Other biological control agents

Harmonia axyridis (Asian lady beetle) feeds on a broad range of pests including scale insects, citrus pests, lepidopteran larvae, mites and numerous aphids [106–108]. *H. axyridis* is known to migrate lengthy distances to locate breeding sites, dormancy sites and to search for new food sources [109]. *H. axyridis* has been widely used as a biological control agent in orchards and crop fields since as early as 1916 and has been commercially available since at least 2000 [106–108]. A few insecticides are known to be highly toxic to *H. axyridis*, including: Abamectin (bio-insecticide), Chlorpyrifos (organophosphate), Fenpropathrin and λ-cyhalothrin (pyrethroids). Many other tested insecticides showed low toxicity, including but not limited to: spinosad (bio-insecticide), pymetrozine (pyridine), imidacloprid (neonicotinoid), pyriproxyfen (insect growth regulator), dichlorbenzuron (chitin synthesis inhibitor) and dicofol (organochlorine) [106].

The predatory mite *Metaseiulus occidentalis* is considered the most important predator of spider mites and has been used as a biological control agent in American orchards and vineyards since 1969 [110,111]. Field-developed organophosphate resistance was reported in *M. occidentalis* populations in orchards and vineyards [112–114]. These resistant populations were used to initiate laboratory colonies which were successfully artificially selected for resistance to permethrin and carbaryl whilst maintaining organophosphate resistance (despite no further exposure to organophosphates) [113,115,116]. Additionally these strains were mass bred and released in American orchards where they successfully established and controlled spider mite populations whilst surviving pesticide applications [116–120]. These resistant strains were easily mass-reared, and were subsequently made commercially available, with over 62 million *M. occidentalis* being mass-reared in 1981 alone [121].

There are at least 230 insect species used as biological control agents [122], however, the aforementioned species are amongst the very few which have publicly available genomes and for which analysis of detoxification genes has been performed (with the notable exception of a variety of parasitoid wasp species which have been sequenced recently [24,123–127]).

5.4 Overview of crop pests

5.4.1 Spider mites

Tetranychus urticae (twospotted spider mite) is well-known for rapidly developing resistance to insecticides, with a large number of documented field-developed resistance cases. The main causes of this resistance are its high fecundity, very short life cycles and arrhenotoky (a phenomenon where unfertilised eggs become haploid males, and as such resistant traits are more quickly fixed in the population) [128]. It is also highly polyphagous with nearly 800 host plants [129] and this exposure to a wide range of plant xenobiotics will likely have promoted

the development of detoxification mechanisms. In *T. urticae*, pyrethroid resistance has generally been explained by P450 and CCE activity [130–132], however, there are examples of target-site resistance in the para sodium channel of a particularly resistant strain [133]. Organophosphate resistance has been associated with AChE point mutations [134,135] as well as AChE gene duplication [136].

5.4.2 Planthoppers

Nilaparvata lugens (brown planthopper) is a monophagous pest which feeds on rice plants [137]. This small range of host plants may have reduced the variety of natural xenobiotics which *N. lugens* has been exposed to, potentially limiting development of detoxification mechanisms pre-insecticide exposure. However, *N. lugens* possesses several characteristics which increase its potential for resistance development, including: long migratory distances [138], short development time and high fecundity [139]. A wide range of insecticides were applied heavily in rice paddy fields, resulting in a large number of field-developed resistance cases in *N. lugens* as a result of several different resistance mechanisms. Pyrethroid resistance has been associated with P450 overexpression [140]. Neonicotinoid resistance has also been associated with P450 overexpression [141] as well as target-site resistance in the nicotinic acetylcholine receptor (nAChR) [142]. Carbamate and organophosphate resistance are associated with acetylcholinesterase (AChE) point mutations [143,144] as well as amplification of a CCE gene [145].

5.4.3 Aphids

Myzus persicae (peach potato aphid) is a polyphagous pest of over 400 plant species, providing exposure to a range of natural xenobiotic compounds. *M. persicae* has been recognised for its

exceptional ability to adapt to new host plants, to the extent that they have even formed new host subspecies [146]. Additionally, *M. persicae's* short generation time and anholocyclic life cycle (meaning they can overwinter as an active life stage and reproduce faster in the spring) enhances its ability to develop resistance mechanisms [147]. Insecticides have been used intensively to control *M. persicae*, resulting in a variety of resistance mechanisms to many insecticide classes. Carbamate resistance has been associated with an AChE point mutation [148]; neonicotinoid resistance is associated with a point mutation in nAChR [149] as well as P450 overexpression [42] and pyrethroid and DDT resistance are associated with point mutations in the para sodium channel, which additionally have been predicted to arise independently in separate populations [150,151]. Amplification of CCEs (E4/FE4) is associated with broad insecticide resistance [43].

Aphis gossypii (melon aphid) is a polyphagous species with more than 600 host species and the ability to form multiple subspecies to adapt to new hosts [152,153]. Whilst they are geographically widely distributed, populations tend to overwinter locally as opposed to undergoing seasonal migration [154]. A large number of field-developed resistance cases have been reported for *A. gossypii*. Organophosphate and carbamate resistance has been associated with an AChE point mutation [155,156] as well as CCE overexpression [157] and UGT overexpression may potentially be involved in neonicotinoid detoxification [158].

5.4.4 Whiteflies

Trialeurodes vaporariorum (glasshouse whitefly) is distributed worldwide and is extremely polyphagous [159], exposing it to a variety of natural plant xenobiotics. Its short life cycle and developmental time have promoted the development of resistance mechanisms and so there are many documented cases of field-developed resistance in *T. vaporariorum*. P450 overexpression is associated with broad spectrum insecticide resistance [160] and para sodium

channel mutations are associated with pyrethroid resistance [161]. Neonicotinoid resistance has also been reported, although the resistance mechanism associated with this is currently uncertain [162].

Bemisia tabaci (sweetpotato whitefly) is distributed worldwide and has a broad host range, exposing it to a variety of natural xenobiotics. Its short life cycle and genetic diversity have helped to promote the development of several resistance mechanisms [163]. There have been multiple reported field-developed cases of resistance in *B. tabaci*. Organophosphate and carbamate resistance is associated with AChE point mutations [164]; para sodium channel mutations are associated with pyrethroid resistance [165] and neonicotinoid resistance is associated with P450 overexpression [166].

5.4.5 Thrips

Thrips palmi (melon thrip) is a highly polyphagous pest which is known to be notoriously difficult to control with insecticides [167]. Many field-developed resistance cases have been reported for organophosphates, pyrethroids and carbamates [168]. An nAChR point mutation is associated with broad spectrum insecticide resistance [169]; pyrethroid resistance is associated with a point mutation in the para sodium channel [170] and neonicotinoid resistance is associated with P450-mediated detoxification as well as nAChR point mutations [171].

In *Frankliniella occidentalis*, carbamate resistance is associated with an altered AChE target site as well as enhanced P450/CCE (and potentially GST) detoxification [172] and pyrethroid resistance is associated with enhanced P450 metabolism [173].

5.4.6 Coleoptera beetles

In *Psylliodes chrysocephala*, pyrethroid resistance has been associated with point mutations in the para sodium channel, one of which was interestingly also found in a museum sample from 1957, potentially as a result of natural pyrethrin exposure in the early 20th century [174]. Additionally, enhanced P450 detoxification has been linked to pyrethroid resistance, and in the UK has been identified as the main cause of a recent significant increase in resistance levels over only two years [175–177].

5.5 Pest versus predator

When looking at the number of documented cases of resistance from the Arthropod Pesticide Resistance Database, there are a significantly higher number of cases to a broader variety of compounds within pest species compared to beneficial predators (table 1). It is important to note that there is likely some bias in the focus of such studies, which are often pest-oriented, however, studies which have looked at beneficial predators have reported minimal field-developed resistance. There have been some successful cases of resistant laboratory-selected strains for *O. laevigatus* [80,81], but field-developed resistance has only been reported in *C. carnea* and to some extent *M. occidentalis* (although it is difficult to know how well the field-developed cases of resistance would have spread throughout the population if they had not been especially selected and mass-reared for subsequent release) [55,56,58,60,115,116].

Table 1. Numbers of documented cases of resistance and the number of active ingredients for which resistance has been reported. Numbers obtained from the 'Arthropod Pesticide Resistance Database' (APRD) in 2021. (www.pesticideresistance.org)

	Species	**# Cases**	**# Active Ingredients**
Beneficial predators	*Orius laevigatus*	8	1
	Chrysoperla carnea	162	21
	Metaseiulus occidentalis	10	8
Crop pests	*Tetranychus urticae*	551	96
	Nilaparvata lugens	448	33
	Myzus persicae	477	81
	Aphis gossypii	293	50
	Trialeurodes vaporariorum	112	28
	Bemisia tabaci	685	66
	Thrips palmi	5	4
	Frankliniella occidentalis	175	30
	Psylliodes chrysocephala	25	4

With the advance in sequencing technology over recent years, we now have access to genomic information for many of these crop pests and beneficial predators. This information can be used to perform comparative analysis which may indicate why crop pests have developed such rapid resistance, and additionally, why some beneficial predators have developed increased resistance than others.

5.6 Comparing assembly quality and annotation methods

5.6.1 Assembly quality and the impact on comparative analyses

In order to perform reliable comparative analyses studies, it is important that any genome assemblies being used are of excellent quality. Any conclusions made from these studies will only be as reliable as the assemblies used to infer them. Unfortunately, assembly quality can vary widely due to a variety of factors, including sequencing technology, read length, genome coverage and the assembly software used. Assembly completeness and contiguity can be assessed to give us an overall picture of assembly quality, however, these will not give a strong indication of whether gene copies have been missed, which is key when studying frequently expanded detoxification gene families.

Insect genomes can be especially challenging to assemble, due to a large number of repetitive sequences and a high level of heterozygosity - which is only exacerbated by the need to pool together multiple individuals to get sufficient DNA for sequencing [178].

Many insect assemblies have been produced using only short-read sequencing data such as Illumina HiSeq. These short-reads are often highly accurate, however, they are not well suited to highly repetitive regions, including tandem duplicated detoxification genes. Additionally, most short-read assemblers (such as All Paths-LG [179] and SOAPdenovo [180]) were not designed with invertebrate genomes in mind [181]. Therefore, resultant assemblies may be of poor quality and will likely have missed some additional copies of recently duplicated and/or highly similar genes.

Table 2. Comparison of assembly strategies and quality for beneficial predators and crop pest genomes.

To standardise quality assessment, the completeness of each assembly was re-assessed using 'BUSCO v4.1.4, insecta_odb10, n = 1367' in 'genome' mode [182].

	Species	Technology	Assembly software	Number of Scaffolds	N50 (bp)	Total size (Mb)	BUSCO (% complete)	Resistance genes manually curated?	GenBank assembly accession
Predators	*Orius laevigatus*	Illumina HiSeq & PacBio Sequel	Falcon; Flye; Canu (432x)	2,050	125,649	151	90.2%	Yes	GCA_018703685.1 [19]
	Chrysoperla carnea	PacBio HiFi & 10X & Hi-C	Hifiasm; SALSA2 (40x)	6 chromosomes (337 scaffolds)	94,407,144	560	97.7%	Yes	GCA_905475395.1 (DToL)
	Sphaerophoria rueppellii	Illumina HiSeq & PacBio Sequel & Hi-C	Flye; Platanus-Allee; 3D-DNA (150x)	8 chromosomes (8,476 scaffolds)	87,097,991	557	96.0%	Yes	GCA_920937365.1 [20]
	Microctonus brassicae	PacBio HiFi	Hifiasm (75x)	109	7,123,232	139	98.5%	Yes	TBC
	Metaseiulus occidentalis	Roche 454 XLR	Celera (17.7x)	2,211	896,831	152	83.4%	Yes	GCA_000255335.1 [22]
	Harmonia axyridis	PacBio RS II, 10X & Hi-C	Hifiasm; SALSA2 (53x)	14	63,675,256	426	97.6%	No	GCA_914767665.1 [21]
Pests	*Frankliniella occidentalis*	Illumina	All Paths-LG; Atlas-Link (158.7x)	6,263	948,890	416	97.9%	Yes	GCA_000697945.4 (i5k initiative)
	Diuraphis noxia	Illumina HiSeq	All Paths-LG (104x)	5,637	397,774	395	89.2%	No	GCA_001186385.1 [28]
	Acyrthosiphon pisum	Sanger & 454	Newbler; phrap; Atlas-Link (15x)	23,924	518,546	542	96.0%	Yes	GCA_000142985.2 [27]
	Myzus persicae	Rocher GS-FLX & 454 - ESTs only		-	-	-	-	No	[183]
	Thrips palmi	PacBio Sequel, Illumina & Hi-C	Wtdbg (123.8x)	17	14,670,875	238	97.0%	Yes	GCA_012932325.1
	Tetranychus urticae	Sanger	Arachne (8x)	641	2,993,488	91	80.5%	No	GCA_000239435.1 [29]
	Aphis gossypii	Illumina HiSeq	All Paths (240x)	4,718	437,960	294	93.2%	Yes (for ABCs [184])	GCA_004010815.1 [185]
	Trialeurodes vaporariorum	Illumina HiSeq	Supernova (100x)	47,417	614,417	842	86.4%	Yes	GCA_009741425.1 [186]
	Bemisia tabaci (MEAM1)	Illumina & PacBio RSII	Platanus (300x)	19,762	3,232,964	615	97.6%	Yes (for GSTs [187])	GCA_001854935.1 [188]

	Bemisia tabaci (MED/Q)	Illumina HiSeq	SOAPdenovo (594.7x)	4,973	436,791	658	91.9%	Yes (for ABCs [189], P450s [190] and CCEs [191])	GCA_003994315.1 [26]
	Nilaparvata lugens	Illumina HiSeq/GA-II	SOAPdenovo (133.3x)	46,558	356,597	1,141	89.4%	No	GCA_000757685.1 [192]

Long-read sequencing data such as PacBio Sequel (high error rate) and PacBio Hifi (low error rate) tend to cope better with the high heterozygosity and repetitive regions often seen in insect genomes. As a result, assemblies tend to have a much higher potential for detecting cases of tandem duplication; however, attempts to get a single haploid genome sequence from diploid individuals can result in copy number information being lost [193].

When looking at currently sequenced crop pests and beneficial predators, it is clear that generally the overall quality of a genome assembly (based on scaffold N50 and BUSCO score) is lower when only short-read data is used (table 2). Additionally, assemblies produced using particularly low coverage read data (<20x) are of a lower quality than those produced using higher coverage data. Many of the crop pest genome assemblies fall into these categories, as they were generated less recently and as such were produced using older technologies and methodologies than more recent species. There could therefore be an underestimate in the number of insecticide detoxification genes reported in less recently sequenced genomes.

Updated assemblies of a much higher quality have been made publicly available for many of these crop pests species over the past few years, and they will likely continue to be updated regularly as technology advances. However, comparative analyses studies which were produced using older versions would need to be re-performed to reap the benefits of this, which would likely be hugely time-consuming.

5.6.2 Annotation pipelines and their impact on comparative analyses

Annotation pipelines (such as the MAKER pipeline [194]) often struggle with annotating tandem duplications, even when transcriptome evidence is used. Due to their high similarity, adjacent gene copies are frequently annotated as a single gene, with the final gene model being far longer than expected and having an incorrect exon/intron and CDS structure (figure 1).

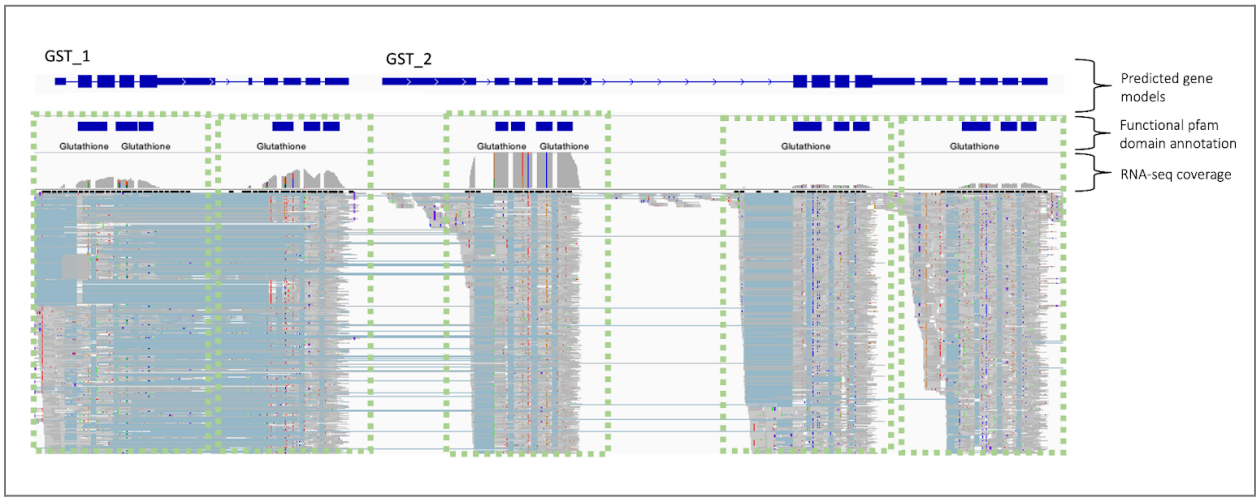


Figure 1. An example of incorrect gene models outputted from the MAKER v2.31.8 annotation pipeline, before manual curation has been performed. Visualised in IGV. GST_1 and GST-2 represent the uncurated MAKER-predicted gene models. Green boxes indicate the correct separation of gene models. RNA-seq data was mapped to the genome using HISAT v2.0.5 [195] and a coverage track indicates the depth of mapped RNA-seq reads. A track showing functional pfam annotation is also included.

Annotation using homology-based evidence often manages a slightly more accurate job of correctly annotating the exon/intron structure of tandem duplicated genes, however, this does require the gene models to have been correctly annotated or manually curated in the closely related species. This is often not the case; for example, a basic search for insect P450s (which are consistently very close to 500aa in length [196,197]) on UniprotKB returns proteins up to 4,000aa in length, many with InterPro/Pfam annotations not found in P450s. These results can

of course be filtered for those which have been manually annotated, but this hugely reduces the number of results. For example, searching UniProtKB for "IPR001128 AND insecta" returns 22,384 unreviewed results and only 109 reviewed. This becomes more of an issue when the search is narrowed further by taxonomy; for example, there were no reviewed P450s for hemiptera, hymenoptera or coleoptera at the time of writing.

Overall, manual curation of detoxification genes prior to any comparative analyses studies is essential and should be performed using RNA-seq data mapped to the genome as well as alignments to closely related genes. Without this, any functional annotation of these genes would likely be inaccurate and there could be an under- or over-estimate of the true number of these genes. Many of the studies included in table 2 did not explicitly mention any manual curation of insecticide resistance genes, and so care should be taken when interpreting any results from these.

5.6.3 How can we improve the reliability of assemblies and annotation for comparative analyses?

Overall, it is difficult to know exactly how complete a genome is or how accurate the gene models are, and this does unfortunately add a layer of uncertainty when inferring the results of comparative analyses, especially for the often-expanded detoxification genes. However, perhaps with large-scale sequencing projects such as the Darwin Tree of Life project [198] which develop specific standard operating procedures for each of the major taxa, newly released genomes may become more standardised, improving the reliability of any conclusions from comparative analyses between them.

In the case of annotation, manual curation of gene models is time-consuming, and whilst there are consortiums which collaborate on improving the quality of annotations, it would be good if

the quality of gene models generated directly from automated annotation pipelines was improved. In the case of tandem duplicated genes, it may help if differences in read depth were taken into account. Figure 1 shows clear differences in coverage between separate genes models and so this could be used to separate adjacent highly similar genes. Additionally, the functional pfam domain annotation track could be used to help infer gene models, as these proved to be reliable indicators of gene models during manual curation. And finally, another option could be to utilise the machine learning aspect of gene prediction. For example, when training gene prediction software, such as SNAP [199] and Augustus [200], only a subset of the outputted gene models from MAKER are usually used. Therefore, by ensuring that this subset contained only correct/curated gene models from tandem duplicated genes, this could improve gene model prediction from subsequent MAKER runs. Whilst this may not be applicable in fully automated pipelines, it could certainly reduce the amount of manual curation required for groups performing their own annotation.

5.7 Overview of detoxification gene families and their links to insecticide resistance

5.7.1 ABC transporters

ATP-binding cassette transporters (ABCs) are the largest known group of active transporters and are able to eliminate by translocation xenobiotic compounds such as secondary metabolites produced by plants or insecticides [201]. The ABC transporters are subdivided into eight subfamilies (ABCA-H), of which ABCB, ABCC and ABCG are the most associated with

resistance to a variety of insecticides including pyrethroids, carbamates, organophosphates and neonicotinoids [202].

5.7.2 Cytochrome P450 monooxygenase

Cytochrome P450 monooxygenase (P450s) are a diverse superfamily capable of metabolizing a huge variety of endogenous and exogenous substrates including insecticides and plant toxins [203,204]. P450s are associated with the resistance to insecticides from a variety of classes, including pyrethroids, carbamates and neonicotinoids. Many examples of resistance are linked to upregulated P450s from the CYP3 and CYP4 clans [205–208]. They are also linked to the activation of organophosphates and other proinsecticides [209] often as a result of downregulation [210,211].

5.7.3 Carboxyl/choline esterases

Carboxyl/choline esterases (CCEs) hydrolyse the ester bonds present in a wide range of insecticides. They are associated with insecticide resistance notably to organophosphates, and to a lesser degree carbamates and pyrethroids [212]. There are many examples of resistance associated with upregulated dietary esterases [44,213–218]

5.7.4 Glutathione-S-transferases

Glutathione-S-transferases (GSTs) have been shown to confer resistance to all main insecticide classes via several different mechanisms. Members of the delta class and epsilon classes have

been linked to pyrethroid resistance through reduction of lipid peroxidation and the direct metabolism of pyrethroids; and additionally both classes are able to metabolize DDT [219–223]. The omega, zeta, sigma and microsomal classes are thought to be involved in various cellular processes, including defense against oxidative stress [224–226]. Upregulation of these classes has been seen in response to organophosphates, pyrethroids and broad spectrum insecticides [104,227,228].

5.7.5 UDP-glycosyltransferases

UDP-glycosyltransferases (UGTs) are phase II detoxification enzymes which are involved in insecticide resistance. The mechanisms of UGT-mediated resistance are for example based on the conjugation of P450-functionalized substrates. They have been linked to pyrethroid resistance [229,230], carbamate resistance [231], diamide resistance [232] and neonicotinoid resistance [233,234] and they also contribute to insecticide detoxification via the elimination of oxidative stress [235].

5.8 Comparative analysis of detoxification genes between beneficial predators and crop pests

5.8.1 Overview of detoxification genes in predators and pests

Expansion of detoxification genes via gene duplication/amplification has been directly linked to insecticide resistance as a method of adaptive evolution in a variety of insect species [41,236]. Both ancient and recent gene duplications have contributed to insecticide resistance,

with intense selection over the past century likely having increased their prevalence in crop pest and beneficial predator populations [193].

Table 3. Total numbers of detoxification gene families in beneficial predator and crop pest species - also including counts for the subfamilies which are currently known to be associated with insecticide resistance.

The data and references used to generate these numbers are included in tables S1-S5.

	Species	UGTs	GSTs	CCEs		P450s		ABCs	
				All	*Dietary*	All	*CYP3/4*	All	*B/C/G*
Beneficial Predators	*Microctonus brassicae*	15	8	25	*13*	64	*50*	44	*34*
	Sphaerophoria rueppellii	46	23	40	*15*	69	*49*	45	*24*
	Orius laevigatus	10	24	32	*0*	58	*47*	64	*41*
	Chrysoperla carnea	63	20	58	*46*	99	*/*	47	*29*
	Metaseiulus occidentalis	/	13	44	*0 *37 in acari-specific class*	89	*42*	/	/
	Harmonia axyridis	/	/	/	/	63	*70*	/	/
Crop pests	*Tetranychus urticae*	81	31	71	*0 *59 in acari-specific class*	86	*33*	/	/
	Nilaparvata lugens	20	/	/	/	/	/	/	/
	Thrips palmi	17	24	/	/	91	*68*	49	*32*
	Frankliniella occidentalis	/	/	50	*28*	81	*59*	70	*46*
	Myzus persicae	101	17	22	*5*	115	*111*	/	/
	Acyrthosiphon pisum	72	19	28	*5*	83	*65*	/	/
	Aphis gossypii	/	/	/	/	/	/	71	*60*
	Trialeurodes vaporariorum	55	15	27	*12*	80	*66*	45	*25*
	Bemisia tabaci	76	23	42	*6*	171	*149*	55	*32*
Beneficial predator average		**33.5**	**17.6**	**39.8**	***18.5****	**73.7**	***51.6***	**50.0**	***32.0***
Crop pest average		**60.3**	**21.5**	**40.0**	***11.2****	**101.0**	***78.7***	**61.2**	***41.8***

**M. occidentalis* and *T. urticae* not included

Overall, crop pests have a higher total number of detoxification genes when compared to beneficial predators (table 3). When taking into account only the subfamilies which have currently been associated with insecticide resistance the pattern is generally the same, with the exception of dietary CCEs which have a slightly higher average in beneficial predators. Crop pests most notably have a higher average number of UGTs and P450s compared to beneficial predators (table 3). This could be related to their diet, as UGTs and P450s are the main detoxification mechanisms used by insect herbivores to metabolise plant toxins [237].

5.8.2 Why do pests have higher numbers of detoxification genes compared to beneficial predators?

Exposure to plant defense toxins has a far longer evolutionary history compared to insecticide exposure. Therefore, the vast majority of detoxification genes are likely to have evolved prior to insecticide use. The notably higher number of detoxification family genes required for plant toxin metabolism (UGTs and P450s) in crop pests compared to beneficial predators would fit in with this theory (table 3). The main driving force of insecticide resistance is likely an increased selection pressure for pre-existing detoxification genes as a result of intensive insecticide use [238]. It is therefore probable that plant-feeding insect species (such as crop pests) may have had an immediate advantage when exposed to insecticides. Trophic dilution would likely reduce the concentration of plant toxins that beneficial predators would be exposed to compared to crop pests - although consumption of large numbers of pests could increase this exposure; and there are a few examples of insects sequestering toxins specifically for protection against natural enemies [239]. Beneficial predators adults often feed on nectar and pollen from plants, however, the toxins are generally more diluted in these compared to the rest of the plant [240].

This 'head start' could explain the increased level of insecticide resistance that is often seen in pest species compared to beneficial predators. This, alongside the larger population sizes of crop pests would mean there is a higher likelihood of duplicated resistance genes or resistance alleles being present within the population. It is then only a matter of strong selection pressure, provided in the form of insecticides, to develop widespread resistance. In addition, beneficial predators are often mass-reared and released each crop cycle, so they may not experience such a continuous selection pressure from insecticide exposure [79].

5.9 Comparative analysis of detoxification genes within beneficial predators

5.9.1 Predicting recent gene expansion

Gene duplication/amplification generally results from unequal crossing over between homologous segments of chromosomes during meiosis or replication slippage during DNA replication. Both of these methods result in tandem repeats, and additionally, these repeats will increase the likelihood of further gene duplication events at these loci [241]. Assuming that gene duplications confer a beneficial increase in dosage, then they will tend to be fixed in a population, for as long as the conditions which lead to this selection remain [242]. Genes for which their amplification has been shown to have a role in adaptation to environmental conditions tend to be amongst the most recently duplicated genes in the genome. The fixation of such paralogs in a population is suggested to have an immediate and direct impact on fitness [243]. It is also hypothesized that there will be an optimum number of gene copies within the genome, according to environmental conditions. Therefore, if the environment is constant - i.e. exposure to insecticides is continuous - then the number of gene duplications may decrease slightly once an equilibrium has been reached. On the other hand, in a variable environment - i.e. with varying exposure to insecticides - cycles of gene duplication and loss often occur. Under this model, duplicated gene copies will often be nearly identical, which is often the case for amplified resistance-associated genes [242, 244].

Based on the assumption that genes for environmental adaptation tend to be recently duplicated and thus will often have a high identity and be arranged in tandem repeats, we can predict which detoxification gene families have most recently expanded for each beneficial predator species. The results of this analysis are shown in table 4.

Table 4. Total numbers of detoxification genes, and the percentage of these which are likely tandem duplications (i.e. recent duplications) based on having a similarity ≥70 and genes being adjacent within the genome. Similarity measured using Blosum45 with threshold 0, and proteins aligned using MAFFT.

	O. laevigatus		*S. rueppellii*		*M. brassicae*		*C. carnea*		Average
	Percentage	Total	Percentage	Total	Percentage	Total	Percentage	Total	Percentage
ABCs	6.25%	**64**	44.44%	45	9.09%	44	27.66%	47	21.86%
P450s	27.59%	58	66.67%	69	70.31%	64	55.56%	**99**	55.03%
CCEs	25.00%	32	47.50%	40	36.00%	25	65.52%	**58**	43.51%
GSTs	54.17%	**24**	69.57%	23	0.00%	8	50.00%	20	43.44%
UGTs	20.00%	10	82.61%	46	66.67%	15	79.37%	**63**	62.16%
Average	26.60%	38	62.16%	45	36.41%	31	55.62%	**56**	

Based on the highest percentage of adjacent high similarity genes, *S. rueppellii* appears to have the greatest degree of recent detoxification gene family expansion, followed by *C. carnea*, *M. brassicae* and lastly *O. laevigatus*. This did not correlate with the largest total of detoxification genes, for which *C. carnea* has the highest total number. The apparently most recently expanded detoxification gene family varied by species: GSTs for *O. laevigatus*, P450s for *M brassicae* and UGTs for *S. rueppellii* and *C. carnea*.

These beneficial predator genomes exhibited a wide range of contiguity, with N50 scores ranging from 125,649bp to 94,407,144bp (table 1). Genomes with a lower contiguity may have captured less tandem duplications due to shorter scaffolds and increased fragmentation which could have resulted in gene clusters potentially being spread across multiple scaffolds. In an attempt to account for this variation of contiguity, OrthoFinder was used to infer the overall number of gene duplications for each species, as this software uses only the protein sequences as input and thus ignores genomic location (figure 2).

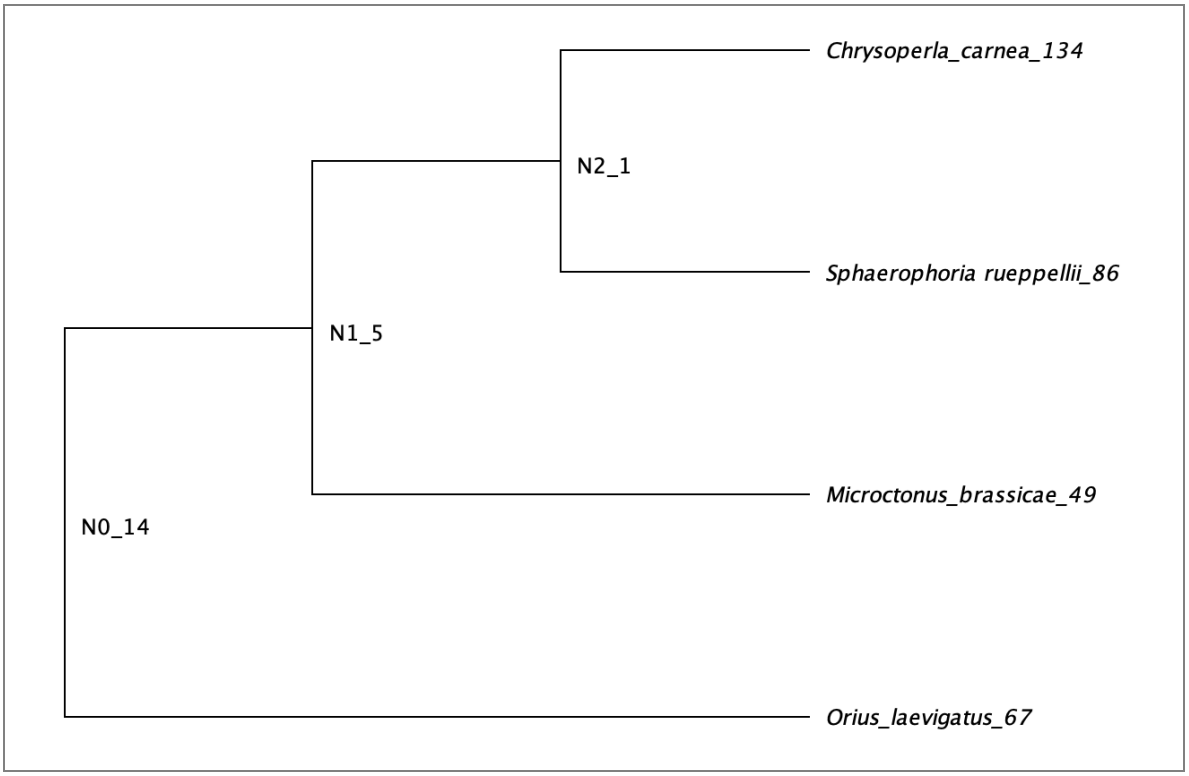


Figure 2. Full-length detoxification genes duplication tree inferred using OrthoFinder. The node name is followed by the number of well-supported gene duplication events mapped to each node in the species tree (considered 'well-supported' if at least 50% of descendant species have retained both copies of the duplicated gene). The species name is followed by the number of lineage-specific duplication events.

From the OrthoFinder analysis it was evident that there are a small number of ancient duplicated detoxification genes which are shared by all beneficial predator species, however, the vast majority are lineage-specific and hence more recent (figure 2). *C. carnea* appears to have the highest number of duplicated detoxification genes, whilst *M. brassicae* has the lowest. The results of OrthoFinder differed slightly from the results of comparing likely recent tandem duplications, which was expected with the differing levels of assembly contiguity. However, it is often the case that duplicated detoxification genes confer a selective advantage to the individual, and therefore it is probable that less recently duplicated genes may still have a high identity whilst being dispersed throughout the genome over time. It is therefore possible that whilst *C. carnea* exhibited the most detoxification expansion overall, *S. rueppellii* has undergone the most recent gene expansion.

5.9.2 Potential causes of differences in detoxification gene family expansion amongst beneficial predators

Detoxification genes play an important role in environmental adaptation. Therefore, a variety of environmental factors and lifestyle may have contributed to the differing degrees of detoxification gene expansion seen amongst different beneficial predator species. This includes differences in behaviour or habitat, as well as more recent pressures, such as insecticide exposure and mass-rearing.

Genetic predisposition

The genetic predisposition of a species contributes to the likelihood that it will develop resistance. This predisposition results from the differing history and ecology of each species, which will have resulted in a specific set of adaptation mechanisms. This includes differing exposure to: plant xenobiotics, bacterium, virus vectors (from which horizontal gene transfer can result in insects gaining novel detoxification genes [188]) and more recently, insecticides. In addition, factors such as genetic heterogeneity and population size can contribute to resistance development. For example, within larger or more heterogeneous populations, there is a greater likelihood that resistance alleles or duplicated detoxification genes are already present at low frequencies prior to pesticide application. These can arise in response to natural toxins; through pleiotropic effects (i.e. as a byproduct of adaptation unrelated to toxin resistance) and also simply by neutral processes (i.e. random mutation and genetic drift) [245].

Habitats and diet

Studies have shown that species which are not exposed to diverse habitats tend to lose duplicated genes, such as those detoxification genes which play an important role in environmental adaptation [246,247]. Host-specific parasitoids such as *M. brassicae* likely fall

under this category [248]. On the other hand, *O. laevigatus*, *S. rueppellii* and *C. carnea* all have a broad host plant and prey range [45–49,62–66,69,82–84].

Adult *C. carnea*, *S. rueppellii* and *M. brassicae* (and *O. laevigatus*, but only if prey is limited) feed on plant products such as pollen, nectar and honey. Insects with a pollen-based diet have been found to have an increased degree of insecticide tolerance, with many of the same genes being upregulated in response to both pollen and to certain insecticides [249].

Migration

Species which migrate will be exposed to more diverse habitats, and so are likely to develop a more diverse set of detoxification genes. Additionally, even if only a small proportion of the population migrates, this is sufficient to increase the genetic homogeneity of geographically distant populations through widespread gene flow [250], and thus will result in a potential spread of resistance genes [251]. Current research suggests migration of *O. laevigatus* and *M. brassicae* to be very limited [73,97] whereas *S. rueppellii* and *C. carnea* are capable of migrating vast distances [53,91]. Indeed, migration has already been shown to cause a high level of genetic variation in predatory hoverfly species [252].

In addition to seasonal migration, female *S. rueppellii* and *C. carnea* adults are both known to migrate in search for oviposition sites which have high prey availability based on the presence of certain semiochemicals, often choosing sites which are distantly located from their site of emergence [52,94].

Varied levels of commercial use/insecticide exposure

Of these four beneficial predators, *C. carnea* has the most extensive use, having been recognised as an effective biological control agent for over 250 years and used commercially

since the late 1970s [46,50,51,253,254]. *O. laevigatus* is the next most heavily used beneficial predator, having been in use as a commercial biological control agent within Europe since 1996 [68]. *S. rueppellii* has only been made available commercially in the past decade [86,87], although flowering plants were used to attract them to greenhouses in the early 2000s [88] and they have been recorded in northern European crops since the late 1900s [89,90]. *M. brassicae* has only been discovered very recently, and has only been reported in two sites in the UK so far [97]. They are not currently commercially available, and as such have not been mass-reared.

The long-standing use of biological control agents such as *C. carnea* increases the likelihood they have been exposed to a wide variety of pesticides, potentially over a long period of time. This prolonged insecticide exposure has likely provided strong selection pressure for detoxification genes.

Population fluctuation and mass-rearing

If beneficial predator populations experience less fluctuation in size there is a reduced chance of genetic bottlenecks which may reduce genetic variation and hence reduce the number and variety of resistance alleles present. Beneficial predators which can feed on multiple food sources are more likely to gain field-developed resistance because they can maintain their population when pest density is low [255].

O. laevigatus populations are able to persist even as crop pest populations fluctuate by feeding on alternative food sources such as pollen and insect eggs [70,71]. *C. carnea* larvae can also feed on alternative food sources such as honeydew and insect eggs [256,257]. *S. rueppellii* larvae can adjust their feeding rate based on aphid population levels, ensuring that low prey availability does not impact mortality [86]. Alternatively, whilst *M. brassicae* does not depend on its host *P. chrysocephala* for food in the same way as other beneficial predators, it does rely

on its host for reproduction and larval development, and therefore when its host population is reduced, *M. brassicae* populations will likely fluctuate accordingly [97].

Whilst the majority of these species can maintain their populations using alternative food sources, factors such as insecticide applications may cause dramatic fluctuations in their population which could result in genetic bottlenecks. Additionally, mass-reared colonies often suffer from the founder effect which causes an initial population bottleneck [258]. This initial bottleneck is likely partly responsible for the reduced genetic variation of mass-reared colonies compared to wild populations. This has been shown to impact their overall fitness, insecticide resistance capabilities and efficacy in pest control [259]. In addition, a lack of habitat diversity and insecticide exposure in mass-reared colonies reduces the selection pressure for detoxification genes, and may result in the loss of resistance alleles. Resistance has frequently been shown to decline without insecticide exposure [260–262]. This is generally attributed to the fitness cost of overexpressed detoxification genes or target-site modifications relative to susceptible individuals in the absence of insecticide exposure [263]. For this reason resistance genes are rarely fixed within populations that are not under intense selection pressure from insecticides and resistance can decline in such populations. However, if the resistance genes are fixed then resistance can remain stable for several years even in the absence of insecticides [264].

O. laevigatus and *S. rueppellii* assemblies were produced from individuals obtained commercially from mass-reared colonies, and so may have reduced fitness. Whereas, *M. brassicae* and *C. carnea* (BioSample ID: SAMEA7520372) assemblies were generated from field-collected individuals and so could be expected to have increased fitness. The degree of insecticide exposure experienced by the populations these individuals were obtained from is unknown.

Host/symbiont protection

M. brassicae is the only parasitoid beneficial predator discussed here and is therefore the only predator which will receive host-protection. This may have reduced the selection pressure for resistance mechanisms in *M. brassicae* [104,105].

Bacterial symbionts such as *Rickettsia* spp., which are commonly found in green lacewings [265], have been shown to confer positive fitness effects such as enhanced detoxification to their host [266,267] and may reduce selection pressure for some resistance mechanisms in their host.

5.9.3 Why does resistance vary amongst beneficial predators?

Lacewings have a large number of field-developed resistance cases [54–58,60,268]. Whereas *O. laevigatus* has little evidence of field-developed resistance, even under repeated pesticide exposure in greenhouses [78]. There is currently very limited information on the resistance status of *S. rueppellii* and *M. brassicae*.

The number of field-developed resistance cases correlates with the higher degree of detoxification gene expansion seen in *C. carnea* compared to *O. laevigatus* (figure 2). It is expected that *C. carnea* would have high levels of resistance: it has prolonged commercial use, undergoes migration over long distances and the adults feed primarily on plant byproducts. On the other hand, whilst *O. laevigatus* does have fairly extensive commercial use, its migration is limited and all life stages feed primarily on insects pests. There are in fact no field-developed cases of resistance in heteropteran predators [255]. This lack of exposure to plant xenobiotics in the diet of heteropteran predators such as *O. laevigatus* has likely resulted in them having

minimal pre-existing detoxification mechanisms prior to insecticide exposure. This in turn may have hindered their development of insecticide resistance.

When looking at our predictions for the most recent expansion, *S. rueppellii* appears to have experienced detoxification gene expansion more recently than *C. carnea* (table 4). This would make sense, as *S. rueppellii* has only been used as a commercial biological control agent over the past decade. We would expect gene expansion to be more frequent in early commercial use / exposure to insecticides, as after prolonged insecticide exposure an optimum number of gene copies would likely be reached. After this point, the fitness cost of extra copies would suppress further expansion. However, this pattern is not consistent for the other species, as *C. carnea* appears to have undergone more recent expansion than *O. laevigatus* which began use as a commercial biological control agent more recently. This could be explained by exposure to new/different classes of insecticides which might trigger further bouts of gene expansion. However, overall it is likely that comparing gene identities is too oversimplified for drawing conclusions on the recent development of resistance mechanisms, as this does not take into account target site resistance, neofunctionalization of genes or pleiotropic effects from ancient duplications.

5.10 Other causes of increased resistance

Gene expansion (and target site mutations) have strong associations with insecticide resistance, but they are not the only mechanisms that can confer insecticide resistance. As a result of the complex interplay of these many resistance mechanisms, gene expansion alone is not a perfect indicator of an individual's detoxification capabilities.

5.10.1 Neofunctionalization

Gene duplication results in increased expression, but can also lead to the neofunctionalization of resistance genes. For example, neofunctionalization of a P450 gene in *N. lugens* lead to individuals possessing paralogs of the gene both with and without a gain-of-function mutation that conferred resistance to a particular insecticide (imidacloprid). Additionally, only the paralog which conferred resistance was highly expressed, as a result of novel cis-acting elements in an upstream region [269].

Neofunctionalization was the most common process maintaining duplicated genes in *Drosophila*, likely as a result of its large population size and thus more efficient natural selection. Species with smaller population sizes, such as *A. mellifera*, tend to have less efficient natural selection, and thus neofunctionalized genes will be fixed less often [270] .

5.10.2 Epigenetics and post-transcriptional regulation

Novel mutations and changes in gene copy number are unlikely to become fixed in insect populations within only a few generations. On the other hand, transgenerational epigenetic inheritance could allow gene expression changes to become widespread in insect populations on a much shorter time scale [271]. Epigenetic changes induced by chemical stressors can be inherited by F1 offspring, and have been shown to persist for at least two further generations without any subsequent exposure to the chemical stressor [272]. In addition, epigenetic mechanisms can regulate the expression of duplicated/amplified detoxification genes to coincide with fluctuations in insecticide exposure. This avoids the fitness costs associated with overexpressing detoxification genes in the absence of any selective advantage [263, 273]. The methylation and consequential upregulation of a P450 gene were associated with resistance in strain of *Bemisia tabaci*, with knockdown of methyltransferases resulting in increased

sensitivity to a neonicotinoid. Additionally, insecticide exposure was found to increase global N^6-methyladenosine (methylation on the N6-position of adenosine) levels [274]. Mutations in cis- or trans-acting regulatory elements which can increase expression of detoxification genes have also been reported in several insect species [275–280].

Post-transcriptional regulation of resistance-associated genes has not been well studied; however, one example is the differential expression of miRNAs which post-transcriptionally regulate detoxification genes in resistant versus susceptible strains of *Drosophila* and *Culex pipiens pallens* [281, 282].

5.10.3 Others

Factors such as climate change can result in the selection of certain climatic adaptation traits which may favour insecticide resistance traits due to a common molecular basis of both mechanisms, or impose a fitness cost on resistance traits which results in their loss [283]. Studies have also reported links between insect gut microbiota and enhanced or decreased resistance [284]. There is also some evidence of behavioural insecticide resistance, with receptor based aversion suggested to be the most likely mechanism responsible for this [285].

5.11 Should we still use gene expansion for comparative analyses?

The existing diversity of detoxification gene superfamilies is largely the result of expansion and diversification over the course of hundreds of millions of years, with many gene family members shared across multiple species. This diversification has helped to facilitate adaptation of each species to their particular ecological niche, for example: diet, host-plant

adaptation and semiochemical signaling (including pheromones and defense chemicals) [203,286–288].

Insecticide resistance is commonly attributed to expansion within these gene families, however this is likely only a secondary effect. Any widespread gene expansion in direct response to insecticides could only have occurred since their introduction in the past couple of hundred years. Even with large population sizes and strong selection pressure from intensive insecticide use, we would still expect that any duplicated genes which evolved in response to this would have extremely high similarity, likely a percentage in the high nineties. Such high identity duplicated genes do exist within insects, for example in *O. laevigatus* and *S. rueppellii*, but only in very low numbers [19,20]. Although, it could be argued that insecticide exposure may have increased the prevalence of some pre-existing expanded detoxification genes due to heightened selection pressure within insect populations.

With insect genomes currently increasing in availability, gene expansion is a good starting point to study insecticide resistance. The fact that tandem duplications of these genes can arise independently around the world [289], can become fixed in populations and sometimes lead to neofunctionalization [41,269], suggests that these duplications serve a useful function, and are therefore a good indicator of potential increases in insecticide resistance, even if this is only a secondary effect. However, we still need a significantly larger number of pest and predator genomes to be available for reliable comparisons to be made. Projects such as the Darwin Tree of Life (DToL) project [198], which aims to sequence 70,000 species in the UK, will help in this endeavor. Increasing the numbers of sequenced individuals would allow for more solid predictions when performing comparative analyses. The DToL project aims to develop specific standard operating procedures for each of the major taxa in order to standardise methods for genome assembly and annotation. This will also allow for more reliable comparisons of gene expansion amongst species. Additionally, the sequencing of close

relatives of insects with genomes currently available, would enable us to perform in-depth molecular clock analyses to estimate the age of certain duplications.

It is, however, important to note that these genomes will only be representative of the population from which the sequenced individuals came. Numbers of duplicated genes could vary heavily amongst populations with differing levels of insecticide exposure, especially in the case of geographically separated populations. It may therefore be useful in the future - with genome sequencing and assembly having become so readily accessible - to produce pan-genomes, as these would give a more representative overview of the differing resistance mechanisms of each species and between resistant and susceptible populations. In addition, standardising the methods used to produce these pan-genomes would enhance the reliability of any subsequent comparative analyses.

It is likely that changes in gene expression are mostly responsible for more immediate changes in resistance level. Epigenetic factors and post-transcriptional regulation can alter the expression of gene copies, potentially resulting in a far different amount of gene product than expected from the number of copies. For example, microsatellite variation can drive rapid adaptive change of gene expression in response to insecticides [290]. It would be difficult to detect and compare many of these mechanisms from the genome alone. This highlights the need for further studies before concluding the phenotypic results of gene expansion, such as expression analysis, homology modeling and metabolism assays. Epigenomes could also provide useful information for studying regulation of resistance gene expression.

Additionally, gain-of-function mutations may not be detected when analysing changes in gene copy number alone, and could cause a far more substantial change in resistance capability compared to increased expression of the original gene copy alone.

Finally, a better understanding of the precise function of each detoxification gene, using knockout and functional analyses may also help us better understand how and why these gene superfamilies evolved.

5.12 Challenges of mass-rearing resistant and effective beneficial predators

Insecticide resistance is likely higher in crop pests compared to beneficial predators as a result of many lifestyle and environmental factors including: shorter generation time, larger population size, increased exposure to plant xenobiotics and more intensive insecticide pressure. Such factors are likely responsible for pests having a greater degree of detoxification gene expansion (and likely a greater number of resistant alleles). However, through the artificial selection and mass-rearing of resistant strains, we can attempt to help beneficial predators 'catch up' with crop pests from an evolutionary perspective.

Artificial selection of resistant beneficial predators has been achieved for several species [81,291–293]. However, mass rearing them for commercial use comes with several challenges. Rearing using an artificial environment and diet can result in altered behavioural responses to the target pest as well as the host plant (which can provide cues for locating the pest) [294,295]. For example, a mass reared colony of *C. carnea* had reduced efficacy once released because the larvae had reduced foraging capabilities and the adults did not respond to the wild sunflowers which they had previously used as a source of nectar [296]. Additionally, genetic diversity will inevitably decrease in mass-reared populations over time as a result of reduced genetic drift and inbreeding [297]. In *Drosophila melanogaster*, inbred populations have a much higher probability of extinction when facing environmental stressors due to a lack of genetic variation, and they are often incapable of surviving when returned to the wild due to

the genetic deterioration of 'wild' fitness [298,299]. There are a few strategies which can be used to reduce the impact of these issues, including: introducing individuals from the wild to the colony to increase genetic variation; splitting the founding population into subpopulations which will each lose some genetic variation but can then be mixed before reintroduction to the wild to restore the original variation; and finally, trying to replicate wild conditions in the artificial rearing environment (such as exposure to their insect prey and host plant) [297]. However, these options may be considered time-consuming and costly when rearing huge numbers of predators for commercial release.

Alongside preserving the efficacy of biological control agents during mass-rearing, there are also difficulties with maintaining high levels of insecticide resistance within colonies. Once insecticide pressure is removed, resistance often declines due to the associated fitness cost. This has been associated with a decrease in frequency of resistance alleles [261] and reduced methylation/expression of amplified detoxification genes. Interestingly, this methylation did not return with a subsequent increase in insecticide pressure [300]. Maintaining insecticide pressure whilst mass-rearing colonies could circumvent these problems, however, if the overall fitness of a colony declines, this can also impact resistance. Increased fitness is associated with an increased resistance to insecticides in *C. carnea* [301–303], and conversely, resistance can also decline due to an overall deterioration of fitness. For example, in a *C. carnea* laboratory culture, resistance to all insecticide classes declined substantially over 3 years [60]. This was associated with an overall deterioration of fitness as a result of inbreeding multiple generations, possibly exacerbated by this study using populations of only ~50 *C. carnea* larvae to start the colonies [304]. In contrast, another study which used ~300 *C. carnea* larvae to initiate a colony managed to maintain resistance for 4 generations without any further insecticide exposure [61]. The increased size of this initial colony compared to the aforementioned study likely resulted in higher fitness levels due to a less severe founder effect [305]. However, even if a large founding population is used to initialize a colony, only some of

the genotypes may be favoured under mass-rearing conditions, which may result in a genetic bottleneck similar to the founder effect [297].

Aside from mass-rearing colonies, beneficial predators can be attracted to crops, greenhouses and orchards using attractive host plants or 'trap cropping', which has proven highly effective in the control of pest populations [306]. These beneficial predators may not be resistant to the insecticides used, but natural populations will likely have a higher efficacy as biological control agents and may limit the need for supplemental insecticide applications, which could anyway trigger secondary pest outbreaks [307].

5.13 Conclusion

Biological control is considered to be the 'cornerstone' of IPM strategies, and so we are certainly headed in the right direction by steadily increasing the amount of genomic information available for beneficial predators. With this information we have been able to perform comparative analyses of beneficials and crop pests, and these have revealed a greater degree of gene expansion in pests compared to predators. In particular, the UGT and P450 detoxification families appear to have undergone more expansion in pests, suggesting that the heavy exposure to plant xenobiotics in their diet has given them an advantage when it comes to developing resistance mechanisms compared to beneficial predators. In addition, there is variation in the levels of resistance between different beneficial predators, which also correlates to some extent with gene expansion. Several factors have likely influenced this variation, including diet, migration and length of commercial use.

Whilst gene expansion alone is not a perfect indicator of resistance levels, it is a good starting point. In the future, a larger number of high quality genomes - as a result of the DToL project

for example - will allow for more accurate comparative analyses. Additionally, epigenomes could give a better indication of how resistance genes are regulated and how epigenetic modifications affect resistance in future generations. Pan-genomes could give a better overview of resistance mechanisms for the species as a whole and could help infer how detoxification families vary between resistant and susceptible populations as well as populations from distinct geographical locations. These resources could help us gain a deeper understanding of why intraspecies resistance is often so varied, the mechanisms behind resistance inheritance and also how resistance is lost and gained. Answers to such questions could help with the selection of targeted insecticides when developing IPM strategies which limit resistance gain in pests, promote resistance in predators, and assist with identifying best practices for the mass-rearing of resistant beneficial predators.

5.14 Materials and Methods

5.14.1 Assessing genome completeness

The completeness of all pest and predator genomes was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) [182] of the insect gene set (insecta odb 10). ‘Genome’ mode was used to assess the assembly. ‘Fly’ was used as the training species for gene prediction. BUSCO assessments were then run with default parameters.

5.14.2 *Chrysoperla carnea* genome annotation

The *C. carnea* genome produced by the Darwin Tree of Life Project (www.darwintreeoflife.org/) available at: GCA_905475395.1, was used for in-house annotation. Illumina RNA-seq data from ENA was used to aid annotation, available through accessions: SRR10012086 and SRR10012087.

Gene prediction was performed using the MAKER v2.31.8 pipeline [194] through the incorporation of both transcriptome evidence and *ab initio* gene prediction as well as a custom repeat library (see below). MAKER was run using Augustus v3.3.1 [200], GeneMark-ES v4.32 [308] and FGeneSH v8.0.0 [309] as well as EVidenceModeler v1.1.1 [310] with default masking options.

A *de novo* species specific repeat library was constructed using RepeatModeller v1.0.7 [311] to identify repeat models. These models were searched against the GenBank non-redundant (*nr*) protein database for Arthropoda (e value $<10^{-3}$) using Blastx to remove any potential protein-coding genes. This was combined with transposon data to create a custom library. Transposons were identified from the transcriptome assembly by running HMMER: hmmscan [312] against the Pfam database [313] and filtering the resultant Pfam descriptions for those containing "transposon". A search for transposons was also performed on transcripts produced from MAKER and these transposons were then added to the custom repeat library which was used for a second round of MAKER. RepeatMasker v4.0.7 [314] was used to mask repeats in the genome assembly using these repeat libraries, as well as to estimate the abundances of all predicted repeats.

RNA-seq reads were mapped to the genome with HISAT2 v2.0.5 [195] for assembly with StringTie v1.0.1 [315]. A de novo assembly was also done using Trinity v2.5.1 [316]. The best

transcripts were selected from the Trinity and StringTie assemblies using Evigene v19.jan01 [317].

Evidence from assembled transcripts was transferred to the genome assembly via MAKER. The output from this was then used to produce a high confidence level gene model training set. Overlapping and redundant gene models were removed. Augustus and GeneMark were trained using this training set prior to being used for *ab initio* gene predictions. FGeneSH was run based on the *Drosophila melanogaster* genome.

The best transcripts (classified by reasonable transcript size and homology to other species) from both the *ab initio* gene prediction annotation and the transcriptome-based annotation were selected using Evigene and combined to create the final annotation.

C. carnea protein sequences were aligned using Blastp against the non-redundant (nr) NCBI protein database for Arthropoda. InterProscan searches were run against several databases (CDD, HAMAP, HMMPAnther, HMMPfam, HMMPIR, FPrintScan, BlastProDom, ProfileScan, HMMTigr) for functional annotation. BLAST2GO [318] was used to assign gene ontology (GO annotations). Infernal v1.1.2 [319] was used to predict and annotate non-coding RNAs.

5.14.3 Identifying and curating *Chrysoperla carnea* detoxification genes

In order to identify genes potentially involved in insecticide resistance, the PFAM domains assigned to gene models during annotation (as described in the 'Genome Annotation' methods section) were used as follows: CCEs (PF00135/IPR002018), GSTs (IPR004045/PF02798), (IPR004046/PF00043), P450s (IPR001128/PF00067), ABCs (IPR003439/PF00005) and UGTs (IPR002213/PF00201). Proteins from UniProtKB for the classes of interest, from neuropteran species, were used for BLAST queries against *C. carnea* to identify any missed genes and to

assist with subfamily assignment within these classes. Subfamily assignment for gene families was finalised using phylogenetic trees which were produced using MAFFT alignments [320,321] and RaxML v8.2.11 [322]. The GAMMA LG protein model [323] was used and a bootstrap consensus tree was inferred from 100 replicates.

Manual checks and curation were performed for genes potentially involved in insecticide resistance. Increased copy numbers of genes linked to insecticide resistance often led to adjacent tandem duplications being incorrectly annotated by MAKER as one gene model; therefore curation was important to prevent incorrect gene numbers being reported in later analyses. The exon/intron boundaries and start/stop codons of the genes were confirmed through visualization in IGV [324] of RNAseq data mapped to the genome using HISAT2 v2.0.5 [195] and the gene models were edited in Geneious where necessary.

5.14.4 Comparative analysis methods

For *O. laevigatus*, *S. rueppellii* and *M. brassicae*, assemblies and annotations had been produced in the author's previous work. The methods used to generate these were covered in the author's previous papers [19,20] (*M. brassicae* paper to be released).

In order to calculate the percentage of detoxification genes which were likely tandem duplications within beneficial predator species, proteins were aligned using MAFFT [320,321]. High similarity genes were those with a similarity ⩾70 when using Blosum45 with threshold 0, and those which were also adjacent within the genome were considered to be tandem duplications.

5.15 Literature Cited

[1] Oerke E-C. Crop losses to pests. J Agric Sci 2006;144:31–43.
[2] Culliney TW. Crop Losses to Arthropods. In: Pimentel D, Peshin R, editors. Integrated Pest Management: Pesticide Problems, Vol.3, Dordrecht: Springer Netherlands; 2014, p. 201–25.
[3] Barzman M, Bàrberi P, Birch ANE, Boonekamp P, Dachbrodt-Saaydeh S, Graf B, et al. Eight principles of integrated pest management. Agron Sustain Dev 2015;35:1199–215.
[4] Stenberg JA. A Conceptual Framework for Integrated Pest Management. Trends Plant Sci 2017;22:759–69.
[5] Prokopy R, Kogan M. Chapter 139 - Integrated Pest Management. In: Resh VH, Cardé RT, editors. Encyclopedia of Insects (Second Edition), San Diego: Academic Press; 2009, p. 523–8.
[6] Ehler LE. Integrated pest management (IPM): definition, historical development and implementation, and the other IPM. Pest Manag Sci 2006;62:787–9.
[7] Perkins JH. Insects, Experts, and the Insecticide Crisis: The Quest for New Pest Management Strategies. Springer Science & Business Media; 2012.
[8] Whitfield, B. J, Doyen, T. J, Purcell, H. A, et al. Daly and Doyen's introduction to insect biology and diversity. New York: Oxford University Press; 2014.
[9] Weston DP, Poynton HC, Wellborn GA, Lydy MJ, Blalock BJ, Sepulveda MS, et al. Multiple origins of pyrethroid insecticide resistance across the species complex of a nontarget aquatic crustacean, Hyalella azteca. Proc Natl Acad Sci U S A 2013;110:16532–7.
[10] Sparks TC, Nauen R. IRAC: Mode of action classification and insecticide resistance management. Pestic Biochem Physiol 2015;121:122–8.
[11] Hainzl D, Cole LM, Casida JE. Mechanisms for selective toxicity of fipronil insecticide and its sulfone metabolite and desulfinyl photoproduct. Chem Res Toxicol 1998;11:1529–35.
[12] Kogan M. Integrated pest management: historical perspectives and contemporary developments. Annu Rev Entomol 1998;43:243–70.
[13] Horne PA, Page J. Integrated Pest Management for Crops and Pastures. Csiro Publishing; 2008.
[14] Galvan TL, Koch RL, Hutchison WD. Effects of spinosad and indoxacarb on survival, development, and reproduction of the multicolored Asian lady beetle (Coleoptera: Coccinellidae). Biol Control 2005;34:108–14.
[15] Van Driesche RG, Lyon S, Nunn C. Compatibility of spinosad with predacious mites (Acari: Phytoseiidae) used to control western flower thrips (Thysanoptera: Thripidae) in greenhouse crops. Fla Entomol 2006.
[16] Hattingh V. The use of insect growth regulators - implications for ipm with citrus in southern africa as an example. Entomophaga 1996;41:513–8.
[17] Cole PG, Cutler AR, Kobelt AJ, Horne PA. Acute and long-term effects of selective insecticides on Micromus tasmaniae Walker (Neuroptera: Hemerobiidae), Coccinella transversalis F. (Coleoptera: Coccinellidae) and Nabis kinbergii Reuter (Hemiptera: Miridae). Aust J Entomol 2010;49:160–5.
[18] Bielza P. Insecticide Resistance in Natural Enemies. In: Horowitz AR, Ishaaya I, editors. Advances in Insect Control and Resistance Management, Cham: Springer International Publishing; 2016, p. 313–29.
[19] Bailey E, Field L, Rawlings C, King R, Mohareb F, Pak K-H, et al. A scaffold-level genome assembly of the pirate bug, Orius laevigatus, and a comparative analysis of insecticide resistance-related gene families with hemipteran crop pests. Research Square 2021. https://doi.org/10.21203/rs.3.rs-537204/v1.
[20] Bailey E, Field L, Rawlings C, King R, Mohareb F, Hassani-Pak K, et al. A near-chromosome level genome assembly of the European hoverfly, Sphaerophoria rueppellii (Diptera: Syrphidae), and a comparative analysis of insecticide resistance-related gene families in hemipteran crop pests and pollinators. Research Square 2021. https://doi.org/10.21203/rs.3.rs-1050342/v1.
[21] Chen M, Mei Y, Chen X, Chen X, Xiao D, He K, et al. A chromosome-level assembly of the harlequin ladybird Harmonia axyridis as a genomic resource to study beetle and invasion biology. Mol Ecol Resour 2021;21:1318–32.
[22] Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. Genome Biol Evol 2016;8:1762–75.
[23] Wang Y, Zhang R, Wang M, Zhang L, Shi C-M, Li J, et al. The first chromosome-level genome assembly of a green lacewing Chrysopa pallens and its implication for biological control. Mol Ecol Resour 2021. https://doi.org/10.1111/1755-0998.13503.
[24] Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science 2010;327:343–8.
[25] Geib SM, Liang GH, Murphy TD, Sim SB. Whole Genome Sequencing of the Braconid Parasitoid Wasp Fopius arisanus, an Important Biocontrol Agent of Pest Tepritid Fruit Flies. G3 2017;7:2407–11.
[26] Xie W, Chen C, Yang Z, Guo L, Yang X, Wang D, et al. Genome sequencing of the sweetpotato whitefly Bemisia tabaci MED/Q. Gigascience 2017;6:1–7.
[27] International Aphid Genomics Consortium. Genome sequence of the pea aphid Acyrthosiphon pisum. PLoS Biol 2010;8:e1000313.
[28] Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, et al. The genome of Diuraphis noxia, a global aphid pest of small

grains. BMC Genomics 2015;16:429.
[29] Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 2011;479:487–92.
[30] Rotenberg D, Baumann AA, Ben-Mahmoud S, Christiaens O, Dermauw W, Ioannidis P, et al. Genome-enabled insights into the biology of thrips as crop pests. BMC Biol 2020;18:142.
[31] Xie W, He C, Fei Z, Zhang Y. Chromosome-level genome assembly of the greenhouse whitefly (Trialeurodes vaporariorum Westwood). Mol Ecol Resour 2020;20:995–1006.
[32] Kakumani PK, Malhotra P, Mukherjee SK, Bhatnagar RK. A draft genome assembly of the army worm, Spodoptera frugiperda. Genomics 2014;104:134–43.
[33] Heckel DG. Insecticide Resistance After Silent Spring. Science 2012;337:1612–4.
[34] Li X, Shi H, Gao X, Liang P. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in Plutella xylostella (L.). Pest Manag Sci 2018;74:695–704.
[35] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[36] Pavlidi N, Vontas J, Van Leeuwen T. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. Curr Opin Insect Sci 2018;27:97–102.
[37] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[38] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[39] Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Current Opinion in Insect Science 2019;31:131–8.
[40] Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. Evolution of supergene families associated with insecticide resistance. Science 2002;298:179–81.
[41] Bass C, Field LM. Gene amplification and insecticide resistance. Pest Manag Sci 2011;67:886–90.
[42] Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. PLoS Genet 2010;6:e1000999.
[43] Field LM, Devonshire AL, Forde BG. Molecular evidence that insecticide resistance in peach-potato aphids (Myzus persicae Sulz.) results from amplification of an esterase gene. Biochem J 1988;251:309–12.
[44] Field LM, Blackman RL, Tyler-Smith C, Devonshire AL. Relationship between amount of esterase and gene copy number in insecticide-resistant Myzus persicae (Sulzer). Biochem J 1999;339 (Pt 3):737–42.
[45] Hassan SA. Über die Massenzucht von Chrysopa carnea Steph. (Neuroptera, Chrysopidae). Z Angew Entomol 1975;79:310–5.
[46] T. R. NEW. The biology of Chrysopidae and Hemerobiidae (Neuroptera), with reference to their usage as biocontrol agents: a review. Trans R Entomol Soc Lond 1975;127:115–40.
[47] Stelzl M, Devetak D. Neuroptera in agricultural ecosystems. Agric Ecosyst Environ 1999;74:305–21.
[48] Canard M, Séméria Y, New TR. Biology of chrysopidae. vol. 27. Springer; 1984.
[49] NEW, T. R. Neuroptera. Aphids Their Biology, Natural Enemies and Control 1988:249–58.
[50] Duelli P, McEwen P, New TR, Whittington AE. Lacewings in the crop environment 2001.
[51] McEwen PK, New TR, Whittington AE. Lacewings in the crop environment. Cambridge University Press; 2007.
[52] Duelli P. Dispersal and opposition strategies in Chrysoperla carnea. Progress in world's neuropterology: proceedings of the 1st International Symposium on Neuropterology in Graz, Austria:(Insecta, Megaloptera, Raphidioptera, Planipennia), 1984.
[53] Chapman JW, Reynolds DR, Brooks SJ, Smith AD, Woiwod IP. Seasonal variation in the migration strategies of the green lacewingChrysoperla carneaspecies complex. Ecol Entomol 2006;31:378–88.
[54] Grafton-Cardwell EE. Intraspecific Variability in Response to Pesticides in the Common Green Lacewing, Chrysoperla Carnea (Stephens) (Neuroptera: Chrysopidae). California Agricultural Experiment Station; 1985.
[55] Plapp FW Jr, Bull DL. Toxicity and selectivity of some insecticides to Chrysopa carnea, a predator of the tobacco budworm. Environ Entomol 1978;7:431–4.
[56] Shour MH, Crowder LA. Effects of Pyrethroid Insecticides on the Common Green Lacewing12. J Econ Entomol 1980;73:306–9.
[57] Ishaaya I, Casida JE. Pyrethroid Esterase(s) May Contribute to Natural Pyrethroid Tolerance of Larvae of the Common Green Lacewing. Environ Entomol 1981;10:681–4.
[58] Lingren PD, Ridgway RL. Toxicity of five insecticides to several insect predators. J Econ Entomol 1967;60:1639–41.
[59] Grafton-Cardwell EE, Hoy MA. Genetic Improvement of Common Green Lacewing, Chrysoperla carnea (Neuroptera: Chrysopidae): Selection for Carbaryl Resistance. Environ Entomol 1986;15:1130–6.
[60] Pree DJ, Archibald DE, Morrison RK. Resistance to Insecticides in the Common Green Lacewing Chrysoperla carnea (Neuroptera: Chrysopidae) in Southern Ontario. J Econ Entomol 1989;82:29–34.
[61] Mansoor MM, Raza ABM, Abbas N, Aqueel MA, Afzal M. Resistance of green lacewing, Chrysoperla carnea Stephens to nitenpyram: Cross-resistance patterns, mechanism, stability, and realized heritability. Pestic Biochem Physiol 2017;135:59–63.
[62] Chambers RJ, Long S, Helyer NL. Effectiveness of Orius laevigatus (Hem.: Anthocoridae) for the control of Frankliniella occidentalis on cucumber and pepper in the UK. Biocontrol Sci Technol 1993;3:295–307.
[63] Riudavets J, Others. Predators of Frankliniella occidentalis (Perg.) and Thrips tabaci Lind.: a review. Wageningen Agricultural University Papers 1995:43–87.
[64] Alvarado P, Baltà O, Alomar O. Efficiency of four heteroptera as predators ofAphis gossypii andMacrosiphum euphorbiae (Hom.:

Aphididae). Entomophaga 1997;42:215–26.
[65] Aragón-Sánchez M, Román-Fernández LR, Martínez-García H, Aragón-García A, Pérez-Moreno I, Marco-Mancebón VS. Rate of consumption, biological parameters, and population growth capacity of Orius laevigatus fed on Spodoptera exigua. Biocontrol 2018;63:785–94.
[66] Arnó J, Roig J, Riudavets J. Evaluation of Orius majusculus and O. laevigatus as predators of Bemisa tabaci and estimation of their prey preference. Biol Control 2008;44:1–6.
[67] Tommasini MG. Evaluation of Orius species for biological control of Frankliniella occidentalis (Pergande)(Thysanoptera: Thripidae) 2003.
[68] van Lenteren JC, Tommasini MG. Mass Production, Storage, Shipment and Quality Control of Natural Enemies. In: Albajes R, Lodovica Gullino M, van Lenteren JC, Elad Y, editors. Integrated Pest and Disease Management in Greenhouse Crops, Dordrecht: Springer Netherlands; 1999, p. 276–94.
[69] Hogmire HW, Altman S, Baugher TA, Biddinger DJ, Biggs AR, Byers RE, et al. Mid-Atlantic Orchard Monitoring Guide (NRAES 75). Https://ecommons.cornell.edu › Handlehttps://ecommons.cornell.edu › Handle 1995.
[70] Vacante V, Cocuzza GE, Clercq P, Veire M, Tirry L. Development and survival ofOrius albidipennis andO. levigatus (Het.: Anthocoridae) on various diets. Entomophaga 1997;42:493–8.
[71] Cocuzza GE, De Clercq P, Van De Veire M, De Cock A, Vacante V. Reproduction of Orius laevigatus and Orius albidipennis on pollen and Ephestia kuehniella eggs. Entomol Exp Appl 1997;82:101–4.
[72] Oveja MF, Arnó J, Gabarra R, Others. Effect of supplemental food on the fitness of four omnivorous predator species. IOBC/WPRS Bull 2012;80:97–101.
[73] Bosco L, Tavella L. Distribution and abundance of species of the genus Orius in horticultural ecosystems of northwestern Italy n.d. https://www.cabi.org/isc/FullTextPDF/2013/20133416565.pdf (accessed November 12, 2021).
[74] Angeli G, Baldessari M, Maines R, Duso C. Side-effects of pesticides on the predatory bug Orius laevigatus (Heteroptera: Anthocoridae) in the laboratory. Biocontrol Sci Technol 2005;15:745–54.
[75] Delbeke F, Vercruysse P, Tirry L, Clercq P, Degheele D. Toxicity of diflubenzuron, pyriproxyfen, imidacloprid and diafenthiuron to the predatory bugOrius laevigatus (Het.: Anthocoridae). Entomophaga 1997;42:349–58. https://doi.org/ 10.1007/bf02769828.
[76] Van de Veire M, Tirry L. Side effects of pesticides on four species of beneficials used in IPM in glasshouse vegetable crops:" worst case" laboratory tests. IOBC WPRS BULLETIN 2003.
[77] Van de Veire M, Sterk G, van der Staaij M, Ramakers PMJ, Tirry L. Biocontrol 2002;47:101–13.
[78] Dáder B, Colomer I, Adán Á, Medina P, Viñuela E. Compatibility of early natural enemy introductions in commercial pepper and tomato greenhouses with repeated pesticide applications. Insect Sci 2020;27:1111–24.
[79] Rami Horowitz A, Ishaaya I. Advances in Insect Control and Resistance Management. Springer International Publishing; 2016.
[80] Balanza V, Mendoza JE, Bielza P. Variation in susceptibility and selection for resistance to imidacloprid and thiamethoxam in Mediterranean populations of Orius laevigatus. Entomol Exp Appl 2019;167:626–35.
[81] Balanza V, Mendoza JE, Cifuentes D, Bielza P. Selection for resistance to pyrethroids in the predator Orius laevigatus. Pest Manag Sci 2021;77:2539–46.
[82] Dunn L, Lequerica M, Reid CR, Latty T. Dual ecosystem services of syrphid flies (Diptera: Syrphidae): pollinators and biological control agents. Pest Manag Sci 2020;76:1973–9.
[83] Bellefeuille Y, Fournier M, Lucas E. Evaluation of Two Potential Biological Control Agents Against the Foxglove Aphid at Low Temperatures. J Insect Sci 2019;19. https://doi.org/ 10.1093/jisesa/iey130.
[84] Hopper JV, Nelson EH, Daane KM, Mills NJ. Growth, development and consumption by four syrphid species associated with the lettuce aphid, Nasonovia ribisnigri, in California. Biol Control 2011;58:271–6.
[85] Pekas A, De Craecker I, Boonen S, Wäckers FL, Moerkens R. One stone; two birds: concurrent pest control and pollination services provided by aphidophagous hoverflies. Biol Control 2020;149:104328.
[86] Amorós-Jiménez R, Pineda A, Fereres A, Marcos-García MÁ. Prey availability and abiotic requirements of immature stages of the aphid predator Sphaerophoria rueppellii. Biol Control 2012;63:17–24.
[87] Amorós-Jiménez R, Pineda A, Fereres A, Marcos-García MÁ. Feeding preferences of the aphidophagous hoverfly Sphaerophoria rueppellii affect the performance of its offspring. Biocontrol 2014;59:427–35.
[88] Pineda A, Marcos-García MÁ. Use of selected flowering plants in greenhouses to enhance aphidophagous hoverfly populations (Diptera: Syrphidae). Ann Soc Entomol Fr 2008;44:487–92.
[89] Niehoff B, Poehling H-M. Population dynamics of aphids and syrphid larvae in winter wheat treated with different rates of pirimicarb. Agric Ecosyst Environ 1995;52:51–5.
[90] Chambers RJ, Adams THL. Quantification of the Impact of Hoverflies (Diptera: Syrphidae) On Cereal Aphids in Winter Wheat: An Analysis of Field Populations. J Appl Ecol 1986;23:895–904.
[91] Tenhumberg B, Poehling H-M. Syrphids as natural enemies of cereal aphids in Germany: Aspects of their biology and efficacy in different years and regions. Agric Ecosyst Environ 1995;52:39–43.
[92] Wotton KR, Gao B, Menz MHM, Morris RKA, Ball SG, Lim KS, et al. Mass Seasonal Migrations of Hoverflies Provide Extensive Pollination and Crop Protection Services. Curr Biol 2019;29:2167–73.e5.
[93] Jauker F, Diekötter T, Schwarzbach F, Wolters V. Pollinator dispersal in an agricultural matrix: opposing responses of wild bees and hoverflies to landscape structure and distance from main habitat. Landsc Ecol 2009;24:547–55.
[94] Amorós-Jiménez R, Robert CAM, Marcos-García MÁ, Fereres A, Turlings TCJ. A Differential Role of Volatiles from Conspecific and Heterospecific Competitors in the Selection of Oviposition Sites by the Aphidophagous Hoverfly Sphaerophoria rueppellii. J Chem

Ecol 2015;41:493–500.
[95] Calvo-Agudo M, González-Cabrera J, Picó Y, Calatayud-Vernich P, Urbaneja A, Dicke M, et al. Neonicotinoids in excretion product of phloem-feeding insects kill beneficial insects. Proc Natl Acad Sci U S A 2019;116:16817–22.
[96] Calvo-Agudo M, González-Cabrera J, Sadutto D, Picó Y, Urbaneja A, Dicke M, et al. IPM-recommended insecticides harm beneficial insects through contaminated honeydew. Environ Pollut 2020;267:115581.
[97] Jordan A, Broad GR, Stigenberg J, Hughes J, Stone J, Bedford I, et al. The potential of the solitary parasitoid Microctonus brassicae for the biological control of the adult cabbage stem flea beetle, Psylliodes chrysocephala. Entomol Exp Appl 2020;168:360–70.
[98] Jordan A, Broad GR, Stigenberg J, Hughes J, Stone J, Bedford I, et al. The potential of the solitary parasitoid Microctonus brassicae for the biological control of the adult cabbage stem flea beetle, Psylliodes chrysocephala. Entomol Exp Appl 2020;168:360–70.
[99] Bartlet E, Williams IH. Factors restricting the feeding of the cabbage stem flea beetle (Psylliodes chrysocephala). Entomol Exp Appl 1991;60:233–8.
[100] Højland DH, Nauen R, Foster SP, Williamson MS, Kristensen M. Incidence, Spread and Mechanisms of Pyrethroid Resistance in European Populations of the Cabbage Stem Flea Beetle, Psylliodes chrysocephala L. (Coleoptera: Chrysomelidae). PLoS One 2015;10:e0146045.
[101] Robert C, Ruck L, Carpezat J, Lauvernay A, Siegwart M, Leflon M, et al. Pyrethroids resistance monitoring in French cabbage stem flea beetle (Psylliodes chrysocephala) and rape winter stem weevil (Ceutorhynchus picitarsis) populations in oilseed rape. Écologie Chimique: nouvelles contributions à la protection des cultures contre les ravageurs et 11e Conférence Internationale sur les Ravageurs et Auxiliaires en Agriculture, 24 au 26 octobre 2017, Montpellier, France, Association Française de Protection des Plantes (AFPP); 2017, p. 196–208.
[102] Stará J, Kocourek F. Cabbage stem flea beetle's (Psylliodes chrysocephala L.) susceptibility to pyrethroids and tolerance to thiacloprid in the Czech Republic. PLoS One 2019;14:e0214702.
[103] Willis CE, Foster SP, Zimmer CT, Elias J, Chang X, Field LM, et al. Investigating the status of pyrethroid resistance in UK populations of the cabbage stem flea beetle (Psylliodes chrysocephala). Crop Prot 2020;138:105316.
[104] Zhang X-R, Zhang J-Q, Shao Y-Y, Xing X-R, Wang J, Liu Z-X, et al. Identification of glutathione-S-transferase genes by transcriptome analysis in Meteorus pulchricornis (Hymenoptera: Braconidae) and their expression patterns under stress of phoxim and cypermethrin. Comp Biochem Physiol Part D Genomics Proteomics 2019;31:100607.
[105] Iqbal M, Wright DJ. Host resistance to insecticides can confer protection to endo-larval parasitoids. Bull Entomol Res 1996;86:721–3.
[106] Koch RL. The multicolored Asian lady beetle, Harmonia axyridis: a review of its biology, uses in biological control, and non-target impacts. J Insect Sci 2003;3:32.
[107] Roy H, Wajnberg E. From biological control to invasion: the ladybird Harmonia axyridis as a model species. Biocontrol 2008;53:1–4.
[108] Gordon RD. Journal of the New York Entomological Society. The Coccinellidae (Coleoptera) of America North of Mexico 1985;93:1–912.
[109] Hodek I, Iperti G, Hodkova M, Others. Long-distance flights in Coccinellidae (Coleoptera). Eur J Entomol 1993;90:403–14.
[110] Flaherty DL, Huffaker CB, Others. Biological control of Pacific mites and Willamette mites in San Joaquin Valley vineyards. I. Role of Metaseiulus occidentalis. Hilgardia 1970;40.
[111] Flaherty DL, Huffaker CB, Others. Biological control of Pacific mites and Willamette mites in San Joaquin Valley vineyards. II. Influence of dispersion patterns of Metaseiulus occidentalis. Hilgardia 1970.
[112] Hoyt SC. Integrated chemical control of insects and biological control of mites on apple in Washington. J Econ Entomol 1969;62:74–86.
[113] Hoy MA, Knop NF. Studies on pesticide resistance in the phytoseiid Metaseiulus occidentalis in California. Recent Advances in Acarology 1979;1:89–94.
[114] Hoy MA, Flaherty D, Peacock W, Culver D. Vineyard and Laboratory Evaluations of Methomyl, Dimethoate, and Permethrin for a Grape Pest Management Program in the San Joaquin Valley of California. J Econ Entomol 1979;72:250–5.
[115] Roush RT, Hoy MA. Genetic improvement of Metaseiulus occidentalis: selection with methomyl, dimethoate, and carbaryl and genetic analysis of carbaryl resistance. J Econ Entomol 1981;74:138–41.
[116] Hoy MA, Knop NF. Selection for and genetic analysis of permethrin resistance in Metaseiulus occidentals: Genetic improvement of a biological control agent. Entomol Exp Appl 1981;30:10–8.
[117] Roush RT, Hoy MA. Laboratory, glasshouse, and field studies of artificially selected carbaryl resistance in Metaseiulus occidentalis. J Econ Entomol 1981;74:142–7.
[118] Hoy MA. Aerial dispersal and field efficacy of a genetically improved strain of the spider mite predator Metaseiulus occidentals. Entomol Exp Appl 1982;32:205–12.
[119] Hoy M, Barnett W, Reil W, Castro D, Cahn D, Hendricks L, et al. Large-scale releases of pesticide-resistant spider mite predators. Calif Agric 1982;36:8–10.
[120] Hoy MA, Westigard PH, Hoyt SC. Release and evaluation of a laboratory-selected, pyrethroid-resistant strain of the predaceous mite Metaseiulus occidentalis (Acari: Phytoseiidae) in southern Oregon pear orchards and a Washington apple orchard. J Econ Entomol 1983;76:383–8.
[121] Hoy MA, Castro D, Cahn D. Two methods for large scale production of pesticide-resistant strains of the spider mite predator Metaseiulus occidentalis (Nesbitt)(Acurina, Phytoseiidae). Z Angew Entomol 1982;1:1–9.
[122] van Lenteren JC. The state of commercial augmentative biological control: plenty of natural enemies, but a frustrating lack of

uptake. Biocontrol 2012;57:1–20.
[123] Gauthier J, Boulain H, van Vugt JJFA, Baudry L, Persyn E, Aury J-M, et al. Chromosomal scale assembly of parasitic wasp genome reveals symbiotic virus colonization. Commun Biol 2021;4:104.
[124] Ye X, Yan Z, Yang Y, Xiao S, Chen L, Wang J, et al. A chromosome-level genome assembly of the parasitoid wasp Pteromalus puparum. Mol Ecol Resour 2020;20:1384–402.
[125] Shi M, Wang Z, Ye X, Xie H, Li F, Hu X, et al. The genomes of two parasitic wasps that parasitize the diamondback moth. BMC Genomics 2019;20:893.
[126] Tvedte ES, Walden KKO, McElroy KE, Werren JH, Forbes AA, Hood GR, et al. Genome of the Parasitoid Wasp Diachasma alloeum, an Emerging Model for Ecological Speciation and Transitions to Asexual Reproduction. Genome Biol Evol 2019;11:2767–73.
[127] Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MR. Whole Genome Sequence of the Parasitoid Wasp Microplitis demolitor That Harbors an Endogenous Virus Mutualist. G3 2018;8:2875–80.
[128] Van Leeuwen T, Vontas J, Tsagkarakou A, Tirry L. Mechanisms of Acaricide Resistance in the Two-Spotted Spider Mite Tetranychus urticae. In: Ishaaya I, Horowitz AR, editors. Biorational Control of Arthropod Pests: Application and Resistance Management, Dordrecht: Springer Netherlands; 2009, p. 347–93.
[129] Migeon A, Nouguier E, Dorkeld F. Spider Mites Web: A comprehensive database for the Tetranychidae. Trends in Acarology, Springer Netherlands; 2010, p. 557–60.
[130] Ay R, Gürkan MO. Resistance to bifenthrin and resistance mechanisms of different strains of the two-spotted spider mite (Tetranychus urticae) from Turkey. Phytoparasitica 2005;33:237–44.
[131] Van Leeuwen T, Van Pottelberge S, Tirry L. Comparative acaricide susceptibility and detoxifying enzyme activities in field-collected resistant and susceptible strains of Tetranychus urticae. Pest Manag Sci 2005;61:499–507.
[132] Van Leeuwen T, Tirry L. Esterase-mediated bifenthrin resistance in a multiresistant strain of the two-spotted spider mite, Tetranychus urticae. Pest Manag Sci 2007;63:150–6.
[133] Tsagkarakou A, Van Leeuwen T, Khajehali J, Ilias A, Grispou M, Williamson MS, et al. Identification of pyrethroid resistance associated mutations in the para sodium channel of the two-spotted spider mite Tetranychus urticae (Acari: Tetranychidae). Insect Mol Biol 2009;18:583–93.
[134] Khajehali J, Van Leeuwen T, Grispou M, Morou E, Alout H, Weill M, et al. Acetylcholinesterase point mutations in European strains of Tetranychus urticae (Acari: Tetranychidae) resistant to organophosphates. Pest Manag Sci 2010;66:220–8.
[135] Kwon DH, Im JS, Ahn JJ, Lee J-H, Marshall Clark J, Lee SH. Acetylcholinesterase point mutations putatively associated with monocrotophos resistance in the two-spotted spider mite. Pestic Biochem Physiol 2010;96:36–42.
[136] Kwon DH, Clark JM, Lee SH. Extensive gene duplication of acetylcholinesterase associated with organophosphate resistance in the two-spotted spider mite. Insect Mol Biol 2010;19:195–204.
[137] Pathak MD, Khan ZR. Rice leafhoppers and planthoppers. Insect Pests of Rice 1994:19–27.
[138] Riley JR, Xia-Nian C, Xiao-Xi Z, Reynolds DR, Guo-Min XU, Smith AD, et al. The long-distance migration of Nilaparvata lugens (Stål) (Delphacidae) in China: radar observations of mass return flight in the autumn. Ecol Entomol 1991;16:471–89.
[139] Zhang X, Liu X, Zhu F, Li J, You H, Lu P. Field evolution of insecticide resistance in the brown planthopper (Nilaparvata lugens Stål) in China. Crop Prot 2014;58:61–6.
[140] Sun H, Yang B, Zhang Y, Liu Z. Metabolic resistance in Nilaparvata lugens to etofenprox, a non-ester pyrethroid insecticide. Pestic Biochem Physiol 2017;136:23–8.
[141] Ding Z, Wen Y, Yang B, Zhang Y, Liu S, Liu Z, et al. Biochemical mechanisms of imidacloprid resistance in Nilaparvata lugens: over-expression of cytochrome P450 CYP6AY1. Insect Biochem Mol Biol 2013;43:1021–7.
[142] Liu Z, Williamson MS, Lansdell SJ, Denholm I, Han Z, Millar NS. A nicotinic acetylcholine receptor mutation conferring target-site resistance to imidacloprid in Nilaparvata lugens (brown planthopper). Proc Natl Acad Sci U S A 2005;102:8420–5.
[143] Kwon DH, Cha DJ, Kim YH, Lee SW, Lee SH. Cloning of the acetylcholinesterase 1 gene and identification of point mutations putatively associated with carbofuran resistance in Nilaparvata lugens. Pestic Biochem Physiol 2012;103:94–100.
[144] Zhang Y, Yang B, Li J, Liu M, Liu Z. Point mutations in acetylcholinesterase 1 associated with chlorpyrifos resistance in the brown planthopper, Nilaparvata lugens Stål. Insect Mol Biol 2017;26:453–60.
[145] Small GJ, Hemingway J. Molecular characterization of the amplified carboxylesterase gene associated with organophosphorus insecticide resistance in the brown planthopper, Nilaparvata lugens. Insect Mol Biol 2000;9:647–53.
[146] Blackman RL. Morphological discrimination of a tobacco-feeding form from Myzus persicae (Sulzer) (Hemiptera: Aphididae), and a key to New World Myzus (Nectarosiphon) species. Bull Entomol Res 1987;77:713–30.
[147] Blackman RL. Life-cycle variation of Myzus persicae (Sulz.) (Hom., Aphididae) in different parts of the world, in relation to genotype and environment. Bulletin of Entomological Research 1974;63:595–607. https://doi.org/10.1017/s0007485300047830.
[148] Nabeshima T, Kozaki T, Tomita T, Kono Y. An amino acid substitution on the second acetylcholinesterase in the pirimicarb-resistant strains of the peach potato aphid, Myzus persicae. Biochem Biophys Res Commun 2003;307:15–22.
[149] Bass C, Puinean AM, Andrews M, Cutler P, Daniels M, Elias J, et al. Mutation of a nicotinic acetylcholine receptor β subunit is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. BMC Neurosci 2011;12:51.
[150] Martinez-Torres D, Foster SP, Field LM, Devonshire AL, Williamson MS. A sodium channel point mutation is associated with resistance to DDT and pyrethroid insecticides in the peach-potato aphid, Myzus persicae (Sulzer) (Hemiptera: Aphididae). Insect Mol Biol 1999;8:339–46.
[151] Anstead JA, Williamson MS, Denholm I. Evidence for multiple origins of identical insecticide resistance mutations in the aphid Myzus persicae. Insect Biochem Mol Biol 2005;35:249–56.

[152] Komazaki S, Toda S. Differences in Host Preference, Life Cycle Pattern, and Insecticide Susceptibility Among Aphis gossypii Clones and Genetic Relationships Inferred from Internal Transcribed Spacer 2 Sequences of rDNA. Ann Entomol Soc Am 2008;101:565–72.
[153] Brévault T, Carletto J, Tribot J, Vanlerberghe-Masutti F. Insecticide use and competition shape the genetic diversity of the aphid Aphis gossypii in a cotton-growing landscape. Bull Entomol Res 2011;101:407–13.
[154] Zhang GX, Zhong TS. Study on life history of several species of aphids. Sinozoologia 1982;2:7–16.
[155] Li F, Han Z. Mutations in acetylcholinesterase associated with insecticide resistance in the cotton aphid, Aphis gossypii Glover. Insect Biochem Mol Biol 2004;34:397–405.
[156] Toda S, Komazaki S, Tomita T, Kono Y. Two amino acid substitutions in acetylcholinesterase associated with pirimicarb and organophosphorous insecticide resistance in the cotton aphid, Aphis gossypii Glover (Homoptera: Aphididae). Insect Mol Biol 2004;13:549–53.
[157] Cao C-W, Zhang J, Gao X-W, Liang P, Guo H-L. Overexpression of carboxylesterase gene associated with organophosphorous insecticide resistance in cotton aphids, Aphis gossypii (Glover). Pestic Biochem Physiol 2008;90:175–80.
[158] Chen X, Tang C, Ma K, Xia J, Song D, Gao X-W. Overexpression of UDP-glycosyltransferase potentially involved in insecticide resistance in Aphis gossypii Glover collected from Bt cotton fields in China. Pest Manag Sci 2020;76:1371–7.
[159] Inbar M, Gerling D. Plant-mediated interactions between whiteflies, herbivores, and natural enemies. Annu Rev Entomol 2008;53:431–48.
[160] Karatolos N, Williamson MS, Denholm I, Gorman K, Ffrench-Constant RH, Bass C. Over-expression of a cytochrome P450 is associated with resistance to pyriproxyfen in the greenhouse whitefly Trialeurodes vaporariorum. PLoS One 2012;7:e31077.
[161] Karatolos N, Gorman K, Williamson MS, Denholm I. Mutations in the sodium channel associated with pyrethroid resistance in the greenhouse whitefly, Trialeurodes vaporariorum. Pest Manag Sci 2012;68:834–8.
[162] Pappas ML, Migkou F, Broufas GD. Incidence of resistance to neonicotinoid insecticides in greenhouse populations of the whitefly, Trialeurodesvaporariorum (Hemiptera: Aleyrodidae) from Greece. Appl Entomol Zool 2013;48:373–8.
[163] Sani I, Ismail SI, Abdullah S, Jalinas J, Jamian S, Saad N. A Review of the Biology and Control of Whitefly, Bemisia tabaci (Hemiptera: Aleyrodidae), with Special Reference to Biological Control Using Entomopathogenic Fungi. Insects 2020;11. https://doi.org/10.3390/insects11090619.
[164] Alon M, Alon F, Nauen R, Morin S. Organophosphates' resistance in the B-biotype of Bemisia tabaci (Hemiptera: Aleyrodidae) is associated with a point mutation in an ace1-type acetylcholinesterase and overexpression of carboxylesterase. Insect Biochem Mol Biol 2008;38:940–9.
[165] Roditakis E, Tsagkarakou A, Vontas J. Identification of mutations in the para sodium channel of Bemisia tabaci from Crete, associated with resistance to pyrethroids. Pestic Biochem Physiol 2006;85:161–6.
[166] Karunker I, Morou E, Nikou D, Nauen R, Sertchook R, Stevenson BJ, et al. Structural model and functional characterization of the Bemisia tabaci CYP6CM1vQ, a cytochrome P450 associated with high levels of imidacloprid resistance. Insect Biochem Mol Biol 2009;39:697–706.
[167] Walker AK, Others. A review of the pest status and natural enemies of Thrips palmi. Biocontrol News and Information 1994;15.
[168] Herron GA, Rophail J, Gullick GC. Laboratory-Based, Insecticide Efficacy Studies on Field-Collected Frankliniella occidentalis (Pergande) (Thysanoptera: Thripidae) and Implications for its Management in Australia. Aust J Entomol 1996;35:161–4.
[169] Bao WX, Narai Y, Nakano A, Kaneda T, Murai T, Sonoda S. Spinosad resistance of melon thrips, Thrips palmi, is conferred by G275E mutation in α6 subunit of nicotinic acetylcholine receptor and cytochrome P450 detoxification. Pestic Biochem Physiol 2014;112:51–5.
[170] Bao WX, Sonoda S. Resistance to cypermethrin in melon thrips, Thrips palmi (Thysanoptera: Thripidae), is conferred by reduced sensitivity of the sodium channel and CYP450-mediated detoxification. Appl Entomol Zool 2012;47:443–8.
[171] Bao WX, Kataoka Y, Fukada K, Sonoda S. Imidacloprid resistance of melon thrips, *Thrips palmi*, is conferred by CYP450-mediated detoxification. J Pestic Sci 2015;40:65–8.
[172] Jensen SE. Mechanisms associated with methiocarb resistance in Frankliniella occidentalis (Thysanoptera: Thripidae). J Econ Entomol 2000;93:464–71.
[173] Espinosa PJ, Contreras J, Quinto V, Grávalos C, Fernández E, Bielza P. Metabolic mechanisms of insecticide resistance in the western flower thrips, Frankliniella occidentalis (Pergande). Pest Manag Sci 2005;61:1009–15.
[174] Nolte H-W. Käfer bedrohen den Raps. Ziemsen; 1954.
[175] Højland DH, Nauen R, Foster SP, Williamson MS, Kristensen M. Incidence, Spread and Mechanisms of Pyrethroid Resistance in European Populations of the Cabbage Stem Flea Beetle, Psylliodes chrysocephala L. (Coleoptera: Chrysomelidae). PLoS One 2015;10:e0146045.
[176] Zimmer CT, Müller A, Heimbach U, Nauen R. Target-site resistance to pyrethroid insecticides in German populations of the cabbage stem flea beetle, Psylliodes chrysocephala L. (Coleoptera: Chrysomelidae). Pestic Biochem Physiol 2014;108:1–7.
[177] Willis CE, Foster SP, Zimmer CT, Elias J, Chang X, Field LM, et al. Investigating the status of pyrethroid resistance in UK populations of the cabbage stem flea beetle (Psylliodes chrysocephala). Crop Prot 2020;138:105316.
[178] Li F, Zhao X, Li M, He K, Huang C, Zhou Y, et al. Insect genomes: progress and challenges. Insect Mol Biol 2019;28:739–58.
[179] 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.
[180] 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.

[181] Richards S, Murali SC. Best Practices in Insect Genome Sequencing: What Works and What Doesn't. Curr Opin Insect Sci 2015;7:1–7.
[182] Simão 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.
[183] Ramsey JS, Wilson ACC, de Vos M, Sun Q, Tamborindeguy C, Winfield A, et al. Genomic resources for Myzus persicae: EST sequencing, SNP identification, and microarray design. BMC Genomics 2007;8:423.
[184] Pan Y, Zeng X, Wen S, Gao X, Liu X, Tian F, et al. Multiple ATP-binding cassette transporters genes are involved in thiamethoxam resistance in Aphis gossypii glover. Pestic Biochem Physiol 2020;167:104558.
[185] Quan Q, Hu X, Pan B, Zeng B, Wu N, Fang G, et al. Draft genome of the cotton aphid Aphis gossypii. Insect Biochem Mol Biol 2019;105:25–32.
[186] Pym A, Singh KS, Nordgren Å, Emyr Davies TG, Zimmer CT, Elias J, et al. Host plant adaptation in the polyphagous whitefly, Trialeurodes vaporariorum , is associated with transcriptional plasticity and altered sensitivity to insecticides. BMC Genomics 2019;20:1–19.
[187] Aidlin Harari O, Santos-Garcia D, Musseri M, Moshitzky P, Patel M, Visendi P, et al. Molecular Evolution of the Glutathione S-Transferase Family in the Bemisia tabaci Species Complex. Genome Biol Evol 2020;12:3857–72.
[188] Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol 2016;14:1–15.
[189] Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, Bemisia tabaci. BMC Genomics 2017;18:1–16.
[190] Ilias A, Lagnel J, Kapantaidaki DE, Roditakis E, Tsigenopoulos CS, Vontas J, et al. Transcription analysis of neonicotinoid resistance in Mediterranean (MED) populations of B. tabaci reveal novel cytochrome P450s, but no nAChR mutations associated with the phenotype. BMC Genomics 2015;16:939.
[191] Xia J, Xu H, Yang Z, Pan H, Yang X, Guo Z, et al. Genome-Wide Analysis of Carboxylesterases (COEs) in the Whitefly, (Gennadius). Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20204973.
[192] Xue J, Zhou X, Zhang C-X, Yu L-L, Fan H-W, Wang Z, et al. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. Genome Biol 2014;15:521.
[193] Heckel DG. Perspectives on gene copy number variation and pesticide resistance. Pest Manag Sci 2021. https://doi.org/10.1002/ps.6631.
[194] Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 2011;12:1–14.
[195] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357–60.
[196] Yu L, Tang W, He W, Ma X, Vasseur L, Baxter SW, et al. Characterization and expression of the cytochrome P450 gene family in diamondback moth, Plutella xylostella (L.). Sci Rep 2015;5:8952.
[197] Ai J, Zhu Y, Duan J, Yu Q, Zhang G, Wan F, et al. Genome-wide analysis of cytochrome P450 monooxygenase genes in the silkworm, Bombyx mori. Gene 2011;480:42–50.
[198] Darwin Tree of Life – Reading the genomes of all life: a new platform for understanding our biodiversity n.d. https://www.darwintreeoflife.org/ (accessed July 22, 2021).
[199] Korf I. Gene finding in novel genomes. BMC Bioinformatics 2004;5:59.
[200] Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 2004;32:W309–12.
[201] Merzendorfer H. Chapter One - ABC Transporters and Their Role in Protecting Insects from Pesticides and Their Metabolites. In: Cohen E, editor. Advances in Insect Physiology, vol. 46, Academic Press; 2014, p. 1–72.
[202] Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol 2014;45:89–110.
[203] Feyereisen R. Insect P450 enzymes. Annu Rev Entomol 1999;44:507–33.
[204] Feyereisen R. Evolution of insect P450. Biochem Soc Trans 2006;34:1252–5.
[205] Karunker I, Benting J, Lueke B, Ponge T, Nauen R, Roditakis E, et al. Over-expression of cytochrome P450 CYP6CM1 is associated with high resistance to imidacloprid in the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). Insect Biochem Mol Biol 2008;38:634–44.
[206] Liang X, Xiao D, He Y, Yao J, Zhu G, Zhu KY. Insecticide-mediated up-regulation of cytochrome P450 genes in the red flour beetle (Tribolium castaneum). Int J Mol Sci 2015;16:2078–98.
[207] Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. PLoS Genet 2010;6:e1000999.
[208] Yang T, Liu N. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, Culex quinquefasciatus. PLoS One 2011;6:e29418.
[209] Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol 1999;29:757–77.
[210] Main BJ, Everitt A, Cornel AJ, Hormozdiari F, Lanzaro GC. Genetic variation associated with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii. Parasit Vectors 2018;11:225.
[211] Vlogiannitis S, Mavridis K, Dermauw W, Snoeck S, Katsavou E, Morou E, et al. Reduced proinsecticide activation by cytochrome P450 confers coumaphos resistance in the major bee parasite Varroa destructor. Proc Natl Acad Sci U S A 2021;118.

https://doi.org/10.1073/pnas.2020380118.
[212] Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. Toxicol Lett 2002;128:215–28.
[213] Minks, K. A, Harrewijn, P. Aphids : their biology, natural enemies, and control. Amsterdam; New York: Elsevier; 1987.
[214] Field LM, Williamson MS, Moores GD, Devonshire AL. Cloning and analysis of the esterase genes conferring insecticide resistance in the peach-potato aphid, Myzus persicae (Sulzer). Biochem J 1993;294 (Pt 2):569–74.
[215] Chen W-L, Sun C-N. Purification and characterization of carboxylesterases of a rice brown planthopper, Nilaparvata lugens Stål. Insect Biochem Mol Biol 1994;24:347–55.
[216] Karunaratne SHPP, Small GJ, Hemingway J. Characterization of the elevated esterase-associated insecticide resistance mechanism in Nilaparvata lugens (Stal) and other planthopper species. Int J Pest Manage 1999;45:225–30.
[217] Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol 2000;45:371–91.
[218] Bisset JA, Marín R, Rodríguez MM, Severson DW, Ricardo Y, French L, et al. Insecticide resistance in two Aedes aegypti (Diptera: Culicidae) strains from Costa Rica. J Med Entomol 2013;50:352–61.
[219] Ranson H, Jensen B, Wang X, Prapanthadara L., Hemingway J, Collins FH. Genetic mapping of two loci affecting DDT resistance in the malaria vector Anopheles gambiae. Insect Mol Biol 2000;9:499–507.
[220] Ranson H, Prapanthadara L-A, Hemingway J. Cloning and characterization of two glutathione S-transferases from a DDT-resistant strain of Anopheles gambiae. Biochemical Journal 1997;324:97–102. https://doi.org/ 10.1042/bj3240097.
[221] Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. Genome Biol 2014;15:R27.
[222] Lumjuan N, Rajatileka S, Changsom D, Wicheer J, Leelapat P, Prapanthadara L-A, et al. The role of the Aedes aegypti Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. Insect Biochem Mol Biol 2011;41:203–9.
[223] Vontas JG, Small GJ, Hemingway J. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens. Biochem J 2001;357:65–72.
[224] Tu C-PD, Akgül B. Drosophila glutathione S-transferases. Methods Enzymol 2005;401:204–26.
[225] Yamamoto K, Zhang P, Miake F, Kashige N, Aso Y, Banno Y, et al. Cloning, expression and characterization of theta-class glutathione S-transferase from the silkworm, Bombyx mori. Comp Biochem Physiol B Biochem Mol Biol 2005;141:340–6.
[226] Hassan F, Singh KP, Ali V, Behera S, Shivam P, Das P, et al. Detection and functional characterization of sigma class GST in Phlebotomus argentipes and its role in stress tolerance and DDT resistance. Sci Rep 2019;9:19636.
[227] Cheng J, Wang C-Y, Lyu Z-H, Lin T. Multiple Glutathione S-Transferase Genes in Heortia vitessoides (Lepidoptera: Crambidae): Identification and Expression Patterns. J Insect Sci 2018;18. https://doi.org/ 10.1093/jisesa/iey064.
[228] Wang YH, Gu ZY, Wang JM, Sun SS, Wang BB, Jin YQ, et al. Changes in the activity and the expression of detoxification enzymes in silkworms (Bombyx mori) after phoxim feeding. Pestic Biochem Physiol 2013;105:13–7.
[229] Vontas J, Blass C, Koutsos AC, David J-P, Kafatos FC, Louis C, et al. Gene expression in insecticide resistant and susceptible Anopheles gambiae strains constitutively or after insecticide exposure. Insect Mol Biol 2005;14:509–21.
[230] Yan M-W, Xing X-R, Wu F-A, Wang J, Sheng S. UDP-glycosyltransferases contribute to the tolerance of parasitoid wasps towards insecticides. Pestic Biochem Physiol 2021:104967.
[231] Silva AX, Bacigalupe LD, Luna-Rudloff M, Figueroa CC. Insecticide resistance mechanisms in the green peach aphid Myzus persicae (Hemiptera: Aphididae) II: Costs and benefits. PLoS One 2012;7:e36810.
[232] Zhao J, Xu L, Sun Y, Song P, Han Z. UDP-Glycosyltransferase Genes in the Striped Rice Stem Borer, (Walker), and Their Contribution to Chlorantraniliprole Resistance. Int J Mol Sci 2019;20. https://doi.org/ 10.3390/ijms20051064.
[233] Tian F, Wang Z, Li C, Liu J, Zeng X. UDP-Glycosyltransferases are involved in imidacloprid resistance in the Asian citrus psyllid, Diaphorina citri (Hemiptera: Lividae). Pestic Biochem Physiol 2019;154:23–31.
[234] Chen X, Tang C, Ma K, Xia J, Song D, Gao X-W. Overexpression of UDP-glycosyltransferase potentially involved in insecticide resistance in Aphis gossypii Glover collected from Bt cotton fields in China. Pest Manag Sci 2020;76:1371–7.
[235] Cui X, Wang C, Wang X, Li G, Liu Z, Wang H, et al. Molecular Mechanism of the UDP-Glucuronosyltransferase 2B20-like Gene (AccUGT2B20-like) in Pesticide Resistance of Apis cerana cerana. Front Genet 2020;11:592595.
[236] Devonshire AL, Field LM. Gene amplification and insecticide resistance. Annu Rev Entomol 1991;36:1–23.
[237] Heckel DG. Insect detoxification and sequestration strategies. Annual Plant Reviews Online 2018:77–114. https://doi.org/10.1002/9781119312994.apr0507.
[238] Palumbi SR. Humans as the world's greatest evolutionary force. Science 2001;293:1786–90.
[239] Robert CA, Zhang X, Machado RA, Schirmer S, Lori M, Mateo P, et al. Sequestration and activation of plant toxins protect the western corn rootworm from enemies at multiple trophic levels. Elife 2017;6. https://doi.org/ 10.7554/eLife.29307.
[240] Irwin RE, Cook D, Richardson LL, Manson JS, Gardner DR. Secondary compounds in floral rewards of toxic rangeland plants: impacts on pollinators. J Agric Food Chem 2014;62:7335–44.
[241] Lee SH, Kwon DH. Gene duplication in insecticide resistance. Gene Duplication 2011.
[242] Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 2010;11:97–108.
[243] Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. Genome Biol 2002;3:RESEARCH0008.
[244] Francino MP, Pilar Francino M. An adaptive radiation model for the origin of new gene functions. Nature Genetics 2005;37:573–8. https://doi.org/10.1038/ng1579.

[245] Hawkins NJ, Bass C, Dixon A, Neve P. The evolutionary origins of pesticide resistance. Biol Rev Camb Philos Soc 2018. https://doi.org/10.1111/brv.12440.
[246] Makino T, Kawata M. Habitat variability correlates with duplicate content of Drosophila genomes. Mol Biol Evol 2012;29:3169–79.
[247] Tamate SC, Kawata M, Makino T. Contribution of nonohnologous duplicated genes to high habitat variability in mammals. Mol Biol Evol 2014;31:1779–86.
[248] Albalat R, Cañestro C. Evolution by gene loss. Nat Rev Genet 2016;17:379–91.
[249] Schmehl DR, Teal PEA, Frazier JL, Grozinger CM. Genomic analysis of the interaction between pesticide exposure and nutrition in honey bees (Apis mellifera). J Insect Physiol 2014;71:177–90.
[250] Peterson MA, Denno RF, Robinson L. Apparent widespread gene flow in the predominantly flightless planthopperTumidagena minuta. Ecol Entomol 2001;26:629–37.
[251] Caprio MA, Tabashnik BE. Gene flow accelerates local adaptation among finite populations: Simulating the evolution of insecticide resistance. J Econ Entomol 1992;85:611–20.
[252] Raymond L, Plantegenest M, Vialatte A. Migration and dispersal may drive to high genetic variation and significant genetic mixing: the case of two agriculturally important, continental hoverflies (Episyrphus balteatus and Sphaerophoria scripta). Mol Ecol 2013;22:5329–39.
[253] Tulisalo U. Adult Angoumois Grain Moths Sitotroga Cerealella Oliv. as a Food Source for Larvae of the Green Lacewing Chrysopa Carnea Steph. in Mass Rearing. Maatalouden tutkimuskeskus; 1977.
[254] Wang R, Nordlund DA. Use of Chrysoperla spp.(Neuroptera: Chrysopidae) in augmentative release programmes for control of arthropod pests. Biocontrol News and Information (United Kingdom) 1994.
[255] Bellows TS, Fisher TW. Handbook of biological controlprinciples and applications of biological control 1999.
[256] Hogervorst PAM, Wäckers FL, Carette A-C, Romeis J. The importance of honeydew as food for larvae of Chrysoperla carnea in the presence of aphids. J Appl Entomol 2008;132:18–25.
[257] Zheng Y, Daane KM, Hagen KS, Mittler TE. Influence of larval food consumption on the fecundity of the lacewing Chrysoperla carnea. Entomol Exp Appl 1993;67:9–14.
[258] Gavrilets S, Hastings A. Founder Effect Speciation: A Theoretical Reassessment. Am Nat 1996;147:466–91.
[259] Paspati A, Ferguson KB, Verhulst EC, Urbaneja A, González-Cabrera J, Pannebakker BA. Effect of mass rearing on the genetic diversity of the predatory mite Amblyseius swirskii. Entomol Exp Appl 2019;167:670–81.
[260] Parrella MP, Trumble JT. Decline of Resistance in Liriomyza trifolii (Diptera: Agromyzidae) in the Absence of Insecticide Selection Pressure. J Econ Entomol 1989;82:365–8.
[261] Vera-Maloof FZ, Saavedra-Rodriguez K, Penilla-Navarro RP, D Rodriguez-Ramirez A, Dzul F, Manrique-Saide P, et al. Loss of pyrethroid resistance in newly established laboratory colonies of Aedes aegypti. PLoS Negl Trop Dis 2020;14:e0007753.
[262] Keil CB, Parrella MP. Characterization of Insecticide Resistance in Two Colonies of Liriomyza trifolii (Diptera: Agromyzidae). J Econ Entomol 1990;83:18–26.
[263] Kliot A, Ghanim M. Fitness costs associated with insecticide resistance. Pest Manag Sci 2012;68:1431–7.
[264] Raymond M, Poulin E, Boiroux V, Dupont E, Pasteur N. Stability of insecticide resistance due to amplification of esterase genes in Culex pipiens. Heredity 1993;70:301–7.
[265] Gerth M, Wolf R, Bleidorn C, Richter J, Sontowski R, Unrein J, et al. Green lacewings (Neuroptera: Chrysopidae) are commonly associated with a diversity of rickettsial endosymbionts. Zoological Lett 2017;3:12.
[266] Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, et al. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. Science 2011;332:254–6.
[267] Cass BN, Himler AG, Bondy EC, Bergen JE, Fung SK, Kelly SE, et al. Conditional fitness benefits of the Rickettsia bacterial symbiont in an insect pest. Oecologia 2016;180:169–79.
[268] Khan Pathan A, Sayyed AH, Aslam M, Razaq M, Jilani G, Saleem MA. Evidence of Field-Evolved Resistance to Organophosphates and Pyrethroids in Chrysoperla carnea (Neuroptera: Chrysopidae). J Econ Entomol 2008;101:1676–84.
[269] Zimmer CT, Garrood WT, Singh KS, Randall E, Lueke B, Gutbrod O, et al. Neofunctionalization of Duplicated P450 Genes Drives the Evolution of Insecticide Resistance in the Brown Planthopper. Curr Biol 2018;28:268–74.e5.
[270] Chau LM, Goodisman MAD. Gene duplication and the evolution of phenotypic diversity in insect societies. Evolution 2017;71:2871–84.
[271] Mukherjee K, Vilcinskas A. Chapter 14 - Transgenerational epigenetic inheritance in insects. In: Tollefsbol TO, editor. Transgenerational Epigenetics (Second Edition), vol. 13, Academic Press; 2019, p. 315–29.
[272] Oppold A, Kreß A, Vanden Bussche J, Diogo JB, Kuch U, Oehlmann J, et al. Epigenetic alterations and decreasing insecticide sensitivity of the Asian tiger mosquito Aedes albopictus. Ecotoxicol Environ Saf 2015;122:45–53.
[273] Bass C, Puinean AM, Zimmer CT, Denholm I, Field LM, Foster SP, et al. The evolution of insecticide resistance in the peach potato aphid, Myzus persicae. Insect Biochem Mol Biol 2014;51:41–51.
[274] Yang X, Wei X, Yang J, Du T, Yin C, Fu B, et al. Epitranscriptomic regulation of insecticide resistance. Sci Adv 2021;7. https://doi.org/10.1126/sciadv.abe5903.
[275] Kalsi M, Palli SR. Transcription factors, CncC and Maf, regulate expression of CYP6BQ genes responsible for deltamethrin resistance in Tribolium castaneum. Insect Biochem Mol Biol 2015;65:47–56.
[276] Wilding CS. Regulating resistance: CncC:Maf, antioxidant response elements and the overexpression of detoxification genes in insecticide resistance. Curr Opin Insect Sci 2018;27:89–96.

[277] McDonnell CM, Brown RP, Berenbaum MR, Schuler MA. Conserved regulatory elements in the promoters of two allelochemical-inducible cytochrome P450 genes differentially regulate transcription. Insect Biochem Mol Biol 2004;34:1129–39.
[278] Brown RP, McDonnell CM, Berenbaum MR, Schuler MA. Regulation of an insect cytochrome P450 monooxygenase gene (CYP6B1) by aryl hydrocarbon and xanthotoxin response cascades. Gene 2005;358:39–52.
[279] Misra JR, Horner MA, Lam G, Thummel CS. Transcriptional regulation of xenobiotic detoxification in Drosophila. Genes Dev 2011;25:1796–806.
[280] Yang X, Deng S, Wei X, Yang J, Zhao Q, Yin C, et al. MAPK-directed activation of the whitefly transcription factor CREB leads to P450-mediated imidacloprid resistance. Proc Natl Acad Sci U S A 2020;117:10246–53.
[281] Seong KM, Coates BS, Pittendrigh BR. Impacts of Sub-lethal DDT Exposures on microRNA and Putative Target Transcript Expression in DDT Resistant and Susceptible Drosophila melanogaster Strains. Front Genet 2019;10:45.
[282] Tian M, Liu B, Hu H, Li X, Guo Q, Zou F, et al. MiR-285 targets P450 (CYP6N23) to regulate pyrethroid resistance in Culex pipiens pallens. Parasitol Res 2016;115:4511–7.
[283] Pu J, Wang Z, Chung H. Climate change and the genetics of insecticide resistance. Pest Manag Sci 2020;76:846–52.
[284] Xia X, Sun B, Gurr GM, Vasseur L, Xue M, You M. Gut Microbiota Mediate Insecticide Resistance in the Diamondback Moth, Plutella xylostella (L.). Front Microbiol 2018;9:25.
[285] Lockwood JA, Byford RL, Story RN, Sparks TC, Quisenberry SS. Behavioral Resistance to the Pyrethroids in the Horn Fly, Haematobia irritans (Diptera: Muscidae). Environ Entomol 1985;14:873–80.
[286] Heckel DG. Insect detoxification and sequestration strategies. Annual Plant Reviews, Chichester, UK: John Wiley & Sons, Ltd; 2014, p. 77–114.
[287] Heidel-Fischer HM, Vogel H. Molecular mechanisms of insect adaptation to plant secondary compounds. Curr Opin Insect Sci 2015;8:8–14.
[288] Gilbert LI, Gill SS. Insect Pharmacology: Channels, Receptors, Toxins and Enzymes. Academic Press; 2010.
[289] Labbé P, Berthomieu A, Berticat C, Alout H, Raymond M, Lenormand T, et al. Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito Culex pipiens. Mol Biol Evol 2007;24:1056–67.
[290] Bass C, Zimmer CT, Riveron JM, Wilding CS, Wondji CS, Kaussmann M, et al. Gene amplification and microsatellite polymorphism underlie a recent insect host shift. Proc Natl Acad Sci U S A 2013;110:19460–5.
[291] Rosenheim JA, Hoy MA. Genetic Improvement of a Parasitoid Biological Control Agent: Artificial Selection for Insecticide Resistance in Aphylis melinus (Hymenopera: Aphelinidae). J Econ Entomol 1988;81:1539–50.
[292] Chaverra-Rodriguez D, Jaramillo-Ocampo N. Artificial selection of insecticide resistance to lambda-cyhalothrin in Aedes aegypti and cross resistance to other insecticides. Revista Colombiana 2012.
[293] Maurya RP, Khan MA, Kumar S, Kumar A. Artificial selection of green lacewing, Chrysoperla carnea Stephens for tolerance to insecticides. Annals of Plant Protection Sciences 2009;17:1–4.
[294] Kruidhof HM, Smid HM, Thiel A, Hoffmeister TS, Vet LEM, Others. Olfactory conditioning of natural enemies: potential benefits for pest control in greenhouse crops. Olfactory Conditioning of Natural Enemies: Potential Benefits for Pest Control in Greenhouse Crops 2014;102:121–6.
[295] Giunti G, Canale A, Messing RH, Donati E, Stefanini C, Michaud JP, et al. Parasitoid learning: Current knowledge and implications for biological control. Biol Control 2015;90:208–19.
[296] Michaud JP. Problems Inherent to Augmentation of Natural Enemies in Open Agriculture. Neotrop Entomol 2018;47:161–70.
[297] Nunney L. Managing captive populations for release: a population-genetic perspective. Quality control and production of biological control agents: theory and testing procedures, Wallingford: CABI; 2003, p. 73–87.
[298] Bijlsma, Bundgaard, Boerema. Does inbreeding affect the extinction risk of small populations?: predictions from Drosophila. J Evol Biol 2000;13:502–14.
[299] Woodworth LM, Montgomery ME, Briscoe DA, Frankham R. Rapid genetic deterioration in captive populations: Causes and conservation implications. Conserv Genet 2002;3:277–88.
[300] Hick CA, Field LM, Devonshire AL. Changes in the methylation of amplified esterase DNA during loss and reselection of insecticide resistance in peach-potato aphids, Myzus persicae. Insect Biochem Mol Biol 1996;26:41–7.
[301] Mansoor MM, Abbas N, Shad SA, Pathan AK, Razaq M. Increased fitness and realized heritability in emamectin benzoate-resistant Chrysoperla carnea (Neuroptera: Chrysopidae). Ecotoxicology 2013;22:1232–40.
[302] Abbas N, Mansoor MM, Shad SA, Pathan AK, Waheed A, Ejaz M, et al. Fitness cost and realized heritability of resistance to spinosad in Chrysoperla carnea (Neuroptera: Chrysopidae). Bull Entomol Res 2014;104:707–15.
[303] Khan Pathan A, Sayyed AH, Aslam M, Liu T-X, Razzaq M, Ahmed Gillani W. Resistance to Pyrethroids and Organophosphates Increased Fitness and Predation Potential of Chrysoperla carnae (Neuroptera: Chrysopidae). J Econ Entomol 2010;103:823–34.
[304] Jones SL, Kinzer RE, Bull DL, Ables JR, Ridgway RL. Deterioration of Chrysopa carnea in mass culture. Ann Entomol Soc Am 1978;71:160–2.
[305] Gilpin M. Minimum viable populations : Processes of species extinction 1986.
[306] Hokkanen HMT. Trap Cropping in Pest Management. Annu Rev Entomol 1991;36:119–38.
[307] Hill MP, Macfadyen S, Nash MA. Broad spectrum pesticide application alters natural enemy communities and may facilitate secondary pest outbreaks. PeerJ 2017;5:e4179.
[308] Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 2005;33:6494–506.
[309] Solovyev V. Statistical approaches in eukaryotic gene prediction. In: D.J. Balding, M. Bishop, C. Cannings, editor. Handbook of

Statistical Genetics, Chichester: John Wiley & Sons Ltd; 2001, p. 83–127.
[310] Haas BJ, 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;9:1–22.
[311] Smit AFA, Hubley R. RepeatModeler Open-1.0 2008-2015. http://www.repeatmasker.org.
[312] Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics 2013;29:2487–9.
[313] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res 2014;42:D222–30.
[314] Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0 2013-2015. http://www.repeatmasker.org.
[315] Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33:290–5.
[316] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644–52.
[317] Gilbert D. EvidentialGene - Evidence Directed Gene Construction for Eukaryotes 2013. https://sourceforge.net/projects/evidentialgene/.
[318] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008;36:3420–35.
[319] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 2013;29:2933–5.
[320] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–80.
[321] Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059–66.
[322] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.
[323] Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol 2008;25:1307–20.
[324] Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.

Chapter 6. Conclusions and Future Work

6.1 Conclusions

In this thesis, we addressed the problem of a shortage of genomic data for beneficial predators by assembling and annotating three high quality beneficial predator genomes - *Orius laevigatus* (minute pirate bug), *Sphaerophoria rueppellii* (European hoverfly) and *Microctonus brassicae* (parasitoid wasp). These genomes are a valuable resource and have been made publicly available, providing a useful reference for the respective families of these species - several of which contain a large number of other beneficial predators with potential for commercial use.

These beneficial predator genomes allowed for comparative analyses of resistance-associated detoxification genes to be performed between predators and the crop pests which they feed on. This comparison aimed to assess if there was any genomic basis which could explain why crop pests have a higher occurrence of resistance cases compared to beneficial predators. The detoxification genes were manually curated to provide highly accurate gene models which were used to perform comparative analyses between beneficial predators and crop pests. These comparative analyses included a comparison of insecticide target-sites and genes encoding metabolic enzymes potentially responsible for insecticide resistance.

The results from the comparative analysis suggested a greater degree of detoxification family gene expansion within crop pests compared to beneficial predators, which could contribute to the higher number of resistance cases seen in pests. This difference was particularly apparent in the UGT and P450 detoxification families which are associated with plant xenobiotic detoxification. This suggests that the plant-based diet of crop pests likely provided an increased selection pressure for these resistance mechanisms prior to the introduction of insecticides. On the other hand, beneficial predators which feed on insects and sometimes pollen/nectar, would have had far less exposure to plant xenobiotics and hence a reduced selection pressure for the associated detoxification mechanisms. Therefore, once insecticides

were introduced, crop pests may have had an advantage over beneficial predators in terms of developing insecticide resistance.

The comparative analysis also showed that variation in the levels of resistance between different beneficial predators correlated to some extent with gene expansion, with several factors having likely had some influence on this, including diet, migration and length of commercial use.

The knowledge gained from this project has contributed to our understanding of insecticide resistance from a genomic perspective and could aid in the development of successful IPM strategies.

6.2 Future work

The results presented in this thesis provide some preliminary observations for a genomic basis which may explain higher rates of resistance in pests compared to predators. However, further work is needed to confirm these claims.

The genomes used for comparative analysis vary greatly in their quality and this could affect the reliability of any claims made from such analyses. With the constant improvement and reducing cost of sequencing technologies, higher quality genomes should be generated where needed to allow for more accurate comparative analyses.

In the future, these genomes could be used to help generate pan-genomes which would give a better overview of resistance mechanisms for the species as a whole and could help infer how detoxification families vary between resistant and susceptible populations. Additionally, acquiring high numbers of individual sequence data for each species would allow for a more thorough analysis of target site mutations.

As a greater variety of closely related species are sequenced, their genomes could be used to perform deeper comparative analyses. These could confirm whether gene expansions are truly lineage specific by calculating the divergence between recently duplicated gene copies in relation to copies from closely related species.

Finally, future studies should look to include epigenetic and post-transcriptional regulation data when investigating resistance mechanisms. Whilst gene expansion may play a large role in resistance levels in the long term, these mechanisms may have a more significant role in the short-term when reacting to recent insecticide exposure. Additionally, such data would provide far more information than gene numbers alone when performing comparative analyses of resistance mechanisms.

This future work could provide further knowledge for the development of more successful IPM strategies which would increase crop yields and feed our growing world population.

Appendices

i) The assembly statistics at each stage in the assembly pipeline of the *Orius laevigatus* genome.

ii) Numbers of proteins in the *Orius laevigatus* genome annotated by the InterPro member databases.

iii) Phylogenetic tree of the *Orius laevigatus* ATP-binding cassette (ABC) transporters.

iv) Phylogenetic tree of *Orius laevigatus, Rhodnius prolixus and Acyrthosiphon pisum* Carboxyl/cholinesterases (CCEs).

v) Numbers of detoxification genes annotated in beneficial predators and crop pests. S1. ABCs; S2. P450s; S3. CCEs; S4. GSTs; S5. UGTs.

Appendix i) The assembly statistics at each stage in the assembly pipeline of the *Orius laevigatus* genome.

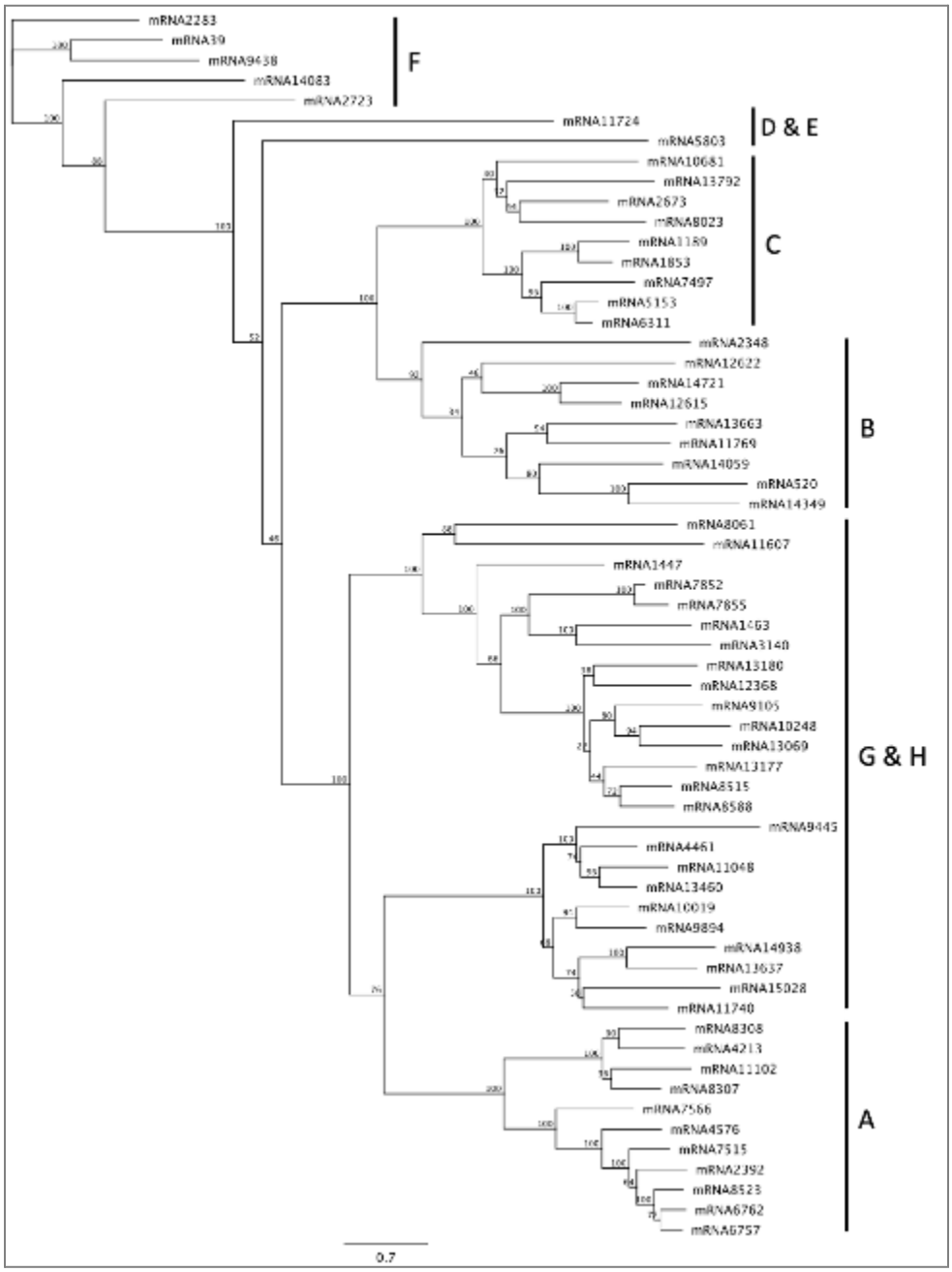
	Scaffold N50	Longest Scaffold	Number of Scaffolds	BUSCO Score C = complete, S = single copy, D = duplicated, F = fragmented, M = missing
Flye	89,878	1,464,728	2,800	C:86.1%[S:83.8%,D:2.3%],F:4.5%,M:9.4%
Flye w/ Rascaf	105,409	1,464,728	2,611	C:86.1%[S:83.7%,D:2.4%],F:4.6%,M:9.3%
Falcon	22,424	1,701,193	7,514	C:54.3%[S:49.2%,D:5.1%],F:10.4%,M:35.3%
Falcon w/ Unzip	43,913	1,667,874	1,665	C:45.5%[S:43.4%,D:2.1%],F:4.3%,M:50.2%
Canu	27,715	859,161	8,818	C:75.8%[S:65.4%,D:10.4%],F:7.2%,M:17.0%
Quickmerge: Falcon w/ Unzip + Flye w/ Rascaf	112,029	1,935,382	2,473	C:85.6%[S:82.8%,D:2.8%],F:4.8%,M:9.6%
Quickmerge: Falcon w/ Unzip + Flye w/ Rascaf + Canu	120,811	2,051,687	2,353	C:84.5%[S:81.7%,D:2.8%],F:4.9%,M:10.6%
1st Round Pilon	121,050	2,051,719	2,353	C:89.8%[S:86.2%,D:3.6%],F:1.9%,M:8.3%
Redundans redundancy removal	128,051	2,051,719	2,018	C:88.9%[S:86.8%,D:2.1%],F:2.1%,M:9.0%
2nd Round Pilon	128,042	2,051,825	2,018	C:88.9%[S:86.7%,D:2.2%],F:2.1%,M:9.0%
Redundans scaffolding	128,042	2,051,825	2,012	C:88.9%[S:86.7%,D:2.2%],F:2.1%,M:9.0%
3rd Round Pilon	128,043	2,051,691	2,012	C:89.0%[S:86.8%,D:2.2%],F:2.1%,M:8.9%
Manually bringing in genes from the original Flye assembly	125,152	2,051,691	2,022	C:90.9%[S:88.7%,D:2.2%],F:1.9%,M:7.2%
Manually bringing in genes from the original Falcon Assembly	125,152	2,051,691	2,023	C:91.1%[S:88.9%,D:2.2%],F:1.7%,M:7.2%
Manually bringing in genes from the original Canu Assembly	125,657	2,051,691	2,050	C:93.1%[S:91.0%,D:2.1%],F:2.0%,M:4.9%
4th Round Pilon (5th showed no improvement)	125,649	2,051,674	2,050	C:93.6%[S:91.3%,D:2.3%],F:1.6%,M:4.8%

Appendix ii) Numbers of proteins in the *Orius laevigatus* genome annotated by the InterPro member databases.

InterPro member database	**Number of proteins annotated**
Pfam	9,799
PANTHER	10,433
CATH-GENE3D	8,278
SUPERFAMILY	8,097

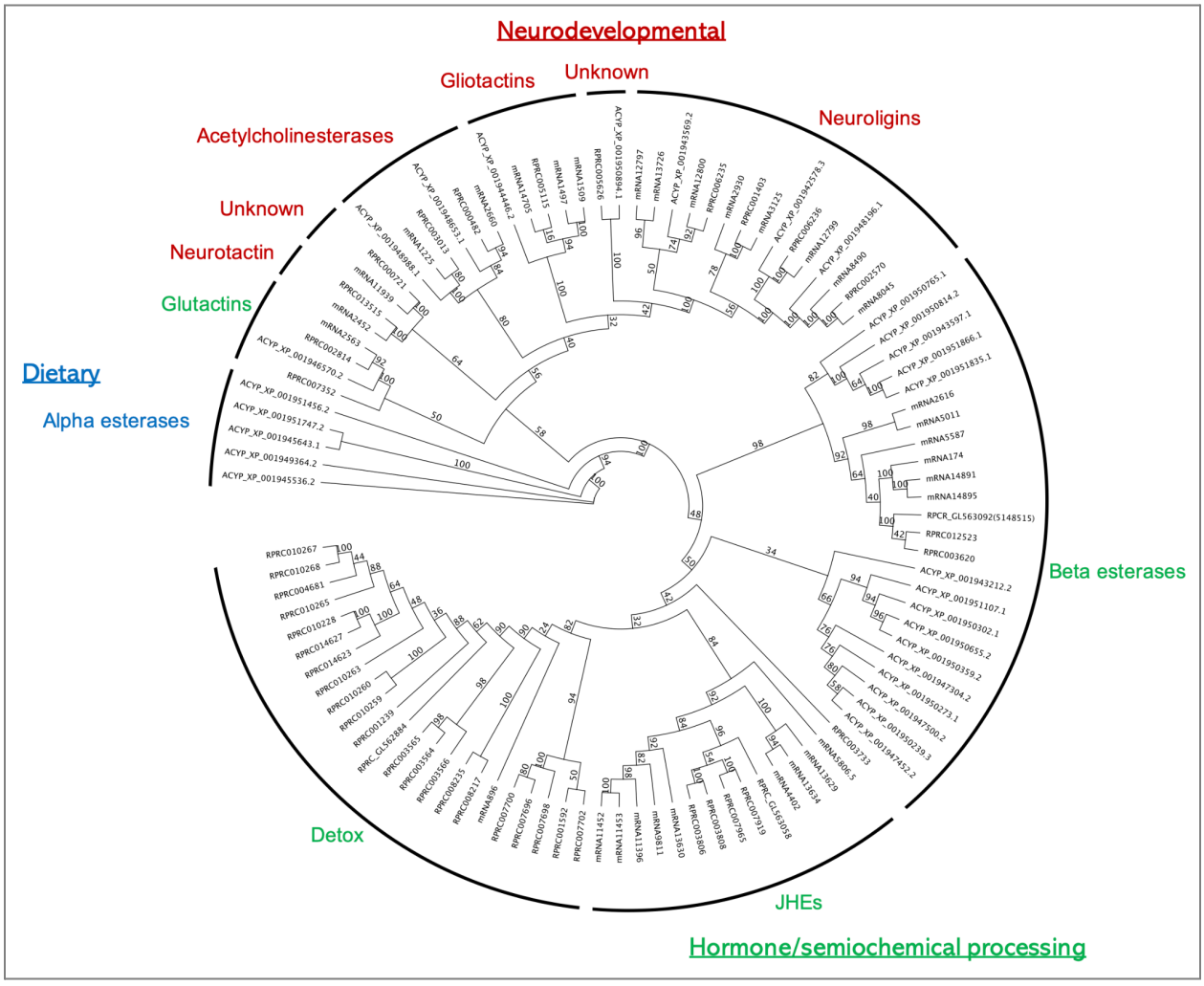
Appendix iii) Phylogenetic tree of the *Orius laevigatus* ATP-binding cassette (ABC) transporters.

Amino acid sequences were aligned using MAFFT and analysed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates.



Appendix iv) Phylogenetic tree of *Orius laevigatus, Rhodnius prolixus and Acyrthosiphon pisum* Carboxyl/cholinesterases.

Amino acid sequences were aligned using MAFFT and analysed using RAxML (the GAMMA LG protein model was used). The bootstrap consensus tree was inferred from 100 replicates. *O. laevigatus* CCEs are those beginning "mRNA…". *R. prolixus (RPRC)* and *A. pisum (ACYP)* CCEs were obtained from UniProt.



Appendix v) Numbers of detoxification genes annotated in beneficial predators and crop pests.

S1. ABCs; S2. P450s; S3. CCEs; S4. GSTs; S5. UGTs. (Reference numbers from section 5.15.)

Table S1. Numbers of ABC transporter genes annotated in *Microctonus brassicae* (author's previous work, to be released), *Orius laevigatus, Sphaerophoria rueppellii* [19,20] *Chrysoperla carnea* (annotated in this study), *Frankliniella occidentalis* [30], *Thrips palmi* [318], *Aphis gossypii* [181], *Trialeurodes vaporariorum* [319], *Diuraphis noxia* and *Bemisia tabaci* [183] and their distribution across subfamilies.

	Species	**ABCA**	**ABCB**	**ABCC**	**ABCD**	**ABCE**	**ABCF**	**ABCG**	**ABCH**	**Total**
Beneficial Predators	*Microctonus brassicae*	4	6	13	2	1	3	15	0	**44**
	Sphaerophoria rueppellii	11	6	8	3	1	3	10	3	**45**
	Orius laevigatus	11	9	9	1	1	5	23	2	**64**
	Chrysoperla carnea	8	5	10	2	3	4	14	1	**47**
Hemiptera crop pests	*Frankliniella occidentalis*	3	5	19	2	1	3	22	13	**70**
	Thrips palmi	3	4	12	2	2	3	16	7	**49**
	Diuraphis noxia	3	6	24	3	1	3	26	11	**77**
	Aphis gossypii	4	5	25	2	1	4	30	0	**71**
	Trialeurodes vaporariorum	3	9	7	4	1	3	9	9	**45**
	Bemisia tabaci	8	3	6	2	1	3	23	9	**55**

Table S2. Total numbers of Cytochrome P450 genes annotated in *Microctonus brassicae* (author's previous work, to be released), *Orius laevigatus, Sphaerophoria rueppellii* [19,20] *Chrysoperla carnea* (annotated in this study), Harmonia axyridis [21], *Metaseiulus occidentalis, Frankliniella occidentalis, Thrips palmi* [30], *Myzus persicae, Acyrthosiphon pisum* [320], *Trialeurodes vaporariorum* [319], *Bemisia tabac*i [184] and *Tetranychus urticae* [321].

	Species	**CYP2**	**CYP3**	**CYP4**	**Mito**	**Total**
Beneficial Predators	*Microctonus brassicae*	7	29	21	7	**64**
	Sphaerophoria rueppellii	6	34	15	14	**69**
	Orius laevigatus	6	34	13	5	**58**
	Chrysoperla carnea	-	-	-	-	**99**
	Harmonia axyridis	10	42	28	9	**89**
	Metaseiulus occidentalis	16	23	19	5	**63**
Hemiptera crop pests	*Frankliniella occidentalis*	12	22	37	10	**81**
	Thrips palmi	12	26	42	11	**91**
	Myzus persicae	3	63	48	1	**115**
	Acyrthosiphon pisum	10	33	32	8	**83**
	Trialeurodes vaporariorum	7	41	25	7	**80**
	Bemisia tabaci	18	76	73	4	**171**
	Tetranychus urticae	48	10	23	5	**86**

Table S3. Numbers of CCE genes annotated in *Microctonus brassicae* (author's previous work, to be released), *Orius laevigatus*, *Sphaerophoria rueppellii* [19,20] *Chrysoperla carnea* (annotated in this study), *Metaseiulus occicdentalis*, *Tetranychus urticae* [321], *Frankliniella occidentalis* [30], *Myzus persicae* [320], *Acyrthosiphon pisum*, *Bemisia tabaci* [185] and *Trialeurodes vaporariorum* [322] and their distribution across classes and clades.

	Species	**Dietary class**	**Hormone/ semiochemical processing class**	**Neurodevelopmental**						**Acari-specific classes**	**Total**
				Glutactins	AChE	uncharacterised	gliotactin	neuroligin	neurotactin		
Beneficial predators	*Microctonus brassicae*	13	4	0	2	0	0	5	1	0	**25**
	Sphaerophoria rueppellii	15	13	4	1	-	1	5	1	0	**40**
	Orius laevigatus	0	16	1	2	1	3	8	1	0	**32**
	Chrysoperla carnea	46	4	0	4	0	0	3	1	0	**58**
	Metaseiulus occidentalis	0	0	0	1	0	1	5	0	37	**44**
Hemiptera crop pests	*Frankliniella occidentalis*	28	7	2	2	2	1	7	1	0	**50**
	Myzus persicae	5	12	0	3	1	1	0	0	0	**22**
	Acyrthosiphon pisum	5	16	0	2	1	1	3	0	0	**28**
	Trialeurodes vaporariorum	12	6	1	2	1	1	3	1	0	**27**
	Bemisia tabaci	6	19	1	4	1	1	10	0	0	**42**
	Tetranychus urticae	0	2	2	1	0	1	5	1	59	**71**

Table S4. Numbers of GST genes annotated in *Microctonus brassicae* (author's previous work, to be released), *Orius laevigatus, Sphaerophoria rueppellii* [19,20] *Chrysoperla carnea* (annotated in this study), *Metaseiulus occicdentalis, Tetranychus urticae* [321], *Thrips palmi* [318], *Myzus persicae, Acyrthosiphon pisum, Trialeurodes vaporariorum and Bemisia tabaci* [182] and their distribution across classes. The 'Mu' class is likely Acari-specific [321].

	Species	Delta	Epsilon	Omega	Sigma	Theta	Zeta	Microsomal	Mu	Unclassified	Detoxification Total *
Beneficial Predators	*Microctonus brassicae*	0	1	1	4	0	1	4	0	1	**8**
	Sphaerophoria rueppellii	4	11	3	1	3	1	0	0	-	**23**
	Orius laevigatus	1	0	2	16	1	1	3	0	-	**24**
	Chrysoperla carnea	5	9	0	3	1	1	1	0	-	**20**
	Metaseiulus occidentalis	3		3	0	0	1	-	5	1	**13**
Hemiptera crop pests	*Thrips palmi*	14	0	1	6	1	2	1	0	-	**24**
	Myzus persicae	3	0	1	12	1	0	2	0	0	**17**
	Acyrthosiphon pisum	11	0	1	5	2	0	2	0	0	**19**
	Trialeurodes vaporariorum	9	1	0	3	0	2	3	0	0	**15**
	Bemisia tabaci	14	0	1	6	0	2	2	0	0	**23**
	Tetranychus urticae	16		2	0	0	1	-	12	0	**31**

(*microsomal not included)

Table S5. Numbers of annotated UDP glycosyltransferase genes found in *Microctonus brassicae* (author's previous work, to be released), *Orius laevigatus*, *Sphaerophoria rueppellii* [19,20] *Chrysoperla carnea* (annotated in this study), *Tetranychus urticae, Nilaparvata lugens, Acyrthosiphon pisum, Bemisia tabaci* [241], *Thrips palmi* [318], *Myzus persicae* [323] and *Trialeurodes vaporariorum* [31].

	Species	**Total**
Beneficial predators	*Microctonus brassicae*	**15**
	Sphaerophoria rueppellii	**46**
	Orius laevigatus	**10**
	Chrysoperla carnea	**63**
Hemiptera crop pests	*Tetranychus urticae*	**81**
	Nilaparvata lugens	**20**
	Thrips palmi	**17**
	Myzus persicae	**101**
	Acyrthosiphon pisum	**72**
	Trialeurodes vaporariorum	**55**
	Bemisia tabaci	**76**