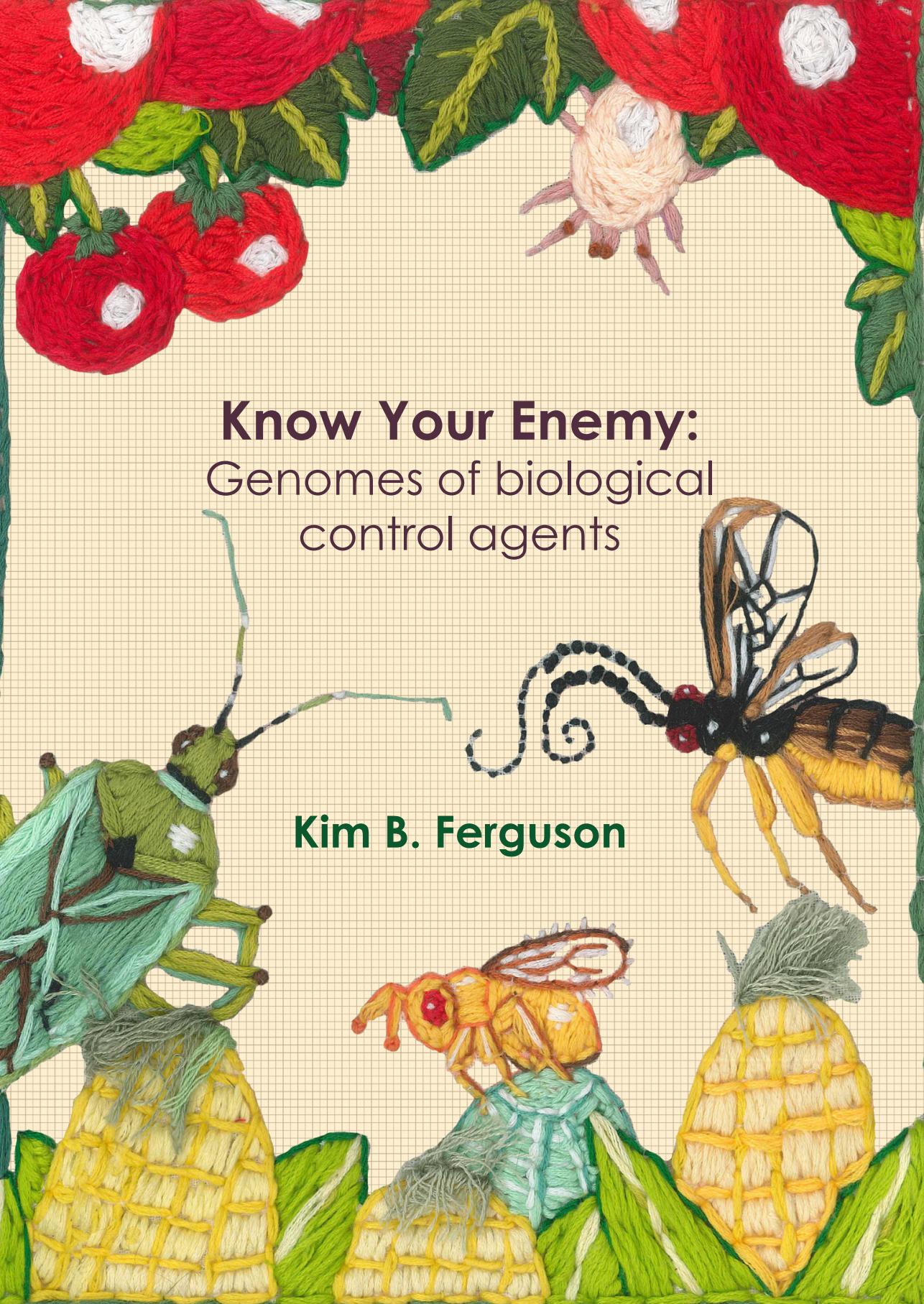
Know Your Enemy:
Genomes of biological control agents



Kim B. Ferguson

Know Your Enemy:
Genomes of biological control agents

Kim B. Ferguson

Thesis Committee

Promotor

Prof. Dr B. J. Zwaan
Professor of Genetics
Wageningen University & Research

Co-promotors

Dr B. A. Pannebakker
Assistant professor, Laboratory of Genetics
Wageningen University & Research

Dr E. C. Verhulst
Assistant professor, Laboratory of Entomology
Wageningen University & Research

Other Members

Prof. Dr M. M. van Oers, Wageningen University & Research
Dr Y. H. Chen, University of Vermont, USA
Prof. Dr J. Ellers, Vrije Universiteit Amsterdam
Prof. Dr B. Wertheim, University of Groningen

This research was conducted under the auspices of the C. T. de Wit Graduate School for Production Ecology and Resource Conservation.

Know Your Enemy:
Genomes of biological control agents

Kim B. Ferguson

Thesis

Submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 25 March 2020
at 4:00 PM in the Aula.

Kim B. Ferguson
Know Your Enemy: Genomes of biological control agents
312 pages

PhD Thesis, Wageningen University, Wageningen, The Netherlands (2020)
With references, with summary in English, Dutch, German, and Spanish

ISBN 978-94-6395-281-1
DOI 10.18174/511993

Table of Contents

CHAPTER 1:
Introduction 7

CHAPTER 2:
Genetic variation of traits in natural enemies relevant for biological control: a systematic review 29

CHAPTER 3:
Effect of mass rearing on the genetic diversity of the predatory mite *Amblyseius swirskii* 67

CHAPTER 4:
Braconidae revisited: *Bracon brevicornis* genome showcases the potential of linked-read sequencing in identifying a putative *complementary sex determiner* gene 91

CHAPTER 5:
Hybrid genome assembly and annotation of egg parasitoid and biological control agent *Trichogramma brassicae* 109

CHAPTER 6:
The wasp and the butterfly: opulation genetics and genomics of *Trichogramma evanescens* hint at role of phoresy in dispersal 129

CHAPTER 7:
Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest 163

CHAPTER 8:
Synthesis 211

APPENDICES 235

References 236

Summary 280

Summary in Dutch, German, and Spanish 285

Acknowledgements 303

About the Author 307

List of Publications 308

PE&RC Training and Education Statement 309

About the Artwork 310



Chapter 1

Introduction

Kim B. Ferguson

Chapter 1

To whom, dear Muse, do you bring these varied fruits of song, or who was it who wrought this garland of poets?[1]

Anthologies

The root of the English word "anthology" is interesting[2], as its meaning is two-fold in the original Greek: *anthologia* ('a gathering of flowers') and *anthologiai* ('a collection of epigrams') (Beal, 2008). Meleager of Gadara (in modern-day Jordan) is credited with creating the first anthology in approximately 100 B.C.E., though the original has been lost for many centuries (Gutzwiller, 1997). Meleager collected epigrams– a type of short and pithy Greek prose– from various well-known poets, as well as his own, and put them all into one binding. The goal was to bring attention to the works that, in Meleager's eyes, were the best and most deserving of praise. Comparing each poet to a flower, he created a "garland" out of these works, and so this work came to be known as Meleager's *Garland*. The *Garland* became the only source for many of these poems, and was itself placed within a larger anthology, the *Anthologia Palatina*. While earlier modern critics of Greek poetry and literature viewed Meleager as unoriginal and overblown, more recent critics see him as more of a very intentional and creative editor. The structure of the *Garland* also includes two epigrams that are deliberately set outside the rest of the collection and function as introduction and epilogue.

Moving into present day, anthologies are collections of works of one or more authors, often on a specific theme or genre (Beal, 2008). Perhaps the most well-known is the Norton set of anthologies, a perennial occupant of North American university libraries and textbook stores (Donadio, 2006). Modern anthologies serve as an accessible yet intentional collection of works and can extend beyond literature. One such example is the ambitious *Environment: An interdisciplinary anthology*, a collection of various essays and articles with sources across sciences, life sciences, and humanities on the theme of the

1 from Meleager's *Garland*, AP 4.1, sourced from http://www.attalus.org/poetry/meleager.html with number references to the original *Anthologia Palatina*

2 Also interesting is that while Dutch usage switches between *anthologie* and *bloemlezing*, both come from the same root idea, but in different languages: Greek and Latin (*florilegium*), respectively (van der Sijs).

environment (Adelson et al., 2008).

In many ways, a doctoral thesis is an anthology. It is specific to a theme (the project of the doctoral candidate), often includes multiple authors (co-authors on each chapter or publication) and contains both an introduction (this chapter) as well as an epilogue (the synthesis), similar to Meleager's *Garland* and other contemporary iterations. The goal then, of this introduction, is simple: to convince you, the reader, that this anthology is of immediate interest to you, both in terms of its varied chapters but also as a unified whole. To do so, I will first introduce two fields in biology that thus far have mostly run in parallel: biological control and genetics/genomics.

Biological control

Biological control, also known as biocontrol, is the introduction of one organism to control the population of another organism. The practice of releasing these "natural enemies" has been in practice for nearly as long as agriculture in some form (van Lenteren and Godfray, 2005), but has been commercially formalised for over 120 years, with a sharp increase in commercial and economic presence in the last two decades (van Lenteren, 2012). Both types of pests and biological control agents (BCAs) available to use against them run the gamut: microbes, plants, nematodes, arthropods, mammals, birds, *et cetera*. Here, I will narrow the focus to arthropod pests and their arthropod BCAs, as these are the focus of my thesis.

Types of biological control

Biological control is typically divided into three categories: conservation, classical, and augmentative. Conservation biological control is about protecting the habitat necessary for natural enemies to thrive and often extends beyond agricultural fields. It is closely related to ecosystem management, and is often happening at the level of large, outdoor landscapes. One example of conservation biological control would be maintaining hedges on the edge of an outdoor tomato field in order to promote a buffer zone for predators and parasitoids against tomato pests (Balzan et al., 2015).

Classical biological control, the more (in)famous of the three, is typically what the general public and biologists alike think of when they hear "biological control." This is where a foreign parasitoid or predator is released against a pest (which can be foreign or native, but most often is foreign). The idea here is that if existing natural enemies in the affected environment have not been able to deal with the invasive pest, the natural enemy that keeps the pest in check in the place of origin could be the solution. The first successful account of classical biological control is often cited as the importation of the vedalia beetle, *Rodolia cardinalis* (Mulsant) (Coleoptera: Coccinellidae) from Australia to control the cottony cushion scale, *Icerya purchasi* (Maskell) (Hemiptera: Margodidae) in California, USA in 1899 (Caltagirone and Doutt, 1989; Mason et al., 2017). The success of the vedalia beetle was recognised by Australia in stamp form in 2003 (Figure 1) ("Cardinal Ladybird (*Rodalia cardinalis*)," 2019).

In turn, the infamy of classical biological control comes from ecological catastrophes that have occurred, such as the well-known cane toad, *Rhinella (Bufo) marinus* (L.) (Amphibia: Anura), situation in Australia. Australia was not the first country where the cane toad was introduced, nor was it the last, but it was certainly the most studied introduction from the first half of the 20th century (Easteal, 1981). It was brought in as predator of the cane beetle, *Dermolepida albohirtum* (Waterhouse) (Coleoptera: Scarabaeidae), one of the main agricultural pests of Australian sugar cane. Over time, the cane toad extended beyond the expected range, led to the decline of several native predators, and also did not do a good job of controlling the pest (Doody et al., 2009). Problem species, such as the cane toad in Australia, are no longer referred to as "biological control agents" but rather "invasive species," despite long-term establishment being a goal of classical biological control. The whole world has not turned their back on the cane toad, however, as in Barbados it was honoured with its own stamp in 1989, following a successful introduction sometime in the early half of the 19th century (Easteal, 1981) (Figure 1) ("Cane Toad (*Bufo marinus*)," 2019). Due to a number of unintended consequences such as the establishment of toxic toads in Australia, the importation of non-native species for classical biological control is now highly regulated, though



Figure 1. Left, The vedalia beetle, Australian postage, 2003; Right, the cane toad, Barbados postage, 1989. Credit: colnect.com

loopholes exist and accidents continue to happen (Stouthamer, 2018).

Augmentative biological control is the commercial rearing and release of BCAs for release in large numbers (inundative) for immediate pest management, or in advance of expected pests (inoculative) (Lommen et al., 2017). More recently, the BCAs used in augmentative biological control are found in the area they are released (van Lenteren, 2012), while others may not have been native to the area but are now widespread regardless and can be released (providing they are not considered a pest) (Stouthamer, 2018). Augmentative biological control also represents a sizable industrial presence in the form of biological control companies, with a global market value of approximately US$ 1.7 billion in 2015 (a drop in the bucket compared to the entire pesticide market, US$ 58 billion) (van Lenteren et al., 2018).

Integrated Pest Management (IPM)

If we consider that biological control involves the application of insect ecology, behavioural ecology, and population ecology, then integrated pest management (IPM) involves the application of augmentative biological control alongside economic entomology. Directly from the Food and Agriculture Organisation of the United Nations, the FAO: "Integrated Pest Management (IPM) is an ecosystem approach to crop production and protection that combines different management strategies and practices to grow healthy crops and minimize the use of pesticides" (Food and Agriculture Organisation of the United Nations, 2019). By bringing economic reasoning into pest management, IPM goes beyond the application of entomology, adding in human values and valuation along the way (Onstad and Knolhoff,

2009). The relatively low monetary cost of pesticides compared to BCAs would seemingly keep BCAs from becoming a viable tool for pest management, or at least keep them from being centred in a pest management program. However, works such as Rachel Carson's *Silent Spring* and consumer demands for less pesticide residue in their food have buoyed an interest in reducing pesticide use in crop production (Feditchkina Tracy, 2014).

To this end, BCAs are often used in IPM programs as a key player in pest management, whereas pesticides (both synthetic and organic) are used as little as possible (Feditchkina Tracy, 2014; van Lenteren, 2012). The action plan of IPM essentially begins with cultural practices to reduce the possibility of pests invading and establishing (Stenberg, 2017). There are various rungs in the action ladder, including augmentative biological control, where natural enemies should reduce pest populations. If not, the final step on the ladder will be pesticide use. The guiding principle of ascending this decision ladder is that of economic thresholds, which describe the pest density capable of causing economic damage (Stern, 1973). The higher the yield, the better the outcome for the grower, within tolerable limits (such as what percentage of crop loss due to pest damage is tolerable to a producer). We cannot eradicate all pests, but we can reduce them to a number where crop loses are tolerable within economic thresholds: the economic injury limit (EIL) (Pedigo, 1986).

IPM researchers have noted that there is the influence of a sort of "other IPM," where instead of integrated pest management, it is integrated pesticide management, where a "quick-fix" mentality impacts the design of IPM strategies where the logical endpoint is pesticide use (Ehler, 2006). In response to this concern, recent calls have been made for a return to so-called "true" IPM, also dubbed "conscious agriculture," warning IPM practitioners that a systems-based approach is necessary (Lewis et al., 1997; van Lenteren et al., 2018). Greenhouses and agricultural fields are complex agro-ecosystems, and understanding ecosystem dynamics and key players therein is of greater benefit for reducing pesticide use and the associated environmental and human damage (van Lenteren et al., 2018). "What makes the pest a pest? How do we boost plant resistance? What is necessary to support predators

and parasitoids?" are just some examples of the questions that need to be addressed (Lewis et al., 1997; van Lenteren et al., 2018).

Augmentative biological control is arguably the most environmentally responsible and economically profitable option for pest management in greenhouses, yet has failed to catch on in a large way globally (Naranjo et al., 2015; van Lenteren, 2012). While 90% of Dutch vegetable-growing greenhouses utilize IPM strategies, this is in stark contrast to the global average of five percent (Feditchkina Tracy, 2014; Pilkington et al., 2010). Crucially, part of that gap is due to the perceived unreliability of biological control as compared to conventional pesticides. Yet, including the concepts such as the ecology and biology of established natural enemies holds much promise. Indeed, there have been plenty of calls for improving the addition of biological control in IPM research (Giles et al., 2017). Essentially, we need to consider that existing BCAs can be improved (Kruitwagen et al., 2018), and one way to ensure that is using genomics. But how?

Genetics and Genomics

The study of genetics is essentially about inheritance and evolution, how traits are passed on from one generation to another and why (or why not), and this happens at a variety of scales. From reading and comparing nucleic acids or their translated proteins, to the comparison between genotypes and/or phenotypes, to looking at trends within and between populations of organisms, all the way to comparing groups of species based on their genetics or relationships through evolution – genetics is about inheritance and evolution across all levels. This can also be studied using discrete units, such as barcode sequences, genes, or whole-genome comparisons.

We are currently situated in what has been called the omics era: proteomics, transcriptomics, and genomics, among others (Ellegren, 2014). This has been bolstered by improved sequencing technologies, often referred to as next generation sequencing (NGS) (Ellegren, 2014). However, despite these technological advancements and the large reduction in costs of sequencing whole genomes, there are concerns that other costs are rising e.g. regarding reliable data storage (Papageorgiou et al., 2018; Wetterstrand, 2017).

Chapter 1

What is a genome?

Over the years, the definition of a genome and how to describe it has been changing. While once used in comparison to a blueprint for an organism, that analogy has since fallen out of favour for a variety of reasons, the largest of which is that it is not an apt metaphor for both the complexity of genomes as well as the additional molecular and biological processes that impact the development of an organism (Pigliucci and Boudry, 2011). Additionally, genomes hold nongenetic functions, which the blueprint definition misses entirely (Bustin and Misteli, 2016).

Tracing back to 1920, the original definition discussed a haploid chromosome set that is the foundation of an organism, and as a term it did not immediately enter the scientific discourse (Goldman and Landweber, 2016). For a modern definition, the National Institutes of Health (National Institutes of Health, 2019) defines a genome as "an organism's complete set of DNA, including all of its genes." Furthermore, "each genome contains all of the information needed to build and maintain that organism." This information is stored in nucleic acids, wound up within strands of DNA that, in eukaryotes, are wound up even tighter into chromosomes and packed within the nucleus. This is the biological understanding of a genome. When DNA is extracted, sequenced, and assembled (more in **Box 1**), this becomes the computational or bioinformatic concept of a genome. In this case, a genome is a working draft, a hypothesis, and a snapshot at a certain point in time, at least at an individual level (Ekblom and Wolf, 2014).

Box 1 Genome projects: From raw data to annotation

Eukaryotic genome projects contain several stages that rely on each other, and naturally they have their own vocabulary.

The genome assembly is the backbone where the nucleotides are arranged in pieces (contigs, scaffolds, or chromosomes, depending on the size and structural information). In order to get an assembly, you need sequence data ("reads") to build the assembly. Genetic material

is sequenced into A's, T's, C's, and G's, on one or more platforms. The type of platform matters, as a variety of types of reads are possible, depending on the quality of the input material. While I have mentioned NGS technologies, these can range in chemistry and biological principle of how genetic material is being extracted. The quality of sequence data also varies depending on the platform. Finally, the length of the sequence matters. As you see in later chapters, they are often referred to as short or long reads, and the length of reads generated will impact the assembly strategy.

Moving on to assembly, this is performed by a computer program– or a pipeline of several– to puzzle these reads together to form an assembly. There is a plethora of assembly pipelines currently available, and these options will continue to expand as more assemblers are built and pipelines constructed. Moreover, in most cases it is never just a single pipeline, but rather several that are used, with the outputs compared with regards to completeness, size, or contiguity.

On top of these arrangements of A's, T's, C's, and G's is the genome annotation. This is an extra layer of analysis that contains gene location information. The annotation may be derived entirely by computer predicted models ("*ab-initio*"), based on evidence such as a transcriptome ("evidence-based"), or based on comparisons to other species ("homology-based"). Most annotations are a combination of these methods. From the annotation, gene and protein sets can be derived, though further investigation is necessary for improvement – this is referred to as "manual annotation". Annotations are important, and while not a strict requirement, the presence of an annotation often depends on the questions being asked and the amount of resources available, where the default minimum is often an *ab-initio* only annotation, itself improving over time with new software and genomes available for deriving gene structures.

Quality of genomes differ, with several model genomes that we can look to as examples. One of the first eukaryotic genomes to be assembled was that of the fruit fly, *Drosophila melanogaster*, largely due to its reduced complexity compared to humans (Papanicolaou, 2016). Being one of the first, and the most complete, it is important to note that genomes such as *D. melanogaster* are essentially the diamond standard of assembled and annotated genomes currently available. *D. melanogaster* is a model organism, used in a large variety of research that extends beyond insect genomics. The amount of time, resources, researchers, and data that went into the initial genome and all additional updates would not be considered the standard for new genome projects to be held to.

Conventionally, the value of a genome is judged based on its utility for the task at hand, not its completeness (Ekblom and Wolf, 2014; Ellegren, 2014). However, the notion of a "complete" genome is also relative, as there will always be a certain amount of a genome that will be unknown (including in *D. melanogaster*), due to highly repetitive segments that evade sequencing along with the insertion of ambiguous nucleotides to span gaps.

So, what are genomes good for? It depends on the question being asked, and once again it depends on scale (Richards, 2015). As reviewed in Richards 2015, in-depth functional analyses such as CRISPR gene editing, RNAi interference, or gene-network modelling are fine-tuned methods where potential interactions can cause problems. For these techniques, accuracy is key. Richards continues by widening the focus to population genomics, where populations can be defined and compared with regards to structure and make-up. Genomes can improve methods such as QTL mapping or phenotype mapping as well as improve genetic selection. While the completeness of the genome is not as important, having access to more data can increase the power of analyses. At the largest scale, comparative genomics move beyond an individual species, where comparisons between species can reveal regions of differentiation, as well as answer evolutionary questions such as correlating gene content and lifestyle (Richards, 2015).

While these are admirable goals and endpoints of genome projects,

along the way data is generated that is useful in its own right, and– more importantly– experiments fail. Within this thesis alone, one chapter began as an attempt to sequence a whole genome, but it was not possible to clear some initial hurdles. But even though experiments fail, and genomes are not actualised, data is generated that can answer questions as well. What may help conceptualise this is the life cycle of a genome project (**Box 2**).

Box 2 The life cycle of a genome project

Proposed by Alexie Papanicolaou in 2016 (Papanicolaou, 2016), the idea of the genome project life cycle builds from Papanicolaou's experiences in multiple genome projects, including the i5k Initiative. This initiative formally launched in 2011 with the goal of sequencing 5000 insect genomes, though the pilot phase had a more modest intention with 30 projects committed to by the Baylor College of Medicine's Human Genome Sequencing Center (Evans et al., 2013; Robinson et al., 2011). As of November 2019, there are 66 insect genomes on the i5K Workspace, with several projects still ongoing (Childers, 2019; Poelchau et al., 2015). Papanicolaou discusses a variety of insights with regards to a genome project having a life cycle, beginning with the question being asked. As discussed previously, the utility of the genome depends on scale and supporting information available, and this reality needs be to be taken into consideration if a genome project will loop through the life cycle at least once.

As seen in Figure 2, a genome project enters a cycle, going from design and sequencing to assembly. Once assembly is complete, two different stages of annotation are started: structural annotation (what genes are where?), followed by functional annotation (what do the genes do?). After annotation comes the most time-consuming step and the most subjective to move past: the community or manual curation. The cycle then continues to dissemination, where the genome is shared with the (scientific)

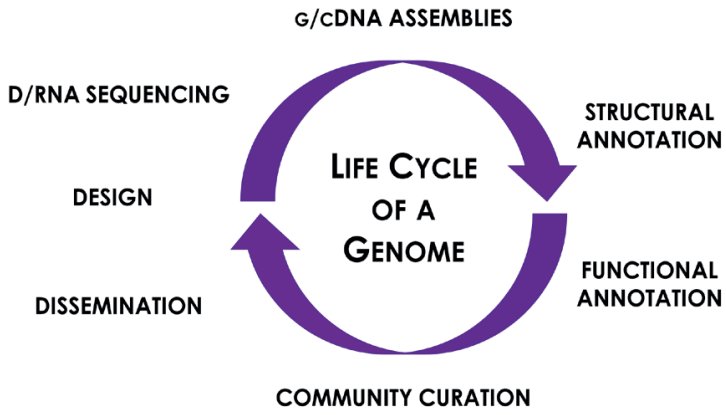


Figure 2. The iterative process of generating a genome sequence can be seen as a life cycle. After Papanicolaou 2016.

public at large. While many would see dissemination as a final goal of a project, Papanicolaou instead imagines it as cyclical, with improvements possible with each iteration of the cycle– and each cycle and the individual steps in them– as useful products as well (Papanicolaou, 2016).

Papanicolaou identifies a variety of insights that can be gained by viewing a project within this framework, including the quality of the work involved, the idea of the "perfect" genome (when is it good enough for release?), and the importance of manual curation through a community approach (Papanicolaou, 2016). This last section may be tricky to scale across projects, as manual curation requires time, skill, and resources that may not be within the wheelhouse of all researchers interested in a genome.

The protocols and workspaces of the i5K are also available for use by those outside of the initiative, but without that large group

of researchers to call upon for specific needs, such as manual annotation, it is difficult to scale the i5K model down to smaller research teams, especially those new to genome work. At some point, you need to go back to the question at hand and see what is required to answer the question and how many resources are available for each part of the genome project life cycle.

Contained within this thesis are projects that have made it through the cycle, though some phases were missed. This begs the question whether the life cycle of a genome project, or at least the one proposed by Papanicolaou (Papanicolaou, 2016), is 1) feasible for every genome project, and 2) necessary.

Bringing biological control and genetics together

There are several intersections between biological control and genetics, such as the application of molecular techniques to identify pests and BCAs alike (Rugman-Jones and Stouthamer, 2017; Stouthamer et al., 1999), track the utility of rearing populations (Ciosi et al., 2014), or address broader enquiries such as evolutionary stable strategies (Holt and Hochberg, 1997) for stable communities. Indeed, the idea to use genetic knowledge and applications to improve biological control is not recent, and has been proposed for all types of biological control over the past twenty years, including IPM (Agunbiade et al., 2013; Kruitwagen et al., 2018; Roderick and Navajas, 2003; Stouthamer et al., 1992; Webber et al., 2015).

More recently, calls for bringing together genomics and biological control have focussed on determining and using existing genetic variation alongside next generation sequencing methods to improve biological control (Lommen et al., 2017). This call from Lommen, de Jong, and Pannebakker is more specific to augmentative biological control, where established and/or native species are given more importance (Lommen et al., 2017). This is partially due to the implementation and perceptions of the Nagoya Protocol (NP), a 2010 supplementary agreement to the Convention on Biological Diversity

(CBD), which aims to provide fair and equitable sharing of benefits that arise from the use of genetic resources (CBD Secretariat, 2011). In addition to the aforementioned risks of introductions, the legal intricacies of NP fueled momentum for projects that focused on improving existing resources (Lommen et al., 2017).

Breeding Invertebrates for Next Generation BioControl (BINGO)

Beginning in 2015, Breeding Invertebrates for Next Generation BioControl (BINGO) was a Marie Skłodowska-Curie Innovative Training Network that aimed to improve established BCAs using genetic and genomic tools (BINGO-ITN, 2015). BINGO focussed on the sustainable and efficient use of BCAs to reduce the impact of invasive and endemic agricultural pests. The numbers for the project are impressive: 24 senior researchers, 13 PhD projects, and 12 partners from academia, industrial, and non-profit organizations throughout the EU.

This EU Horizon 2020 funded project contained several projects that together aimed to improve the production and performance of BCAs using genetic variation, including the use of genomic techniques. Improvement of BCAs in this context includes not only the performance of BCAs, but also the monitoring, risk-assessment, production, and transportation. The questions that could be answered within the BINGO-ITN included tracking cryptic species with molecular markers, determining the relative genetic variation in commercial lines *versus* their wild counterparts, and understanding the genetic underpinnings of basic biology and ecology of certain BCAs.

The BINGO project formally finished in January 2019 and this thesis represents the outcome of one of those projects, RP01, originally titled, "Population genomics of natural enemies." At the time of the start of this project, there were two registered BCAs with a published genome in the NCBI's Genome Database (NCBI, 2019), and these represented a starting point for my project in particular. While the primary goals of the project were the generation of genomes for three BCAs, alongside these genomes was the possibility of collaboration with other BINGO researchers, as they were working on the same organisms. The original three species were *Trichogramma brassicae*,

Nesidiocoris tenuis, and *Amblyseius swirskii*. Two additional species entered the project later: *Trichogramma evanescens*, and *Bracon brevicornis*.

Trichogramma spp.

The chalcidoid *Trichogramma brassicae* (Bezdenko) (Hymenoptera: Trichogrammatidae) (Figure 3) is an egg parasitoid that is minute in size (~0.5 mm in length) (Smith, 1996). It is found throughout Europe and Asia, and is used world-wide as a BCA against a variety of lepidopteran pests in a wide range of crops, including the European corn borer (*Ostrinia nubilalis*) (Polaszek, 2009; Smith, 1996). Beyond biological control applications, *T. brassicae* has been used in a variety of genetic (Laurent et al., 1998; Wajnberg, 1993) and ecological studies (Cusumano et al., 2015; Fatouros and Huigens, 2012). In collaboration with BINGO researchers and partners in Germany, we obtained a sample of *T. brassicae* that was infected with *Wolbachia*. This well-known endosymbiont of insects (Werren et al., 2008) likely plays a role in the *T. brassicae* line being able to reproduce asexually via thelytokous parthenogenesis, where an unfertilized egg results in a diploid female instead of the expected haploid male.



Figure 3. *Trichogramma brassicae*, credit: Nina Fatouros

Along the way, in addition to *T. brassicae*, we were given access to both commercial and wild-caught lines of congeneric *Trichogramma evanescens* (Westwood) (Hymenoptera: Trichogrammatidae) (Figure 4). Similar biological control research has been performed with *T. evanescens* (Fatouros et al., 2007; Fatouros and Huigens, 2012; Huigens et al., 2009), and has been used to study insect neurology (van der Woude et al., 2013; van der Woude and Smid, 2016). *T. evanescens* has a similar sized geographic range as *T. brassicae*, though is found in more Western European regions and Northern Africa but not so much of Asia (Polaszek, 2009). The *T. evanescens* lines in our work were collected in German cabbage fields with no previous history of biological control, while other lines



Figure 4. *Trichogramma evanescens*, credit: Nina Fatouros

are from a commercial source. This collection of *T. evanescens* lines represents different locations, population structures, and time spent within a laboratory or commercial breeding setting.

It is worth noting that the species status of both *T. brassicae* and *T. evanescens* is difficult to clarify. This is partly due to *Trichogramma* being a cryptic genus and partly due to inconsistent accounts in the original collections: *T. evanescens* in the first half of the 19th century and *T. brassicae* in the first half of the 20th century (Polaszek, 2009; Triapitsyn, 2015). While *T. brassicae* only has one scientific synonym to its name, the now defunct *T. maidis* (debated, Li *et al*., 2004), *T. evanescens* has at least 15 scientific synonyms (Polaszek, 2009). From Polaszek, "[t]he true identity of *T. evanescens* is still uncertain," and this uncertainty also affects the molecular identification of both species. The difficulty in identifying *Trichogramma* species through molecular means is well documented, going through several different stages of development and practical application (Stouthamer et al., 1999; Rugman-Jones and Stouthamer, 2017), but nearly always requiring a reliable library of voucher sequences to compare results to (Nasir et al., 2013; Venkatesan et al., 2016).

Nesidiocoris tenuis

Nesidiocoris tenuis (Reuter) (Heteroptera: Miridae) (Figure 5) is a zoophytophagous mirid used throughout the Mediterranean Basin in tomato greenhouses and open-air fields against pests such as whitefly, thrips, and tomato leafminer (Calvo et al., 2009; Mollá et al., 2014). While *N. tenuis* is an important BCA in Spain, it is often considered a pest in other countries (Calvo et al., 2009). This brings us back to what "zoophytophagous" means: feeding mostly on animals but can also feed on plants. When the prey density gets below a certain threshold, the mirid will switch from hunting

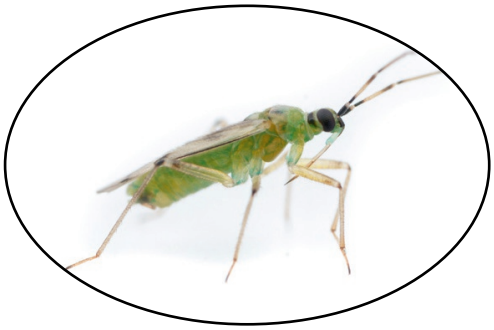


Figure 5. *Nesidiocoris tenuis*, credit: Jitte Groothuis

prey to consuming the tomato it is supposed to be protecting (Sanchez, 2009). The damage ranges from necrotic rings and leaf wilt to spontaneous flower abortion (eventual fruit loss), all in addition to facilitating the spread of plant pathogens due to this damage (Pérez-Hedo and Urbaneja, 2016). This tomato-*Nesidiocoris* problem is well-documented, making up the largest proportion of research on the damage caused by omnivorous arthropods (Puentes et al., 2018). In determining whether there is a possibility to select for *N. tenuis* that switch to phytophagy at a lower prey density, knowing more about the genome and genes in particular can help any selective adaptation programs. Recent work looking into phytophagy and zoophagy indicated that there is genetic variation within these phenotypic differences, and selection may be possible (M. Chincilla-Ramírez, pers. comm).

Amblyseius swirskii

Amblyseius swirskii (Athias-Henriot) (Acari: Phytoseiidae) (Figure 6) is a predatory mite that has an unknown natural range, though the source of the commercial line and other collections is from Eastern Mediterranean populations (Paspasti et al 2019; *cf* **Chapter 3**). *A. swirskii* is a major BCA, becoming both a commercial and biological control success within ten years of its availability on the market (Calvo et al., 2015; Knapp et al., 2018). This is largely due to its effectiveness against two of the most problematic greenhouse pests, whitefly and thrips, while being able to create a "standing army" when



Figure 6. *Amblyseius swirskii*, credit: Jitte Groothuis

pests are scarce with supplemental feeding on pollen (Calvo et al., 2015). However, crops with trichomes on the stem tend to complicate things for the establishment and pest control ability of *A. swirskii* (Buitenhuis et al., 2014). Due to both the exudates and physical barrier presented by some types of trichomes, *A. swirskii* is not able to establish on crops such as tomato (A. Paspati, pers. comm.).

Bracon brevicornis

Similar to its previously mentioned *Trichogramma* parasitoids, *Bracon brevicornis* (Wesmael) (Hymenoptera: Braconidae) (Figure 7) comes from a messy classification history involving changing genus names and eventual synonymizing with sister species *Bracon hebetor* (Kittel and Maeto, 2019). However, a recent revalidation study indicates that there are indeed two separate species, *B. brevicornis* and *B. hebetor* (Kittel and Maeto, 2019), undoing this synonymizing. Research into the biological control potential of *B. brevicornis* is positive: it has a wide range of larval hosts (Temerak, 1983), can be used against storage pests (Kares et al., 2010), and is cosmopolitan in distribution (Chen and van Achterberg, 2019). Research currently focuses on determining basic ecology (Villacañas de Castro and Thiel, 2017), setting the stage for future biological control use following registration. *B. brevicornis* also offers insight into complementary sex determination (CSD) (Speicher and Speicher, 1940), a form of sex determination that involves many alleles at a single locus that is found in a range of Hymenoptera (Heimpel and de Boer, 2008). CSD limits inbreeding because it results in the production of just one sex (Heimpel and de Boer, 2008), which has both implications for commercial biological control and genetic studies alike.



Figure 7. *Bracon brevicornis*, credit: Nils Linek

Know your enemy: Genomes of biological control agents

Aims of the thesis

The aim of this thesis is to present and detail the genomic results and resources that have been generated from these projects, complete with context, utilisation, and suggestions for future applications. The improvement of invertebrates used for biological control is not only possible, but necessary, and one area to begin is by generating genomic resources, including but not limited to genomes. My approach has been to focus on the applied side of genomics. By using existing software, pipelines, and web-based tools, I hope to show what is possible and achievable for genome studies within the biological control context alongside efforts to make my science as open as possible.

Thesis outline

In **Chapter 2**, I present the results of a systematic review on the genetic variation of traits relevant to biological control. This review was born out of a position paper from the BINGO-ITN Network (Leung et al., 2019a) and seeks to determine the current state of research on the genetic variation of phenotypic traits within biological control research. It is important to know what has been done and highlight any gaps. In this chapter, I establish that while there are a variety of BCAs and traits being assessed, there is a general lack of reporting for important values such as heritability or evolvability. In order to facilitate the improvement of BCAs using existing variation, it is important to first outline what traits are being studied, whether they are desirable for biological control, and whether they are heritable.

The rest of this thesis, this anthology, has been arranged in order of end product: what genomic resources were achieved within a project. This order is not in order of importance or size of output, as the true impact of each project will be felt in the future uptake of the resources generated. **Chapter 3** does not contain a genome, but rather genomic resources. Both **Chapter 4** and **Chapter 5** are written as genome reports, short missives that are intended for peer-review but do not need additional experiments or investigations for a traditional scientific article. In **Chapter 6**, genomic resources from the

preceding chapter are used in addition to a *de novo* genome assembly for a population genetic/genomic study. Finally, **Chapter 7** is what I would consider "genome+": a fully assembled and annotated genome with additional experiments and investigations to support the genome assembly; the classic genome manuscript.

In **Chapter 3**, I worked with *A. swirskii* and a relatively new sequencing technology, the Oxford Nanopore MinION flowcells and benchtop sequencer. Using an inbred line, the overall output of the whole genome sequencing was initially disconcerting, as it was not enough for genome assembly. However, it was more than enough to mine microsatellite data, and was used in pooled microsatellite analysis to compare wild-caught populations of *A. swirskii* to the commercial counterpart.

Chapter 4 presents the *Bracon brevicornis* genome, a linked-reads assembly from a fairly inbred line maintained in Germany. In using linked-reads, we were able to use a reduced number of individuals, resulting in a largely homozygous read library. This genome features an *ab-initio* annotation, building on the existing annotations of other braconid genomes. Though largely unknown, the gene or mechanism for CSD was searched for within the assembly in zones of heterozygosity by using the pseudohaploid arms of the initial assembly, a feature unique to linked-read methods.

In **Chapter 5**, a hybrid sequencing strategy was deployed for *Trichogramma brassicae*, where short and long reads were used in conjunction to create an assembly. Additional sequencing of the transcriptome allowed for a combined annotation strategy of *ab-initio*, homology, and evidence-based annotation. Some additional assessments include protein comparison with congeneric *T. pretiosum*, as well as confirming the homozygous nature of our *Wolbachia*-aided line.

Chapter 6 is an initial exploration into the population genetics and genomics in several wild-caught and commercial *T. evanescens* lines. This was achieved through two methods. The first used sequence data from a sister species, obtained in the previous chapter, to mine for microsatellites. The second involved pooled sequencing that resulted in an unannotated

de novo genome for *T. evanescens* alongside sliding window analyses for genome-wide comparisons. These two approaches, individual genotyping and pooled samples, offer different insights into the different lines, namely how genetically variable they are, and will yield more insights down the road.

The most detailed genome study in this thesis can be found in **Chapter 7** with the linked-read genome of *Nesidiocoris tenuis*. This genome was derived from a single adult female, drastically reducing the time and physical resources that would have been necessary for setting up an inbred line. Decontamination of the assembly resulted in the identification of potential endosymbionts, as well as the indication of regions of potential lateral gene transfer for future investigation. Additional supporting investigations include cytogenetics to determine the karyotype and sex chromosome complement of *N. tenuis,* to localize unique repetitive probes as well as investigate the absence or presence of telomeric repeats in *N. tenuis*. Finally, the investigations are completed by a population genomic analysis using previously generated whole genome sequencing to test the utility of the genome.

In **Chapter 8**, I will wrap up this anthology and address the chapters as a whole, how they connect to each other, and any insights gained from the project in general. Furthermore, I will take the outcomes of these chapters and speculate on future research, both necessary or possible. Finally, I intend to address some lingering aspects of genome research and biological control in general that have been brought up in this introduction.

To my friends I make the gift, but this sweet-voiced garland of the Muses is common to all the initiated[3]

3 from Meleager's *Garland*, AP 4.57, sourced from http://www.attalus.org/poetry/meleager.html with reference to the original *Anthologia Palatina*


THEN, IN 2019 ...
WHAT??
WE READ 3,000 PAPERS AND STILL DON'T HAVE ENOUGH DATA ...
... FOR A META ANALYSIS!!!
WELL, TO BE FAIR IT'S ONLY 2,927 PAPERS!
SO, HOW ABOUT A SYSTEMATIC REVIEW?
www.comiclaboratory.com
Stefan Christ

Chapter 2

Genetic variation of traits in natural enemies relevant for biological control: a systematic review

Kim B. Ferguson[a], Sophie R. Chattington[b], Wouter N. Plouvier[a,c], and Bart A. Pannebakker[a]

[a] Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands
[b] University of Bremen, Institute of Ecology, Population and Evolutionary Ecology Group, Bremen, Germany
[c] INRA, CNRS, Sophia Antipolis, France

This chapter has been submitted for publication and is available as a preprint, DOI: 10.20944/preprints202001.0276.v1

Abstract

The concept of genetic improvement in relation to biological control involves the exploitation of natural genetic variation for the benefit of existing biological control agents (BCAs). Despite recent calls for this process to be adopted in biological control research, there is no clear overview of the current state of research into genetic variation within a biological control context, including quantifiable estimates such as narrow-sense heritability (h^2). In this systematic review, we aim to determine the current state of research on the genetic variation of biological control traits in natural enemies. After the searching process, screening for papers that can deliver on our research question reduced the initial 2,927 search hits to a mere 69 papers for data extraction. Of these, the majority (73.6%) did not report quantitative values for genetic variation. Extracting the traits measured in these papers, we categorized them according to two approaches; the first related to fitness components, and the second related to biological control importance. This systematic review highlights the need for more rigorous reporting of the quantitative values of genetic variation to enable the successful genetic improvement of biological control agents.

Introduction

With a rising global population, as well as a changing global climate, food production is increasingly under pressure (IPCC, 2007; Misra, 2014). Rising arthropod pest pressure (due to modern monoculture practice as well as expanding pest ranges related to global change (Das et al., 2011)), coupled with a desire for more environmentally sustainable agriculture (i.e. using less pesticides) (Rigby and Cáceres, 2001), has led to an increased interest in biological control for pest management (van Lenteren et al., 2018). The use of arthropod natural enemies as biological control agents (BCAs) is not new, with records of first releases dating back over 100 years (Barratt et al., 2018).

For several decades, imported exotic natural enemies were relied upon to control non-native pests, which are becoming increasingly more prevalent due to expanding or shifting ranges linked with changing climates (Aguilar-Fenollosa and Jacas, 2014; van Lenteren et al., 2006). By taking the natural

enemy from the native range of the pest, it was considered to be better adapted to tackling the pest due to their shared evolutionary past. Due to knowledge gained on the risks of releasing non-native species, along with stricter regulations on the import and export of species (such as the Nagoya Protocol on Access and Benefit Sharing, 2010), this practice has become less desirable (CBD Secretariat, 2011; Lommen et al., 2017). These recent developments in combination with an expansion of pest ranges due to global warming and the anthropogenic movement of produce across the globe, make it more important than ever to improve the efficacy of those native natural enemies that are already at our disposal (Leung et al., 2019a; Lommen et al., 2017).

Selection of effective BCAs is usually determined at species level, by looking at interspecific variation (Lommen et al., 2017), i.e. finding the best species to tackle a certain pest. More recently, there is increased attention on exploiting the large amount of intraspecific variation between, and even within, populations (Kruitwagen et al., 2018; Lommen et al., 2017). Recent studies have also suggested the use of genetic modification for increased effectiveness of BCAs, particularly in fungal and pathogen-based agents (Karabörklü et al., 2018; Lovett and St. Leger, 2018; St. Leger and Wang, 2010), as well as insects (Poppy and Powell, 2009; Routray et al., 2016).

However, as with using genetically modified crop plants, the release of genetically modified natural enemies would come under strict regulation (Barratt et al., 2018; Reeves et al., 2012) and does not seem compatible with the eco-friendly mission of biological control. On the other hand, artificial selection of traits is an accepted method already commonly used across agriculture, such as in livestock and plant breeding (Brotherstone and Goddard, 2005), which can be exploited to effectively improve BCAs (Kruitwagen et al., 2018; Lommen et al., 2017).

A recent review on the use of experimental evolution in biological control supports the potential of artificial selection for improving BCAs (Lirakis and Magalhães, 2019). The key prerequisite for artificial selection is the presence of genetic variation for the trait of interest. The prime estimate of genetic

variation is heritability, which is defined as the proportion of the total variation between individuals in a defined population that is due to genetic variation (Lommen et al., 2017).

Heritability or evolvability is regularly studied within the realm of basic and evolutionary research, and often so with insects due to the ease of their rearing and workability in a controlled laboratory setting. However, it has been studied to a lesser extent within an applied context. This can be to the detriment of BCAs and their effective use, both in regards to deleterious impacts of genetic drift and subsequent loss of efficacy (Paspati et al., 2019; Szűcs et al., 2019), and that without variation data for traits of interest, there is no way of knowing if the trait in question can be improved through genetic selection (Houle, 1992). Full reporting or values such as variance and determining evolvability in place of heritability values was proposed more than twenty years ago (Hansen et al., 2011; Houle, 1992) as a more appropriate measure, yet the uptake has been slow and inaccurate calculations remain a concern (Garcia-Gonzalez et al., 2012).

With the increased interest in the role of experimental evolution and artificial selection in the improvement of biological control (Kruitwagen et al., 2018; Leung et al., 2019a; Lirakis and Magalhães, 2019; Lommen et al., 2017; Routray et al., 2016), it is necessary to determine what has already been assessed with regards to genetic variation and heritable traits of BCAs, and where there are gaps in our knowledge.

In this systematic review, our goal is to compile and summarize findings from across literature regarding genetic variation in BCAs for traits that are potentially useful for biological control, regardless of whether these studies have an applied focus or not. This will identify the advances already made with regards to identifying suitable heritable traits within arthropod BCAs, and where more work is necessary.

Methods

Search Criteria

As the steps of a meta-analysis and systematic review are similar up until the data extraction phase (Côté and Jennions, 2013) for searching, screening,

and the extraction protocol, we followed the steps outlined in Chapter 2 of the *Handbook of Meta-Analysis in Ecology and Evolution* (Koricheva et al., 2013). An overview of the process is provided in Figure 1. Our question was refined to a single search term that focused on two aspects: BCAs and genetic variation or heritability. We did not wish to limit our initial search results by *a priori* defining traits of interest.

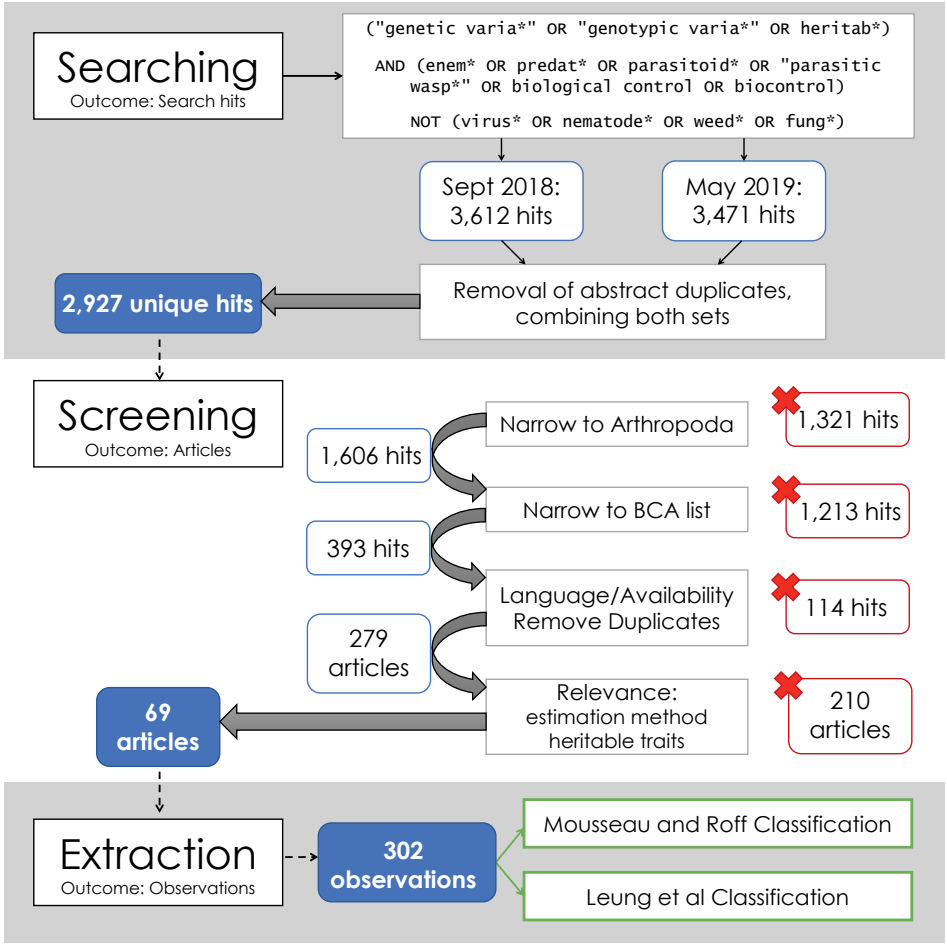


Figure 1. Flow diagram of the searching, screening, and extraction process used in this systematic review. Classification step is based on fitness components from Mousseau and Roff (Mousseau and Roff, 1987) and aspects of biological control research from Leung *et al.* (Leung et al., 2019a)

The databases used were CAB Abstracts (1910 to 2019), the Biological Abstracts database (1969 to 2016), and the Agricola database (1970 to April 2019), which were all accessed via the CAB Abstracts search platform. Additional search parameters were chosen, including using the "field= keyword" search field (searching for terms in the title, abstract, or keywords), as well as limiting document type to 'Article' and 'Review.' No language restrictions were imposed at the time of the search.

To validate our search term, a positive control group was set up of 15 papers that would suit the purpose of initial screening for our research question. This set of controls can be found in the supplementary materials (Supplementary Table 1).

As we were interested in finding publications that assessed genetic variation or heritability of different traits of BCAs, the search term contained various terms and wildcard symbols (*) to aid in the search, as well as the AND/OR operator to tie the two ideas together. To limit the search to arthropod BCAs, the NOT operator was used in conjunction to avoid the inclusion of several undesired returns.

Consequently, the final search term was as follows:

("genetic varia*" OR "genotypic varia*" OR heritab*)

AND (enem* OR predat* OR parasitoid* OR "parasitic wasp*" OR biological control OR biocontrol)

NOT (virus* OR nematode* OR weed* OR fung*)

Initial search hits were deduplicated using the duplication removal tool found within the CAB Abstracts search tool (based on Abstract). The total bibliographic information for these hits were downloaded and arranged into a table for further screening. Search queries were performed on September 17, 2018 and May 16, 2019. Duplicate entries overlapping both search periods were removed manually, while some duplicates remained in the dataset until screening. Importantly, all positive controls were returned in the search results.

Screening search results

Unique entries were subject to a preliminary screening that sought to determine, based on title, abstract, and any keywords whether the entries were appropriate for the systematic review. This initial screening was performed by a team of 16 readers, and was based solely on content related to study species: first, if the species of study (as determined in the title or abstract) was truly an arthropod; and second, if the species of study was classified as a BCA by checking whether it appeared on either the EPPO list of BCAs available for use in Europe (PM 6/003, both 2016 and 2019 versions, https://gd.eppo.int/standards/PM6/), or on the list of BCA species compiled in previous reviews (van Lenteren, 2012; van Lenteren et al., 2018). This list contains 361 species names, some of which are former scientific names that are no longer in use but allow for scanning older papers.

Two additional screening measures occurred: on language and availability. First, language was either listed as a variable from the database bibliographical information or was assessed by viewing the full text of the entry. Languages that were understood by our team of readers were assessed, while languages not understood by the team remained unassessed. Second, unavailable entries were listed as 'missing', that is, while the entry existed in one of the three databases used for the initial search, neither electronic nor physical records in the Wageningen University Library network were available and additional online queries yielded no records. Remaining duplicates (based on entry title), as well as conference proceedings, were removed manually.

At this point, each entry in the dataset corresponds to a single peer-reviewed article. The final screening category was 'relevance'. Relevance to the review terms was evaluated based on the article's abstract, title, or full-text in some cases, and was determined based on two parameters: 'methods' and 'trait of interest'. If the methods allowed for determining of genetic variation, the article was deemed relevant. Such methods would include the use of isofemale lines, tracking over multiple generations, common garden set-ups, parent-offspring regressions, or a type of sibling or crossing strategy. For the second parameter, 'traits of interest', it was the presence of any kind of phenotypic trait being studied that was important. For instance, articles

that solely focused on population structure via neutral molecular markers, but were not looking at phenotypic traits, were deemed unrelated to this systematic review. All articles determined to be relevant were then subjected to full-text assessment and data extraction.

Data extraction and categorization

In-depth data extraction was performed by six readers, with a follow-up by two readers to double-check the values of the extractions. Each article was viewed in print and assessed for the following aspects: 1) estimation method, 2) trait of interest, 3) species, and 4) population values (such as size and number of populations). When applicable, a wide range of variables related to each trait and population that were assessed in the study were extracted as well. Interest was paid to values directly informative of genetic variation, such as narrow-sense heritability (h^2), broad-sense heritability (H^2), or evolvability (CV_A) estimates. For a single article, several unique data extraction outputs (termed 'observations') are possible, and all received a unique identification number in order to track the total data extracted. This granularity allows for assessing how often a trait is measured for different populations or species in various studies, as multiple unique values related to genetic variation of traits of interest are possible for a single article.

For each trait of interest, categories were assigned to each observation to refine the extraction data to clear themes. Two categorical systems were used. The first system uses the categories of phenotypic traits based on biological function as defined by Mousseau and Roff (1987): physiology (P), morphology (M), life history (L), and behaviour (B) (Mousseau and Roff, 1987). These categories are explicit, and each trait can only be attributed to a single category. The second system is based on the application of biological control and uses categories as defined in Leung *et al.* 2019, which posits four aspects of biological control that could be improved using genetic knowledge: pest suppression ability (1), adaptation to abiotic factors (2), reducing ecological risk (3), and improving mass rearing conditions (4). Three additional categories were created to address the outcomes of this extraction, and include: insecticide resistance (5), adaptation to biotic factors (6), and a null

category named 'unrelated' (7), as to avoid forcing a trait into one of the other six categories. This second classification is not explicit, and one trait may be attributed to multiple categories. Simple calculations (counts and means) were performed manually, while figure generation was performed using R (version 3.5.0), R Studio (version 1.2.1335), and ggplot2 (version 3.2.0)(R Core Team, 2015; RStudio Team, 2015; Wickham, 2016).

Results

Search results

The first search using the full search term yielded 3,612 hits but was reduced to 2,880 following deduplications (Figure 1). The second search, carried out eight months later, yielded 3,471 hits, and deduplication reduced it to 2,891. This includes new articles published as well as any entries added to any of the three databases within the intervening period. From the second search, 87 unique hits were found that cover the period between the first and second searches (September 2018 to April 2019). Overlapping hits were removed by hand, resulting in a total of 2,927 hits for the systematic review. It is worth noting that some hits that were in the first search were no longer present in the database for the second search. The complete dataset of search returns at this point are available in the supplementary materials (Supplementary Table 2) in the format delivered by the CAB Abstracts Database.

Results of this combined search return, according to year of publication, can be found in Figure 2. Year of publication ranges from 1914 to 2019. Several years in the earlier time of the search returned no hits, representing the true minimum. Fifteen different years are represented by only a single hit, while the maximum belongs to 2014 with 181 entries. Values skew closer to the present-day, likely attributed to both the well-documented increase in scientific publishing (Larsen and von Ins, 2010) and a general increase in biological control studies. Similar results occurred in a recent review on zoophytophagous predators (Puentes, Stephan, and Björkman, 2018). The majority of returns are from the CAB Abstracts database (1304), which was chosen as the preferred database for the search and deduplication, followed by Biological Abstracts (969), and Agricola (654).

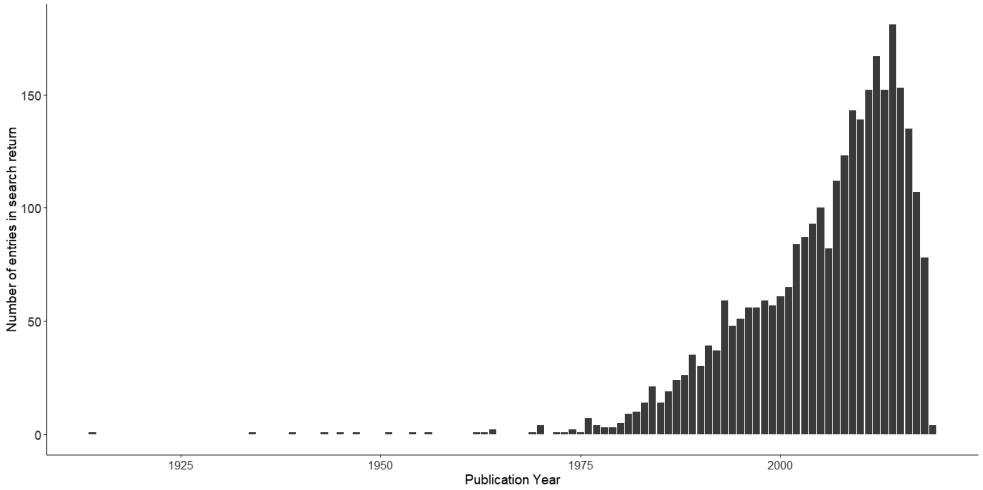


Figure 2. Initial search returns displayed according to publication year (n = 2, 927). Absent bar indicates zero search hits for that year.

SCREENING

Preliminary screening was performed to remove all the entries that were non-arthropods, not on the EPPO or van Lenteren lists, or those that were problematic (duplicates, missing, or languages unavailable to readers). From the 2,927 entries, only 1,606 were entries containing arthropod research (1,321 non-arthropods were removed). From these 1,606, only 393 were determined to be species that were on the combined EPPO and van Lenteren 2012 list of BCAs. All remaining duplicates were removed manually (94) along with conference proceedings (3), resulting in a total of 296 articles.

At this point, several articles were identified as problematic, either for availability or language. Four papers were identified as missing and were removed from the dataset. Eighteen articles were not in English. Most of these articles were in Chinese (9), followed by Portuguese (3), Russian (3), Czech (1), German (1), Persian (1), Spanish (1), and French (1). Where possible, articles were assessed by native speakers with a biological background to determine relevance, followed by non-native but fluent speakers with a biological background. All foreign language papers could be assessed with the exception of all three Portuguese papers and one Russian paper. These

four unassessed papers were deemed unrelated, followed by an additional eight papers that were assessed. Of the 20 non-English papers, seven were assessed and kept for further screening. This final step reduced the articles in this dataset to 279, all of which are listed in the supplementary materials (Supplementary Table 3).

Extraction

Of the 279 screened articles, only 69 were ultimately deemed relevant to this systematic review, based on the estimation methods and presence of measured traits, and are listed in Table 1. Results from the extraction process, with reduced variable extractions for each extraction output, can be found in Table 2. The amount of observations depends on the species studied, the number of populations assessed, and the trait of interest. For example, one paper may look at a single trait in two species using isofemale lines, which would result in two observations. Within our results, there are 302 observations in total: 137 for parasitoids, and 165 for predators. Per paper, the average number of observations is 4.4, however, a small number of papers exceed 10 observations, and can therefore account for any skewing of popularity in estimation methods or species interest. For example, the most observations per species belong to *Harmonia axyridis* with 50 observations, while only being found in five papers. While the full dataset is informative for traits of interest, other parameters such as year of publication or estimation method will be better observed by reducing to unique observations.

Table 1. List of 69 studies examining the genetic heritability of traits in biological control agents (BCAs) that were extracted for this systematic review. Table continues over 3 pages.

Paper ID[a]	In-text reference	Focal BCA	BCA Type
88	Hufbauer, 2001	*Aphidius ervi*	Parasitoid
299	Guzmán-Larralde et al., 2014	*Trichogramma pretiosum*	Parasitoid
340	Bilodeau et al., 2013	*Aphidius ervi*	Parasitoid
471	Astles et al., 2005	*Harmonia axyridis*	Predator
515	Obrycki et al., 2001	*Hippodamia convergens*	Predator
525	Liu and Smith, 2000	*Trichogramma minutum*	Parasitoid
543	Grill et al., 1997	*Harmonia axyridis*	Predator
564	Rodríguez et al., 1994	*Trichogramma fuentesi*	Parasitoid

Chapter 2

Paper ID[a]	In-text reference	Focal BCA	BCA Type
571	Marples et al., 1993	*Adalia bipunctata*	Predator
577	Sequeira and Mackauer, 1992	*Aphidius ervi*	Parasitoid
768	Navia et al., 2014	*Amblyseius largoensis*	Predator
947	Facon et al., 2011	*Harmonia axyridis*	Predator
962	Sayyed et al., 2010	*Chrysoperla carnea*	Predator
1092	Fukunaga and Akimoto, 2007	*Harmonia axyridis*	Predator
1174	Wang et al., 2004	*Cotesia glomerata*	Parasitoid
1226	Thomson and Hoffmann, 2009	*Trichogramma carverae*	Parasitoid
1286	Hufbauer and Via, 1999	*Aphidius ervi*	Parasitoid
1311	Bennett and Hoffmann, 1998	*Trichogramma carverae*	Parasitoid
1379	van Houten et al., 1995	*Amblyseius barkeri, A. cucumeris*	Predator
1401	Bruins et al., 1994	*Trichogramma brassicae*	Parasitoid
1439	Antolin, 1992	*Muscidifurax raptor*	Parasitoid
1466	Legner, 1991a	*Muscidifurax raptorellus*	Parasitoid
1514	Legner, 1988	*Muscidifurax raptorellus*	Parasitoid
1520	Legner, 1987	*Muscidifurax raptorellus*	Parasitoid
1528	Vianen and Lenteren, 1986	*Encarsia formosa*	Parasitoid
1536	Parker and Orzack, 1985	*Nasonia vitripennis*	Parasitoid
1642	Parreño et al., 2017	*Lysiphlebus fabarum*	Parasitoid
1675	Sepúlveda et al., 2017	*Aphidius ervi*	Parasitoid
1684	Mansoor et al., 2017	*Chrysoperla carnea*	Predator
1851	Kamala Jayanthi et al., 2014	*Cryptolaemus montrouzieri*	Predator
1868	Abbas et al., 2014	*Chrysoperla carnea*	Predator
1899	Paolucci et al., 2013	*Nasonia vitripennis*	Parasitoid
1937	Mansoor et al., 2013	*Chrysoperla carnea*	Predator
2009	Wajnberg et al., 2012	*Trichogramma brassicae*	Parasitoid
2106	Sandrock et al., 2010	*Lysiphlebus fabarum*	Parasitoid
2108	Henry et al., 2010	*Aphidius ervi*	Parasitoid
2141	Nachappa et al., 2010	*Phytoseiulus persimillis*	Predator
2236	El-Heneidy and Shoeb, 2007	*Trichogramma evanescens*	Parasitoid
2278	Pérez-Maluf et al., 2008	*Leptopilina heterotoma*	Parasitoid
2280	Pannebakker et al., 2008	*Nasonia vitripennis*	Parasitoid
2307	Shuker et al., 2007	*Nasonia vitripennis*	Parasitoid
2312	Kamping et al., 2007	*Nasonia vitripennis*	Parasitoid

Paper ID[a]	In-text reference	Focal BCA	BCA Type
2453	Wang et al., 2003	*Cotesia glomerata*	Parasitoid
2465	Wajnberg et al., 2004	*Trissolcus basalis*	Parasitoid
2470	Jia et al., 2002	*Phytoseiulus persimillis*	Predator
2595	Maeda et al., 1999	*Amblyseius womersleyi*	Predator
2601	Wagner et al., 1999	*Harmonia axyridis*	Predator
2633	Lesna and Sabelis, 1999	*Hypoaspis aculeifer*	Predator
2645	Mason and Hopper, 1997	*Aphelinus asychis*	Parasitoid
2657	Wajnberg and Colazza, 1998	*Trichogramma brassicae*	Parasitoid
2675	Gilchrist, 1996	*Aphidius ervi*	Parasitoid
2689	Margolies et al., 1997	*Phytoseiulus persimillis*	Predator
2696	Sorati et al., 1996	*Trichogramma brassicae*	Parasitoid
2707	Henter et al., 1996	*Encarsia formosa*	Parasitoid
2736	Holloway et al., 1995	*Adalia bipunctata*	Predator
2738	Fleury et al., 1995	*Leptopilina heterotoma*	Parasitoid
2741	Henter, 1995	*Aphidius ervi*	Parasitoid
2755	Pompanon et al., 1994	*Trichogramma brassicae, T. cacoeciae*	Parasitoid
2759	Geden et al., 1992	*Muscidifurax raptor*	Parasitoid
2763	Orzack and Gladstone, 1994	*Nasonia vitripennis*	Parasitoid
2765	Holloway et al., 1993	*Adalia bipunctata*	Predator
2781	Tauber and Tauber, 1992	*Chrysoperla carnea*	Predator
2782	Legner, 1993	*Muscidifurax raptorellus*	Parasitoid
2791	Legner, 1991b	*Muscidifurax raptorellus*	Parasitoid
2799	Orzack et al., 1991	*Nasonia vitripennis*	Parasitoid
2845	Boulétreau and Wajnberg, 1986	*Leptopilina heterotoma*	Parasitoid
2846	Tauber and Tauber, 1986	*Chrysoperla carnea*	Predator
3001	Al-Khateeb et al., 2018	*Cryptolaemus montrouzieri*	Predator
3007	Rasmussen et al., 2018	*Orius majusculus*	Predator

[a] Paper identification numbers are unique to each paper and are cross-referenced in Table 2.

Focusing on publication year, the spread of articles that reach the extraction phase is relatively well-distributed from 1979 to 2018 with no clear bias towards the present day (Table 1). Moving on to the BCAs themselves, there are 29 different species studied in this dataset, 13 of which only appeared once (Table 2). The most studied species in our dataset is the parasitoid *Aphidius*

ervi, which was covered in eight articles. The parasitoid *Trichogramma* is the most studied genus of the extraction with 12 articles, followed by *Aphidius*, with eight articles. Considering the genetic variation estimation methods (Figure 3), just over a quarter of observations used isofemale populations to estimate inheritance (27.1%), followed by artificial selection (17.1%).

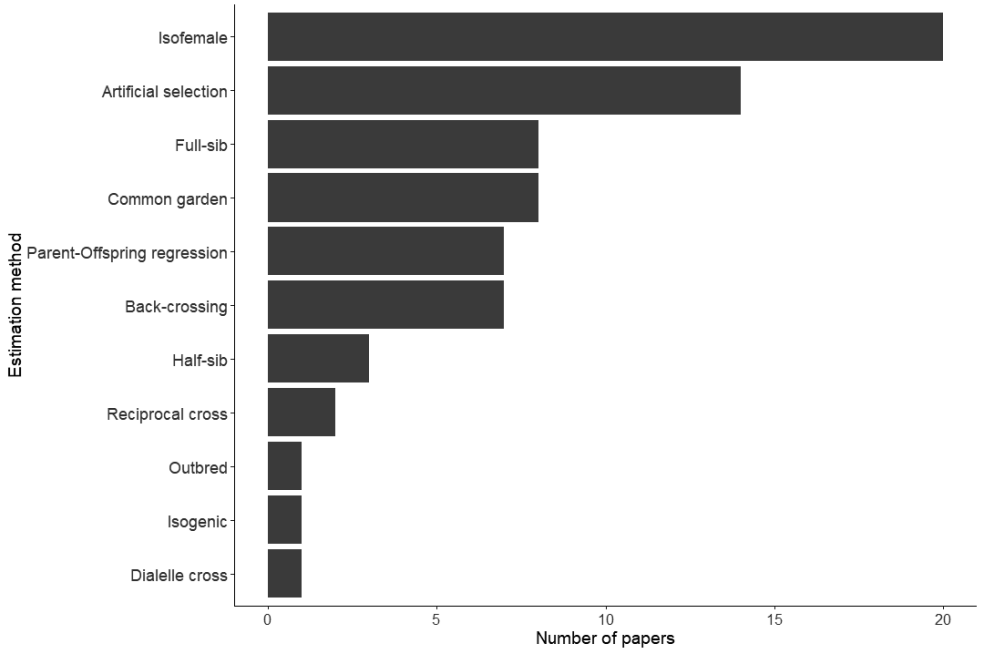


Figure 3. Estimation methods of extracted papers, reduced to unique species per paper (n = 72).

Despite the use of these estimation methods, the majority of papers did not provide heritability estimates. We limited this to values of broad-sense heritability (H^2) and narrow-sense heritability (h^2), as well as evolvability (CV_A). In some cases, these estimates were not reported directly, but we were able to convert the reported genetic or dominance variance into these estimates using other data reported in the paper. Of our 302 observations, only 105 were supported by measures of genetic variation. When reduced to unique species per paper (72, as a few papers cover two species), 19 provided a heritability estimate for the trait assessed as opposed to the 53 (73.6%) that

did not (Figure 4). The most commonly used were either h^2 (9.7%) or H^2 (8.3%), followed by either in combination with CV_A (h^2 and CV_A at 4.2%, and H^2 and CV_A at 2.8%), CV_A on its own (1.4%), or both broad and narrow sense heritability at the same time (1.4%). These amounts include papers where we were able to calculate estimates using reported data in the text. Papers that reported on variations of CV_A nearly always reported CV_G, but in the cases referenced here these were found to be referring to the same measure (CV_A). Meanwhile, the majority of papers did not report any heritability estimates or include values necessary for their calculation.

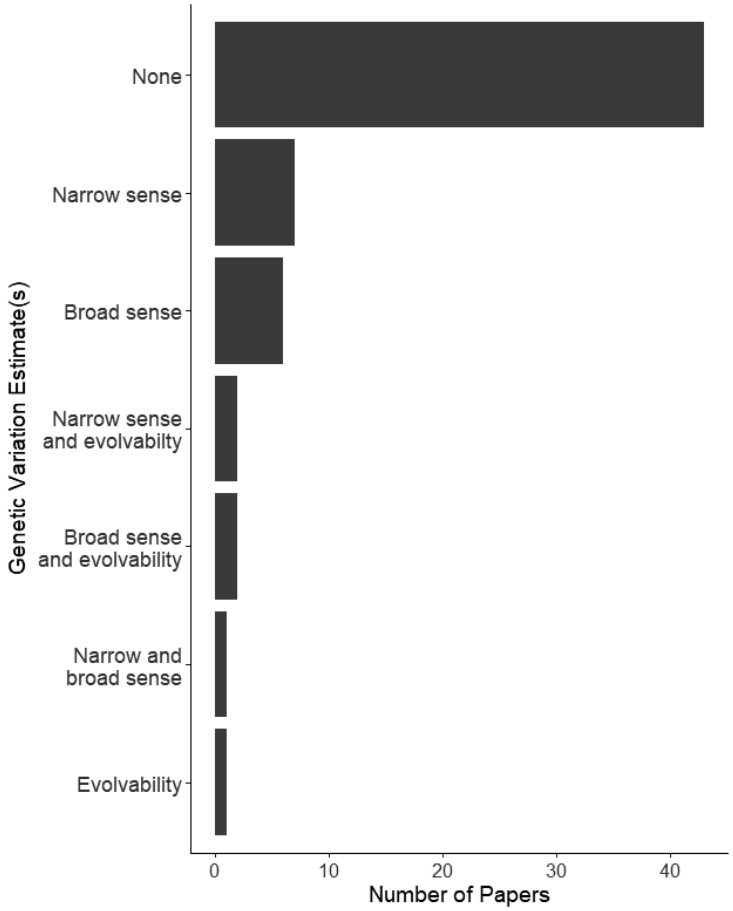


Figure 4. Heritability estimates found in extracted papers (n = 69).

An important goal of our analysis is the categorization of traits according to fitness component, or importance to biological control research. Given the variety of traits being measured, both in nomenclature and formulation, categories were assigned and pooled by observation count in two ways: the Mousseau and Roff assessment based on biological function used in their review of the heritability of fitness components, and the Leung *et al*. 2019 assessment based on biological control application (Mousseau and Roff 1987; Leung et al. 2019). The outcome of the Mousseau and Roff assessment indicates that the majority of the 302 observations were of life history traits (53%), followed by behaviour (21%), morphology (19%), and physiology (7%; Figure 5). Also within the different BCA types, parasitoid or predator, life history is still the most accounted for trait for both types, while behaviour and morphology switch middling positions depending on BCA type.

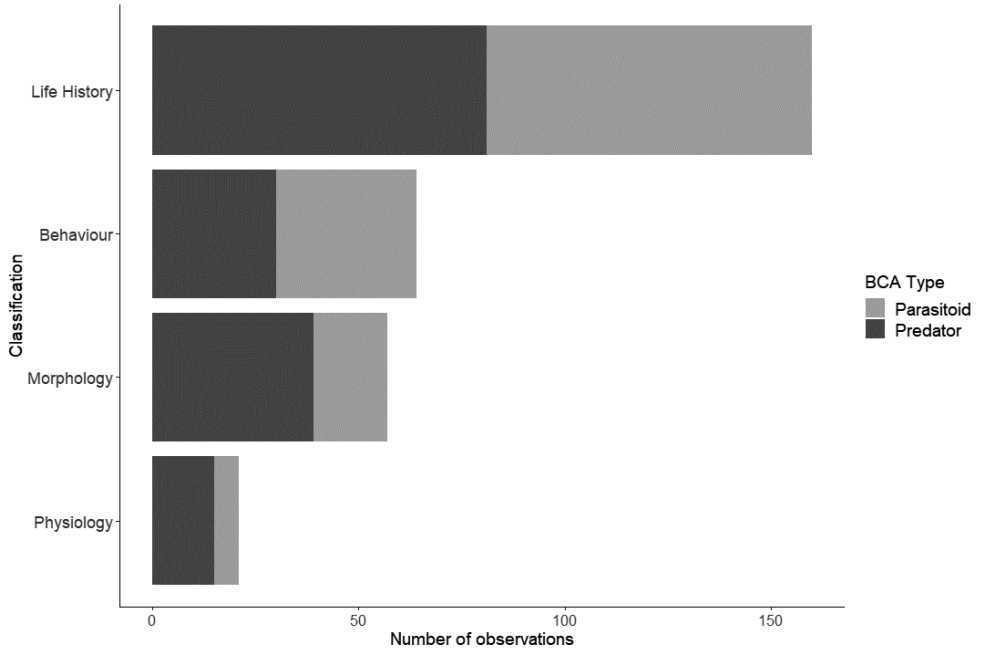


Figure 5. Classification of heritable traits related to fitness, based on Mousseau and Roff (1976), in extracted papers according to BCA type (n = 302).

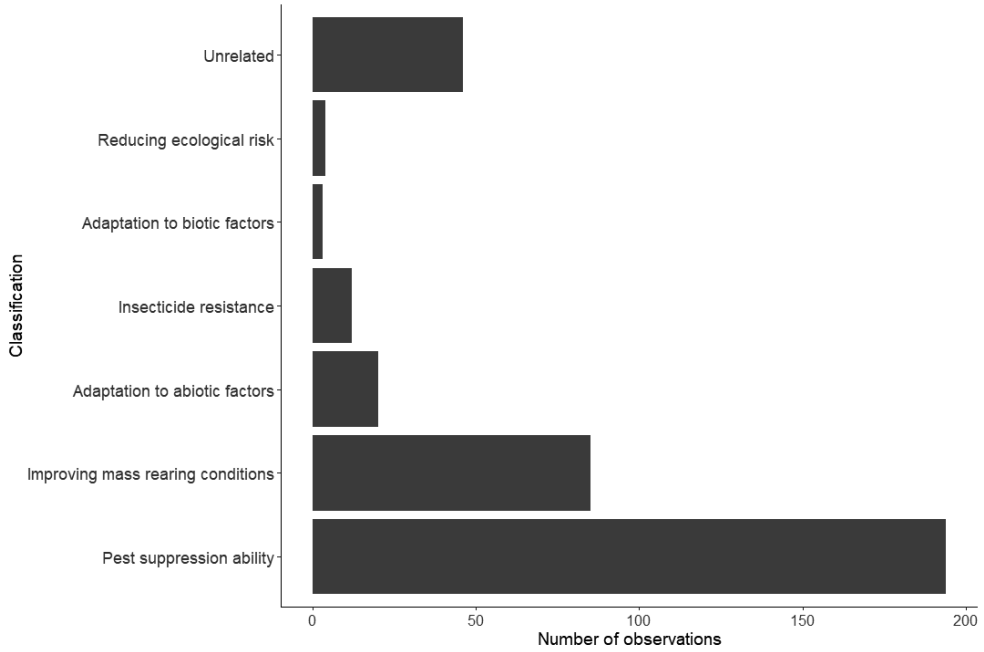


Figure 6. Classification of heritable traits related to importance to biological control research, based on Leung *et al.* (2019), in extracted papers (n = 302).

The second assessment method, based on biological control application (Leung et al 2019), allowed for multiple classifications per observation (364 total). As shown in Figure 6, the majority of the traits measured were related to pest suppression ability (53%), followed by improving mass rearing conditions (23%). The third most common classification of the observations was the 'unrelated' category with 13% of the observations and the traits associated with them were most likely not useful for improving biological control directly.

Discussion

This systematic review sought to identify studies on BCAs that look at the genetic variation and heritability of traits and ordered these traits by two classification systems. According to the Mousseau and Roff classification, a majority of studies dealing with trait heritability focused on life history traits, regardless of BCA type. The other classifications of Mousseau and Roff (morphology, behaviour, and physiology) were also approached in various

Table 2. Extracted observations from 69 studies examining the genetic heritability of life history traits in biological control agents (BCAs), listed according to type of BCA and then by species (n=302). Table continues over 16 pages.

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*Aphelinus asychis*	walking speed	common	none	B	7	2645
	Aphidius ervi	body size, female offspring	full-sibling	h^2	L	7	577
		body size, male offspring	full-sibling	h^2	L	7	577
		body size, offspring mean	full-sibling	h^2	L	7	577
		emergence	parent-offspring reg	h^2	L	1	2108
		emergence	parent-offspring reg	h^2	L	1	2108
		emergence	common	none	L	1	340
		female sex ratio	parent-offspring reg	h^2	L	1;4	2108
		female sex ratio	parent-offspring reg	h^2	L	1;4	2108
		fitness	artificial selection	none	L	1	1675
		handling time	artificial selection	none	B	1	1675
		host adaptation	parent-offspring reg	h^2	L	1;3	2108
		host adaptation	parent-offspring reg	h^2	L	1;3	2108
		mass	full-sibling	none	M	1	2675
		ovipositor contact	common	none	B	1	340
		parasitism	half-sibling	h^2, CV_A	L	1	2741
		parasitism	common	none	L	1	1286
		performance breadth	full-sibling	none	P	2	2675
		prey recognition	artificial selection	none	B	1;3	1675

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*A. ervi* cont.	time to oviposition	artificial selection	none	B	1	1675
		total parasitism	parent-offspring reg	h^2	L	1	2108
		total parasitism	parent-offspring reg	none	L	1	2108
		virulence	common	none	P	1	88
		walking speed	full-sibling	none	B	7	2675
		development time	half-sibling	none	L	7	1174
		efficiency of parasitism	half-sibling	none	L	1	1174
		females per brood	half-sibling	none	L	1:4	1174
		flight orientation	artificial selection	none	B	7	2453
		landing success	artificial selection	none	B	7	2453
		tibia length	half-sibling	none	M	1	1174
	Encarsia formosa	contacts with host	common	none	B	7	2707
		drumming behaviour	common	none	B	7	2707
		encounters	common	none	B	7	2707
		head width	parent-offspring reg	none	M	1:4	1528
		host acceptance	common	none	B	1:3	2707
		host feeding	common	none	B	1	2707
		host rejection	common	none	B	1	2707
		ovariole count	parent-offspring reg	none	L	7	1528
		oviposition acceptance	common	none	B	1	2707
		oviposition posture	common	none	B	7	2707
		successful parasitization	common	none	L	1	2707

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*Leptopilina heterotoma*	cumulative duration of oviposition	common	none	B	1	2278
		locomotor activity	isofemale	none	B	7	2738
		parasitism rate	common	none	L	1	2845
		probing duration	common	none	B	1	2278
		probing latency	common	none	B	1	2278
		response to olfactory stimuli	common	none	B	1	2278
		survival rate	common	none	L	1	2845
		time to find larvae	common	none	B	1	2278
		time to oviposition	common	none	B	1	2278
	Lysiphlebus fabarum	parasitism	isofemale	H^2	L	1	2106
		wing shape	isofemale	none	M	7	1642
		wing size	isofemale	none	M	7	1642
	Muscidifurax raptor	development time egg to adult	dialelle cross	H^2	L	7	1439
		fecundity	dialelle cross	H^2	L	1;4	1439
		fecundity	parent-offspring reg	h^2	L	1;4	2579
		reproductive lifespan	dialelle cross	H^2	L	1;4	1439
		sex ratio	dialelle cross	H^2	L	1;4	1439
		wing length	parent-offspring reg	h^2	M	7	2579
		wing length	parent-offspring reg	h^2	M	7	2579
	Muscidifurax raptorellus	eggs laid	backcrossing	none	L	1	1514

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*M. raptorellus* cont	eggs per gregarious oviposition	backcrossing	none	L	1	1514
		eggs per gregarious oviposition	backcrossing	none	L	1	1520
		eggs per gregarious oviposition	backcrossing	none	L	1	2791
		gregarious oviposition	backcrossing	none	L	1	1514
		gregarious oviposition	backcrossing	none	L	1	2782
		gregarious oviposition	backcrossing	none	L	1	1520
		gregarious oviposition	backcrossing	none	L	1	2791
		gregarious oviposition, total	backcrossing	none	L	1	2791
		hosts killed	backcrossing	none	P	1	1520
		hosts parasitized	backcrossing	none	L	1	1520
		longevity females	backcrossing	none	L	1	1520
		parasitism	isofemale	h^2, H^2	L	1	1466
		parasitism	isofemale	h^2, H^2	L	1	1466
		parasitism	backcrossing	none	L	1	2782
		parasitism	backcrossing	none	L	1	2791
		parasitism rate	backcrossing	none	L	1	1514
		sex ratio	backcrossing	none	L	1;4	1520
		sex ratio	backcrossing	none	L	1;4	2791
		total eggs laid	backcrossing	none	L	1	1514
		total eggs laid	backcrossing	none	L	1	2791

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*M. raptorellus* cont	total progeny	backcrossing	none	L	1	1520
	Nasonia vitripennis	copulation duration	half-sibling	h^2, CV_A	B	7	2307
		courtship duration	half-sibling	h^2, CV_A	B	7	2307
		critical photoperiod	isofemale	none	P	1;4	1899
		diapause response	isofemale	none	L	4	1899
		gynandromorphism	isofemale	none	L	7	2312
		lifespan	isofemale	none	L	1;4	1899
		receptivity at 1st courtship	half-sibling	h^2, CV_A	B	7	2307
		receptivity within 10 min	half-sibling	h^2, CV_A	B	7	2307
		second foundress sex ratio	isofemale	H^2	L	1	2280
		sex ratio	isofemale	none	L	1;4	1536
		sex ratio	isofemale	none	L	1;4	2763
		sex ratio	isofemale	none	L	1;4	2799
		single foundress sex raio	isofemale	H^2	L	1;4	2280
	Trichogramma brassicae	area searched	isofemale	none	B	1	2657
		body length	isofemale	none	M	1	2696
		body length	outbred	none	M	1	2696
		fecundity	isofemale	none	L	1;4	2696
		fecundity	outbred	none	L	1;4	2696
		head width	isofemale	none	M	1;4	2696
		head width	outbred	none	M	1;4	2696

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*T. brassicae* cont	hind tibia length	isofemale	none	M	1;4	2696
		hind tibia length	outbred	none	M	1;4	2696
		locomotor activity	isofemale	none	B	7	2755
		male mating	isofemale	none	L	7	2696
		male mating	outbred	none	L	7	2696
		ovigeny index	parent-offspring reg	none	L	1	2009
		reactive distance	full-sibling	none	B	7	1401
		reproductive concentration index	parent-offspring reg	none	L	1	2009
		walking speed	isofemale	none	B	7	2657
		locomotor activity	isofemale	none	B	7	2755
		fecundity	isofemale	H^2	L	1;4	1226
		forewing length	parent-offspring reg	none	M	7	1311
		forewing width	parent-offspring reg	none	M	7	1311
		head width	parent-offspring reg	none	M	1;4	1311
		hind tibia length	parent-offspring reg	none	M	1;4	1311
		hindwing length	parent-offspring reg	none	M	7	1311
	Trichogramma evanescens	developmental period	isofemale	none	L	4	2236
		emergence rate	isofemale	none	L	1;4	2236
		parasitism	isofemale	none	L	1	2236
		sex ratio	isofemale	none	L	1;4	2236
		thermal requirements	isofemale	none	P	2	2236

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Parasitoid	*T. evanescens* cont	host egg color	isofemale	none	P	7	564
		emergence	isogenic	H^2	L	1	525
		fecundity	isogenic	H^2	L	1;4	525
		longevity, females	isogenic	H^2	L	1	525
		number female offspring	isogenic	H^2	L	1	525
		number male offspring	isogenic	H^2	L	1	525
		sex ratio	isogenic	H^2	L	1;4	525
		fecundity	isofemale	none	L	1;4	299
		female offspring	isofemale	none	L	1	299
		mean generation time	isofemale	none	L	1;4	299
		reproductive rate	isofemale	none	L	1	299
		sex ratio	isofemale	none	L	1;4	299
	Trissolcus basalis	patch residency	isofemale	none	B	1	2465
		patch-leaving tendency	isofemale	none	B	1	2465
Predator	*Adalia bipunctata*	adaline concentration	parent-offspring reg	none	P	6	2765
		body weight	parent-offspring reg	none	M	1	2765
		growth rate	parent-offspring reg	none	M	4	2765
		reflex fluid	parent-offspring reg	none	P	2	2765
		spot 1 female	parent-offspring reg	none	M	2	2736
		spot 1 male	parent-offspring reg	none	M	2	2736
		spot 2 female	parent-offspring reg	none	M	2	2736

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*A. bipunctata* cont	spot 2 male	parent-offspring reg	none	M	2	2736
		spot 3 female	parent-offspring reg	none	M	2	2736
		spot 3 male	parent-offspring reg	none	M	2	2736
		spot 4 female	parent-offspring reg	none	M	2	2736
		spot 4 male	parent-offspring reg	none	M	2	2736
		wingless	outbred	none	M	1	571
	Amblyseius arkeri	oviposition rate	artificial selection	none	B	1	1379
		predation rate	artificial selection	none	L	1	1379
	Amblyseius cucumeris	eggs laid per female per day	artificial selection	none	L	1	1379
		eggs laid per female per day	artificial selection	none	L	1	1379
		eggs laid per female per day	artificial selection	none	L	1	1379
		eggs laid per female per day	artificial selection	none	L	1	1379
		eggs laid per female per day	artificial selection	none	L	1	1379
		larvae killed per female per day	artificial selection	none	B	1	1379
		larvae killed per female per day	artificial selection	none	B	1	1379
		larvae killed per female per day	artificial selection	none	B	1	1379
		larvae killed per female per day	artificial selection	none	B	1	1379

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*A. cucumeris* cont	larvae killed per female per day	artificial selection	none	B	1	1379
		oviposition rate	artificial selection	none	B	1	1379
		predation rate	artificial selection	none	L	1	1379
	Amblyseius largoensis	egg viability	backcrossing	none	L	1;4	768
		eggs laid per female per day	backcrossing	none	L	1	768
		number of eggs	backcrossing	none	L	1	768
		sex ratio	backcrossing	none	L	1;4	768
	Amblyseius womersleyi	dispersal ratio over time	common	none	B	1	2595
		patch choice	common	none	B	1	2595
	Chrysoperla carnea	benzoate resistance	artificial selection	none	P	2	1937
		development time egg to adult	artificial selection	none	L	4	1868
		development time egg to adult	artificial selection	none	L	4	1937
		diapause	full-sibling	none	L	4	2781
		diapause	full-sibling	none	L	4	2846
		eggs laid per female	artificial selection	none	L	1	1868
		eggs laid per female	artificial selection	none	L	1	1937
		emergence rate of healthy adults	artificial selection	none	L	1;4	1868
		emergence rate of healthy adults	artificial selection	none	L	1;4	1937

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*C. carnea* cont	fecundity	full-sibling	none	L	1;4	2781
		hatchability	artificial selection	none	L	1;4	1868
		hatchability	artificial selection	none	L	1;4	1937
		larval development time	artificial selection	none	L	4	1868
		larval development time	artificial selection	none	L	4	1937
		larval survival rate	artificial selection	none	L	5	1868
		larval survival rate	artificial selection	none	L	5	1937
		LD50	artificial selection	none	P	5	1684
		LD50	artificial selection	none	P	5	1684
		LD50	artificial selection	none	P	5	1684
		LD50	artificial selection	none	P	5	1684
		net replacement rate	artificial selection	none	L	1;4	1868
		net replacement rate	artificial selection	none	L	1;4	1937
		nondiapause preoviposition period	full-sibling	none	L	1;4	2781
		number of next gen larvae	artificial selection	none	L	1;4	1868
		number of next gen larvae	artificial selection	none	L	1;4	1937
		pupal duration	artificial selection	none	L	4	1868
		pupal duration	artificial selection	none	L	4	1937
		pupal weight	artificial selection	none	M	7	1868
		pupal weight	artificial selection	none	M	7	1937
		pupation rate	artificial selection	none	L	7	1868
		pupation rate	artificial selection	none	L	7	1937

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*C. carnea* cont	relative fitness	artificial selection	none	L	1	1868
		relative fitness	artificial selection	none	L	1	1937
		resistance alphametrin	artificial selection	h^2	P	5	962
		resistance chlorpyrifos2	artificial selection	h^2	P	5	962
		resistance deltamethrin2	artificial selection	h^2	P	5	962
		resistance lambdacyhalothrin2	artificial selection	h^2	P	5	962
		resistance profenofos2	artificial selection	h^2	P	5	962
		spinosad resistance	artificial selection	none	P	5	1868
	Cryptolaemus montrouzieri	developmental period	artificial selection	none	L	4	3001
		fecundity	artificial selection	none	L	1;4	3001
		female length	artificial selection	none	M	1	3001
		larval developmental time	artificial selection	none	L	4	3001
		prey eaten per day	artificial selection	none	B	1	1851
		prey potential of 3rd instar	artificial selection	none	B	1	3001
		prey potential of female	artificial selection	none	B	1	3001
		reproduction	artificial selection	none	L	1	3001
		survival rate	artificial selection	none	L	1;4	3001
	Harmonia axyridis	abdomen length	full-sibling	H^2, CV_A	M	1	471
		adult development time	artificial selection	h^2, CV_A	L	1	2601
		adult development time	artificial selection	h^2, CV_A	L	1	2601
		adult mass	full-sibling	H^2, CV_A	M	1	471

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*H. axyridis* cont	adult size	artificial selection	h^2, CV_A	M	1	2601
		adult size	artificial selection	h^2, CV_A	M	1	2601
		body length	reciprocal cross	CV	M	1	947
		body length	reciprocal cross	CV	M	1	947
		body length	reciprocal cross	CV	M	1	947
		body length	reciprocal cross	CV	M	1	947
		cannibalism rate	artificial selection	h^2, CV_A	B	4	2601
		cannibalism rate	artificial selection	h^2, CV_A	B	4	2601
		development	full-sibling	H^2, CV_A	L	4	543
		development	full-sibling	H^2, CV_A	L	4	543
		development time	full-sibling	H^2, CV_A	L	4	471
		development time egg to adult	reciprocal cross	CV_A	L	4	947
		development time egg to adult	reciprocal cross	CV_A	L	4	947
		development time egg to adult	reciprocal cross	CV_A	L	4	947
		development time egg to adult	reciprocal cross	CV_A	L	4	947
		female reproductive investment	reciprocal cross	CV_A	L	1	947
		female reproductive investment	reciprocal cross	CV_A	L	1	947
		female reproductive investment	reciprocal cross	CV_A	L	1	947

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*H. axyridis* cont	female reproductive investment	reciprocal cross	CV_A	L	1	947
		hatching time	full-sibling	H^2, CV_A	L	4	471
		larval development time	artificial selection	h^2, CV_A	L	4	2601
		larval development time	artificial selection	h^2, CV_A	L	4	2601
		larval size	artificial selection	h^2, CV_A	M	1	2601
		larval size	artificial selection	h^2, CV_A	M	1	2601
		larval survival to adult-hood	reciprocal cross	CV_A	L	7	947
		larval survival to adult-hood	reciprocal cross	CV_A	L	7	947
		larval survival to adult-hood	reciprocal cross	CV_A	L	7	947
		larval survival to adult-hood	reciprocal cross	CV_A	L	7	947
		larval weight	full-sibling	H^2	M	1	1092
		larval weight	full-sibling	H^2	M	1	1092
		orange wing reflectance	full-sibling	H^2, CV_A	M	2	543
		orange wing reflectance	full-sibling	H^2, CV_A	M	2	543
		pronotum width	full-sibling	H^2, CV_A	M	7	471
		pronotum width	full-sibling	H^2, CV_A	M	7	543
		pronotum width	full-sibling	H^2, CV_A	M	7	543
		red wing reflectance	full-sibling	H^2, CV_A	M	2	543
		red wing reflectance	full-sibling	H^2, CV_A	M	2	543
		starvation resistance	reciprocal cross	CV_A	L	1;4	947

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*H. axyridis* cont	starvation resistance	reciprocal cross	CV_A	L	1;4	947
		starvation resistance	reciprocal cross	CV_A	L	1;4	947
		starvation resistance	reciprocal cross	CV_A	L	1;4	947
		survival rate in quiescent conditions	reciprocal cross	CV_A	L	1;4	947
		survival rate in quiescent conditions	reciprocal cross	CV_A	L	1;4	947
		survival rate in quiescent conditions	reciprocal cross	CV_A	L	1;4	947
		survival rate in quiescent conditions	reciprocal cross	CV_A	L	1;4	947
		thorax length	full-sibling	H^2, CV_A	M	1	471
	Hippodamia convergens	development time egg to adult	backcrossing	none	L	4	515
		elytra size	backcrossing	none	M	2	515
		fecundity	backcrossing	none	L	1;4	515
		fertility	backcrossing	none	L	1;4	515
		pronotum size	backcrossing	none	M	7	515
		sex ratio	backcrossing	none	L	1	515
		survival	backcrossing	none	L	1;4	515
		weight, female	backcrossing	none	M	1	515
		weight, male	backcrossing	none	M	1	515
	Hypoaspis aculeifer	mate choice	backcrossing	none	B	4	2633
	Orius majusculus	basal activity	reciprocal cross	none	B	2	3007

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*O. majusculus* cont	body size	reciprocal cross	none	M	1	3007
		critical thermal maximum	reciprocal cross	none	P	2	3007
		predation rate	reciprocal cross	none	B	1	3007
		starvation tolerance	reciprocal cross	none	L	1;4	3007
		temperature of maximal activity	reciprocal cross	none	P	2	3007
		consumption rate	artificial selection	h^2	B	1	2141
		consumption rate	artificial selection	h^2	B	1	2141
		consumption rate	artificial selection	h^2	B	1	2141
		consumption rate	isofemale	h^2	B	1	2470
		conversion efficiency	artificial selection	h^2	L	1	2141
		conversion efficiency	artificial selection	h^2	L	1	2141
		conversion efficiency	artificial selection	h^2	L	1	2141
		dispersal response to prey density	artificial selection	h^2	B	1	2141
		dispersal response to prey density	artificial selection	h^2	B	1	2141
		dispersal response to prey density	artificial selection	h^2	B	1	2141
		eggs laid per female per day	artificial selection	none	L	1	2689
		olfactory attraction to pest induced plant volatiles	artificial selection	h^2	B	6	2141

Type of BCA	Species	Trait of interest as reported	Estimation method	Genetic variation measure	Fitness component[a]	Biological control relevance[b]	Paper ID[c]
Predator	*O. majusculus* cont	olfactory attraction to pest induced plant volatiles	artificial selection	h^2	B	6	2141
		oviposition rate	isofemale	h^2	L	1	2470
		patch residency	isofemale	h^2	B	1	2470
		prey consumption	artificial selection	none	B	1	2689
		prey location	isofemale	h^2	B	1	2470
		residence time	artificial selection	none	B	1	2689

[a] Classification based on Mousseau and Roff (Mousseau and Roff, 1987): behaviour (B), life history (L), morphology (M), and physiology (P)

[b] Classification based on Leung *et al.* (Leung et al., 2019a): pest suppression ability (1), adaptation to abiotic factors (2), reducing ecological risk (3), improving mass rearing conditions (4). Three additional categories were created to address the outcomes of this extraction: insecticide resistance (5), adaptation to biotic factors (6), and a null category named 'unrelated' (7)

[c] Paper identification numbers are unique to each paper, and are cross-referenced in Table 1.

heritability studies, but to a much lesser extent than those focusing on life history traits. This is most probably due to the importance of life history traits for biological control, which resulted in a large body of descriptive literature for BCAs (Mayhew, 2016). Furthermore, life history traits are generally considered to be an established measure of an organism's reproductive strategy and, ultimately, fitness. As such, they are considered important in the establishment and success of BCAs (Plouvier and Wajnberg, 2018). They are also wide ranging and can vary from obvious characteristics such as longevity and number of offspring, to more specific attributes such as conversion efficiency, certain of which are proven to be very useful in particular applied settings. Of the papers in the final selection, 30.6% were found to include usable heritability estimates or the means to calculate them.

The functionality of traits in a biological control setting was further explored using the second classification. This classification gives a more targeted approach to labelling traits as to their potential utility in a biological control setting. The majority of traits analyzed within this set were attributed to 'pest suppression ability'. This attribute would include traits classed as life history traits in the Mousseau and Roff classification (i.e. number of offspring or sex ratio) but could also include certain behavioural traits (i.e. patch residence time). In terms of number of observations, the next group of traits are those 'useful in improving mass rearing conditions', which would include life history traits such as sex ratio. These are potentially two of the more important classes within the classification for biological control application, indicating that the traits outlined in Leung *et al*. are in line with the current trends in biological control research (Leung et al., 2019a). It is also important to note that certain traits could fit into several of the biological control trait classifications, whereas the Mousseau and Roff classifications are clear-cut. In this way, the classification for biological control application could be deemed more relevant or concise when narrowing down traits for their importance in biological control, and by doing so in such a study as ours, pinpoints those that are also heritable. Our results further show that traits related to 'adaptation to biotic factors,' and 'reducing ecological risk' are relatively understudied compared to the others, which could be seen as an incentive for more studies.

The lack of data on genetic variation in our results does not necessarily represent a lack of heritability studies - indeed in our original search we found 69 papers concerning heritability of life history traits in BCAs with appropriate estimation methods. Of these, however, only 19 papers contained heritability values, either in the form of broad-sense heritability, narrow-sense heritability, evolvability, or the genetic and environmental variance estimates required for calculation. As for estimation methods, nearly 9 out of 10 papers were removed from the analysis during our screening due to opaque reporting of methods and results, yet the papers still referenced heritability or genetic variation. This opacity has been documented previously, such as in a recent review on the use of experimental evolution to improve biological control, where the majority of studies found did not carry out replicates or maintain control populations (Lirakis and Magalhães, 2019). Discrepancies such as these are indications of the flexibility of terms and experimental design between research disciplines; the full reporting of values is simply a recommendation for improving research on genetic variation that has been made before (see Houle, 1992). Reporting sample size is necessary for variance calculations, whereas calculating heritability values is encouraged in order to add quantitative data to a discussion on heritability. With this in mind, it became clear during the extraction of data that further analysis, such as a meta-analysis of the heritability results, would be intractable. This is due to the wide spread of species and traits of interests juxtaposed to a lack of heritability values and other values necessary for such an analysis. This lack of reporting could be due to oversight or publication restrictions (such as limitation on manuscript length or perceived triviality of some data), but another possibility is that the perceived importance of reporting these values is not high. We hope that recent calls on the importance of genetic variation for improving BCAs, such as Leung *et al*. (2019), and the need for these values will shift towards studies publishing more basic information and additional data in the future.

Finally, there are fairly similar number of papers on parasitoids and predators, both before and after extraction. While initially there looked to be a slight bias towards parasitoids, there is a larger bias in the registered and published

BCA types in use, with far more parasitoids than predators (van Lenteren, 2012), making this more-or-less equal outcome more surprising.

This systematic review was carried out across three different reference databases with an intentional search term that was tested against positive and negative controls. Once the search results returned, assessment of the study for its inclusion was based on both objective and subjective measures. In this respect, certain papers that would fit the desired outcome of the search may have been absent due to database or search term or dismissed prematurely due to human error. Taxonomic changes could also play a part in missing papers that would otherwise be suitable for this analysis. Finally, refining the species studied to the combined list of the 2016 and 2019 EPPO lists of commercial BCAs (PM 6/003) as well as using a commercial list from 2012 that was updated in 2018 (van Lenteren, 2012; van Lenteren et al., 2018) may remove any species that are currently being investigated as a potential BCA.

Conclusion

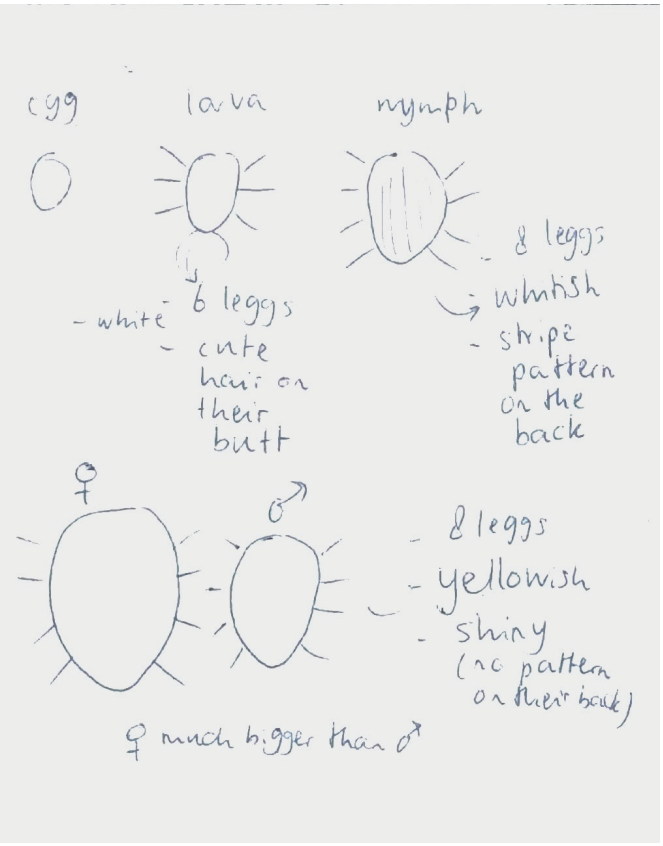
The results of this systematic review indicate a variety of research and interest in studying the heritability of traits in BCAs and can act as a starting point for improving natural enemies. The presence of genetic variation for traits related to biological control supports the calls for improvement of biological control agents via natural genetic variation made previously (Kruitwagen et al., 2018; Leung et al., 2019a; Lirakis and Magalhães, 2019; Lommen et al., 2017). Indeed, a large variety of groundwork has been established in these fields related to species, estimation methods, and traits of interest. However, studies that report on genetic variation cannot necessarily add to the advancement of biological control research if heritability estimates and associated values are not reported. With this in mind, this aspect of biological control research is ripe with opportunities, and several of the studies we have profiled here are an excellent starting point for the genetic improvement of the efficacy of biological control agents.

ACKNOWLEDGEMENTS

This work was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 641456. We especially want to thank Simone Ariëns for her contributions in the initial stages of this work, and Bas Zwaan for his valuable comments that helped to improve our manuscript. Thank you to the WUR library for their assistance with retrieving articles. Special thanks to our team of readers: Shuwen Xia, Eveline Verhulst, Bas Zwaan, Leo Beukeboom, Andra Thiel, Anna Alekseeva, Alejandra Centurión, Milena Chinchilla Ramírez, Giota Koskinioti, Kelley Leung, Sophie Le Hesran, Manolis Lirakis, Angeliki Paspati, Laura Perez Martin, Judith Stahl, and Sander Visser.

SUPPLEMENTARY MATERIALS

Supplementary material indicated in this chapter (Supplementary Tables 1, 2, and 3) are available on the DANS EASY Repository,
DOI: 10.17026/dans-zvv-d2dr



A guide to identifying juvenile and adult *Amblyseius swirskii*, as rendered by Kirsten Oude Lenferink.

Chapter 3

Effect of mass rearing on the genetic diversity of the predatory mite *Amblyseius swirskii*

Angeliki Paspati[a], Kim B. Ferguson[b], Eveline C. Verhulst[c], Alberto Urbaneja[a], Joel González-Cabrera[d] & Bart A. Pannebakker[b]

[a]Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Moncada, Valencia, Spain
[b]Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands
[c]Wageningen University, Laboratory of Entomology, Wageningen, The Netherlands
[d]Unidad Mixta Gestión Biotecnológica de Plagas UV-IVIA, Department of Genetics, Universitat de València, Estructura de Recerca Interdisciplinar en Biotecnología i Biomedicina (ERI-BIOTECMED), Burjassot, Spain

Published in Entomologia Experimentalis et Applicata as:
Paspati, A., Ferguson, K.B., Verhulst, E.C., Urbaneja, A., González-Cabrera, J., Pannebakker, B.A., 2019. Effect of mass rearing on the genetic diversity of the predatory mite *Amblyseius swirskii* Athias-Henriot (Acari: Phytoseiidae). Entomol. Exp. Appl. 167, 670–681. https://doi.org/10.1111/eea.12811

ABSTRACT

Amblyseius swirskii Athias-Henriot (Acari: Phytoseiidae) is a predatory mite used to control whiteflies and thrips in protected crops. This biological control agent, originating from the Eastern Mediterranean region, has been mass-reared for commercial use since 2005 and is widely used in augmentative biological control programs. As a polyphagous predator, it has to cope with different biotic and abiotic factors. However, possible adaptation to mass rearing for production might be hindering its resilience and capacity for optimum performance in the field. In this study, we investigated the effect of long-term mass rearing on the genetic diversity of *A. swirskii*. We identified six microsatellite loci from whole-genome nanopore sequencing of *A. swirskii* and used these in a comparative analysis of the genetic diversity and differentiation in eight wild populations collected from Israel in 2017 and a commercially available population. Our results indicate that the commercial population is 2.5 times less heterozygous than the wild *A. swirskii*. Furthermore, the commercial population has the highest genetic differentiation from all the natural populations, as indicated by higher pairwise F_{ST} values. Overall, we show that commercially reared *A. swirskii* have reduced genetic variation compared to their wild counterparts, which may reduce their performance when released to control pests in an integrated pest management (IPM) context.

INTRODUCTION

Food security and safety should be the cornerstone of modern agriculture and the overreliance on chemicals to control pests is jeopardising this goal. The accumulation of residues in the environment and the food chain, their negative impact on human health, biodiversity, soil and water quality, and the reduced efficacy due to resistance evolution have put the spotlight on biological control as the environmentally sound alternative for pest control. In augmentative biological control, mass-reared natural enemies are released in large numbers to protect the crops from the negative impact of certain pests (Eilenberg et al., 2001). Today, this approach is key for the success of integrated pest management (IPM) programs for many fruit and vegetable crops, such as cereals, maize, cotton, sugarcane, soybean, grapes, and

many greenhouse crops on more than 30 million ha worldwide (van Lenteren et al., 2018).

The most widely used predatory mite in augmentative biological control is the polyphagous predator *Amblyseius swirskii* Athias-Henriot (Acari: Phytoseiidae), followed by two other phytoseiids, *Phytoseiulus persimilis* Athias-Henriot, and *Neoseiulus californicus* (McGregor) (Knapp et al., 2018). *Amblyseius swirskii* was originally described in 1962 from almond trees (*Prunus dulcis* (Mill.) DA Webb) in Israel (Athias Henriot, 1962), where it naturally occurs on various other annual and perennial crops, such as citrus, grapes, vegetables, and cotton (Swirski & Amitai, 1997). When whitefly resistance to pesticides caused outbreaks in greenhouses at the beginning of this century (e.g., Stansly et al., 2004), attention was drawn to *A. swirskii* because of its fast reproduction and high performance against the key pest *Bemisia tabaci* (Gennadius) (Nomikou et al., 2001; Calvo et al., 2015). Experiments showed that *A. swirskii* was able to suppress *B. tabaci* populations and also provided good control of thrips *Frankliniella occidentalis* (Pergande) on cucumber plants (Nomikou et al., 2002; van Houten et al., 2005).

Today, this species is used to control whiteflies and thrips in greenhouse vegetables, fruits, and ornamentals in various parts of the world (Calvo et al., 2015). The importance of *A. swirskii* in current agriculture is showcased in protected sweet pepper crops in south-eastern Spain, where the successful integration of this predatory mite in the IPM strategy against whiteflies and thrips has led to a sharp decrease in the use of chemical pesticides (Calvo et al., 2011; van Lenteren et al., 2018). In addition to its high efficacy in managing whitefly and thrips infestations, the success of *A. swirskii* has been further encouraged by its early establishment on crops before the target pests arrive, using pollen or factitious prey as a food source, and by its mass rearing on a plantless system using the stored-product mite *Carpoglyphus lactis* (L.) as factitious prey (Bolckmans & van Houten, 2006; Calvo et al., 2015).

The effects of long-term mass rearing on the quality and performance of commercial biological control agents such as *A. swirskii* in natural or semi-natural environments are largely unknown (e.g., Tayeh et al., 2012; Guzmán-

Larralde et al., 2014; Rasmussen et al., 2018). Selection, inbreeding, and random genetic drift may lead to loss of genetic variability, loss of fitness, and reduced field performance of the mass-reared biological control agents (Mackauer, 1976). The constant and regular commercial mass rearing conditions, which are optimized to obtain the maximum number of individuals in the shortest possible time, differ considerably from the greenhouse biotic and abiotic conditions. Furthermore, under these artificial conditions, searching time for preying, mating, or dispersal is intentionally restricted to accelerate and synchronize the breeding batches as much as possible.

Therefore, long-term laboratory rearing eventually leads to selection for the facility conditions (domestication), which in turn affects the reproductive or behavioural traits of the arthropods (Hoffmann & Ross, 2018). For example, the reproductive fitness of *Drosophila melanogaster* Meigen populations captive for 50 generations, was reduced in small populations because of inbreeding depression and in large populations due to genetic adaptation (Woodworth et al., 2002). Although domestication is likely to play a key role in the rearing and success of biological control agents, only few studies on natural enemies have addressed this process. For instance, fecundity and emergence of the egg parasitoid *Trichogramma galloi* (Zucchi) were reduced on the target pest, when the parasitoids were maintained on factitious compared to natural hosts (Bertin et al., 2017).

Long-term laboratory rearing of the codling moth parasitoid, *Mastrus ridens* (Horstmann), resulted in lower genetic diversity and higher occurrence of diploid males, most likely as an effect of the reduced allelic diversity on a complementary sex determination locus (Retamal et al., 2016). In inbred lines of this parasitoid, the sex ratio of males to females and the proportion of diploid males was higher, and fewer daughters were produced compared to outbred lines (Zaviezo et al., 2018). Also, inbred populations of the mass-reared parasitoid *Trichogramma pretiosum* (Riley) produced fewer offspring than genetically variable populations (Guzmán-Larralde et al., 2014). A severe lack of genetic diversity and phenotypic differentiation in wing size and abdomen colour were found for one inbred laboratory colony of the crop pollinator *Eristalis tenax* (L.) compared with natural populations (Francuski et

al., 2014). In another case, reproduction parameters and virulence of the entomopathogenic nematode *Steinernema glaseri* (Steiner) were reduced after laboratory adaptation took place (Stuart & Gaugler, 1996).

The performance of biological control agents under variable field conditions depends on its genetic variability, which is defined by the initial size, origin, and degree of inbreeding of the founder colony of a biological control agent (Mackauer, 1976). If the commercial population is established from a limited number of individuals, then the genetic variation is expected to be low and further reduced due to inbreeding. The genetic variation of commercial populations may be reduced by random genetic drift during rearing as well, such as splitting the colony for commercial distribution, or the occurrence of bottlenecks (Nunney, 2003). One example is the aphid parasitoid *Diaeretiella rapae* (M'Intosh), which experienced a significant founder effect and a strong reduction in the genetic diversity upon its introduction in Australia, compared to the populations from the Old World (Baker et al., 2003). Finally, six biofactory populations of the cereal stemborer parasitoid *Cotesia flavipes* (Cameron) in Brazil showed genetic differentiation among each other, resulting in new guidelines for the mixture of populations in the mass-rearing strategy to enhance the genetic variability (Freitas et al., 2018).

The commercial *A. swirskii* was originally collected in Israel and is being reared on a large scale without its natural prey since 2005 (YM van Houten, Koppert Biological Systems, The Netherlands, pers. comm.). The aim of this study was to investigate the population genetic structure of field *A. swirskii* from Israel, as well as the genetic variation in a commercial, long-term reared population of this biological control agent. For this, we developed microsatellite markers for *A. swirskii*, using a cost-effective, in-house sequencing process, the MinION Nanopore sequencer for whole genome sequencing, which was applied and evaluated for the first time in sequencing predatory mites. Also, in our study we explore the application of pooled microsatellite analysis to predatory mites, thereby offering a cost-effective solution to low individual DNA yields in these minute organisms (Skalski et al., 2006).

Materials and methods

Populations and DNA preparation

For the whole-genome sequencing, an inbred population from the commercially available Swirski-Mite® (Koppert Biological Systems, Berkel en Rodenrijs, The Netherlands) was established. Five females and five males from the general population were allowed to reproduce for approximately 10 days in a rearing system with cattail pollen (*Typha latifolia* L.) as food source, where eggs were collected and isolated on a new rearing setup. Subsequently five F1 females and five F1 males were used to start a new generation. This was repeated for a total of 10 generations. Over 200 individuals from the 10th generation of the inbred population were collected in 96% ethanol. DNA extraction was performed using the Qiagen MagAttract HMW DNA Kit (Qiagen, Hilden, Germany) according to the manual, whereas the final DNA product was eluted with nuclease-free water.



Figure 1. Map of *Amblyseius swirskii* sampling locations across Israel.

For microsatellite analysis, in total eight populations were sampled, from five locations across Israel, from various host crops (citrus, kiwi, sweet pepper, Jerusalem artichoke, cotton) (Figure 1, Table 1). Mites were collected from the field and transferred to the laboratory where isolated colonies were established and named based on the plant and location found, with *Carpobrotus edulis* (L.) NE Brown pollen as food source. After approximately 18 generations (6 months), 100 female mites were collected from each colony and placed in 96% ethanol at –20 °C for DNA extraction. A single batch of 100 *A. swirskii* individuals from a commercially available Swirski-Mite bottle (Koppert) was collected directly and placed in 96% ethanol at –20 °C.

Table 1. Sampling site locations (cf. Figure 1), source (host plant in the field or commercial product), and number of mites initially collected for all *Amblyseius swirskii* populations

Location	Population	Source	Sample size
Hula Valley (HV)	HV-1	Citrus	100
Beit HaEmek (BHE)	BHE-2	Kiwi	30
Ezor Yokne'am (EY)	EY-3	Citrus	100
Hadera (Had)	Had-4	Sweet pepper	20
	Had-5	Jerusalem artichoke	10
	Had-6	Citrus	15
Hof Ashkelon (HAs)	HAs-7	Cotton	40
	HAs-8	Citrus	8
-	KBS-9	Koppert Biological Systems	100

The commercial population was not reared in the laboratory before sampling, because this could cause a reduction of its genetic variation. Due to the small size of the mites (0.5 mm) and to reduce genotyping costs (Skalski et al., 2006), we performed pooled DNA extractions for individuals from the same population. Tissue homogenization was achieved by placing dry snap-frozen samples with beads in a shaker for 20 s. Finally, DNA was extracted using a high salt extraction protocol (Maniatis et al., 1982).

Genomic sequencing, markers, and experimental procedures

The inbred *A. swirskii* population was sequenced using Oxford Nanopore Technologies (ONT) MinION sequencer (Ligation Sequencing Kit v.108, SpotON Flow Cell FLO-MIN107 R9.5) (ONT, Oxford, UK). Approximately 30 000 reads were generated through direct base calling with the MinKNOW platform (v.1.10.16) or offline base calling using Albacore (v.2.1.3) (ONT). All raw and uncorrected read files (both FAST5 and FASTQ formats) are available for download (Ferguson, 2018). The base-called reads, in FASTQ format, were initially trimmed in CLC Genomics Workbench v.11.0 (Qiagen, Hilden, Germany) for a minimum read length of 100 bp with low-quality sequences removed (limit = 0.05), resulting in 11 511 trimmed reads (Qiagen). These trimmed FASTQ reads were then corrected using Canu v.1.4 pipeline (`parameters: genomeSize = 100 m, correctedErrorRate = 0.120,`

`stopOnReadQuality = false`) (Koren et al., 2017). This resulted in a total of 574 corrected reads, all longer than 1 000 bp. These corrected, raw reads are deposited in the NCBI Sequence Read Archive (PRJNA433466).

The corrected sequences were used for mining di-, tri-, tetra-, penta-, and hexanucleotide repeats using MsatCommander v.0.8.2 with the default settings, except for the number of repeats that was set to a minimum of three for tri-, tetra-, penta-, and hexanucleotides (Faircloth, 2008). The MsatCommander program uses Primer3 v.1.1.1 to design locus-specific, flanking primers. The settings used for Primer3 were: no perfect repeats, product size 75-500, primer size: min 16, opt 20, max 24 bases, and primer annealing temperature: 56-64 °C. Poor primers were removed based on the following criteria: duplicate, positive for hairpins, high complementarity, and selfing (Rozen & Skaletsky, 2000). The markers with their corresponding primer pairs were sorted by the count of repeats in descending order, as markers with fewer repeats can be less polymorphic and the first 24 primer pairs of microsatellite loci were selected to be tested for their efficiency using PCR.

A total of six microsatellite loci primer pairs amplified efficiently only the desired microsatellite marker and were selected for analysis, whereas the remaining were rejected for either low efficiency or low specificity. Primer specificity was considered low when the PCR resulted in products different from the expected size. The sequence data of the loci were deposited in the NCBI's GenBank with accession numbers MK267176-MK267181 (Supplementary Table 1). The six primer pairs used for the microsatellite amplification and analysis were fluorescently labelled (Supplementary Table 1). The microsatellite markers were amplified for the nine populations of *A. swirskii* according to the following protocol. The PCR reactions were performed in a total volume of 50 µl containing 60 ng of genomic DNA, 1x GoTaq polymerase buffer, 200 µM dNTP, 1.25 U GoTaq polymerase (Promega, Madison, WI, USA), and 0.4 µM of each primer. The PCR protocol was as follows: 5 min denaturation at 95 °C, 35 cycles of 95 °C for 30 s, 50-60 °C for 30 s, 72 °C for 60 s, followed by 10 min at 72 °C. Primer sequences and annealing temperatures are indicated in Supplementary Table 1. For analysis, microsatellite amplicons were diluted 250x and pooled into sets of 2-3 in such a way that overlapping of alleles was

avoided even if markers differed for the fluorescent dye. The samples were denatured at 95 °C and loaded with an internal size standard (GeneScan 500 LIZ) onto an ABI3730 capillary automated sequencer (Thermo Fisher Scientific, Waltham, MA, USA).

MICROSATELLITE ANALYSIS

Genotyping of the pooled DNA samples was performed using the GENEMAPPER software (Thermo Fisher Scientific). Instead of allele frequencies, peak frequencies were calculated based on peak areas, which is more reliable than to correct for stutter bands in pooled microsatellite samples (Crooijmans et al., 1996). Microsatellite peak scoring was used to estimate allele frequencies as explained in Hillel et al. (2003). Peak frequencies lower than 0.05 were discarded and peak frequencies higher than 0.05 were subsequently recalculated to add up to 1 to correct for stutter bands that resulted from incomplete PCR cycles (Megens et al., 2008).

DATA ANALYSIS

Neighbour Joining clustering was performed for the microsatellite dataset, using Nei's D_A genetic distance in POPTREEW with 1,000 bootstraps (Takezaki et al., 2014). Gene diversity was measured as heterozygosity for all loci and for all populations by using peak frequencies instead of allele frequencies according to Hillel et al. (2003). An average heterozygosity was calculated for each population across markers (H) and for each marker across populations (H_m). Similarly, the average allelic richness for each population across markers (A) and for each marker across populations (A_m), and the number of private alleles (N_p) for each population were calculated. Three genetic distance measures for various evolutionary models were calculated based on peak frequencies: Nei's genetic distance (D_A), Cavalli-Sforza chord measure (D_{CS}), and Reynolds' genetic distance (D_R) (Cavalli-Sforza & Edwards, 1967; Nei et al., 1983; Reynolds et al., 1983).

Pairwise distances between each pair of the nine populations (81 estimates) were estimated for each measure. D_A was chosen to be used in the downstream analysis as more appropriate, because this measure assumes that genetic differences are caused by mutations and genetic drift, contrary

to the other measures, which do not take into account mutation (Cavalli-Sforza & Edwards, 1967; Nei et al., 1983; Reynolds et al., 1983). The distance matrices were used for multidimensional scaling (Gower, 1966) in R v.3.4.3 (R Core Team, 2018). Isolation-by-distance analysis with a Mantel (1967) correlation test based on 9,999 replicates was also performed using the ADE4 package (Dray et al., 2007) in R v.3.4.3. Last, F_{st} measures of genetic differentiation were calculated manually according to Wright (1984).

RESULTS

Amblyseius swirskii genomic sequencing with the MINION generated 256 Mb total sequence data after base calling, or approximately 30,000 reads. After sequence trimming and correction, 574 corrected reads of at least 1,000 bp long where retained (5.2 Mb). Microsatellite mining of the genome sequences identified 2,423 microsatellite loci on 532 reads. The density of microsatellites in the genome is 466 microsatellites per Mb and 92.6% of our reads contain microsatellite loci. The most common type of microsatellites are dinucleotide repeats (72.1%), followed by trinucleotide repeats (23.7%), whereas the least common microsatellites found were the hexanucleotide repeats (0.12%) (Supplementary Table 2). Primers designed using PRIMER3 were filtered for quality, resulting in 622 valid primer pairs. Of those, 24 primer pairs of the markers with the largest repeat count were selected and then tested for their specificity and efficiency. Finally, a set of six microsatellite markers was found to be specific and efficient for using in the microsatellite analysis of population pools. The nucleotide motifs of these six microsatellite markers are: di- (Asw1), tri- (Asw3, Asw6), and tetranucleotide repeats (Asw2, Asw4, Asw5) (Supplementary Table 1).

In total, 32 different alleles were scored across the six microsatellite loci for the nine populations, with four of those alleles being private, exclusively found in equal number of wild populations (Table 2). All markers were polymorphic in at least four of the nine populations and all of the populations had at least three markers polymorphic (Table 2). The average gene diversity H_m for the six markers analysed was 0.42 and the mean allelic richness per locus A_m was 5.3 across populations and 2.9 within populations (Table 3). The most

Table 2. Frequency of polymorphic markers (P), average heterozygosity (*H*), average allelic richness (*A*), the mean genetic distance estimates (MGD) of Nei (D_A), Cavalli-Sforza (D_{CS}), and Reynolds (D_R), and the number of private alleles (N_p) for nine populations of *Amblyseius swirskii* (eight wild and one commercial, KBS-9)

Population	P	*H*	*A*	MGD_A	MGD_{CS}	MGD_R	N_p
BHE-2	0.83	0.34	2.5	0.499	0.333	0.386	1
EY-3	1	0.50	3.7	0.254	0.176	0.210	0
HV-1	1	0.48	2.7	0.323	0.209	0.252	0
HAs-7	0.83	0.39	3	0.430	0.293	0.336	0
HAs-8	1	0.52	3.7	0.423	0.274	0.287	1
Had-4	1	0.53	2.5	0.418	0.250	0.279	0
Had-5	0.83	0.42	3.3	0.362	0.244	0.295	1
Had-6	0.83	0.41	2.7	0.401	0.315	0.318	1
KBS-9	0.50	0.21	1.8	0.430	0.300	0.418	0

Table 3. Heterozygosity and allelic richness for all the microsatellite markers per locus (H_m/A_m), and within populations (*H*/*A*) across nine populations of *Amblyseius swirskii* (eight wild and one commercial, KBS-9)

Population	Asw1	Asw2	Asw3	Asw4	Asw5	Asw6	Average *H*/*A*
BHE-2	0.69/5	0.37/3	0.50/2	0.31/2	0.19/2	0.00/1	0.34/2.5
EY-3	0.63/8	0.50/2	0.50/2	0.63/5	0.13/2	0.63/3	0.50/3.7
HV-1	0.48/4	0.50/2	0.40/2	0.51/3	0.40/2	0.57/3	0.48/2.7
HAs-7	0.74/7	0.51/3	0.46/2	0.33/3	0.00/1	0.27/2	0.39/3.0
HAs-8	0.79/8	0.47/2	0.25/2	0.59/5	0.50/2	0.52/3	0.52/3.7
Had-4	0.55/3	0.38/2	0.50/2	0.65/3	0.48/2	0.62/3	0.53/2.5
Had-5	0.73/6	0.29/2	0.25/2	0.65/6	0.00/1	0.62/3	0.42/3.3
Had-6	0.71/7	0.49/2	0.50/2	0.32/2	0.00/1	0.43/2	0.41/2.7
KBS-9	0.65/4	0.00/1	0.00/1	0.28/2	0.00/1	0.35/2	0.21/1.8
Average H_m/A_m	**0.66/5.8**	**0.39/2.1**	**0.37/1.9**	**0.47/3.4**	**0.19/1.6**	**0.45/2.4**	**0.42/2.9**

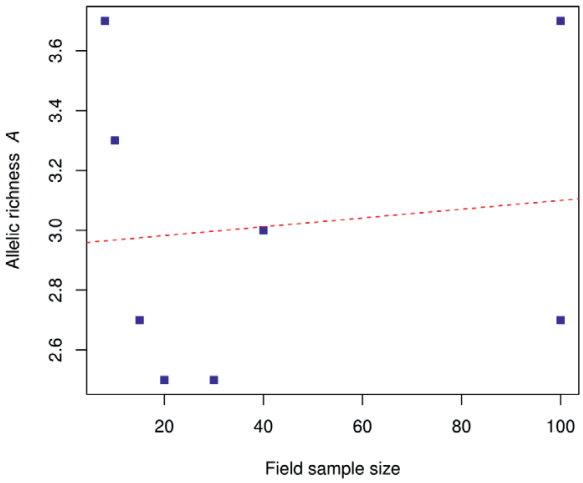


Figure 2. Allelic richness (*A*) of each population vs. field sample size for eight wild *Amblyseius swirskii* populations. The dashed line represents the linear regression fit (R^2 = 0.01266, P = 0.79).

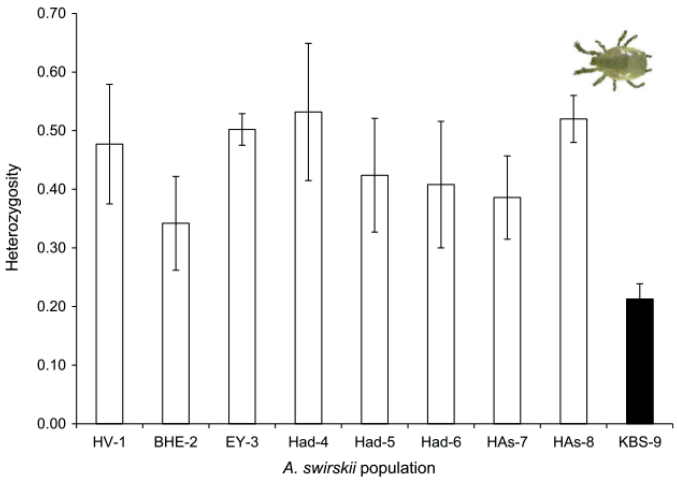


Figure 3. Mean (± SE) heterozygosity (*H*) across six microsatellite markers for eight wild and one commercial (KBS-9) population of *Amblyseius swirskii*.

polymorphic marker was Asw1 with 12 alleles across populations and a mean allelic richness of A_m = 5.8 per population (Table 3).

The gene diversity H_m of marker Asw1 was 0.66 and all of the populations were polymorphic for this marker (Table 3). On the other hand, the least polymorphic marker was Asw5, with two alleles over all populations, with an average A_m of 1.6 alleles per population and gene diversity H_m = 0.19, and it was polymorphic in five of the nine populations (Table 3). No bias of field sample size on the allelic richness was found (Figure 2).

When comparing populations, the lowest frequency of polymorphic markers P, average heterozygosity, *H*, and allelic richness, *A*, across loci, were found in the commercial population KBS-9 (P = 0.5, *H* = 0.21, *A* = 1.8). The average heterozygosity (*H*) in KBS-9 was 1.6–2.5 times lower than the average heterozygosity present in wild *A. swirskii* (*H* = 0.34–0.53, P = 0.83–1, *A* = 2.5–3.7) (Table 2, Figure 3). The heterozygosity H_m of the commercial population KBS-9 ranged from 0.65 and A_m = 4 alleles for marker Asw1, to 0 for markers Asw2, Asw3, and Asw5 (Table 3). The highest average heterozygosity across loci was found in *A. swirskii* population Had-4 (*H* = 0.53), collected from sweet pepper crops in Israel, and ranged from H_m = 0.65 and A_m = 3 alleles for marker Asw4, to H_m = 0.38 and A_m = 2 alleles for marker Asw2 (Table 3). The commercial population KBS-9 had the highest genetic differentiation from all the natural populations, as indicated by the pairwise F_{ST} values (Table 4).

The genetic distance measures Nei's D_A, Reynolds' D_R, and Cavalli-Sforza chord D_{CS} were calculated between populations and the highest average distances were found for the commercial population and for the population collected on kiwi (Table 2). Multidimensional scaling and Neighbour joining clustering of the Nei's D_A genetic distances derived from the microsatellite genotyping of the wild *A. swirskii* populations suggest three possible clusters, one is the population BHE-2-Kiwi separated from the other populations, a second including the populations Had-6-Citrus, HAs-7-Cotton, and HAs-8-Citrus, and last a cluster of the populations Had-5-Jerusalem artichoke, Had-4-Sweet pepper, HV-1-Citrus, and EY-3-Citrus – the commercial population KBS-9 falls in the latter cluster of wild populations (Figures 4 and 5).

Table 4. Pairwise F_{ST} (below the diagonal) and Nei's D_A pairwise genetic distances (above the diagonal) among nine populations of *Amblyseius swirskii* (eight wild and one commercial, KBS-9)

	BHE-2	EY-3	HV-1	HAs-7	HAs-8	Had-4	Had-5	Had-6	KBS-9
BHE-2		0.376	0.507	0.581	0.560	0.471	0.403	0.506	0.586
EY-3	0.0013		0.039	0.303	0.337	0.247	0.222	0.271	0.236
HV-1	0.0312	0		0.410	0.404	0.253	0.278	0.363	0.330
HAs-7	0.1387	0	0		0.301	0.641	0.404	0.309	0.488
HAs-8	0	0	0	0		0.466	0.446	0.443	0.429
Had-4	0	0	0	0	0		0.369	0.431	0.466
Had-5	0.0928	0	0	0.0417	0	0		0.376	0.397
Had-6	0.1117	0	0	0.0605	0	0	0.0147		0.506
KBS-9	0.3428	0.1542	0.1841	0.2916	0.1332	0.1187	0.2457	0.2646	

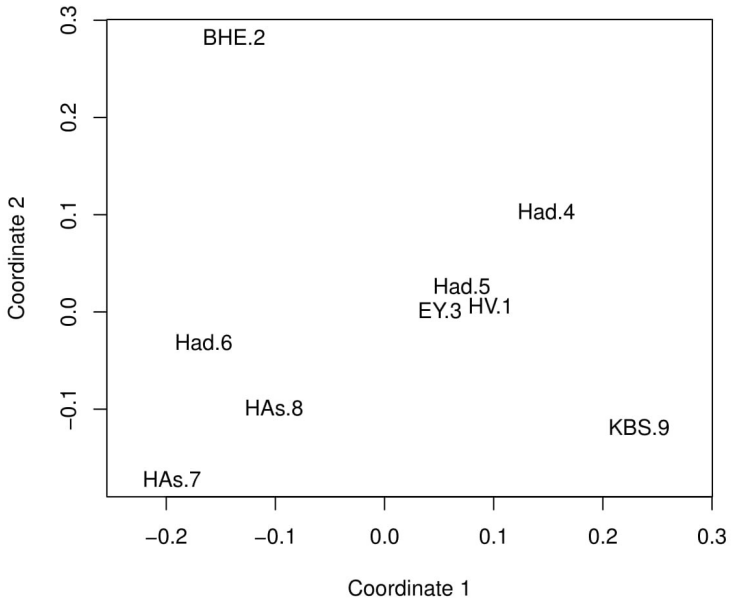


Figure 4. Multidimensional scaling using the Nei's D_A genetic distance matrix for nine populations of *Amblyseius swirskii* (eight wild and one commercial, KBS-9).

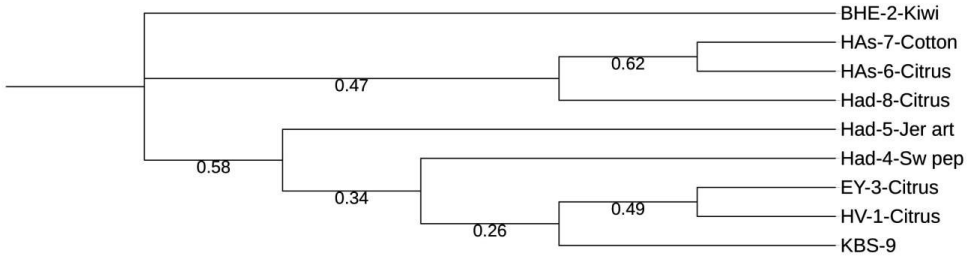


Figure 5. Neighbour-joining clustering of six microsatellite markers with 1,000 bootstrap using Nei's D_A genetic distance matrix for nine populations of *Amblyseius swirskii* (eight wild and one commercial, KBS-9).

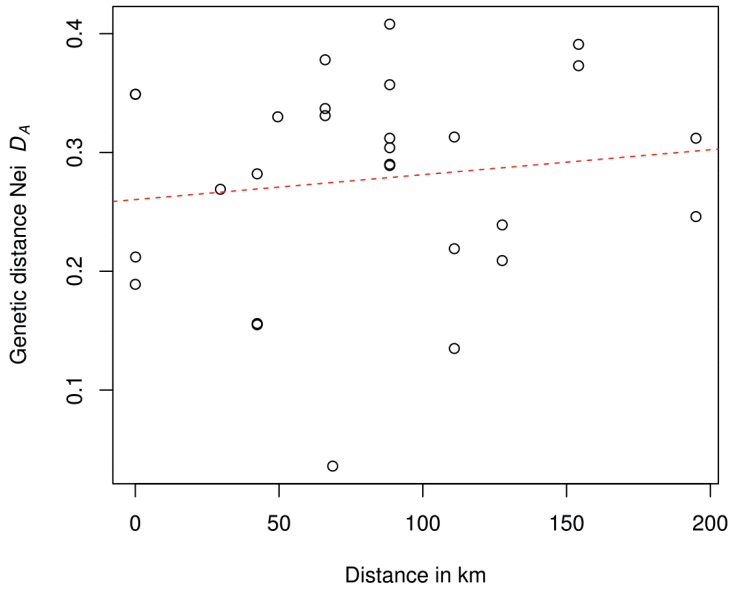


Figure 6. Isolation-by-distance using pairwise Nei's D_A genetic distances vs. geographic distances between eight wild populations of *Amblyseius swirskii* collected in Israel. Dashed line represents the linear regression fit (R^2 = 0.01653, P = 0.51).

The Neighbour Joining analysis yielded results very similar to the Multidimensional scaling; however, moderate bootstrap values (<70%) were computed, because the number of microsatellite loci employed in the analysis was lower than the number of populations analysed (Figure 5). No evidence for isolation-by-distance was found among the wild A. swirskii populations, according to the Mantel correlation test (simulated P = 0.30) and the linear regression fit (Figure 6).

Discussion

The objective of our study was to estimate and compare the genetic diversity in natural populations and a mass-reared commercial population of the biological control agent *A. swirskii* by analysing microsatellite markers. In absence of a published genome of *A. swirskii*, the identification of suitable markers for the genetic population inference required the de novo sequencing of the genome. However, next-generation sequencing approaches, such as 'whole genome shotgun' (WGS) sequencing, need high DNA yield from a highly inbred population, which is intricate to obtain for this minute predatory mite because of its size. In this study, we selected the MinION sequencing platform because it allows small DNA input, and generates long sequence reads (Goodwin et al., 2015). Furthermore, the entire sequencing process, including library preparation, can be done in-house and is cost-effective.

Despite the benefits of nanopore sequencing in generating genomic sequences from the inbred *A. swirskii* strain, the 30,000 sequences (256 Mb) were not enough to generate a genome assembly, because these sequences did not provide an adequate coverage of the whole genome. We therefore refrained from further assembling the reads into scaffolds, and used the raw reads as input for our microsatellite analysis. However, these reads were beneficial for mining and developing microsatellites for population genetic studies, and add to the low amount of publicly available sequence data for *A. swirskii*, and to the microsatellite panels developed for phytoseiids of economic importance (e.g., Sabater-Muñoz et al., 2012).

Microsatellite mining of the raw reads, primer design, and filtering yielded 2,w423 microsatellite markers and 622 candidate primer pairs for population genetic analysis. However, when 24 of those primers were tested for efficiency

and specificity, only a small fraction (25%) complied with both criteria. One explanation for the low success rate of our microsatellites could be the high error rate of the nanopore sequencing platform with 10% of the bases called wrongly (Laver et al., 2015) combined with insufficient read depth to properly correct for these errors.

Another issue is the likely contamination of the sequencing starting material by the food source cattail pollen (*T. latifolia*). Contamination by the food source may have affected the outcome of the genome sequencing of *A. swirskii* providing sequence reads of the contaminant organism (in this case the *T. latifolia*) as well, and has been outlined as a putative hazard for predatory mite studies before (Hoy et al., 2013). Predatory mites are tiny organisms whose gut occupies most of their total volume and it is very difficult to be dissected and excluded, as it is the common technique in larger arthropod species studies. Also, the detection window of prey DNA in the guts of arthropod predators (analysed by PCR) can range from a few hours to various days post-feeding (King et al., 2008). Unfortunately, the genome of *T. latifolia* is not known yet and its sequences cannot be filtered out from the consensus genomic sequences. Hence, in order to improve the methodology, the food source provided to the predatory mites should be a species with a genome sequence available, as those sequences can be filtered out and removed from the sequencing reads. Furthermore, the coverage of the genome should be improved in further genome studies.

Individual genotyping of mites for population genetic inferences is a costly and elaborate method, mainly because of their tiny size and the low amount of DNA available for extraction, which limits the number of loci analysed per individual. However, DNA pooling for assessing relative differences of allele frequency among populations can overcome these limitations and provide a cost-effective alternative (Skalski et al., 2006). Although the effective number of peaks obtained from DNA pools is systematically overestimated compared to the actual number of alleles obtained from individual typing, those estimates can still detect relative differences in allele frequencies among DNA pools (Skalski et al., 2006). Microsatellite-based genetic diversity estimates obtained from DNA pools were found to strongly correlate with those obtained from individually typing in previous studies in chickens and pigs

(Hillel et al., 2003; Megens et al., 2008). In those studies, the effective number of peaks correlated strongly with the effective number of alleles, as did the respective heterozygosity estimates. Hence, the use of DNA pools provide reliable estimates for the population's diversity (Hillel et al., 2003; Skalski et al., 2006; Megens et al., 2008). Also, the three genetic distance measures –D_A, D_{CS}, and D_R– for different evolutionary models correlate significantly between them, similar to what was found in chicken and pig lines for pooled samples, suggesting that the downstream inference based on D_A is robust under alternative evolutionary scenarios as well (Hillel et al., 2003; Megens et al., 2008).

A small number of microsatellite loci (2,423) was identified in the genomic sequences of *A. swirskii*, similar to the low number of loci found in *Tetranychus urticae* Koch genome (Grbić et al., 2011). However, the actual number of microsatellite loci is expected to be higher, because the genome of *A. swirskii* was partially covered, yet our estimate is useful for the calculation of the microsatellite density on the genome. For two other Acari species – the phytoseiid *Amblyseius fallacis* (Garman) and the ixodid tick *Ixodes scapularis* (Say) – a low abundance of microsatellites has been proposed as well, following the generation of plasmid libraries enriched for microsatellites (Navajas et al., 1998; Fagerberg et al., 2001), although this method does not provide absolute estimates of microsatellite density to compare.

In *A. swirskii*, dinucleotide repeats are the most abundant type of microsatellite, trinucleotides are less than half frequent, and longer repeats, such as tetra-, penta-, and hexanucleotides, are found markedly less, similar to most arthropod species (Pannebakker et al., 2010) and different from *T. urticae*, in which trinucleotides are the most abundant repeat motif (Grbić et al., 2011). It is unclear why mites have a low microsatellite density compared to other arthropod species (e.g., Pannebakker et al., 2010; Abe & Pannebakker, 2017), but this does not need to reflect an overall low genomic diversity. For instance, despite the low microsatellite abundance, Van Zee et al. (2013) found a very high SNP density in the genome of the tick *I. scapularis*, suggesting SNPs to be a more suitable marker for population genetic analysis. Nevertheless, we did observe high levels of polymorphism in the six microsatellite markers employed in *A. swirskii* in the current study, as reflected by the finding that

the least polymorphic marker was polymorphic in 50% of the wild populations and six out of eight markers were polymorphic in all of the wild populations. Our finding is in accordance with the elevated molecular evolution found for the phytoseiid mite *Metaseiulus occidentalis* (Nesbitt) (Hoy et al., 2016).

High genetic variation is observed in natural populations of *A. swirskii*, with the mean observed heterozygosity, *H*, ranging from 0.34 to 0.53, when compared to the commercial population (0.21), despite the fact that some natural populations were established by a low number of individuals. Heterozygosity estimates based on pooled microsatellite analysis found in our study are comparable to the average observed heterozygosity found in a field population of the predatory mite *Neoseiulus womersleyi* (Schicha) (Hinomoto et al., 2011). Considerable genetic variation was found also for tetranychid mites; *T. urticae* field populations from Europe had a spatial genetic structure along a latitudinal gradient (Carbonnelle et al., 2007). *Tetranychus turkestani* (Ugarov & Nikolskii) collected on crops and weeds in southern France had comparable average heterozygosity and mites living on different host plants did not demonstrate clear evidence for genetic differentiation, similar to our results (Bailly et al., 2004). Our results did not show a correlation between host plant and genetic diversity in the wild populations; however, the BHE-2 population collected on kiwi plants did show a lower genetic diversity, which might be due to the presence of high trichome density on this plant. Plant trichomes can affect negatively the predators by impeding their movement and causing their entrapment (Riddick & Simmons, 2014). Hence, the populations of predatory mites can be smaller on plants with high pubescence compared to glabrous plants.

Natural populations are often characterized by higher levels of genetic variation compared to captive-bred organisms because genetic diversity is crucial for their survival under fluctuating environmental conditions and diverse ecosystems (Barrett & Schluter, 2008). For example, population genetic analysis of chicken using microsatellites showed that wild populations of the red junglefowl (*Gallus gallus gallus* (L.)) had 3 times higher average heterozygote frequencies than a European, domesticated chicken population (Granevitze et al., 2007). Commercial *A. swirskii* were polymorphic only in 50% of the microsatellite markers and had lower average

allelic richness and lower heterozygosity estimates compared to their wild counterparts, indicating a very low genetic diversity. Low genetic variation of a commercial biological control agent may affect their resilience when unexpected and even minor changes or fluctuations of the environmental conditions take place (Lommen et al., 2017; Wright & Bennett, 2018). For instance, in *D. melanogaster* it has been demonstrated that for higher inbreeding levels, the impact of environmental stress becomes significantly greater and can even lead to extinction (Bijlsma et al., 2000).

The decline in the genetic variation in the mass-reared population of this biological control agent may affect traits that are important in its performance, such as reproductive parameters, prey preference, or predation rate. Our laboratory colonies of wild *A. swirskii* populations were founded by low numbers of mites in some cases (8-100); however, still in all cases their genetic diversity was higher than in the commercial strain. Despite the very high numbers at which commercial *A. swirskii* populations are reared, genetic diversity can be low if the initial founding population is low (Mackauer, 1976; Bartlett, 1993). Furthermore, even if a large founding population was used, it may have included a few genotypes that are favoured under mass-rearing conditions, which can result in a bottleneck that reduces the genetic diversity (Nunney, 2003). As the field environment is different from the mass-rearing conditions, adaptation to the rearing environment can have a negative impact on the fitness of mass-reared biological control agents in the field (Sørensen et al., 2012). Genetic change in mass-reared populations has been described in many arthropods (Nunney, 2003), including experimental populations of *D. melanogaster* (Woodworth et al., 2002), but also in parasitoids reared for biological control (Guzmán-Larralde et al., 2014; Retamal et al., 2016; Bertin et al., 2017) and various insects reared for 'sterile insect technique' programmes since the 1980s (Rössler, 1975; Bush & Neck, 1976; Wong & Nakahara, 1978; Vargas & Carey, 1989).

To further improve the predatory mite *A. swirskii* for augmentative biological control and prevent the decrease in traits related to its fitness, we would suggest enhancing the low genetic diversity present in the mass-reared commercial *A. swirskii* by adding genetic variation from natural populations (Nunney, 2003). However, a better understanding of the levels of genetic

variation in mass-reared *A. swirskii* populations should be obtained first, before new genetic material is added to mass-rearing. The present study only tested a single commercial population limiting the scope of the conclusions. Additional genetic analysis of commercial populations can further assess the impact of genetic diversity on the performance of *A. swirskii* as a biological control agent, which remains one of the most successful natural enemies in augmentative biological control of whiteflies and thrips in greenhouse vegetables, fruits, and ornamentals in several parts of the world (Calvo et al., 2015; Knapp et al., 2018).

Augmentative biological control of agricultural pests is the environmentally sustainable alternative to synthetic pesticides for safe food production and healthy ecosystems. We found that the genetic diversity of a widely used commercial biological control agent is limited compared to the natural conspecifics and this could have implications when the biological control agents need to adapt to adverse or novel conditions such as when being released to the field or greenhouse. Also, we showed that pooled microsatellite analysis provides a cost-effective method to determine the genetic diversity of minute biological control agents. Hence, adding to our knowledge approaches to determine the existing genetic variance of commercial biological control agents is imperative for the long-term success of IPM strategies and control programmes.

Acknowledgements

We thank Arnon Tabic from Biobee Israel for his help in collecting the field samples, Eric Palevsky from the ARO-Newe-Ya'ar Research Center (Volcani Center, Israel) for providing valuable information and material, and Kirsten Oude Lenferink from Koppert Biological Systems for her technical assistance with the inbred lines. This work was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 641456. JGC was supported by the Spanish Ministry of Economy and Competitiveness, Ramón y Cajal Program (RYC-2013-13834).

Supplementary Materials

Additional supporting information as included in the online version of this article follows.

Supplementary Table 1. Primer pairs for the mitochondrial and microsatellite markers with corresponding GenBank accession numbers, melting temperature (T_m), PCR product sizes, number of repeats, motifs, and fluorescent dyes

Marker	Accession no.	Primer direction	Primer sequence (5'-3')	T_m (°C)	PCR product size (bp)	No. repeats	Motif	Dye
Asw1	MK267176	F	TCTCGGTGGGTTCAAGGATG	55.5	203-225	13-24	AG	ATTO
		R	AACGTCGGAAATTGAGCTGG	55.5				
Asw2	MK267177	F	TAACCTCTTGCACCCTCCTG	55.6	237-253	4-8	AAAG	HEX
		R	TCCTTCACTCTGTCTCGACG	56.9				
Asw3	MK267178	F	CTAAGAGGTAGCAGCCACCG	58.3	440-443	5-6	AGC	ATTO
		R	TCGCCATGTTTGCTGTGTAC	55.2				
Asw4	MK267179	F	GAGGGAGCGAGCGTATCTTC	58.4	425-457	3-9	ACCG	HEX
		R	GTGGGTACGATGTTTGGCAC	54				
Asw5	MK267180	F	GCCTGCTCTTCGTCTTTGAC	56.7	419-423	5-6	AGAT	6FAM
		R	GGTCGCGTTACTTGGCTTAC	55.8				
Asw6	MK267181	F	CATCAGACAGCGATGCGATC	58.1	277-286	5-8	AGC	ATTO
		R	CAAGATGACGGCGGAAACTC	56.5				

Supplementary Table 2. The composition of the microsatellite repeats found in the genome of *Amblyseius swirskii*.

Motif type	Number	Frequency (%)
Dinucleotide	1747	72.10
Trinucleotide	573	23.65
Tetranucleotide	94	3.88
Pentanucleotide	6	0.25
Hexanucleotide	3	0.12
Total	2423	100

©E.RAS

Chapter 4

Braconidae revisited: *Bracon brevicornis* genome showcases the potential of linked-read sequencing in identifying a putative *complementary sex determiner* gene

Kim B. Ferguson[a], Bart A. Pannebakker[a], Alejandra Centurión[b], Joost van den Heuvel[a], Ronald Nieuwenhuis[c], Frank F. M. Becker[a], Elio Schijlen[c], Andra Thiel[b], Bas J. Zwaan[a], and Eveline C. Verhulst[d]

[a] Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands
[b] University of Bremen, FB02, Institute of Ecology, Population and Evolutionary Ecology Group, Bremen, Germany
[c] Wageningen University & Research, Bioscience, Wageningen, The Netherlands
[d] Wageningen University, Laboratory of Entomology, Wageningen, The Netherlands

ABSTRACT

Bracon brevicornis is an ectoparasitoid of a wide range of larval-stage Lepidopterans, including several pests of important crops, such as the corn borer, *Ostrinia nubilalis*. It is also one of the earliest documented cases of complementary sex determination in Hymenoptera. Here, we present the linked-read genome of *B. brevicornis*, complete with an *ab initio*-derived annotation and protein comparisons with fellow braconids, *Fopius arisanus* and *Diachasma alloem*. We demonstrate the potential of linked-read assemblies in exploring regions of heterozygosity and search for structural and homology-derived evidence of the *complementary sex determiner* gene (*csd*).

INTRODUCTION

Bracon brevicornis (Wesmael) is a gregarious ectoparasitoid of various Lepidoptera larvae, including many important pests, and is considered a cosmopolitan species (Temerak 1983b; Venkatesan, Jalali, and Srinivasamurthy 2009). In the past *B. brevicornis* has been classified under the genus *Habrobracon* (Speicher and Speicher 1940), *Microbracon* (Narayanan et al. 1954), or classified as one species with *Habrobracon/Bracon hebetor* (Puttarudriah and Basavanna 1956), however recent research shows that *B. brevicornis* and *B. hebetor* are genetically two distinct species (Kittel and Maeto 2019). In the field, *B. brevicornis* has shown potential as a biological control agent against important pest species in stored corn stalks, such as *Ostrinia nubilalis* and *Sesamia cretica* (Kares et al. 2010), or against the coconut moth, *Opisinia arenosella* (Venkatesan et al. 2009). In the laboratory, *B. brevicornis* attacks a wide range of larval host such as *Ephesthia kuehniella*, *Galleria* spp., and *Spodoptera* spp. (Temerak 1983a).

Work on *B. brevicornis* has included both laboratory and semi-field set-ups to determine both its efficacy as a biological control agent as well as its suitability as a study system. There are several studies on the biology of *B. brevicornis*, e.g. on population growth potential (Srinivasan and Chandrikamohan 2017), their host range (Temerak 1983a), interspecific competition (Venkatesan et al. 2009), clutch size and fitness (Villacañas de Castro and Thiel 2017), mate

choice (Thiel and Weeda 2014), diet (Temerak 1983b), and efficacy (Kares et al. 2010).

Within a phylogenetic perspective, *B. brevicornis* falls within the subfamily Braconinae, the largest of the cyclostome-forming braconid wasps (Chen and van Achterberg 2019). The presence of a cyclostome (round mouthpart) is a defining feature within braconid wasps, as it represents an unresolved evolutionary and systematic question: is the cyclostome a derived trait within certain branches, or an ancestral trait that has been lost in others (Chen and van Achterberg 2019)? Within the Braconinae, there have been multiple switches from ectoparasitism to endoparasitism and *vice versa*, and this combination of cyclostome and endoparasitism has been described as a "controversial topic" by braconid researchers and taxonomists (Chen and van Achterberg 2019). These systematic issues are far from being resolved, and more genomic data would be useful for future phylogenetic analyses (Chen and van Achterberg 2019). Yet, a representative genome for the Braconinae is currently lacking. As previously stated, *B. brevicornis* is an ectoparasitoid, and its position within a family that contains both types of parasitism lifestyles holds promise for further phylogenetic comparisons.

In addition, as being part of the order Hymenoptera, *B. brevicornis* has a haplodiploid sex determination system where males develop from unfertilized eggs and females develop from fertilized eggs (Cook and Crozier 1995; Heimpel and de Boer 2008). From a genetic perspective, *B. brevicornis* belongs to an interesting genus where sex determination and diploid male production have been widely studied (*B. hebetor*, Whiting and Whiting 1925; *B. brevicornis*, Speicher and Speicher 1940; *B. serinopae* Clark, Bertrand, and Smith 1963; reviewed in van Wilgenburg, Driessen, and Beukeboom 2006, *B.* spec. near *hebetor,* Holloway *et al*. 1999; and *B. variator*, A. Thiel, pers. comm.). Indeed, the first description of the complementary sex determination (CSD) mechanism was provided for *B. hebetor* (= *B. juglandis* by Whiting 1940, reviewed in Antolin *et al.* 2003), and recent work on *B. brevicornis* and polyploidy studies include diploid male fitness as well as ploidy-dependent mate choice behaviour (Thiel and Weeda 2014).

While straightforward to detect phenotypically through the formation of diploid males following inbreeding (van Wilgenburg et al. 2006), the molecular mechanism underlying CSD has thus far only been resolved in the honeybee *Apis mellifera* (L.) (Hymenoptera: Apidae), with the identification of the *complementary sex determiner* (*csd*) gene. Heterozygosity at this gene leads to female development, while hemi- and homozygous individuals develop into haploid and diploid males respectively (Beye et al. 2003). Therefore, inbreeding often leads to diploid male production in species with a CSD mechanism as it increases homozygosity. *Csd* is a duplication of *feminizer* (*fem*), a *transformer* (*tra*) ortholog (Hasselmann et al. 2008) that is conserved across many insect orders as part of the sex determination cascade (Geuverink and Beukeboom 2014). When heterozygous, *csd* initiates the female-specific splicing of *fem*, which then autoregulates its own female-specific splicing, ultimately resulting in female development. Within the Hymenoptera, more duplications of *tra*/*fem* have been identified in species that are presumed to have CSD (Geuverink and Beukeboom 2014), but these *tra*/*fem* duplications have not been analysed for potential heterozygosity. Also, additional hymenopteran genomes are necessary to understand the evolutionary history of *tra*/*fem* duplications and identify the genes underlying CSD. However, an assembled genome is usually haploid as areas of heterozygosity are collapsed in the final stages of assembly. Yet recent advances in sequencing and analysis gave us the ability to view heterozygous regions, known as "phases" in diploid assemblies, within a genome which allow us to investigate potential *csd* regions.

Here we report on the whole-genome sequencing of a pool of females from an isolated *B. brevicornis* strain using 10X Genomics technology that relies on linked-read sequencing (10x Genomics Inc., Pleasanton, CA, USA). Due to their long history of genetic isolation during laboratory rearing, the females in this strain are assumed to have a high level of homozygosity, whereas a *csd* locus would retain its heterozygosity. The 10X Genomics technology allows for generating phased data in which allelic variants can be identified after assembly. High-molecular weight DNA is partitioned into small droplets containing a unique barcode and adapter in such a way that only a few

DNA molecules are present within each droplet. Within each droplet the DNA is broken into pieces and the barcode (Gel Bead-in-Emulsion, "GEM") is ligated to each of the DNA fragments. This resulting library can then be sequenced on an Illumina sequence platform. In the assembly step the reads originating from the same fragment are organized by barcode and put together into synthetic long-read fragments. Importantly, it is nearly impossible that two fragments with opposing allelic-variances are together in the same droplet are nearly impossible (Weisenfeld et al. 2017). This technique therefore allowed us to identify potential *csd* candidates in the female-derived *B. brevicornis* genome after sequencing by studying the phased data containing the different haplotypes. Moreover, as *B. brevicornis* is a potential biological control agent of several pests, the availability of a full genome may provide effective ways to study and improve this species to grow it into an established biological control agent for Lepidopteran pests.

Methods

Species description and general rearing

Individuals of *B. brevicornis* were taken from the laboratory colony L06. The colony was initiated in 2006 from naturally parasitized *O. nubilalis* larvae collected in maize fields near Leipzig, Germany. Species identification was first carried out by Matthias Schöller and Cornelis van Achterberg based on morphological characteristics (B. Wührer, AMW Nützlinge, pers. comm.) Since collection, parasitoids have been reared on late instar larvae of the Mediterranean flour moth, *E. kuehniella* (Thiel and Weeda 2014). The species identity of strain L06 was recently revalidated based on molecular data and is entirely separate from its congeneric *B. hebetor* (Kittel and Maeto 2019).

DNA Extraction

Immediately following emergence, 100 to 120 female wasps were flash frozen in liquid nitrogen and ground with a mortar and pestle. Genomic DNA was extracted using a protocol modified from Chang, Puryear, and Cairney (Chang, Puryear, and Cairney 1993). Modifications include adding 300 µL BME to extraction buffer just before use. Instead of 10M LiCl, 0.7 volume isopropanol (100%) was added to the initial supernatant, after which it was divided into

1.5 mL Eppendorf tubes as 1 mL aliquots for subsequent extractions. The initial centrifugation step occurred at a slower rate and for a longer period of time to adjust for machine availability. Final pellets were dissolved in 50 µL autoclaved MQ and recombined at the end of the extraction process (1.0 mL). DNA concentration was measured with an Invitrogen Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA) with final assessments for DNA quality, amount, and fragment size confirmed via BioAnalyzer 2100 (Agilent, Santa Clara, California, USA).

10X Genomics Library Preparation and Sequencing

As the genome of *B. brevicornis* is relatively small for the scale of the 10X platform, there is a higher risk of overlapping fragments within single GEMs. In order to reduce this risk, genomic DNA of a larger and previously analysed genome (Tomato, *Solanum lycopersicon* (L.) (Solanaceae), commercial variety Heinz 1607) (Hosmani et al. 2019) was used as 'carrier DNA'. DNA extraction of *S. lycopersicon* followed the protocol of Hosmani *et al*. (Hosmani et al. 2019). The DNA of both *B. brevicornis* and *S. lycopersicon* was pooled in a 1:4 molar ratio.

One nanogram of this pooled DNA was used for 10X Genomics linked read library preparation following the Chromium Genome Reagent Kits Version 1 User Guide (CG-00022) (10x Genomics, Pleasanton, USA). Barcoded linked read DNA fragments were recovered for final Illumina library construction (Illumina, San Diego, USA). The library was used for 2 x 150 bp paired-end sequencing on one lane of an Illumina HiSeq 2500 at the business unit Bioscience of Wageningen University and Research (Wageningen, The Netherlands). Sequencing data was then used for basecalling and subsequent demultiplexing using Longranger (v2.2.2) (10X Genomics) (command `-mkfastq`), yielding 212,910,509 paired-end reads with a read length of 150 bp.

Assembly

To filter sequence data from Heinz tomato (*S. lycopersicon*) carrier DNA sequences, 23bp (16bp GEM + 7 bp spacer) were removed from forward reads and all reads were subsequently mapped to an in-house high quality

reference assembly of the Heinz genome using BWA-MEM v0.7.17 (Li 2013). Using SAMTOOLS v1.9 (Li et al. 2009), all unaligned read pairs (-F=12) were extracted and labelled non-Heinz. The assembly of the non-Heinz labelled read set was performed with SUPERNOVA assembler v2.1.0 (10X Genomics), using default settings including commands for both pseudohap (`--style=pseudohap`) and pseudohap2 (`--style=pseudohap2`) outputs (Weisenfeld et al. 2017). These commands determine the output from Supernova, the first being the final scaffold output (pseudohap), while the second is the so-called 'parallel pseudohaplotype' (pseudohap2) scaffolds that represent areas of divergence or phases (Weisenfeld et al. 2017). Phasing is flattened in the pseudohap output by selecting the region with higher mapping coverage, whereas in the pseudohap2 output is differentiated by ".1" and ".2" at the end of each scaffold name to denote phasing, though not all scaffolds are phased at this point due to lack of divergence during assembly.

To verify whether there were no Heinz leftovers in the assembly, MINIMAP2 v2.17-r941 (Li 2018) was used to align the assembly against the same Heinz assembly. Further examination on presence of possible non-*B. brevicornis* scaffolds, i.e. bacterial scaffolds from sample microbiome, was performed with BLOBTOOLS (v1.0) (Laetsch and Blaxter 2017), relying on MEGABLAST against the NCBI NT-NR database (Acland et al. 2014)(2018-11-19) (`max_target_seqs=1`, `max_hsps=1`, `evalue=1e-25`) for taxonomical classification and BWA-MEM mapping of reads against scaffolds for coverage statistics. Reads mapping only against "Arthropoda" classified scaffolds were then extracted and used for a final k-mer analysis using JELLYFISH v2.1.1 (`-C -m=21 -s=2000000000`) (Marçais and Kingsford 2011) and GENOMESCOPE (Vurture et al. 2017) to infer heterozygosity.

Assembly completeness was determined using BUSCO (v3.0.2) with the `insect_odb9` ortholog set and the fly training parameter (Simão et al. 2015) while assembly statistics were determined using QUAST (Gurevich et al. 2013). The aforementioned pseudohap2 scaffolds were used in *csd* analysis, while the pseudohap scaffolds are now the assembly used for annotation.

Ab initio gene finding and protein comparison

The coding sequences of two additional braconids (members of the subfamily Opiinae, and similar to the Braconinae belonging to the cyclostome subgroup (Chen and van Achterberg 2019; Li, van Achterberg, and Tan 2013)) were used for gene prediction and protein comparisons: *Fopius arisanus* (Sonan) (Hymenoptera: Braconidae) and *Diachasma alloeum* (Muesebeck) (Hymenoptera: Braconidae). Both sets of coding sequences were retrieved from the NCBI Assembly Database, version ASM8063v1 for *F. arisanus* and version Dall2.0 for *D. alloem* (Acland et al. 2014; Geib et al. 2017; Tvedte et al. 2019).

For gene prediction, Augustus (v2.5.5) was first used to predict genes from the *B. brevicornis* assembly (Stanke and Morgenstern 2005). Using BLAST, coding sequences of *F. arisanus* were set as a query to the genome of *B. brevicornis* using default parameters (except `minIdentity=50`) (Camacho et al. 2009). The result was converted into a hints file that was used to predict the genes of *B. brevicornis* using *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae) as the species parameter in Augustus (`--species=nasonia --extrinsiccCfgFile=extrinsic.E.cfg`).

After prediction, the protein sequences were retrieved and compared to both *F. arisanus* and *D. alloeum* (version Dall2.0) using Proteinortho (v6.0, `-p=blastp, -e=0.001`) (Lechner et al. 2011). From the orthology grouping generated by Proteinortho, gene names could be allocated to the predicted genes. Lengths of both these *B. brevicornis* genes and the orthologs of *F. arisanus* and *D. alloem* were retrieved using samtools for comparison (Li et al. 2009).

In silico identification of *feminizer* as a putative *csd* locus

The pseudohap2 files were deduplicated using the dedupe tool within BBTools (sourceforge.net/projects/bbmap/) (`ac=f`) to remove all parallel pseudohaplotypes that were complete duplicates as these scaffolds were not heterozygous. The remainder of the set contained both scaffolds that previously had a duplicate, as well as solitary scaffolds that did not have a partner scaffold. These unique scaffolds were removed using the "filter

by name" tool in BBTools, leaving 258 scaffolds, or 129 pairs of pseudohap2 scaffolds. Pairs were pairwise aligned in CLC Genomics Workbench 12 (Qiagen, Hilden, Germany) using default settings (`gap open cost=10, gap extension cost=1, end gap cost=free, alignment=very accurate`).

A local tBLASTn search against the entire *B. brevicornis* assembly was performed using the *Apis mellifera* Feminizer protein (NP_001128300) as query in Geneious Prime v2019.1.3 (http://www.geneious.com, (Kearse et al. 2012)). The protein of gene BBRV_07607 was used in an NCBI BLASTp against the nr database with default settings (Acland et al. 2014; Camacho et al. 2009). Next a region stretching from ~10Kbp upstream and downstream of the first and last tBLASTn hit in scaffold 12, respectively, was annotated using HMM plus similar protein-based gene prediction (FGENESH+, Softberry, http://www.softberry.com/) with *Nasonia vitripennis tra* (NP_001128299) and *N. vitripennis* for the specific gene-finding parameters (Solovyev 2007). Only this combination of settings resulted in a full-length annotation from TSS to poly-A with seven exons. The resulting protein prediction was used in a BLASTp search with default settings against the nr database. To annotate the potential *fem* duplication, a stretch of ~10Kbp directly upstream of the annotated putative fem was again annotated using FGENESH+ (Softberry) with *Nasonia vitripennis tra* (NP_001128299) and *N. vitripennis* for the specific gene-finding parameters (Solovyv et al. 2007). The predicted annotation contained five exons but lacked the last coding segment with stop codon. A protein alignment was made in Geneious Prime v2019.1.3 with *A. mellifera csd* (ABU68670) and *fem* (NP_001128300)*; N. vitripennis tra* (XP_001604794) *and B. brevicornis* putative *fem* and *B. brevicornis* putative *fem* duplicate (*fem1*), using MAFFT v7.450 with the following settings: `Algorithm=auto, Scoring matrix=BLOSUM62, Gap open penalty=1.53, Offset value=0.123` (Katoh 2002; Katoh and Standley 2013).

Microsynteny analysis

A microsynteny analysis was achieved by comparing the arrangement of a set of homologous genes directly upstream and downstream of *tra* or *fem* in *A. mellifera* and *N. vitripennis* using a combination of the online tool SimpleSynteny (Veltri, Wight, and Crouch 2016) and tBLASTn searches using default settings in

Geneious Prime. The scaffolds containing *fem* (*A. mellifera,* scaffold CM000059.5, 13.2Mbp in length), *tra* (*N. vitripennis,* scaffold NW_001820638.3, 3.7Mbp in length) or the putative *fem* (*B. brevicornis,* scaffold 12, 4.5 Mbp in length) were extracted from their respective genomes (*Apis*: GCA_000002195.1_Amel_4.5_genomic, *Nasonia*: nvi_ref_Nvit_2.1, *Bracon*: *B. brevicornis* assembly from this study) and searched with protein sequence from the following genes: *tra* (GeneID: 00121203), LOC100121225, LOC100678616, LOC100680007 originating from *N. vitripennis*; and *fem* (GeneID:724970), *csd* (GeneID:406074), LOC408733, LOC551408, LOC724886 originating from *A. mellifera*. The advanced settings for SimpleSynteny were as follows: `BLAST E-value Threshold=0.01`, `BLAST Alignment type=Gapped`, `Minimum Query Coverage Cutoff=1%`, `Circular Genome Mode=Off`. If the gene was not found within the extracted scaffold, it was searched for in the full genome assembly. For the image settings, `Gene Display Mode=Project Full-Length Gene`. This generated image was used together with results from the tBLASTn searches as template to draw the final figure. The final figure that we present in the Results and Discussion section depicts ~0.9Mbp of genomic region for all three species.

Data Availability

Raw sequence data for *B. brevicornis* after removal of carrier DNA and contamination, as well as the annotated assembly, can be found in the EMBL-EBI European Nucleotide Archive (ENA) under BioProject PRJEB35412. Contaminated pseudohap scaffolds are available for download alongside the two pseudohap2 FASTA files, more details are provided in the supplementary materials.

Results and Discussion

A total of 172 ng of *B. brevicornis* DNA was extracted, which was then reduced to 1 ng/µL for library preparation. Sequencing of the Heinz diluted library resulted in a total yield of 54 Gbp of data (corrected for 10X 23bp segment of forward reads). Mapping against the Heinz genome assembly showed a mapping percentage of 84.8%. There was a total of 30,278,915 unmapped pairs, comprising ~8.39 Gbp of data. This corresponds to the 4:1

ratio between Heinz and *B. brevicornis* DNA in the library. Further scaffold decontamination with BlobTools resulted in a separation of the assembly into *B. brevicornis* scaffolds and microbiome scaffolds. The final genome is 123,126,787 bp (123 Mbp) in size, comprised of 353 scaffolds (5.5% ambiguous nucleotides). This is similar to the projected physical genome size of 133 Mbp (J. G. de Boer, unpublished data, flow cytometry). BUSCO analysis indicates a completeness of 98.7% (single orthologs 97.0%, duplicate orthologs 1.7%).

K-mer analysis of the *B. brevicornis*-only read set showed an expected haploid genome length of ~115 mbp (105 Mbp unique, 10 Mbp repeat) and a heterozygosity of ~0.54%. Peak coverage was 27x.

Ab initio gene finding and protein comparison

In total, 12,688 genes were predicted, with an average coding sequence length of 529.86 amino acids. The number of genes correspond well to those found in *F. arisanus* (11,775) and *D. alloem* (13,273), the two closest relatives of *B. brevicornis* for which public data is available. ProteinOrtho analysis resulted in 7660 three-way orthology groups (7,830 *B. brevicornis* genes), while 362 othology groups contained proteins of *B. brevicornis* and *F. arisanus* (382 *B. brevicornis* genes), and 451 groups contained *B. brevicornis* and *D. alloem* genes (479 *B. brevicornis* genes). A large number of orthology groups (2,492) had no *B. brevicornis* genes, while 3,997 predicted genes remain ungrouped.

Compared to *F. arisanus*, the mean relative length of predicted *B. brevicornis* genes was 1.016, while the mean relative length for the two- and three-way orthology groups was 0.996. Similar results were obtained for comparisons to *D. alloem,* where mean relative length for *B. brevicornis* genes was 1.011 and 0.988 for the two- and three-way orthology groups. Furthermore, the pairwise lengths of all these proteins resemble each other very well (Figure 1).

Identification of a putative *feminizer* ortholog and duplication event

After deduplicating the similar parallel pseudohaplotype files, 6,706 scaffolds in total, the remainder of the set contained 3,420 scaffolds, of which 3,286 scaffolds were solitary and did not have a counterpart pseudohap2 for comparison. Some had had a previous duplicate removed in the

deduplication, while others never had a partner scaffold in the first place. These unique scaffolds were removed, leaving 258 scaffolds, or 129 pairs of pseudohap2 scaffolds. These putatively heterozygous scaffolds were good candidates to search for potential *csd* loci as these are presumed to be heterozygous in females.

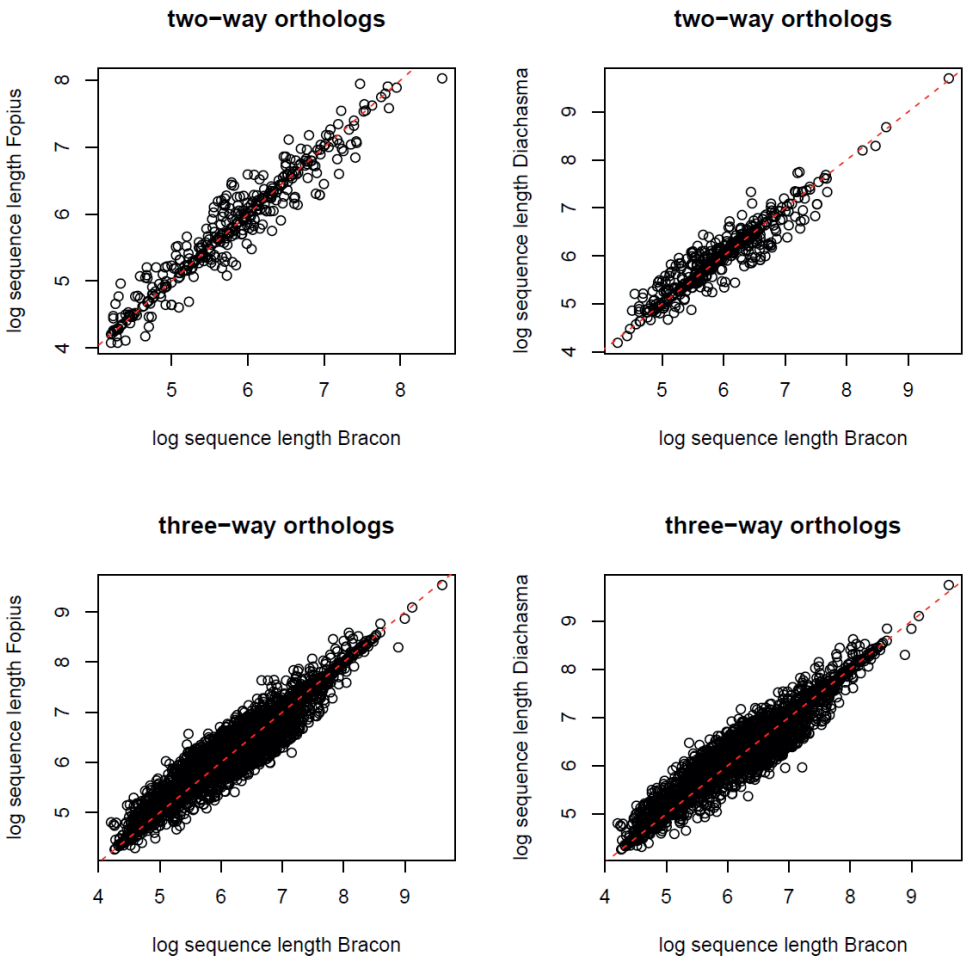


Figure 1. Protein length comparison between *Bracon brevicornis* and *Fopius arisanus*, a) two- and b) three-way orthologs, and *B. brevicornis* and *Diachasma alloem*, c) two- and d) three-way orthologs. Sequence lengths have been log-transformed; red dashed line indicates synteny.

So far, a *csd* gene has been sequenced only in species of bees of the genus *Apis,* and it is highly polymorphic, even within subspecies (Wang et al. 2012). It is located adjacent to the more conserved *feminizer* (*fem*) (Hasselmann et al. 2008), and we therefore started with localizing *feminizer* in the genome. As *feminizer* (or its ortholog *transformer*, *tra*) was not identified in the *ab-initio* annotation, we used a local tBLASTn search to find *fem* in the assembly. Four hits with E-value from 5.86e-04 to 8.59e-08 were found in scaffold 12. Searching the annotation using part of the tBLASTn result shows that it is annotated as BBRV_07607 which gave a first hit with protein O-glucosyltransferase 2 (*Diachasma alloeum*) after a BLASTp search, and no *fem* or *tra* hits were found. A closer inspection showed that BBRV_07607 is annotated as fusion protein with the N-terminal part resembling *fem* and the C-terminal part putatively encoding *O-glucosyltransferase 2*. Next, we used FGENESH+ to re-annotate the genomic region, resulting in a full-length putative *B. brevicornis feminizer* (*Bbfem*) ortholog containing seven exons (Figure 2). We found that the two *fem*/*tra* signature domains in Hymenoptera, the Hymenoptera domain (Verhulst, van de Zande, and Beukeboom 2010) and CAM domain (putative autoregulatory domain) (Hediger et al. 2010), are present in the putative *fem* ortholog, but are also duplicated upstream of putative *Bbfem*. A second manual re-annotation step showed that a partial *fem*-duplicate is encoded directly upstream of putative *Bbfem* containing five exons (Figure 2), which we denote here as *Bbfem1* as suggested by Koch *et al*. (Koch et al. 2014). The level of potential heterozygosity in the area encoding *Bbfem* and *Bbfem1* is the highest when compared across all 129 pairs of pseudohap2 scaffolds (Figure 2).

A protein alignment showed that the full-length putative *Bbfem* as well as *Bbfem1* are highly similar to each other and both contain all known *fem*/*tra* domains (Supplementary Figure 1). *Bbfem1* lacks a notably long first Arginine/Serine (RS)-rich region which is present only in *Bbfem* (124-153aa), but it otherwise appears to encode for a full-length protein. The *csd*-specific hypervariable domain (Supplementary Figure 1, purple text; (Beye et al. 2003)) is not present in *Bbfem* nor in *Bbfem1*.

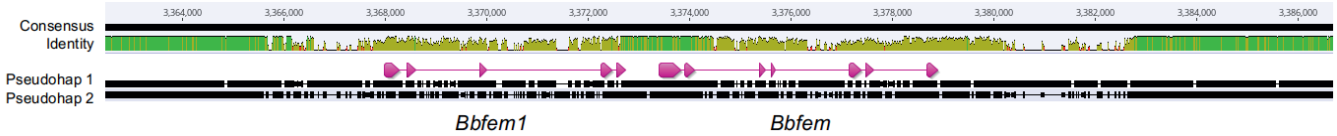


Figure 2. Annotation of *Bbfem* and *Bbfem1* on the alignment of pseudohaplotype track 1 and 2 in Geneious Prime v2019.1.3 (http://www.geneious.com, (Kearse et al. 2012)). Within the assembled genome, this section corresponds to a region on scaffold 12. The *Bbfem1* annotation lacks the last coding segment with stop codon. The identity track shows the amount of sequence identity across an arbitrary window (depending on zoom setting) and can be used as a proxy for heterozygosity. Green is identical, yellow is mismatch, red is no match due to introduced gaps during alignment. The coding regions of *Bbfem1* and *Bbfem* are in a high putatively heterozygous region.

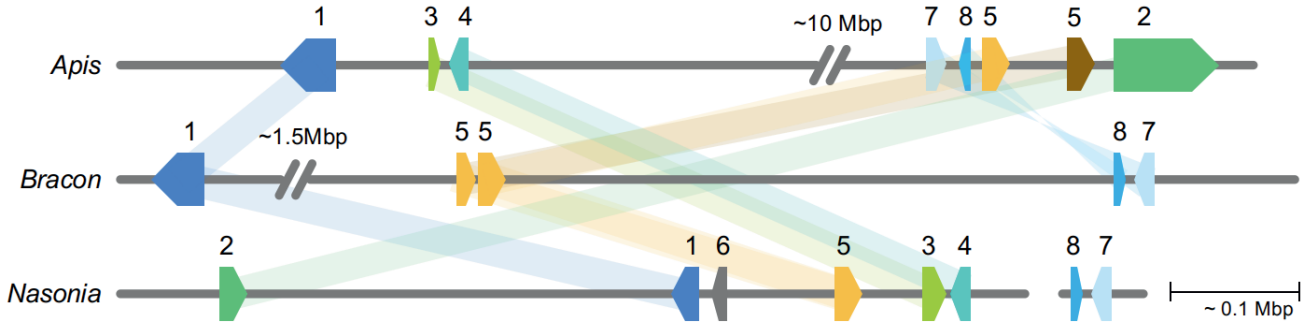


Figure 3. Microsynteny of genomic regions containing *tra/fem* paralogues. Shown is ~0.9 Mbp of genomic region of *Apis mellifera*, *Bracon brevicornis* and *Nasonia vitripennis*, containing the approximate coding region for 1. LOC100680007 (dark blue), 2. LOC408733 (green), 3. LOC100121225 (lime), 4. LOC100678616 (cyan), 5. *tra/fem/fem1* (yellow) and *csd* (brown), 6. LOC107980471 (gray), 7. LOC724886 (blue), 8. LOC551408 (light blue). Locus 2 is located on a different scaffold in *B. brevicornis*, locus 3 and 4 are not present in *B. brevicornis*. Locus 6 is unique to *N. vitripennis*, and locus 7 and 8 are located on a different scaffold in *N. vitripennis*, which is depicted on the right. Both 7 and 8 are in the same order and orientation as in *B. brevicornis*, but reversed in *A. mellifera*.

MICROSYNTENY ANALYSIS OF PUTATIVE *FEM* ENCODING REGION

We compared the orthologous gene arrangement of a number of genes up- and downstream of *N. vitripennis* tra and *A. mellifera fem* and *csd,* with the genomic organization of the *Bbfem* region (Figure 3). *N. vitripennis* LOC100680007 is present in the *tra*/*fem* containing scaffolds of all three genomes, while *A. mellifera* LOC408733 has both translocated closer to *Nasonia tra* and to a different scaffold in *B. brevicornis. N. vitripennis* LOC100121225 and LOC100678616 are encoded in opposing directions in both *A. mellifera* and *N. vitripennis* but are both downstream of *tra* in *N. vitripennis* and upstream of *fem* and *csd* in *A. mellifera.* There is no match for both genes in *B. brevicornis*. *A. mellifera* LOC724886 and LOC551408 are encoded in opposing directions with the same orientation in both *N. vitripennis* and *A. mellifera* but are reversed in *B. brevicornis* and downstream of *Bbfem* and *Bbfem1* while they are upstream of *csd* and *fem* in *A. mellifera*. In *N. vitripennis* both genes are not located in the *tra* containing scaffold but in another scaffold indicating that this region has undergone chromosomal rearrangements.

CONCLUSIONS AND PERSPECTIVES

Here, we present the genome of the braconid wasp *Bracon brevicornis*, a parasitoid wasp that not only has biological control applications, but also offers potential as a study system for future analyses into braconid phylogenetics and gene evolution. With no previous genomes available for the subfamily Braconinae, the most specious of the braconid wasps, the resources and investigations presented here fill this gap. Our linked-read library, assisted by carrier DNA of *S. lycopersicon,* has resulted in a highly contiguous, very complete assembly, comprised of just 353 scaffolds and 12,688 genes. This gene count is similar to related species, and in further protein length comparisons, the proteins are highly similar. This indicates that the predicted genes are highly complete, a necessary feature for any future phylogenetic comparisons between species or families.

We utilized the 10X Genomics linked-read approach to obtain pseudohaploid information that would allow us to search for potential *csd* loci *in silico.* As a

substantial number of scaffolds were putatively heterozygous, we used the notion that in *A. mellifera, csd* is located adjacent to *fem* (Hasselmann et al. 2008) to limit our search for *csd* candidates. We manually annotated a putative *B. brevicornis fem* and a partial *Bbfem* duplicate that is highly similar, and both genes encode all known *tra/fem* protein domains (Supplementary Figure 1) (Verhulst et al. 2010). Both genes are in a small region that is highly heterozygous, especially when compared to the remainder of the scaffold, which would suggest true heterozygosity and not assembly error, but also when compared to the level of heterozygosity in the other 128 aligned pseudohap2 scaffolds.

Our synteny analysis showed only little structural conservation between *B. brevicornis*, and *A. mellifera* and *N. vitripennis* with the translocation of LOC408733 (*A. mellifora*) and the absence of LOC100121225 and LOC100678616 (*N. vitripennis*) in the *B. brevicornis* genome region. It is known that genomic regions encoding sex determination genes are dynamic in nature, showing both duplications and translocations (Dechaud et al. 2019). Also, *tra/fem* duplications have been shown in CSD systems before, most notably in *A. mellifera* where a *fem* gene duplication event resulted in it becoming a *csd* locus (Gempe et al. 2009; Hasselmann et al. 2008). However, also in non-CSD systems *tra* duplications have been observed (Geuverink et al. 2018; Geuverink and Beukeboom 2014; Jia et al. 2016). Although there is some debate on whether *fem* paralogs originated due to a single duplication event and functions as *csd* (Schmieder, Colinet, and Poirié 2012), or evolved multiple times independently and may have other functions (Koch et al. 2014), we suggest that the *Bbfem* paralog, *Bbfem1*, is a good *csd* gene candidate in *B. brevicornis*. However, in-depth analyses are required to verify this. Ultimately, our presented genome with its pseudohaploid information provides multiple opportunities for future studies, such as to improve the biological control opportunities with this species, but also to shed light on the evolutionary history of complementary sex determination systems.

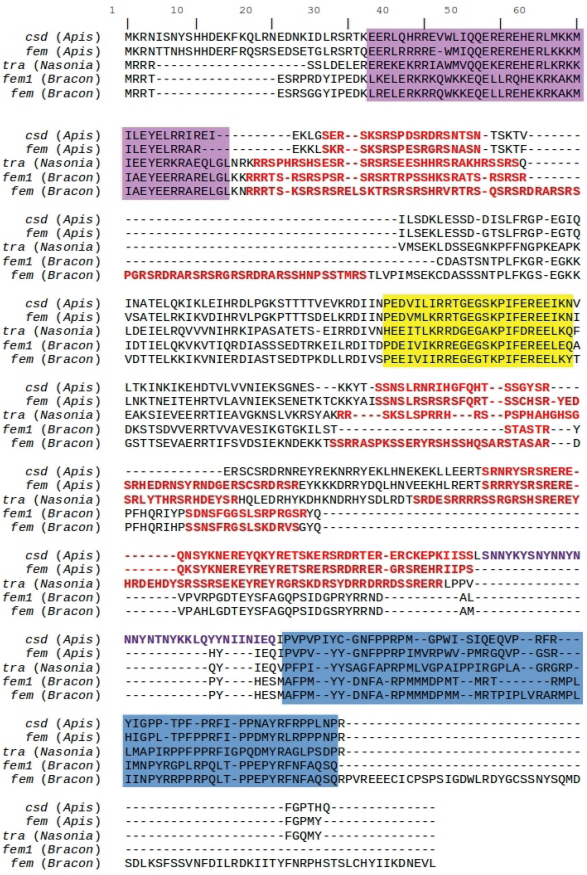
Acknowledgements and funding

We would like to acknowledge Jetske de Boer for information on *B. brevicornis* genome size, Martin Hasselmann for discussion on honeybee *csd*, and Elzemiek Geuverink for discussions on *tra/fem* duplicates in Hymenoptera. This project was

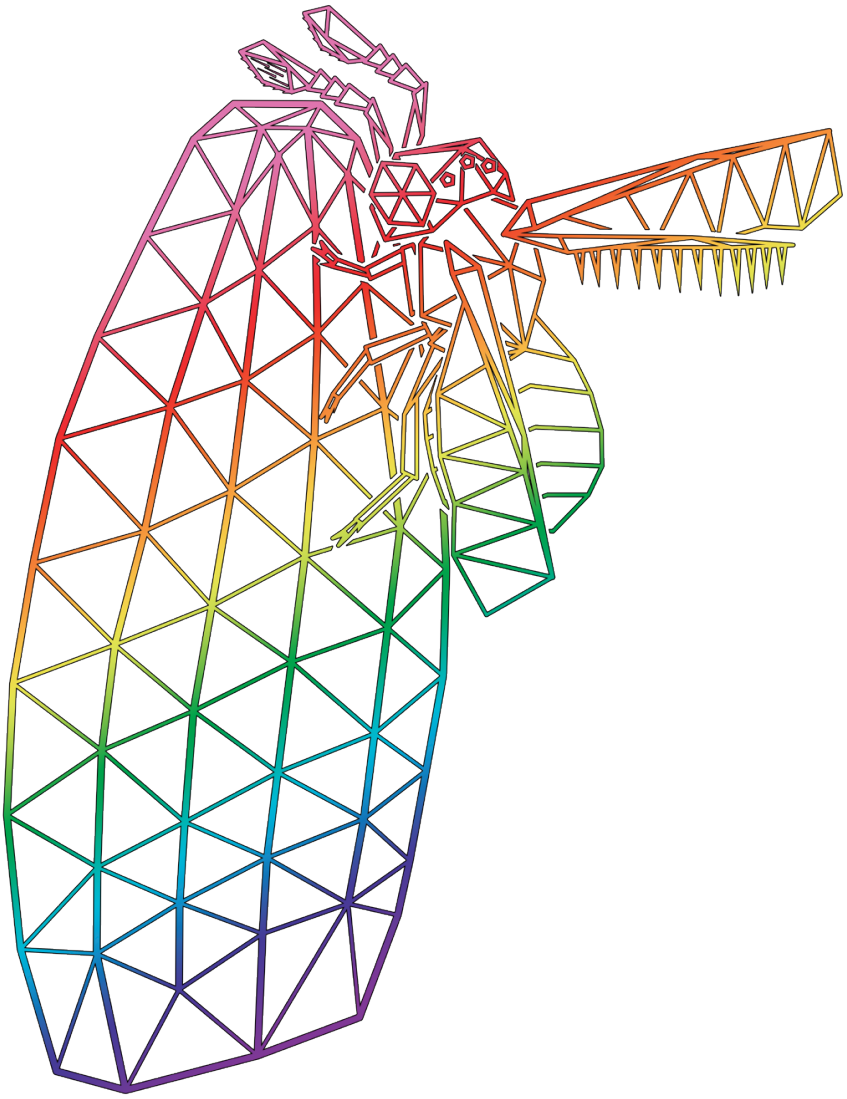
funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 641456.

SUPPLEMENTARY MATERIALS

Additional supplementary material from this study (contaminated scaffolds, pseudohap2 scaffolds) are available on the DANS EASY Repository, DOI:10.17026/dans-xn6-pjm8



Supplementary Figure 1. Protein alignment of *Apis mellifera csd* (ABU68670) and fem (NP_001128300), *Nasonia vitripennis tra* (XP_001604794), *Bracon brevicornis fem* and *fem1*. Purple shading indicates Hymenoptera domain (Verhulst et al. 2010), yellow shading indicates CAM domain (Hediger et al. 2010), blue shading indicates Proline (P)-rich region, red text colour indicates Arginine/Serine (RS)-rich regions, and purple text colour indicates hypervariable region in *csd* (Beye et al. 2003).



Chapter 5

Hybrid genome assembly and annotation of egg parasitoid and biological control agent *Trichogramma brassicae*

Kim B. Ferguson[a], Tore Kursch-Metz[b,c], Eveline C. Verhulst[d], and Bart A. Pannebakker[a]

[a] Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands
[b] Technische Universität Darmstadt, Department of Biology, Darmstadt, Germany
[c] AMW Nützlinge GmbH, Pfungstadt, Germany
[d] Wageningen University, Laboratory of Entomology, Wageningen, The Netherlands

Abstract

Trichogramma brassicae (Bezdenko) are egg parasitoids that are used throughout the world as biological control agents and in laboratories as model species. Despite this ubiquity, few genetic resources exist beyond COI, ITS2, and RAPD markers. Aided by a *Wolbachia* infection, a wild-caught strain from Germany was reared for low heterozygosity and sequenced in a hybrid de novo strategy, after which several assembling strategies were evaluated. The best assembly, derived from a DBG2OLC-based pipeline, yielded a genome of 235 Mbp in 1,572 contigs with an N50 of 556,663 bp. Following a rigorous *ab initio*-, homology-, and evidence-based annotation, 16,905 genes were annotated and functionally described. As an example of the utility of the genome, a simple ortholog cluster analysis was performed with sister species *T. pretiosum*, revealing over 6000 shared clusters and under 400 clusters unique to each species. The genome and transcriptome presented here provides an essential resource for comparative genomics of the commercially relevant genus *Trichogramma*, but also for research into molecular evolution, ecology, and breeding of *T. brassicae*.

Introduction

The chalcidoid *Trichogramma brassicae* (Bezdenko) (Hymenoptera: Trichogrammatidae) is a minute parasitoid wasp (~0.5 mm in length) that develops within the eggs of other insects (Smith, 1996). For over 50 years, it has been in use world-wide as a biological control agent as many lepidopteran pests of different crops are suitable hosts (Polaszek, 2009). The most common application of *T. brassicae* in Europe is against *Ostrinia nubilalis* (Hubner) (Lepidoptera: Pyralidae), the European corn borer. For example, in 2003 alone, over 11,000 ha of maize in Germany was treated with *T. brassicae* (Zimmermann, 2004). It is also released against lepidopteran pests in spinach fields as well as in greenhouses (e.g. tomato, pepper, and cucumber) (Klug and Meyhöfer, 2009). With its wide application in biological control, *T. brassicae* is a well-studied species. Field trials have been conducted on several aspects, such as host location and dispersal behaviour (Suverkropp et al., 2010, 2009), overwintering ability (Babendreier et al., 2003), while other biological control related studies considered issues related to low temperature storage (Lessard

and Boivin, 2013), reaction to insecticides (Delpuech and Delahaye, 2013; Ghorbani et al., 2016; Jamshidnia et al., 2018; Liu and Zhang, 2012; Thubru et al., 2018), or risk assessment (Kuske et al., 2004).

Next to its application as a biological control agent, this tiny parasitoid has been used in other research, both in genetic studies (Cruaud et al., 2018; Laurent et al., 1998; Wajnberg, 1993) and ecological studies (Cusumano et al., 2015; Fatouros and Huigens, 2012; Huigens et al., 2009). In addition, several initiatives investigate the infection of *T. brassicae* with *Wolbachia* bacteria (Ivezić et al., 2018; Poorjavad et al., 2012) and the consequences of such an infection (Farrokhi et al., 2010; Poorjavad et al., 2018; Rahimi-Kaldeh et al., 2018). As *T. brassicae* is a cryptic species with several other congenerics, misidentification and misclassification is a known issue (Polaszek, 2009). In response, molecular identification of trichogrammatids is well studied and established (Ivezić et al., 2018; Rugman-Jones and Stouthamer, 2017; Stouthamer et al., 1999; Sumer et al., 2009). Recently, several RADseq libraries were constructed from single *T. brassicae* wasps to aide in resolving the aforementioned phylogenetic issues within *Trichogramma* (Cruaud et al., 2018). Otherwise, the genomics of *T. brassicae* have largely been neglected even though a well annotated genome would allow researchers and biological control practitioners access to a wealth of information and open new avenues for comparative genomics and transcriptomics for evolutionary, ecological, and applied research.

Here, we report the whole-genome sequencing and annotation of a *T. brassicae* strain infected by *Wolbachia* that had thelytokous reproduction, in which females arise from unfertilized eggs. A hybrid *de novo* sequencing strategy was chosen to address two common issues: we used long PacBio Sequel reads to bridge the large segments of repetitive sequences often found in Hymenoptera, while countering the error bias of long read technology with the accuracy of Illumina short reads. A similar strategy was recently applied to improve the *Apis mellifera* genome, where the long PacBio reads were the backbone that boosted the overall contiguity of the genome, alongside the incorporation of repetitive regions (Wallberg et al., 2019).

In this report, we present the hybrid *de novo* genome of *T. brassicae*. Three different assemblers were evaluated, and the most complete genome assembly was used for decontamination and *ab initio*-, homology-, and evidence-based annotation. The resulting annotation was functionally described using gene ontology analysis. Finally, a heterozygosity comparison and simple ortholog cluster analysis with the congeneric *T. pretiosum* was performed, which can be considered a starting point for future comparative genomics of the commercially important genus *Trichogramma*.

Methods

Species origin and description

Individuals of *Trichogramma brassicae* were acquired by AMW Nützlinge GmbH (Pfungstadt, Germany). The strain was baited in May 2013 in an apple orchard near Eberstadt, Germany. The orchard was surrounded by blackberry hedges, forest, and other orchards. For baiting, the eggs of *Sitotroga cerealella* (Olivier) (Lepidoptera: Gelechiidae) (Mega Corn Ltd., Bulgaria) were glued on paper cards (AMW Nützlinge GmbH, Germany), usually used for releasing *Trichogramma* sp. in maize fields and households. These cards were placed directly into the trees, approximately two meters above ground. After five days in the field, baiting cards were collected and incubated together at 25°C. Following emergence, individuals were kept together, offered *S. cerealella* eggs, and reared in a climate chamber (27±2°C, L:D=24:0h for four days, then transferred to16±2°C, L:D=0:24h until emergence).

In 2016, the offspring of twenty isolated females were transferred to Wageningen University (The Netherlands) to be reared for low heterozygosity. The resulting offspring were reared in a single general population on irradiated *Ephestia kuehniella* (Zeller) (Lepidoptera: Pyralidae) eggs as factitious hosts under laboratory conditions in a climate chamber (20 ± 5˚C, RH 50 ± 5%, L:D=12:12 h). *Wolbachia* presence was determined following the PCR amplification protocol of Zhou *et al.* 1998 in a presence/absence assessment with known positive and negative control samples (Zhou et al., 1998). Natural

Wolbachia infections have previously been detected in Iranian populations of *T. brassicae* (Farrokhi et al., 2010), but none of the Eurasian populations have been known to support this symbiosis (Stouthamer, 1997; Stouthamer and Huigens, 2003). Sequence coverage was calculated using the previously established genome size estimate for *T. brassicae* of 246 Mbp (Johnston et al., 2004).

Isofemale line

Following confirmation of *Wolbachia* infection (Supplementary Figure 1) a single female from the general population was isolated (generation 0, G0), and given eggs *ad libitum*. In the resulting generation (G1), unmated females were isolated and reared with eggs *ad libitum*. Offspring of the initial isolations G0 and G1 were confirmed to be entirely female, suggesting thelytokous parthenogenetic reproduction. Combined with isolating single females, this maximizes genetic similarity of the following generation (G2) of these G1 females. One of these G2 strains, S301, was boosted for multiple generations over the period of one year. By the time of collection for sequencing, both the S301 and general population no longer harboured *Wolbachia* at detectable levels (Supplementary Figure 2).

gDNA Extraction

Three separate extractions were prepared in 1.5 mL safelock tubes with each several hundred *Trichogramma brassicae*. The tubes were frozen in liquid nitrogen with approximately six 1-mm glass beads and shaken for 30 s in a Silamat S6 shaker (Ivoclar Vivadent, Schaan, Liechtenstein). DNA was then extracted using the Qiagen MagAttract Kit (Qiagen, Hilden, Germany). Following an overnight lysis step with Buffer ATL and proteinase K at 56˚C, extraction was performed according to the MagAttract Kit protocol. Elutions were performed in two steps with Buffer AE (Tris-EDTA) each time (first 60 µL, then 40 µL), yielding 100 µL. The two extractions yielding the largest amount of DNA (5.49 µg and 8.24 µg) were combined for long-read sequencing, while the remaining extraction (1.67 µg) was used for short-read sequencing. DNA concentration was measured with an Invitrogen Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA) while

fragment length was confirmed on gel.

Library Preparation and Sequencing

Library preparation and sequencing was performed by Novogene Bioinformatics Technology Co., Ltd., (Beijing, China). For Illumina sequencing, gDNA was used to construct one paired-end (PE) library according to the standard protocol for Illumina with an average insert size of 150 bp and was sequenced using an Illumina HiSeq 2000 (Illumina, San Diego, USA). For Single Molecule Real Time (SMRT) sequencing, gDNA was selected for optimal size using a Blue Pippin size selection system (Sage Science, Beverley, USA) following a standard library preparation. The library was then sequenced on a PacBio Sequel (Pacific Biosciences, Menlo Park, USA) with 16 SMRT cells.

Assembly and decontamination

Prior to assembly, Illumina reads were assessed for quality using FASTQC (Andrews et al., 2015), then trimmed for quality in CLC Genomics Workbench 11 using default settings (Qiagen). Trimmed Illumina reads were paired for subsequent analysis.

In order to achieve the best possible assembly, three assembly pipelines were evaluated: one for PacBio-only reads and two hybrid assemblers. The PacBio-only were assembled with Canu (v1.6) with modifications based on PacBio Sequel reads, correctedErrorRate=0.085 corMhapSensitivity=normal alongside corMhapSensivity=normal (Koren et al., 2017). This is assembly version v1.0 in the subsequent discussion.

The first hybrid assembly pipeline using both long and short sequencing read sets was SPAdes (v3.11.1) (Bankevich et al., 2012). The SPAdes genome toolkit supports hybrid assemblies with the hybridSPAdes algorithm (Antipov et al., 2016). Three iterations of the SPAdes pipeline were run with varying k-mer sizes resulting in three different assembly versions: 21, 33, 55 (default, v2.1); k-mer sizes 21, 33, 55, 77 (v2.2); and a single k-mer size of 127 (v2.3).

The second hybrid assembly pipeline was DBG2OLC (Ye et al., 2016). The DBG2OLC pipeline can be readily tweaked with other programs depending on the job (Chakraborty et al., 2016). Following the DBG2OLC pipeline, de

Bruijn graph contigs were generated using SparseAssembler using default settings and setting the expected genome size to 750 Mbp to ensure a genome size output that is unrestricted (Ye et al., 2012). Contigs were transformed into read overlaps using DBG2OLC with settings suggested for large genomes and PacBio Sequel data (`k=17; AdaptiveTh=0.01; KmerCovTh=2; MinOverlap=20; RemoveChimera=1`), according to the DBG2OLC manual (https://github.com/yechengxi/DBG2OLC). This creates an assembly backbone of the best overlaps between the short-read de Bruijn contigs and the long reads. minimap2 (v2.9) and Racon (v1.0.2) were used for consensus calling remaining overlaps to the assembly backbone (Li, 2018; Vaser et al., 2017). The resulting consensus assembly was polished twice using the Illumina reads with Pilon (Walker et al., 2014). This final assembly is v3.0 in subsequent discussion.

The best of the five assemblies generated was determined on the basis of N50, genome size, and completeness (Table 1). Genome statistics such as N50, number of contig, and genome size were determined using QUAST (Gurevich et al., 2013). Assembly completeness was assessed using BUSCO (v3.0.2) with the `insect_odb9` ortholog set and the fly training parameter (Simão et al., 2015). Based on these characteristics, the decision was made to move forward with assembly v3.0, which was then decontaminated for microbial sequences using NCBI BLASTn v2.2.31+ against the NCBI nucleotide collection (nr).

Wolbachia contamination

Two contigs contained a large amount of *Wolbachia* content, with over 80% of the scaffold containing material with 75% or higher homology to *Wolbachia*. These contigs were assessed for homology against the NCBI nucleotide collection (nr) and removed from the assembly (see supplementary materials). Post-decontamination, the assembly is referred to as v3.5.

RNA extraction, library construction, and sequencing

T. brassicae wasps from the S301 line were collected for RNAseq for evidence-based annotation. Hundreds of adult individuals (male and female) were killed by freezing at -80˚C, then frozen in liquid nitrogen in a single 1.5 mL

safelock tube with approximately six 1-mm glass beads and shaken for 30 s in a Silamat S6 shaker (Ivoclar Vivadent). The RNeasy Blood and Tissue Kit (Qiagen) was used according to manufacturer's instructions, and final column elution was achieved using 60 µL sterilized H_2O. The sample was measured for quality and RNA quantity using an Invitrogen Qubit 2.0 fluorometer and the RNA BR Assay Kit (Thermo Fisher Scientific). The RNA sample was then processed by Novogene Bioinformatics Technology Co., Ltd., (Beijing, China) using poly(A) selection followed by cDNA synthesis with random hexamers and library construction with an insert size of 300 bp. Paired-end sequencing was performed on an Illumina HiSeq 4000 according to manufacturer's instruction. Quality filtering was applied to remove adapters, reads with more than 10% undetermined bases, and reads of low quality for more than 50% of the total bases (Qscore less than or equal to 5).

AB INITIO GENE FINDING, TRANSCRIPTOME ASSEMBLY, AND ANNOTATION

For the *ab initio* gene finding, a training set was established using the reference genome of *Drosophila melanogaster* (Meigen) (Diptera: Drosophilidae) (Genbank: GCA_000001215.4; Release 6 plus ISO1 MT) and the associated annotation (Adams et al., 2000; Dos Santos et al., 2015). The training parameters were used by GLIMMERHMM v3.0.1 for gene finding in the *T. brassicae* genome assembly v3.5 (Majoros et al., 2004). For homology-based gene prediction, GEMOMA v1.6 was used with the *D. melanogaster* reference genome alongside our RNAseq data as evidence for splice site prediction (Keilwagen et al., 2016). For evidence-based gene finding, the pooled RNAseq data was mapped to the to the *T. brassicae* genome separately with TOPHAT v2.0.14 with default settings (Trapnell et al., 2009). After mapping, CUFFLINKS v2.2.1 was used to assemble transcripts (Trapnell et al., 2010). CODINGQUARRY v1.2 was used for gene finding in the genome using the assembled transcripts, with the strandness setting set to 'unstranded' (Testa et al., 2015).

The tool EVIDENCEMODELER *(EVM)* v1.1.1 was used to combine the *ab initio*, homology-based, and evidence-based information, with evidence-based weighted 1, *ab initio* weighted 2, and homology-based weighted 3 (Haas et al. 2008). We annotated the predicted proteins with BLASTP v2.2.31+

on a custom database containing all SwissProt and Refseq genes of *D. melanogaster* (Acland et al., 2014; Boutet et al., 2008; Camacho et al., 2009), followed by an additional search in the NCBI non-redundant protein database (nr) to obtain additional homology data.

GO Term Analysis

A list of genes was constructed for Gene Ontology (GO) term classification by deduplicating the annotated proteins and removing the non-annotated proteins. These accession IDs were converted into UniProtKB accession IDs using the UniProt ID mapping feature and deduplicated a final time (Boutet et al., 2008). These UniProtKB accession IDs were in turn used with the DAVID 6.8 Functional Annotation Tool to assign GO terms to each accession ID with the *D. melanogaster* background and generate initial functional analyses (Huang et al., 2009a, 2009b) (see supplementary materials for DAVID input list).

Heterozygosity estimates

The heterozygosity of the S301 line was assessed using sequence reads and k-mer counting, and compared to the congeneric *Trichogramma pretiosum* (Riley) (Hymenoptera: Trichogrammatidae), for which sequence data exists for both a thelytokous (asexual) *Wolbachia*-infected strain as well as an inbred arrhenotokous (sexual) line (Lindsey et al., 2018). Using jellyfish v2.3.0 to count k-mers, the same trimmed and paired Illumina reads used for assembly were assessed using the default k-mer size of 21 (m=21), with results exported to a histogram (Marçais and Kingsford, 2011). This histogram file was then used with GenomeScope v1.0 to estimate heterozygosity of the reads based on a statistical model, where a Poisson distribution is expected for a homozygous sample while a bimodal distribution is expected for a homozygous distribution (Vurture et al., 2017). This genome profiling gives a reliable estimate for heterozygosity as well as estimates of repetitive content. The same jellyfish and GenomeScope analyses were performed on *T. pretiosum* short-read sequence data for the thelytokous strain (NCBI SRA database, SRR1191749) and the arrhenotokous line (SRR6447489), with adaptions for reported insert sizes (Lindsey et al., 2018).

Ortholog cluster analysis

The complete gene set of *T. brassicae* was compared to that of *T. pretiosum* (Lindsey et al., 2018), which was retrieved from the i5K Workspace (Poelchau et al., 2016). An ortholog cluster analysis was performed on both gene sets via OrthoVenn2 with the default settings of E-values of 1e-5 and an inflation value of 1.5 (Xu et al., 2019). For *T. brassicae* protein set, see supplementary materials.

Data Availability

All sequence data, including raw reads, assembly, and annotation, are available at the EMBL-EBI European Nucleotide Archive (ENA) under BioProject PRJEB35413. Additional data, such as gel images, the *Wolbachia* contaminated contigs, input gene list for DAVID, GenomeScope images, and complete protein set are available via the supplementary materials.

Results and Discussion

Sequencing, assembly, and decontamination

Sequencing of the Illumina 150 bp paired-end library yielded 80,489,816 reads. After quality filtering and trimming, 80,483,128 paired-end reads were retained. Sequencing the PacBio Sequel library yielded 2,500,204 subreads with an average length of 6377 bp. The genome size estimate for *T. brassicae* is 246 Mbp (Johnston et al., 2004) indicating that short-read coverage was 98x while long-read coverage was 64x, resulting in a total coverage of 162x. Three assembly pipelines were used, resulting in five potential assemblies where one, v3.0, was eventually selected for further use. Results of these assemblies are detailed in Table 1.

The first draft assembly generated with Canu with the altered settings for PacBio Sequel data resulted in an assembly of approximately 70 Mbp in size, drastically smaller than the 246 Mbp expected, and contained a total of 3,007 contigs with an N50 of 27,303. The longest contig was 126,800 bp in size.

The second assembly strategy relied on hybrid assembly pipelines, and *SPAdes* was used with the default k-mer settings, which resulted in an assembly of approximately 227 Mbp in size with an N50 of 36,870 and a *BUSCO*

completeness of 96.8%. Three different assembly runs were done with differing k-mer sizes: the default k-mer sizes of 21, 33, 55 (v2.1); default k-mer sizes plus 77 (v2.2); or the highest possible k-mer size of 127 (v2.3). Increasing the k-mer size only improved N50 scores to a point, along with decreasing the number of contigs, and stable BUSCO scores, however, the assembled genome size drops dramatically with the third attempt shrinking down to 211 Mbp. Based on BUSCO scores and N50 alone, the second SPAdes attempt, v2.2, would be the best of the three, though all three are similar in most measures.

Table 1. Statistics for five assemblies of *Trichogramma brassicae*. The first strategy was PacBio-only in Canu, while three hybrid assembly strategies were based on SPAdes and modulating k-mer sizes, and an additional hybrid assembly was based on an adapted DBG2OLC+Racon+Pilon protocol. BUSCO score is based on the insect_db09 dataset (Simão et al., 2015).

Assembler	**Version**	**Size** (bp)	**Contigs**	**Longest contig** (bp)	**N50** (bp)	**BUSCO** (Complete %)
Canu	v1.0	69,522,446	3,007	126,800	27,303	18.7
SPAdes (k=21, 33, 55)	v2.1	227,096,967	282,988	474,998	36,870	96.8
SPAdes (k=21,33, 55, 77)	v2.2	226,864,253	189,696	548,753	49,096	97.1
SPAdes (k=127)	v2.3	211,402,326	73,567	537,817	63,558	96.4
DBG2OLC+ Racon+Pilon	v3.0	235,413,774	1,572	2,953,580	556,663	95.5

The third assembly strategy used the DGB2OLC+Racon+Pilon pipeline, which resulted in assembly v3.0. Here, there is a large difference compared to the previous SPAdes assemblies. Particularly, the number of contigs is reduced dramatically from the 70,000 to 280,000 range of the *SPAdes* output down to a mere 1,572. Meanwhile, the assembled genome size is now 235 Mbp and with an N50 of 556,663 and a BUSCO score of 95.5%. The full completeness score for this assembly, using the 1658 BUSCO groups within the insect_od09 BUSCO set, returned 1531 (92.3%) complete and single-copy BUSCOs, 53 (3.2%) complete and duplicated BUSCOs, 22 (1.3%) fragmented BUSCOs, and 52 (3.2%) missing BUSCOs (Simão et al., 2015).

While the PacBio-only assembly in Canu could have been improved using

different settings or additional tools, we decided to focus on using the additional sequence information of the Illumina reads in the subsequent hybrid assembly strategies. The SPAdes assemblies (v2.1-3) were already decent but could have been further improved using *Pilon*, a tool that improves assemblies at the base pair level using high quality Illumina data. However, the v3.0 assembly was by far the best assembly based on assembled genome size, N50, and BUSCO scores and therefore we chose this strategy for our *T. brassicae* genome assembly.

Decontamination of this assembly (v3.0) resulted in the removal of two contigs as the homology analysis using BLASTn with the NCBI nr database indicated that both contigs were confirmed to be largely composed of *Wolbachia* genomic content. Contig "Backbone_1176" is 9,448 bp in length and two areas of the contig, representing over 80% of its length, showed high homology to *Wolbachia*. Similarly, contig "Backbone_1392" is 17,350 bp and three separate areas representing over 80% showed similar levels of homology to *Wolbachia*. After decontamination this final assembly (v3.5) was used for annotation.

Ab initio gene finding, transcriptome assembly, and annotation

In our RNA sequencing experiment, we generated 26,479,830 150bp paired-end cDNA reads. Filtering the reads for quality retained 99.3% of these reads to be used for evidence-based gene finding via transcriptome assembly.

The annotations from the evidence-based gene finding were used alongside homology-based findings and *ab initio* annotations in a weighted model, resulting in a complete annotation for the assembly. In 865 mRNA tracks, representing approximately 5.1% of the official gene set, a gene model could not be annotated via the SwissProt database, and these tracks are named "No_blast_hit." The majority of tracks are annotated with reference to SwissProt or GenBank accession number of the top BLASTp hit.

Transcriptome assembly and mapping resulted in 45,876,158 mapped transcripts (48,327,134 total). CodingQuarry predicted 45,454 evidence-based genes from these mapped transcripts, while *ab initio* gene finding using GlimmerHMM resulted in 16,877 genes and homology-based gene finding with

GeMoMa resulted in 6,675 genes. The final complete gene set was created using EVidenceModeler, where a weighted model using all three inputs resulted in a complete gene set of 16,905 genes.

GO term analysis

The complete gene set of 16,905 genes was deduplicated and genes with no correlating BLASTp hit were removed from this analysis. The remaining 9,373 genes were subjected to UniProtKB ID mapping, resulting in 8,247 genes with a matching ID after another round of deduplication (828 duplicates found). The remaining 755 accession IDs were not able to be matched, half of which are obsolete proteins within the UniParc database (377).

The DAVID Functional Annotation Tool used 6,585 genes for the analysis and showed that 80.8% (5,320) contribute to 530 biological processes, 77.5% (5,104) contribute to 115 different cellular component categories, and 74.2% (4,889) contribute to 93 molecular functions (genes can code to multiple GO terms). The remaining 1,662 genes are uncategorised.

Heterozygosity estimates

Using short-read data and k-mer counting, heterozygosity was estimated for our isofemale S301 line and compared to both a parthenogenesis inducing *Wolbachia*-infected strain and an arrhenotokous line of *T. pretiosum* (Lindsey et al., 2018). The average estimated heterozygosity for our S301 *T. brassicae* line is 0.0332% with approximately 0.608% repetitive content (for full details, see Table 2). This is similar to the thelytokous *T. pretiosum* line, which has a slightly lower estimated heterozygosity (0.0289%) and a lower amount of repetitive content (0.482%). Both have a very distinct Poisson distribution, indicating a low heterozygosity (Supplementary Figures 3 and 4). The arrhenotokous *T. pretiosum* showed a higher estimated heterozygosity (0.863%), a larger amount of repetitive content (2.64%), and a slightly bimodal distribution (Supplementary Figure 5).

The fact that both thelytokous *Trichogramma* species have a similar low level of heterozygosity when compared to the arrhenotokous *T. pretiosum* suggests that in both cases *Wolbachia* infection had a severe effect on

Table 2. Heterozygosity and repetitive content analysis of *Trichogramma brassicae* (thelytokous), *Trichogramma pretiosum* (thelytokous), and *T. pretiosum* (arrhenotokous) lines based on sequence data.

	Heterozygosity (%)	**Repetitive content** (%)	**Source of sequence data**
T. brassicae, thelytokous S301 line	0.0332	0.608	This chapter
T. pretiosum, thelytokous *Wolbachia* line	0.0289	0.482	Lindsey *et al*., 2018
T. pretiosum, arrhenotokous inbred line	0.863	2.64	Lindsey *et al*., 2018

genetic diversity. As the canonical mechanism of parthenogenesis-induction in other *Wolbachia* infected thelytokous *Trichogramma* species is gamete duplication (Pannebakker et al., 2004; Stouthamer and Kazmer, 1994), in which unfertilized eggs are diploidized and results in fully homozygous progeny in a single generation, the low genomic heterozygosity rate suggests a similar mechanism for *Wolbachia*-induced parthenogenesis in *T. brassicae*. However, the involvement of *Wolbachia* in causing all-female offspring in this *T. brassicae* strain and the presence and mechanisms of *Wolbachia* in other thelytokous *T. brassicae* strains (Farrokhi et al., 2010; Poorjavad et al., 2018, 2012) does require further investigation.

Ortholog cluster analysis

The complete gene set of *T. brassicae* was compared to that of *T. pretiosum* using OrthoVenn2 (full output in Table 3). Both species have a similar range of proteins (16,905 in *T. brassicae* and 13,200 in *T. pretiosum*) that form a similar number of clusters (6,537 in *T. brassicae* and 6,489 in *T. pretiosum*). The two species share 6,158 clusters (of 16,899 proteins), while *T. brassicae* has 379 unique clusters (1,726 proteins) and *T. pretiosum* has 331 unique clusters (1,005 proteins), as shown in Figure 1. These unique clusters account for approximately 5% of the entire cluster set for both species, and may both indicate true areas of differentiation, or result from differences in the annotation strategies. There is a similar amount of singleton clusters (proteins that do not cluster with others) in *T. brassicae* (5,291) and *T. pretiosum* (5,184). Both the unique clusters and the unique single-copy genes could be novel

proteins, regions of contamination, evidence of unique horizontal gene transfer, or pseudogenes. More investigation into these protein clusters in addition to a more comprehensive manual annotation should shed some light on the differences between these closely related yet geographically

Table 3. Output of OrthoVenn2 ortholog cluster analysis of *Trichogramma brassicae* and *Trichogramma pretiosum*.

Species	Proteins	Clusters	Singletons	Source of gene set
T. brassicae	16,905	6,537	5,291	This chapter (supplementary)
T. pretiosum	13,200	6,489	5,184	Lindsey *et al.*, 2018; Poelchau *et al.*, 2015

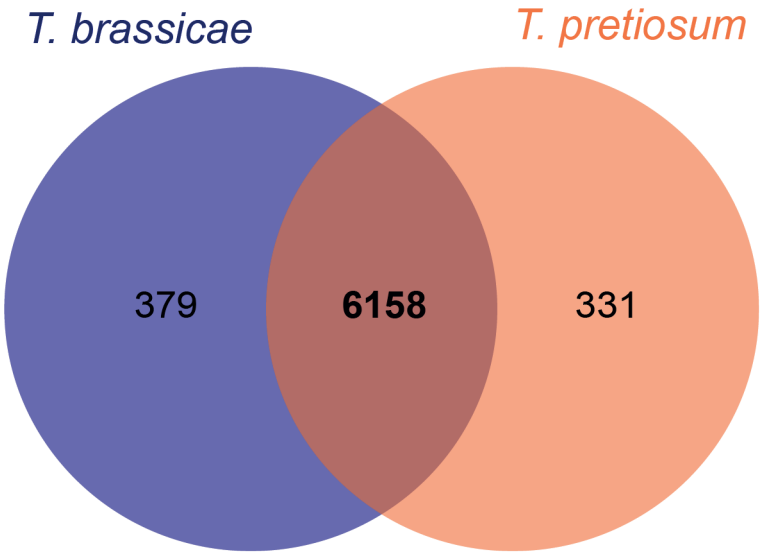


Figure 1. Ortholog clusters analysis between *Trichogramma brassicae* and *Trichogramma pretiosum* using OrthoVenn2 (Xu et al., 2019). The number of clusters shared between the two organisms are in bold.

distinct parasitoid wasps.

Conclusions and Perspectives

Here, we present the genome of biological control agent *Trichogramma brassicae*, a chalcidoid wasp used throughout the world for augmentative biological control as well as genetic and ecological research. This unique strain hosted a parthenogenesis-inducing *Wolbachia* infection and is the first European *Trichogramma* genome to be published, allowing for comparative analyses with other *Trichogramma* genomes, as we have shown. Our genomic data also illuminates the possible mechanism of parthenogenesis-induction by *Wolbachia* in this strain. Furthermore, the variety of genomic and transcriptomic data generated for this genome provide much-need resources to bring *T. brassicae* into the -omics era of biological research.

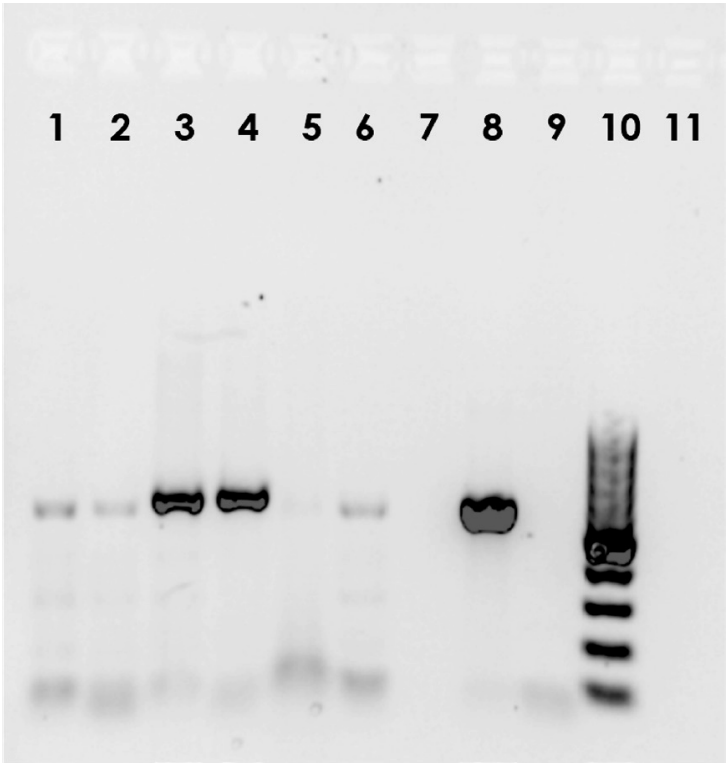
A hybrid approach was used, resulting in a highly contiguous assembly of 1,572 contigs and 16,905 genes based on *ab initio*, homology-based, and evidence-based annotation, for a total assembly size of 235 Mbp. Two scaffolds were identified that were of *Wolbachia* origin and removed. Ortholog cluster analysis showed 379 unique protein clusters containing 1,726 proteins. Future studies are needed to show whether these clusters are truly unique. This genome and annotation provides the basis for future, more in-depth comparative studies into the genetics, evolution, ecology, and biological control use of *Trichogramma* species.

Acknowledgements

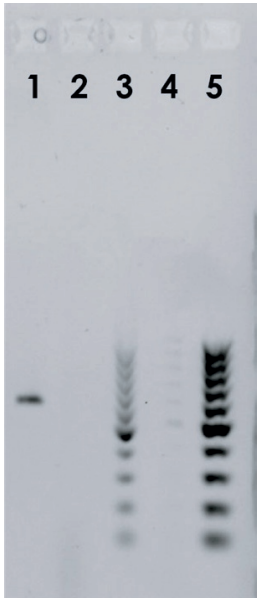
We would like to thank Bernd Wührer (AMW Nützlinge) for providing access to specimens; Gabrielle Bukovinszkine Kiss, José van de Belt, Frank Becker (Wageningen University), and Lorraine Latchoumane for assistance with rearing, DNA, and RNA extraction; Richard Stouthamer and Nina Fatouros for sharing their *Trichogramma* knowledge; and Sophie Chattington, Andra Thiel (University of Bremen), and Bas Zwaan (Wageningen University) for their assistance in this project. Decontamination and annotation analysis was performed by GenomeScan B.V. (Leiden, NL). This work has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 641456.

Supplementary Material

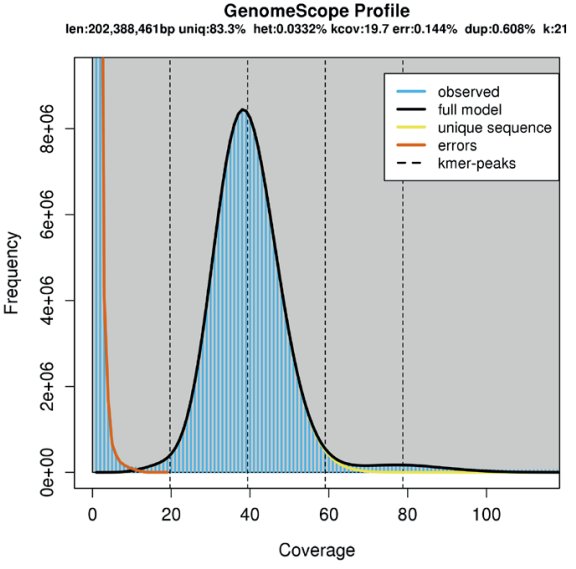
Additional supplementary material from this study (contaminated contigs, DAVID input data and initial analyses, protein set) are available on the DANS EASY Repository, DOI: 10.17026/dans-23w-a9tn



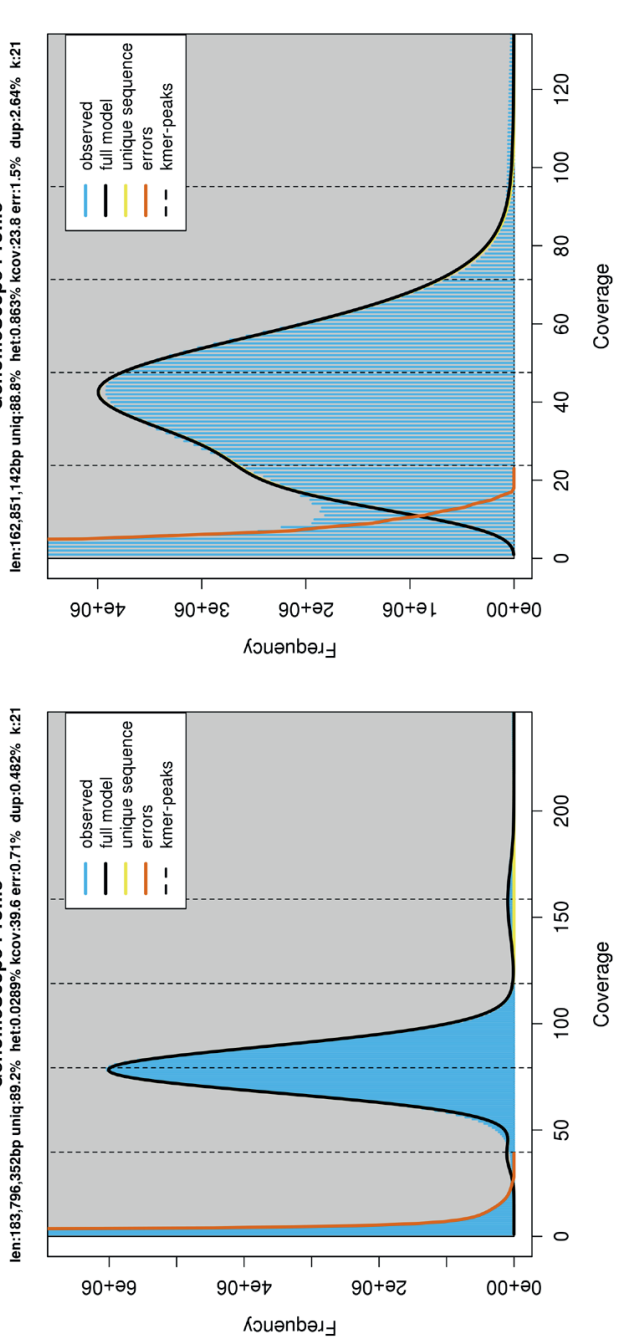
Supplementary Figure 1. Initial *Wolbachia* detection via PCR. *Wolbachia* detection was carried out on isolated DNA from individual males and female wasps, as well as population samples of several individuals, according to the PCR protocol of Zhou, Rousset and Neill, 1998. Lanes 1 and 2 are male *T. brassicae*, 3 and 4 are female *T. brassicae*, and 5 and 6 are pooled populations from the general population of *T. brassicae*. Lane 7 was a negative PCR control (sterilised water). Lane 8 is positive *Wolbachia* control *Muscidifurax uniraptor* (Hymenoptera: Pteromalidae) from a laboratory population confirmed to have thelytoky-inducing *Wolbachia* (Gottlieb and Zchori-Fein, 2001). Lane 9 is negative *Wolbachia* control *Nasonia vitripennis* from the Wolbachia cured AsymCx line (refer to Werren and Loehlin (2009) for line information). Lane 10 is a 100bp ladder (Invitrogen, Thermo Fisher, Waltham, Massachusetts, USA), while Lane 11 is empty.



Supplementary Figure 2. Loss of *Wolbachia* infection, detection via PCR. Detection of *Wolbachia* was carried out according to the PCR protocol of Zhou, Rousset and Neill (1998). Lane 1 is positive *Wolbachia* control *M. uniraptor* as in Supplementary Figure 1. Lane 2 is negative *Wolbachia* control *N. vitripennis* from the AsymCx as in Supplementary Figure 1. Lanes 3 and 5 are a 100bp ladder (Invitrogen). Lane 4 is a pooled sample of the general population, showing no products as expected for *Wolbachia*. Similar results were obtained for the S301 inbred line when it was previously used as positive control for other experiments, only to no longer display bands indicative of *Wolbachia* (KBF, unpublished results).

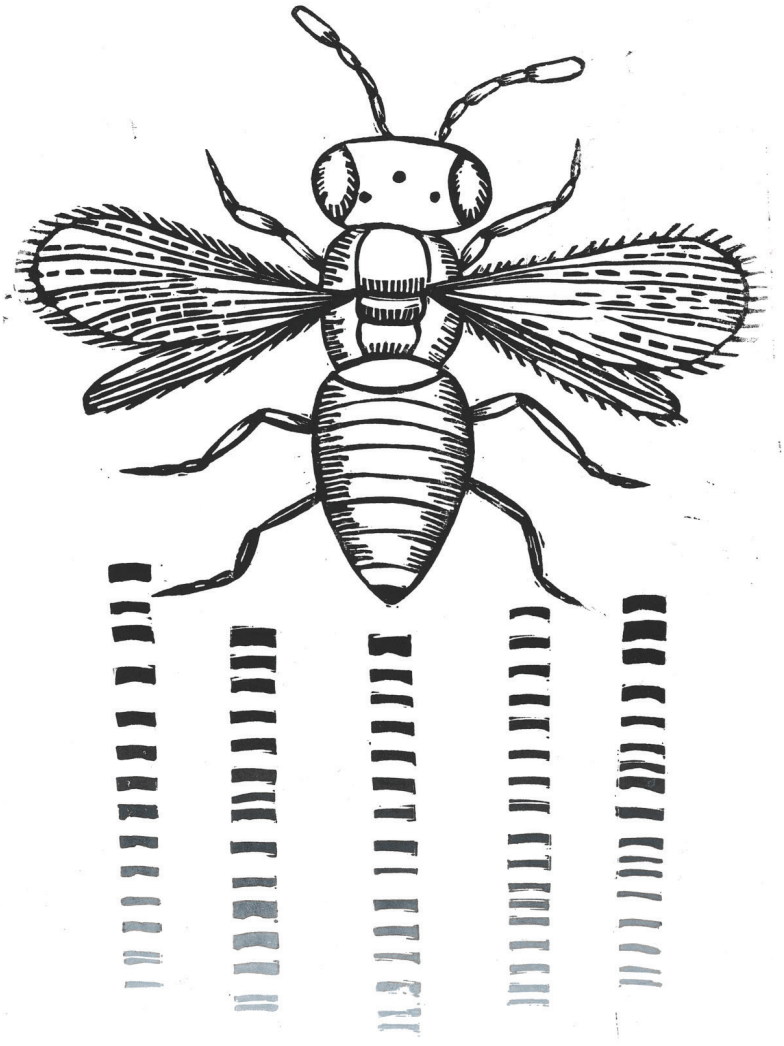


Supplementary Figure 3. GenomeScope results for thelytokous *Trichogramma brassicae* strain S301, based on reads generated in this chapter (BioProject PRJEB35413).



Supplementary Figure 4. GenomeScope results for thelytokous strain of *Trichogramma pretiosum*, based on reads from Lindsey *et al* 2018, NCBI SRA database SRR1191749 (Lindsey et al., 2018).

Supplementary Figure 5. GenomeScope results for arrhenotokous strain of *Trichogramma pretiosum*, based on reads from Lindsey *et al* 2018, NCBI SRA database SRR6447489 (Lindsey et al., 2018).



Going GAGA for Trichogramma

Chapter 6

The wasp and the butterfly: Population genetics and genomics of *Trichogramma evanescens* hint at role of phoresy in dispersal

Kim B. Ferguson[a], Bas J. Zwaan[a], Eveline C. Verhulst[b], and Bart A. Pannebakker[a]

[a] Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands
[b] Wageningen University, Laboratory of Entomology, Wageningen, The Netherlands

Abstract

Here we present the population genetics and genomics of the parasitoid wasp *Trichogramma evanescens* to better understand the population structure and genetic differentiation of native populations collected in Germany and The Netherlands. This was achieved using two levels of investigation: microsatellites and pooled DNA sequencing. Microsatellite mining of sequence data from a sister species produced 18,959 microsatellites, which was narrowed down to seven microsatellites that could be used on all populations (249 individuals). Alongside assembling a *de novo* genome for *T. evanescens* (190 Mbp), 590 million reads were generated over nine pools of 30 females for comparison via pooled DNA sequencing. There are clear indications of clusters (K=5) within the populations, meaning that the individuals do not constitute a single population, although these clusters do not correspond with the locations of capture. Our results from both microsatellite and pooled sequencing analyses indicated that there is clear differentiation between the populations, but only slight isolation-by-distance. We discuss the possible causes of these patterns of genetic differentiation, including the role of phoresy via the butterflies which eggs are parasitized by *Trichogramma*.

Introduction

In recent years, the use of next generation sequencing to improve several aspects of biological control has been proposed, such as production (rearing and storage), non-target testing, traits of interest, or monitoring (Lommen et al., 2017; Kruitwagen et al., 2018; Leung et al., 2019a). Within the context of monitoring, this not only includes assessing the genetic diversity of existing stock, but also determining the likelihood of invasion in existing, native, populations (Stouthamer and Nunney, 2014). This is more of a risk for open-field augmentative biological control, where commercially produced biological control agents, such as parasitoids, are reared in large volumes on factitious hosts, and then released in large numbers in order to inundate local pest populations (Stouthamer and Nunney, 2014).

At the moment of release, the most important question is if the released parasitoid is a successful biological control agent. Yet another question

remains, and that is what happens to existing conspecifics that are already established in the area (Baker et al., 2003; Hopper et al., 1993). Recent investigations into large-scale augmentative releases indicated that non-target effects, such as the introduction of deleterious alleles, are a risk, but require more study (Sethuraman et al., 2015). To that end, not only is it important to understand the population structure and genetic variation within commercial biological control populations, but also the population structure of wild populations within the area of release.

Trichogramma evanescens (Westwood) (Hymenoptera: Trichogrammatidae) is a tiny parasitoid wasp (~0.5 mm in length) that develops within the eggs of other insects (Smith, 1996). Found throughout Western Europe and Northern Africa, *T. evanescens* has a complicated taxonomic legacy, with at least 15 scientific synonyms (Polaszek, 2009). This is mainly due to the fact that the *Trichogramma* genus is largely cryptic, as it exhibits non-specific colouration (exception in some species, such as *T. cacoeciae*) and is minute in size. In spite of this, *T. evanescens* is used as biological control agent throughout Asia, Northern Africa, and Europe against lepidopteran pests (van Lenteren, 2012), and is well-studied for e.g. its ability to use herbivore-induced volatile chemicals as cues within a variety of contexts, including patch-finding (Cusumano et al., 2015; Fatouros et al., 2005a, 2007, 2014).

While the direct dispersal range of *T. evanescens* is relatively small (< 20m) and strongly affected by wind (Smith, 1996; Fournier and Boivin, 2000), there is also evidence of dispersal through phoresy. For this, the wasps are able to sense the anti-aphrodisiac that is passed from male to female butterflies post-mating, and ride on the body of the female butterfly (Fatouros et al., 2005b). This "hitch-hiking" allows for wasps to parasitize the freshly-laid butterfly eggs and also acts as a means for dispersal, as is well documented in *T. evanescens* both in the laboratory and in the field (Fatouros et al., 2005b; Fatouros and Huigens, 2012). Field sampling has demonstrated that at least five species of butterflies can act as phoretic hosts for *Trichogramma*: pierids *Pieris rapae*, *P. brassicae*, and *P. napi*, and nymphalids *Maniola jurtina* and *Vanessa cardui* (Fatouros and Huigens, 2012). In laboratory experiments, both *P. brassicae* and *P. rapae* have been observed to act as vehicles for *T.*

evanescens phoretic activity, albeit with varying success based on innate or learned behaviour (Huigens et al., 2010). Even though the direct dispersal of *T. evanescens* may be relatively contained, both the large cabbage white (*P. brassicae*) and the small cabbage white (*P. rapae*) are long-distance migratory butterflies with distributions throughout Europe (Tolman, 2008), and allow *T. evanescens* to travel substantial distance through the involuntary help of others.

In 2016 and 2017, field populations of *T. evanescens* were established from collections throughout central and northern Germany as part of a larger project to determine *Trichogramma* presence and distribution in the region. Given our knowledge on the two modes of dispersal, this begs the question: How distinct are these *T. evanescens* populations, and by extension all populations, if there are multiple modes of dispersal? And what does this mean for the likely spread of commercially reared and released *T. evanescens* throughout Europe?

Microsatellite markers are a well-established tool for population genetics that relies on tracking selectively neutral repetitive sequences, and have already been employed throughout biological control research (Stouthamer and Nunney, 2014; van Nouhuys, 2016), both in individual genotyping (Sanchez et al., 2012; Sethuraman et al., 2015; Streito et al., 2017), as well as in pooled analyses (Paspati et al., 2019). Pooled DNA sequencing (pool-seq), on the other hand, is a relatively new population genomic technique, heralded as the low-cost advancement in population analyses (Futschik and Schlötterer, 2010; Schlötterer et al., 2014). Performed in model species (Hoedjes et al., 2019) and non-model species (Nielsen et al., 2018) alike, pool-seq exchanges the individual-level precision of microsatellites for increased, genome-wide marker resolution. This may result in different calculations between the two, but results from microsatellites and pool-seq can be compared across certain parameters, such as F_{ST} (fixation index), depending on the method of calculation and methods of determining significance.

Without long-distance dispersal through phoresy, the expectation is that the (flight) biology of *T. evanescens* would result in extensive isolation-by-distance

of its populations. Distance between populations is thus a strong predictor of genetic differentiation, and populations from the same collection location are less differentiated than between collection locations. However, if phoresy is an extensively used mode of long-distance dispersal, geographical distance between populations becomes an unreliable indicator of genetic isolation. Within this context, we will be using both microsatellites and pool-seq, to determine the genetic structure and differentiation between and within populations of *T. evanescens*.

Methods

Wasp collection and rearing

The German wild-caught populations of *T. evanescens* were obtained through field collections in 2016 and 2017 (S. R. Chattington, pers. comm.). Collections took place in the period of June to September, on outdoor farms across Western and Central Germany, and were a combination of collecting *in situ* and using bait cards containing *Ephestia kuehniella* (Zeller) (Lepidoptera: Pyralidae) eggs (Figure 1, Table 1). The collection sites had no prior biological control releases, reducing the likelihood of collecting commercially released *Trichogramma* populations. The Dutch wild-caught population, BB, was collected *in situ* in 2018, in a residential area of the city of Wageningen. The final population in this study is an isofemale line that was derived from an *in situ* collection in 2006, in the Grebbedijk area just south of the city of Wageningen (Huigens et al., 2009). All populations have been reared in isolation with no crossing or reintroductions.

All German populations were initially reared in the Population and Evolutionary Ecology Group at the University of Bremen (DE) before being transferred to the Laboratory of Genetics at Wageningen University (NL) for the present study. All ten populations were maintained on irradiated *E. kuehniella* eggs under laboratory conditions in a climate chamber (20 ± 5˚C, RH 50 ± 5%, L:D=12:12 h).

Microsatellite mining and testing

Microsatellites were mined using a subset of the *Trichogramma brassicae* (Bezdenko) (Hymenoptera: Trichogrammatidae) trimmed reads resulting

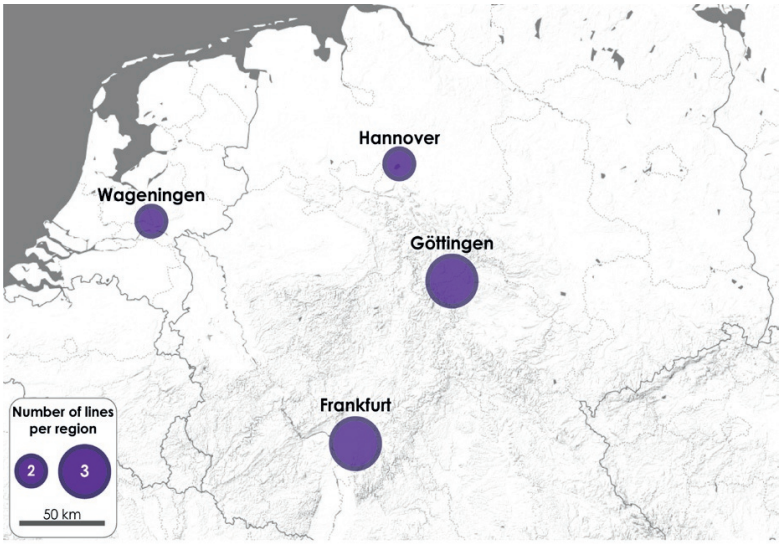


Figure 1. Map of Northern/Central Germany and The Netherlands, showing collection area for *T. evanescens*. Collection regions are indicated in circles, where size of circle is scaled to number of populations from each region. Refer to Table 1 for information on populations within each region and their year of collection.

Table 1. Information on all 10 populations used, number of individual *Trichogramma evanescens* females used in microsatellite analysis after quality filtering, and number of individuals used in pooled sequencing analysis, arranged according to region.

Collection Year	Region [a]	Population ID	Females in microsatellite analysis
2006	NL	GD011[b]	23
2018	NL	BB	46
2017	HA	KL90	20
2017	HA	KL147	15
2016	GO	BaWc	20
2017	GO	Ba154	21
2017	GO	MO8	17
2016	FR	GRRC16	24
2016	FR	GRWC17	38
2016	FR	GRWC27	25

[a] Region abbreviations are as follows: NL, The Netherlands; HA, Hannover; GO, Göttingen; FR, Frankfurt
[b] Indicates isogenic line

from whole genome Illumina sequencing (*cf* Chapter 5). As the original set was a rather large dataset (over 80 million paired end reads) that would have been computationally burdensome, therefore 800,000 paired reads from the high throughput sequencing set were randomly subsampled using CLC Genomics Workbench 11 (Qiagen, Hilden, Germany). These reads were mined for repeats using MsatCommander v.0.8.2 (Faircloth, 2008), where the number of repeats was set to a minimum of eight for di- and tri-nucleotides and a minimum of six for tetra- and penta-nucleotides. From our subset of reads, 18,959 microsatellites were identified. The MsatCommander program uses Primer3 v1.1.1 to design locus-specific, flanking primers (Rozen and Skaletsky, 1999). The settings used for Primer3 were: no perfect repeats; product size 75-140; primer size: min 16, opt 20, max 24 bases; primer melting temperature: 56 °C - 64 °C. This yielded 3,286 sets of primers. Poor primers were removed based on the following criteria: duplicate, positive for hairpins, high complementarity, and high likelihood of selfing, after which 1,196 primer sets remained.

Fifty-seven primer sets were selected for initial testing in single-reaction PCRs to verify efficacy and annealing temperature with pooled extractions of the isofemale *T. brassicae* line S301 (*cf* Chapter 5). Following the Chelex protocol from Fatouros and Huigens (2012), DNA isolations were prepared for i) *T. brassicae* S301 samples for initial primer set testing, ii) population-specific pooled samples for the 2016 German populations of *T. evanescens*, and iii) individual female wasp samples for the 2016 German populations of *T. evanescens* (Fatouros and Huigens, 2012). With this set-up, the accuracy of the primers as well as the size of the microsatellite were confirmed in step (i) where 33 primer sets of the original 57 successfully yielded product of expected size. Both the success of the primers as well as the presence of polymorphic microsatellites were confirmed in step (ii), while step (iii) screened out rare loci that could not be reliably reproduced in individual wasps. In both steps, 24 primer pairs were able to amplify microsatellites. Singleplex PCRs were performed with each primer set and 1 µL of the various DNA isolations to test primer sets.

PCRs were performed with GoTaq PCR kits according to manufacturer's instructions (Promega, Madison, Wisconsin, USA). The cycling program was 3

min at 95°C, followed by 35 cycles of 35 s at 95°C, 1 min at 53-57°C, and 45 s at 72°C, followed by 7 min at 72°C after the final cycle. Of the initial 57 primer sets, 33 successfully amplified with DNA of the isofemale *T. brassicae* line in step (i), which was reduced further to 24 primer sets that were successful on all four 2016 German populations in pooled (ii) and single-wasp (iii) isolations. Details on these 24 primer sets can be found in Supplementary Table 1.

Microsatellite multiplex PCR

Of these 24 primer sets, ten were selected for a variety of product sizes and were used in combinations of five in two rounds of multiplex reactions. We used expected product size to determine which fluorescent probes to use for the multiplex sets in order to reduce overlap of emission spectra in subsequent fluorescent fragment analysis. Three fluorescent labels were attached to a combination of primers for these microsatellites: 6FAM (emission range: 519 λ_{max}/nm), HEX (559 λ_{max}/nm), and ATTO-550 (576 λ_{max}/nm) (Biolegio, Nijmegen, The Netherlands). Details on these sets of primers can be found in Table 2.

DNA from over 264 individual wasps in various amounts of the ten different populations was extracted using the aforementioned Chelex protocol (Fatouros and Huigens, 2012). Multiplex reactions were carried out on both sets of primer combinations using the Qiagen Multiplex PCR kit (Qiagen), modified accordingly: For each well of a 96-well plate, the reaction contained 25 µl of master mix, 1 µL of template DNA (~30 ng), 5 µl of primer mix (each primer at 0.2 µg), and 19 µL of MQ. Cycling was run for 15 min at 95°C, 35 cycles of 35 s at 95°C, 1 minute at 54 °C and 45 s at 72°C, followed by 7 min at 72°C after the last cycle. PCR products were diluted 200 times and then analysed on an ABI-3730 Bioanalyzer (Agilent, Santa Clara, USA) alongside a GeneScan 500 LIZ ladder (GeneScan, Freiburg, Germany).

Initial microsatellite analysis

Microsatellite results were visualised and analysed using Geneious v10.3 and Geneious Prime v2019.2.1, where peaks were initially called using the automatic peak detection within Geneious v10.3 (http:// www.geneious.com, (Kearse et al., 2012)). Peaks were then confirmed manually using the guide from T. K. Luckau (OpenWetWare Contributers, 2013). Runs of poor quality following

these checks were removed from the analyses.

Results for two primer sets, TrBr0985 and TrBr1003, were discarded due to overlapping fragments. Prior to calculations, individuals containing more than five null alleles were removed from analysis to reduce the likelihood of interpreting errors. Additionally, it was observed that locus TrBr0021 was heavily influencing the error rate for calculations due to a large number of null alleles for this locus, specifically within the BB population. Therefore, locus TrBr0021 was removed for the entire dataset. The final dataset is thus composed of the information of seven microsatellite loci for 249 individual wasps, with full individual numbers for each population in Table 1.

This curated dataset was then assessed for genetic distance and other population genetics calculations using the Microsoft Excel add-on GenAlEx v6.51b2 (Peakall and Smouse, 2006, 2012). Three main tests were performed using GenAlEx: genetic and geographic matrices to perform Mantel's test, and two sets of G-statistics, including the estimate of differentiation (D_{EST}), and G_{ST}-adjusted F_{ST}. G-statistics are specific for microsatellites, enabling us to determine the significance of microsatellite-derived F_{ST}. The decision to use D_{EST} to determine differentiation was chosen in consultation with contemporary discourse on the topic (Kane, 2011), as well as the original publication (Jost, 2008). Essentially, D_{EST} is useful for establishing the relative degree of differentiation (D_{EST} = 1 at complete differentiation) of allele frequencies within a population, and is more useful for determining population differences between populations than other measures such as G'_{ST} (Jost, 2009). Finally, an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree was built using D_{EST} to visualise the amount of differentiation, performed in MEGA7 (Kumar et al., 2016).

Structure analysis

The final microsatellite data set was analysed using Structure v2.3.4 (Pritchard et al., 2000), which uses Bayesian inferences to determine the likely number of ancestral clusters (*K*) within population information (50,000 MCMC and 50,000 burn-in; admixture assumed; population information used; 15 iterations) for *K*=2 to *K*=11 (n+1). These initial Structure results aim to determine the optimal

K, which was performed using STRUCTURE HARVESTER web version v0.6.94 (Earl and vonHoldt, 2012), which relies on the Evanno method (Evanno et al., 2005). This web tool utilises the STRUCTURE output to calculate Δ*K*, a calculated likelihood for each *K* size used, where the highest Δ*K* is the optimal *K* size for the dataset. The optimal *K* size was then used to re-analyse STRUCTURE results, this time with the *K* size set to the optimal cluster, an MCMC and burn-in of both 200,000, and 100 iterations, following Hinimoto *et al*. (2011) (Hinomoto et al., 2011). These hundred iterations on the optimal *K* size were interpreted using CLUMPAK (Kopelman et al., 2015) to refine the results into major and minor STRUCTURE outputs.

POOLED WHOLE GENOME SEQUENCING

Pools of 30 females across two generations were collected after emergence from each of the ten populations. In one case, when females were unavailable, two males were used in place of one female to maintain N=30 diploid genomes in the pool. This is an established modification when pooling haplodiploids for genome-wide analyses (Gebiola et al., 2019), and specific counts are available in Table 5.

Pooled samples were placed in 1.5 mL safelock tubes with 5-8 one mm glass beads, frozen in liquid nitrogen, and shaken for 30 seconds in a Silamat S6 shaker (Ivoclar Vivadent, Schaan, Liechtenstein). DNA was then extracted using the Qiagen MagAttract Kit (Qiagen). Following an overnight lysis step with Buffer ATL and proteinase K at 56˚C, extraction was performed according to MagAttract Kit protocol. Elution was performed in two steps with 50 µL of Buffer AE (Tris-EDTA) each time, yielding genomic DNA in 100 µL as measured with an Invitrogen Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA). DNA quantity is presented in Supplementary Table 2.

Intended coverage for all but one pooled sample was 30X, to match the number of females (or female proxies) within each pooled sample. As POPOOLATION requires a genome for comparison in the sliding-window analysis, 100X coverage was intended for GD011 to aide in constructing a *de novo* genome for this purpose. Library preparation and sequencing was performed

by Novogene Bioinformatics Technology Co., Ltd., (Beijing, China). Here, DNA was used to construct ten paired-end (PE) libraries according to the standard protocol for Illumina with an average insert size of 150 bp and these libraries were sequenced using an Illumina Novaseq6000 (Illumina, San Diego, USA). Prior to assembly, demultiplexed reads were assessed for quality using FASTQC (Andrews et al., 2015), then trimmed for quality in CLC Genomics Workbench 11 using default settings and unpaired reads removed (Qiagen).

To check whether the GD011 reads were homozygous enough for an assembly, JELLYFISH v2.3.0 was used with results exported to a histogram (Marçais and Kingsford, 2011). These histogram files were then used with GENOMESCOPE v1.0 to estimate heterozygosity of the reads based on a statistical model, where a Poisson distribution is expected for a homozygous sample while a bimodal distribution is expected for a homozygous distribution (Vurture et al., 2017). This was performed on all sets of pooled reads to assess heterozygosity.

DE NOVO GENOME ASSEMBLY

A *de novo* genome was constructed from the GD011 reads using ABYSS v2.2.3, an assembler that is specialised for short reads (Jackman et al., 2017). Following the removal of unpaired reads, two k-mer sizes were tested for an ABYSS assembly. First, the default k-mer (96), which is also the largest k-mer size possible with ABYSS, followed by k-mer size 77, following our prior experience in the the *T. brassicae* assembly (cf Chapter 5). ABYSS creates both contig and scaffold assemblies, all four of which were assessed using QUAST for the best N50, least amount of pieces, and total size (Gurevich et al., 2013).

Assembly completeness was assessed using BUSCO (v3.0.2) with the `insect_odb9` ortholog set and the fly training parameter (Simão et al. 2015). Finally, to improve the sliding window analyses, scaffolds less than 4000 bp in length were removed, as calculating windows that are larger than the scaffold results in uninformative data and reducing the size of the sliding window, reducing the power of the analysis.

POOL-SEQ ANALYSIS ON WILD-CAUGHT POPULATIONS

Two pipelines were used to analyse our pooled samples: POPOOLATION and

PoPoolation2. PoPoolation performs sliding window analyses within populations, delivering diversity estimates such as nucleotide diversity (*Tajima's π*) and neutrality (*Tajima's D*) (Kofler et al., 2011a). PoPoolation2 performs sliding window analyses as well, but performs pair-wise comparisons across populations to calculate F_{ST} values for each window (Kofler et al., 2011b).

Using the PoPoolation v1.2.2 pipeline, the pooled reads of all the wild-caught populations were aligned to the trimmed GD011 genome and aligned reads were binned into windows using the bwa and samtools packages (Li et al., 2009). The "`basic-pipeline/mask-sam-indelregions.pl`" pipeline in PoPoolation was used to mask indel regions of the SAM file for each population, ensuring that indel regions are not calculated in subsequent analyses (Kofler et al., 2011a). Following the PoPoolation pipeline documentation, SAM files are then sorted and stored as BAM files using samtools before being converted into `.pileup` files. Pileup files and the scripts from the PoPoolation pipeline were used to produce sliding windows analyses for *Tajima's D* and *π* with a pool size of 60, windows of 2000 bp, a step size of 2000 bp, minimum count of 4, minimum coverage of 20, minimum quality of 20, and max coverage of 2%.

As for the PoPoolation2 v1.2.01 pipeline, the same masked indel BAM files from PoPoolation were used to create a single `.mpileup` file that was synced and analysed for F_{ST} values. This was achieved using windows of 2000bp and a step size of 2000bp, alongside a minimum count of 4, minimum coverage of 20, and suppressing non-informative values (i.e. windows of insufficient coverage for all populations were not included in analysis). Fisher's Exact Test was performed on these results using the `fisher-test.pl` pipeline from Popoolation2, and only statistically significant ($p<0.05$) F_{ST} values were used for analysis.

The default F_{ST} calculation in the `PoPoolation2` pipeline uses the formula listed in Hartl and Clark, which uses allele frequency (as opposed to allele count) (Hartl and Clark, 1997; Kofler et al., 2011b). This is essentially Nei's F_{ST} calculation, which is used by GenAlEx for the calculation of F_{ST}, allowing for correction under G_{ST} (Nei, 1977; Peakall and Smouse, 2006, 2012). Therefore, statistically-supported F_{ST} values from both microsatellite data and pool-seq

data are comparable under these conditions, as seen in other analyses (Kurland et al., 2019).

Results

All raw pooled sequence data and unannotated genome are all available on the EMBL-EBI European Nucleotide Archive (ENA) repository under BioProject PRJEB36429, while microsatellite data is available on the ENA under accessions LR758000-LR758009. Additional data and results can be found in the supplementary data when indicated.

Microsatellite analysis

Our final dataset for all populations includes 249 individuals and 7 microsatellite loci (Table 2), with the number of individuals used per population in Table 1. The lowest number of individuals was in population MO8 (17 females), while the highest was population BB (46 females). The full dataset as formatted for GenAlEx is available to download, more details in the supplementary material.

All seven microsatellite loci were polymorphic across all populations. The lowest number of polymorphic loci within populations was in GRRC16 (2 loci)

Table 2. Microsatellite loci details and their associated primers, along with fluorescent probe emission spectra. Presence in final microsatellite analysis is indicated. European Nucleotide Archive accession entries LR758000-LR758009.

Locus	Predicted size (bp)	Microsatellite	Emission spectra of fluorescent probe (λ_{max}/nm)	Multiplex round	Final analysis
TrBr1103	76	(AC)11	576	1	present
TrBr0405	88	(AG)8	519	1	present
TrBr0881	125	(AG)14	519	1	present
TrBr0543	131	(AG)15	559	1	present
TrBr0484	90	(AC)9	519	2	present
TrBr0620	121	(AC)8	519	2	present
TrBr0065	133	(ACG)9	559	2	present
TrBr0021	102	(AGC)10	576	1	removed
TrBr0985	80	(AG)8	576	2	removed
TrBr1003	102	(AC)23	576	2	removed

while the highest was in both GD011 and BB (6 loci). Our microsatellite analysis revealed population differentiation among the ten populations sampled, showing significant values of D_{EST} for all pairwise comparisons (Table 3, $p<0.05$ for all values). Similar results can be seen in the G_{ST} analysis for F_{ST} values (Table 4, 999 permutations, all values significant, p <0.05).

The values of D_{EST} indicate that GD011 is the most different from the other populations, with ranges between 0.724 (with BB) and 0.930 (KL90) (Table 3). The range for other populations was less extreme (0.123-0.657). Looking at the populations GRRC16, GRWC17, and GRWC27 that are all collected in the region of Frankfurt (all at the same farm, Großostheim, D_{EST} is relatively small (GRRC16:GRWC17= 0.308; GRWC16:GRWC27 = 0.162; and GRWC17:GRWC27 = 0.123), which would be expected of populations collected at the same location. The observed population differentiation is further reflected in a significant linear relationship with a slight positive slope (Mantel test, y=2.1983x + 10.766, R^2= 0.14, $p<0.05$) between the genetic distance and (log-transformed) geographic distance (Figure 2). The UPGMA tree built, based on the D_{EST} estimates (Figure 3a), indicates that GD011 indeed stands alone, as do Ba154 and KL147 to a lesser extent. With the exception of the GR populations from Frankfurt, there is no clear pattern based on the region of the collection site or the year of collection, the former of which is in agreement with the Mantel test results.

STRUCTURE ANALYSIS

From the initial STRUCTURE results the largest ΔK belonged to K = 5 (ΔK = 3.028) followed by K = 4 (ΔK = 2.412), and further STRUCTURE analyses were performed with K = 5. From these results, CLUMPAK determined a major STRUCTURE mode that occurred in 70 out of 100 simulations, arranged according to the UPGMA-D_{EST} tree (Figure 3b). According to geography, there is no latitudinal pattern. GD011 stands alone, as expected, while BB appears more like Ba154 and MO8. There are indications of population differences within the GR populations, as GRRC16 has very little of the white cluster compared to the other GR populations, GRWEC17 and GRWC27. Outputs for the minor cluster (16/100) are available in supplementary material (Supplementary Figure 1).

Table 3. Measure of differentiation, D_{EST}, using microsatellite data for 7 loci across 10 populations of *Trichogramma evanescens*.

	GD011	BB	KL90	KL147	BaWc	Ba154	MO8	GRRC16	GRWC17	GRWC27
GD011		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
BB	0.724		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
KL90	0.930	0.438		0.001	0.001	0.001	0.001	0.001	0.001	0.001
KL147	0.923	0.397	0.381		0.001	0.001	0.001	0.001	0.001	0.001
BaWc	0.879	0.415	0.356	0.478		0.001	0.001	0.001	0.001	0.001
Ba154	0.816	0.353	0.621	0.421	0.421		0.001	0.001	0.001	0.001
MO8	0.886	0.255	0.426	0.537	0.327	0.473		0.001	0.001	0.001
GRRC16	0.821	0.521	0.619	0.360	0.556	0.657	0.643		0.001	0.001
GRWC17	0.907	0.311	0.334	0.339	0.344	0.467	0.393	0.308		0.001
GRWC27	0.849	0.383	0.385	0.317	0.407	0.526	0.480	0.162	0.123	

D_{EST} values are below the diagonal, while p-values (G-statistics) are above the diagonal. All values for D_{EST} are statistically significant ($p<0.05$). Values were calculated in GenAlEx 6.5.51b2 (Peakall and Smouse, 2006, 2012).

Table 4. Fixation index, F_{ST}, corrected with G-statistics, using microsatellite data for 7 loci across 10 populations of *Trichogramma evanescens*.

	GD011	BB	KL90	KL147	BaWc	Ba154	MO8	GRRC16	GRWC17	GRWC27
GD011		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
BB	0.491		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
KL90	0.722	0.357		0.001	0.001	0.001	0.001	0.001	0.001	0.001
KL147	0.686	0.313	0.458		0.001	0.001	0.001	0.001	0.001	0.001
BaWc	0.618	0.284	0.384	0.424		0.001	0.001	0.001	0.001	0.001
Ba154	0.643	0.278	0.561	0.432	0.380		0.001	0.001	0.001	0.001
MO8	0.612	0.196	0.419	0.445	0.289	0.401		0.001	0.001	0.001
GRRC16	0.758	0.442	0.681	0.507	0.546	0.633	0.572		0.001	0.001
GRWC17	0.524	0.177	0.281	0.265	0.235	0.318	0.255	0.301		0.001
GRWC27	0.547	0.231	0.344	0.281	0.294	0.378	0.323	0.214	0.088	

F_{ST} values are below the diagonal, while p-values (G-statistics) are above the diagonal. All values for F_{ST} are statistically significant ($p<0.05$). Values were calculated in GenAlEx 6.5.51b2 (Peakall and Smouse, 2006, 2012).

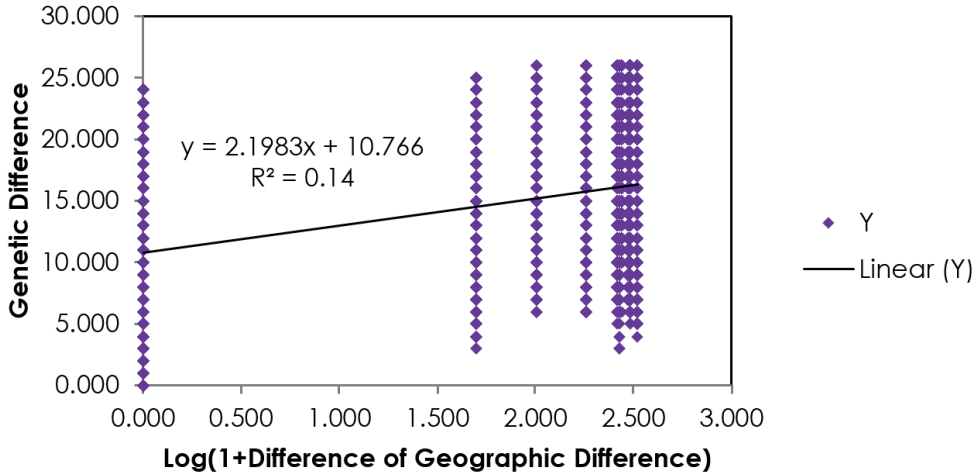


Figure 2. Results from Mantel test of microsatellite data and location. As geographic distance increases, the genetic difference increases slightly, indicating that isolation-by-distance is likely not at play between these populations at this geographic scale. Output from GenAlEx 6.5.51b2 (Peakall and Smouse, 2006, 2012).

Pool-seq and *de novo* genome assembly

Heterozygosity results showed that the reads for GD011 are very homozygous, with estimates ranging from 0.007% to 0.039% depending on the k-mer size and confidence intervals. Both assessments indicate a haploid genome size of approximately 195.5 Mbp. Heterozygosity estimates of the other pooled libraries based on sequence data are available in the supplementary materials (Supplementary Table 2). Using the short-read assembler ABySS, we assembled a *de novo* genome for *T. evanescens* based on pool-seq of GD011, which is 190,828,473 bp in length (190.8 Mbp) and comprised of 14,838 scaffolds with an N50 of 45,630. Compared to the haploid genome estimates from sequence data above, this means that our assembly is slightly smaller than expected, though this is often the case for assembled genomes, see also the *T. brassicae* genome (*cf* Chapter 5). Finally, the full completeness score for this assembly, using the 1,658 BUSCO groups within the `insect_od09` BUSCO set, returned 1557 (93.9%) complete and single-copy BUSCOs, 52 (3.1%) complete and duplicated BUSCOs, 14 (0.8%) fragmented BUSCOs, and 35 (2.2%) missing BUSCOs (Simão et al., 2015). Overall, both the assembly

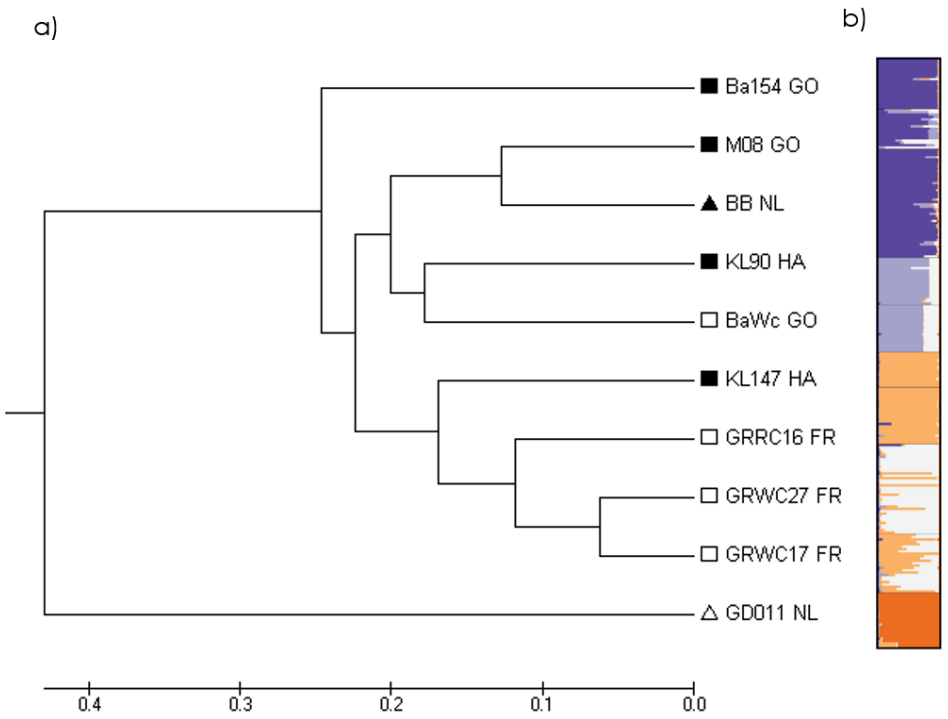


Figure 3. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram depicting differentiation between different populations of *T. evanescens* based on microsatellite data juxtaposed with STRUCTURE chart. a) UPGMA tree, name of population and region are preceded by symbol showing year of collection: empty triangle (2006), empty circle (2016), filled circle (2017), and filled triangle (2018). b) the STRUCTURE-derived clusters of likely ancestral populations for *T. evanescens* populations, data from 7 microsatellite loci, arranged to line up with populations in tree. Based on k = 5, 100 iterations were performed where the major cluster occurred 70 times. Colours represent the likely ancestral population (K),

Scale is to D_{EST} values used to construct tree. UPGMA tree constructed using values for D_{EST} (Jost, 2008) in GENEALEX (Peakall and Smouse, 2006, 2012) and drawn in MEGA7 (Kumar et al., 2016). Calculations occurred in STRUCTURE (Pritchard et al., 2000), followed by visualisation using CLUMPAK (Kopelman et al., 2015).

size and the high BUSCO scores indicate our *T. evanescens* genome to be reasonably complete and of the expected size.

PoPoolation and PoPoolation2

To accommodate the 2000 bp windows for PoPoolation analyses, the genome was trimmed of scaffolds that were less than 4000 bp in length. While this retained just 4.7% of the scaffolds (6,881 in total), it retained 94% of its size (~190.0 Mbp). This trimmed assembly was used in all pool-seq analysis going forward.

To analyse the pool-seq reads, both PoPoolation and PoPoolation2 were used. Both analyses used the same sets of reads, both with masked indel regions, and both used similar size of sliding windows and step (2000 bp). The number of reads used, subsequent masking, coverage calculations after masking the top 2% of reads, and the full outputs of all PoPoolation and PoPoolation2 analyses are available for download (see supplementary material for more information).

Reduced PoPoolation results can be seen in Table 5 as mean values over the entire genome. The coverage for all populations at 2% fell within a range of 41 (KL90) to 52 (BaWc), and this ranges between 1.36X and 1.73X coverage per individual, respectively. These results are genome-wide and can be used to compare the populations as a whole. Further visualisation of the results for *Tajima's D* for each population for the ten largest scaffolds are in Supplementary Figure 2. Here, we can see differences between the populations in specific regions, as well as certain windows within each population where the *Tajima's D* is lower than -2 or higher than 2. PoPoolation2 reports pairwise F_{ST} values between populations. Table 6 presents the mean pairwise significant F_{ST} (Fisher's Exact Test, $p<0.05$). These values can be directly compared to the microsatellite F_{ST} values, as both are calculated in the same manner, and have been statistically validated (albeit by different methods due to the different types of data). The F_{ST} values from the PoPoolation2 analysis are nearly always higher than the microsatellite F_{ST} values but show variation in the amount of increase (Figure 4). Only four comparisons – GRRC16 to BB, Ba154, and MO8, and between KL147 and MO8 show lower F_{ST} values

from the PoPoolation2 analysis compared to the microsatellite data, though in most cases by a fairly small amount.

In terms of a latitudinal gradient, the F_{ST} values do not increase with geographic distance, but show rather that some populations from the same location have a higher fixation than others from that location. For instance, the locations KL from Hannover and GR from Frankfurt are the furthest apart, and while GRRC16 and KL90 have the largest amount of fixation in this table (0.699), similar values are observed in comparison of GRRC16 and GRWC27 (0.579) from the same farm in Frankfurt, and GRWC27 and KL90 (0.57) that are the same distance apart as GRRC16 and KL90 (Figure 1). This supports that there is slight isolation-by-distance at play within these populations, however it is clear that the amount of fixation varies between populations.

Discussion

Here we studied ten populations of parasitoid *T. evanescens*. Eight were wild-caught German populations, collected throughout central Germany in 2016 (GRRC16, GRWC17, GRWC27, and BaWc), and 2017 (KL90, KL147, MO8, and Ba154). Additionally, we studied two Dutch populations, where one was collected in 2018 (BB), while the other is an isogenic population that was part of a 2006 collection (GD011). All populations have been maintained without introgression of new genetic material since their collection, though one (GD011) was intentionally bottlenecked to become an isofemale line in 2006. Our main question was to determine the genetic differentiation between the ten populations and determine if isolation-by-distance likely plays a role in the amount of differentiation observed.

An additional question posed in our work relates to the types of information and insights that can be gained from using different genetic and genomic tools. Microsatellites are an established tool for genotyping, though the number of microsatellites required to differentiate closely related populations remains a concern (Hale et al., 2012; van Nouhuys, 2016). Especially in a biological control context, microsatellites are useful for determining the diversity within a commercial line for rearing, or for tracking the introgression of commercial lines into wild populations (Leung et al., 2019a). And while

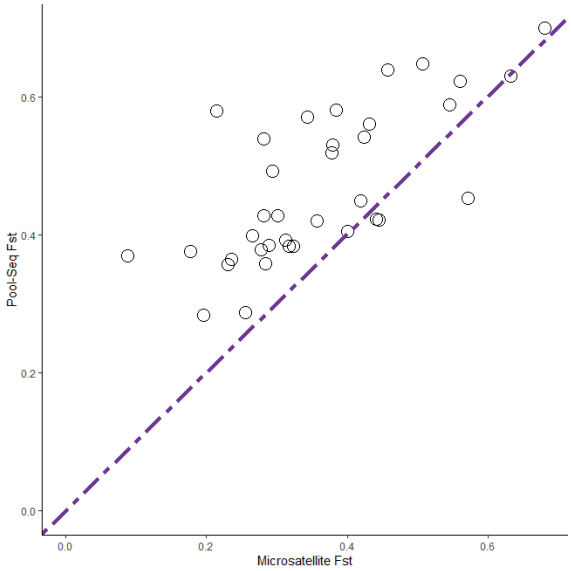


Figure 4. Comparison of microsatellite derived F_{ST} values (see Table 4) and pool-seq derived F_{ST} values (see Table 6). Purple line is identity.

microsatellites offer an individual-level of information but at low marker densities, pool-seq analysis increases the marker density but reduces the amount of individual-level information. This can increase the resolution of the analysis in some respects and may answer questions about the genetic diversity of a population without the cost and time associated with individual genotyping (both microsatellites and whole-genome genotyping).

Little support for isolation-by-distance

From our Mantel's test, results indicated that while there is a slight positive slope, the genetic variation explained by geographic distance was only very small as shown in the low R^2, thereby offering little support for isolation-by-distance. This is corroborated by the D_{EST} results, where there is no clear pattern of differentiation based on location. Indeed, some populations show more differentiation from other populations collected at the same farm as opposed to populations collected further away in different regions. The Structure analysis also shows this by indicating five ancestral clusters not based on location. So while there is little support for isolation-by-distance, there is also

Table 5. Sliding window variance analysis results for pooled sequencing across 9 wild-caught populations of *Trichogramma evanescens* with number of sufficiently covered windows, number of SNPs in sufficiently covered windows, average *Tajima's D*, and average *Tajima's π*.

Population	Individual wasps (female/male)	# of windows	# of SNPs	*Tajima's D*	*Tajima's π*
BB	30/0	42,492	385,020	-0.0533	0.0032
KL90	30/0	38,125	160,458	-0.2505	0.0017
KL147	30/0	40,094	177,941	-0.1716	0.0017
BaWc	27/6	41,923	243,690	-0.1227	0.0021
Ba154	30/0	41,976	258,860	-0.3977	0.0022
MO8	30/0	41,940	326,251	-0.0222	0.0028
GRRC16	30/0	40,892	162,838	-0.2512	0.0016
GRWC17	30/0	40,252	311,331	-0.0584	0.0028
GRWC27	30/0	38,054	306,780	-0.2818	0.003

Results were calculated using the PoPoolation v1.2.2 (Kofler et al., 2011a) pipeline for a pool size of 60, window and step size of 2000 bp, with indel regions and the top 2% of coverage masked. All populations were sequenced at 30X coverage.

Table 6. Fixation index, F_{ST}, in genome-wide sliding windows analysis of 9 populations of wild-caught *Trichogramma evanescens*.

	BB	KL90	KL147	BaWc	Ba154	MO8	GRRC16	GRWC17	GRWC27
BB									
KL90	0.420								
KL147	0.392	0.639							
BaWc	0.358	0.581	0.541						
Ba154	0.379	0.623	0.561	0.530					
MO8	0.283	0.449	0.422	0.385	0.405				
GRRC16	0.423	0.699	0.648	0.588	0.630	0.453			
GRWC17	0.376	0.428	0.399	0.364	0.384	0.287	0.428		
GRWC27	0.357	0.570	0.539	0.492	0.519	0.384	0.579	0.369	

Mean F_{ST} values are below the diagonal. Values are the average for all F_{ST} values that are statistically significant (Fisher's Exact Test, $p<0.05$). Values were calculated using PoPoolation2 v1.2.01 (Kofler et al., 2011b), with sliding window and step sizes of 2000 bp, masking indel regions, minimum count of 4, minimum coverage of 20, maximum of 2% coverage masked, and suppressing non-informative values.

little evidence of one large unstructured European-wide population. Finally, turning to the pool-seq results, a slight indication of isolation-by-distance can be seen in the pairwise F_{ST} values, where similar to the D_{EST}, there are higher levels of fixation between populations collected at the same farm or region than between different regions (Figure 1).

These results indicate that the influence of geography on the population structure of *T. evanescens* is fairly low and while this may be due to the restricted scale of our analysis, it is equally likely that large-scale phoresy on butterflies may be affecting the relationship between relatedness and proximity. Most studies and reviews indicate a small, less than 20m, and largely downwind dispersal range for *T. evanescens* (Fournier and Boivin, 2000; Smith, 1996). However, previous investigations reveal that phoresy may contribute to a larger dispersal range than thought possible (Fatouros and Huigens, 2012; Huigens et al., 2009). With the exception of *T. evanescens*, dispersal through phoresy within *Trichogramma* species is often alluded to but never fully investigated (Chapman et al., 2009; Smith, 1996). Indeed, when assessed for non-target effects for biological control risk assessment, the only dispersal modes considered for *T. brassicae* is direct or wind-assisted dispersal (Babendreier et al., 2003).

While we did not include *T. evanescens* strains reared for biological control in our study, it can be interesting to compare our study to population genetic studies of other native insects released for biological control. Mirid *Macrolophus pygmaeus* (Rambur) (Hemiptera: Miridae) is a biological control agent of many greenhouse pests throughout temperate and Mediterranean Europe (Vandekerkhove et al., 2011), and the population structure of similar scales has been assessed in two separate studies. In 2012, Sanchez *et al.* found isolation-by-distance between *M. pygmaeus* populations across Europe using nine microsatellite loci across 15 European sample sites (Sanchez et al., 2012). When looking at a similar geographic range (13 sample sites across Europe with seven microsatellite loci), Streito *et al*. found very little evidence for isolation-by-distance (Streito et al., 2017). With a Palearctic range alongside European-wide biological control releases, determining the difference between wild *M. pygmaeus* and released *M. pygmaeus* appears

to be the point of contention here. As Sanchez *et al.* used one commercial strain for comparison, while Streito *et al.* used four (some had shared initial sources), this suggests that some populations that Sanchez *et al.* had assumed to be wild could have been derived from commercial ancestors, leading to population differentiation (Sanchez et al., 2012; Streito et al., 2017). Without knowing the population structure of *M. pygmaeus* before human-aided release (*M. pygmaeus* has been available in Europe for greenhouse release since 1994, (van Lenteren, 2012)), it is difficult to determine the admixture of commercial and wild populations, and the deleterious effects it may have. For *M. pygmaeus*, this is largely in the context of greenhouse released biological control agents that escape the greenhouse. For *T. evanescens*, releases are open-air. Although in some cases of inundative biological control release, establishment in the surrounding environment is preferable in order to facilitate multiple generations (Rahimi-Kaldeh et al., 2018), in other cases it is used in risk assessments to determine risk of uncontrolled spread (Kuske et al., 2004).

Another difference between *M. pygmaeus* and *T. evanescens* may be related to the rearing practices for predators and parasitoids in general. Predators are often reared in genetic isolation for long periods of time, reducing genetic variation (Paspati et al., 2019; Rasmussen et al., 2018). For parasitoids, different methods are employed to avoid this expected reduction, which can include consistent genetic introgression with fresh field material, or setting up isofemale lines (Guzmán-Larralde et al., 2014). These methods would result in drastically different amounts of genetic variation. Without knowing how commercial lines are produced and maintained, or indeed how many different biological control operators are working in the field, it is difficult to determine whether released *T. evanescens* are more or less genetically distinct from the wild-caught populations reported here.

We posit that phoresy is an instrumental method of dispersal of *T. evanescens*, however, our study did not include commercial lines and our intention of capturing only "wild" wasps by visiting farms without history of biological control use may have been insufficient. It may be that due to years of releases throughout Europe, biological control lines have already mixed with

wild populations more readily and widely than previously expected, partially due to phoresy-aided dispersal. Therefore, having access to commercial lines from multiple operators that operate in central and northern Germany would be necessary to determine whether commercial and wild populations have introgressed, and what the potential outcome would be in terms of genetic differentiation and our isolation-by-distance calculations.

Genetic differentiation

The differences between the microsatellite-derived F_{ST} and pool-seq F_{ST}, as compared between Tables 4 and 6, displayed in Figure 4, go against previous comparison studies. For instance, in Kurland *et al.* the results show that F_{ST} values from pool-seq were lower than microsatellite-derived F_{ST} values (Kurland et al., 2019). This may be due to the polymorphic nature of our microsatellites, or it could be that we had fewer individuals genotyped by microsatellite than in most pooled samples, as was the case in the other study (Kurland et al., 2019). Kurland *et al.* looked at wild-caught and introduced stocks of salmon and trout, and found that while some F_{ST} values from pool-seq were lower than their genotyping data, this only occurred in the introduced populations where genetic variation was expected to be lower. While there may be a biological explanation for the differences observed in our study, it may also be technical or related to the bioinformatic processes. Therefore, more subsampling of our pool-seq data could be used to get a clearer picture on the differences between microsatellite and pooled F_{ST} values, and on the differences between studies.

Finally, large scale comparisons between populations can also give insight into the evolutionary forces acting upon the genome. The genome-wide *Tajima's D* for the populations fall within the range of less than 2 but more than -2, indicating that no strong selection has occurred (Reddiex et al., 2018; Tajima, 1989). As seen in Supplementary Figure 2, there are regions in the ten largest scaffolds where certain populations have windows that are above or below the *Tajima's D* thresholds, which may be interesting to look into further, especially if gene annotation data becomes available. Recent pool-seq projects in other species offer intriguing directions for analysis with regards to geographic comparisons and population differentiation (Kapun

et al., 2018).

Use of existing genomic data *versus* purpose-driven data generation

Here we used reads from the related species *T. brassicae* to generate microsatellite markers for use in analysis. We were able to mine a large number of microsatellites (18,959) for which we could produce 1,196 primer sets, of which we tested 57, i.e. the majority remains to be tested and could be of interest to other researchers.

All of the loci used in the final analysis were polymorphic in all populations, and may be useful for future genotyping of *T. evanescens* populations from field collections in this region. Therefore, our results support that the utility of whole genome sequence data often extends beyond the original purpose or even, as in this case, the original species.

Future directions

Our initial analysis is currently descriptive, but further investigation will involve isolating population-specific SNPs and extending the population genetic and genomic analysis, for instance through other parts of the PoPoolation2 package, as well as using the annotation from *T. brassicae* to annotate this *T. evanescens* genome. The microsatellite and pool-seq data that we have generated here will be analysed further alongside the more location-specific context (S. R. Chattington, pers. comm.). For example, a haplotype analysis alongside a measure of migration, such as G'_{ST}, would be beneficial to try to determine the possible scenarios for population spread. Similar analyses performed on other insects offer excellent possibilities in this regard (Valentin et al., 2017; Xun et al., 2016).

There are numerous factors to consider when determining genetic variation of wild-caught parasitoid populations, such as the presence of endosymbionts that can influence the genetic isolation between populations. One such endosymbiont is the well-studied and well-documented facultative endosymbiont *Wolbachia* (Werren et al., 2008), which has been detected in some Portuguese *T. evanescens* populations (Gonçalves et al., 2006). The *Wolbachia* infection status of our German and Dutch *T. evanescens* samples

is currently being determined and will be included in the eventual publication for this work, as it may be altering our interpretation of genetic differentiation between these populations and within regions.

Additionally, we are not able to uncouple the likelihood of initial genetic bottlenecks that occurred in the collection of our populations, as well as any other bottlenecks that occurred due to transfer or rearing fluctuations. This has been observed in other biological control populations (Ciosi et al., 2014; Gebiola et al., 2019; Paspati et al., 2019), and research lines are no less exposed to this risk (Szűcs et al., 2019). Furthermore, the combination of haplodiploidy and a potentially small number of foundresses may have led to increased bottlenecks within the populations, the latter of which is difficult to account for as we do not know the number of foundresses in *in situ* collections. However, additional analysis of our microsatellite data might allow us to determine the number of foundresses for each population (Grillenberger et al., 2008). Overall, additional field collections would be required to complete the picture, including netting live individuals that are studied genetically in addition to setting up populations, though this would require considerable resources to achieve.

Conclusions

Using both microsatellites and pool-seq applications, we determined the population structure and genetic differentiation of the German and Dutch *T. evanescens* populations presented here, finding little evidence of isolation-by-distance, but only very little of the genetic variation was explained by geographic distance. This supports the inclusion of phoresy as an important mode of dispersal in *T. evanescens*, though it is likely that the human-induced spread of *T. evanescens* for biological control purposes has also aided this. Additionally, differences in genetic differentiation between regions and collection locations suggest that there is clearly no German or European-wide *T. evanescens* population or ancestral cluster.

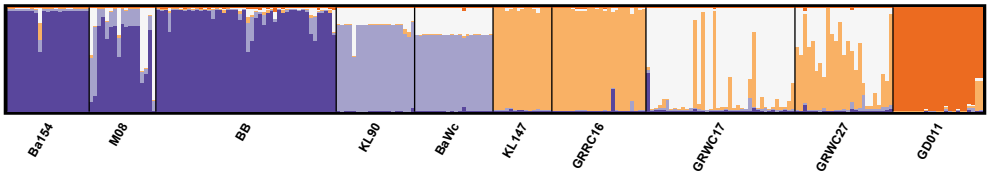
Acknowledgements

The authors would like to first and foremost acknowledge Sophie Chattington for the German populations, all collection and location information, and valuable feedback during this project. Thanks to Nina Fatouros for her *Trichogramma* expertise and knowledge, as well as for providing the GD011 line (2006-2019), may it be remembered fondly for this, its last experiment. Thanks to Gabrielle Bukovinszkine Kiss for rearing assistance throughout the years and to Joost van den Heuvel for his advice on pooled sequencing. HUGE thank you to our lovely students who worked on this project: PCR-Rockstar, Lorraine Latchoumane; MVP of Microsats, Susanne Borgman; and BB Extraordinaire, Joris Santegoets. Final thanks to Richard Stouthamer and Eric Wajnberg for sharing their *Trichogramma* expertise early on in this project. This work has received funding from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 641456.

Supplementary materials

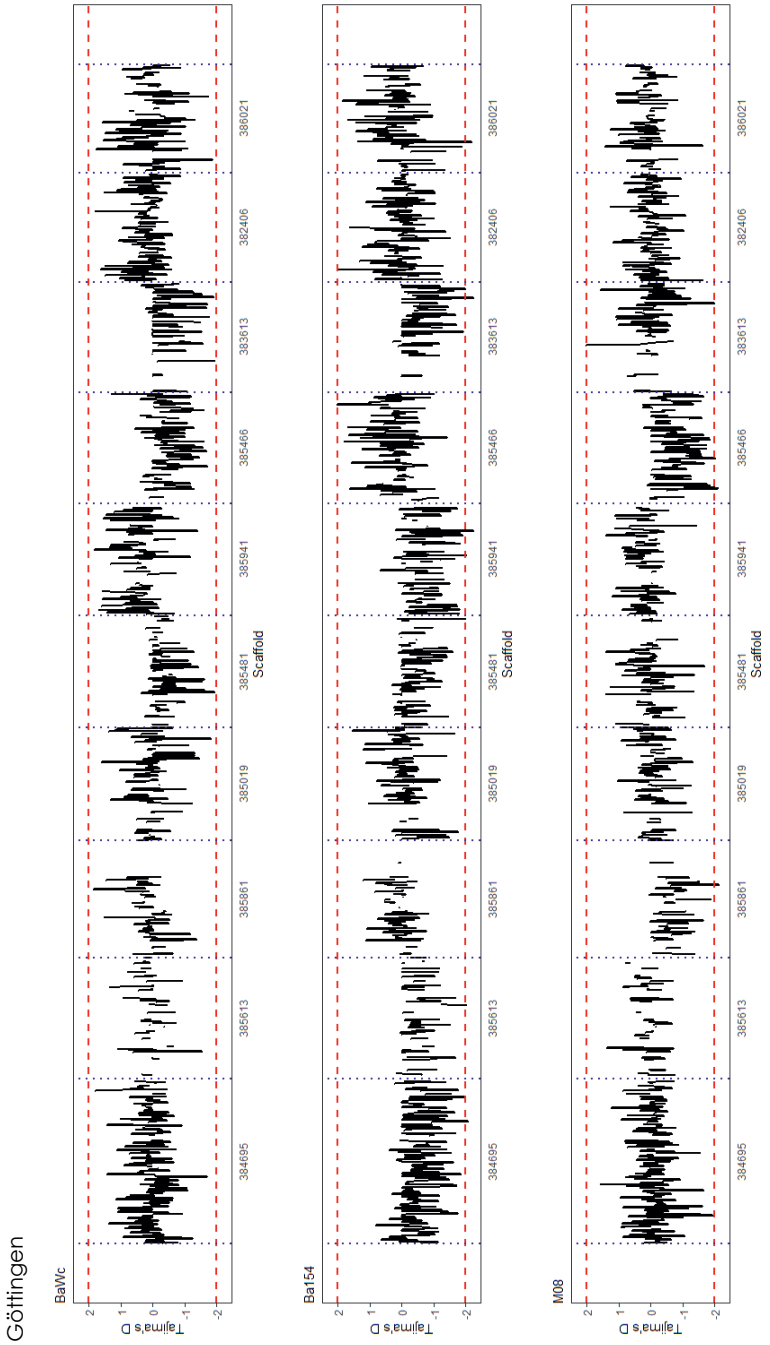
Full microsatellite results, formatted for GenAlEx, available at figshare DOI: 10.6084/m9.figshare.11340104

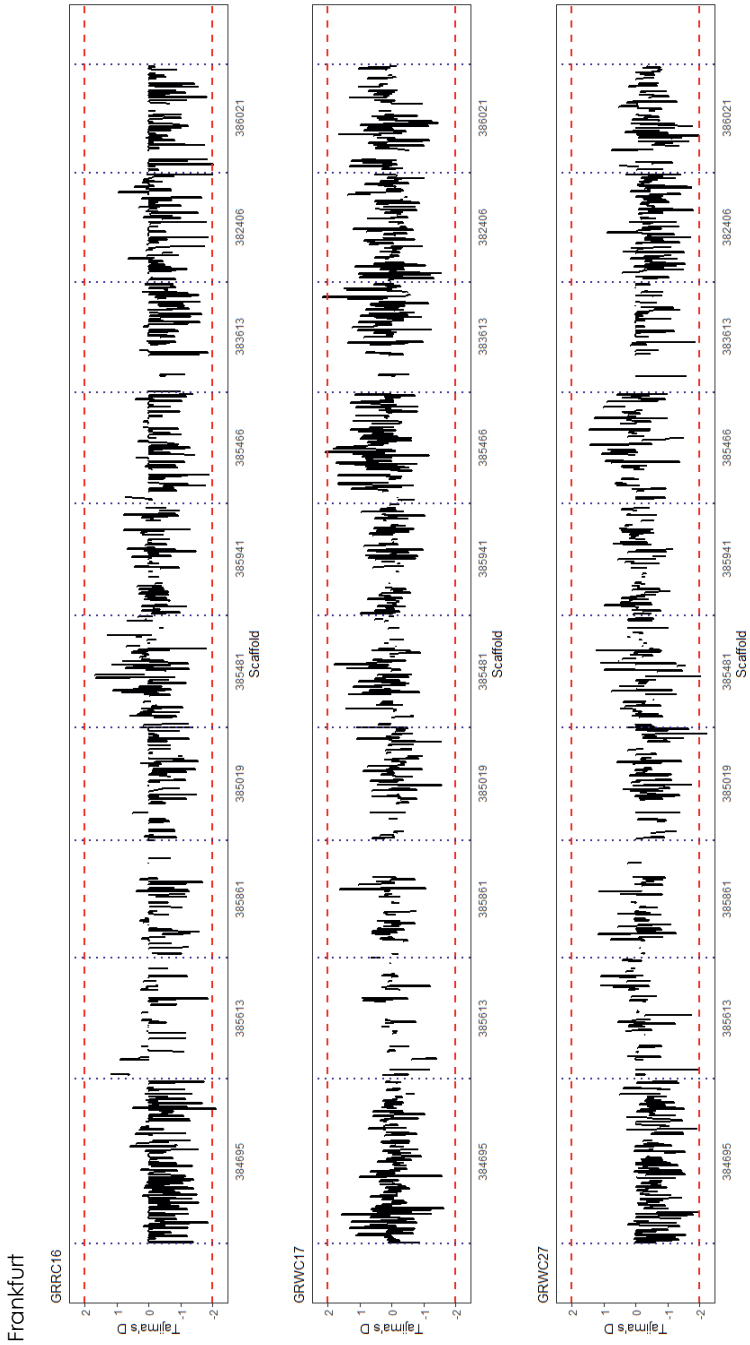
Full PoPoolation and PoPoolation2 results, available at figshare DOI: 10.6084/m9.figshare.11340152



Supplementary Figure 1. Minor cluster 1 for *k*=5, 16/100 iterations, arranged to match the UPGMA tree based on D_{EST} found in Figure 3. Colours represent the likely ancestral population (*K*), names refer along bottom refer to population. Width of column for each population relates to number of individuals genotypes. Calculations occurred in Structure (Pritchard et al., 2000), followed by visualisation using CLUMPAK (Kopelman et al., 2015).

Supplementary Figure 2. Genetic diversity of nine wild-caught populations of *Trichogramma evanescens* according to *Tajima's D*, arranged according to region. Red dashed lines indicate the range of *Tajima's D* values between -2 and 2. Ten largest scaffolds, ordered according to size, demarcated by blue dotted lines, scaffold name beneath on the x-axis. Continues over 3 pages.


Göttingen
BaWc
Ba154
M08
Tajima's D
Scaffold
384695
385613
385861
385019
385481
385941
385466
383613
382406
386021


Frankfurt
GRRC16
GRWC17
GRWC27
Tajima's D
Scaffold
384695
385613
385861
385019
385481
385941
385466
383613
382406
386021

Supplementary Table 1. 24 sets of microsatellite primers that successfully amplified on individual *Trichogramma evanescens* wasps

Primer Set	Microsatellite	Forward primer	Reverse primer	Expected product size
TrBr0021	$(AGC)^{10}$	AGGGTACACAGTCCATACAGC	GCTCGTTTGACAAGGTCCTC	102
TrBr0065	$(ACG)^{9}$	AAGCCAGAGAAACGACGATG	GGTTGGCCTATACGTTGTGG	133
TrBr0069	$(AC)^{9}$	ACGGAGCTTTCACTGATTCG	TGACCCAAGGAATTTCAAGTCG	114
TrBr0079	$(AG)^{11}$	TCCAGCAGAGAGGTAAGTGC	TGTGACGTTTCGACCTTGAC	101
TrBr0090	$(ACCT)^{7}$	TGGAAATGCGGTGAAACAGTC	TCACGGTTAGGTCGGTTGG	78
TrBr0204	$(AG)^{14}$	CGAGCTATTCTAACGAGCGC	ATAGTTCCAGGTCGTTTATCCC	100
TrBr0232	$(AG)^{13}$	GACGACGTGTATAGCAGTGAG	CTTTCCACCTGCCTCGTTTC	84
TrBr0340	$(AG)^{8}$	GACGCTACATTATAATTGCCGC	GACTCGGACTTGTTGCCTTC	123
TrBr0405	$(AG)^{8}$	GACGTACAGTACACCGTTGG	TGTTGCTGCACCATCATCAG	88
TrBr0484	$(AC)^{9}$	ACATGTATGCGATCCCTCCC	ATCGACTCCACTCTCGTTTG	90
TrBr0504	$(AC)^{10}$	AGCCCAACCGAATGATGAAG	ACGATGATTGTTCCTTGCGC	118
TrBr0543	$(AG)^{15}$	TCGACCTTCCCGATGCTG	GTCAACGCATATTGTACTCTCG	131
TrBr0620	$(AC)^{8}$	GCTGAGCTGAGAAACTCCAG	TGCAAAGACTCGGGTATAAGG	121
TrBr0639	$(AG)^{11}$	TCGCCAGTTTCCATTTCTCTC	GCAAGTTTGTCCTCTATATGCG	86
TrBr0734	$(AG)^{10}$	GCAAAGGATCTGTACCACCC	GCCGGTGTAGACTCCTCTC	79
TrBr0749	$(AC)^{9}$	AGTCTCCTGCATACAAACGC	GTTCTAAGCCTGCAAACCCG	76
TrBr0846	$(AC)^{9}$	CCACTTTCGTCGGTTCCTAC	CAGTCCACGCAATTTCTCCG	100
TrBr0864	$(AAAG)^{6}$	CGTTCGTGGTCATCGGAATTAG	CCATGTGTGTCTCAGGCAATTC	131
TrBr0881	$(AG)^{14}$	ACATTCCAGAGTGTTCGCAAG	ATGTGTTTGTGTCTTGAGCAC	125
TrBr0985	$(AG)^{8}$	CGGGTAAATATTGAATTCGGCC	TGTCTACAAGGAGCTTTCGG	80
TrBr0993	$(AGC)^{10}$	GCAGTCGAGTAAACAACGC	GCTCCGAGGTAAATTATTGCTC	125
TrBr1003	$(AC)^{23}$	CGAGGTGTCAAAGAGCTGTAG	AGTTCAGAGCGGTTTGTTTATG	102
TrBr1103	$(AC)^{11}$	CAATCCTTCCGCTCGTGTTC	AATGTGTCTATGAAAGTGCAGC	76
TrBr1171	$(AG)^{12}$	TATCATTGTTCACCGCCCAC	CTCGCGCATACATTTAGAGAGG	85

Successful amplifications were carried on with an annealing temperature of 54˚C on individuals from 2016 populations

Supplementary Table 2. DNA recovery, sequence read amount, and average sequence heterozygosity for 10 pooled *Trichogramma evanescens* populations

Population	DNA recovered (ng)	Paired reads used	Average sequence heterozygosity (%)
BB	278	67,263,368	0.57
KL90	367	69,333,152	0.141
KL147	366	55,069,496	0.271
BaWc	272	62,340,576	0.253
Ba154	278	74,690,166	0.195
MO8	323	61,848,886	0.334
GRRC16	288	64,778,870	0.244
GRWC17	323	58,117,192	0.671
GRWC27	376	81,086,106	0.253
GD011	264	180,088,002	0.0365

All reads for pooled sequencing comparison were sequenced at 30x 150 bp PE, while GD011 was sequenced at 100x 150 bp PE for *de novo* genome assembly.

Heterozygosity was calculated using JELLYFISH (v2.3.0) (Marçais and Kingsford, 2011) and GENOMESCOPE (v.1) (Vurture et al., 2017).



Chapter 7

Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest

Kim B. Ferguson[a], Sander Visser[b,c], Martina Dalíková[b,c], Irena Provazníková[b,c,†], Alberto Urbaneja[d], Meritxell Pérez-Hedo[d], František Marec[b], John H. Werren[e], Bas J. Zwaan[a], Bart A. Pannebakker[a], and Eveline C. Verhulst[f]

[a] Wageningen University, Laboratory of Genetics, Wageningen, The Netherlands
[b] Biology Centre CAS, Institute of Entomology, České Budějovice, Czech Republic
[c] University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic
[d] Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Moncada, Valencia, Spain
[e] University of Rochester, Department of Biology, Rochester, NY, USA
[f] Wageningen University, Laboratory of Entomology, Wageningen, The Netherlands
[†] Current address: European Molecular Biology Laboratory, Heidelberg, Germany

Abstract

Nesidiocoris tenuis (Reuter) is an efficient predatory biological control agent used throughout the Mediterranean Basin in tomato crops but regarded as a pest in northern European countries. Belonging to the family Miridae, it is an economically important insect yet very little is known in terms of genetic information – no published genome, karyotype, population studies, or RNA transcripts. It is a relatively small and long-lived diploid insect, characteristics that complicate genome sequencing. Here, we circumvent these issues by using a linked-read sequencing strategy on a single female *N. tenuis*. From this, we assembled the 355 Mbp genome and delivered an *ab initio*, homology-based, and evidence-based annotation. Along the way, the bacterial "contamination" was removed from the assembly, which also revealed potential symbionts. Additionally, bacterial lateral gene transfer (LGT) candidates were detected in the *N. tenuis* genome. The complete gene set of 24,688 genes and the associated proteins were compared to other hemipterans (*Cimex lectularis*, *Halyomorpha halys*, and *Acyrthosiphon pisum*), resulting in an initial assessment of unique and shared protein clusters. We visualised the genome using various cytogenetic techniques, such as karyotyping, indicating a karyotype of 2n=32 with a male-heterogametic XX/XY system. Additional analyses include, CGH, GISH, and the location of unique satellite probes via FISH techniques. Finally, population genomics via pooled sequencing further showed the utility of this genome. This is the first mirid genome to be released and represents a step forward in integrating genome sequencing strategies with biological control research.

Introduction

Hemiptera is the fifth largest insect order and the most speciose hemimetabolous order with over 82,000 described species (Panfilio and Angelini, 2018). While recent sequencing projects have presented a variety of information about hemipteran genomes, large families such as the plant bugs Miridae still lack genomic resources. This is despite the diverse life histories present, as it contains not only some of the most notorious agricultural pests but also predators that are often used in biological control (van Lenteren et al., 2018). In addition, Hemiptera are known for their intriguing karyotype

evolution involving holocentric (holokinetic) chromosomes but there is a lack of cytogenetic information on Miridae. Characteristics such as karyotype, sex chromosome system, and presence or absence of telomeric repeats are currently unknown. The absence of the ancestral $TTAGG_n$ telomeric repeat have been reported for mirids *Macrolophus* spp., *Deraeocoris* spp., and *Megaloceroea recticornis* (Geoffroy) (Grozeva et al., 2019, 2011; Jauset et al., 2015) but more knowledge of this trait is necessary for evolutionary studies of genomes and karyotypes. Furthermore, the taxonomic issues that lie within both Miridae and Hemiptera could better be resolved using protein and transcriptome-based analysis, but there is a noted lack of data in this regard as well (Panfilio and Angelini, 2018). While there is a relatively large amount of research into mirids and their use in biological control compared to other predators (Puentes et al., 2018), sequencing projects, if any, often focus on pest species and not on biological control agents (Panfilio and Angelini, 2018). For more advanced molecular methods such as RNAi and CRISPR-based gene editing strategies, it is necessary to have access to genomic and transcriptomic resources of the target species, and so these methods are currently out of reach for *N. tenuis* researchers. This lack in resources on both agricultural pests and biological control agents in the Miridae prompted us to generate genomic and cytogenetic resources of a mirid species that is both.

Nesidiocoris tenuis (Reuter) (Hemiptera: Miridae) is a zoophytophagous mirid used worldwide, including in Spain, the Mediterranean Basin, and China (Pérez-Hedo and Urbaneja, 2016; Xun et al., 2016). Throughout the Mediterranean Basin, *N. tenuis* is used in tomato greenhouses and open fields as a biological control agent against whiteflies (Hemiptera: Aleyrodidae), and the South American tomato pinworm, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) (Calvo et al., 2009; Mollá et al., 2014). In addition, due to its high degree of polyphagous behaviour, it is able to prey on other pest species such as thrips, leaf miners, leafhoppers, aphids, spider mites, and lepidopteran pests (Pérez-Hedo and Urbaneja, 2016). While *N. tenuis* is an important biological control agent in Mediterranean countries (Urbaneja et al., 2012), it is often cited as a pest in other contexts and countries (Calvo et al., 2009; Pérez-Hedo and Urbaneja, 2016). When prey is scarce in tomatoes,

due to its phytophagy, *N. tenuis* can cause plant lesions such as brown discolouration around tender stems, known as necrotic rings, in addition to leaf wilt, and flower abortion (Arnó et al., 2010). This switch to phytophagy has been observed to be inversely proportional to the availability of prey (Sanchez, 2009). Therefore, much of the research thus far has focused on characterizing *N. tenuis* biology and ecology, classifying the induced damage, and attempting to reduce it (Biondi et al., 2015; Castañé et al., 2011; Garantonakis et al., 2018; Martínez-García et al., 2016; Urbaneja-Bernat et al., 2019). Despite its associated plant damage, *N. tenuis* is widely used across South-eastern Spain as it is an efficient predator against the various pests it controls (Arnó et al., 2010). Furthermore, the aforementioned phytophagy has been demonstrated to have benefits by triggering predator-induced defences, including attracting parasitoids, repulsing other herbivorous pests, and restricting accumulation of viruses (Bouagga et al., 2019; Pérez-Hedo et al., 2018, 2015).

In recent years, the controversial success of *N. tenuis* has encouraged the scientific community to research this predatory mirid (Puentes et al., 2018). However, some issues remain to be addressed, such as the genetic variation in commercial stocks of similar biological control agents when compared to wild populations, with the former often diminished in comparison to the latter as seen in other biological control agents (Paspati et al., 2019; Rasmussen et al., 2018; Streito et al., 2017). In order to compare biological control stock to wild (or wild-caught) populations, determining the current diversity and genetic variation of the commercial stock is important. Finally, *N. tenuis* is known to host bacterial symbionts, including *Wolbachia* and *Rickettsia*, though the effect of these bacteria on their host is relatively unknown (Caspi-Fluger et al., 2014). Sequence data can provide additional insight into potential symbionts as well as identify potential LGTs (lateral gene transfers) between host and symbiont.

With all of these fascinating avenues of research in mind, it may be surprising to learn that, aside from a mitogenome (Dai et al., 2012), a regional population analysis (Xun et al., 2016), and more recent work shedding light on evidence of LGT (P. Xu et al., 2019), little genomic information exists for *N. tenuis* and

there is no published *N. tenuis* genome. A likely reason for this absence is that advances made in sequencing technology are often juxtaposed to the complexities of insect life cycles and difficulties in obtaining enough high quality genomic material due to size and exoskeleton (Leung et al., 2019a; Richards and Murali, 2015). Additionally, current assembly tools have a hard time dealing with heterozygosity; therefore, a genome assembly is benefited by sequencing material of reduced genetic heterozygosity for a more contiguous assembly. Reduced heterozygosity is often difficult to achieve in diploid insects where the genetic variation within a population is unknown or the species cannot be inbred (Keeling et al., 2013).

Generating the genomes of highly heterozygous, diploid, and relatively small insects is tricky; researchers have to be prepared to balance their expectations and the available technology (Ellegren, 2014; Leung et al., 2019a). While a single diploid individual may yield enough material for an Illumina-only library, assembly may be difficult due to large repeat regions that extend beyond the insert size of the library. Conversely, enough material could be obtained for sequencing on a long-read platform, but may require pooling material from multiple individuals, potentially complicating assembly due to the heterozygosity of the population. While possible solutions include estimating the heterozygosity or setting up inbred populations (which can be nearly impossible if deleterious effects of inbreeding need to be avoided or if the presence of a complementary sex determining system limits inbreeding (Szűcs et al., 2019; van Wilgenburg et al., 2006)), an alternative is to create a linked-read library. The 10x Genomics platform creates a microfluidic partitioned library that individually barcodes minute amounts of long strands of DNA for further amplification (10x Genomics Inc., Pleasanton, CA, USA). This library is then sequenced on a short-read sequencing platform and then assembled using the barcodes to link reads together into the larger fragment (i.e. Chin et al. 2016; Jones et al. 2017). This method allows for a library to be constructed from a single individual that contains additional structural information to aid assembly (such as phasing), removing the need for pooling multiple individuals and avoiding assembly difficulties in repetitive regions. Additional information, such as karyotype, can further improve genomes in

the assembly stage as well as inform further directions of research by providing chromosome-level context, encouraging further improvement of a genome beyond its initial release.

Here we present the first mirid genome of *Nesidiocoris tenuis* achieved by sequencing a linked-read library of a single adult female bug, along with an annotation based on transcriptome, homology-based, and *ab initio* predictions. In addition to the genome, various avenues for future research are initiated to raise the profile of *N. tenuis* as a research organism, including cytogenetic analyses, protein cluster analysis, and a genome-wide pooled sequencing population genetics analysis. These resources benefit biological control research, as more knowledge becomes available to use in research as well as knowledge of the species for taxonomic and phylogenetic purposes.

Methods

Species origin and description

Individuals of *N. tenuis* were received either from the commercial biological control stock at Koppert Biological Systems, S. L. (Águilas, Murcia, Spain) or from the population maintained for less than a year at WUR Greenhouse Horticulture (Bleiswijk, The Netherlands), which in turn were originally sourced from the Koppert Biological Systems commercial population. Material used for DNA sequencing, PCR testing, pooled sequencing, and cytogenetics was from the Koppert Biological Systems population, while material used for RNA sequencing was from the WUR Greenhouse Horticulture population. Additional species used for cytogenetic comparison purposes were sourced from two separate laboratory populations within the Biology Centre CAS in České Budějovice, Czech Republic: *Triatoma infestans* (Klug) (Hemiptera: Reduviidae) individuals were obtained from a laboratory colony at the Institute of Parasitology that was originally sourced from Bolivia (Schwarz et al., 2014), while *Ephestia kuehniella* (Zeller) (Lepidoptera: Pyralidae) individuals were obtained from a wild-type laboratory colony at the institute of Entomology (Marec and Shvedov, 1990). Species identification of the Koppert Biological Systems population was confirmed via COI sequencing using a PCR amplification protocol (Itou et al., 2013), in addition to testing for

the presence of *Wolbachia* via PCR amplification protocol (Zhou et al., 1998).

Flow cytometry

Genome size was estimated with flow cytometry on propidium-iodide stained nuclei. Individuals from a mixed *Drosophila melanogaster* (Meigen) (Diptera: Drosophilidae) laboratory population (May et al., 2019) were used as the standard for genome size comparison. Following established preparation protocols (De Boer et al., 2007), three samples of single *D. melanogaster* heads, two samples of single *N. tenuis* heads, and one sample of a single *N. tenuis* head pooled with a single *D. melanogaster* head were analysed in a FACS flow cytometer (BD FACSAria™ III Fusion Cell Sorter, BD Biosciences, San Jose, USA). With the known genome size of *D. melanogaster* of 175 Mbp, we could calculate an approximate genome size relative to the amount of fluorescence (Hare and Johnston, 2011).

gDNA Extraction

A single female *N. tenuis* was placed in a 1.5 mL safelock tube with 5-8 one mm glass beads and frozen in liquid nitrogen and shaken for 30 s in a Silamat S6 shaker (Ivoclar Vivadent, Schaan, Liechtenstein). DNA was then extracted using the Qiagen MagAttract Kit (Qiagen, Hilden, Germany). Following an overnight lysis step with Buffer ATL and proteinase K at 56˚C, extraction was performed according to MagAttract Kit protocol. Elution was performed in two steps with 50 µL of Buffer AE (Tris-EDTA) each time, yielding 424 ng of genomic DNA (gDNA) in 100 µL as measured with an Invitrogen Qubit 2.0 fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA).

Library Preparation and Sequencing

Following extraction, gDNA was further diluted to 1 ng/µl following the Chromium Genome Reagent Kits Version 1 User Guide (vCG-00022) (10x Genomics, Pleasanton, USA). A library of Genome Gel Beads was combined with 1 ng of gDNA, Master Mix, and partitioning oil to create Gel Bead-In-EMulsions (GEMs). The GEMs underwent an isothermal amplification step and barcoded DNA fragments were recovered for Illumina library construction

(Illumina, San Diego, USA). The library was then sequenced on an Illumina HiSeq 2500 at the Bioscience Omics Facility at Wageningen University and Research (Wageningen, The Netherlands), yielding 212,910,509 paired-end reads with a read length of 150 bp. After adapter trimming, read size was 126 bp. The first 23 bp of each forward read is a 10X GEM barcode used in the assembly process. Forward read quality was similar to that of the reverse reads, and no reads were flagged for poor quality in a FastQC assessment (Andrews et al., 2015).

Assembly

Using the reads, a k-mer count analysis was performed using GenomeScope on k-mer sizes of 21 and 48, which was used to infer heterozygosity (Vurture et al., 2017). Assembly was performed using all available reads with the GEM barcodes incorporated during the Chromium library preparation in Supernova v2.1.1 (10X Genomics, Pleasanton, USA), with default settings (Weisenfeld et al., 2017). This assembly, v1.0, underwent a preliminary decontamination using NCBI BLASTn v2.2.31+ against the NCBI nucleotide collection (nt) focusing on scaffolds with over 95% homology to bacteria (Camacho et al., 2009), followed by the more elaborate method described below (Detecting contamination and LGT events). Finally, 100% duplicate scaffolds were identified using the dedupe tool within BBTools (sourceforge.net/projects/bbmap/), and removed alongside the contaminated scaffolds, resulting in assembly v1.5. Attempts at further deduplication by adjusting the threshold (such as 95% duplication) resulted in further deletions, but at larger scaffold size, percentage is a rather blunt tool and any percentage is an arbitrary cut-off, so we decided to only remove true duplicates. Assembly completeness for both assemblies were determined using BUSCO v3.0.2 and the `insect_odb9` ortholog set (Simão et al., 2015), while assembly statistics were determined using QUAST (Gurevich et al., 2013).

Detecting contamination and LGT events

Lateral gene transfers (LGTs) from bacteria into metazoan genomes was once thought to be rare or non-existent, but are now known to be relatively common and can evolve into functional genes (Dunning Hotopp et al., 2007;

Husnik and McCutcheon, 2018). We therefore screened our insect genome for LGTs from bacteria. As insect genome assemblies often contain scaffolds from associated bacteria, we first screened for such "contaminating" scaffolds and moved them into a separate metagenomic data files (online supplementary material).

We used a DNA based computational pipeline to both identify likely contaminating bacterial scaffolds in the assembly, and to detect potential LGT from bacteria into the insect genome. The LGT pipeline was modified from an earlier version developed by David Wheeler and John Werren (Wheeler et al., 2013), and has been used to screen for bacterial "contamination" and LGTs in a number of arthropod genomes before (e.g. bedbug *Cimex lectularis* L. (Hemiptera: Cimicidae) (Benoit et al., 2016), parasitoid wasp *Trichogramma pretiosum* Riley (Hymenoptera: Trichogrammatidae) (Lindsey et al., 2018), and the milkweed bug *Oncopeltus fasciatus* (Dallas) (Hemiptera: Lygaeidae) (Panfilio et al., 2019)). In some cases, entire or nearly complete bacterial genomes have been retrieved from arthropod genome projects (e.g. Benoit *et al*., 2016; Lindsey *et al*., 2016)).

Detection of Bacterial Scaffolds in the Assembly

To detect bacterial contaminating scaffolds, the following method was used after the preliminary bacterial contamination assessment described above. First, each scaffold was broken into 1 Kbp fragments and each fragment was subsequently searched with BLASTN against an in-house reference database that contains 2,100 different bacterial species (complete list in supplementary materials) which was masked for low complexity regions using the NCBI Dustmasker function (Morgulis et al., 2006). We recorded each bacterial match with bitscore > 50, the number of bacterial matches, total bacterial coverage in the scaffold, proportion of the scaffold covered, total hit width of coverage (the distance between the leftmost and rightmost bacteria hit proportional to the scaffold size) and the bacterial species with the greatest number of matches within the scaffold from the in-house bacterial data base. It should be noted that the latter method does not indicate the actual bacterial species from which the scaffold was derived, as it is based on

similarity to a curated database – that determination would require follow-up analysis, which was not performed in this study.

Any criterion for deciding whether a scaffold comes from a bacterium is unavoidably arbitrary: Too stringent and insect scaffolds are included; too lax and insect scaffolds are inappropriately removed. We applied a cut-off of ≥ 0.40 proportion bacterial hit width, which has performed well to remove contamination in a few test cases where we have manually examined scaffolds near the cut-off. All instances of contaminated scaffolds were removed from the assembly and are available in supplementary materials as a list and a multi-FASTA file.

Identifying LGT Candidate Regions

We used the same DNA based computational pipeline to identify potential LGTs from bacteria into the insect genome. The basic method is as follows: as before, scaffolds from the genome assembly are broken into 1 Kbp intervals, which are searched against a bacterial genome database. Any positive bacterial hit in a 1 Kbp region (bitscore > 50) was then searched against a database containing transcripts from the following eukaryotes: *Xenopus*, *Daphnia*, *Strongylocentrotus*, *Mus*, *Homo sapiens*, *Aplysia*, *Caenorhabditis*, *Hydra*, *Monosiga,* and *Acanthamoeba* (ftp://ftp.hgsc.bcm.edu/I5K-pilot/LGT_analysis/All_species_genomes/lgt_finder_blastn_database_directories/). The purpose of this eukaryotic screening is to identify highly conserved genes that are shared between eukaryotes and bacteria and exclude these from further analysis. To focus our attentions on the most likely LGT candidates, we selected hits with a bitscore = 0 in the corresponding reference eukaryote database and bitscore > 75 from the bacterial database. We also screened the output for adjacent 1 Kbp pieces that contain bacterial matches and reference eukaryote bitscore = 0 and fused these adjoining pieces for analysis.

LGT candidate regions were then manually curated as follows: each candidate region was searched with BLAST$_N$ to the NCBI nr/nt database. If this search indicated that the region's nucleotide sequence was similar or identical to the nucleotide sequence of a known gene in related insects, it

was discarded as a likely conserved insect gene. Regions were retained only when the matches to other insects were sporadic, as our experience has indicated that these can be independent LGTs into different lineages. If no match was found, the region was additionally searched with BLASTx to the NCBI nr/nt database. If this second search also resulted in no hits to multiple insect proteins, it was called an LGT candidate. In this case, we additionally identified the best bacterial match using the NCBI nr and protein databases. Using the gene annotation information, we then evaluated the flanking genes within the scaffold to determine whether they were eukaryotic or bacterial, we determined whether the LGT region was associated with an annotated gene within the insect genome, and we observed with transcriptome data if RNA sequencing data showed evidence of transcriptional activity in the LGT region. This short list is available in the supplementary materials.

RNA extraction, library construction, and sequencing

Juveniles, adult males, and adult females (approximately 4-5 of each) were prepared for RNAseq using the RNeasy Blood and Tissue Kit (Qiagen). Individuals were placed in a 1.5 mL safelock tube along with 5-8 one mm glass beads placed in liquid nitrogen and then shaken for 30 s in a Silamat S6 shaker (Ivoclar Vivadent). RNeasy Blood and Tissue Kit (Qiagen) was used according to manufacturer's instructions. Samples were measured for quality and RNA quantity using an Invitrogen Qubit 2.0 fluorometer and the RNA BR Assay Kit (Thermo Fisher Scientific). These three RNA samples were then processed by Novogene Bioinformatics Technology Co., Ltd., (Beijing, China) using poly(A) selection followed by cDNA synthesis with random hexamers and library construction with an insert size of 550-600 bp. Paired-end sequencing was performed on an Illumina HiSeq 4000 according to manufacturer's instruction.

Gene finding, transcriptome assembly, and annotation

For the *ab initio* gene finding, a training set was established using the reference genome of *D. melanogaster* (Genbank: GCA_000001215.4; Release 6 plus ISO1 MT) and the associated annotation. The training parameters were used by GlimmerHMM v3.0.1 for gene finding in the *N. tenuis* genome assembly v1.5 (Majoros et al., 2004). For homology-based gene prediction, GeMoMa

v1.6 was used with the *D. melanogaster* reference genome alongside our RNAseq data as evidence for splice site prediction (Keilwagen et al., 2016). For evidence-based gene finding, each set of RNAseq data (male, female, and juvenile) was mapped to the *N. tenuis* genome separately with TopHat v2.0.14 with default settings (Trapnell et al., 2009). After mapping, Cufflinks v2.2.1 was used to assemble transcripts (Trapnell et al., 2010). CodingQuarry v1.2 was used for gene finding in the genome using the assembled transcripts, with the strandness setting set to 'unstranded' (Testa et al., 2015).

The tool EVidenceModeler (EVM) v1.1.1 was used to combine the *ab initio*, homology-based, and evidence-based information, with evidence-based weighted 1, *ab initio* weighted 2, and homology-based weighted 3 (Haas et al., 2008). The resulting amino acid sequences were searched with BLASTp v2.2.31+ on a custom database containing all SwissProt and Refseq genes of *D. melanogaster* (Acland et al., 2014; Boutet et al., 2008; Camacho et al., 2009). The top hit for each amino acid sequence/gene was retained and its Genbank accession number and name are found within the annotation. If no hit was found, an additional search in the NCBI non-redundant protein database (nr) was performed to obtain additional homology data.

Functional annotation and GO term analysis

Gene attributes from the annotation were used to construct a list of genes to be used in Gene Ontology (GO) term classification. Duplicate accession numbers were removed alongside cases where no BLAST hit was found. The remaining accession IDs were converted into UniProtKB accession IDs using the UniProt ID mapping feature (Huang et al., 2011). These UniProtKB accession IDs were in turn used with the DAVID 6.8 Functional Annotation Tool to assign GO terms to each accession ID with the *D. melanogaster* background and generate initial functional analyses (Huang et al., 2009a, 2009b).

Ortholog cluster analysis and comparison

The complete gene set of *N. tenuis* was compared to those of three additional hemipteran species: the bed bug *C. lectularis* (Hemiptera: Cimicidae), the brown marmorated stinkbug *Halyomorpha halys* (Stål) (Hemiptera: Pentatomidae), and the pea aphid *Acyrthosiphon pisum* (Harris) (Hemiptera:

Aphididae) using OrthoVenn2 (L. Xu et al., 2019). The gene set of *A. pisum* is the 2015 version from AphidBase (Legeai et al., 2010; Richards et al., 2010) as maintained on the OrthoVenn2 server. The *H. halys* 2.0 complete gene set was used (Lee et al., 2009) along with the complete gene set of *C. lectularis* (Clec 2.1, OGSv1.3) (Benoit et al., 2016; Thomas et al., 2019), both of which were retrieved from the i5K Workspace (Poelchau et al., 2015). An ortholog cluster analysis was performed on all four gene sets via OrthoVenn2 with the default settings of E-values of 1e-5 and an inflation value of 1.5.

Cytogenetic analysis

Slide preparations

To determine karyotype, *N. tenuis* individuals were obtained from the Koppert Biological Systems population and prepared for cytogenetic experiments. Chromosomal preparations were prepared from the female and male reproductive organs of adults and juveniles by spreading technique according to Traut (1976) with modifications from Mediouni *et al.* (Mediouni et al., 2004; Traut, 1976). After inspection via stereomicroscope to confirm presence of chromosomes, slides were washed and dehydrated in ethanol series (70, 80, and 100%, 30 s each) and stored at -20°C for future use.

18s rDNA probe preparation and Fluorescence In Situ Hybridisation (FISH)

To confirm the presence of 18S rDNA sequences in the assembled genome, the previously published partial 18S rDNA sequence of *N. tenuis* (GU194646, Jung and Lee, 2012) was used as a BLAST query against the *N. tenuis* v1.5 genome. To verify sequence homology, the obtained 18S rDNA sequences were subsequently compared to the previously published sequence.

For preparation of the probe, we isolated gDNA from two *N. tenuis* females with the NucleoSpin DNA Insect Kit (Macherey-Nagel, Düren, Germany) according the manufacturer's protocol. gDNA was used as template in PCR to amplify the 18S rDNA sequence using primers 18S-1 and 18S-4 as described in Jung and Lee (2012). Obtained products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA), and subsequently cloned using the pGEM-T Easy Vector System (Promega, Madison, WI) according to the manufacturer's protocol. Plasmids were

extracted from positive clones with the NucleoSpin Plasmid kit (Macherey-Nagel) following the manufacturer's protocol, confirmed by sequencing (SEQme, Dobříř, Czech Republic), and used as template in PCR with the 18S-1 and 18S-4 primers. PCR-products were purified, and used as template for labelling by a modified nick translation protocol as described by Kato *et al*. (2006) with modifications described in Dalíková *et al*. (2017), using biotin-16-dUTP (Jena Bioscience, Jena, Germany) and an incubation time of 35 minutes at 15°C (Dalíková et al., 2017a; Kato et al., 2006). Fluorescence *in situ* hybridization (FISH) was performed as described in Sahara *et al*. (1999) with modifications described in Zrzavá *et al*. (2018) (Sahara et al., 1999; Zrzavá et al., 2018).

Sex chromosome identification

Determination of the sex chromosome constitution is important for the assembly of the *N. tenuis* genome to identify any potential missing information due to sequencing a single sex, as well as add to knowledge on sex chromosomes in Miridae. Comparative Genomic Hybridization (CGH), and Genomic *In Situ* Hybridization (GISH) were, therefore, used to identify the sex chromosomes of *N. tenuis*. The reproductive organs of adult females were dissected out to avoid potential male gDNA contamination, as the mated status was unknown, after which remaining tissue was snap-frozen in liquid nitrogen and stored at -20°C until further use. Adult males were not dissected but otherwise treated the same. Female and male gDNA was extracted from 10-20 pooled individuals using cetyltrimethylammonium bromide (CTAB) gDNA isolation with modifications (Doyle and Doyle, 1990).

Samples were mechanically disrupted in extraction buffer (2% CTAB, 100 mM Tris-HCl pH 8.0, 40 mM EDTA, 1.4 M NaCl, 0.2% β-mercaptoethanol, 0.1 mg/mL proteinase K), and incubated overnight at 60°C with light agitation. An equal volume chloroform was added, tubes were inverted for 2 min, and samples were centrifuged 10 min at maximum speed. The aqueous phase was transferred to a new tube, RNase A (200 ng/µL) was added and samples were incubated 30 min at 37°C to remove RNA. DNA was precipitated by adding 2/3 volume isopropanol, gently inverting the tubes, and centrifugation for 15 min at maximum speed. Pellets were washed twice with 70% ethanol, air-

dried briefly, and dissolved overnight in sterile water. DNA was stored at -20°C until further use. Probes were prepared with 1 µg gDNA using Cy3-dUTP (for female gDNA), or fluorescein-dUTP (for male gDNA) (both Jena Bioscience, Jena, Germany) by nick translation mentioned above with an incubation time of 2-2.5 hours at 15°C. CGH and GISH were performed according to Traut *et al*. (1999) with modifications described in Dalíková *et al*. (2017) (Dalíková et al., 2017b; Traut et al., 1999).

Detecting a telomeric motif

Initially, we searched both the raw sequencing data and the assembled genome for presence of the ancestral insect telomere motif (TTAGG), which is known to be absent in several Miridae species (Grozeva et al., 2019; Kuznetsova et al., 2011), and tested for its presence using Southern dot blot in *N. tenuis*. gDNA was isolated from *N. tenuis*, and positive controls *E. kuehniella* and *T. infestans*, using CTAB DNA isolation described above. DNA concentrations were measured by Qubit 2.0 (Broad Spectrum DNA Kit) (Invitrogen) and diluted to equalize concentrations, after which 500 ng and 150 ng of each specimen was spotted on a membrane and hybridized as described in (Dalíková et al., 2017b). As a negative control, an equal amount of sonicated DNA from the chum salmon *Oncorhynchus keta* (Walbaum) (Salmoniformes: Salmonidae) (Sigma-Aldrich, St. Louis, MO, USA), was spotted on the same membrane.

Probe template was prepared using non-template PCR according to Sahara *et al*. (1999) and labelling with digoxigenin-11-dUTP (Jena Bioscience) was performed using nick translation, with an incubation time of 50 min according to Dalíková *et al*. (2017) (Dalíková et al., 2017a; Sahara et al., 1999). Absence of the insect telomere motif $(TTAGG)_n$ was confirmed by dot blot, and three sequence motifs, $(TATGG)_n$, $(TTGGG)_n$, and $(TCAGG)_n$, were selected as potential telomeric motifs in *N. tenuis* based on high copy numbers in the genome and sequence similarity to the ancestral insect telomere motif. Copy numbers were determined by Tandem Repeat Finder (TRF, v4), on collapsed quality filtered reads corresponding to 0.5x coverage with default numeric parameters except maximal period size, which was set to 25 bp (Benson, 1999). TRF output was further analysed using Tandem Repeat

ANALYSIS PROGRAM (Sobreira et al., 2006). Probe template, and subsequent labelling of the probes, was done as described above with slight alterations. To obtain optimal length of fragments for labelling, non-template PCR was performed with reduced primer concentrations (50 nM for each primer). In addition, probes were labelled by biotin-16-dUTP (Jena Bioscience) using nick translation as described above, with an incubation time of 50 min. FISH was performed as described above for 18S rDNA.

Repeat identification and visualization

To assess the repetitive component of the *N. tenuis* genome, we used REPEATEXPLORER, v2, on trimmed and quality-filtered reads with default parameters (Novák et al., 2013). Repeats with high abundance in the genome were selected and amplified using PCR. These products, named **Nt_rep1**, were additionally cloned and template for probe labelling was prepared from plasmid DNA as described above, see 18S rDNA probe preparation. Probes were labelled by PCR in a volume of 25 µL consisting of 0.625 U ExTaq polymerase (TaKaRa, Otsu, Japan), 1x ExTaq buffer, 40 µM dATP, dCTP, and dGTP, 14.4 µM dTTP, 25.6 µM biotin-16-dUTP (Jena Bioscience), 400 nM of forward and reverse primer, and 1 ng of purified PCR-product. The amplification program consisted of an initial denaturing step at 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, and a final extension at 72°C for 2 min. The FISH procedure was performed as described for 18S rDNA. Abundance and distribution of **Nt_rep1** in the assembled genome was assessed using NCBI GENOME WORKBENCH v2.13.0. The complete list of primers used in this study can be found in Supplementary Table 2.

POOLED SEQUENCING AND POPULATION ANALYSIS

For this project, it was important to use existing data wherever possible to test the utility of the genome and possible research avenues. Therefore, we analysed whole genome sequence data originally generated for another *N. tenuis* genome assembly project, that has not been published before. In the original set-up, ten females were collected from the Koppert Biological Systems population for pooled sequence analysis. DNA was isolated from

this pooled cohort using the "salting-out" method as described in Sunnucks and Hales with a final aliquot of 20 µL (Sunnucks and Hales, 1996) and then treated with 2 µL RNase. The paired-end library was sequenced on a Illumina HiSeq2500 platform by Macrogen Inc. (Seoul, Korea) with read sizes of 100 bp. Reads were assessed for quality using FastQC (Andrews et al., 2015) and adapters were trimmed with Trimmomatic (Bolger et al., 2014). Following quality filtering, reads with phred scores lower than 20 were discarded. Heterozygosity was calculated using jellyfish v2.3.0 and GenomeScope v1.0 with a k-mer size of 21 and default parameters (Marçais and Kingsford, 2011; Vurture et al., 2017).

Instead of genome assembly, these whole genome sequence reads can be used in a pooled sequencing (pool-seq) population analysis with our genome, with some adaptations. First, the reads were randomly subsampled to a coverage of 10X (in a pool with 10 females, this results in approximately 1X coverage per female) using CLC Genomics Workbench 12 (Qiagen). Using the PoPoolation v1.2.2 pipeline (Kofler et al., 2011), these reads were aligned to an adapted v1.5 genome, where scaffolds smaller than 10,000 bp were removed, and aligned reads were binned into windows using the bwa and samtools packages (Li et al., 2009). Pileup files and the scripts from the PoPoolation pipeline were used to produce variance sliding windows analyses of neutrality, *Tajima's D*, and nucleotide diversity, *Tajima's Pi* (π), with default settings and a pool size of 40. Window and step sizes of both 10,000 and 5,000 were tested, as well as using the "`basic-pipeline/mask-sam-indelregions.pl`" pipeline to mask indel regions of the SAM file – this ensures that indel regions are not calculated. Of the 18,000,000 reads, 817,226 had regions of indels masked.

Results

Species origin, description, and data availability

The Koppert Biological Systems biological control stock was confirmed to be *N. tenuis*, and the presence of *Wolbachia* in individuals from this colony was confirmed (Supplementary Figure 1). All sequence data generated, including raw reads, assembly, and annotation, can be found in the EMBL-EBI European Nucleotide Archive (ENA) under BioProject PRJEB35378.

Genome assembly and size

The single adult female *N. tenuis* yielded 424 ng total DNA. The 10X Genomics Chromium reaction and subsequent Illumina sequencing resulted in more than 212 million paired-end linked-reads. The inferred heterozygosity, based on GenomeScope, was between 1.675% and 1.680% for a k-mer size of 21, and between 1.250% and 1.253% for a k-mer size of 48. Genome size estimates at this point were between 306 Mbp (k-mer=21) and 320 Mbp (k-mer=48). Following assembly with Supernova, assembly v1.0 was approximately 388 Mbp in size and comprised of 44,273 scaffolds (5.91% ambiguous nucleotides).

Assembly v1.0 was then assessed for contamination with a preliminary search against the NCBI for bacterial homology. Several scaffolds with high amounts of bacterial sequence contamination were identified, indicating that further decontamination of the assembly was required. A decontamination pipeline was used to identify and remove a total of 3,043 scaffolds, while those identified as potential examples of LGT were kept. From the remainder, an additional 4,717 were identified as being identical duplicates and were removed. At this point, the resulting assembly was finalised and designated v1.5. This assembly is 355 Mbp in size in 36,513 scaffolds (6.29% ambiguous nucleotides). Quality and completeness of v1.5 using BUSCO indicated a completeness of 87.5% (65.6% single copy orthologs, 21.9% duplicated orthologs), while 7.1% orthologs were fragmented and 5.4% were missing (n=1658).

Initially, the genome size of *N. tenuis* was estimated by flow cytometry to be 232 Mbp, with a confidence interval of 20 Mbp (See supplementary material for more details). Further estimates via k-mer analysis of sequence data in GenomeScope indicated an expected genome size of 306 Mbp (k-mer=21) or 320 Mbp (k-mer=48). Both the flow cytometry and sequence data estimates are smaller than the 355 Mbp of the final assembly (v1.5). In total, the *N. tenuis* genome has 36,513 scaffolds, with the largest scaffold being 1.39 Mbp, though the majority of scaffolds are under 50,000 bp in size. The number of gaps per 100 kbp is 6292.10 (6.29% of the genome). Details on the assemblies can be found in Table 1.

Table 1. Assembly statistics for both versions of the *Nesidiocoris tenuis* assembly, pre- and post-decontamination

Assembly Version	Size (bp)	No. of Scaffolds	N50	Largest scaffold (bp)	No. of N's per 100 kbp (% of genome)	BUSCO score, Complete% (Single%, Duplicate%)
1.0	387,724,797	44,273	27,195	1,392,896	5912.60 (5.91)	81.3 (60.6, 20.7)
1.5 (final assembly)	355,120,802	36,513	28,732	1,392,896	6292.10 (6.29)	87.5 (65.6, 21.9)

Assessment of potential symbionts and LGT candidates

Potential symbionts

The initial assembly (v1.0) was decontaminated using two bacterial decontamination pipelines: the first pipeline broadly utilised BLASTN to identify scaffolds with high amounts of bacterial sequences against the NCBI nr database, while the second pipeline is more specified and uses BLASTN against a list of known contaminants and symbionts and is adapted from previous work (Wheeler et al., 2013). The first decontamination pipeline identified and removed 1,443 scaffolds with high bacteria content, and the second decontamination pipeline identified and removed an additional 1,600 scaffolds alongside potential LGT events. All removed scaffolds are available in online supplementary materials. The hits from the second pipeline were used to create a list of potential contaminants or symbionts of this particular *N. tenuis* individual used for whole genome sequencing according to genus, basepair content, and number of scaffolds affected (Table 2). The majority of these scaffolds (1,470) are under 5 Kbp in length, with an additional 61 scaffolds falling between 5-10 Kbp. The ten largest scaffolds are putatively associated with *Pantoea* and relatives (three of 561,7472 bp, 205,621 bp, and 131,905 bp), *Sodalis* (326,101 bp), *Erwinia* (254,660 bp; 220,307 bp; 154,581 bp), and *Citrobacter* (239,269 bp; 190,839 bp).

We emphasize that these "calls" are very preliminary, as they are based on the most frequent hits in the bacterial matches in each scaffold, rather than comprehensive gene annotations. Nevertheless, they do indicate a range of bacterial types associated with *N. tenuis*, and the scaffold assemblies are likely to contain some complete or near complete bacterial genomes of

interest.

Sorting scaffolds across the range of bacterial genera matches gives 131 genera with some substantial representation: *Erwinia* (2,078,531 bp), *Pantoea* (2,226,778 bp), *Citrobacter* (594,902 bp), *Sodalis* (355,847 bp), *Cronobacter* (314,511 bp), and *Rickettsia* (483,217 bp) (Table 2). In addition to known symbiont *Rickettsia*, previously established via PCR and known symbiont *Wolbachia* is also present in the results (137,109 bp) (Table 2). Multiple genera of bacteria can be found on a single scaffold, likely due to misassembly. The full list of bacterial scaffolds and multiFASTA file is available with details in supplementary materials.

LGT Candidates

We continued with our detection of potential LGT events by further assessing a handful of strong candidates. Two of these regions occur on scaffolds 22012 and 22013, which are of similar length (22,634 bp and 22,957 bp, respectively) and are highly similar on a nucleotide level. Scaffold 22013 appears to have additional nucleotides on each flanking side, with some indels and SNPs between the two scaffolds. The putative LGT region in question is belonging to or gained from a *Sodalis* species, coding for phenazine biosynthesis protein *PhzF* (OIV46256.1). This region also showed transcriptional support, and is flanked by conserved insect genes, most immediately *Rab19* (NP_523970.1) on one side and an uncharacterized protein, Dmel_CG32112 (NP_729820.2), on the other side.

Two additional LGTs were found in the current assembly that match *Rickettsia* sequences (scaffolds 4712 and 27281), which contain a segment of the rickettsial genes *elongation factor G* and *AAA family ATPase* genes, respectively. One corresponds to a gene model, while the other does not, and there is no evidence of expression for either in the current male, female, and mixed sex juvenile RNA sequencing data. More information on these candidate regions can be found in online supplementary materials.

Ab initio gene finding, transcriptome assembly, and annotation

To obtain a comprehensive set of transcripts for *N. tenuis*, three separate libraries of multiple individuals were prepared – males, females, and juveniles

Table 2. Genera of potential symbionts or contaminants as determined by decontamination pipeline based on known contaminants and symbionts against *Nesidiocoris tenuis* assembly v1.0. Identification is according to largest hit percentage, multiple bacterial sections possible in each scaffold. Affected scaffolds were removed leading to assembly v1.5 and are available in online supplementary material, along with full list of hits.

Bacteria Genus	Total amount in genome (bp)	Number of scaffolds affected
Pantoea	1,379,962	307
Erwinia	1,342,298	456
Citrobacter	349,562	40
Rickettsia	204,934	50
Sodalis	189,471	17
Cronobacter	180,427	33
Wolbachia	137,109	47
Enterobacter	110,104	70
Serratia	100,197	49
Klebsiella	44,248	30
Pseudoalteromonas	39,779	118
Pectobacterium	28,130	16
Shigella	25,221	24
Yersinia	24,052	21
Dickeya	23,312	18
Salmonella	19,243	15
Photorhabdus	18,283	14
Escherichia	16,301	11
Rahnella	12,479	10
Burkholderia	7,825	17
Xenorhabdus	6,406	8
Arsenophonus	5,070	8
Pseudomonas	3,427	2
Vulcanisaeta	3,420	3
Ralstonia	3,370	7
Paenibacillus	3,070	8
Others (105)	56,287	201
Total	**4,333,987**	**1,600**

from different stages of mixed sex. More than 77 million 150 bp pair-end reads were generated. Filtering the reads for quality led to a slightly reduced total of 76,711,096 paired reads (male: 28,413,231 paired reads; female: 24,075,901 paired reads; juvenile: 24,221,964 paired reads) to be used for evidence-based gene finding. The mapping and assembling of reads of the three individual samples as well as the pooled reads resulted in four transcriptomes: male, female, juvenile, and the combined transcriptome.

The male, female, juvenile, and combined annotations from the evidence-based gene finding was used alongside homology-based findings and *ab initio* annotations in a weighted model, resulting in complete annotations for the assembly. When gene name assignment via the SwissProt database resulted in "no hit," tracks are named "No_blast_hit." This occurred in 1,556 mRNA tracks and represents approximately 6% of the official gene set. The majority of tracks were annotated with reference to SwissProt or GenBank accession number of the top BLASTp hit.

CodingQuarry predicted 56,309 genes from the mapped transcript evidence, while *ab initio* gene finding using GlimmerHMM resulted in 39,888 genes and homology-based gene finding with GeMoMa resulted in 6,028 genes. The complete gene set for *N. tenuis* was created using EVidenceModeler, where a weighted model using all three inputs resulted in a complete gene set of 24,688 genes.

Functional annotation and GO term analysis

The complete gene set of 24,668 genes was deduplicated and genes with no correlating BLASTp hit were removed. The remaining 11,724 genes were mapped to UniProtKB IDs, resulting in 11,261 genes with a matching ID after another round of deduplication (80 duplicates found). The remaining 383 genes either did not match to a UniProt KB ID or were considered obsolete proteins within the UniParc database.

DAVID used 8,920 genes for the functional annotation analysis, of which 78.4% (6503) contribute to 19 biological processes, 75.8% (6826) contribute to 100 different cellular components, and 72.8% (6032) contribute to 91 categories of molecular functions (genes can code to multiple GO terms).

The remaining genes were uncategorized. Data linking the genes to the GO terms, the DAVID GENE LIST REPORT, and the DAVID GENE REPORT are available in supplementary materials.

ORTHOLOG CLUSTER ANALYSIS

The complete gene set of *N. tenuis* was compared to those of three additional species: the bed bug *C. lectularis*, the brown marmorated stinkbug *H. halys*, and the pea aphid *A. pisum* using ORTHOVENN2. The ortholog analysis summary is presented in Table 3 and visualized in Figure 1. *N. tenuis* has a similar number

Table 3. Output of ORTHOVENN2 ortholog cluster analysis of *Nesidiocoris tenuis*, *Cimex lectularis*, *Halyomorpha halys*, and *Acyrthosiphon pisum*.

Species	Proteins	Clusters	Singletons	Source of complete gene set
N. tenuis	24,668	8,174	9,136	This chapter (supplementary)
C. lectularis	12,699	7,989	7,989	Benoit *et al.*, 2016; Poelchau *et al.*, 2015; Thomas *et al.*, 2019
H. halys	25,026	9,584	2,170	Lee *et al.*, 2009; Poelchau *et al.*, 2015
A. pisum	36,195	8,765	7,298	Legeai *et al.*, 2010; L. Xu *et al.*, 2019

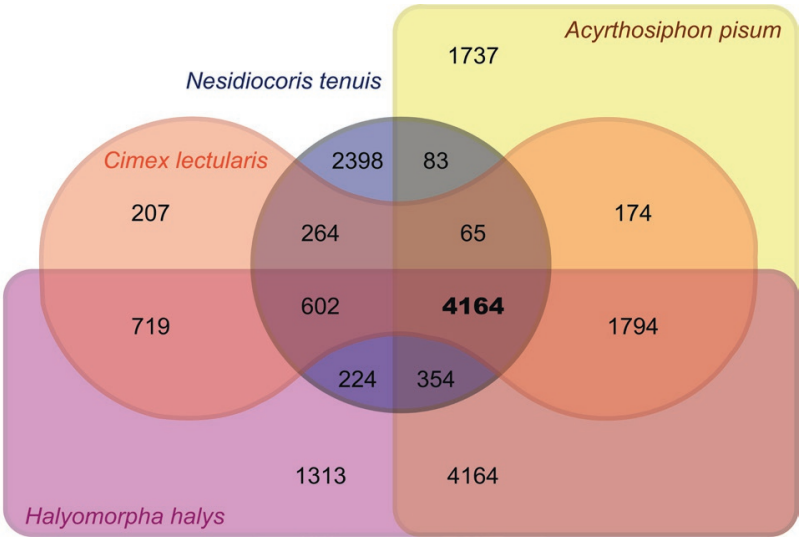


Figure 1. Ortholog cluster analysis of *Nesidiocoris tenuis* with three other hemipterans (*Cimex lectularis*, *Halyomorpha halys*, and *Acyrthosiphon pisum*). Numbers indicate the number of ortholog clusters in each grouping, with the clusters shared by all four species in bold.

of clusters (8,174) as compared to *C. lectularis, H. halys and A. pisum* (7,989; 9,584; and 8,765, respectively). In total 14,512 clusters are assigned, 12,964 of which are orthologous clusters (contains at least two species), and the remaining 1,548 are single-copy gene clusters. There are 9,136 singleton clusters in *N. tenuis*, 3,573 in *C. lectularis*, 2,170 in *H. halys*, and 7,298 in *A. pisum*.

The amount of singleton clusters, i.e. proteins that do not cluster, indicate that *N. tenuis* differs the most from the other species, as 37.04% of the proteins are singletons. Just over half of the orthologs cluster with *N. tenuis*, where 6,338 clusters are outside of *N. tenuis* as compared to the 8,174 clusters within *N. tenuis*. The final protein set from *N. tenuis* used in this analysis is available via supplementary materials.

Karyotype analysis

Karyotype analysis revealed 2n=32 chromosomes in both females and males (Figure 2a and b). All chromosomes are relatively small with one larger pair of submetacentric chromosomes in females (Figure 2a). In males (Figure 2b), we were unable to obtain mitotic chromosomes of reasonable quality as in females and therefore we were unable to clearly identify these larger chromosomes. Screening of multiple nuclei showed sporadic deviations of the karyotype in some individuals. This was the result of supernumerary chromosomes (B chromosomes) which were clearly visible in (meiotic) pachytene stage as distinctly smaller chromosomes (Figure 2c, three B chromosomes). The count of B chromosomes varied between individuals and these B chromosomes were not present in all individuals surveyed.

Analysis and localization of 18S rDNA

The 18S rRNA gene is often used as a cytogenetic marker in comparative evolutionary studies due to its ease of visualization on the chromosomes caused by high copy number and cluster organisation in animal (Sochorová et al., 2018) and plant (Gomez-Rodriguez et al., 2013) genomes. The published partial 18S sequence of *N. tenuis* (GU194646.1) and the 18S sequence identified in this study were compared to each other revealing some differences between the sequences. The published sequence consists of two

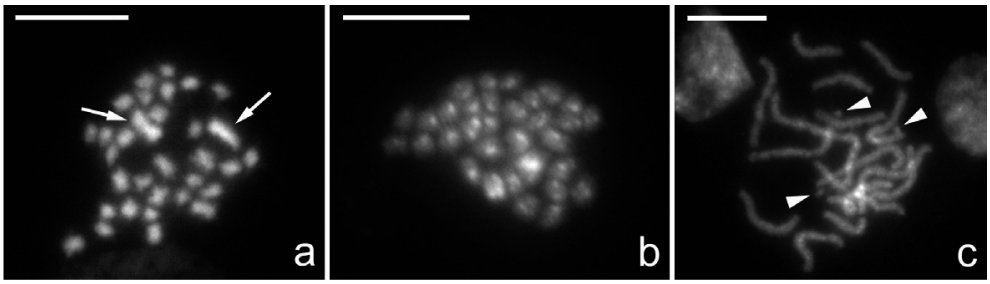


Figure 2. Cytogenetic analysis of *Nesidiocoris tenius* karyotype. Chromosomes were counterstained by DAPI (grey). (a) Female mitotic metaphase consisting of 32 chromosomes (2n=32) with two large chromosomes indicated (arrows). (b) Male mitotic metaphase consisting of 32 chromosomes (2n=32). (c) Female pachytene nucleus with B chromosomes (arrowheads). Scale bar = 10 µm.

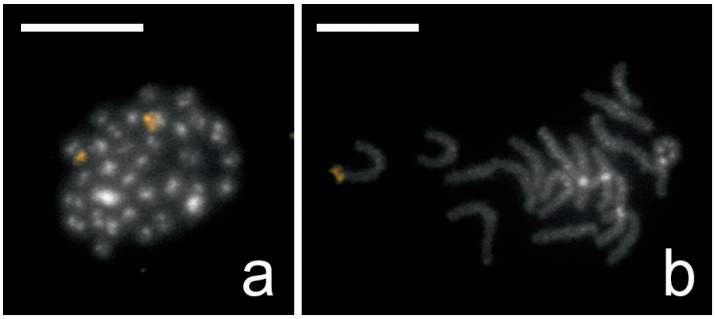


Figure 3. Results of fluorescence *in situ* hybridization with 18S rDNA probe labelled by biotin and visualised in *Nesidiocoris tenuis* chromosome preparations by detection with Cy3-conjugated streptavidin (gold). Chromosomes were counterstained by DAPI (grey). (a) Male mitotic metaphase; probe identified a cluster of 18S rDNA on two homologues chromosomes. (b) Female pachytene complement with one terminal cluster of 18S rDNA genes on a bivalent. Scale bar = 10 µm.

fragments of 869 bp and 739 bp, which are, respectively, 99.7% and 94.2% homologous to our identified partial 18S sequence. Interestingly, the second half of our isolated 18S sequence is more homologous to a *Macrolophus sp*. partial 18S sequence (EU683153.1), i.e. 97.8%, than to the previously published *N. tenuis* sequence. A BLAST search against the *N. tenuis* genome with either of the *N. tenuis* 18S sequences resulted in four gene copies in both cases, each located on a different scaffold. However, RepeatExplorer estimated 98 18S rDNA copies with the obtained genome size of 355 Mbp. Using FISH with the 18S rDNA probe we finally showed that the major rDNA forms a single cluster located terminally on a pair of homologous chromosomes (Figure 3).

IDENTIFICATION OF SEX CHROMOSOMES

The common sex chromosome constitution in Miridae is the male-heterogametic XX/XY system. To identify the sex chromosome constitution and estimate sex chromosome differentiation in N. *tenuis* we employed GISH and CGH experiments. The GISH results clearly revealed a single chromosome densely labelled by the male-derived probe, caused by male-enriched repetitive DNA and/or male-specific sequences which is typical for

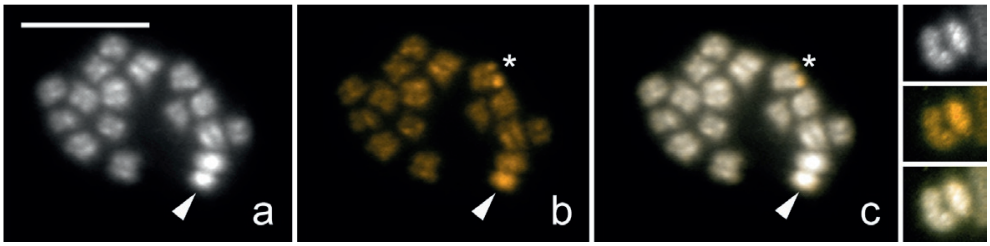


Figure 4. Genomic *in situ* hybridization (GISH) on male chromosomal preparation of *Nesidiocoris tenuis*. Panel (a) shows DAPI counterstaining (grey), panel (b) hybridisation signals of the male derived genomic probe labelled by Cy3 (gold) together with competitor generated from unlabelled female genomic DNA, and panel (c) a merged image. (a, b, c, detail) Meiotic metaphase I, male derived probe highlighted the Y chromosome (arrowhead) more (b, c) compared to autosomes and the X chromosome. Note highlighted terminal regions of one of the bivalents caused by presence of major rDNA genes (asterisk). (detail) Detail picture of XY bivalent; Y chromosome labelled by male derived probe. Note that the Y chromosome is smaller in size and showing more heterochromatin compared to the X chromosome (and autosomes). Scale bar = 10 µm.

the Y chromosome (Figure 4). In addition, the Nucleolus Organizer Region (NOR; including 18S rDNA) was observed as well, as is often the case in GISH experiments due to the presence of highly repetitive sequences in the rDNA cluster. The NOR is clearly located terminally on a pair of autosomes, corroborating our 18S rDNA FISH results.

To further study the differentiation of the sex chromosomes we carried out CGH experiments on chromosome preparations of both sexes (Figure 5). All chromosomes were labeled evenly by the female and male probes with the exception of the largest chromosome pair. Both sex chromosomes were highlighted with DAPI (Figure 5a, e), indicating that they are both A-T rich and largely composed of heterochromatin. In females, the largest chromosome

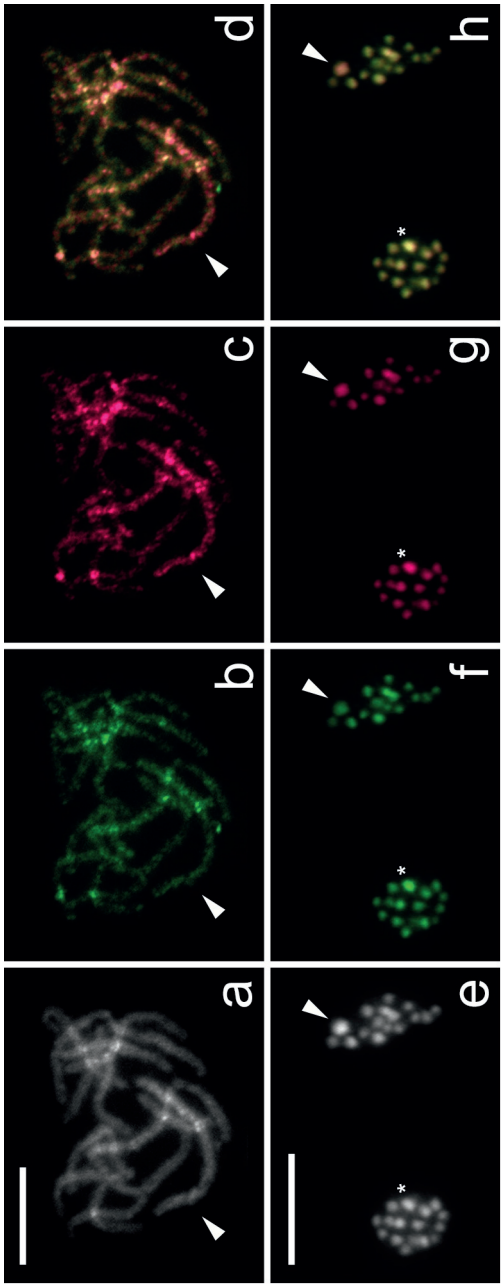


Figure 5. Comparative genomic hybridization (CGH) on female (a, b, c, d) and male meiotic metaphase II (e, f, g, h) chromosomes of *Nesidiocoris tenuis*. **Note: alternate colouration in supplementary materials, Supplementary Figure 4.** Panels (a, e) show chromosomes counterstained by DAPI (grey), panels (b, f) hybridization signals of the male derived genomic probe labelled by fluorescein (green), panels (c, g) hybridization signals of the female derived genomic probe labelled by Cy3 (magenta), and panels (d, h) merged images. (c, d) Note that the X chromosome bivalent (arrowhead) in female pachytene complement was highlighted more by female probe compared to the autosomal bivalents; (b, d) male probe labelled all chromosomes equally. (h) Two sister nuclei in meiotic metaphase II showed equal hybridization patterns of both probes on autosomes; in one of the forming nuclei, the X chromosome (arrowhead) was highlighted by female derived genomic probe (g, h) and in the second nucleus the Y chromosome (asterisk) was strongly highlighted by male derived genomic probe compared to autosomes (f, h) and less highlighted by female derived probe (g, h). (e) Note that the sex chromosomes are the biggest and most heterochromatic elements in the nucleus. Scale bar=10 µm.

pair was labelled more by the female probe than the male probe indicating that these chromosomes contain sequences with higher copy numbers in females, and are thus the X chromosomes, as seen in Figure 5a-d. In male meiotic nuclei (Figure 5e-h), two types of nuclei can be discerned, where the largest chromosome was labelled more by either the female probe or the male probe corresponding to the X, and Y chromosome, respectively, whereas the autosomes were labelled equally by both probes.

Identification and mapping of abundant repeats

RepeatExplorer software was used on reads with GEM barcodes removed to identify the most abundant repeats in the genome of *N. tenuis* (results available in supplementary materials, see Supplementary Table 1). The most abundant repeat, **Nt_rep1**, makes up approximately 3% of the genome estimated by RepeatExplorer. Analysis on the assembled genome, using a coverage cut-off value of 70%, reveals that **Nt_rep1** is present on 3190 scaffolds (8.737% of the assembled scaffolds), with a maximum of 17 copies on a single scaffold. According to the assembled genome, **Nt_rep1** makes up approximately 0.8% of the entire genome (Supplementary Figure 2). We subsequently mapped **Nt_rep1** to the chromosomes of *N. tenuis* using FISH. The repeat is located on most chromosomes and is accumulated in sub-telomeric regions (Figure 6). Additional signals were identified on the X chromosome indicating a higher number of this repeat (Figure 6a-c). This increase in frequency is specific to the X chromosome and is not found on the Y chromosome of *N. tenuis* (Figure 6d-f).

Testing of candidate telomere motifs

Analysis of the raw sequencing data and the assembled genome both revealed low numbers of the insect telomere motif $(TTAGG)_n$ (Frydrychová et al., 2004) in *N. tenuis*, i.e. approximately 98 repeats per haploid genome. This translates into approximately three copies of the repeat per chromosome end, much lower than expected for a telomeric motif. These low copy numbers were additionally confirmed using Southern dot blot (Supplementary Figure 3). Other candidate telomere motifs previously identified by TRF analysis, being $(TATGG)_n$, $(TTGGG)_n$, and $(TCAGG)_n$, were examined by FISH

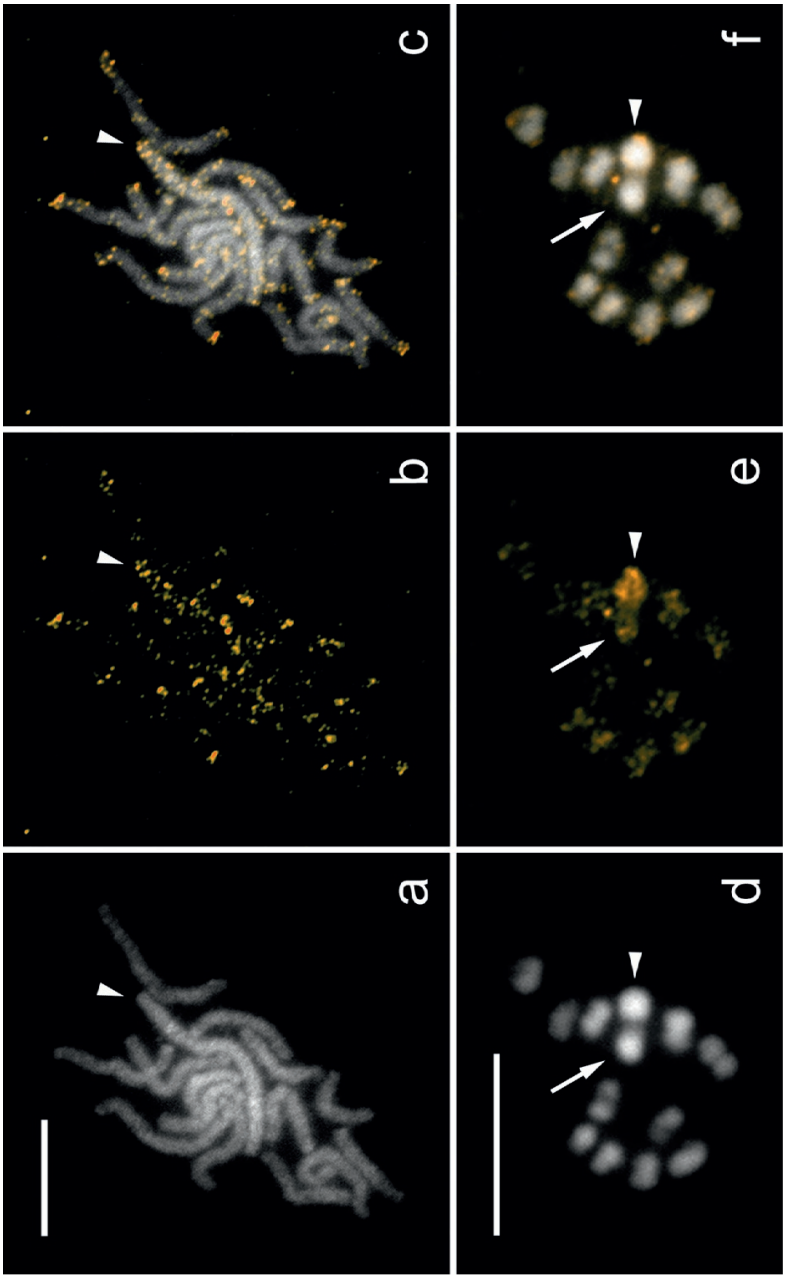


Figure 6. Fluorescence *in situ* hybridization with **Nt_rep1** probe labelled by biotin (gold) on female (a, b, c) and male (d, e, f) chromosomes of *Nesidiocoris tenuis* counterstained with DAPI (grey). (a, b, c) Female pachytene chromosomes; **Nt_rep1** probe highlighted the pair of X chromosomes (arrowhead). (b, c) Note that terminal regions of all bivalents were also labelled by probe, probably due to presence of this sequence in sub-telomeric regions. (d, e, f) Incomplete male nucleus in meiotic metaphase I; probe highlighted the X chromosome (arrowhead) more compared to autosomes and Y chromosome (arrow). (b, c, e, f) Strong hybridization signals on X chromosomes in both sexes were caused by enrichment of **Nt_rep1** sequence on the X chromosomes. Scale bar = 10µm.

for their distribution in the genome. They were found scattered throughout the genome but lacked a clear accumulation at the terminal regions of the chromosomes (not shown). Therefore, these sequences can also be excluded as telomeric motifs in *N. tenuis*.

Pooled sequencing analysis

Using previously generated whole genome sequencing of ten females from the Koppert Biological Systems population, we were able to estimate genetic diversity of the commercial population via a pool-seq population analysis. Read coverage was randomly subsampled to 10X coverage (18,000,000 reads). Additionally, we used a modified v1.5 *N. tenuis* genome with scaffolds of less than 10,000 bp removed. This was to ensure that window sliding was not being inflated on scaffolds smaller than the window size. This reduced the genome from 36,513 scaffolds to 7,076, however, the reduced genome still contained 72.23% of the genome in terms of size (256,487,768 bp).

Three runs of PoPoolation were performed with varied window size, step size, and the masking of indel regions. The default setting, window size and step size of 10,000, yielded similar results as the adjusted window size and step size of 5,000, while differences were apparent when indel regions were masked. As such, results of window size and step size 10,000 with indel regions mapped are reported here (other results available in online supplementary materials). The variance sliding program created 28,833 windows of 10,000 bp with mapped reads, of which 5,913 were sufficiently covered with reads to calculate values per window (coverage ≥ 0.60). Genome-wide, the nucleotide diversity (*Tajima's* π) is 0.0080 and *Tajima's D* is -0.0355. Figure 7 shows the *D* (a) and π (b) for the ten largest scaffolds, all containing gene annotations, arranged in order of size. These ten scaffolds represent approximately 1.7% of the genome (6,135,756 bp), and varied in terms of window coverage (from no coverage to full coverage) as well as both *D* and π. These ten scaffolds are a snapshot of the whole genome, summarised in Table 4, whereas genome-wide results can be found in supplementary materials.

Discussion

Assembly and Annotation

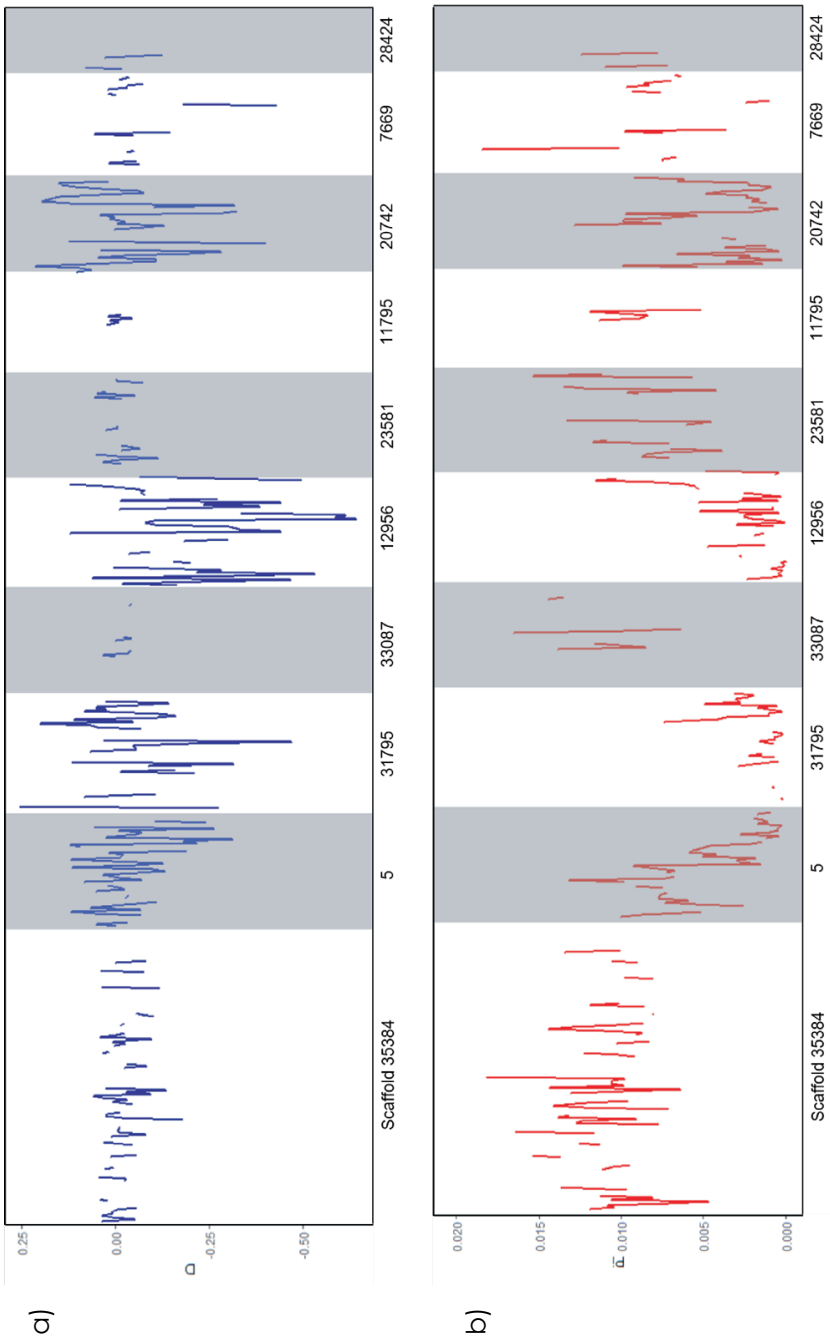


Figure 7. Genetic diversity of Koppert Biological Systems commercial *Nesidiocoris tenuis* population according to *Tajima's D* (7a), and nucleotide diversity, *Tajima's π* (7b). Scaffolds are ordered according to size, which each name beneath on the x-axis.

Table 4. PoPoolation analysis on commercial Koppert Biological Systems *Nesidiocoris tenuis* population (n=10 females), with 10 largest scaffolds according to size. Coverage is ≥ 0.60 and indel regions are masked.

Scaffold	Size (bp)	Windows (10 Kbp)	Number of sufficiently covered windows	Average coverage of sufficiently covered windows	Average *Tajima's π* across scaffold	Average *Tajima's D* across scaffold
35384	1,392,896	140	70	0.68	0.010682	-0.02091
5	613,435	62	48	0.76	0.004932	-0.03869
31795	577,751	58	38	0.74	0.001753	-0.04655
33087	539,928	54	17	0.66	0.012571	-0.00458
12956	519,254	52	39	0.80	0.002541	-0.21012
23581	513,368	52	28	0.70	0.008417	-0.04658
11795	508,155	51	14	0.68	0.011553	-0.01147
20742	504,856	51	39	0.78	0.004165	-0.01963
7669	488,533	49	24	0.67	0.007856	-0.0551
28424	477,580	48	5	0.66	0.009563	-0.00998
Total	**256,487,768**	**28,833**	**5,913**	**0.70**	**0.0080**	**-0.0355**

Presented here is the first mirid genome of *N. tenuis,* a biological control agent used throughout the Mediterranean in tomato crops. We chose to use 10X Genomics sequencing strategy as it best suited the challenges that come with working with a relatively small and long-lived mirid such as *N. tenuis*. Assembling a genome is easiest with reduced heterozygosity in the input sample, often through single individual sampling or inbreeding (Ekblom and Wolf, 2014; Richards and Murali, 2015). This proved an initial challenge for the sequencing strategy of *N. tenuis*, as they are too small for a single individual to yield the minimum amount of DNA required for a traditional NGS library, and an inbred population was not readily available for sequencing. Therefore, 10X Genomics and a linked-read library was the immediate solution for which a small amount of input DNA from a single individual would yield a highly contiguous genome.

Assemblies v1.0 and v1.5 contain 5.91% and 6.29% ambiguous nucleotides, while still offering a relatively high BUSCO score, with the final decontaminated assembly (v1.5) having a completeness of 87.5% of the `insect_0db9` ortholog

dataset. However, the final assembled genome size is approximately 150 Mbp larger than was expected based on flow cytometry data, and we suggest the assembly presented here can best be improved in terms of accuracy and contiguity with long reads from an inbred sample. This discrepancy between estimated genome size and assembled genome size may also be due to the ambiguous nucleotides inserted into the genome during the assembly process. Making up just over 6% of the final assembled genome, that is approximately 2.2 Mbp of ambiguous nucleotides. However, most of the genome inflation is likely due to residual contamination along with duplicate scaffolds that remain after removing 100% identical ones.

Annotation via evidence-based, homology-based, and *ab initio* models resulted in 24,668 genes. Compared to other assemblies within the hemipteran order, such as *C. lectularis*, with a genome size of 650 Mbp and 12,699 genes (Thomas et al., 2019) or *A. pisum*, with a draft genome size of 464 Mbp and 36,195 genes (Richards et al., 2010), *N. tenuis* sits, in the middle in terms of genome size and number of genes. It is worth noting that of the 24,668 genes within the complete gene set, only 11,261 (45.7%) remained after UniProtKB mapping, of which 8,920 (36.2% of total) were used by DAVID for functional analysis. This is relatively low compared to similar genome projects, such as *Aphys gossypii* (Glover) (Hemiptera: Aphididae), where 49.2% of the gene set could be used for GO term analysis (Quan et al., 2019). However, we used different methods which may explain the difference. The next step for the *N. tenuis* genome is manual annotation and curation, which would likely improve the GO term analysis, but this requires time and expertise. Still, we hope that other researchers will use and add to the annotation.

Comparing the current gene set of *N. tenuis* to other Hemipterans, the clustering identified considerable overlap, as 71% of the clusters that are found in *N. tenuis* were shared between the other species in the comparative analysis. Despite being more closely related to *C. lectularis* in terms of phylogeny, in terms of lifestyle, *N. tenuis* is far more similar to *A. pisum* and *H. halys*, and this is likely reflected in absolute number of proteins and clusters shared between the four species. The remaining 29% of clusters, as well as the singleton proteins, are indications for proteins unique to either *N. tenuis* or

Miridae in general. Through the OrthoVenn2 website, the analysis performed here can be easily replicated, altered with other species of interest, and even improved upon if the complete gene sets are updated or with a newer software version. In our iteration, the 24,668 proteins of *N. tenuis* group into 8,174 clusters. 2,398 clusters are unique to *N. tenuis*, however, some of these genes have a relatively strong homology to genes of one of the other species used in the analysis and could be incorrectly flagged as being unique. Reasons could be poor gene annotation quality resulting in a poor *in silico* protein translation, or too stringent clustering settings. Regardless, these 2,398 clusters may be of interest to researchers working on zoophytophagy, the negative effects of *N. tenuis* on tomato as compared to other mirids, or broader questions such as phylogeny of the Hemiptera.

Characterizing the Genome

Every sequence and assembly strategy has benefits and drawbacks, and the 10X Genomics linked-read strategy is no exception. The technique requires only few nanograms of DNA for library preparation which allowed us to use a single individual and removed the need for inbreeding to reduce variation in the sequencing population. However, using a single individual from a closed and proprietary rearing process presented other challenges. These challenges were threefold: we had to deal with bacterial contamination as antibiotic treatment is not possible, we had to ensure that the single individual-derived assembly reflects reality in terms of genes present and structure, and we had to ensure that a single female-derived assembly is applicable for population-level analyses.

Contamination of genomes is a constant concern, and sequencing strategies should attempt to address the risks in the best way possible to deliver reliable genomes (Ekblom and Wolf, 2014). Equally so is the desire for inbred strains if multiple individuals are required to reach the micrograms of DNA necessary for NGS platforms. The inability to remove symbionts and microbiota using antibiotics administered to a few successive generations as well as the difficulty or inability to inbreed a strain is not restricted to *N. tenuis*. The sequencing strategy chosen for the mountain pine beetle, *Dendroctonus ponderosae*

(Hopkins) (Coleoptera: Curculionidae), relied on assuming the relatedness of several individuals as well as isolating the gut during the extraction process, and still additional post-assembly decontamination was required (Keeling et al., 2013). A linked-read strategy with low input requirements, such as the 10X Genomics library that was chosen here, negates the need for a pool of inbred samples or controlling for relatedness. However, another potential benefit of controlled rearing such as those used in inbreeding (as opposed to be limited to wild-caught specimens, for example) is the ability to treat with antibiotics for multiple generations. Without the ability to do so, sequencing and assembly strategies rely heavily on post-sequencing decontamination strategies (both pre- and post-assembly are possible). That such post-assembly filtering strategies as used here for *N. tenuis* can be successful was shown by a less contaminated assembly and by the identification of potential LGTs.

Beyond the Genome: Potential symbionts and LGT events

The list of potential symbionts or pathogens generated in Table 2 represent both insect and plant pathogens, as well as potential environmental contaminants. In addition to the positive test for *Wolbachia* in the Koppert Biological Systems population used here, *N. tenuis* is known to potentially harbour *Rickettsia* as an endosymbiont in addition to *Wolbachia* (Caspi-Fluger et al., 2014). *Rickettsia* genome sizes can range from 0.8 to 2.3 Mbp, also reflecting variation in levels of reductive evolution (Sachman-Ruiz and Quiroz-Castañeda, 2018). However, the total scaffold length identified here as from a Rickettsia falls below this range, likely indicating incomplete recovery of the genome from the insect sequencing. The potential symbionts revealed included not only *Wolbachia* and *Rickettsia*, but also other known insect symbionts, in addition to the usual lab contamination suspects.

Sodalis is a genus of bacterium symbiotic with various insects, including the tsetse fly and louse fly, louse and hemipteran species (Boyd et al., 2016). Genome sizes of *Sodalis* and close relatives range from 0.35 to 4.57 Mbp (Santos-Garcia et al., 2017). The relatively small total scaffold size found in our results (0.36 Mbp) likely reflects incomplete genome recovery in the assembly, but could also be due to genome size reduction, and is worthy of further investigation. *Erwinia* and *Pantoea* are closely related bacteria that are

associated with plant pathology (Kamber et al., 2012; Zhang and Qiu, 2015) and both have been found in the midgut of stink bugs as vertically transferred plant-associated bacteria that become temporary endosymbionts of stink bugs until later replacement with another endosymbiont (Prado and Almeida, 2009). The genome sizes of *Erwinia* and *Pantoea* species typically range from 3.8-5.1 Mbp. Our total scaffold size for the *Erwinia* and *Pantoea* are substantially smaller (2.078 Mbp and 2.22 Mbp), but it is possible that these scaffolds belong to the same bacterium. In any case, the association of a zoophytophagous mirid bug with potential plant pathogens is noteworthy, especially in a biological control context.

As for *Serratia*, *S. marcescens* is both a common Gram-negative human-borne pathogen and a causal agent of cucurbit yellow vine disease (CYVD) (Abreo and Altier, 2019; Bruton et al., 2007). It is worth noting that in cases of CYVD, the transmission of *Serratia marcescens* from its vector the squash bug, *Anasa tristis* (De Geer) (Hemiptera: Coreidae), to host crops is via the phloem. Other *Serratia* spp. have been identified as insect symbionts previously, as have other potential symbionts found in the contaminated scaffolds, such as *Cedecea* spp. (Jang and Nishijima, 1990). The presence of *Dickeya* is an interesting find, as *Dickeya dadantii* has been established as a pathogen of *A. pisum*, while the pea aphid itself is a potential vector for the bacterium with regards to plants (Costechareyre et al., 2012). *Dickeya* spp. cause soft rot in various crops, including tomato. In a similar vein, the identification of *Ralstonia,* as some members of this order, such as *Ralstonia solanacearum,* are soil-borne pathogens that causes wilt in several crop plants, including tomato (Lowe-Power et al., 2018). However, both *Dickeya* spp. and *R. solanacearum* infect the xylem while *N. tenuis* is a phloem-feeder. The scaffold lengths for *Citrobacter* and *Cronobacter* are also considerably below their typical genome sizes, and likely represent incomplete sequence recovery in the metagenomic sample.

All of these described associations are preliminary, as follow-up analyses against the entire NCBI database and proper bacterial gene annotation are lacking. Nevertheless, these putative bacterial associations of *N. tenuis*, their distribution within the insect, and their possible biological significance,

warrants further investigation. It is important to note that some of these bacterial "contaminants" may actually represent large LGTs, which can be confirmed by identifying flanking sequences (e.g. using long-read technologies) and/or *in situ* chromosome hybridization analyses, such as done for the large LGT in *Drosophila ananassae* (Doleschall) (Diptera: Drosophilidae) (Dunning Hotopp et al., 2007). More research into the symbionts of *N. tenuis* via metagenomics would certainly shed some light on true symbionts (or pathogens) versus true contaminants, with potential implications for biological control and related research.

One of the LGT candidate genes that were detected following manual curation corresponds to phenazine biosynthesis protein *PhzF* (OIV46256.1), with the likely microbial source being a *Sodalis* species. Phenazines are heterocyclic metabolites with "antibiotic, antitumor, and antiparasitic activity," but are also toxic when excreted by bacteria (Blankenfeldt et al., 2004). This LGT region exhibits gene expression and is flanked by conserved insect genes, providing further support for it being a legitimate LGT, though further research into this region will be necessary to confirm this. The gene occurs on two different scaffolds, 22012 and 22013, which are highly similar to each other in some regions at the nucleotide level. These could represent homologous regions that differ sufficiently to assemble as different scaffolds, or alternatively a duplication in two different regions of the genome. Future work should focus on its expression patterns in different tissues (e.g. salivary glands, in interest of *PhzF*) and potential functional role in *N. tenuis*.

Beyond the Genome: Cytogenetics

We determined the karyotype of *N. tenuis* to be 2n=32 (30+XY in males) chromosomes which is the second most common chromosome number in the family Miridae (Kuznetsova et al., 2011). In addition, we have shown that *N. tenuis* has an XX/XY sex chromosome constitution, with the sex chromosomes being the largest elements in the karyotype. This is different from the closely related *Macrolophus costalis* (Fieber) (Hemiptera: Miridae) (2n=24+X_1X_2Y), and *M. pygmaeus* (Rambur) (2n=26+XY) where two pairs of autosomes are larger than the sex chromosomes, yet similar to *M. melanotoma* (Costa)

which only differs from *N. tenuis* in the number of autosomes, 2n=32+XY (Jauset et al., 2015). As we sequenced a single female, sequence information of the Y chromosome is missing from our genome assembly. While analyzing the *N. tenuis* karyotype we discovered the sporadic presence of B chromosomes in the Koppert Biological Systems population. B chromosomes are supernumerary chromosomes that are dispensable to the organism, and are often present in only a subset of individuals from a population (Banaei-Moghaddam et al., 2013). Supernumerary chromosomes are common in Heteroptera, yet only a few species of Miridae have been identified to carry supernumerary chromosomes (Grozeva et al., 2011). Presence of B chromosomes in high numbers within an individual is often found to be detrimental, though in lower numbers they are often considered neutral or, in some cases, beneficial (Camacho et al., 2000; Jones and Rees, 1982). The abundance of B chromosomes in *N. tenuis* biological control populations is currently unknown but determining their potential effects on fitness-relevant traits might reveal beneficial information for the optimization of mass-reared populations.

The hemizygous sex chromosomes of most organisms have a high content of repetitive DNA, consisting of multiple different repetitive sequences that are less frequent found on autosomes (Charlesworth and Charlesworth, 2000; Traut et al., 1999). Therefore, the use of cytogenetic techniques, such as CGH and GISH, in the identification of hemizygous sex chromosomes is a powerful tool and is well established in different groups of organisms, e.g. Lepidoptera (Carabajal Paladino et al., 2019; Dalíková et al., 2017a; Zrzavá et al., 2018), Orthoptera (Jetybayev et al., 2017), fish (Sember et al., 2018), and frogs (Gatto et al., 2018). However, to our knowledge, this is the first time these techniques have been used in the family Miridae. The X and Y chromosome of *N. tenuis* are similar in size, with the X chromosome being slightly bigger, and are difficult to distinguish from each other based solely on their appearance without special probing. Our CGH and GISH results showed relatively weak hybridization signals of genomic probes on the sex chromosomes compared to other species indicating little differentiation of sequence content between the X and Y chromosomes, and/or between the

sex chromosomes and the autosomes. Though the hybridization signals are relatively weak, not only in the Y chromosome but also the homogametic sex chromosome, the X chromosome, is distinguishable in the CGH results, which shows X-enriched or X-specific repetitive DNA, similar to what was found on the Z chromosome in *Abraxas* spp. (Zrzavá et al., 2018). Mapping of the most abundant repeat in the genome revealed that one such X- enriched repeats is **Nt_rep1**, confirming the outcomes of our CGH results.

The low copy numbers of 18S rDNA identified in the assembled genome were surprising. The NOR is usually composed of tens to hundreds of copies, and is therefore used in heteropteran cytogenetic studies due to its easy visualization (Kuznetsova et al., 2011). Analysis of the raw data estimates 98 copies of 18S rDNA are present in the genome, yet the majority of these copies are missing from the final assembly. The FISH results show that 18S rDNA is present as a single cluster in the genome, indicating that there is a limit to the genome assembler Supernova, and 10X Genomics by extension, and its ability to assemble highly repetitive regions of the genome. Similarly, the FISH results of **Nt_rep1** and the analysis of the copy numbers and distribution of the repeat in the genome assembly do not corroborate. Though many copies of the repeat are present in the assembled genome, most scaffolds contain one or few copies of the repeat. The FISH results, however, show multiple clusters scattered across most chromosomes each containing high copy numbers, revealing a lack of scaffolds containing high copy numbers of **Nt_rep1** in the assembled genome. Therefore, analyses on repetitive DNA content are currently more reliable using the short sequence reads rather than the assembled genome as it underestimates repeat content. Long read sequencing methods would be able to overcome such problems with repetitive DNA, not only in *N. tenuis* but in any species, and would be better suited to analyse repetitive regions of genomes. As mentioned before, a hybrid assembly strategy combining our 10X sequencing data with long reads, obtained by e.g. Oxford Nanopore or PacBio sequencing, would presumably improve the assembly, though in this aspect for particular segments of the genome that are high in repetitive DNA. This should be kept in mind for other 10X Genomics/Supernova-derived genomes: the true number of repeats may be underestimated.

Screening of the genome and Southern blot assay suggests the absence of the ancestral insect telomere motif, (TTAGG)n, *in N. tenuis*, as the case in other species from the family Miridae (Grozeva et al., 2019; Kuznetsova et al., 2011). The telomeric motif was present in our Tandem Repeat Finder results, but in much lower numbers than expected for telomeric sequences. Additional attempts of identifying the telomeric repeat motif did not resolve this question. Three additional repeats we identified in the *N. tenuis* genome were tested via FISH, i.e. $(TATGG)_n$, $(TTGGG)_n$, and $(TCAGG)_n$, but did not localise near the ends of the chromosomes. Notably though, mapping the most abundant repeat in the genome, **Nt_rep1**, did reveal accumulation in the sub-telomeric regions of chromosomes (Figure 6). Therefore, our approach to identify potential telomere motifs, though presently unsuccessful, would presumably be effective if more repeats would be screened. In addition, a similar approach was used by Pita *et al*. (2016) in *T. infestans*, where the insect telomere motif, $(TTAGG)_n$, was successfully identified from the raw sequencing data (Pita et al., 2016). It must be noted, however, that the telomeres of *N. tenuis* might consist of different types of repeats other than short tandem repeats (as found in, for example, *Drosophila*; Traverse & Pardue, 1988) which would not be identified using Tandem Repeat Finder (Traverse and Pardue, 1988). Therefore, the identity (or even presence) of the telomeric repeat in *N. tenuis,* and by extension Miridae, remains unknown.

Beyond the Genome: Population Genomics

Pooled sequence data of ten females from the Koppert Biological Systems population were compared against the genome and provide interesting population-level effects. The overall negative *Tajima's D* would seem to indicate an abundance of rare alleles and is possible evidence of selective sweeps or population expansion, as seen in some populations *of Drosophila serrata* (Malloch) (Reddiex et al., 2018), however, this is generally true in case of values near -1 or -2. While overall negative, the absolute value of *D* in our results is small in comparison (total range from -0.89 to 0.56). To best assess the state of the commercial population, monitoring the genetic variation over time would indicate if the population is undergoing an expansion after a bottleneck ($D < 0$) or contracting ($D > 0$), whereas when $D = 0$, we assume no

selection. We can then assume that there is no selection is currently at play in the commercial population now, though it may be recovering from a recent expansion post-bottleneck, as seen in the relatively small, near zero negative *D*. In a commercial biological control rearing setting, this would make sense. Individuals are collected in relatively small numbers in the wild, reared to large numbers with unlimited resources, no predators, and no additional genetic input in the form of immigration. This initial bottleneck leads to a population expansion, which should eventually level off, as the commercial rearing is a limited space and controlled emigration, though risk of genetic drift exists (Ciosi et al., 2014). The few studies that have looked at genetic diversity within biological control populations have primarily been reduced representation analyses, such by genotyping with microsatellites (Paspati et al., 2019). Here, a pool-seq approach offers a genome-wide look at the population and can give indications of the genetic diversity of the population; this could be a useful tool for monitoring population levels efficiently.

Both genetic diversity values calculated here can also be used in population comparisons between the biological control population and wild populations. For instance, Xun *et al*. used mitochondrial and nuclear barcoding regions to haplotype 516 individuals across 37 populations into two regional groups, southwest China (SWC) and other regions in China (OC) (Xun et al., 2016). π was 0.0048 (SWC) and 0.904 (OC), while *D* was -0.112 (SWC) and -1.998 (OC). It was concluded that the SWC population was stable while, similar to the Koppert Biological Systems population here, the OC population was undergoing sudden population explosion. Pooled sequencing could be a useful tool for comparing wild Mediterranean populations to the commercial population to determine disparities in genetic variation as well as to understand the dynamics of the wild populations.

There is a concern in using PoPoolation in this context: are ten individual females sufficient for determining population variation? Here we used existing population sequence data to better utilize resources, reduced to an appropriate coverage with masked indel regions. This enabled us to show population-level impacts at the very least, which can then pave the way for further studies, with better constructed sampling methods and sample sizes;

the lack of perfect data should not preclude preliminary studies from being pursued.

Conclusion

Reported here is the genome for *N. tenuis*, a mirid that is both used throughout the Mediterranean Basin as a biological control agent and reported as a greenhouse pest in other European countries. The assembled genome is 355 Mbp in length, composed of 42,830 scaffolds with an N50 of 27,074. The goal of this project was to not only provide a genome, but also to highlight possible avenues of research now available with *N. tenuis*. A protein analysis has provided interesting prospects for mirid-specific proteins, while examples of potential LGT call for further inquiry. Putative symbionts were identified while filtering out contamination, creating a precursor for future metagenomic analysis. The cytogenetic analyses of *N. tenuis* here shed some light on Mirid cytogenetics, such as the karyotype and sex determination system, but also solicits more questions. As for the commercial population, now that there is a baseline level of genetic variation documented through our pooled sequencing, what remains to be seen is how it compares to other populations, such as other commercial populations, wild, or invasive populations. To this end, future exploration on these themes, among others, are now greatly facilitated with our release of this genome.

Acknowledgements

We would like to express special thanks to Milena Chinchilla Ramírez for her advice and *N. tenuis* expertise. Thanks to Markus Knapp (Koppert BV), Javier Calvo (Koppert BS) and Gerben Messelink (WUR) for providing *N. tenuis* specimens. Thanks to both José van de Belt, Frank Becker (WUR), and Carolina Gallego (IVIA) for their technical assistance in DNA and RNA extraction. We acknowledge the assistance of Magda Zrzavá, Anna Voleníková, Martina Flegrová, and Diogo Cabral-de-Mello (Institute of Entomology BC CAS) for their cytogenetic expertise and assistance. JHW acknowledges Sammy Cheng for assistance with the LGT pipeline. KBF acknowledges Jetske de Boer for her assistance with the flow cytometry and Joost van den Heuvel for his assistance with PoPoolation. Preliminary decontamination and annotation

was performed by GenomeScan B.V. (Leiden, NL). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated. This project was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 641456. JHW acknowledges the US-NSF-IOS-1456233 and Nathaniel & Helen Wisch Chair for funding support. The research leading to these results was partially funded by the Spanish Ministry of Economy and Competitiveness MINECO (RTA2017-00073-00-00). Cytogenetic experiments were financed by grants 17-13713S (SV and FM) and 17-17211S (MD and IP) of the Czech Science Foundation.

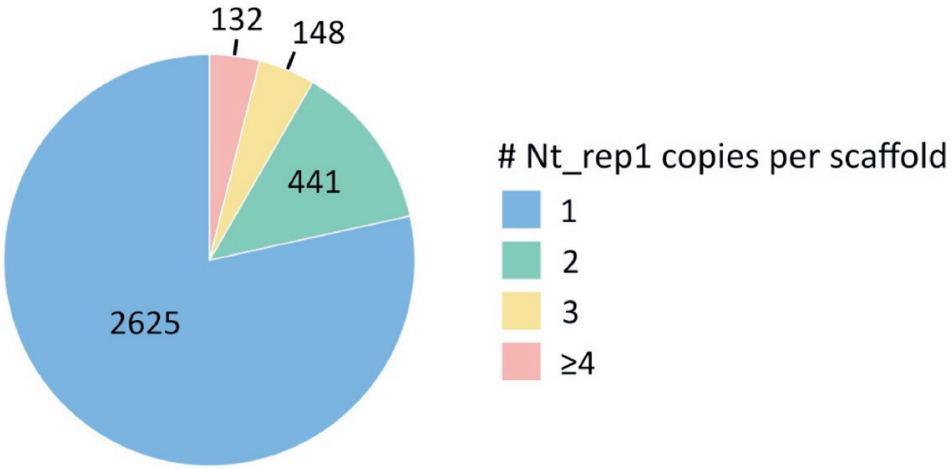
SUPPLEMENTARY MATERIAL

Most supplementary materials are deposited online. These files contain flow cytometry data, decontamination and putative LGT identification, gene lists and DAVID reports, protein sets, and pool-seq results (different sliding window sizes and indel masking steps). DANS EASY, DOI:10.17026/dans-z5z-zec9

Additional information related to online content available upon request. Additional supplementary materials are available below.



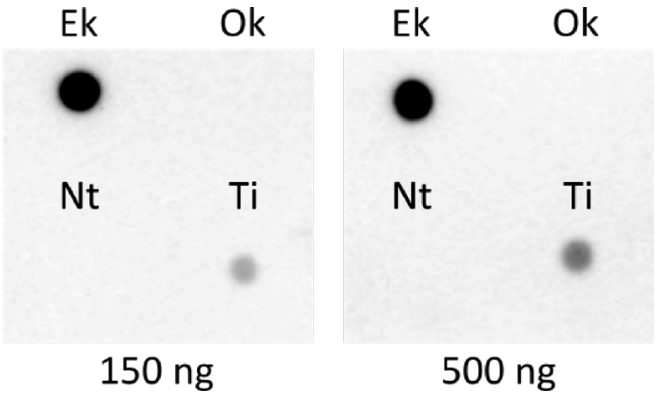
Supplementary Figure 1. *Wolbachia* detection was carried out according to the PCR protocol of Zhou *et al.* (Zhou, Rousset, and Neill 1998). Lane 1 contains a 100bp ladder, while lanes 2 and 3 contain PCR products for *Nesidiocoris tenuis* DNA (see 18S procedure). Lanes 4 and 5 contain sterilized water as a negative control, while lanes 6 and 7 contain PCR products for *Ephestia kuehniella* DNA (see Southern Blot procedure).



Supplementary Figure 2. Distribution of **Nt_rep1** on the assembled scaffolds of *Nesidiocoris tenius*. Note that most of the scaffolds containing **Nt_rep1** carry only a single copy.

Supplementary Table 1. Abundance and distribution of **Nt_rep1** in the assembled *Nesidiocoris tenuis* genome

# of *Nt_rep1* copies per scaffold	Number of scaffolds in the assembly	% of total
1	2625	78.45
2	441	13.18
3	148	4.42
4	74	2.21
5	32	0.96
6	11	0.33
7	3	0.09
8	6	0.18
9	3	0.09
11	2	0.06
17	1	0.03



Supplementary Figure 3. Southern dot blot assay using the insect telomere motif $(TTAGG)_n$ in *Ephestia kuehniella*, *Oncorhynchus keta*, *Nesidiocoris tenuis*, and *Triatoma infestans*. *E. kuehniella* (Ek) and *T. infestans* (Ti) show hybridization signals, whilst *N. tenuis* (Nt) and the negative control *O. keta* (Ok) display no detectable hybridization signals. Both quantities of DNA (150 ng and 500 ng) show comparable results.

Supplementary Table 2. Overview of cytogenetic primers used in this study.

Name	Sequence 5'-3'	Function	Reference
18S-1	CTG GTT GAT CCT GCC AGT AGT	18S rDNA partial	(Jung and Lee 2012)
18S-4	GAT CCT TCT GCA GGT TCA CC	18S rDNA partial	(Jung and Lee 2012)
"Ins_telo_F"	TAG GTT AGG TTA GGT TAG GT	Insect telomere motif	(Sahara, Marec, and Traut 1999)
"Ins_telo_R"	CTA ACC TAA CCT AAC CTA AC	Insect telomere motif	(Sahara, Marec, and Traut 1999)
Nt_pt1_F	ATG GTA TGG TAT GGT ATG GT	*N. tenuis* potential telomere motif 1	This study
Nt_pt1_R	CAT ACC ATA CCA TAC CAT AC	*N. tenuis* potential telomere motif 1	This study
Nt_pt2_F	TGG GTT GGG TTG GGT TGG GT	*N. tenuis* potential telomere motif 2	This study
Nt_pt2_R	CCA ACC CAA CCC AAC CCA AC	*N. tenuis* potential telomere motif 2	This study
Nt_pt3_F	CAG GTC AGG TCA GGT CAG GT	*N. tenuis* potential telomere motif 3	This study
Nt_pt3_R	CTG ACC TGA CCT GAC CTG AC	*N. tenuis* potential telomere motif 3	This study
Nt_rep1_F	TTC GCC CAA AAT GAA AAA ACG C	**Nt_rep1**	This study
Nt_rep1_R	TCC TGA ACA AGT GTC TGT GTG T	**Nt_rep1**	This study

Supplementary Figure 4. Alternate colouration for Figure 5 in-text. Comparative genomic hybridization (CGH) on female (a, b, c, d) and male meiotic metaphase II (e, f, g, h) chromosomes of *Nesidiocoris tenuis*. Panels (a, e) show chromosomes counterstained by DAPI (grey), panels (b, f) hybridization signals of the male derived genomic probe labelled by fluorescein (blue), panels (c, g) hybridization signals of the female derived genomic probe labelled by Cy3 (gold), and panels (d, h) merged images. (c, d) Note that the X chromosome bivalent (arrowhead) in female pachytene complement was highlighted more by female probe compared to the autosomal bivalents; (b, d) male probe labelled all chromosomes equally. (h) Two sister nuclei in meiotic metaphase II showed equal hybridization patterns of both probes on autosomes; in one of the forming nuclei, the X chromosome (arrowhead) was highlighted by female derived genomic probe (g, h) and in the second nucleus the Y chromosome (asterisk) was strongly highlighted by male derived genomic probe compared to autosomes (f, h) and less highlighted by female derived probe (g, h). (e) Note that the sex chromosomes are the biggest and most heterochromatic elements in the nucleus. Scale bar = 10µm.

Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest



Chapter 8
Synthesis

Kim B. Ferguson

The aim of this thesis was to present the genomes with associated resources of important biological control agents (BCAs), complete with context, utility, and future prospects. This was approached through an applied lens, where I focused on using existing tools and making my science as open as possible.

In **Chapter 1,** I presented two complementary lines of research – biological control and genetics – and how my thesis is an intersection of the two. I briefly covered the three types of biological control and discussed Integrated Pest Management (IPM). With those two lines of research covered, I presented the BINGO project and how my thesis project fell within it. Before moving on to my study species, I briefly mentioned the Nagoya Protocol and the Convention on Biological Diversity. In Box 2 of Chapter 1, I introduced a concept that I used as a guide for my work, the life cycle of a genome project (Papanicolaou, 2016).

In this final chapter of my anthology, I would like to address three things. First, the life cycle of a genome. It turned out that the life cycle of a genome concept was more like a set of signposts along a path to guide the way, and that bears some discussion on how helpful the concept is, and where I strayed from the path, including on making my science open (**Box 2**). Secondly, as for the Nagoya Protocol, I will dive a bit further into this legislation and its current context within biological control research, as well as the responsibility that scientists have in upholding the ideals behind it. Finally, I will discuss potential future directions from this thesis in addition to some species-specific ideas and reflect on whether this anthology has delivered on its promises.

From genetic variation to genomes

From **Chapter 2** it became clear that there is much to be done in the reporting of genetic variation as studied in BCAs. We covered the state of (genetic) affairs for species on the EPPO lists of 2016 and 2019 for registered BCAs for use in Europe, as well as an extensive list from van Lenteren and colleagues (van Lenteren, 2012; van Lenteren et al., 2018). The majority of papers that ended up in our extraction phase did not provide genetic variation estimates such as heritability or evolvability, or enough information to derive these values. Where possible, we looked for supplementary data, and for the most

part we were not able to find any that delivered additional answers. While the lack of quantifiable genetic variation values is disappointing, our review manages to provide a tremendous overview on what the current state of the art is in biological control research, as well as highlight the urgent need for, and potential areas of, growth. Related to this is the number of BCAs with publicly available genomes (**Box 1**).

Chapters 4 and 5 are written in the style of genome reports. Genome reports are intended to be short peer-reviewed articles that have become more common over the past five years as the novelty of genome papers begins to wear off (Smith, 2016). They aim to reduce the clog of the usual publication avenues and the over-emphasis on presumed "specialities" of the genomes, while still arranging peer review and eventual publication. More importantly, the amount of additional analysis is less than traditional genome papers. While it is important to showcase the utility of a genome and any analyses made possible by it, such as the heterozygosity and protein comparisons in **Chapter 4** or the inquiries into the CSD mechanism as made in **Chapter 5**, equally important is getting these genomes out into "the wild" with background and inspiration for future use by other researchers. These two chapters had different approaches to the sequencing strategy, genome assembly, decontamination protocol, and annotation method. Although the chapters are both the genome reports of parasitoid wasps, that may be the only parallel to draw between the two chapters.

Box 1 What about the genomes?

Within our systematic review (**Chapter 2**), the extraction data was listed according to BCA agent followed by species. In this regard, one question to ask, as we did, is what knowledge or research exists for heritable traits of interest in these studies. Another question, given the goals of this thesis: How many of them have genomes? In **Chapter 1**, I mentioned that at the start of the BINGO project, only two registered or classically used biological control agents had

published genomes. In the beginning of 2015, this was the parasitoid wasp *Nasonia vitripennis* and the predatory mite *Galendromis occidentalis*. By the end of 2019, this number has increased to eleven, and exempting the four presented in this thesis, this represents a 250% increase (NCBI, 2019). Notably, of these eleven species, six were represented in the final extraction list found in **Chapter 2**. How does rate of increase match up to arthropod genomes in general? Within the same time period (beginning 2015 to end 2019), arthropod genomes (first published version, not including updates) increased by 275%, from 141 to 530. So while the increase in the availability of BCA genomes is certainly impressive, its pace is in line with the increase in published arthropod genomes[4].

Evaluating the life cycle of these genomes

Referring back to **Chapter 1** and the life cycle of a genome project as proposed by Papanicolaou (Papanicolaou, 2016), each project in this thesis entered the cycle at the Design stage.

Here, it was important to figure out what quantity, quality, and length of DNA could be obtained from our prospective species. For example, the amount of DNA recovered per individual informs the necessity of pooling (which goes on to determine inbreeding, if possible), while the extraction techniques available and resulting fragment size dictate the most appropriate sequencing method (i.e. long-read technologies need long fragments). Overshadowing all of this is the budget available, as the costs can vary drastically between different sequencing strategies. Over time, it became more useful to think of this in terms of a dichotomous key, where each option has two possible

4 These are genomes accessed by the NCBI Genome Database which includes those published in the ENA, DDBJ, and NCBI. This does not include in-house, un-published genomes that are either going through the publication process, are not going to be published any time soon, or worse, will not be published at all. And while it is difficult to illustrate the absence of genomes from public databases, this is a non-zero number.

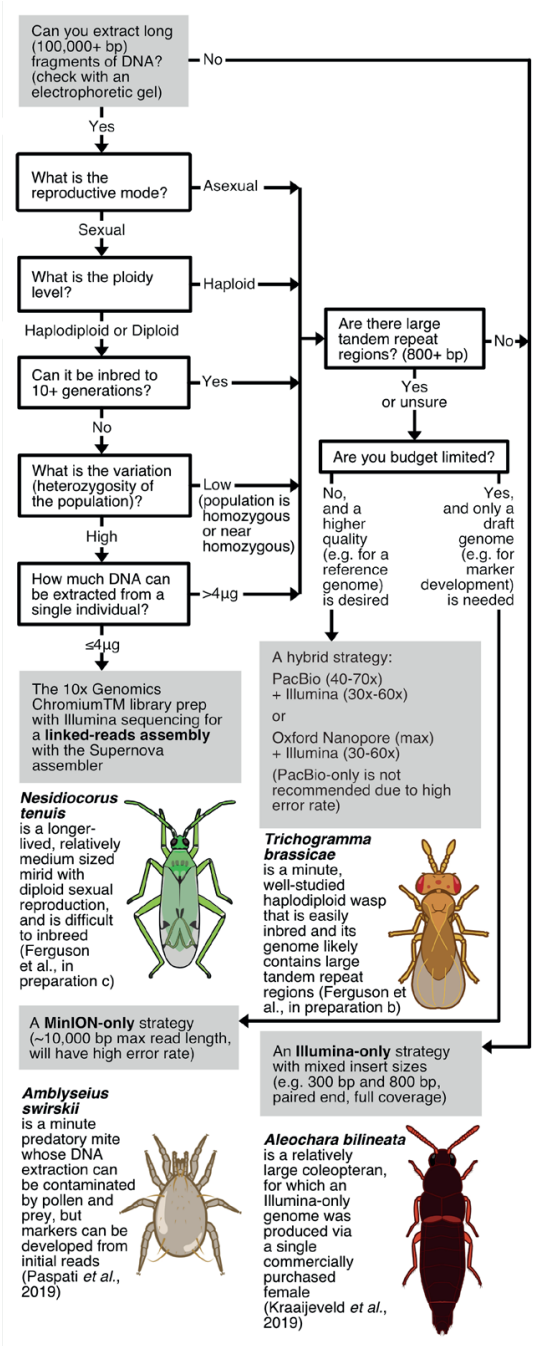


Figure 1. Sequencing strategy key, aimed specifically at sequencing arthropods in general and BCAs in particular. Here, you can imagine your arthropod of choice, and follow the flow chart, beginning in the upper left corner. Following answers along the arrows, you should arrive at the most appropriate sequencing strategy for your arthropod BCA of choice. With each strategy, there is an example, including three from this thesis (**Chapters 3, 5, and 7**), and one from a recent project (Kraaijeveld et al 2019). After Leung *et al*. 2019a.

answers, and while there is a specific order to be followed, there are multiple end points possible. This immediately reminded me of the Choose Your Own Adventure series of books and ended up inspiring the design for the first poster I prepared for this key (Ferguson et al., 2018). A few iterations later, the key is now in a recent preprint and hopefully soon-to-be-published position paper from the BINGO-ITN, the same paper that was used in **Chapter 2** to define biological control relevant traits (Leung et al., 2019a). The current key, shown here in Figure 1, begins at the upper left corner, following the decision tree until a (hopefully) optimal solution is reached. Three of the four endpoints are the genome sequencing strategies used in **Chapters 3, 5, and 7**.

Leaving the cycle early

The result of the whole-genome sequencing of *Amblyseius swirskii* (Figure 2) in **Chapter 3** was largely the outcome of when things do not go quite right in genome assembly but go fairly well in other respects. As alluded to in Figure 1, we attempted to make a MinION-only genome assembly of *A. swirskii*, following scrupulous inbreeding and doing our best to limit contamination from the pollen used as food for the mites in production. In spite of this care and attention, the benchtop sequencing yielded a small amount of results that were made even smaller following clean-up and correction. In a true "make lemonade" situation, we were able to take these lemons of sequencing results and generate microsatellites to use for population genetics of *A. swirskii*, including the commercial line and wild-caught populations that lead author and BINGO colleague, Angeliki Paspati, had obtained from field collections in Israel.

Had this been a stand-alone project, these reads might have been left in a directory folder named "Failed Run", never to see the light of day. I hope that those working with whole sequence data see this study and are encouraged to share their not-so-perfect sequencing data. That being said, I do believe that the extra step of clean-up and correction



Figure 2. At this point in the cycle, we bid farewell to *Amblyseius swirskii*, predatory mite. Credit: Jitte Groothuis

was necessary, and while the reads used for microsatellite mining are in the NCBI database, the raw data MinION data is not in the NCBI or any similar database. They are, however, publicly available and linked within the resulting publication in their pre-call format.

In terms of the genome project life cycle, either the *A. swirskii* project is stuck at the Design stage or, more likely, has exited the cycle entirely. A genome is possible for *A. swirskii*, but I would argue that it is not entirely necessary, at least for the project as it was originally intended, as the existing sequence data and resulting microsatellites as described in **Chapter 3** hold promise for future utilisation.

Some (re-)assembly required

As mentioned in **Chapter 1** and Papanicolaou's original work (Papanicolaou, 2016), genome assemblies are usually iterative phases themselves, where multiple assemblers and settings are tested for the optimal assembly. This was certainly the case for the *Trichogramma brassicae* (Figure 3) genome described in **Chapter 5**, where three different assemblers were tested: Canu (Koren et al., 2017), which used long-reads only; SPAdes (Bankevich et al., 2012), used with both long and short-read data as hybridSPAdes (Antipov et al., 2016) for a hybrid strategy; and DBG2OLC (Ye et al., 2016), in addition to the tools minimap2 (Li, 2018), Racon (Vaser et al., 2017), and Pilon (Walker et al., 2014), a hybrid pipeline that begat the final genome that went on to assembly.



Figure 3. *Trichogramma brassicae*, credit: Nina Fatouros

One aspect of genome assembly that was missing in Papanicolaou's guide to genome projects was benchmarking, though it can be found in more recent guides (Richards, 2019). Assemblies, including the *T. brassicae* genome in **Chapter 5**, are informally benchmarked according to a selection of measures, some of which are more informative than others. A key benchmark, as pointed out in reviews specific to working with insect genomes (Richards and Murali,

2015), is the BUSCO score. BUSCO, short for Benchmarking Universal Single-Copy Orthologs (Simão et al., 2015), is about genome completeness, where a reference of conserved gene orthologs (different options available based on phyla) is compared against the genome, and the amount of complete, duplicated, fragmented, or missing orthologs is returned, resulting in a BUSCO score ranging from 0 to 100% completeness. For reference, the BUSCO score for the final *T. brassicae* genome in Chapter 5, based on the insect ortholog set of 1,658 orthologs, is 95.5% complete (92.3% single copy, 3.2% duplicated). However, other assembly iterations had better BUSCO scores. The next best measures of this final assembly, such as the number of scaffolds and N50, were by far better than the other assemblies, so in this case, the BUSCO score took the back seat for informing the best genome. Now is the time to point out that while BUSCO scores are percentages, the ability to achieve a 100% BUSCO score is limited, not just by genome or transcriptome assembler technology, but also by gene finding technology (Waterhouse et al., 2018). Assembling a genome means accepting that things are incomplete. As I pointed out in **Chapter 1**, even the best of genomes should be considered a working draft.

This acceptance, however, is not possible when a key part of the genome is missing from the assembly. Such as, say, the 18S rRNA gene that is often used as a molecular marker in cytogenetic studies (Sochorová et al., 2018). At the beginning of our cytogenetic work on *Nesidiocoris tenuis* (Figure 4) in **Chapter 7**, we were able to recover the 18S sequence fragments in the raw sequence data, however, it was missing in the assembly.

The initial assembly, which is not referred to within Chapter 7 for brevity's sake, was performed on a previous version of SUPERNOVA (v1.1.5), the proprietary assembler from 10X Genomics for use with GEM linked-read data (10X Genomics Inc., Pleasanton, CA, USA). At the time, this assembler was costly third-party software, but by

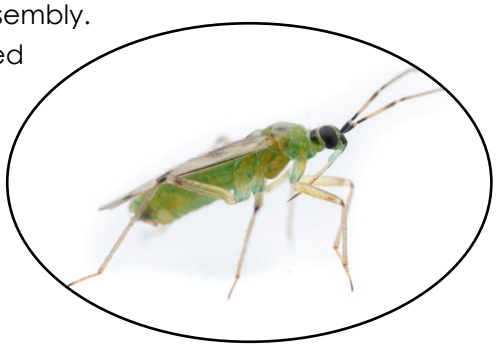


Figure 4. *Nesidiocoris tenuis*. Is it Dr. Jekyll or Mr. Hyde? Credit: Jitte Groothuis

the time it became known that the 18S was missing, a newer version was out and was freely available for download. The new version of SUPERNOVA (v2.1.1), used in Chapter 7, was able to create a more intact assembly that included the 18S rRNA gene. This was an initial set-back, but overall it led to a better assembly and annotation.

CONTAMINATION OR LGT?

Another aspect missing from the original concept of the genome project life cycle is contamination. Essentially, this is the presence of genetic material in sequence data, genomes, or annotations that do not belong to the study organism. Most often this is in reference to bacterial symbionts but can also refer to prey or environmental contamination (such as our pollen as food in the *A. swirskii* project). Although absent from the original genome project life cycle, this is a known issue within genome projects in general, and can be addressed at multiple stages of a project (Ekblom and Wolf, 2014; Richards, 2019).

Earlier genome projects relied on antibiotic treatment or the removal of the gut to avoid contamination, however, more recent advice trends towards keeping all potential symbionts and contaminants in until after sequencing, followed by either pre-assembly or post-assembly removal (Richards, 2019). Aside from the fact that complete decontamination is difficult no matter the stage at which it is performed, above all it allows for the recovery of clues towards symbiont identity, as we saw in **Chapters 5 and 7**. The *N. tenuis* genome decontamination in particular yielded a list of over 131 bacteria genera using the modified pipeline that performs both lateral gene transfer (LGT) and contaminant identification (Wheeler et al., 2013). With regards to both potential endosymbionts and possible LGT events, these lists are putative. Still, they offer intriguing possibilities for future work on *N. tenuis*, as outlined in **Chapter 7**. While Chapters 5 and 7 used post-assembly decontamination, the *Bracon brevicornis* genome (Figure 5) in **Chapter 4** relied on pre-assembly decontamination, and was also able to deliver sequences that, while untested so far, are publicly available for potential analysis for endosymbionts.

By no means does that imply that the genomes within this thesis are bacteria-



Figure 5. *Bracon brevicornis*, credit: Nils Linek

free. There will still be contamination within the genome and sequence data, but this should not be cause for alarm. Even when projects go to great lengths to control for contamination, by e.g. gut removal of to-be-sequenced specimens, a genome is not guaranteed to be 100% bacteria free (Brand et al., 2018). In addition, as we pointed out in **Chapter 7**, the area between bacterial contamination and possible LGT is grey. Moreover, in case temptation arises to blame bacteria for spurious proteins within databases, fear not, for bacterial genomes are also contaminated by human sequences and face similar issues (Breitwieser et al., 2019). Reduction of contamination is important, either directly through antibiotics or by gut removal, or indirectly through pre or post-assembly decontamination that then allows for further metagenomic/LGT analysis. At this point, the presence of bacterial contamination in publicly available genomes is no secret and should be mentioned in publications and guides, lest one goes down the rabbit hole of complete bacterial removal from a genome.

Annotation

Annotation spans over three steps of the original life cycle of a genome project: Structural annotation, functional annotation, and community curation (Papanicolaou, 2016). As the *de novo T. evanescens* genome produced in **Chapter 6** is unannotated, it has essentially left the cycle at this point, though I want to note that it is still being disseminated along with associated resources.

All of the annotated genomes within this thesis, in **Chapters 4, 5, and 7**, received structural annotation, with the genomes of *T. brassicae* and *N. tenuis* receiving a combination of *ab initio*, homology, and evidence-based annotation and the genome of *B. brevicornis* receiving only *ab initio*-based annotation. This difference in strategies can be traced back to the goals of each genome project and the resources available. However, of all the genomes, the *B.*

brevicornis genome received the most manual annotation, albeit only related to our investigations into the CSD mechanism. In comparison, the genomes of *T. brassicae* and *N. tenuis* received little manual annotation and a basic functional annotation in the form of GO term analysis.

This lack of manual and community-driven curation is in direct opposition to the advice found within Papanicolaou's work on the life cycle of a genome (Papanicolaou, 2010). And while the idea of community curation found within the life cycle of the genome project is an admirable goal and necessary for high quality genomes, Papanicolaou drew from his experience in several large genome consortia. Consortia such as the i5K Initiative, a plan to sequence 5,000 insect genomes (including an initial small pilot), are tremendous in their scope, and show what is possible when a large scientific community works towards a common goal (Evans et al., 2013; Robinson et al., 2011).

But what happens when a research community does not have the resources for a large genome sequencing project, or does not have the expertise required to take on manual annotation alone? These are questions I have dwelled on throughout this project, and it became clear that focusing on creating solid and reliable assemblies with a straightforward annotation strategy would lead to a greater likelihood of these genomes are disseminated for use within the biological control community. Here is where the i5K Initiative can lend its experience more directly: these good quality genomes and initial annotations can be taken on by members of their respective communities into the i5K workspace. As I mentioned in **Chapter 1**, it may be difficult for smaller genome teams to scale down the effort of the i5K model. But why do that in the first place, when the i5K project has already provided a server and web-based community curation platform (Poelchau et al., 2015)? This is a recent suggestion aimed at improving insect genomes (Li et al., 2019), and would allow for both the necessary improvement of the genomes presented here, while still making the resources available via the usual dissemination (i.e. all genomes are currently publicly available) in the meantime. It is my hope that some or all of the genomes presented here will be taken up by

enthusiastic researchers and make their way to the i5K server (or similar) to continue their trip through the cycle.

A new step to add to the cycle

I have taken you through the cycle as originally proposed (Papanicolaou, 2016), however, I think that there is one more step to be added: Acceptance. Once disseminated, once made publicly available via databases and given context by publications (or similar, such as preprints), genomes need to be accepted by the research communities that they are to benefit. Otherwise, their production would have been for naught. One way to ensure acceptance is to give context to the genome by making initial investigations with the generated genomic resources. This sows the seeds of possibility, whether it is selecting for a more efficient strain, improving rearing conditions, or unravelling the mechanisms behind sex determination.

As for those looking to take on a genome project, either for the first trip through the cycle or for the improvement of an existing genome, a research community needs to come together. As shown by the work within this thesis, this was a group effort, spanning multiple groups, countries, and types of research institutes. Therefore, I encourage biological control researchers out there who are considering a genome project to make it a community effort, just as we did with BINGO, if not to share the load, than at least to get a sense of what problems are out there waiting and what tools would be beneficial.

Uses for genomes for biological control

Within this thesis, we have demonstrated a large variety of investigations possible with genomes and associated genomic resources. As previously outlined, the whole genome sequencing of *A. swirskii* (**Chapter 3**) resulted in whole genome sequence data that was used to mine microsatellites. These microsatellites were used in a pooled analysis that demonstrated the relatively large amount of genetic variation found in wild-caught populations of *A. swirskii* in comparison to the commercial line. This reduced genetic variation, seen here as low heterozygosity alongside high genetic differentiation, may hamper the ability of the commercial lines to adapt to unanticipated conditions. As such, we were able to deliver recommendations

for the monitoring and rearing of *A. swirskii*, a direct application of genomic resources and population genetics for improving biological control.



Figure 6. *Trichogramma evanescens,* credit: Nina Fatouros

The *T. brassicae* sequence data from **Chapter 5** was used to generate microsatellites that also worked on sister species, *Trichogramma evanescens* (Figure 6), proving an essential tool for analysis within **Chapter 6** alongside pooled DNA sequencing (pool-seq). Our results indicate that phoresy, in this case the transportation of wasps on the legs or wings of butterflies whose eggs they will parasitize (Fatouros and Huigens, 2012), may have a larger part to play in *Trichogramma* dispersal than previously thought. Furthermore, this represents a small slice of the usefulness of the microsatellites, *de novo* genome, and pool-seq data generated in this chapter.

Speaking of pool-seq, in **Chapter 7**, we were able to make use of previously sequenced *N. tenuis* data that, while not explicitly for pool-seq analysis, yielded interesting observations on the nucleotide diversity and population dynamics within the commercial line. Chapter 7 is rich in other supporting investigations intended to suggest research avenues now possible with the genome of *N. tenuis*. Some of these investigations may not have a direct application within a biological control context, but by raising the profile of *N. tenuis* in other fields of research, such as in cytogenetics, LGT investigation, or comparative genomics, we increased the overall research profile of this mirid.

Belonging to the most often used genus of BCAs, the profile of *T. brassicae* is already quite high, which makes the lack of a genome (until now) slightly bewildering. The biological control applicability of **Chapter 5** is not immediate upon reading, though the protein comparison to sister species *T. pretiosum* offers potential for novel gene discovery and phylogenetic comparisons.

Nevertheless, the whole sequence data of *T. brassicae* helped produce the microsatellites used in **Chapter 6** and the resulting analyses are important for several biological control applications and practices.

Certainly, this thesis can be used as a guide for biological control researchers interested in genome-assisted work, both for the variety of approaches we utilised, as well as the plethora of applications presented here and alluded to with each chapter. One final recommendation, then, in addition to the guides and papers I have referred to previously, is the recently released book *Insect Genomics* (Brown and Pfrender, 2019), which would have been very useful to me in the beginning of this project, but, as the Dutch saying goes, *helaas pindakaas*[5].

Box 2 Open science and biological control genomics

One of the aims of this project was to approach it with application in mind, as well as keeping my science as open as possible. But what exactly does that mean in this framework? First, all of my presentations given at public conferences are available, with notes, on the figshare repository with a link to my ResearchGate profile. Furthermore, each chapter of this thesis (and subsequent publications) contains publicly available supplementary data important to reproducing the analyses provided here. All sequence data, genome assembly, and annotation should be online at the time of printing of this thesis, barring disaster. Finally, each manuscript resulting from this project that I am lead author on will first exist as a preprint while awaiting publication in an open access journal.

As for the analyses performed within this thesis, the majority of bioinformatic analyses, such as assembly, annotation, comparative genomics, protein comparisons, and the pool-seq analyses were

5 Translation into English: "Too bad, peanut butter"

carried out on publicly available or web-based tools, open-source wherever possible. However, I did use proprietary software as well, and this was a trade-off between my own ease of use as well as what was available within my research group. Furthermore, I did not adhere to the Fort Lauderdale Agreement for genomic resource producers (Wellcome Trust, 2003). Specifically, I did not make this data publicly available as soon as possible. This was not due to the fear of being scooped. There were some hit and misses during the sequencing and assembly stages of a few projects, and as the main researcher responsible for the data, I needed to convince myself of the quality of the data before putting it out into the world. With that in mind, the Fort Lauderdale Agreement was written in the early 2000s, and may reflect a different reality than my own: rather than a large sequencing consortium, I am one person, so I think there is some leeway on how soon "immediately" is.

The Nagoya Protocol and biological control

Previously introduced in **Chapter 1**, the Nagoya Protocol (NP) is a supplementary agreement to the United Nation's Convention on Biological Diversity (CBD). The CBD was passed in 1992 but has not been signed on to and/or ratified by every nation – the two holdouts are the United States of America, who signed it but never ratified it, and The Vatican, who did neither (CBD Secretariat, 2011; Greiber, 2019). As for signing on to the optional NP, as of November 2019, there are 120 Parties to the Nagoya Protocol, three ratified members that are not yet Parties, and 78 Non-Parties (CBD Secretariat, 2019). Members of the Parties include Argentina, China, India, Indonesia, The Netherlands, and Zambia, while Non-Parties include Australia, Brazil, Canada, Nigeria, Poland, Russia, and, as it has not signed on to the CBD, The United States of America (CBD Secretariat, 2019).

The NP aims to provide fair and equitable sharing of benefits that arise from the use of "genetic resources (GR)" (Kariyawasam and Tsai, 2018). If a company or researcher wants access to search for a GR, they are the "user" under

NP, while "providers" are the regional government or Party of where the GR is located (Kariyawasam and Tsai, 2018). Under the NP, Access Benefits and Sharing (ABS) is the intended outcome, achieved by users obtaining "prior informed consent (PIC)" from providers, followed by "mutually agreed terms (MATs)" (Kariyawasam and Tsai, 2018). Additional considerations for ABS include the use of "traditional knowledge (TK)" that belongs to the provider and is directly linked to supplying the GR for the user and its intended utilisation (Morgera et al., 2014). It is up to every Party to determine their own ABS policy and make it known, however, only eight of the 120 Parties currently have a policy on the website of The Access and Benefits-Sharing Clearing-House (ABSCH) (CBD Secretariat, 2019). These agreements are to be bilateral (between two Parties), and not multilateral agreements (between multiple parties), except in cases where PIC is untenable or in transboundary situations, such as taxonomic collections from all over the world or migratory species that span the territory of multiple Parties (Morgera et al., 2014).

A common criticism of the NP, across a variety of disciplines, is that it is vague and confusing. Suffice it to say, there are tricky waters to navigate when it comes to securing legal access to GR, including biological control agents. This uncertainty has sparked several perspectives within biological control and other fields, though many do not address a key goal of the NP. The NP is, among other things, intended to deter biopiracy and bioprospecting, and attempts to address (neo)colonialism (Dörr, 2019; Efferth et al., 2019). It is arguable that using intellectual property rights as a mechanism would even be useful to this end (Dutfield and Suthersanen, 2019), there is also concern that the NP will lead to more global inequality rather than mitigating it (Deplazes-Zemp et al., 2018). More importantly, several Indigenous peoples and groups have voiced concerns over the NP throughout the process and its ratification (Teran, 2016), as well as after the protocol was drafted (Joint Submission of Grand Council of the Crees (Eeyou Istchee) et al, 2011). Specific concerns include criticisms that the language used on access rights is in conflict with other UN legislation, such as the UN Declaration on the Rights of Indigenous Peoples (UNDRIP) while actively referencing UNDRIP as a guiding document (Morgera et al., 2014).

In the context of biological control, the intention of the NP and the aforementioned concerns are rarely discussed. Rather, there are criticisms of it being vague and leading to restrictive laws, which in turn spurs on the demands for exceptions for biological control (Barratt et al., 2018; Cock et al., 2010). When the NP was first outlined, the IOBC (International Organization for Biological Control of Noxious Animals and Plants) Global Commission on Biological Control and Access and Benefit Sharing prepared and published a position paper (Cock et al., 2010). Titled, "Do new Access and Benefit Sharing procedures under the Convention on Biological Diversity threaten the future of biological control?", the paper asserts that, while the IOBC "respects the concerns surrounding ABS and the sovereign rights of countries", "Indigenous/ traditional knowledge has not been relevant" for the development of biological control (Cock et al., 2010).

However, several examples exist where Indigenous/traditional knowledge has benefitted biological control research, such as weaver ant husbandry in the Mekong delta in Vietnam now being used in Australia on cashew plantations (Barzman et al., 1996; Peng et al., 2004), or validation research on traditional farming methods from India for use against plant pathogens (Kumar and Purohit, 2012). Reading this paper on its own, there is no mention of the intent or context behind the NP, and that ABS as outlined should have nothing to do with biological control. The single mention of Indigenous peoples is to state that traditional knowledge "should not be confused with local scientific knowledge about habitats, fauna and flora, which clearly can assist in finding appropriate locations for surveys and collections" (Cock et al., 2010). To unpack this statement, and why it is problematic, I need to briefly[6] discuss (de)colonisation and its role in academia and science (**Box 3**).

6 This is a topic that deserves far more time and perspectives to discuss than there is room in this synthesis. Several of the references used here are intended as accessible starting points for readers interested in learning more, as well as Chanda Prescod-Weinstein's Decolonizing Science Reading List (Prescod-Weinstein, 2015).

Box 3 Decolonising academia and diversifying science

The distinction between "traditional" or "Indigenous" knowledge and "scientific" knowledge is not new, but it is a false dichotomy (Green, 2008; Nicholas, 2018; Ogungbure, 2013). While there is no singular "Indigenous knowledge", one way of describing the difference is not in the truths or facts behind the knowledge, but rather the way of knowing (Nicholas, 2018). In most cases, "Science" often decides what is knowledge and what is not (Green, 2008; Ogungbure, 2013). Instead, traditional knowledge and science exist within a "diversity of knowledge" rather than as disparate entities (Green, 2008). And while incorporating one into the other is important in the context of agricultural research (Ammann, 2007), it should always be done with permission and on equitable terms, coming back to the concepts of PIC and MATs within the NP.

Furthermore, differentiating between "Indigenous knowledge" and "scientific knowledge" within discussions on access to genetic resources also fails to address the centuries of imperialism and colonialism that benefitted scientific progress (Nagtegaal and de Bruin, 1994). Colonialism helped build Western libraries, collections, and museums (Deb Roy, 2018; Nagtegaal and de Bruin, 1994)[7]. No field of science is left untouched by this truth, including genetics, where we are still coming to terms with important figures from our past and re-examining them in a different light, such as Francis Galton, geographer, statistician, biologist, and also known as the "father of eugenics" (Arney, 2019; McKie, 2019; Saini, 2019).

As for decolonising academia, there is no one-size-fits-all solution. It will include looking at the structures set up within an institution,

7 For a perspective on colonialism within the Netherlands and possible avenues to decolonising the academy, I recommend Gloria Wekker's *White Innocence: Paradoxes of Colonialism and Race* (Wekker, 2016).

how curricula are determined and taught, and the hiring processes of and existing support for students, staff, and scientists who are from underrepresented minorities; who holds power within our institutions and who does not? The process itself will not be swift, and it will never be complete (Karabinos, 2019).

Within The Netherlands, the Netherlands Institute for Advanced Study (NIAS) is opening up the discussion on decolonizing science and academia (Vince, 2019) though there has been previous reluctance in some fields about examining the trappings of colonialism (Hira, 2012). Discussions have been playing out in Dutch museums and collections as well. Recently, the Rijksmuseum entered talks with Sri Lanka and Indonesia about returning stolen artifacts, though there have been criticisms about the less-than-proactive approach by the museum (Boffey, 2019). The horticultural collections held by Leiden University Library and Naturalis Biodiversity Center, with herbaria entries and drawings from around the world including former Dutch colonies, are being digitised and then made public, though not necessarily returned (Universiteit Leiden, 2017).

Progress is being made, but more work will be required. The substantive structural changes necessary to decolonise academia should not be understated. At the institutional level, it is important to support research programs that involve departments outside of the usual realm of social studies that seek to address and dismantle colonialism alongside their science. At the University of Alberta, Kim TallBear leads the Indigenous STS (Science, Technology, and Society) research and teaching hub, where one of the projects, the Summer INternship in Genomics Canada (SING Canada), builds upon Indigenous capacity by training students and postdoctoral and community fellows in the basics of genomics alongside decolonial bioethics (Indigenous STS, 2019). Here in The

Netherlands, Utrecht University is home to Decolonisation Group, a platform recently built to engage academics, activists, archivists, and teachers together to learn about decolonisation and how to decolonise their expertise (Utrecht University Centre for Global Challenges, 2019a, 2019b).

Above all, we need to ensure that efforts to decolonise academia does not become a talking point rather than an action plan. As pointed out by Dutch researcher Michael Karabinos when discussing the recent opening of the Dutch East India Company (VOC) archives, "There are questions that any institution — such as an archive — that chooses to invoke the word 'decolonise' must ask itself. Why does it want to frame its work as decolonisation? What does it mean by the term? How far is it willing to go to achieve this goal? How often is it willing to listen, and who will it listen to? Does decolonisation have budgetary constraints?" (Karabinos, 2019).

Another aspect of decolonising academia, and science in particular, is looking at the way that universities, institutions, and researchers approach research in developing nations. Helicopter research– where foreign researchers go into a community, collect samples and data, and then leave to analyse and publish without any meaningful involvement of local communities and researchers –is no longer an option (Nordling, 2018). Projects that include stakeholders and researchers from the region, equitably and from the start, are not only more ethical and responsible, they gain an added context and promote co-creation of resources and sharing of ideas.

For example, *Gynandropsis gynandra* is a regionally important and nutritionally rich leafy green vegetable found throughout Africa, but is also an orphan crop, a term for a regional crop that

receives little research attention (Sogbohossou et al., 2018). Or rather, it was an orphan crop. As the title of her thesis "Orphan No More" suggests, E. O. Dêêdi Sogbohossou, alongside local farmers, consumers, and researchers in Europe, West Africa, and East Africa, generated genomic resources and developed breeding techniques for *G. gynandra* that has encouraged complementary research in the region (Sogbohossou, 2019). Another example involves *mabisi* and *munkoyo*, two traditionally fermented food products found in Zambia (Moonga et al., 2019; Phiri et al., 2019). Within an interdisciplinary project led by Sijmen Schoustra alongside several researchers and stakeholders in Zambia and The Netherlands, evolutionary biology, microbiology, and food science is used to describe the spontaneous fermentation processes of these regional beverages within the context of improving food security and local resiliency (Schoustra, 2019; Wageningen University and Research, 2019). The scope of the project has since been expanded, and including new beverages and local researchers and stakeholders in Benin and Zimbabwe (Schoustra, 2019; Wageningen University and Research, 2019). In both of these projects, the diversity of knowledge used includes traditional knowledge that is accessed and documented with participation and permission of the communities involved. There are direct benefits for all scientists involved, local and Europe-based, in the knowledge gained as well as for the local stakeholders in the direct application of the research within the region.

That first IOBC position paper on the Nagoya Protocol was published in 2010, and recent publications within its oversight continue to push for biological control exceptionalism (Barratt et al., 2018). I am all for criticism of the Nagoya Protocol, but it needs to be constructive and in recognition of its original goals. As pointed out earlier, it is not a perfect document in either

its consultation process or implementation. However, I believe that there is room within the field of biological control to both advocate for clear ABS guidelines and best practices as users while at the same time addressing the concerns of provider Parties and others. Beyond the NP, there are plenty of opportunities for biological control and integrated pest management (IPM) researchers to work towards stronger, more inclusive, and decolonised science. This includes the involvement of stakeholders and growers, such as those suggested in a call to bring genomics and other -omics into IPM alongside farmers in the context of cowpea pests in West Africa (Agunbiade et al., 2013). The focus of outreach is currently on science communication and increasing public knowledge. Equally important would be increasing stakeholder and local researcher participation in the framing and design of a project, not just the outcomes.

As mentioned in **Chapter 1**, there is a call from within to return to a so-called "true" IPM, where the focus is less on pesticide management and more on a whole systems approach to pest management, a sort of "conscious agriculture" (Lewis et al., 1997; van Lenteren et al., 2018). Instead, I am suggesting that we go a step further and that some IPM and biological control researchers and organizations could benefit from reframing their take on the Nagoya Protocol. Rather than seeing the Nagoya Protocol as a threat to the application of biological control, its practitioners should acknowledge the intent behind it and work to achieve it. Calling for multilateral exemptions for biological control could be a path forward, but not in poor faith by refusing to acknowledge the role of traditional knowledge in biological control and the right for self-determination of Parties to the NP and Indigenous peoples.

Future directions

The most immediate project that will be wrapped up will be going further with the work presented in Chapter 6, using haplotype data and collection-site specific information to better address long-term variation and movement between the lines, so look forward to that soon.

If I were a ~~perfectionist~~ "completionist", there are two other directions to take here: Make it a solid five genomes! First, I would re-attempt inbreeding

and sequencing of *A. swirskii,* from **Chapter 3**. I have a hunch that the latest generation of desktop sequencers, flowcells, protocols, and assemblers thereof would be able to deal with our pollen problem. Surely, it will work this time! While that pipe dream may be a bit further off then I would like, the second direction would be the annotation of the *T. evanescens* genome. It could be achieved using *ab initio* annotation methods described in the annotation of *B. brevicornis* in **Chapter 4**, or by using the annotation from *T. brassicae* in **Chapter 5**. The latter options will be much more likely than a genome of *A. swirskii*. And as stated earlier, a genome for *A. swirskii* may not be necessary at this point given the success of the microsatellites.

But beyond those two loose ends, what else remains to be tied up? From **Chapter 4**, certainly getting functional molecular genetic evidence of the *BBfem* and *BBfem1* genes would be the next step in investigating the possible mechanism of CSD in *B. brevicornis*. This would have both theoretical and applied applications, the former in the realm of insect sex determination, while the latter relates to the (in)breeding and rearing of *B. brevicornis* and other hymenopterans with a CSD mechanism. With *N. tenuis* (**Chapter 7**), the next step would be to explore the possible LGT events that we isolated by designing primers and amplifying the regions. And what is going on with those B-chromosomes? More visualisations there would be interesting in order to see if it is just in the commercial line, or if it is a feature found throughout *N. tenuis* populations.

Some final words on this anthology

In this thesis, I present a selection of biological control agents and their associated genomic resources, straight from acquisition to application. While the scope of my thesis changed over time, growing from three species to five with alterations within, the aim remained fixed: generate genomic resources, assemble genomes whenever possible, and contextualise with applications of the genome and genomic resources. And most importantly, the various resources generated in this thesis have been made publicly available; they are now common to all the initiated, to borrow from Meleager.

The goal of BINGO was to improve the production and performance of BCAs

using genetic variation, including the generation of genomic resources. Even though several of the early stage researchers (ESRs, who in the case of BINGO are all PhD candidates) are still wrapping up and writing up, the results so far speak possibilities for incorporating next generation approaches into biological control research (Koskinioti et al., 2019; Le Hesran et al., 2019a, 2019b, 2019c; Leung et al., 2019b; Lirakis et al., 2018; Lirakis and Magalhães, 2019; Paspati et al., 2019; Plouvier and Wajnberg, 2018; Ras et al., 2017; Stahl et al., 2018, 2019a, 2019b, 2019c). For each species within this thesis, I was able to work alongside colleagues who were experts and shared their knowledge with me. This provided necessary context for each genome, as well as essential feedback and ideas, and each chapter is all the better for it.

As for this thesis as an anthology, while there is value in each chapter on their own, I would hope that the value of the whole is greater than the sum of its parts. Taken together, this anthology provided various case studies of generating genomic resources for biological control agents along the necessary context to imagine what is possible with these genomes. The goal was to inspire more genome projects, no matter how far along they get, and that anything less than an annotated genome is not a failure.

Most importantly, I hope that this anthology left you with some food for thought in general, something to chew on after the book has been closed.



Curled in coils like the back of a snake, I am set here enthroned beside the last lines of [her] learned work.[8]

8 from Meleager's *Garland*, AP 12.257, sourced from http://www.attalus.org/poetry/meleager.html with number references to the original *Anthologia Palatina*. The mark to the left is a coronis, used in Greek poetry to mark the end of a work. Here, Meleager is giving it a voice, making it a character within the work as well as a functional object on the page (Gutzwiller, 1997). Edited to reflect the current author.

Appendices

REFERENCES

Abbas, N., Mansoor, M.M., Shad, S.A., Pathan, A.K., Waheed, A., Ejaz, M., Razaq, M., Zulfiqar, M.A., 2014. Fitness cost and realized heritability of resistance to spinosad in *Chrysoperla carnea* (Neuroptera: Chrysopidae). Bull. Entomol. Res. 104, 707–715. https://doi.org/10.1017/S0007485314000522

Abe, J., Pannebakker, B.A., 2016. Development of microsatellite markers and estimation of inbreeding frequency in the parasitoid wasp *Melittobia*. Nat. Publ. Gr. 1–7. https://doi.org/10.1038/srep39879

Abreo, E., Altier, N., 2019. Pangenome of *Serratia marcescens* strains from nosocomial and environmental origins reveals different populations and the links between them. Sci. Rep. 9, 1–8. https://doi.org/10.1038/s41598-018-37118-0

Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bryant, S.H., Canese, K., Church, D.M., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Johnson, M., Kelly, C., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Krasnov, S., Kuznetsov, A., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Ostell, J., O'Sullivan, C., Panchenko, A., Phan, L., Pruitt, D.P.K.D., Rubinstein, W., Sayers, E.W., Schneider, V., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., John Wilbur, W., Yaschenko, E., Zbicz, K., 2014. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 42, 8–13. https://doi.org/10.1093/nar/gkt1146

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H.C., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P. V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Michael Cherry, J., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Deslattes Mays, A., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Harley Gorrell, J., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb,

J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D.C., Scheeler, F., Shen, H., Christopher Shue, B., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Alison Yao, Q., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Craig Venter, J., 2000. The genome sequence of *Drosophila melanogaster*. Science (80-.). 287, 2185–2195. https://doi.org/10.1126/science.287.5461.2185

Adelson, G., Engell, J., Ranalli, B., Van Anglen, K.P. (Eds.), 2008. Environment: an interdisciplinary anthology. Yale University Press, New Haven.

Aguilar-Fenollosa, E., Jacas, J.A., 2014. Can we forecast the effects of climate change on entomophagous biological control agents? Pest Manag. Sci. 70, 853–859. https://doi.org/10.1002/ps.3678

Agunbiade, T., Steele, L., Coates, B., Gassmann, A., Margam, V., Ba, M., 2013. IPM-omics: from genomics to extension for integrated pest management of cowpea, in: Boukar, O., Coulibaly, O., Fatokun, C., Lopez, K., Tamò, M. (Eds.), Proceedings of the Fifth World Cowpea Conference. pp. 231–248.

Al-Khateeb, N., Asslan, L., Ibrahim, G., El-Heneidy, A., 2018. A selection index of multiple quantitative traits of the predator *Cryptolaemus montrouzieri* (Mulsant) (Coleoptera: Coccinellidae) through 12 inbreeding generations. Egypt. J. Biol. Pest Control 28. https://doi.org/10.1186/s41938-018-0049-9

Ammann, K., 2007. Reconciling Traditional Knowledge with Modern Agriculture : A Guide for Building Bridges, in: Krattiger, A., Mahoney, R.T., Nelsen, L., Thomson, J., Bennett, A., Satyanarayana, K., Graff, G., Fernandez, C., Kowalski, S. (Eds.), Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices. pp. 1539–1559.

Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F., Wingett, S., 2015. FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics. Babraham Inst.

Antipov, D., Korobeynikov, A., McLean, J.S., Pevzner, P.A., 2016. HybridSPAdes: An algorithm for hybrid assembly of short and long reads. Bioinformatics 32, 1009–1015. https://doi.org/10.1093/bioinformatics/btv688

Antolin, M.F., 1992. Sex Ratio Variation in a Parasitic Wasp II. Diallel Cross. Evolution (N. Y). 46, 1511. https://doi.org/10.2307/2409954

Antolin, M.F., Ode, P.J., Heimpel, G.E., O'Hara, R.B., Strand, M.R., 2003. Population structure, mating system, and sex-determining allele diversity of the parasitoid wasp *Habrobracon hebetor*. Heredity (Edinb). 91, 373–381. https://doi.org/10.1038/sj.hdy.6800337

Arney, K., 2019. The darkest side of genetics [WWW Document]. Genet. Genet. Soc. Pod. URL https://geneticsunzipped.com/news/2019/8/15/eugenics-the-darkest-side-of-genetics (accessed 12.11.19).

Appendices

Arnó, J., Castañé, C., Riudavets, J., Gabarra, R., 2010. Risk of damage to tomato crops by the generalist zoophytophagous predator *Nesidiocoris tenuis* (Reuter) (Hemiptera: Miridae). Bull. Entomol. Res. 100, 105–115. https://doi.org/10.1017/S0007485309006841

Astles, P.A., Moore, A.J., Preziosi, R.F., 2005. Genetic variation in response to an indirect ecological effect. Proc. R. Soc. B Biol. Sci. 272, 2577–2581. https://doi.org/10.1098/rspb.2005.3174

Athias Henriot, C., 1962. *Amblyseius swirskii*, un nouveau phytoséiide voisin d'A. andersoni (Acariens anactinotriches). Ann Ec Natl Agric Alger 1–7.

Babendreier, D., Rostas, M., Höfte, M.C.J., Kuske, S., Bigler, F., 2003. Effects of mass releases of *Trichogramma brassicae* on predatory insects in maize. Entomol. Exp. Appl. 108, 115–124. https://doi.org/10.1046/j.1570-7458.2003.00075.x

Bailly, X., Migeon, A., Navajas, M., 2004. Analysis of microsatellite variation in the spider mite pest *Tetranychus turkestani* (Acari: Tetranychidae) reveals population genetic structure and raises questions about related ecological factors. Biol. J. Linn. Soc. 82, 69–78. https://doi.org/10.1111/j.1095-8312.2004.00316.x

Baker, D.A., Loxdale, H.D., Edwards, O.R., 2003. Genetic variation and founder effects in the parasitoid wasp, *Diaeretiella rapae* (M'intosh) (Hymenoptera: Braconidae: Aphidiidae), affecting its potential as a biological control agent. Mol. Ecol. 12, 3303–3311. https://doi.org/10.1046/j.1365-294X.2003.02010.x

Balzan, M. V., Bocci, G., Moonen, A.C., 2015. Landscape complexity and field margin vegetation diversity enhance natural enemies and reduce herbivory by Lepidoptera pests on tomato crop. BioControl 61, 1–14. https://doi.org/10.1007/s10526-015-9711-2

Banaei-Moghaddam, A.M., Meier, K., Karimi-Ashtiyani, R., Houben, A., 2013. Formation and expression of pseudogenes on the B chromosome of rye. Plant Cell 25, 2536–2544. https://doi.org/10.1105/tpc.113.111856

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V, Sirotkin, A. V, Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–77. https://doi.org/10.1089/cmb.2012.0021

Barratt, B.I.P., Moran, V.C., Bigler, F., van Lenteren, J.C., 2018. The status of biological control and recommendations for improving uptake for the future. BioControl 63, 155–167. https://doi.org/10.1007/s10526-017-9831-y

Barrett, R.D.H., Schluter, D., 2008. Adaptation from standing genetic variation. Trends Ecol. Evol. 23, 38–44. https://doi.org/10.1016/J.TREE.2007.09.008

Bartlett, A.C., 2018. Maintaining genetic diversity in laboratory colonies of parasites and predators, in: Bartlett, A.C., Narang, S.K., Faust, R.M. (Eds.), Applications of Genetics to Arthropods of Biological Control Significance. CRC Press, Boca Raton, pp. 133–145. https://doi.org/10.1201/9781351069762

Barzman, M.S., Mills, N.J., Thu Cuc, N.T., 1996. Traditional knowledge and rationale for weaver ant husbandry in the Mekong delta of Vietnam. Agric. Human Values 13, 2–9. https://doi.org/10.1007/BF01530519

Beal, P., 2008. Anthology, in: A Dictionary of English Manuscript Terminology 1450–2000. Oxford University Press.

Bennett, D.M., Hoffmann, A.A., 1998. Effects of size and fluctuating asymmetry on field fitness of the parasitoid *Trichogramma carverae* (Hymenoptera: Trichogrammatidae). J. Anim. Ecol. 67, 580–591. https://doi.org/10.1046/j.1365-2656.1998.00218.x

Benoit, J.B., Adelman, Z.N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E.C., Szuter, E.M., Hagan, R.W., Gujar, H., Shukla, J.N., Zhu, F., Mohan, M., Nelson, D.R., Rosendale, A.J., Derst, C., Resnik, V., Wernig, S., Menegazzi, P., Wegener, C., Peschel, N., Hendershot, J.M., Blenau, W., Predel, R., Johnston, P.R., Ioannidis, P., Waterhouse, R.M., Nauen, R., Schorn, C., Ott, M.-C., Maiwald, F., Johnston, J.S., Gondhalekar, A.D., Scharf, M.E., Peterson, B.F., Raje, K.R., Hottel, B.A., Armisén, D., Crumière, A.J.J., Refki, P.N., Santos, M.E., Sghaier, E., Viala, S., Khila, A., Ahn, S.-J., Childers, C., Lee, C.-Y., Lin, H., Hughes, D.S.T., Duncan, E.J., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Wheeler, D., Panfilio, K.A., Vargas Jentzsch, I.M., Vargo, E.L., Booth, W., Friedrich, M., Weirauch, M.T., Anderson, M.A.E., Jones, J.W., Mittapalli, O., Zhao, C., Zhou, J.-J., Evans, J.D., Attardo, G.M., Robertson, H.M., Zdobnov, E.M., Ribeiro, J.M.C., Gibbs, R.A., Werren, J.H., Palli, S.R., Schal, C., Richards, S., 2016. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat. Commun. 7, 10165. https://doi.org/10.1038/ncomms10165

Benson, G., 1999. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580. https://doi.org/10.1093/nar/27.2.573

Bertin, A., Pavinato, V.A.C., Parra, J.R.P., 2017. Fitness-related changes in laboratory populations of the egg parasitoid *Trichogramma galloi* and the implications of rearing on factitious hosts. BioControl. https://doi.org/10.1007/s10526-017-9795-y

Beye, M., Hasselmann, M., Fondrk, M.K., Page, R.E., Omholt, S.W., 2003. The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114, 419–429. https://doi.org/10.1016/S0092-8674(03)00606-8

Bijlsma, R., Bundgaard, J., Boerema, A.C., 2000. Does inbreeding affect the extinction risk of small populations?: predictions from Drosophila. J. Evol. Biol. 13, 502–514. https://doi.org/10.1046/j.1420-9101.2000.00177.x

Bilodeau, E., Simon, J.C., Guay, J.F., Turgeon, J., Cloutier, C., 2013. Does variation in host plant association and symbiont infection of pea aphid populations induce genetic and behaviour differentiation of its main parasitoid, *Aphidius ervi*? Evol. Ecol. 27, 165–184. https://doi.org/10.1007/s10682-012-9577-z

BINGO-ITN, 2015. BINGO-ITN [WWW Document]. BINGO-ITN. URL https://www.bingo-itn.eu/en/bingo.htm (accessed 12.14.19).

Biondi, A., Zappalà, L., Di Mauro, A., Tropea Garzia, G., Russo, A., Desneux, N., Siscaro, G.,

2015. Can alternative host plant and prey affect phytophagy and biological control by the zoophytophagous mirid *Nesidiocoris tenuis*? BioControl 79–90. https://doi.org/10.1007/s10526-015-9700-5

Blankenfeldt, W., Kuzin, A.P., Skarina, T., Korniyenko, Y., Tong, L., Bayer, P., Janning, P., Thomashow, L.S., Mavrodi, D. V., 2004. Structure and function of the phenazine biosynthetic protein *PhzF* from *Pseudomonas fluorescens*. Proc. Natl. Acad. Sci. U. S. A. 101, 16431–16436. https://doi.org/10.1073/pnas.0407371101

Boffey, D., 2019. Rijksmuseum laments Dutch failure to return stolen colonial art. Guard.

Bolckmans, K., YM Van Houten - US Patent 8, 097,248, 2012, undefined, n.d. Mite composition, use thereof, method for rearing the phytoseiid predatory mite *Amblyseius swirskii*, rearing system for rearing said phytoseiid mite and methods for. Google Patents.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouagga, S., Urbaneja, A., Depalo, L., Rubio, L., Pérez-Hedo, M., 2019. Zoophytophagous predator-induced defences restrict accumulation of the tomato spotted wilt virus. Pest Manag. Sci. ps.5547. https://doi.org/10.1002/ps.5547

Boulétreau, M., Wajnberg, E., 1986. Comparative responses of two sympatric parasitoid cynipids to the genetic and epigenetic variations of the larvae of their host, *Drosophila melanogaster*. Entomol. Exp. Appl. 41, 107–114. https://doi.org/10.1111/j.1570-7458.1986.tb00516.x

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A., 2008. UniProtKB/Swiss-Prot: The manually annotated section of the UniProt KnowledgeBase. Methods Mol. Biol. 406, 89–112. https://doi.org/10.1007/978-1-59745-535-0

Boyd, B.M., Allen, J.M., Koga, R., Fukatsu, T., Sweet, A.D., Johnson, K.P., Reed, D.L., 2016. Two Bacterial Genera, *Sodalis* and *Rickettsia*, Associated with the Seal Louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). Appl. Environ. Microbiol. 82, 3185–3197. https://doi.org/10.1128/AEM.00282-16

Brand, P., Lin, W., Johnson, B.R., 2018. The draft genome of the invasive walking stick, *Medauroidea extradendata*, reveals extensive lineage-specific gene family expansions of cell wall degrading Enzymes in Phasmatodea. G3 Genes, Genomes, Genet. 8, 1403–1408. https://doi.org/10.1534/g3.118.200204

Breitwieser, F.P., Pertea, M., Zimin, A. V., Salzberg, S.L., 2019. Human contamination in bacterial genomes has created thousands of spurious proteins. Genome Res. 29, 954–960. https://doi.org/10.1101/gr.245373.118

Brotherstone, S., Goddard, M., 2005. Artificial selection and maintenance of genetic variance in the global dairy cow population. Philos. Trans. R. Soc. B Biol. Sci. 360, 1479–1488. https://doi.org/10.1098/rstb.2005.1668

Brown, S.J., Pfrender, M.E. (Eds.), 2019. Insect Genomics, Methods in Molecular Biology. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4939-8775-7

Bruins, E.B.A.W., Wajnberg, E., Pak, G.A., 1994. Genetic variability in the reactive distance in *Trichogramma brassicae* after automatic tracking of the walking path. Entomol. Exp. Appl. 72, 297–303. https://doi.org/10.1111/j.1570-7458.1994.tb01830.x

Bruton, B.D., Mitchell, F., Fletcher, J., Pair, S.D., Wayadande, A., Melcher, U., Brady, J., Bextine, B., Popham, T.W., 2007. *Serratia marcescens*, a Phloem-Colonizing, Squash Bug -Transmitted Bacterium: Causal Agent of Cucurbit Yellow Vine Disease. Plant Dis. 87, 937–944. https://doi.org/10.1094/pdis.2003.87.8.937

Buitenhuis, R., Shipp, L., Scott-Dupree, C., Brommit, A., Lee, W., 2014. Host plant effects on the behaviour and performance of *Amblyseius swirskii* (Acari: Phytoseiidae). Exp. Appl. Acarol. 62, 171–180. https://doi.org/10.1007/s10493-013-9735-1

Bush, G.L., Neck, R.W., 1976. Ecological Genetics of the Screwworm Fly, *Cochliomyia hominivorax* (Diptera: Calliphoridae) and Its Bearing on the Quality Control of Mass-reared Insects. Environ. Entomol. 5, 821–826. https://doi.org/10.1093/ee/5.5.821

Bustin, M., Misteli, T., 2016. Nongenetic functions of the genome. Science 352, 671, aad6933. https://doi.org/10.1126/science.aad6933

Caltagirone, L.E., Doutt, R.L., 1989. The history of the vedalia beetle importation to California and its impact on the development of biological control. Annu. Rev. Entomol. 34, 1–16. https://doi.org/http://dx.doi.org/10.1146/annurev.en.34.010189.000245

Calvo, F.J., Bolckmans, K., Belda, J.E., 2011. Control of *Bemisia tabaci* and *Frankliniella occidentalis* in cucumber by *Amblyseius swirskii*. BioControl. https://doi.org/10.1007/s10526-010-9319-5

Calvo, J.F., Bolckmans, K., Stansly, P. a., Urbaneja, A., 2009. Predation by *Nesidiocoris tenuis* on *Bemisia tabaci* and injury to tomato. BioControl 54, 237–246. https://doi.org/10.1007/s10526-008-9164-y

Calvo, J.F., Knapp, M., van Houten, Y.M., Hoogerbrugge, H., Belda, J.E., 2015. *Amblyseius swirskii*: What made this predatory mite such a successful biocontrol agent? Exp. Appl. Acarol. 65, 419–433. https://doi.org/10.1007/s10493-014-9873-0

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Camacho, J.P.M., Sharbel, T.F., Beukeboom, L.W., 2000. B-chromosome evolution. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 355, 163–178. https://doi.org/10.1098/rstb.2000.0556

Cane Toad (*Bufo marinus*) [WWW Document], 2019. Colnect.com. URL https://colnect.com/en/stamps/stamp/336916-Cane_Toad_Bufo_marinus-Wildlife_Preservation-Barbados (accessed 11.27.19).

Carabajal Paladino, L.Z., Provazníková, I., Berger, M., Bass, C., Aratchige, N.S., López, S.N., Marec, F., Nguyen, P., 2019. Sex Chromosome Turnover in Moths of the Diverse Superfamily Gelechioidea. Genome Biol. Evol. 11, 1307–1319. https://doi.org/10.1093/gbe/evz075

Carbonnelle, S., Hance, T., Migeon, A., Baret, P., Cros-Arteil, S., Navajas, M., 2007. Microsatellite

markers reveal spatial genetic structure of *Tetranychus urticae* (Acari: Tetranychidae) populations along a latitudinal gradient in Europe. Exp. Appl. Acarol. 41, 225–241. https://doi.org/10.1007/s10493-007-9068-z

Cardinal Ladybird (*Rodalia cardinalis*) [WWW Document], 2019. Colnect.com. URL https://colnect.com/en/stamps/stamp/137993-Cardinal_Ladybird_Rodolia_cardinalis-Bugs_and_Butterflies-Australia (accessed 11.27.19).

Caspi-Fluger, A., Inbar, M., Steinberg, S., Friedmann, Y., Freund, M., Mozes-Daube, N., Zchori-Fein, E., 2014. Characterization of the symbiont Rickettsia in the mirid bug *Nesidiocoris tenuis* (Reuter) (Heteroptera: Miridae). Bull. Entomol. Res. 104, 681–688. https://doi.org/10.1017/S0007485314000492

Castañé, C., Arnó, J., Gabarra, R., Alomar, O., 2011. Plant damage to vegetable crops by zoophytophagous mirid predators. Biol. Control 59, 22–29. https://doi.org/10.1016/j.biocontrol.2011.03.007

Cavalli-Sforza, L.L., Edwards, A.W.F., 1967. Phylogenetic analysis: models and estimation procedures. Evolution (N. Y). 21, 550–570. https://doi.org/10.1111/j.1558-5646.1967.tb03411.x

CBD Secretariat, 2019. ABSCH: The Access and Benefit-Sharing Clearing-House [WWW Document]. URL https://absch.cbd.int/

CBD Secretariat, 2011a. About the Nagoya Protocol [WWW Document]. URL https://www.cbd.int/abs/about/ (accessed 11.14.19).

CBD Secretariat, 2011b. Country Profiles [WWW Document]. URL https://www.cbd.int/countries/ (accessed 11.29.19).

Chakraborty, M., Baldwin-Brown, J.G., Long, A.D., Emerson, J.J., 2016. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 44, 1–12. https://doi.org/10.1093/nar/gkw654

Chang, S., Puryear, J., Cairney, J., 1993. A simple and efficient method for isolating RNA from pine trees. Plant Mol. Biol. Report. 11, 113–116. https://doi.org/10.1007/BF02670468

Chapman, A. V., Kuhar, T.P., Schultz, P.B., Brewster, C.C., 2009. Dispersal of *Trichogramma ostriniae* (Hymenoptera: Trichogrammatidae) in Potato Fields. Environ. Entomol. 38, 677–685. https://doi.org/10.1603/022.038.0319

Charlesworth, B., Charlesworth, D., 2000. The degeneration of Y chromosomes. Philos. Trans. R. Soc. B Biol. Sci. 355, 1563–1572. https://doi.org/10.1098/rstb.2000.0717

Chen, X., van Achterberg, C., 2019. Systematics, Phylogeny, and Evolution of Braconid Wasps: 30 Years of Progress. Annu. Rev. Entomol. 64, 335–358. https://doi.org/10.1146/annurev-ento-011118-111856

Childers, A.K., 2019. i5K, Sequenced Insect Genomes [WWW Document]. URL http://i5k.github.io/arthropod_genomes_at_ncbi (accessed 11.14.19).

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo,

C., Ecker, J.R., Cantu, D., Rank, D.R., Schatz, M.C., 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13, 1050–1054. https://doi.org/10.1038/nmeth.4035

Ciosi, M., Masiga, D.K., Turner, C.M.R., 2014. Laboratory Colonisation and Genetic Bottlenecks in the Tsetse Fly *Glossina pallidipes*. PLoS Negl. Trop. Dis. 8. https://doi.org/10.1371/journal.pntd.0002697

Clark, A.M., Bertrand, H.A., Smith, R.E., 1963. Life Span Differences between Haploid and Diploid Males of *Habrobracon serinopae* after Exposure as Adults to X Rays. Am. Nat. 97, 203–208. https://doi.org/10.1086/282271

Cock, M.J.W., Lenteren, J.C. Van, Brodeur, J., Haas, F., Mason, P.G., van Lenteren, J.C., Brodeur, J., Barratt, B.I.P., Bigler, F., Bolckmans, K., Cônsoli, F.L., Haas, F., Mason, P.G., Parra, J.R.P., 2010. Do new Access and Benefit Sharing procedures under the Convention on Biological Diversity threaten the future of biological control? BioControl 55, 199–218. https://doi.org/10.1007/s10526-009-9234-9

Cook, J.M., Crozier, R.H., 1995. Sex determination and population biology in the hymenoptera. Trends Ecol. Evol. 10, 281–286. https://doi.org/10.1016/0169-5347(95)90011-X

Costechareyre, D., Balmand, S., Condemine, G., Rahbé, Y., 2012. *Dickeya dadantii*, a plant pathogenic bacterium producing cyt-like entomotoxins, causes septicemia in the pea aphid *Acyrthosiphon pisum*. PLoS One 7. https://doi.org/10.1371/journal.pone.0030702

Côté, I.M., Jennions, M.D., 2013. The procedure of meta-analysis in a nutshell, in: Koricheva, J., Gurevitch, J., Mengersen, K. (Eds.), Handbook of Meta-Analysis in Ecology and Evolution. Princeton University Press, Oxford, pp. 14–24. https://doi.org/10.1515/9781400846184-004

Crooijmans, R.P.M.A., Groen, A.F., Van Kampen, A.J.A., Van Der Beek, S., Van Der Poel, J.J., Groenen, M.A.M., 1996. Microsatellite Polymorphism in Commercial Broiler and Layer Lines Estimated Using Pooled Blood Samples. Poult. Sci. 75, 904–909. https://doi.org/10.3382/ps.0750904

Cruaud, A., Groussier, G., Genson, G., Sauné, L., Polaszek, A., Rasplus, J.-Y., 2018. Pushing the limits of whole genome amplification: successful sequencing of RADseq library from a single microhymenopteran (Chalcidoidea, *Trichogramma*). PeerJ 6, e5640. https://doi.org/10.7717/peerj.5640

Cusumano, A., Weldegergis, B.T., Colazza, S., Dicke, M., Fatouros, N.E., 2015. Attraction of egg-killing parasitoids toward induced plant volatiles in a multi-herbivore context. Oecologia 179, 163–174. https://doi.org/10.1007/s00442-015-3325-3

Dai, X., Xun, H., Chang, J., Zhang, J., Hu, B., Li, H., Yuan, X., Cai, W., 2012. The complete mitochondrial genome of the plant bug *Nesidiocoris tenuis* (Reuter) (Hemiptera: Miridae: Bryocorinae: Dicyphini). Zootaxa 3554, 30–44. https://doi.org/10.11646/zootaxa.3554.1.2

Dalíková, M., Zrzavá, M., Hladová, I., Nguyen, P., Šonský, I., Flegrová, M., Kubíčková, S., Voleníková, A., Kawahara, A.Y., Peters, R.S., Marec, F., Sayres, M.W., 2017a. New Insights into the Evolution of the W Chromosome in Lepidoptera. J. Hered. 108, 709–719. https://doi.org/10.1093/jhered/esx063

Dalíková, M., Zrzavá, M., Kubíčková, S., Marec, F., 2017b. W-enriched satellite sequence in the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Pyralidae). Chromosom. Res. 25, 241–252. https://doi.org/10.1007/s10577-017-9558-8

Das, D.K., Singh, J., Vennila, S., 2011. Emerging Crop Pest Scenario under the Impact of Climate Change – A Brief Review. J. Agric. Phys. 11, 13–20. https://doi.org/ISSN 0973-032X

De Boer, J.G., Ode, P.J., Vet, L.E.M., Whitfield, J.B., Heimpel, G.E., 2007. Diploid males sire triploid daughters and sons in the parasitoid wasp *Cotesia vestalis*. Heredity (Edinb). 99, 288–294. https://doi.org/10.1038/sj.hdy.6800995

Deb Roy, R., 2018. Decolonise science - time to end another imperial era [WWW Document]. Conversat. URL https://theconversation.com/decolonise-science-time-to-end-another-imperial-era-89189/?xid=PS_smithsonian (accessed 12.11.19).

Dechaud, C., Volff, J.N., Schartl, M., Naville, M., 2019. Sex and the TEs: Transposable elements in sexual development and function in animals. Mob. DNA 10, 42. https://doi.org/10.1186/s13100-019-0185-0

Delpuech, J.M., Delahaye, M., 2013. The sublethal effects of deltamethrin on *Trichogramma behaviors* during the exploitation of host patches. Sci. Total Environ. 447, 274–279. https://doi.org/10.1016/j.scitotenv.2012.12.096

Deplazes-Zemp, A., Abiven, S., Schaber, P., Schaepman, M., Schaepman-Strub, G., Schmid, B., Shimizu, K.K., Altermatt, F., 2018. The Nagoya Protocol could backfire on the Global South. Nat. Ecol. Evol. 2, 917–919. https://doi.org/10.1038/s41559-018-0561-z

Donadio, R., 2006. Keeper of the Canon. New York Times 7007027.

Doody, J.S., Green, B., Rhind, D., Castellano, C.M., Sims, R., Robinson, T., 2009. Population-level declines in Australian predators caused by an invasive species. Anim. Conserv. 12, 46–53. https://doi.org/10.1111/j.1469-1795.2008.00219.x

Dörr, D., 2019. Biopiracy and the right to self-determination of indigenous peoples. Phytomedicine 53, 308–312. https://doi.org/10.1016/j.phymed.2018.10.019

Dos Santos, G., Schroeder, A.J., Goodman, J.L., Strelets, V.B., Crosby, M.A., Thurmond, J., Emmert, D.B., Gelbart, W.M., Brown, N.H., Kaufman, T., Werner-Washburne, M., Cripps, R., Broll, K., Gramates, L.S., Falls, K., Matthews, B.B., Russo, S., Zhou, P., Zytkovicz, M., Adryan, B., Attrill, H., Costa, M., Marygold, S., McQuilton, P., Millburn, G., Ponting, L., Stefancsik, R., Tweedie, S., Grumbling, G., 2015. FlyBase: Introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. 43, D690–D697. https://doi.org/10.1093/nar/gku1099

Doyle, JJ, Doyle, JL, 1990. Isolation of plant DNA from fresh tissue. Focus (Madison). 12, 13–15.

Dray, S., Dufour, A.-B., Chessel, D., 2007. The ade4 Package - II: Two-Table and K-Table Methods. R News 7, 47–52.

Dunning Hotopp, J.C., Clark, M.E., Oliveira, D.C.S.G., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R. V., Shepard, J., Tomkins, J., Richards, S., Spiro, D.J., Ghedin, E., Slatko, B.E., Tettelin, H., Werren, J.H., 2007.

Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science (80-.). 317, 1753–1756. https://doi.org/10.1126/science.1142490

Dutfield, G., Suthersanen, U., 2019. Traditional Knowledge and Genetic Resources: Observing Legal Protection through the Lens of Historical Geography and Human Rights. Washburn Leg. J. 58, 399–448.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Easteal, S., 1981. The history of introductions of *Bufo marinus* (Amphibia: Anura); a natural experiment in evolution. Biol. J. Linn. Soc. 16, 93–113. https://doi.org/10.1111/j.1095-8312.1981.tb01645.x

Efferth, T., Banerjee, M., Abu-Darwish, M.S., Abdelfatah, S., Böckers, M., Bhakta-Guha, D., Bolzani, V., Daak, S., Demirezer, Ö.L., Dawood, M., Efferth, M., El-Seedi, H.R., Fischer, N., Greten, H.J., Hamdoun, S., Hong, C., Horneber, M., Kadioglu, O., Khalid, H.E., Khalid, S.A., Kuete, V., Mahmoud, N., Marin, J., Mbaveng, A., Midiwo, J., Nakagawa, H., Naß, J., Ngassapa, O., Ochwang'i, D., Omosa, L.K., Ooko, E.A., Özenver, N., Poornima, P., Romero, M.R., Saeed, M.E.M., Salgueiro, L., Seo, E.-J., Yan, G., Yasin, Z., Saeed, E.M., Paul, N.W., 2019. Biopiracy versus One-World Medicine–From colonial relicts to global collaborative concepts. Phytomedicine 53, 319–331. https://doi.org/10.1016/j.phymed.2018.06.007

Ehler, L.E., 2006. Integrated pest management (IPM): Definition, historical development and implementation, and the other IPM. Pest Manag. Sci. 62, 787–789. https://doi.org/10.1002/ps.1247

Ekblom, R., Wolf, J.B.W., 2014. A field guide to whole-genome sequencing, assembly and annotation. Evol. Appl. 1026–1042. https://doi.org/10.1111/eva.12178

El-Heneidy, A.H., Shoeb, M.A., 2007. Comparative biological aspects of two strains from the egg parasitoid, *Trichogramma evanescens* Westwood (Hymenoptera: Trichogrammatidae) in Egypt. Egypt. J. Biol. Pest Control 17, 99–106.

Ellegren, H., 2014. Genome sequencing and population genomics in non-model organisms. Trends Ecol. Evol. 29, 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Evans, J.D., Brown, S.J., Hackett, K.J.J., Robinson, G., Richards, S., Lawson, D., Elsik, C., Coddington, J., Edwards, O., Emrich, S., Gabaldon, T., Goldsmith, M., Hanes, G., Misof, B., Muñoz-Torres, M., Niehuis, O., Papanicolaou, A., Pfrender, M., Poelchau, M., Purcell-Miramontes, M., Robertson, H.M., Ryder, O., Tagu, D., Torres, T., Zdobnov, E., Zhang, G., Zhou, X., 2013. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. J. Hered. 104, 595–600. https://doi.org/10.1093/jhered/est050

Facon, B., Crespin, L., Loiseau, A., Lombaert, E., Magro, A., Estoup, A., 2011. Can things

get worse when an invasive species hybridizes? The harlequin ladybird *Harmonia axyridis* in France as a case study. Evol. Appl. 4, 71–88. https://doi.org/10.1111/j.1752-4571.2010.00134.x

Fagerberg, A.J., Fulton, R.E., Black Iv, W.C., 2001. Microsatellite loci are not abundant in all arthropod genomes: analyses in the hard tick, *Ixodes scapularis* and the yellow fever mosquito, *Aedes aegypti*. Insect Mol. Biol. 10, 225–236. https://doi.org/10.1046/j.1365-2583.2001.00260.x

Faircloth, B.C., 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol. Ecol. Resour. 8, 92–94. https://doi.org/10.1111/j.1471-8286.2007.01884.x

Farrokhi, S., Ashouri, A., Shirazi, J., Allahvari, H., Huigens, M.E.E., 2010. A comparative study on the functional response of *Wolbachia*-infected and uninfected forms of the parasitoid wasp *Trichogramma brassicae*. J. Insect Sci. 10, 167. https://doi.org/10.1673/031.010.14127

Fatouros, N.E., Bukovinszkine'Kiss, G., Dicke, M., Hilker, M., 2007. The Response Specificity of *Trichogramma* Egg Parasitoids towards Infochemicals during Host Location. J. Insect Behav. 20, 53–65. https://doi.org/10.1007/s10905-006-9062-z

Fatouros, N.E., Bukovinszkine'Kiss, G., Kalkers, L.A., Gamborena, R.S., Dicke, M., Hilker, M., 2005a. Oviposition-induced plant cues: Do they arrest *Trichogramma* wasps during host location? Entomol. Exp. Appl. 115, 207–215. https://doi.org/10.1111/j.1570-7458.2005.00245.x

Fatouros, N.E., Huigens, M.E., 2012. Phoresy in the field: natural occurrence of *Trichogramma* egg parasitoids on butterflies and moths. BioControl 57, 493–502. https://doi.org/10.1007/s10526-011-9427-x

Fatouros, N.E., Huigens, M.E., van Loon, J.J.A., Dicke, M., Hilker, M., 2005b. Butterfly anti-aphrodisiac lures parasitic wasps. Nature 433, 704–704. https://doi.org/10.1038/433704a

Fatouros, N.E., Pineda, A., Huigens, M.E., Broekgaarden, C., Shimwela, M.M., Candia, I.A.F., Verbaarschot, P., Bukovinszky, T., 2014. Synergistic effects of direct and indirect defences on herbivore egg survival in a wild crucifer.

Feditchkina Tracy, E., 2014. The promise of biological control for sustainable agriculture: a stakeholder-based analysis. J. Sci. Policy Gov. 5, 1–13.

Ferguson, K.B., Verhulst, E.C., Zwaan, B.J., Pannebakker, B.A., 2018. Choose Your Own Insect Genome Adventure. https://doi.org/10.6084/m9.figshare.5777769.v2

Ferguson, K.B., 2018. *Amblyseius swirskii* I2 ONT Minion Raw and Uncorrected Data. https://doi.org/https://doi.org/10.6084/m9.figshare.c.3699148.v1

Fleury, F., Allemand, R., Fouillet, P., Boulétreau, M., 1995. Genetic variation in locomotor activity rhythm among populations of *Leptopilina heterotoma* (Hymenoptera: Eucoilidae), a larval parasitoid of *Drosophila* species. Behav. Genet. 25, 81–89. https://doi.org/10.1007/BF02197245

Food and Agriculture Organisation of the United Nations, 2019. AGP - Integrated Pest Management [WWW Document]. URL http://www.fao.org/agriculture/crops/core-

themes/theme/pests/ipm/en/ (accessed 11.14.19).

Fournier, F., Boivin, G., 2000. Comparative Dispersal of *Trichogramma evanescens* and *Trichogramma pretiosum* (Hymenoptera: Trichogrammatidae) in Relation to Environmental Conditions. Environ. Entomol. 29, 55–63. https://doi.org/10.1603/0046-225X-29.1.55

Francuski, L., Djurakic, M., Ludoški, J., Hurtado, P., Pérez-Bañón, C., Ståhls, G., Rojo, S., Milankov, V., 2014. Shift in Phenotypic Variation Coupled With Rapid Loss of Genetic Diversity in Captive Populations of *Eristalis tenax* (Diptera: Syrphidae): Consequences for Rearing and Potential Commercial Use. J. Econ. Entomol. 107, 821–832. https://doi.org/10.1603/EC13243

Freitas, F.C., Morales-Corrêa e Castro, A.C., Barbosa, N.C.C.P., Fernandes, O.A., 2018. Characterization and Comparison of Genetic Variation in *Cotesia flavipes* (Hymenoptera: Braconidae) Mass Reared for Biological Pest Control Using Microsatellite Markers. Neotrop. Entomol. 47, 433–439. https://doi.org/10.1007/s13744-017-0531-1

Frydrychová, R., Grossmann, P., Trubac, P., Vítková, M., Marec, F., 2004. Phylogenetic distribution of TTAGG telomeric repeats in insects. Genome 47, 163–178. https://doi.org/10.1139/g03-100

Fukunaga, Y., Akimoto, S.I., 2007. Toxicity of the aphid *Aulacorthum magnoliae* to the predator *Harmonia axyridis* (Coleoptera: Coccinellidae) and genetic variance in the assimilation of the toxic aphids in *H. axyridis* larvae: Original article. Entomol. Sci. 10, 45–53. https://doi.org/10.1111/j.1479-8298.2006.00197.x

Futschik, A., Schlötterer, C., 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics. https://doi.org/10.1534/genetics.110.114397

Garantonakis, N., Pappas, M.L., Varikou, K., Skiada, V., Broufas, G.D., Kavroulakis, N., Papadopoulou, K.K., 2018. Tomato Inoculation With the Endophytic Strain *Fusarium solani* K Results in Reduced Feeding Damage by the Zoophytophagous Predator *Nesidiocoris tenuis*. Front. Ecol. Evol. 6, 1–7. https://doi.org/10.3389/fevo.2018.00126

Garcia-Gonzalez, F., Simmons, L.W., Tomkins, J.L., Kotiaho, J.S., Evans, J.P., 2012. Comparing evolvabilities: Common errors surrounding the calculation and use of coefficients of additive genetic variation. Evolution (N. Y). 66, 2341–2349. https://doi.org/10.1111/j.1558-5646.2011.01565.x

Gatto, K.P., Mattos, J. V., Seger, K.R., Lourenço, L.B., 2018. Sex chromosome differentiation in the frog genus *Pseudis* involves satellite DNA and chromosome rearrangements. Front. Genet. 9, 1–12. https://doi.org/10.3389/fgene.2018.00301

Gebiola, M., Streicher, J.W., Rugman-Jones, P.F., Morse, J.G., Stouthamer, R., 2019. Genome-wide analyses of single nucleotide polymorphisms reveal the consequences of traditional mass-rearing on genetic variation in *Aphytis melinus* (Hymenoptera: Aphelinidae): the danger of putting all eggs in one basket. Pest Manag. Sci. https://doi.org/10.1002/ps.5427

Geden, C.J., Smith, L., Long, S.J., Rutz, D.A., 1992. Rapid Deterioration of Searching Behavior, Host Destruction, and Fecundity of the Parasitoid *Muscidifurax raptor* (Hymenoptera:

Pteromalidae) in Culture. Ann. Entomol. Soc. Am. 85, 179–187. https://doi.org/10.1093/aesa/85.2.179

Geib, S.M., Liang, G.H., Murphy, T.D., Sim, S.B., 2017. Whole Genome Sequencing of the Braconid Parasitoid Wasp *Fopius arisanus*, an Important Biocontrol Agent of Pest Tepritid Fruit Flies. G3 Genes, Genomes, Genet. 7, 2407–2411. https://doi.org/10.1534/g3.117.040741

Gempe, T., Hasselmann, M., Schiøtt, M., Hause, G., Otte, M., Beye, M., 2009. Sex determination in honeybees: Two separate mechanisms induce and maintain the female pathway. PLoS Biol. 7, e1000222. https://doi.org/10.1371/journal.pbio.1000222

Geuverink, E., Beukeboom, L.W., 2014. Phylogenetic Distribution and Evolutionary Dynamics of the Sex Determination Genes *doublesex* and *transformer* in Insects. Sex. Dev. 8, 38–49. https://doi.org/10.1159/000357056

Geuverink, E., Kraaijeveld, K., van Leussen, M., Chen, F., Pijpe, J., Linskens, M.H.K., Beukeboom, L.W., van de Zande, L., 2018. Evidence for involvement of a *transformer* paralogue in sex determination of the wasp *Leptopilina clavipes*. Insect Mol. Biol. 27, 780–795. https://doi.org/10.1111/imb.12522

Ghorbani, M., Saber, M., Bagheri, M., Vaez, N., 2016. Effects of diazinon and fipronil on different developmental stages of *Trichogramma brassicae* Bezdenko (Hym.; Trichogrammatidae). J. Agric. Sci. Technol. 18, 1267–1278.

Gilchrist, G.W., 1996. A Quantitative Genetic Analysis of Thermal Sensitivity in the Locomotor Performance Curve of *Aphidius ervi*. Evolution (N. Y). 50, 1560. https://doi.org/10.2307/2410892

Giles, K.L., McCornack, B.P., Royer, T.A., Elliott, N.C., 2017. Incorporating biological control into IPM decision making. Curr. Opin. Insect Sci. 20, 84–89. https://doi.org/10.1016/j.cois.2017.03.009

Goldman, A.D., Landweber, L.F., 2016. What Is a Genome? PLoS Genet. 12, e1006181. https://doi.org/10.1371/journal.pgen.1006181

Gonçalves, C.I., Huigens, M.E., Verbaarschot, P., Duarte, S., Mexia, A., Tavares, J., 2006. Natural occurrence of *Wolbachia*-infected and uninfected *Trichogramma* species in tomato fields in Portugal. Biol. Control 37, 375–381. https://doi.org/10.1016/j.biocontrol.2005.11.015

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M.C., McCombie, W.R., 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res. 25, 1750–6. https://doi.org/10.1101/gr.191395.115

Gottlieb, Y., Zchori-Fein, E., 2001. Irreversible thelytokous reproduction in *Muscidifurax uniraptor*. Entomol. Exp. Appl. 100, 271–278. https://doi.org/10.1046/j.1570-7458.2001.00874.x

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–338. https://doi.org/10.1093/biomet/53.3-4.325

Granevitze, Z., Hillel, J., Chen, G.H., Cuc, N.T.K., Feldman, M., Eding, H., Weigend, S., 2007. Genetic diversity within chicken populations from different continents and management

histories. Anim. Genet. 38, 576–583. https://doi.org/10.1111/j.1365-2052.2007.01650.x

Grbić, M., Leeuwen, T. Van, Clark, R., Rombauts, S., 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. Nature.

Green, L.J.F., 2008. 'Indigenous Knowledge' and 'Science': Reframing the Debate on Knowledge Diversity. Archaeologies 4, 144–163. https://doi.org/10.1007/s11759-008-9057-9

Greiber, T., 2019. Implementation of the Nagoya Protocol in the European Union and in Germany. Phytomedicine 53, 313–318. https://doi.org/10.1016/j.phymed.2018.10.020

Grill, C.P., Moore, A.J., Brodie, E.D., 1997. The genetics of phenotypic plasticity in a colonizing population of the ladybird beetle, *Harmonia axyridis*. Heredity (Edinb). 78, 261–269. https://doi.org/10.1038/sj.hdy.6881030

Grillenberger, B.K., Koevoets, T., Burton-Chellew, M.N., Sykes, E.M., Shuker, D.M., Van De Zande, L., Bijlsma, R., Gadau, J., Beukeboom, L.W., 2008. Genetic structure of natural *Nasonia vitripennis* populations: Validating assumptions of sex-ratio theory. Mol. Ecol. 17, 2854–2864. https://doi.org/10.1111/j.1365-294X.2008.03800.x

Grozeva, S., Anokhin, B.A., Simov, N., Kuznetsova, V.G., 2019. New evidence for the presence of the telomere motif (TTAGG)n in the family Reduviidae and its absence in the families Nabidae and Miridae (Hemiptera, Cimicomorpha). Comp. Cytogenet. 13, 283–295. https://doi.org/10.3897/compcytogen.v13i3.36676

Grozeva, S., Kuznetsova, V.G., Anokhin, B.A., 2011. Karyotypes, male meiosis and comparative FISH mapping of 18S ribosomal DNA and telomeric (TTAGG)n repeat in eight species of true bugs (Hemiptera, Heteroptera). Comp. Cytogenet. 5, 97–116. https://doi.org/10.3897/CompCytogen.v5i4.2307

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gutzwiller, K., 1997. The Poetics of Editing in Meleager's Garland. Trans. Am. Philol. Assoc. 127, 169. https://doi.org/10.2307/284391

Guzmán-Larralde, A., Cerna-Chávez, E., Rodríguez-Campos, E., Loyola-Licea, J.C., Stouthamer, R., 2014. Genetic variation and the performance of a mass-reared parasitoid, *Trichogramma pretiosum* (Hymenoptera: Trichogrammatidae), in laboratory trials. J. Appl. Entomol. 138, 346–354. https://doi.org/10.1111/jen.12073

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Robin, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9, 1–22. https://doi.org/10.1186/gb-2008-9-1-r7

Hale, M.L., Burg, T.M., Steeves, T.E., 2012. Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. PLoS One 7. https://doi.org/10.1371/journal.pone.0045170

Hansen, T.F., Pélabon, C., Houle, D., 2011. Heritability is not Evolvability. Evol. Biol. 38, 258–277.

https://doi.org/10.1007/s11692-011-9127-6

Hare, E.E., Johnston, J.S., 2011. Genome Size Determination Using Flow Cytometry of Propidium Iodide-Stained Nuclei, Molecular Methods for Evolutionary Genetics. https://doi.org/10.1007/978-1-61779-228-1

Hartl, D., Clark, A., 1997. Principles of Population Genetics, 3rd ed. Sinauer Associates, Inc, Sunderland.

Hasselmann, M., Gempe, T., Schiøtt, M., Nunes-Silva, C.G., Otte, M., Beye, M., 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. Nature 454, 519–522. https://doi.org/10.1038/nature07052

Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G., Bopp, D., 2010. Molecular Characterization of the Key Switch F Provides a Basis for Understanding the Rapid Divergence of the Sex-Determining Pathway in the Housefly. Genetics 184, 155–170. https://doi.org/10.1534/genetics.109.109249

Heimpel, G.E., de Boer, J.G., 2008. Sex Determination in the Hymenoptera. Annu. Rev. Entomol. 53, 209–230. https://doi.org/10.1146/annurev.ento.53.103106.093441

Henry, L.M., May, N., Acheampong, S., Gillespie, D.R., Roitberg, B.D., 2010. Host-adapted parasitoids in biological control: Does source matter? Ecol. Appl. 20, 242–250. https://doi.org/10.1890/08-1869.1

Henter, H.J., 1995. The Potential for Coevolution in a Host-Parasitoid System. II. Genetic Variation within a Population of Wasps in the Ability to Parasitize an Aphid Host. Evolution (N. Y). 49, 439. https://doi.org/10.2307/2410268

Henter, H.J., Brasch, K., Lenteren, J.C., 1996. Variation between laboratory populations of *Encarsia formosa* in their parasitization behavior on the host *Bemisia tabaci*. Entomol. Exp. Appl. 80, 435–441. https://doi.org/10.1111/j.1570-7458.1996.tb00957.x

Hillel, J., Groenen, M.A.M., Tixier-Boichard, M., Korol, A.B., David, L., Kirzhner, V.M., Burke, T., Barre-Dirie, A., Crooijmans, R.P.M.A., Elo, K., Feldman, M.W., Freidlin, P.J., Mäki-Tanila, A., Oortwijn, M., Thomson, P., Vignal, A., Wimmers, K., Weigend, S., 2003. Biodiversity of 52 chicken populations assessed by microsatellite typing of DNA pools. Genet. Sel. Evol 35, 533–557. https://doi.org/10.1051/gse:2003038

Hinomoto, N., Todokoro, Y., Higaki, T., 2011. Population structure of the predatory mite *Neoseiulus womersleyi* in a tea field based on an analysis of microsatellite DNA markers. Exp. Appl. Acarol. 53, 1–15. https://doi.org/10.1007/s10493-010-9384-6

Hira, S., 2012. Decolonizing the Mind: The Case of the Netherlands. Hum. Archit. J. Sociol. Self-knowledge 10, 53–68.

Hoedjes, K.M., van den Heuvel, J., Kapun, M., Keller, L., Flatt, T., Zwaan, B.J., 2019. Distinct genomic signals of lifespan and life history evolution in response to postponed reproduction and larval diet in *Drosophila*. Evol. Lett. https://doi.org/10.1002/evl3.143

Hoffmann, A.A., Ross, P.A., 2018. Rates and Patterns of Laboratory Adaptation in (Mostly) Insects. J. Econ. Entomol. 1–9. https://doi.org/10.1093/jee/toy024

Holloway, A.K., Heimpel, G.E., Strand, M.R., Antolin, M.F., 1999. Survival of diploid males in *Bracon* sp. near *hebetor* (Hymenoptera: Braconidae). Ann. Entomol. Soc. Am. 92, 110–116. https://doi.org/10.1093/aesa/92.1.110

Holloway, G.J., Brakefield, P.M., de Jong, P., Ottenheim, M.M., De Vos, H., Kesbeke, F., Peynenburg, L., 1995. A quantitative genetic analysis of an aposematic colour pattern and its ecological implications. Philos. Trans. R. Soc. B Biol. Sci. 348, 373–379. https://doi.org/10.1098/rstb.1995.0075

Holloway, G.J., de Jong, P.W., Ottenheim, M., 1993. The Genetics and Cost of Chemical Defense in the Two-Spot Ladybird (*Adalia bipunctata* L.). Evolution (N. Y). 47, 1229. https://doi.org/10.2307/2409988

Holt, R.D., Hochberg, M.E., 1997. When is biological control evolutionarily stable (or is it)? Ecology 78, 1673–1683. https://doi.org/10.1890/0012-9658(1997)078[1673:WIBCES]2.0.CO;2

Hopper, K.R., Roush, R.T., Powell, W., 1993. Management of genetics of biological-control introductions. Annu. Rev. Entomol. 38, 27–51. https://doi.org/10.1146/annurev.ento.38.1.27

Hosmani, P.S., Flores-Gonzalez, M., van de Geest, H., Maumus, F., Bakker, L. V, Schijlen, E., van Haarst, J., Cordewener, J., Sanchez-Perez, G., Peters, S., Fei, Z., Giovannoni, J.J., Mueller, L.A., Saha, S., 2019. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv 767764. https://doi.org/https://doi.org/10.1101/767764

Houle, D., 1992. Comparing evolvability and variability of quantitative traits. Genetics 130, 195–204.

Hoy, M.A., Waterhouse, R.M., Wu, K., Estep, A.S., Ioannidis, P., Palmer, W.J., Pomerantz, A.F., Simã O, F.A., Thomas, J., Jiggins, F.M., Murphy, T.D., Pritham, E.J., Robertson, H.M., Zdobnov, E.M., Gibbs, R.A., Richards, S., 2016. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution GBE. Genome Biol. Evol 8, 1762–1775. https://doi.org/10.1093/gbe/evw048

Hoy, M.A., Yu, F., Meyer, J.M., Tarazona, O.A., Jeyaprakash, A., Wu, K., 2013. Transcriptome sequencing and annotation of the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae): a cautionary tale about possible contamination by prey sequences. Exp. Appl. Acarol. 59, 283–296. https://doi.org/10.1007/s10493-012-9603-4

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13. https://doi.org/10.1093/nar/gkn923

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, H., McGarvey, P.B., Suzek, B.E., Mazumder, R., Zhang, J., Chen, Y., Wu, C.H., 2011. A comprehensive protein-centric ID mapping service for molecular data integration. Bioinformatics 27, 1190–1191. https://doi.org/10.1093/bioinformatics/btr101

Hufbauer, R.A., 2001. Pea aphid-parasitoid interactions: Have parasitoids adapted to differential resistance? Ecology 82, 717–725. https://doi.org/10.1890/0012-9658(2001)082[0717:PAPIHP]2.0.CO;2

Hufbauer, R.A., Via, S., 1999. Evolution of an Aphid-Parasitoid Interaction: Variation in Resistance to Parasitism among Aphid Populations Specialized on Different Plants. Evolution (N. Y). 53, 1435. https://doi.org/10.2307/2640890

Huigens, M.E., Pashalidou, F.G., Qian, M.-H., Bukovinszky, T., Smid, H.M., van Loon, J.J.A., Dicke, M., Fatouros, N.E., 2009. Hitch-hiking parasitic wasp learns to exploit butterfly antiaphrodisiac. Proc. Natl. Acad. Sci. U. S. A. 106, 820–825. https://doi.org/10.1073/pnas.0812277106

Huigens, M.E., Woelke, J.B., Pashalidou, F.G., Bukovinszky, T., Smid, H.M., Fatouros, N.E., 2010. Chemical espionage on species-specific butterfly anti-aphrodisiacs by hitchhiking *Trichogramma* wasps. Behav. Ecol. 21, 470–478. https://doi.org/10.1093/beheco/arq007

Husnik, F., McCutcheon, J.P., 2018. Functional horizontal gene transfer from bacteria to eukaryotes. Nat. Rev. Microbiol. 16, 67–79. https://doi.org/10.1038/nrmicro.2017.137

Indigenous STS, 2019. Indigenous STS [WWW Document]. Univ. Alberta, Fac. Nativ. Stud. URL https://indigenoussts.com/ (accessed 12.13.19).

IPCC, 2007. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107415324.015

Itou, M., Watanabe, M., Watanabe, E., Miura, K., 2013. Gut content analysis to study predatory efficacy of *Nesidiocoris tenuis* (Reuter) (Hemiptera: Miridae) by molecular methods. Entomol. Sci. 16, 145–150. https://doi.org/10.1111/j.1479-8298.2012.00552.x

Ivezić, A., Rugman-Jones, P., Stouthamer, R., Ignjatović-Ćupina, A., 2018. Molecular identification of *Trichogramma* egg parasitoids of *Ostrinia nubilalis* in northeastern Serbia. Arch. Biol. Sci. 70, 425–432. https://doi.org/10.2298/ABS171103002I

Jackman, S.D., Vandervalk, B.P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S.A., Jahesh, G., Khan, H., Coombe, L., Warren, R.L., Birol, I., 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27, 768–777. https://doi.org/10.1101/gr.214346.116

Jamshidnia, A., Abdoli, S., Farrokhi, S., Sadeghi, R., 2018. Efficiency of spinosad, *Bacillus thuringiensis* and *Trichogramma brassicae* against the tomato leafminer in greenhouse. BioControl 63, 619–627. https://doi.org/10.1007/s10526-018-9893-5

Jang, E.B., Nishijima, K.A., 1990. Identification and Attractancy of Bacteria Associated with *Dacus dorsalis* (Diptera: Tephritidae). Environ. Entomol. 19, 1726–1731. https://doi.org/10.1093/ee/19.6.1726

Jauset, A.M., Edo-Tena, E., Castañé, C., Agustí, N., Alomar, O., Grozeva, S., 2015. Comparative cytogenetic study of three *Macrolophus* species (Heteroptera, Miridae). Comp. Cytogenet. 9, 613–623. https://doi.org/10.3897/CompCytogen.v9i4.5530

Jetybayev, I.Y., Bugrov, A.G., Ünal, M., Buleu, O.G., Rubtsov, N.B., 2017. Molecular cytogenetic analysis reveals the existence of two independent neo-XY sex chromosome systems in Anatolian Pamphagidae grasshoppers. BMC Evol. Biol. 17, 1–12. https://doi.org/10.1186/s12862-016-0868-9

Jia, F., Margolies, D.C., Boyer, J.E., Charlton, R.E., 2002. Genetic variation in foraging traits among inbred lines of a predatory mite. Heredity (Edinb). 89, 371–379. https://doi.org/10.1038/sj.hdy.6800145

Jia, L.-Y., Xiao, J.-H., Xiong, T.-L., Niu, L.-M., Huang, D.-W., 2016. The *transformer* genes in the fig wasp *Ceratosolen solmsi* provide new evidence for duplications independent of complementary sex determination. Insect Mol. Biol. 25, 191–201. https://doi.org/10.1111/imb.12210

Johnston, J.S., Ross, L.D., Beani, L., Hughes, D.P., Kathirithamby, J., 2004. Tiny genomes and endoreduplication in Strepsiptera. Insect Mol. Biol. 13, 581–585. https://doi.org/10.1111/j.0962-1075.2004.00514.x

Joint Submission of Grand Council of the Crees (Eeyou Istchee) et al, 2011. Nagoya Protocol on Access and Benefit Sharing: Substantive and Procedural Injustices relating to Indigenous Peoples' Human Rights [WWW Document]. CBD. URL https://www.cbd.int/abs/doc/protocol/icnp-1/joint-submission-grand-council-and-others-en.pdf (accessed 12.12.19).

Jones, R.N., Rees, H., 1982. B Chromosomes. Academic Press, New York.

Jones, Steven, Taylor, G., Chan, S., Warren, R., Hammond, S., Bilobram, S., Mordecai, G., Suttle, C., Miller, K., Schulze, A., Chan, A., Jones, Samantha, Tse, K., Li, I., Cheung, D., Mungall, K., Choo, C., Ally, A., Dhalla, N., Tam, A., Troussard, A., Kirk, H., Pandoh, P., Paulino, D., Coope, R., Mungall, A., Moore, R., Zhao, Y., Birol, I., Ma, Y., Marra, M., Haulena, M., 2017. The Genome of the Beluga Whale (*Delphinapterus leucas*). Genes (Basel). 8, 378. https://doi.org/10.3390/genes8120378

Jost, L., 2009. D vs. GST: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Mol. Ecol. 18, 2088–2091. https://doi.org/10.1111/j.1365-294X.2009.04186.x

Jost, L., 2008. GST and its relatives do not measure differentiation. Mol. Ecol. 17, 4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Jung, S., Lee, S., 2012. Molecular phylogeny of the plant bugs (Heteroptera: Miridae) and the evolution of feeding habits. Cladistics 28, 50–79. https://doi.org/10.1111/j.1096-0031.2011.00365.x

Kamala Jayanthi, P.D., Sangeetha, P., Verghese, A., 2014. Study of inheritance of feeding potential in natural populations of predatory coccinellid *Cryptolaemus montrouzieri* Mulsant using isofemale strains. J. Genet. 93, 113–122. https://doi.org/10.1007/s12041-014-0350-7

Kamber, T., Smits, T.H.M., Rezzonico, F., Duffy, B., 2012. Genomics and current genetic understanding of *Erwinia amylovora* and the fire blight antagonist *Pantoea vagans*. Trees - Struct. Funct. 26, 227–238. https://doi.org/10.1007/s00468-011-0619-x

Kamping, A., Katju, V., Beukeboom, L.W., Werren, J.H., 2007. Inheritance of gynandromorphism in the parasitic wasp *Nasonia vitripennis*. Genetics 175, 1321–1333. https://doi.org/10.1534/genetics.106.067082

Kane, N., 2011. Should I use Fst, G'st or D? [WWW Document]. Mol. Ecol. URL https://www.molecularecologist.com/2011/03/should-i-use-fst-gst-or-d-2/ (accessed 11.18.19).

Kapun, M., Barrón, M.G., Staubach, F., Vieira, J., Obbard, D.J., Goubert, C., Rota-Stabelli, O., Kankare, M., Bogaerts-Márquez, M., Haudry, A., Wiberg, R.A.W., Waidele, L., Kozeretska, I., Pasyukova, E.G., Loeschcke, V., Pascual, M., Vieira, C.P., Serga, S., Montchamp-Moreau, C., Abbott, J., Gibert, P., Porcelli, D., Posnien, N., Sánchez-Gracia, A., Grath, S., Sucena, É., Bergland, A.O., Guerreiro, M.P.G., Onder, B.S., Argyridou, E., Guio, L., Schou, M.F., Deplancke, B., Vieira, C., Ritchie, M.G., Zwaan, B.J., Tauber, E., Orengo, D.J., Puerma, E., Aguadé, M., Schmidt, P.S., Parsch, J., Betancourt, A.J., Flatt, T., González, J., 2018. Genomic analysis of European Drosophila populations reveals major longitudinal structure, continent-wide selection, and unknown DNA viruses. bioRxiv 313759. https://doi.org/10.1101/313759

Karabinos, M., 2019. Decolonisation in Dutch Archives: Defining and Debating. BMGN - Low Ctries. Hist. Rev. 134, 129. https://doi.org/10.18352/bmgn-lchr.10687

Karabörklü, S., Azizoglu, U., Azizoglu, Z.B., 2018. Recombinant entomopathogenic agents: a review of biotechnological approaches to pest insect control. World J. Microbiol. Biotechnol. 34, 14. https://doi.org/10.1007/s11274-017-2397-0

Kares, E.A., El-Sappagh, I.A., Ebaid, G.H., Sabra, I.M., 2010. Efficacy of releasing *Bracon brevicornis* Wesm. (Hymenoptera: Braconidae) for controlling hibernated *Ostrinia nubilalis* (Hübner) and *Sesamia cretica* led. larvae in stored corn stalks. Egypt. J. Biol. Pest Control 20, 155–159.

Kariyawasam, K., Tsai, M., 2018. Access to genetic resources and benefit sharing: Implications of Nagoya Protocol on providers and users. J. World Intellect. Prop. 21, 289–305. https://doi.org/10.1111/jwip.12095

Kato, A., Albert, P.S., Vega, J.M., Birchler, J.A., 2006. Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. Biotech. Histochem. 81, 71–78. https://doi.org/10.1080/10520290600643677

Katoh, K., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keeling, C.I., Yuen, M.M.S., Liao, N.Y., Roderick Docking, T., Chan, S.K., Taylor, G.A., Palmquist, D.L., Jackman, S.D., Nguyen, A., Li, M., Henderson, H., Janes, J.K., Zhao, Y., Pandoh, P., Moore, R., Sperling, F.A.H., W Huber, D.P., Birol, I., Jones, S.J.M., Bohlmann, J., 2013. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol. 14, R27. https://doi.org/10.1186/gb-2013-14-3-r27

Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., Hartung, F., 2016. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 44. https://doi.org/10.1093/nar/gkw092

King, R.A., Read, D.S., Traugott, M., Symondson, W.O.C., 2008. Molecular analysis of predation: a review of best practice for DNA-based approaches. Mol. Ecol. 17, 947–963. https://doi.org/10.1111/j.1365-294X.2007.03613.x

Kittel, R.N., Maeto, K., 2019. Revalidation of *Habrobracon brevicornis* stat. rest. (Hymenoptera: Braconidae) Based on the CO1, 16S, and 28S Gene Fragments. J. Econ. Entomol. 112, 906–911. https://doi.org/10.1093/jee/toy368

Klug, T., Meyhöfer, R., 2009. Performance of two *Trichogramma brassicae* strains under greenhouse and field conditions for biocontrol of the silver Y moth in spinach cultures. J. Pest Sci. (2004). 82, 73–79. https://doi.org/10.1007/s10340-008-0224-y

Knapp, M., Houten, Y. van, Baal, E. van, Groot, T., 2018. Use of predatory mites in commercial biocontrol: Current status and future prospects. Acarologia 58, 72–82. https://doi.org/10.24349/acarologia/20184275

Koch, V., Nissen, I., Schmitt, B.D., Beye, M., 2014. Independent evolutionary origin of femparalogous genes and complementary sex determination in hymenopteran insects. PLoS One 9, e91883. https://doi.org/10.1371/journal.pone.0091883

Kofler, R., Orozco-terWengel, P., de Maio, N., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., Schlötterer, C., 2011a. PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6. https://doi.org/10.1371/journal.pone.0015925

Kofler, R., Pandey, R.V., Schlotterer, C., 2011b. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27, 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736. https://doi.org/10.1101/gr.215087.116

Koricheva, J., Gurevitch, J., Mengersen, K., 2013. Handbook of meta-analysis in ecology and evolution, Handbook of Meta-analysis in Ecology and Evolution. Princeton University Press, Oxford. https://doi.org/10.23943/princeton/9780691137285.001.0001

Koskinioti, P., Ras, E., Augustinos, A.A., Tsiamis, G., Beukeboom, L.W., Caceres, C., Bourtzis, K., 2019. The effects of geographic origin and antibiotic treatment on the gut symbiotic communities of *Bactrocera oleae* populations. Entomol. Exp. Appl. 167, 197–208. https://doi.org/10.1111/eea.12764

Kraaijeveld, K., Neleman, P., Mariën, J., de Meijer, E., Ellers, J., 2019. Genomic Resources for *Goniozus legneri*, *Aleochara bilineata* and *Paykullia maculata*, Representing Three Independent Origins of the Parasitoid Lifestyle in Insects. G3 Genes, Genomes, Genet. 9, g3.300584.2019. https://doi.org/10.1534/g3.119.300584

Kruitwagen, A., Beukeboom, L.W., Wertheim, B., 2018. Optimization of native biocontrol agents, with parasitoids of the invasive pest *Drosophila suzukii* as an example. Evol. Appl. 11, 1473–1497. https://doi.org/10.1111/eva.12648

Kumar, A., Purohit, A.K., 2012. The Role of Indigenous Knowledge in Biological Control of Plant Pathogens: Logistics of New Research Initiatives, in: Mérillon, J.M., Ramawat, K.G. (Eds.), Plant Defence: Biological Control. Springer Netherlands, Dordrecht, pp. 161–194. https://doi.org/10.1007/978-94-007-1933-0

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Kurland, S., Wheat, C.W., Paz Celorio Mancera, M., Kutschera, V.E., Hill, J., Andersson, A., Rubin, C., Andersson, L., Ryman, N., Laikre, L., 2019. Exploring a Pool-seq-only approach for gaining population genomic insights in nonmodel species. Ecol. Evol. 11448–11463. https://doi.org/10.1002/ece3.5646

Kuske, S., Babendreier, D., Edwards, P.J., Turlings, T.C.J., Bigler, F., 2004. Parasitism of non-target lepidoptera by mass released *Trichogramma brassicae* and its implication for the larval parasitoid *Lydella thompsoni*. BioControl 49, 1–19. https://doi.org/10.1023/B:BICO.0000009379.13685.47

Kuznetsova, V.G., Grozeva, S.M., Nokkala, S., Nokkala, C., 2011. Cytogenetics of the true bug infraorder Cimicomorpha (Hemiptera, Heteroptera): A review. Zookeys 154, 31–70. https://doi.org/10.3897/zookeys.154.1953

Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research 6, 1287. https://doi.org/10.12688/f1000research.12232.1

Larsen, P.O., von Ins, M., 2010. The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics 84, 575–603. https://doi.org/10.1007/s11192-010-0202-z

Laurent, V., Wajnberg, E., Mangin, B., Schiex, T., Gaspin, C., Vanlerberghe-Masutti, F., 1998. A composite genetic map of the parasitoid wasp *Trichogramma brassicae* based on RAPD markers. Genetics 150, 275–82.

Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., Studholme, D.J., 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol. Detect.

Quantif. 3, 1–8. https://doi.org/10.1016/J.BDQ.2015.02.001

Le Hesran, S., Groot, T., Knapp, M., Bukovinszky, T., Forestier, T., Dicke, M., 2019a. Phenotypic variation in egg survival in the predatory mite *Phytoseiulus persimilis* under dry conditions. Biol. Control 130, 88–94. https://doi.org/10.1016/j.biocontrol.2018.10.007

Le Hesran, S., Groot, T., Knapp, M., Bukovinszky, T., Nugroho, J.E., Beretta, G., Dicke, M., 2019b. Maternal effect determines drought resistance of eggs in the predatory mite *Phytoseiulus persimilis*. Oecologia. https://doi.org/10.1007/s00442-019-04556-0

Le Hesran, S., Groot, T., Knapp, M., Nugroho, J.E., Beretta, G., Salomé-Abarca, L.F., Choi, Y.H., Vancová, M., Moreno-Rodenas, A.M., Dicke, M., 2019c. Proximate mechanisms of drought resistance in *Phytoseiulus persimilis* eggs. Exp. Appl. Acarol. 79, 279–298. https://doi.org/10.1007/s10493-019-00442-9

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., Prohaska, S.J., 2011. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12, 124. https://doi.org/10.1186/1471-2105-12-124

Lee, W., Kang, J., Jung, C., Hoelmer, K., Lee, S.H., Lee, S., 2009. Complete mitochondrial genome of brown marmorated stink bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and phylogenetic relationships of hemipteran suborders. Mol. Cells 28, 155–165. https://doi.org/10.1007/s10059-009-0125-9

Legeai, F., Shigenobu, S., Gauthier, J.P., Colbourne, J., Rispe, C., Collin, O., Richards, S., Wilson, A.C.C., Murphy, T., Tagu, D., 2010. AphidBase: A centralized bioinformatic resource for annotation of the pea aphid genome. Insect Mol. Biol. 19, 5–12. https://doi.org/10.1111/j.1365-2583.2009.00930.x

Legner, E.F., 1993. Theory for quantitative inheritance of behavior in a protelean parasitoid, *Muscidifurax raptorellus* (Hymenoptera: Pteromalidae). Eur. J. Entomol. 90, 11–21.

Legner, E.F., 1991a. Estimations of number of active loci, dominance and heritability in polygenic inheritance of gregarious behavior in *Muscidifurax raptorellus* [Hymenoptera: Pteromalidae]. Entomophaga 36, 1–18. https://doi.org/10.1007/BF02374633

Legner, E.F., 1991b. Recombinant males in the parasitic wasp *Muscidifurax raptorellus* [Hymenoptera: Pteromalidae]. Entomophaga 36, 173–181. https://doi.org/10.1007/BF02374553

Legner, E.F., 1988. *Muscidifurax raptorellus* (Hymenoptera: Pteromalidae) Females Exhibit Postmating Oviposition Behavior Typical of the Male Genome. Ann. Entomol. Soc. Am. 81, 522–527. https://doi.org/10.1093/aesa/81.3.522

Legner, E.F., 1987. Inheritance of gregarious and solitary oviposition in *Muscidifurax raptorellus* Kogan and Legner (Hymenoptera: Pteromalidae). Can. Entomol. 119, 791–808. https://doi.org/10.4039/Ent119791-9

Lesna, I., Sabelis, M.W., 1999. Diet-dependent female choice for males with 'good genes' in a soil predatory mite. Nature 401, 581–584. https://doi.org/10.1038/44125

Lessard, E., Boivin, G., 2013. Effect of low temperature on emergence, fecundity, longevity and

host-feeding by *Trichogramma brassicae*. BioControl 58, 319–329. https://doi.org/10.1007/s10526-012-9493-8

Leung, K., Ras, E., Ferguson, K.B., Ariëns, S., Babendreier, D.B., Bijma, P., Bourtzis, K., Brodeur, J., Bruins, M., Centurión, A., Chattington, S., Chinchilla-Ramírez, M., Dicke, M., Fatouros, N., González Cabrera, J., Groot, T., Haye, T., Knapp, M., Koskinioti, P., Le Hesran, S., Lirakis, M., Paspati, A., Pérez-Hedo, M., Plouvier, W., Schlötterer, C., Stahl, J., Thiel, A., Urbaneja, A., van de Zande, L., Verhulst, E., Vet, L., Visser, S., Werren, J., Xia, S., Zwaan, B., Magalhães, S., Beukeboom, L., Pannebakker, B., 2019a. Next Generation Biological Control: the Need for Integrating Genetics and Evolution. Preprints 1–34. https://doi.org/10.20944/preprints201911.0300.v1

Leung, K., Zande, L., Beukeboom, L.W., 2019b. Life-history traits of the Whiting polyploid line of the parasitoid *Nasonia vitripennis*. Entomol. Exp. Appl. eea.12808. https://doi.org/10.1111/eea.12808

Lewis, W.J., van Lenteren, J.C., Phatak, S.C., Tumlinson, J.H., 1997. A total system approach to sustainable pest management. Proc. Natl. Acad. Sci. 94, 12243–12248. https://doi.org/10.1073/pnas.94.23.12243

Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z., Walters, J.R., 2019. Insect genomes: progress and challenges. Insect Mol. Biol. https://doi.org/10.1111/imb.12599

Li, H., 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X.Y., van Achterberg, C., Tan, J.C., 2013. Revision of the subfamily Opiinae (Hymenoptera, Braconidae) from Hunan (China), including thirty-six new species and two new genera. Zookeys 268, 1–186. https://doi.org/10.3897/zookeys.268.4071

Li, Z.X., Zheng, L., Shen, Z.R., 2004. Using internally transcribed spacer 2 sequences to re-examine the taxonomic status of several cryptic species of *Trichogramma* (Hymenoptera: Trichogrammatidae). Eur. J. Entomol. 101, 347–358.

Lindsey, A.R.I., Kelkar, Y.D., Wu, X., Sun, D., Martinson, E.O., Yan, Z., Rugman-Jones, P.F., Hughes, D.S.T., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H.V., Worley, K.C., Muzny, D.M., Ye, G., Gibbs, R.A., Richards, S., Yi, S. V., Stouthamer, R., Werren, J.H., 2018. Comparative genomics of the miniature wasp and pest control agent *Trichogramma pretiosum*. BMC Biol. 16, 1–20. https://doi.org/10.1186/s12915-018-0520-9

Lindsey, A.R.I., Werren, J.H., Richards, S., Stouthamer, R., 2016. Comparative Genomics of a Parthenogenesis-Inducing *Wolbachia* Symbiont. G3 Genes, Genomes, Genet. 6, 2113–2123. https://doi.org/10.1534/g3.116.028449

Lirakis, M., Dolezal, M., Schlötterer, C., 2018. Redefining reproductive dormancy in *Drosophila*

as a general stress response to cold temperatures. J. Insect Physiol. 107, 175–185. https://doi.org/10.1016/j.jinsphys.2018.04.006

Lirakis, M., Magalhães, S., 2019. Does experimental evolution produce better biological control agents? A critical review of the evidence. Entomol. Exp. Appl. 584–597. https://doi.org/10.1111/eea.12815

Liu, F.H., Smith, S.M., 2000. Measurement and selection of parasitoid quality for mass-reared *Trichogramma minutum* Riley used in inundative release. Biocontrol Sci. Technol. 10, 3–13. https://doi.org/10.1080/09583150029332

Liu, T.X., Zhang, Y., 2012. Side effects of two reduced-risk insecticides, indoxacarb and spinosad, on two species of *Trichogramma* (Hymenoptera: Trichogrammatidae) on cabbage. Ecotoxicology 21, 2254–2263. https://doi.org/10.1007/s10646-012-0981-5

Lommen, S.T.E., de Jong, P.W., Pannebakker, B.A., 2017. It is time to bridge the gap between exploring and exploiting: prospects for utilizing intraspecific genetic variation to optimize arthropods for augmentative pest control – a review. Entomol. Exp. Appl. 162, 108–123. https://doi.org/10.1111/eea.12510

Lovett, B., St. Leger, R.J., 2018. Genetically engineering better fungal biopesticides. Pest Manag. Sci. 74, 781–789. https://doi.org/10.1002/ps.4734

Lowe-Power, T.M., Khokhani, D., Allen, C., 2018. How *Ralstonia solanacearum* Exploits and Thrives in the Flowing Plant Xylem Environment. Trends Microbiol. 26, 929–942. https://doi.org/10.1016/j.tim.2018.06.002

Mackauer, M., 1976. Genetic problems in the production of biological control agents. Annu. Rev. Entomol. 21, 369–385. https://doi.org/doi:10.1146/annurev.en.21.010176.002101

Maeda, T., Takabayashi, J., Yano, S., Takafuji, A., 1999. Response of the predatory mite, *Amblyseius womersleyi* (Acari: Phytoseiidae), toward herbivore-induced plant volatiles: Variation in response between two local populations. Appl. Entomol. Zool. 34, 449–454. https://doi.org/10.1303/aez.34.449

Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular cloning: a laboratory manual., 545th ed. Cold spring harbor laboratory, NY, NY.

Mansoor, M.M., Abbas, N., Shad, S.A., Pathan, A.K., Razaq, M., 2013. Increased fitness and realized heritability in emamectin benzoate-resistant *Chrysoperla carnea* (Neuroptera: Chrysopidae). Ecotoxicology 22, 1232–1240. https://doi.org/10.1007/s10646-013-1111-8

Mansoor, M.M., Raza, A.B.M., Abbas, N., Aqueel, M.A., Afzal, M., 2017. Resistance of green lacewing, *Chrysoperla carnea* Stephens to nitenpyram: Cross-resistance patterns, mechanism, stability, and realized heritability. Pestic. Biochem. Physiol. 135, 59–63. https://doi.org/10.1016/j.pestbp.2016.06.004

Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of

occurrences of k-mers. Bioinformatics 27, 764–770. https://doi.org/10.1093/bioinformatics/btr011

Marec, F., Shvedov, A.N., 1990. Yellow eye, a new pigment mutation in *Ephestia kuehniella* Zeller (Lepidoptera: Pyralidae). Hereditas 113, 97–100. https://doi.org/10.1111/j.1601-5223.1990.tb00072.x

Margolies, D.C., Sabelis, M.W., Boyer, J.E., 1997. Response of a phytoseiid predator to herbivore-induced plant volatiles: Selection on attraction and effect on prey exploitation. J. Insect Behav. 10, 695–709. https://doi.org/10.1007/BF02765387

Marples, N.M., de Jong, P.W., Ottenheim, M.M., Verhoog, M.D., Brakefield, P.M., 1993. The inheritance of a wingless character in the 2 spot ladybird (*Adalia bipunctata*). Entomol. Exp. Appl. 69, 69–73. https://doi.org/10.1111/j.1570-7458.1993.tb01729.x

Martínez-García, H., Román-Fernández, L.R., Sáenz-Romo, M.G., Pérez-Moreno, I., Marco-Mancebón, V.S., 2016. Optimizing *Nesidiocoris tenuis* (Hemiptera: Miridae) as a biological control agent: mathematical models for predicting its development as a function of temperature. Bull. Entomol. Res. 106, 215–224. https://doi.org/10.1017/S0007485315000978

Mason, P.G., Cock, M.J.W., Barratt, B.I.P., Klapwijk, J.N., van Lenteren, J.C., Brodeur, J., Hoelmer, K.A., Heimpel, G.E., 2017. Best practices for the use and exchange of invertebrate biological control genetic resources relevant for food and agriculture. BioControl 63, 1–6. https://doi.org/10.1007/s10526-017-9810-3

Mason, P.G., Hopper, K.R., 1997. Temperature Dependence in Locomotion of the Parasitoid *Aphelinus asychis* (Hymenoptera: Aphelinidae) from Geographical Regions with Different Climates. Environ. Entomol. 26, 1416–1423. https://doi.org/10.1093/ee/26.6.1416

May, C.M., Heuvel, J., Doroszuk, A., Hoedjes, K.M., Flatt, T., Zwaan, B.J., 2019. Adaptation to developmental diet influences the response to selection on age at reproduction in the fruit fly. J. Evol. Biol. 32, 425–437. https://doi.org/10.1111/jeb.13425

Mayhew, P.J., 2016. Comparing parasitoid life histories. Entomol. Exp. Appl. 159, 147–162. https://doi.org/10.1111/eea.12411

McKie, R., 2019. Top university split in row over erasing 'racist' science pioneers from the campus. Guard.

Mediouni, J., Fuková, I., Frydrychová, R., Marec, F., Fuková, I., Marec, F., Dhouibi, M.H., 2004. Karyotype, sex chromatin and sex chromosome differentiation in the carob moth, *Ectomyelois ceratoniae* (Lepidoptera: Pyralidae). Caryologia 57, 184–194. https://doi.org/10.1080/00087114.2004.10589391

Megens, H.J., Crooijmans, R., Cristobal, M.S., Hui, X., Li, N., Groenen, M.A.M., 2008. Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. Genet. Sel. Evol. 40, 103–128. https://doi.org/10.1051/gse:2007039

Misra, A.K., 2014. Climate change and challenges of water and food security. Int. J. Sustain. Built Environ. 3, 153–165. https://doi.org/10.1016/j.ijsbe.2014.04.006

Mollá, O., Biondi, A., Alonso-Valiente, M., Urbaneja, A., 2014. A comparative life history study of two mirid bugs preying on *Tuta absoluta* and *Ephestia kuehniella* eggs on tomato crops: Implications for biological control. BioControl 59, 175–183. https://doi.org/10.1007/s10526-013-9553-8

Moonga, H.B., Schoustra, S.E., Linnemann, A.R., Kuntashula, E., Shindano, J., Smid, E.J., 2019. The art of mabisi production: A traditional fermented milk. PLoS One 14, e0213541. https://doi.org/10.1371/journal.pone.0213541

Morgera, E., Tsioumani, E., Buck, M., 2014. Unraveling the Nagoya Protocol: a commentary on the Nagoya Protocol on access and benefit-sharing to the Convention on Biological Diversity. Brill.

Morgulis, A., Gertz, E.M., Schäffer, A.A., Agarwala, R., 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. J. Comput. Biol. 13, 1028–1040. https://doi.org/10.1089/cmb.2006.13.1028

Mousseau, T.A., Roff, D.A., 1987. Natural selection and the heritability of fitness components. Heredity (Edinb). 59, 181–197. https://doi.org/10.1038/hdy.1987.113

Nachappa, P., Margolies, D.C., Nechols, J.R., Morgan, T.J., 2010. Response of a complex foraging phenotype to artificial selection on its component traits. Evol. Ecol. 24, 631–655. https://doi.org/10.1007/s10682-009-9318-0

Nagtegaal, L.W., de Bruin, R.E., 1994. The French connection and other neo-colonial patterns in the global network of science. Res. Eval. 4, 119–127. https://doi.org/10.1093/rev/4.2.119

Naranjo, S.E., Ellsworth, P.C., Frisvold, G.B., 2015. Economic Value of Biological Control in Integrated Pest Management of Managed Plant Systems. Annu. Rev. Entomol. 60, 621–645. https://doi.org/10.1146/annurev-ento-010814-021005

Narayanan, E.S., Angalet, G.W., Subba Rao, B.R., D'Souza, G.I., 1954. Effect of refrigeration of the pupae of *Microbracon brevicornis* Wesm. on the pigmentation of the adult. Nature 173, 503–504. https://doi.org/10.1038/173503b0

Nasir, M.F., Hagedorn, G., Büttner, C., Reichmuth, C., Schöller, M., 2013. Molecular identification of *Trichogramma* species from Pakistan, using ITS-2 region of rDNA. BioControl 58, 483–491. https://doi.org/10.1007/s10526-013-9509-z

National Institutes of Health, 2019. What is a genome? [WWW Document]. URL https://ghr.nlm.nih.gov/primer/hgp/genome (accessed 11.14.19).

Navajas, M.J., Thistlewood, H.M.A., Lagnel, J., Hughes, C., 1998. Microsatellite sequences are under-represented in two mite genomes. Insect Mol. Biol. 7, 249–256. https://doi.org/10.1111/j.1365-2583.1998.00066.x

Navia, D., Domingos, C.A., Mendonça, R.S., Ferragut, F., Rodrigues, M.A.N., de Morais, E.G.F., Tixier, M.S., Gondim, M.G.C., 2014. Reproductive compatibility and genetic and morphometric variability among populations of the predatory mite, *Amblyseius largoensis* (Acari: Phytoseiidae), from Indian Ocean Islands and the Americas. Biol. Control 72, 17–29. https://doi.org/10.1016/j.biocontrol.2014.01.011

Appendices

NCBI, 2019. NCBI: Genome [WWW Document]. NCBI. URL https://www.ncbi.nlm.nih.gov/genome/ (accessed 11.18.19).

Nei, M., 1977. F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Genet. 41, 225–233. https://doi.org/10.1111/j.1469-1809.1977.tb01918.x

Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19, 153–170. https://doi.org/10.1007/BF02300753

Nicholas, G., 2018. It's taken thousands of years, but Western science is finally catching up to Traditional Knowledge [WWW Document]. Conversat. URL https://theconversation.com/its-taken-thousands-of-years-but-western-science-is-finally-catching-up-to-traditional-knowledge-90291 (accessed 12.13.19).

Nielsen, E.S., Henriques, R., Toonen, R.J., Knapp, I.S.S., Guo, B., von der Heyden, S., 2018. Complex signatures of genomic variation of two non-model marine species in a homogeneous environment. BMC Genomics 19, 1–16. https://doi.org/10.1186/s12864-018-4721-y

Nomikou, M., Janssen, A., Schraag, R., Sabelis, M.W., 2002. Phytoseiid Predators Suppress Populations of *Bemisia tabaci* on Cucumber Plants with Alternative Food. Exp. Appl. Acarol. 27, 57–68. https://doi.org/10.1023/A:1021559421344

Nomikou, M., Janssen, A., Schraag, R., Sabelis, M.W., 2001. Phytoseiid predators as potential biological control agents for *Bemisia tabaci*. Exp. Appl. Acarol. 25, 271–291. https://doi.org/10.1023/A:1017976725685

Nordling, L., 2018. African scientists call for more control of their continent's genomic data. Nature. https://doi.org/10.1038/d41586-018-04685-1

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., MacAs, J., 2013. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29, 792–793. https://doi.org/10.1093/bioinformatics/btt054

Nunney, L., 2003. Managing captive populations for release: a population-genetic perspective., in: Quality Control and Production of Biological Control Agents: Theory and Testing Procedures. CABI, Wallingford, pp. 73–87. https://doi.org/10.1079/9780851996882.0073

Obrycki, J.J., Krafsur, E.S., Bogran, C.E., Gomez, L.E., Cave, R.E., 2001. Comparative Studies of Three Populations of the Lady Beetle Predator *Hippodamia convergens* (Coleoptera: Coccinellidae). Florida Entomol. 84, 55. https://doi.org/10.2307/3496663

Ogungbure, A.A., 2013. African indigenous knowledge: scientific or unscientific? Inkanyiso J. Humanit. Soc. Sci. 5, 12–20.

Onstad, D.W., Knolhoff, L.M., 2009. Finding the Economics in Economic Entomology. J. Econ. Entomol. 102, 1–7. https://doi.org/10.1603/029.102.0101

OpenWetWare Contributers, 2013. Luckau Protocols: Scoring [WWW Document]. OpenWetWare. URL https://openwetware.org/mediawiki/index.php?title=Luckau_

Protocols:Scoring&oldid=700763 (accessed 11.30.19).

Orzack, S.H., Gladstone, J., 1994. Quantitative genetics of sex ratio traits in the parasitic wasp, *Nasonia vitripennis*. Genetics 137, 211–220.

Orzack, S.H., Parker, E.D., Gladstone, J., 1991. The comparative biology of genetic variation for conditional sex ratio behavior in a parasitic wasp, *Nasonia vitripennis*. Genetics 127, 583–599.

Panfilio, K.A., Angelini, D.R., 2018. By land, air, and sea: hemipteran diversity through the genomic lens. Curr. Opin. Insect Sci. 25, 106–115. https://doi.org/10.1016/j.cois.2017.12.005

Panfilio, K.A., Vargas Jentzsch, I.M., Benoit, J.B., Erezyilmaz, D., Suzuki, Y., Colella, S., Robertson, H.M., Poelchau, M.F., Waterhouse, R.M., Ioannidis, P., Weirauch, M.T., Hughes, D.S.T., Murali, S.C., Werren, J.H., Jacobs, C.G.C., Duncan, E.J., Armisén, D., Vreede, B.M.I., Baa-Puyoulet, P., Berger, C.S., Chang, C., Chao, H., Chen, M.-J.M., Chen, Y.-T., Childers, C.P., Chipman, A.D., Cridge, A.G., Crumière, A.J.J., Dearden, P.K., Didion, E.M., Dinh, H., Doddapaneni, H.V., Dolan, A., Dugan, S., Extavour, C.G., Febvay, G., Friedrich, M., Ginzburg, N., Han, Y., Heger, P., Holmes, C.J., Horn, T., Hsiao, Y., Jennings, E.C., Johnston, J.S., Jones, T.E., Jones, J.W., Khila, A., Koelzer, S., Kovacova, V., Leask, M., Lee, S.L., Lee, C.-Y., Lovegrove, M.R., Lu, H., Lu, Y., Moore, P.J., Munoz-Torres, M.C., Muzny, D.M., Palli, S.R., Parisot, N., Pick, L., Porter, M.L., Qu, J., Refki, P.N., Richter, R., Rivera-Pomar, R., Rosendale, A.J., Roth, S., Sachs, L., Santos, M.E., Seibert, J., Sghaier, E., Shukla, J.N., Stancliffe, R.J., Tidswell, O., Traverso, L., van der Zee, M., Viala, S., Worley, K.C., Zdobnov, E.M., Gibbs, R.A., Richards, S., 2019. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. Genome Biol. 20, 64. https://doi.org/10.1186/s13059-019-1660-0

Pannebakker, B.A., Halligan, D.L., Reynolds, K.T., Ballantyne, G.A., Shuker, D.M., Barton, N.H., West, S.A., 2008. Effects of spontaneous mutation accumulation on sex ratio traits in a parasitoid wasp. Evolution (N. Y). 62, 1921–1935. https://doi.org/10.1111/j.1558-5646.2008.00434.x

Pannebakker, B.A., Niehuis, O., Hedley, A., Gadau, J., Shuker, D.M., 2010. The distribution of microsatellites in the *Nasonia* parasitoid wasp genome. Insect Mol. Biol. 19, 91–98. https://doi.org/10.1111/j.1365-2583.2009.00915.x

Pannebakker, B.A., Pijnacker, L.P., Zwaan, B.J., Beukeboom, L.W., 2004. Cytology of *Wolbachia*-induced parthenogenesis in *Leptopilina clavipes* (Hymenoptera: Figitidae). Genome 303, 299–303. https://doi.org/10.1139/G03-137

Paolucci, S., Van de Zande, L., Beukeboom, L.W., 2013. Adaptive latitudinal cline of photoperiodic diapause induction in the parasitoid *Nasonia vitripennis* in Europe. J. Evol. Biol. 26, 705–718. https://doi.org/10.1111/jeb.12113

Papageorgiou, L., Eleni, P., Raftopoulou, S., Mantaiou, M., Megalooikonomou, V., Vlachakis, D., 2018. Genomic big data hitting the storage bottleneck. EMBnet.journal 24, 910. https://doi.org/10.14806/ej.24.0.910

Papanicolaou, A., 2016. The life cycle of a genome project: perspectives and guidelines inspired by insect genome projects [version 1; peer review: 2 approved, 1 approved with

reservations]. F1000Research 5, 18. https://doi.org/10.12688/f1000research.7559.1

Parker, E.D., Orzack, S.H., 1985. Genetic variation for the sex ratio in *Nasonia vitripennis*. Genetics 110, 93–105.

Parreño, M.A., Ivanovic, A., Petrovic, A., Zikic, V., Tomanovic, Z., Vorburger, C., 2017. Wing shape as a taxonomic trait: Separating genetic variation from host-induced plasticity in aphid parasitoids. Zool. J. Linn. Soc. 180, 288–297. https://doi.org/10.1111/zoj.12490

Paspati, A., Ferguson, K.B., Verhulst, E.C., Urbaneja, A., González-Cabrera, J., Pannebakker, B.A., 2019. Effect of mass rearing on the genetic diversity of the predatory mite *Amblyseius swirskii* Athias-Henriot (Acari: Phytoseiidae). Entomol. Exp. Appl. 167, 670–681. https://doi.org/10.1111/eea.12811

Peakall, R., Smouse, P.E., 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Peakall, R., Smouse, P.E., 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pedigo, L., 1986. Economic Injury Levels in Theory and Practice. Annu. Rev. Entomol. 31, 341–368. https://doi.org/10.1146/annurev.ento.31.1.341

Peng, R.K., Christian, K., Gibb, K., 2004. Implementing ant technology in commercial cashew plantations and continuation of transplanted green ant colony monitoring, Rural Industries Research and Development Corporation.

Pérez-Hedo, M., Arias-Sanguino, Á.M., Urbaneja, A., 2018. Induced tomato plant resistance against *Tetranychus urticae* triggered by the phytophagy of *Nesidiocoris tenuis*. Front. Plant Sci. 9, 1–8. https://doi.org/10.3389/fpls.2018.01419

Pérez-Hedo, M., Urbaneja-Bernat, P., Jaques, J.A., Flors, V., Urbaneja, A., 2015. Defensive plant responses induced by *Nesidiocoris tenuis* (Hemiptera: Miridae) on tomato plants. J. Pest Sci. (2004). 88, 543–554. https://doi.org/10.1007/s10340-014-0640-0

Pérez-Hedo, M., Urbaneja, A., 2016. The Zoophytophagous Predator *Nesidiocoris tenuis*: A Successful But Controversial Biocontrol Agent in Tomato Crops, in: Advances in Insect Control and Resistance Management. Springer International Publishing, Cham, pp. 121–138. https://doi.org/10.1007/978-3-319-31800-4_7

Pérez-Maluf, R., Rafalimanana, H., Campan, E., Fleury, F., Kaiser, L., 2008. Differentiation of innate but not learnt responses to host-habitat odours contributes to rapid host finding in a parasitoid genotype. Physiol. Entomol. 33, 226–232. https://doi.org/10.1111/j.1365-3032.2008.00636.x

Phiri, S., Schoustra, S.E., van den Heuvel, J., Smid, E.J., Shindano, J., Linnemann, A., 2019. Fermented cereal-based Munkoyo beverage: Processing practices, microbial diversity and aroma compounds. PLoS One 14, e0223501. https://doi.org/10.1371/journal.pone.0223501

Pigliucci, M., Boudry, M., 2011. Why Machine-Information Metaphors are Bad for Science and Science Education. Sci. Educ. 20, 453–471. https://doi.org/10.1007/s11191-010-9267-6

Pilkington, L.J., Messelink, G.J., van Lenteren, J.C., Le Mottee, K., 2010. "Protected Biological Control" – Biological pest management in the greenhouse industry. Biol. Control 52, 216–220. https://doi.org/10.1016/j.biocontrol.2009.05.022

Pita, S., Panzera, F., Mora, P., Vela, J., Palomeque, T., Lorite, P., 2016. The presence of the ancestral insect telomeric motif in kissing bugs (Triatominae) rules out the hypothesis of its loss in evolutionarily advanced Heteroptera (Cimicomorpha). Comp. Cytogenet. 10, 427–437. https://doi.org/10.3897/compcytogen.v10i3.9960

Plouvier, W.N., Wajnberg, E., 2018. Improving the efficiency of augmentative biological control with arthropod natural enemies: A modeling approach. Biol. Control 125, 121–130. https://doi.org/10.1016/j.biocontrol.2018.05.010

Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.Y., Lin, H., Lin, J.W., Hackett, K., 2015. The i5k Workspace@NAL-enabling genomic data access, visualization and curation of arthropod genomes. Nucleic Acids Res. 43, D714–D719. https://doi.org/10.1093/nar/gku983

Poelchau, M.F., Coates, B.S., Childers, C.P., Peréz De León, A.A., Evans, J.D., Hackett, K., Shoemaker, D.W., 2016. Agricultural applications of insect ecological genomics. Curr. Opin. Insect Sci. 13, 61–69. https://doi.org/10.1016/j.cois.2015.12.002

Polaszek, A., 2009. Species Diversity and Host Associations of Trichogramma in Eurasia, in: Consoli, F.L., Parra, J.R.P., Zucchi, R.A. (Eds.), Egg Parasitoids in Agroecosystems with Emphasis on *Trichogramma*. Springer Netherlands, Dordrecht, pp. 237–266. https://doi.org/10.1007/978-1-4020-9110-0_9

Pompanon, F., Fouillet, P., Bouletreau, M., 1994. Locomotor behaviour in females of two *Trichogramma* species: description and genetic variability. Nor. J. Agric. Sci. 185–190.

Poorjavad, N., Goldansaz, S.H., Machtelinckx, T., Tirry, L., Stouthamer, R., van Leeuwen, T., 2012. Iranian *Trichogramma*: ITS2 DNA characterization and natural *Wolbachia* infection. BioControl 57, 361–374. https://doi.org/10.1007/s10526-011-9397-z

Poorjavad, N., Goldansaz, S.H., van Leeuwen, T., 2018. Fertility life table parameters, COI sequences and *Wolbachia* infection in populations of *Trichogramma brassicae* collected from *Chilo suppressalis*. Bull. Insectology 71, 89–96.

Poppy, G.M., Powell, W., 2009. Genetic manipulation of natural enemies: can we improve biological control by manipulating the parasitoid and/or the plant?, in: Genetics, Evolution and Biological Control. CABI, Wallingford, pp. 219–233. https://doi.org/10.1079/9780851997353.0219

Prado, S.S., Almeida, R.P.P., 2009. Phylogenetic placement of pentatomid stink bug gut symbionts. Curr. Microbiol. 58, 64–69. https://doi.org/10.1007/s00284-008-9267-9

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Puentes, A., Stephan, J.G., Björkman, C., 2018. A Systematic Review on the Effects of Plant-Feeding by Omnivorous Arthropods: Time to Catch-Up With the Mirid-Tomato Bias? Front. Ecol. Evol. 6. https://doi.org/10.3389/fevo.2018.00218

Puttarudriah, M., Basavanna, G.P.C., 1956. A study on the identity of *Bracon hebetor* Say and *Bracon brevicornis* Wesmael (Hymenoptera: Braconidae). Bull. Entomol. Res. 47, 183–191. https://doi.org/10.1017/S0007485300046617

Quan, Q., Hu, X., Pan, B., Zeng, B., Wu, N., Fang, G., Cao, Y., Chen, X., Li, X., Huang, Y., Zhan, S., 2019. Draft genome of the cotton aphid *Aphis gossypii*. Insect Biochem. Mol. Biol. 105, 25–32. https://doi.org/10.1016/j.ibmb.2018.12.007

R Core Team, 2015. R Development Core Team. R A Lang. Environ. Stat. Comput. 55, 275–286.

Rahimi-Kaldeh, S., Ashouri, A., Bandani, A., 2018. Does *Wolbachia* Infection Change the Overwintering Ability of *Trichogramma brassicae* (Hymenoptera: Trichogrammatidae)? Neotrop. Entomol. 47, 583–590. https://doi.org/10.1007/s13744-017-0549-4

Ras, E., Beukeboom, L.W., Cáceres, C., Bourtzis, K., 2017. Review of the role of gut microbiota in mass rearing of the olive fruit fly, *Bactrocera oleae*, and its parasitoids. Entomol. Exp. Appl. 164, 237–256. https://doi.org/10.1111/eea.12609

Rasmussen, L.B., Jensen, K., Sørensen, J.G., Sverrisdóttir, E., Nielsen, K.L., Overgaard, J., Holmstrup, M., Kristensen, T.N., 2018. Are commercial stocks of biological control agents genetically depauperate? – A case study on the pirate bug *Orius majusculus* Reuter. Biol. Control 127, 31–38. https://doi.org/10.1016/j.biocontrol.2018.08.016

Reddiex, A.J., Allen, S.L., Chenoweth, S.F., 2018. A Genomic Reference Panel for *Drosophila serrata*. G3 Genes, Genomes, Genet. 8, 1335–1346. https://doi.org/10.1534/g3.117.300487

Reeves, R.G., Denton, J.A., Santucci, F., Bryk, J., Reed, F.A., 2012. Scientific standards and the regulation of genetically modified insects. PLoS Negl. Trop. Dis. 6. https://doi.org/10.1371/journal.pntd.0001502

Retamal, R., Zaviezo, T., Malausa, T., Fauvergue, X., Goff, I. Le, Toleubayev, K., 2016. Genetic analyses and occurrence of diploid males in field and laboratory populations of *Mastrus ridens* (Hymenoptera: Ichneumonidae), a parasitoid of the codling moth. Biol. Control.

Reynolds, J., Weir, B.S., Cockerham, C.C., 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105.

Richards, S., 2019. Arthropod Genome Sequencing and Assembly Strategies, in: Brown, S.J., Pfrender, M.E. (Eds.), Methods in Molecular Biology. Springer Science + Business Media, pp. 1–14. https://doi.org/10.1007/978-1-4939-8775-7_1

Richards, S., 2015. It's more than stamp collecting: how genome sequencing can unify biological research. Trends Genet. 31, 411–421. https://doi.org/10.1016/j.tig.2015.04.007

Richards, S., Gibbs, R.A., Gerardo, N.M., Moran, N., Nakabachi, A., Stern, D., Tagu, D., Wilson, A.C.C., Muzny, D., Kovar, C., Cree, A., Chacko, J., Chandrabose, M.N., Dao, M.D., Dinh, H.H., Gabisi, R.A., Hines, S., Hume, J., Jhangian, S.N., Joshi, V., Lewis, L.R., Liu, Y.S., Lopez, J., Morgan, M.B., Nguyen, N.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Wright, R.A., Fowler,

G.R., Hitchens, M.E., Lozado, R.J., Moen, C., Steffen, D., Warren, J.T., Zhang, J., Nazareth, L. V., Chavez, D., Davis, C., Lee, S.L., Patel, B.M., Pu, L.L., Bell, S.N., Johnson, A.J., Vattathil, S., Williams, R.L., Shigenobu, S., Dang, P.M., Morioka, M., Fukatsu, T., Kudo, T., Miyagishima, S.Y., Jiang, H., Worley, K.C., Legeai, F., Gauthier, J.P., Collin, O., Zhang, L., Chen, H.C., Ermolaeva, O., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Maglott, D., Murphy, T., Pruitt, K., Sapojnikov, V., Souvorov, A., Thibaud-Nissen, F., Câmara, F., Guigó, R., Stanke, M., Solovyev, V., Kosarev, P., Gilbert, D., Gabaldón, T., Huerta-Cepas, J., Marcet-Houben, M., Pignatelli, M., Moya, A., Rispe, C., Ollivier, M., Quesneville, H., Permal, E., Llorens, C., Futami, R., Hedges, D., Robertson, H.M., Alioto, T., Mariotti, M., Nikoh, N., McCutcheon, J.P., Burke, G., Kamins, A., Latorre, A., Ashton, P., Calevro, F., Charles, H., Colella, S., Douglas, A.E., Jander, G., Jones, D.H., Febvay, G., Kamphuis, L.G., Kushlan, P.F., Macdonald, S., Ramsey, J., Schwartz, J., Seah, S., Thomas, G., Vellozo, A., Cass, B., Degnan, P., Hurwitz, B., Leonardo, T., Koga, R., Altincicek, B., Anselme, C., Atamian, H., Barribeau, S.M., De Vos, M., Duncan, E.J., Evans, J., Ghanim, M., Heddi, A., Kaloshian, I., Vincent-Monegat, C., Parker, B.J., Pérez-Brocal, V., Rahbé, Y., Spragg, C.J., Tamames, J., Tamarit, D., Tamborindeguy, C., Vilcinskas, A., Bickel, R.D., Brisson, J.A., Butts, T., Chang, C.C., Christiaens, O., Davis, G.K., Duncan, E., Ferrier, D., Iga, M., Janssen, R., Lu, H.L., McGregor, A., Miura, T., Smagghe, G., Smith, J., Van Der Zee, M., Velarde, R., Wilson, M., Dearden, P., Edwards, O.R., Gordon, K., Hilgarth, R.S., Rider, S.D., Srinivasan, D., Walsh, T.K., Ishikawa, A., Jaubert-Possamai, S., Fenton, B., Huang, W., Rizk, G., Lavenier, D., Nicolas, J., Smadja, C., Zhou, J.J., Vieira, F.G., He, X.L., Liu, R., Rozas, J., Field, L.M., Campbell, P., Carolan, J.C., Fitzroy, C.I.J., Reardon, K.T., Reeck, G.R., Singh, K., Wilkinson, T.L., Huybrechts, J., Abdel-Latief, M., Robichon, A., Veenstra, J.A., Hauser, F., Cazzamali, G., Schneider, M., Williamson, M., Stafflinger, E., Hansen, K.K., Grimmelikhuijzen, C.J.P., Price, D.R.G., Caillaud, M., Van Fleet, E., Ren, Q., Gatehouse, J.A., Brault, V., Monsion, B., Diaz, J., Hunnicutt, L., Ju, H.J., Pechuan, X., Aguilar, J., Cortés, T., Ortiz-Rivas, B., Martínez-Torres, D., Dombrovsky, A., Dale, R.P., Davies, T.G.E., Williamson, M.S., Jones, A., Sattelle, D., Williamson, S., Wolstenholme, A., Cottret, L., Sagot, M.F., Heckel, D.G., Hunter, W., 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8. https://doi.org/10.1371/journal.pbio.1000313

Richards, S., Murali, S.C., 2015. Best practices in insect genome sequencing: what works and what doesn't. Curr. Opin. Insect Sci. 7, 1–7. https://doi.org/10.1016/j.cois.2015.02.013

Riddick, E.W., Simmons, A.M., 2014. Do plant trichomes cause more harm than good to predatory insects? Pest Manag. Sci. 70, 1655–1665. https://doi.org/10.1002/ps.3772

Rigby, D., Cáceres, D., 2001. Organic farming and the sustainability of agricultural systems. Agric. Syst. 68, 21–40. https://doi.org/10.1016/S0308-521X(00)00060-3

Robinson, G.E., Hackett, K.J., Purcell-Miramontes, M., Brown, S.J., Evans, J.D., Goldsmith, M.R., Lawson, D., Okamuro, J., Robertson, H.M., Schneider, D.J., 2011. Creating a buzz about insect genomes. Science 331, 1386. https://doi.org/10.1126/science.331.6023.1386

Roderick, G.K., Navajas, M., 2003. Genes in new environments: Genetics and evolution in biological control. Nat. Rev. Genet. 4, 889–899. https://doi.org/10.1038/nrg1201

Rodríguez, J.R., Pintureau, B., Galán, M., 1994. Déterminisme de la couleur des hôtes parasités par *Trichogramma fuentesi*. Entomol. Exp. Appl. 70, 121–128. https://doi.org/10.1111/j.1570-7458.1994.tb00740.x

Rössler, Y., 1975. Reproductive Differences between Laboratory-reared and Field-collected Populations of the Mediterranean Fruitfly, *Ceratitis capitata*. Ann. Entomol. Soc. Am. 68, 987–991. https://doi.org/10.1093/aesa/68.6.987

Routray, S., Dey, D., Baral, S., Das, A.P., Mahantheshwara, B., 2016. Genetic improvement of natural enemies: A review. Agric. Rev. 37. https://doi.org/10.18805/ag.v37i4.6463

Rozen, S., Skaletsky, H., 1999. Primer3 on the WWW for General Users and for Biologist Programmers, in: Misener, S., Krawetz, S.A. (Eds.), Bioinformatics Methods and Protocols. Humana Press, New Jersey, pp. 365–386. https://doi.org/10.1385/1-59259-192-2:365

RStudio Team, 2015. RStudio: Integrated Development for R.

Rugman-Jones, P.F., Stouthamer, R., 2017. High-resolution melt analysis without DNA extraction affords rapid genotype resolution and species identification. Mol. Ecol. Resour. 17, 598–607. https://doi.org/10.1111/1755-0998.12599

Sabater-Muñoz, B., Pascual-Ruiz, S., Gómez-Martínez, M.A., Jacas, J.A., Hurtado, M.A., 2012. Isolation and characterization of polymorphic microsatellite markers in *Tetranychus urticae* and cross amplification in other Tetranychidae and Phytoseiidae species of economic importance. Exp. Appl. Acarol. 57, 37–51. https://doi.org/10.1007/s10493-012-9529-x

Sachman-Ruiz, B., Quiroz-Castañeda, R.E., 2018. Genomics of Rickettsiaceae: An Update, in: Farm Animals Diseases, Recent Omic Trends and New Strategies of Treatment. InTech, p. 13. https://doi.org/10.5772/intechopen.74563

Sahara, K., Marec, F., Traut, W., 1999. TTAGG telomeric repeats in chromosomes of some insects and other arthropods. Chromosom. Res. 7, 449–460. https://doi.org/10.1023/A:1009297729547

Saini, A., 2019. Superior: the return of race science. Harper Collins Publishers Ltd, London.

Sanchez, J.A., 2009. Density thresholds for *Nesidiocoris tenuis* (Heteroptera: Miridae) in tomato crops. Biol. Control 51, 493–498. https://doi.org/10.1016/j.biocontrol.2009.09.006

Sanchez, J.A., La Spina, M., Perera, O.P., 2012. Analysis of the population structure of *Macrolophus pygmaeus* (Rambur) (Hemiptera: Miridae) in the Palaearctic region using microsatellite markers. Ecol. Evol. 2, 3145–3159. https://doi.org/10.1002/ece3.420

Sandrock, C., Gouskov, A., Vorburger, C., 2010. Ample genetic variation but no evidence for genotype specificity in an all-parthenogenetic host-parasitoid interaction. J. Evol. Biol. 23, 578–585. https://doi.org/10.1111/j.1420-9101.2009.01925.x

Santos-Garcia, D., Silva, F.J., Morin, S., Dettner, K., Kuechler, S.M., 2017. The all-rounder sodalis: A new bacteriome-associated endosymbiont of the lygaeoid bug *Henestaris halophilus* (Heteroptera: Henestarinae) and a critical examination of its evolution. Genome Biol. Evol. 9, 2893–2910. https://doi.org/10.1093/gbe/evx202

Sayyed, A.H., Pathan, A.K., Faheem, U., 2010. Cross-resistance, genetics and stability of resistance to deltamethrin in a population of *Chrysoperla carnea* from Multan, Pakistan. Pestic. Biochem. Physiol. 98, 325–332. https://doi.org/10.1016/j.pestbp.2010.07.004

Schlötterer, C., Tobler, R., Kofler, R., Nolte, V., 2014. Sequencing pools of individuals-mining

genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15, 749–763. https://doi.org/10.1038/nrg3803

Schmieder, S., Colinet, D., Poirié, M., 2012. Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. Nat. Commun. 3, 895. https://doi.org/10.1038/ncomms1898

Schoustra, S.E., 2019. Zambian traditional fermented foods [WWW Document]. Food Bus. Knowl. Platf. URL https://knowledge4food.net/research-project/gcp1-zambian-traditional-fermented-foods/ (accessed 12.13.19).

Schwarz, A., Medrano-Mercado, N., Schaub, G.A., Struchiner, C.J., Bargues, M.D., Levy, M.Z., Ribeiro, J.M.C., 2014. An Updated Insight into the Sialotranscriptome of *Triatoma infestans*: Developmental Stage and Geographic Variations. PLoS Negl. Trop. Dis. 8. https://doi.org/10.1371/journal.pntd.0003372

Sember, A., Bertollo, L.A.C., Ráb, P., Yano, C.F., Hatanaka, T., de Oliveira, E.A., Cioffi, M. de B., 2018. Sex chromosome evolution and genomic divergence in the fish *Hoplias malabaricus* (Characiformes, Erythrinidae). Front. Genet. 9, 1–12. https://doi.org/10.3389/fgene.2018.00071

Sepúlveda, D.A., Zepeda-Paulo, F., Ramírez, C.C., Lavandero, B., Figueroa, C.C., 2017. Loss of host fidelity in highly inbred populations of the parasitoid wasp *Aphidius ervi* (Hymenoptera: Braconidae). J. Pest Sci. (2004). 90, 649–658. https://doi.org/10.1007/s10340-016-0798-8

Sequeira, R., Mackauer, M., 1992. Quantitative genetics of body size and development time in the parasitoid wasp *Aphidius ervi* (Hymenoptera: Aphidiidae). Can. J. Zool. 70, 1102–1108. https://doi.org/10.1139/z92-154

Sethuraman, A., Janzen, F.J., Obrycki, J., 2015. Population genetics of the predatory lady beetle *Hippodamia convergens*. Biol. Control 84, 1–10. https://doi.org/10.1016/j.biocontrol.2015.01.002

Shuker, D.M., Phillimore, A.J., Burton-Chellew, M.N., Hodge, S.E., West, S.A., 2007. The quantitative genetic basis of polyandry in the parasitoid wasp, *Nasonia vitripennis*. Heredity (Edinb). 98, 69–73. https://doi.org/10.1038/sj.hdy.6800897

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Skalski, G.T., Couch, C.R., Garber, A.F., Weir, B.S., Sullivan, C. V, 2006. Evaluation of DNA pooling for the estimation of microsatellite allele frequencies: a case study using striped bass (*Morone saxatilis*). Genetics 173, 863–75. https://doi.org/10.1534/genetics.105.053702

Smith, D.R., 2016. Goodbye genome paper, hello genome report: the increasing popularity of 'genome announcements' and their impact on science 16, 156–162. https://doi.org/10.1093/bfgp/elw026

Smith, S.M., 1996. Biological control with *Trichogramma*: advances, successes, and potential of their use. Annu. Rev. Entomol. 41, 375–406. https://doi.org/10.1146/annurev.ento.41.1.375

Sobreira, T.J.P., Durham, A.M., Gruber, A., 2006. TRAP: Automated classification, quantification and annotation of tandemly repeated sequences. Bioinformatics 22, 361–362. https://doi.org/10.1093/bioinformatics/bti809

Sochorová, J., Garcia, S., Gálvez, F., Symonová, R., Kovařík, A., 2018. Evolutionary trends in animal ribosomal DNA loci: introduction to a new online database. Chromosoma 127, 141–150. https://doi.org/10.1007/s00412-017-0651-8

Sogbohossou, E.O.D., 2019. Orphan no more: Ethnobotany and genetic analysis of leaf yield and secondary metabolites content in *Gynandropsis gynandra* (Cleomaceae). Wageningen University. https://doi.org/10.18174/501019

Sogbohossou, E.O.D., Achigan-Dako, E.G., Maundu, P., Solberg, S., Deguenon, E.M.S., Mumm, R.H., Hale, I., Van Deynze, A., Schranz, M.E., 2018. A roadmap for breeding orphan leafy vegetable species: A case study of *Gynandropsis gynandra* (Cleomaceae). Hortic. Res. 5, 2. https://doi.org/10.1038/s41438-017-0001-2

Solovyev, V., 2007. Statistical Approaches in Eukaryotic Gene Prediction, in: Balding, D., Cannings, C., Bishop, M. (Eds.), Handbook of Statistical Genetics, Wiley-Interscience. pp. 97–159. https://doi.org/doi:10.1002/9780470061619.ch4

Sorati, M., Newman, M., Hoffmann, A.A., 1996. Inbreeding and incompatibility in *Trichogramma* nr. *brassicae*: Evidence and implications for quality control. Entomol. Exp. Appl. 78, 283–290. https://doi.org/10.1111/j.1570-7458.1996.tb00792.x

Sørensen, J.G., Addison, M.F., Terblanche, J.S., 2012. Mass-rearing of insects for pest management: Challenges, synergies and advances from evolutionary physiology. Crop Prot. 38, 87–94. https://doi.org/10.1016/j.cropro.2012.03.023

Speicher, B.R., Speicher, K.G., 1940. The Occurrence of Diploid Males in *Habrobracon brevicornis*. Am. Nat. 74, 379–382. https://doi.org/10.1086/280904

Srinivasan, T., Chandrikamohan, 2017. Population growth potential of *Bracon brevicornis* Wesmael (Braconidae: Hymenoptera): A life table analysis. Acta Phytopathol. Entomol. Hungarica 52, 123–130. https://doi.org/10.1556/038.52.2017.010

St. Leger, R.J., Wang, C., 2010. Genetic engineering of fungal biocontrol agents to achieve greater efficacy against insect pests. Appl. Microbiol. Biotechnol. 85, 901–907. https://doi.org/10.1007/s00253-009-2306-z

Stahl, J.M., Babendreier, D., Haye, T., 2019a. Life history of *Anastatus bifasciatus*, a potential biological control agent of the brown marmorated stink bug in Europe. Biol. Control 129, 178–186. https://doi.org/10.1016/j.biocontrol.2018.10.016

Stahl, J.M., Babendreier, D., Haye, T., 2018. Using the egg parasitoid *Anastatus bifasciatus* against the invasive brown marmorated stink bug in Europe: can non-target effects be ruled out? J. Pest Sci. (2004). 91, 1005–1017. https://doi.org/10.1007/s10340-018-0969-x

Stahl, J.M., Babendreier, D., Marazzi, C., Caruso, S., Costi, E., Maistrello, L., Haye, T., 2019b. Can *Anastatus bifasciatus* be used for augmentative biological control of the brown marmorated stink bug in fruit orchards? Insects 10, 108. https://doi.org/10.3390/insects10040108

Stahl, J.M., Gariepy, T.D., Beukeboom, L.W., Haye, T., 2019c. A molecular tool to identify *Anastatus* parasitoids of the brown marmorated stink bug. Entomol. Exp. Appl. eea.12809. https://doi.org/10.1111/eea.12809

Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33, W465–W467. https://doi.org/10.1093/nar/gki458

Stansly, P.A., Sánchez, P.A., Rodrıguez, J.M., Cañizares, F., Nieto, A., Leyva, M.J.L., Fajardo, M., Suárez, V., Urbaneja, A., 2004. Prospects for biological control of *Bemisia tabaci* (Homoptera, Aleyrodidae) in greenhouse tomatoes of southern Spain. Crop Prot. 23, 701–712. https://doi.org/10.1016/J.CROPRO.2003.11.016

Stenberg, J.A., 2017. A Conceptual Framework for Integrated Pest Management. Trends Plant Sci. 22, 759–769. https://doi.org/10.1016/j.tplants.2017.06.010

Stern, V.M., 1973. Economic Thresholds. Annu. Rev. Entomol. 18, 259–280. https://doi.org/10.1146/annurev.en.18.010173.001355

Stouthamer, R., 2018. Improving the genetics of natural enemies for biological control with lessons not yet learned from the last century, in: XI European Congress of Entomology. p. C0081.

Stouthamer, R., 1997. *Wolbachia*-induced parthenogenesis, in: O'Neill, S.L., Hoffmann, A.A., Werren, J.H. (Eds.), Influential Passengers: Inherited Microorganisms and Arthropod Reproduction. Oxford University Press, Oxford, UK, pp. 102–124.

Stouthamer, R., Hu, J., Van Kan, F.J.P.M.P.M., Platner, G.R., Pinto, J.D., 1999. The utility of internally transcribed spacer 2 DNA sequences of the nuclear ribosomal gene for distinguishing sibling species of *Trichogramma*. BioControl 43, 421–440. https://doi.org/10.1023/A:1009937108715

Stouthamer, R., Huigens, M., 2003. Parthenogenesis Associated With *Wolbachia*, in: Bourtzis, K., Miller, T.A. (Eds.), Insect Symbiosis. CRC Press, Boca Raton, pp. 247–266. https://doi.org/10.1201/9780203009918.ch15

Stouthamer, R., Kazmer, D.J., 1994. Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in *Trichogramma* wasps. Heredity (Edinb). 73, 317–327. https://doi.org/10.1038/hdy.1994.139

Stouthamer, R., Luck, R.F., Werren, J.H., 1992. Genetics of sex determination and the improvement of biological control using parasitoids. Environ. Entomol. 21, 427–435. https://doi.org/10.1093/ee/21.3.427

Stouthamer, R., Nunney, L., 2014. Can Neutral Molecular Markers be Used to Determine the Success of an Introduction of a Better Strain Into an Established Population of a Biocontrol Parasitoid? J. Econ. Entomol. 107, 483–495. https://doi.org/10.1603/EC13444

Streito, J.C., Clouet, C., Hamdi, F., Gauthier, N., 2017. Population genetic structure of the biological control agent *Macrolophus pygmaeus* in Mediterranean agroecosystems. Insect Sci. 24, 859–876. https://doi.org/10.1111/1744-7917.12370

Stuart, R.J., Gaugler, R., 1996. Genetic adaptation and founder effect in laboratory populations of the entomopathogenic nematode *Steinernema glaseri*. Can. J. Zool. 74, 164–170. https://doi.org/10.1139/z96-021

Sumer, F., Tuncbilek, A.S., Oztemiz, S., Pintureau, B., Rugman-Jones, P., Stouthamer, R., 2009. A molecular key to the common species of *Trichogramma* of the Mediterranean region. BioControl 54, 617–624. https://doi.org/10.1007/s10526-009-9219-8

Sunnucks, P., Hales, D.F., 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Mol. Biol. Evol. 13, 510–524. https://doi.org/10.1093/oxfordjournals.molbev.a025612

Suverkropp, B.P., Bigler, F., van Lenteren, J.C., 2010. Movement and host finding of *Trichogramma brassicae* on maize plants. Bull. Insectology 63, 115–127.

Suverkropp, B.P., Bigler, F., van Lenteren, J.C., 2009. Dispersal behaviour of *Trichogramma brassicae* in maize fields. Bull. Insectology 62, 113–120.

Swirski, E., 1967. Laboratory studies on the feeding, development and reproduction of the predaceous mites *Amblyseius rubini* Swirski and Amitai and *Amblyseius swirskii* Athias (Acarina: Phytoseiidae) on various kinds of food substances. Isr. J. Agric. Res. 17, 101–119.

Swirski, E., Amitai, S., 1997. Annotated list of phytoseiid mites (Mesostigmata: Phytoseiidae) in Israel. Isr. J. Entomol. 31, 21–46.

Szűcs, M., Vercken, E., Bitume, E. V., Hufbauer, R.A., 2019. The implications of rapid eco-evolutionary processes for biological control - a review. Entomol. Exp. Appl. eea.12807. https://doi.org/10.1111/eea.12807

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–696.

Takezaki, N., Nei, M., Tamura, K., 2014. POPTREEW: Web Version of POPTREE for Constructing Population Trees from Allele Frequency Data and Computing Some Other Quantities. Mol. Biol. Evol. 31, 1622–1624. https://doi.org/10.1093/molbev/msu093

Tauber, C.A., Tauber, M.J., 1992. Phenotypic Plasticity in *Chrysoperla*: Genetic Variation in the Sensory Mechanism and in Correlated Reproductive Traits. Evolution (N. Y). 46, 1754. https://doi.org/10.2307/2410029

Tauber, C.A., Tauber, M.J., 1986. Genetic variation in all-or-none life-history traits of the lacewing *Chrysoperla carnea*. Can. J. Zool. 64, 1542–1544. https://doi.org/10.1139/z86-229

Tayeh, A., Estoup, A., Hufbauer, R.A., Ravigne, V., Goryacheva, I., Zakharov, I.A., Lombaert, E., Facon, B., 2013. Investigating the genetic load of an emblematic invasive species: The case of the invasive harlequin ladybird *Harmonia axyridis*. Ecol. Evol. 3, 864–871. https://doi.org/10.1002/ece3.490

Temerak, S.A., 1983a. Host preferences of the parasitoid *Bracon brevicornis* Wesmael (Hym., Braconidae) and host sensitivity to its venom. Zeitschrift für Angew. Entomol. 96, 37–41. https://doi.org/10.1111/j.1439-0418.1983.tb03638.x

Temerak, S.A., 1983b. Longevity of *Bracon brevicornis* [Hym: Braconidae] adults as influenced

by nourishment on artificial and natural foods. Entomophaga 28, 145–150. https://doi.org/10.1007/BF02372138

Teran, M.Y., 2016. The Nagoya Protocol and Indigenous peoples. Int. Indig. Policy J. 7. https://doi.org/10.18584/iipj.2016.7.2.6

Testa, A.C., Hane, J.K., Ellwood, S.R., Oliver, R.P., 2015. CodingQuarry: Highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics 16, 1–12. https://doi.org/10.1186/s12864-015-1344-4

Thiel, A., Weeda, A.C., 2014. Haploid, diploid, and triploid - Discrimination ability against polyploid mating partner in the Parasitic Wasp, *Bracon brevicornis* (Hymenoptera: Braconidae). J. Insect Sci. 14, 1–7. https://doi.org/10.1093/jisesa/ieu153

Thomas, G.W.C., Dohmen, E., Hughes, D.S.T., Murali, S.C., Poelchau, M., Glastad, K., Anstead, C.A., Ayoub, N.A., Batterham, P., Bellair, M., Binford, G.J., Chao, H., Chen, Y.H., Childers, C., Dinh, H., Doddapaneni, H.V., Duan, J.J., Dugan, S., Esposito, L.A., Friedrich, M., Garb, J., Gasser, R.B., Goodisman, M.A.D., Gundersen-Rindal, D.E., Han, Y., Handler, A.M., Hatakeyama, M., Hering, L., Hunter, W.B., Ioannidis, P., Jayaseelan, J.C., Kalra, D., Khila, A., Korhonen, P.K., Lee, C.E., Lee, S.L., Li, Y., Lindsey, A.R.I., Mayer, G., McGregor, A.P., McKenna, D.D., Misof, B., Munidasa, M., Munoz-Torres, M., Muzny, D.M., Niehuis, O., Osuji-Lacy, N., Palli, S.R., Panfilio, K.A., Pechmann, M., Perry, T., Peters, R.S., Poynton, H.C., Prpic, N.-M., Qu, J., Rotenberg, D., Schal, C., Schoville, S.D., Scully, E.D., Skinner, E., Sloan, D.B., Stouthamer, R., Strand, M.R., Szucsich, N.U., Wijeratne, A., Young, N.D., Zattara, E.E., Benoit, J.B., Zdobnov, E.M., Pfrender, M.E., Hackett, K.J., Werren, J.H., Worley, K.C., Gibbs, R.A., Chipman, A.D., Waterhouse, R.M., Bornberg-Bauer, E., Hahn, M.W., Richards, S., 2020. Gene content evolution in the arthropods. Genome Biol. 21, 15. https://doi.org/10.1186/s13059-019-1925-7

Thomson, L.J., Hoffmann, A.A., 2009. Laboratory Fecundity as Predictor of Field Success in *Trichogramma carverae* (Hymenoptera: Trichogrammatidae). J. Econ. Entomol. 95, 912–917. https://doi.org/10.1603/0022-0493-95.5.912

Thubru, D.P., Firake, D.M., Behere, G.T., 2018. Assessing risks of pesticides targeting lepidopteran pests in cruciferous ecosystems to eggs parasitoid, *Trichogramma brassicae* (Bezdenko). Saudi J. Biol. Sci. 25, 680–688. https://doi.org/10.1016/j.sjbs.2016.04.007

Tolman, T., 2008. Collins Field Guide for the Butterflies of Britain and Europe. Harper Collins Publishers Ltd, London.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–5. https://doi.org/10.1038/nbt.1621

Traut, W., 1976. Pachytene mapping in the female silkworm, *Bombyx mori* L. (Lepidoptera). Chromosoma 58, 275–284. https://doi.org/10.1007/BF00292094

Traut, W., Sahara, K., Otto, T.D., Marec, F., 1999. Molecular differentiation of sex chromosomes probed by comparative genomic hybridization. Chromosoma 108, 173–180. https://doi.org/10.1007/s004120050366

Traverse, K.L., Pardue, M.L., 1988. A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. Proc. Natl. Acad. Sci. U. S. A. 85, 8116–8120. https://doi.org/10.1073/pnas.85.21.8116

Triapitsyn, S. V., 2015. Problems regarding the taxonomy of some palaearctic species of *Trichogramma*: Background information, recent developments, and approaches to their solution, in: Vinson, S.B., Greenberg, S.M., Liu, T.X., Rao, A., Volosciuk, L.F. (Eds.), Biological Control of Pests Using *Trichogramma*: Current Status and Perspectives. Northwest A&F University Press, Xianyang, pp. 1–12.

Tvedte, E.S., Walden, K.K.O., McElroy, K.E., Werren, J.H., Forbes, A.A., Hood, G.R., Logsdon, J.M., Feder, J.L., Robertson, H.M., 2019. Genome of the parasitoid wasp *Diachasma alloeum*, an emerging model for ecological speciation and transitions to asexual reproduction. Genome Biol. Evol. 11, 2767–2773. https://doi.org/10.1093/gbe/evz205

Universiteit Leiden, 2017. "Decolonise the botanical treasure house" [WWW Document]. Univ. Leiden. URL https://www.universiteitleiden.nl/en/news/2017/01/decolonise-the-botanical-treasure-house (accessed 12.12.19).

Urbaneja-Bernat, P., Bru, P., González-Cabrera, J., Urbaneja, A., Tena, A., 2019. Reduced phytophagy in sugar-provisioned mirids. J. Pest Sci. (2004). 92, 1139–1148. https://doi.org/10.1007/s10340-019-01105-9

Urbaneja, A., González-Cabrera, J., Arnó, J., Gabarra, R., 2012. Prospects for the biological control of *Tuta absoluta* in tomatoes of the Mediterranean basin. Pest Manag. Sci. 68, 1215–1222. https://doi.org/10.1002/ps.3344

Utrecht University Centre for Global Challenges, 2019a. Decolonisation group- the continued colonialism platform (2019) [WWW Document]. Utr. Univ. URL https://www.uu.nl/en/organisation/centre-for-global-challenges/projects/projects-2019/decolonisation-group-the-continued-colonialism-platform-2019 (accessed 12.13.19).

Utrecht University Centre for Global Challenges, 2019b. The Decolonisation Group at Utrecht University [WWW Document]. Utr. Univ. URL https://www.decolonisationgroup.com/ (accessed 12.13.19).

Valentin, R.E., Nielsen, A.L., Wiman, N.G., Lee, D.H., Fonseca, D.M., 2017. Global invasion network of the brown marmorated stink bug, *Halyomorpha halys*. Sci. Rep. 7, 1–12. https://doi.org/10.1038/s41598-017-10315-z

van der Sijs, N., 2018. "Bloemlezing - (Anthologie)" [WWW Document]. Meertens Inst. Etymologie Bank. URL http://www.etymologiebank.nl/trefwoord/bloemlezing (accessed 11.27.19).

van der Woude, E., Smid, H.M., 2016. How to escape from haller's rule: Olfactory system complexity in small and large *Trichogramma evanescens* parasitic wasps. J. Comp. Neurol. 524, 1876–1891. https://doi.org/10.1002/cne.23927

van der Woude, E., Smid, H.M., Chittka, L., Huigens, M.E., 2013. Breaking Haller's Rule: Brain-Body Size Isometry in a Minute Parasitic Wasp. Brain. Behav. Evol. 81, 86–92. https://doi.org/10.1159/000345945

van Houten, Y.M., Ostlie, M.L., Hoogerbrugge, H., Bolckmans, K., 2005. Biological control of western flower thrips on sweet pepper using the predatory mites *Amblyseius cucumeris*, *Iphiseius degenerans*, *A. andersoni* and *A. swirskii*. IOBC/wprs Bull 28, 283–286.

van Houten, Y.M., van Stratum, P., Bruin, J., Veerman, A., 1995. Selection for non-diapause in *Amblyseius cucumeris* and *Amblyseius barkeri* and exploration of the effectiveness of selected strains for thrips control. Entomol. Exp. Appl. 77, 289–295. https://doi.org/10.1111/j.1570-7458.1995.tb02326.x

van Lenteren, J.C., 2012. The state of commercial augmentative biological control: plenty of natural enemies, but a frustrating lack of uptake. BioControl 57, 1–20. https://doi.org/10.1007/s10526-011-9395-1

van Lenteren, J.C., Bale, J., Bigler, F., Hokkanen, H.M.T., Loomans, A.J.M., 2006. Assessing Risks of Releasing Exotic Biological Control Agents of Arthropod Pests. Annu. Rev. Entomol. 51, 609–634. https://doi.org/10.1146/annurev.ento.51.110104.151129

van Lenteren, J.C., Bolckmans, K., Köhl, J., Ravensberg, W.J., Urbaneja, A., 2018. Biological control using invertebrates and microorganisms: plenty of new opportunities. BioControl 63, 39–59. https://doi.org/10.1007/s10526-017-9801-4

van Lenteren, J.C., Godfray, H.C.J., 2005. European science in the Enlightenment and the discovery of the insect parasitoid life cycle in The Netherlands and Great Britain. Biol. Control 32, 12–24. https://doi.org/10.1016/j.biocontrol.2004.08.009

van Nouhuys, S., 2016. Diversity, population structure, and individual behaviour of parasitoids as seen using molecular markers. Curr. Opin. Insect Sci. 14, 94–99. https://doi.org/10.1016/j.cois.2016.02.006

van Wilgenburg, E., Driessen, G., Beukeboom, L.W., 2006. Single locus complementary sex determination in Hymenoptera: an "unintelligent" design? Front. Zool. 3, 1. https://doi.org/ https://doi.org/10.1186/1742-9994-3-1

Van Zee, J., Black, W.C., Levin, M., Goddard, J., Smith, J., Piesman, J., 2013. High SNP density in the blacklegged tick, *Ixodes scapularis*, the principal vector of Lyme disease spirochetes. Ticks Tick. Borne. Dis. 4, 63–71. https://doi.org/10.1016/J.TTBDIS.2012.07.005

Vandekerkhove, B., De Puysseleyr, V., Bonte, M., De Clercq, P., 2011. Fitness and predation potential of *Macrolophus pygmaeus* reared under artificial conditions. Insect Sci. 18, 682–688. https://doi.org/10.1111/j.1744-7917.2011.01414.x

Vargas, R.I., Carey, J.R., 1989. Comparison of Demographic Parameters for Wild and Laboratory-Adapted Mediterranean Fruit Fly (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 82, 55–59. https://doi.org/10.1093/aesa/82.1.55

Vaser, R., Sovic, I., Nagarajan, N., Sikic, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737–746. https://doi.

org/10.1101/068122

Veltri, D., Wight, M.M., Crouch, J.A., 2016. SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. Nucleic Acids Res. 44, W41–W45. https://doi.org/10.1093/nar/gkw330

Venkatesan, T., Jalali, S.K., Srinivasamurthy, K., 2009. Competitive interactions between *Goniozus nephantidis* and *Bracon brevicornis*, parasitoids of the coconut pest *Opisina arenosella*. Int. J. Pest Manag. 55, 257–263. https://doi.org/10.1080/09670870902914155

Venkatesan, T., More, R.P., Baskar, R., Jalali, S.K., Lalitha, Y., Ballal, C.R., 2016. Differentiation of some indigenous and exotic trichogrammatids (Hymenoptera: Trichogrammatidae) from India based on Internal transcribed spacer-2 and cytochrome oxidase-I markers and their phylogenetic relationship. Biol. Control 101, 130–137. https://doi.org/10.1016/j.biocontrol.2016.07.005

Verhulst, E.C., van de Zande, L., Beukeboom, L.W., 2010. Insect sex determination: it all evolves around transformer. Curr. Opin. Genet. Dev. 20, 376–383. https://doi.org/10.1016/j.gde.2010.05.001

Vianen, A. van, Lenteren, J.C. van, 1986. The parasite-host relationship between *Encarsia formosa* (Gahan) (Hym., Aphelinidae) and *Trialeurodes vaporariorum* (Westwood) (Hom., Aleyrodidae). XIV. Genetic and environmental factors influencing body-size and number of ovarioles of *Encarsia formosa*. J. Appl. Entomol. 101, 321–331. https://doi.org/http://dx.doi.org/10.1111/j.1439-0418.1986.tb00866.x

Villacañas de Castro, C., Thiel, A., 2017. Resource-Dependent Clutch Size Decisions and Size-Fitness Relationships in a Gregarious Ectoparasitoid Wasp, *Bracon brevicornis*. J. Insect Behav. 30, 454–469. https://doi.org/10.1007/s10905-017-9632-2

Vince, N., 2019. Decolonising the Academy, the Curriculum and the History of Decolonisation [WWW Document]. NIAS. URL https://nias.knaw.nl/projects/decolonising-the-academy-the-curriculum-and-the-history-of-decolonisation/ (accessed 12.12.19).

Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Schatz, M.C., Gurtowski, J., Underwood, C.J., Vurture, G.W., Fang, H., 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics 33, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Wageningen University and Research, 2019. Traditional fermented foods to promote food and nutrition security in Africa [WWW Document]. Wageningen Univ. Res. URL https://www.wur.nl/en/project/Traditional-fermented-foods-to-promote-food-and-nutrition-security-in-Africa.htm (accessed 12.12.19).

Wagner, J.D., Glover, M.D., Moseley, J.B., Moore, A.J., 1999. Heritability and fitness consequences of cannibalism in *Harmonia axyridis*. Evol. Ecol. Res. 1, 375–388.

Wajnberg, E., 1993. Genetic variation in sex allocation in a parasitic wasp: variation in sex pattern within sequences of oviposition. Entomol. Exp. Appl. 69, 221–229. https://doi.org/10.1111/j.1570-7458.1993.tb01745.x

Wajnberg, E., Colazza, S., 1998. Genetic variability in the area searched by a parasitic wasp: Analysis from automatic video tracking of the walking path. J. Insect Physiol. 44, 437–444.

https://doi.org/10.1016/S0022-1910(98)00032-8

Wajnberg, E., Curty, C., Colazza, S., 2004. Genetic variation in the mechanisms of direct mutual interference in a parasitic wasp: Consequences in terms of patch-time allocation. J. Anim. Ecol. 73, 1179–1189. https://doi.org/10.1111/j.0021-8790.2004.00892.x

Wajnberg, E., Curty, C., Jervis, M., 2012. Intra-Population Genetic Variation in the Temporal Pattern of Egg Maturation in a Parasitoid Wasp. PLoS One 7, e45915. https://doi.org/10.1371/journal.pone.0045915

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9. https://doi.org/10.1371/journal.pone.0112963

Wallberg, A., Bunikis, I., Pettersson, O.V., Mosbech, M.-B., Childers, A.K., Evans, J.D., Mikheyev, A.S., Robertson, H.M., Robinson, G.E., Webster, M.T., 2019. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. BMC Genomics 20, 275. https://doi.org/10.1186/s12864-019-5642-0

Wang, Q., Gu, H., Dorn, S., 2004. Genetic relationship between olfactory response and fitness in *Cotesia glomerata* (L.). Heredity (Edinb). 92, 579–584. https://doi.org/10.1038/sj.hdy.6800464

Wang, Q., Gu, H., Dorn, S., 2003. Selection on olfactory response to semiochemicals from a plant-host complex in a parasitic wasp. Heredity (Edinb). 91, 430–435. https://doi.org/10.1038/sj.hdy.6800340

Wang, Z., Liu, Z., Wu, X., Yan, W., Zeng, Z., 2012. Polymorphism analysis of csd gene in six *Apis mellifera* subspecies. Mol. Biol. Rep. 39, 3067–3071. https://doi.org/10.1007/s11033-011-1069-7

Waterhouse, R.M., Seppey, M., Simao, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. 35, 543–548. https://doi.org/10.1093/molbev/msx319

Webber, B.L., Raghu, S., Edwards, O.R., 2015. Opinion: Is CRISPR-based gene drive a biocontrol silver bullet or global conservation threat? Proc. Natl. Acad. Sci. U. S. A. 112, 10565–10567. https://doi.org/10.1073/pnas.1514258112

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., Jaffe, D.B., 2017. Direct determination of diploid genome sequences. Genome Res. 27, 757–767. https://doi.org/10.1101/gr.214874.116.Freely

Wekker, G., 2016. White Innocence: Paradoxes of colonialism and race. Duke University Press, Durham and London.

Wellcome Trust, 2003. Sharing Data from Large-scale Biological Research Projects : A System of Tripartite Responsibility. Rep. a Meet. Organ. by Wellcome Trust held 14–15 January 2003 Fort Lauderdale, USA. https://doi.org/10.1093/hmg/ddl213

Werren, J.H., Baldo, L., Clark, M.E., 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6, 741–751. https://doi.org/10.1038/nrmicro1969

Werren, J.H., Loehlin, D.W., 2009. Curing *Wolbachia* infections in *Nasonia* (parasitoid wasp). Cold Spring Harb. Protoc. 2009, pdb.prot5312-pdb.prot5312. https://doi.org/10.1101/pdb.prot5312

Wetterstrand, K.A., 2017. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [WWW Document]. Natl. Hum. Genome Res. Inst. URL http://www.genome.gov/sequencingcosts/ (accessed 11.14.19).

Wheeler, D., Redding, A.J., Werren, J.H., 2013. Characterization of an Ancient Lepidopteran Lateral Gene Transfer. PLoS One 8. https://doi.org/10.1371/journal.pone.0059262

Whiting, P.W., 1940. Multiple alleles in sex determination of *Habrobracon*. J. Morphol. 66, 323–355. https://doi.org/10.1002/jmor.1050660208

Whiting, P.W., Whiting, A.E., 1925. Diploid males from fertilized eggs in hymenoptera. Science (80-.). 62, 437. https://doi.org/10.1126/science.62.1611.437

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.

Wong, T.T.Y., Nakahara, L.M., 1978. Sexual Development and Mating Response of Laboratory-Reared and Native Mediterranean Fruit Flies1,2. Ann. Entomol. Soc. Am. 71, 592–596. https://doi.org/10.1093/aesa/71.4.592

Woodworth, L.M., Montgomery, M.E., Briscoe, D.A., Frankham, R., 2002. Rapid genetic deterioration in captive populations: Causes and conservation implications. Conserv. Genet. 3, 277–288. https://doi.org/10.1023/A:1019954801089

Wright, M.G., Bennett, G.M., 2018. Evolution of biological control agents following introduction to new environments. BioControl 63, 105–116. https://doi.org/10.1007/s10526-017-9830-z

Wright, S., 1984. Evolution and the genetics of populations. University of Chicago Press.

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y.Q., Coleman-Derr, D., Xia, Q., Wang, Y., 2019. OrthoVenn2: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 47, W52–W58. https://doi.org/10.1093/nar/gkz333

Xu, P., Lu, B., Liu, J., Chao, J., Donkersley, P., Holdbrook, R., Lu, Y., 2019. Duplication and expression of horizontally transferred polygalacturonase genes is associated with host range expansion of mirid bugs. BMC Evol. Biol. 19, 12. https://doi.org/10.1186/s12862-019-1351-1

Xun, H., Li, H., Li, S., Wei, S., Zhang, L., Song, F., Jiang, P., Yang, H., Han, F., Cai, W., 2016. Population genetic structure and post-LGM expansion of the plant bug *Nesidiocoris tenuis* (Hemiptera: Miridae) in China. Sci. Rep. 6, 26755. https://doi.org/10.1038/srep26755

Ye, C., Hill, C.M., Wu, S., Ruan, J., Ma, Z., 2016. DBG2OLC: Efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. 6, 1–9. https://doi.org/10.1038/srep31900

Ye, C., Ma, Z.S., Cannon, C.H., Pop, M., Yu, D.W., 2012. Exploiting sparseness in *de novo* genome assembly. BMC Bioinformatics 13 Suppl 6, S1. https://doi.org/10.1186/1471-2105-13-S6-S1

Zaviezo, T., Retamal, R., Urvois, T., Fauvergue, X., Blin, A., Malausa, T., 2018. Effects of inbreeding on a gregarious parasitoid wasp with complementary sex determination. Evol. Appl. 11, 243–253. https://doi.org/10.1111/eva.12537

Zhang, Y., Qiu, S., 2015. Examining phylogenetic relationships of *Erwinia* and *Pantoea* species using whole genome sequence data. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 108, 1037–1046. https://doi.org/10.1007/s10482-015-0556-6

Zhou, W., Rousset, F., O'Neill, S., 1998. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. R. Soc. B Biol. Sci. 265, 509–515. https://doi.org/10.1098/rspb.1998.0324

Zimmermann, O., 2004. Der Einsatz von *Trichogramma*-Schlupfwespen in Deutschland: Zum aktuellen Stand der Forschung und Nutzung von Eiparasitoiden gegen Schadlepidopteren im biologischen Pflanzen- und Vorratsschutz. Gesunde Pflanz. 56, 157–166. https://doi.org/10.1007/s10343-004-0031-1

Zrzavá, M., Hladová, I., Dalíková, M., Šíchová, J., Õunap, E., Kubíčková, S., Marec, F., 2018. Sex chromosomes of the iconic moth *Abraxas grossulariata* (Lepidoptera, Geometridae) and its congener *A. sylvata*. Genes (Basel). 9, 1–16. https://doi.org/10.3390/genes9060279

SUMMARY

Biological control is the use of an organism, the biological control agent (BCA), to control the population of another organism, the pest. Biological control is used within a variety of contexts, such as in open-field agriculture such as maize, or in greenhouses for crops such as tomatoes or cucumbers. BCAs are judged not only by their efficiency and effectiveness in controlling pests, but also by their ease or ability to be reared in commercial settings, stored for release, or the likelihood of what are called non-target effects (undesired effects such as preying upon native insects that are not pests).

While BCAs can be used isolated in several different agricultural practices, they have especially high potential as a key part of Integrated Pest Management (IPM). IPM is an ecosystem approach to agriculture, where the entire system in question is assessed and the main goal is more healthy crops and less pesticides. At face value, this sounds like a win-win situation as reducing pesticide use is necessary to reduce the negative impact of agriculture on the environment, and consumers (in general) are increasingly moving towards products with less pesticides or to organic products. Indeed, BCAs are often part of organically produced crops and products. However, the total uptake of IPM and other use cases of BCAs is low, partially due to the perceived unreliability of BCAs as compared to conventional pesticides.

One way to improve BCAs is to use a genetics approach with next-generation sequencing and genomics. The study of genetics is essentially about evolution and inheritance, and its application on BCAs is fairly straightforward: are the traits that we are interested in, such as parasitism rate for parasitoid wasps or starvation resistance for predatory bugs, 1) heritable, and 2) able to be improved without deleterious (side-)effects? Is there enough genetic variation within a population to select for improvement of these traits? While these scenarios can be used for improving the BCAs themselves, the way that BCAs are monitored, stored, and assessed for non-target effects can also be improved using genetics and/or genomics.

The EU-funded research training network Breeding Invertebrates for Next Generation BioControl (BINGO-ITN) contained several projects that together aimed to improve the production and performance of BCAs using genetic variation, including the use of genomic techniques. The project involved 24 senior researchers, 13 PhD projects, and 12 partners from academia, and

industrial and non-profit organizations throughout the EU. This thesis contains the work of one of those projects. Furthermore, this thesis is framed as an anthology of works on the possibilities and potential of genomics and genomic resources of BCAs.

The intention of this thesis was to generate genomic results and resources for five biological control agents, complete with context and suggestions for future directions. My approach was to focus on the direct application of each species, giving clear indications of the possible uses of these newly generated genomic resources, while trying to be as open as possible with my science. I also used the concept of the life cycle of a genome project as a guide to completing or at least advancing genome projects.

In **Chapter 2**, we perform a systematic review where we delved into literature on genetic variation of BCAs. We focussed our search term for papers that mentioned genetic variation and/or heritability, as well as biological control and all possible permutations therein. While our search initially returned nearly 3,000 hits, this was quickly narrowed down to include only species known to be used in biological control, as well as to the purpose of our study question. In the end, 69 papers fitted the bill, though the majority of these papers did not mention quantifiable measures of genetic variation, such as heritability (h^2 or H^2) or evolvability (CV_A). From this review, we were able to shed light on the traits currently being studied in relation to those potentially important for improving biological control. Additionally, we made a case for including quantifiable measures of genetic variation in associate research and more transparent reporting practices in general.

The remaining research chapters, **Chapters 3-7**, are arranged in order of end product, as a genome was not generated for all five species. In **Chapter 3**, we were working with *Amblyseius swirskii*, a predatory mite from the Eastern Mediterranean which is used around the world in a large variety of greenhouse crops. However, its initial collection was a fairly small source population, and concerns about resiliency and field performance led us to investigate the genetic variation of the commercial population. To do so, we set up an inbred line from the commercial population using a mother-son mating scheme over ten generations, while collecting eight wild populations in Israel for comparison purposes. Using whole-genome nanopore sequencing, we obtained just 512 Mbp of clean, corrected reads from the inbred line. While

this was not enough for genome construction, as was the original goal, it was more than enough for microsatellite mining. Using six microsatellite loci, DNA was pooled for microsatellite analysis, a cost-effective alternative to individual genotyping that proved effective for this study. Our findings indicate that the commercial population had reduced genetic variation and far more differentiated than its wild counterparts. Given these results, we recommended increasing the scope to more commercial populations, more consistent monitoring of commercial lines for genetic variation, and to consider introgressing new material into the commercial lines. We recommend that the latter is done first in a test population to see whether genetic variation can be increased without hampering the biological control performance.

In **Chapter 4** we present the linked-read genome sequencing and assembly approach for *Bracon brevicornis*, an ectoparasitoid wasp that is currently being investigated around the world for biological control. For sequencing and assembly strategy, we went with a linked-reads approach, where a small amount of input material binds with barcodes to aid in an assembly that has low coverage but high accuracy. In addition to *B. brevicornis* material, we used a well-studied and sequenced organism (*Solanum lycopersicum*, tomato) as "carrier DNA" in the library preparation step. The resulting genome was 123 Mbp in size. We chose to use a linked-read assembly as it requires far less input material and can provide "phased" genomes, where areas of heterozygosity can be visualised instead of being removed as in other methods. The difficulty in getting enough genetic material from these wasps is partially due to the complementary sex determination (CSD) system within *B. brevicornis* that complicates inbreeding as it results in sterile diploid males. In the end, our solution for dealing with a low amount of input DNA also allowed us to identify a genomic region that is likely linked to the CSD mechanism in this species. Along with this in-depth investigation into CSD, we performed a protein comparison between *B. brevicornis* and two other braconid wasps to highlight the possibilities for comparative genomics with this genome, as well as an assessment of the accuracy of our *ab initio*-only assembly.

The genome of the widely used parasitoid wasp *Trichogramma brassicae* (**Chapter 5**) was achieved through a hybrid approach, where short and long-

read sequencing technology was used. The homozygosity of our inbred line, S301, was likely caused by a *Wolbachia* infection. This bacterial endosymbiont is found in a variety of insect hosts, and in some cases it can lead to female-only populations, where unfertilized eggs become female wasps instead of the usual male wasps. Three different assemblers were used, and five potential assemblies were narrowed down to one that went on to receive *ab initio*-, homology-, and evidence-based annotation. The final assembly size is 235 Mbp distributed over 1,572 contigs and contains 16,905 genes.

The whole-genome sequencing from Chapter 5 went on to generate microsatellites in **Chapter 6** on sister species *Trichogramma evanescens*. The question at hand here was related to ongoing efforts of monitoring both BCAs and their wild counterparts. The dispersal modes of *Trichogramma* spp. are generally considered to be either through direct or wind-based dispersal, and as such its dispersal range is quite small. However, recent observations of phoresy (a form of biological "hitch-hiking" behaviour) would increase the dispersal range to that of any butterflies that wasps are hitching a ride on. With eight German wild-caught lines and two Dutch lines, we used a mixture of population genetics and population genomics to explore this question. Microsatellites, an unannotated genome of *T. evanescens*, and pooled sequencing were used, and our analyses indicate that the populations show slight isolation-by-distance and strong differentiation between lines and within collection sites, but no clear latitudinal cline. While more investigation is necessary, our results combined with several in-field observations of phoresy, suggest that the dispersal range of *Trichogramma* may be larger than previously thought.

The final genome within this thesis (**Chapter 7**) belongs to *Nesidiocoris tenuis*, a predatory bug used in greenhouses throughout the Mediterranean Basin, and is the second linked-read genome within this thesis. A single *N. tenuis* female was the basis for this genome, along with potential bacterial contaminants. These tag-alongs were removed through two decontamination pipelines, one of which also identified putative regions of lateral gene transfer (LGT). The total genome assembly size is 355 Mbp. Post decontamination, we performed an *ab initio*-, homology-, and evidence-based assembly that yielded 24,688 genes, prompting a comparative analysis with other Hemipteran genomes. Karyotyping and sex chromosome analysis indicates that *N. tenuis* has

a 32-chromosome constitution with an XX/XY sex determination system. Additional cytogenetic analysis looked at sex chromosome composition as well as visualising two probes, the 18s rDNA cluster and a unique satellite from the sequence reads. Lastly, the chromosomes were checked for ancestral insect telomeric repeats, which are thought to be lost in Hemipterans such as *N. tenuis*, and indeed, we were not able to visualise these or similar repetitive probes. The variety of analyses performed for this genome was to inspire future genomic and genetic work with *N. tenuis*, as well as to showcase what is possible with an annotated genome.

In the synthesis in **Chapter 8**, I bring back the concept of the genome project life cycle, discussing how each genome project in this thesis moved through the cycle, or did not, and I also make suggestions for additions that I would make to the cycle. I briefly touch on the openness of resources generated within and from this thesis, before discussing the Nagoya Protocol. I cover how the international agreement is currently perceived by biological control practitioners, and how biological control and Dutch universities alike would benefit from incorporating decolonisation into their framing and approach to science. At the end, I briefly reflect upon the BINGO-ITN and close out the anthology.

SAMENVATTING

Biologische bestrijding is het gebruik van een organisme, de biologische bestrijder (BB), om de populatie van een ander organisme, de plaag, te bestrijden. Biologische bestrijding wordt gebruikt in verschillende contexten, zoals in de akkerbouw voor gewassen zoals maïs, of in de kasteelt voor gewassen zoals tomaten en komkommers. BB's worden niet alleen beoordeeld op hun efficiëntie en effectiviteit bij het bestrijden van schadelijke organismen, maar ook op hoe makkelijk het is om ze te kweken in commerciële instellingen, op hun opslagmogelijkheden, en op mogelijke ongewenste effecten zoals het aanvallen van inheemse insecten die geen ongedierte zijn.

Hoewel BB's op zichzelf kunnen worden gebruikt in verschillende landbouwpraktijken, hebben ze vooral veel potentie als een belangrijk onderdeel van geïntegreerde gewasbescherming, ook wel *Integrated Pest Management* (IPM) genoemd. IPM is een ecosysteem-benadering van de landbouw, waarbij het hele systeem wordt meegenomen met als hoofddoel gezondere gewassen met minder pesticiden te krijgen. Op het eerste gezicht klinkt dit als een win-win situatie, aangezien het verminderen van het gebruik van pesticiden noodzakelijk is om de negatieve impact van de landbouw op het milieu te verminderen, en consumenten steeds meer op weg zijn naar producten met minder pesticiden of naar biologische producten. BB's maken inderdaad vaak deel uit van de biologische productie van gewassen en andere biologische producten. Het totale gebruik van IPM en BB's is echter laag, gedeeltelijk vanwege de reputatie dat BB's onbetrouwbaar zijn in vergelijking met conventionele pesticiden.

BB's kunnen worden verbeterd via een genetisch aanpak met *next-generation sequencing* en genomica. Genetica gaat in essentie over evolutie en het erven van eigenschappen, en de toepassing ervan op BB's is vrij eenvoudig: zijn de eigenschappen waarin we geïnteresseerd zijn, zoals parasitisme voor sluipwespen of hongerbestendigheid voor roofwantsen, 1) erfelijk en 2) kunnen ze worden verbeterd zonder schadelijke effecten? Is er voldoende genetische variatie binnen een populatie om te kunnen selecteren voor verbetering van deze eigenschappen? Hoewel deze scenario's kunnen worden gebruikt om de BB's te verbeteren, kan de manier waarop BB's worden gemonitord, opgeslagen en beoordeeld op niet-doeleffecten ook

worden verbeterd met behulp van genetica en/of genomica.

Het door de EU gefinancierde netwerk *Breeding Invertebrates for Next Generation BioControl* (BINGO-ITN) omvatte verschillende projecten die samen gericht waren op het verbeteren van de productie en prestaties van BB's met behulp van genetische variatie, waaronder het gebruik van genomische technieken. Bij het project waren 24 senior onderzoekers, 13 PhD-projecten en 12 partners uit de academische wereld en industriële en non-profitorganisaties uit de hele EU betrokken. Dit proefschrift bevat het werk van een van deze projecten. Verder is dit proefschrift gekaderd als een anthologie van werken over de mogelijkheden en het potentieel van genomica en genomische resources van BB's.

Het doel van dit proefschrift was om genomische resultaten en middelen te genereren voor vijf biologische bestrijders, compleet met context en suggesties voor toekomstige onderzoeksrichtingen. Mijn aanpak was om me te concentreren op de directe toepassing van elke soort, door duidelijke aanwijzingen te geven voor het mogelijke gebruik van deze nieuw gegenereerde genomische hulpbronnen, terwijl ik probeerde zo open mogelijk te zijn met mijn wetenschap. Ik heb ook het concept van de levenscyclus van een genoomproject gebruikt als gids voor het voltooien of op zijn minst bevorderen van genoomprojecten.

In **Hoofdstuk 2** voeren we een systematische review uit waarin we ons verdiepen in literatuur over de genetische variatie van BB's. We hebben onze zoektermen geconcentreerd op publicaties waarin genetische variatie en/of erfelijkheid wordt genoemd, in combinatie met biologische controle en alle mogelijke permutaties daarvan. Hoewel onze zoekopdracht aanvankelijk bijna 3.000 hits opleverde, werd dit aantal snel minder toen we ons beperkten tot het doel van onze onderzoeksvraag, en soorten waarvan bekend is dat ze worden gebruikt bij biologische bestrijding. Uiteindelijk vonden we 69 geschikte publicaties, hoewel de meerderheid van deze artikelen geen meetbare resultaten van genetische variatie vermeldde, zoals erfelijkheid (h^2 of H^2) of evolueerbaarheid (CV_A). Door deze review konden we licht werpen op de eigenschappen die momenteel worden bestudeerd in relatie tot eigenschappen die mogelijk belangrijk zijn voor het verbeteren van biologische bestrijding. Daarnaast hebben we gepleit voor het opnemen van kwantificeerbare metingen van genetische variatie in dit

type onderzoek, en voor transparantere rapportagepraktijken.

De resterende onderzoekhoofdstukken, **Hoofdstukken 3-7**, zijn gerangschikt in volgorde van het eindproduct, omdat niet voor alle vijf soorten een genoom is gegenereerd. In **Hoofdstuk 3** werkten we met *Amblyseius swirskii*, een roofmijt uit het oostelijke Middellandse Zeegebied, die over de hele wereld wordt gebruikt in veel verschillende kasgewassen. Deze roofmijt is oorspronkelijk gevonden in een kleine bronpopulatie, en zorgen over veerkracht en veldprestaties brachten ons ertoe de genetische variatie van de commerciële populatie te onderzoeken. Hiervoor hebben we een inteeltlijn opgezet van de commerciële populatie met behulp van een moeder-zoon-paringsschema voor tien generaties, terwijl we acht wilde populaties in Israël verzamelden om ze te kunnen vergelijken. Met het gebruik van *whole-genome nanopore sequencing* hebben we slechts 512 Mpb schone, gecorrigeerde reads van de inteeltlijn gekregen. Hoewel dit niet genoeg was voor genoomconstructie, zoals het oorspronkelijke doel was, was het meer dan genoeg om microsatellieten te kunnen vinden. Met behulp van zes microsatellieten werd DNA samengevoegd voor microsatellietanalyse, een kosteneffectief alternatief voor individuele genotypering dat effectief bleek voor deze studie. Onze resultaten laten zien dat de commerciële populatie minder genetische variatie had en meer gedifferentieerd was dan de wilde tegenhangers. Gebaseerd op deze resultaten hebben we aanbevolen om consistentere monitoring van commerciële lijnen op genetische variatie uit te voeren en om te overwegen om nieuw materiaal in de commerciële lijnen te introduceren. We raden aan om dit laatste eerst in een testpopulatie te doen om te zien of genetische variatie kan worden verhoogd zonder het succes van de biologische bestrijding te belemmeren.

In **Hoofdstuk 4** presenteren we het *linked-read* genoom en de assembly strategie voor *Bracon brevicornis*, een ectoparasitoïde wesp die momenteel wereldwijd wordt onderzocht als biologische bestrijder. Voor het sequencen en de assembly hebben we gekozen voor een methode met *linked-reads*, waarbij een kleine hoeveelheid invoermateriaal bindt met barcodes om te helpen bij het samenvoegen met een lage dekking maar een hoge nauwkeurigheid. Naast *B. brevicornis*-materiaal gebruikten we een goed bestudeerd en gesequenced organisme (*Solanum lycopersicum*, tomaat) als "drager-DNA" in de voorbereidingsstap van de library. Het resulterende

genoom was 123 Mbp groot. We hebben ervoor gekozen om een *linked-reads* assembly te gebruiken, omdat dit veel minder beginmateriaal vereist en "gefaseerde" genomen kan maken, waar gebieden met heterozygositeit kunnen worden gevisualiseerd in plaats van te worden verwijderd zoals bij andere methoden. De moeilijkheid om voldoende genetisch materiaal van deze wespen te krijgen, is gedeeltelijk te wijten aan het systeem van complementaire geslachtsbepaling (CGB) in *B. brevicornis* dat inteelt bemoeilijkt omdat het resulteert in steriele diploïde mannetjes. Uiteindelijk heeft onze oplossing voor het omgaan met een lage hoeveelheid input-DNA ons ook in staat gesteld een genomisch gebied te identificeren dat waarschijnlijk is gekoppeld aan het CGB-mechanisme in deze soort. Samen met dit diepgaande onderzoek naar CGB hebben we een eiwitvergelijking tussen *B. brevicornis* en twee andere braconide wespen uitgevoerd om de mogelijkheden voor vergelijkende genomica met dit genoom te highlighten, en ook als beoordeling van de nauwkeurigheid van onze assembly die alleen *ab initio* geannoteerd is.

Het maken van een genoom van de veel gebruikte sluipwesp *Trichogramma brassicae* (**Hoofdstuk 5**) werd bereikt door een hybride aanpak, waarbij gebruik werd gemaakt van sequencingtechnologie met short- en long-reads. De homozygositeit van onze ingeteelde lijn, S301, werd waarschijnlijk veroorzaakt door een *Wolbachia*-infectie. Deze bacteriële endosymbiont bestaat in verschillende insecten en kan in sommige gevallen leiden tot populaties die volledig vrouwelijk zijn, waar onbevruchte eieren vrouwelijke wespen worden in plaats van de gebruikelijke mannelijke wespen. Drie verschillende assemblers werden gebruikt, en vijf potentiële assemblies werden teruggebracht tot één die vervolgens *ab initio*, homologie, en op bewijs gebaseerde annotatie ontving. De uiteindelijke assembly is 235 Mbp groot, verdeeld over 1.572 contigs en bevat 16.905 genen.

Whole-genome sequencing van Hoofdstuk 5 werd gebruikt om microsatellieten te genereren in **Hoofdstuk 6** voor de zustersoort *Trichogramma evanescens*. De vraag was gerelateerd aan voortdurende inspanningen om zowel BB's als hun wilde tegenhangers te monitoren. De verspreidingswijzen van *Trichogramma* spp., dacht men tot nu toe, zijn ofwel directe dispersie of dispersie door de wind, en daardoor zou het verspreidingsbereik vrij klein zijn. Recente waarnemingen van *phoresy* (een vorm van "liftend" gedrag

in de natuur) zouden het verspreidingsbereik echter zo groot maken als het bereik van alle vlinders waar wespen op mee kunnen liften. Met acht Duitse lijnen uit het wild en twee Nederlandse lijnen gebruikten we een combinatie van populatiegenetica en -genomica om deze vraag te onderzoeken. Microsatellieten, een ongeannoteerd genoom van *T. evanescens* en gepoolde sequencing werden gebruikt, en onze analyses laten zien dat de populaties een lichte isolatie over afstand hebben, en sterke differentiatie tussen lijnen in binnen verzamelplaatsen, maar dat er geen duidelijke trend is over de geografische breedtegraad. Hoewel meer onderzoek nodig is, suggereren onze resultaten in combinatie met verschillende veldobservaties van phoresy dat het verspreidingsbereik van *Trichogramma* mogelijk groter is dan eerder gedacht.

Het laatste genoom in dit proefschrift (**Hoofdstuk 7**) is van *Nesidiocoris tenuis*, een roofwants die wordt gebruikt in kassen in het Middellandse Zeegebied, en is het tweede linked-read genoom in dit proefschrift. Eén vrouwelijke *N. tenuis* vormde de basis voor het genoom, samen met mogelijke bacteriële verontreinigingen. Deze aanhangsels werden verwijderd via twee decontaminatie pipelines, waarvan er één ook vermoedelijke gebieden van laterale genoverdracht identificeerde. De totale genoomassembly is 355 Mbp groot. Na decontaminatie hebben we een *ab initio*, homologie, en evidence-based assembly uitgevoerd die 24.688 genen opleverde, wat naar een vergelijkende analyse met andere Hemiptera genomen leidde. Karyotyping en geslachtschromosoomanalyse geven aan dat *N. tenuis* een 32 chromosomen heeft met een XX/XY geslachtsbepalingssysteem. Aanvullende cytogenetische analyse omvatte de geslachtschromosoomsamenstelling en visualiseerde twee probes, de 18s rDNA-cluster en een unieke satelliet uit de reads van de sequentie. Ten slotte werden de chromosomen gecontroleerd op telomerische herhalingen van vooroudelijke insecten, waarvan wordt gedacht dat ze verloren zijn in Hemipterans zoals *N. tenuis*, en we waren inderdaad niet in staat om deze of soortgelijke repetitieve probes te visualiseren. De verscheidenheid aan analyses die voor dit genoom werd uitgevoerd, was om toekomstig genomisch en genetisch werk met *N. tenuis* te inspireren, en om te laten zien wat de mogelijkheden zijn van een geannoteerd genoom.

In de synthese in **Hoofdstuk 8** breng ik het concept van de levenscyclus van

het genoomproject terug, waarbij ik bespreek hoe elk genoomproject in dit proefschrift de cyclus heeft doorlopen, of juist niet, en ik doe ook suggesties voor toevoegingen aan de cyclus. Ik zal kort ingaan op de openheid van resources die binnen en door dit proefschrift worden gegenereerd, waarna ik het Nagoya protocol bespreek. Ik bespreek hoe de internationale overeenkomst momenteel wordt waargenomen door beoefenaars van biologische bestrijding, en hoe biologische bestrijding en Nederlandse universiteiten er baat bij zouden hebben om dekolonisatie in hun benadering van de wetenschap op te nemen. Ter afsluiting reflecteer ik kort op het BINGO-ITN en sluit ik de anthologie af.

ZUSAMMENFASSUNG

Wird ein lebender Organismus benutzt, um die Ausbreitung schädlicher Organismen zu kontrollieren, spricht man von Biologischer Schädlingsbekämpfung (biological control agent, BCA). Betrieben wird Biologische Schädlingsbekämpfung in einer Vielzahl von Kontexten, unter Anderem beim Anbau von Feldfrüchten wie Mais oder für Produkte aus dem Gewächshaus wie Tomaten oder Gurken. BCAs werden nicht nur nach ihrer Effektivität und Effizienz bei der Schädlingsbekämpfung bewertet. Auch einfache Handhabung bei Aufzucht und Lagerung in einem kommerziellen Kontext spielen eine Rolle, sowie die Wahrscheinlichkeit für mögliche Folgen (z.B. durch Prädation) für Arten, die nicht im Fokus der Behandlung stehen (*non-target* Effekte).

In vielen verschiedenen landwirtschaftlichen Verfahren werden BCAs allein eingesetzt, obwohl besonderes Potential in der Nutzung von BCAs als Schlüsselkomponente des integrierten Pflanzenschutzes (Integrated Pest Management, IPM) liegt. Dabei handelt es sich um einen ganzheitlichen Ansatz in der Landwirtschaft. Das betreffende Ökosystem wird eingehend geprüft mit dem Ziel gesündere Pflanzen und größere Erträge zu erzielen. Außerdem soll es den Einsatz von Pestiziden vermindern. Rational betrachtet ergibt sich daraus eine Win-Win-Situation. Neben dem (generell) veränderten Konsumverhalten hin zu Produkten mit geringer Pestizidbelastung oder aus biologischer Herstellung, ist die Einschränkung des Einsatzes von Pestiziden notwendig, um die negativen Auswirkungen der Landwirtschaft auf die Umwelt zu reduzieren. Tatsächlich sind BCAs häufig Teil der Herstellung biologischer Produkte oder im biologischen Ackerbau. Insgesamt kommen IPMs und BCAs aber nur in geringem Maß zum Einsatz, was teilweise auf die scheinbare Unzuverlässigkeit von BCAs im Vergleich zu konventionellen Pestiziden zurückzuführen ist.

Eine genetische Herangehensweise durch *Next-Generation-Sequencing* (NGS) und Genomik bietet die Möglichkeit BCAs zu verbessern. Im Mittelpunkt stehen Evolution und Vererbung. Fragestellungen bzgl. der BCAs, die sich daraus ergeben sind simpel z.B.: Sind interessante Eigenschaften, wie die Parasitierraten von parasitoiden Wespen oder die Toleranz von prädatorischen Wanzen gegenüber Nahrungsknappheit 1) vererbbar und 2) besitzen sie das Potential verbessert zu werden, ohne dabei schädliche (Neben-)Wirkungen

zu erzeugen? Ist die genetische Variation einer Population groß genug, um Merkmale von Interesse selektiv zu verbessern? Diese Fragen können auf die Art und Weise der Überwachung und Lagerung der BCAs ausgeweitet werden. Des Weiteren können genetische Techniken die Beurteilung von *non-target* Effekten verbessern.

Breeding Invertebrates for Next Generation BioControl (BINGO-ITN) ist ein durch die EU finanziertes Netzwerk für Wissenschaft und Ausbildung im Bereich der Zucht der nächsten Generation von Wirbellosen zur Schädlingsbekämpfung. BINGO-ITN umfasst mehrere Projekte, die genetische Variation und Techniken aus der Genomik nutzen um BCAs leistungsstärker zu machen und ihre Produktion zu optimieren. EU-weit beteiligten sich 24 leitende Forscher, 13 Doktoranden und 12 Partner aus dem akademischen Umfeld, sowie Partner aus der Industrie und von gemeinnützigen Organisationen. Diese Dissertation entstand als eins der Promotionsprojekte unter BINGO und soll als Sammlung von Ausbau- und Anwendungsmöglichkeiten der Genomik und genomischer Ressourcen in Bezug auf BCAs gelten.

Das Ziel dieser Dissertation war es Genomressourcen und genomische Daten von fünf Organismen zur biologischen Schädlingsbekämpfung zu erzeugen, und mit Hintergrund und potentiellen Anwendungsmöglichkeiten zu präsentieren. Mein Ansatz war auf die direkte Anwendbarkeit der einzelnen Arten ausgerichtet, darauf hinzuweisen, welche Einsatzmöglichkeiten diese neu generierten genomischen Daten eröffnen und gleichzeitig möglichst transparent und zugänglich nach den Prinzipien von *Open Science* zu forschen. Das Konzept des „*Lifecycles of a Genome Project*“ (zyklisches Verbessern eines Genomprojekts) diente mir als Leitfaden, um Genomprojekte abzuschließen oder zumindest voranzubringen.

In **Kapitel 2** geben wir einen systematischen Überblick, der vertieft auf die genetische Variation von BCAs eingeht. Wir schränkten unsere Suche auf Publikationen ein, welche „*genetic variation*“ (genetische Variation) und/ oder „*heritability*“ (Vererblichkeit), sowie „*biological control*“ (Biologische Schädlingsbekämpfung) und Kombinationen dieser Begriffe beinhalteten. Die nahezu 3000 anfänglichen Treffer unserer Suche wurden auf solche eingegrenzt, die zu unserer Forschungsfrage passten und Arten enthielten, die bekanntlich zur biologischen Schädlingsbekämpfung eingesetzt werden. Letztendlich entsprachen 69 Artikel den Anforderungen, wobei die Mehrheit

kein quantifizierbares Maß für genetische Variation, wie Heredität (h^2 or H^2) oder Evolvierbarkeit (CV_A) nutzten. Der daraus entstandene Review-Artikel beleuchtet Merkmale, an denen derzeit geforscht wird, im Vergleich zu solchen, welche möglicherweise nützlich wären für die Verbesserung biologischer Schädlingsbekämpfung. Darüber hinaus sprachen wir uns für die Aufnahme eines quantifizierbaren Maßes für genetische Variation in relevanter Forschung und für eine transparentere Berichterstattung aus.

Die verbleibenden Kapitel 3-7 sind nach ihrer Vollständigkeit geordnet, da nicht alle fünf Genome vollständig erarbeitet werden konnten. In **Kapitel 3** wurde an *Amblyseius swirskii*, einer Raubmilbe aus dem östlichen Mittelmeerraum, gearbeitet, welche weltweit zum Schutz einer Vielzahl von Pflanzen im Gewächshaus eingesetzt wird. Allerdings war die Population, aus der die Zuchtkulturen stammen, relativ klein. Bedenken hinsichtlich der Belastbarkeit und der Leistungsfähigkeit im Feld brachten uns dazu die genetische Variation innerhalb der kommerziell genutzten Population zu untersuchen. Zu diesem Zweck haben wir über 10 Generationen eine Inzuchtlinie aus der kommerziellen Linie aufgebaut, unter Verwendung eines Mutter-Sohn-Verpaarungsschemas. Zum Vergleich sammelten wir währenddessen acht Wildpopulationen in Israel. Durch Nanopore-Sequenzierung erzeugte Gesamt-Genom-Analysen der Inzuchtlinie lieferten nach Korrektur und Aufreinigung nur 512 Mbp *Reads* (sequenzierte DNA-Fragmente). Für die ursprünglich geplante Genomrekonstruktion ist dies zu wenig, aber es ist mehr als ausreichend um darin Mikrosatelliten zu identifizieren. Gepoolte DNA-Proben wurden an 6 Mikrosatelliten *Loci* analysiert, was in unserem Fall eine effektive und kostengünstige Alternative zur individuellen Genotypisierung war. Aus unseren Ergebnissen lässt sich ableiten, dass die kommerzielle Population eine deutlich verminderte genetische Variation aufweist und stärkere Abweichungen von anderen Populationen (Differenzierung) zeigt als die Wildpopulationen. Ausgehend davon, empfahlen wir: die Ausweitung kommerzieller Populationen, die genetische Variation der kommerziellen Linien konsequenter zu überwachen und die Einführung neuen Materials in die kommerziellen Linien zu erwägen. Weiterhin halten wir es für ratsam letzteres zunächst innerhalb einer Population zu testen, um zu sehen, ob die genetische Variation erhöht werden kann, ohne die Leistungsfähigkeit als biologisches Schädlingsbekämpfungsmittel negativ zu beeinflussen.

Appendices

In **Kapitel 4** stellen wir den Ansatz zur *Linked-Read*-Genom-Sequenzierung und Assemblierung (*Assembly, Zusammensetzen der DNA-Fragmente*) bei *Bracon brevicornis* vor. Diese ektoparasitoide Wespe steht momentan weltweit im Forschungsinteresse als möglicher Kandidat für die biologische Schädlingsbekämpfung. Wir wählten einen *linked-read* Ansatz für Sequenzierung und *Assembly*, dabei ermöglichen kleine Mengen des Ausgangsmaterials mit *Barcodes* ein *Assembly* mit hoher Genauigkeit aber geringer *Coverage* (statistische Redundanz einzelner Basen). Während des Erstellens der *Library* (Sammlung von DNA-Fragmenten) nutzten wir neben *B. brevicornis* auch Material eines bekannten und bereits sequenzierten Organismus (*Solanum lycopersicum*, Tomate) als Träger-*DNA*. Das resultierende Genom war 123 Mbp lang. Wir wählten das *Linked-Read-Assembly* da es weniger Anfangsmaterial benötigt und ein „typisches Genom" (*phased genome*) beinhalten kann, welches heterozygotische Abschnitte darstellen kann, anstatt sie, wie in anderen Verfahren, zu entfernen. Die Schwierigkeit, genügend genetisches Material von diesen Wespen zu erhalten, ist teilweise auf das System der komplementären Geschlechtsbestimmung (*complementary sex determination*, CSD) innerhalb von *B. brevicornis* zurückzuführen. Es erschwert Inzucht, da es zu sterilen diploiden Männchen führt. Unsere Lösung, um mit der geringen Menge an DNA umzugehen, ermöglichte es uns schlussendlich sogar eine Region im Genom zu identifizieren, die wahrscheinlich mit dem CSD-Mechanismus dieser Spezies in Verbindung steht. Neben dieser eingehenden Untersuchung der CSD, führten wir auch einen Proteinvergleich zwischen *B. brevicornis* und zwei anderen Brackwespen durch, um die Möglichkeiten der Vergleichenden-Genomik mit diesem Genom hervorzuheben und um die Genauigkeit unseres reinen *ab initio Assemblies* zu überprüfen.

Das Genom der parasitoiden Wespe *Trichogramma brassicae* (**Kapitel 5**), deren Nutzung weit verbreitet ist, wurde durch einen gemischten Ansatz erzielt. Dabei wurden *short-Read* und *long-Read* Sequenzierungstechniken genutzt. Die Homozygotie unserer Inzuchtslinie, S301, wurde höchstwahrscheinlich durch eine Infektion mit *Wolbachia* verursacht. Dieses Bakterium ist ein Endosymbiont, welches in einer Vielzahl von Insekten gefunden werden kann. In manchen Fällen führt *Wolbachia* zu rein weiblichen Populationen, wobei sich unbefruchtete Eier zu weiblichen Wespen entwickeln, anstatt wie üblich zu männlichen Wespen. Drei unterschiedliche Assembler Programme wurden

genutzt und fünf potentielle *Assemblies* wurden zu einem vereint, welches im nächsten Schritt *ab initio*-, Homologie- und Evidenzbasiert annotiert wurde. Schlussendlich ist das *Assembly* 235 Mbp lang, verteilt über 1.572 *Contigs* und beinhaltet 16.905 Gene.

Die Sequenzierung des gesamten Genoms aus Kapitel 5 ermöglichte es Mikrosatelliten für **Kapitel 6** zu erstellen**,** welches von der Schwesterart *Trichogramma evanescens* handelt. Die hier vorliegende Fragestellung ist laufenden Bestrebungen entlehnt sowohl BCAs als auch ihre in Freiheit lebenden Pendants zu überwachen. Es wird allgemein angenommen, dass die Verbreitungsmodi von *Trichogramma* spp. entweder direkt oder windbedingt sind, dementsprechend kurz ist die Reichweite der Verbreitung. Jüngste Beobachtungen der Phorese (eine Form von biologischem „per Anhalter fahren") würden jedoch die Verbreitungsreichweite auf die aller betreffenden Schmetterlingsarten erhöhen. Wir gingen dieser Frage nach mit im Freiland gefangenen Linien aus Deutschland (8) und den Niederlanden (2). Dabei nutzten wir eine Mischung aus Populationsgenetik und Populationsgenomik. Es wurden Mikrosatelliten, ein nicht annotiertes Genom von *T. evanescens* und Sammelproben-Sequenzierung genutzt. Unsere Analysen weisen auf Populationen hin, die leicht entfernungsisoliert (*isolation-by-distance*) sind und starke Differenzierung zwischen den einzelnen Linien und innerhalb der Stelle der Probenentnahme aufweisen. Sie folgen keinem klaren Gradienten entlang der Breitengrade. Nachfolgeuntersuchungen sind notwendig um zu bestätigen was unsere Ergebnisse kombiniert mit einigen Freilandobservationen von Phorese zeigen, nämlich, dass die Ausbreitungsreichweite von *Trichogramma* weitaus größer sein könnte als bisher angenommen.

Das letzte Genom dieser Dissertation (**Kapitel 7**) gehört zu *Nesidiocoris tenuis*, einer räuberisch lebenden Wanze, die im gesamten Mittelmeerraum in Gewächshäusern verwendet wird. Es ist das zweite Genom innerhalb dieser Dissertation welches per *Linked-Reads* erstellt wurde. Dieses Genom basiert auf einem einzelnen weiblichen Exemplar und potentiellen bakteriellen Verunreinigungen. Diese Verunreinigungen wurden durch zwei Dekontaminations-*Pipelines* entfernt, wobei auch vermeintliche Regionen für lateralen Gentransfer (LGT) identifiziert wurden. Das komplette Genom *Assembly* ist 355 Mbp lang. Nach dem Entfernen der Verunreinigungen,

führten wir eine *ab initio*, Homologie- und Evidenzbasierte Assemblierung durch, die 24.688 Gene ergab, was eine vergleichende Analyse mit anderen Hemiptera-Genomen ermöglichte. Die Karyotypisierung und Geschlechtschromosomenanalyse zeigt, dass *N. tenuis* 32 Chromosomen mit einem XX/XY Geschlechtsdeterminationssystem hat. Zusätzliche zytogenetische Analysen untersuchten die Zusammensetzung der Geschlechtschromosomen und veranschaulichten zwei Regionen, das 18S rDNA-Cluster und einen einzigartigen Satelliten. In einem letzten Schritt wurden die Chromosomen nach teleomerischen *Repeats* durchsucht, die typisch sind für Insekten, von denen jedoch angenommen wird, dass sie in der Ordnung der Hemiptera verloren gegangen sind. Tatsächlich waren wir nicht in der Lage diese oder ähnliche repetitive Regionen zu visualisieren. Die Vielzahl der Analysen an diesem Genom sollte zukünftiges genomisches und genetisches Arbeiten mit *N. tenuis* inspirieren und die Möglichkeiten, die ein annotiertes Genom bietet, exemplarisch aufzeigen.

In der Synthese in **Kapitel 8** komme ich noch einmal auf das Konzept des *Genome Project Life-cycles* zu sprechen. Es werden die Genomprojekte dieser Dissertation diskutiert und inwiefern sie den Genomzyklus durchlaufen haben. Außerdem mache ich Vorschläge für Ergänzungen, die ich in den Zyklus einbringen würde. Ich gehe kurz auf die freie Verfügbarkeit der Ressourcen ein, die innerhalb und durch diese Dissertation generiert wurden. Des Weiteren gehe ich darauf ein, wie das internationale Abkommen „Nagoya Protocol" derzeit von Anwendern der biologischen Schädlingsbekämpfung wahrgenommen wird. Die Universitäten der Niederlande und die biologische Schädlingsbekämpfung würden gleichsam davon profitieren, in Herangehensweise und bei der Rahmensetzung ihrer wissenschaftlichen Arbeiten, das Konzept der Dekolonisation zu berücksichtigen. Mit einer Reflexion über das BINGO-ITN Projekt schließe ich die Anthologie.

Resumen

El control biológico es el uso de un organismo, el agente de control biológico (ACB), para controlar la población de otro organismo, la plaga. El control biológico se usa en una variedad de contextos, desde la agricultura en campo abierto, como en el cultivo de maíz, o en invernaderos para cultivos, como el tomate o el pepino. Los ACB se evalúan no solo por su eficiencia y efectividad en el control de plagas, sino también por su facilidad o capacidad de cría en condiciones comerciales, su almacenaje antes de ser liberados, o la probabilidad de que ataquen o desplacen organismos no diana (especies nativas de insectos que no son plaga).

Si bien los ACB se pueden usar de forma aislada en varias prácticas agrícolas, tienen un potencial especialmente alto como parte clave del Manejo Integrado de Plagas (MIP). El MIP es un enfoque ecosistémico para la agricultura, donde se evalúa todo el sistema en cuestión, y el objetivo principal es la obtención de cultivos más saludables y un menor uso de pesticidas. A primera vista, esto parece una situación beneficiosa, tanto para el medio ambiente, pues es necesario reducir el uso de pesticidas para reducir el impacto negativo de la agricultura, como para los consumidores, que generalmente están optando cada vez más por productos orgánicos o con un menor uso de pesticidas. De hecho, los ACB a menudo se emplean en cultivos y productos producidos orgánicamente. Sin embargo, la aplicación total de métodos de MIP y otros usos de ACB es baja, en parte debido a la percepción de los ACB como poco fiables, en comparación con los pesticidas convencionales.

Una forma de mejorar los ACB es utilizando un enfoque genético con secuenciación de nueva generación y genómica. El estudio de la genética tiene que ver esencialmente con la evolución y la herencia. Así, su aplicación en los ACB es bastante sencilla: ¿son los rasgos que nos interesan, como la tasa de parasitismo de las avispas parasitoides o la resistencia a la inanición de los insectos depredadores 1) heredables y 2) mejorables sin efectos nocivos (secundarios)? ¿Hay suficiente variación genética dentro de una población para seleccionar la mejora de estos rasgos? Si bien estos escenarios se pueden usar para mejorar los ACB en sí mismos, la forma en que se monitorean, almacenan y evalúan los efectos de los ACB en la fauna nativa también se puede mejorar mediante la genética y/o la genómica.

Appendices

La red de formación científica BINGO-ITN (del inglés Breeding Invertebrates for Next Generation BioControl Training Network), financiada con fondos europeos, desarrolló varios proyectos que, en conjunto, tenían como objetivo mejorar la producción y el rendimiento de los ACB mediante la variación genética, incluido el uso de técnicas genómicas. Este proyecto implicó a 24 investigadores *senior*, 13 proyectos doctorales y 12 socios académicos, así como organizaciones industriales y entidades sin ánimo de lucro de toda la UE. Esta tesis expone el trabajo realizado en uno de esos proyectos. Además, la tesis está enmarcada como una antología de trabajos sobre las posibilidades y el potencial de la genómica y los recursos genómicos de los ACB.

El objetivo de esta tesis era generar resultados y recursos genómicos para cinco agentes de control biológico, incluyendo su contexto y sugerencias para futuras direcciones. Mi enfoque se centró en la aplicación directa de cada especie, dando indicaciones claras de los posibles usos de estos recursos genómicos recién generados, mientras trataba de ser lo más abierta, pública y transparente posible con mi ciencia (basándome en el concepto de *open science*). También utilicé el concepto "ciclo de vida" de un proyecto genómico, como guía para completar o avanzar proyectos genómicos.

En el **Capítulo 2**, realizamos una revisión bibliográfica sistemática, en la que ahondamos en la literatura científica publicada sobre la variación genética de los ACB. Fijamos nuestro término de búsqueda en artículos que mencionasen variación genética y/o heredabilidad, así como control biológico y todas las permutaciones posibles sobre el mismo. Inicialmente nuestra búsqueda produjo cerca de 3.000 resultados, que logramos reducir para incluir sólo las especies de las que se conoce su uso como control biológico, así como las que sirvieran al propósito de nuestra pregunta de estudio. Finalmente, 69 artículos cumplieron con los requisitos, aunque la mayoría de dichos artículos no mencionaban medidas cuantificables de variación genética, como la heredabilidad (h^2 o H^2) o la evolucionabilidad (coeficiente de variación genética aditiva, CV_A). Con este artículo de revisión, conseguimos arrojar luz sobre aquellos rasgos que actualmente se están estudiando en relación con los potencialmente importantes para mejorar el control biológico. Además, argumentamos en defensa de la

inclusión de medidas de variación genética cuantificables en investigación, así como prácticas en general más transparentes a la hora de presentar informes.

Los capítulos de investigación restantes, **Capítulos 3-7**, están organizados según el producto final, ya que no se generó un genoma para las cinco especies. En el **Capítulo 3**, trabajamos con *Amblyseius swirskii*, un ácaro depredador del Mediterráneo Oriental usado mundialmente en una gran variedad de cultivos de invernadero. Sin embargo, su recolección inicial procedía de una fuente poblacional bastante pequeña, y diversas preocupaciones sobre su resiliencia y rendimiento en el campo nos llevaron a investigar la variación genética de la población comercial. Para hacerlo, establecimos una línea endogámica de la población comercial utilizando un procedimiento de apareamiento madre-hijo durante diez generaciones. De manera simultánea, recolectamos muestras de ocho poblaciones silvestres en Israel con fines comparativos. Utilizando la secuenciación por nanoporos para el genoma completo, obtuvimos lecturas limpias y corregidas de la línea endogámica de solo 512 Mpb. Si bien esto no fue suficiente para la construcción del genoma, como era originalmente el objetivo, fue más que suficiente para el minado de microsatélites. Utilizando seis loci de microsatélites, se creó un *pool* de ADN para el análisis de microsatélites, una alternativa rentable a la genotipificación individual que resultó ser eficaz para este estudio. Nuestros hallazgos indican que la población comercial mostraba menor variación genética y mucho más diferenciada que sus contrapartes silvestres. Teniendo en cuenta estos resultados, recomendamos aumentar el alcance a más poblaciones comerciales, un monitoreo más consistente de las líneas comerciales para detectar variaciones genéticas y considerar la introgresión genética de material nuevo en dichas líneas. Recomendamos que esto último se realice primero en una población experimental para comprobar que la variación genética pueda aumentarse sin obstaculizar su rendimiento como control biológico.

En el **Capítulo 4** presentamos la técnica de lectura ligada (en inglés *linked-read*), usada para la secuenciación y ensamblado del genoma *Bracon brevicornis*, una avispa ectoparasitoide que actualmente está siendo investigada en todo el mundo para su uso como control biológico. Para la secuenciación y ensamblado, optamos por esta nueva tecnología,

mediante la cual una pequeña cantidad de material genético inicial se une a códigos de barra genéticos, para facilitar un ensamblado de baja cobertura pero de alta precisión. Además del material genético de *B. brevicornis*, utilizamos un organismo bien estudiado y secuenciado (*Solanum lycopersicum*, tomatera) como "ADN portador" en la fase de preparación de la librería. El genoma resultante tenía un tamaño de 123 Mpb. Elegimos usar un ensamblado de lectura ligada, ya que requiere mucho menos material inicial y puede proporcionar genomas de los que se conoce su haplotipo ("*phased" genomes*) y donde se pueden visualizar áreas de heterocigosidad en lugar de eliminarlas como en otros métodos. La dificultad para obtener suficiente material genético de estas avispas se debe en parte al sistema de determinación sexual complementaria (CSD, por sus siglas en inglés) de *B. brevicornis*, que dificulta la consanguinidad al resultar en machos diploides estériles. Finalmente, nuestra solución para lidiar con una baja cantidad de ADN inicial también nos permitió identificar una región genómica que probablemente está vinculada al mecanismo CSD en esta especie. Además de esta investigación en profundidad sobre el CSD, realizamos una comparación de las proteínas de *B. brevicornis* y otras dos avispas bracónidas para resaltar las posibilidades de la genómica comparativa con el genoma obtenido, así como una evaluación de la precisión de nuestro ensamblado *ab initio*.

El genoma de la ampliamente utilizada avispa parasitoide *Trichogramma brassicae* (**Capítulo 5**) se logró a través de un enfoque híbrido, utilizando tecnología de secuenciación de corta y larga lectura. La homocigosidad de nuestra línea endogámica, S301, probablemente fue causada por una infección de *Wolbachia*. Este endosimbionte bacteriano se encuentra en una variedad de insectos hospedadores, y en algunos casos puede conducir a poblaciones compuestas solo por hembras, donde los huevos no fertilizados dan lugar a avispas hembras en lugar de a avispas machos, como es habitual. Se utilizaron tres ensambladores diferentes, y de cinco ensamblados potenciales se redujo a uno, que pasó a ser anotado por los métodos *ab initio*, métodos basados en homología, y anotación basada en evidencias. El tamaño final del ensamblado es de 235 Mpb distribuido en 1,572 *contigs* (conjuntos de fragmentos solapantes) y contiene 16,905 genes.

La secuenciación del genoma completo del Capítulo 5 generó en el

Capítulo 6 microsatélites de la especie hermana *Trichogramma evanescens*. La cuestión aquí tratada estaba relacionada con los continuos esfuerzos de monitoreo, tanto de ACB como de sus contrapartes silvestres. Generalmente se considera que el modo de dispersión de *Trichogramma spp*. es por viento o por dispersión directa, y por lo tanto su rango de dispersión es bastante bajo. Sin embargo, observaciones recientes de foresia (un tipo de comportamiento biológico en el que animales pequeños son polizones de otros más grandes, usándolos así como medio de transporte) aumentarían el rango de dispersión al de las mariposas en las que se montan estas avispas. Usamos ocho líneas alemanas silvestres y dos líneas holandesas para explorar esta cuestión, empleando una mezcla de genética y genómica de poblaciones. Se utilizaron microsatélites, un genoma no anotado de *T. evanescens* y la secuenciación del *pool* de ADN. Nuestros análisis indican que las poblaciones muestran un ligero aislamiento por distancia, y una fuerte diferenciación entre líneas y dentro de los sitios de recolección, pero no hay una clara clina latitudinal. Si bien es necesario seguir investigando, nuestros resultados, combinados con varias observaciones de foresia en el campo, sugieren que el rango de dispersión de *Trichogramma* puede ser mayor de lo que se pensaba anteriormente.

El genoma final dentro de esta tesis (**Capítulo 7**) pertenece a *Nesidiocoris tenuis*, un hemíptero depredador de la familia de los míridos (comúnmente llamado chinche), utilizado en invernaderos de toda la cuenca del Mediterráneo, y es el segundo genoma de lectura ligada de esta tesis. Una sola hembra de *N. tenuis* sirvió de base para este genoma, junto con posibles contaminantes bacterianos. Estos componentes no deseados se eliminaron a través de dos *pipelines* de descontaminación, uno de los cuales también identificó regiones putativas de transferencia genética lateral (TGL). El tamaño total del ensamblado del genoma es de 355 Mpb. Después de la descontaminación, realizamos un ensamblado de secuencias basado en evidencias, homología y *ab initio* que arrojó 24,688 genes, lo que impulsó un análisis comparativo con otros genomas de hemípteros. El cariotipado y análisis de cromosomas sexuales indican que *N. tenuis* tiene una constitución genética de 32 cromosomas con un sistema de determinación del sexo XX/XY. Adicionalmente, el análisis citogenético examinó la composición de los cromosomas sexuales, así como la visualización de dos sondas, el *cluster* de ADNr 18S y un único satélite de las lecturas de secuencias. Por último,

los cromosomas fueron revisados para detectar repeticiones teloméricas ancestrales de insectos, las cuales se creen perdidas en hemípteros como *N. tenuis*. Efectivamente, no pudimos visualizar tales repeticiones teloméricas ni secuencias repetitivas similares. El motivo de la variedad de análisis realizados usando este genoma fue inspirar futuros trabajos genómicos y genéticos con la especie *N. tenuis*, así como demostrar las posibilidades que ofrece un genoma anotado.

En la síntesis del **Capítulo 8**, recupero el concepto de "ciclo de vida" de un proyecto genómico, argumentando cómo cada proyecto genómico incluido en esta tesis fue avanzando o no a través del ciclo, e incluyo sugerencias de adiciones que aplicaría al ciclo. Menciono brevemente la accesibilidad de los recursos generados en y a partir de esta tesis, antes de discutir el Protocolo de Nagoya. Hablo sobre cómo los profesionales de control biológico perciben actualmente dicho acuerdo internacional, y cómo tanto el control biológico como las universidades holandesas se beneficiarían al incorporar la descolonización en su estructura institucional y en su manera de enfocar la ciencia. Por último, reflexiono brevemente sobre la red de formación BINGO-ITN y cierro la antología.

ACKNOWLEDGEMENTS

When I first began this PhD project in September of 2015, the end was nowhere in sight, and that was exactly the way I liked it. Having been a profound generalist throughout my academic career, I was looking forward to a four-year (ish) long interdisciplinary project that was so different from what I previously studied, I didn't really bother to think about what the ending would be like. That was something for Future Kim to worry about! Well guess what, Past Kim: It's time to wrap up this thesis and deliver it unto the world. So let get this ending started. This section is best read while listening to "Now or Never Now," by Metric (and the whole *Art of Doubt* album).

The work within and around this thesis would not have been possible without the support, creativity, and guidance of my supervisors. **Bart**, you've worn many hats throughout the years, some of which include being my daily supervisor as well as the BINGO Coordinator, but also include being a friend and colleague. You made sure that I celebrated successes and that I didn't get stuck in a rut when things didn't work out. **Eveline**, thank you for your generosity and always asking me the tough questions. I've learned so much from you over the years, beyond the topic of this thesis, and I hope that we can work together again soon. And last but not least, **Bas**. Over the years, you ensured that while I focused on the details of the various projects that I also kept the bigger picture in mind, and your direction these last few months has been invaluable. To my supervisors, I cannot thank the three of you enough for your mentorship these last four years, as well as the time that you were able to commit to the crunch of the final months of writing. Our regular meetings over the years were (nearly) always something I looked forward to in my week, even when the results were shit or I was having a bad day, because of the lighthearted approach and the jokes that we were able to share.

To my BINGO Network compatriots throughout the project, I want to thank you all for the great times that we had. Having spoken to folks in other ITNs, when it comes time for the yearly get-togethers, did you know that they don't all rent a hostel or a house or a share of flats and hang out together? Bizarre! Through all of our summer schools and general meetings, we had a pretty happy little group and I wouldn't have it any other way. To all my fellow ESRs – **Shuwen**, **Kelley**, **Sophie LH**, **Alé**, **Simone A**, **Sophie C**, **Manolis**, **Erica**, **Giota**, **Judith**, **Angeliki**, **Milena**, **Sander**, and **Wouter** – we went through this experience together in a variety of ways, and I'm grateful to all of you for your kindness and tenacity these past years. Every time I visited one of you at your city or institute, I knew that there would be a friendly face when I arrived. A huge thanks to **Margreet**, for keeping the BINGO on course and juggling multiple duties. Thanks to **František**, **Tom G**, **Eric**, **Tim**, **Dirk**, **Bernd**, **Nina**, **Andra**, and numerous other collaborators, we put a lot into this project.

A special thank you to **František** and his team at the Institute of Entomology at the Czech Academy of Sciences in České Budějovice. **Irena**, **Martina**, **Anna**, **Lafi**, **Magda**, **Diogo**, and especially **Sander**, thanks for having me over (twice!) and sharing your experience and knowledge with me, as well as bbqs and Czech beer. And let us not overlook the German beers I shared with my Uni Bremen colleagues in my old "home town," thanks to **Sophie C**, **Alé**, **Stefan**, **Carmen**, **Stefanos**, **Valentina**, **Filip**, and **Toby** for the times you invited me to give talks and gave me a place to stay during my visits.

Thanks to the artists who contributed to this thesis: **Erica**, **Jitte**, **Kelley**, **Sarah Z**, **Stefan**, **Tessina**, and in loving memory of **Kirsten Oude Lenferink** who drew her tips for identifying *A. swirskii* several years ago. Thanks to **Carmen**, **Elisabeth**, **Helena**, **Milena**, and **Simone W** for their translation work for the summary. Thank you to my co-authors for your collaboration in addition to your time, patience, and insight.

To my fantastic students whose work can be found throughout this thesis, thanks for hard work and due diligence. **Lorraine**, **Susanne**, and **Joris**, make sure to read the acknowledgements in Chapter 6, and thanks for all of your hard work with the *Trichogramma*. **Marijn**, thanks for all your work with the initial *Nesidiocoris tenuis* horizontal gene transfer results.

To my colleagues in the Laboratory of Genetics here at Wageningen University, thank you all for the support and enthusiasm over these last four years. First and foremost, thank you **Wytske**, for your diligence and determination over the years. To **Margo**, **Anneloes**, **Sabine**, **Ramon**, **Lennart** – we made it through the open office jungle in those early, wall-less days, and I couldn't have asked for better deskmates. Thanks to **Gabriella** for helping with the wasps – you're an amazing person, an anchor of our little insect genetics group. Shout out to **Krithi**, **Suzette**, **Manja**, and **Jianhua** – it's not cheating the escape room if they write things in Chinese that end up being the answers to the puzzles. To **Eveline S**, good luck with your future endeavours! Thanks to **Corrie**, **Eric B**, **Frank**, **Jordy**, and **José** for the good times as well as helping me out in the lab throughout the years. Thanks to **Joost vd H** for all of your bioinformatics help throughout the years, this wouldn't have worked without you. Thanks to **Ben** and **Mathijs**, for some reason you both agreed to have the only sudo rights for our server, which also means you agreed to install every piece of software I needed throughout the last year of my PhD – well-documented or not. To **Alex GrGr**, **Mark Z**, and **Marjon**, I understand why you went on to start up new and exciting things, but I don't have to like it – the three of you are sorely missed!

Final thanks to **Alex TL**, **Anna**, **Andy**, **Anouk**, **Arjan**, **Bernard**, **Bwalya**, **Claudio**, **Cris**, **Eric W**, **Helena**, **Jelle**, **Jitpanu**, **Job**, **Joost K**, **Justin**, **Klaas**, **Lidia**, **Mariska**, **Mark A**, **Murambiwa**, **Philip**, **Phuong**, **Raphael**, **René**, **Robert**, **Roel**, **Sarah D**, **Sijmen**, **Sydney**, **Tom**, **Valeria**, and **Vanessa** for the time we shared together and the probably-a-bit-too-long coffee and lunch breaks, they were little

vacations away from the lab or the computer screen. And thanks to my colleagues in Biosystematics and Entomology: **Dêêdi**, **Eric S**, **Freek**, **Niccolo**, **Nina**, **Patrick**, **Sabrina**, **Sarah**, and **Thijmen** from Biosystematics, and **Filippo**, **Jeroen**, **Thibault**, and **Yidong** from Entomology.

My two fantastic paranymphs, Margo and Sophie, are fearless. I don't believe in fate, but heck, **Margo Spelt-Wisselink** and I began our projects on the same day and we hit it off immediately. Thank you for all of our chats throughout the years, and your support back when I was at the peak of my anxiety woes. You gave me the permission to be upset and take the time to pause that I couldn't give myself. The fate that I don't believe in also led me to this BINGO project and meeting amazing folks, including the incomparable **Sophie Chattington**, who I cannot thank enough for travelling from Germany and being a part of my defense. I can always count on you for having black tea in this land of herbal infusions, as well as for having my back when things get tricky.

A special shout-out to everyone at **Loading Ready Run** – thanks for all you do and for the laughs over the years. Thanks to **Kelsey Byers** and countless others in the #disabledinSTEM/#disabledandSTEM/@chronic_ac community for stoking my activism in the area of disability/chronic illness and STEM.

Thanks to all of my friends from my Bremen days: **Tessina**, **Jimena**, **Carmen**, **Catherina**, **Miriam**, **Teresa**, **Stefanos**, **Shivan**, **Asif**, **Suha**, **Hannah**, **Kathrin**, **Friederike**, **Toni**, and **Anne**. Thanks to both **Michael B** for the good memories in and support from Münster, and to the hilarious **Olga,** our Hilversum correspondent and inappropriate joke machine. Lovely **Simone W**, our little pubquiz team-turned-movie club "Bed and Brexit" is a cornerstone of my life, thanks for your care and compassion throughout the years.

My friends and family back home and abroad mean the world to me, and while we've said it to each other before, I'll repeat it here: even though we're apart, we're never truly gone from each other's lives. **Ngina**, I miss you dearly, thanks for all the support and help that you've been able to give me over the years. **Christine** and **Lukas**, thanks for the letters, postcards, and pictures from your acreage. Whenever I miss Alberta or try to explain what Aspen parkland is, I now have some reference material. Three cheers for **Carson**, the copy editor of Chapter 1, as well as the thesis layout coordinator and transcription service for the live defence – thanks for everything. Thanks to **Dean**, **Eric W**, **Jenn**, **Jess**, **Natalie**, and **Sarah Z** for the drive-by dinners and meetups in Edmonton every year or so since I moved away, I really appreciate it.

To **Joe**, **Cindy**, **Allen**, and **Darla**, thanks for all of your love and support over the years. Love to **Candace**, **Nicki**, **Adam**, little **Finnick**, **Lil**, **Jean**, **Jeff**, **Brian**, **Kyle**, **Sandra**, **Tony**, **Janice**, **Clay**, **Rebecca**, **Shane**, **Wayne**, and all of my family back in North America. **Grandma Julie**, you mean the world to me and I cannot thank you enough for every opportunity that you've encouraged me to take. **Grandpa Ron**, **Grandma Dorothy**, and great **Uncle Fred**, I'm sorry

I wasn't there to say a last goodbye. I cannot imagine being here without the support you gave me through the years.

Aan mijn Nederlandse familie, de Blankvoorts– **Marian**, **Fons**, **Tim**, and **Michelle** –heel erg bedankt. Jullie hebben mij met open armen ontvangen bij jullie thuis en hebben mij veel geholpen door de jaren heen. We hebben al meerdere keren Sinterklaas, Kerstmis, en verjaardagen gevierd, en daar kijk ik altijd naar uit!

Thanks to my lovely sisters, Kathryn and Pam. **Kat**, I'm looking forward to showing you around the Netherlands and introducing you to the concept of *gezelligheid*. **Pam**, you're my best friend and I can't thank you enough for your love and support, I wish you the best in everything that you do. And thanks to my niece **Haley**, who's boundless enthusiasm gives me hope. To my parents, **Cheryl** and **Dave**, I cannot thank you enough for the endless love and support that you've provided over the years. You're both to blame/ thank for my stubborn nature, something that I needed to rely on to get this project to the finish line.

And finally, to my lovely **Erik Blankvoort**, my *suikerbietje*, you mean the world to me and I can't thank you enough for the love and support that you've provided through the years. You introduced me to so much– Utrecht, prog rock, Terry Pratchett –that has made my life fuller and happier. During the final months of the thesis you took care of me and our little piece of domesticity, all the while being woken up far too early by my alarm and turning in way too late when I was nearly a zombie. You're my favourite person in the whole wide world. Thank you.

About the Author



Kimberley Brianne Ferguson was born on March 24th, 1989 in Edmonton, Alberta, Canada. Kim grew up in Edmonton (ᐊᒥᐢᑲᐧᒋᐢᑲᐤᐃᑲᐣ or *amiskwaciwâskahikan*, in Plains Cree) and received her high school diploma from L'Academie Vimy Ridge Academy in 2007. She then went on to pursue a BSc General degree from the University of Alberta.

She obtained her BSc in 2012 with a major in Biological Sciences and a minor in English. During her last year, she studied northern ecology, political ecology, and conservation ecology, and as such was really into the idea of studying walruses along the Arctic coast.

Following a brief and enjoyable summer course abroad in Kassel, Germany during her Bachelor's, Kim traded root beer for *apfelsaftschorle* and undertook an International MSc in Ecology at Universität Bremen in autumn 2012. Through the course of her Masters, she worked with parasitoid wasps, and as cool as wasps are, Kim really wanted to work with mammals. She had given up on studying walruses, but was determined to keep that warm-blooded connection.

On a fortuitous visit the Netherlands Institute for Ecology (NIOO) in Wageningen, Kim learned of the *Licht op Natuur* project, where the possibility of working (indirectly) with bats was introduced. Fast forward a year later, and aided with a small grant from Bat Conservation International, Kim began her thesis work on monitoring bat hunting behaviour and insect abundance, under the oversight of Kamiel Spoelstra (NIOO). Kim graduated with her MSc in 2015. During her MSc thesis work, Kim met her partner, Erik, and fell in love with him and the Netherlands. She was determined to stay.

Ever the generalist, Kim sought out an interdisciplinary PhD that would expand her skillset and knowledge. Kim abandoned mammals entirely and began her PhD on insect genetics in autumn 2015 as part of an EU-funded Marie Skłodowska-Curie Action Innovative Training Network, the BINGO-ITN. Based in the Laboratory of Genetics at Wageningen University, her work on this project dealt with the genomes and genomics of natural enemies and is described in this thesis. During her PhD, Kim was an active member of the department's activity committee, was a member and chair of the Wageningen Evolution and Ecology Seminars (WEES), and helped organize the 4th Annual Wageningen PhD Symposium.

As fulfilling as her work on insect genomics and genomics has been, Kim will make a jubilant return to studying bats (and insects) for a brief post-doctoral position split between the NIOO and Wageningen University & Research. Beyond that, Kim will certainly stay in The Netherlands, if only because she has lost her cold tolerance and can no longer abide the Edmonton winter.

Appendices

List of Publications

Paspati, A., **Ferguson, K.B.**, Verhulst, E.C., Urbaneja, A., González-Cabrera, J., Pannebakker, B.A., 2019. Effect of mass rearing on the genetic diversity of the predatory mite *Amblyseius swirskii* Athias-Henriot (Acari: Phytoseiidae). Entomol. Exp. Appl. 167, 670–681. https://doi.org/10.1111/eea.12811

Spoelstra, K., van Grunsven, R.H.A., Ramakers, J.J.C., **Ferguson, K.B.**, Raap, T., Donners, M., Veenendaal, E.M., Visser, M.E., 2017. Response of bats to light with different spectra: light-shy and agile bat presence is affected by white and green, but not red light. Proc. R. Soc. B Biol. Sci. 284, 20170075. https://doi.org/10.1098/rspb.2017.0075

Submitted, available as preprint

Ferguson, K.B., Chattington, S.R., Plouvier, W.N., Pannebakker, B.A., 2020. Genetic Variation of Traits in Natural Enemies Relevant for Biological Control: A Systematic Review. Preprints: 2020010276. https://doi.org/10.20944/PREPRINTS202001.0276.V1

Leung, K., Ras, E., **Ferguson, K.B.**, Ariëns, S., Babendreier, D.B., Bijma, P., Bourtzis, K., Brodeur, J., Bruins, M., Centurión, A., Chattington, S., Chinchilla-Ramírez, M., Dicke, M., Fatouros, N., González Cabrera, J., Groot, T., Haye, T., Knapp, M., Koskinioti, P., Le Hesran, S., Lirakis, M., Paspati, A., Pérez-Hedo, M., Plouvier, W., Schlötterer, C., Stahl, J., Thiel, A., Urbaneja, A., van de Zande, L., Verhulst, E.C., Vet, L., Visser, S., Werren, J., Xia, S., Zwaan, B.J., Magalhães, S., Beukeboom, L., Pannebakker, B.A., 2019. Next Generation Biological Control: the Need for Integrating Genetics and Evolution. Preprints: 2019110300. https://doi.org/10.20944/preprints201911.0300.v1

In preparation

Ferguson, K.B., Pannebakker, B.A., Centurion, A., van den Heuvel, J., Nieuwenhuis, R., Becker, F.F.M., Schijlen, E., Thiel, A., Zwaan, B.J., Verhulst, E.C. Braconidae revisited: *Bracon brevicornis* genome showcases the potential of linked-read sequencing in identifying a putative complementary sex determiner gene.

Ferguson, K.B., Kursch-Metz, T., Verhulst, E.C., Pannebakker, B.A. Hybrid genome assembly and annotation of egg parasitoid and biological control agent *Trichogramma brassicae*.

Ferguson, K.B., Visser, S., Dalíková, M., Provazníková, I., Urbaneja, A., Pérez-Hedo, M., Marec, F., Werren, J. H., Zwaan, B.J., Pannebakker, B.A., Verhulst, E.C. Jekyll or Hyde? The genome (and more) of *Nesidiocoris tenuis*, a zoophytophagous predatory bug that is both a biological control agent and a pest.

PE&RC Training and Education Statement



With the training and education activities listed below, the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (4.5 ECTS)

Getting the lay of the land: a systematic review into current research on genetic heritability of phenotypic traits in natural enemies

Post-graduate courses (3.8 ECTS)

BINGO Summer school; University Bremen (2015)
BINGO Summer school; Koppert Biological Systems (2016)
Population genomics in R; University of Lausanne (2016)

Laboratory training and working visits (3 ECTS)

Cytogenetic techniques in insects; Czech Academy of Sciences, Institute of Entomology (2018)

Invited review of (unpublished) journal manuscript (2 ECTS)

Scientific Reports: *Trichogramma dendrolimi* rearing conditions (2017)
Entomologia Generalis: *Bradysia odoriphaga* and microsatellites (2017)

Competence strengthening / skills courses (2.9 ECTS)

Competence assessment; WGS (2016)
Workshop carousel; WGS (2017)
Writing scientific articles and proposals & leading effective discussions; BINGO Summer School (2017)
Workshop carousel; WGS (2018)
Ethics in plant and environmental sciences; WGS (2019)

PE&RC Annual meetings, seminars and the PE&RC weekend (2.7 ECTS)

PE&RC First year's weekend (2016)
PE&RC Day (2016-2019)
PE&RC Last year's weekend (2019)

Discussion groups / local seminars / other scientific meetings (5.8 ECTS)

Wageningen PhD symposium (2017)
Third FAO/IAEA International conference on area-wide management of insect pests; Vienna, AUT (2017)
WEES: Wageningen Ecology and Evolution Seminars (Chair, 2017-2018)
Entomologendag, NL (2018)
XI European congress of entomology; Napoli, ITA (2018)

International symposia, workshops and conferences (8.8 ECTS)

RES Special Interest Group: insect genomics meeting; oral presentation; UK (2017)
European Society for Evolutionary Biology XVI congress; oral presentation; NL (2017)
5th International Entomophagous Insects conference; oral presentation; JP (2017)
ESA, ESC, and ESBC Joint Annual Meeting Entomology; oral presentation; CAN (2018)

Lecturing / supervision of practicals / tutorials (3 ECTS)

Molecular and evolutionary ecology (2018)

Supervision of BSc students (3 ECTS)

Population genetics in a wild-caught *Trichogramma evanescens* population
In vivo support for horizontal gene transfer in *Nesidiocoris tenuis*

About the Artwork

Cover and inset pages, Chapters 1 and 8
Kelley Leung
The biocontrol insects, plant hosts, and provincial flowers of Kim B. Ferguson, 2019
Embroidery on linen
45 x 38 cm

Chapter 2
Stefan A. Christ
Bingo! A systematic review, 2020
Ink drawing/digital edited artwork

Chapter 3
Kirsten Oude Lenferink
Untitled, 2016
Ink on paper
9 x 9 cm

Chapter 4
Erica Ras
Unraveled, 2019
Papercut
21 x 30 cm

Chapter 5
Jitte Groothuis
What's the Tea?, 2019
Digital vector drawing

Chapter 6
Zhuyin (Sarah) Zhao
Going GAGA for Trichogramma, 2019
Linocut
20 x 25 cm

Chapter 7
Tessina De Lille
N. tenuis, 2019
Watercolour on paper
20 x 42 cm

Colophon

The research in this thesis was carried out at the Laboratory of Genetics at Wageningen University & Research, The Netherlands, and was financially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 641456.

Cover design: Kelley Leung
Thesis layout: Carson McConnell
Printed by: Digiforce

