

## To pest, or not to pest:

comparative transcriptomics  
and genomics towards  
an understanding of  
polyphagous pest Noctuidae

**Thijmen  
Breeschoten**

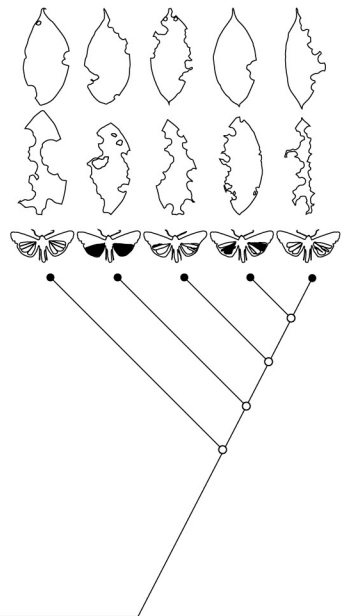
# Propositions

1. Species-specific transcriptional plasticity observed in polyphagous Noctuidae gives leads for sustainable pest-control.  
(this thesis)
2. Incorporating phylogenetic context into comparative transcriptomic and genomic studies is critical for understanding biological observations and evolutionary trends.  
(this thesis)
3. Total inclusion, respect and openness towards all forms of diversity will lead to maximum “scientific” production in academia.
4. Similar to arts and architecture, science benefits from minimalism and abstraction.
5. There is a lack of proper nature-based education and upbringing which causes disrespect, ignorance and abuse of our natural environment.
6. Covid measurements are a blessing for introverted people.

Propositions belonging to the thesis, entitled

To pest, or not to pest:  
comparative transcriptomics and genomics  
towards an understanding of polyphagous  
pest Noctuidae

Thijmen Breeschoten  
Wageningen, 8 September 2021



# **To pest, or not to pest:**

comparative transcriptomics and genomics towards an  
understanding of polyphagous pest Noctuidae

**Thijmen Breeschoten**

## **Thesis committee**

### **Promotor**

Prof. Dr M.E. Schranz  
Professor of Biosystematics  
Wageningen University & Research

### **Co-promotor**

Dr S. Simon  
Assistant Professor, Biosystematics Group  
Wageningen University & Research

### **Other members**

Prof. Dr Bas J. Zwaan, Wageningen University & Research, the Netherlands  
Prof. Dr Astrid T. Groot, University of Amsterdam, the Netherlands  
Dr Franziska Beran, Max Planck Institute for Chemical Ecology, Jena, Germany  
Dr Christopher W. Wheat, Stockholm University, Sweden

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences



# **To pest, or not to pest:**

comparative transcriptomics and genomics towards an  
understanding of polyphagous pest Noctuidae

**Thijmen Breeschoten**

## **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 8 September 2021  
at 11 a.m. in the Aula.

Thijmen Breeschoten

To pest, or not to pest: comparative transcriptomics and genomics towards an understanding of polyphagous pest Noctuidae, 176 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)

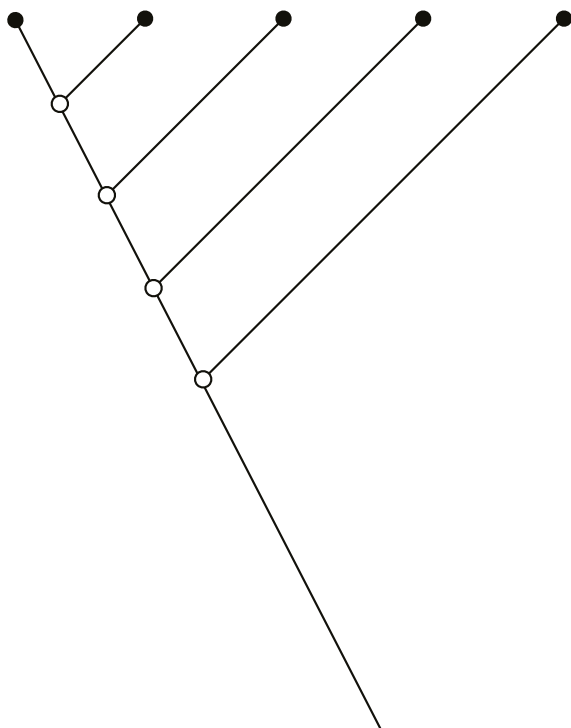
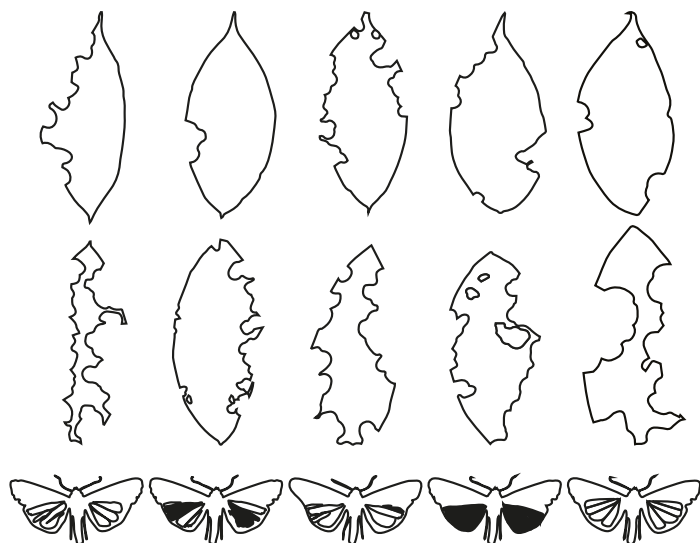
With references, with summaries in English and Dutch

ISBN: 978-94-6395-915-5

DOI: <http://doi.org/10.18174/550726>

## Contents

<b>Chapter 1</b>	General introduction	<b>7</b>
<b>Chapter 2</b>	An influential meal: host plant dependent transcriptional variation in the beet armyworm, <i>Spodoptera exigua</i> (Lepidoptera: Noctuidae)	<b>21</b>
<b>Chapter 3</b>	Family dinner: transcriptional plasticity of five Noctuidae (Lepidoptera) feeding on three host plant species	<b>45</b>
<b>Chapter 4</b>	Expanding the menu: are polyphagy and gene family expansions linked across Lepidoptera?	<b>75</b>
<b>Chapter 5</b>	Genome and transcriptome analysis of the beet armyworm <i>Spodoptera exigua</i> reveals targets for pest control	<b>99</b>
<b>Chapter 6</b>	General discussion	<b>127</b>
	References	<b>141</b>
	Summary	<b>163</b>
	Samenvatting	<b>165</b>
	Acknowledgements	<b>168</b>
	About the author	<b>171</b>
	List of publications	<b>172</b>
	Education statement	<b>173</b>



# Chapter 1

General introduction

Thijmen Breeschoten



## Background

The coevolutionary relationship between land plants and herbivorous insects dates back over 400 million years (Grimaldi & Engel, 2005; Kant et al., 2015; Schoonhoven et al., 2005). Plants (Kingdom Plantae), being the primary energy producers for heterotrophic animal life on earth, form an indispensable part of the ecosystem. Insects (Class Insecta) are an extremely species-rich clade with species estimates, combining both described and undescribed species, of 2 to more than 30 million (Erwin, 1997; Grimaldi & Engel, 2005; Ødegaard, 2000). Nearly half of all described eukaryotes are herbivorous insects (Birnbaum & Abbot, 2020; Strong et al., 1984). The fossil record of insects and their diversity in mouthparts indicate the early relationship and specialised feeding on Carboniferous plants (Grimaldi & Engel, 2005). Despite the large-scale diversity decrease during the Cretaceous mass extinction, both insects and plants further diversified over the last 65 million years. The recovery and diversification of herbivorous insect groups occurred in close association with the radiation of flowering plants (Campbell et al., 2008; Grimaldi & Engel, 2005). Indeed, the radiation of angiosperms took place during the Palaeocene, after the Cretaceous mass extinction (Lupia et al., 1999; Manchester et al., 2015; Ramírez-Barahona et al., 2020). It may have been the speciose insects, with their diversity in life forms and selective forces on plant diversity, that have influenced the evolution of flowering plants (Marquis, 2004).

The close relationship and the evolutionary interactions between plants and insects was already referenced by Charles Darwin in *On the Origin of Species* (1859). However, it was the influential paper of Ehrlich and Raven (1964) when the term “coevolution” was introduced, using butterflies as a model, to describe the close insect-plant interactions and impact on the evolution and diversity of these clades. The butterflies and moths (Lepidoptera) have been a focus group for the study of herbivorous insect-plant relationships and evolution. Ehrlich and Raven (1964) focused on how the close associations between butterflies and plants were shaped by coevolutionary forces, where plants on one side evolve defences while herbivores evolve new ways to cope with these. Plants with a successful defence, in the absence of herbivory, are able to diversify, while the successful insects through adaptation are able to diversify on related plants with similar defences. However, reanalysis and new insights have re-shaped this view on the link between diversification of insects and expanding angiosperms and the close associations between butterflies and plants (Labandeira & Sepkoski, 1993).

The patterns of butterfly-host plant associations were primarily shaped by host shifts between closer related plants and a higher tendency of recolonising ancestral hosts (Janz & Nylin, 1998; Jermy, 1976, 1984). The similarity in biochemical defences employed by plants is a major determinant for successful host shifts, which is partially correlated by phylogenetic relatedness of the plants (Menken et al., 2010; Wheat et al., 2007). Further, it was shown that plant groups colonised by butterflies and moths are often much older and that they diversified before the colonisation event (Janz & Nylin,

1998; Labandeira & Sepkoski, 1993; Lopez-Vaamonde et al., 2006; Menken et al., 2010). Underutilisation of plant groups by Lepidoptera can be explained due to evolved defences against other insects that worked effectively against butterflies. Some plant groups may never have been colonised by butterflies because phylogenetic distant colonisations are rare (Janz & Nylin, 1998). As a consequence, the butterfly-plant associations are rather shaped by host shifts, through colonisation and specialisation processes (Braby & Trueman, 2006; Janz & Nylin, 1998; Jermy, 1984; Menken et al., 2010), likely affected by similarity of biochemical defences of the host plants (Menken et al., 2010).

Further, the diversity seen in butterflies, and herbivorous insects in general, cannot be explained by specialisation. Specialisation does not generate diversity by itself, and therefore it is argued that it is likely that diversity of host use (e.g. the variation in host colonisations and ranges) drives species richness (Janz & Nylin, 2008; Janz et al., 2006; Nylin et al., 2014). Janz et al. (2006) proposed that diversification of herbivorous insects was driven by oscillations in host plant range, host expansions allow the insects to increase their distributions which may be followed by secondary specialisation on different hosts, e.g. by allopatric isolation. In conclusion, the success of herbivorous insects in terms of species richness may not be due to stable processes of specialisation and cospeciation but rather by dynamic oscillating host ranges (Janz & Nylin, 2008; Janz et al., 2006).

## The mechanisms and dynamics of insect-plant interactions

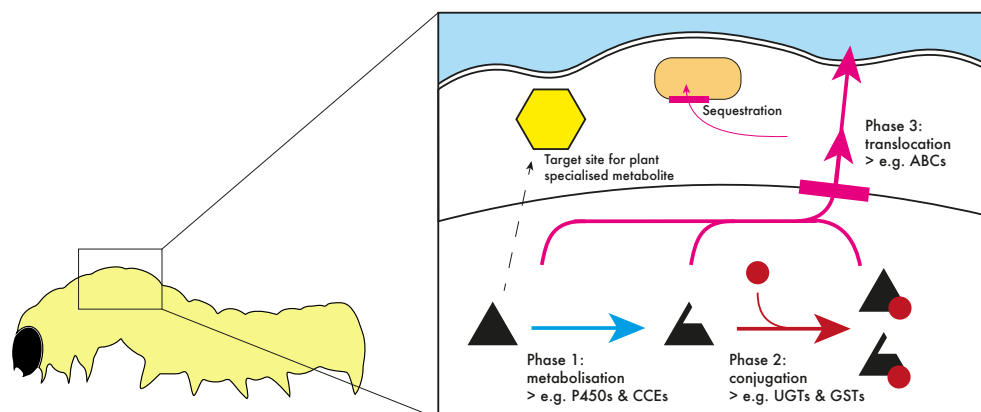
The influential paper by Ehrlich and Raven (1964) specifically stressed the importance of biochemical (specialised) defences employed by plants in insect-plant interactions:

*“The evolution of secondary plant substances and the stepwise evolutionary responses to these by phytophagous organisms have clearly been the dominant factors in the evolution of butterflies and other phytophagous groups.”*

Herbivorous selective forces on plants have resulted in a wide range of adaptations against herbivory that can be roughly divided into ‘constitutive defences’ (physical barriers that are present at all times like thorns) and ‘induced defences’ (defences that become operational upon attack) (Howe & Jander, 2008; Kant et al., 2015). The production of chemical defences, or specialised metabolites (also often called ‘secondary metabolites’, ‘plant allelochemicals’ or defensive ‘phytochemicals’), are an important induced defence system that can deter or poison attacking herbivores, e.g. by interfering with its metabolism (Whittaker & Feeny, 1971). The specialised metabolites employed by plants are important components of an active defence mechanism (Bennett & Wallsgrove, 1994). There is an enormous diversity of specialised metabolites in plants which can be widespread or taxon-specific. Still, based on their biosynthetic composition they can be classified into three basic groups: phenolics, terpenoids and N-containing compounds (Kant et al., 2015).

Insects evolved adaptive ways to handle specialised metabolites encountered during plant feeding. The responses are important selective forces in the interaction with plants and are roughly divided into three main groups: behavioural (by adjusting feeding patterns), physiological (sequestration or storage), and biochemical (metabolising specialised metabolites) (Brattsten, 1988). The biochemical response includes the detoxification of ingested chemical defences by the insect.

Detoxification of specialised metabolites most commonly occurs in three phases (Figure 1). Phase 1 is metabolisation, a nucleophilic functional group (hydroxyl, carboxyl or amine group) is incorporated to the metabolite making it more polar/hydrophilic. Phase 2 is conjugation, the metabolite is conjugated to endogenous molecules (e.g. glutathione or sugar), further increasing the polarity and hydrophilicity of the compound. Finally, in phase 3 the conjugated compound is transported away for excretion by cellular transporters (Brattsten, 1988; Brattsten, 1992; Després et al., 2007; Kant et al., 2015).



**FIGURE 1** | The herbivore insect resistance response to plant specialised metabolites. Herbivorous insects have adaptive ways to deal with plant produced toxins, 1) behavioural, e.g. by avoidance (not shown), 2) physiological, e.g. by sequestration, and 3) biochemical, e.g. by metabolising the toxin by use of the detoxification pathway. This detoxification pathway metabolises the toxin in three phases: metabolisation by e.g. enzymes from cytochrome P450 monooxygenases (P450s) and carboxyl- and cholinesterases (CCEs), conjugation by e.g. enzymes from UDP-glycosyltransferases (UGTs) and glutathione-S-transferases (GSTs) and transportation by e.g. enzymes from ATP-binding cassettes (ABCs). During phase 1 and 2 the xenobiotic metabolite becomes more water soluble and easier excreted during phase 3.

Enzymes involved in this pathway often belong to commonly occurring multi-gene families. These large ubiquitous gene families involved in detoxification are generally present in all insects, but often a smaller number of enzymes can metabolise specific specialised metabolites (Heidel-Fischer & Vogel, 2015). Cytochrome P450 monooxygenases (P450s) might be the best-known detoxification gene family, and is associated with adaptations to plant defences in herbivorous insects (Feyereisen, 1999; Schuler, 2011, 2012). The P450 gene family is a diverse and large family, occurring



in all organisms, and a large part of its members are involved in phase 1 of detoxification in herbivorous insects (Dermauw et al., 2020; Feyereisen, 1999; Feyereisen, 2006; Feyereisen, 2011). Additionally, carboxyl- and cholinesterases (CCEs) are important enzymes also involved in this first phase of detoxification. Enzymes of the CCE gene family have a wide range of roles, but are especially important in the metabolism of both endogenous and exogenous compounds (Montella et al., 2012). UDP-glycosyltransferases (UGTs) are important enzymes of phase 2 detoxification, that in insects catalyse the conjugation of the metabolite mainly with glucose (Bock, 2016). It is suggested that UGTs fulfil a variety of rolls including the deactivation of pheromones but their role in detoxification of xenobiotics is shown in all living organisms (Bock, 2016; Hu et al., 2019). Another group of enzymes involved in phase 2 are the glutathione-S-transferases (GSTs). These enzymes are well recognised to be involved in detoxification of plant defences and have received much attention for their role in insect resistance against DDT (Armstrong, 1997; Brattsten, 1988; Francis et al., 2005; Shi et al., 2012). Finally, a fifth gene family well known to be involved in the final part of the detoxification pathway are the ATP-binding cassettes (ABCs) that translocate the conjugated compound across membranes for excretion (Bretschneider et al., 2016; Dermauw & Van Leeuwen, 2014; Sorensen & Dearing, 2006).

## There are multiple ways to be a herbivorous insect

While the defences employed by plants will effectively deter many herbivorous insects, generally one or a few insect species have evolved ways to cope with plant-specific defences and have specialised on feeding on any given plant. Indeed, the majority of herbivorous insects are specialists, or monophagous species, on a small number of related host plants (Bernays & Graham, 1988; Heckel, 2018; Schoonhoven et al., 2005). Within phytophagous insects, the evolutionary trend is to specialise, with transition rates from generalisation to specialisation significantly higher than the reverse (Nosil, 2002). I consider monophagous species as herbivorous insects that feed on plants within a single host plant family. Monophagous species have specialised on detoxification of a small range of specialised metabolites. The insect may deploy its countermeasures even constitutively and might use these compounds for its own benefits, e.g. for feeding stimulants or sequestration purposes (Heckel, 2018). Often, monophagous species have evolved biochemical adaptations to efficiently detoxify encountered chemical defences of their host plants. In Pieridae butterflies, the whites, such an adaptation to specific chemical defences has evolved after the evolution of their hosts, the Brassicales (Wheat et al., 2007). The Brassicales employ the glucosinolate-myrosinase system as their chemical defence against herbivory. Feeding damage ensures a chemical response between the myrosinase enzymes with glucosinolates resulting in formation of toxins like isothiocyanates (ITCs) (Fahey et al., 2001; Halkier & Gershenzon, 2006; Kliebenstein et al., 2005). In Pieridae, a nitrile-specifier protein

(NSP) coevolved to modify the toxic degradation product of glucosinolates into less toxic nitriles (Fischer et al., 2008; Okamura et al., 2019; Wheat et al., 2007; Wittstock et al., 2004). A different adaptation against the glucosinolate-myrosinase system was identified within another monophagous species on Brassicaceae, the diamondback moth (*Plutella xylostella*). Within this species a glucosinolate sulfatase enzyme (GSS) evolved that desulfates the glucosinolates, preventing the hydrolysis by myrosinases and thus prevents the formation of toxins (Heidel-Fischer et al., 2019; Ratzka et al., 2002).

In contrast to monophagous species, a minority of herbivorous insects are generalists, or polyphagous species, feeding on a larger range of host plant species belonging to different plant families (Heckel, 2018; Schoonhoven et al., 2005). I consider polyphagous species as herbivorous insects that feed on plants belonging to different host plant families. The ability of polyphagous species to cope with the large variety in chemical defences employed by different plant families is remarkable (Agrawal et al., 1999; Kant et al., 2015). Previously, it was suggested that polyphagous species had a higher capacity of their detoxifying enzymes in order to detoxify all encountered biochemical defences (Krieger et al., 1971). Later this was refined to suggest that polyphagous species may employ a detoxification system with broader substrate ranges (e.g. in P450 enzymes (Berenbaum et al., 1992; Li et al., 2004)). Indeed, monophagous species encounter known specialised metabolites and thus their detoxification mechanism needs less flexibility (Li et al., 2004). Polyphagous species need to use diverse and flexible detoxification strategies to be able to detect and cope with the variety of specialised metabolites in their diet (Heckel, 2018). This flexibility is shown in plasticity of the transcriptional detoxifying response, which is likely needed to overcome the diversity in plant defences (Koenig et al., 2015; Mathers et al., 2017; Orsucci et al., 2018). However, even though polyphagous insects have the ability to cope with chemical defences from a wide variety of plant families, a degree of specialisation may occur. For example, populations of polyphagous species may have a restricted host plant choice in terms of successful feeding and development within their larger range of host plant families (Foss & Rieske, 2003; Gouin et al., 2017; Howard et al., 1994; Schoonhoven et al., 2005). This difference in host use and feeding success is of importance to understand and anticipate potential pest formations (species forming pests) by polyphagous insects. However, it is largely unknown if the transcriptional plastic pattern in response to successful host plant feeding is shared across polyphagous insects.

## Insect pest formations

Monophagous and polyphagous insects can compete with humans for crop plants. If a crop falls within the range of favourable host plants of monophagous and/or polyphagous insects and is accepted as a host, e.g. within agricultural environments,

we identify this species as a pest. Strictly speaking a species cannot “be” a pest but rather forms a pest, since in absence of the attacked crop the insect can no longer be a pest. The concept of pests is a human construct, lacking any objective criteria, based on cultural values and economic importance (Kim, 1993). In a broad sense, pests are defined as organisms that form a hindrance of any form to humans (Kim, 1993). Recent concerns of the biodiversity crisis extended this definition further to organisms that cause a risk for native species and natural ecosystems (Worner, 2002; Worner & Gevrey, 2006).

Within agricultural environments pest insects are competing with humans for common crop resources. Herbivorous insects may become pests if a host plant species becomes available as a human-introduced crop, e.g. by crop-plant introductions. For insects to form pests means that herbivore success for the specific crop is high, in terms of successful feeding and development. There are various aspects of an insect species’ physiological and behavioural characteristics that contribute to the likelihood of becoming a pest. For example, the host plant range should include the crop-plant species of interest. Further, multiple generations per year (multivoltinism) enables a rapid growth of pest populations and the resulting dense population sizes are accepted and preferred by the insect species (Chew, 1995; Schoonhoven et al., 2005).

The number of insect species forming pests is relatively small compared to the large number of insect species described. Worldwide, around 9,000 insect species may attack agricultural crops, of which less than 5% form serious threats (Hill, 2012; Schoonhoven et al., 2005). Insect pests can both be native and invasive species, but the majority of pests are formed by native species (Pimentel, 1991, 1993). However, due to the impact of climate change the number of invasive pest species have increased in recent times, for example due to altered distribution ranges and life history traits (Battisti & Larsson, 2015; Dukes et al., 2009; Yan et al., 2017; Ziska et al., 2011). Indeed, invasive pests often cause extreme high damage rates and form major treats to economical important crops worldwide. For example, the fall armyworm, *Spodoptera frugiperda*, is currently perhaps the most devastating pest worldwide. The species is invasive to large parts of the world including the African, Asian and Australian continents where it forms a major treat to the agricultural sector (Early et al., 2018; EPPO, 2019; Goergen et al., 2016; Jing et al., 2020; Kalleshwaraswamy et al., 2018).

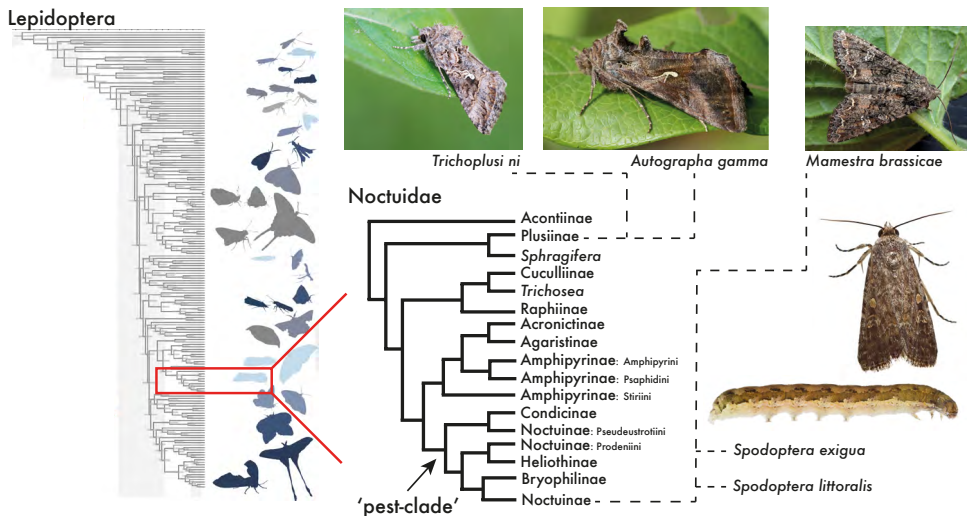
## Lepidoptera genomics and transcriptomics

Since the discovery of Sanger sequencing of DNA (Sanger et al., 1977) and especially since fast developments in high-throughput sequencing techniques (Goodwin et al., 2016), Lepidoptera research quickly advanced to a genomic scale (Triant et al., 2018). Studies implementing genome data has brought us new insights in lepidopteran evolution, from chromosomal rearrangements (Hill et al., 2019), to potential large

scale duplication events (Li et al., 2018) and gene family expansion rates (Thomas et al., 2020). The success and evolution of Lepidoptera is closely linked to their relation with plants. The coevolutionary interactions between plants and Lepidoptera, with evolving plant defences and counter adaptations by the herbivore, has resulted in gene and genome dynamics that fuel adaptations and diversification rates (Allio et al., 2021; Edger et al., 2015; Gloss et al., 2019; Simon, d'Alencon, Guy, Jacquin-Joly, Jaquier, et al., 2015). Indeed, the origin of plant defences and evolved counterresponses are associated with shifts in diversification rates in Lepidoptera (Edger et al., 2015). Furthermore, the dynamics of plant use in terms of host colonisations and host shifts have played profound roles in accelerated diversification rates in Lepidoptera (Allio et al., 2021; Fordyce, 2010; Menken et al., 2010; Nylin et al., 2018).

Butterflies and moths are placed in the insect order Lepidoptera and form one of the major radiations of insects, comprising ~160,000 species (Grimaldi & Engel, 2005; Pogue, 2009; van Nieuwerkerken et al., 2011). Adult Lepidoptera have various highly distinguishable characteristics, making them often easy to recognise, which includes a proboscis to suck up liquids for feeding and scale-covered wings and body. These scales, modified hairs, form often brightly coloured species-specific patterns (Resh & Cardé, 2009). The Lepidoptera is a highly versatile and diverse group with species living in almost all terrestrial ecosystems and showing a large variation of ecological adaptations. Phylogenetic reconstructions showed that butterflies likely became day active to feed on day-blooming flowers, showing their close connection with flowering plants (Kawahara et al., 2019). The common ancestor of Lepidoptera was likely an internal feeder of non-vascular land plants during the late Carboniferous, while the majority of Lepidoptera diversified with the rise of angiosperms in the Cretaceous (Bazinet et al., 2017; Kawahara et al., 2019; Mitter et al., 2017; Wiegmann et al., 2000). This close connection and shared history with plants make Lepidoptera ideal candidates for studies related to herbivory, and thus serve as model organisms for research in the field of ecology, pollination and agriculture (Roe et al., 2009; Triant et al., 2018). The role of Lepidoptera in terrestrial ecosystems, the importance as model in plant-insect research and the impact many species have as pests likely promoted studies on their phylogeny and evolution (e.g. Kawahara et al., 2019; Mayer et al., 2021; Mitchell et al., 2006; Regier et al., 2017; Regier et al., 2009; Zahiri et al., 2013). A robust phylogeny is of importance in order to incorporate evolutionary relationships among species in comparative (genomic) studies and to understand the biology of species within a phylogenetic context (Dunn et al., 2018). The inclusion of a robust comparative phylogenomic framework is of specific interest in predictions on gene and genome evolution and insect-plant studies focussing on adaptive evolutionary patterns (Dunn et al., 2018; Edger et al., 2015; Wheat et al., 2007). Within Lepidoptera much phylogenetic work is done on the cutworm moth family, Noctuidae, likely in part due to their importance for agriculture (Keegan et al., 2021; Mitchell et al., 2006; Regier et al., 2017; Zahiri et al., 2013).

The Noctuidae are one of the most speciose radiations of butterflies and moths, comprising ~11,700 species (Regier et al., 2017; van Nieukerken et al., 2011). Noctuidae, as most Lepidoptera, are plant feeding during their larval phase (commonly called caterpillars) and nectar feeding as adults, and form a prominent element of ecosystems. Many Noctuidae species are polyphagous, being able to feed on a large range of host plant families, and form one of the most damaging groups to agriculture (Mitchell et al., 2006; Regier et al., 2009). This even led to the naming of a ‘pest clade’ in recognition of the high level of risk for agriculture formed by Heliothinae, Noctuinae and several smaller groups, such as *Spodoptera* (Mitchell et al., 2006; Regier et al., 2017) (Figure 2). The major impact that Noctuidae, and other herbivorous insect clades, have on agriculture likely resulted in an increased interest in studying the molecular and genetic mechanisms underlying herbivory.



**FIGURE 2** | Phylogenetic placement of the cutworm moths, Noctuidae, within the Lepidoptera (adjusted from Kawahara et al. (2019)). Phylogeny of the Noctuidae (adjusted from Regier et al. (2017)), with an indication of the ‘pest-clade’. Further, placement of the five cutworm moths studied in Chapter 3 of this thesis are indicated: the cabbage looper (*Trichoplusia ni*), the silver-Y moth (*Autographa gamma*), the cabbage moth (*Mamestra brassicae*), the African cotton leafworm (*Spodoptera littoralis*; not shown) and the beet armyworm (*Spodoptera exigua*; larvae and adult). Photo credits: De Vlinderstichting (J. Voogd, T.n.; J.v. Deijk, A.g.; K. Wingelaar, M.b.) and C.F.H. v.d. Linden (S.e).

The implementation of genome data has put studies focusing on the genetic basis of host plant use and host shifts into a new perspective (Allio et al., 2021; Nallu et al., 2018). For example, the genomes of polyphagous species are often associated with expansions in gene families involved in plant feeding, such as chemoreception and detoxification (Calla et al., 2017; Xu et al., 2016). Observed expansions in major

polyphagous species of the cutworm moth genus *Spodoptera* are suggested to be adaptations to enable polyphagy in order to deal with the large variation in host plants and their chemical defences (Cheng et al., 2017; Gouin et al., 2017).

Further, transcriptome data sheds light on the role of transcriptional plasticity in plant use by herbivorous insects. Indeed, understanding the role of gene expression in plant feeding and host plant range will help us understand the interactions between plants and insects and eventually the mechanisms that generate diversification (Birnbaum & Abbot, 2020). These studies use different herbivorous insect species as a model, but most findings may be applied to herbivorous insects in general, including Lepidoptera.

Much effort has been put into deciphering the differences between monophagous and polyphagous species in transcriptional response to plant feeding (Ali & Agrawal, 2012). Polyphagous insects generally display a plastic transcriptional response which facilitates a broad niche. Indeed, polyphagous insects demonstrate a greater expression plasticity in comparison to monophagous species (Birnbaum & Abbot, 2020; Govind et al., 2010; Roy et al., 2016; Schweizer et al., 2017). In contrast, a targeted gene repertoire towards specific plant defences is suggested to be employed by monophagous species (Celorio-Mancera et al., 2013; Govind et al., 2010; Schweizer et al., 2017). Their detoxification mechanism is suggested to have a more targeted specificity which may reflect host plant adaptation (Heckel, 2018). Besides differences between monophagous and polyphagous species, much focus of transcriptome studies has been put to understand gene expression and plasticity in (novel) host plant adaptations (e.g. Dermauw, Wybouw, et al., 2013; Hoang et al., 2015; Mathers et al., 2017; Orsucci et al., 2018; Wybouw et al., 2015; see Birnbaum and Abbot, 2020 for an extensive overview). It was shown that primarily the expression differences in genes related to digestion and detoxification facilitates host shifts (Birnbaum & Abbot, 2020). Indeed, host plants represent different diets and thus gene expression must allow some plasticity to allow dietary differences (Birnbaum & Abbot, 2020).

Another point of focus in plant-insect transcriptomic studies are species for interest for agriculture. Given the impact of insect pest formations on agriculture, various studies focus on the genomics and transcriptomics that shape and form insect pests (Kirk et al., 2013). The comparison of the molecular and evolutionary mechanisms of insect response to specialised metabolites may help in controlling insect pest formations (Després et al., 2007). Many studies focus on Lepidoptera given their importance in pest formations. For example, it was suggested that a population of the cabbage white butterfly, *Pieris rapae*, is adapted to an agricultural environment through selection on standing genetic variation (Sikkink et al., 2017). Further, there is a particular focus on polyphagous Lepidoptera, likely due to their role as agricultural pests (Birnbaum & Abbot, 2020), with a specific interest in members of the moth family Noctuidae. Indeed, various genomic adaptations, including expanded gene families, are suggested to be linked to polyphagy and pest potential within species of

this family (e.g. Cheng et al., 2017; Fritz et al., 2018; Gouin et al., 2017; Pearce et al., 2017; Xiao et al., 2020). The question what genetic basis underlines polyphagy and pest formations and to what degree this is shared among clades is of importance for both (co)evolutionary and agricultural point of view.

## Thesis aim and outline

My thesis aims to elucidate the genomic and evolutionary basis of polyphagous pest insect species. Many pest insects that are most damaging for agriculture are polyphagous herbivorous species. I focus on the Lepidoptera, with a particular focus on the cutworm moth family Noctuidae as a model group. Many species within the Noctuidae are major polyphagous species and form pests most injurious to agriculture.

Our knowledge on plant-insect interactions have greatly expanded in the past years, with much progress in understanding the genetics and adaptations of plant defence mechanisms and the way herbivorous insects adapt to these. However, there is a dearth of robust knowledge on what defines a polyphagous insect pest on a genetic scale and the differences and commonalities between pest forming species. Therefore, I aim to understand the pest concept on a genetic basis by examination of transcriptional dynamics related to polyphagous herbivory and pest formations within a phylogenomic framework. Further, I study gene family evolution patterns that may be related to polyphagy and of potential importance to pest formations. Moreover, I present preliminary steps for the identification of target genes for pest control in *Spodoptera* by analysing the genome of the beet armyworm, *Spodoptera exigua*, as presented here. The chapters in this thesis have their own focus but built upon each other. Combined they add to our knowledge on the transcriptional and genetic mechanisms of polyphagy linked to pest formations.

**Chapter 2** focuses on the gene expression of the polyphagous herbivore noctuid *Spodoptera exigua* (Figure 3) feeding on three selected host plant species and a control. Herbivore success of the larvae for each host plant species, as a quantification for ‘pest level’, is linked to host plant-specific expression patterns. By studying the transcriptional plasticity while feeding on different host species I am able to examine the genetic mechanisms of a polyphagous feeding style linked to herbivore success levels.

**Chapter 3** builds upon the knowledge of the transcriptional plasticity deployed by *S. exigua* feeding on a range of host plant species by extending the study system. I establish and utilise a comparative phylogenomic framework of five closely related species from the Noctuidae: the cabbage looper (*Trichoplusia ni*), the silver-Y moth (*Autographa gamma*), the cabbage moth (*Mamestra brassicae*), the African cotton leafworm (*Spodoptera littoralis*) and the beet armyworm (*Spodoptera exigua*). All species are notorious pests and major polyphagous species. However, within their

range of accepted host plant species they show a level of specialisation for particular host plant species. This high herbivore success for particular hosts, which again can be seen as a quantification for ‘pest level’, does differ between the moths. This allows me to examine and compare gene expression patterns employed by these polyphagous moth species feeding on different host plant species under various level of herbivore success. The comparison is done within a comparative framework that made it possible to discriminate transcriptional patterns due to a shared evolutionary history from patterns linked to high herbivore success. Placing transcriptomic data in a comparative phylogenomic framework, by comparing two closely related *Plusiinae* species with three closely related *Noctuinae* species (Figure 2), allows me to identify genetic signatures specific to pest formations among different lineages. Analysing the gene expression activity in the context of this framework additionally enables the identification of convergent or divergent patterns between the species that are subject to similar or different selection pressures. I evaluate if successful herbivory is determined by a shared or lineage-specific transcriptional response. This study not only sheds light on the transcriptional plasticity of polyphagy but also examines the genetic and transcriptional basis of pest formations.

In **Chapter 4** I look at adaptive genomic evolution patterns linked to polyphagy. I test if gene family evolution, in terms of gene gains and losses, can be linked to a polyphagous feeding style. I focus on gene families involved in plant feeding. Gene family evolution patterns that enable polyphagy might enable pest formations, and thus the study of these adaptive genomic changes is of interest here. Gene family expansions of families involved in plant feeding, such as detoxification, are hypothesised to enable a polyphagous feeding style. I use an extensive comparative genomic approach of 37 lepidopteran genomes to study the presence of the association between polyphagy and gene family expansions.

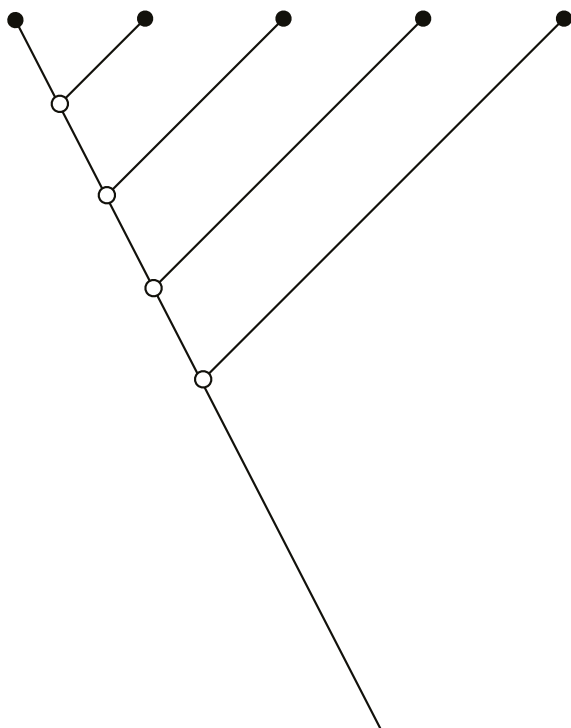
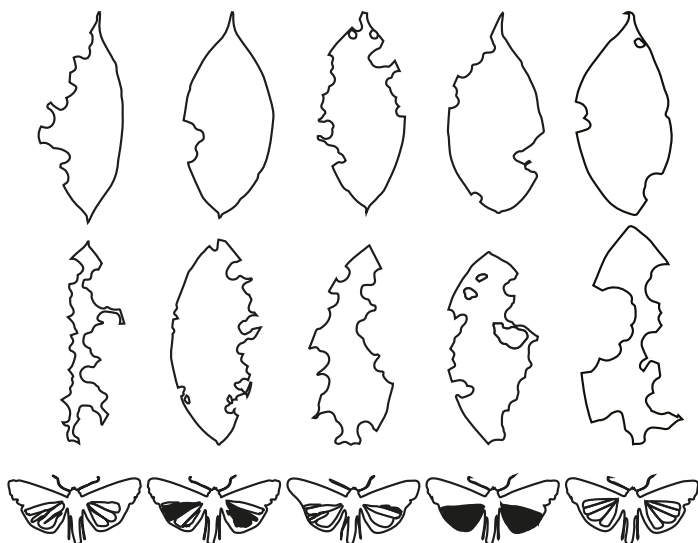
In **Chapter 5** I return to the polyphagous pest herbivore *Spodoptera exigua* of Chapter 2 and dive deeper into the genomics of this species. I present the genome assembly and perform various analyses comparing published *Spodoptera* and other lepidopteran genomes. I further perform a developmental transcriptional analysis and examine *Spodoptera*-specific genes as preliminary steps to identify genes for targeted pest control in *Spodoptera*.

Finally, in **Chapter 6** I discuss the combined results found. I cover the implications and incorporate the new insights on our knowledge of the genetic mechanisms of polyphagy, implications for the pest definition, and future perspectives.





**FIGURE 3** | Different life stages of the beet armyworm, *Spodoptera exigua*. Shown are an egg clutch, larvae in third developmental stage, pupa, and female adult (not to scale). Photos made by Corné F. H. van der Linden.



# Chapter 2

An influential meal: host plant dependent transcriptional variation in the beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae)

Thijmen Breeschoten<sup>1,\*</sup>,  
Vera I. D. Ros<sup>2</sup>,  
M. Eric Schranz<sup>1</sup>,  
Sabrina Simon<sup>1</sup>

<sup>1</sup>Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>2</sup>Laboratory of Virology, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands



To understand the genetic mechanisms of insect herbivory, the transcriptional response of insects feeding on different host plant species has to be studied. Here, we generated gene expression data of the generalist herbivore *Spodoptera exigua* (Hübner) feeding on three selected host plant species and a control (artificial diet). The host plant species used in this study –cabbage (*Brassica oleracea*), maize (*Zea mays*) and tobacco (*Nicotiana tabacum*)– are members of different plant families that each employ specific defence mechanisms and toxins.

*Spodoptera exigua* larvae had a higher growth rate, indicator for herbivore success, when feeding on *Z. mays* compared to larvae feeding on *B. oleracea* or *N. tabacum*. Larvae feeding on the different host plant species showed divergent transcriptional responses. We identified shared and unique gene expression patterns dependent of the host plant species the larvae fed on. Unique gene expression patterns, containing uniquely upregulated transcripts including specific detoxification genes, were found for larvae feeding on either *B. oleracea* or *N. tabacum*. No diet-specific gene cluster was identified for larvae feeding on the host for which larvae showed optimal herbivore success, *Z. mays*, or artificial diet. In contrast, for larvae feeding on hosts for which they showed low herbivore success, specific diet-dependent gene clusters were identified. Functional annotation of these clusters indicates that *S. exigua* larvae deploy particular host plant-specific genes for digestion and detoxification.

The lack of a host plant-specific gene activity for larvae feeding on *Z. mays* and the artificial diet suggest a general and non-specific gene activity for host plants with optimal herbivore success. Whereas the finding of specific gene clusters containing particular digestion and detoxifying genes expressed in larvae feeding on *B. oleracea* and *N. tabacum*, with low herbivore success, imply a host plant-specific gene activity for larvae feeding on host plants with suboptimal herbivore success. This observation leads to the conclusion that a polyphagous herbivore is able to feed on a large variation of host plants due to the flexibility and diversity of genes involved in digestion and detoxification that are deployed in response to particular host plant species.

#### Keywords

Transcriptomics; gene expression; detoxification; herbivory; generalist; host specialisation; polyphagy; RNA-seq

#### Abbreviations

Cytochrome P450 monooxygenase (P450), Carboxyl/Cholinesterase (CCE), Glutathione-S-transferase (GST), UDP-glycosyltransferase (UGT), ATP-binding cassette transporters (ABC), (GO category) Biological Processes (BP), (GO category) Molecular Functions (MF), Differentially Expressed (DE), Count Per Million (CPM), Trimmed Mean of M values (TMM)

## Introduction

The 400 million years of interaction and coevolution between plants and insects has led to a wide diversity of plant defences, to which herbivorous insects in turn have evolved a diverse array of resistance and detoxification mechanisms (Kant et al., 2015). Numerous herbivorous insects have evolved the ability to feed on a large range of host plant species (polyphagy), thereby encountering a variety of plant-specific defence compounds (Després et al., 2007; Schoonhoven et al., 2005). The ability of an herbivorous insect to feed on different host plants does not imply equal herbivore success on each of these plants. This success is partly dependent on nutrient content and plant defence resistance and is reflected by growth- and developmental rate of the insect (Merkx-Jacques et al., 2008; Saeed et al., 2017). In herbivorous insects the detoxification of plant defence compounds follow a three phased pathway involving members of several known enzyme families. In phase I of the detoxification process, cytochrome P450 monooxygenases (P450s) and carboxyl/cholinesterases (CCEs) metabolise toxins (Kant et al., 2015). The metabolised product is conjugated by phase II enzymes such as UDP-glycosyltransferases (UGTs) and glutathione-S-transferases (GSTs), and transported out of the cell by transporters like ATP-binding cassettes (ABCs) and solute carrier proteins (SLC) in phase III (Heidel-Fischer & Vogel, 2015; Kant et al., 2015; Li et al., 2007).

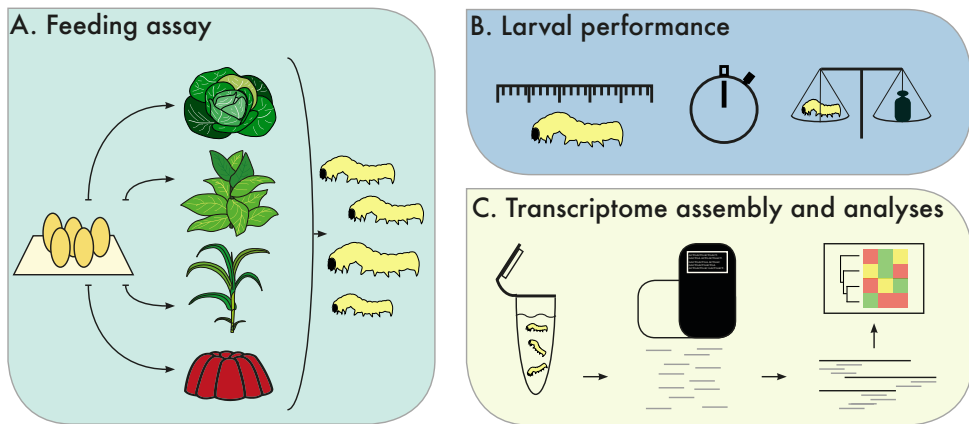
The ability of an herbivorous insect to feed on a broad host range largely depends on the flexibility and diversity of the insect's digestion and detoxification system. A recent comparative genomic study by Pearce et al. (Pearce et al., 2017) showed major expansions of gene families involved in detoxification and digestion including P450s, GSTs and CCEs when comparing two polyphagous moth species to two monophagous species. The evolution of polyphagy is hypothesised to be associated with expansions of gene families involved in host plant use, due to gene duplication and neofunctionalisation, (e.g. Cheng et al., 2017; Grbić et al., 2011; Pearce et al., 2017; Wen et al., 2006). Indeed, expansions of detoxification and digestion related gene families have been observed in multiple polyphagous arthropod species, such as the spider mite (*Tetranychus urticae*), known to feed on over 1,000 different host plant species (Ahn et al., 2014; Dermauw, Wybouw, et al., 2013; Grbić et al., 2011), the tobacco cutworm (*Spodoptera litura*) (Cheng et al., 2017), the fall armyworm (*Spodoptera frugiperda*) (Gouin et al., 2017) or the whitefly (*Bemisia tabaci*) (Chen et al., 2016).

Empirical evidence for the role of genes involved in the detoxifying ability of insects is mainly based on experimental studies using pesticide and isolated toxin treatments (e.g. Carino et al., 1994; Cheng et al., 2017; Feyereisen et al., 1989; Hu et al., 2019). More recently, plants are incorporated into molecular studies on the transcriptional response of insects, mimicking a more natural system. These studies have shown a differential genetic response of insects after host plant switches (Müller et al., 2017), or the transcriptional response of feeding on specific host plants (Koenig et al., 2015;

Orsucci et al., 2018; Roy et al., 2016; Schweizer et al., 2017). Yet, a multi-comparison of the genetic response of a polyphagous insect on multiple hosts from diverse plant families would provide insights in shared and unique gene activity linked to specific host plant usage.

In the present study we analysed the gene expression of the polyphagous beet armyworm, *Spodoptera exigua* (Hübner), feeding and developing on three of its recorded host plants and artificial diet (Figure 1A). This species is a member of the family Noctuidae and occurs worldwide except for cold regions (Pogue, 2002). *Spodoptera exigua* is a polyphagous herbivore being able to accept over 130 host plants representing more than 30 families (Merkx-Jacques et al., 2008; G. S. Robinson et al., 2010). Many of these host plants are considered economically important crops, making *S. exigua* an agricultural pest species (EPPO, 2019; Greenberg et al., 2002; Saeed et al., 2017). In our study we used three host plant species: cabbage (*Brassica oleracea*), maize (*Zea mays*) and tobacco (*Nicotiana tabacum*). They are members of three distantly related families: the crucifers (Brassicaceae), the grasses (Poaceae) and the nightshades (Solanaceae), respectively, and employ different defence mechanisms. The diverse plant families, as represented by the selected host species in the experimental comparison, are known for their specific defence compounds: *B. oleracea* produces glucosinolates, which are most dominant in the Brassicaceae family (Halkier & Gershenzon, 2006), *Z. mays* produces benzoxazinoids (Frey et al., 2009; Schuler, 2011) and *N. tabacum* produces various alkaloids including nicotine (Baldwin, 1988). We aimed to identify shared and/or unique gene expression patterns in relation to the different host plants. The identification of these expression patterns provides us information on the general genetic mechanism of herbivory, and, moreover, shows the diversity in transcriptional responses of insects while feeding on alternative host plants. Eventually, this is of importance in the understanding of the evolutionary and molecular mechanisms of herbivory as a whole. We analysed larval performance after feeding assays on the selected host plant species to quantify herbivore success and adaptability to the specific host defences (Figure 1B). Furthermore, we studied and compared the transcriptional response of *S. exigua* larvae feeding on the host plants using RNA-Seq (Figure 1C). We identified differential gene expression patterns as a result of feeding on different hosts.

Our study revealed unique clusters of genes with a diet-dependent expression in larvae feeding on hosts with sub-optimal herbivore success: *B. oleracea* and *N. tabacum*. A specific expression pattern containing uniquely expressed genes was not observed for larvae feeding on *Z. mays*, for which the feeding assay showed optimal herbivore success.

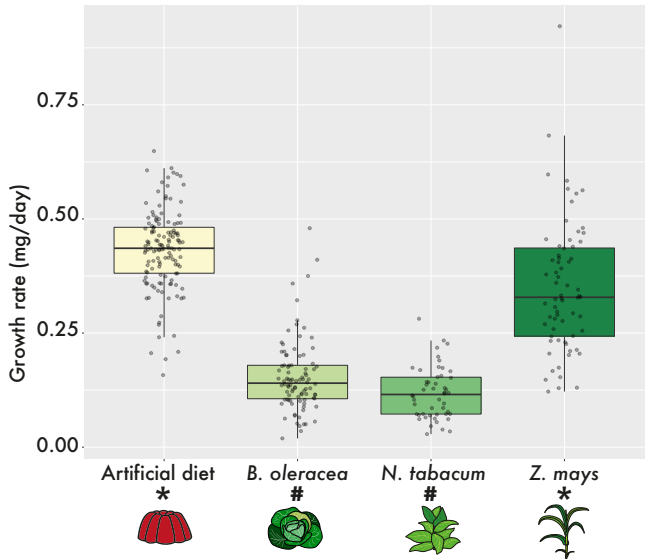


**FIGURE 1** | Overview of the experimental design to study the gene expression differences in *Spodoptera exigua* larvae feeding on different host plants. A. Feeding assay: larvae hatched and developed on three different host plants (*Brassica oleracea*, *Nicotiana tabacum* and *Zea mays*) and artificial diet as control. B. larval weight and developmental time was recorded to quantify herbivore success (growth rate in mg/day). C. RNA-Seq data was generated of each diet treatment, followed by *de novo* assembly and differential gene expression analyses.

## Results

### Feeding assays

Herbivore success of *S. exigua* larvae on the different host plant species was assessed in feeding assays comparing growth rates (weight/developmental time). The weight differences and the growth rates (mg/day) between the diet treatments were significantly different (ANOVA,  $p$ -value  $< 1e^{-16}$ ). Larvae on *Z. mays* proved to have the highest herbivore success (Figure 2;  $N = 68$ ; weight:  $2.103 \pm 0.56$  mg; growth rate:  $0.35 \pm 0.15$  (mean  $\pm$  SD) mg/day; developmental time: six days) both the weight and growth rate are significantly higher compared to all other host plant treatments as calculated with a post-hoc Dunn-test. The weight and growth rates of larvae developing on *B. oleracea* ( $N = 92$ ; weight:  $1.08 \pm 0.87$  mg; growth rate:  $0.154 \pm 0.08$  mg/day; developmental time: seven days) were not significantly different from larvae developing on *N. tabacum*, under the assay conditions as described in the methods part ( $N = 50$ ; weight:  $1.065 \pm 0.51$  mg; growth rate:  $0.118 \pm 0.06$  mg/day; developmental time: nine days). Larvae developing on the artificial diet ( $N = 134$ ; weight:  $2.162 \pm 0.44$  mg; growth rate:  $0.433 \pm 0.09$  mg/day; developmental time: five days) had significantly the highest growth rate as compared to all other diets but larval weight was not significantly different from larvae feeding on *Z. mays*. See Supplementary Table 1 for details of feeding assay results and Supplementary Table 2 for results of the feeding assay statistics (all supplementary material available under DOI: 10.1186/s12864-019-6081-7).



**FIGURE 2** | Graph showing the growth rates of *Spodoptera exigua* larvae collected in the third larval stage developing on different hosts, *Zea mays*, *Brassica oleracea*, *Nicotiana tabacum* and artificial diet. Growth rate (mg/day) was calculated by dividing the total larval weight (mg) by the number of days the larvae were allowed to feed (point of collecting: first observed larva reaching third larval phase). Asterisks (\*) indicate significant different treatments compared to all others. The hashtag (#) indicate treatments that are not significantly different from each other. The graph was created with the R function ggplot2 (Wickham, 2016) with spacing between datapoints.

## Transcriptome assembly

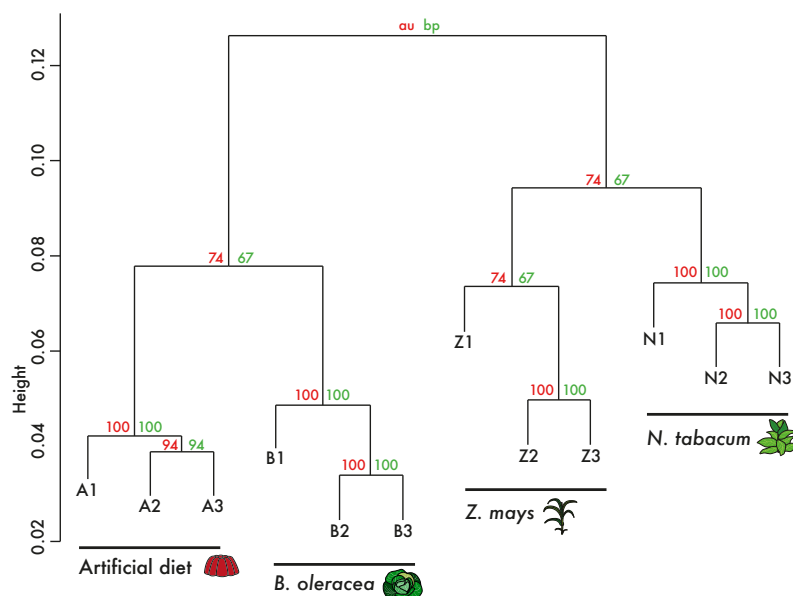
For each of the four treatments three replicates consisting of five larvae each were created. The 12 RNA libraries were sequenced on an Illumina HiSeq platform and the raw reads were assembled using Trinity v.2.5.1. (Grabherr et al., 2011). See Supplementary Data 1 for a description of the transcriptome assembly statistics and Supplementary Table 3 for a total overview of the number of raw reads per library and the final number of reads after trimming, cleaning and contamination checks. The assembly has been submitted to NCBI TSA database (GGRZ000000000) and has been used as a reference for the transcript expression quantification.

A completeness analysis of the final assembly using BUSCO v.3.0.2 against the *Insecta* gene set (Simão et al., 2015) indicated a high completeness level of expressed genes of 97.3%, with 1,088 complete single-copy BUSCOs, 526 complete duplicated BUSCOs, 21 fragmented BUSCOs and 23 missing BUSCOs of the in total 1,658 total BUSCO groups that were searched.



## Transcript expression quantification

The R package PVClust v.2.0 (Suzuki & Shimodaira, 2006) was used to analyse the sample relationships based on the filtered and normalised count matrix. The 10,000 bootstrap replicate based hierarchical clustering showed that replicates of the different diet treatment were more similar to each other than to samples of other treatments (Figure 3), proving the differentiation of the samples according to diet. Sample A2 within the *Z. mays* diet treatments did show increased variation with the other samples compared to all other treatments but was forming a cluster with the other maize replicates based on the PVClust analysis. The expression pattern of larvae feeding on *Z. mays* was most closely resembling the *N. tabacum* diet treatments. Furthermore, *B. oleracea* diet treatments clustered together with larvae of the control group feeding on the artificial diet, while *Z. mays* diet treatments clustered together with *N. tabacum* diet treatments.



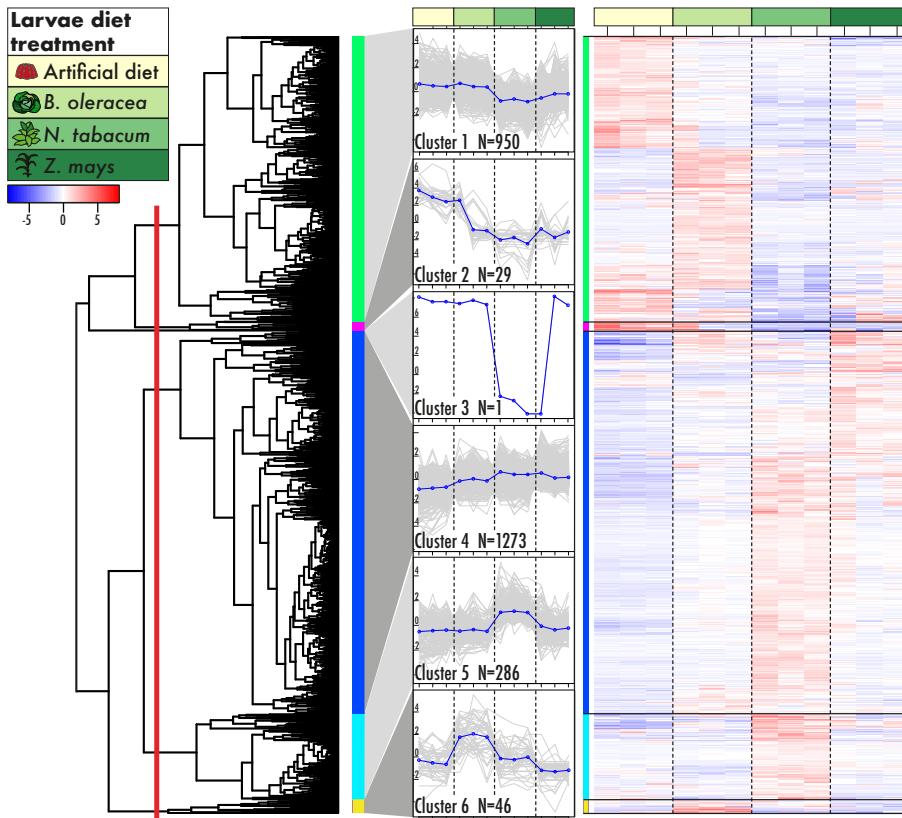
**FIGURE 3** | Hierarchical cluster dendrogram using multiscale bootstrap resampling of *Spodoptera exigua* larvae samples with different diet treatments. The filtered (gene count of  $\geq 10$  and occurrence  $\geq 2$  samples) and normalised (CPM + TMM normalisation) gene count matrix was used as input, including 58,749 genes. The number of bootstrap replications was 10,000. Given are Bootstrap Probability (BP) and the Approximately Unbiased (AU) values based on complete hierarchical clustering using the correlation distance measure.

We have used the trinity pipeline to perform a differential gene expression analysis with DESeq2 v.1.18.1 (Love et al., 2014) to identify transcripts differentially expressed (DE) across samples. Transcripts were considered DE with a fold change of  $2^2$  and  $p\text{-value} \leq 1e^{-3}$ . The transcripts were clustered based on the expression patterns. A cut-off of 50% for the hierarchical-clustering dendrogram was used. This resulted in the 2,585 DE genes being clustered into six groups (graphical representation and visualisation of the expression patterns of the DE genes and differences between the diet treatments, Figure 4). In two clusters a treatment-dependent gene expression was observed. Cluster 5 consisted of 286 transcripts that were upregulated in the *N. tabacum* treatment. Cluster 6 consisted of 46 transcripts upregulated in larvae feeding on *B. oleracea*. Transcripts in cluster 2 and 3 showed upregulation in samples from multiple treatments. Upregulation of genes in multiple treatments and/or absence of a response in part of the samples within a treatment are interpreted as non-host plant species-specific gene expression effects. Cluster 2 consisted of 29 transcripts that showed upregulation in all samples from the control group and a single sample from the *B. oleracea* treatment. Cluster 3 consisted of a single transcript only, downregulated in *N. tabacum* samples and a *Z. mays* sample. Cluster 1 and 4 contained the highest number of transcripts. Cluster 1 consisted of 950 transcripts that showed upregulation in larvae feeding on the artificial diet and *B. oleracea*. Cluster 4 consisted of 1,273 transcripts with upregulation in all host plant diet treatments (Figure 4).

## Transcript annotation

The reference transcriptome assembly was annotated using the Trinotate pipeline v.3.0 (Haas et al., 2013) (See Supplementary Table 4 for the Trinotate annotation report). Additionally, all DE transcripts were used as a query in a local BLASTX search against a local database containing all Arthropoda protein sequences as retrieved from the NCBI protein database (see Supplementary Table 5 for the BLASTX annotated DE transcripts using the local Arthropoda database).

Of the 2,586 DE genes 940 retrieved a GO annotation. Both cluster 3 (transcript downregulated in *N. tabacum* samples and a *Z. mays* sample) and cluster 6 (transcripts upregulated in larvae feeding on *B. oleracea*) had no significant GO terms overrepresented. The GO terms that showed statistical enrichment in the remaining clusters were compiled into broader generic GO slim categories for the ontology categories of biological processes (BP) and molecular functions (MF). Since the majority of the MF terms were summarised into the GO slim ‘molecular function’ (GO:0003674), GO terms were additionally compiled into the child terms of the GO slim ‘molecular function’, to increase specificity. GO terms of cluster 1 (transcripts upregulated in larvae feeding on the artificial diet and *B. oleracea*) and cluster 4 (transcripts upregulated in all host plant diet treatments) were summarised into both BP and MF category GO slims while for cluster 2 (transcripts upregulated in control samples and a *B. oleracea* sample) and cluster 5 (transcripts upregulated in the *N. tabacum* treatment) GO slims were only retrieved from the MF category. See Supplementary Table 6 for an overview of all GO slims per cluster.



**FIGURE 4** | Hierarchical clustering dendrogram of all differentially expressed genes (DE) (left) in *Spodoptera exigua* larvae with different diet treatments clusters the 2,585 DE in 6 clusters using a cut-off of at 50% (red line). Expression patterns were visualised and the number of genes (N=#) per cluster is indicated (middle). Green colour coding indicates host diet treatment, larvae fed with *B. oleracea*, *N. tabacum* and *Z. mays* are shown with increasing darker shades of green. The heatmap (right) shows the expression pattern of the DE genes, black lines indicate the cluster division. Each column corresponds to larvae from different diet treatments as indicated by the colour scheme.

Major represented GO terms according to the BP GO slim in cluster 1 (transcripts upregulated in larvae feeding on the artificial diet and *B. oleracea*) were ‘immune system process’ (GO:0002376), ‘response to stress’ (GO:0006950), ‘signal transduction’ (GO:0007165), ‘reproduction’ (GO:0000003) and ‘catabolic process’ (GO:0009056). Major MF GO slim were ‘peptidase activity’ (GO:0008233), ‘oxidoreductase activity’ (GO:0016491) and ‘molecular function’ (GO:0003674) with major significant child terms of the GO slim ‘molecular function’: ‘catalytic activity’ (GO:0003824), ‘binding’ (GO:0005488) and ‘structural molecule activity’ (GO:0005198).

In cluster 4 (transcripts upregulated in all host plant diet treatments) major GO slim terms were ‘circulatory system process’ (GO:0003013), ‘reproduction’ (GO:0000003) and ‘nervous system process’ (GO:0050877) for the BP category. Major MF GO slim terms were ‘transmembrane transporter activity’ (GO:0022857) and ‘molecular function’ (GO:0003674) with major significant child terms of the GO slim ‘molecular function’: ‘transporter activity’ (GO:0005215), ‘catalytic activity’ (GO:0003824), ‘binding’ (GO:0005488) and ‘molecular transducer activity’ (GO:0060089).

For cluster 5 (transcripts upregulated in the *N. tabacum* treatment) MF GO slim terms of this cluster were ‘oxidoreductase activity’ (GO:0016491), ‘transferase activity, (alkyl/aryl groups)’ (GO:0016765), ‘ion binding’ (GO:0043167) and ‘molecular function’ (GO:0003674) of which the significant child terms were grouped in the GO slim ‘catalytic activity’ (GO:0003824) and ‘binding’ (GO:0005488). For cluster 6 (transcripts upregulated in larvae feeding on *B. oleracea*) no significant overrepresented GO terms were present.

## Detoxification genes

The reference transcriptome was further analysed for putative detoxification genes of five major selected gene families: cytochrome P450s (P450), carboxyl/cholinesterases (CCE), glutathione S-transferases (GST), UDP-glycosyltransferases (UGT) and ATP-binding cassette transporters (ABC transporters). In total, 210,101 transcripts were searched of which 3,772 transcripts were annotated as putative detoxification genes from one of the five major selected gene families. In total 208 putative detoxification genes were DE (fold change of  $2^2$ ,  $p\text{-value} \leq 1e^{-3}$ ). The majority of transcripts in the set of DE transcripts were members of the ABC transporter enzyme family (75 transcripts). The four other gene families were less often identified: P450 (51 transcripts), GST (26 transcripts), CCE (37 transcripts), and UGT (19 transcripts). Each cluster, except for cluster 3, contained multiple annotated transcripts identified as genes from one or multiple detoxification families. In cluster 1 (transcripts upregulated in larvae feeding on the artificial diet and *B. oleracea*; 66 annotated transcripts), cluster 4 (transcripts upregulated in all host plant diet treatments; 91 identified transcripts) and cluster 5 (transcripts upregulated in the *N. tabacum* treatment; 41 annotated transcripts) all five detoxification families were present. Cluster 2 (transcripts upregulated in control samples and a *B. oleracea* sample; 4 annotated transcripts in total) only contained ABC genes. In cluster 6 (transcripts upregulated in larvae feeding on *B. oleracea*) in total six transcripts were identified from the ABC, CCE and GST families. See Supplementary Table 7 for the BLASTX annotated DE genes using the local detoxification gene database and Table 1 for an overview of transcripts identified as detoxification genes per gene expression cluster.

**TABLE 1** | Transcripts annotated as detoxification gene family member per gene expression cluster.

Cluster	Gene family	Group/Clade/Clan	#
<b>Cluster 1</b>	P450	clan-3	4
		clan-4	2
	UGT	UGT33	7
		UGT34	1
		UGT39	1
	GST	Sigma	3
	CCE		10
	ABC	ABCA	24
		ABCB	2
		ABCC	2
		ABCD	9
		ABCG	1
<b>Cluster 2</b>	ABC	ABCA	4
<b>Cluster 3</b>	-		
<b>Cluster 4</b>	P450	clan-2	1
		clan-3	21
		clan-4	11
		clan-M	1
	UGT	UGT33	4
		UGT39	1
		UGT40	3
	GST	Sigma	1
		Delta	1
		Omega	2
		Other/ Unknown	7
	CCE		15
	ABC	ABCA	6
		ABCD	11
		ABCG	7
<b>Cluster 5</b>	P450	clan-2	1
		clan-3	8
		clan-4	2
	UGT	UGT40	2
	GST	Sigma	3
		Other/ Unknown	9
	CCE		8
	ABC	ABCB	2
		ABCC	1
		ABCD	2
		ABCF	1
		ABCG	2
<b>Cluster 6</b>	GST	Other/ Unknown	1
	CCE		4
	ABC	ABCG	1

## Discussion

We compared the gene expression patterns of larvae of the polyphagous beet armyworm, *Spodoptera exigua* (Hübner) feeding and developing on different host plant species. We aimed to identify shared and/or unique gene expression patterns to study the general genetic mechanisms of herbivory and observe transcriptional response due to host plant differences. All treatment specific results are due to experimental setup design host plant specific and are thus diet-dependent.

Given the highest growth rates in larvae feeding on *Z. mays* compared to the other host plant species indicate that *S. exigua* larvae from this population were better able to utilise the nutrients available in *Z. mays* and/or cope with the specific defences. Diet treatment-dependent gene clusters with uniquely upregulated genes, including cluster-specific detoxification and digestion related genes, were only found in larvae feeding on *B. oleracea* or *N. tabacum*. This observation shows that *S. exigua*, considered as a major polyphagous insect, relies on the deployment of specific genes against particular host plant species. Specimens of *S. exigua* as used in this study originated from a lab culture recurrently refreshed with individuals from other populations. Different populations of generalist insects often demonstrate certain degrees of host specialisation on a local level (Schoonhoven et al., 2005). By using a lab population, both the background and feeding history is known which is important for studying the gene plasticity of feeding on different host plant species. Previous studies have shown that lab populations without regular renewal of individuals do adapt to their specific rearing conditions (Li et al., 2014). The strain of *S. exigua* was kept on an artificial diet, not resembling the live plant material the larvae are exposed to during the feeding assays. The artificial diet does contain plant-based products derived from seeds but is heavily produced and does not contain (whole or detached) plant parts such as leaves. Here we identify the transcriptional responses of polyphagy within this population of *S. exigua*. Comparing wild local populations is an interesting next step for understanding within-species differences.

The selected host plants are distantly related and produce a variety of different secondary metabolites for defence: *Z. mays*: benzoxazinoids, *B. oleracea*: glucosinolates and *N. tabacum*: various alkaloids like nicotine (Bennett & Wallsgrove, 1994). *Spodoptera exigua* is considered a major polyphagous species, being able to feed and develop on a large variety of host plants (G. S. Robinson et al., 2010). However, the feeding assay indicates a different degree of host plant usage and response to the different plant defences. As expected, larvae feeding on artificial diet showed an optimal development with the highest weight gain per day in comparison to larvae feeding on any of the host plants. The artificial diet is optimal in nutrient content and lack of any herbivory defences, like secondary metabolites. In addition, previous generations have been reared on the artificial diet in the stock rearing (see Methods section).

For the *N. tabacum* feeding assay a slightly adjusted experimental design has been used with an initial feeding period on detached leaves preceding a transfer to full plants (see for details Methods section). This adjusted treatment was used to lower larval mortality in the first larval phase on *N. tabacum*. These larvae have not been exposed to induced plant defences in response to herbivory as larvae feeding on the other plants had. Still, larvae were exposed to the defences present in the host plant during collecting. Plants alter physical and chemical defensive mechanisms upon herbivory (Kant et al., 2015; Tzin et al., 2017; Zhou et al., 2015). The increased defences of *N. tabacum* might have caused the high mortality during the critical first larval stage (Saeed et al., 2017). In initial trials all larvae feeding on full plants of *N. tabacum* died within 24 hours after hatching. Providing larvae detached leaves during the first larval stage prevented the possibility of the *N. tabacum* plant to increase defence compounds. Yet, transferring the larvae back on the full plant in the second larval stage is important to restore induced interactions (Karban & Myers, 1989). Providing first instar larvae detached leaves proved necessary due to high mortality. The interplay between induced defences of living plants in response of feeding insects shows the importance of using full living plants in studies analysing plant-insect interactions.

## Analysis of DE transcripts

The differential gene analysis showed that gene expression in larvae of *S. exigua* is dependent on the host plant. Two out of the six clusters showed upregulation of transcripts for only a single host-plant; cluster 5 for *N. tabacum* and cluster 6 for *B. oleracea*. Interestingly, such a host-specific gene expression activity pattern is missing for the other two diet types where *S. exigua* larvae showed optimal levels of herbivory: artificial diet and the host plant *Z. mays*.

Transcripts in cluster 1 showed upregulation in larvae with the artificial diet and *B. oleracea* diet treatments. Most GO terms were summarised into the GO slim ‘immune system process’ (GO:0002376) which includes genes involved in the development and role in the immune system. Various transcripts are identified as antimicrobial peptides including attacin (three transcripts), gloverin (one transcript) and cecropin (three transcripts) involved in acting against microbial intrusion in the hemolymph by targeting the microbial membrane (Jo et al., 2018; Lee & Lee, 2015). Another group of proteins involved in the immune system response and identified within cluster 1 are various REPAT proteins. Although cluster 2 and 6 both contained unique REPAT genes as well (cluster 2: REPAT 45, cluster 6: REPAT 10 and 12), cluster 1 contained 20 transcripts identified as REPAT from 16 different REPAT types. REPAT genes showed increased expression in response to pathogens inside the midgut (Herrero et al., 2007). However, it is speculated that besides this response to pathogens part of the REPAT proteins are involved in different physiological processes of the midgut (Navarro-Cerrillo et al., 2013).

The expression profile of cluster 4 included transcripts with upregulation in all plant diet treatments as compared to the artificial diet. Among the transcripts annotated in this cluster were 39 (BLASTX identification; 41 with Trinotate) encoding for cuticle-related proteins. An upregulation of genes involved in cuticle formation can be an indication for diet response. The insect cuticle and peritrophic matrix of the midgut serve in the correct functioning and movement of these organs (Kumar et al., 2018). Being in direct contact with abrasive food particles, the peritrophic matrix forms a biochemical and physical barrier and is involved in toxin inactivation (Hegedus et al., 2009). Alteration of the cuticle protein composition can lead to a more resilient and protective cuticle (Agrawal et al., 2014; Kelkenberg et al., 2015). Similarly to cluster 4, multiple transcripts in cluster 1 (nine according to BLASTX annotation; 10 according to Trinotate annotation) are identified as being involved in cuticle development. This cluster shows upregulation of transcripts in larvae fed with artificial diet and *B. oleracea*. Indeed, both clusters 1 and 4 are characterised by GO terms linked to the development of the cuticle, e.g. 'anatomical structure development'. Yet, the higher number of cuticle-related genes in cluster 4, upregulated in larvae fed on host plants, indicates that this mechanism might play a general role in coping with the various plant defence toxins. The overrepresentation of chitin related genes is not unique for *S. exigua* and has been found in other Noctuidae species with different host plants (e.g. Koenig et al., 2015; Schweizer et al., 2017).

We further focussed on the gene activity of five major detoxification gene families. This is of special interest for gene clusters showing upregulation in larvae feeding on a single host plant like clusters 5 and 6. Genes uniquely expressed in a single diet treatment are indicative for a potential host-specific response, e.g. detoxification genes acting specifically on alkaloids, the defensive compounds of *N. tabacum*. Members of all five detoxification families were present in cluster 5, with transcripts upregulated in larvae feeding on *N. tabacum*. Cytochrome P450 monooxygenases are involved in metabolising toxins in the first phase of the detoxification pathway. In cluster 5, 11 P450 genes were identified of which 10 were found uniquely expressed in this cluster. The majority of the genes are clade-3 P450s, well known for their role in the detoxification of plant defence secondary metabolites (e.g. Chandra et al., 2016; Wen et al., 2006; Zhu & Liu, 2008). The specific expression of the identified CYP6AE and CYP321 members within larvae feeding on *N. tabacum* might indicate a role in alkaloid detoxification. This confirms recent findings for the role of these genes in xenobiotic detoxification within the cotton bollworm (*Helicoverpa armigera*) and the tobacco cutworm (*Spodoptera litura*) (Shi et al., 2018; Wang et al., 2017a; Wang et al., 2017b). Of the eight transcripts identified as carboxyl/cholinesterases that also act in the first detoxification phase, two genes were uniquely expressed in this cluster. CCEs have been recognised as an important component of xenobiotic detoxification in insects (Oakeshott et al., 2005). Expansions of the gene families in polyphagous Noctuidae suggest a link with increased detoxification ability (Cheng et al., 2017; Pearce et al., 2017).



Both UDP-glycosyltransferases and glutathione-S-transferases are active in the second phase of the detoxification process. Of both families gene members have been identified in cluster 5. The two transcripts annotated as UGT proteins were both identified as UGT40U2. Members of the UGT40 family have been found particularly upregulated in digestion and detoxification-related body parts of *Helicoverpa* species (Pearce et al., 2017). Tissue-specific expression profiles of UGT genes in *S. exigua* showed high expression of UGT40U2 in the fat body of larvae (Hu et al., 2019). However, this gene is also present in cluster 4 showing increased expression in all host treatments suggesting a broad gene activity.

Moreover, we identified 12 transcripts in cluster 5 as GST genes; one of these was uniquely expressed in larvae feeding on *N. tabacum*. This GST gene is a member of the microsomal subfamily that is thought to be involved in oxidative stress response (Castello et al., 2007; Nunes et al., 2018).

ABC transporters are active in the third phase of detoxification. ABCs are involved in the transport of different compounds, many members show increased activity upon xenobiotic treatment (Cheng et al., 2017; Dermauw & Van Leeuwen, 2014; and references therein). ABCs involved in xenobiotic detoxification processes have been identified to belong to the ABC-B, ABC-C and ABC-G subfamilies (Labbé et al., 2011; Liu et al., 2011). Eight transcripts were identified as ABC genes within cluster 5. Only two ABC genes were uniquely expressed in the *N. tabacum* diet treatment samples: ABC-C2 and ABC-F4. ABC-C2 has been identified as the receptor of several specific Bt (Cry) toxins, a widely used biopesticide, in many Lepidoptera taxa including *S. exigua* (Endo et al., 2017; Pinos et al., 2019; Ren et al., 2016). ABC-C2 is only upregulated in larvae feeding on *N. tabacum* absent of Cry proteins, suggesting that ABC-C2 only has a role in alkaloid resistance.

All transcripts grouped in cluster 6 showed upregulation in larvae feeding on *B. oleracea*. Genes of only three out of five detoxification families were identified: CCE, GST and ABC transporters. *B. oleracea* uses glucosinolates as defence, in contrast to the other hosts. The detoxification genes grouped within this cluster might act upon glucosinolate presence (Jeschke et al., 2016). The detoxification genes upregulated in larvae feeding on a specific host, *N. tabacum* in cluster 5 and *B. oleracea* in cluster 6, are indicative for a host-specific role in detoxification. Detoxification genes identified in clusters with stable expression across hosts suggests a general physiological role or broad toxin acceptance.

In addition to detoxification-related genes in cluster 6, transcripts involved in the general digestion of plant material have been annotated. Three transcripts are annotated as trypsins, which together with chymotrypsins are involved in the digestion of proteins. The trypsins in cluster 6, according to the BLASTX annotation are of the alkaline c-type. Furthermore, alkaline c-type trypsins have been identified in the clusters 1, 3 and 4. Trypsins play an important role in insect survival. Various *Nicotiana* species successfully use trypsin inhibitors as defence against insect herbivores (Birk, 2003). But insects evolved adaptive responses (Zhu-Salzman & Zeng, 2015). In all

clusters trypsins and/or chymotrypsins have been identified but they are absent in cluster 5, showing upregulation of genes in larvae feeding on *N. tabacum*.

In summary, the identified gene expression clusters show the transcriptional response of *S. exigua* larvae feeding on different host plants that utilise species-specific defence mechanisms. The results of the feeding assays, quantifying the herbivore success for each host, showed that larvae of the beet armyworm had the highest growth rate when feeding on *Z. mays* compared to *B. oleracea* and *N. tabacum* fed larvae. This indicates that the used population of *S. exigua* is better able to cope with the *Z. mays*-specific plant defences and conversion of the material into biomass in comparison to the other host plant species. We found six clusters of transcripts differing in their expression pattern. Clusters with genes upregulated in a single diet treatment were observed for larvae feeding on either *B. oleracea* or *N. tabacum*, the hosts for which larvae showed low herbivore success. Among the genes uniquely upregulated in either *B. oleracea* or *N. tabacum* diet treatments were multiple members from major gene families putatively involved in plant toxin detoxification: P450 monooxygenases, carboxyl/cholinesterases, UDP-glycosyltransferases, glutathione-S-transferases and ABC-transporters. No specific diet-dependent gene clusters were identified for larvae feeding on the diet treatments with optimal herbivore success, *Z. mays* and the artificial diet. This indicates that for *S. exigua* larvae feeding on host plant species under optimal growth rates a general gene activity is used in the digestion and detoxification of the plant material. On the contrary, *S. exigua* larvae feeding on host plants with low herbivore success deploy particular host plant-specific (detoxification) genes for digestion and survival.

## Conclusions

Our work shows how a major polyphagous insect relies on the deployment of various host plant-specific genes. The importance of diverse and flexible gene expression involved in digestion and detoxification for a major polyphagous herbivore is evident from this study. We conclude that a generalist herbivore like *S. exigua* deploys a flexible set of genes in order to be able to survive on a wide array of host plant species. This work is based on the gene activity of a single species. Clearly, comparative transcriptomic studies of related polyphagous insect species using a similar set-up are needed to verify this finding and elucidate the gene activity of major polyphagous herbivore insects feeding on different host plants. Moreover, the increasing amount of genomic data makes large-scale comparative work possible to study and understand the molecular and evolutionary mechanisms of insect herbivory.

## Materials and methods

### Host plants and *Spodoptera exigua* rearing

*Spodoptera exigua* specimens originated from a stock rearing of the Laboratory of Virology, Wageningen University & Research, kept on artificial diet at 27°C with 50% relative humidity and a 14:10 h light:dark photoperiod. This population is recurrently renewed with specimens from other populations. The artificial diet consisted of water, corn flower, agar, yeast, wheat germ, sorbic acid, methylparaben, ascorbic acid and streptomycin sulphate. In the stock rearing, adult moths were kept in cylindrical containers, lined with paper sheets for egg deposition.

For the host plant experiments three different plant species were used in this study: *Zea mays* L. (accession B73, PI550473 –lot 94ncai02; propagated by selfing at the University of Amsterdam, seeds provided by Dr M.E. Stam), *Brassica oleracea* var. *gemmifera* L. cultivar Cyrus (provided by Unifarm Wageningen University & Research, seeds from Syngenta Seeds, The Netherlands), and *Nicotiana tabacum* L. (accession TC325, PI552514; provided by Dr J.M. Nifong, US Nicotiana Germplasm Collection). Plants were sowed and grown under optimal species-specific conditions at the Unifarm Wageningen University & Research greenhouse facilities until use in feeding assays.

### Feeding assays

To quantify the success of herbivory, larval performance was recorded in feeding assays. Feeding assays consisted of larvae developing and feeding on a single individual host plant, which was placed in a separate breeding cage, or on artificial diet in a plastic container (control group). The feeding assays took place in a temperature controlled greenhouse compartment at Wageningen University & Research (18°C:20°C, controlled lowest minimum night:day temperature; and 16:8 light:dark photoperiod). Larvae never had a lack of food nor were they reared under high population densities to avoid aggression and cannibalism (Orrock et al., 2017).

Multiple egg clutches were selected counting up to a total of  $\pm 300$  eggs using a dissection microscope and eggs were placed on top of leaves of whole plants inside the cages, or in plastic containers with artificial diet (control). The plants used for feeding assays were 5-9 weeks old. Hatched larvae were allowed to feed on the host plant or artificial diet until reaching the third larval stage. As soon as the first larva reached the third larval stage all larvae were collected on ice, frozen and weighed on a Sartorius MSE3.6P-000-DM Cubis Micro Balance (Sartorius, Göttingen, Germany). The collection point was chosen to have a similar time point for all treatments, based on development. The growth rate per day up to larval collection (mg/day) was calculated for each individual larva by dividing the larval weights by the recorded developmental time (days passed from hatching until larvae collecting).

Due to 100% mortality of larvae during the first larval stage after hatching on whole living *N. tabacum* plants, we followed a slightly modified experimental design for this host plant: eggs were placed on detached leaves and the larvae were allowed to develop on detached leaves under controlled conditions (27°C, 50% humidity, 14:10 h light:dark photoperiod) until reaching the late-second larval stage. This approach ensured larvae were facing the specific plant material at all times but without increased levels of induced defences at the more sensitive early stage. Afterwards larvae were relocated to the whole plant set-up, restoring the natural plant-herbivore interactions. Larvae then moulted to the third larval stage and continued feeding; solely larvae with proven feeding (close-by feeding damage visible on the plant) were collected after 48 hours.

The weight and growth rates of all four different diet treatments were processed and checked for significance using a non-parametric one-way ANOVA test followed by a Dunn-test for pairwise comparison using the Bonferroni method for p-value adjustment in R v.3.4.3.

## Larvae rearing for RNA-Seq

Larvae used for the RNA sequencing were treated exactly similar as larvae in the feeding assays. Larvae were collected on the first day of reaching the third larval stage, directly ground and frozen in RNeasy lysis reagent (QIAGEN, Hilden, Germany) and kept at -80°C until RNA isolation. Only larvae in the third larval stage were included. Morphological characteristics were used to confirm the correct larval stage: at end of the L2 stage a black stripe is visible, head of freshly moulted larvae is as wide as its body and the spiracular line becomes visible. For each treatment three biological replicates of five larvae each were created, based on initial survival tests.

## RNA sequencing

All samples for RNA-Seq were sent to Novogene (Novogene Co., Ltd.) for RNA extraction, cDNA library preparation and sequencing. Total RNA was extracted from each individual sample using TRIzol Reagent following the manufacturer's instructions (Invitrogen Co. Ltd, San Diego, USA). RNA degradation was checked for all samples on 1% agarose gel electrophoresis. The quality and integrity of the RNA samples was checked using the Nano 6000 Assay Kit of the Bioanalyzer 2100 (Agilent technologies, CA, USA) and the NanoPhotometer spectrophotometer (Implen, CA, USA). The RNA concentration was measured using the Qubit RNA Assay Kit on a Qubit 2.0 Fluorometer (Invitrogen Co. Ltd, San Diego, USA). RNA concentrations are given in Supplementary Table 8, indicating the quality and consistency of the RNA samples used.

The NEBNext Ultra RNA Library Prep Kit for Illumina sequencing (NEB, MA, USA) following the protocol as provided by the manufacturer was used for library preparation,

index sequences were added to trace back the sequences to each sample. In short, a total of 3 µg total RNA extract was used for each library preparation. The mRNA was purified using poly-T oligo-attached magnetic beads followed by fragmentation using NEBNext First Strand Synthesis Reaction Buffer (5x). Synthesis of first strand cDNA was performed using random hexamer primers and M-MuLV Reverse Transcriptase (RNase H<sup>-</sup>), subsequently second strand cDNA was synthesised using DNA polymerase I and RNase H. After terminal repair of the sequences, the 3' ends were adenylated and NEBnext adapters (NEB, MA, USA) were ligated. The library fragments were purified, for a 150-200 bp size range, with AMPure XP purification kit (Beckman Coulter, MA, USA). To the size-selected cDNA, 3 µl USER Enzyme (NEB, MA, USA) was added and heated to 37°C for 15 min followed by 5 min at 95°C. The cDNA was enriched by PCR reactions using Phusion High-Fidelity DNA polymerase, universal PCR primers and index primers. The PCR products were purified with the AMPure XP kit and the quality of the libraries was evaluated on the Agilent Bioanalyzer 2100 (Agilent technologies, CA, USA). Clustering of the index-coded samples was done on a cBot Cluster Generation System using a TruSeq PECluster Kit 3-cBot-HS (Illumina Inc., CA, USA). Finally, the prepared samples were sequenced on Illumina HiSeq 4000 platform with 150 bp paired-end (PE) reads (see Supplementary Table 3 for an overview of number of raw reads per treatment).

## Transcriptome assembly

The sequencing reads were quality checked using FastQC v.0.10.1 (Andrews, 2010). Adapter sequences were removed and quality-filtered using Trimmomatic v.0.36 (Bolger et al., 2014), with parameters set: ILLUMINACLIP: *file fq*:2:30:10, LEADING: 5, TRAILING: 5, SLIDINGWINDOW:4:20, and removing all reads of less than 36 bp in length.

Before transcriptome assembly the raw reads were checked for potential contamination and remaining adapter sequences using local BLAST (NCBI-BLAST+ v.2.6.0 (Camacho et al., 2009) was used for all local BLAST searches) against the UniVec database (<ftp.ncbi.nlm.nih.gov/pub/UniVec/>, accessed: 01 February 2018) using BLASTN with parameters: -reward 1 -penalty 5 -gapopen 3 -gapextend 3 -dust yes -soft\_masking true -evalue 700 -searchsp 1750000000000. Raw reads were submitted to the NCBI SRA database under accession number PRJNA477295. Reads were assembled *de novo* using Trinity v.2.5.1 (Grabherr et al., 2011) using PE mode and without *in silico* read normalisation.

The assembled transcripts were checked for contamination using DeconSeq v.0.4.3 (Schmieder & Edwards, 2011) with a coverage of 90% and an alignment identity threshold of 98% against a set of bacterial, human, *B. oleracea*, *Z. mays*, and *N. tabacum* reference sequence sets. Sequence reference sets were obtained from various sources: the bacterial set is based on 1,544 full representative 'RefSeq' bacterial genomes, from <https://www.ncbi.nlm.nih.gov/assembly>, accessed 24

May 2018; the Human Reference genome (GRCh38.p12) from [ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo\\_sapiens/Assembled\\_chromosomes/seq/](ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/Assembled_chromosomes/seq/), accessed 24 May 2018; and the three host plant sets were all obtained from the NCBI genome database (*B. oleracea*, from [ftp://ftp.ncbi.nih.gov/genomes/Brassica\\_oleracea/RNA/Gnomon\\_mRNA.fsa.gz](ftp://ftp.ncbi.nih.gov/genomes/Brassica_oleracea/RNA/Gnomon_mRNA.fsa.gz); *N. tabacum*, from [ftp://ftp.ncbi.nih.gov/genomes/Nicotiana\\_tabacum/RNA/Gnomon\\_mRNA.fsa.gz](ftp://ftp.ncbi.nih.gov/genomes/Nicotiana_tabacum/RNA/Gnomon_mRNA.fsa.gz); and *Z. mays*, from [ftp://ftp.ncbi.nih.gov/genomes/Zea\\_mays/RNA/Gnomon\\_mRNA.fsa.gz](ftp://ftp.ncbi.nih.gov/genomes/Zea_mays/RNA/Gnomon_mRNA.fsa.gz), all three accessed on 20 June 2018). Moreover, the assembly was cross-checked against various insect genomes: *Drosophila melanogaster* ([ftp://ftp.flybase.net/genomes/Drosophila\\_melanogaster/dmel\\_r6.21\\_FB2018\\_02/fasta/](ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmel_r6.21_FB2018_02/fasta/), accessed 05 June 2018), *B. mori* ([ftp://ftp.ensemblgenomes.org/pub/metazoa/release-39/fasta/bombyx\\_mori/dna](ftp://ftp.ensemblgenomes.org/pub/metazoa/release-39/fasta/bombyx_mori/dna), accessed 05 June 2018) and *S. litura* (ASM27686v1; [ftp://ftp.ncbi.nih.gov/genomes/Spodoptera\\_litura/Assembled\\_chromosomes/seq/](ftp://ftp.ncbi.nih.gov/genomes/Spodoptera_litura/Assembled_chromosomes/seq/), accessed 05 June 2018). Suspicious transcripts were subsequently removed. The final reference transcriptome was submitted to the NCBI TSA database and is available under accession GGRZ000000000, version GGRZ010000000 is used in this study.

Completeness of the final reference transcriptome assembly was accessed using BUSCO v.3.0.2 using the *Insecta\_odb9* set for comparison (Simão et al., 2015).

## Transcript expression quantification

To estimate transcript expression, reads of all samples from each treatment were separately mapped to the reference transcriptome using Bowtie2 v.2.3.4 (Langmead & Salzberg, 2012). The isoform and gene abundance estimation was done using RSEM v.1.3.0 (Li & Dewey, 2011), resulting in a per-treatment count matrix which was fed into the Trinity package perl script '*abundance\_estimates\_to\_matrix.pl*' to construct a raw (non-normalised) count matrix.

The raw count matrix was filtered by abundance based on count-per-million values (CPM; to account for library size differences between samples) using edgeR v.3.20.8 (M. D. Robinson et al., 2010) in R v.3.4.3. Only genes with a minimum of 10 counts in at least two of the samples were considered expressed and retained in the dataset. Moreover, the filtered count matrix was cross-sample normalised using the 'calcNormFactors' function in edgeR using trimmed mean of M values (TMM) (Robinson & Oshlack, 2010). See Supplementary Table 9 for the CPM, TMM cross-sample normalised and filtered count matrix, and Supplementary Table 10 for the subset of DE genes. Venny v.2.1 (Oliveros, 2007) was used to evaluate gene similarity across the diet treatments using the filtered and normalised gene matrix (Venn diagram is provided in Supplementary Figure 1, results are presented in Supplementary Data 1). To estimate sample and biological replicate relationships, the filtered and normalised count matrix was analysed using cluster bootstrapping with 10,000 replications with the R package PVClust v.2.0 (Suzuki & Shimodaira, 2006).

The differential expression analysis was performed using DESeq2 v.1.18.1 (Love et al., 2014) as implemented in the Trinity package under default settings; `min_rowSum_counts = 2` and using the median of ratios method for normalisation, taking into account sequencing depth and RNA composition. Transcripts were considered significantly differentially expressed with a minimal fold-change of four between any of the treatments and a false discovery rate (FDR) of  $p\text{-value} \leq 1e-3$ . The CPM and TMM normalised expression values of all differentially expressed transcripts were hierarchically clustered and cut at 50% using the Trinity provided script '`define_clusters_by_cutting_tree.pl`'. This resulted in six clusters of differentially expressed transcripts with similar expression patterns that were used in the cluster-specific GO analysis.

### Transcript annotation

The sequences, on isoform level, were annotated using the Trinotate pipeline v.3.0 (Haas et al., 2013). TransDecoder v.5.0.2 from the Trinity package was used to identify candidate coding regions within the assembled transcripts (Haas et al., 2013), identifying open reading frames (ORFs) with a minimal length of 100 amino acids by default. The TransDecoder-predicted protein sequences were used as a query for a BLASTP search using the manually annotated and non-redundant Swiss-Prot database ([ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/complete/uniprot\\_sprot.dat.gz](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.dat.gz); release 2018\_06, accessed June 2018). The transcripts from the assembly were used as a query for a BLASTX search using the same protein database. Moreover, a protein domain search was performed using '`hmmsearch`' from HMMER v.3.1b2 against the Pfam-A database v.31. Signal peptides were predicted using SignalP 4.1 server (Petersen et al., 2011) and TMHMM v.2.0 was used for transmembrane region prediction (Krogh et al., 2001). See Supplementary Table 4 for the Trinotate-pipeline derived annotations. Additionally, a local BLASTX search (settings: `max_hsps` of 1, `best_hit_overhang` set on 0.1, E-value cutoff  $\leq 1e-3$ ) was conducted using the assembled transcripts as a query against a database consisting of all local Arthropoda protein sequences downloaded from the NCBI protein database (accessed, February 2018).

Gene ontology (GO) analysis was performed using the GO annotations from the Trinotate pipeline. GO terms, including all ancestral terms were extracted from the annotation report. The R package Goseq v.1.30.0 was used for the analysis, which takes gene length bias into account (Young et al., 2010). Only statistically overrepresented GO terms were taken into account that had an adjusted  $p\text{-value} < 0.05$ , according to Benjamini-Hochberg (BH) correction for multiple hypothesis testing (Benjamini & Hochberg, 1995). For each of the six identified DE clusters the statistically overrepresented GO terms (BH corrected  $p\text{-value} < 0.05$ ) were calculated relative to the reference of all other GO annotated transcripts of the reference assembly. The retrieved significant GO terms were compiled into broader terms,

generic GOslim categories (subset used: goslim\_generic.obo, the Gene Ontology Consortium; accessed August 2018), using the R package GOstats v.2.44.0 (Falcon & Gentleman, 2007).

The reference transcriptome was screened for genes of five major gene families known to be involved in toxin detoxification: P450, CCE, GST, UGT and ABC transporters. Annotated genes from five detoxification gene families from Lepidoptera members of the Noctuidae, Plutellidae, Sphingidae and Bombycidae were combined in a local protein BLAST database. The database consisted of 2,369 identified protein sequences derived from the following eight species: the fall armyworm (*S. frugiperda*) (both corn and rice populations, OGS2.2; Gouin et al., 2017), tobacco cutworm (*S. litura*) (protein set 002706865.1 genome ASM270686v1, from NCBI Genome database; Cheng et al., 2017), the beet armyworm (*S. exigua*) (all P450 and UGT genes; Hu et al., 2019), the silkworm (*Bombyx mori*) (protein set 000151625.1 genome ASM15162v1, from NCBI Genome database; International Silkworm Genome Consortium, 2008), the diamondback moth (*Plutella xylostella*) (OGSv2; Jouraku et al., 2013), the tobacco hornworm (*Manduca sexta*) (OGS2; Koenig et al., 2015), and both the cotton bollworm (*Helicoverpa armigera*) and corn earworm (*Helicoverpa zea*) (Pearce et al., 2017). Transcripts from the reference assembly were run as query in a BLASTX run against the created detoxification protein database. For each group of isoforms the highest expressed transcript was selected as gene representative for the detoxification-gene annotation. Only BLAST hits with an E-value  $>1e^{-5}$  were selected as putative detoxification genes.

## Acknowledgements

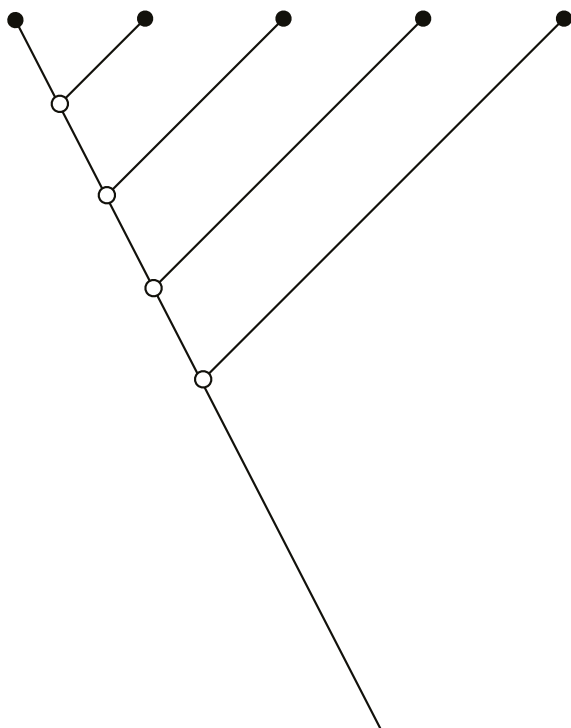
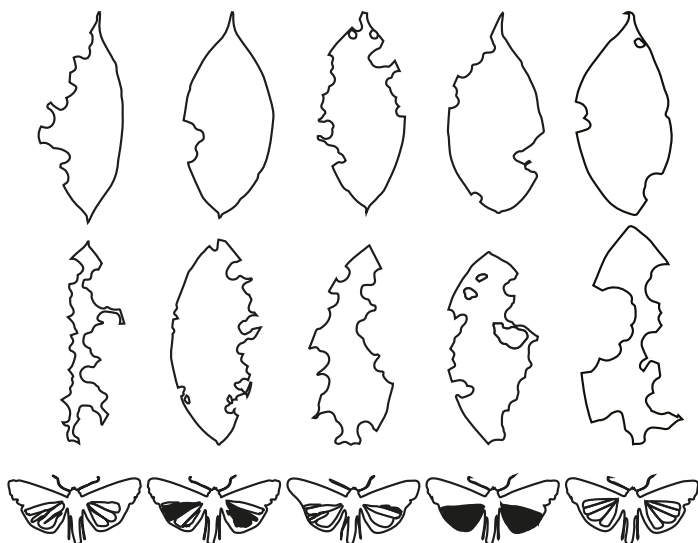
We want to thank Els Roode, Han Yue and in specific we gratefully acknowledge the late Hanke Bloksma (Virology, WUR) for advice and help with the *S. exigua* rearing and experiments. Further we are grateful to Dr M. Stam (UvA) for providing *Z. mays* B73 seeds and Dr J.M. Nifong (US Nicotiana Germplasm Collection) for providing *N. tabacum* seeds. Finally we want to thank the Unifarm Wageningen University & Research personnel for providing *B. oleracea* plants and overall plant care.



## Supplementary data

All supplementary data is available under DOI: 10.1186/s12864-019-6081-7.

- SUPPLEMENTARY TABLE 1** | Overview of the results of the feeding assays.
- SUPPLEMENTARY TABLE 2** | Overview feeding assay statistics.
- SUPPLEMENTARY TABLE 3** | Overview of the number of raw reads per library and the final number of reads after trimming, cleaning and contamination checks.
- SUPPLEMENTARY TABLE 4** | Trinotate annotation report.
- SUPPLEMENTARY TABLE 5** | BLASTX annotated DE genes using local Arthropoda database.
- SUPPLEMENTARY TABLE 6** | Overview of all GO slims per cluster.
- SUPPLEMENTARY TABLE 7** | BLASTX annotated DE genes using the local detoxification gene database.
- SUPPLEMENTARY TABLE 8** | RNA concentrations for all samples.
- SUPPLEMENTARY TABLE 9** | CPM, TMM cross-sample normalised and filtered count matrix.
- SUPPLEMENTARY TABLE 10** | CPM, TMM cross-sample normalised and filtered count matrix only including DE genes.
- SUPPLEMENTARY FIGURE 1** | Venn diagram indicating the number of expressed transcripts shared or unique in *Spodoptera exigua* larvae feeding on different diet treatments based on the filtered and normalised count matrix. Larvae developed on *Zea mays*, *Brassica oleracea*, *Nicotiana tabacum* or artificial diet until reaching the third larval stage. The total number of transcripts included was 58,749.
- SUPPLEMENTARY DATA 1** | Additional results of transcriptome assembly statistics and annotation.



# Chapter 3

Family dinner: transcriptional plasticity of five  
Noctuidae (Lepidoptera) feeding on three host  
plant species

Thijmen Breeschoten<sup>1,\*</sup>,  
M. Eric Schranz<sup>1</sup>,  
Erik H. Poelman<sup>2</sup>,  
Sabrina Simon<sup>1,\*</sup>

<sup>1</sup> Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>2</sup> Laboratory of Entomology, Wageningen University & Research, PO Box 16, 6700 AA Wageningen, The Netherlands



Polyphagous insects often show specialisation in feeding on different host plants in terms of survival and growth and thus be considered minor or major pests of particular hosts. Whether polyphagous insects employ a common transcriptional response to cope with defences from diverse host plants is scarcely studied. We focused on patterns of transcriptional plasticity in polyphagous moths (Noctuidae), of which many species are notorious pests, in relation to herbivore performance on different host plants. We compared the transcriptional plasticity of five polyphagous moth species feeding and developing on three different host plant species. Using a comparative phylogenetic framework, we evaluated if successful herbivory is determined by a shared or lineage-specific transcriptional response. Based on feeding assays, we identified varying degrees of herbivore success. The transcriptional activity of larvae feeding on the different host plants was highly plastic and moth species-specific. Specialisation, defined as high herbivore success for specific host plants, was not generally linked to a lower number of induced genes. Moths with high herbivore success for certain host plants showed shared expression of multiple homologous genes, indicating convergence. We further observed specific transcriptional responses within phylogenetic lineages. These expression patterns for specific host plant species are likely caused by shared evolutionary histories and could therefore not be associated with herbivore success alone. Multiple gene families were widely expressed in response to host plant feeding, showing their importance for herbivory, but their expression pattern was highly moth species-specific. Consequently, high herbivore success for specific host plants is also driven by species-specific transcriptional plasticity. Thus, potential pest moths display a complex and species-specific transcriptional plasticity.

#### Keywords

Polyphagy; generalists; detoxification genes; RNA-Seq; transcriptomics; phylogenetic framework

#### Abbreviations

Cytochrome P450 monooxygenase (P450), Carboxyl/Cholinesterase (CCE), Glutathione-S-transferase (GST), UDP-glycosyltransferase (UGT), ATP-binding cassette transporters (ABC), Glucosinolate sulfatase (GSS), Nitrile-specifier protein (NSP), Glucose-methanol-choline oxidoreductase (GMC), Differentially Expressed (DE), Count Per Million (CPM), Trimmed Mean of M values (TMM), Orthogroup (OG).

## Introduction

Herbivorous insects and their host plants have evolved various forms of specialisation, often forming the foundation of ecological and evolutionary theories (Funk et al., 2002; Kawecki, 1998). Host specialisation among herbivorous insects is common. Specialisation is often seen as a consequence of the coevolution of plant defences and herbivore resistances, resulting in both species and gene evolution (Edger et al., 2015; Ehrlich & Raven, 1964; Heidel-Fischer & Vogel, 2015). Plants defend themselves from herbivores by both physical barriers and by synthesising toxins (allelochemicals/specialised metabolites) (Després et al., 2007; War et al., 2018). Despite these defences many herbivorous insects are polyphagous and have maintained and / or evolved the ability to feed on a broad set of host plants, even including multiple plant families (Bernays & Graham, 1988; Schoonhoven et al., 2005).

A general molecular mechanism for herbivores to cope with plant chemical defences is through detoxification (Després et al., 2007). Various major gene families, well known to be involved in the three-phased detoxification pathway (Kant et al., 2015), are generally conserved across insects. In the first phase, gene families including cytochrome P450 monooxygenases (P450s), oxidoreductases and carboxyl- and cholinesterases (CCEs) metabolise the toxin (Després et al., 2007; Heckel, 2018; Li et al., 2007; Müller et al., 2017; Schuler, 2011; Zhang et al., 2013). In the second phase, they are conjugated by families including UDP-glycosyltransferases (UGTs) and glutathione-S-transferases (GSTs) before they, in the third phase, get transported away by transporters including ATP-binding cassettes (ABCs) (Bock, 2016; Francis et al., 2005; Heidel-Fischer & Vogel, 2015; Kant et al., 2015).

Detoxification is not restricted to these well-known gene families, various insect species, often specialising on narrow groups of plant species, have evolved specialised resistance methods differing from those described above, such as the nitrile-specifier protein (NSP) in Pierinae (Fischer et al., 2008; Wheat et al., 2007; Wittstock et al., 2004) and glucosinolate sulfatase (GSS) in the diamondback moth (*Plutella xylostella*) (Heidel-Fischer et al., 2019; Ratzka et al., 2002). The detoxification and metabolism of plant defence toxins is most likely due to concerted functioning of genes from these various shared and specific gene families and is expected to vary with different toxins.

A polyphagous herbivore requires plasticity in its transcriptional response to overcome the diversity in plant defences (Koenig et al., 2015; Orsucci et al., 2018). Transcriptional changes of genes involved in xenobiotic detoxification and digestion especially play a central role when facing different host plant species (Celorio-Mancera et al., 2012; Celorio-Mancera et al., 2013; Dermauw, Wybouw, et al., 2013; Rivard et al., 2004; Wybouw et al., 2015). For example, in the green peach aphid (*Myzus persicae*) transcriptional plasticity facilitates the shift to a new host plant by differential regulation of specific gene clusters (Mathers et al., 2017). Moreover, in aphids the co-regulated expression of effector genes were suggested to form a transcriptional mechanism enabling plant feeding (Thorpe et al., 2018). Transcriptional plasticity is

also shown in the beet armyworm (*Spodoptera exigua*), where specific upregulated expression clusters are observed in moths feeding on host plant species which are sub-optimal (Breeschoten et al., 2019). It has been found that Lepidoptera species feeding on a wide host plant range display broad plasticity in transcriptional responses (Celorio-Mancera et al., 2013; Pearce et al., 2017; Schweizer et al., 2017). However, if a general transcriptional plastic pattern in response to host plant feeding exists across polyphagous insects is largely unknown.

It is not well understood if successful herbivory of polyphagous insects feeding on different host plant species is determined by shared transcriptional responses, or rather involves lineage-specific gene expression patterns. This is of importance to understand and anticipate on pest formations by polyphagous insects. Within polyphagous insect species, populations can be restricted in their host plant choices in terms of successful feeding and development (Howard et al., 1994; Pashley, 1986; Schoonhoven et al., 2005), thus, a degree of specialisation occurs. This ‘host-specialisation’ could eventually lead to pest-formations in certain insect species (Gouin et al., 2017).

In order to elucidate a potential shared transcriptional pattern that differentiates pest formations from unsuccessful herbivory, multi-species comparisons need to be conducted. For the multi-species comparisons, the polyphagous target species should display variable levels of herbivore success. It is also of importance to account for the evolutionary relationships of the target species and consequently perform comparison in a phylogenetic framework. Data gathered from multiple species is not independent and will be affected by evolutionary relatedness. If evolutionary relatedness is not considered, the statistical assumption of independence is violated and shared traits, such as responsiveness or as presented here transcriptional expression, are wrongly interpreted (Dunn et al., 2013; Dunn et al., 2018). Further, the inclusion of multiple species is of importance to test for general patterns and to avoid the classical two-species comparison-pitfall because any two species differ (Ali & Agrawal, 2012). By increasing the number of species, patterns of behaviour, transcriptional response and evolutionary history become visible and evident.

In this study we implement a comparative phylogenetic framework in order to study the transcriptional mechanisms of (un)successful herbivory in polyphagous moth species. Our experimental system included five species, the cabbage looper (*Trichoplusia ni*), the silver-Y moth (*Autographa gamma*), the cabbage moth (*Mamestra brassicae*), the African cotton leafworm (*Spodoptera littoralis*) and the beet armyworm (*Spodoptera exigua*) (Breeschoten et al., 2019) while feeding and developing on three different recorded host plant species: cabbage (*Brassica oleracea*), maize (*Zea mays*) and tobacco (*Nicotiana tabacum*), and artificial diet. The selected host plant species are members of different families that are known to employ different defence mechanisms including specialised metabolites: benzoxazinoids in *Z. mays* (Frey et al., 2009), glucosinolates in *B. oleracea* (Halkier & Gershenzon, 2006) and various alkaloids such as nicotine in *N. tabacum* (Baldwin, 1988). All included moth species

belong to the cutworm moth family Noctuidae, which is a cosmopolitan species-rich family including ~1,089 genera and ~11,772 species (van Nieukerken et al., 2011). Many species in the family are major polyphagous species and form notorious pests, which has led to the description of a “pest-clade” (Mitchell et al., 2006). Especially in the genus *Spodoptera*, major polyphagous species occur which cause infestations in many parts of the world that destroy harvests and diminish crop yields (e.g. EPPO, 2019; Goergen et al., 2016; Kalleshwaraswamy et al., 2018).

In our comparative phylogenetic framework, we first analysed larval performance by feeding assays to quantify herbivore success on the different plant species. We then analysed and compared the transcriptional plasticity using RNA sequencing of feeding larvae in relation to their herbivore success. We focused on transcriptional response differences between the Noctuidae moths on different host-plant species which could be correlated to high herbivore success and thus pest potential.

## Materials and methods

### Noctuidae species selection, rearing and host plant selection

Species used in the experimental feeding assays were obtained from various stock rearings. The parent generation of all Noctuidae specimens, including *S. exigua* from a former study (Breeschoten et al., 2019), used in the feeding assays and RNA sequencing were raised on the same artificial diet at Unifarm, WUR greenhouse facilities. The subsequent generation was then used in the feeding assays and RNA sequencing as presented here.

The specimens used for the *Spodoptera littoralis* originated from Andermatt Biocontrol, Grossdietwil, Switzerland. Specimens of *Trichoplusia ni* originated from Great Lakes Forestry Centre, Ontario, Canada. Both *Autographa gamma* and *Mamestra brassicae* originated from the Laboratory of Entomology, Wageningen University & Research (WUR), The Netherlands. Rearing and further experiments were done in temperature and light controlled greenhouse compartments at WUR (*S. littoralis*: 26°C/16:8 h light:dark photoperiod/50% humidity; *A. gamma*, *M. brassicae* and *T. ni*: 18°C/20°C minimum night:day temperature/16:8 light:dark photoperiod). We additionally used the data from *Spodoptera exigua* presented in Breeschoten et al. (2019), for which specimens originated from the stock rearing of the Laboratory of Virology, Wageningen University & Research. Rearing was similar for all five species. Eggs were received and placed in plastic containers. Larvae were kept in these containers providing only artificial diet, consisting of water, cornflour, agar, yeast, wheat germ, sorbic acid, methylparaben, ascorbic acid and streptomycin sulphate. The food did contain seed-derived products but is heavily processed and is not resembling live plant parts the larvae are exposed to in the feeding assays. Larvae were transferred to vermiculite filled containers for pupating. Adult moths were kept

in cylindrical containers; eggs were deposited on paper sheets that were placed in the containers. For *M. brassicae*, adult moths were kept in cylindrical containers covered with paper sheets for horizontal egg deposition, while *T. ni* moths primarily deposited their eggs on gauze fabric placed inside the containers. Respective egg clutches were collected and finally used for the feeding assays (see below).

Three different plant species were used in the feeding assays: *Zea mays* L. (accession B73, PI550473 -lot 94ncai02; propagated by self-pollination at the University of Amsterdam, seeds provided by Dr M.E. Stam), *Brassica oleracea* var. *gemmifera* L. cultivar Cyrus (provided by Unifarm Wageningen University & Research, seeds from Syngenta Seeds, The Netherlands), and *Nicotiana tabacum* L. (accession TC325, PI552514; provided by Dr J.M. Nifong, US Nicotiana Germplasm Collection). The plant growth was similar as in Breeschoten et al. (2019). Plants were sown and grown under optimal species-specific conditions at the Unifarm Wageningen University & Research greenhouse facilities until use in the feeding assays.

### Larval performance and collecting for RNA-Seq

Larval performance, as an indicator for herbivore success, was quantified in feeding assays. Larvae developed and fed on a single host plant within separate breeding cages. The control group was kept on artificial diet in a plastic container. The plants used for the feeding assays were ~5-9 weeks old and still undergoing vegetative growth (e.g., not yet flowering). Multiple egg clutches, totalling ~300-400 eggs, were selected and placed on top of the leaves of the three host plants or inside the plastic container for the control group (artificial diet) setup. The larvae hatched and were allowed to develop and feed from the host plants or artificial diet until reaching the third larval stage.

A subgroup of the larvae reaching the third larval stage were used for RNA sequencing. Larvae used for RNA sequencing were collected on the first day of reaching the third larval stage, ground and frozen in RNAlater reagent (QIAGEN, Hilden, Germany) and kept at -80°C until RNA isolation. For each species, three biological replicates of five larvae each were sampled for each treatment.

Surviving larvae, not used for sequencing and to a maximum of about 100 individuals, were used for larval performance measurements. The larvae, at the timepoint first individuals reached third larval stage, were collected on ice, frozen and weighed within a single day on a Sartorius MSE3.6P-000-DM Cubis Micro Balance (Sartorius, Göttingen, Germany). Developmental time (period from hatching until reaching third larval stage, in days) was recorded to calculate the growth rate in mg per day (mg/day).

Due to previous observed high larvae mortality rate in the whole-plant set-up of *N. tabacum*, a slightly modified experimental design as outlined in detail in Breeschoten et al. (2019) was adopted for all specimens. In brief, eggs were placed on detached leaves of *N. tabacum* in a plastic container. After hatching, leaves were regularly



refreshed until larvae reached the late-second larval stage. Afterwards, larvae were transferred back to the whole-plant set-up, continuing their development and collected after 48 hours. Only larvae moulted to the third larval stage and with clear feeding marks (i.e. visible feeding damage) were collected.

### Feeding assay statistics

Weight and growth rates for all diet treatments and species were checked for significance using a Kruskal-Wallis test followed by a Dunn-test for pairwise comparisons using the Bonferroni method for *p*-value adjustment in R v.3.4.3 (R Development Core Team, 2020).

For comparison of growth rates on the various host plants, the growth rates of each species were normalised using the growth rates of the respective control group. Normalisation is needed in order to compare the species-specific growth rates. We assumed the artificial diet to be an optimal food source lacking any form of defence toxins. For normalisation, we calculated the average growth rate of larvae feeding on the artificial diet (= ‘neutral’ growth rate). Second, for every host plant, the growth rates were calculated by using the deviation in percentage from the respective species-specific average ‘neutral’ growth rate.

### RNA sequencing & assembly

RNA extraction, cDNA library preparation and sequencing were conducted by Novogene (Novogene Co., Ltd., Hong Kong) and are described in detail by Breeschoten et al. (2019). In brief, RNA was extracted using TRIzol Reagent (Invitrogen Co. Ltd, San Diego, USA) and the NEBNext Ultra RNA Library Prep Kit for Illumina sequencing (NEB, Massachusetts, USA) was used for cDNA library preparation following the manufacturer’s protocol. RNA concentrations are given in Supplementary Table 1, indicating the quality and consistency of RNA samples used. All Supplementary Data is uploaded to the 4TU Centre for Research Data repository and available online: <https://figshare.com/s/f3c6a1bfacc06cfc239d>. Library quality was evaluated on an Agilent Bioanalyzer 2100 (Agilent technologies, CA, USA) and libraries were sequenced on an Illumina HiSeq 4000 platform with 150 bp paired-end (PE) reads. In total 48 RNA libraries were sequenced including three replicates, consisting of five larvae, for all treatments.

The transcriptome assembly procedure and settings were described in detail in Breeschoten et al. (2019). In brief, quality checks of sequencing reads were done using FastQC v.0.10.1 (Andrews, 2010) and filtering and adapter trimming were done using Trimmomatic v.0.36 (Bolger et al., 2014). Raw reads were additionally checked for remaining adapter sequences and potential contamination using a local BLASTN search (NCBI-BLAST+ v.2.6.0 (Camacho et al., 2009)) against the UniVec database (<ftp.ncbi.nlm.nih.gov/pub/UniVec/>, accessed: 01 February 2018).

Built: 10.0). Trinity v.2.5.1 (Grabherr et al., 2011) was used for *de novo* transcriptome assembly. Raw reads were submitted to the NCBI SRA database under accession numbers PRJNA512462 (*S. littoralis*), PRJNA556816 (*A. gamma*), PRJNA543970 (*M. brassicae*) and PRJNA548256 (*T. ni*). Each assembled transcriptome was checked for contaminated sequences using DeconSeq v.0.4.3 (Schmieder & Edwards, 2011) and subsequently suspicious transcripts were removed. All reference transcriptomes were submitted to the NCBI TSA database: GHFD000000000 (*S. littoralis*; described here: GHFD010000000), GHUD000000000 (*A. gamma*; described here: GHUD010000000), GHNQ000000000 (*M. brassicae*; described here: GHNQ010000000) and GHOK000000000 (*T. ni*; described here: GHOK010000000). All material for *S. exigua* can be found under NCBI BioProject PRJNA477295, and the transcriptome under accession GGRZ000000000, version GGRZ010000000 is used here (Breeschoten et al., 2019). See Supplementary Table 2 for an overview of the number of raw reads per library and the number of reads after trimming, cleaning and contamination checks, as well as number of contigs per transcriptome before and after contamination. Completeness of the final reference transcriptome assemblies were accessed using BUSCO v.3.0.2. using the *Insecta\_odb9* set for comparison (Simão et al., 2015) (Supplementary Table 3).

### Expression quantification

Transcript expression analysis was done for each species separately. Reads from all sample treatments per species were mapped to the respective reference transcriptome using Bowtie v.2.3.4 (Langmead & Salzberg, 2012). Isoform and gene abundance estimates were calculated using RSEM v.1.3.0 (Li & Dewey, 2011). The per-treatment count matrix was fed into a perl-based script available in the Trinity package (*abundance\_estimates\_to\_matrix.pl*) to produce a non-normalised count matrix per species and summarising all treatments. This species-specific count matrix was used for the individual differential expression analyses. Differential expression analysis for each species was done with DESeq2 v.1.18.1 (Love et al., 2014) implemented in Trinity using default settings (min\_rowSum\_counts of 2 and considering sequencing depth and RNA composition by using the median of ratios method for normalisation).

We tested for each species the sample and biological replicate relationships using principal components analyses (PCA) in the R package *stats* v.3.4.3 (R Development Core Team, 2020). We used a filtered, CPM-TMM, log<sub>2</sub> transformed and centred dataset (Supplementary Figure 4a-d). Centring is needed to focus on differences in expression rather than general physiological and metabolic functions. Therefore, the raw count matrices were filtered on abundance and normalised using count-per-million values (CPM; accounting for library size differences between samples) using edgeR v.3.20.8 (M. D. Robinson et al., 2010) in R v.3.4.3 (R Development Core Team, 2020) and trimmed mean of M values (TMM; cross-sample normalisation) (Robinson & Oshlack, 2010) using 'calcNormFactors' in edgeR v.3.20.8. Only genes with a

minimum of 10 counts in at least two samples were considered expressed and retained for further analysis.

Finally, we extracted and clustered the differentially expressed transcripts according to their patterns of expression across the samples. We considered transcripts significantly differentially expressed with a minimal fold-change of four between any of the treatments and a false discovery rate (FDR) of  $p$ -value  $\leq 1e-3$ . Hierarchical clustering of the CPM and TMM normalised expressed values of all differentially expressed transcripts was done using the Trinity script (*define\_clusters\_by\_cutting\_tree.pl*). The resulting dendrogram was pruned at 50% to define clusters with similar expression patterns.

### Transcript annotation

Different strategies were applied for transcript annotation. All isoforms of the species-specific *de novo* transcriptomes were annotated using the Trinotate pipeline v.3.0 (Haas et al., 2013). First, candidate coding regions were identified using Transdecoder v.5.0.2, from the Trinity package (Haas et al., 2013). A default length of minimal 100 amino acids (aa) for open reading frames (ORFs) was used. A BLASTP and BLASTX search was conducted, using either the translated protein sequences or the nucleotide-coded transcripts as a query against the manually annotated and non-redundant SwissProt database ([ftp://ftp.uniprot.org/pub/databases/uniprot/current\\_release/knowledgebase/complete/uniprot\\_sprot.dat.gz](ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.dat.gz); release 2019\_01, accessed 08/02/2019). A protein domain search was performed using ‘*hmmscan*’ in HMMER v.3.1b2 against the Pfam-A database v.31. Signal peptides were predicted using SignalP 4.1 server (Petersen et al., 2011), transmembrane domains were annotated using TMHMM v.2.0 (Krogh et al., 2001) and rRNA transcripts were identified using RNAMMER v.1.2 (Lagesen et al., 2007).

For all remaining annotation strategies, we annotated the highest expressed isoform per gene. Protein families were identified using InterProScan v.5.36-75 (-appl Pfam --goterms) (Jones et al., 2014) and a BLASTP search (E-value cutoff  $\leq 1e-3$ ) against UniRef90 (<ftp://ftp.uniprot.org/pub/databases/uniprot/uniref/uniref90/uniref90.fasta.gz>; release version 31/07, accessed 08/08/2019) (UniProt Consortium, 2019). We further conducted local BLASTX searches (max\_hsps 1, best\_hit\_overhang 0.1 and E-value cutoff  $\leq 1e-5$ ) against different protein databases. Two custom local protein BLAST databases were designed to specifically annotate putative detoxification genes of selected families: Cytochrome P450s (P450), carboxyl/cholinesterases (CCE), glutathione S-transferases (GST), UDP-glycosyltransferases (UGT), ATP-binding cassette transporters (ABC transporters) and glucose-methanol-choline oxidoreductases (GMC). And additionally, arylsulfatase genes for which a specific member, glucosinolate sulfatase (GSS), evolved detoxifying properties in *P. xylostella*; and the NSP-like gene family for which a member, nitrile-specifier protein (NSP), evolved detoxifying properties within Pierinae (whites). The first protein database of

these, was compiled from detoxification genes of eight different reference Lepidoptera species as described in Breeschoten et al. (2019). This database was lacking representatives of GMCs, arylsulfatases (or GSS-like genes) and NSP-like genes and was only used to identify differentially expressed genes. The second database was an OrthoDB v.10 (Kriventseva et al., 2019) (accessed 08-11/2019) derived protein database using specific keywords (Supplementary Table 5) including all selected detoxification gene families (overview detoxification genes in Supplementary Table 6). A third database was constructed based on all Arthropoda protein sequences downloaded from the NCBI protein database (accessed, 31/01/2019). Annotation results can be found in Supplementary Table 7-12.

### Gene expression comparisons across four Noctuidae species

Putative orthologous genes were identified using OrthoFinder v.2.2.7 (Emms & Kelly, 2015) under default settings, using as input the amino acid (aa) sequences of the *de novo* assemblies according to the Transdecoder translations. Resulting orthogroups (OGs) were used in the species comparisons, forming a group of genes descended from a single gene in the last common ancestor for this group of species, and thus including both orthologs and paralogs. The OGs were annotated based on individual gene identifications (Supplementary Table 13). OGs consisted of homologous genes and in some cases the annotation strategy revealed different annotations for the genes clustered within a single OG. We provide all annotations for each OG (Supplementary Table 13 and 21).

We applied the phylogenetic ANOVA method using the EVE model for studying gene expression evolution and population variance (Rohlf & Nielsen, 2015). Phylogenetic ANOVA can detect genes with increased or decreased ratios of expression while taking phylogenetic relationships into account and controlling for a phylogenetic signal, and therefore reduces bias caused by false positives. We used fold change (FC) values to compare expression levels of orthologous genes. Using FC instead of direct expression values circumvents cross-species normalisation (Dunn et al., 2013). We used the FC calculated by DESeq2 for pairwise sample comparisons between each plant treatment and the control group for each moth species individually. The FC values from orthologous genes of the different moths were then used for the between-moth comparisons. We only included one-to-one orthologs (3,369), also called strict orthologs (Fernández et al., 2020) without missing values for all samples. Clustering of the FC data matrix was done using 'hclust' from the R package *stats* v.3.4.3 (R Development Core Team, 2020) to evaluate if a phylogenetic signal was present in the expression data and, consequently, if the application of phylogenetic ANOVA using the EVE model was appropriate for our data.

Further, we employed Xspecies, a cross-species meta-analysis of gene expression (Kristiansson et al., 2013). Xspecies takes homology into account and compares expression data from genes with any number of orthologs and paralogs (Kristiansson

et al., 2013). In short, this method compares the gene-specific  $p$ -values from the individual differential gene expression analysis within OGs between the species. Per moth species and for each gene in every OG, the lowest, most significant  $p$ -value was extracted from the pairwise comparison between each plant treatment and the control group. Multiple paralogs can exist per gene in each OG and thus the lowest  $p$ -value is selected. Each species is thus represented by one  $p$ -value per OG. Based on these  $p$ -values from the individual species, a single combined  $p$ -value for each OG was calculated which gives the level of significance for differential expression of homologous genes between the species. Because the method compares significance levels in terms of  $p$ -values, it circumvents the need for normalisation.

First,  $p$ -values were extracted from the individual pairwise sample comparisons between each plant treatment and the control group as calculated with DESeq2. Thus, each moth species consisted of three samples, showing the significance of expression difference: *Z. mays* vs. control, *B. oleracea* vs. control and *N. tabacum* vs. control. In total, the dataset consisted of 9,674 OGs with missing data allowed.

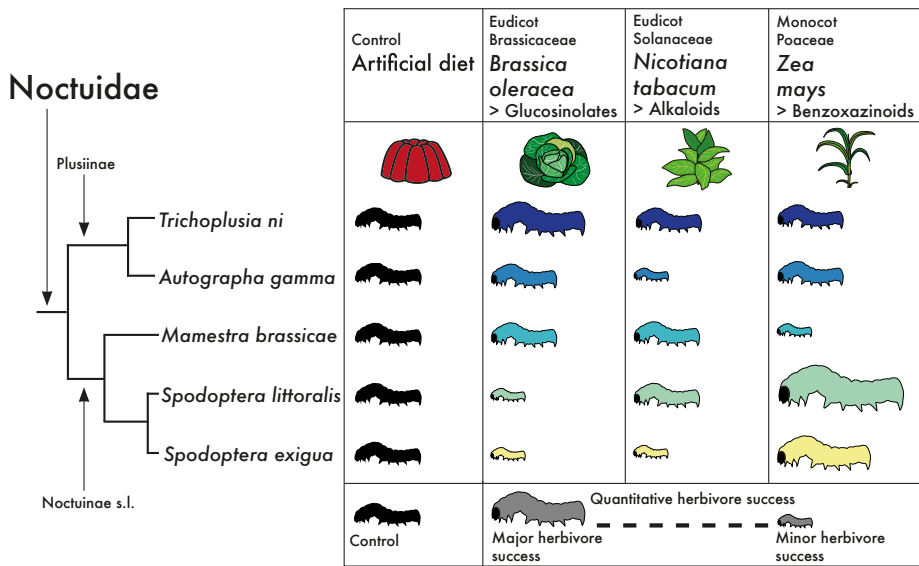
Based on the  $p$ -values of homologous genes grouped in OGs, Xspecies calculates if the compared species have a different expression. The significance of the difference in expression is calculated similar to Fisher's combined probability test and results in a  $p$ -value describing the differential expression of homologous genes between the different species. OGs were considered significantly differentially expressed between species with a false discovery rate (FDR) of  $\leq 0.05$ . Heatmaps were created using ggplot2 in R (Wickham, 2016).

## Results

### Feeding assays

The herbivore success of all Noctuidae larvae on the different host plant species was assessed by the calculated growth rate (weight (mg) / developmental time (day)) (Supplementary Table 14). The moth species showed variation in herbivore success for the different host plants and thus had variable levels of host plant specialisation (Figure 1). For *A. gamma*, the growth rates of all diet treatments were significantly different, with larvae feeding on *Z. mays* showing the highest herbivore success ( $N=73$ ; growth rate:  $0.235 \pm 0.1$  (mean  $\pm$  SD) mg/day) (Supplementary Table 15). *Mamestra brassicae* had the highest herbivore success on *B. oleracea* ( $N=105$ ; growth rate:  $0.241 \pm 0.12$  mg/day) but was not significantly different from the herbivore success on *N. tabacum* ( $N=105$ ; growth rate:  $0.221 \pm 0.18$  mg/day) and the latter not from the herbivore success on *Z. mays* ( $N=98$ ; growth rate:  $0.174 \pm 0.13$  mg/day). Larvae of *T. ni* similarly were most successful on *B. oleracea* ( $N=100$ ; growth rate:  $0.245 \pm 0.1$  mg/day). For *T. ni* the herbivore success of larvae feeding on *N. tabacum* and *Z. mays* were not significantly different from each other. Finally, similarly as *S. exigua* described in

Breeschoten et al. (2019), *S. littoralis* had the highest herbivore success on *Z. mays* (N=100; growth rate:  $0.743 \pm 0.41$  mg/day). Larvae of *S. littoralis* feeding on *Z. mays* were not significantly more successful than larvae from the control group. For all the other species, larvae in the control group, feeding on the artificial diet, showed the significant highest herbivore success.



**FIGURE 1 |** Results of the feeding assays comparing the growth rates (= herbivore success, a quantification for specialisation and pest level) of larvae feeding on the control and three different host plant species: *Brassica oleracea*, *Nicotiana tabacum* and *Zea mays* across the selected Noctuidae. In total five moth species are included (*Spodoptera exigua*, *Spodoptera littoralis*, *Mamestra brassicae*, *Autographa gamma* and *Trichoplusia ni*) for which the phylogenetic relatedness and higher taxonomic groups are shown on the left. Sizes represent herbivore success according to growth rate except for the control groups that are kept equal. Larval sizes reflect significant herbivore success differences between moth species. Size differences do not reflect significance for *S. exigua* and *A. gamma* feeding on *Z. mays* ( $p$ -value =  $5.2e-2$ ). As a consequence, sizes represent growth rate differences within moths except for *A. gamma* feeding on *B. oleracea* and *Z. mays* ( $p$ -value =  $4.8e-2$ ) and *M. brassicae* feeding on *N. tabacum* and *Z. mays* ( $p$ -value =  $4.9e-1$ ).

To compare herbivore success between Noctuidae species (Figure 1), we normalised the species-specific growth rates per host plant. For the host plant *B. oleracea* herbivore success between *S. exigua* – *S. littoralis* and between *A. gamma* – *M. brassicae* showed no significant differences- (Supplementary Table 15). For *Z. mays* there was no significant difference in herbivore success between *A. gamma* – *S. exigua* and between *A. gamma* – *T. ni*. Herbivore success difference between the Noctuidae species was much smaller for larvae feeding on *N. tabacum* than for the other host plants, no significance was found for the species pairs: *A. gamma* – *S. exigua*, *S. littoralis* – *M. brassicae*, *S. littoralis* – *T. ni* and *M. brassicae* – *T. ni* (Supplementary Table 15).

We were mainly interested in identifying genes potentially related to herbivore success. Therefore, we assigned species groups discriminating the major herbivore to minor herbivore success according to the growth rates while avoiding single species comparisons, and studied these further by gene expression analysis. For *B. oleracea*, the moths *T. ni*, *A. gamma* and *M. brassicae* showed highest herbivore success in comparison to both *Spodoptera* species (Figure 1). For *N. tabacum*, the species *T. ni*, *M. brassicae* and *S. littoralis* had the highest herbivore success. Finally, for *Z. mays* four different levels of herbivore success were found and thus we had to select three groups to avoid single-species comparisons. We formed three groups of species with highest herbivore success: the first group consisted of *S. exigua* and *S. littoralis*, the second of *S. exigua*, *S. littoralis* and *T. ni*, and finally, the third group consisted of *S. exigua*, *S. littoralis* and *A. gamma*.

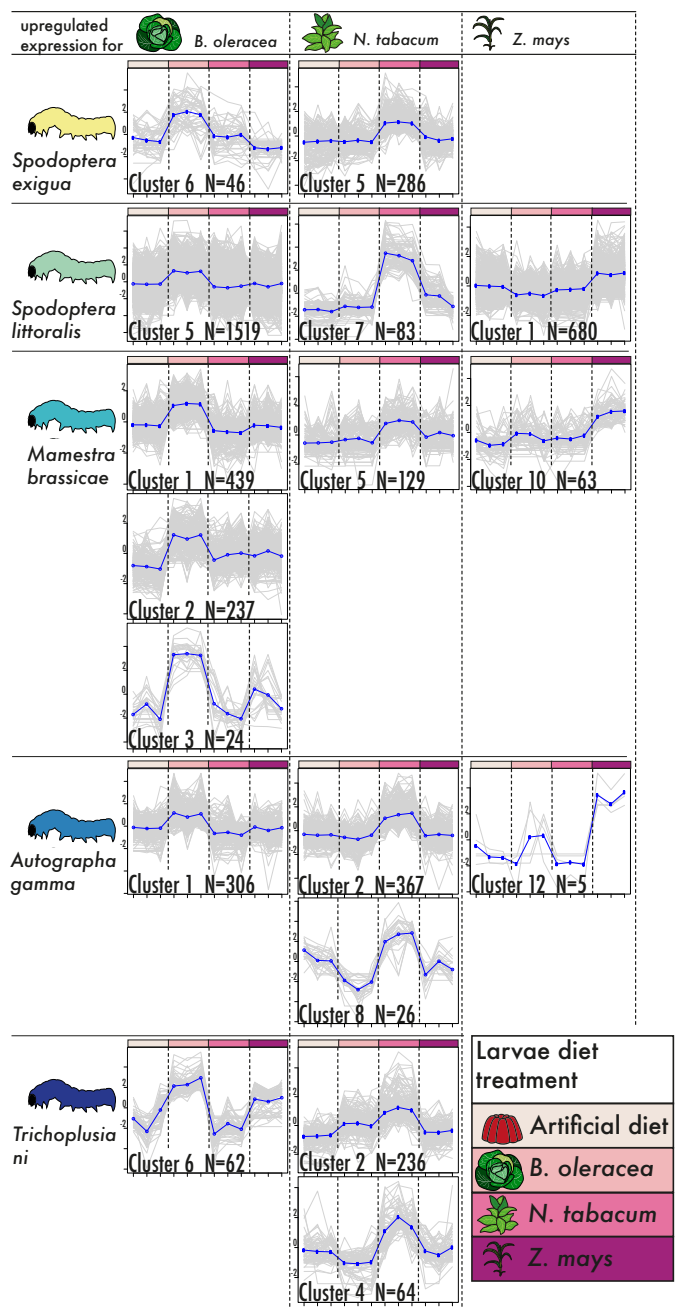
### Expression quantification

DESeq2 was used for differential gene expression analysis. The differential gene expression analyses of *A. gamma* identified 14 clusters using a 50% cut-off of in total 1,541 differentially expressed (DE) genes, for *M. brassicae* 12 clusters of in total 1,429 DE genes, for *T. ni* 11 clusters of in total 1,735 DE genes, and for *S. littoralis* 10 clusters of in total 4,384 DE genes. The CPM, TMM cross-sample normalised and filtered count matrices are available as Supplementary Tables 16a-d.

In our description of the DE genes we focus on upregulated expression patterns because genes induced by plant feeding are potentially of importance to herbivory and pest formations. For *S. littoralis*, *A. gamma* and *M. brassicae* we identified clusters with upregulated genes for all three host plants. In contrast, for *T. ni* and *S. exigua* no cluster with upregulated genes was identified for *Z. mays* (Figure 2 and Supplementary Figures 17a-d).

### Multi-species gene expression comparison

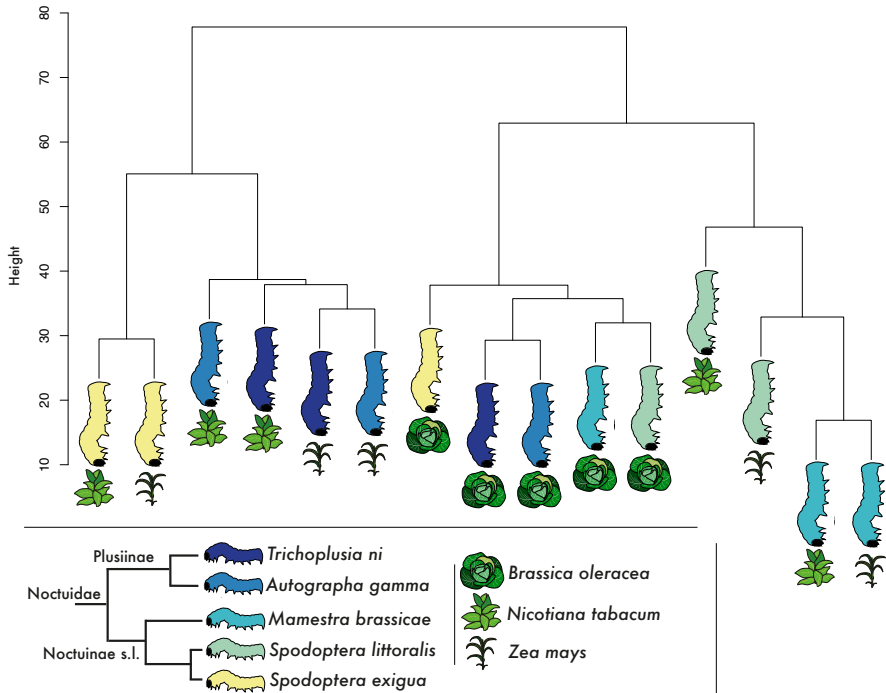
OrthoFinder returned 52,921 OGs in total, with the largest OG including 99 genes and the smallest consisting of one single gene (Supplementary Table 18). The phylogenetic ANOVA method using the EVE model resulted in unusual high *beta* values, which could indicate an absence of a phylogenetic signal (Supplementary Table 19). The *beta* parameter gives the ratio of within-species variance (diversity) to the expression divergence. Given the high *beta* values, we clustered the FC expression matrix in order to test presence of a phylogenetic signal. Indeed, the hierarchical clustering approach did show an absence of a phylogenetic signal with species clustering more closely together according to host plant diet than their phylogenetic relatedness (Figure 3). For example, all moths feeding on *B. oleracea* cluster together. However, Plusiinae cluster together for the *N. tabacum* and *Z. mays* diet treatment. This grouping - according to phylogenetic relatedness - could be indeed a result of phylogenetic signal inherent in the data. Nevertheless, the phylogenetic signal was too weak to be picked up by the



**FIGURE 2 |** Selected diet treatment-specific gene expression cluster plots. Shown are identified clusters with upregulated genes according to the differential gene expression analyses done for each species separately. Expression patterns were visualized and the number of genes (N = #) per cluster is indicated, with all samples represented on the x-axis and the normalised expression values (FPKM, median centred, log2 transformed) on the y-axis. All gene expression clusters are given in Supplementary Figures 17a-d.



statistical analysis (shared  $\beta = 99.99$ ) and thus a phylogenetic model is not appropriate for our data. We therefore implemented Xspecies, a cross species gene expression analyses based on comparison of significance levels (Kristiansson et al., 2013).



**FIGURE 3 |** Hierarchical cluster plot of the fold change expression matrix including all identified strict orthologs (N=3,369) without missing data.

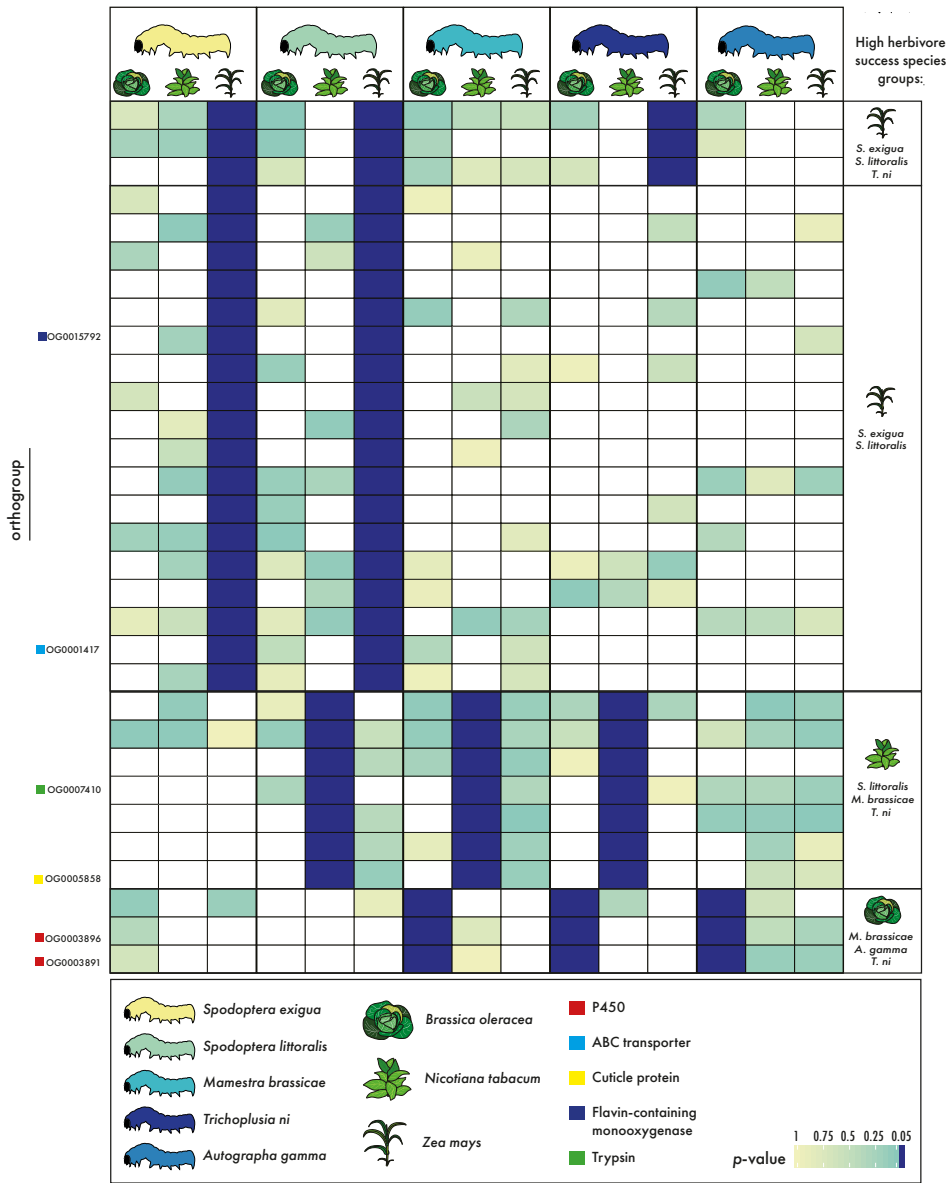
Xspecies identified in total 9,675 significant OGs (Benjamini-Hochberg false discovery rate (FDR)  $\leq 0.05$ ), of which 7,728 OGs retrieved at least one annotation (Supplementary Table 20). We identified expression patterns shared among species with high herbivore success levels and potential pest formations (Figure 4). Significant differentially expressed homologous genes (grouped in OGs) between the different moths, as calculated with Xspecies, were assigned to high herbivore success species groups. These homologous genes were significantly differentially expressed within moths showing major herbivore success for a specific host plant, while insignificant ( $p > 0.05$ ) for the remaining species with minor herbivore success levels. This revealed patterns of specific genes that are expressed by moth species showing high herbivore success for specific host plants (Figure 4).

For *Z. mays* three different groups were selected to avoid single-species comparisons, but for only two groups significant OGs were found (Figure 4). The

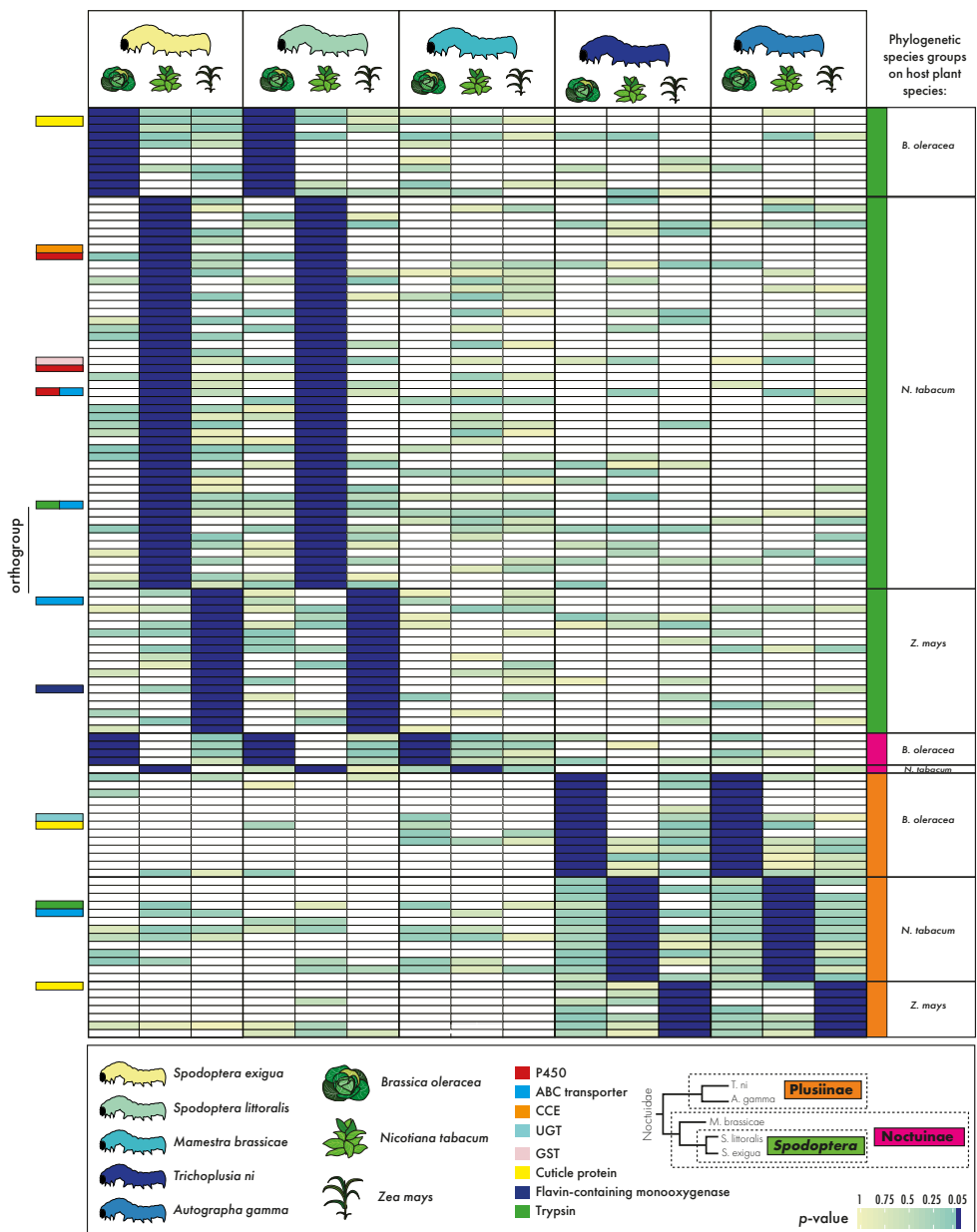
first species group, consisting of *S. littoralis*, *S. exigua* and *T. ni*, had three OGs that were differentially expressed. A similar group, consisting of *S. littoralis*, *S. exigua* and *A. gamma*, was lacking any significant OGs. The third group, with *S. exigua* and *S. littoralis*, had 18 OGs that were significantly differentially expressed. The species group showing major herbivore success for *N. tabacum* consisted of *S. littoralis*, *M. brassicae* and *T. ni*. For this group a total of seven significant OGs were found. For *B. oleracea* and the species group *M. brassicae*, *T. ni* and *A. gamma* only three OGs were identified as significantly differentially expressed (Figure 4). The majority of the genes in all OGs are involved in general physiological, developmental and metabolic functions. However, various OGs belong to gene families involved in digestion and detoxification (Supplementary Table 21). Focusing on detoxification genes within these high herbivore success species groups, only three OGs were annotated as putative detoxification genes. In the high herbivore success species group on *Z. mays* only a single ABC gene was annotated. Further, two P450 genes were found differentially expressed in the species group with high herbivore success for *B. oleracea* and those with low herbivore success (Supplementary Table 21).

Beside high herbivore success, we grouped the significantly differentially expressed genes according to phylogenetic relatedness (Figure 5). This revealed expression patterns which are likely caused by shared evolutionary histories and could therefore not be correlated to herbivore success alone. We selected three phylogenetic species groups, *Spodoptera* (*S. exigua* and *S. littoralis*), Noctuidae (*S. exigua*, *S. littoralis* and *M. brassicae*) and Plusiinae (*T. ni* and *A. gamma*). For each phylogenetic species group, we selected significantly differentially expressed OGs per host plant species, again based on the Xspecies results. The homologous genes selected within each phylogenetic clade proved significantly differentially expressed in moths feeding on the host plant compared to the control food, while insignificant ( $p > 0.05$ ) for the remaining moth species outside the clade. This is similar to the high herbivore success selection criterium and does reveal expression patterns related to specific host plant species response, in this case shared among related species.

For the phylogenetic species group *Spodoptera*, 18 OGs were significantly differentially expressed in their response to feeding on *Z. mays* as compared to the control food. Further, for the *Spodoptera* species group 11 OGs were significantly differentially expressed when feeding on *B. oleracea* and 49 OGs were significantly differentially expressed when feeding on *N. tabacum*. The species within the phylogenetic group Noctuidae had no OGs significantly differentially expressed when feeding on *Z. mays*, while four OGs were significantly differentially expressed when feeding on *B. oleracea* and one single OG significantly differentially expressed when feeding on *N. tabacum*. Finally, for the Plusiinae species group seven OGs were significantly differentially expressed when feeding on *Z. mays* and 13 OGs were significantly differentially expressed when feeding on *B. oleracea* or on *N. tabacum*.



**FIGURE 4 |** Selection for genes with expression patterns linked to high herbivore success (herbivore success species groups). Heatmap showing significantly differentially expressed homologous genes (grouped in orthogroups) between the moths as calculated with Xspecies. Genes are selected that show expression patterns related to high herbivore success for each host plant species (right). Genes are significantly upregulated in larvae feeding on the host plant as compared to the artificial food for those moths showing high herbivore success ( $\leq 0.05$ , dark blue). For moths with lowered herbivore success these homologous genes do not show significant upregulation in larvae feeding on the host compared to the artificial food (yellow to green). Heatmap colouring corresponds to intensity of this difference by significance level of the  $p$ -value. Various orthogroups are highlighted as discussed in the text. For a full list of annotated orthogroups see Supplementary Table 21.



**FIGURE 5 |** Selection for genes with expression patterns potentially caused by evolutionary history (phylogenetic species groups). Heatmap showing significantly differentially expressed homologous genes (grouped in orthogroups) between the moths as calculated with Xspecies. Genes are selected that show expression patterns that can be linked to evolutionary history, for each host plant species independently (right). The tested phylogenetic species groups are depicted in the phylogeny (lower right) with colours coding for each clade that link to the species groups in the heatmap (green = *Spodoptera*, pink = Noctuinae and orange = Plusiinae). Genes are significantly upregulated in larvae feeding on the host plant as compared to the artificial food for those moths within the phylogenetic

Multiple OGs in these phylogenetic species groups are identified as detoxification gene family members. Within the *Spodoptera* group feeding on *N. tabacum* one OG is identified as a CCE family member, three OGs as putative P450s, one GST, and one ABC transporter. Similar to the high herbivore success species group, one OG was identified as ABC in *Spodoptera* feeding on *Z. mays*. Only a single OG annotated as UGT is differentially expressed within Plusiinae feeding on *B. oleracea* compared to the other species, and one ABC transporter within Plusiinae on *N. tabacum* (Supplementary Table 21).

In general, there are no OGs shared between high herbivore success species groups and phylogenetic species groups, and thus they are comprised of different homologous genes. The one exception is the high herbivore success group on *Zea mays* consisting of both closely related *Spodoptera* species, *S. exigua* and *S. littoralis*. This group is identical to the phylogenetic species group “*Spodoptera*” feeding on *Z. mays*. Distinguishing between herbivory success and evolutionary history (phylogeny) is not possible in this case. The expression of these genes may play important roles in the feeding success of *Spodoptera* on *Z. mays* and/or may be due to shared evolutionary history.

## Discussion

We performed a multi-species comparison of gene expression patterns of larvae of five different cutworm moth species (Noctuidae), feeding and developing on three different host plant species. The polyphagous Noctuidae had different herbivore success rates per host plant species as quantified by feeding assays (Figure 1). In general, polyphagous herbivore insect species showed a level of host plant species specialisation (Schoonhoven et al., 2005), which eventually could lead to pest formations. In order to understand this specialisation, we focused on transcriptional plasticity related to high herbivore success.

To study and interpret gene expression data in a biological and evolutionary context, comparable data from closely related species needs to be collected and analysed within a multi-species comparison implementing the phylogenetic perspective (Dunn et al., 2013). Failing to incorporate phylogenetic relationships in a multi-species comparison can result in misleading and/or erroneous conclusions (Dunn et al., 2018). The explicit phylogenetic framework as applied here discriminated between dependent gene expression patterns shared as a result of evolutionary history

species groups ( $\leq 0.05$ , dark blue), while for larvae outside the clades these homologous genes are not significantly different expressed between the host plant and artificial food (yellow to green). Heatmap colouring corresponds to intensity of this difference by significance level of the *p*-value. Various orthogroups are highlighted as discussed in the text. For a full list of annotated orthogroups see Supplementary Table 21.

from those expression patterns independent from their phylogenetic relationships. Consequently, our applied multi-species comparison revealed shared and unique gene expression patterns of the five different moth species while feeding on different host plants regardless of their evolutionary history.

### Feeding assays

Herbivore success for the different host plants differed among the five species (Figure 1), although all five moth species are polyphagous herbivores able to feed on the selected plants (EPPO, 2019). The herbivore success of polyphagous insect species often varies within the range of host plant species they feed on (Schoonhoven et al., 2005). This is also evident from our feeding assay results: the levels of herbivore success for the selected host plant species are variable and species-specific. Therefore, we have focused on the transcriptional response behind this variation in herbivore success. The response to host plants can be influenced by host plant experience of parent generations (Müller et al., 2017). To reduce this bias, we have excluded insect parent host plant experience as a cause for the difference by first rearing the paternal generation of all moth species on the control diet (e.g., no anti-herbivory compounds).

### Expression data analysis and transcript annotation

Differential gene expression analyses within the individual moth species revealed diet-specific gene clusters for all host plants and the artificial food in *S. littoralis*, *A. gamma* and *M. brassicae* (Figure 2). These clusters of upregulated genes indicate a flexible transcriptional response influenced by diet. Plant-specific transcriptional responses have been observed in several polyphagous insects, including Lepidoptera (Breeschoten et al., 2019; Celorio-Mancera et al., 2012), Coleoptera (Müller et al., 2017), Hemiptera (Mathers et al., 2017), and other arthropods such as the spider mite *Tetranychus urticae* (Grbić et al., 2011).

In an earlier study, we found that *S. exigua* relies on non-diet specific expressed genes when feeding on host plants with optimal suitability, showing high herbivore success (Breeschoten et al., 2019). Larvae of *S. exigua* feeding on *Z. mays* showed high herbivore success and absence of diet-specific gene clusters, indicating that detoxification and digestion was potentially sufficient with a general gene activity. This ‘specialisation’ and absence of a diet-specific cluster as observed in *S. exigua* shares similarities with the transcriptional responses of monophagous species to their host plants. Monophagous species show a host plant specific transcriptional response, often with evolved adaptations, to efficiently detoxify the defences of their hosts (Heidel-Fischer et al., 2019; Wheat et al., 2007). Moreover, molecular studies have shown that monophagous species generally employ a lower number of expressed genes compared to polyphagous species feeding on host plant species with similar

defences (Govind et al., 2010; Ragland et al., 2015; Roy et al., 2016; Schweizer et al., 2017), relying more on a constitutive and targeted response (Berenbaum, 2002).

The ability of a polyphagous species to feed on various plants is potentially due to a greater transcriptional plasticity than of a monophagous insect (Birnbaum & Abbot, 2020). For example, the polyphagous spider mite (*T. urticae*) showed a 3-fold increase of DE genes when shifting to a new, and thus less adapted/optimal host plant (Dermauw, Wybouw, et al., 2013). Thus, the absence of a diet-specific expression cluster in *S. exigua* feeding on *Z. mays* might be due to 'specialisation' for this host plant (Breeschoten et al., 2019). However, this non-diet specific response seems to be species-specific to *S. exigua*. For example, *S. littoralis* showed also the highest herbivore success on *Z. mays* (Figure 1), but in contrast to *S. exigua* we could identify for *S. littoralis* also a large number of upregulated genes (Figure 2). In fact, *S. littoralis* showed the largest number of upregulated genes in combination with highest herbivore success for *Z. mays*. We conclude that the number of induced genes, as adaptive response to an optimal host plant, is rather species-specific and there is no general pattern within polyphagous Noctuidae. However, the broad, species-specific transcriptional plasticity might enable the wide host plant usage of polyphagous insects (Birnbaum & Abbot, 2020).

The expression of detoxification genes within diet-dependent gene clusters is of particular interest in insect-plant interaction studies (further discussed in Supplementary File 22). For each moth species we focused on the expression of eight gene families (P450, CCE, GST, UGT, ABC, GMC, NSP-like gene family and GSS-like genes) involved in detoxification of plant defence compounds (Heidel-Fischer & Vogel, 2015; Kant et al., 2015; Li et al., 2007). Detoxification activity of nitrile-specifier protein (NSP) and glucosinolate sulfatases (GSS) is known from specific butterfly / moth clades (Pierinae and *Plutella*). In Pierinae (whites), specialised on glucosinolate containing crucifers, a NSP gene evolved to modify the toxic degradation product of glucosinolates into less toxic nitriles (Fischer et al., 2008; Wheat et al., 2007; Wittstock et al., 2004). The NSP-like gene family is an insect-specific family, but only in Pierinae members an evolved NSP gene is identified with a glucosinolate detoxifying mechanism (Fischer et al., 2008; Wheat et al., 2007). For the NSP-like family we identified gene members in all five Noctuidae species, but none was differentially expressed (see NSP, Supplementary Table 6). Our results indicated that there is no *B. oleracea* specific detoxification activity, as expected from the strict Pierinae response reported earlier (Fischer et al., 2008; Wheat et al., 2007). The diamondback moth (*Plutella xylostella*) evolved a GSS enzyme which is part of an arylsulfatase class of genes that prevents the formation of toxic hydrolysis products by Brassicaceae plants (Heidel-Fischer et al., 2019; Ratzka et al., 2002). We looked at the expression of arylsulfatase genes, but a host-specific response was absent in all Noctuidae even though members were differentially expressed in *T. ni* and *S. littoralis* (see GSS, Supplementary Table 6). The general expression pattern of arylsulfatases (or GSS-like) and NSP-like gene family members reconfirms the clade-specific detoxifying function in *P. xylostella* and Pierinae, since a detoxifying role in Noctuidae was absent.

For the other gene families, detoxifying properties were not restricted to specific species but generally recognised in herbivorous insects (Heckel, 2018; Kant et al., 2015; Zhang et al., 2013). The differentially expressed glucose-methanol-choline oxidoreductases (GMCs) were identified in all moth species feeding on multiple host plants (see GMC, Supplementary Table 6). Oxidoreductases have been described to degrade various groups of plant specialised metabolites (Müller et al., 2017; Zhang et al., 2013). According to our results, there is no evidence for a host plant species-specific expression of GMCs or herbivore success-related response and GMCs seems in general to be involved in the degradation of specialised metabolites of plants. However, oxidoreductases have been associated with immunity (Sun et al., 2012), which would also explain the observed wide expression pattern.

Further, all Noctuidae species had members of P450, CCE, GST, UGT and ABC families upregulated in individuals feeding on *B. oleracea*, except for *S. exigua*, lacking significantly upregulated P450s, and *T. ni*, lacking upregulated CCEs (Supplementary Table 6). Since both P450s and CCEs are involved in hydrolyses, detoxification of glucosinolates is potentially dominated by a single family. The main defence strategy of *B. oleracea* against herbivores is by formation of toxins like isothiocyanates (ITCs) after activation of glucosinolates by myrosinases (e.g. due to chewing) (Kliebenstein et al., 2005). Many different forms of counteradaptations by herbivore insects are described (Jeschke et al., 2016; Jeschke et al., 2017). The general detoxification strategy, besides specialised responses like sequestration (Beran et al., 2014; Zagrobelny & Møller, 2011), involves diverse gene families from the detoxification pathway (Jeschke et al., 2016). Detoxification can potentially take place by direct conjugation (phase II) of ITC due to its highly reactive state and skipping the first phase of activating the metabolite (Jeschke et al., 2016; correspondence Katharina Schramm, November 2020). This could explain the absence of upregulated P450s or CCEs in *S. exigua* and *T. ni* in moths feeding on *B. oleracea*. Gene members of the GST family were found upregulated in all moths feeding on *B. oleracea* (Supplementary Table 6), which supports the suspected role of GSTs in the conjugation reaction (Jeschke et al., 2016; Schramm et al., 2012). However, several studies showed that a different detoxification strategy by polyphagous moths is potentially through direct conjugation of ITC by L-glutathione (GSH) (Jeschke et al., 2016; Schramm et al., 2012). It was shown that this conjugation mechanism is possibly shared among many polyphagous Noctuidae species, including *M. brassicae*, *T. ni*, *S. exigua* and *S. littoralis* based on glutathione conjugates in larval faeces (Jeschke et al., 2017; Schramm et al., 2012). Also, the clustering of all Noctuidae species in the *B. oleracea* diet treatment using the fold change (FC) expression matrix (Figure 3) further indicates a shared response to the toxins employed by *B. oleracea* and might indicate a shared ITC conjugation process.



## Multi-species comparison

In multi-species comparisons, phylogenetic relationships can - in pairwise comparisons - result in a biased outcome if they are neglected. Pairwise comparisons are not independent in a multi-species comparison because some species pairs are evolutionarily more closely related than others and due to evolutionary history might show a similar expression profile (Dunn et al., 2018). Therefore, using the phylogenetic ANOVA approach, we first tested for the presence of similar expression profiles due to shared evolutionary history. However, the hierarchical clustering of the strict orthologs (Figure 3) and the high *beta* values according to the phylogenetic ANOVA analyses (Supplementary Table 19) indicated that the species under comparison were too closely related to observe any effects on selection and/or the absence of a clear phylogenetic signal in the expression data. The absence of a clear phylogenetic signal in the expression data makes implementing the phylogenetic ANOVA with the EVE model inappropriate (correspondence developer EVE model, Lars Grønvold, May 2020). Consequently, we implemented Xspecies, a cross species gene expression analyses based on comparison of significance levels (Kristiansson et al., 2013).

## Expression comparison using Xspecies

We compared the gene expression patterns of the polyphagous moth species to study the transcriptional response contributing to polyphagy and herbivore success. From the significantly differentially expressed homologous genes (grouped in OGs) as calculated with Xspecies, we selected for each host plant OGs with shared expression patterns related to high herbivore success (see results for selection details) (Figure 4).

The OGs identified showed that moths feeding with high herbivore success on host plants employ host plant specific OGs (Figure 4; Supplementary Table 21). We further focused on the shared expression of detoxification genes for species with high herbivore success (=high herbivore success species groups), given the important role of detoxification in successful plant feeding and resisting specialised metabolites.

Three OGs within the high herbivore success species groups belonged to one of the analysed detoxification gene families (P450, CCE, GST, UGT, ABC) (Figure 4). This low number of 'detoxification OGs' with shared response (three in total) indicated that expression of detoxification gene family members was highly species-specific because each moth transcribed a larger number of DE detoxification genes (Supplementary Table 6). Nevertheless, the shared response of the three OGs could be correlated to host plant adaptation of species with high herbivore success. Two OGs were annotated as P450 members, a gene family broadly involved in detoxification (Feyereisen, 2005; Schuler, 2011). Activity of these genes could be associated with high herbivore success of *M. brassicae*, *A. gamma* and *T. ni* while feeding on *B. oleracea* (Figure 4). Further, we identified one OG as ABC transporter, a family involved in the detoxification pathway (Bretschneider et al., 2016; Dermauw, Osborne, et al., 2013; Dermauw & Van Leeuwen, 2014). This OG was significantly upregulated in both *Spodoptera* species

while feeding on *Z. mays* with high herbivore success. However, given the shared evolutionary history of these species we could not correlate the activity of this OG to herbivore success alone (Figure 4 and 5).

Other gene families than these ‘common’ detoxification families could also play key roles in high herbivore success. For example, OG0015792, annotated as Flavin-containing monooxygenase (FMO) based on Pfam domains and senecionine N-oxygenase by Uniref90 (Supplementary Table 21), was significantly differentially expressed in both, *S. exigua* and *S. littoralis* while feeding on *Z. mays* with high herbivore success and absent in most other species (Figure 4). The FMOs are known to oxidise xenobiotics in a similar way as P450s (Eswaramoorthy et al., 2006), and are involved in alkaloid detoxification (Lindigkeit et al., 1997; Naumann et al., 2002). The activity of the FMO annotated gene members in OG0015792 might indicate a potential role in herbivory of Noctuidae.

We also selected host plant species-specific expression patterns that are found only within three phylogenetic species groups (*Spodoptera*, Noctuidae and Plusiinae). These expression patterns were likely caused by shared evolutionary history, and as a consequence might not be purely correlated to herbivore success (Figure 5). The number of shared OGs in phylogenetic species groups (average = 15 OGs) was larger than within high herbivore success species groups (average = 4 OGs), indicating the importance of lineage-specificity for host plant response. However, the larger number of shared OGs was not found for the Noctuidae groups (*S. exigua*, *S. littoralis* and *M. brassicae*; 4 OGs for *B. oleracea*, 1 OG for *N. tabacum* and none for *Z. mays*; Figure 5), which could be explained by the larger phylogenetic distance due to inclusion of a third distant related species (*M. brassicae*).

Of the annotated detoxification gene families, all families (P450, CCE, UGT, GST and ABC) were present in the shared OGs of the phylogenetic species groups (Figure 5). However, the detoxification genes families CCE, GST, and UGT were represented by a single OG only. Indeed, based on the individual moth differential expression analysis we identified many DE detoxification genes (Supplementary Table 6). However, the shared response of only a single OG of these gene families within phylogenetic species groups (or three or four OGs for P450 and ABC) did indicate that the individual moths rely for a large degree on a species-specific employment of detoxification genes.

Most gene families identified as significantly expressed by the multi-species comparison have no clear role in detoxification based on the annotations (Supplementary Table 21). Indeed, various OGs were involved in processes like physiological, metabolic and developmental functions, indicating a potential impact of host plant feeding on these processes. However, a shared transcriptional response of homologous genes from these gene families, both within high herbivore success species groups (Figure 4) or within phylogenetic species groups (Figure 5), could indicate a role in herbivory. The role of gene families not known for detoxifying properties have been shown to be potentially important in feeding on newly introduced host plant species (Dermauw, Wybouw, et al., 2013; Wybouw et al., 2015). Indeed,

various gene families were shared between high herbivore success species groups and phylogenetic species groups. The wide occurrence and general expression of members of these gene families potentially showed a general role of these families in herbivory. However, the unique expression of specific OGs of these widely deployed gene families, associated with high herbivore success, indicates their importance for host plant feeding and potential pest formations.

An example of a widely occurring gene family is trypsin which is involved in the digestion of plant material by hydrolyses of proteins (Muhlia-Almazán et al., 2008). The unique occurrence of the trypsin homologs containing OG (OG0007410), identified in the high herbivore success species group consisting of *S. littoralis*, *M. brassicae* and *T. ni*, may indicate the important digestive roles of trypsins while feeding on *N. tabacum* (Figure 4). However, trypsin family members were also widely identified within all individual moth gene expression analyses (Supplementary Table 7-12). Also within the two phylogenetic species groups, *Spodoptera* and *Plusiinae*, trypsin family members were shared while feeding on *N. tabacum* (OG0020891 and OG0008359; Figure 5). Indeed, the wide expression of trypsins within all moth species indicated a general importance for herbivory.

We observed a similar pattern for the insect cuticle protein family. A strengthened cuticle of the peritrophic matrix and midgut, that comes into contact with toxins and abrasive food particles, may improve the protective function by forming a protective physical barrier for biochemical toxins (Agrawal et al., 2014; Hegedus et al., 2009; Kelkenberg et al., 2015; Kumar et al., 2018). Differential expression of transcripts coding for structural constituents of the cuticle in response to the diet is found in larvae of various herbivorous insect species (Breeschoten et al., 2019; Celorio-Mancera et al., 2013; Hoang et al., 2015; Müller et al., 2017; Orsucci et al., 2018). In our multi-species comparison, we identified three OGs as members of the insect cuticle protein family. However, only one OG can be associated with high herbivore success on *N. tabacum* (OG0005858, Figure 4). The two other OGs (OG0011643 and OG0007820, Figure 5) share lineage-specific expression within the phylogenetic species group *Plusiinae*. In general, the wide occurrence of certain families, such as trypsin and cuticle protein families show their importance for herbivory. The employment seems highly lineage and species-specific. Only the expression of a few specific OGs of these gene families can be associated with high herbivore success. These genes are of high interest for future studies aiming for a deeper understanding of the genetic mechanisms of potential pest formations in Noctuidae moths.

Transcriptomic studies have been important in understanding gene expression and transcriptional plasticity in plant feeding and host adaptation (Birnbaum & Abbot, 2020). The implementation of genome data in plant-insect interactions provided insight in many aspects of plant-insect interactions, for example the transcriptional differences between monophagous and polyphagous insects (Govind et al., 2010; Schweizer et al., 2017), and pesticide resistance (Dermauw et al., 2018; Dermauw, Wybouw, et al., 2013). Based on these studies, certain genes and gene families have

been identified as important, e.g., for adaptation to new host plants (Dermauw, Wybouw, et al., 2013) or to play a key role in herbivory for a range of butterfly species (Nallu et al., 2018). Our results provide insight in the genetic mechanism of successful herbivory; the initial step for pest formations. The polyphagous moth species in the comparison are infamous for their ability to form pests. Many Noctuidae are well known pest species (EPPO, 2019; Reed & Pawar, 1982; Tay et al., 2013), in particular members of the genus *Spodoptera* (Early et al., 2018; Goergen et al., 2016; Montezano et al., 2018). Under influence of global climate change the number of pest outbreaks, of in particular invasive species, have increased (Battisti & Larsson, 2015; Dukes et al., 2009; Yan et al., 2017; Ziska et al., 2011). Expanding our understanding of the genetic basis of pest formations is therefore of biological and practical importance. We aimed to identify if successful herbivory is determined by a shared or lineage-specific transcriptional response in Noctuidae. By implementation of a phylogenomic framework we discerned the shared expression of homologous genes due to high herbivore success from genes shared under influence of the relatedness of the moth species.

## Conclusions

We studied the gene expression of polyphagous Noctuidae species feeding on different host plants associated with varying levels of herbivore success as quantified with feeding assays. Our work shows how major polyphagous insects rely on the deployment of some widely employed gene families, indicating their importance for herbivory and polyphagy. However, transcriptional plasticity was high and moth species-specific. By implementation of a phylogenomic framework we identified groups of homologous genes with shared expression within clades of related species (*Spodoptera*, *Plusiinae* and *Noctuinae*). Furthermore, we identified shared expression patterns of homologous genes between moth species associated with high herbivore success, independent of phylogeny and thus indicating convergence.

Our main aim was to evaluate if successful herbivory is determined by a shared or lineage-specific transcriptional response. We conclude that successful polyphagous herbivores, or potential pests, have shared expression of groups of homologous genes but rely also on species-specific transcriptional plastic expression. The shared expression indicated the potential role of these genes in reaching high herbivore success on specific plants. Clearly, similar comparative studies are needed to verify the shared gene activity in related (polyphagous) insect clades forming pests. As a whole, our results provide an initial overview of the genetic basis of polyphagy and pest formations. The lineage-specific shared expression, putatively important in plant feeding among related species is of interest for the understanding of the evolution and genetic basis of polyphagy. Pinpointing the shared expression of genes associated with high herbivore success is a promising step towards the development of sustainable ways of coping with, and genetic understanding of, herbivorous insects forming pests.

## Acknowledgements

We want to thank the Unifarm Wageningen University & Research personnel for providing, rearing and overall plant care. Further we are grateful to Dr Maike Stam (UvA) for providing *Z. mays* B73 seeds and Dr Jessica M. Nifong (US Nicotiana Germplasm Collection) for providing *N. tabacum* seeds. Finally, we want to thank Paul W. Goedhart from Biometris, Wageningen University & Research for advice on the feeding assay statistics and comparison.

## Supplementary data

All Supplementary Data is uploaded to the 4TU Centre for Research Data repository and available online: <https://figshare.com/s/f3c6a1bfacc06cfc239d>.

**SUPPLEMENTARY TABLE 1** | RNA concentrations of all samples tested, indicating the quality and consistency of RNA samples used.

**SUPPLEMENTARY TABLE 2** | Overview of the number of raw reads per library and the final number of reads after trimming, cleaning and contamination checks for all four new species presented in this study: *Spodoptera littoralis*, *Autographa gamma*, *Trichoplusia ni*, *Mamestra brassicae*.

**SUPPLEMENTARY TABLE 3** | BUSCO scores of all final used transcriptome assemblies.

**SUPPLEMENTARY FIGURE 4A** | PCA analysis results of sample relationships based on the filtered, CPM-TMM and log2 transformed centred dataset of *Spodoptera littoralis*.

**SUPPLEMENTARY FIGURE 4B** | PCA analysis results of sample relationships based on the filtered, CPM-TMM and log2 transformed centred dataset of *Autographa gamma*.

**SUPPLEMENTARY FIGURE 4C** | PCA analysis results of sample relationships based on the filtered, CPM-TMM and log2 transformed centred dataset of *Mamestra brassicae*.

**SUPPLEMENTARY FIGURE 4D** | PCA analysis results of sample relationships based on the filtered, CPM-TMM and log2 transformed centred dataset of *Trichoplusia ni*.

**SUPPLEMENTARY TABLE 5** | Keywords used to gather annotated orthologous proteins from OrthoDB v.10 from selected gene families.

**SUPPLEMENTARY TABLE 6** | Overview of putative detoxification genes divided per family and expression cluster as annotated with InterProScan and BLASTX using both detoxification gene databases.

**SUPPLEMENTARY TABLE 7A** | Trinotate annotation report of *Spodoptera littoralis*.

**SUPPLEMENTARY TABLE 7B** | Trinotate annotation report of *Autographa gamma*.

**SUPPLEMENTARY TABLE 7C** | Trinotate annotation report of *Mamestra brassicae*.

**SUPPLEMENTARY TABLE 7D** | Trinotate annotation report of *Trichoplusia ni*.

**SUPPLEMENTARY TABLE 8A** | Interproscan annotation report of *Spodoptera exigua*.

**SUPPLEMENTARY TABLE 8B** | Interproscan annotation report of *Spodoptera littoralis*.

**SUPPLEMENTARY TABLE 8C** | Interproscan annotation report of *Autographa gamma*.

**SUPPLEMENTARY TABLE 8D** | Interproscan annotation report of *Mamestra brassicae*.

**SUPPLEMENTARY TABLE 8E** | Interproscan annotation report of *Trichoplusia ni*.

**SUPPLEMENTARY TABLE 9A** | BLASTP annotated genes of *Spodoptera exigua* using UniRef90 database.

**SUPPLEMENTARY TABLE 9B** | BLASTP annotated genes of *Spodoptera littoralis* using UniRef90 database.

**SUPPLEMENTARY TABLE 9C** | BLASTP annotated genes of *Autographa gamma* using UniRef90 database.

**SUPPLEMENTARY TABLE 9D** | BLASTP annotated genes of *Mamestra brassicae* using UniRef90 database.

**SUPPLEMENTARY TABLE 9E** | BLASTP annotated genes of *Trichoplusia ni* using local UniRef90 database.

**SUPPLEMENTARY TABLE 10A** | BLASTX annotated genes of *Spodoptera exigua* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10B** | BLASTX annotated genes of *Spodoptera littoralis* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10C** | BLASTX annotated genes of *Autographa gamma* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10D** | BLASTX annotated genes of *Mamestra brassicae* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10E** | BLASTX annotated genes of *Trichoplusia ni* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10A2** | BLASTX annotated DE genes of *Spodoptera exigua* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10B2** | BLASTX annotated DE genes of *Spodoptera littoralis* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10C2** | BLASTX annotated DE genes of *Autographa gamma* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10D2** | BLASTX annotated DE genes of *Mamestra brassicae* using local Arthropoda database.

**SUPPLEMENTARY TABLE 10E2** | BLASTX annotated DE genes of *Trichoplusia ni* using local Arthropoda database.

**SUPPLEMENTARY TABLE 11A** | BLASTX annotated genes of *Spodoptera exigua* using local detoxification gene family database.

**SUPPLEMENTARY TABLE 11B** | BLASTX annotated genes of *Spodoptera littoralis* using local detoxification gene family database.

**SUPPLEMENTARY TABLE 11C** | BLASTX annotated genes of *Autographa gamma* using local detoxification gene family database.

**SUPPLEMENTARY TABLE 11D** | BLASTX annotated genes of *Mamestra brassicae* using local detoxification gene family database.

**SUPPLEMENTARY TABLE 11E** | BLASTX annotated genes of *Trichoplusia ni* using local detoxification gene family database.

**SUPPLEMENTARY TABLE 12A** | BLASTX annotated DE genes of *Spodoptera exigua* using a local detoxification gene family database compiled from 8 different Lepidoptera species (see Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 12B** | BLASTX annotated DE genes of *Spodoptera littoralis* using a local detoxification gene family database compiled from 8 different Lepidoptera species (see Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 12C** | BLASTX annotated DE genes of *Autographa gamma* using a local detoxification gene family database compiled from 8 different Lepidoptera species (see Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 12D** | BLASTX annotated DE genes of *Mamestra brassicae* using a local detoxification gene family database compiled from 8 different Lepidoptera species (see Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 12E** | BLASTX annotated DE genes of *Trichoplusia ni* using a local detoxification gene family database compiled from 8 different Lepidoptera species (see Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 13** | Overview of annotated orthogroups based on identifications of the individual genes using different identification techniques including InterProScan and various BLASTX runs implementing local detoxification databases and UniRef90.

**SUPPLEMENTARY TABLE 14** | Overview of the results of the feeding assays for all Noctuidae compared in this study: *Spodoptera littoralis*, *Autographa gamma*, *Trichoplusia ni*, *Mamestra brassicae*, *Spodoptera exigua* (from Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 15** | Overview feeding assay statistics for all Noctuidae compared in this study: *Spodoptera littoralis*, *Autographa gamma*, *Trichoplusia ni*, *Mamestra brassicae*, *Spodoptera exigua* (from Breeschoten et al. 2019).

**SUPPLEMENTARY TABLE 16A** | CPM, TMM cross-sample normalised and filtered count matrix of *Spodoptera littoralis*.

**SUPPLEMENTARY TABLE 16B** | CPM, TMM cross-sample normalised and filtered count matrix of *Autographa gamma*.

**SUPPLEMENTARY TABLE 16C** | CPM, TMM cross-sample normalised and filtered count matrix of *Mamestra brassicae*.

**SUPPLEMENTARY TABLE 16D** | CPM, TMM cross-sample normalised and filtered count matrix of *Trichoplusia ni*.

**SUPPLEMENTARY TABLE 17A** | Overview of all expression pattern clusters of differentially expressed genes for *S. littoralis*. The DE genes were partitioned based on expression pattern similarity using a 50% cut-off for the hierarchical-clustering dendrogram. Each host plant species is represented by three larvae samples: Control group (Lit23, Lit24, Lit25), *Z. mays* (Lit27, Lit28, Lit29), *N. tabacum* (Lit35, Lit36, Lit37), *B. oleracea* (Lit31, Lit32, Lit22).

**SUPPLEMENTARY TABLE 17B** | Overview of all expression pattern clusters of differentially expressed genes for *A. gamma*. The DE genes were partitioned based on expression pattern similarity using a 50% cut-off for the hierarchical-clustering dendrogram. Each host plant species is represented by three larvae samples: Control group (Gam80, Gam81, Gam82), *Z. mays* (Gam85, Gam86, Gam87), *N. tabacum* (Gam101, Gam102, Gam103), *B. oleracea* (Gam89, Gam90, Gam91).

**SUPPLEMENTARY TABLE 17C** | Overview of all expression pattern clusters of differentially expressed genes for *M. brassicae*. The DE genes were partitioned based on expression pattern similarity using a 50% cut-off for the hierarchical-clustering dendrogram. Each host plant species is represented by three larvae samples: Control group (Bra22, Bra23, Bra24), *Z. mays* (Bra35, Bra36, Bra37), *N. tabacum* (Bra27, Bra28, Bra29), *B. oleracea* (Bra32, Bra33, Bra39).

**SUPPLEMENTARY TABLE 17D** | Overview of all expression pattern clusters of differentially expressed genes for *T. ni*. The DE genes were partitioned based on expression pattern similarity using a 50% cut-off for the hierarchical-clustering dendrogram. Each host plant species is represented by three larvae samples: Control group (Bra52, Bra53, Bra54), *Z. mays* (Bra67, Bra68, Bra69), *N. tabacum* (Bra57, Bra58, Bra59), *B. oleracea* (Bra62, Bra63, Bra64).

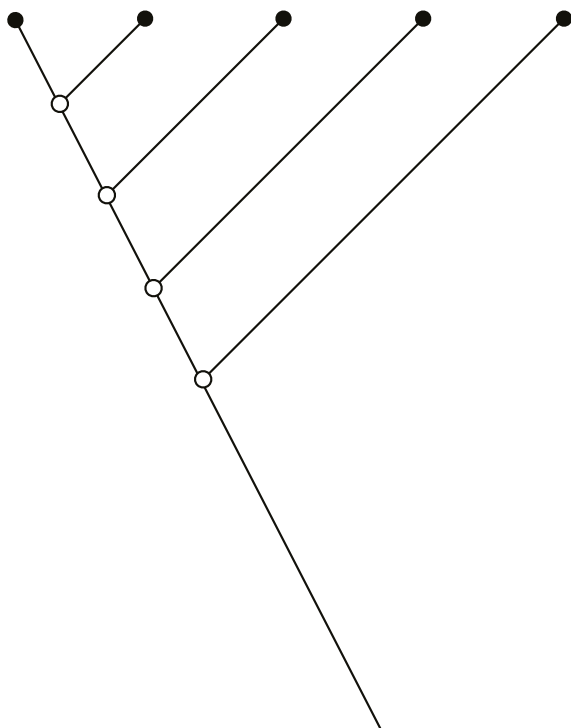
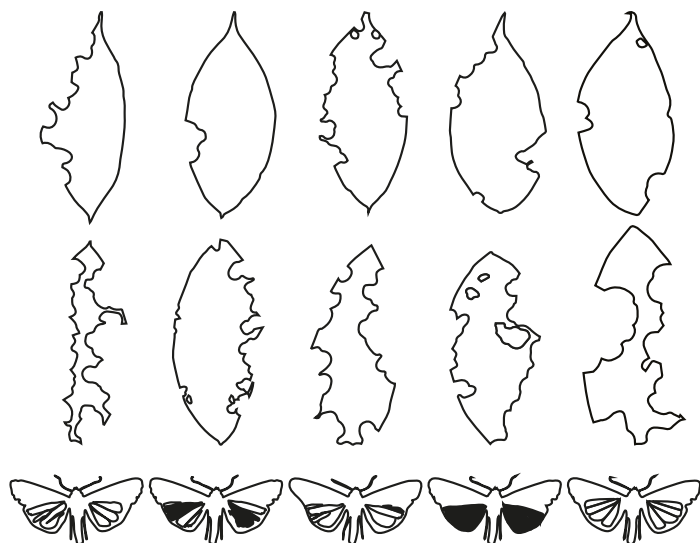
**SUPPLEMENTARY FILE 18** | Orthogroups identified using OrthoFinder v.2.2.7 including translated proteins from all five Noctuidae species included in this study.

**SUPPLEMENTARY TABLE 19** | Output of the phylogenetic ANOVA (Beta shared test) including the calculated Beta values for all orthogroups consisting of strict orthologs.

**SUPPLEMENTARY TABLE 20** | Raw output from the Xspecies analysis run showing statistics of each orthogroup for the multi-species gene expression analysis. With first letter representing the species (E=*S. exigua*; L=*S. littoralis*; M=*M. brassicae*; T=*T. ni*; A=*A. gamma*) and second letters standing for the host plant species (B=*B. oleracea*; N=*N. tabacum*; Z=*Z. mays*).

**SUPPLEMENTARY TABLE 21** | Orthogroups significantly differentially expressed between the control- and host plant species treatments within selected species groups as compared to the other Noctuidae, detected with Xspecies and annotated using various annotation strategies including UniRef90, InterProScan and various BLASTX runs.

**SUPPLEMENTARY DATAFILE 22** | Supplementary results and discussion.





# Chapter 4

Expanding the menu: are polyphagy and gene family expansions linked across Lepidoptera?

Thijmen Breeschoten<sup>1</sup>,  
Corné F. H. van der Linden<sup>1</sup>,  
Vera I. D. Ros<sup>2</sup>,  
M. Eric Schranz<sup>1</sup>,  
Sabrina Simon<sup>1</sup>

<sup>1</sup> Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>2</sup> Laboratory of Virology, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands



Evolutionary expansions and contractions of gene families are often correlated with key innovations and/or ecological characteristics. In butterflies and moths (Lepidoptera), expansions of gene families involved in detoxification of plant specialised metabolites are hypothesised to facilitate a polyphagous feeding style. However, analyses supporting this hypothesis are mostly based on a limited number of lepidopteran species. We applied a phylogenomics approach, using 37 lepidopteran genomes, to analyse if gene family evolution (gene gain and loss) is associated with the evolution of polyphagy. Specifically, we compared gene counts and evolutionary gene gain and loss rates of gene families involved in adaptations to plant feeding. We correlated gene evolution to host plant family range (phylogenetic diversity) and specialised metabolite content of plant families (functional metabolite diversity). We found a higher rate for gene loss than gene gain in Lepidoptera, a potential consequence of fractionation after a duplication event. Gene family expansions and contractions varied across lepidopteran families, and were associated with host plant use and specialisation levels. Within the family Noctuidae, a higher expansion rate for gene families involved in detoxification can be related to the large number of polyphagous species. However, gene family expansions are observed in both polyphagous and monophagous lepidopteran species and thus seems to be species-specific. Nevertheless, a significant positive correlation of gene counts of the carboxyl- and cholinesterase and glutathione S-transferase detoxification gene families with the level of polyphagy was identified across Lepidoptera.

#### Keywords

Gene family evolution; gene family expansion; herbivory; butterfly-plant interactions; Lepidoptera

#### Abbreviations

Phylogenetic Diversity (PD), Functional Metabolite Diversity (FMD), Cytochrome P450 monooxygenase (P450), Carboxyl/Cholinesterase (CCE), Glutathione-S-transferase (GST), UDP-glycosyltransferase (UGT), ATP-binding cassette transporters (ABC), Orthogroup / Orthologous protein Group (OG)

#### Significance statement

Major expansions of gene families involved in plant feeding, such as detoxification of plant specialised metabolites, are hypothesised to facilitate and enable polyphagy in herbivorous insects. To test this hypothesis, we apply a comparative phylogenomics framework. We find gene family expansions to occur in both monophagous and polyphagous Lepidoptera, but we do find a significant positive correlation between the size of the detoxification gene families CCE and GST with level of polyphagy. Thus, gene family sizes are variable across monophagous and polyphagous Lepidoptera but expansions can be correlated to host plant breadth for specific gene families.

## Introduction

The insect order Lepidoptera (butterflies and moths) form an insect super radiation with ~160,000 described species found in nearly all ecosystems on earth (Grimaldi & Engel, 2005; Pogue, 2009; van Nieukerken et al., 2011). Furthermore, Lepidoptera are one of the largest radiations of plant-feeding insects, with plant feeding having evolved from inner plant tissue (via concealed external feeding) to exposed folivory (Menken et al., 2010; Mitter et al., 2017). The diversity of Lepidoptera and the expansion of the order have been linked to the close association with- and rise of angiosperms (Allio et al., 2021; Ehrlich & Raven, 1964; Mitter et al., 2017). The coevolutionary relationship through a process termed “escape-and-radiate” has long been hypothesised to be the driving force for the diversification of both flowering plants and butterflies (Ehrlich & Raven, 1964; Thompson, 1989). However, Lepidoptera primarily feed upon plant families that radiated before most butterfly and moth families did (Janz & Nylin, 1998; Labandeira & Sepkoski, 1993). Therefore, studies argue that host-shifts, through colonisation and specialisation, have likely shaped the patterns of insect-plant associations (Braby & Trueman, 2006; Janz & Nylin, 1998; Jermy, 1984). The study of the close interactions and adaptations of Lepidoptera to their host plants have benefited from genomic analysis (Birnbaum and Abbot 2020; Simon, et al. 2015; Triant, et al. 2018).

Genomic changes can be correlated to adaptive changes and ecological characteristics associated with plant feeding (Edger, et al. 2015; Gloss, et al. 2019; Simon, et al. 2015). Correlating genomic changes to evolutionary processes, like radiation events and dietary shifts, can reveal the genomic drivers of these processes (Allio et al., 2021; Seppey, Ioannidis, et al., 2019). Genomic changes vary from point mutations, to expansions of specific gene-families up to genome duplications. Duplicated gene copies can lead to a selective advantage and may eventually be preserved by selective forces, or alternatively be non-beneficial and thus lost and/or deleted (Innan & Kondrashov, 2010). The selective advantage of duplicated genes can be due to increased gene dosage and/or gene neo-functionalisation. For example, duplicated and neo-functionalised genes might facilitate the detoxification of novel plant defence toxins and thus potentially expand the breadth of accepted host plant species (e.g. Heidel-Fischer et al., 2019; Wen et al., 2006).

Host specialisation, single or a few host plant species within one plant family, is most common among herbivorous insects. Whereas some herbivorous insects, including some of the most devastating pest species, are polyphagous meaning they are able to feed on a variety of plant species belonging to different families (Schoonhoven et al., 2005; Voelckel & Jander, 2014). Polyphagous species likely evolved and maintained detoxification mechanisms with a broad substrate specificity as a counter-response to the large variety of plant defence toxins, or specialised metabolites, they encounter (Heidel-Fischer & Vogel, 2015).

A general insect detoxification mechanism occurs via the three-step detoxification pathway for which a series of generally recognised gene families are involved (Brattsten, 1988; Kant et al., 2015). In the first step, P450 monooxygenases (P450s) and carboxyl- and cholinesterases (CCEs) make the plant toxin more hydrophilic. During the second step, UDP-glycosyltransferases (UGTs) and glutathione S-transferases (GSTs) conjugate the compounds to endogenous molecules increasing the polarity and hydrophilicity of the molecules even further, after which in the final third step membrane transporters like ATP-binding cassettes (ABCs) move the compounds for excretion (Dermauw et al., 2020; Feyereisen, 1999; Heidel-Fischer & Vogel, 2015; Kant et al., 2015; Voelckel & Jander, 2014). While monophagous species have often evolved specialised gene functions to target specific host plant defences (e.g. Fischer et al., 2008; Heidel-Fischer et al., 2019; Ratzka et al., 2002; Wheat et al., 2007; Wittstock et al., 2004), polyphagy has often been associated with gene family expansions. Genome studies of polyphagous arthropods show the occurrence of major gene family expansions of detoxification and digestion related families, for example, in Lepidoptera (Xu et al., 2016), Hemiptera (Chen et al., 2016), and Acari (Dermauw, Osborne, et al., 2013). In the fall armyworm, *Spodoptera frugiperda*, observed expansions were primarily due to tandem duplications forming a suggested adaptation mechanism to enable polyphagy (Gouin et al., 2017). Similarly, gene expansions have been linked to polyphagy for the cotton leafworm, *Spodoptera litura* (Cheng et al., 2017).

Gene family expansions are hypothesised to be causal for the emergence of polyphagy in Lepidoptera (Cheng et al., 2017; Gouin et al., 2017). Specific gene copies and functional diversity of Clan 3 P450 subfamilies have been linked to diet complexity in multiple Lepidoptera (Calla et al., 2017). This indicated a correlation between gene family dynamics (e.g. duplication and functionalisation level) and xenobiotic metabolism (Calla et al., 2017). Nonetheless, in a larger comparative study a relationship between detoxification gene family sizes (P450, CCE and GST) and feeding preference was found in multiple insect orders but not in the studied seven Lepidoptera species (Rane et al., 2016). However, studies on the association between gene family expansions and herbivory among a larger number of lepidopteran species has not yet been reported.

In this study, we apply a comparative phylogenomic approach using available high-quality lepidopteran genomes (37) spanning the lepidopteran tree of life in order to examine the evolution of gene family expansions associated with host plant use. The lepidopteran species vary in many characteristics such as feeding habit, host plant species range, specialised metabolite acceptance and pest status. Therefore, we examined the correlation between these herbivory-characteristics and genomic changes.

Gene family evolution, in terms of gene gain and loss rates, differed across lepidopteran families and were associated with ecological and evolutionary characteristics. Gene family expansions occur in both polyphagous and monophagous species. However, we found a significant positive correlation between the size of

the detoxification gene families CCE and GST with levels of polyphagy. In summary, detoxification gene family sizes are variable across monophagous and polyphagous Lepidoptera but expansions can be correlated to level of polyphagy for specific gene families.

## Materials and methods

### Data sources and quality assessment

We analysed 37 Lepidoptera genomes for which complete gene sets were available (on 09/2019) and included one outgroup represented by the sister clade Trichoptera. Annotation files and gene sets (protein translations) were downloaded from various databases, including Ensemble LepBase release v. 4 (Challi et al., 2016) and NCBI (Sayers et al., 2020). The included species, data sources and accession dates are reported in Supplementary Table 1 (All Supplementary Data is uploaded to the 4TU Centre for Research Data repository and available online: <https://figshare.com/s/0de53657df61d5e4d7c6>).

When genes were represented by multiple isoforms per gene (e.g. based on the sequence names), sequence files were edited using the Trinity based perl script ‘get\_longest\_isoform\_seq’ to ensure a single representative longest isoform. Completeness of genome gene sets were assessed using the *Insecta\_odb9* gene set, consisting of 1,658 Benchmarking Universal Single-Copy Orthologs (BUSCO) in BUSCO v. 3.0.2. (Simão et al., 2015). BUSCO results showing high duplication levels in the gene set could indicate the presence of a high number of isoforms. In case high duplication levels were found, we checked the full genome assembly for the degree of gene duplication to see if high duplication levels actually reflected true duplications. In case a large difference between the protein set and genome assembly was shown, we assumed multiple isoforms per gene were still present and assessed CD-HIT-EST v. 4.8.1. (Li & Godzik, 2006) using a 95% identity threshold. We applied CD-HIT-EST on *Heliconius melpomene melpomene*, *Heliconius erato demophoon*, *Leptidea sinapis* and *Heliothis virescens*.

### Functional annotation and orthology prediction

Peptide sequences were cleaned of diverse characters like “\*” and “:”, to avoid the use of illegal characters for the annotation analysis (e.g. InterProScan). We used InterProScan v. 5.36-75 (-appl Pfam --goterms) (Jones et al., 2014) for general annotation and identification of protein families. Further, we ran a local BLASTP v. 2.6.0 (Camacho et al., 2009) against the UniRef50 database (<ftp.uniprot.org/pub/databases/uniprot/uniref/uniref50/uniref50.fasta.gz>; release version 31/07/2019, accessed 20/08/2019) (UniProt Consortium, 2019) using a cut-off e value of 1e-3.

The annotated proteins using InterProScan and local BLASTP were used to retrieve gene counts for the gene families of interest. Further, OrthoFinder v. 2.2.7 (Emms & Kelly, 2015) was used to predict orthologous protein groups (OGs). An OG is a group of genes descended from a single gene in the last common ancestor of a group of species. The protein sequence files were used as input and OrthoFinder was run under default settings. We used the resulting orthologous protein groups as input for CAFE v. 4.2.1 (De Bie et al., 2006; Hahn et al., 2005). Since we focused on various gene families involved in plant feeding, we selected candidate OGs based on the BLASTP and InterProScan identifications. We selected OGs of gene families of interest if genes matched one of the Uniref50 cluster terms, Pfam families or InterProScan identifiers specific for each gene family (Supplementary Table 2). The gene families of interest were: P450 monooxygenases (P45Os), carboxyl- and cholinesterases (CCEs), UDP-glycosyltransferases (UGTs), glutathione S-transferases (GSTs), ATP-binding cassettes (ABCs), trypsin, and the insect cuticle protein family.

### Time-calibrated species phylogeny

The CAFE analyses (Computational Analysis of gene Family Evolution) required an ultrametric phylogeny of the Lepidoptera. We used the protein sequences of single-copy BUSCO genes to generate alignments of orthologous genes. All 1,367 Single-copy BUSCO proteins were individually retrieved for every species on the amino acid (aa) level using BUSCO v. 4.0.5., applying the *insecta\_odb10* as a reference lineage dataset (Simão et al., 2015). Individual alignments were generated for every BUSCO-identified ortholog using MAFFT v. 7.305 (Katoh & Standley, 2013) using the L-INS-i algorithm. For the identification of putative ambiguously aligned or randomised multiple sequence alignment (MSA) sections, we used Aliscore v. 1.2 (Kück et al., 2010; Misof & Misof, 2009) on each MSA with the default sliding window size, the maximal number of pairwise sequence comparisons and a special scoring for gap-rich amino acid data (options -r and -e). After exclusion of the identified putative ambiguously aligned or randomised MSA sections with ALICUT v. 2.3 (Kück et al., 2010), the final MSAs were concatenated into a supermatrix using FASconCAT-G v. 1.02 (Kück & Longo, 2014). The resulting dataset comprised 1,367 gene partitions and 687,494 amino acid positions.

Prior to the tree reconstruction, the best scoring amino acid substitution matrix for each gene partition was selected with ModelFinder as implemented in IQ-TREE v. 1.6.12 (Kalyaanamoorthy et al., 2017). We restricted the search of the best fitting model to eight amino acid substitution matrices appropriate for nuclear markers: DCMut (Kosiol & Goldman, 2005), JTT (Jones et al., 1992), LG (Le & Gascuel, 2008), Poisson, PMB (Veerassamy et al., 2003), VT (Muller & Vingron, 2000), and WAG (Whelan & Goldman, 2001). We additionally included the protein mixture model LG4X (Le et al., 2012), which accounts for FreeRate heterogeneity. Furthermore, we allowed testing the default rate heterogeneity types (E, I, G, I+G, and FreeRates: R) (Gu et

al., 1995; Soubrier et al., 2012; Yang, 1994), with or without empirical rates (-F, -FU) as well as testing the number of rate categories (-cmin 4 -cmax 15). The best model for each gene partition was selected according to the best second-order or corrected Akaike Information Criterion (AICc) score (Hurvich & Tsai, 1989). Dataset and partition scheme including selected models can be found at the 4TU Centre for Research Data repository available online: <https://figshare.com/s/0de53657df61d5e4d7c6> (reserved DOI: 10.4121/14529360).

Phylogenetic relationships were inferred under the maximum likelihood (ML) optimality criterion as implemented in IQ-TREE v. 1.6.12 (Chernomor et al., 2016; Nguyen et al., 2015) using the best scoring amino acid substitution matrix for each gene partition and the edge-proportional partition model allowing partitions to have different evolutionary rates (option -ssp). We performed 50 independent tree searches (25 searches with a random and 25 with a parsimony start tree). The resulting number of unique tree topologies was assessed with Unique Tree v. 1.9, kindly provided by Thomas Wong and available upon request. We used the ML tree with the best log-likelihood to obtain an ultrametric tree using the chronos function of the R package ape v. 5.4 on R v. 3.6.3, relaxed model (Paradis et al., 2004; R Development Core Team, 2020). The tip to root length was adjusted to match the approximately 299.5 million-year evolutionary history of crown group Lepidoptera (Kawahara et al., 2019).

## Ecological host data and diversity metrics

For each lepidopteran species we collected host plant specialisation level, host plant family and species range, pest status and specialised metabolite content within the accepted host plant range. Data was collected by browsing literature for host plant species accepted by each of our butterflies and moths studied. We used this information to determine for each lepidopteran species the range of host plant acceptance, and subsequently classified diet breadth (level of polyphagy or monophagy; Supplementary Table 3). Additionally, we recorded pest status of the lepidopteran species if the species was a described pest in the literature searched or if included in the EPPO or CABI databases (CABI, 2020; EPPO, 2019). We considered all lepidopteran species that accept host plant species from a single plant family to be a monophagous species. Species recorded as polyphagous were those feeding on species from multiple plant families.

To quantify the phylogenetic diversity of a particular lepidopteran species' diet, we calculated the Faith's measure of phylogenetic diversity (PD) (Faith, 1992). To calculate the PD for each range of host plant families we used the package Picante v. 1.8.2 (Kembel et al., 2010) in R v.3.6.2 (R Development Core Team, 2020). This metric quantifies the degree of host plant range diversity by calculating the distance between plant families according to branch lengths of a reference phylogeny. As a reference phylogeny, we used the recent angiosperm phylogeny of Ramírez-Barahona et al. (2020), pruned for lepidopteran host plant families. Calculated PD values were scaled so that monophagous species had a PD = 1 (all PDs divided by 374.14, the value for

single gene family acceptance). Two families included in our list of recorded host plant families, Aspleniaceae and Araucariaceae, were missing in the reference phylogeny. These plant families are hosts for only two highly polyphagous species in our analysis (*S. frugiperda* (Aspleniaceae) and *S. exigua* (Araucariaceae)). Thus, we expected that their exclusion would have a very small impact on the dietary PD. Accordingly, we removed Aspleniaceae from our dataset while we replaced Araucariaceae by Cupressaceae, the next most closely related family in our reference phylogeny.

We compiled reported specialised metabolites for each host plant family. We collected information for the three main groups of secondary metabolites, as classified in Schoonhoven et al. (2005): phenolics, terpenoids, and nitrogen-containing compounds. For each host plant family, we recorded the secondary metabolite type, chemical class, subclass and, if present, any additional sub-level (Supplementary Table 4). Metabolites belonging to the same chemical type or class are by definition more similar. Thus, we used a hierarchical structure to calculate the degree of ‘functional metabolite diversity’ (FMD) of specialised metabolites encountered by the lepidopteran species in their range of accepted host plant families. Plant families with similar chemical compositions are likely to be detoxified by similar mechanisms. Lepidopteran species feeding on plants with diverse specialised metabolites will thus have a higher value for the FMD (e.g. polyphagous species).

We used the database of specialised metabolite records per plant family to create a trait matrix (Supplementary Table 5), the first step to calculate a functional diversity index (Petchey & Gaston, 2002, 2006). Afterwards we calculated the dietary FMD of each lepidopteran species with a dendrogram-based method using the script by J. Schumacher and O. Petchey (accessed 02/2021; [http://github.com/opetchey/dumping\\_ground/tree/master/functional\\_diversity/FD.example.2](http://github.com/opetchey/dumping_ground/tree/master/functional_diversity/FD.example.2)) as described in Petchey and Gaston (2002). As a consequence of the dendrogram-based calculation method, the FMD could only be calculated for polyphagous species because of the range of accepted metabolites. Measures of PD and FMD could not be calculated for the Indian meal moth, *Plodia interpunctella*, because this species feeds exclusively on dried products such as stored and processed food, and thus the influence of specialised metabolites is limited.

We calculated a Spearman rank correlation coefficient to examine the correlation between degree of polyphagy, using the PD and FMD metrics, and gene counts of gene families involved in plant feeding. Specifically, we used the gene counts of plant detoxification related gene families (P450, CCE, UGT, GST and ABC) and the trypsin and insect cuticle protein families. Correlation analyses of gene family counts (Supplementary Table 6) and both PD and FMD (Supplementary Table 7 and 8) were analysed. Correlation statistics were calculated using the function ‘cor.test’ in the package Stats v. 3.6.2 in R v. 3.6.2 (R Development Core Team, 2020). *Spodoptera frugiperda* is represented in our dataset by both the rice and corn strain, belonging to the same species. Therefore, we additionally tested the correlation significance when only a single *S. frugiperda* strain (rice population, with the lowest gene counts) was included.



## CAFE analysis

We used Computational Analysis of gene Family Evolution, CAFE v. 4.2.1 (De Bie et al., 2006; Hahn et al., 2005) to analyse gene family evolution (gene gains and losses) in a phylogenetic context. CAFE uses a birth and death process to model gene gain and loss across a ultrametric phylogenetic tree. Based on the results of OrthoFinder, gene counts per species were used as input for the CAFE analyses.

Gene families that have large variance in gene copy numbers across species can cause the parameter calculations to be non-informative (CAFE tutorial documentation v. 20 January 2016). From a computational perspective filtering out high variance OGs is needed in order to let the statistical analyses reach saturation. Therefore, the gene count dataset as derived from the OrthoFinder run was filtered for OGs with high variance levels. We filtered out all OGs which showed  $\geq 100$  copies, as recommended by the developers (CAFE tutorial documentation v. 20 January 2016) in at least one species. After testing if CAFE reached convergence with multiple thresholds, we finally removed the top 2% OGs with highest variance. This resulted into the dataset including OGs comprising all gene families, called hereafter the *all gene families dataset*. Based on this dataset we calculated the error model because errors in genome assembly and gene annotation potentially result in biased evolutionary rate estimates (Han et al., 2013). We used *cafererror.py*, as provided in CAFE, to calculate the error in our dataset due to assembly and annotation mistakes. The method accounts for errors by modelling the uncertainty of observed family sizes at the tips of the tree (Han et al., 2013). The resulting model based on the *all gene families dataset* was used in all CAFE runs analysing the datasets as described below.

We generated one additional dataset which was filtered for OGs belonging to five specific gene families involved in specialised metabolite detoxification: P450s, CCEs, UGTs, GSTs and ABCs, called hereafter *5 gene families dataset*. In addition, we applied CAFE on selected *single gene family datasets* to study rates of change per gene family. These selected single gene families were the five detoxification families (P450, CCE, UGT, GST and ABC) and two additional families potentially important for (polyphagous) herbivory: the insect cuticle protein family and the trypsin gene family. These datasets are called hereafter *single gene family datasets*.

The CAFE runs included the calculation of the single rate of change ( $\lambda$ ), and a second mode where gains ( $\lambda$ ) and losses ( $\mu$ ) were calculated separately. The p-value threshold was kept at the default value (0.01), and the top 2% high variance OGs were removed in all datasets in order to let CAFE reach saturation. Multiple runs of CAFE were used to test for convergence. To reach convergence, only a single rate of change ( $\lambda$ ) was calculated for the *single gene family datasets* while both gain- ( $\lambda$ ) and loss- ( $\mu$ ) rates were calculated for the *all gene families dataset* and *5 gene families dataset*. In all the analyses the trichopteran outgroup, *Limnephilus lunatus*, was treated as a separate group calculating an individual  $\lambda$  and  $\mu$ . We similarly treated *Maniola jurtina*, the meadow brown, as a separate group within Nymphalidae because it had a higher number of predicted genes than the other species (*M. jurtina*: 36,294, Nymphalidae

average excluding *M. jurtina*: 17,554) and thus expected to have a different rate of change.

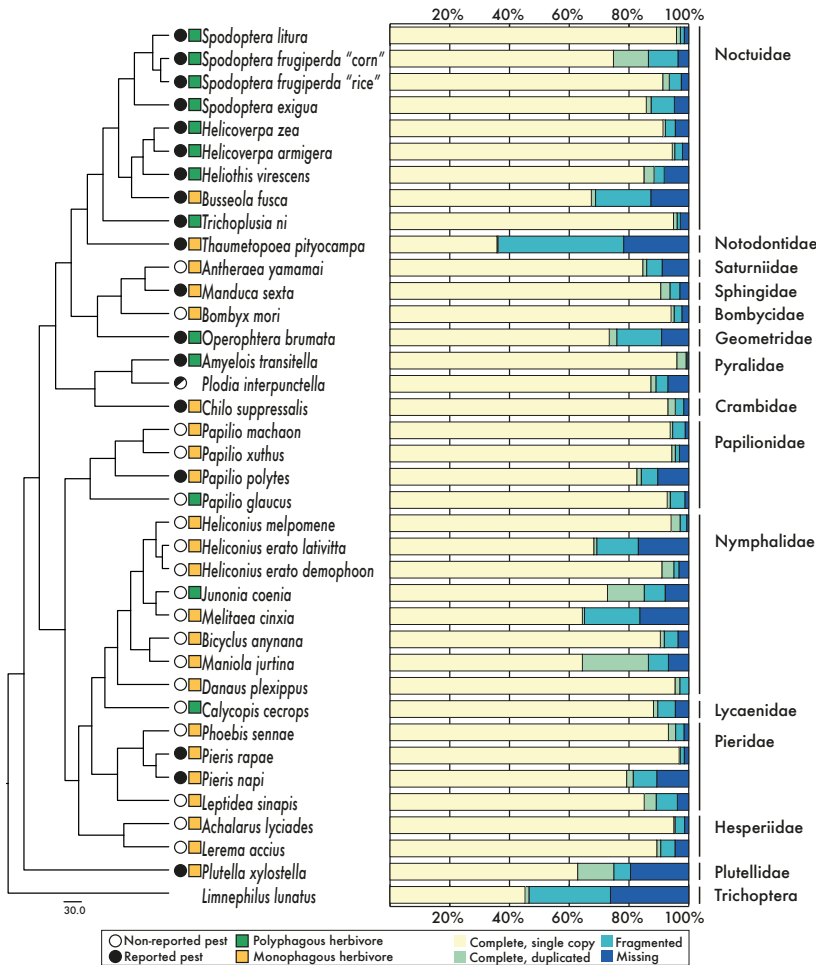
Single expansion and contraction rates based on the entire phylogeny were only calculated for the *all gene families dataset*. In addition, we also ran several analyses calculating separate rates of change for specific clades in the tree to address the evolution of polyphagous herbivory. Specifically, we analysed four target lepidopteran families for which more than two species were included in the dataset: cutworm moths (Noctuidae), swallowtails (Papilionidae), brush-footed butterflies (Nymphalidae) and whites (Pieridae). For these four families expansion and contraction rates were calculated using the *all gene families dataset*, *5 gene families dataset* and the *single gene family datasets*.

## Results

### Genomes, gene families and species tree reconstruction

The average number of protein-coding sequences in the 37 Lepidoptera and the one Trichoptera outgroup genome was 17,589 genes and ranged from 12,240 to 29,415 per species (Supplementary Table 1; Table 1). Based on the BUSCO analyses, the majority of species (85%) had a completeness of >75% with an average completeness of 86.8% (Figure 1). The number of functionally annotated protein sequences from InterProScan ranged from 10,723 to 32,131 (Supplementary Table 9) and from BLASTP against the UniRef50 database from 13,279 to 40,328 (Supplementary Table 10). We calculated the gene number of various herbivory related gene families (P450s, CCEs, UGTs, GSTs, ABCs, trypsins and insect cuticle proteins; Figure 2; Supplementary Table 6) based on InterProScan and Uniref50 identifiers (Supplementary Table 2).

OrthoFinder identified 21,610 orthologous groups (OGs) (Supplementary Table 11). The gene count datasets used in the CAFE analyses were derived from this OrthoFinder run (Supplementary Table 12) after filtering for high variance groups. The *all gene families dataset* consisted of 21,148 OGs (Supplementary Table 13) and the *5 gene families dataset* consisted of 574 OGs (Supplementary Table 14). The *single gene family datasets* consisted of 197 OGs for the P450 gene family, 148 OGs for CCE, 64 OGs for UGT, 32 OGs for GST, 154 OGs for ABC, 383 OGs for trypsin and 203 OGs for the insect cuticle gene family (Supplementary Table 15). The species phylogeny was constructed using the protein sequences of 1,367 single-copy and complete BUSCO genes (Figure 1, left). The 50 independent ML tree searches returned one unique tree topology. Our phylogeny contained six lepidopteran superfamilies of which four consisted of more than one species, and 14 families of which six consisted of more than one species. This resulted in a Lepidoptera-clade wide species representation which is consistent with the comprehensive phylogeny by Kawahara et al. (2019).



**FIGURE 1** | ML tree topology based on 1,367 single copy BUSCOs from 37 lepidopteran genomes. Species pest status and feeding style is given, discriminating between monophagous and polyphagous species (Supplementary Table 3). Feeding style is not provided for *Plodia interpunctella*, since this species feeds on dried products. For every species the taxonomic family is given (right). Stacked bar graphs present the BUSCO quality assessment of the genome gene sets used in this study.

## Ecological host data and diversity metrics

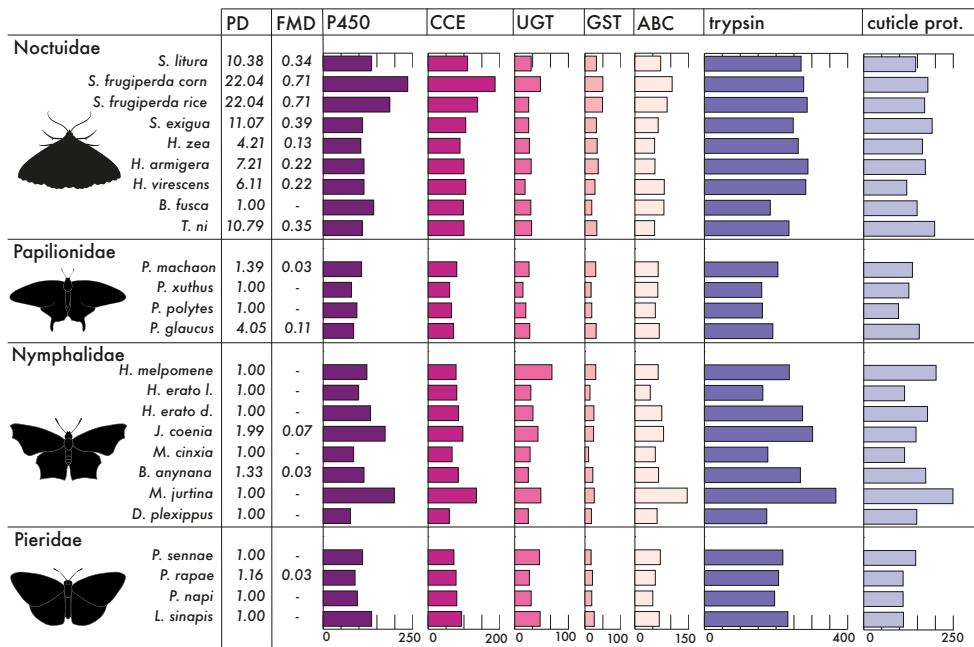
Based on host plant family range per lepidopteran species (Supplementary Table 3), we calculated a phylogenetic diversity (PD) index (Supplementary Table 7, Figure 2). The scaled PDs ranged from 1, for monophagous species feeding on host(s) within a single plant family, to 22.04 for the major polyphagous species *S. frugiperda*, feeding on hosts from 74 different plant families.

Our dataset of specialised metabolite content per host plant species consisted of 3,831 entries, and is based on ~750 literature sources (Supplementary Table 4; on plant family level). Further, based on the host plant acceptance range per lepidopteran species we calculated a ‘functional metabolite diversity’ (FMD) index (Supplementary Table 8, Figure 2). The FMD as calculated for all polyphagous species ranged from 0.03 for *P. machaon*, able to metabolise the specialised metabolites from host plant species within Apiaceae and Rutaceae, to 0.71 for *S. frugiperda*.

## Gene family expansions and contractions

We calculated the gene counts of the seven target gene families involved in plant feeding (P450, CCE, UGT, GST, ABC, trypsin and insect cuticle) for all 38 genomes. The gene family sizes varied across the species with P450 family ranging from 236 genes in *S. frugiperda* to 73 in *Chilo suppressalis*, CCE ranging from 187 in *S. frugiperda* to 59 in *Danaus plexippus*, UGT from 104 in *H. melpomene* to 23 in *Papilio xuthus*, GST from 50 in *S. frugiperda* to 10 in *Melitaea cinxia*, ABC from 146 in *M. jurtina* to 43 in *H. erato lativitta*, trypsin from 367 in *M. jurtina* to 140 in *Thaumetopoea pityocampa*, and finally the insect cuticle protein family with 249 genes in *M. jurtina* to 97 genes in *Papilio polytes* (Supplementary Table 6). We focussed on the gene counts of these gene families within the four focal lepidopteran families.

The butterfly families Noctuidae, Papilionidae, Nymphalidae and Pieridae differed in herbivorous traits and range of polyphagy and each was represented by at least >2 species. The average number of total gene counts for the seven gene families was greatest for the Noctuidae (831.56; Table 1, Figure 2), in concordance with the widest range of accepted host plants (PD, ranging between 1 and 22.04, and FMD, ranging between 0.13 and 0.71; Figure 2). The Noctuidae also had the highest average number of genes when only the five detoxification gene families were included (406.67), with the second largest number of genes found for the Nymphalidae (773.88 and 363.13). However, a high gene count can be the result of an overall larger number of predicted genes dependent on the quality of the genome annotation. Therefore, we normalised the number of genes from the target gene families using the percentage of the total number of predicted genes (Table 1). Again, the size of the gene families was highest in Noctuidae (4.68%) but the order of the second largest shifted to Pieridae (4.32%). However, the differences were small and disproportionate to the differences in PD and FMD values (Figure 2), which reflect the level of polyphagy.

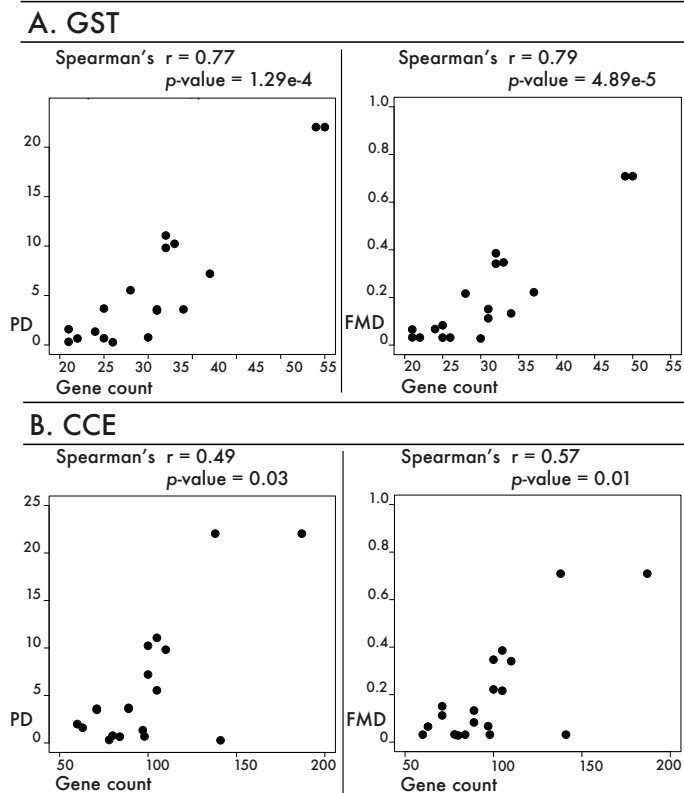


**FIGURE 2** | Graph showing the gene counts of seven gene families for four lepidopteran families: Noctuidae, Papilionidae, Nymphalidae and Pieridae. Gene families include five families involved in metabolite detoxification: P450 monooxygenase (P450), carboxyl- and cholinesterase (CCE), UDP-glycosyltransferase (UGT), glutathione S-transferase (GST) and ATP-binding cassette (ABC), one family involved in digestion: trypsin and one family putatively involved in protection of the insect midgut: insect cuticle protein. PD values represent the phylogenetic diversity of the host plant families within each lepidopteran diet. The PD values are scaled, with 1 being monophagous and >1 polyphagous species. FMD values show the functional metabolite diversity of metabolites employed by the host plant families within the lepidopteran diets.

We calculated the correlation and level of significance between the PD values and gene counts of the seven target gene families across polyphagous Lepidoptera (Supplementary Table 16). There was a significant positive correlation between gene counts of the detoxification gene families CCE ( $r = 0.49$ ,  $p$ -value = 0.03) and GST ( $r = 0.77$ ,  $p$ -value =  $1.29 \times 10^{-4}$ ) in polyphagous species and the level of polyphagy as represented by the phylogenetic diversity index (Figure 3A). The GST gene family was also significant positively correlated in the additional tests when only the single *S. frugiperda* rice strain was included (Supplementary Table 16).

**TABLE 1** | Shown is an overview of four major Lepidoptera families (Noctuidae, Papilionidae, Nymphalidae and Pieridae): number of predicted genes based on the genome annotations; number of annotated detoxification genes (from families P450, CCE, UGT, GST and ABC); the number of annotated detoxification genes as percentage of the predicted genes; number of annotated detoxification, trypsin, and insect cuticle genes; and the number of annotated detoxification, trypsin and insect cuticle genes as percentage of the predicted genes. Further, the averages for each Lepidoptera family are given.

Family	Species	Predicted genes	Annotated detoxification genes	% Detox. genes	Annotated detoxification, trypsin and cuticle genes	% Detox., trypsin and cuticle genes
Noctuidae	<i>Spodoptera litura</i>	15,317	395	2.58	810	5.29
	<i>Spodoptera frugiperda</i> “corn”	21,779		2.98	1105	5.07
	<i>Spodoptera frugiperda</i> “rice”	26,356	502	1.9	959	3.64
	<i>Spodoptera exigua</i>	18,477	351	1.9	790	4.28
	<i>Helicoverpa zea</i>	15,128	324	2.14	750	4.96
	<i>Helicoverpa armigera</i>	17,082	353	2.07	814	4.77
	<i>Heliothis virescens</i>	15,099	358	2.37	761	5.04
	<i>Busseola fusca</i>	19,417	384	1.98	717	3.69
Average per family		14,384	344	2.39	778	5.41
Papilionidae			406.67	2.26	831.56	4.68
	<i>Papilio machaon</i>	15,497	322	2.08	663	4.28
	<i>Papilio xuthus</i>	13,102	243	185	529	4.04
	<i>Papilio polytes</i>	12,244	266	2.17	525	4.29
Average per family		15,692	297	1.89	643	4.1
Nymphalidae			282	2	590	4.18
	<i>Heliconius melpomene</i>	20,075	399	1.99	838	4.17
	<i>Heliconius erato lativitta</i>	14,613	281	1.92	558	3.82
	<i>Heliconius erato demophoon</i>	14,517	368	2.53	820	5.65
	<i>Junonia coenia</i>	19,234	439	2.28	887	4.61
	<i>Melitaea cinxia</i>	16,667	262	1.57	553	3.32
	<i>Bicyclus anynana</i>	22,642	324	1.43	765	3.38
	<i>Maniola jurtina</i>	36,294	579	1.6	1195	3.29
Average per family		15,130	253	1.67	575	3.8
Pieridae			363.13	1.87	773.88	4.01
	<i>Phoebis sennae</i>	16,493	339	2.06	703	4.26
	<i>Pieris rapae</i>	13,188		2.17	603	4.57
	<i>Pieris napi</i>	13,622	290	2.13	596	4.38
Average per family		18,049	393	2.18	737	4.08
Average per family			327	2.13	659.75	4.32



**FIGURE 3** | Scatterplots showing the distribution of gene counts of GST genes (A, x-axes) or CCE genes (B, x-axes) against the phylogenetic diversity (PD) index values (y-axes, left) or functional metabolite diversity (FMD) index values (y-axes, right) of all polyphagous Lepidoptera species. The Spearman correlation coefficient and  $p\text{-value}$  is given above each plot.

Further, we also calculated the correlation coefficient and level of significance between the FMD and gene counts of the seven target gene families (Supplementary Table 16). Again, there was a significant positive correlation between the gene counts of gene families CCE ( $r = 0.57$ ,  $p\text{-value} = 0.01$ ) and GST ( $r = 0.79$ ,  $p\text{-value} = 4.89\text{e-}5$ ) in polyphagous species and the level of polyphagy as represented by the FMD (Figure 3B). Further, both CCE and GST gene families remained positively correlated when only the single *S. frugiperda* rice strain was included (Supplementary Table 16). Finally, in all cases (both for PD and FMD values) we tested for correlation significance of the seven gene families as fraction of the total number of annotated genes. The GST gene family was significant positively correlated for all analyses (Supplementary Table 16).

## Gene family evolution

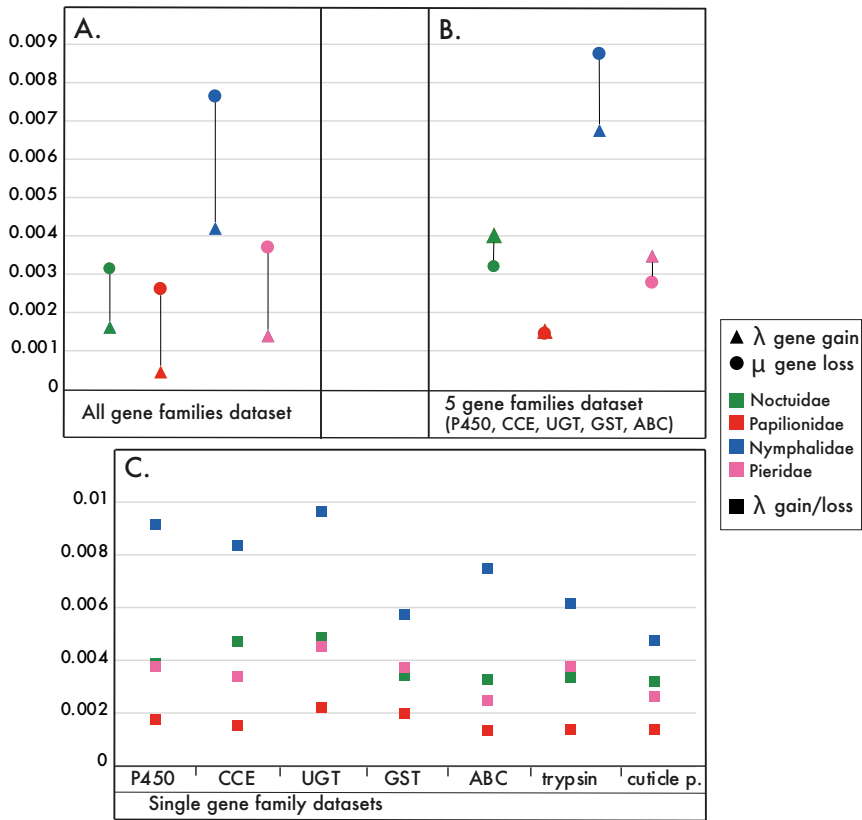
The analyses of gene family expansions and contractions using CAFE and inclusion of all gene families, using the *all gene families dataset*, resulted in an overall rate of change,  $\lambda$ , of 0.0023 (likelihood score -641908; Supplementary Table 17). Calculating a distinct rate for gene gain,  $\lambda = 0.0015$  gain/gene/million years, and gene loss,  $\mu = 0.0032$  loss/gene/million years, resulted in a greater likelihood score (-628685; Supplementary Table 17) and thus was preferred over calculating a single rate of change (Hahn et al., 2005).

We associated gene expansion and contraction rates with the ecology and herbivorous characteristics for the four lepidopteran families, Noctuidae, Papilionidae, Nymphalidae and Pieridae, separately. The  $\lambda$  (gain) and  $\mu$  (loss) values calculated when all gene families were included, using the *all gene families dataset*, showed a higher rate for gene loss for all butterfly families (Figure 4A; Supplementary Table 17). Both  $\lambda$  and  $\mu$  rates were highest for Nymphalidae compared to the other families, with the rate of gene loss ( $\mu = 0.0076$ ), almost twice as large as the highest second value ( $\mu = 0.0036$ ) for Pieridae (Figure 4A).

The gene gain and loss rates by inclusion of only the five detoxification gene families (P450, CCE, UGT, GST and ABC), using the *5 gene families dataset*, was again highest for Nymphalidae compared to the other families, with a higher rate for gene loss ( $\lambda = 0.0067$ ,  $\mu = 0.0087$ ). Papilionidae had a similar rate for  $\lambda$  (0.0015) and  $\mu$  (0.0014) while both Noctuidae ( $\lambda = 0.0040$ ,  $\mu = 0.0032$ ) and Pieridae ( $\lambda = 0.0035$ ,  $\mu = 0.0028$ ) showed a higher rate for gene gain over gene loss (Figure 4B; Supplementary Table 17).

Finally, the single rate of change ( $\lambda$ ) as calculated for each of the seven gene families (including the trypsin and cuticle protein families), using the *single gene family datasets*, differed across the Lepidoptera families. The calculated  $\lambda$  was consistently highest for the Nymphalidae (P450  $\lambda = 0.0091$ , CCE  $\lambda = 0.0083$ , UGT  $\lambda = 0.0096$ , GST  $\lambda = 0.0057$ , ABC  $\lambda = 0.0075$ , trypsin  $\lambda = 0.0061$ , insect cuticle  $\lambda = 0.0047$ ), while Papilionidae (P450  $\lambda = 0.0017$ , CCE  $\lambda = 0.0015$ , UGT  $\lambda = 0.0022$ , GST  $\lambda = 0.002$ , ABC  $\lambda = 0.0013$ , trypsin  $\lambda = 0.0013$ , insect cuticle  $\lambda = 0.0014$ ) had the lowest rate of change for all studied gene families. Both Pieridae (P450  $\lambda = 0.0037$ , CCE  $\lambda = 0.0033$ , UGT  $\lambda = 0.0045$ , GST  $\lambda = 0.0037$ , ABC  $\lambda = 0.0024$ , trypsin  $\lambda = 0.0037$ , insect cuticle  $\lambda = 0.0026$ ) and Noctuidae (P450  $\lambda = 0.0038$ , CCE  $\lambda = 0.0047$ , UGT  $\lambda = 0.0048$ , GST  $\lambda = 0.0034$ , ABC  $\lambda = 0.0032$ , trypsin  $\lambda = 0.0033$ , insect cuticle  $\lambda = 0.0032$ ) showed similar  $\lambda$ s for most gene families but for CCE, ABC and the insect cuticle protein family the difference in rate of change was larger (Figure 4C; Supplementary Table 17).





**FIGURE 4** | Estimates of gene family evolution rates as calculated with CAFE. The parameters are calculated for the four lepidopteran families Noctuidae, Papilionidae, Nymphalidae and Pieridae. Rates for gene loss (circles, loss/gene/million years,  $\mu$ ) and gene gain (triangles, gain/gene/million years,  $\lambda$ ) calculated for: A) *all gene families dataset*; and B) *5 gene families dataset*, which include the detoxification gene families P450 monooxygenase (P450), carboxyl- and cholinesterase (CCE), UDP-glycosyltransferase (UGT), glutathione S-transferase (GST) and ATP-binding cassette (ABC). Single rates of change (squares, either gain or loss /gene/million years,  $\lambda$ ) calculated for: C) *single gene family datasets* of the five main detoxification gene families, and trypsin and insect cuticle protein families.

## Discussion

In this study, we evaluated if gene family expansions are correlated with polyphagy across Lepidoptera. We examined and associated genomic data of 37 lepidopteran genomes with the range of accepted host plants and their specialised metabolite contents. Specifically, we investigated gene family repertoires and expansion and contraction rates of gene families putatively involved in metabolite detoxification and digestion.

## Lepidopteran phylogenomic framework

Predictions on gene and genome evolution across a range of species depends on the robustness and accuracy of the species phylogeny. Our targeted phylogenetic reconstruction of lepidopteran species with completed genomes (Figure 1) was consistent with the comprehensive phylogeny by Kawahara et al. (2019). Further, the compared lepidopteran genomes should be of similar quality and completeness to avoid biases. The mean number of predicted proteins was 17,590 (SD = 4,785.73) which falls within the expected range of insect genomes (Waterhouse, 2015). For a few species the number of reported predicted proteins was higher than the average. For example, 29,415 proteins in the pine processionary moth (*Thaumetopoea pityocampa*) (Gschloessl et al., 2018) and 36,294 predicted proteins in the meadow brown butterfly (*M. jurtina*) (Singh et al., 2020). However, this difference was reduced due to the selection of the 21,610 orthogroups, excluding ungrouped and unplaced sequences, specific sub-selections of particular gene families, and selection and focus on specific lepidopteran families.

## Gene evolution in Lepidoptera

Using our lepidopteran phylogenomic framework and inclusion of all gene families, we estimated an overall rate of change,  $\lambda$ , of 0.0023 (gains/losses per million years). Our estimate was consistent with gene turnover estimates of other insect clades including *Drosophila* ( $\lambda = 0.0012$ ; (Hahn et al., 2007)) and *Anopheles* ( $\lambda = 0.0031$ ; (Neafsey et al., 2015)), and other taxa, such as yeast ( $\lambda = 0.002$ ; (Hahn et al., 2005)) and mammals ( $\lambda = 0.0016$ ; (Demuth et al., 2006)). When we calculated a separate value for gene gain and loss, the overall loss rate ( $\mu = 0.0032$ ) was higher than the gene gain rate ( $\lambda = 0.0015$ ). This individual rate for gene gain ( $\lambda$ ) was similar to the single estimated parameter for gene gain/loss calculated in Lepidoptera based on five genomes in a recent study ( $\lambda = 0.0014$ ) (Thomas et al., 2020). Both of our calculated turnover estimates were close to the general rates in other taxa but the difference in  $\lambda$  and  $\mu$  are larger than in estimates of beetles, Coleoptera ( $\lambda = 0.0019$ ,  $\mu = 0.0018$ ) (Seppey et al., 2019). This shows a higher rate of gene loss over gene gain within Lepidoptera. Indeed, gene loss can be seen as an important aspect in the evolution of species in terms of adaptive and/or neutral evolution (Albalat & Cañestro, 2016).

It has been suggested that in the ancestry of Lepidoptera a large-scale genome duplication event occurred (Li et al., 2018). This was proposed to be an ancient event before the radiation of Lepidoptera (< 300 million years ago), and likely followed by genome fractionation which reduces genome size. Indeed, the genome size of extant Lepidoptera is similar to other insect orders (Gregory, 2020; Hanrahan & Johnston, 2011; Triant et al., 2018). Moreover, in a recent study on gene content evolution in Arthropoda, the common ancestor of Lepidoptera had the highest number of emergent gene families in comparison to all other insect clades (Thomas et al., 2020). Genome fraction (gene loss) and diploidisation after the duplication event in the lepidopteran

ancestor could have resulted in the loss of a large number of dispensable genes (Albalat & Cañestro, 2016). This would result in a higher rate for gene loss compared to gene gain as shown by our analyses ( $\mu = 0.0032 > \lambda = 0.0015$ ).

## Gene family expansions and contractions in four lepidopteran families

We further focused on the expansion rates within four lepidopteran families. The cutworm moths (Noctuidae) are a large cosmopolitan and species rich radiation of moths of which many species are major polyphagous herbivores (Regier et al., 2017; van Nieukerken et al., 2011). Numerous major polyphagous species, such as the genus *Spodoptera*, are considered notorious pests causing significant agricultural damage worldwide (Cho et al., 2008; Kalleshwaraswamy et al., 2018; Pogue, 2002; Stacke et al., 2018). All Noctuid species included in our dataset are reported pests and all except of the maize stalk borer, *Busseola fusca*, have a polyphagous feeding habit (Figure 1, Supplementary Table 3). The other families included in our comparison (Papilionidae, Nymphalidae and Pieridae) primarily consist of monophagous, non-pest reported species (Figure 1). The majority of the species within these three families feed on a narrow host plant range, as indicated by the low PD and FMD values (Supplementary Table 3; Figure 2). For Papilionidae only 21% of the 281 species have a polyphagous feeding habit, accepting more than one plant family (Scriber et al., 1991), while the Pieridae primarily feed on a restricted range of plants within three Angiosperm orders: Fabales, Santalales, and Brassicales (Braby & Trueman, 2006). Within Nymphalidae, major polyphagous species do occur, but most species have a limited host plant range (Celorio-Mancera et al., 2016; Nylin et al., 2014).

The Nymphalidae show high dynamic genome evolution rates when looking at rates of gene gains and losses, as calculated using all different datasets, in comparison to the other families (Figure 4). This is consistent with Nylin et al. (2014) who found that polyphagy in Nymphalidae was transient and that selection favoured the host plant specialisation with similar specialised metabolites. This oscillation of host plant ranges, termed the ‘oscillation hypothesis’, may contribute to phytophagous insect diversification when ancestral specialists give rise to plastic generalists that in turn adapt, diversify and again specialise (Janz & Nylin, 2008; Janz et al., 2006). When host specialisation is currently selected for and becomes more dominant in Nymphalidae, than this may be associated with the higher rate for gene family contraction (Figure 4).

We specifically looked at five gene families involved in detoxification of specialised metabolites (Heidel-Fischer & Vogel, 2015; Kant et al., 2015; Schuler, 2011), and compared the expansion and contraction rates, calculated using the *5 gene families dataset*, across the lepidopteran families. In contrast to Nymphalidae, the Noctuidae and Pieridae both had a higher rate for gene gain (Figure 4B). The overall PD, and FMD values of Noctuidae indicate a wider range of accepted plant families in comparison to the other lepidopteran families (PD (1 - 22.04) and FMD (0.13 - 0.71), Figure 2). The higher expansion rates of detoxification gene families in Noctuidae suggests a

correlation between their expansion and the evolution of polyphagy. However, the higher expansion rate in monophagous Pieridae might indicate that expansions are not exclusive to major polyphagous lineages. The lower PD (1 - 1.16) and FMD (0.03) values in Pieridae (and the overall high occurrence of monophagous species) does not explain the higher rate for gene family expansion given their restricted host ranges (Braby & Trueman, 2006). Indeed, gene duplications occur in all organisms and can result in selective advantages due to sub-functionalisation and/or neo-functionalisation (He & Zhang, 2005; Heidel-Fischer et al., 2019; Rastogi & Liberles, 2005). For example, in Pierinae (a subfamily within Pieridae) gene duplication followed by neofunctionalisation resulted in the evolution of the nitrile-specifier protein involved in the detoxification of glucosinolates produced by Brassicaceae plants (Fischer et al., 2008; Wheat et al., 2007; Wittstock et al., 2004).

Further, we looked at the rate of change ( $\lambda$ ) of individual gene families involved in detoxification and digestion, calculated using the *single gene family datasets* (Figure 4C). Besides the selected five detoxification families we added the trypsin and insect cuticle protein gene families. Trypsin, a family of serine proteases is involved in the hydrolyses of proteins and plays a role in the digestion of plant material in herbivorous invertebrates (Muhlia-Almazán et al., 2008; Rawlings & Barrett, 1994). The insect cuticle protein family, involved in formation of the exoskeleton, is suggested to play a role in increased protection of the peritrophic matrix and midgut, forming a physical barrier for biochemical toxins (Agrawal et al., 2014; Hegedus et al., 2009; Kelkenberg et al., 2015; Kumar et al., 2018). The distribution of the calculated  $\lambda$ s between the Lepidoptera families is similar for all gene families, with Nymphalidae showing the highest rates (average = 0.0073) followed by either Noctuidae (average = 0.0038) or Pieridae (average = 0.0034) and Papilionidae (average = 0.0016). This may indicate that selection forces act similar on gene families involved in plant feeding within Lepidoptera families. However, the differences in rate of change were not in proportion or equal to the level of polyphagy when comparing the PD and FMD values (Figure 2). For example, the Noctuidae that include species with the highest level of polyphagy (PD ranging between 1 - 22.04 and FMD between 0.13 - 0.71) did not correspond with highest rates of change,  $\lambda$ . This might indicate that diet breadth is not the dominant factor contributing to high gene turnover rates in these seven gene families and that other factors, like oscillating host ranges, are equally important.

## Gene family expansions and diet breadth

We analysed gene family expansions of the seven gene families involved in plant feeding and diet breadth to test the correlation between gene family size and level of polyphagy. We observed size differences of gene families across the species of the four focal lepidopteran families (Figure 2). Indeed, Noctuidae have the widest range of accepted host plant families (PD (1 - 22.04) and FMD (0.13 - 0.71), Figure 2), and holds the highest number of genes for all the compared gene families (Table 1).

Within Noctuidae, *Spodoptera* (*S. litura* and *S. frugiperda*) had the highest overall gene count (Table 1). Looking at the range of accepted plant families, *Spodoptera* has the largest host family range of all tested species (*S. frugiperda*: 74 families, PD = 22.04, FMD = 0.71; *S. litura*: 28 families, PD = 10.38, FMD = 0.34; *S. exigua*: 35 families, PD = 11.07, FMD = 0.39; Figure 2, Supplementary Table 3). Genome analyses of these species (Cheng et al., 2017; Gouin et al., 2017; Gui et al., 2020; Xiao et al., 2020), and this study showed expansions in gene families involved in detoxification (Figure 2, Table 1, Supplementary Table 6). These expansions are in line with the large breadth of host plant families and might enable the level of polyphagy. In contrast, the cabbage looper (*Trichoplusia ni*), another major polyphagous species (33 families, PD = 10.79, FMD = 0.35) has lower gene counts for all detoxification families (Figure 2, Table 1, Supplementary Table 6), indicating that expansion levels vary among major polyphagous Noctuidae.

Further, whereas the range of the seven gene family sizes was relatively similar in Papilionidae and Pieridae, there were larger differences in family sizes across Nymphalidae (Figure 2). The majority of the included nymphalids are monophagous (stable PD and FMD values), while gene family sizes are highly variable (Figure 2; Supplementary Table 6). This indicates that expansions in gene families involved in plant feeding are not restricted to polyphagous species. For example, the meadow brown butterfly (*M. jurtina*), which showed the largest number of total gene counts (Table 1), is found in grasslands, open woodland areas, and forest- and field- edges throughout the Palearctic region and is specialising on grasses (Poaceae), and thus considered a monophagous species (Singh et al., 2020; Tshikolovets, 2011) (Supplementary Table 3).

## Testing the correlation between gene family expansions and polyphagy

Previous studies have reported major gene family expansions associated with polyphagy in individual species, such as ABC genes in the spider mite, *Tetranychus urticae* (Dermauw et al., 2013; Dermauw & Van Leeuwen, 2014), GST genes in the red flour beetle, *Tribolium castaneum* (Shi et al., 2012) and P450 genes in the fall armyworm *Spodoptera frugiperda* (Gouin et al., 2017; Gui et al., 2020; Xiao et al., 2020). Here, we found a significant positive correlation between the gene family sizes of the detoxification gene families CCE and GST and level of polyphagy (Figure 3). Carboxyl- and cholinesterases (CCEs) are involved in the first phase of specialised plant metabolite detoxification by modifying the metabolite through hydrolysis (Montella et al., 2012; Oakeshott et al., 2005). This activates the compound for the second detoxification phase involving glutathione S-transferases (GSTs) that catalyse the conjugation of the tripeptide L-glutathione (GSH) and electrophiles, which increases the solubility of the compound and thus increases the ease of excretion (Armstrong, 1997; Francis et al., 2005; Shi et al., 2012). The significant positive correlation indicates that a higher gene count of the CCE and GST families may increase the flexibility

and range of host plant families by detoxification of a wider range of metabolites in polyphagous Lepidoptera. We did not find a significant correlation for the other gene families. This is in contrast to an earlier comparison of seven lepidopteran genomes where a correlation was found for subfamilies within clan 3 of the P450 gene family and host plant breadth (Calla et al., 2017). We focussed on complete gene families and therefore we acknowledge that gene members might be included in the gene family counts that are not involved in detoxification. It is important to note that the study of Calla et al. (2017) indicate that within gene families expansion/loss rates might differ between smaller groups of genes (such as subfamilies within clan 3) which could correlate with host plant breadth. This could also lead to the different outcome of the studies.

Our data shows that putative expansions of gene families involved in plant feeding are species-specific and not restricted to (major) polyphagous species alone. A significant correlation is only found for the CCE and GST gene families in polyphagous Lepidoptera. Expansion in these families is correlated with an increased level of polyphagy and may enable increased levels of polyphagy.

## Conclusions

Using available whole genome data, we studied the association between polyphagy and gene family expansions across Lepidoptera. For each species we calculated the phylogenetic diversity (PD) and specialised metabolite content (FMD) of the host plants within each butterfly/moth diet to quantify level of polyphagy. Expansions of gene families involved in plant feeding were found in both monophagous and polyphagous species. Evolutionary expansion rates varied across Lepidoptera families, but were not proportionally higher in the Noctuidae, a lepidopteran family with widest host plant ranges (highest PD and FMD values). However, we observed a significant positive correlation between the gene expansion of CCE and GST detoxification families and host plant family range (PD and FMD values) across polyphagous Lepidoptera. We therefore conclude that expansions of gene families involved in plant feeding are species-specific and occur in both monophagous and polyphagous species, but particular gene families, CCE and GST, were positively correlated with level of polyphagy.

## Supplementary data

All Supplementary Data is uploaded to the 4TU Centre for Research Data repository and available online: <https://figshare.com/s/Ode53657df61d5e4d7c6>.

**SUPPLEMENTARY TABLE 1** | Overview of genomes including source location and accession dates.

**SUPPLEMENTARY TABLE 2** | Gene family identifiers using InterProScan identifiers and UniRef cluster terms. Based on these identifiers, putative gene members from gene families P450, CCE, UGT, GST, ABC, trypsins and insect cuticle proteins were selected for the orthogroup candidates.

**SUPPLEMENTARY TABLE 3** | Overview of host plant range, feeding style and pest status for all species included in this study.

**SUPPLEMENTARY TABLE 4** | Overview of specialised metabolite content per host plant family. For each family the metabolite content on metabolite type, class, subclass and sub-level is provided.

**SUPPLEMENTARY TABLE 5** | Trait matrix based on the database of specialised metabolite records per plant family as used to calculate the functional metabolite diversity index (FMD).

**SUPPLEMENTARY TABLE 6** | Number of gene members annotated using InterProScan and a local BLASTP against the UniRef50 protein reference database for the five main detoxification gene families P450 monooxygenases (P450s), carboxyl- and cholinesterases (CCEs), UDP-glycosyltransferases (UGTs), glutathione S-transferases (GSTs) and ATP-binding cassettes (ABCs) and the gene families trypsin and insect cuticle protein.

**SUPPLEMENTARY TABLE 7** | Faith's phylogenetic diversity index (PD) per lepidopteran species based on the accepted host plant family range.

**SUPPLEMENTARY TABLE 8** | Functional metabolite diversity index (FMD) per lepidopteran species based on metabolite content for host plant family range.

**SUPPLEMENTARY TABLE 9 (A-C)** | InterProScan annotation report for all 38 specimens.

**SUPPLEMENTARY TABLE 10 (A-C)** | Annotation report of the proteins from all 38 specimens of the BLASTP run against the UniRef50 reference database.

**SUPPLEMENTARY TABLE 11** | Orthogroups identified using OrthoFinder v. 2.2.7 including translated proteins from all 37 Lepidoptera, and single Trichoptera species as included in this study.

**SUPPLEMENTARY TABLE 12** | Gene counts per orthogroup of all orthogroups identified using OrthoFinder.

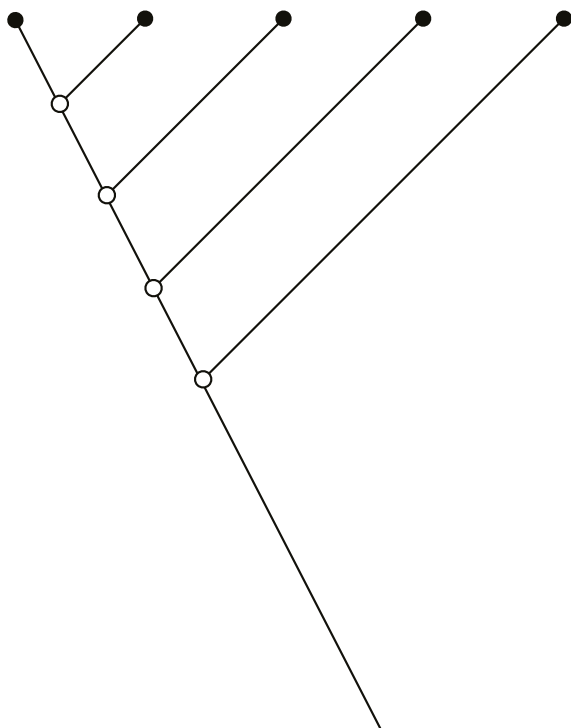
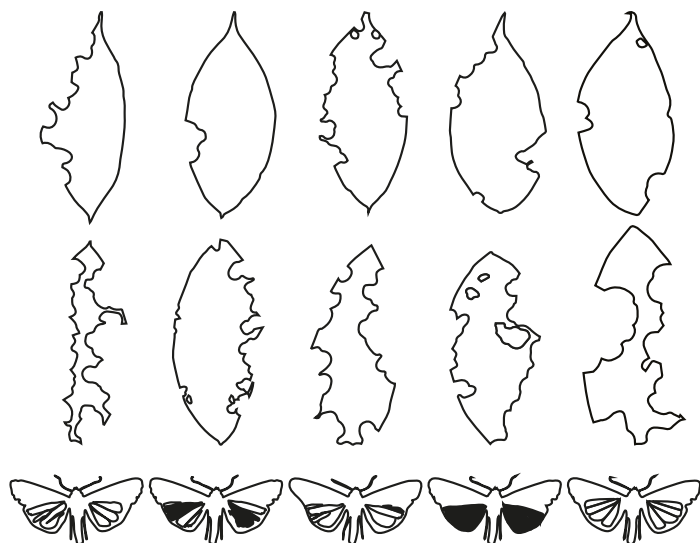
**SUPPLEMENTARY TABLE 13** | *All gene families dataset* as used in the CAFE analyses analysing all gene families. This dataset is filtered for high variance OGs, and also used to calculate the error model.

**SUPPLEMENTARY TABLE 14** | *5 gene families dataset* as used in the CAFE analyses to analyse gene evolution of the 5 selected detoxification gene families (P450, CCE, UGT, GST and ABC).

**SUPPLEMENTARY TABLE 15** | *Single gene family datasets* as used in the CAFE analyses of the seven selected gene families involved in plant feeding (P450, CCE, UGT, GST, ABC, trypsin and insect cuticle protein).

**SUPPLEMENTARY TABLE 16** | Overview of correlation and significance tests between level of polyphagy (quantified by PD and FMD index values) and gene counts of various gene families (P450, CCE, UGT, GST, ABC, trypsin and insect cuticle protein) in polyphagous lepidopteran species.

**SUPPLEMENTARY TABLE 17** | Overview of CAFE results for the analyses of gene family evolution in terms of gene gain and loss rates.





# Chapter 5

Genome and transcriptome analysis of the beet armyworm *Spodoptera exigua* reveals targets for pest control

Sabrina Simon<sup>1,\*</sup>,  
Thijmen Breeschoten<sup>1,\*</sup>,  
Hans J. Jansen<sup>3</sup>,  
Ron P. Dirks<sup>3</sup>,  
M. Eric Schranz<sup>1</sup>  
Vera I.D. Ros<sup>2,\*</sup>

<sup>1</sup>Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands.

<sup>2</sup>Laboratory of Virology, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands.

<sup>3</sup>Future Genomics Technologies, Leiden, The Netherlands.

\*Contributed equally to this study



---

Submitted to G3: Genes, Genomes, Genetics

Preprint available on *bioRxiv*  
DOI: 10.1101/2021.05.14.444148

The genus *Spodoptera* (Lepidoptera: Noctuidae) includes some of the most infamous insect pests of cultivated plants including *Spodoptera frugiperda*, *Spodoptera litura* and *Spodoptera exigua*. To effectively develop targeted pest control strategies for diverse *Spodoptera* species, genomic resources are highly desired. To this aim, we provide the genome assembly and developmental transcriptome comprising all major life stages of *S. exigua*, the beet armyworm. *Spodoptera exigua* is a polyphagous herbivore that can feed from > 130 host plants including several economically important crops.

The 419 Mb beet armyworm genome was sequenced from a female *S. exigua* pupa. Using a hybrid genome sequencing approach (Nanopore long read data and Illumina short read), a high-quality genome assembly was achieved (N50=1.1 Mb). An official gene set (OGS, 18,477 transcripts) was generated by automatic annotation and by using transcriptomic RNA-Seq data sets of 18 *S. exigua* samples as supporting evidence. In-depth analyses of developmental stage-specific expression in combination with gene tree analyses of identified homologous genes across Lepidoptera genomes revealed potential *Spodoptera*-specific genes of interest such as mg7 and REPAT46 upregulated during 1<sup>st</sup> and 3<sup>rd</sup> instar larval stages for targeted pest-outbreak management.

The beet armyworm genome sequence and developmental transcriptome covering all major developmental stages provides critical insights into the biology of this devastating polyphagous insect pest species with a worldwide distribution. In addition, comparative genomic analyses across Lepidoptera significantly advance our knowledge to further control other invasive *Spodoptera* species and reveals potential lineage-specific target genes for pest control strategies.

#### Keywords

Beet armyworm; *Spodoptera exigua*; whole genome; transcriptomics; gene expression; pest control

## Introduction

Analysis of genome and transcriptome data can be used to study many important questions ranging from species-specific mutations to comparative genomic evolutionary patterns. The genus *Spodoptera* is known for the high number of notorious pest species causing enormous agricultural damage resulting in economic losses worldwide, including *Spodoptera exigua*, *Spodoptera frugiperda* and *Spodoptera litura* (Cheng et al., 2017; EPPO, 2019; Goergen et al., 2016; Pogue, 2002). The beet armyworm, *S. exigua* (Hübner) (Lepidoptera: Noctuidae) is a devastating polyphagous insect pest with a worldwide distribution (Fu et al., 2017; Mehrkhou et al., 2012), being able to feed on more than 130 plant species from at least 30 families including several economically important crops such as sugar beet, cotton, soybean, cabbage, maize and tomato (Fu et al., 2017; Mehrkhou et al., 2012; Merks-Jacques et al., 2008; G. S. Robinson et al., 2010). *Spodoptera exigua* originated in Southern Asia and was subsequently introduced to other parts of the world including North America and Europe (Fu et al., 2017; Mehrkhou et al., 2012). It is widely distributed in the tropical and subtropical regions and migrates into more temperate regions throughout the growing season (Pogue, 2002). Its long-distance migration likely played a major role in the geographic expansion of populations and its spread across the world (Fu et al., 2017). In temperate regions it can be abundant in greenhouses (Smits et al., 1986).

Successful control of *S. exigua* is challenging due to its broad host range, rapid growth rate, its migratory dispersal and its ability to rapidly evolve resistance to pesticides (Fu et al., 2017; Hu et al., 2021; Huang et al., 2021). Moreover, the use of conventional chemical pesticides causes health and environmental issues and is generally less accepted (Omkar, 2016; Wheeler, 2002). Therefore, there is a pressing need for other, more sustainable, strategies to control *S. exigua* and other *Spodoptera* species. A promising approach includes RNA interference (RNAi)-based insect management (Burand & Hunter, 2013; Renuka et al., 2017; Scott et al., 2013). One of the major challenges is to find target genes for RNAi to control specific pest species or a range of closely related pest species (Bi et al., 2016; Li et al., 2013; Tian et al., 2019). One way to select potential species-specific candidate genes is by carefully analysing orthologous relationships of homologous genes in related species. Targeting specific gene(s) of single species using RNAi approaches could be an extremely powerful tool to diminish a specific pest outbreak without harming other (closely-related) arthropod species (Price & Gatehouse, 2008; Scott et al., 2013), which often does occur when applying general insecticides (Schulz, 2004). Given the high pest potential of many *Spodoptera* species, lineage-specific genes should be identified that can be targeted during pest outbreaks. However, genomic studies have been focused mainly on *S. frugiperda* (Gouin et al., 2017; Kakumani et al., 2014) while other *Spodoptera* species have largely been neglected. To address this gap, we present the *S. exigua* genome assembly and official gene set.

In this study we obtained an RNA sequencing (RNA-Seq) profile across all major life stages of *S. exigua*. We performed in-depth analysis of gene expression patterns during the different developmental stages and identified four *Spodoptera*-specific candidate genes for RNAi-based pest management strategies. Furthermore, we produced a *de novo* assembled genome draft of *S. exigua*, based on one female pupa.

## Results

### Genome annotation and comparison to other Lepidoptera genomes

For the *de novo* assembly of the *S. exigua* genome both Oxford Nanopore long read data and Illumina 2x150nt short read data were used. The TULIP assembly strategy was employed to assemble the genome from the long read data (Jansen et al., 2017). This was followed by two rounds of error correction with the long read data using Racon (Loman et al., 2015) and two rounds of Pilon polishing with the Illumina data (Goodwin et al., 2015; Walker et al., 2014).

The total size of the final polished assembled genome was 419 Mb, which was divided over 946 contigs (largest contig = 4.15Mb) with N50=1.1 Mb (Table 1). To confirm the assembly genome size, a k-mer counting approach was used. After counting the 21 and 27 mers in the Illumina data set, the count tables were analysed with GenomeScope. The genome size as estimated by k-mer counting was ~370 Mbp, which correlated with the Nanopore assembly size (which is slightly larger). The genome size of *S. exigua* presented here, as well as the GC content (given in %), is comparable to other published *Spodoptera* genomes (Table 1).

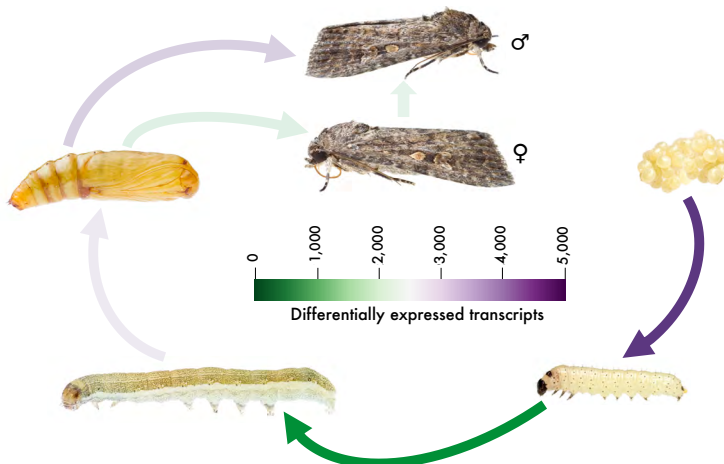
The Benchmarking Universal Single-Copy Ortholog (BUSCO) (v. 3) assessments indicated that the quality and completeness of our *de novo* assembly was good (complete: 96.8%; fragmented: 1.0%; missing: 2.2%) and comparable to other lepidopteran genomes (Supplementary file 1: Figure S1; all Supplementary Data is uploaded to the 4TU Centre for Research Data repository and available online: <https://figshare.com/s/Oe12d9153800dcdf8927>). By these quality metrics, the *S. exigua* assembly is comparable to those of fellow lepidopterans, facilitating comparative genomic analyses.

Using our final assembly, an official gene set (OGS) was generated by automatic annotation and transcriptomic RNA-Seq data sets of 18 *S. exigua* samples (see below) as supporting evidence. The official gene set, OGS v. 1.1, consists of 18,477 proteins and is provided at the Dryad digital repository (<https://datadryad.org/stash/share/aUEyO49FeYl7cFZOE72MjLZnhV4N7enuapuxfSpL3cw>).

## Gene expression analyses across the whole life-cycle of *Spodoptera exigua*

The major developmental stages across the whole life-cycle of *S. exigua*, namely embryonic stage (egg), early 1<sup>st</sup> instar larva (L1), early 3<sup>rd</sup> instar larva (L3), pupa, and adult (both sexes: female and male), were sequenced on an Illumina NovaSeq 6000 system at an average of 13.4 million PE2x150nt reads (6.9-22.5 million reads per sample) (Supplementary file 2: Table S1.3). Based on these reads, we performed differential expression analyses using our *de novo* assembled *S. exigua* genome as a reference.

We first compared gene expression from subsequent different developmental stages and sexes based on pairwise comparisons to determine the dynamic changes in gene expression during development. We used DESeq2 v. 1.18.1 (Love et al., 2014) to discern significant changes in gene expression throughout the whole life cycle. A striking number of significantly differentially expressed transcripts ( $n=4,974$  transcripts) was detected during early embryonic development (between the egg and the 1<sup>st</sup> instar larval stage) (Figure 1). Notably, this rapid change in the expression dynamics of *S. exigua* was the largest during the entire life cycle (Figure 1 and Supplementary file 3: Table S2). In contrast, the smallest change in gene expression was between 1<sup>st</sup> and 3<sup>rd</sup> instar larvae ( $n=1,222$  transcripts). A larger change in gene expression was also observed between pupa and adult male ( $n=3,112$  transcripts) compared to pupa to adult female ( $n=2,061$  transcripts), likely due to the fact that female pupae were analysed. For an overview of relationships between the different life stages based on identified significant changes in gene expression see Supplementary file 4: Figure S2.



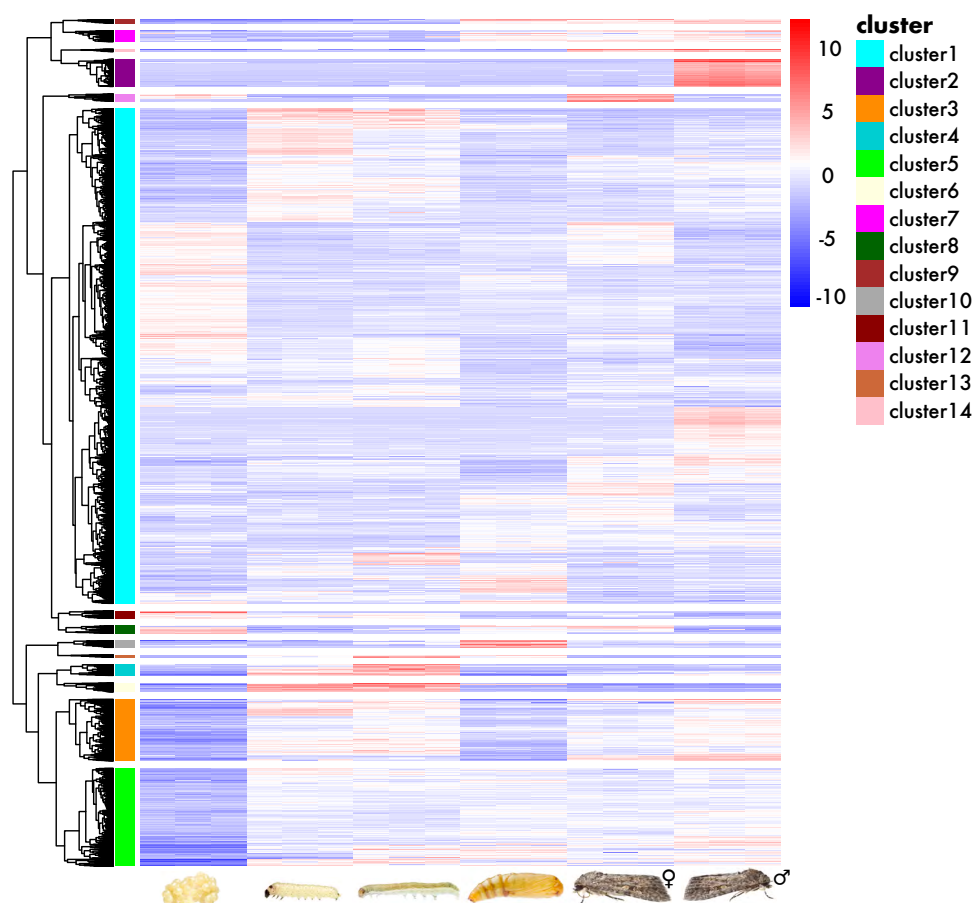
**FIGURE 1** | *Spodoptera exigua* life cycle and gene expression profile. The major developmental stages and sexes sequenced for *Spodoptera exigua* are shown, starting from an egg (embryonic stage) and proceeding two larval stages, namely 1<sup>st</sup> instar and 3<sup>rd</sup> instar. After the pupal stage, there is the final differentiation into adult male and female. The colour of the arrows is proportional to the number of statistically significant differentially expressed genes (FDR = 0.001, minimal fold-change of four). Note that the size of the developmental stages is not proportional.

**Table 1.** Genome metrics of *Spodoptera exigua* and other published *Spodoptera* genomes.

Sequencing information		
Species	Notes	Method
<i>Spodoptera exigua</i>	Female pupae	Nanopore + Illumina
<i>Spodoptera litura</i>	Male adults	Illumina
<i>Spodoptera frugiperda</i>	Sf21 cell line	Illumina
<i>Spodoptera frugiperda</i> “corn”	Two male larvae	Illumina
<i>Spodoptera frugiperda</i> “rice”	Single male larva	Illumina
<i>Spodoptera frugiperda</i>	Sf9 cell line	PacBio
<i>Spodoptera frugiperda</i> “corn”	Two male larvae	PacBio + Illumina + Hi-C
<i>Spodoptera frugiperda</i> “rice”	Single male larva	PacBio + Illumina + Hi-C
<i>Spodoptera frugiperda</i>	Single male adult	MGISEQ + Hi-C
<i>Spodoptera frugiperda</i>	Single male adult	PacBio + Illumina + Hi-C
<i>Spodoptera frugiperda</i>	Female pupa	PacBio + Hi-C

Reference	Genome assembly				
	Genome assembly total length	Contig N50	# contigs	GC content	proteins
This study	419.3 MB	1.1 Mb	946	36.52	18,477
Cheng et al. 2017	438.3 Mb	0.068 Mb	13,636	37	15,317
Kakumani et al. 2014	358.0 Mb	0.008 Mb	97,607	32.97	11,595
Gouin et al. 2017	437.9 Mb	21.6 Kb	-	36	21,700
Gouin et al. 2017	371.0 Mb	25.4 Kb	-	36	26,329
Nandakumar et al., 2017	451.0 Mb	0.25 Mb	4,577	36.53	25,699
Nam et al. preprint 2018, preprint 2019	384.4 Mb	-	-	36.34	21,839
Nam et al. preprint 2018, preprint 2019	379.9 Mb	-	-	36.37	22,026
Gui et al. 2020	543.7 Mb	0.09 Mb	29,584	36.52	22,201
Zhang et al., 2020	390.4 Mb	5.6 Mb	776	36.40	22,260
Xiao, 2020	486.3 Mb	1.1 Mb	618	36.40	22,623

We further identified 9,896 transcripts as differentially expressed (DE) across all pairwise comparisons. Hierarchical clustering revealed 14 clusters of DE transcripts with similar expression patterns (Figure 2). Of these, the gene expression of eight clusters could be associated with a single developmental stage or to similar subsequent developmental stages, e.g. one cluster for the larval stage (see also Supplementary file 5: Figure S3). For these eight clusters, statistically over-represented Gene Ontology (GO) terms were identified using FDR-adjusted  $p$ -value ( $< 0.05$ ) and were further summarised to generic GO slim categories (Figure 3).



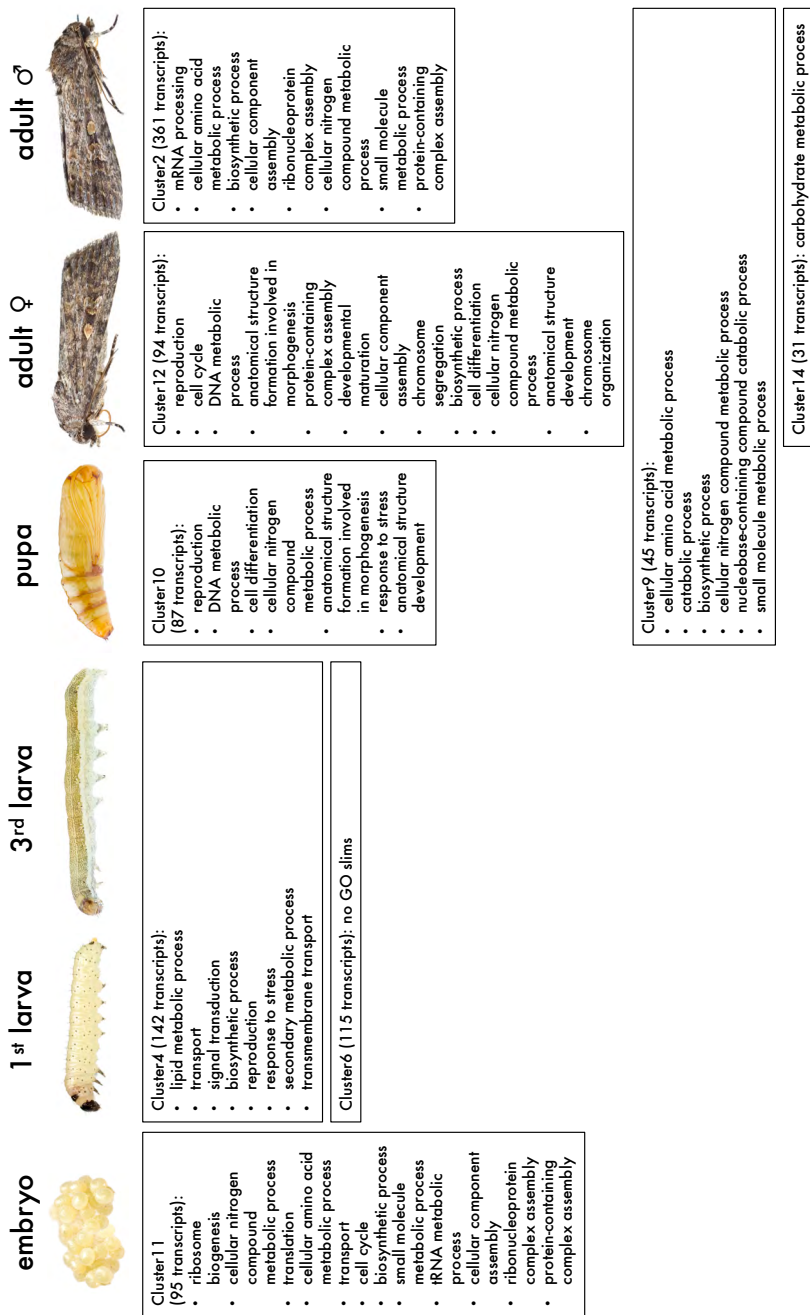
**FIGURE 2** | Hierarchical clustering dendrogram of all differentially expressed (DE) genes in the life cycle of *Spodoptera exigua*. Heatmap shows 9,896 transcripts which have been identified differentially expressed (minimal fold-change of four,  $\text{FDR} \leq 1\text{e-}3$ ) between the six developmental stages / sexes including three replicates each (left to right: embryo, 1<sup>st</sup> instar larva, 3<sup>rd</sup> instar larva, pupa, female adult, male adult). Transcripts form 14 distinct clusters using a cut-of at 50% (right dendrogram). The colour key of the heatmap indicates low (blue) to high (red) expression values for transcripts.



For the embryonic stage (cluster 11, Figure 3), there was an enrichment of GO categories associated with ribosome biogenesis (GO:0042254), ribonucleoprotein complex assembly (GO:0022618), tRNA metabolic process (GO:0006399), translation (GO:0006412), and cell cycle (GO:0007049). The enrichment of these categories highlights the rapid succession of cell cycles associated with chromatin replication and initiation of transcription and translation for embryo patterning (Koutsos et al., 2007). Detailed investigation of DEs gene annotations based on the Arthropoda database (Supplementary file 6: Table S3 and Supplementary file 7: Table S4) revealed several known genes important in morphogenesis, for example during the embryonic stage Krüppel-like transcription factors (KLFs) (Kaczynski et al., 2003; McCulloch & Koenig, 2020), specificity proteins (SP) (Kennedy et al., 2016) and several WD-repeat-containing proteins (Smith, 2008).

We did not identify a specific cluster for the 1<sup>st</sup> larval stage nor for the 3<sup>rd</sup> larval stage, but rather one cluster including both larval stages (=larval stage cluster, cluster 4, Figure 3). The larval stage was enriched for genes involved in general metabolic processes, such as signal transduction (GO:0007165), biosynthetic processes (GO:0009058), and secondary metabolic processes (GO:0019748). Several genes having a key role in the digestion of plant material and herbivore success were significantly differentially expressed within the larval stage (see Supplementary file 6: Table S3). These include REPAT genes (Herrero et al., 2007; Navarro-Cerrillo et al., 2013), trypsins (Muhlia-Almazán et al., 2008), cuticle proteins (Breeschoten et al., 2019; Celorio-Mancera et al., 2013; Müller et al., 2017; Orsucci et al., 2018), and members of prominent detoxification gene families such as cytochrome P450s (P450), carboxyl/cholinesterases (CCE), glutathione S-transferases (GST), and UDP-glycosyltransferases (UGT). The pupal stage varied from the larval stage in that there was significant enrichment in processes associated with cell differentiation (GO:0030154), anatomical structure formation involved in morphogenesis (GO:0048646), and anatomical structure development (GO:0048856).

We further identified several pupal cuticle proteins as significantly differentially expressed within this pupal stage. The female adult stage (cluster 12) was enriched for genes involved in e.g., cell cycle (GO:0007049), chromosome segregation (GO:0007059) and chromosome organisation (GO:0051276), anatomical structure development (GO:0048856), and biosynthetic process (GO:0009058) and we identified orthologs of several homeotic genes(-like), such as *Bicaudal C*, *Sex combs reduced* (*Scr*) and *proboscipedia* (*pb*). For the male adult stage (cluster 2, Figure 3), there was an enrichment of GO categories associated with e.g. mRNA processing (GO:0006397), cellular amino acid metabolic process (GO:0006520), cellular component assembly (GO:0022607), and biosynthetic process (GO:0009058). For the female and the male adult stage we further identified several sex-specific genes as differentially expressed, such as vitellogenin and vitellogenin receptor in the female (Rotllant et al., 2017) and testis-specific serine/threonine-protein kinase 2 (Kim et al., 2019) or ejaculatory bulb-specific protein (Liu et al., 2020) in the male stage, respectively.



**FIGURE 3** | Upregulated GO slims (Biological Process) per development stages. Shown are only the eight clusters of DE transcripts which could be assigned to one developmental stage or sex or to subsequent developmental stages. The cluster number is according to the formed clusters as indicated in Figure 2. The number of transcripts is provided in parentheses as well as the statistically over-represented GO terms ( $FDR \leq 0.05$ ) which have been summarised to generic GO slim categories.

One cluster (cluster 14) was specific for both adult sexes but was enriched only for carbohydrate metabolic process (GO:0005975). In contrast, cluster 9 (comprised of the pupa and both adult sexes) was enriched for several GO categories: cellular amino acid metabolic process (GO:0006520), catabolic process (GO:0009056), biosynthetic process (GO:0009058), and cellular nitrogen compound metabolic process (GO:0034641) (see Figure 3 and Supplementary file 8: Table S5).

## Lepidopteran phylogenomics and detoxification gene content evolution

Phylogenomic analysis correctly placed *S. exigua* within the *Spodoptera* clade and as the sister-group to the clade containing *S. litura* and *S. frugiperda* (Figure 4) (Kergoat et al., 2021; Le Ru et al., 2018). In addition, the inferred species relationships within Lepidoptera were in agreement with previous findings (Kawahara et al., 2019). We further scanned all lepidopteran genomes for gene families associated with detoxification functions. This included: gene families involved in phase I of the detoxification pathway such as cytochrome P450 monooxygenase (P450) and carboxyl- and cholinesterase (CCE) (Kant et al., 2015); gene families involved in phase II, such as UDP-glycosyltransferase (UGT) and glutathione-S-transferase (GST); and the gene family ATP-binding cassette (ABC) involved in phase III (Heidel-Fischer & Vogel, 2015; Kant et al., 2015; Li et al., 2007). Based on the annotation of the lepidopteran genomes, we searched for expanded detoxification-related genes (Figure 4 and Supplementary file 9: Table S6). Expansion of major genes families involved in detoxification was mainly visible for *S. frugiperda* (“corn” strain) within the Noctuidae. In the following, we analysed in greater detail several lineage-specific genes.

## Potential lineage-specific and stage-specific candidate genes as targets for pest-control

We used OrthoFinder v. 2.3.11 (Emms & Kelly, 2015) to identify homologous gene sequences in the genomes of eight closely related but diverse lepidopteran species including three *Spodoptera* species, *S. exigua*, *S. litura*, and *S. frugiperda*. We aimed to identify *Spodoptera*-specific orthogroups, as such lineage-specific genes would be candidates for targeted pest-outbreak management development. We identified in total 119 orthogroups (OGs) containing genes from only the three *Spodoptera* species (Supplementary file 10: Table S7.1).

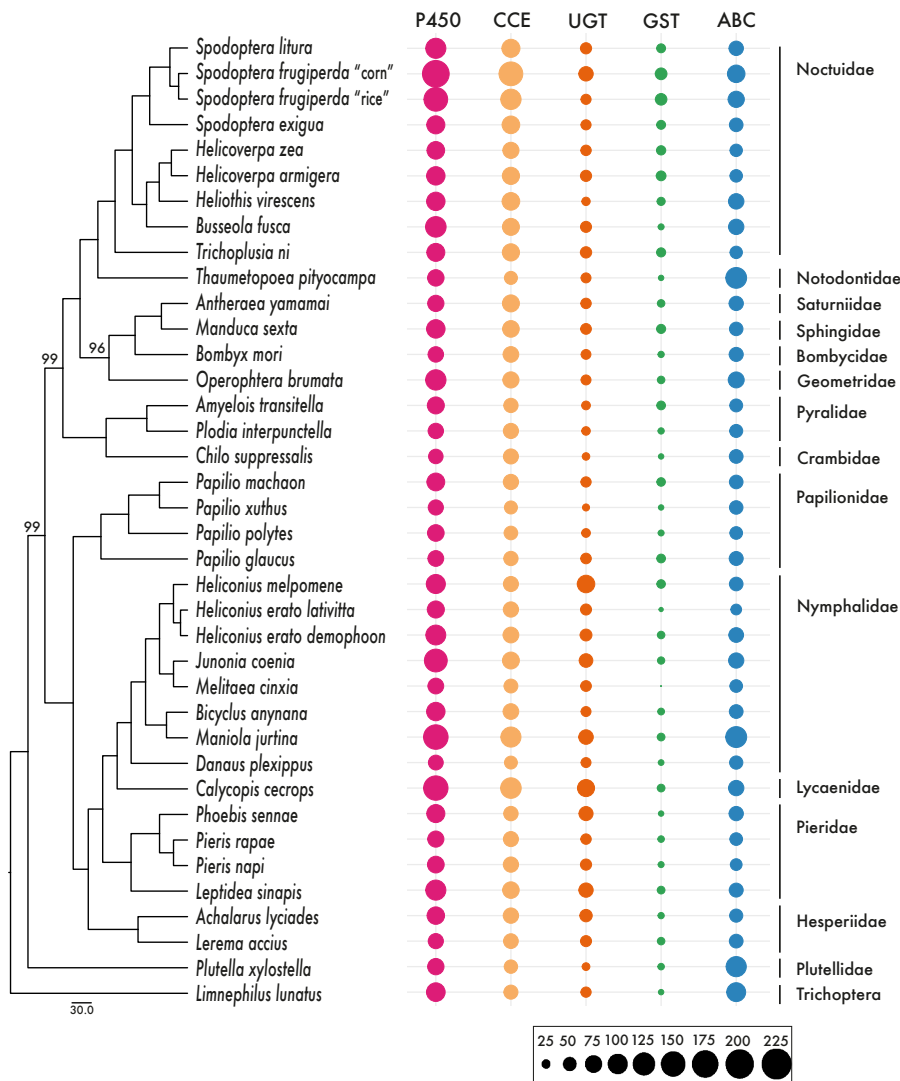
Since the larval feeding stage of *Spodoptera* is the most detrimental to crops, we further selected seven OGs for which the *S. exigua* gene representative is differentially expressed in the larval stage cluster (cluster 4). For three of the seven genes, the closest homologs were “uncharacterised” proteins (Supplementary file 10: Table S7.2). The four remaining genes were annotated as: nuclear complex protein (OG0013351), REPAT46 (OG0014254), trypsin alkaline-c type protein (OG0014208) and mg7 (OG0014260) (Supplementary file 10: Table S7.2). We confirmed the expression of all

seven genes by checking the number of RNA-Seq reads mapped to each assembled transcript based on the results of the transcript abundance estimation with RSEM. The read count in the larval stages (L1 and L3 stages) was higher than in the other stages (Supplementary file 11: Table S8). Several reads derived from other stages mapped to the protein sequences, but the number of these mapped reads was low (Supplementary file 11: Table S8).

For the four putative lineage-specific and stage-specific annotated genes, we validated their *Spodoptera*-specificity by constructing gene trees of *Spodoptera* sequences with their most similar sequences identified from other lepidopteran species. For two of the annotated genes (mg7 and REPAT), we constructed two different gene trees that differed in dataset composition. The identification of putative homologs in related species varied per gene as well as the number of included sequences and species for the gene tree analyses (nuclear complex protein (OG0013351): 20 sequences, 3,494 amino acid (aa) positions, REPAT46 (OG0014254) extended dataset containing both  $\alpha$ REPAT and  $\beta$ REPAT clusters: 153 sequences, 863 aa positions, reduced dataset containing only the  $\beta$ REPAT cluster: 91 sequences, 717 aa positions, trypsin alkaline-c type protein (OG0014208): 69 sequences, 1,101 aa positions, and mg7 (OG0014260): extended dataset: 27 sequences, 368 aa positions, reduced dataset: 17 sequences, 350 aa positions). All alignment files are provided at the Dryad digital repository (<https://datadryad.org/stash/share/aUEyO49FeYl7cFZOE72MjLZnhV4N7enuapuxfSpL3cw>).

The gene tree of the nuclear pore complex proteins showed that the *Spodoptera*-specific genes form a single cluster, nested within lepidopteran DDB\_G0274915-like nuclear pore complex proteins and sister to *Helicoverpa* sequences (Supplementary file 12: Figure S4). The reduced mg7 dataset comprised sequences from the *Spodoptera*-specific OG in addition to the ortholog group “15970at7088” from OrthoDB. For the extended mg7 dataset we additionally included all ‘mg’ protein sequences according to He et al. (2012). The ortholog group “15970at7088” included nine single copy genes present in other butterfly species and we found two paralogous copies in *S. litura*, likely due to a specific gene duplication. In order to evaluate whether other paralogs were present in any of the *Spodoptera* gene sets, we blasted the protein sequences against a local blast database of mg7 sequences comprising the sequences from OrthoDB, OG0014260, and He et al. (2012). In *S. exigua* we identified three paralogs, which according to the GFF file are located (mRNAs) consecutively on the genome: 1268792-1275628, 1276053-1279376, 1280841-1286731. Similarly, in *S. litura* we identified two paralogs and three paralogs in *S. frugiperda*. Running the same blast searches but using the protein sets of *Bombyx mori*, *Helicoverpa armigera*, *Helicoverpa zea*, *Heliothis virescens* and *Trichoplusia ni* all detected a single gene copy with reliable BLAST scores. Both the reduced and the extended mg7 gene trees included all identified *Spodoptera* paralogs. The reduced mg7 gene tree including all paralog *Spodoptera* genes and the single copy homologs from OrthoDB showed that *Spodoptera*-specific OG sequences were clustered together (Supplementary file 13:

Figure S5). This cluster formed a sister clade to all remaining *Spodoptera* paralogs and the *H. armigera* gene. In the extended mg7 gene tree the *Spodoptera*-specific OG sequences did not form a monophyletic clade but did cluster together with the mg7 genes of *Choristoneura fumiferana*, *H. armigera* and *S. litura* derived from He et al. (2012) (Supplementary file 14: Figure S6).



**FIGURE 4** | Comparison of Lepidoptera genomes and inferred phylogenetic relationships. Shown is the maximum likelihood phylogeny based on 1,367 single copy BUSCOs (left, all nodes have 100% support unless otherwise noted). Number of detoxification gene members of five main detoxification families, P450 monooxygenases (P450s), carboxyl- and cholinesterases (CCEs), UDP-glycosyltransferases (UGTs), glutathione S-transferases (GSTs), ATP-binding cassettes (ABCs), are presented per species in a bubbleplot generated with ggplot2.

For the REPAT gene analyses we compiled two datasets. Both datasets consisted of sequences derived from the *Spodoptera*-specific OG, the MBF2 ortholog group “16151at7088” from OrthoDB and all protein sequences according to Navarro-Cerrillo et al. (2013). The reduced dataset only contained protein sequences belonging to the  $\beta$ REPAT class while the extended dataset included both  $\alpha$ REPAT and  $\beta$ REPAT classes. In both gene tree analyses, the *Spodoptera*-specific OG sequences clustered together with the annotated REPAT46 gene from *S. exigua* (Supplementary file 15: Figure S7, Supplementary file 16: Figure S8). The *Spodoptera*-specific orthogroup is placed in the  $\beta$ REPAT cluster, *sensu* Navarro-Cerrillo et al. (2013), where it is placed within group VI (Navarro-Cerrillo et al., 2013). Further, in total 54 putative REPAT proteins have been identified in the *S. exigua* protein set which were included in both gene tree datasets (Supplementary file 17: Table S9).

The gene tree of the trypsin proteins showed a monophyletic clustering of all Lepidoptera-derived trypsin genes (Supplementary file 18: Figure S9). In addition, all *Spodoptera* trypsins were clustered within one monophyletic clade, with the *Spodoptera*-specific orthogroup nested within. Trypsins occurred in all Lepidoptera species in large numbers, thus we compared various OrthoFinder runs under different stringency settings (varying the inflation parameter from 1, 1.2, 1.5 (default), 3.1 and 5) to test the degree of “*Spodoptera*-specificity” of this orthogroup. In all five runs, the OG containing the *Spodoptera* trypsin genes was stable (e.g. lineage-specific) and remained unchanged.

## Discussion

Using a combination of Oxford Nanopore long read data and Illumina short read data for the genome sequencing approach, we generated a high-quality genome and transcriptome of the beet armyworm, *Spodoptera exigua*. These resources will be beneficial for future research on *S. exigua* and other noctuid pest species. The developmental gene expression profile of *S. exigua* demonstrated that the transition from embryo to larva is the most dynamic period of the beet armyworm’s transcriptional activity. Within the larval stage the transcriptional activity was highly similar between early (1<sup>st</sup>) and late (3<sup>rd</sup>) instars, making the early larval stage to an ideal stage for pest-control (see below). Within the larval stage, genes involved in secondary metabolic process (GO:0019748) were uniquely expressed during the whole development (Figure 3). In addition, several prominent genes involved in digestion and detoxification, including cytochrome P450s and UDP-glycosyltransferases, and potential target genes for pest control could be identified which are specifically expressed in the larval stage (Supplementary file 6: Table S3).

The significant enrichment in the pupal stage in processes associated with anatomical structure development reflects the dramatic structural changes of the larva to the adult (Truman & Riddiford, 2019). The identified pupal cuticle proteins

within the pupal stage have been reported previously by other studies and reflect the morphological changes in wing disc and the larva-to-pupa metamorphosis (Gu et al., 2013; Ou et al., 2014).

The gene expression analyses of the developmental transcriptome of *S. exigua* revealed larval stage-specific upregulated genes (cluster 4, Figure 2&3). These identified genes are strong candidates for targeted RNAi of feeding larvae. Targeted RNAi of genes involved in vital functions of the most important larval stage can be an efficient strategy to minimise the detrimental effect of pest species (Xue et al., 2012). The larva stages of Noctuidae insects are the most damaging to plants. Our homology search revealed seven *Spodoptera*-specific genes with upregulation in the 1<sup>st</sup> and 3<sup>rd</sup> instar larval stages, and highest expression levels in the 3<sup>rd</sup> instar stage (Supplementary file 11: Table S8). Four of these seven genes were annotated and we confirmed *Spodoptera*-specificity by gene tree analyses.

One putative *Spodoptera*-specific orthogroup consisted of nuclear pore complex proteins. These proteins are involved in the transport of particles through the nuclear envelope (Alber et al., 2007). Although the gene tree did not follow well established lepidopteran relationships (Kawahara et al., 2019), e.g. Noctuoidea nested within Papilionoidea (Supplementary file 12: Figure S4), all identified *Spodoptera* nuclear pore complex proteins clustered together. This is an initial requisite for potential target genes, showing a clear separation of *Spodoptera*-derived sequences to sequences of other species.

We identified mg7 as potential target gene for RNAi. This gene was previously reported to be highly upregulated in all larval stages in the midgut of *S. litura* with an expression peak after larvae have molted into the 6<sup>th</sup> larval stage (He et al., 2012). Our results show a similar pattern with an increased expression towards the 3<sup>rd</sup> instar larva (Supplementary file 11: Table S8). Expression in the midgut suggests a role in digestion-related processes (He et al., 2012). The gene tree based on the reduced dataset showed a clustering of *Spodoptera*-specific mg7 genes (Supplementary file 13: Figure S5). He et al. (2012) reported several homologs, mg2, mg7, mg9, mg17, in related species which we included in the extended gene tree reconstruction (Supplementary file 14: Figure S6). The genes derived from the *Spodoptera*-specific orthogroup form a monophyletic group with the mg7 genes of *C. fumiferana*, *H. armigera* and *S. litura* derived from He et al. (2012), establishing orthology of Noctuidae and Tortricidae sequences. The spruce budworm, *C. fumiferana* is a notorious conifer-feeding pest restricted to the Nearctic region where it is considered one of the most destructive insect defoliators (Lumley & Sperling, 2010; Volney & Fleming, 2007). The extended phylogeny identified homologous clusters (although with low support values) of “mg” genes (mg7, mg17 and mg9) in related lepidopteran species. The close-relationship of additional gene family members from other lepidoptera make mg7 a potential candidate for RNAi-based pest-formation control in a wider range of lepidopteran pest species with the caveat that more work is needed to resolve specificity.

Finally, a strong potential target gene for biocontrol are the  $\alpha$ REPAT proteins which are involved in various physiological processes and can be induced in response to infections, bacterial toxins and other microbial pathogens within the larval midgut (Herrero et al., 2007; Navarro-Cerrillo et al., 2013). Upregulation of REPAT genes has been identified in response to the entomopathogenic *Bacillus thuringiensis* (Herrero et al., 2007). In *S. frugiperda*, REPAT genes were associated with defence functions in other tissues than the midgut and found to be likely functionally diverse with roles in cell envelope structure, energy metabolism, transport and binding (Machado et al., 2016).

REPAT genes are divided in two classes based on conserved domains. Homologous genes of the  $\alpha$ REPAT class are identified in closely related *Spodoptera* and *Mamestra* species, while  $\beta$ REPAT class homologs are identified in distantly related species, for example HMG176 in *H. armigera* and MBF2 (Multiprotein bridge factor 2) in *B. mori* (Navarro-Cerrillo et al., 2013). Our analyses found that REPAT genes (and homologs like MBF2 members) from distantly related species are nested within the  $\beta$ REPAT cluster, while the  $\alpha$ REPAT class are exclusive for *Spodoptera* and very closely related species like *Mamestra* spp. (Navarro-Cerrillo et al., 2013; Zhou et al., 2016) (Supplementary file 15: Figure S7, Supplementary file 16: Figure S8). In contrast to Navarro-Cerrillo et al. (2013) where  $\alpha$ REPAT and  $\beta$ REPAT form sister clades, our tree topology show  $\alpha$ REPAT genes to be nested within  $\beta$ REPAT.

Previously, 46 REPAT genes were reported for *S. exigua* (Navarro-Cerrillo et al., 2013), while we detected 54 REPAT genes in the *S. exigua* genome (Supplementary file 6: Table S3). The genes of *S. exigua*, *S. litura* and *S. frugiperda* from the *Spodoptera*-specific orthogroup as identified here cluster together with REPAT46 from *S. exigua* and thus are group VI  $\beta$ REPAT genes (Supplementary file 15: Figure S7). As shown in Navarro-Cerrillo et al. (2013) and here (Supplementary file 15: Figure S7), group VI  $\beta$ REPATs are comprised of *Spodoptera*- and other noctuid-derived genes, like *Helicoverpa* and *Mamestra*. The Noctuidae family is one of the most damaging groups of pests to agriculture, which is recognised by naming of a “pest clade” where species from the genera *Spodoptera*, *Helicoverpa* and *Mamestra* are included (Mitchell et al., 2006; Regier et al., 2017). Overall, the results presented here show that REPAT gene members of especially the  $\alpha$ REPAT class and the group VI  $\beta$ REPATs are putatively promising candidates for targeted RNAi in notorious pest species belonging to *Spodoptera* and closely related genera in Noctuidae, given their *Spodoptera*- and/or Noctuidae-specificity.

## Conclusions

The genome and developmental transcriptome including all major stages: embryonic, larval, pupal and adult stages of both sexes, of the beet armyworm *Spodoptera exigua* provides a valuable genomic resource for this important pest species. Using



a dual sequencing approach including long read and short read data, we were able to provide a genome which is comparable to fellow lepidopterans, strongly supporting the use of these resources in further genome comparisons. Based on the differential gene expression analyses we identified developmental stage-specific (embryonic, larva, pupa or adult) or sex-specific (female, male adult) transcriptional profiles. Of particular interest are the identified genes specifically upregulated in the larval stages because those stages are most detrimental to the host plants. We have further validated these larva-specific genes for their suitability for RNAi-based targeted pest control by comparative genome analyses. RNAi-mediated insect control can be a powerful tool if selected target gene(s) are essential genes in insect tissues to trigger toxic effects. In addition, the target gene(s) should be pest species-specific or specific to a range of closely related pest species and should not harm non-target organisms. In this context, *Spodoptera* lineage-specific target gene(s) are of high interest due to the high number of notorious pest species in this genus causing enormous agricultural damage resulting in economic losses worldwide. Analysing the homologous relationships of the identified potential target genes and including a broad selection of other insect species allowed us to verify the specificity of four candidate genes for the genus *Spodoptera*. Additional in-depth research may further confirm the clade-specificity of these genes and their potential application in RNAi mediated pest-outbreak management.

## Materials and methods

### Breeding and sample collection

*Spodoptera exigua* specimens originated from a stock rearing of the Laboratory of Virology, Wageningen University & Research, which was initiated in July 2014 using pupae from a large continuous rearing, kindly provided by Andermatt Biocontrol (Switzerland). The rearing was kept on artificial diet at 27°C with 50% relative humidity and a 14:10 h light:dark photoperiod. The artificial diet consisted of water, corn flower, agar, yeast, wheat germ, sorbic acid, methylparaben, ascorbic acid and streptomycin sulphate. Disposable plastic trays covered with paper tissues and a lid were used as rearing containers for groups of maximum 35 larvae (for larger stages). Late 5<sup>th</sup> instars were transferred to a plastic tray containing vermiculite to facilitate pupation. Pupae were collected and transferred to cylindrical containers lined with paper sheets for egg deposition, with around 45 pupae per cylinder. Adult moths were provided with water. Collected eggs were surface sterilised with formaldehyde vapor to eliminate external microbial contamination.

High molecular weight (HMW) chromosomal DNA was extracted from a female *S. exigua* pupa using the Qiagen Genomic-tip 100/G kit according to the manufacturer's instructions (Qiagen, Venlo, The Netherlands). Quality of the extracted HMW DNA was

analysed on an Agilent 4200 TapeStation System using Genomic DNA ScreenTape (Agilent, Amstelveen, The Netherlands).

To retrieve samples for RNA sequencing (RNA-Seq), a newly hatched male and female from the continuous rearing were mated in a plastic cup. Offspring of this couple was used for RNA-Seq, six stages were collected: embryonic (eggs), 1<sup>st</sup> instar larvae, 3<sup>rd</sup> instar larvae, pupae, male adults, female adults, with three replicates (individuals) per stage except for the embryonic stage were three clusters of each ~100 eggs were taken. To obtain the samples, eggs were harvested, and larvae were reared as above. For the embryonic stage, egg clusters (laid within 21 hours) were cut out of paper, transferred to Eppendorf tubes, snap frozen in liquid nitrogen and transferred to -80 °C until shipment on dry ice to Future Genomics Technologies for further RNA extraction and sequencing. Synchronised newly hatched (non-fed) 1<sup>st</sup> instar larvae, early 3<sup>rd</sup> instar larvae, second day pupae and newly emerged (non-mated) female and male adults were collected. Individuals were transferred to Eppendorf tubes and snap frozen as before.

For an overview of all samples please refer to Supplementary file 2: Table S1. Please refer also to Figure 1 for an overview of the developmental stages.

## Sequencing and assembly of the *S. exigua* genome

A dual sequencing approach was used for *de novo* assembly of the *S. exigua* genome sequence. In total ~100 Gb of raw Nanopore long read data (Oxford Nanopore Technologies, Oxford, UK) and ~73 Gb of raw Illumina 2x150nt short read data were generated. Long sequence read data was generated using the Oxford Nanopore Technologies platform. Prior to library preparation, HMW DNA was sheared to ~12.5 kb fragments using Covaris gTUBE (Covaris Inc., Woburn, MA, USA). Quality was checked on the Agilent TapeStation. Library preparation was done with the SQK-LSK109 1D ligation kit from Oxford Nanopore Technologies (ONT). Samples were sequenced using one run on an ONT MinION R9.4.1 flowcell and one run on an ONT PromethION R9.4.1 flowcell, respectively. Basecalling was done with Guppy v2.2.2 (ONT MinION) and v1.6.0 (ONT PromethION), respectively. Basecalled reads were used for further processing and assembly.

In addition to long sequence read data, short read data were generated using the Illumina NovaSeq 6000 system. Library preparation was done with the Nextera DNA Flex Library Prep Kit following manufacturers' protocol (Illumina Inc. San Diego, CA, USA) and quality was checked using the Agilent Bioanalyzer 2100 High Sensitivity DNA Kit (Agilent, Amstelveen, The Netherlands). The genomic paired-end (PE) library was sequenced with a read length of 2x150 nucleotides. Image analysis and basecalling was done by the Illumina pipeline. Please refer to Supplementary file 19: Table S10 for an overview of the DNA sequencing approach. All raw reads from the Illumina, MinION and PromethION sequencing runs were submitted to the NCBI SRA database under accession number PRJNA623582 under sample number SAMN14550570.

To assemble the *S. exigua* genome sequence, only long sequence read data were used. First, all reads with a quality score lower than  $qv=7$  were removed from the long sequence read data set. Then, the SEA program (Future Genomics Technologies BV, Leiden, The Netherlands) (Jansen et al., 2017) was used to prepare seed sequences from the longest reads. In total ~30x estimated coverage of the longest reads were then aligned to these seeds. Reads, alignments and seed files were used to run Tulip v. 1.0.0 (Future Genomics Technologies BV, Leiden, The Netherlands) (Jansen et al., 2017) to obtain an assembly. The assembly results were used to further optimise the assembly parameters. After this optimisation, the total size of the assembled genome was 419 Mb, which was divided over 946 contigs (largest contig = 4.08Mb) with a contig N50 of 1.10 Mb. To further optimise the genome assembly, Racon (Loman et al., 2015) was used (two rounds) to correct mistakes in the assembly and then two rounds of Pilon polishing (Goodwin et al., 2015; Walker et al., 2014) were used to polish the assembly based on the genomic Illumina reads and to reach a high accuracy of the *de novo* assembly that was the basis for genome annotation. The final genome assembly was submitted to the NCBI GenBank database and is available under accession JACEFF000000000, version JACEFF010000000 is used in this study. As a quality check a BUSCO v. 3.0.2 analysis was done on the polished *de novo* assembly using the “insecta\_odb9” dataset.

## Sequencing the developmental transcriptome of *Spodoptera exigua*

Following the Illumina Truseq stranded mRNA library prep protocol (150-750 bp inserts), we prepared 18 different indexed RNA-Seq libraries representing the different developmental stages (namely embryonic stage, early 1<sup>st</sup> instar larva, early 3<sup>rd</sup> instar larva, pupa, adult (female and male), and including three biological replicates per stage/sex (Supplementary file 2: Table S1.1). Libraries were sequenced on an Illumina NovaSeq 6000 system at an average of 13.4 million PE2x150nt reads (6.9-22.5 million reads) per sample at Future Genomics Technologies BV, Leiden, The Netherlands. For an overview of the number of raw reads per samples please refer to Supplementary file 2: Table S1.3. The sequencing reads were quality checked using FastQC v. 0.10.1 (Andrews, 2010). Adapter sequences were removed and quality-filtered using Trimmomatic v. 0.36 (Bolger et al., 2014), with parameters set: TruSeq3-PE-2.fa:2:30:10, LEADING: 5, TRAILING: 5, SLIDINGWINDOW:4:20, and removing all reads of less than 36 bp in length. All raw reads from the Illumina RNA sequencing approach were submitted to the NCBI SRA database under accession number PRJNA623582.

## Annotation of the *Spodoptera exigua* genome sequence

The assembled and polished genome was annotated using the maker3 pipeline (maker-3.01.02-beta). As a first step in this analysis a repeat library was constructed with RepeatModeler (RepeatModeler-open-1.0.11; -database *Spodoptera exigua*). This species-specific library was used in addition to the RepeatMasker library (Lepidoptera). For gene prediction, Augustus v. 3.3.2 was used which used the model from *heliconius\_melpomene*1 to find genes. As additional evidence for gene models the protein sequences for the family of the Noctuidae were extracted from UniProt (accessed March 7, 2019). Also the RNA-Seq data sets of our 18 *S. exigua* samples were used as supporting evidence. This data set was first assembled using the De Bruijn graph-based *de novo* assembler implemented in the CLC Genomics Workbench version 4.4.1 (CLC bio, Aarhus, Denmark). The available *S. exigua* mRNA nucleotide data from NCBI Genbank (accessed March 7, 2019) was added to this data. After running the pipeline, maker3 annotated a total of 18,477 transcripts. Gene annotations, predicted mRNA and proteins, and assemblies for gene annotations are also provided at the Dryad digital repository (<https://datadryad.org/stash/share/aUEyO49FeYl7cFZOE72MjLZnhV4N7enuapuxfSpL3cw>).

*Spodoptera exigua* proteins from the official gene set OGS v. 1.1 were further annotated using InterProScan (v. 5.36-75) with several approaches including Go Term annotation (Jones et al., 2014). Of the 18,477 transcripts, 16,718 transcripts retrieved annotations (Supplementary file 20: Table S11). Furthermore, the transcript OGS was used in a local BLASTX search v. 2.6.0 (Camacho et al., 2009) (max\_hsps 1, best\_hit\_overhang 0.1 and E-value cutoff  $\leq 1e-3$ ) against a locally constructed database of all Arthropoda protein sequences downloaded from the NCBI protein database (accessed, 31/01/2019). The translated proteins were additionally used in a BLASTP search v. 2.6.0 (Camacho et al., 2009) against the same Arthropoda database and parameters (Supplementary file 6: Table S3, Supplementary file 7: Table S4).

## Transcript expression quantification

To estimate transcript expression, reads of all samples from each developmental stage were separately mapped to the newly generated *S. exigua* genome (version JACEFFO10000000) using Bowtie2 v. 2.3.4 (Langmead & Salzberg, 2012). The isoform and gene abundance estimations were done using RSEM v. 1.3.0 (Li & Dewey, 2011). A raw (non-normalised) count matrix was created using the perl script 'abundance\_estimates\_to\_matrix.pl' implemented in the Trinity v. 2.5.1 package (Grabherr et al., 2011). The count matrix was cross-sample normalised using the 'calcNormFactors' function in edgeR v.3.20.8 (M. D. Robinson et al., 2010) (R v. 3.4.3) using trimmed mean of M values (TMM) (Robinson & Oshlack, 2010). See Supplementary file 21: Table S12 for the raw counts matrix of isoforms in the samples. The normalised count matrix was further filtered by abundance based on count-per-million values (CPM; to account for library size differences between samples) using edgeR v. 3.20.8 (M. D.

Robinson et al., 2010). Only genes with a minimum of 5 counts in at least two of the samples were considered expressed and retained in the dataset, see Supplementary file 23: Table S13.

To measure the similarity of the samples covering the developmental stages and to verify the biological replicates, we implemented the trinity-provided perl script “PtR”. The PCA plot is generated based on the raw non-normalised isoform count matrix which we centred, CPM normalised, log transformed and filtered using a minimum count of 10 (Supplementary file 24: Figure S10).

The differential expression analysis was performed using DESeq2 v. 1.18.1 (Love et al., 2014) as implemented in the Trinity package. Transcripts were considered differentially expressed with a minimal fold-change of four between any of the treatments and a false discovery rate (FDR) of  $p\text{-value} \leq 1e-3$ . The CPM and TMM normalised expression values of all differentially expressed transcripts were hierarchically clustered and cut at 50% using the Trinity provided script ‘define\_clusters\_by\_cutting\_tree.pl’. This resulted in 14 clusters of differentially expressed transcripts with similar expression patterns that were used in the cluster-specific Gene Ontology (GO) analysis. See Supplementary file 25: Table S14 for an overview of cluster membership of all 9,896 DE isoforms and Figure 2 and Supplementary file 5: Figure S3 for expression patterns.

GO analysis was performed using the Goseq package adjusting for transcript length bias in deep sequencing data (Young et al., 2010) and using the GO annotation retrieved from the Interpro annotation. See Supplementary file 26: Table S15 for overview of GO annotations within the clusters. GO terms were further summarised to generic GOSlim categories using the R package GOSTats (Falcon & Gentleman, 2007). For the identified DE genes, statistically over-represented GO terms in each cluster were identified and further summarised to generic GOSlim categories (Supplementary file 8: Table S5).

## Phylogenomic analyses and comparative genome analyses

We used BUSCO v. 4.0.5 applying the insecta\_odb10 as a reference lineage dataset (Seppey, Manni, et al., 2019) and comprising in total 1,367 BUSCOs, to extract single copy complete BUSCOs on the amino acid (aa) level for *S. exigua* and another 37 lepidopteran genomes (Supplementary file 27: Table S16).

For the phylogenomic analysis, first, amino acid (aa) sequences of single-copy BUSCO genes were separately aligned using MAFFT v. 7.305 (Katoh & Standley, 2013) using the L-INS-i algorithm. For the identification of putative ambiguously aligned or randomised multiple sequence alignment (MSA) sections, we used Aliscore v. 1.2 (Kück et al., 2010; Misof & Misof, 2009) on each MSA with the default sliding window size, the maximal number of pairwise sequence comparisons and a special scoring for gap-rich amino acid data (options -r and -e). After exclusion of the identified putative ambiguously aligned or randomised MSA sections with ALICUT v. 2.3 (Kück et al., 2010), the final MSAs were concatenated into supermatrices using FASconCAT-G v.

1.02 (Kück & Longo, 2014). The resulting dataset comprised 1,367 gene partitions and 687,494 amino acid positions.

Prior to the tree reconstruction, the best scoring amino-acid substitution matrix for each gene partition was selected with ModelFinder as implemented in IQ-TREE v. 1.6.12 (Kalyaanamoorthy et al., 2017). We restricted the search of the best fitting model to eight amino-acid substitution matrices appropriate for nuclear markers: DCMut (Kosiol & Goldman, 2005), JTT (Jones et al., 1992), LG (Le & Gascuel, 2008), Poisson, PMB (Veerassamy et al., 2003), VT (Muller & Vingron, 2000), and WAG (Whelan & Goldman, 2001). We additionally included the protein mixture model LG4X (Le et al., 2012), which accounts for FreeRate heterogeneity. Furthermore, we allowed testing the default rate heterogeneity types (E, I, G, I+G, and FreeRates: R) (Gu et al., 1995; Soubrier et al., 2012; Yang, 1994), with or without empirical rates (-F, -FU) as well as testing the number of rate categories (-cmin 4 -cmax 15). The best model for each gene partition was selected according to the best second-order or corrected Akaike Information Criterion (AICc) score (Hurvich & Tsai, 1989). Dataset and partition scheme including selected models are provided at the Dryad digital repository (<https://datadryad.org/stash/share/aUEyO49FeYl7cFZOE72MjLZnhV4N7enuapuxfSpL3cw>).

Phylogenetic relationships were inferred under the maximum likelihood (ML) optimality criterion as implemented in IQ-TREE v. 1.6.12 (Chernomor et al., 2016; Nguyen et al., 2015) using the best scoring amino-acid substitution matrix for each gene partition and the edge-proportional partition model allowing partitions to have different evolutionary rates (option -ssp). We performed 50 independent tree searches (25 searches with a random and 25 with a parsimony start tree). The resulting number of unique tree topologies was assessed with Unique Tree v. 1.9, kindly provided by Thomas Wong and available upon request. Node support was estimated via non-parametric bootstrapping of 100 bootstraps replicates in IQ-TREE and mapped onto the ML tree with the best log-likelihood.

We further scanned all these lepidopteran protein sets for several gene families associated with detoxification function, namely P450 monooxygenases (P450s), carboxyl- and cholinesterases (CCEs), UDP-glycosyltransferases (UGTs), glutathione S-transferases (GSTs), ATP-binding cassettes (ABCs). We identified the protein families of all proteins by running InterProScan v. 5.36-75 (-appl Pfam --goterms) (Jones et al., 2014), additionally we ran a local BLASTP against the UniRef50 database (<ftp.uniprot.org/pub/databases/uniprot/uniref/uniref50/uniref50.fasta.gz>; release version 31/07/2019, accessed 20/08/2019) using an e-value cutoff of 1e-3. Based on these annotations, genes were selected to belong to any of the gene families of interest if it had a match to one of the Uniref50 cluster terms or Pfam- or InterProScan identifiers (Supplementary file 28: Table S17). The number of detoxification gene members of the five main detoxification families were plotted for each species in a bubbleplot generated with ggplot2 (Wickham, 2016) (Figure 4).

## Comparative analysis of *Spodoptera*-specific genes

We used OrthoFinder v. 2.3.11 using default settings (Emms & Kelly, 2015) to identify homologs within the *Spodoptera* clade. We included the genome protein sequence files from three *Spodoptera* species: *S. exigua* (this study), *S. litura* (direct receival OGSv1 28/09/2019 from authors (Cheng et al., 2017)) and *S. frugiperda* ([ftp://ftp.cngb.org/pub/CNSA/CNP0000513/CNS0099235/CNA0003276/Sf\\_20190612ynM\\_v1.pep](ftp://ftp.cngb.org/pub/CNSA/CNP0000513/CNS0099235/CNA0003276/Sf_20190612ynM_v1.pep), accessed 20/09/2019; (Liu et al., 2019)). In addition, we included five closely related but diverse Lepidoptera species: *Heliothis virescens* ([ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/382/865/GCA\\_002382865.1\\_K63\\_refined\\_pacbio/GCA\\_002382865.1\\_K63\\_refined\\_pacbio\\_protein.faa.gz](ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/382/865/GCA_002382865.1_K63_refined_pacbio/GCA_002382865.1_K63_refined_pacbio_protein.faa.gz), accessed 20/09/2019; (Fritz et al., 2018)), *Helicoverpa zea* (<https://data.csiro.au/collections/#collection/CiCSIRO:23812v3>, accessed 21/08/2019; (Pearce et al., 2017)), *Helicoverpa armigera* (<https://data.csiro.au/collections/#collection/CiCSIRO:23812v3>, accessed 21/08/2019; (Pearce et al., 2017)), *Trichoplusia ni* ([ftp://www.tnibase.org/pub/tni/tni\\_protein\\_v1.faa.gz](ftp://www.tnibase.org/pub/tni/tni_protein_v1.faa.gz), accessed 20/09/2019; (Chen et al., 2019)), and *Bombyx mori* (<http://silkbases.ab.a.u-tokyo.ac.jp/cgi-bin/download.cgi>, accessed 20/08/2019; (International Silkworm Genome Consortium, 2018)).

We identified 119 orthogroups (OGs) containing sequences only from the three *Spodoptera* species (Supplementary file 10: Table S7.1). Of these 119 OGs, only seven OGs were differentially expressed in the larval stage (cluster 4, Supplementary file 10: Table S7.2). Of these seven OGs, three OGs were “uncharacterised” protein and four OGs were annotated as: nuclear complex protein (OG0013351), REPAT46 (OG0014254), trypsin alkaline-c type protein (OG0014208), and mg7 (OG0014260) (Supplementary file 10: Table S7) for which we performed gene tree analyses. For the gene tree analyses, we extended our dataset based on the original OrthoFinder run by including similar sequences from related species to additionally verify the lineage-specificity of these genes. Using the identified *S. exigua* sequences within the lineage-specific OGs as queries, we searched for close homologs using BLASTX (Bravo et al., 2019) against the NCBI protein database online (Sayers et al., 2020). Thus, the resulting datasets used to construct gene trees were compiled with some differences. The gene tree of nuclear pore complex proteins was composed of *Spodoptera* orthogroup sequences and all Lepidoptera nuclear complex DDB\_G0274915 proteins from the NCBI-nr database (accessed 2/10/2020, keyword “DDB\_G0274915”). The initial BLAST identifications of *Spodoptera*-specific orthogroup sequences showed high similarity with DDB\_G0274915-like nuclear pore complex proteins. For the remaining three datasets, we additionally included clusters of homologous genes from OrthoDB v. 10 (Kriventseva et al., 2019). For the REPAT protein dataset, we added the ortholog cluster (“16151at7088”) consisting of “MBF2” orthologs. MBF2 proteins are described to be homologs of REPAT genes in other Lepidoptera species, and have been therefore included (Navarro-Cerrillo et al., 2013). The REPAT protein gene tree dataset included all protein sequences from Navarro-Cerrillo et al. (2013). For a second REPAT tree, we only analysed sequences from the  $\beta$ REPAT class (Navarro-Cerrillo et al., 2013). For both, the trypsin and mg7 gene tree datasets we

included clusters of homologous genes from OrthoDB v. 10 based on the linked cluster to our closest BLAST hit via the online NCBI protein database. For the trypsin gene tree dataset, we added the ortholog cluster “118933at50557” consisting of “serine protease” orthologs. These homologous sequences were selected because the *S. litura* sequence (“SWUSL0076430”) from the *Spodoptera*-specific orthogroup formed a member of this group. All insect orthologs were included. Finally, the mg7 gene tree dataset included the ortholog group “15970at7088” from OrthoDB v. 10 (accessed 15/09/2020), because the *S. litura* sequence (“SWUSL0113290”) was an ortholog member. For a second tree we included all genes derived from He et al. (2012), where the expression of mg7 in the midgut of *S. litura* was studied and homologs in related lepidopteran species were analysed. Finally, we searched for potential paralogs of all target genes in the protein sets of *S. exigua*, *S. litura* and *S. frugiperda* using BLASTP (max\_hsps 1, best\_hit\_overhang 0.1 and E-value cutoff  $\leq 1e-5$ ) with NCBI-BLAST+ v. 2.6.0 (Camacho et al., 2009) against a local BlastDB of above gene tree datasets of nuclear pore complex, REPAT, trypsin and mg7 proteins.

For all genes, sequences were aligned using MAFFT v. 7.471 with the L-INS-i method and default settings (Katoh & Standley, 2013). Gene trees were reconstructed using IQ-TREE v. 1.6.12 (Chernomor et al., 2016; Nguyen et al., 2015) using the maximum likelihood method and implementing bootstrap with 100 replications. The preferred model was applied as based on the model selection (Kalyaanamoorthy et al., 2017). For the nuclear pore complex gene tree, the best-fit model was “WAG+F+G4”, for REPAT including both  $\alpha$ REPAT and  $\beta$ REPAT proteins “WAG+F+R4”, for the gene tree consisting only  $\beta$ REPAT proteins “VT+G4”, for the trypsin gene tree “WAG+F+R5” finally for both mg7 based gene trees “LG+G4”.

The gene trees were rooted dependent of included species and gene composition, aiming for earliest branching genes or species, e.g. by selecting earliest branching lineages from Kawahara et al. (2019). For the nuclear pore complex protein gene tree, *Papilio xuthus* was used for rooting since it branched early within Papilionidae (Kawahara et al., 2019). For the REPAT gene tree, we used the same approach as Navarro-Cerrillo et al. (2013), which rooted the tree using the REPAT-like27 and REPAT-like28 cluster. However, for the limited REPAT gene tree only including  $\beta$ REPAT class genes, we rooted using group V of the  $\beta$ REPAT class according to the first group branching off (Navarro-Cerrillo et al., 2013). The trypsin tree was rooted using the branch giving rise to a Hymenoptera specific cluster. Finally, the mg7 gene trees were rooted using either *Choristoneura fumiferana* (Tortricidae) (mg9 cluster) or, if absent, *Amyelois transitella* (Pyralidae) (Kawahara et al., 2019).

## Acknowledgements

We thank Els Roode and the late Hanke Bloksma for help with the *S. exigua* rearing and sample collection. We also thank Corné van der Linden for providing *S. exigua* photographs. We thank Entocare for their support to this project.



## Supplementary data

All Supplementary Data is uploaded to the 4TU Centre for Research Data repository and available online: <https://figshare.com/s/Oe12d9153800dcdf8927>.

Further genome datasets and other datasets generated during the current study are provided at the Dryad digital repository: <https://datadryad.org/stash/share/aUEyO49FeYl7cFZOE72MjLZnhV4N7enuapuxfSpL3cw>.

**SUPPLEMENTARY FILE 1: FIGURE S1** | BUSCO assessment of *Spodoptera exigua* compared to other Lepidoptera genomes for ortholog presence and copy number. The relationships are according to Figure 4.

**SUPPLEMENTARY FILE 2: TABLE S1** | Overview of developmental stages used for the RNA-Seq approach and sequencing results.

**TABLE S1.1** | Developmental stages and samples used for the RNA-Seq approach.

**TABLE S1.2** | Quality of developmental stages and samples used for the RNA-Seq approach.

**TABLE S.3** | Details on RNA-Seq run results.

**SUPPLEMENTARY FILE 3: TABLE S2** | Comparison of differential gene expression between *Spodoptera exigua* developmental stages.

**SUPPLEMENTARY FILE 4: FIGURE S2** | Clustered heatmap showing Pearson's correlation for pairwise sample comparisons based on differentially expressed genes.

**SUPPLEMENTARY FILE 5: FIGURE S3** | Visualisation of expression patterns with the number of genes per cluster.

**SUPPLEMENTARY FILE 6: TABLE S3** | BLASTX annotation report of *Spodoptera exigua* proteins from the official gene set OGS v1.1 using local Arthropoda database.

**SUPPLEMENTARY FILE 7: TABLE S4** | BLASTP annotation report of *Spodoptera exigua* proteins from the official gene set OGS v1.1 using local Arthropoda database.

**SUPPLEMENTARY FILE 8: TABLE S5** | Go-slits and frequency in respective clusters.

**SUPPLEMENTARY FILE 9: TABLE S6** | Proteins annotated using InterProScan for the five main detoxification gene families P450 monooxygenases (P45Os), carboxyl- and cholinesterases (CCEs), UDP-glycosyltransferases (UGTs), glutathione S-transferases (GSTs), ATP-binding cassettes (ABCs).

**SUPPLEMENTARY FILE 10: TABLE S7** | Table S7.1: List of 119 *Spodoptera*-specific orthogroups (OG) with gene IDs.

**TABLE S7.2** | Within the 119 OGs, seven *Spodoptera exigua* genes were differentially expressed and clustered in expression cluster 4.

**SUPPLEMENTARY FILE 11: TABLE S8** | Read counts of all samples for the seven selected genes based on the ortholog search and specificity for cluster 4.

**SUPPLEMENTARY FILE 12: FIGURE S4** | Gene tree of the nuclear pore complex sequences consisting of: the *Spodoptera*-specific orthogroup sequences (OG0013351) and the Lepidoptera sequences from the protein group 'DDB\_G0274915' as derived from the NCBI protein database. The sequences from the *Spodoptera*-specific orthogroup are marked by the red asterisks, the red frame highlights the *Spodoptera*-specific clade, the red arrow indicates the specific *Spodoptera exigua* differentially expressed gene from expression cluster 4. All sequences derived from the *S. exigua* genome (this study) are underlined. Sequences derived from the protein group 'DDB\_G0274915' are given their full species names and gene ID.

**SUPPLEMENTARY FILE 13: FIGURE S5** | Gene tree of the mg7 sequences (reduced dataset) consisting of: the *Spodoptera*-specific orthogroup sequences (OG0014260) and sequences from OrthoDB cluster '1597Oat7088'. The sequences from the *Spodoptera*-specific orthogroup are marked by the red asterisks, the red frame highlights the *Spodoptera*-specific clade, the red arrow indicates the specific *Spodoptera exigua* differentially expressed gene from expression cluster 4. All sequences derived from the *Spodoptera exigua* genome (this study) are underlined. Sequences derived from the OrthoDB cluster are given their full species names.

**SUPPLEMENTARY FILE 14: FIGURE S6** | Gene tree of the mg7 sequences (extended dataset) consisting of: the *Spodoptera*-specific orthogroup sequences (OG0014260), sequences from OrthoDB cluster '15970at7088' and all 'mg' gene sequences from He et al. (2012). The sequences from the *Spodoptera*-specific orthogroup are marked by the red asterisks, the red frame highlights the clade, the red arrow indicates the specific *Spodoptera exigua* differentially expressed gene from expression cluster 4. All sequences derived from the *Spodoptera exigua* genome (this study) are underlined. Sequences derived from the OrthoDB cluster are given their full species names only. Genes from He et al. (2012) are given their species names, followed by their gene ID.

**SUPPLEMENTARY FILE 15: FIGURE S7** | Gene tree of  $\beta$ REPAT sequences (reduced dataset) consisting of the *Spodoptera*-specific orthogroup (OG0014254), sequences from OrthoDB cluster '16151at7088 - "MBF2"', and  $\beta$ REPAT gene sequences from Navarro-Cerrillo et al. (2013). The sequences from the *Spodoptera*-specific orthogroup are marked by the red asterisks, the red frame highlights the *Spodoptera*-specific clade, while the red arrow indicates the specific *Spodoptera exigua* differentially expressed gene from expression cluster 4. All sequences derived from the *S. exigua* genome (this study) are underlined. Sequences derived from the OrthoDB cluster are given their full species names and cluster ID "MBF2". Genes from Navarro-Cerrillo et al. (2013) are given the initials of the species names, followed by the REPAT identifier.

**SUPPLEMENTARY FILE 16: FIGURE S8** | Gene tree of the REPAT sequences (extended dataset) consisting of: the *Spodoptera*-specific orthogroup sequences (OG0014254), sequences from OrthoDB cluster '16151at7088 - "MBF2"', and REPAT gene sequences derived from Navarro-Cerrillo et al. (2013). Both  $\beta$ REPAT and  $\alpha$ REPAT sequences are included in this gene tree. The sequences from the *Spodoptera*-specific orthogroup are marked by the red asterisks, the red frame highlights the *Spodoptera*-specific clade, the red arrow indicates the specific *Spodoptera exigua* differentially expressed gene from expression cluster 4. All sequences derived from the *S. exigua* genome (this study) are underlined. Sequences derived from the OrthoDB cluster are given their full species names and cluster ID "MBF2". Genes from Navarro-Cerrillo et al. (2013) are given the initials of the species names, followed by the REPAT identifier.

**SUPPLEMENTARY FILE 17: TABLE S9** | BLASTP annotated REPAT genes in the *Spodoptera exigua* protein set using reference REPAT sequences from Navarro-Cerrillo et al. (2013).

**SUPPLEMENTARY FILE 18: FIGURE S9** | Gene tree of the trypsin sequences consisting of: the *Spodoptera*-specific orthogroup sequences (OG0014208) and sequences from OrthoDB cluster '118933at50557'. The sequences from the *Spodoptera*-specific orthogroup are marked by the red asterisks, the red frame highlights the *Spodoptera*-specific clade, the red arrow indicates the specific *Spodoptera exigua* differentially expressed gene from expression cluster 4. All sequences derived from the *Spodoptera exigua* genome (this study) are underlined. Sequences derived from the OrthoDB cluster are given their full species names and gene ID.

**SUPPLEMENTARY FILE 19: TABLE S10** | Overview of DNA-Seq approach and sequencing results.

**TABLE S10.1** | Results of the Oxford Nanopore long read sequencing run

**TABLE S10.2** | Results of the Illumina short read sequencing run

**TABLE S10.3** | Results of the trimming and cleaning steps.

**SUPPLEMENTARY FILE 20: TABLE S11** | InterProScan annotation report of *Spodoptera exigua* proteins from the official gene set OGS v. 1.1.

**SUPPLEMENTARY FILE 21: TABLE S12** | Raw read counts matrix of isoforms in the different samples.

**SUPPLEMENTARY FILE 23: TABLE S13** | CPM, TMM cross-sample normalised and filtered count matrix.

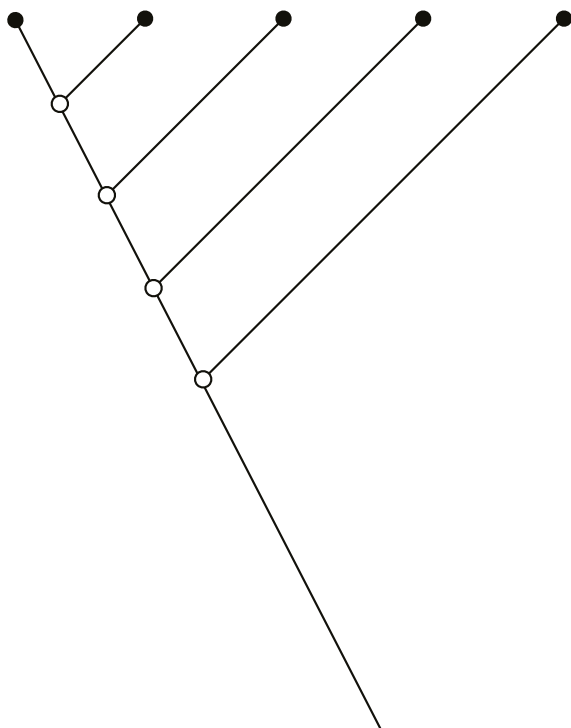
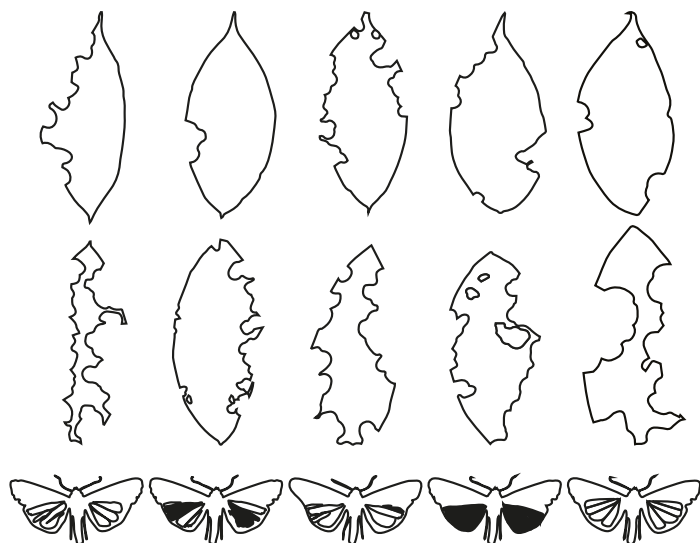
**SUPPLEMENTARY FILE 24: FIGURE S10** | PCA plot showing sample relationships based on the filtered, CPM-TMM and log<sub>2</sub> transformed centred dataset.

**SUPPLEMENTARY FILE 25: TABLE S14** | Genes with included in the 14 clusters.

**SUPPLEMENTARY FILE 26: TABLE S15** | Cluster membership with GO annotations

**SUPPLEMENTARY FILE 27: TABLE S16** | Overview of Lepidoptera genomes and protein sequence files including source location and accession date used for the phylogenomic analyses.

**SUPPLEMENTARY FILE 28: TABLE S17** | Gene family identifiers, using InterProScan identifiers and UniRef cluster terms, used to identify putative gene members from families P450, CCE, UGT, GST and ABC.



# Chapter 6

General discussion

Thijmen Breeschoten

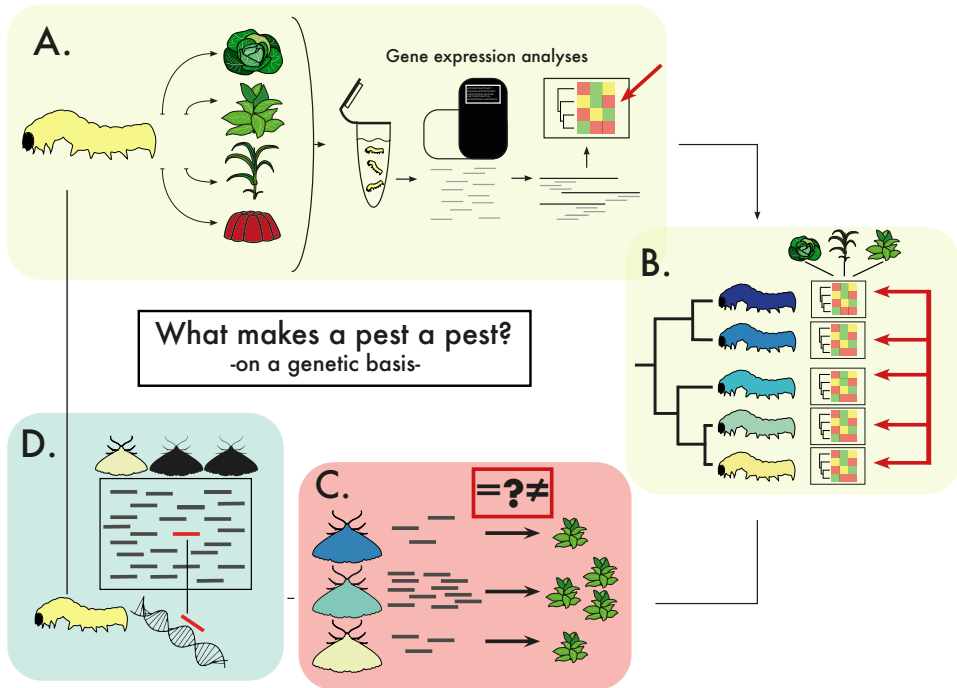


My thesis aimed to elucidate the genomic and evolutionary basis of polyphagous pest insect species. By studying multiple facets of polyphagous pest formations in Lepidoptera I aimed to, on a genetic basis, address the question: ‘what makes a pest a pest?’. The different chapters of my thesis each stand alone, but build upon one other by focusing on different aspects of polyphagy in Lepidoptera and polyphagous pest formations in Noctuidae moths (Figure 1).

Firstly, I examined the transcriptional dynamics and gene family evolution patterns related to herbivory to understand the factors that contribute to the success of polyphagous insect pest formations. The gene expression of major polyphagous cutworm moths, Noctuidae, feeding on different host plant species was compared. By linking herbivore success to gene expression patterns, I identified the transcriptional plasticity related to polyphagy and potential pest formations in Noctuidae moths. Further, I performed comparative genomics to study the evolutionary patterns of gene families involved in plant feeding, such as detoxification related processes. Additionally, I presented the genome assembly and developmental transcriptome of the polyphagous pest moth, the beet armyworm, *Spodoptera exigua*, and performed comparative genomic analyses within Lepidoptera. The outcome of the genomic and transcriptomic analyses was used to identify lineage-specific candidate genes for targeted pest control of *Spodoptera* larvae. Finally, in this General discussion I discuss and reflect on my findings, discuss the applied methods and analyses, and provide future perspectives.

## Main findings and conclusions

This thesis started with the transcriptional analyses of the beet armyworm, *Spodoptera exigua* feeding on a control diet and different host plant species in **Chapter 2**. Larvae feeding on the different plant species showed divergent transcriptional responses. With feeding assays, I examined the herbivore success of larvae on each plant. Herbivore success, as an indicator for ‘specialisation’ on a particular host and quantification for level of pest formation, was highest on maize (*Zea mays*) compared to cabbage (*Brassica oleracea*) and tobacco (*Nicotiana tabacum*). I found gene expression patterns, containing uniquely upregulated genes, for larvae feeding on *B. oleracea* and *N. tabacum* but not for *Z. mays*. The specific upregulated gene activity employed for feeding on host plant species with low herbivore success implied a specific response for suboptimal hosts while a general and non-specific gene activity was employed for a host plant with optimal herbivore success. I concluded that *S. exigua* employed transcriptional plasticity, which included genes involved in detoxification and digestion, to enable a polyphagous feeding style.



**FIGURE 1** | Overview of the different approaches implemented in this thesis aiming to answer the main research question, ‘What makes a pest a pest, on a genetic scale?’. Each chapter stands alone, but build upon each other by focussing on different aspect of polyphagy in Lepidoptera and polyphagous pest formations in Noctuidae moths. (A.) Research Chapter 2. Study of the gene expression of the beet armyworm, *Spodoptera exigua*, in feeding on different host plants in order to study the transcriptional plasticity employed by a polyphagous Noctuidae species. (B.) Research Chapter 3. Multi-species comparison of the transcriptional plastic response of five major polyphagous Noctuidae species feeding on different host plant species under a comparative phylogenetic framework. I studied the transcriptional mechanisms which might contribute to polyphagy and herbivore success and evaluated if successful herbivory is determined by a shared or lineage specific transcriptional response. (C.) Research Chapter 4. Comparative genomics of 37 lepidopteran genomes to study genomic evolution patterns linked to polyphagy in butterflies and moths. I tested if gene family evolution, in terms of gene gains and losses, can be associated with a polyphagous feeding style. (D.) Research Chapter 5. I presented the genome of the polyphagous Noctuidae beet armyworm, *Spodoptera exigua*. By performing developmental transcriptional analysis in combination with comparative genomics, I identified potential lineage-specific candidate genes for targeted RNA interference (RNAi)-based control of *Spodoptera* pest species.

In **Chapter 3**, I extended my transcriptional analysis of insect pests. I examined and compared the transcriptional plasticity of five major polyphagous Noctuidae cutworm moth species (*Trichoplusia ni*, *Autographa gamma*, *Mamestra brassicae*, *Spodoptera littoralis*, and *S. exigua*) feeding on the three different host plant species and a control. The major polyphagous herbivores relied on the deployment of diverse and flexible gene expression, including genes involved in plant feeding such as digestion and detoxification, for their polyphagous feeding style. The feeding assays showed that

the moths had species-specific herbivore success levels for the different host plant species. In Chapter 2, I showed that specific upregulated gene activity was employed by *S. exigua* for feeding on host plants with low herbivore success. Comparative analyses with the other moth species showed host plant-specific gene expression patterns independent of herbivore success. The transcriptional plastic pattern was moth species-specific and showed the importance of a diverse and flexible response for host plant feeding.

I compared the expression of homologous genes between the moths in this study. I implemented a phylogenomic framework to distinguish shared expression patterns related to high herbivore success, potential pest formations, from patterns shared under influence of evolutionary relatedness. Homologous genes were identified, that showed similar expression in moths feeding on plants with high herbivore success levels, indicating convergence. The shared expression indicated the potential role of these genes in reaching high herbivore success on specific plants. Moreover, by implication of the phylogenomic framework, I identified homologous genes that were similarly expressed for feeding on specific host plant species within clades of closely related moths. The common expression pattern was likely caused by shared evolutionary histories and could therefore not be associated with herbivore success alone. The similarly expressed homologous genes within related moths, shows the influence of lineage-specificity on the adaptive response to specific host plants. In short, polyphagous moths displayed a complex species-specific transcriptional plastic response while feeding on different host plants. Two groups of homologous genes were identified: shared among related species (lineage-specific), or independent of evolutionary history but linked to high herbivore success (pest formation) on specific plants, indicating convergence.

The importance of various ubiquitous gene families for plant feeding became apparent in the analyses of Chapter 3. The expansion of gene families involved in plant feeding such as detoxification and digestion processes, has been hypothesised to enable polyphagy. To test the suggested association, I compared the gene counts of various gene families involved in adaptations to plant feeding and correlated gene family sizes to diet breadth in **Chapter 4**. I analysed 37 lepidopteran genomes from 35 different species. For each species I calculated the phylogenetic diversity (PD) and specialised metabolite content (FMD) of the host plants within each lepidopteran diet to quantify the level of polyphagy. I found that gene family expansions are not limited to polyphagous species and that total number of genes is not necessarily highest in butterflies and moths with larger host ranges. However, for the detoxification gene families carboxyl- and cholinesterase (CCE) and glutathione S-transferase (GST), a significant positive correlation between gene family size and host range (PD and FMD values) across Lepidoptera was found.

Further, I studied the evolutionary rates of gene family sizes in terms of gene losses and gains. A higher rate for gene family contraction was observed in Lepidoptera, which is a potential consequence of fractionation after a duplication event. The



evolutionary rates varied across lepidopteran families, with a higher expansion rate for gene families involved in detoxification within the Noctuidae, potentially due to the large number of polyphagous species within this family. However, a higher expansion rate was also observed for the Pieridae consisting of primarily monophagous species. I concluded that gene family expansions of gene families involved in plant feeding are species-specific, occurring in both monophagous and polyphagous species, but the size of particular gene families, CCE and GST, are positively correlated with level of polyphagy across Lepidoptera.

In the final research chapter, **Chapter 5**, I presented the genome of the polyphagous Noctuidae beet armyworm, *Spodoptera exigua*. By performing developmental transcriptional analysis in combination with comparative genomics across Lepidoptera, I identified seven *Spodoptera*-specific genes expressed during the larval phase. I was able to identify four genes putatively involved in transport, pathogen response or digestion. *Spodoptera* is arguably the most destructive insect genus to agriculture. The genes presented may serve as candidates for targeted RNA interference (RNAi) in order to control the larvae of *Spodoptera* that are most damaging to plants.

## Methodology - important considerations and system set-up

Throughout my thesis I implemented various study systems, set-ups and made various considerations to appropriately analyse the genomic/transcriptomic data. Firstly, in the comparative study focusing on the transcriptional dynamics of polyphagy and successful herbivory (Chapters 2 and 3), I implemented feeding assays. During the feeding assays, eggs were placed on entire living plants and larvae were allowed to develop and feed until collection for measurements and sequencing. Plants alter or increase their production of specialised metabolites upon herbivory (Kant et al., 2015; Karban, 2020; Karban & Myers, 1989). The use of entire plants is important in studying the natural response of larvae to plant defences, e.g. for the development of efficient and targeted pest control strategies (Després et al., 2007). I was interested to capture the transcriptional response of naturally feeding larvae coping with all defences employed by host plants. Examining the insect response by artificially testing the plant defences or forcing larvae to feed or stay on specific plant parts, e.g. by use of clip cages, will likely influence the results.

However, due to 100% mortality of *S. exigua* larvae in the first hours after hatching on *N. tabacum*, I slightly modified the set-up. Observations indicated that hatched larvae stuck to secretory glandular hairs, making the larvae immobile and eventually causing death. For the modified and final set-up, eggs were placed on detached leaves and larvae were offered fresh leaves until reaching the second larval stage. Then, larvae were relocated to an entire-plant set-up, restoring the natural plant-herbivore interactions, upon collection in the third larval phase. The updated approach ensured that the larvae were facing the *N. tabacum* defences at all times but avoided the

increased levels of excretion at the more sensitive early larval stage. More important, the natural dynamics of plant-insect interactions, in the period before larval collection, was restored.

Further, I compared transcriptional dynamics by establishing and utilising a comparative phylogenomic framework of the five closely related Noctuidae moth species. Any two species differ and thus a multi-species comparison of more than two species is of importance in order to make any conclusions and do reliable observations (Ali & Agrawal, 2012). Examining the moths in the context of a comparative phylogenomic framework, expression patterns related to herbivore success, pest-level, could be studied independent of shared evolutionary histories.

Previous studies showed the importance of implementing a comparative phylogenomic framework in gaining insight into evolutionary origins of insect detoxification mechanisms and to unravel the patterns of adaptive evolution (Edger et al., 2015; Fischer et al., 2008; Nallu et al., 2018; Wheat et al., 2007). The absence or wrongly incorporation of evolutionary relationships in a multi-species comparison can lead to misleading and/or erroneous conclusions (Dunn et al., 2018). The explicit phylogenetic framework applied in Chapter 3 made it possible to identify patterns caused by shared evolutionary histories which, as a consequence, could not exclusively be linked to herbivore success and pest formations. Studying and interpreting gene expression data in both a biological and evolutionary meaningful way is achieved by analysing and collecting similar data from closely related species in a multi-species comparison implementing a phylogenetic perspective (Dunn et al., 2013). By implementation of a phylogenomic framework I was able to take phylogenetic relatedness into account. Incorporating information on the evolutionary relationships of species is of great importance in a comparative study that aims to evaluate if successful herbivory, and thus pest formations, is determined by a shared or lineage/species-specific transcriptional response.

Furthermore, the expression of homologous genes between the moth species was compared. One of the main challenges in phylogenetic comparative analyses of expression patterns is normalisation of the data to allow for comparison between species (Dunn et al., 2013). Normalisation of expression data, by removing systematic variation and reduction of noise, is needed because read counts are proportional to expression but impacted by other effects that likely differ across genes and species, like gene length (Conesa et al., 2016). In the case of normalisation within samples (e.g. by applying 'TPM' normalisation) or within single species (e.g. by applying 'TMM' normalisation) much progress has been made in the development of methods and how to correctly apply them (Bolstad et al., 2003; Robinson & Oshlack, 2010; Zhao et al., 2020). But normalisation across species is still challenging, since not only the total read counts but also the variation in gene number and gene lengths need to be considered (Dunn et al., 2013; Zhou et al., 2019). A simple solution I adopted in Chapter 3 was the use of expression ratios. By using ratios I avoided the direct use of count data and the need for normalisation (Dunn et al., 2013). I used Fold Change

(FC) to compare expression levels of orthologous genes between the moths. For each moth species, I used the difference in FC of orthologous genes between feeding on the control diet and each of the host plant species. I was interested in the expression response of larvae coping with the plant-specific defences. The plant dependent gene expression is expected to be absent or lower in larvae feeding on the control food, reflected by the FC. I implemented the phylogenetic ANOVA method to detect genes with increased or decreased ratios of expression between the moth species while accounting for phylogenetic relationships (Rohlf & Nielsen, 2015). However, the phylogenetic signal proved too weak, as visualised by the hierarchical clustering which showed diet-based rather than phylogeny-based clustering of the data. I implemented a method developed by Kristiansson et al. (2013), based on the Fisher's combined probability test, a widely used statistical method for meta-analysis (Fisher, 1948; Kristiansson et al., 2013). The method relies on comparison of significance levels of expression differences between samples across species. Similar to the FC values as explained above, these significance levels are based on the gene expression comparison of larvae feeding on the control food compared to larvae feeding on any of the host plant species. Instead of using the FC, the method compares the level of significance in expression differences in terms of  $p$ -values, and circumvents the need for normalisation. Moreover, the method takes homology into account and compares expression data from genes with any number of orthologs and paralogs (Kristiansson et al., 2013). Taking homology into account allows for a much larger number of compared genes between the moth species, because the use of strict orthologs and exclusion of missing values for the phylogenetic ANOVA reduced the dataset considerably (9,674 vs 3,369).

Finally, throughout my thesis I put a strong focus on the five larger and well-known gene families (P450, CCE, GST, UGT and ABC) involved in the detoxification pathway of plant specialised metabolites in herbivorous insects. The importance of these families in insects feeding on their host plants has been shown repeatedly. These gene families likely are just one aspect of the adaptation and specialisation of herbivores that evolved as part of the response to plant defences (Birnbaum & Abbot, 2020; Heidel-Fischer & Vogel, 2015). Still, the novel aspects and comparative approaches implemented in my thesis allowed a focus on the gene families involved in detoxification. Nonetheless, the complete plastic transcriptional response of the moth larvae feeding on different host plants is studied and overall gene activity is examined.

## Polyphagy in Lepidoptera

Monophagous species likely form pests on species within, or highly similar to, their host range (Ryan et al., 2019; Sarfraz et al., 2005; Sikkink et al., 2017). The genetic response and ways to cope with the encountered plant defences are expected to be similar for the 'common' host plant as well as for the newly attacked host. The

detoxification strategy of monophagous species likely needs less flexibility, since plants within the accepted host range will rely on similar defence strategies (Li et al., 2004).

In contrast, polyphagous species need to cope with a large variety of different plant defence strategies employed by plants across their host range. They likely employ a diversity of detoxification strategies (Heckel, 2018). Since a polyphagous niche is more heterogeneous and broader than the niche of a monophagous species, more genomic flexibility is expected (Greenspan, 2001; Sexton et al., 2017). The expansion of gene families involved in plant feeding processes, such as detoxification, is often linked to level of host plant breadth in polyphagous species (Chen et al., 2016; Cheng et al., 2017; Gouin et al., 2017; Grbić et al., 2011; Xu et al., 2016). Gene family expansions can be the result of a variety of origins, including neofunctionalisation of gene copies and lateral gene transfers (Heidel-Fischer et al., 2019; Wybouw et al., 2016). Supporting evidence was found for the association between gene family size of detoxification families P450, CCE and GST and feeding preference of various species in Diptera and Hymenoptera (Rane et al., 2016). Highest gene counts were especially found in omnivorous species with a larger and complex diet breadth. A similar association was present in some species of Coleoptera and Orthoptera but absent in seven lepidopteran species compared, ranging in diet breadth (Rane et al., 2016).

More recently, a large number of new Lepidoptera genomes became available, which allowed me to test the correlation between gene family expansions and polyphagy using a larger set of species (Chapter 4). Expansions were present in both polyphagous and monophagous species and expansion rates did not differ much in families represented by primarily polyphagous or monophagous butterflies and moths. However, a positive correlation was found between expansions of gene families CCE and GST and host plant breadth. The expansions of these gene families may enable feeding on a more complex and larger host breadth in polyphagous Lepidoptera.

However, not all members of the gene families I focused on in Chapter 4 are involved in processes related to plant use and specialised metabolite detoxification. Various members of the gene families P450, CCE, GST, UGT and ABC are involved in a variety of roles, such as detoxification of endogenous compounds, intracellular transport and biosynthesis of hormones (Bock, 2016; Dermauw & Van Leeuwen, 2014; Feyereisen, 2006; Montella et al., 2012; Shi et al., 2012). By looking at total gene family size, I gave equal weighting to all gene members with varying functions. Correlations between expansion size and host breadth that may exist in parts of the gene family may get overlooked. For example, in subfamilies of P450, such as CYP6B, CYP6AB, CYP6AE, CYP9A and CYP9G that are repeatedly reported to be involved in xenobiotic detoxification show high diversification levels among Lepidoptera (Calla et al., 2017). Furthermore, besides gene counts and expansion rates, the regulation of gene expression forms a major determinant of the contribution to herbivory, polyphagy and potential pest formations as is shown in Chapters 2 and 3 of this thesis.

Polyphagous species showed high levels of transcriptional plasticity in response to plant feeding (Govind et al., 2010; Schweizer et al., 2017), likely due to the heterogeneity of their niche (Birnbaum & Abbot, 2020). The transcriptional analyses presented in Chapter 2 and 3 showed the highly plastic, species-specific expression pattern of moths feeding on different host plant species. For example, even species within the same genus, *Spodoptera*, showed two opposite expression patterns in response to feeding on *Z. mays* for which both species showed high pest potential. Larvae of *S. littoralis* deployed a large set of upregulated genes, while no specific upregulation of genes was observed for *S. exigua*.

I assumed the absence of plant-specific upregulated expressed genes indicated the importance of a general and non-specific gene activity for host plant feeding. However, by doing so, the constitutive gene expression was assumed of importance in absence of a plastic response and the difference across species was not studied. In fact, studies assessing the difference between plastic and constitutive expression and the importance for herbivory are rare (Birnbaum & Abbot, 2020). In the case of genetic assimilation, transcriptional plasticity that helped in surviving on novel host plants becomes constitutively expressed by selection (Crispo, 2007). In a comparative study of populations of the Sonoran Desert fly, *Drosophila mettleri*, on native and novel host plants, only four genes were found demonstrating genetic assimilation while the vast majority of genes (5,803) showed high plasticity levels without signs of adaptive divergence (Hoang et al., 2015). Further, 64 genes showed patterns reflecting the Baldwin effect; when directional selection is further expected to alter plasticity in the direction of the optimum (Crispo, 2007; Hoang et al., 2015). The results of the study showed that the majority of differentially expressed genes were plastic without divergence, even in a population adapted to a derived host, indicating the importance of transcriptional plasticity in host use and shifts (Birnbaum & Abbot, 2020; Hoang et al., 2015). The *D. mettleri* study allowed comparison of transcriptional plasticity after long-term host adaptation. With the experimental set up I implemented it was not possible to look at long-term adaptation effects. The parental generation of all moths in the comparison were raised on the artificial control food and their offspring was introduced to the host plant species. However, differences between short-term and long-term adaptations are expected since difference were found in polyphagous Acari (Dermauw, Wybouw, et al., 2013; Wybouw et al., 2015) and Coleoptera (Müller et al., 2017). However, my main focus was transcriptional dynamics of potential pest formations in polyphagous Noctuidae. Arguably, polyphagous species forming pests can be quickly developing and adapting (Kirk et al., 2013), thus it is short-term transcriptional adaptations that are of particular interest here.

## Transcriptional plasticity related to pest formations

Polyphagous species are often linked to pest formations and insecticide resistance. Insecticide resistance must be seen as a specific case of insect adaptation to toxic compounds (Alyokhin & Chen, 2017). Indeed, a majority of arthropods on the top list of most resistant species are considered polyphagous species (Sparks & Nauen, 2015). Gordon (1961) was the first to argue that insecticide resistance in herbivorous insects could be a result of utilising existing enzymes that evolved for the detoxification of plant specialised metabolites. Thus, the evolved detoxification response of polyphagous herbivores to the range of plant toxins preadapted them to pesticides. Various studies indeed supported the link between host plant adaptation and pesticide resistance (e.g. Ahmad, 1986; Zeng et al., 2007). Moreover, by implementing a genomics approach, specific paralogs within detoxification gene families were identified to detoxify pesticides (e.g. Cheng et al., 2017; Joußen et al., 2012; Wang et al., 2017b).

However, in a recent review Dermauw et al. (2018) re-analysed evidence for the ‘pre-adaptation’ hypothesis. They concluded that rather the intense and long-term use of pesticides against important polyphagous pest species likely resulted in field-evolved resistance, and a biased view on insecticide adaptation. Nevertheless, the evolutionary biology of both insecticide resistance and host plant adaptation is based on the same selective forces (Dermauw et al., 2018). Studying the genetic basis of herbivory and pest formations are crucial to understand and develop sustainable ways of coping with herbivorous insects forming pests. The objective of my thesis, studying the genetic basis of polyphagous pest formations, was achieved by comparison of transcriptional dynamics involved in plant use and successful herbivory.

Similar expressed homologous genes in Noctuidae species feeding on host plants with high herbivore success indicated that potential pest formations were associated with the expression of specific genes. Many of these homologous genes showed to be members of ubiquitous gene families involved in a variety of processes including detoxification, digestion and metabolism. The shared expression of the homologous genes indicated their potential role in pest formations.

Genomic and transcriptomic approaches provide the opportunity to study specific candidate genes and pathways involved in host plant use (Birnbaum & Abbot, 2020). Moreover, these approaches can be used to identify overall genetic mechanisms and patterns associated with herbivory and pests (Després et al., 2007; Groen & Whiteman, 2016). For example, by combining a genome-wide association study (GWAS) and transcriptomics, Nallu et al. (2018) identified genomic regions in the monophagous small white, *Pieris rapae*, associated with herbivory on its host plant. They identified one region consisting of seven genes significantly associated with herbivory. Moreover, they performed a transcriptomic analysis with four monophagous butterfly species, from three families, and identified a core set of 17 orthologous genes involved in feeding on the respective host species. The expression patterns of the genes expressed in response to herbivory in all selected divergent species proved highly species-specific (Nallu et al., 2018).

Similarly, a species-specific expression pattern was observed in the Noctuidae moths in response to host plant feeding in Chapter 3. Additionally, we identified homologous genes with shared expression among closely related species within *Spodoptera* and subfamilies Noctuinae and Plusiinae. The shared expression as observed confirms the importance of applying a phylogenomic framework in comparative studies since shared evolutionary history does influence the transcriptional plastic response that herbivorous species evolved and/or maintained in their interaction with host plants. However, the plastic expression patterns indicated that over evolutionary time the genetic dynamics becomes highly species-specific. The lineage-specific response proves that specific genetic underpinnings identified in different species or model systems cannot be generalised to herbivorous insects as a whole (Nallu et al., 2018).

Lineage-specific expression patterns provides opportunities for targeted control of pest formations. Since the non-targeted effect of pesticides causes, besides pesticide resistance, undesired effects on other organisms and the environment (Schulz et al., 2021). The undesired effects of pesticides on non-target organisms are large and both lethal and sublethal effects are visible in a large variety of organisms and ecosystems (Anju et al., 2010; Davidson & Knapp, 2007; Desneux et al., 2007; Schulz et al., 2021; Stanley et al., 2016; Theiling & Croft, 1988).

Sequence-specific gene silencing via RNA interference (RNAi) holds great promise in targeted effective management of pest species in agricultural environments (Christiaens et al., 2020; Mamta & Rajam, 2017; Mezzetti et al., 2020). However, the outcome of RNAi effects depends on the selected target gene and its role and influence on important processes related to insect survival (Mamta & Rajam, 2017; Terenius et al., 2011). Gene silencing of genes involved in successful herbivory, as identified in Chapter 3, may form ideal candidates for targeted control of pest formations. Besides genes involved in plant feeding optimal candidates are genes with vital roles in development and survival. By RNA sequencing of all major developmental life stages of *S. exigua* I identified genes expressed in larvae, the life stage most detrimental to agriculture. The identification of these potential candidates forms preliminary steps for a targeted, sustainable control of one of the most detrimental insect pest genera worldwide: *Spodoptera*.

## Reflections and perspectives

In this thesis I addressed the main question, ‘What makes a pest a pest, on a genetic basis?’, from multiple angles. I would like to emphasise that in my view a species ‘being’ a pest is non-existent, as introduced in Chapter 1. Species compete for common crop resources with our own species, *Homo sapiens*, and it’s their good right to do so! *Homo sapiens* are part of an ecosystem as much as these “pest” insect species. It is our responsibility to find sustainable, respectful and ethical ways of preventing

and minimising the loss of crop harvests. Especially in times of biodiversity crises and decreasing insect populations (Forister et al., 2019), any form of untargeted attacks and eradication of insect species, populations and individuals with the use of insecticides and pesticides should be avoided. I hope that with my research of the past years, as presented in this thesis, I have contributed towards a genetic understanding of pest formations of polyphagous butterflies and moths and, moreover, formed some preliminary steps towards sustainable ways of preventing pest formations in the future by applying targeted control strategies.

My findings give us the opportunity to place polyphagous pest insect species in a broader evolutionary perspective. Moreover, it can be seen as initial steps in forming new pest control strategies, increase our understanding of pest formations and allow better assessment of insect pest risks. Studying the evolutionary success of herbivorous insects is of biological and practical importance and provides insights into fundamental aspects of insect-plant interactions (Edger et al., 2015). Increased globalisation and the growing demand for food production due to an increase of the human population makes understanding and controlling insect pest species of critical importance. Further, global climate change will likely have implications in the field of herbivorous insects and pest formations (Bale et al., 2002; Boggs, 2016; Karuppaiah & Sujayanad, 2012). It is expected that insect pests will generally become more abundant due to higher reproductive rates, rapid growth, increased migration, and less reliance on overwintering strategies. Invasive pest species are expected to respond more quickly to global climate change than plants and therefore likely to colonise newly available crops and habitats (Yan et al., 2017; Ziska et al., 2011). With the foresight of increasing pest formations, by both invasive and local species, an (genetic) understanding of pest formations is a necessity to predict, manage and sustainably control pest species and outbreaks.

Insect pest risk analyses are an important tool to assess the pest potential of herbivorous insect species. Assessments are conducted for species assigned high risk due to, e.g. climatic changes, feeding on agricultural crops, expanding distribution ranges and/or migration to new environments (e.g. Gruffudd et al., 2019; Strauss, 2010; Sutherst, 1991). The assessments are often based on models (Gruffudd et al., 2019), surveys (Hasan et al., 2021), and (periodical) observed changes in biological and/or economic trends from different localities (MacLeod et al., 2002). The accurate, rapid identification of species through DNA barcodes greatly improved and facilitated appropriate management strategies (Ball & Armstrong, 2006; Hebert & Gregory, 2005). In contrast, assessing the pest potential of herbivorous insect species still largely depends on life-history characteristics such as feeding style and dispersal ability (Nahrung & Swain, 2015; Nylin, 2001). The results of my thesis can be seen as the first, fundamental steps to close this gap towards our understanding and identification of genetic mechanisms and factors contributing and promoting pest potential and formations. Increased knowledge of the success, evolution and genetic basis of pest formations will lead us to:

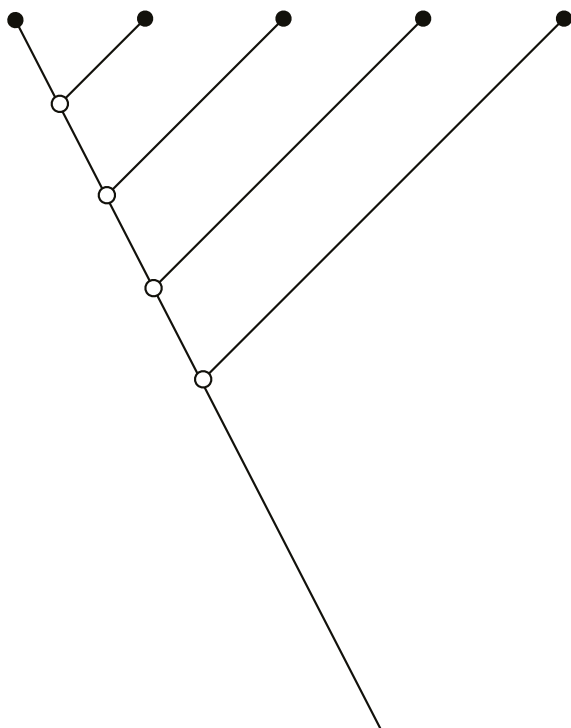
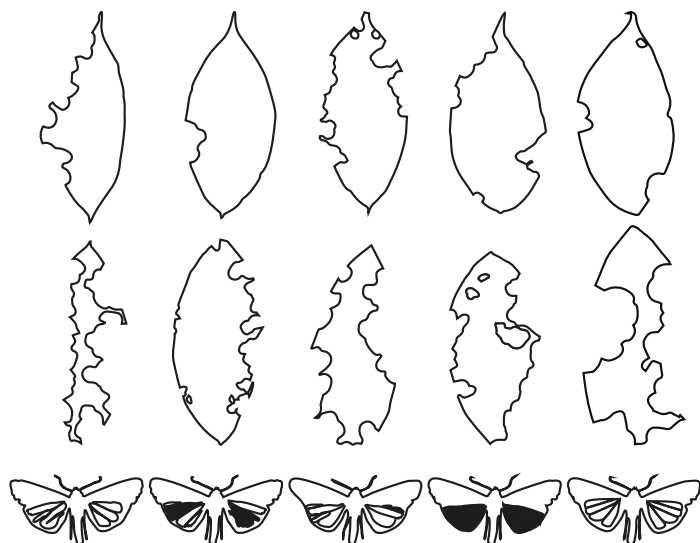


- Refinement of biological knowledge related to (polyphagous) insect pest formations;
- Establishment of a scientific strategy for insect pest risk analyses on an independent genetic basis;
- Improvement of current (polyphagous) insect pest formation risk assessment strategies and predictive risk models;
- Development of appropriate sustainable and targeted measures to prevent, mitigate and control (polyphagous) insect pest formations and their impact.

A complete understanding of the evolution of 'pest-factors' and definition of pest-forming species on a genetic basis will assist in rapid detection of pests and accurate deployment of appropriate management strategies. The presented findings have brought us a step closer to the understanding of the genetic basis and transcriptional plasticity of polyphagy and polyphagous pest formations. The results can be used for further research in order to study pest species within a broader evolutionary perspective across insects. The pest concept and complex interactions between plants and polyphagous insects are challenging. However, the lepidopteran family Noctuidae could serve as an example model group to further study the genetic basis of pest formations by applying the comparative phylogenomic approach across insect and arthropod groups. The differences and commonalities between clades of herbivorous arthropods could further shed light on the evolutionary origin and adaptive evolution of pest-related factors and genes.

A broad evolutionary perspective of the comparative system can be applied to study inter-species transcriptional plasticity related to pest formations. Populations of herbivorous insect species differ in terms of acceptance and use of host plant species and herbivore success. By studying inter-species transcriptional plasticity related to pest formations across populations, our view on the pest concept and genetic basis associated with plant use may get refined on population level.

The next steps in research addressing the genetic mechanisms of pests could be to functionally characterise genes crucial for the success of pest formations. Additional in-depth research may further pinpoint genes of importance for survival of notorious pest-forming species. Future studies should address and confirm the species- or clade-specificity of these genes to select promising targets for application in RNAi mediated pest-outbreak management to avoid harm being done to non-target organisms and the ecosystem. I hope that research on the genetic basis of polyphagous herbivory and pest formations within an evolutionary perspective will continue. The outcome of this thesis may serve in gaining a genetic understanding of insect pest formations and polyphagous herbivory and help in finding sustainable and respectful ways of controlling pest formations.



## References



## A

- Agrawal, A. A., Tuzun, S., & Bent, E. (1999). *Induced plant defenses against pathogens and herbivores: biochemistry, ecology, and agriculture*. APS Press.
- Agrawal, S., Kelkenberg, M., Begum, K., Steinfeld, L., Williams, C. E., Kramer, K. J., Beeman, R. W., Park, Y., Muthukrishnan, S., & Merzendorfer, H. (2014). Two essential peritrophic matrix proteins mediate matrix barrier functions in the insect midgut. *Insect Biochemistry and Molecular Biology*, 49, 24-34.
- Ahmad, S. (1986). Enzymatic adaptations of herbivorous insects and mites to phytochemicals. *Journal of Chemical Ecology*, 12(2), 533-560.
- Ahn, S.-J., Dermauw, W., Wybouw, N., Heckel, D. G., & Van Leeuwen, T. (2014). Bacterial origin of a diverse family of UDP-glycosyltransferase genes in the *Tetranychus urticae* genome. *Insect Biochemistry and Molecular Biology*, 50, 43-57.
- Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., & Chait, B. T. (2007). The molecular architecture of the nuclear pore complex. *Nature*, 450(7170), 695-701.
- Ali, J. G., & Agrawal, A. A. (2012). Specialist versus generalist insect herbivores and plant defense. *Trends in Plant Science*, 17(5), 293-302.
- Allio, R., Nabholz, B., Wanke, S., Chomicki, G., Pérez-Escobar, O. A., Cotton, A. M., Clamens, A.-L., Kergoat, G. J., Sperling, F. A., & Condamine, F. L. (2021). Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. *Nature Communications*, 12(1), 1-15.
- Alyokhin, A., & Chen, Y. H. (2017). Adaptation to toxic hosts as a factor in the evolution of insecticide resistance. *Current Opinion in Insect Science*, 21, 33-38.
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Anju, A., Ravi S, P., & Bechan, S. (2010). Water pollution with special reference to pesticide contamination in India. *Journal of Water Resource and Protection*, 2(5).
- Armstrong, R. N. (1997). Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chemical Research in Toxicology*, 10(1), 2-18.

## B

- Baldwin, I. T. (1988). The alkaloidal responses of wild tobacco to real and simulated herbivory. *Oecologia*, 77(3), 378-381.
- Bale, J. S., Masters, G. J., Hodkinson, I. D., Awmack, C., Bezemer, T. M., Brown, V. K., Butterfield, J., Buse, A., Coulson, J. C., & Farrar, J. (2002). Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biology*, 8(1), 1-16.
- Ball, S. L., & Armstrong, K. F. (2006). DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). *Canadian Journal of Forest Research*, 36(2), 337-350.
- Battisti, A., & Larsson, S. (2015). Climate change and insect pest distribution range. In Björkman, C. & Niemelä, P. (Eds.), *Climate Change and Insect Pests* (pp. 1-15), CABI.
- Bazin, A. L., Mitter, K. T., Davis, D. R., Van Nieukerken, E. J., Cummings, M. P., & Mitter, C. (2017). Phylotranscriptomics resolves ancient divergences in the Lepidoptera. *Systematic Entomology*, 42(2), 305-316.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 289-300.
- Bennett, R. N., & Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytologist*, 127(4), 617-633.

- Beran, F., Pauchet, Y., Kunert, G., Reichelt, M., Wielsch, N., Vogel, H., Reinecke, A., Svatoš, A., Mewis, I., & Schmid, D. (2014). *Phyllotreta striolata* flea beetles use host plant defense compounds to create their own glucosinolate-myrosinase system. *Proceedings of the National Academy of Sciences*, 111(20), 7349-7354.
- Berenbaum, M. R. (2002). Postgenomic chemical ecology: from genetic code to ecological interactions. *Journal of Chemical Ecology*, 28(5), 873-896.
- Berenbaum, M. R., Cohen, M. B., & Schuler, M. A. (1992). Cytochrome P450 monooxygenase genes in oligophagous Lepidoptera. In C. A. Mullin & J. G. Scott (Eds.), *Molecular Mechanisms of Insecticide Resistance*. American Chemical Society.
- Bernays, E., & Graham, M. (1988). On the evolution of host specificity in phytophagous arthropods. *Ecology*, 69(4), 886-892.
- Bi, H. L., Xu, J., Tan, A. J., & Huang, Y. P. (2016). CRISPR/Cas9-mediated targeted gene mutagenesis in *Spodoptera litura*. *Insect Science*, 23(3), 469-477.
- Birk, Y. (2003). *Plant protease inhibitors: significance in nutrition, plant protection, cancer prevention and genetic engineering*. Springer Science & Business Media.
- Birnbaum, S. S. L., & Abbot, P. (2020). Gene expression and diet breadth in plant-feeding insects: summarizing trends. *Trends in Ecology & Evolution*, 35(3), 259-277.
- Bock, K. W. (2016). The UDP-glycosyltransferase (UGT) superfamily expressed in humans, insects and plants: animal-plant arms-race and co-evolution. *Biochemical Pharmacology*, 99, 11-17.
- Boggs, C. L. (2016). The fingerprints of global climate change on insect populations. *Current Opinion in Insect Science*, 17, 69-73.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2), 185-193.
- Braby, M., & Trueman, J. (2006). Evolution of larval host plant associations and adaptive radiation in pierid butterflies. *Journal of Evolutionary Biology*, 19(5), 1677-1690.
- Brattsten, L. (1988). Enzymic adaptations in leaf-feeding insects to host-plant allelochemicals. *Journal of Chemical Ecology*, 14(10), 1919-1939.
- Brattsten, L. B. (1992). Metabolic defenses against plant allelochemicals. In Rosenthal, G. A., & Berenbaum, M. R., *Herbivores: their interaction with secondary plant metabolites* (Vol. 2, pp. 175-241), Academic Press.
- Bravo, G. A., Antonelli, A., Bacon, C. D., Bartoszek, K., Blom, M. P., Huynh, S., Jones, G., Knowles, L. L., Lamichhaney, S., & Marcussen, T. (2019). Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. *PeerJ*, 7, e6399.
- Breeschoten, T., Ros, V. I., Schranz, M. E., & Simon, S. (2019). An influential meal: host plant dependent transcriptional variation in the beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae). *BMC Genomics*, 20(1), 1-15.
- Bretschneider, A., Heckel, D. G., & Vogel, H. (2016). Know your ABCs: characterization and gene expression dynamics of ABC transporters in the polyphagous herbivore *Helicoverpa armigera*. *Insect Biochemistry and Molecular Biology*, 72, 1-9.
- Burand, J. P., & Hunter, W. B. (2013). RNAi: future in insect management. *Journal of Invertebrate Pathology*, 112(1), 68-74.

## C

- Calla, B., Noble, K., Johnson, R. M., Walden, K. K., Schuler, M. A., Robertson, H. M., & Berenbaum, M. R. (2017). Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: birth, death and adaptation. *Molecular Ecology*, 26(21), 6021-6035.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1), 421.
- Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2008). *Biology* (B. Wilbur, Ed. 8th ed.). Pearson Benjamin Cummings.
- Carino, F., Koener, J., Plapp, F., & Feyereisen, R. (1994). Constitutive overexpression of the cytochrome P450 gene CYP6A1 in a house fly strain with metabolic resistance to insecticides. *Insect Biochemistry and Molecular Biology*, 24(4), 411-418.
- Castello, P. R., Drechsel, D. A., & Patel, M. (2007). Mitochondria are a major source of paraquat-induced reactive oxygen species production in the brain. *Journal of Biological Chemistry*, 282(19), 14186-14193.
- Celorio-Mancera, M. d. I. P., Heckel, D. G., & Vogel, H. (2012). Transcriptional analysis of physiological pathways in a generalist herbivore: responses to different host plants and plant structures by the cotton bollworm, *Helicoverpa armigera*. *Entomologia Experimentalis et Applicata*, 144(1), 123-133.
- Celorio-Mancera, M. d. I. P., Wheat, C. W., Vogel, H., Söderlind, L., Janz, N., & Nylin, S. (2013). Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-seq. *Molecular Ecology*, 22(19), 4884-4895.
- Celorio-Mancera, M. d. I. P., Wheat, C. W., Huss, M., Vezzi, F., Neethiraj, R., Reimegård, J., Nylin, S., & Janz, N. (2016). Evolutionary history of host use, rather than plant phylogeny, determines gene expression in a generalist butterfly. *BMC Evolutionary Biology*, 16(1), 59.
- Chandra, G. S., Asokan, R., Manamohan, M., & Sita, T. (2016). Cytochrome P450 isoforms transcriptional, larval growth and development responses to host allelochemicals in the generalist herbivore, *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae). *Current Science*, 111(5), 901-906.
- Chen, W., Hasegawa, D. K., Kaur, N., Kliot, A., Pinheiro, P. V., Luan, J., Stensmyr, M. C., Zheng, Y., Liu, W., & Sun, H. (2016). The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology*, 14(1), 110.
- Chen, W., Yang, X., Tetreau, G., Song, X., Coutu, C., Hegedus, D., Blissard, G., Fei, Z., & Wang, P. (2019). A high-quality chromosome-level genome assembly of a generalist herbivore, *Trichoplusia ni*. *Molecular Ecology Resources*, 19(2), 485-496.
- Cheng, T., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., Feng, L., Li, W., Chen, Z., & Guo, H. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nature Ecology & Evolution*, 1(11), 1747.
- Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, 65(6), 997-1008.
- Chew, F. S. (1995). From weeds to crops: changing habitats of pierid butterflies (Lepidoptera: Pieridae). *Journal of the Lepidopterists Society*, 49(4), 285-303.
- Cho, S., Mitchell, A., Mitter, C., Regier, J., Matthews, M., & Robertson, R. (2008). Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. *Systematic Entomology*, 33(4), 581-594.
- Christiaens, O., Niu, J., & Nji Tizi Tining, C. (2020). RNAi in insects: a revolution in fundamental research and pest control applications. *Insects*, 11(7), 415.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., & Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 1-19.
- Crispo, E. (2007). The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution: International Journal of Organic Evolution*, 61(11), 2469-2479.

## D

- Darwin, C. (1859). *On the origin of the species by means of natural selection, or, preservation of favoured races in the struggle for life*. John Murray.
- Davidson, C., & Knapp, R. A. (2007). Multiple stressors and amphibian declines: dual impacts of pesticides and fish on yellow-legged frogs. *Ecological Applications*, 17(2), 587-597.
- Dermauw, W., Osborne, E. J., Clark, R. M., Grbić, M., Tirry, L., & Van Leeuwen, T. (2013). A burst of ABC genes in the genome of the polyphagous spider mite *Tetranychus urticae*. *BMC Genomics*, 14(1), 317.
- Dermauw, W., Pym, A., Bass, C., Van Leeuwen, T., & Feyereisen, R. (2018). Does host plant adaptation lead to pesticide resistance in generalist herbivores? *Current Opinion in Insect Science*, 26, 25-33.
- Dermauw, W., & Van Leeuwen, T. (2014). The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. *Insect Biochemistry and Molecular Biology*, 45, 89-110.
- Dermauw, W., Van Leeuwen, T., & Feyereisen, R. (2020). Diversity and evolution of the P450 family in arthropods. *Insect Biochemistry and Molecular Biology*, 103490.
- Dermauw, W., Wybouw, N., Rombauts, S., Menten, B., Vontas, J., Grbić, M., Clark, R. M., Feyereisen, R., & Van Leeuwen, T. (2013). A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proceedings of the National Academy of Sciences*, 110(2), 113-122.
- Desneux, N., Decourtye, A., & Delpuech, J.-M. (2007). The sublethal effects of pesticides on beneficial arthropods. *Annual Review of Entomology*, 52, 81-106.
- Després, L., David, J.-P., & Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution*, 22(6), 298-307.
- Dukes, J. S., Pontius, J., Orwig, D., Garnas, J. R., Rodgers, V. L., Brazee, N., Cooke, B., Theoharides, K. A., Stange, E. E., & Harrington, R. (2009). Responses of insect pests, pathogens, and invasive plant species to climate change in the forests of northeastern North America: what can we predict? *Canadian Journal of Forest Research*, 39(2), 231-248.
- Dunn, C. W., Luo, X., & Wu, Z. (2013). Phylogenetic analysis of gene expression. *Integrative and Comparative Biology*, 53(5), 847-856.
- Dunn, C. W., Zapata, F., Munro, C., Siebert, S., & Hejnal, A. (2018). Pairwise comparisons across species are problematic when analyzing functional genomic data. *Proceedings of the National Academy of Sciences*, 115(3), 409-417.

## E

- Early, R., González-Moreno, P., Murphy, S. T., & Day, R. (2018). Forecasting the global extent of invasion of the cereal pest *Spodoptera frugiperda*, the fall armyworm. *NeoBiota*, 40, 25-50.
- Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glöckner, G., Platts, A. E., Heckel, D. G., Der, J. P., Wafula, E. K., & Tang, M. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences*, 112(27), 8362-8366.
- Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: a study in coevolution. *Evolution*, 18(4), 586-608.
- Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1), 157.
- Endo, H., Tanaka, S., Imamura, K., Adegawa, S., Kikuta, S., & Sato, R. (2017). Cry toxin specificities of insect ABCC transporters closely related to lepidopteran ABCC2 transporters. *Peptides*, 98, 86-92.
- EPPO. (2019). *EPPO global pest database*. <https://gd.eppo.int>

- Erwin, T. L. (1997). Biodiversity at its utmost: tropical forest beetles. In Reaka-Kudla M. L., Wilson, D. E. & Wilson, E. O. (Eds.), *Biodiversity II: understanding and protecting our biological resources* (pp. 27-40). Joseph Henry Press.
- Eswaramoorthy, S., Bonanno, J. B., Burley, S. K., & Swaminathan, S. (2006). Mechanism of action of a flavin-containing monooxygenase. *Proceedings of the National Academy of Sciences*, 103(26), 9832-9837.

## F

- Fahey, J. W., Zalcman, A. T., & Talalay, P. (2001). The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry*, 56(1), 5-51.
- Falcon, S., & Gentleman, R. (2007). Using GStats to test gene lists for GO term association. *Bioinformatics*, 23(2), 257-258.
- Fernández, R., Gabaldon, T., & Dessimoz, C. (2020). Orthology: definitions, prediction, and impact on species phylogeny inference. In Scornavacca, C., Delsuc, F. d. r. & Galtier, N. (Eds.), *Phylogenetics in the genomic era* (pp. 1-14). No commercial publisher, authors open access book.
- Feyereisen, R. (1999). Insect P450 enzymes. *Annual Review of Entomology*, 44(1), 507-533.
- Feyereisen, R. (2005). Insect cytochrome P450. *Comprehensive Molecular Insect Science*, 4, 1-77.
- Feyereisen, R. (2006). Evolution of insect P450. *Biochemical Society Transactions*, 34(6), 1252-1255.
- Feyereisen, R. (2011). Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(1), 19-28.
- Feyereisen, R., Koener, J. F., Farnsworth, D. E., & Nebert, D. W. (1989). Isolation and sequence of cDNA encoding a cytochrome P-450 from an insecticide-resistant strain of the house fly, *Musca domestica*. *Proceedings of the National Academy of Sciences*, 86(5), 1465-1469.
- Fischer, H. M., Wheat, C. W., Heckel, D. G., & Vogel, H. (2008). Evolutionary origins of a novel host plant detoxification gene in butterflies. *Molecular Biology and Evolution*, 25(5), 809-820.
- Fisher, R. A. (1948). Answer to Question 14 on Combining independent tests of significance. *The American Statistician* 2, 30.
- Fordyce, J. A. (2010). Host shifts and evolutionary radiations of butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701), 3735-3743.
- Forister, M. L., Pelton, E. M., & Black, S. H. (2019). Declines in insect abundance and diversity: we know enough to act now. *Conservation Science and Practice*, 1(8), e80.
- Foss, L., & Rieske, L. (2003). Species-specific differences in oak foliage affect preference and performance of gypsy moth caterpillars. *Entomologia Experimentalis et Applicata*, 108(2), 87-93.
- Francis, F., Vanhaelen, N., & Haubruge, E. (2005). Glutathione S-transferases in the adaptation to plant secondary metabolites in the *Myzus persicae* aphid. *Archives of Insect Biochemistry and Physiology*, 58(3), 166-174.
- Frey, M., Schullehner, K., Dick, R., Fiesselmann, A., & Gierl, A. (2009). Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry*, 70(15-16), 1645-1651.
- Fritz, M. L., DeYonke, A. M., Papanicolaou, A., Micinski, S., Westbrook, J., & Gould, F. (2018). Contemporary evolution of a Lepidopteran species, *Heliothis virescens*, in response to modern agricultural practices. *Molecular Ecology*, 27(1), 167-181.
- Fu, X., Feng, H., Liu, Z., & Wu, K. (2017). Trans-regional migration of the beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae), in North-East Asia. *PLOS ONE*, 12(8), e0183582.
- Funk, D. J., Filchak, K. E., & Feder, J. L. (2002). Herbivorous insects: model systems for the comparative study of speciation ecology. In Etges, W. J. & Noor, M. A. (Eds.), *Genetics of mate choice: from sexual selection to sexual isolation* (pp. 251-267). Springer.



## G

- Gloss, A. D., Abbot, P., & Whiteman, N. K. (2019). How interactions with plant chemicals shape insect genomes. *Current Opinion in Insect Science*, 36, 149-156.
- Goergen, G., Kumar, P. L., Sankung, S. B., Togola, A., & Tamò, M. (2016). First report of outbreaks of the fall armyworm *Spodoptera frugiperda* (JE Smith) (Lepidoptera, Noctuidae), a new alien invasive pest in West and Central Africa. *PLOS ONE*, 11(10), e0165632.
- 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 Research*, 25(11), 1750-1756.
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333.
- Gordon, H. (1961). Nutritional factors in insect resistance to chemicals. *Annual Review of Entomology*, 6(1), 27-54.
- Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J.-M., Duvic, B., Hilliou, F., Durand, N., Montagné, N., & Darboux, I. (2017). Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Scientific Reports*, 7(1), 11816.
- Govind, G., Mittapalli, O., Griebel, T., Allmann, S., Böcker, S., & Baldwin, I. T. (2010). Unbiased transcriptional comparisons of generalist and specialist herbivores feeding on progressively defenseless *Nicotiana attenuata* plants. *PLOS ONE*, 5(1), e8735.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., & Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644-652.
- Grbić, M., Van Leeuwen, T., Clark, R. M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E. J., Dermauw, W., Ngoc, P. C. T., & Ortego, F. (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature*, 479(7374), 487.
- Greenberg, S., Sappington, T., Sétamou, M., & Liu, T.-X. (2002). Beet armyworm (Lepidoptera: Noctuidae) host plant preferences for oviposition. *Environmental Entomology*, 31(1), 142-148.
- Greenspan, R. J. (2001). The flexible genome. *Nature Reviews Genetics*, 2(5), 383-387.
- Grimaldi, D., & Engel, M. S. (2005). *Evolution of the insects*. Cambridge University Press.
- Groen, S. C., & Whiteman, N. K. (2016). Using *Drosophila* to study the evolution of herbivory and diet specialization. *Current Opinion in Insect Science*, 14, 66-72.
- Gruffudd, H. R., Schröder, T., Jenkins, T. A. R., & Evans, H. F. (2019). Modelling pine wilt disease (PWD) for current and future climate scenarios as part of a pest risk analysis for pine wood nematode *Bursaphelenchus xylophilus* (Steiner and Buhner) Nickle in Germany. *Journal of Plant Diseases and Protection*, 126(2), 129-144.
- Gu, J., Huang, L. X., Gong, Y. J., Zheng, S. C., Liu, L., Huang, L. H., & Feng, Q. L. (2013). De novo characterization of transcriptome and gene expression dynamics in epidermis during the larval-pupal metamorphosis of common cutworm. *Insect Biochemistry and Molecular Biology*, 43(9), 794-808.
- Gu, X., Fu, Y. X., & Li, W. H. (1995). Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Molecular Biology and Evolution*, 12(4), 546-557.
- Gui, F., Lan, T., Zhao, Y., Guo, W., Dong, Y., Fang, D., Liu, H., Li, H., Wang, H., & Hao, R. (2020). Genomic and transcriptomic analysis unveils population evolution and development of pesticide resistance in fall armyworm *Spodoptera frugiperda*. *Protein & Cell*, 1-19.

## H

- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., & Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494-1512.
- Hahn, M. W., De Bie, T., Stajich, J. E., Nguyen, C., & Cristianini, N. (2005). Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Research*, 15(8), 1153-1160.
- Halkier, B. A., & Gershenzon, J. (2006). Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology*, 57, 303-333.
- Hasan, S., Haque, M. E., Afrad, M. S. I., Alam, M. Z., Hoque, M. Z., & Islam, M. R. (2021). Pest risk analysis and management practices for increasing profitability of lemon production. *Journal of Agriculture and Ecology Research International*, 26-35.
- He, W.-y., Rao, Z.-c., Zhou, D.-h., Zheng, S.-c., Xu, W.-h., & Feng, Q.-l. (2012). Analysis of expressed sequence tags and characterization of a novel gene, *Slmg7*, in the midgut of the common cutworm, *Spodoptera litura*. *PLOS ONE*, 7(3), e33621.
- He, X., & Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*, 169(2), 1157-1164.
- Hebert, P. D., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. *Systematic Biology*, 54(5), 852-859.
- Heckel, D. G. (2018). Insect detoxification and sequestration strategies. In Voelckel, C., & Jander, G. (Eds.), *Insect-Plant Interactions* (Vol. 47, pp. 77-114). Wiley Blackwell.
- Hegedus, D., Erlandson, M., Gillott, C., & Toprak, U. (2009). New insights into peritrophic matrix synthesis, architecture, and function. *Annual Review of Entomology*, 54, 285-302.
- Heidel-Fischer, H. M., Kirsch, R., Reichelt, M., Ahn, S.-J., Wielsch, N., Baxter, S. W., Heckel, D. G., Vogel, H., & Kroymann, J. (2019). An insect counteradaptation against host plant defenses evolved through concerted neofunctionalization. *Molecular Biology and Evolution*, 36(5), 930-941.
- Heidel-Fischer, H. M., & Vogel, H. (2015). Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*, 8, 8-14.
- Herrero, S., Ansems, M., Van Oers, M. M., Vlak, J. M., Bakker, P. L., & de Maagd, R. A. (2007). REPAT, a new family of proteins induced by bacterial toxins and baculovirus infection in *Spodoptera exigua*. *Insect Biochemistry and Molecular Biology*, 37(11), 1109-1118.
- Hill, D. S. (2012). *The economic importance of insects*. Springer Science & Business Media.
- Hill, J., Rastas, P., Hornett, E. A., Neethiraj, R., Clark, N., Morehouse, N., de la Paz Celorio-Mancera, M., Cols, J. C., Dirksen, H., & Meslin, C. (2019). Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Science Advances*, 5(6), eaau3648.
- Hoang, K., Matzkin, L. M., & Bono, J. M. (2015). Transcriptional variation associated with cactus host plant adaptation in *Drosophila mettleri* populations. *Molecular Ecology*, 24(20), 5186-5199.
- Howard, J. J., Raubenheimer, D., & Bernays, E. A. (1994). Population and individual polyphagy in the grasshopper *Taeniopoda eques* during natural foraging. *Entomologia Experimentalis et Applicata*, 71(2), 167-176.
- Howe, G. A., & Jander, G. (2008). Plant immunity to insect herbivores. *Annual Review of Plant Biology*, 59, 41-66.
- Hu, B., Huang, H., Hu, S., Ren, M., Wei, Q., Tian, X., Esmail Abdalla Elzaki, M., Bass, C., Su, J., & Reddy Palli, S. (2021). Changes in both trans- and cis-regulatory elements mediate insecticide resistance in a lepidopteran pest, *Spodoptera exigua*. *PLoS Genetics*, 17(3), e1009403.
- Hu, B., Zhang, S. H., Ren, M. M., Tian, X. R., Wei, Q., Mburu, D. K., & Su, J. Y. (2019). The expression of *Spodoptera exigua* P450 and UGT genes: tissue specificity and response to insecticides. *Insect Science*, 26(2), 199-216.

- Huang, J. M., Zhao, Y. X., Sun, H., Ni, H., Liu, C., Wang, X., Gao, C. F., & Wu, S. F. (2021). Monitoring and mechanisms of insecticide resistance in *Spodoptera exigua* (Lepidoptera: Noctuidae), with special reference to diamides. *Pesticide Biochemistry and Physiology*, 174, 104831.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and time-series model selection in small samples. *Biometrika*, 76(2), 297-307.

## I

- Innan, H., & Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics*, 11(2), 97-108.
- International Silkworm Genome Consortium (2008). The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 38(12), 1036-1045.

## J

- Jansen, H. J., Liem, M., Jong-Raadsen, S. A., Dufour, S., Weltzien, F. A., Swinkels, W., Koelewijn, A., Palstra, A. P., Pelster, B., Spaik, H. P., Thillart, G., Dirks, R. P., & Henkel, C. V. (2017). Rapid de novo assembly of the European eel genome from nanopore sequencing reads. *Scientific Reports*, 7(1), 7213.
- Janz, N., & Nylin, S. (1998). Butterflies and plants: a phylogenetic study. *Evolution*, 52(2), 486-502.
- Janz, N., & Nylin, S. (2008). The oscillation hypothesis of host-plant range and speciation. In Tilmon, K. J. (Ed.), *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects* (pp. 203- 215). Univ. of California Press.
- Janz, N., Nylin, S., & Wahlberg, N. (2006). Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology*, 6(1), 4.
- Jerny, T. (1976). Insect-host-plant relationship – co-evolution or sequential evolution? In Jerny, T., *The host-plant in relation to insect behaviour and reproduction* (pp. 109-113). Springer.
- Jerny, T. (1984). Evolution of insect/host plant relationships. *The American Naturalist*, 124(5), 609-630.
- Jeschke, V., Gershenzon, J., & Vassão, D. (2016). Insect detoxification of glucosinolates and their hydrolysis products. In *Advances in Botanical Research* (Vol. 80, pp. 199-245). Elsevier.
- Jeschke, V., Kearney, E. E., Schramm, K., Kunert, G., Shekhov, A., Gershenzon, J., & Vassão, D. G. (2017). How glucosinolates affect generalist lepidopteran larvae: growth, development and glucosinolate metabolism. *Frontiers in Plant Science*, 8, 1995.
- Jing, D. P., Guo, J. F., Jiang, Y. Y., Zhao, J. Z., Sethi, A., He, K. L., & Wang, Z. Y. (2020). Initial detections and spread of invasive *Spodoptera frugiperda* in China and comparisons with other noctuid larvae in cornfields using molecular techniques. *Insect Science*, 27(4), 780-790.
- Jo, Y. H., Park, S., Park, K. B., Noh, M. Y., Cho, J. H., Ko, H. J., Kim, C. E., Patnaik, B. B., Kim, J., & Won, R. (2018). In silico identification, characterization and expression analysis of attacin gene family in response to bacterial and fungal pathogens in *Tenebrio molitor*. *Entomological Research*, 48(1), 45-54.
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, 8(3), 275-282.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., & Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236-1240.
- Jouraku, A., Yamamoto, K., Kuwazaki, S., Urio, M., Suetsugu, Y., Narukawa, J., Miyamoto, K., Kurita, K., Kanamori, H., & Katayose, Y. (2013). KONAGabase: a genomic and transcriptomic database for the diamondback moth, *Plutella xylostella*. *BMC Genomics*, 14(1), 464.
- Joußen, N., Agnolet, S., Lorenz, S., Schöne, S. E., Ellinger, R., Schneider, B., & Heckel, D. G. (2012). Resistance of Australian *Helicoverpa armigera* to fenvalerate is due to the chimeric P450 enzyme CYP337B3. *Proceedings of the National Academy of Sciences*, 109(38), 15206-15211.

## K

- Kaczynski, J., Cook, T., & Urrutia, R. (2003). Sp1- and Kruppel-like transcription factors. *Genome Biology*, 4(2), 206.
- Kakumani, P. K., Malhotra, P., Mukherjee, S. K., & Bhatnagar, R. K. (2014). A draft genome assembly of the army worm, *Spodoptera frugiperda*. *Genomics*, 104(2), 134-143.
- Kalleshwaraswamy, C., Asokan, R., Swamy, H., Maruthi, M., Pavithra, H., Hegde, K., Navi, S., Prabhu, S., & Goergen, G. (2018). First report of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), an alien invasive pest on maize in India. *Pest Management in Horticultural Ecosystems*, 24(1), 23-29.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587-589.
- Kant, M., Jonckheere, W., Knecht, B., Lemos, F., Liu, J., Schimmel, B., Villarroel, C., Ataíde, L., Dermauw, W., & Glas, J. (2015). Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany*, 115(7), 1015-1051.
- Karban, R. (2020). The ecology and evolution of induced responses to herbivory and how plants perceive risk. *Ecological Entomology*, 45(1), 1-9.
- Karban, R., & Myers, J. H. (1989). Induced plant responses to herbivory. *Annual Review of Ecology and Systematics*, 20(1), 331-348.
- Karuppaiah, V., & Sujayanad, G. (2012). Impact of climate change on population dynamics of insect pests. *World Journal of Agricultural Sciences*, 8(3), 240-246.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772-780.
- Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F., Donath, A., Gimnich, F., Frandsen, P. B., Zwick, A., & dos Reis, M. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences*, 116(45), 22657-22663.
- Kawecki, T. J. (1998). Red queen meets Santa Rosalia: arms races and the evolution of host specialization in organisms with parasitic lifestyles. *The American Naturalist*, 152(4), 635-651.
- Keegan, K. L., Rota, J., Zahiri, R., Zilli, A., Wahlberg, N., Schmidt, B. C., Lafontaine, J. D., Goldstein, P. Z., & Wagner, D. L. (2021). Toward a stable global Noctuidae (Lepidoptera) taxonomy. *Insect Systematics and Diversity*, 5(3).
- Kelkenberg, M., Odman-Naresh, J., Muthukrishnan, S., & Merzendorfer, H. (2015). Chitin is a necessary component to maintain the barrier function of the peritrophic matrix in the insect midgut. *Insect Biochemistry and Molecular Biology*, 56, 21-28.
- Kennedy, M. W., Chalamalasetty, R. B., Thomas, S., Garriock, R. J., Jailwala, P., & Yamaguchi, T. P. (2016). Sp5 and Sp8 recruit  $\beta$ -catenin and Tcf1-Left to select enhancers to activate Wnt target gene transcription. *Proceedings of the National Academy of Sciences*, 113(13), 3545-3550.
- Kergoat, G. J., Goldstein, P. Z., Le Ru, B., Meagher, R. L., Zilli, A., Mitchell, A., Clamens, A.-L., Gimenez, S., Barbut, J., Nègre, N., d'Alençon, E., & Nam, K. (2021). A novel reference dated phylogeny for the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae: Noctuinae): new insights into the evolution of a pest-rich genus. *Molecular Phylogenetics and Evolution*, 161, 107161.
- Kim, E. J., Kim, S. J., Park, C. J., & Nam, Y. K. (2019). Characterization of testis-specific serine/threonine kinase 1-like (TSSK1-like) gene and expression patterns in diploid and triploid Pacific abalone (*Haliotis discus hannai*; Gastropoda; Mollusca) males. *PLOS ONE*, 14(12), e0226022.
- Kim, K. C. (1993). Insect pests and evolution. In Kim, K. C., & McPheron, B. (Eds.), *Evolution of insect pests: patterns of variation* (pp. 3-26). John Wiley & Sons.
- Kirk, H., Dorn, S., & Mazzi, D. (2013). Molecular genetics and genomics generate new insights into invertebrate pest invasions. *Evolutionary Applications*, 6(5), 842-856.
- Kliebenstein, D. J., Kroymann, J., & Mitchell-Olds, T. (2005). The glucosinolate-myrosinase system in an ecological and evolutionary context. *Current Opinion in Plant Biology*, 8(3), 264-271.

- Koenig, C., Bretschneider, A., Heckel, D. G., Grosse-Wilde, E., Hansson, B. S., & Vogel, H. (2015). The plastic response of *Manduca sexta* to host and non-host plants. *Insect Biochemistry and Molecular Biology*, 63, 72-85.
- Kosiol, C., & Goldman, N. (2005). Different versions of the Dayhoff rate matrix. *Molecular Biology and Evolution*, 22(2), 193-199.
- Koutsos, A. C., Blass, C., Meister, S., Schmidt, S., MacCallum, R. M., Soares, M. B., Collins, F. H., Benes, V., Zdobnov, E., Kafatos, F. C., & Christophides, G. K. (2007). Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 104(27), 11304.
- Krieger, R. I., Feeny, P. P., & Wilkinson, C. F. (1971). Detoxication enzymes in the guts of caterpillars: an evolutionary answer to plant defenses? *Science*, 172(3983), 579-581.
- Kristiansson, E., Österlund, T., Gunnarsson, L., Arne, G., Larsson, D. J., & Nerman, O. (2013). A novel method for cross-species gene expression analysis. *BMC Bioinformatics*, 14(1), 70.
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 47(1), 807-811.
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3), 567-580.
- Kück, P., & Longo, G. C. (2014). FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, 11(1), 81.
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Wagele, J. W., & Misof, B. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, 7, 10.
- Kumar, M., Vivekanand, V., & Pareek, N. (2018). Structure, regulation, and potential applications of insect chitin-metabolizing enzymes. In Kumar, D. & Gong, C. (Eds.), *Trends in Insect Molecular Biology and Biotechnology* (pp. 295-316). Springer International Publishing.

## L

- Labandeira, C. C., & Sepkoski, J. J. (1993). Insect diversity in the fossil record. *Science*, 261(5119), 310-315.
- Labbé, R., Caveney, S., & Donly, C. (2011). Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. *Insect Molecular Biology*, 20(2), 243-256.
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100-3108.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357.
- Le Ru, B., Barbut, J., Capdevielle-Dulac, C., Gofitshu, M., & Kergoat, G. J. (2018). Re-establishment of *Spodoptera teferii* Laporte in Rougeot (Lepidoptera: Noctuidae: Noctuidae), with an updated molecular phylogeny for the genus *Spodoptera* Guenée. *Annales de la Société entomologique de France (N.S.)*, 54(6), 497-510.
- Le, S. Q., Dang, C. C., & Gascuel, O. (2012). Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Molecular Biology and Evolution*, 29(10), 2921-2936.
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307-1320.
- Lee, J., & Lee, D. G. (2015). Antimicrobial peptides (AMPs) with dual mechanisms: membrane disruption and apoptosis. *Journal of Microbiology and Biotechnology*, 25(6), 759-764.
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), 323.

- Li, H., Guillemaud, T., French, B., Kuhlmann, U., & Toepfer, S. (2014). Phenotypic trait changes in laboratory-reared colonies of the maize herbivore, *Diabrotica virgifera virgifera*. *Bulletin of Entomological Research*, 104(1), 97-115.
- Li, H., Jiang, W., Zhang, Z., Xing, Y., & Li, F. (2013). Transcriptome analysis and screening for potential target genes for RNAi-mediated pest control of the beet armyworm, *Spodoptera exigua*. *PLOS ONE*, 8(6), e65931.
- Li, X., Baudry, J., Berenbaum, M. R., & Schuler, M. A. (2004). Structural and functional divergence of insect CYP6B proteins: from specialist to generalist cytochrome P450. *Proceedings of the National Academy of Sciences*, 101(9), 2939-2944.
- Li, X., Schuler, M. A., & Berenbaum, M. R. (2007). Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, 52, 231-253.
- Li, Z., Tiley, G. P., Galuska, S. R., Reardon, C. R., Kidder, T. I., Rundell, R. J., & Barker, M. S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proceedings of the National Academy of Sciences*, 115(18), 4713-4718.
- Lindigkeit, R., Biller, A., Buch, M., Schiebel, H. M., Boppré, M., & Hartmann, T. (1997). The two faces of pyrrolizidine alkaloids: the role of the tertiary amine and its N-oxide in chemical defense of insects with acquired plant alkaloids. *European Journal of Biochemistry*, 245(3), 626-636.
- Liu, H., Lan, T., Fang, D., Gui, F., Wang, H., Guo, W., Cheng, X., Chang, Y., He, S., & Lyu, L. (2019). Chromosome level draft genomes of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), an alien invasive pest in China. *bioRxiv*, 671560.
- Liu, P. C., Hao, D. J., Hu, H. Y., & Wei, J. R. (2020). Sexual dimorphism and sex-biased gene expression in an egg parasitoid species, *Anastatus disparis*. *BMC Genomics*, 21(1), 492.
- Liu, S., Zhou, S., Tian, L., Guo, E., Luan, Y., Zhang, J., & Li, S. (2011). Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*. *BMC Genomics*, 12(1), 491.
- Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, 12(8), 733-735.
- Lopez-Vaamonde, C., Wikström, N., Labandeira, C., Godfray, H. C. J., Goodman, S. J., & Cook, J. M. (2006). Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, 19(4), 1314-1326.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
- Lumley, L. M., & Sperling, F. A. H. (2010). Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (*Choristoneura fumiferana*) cryptic species complex (Lepidoptera: Tortricidae). *Systematic Entomology*, 35(3), 416-428.
- Lupia, R., Lidgard, S., & Crane, P. R. (1999). Comparing palynological abundance and diversity: implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology*, 305-340.

## M

- Machado, V., Serrano, J., & Galian, J. (2016). Identification and characterization of pathogen-response genes (repat) in *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Folia Biologica (Kraków)*, 64(1), 23-29.
- MacLeod, A., Evans, H. F., & Baker, R. H. A. (2002). An analysis of pest risk from an Asian longhorn beetle (*Anoplophora glabripennis*) to hardwood trees in the European community. *Crop Protection*, 21(8), 635-645.
- Mamta, B., & Rajam, M. (2017). RNAi technology: a new platform for crop pest control. *Physiology and Molecular Biology of Plants*, 23(3), 487-501.
- Manchester, S. R., Grimsson, F., & Zetter, R. (2015). Assessing the fossil record of asterids in the context of our current phylogenetic framework. *Annals of the Missouri Botanical Garden*. *Missouri Botanical Garden*, 100(4), 329.

- Marquis, R. J. (2004). Herbivores rule. *Science*, 305(5684), 619-621.
- Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., & Collin, O. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biology*, 18(1), 27.
- Mayer, C., Dietz, L., Call, E., Kukowka, S., Martin, S., & Espeland, M. (2021). Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens. *Systematic Entomology*, 46, 649-671.
- McCulloch, K. J., & Koenig, K. M. (2020). Kruppel-like factor/specificity protein evolution in the Spiralia and the implications for cephalopod visual system novelties. *Proceedings of the Royal Society B: Biological Sciences*, 287(1937), 20202055.
- Mehrkhou, F., Talebi, A. A., Moharramipour, S., & Naveh, V. H. (2012). Demographic parameters of *Spodoptera exigua* (Lepidoptera: Noctuidae) on different soybean cultivars. *Environmental Entomology*, 41(2), 326-332.
- Menken, S. B., Boomsma, J. J., & Van Nieukerken, E. J. (2010). Large-scale evolutionary patterns of host plant associations in the Lepidoptera. *Evolution: International Journal of Organic Evolution*, 64(4), 1098-1119.
- Merkx-Jacques, M., Despland, E., & Bede, J. C. (2008). Nutrient utilization by caterpillars of the generalist beet armyworm, *Spodoptera exigua*. *Physiological Entomology*, 33(1), 51-61.
- Mezzetti, B., Smaghe, G., Arpaia, S., Christiaens, O., Dietz-Pfeilstetter, A., Jones, H., Kostov, K., Sabbadini, S., Opsahl-Sorteberg, H.-G., & Ventura, V. (2020). RNAi: what is its position in agriculture? *Journal of Pest Science*, 93, 1125-1130.
- Misof, B., & Misof, K. (2009). A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology*, 58(1), 21-34.
- Mitchell, A., Mitter, C., & Regier, J. C. (2006). Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): evidence from two protein-coding nuclear genes. *Systematic Entomology*, 31(1), 21-46.
- Mitter, C., Davis, D. R., & Cummings, M. P. (2017). Phylogeny and evolution of Lepidoptera. *Annual Review of Entomology*, 62, 265-283.
- Montella, I. R., Schama, R., & Valle, D. (2012). The classification of esterases: an important gene family involved in insecticide resistance - a review. *Memorias do Instituto Oswaldo Cruz*, 107(4), 437-449.
- Montezano, D. G., Specht, A., Sosa-Gómez, D. R., Roque-Specht, V. F., Sousa-Silva, J. C., Paula-Moraes, S. d., Peterson, J. A., & Hunt, T. (2018). Host plants of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in the Americas. *African Entomology*, 26(2), 286-300.
- Muhlia-Almazán, A., Sánchez-Paz, A., & García-Carreño, F. L. (2008). Invertebrate trypsins: a review. *Journal of Comparative Physiology B*, 178(6), 655-672.
- Müller, C., Vogel, H., & Heckel, D. G. (2017). Transcriptional responses to short-term and long-term host plant experience and parasite load in an oligophagous beetle. *Molecular Ecology*, 26(22), 6370-6383.
- Müller, T., & Vingron, M. (2000). Modeling amino acid replacement. *Journal of Computational Biology*, 7(6), 761-776.

## N

- Nahrung, H. F., & Swain, A. J. (2015). Strangers in a strange land: do life history traits differ for alien and native colonisers of novel environments? *Biological Invasions*, 17(2), 699-709.
- Nallu, S., Hill, J. A., Don, K., Sahagun, C., Zhang, W., Meslin, C., Snell-Rood, E., Clark, N. L., Morehouse, N. I., & Bergelson, J. (2018). The molecular genetic basis of herbivory between butterflies and their host plants. *Nature Ecology & Evolution*, 2, 1418-1427.



- Naumann, C., Hartmann, T., & Ober, D. (2002). Evolutionary recruitment of a flavin-dependent monooxygenase for the detoxification of host plant-acquired pyrrolizidine alkaloids in the alkaloid-defended arctiid moth *Tyria jacobaeae*. *Proceedings of the National Academy of Sciences*, 99(9), 6085-6090.
- Navarro-Cerrillo, G., Hernández-Martínez, P., Vogel, H., Ferré, J., & Herrero, S. (2013). A new gene superfamily of pathogen-response (repat) genes in Lepidoptera: classification and expression analysis. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 164(1), 10-17.
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274.
- Nosil, P. (2002). Transition rates between specialization and generalization in phytophagous insects. *Evolution*, 56(8), 1701-1706.
- Nunes, M. E., Müller, T. E., Murussi, C., do Amaral, A. M., Gomes, J. L., Marins, A. T., Leitemperger, J., Rodrigues, C. C., Fiuza, T. L., & Costa, M. D. (2018). Oxidative effects of the acute exposure to a pesticide mixture of cypermethrin and chlorpyrifos on carp and zebrafish– a comparative study. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 206, 48-53.
- Nylin, S. (2001). Life history perspectives on pest insects: what's the use? *Austral Ecology*, 26(5), 507-517.
- Nylin, S., Agosta, S., Bensch, S., Boeger, W. A., Braga, M. P., Brooks, D. R., Forister, M. L., Hambäck, P. A., Hoberg, E. P., & Nyman, T. (2018). Embracing colonizations: a new paradigm for species association dynamics. *Trends in Ecology & Evolution*, 33(1), 4-14.
- Nylin, S., Slove, J., & Janz, N. (2014). Host plant utilization, host range oscillations and diversification in nymphalid butterflies: a phylogenetic investigation. *Evolution*, 68(1), 105-124.

## O

- Oakeshott, J., Claudianos, C., Campbell, P., Newcomb, R., Russell, R., & Gilbert, L. (2005). Biochemical genetics and genomics of insect esterases. In *Insect pharmacology: channels, receptors, toxins and enzymes* (Vol. 5). Academic Press.
- Ødegaard, F. (2000). How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society*, 71(4), 583-597.
- Okamura, Y., Sato, A., Tsuzuki, N., Murakami, M., Heidel-Fischer, H., & Vogel, H. (2019). Molecular signatures of selection associated with host-plant differences in *Pieris* butterflies. *bioRxiv*, 627182.
- Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams.
- Omkar, K. K. (2016). *Ecofriendly pest management for food security*. Academic Press.
- Orrock, J., Connolly, B., & Kitchen, A. (2017). Induced defences in plants reduce herbivory by increasing cannibalism. *Nature Ecology & Evolution*, 1(8), 1205.
- Orsucci, M., Audiot, P., Dorkeld, F., Pommier, A., Vabre, M., Gschloessl, B., Rialle, S., Severac, D., Bourguet, D., & Streiff, R. (2018). Larval transcriptomic response to host plants in two related phytophagous lepidopteran species: implications for host specialization and species divergence. *BMC Genomics*, 19(1), 265.
- Ou, J., Deng, H. M., Zheng, S. C., Huang, L. H., Feng, Q. L., & Liu, L. (2014). Transcriptomic analysis of developmental features of *Bombyx mori* wing disc during metamorphosis. *BMC Genomics*, 15, 820.



## P

- Pashley, D. P. (1986). Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): a sibling species complex? *Annals of the Entomological Society of America*, 79(6), 898-904.
- Pearce, S. L., Clarke, D. F., East, P. D., Elfekih, S., Gordon, K., Jermini, L. S., McGaughran, A., Oakeshott, J. G., Papanikolaou, A., & Perera, O. P. (2017). Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biology*, 15(1), 63.
- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785.
- Pimentel, D. (1991). Diversification of biological control strategies in agriculture. *Crop Protection*, 10(4), 243-253.
- Pimentel, D. (1993). *Habitat factors in new pest invasions*. New York: John Wiley & Sons.
- Pinos, D., Martínez-Solís, M., Herrero, S., Ferré, J., & Hernández-Martínez, P. (2019). The *Spodoptera exigua* ABCC2 acts as a Cry1A receptor independently of its nucleotide binding domain II. *Toxins*, 11(3), 172.
- Pogue, M. (2002). World revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae). *Memoirs of the American Entomological Society*, 43, 1-202.
- Pogue, M. G. (2009). Lepidoptera biodiversity. In *Insect biodiversity: science and society* (pp. 263-293). Blackwell Science Publishing.
- Price, D. R., & Gatehouse, J. A. (2008). RNAi-mediated crop protection against insects. *Trends in Biotechnology*, 26(7), 393-400.

## R

- R Development Core Team (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Ragland, G. J., Almskaar, K., Vertacnik, K. L., Gough, H. M., Feder, J. L., Hahn, D. A., & Schwarz, D. (2015). Differences in performance and transcriptome-wide gene expression associated with *Rhagoletis* (Diptera: Tephritidae) larvae feeding in alternate host fruit environments. *Molecular Ecology*, 24(11), 2759-2776.
- Ramírez-Barahona, S., Sauquet, H., & Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution*, 4(9), 1232-1238.
- Rane, R. V., Walsh, T. K., Pearce, S. L., Jermini, L. S., Gordon, K. H., Richards, S., & Oakeshott, J. G. (2016). Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Current Opinion in Insect Science*, 13, 70-76.
- Rastogi, S., & Liberles, D. A. (2005). Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evolutionary Biology*, 5(1), 1-7.
- Ratzka, A., Vogel, H., Kliebenstein, D. J., Mitchell-Olds, T., & Kroymann, J. (2002). Disarming the mustard oil bomb. *Proceedings of the National Academy of Sciences*, 99(17), 11223-11228.
- Rawlings, N. D., & Barrett, A. J. (1994). Families of serine peptidases. *Methods in Enzymology*, 244, 19-61.
- Reed, W., & Pawar, C. (1982). *Heliothis*: a global problem. In W. Reed & V. Kumble (Eds.), *Proceedings of the international workshop on Heliothis management* (pp. 9-14). ICRISAT.
- Regier, J. C., Mitter, C., Mitter, K., Cummings, M. P., Bazinet, A. L., Hallwachs, W., Janzen, D. H., & Zwick, A. (2017). Further progress on the phylogeny of Noctuoidea (Insecta: Lepidoptera) using an expanded gene sample. *Systematic Entomology*, 42(1), 82-93.
- Regier, J. C., Zwick, A., Cummings, M. P., Kawahara, A. Y., Cho, S., Weller, S., Roe, A., Baixeras, J., Brown, J. W., & Parr, C. (2009). Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evolutionary Biology*, 9(1), 1-21.

- Ren, X.-L., Jiang, W.-L., Ma, Y.-J., Hu, H.-Y., Ma, X.-Y., Ma, Y., & Li, G.-Q. (2016). The *Spodoptera exigua* (Lepidoptera: Noctuidae) ABCC2 mediates Cry1Ac cytotoxicity and, in conjunction with cadherin, contributes to enhance Cry1Ca toxicity in Sf9 cells. *Journal of Economic Entomology*, 109(6), 2281-2289.
- Renuka, P., Madhav, M. S., Padmakumari, A. P., Barbadikar, K. M., Mangrauthia, S. K., Vijaya Sudhakara Rao, K., Marla, S. S., & Ravindra Babu, V. (2017). RNA-seq of rice yellow stem borer *Scirpophaga incertulas* reveals molecular insights during four larval developmental stages. *G3: Genes, Genomes, Genetics*, 7(9), 3031-3045.
- Resh, V. H., & Cardé, R. T. (2009). *Encyclopedia of insects*. Academic press.
- Rivard, D., Cloutier, C., & Michaud, D. (2004). Colorado potato beetles show differential digestive compensatory responses to host plants expressing distinct sets of defense proteins. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, 55(3), 114-123.
- Robinson, G. S., Ackery, P. R., Kitching, I. J., Beccaloni, G. W., & Hernández, L. M. (2010). HOSTS - a database of the world's lepidopteran hostplants. <https://www.nhm.ac.uk/our-science/data/hostplants/>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25.
- Roe, A. D., Weller, S. J., Baixeras, J., Brown, J., Cummings, M. P., Davis, D., Kawahara, A. Y., Parr, C., Regier, J. C., & Rubinoff, D. (2009). Evolutionary framework for Lepidoptera model systems. In Goldsmith, M. R., Marec, F. (Eds.), *Genetics and molecular biology of Lepidoptera* (pp. 1-24), CRC Press.
- Rohlf, R. V., & Nielsen, R. (2015). Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Systematic Biology*, 64(5), 695-708.
- Rotllant, G., Nguyen, T. V., Sbragaglia, V., Rahi, L., Dudley, K. J., Hurwood, D., Ventura, T., Company, J. B., Chand, V., Aguzzi, J., & Mather, P. B. (2017). Sex and tissue specific gene expression patterns identified following de novo transcriptomic analysis of the Norway lobster, *Nephrops norvegicus*. *BMC Genomics*, 18(1), 622.
- Roy, A., Walker III, W., Vogel, H., Chattington, S., Larsson, M., Anderson, P., Heckel, D. G., & Schlyter, F. (2016). Diet dependent metabolic responses in three generalist insect herbivores *Spodoptera* spp. *Insect Biochemistry and Molecular Biology*, 71, 91-105.
- Ryan, S. F., Lombaert, E., Espeset, A., Vila, R., Talavera, G., Dincă, V., Doellman, M. M., Renshaw, M. A., Eng, M. W., & Horne, E. A. (2019). Global invasion history of the agricultural pest butterfly *Pieris rapae* revealed with genomics and citizen science. *Proceedings of the National Academy of Sciences*, 116(40), 20015-20024.

## S

- Saeed, Q., Ahmad, F., & Saeed, S. (2017). Development and survival of *Spodoptera exigua* (Lepidoptera: Noctuidae) on alternate crops in cotton cropping pattern, with implications to integrated pest management. *Environmental Entomology*, 46(3), 595-601.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- Sarraz, M., Keddie, A. B., & Dosdall, L. M. (2005). Biological control of the diamondback moth, *Plutella xylostella*: a review. *Biocontrol Science and Technology*, 15(8), 763-789.
- Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., Funk, K., Ketter, A., Kim, S., & Kimchi, A. (2020). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 48(1), D9.
- Schmieder, R., & Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS ONE*, 6(3), e17288.

- Schoonhoven, L. M., Van Loon, B., van Loon, J. J., & Dicke, M. (2005). *Insect-plant biology*. Oxford University; Press on Demand.
- Schramm, K., Vassão, D. G., Reichelt, M., Gershenzon, J., & Wittstock, U. (2012). Metabolism of glucosinolate-derived isothiocyanates to glutathione conjugates in generalist lepidopteran herbivores. *Insect Biochemistry and Molecular Biology*, 42(3), 174-182.
- Schuler, M. A. (2011). P450s in plant-insect interactions. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814(1), 36-45.
- Schuler, M. A. (2012). Insect P450s: mounted for battle in their war against toxins. *Molecular Ecology*, 21(17), 4157-4159.
- Schulz, R. (2004). Field studies on exposure, effects, and risk mitigation of aquatic nonpoint-source insecticide pollution: a review. *Journal of Environmental Quality*, 33(2), 419-448.
- Schulz, R., Bub, S., Petschick, L. L., Stehle, S., & Wolfram, J. (2021). Applied pesticide toxicity shifts toward plants and invertebrates, even in GM crops. *Science*, 372(6537), 81-84.
- Schweizer, F., Heidel-Fischer, H., Vogel, H., & Reymond, P. (2017). *Arabidopsis* glucosinolates trigger a contrasting transcriptomic response in a generalist and a specialist herbivore. *Insect Biochemistry and Molecular Biology*, 85, 21-31.
- Scott, J. G., Michel, K., Bartholomay, L. C., Siegfried, B. D., Hunter, W. B., Smagghe, G., Zhu, K. Y., & Douglas, A. E. (2013). Towards the elements of successful insect RNAi. *Journal of Insect Physiology*, 59(12), 1212-1221.
- Scriber, J., Lederhouse, R., & Hagen, R. (1991). Foodplants and evolution within *Papilio glaucus* and *Papilio troilus* species groups (Lepidoptera: Papilionidae). In L. T. Price PW, Fernandes GW, Benson WW (Eds.), *Plant-animal interactions: evolutionary ecology in tropical and temperate regions*. (pp. 341-374). John Wiley & Sons Inc.
- Seppey, M., Ioannidis, P., Emerson, B. C., Pitteloud, C., Robinson-Rechavi, M., Roux, J., Escalona, H. E., McKenna, D. D., Misof, B., & Shin, S. (2019). Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. *Genome Biology*, 20(1), 98.
- Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962, 227-245.
- Sexton, J. P., Montiel, J., Shay, J. E., Stephens, M. R., & Slatyer, R. A. (2017). Evolution of ecological niche breadth. *Annual Review of Ecology, Evolution, and Systematics*, 48, 183-206.
- Shi, H., Pei, L., Gu, S., Zhu, S., Wang, Y., Zhang, Y., & Li, B. (2012). Glutathione S-transferase (GST) genes in the red flour beetle, *Tribolium castaneum*, and comparative analysis with five additional insects. *Genomics*, 100(5), 327-335.
- Shi, Y., Wang, H., Liu, Z., Wu, S., Yang, Y., Feyereisen, R., Heckel, D. G., & Wu, Y. (2018). Phylogenetic and functional characterization of ten P450 genes from the CYP6AE subfamily of *Helicoverpa armigera* involved in xenobiotic metabolism. *Insect Biochemistry and Molecular Biology*, 93, 79-91.
- Sikkink, K. L., Kobiela, M. E., & Snell-Rood, E. C. (2017). Genomic adaptation to agricultural environments: cabbage white butterflies (*Pieris rapae*) as a case study. *BMC Genomics*, 18(1), 412.
- 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(19), 3210-3212.
- Simon, J.-C., d'Alençon, E., Guy, E., Jacquín-Joly, E., Jaquière, J., Nouhaud, P., Peccoud, J., Sugio, A., & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, 14(6), 413-423.
- Singh, K. S., Hosken, D. J., Wedell, N., Bass, C., Baxter, S., Paszkiewicz, K., & Sharma, M. D. (2020). De novo genome assembly of the meadow brown butterfly, *Maniola jurtina*. *G3: Genes, Genomes, Genetics*, 10(5), 1477-1484.
- Smith, T. F. (2008). Diversity of WD-Repeat proteins. In C. S. Clemen, L. Eichinger, & V. Rybakín (Eds.), *The coronin family of proteins* (Vol. 48). Subcellular Biochemistry.
- Smits, P. H., Van De Vrie, M., & Vlak, J. M. (1986). Oviposition of beet armyworm (Lepidoptera: Noctuidae) on greenhouse crops. *Environmental Entomology*, 15(6), 1189-1191.

- Sorensen, J. S., & Dearing, M. D. (2006). Efflux transporters as a novel herbivore countermechanism to plant chemical defenses. *Journal of Chemical Ecology*, 32(6), 1181-1196.
- Soubrier, J., Steel, M., Lee, M. S., Der Sarkissian, C., Guindon, S., Ho, S. Y., & Cooper, A. (2012). The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution*, 29(11), 3345-3358.
- Sparks, T. C., & Nauen, R. (2015). IRAC: mode of action classification and insecticide resistance management. *Pesticide Biochemistry and Physiology*, 121, 122-128.
- Stacke, R. F., Arnemann, J. A., Rogers, J., Stacke, R. S., Strahl, T. T., Perini, C. R., Dossin, M. F., Pozebon, H., de Arruda Cavallin, L., & Guedes, J. V. (2018). Damage assessment of *Helicoverpa armigera* (Lepidoptera: Noctuidae) in soybean reproductive stages. *Crop Protection*, 112, 10-17.
- Stanley, J., Preetha, G., & Stanley. (2016). *Pesticide toxicity to non-target organisms*. Springer.
- Strauss, G. (2010). Pest risk analysis of *Metcalfa pruinosa* in Austria. *Journal of Pest Science*, 83(4), 381-390.
- Strong, D. R., Lawton, J. H., & Southwood, S. R. (1984). *Insects on plants. Community patterns and mechanisms*. Blackwell Scientific Publications.
- Sun, W., Shen, Y.-H., Yang, W.-J., Cao, Y.-F., Xiang, Z.-H., & Zhang, Z. (2012). Expansion of the silkworm GMC oxidoreductase genes is associated with immunity. *Insect Biochemistry and Molecular Biology*, 42(12), 935-945.
- Sutherst, R. W. (1991). Predicting the survival of immigrant insect pests in new environments. *Crop Protection*, 10(5), 331-333.
- Suzuki, R., & Shimodaira, H. (2006). Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542.

## T

- Tay, W. T., Soria, M. F., Walsh, T., Thomazoni, D., Silvie, P., Behere, G. T., Anderson, C., & Downes, S. (2013). A brave new world for an old world pest: *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Brazil. *PLOS ONE*, 8(11), e80134.
- Terenius, O., Papanicolaou, A., Garbutt, J. S., Eleftherianos, I., Huvenne, H., Kanginakudru, S., Albrechtsen, M., An, C., Aymeric, J.-L., & Barthel, A. (2011). RNA interference in Lepidoptera: an overview of successful and unsuccessful studies and implications for experimental design. *Journal of Insect Physiology*, 57(2), 231-245.
- Theiling, K. M., & Croft, B. (1988). Pesticide side-effects on arthropod natural enemies: a database summary. *Agriculture, Ecosystems & Environment*, 21(3-4), 191-218.
- 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 Biology*, 21(1), 15.
- Thompson, J. N. (1989). Concepts of coevolution. *Trends in Ecology & Evolution*, 4(6), 179-183.
- Thorpe, P., Escudero-Martinez, C. M., Cock, P. J. A., Eves-van den Akker, S., & Bos, J. I. B. (2018). Shared transcriptional control and disparate gain and loss of aphid parasitism genes. *Genome Biology and Evolution*, 10(10), 2716-2733.

- Tian, L., Zeng, Y., Xie, W., Wu, Q., Wang, S., Zhou, X., & Zhang, Y. (2019). Genome-wide identification and analysis of genes associated with RNA interference in *Bemisia tabaci*. *Pest Management Science*, 75(11), 3005-3014.
- Triant, D. A., Cinel, S. D., & Kawahara, A. Y. (2018). Lepidoptera genomes: current knowledge, gaps and future directions. *Current Opinion in Insect Science*, 25, 99-105.
- Truman, J. W., & Riddiford, L. M. (2019). The evolution of insect metamorphosis: a developmental and endocrine view. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1783), 20190070.
- Tshikolovets, V. V. (2011). *Butterflies of Europe & the Mediterranean area*. Tshikolovets Publications.
- Tzin, V., Hojo, Y., Strickler, S. R., Bartsch, L. J., Archer, C. M., Ahern, K. R., Zhou, S., Christensen, S. A., Galis, I., & Mueller, L. A. (2017). Rapid defense responses in maize leaves induced by *Spodoptera exigua* caterpillar feeding. *Journal of Experimental Botany*, 68(16), 4709-4723.

## U

- UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(1), 506-515.

## V

- van Nieukerken, E. J., Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., Mitter, C., Mutanen, M., Regier, J. C., & Simonsen, T. J. (2011). Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) *Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness*. *Zootaxa*, 3148(1), 212-221.
- Veerassamy, S., Smith, A., & Tillier, E. R. (2003). A transition probability model for amino acid substitutions from blocks. *Journal of Computational Biology*, 10(6), 997-1010.
- Voelckel, C., & Jander, G. (2014). *Insect-Plant Interactions* (Vol. 47). John Wiley & Sons.
- Volney, W. J. A., & Fleming, R. A. (2007). Spruce budworm (*Choristoneura* spp.) biotype reactions to forest and climate characteristics. *Global Change Biology*, 13(8), 1630-1643.

## W

- 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(11), e112963.
- Wang, R.-L., He, Y.-N., Staehelin, C., Liu, S.-W., Su, Y.-J., & Zhang, J.-E. (2017a). Identification of two cytochrome monooxygenase P450 genes, CYP321A7 and CYP321A9, from the tobacco cutworm moth (*Spodoptera litura*) and their expression in response to plant allelochemicals. *International Journal of Molecular Sciences*, 18(11), 2278.
- Wang, R.-L., Zhu-Salzman, K., Baerson, S. R., Xin, X. W., Li, J., Su, Y. J., & Zeng, R. S. (2017b). Identification of a novel cytochrome P450 CYP321B1 gene from tobacco cutworm (*Spodoptera litura*) and RNA interference to evaluate its role in commonly used insecticides. *Insect Science*, 24(2), 235-247.
- War, A. R., Taggar, G. K., Hussain, B., Taggar, M. S., Nair, R. M., & Sharma, H. C. (2018). Plant defence against herbivory and insect adaptations. *AoB Plants*, 10(4), 037.
- Wen, Z., Rupasinghe, S., Niu, G., Berenbaum, M. R., & Schuler, M. A. (2006). CYP6B1 and CYP6B3 of the black swallowtail (*Papilio polyxenes*): adaptive evolution through subfunctionalization. *Molecular Biology and Evolution*, 23(12), 2434-2443.
- Wheat, C. W., Vogel, H., Wittstock, U., Braby, M. F., Underwood, D., & Mitchell-Olds, T. (2007). The genetic basis of a plant-insect coevolutionary key innovation. *Proceedings of the National Academy of Sciences*, 104(51), 20427-20431.

- Wheeler, W. (2002). *Pesticides in agriculture and the environment*. Marcel Dekker Inc.
- Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*, 18(5), 691-699.
- Whittaker, R. H., & Feeny, P. P. (1971). Allelochemicals: chemical interactions between species. *Science*, 171(3973), 757-770.
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.
- Wiegmann, B. M., Mitter, C., Regier, J. C., Friedlander, T. P., Wagner, D. M., & Nielsen, E. S. (2000). Nuclear genes resolve Mesozoic-aged divergences in the insect order Lepidoptera. *Molecular Phylogenetics and Evolution*, 15(2), 242-259.
- Wittstock, U., Agerbirk, N., Stauber, E. J., Olsen, C. E., Hippler, M., Mitchell-Olds, T., Gershenzon, J., & Vogel, H. (2004). Successful herbivore attack due to metabolic diversion of a plant chemical defense. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), 4859-4864.
- Worner, S. (2002). Predicting the invasive potential of exotic insects. In Hallman, G. J. & Schwalbe, C. P., *Invasive arthropods in agriculture* (pp. 119-137). Science Publishers.
- Worner, S., & Gevrey, M. (2006). Modelling global insect pest species assemblages to determine risk of invasion. *Journal of Applied Ecology*, 43(5), 858-867.
- Wybouw, N., Pauchet, Y., Heckel, D. G., & Van Leeuwen, T. (2016). Horizontal gene transfer contributes to the evolution of arthropod herbivory. *Genome Biology and Evolution*, 8(6), 1785-1801.
- Wybouw, N., Zhurov, V., Martel, C., Bruinsma, K. A., Hendrickx, F., Grbić, V., & Van Leeuwen, T. (2015). Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host. *Molecular Ecology*, 24(18), 4647-4663.

## X

- Xiao, H., Ye, X., Xu, H., Mei, Y., Yang, Y., Chen, X., Yang, Y., Liu, T., Yu, Y., & Yang, W. (2020). The genetic adaptations of fall armyworm *Spodoptera frugiperda* facilitated its rapid global dispersal and invasion. *Molecular Ecology Resources*, 20(4), 1050-1068.
- Xu, W., Papanicolaou, A., Zhang, H.-J., & Anderson, A. (2016). Expansion of a bitter taste receptor family in a polyphagous insect herbivore. *Scientific Reports*, 6, 23666.
- Xue, X.-Y., Mao, Y.-B., Tao, X.-Y., Huang, Y.-P., & Chen, X.-Y. (2012). New approaches to agricultural insect pest control based on RNA interference. In Jokusch, E. L., *Advances in insect physiology* (Vol. 42, pp. 73- 117). Academic Press.

## Y

- Yan, Y., Wang, Y.-C., Feng, C.-C., Wan, P.-H. M., & Chang, K. T.-T. (2017). Potential distributional changes of invasive crop pest species associated with global climate change. *Applied Geography*, 82, 83-92.
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39(3), 306-314.
- Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11(2), 14.

## Z

- Zagobelny, M., & Møller, B. L. (2011). Cyanogenic glucosides in the biological warfare between plants and insects: the burnet moth-birdsfoot trefoil model system. *Phytochemistry*, 72(13), 1585-1592.
- Zahiri, R., Lafontaine, D., Schmidt, C., Holloway, J. D., Kitching, I. J., Mutanen, M., & Wahlberg, N. (2013). Relationships among the basal lineages of Noctuidae (Lepidoptera, Noctuoidea) based on eight gene regions. *Zoologica Scripta*, 42(5), 488-507.
- Zeng, R. S., Wen, Z., Niu, G., Schuler, M. A., & Berenbaum, M. R. (2007). Allelochemical induction of cytochrome P450 monooxygenases and amelioration of xenobiotic toxicity in *Helicoverpa zea*. *Journal of Chemical Ecology*, 33(3), 449-461.
- Zhang, M., Fang, T., Pu, G., Sun, X., Zhou, X., & Cai, Q. (2013). Xenobiotic metabolism of plant secondary compounds in the English grain aphid, *Sitobion avenae* (F.) (Hemiptera: Aphididae). *Pesticide Biochemistry and Physiology*, 107(1), 44-49.
- Zhao, S., Ye, Z., & Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *Rna*, 26(8), 903-909.
- Zhou, C. Y., Zha, X. F., Liu, C., Han, M. J., Zhang, L. Y., Shi, P. P., Wang, H., Zheng, R. W., & Xia, Q. Y. (2016). Identification of MBF2 family genes in *Bombyx mori* and their expression in different tissues and stages and in response to *Bacillus bombysepticus* infection and starvation. *Insect Science*, 23(4), 502-512.
- Zhou, S., Lou, Y.-R., Tzin, V., & Jander, G. (2015). Alteration of plant primary metabolism in response to insect herbivory. *Plant Physiology*, 169(3), 1488-1498.
- Zhou, Y., Zhu, J., Tong, T., Wang, J., Lin, B., & Zhang, J. (2019). A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinformatics*, 20(1), 163.
- Zhu, F., & Liu, N. (2008). Differential expression of CYP6A5 and CYP6A5v2 in pyrethroid-resistant house flies, *Musca domestica*. *Archives of Insect Biochemistry and Physiology*, 67(3), 107-119.
- Zhu-Salzman, K., & Zeng, R. (2015). Insect response to plant defensive protease inhibitors. *Annual Review of Entomology*, 60, 233-252.
- Ziska, L. H., Blumenthal, D. M., Runion, G. B., Hunt, E. R., & Diaz-Soltero, H. (2011). Invasive species and climate change: an agronomic perspective. *Climatic Change*, 105(1), 13-42.





## Summary

The coevolutionary relationship between herbivorous insects and plants dates back more than 400 million years. Plant-feeding insects have adapted to cope with plant anti-herbivory defences. A specific type of plant defence is the use of toxins, or specialised metabolites, that plants produce and use in defence to deter or poison attacking herbivores. A common mechanism of herbivorous insects is the use of a three-phased detoxification pathway to detoxify the specialised metabolites in response. Monophagous herbivores, specialised on feeding on a single or small group of plant species, often evolved adaptations to cope with these host plant-specific defences. In contrast, polyphagous species are able to feed on a large range of plants belonging to different plant families. Polyphagous species need to employ a diversity of detoxification strategies to cope with the variation of specialised metabolites, and moreover, being able to flexibly use these strategies whenever necessary. However, how polyphagous insect species cope with the variety of defences on a genetic scale is little understood. Moreover, many polyphagous insect species form notorious pests, causing major infestations and damage to agriculture worldwide, yet we know little about the genetic differences and commonalities of these insect species forming pests.

My thesis aims to elucidate the genomic and evolutionary basis of polyphagous pest insect species. I studied polyphagous pest formations (species forming pests) by examining which genes and mechanisms enable feeding on a range of host plant species in Lepidoptera. In short, I tried to answer the question: ‘What makes a pest a pest on a genetic scale?’. I studied butterflies and moths, Lepidoptera, with a specific focus on the cutworm moth family Noctuidae. Many species within the Noctuidae are polyphagous species, including some of the most devastating agricultural pests worldwide.

In my experimental set up, I studied five different polyphagous Noctuidae species: the cabbage looper (*Trichoplusia ni*), the silver-Y moth (*Autographa gamma*), the cabbage moth (*Mamestra brassicae*), the African cotton leafworm (*Spodoptera littoralis*) and the beet armyworm (*Spodoptera exigua*). I focused on the gene expression of these moths feeding on three selected host plant species, cabbage (*Brassica oleracea*), maize (*Zea mays*) and tobacco (*Nicotiana tabacum*), and an artificial diet-based control. By studying the transcriptional plasticity while feeding on different host species I examined the genetic mechanisms of polyphagy linked to herbivore success levels (or pest formations). I implemented a comparative phylogenomic approach to distinguish patterns shared due to evolutionary relatedness from those convergent patterns that could be associated with herbivore success.

The transcriptional activity of the moths feeding on the different host plants was highly plastic and moth species-specific. Nonetheless, the expression of groups of homologous genes was shared among related lineages, *Spodoptera*, Noctuidae and Plusiinae, in response to specific host plant species (lineage-specific patterns). Moreover, multiple homologous genes were identified with shared expression among

species with high herbivore success for specific host plant species independent of phylogenetic placement, indicating convergence. This shows that ‘successful’ polyphagous Noctuidae moths, or potential pests, share a common gene expression response which may be of importance in pest formations. These genes are of particular interest for future studies aiming for a deeper understanding of the genetic mechanisms of potential pest formations in herbivorous insects and the development of sustainable, targeted ways of controlling pest formations.

Further, I performed a comparative genomics study of 37 lepidopteran genomes to investigate the genomic evolutionary patterns linked to polyphagy in butterflies and moths. I tested if gene family evolution, in terms of gene gains and losses, can be associated with polyphagy as has been hypothesized in the literature. To quantify host range, or level of polyphagy, I calculated the phylogenetic diversity (PD) and specialised metabolite content (FMD) of plant species in the diet of each lepidopteran species. I focused on expansion levels of gene families involved in plant feeding, such as detoxification genes. The expansions were observed in both polyphagous and monophagous lepidopteran species and were species-specific. Moreover, evolutionary expansion rates varied across lepidopteran families. Nevertheless, I also observed a significant positive correlation between the expansion of carboxyl- and choline esterases (CCE) and glutathione S-transferases (GST) detoxification gene families and host plant family range across polyphagous lepidopteran species. The outcomes of the study are important for understanding the evolution of polyphagy, and hints at adaptive evolutionary ties between the expansion of CCE and GST gene families and wider host range in Lepidoptera.

Finally, I presented the genome of the polyphagous Noctuidae beet armyworm, *Spodoptera exigua*. Many species in the genus *Spodoptera* are considered major polyphagous pest species causing infestations, destroying harvests and diminishing crop yields in many parts of the world. By performing developmental transcriptional analysis in combination with comparative genomics, I identified potential lineage-specific candidate genes for targeted RNA interference (RNAi)-mediated control of *Spodoptera* pest species. RNAi-mediated control of insect pest formations can be an efficient and effective tool if selected gene(s) are essential for development or survival. Additionally, RNAi can be a promising sustainable technique if targeted gene(s) are pest species-specific, without harming non-target organisms and the ecosystem. I was able to identify seven *Spodoptera*-specific genes expressed during the destructive larval phase, of which four were annotated to be involved in transport, pathogen response and digestion. The *Spodoptera*-specificity of the identified target genes was additionally verified by analysing homologous relationships of a broad selection of other insect species. I presented these genes as potential candidates for sustainable, targeted RNAi in order to control the destructive larvae of *Spodoptera*.

This research focused on the genetic basis of polyphagous herbivory and pest formations in Lepidoptera within an evolutionary perspective. The findings help in the understanding of insect pest formations and polyphagous herbivory and may serve in finding sustainable and respectful ways of controlling pest formations.

## Samenvatting

De nauwe co-evolutionaire relatie die herbivore insecten en planten met elkaar hebben gaat meer dan 400 miljoen jaar terug. Planten hebben een grote variatie aan afweermechanismen ontwikkeld waarbij de productie en het gebruik van plant toxines één van de belangrijkste is. De plant gebruikt deze toxines om insecten te vergiften of verjagen als afweer tegen herbivorie. Op hun beurt hebben insecten een grote verscheidenheid aan mechanismen ontwikkeld om zich tegen deze verdedigingen te wapenen. Een algemene strategie is het ontgiften van deze toxines in drie fasen door middel van een detoxificatie mechanisme. Monofage insectensoorten, welke voor hun voeding afhankelijk zijn van één of enkele, tot dezelfde familie behorende, plantensoorten, hebben veelal gespecialiseerde mechanismen ontwikkeld tegen deze plant specifieke verdedigingen. Polyfage insecten daarentegen kunnen zich voeden met een grote variatie aan waardplanten afkomstig van meerdere plantenfamilies en moeten daarom flexibel kunnen omgaan met een grote variatie aan plantverdedigingsmechanismen. Ondanks deze mechanismen tegen herbivorie zijn er polyfage insecten die wereldwijd plagen veroorzaken en daarmee op grote schaal schade toebrengen aan landbouwgewassen. Hoe deze polyfage insecten op genetische schaal omgaan met de grote verscheidenheid aan afweermechanismen is nog grotendeels onbekend. Bovendien is onbekend of plaagvormende insecten gebruik maken van een algemeen voorkomend en gedeeld genetisch mechanisme als reactie op plantenafweer of dat er grote verschillen tussen soorten voorkomen.

In mijn proefschrift heb ik onderzoek gedaan naar de genetische en evolutionaire basis van plaagvorming in polyfage insecten teneinde de vraag: “Wat maakt een plaaginsect tot een plaaginsect, op een genetisch niveau?” te beantwoorden. Ik heb onderzocht welke genetische mechanismen van belang zijn en worden gebruikt bij potentiële plaagvorming in (nacht)vinders, de Lepidoptera. Nadruk van mijn onderzoek lag onder meer op de genexpressie van nachtvinders uit de familie van de ‘uilen’, de Noctuidae. Veel soorten binnen deze familie zijn polyfaag en wereldwijd berucht om hun vermogen plagen te vormen op bekende en belangrijke landbouwgewassen.

Voor dit onderzoek heb ik de genexpressie vergeleken van vijf verschillende polyfage nachtvinders: de ni-uil (*Trichoplusia ni*), de gamma-uil (*Autographa gamma*), de kooluil (*Mamestra brassicae*), de katoenuil (*Spodoptera littoralis*) en de Florida-uil (*Spodoptera exigua*), etend van verschillende waardplanten: kool (*Brassica oleracea*), mais (*Zea mays*) en tabak (*Nicotiana tabacum*). Aan de hand van deze vergelijking was het mogelijk om de genexpressie van verschillende motten op verschillende waardplanten, met ieder hun specifieke afweermechanismen, te onderzoeken en te koppelen aan de mate van herbivorie, en dus de mate van plaagvorming. Doordat de evolutionaire verwantschappen van de nachtvinders bekend zijn in de vergelijking, en werden meegenomen in de analyse, konden de overeenkomende genexpressiepatronen onder invloed van verwantschap geïdentificeerd worden (“clade-specifieke patronen”). Op deze manier konden genen worden geïdentificeerd waarvan de activiteit gelijk is

in plaagvormende nachtvlinders op één van de waardplanten, onafhankelijk van een mogelijk gedeelde evolutionaire historie (“convergente patronen”).

De gemeten genexpressie in de individuele nachtvlindersoorten etend van de waardplanten bleek zeer variabel en soort-specifiek. De expressie van meerdere groepen homologe genen werd gedeeld tussen nauw verwante soorten binnen de clades *Spodoptera*, Noctuidae en Plusiinae in respons op de planten (=clade-specifieke patronen). Daarnaast werden, onafhankelijk van verwantschap, meerdere homologe genen geïdentificeerd waarvan de gen activiteit gelijk was in nachtvlinders met grootste ‘vraat succes’ (of herbivorie succes) op de specifieke waardplanten (=convergente patronen). Deze genen zijn een indicatie voor een overeenkomende regulatie van gedeelde genen tussen polyfage Noctuidae nachtvlinders die mogelijk belangrijk zijn voor plaagvorming. De identificatie van deze genen is enerzijds interessant om de genetische mechanismen achter polyfagie en plaagvorming te ontrafelen en anderzijds voor de ontwikkeling van duurzame en doelgerichte plaagbestrijding.

Daarnaast heb ik vergelijkend onderzoek gedaan aan 37 (nacht)vlindergenomen. Ik heb de correlatie tussen de evolutie van genfamilies door expansie (toename van het aantal genen per familie) of contractie (afname van het aantal genen) en polyfagie onderzocht. Hiervoor werd voor iedere (nacht)vlinder soort in de vergelijking de graad van polyfagie gekwantificeerd op basis van: 1) de fylogenetische diversiteit van de waardplanten in het dieet (PD waarde) en 2) de diversiteit aan plant toxines in het dieet van iedere (nacht)vlindersoort (FMD waarde). Ik heb mij met name gericht op de associatie tussen expansie van genfamilies die betrokken zijn bij herbivorie, zoals genen die betrokken zijn bij de detoxificatie van plant toxines, en polyfagie. Genfamilie-expansies worden vaak in verband gebracht met het mogelijk maken van polyfagie bij insecten, en is daarmee van belang voor dit onderzoek.

Genfamilie-expansies werden aan de hand van de analyses aangetroffen in zowel polyfage als monofage soorten en lijken daarmee soort-specifiek voor te komen. Specifiek heb ik de mate van expansies en contracties in de loop van de evolutie onderzocht en vergeleken tussen (nacht)vlinder families. De mate van expansies verschilde sterk tussen de (nacht)vlinder families. Desondanks werd een significant positieve correlatie gevonden tussen expansie van de detoxificatie genfamilies carboxyl- en cholinesterases (CCE) en glutathion S-transferases (GST) en hoeveelheid plantfamilies in het dieet van polyfage (nacht)vlinders. De resultaten zijn van belang om de evolutie van polyfagie te begrijpen, en laat de mogelijke adaptieve evolutionaire associatie zien tussen de expansie van de detoxificatie genfamilies CCE en GST en een grotere mate van polyfagie.

Tot slotte analyseerde ik het genoom van de polyfage Florida-uil, *Spodoptera exigua*. Veel soorten binnen het geslacht *Spodoptera* zijn polyfage soorten en zijn berucht om het vormen van wereldwijde plagen op belangrijke landbouwgewassen. Met behulp van een combinatie van genexpressie analyses van verschillende ontwikkelingsstadia en vergelijkend onderzoek aan de genomen van verwante (nacht)vlinder soorten

heb ik meerdere *Spodoptera*-specifieke genen geselecteerd die mogelijk gebruikt kunnen worden in doelgerichte plaagbestrijding van *Spodoptera*-plagen met behulp van RNA-interferentie technieken (RNAi). Doelgerichte plaagcontrole met behulp van RNAi kan zeer efficiënt en effectief zijn als de beoogde genen belangrijk zijn voor de ontwikkeling van het organisme. Daarnaast kan RNAi zeer duurzaam worden toegepast als de geselecteerde genen uniek en specifiek zijn voor de doelsoort, waardoor andere organismen en het ecosysteem niet worden beïnvloed. In totaal heb ik zeven *Spodoptera*-specifieke genen geïdentificeerd die specifiek in het larvale stadium, het meest destructieve levensstadium voor de landbouw, verhoogde expressie toonden. Vier van deze genen werden verder geïdentificeerd en bleken betrokken bij processen behorende tot transport, pathogeen respons en vertering. Ten slotte werd de *Spodoptera*-specificiteit verder bevestigd door analyse en vergelijking van de evolutionaire relaties van homologe genen van andere insectensoorten. De geselecteerde genen presenteer ik als potentiële kandidaten voor een duurzame, doelgerichte plaagcontrole van *Spodoptera* larven met behulp van RNAi.

Dit onderzoek heeft zich gericht op de genetische basis van polyfagie en plaagvorming binnen Lepidoptera met gebruik van een evolutionair onderzoekskader. De resultaten zijn van belang om de genetica achter plaagvorming door insecten en polyfagie beter te begrijpen en als initiële stappen voor het vinden van duurzame controle en inperking van plaagvorming.

## Acknowledgements

Years ago, around 1995 when I was 5 years old, I was fascinated by the sight of a large shiny beetle in my parents' garden. The beetle, likely a large *Carabus* ground beetle, was collected in a large bucket and closely observed: the young naturalist Thijmen was born! In the coming years, many different garden- critters and creatures ended up in that bucket. Fair and painful, to say: not all of them left the bucket alive due to ignorance, negligence or because of higher trophic levels interacting (my apologies to them all).

In the years to follow, I went through phases of rearing beetles in the kitchen, cleaning skulls in the garden (with great pleasure of our direct neighbours), making herbariums and collecting insects. The latter stayed and became something permanent, with a strong interest and passion on beetles. Many family holidays were spent in nature observing and collecting insects, with an intricate system of running and catching by me, keeping up with the administration and data collection by my Mom and additional collecting and photographing by my Dad.

The open-mindedness, respect and appreciation I received from my parents have shaped who I am today. The constant flow of visits to museums, archaeological sites and hiking trips in nature have thought me to be and stay curious and interested in the world around me. I am certain that this initiated my motivation and perseverance to do research and eventually do and finish my PhD. I want to thank you, my Mom and Dad, Gerrie and Chris, for the childhood that you gave me. Especially in the past years, I finally allowed myself to be who I really am and I want to thank you for being there for me during all the tough moments!

It has also been in my PhD-years that I have met you, Niall, the beautiful, super-enthusiastic, strong-minded person that you are. Thanks to the Covid-19 restrictions, we were able to stay with each other and live together much sooner than expected. We moved every week from our “summer residence” in Wageningen to our “city apartment” in The Hague, together with Puck and, a little later, Willemijn. Thanks for being there for me at all times, and especially in the final period of my PhD where you helped me find the motivation to continue. We are similar in many aspects, and yet we are so different. I think this is the secret of our love and success. Together we can be ourselves, wherever this may be on this world!

I also want to thank all the support I got from my sisters, Ellen and Eveline, and their families; Silvester, Amy and Mick, and; Peter, Bram, Jet and Cato. We have shared so many great family-moments in the past, and we will in the future! It has been (and still is) so nice to be so close to you and to see all my nieces and nephews grow up!

This thesis would not have been completed without the daily guidance of my supervisors. I really want to thank my co-promotor Sabrina Simon, for supervising me day in and day out. A huge thanks for guiding me along the path of becoming an independent researcher. I especially want to thank you for being such an inspirational scientist, I have learned a great deal from your perseverance, punctuality, eye for

detail and the never-ending advice to be precise and clear in the things that I actually want to say! I do not think that I am going to miss the “unclear, rephrase!” comments in my manuscripts, but this advice will stick to my mind forever and will for sure improve my writing for always! I know I was your first PhD student and it seems I made it!

Further I want to thank my promotor, Eric Schranz. Your great enthusiasm, your amazing ability to connect topics with each other, and broad interest has been a huge inspiration. The fact that you are such a relaxed person, allowing everyone within the Biosystematics Group to work and think freely without strict office rules, makes the group as pleasant as it is. I want to thank you for always reminding me of the positive parts of my research, and the enthusiasm in times I did not see how to continue. Your quote “Create your own reality” got a permanent position on my desk, and it will stay there!

I want to thank everyone at the Biosystematics Group, both the people that have left already or that are still there. I have great memories of the numerous discussions and times off with each other. In particular I won’t forget the various lab outings including the inspirational video at the butterfly garden and the tour through Amerongen! Nina Fatouros, Klaas Bouwmeester, Marco Busscher, Kitty Vijverberg, Freek Bakker, Lars Chatrou, Robin van Velzen, Casper Quist, Patrick Verbaarschot, Nora Walden, Nynke Groendijk-Wilders and Roel Lemmens, I am extremely thankful for the great time I had and all the advice I received from all of you.

I further want to thank all the fellow former and current PhD students I had the pleasure with to work together, share thoughts and ideas, or just complain about anything: Floris Breman, Niccolò Bassetti, Wei Xiong, Sara van de Kerke, Setareh Mohammadin, Dêêdi Sogbohossou, Eddy Griese, Tao Zhao and Wouter Makkinje.

Wilma Twigt, thanks so much for the many chit-chats, vegan recipes and yoga sessions. It is great to have had a stable and relaxed coffee/tea-mate in the group. I am convinced that the whole Biosystematics machinery will get stuck without you!

And then, early 2020, Covid-restrictions were introduced and we all had to work from home. To be honest, this brought me many good things but did prevent me to work together and share office-time with all “old” colleagues and get to know so many new people at Biosystematics: what a shame. I hope you have a great time ahead of you. I am definitely going to miss the daily breaks, casual talks and discussions that often-formed relaxed moments to share our opinions and thoughts (whether it be politics, mowing regimes or museum visits). Often these laid-back discussions were the initial steps for *eureka!* moments. Additionally, I want to thank all former Radix-neighbours from the CGN which I shared many of these breaks and good laughs with. And thanks to my colleagues at the Laboratory of Genetics and Laboratory of Entomology, either with your advice, the discussions that we had and/or the shared summer and Christmas dinners.

Further I want to thank all my co-authors for their help and contributions. In particular, I want to thank Vera Ros (co-authoring 3 papers!) for your help and comments that helped improving my work. Thanks for introducing me to the world of *Spodoptera exigua*. I also want to thank all people that further helped, guided and

supported me in the world of plant breeding, insect rearing and feeding assays. I want to thank the Unifarm personnel for their great plant care in times I did not. I want to thank Maike Stam (UvA), Jessica Nifon (US Nicotiana Germplasm Collection), Erik Poelman (WUR), Els Roode (WUR), the late Hanke Bloksma (WUR), and John Dedes (Great Lakes Forestry Centre, Canada) that all have kindly provided me with plant or moth material and loads of advice to do my own rearing and experiments. I want to thank my students Jorn School and Daniel de Groot for their contributions.

During my PhD project I had the great pleasure to be involved in teaching, which I enjoyed a lot. In the first year of the “Biodiversity of the Netherlands” course I had the pleasure to meet the ultra-Wageningen biologists Eva Drukker, Hans Linssen and Corné van der Linden that I can now call my friends! Thanks so much for the great moments and crazy “expeditions” we had together, and dragging me out of my beetle-bubble. All the best with your PhD’s, whether it already started or still has to start. Let’s keep the expeditions coming! And: lets continue with our Papua plans... ;-)

Eva, I am happy with you as my paranymph! Your never-ending enthusiasm is very inspiring. Further I am grateful that you, Lotte Caarls, will be my paranymph. I got to know you as a Postdoc at Biosystematics and in the meantime, you became a senior scientist, well done! During our coffee and lunch walks and the shared moments in the greenhouse we always had a lot of fun and had a lot of discussions that varied from science related issues to dating gossips! Thanks for allowing me to often discuss minor or major issues. I hope to have picked up some of your calmness, perseverance and humour.

I do believe that the time before my PhD has been very important as preparation for my PhD and for my personal growth. During my BSc and MSc studies, and a little later as project researcher at Naturalis, I had the pleasure to study and work with loads of fellow students, colleagues and learn from many different researchers. It has been an incredible time and it was great to be able to learn so much from all aspects of biology. I am grateful to have worked with or get supervised by many incredible scientists. I learned a lot from my period at the Natural History Museum in London at the group of Alfried Vogler. I have also great memories to have worked together with Hans Huijbregts and Oscar Vorst during research projects and in the Naturalis beetle collections. The knowledge of Coleoptera (and beyond) of both of you is immense and I have learned a lot. Hans, our trips to Southeast Asia have been a lot of fun and a great experience I did not wanted to miss!

During my studies, I was also supervised various times by Menno Schilthuis. Menno, you are, and have been, a great inspiration to me. In the first place your lectures and books made me fascinated with evolutionary biology. During the research projects you have been a great mentor and I can truly say that you have played a big role for my enthusiasm for science and pursuing a life in academia.

Finally, I want to thank all the members of my PhD committee: Franziska Beran, Astrid Groot, Christopher Wheat and Bas Zwaan. I am honoured you took the time and effort to critically read my thesis. Thank you so much for helping and participating in this critical step in my scientific development and metamorphosis.



## About the author

Thijmen Breeschoten was born in Eindhoven, the Netherlands on the 18<sup>th</sup> of November 1990, but mainly grew up in Groot-Ammers and Voorhout. Thijmen has always been an avid naturalist and developed a main interest in beetles, the Coleoptera, and dung beetles in particular. He is an active member of the Netherlands Entomological Society (NEV) and participates in their yearly biodiversity inventories.



After receiving his high school diploma at Northgo College in Noordwijk, he studied Biology at Leiden University, the Netherlands. Thijmen completed his BSc research project under the supervision of Prof. Menno Schilthuizen at Naturalis Biodiversity Center in Leiden. He worked on the evolutionary patterns of asymmetric genitalia in Cyclocephalini beetles, which was published in *Contributions to Zoology* in 2013. Thijmen received the ‘Professor Kees Bakker Award 2014’ for best BSc thesis in evolutionary biology at Leiden University.

He continued his studies in Leiden and in 2015 he finished his MSc Biology with *cum laude* (majoring in Evolution, Biodiversity and Conservation). He completed two research projects. His first project focused on the evolution and species delimitation of a species complex of arboreal dung beetles of the genus *Onthophagus* in Southeast Asia using molecular sequence data, under supervision of Mr Hans Huijbregts and Prof. Menno Schilthuizen at Naturalis. For this project, fieldwork was conducted in Borneo and Peninsular Malaysia. This work is still ongoing, with additional research being done on the morphological differentiation.

Thijmen did his second research project jointly at Imperial College and The Natural History Museum in London, England, at the Molecular Systematics group of Prof. Alfried Vogler. He implemented a novel technique of pooling DNA extracts to sequence and assemble mitochondrial genomes. He then applied this ‘meta-mitogenomics approach’ to study the phylogenetics and biogeography of the cosmopolitan dung beetle genus *Onthophagus*. The results of the study were published in *Molecular Phylogenetics and Evolution* in 2016.

After his studies, Thijmen worked as project researcher at the Naturalis Biodiversity Center in Leiden, where he was involved in analysing the data generated during the barcoding project. Later, he was involved in the citizen science-based research on the evolution of shell colour in *Cepaea nemoralis* snails as published in *Communications Biology* in 2019.

In 2017, Thijmen started his PhD project at the Biosystematics Group of Wageningen University and Research, supervised by Prof. Eric Schranz and Dr Sabrina Simon. For this project he worked on the evolutionary genetic basis of polyphagous pest formations in Lepidoptera, with a focus on Noctuidae moths, by implementation of comparative transcriptomics and genomics approaches. The main findings of this research are presented in this thesis.

## List of publications

- Breeschoten, T.**, Schranz, M.E., Poelman, E.H., Simon, S. (In revision). Family dinner: transcriptional plasticity of five Noctuidae (Lepidoptera) feeding on three host plant species. *Molecular Ecology*.
- Breeschoten, T.**, van der Linden, C.F.H., Ros, V.I., Schranz, M.E., Simon, S. (In revision). Expanding the menu: exploring the link between polyphagy and gene family expansions across Lepidoptera. *Genome Biology and Evolution*.
- Simon,<sup>\*</sup> S., **Breeschoten,<sup>\*</sup> T.**, Jansen, H.J., Dirks, R.P., S., Schranz, M.E., Ros<sup>\*</sup> V.I. (In revision). Genome and transcriptome analysis of the beet armyworm *Spodoptera exigua* reveals targets for pest control. *G3: Genes, Genomes, Genetics*.
- Breeschoten, T.**, Ros V.I., Schranz, M.E., Simon, S. (2019). An influential meal: host plant dependent transcriptional variation in the beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae). *BMC Genomics*, 20, 1-15.
- Breeschoten, T.**, Doorenweerd, C., Tarasov, S., Vogler, A.P. (2019). Incorporating older literature into genomic studies: A response to Zunino & Halffter. *Molecular Phylogenetics and Evolution*, 133, 164-165.
- Kerstes, N.A.G., **Breeschoten, T.**, Kalkman, V.J., Schilthuizen, M. (2019). Snail shell colour evolution in urban heat islands detected via citizen science. *Communications Biology*, 2, 1-11.
- Rui-E, N., **Breeschoten, T.**, Timmermans, M.J.T.N., Nadein, K., Huai-Jun, X., Ming, B., Yuan, H., Xing-Ke, Y., Vogler, A.P. (2018). The phylogeny of Galerucinae (Coleoptera: Chrysomelidae) and the performance of mitochondrial genomes in phylogenetic inference compared to nuclear rRNA genes. *Cladistics*, 34, 113-130.
- Linard, B., Crampton-Platt, A., Moriniere, J., Timmermans, M.J., Andujar, C., Arribas, P., Miller, K.E., Lipecki, J., Favreau, E., Hunter, A., Gómez-Rodríguez, C., Barton, C, Nie, R., Gillett, C.P.D.T, **Breeschoten, T.**, Bocak, L., Vogler, A.P., et al. (2018). The contribution of mitochondrial metagenomics to large-scale data mining and phylogenetic analysis of Coleoptera. *Molecular Phylogenetics and Evolution*, 128, 1-11.
- Breeschoten, T.**, Doorenweerd, C., Tarasov, S., Vogler, A.P. (2016). Phylogenetics and biogeography of the dung beetle genus *Onthophagus* inferred from mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 105, 86-95.
- Breeschoten, T.**, Clark, D.R., Schilthuizen, M. (2013). Evolutionary patterns of asymmetric genitalia in the beetle tribe Cyclocephalini (Coleoptera: Scarabaeidae: Dynastinae). *Contributions to Zoology*, 82, 95-106.

<sup>\*</sup>Contributed equally to this study and shared first authorship.

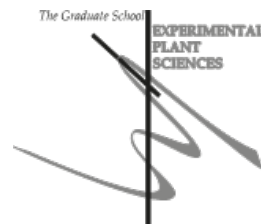
## Education Statement of the Graduate School Experimental Plant Sciences

**Issued to:** Thijmen Breeschoten

**Date:** 8 September 2021

**Group:** Biosystematics

**University:** Wageningen University & Research



1) Start-Up Phase	<u>date</u>	<u>cp</u>
► <b>First presentation of your project</b>		
“A comparative phylogenomic framework to study the evolution and molecular basis of ‘pestiness’ in pest insects”	12 Jul 2017	1,5
► <b>Writing or rewriting a project proposal</b>		
Project proposal: “What makes a pest a pest? A comparative transcriptomic approach to study ‘pestiness’ in insects”	Nov 2017	1,0
<i>Subtotal Start-Up Phase</i>		2,5

2) Scientific Exposure	<u>date</u>	<u>cp</u>
► <b>EPS PhD student days</b>		
EPS PhD student days - Get2Gether 2018, Soest, The Netherlands	15-16 Feb 2018	0,6
EPS PhD student days - Get2Gether 2019, Soest, The Netherlands	11-12 Feb 2019	0,6
► <b>EPS theme symposia</b>		
EPS theme 4 symposium “Genome Biology”, UvA	25 Sep 2018	0,3
EPS theme 2 symposium & Willie Commelin Scholten Day “Interactions between plants and biotic agents”, WUR	01 Feb 2019	0,3
► <b>Lunteren Days and other national platforms</b>		
Annual Meeting Experimental Plant Sciences, Lunteren, The Netherlands	10-11 Apr 2017	0,6
Annual Meeting Experimental Plant Sciences, Lunteren, The Netherlands	9-10 Apr 2018	0,6
2nd NLSEB, Ede, The Netherlands	16 Apr 2019	0,3
29th Entomologendag (NEV), Ede, The Netherlands	15 Dec 2017	0,3
30th Entomologendag (NEV), Ede, The Netherlands	14 Dec 2018	0,3
31th Entomologendag (NEV), Ede, The Netherlands	13 Dec 2019	0,3
32nd Entomologendag (NEV) - Online	18 Dec 2020	0,2
► <b>Seminars (series), workshops and symposia</b>		
<i>Symposium:</i> Insect-Microbe-Plant interactions (organised by Dr. Nina Fatouros & Dr. Ana Pineda), WUR	07 Apr 2017	0,3
<i>EPS Seminar:</i> Dr. Martin Kaltenpoth - ‘Outsourcing immunity: Microbial symbionts for pathogen defense in insects’, WUR	07 Apr 2017	
<i>WEES Seminar:</i> Prof. Michael Singer - ‘One butterfly species tumbles off an adaptive peak and enters a lethal trap in the course of six host shifts observed across half a century’, WUR	12 Apr 2017	0,1

<i>WEES Seminar</i> : Dr. Papanicolaou - 'Identification of a new sensory neuron membrane gene and why phylogenomics is important', WUR	22 Jun 2017	0,1
<i>EPS Seminar</i> : Prof.Dr. U. Wyss - 'Highlights of hidden Insect worlds', WUR	02 Oct 2017	0,1
<i>Workshop</i> : 12th workshop Plant-Insect Interactions (PII), Wageningen, The Netherlands	07 Nov 2017	0,3
<i>Workshop</i> : 13th workshop Plant-Insect Interactions (PII), Leiden, The Netherlands	09 Oct 2018	0,3
<i>Workshop</i> : 14th workshop Plant-Insect Interactions (PII), Utrecht, The Netherlands	14 Nov 2019	0,3
► <b>Seminar plus</b>		
<i>WEES workshop</i> : Dr. Papanicolaou 'The genome project life cycle: no longer an aim but a tool', WUR	22 Jun 2017	0,1
► <b>International symposia and congresses</b>		
Originally planned attendance at International Congress of Entomology (ICE2020/ICE2022)	Postponed to 2022	0,0
Originally planned attendance at International Symposium on Insect-Plant Interactions (SIP2020/SIP2021)	Postponed to 2021	0,0
► <b>Presentations</b>		
<i>Talk</i> : "A comparative phylogenomic framework to study the evolution and molecular basis of 'pestiness' in insects", at 12th workshop Plant-Insect Interactions	07 Nov 2017	1,0
<i>Poster</i> : "What makes a pest a pest? A comparative transcriptomic approach to study 'pestiness' in insects", at 29th Entomologendag (NEV)	15 Dec 2017	1,0
<i>Poster</i> : "Influential dinner: host plant dependent gene expression in <i>Spodoptera exigua</i> ", at Annual Meeting Experimental Plant Sciences	09 Apr 2018	1,0
<i>Poster</i> : "Influential meal: a story about <i>Spodoptera exigua</i> and its hosts", at 30th Entomologendag (NEV)	14 Dec 2018	1,0
<i>Presentation &amp; discussion</i> : with Director Plant Sciences Group about PhD project, WUR	09 Jan 2019	1,0
<i>Poster</i> : "Flexibility makes the generalist: evolution of herbivory in generalist moths (Lepidoptera: Noctuidae)", at NLSEB meeting	16 Apr 2019	1,0
<i>Talk</i> : "Food matters: gene expression variation in moths", at 31th Entomologendag (NEV)	13 Dec 2019	1,0
► <b>Excursions</b>		
Online company visit Bejo seeds	14 Dec 2020	0,2
<i>Subtotal Scientific Exposure</i>		13,2

### 3) In-Depth Studies

	<i>date</i>	<i>cp</i>
► <b>Advanced scientific courses &amp; workshops</b>		
2018 Workshop on Genomics, Cesky Krumlov, Czech Republic	07-20 Jan 2018	3,2
The Power of RNA-seq, WUR	11-13 Jun 2018	0,9
Software Carpentry Workshop on Python programming, UvA	14-15 Jan 2019	0,6
► <b>Journal club</b>		
Biosystematics Journal Club	2017-2021	3,0
<i>Subtotal In-Depth Studies</i>		7,7

<b>4) Personal Development</b>	<u>date</u>	<u>cp</u>
► General skill training courses		
Competence Assessment, WUR	19 Sep 2017	0,3
EPS Introduction Course, WUR	26 Sep 2017	0,3
Course Scientific Artwork - Vector graphics and images, WUR	02-03 Oct 2018	0,6
Course Adobe InDesign Essential Training, WUR	22-23 Nov 2018	0,6
Course Infographics and Iconography - Data visualization, WUR	04 Dec 2018	0,3
Brain-friendly Working and Writing, WUR	04 Jun 2019	0,3
WGS PhD Workshop Carousel, WUR	25 May 2018	0,3
WGS PhD Workshop Carousel, WUR	24 May 2019	0,3
Brain training, WUR	26 Jun 2019	0,3
Workshop Applying for Marie Skłodowska-Curie Fellowship: from proposal to project, WUR	10 Dec 2020	0,2
► <b>Organisation of meetings, PhD courses or outreach activities</b>		
Brainstorming + initiating python/genomics course, WUR	2018	0,1
Assisting in Genomics Workshop, Data Carpentry, WUR	05-06 Feb 2019	0,6
<i>Subtotal Personal Development</i>		4,2

<b>5) Teaching &amp; Supervision Duties</b>	<u>date</u>	<u>cp</u>
► <b>Courses</b>		
BIS-10306 Biodiversity of the Netherlands, WUR	2017	3,0
BIS-10306 Biodiversity of the Netherlands, WUR	2018	
BIS-10306 Biodiversity of the Netherlands, WUR	2019	
BIF-30806 Advanced Bioinformatics, WUR	2019	
► <b>Supervision of BSc/MSc students</b>		
Supervision BSc student J. School	Sep-Dec 2017	3,0
Supervision BSc student D. de Groot	Mar-Apr 2018	
Supervision MSc student for Capita Selecta C. v.d. Linden	Oct 2020 - Mar 2021	
<i>Subtotal Teaching &amp; Supervision Duties</i>		6,0

<b>TOTAL NUMBER OF CREDIT POINTS*</b>	<b>33,6</b>
---------------------------------------	-------------

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

\* A credit represents a normative study load of 28 hours of study.

The research presented in this thesis was performed at the Biosystematics Group of Wageningen University & Research (WUR) and was supported by Graduate School Experimental Plant Sciences (EPS) Strategic Funding awarded to S. Simon.

Cover design: Thijmen Breeschoten & Iliana Boshoven-Gkini | [AgileColor.com](https://AgileColor.com)

Layout design: Iliana Boshoven-Gkini | [AgileColor.com](https://AgileColor.com)

Printed by: GVO drukkers & vormgevers | [gvo.nl](https://gvo.nl)

