UDP-glycosyltransferases: Comparative genomic analysis in insects and capsaicin glucosylation in *Helicoverpa* spp.

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem
Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von Master of Science
Seung-Joon Ahn

geboren am 2.2.1976
in Ulsan, Korea

Gutachter

1. Prof. Dr. David G. Heckel (Max-Planck-Institut für chemische Ökologie, Jena, Deutschland)
2. Prof. Dr. Wolfgang W. Weisser (Friedrich-Schiller-Universität, Jena, Deutschland)
3. Prof. Philip Batterham (University of Melbourne, Victoria, Australia)

Datum der öffentlichen Verteidigung: 8. Dezember 2011

Table of Contents

1. INTRODUCTION

1.1. Metabolic detoxification in insects

1.1.1 Plant secondary metabolites and insect-plant interactions

Plant secondary metabolites play an important role in insect-plant interactions. Many of them are detrimental or toxic to herbivorous insects, therefore having defensive properties. Insects feeding on the plants containing such noxious compounds, however, have strategies to cope with the plant's defense. Therefore, the plant secondary compounds act as a driving force for specialization of herbivores on their host plants through the 'evolutionary arms race' (Ehrlich and Raven, 1964). These allelochemicals can have multiple effects on herbivores, influencing behavior as well as physiology. They can deter feeding by non-specialists or stimulate feeding by specialists which associate the chemical with an exploitable food source. Once they are consumed, allelochemicals may influence the postingestive utilization of nutrients through various physiological and biochemical mechanisms. After being absorbed, many allelochemicals exert deleterious effects through a variety of pharmacological modes of action (Slansky, 1992).

1.1.2. Detoxification of plant allelochemicals in insect

Insect herbivores confronted with a variety of noxious chemicals in their food plants have evolved various counter defense mechanisms to cope with their harmful effects. Among those are (1) to avoid continuous contact, (2) to excrete the unwanted compounds rapidly, (3) to modify them enzymatically, (4) to sequester them for further utilization, or (5) to develop target-site insensitivity (Després et al., 2007)Schoonhoven et al, 2005).

1.1.3. Enzymatic detoxification

Enzymatic detoxification of ingested plant allelochemicals is one of the important mechanisms by which insects can neutralize, degrade, or modify a variety of xenobiotics including plant allelochemicals (Brattsten, 1992). The most extensively studied detoxicative enzymes are the cytochrome P450s (P450s), the phase I metabolic enzymes, capable of converting the compounds into more polar and reactive compounds, produced by several different reactions including oxidation, carbon hydroxylations, *N*- and *O*-dealkylations, or epoxidations (Feyereisen, 2005). The metabolites can be further processed by phase II metabolic enzymes, where group

transferring reactions are usually occurring by, for example, the glutathione S-transferases (GSTs) or the UDP-glycosyltransferases (UGTs).

1.2. UDP-glycosyltransferase and insect defense

1.2.1. UDP-glycosyltransferase (UGT)

UDP-glycosyltransferases (UGTs) are a family of enzymes that catalyses the transfer of the glycosyl group from a nucleotide sugar, such as UDP-glucuronic acid, UDP-glucose, UDP-galactose, or UDP-xylose, to a variety of small hydrophobic molecules (aglycones), resulting in more hydrophilic compounds that are easily excreted (Mackenzie et al., 1997). UGTs are ubiquitous in all living organisms from viruses to bacteria, plants and animals. In plants, a variety of UGTs play an important role in the modification of secondary metabolites, thereby enhancing their solubility and stability, and determining their bioactivity (Bowles et al., 2005). In vertebrates, UGTs are regarded as a major member of the phase II drug metabolizing enzymes, conjugating a large number of xenobiotics as well as endobiotics, such as bilirubin and steroid hormones with UDP-glucuronic acid (Bock, 2003). In insects, the glycosylation of small lipophilic compounds has been considered as a minor enzymatic detoxification mechanisms for half a century (Ahmad et al., 1986; Brattsten, 1988; Després et al., 2007; Smith, 1962). So far, more than 800 UGT sequences named by the UGT Nomenclature Committee are present in animals, plants, yeasts, bacteria and viruses (http://www.flinders.edu.au/medicine/sites/clinical-pharmacology/ugt-homepage.cfm).

1.2.2. Insect UGTs

Insect UGTs are relatively less known compared to mammalian or plant UGTs. However, it was already observed more than half a century ago that glucosides of various exogenous and endogenous compounds are formed in insects (Smith, 1955). This glucoside conjugation in insects attracted a lot of attention as it was the only major detoxification mechanism which is qualitatively different in insects and mammals, in other words, glucose conjugation in insect versus glucuronic acid conjugation in mammals (Smith, 1962). Endogenous compounds especially involved in cuticle tanning had been investigated to some extent, as several precursors for tanning, for example protocatechuic acid and tyrosine, are stored as glucoside forms, and the conjugates serve in cuticle tanning or sclerotization processes during metamorphosis (Hopkins and Kramer, 1992; Kramer and Hopkins, 1987). Overall biochemical properties of insect UGT

enzymes were extensively studied by Ahmad (1994) mainly with *Manduca sexta* as a model insect. The *Drosophila* Genome Project provided the first genomic view of the insect UGT multigene family, although functional analysis was incomplete (Luque and O'Reilly, 2002). An additional genomic analysis of the insect UGTs was conducted from the silkworm genome, which provided a more interesting picture of the diverse gene family in insects (Huang et al., 2008).

1.2.3. Protein structure and cellular localization

In animals, UGTs are membrane-bound proteins located in the endoplasmic reticulum (ER) facing the lumen, whereas plant UGTs are cytosolic. The primary structure of the protein is divided into two main domains, the N-terminal aglycone substrate binding domain and the C-terminal UDP-glycoside binding domain. The N-terminal end of the animal UGTs has a signal peptide mediating the integration of the protein precursor into the ER compartment. The signal peptide is subsequently cleaved and then the protein is *N*-glycosylated. The mature protein is retained in the ER membrane by its hydrophobic transmembrane domain at the C-terminal end, followed by a short cytoplasmic 'tail' (Magdalou et al., 2010). Although UGTs have been investigated from many other perspectives like drug metabolism, few studies have examined the structural properties of UGTs. However, the recent crystallization of the C-terminal domain of human UGT2B7 provides detailed information on the UDP-glucuronic acid binding domain (Miley et al., 2007). Furthermore, x-ray crystal structures of a number of plant UGTs have also been determined, providing additional valuable information on their structure (Osmani et al., 2009).

1.2.4. Functional diversity in insects

Biochemical properties of insect UGTs have been widely investigated in many insects; the housefly *Musca domestica* (Morello and Repetto, 1979), the fruitfly *Drosophila melanogaster* (Real et al., 1991), the tobacco hornworm *Manduca sexta* (Ahmad and Hopkins, 1992), the silkworm *Bombyx mori* (Luque et al., 2002), and other insects (Ahmad and Hopkins, 1993b). These biochemical studies have shown that the insect UGT enzymes typically use UDP-glucose as a main sugar donor unlike vertebrate UGTs which conjugate UDP-glucuronic acid, but are probably bound to the endoplasmic reticulum in a similar manner as in vertebrates. Enzyme activities of the insect UGTs are detected in the fat body, midgut and other tissues (Ahmad and Hopkins, 1993b), and are directed towards a variety of plant allelochemicals (Ahmad and Hopkins, 1993a; Luque et al., 2002; Sasai et al., 2009). Interestingly, the enzymes were also detected in the antenna of *D. melanogaster*, but their function has not been characterized yet (Wang et al., 1999). In addition, many endogenous compounds, like ecdysteroid hormones (Svoboda and Weirich, 1995) and cuticle tanning precursors (Ahmad et al., 1996; Hopkins and

Kramer, 1992), are glycosylated by the UGT enzymes. Furthermore, in a lycaenid butterfly dietary flavonoids have been shown to be sequestered as a glucose conjugate to impart color to the wings (Wiesen et al., 1994) or in *B. mori* to be glycosylated to make green cocoon coloration with UV-shielding properties (Daimon et al., 2010). These findings suggest multiple roles of the insect UGT enzymes in detoxification, olfaction, endobiotic modulation, and sequestration.

1.2.5. UGT multigene family

Gene families are groups of genes descended from a common ancestor that retain similar sequences and often similar functions (Demuth and Hahn, 2009). Gene duplications are known to be a major source of gene family members, thereby producing evolutionary novelties and phenotypic variation. Duplicated genes can originate from whole genome-, segmental- or tandem duplications (Li, 1997).

In an era of genome sequences, many insect gene families are being discovered. At the time of this manuscript preparation, there are 33 insect species of which genome sequences have been either completed or at least assembled (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi). Multigene families in insects especially involved in detoxification and metabolism, like cytochrome P450s (P450s), carboxyl/cholinesterases (CCEs), glutathione transferases (GSTs), have been identified from *Drosophila melanogaster* (Tijet et al., 2001), *Aedes aegypti* (Strode et al., 2008), *Acyrthosiphon pisum* (Ramsey et al., 2010), *Nasonia vitripennis* (Oakeshott et al., 2010), *Apis mellifera* (Claudianos et al., 2006), and *Bombyx mori* (Tsubota and Shiotsuki, 2010; Yu et al., 2008). The molecular identities of the insect UGTs are, however, relatively unknown compared to the other detoxification gene families in insects or to the vertebrate UGT gene families. Although the UGT gene families in *D. melanogaster* (Luque and O'Reilly, 2002) and in *B. mori* (Huang et al., 2008) have been reported, they have not been compared in the context of the diversity of other insects.

The first genome-wide study on the insect UGT gene family revealed that the UGT gene family is composed of 33 gene sequences in *D. melanogaster* (Luque and O'Reilly, 2002). They all have the characteristic signature sequence in their C-terminal half which binds the sugar donor, which tends to be more highly conserved than the N-terminal half that is believed to be responsible for binding the aglycone. These are highly diverse, explaining the low conservation in the N-terminal region. More recently the second genome-wide study on UGTs of an insect was published; the domesticated silkworm *B. mori* was reported to possess 42 UGT genes (Huang et al., 2008). Phylogenetic analysis, genomic organization and expression profiles provided an overview for the silkworm UGTs to facilitate their functional studies in the future.

1.3. Capsaicin and insect interactions

1.3.1 Capsaicin, a pungent compound in the *Capsicum* fruit

Hot peppers (*Capsicum* spp.) produce capsaicinoids in the fruits, which are responsible for the burning taste experienced by mammals (Caterina et al., 1997). Capsaicin (8-methyl-N-vanillyl-6-nonenamide, Fig. 1) is the predominant form among more than ten capsaicinoids (Mazourek et al., 2009). The amount of capsaicin in the fruits is enormously variable in the wild, depending on the genetic, environmental and developmental conditions (Appendino, 2008; Aza-González et al., 2010; Tewksbury et al., 2006). The pungency is caused by activation of a transient receptor potential (TRP) cation channel, TRPV1, on sensory nerve endings (Caterina et al., 1997).

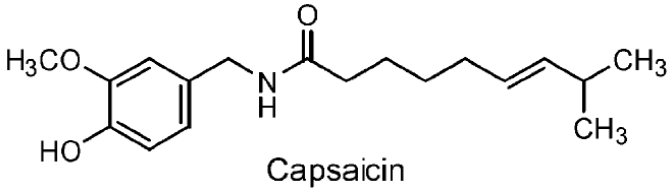


Figure 1. Chemical structure of capsaicin (*trans*-8-methyl-*N*-vanillyl-6-nonenamide)

1.3.2. Chemical ecology of capsaicin

Studies on the ecological role of capsaicin have been already reported with wild *Capsicum* spp. and their vertebrate frugivores, suggesting a seed dispersal role by birds, which are not sensitive to capsaicin compared to mammalian frugivores (Levey et al., 2006; Tewksbury and Nabhan, 2001). Interestingly, an avian TRPV1 ortholog from chicken sensory neurons is not sensitive to capsaicin, which provides a molecular basis for the ecological differences between mammals and birds (Jordt and Julius, 2002). Recently, variation in the pungency of wild chilies was suggested to be an adaptive response to a fungal pathogen of the *Capsicum* fruit (Tewksbury et al., 2008). The capsaicin production is directly linked to the variation in the damage caused by *Fusarium* fungus, which is the main cause of the seed mortality.

1.3.3. Effect of capsaicin in insects

There are several studies on the effects of capsaicin on insects. Capsaicin is known to deter oviposition in the onion fly *Delia antiqua* (Diptera: Anthomyiidae) (Cowles et al., 1989), to

inhibit the feeding of a ladybird beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae) (Hori et al., 2011), and to retard larval growth in the spiny bollworm *Earias insulana* (Lepidoptera: Noctuidae) (Weissenberg et al., 1986). *Capsicum* extracts have larvicidal activity to two species of mosquitoes, *Anopheles stephensi* and *Culex quinquefasciatus* (Madhumathy et al., 2007), and synergistic effects with insecticides on the green peach aphid, *Myzus persicae* (Edelson et al., 2002). While most of such studies have been performed in order to evaluate the usage of capsaicin or *Capsicum* extracts as alternative pesticides, only a few studies have dealt with capsaicin in the context of host plant adaptation by herbivorous insects (Larue and Welty, 2010).

1.3.4. *Helicoverpa assulta,* a host-plant specialist feeding on *Capsicum* fruit

The Oriental tobacco budworm, *Helicoverpa assulta* (Guenée) (Lepidoptera: Noctuidae), is one of the few insects that can successfully feed on hot pepper fruits. It is a specialist on the family Solanaceae, feeding also on *Lycopersicon*, *Nicotiana*, *Physalis*, and *Solanum* in Korea, Japan, China, Australia, and Africa (Cho et al., 2008; Matthews, 1999; Mitter et al., 1993; Yang et al., 2004). The relationship between *H. assulta* and hot pepper has been described by studies on egg distribution (Han et al., 1994), larval feeding preference (Choi and Boo, 1989) and feeding damage (Baek et al., 2009). Effects of capsaicin on *H. assulta* larvae, however, have not yet been described. I hypothesized that capsaicin is less detrimental to this specialist herbivore than to other insects.

1.4. Purpose of this study

This study was intended to figure out how the host-specialist, *H. assulta*, was able to adapt to hot pepper and specialize on the fruit as a larval food. My initial work investigated the effect of capsaicin on growth and survivorship of *H. assulta*, and compared it with other noctuid moths. The results of the initial investigation led the study to focus on the metabolic detoxification mechanism of dietary capsaicin in the insects, revealing production of a capsaicin glycoside in *Helicoverpa* species. Since UDP-glycosyltransferase (UGT) was implicated in the capsaicin glucosylation, the study further concentrated on the genomic analysis of the multigene family in

Lepidoptera, with three genome sequences available for analysis. For these investigations, this study includes:

(1) Investigation of the effects of capsaicin on different insect species including *H. assulta*: The effects were specified by comparing various developmental measures like larval growth, development time, survivals, and pupal weights among six different noctuid species. Nutritional indices were compared to characterize the effects on the food uptake in more detail. Capsaicin injection was conducted in order to distinguish direct effects in the internal cavity from those in the gut on the larval development. The acute toxicity of capsaicin was also compared by injection of high doses.

(2) Identification of a capsaicin metabolite, capsaicin glucoside, in the feces of three *Helicoverpa* spp.: Chemical identification of the metabolite was performed by various analytical methods, LC-MS, NMR, and enzymatic hydrolysis. Quantitative measurements of the metabolite from three different species gave insights into how differently generalists and the specialist metabolize capsaicin. Biochemical characterization of the crude enzymes involved in the glucose conjugation suggested that UGTs might be responsible for the reaction.

(3) Genomic analysis of the UGT multigene family from *Helicoverpa armigera* and other insects. The identification of the gene family members serves as an initial step toward investigations of individual enzymes, their substrate specificities, and functional characterization. In addition, phylogenetic analysis of the gene family sheds light on evolutionary perspectives of the detoxicative and metabolic enzyme family in insects.

2. OVERVIEW OF THE MANUSCRIPTS

Manuscript I

A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species

Seung-Joon Ahn, Francisco R. Badenes-Pérez, David G. Heckel

in press, *Journal of Insect Physiology* (2011) doi:10.1016/j.jinsphys.2011.05.015

Description:

In this manuscript, we examine the effects of capsaicin on six different insects including a host-specialist, *Helicoverpa assulta* in order to test whether capsaicin is detrimental to insects or not. By feeding and injection experiments, we discovered that capsaicin has a growth-inhibiting effect on the insects except *H. assulta*, which showed higher tolerance to capsaicin than other insects.

Authors' contributions:

S.-J. Ahn planned and performed the experiments, analyzed the data, and wrote the manuscript. F. R. Badenes-Pérez commented on and edited the manuscript. D. G. Heckel supervised the project and edited the manuscript.

Manuscript II

Metabolic detoxification of capsaicin by UDP-glycosyltransferase in three *Helicoverpa* species

Seung-Joon Ahn, Francisco R. Badenes-Pérez, Michael Reichelt, Aleš Svatoš, Bernd Schneider, Jonathan Gershenzon and David G. Heckel

submitted to *Archives of Insect Biochemistry and Physiology* (July, 1, 2011)

Description:

Manuscript II reports the chemical identification of capsaicin glucoside in the larval feces from *Helicoverpa armigera*, *H. assulta* and *H. zea*. The three species excreted capsaicin glucoside at different rates. The enzyme responsible for the glucoside conjugation, UDP-glycosyltransferase, was biochemically investigated, suggesting different enzyme activities among different species as well as among different larval tissues.

Authors' contributions:

S.-J. Ahn and F. R. Badenes-Pérez designed and performed the experiments. A. Svatoš contributed to the initial identification of capsaicin glucoside, M. Reichelt and J. Gershenzon performed LC-MS and provided technical comments for the experiments, and B. Schneider conducted NMR analysis. S.-J. Ahn wrote the manuscript. D. G. Heckel supervised the project and edited the manuscript.

Manuscript III

Comparative analysis of the UDP-glycosyltransferase multigene family in insects

Seung-Joon Ahn, Heiko Vogel and David G. Heckel

submitted to *Insect Biochemistry and Molecular Biology* (June 30, 2011)

Description:

This manuscript describes the identification of transcripts of the UDP-glycosyltransferase multigene family from *Helicoverpa armigera*. Phylogenetic analysis of the *H. armigera* UGT sequences was performed in comparison with *Bombyx mori* UGT sequences. In addition, transcript diversity of *H. armigera* UGTs and their structure prediction were also explained.

Authors' contributions:

S.-J. Ahn and D. G. Heckel planned the project, S.-J. Ahn compiled the sequences from private EST databases of *H. armigera*, performed data analyses and wrote the manuscript. H. Vogel constructed the cDNA library and contributed to the phylogenetic analysis.

Manuscript IV

Genomic analysis of the UDP-glycosyltransferase multigene family in *Helicoverpa armigera*

Seung-Joon Ahn, Heiko Vogel and David G. Heckel

unpublished results *(in preparation)*

Description:

Due to the constraints to publication of *H. armigera* genome sequences by the Helicoverpa Genome Consortium, this manuscript is prepared with unpublished results with permission of the Consortium. This manuscript describes the first and the most extensive genome-wide analysis of the insect UDP-glycosyltransferase multigene family; it deals with about 320 UGT sequences from nine insect species belonging to five different Orders. Phylogenetic analysis, chromosomal distribution, and exon-intron structures were compared with *Bombyx mori* UGT sequences in particular, suggesting a variety of phylogenetic patterns; from conserved families to large taxon-specific families in Lepidoptera, Diptera, an aphid, and a beetle.

Authors' contributions:

S.-J. Ahn and D. G. Heckel planned the project, S.-J. Ahn compiled the sequences from public and private databases, performed data analyses and wrote the manuscript. H. Vogel constructed the cDNA library and contributed to the phylogenetic analysis. D. G. Heckel edited the manuscript.

3. MANUSCRIPTS

3.1. Manuscript I

A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species

Seung-Joon Ahn[a,*], Francisco R. Badenes-Pérez[a,b], David G. Heckel[a]

Journal of Insect Physiology (2011) in press

[a] Department of Entomology, Max-Planck Institute for Chemical Ecology, 07745 Jena, Germany
[b] Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

*Corresponding author.
Seung-Joon Ahn, Department of Entomology, Max-Planck Institute for Chemical Ecology, 07745 Jena, Germany, Tel.: +49 3641 57 1560; fax: +49 3641 57 1502. *E-mail address*: sahn@ice.mpg.de (S.-J. Ahn)

ARTICLE IN PRESS

Journal of Insect Physiology xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Journal of Insect Physiology

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

A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species

Seung-Joon Ahn [a,*], Francisco R. Badenes-Pérez [a,b], David G. Heckel [a]

[a] Department of Entomology, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany
[b] Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Cientificas, 28006 Madrid, Spain

ARTICLE INFO

Article history:
Received 1 March 2011
Received in revised form 23 May 2011
Accepted 24 May 2011
Available online xxxx

Keywords:
Helicoverpa assulta
Capsaicin
Toxicity
Nutritional index
Injection
Noctuidae

ABSTRACT

Plant secondary compounds not only play an important role in plant defense, but have been a driving force for host adaptation by herbivores. Capsaicin (8-methyl-*N*-vanillyl-6-nonenamide), an alkaloid found in the fruit of *Capsicum* spp. (Solanaceae), is responsible for the pungency of hot pepper fruits and is unique to the genus. The oriental tobacco budworm, *Helicoverpa assulta* (Lepidoptera: Noctuidae), is a specialist herbivore feeding on solanaceous plants including *Capsicum annuum*, and is one of a very few insect herbivores worldwide capable of feeding on hot pepper fruits. To determine whether this is due in part to an increased physiological tolerance of capsaicin, we compared *H. assulta* with another specialist on Solanaceae, *Heliothis subflexa*, and four generalist species, *Spodoptera frugiperda*, *Heliothis virescens*, *Helicoverpa armigera*, and *Helicoverpa zea*, all belonging to the family Noctuidae. When larvae were fed capsaicin-spiked artificial diet for the entire larval period, larval mortality increased in *H. subflexa* and *H. zea* but decreased in *H. assulta*. Larval growth decreased on the capsaicin-spiked diet in four of the species, was unaffected in *H. armigera* and increased in *H. assulta*. Food consumption and utilization experiments showed that capsaicin decreased relative consumption rate (RCR), relative growth rate (RGR) and approximate digestibility (AD) in *H. zea*, and increased AD and the efficiency of conversion of ingested food (ECI) in *H. armigera*; whereas it did not significantly change any of these nutritional indices in *H. assulta*. The acute toxicity of capsaicin measured by injection into early fifth instar larvae was less in *H. assulta* than in *H. armigera* and *H. zea*. Injection of high concentrations produced abdominal paralysis and self-cannibalism. Injection of sub-lethal doses of capsaicin resulted in reduced pupal weights in *H. armigera* and *H. zea*, but not in *H. assulta*. The results indicate that *H. assulta* is more tolerant to capsaicin than the other insects tested, suggesting that this has facilitated expansion of its host range within Solanaceae to *Capsicum* after introduction of the latter to the Old World about 500 years ago. The increased larval survival and growth due to chronic dietary exposure to capsaicin suggests further adaptation of *H. assulta* to that compound, the mechanisms of which remain to be investigated.

1. Introduction

Plant secondary metabolites play an important role in insect-plant interactions. Compounds present in a limited range of plants perform a role in plant defense against herbivores, and in so doing also act as a driving force for specialization of herbivores on their host plants through the 'evolutionary arms race' (Ehrlich and Raven, 1964). These allelochemicals can have multiple effects on herbivores, influencing behavior as well as physiology. They can deter feeding by non-specialists or stimulate feeding by specialists which associate the chemical with an exploitable food source. Once they are consumed, allelochemicals may influence the postingestive utilization of nutrients through various physiological and biochemical mechanisms. After being absorbed, many allelochemicals exert deleterious effects through a variety of pharmacological modes of action (Slansky, 1992).

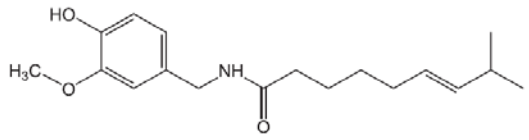
Hot peppers (*Capsicum* spp.) produce capsaicinoids in the fruits, which are responsible for the burning taste experienced by mammals (Caterina et al., 1997). Capsaicin (8-methyl-*N*-vanillyl-6-nonenamide, Fig. 1) is the predominant form among more than ten capsaicinoids (Mazourek et al., 2009). The amount of capsaicin in the fruits is enormously variable in the wild, depending on the genetic, environmental and developmental conditions (Appendino, 2008; Aza-González et al., 2011; Tewksbury et al., 2006). *Capsicum annuum* cultivars where *Helicoverpa assulta* is usually found in the field particularly in Korea contains 1.7–3.5 mg/g (dry weight) of capsaicinoids, occasionally up to 5 mg/g (Jung et al., 2006).

Because of its obvious pungent and feeding deterrent activity to mammals including humans, and because of its pharmaceutical potential, many studies on the effects of capsaicin have been

* Corresponding author. Tel.: +49 3641 57 1560; fax: +49 3641 57 1502.
E-mail address: sahn@ice.mpg.de (S.-J. Ahn).

0022-1910/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jinsphys.2011.05.015

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015



Capsaicin

Fig. 1. Chemical structure of capsaicin (8-methyl-*N*-vanillyl-6-nonenamide).

conducted with mammals (Surh and Kundu, 2011). Studies on the ecological role of capsaicin have been already reported with wild *Capsicum* spp. and their vertebrate frugivores, suggesting a seed dispersal role by birds, which are not sensitive to capsaicin compared to mammalian frugivores (Levey et al., 2006; Tewksbury and Nabhan, 2001). Recently, variation in the pungency of wild chilies was suggested to be an adaptive response to a fungal pathogen of the *Capsicum* fruit (Tewksbury et al., 2008).

There are only a few studies on the effects of capsaicin on insects. Capsaicin is known to deter oviposition in the onion fly *Delia antiqua* (Diptera: Anthomyiidae) (Cowles et al., 1989), to inhibit the feeding of a ladybird beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae) (Hori et al., 2011), and to retard larval growth in the spiny bollworm *Earias insulana* (Lepidoptera: Noctuidae) (Weissenberg et al., 1986). *Capsicum* extracts have larvicidal activity to two species of mosquitoes, *Anopheles stephensi* and *Culex quinquefasciatus* (Madhumathy et al., 2007), and synergistic effects with insecticides on the green peach aphid, *Myzus persicae* (Edelson et al., 2002). While most of such studies have been performed in order to evaluate the usage of capsaicin or *Capsicum* extracts as alternative pesticides, only a few studies have dealt with capsaicin in the context of host plant adaptation by herbivorous insects (Larue and Welty, 2010).

The Oriental tobacco budworm, *H. assulta* (Guenée) (Lepidoptera: Noctuidae), is one of the few insects that can successfully feed on hot pepper fruits. It is a specialist on the family Solanaceae, feeding also on *Lycopersicon*, *Nicotiana*, *Physalis*, and *Solanum* in Korea, Japan, China, Australia, and Africa (Cho et al., 2008; Matthews, 1999; Mitter et al., 1993; Yang et al., 2004). The relationship between *H. assulta* and hot pepper has been described by studies on egg distribution (Han et al., 1994), larval feeding preference (Choi and Boo, 1989) and feeding damage (Baek et al., 2009). Effects of capsaicin on *H. assulta* larvae, however, have not yet been described. We hypothesize that capsaicin is less detrimental to this specialist herbivore than to other insects. In order to test this hypothesis, the effects of capsaicin on *H. assulta* were compared with five other noctuids. Species tested include four generalists, *Helicoverpa armigera* (Hübner), *Helicoverpa zea* (Boddie), *Heliothis virescens* (F.), and *Spodoptera frugiperda* (Smith), and one specialist, *Heliothis subflexa* (Guenée). *H. armigera* feeds on more than 100 host plants, including cotton, tomato, tobacco, pigeon pea, chickpea, maize and sunflower, belonging to more than 40 plant families; *H. zea* feeds also on more than 100 plant species, mainly on corn, cotton and tomato; *H. virescens* also feeds about 40 different plants, including tobacco, cotton and soybean, belonging to 15 different families (Fitt, 1989); and *S. frugiperda* has over 180 host plants belonging to 42 different plant families, mainly grasses (Casmuz et al., 2010). *H. subflexa* feeds primarily on fruits of *Physalis* spp. (Solanaceae) (Bateman, 2006). We compared larval development and survival of all six species, continuously consuming capsaicin incorporated into the same artificial diet. In order to analyze the basis for *H. assulta*'s higher tolerance of dietary capsaicin, food consumption and utilization was measured and acute toxicity assessed by injection assays. The results provide preliminary evidence that adaptation to capsaicin has played a role in the expansion of the host range of the solanaceous specialist *H. assulta* to include *Capsicum*.

2. Materials and methods

2.1. Insects

H. armigera and *H. assulta* were collected from Queensland (Australia) and Suwon (Korea), respectively. *S. frugiperda* from Florida (USA) were provided by Dr. Robert Meagher, USDA-ARS Gainesville, and the other three species from North Carolina (USA) were collected from North Carolina (USA) and were provided by Dr. Fred Gould, North Carolina State University. Each species was maintained under laboratory conditions (26 °C, 55% relative humidity and 16:8 h L:D) with artificial diets (see below) in Jena, Germany. When adults emerged, single-pair matings were set up in paper cups (473 ml, SOLO, USA) and provided with 10% honey solution, having a mesh cloth on top of the mating cup for collecting eggs. The mesh cloths with fertile eggs were collected daily and placed in Petri-dishes. Neonates from at least three different mating families were used for the feeding experiments with the six different insect species. Freshly molted fifth instar larvae from three *Helicoverpa* species, *H. armigera*, *H. zea* and *H. assulta*, were used for the nutritional index measurements and the injection experiments (see below).

2.2. Artificial diet

Pinto bean based artificial diet was prepared with the following ingredients in a 2-liter batch: 125 g pinto bean powder (Huber-Mühle, Hohberg, Germany); 100 g wheat germ, 50 g soy protein, 5 g methyl paraben, 35 g agar, and 50 g casein, 62.5 g Torula yeast (Bio-Serv, Frenchtown, NJ, USA); 6 g ascorbic acid, 0.25 g tetracycline (Roth, Karlsruhe, Germany); 3 g sorbic acid, 10 g Vanderzant vitamin mixture (Sigma-Aldrich, MO, USA); and 1850 ml distilled water. This basal diet was chosen because it produced good growth and survivorship for most of the species, and was used for rearing the caterpillars until early fifth instar for the nutritional index measurements and the injection experiments. The capsaicin-spiked diet for the feeding experiments was prepared by supplementing a range of different concentrations of capsaicin (Fig. 1, ⩾98%, Fluka, USA) solution dissolved in ethanol into the diet right before it solidified (ca. 50 °C) and mixing for an additional 1 min. The final concentration of ethanol in total diet volume was 0.5% (v/v). For the control diet, only the corresponding amount of ethanol was added. All the diet treatments were prepared from a single batch in order to make the basal diet as even as possible.

2.3. Feeding assays

2.3.1. Larval development from neonate to pupa

Newly hatched neonates from at least three single-paired families of each species were provided with the artificial diet spiked with 0, 20 and 200 mg/l (w/v) of capsaicin prepared as above. This range of capsaicin concentrations corresponds to *Capsicum* cultivars where *H. assulta* is usually found in the field (Jung et al., 2006). A single neonate larva was placed in a single plastic cup (28 ml volume, SOLO, USA) with a cube of the test diets (ca. 1.5 g fresh weight). The cups were closed with plastic snap-top lids, which were maintained intact without any holes for the first 3 days in order to prevent larvae from escaping, and then were perforated with a small gimlet (ca. 1 mm diameter) to provide sufficient air for respiration during the remaining period of development. A fresh diet cube was given six days after the experiments had started. Larval fresh weights were individually measured at

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

Table 1
Nutritional indices calculated for three species. *I*: dry weight of food eaten (mg), *B*: dry weight gain of larva during the feeding period (mg), B_m: mean dry weight of the initial and final larva during the feeding period (mg), *T*: duration of feeding period (days), *F*: dry weight of feces produced (mg) (Waldbauer, 1968).

Abbrev.	Nutritional index	Formula	Explanation
RCR	Relative consumption rate	$I/(B_m \times T)$	Rate of food consumed per mean body weight per day
RGR	Relative growth rate	$B/(B_m \times T)$	Rate of body weight increase per mean body weight per day (= RCR × ECI)
AD	Approximate digestibility	$(I - F)/I$	Assimilated portion out of food ingested (%)
ECI	Efficiency of conversion of ingested food	B/I	Gross growth efficiency (%) (= RGR/RCR)
ECD	Efficiency of conversion of digested food	$B/(I - F)$	Net growth efficiency (%)

days 3, 6, 9, 12, and 15 by using a fine scale with a precision of 0.1 mg (Mettler-Toledo, USA). Larval survival was also checked every 3 days by observing if larvae moved when touched with a forceps. Dates of pupation and adult emergence were recorded by daily observation and fresh pupae were weighed a day after pupation due to the fragility of the fresh pupa. All measurements were stopped 30 days after the experiment had started.

2.3.2. Food consumption and utilization of fifth instar larvae

Newly molted fifth instar larvae that had been reared on the normal artificial diet were randomly chosen and starved for 24 h before the experimental diets were provided. The diets spiked with 0, 20 or 200 mg/l (w/v) capsaicin (see above for the diet preparation) were cut into a cube shape (ca. 1000 mg-fresh weight). After 48 h of feeding, larva, feces and diet remaining in the cup were separated and dried in an oven (60 °C) for 2 days until completely dried. The initial dry weights of diet and larvae were estimated by linear regression (R^2 = 0.9995 for diet, R^2 = 0.9508–0.9945 for three different larval species) of weights of fresh diet pieces (N = 15) and a cohort of larvae (N = 10), before and after drying in a drying oven. The nutritional index values were calculated based on dry weights as described in Table 1. The weight gain of larvae (B) was calculated by subtracting the initial dry weight estimate of a larva from the final dry weight. The final dry weight was directly obtained by freezing the larvae at −20 °C for 30 min and then drying them in the oven for 2 days. The dry weight of food eaten (I) by each larva was calculated by subtracting the weight of remaining food from the weight of food given initially. The dry weight of feces (F) remaining in the cup was measured by direct drying in the oven.

2.4. Injection assays

2.4.1. Acute toxicity at high doses

Due to the low water solubility of capsaicin, a range of capsaicin solutions was prepared by dissolving capsaicin in dimethyl sulfoxide (DMSO) (Sigma, MO, USA) and by serially diluting it to prepare solutions containing from 0.1 to 1000 μg capsaicin used in a 2 μl-injection dose. Each dose of capsaicin solution was injected through a posterior abdominal proleg of newly molted fifth instar larvae (N = 10) from three *Helicoverpa* spp. using the Microliter™ syringe (10 μl vol, GASTIGHT No. 1701, Hamilton, USA) attached to a fine needle (26 gauge, No. 7758-04, Hamilton, USA). An equal amount of DMSO was injected as a control. Injected larvae were allowed to feed on normal artificial diet. From the pilot experiments, we came to know that the capsaicin injection led larvae to a typical symptom of abdominal paralysis, with some recovering afterward and others failing to survive. Since it was not appropriate to use the mortality data due to the recovery from the paralysis and death by 'self-eating' behavior (see Section 4), we counted the number of paralyzed larvae two days after the injection and calculated the median effective dose (ED_{50}), instead of the median lethal dose (LD_{50}). Paralysis was scored as a failure to respond to pinching of the abdomen with forceps.

2.4.2. Long-term effects at sub-lethal doses

Sub-lethal amounts of capsaicin were additionally injected in order to investigate its long-term growth inhibiting effect, bypassing any possible detoxification by the digestive system. Capsaicin stock solution was prepared in ethanol first, and then diluted ten-fold into PIPES saline solution (5 mM PIPES, 21 mM KCl, 12 mM NaCl, 3 mM $CaCl_2$, 18 mM $MgCl_2$, 170 mM glucose, pH 6.6) to make solutions of 30 and 300 μg/ml of capsaicin. From these solutions of capsaicin, 1 μl (30 and 300 ng capsaicin equivalent, respectively) was injected through a posterior abdominal proleg of newly molted fifth instar larvae from *H. assulta* (N = 10), *H. armigera* (N = 10) and *H. zea* (N = 13), as explained above (Section 2.4.1). The same amount of saline solution mixed with ethanol (10% ethanol in final) was injected as a control. Injected larvae were allowed to feed on the basal artificial diet under standard rearing conditions until they pupated. The weight of the pupa was measured one day after pupation.

2.5. Statistical analysis

Larval performance (larval weight, pupal weight, and development time) and nutritional indices (RGR, RCR, AD, ECI, and ECD) were analyzed by one-way analysis of variance (ANOVA) in order to determine the influence of capsaicin in the same species of insect. Post-hoc tests were used to distinguish the differences among capsaicin concentrations. Larval survivorship was analyzed by χ^2-test. The ED_{50} of capsaicin was calculated using the paralyzing effect as a parameter by Probit analysis (Finney, 1971) using SPSS Statistics software.

3. Results

3.1. Capsaicin feeding effects on larval development

3.1.1. Larval survivorship and development

In order to compare the effects of chronic dietary exposure on six species of noctuids, larvae were reared on the same artificial diet containing 0, 20, or 200 mg capsaicin per liter from hatching to pupation. This pinto-bean-based artificial diet was not equally suited to all six species; mortality on diet lacking capsaicin was 31%, 50%, and 55%, over 30 days for *H. subflexa*, *H. zea* and *H. assulta* respectively, but less than 10% for the other three species (Fig. 2). Added capsaicin produced different effects depending on the insect species (Fig. 2). In *S. frugiperda* (Fig. 2A) only 3 and 2 larvae died during the feeding test with 20 and 200 mg/l capsaicin diet, respectively, whereas there was no mortality for the control diet. Larval growth and pupation time were delayed by both capsaicin concentrations. Unexpectedly, the lower concentration of capsaicin reduced the larval growth of *S. frugiperda* more than the higher concentration. In addition, pupal development time (days from

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

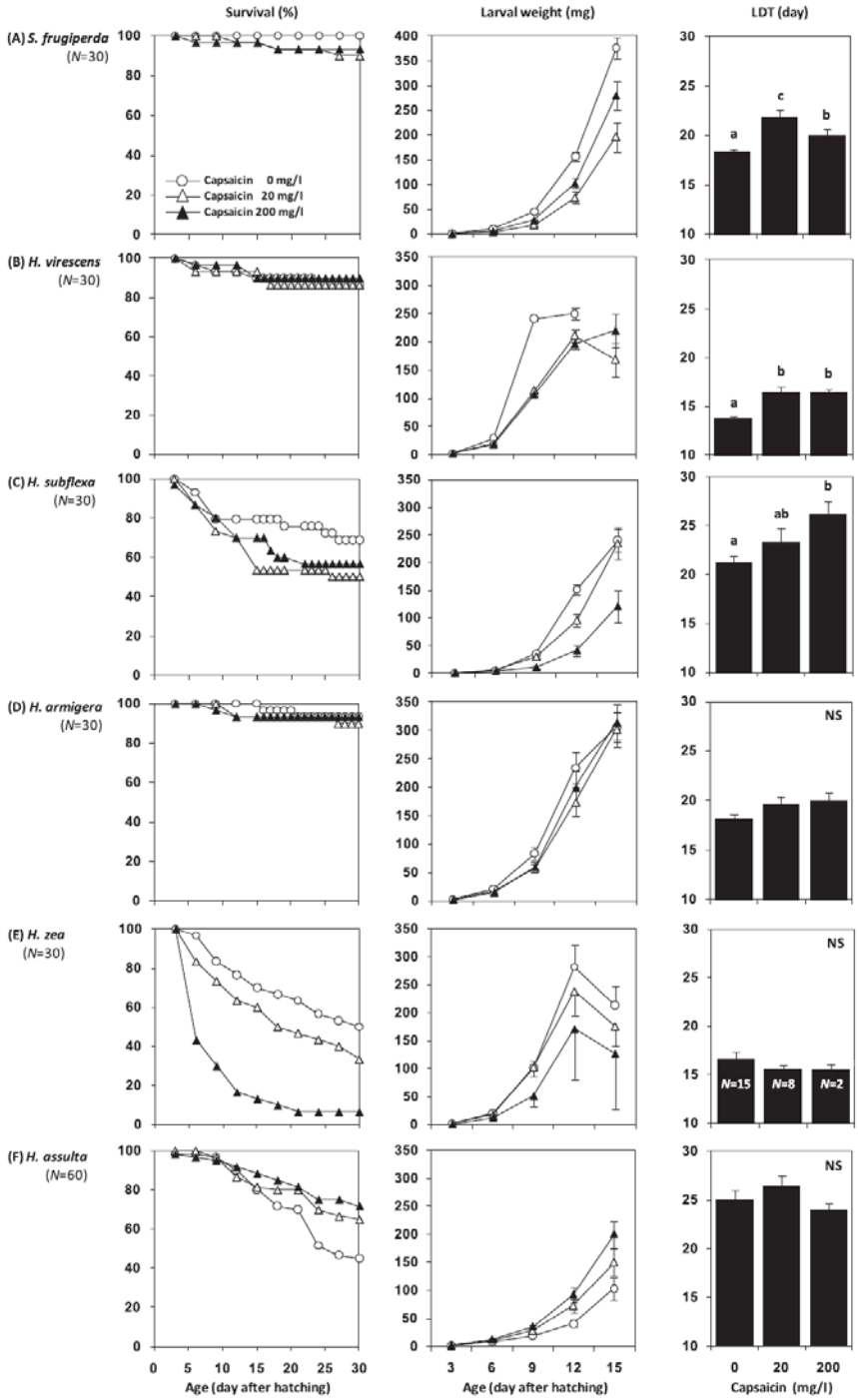


Fig. 2. The effects of capsaicin feeding on larval development of six different noctuid species. Capsaicin (20 or 200 mg/l) was mixed into artificial diet and provided to neonates individually. Larval survival and weights were recorded every 3 days. Pupation date was recorded on a daily basis. Larval development time (LDT) is the number of days from hatching to pupation. All bars in larval weight and LDT graphs represent the mean (±SE) of individuals. Different letters in LDT graphs indicate statistically significant differences ($P < 0.05$), and NS refers to not significant.

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

pupation to adult emergence) was significantly delayed in *S. frugiperda* at the higher capsaicin concentrations (about 1 day delayed, $P = 0.021$). For *H. virescens* (Fig. 2B), although survivorship was not different among treatments, larval growth was negatively affected by both capsaicin treatments. Retarded growth of *H. virescens* at Day 12 in the control and at Day 15 in both treatments caused body weight decrease at the pre-pupa stage. Time to pupation was longer in both capsaicin treatments, and time to adult from pupa was also longer at both capsaicin treatments (about 1 day delayed, $P < 0.01$). In the other specialist *H. subflexa* (Fig. 2C), both concentrations of capsaicin doubled the larval mortality (χ^2-test, $P < 0.01$ after Day 15) and higher amounts of capsaicin retarded larval growth resulting in the delayed pupation. Pupal development time, however, was not delayed by capsaicin (data not shown). *H. armigera* (Fig. 2D) was not significantly affected by capsaicin in survivorship, growth, or development time, however pupal weight decreased at 200 mg/l capsaicin diet (7.1% reduction compared to control pupal weight, $P = 0.028$). *H. zea* (Fig. 2E) suffered the highest mortality on capsaicin-spiked diet compared to the other insects; only two larvae successfully pupated at 200 mg/l capsaicin diet at the end, thus the larval growth graph and LDT bar graph apply only to the few survivors of this species. The control diet itself seems not to be optimal for *H. zea*, but the effect of capsaicin on reducing survivorship (6.7% at Day 30) was obvious compared to the control (50%). *H. assulta* (Fig. 2F) showed a markedly different response to dietary capsaicin than the other insects tested. Added capsaicin at either concentration nearly doubled larval survivorship at 30 days. Larval growth on control diet was lowest of the six species, but growth increased on diet with 20 mg/l and even more on diet with 200 mg/l capsaicin. On capsaicin diets, the weight of *H. assulta* larvae at Day 3 had already started surpassing that on the control diet ($F_{2, 175} = 23.9$, $P < 0.01$). Larval development time, however, was not different among treatments.

3.1.2. Nutritional indices of fifth-instar larvae

Growth and energy conversion indices were estimated for fifth-instar larvae that fed on capsaicin-containing diet for 48 h. None of the five nutritional indices were significantly different among the capsaicin treatments in *H. assulta* (Table 2). They were not different in *H. armigera* either, except AD and ECI, which increased from 43.0% to 47.0% and from 31.3% to 33.6%, respectively, in the presence of 30 mg/l capsaicin compared to the control, but they were not different at the higher concentration (300 mg/l). In *H. zea*, RCR, RGR and AD were significantly decreased at the 30 mg/l capsaicin, but ECI and ECD were not statistically different although the net values were smaller than the control. At 300 mg/l, all indices except ECD were smaller than the control; the differences were greater than for the other two species at the high concentration but not statistically significant, likely because of the lower sample size for *H. zea*.

A

Insect	ED_{50} (95% CL*) (μg/larva)
H. assulta	322 (178 — 641)
H. armigera	120 (51 — 358)
H. zea	49 (16 — 92)

B

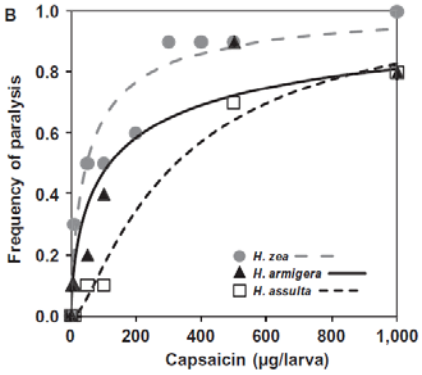


Fig. 3. Probit analysis of median effective dose (ED_{50}) of capsaicin against the larval paralysis upon injection. Paralyzed or dead larvae out of ten were counted two days after injection for each dose of capsaicin. (A) ED_{50} (95% CL) of capsaicin for paralyzing fifth instar larvae of *H. assulta*, *H. armigera*, and *H. zea* ($N = 10$) injected with a range of capsaicin concentrations (dissolved in DMSO). (B) Probit analysis (Finney, 1971) was used to estimate ED_{50}. *95% confidence limits.

3.2. Capsaicin injection effects

3.2.1. Acute toxicity at high doses

H. assulta was more tolerant to capsaicin injection than *H. armigera* or *H. zea*. The median effective dose (ED_{50}) values of capsaicin to *H. assulta* were 2.7 and 6.5 times higher than those of *H. armigera* and *H. zea*, respectively (Fig. 3). *H. zea* larvae were the most sensitive to capsaicin ($ED_{50} = 49$ μg/larva) among the three species. The larvae subjected to injection of the effective dose of capsaicin showed a series of symptoms. Some larva trembled convulsively a few minutes after the injection and then became partially paralyzed, unable to move the abdomen (Fig. 4). In many cases larvae were still able to move the head and thorax to feed. Some paralyzed larvae succeeded in recovering within a couple of days, but others failed to survive, sometimes eating their own paralyzed abdomen, leading to bleeding and death (Fig. 4).

3.2.2. Long-term effects at sub-lethal doses

Injection of the lower doses of capsaicin (30 and 300 ng/larva) was not lethal and produced no obvious behavioral effects. Pupal

Table 2
Nutritional indices (mean ± SE) of the fifth instar larvae of *H. assulta*, *H. armigera*, and *H. zea* feeding on artificial diet containing capsaicin for 2 days. Means among capsaicin treatments were not significantly different within parameters in the same species, unless followed by different letters. ANOVA followed by Duncan's test ($P < 0.05$) was performed for multiple comparisons (see Table 1 for the abbreviations of nutritional indices).

Insect	Capsaicin (mg/l)	*N*	RCR (mg/mg/day)	RGR (mg/mg/day)	AD (%)	ECI (%)	ECD (%)
H. assulta	0	11	1.84 ± 0.053^{a}	0.52 ± 0.012^{a}	49.7 ± 0.74^{a}	28.1 ± 0.40^{a}	56.6 ± 0.94^{a}
	30	11	1.80 ± 0.048^{a}	0.51 ± 0.020^{a}	49.9 ± 0.52^{a}	28.5 ± 0.57^{a}	57.2 ± 1.19^{a}
	300	11	1.79 ± 0.051^{a}	0.48 ± 0.012^{a}	49.7 ± 0.75^{a}	27.2 ± 0.61^{a}	54.7 ± 0.99^{a}
H. armigera	0	14	1.67 ± 0.058^{a}	0.52 ± 0.022^{a}	43.0 ± 1.29^{b}	31.3 ± 0.73^{b}	73.0 ± 1.27^{a}
	30	14	1.67 ± 0.036^{a}	0.56 ± 0.011^{a}	47.0 ± 0.70^{a}	33.6 ± 0.62^{a}	71.5 ± 0.76^{a}
	300	14	1.69 ± 0.056^{a}	0.54 ± 0.021^{a}	45.3 ± 1.16^{ab}	32.3 ± 0.70^{ab}	71.5 ± 1.36^{a}
H. zea	0	5	1.98 ± 0.078^{a}	0.63 ± 0.037^{a}	50.2 ± 1.09^{a}	31.9 ± 1.22^{a}	63.5 ± 2.56^{a}
	30	5	1.51 ± 0.191^{b}	0.39 ± 0.105^{b}	44.8 ± 1.49^{b}	23.3 ± 4.51^{a}	52.1 ± 9.57^{a}
	300	5	1.75 ± 0.090^{ab}	0.53 ± 0.035^{ab}	46.8 ± 1.30^{ab}	30.7 ± 1.84^{a}	65.4 ± 2.96^{a}

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

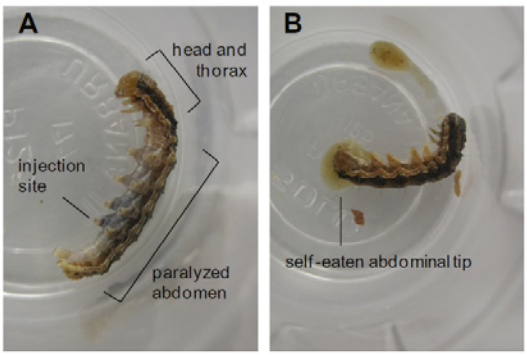


Fig. 4. Typical symptoms of capsaicin toxicity when it was injected into hemocoel of larvae in the three *Helicoverpa* species. As an example, a fifth instar larva of *H. armigera* was injected with 400 μg capsaicin into a posterior proleg. (A) Abdomen is paralyzed and turns blackish. (B) It was frequently observed that the affected larva feeds on the tip of its own abdomen, which usually leads to bleeding and death.

weights after the larval development period were, however, lower at the capsaicin treatments in *H. armigera* and *H. zea*, but not in *H. assulta* (Fig. 5). The effect on the pupal weight appears to be dose-dependent in *H. armigera* and *H. zea*. It took a longer time to reach pupation in the capsaicin treatment than in the control in *H. armigera* and *H. zea*, but not in *H. assulta* (data not shown), suggesting that sub-lethal doses of capsaicin hindered larval development.

4. Discussion

The results of the present study show that long-term dietary exposure to capsaicin has a growth-inhibiting effect on *S. frugiperda*, *H. virescens*, *H. subflexa*, and *H. zea*, little effect on *H. armigera*, and a growth-stimulating effect on *H. assulta* larvae (Fig. 2). The first four species compensate for slower larval growth by increasing the length of the larval period so that the same final pupal weight is attained, while *H. armigera* produces smaller pupae at the higher capsaicin concentrations. The long-term feeding experiments are complicated by the unequal performance of the six species on the artificial diet used. The pinto bean diet seems to be the most beneficial for *H. virescens*, which reached its maximum larval weight at the ninth day after hatching, followed by *H. armigera* and *S. frugiperda*. These three species experienced very little mortality, but nearly all larvae in a given treatment showed about the same degree of growth retardation. *H. subflexa* and *H. zea* however experienced higher mortality, which may have operated differentially producing more vigorous survivors showing less growth inhibition. The use of different diets optimized for each species might have equalized larval performance in the absence of capsaicin, but then species differences in response to capsaicin would be confounded with possible interactions of capsaicin with different diet ingredients. For the other five species, the effects of capsaicin did not seem to be related to how well they performed on capsaicin-free diet, except for *H. zea* where 200 mg/l had a disproportionately strong effect in reducing survivorship. Capsaicin slightly reduced larval growth rate and prolonged the larval period in *H. armigera* which is well-adapted to this diet; these reductions were not statistically significant but the decrease in pupal weight was. It might be argued that any positive effect of capsaicin, if it existed in *H. armigera*, would be harder to detect if growth is already as high as it can be on such a suitable diet. Repeating the feeding experiments with a different common diet deliberately chosen to be sub-optimal for *H. armigera* would be one way to test this idea.

The short-term feeding experiments in which nutritional indices were measured were originally intended to investigate whether stimulation of *H. assulta* growth was due to increased consumption of food due to capsaicin acting as a feeding stimulus, or whether insensitivity of *H. assulta* to toxic effects of capsaicin enabled it to more efficiently utilize the same amount of food. Their fairly low resolution may be due to restriction to two days of measurement, and due to the first exposure to capsaicin being at the fifth instar, when detoxicative enzymes are most highly expressed. The reduced consumption (RCR) and digestibility (AD) shown by *H. zea* resulted in a lower relative growth rate (RGR), consistent with its pronounced long-term responses as the most sensitive species in the first experiment. *H. assulta* showed neither of the responses hypothesized from the long-term feeding experiment. *H. armigera* showed a higher AD and conversion efficiency (ECI) at the lower capsaicin concentration, but these were not reflected in short-term RGR or in long-term growth (if these same rates had applied to younger larvae). Provision of a suboptimal diet to *H. armigera* might have amplified these subtle efficiency advantages into a detectable growth effect, which was not seen in long-term feeding of an optimal diet.

The injection experiment bypasses the digestive system and enables a more direct assay of toxic effects on other tissues. This showed that capsaicin is less acutely toxic to *H. assulta*, and that sub-lethal effects due to presence of capsaicin in the circulatory system are also less in this species. This suggests that *H. assulta*, a specialist on Solanaceae feeding on the fruits of hot pepper in the field, has physiological mechanisms for coping with this

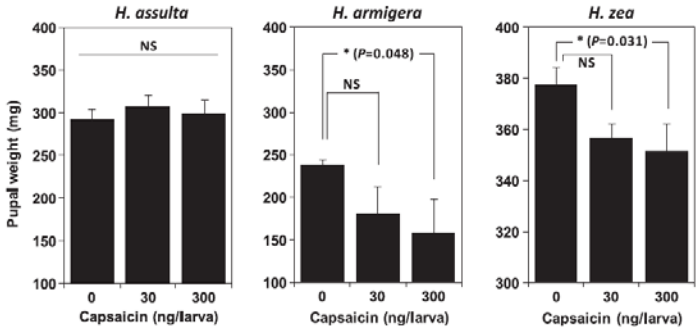


Fig. 5. Effect of sublethal-dose capsaicin injection on pupal weight. The early fifth instar larvae were injected with 30 or 300 ng of capsaicin delivered in 1 μl insect saline solution and left to feed on artificial diet until pupation. Mean pupal weights (±SE) were analyzed by ANOVA, followed by Dunnett's Post-hoc test ($P < 0.05$) to compare the other two capsaicin treatments to control. *Significantly different at $P < 0.05$ and NS refers to not significant.

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

noxious compound from the host plant. Moreover, the ability of *H. assulta* larvae to increase growth in the presence of capsaicin, indicates mechanisms for utilizing it for a feeding stimulus or metabolic benefit over the long term. In the short-term food consumption and utilization experiments, no nutritional indices increased when the fifth instar larva of *H. assulta* was fed with capsaicin for 2 days (Table 2). Measurement of nutritional indices over a longer period starting with neonates would be necessary to determine whether the long-term growth benefit is attributable to increased food consumption, conversion efficiency, or both.

Negative effects of capsaicin on the other five noctuid species might be related to their host ranges. Although the host range of the generalists tested is very broad, they have no coherent relationship with hot pepper in the field (Casmuz et al., 2010; Fitt, 1989). *H. armigera*, which is often collected in the fruit of hot pepper plant (Liu et al., 2004), however, seems to be relatively tolerant of capsaicin. It is noteworthy that no significant inhibition was found in the larval development of *H. armigera*, although pupal weight decreased when *H. armigera* larvae were injected with sub-lethal doses of capsaicin. Increase of AD and ECI in nutritional indices (Table 2) suggests higher digestibility and higher efficiency of conversion of ingested food at low capsaicin concentration (30 mg/l), but not at high (300 mg/l). Tolerance to capsaicin in the generalist *H. armigera* is evident in some aspects, but its tolerance seems to be different from that of the specialist, *H. assulta*.

The convulsive behavior caused by capsaicin injection might be provoked by neuronal activation via a type of capsaicin receptor (Caterina et al., 1997); the behavior was obviously different from an avoidance behavior. During abdominal paralysis, the head and thorax retained mobility, enabling movement and even feeding. Paralysis of the abdomen is probably due to the topical injection of capsaicin at the very rear part of the body, or due to the restricted localization of the potential capsaicin target site in the insect. Some larvae from all three species with paralyzed abdomen were able to move their head and thorax, and sometimes bite their own abdomen, leading to leakage of hemolymph from the hemocoel and death. This suggests that the affected larva had lost the sensation of pain in the abdomen. Alternatively, the biting behavior may be a form of self-defense in response to a painful feeling in the abdomen. Larvae will attempt to bite the fingers of a human who is picking them up. The cause of the blackish color appearing in the affected abdomen is also not understood.

The molecular mechanism of capsaicin pungency in mammals is known to be related to a transient receptor potential V1 (TRPV1) channel protein, which responds to vanilloids including capsaicin as well as to heat (43 °C) (Caterina et al., 1997). A similar channel protein might be responsible for the convulsive behavior of the larva observed in the present study, but it is still unknown whether insects have a homolog of the mammalian channel with similar properties. The *Drosophila melanogaster* genome encodes two TRPV subfamily proteins, Nanchung (CG5842) and Inactive (CG4536) involved in hearing. Other insect genomes also contain the homologs of the two TRPV genes (Matsuura et al., 2009), suggesting the conserved presence of the two paralogs of TRPV genes in insects. Capsaicin, however, did not trigger any cellular response by Nanchung or Inactive (Gong et al., 2004), but the wild-type adult fly showed a preference for capsaicin when it was provided with sucrose in a two-choice behavioral assay (Al-Anzi et al., 2006). Furthermore, capsaicin caused a shift in the preferred temperature of the American cockroach, *Periplaneta americana* (Olszewska, 2010). With the exception of *D. melanogaster*, none of the insect TRPV genes has been subjected to functional studies.

The paralyzed abdomen was sometimes restored to the normal condition, after injection with the low dose of capsaicin. Such a reversible paralysis could imply that the capsaicin effect can be inactivated by unknown mechanisms. Such a recovery could be achieved by detoxification processes. In mammals, capsaicin undergoes epoxidation, oxidation, *O*-methylation, dehydrogenation, and glucuronide conjugation to produce various kinds of metabolites (Chanda et al., 2008; Noami et al., 2006; Reilly and Yost, 2006). In microorganisms, capsaicin goes through degradation into vanillin, vanillylamine and 8-methylnonenoic acid by *Capsicum*-associated bacteria (Flagan and Leadbetter, 2006; Romano et al., 2011). Peroxidase-catalyzed oxidation (Díaz et al., 2004) and glucose conjugation (Shimoda et al., 2007) of capsaicin are known in plants. However, nothing is known in insects with regards to the detoxification or metabolism of capsaicin. Further investigations should shed light on the mechanism underlying the recovery from capsaicin effect as well as the acute toxicity.

No external symptoms were observed when the small amounts of capsaicin (30 and 300 ng/larva) were injected, but smaller pupae compared to the control were produced in *H. armigera* and *H. zea*, but not in *H. assulta* (Fig. 5), meaning that only about 400 and 160 times less amount of capsaicin than the ED_{50} values (Fig. 3) were enough to affect larval growth without acute toxicity for *H. armigera* and *H. zea*, respectively. Pupal weight is dependent on the maximum weight which a larva attains at the last instar and this is in turn dependent on the amount of food ingested and the digestibility during the developmental period (Nijhout, 2003). Thus, the low pupal weights caused by the injection of small amounts of capsaicin in *H. armigera* and *H. zea* may result from either lower amount of food ingested, lower digestibility caused by the disturbance of the physiological state of the larva, or higher consumption of metabolic energy in order to detoxify the xenobiotic. These are all responses acting over a long period of the larval period, not acute responses. Smaller amounts of capsaicin appear to impose some sort of metabolic load on the other two species, but not *H. assulta*.

As shown in this study, host specialization of *H. assulta* seems to be related to the response to the host allelochemical. According to another study on nutritional tests with gossypol, nicotine, tomatine, or capsaicin on *H. assulta* and *H. armigera* larvae, *H. assulta* is tolerant particularly to nicotine and capsaicin whereas *H. armigera* shows broad tolerance to the four allelochemicals (Dong et al., 2002). Capsaicin is also known to stimulate the oviposition of *H. assulta* adults (Lee et al., 2006).

Monotonic dose-dependency was not always seen in this study. The low dose of capsaicin was more detrimental than the higher dose to *S. frugiperda* larval growth (Fig. 2A), and nutritional indices were significantly different at the lower dose than the higher one in *H. armigera* and *H. zea* (Table 2). In these cases, capsaicin is uniformly detrimental, but paradoxically less so at higher concentrations. Another exception to monotonic dose-dependency is hormesis (Calabrese and Blain, 2005), in which low doses are beneficial but high doses detrimental. Examples involving plant allelochemicals are given by the pine sawfly *Neodiprion sertifer* where female weight increased due to a small dose of resin acid derived from a host plant, but decreased in response to a higher dose (Björkman, 1997); and gossypol from cotton glands produced such a dual effect on the larval weight of *H. virescens* (Stipanovic et al., 1986). In these cases the benefit at low doses is relatively small compared with the detriment at high doses. Whether the strongly positive effect of capsaicin on *H. assulta* larval growth at relatively high doses would continue monotonically was not tested here; this would involve feeding larvae substantially higher concentrations than normally encountered in hot pepper fruits. However, since high doses by injection eventually produced paralysis it is likely that high enough concentrations fed to larvae would inhibit growth after all.

In conclusion, this paper describes the comparative analysis of the differential effects of capsaicin among insects with different host ranges. Dietary capsaicin was detrimental to the larval development

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

of *S. frugiperda*, *H. virescens*, *H. subflexa*, and *H. zea*, but it was relatively neutral to that of *H. armigera*. *H. assulta*, however, performed even better in the presence of capsaicin. Capsaicin injection indicated that *H. assulta* is more tolerant to capsaicin than the other two *Helicoverpa* spp. This study suggests that capsaicin can act as a defensive allelochemical against insects which are not adapted to the *Capsicum* plant. Furthermore, the beneficial effect of dietary capsaicin on the larval development of *H. assulta* leaves an open question on the physiological role of capsaicin in the adaptation of *H. assulta* to a host on which it is the primary agricultural pest. This study lays a stepping stone for further investigation of the molecular toxicity and detoxification of capsaicin in insects. The study on the ecological and physiological functions of capsaicin in the context of insect-plant interactions could shed light on the *raison d'être* of one of the most pungent compounds in the plant world.

Acknowledgements

This work was supported by the Max-Planck-Gesellschaft and the International Max Planck Research School (IMPRS). The authors thank Dr. Jum-Rae Cho (RDA, Korea) for helping to collect *H. assulta* and Dr. Lisa Knolhoff for comments on the original manuscript.

References

Al-Anzi, B., Tracey, W.D., Benzer, S., 2006. Response of *Drosophila* to wasabi is mediated by *painless*, the fly homolog of mammalian TRPA1/ANKTM1. Current Biology 16, 1034–1040.

Appendino, G., 2008. Capsaicin and capsaicinoids. In: Fattorusso, E., Taglialatela-Scafati, O. (Eds.), Modern Alkaloids. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 73–109.

Aza-González, C., Núñez-Palenius, H., Ochoa-Alejo, N., 2011. Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). Plant Cell Reports 30, 695–706.

Baek, S., Cho, K., Song, Y.H., Lee, J.H., 2009. Sampling plans for estimating pepper fruit damage levels by Oriental tobacco budworm, *Helicoverpa assulta* (Guenée), in hot pepper fields. Journal of Asia-Pacific Entomology 12, 175–178.

Bateman, M.L., 2006. Impact of plant suitability, biogeography, and ecological factors on associations between the specialist herbivore *Heliothis subflexa* G. (Lepidoptera: Noctuidae) and the species in its host genus, *Physalis* L. (Solanaceae), in west-central Mexico. Ph.D. Thesis, North Carolina State University, Raleigh, North Carolina.

Björkman, C., 1997. A dome-shaped relationship between host plant allelochemical concentration and insect size. Biochemical Systematics and Ecology 25, 521–526.

Calabrese, E.J., Blain, R., 2005. The occurrence of hormetic dose responses in the toxicological literature, the hormesis database: an overview. Toxicology and Applied Pharmacology 202, 289–301.

Casmuz, A., Laura Juarez, M., Guillermina Socias, M., Gabriela Murua, M., Prieto, S., Medina, S., Willink, E., Gastaminza, G., 2010. Review of the host plants of fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). Revista de la Sociedad Entomologica Argentina 69, 209–231.

Caterina, M.J., Schumacher, M.A., Tominaga, M., Rosen, T.A., Levine, J.D., Julius, D., 1997. The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature 389, 816–824.

Chanda, S., Bashir, M., Babbar, S., Koganti, A., Bley, K., 2008. *In vitro* hepatic and skin metabolism of capsaicin. Drug Metabolism and Disposition 36, 670–675.

Cho, S., Mitchell, A., Mitter, C., Regier, J., Matthews, M., Robertson, R.O.N., 2008. Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. Systematic Entomology 33, 581–594.

Choi, K.S., Boo, K.S., 1989. Food attractancy of the oriental tobacco budworm, *Heliothis assulta*, larvae. Korean Journal of Applied Entomology 28, 88–92.

Cowles, R.S., Keller, J.E., Miller, J.R., 1989. Pungent spices, ground red pepper, and synthetic capsaicin as onion fly ovipositional deterrents. Journal of Chemical Ecology 15, 719–730.

Díaz, J., Pomar, F., Bernal, A., Merino, F., 2004. Peroxidases and the metabolism of capsaicin in *Capsicum annuum* L.. Phytochemistry Reviews 3, 141–157.

Dong, J.-F., Zhang, J.-H., Wang, C.-Z., 2002. Effects of plant allelochemicals on nutritional utilization and detoxication enzyme activities in two *Helicoverpa* species. Acta Entomologica Sinica 45, 296–300.

Edelson, J.V., Duthie, J., Roberts, W., 2002. Toxicity of biorational insecticides: activity against the green peach aphid, *Myzus persicae* (Sulzer). Pest Management Science 58, 255–260.

Ehrlich, P.R., Raven, P.H., 1964. Butterflies and plants: a study in coevolution. Evolution 18, 586–608.

Finney, D.J., 1971. Probit Analysis, third ed. Cambridge University Press, Cambridge, UK.

Fitt, G.P., 1989. The ecology of *Heliothis* species in relation to agroecosystems. Annual Review of Entomology 34, 17–53.

Flagan, S.F., Leadbetter, J.R., 2006. Utilization of capsaicin and vanillylamine as growth substrates by *Capsicum* (hot pepper)-associated bacteria. Environmental Microbiology 8, 560–565.

Gong, Z., Son, W., Chung, Y.D., Kim, J., Shin, D.W., McClung, C.A., Lee, Y., Lee, H.W., Chang, D.-J., Kaang, B.-K., Cho, H., Oh, U., Hirsh, J., Kernan, M.J., Kim, C., 2004. Two interdependent TRPV channel subunits, Inactive and Nanchung, mediate hearing in *Drosophila*. The Journal of Neuroscience 24, 9059–9066.

Han, M.-W., Lee, J.-H., Son, J.-S., 1994. Intra- and inter-plant distribution of *Helicoverpa assulta* (Lepidoptera: Noctuidae) eggs in red pepper and tobacco fields. Korean Journal of Applied Entomology 33, 6–11.

Hori, M., Nakamura, H., Fujii, Y., Suzuki, Y., Matsuda, K., 2011. Chemicals affecting the feeding preference of the Solanaceae-feeding lady beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae). Journal of Applied Entomology 135, 121–131.

Jung, J.W., Cho, M.C., Cho, Y.S., 2006. Fruit quality of once-over harvest pepper (*Capsicum annuum*) cultivar 'Saengryeg No. 211' and 'Saengryeg No. 213'. Korean Journal of Horticultural Science & Technology 24, 205–209.

Larue, E., Welty, C., 2010. Oviposition behavior and larval development of the European corn borer (Lepidoptera: Crambidae) on sweet versus hot peppers. Journal of Entomological Science 45, 353–365.

Lee, H.S., Hieu, T.T., Ahn, Y.J., 2006. Oviposition-stimulating activity of (*E*)-capsaicin identified in *Capsicum annuum* fruit and related compounds towards *Helicoverpa assulta* (Lepidoptera: Noctuidae). Chemoecology 16, 153–157.

Levey, D.J., Tewksbury, J.J., Cipollini, M.L., Carlo, T.A., 2006. A field test of the directed deterrence hypothesis in two species of wild chili. Oecologia 150, 61–68.

Liu, Z.D., Li, D.M., Gong, P.Y., Wu, K.J., 2004. Life table studies of the cotton bollworm, *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae), on different host plants. Environmental Entomology 33, 1570–1576.

Madhumathy, A.P., Aivazi, A.-A., Vijayan, V.A., 2007. Larvicidal efficacy of *Capsicum annum* against *Anopheles stephensi* and *Culex quinquefasciatus*. Journal of Vector Borne Diseases 44, 223–226.

Matsuura, H., Sokabe, T., Kohno, K., Tominaga, M., Kadowaki, T., 2009. Evolutionary conservation and changes in insect TRP channels. BMC Evolutionary Biology 9, 228.

Matthews, M., 1999. Heliothine Moths of Australia. CSIRO Publishing, Collingwood.

Mazourek, M., Pujar, A., Borovsky, Y., Paran, I., Mueller, L., Jahn, M.M., 2009. A dynamic interface for capsaicinoid systems biology. Plant Physiology 150, 1806–1821.

Mitter, C., Poole, R.W., Matthews, M., 1993. Biosystematics of the Heliothinae (Lepidoptera: Noctuidae). Annual Review of Entomology 38, 207–225.

Nijhout, H., 2003. The control of body size in insects. Developmental Biology 261, 1–9.

Noami, M., Igarashi, K., Kasuya, F., Ohta, M., Kanamori-Kataoka, M., Seto, Y., 2006. Study on the analysis of capsaicin glucuronide in rat urine by liquid chromatography–mass spectrometry after enzymatic hydrolysis. Journal of Health Science 52, 660–665.

Olszewska, J., 2010. Vanilloid receptors – comparison of structure and functions in mammals and invertebrates. Folia Biologica (Kraków) 58, 1–7.

Reilly, C.A., Yost, G.S., 2006. Metabolism of capsaicinoids by P450 enzymes: a review of recent findings on reaction mechanisms, bio-activation, and detoxification processes. Drug Metabolism Reviews 38, 685–706.

Romano, D., Gandolfi, R., Guglielmetti, S., Molinari, F., 2011. Enzymatic hydrolysis of capsaicins for the production of vanillylamine using ECB deacylase from *Actinoplanes utahensis*. Food Chemistry 124, 1096–1098.

Shimoda, K., Kwon, S., Utsuki, A., Ohiwa, S., Katsuragi, H., Yonemoto, N., Hamada, H., Hamada, H., 2007. Glycosylation of capsaicin and 8-nordihydrocapsaicin by cultured cells of *Catharanthus roseus*. Phytochemistry 68, 1391–1396.

Slansky Jr., F., 1992. Allelochemical-nutrient interactions in herbivore nutritional ecology. In: Rosenthal, G.A., Berenbaum, M.R. (Eds.), Herbivores: Their Interactions with Secondary Plant Metabolites, . second ed., Ecological and Evolutionary Processes second ed., vol. II Academic Press Inc., Academic Press Ltd., pp. 135–174.

Stipanovic, R.D., Williams, H.J., Smith, L.A., 1986. Cotton Terpenoid Inhibition of *Heliothis virescens* Development. Natural Resistance of Plants to Pests. American Chemical Society, pp. 79–94.

Surh, Y.-J., Kundu, J.K., 2011. Molecular mechanisms of chemoprevention with capsaicinoids from chili peppers. In: Mutanen, M., Pajari, A.-M. (Eds.), Vegetables, Whole Grains, and Their Derivatives in Cancer Prevention. Springer, Netherlands, pp. 123–142.

Tewksbury, J.J., Manchego, C., Haak, D.C., Levey, D.J., 2006. Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. Journal of Chemical Ecology 32, 547–564.

Tewksbury, J.J., Nabhan, G.P., 2001. Seed dispersal: directed deterrence by capsaicin in chillies. Nature 412, 403–404.

Tewksbury, J.J., Reagan, K.M., Machnicki, N.J., Carlo, T.A., Haak, D.C., Peñaloza, A.L.C., Levey, D.J., 2008. Evolutionary ecology of pungency in wild chilies. Proceedings of the National Academy of Sciences 105, 11808–11811.

Waldbauer, G.P., 1968. The consumption and utilization of food by insects. Advances in Insect Physiology 5, 229–288.

Weissenberg, M., Klein, M., Meisner, J., Ascher, K.R.S., 1986. Larval growth inhibition of the spiny bollworm, *Earias insulana*, by some steroidal secondary plant compounds. Entomologia Experimentalis et Applicata 42, 213–217.

Yang, C.-Y., Jeon, H.-Y., Cho, M.-R., Kim, D.-S., Yiem, M.-S., 2004. Seasonal occurrence of oriental tobacco budworm (Lepidoptera: Noctuidae) male and chemical control at red pepper fields. Korean Journal of Applied Entomology 43, 49–54.

Please cite this article in press as: Ahn, S.-J., et al. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology (2011), doi:10.1016/j.jinsphys.2011.05.015

3.2. Manuscript II

Metabolic detoxification of capsaicin by UDP-glycosyltransferase in three *Helicoverpa* species

Seung-Joon Ahn[1], Francisco R. Badenes-Pérez[1,5], Michael Reichelt[2], Aleš Svatoš[3], Bernd Schneider[4], Jonathan Gershenzon[2] and David G. Heckel[1,*]

Submitted to *Archives of Insect Biochemistry and Physiology* (June 30, 2011)

[1]Department of Entomology, [2]Department of Biochemistry, [3]Mass Spectrometry Research Group, [4]Biosynthesis/NMR Research Group, Max Planck Institute for Chemical Ecology, Jena, 07745, Germany

[5]Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, Madrid, 28006, Spain

Running headline

Capsaicin glucoside and UGT in three *Helicoverpa* spp.

Corresponding author

David G. Heckel, Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, 07745, Germany, Tel.: +49 3641 571501; fax: +49 3641 571502. E-mail: heckel@ice.mpg.de

ABSTRACT

Capsaicin β-glucoside was isolated from the feces of *Helicoverpa armigera*, *H. assulta* and *H. zea* that fed on capsaicin-supplemented artificial diet. The chemical structure was identified by NMR spectroscopic analysis as well as by enzymatic hydrolysis. The excretion rates of the glucoside were different among the three species; those in the both two generalists, *H. armigera* and *H. zea*, were higher than in a specialist, *H. assulta*. UDP-glycosyltransferases (UGT) enzyme activities measured from the whole larval homogenate of the three species with capsaicin and UDP-glucose as substrates were also higher in the two generalists. Compared among five different larval tissues (labial glands, testes from male larvae, midgut, the Malpighian tubules, and fatbody) from the three species, the formation of the capsaicin glucoside by one or more UGT is high in the fat body of all the three species as expected, as well as in *H. assulta* Malpighian tubules. Optimization of the enzyme assay method is also described in detail. Although the lower excretion rate of the unaltered capsaicin in *H. assulta* indicates higher metabolic capacity toward capsacin than in the other two generalists, the glucosylation per se seems to be insufficient to explain the decrease of capsaicin in the specialist, suggesting *H. assulta* might have another important mechanism to deal with capsaicin more specifically.

Keywords: capsaicin; capsaicin glucoside; metabolic detoxification; UDP-glycosyltransferase; *Helicoverpa armigera*; *Helicoverpa assulta*; *Helicoverpa zea*

INTRODUCTION

Herbivorous insects are faced with a large amount of noxious chemicals in their host plants and have evolved various detoxification mechanisms to avoid their harmful effects (Schoonhoven et al., 2005). Metabolic detoxification is considered one of the important mechanisms by which insects can cope with a variety of xenobiotics including plant allelochemicals (Brattsten, 1992). Glycosylation, among the various mechanisms, converts lipophilic aglycones into more hydrophilic glycosides, facilitating excretion or sequestration for further utilization (Wilkinson, 1986). UDP-glycosyltransferase (UGT) is responsible for the glycosylation by catalyzing the conjugation of a glycosyl group from an activated sugar donor, UDP-glycoside, with various small hydrophobic molecules. UGT is known to be involved not only in detoxification, but also in cuticular tanning (Hopkins and Kramer, 1992), pigmentation (Hopkins and Ahmad, 1991; Mizokami and Yoshitama, 2009; Wiesen et al., 1994), and olfaction (Robertson et al., 1999; Wang et al., 1999) in different insects. Compared to other detoxification enzymes such as cytochrome P-450s (P450), carboxylesterases and glutathione transferases (GST), only limited information is available about UGT-mediated metabolic detoxification in insects (Després et al., 2007).

Capsaicin (8-methyl-*N*-vanillyl-6-nonenamide; **1**), an alkaloid found only in *Capsicum* spp. (Solanaceae), is responsible for the pungency of hot pepper fruits at least to mammals (Caterina et al., 1997). Capsaicin plays an important ecological role by selectively deterring mammalian fruit eaters that otherwise destroy the seeds, but not preventing birds from dispersing the seeds (Tewksbury et al., 2006). Capsaicin also inhibits the growth of a fungus harming the seeds (Tewksbury et al., 2008). There are several studies on the effects of capsaicin on insects. Capsaicin is known to deter oviposition in the onion fly *Delia antiqua* (Diptera: Anthomyiidae) (Cowles et al., 1989), to inhibit the feeding of a ladybird beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae) (Hori et al., 2011), and to retard larval growth in the spiny bollworm *Earias insulana* (Lepidoptera: Noctuidae) (Weissenberg et al., 1986). *Capsicum* extracts have larvicidal activity to two species of mosquitoes, *Anopheles stephensi* and *Culex quinquefasciatus* (Madhumathy et al., 2007), and synergistic effects with insecticides on the green peach aphid, *Myzus persicae* (Edelson et al., 2002). Larval development time of *Ostrinia nubilalis* was significantly delayed on pungent rather than non-pungent peppers (Larue and Welty, 2010).

The Oriental tobacco budworm, *Helicoverpa assulta* (Guenée) (Lepidoptera: Noctuidae), is one of the few insects that can successfully feed on hot pepper fruits. It is a specialist on the family Solanaceae, feeding also on *Lycopersicon*, *Nicotiana*, *Physalis*, and *Solanum* in Korea, Japan, China, Australia, and Africa (Cho et al., 2008; Matthews, 1999; Mitter et al., 1993; Yang et al., 2004). The relationship between *H. assulta* and hot pepper has been described by studies on egg distribution (Han et al., 1994), larval feeding preference (Choi and Boo, 1989) and feeding damage (Baek et al., 2009). Recently, we have found that *H. assulta* is more tolerant to capsaicin than other noctuid species

including *H. armigera* and *H. zea*, suggesting further investigations on the detoxification mechanism of capsaicin in the host-specialist (Ahn et al., 2011).

Here we report a capsaicin glucoside as a novel metabolite of capsaicin in insects, and we also demonstrate the conjugation rates are different among three *Helicoverpa* species, *H. armigera*, *H. assulta* and *H. zea*. In addition, the UGT activity toward capsaicin was studied *in vitro* in different larval tissues.

MATERIALS AND METHODS

Insects

Three different *Helicoverpa* moths were used in this study; *H. armigera* and *H. assulta* were collected from Queensland (Australia) and Suwon (Korea), respectively; and *H. zea* from North Carolina was provided by Dr. Fred Gould, North Carolina State University. Each species was maintained under laboratory conditions (26 °C, 55% relative humidity and 16:8 h L:D) with artificial diets in Jena, Germany. For details of the rearing procedure and artificial diet composition, please refer to Ahn et al. (2011).

Extraction and Isolation of Capsaicin Glucoside from the Larval Feces

The larvae of *H. assulta* were used for the initial isolation of the unknown metabolite of capsaicin. The feces were collected from the early fifth instar larvae fed on 200 mg/l capsaicin-spiked artificial diet for 2 days, and the control feces were also collected from the larvae fed on normal artificial diet for comparison. A candidate peak in LC-MS was detected that was seen only in the treatment feces extract, but not in the control feces extract. In order to obtain enough of the candidate metabolite for structure elucidation, the scale was increased to collect the treatment feces from 300 larvae (ca. 80 g, wet weight) and to extract it with 300 ml methanol by shaking at 200 rpm for 24 hr at room temperature. The extracts were filtered with filter paper and the extraction procedure was repeated. The combined filtered extracts were concentrated by rotary evaporator (BÜCHI Rotavapor R-114, BÜCHI Labortechnik AG, Switzerland) under vacuum. The final concentrate was separated into ethyl acetate (1 vol) and water (0.5 vol) phases and the aquatic phase was discarded. The ethyl acetate phase was evaporated to dryness, resuspended in methanol-water (50:50, v:v), and fractionated by a semipreparative HPLC (Agilent 1100 Series system) equipped with Supelcosil LC18-DB column (250 ×10 mm, 5 μm) to obtain a pure candidate peak for further analyses.

LC-MS and NMR Spectroscopy

Samples were analyzed by HPLC (Agilent 1100 Series) equipped with a NUCLEODUR Sphinx RP column (250 × 4.6 mm, 5 μm) using gradient mobile phase with 0.2 % formic acid (A) and acetonitrile

(B) as follows: 10 – 85 % (v/v) B for 25 min, 85 – 100 % (v/v) B for 6 sec, 100 % B for 2 min, 100 – 10 % (v/v) B for 6 sec, 10 % B for 3 min 54 sec. The flow rate was 1.0 ml/min, and the eluate was monitored at 280 nm UV absorbance. A selected set of samples was analyzed by HPLC-ESI-MS equipped with Esquire 6000 Ion trap mass spectrometer, Bruker (Bruker Daltonics, Germany) operated in positive mode in the range *m/z* 50-1200. Skimmer voltage, 40 eV; capillary exit voltage, 114eV; capillary voltage, -4,200 V; nebulizer pressure, 35 psi; drying gas, 11 l/min; gas temperature, 330 °C. NMR spectra of the purified metabolite were measured on a Bruker DRX 500 (Bruker Biospin, Rheinstetten, Germany). ^{1}H NMR, ^{13}C NMR, DEPT, ^{1}H–^{1}H COSY, HMBC, HSQC and ROESY spectra were measured using an inverse-detection probe (5 mm). The operating frequencies were 500.13MHz for acquiring ^{1}H NMR and 125.75MHz for acquiring ^{13}C NMR spectra. Samples were measured at 300 K in $CDCl_3$ with TMS as the internal standard.

Quantitative Analysis of Capsaicin Glucoside from Individual Larvae

Once the metabolite had been identified as capsaicin glucoside, the extraction and analysis methods were scaled down to investigate the glucosylation rate of individual larvae. After feeding the capsaicin-diet (ca. fresh weight 1,000 mg) for 2 days to *H. armigera*, *H. assulta*, or *H. zea*, the fifth instar larva and its feces were separately collected from the remaining diet and dried at 60 °C for 2 days. We measured the dry weights of diet eaten and the amount of feces excreted by individual larvae for 2 days. Based on this as well as the virtual content of capsaicin in the diet analyzed by HPLC, the amount of capsaicin eaten by larvae was calculated and the glucosylation rate was analyzed on the basis of molar content. For the extraction of capsaicin and capsaicin glucoside, the dried materials were ground in a mortar and pestle, collected into a 1.5 ml tube with fine brush, and the working weight of the powder was again measured. After 0.3 ml acetonitrile was added and vortexed for 5 min, it was centrifuged at 10,000 rpm for 30 min at room temperature. Supernatant was collected and the extraction procedure was repeated. The combined acetonitrile extract (ca. 0.6 ml) was concentrated by speed vacuum concentrator (Concentrator 5301, Eppendorf, Germany) and then dissolved in 0.3 ml methanol. The filtrate passed through a 0.45 μm membrane filter by using a 1 ml-syringe was used for quantitative analysis by LC-UV. External capsaicin standard curve was used to quantify capsaicin as well as capsaicin glucoside, by using the same UV-detector mode.

Enzymatic Hydrolysis of Capsaicin Glucoside

To verify the glycosidic linkage conformation of the conjugate, the putatively purified capsaicin glucoside was either treated with 10 units/ml of α-glucosidase (from yeast, Sigma-Aldrich, USA) at 37 ºC for 17 hr in 0.1 M phosphate buffer (pH 7.0) or treated with 1 unit/ml of β-glucosidase (from almond, Sigma-Aldrich, USA) at 37 ºC for 17 hr in 0.1 M acetate buffer (pH 5.0). The reaction mixture was boiled at 100 ºC for 5 min to inactivate the enzyme and cooled down on ice for another 5

min. After centrifugation at 13,000 rpm for 2 min, the supernatant was filtered and analyzed by LC-MS.

Preparation of Enzyme Extract

Fifth instar larvae of the three species were individually dissected in chilled 0.85% KCl solution into midgut, Malpighian tubules, fat body, labial glands, and testes (only from male larva). Each isolated tissue was washed in the chilled KCl solution and homogenized in 0.1 M sodium phosphate buffer (pH 7.0) containing 0.8% sodium cholate as a detergent by using a hand-driven plastic grinder on ice for 1 min. The homogenate was centrifuged at 15,000g for 30 min at 4°C and the supernatant crude enzyme extract was used for following enzyme assays. For the preparation of whole larva homogenate, the fresh fifth instar larva was individually ground by motor-driven Teflon grinder (10 strokes, 3 times) on ice and the other steps were the same as above.

In vitro Enzyme Assay

The standard reaction mixture contained, in a final volume of 100 µl, 0.1 M sodium phosphate buffer (pH 7.0), 5 mM UDP-glucose, 0.3 mM capsaicin (3 µl of a 10 mM stock solution dissolved in DMSO), 25 mM $MgCl_2$, 5 mM D-gluconolactone (as an inhibitor of potential β-glucosidase), and enzyme extract in an amount of less than 0.1 mg of total protein. The reaction was started by the addition of the enzyme source to the incubation mixture and conducted at 42 °C for 20 min, where the velocity of the enzyme reaction was linear (see Results section). Control was run in a same way in the absence of enzyme. The reaction was terminated by the addition of 200 µl of chilled methanol and centrifuged at 15,000g for 30 min at 4 °C. The supernatant was filtered and used for LC-UV analysis of capsaicin glucoside at the analytical condition described above. The specific enzyme activity was calculated as the amount of capsaicin glucoside (nmol) per time (min) per total protein (mg). Protein was quantified according to bicinchoninic acid (BCA) method (Smith et al., 1985).

RESULTS

Identification of Capsaicin Glucoside in Larval Feces

A candidate peak of a capsaicin metabolite was observed in the feces of *H. armigera* larvae that had been fed on the capsaicin-spiked diet that was not present in feces from control fed larvae. The fractionated and concentrated peak was determined to be capsaicin β-glucoside (**2**) by LC-MS analysis, NMR spectroscopy, and enzymatic hydrolysis. The molecular weight of the metabolite was measured by LC-MS to be 467 according to the ion mass detected at *m/z* 490 $[M + Na]^+$ in the positive mode and at *m/z* 512 $[M - H + HCO_2H]^-$ in the negative mode. ^{1}H-, ^{13}C- and 2D-NMR analysis (^{1}H,^{1}H-COSY, HSQC, HMBC) confirmed the chemical structure of *trans*-capsaicin β-glucoside (Table 1), including the β-configuration at the anomeric centre of the glucose unit ($^3J_{H\text{-}1''\text{–}H\text{-}2''}$ = 7.7 Hz). The *trans*-configuration of the double bond was established by comparing the chemical shifts of H-6', H-7', and H-8' (Table 1) with those reported for *cis*- and *trans*-capsiacin (Lin et al., 1993). In addition, the conjugate was hydrolyzed by β-glucosidase treatment to release the aglycone, capsaicin, but not by α-glucosidase, confirming once again that the glucose moiety is conjugated to capsaicin by a β-linkage (Fig. 2).

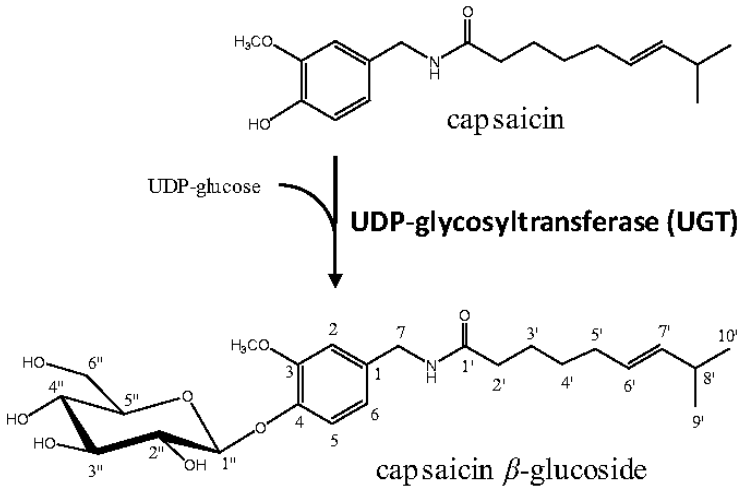


Figure 1. Capsaicin (**1**) is conjugated with UDP-glucose to produce capsaicin β-glucoside (**2**) catalyzed by UDP-glycosyltransferase.

Table 1. ^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) data of capsaicin β-glucoside (MeOH-d_4, TMS).

Position	^{1}H NMR	^{13}C NMR
	δ, J (Hz)	δ
1		134.8
2	6.93, d, 2.1	112.9
3		150.6
4		147.1
5	7.10, d, 8.3	117.8
6	6.81, dd, 8.3, 2.1	121.0
7	4.29, s	43.6
8	3.84, s	56.5
1'		176.0
2'	2.21, t, 7.4	36.8
3'	1.61, m	26.3
4'	1.37, m	30.2
5'	1.99, dt, 6.6, 6.6	33.1
6'	5.34, m	127.6
7'	5.37, m	138.9
8'	2.20, m	32.2
9', 10'	0.95, d, 6.8	22.9
1"	4.85, d, 7.7	102.7
2"	3.47 [1]	74.7
3"	3.44 [1]	77.6
4"	3.37 [1]	71.1
5"	3.37 [1]	78.0
6"a	3.68, dd 11.1, 5.1	62.2
6"b	3.85, dd, 11.1, 1.8	

[1] Chemical shifts from HSQC spectrum.

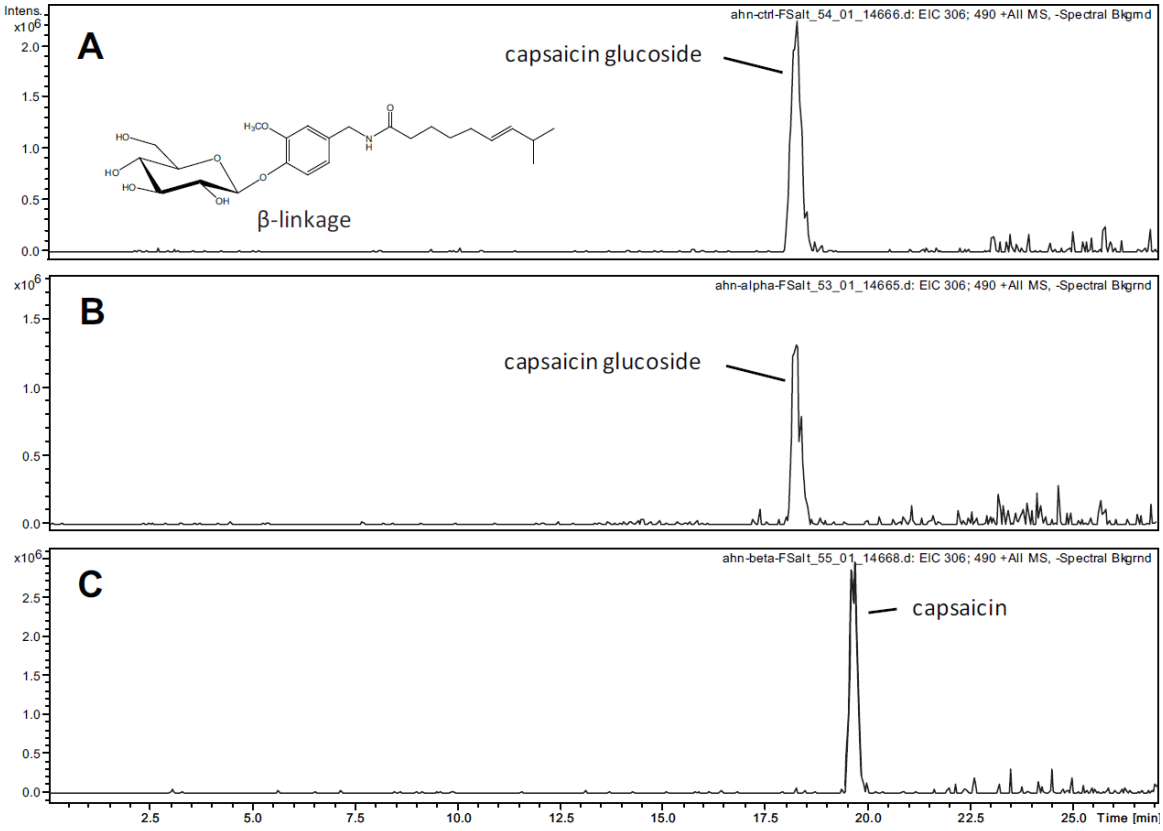


Figure 2. Extracted LC-MS chromatograms of purified capsaicin glucoside treated with (A) control, (B) α-glucosidase or (B) β-glucosidase. Displayed are the extracted ion traces of m/z 306 + m/z 490 in positive ionization mode where m/z 306 is $[M+H]^+$ for capsaicin and m/z 490 is $[M+Na]^+$ for capsaicin glucoside. Capsaicin glucoside was not hydrolyzed by α-glucosidase, but by β-glucosidases.

In vivo Glucosylation in three Helicoverpa spp.

The amounts of capsaicin glucoside and un-metabolized capsaicin in the feces of individual fifth instar larvae fed on capsaicin were compared among *H. armigera*, *H. assulta* and *H. zea*. All of the three *Helicoverpa* species produced the glucose conjugate of capsaicin in their larval feces when they were fed on capsaicin-spiked artificial diet. The amount of capsaicin glucoside, however, was significantly different among the species, while the amount of capsaicin ingested by larvae was not different among three species (data not shown); the glucosylation rate calculated by molar content was 7.2% and 7.7% in *H. armigera* and *H. zea*, respectively, whereas it was 2.3% in *H. assulta*, which was about 3 times less than the two generalists (Fig. 3A). In addition, the amount of unaltered capsaicin excreted in the feces of *H. assulta* was also significantly smaller (5.9%) than in that of *H. armigera* and *H. zea* (31.8 and 38.9%, respectively) (Fig. 3B).

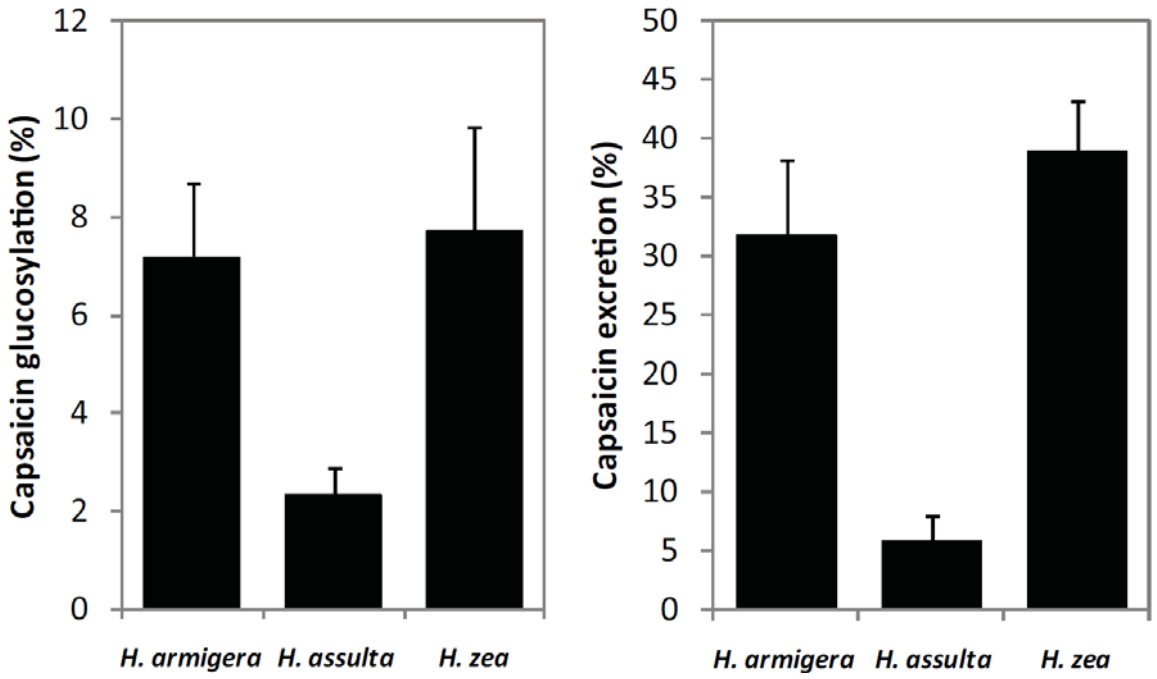


Figure 3. Comparison of excretion rates of (A) capsaicin glucoside (nmol capsaicin glucoside excreted per nmol capsaicin ingested) and (B) capsaicin (nmol capsaicin excreted per nmol capsaicin ingested) among *H. armigera*, *H. assulta*, and *H. zea*.

Optimization of in vitro UDP-Glucosyltransferase Activity

Effect of temperature and incubation time Fat body from *H. armigera* was used for the optimization experiments to test proper conditions for the capsaicin UGT activity measurement, since activity in the fat body was highest of all insect tissues. The activity linearly increased up to 30 min at 47 °C and up to 60 min at 42 °C. As the temperature got higher than 47 °C, the activity decreased significantly. When the incubation time was longer than 30 min, the linearity of the activity started to drop at lower temperature. Therefore, the optimized condition was set at 42 °C for 20 min, where the reaction velocity was linear with respect to time and temperature (Fig. 4A).

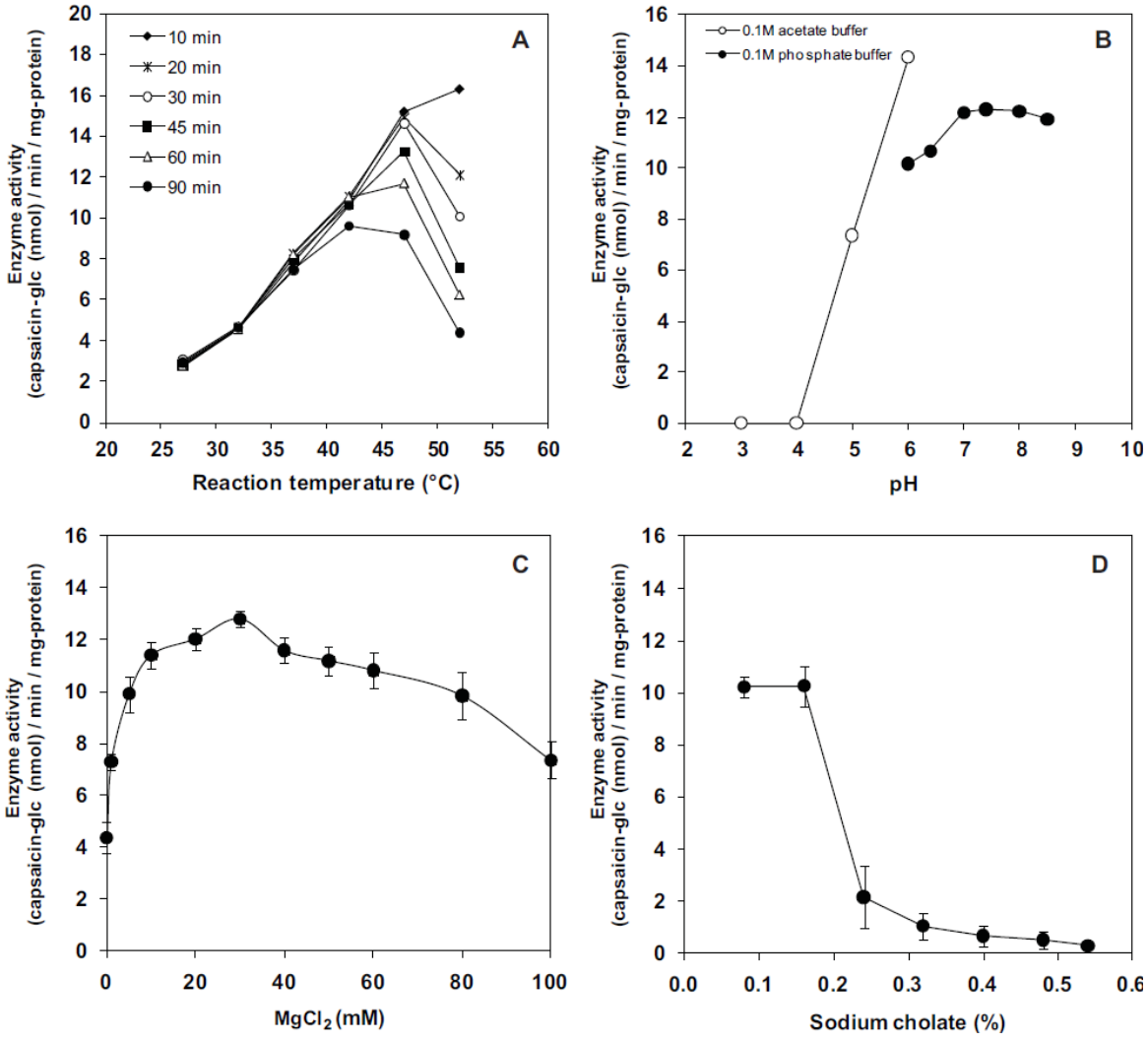


Figure 4. Optimization of capsaicin:UGT activity in preparations from fat body of fifth instar larvae of *Helicoverpa armigera*, depending on (A) temperature and incubation time, (B) pH, (C) $MgCl_2$, and (D) sodium cholate. Fat body collected from15 larvae was pooled and homogenated, being used as one biological replicate. Each point represents a mean value of three biological replicates.

Effect of pH The enzyme activity in the *H. armigera* fat body was detected starting from pH 5.0, increased steeply and reached a maximum at pH 7.0 - 7.5 expected for an intra-cellular enzyme, and then the activity decreased at pH 8.5, but higher pH values were not examined. Acetate buffer gave higher enzyme activity at pH 7.0 than phosphate buffer. However, 0.1 M phosphate buffer at pH 7.0 was used in further experiments, due to its broad buffering capacity (Fig. 4B).

Effect of Mg^{2+} concentration The highest enzyme activity was observed at 30 mM Mg^{2+}. The activity was more than three times higher than the no Mg^{2+} control, suggesting the divalent cation is important for the enzyme activity. It decreased steadily at higher concentration (Fig. 4C).

Effect of sodium cholate Sodium cholate, an ionic detergent used for cell lysis and membrane protein isolation, strongly inhibited the enzyme activity at concentration higher than 0.16%. The control with no sodium cholate was not tested because of the minimum requirement (0.08%) of sodium cholate for the enzyme preparation. The following enzyme activity experiments were conducted at 0.08% of sodium cholate (Fig. 4D).

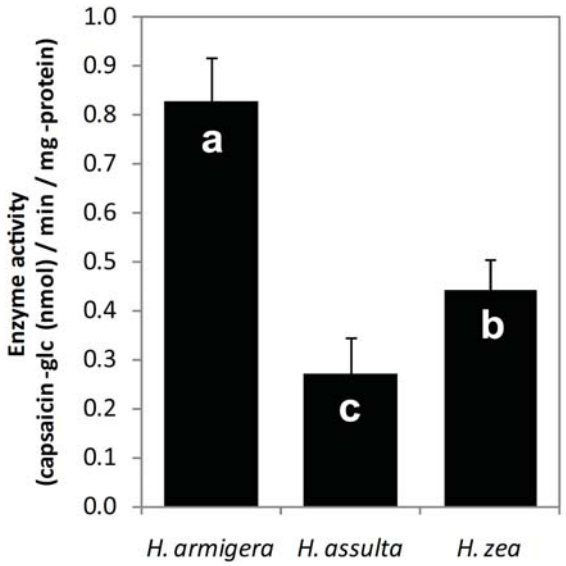


Figure 5. Capsaicin:UGT activity in preparations from the fifth instar larval homogenates (*N*=3) from *H. armigera*, *H. assulta*, and *H. zea*. The whole body was individually homogenated.

Species and Tissue Differences of the Enzyme Activity

Capsaicin UDP-glycosyltransferase (UGT) enzyme activities were compared among five different larval tissues from three *Helicoverpa* spp. The activity in whole-larva homogenates was lowest in *H. assulta* and highest in *H. armigera* (Fig. 5). Among five different larval tissues, fat body (FB) was the tissue showing the highest enzyme activity per total protein in *H. armigera*, followed by the Malpighian tubules (MT), testis (TS), midgut (MG), and labial glands (LG) (Fig. 6). FB was the main

source of the activity among the tissues also in *H. assulta*, and MT contained as high activity as FB, whereas the MG had almost negligibly low activity. Similarly, *H. zea* contained high activity in FB, whereas relatively low activities were detected in the MT, TS, and LG. It is noteworthy that the enzyme activity difference in the whole larva homogenate from the three species (Fig. 5) seems to be similar to those of the MG preparation that reflects virtually most of the protein source of the larval body (Fig. 6). Furthermore, the species difference of the *in vitro* activity is consistent with the *in vivo* glucosylation activity in the feeding bioassay with the three species (Fig. 3A).

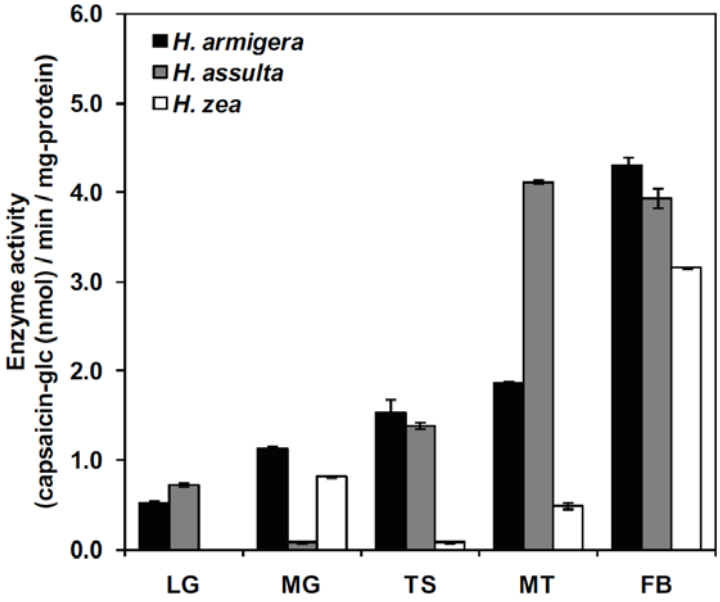


Figure 6. Capsaicin:UGT activity in different tissues of the fifth instar larvae of *H. armigera*, *H. assulta*, and *H. zea* (LG; labial glands, MG; midgut, TS; testis, MT; Malpighian tubules, FB; fat body). Each tissue collected from15 larvae was pooled and homogenated, being used as one biological replicate. Each point represents a mean value of three biological replicates.

DISCUSSION

General discussion

We described the identification of a novel metabolite of capsaicin, capsaicin β-glucoside (**2**), in the feces of three *Helicoverpa* spp. larvae that had been fed on capsaicin-spiked artificial diet. To our knowledge, this is the first report on the identification of any capsaicin metabolite in insects, and the first capsaicin glucoside in insects. We also found species-differential glucosylation activity; *H. assulta* conjugated capsaicin with glucose less readily than *H. armigera* and *H. zea*, which is consistent with UGT enzyme activity toward capsaicin from the whole larva homogenate *in vitro*. Fat body compared to the other four tissues was the main source of the enzyme activity in the three species consistently, but species-differential tissue distribution of the activity was also observed. Our study suggests that the putative detoxification mechanism of the capsaicin glucosylation is likely to be catalyzed by UGT, and that the activity seems to be stronger in the two generalists than the host-specialist, *H. assulta*.

Capsaicin glucoside in other species

In the rat, ingested capsaicin is conjugated with glucuronic acid to produce capsaicin glucuronide in urine (Noami et al., 2006). In *Capsicum annuum*, capsaicin glucoside is present in the fruit, and its content is positively correlated with the aglycone content, although the former amount is 2,000 times less than the latter (Higashiguchi et al., 2006). Capsaicin glucoside can be produced by cultured plant cells of e.g. *Catharanthus roseus* with supplemented capsaicin in medium (Shimoda et al., 2007), although such a biotransformation has been investigated for the purpose of enhancement of bioavailability and drug-like properties of capsaicinoids. Capsaicin glucoside appears to be more water soluble than its aglycone, thereby more easily excreted. Furthermore, it may be less toxic and irritant. In fact, capsaicin glucuronide is approximately 100 times less pungent than the aglycone in panel tests with humans (Kometani et al., 1993). Although capsaicin glucoside is not yet known to be less toxic to insects than capsaicin, the glucoside is more water soluble, because it showed up earlier than the aglycone in the reverse-phase chromatogram (Fig. 2). This suggests that the metabolite may be a detoxification product in insects.

Different conjugation rate of capsaicin in the three species in vivo

Since the amounts of capsaicin ingested were not significantly different among the three species (data not shown), the different excretion rate of the unaltered capsaicin might be caused by post-ingestive metabolic capacity toward capsaicin. Only a small portion of ingested capsaicin was recovered in the feces of *H. assulta* compared to the other two species (Fig. 3B), suggesting higher metabolism of

capsaicin in the host-specialist. In contrast, the two generalists interacting less often with capsaicin in nature excrete more unaltered capsaicin, suggesting that they have a lower overall detoxicative ability.

However, the glucose conjugation rate of capsaicin was lower in *H. assulta* than the other two species (Fig. 3A), suggesting that the glucosylation might not be a major mechanism of dealing with capsaicin in the specialist. Rather, the generalists seem to use glucosylation more actively than the specialist when they deal with capsaicin. However, less than 10% of capsaicin was conjugated with glucose in the generalists. Therefore, capsaicin metabolism in these insects cannot be completely explained by glucose conjugation, but other detoxification mechanisms must exist. Such a question can be tackled by investigating the metabolic fate of the remaining portion of capsaicin ingested. In fact, capsaicin is known to be oxidized into hydroxyl capsaicin, capsaicin oxide, or capsaicin quinone by P450s in mammals (Chanda et al., 2008; Reilly and Yost, 2006; Surh and Lee, 1995) and it can be also hydrolyzed into vanillylamine and 8-methyl-6-*trans*-nonenoic acid by carboxylesterase (Chanda et al.,, 2008). The vanillylamine can be further modified to vanillin, vanillyl alcohol, or vanillic acid, and each of them can form a conjugate further. Bacterial strains isolated from the hot pepper plants are capable of degrading capsaicin as a carbon and energy source, or utilizing its hydrolyzed vanillylamine as a nitrogen source (Flagan and Leadbetter, 2006). In the hot pepper, capsaicin is oxidized to produce capsaicin dimers by peroxidases (Díaz et al., 2004). Therefore, it is possible also in insects that other mechanisms underlying in such a high capsaicin processing capability could be revealed.

It is noteworthy that a small but significant amount of capsaicin was retained in the *H. assulta* larval body, whereas it was not detected in *H. zea* and was negligible in *H. armigera* (data not shown). Although it is not known whether the retained capsaicin in the *H. assulta* larva is found in the gut content or other tissues, it seems that capsaicin is being processed or sequestered in the body of *H. assulta* rather than rapidly excreted. A further experiment to trace capsaicin in different larval tissues might be necessary clarify this issue.

In vitro UGT activity: tissue distribution and species comparison

UGT enzyme activity was detected in a wide range of tissues. The highest activity per total protein was associated with FB, MT and MG. FB contained relatively higher activity in the three species (Fig. 6), which is consistent in other insects where FB is known as the major source of UGT activity toward a variety of compounds including plant allelochemicals and endogenous compounds (Ahmad and Hopkins, 1992) (Ahmad and Hopkins, 1993; Ahmad et al., 1996). It is interesting to note that the MT of *H. assulta* contained extraordinarily high activity, suggesting the species-specific high expression of the host toxin metabolizing enzyme in a particular tissue, although the overall enzyme activity to capsaicin was very low in *H. assulta* (Fig. 5). Since MG accounts for the majority of enzyme activity in terms of protein content, it appears to reflect the overall activity. On the basis of this assumption, the two generalists seem to be able to detoxify most of capsaicin before it passes through in the gut;

whereas the host-specialist lacks the primary defensive ability in its gut. More capsaicin could be moved to the hemocoel due to its hydrophobicity, where MT is probably more responsible for the capsaicin glucosylation than the other generalists. Alternatively, capsaicin in the gut of *H. assulta* might be metabolized by unknown detoxification enzymes, for example P450.

In addition, it is also noteworthy that testes contain moderate levels of the enzyme activity in *H. armigera* and *H. assulta* (Fig. 6). Although the activity was measured with capsaicin, other substrates which are physiologically relevant could be conjugated in this particular tissue due to the substrate binding redundancy in N-terminal domain. Since ecdysteroidogenesis is known to occur also in insect testes (Brown et al., 2009), it is possible that UGT enzyme(s) expressed in testes might be involved in the biosynthesis or regulation of ecdysteroids in these insects tested. Of the three *Helicoverpa* species, *H. zea* consumes *Capsicum* fruits much less often, and shows lower UGT activity in all the tissues except MG, compared to the other two *Capsicum*-feeding species (Fig. 6).

Perspectives

UGT-mediated enzymatic detoxification is relatively unknown in insects, compared to other detoxification enzymes like P450 and GST. It is, however, regarded as an important component of metabolic detoxification in insects, playing crucial roles in cuticular tanning (Hopkins and Kramer, 1992), pigmentation (Hopkins and Ahmad, 1991; Mizokami and Yoshitama, 2009; Wiesen et al.,, 1994), as well as detoxification. UGTs are encoded by a multigene family, which has been recently analyzed in several insects from which genomic information is available (Huang et al., 2008; Luque and O'Reilly, 2002). Further studies are necessary not only to determine which gene(s) is involved in the glucosylation of capsaicin, but also to understand its regulatory mechanism, in relation to the host-plant adaptation of herbivores.

ACKNOWLEDGEMENTS

This work was supported by the Max-Planck-Gesellschaft and the International Max Planck Research School (IMPRS).

LITERATURE CITED

Ahmad SA, Hopkins TL. 1992. Phenol β-glucosyltransferase and β-glucosidase activities in the tobacco hornworm larva *Manduca sexta* (L.): Properties and tissue localization. Arch Insect Biochem Physiol 21: 207-224.

Ahmad SA, Hopkins TL. 1993. β-Glucosylation of plant phenolics by phenol β-glucosyltransferase in larval tissues of the tobacco hornworm, *Manduca sexta* (L.). Insect Biochem Mol Biol 23: 581-589.

Ahmad SA, Hopkins TL, Kramer KJ. 1996. Tyrosine β-glucosyltransferase in the tobacco hornworm, *Manduca sexta* (L): Properties, tissue localization, and developmental profile. Insect Biochem Mol Biol 26: 49-57.

Ahn S-J, Badenes-Pérez FR, Heckel DG. 2011. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. J Insect Physiol doi: 10.1016/j.jinsphys.2011.05.015.

Baek S, Cho K, Song YH, Lee J-H. 2009. Sampling plans for estimating pepper fruit damage levels by Oriental tobacco budworm, *Helicoverpa assulta* (Guenée), in hot pepper fields. J Asia-Pacific Entomol 12: 175-178.

Brattsten LB. 1992. Metabolic defenses against plant allelochemicals. In: Rosenthal GA, Berenbaum MR, editors. Herbivores: Their interactions with secondary plant metabolites, Second edition, Vol II: Ecological and evolutionary processes: Academic Press, Inc.; Academic Press Ltd. p 175-242.

Brown, M.R., Sieglaff, D.H., Rees, H.H., 2009. Gonadal ecdysteroidogenesis in arthropoda: Occurrence and regulation. Annu Rev Entomol 54: 105-125.

Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD, Julius D. 1997. The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature 389: 816-824.

Chanda S, Bashir M, Babbar S, Koganti A, Bley K. 2008. In vitro hepatic and skin metabolism of capsaicin. Drug Metab Dispos 36: 670-675.

Cho S, Mitchell A, Mitter C, Regier J, Matthews M, Robertson RON. 2008. Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. Syst Entomol 33: 581-594.

Choi KS, Boo KS. 1989. Food attractancy of the oriental tobacco budworm, *Heliothis assulta*. Korean J Appl Entomol 28: 88-92.

Cowles RS, Keller JE, Miller JR. 1989. Pungent spices, ground red pepper, and synthetic capsaicin as onion fly ovipositional deterrents. J Chem Ecol 15: 719-730.

Díaz J, Pomar F, Bernal A, Merino F. 2004. Peroxidases and the metabolism of capsaicin in *Capsicum annuum* L. Phytochemistry Reviews 3: 141-157.

Després L, David JP, Gallet C. 2007. The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol Evol 22: 298-307.

Edelson JV, Duthie J, Roberts W. 2002. Toxicity of biorational insecticides: activity against the green peach aphid, *Myzus persicae* (Sulzer). Pest Manag Sci 58: 255-260.

Flagan SF, Leadbetter JR. 2006. Utilization of capsaicin and vanillylamine as growth substrates by *Capsicum* (hot pepper)-associated bacteria. Environ Microbiol 8: 560-565.

Han M-W, Lee J-H, Son J-S. 1994. Intra- and inter-plant distribution of *Helicoverpa assulta* (Lepidoptera: Noctuidae) eggs in red pepper and tobacco field. Korean J Appl Entomol 33: 6-11.

Higashiguchi F, Nakamura H, Hayashi H, Kometani T. 2006. Purification and structure determination of glucosides of capsaicin and dihydrocapsaicin from various Capsicum fruits. J Agric Food Chem 54: 5948-5953.

Hopkins TL, Ahmad SA. 1991. Flavonoid wing pigments in grasshoppers. Cellular and Molecular Life Sciences 47: 1089-1091.

Hopkins TL, Kramer KJ. 1992. Insect cuticle sclerotization. Annu Rev Entomol 37: 273-302.

Hori M, Nakamura H, Fujii Y, Suzuki Y, Matsuda K. 2011. Chemicals affecting the feeding preference of the Solanaceae-feeding lady beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae). J Appl Entomol 135: 121-131.

Huang FF, Chai CL, Zhang Z, Liu ZH, Dai FY, Lu C, Xiang ZH. 2008. The UDP-glucosyltransferase multigene family in *Bombyx mori*. BMC Genomics 9: 563.

Kometani T, Tanimoto H, Nishimura T, Kanbara I, Okada S. 1993. Glucosylation of capsaicin by cell suspension cultures of *Coffea arabica*. Biosci Biotech Bioch 57: 2192-2193.

Larue E, Welty C. 2010. Oviposition behavior and larval development of the European corn borer (Lepidoptera: Crambidae) on sweet versus hot peppers. J Entomol Sci 45: 353-365.

Lin LZ, West DP, Cordell G. 1993. NMR assignments of *cis*- and *trans*-capsaicin. Nat Prod Lett 3: 5-8.

Luque T, O'Reilly DR. 2002. Functional and phylogenetic analyses of a putative *Drosophila melanogaster* UDP-glycosyltransferase gene. Insect Biochem Mol Biol 32: 1597-1604.

Madhumathy AP, Aivazi A-A, Vijayan VA. 2007. Larvicidal efficacy of *Capsicum annum* against *Anopheles stephensi* and *Culex quinquefasciatus*. J Vector Borne Diseases 44: 223-226.

Matthews M. 1999. Heliothine moths of Australia. Collingwood: CSIRO Publishing.

Mitter C, Poole RW, Matthews M. 1993. Biosystematics of the Heliothinae (Lepidoptera: Noctuidae). Annu Rev Entomol 38: 207-225.

Mizokami H, Yoshitama K. 2009. Sequestration and metabolism of host-plant flavonoids by the Pale Grass Blue, Pseudozizeeria maha (Lepidoptera: Lycaenidae). Entomol Science 12: 171-176.

Noami M, Igarashi K, Kasuya F, Ohta M, Kanamori-Kataoka M, Seto Y. 2006. Study on the analysis of capsaicin glucuronide in rat urine by liquid chromatography-mass spectrometry after enzymatic hydrolysis. J Health Sci 52: 660-665.

Reilly CA, Yost GS. 2006. Metabolism of capsaicinoids by P450 enzymes: a review of recent findings on reaction mechanisms, bio-activation, and detoxification processes. Drug Metabol Reviews 38: 685-706.

Robertson HM, Martos R, Sears CR, Todres EZ, Walden KKO, Nardi JB. 1999. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. Insect Mol Biol 8: 501-518.

Schoonhoven LM, Loon JJAv, Dicke M. 2005. Insect-Plant Biology: Oxford University Press. 440 p.

Shimoda K, Kwon S, Utsuki A, Ohiwa S, Katsuragi H, Yonemoto N, Hamada H, Hamada H. 2007. Glycosylation of capsaicin and 8-nordihydrocapsaicin by cultured cells of *Catharanthus roseus*. Phytochem 68: 1391-1396.

Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. 1985. Measurement of protein using bicinchoninic acid. Anal Biochem 150: 76-85.

Surh Y-J, Lee SS. 1995. Capsaicin, a double-edged sword: Toxicity, metabolism, and chemopreventive potential. Life Sci 56: 1845-1855.

Tewksbury JJ, Manchego C, Haak DC, Levey DJ. 2006. Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. J Chem Ecol 32: 547-564.

Tewksbury JJ, Reagan KM, Machnicki NJ, Carlo TA, Haak DC, Peñaloza ALC, Levey DJ. 2008. Evolutionary ecology of pungency in wild chilies. Proc Natl Acad Sci USA 105: 11808-11811.

Wang Q, Hasan G, Pikielny CW. 1999. Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. J Biol Chem 274: 10309-10315.

Weissenberg M, Klein M, Meisner J, Ascher KRS. 1986. Larval growth inhibition of the spiny bollworm, *Earias insulana*, by some steroidal secondary plant compounds. Entomol Exp Appl 42: 213-217.

Wiesen B, Krug E, Fiedler K, Wray V, Proksch P. 1994. Sequestration of host-plant-derived flavonoids by lycaenid butterfly *Polyommatus icarus*. J Chem Ecol 20: 2523-2538.

Wilkinson CF. 1986. Xenobiotic Conjugation in Insects. Xenobiotic Conjugation Chemistry: American Chemical Society. p 48-61.

Yang C-Y, Jeon H-Y, Cho M-R, Kim D-S, Yiem M-S. 2004. Seasonal occurrence of oritental tobacco budworm (Lepidoptera: Noctuidae) male and chemical control at red pepper fields. Korean J Appl Entomol 43: 49-54.

3.3. Manuscript III

Comparative analysis of the UDP-glycosyltransferase multigene family in insects

Seung-Joon Ahn, Heiko Vogel and David G. Heckel

submitted to *Insect Biochemistry and Molecular Biology* (June 30, 2011)

Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, 07745, Germany

*Corresponding author. Tel.: +49 3641 571501; fax: +49 3641 571502. *E-mail address*: heckel@ice.mpg.de (D.G. Heckel)

ABSTRACT

UDP-glycosyltransferases (UGT) catalyze the conjugation of a range of diverse small lipophilic compounds with sugars to produce glycosides, playing an important role in the detoxification of xenobiotics and in the regulation of endobiotics in insects. Recent progress in genome sequencing has revealed the extent of the UGT multigene family in insects. Here we report over 310 putative UGT genes identified from genomic databases of eight different insect species together with a transcript database from the lepidopteran *Helicoverpa armigera*. Phylogenetic analysis of the insect UGTs showed Order-specific gene diversification and inter-species conservation of this multigene family. A sub-tree constructed with 40 *H. armigera* UGTs and 44 *Bombyx mori* UGTs revealed that lineage-specific expansions of some families in both species appears to be driven by diversification in the N-terminal substrate binding domain, increasing the range of compounds that could be detoxified or regulated by glycosylation. By comparison of the deduced protein sequences, several important domains were predicted, including the N-terminal signal peptide, UGT signature motif, and C-terminal transmembrane domain. Furthermore, several conserved residues putatively involved in sugar donor binding and catalytic mechanism were also identified by comparison with human UGTs. The comparative analyses of UGTs from two lepidopteran species demonstrated the conservation as well as diversification dynamics of the multigene family. This study also provides primary information for functional and evolutionary studies on insect UGTs in the future.

Keywords: UDP-glycosyltransferase, *Helicoverpa armigera*, *Bombyx mori*, multigene family, phylogenetic analysis, detoxification

1. Introduction

Glycoside conjugation is one of the most important metabolic pathways for biotransformation of a number of lipophilic xenobiotics and endobiotics. UDP-glycosyltransferases (UGTs) catalyze the conjugation of a sugar donated by a UDP-glycoside to a typically lipophilic molecule, generating water-soluble products which can be easily excreted as well as stably managed, therefore protecting the cellular system from being damaged by toxic foreign compounds and regulating internal molecules more easily (Bock, 2003).

UGTs are membrane-bound proteins located in the endoplasmic reticulum (ER) facing the lumen in animals. The UGT protein structure is divided into two main parts: the N-terminal aglycone substrate binding domain and the C-terminal UDP-glycoside binding domain. The N-terminal end of the animal UGTs has a signal peptide mediating the integration of the protein precursor into the ER compartment. The signal peptide is subsequently cleaved and then the protein is *N*-glycosylated. The mature protein is retained in the ER membrane by its hydrophobic transmembrane domain at the C-terminal end, followed by a short cytoplasmic tail (Magdalou et al., 2010).

UGTs are ubiquitous in all living organisms from virus to bacteria, plants and animals. In plants, a variety of soluble cytosolic UGTs play an important role in the modification of secondary metabolites, thereby enhancing their solubility and stability, and determining their bioactivity (Bowles et al., 2005). In vertebrates, membrane-bound UGTs are regarded as a major member of the phase II drug metabolizing enzymes, conjugating a large number of xenobiotics as well as endobiotics, such as bilirubin and steroid hormones with UDP-glucuronic acid (Bock, 2003). In insects, the glycosylation of small lipophilic compounds has been considered as a minor enzymatic detoxification mechanisms for half a century (Ahmad et al., 1986; Brattsten, 1988; Després et al., 2007; Smith, 1962). Insect UGT enzyme activity has been investigated in the housefly *Musca domestica* (Morello and Repetto, 1979), in the fruitfly *Drosophila melanogaster* (Real et al., 1991), in the tobacco hornworm *Manduca sexta* (Ahmad and Hopkins, 1992), in the silkworm *Bombyx mori* (Luque et al., 2002), and in other insects (Ahmad and Hopkins, 1993b). These biochemical studies have shown that the insect UGT enzymes typically use UDP-glucose as the main sugar donor unlike vertebrate UGTs, but are probably bound to the endoplasmic reticulum in a similar manner. Enzyme activities of the insect UGTs are detected in the fat body, midgut and other tissues (Ahmad and Hopkins, 1993b), and are directed towards a variety of plant allelochemicals (Ahmad and Hopkins, 1993a; Luque et al., 2002; Sasai et al., 2009). Interestingly, the enzymes were also detected in the antenna of *D. melanogaster*, but their function has not been characterized yet (Wang et al., 1999). In addition, many endogenous compounds, like ecdysteroid hormones (Svoboda and Weirich, 1995) and cuticle tanning precursors (Ahmad et al., 1996; Hopkins and Kramer, 1992) are glycosylated by the UGT enzymes. Furthermore, in a lycaenid butterfly dietary flavonoids have been shown to be sequestered as a glucose conjugate to impart color to the wings (Wiesen et al., 1994) or in *B. mori* be glycosylated to make green cocoon

color with UV-shielding properties (Daimon et al., 2010). These findings suggest multiple roles of the insect UGT enzymes in detoxification, olfaction, endobiotic modulation, and sequestration.

At the time of this manuscript preparation, there are 33 insect species of which genome sequences have been either completed or at least assembled (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi). With this information, many gene families involved in detoxification and metabolism in insects, like cytochrome P450s (P450s), carboxyl/cholinesterases (CCEs), and glutathione transferases (GSTs) have been identified from genome sequences of *Drosophila melanogaster* (Tijet et al., 2001), *Aedes aegypti* (Strode et al., 2008), *Acyrthosiphon pisum* (Ramsey et al., 2010), *Nasonia vitripennis* (Oakeshott et al., 2010), *Apis mellifera* (Claudianos et al., 2006), and *Bombyx mori* (Tsubota and Shiotsuki, 2010; Yu et al., 2008). The molecular identities of the insect UGTs are, however, relatively unknown compared to the other detoxification gene families in insects or to the vertebrate UGT gene families. Although the UGT gene families in *D. melanogaster* (Luque and O'Reilly, 2002) and in *Bombyx mori* (Huang et al., 2008) have been reported, they have not been compared in the context of the diversity of other insects.

The cotton bollworm *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae) is a polyphagous herbivore feeding more than 100 different species belonging to about 46 different plant families, and it is regarded as a serious agricultural pest in many regions in the world (Fitt, 1989). Thus it has been extensively studied in order to understand its ecology and physiology especially in relation to host plant adaptation and insecticide resistance (Heckel, 2010). Recently the *H. armigera* CCE gene family was described based on transcriptome analysis (Teese et al., 2010), which is likely incomplete based on comparison to the *Bombyx* genome; no other gene families related to detoxification have yet been described from this species.

In this study, we identified all the putative UGT genes from our *H. armigera* cDNA libraries and from *B. mori* genome databases, as well as from another seven insect genomes available in NCBI. Here we describe the genomic and phylogenetic analyses of the *H. armigera* and *B. mori* UGTs in the context of the insect UGT diversity, and make predictions of the UGT protein structure by comparing several important domains and residues with biochemically characterized human UGT sequences.

2. Materials and methods

2.1. UGT gene sequences from *H. armigera*

Putative UGT gene sequences were identified from a public database (NCBI GenBank) and from a locally generated EST database together with 454-pyrosequencing data of *H. armigera*. The *H. armigera* cDNA libraries in the Department of Entomology, Max Planck Institute for Chemical Ecology were produced from adult, pupa or various larval tissues (gut, fat body, salivary glands, hemocytes, integument, or rest of body) of TWB strain originally from Toowoomba, Queensland, Australia. In total, 353 cDNA contigs were retrieved by tBLASTn searches, and the collected sequences were assembled by using Sequencher (Gene Codes Corporation, MI, USA). To obtain full-length sequences, RNA Ligase-Mediated Rapid Amplification of cDNA Ends (RLM-RACE) was carried out employing GeneRacer Kit (Invitrogen) according to the manufacturer's instructions. Gene-specific primers were designed from the partial cDNA sequences assembled. Total RNA was extracted from a fifth instar larva using Trizol (Invitrogen) followed by purification with an RNAeasy kit (Qiagen). The RACE products were analyzed by 1.5% agarose gel electrophoresis, the excised bands were purified using QIAquick spin columns (QIAGEN) and sequenced by DNA analyzer (Applied Biosystems, USA).

2.2. BLAST searches of insect UGT sequences from public databases

The *B. mori* UGT sequences were collected from KAIKObase (http://sgp.dna.affrc.go.jp/KAIKObase) and SilkDB blast (http://www.silkdb.org) based on the gene names on a previous report (Huang et al., 2008), which however did not give the actual sequences. BLAST searches of unannotated genomic contigs in NCBI were also performed to obtain additional sequences. Four UGT sequences reported from the Six-spot burnet moth *Zygaena filipendulae* were also included (Zagrobelny et al., 2009). In addition, non-lepidopteran UGT sequences were retrieved from seven representative insect species, of which genome sequences have been either completed or at least assembled. Among them are *Drosophila melanogaster, Anopheles gambiae, Aedes aegypti, Acyrthosiphon pisum, Apis mellifera, Tribolium castaneum*, and *Nasonia vitripennis.* BLAST searches were performed against assembled RefSeq as well as against WGS database from each genome. Individual genes were manually annotated using genomic sequences together with EST sequences. When no cDNA sequences were available, the open reading frame was predicted from genomic sequence by comparison with known UGT sequences. In addition, UGT sequences from several baculoviruses were obtained by BLAST searches of NCBI.

2.3. Nomenclature

According to the UGT nomenclature guidelines (Mackenzie et al., 1997; Ross et al., 2001), families were defined at 40% amino acid sequence identity (aaID) or greater and subfamilies defined at 60% aaID or greater. Multiple sequence alignment was performed by using CLUSTAL and adjusted manually. Preliminary grouping was done using the program H-CD-HIT at 60% and 40% sequence identity as cutoff values, and preliminary family and subfamily names were assigned on this basis. A neighbor-joining tree was constructed from the sequence alignment using MEGA5 and plotted using the preliminary names. Groups were examined for consistency, and groups on the borderline of 40% or 60% were examined using pairwise p-distances calculated by MEGA5 (Tamura et al., 2011). In a few cases the family criterion of 40% was difficult to apply due to some pairwise comparisons being 41-42% while others were 38-39%, and the family criterion was relaxed to 37-39% if doing so created a coherent group on the neighbor-joining tree. Preliminary names were re-assigned and the entire process was repeated. Partial sequences were examined to ensure that they were not incorrectly grouped. The nomenclature was submitted to the UGT Nomenclature Committee (Mackenzie et al., 1997) for approval, which is still pending (i.e names used in this version of the manuscript are provisional).

2.4. Phylogenetic analysis

Predicted amino acid sequences of over 320 insect UGT sequences identified from nine insect species were aligned by CLUSTAL (Thompson et al., 1997) using MEGA5. The alignment was manually edited to minimize gaps. A consensus phylogenetic tree was constructed using the neighbor-joining method. Distance calculations were performed after Tajima & Nei and bootstrap analysis, running 1000 bootstrap samples (Felsenstein, 2004). For the comparative study in lepidopteran UGTs, a sub-tree was reconstructed with the *H. armigera* and *B. mori* UGTs together with four *Z. filipendulae* UGTs, following the methods mentioned above.

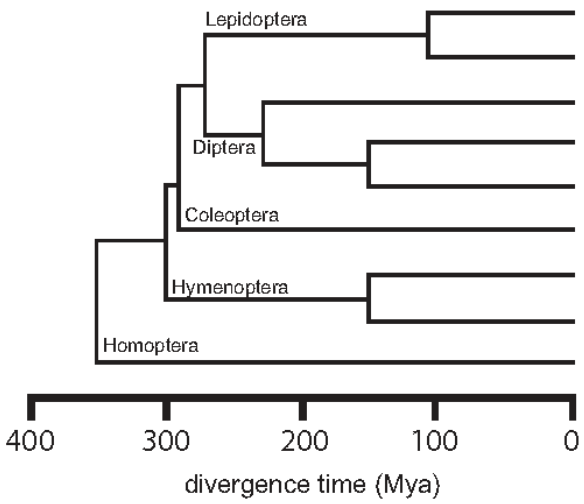
2.5. Protein structure prediction

For the overall homology comparison, ten representative protein sequences from the *H. armigera* UGTs were aligned with four human UGTs by the Clustal X algorithm using Geneious software. Signal peptides were predicted by SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP).

3. Results and discussion

3.1. Number of UGT genes in insects

Using the amino acid sequences of *D. melanogaster* and *B. mori* UGTs as queries, we identified 42 putative UGT sequences from our *H. armigera* cDNA libraries, and using RACE recovered 40 full length UGT sequences and two partial sequences from *H. armigera* (Table 1). This number is similar to that of *B. mori* UGTs (45 UGTs) and *T. castaneum* UGTs (43 UGTs), but larger than that of three Diptera and two Hymenoptera (Fig. 1). A hemipteran species, the pea aphid (*Acyrthosiphon pisum*), contained the largest UGT repertoire (58 UGTs) among insect genomes compared in this study, more than expected compared with other detoxification enzyme families. For comparison, the numbers of P450s (83), CCEs (29), and GSTs (20) in *Ac. pisum* (Ramsey et al., 2010) are not more than those in *D. melanogaster* (85 P450s, 35 CCEs, 38 GSTs) (Claudianos et al., 2006). On the other hand, the lower number in the honeybee (12 UGTs) is consistent with a deficit of the other xenobiotic detoxifying enzymes in this species. The honeybee genome contains only 46 P450s, 24 CCEs, and 10 GSTs. The degree of gene deficiency in phase II enzymes, like GSTs and UGTs, is higher than for phase I enzymes, like P450s and CCEs in the honeybee. The highly specialized living environment, lack of exposure to food-derived toxins and peculiar life history of the honeybee might be related to the decreased number of detoxification genes (Honeybee Genome Sequencing Consortium, 2006).



Insect	No. of UGT genes
Helicoverpa armigera	42*
Bombyx mori	45
Drosophila melanogaster	34
Anopheles gambiae	26
Aedes aegypti	34
Tribolium castaneum	43
Apis mellifera	12
Nasonia vitripennis	22
Acyrthosiphon pisum	58
Total	**316**

Figure 1. Species tree and number of insect UGTs. The species tree was modified from Honeybee Genome Sequencing Consortium (2006). The UGT gene numbers were identified based on genome-wide surveillance, except *H. armigera* UGTs (*) which were identified

Table 1. Summary of *H. armigera* UGT sequences and tissue distribution of transcript origin.

Name	Length (aa)	No. transcripts	Tissue						
			MG	FB	LG	HC	CT	RB	PA
UGT33B1	529	9		x					
UGT33B2	505	7						x	
UGT33B3	513	10		x					
UGT33B4	512	23						x	
UGT33B5	513	1						x	x
UGT33B7	512	2						x	
UGT33B8	511	4						x	x
UGT33B9	513	2						x	
UGT33F1	520	7						x	
UGT33F3	519	2						x	
UGT33S1	520	7						x	
UGT33J1	519	8					x		x
UGT33M1	517	10	x	x	x				x
UGT33T1	524	5						x	
UGT33B11	513	1						x	
UGT33B12	513	4						x	
UGT34A3	525	16	x	x	x				
UGT39B2	525	10	x	x					
HarUGT_43	(135)	2						x	
UGT40D1	521	23	x	x					
UGT40D2	521	5	x						
UGT40F1	516	11	x						x
UGT40F2	520	13		x	x				x
UGT40L1	520	9				x			
UGT40M1	519	17	x	x					
UGT40Q1	518	17	x	x		x	x		
UGT40R1	519	16	x	x			x		
HarUGT_28	(215)	1							x
UGT41B1	513	16		x	x	x			
UGT41B2	516	11	x						
UGT41B3	513	4						x	
UGT41D1	519	25	x	x	x		x		x
UGT42B2	521	7	x	x					
UGT42C1	509	3						x	
UGT43A1	520	13			x	x			
UGT44B1	526	6		x					
UGT46A3	527	11		x	x				x
UGT46A4	527	16	x	x	x		x		
UGT46B1	515	1						x	
UGT47A2	535	21	x	x			x		
UGT48A1	520	1						x	
UGT50A2	520	1						x	
Total 42		378	14	17	8	4	6	18	9

( ) indicates partial sequences. MG, midgut; FB, fatbody; LG, labial glands; HC, hemocytes; CT, cuticle; RB, rest of body; PA, pupa and adult.

3.1.2. *B. mori* UGTs

We identified 45 putative UGT genes in the partially annotated genome of *B. mori* (Table 2). A recent study reported 42 UGTs in this species but did not provide the sequences (Huang et al., 2008). Our blast searches and manual annotation efforts improved the quality of the BGI annotations and provided three additional UGT genes (BGIBMGA010432, KAIKOGA083789, and BGIBMGA009787P). Among 45 *B. mori* UGTs, there are three partial *B. mori* UGT genes (BGIBMGA010098, KAIKOGA083789, and BGIBMGA013858) due to incompleteness in the genome assembly. There are five putative pseudogenes. BGIBMGA009787P is composed of an intact long exon 1, but it was truncated in the middle of exon 2 by a non-LTR transposable element. Furthermore, high nucleotide sequence similarity of exon 1 of BGIBMGA009787P and exon 1 of a flanking BGIBMGA009788P (96.4%) suggests BGIBMGA009787P has been recently duplicated from BGIBMGA009788P. BGIBMGA010099P is also disrupted by a non-LTR retrotransposon in the beginning of exon 1. Each of the other three pseudogenes (BGIBMGA009788P, BGIBMGA010287-2P, and BGIBMGA013836P) has a single inactivating point mutation, occurring in exon 3, start codon, and exon 1, respectively. Whether these are truly pseudogenes, or rather null alleles of genes that are functional in strains other than the Dazao strain that was sequenced, can only be decided when sequence of additional strains becomes available.

3.2. Nomenclature

Each of the putative UGT genes identified from *H. armigera* in this study has been classified according to the nomenclature guidelines recommended by the UGT Nomenclature Committee (http://www.flinders.edu.au/medicine/sites/clinical-pharmacology/ugt-homepage.cfm). In addition, all of the other insect UGTs retrieved from public databases were also named accordingly (Fig. 2). More than 310 insect UGTs were assigned to be categorized into families UGT33 - UGT50, and since UGT51 and higher are reserved for fungi, the series was extended by using UGT301 - UGT349. If more insect UGT family numbers are required in the future, it is recommended to use up to UGT499, and again from UGT3001. Finally, the previously utilized UGT31 family from nucleopolydrosis viruses was augmented by the addition of UGT32 from granuloviruses. These viruses infect Lepidoptera and their UGTs are most similar to the UGT33 family. suggesting that a common ancestor of these viruses obtained the coding sequence from a lepidopteran host.

Table 2. Summary of *B. mori* UGT sequences.

Name	BGI number	Length (aa)	No. exons	Chr.
UGT33D1	BGIBMGA013830*	513	4	28
UGT33D2	BGIBMGA013831*	515	4	28
UGT33D3	BGIBMGA013833*	515	4	28
UGT33D4	BGIBMGA013859	520	4	28
UGT33D5	BGIBMGA013860-1*	520	4	28
UGT33D6	BGIBMGA013860-2*	515	4	28
UGT33D7	BGIBMGA013861*	515	4	28
UGT33D8	BGIBMGA013829	514	4	28
UGT33K1	BGIBMGA013836-2P*	516	4	28
UGT33N1	BGIBMGA013836-1*	519	4	28
UGT33Q1	BGIBMGA013858*	(419)	4(3)	28
UGT33R1	BGIBMGA007327*	504	4	3
UGT33R2	BGIBMGA009788P*	510	4	2
BmUGT009787P*	BGIBMGA009787P*	(271)	4(2)	2
UGT340C1	BGIBMGA013834-2*	521	4	28
UGT340C2	BGIBMGA013834-1*	524	4	28
UGT34A2	BGIBMGA004965	525	4	25
UGT39B1	BGIBMGA005443*	520	4	8
UGT39C1	BGIBMGA005442*	525	4	8
UGT40A1	BGIBMGA010294	520	8	7
UGT40B1	BGIBMGA010098*	(499)	8(7)	7
UGT40B2	BGIBMGA010099-1*	518	8	7
UGT40B3	BGIBMGA010099-2P*	(474)	8(7)	7
UGT40B4	BGIBMGA010295*	518	8	7
UGT40G1	BGIBMGA010287-1	514	8	7
UGT40G2	BGIBMGA010287-2P	514	8	7
UGT40H1	BGIBMGA010289-1	516	8	7
UGT40K1	BGIBMGA010286	522	8	7
UGT40N1	BGIBMGA010100	519	8	7
UGT40P1	BGIBMGA010288	519	8	7
UGT40S1	BGIBMGA010289-2	516	8	7
UGT41A1	BGIBMGA001338*	518	9	24
UGT41A2	BGIBMGA003817*	517	9	24
UGT41A3	BGIBMGA003835*	516	9	24
UGT42A1	BGIBMGA008508-3*	512	4	18
UGT42A2	BGIBMGA014622*	509	4	Un.
UGT42B1	BGIBMGA008508-2*	508	4	18
UGT43B1	BGIBMGA008508-1*	516	4	18
UGT44A1	BGIBMGA008508-4*	525	4	18
UGT46A1	BGIBMGA010432*	527	4	12
UGT46A2	BGIBMGA010433*	525	4	12
UGT46C2	BGIBMGA083789*	(448)	4(3)	12
UGT47A1	BGIBMGA005046*	536	6	25
UGT48C1	BGIBMGA002854*	506	8	10
UGT50A1	BGIBMGA008381*	540	6	18
Total 45				

( ) indicates partial sequences, or partially identified exon numbers.

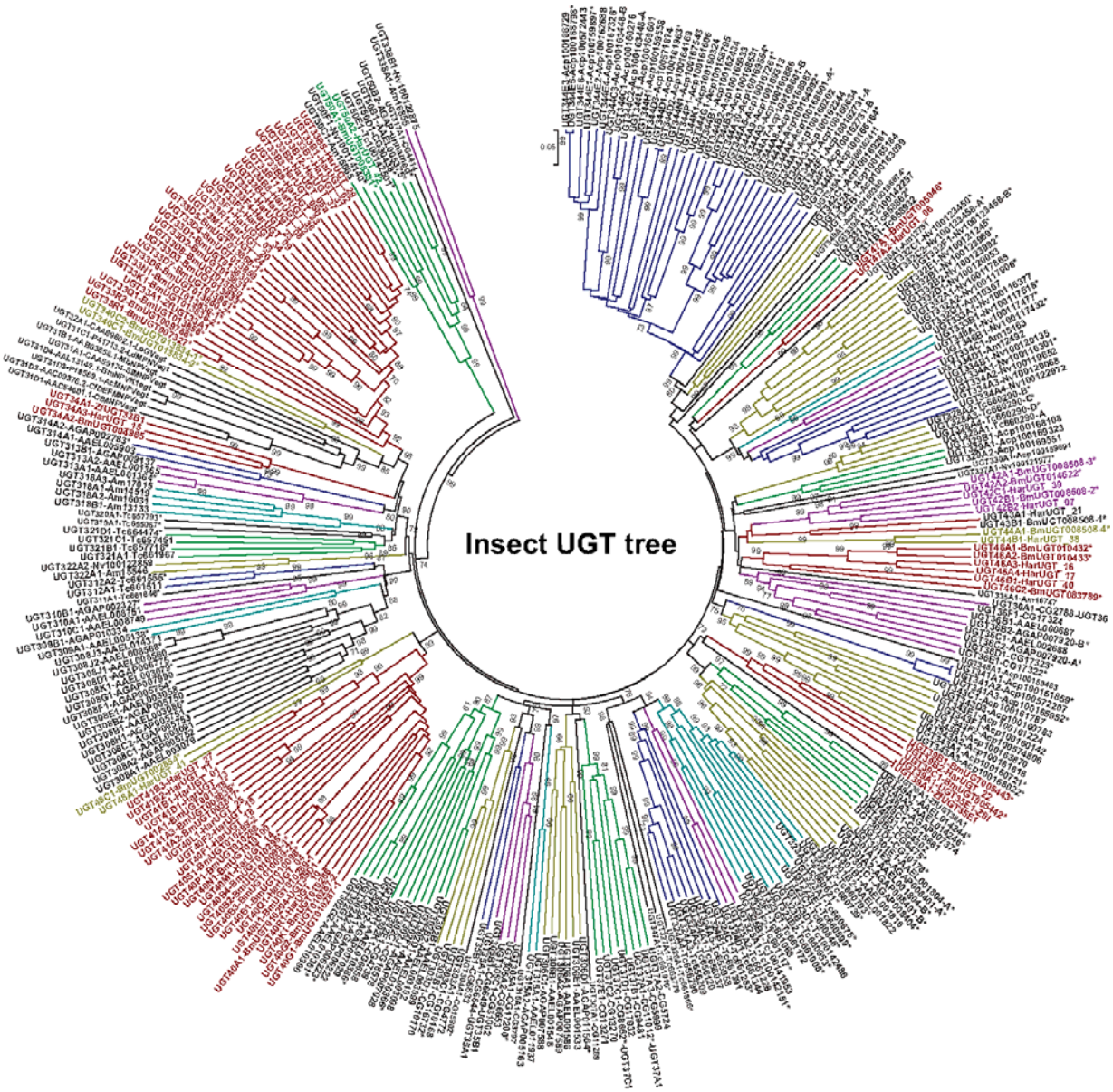


Figure 2.. Phylogenetic tree of insect UGTs. Multiple sequence alignment of more than 310 insect UGT protein sequences was performed by using CLUSTAL and adjusted manually to minimize gaps. A consensus phylogenetic tree was constructed using the neighbor-joining method. Distance calculations were performed after Tajima & Nei and bootstrap analysis. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates are collapsed. The evolutionary distances were computed using the p-distance method. Evolutionary analyses were conducted in MEGA5.

3.3. Phylogenetic analysis

3.3.1. Overall phylogenetic tree of insect UGTs

We have constructed an overall phylogenetic tree using over 310 UGTs from eight different insect genomes plus *H. armigera* transcript sequences (Fig. 2). The UGT families show a variety of phylogenetic patterns; from conserved families with one member in each species like UGT50 (a possible homolog of mammalian UGT8); to large taxon-specific families such as UGT33 in two species of Lepidoptera and UGT36 in three species of Diptera, and large lineage-specific expansions such as UGT344 in aphid (*Acyrthosiphon pisum*), and UGT324, 325, 326 in flour beetle (*Tribolium castaneum*). In addition, UGT332, 333, 334 and 336 separate together a Hymenoptera-specific group.

3.3.2. Comparative phylogenetic analysis between *H. armigera* and *B. mori* UGTs

A consensus phylogenetic tree constructed with deduced amino acid sequences from *H. armigera* (40 UGTs), *B. mori* (44 UGTs) and *Z. filipendulae* (4 UGTs) revealed patterns of inter-specific conservation and lineage-specific expansion of the gene families (Fig. 3). The largest UGT family (**UGT33**) containing 17 *H. armigera* UGTs and 14 *B. mori* UGTs (all but one on BmChr28, i.e. Chromosome 28) accounts for about 40 % and 31 % of all UGT numbers in each species, respectively. This family shows a pattern of recent lineage-specific gene divergence; suggesting two independent gene expansions in *H. armigera* and one in *B. mori* driven by diversification in the N-terminal substrate binding domain, likely increasing the range of compounds that could be detoxified or regulated by glycosylation. Although **UGT340** composed of two *B. mori* UGTs are found in the same BmChr28, they seem to be diversified enough to be separated into different family. **UGT34** composed of single genes from each species, UGT34A1 (previously known as *Z. filipendulae* UGT33B1), UGT34A3 (HarUGT15), UGT34A2 (BGIBMGA004965) on BmChr25, occupies a basal position in this clan, suggesting that a gene duplication followed by transposition from BmChr25 to BmChr28 was followed by further diversification of the latter group.

The second largest family **UGT40** is composed of 9 *H. armigera* UGTs and 12 *B. mori* UGTs. Orthologous pairs are more common in this family, except for four recently diverged *B. mori* UGTs (BGIBMGA010098, BGIBMGA010099, BGIBMGA010099-2, and BGIBMGA010295). The *H. armigera* homolog of these four appears to be HarUGT_28, a partial sequence which is not shown in the tree. **UGT41** composed of 4 *H. armigera* UGTs and 3 *B. mori* UGTs occupies a separate position in the phylogenetic tree next to UGT40. Each of the three *B. mori* UGTs contains 8 introns in the coding region of genome, which is the largest number of introns ever found in animals, while the neighboring UGT40 members consistently have 7 introns. Interestingly, UGT48C1 (BGIBMGA002854), an outgroup of this clade, also contains 7 introns, the positions of which are

identical to those of UGT40 members, suggesting that UGT48 is probably the ancestral group in this clade, and has diversified into UGT40 by gene duplication and into UGT41 by intron gain (Fig. 3).

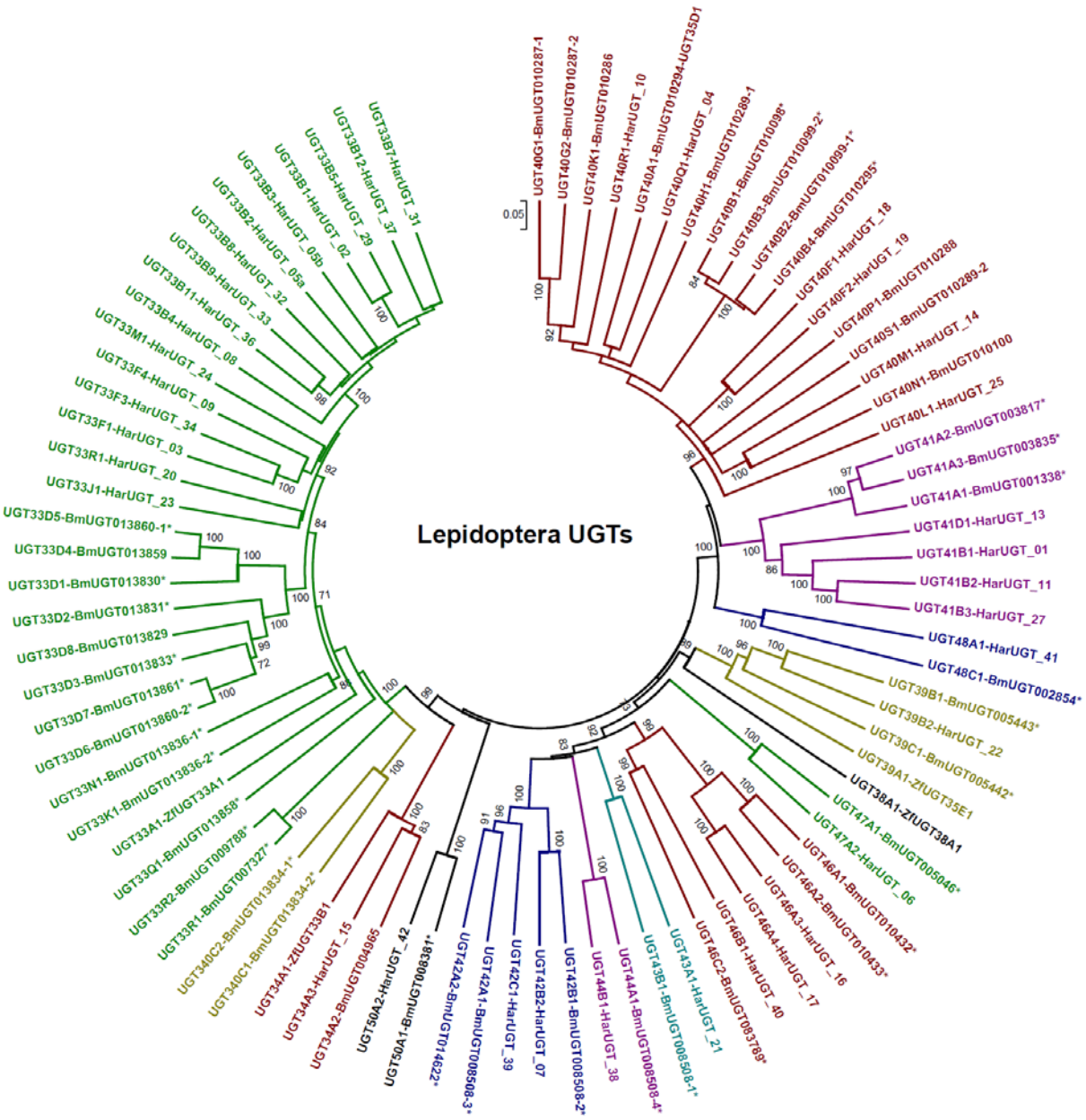


Figure 3. Phylogenetic tree of *H. armigera* and *B. mori* UGTs. Deduced protein sequences of 40 *H. armigera* UGTs and 44 *B. mori* UGTs were aligned by using CLUSTAL and the alignment was manually adjusted to minimize gaps. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates are collapsed. The evolutionary distances were computed using the p-distance method. Evolutionary analyses were conducted in MEGA5.

UGT42, 43, and **44** members in *B. mori* all map to the same chromosome, BmChr18, except UGT42A2 (BGIBMGA014622) which is not assigned to any chromosome yet due to the incompleteness of the genome assembly. Conserved orthologous relationships between the two lepidopteran insect species are shown in this superfamily. **UGT46 family** composes 3 *H. armigera* UGTs and 3 *B. mori* UGTs (BmChr12), which also occur in orthologous pairs; UGT46B1 is an ortholog of UGT46C2 (BGIBMGA083789), whereas UGT46A3 and UGT46A4 correspond to UGT46A1 (BGIBMGA010432) and UGT46A2 (BGIBMGA010433). Interestingly, UGT46A1 and UGT46A2 are produced by alternative splicing of one of two alternative forms of exon 1 to common downstream exons 2, 3, and 4. Since only differences between the cDNA sequences of UGT46A3 and UGT46A4 occur in exon 1, these two *H. armigera* transcripts might also be produced alternative splicing of the same gene. All of the *B. mori* UGT genes in UGT42, 43, 44, and 46 have similar intron-exon structures like the largest family UGT33.

UGT39 is composed of two UGTs from each species, having 4 introns and mapping to BmChr8 in *B. mori* UGTs, although the partial sequence HarUGT_43 is not shown in the tree. One *Z. filipendulae* UGT (UGT39A1, previously known as ZfUGT35E1) is also a member in this family. **UGT38** contains only a single gene (UGT38A1) from *Z. filipendulae*, but no genes from *H. armigera* or *B. mori*, suggesting the family might be unique to *Z. filipendulae*, a species that manipulates cyanogenic glucosides in nature (Zagrobelny et al., 2009).

UGT47 contains a single member from each species, UGT47A2 and UGT47A1 (BGIBMGA005046) on BmChr25. UGT47A1 uniquely has 5 introns, two of which occur within the first large exon present in UGT42-46 and the other three coinciding with introns 1-3 in those families.

UGT50 is a conserved family not only in Lepidoptera but in other insects, composed of one member in each species from all the insects except the pea aphid (Fig. 4). It includes UGT50A2 (*H. armigera*) and UGT50A1 (BGIBMGA008381) in Lepidoptera, as well as UGT50B1 (AAEL005375, *Ae. aegypti*), UGT50B2 (*An. gambiae*), UGT50D1 (CG30438, *D. melanogaster*), UGT50E1 (TcLOC100142501, *T. castaneum*), UGT50C1 (Am15665, *Ap. mellifera*), and UGT50F1 (NvLOC100114140, *N. vitripennis*). The branching pattern of UGT50 mirrors the phylogeny of the species, and the overall pairwise identity is over 50 %, suggesting a common, essential function for this enzyme. No other UGT family in insects shows this pattern. A human homolog (UGT8A1, NP003351) of this conserved group showed higher protein sequence similarities with UGT50A2 (29 %) and with UGT50A1 (28 %) than with other insect UGT families. The human UGT8 is known as a 2-hydroxyacylsphingosine 1-β-galactosyltransferase, or a ceramide UDP-galactosyltransferase, catalyzing the galactose conjugation of a sphingosine (Sprong et al., 1998). Sphingosine (a ceramide) is not only a membrane lipid component, but has been recently recognized as an important endogenous regulator of apoptosis and basic cellular processes (Ruvolo, 2003), suggesting this conserved family within insects may be also responsible for the sphingosine modulation. On the other hand, it is noteworthy that there is no homolog of this family from the pea aphid, as confirmed by a search of the

WGS assembly. Moreover, the predicted *Nasonia vitripennis* homolog (NvLOC100114140) lacks 32 aa residues in the middle of the substrate binding domain compared to the other homologs, and the same truncated homologs were also found in the other two congeneric species, *N. giraulti*, and *N. longicornis*. In spite of the truncation, all the homologs from the three species of *Nasonia* wasp were predicted to produce full length proteins because no frame shift occurs. However, the missing region corresponds to the beginning of exon 3 and there is a gap in the genome assembly at this position in all three species.

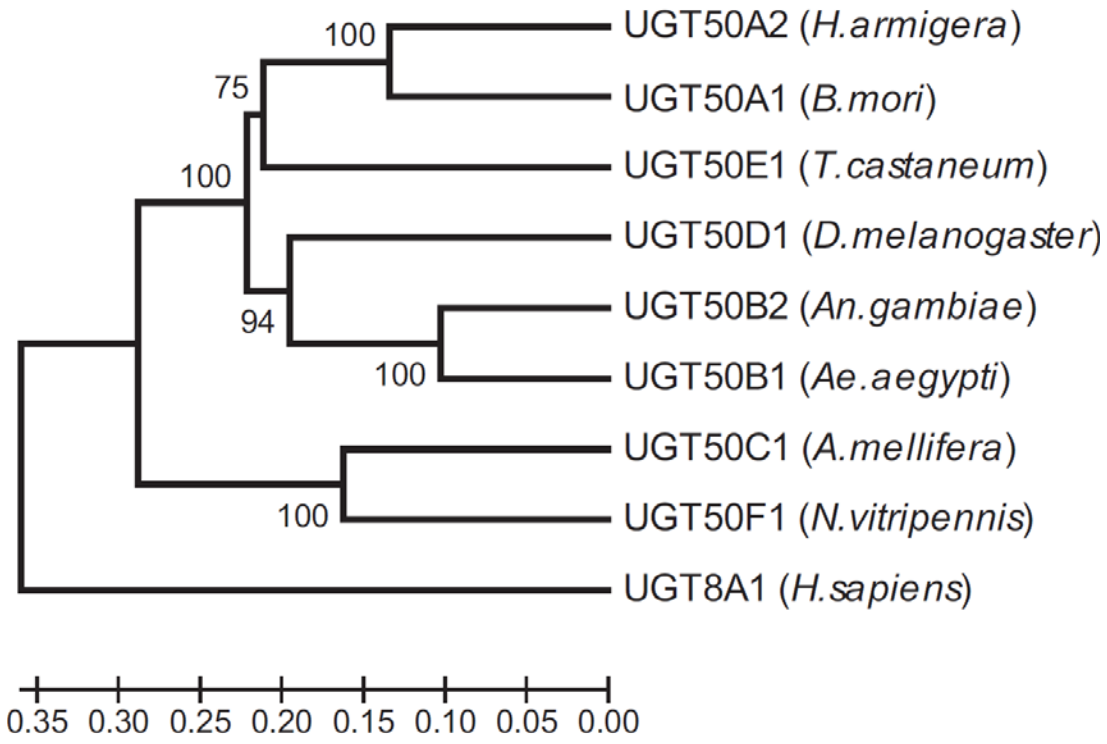


Figure 4. Phylogenetic tree of UGT50 family. The evolutionary history of the conserved family, UGT50, was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method. Evolutionary analyses were conducted in MEGA5.

3.4. Transcript diversity of *H. armigera* UGTs

3.4.1. Tissue distribution

As shown in Table 1, the 40 putative *H. armigera* UGT transcripts were identified from various tissues and developmental stages. Most of UGT33 members were expressed in the rest of body (RB), i.e. larval body exclusive of midgut (MG) and fat body (FB). Therefore, UGT33 doesn't seem to be highly expressed in MG or FB except for three particular genes. In fact, two other lepidopteran UGT genes belonging to this same family showed higher expression in the antenna. UGT33D8 (BGIBMGA013829, *B. mori*) is known to be highly expressed in both larval and adult antennae, although the expression of this gene is not antenna-specific (Huang et al., 2008). A partial EST sequence (AI234470.1) which is very similar to the UGT33 family was expressed in the *Manduca sexta* male antennae (Robertson et al., 1999). This suggests that at least some members of this family might be involved in olfaction (Wang et al., 1999).

Among defined larval tissues, FB and MG are the tissues where UGTs are most highly expressed: 17 and 14 out of 42 *H. armigera* UGTs were expressed in FB and MG, respectively (Table 1). Single tissue-specific UGTs are as follows: MG-specific UGTs were two (UGT40D2 and UGT41B2) and FB-specific UGTs were three (UGT33B1, UGT33B3, and UGT44B1), while UGT40L1 was specific to HC and UGT33J1 was specific to CT. There were also nine UGTs expressed in the pupal and/or adult stages, although specific tissue is not defined.

3.4.2. Transcript variants

UGT33B2 produced two different transcript variants; five out of seven cDNA sequences had an internal deletion of 220 bp corresponding to exon 3 of *B. mori* UGTs in the same family, whereas the other two cDNA sequences were intact (Fig. 5A). Exon-intron splicing sites were predicted by comparison with corresponding splicing sites in *B. mori*. Skipping exon 3 shifts the coding frame and produces a premature stop codon in the middle of putative exon 4. Whether such a truncated UGT could have enzyme activity is unknown. Since more than half the N-terminal sequence could be translated, it might bind at least to a sugar acceptor, thereby trapping substrates. However, it doesn't have the ER-retention signal (e.g. KDEL) at the C-terminal end of the truncated protein, so it might be secreted from the ER.

Retained (unspliced) introns were identified from the transcripts of three *H. armigera* UGTs; UGT40R1, UGT34A3, and UGT43A1 (Fig. 5B). The variants with retained introns, if translated, would also produce truncated proteins, which retain either the N-terminal half (UGT43A1_v2), or more (UGT34A3_v2), or much less (UGT40R1_v2). The correctly spliced variants for all 3 genes represent ~80 % of transcripts and the retained introns are quite short, suggesting they might remain due to a mechanistic error of splice site recognition.

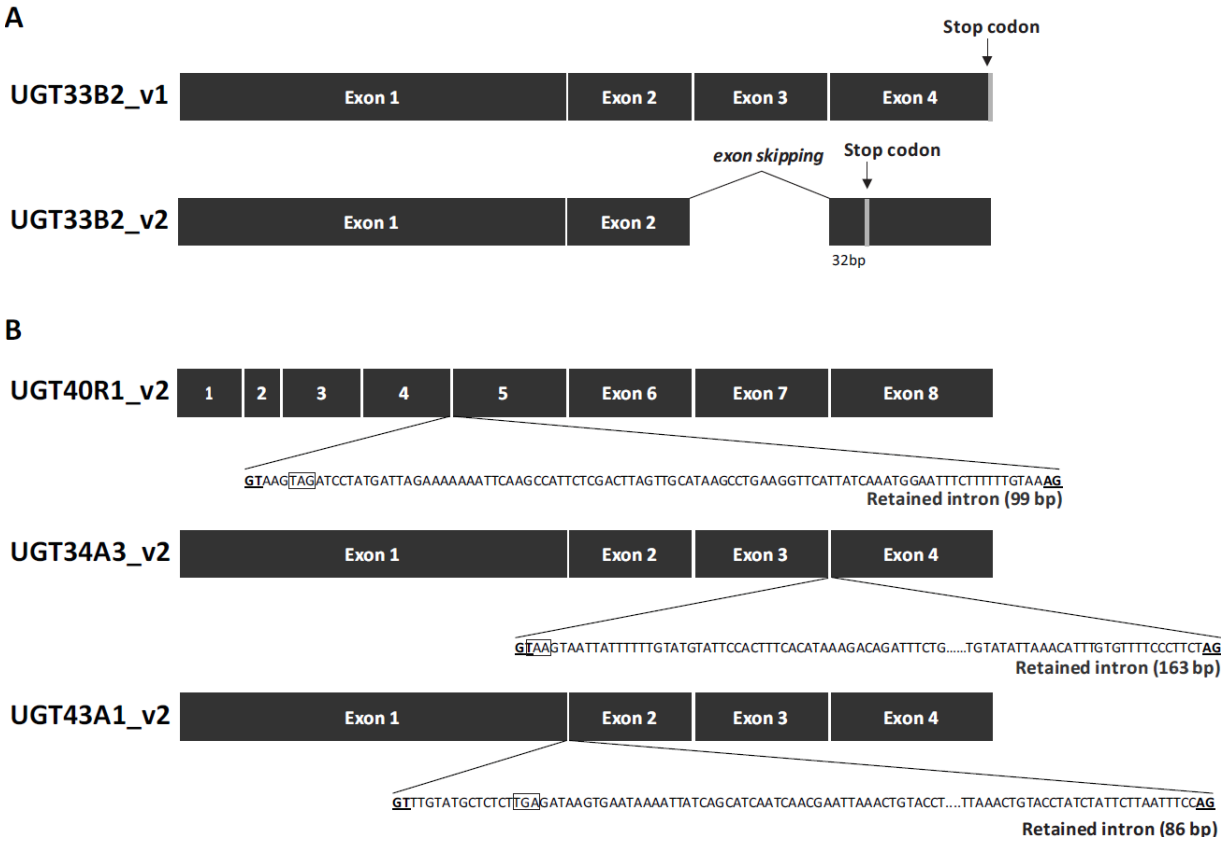


Figure 5. Transcript diversity of *H. armigera* UGTs. (A) Exon skipping. A transcript variant UGT33B2_v2 was truncated by 220 bp in exon 3 position, which is shown in a correctly spliced variant UGT33B2_v1. The putative translation of the second variant seems to be terminated immaturely due to the frame shift. (B) Intron retention. Transcript variants retaining putative intron sequences were identified in UGT40R1, UGT34A3, and UGT43A1; they seem to produce immaturely-translated proteins. Exon-intron splicing sites were predicted by comparison with the *B. mori* orthologous sequences.

3.4.3. Lineage-specific conserved domain in UGT33

A highly conserved nucleotide sequence domain was found in the largest *H. armigera* UGT family, UGT33. All of 16 members of the family except two (UGT33J1 and UGT33R1) contained a 52 bp-long nucleotide sequence that is absolutely conserved just in front of the transmembrane domain at the C-terminal end (Fig. 6). No third-codon-position substitutions have occurred since the divergence of these genes. This suggests there might be binding interactions between the genomic DNA or mRNA in these regions and another protein. In contrast, the *B. mori* paralogs in the UGT33 family don't have such a highly conserved region.

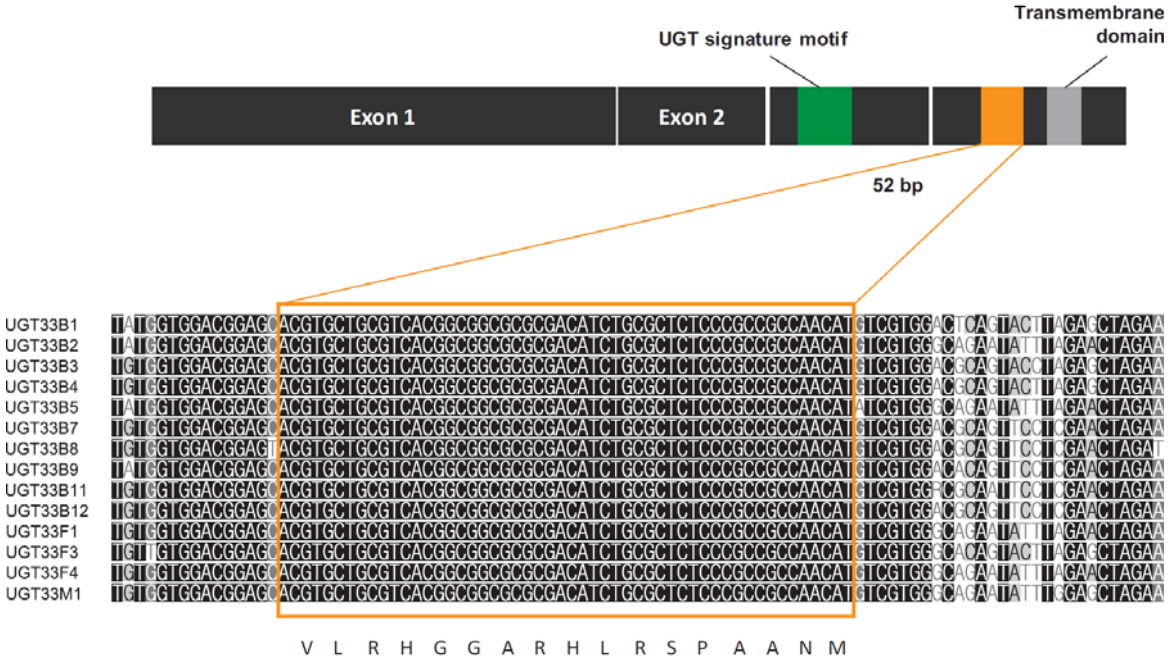


Figure 6. Lineage-specific nucleotide conservation in *H. armigera* UGT33 family. A nucleotide sequence region spanning 52 bp is completely conserved in 14 UGT33s in *H. armigera*. The region lies in front of transmembrane domain at C-terminal end.

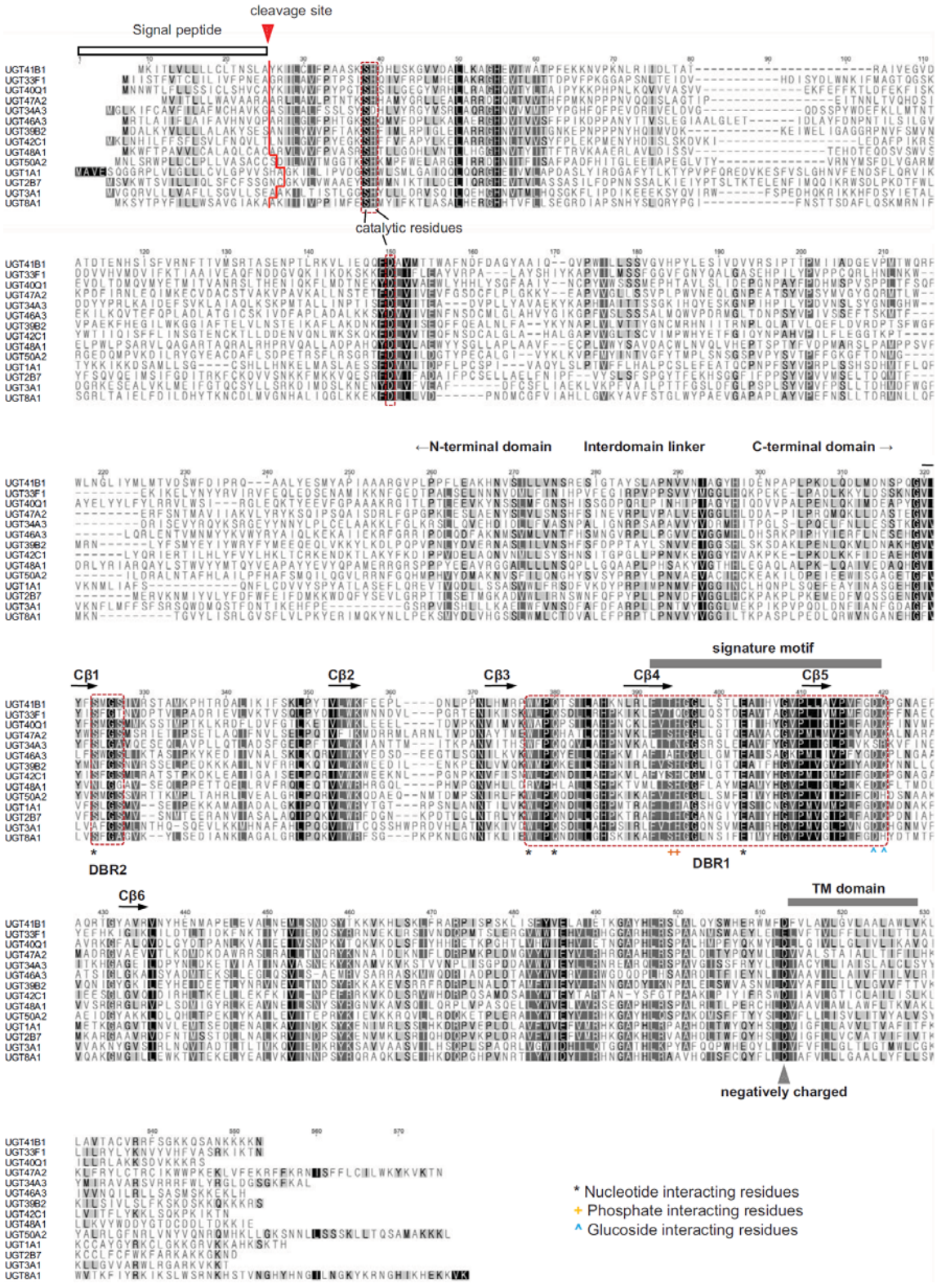


Figure 7. Multiple alignment of the 10 representative *H. armigera* UGTs with four human UGTs. Multiple alignment was performed by using CLUSTAL in Geneious software. Signal peptide in N-terminal predicted by SignalP 3.0 is shown in white bar, and signature motif and transmembrane (TM) domain in C-terminal domain are shown in gray bar above the alignment. Catalytic residues composed of S, H, and D are shown in red-dotted boxes. All of the putative β-sheets in C-terminal half predicted by comparison with a human UGT2B7 crystal structure (Miley et al., 2007) are denoted as Cβ+number with an arrow. DBR refers to donor binding region, and several important residues interacting with the donor components are shown in different indicators (*, +, or ^).

3.5. Structure prediction of the UGT proteins

3.5.1. Structural motifs of the *H. armigera* UGT proteins

Multiple alignment of 10 representative *H. armigera* UGT amino acid sequences reveals the two major domains; the N-terminal substrate binding domain is highly variable whereas the C-terminal sugar-donor binding domain is more conserved (Fig. 7). All *H. armigera* UGTs consist of a 14-22 amino acid-long signal peptide found at the N-terminal end, which is presumably cleaved after integration into the ER compartment (Seppen et al., 1996). The signal peptide cleavage sites predicted by SignalP 3.0 (Bendtsen et al., 2004) are shown in Figure 7. The UGT signature motif (also called PSPG motif, Plant Secondary Product Glycosyltransferase motif, particularly in plants) is found in the middle of the C-terminal domain, which is conserved throughout all organisms. A single short transmembrane domain, composed of about 16 hydrophobic amino acid residues, was found close to the end of the C-terminal domain, confirming that the insect UGTs are membrane bound, unlike plant UGTs which are cytosolic (Osmani et al., 2009). A negatively charged amino acid residue (mostly aspartic acid, D or glutamic acid, E) right in front of the transmembrane domain on the luminal side is highly conserved, suggesting an important role in membrane association of the protein. A short carboxyl terminal stretch of 20-25 amino acids is thought to be exposed to the cytoplasm, probably involved in ER retention (Magdalou et al., 2010), although long cytoplasmic tails (45-46 aa) in UGT47A2 and UGT50A2 (and their *B. mori* orthologs) are unusual. The proposed ortholog to UGT50, human UGT8A1 (ceramide UDP-galactosyltransferase), also has a long C-terminal cytoplasmic tail.

3.5.2. Prediction of sugar binding residues and catalytic triad

The recent crystallization of the C-terminal domain of human UGT2B7 provides detailed information on the sugar donor binding site (Miley et al., 2007). Predicted UDP-glucuronic acid binding regions (donor binding regions, DBR1 and DBR2) are also conserved in insect UGTs (Fig. 7). More specifically, the nucleotide component of the UDP-glucuronic acid is known to be associated with S324, W376, Q380, and E403A; the phosphate groups interact with T394 (α-phosphate) and H395 (β-phosphate); and the glucuronic acid moiety interacts with D419 (O2'/O3') and Q420 (O3'/O4') (Radominska-Pandya et al., 2010). These important residues for the sugar donor binding are also conserved in insect UGTs, suggesting each of these residues in insects might also play an important role in sugar binding. Since insect UGTs use UDP-glucose as a major sugar donor rather than UDP-glucuronic acid, the binding topology could be a little bit different between them. However, since D419 and Q420 residues interact with common oxygens between glucose and glucuronic acid, and not with oxygen in the carboxyl group unique to glucuronic acid, the conserved DQ in insects may also associate with the glucose moiety. On the other hand, UGT50A2 (and the *B. mori* ortholog

UGT50A1-BGIBMGA008381) together with human UGT8A1 (a ceramide UDP-galactosyltransferase) has DH instead of DQ at this position, suggesting H420 might be interacting with oxygens at O3'/O4' position of galactose because the O4' position is the only difference between glucose and galactose. In addition, UGT34A3 (and the *B. mori* ortholog UGT34A2 (BGIBMGA004965), and *Z. filipendulae* ortholog UGT34A1) has SR instead of DQ, indicating a quite different sugar donor might be used with these particular UGTs. Since UGT48A1 shows also a divergence at three key residues out of four for the nucleotide binding (N324, R376, and H380), we would predict this UGT to interact with another nucleotide-sugar donor rather than with uridine diphosphate sugar. The catalytic triad composed of serine (S), aspartic acid (D), and histidine (H) is found at S38, H39, and D150, supposedly involved in the direct mechanism of substrate catalysis. All of these predictions made by primary structures can only be verified when investigations on the crystal structure of insect UGTs are pursued.

3.6. Conclusions

We have identified over 310 insect UGT genes from the genome databases of eight different insect species and the transcript database of *H. armigera.* A comparative genomic analysis revealed the extent of the UGT multigene family in insects. Phylogenetic analyses showed lineage-specific gene diversification of some UGT subfamilies, as well as conservation of a certain family across many organisms. Comparative study of the *H. armigera* UGTs with *B. mori* UGTs gave an insight on how conserved a certain family could be, whereas lineage-specific gene expansion was also characterized in the lepidopteran UGTs. Alternative splicing events such as exon skipping and intron retention increased transcript diversity in *H. armigera* UGTs. Sequence alignments predicted important structural domains and catalytic residues. Knowledge of the insect UGT sequences and organization in genome will shed light on the future research of this multigene family in combination with xenobiotic detoxification and endobiotic modification in insects.

Acknowledgements

This study was funded by the Max Planck Society. Seung-Joon Ahn holds a scholarship of the IMPRS program of the Max Planck Society. We thank Steffi Gebauer-Jung for the database maintenance and Domenica Schnabelrauch for DNA sequencing.

References

Ahmad, S., Brattsten, L.B., Mullin, C.A., Yu, S.J., 1986. Enzymes involved in the metabolism of plant allelochemicals, in: Brattsten, L.B., Ahmad, S. (Eds.), Molecular aspects of insect-plant associations. Plenum Press, New York & London, pp. 73-151.

Ahmad, S.A., Hopkins, T.L., 1992. Phenol β-glucosyltransferase and β-glucosidase activities in the tobacco hornworm larva *Manduca sexta* (L.): Properties and tissue localization. Arch. Insect Biochem. Physiol. 21, 207-224.

Ahmad, S.A., Hopkins, T.L., 1993a. β-Glucosylation of plant phenolics by phenol β-glucosyltransferase in larval tissues of the tobacco hornworm, *Manduca sexta* (L.). Insect Biochem. Mol. Biol. 23, 581-589.

Ahmad, S.A., Hopkins, T.L., 1993b. Phenol β-glucosyltransferases in six species of insects: properties and tissue localization. Comp. Biochem. Physiol. B: Comp. Biochem. 104, 515-519.

Ahmad, S.A., Hopkins, T.L., Kramer, K.J., 1996. Tyrosine β-glucosyltransferase in the tobacco hornworm, *Manduca sexta* (L): Properties, tissue localization, and developmental profile. Insect Biochem. Mol. Biol. 26, 49-57.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783-795.

Bock, K.W., 2003. Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. Biochem. Pharmacol. 66, 691-696.

Bowles, D., Isayenkova, J., Lim, E.K., Poppenberger, B., 2005. Glycosyltransferases: managers of small molecules. Curr. Opin. Plant Biol. 8, 254-263.

Brattsten, L.B., 1988. Enzymic adaptations in leaf-feeding insects to host-plant allelochemicals. J. Chem. Ecol. 14, 1919-1939.

Claudianos, C., Ranson, H., Johnson, R.M., Biswas, S., Schuler, M.A., Berenbaum, M.R., Feyereisen, R., Oakeshott, J.G., 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol. Biol. 15, 615-636.

Daimon, T., Hirayama, C., Kanai, M., Ruike, Y., Meng, Y., Kosegawa, E., Nakamura, M., Tsujimoto, G., Katsuma, S., Shimada, T., 2010. The silkworm Green b locus encodes a quercetin 5-O-glucosyltransferase that produces green cocoons with UV-shielding properties. Proc. Natl. Acad. Sci. USA. 107, 11471-11476.

Després, L., David, J.P., Gallet, C., 2007. The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol. Evol. 22, 298-307.

Felsenstein, J., 2004. Inferring phylogenies. Sinauer Associates Inc.

Fitt, G.P., 1989. The ecology of *Heliothis* species in relation toagroecosystems. Annu. Rev. Entomol. 34, 17-53.

Heckel, D.G., 2010. Molecular genetics of insecticide resistance in Lepidoptera, in: Goldsmith, M.R., Marec, F. (Eds.), Molecular biology and genetics of the Lepidoptera. Taylor & Francis, pp. 239-269.

Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931-949.

Hopkins, T.L., Kramer, K.J., 1992. Insect cuticle sclerotization. Annu. Rev. Entomol. 37, 273-302.

Huang, F.F., Chai, C.L., Zhang, Z., Liu, Z.H., Dai, F.Y., Lu, C., Xiang, Z.H., 2008. The UDP-glucosyltransferase multigene family in *Bombyx mori*. BMC Genomics 9, 563.

Luque, T., O'Reilly, D.R., 2002. Functional and phylogenetic analyses of a putative *Drosophila melanogaster* UDP-glycosyltransferase gene. Insect Biochem. Mol. Biol. 32, 1597-1604.

Luque, T., Okano, K., O'Reilly, D.R., 2002. Characterization of a novel silkworm (*Bombyx mori*) phenol UDP-glucosyltransferase. Eur. J. Biochem. 269, 819-825.

Mackenzie, P.I., Owens, I.S., Burchell, B., Bock, K.W., Bairoch, A., Belanger, A., FournelGigleux, S., Green, M., Hum, D.W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J.R., Ritter, J.K., Schachter, H., Tephly, T.R., Tipton, K.F., Nebert, D.W., 1997. The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7, 255-269.

Magdalou, J., Fournel-Gigleux, S., Ouzzine, M., 2010. Insights on membrane topology and structure/function of UDP-glucuronosyltransferases. Drug Metab. Rev. 42, 159-166.

Miley, M.J., Zielinska, A.K., Keenan, J.E., Bratton, S.M., Radominska-Pandya, A., Redinbo, M.R., 2007. Crystal Structure of the Cofactor-Binding Domain of the Human Phase II Drug-Metabolism Enzyme UDP-Glucuronosyltransferase 2B7. J. .Mol. Biol. 369, 498-511.

Morello, A., Repetto, Y., 1979. UDP-glucosyltransferase activity of housefly microsomal fraction. Biochem. J. 177, 809-812.

Oakeshott, J.G., Johnson, R.M., Berenbaum, M.R., Ranson, H., Cristino, A.S., Claudianos, C., 2010. Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. Insect Mol. Biol. 19, 147-163.

Osmani, S.A., Bak, S., Møller, B.L., 2009. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70, 325-347.

Radominska-Pandya, A., Bratton, S.M., Redinbo, M.R., Miley, M.J., 2010. The crystal structure of human UDP-glucuronosyltransferase 2B7 C-terminal end is the first mammalian UGT target to be revealed: the significance for human UGTs from both the 1A and 2B families. Drug Metab. Rev. 42, 133-144.

Ramsey, J.S., Rider, D.S., Walsh, T.K., De Vos, M., Gordon, K.H.J., Ponnala, L., Macmil, S.L., Roe, B.A., Jander, G., 2010. Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. Insect Mol. Biol. 19, 155-164.

Real, M.D., Ferré, J., Chapa, F.J., 1991. UDP-glucosyltransferase activity toward exogenous substrates in *Drosophila melanogaster*. Anal. Biochem. 194, 349-352.

Robertson, H.M., Martos, R., Sears, C.R., Todres, E.Z., Walden, K.K.O., Nardi, J.B., 1999. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. Insect Mol. Biol. 8, 501-518.

Ross, J., Li, Y., Lim, E.-K., Bowles, D.J., 2001. Higher plant glycosyltransferases. Genome Biology 2, reviews3004.1–3004.6

Ruvolo, P.P., 2003. Intracellular signal transduction pathways activated by ceramide and its metabolites. Pharmacol. Res. 47, 383-392.

Sasai, H., Ishida, M., Murakami, K., Tadokoro, N., Ishihara, A., Nishida, R., Mori, N., 2009. Species-specific glucosylation of DIMBOA in larvae of the rice armyworm. Biosci. Biotech. Bioch. 73, 1333-1338.

Seppen, J., Steenken, E., Lindhout, D., Bosma, P.J., Oude Elferink, R.P.J., 1996. A mutation which disrupts the hydrophobic core of the signal peptide of bilirubin UDP-glucuronosyltransferase, an endoplasmic reticulum membrane protein, causes Crigler-Najjar type IIs. FEBS Lett. 390, 294-298.

Smith, J.N., 1962. Detoxication mechanisms. Annu. Rev. Entomol. 7, 465-480.

Sprong, H., Kruithof, B., Leijendekker, R., Slot, J.W., van Meer, G., van der Sluijs, P., 1998. UDP-galactose:ceramide galactosyltransferase is a class I integral membrane protein of the endoplasmic reticulum. J. Biol. Chem. 273, 25880-25888.

Strode, C., Wondji, C.S., David, J.P., Hawkes, N.J., Lumjuan, N., Nelson, D.R., Drane, D.R., Karunaratne, S., Hemingway, J., Black, W.C., Ranson, H., 2008. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. Insect Biochem. Mol. Biol. 38, 113-123.

Svoboda, J., Weirich, G., 1995. Sterol metabolism in the tobacco hornworm, *Manduca sexta* - A review. Lipids 30, 263-267.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. (In Press).

Teese, M.G., Campbell, P.M., Scott, C., Gordon, K.H.J., Southon, A., Hovan, D., Robin, C., Russell, R.J., Oakeshott, J.G., 2010. Gene identification and proteomic analysis of the esterases of the cotton bollworm, *Helicoverpa armigera*. Insect Biochem. Mol. Biol. 40, 1-16.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

Tijet, N., Helvig, C., Feyereisen, R., 2001. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: Annotation, intron-exon organization and phylogeny. Gene 262, 189-198.

Tsubota, T., Shiotsuki, T., 2010. Genomic analysis of carboxyl/cholinesterase genes in the silkworm *Bombyx mori*. BMC Genomics 11, 377

Wang, Q., Hasan, G., Pikielny, C.W., 1999. Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. J. Biol. Chem. 274, 10309-10315.

Wiesen, B., Krug, E., Fiedler, K., Wray, V., Proksch, P., 1994. Sequestration of host-plant-derived flavonoids by lycaenid butterfly *Polyommatus icarus*. J. Chem. Ecol. 20, 2523-2538.

Yu, Q.Y., Lu, C., Li, B., Fang, S.M., Zuo, W.D., Dai, F.Y., Zhang, Z., Xiang, Z.H., 2008. Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1158-1164.

Zagrobelny, M., Scheibye-Alsing, K., Jensen, N.B., Moller, B.L., Gorodkin, J., Bak, S., 2009. 454 pyrosequencing based transcriptome analysis of *Zygaena filipendulae* with focus on genes involved in biosynthesis of cyanogenic glucosides. Bmc Genomics 10, 574.

3.4. Manuscript IV

Genomic analysis of the UDP-glycosyltransferase multigene family in *Helicoverpa armigera*

Seung-Joon Ahn, Heiko Vogel and David G. Heckel

unpublished results

Department of Entomology, Max-Planck Institute for Chemical Ecology, Jena, 07745, Germany

*Corresponding author. Tel.: +49 3641 571501; fax: +49 3641 571502. *E-mail address*: heckel@ice.mpg.de (D.G. Heckel)

Abstract

UDP-glycosyltransferases (UGT) catalyze the conjugation of a range of diverse small lipophilic compounds with sugars to produce glycosides, playing an important role in the detoxification of xenobiotics and in the regulation of endobiotics in insects. Recent progress in genome sequencing has revealed the extent of the UGT multigene family in insects. Here we report over 320 putative UGT genes identified from genomic databases of 9 different insect species including two lepidopterans, *Helicoverpa armigera* and *Bombyx mori*. Phylogenetic analysis of the insect UGTs showed Order-specific gene diversifications and inter-species conservations of the multigene family. A sub-tree constructed with 46 *H. armigera* UGTs and 45 *B. mori* UGTs revealed that phylogenetic groupings of the lepidopteran UGTs are consistent with the intron-exon structure and chromosomal organization. Lineage-specific expansions of some subfamilies in both species appears to be driven by diversification in the N-terminal substrate binding domain, increasing the range of compounds that could be detoxified or regulated by glycosylation. Investigation of intron-exon organization in the UGTs from both species revealed that the intron positions and phases were highly conserved among the UGTs from the same species as well as between two lepidopteran species. The comparative analyses of UGTs from two lepidopteran species genome sequences demonstrated the conservation as well as diversification dynamics of the multigene family. This study also provides primary information for functional and more detailed evolutionary studies on insect UGTs in the future.

Keywords: UDP-glycosyltransferase, *Helicoverpa armigera*, *Bombyx mori*, multigene family, phylogenetic analysis, intron map, detoxification

1. Introduction

At the time of this manuscript preparation, there are 33 insect species of which genome sequences have been either completed or at least assembled (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi). In an era of a surge of genome sequences, many insect gene families involved in detoxification and metabolism, like cytochrome P450s (P450s), carboxyl/cholinesterases (CCEs), glutathione transferases (GSTs), have been identified from *Drosophila melanogaster* (Tijet et al., 2001), *Aedes aegypti* (Strode et al., 2008), *Acyrthosiphon pisum* (Ramsey et al., 2010), *Nasonia vitripennis* (Oakeshott et al., 2010), *Apis melifera* (Claudianos et al., 2006), and *Bombyx mori* (Tsubota and Shiotsuki, 2010; Yu et al., 2008). The molecular identities of the insect UGTs are, however, relatively unknown compared to the other detoxification gene families in insects or to the vertebrate UGT gene families. Although the UGT gene families in *D. melanogaster* (Luque and O'Reilly, 2002) and in *Bombyx mori* (Huang et al., 2008) have been reported, they have not been compared in the context of the diversity of other insects.

The cotton bollworm *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae) is a polyphagous herbivore feeding more than 100 different species belonging to about 46 different plant families, and it is regarded as a serious agricultural pest in many regions in the world (Fitt, 1989). Thereby, extensive attentions have been paid to this insect in order to understand its ecology and physiology especially in relation to host plant adaptation and insecticide resistance (Heckel, 2010). Recently, the *H. armigera* CCE gene family was reported based on the transcriptome analysis (Teese et al., 2010), but the genome-wide study on any gene family including detoxification-related one from this species has not been tried because genome sequences are not available. Since the Helicoverpa Genome Project was launched in 2009, a lot of genomic contigs have been generated, although they are not assembled yet. The fragmented genomic contigs are, however, useful enough to assemble them mannually in order to investigate a specific gene family. Furthermore, it would be easier to assemble if a closely-related model organism, for example *Bombyx mori*, is compared.

In this study, we identified all the putative UGT genes from the *H. armigera* and *B. mori* genome databases, as well as from another 7 public insect genomes. Here we describe the genomic and phylogenetic analyses of the *H. armigera* and *B. mori* UGTs in the context of the insect UGT diversity, and we analyze the intron-exon organization based on the comparative viewpoints with two lepidopteran species. Finally, the prediction of the UGT protein structure is performed by comparing several important domains and residues with human UGT sequences.

2. Materials and methods

2.1. Identification and annotation of *Helicoverpa armigera* UGT sequences

The UGT-like sequences were retrieved from *H. armigera* genomic and transcript databases in MPI-CE private library. The *H. armigera* genomic contigs had been obtained from the Helicoverpa Genome Project (http://insectacentral.org/helicoverpa), and the transcripts had been obtained from cDNA libraries using tissues from different developmental stages of the Toowoomba strain. The collected sequences of UGT genes were assembled in Sequencher™. Each assembly representing a single gene was arbitrarily named and the assembly work was continued until no more contigs were left. The 5'- or 3'-UTR sequences were assembled with flanking genomic contigs and extened as long as possible in order to identify neighboring genes for the synteny comparison with *B. mori* genome. Coding sequences and intron-exon boundaries were reconciled using cDNA sequences. For genes without cDNA support, the splicing sites were predicted by protein alignments with previously annotated sequences based on the GT-AG rule.

2.2. Phylogenetic analysis

The phylogenetic reconstruction implemented two methods. Predicted amino acid sequences of the *H. armigera* and *B. mori* UGTs together with four *Zygaena filipendulae* UGTs were aligned by Clustal X algorithm (Thompson et al., 1997) using MEGA 4.0 software (Tamura et al., 2007). The alignment was manually edited to minimize gaps. A consensus phylogenetic tree was constructed using the neighbor-joining method. Distance calculations were performed after Tajima & Nei and bootstrap analysis, running 1000 bootstrap samples (Felsenstein, 2004). Bayesian interference using a GTR+I+G nucleotide substitution model was implemented in Mr. Bayes 3.1 (Ronquist and Huelsenbeck, 2003). The tree was constructed using a *Drosophila* UGT sequence as outgroup. The Marcov Chain Monte Carlo runs were carried out for 1,000,000 generations, after which log likelihood values showed that equilibrium had been reached after the first 400 generations in all cases, and those data were discarded from each run and considered as "burnin". Two runs were conducted per dataset showing agreement in topology and likelihood scores. The Neighbour-joining and the Bayesian tree topologies including their general subfamily relationships were in agreement.

2.3. Intron mapping

Intron positions of UGTs were mapped on the dashed lines representing ten amino acid residues as one dash. A phase 0 (|) splice site lies between two codons, while a phase 1 ( ]) site lies one base inside the codon in the 3 prime direction and the phase 2 ([ ) intron lies two base inside a codon in the 3 prime direction.

3. Results and discussion

3.1. Identification of *H. armigera* UGTs

Iterative assemblies of fragmented genomic contigs and transcript sequences from the MPI-CE sequence library revealed 46 putative full and partial *H. armigera* UGT sequences (Table 1). All the sequences were supported by cDNA contigs except three genes (UGT33S1-HarUGT_26, UGT33B6-HarUGT_30 and UGT33B10) that were predicted by genomic contigs only. Two partial genes (HarUGT_28 and HarUGT_43) were obtained only by short cDNA contigs, but not by genomic contigs. A putative pseudogene (UGT33B10-HarUGT_35P) is disrupted by LTR-type retrotransposon in the middle of exon 1. Another pseudogene (HarUGT_24aP), having only an 879 base-long exon 1, is located in the intergenic genomic region between UGT33M1-HarUGT_24 and UGT33F2-HarUGT_09. It seems to be disrupted by a NOF-FB-like transposable element in the position between exon 1s of HarUGT_24aP and UGT33F2-HarUGT_09. High nucleotide sequence similarity of exon 1 of HarUGT_24aP and exon 1 of an adjacent UGT33F2-HarUGT_09 (76 %) suggests HarUGT_24aP has been recently derived from UGT33F2-HarUGT_09 exon 1 by domain duplication.

3.2. Phylogenetic analysis of insect UGTs

3.2.1. Overall phylogenetic tree of insect UGTs

We have constructed an overall phylogenetic tree using over 320 UGTs from nine different insect genomes including the two lepidopteran species (Fig. 1). The identified UGT sequences also included four UGTs from *Z. filipendulae*, three of them had been reported by Zagrobelny et al. (2009) and one of them were collected from GenBank. The phylogenetic tree (Fig. 1) segregated the insect UGTs into several Order-specific groups as well as orthologous groups among species. Two separate groups specific to the Homoptera (the pea aphid) were composed of 44 UGTs and 13 UGTs separately, accounting for 97 % of all 59 *Ac. pisum* UGTs together. A large Diptera specific group was composed of 23 *D. melanogaster* UGTs, 8 *An. gambiae* UGTs, and 11 *Ae. aegypti* UGTs, and another Diptera specific group was composed of 5 *D. melanogaster* UGTs, 4 *An. gambiae* UGTs, , and 6 *Ae. aegypti* UGTs. Two separate groups specific to the two mosquitoes were also found in one group with 9 *An. gambiae* UGTs and 11 *Ae. aegypti* UGTs; and the other with 2 and 3, respectively. An 1:1 orthologous relationship of the UGTs between the two mosquitoes was shown in these groups as well as in the Diptera specific groups, except some recently duplicated additional UGTs found only in *Ae. aegypti*. One Hymenoptera specific group was composed of 18 *N. vitripennis* UGTs and 3 *A. mellifera* UGTs. A Coleoptera specific group with 25 *T. castaneum* UGTs accounting 58 % of 43 UGTs was segregated. Two large Lepidoptera specific groups were also found that is discussed in the following section in more detail. The other UGTs not included in these Order-specific groups seem to be dispersed to take orthologous positions with other lineages.

Table 1. Summary of the *H. armigera* UGTs and *B. mori* UGTs

Family	*H. armigera*					*B. mori*				
	Official name	gene name	protein length	No. exons	No. transcripts	Official name	gene name	protein length	No. exons	BmChr.
UGT41	UGT41B1	HarUGT_01	513	9	16	UGT41A1	BGIBMGA001337+38	518	9	24
	UGT41B2	HarUGT_11	516	9	11	UGT41A2	BGIBMGA003817	517	9	24
	UGT41B3	HarUGT_27	513	9	4	UGT41A3	BGIBMGA003835	516	9	24
	UGT41D1	HarUGT_13	519	9	25					
UGT48	UGT48A1	HarUGT_41	520	8	1	UGT48C1	BGIBMGA002854	506	8	10
UGT40	UGT40D1	HarUGT_12a	521	8	23	UGT40A1	BGIBMGA010294	520	8	7
	UGT40D2	HarUGT_12b	521	8	5	UGT40B1	BGIBMGA010098	499	8(7)	7
	UGT40F1	HarUGT_18	516	8	11	UGT40B2	BGIBMGA010099P	474	8(7)	7
	UGT40F2	HarUGT_19	520	8	13	UGT40B3	BGIBMGA010099-2	518	8	7
	UGT40L1	HarUGT_25	520	8	9	UGT40B4	BGIBMGA010295	518	8	7
	UGT40M1	HarUGT_14	519	8	17	UGT40G1	BGIBMGA010287	514	8	7
	UGT40Q1	HarUGT_04	518	8	17	UGT40G2	BGIBMGA010287-2P	514	8	7
	UGT40R1	HarUGT_10	519	8	16	UGT40H1	BGIBMGA010289	516	8	7
		HarUGT_28	215	8(4)	1	UGT40K1	BGIBMGA010286	522	8	7
						UGT40N1	BGIBMGA010100	519	8	7
						UGT40P1	BGIBMGA010288	519	8	7
						UGT40S1	BGIBMGA010289-2	516	8	7
UGT47	UGT47A2	HarUGT_06	535	6	21	UGT47A1	BGIBMGA005046	536	6	25
UGT46	UGT46A3	HarUGT_16	527	4	11	UGT46A1	BGIBMGA010432	527	4	12
	UGT46A4	HarUGT_17	527	4	16	UGT46A2	BGIBMGA010433+34	525	4	12
	UGT46B1	HarUGT_40	515	4	1	UGT46C2	KAIKOGA083789	448	4(3)	12
UGT43	UGT43A1	HarUGT_21	520	4	13	UGT43B1	BGIBMGA008508	516	4	18
UGT42	UGT42B2	HarUGT_07	521	4	7	UGT42A1	BGIBMGA008508-3	512	4	18
	UGT42C1	HarUGT_39	509	4	3	UGT42A2	BGIBMGA014622	509	4	18
						UGT42B1	BGIBMGA008508-2	508	4	18
UGT44	UGT44A2	HarUGT_38	526	4	6	UGT44A1	BGIBMGA008508-4	525	4	18
UGT34	UGT34A3	HarUGT_15	525	4	16	UGT34A2	BGIBMGA004965	525	4	25
UGT33	UGT33F1	HarUGT_03	520	4	7	UGT340C1	BGIBMGA013834-2	521	4	28
	UGT33F2	HarUGT_09	520	4	7	UGT340C2	BGIBMGA013834	524	4	28
	UGT33F3	HarUGT_34	519	4	2	UGT33K1	BGIBMGA013836-2	519	4	28
	UGT33J1	HarUGT_23	519	4	8	UGT33N1	BGIBMGA013836P	516	4	28
	UGT33M1	HarUGT_24	517	4	10	UGT33Q1	BGIBMGA013858	419	4(3)	28
		HarUGT_24aP	293	1	0	UGT33R1	BGIBMGA007327	504	4	3
	UGT33S1	HarUGT_26	514	4	0	UGT33R2	BGIBMGA009788P	510	4	2
	UGT33T1	HarUGT_20	524	4	5		BGIBMGA009787P	271	4(1)	2
	UGT33B1	HarUGT_02	529	4	9	UGT33D1	BGIBMGA013830	513	4	28
	UGT33B2	HarUGT_05a	505	4	7	UGT33D2	BGIBMGA013831	515	4	28
	UGT33B3	HarUGT_05b	513	4	10	UGT33D3	BGIBMGA013833	515	4	28
	UGT33B4	HarUGT_08	512	4	23	UGT33D4	BGIBMGA013859	520	4	28
	UGT33B5	HarUGT_29	513	4	1	UGT33D5	BGIBMGA013860	520	4	28
	UGT33B6	HarUGT_30	512	4	0	UGT33D6	BGIBMGA013860-2	515	4	28
	UGT33B7	HarUGT_31	512	4	2	UGT33D7	BGIBMGA013861+62	515	4	28
	UGT33B8	HarUGT_32	511	4	4	UGT33D8	BGIBMGA013829	514	4	28
	UGT33B9	HarUGT_33	513	4	2					
	UGT33B10	HarUGT_35P	512	4	0					
	UGT33B11	HarUGT_36	513	4	1					
	UGT33B12	HarUGT_37	513	4	4					
UGT39	UGT39B2	HarUGT_22	525	5	10	UGT39B1	BGIBMGA005443	520	4	8
		HarUGT_43	135	5(2)	2	UGT39C1	BGIBMGA005442	525	4	8
UGT50	UGT50A2	HarUGT_42	520	6	1	UGT50A1	BGIBMGA008381	540	6	18
Total		46			378		45			

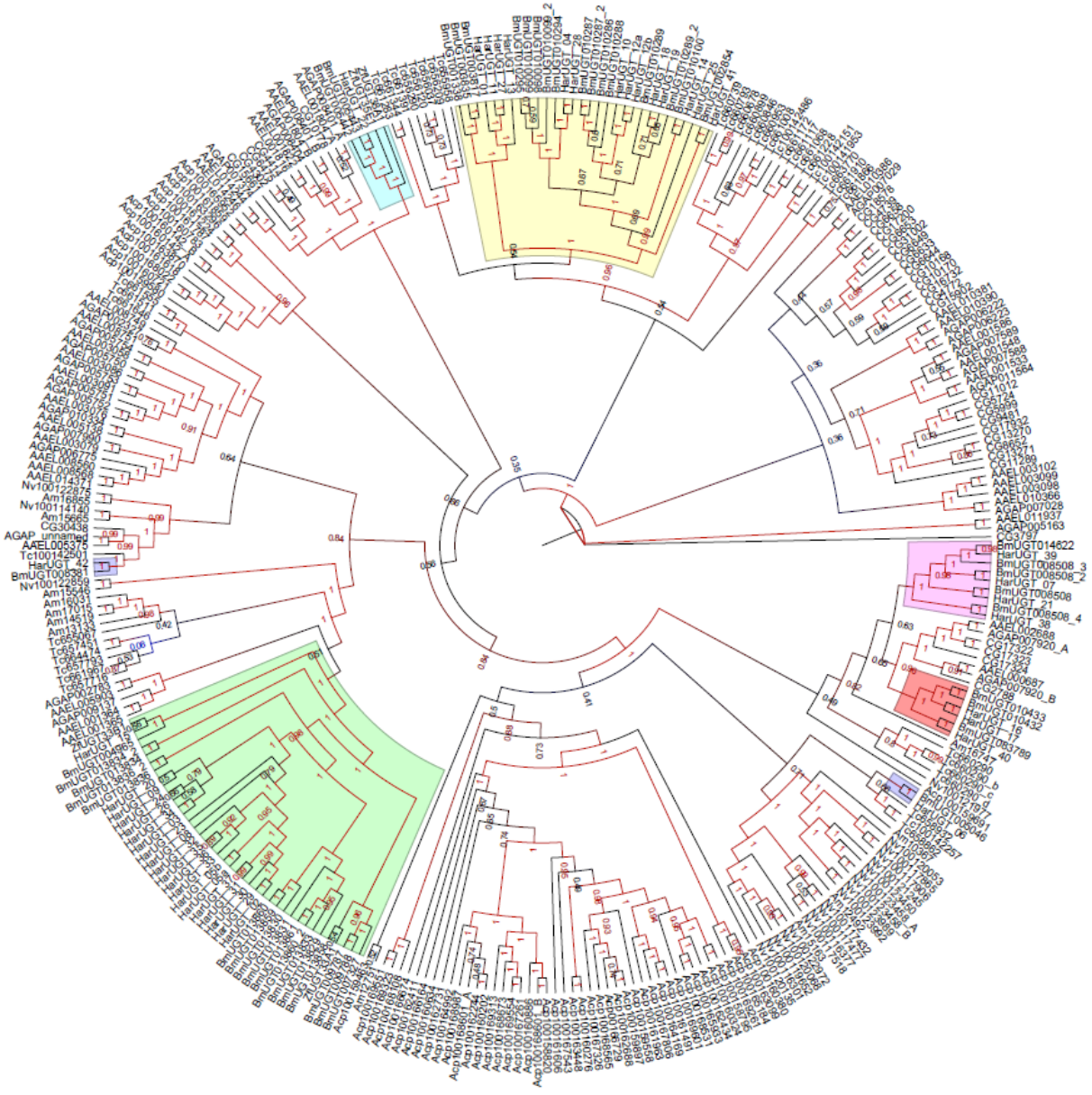


Figure 1. Phylogenetic tree of the insect UGTs. Bayesian interference using a GTR+I+G nucleotide substitution model was implemented in Mr. Bayes 3.1 (Ronquist and Huelsenbeck, 2003). The tree was constructed using a *Drosophila* UGT sequence as outgroup. The Marcov Chain Monte Carlo runs were carried out for 1,000,000 generations, after which log likelihood values showed that equilibrium had been reached after the first 400 generations in all cases, and those data were discarded from each run and considered as "burnin". Two runs were conducted per dataset showing agreement in topology and likelihood scores.

3.2.2. Comparative Phylogenetic analysis of H. armigera *and* B. mori *UGTs*

A consensus phylogenetic tree constructed with only the *H. armigera* and *B. mori* UGTs revealed that the subfamilies share the same intron-exon structure and chromosomal organization (Fig. 2). The largest UGT cluster (**UGT33**) containing 21 *H. armigera* UGTs and 16 *B. mori* UGTs (all but one on BmChr28, i.e. Chromosome 28) accounts for 46 % and 36 % of all UGT numbers in each species, respectively. This subfamily shows a pattern of recent lineage-specific gene divergence; suggesting two independent gene expansions in *H. armigera* and one in *B. mori* driven by diversification in the N-terminal substrate binding domain, likely increasing the range of compounds that could be detoxified or regulated by glycosylation. UGT34A3-HarUGT_15 and UGT34A2- BGIBMGA004965 (on BmChr25) both occupy a basal position in this clade, suggesting that a gene duplication followed by transposition to BmChr28 was followed by further diversification of the latter group.

The second largest UGT cluster is found in **UGT40**, which is composed of 9 *H. armigera* UGTs and 12 *B. mori* UGTs consistently having 8 exons. Orthologous pairs are more in this clade, except for four recently diverged *B. mori* UGTs (UGT40B1-BGIBMGA010098, UGT40B2-BGIBMGA010099, UGT40B3-BGIBMGA010099-2, and UGT40B4-BGIBMGA010295). The *H. armigera* homolog of these four appears to be HarUGT_28, a partial sequence superimposed with a broken line on the tree (the partial sequence was not used in the construction of the tree).

UGT41 composed of 4 *H. armigera* UGTs (UGT41B1-HarUGT_01, UGT41B2-HarUGT_11, UGT41B3-HarUGT_27, and UGT41D1-HarUGT_13 and 3 *B. mori* UGTs (UGT41A1-BGIBMGA001138, UGT41A3-BGIBMGA003835, and UGT41A2-BGIBMGA003817) occupies a separate position in the phylogenetic tree, consistently containing 9 exons. **UGT48**A1-HarUGT_41 and UGT48C1-BGIBMGA002854 (on BmChr10) both occupy a basal position and may be more ancient than the other UGTs in this clade.

UGT39 is composed of two and two UGTs from both species, having 5 exons. Since HarUGT_43 was partially annotated due to incompleteness in the genome assembly, it was superimposed on the tree by sequence similarity after the tree was constructed.

UGT46 is well conserved between the two species, which is composed of 3 and 3 genes from each species. UGT46A4-HarUGT_17 and UGT46A3-HarUGT_16 pair with UGT46A1-BGIBMGA010432 and UGT46A2-BGIBMGA010433 in terms of phylogeny as well as chromosomal orientation. These two genes share three exons corresponding sugar binding C-terminal domain, whereas have different first exons that are endoding substrate binding N-terminal domain. **UGT43**, **UGT44**, and **UGT47** are composed of single gene from each species. Gene structure featured by exon-intron distribution is unique in each family. **UGT42** comprises of two *H. armigera* and three *B. mori* UGTs, where UGT42C1-HarUGT_39 pairs with two *B. mori* orthologs, UGT42A1-BGIBMGA008508-3 and UGT42A2-BGIBMGA014622. Therefore, we even expect one more *H. armigera* UGT to be found in this cluster (see below for chromosomal location).

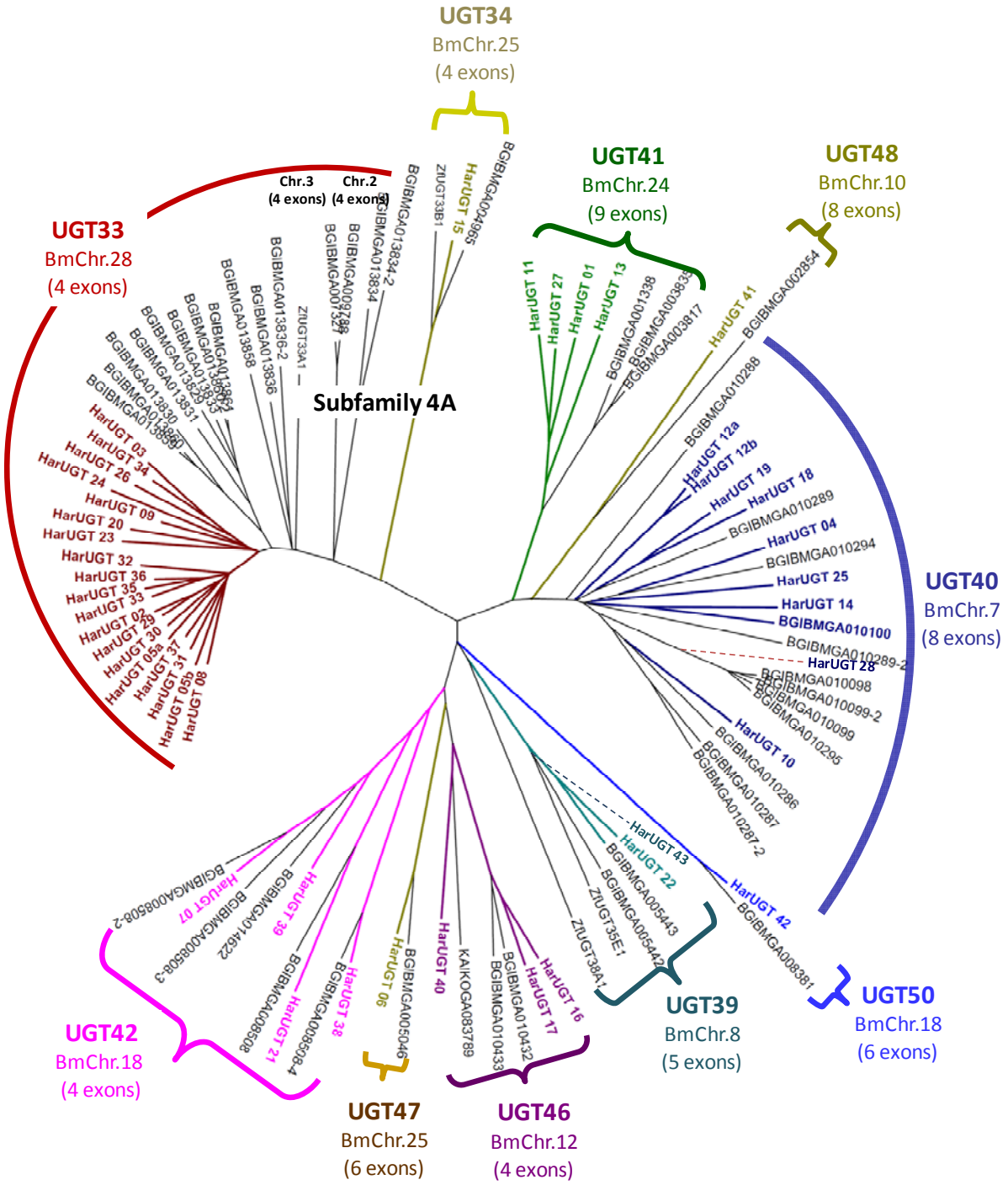


Figure 2. Phylogenetic tree of the *H. armigera* and *B. mori* UGTs. Neighbor-Joining method was used and bootstrap test (1000 replicates) was conducted. A sub-tree constructed with 46 *H. armigera* UGTs and 45 *B. mori* UGTs revealed that phylogenetic groupings of the lepidopteran UGTs are consistent with the intron-exon structure and chromosomal organization.

UGT50 is a conserved family among at least eight insect species, comprised of single orthologs from each species except *Ac. pisum.* It includes UGT50A2-HarUGT_42 and UGT50A1-BGIBMGA008381, as well as UGT50B1-AAEL005375 (*Ae. aegypti*), UGT50B2-AGAP-unnamed (*An. gambiae*), UGT50B3-CG30438 (*D. melanogaster*), UGT50D1-TcLOC100142501 (*T. castaneum*), UGT50C1-Am15665 (*Ap. mellifera*), and UGT50C2-NvLOC100114140 (*N. vitripennis*). The overall pairwise identity of all these UGTs across all 8 insects is over 50 %, suggesting the members of this conserved group have been preserved through evolution. Since a human homolog (UGT8A1) of this conserved group showed high protein sequence similarities with UGT50A2-HarUGT_42 (29 %) and with UGT50A1-BGIBMGA008381 (28 %), we propose that UGT8 is an orthologous family of the insect UGT50. The human UGT8 is known as a 2-hydroxyacylsphingosine 1-β-galactosyltransferase, or a ceramide UDP-galactosyltransferase, catalyzing the galactose conjugation of a sphingosine (Sprong et al., 1998). Sphingosine (a ceramide) is not only a membrane lipid component, but has been recently recognized as an important endogenous regulator of apoptosis and basic cellular processes (Ruvolo, 2003), suggesting this conserved family within insects may be also responsible for the sphingosine modulation.

On the other hand, it is noteworthy that there is no homolog from the pea aphid. Moreover, the predicted *Nasonia vitripennis* homolog (UGT50C2-NvLOC100114140) lacks 32 aa residues in the middle of the substrate binding domain compared to the other homologs, and the same truncated homologs were also found in the other two congeneric species, *N. giraulti*, and *N. longicornis*. In spite of the truncation, all the homologs from the three species of *Nasonia* wasp were predicted to produce full length proteins because no frame shift occurs. However, the missing region corresponds to the beginning of exon 3 and there is a gap in the genome assembly at this position in all three species.

3.3. Chromosomal location of *B. mori* UGTs

3.3.1. Mapping the orphan and new B. mori *UGT genes.*

Bombyx mori UGT genes were previously shown to be distributed among 10 different chromosomes, with three genes still unmapped (Huang et al., 2008). We found additional evidences allowing assignment of the three previously orphaned genes to specific chromosomal locations (Fig. 3); UGT42A2-BGIBMGA014622 is positioned on BmChr18, UGT41A1-BGIBMGA001338 on BmChr24, and UGT50A1-BGIBMGA008381 (previously BmUGT1566) on BmChr18. These re-positionings were supported by ESTs and BAC ends as mapped onto the genome in Kaikobase (http://sgp.dna.affrc.go.jp/KAIKObase). For example, each pair of the 5 BAC ends used in the orphan scaffold assembly of UGT42A2-BGIBMGA014622 is found on BmChr18 and located very near to the UGT gene cluster on this chromosome, and this orphan scaffold (~ 12.3 Kb) probably positions in a large genomic gap between UGT42A1-BGIBMGA008508-3 and UGT44A1-BGIBMGA008508-4. Another orphan scaffold (~ 38 Kb) containing UGT41A1-BGIBMGA001338 was positioned in a large

gap (~ 31 Kb) between UGT41A3-BGIBMGA003835 and UGT41A2-BGIBMGA003817 on BmChr24, because exon8 and exon9 of UGT41A3-BGIBMGA003835 were found at the end of the orphan scaffold and it is supported by an EST sequence (DB674183) found in NCBI GenBank. UGT50A1-BGIBMGA008381 was located in a similar way. In addition, three novel *B. mori* UGTs were also positioned; UGT46A1-BGIBMGA010432 is found in upstream region of UGT46A2-BGIBMGA010433 with common sharing exon 2, 3, and 4 on BmChr12, UGT46C2-KAIKOGA083789 is at the other end of BmChr12, and UGT50A1-BGIBMGA008381 is on BmChr18.

UGT33R2-BGIBMGA009788 (BmChr2) and UGT33R1-BGIBMGA007327 (BmChr3) seem to be incorrectly positioned in the genome as currently depicted in Kaikobase. UGT33R1-BGIBMGA007327, previously indicated on BmChr3 instead seems more likely to lie next to UGT33R2-BGIBMGA009788 on BmChr2, because a BAC end lying within UGT33R1-BGIBMGA007327 is found near to UGT33R2-BGIBMGA009788. Since the genomic congruency around UGT33R2-BGIBMGA009788 is higher than that around UGT33R1-BGIBMGA007327, it seems to be likely that UGT33R1-BGIBMGA007327 should be positioned next to UGT33R2-BGIBMGA009788 rather than vice versa. Therefore, based on these data, the most congruent picture is that *B. mori* UGTs are distributed on 9 different chromosomes.

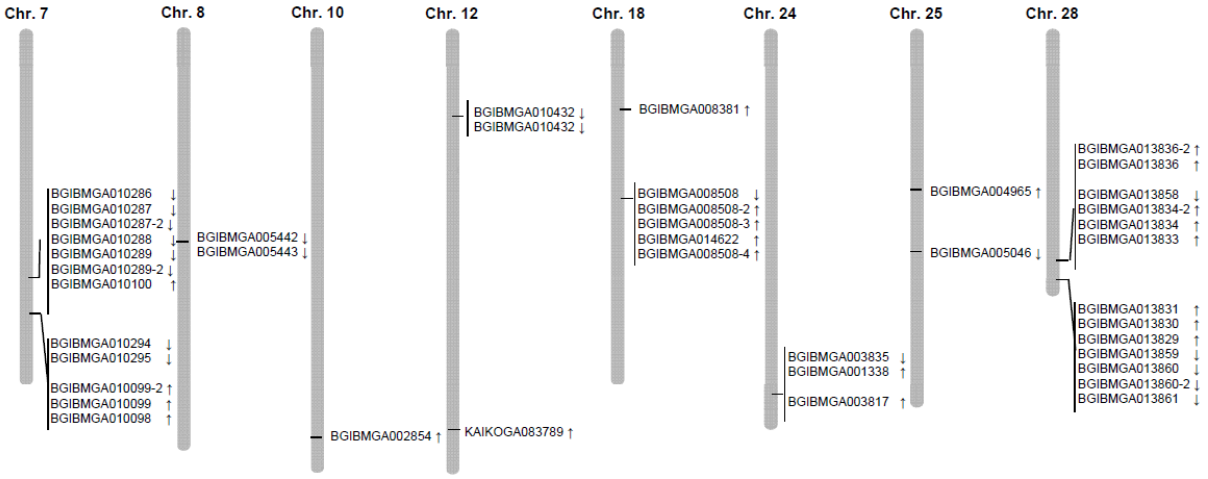


Figure 3. Chromosomal locations of the *B. mori* UGT genes.

3.3.2. B. mori *UGT gene clusters*

There are two large UGT clusters in the *B. mori* genome; 13 UGTs on BmChr28 and 12 UGTs on BmChr7 (Fig. 3). These 25 UGTs account for 56 % of all the UGT numbers in *B. mori* found. BmChr18 has also a UGT cluster lined up with 5 UGTs and BmChr24 contains a 3-UGT cluster. The

chromosomal location of *B. mori* UGT genes is almost consistent with Huang et al. (2008), except the gene orientation of the UGT cluster on BmChr18 and some newly mapped UGTs. The *B. mori* UGT gene clusters were used as a reference to construct *H. armigera* UGT gene clusters as explained below.

3.4. Microsynteny between *H. armigera* and *B. mori* UGT gene clusters

3.4.1. Syntenic relationship

The syntenic relationship between *H. armigera* and *B. mori* genomes was analyzed at the regions of UGT gene clusters as well as neighboring other genes. The genomic orientation of *H. armigera* UGT genes is highly conserved with the chromosomal location of *B. mori* UGT genes. In one of the largest UGT clusters of *H. armigera* corresponding to the one in *B. mori* BmChr7 (Fig. 4a), for example, 7 UGTs are tandemly arranged matching one by one, although three of them (UGT40F2-HarUGT_19, UGT40M1-HarUGT_14, and UGT40F1-HarUGT_18) are oriented in reverse compared to those in *B. mori*. UGT40Q1-HarUGT_04 and HarUGT_28 are not clustered, but the phylogenetic analysis and sequence similarity indicate that they seem to be located in this cluster. As another example of the UGT cluster, 4 *H. armigera* UGT genes reside in tandem, comparable to the respective genomic region of BmChr18, where 5 *B. mori* UGT genes are assembled together with an orphan scaffold containing UGT42A2-BGIBMGA014622 (Fig. 4b). Fragments of an additional UGT were found between UGT43A1-HarUGT_21 and UGT42B2-HarUGT_07, which may be due to incompleteness of the current assembly. Together with a neighbor gene (Tim10/DDP), the synteny of these genomic regions is highly conserved among these two Lepidoptera.

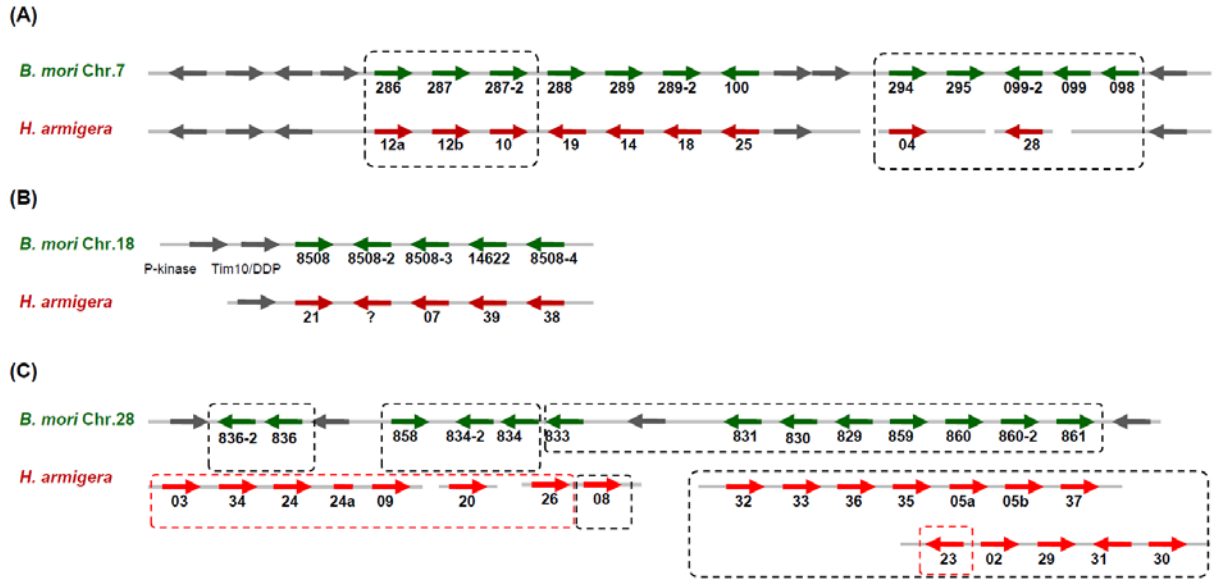


Figure 4. Microsynteny between *H. armigera* and *B. mori* genomes in UGT gene clusters. Three biggest gene clusters from (A) BmChr7, (B) BmChr18, and (C) BmChr28.

Such a high syntenic relationship between the two lepidopteran species identified here was previously studied on a larger scale in a comparison of *H. armigera* BAC sequences with the corresponding region of *B. mori* genome (d'Alencon et al., 2010). Although the analysis has scanned nearly 300 genes across 15 genomic regions distributed on 11 *B. mori* chromosomes, no UGTs were included in that analysis. Therefore, our present study can complement the syntenic comparison, which might be ultimately helpful in the *H. armigera* genome assembly in the future.

3.4.2. Gene duplication in action

In addition to the high levels of synteny in these regions, there are other examples of UGT clusters displaying lineage-specific gene divergence by duplication (UGT33 in Fig. 2, for example). Gene duplication events often occur under the physical constraint of the chromosomal region. Duplicated genes are usually located next to the parental gene; this is how gene clusters have been structured. As an example, the 13 UGT genes of the largest *B. mori* UGT clusters on BmChr28 are subtly segregated into three sub-clusters according to their local position in this chromosomal region. UGT33N1-BGIBMGA013836 and UGT33K1-BGIBMGA013836-2 are positioned at the very distal part, UGT33D3-BGIBMGA013833, UGT340C2-BGIBMGA013834, UGT340C1-BGIBMGA013834-2, and UGT33Q1-BGIBMGA013858 are organized in the middle, and the other 7 UGTs are in tandem on the other end. Such a physical sub-clustering is in almost complete accordance with the phylogenetic relationships, except for BGIBMGA013833 belonging to the middle sub-cluster, which is found in other phylogenetic sub-group. This suggests that UGT33D3-BGIBMGA013833 may have been a 'seed' of the gene duplication leading the local expansion of UGT genes next to it.

Similarly, the largest *H. armigera* UGT cluster with 20 UGT genes, including a pseudogene (HarUGT_24aP), corresponding to the largest *B. mori* UGT cluster on BmChr28, can be divided into two sub-groups in the phylogenetic tree (Fig. 2). Although the physical clustering of the 20 UGTs is not completed (Fig. 4c), it wouldn't be much deviated from the phylogenetic grouping, meaning that two large sub-groups are expected.

On the other hand, phylogenetically distant genes can be found in the same chromosome, but not that closely located. First, both UGT34A2-BGIBMGA004965 (corresponding to UGT34A3-HarUGT_15) and UGT47A1-BGIBMGA005046 (corresponding to UGT47A2-HarUGT_06) are found on BmChr25 (Fig. 3). The physical distance between two is about 2.4 Mb, which is quite far to be physically correlated, and intron-exon structures are different; the former has 4 exons whereas the latter has 6 exons. Second, UGT50A1-BGIBMGA008381 (UGT50A2-HarUGT_42 ortholog) is located about 4.7 Mb away from a cluster comprised of five *B. mori* UGTs (UGT43B1-BGIBMGA008508, UGT42B1- BGIBMGA008508-2, UGT42A1- BGIBMGA008508-3, and UGT44A1- BGIBMGA008508-4; and UGT42A2-BGIBMGA014622) on BmChr18, but the phylogenetic relationship is quite distant (the former has 6 exons whereas the latter four have 4 exons).

3.5. Intron-exon organization

3.5.1. Intron numbers

H. armigera and *B. mori* UGTs have four to nine introns each. No intronless UGT genes were found in either species. Regarding intron numbers per gene, the three-intron structure accounts for the majority in the UGTs of these two species; 64 % in *H. armigera* and 59 % in *B. mori* UGTs in number. In detail, each of them has a consistent gene structure having a long exon 1, coding the N-terminal substrate binding domain, and having relatively short exon 2, 3 and 4. Genes with 7 introns take the second majority having gained additional 3 introns in the N-terminal region.

All of the 321 insect UGT genes we found contain introns, except for two intronless sequences, UGT37C2-CG13270 and UGT37A1-CG11012 from *D. melanogaster*. As a whole, the majority of UGT genes possess 3 introns (42.6%) in their coding sequences, followed by 4 introns and 2 introns. Species differ in the modal intron number, for example most *D. melanogaster* UGTs contain one intron (56 %); UGTs of the mosquitoes *An. gambiae* and *Ae. aegypti* mainly have 2 introns (46 and 41 %, respectively); most *Ac. pisum*, *T. castaneum*, *B. mori*, and *H. armigera* UGTs contain 3 introns (70, 56, 59, and 60 %, respectively); and two Hymenoptera, *N. vitripennis* and *Ap. mellifera*, UGTs have 4 introns most frequently (68 and 50 %, respectively). The intron numbers seem to be related to insect order, meaning Diptera has 1-2 introns mainly, Hemiptera, Coleoptera and Lepidoptera have 3 introns usually, and Hymenoptera has 4 introns as a major. The distribution of intron numbers in each species is unimodal, except for the two Lepidoptera species where a second peak occurs at 7 introns. The 10 *H. armigera* UGTs and 13 *B. mori* UGTs containing 7 introns together with 4 *H. armigera* UGTs and 3 *B. mori* UGTs containing 8 introns constitute separate groups, both within insects in general and within Lepidoptera.

3.5.2. Intron positions and phases

The analysis of the intron-exon organization in *H. armigera* and *B. mori* UGTs revealed that their intron positions and splicing site phases were highly conserved (Fig. 5). Especially in the C-terminal halves are there thoroughly conserved intron positions and phases, whereas in the N-terminal might intron gain has happened, producing additional diversity at the substrate binding domains. A minimum of eight independent intron insertion events appear to have happened in the course of evolution of *H. armigera* and *B. mori* UGT gene families. The three intron positions (position 5, 6 and 7 marked with inverted triangle in Fig. 5) are conserved among all most members of UGTs from both species, with position 7 absent only in three UGTs from each species (UGT39B2-HarUGT_22, HarUGT_28, and UGT50A2-HarUGT_42; UGT39C1-BGIBMGA005442, UGT39B1-BGIBMGA005443, and UGT50A1-BGIBMGA008381), which seem to have lost the introns. UGTs with 7 introns and 8 introns shared all conserved splicing sites except the additional intron position at the very end of the

C-terminal. The last introns found in 8 intron-UGTs seem to have been gained. The splicing site phases were also highly conserved at each consensus intron position among the UGTs within species as well as between species. The consensus intron sites 1 and 5 are phase 0, intron site 3, 4, and 7 are phase one, and intron site 2, 6, and 8 are phase two. The orthologous pairs UGT47A2-HarUGT_06 and UGT47A1-BGIBMGA005046; and UGT50A2-HarUGT_42 and UGT50A1-BGIBMGA008381 violated the consistency of intron positions and phases in their N-terminal substrate binding domains. In these regions their sequences are more divergent from the majority of other UGTs, suggesting that novel exons may have been created from previously intronic sequences, further increasing substrate binding diversity.

3.6. Conclusions

We identified over 320 insect UGT genes from the genome databases of 9 different insect species, including 46 UGTs from *H. armgiera.* A comparative genomic analysis revealed the extent of the UGT multigene family in insects. Phylogenetic analyses showed lineage-specific gene diversification of some UGT subfamilies in insects, as well as conservation of a certain subfamily across many organisms. UGTs from the two lepidopteran species shared the same intron-exon structures and genomic organization in their subfamilies, whereas lineage-specific gene expansion was also characterized. Sequence alignments predicted important structural domains and catalytic residues. This is the first genome-wide study on a gene family from *H. armigera*. Knowledge of the insect UGT sequences and organization in genome might shed light on the future research of the multigene family in combination with xenobiotic detoxification and endobiotic modification in insects.

Acknowledgements

This study was funded by the Max Planck Society. Seung-Joon Ahn holds a scholarship of the IMPRS programme of the Max Planck Society. We thank the Helicoverpa Genome Consortium for allowing us to use *H. armigera* genome sequences in this manuscript. We thank Steffi Gebauer-Jung for database maintenance.

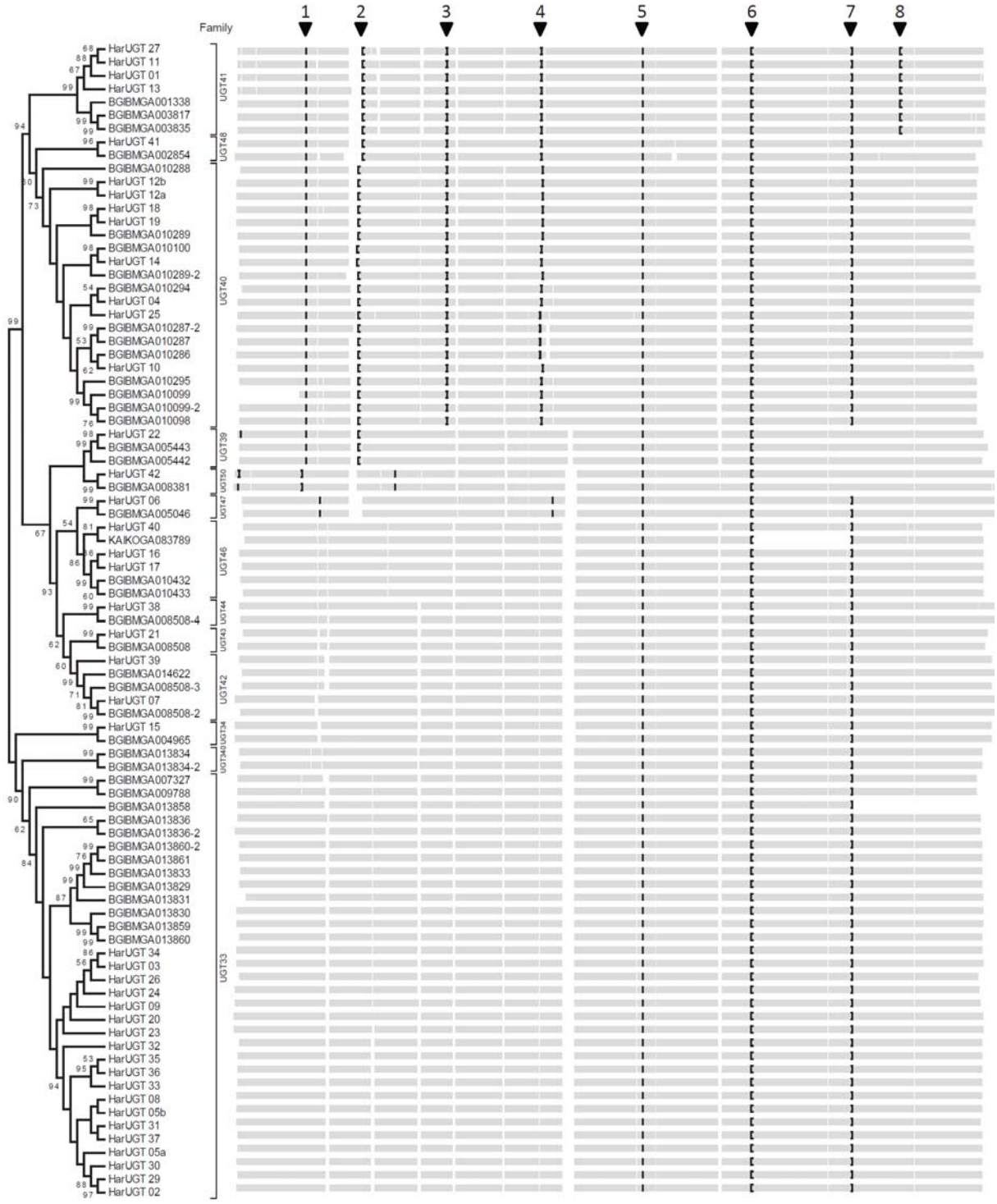


Figure 5. Unrooted NJ phylogenetic tree and intron map of 87 UGT genes from *H. armigera* and *B. mori*. The bootstrap value greater than 50% is indicated for each branch and the subfamilies are shown next to gene name. Protein sequence alignment is shown in gray bar with gaps in white, and the intron position of sequences is marked with (|) for a phase 0 intron, ([ ) for a phase 1 intron and ( ]) for a phase 2 intron. Inverted triangles combined with numbers indicate intron positions that are commonly found in the corresponding genes.

References

Ahmad, S., Brattsten, L.B., Mullin, C.A., Yu, S.J., 1986. Enzymes involved in the metabolism of plant allelochemicals, in: Brattsten, L.B., Ahmad, S. (Eds.), Molecular aspects of insect-plant associations. Plenum Press, New York & London, pp. 73-151.

Ahmad, S.A., Hopkins, T.L., 1992. Phenol β-glucosyltransferase and β-glucosidase activities in the tobacco hornworm larva *Manduca sexta* (L.): Properties and tissue localization. Arch. Insect Biochem. Physiol. 21, 207-224.

Ahmad, S.A., Hopkins, T.L., 1993a. β-Glucosylation of plant phenolics by phenol β-glucosyltransferase in larval tissues of the tobacco hornworm, *Manduca sexta* (L.). Insect Biochemistry and Molecular Biology 23, 581-589.

Ahmad, S.A., Hopkins, T.L., 1993b. Phenol β-glucosyltransferases in six species of insects: properties and tissue localization. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 104, 515-519.

Ahmad, S.A., Hopkins, T.L., Kramer, K.J., 1996. Tyrosine β-glucosyltransferase in the tobacco hornworm, *Manduca sexta* (L): Properties, tissue localization, and developmental profile. Insect Biochemistry and Molecular Biology 26, 49-57.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340, 783-795.

Bock, K.W., 2003. Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. Biochem Pharmacol 66, 691-696.

Bowles, D., Isayenkova, J., Lim, E.K., Poppenberger, B., 2005. Glycosyltransferases: managers of small molecules. Curr. Opin. Plant Biol. 8, 254-263.

Brattsten, L.B., 1988. Enzymic adaptations in leaf-feeding insects to host-plant allelochemicals. J Chem Ecol 14, 1919-1939.

Claudianos, C., Ranson, H., Johnson, R.M., Biswas, S., Schuler, M.A., Berenbaum, M.R., Feyereisen, R., Oakeshott, J.G., 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Mol. Biol. 15, 615-636.

d'Alencon, E., Sezutsu, H., Legeai, F., Permal, E., Bernard-Samain, S., Gimenez, S., Gagneur, C., Cousserans, F., Shimomura, M., Brun-Barale, A., Flutre, T., Couloux, A., East, P., Gordon, K., Mita, K., Quesneville, H., Fournier, P., Feyereisen, R., 2010. Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. P Natl Acad Sci USA 107, 7680-7685.

Daimon, T., Hirayama, C., Kanai, M., Ruike, Y., Meng, Y., Kosegawa, E., Nakamura, M., Tsujimoto, G., Katsuma, S., Shimada, T., 2010. The silkworm Green b locus encodes a quercetin 5-O-glucosyltransferase that produces green cocoons with UV-shielding properties. P Natl Acad Sci USA 107, 11471-11476.

Després, L., David, J.P., Gallet, C., 2007. The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol. Evol. 22, 298-307.

Fitt, G.P., 1989. The ecology of *Heliothis* species in relation toagroecosystems. Annual Review of Entomology 34, 17-53.

Heckel, D.G., 2010. Molecular genetics of insecticide resistance in Lepidoptera, in: Goldsmith, M.R., Marec, F. (Eds.), Molecular biology and genetics of the Lepidoptera. Taylor & Francis, pp. 239-269.

Heckel, D.G., Gahan, L.J., Daly, J.C., Trowell, S., 1998. A genomic approach to understanding *Heliothis* and *Helicoverpa* resistance to chemical and biological insecticides. Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci. 353, 1713-1722.

Honeybee Genome Sequencing, C., 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931-949.

Hopkins, T.L., Kramer, K.J., 1992. Insect cuticle sclerotization. Annual Review of Entomology 37, 273-302.

Huang, F.F., Chai, C.L., Zhang, Z., Liu, Z.H., Dai, F.Y., Lu, C., Xiang, Z.H., 2008. The UDP-glucosyltransferase multigene family in *Bombyx mori*. BMC Genomics 9, 563.

Luque, T., O'Reilly, D.R., 2002. Functional and phylogenetic analyses of a putative *Drosophila melanogaster* UDP-glycosyltransferase gene. Insect Biochemistry and Molecular Biology 32, 1597-1604.

Luque, T., Okano, K., O'Reilly, D.R., 2002. Characterization of a novel silkworm (*Bombyx mori*) phenol UDP-glucosyltransferase. Eur. J. Biochem. 269, 819-825.

Magdalou, J., Fournel-Gigleux, S., Ouzzine, M., 2010. Insights on membrane topology and structure/function of UDP-glucuronosyltransferases. Drug Metabolism Reviews 42, 159-166.

Miley, M.J., Zielinska, A.K., Keenan, J.E., Bratton, S.M., Radominska-Pandya, A., Redinbo, M.R., 2007. Crystal Structure of the Cofactor-Binding Domain of the Human Phase II Drug-Metabolism Enzyme UDP-Glucuronosyltransferase 2B7. Journal of Molecular Biology 369, 498-511.

Morello, A., Repetto, Y., 1979. UDP-glucosyltransferase activity of housefly microsomal fraction. Biochem. J. 177, 809-812.

Oakeshott, J.G., Johnson, R.M., Berenbaum, M.R., Ranson, H., Cristino, A.S., Claudianos, C., 2010. Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. Insect Mol. Biol. 19, 147-163.

Osmani, S.A., Bak, S., Møller, B.L., 2009. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70, 325-347.

Radominska-Pandya, A., Bratton, S.M., Redinbo, M.R., Miley, M.J., 2010. The crystal structure of human UDP-glucuronosyltransferase 2B7 C-terminal end is the first mammalian UGT target to be revealed: the significance for human UGTs from both the 1A and 2B families. Drug Metabolism Reviews 42, 133-144.

Ramsey, J.S., Rider, D.S., Walsh, T.K., De Vos, M., Gordon, K.H.J., Ponnala, L., Macmil, S.L., Roe, B.A., Jander, G., 2010. Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. Insect Mol. Biol. 19, 155-164.

Real, M.D., Ferré, J., Chapa, F.J., 1991. UDP-glucosyltransferase activity toward exogenous substrates in *Drosophila melanogaster*. Anal Biochem 194, 349-352.

Ruvolo, P.P., 2003. Intracellular signal transduction pathways activated by ceramide and its metabolites. Pharmacological Research 47, 383-392.

Sasai, H., Ishida, M., Murakami, K., Tadokoro, N., Ishihara, A., Nishida, R., Mori, N., 2009. Species-specific glucosylation of DIMBOA in larvae of the rice armyworm. Biosci Biotech Bioch 73, 1333-1338.

Seppen, J., Steenken, E., Lindhout, D., Bosma, P.J., Oude Elferink, R.P.J., 1996. A mutation which disrupts the hydrophobic core of the signal peptide of bilirubin UDP-glucuronosyltransferase, an endoplasmic reticulum membrane protein, causes Crigler-Najjar type IIs. Febs Lett 390, 294-298.

Smith, J.N., 1962. Detoxication mechanisms. Annual Review of Entomology 7, 465-480.

Strode, C., Wondji, C.S., David, J.P., Hawkes, N.J., Lumjuan, N., Nelson, D.R., Drane, D.R., Karunaratne, S., Hemingway, J., Black, W.C., Ranson, H., 2008. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. Insect Biochemistry and Molecular Biology 38, 113-123.

Svoboda, J., Weirich, G., 1995. Sterol metabolism in the tobacco hornworm, *Manduca sexta* - A review. Lipids 30, 263-267.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596-1599.

Teese, M.G., Campbell, P.M., Scott, C., Gordon, K.H.J., Southon, A., Hovan, D., Robin, C., Russell, R.J., Oakeshott, J.G., 2010. Gene identification and proteomic analysis of the esterases of the cotton bollworm, *Helicoverpa armigera*. Insect Biochemistry and Molecular Biology 40, 1-16.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

Tijet, N., Helvig, C., Feyereisen, R., 2001. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: Annotation, intron-exon organization and phylogeny. Gene 262, 189-198.

Tsubota, T., Shiotsuki, T., 2010. Genomic analysis of carboxyl/cholinesterase genes in the silkworm *Bombyx mori*. BMC Genomics 11.

Wang, Q., Hasan, G., Pikielny, C.W., 1999. Preferentialexpression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. J Biol Chem 274, 10309-10315.

Wang, X., 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. Febs Lett 583, 3303-3309.

Wiesen, B., Krug, E., Fiedler, K., Wray, V., Proksch, P., 1994. Sequestration of host-plant-derived flavonoids by lycaenid butterfly *Polyommatus icarus*. J Chem Ecol 20, 2523-2538.

Yu, Q.Y., Lu, C., Li, B., Fang, S.M., Zuo, W.D., Dai, F.Y., Zhang, Z., Xiang, Z.H., 2008. Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*. Insect Biochemistry and Molecular Biology 38, 1158-1164.

4. OVERALL DISCUSSION

Herbivorous insects have specialized on their host plants by developing various mechanisms to cope with plant defense compounds. Among many other metabolic detoxification mechanisms, glucose conjugation by UGT plays an important role in modifying exogenous or endogenous compounds in facilitating excretion of toxic compounds from the body or to utilize them for other purposes. However, the UGTs have received less attention for a long period of time than other detoxification enzymes. The present thesis provides an example of UGT-mediated detoxification by using a model system, capsaicin and *Capsicum*-feeding specialist moth, *H. assulta*. The results of this study establish that the host specialist is more tolerant to capsaicin than other noctuid species (Manuscript I). Caterpillars excrete capsaicin glucoside in their feces when they are fed capsaicin, and the metabolism is catalyzed by UGT enzymes (Manuscript II). Genomic analysis of the UGT multigene family from *Helicoverpa armigera* genome as well as other insect genomes gives an evolutionary insight of the detoxicative enzyme family (Manuscript IV). Furthermore, the transcript diversity of the *H. armigera* UGTs shows dynamic patterns of the multigene's expression (Manuscript III). Here I provide an overall discussion of these individual topics within a frame of the major thesis theme.

4.1. UGT: an emerging enzyme family in insect detoxification study

UGTs belong to the glycosyltransferase superfamily (GT, Enzyme Commission number EC 2.4). In general, GTs play a key role in glycosidic bond formation to a variety of biomolecules, like carbohydrates, protein residues, lipids, or small hydrophobic molecules. This superfamily GT is composed of more than 90 families according to their amino acid sequence similarities. GTs can be also classified into either retaining or inverting enzymes depending on the stereochemistry of the substrates and reaction products, or can be classified into either GT-A or GT-B folds by their three-dimensional structures. UGTs are a member of the biggest GT1 family as well as a member of the inverting group and the GT-B fold type, mainly catalyzing the conjugation of UDP-sugar to various small hydrophobic compounds.

UGTs, as the name implies, utilize UDP-glycoside as a sugar donor. UDP-glucuronic acid (UDPGA) is mainly used by mammalian UGTs, whereas UDP-glucose (UDPG) is a predominant sugar donor in plants and insects. The formation of glucuronide was unsuccessful in insects (Smith, 1968). Although very minute activity was detected in a certain insect species, it is not known whether the glucuronide formation in insects was due to structural similarity with UDPG or the results of the presence of UDP-glucuronosyltransferase in insects (Ahmad and Hopkins, 1993). This difference between insect UGTs and mammalian ones could be important in pest control, because such a

mechanism could be useful for the development of insect-specific control agents. Verification of this difference in co-substrate usage on the molecular level would shed light on the mechanism of the co-substrate specificity in insect.

Insect UGTs play important roles primarily in the detoxification of xenobiotics, and also in other physiological roles like sequestration, pigmentation, olfaction, and endobiotic regulation. However, insect UGTs are relatively unknown and have received less attention, compared to other detoxicative enzymes like P450s and GSTs (Després et al., 2007). In this situation, the genomic information acquired from this study could fuel sequential investigations into this enzyme family. Functional expression and characterization of individual genes could be addressed by using the sequences identified. Phylogenetic analysis could give further clues about genes of interest especially in the UGT gene families where lineage-specific gene diversification or interspecific-conservation is observed.

A gene family is referred to groups of genes showing similarity with each other, reflecting the assumption that all arose from a common ancestor. Gene families are derived from a direct consequence that all new genes arise by gene duplication. Therefore, gene families provide important information on how new genes arise and diversify, how the genes in certain families are regulated, and also how novel evolutionary forces are generated by such multiple gene copies, which might have not seen by considering individual genes (Walsh and Stephan, 2008). Likewise, the *H. armigera* UGT multigene family provides evolutionary perspectives; for instance, one of the most interesting features is the lineage-specific gene expansion in the UGT33 family. an independent lineage-specific expanision within UGT33 is also found in *B. mori*. This seems to have occurred very recently driven by diversification in the N-terminal substrate binding domain, thereby increasing the range of compounds that could be metabolized by glycosylation. On the other hand, the C-terminal domain is highly similar among these subfamily members; furthermore a segment (ca. 18 aa) in front of the transmembrane domain is completely conserved at the nucleotide sequence level. In addition, the fact that one of genes in this subfamily is highly expressed in antennae could add another clue about the functional significance in this gene cluster.

This study revealed the most extensive genome-wide investigation of the insect UGT multigene family from nine insects including *H. armigera* to date; the identification and phylogenetic analyses of the gene sequences provided a diverse repertoire of the multigene family. In fact, the model system in this study, i.e. the interaction between a host specialist, *H. assulta*, and capsaicin from a host plant, employed the UGT-mediated detoxification system. Knowledge of the insect UGT sequences and organization in the genome will shed light on the future research of the multigene family in combination with xenobiotic detoxification and endobiotic modification in insects.

4.2. Nomenclature of UGTs

Classifying and naming more than 320 insect UGTs was a challenging task. Vertebrate UGTs are well characterized in their functions as well as genomic organization, thereby their nomenclature is relatively well established (Mackenzie et al., 2005). Besides the lack of functional information, the high degree of diversity of the insect UGTs made it more difficult to perform the classification and the nomenclature. While vertebrate UGTs from 17 species are distributed into only 4 families (UGT1, UGT2, UGT3, and UGT8) according to the official Nomenclature Committee, insect UGTs identified from nine species are putatively classified into 67 families based on the same nomenclature scheme. The great diversity of the insect UGT gene family implies the diverse interaction with a variety of substrates including manifold plant secondary compounds and endogenous molecules.

For the nomenclature of diverse insect UGT sequences, we first considered the extent to which the currently named insect sequences by the Committee with the published guidelines. We then proposed a preliminary naming scheme for the new sequences that follows these guidelines (Mackenzie et al., 1997; Ross et al., 2001). The criteria for family and subfamily definition seem to have changed over the years. We have used the numbers in Ross et. al. (2001) where families were defined at 40% amino acid sequence identity (aaID) or greater and subfamilies at 60% aaID or greater. Cut-off values are arbitrary; the actual decision depends on how a particular sequence clusters on a tree and not so much on the absolute sequence identity (Nelson, 2006). Hence, preliminary grouping was done using the program H-CD-HIT (Huang et al., 2010) at 40% and 60% sequence identity as cut-off thresholds, and provisional family and subfamily names were assigned on this basis. A neighbor-joining tree was constructed from the sequence alignment. Groups were examined for consistency, and groups on the borderline of 40% or 60% were examined using pairwise p-distances calculated by MEGA. In a few cases the family criterion of 40% was difficult to apply due to some pairwise comparisons being 41-42% while others were 38-39%, and the family criterion was relaxed to 37-39% if doing so created a coherent group on the neighbor-joining tree. Preliminary names were re-assigned and the entire process was repeated. Partial sequences were examined to ensure that they were not incorrectly grouped. At the end, the insect UGT sequences occupied 67 UGT families, from UGT33 to UGT50; from UGT301 to UGT349. Considering the surge of genomic data produced in the near future, the naming ranges will definitely expand up to 4-digit or even 5-digit names, which would make the current nomenclature system more complicated. The great abundance and diversity of sequences to be named was not anticipated by the developers of the nomenclature system, and eventually the entire framework might need to be revised to accommodate them all.

4.3. Capsaicin-*H. assulta* interaction: an unusual but unique model system

Capsaicin is a unique compound found only in *Capsicum* fruits. The hot peppers, *Capsicum* spp., were introduced into the Old World including Asian countries only about 500 years ago. However, *H.*

assulta is now the most destructive pest on the fruits of cultivated *Capsicum* plants at least in East Asia, especially in Korea, China and Japan. On the other hand, it is very rare to find any other insects feeding the fruits in the native continent, South America. This raises an open question, how *H. assulta* was able to colonize the *Capsicum* plant so quickly. There must be many factors to consider, like competition, predation, plant defense as well as insect counter defense. Although the laboratory feeding trials performed in this study address only limited aspects of these factors, capsaicin is likely to have been an important factor inhibiting the initial colonization of *Capsicum* by other noctuid species, except for the specialist on Solanaceae, *H. assulta*. Whether *H. assulta* underwent additional adaptation to *Capsicum* is unknown, but could be addressed if non-*Capsicum* feeding populations can be identified and compared.

In this perspective, this specialist could be a good model insect to study many related topics in terms of host specialization, such as host orientation, feeding preferences, ovipositional choices, adaptation, etc. As the common name of this species implies, the Oriental tobacco budworm also feeds on cultivated tobacco which was introduced in Asia at a similar period together with *Capsicum*. Therefore the same question could be asked: how the specialist has evolved to the noxious compound, for example, nicotine from the alternative host plant. In fact, nicotine detoxification has been an interesting research topic in entomology, although the detoxification mechanism is not clearly revealed yet. Studies using the relationships between *H. assulta* and its host plants as a model system could provide extensive and interesting advancement of knowledge on the investigation of insect-plant interactions.

4.4. Generalists versus specialists

Several generalist herbivores were compared in this study with *H. assulta*. Except for *H. armigera* which occasionally feeds on *Capsicum* fruits in the field, all the other insects tested have virtually nothing to do with *Capsicum* fruit or capsaicin in nature. The growth-inhibiting effect of capsaicin on *S. frugiperda*, *H. virescens*, *H. subflexa*, and *H. zea* was contrasted to the growth-promoting effect on *H. assulta*, whereas little effect on *H. armigera*, a generalist feeding occasionally *Capsicum* fruit, was observed. The generalist, as defined by a wide range of its hosts, might also be able to detoxify a broad spectrum of plant allelochemicals. On the other hand, the specialist, as defined by the narrow range of its hosts, may have specialized on its detoxification spectrum as well. Moreover, the positive effect of capsaicin on *H. assulta* suggests the compound seems to be beneficial for the growth, but it might be valid only in the particular experimental conditions where a non-optimal diet was used. Repeating the feeding experiments with a different common diet deliberately chosen to be sub-optimal for *H. armigera* would be one way to test this idea. Therefore, it is still not clear how capsaicin boosted larval performance of *H. assulta*. As shown in this study, host specialization of *H. assulta* seems to be related to the response to the host allelochemical. According to another study on

nutritional tests with gossypol, nicotine, tomatine, or capsaicin on *H. assulta* and *H. armigera* larvae, *H. assulta* is tolerant particularly to nicotine and capsaicin whereas *H. armigera* shows broad tolerance to the four allelochemicals (Dong et al., 2002). This study shows a good example of such a difference between a generalist and specialist from the viewpoint of detoxification.

4.5. Capsaicin glucoside: a novel metabolite from insects

We described the identification of a novel metabolite of capsaicin, capsaicin β-glucoside, in the feces of three *Helicoverpa* spp. larvae that had been fed on capsaicin-spiked artificial diet. This is the first report on the identification of any capsaicin metabolite in insects, and the first capsaicin glucoside in insects. We also found species-differential glucosylation activity; *H. assulta* conjugated capsaicin with glucose less readily than *H. armigera* and *H. zea*, which is consistent with UGT enzyme activity toward capsaicin from the whole larva homogenate *in vitro*. Fat body compared to the other four tissues was the main source of the enzyme activity in the three species consistently, but species-differential tissue distribution of the activity was also observed. Our study suggests that the putative detoxification mechanism of the capsaicin glucosylation is likely to be catalyzed by UGT, and that the activity seems to be stronger in the two generalists than the host-specialist, *H. assulta*. Further studies are necessary not only to determine which gene(s) is involved in the glucosylation of capsaicin, but also to understand its regulatory mechanism, in relation to the host-plant adaptation of herbivores.

4.6. What is the fate of capsaicin in the specialist?

Although the UGT-mediated glucosylation of capsaicin is less active in *H. assulta* than the other generalists, it unaltered capsaicin is excreted less in the specialist and a significant amount of capsaicin remains in the body compared to the generalists. This suggests that *H. assulta* has additional, yet to be discovered mechanisms for dealing with capsaicin, since it appears that UGT is not a major detoxicative enzyme for capsaicin in this species. Although some tolerance of capsaicin due to target-site insensitivity cannot be ruled out, there must be additional metabolism because most dietary capsaicin disappears from the body and yet is not excreted as the glucoside or the unaltered aglucone. Among other detoxicative enzymes are P450s, CCEs, GSTs, etc. In fact, capsaicin is known to be oxidized into hydroxyl capsaicin, capsaicin oxide, or capsaicin quinone by P450s (Chanda et al., 2008; Reilly and Yost, 2006; Surh and Lee, 1995) and it can be also hydrolyzed into vanillylamine and 8-methyl-6-*trans*-nonenoic acid by carboxylesterase in mammals (Chanda et al.,, 2008). The vanillylamine can be further modified to vanillin, vanillyl alcohol, or vanillic acid, and each of them can form a conjugate further. Bacterial strains isolated from the hot pepper plants are capable of degrading capsaicin as a carbon and energy source, or utilizing its hydrolyzed vanillylamine as a nitrogen source (Flagan and Leadbetter, 2006). In the hot pepper, capsaicin is oxidized to produce

capsaicin dimers by peroxidases (Díaz et al., 2004). Therefore, it is possible also in insects that other mechanisms underlying in such a high capsaicin processing capability could be revealed.

4.7. Power of comparative study

Comparative methods were used throughout this study. Six different species of noctuid insects were compared when the effects of dietary capsaicin were tested. Three *Helicoverpa* spp. were compared to evaluate the food consumption and utilization capacities as well as to determine the direct toxicity by injection experiment. The glucose conjugation and excretion of capsaicin were measured in the three species, and UGT enzyme activities were also compared among them. Genomic analysis of the UGT multigene family was even more extensively compared among nine different species of insects, providing a global view of this gene family for the first time. Comparison between *H. armigera* and *B. mori* UGT gene families gave detailed insights within the same insect Order. Such a comparative approach has provided a powerful tool to distinguish the specialist characteristics from the other insects. Genomic study of the UGTs can also benefit from the comparative method to reveal how genomes are built according to the biology of particular insects.

4.8. Concluding remarks

The present study provides some important and new findings in the field of metabolic detoxification in insects. (1) The effects of capsaicin on different insect species including *H. assulta* were investigated. The effects were specified by comparing various developmental measures like larval growth, development time, survival, and pupal weights among six different noctuid species. Nutritional indices were compared to characterize the effects on the food uptake in more detail. Capsaicin injection was conducted in order to distinguish direct effects in the internal cavity from those in the gut on the larval development. The acute toxicity of capsaicin was also compared by high-dose injection administration. (2) A capsaicin metabolite, capsaicin glucoside, in the feces of three *Helicoverpa* spp. was identified. Chemical identification of the metabolite was performed by various analytical methods, LC-MS, NMR, and enzymatic hydrolysis. Quantitative measurements of the metabolite from three different species gave an insight how differently generalists and the specialist metabolize capsaicin. Biochemical characterization of the crude enzymes involved in the glucose conjugation suggested that UGTs might be responsible for the reaction. (3) The UGT multigene family from *H. armigera* and other insects was identified and analyzed. The identification of the gene family members could be an initial step toward investigations of individual enzymes, their substrate specificities, and functional characterization. In addition, phylogenetic analysis of the gene family could shed light on evolutionary perspectives of the detoxicative and metabolic enzyme family in insects.

5. SUMMARY

Throughout their long evolutionary history, herbivorous insects have developed various strategies to cope with a variety of noxious compounds from host plants. Enzymatic detoxification is one of the important mechanisms to neutralize, degrade, or modify such plant allelochemicals. Among many other phase II metabolic enzymes, UDP-glycosyltransferases (UGTs) are involved in the conjugation of glucose to small hydrophobic molecules, facilitating excretion from the system or utilization in the body. Therefore, the insect UGTs are considered to play important roles in the detoxification of xenobiotics. Other physiological roles like sequestration, pigmentation, olfaction, and endobiotic regulation have also been reported in several insects. However, insect UGTs are relatively unknown in terms of biochemical mechanisms and genomic structure compared to other detoxicative enzymes like P450s and GSTs. This study revealed over 320 insect UGT genes from the genome databases of nine different insect species, including 46 UGTs from *H. armigera.* A comparative genomic analysis revealed the extent of the UGT multigene family in insects. Phylogenetic analyses showed lineage-specific gene diversification of some UGT subfamilies in insects, as well as conservation of a certain subfamily across many organisms. UGTs from the two lepidopteran species shared the same intron-exon structures and genomic organization, whereas lineage-specific gene expansion was also characterized. In addition, alternative splicing events such as alternative promoters, exon skipping and intron retention increased transcript diversity in *H. armigera* UGTs. Sequence alignments predicted important structural domains like the N-terminal substrate binding domain and the C-terminal UDP-sugar binding domain. Furthermore, other important structural motives like signal peptides, UGT signature motif, transmembrane domain, and catalytic residues were identified in detail. This is the first genome-wide study on a gene family from *H. armigera* as well as other insects.

Capsaicin is an alkaloid found only in *Capsicum* spp. (Solanaceae), which is responsible for the pungency of hot pepper fruits. The Oriental tobacco budworm, *Helicoverpa assulta* (Guenée) (Lepidoptera: Noctuidae), is one of the few insects that can successfully feed on hot pepper fruits. The relationship between *H. assulta* and hot pepper is thus a good model system for studying host specialization. Indeed, *H. assulta* was more tolerant to capaicin than other noctuid species. Capsaicin was even beneficial to the specialist's development in a particular experimental circumstance. This study suggests that capsaicin can act as a defensive allelochemical against insects which are not adapted to the *Capsicum* plant. Furthermore, the beneficial effect of dietary capsaicin on the larval development of *H. assulta* suggests a physiological role of capsaicin in the adaptation of *H. assulta* to a host on which it is the primary agricultural pest. This study guided the next investigation of the detoxification of capsaicin in insects.

Detecting specific metabolites provides a direct way to elucidate the mechanism of detoxification. As a result, capsaicin glucoside was identified from the feces of *Helicoverpa armigera*, *H. assulta* and *H. zea* that fed on capsaicin-supplemented artificial diet. The species differences in the excretion rates of unaltered capsaicin suggested that *H. assulta* has a higher capacity to deal with capsaicin than the other two generalists. However, the excretion rates of the capsaicin glucoside were higher in *H. armigera* and *H. zea* than *H. assulta*, suggesting the detoxification mechanism by glucose conjugation is more active in *H. armigera* and *H. zea* than *H. assulta*. Results of UGT assays with crude enzyme extracts and capsaicin were reconciled with *in vivo* glucosylation, suggesting that capsaicin glucosylation is not a major detoxification mechanism in *H. assulta*. Compared among five different larval tissues (labial glands, testes from male larvae, midgut, the Malpighian tubules, and fatbody) from the three species, the formation of the capsaicin glucoside by one or more UGTs is high in the fat body of all the three species as expected. Although the lower excretion rate of the unaltered capsaicin in *H. assulta* indicates a higher metabolic capacity toward capsacin than in the other two generalists, the glucosylation *per se* seems to be insufficient to explain the decrease of capsaicin in the specialist, suggesting *H. assulta* might have another important mechanism to deal with capsaicin more specifically.

Detoxicative enzymes are important in the interaction between herbivorous insects and their host plants, especially from the perspectives of host specialization. They are usually composed of gene families, and their pattern of diversification may reveal some aspects of plant-insect coevolution to current investigators. We argue that UGTs are also good materials to study from that viewpoint as other detoxicative enzymes. The study on the interaction between *Helicoverpa assulta* and capsaicin provided a useful model system to investigate the coherent relationship between the host-specialist and a noxious compound from its host plant. The metabolite of capsaicin was revealed as a glucoside from the feces of larvae, and the enzyme responsible for this conjugation was examined and compared among other insect species. Further investigations of the UGT multigene family will lead to identification of one or more genes that are directly responsible not only for the glucosylation of capsaicin, but also manifold examples of other xenobiotic or endobiotic glucosylation in insects, which are likely to contribute to widen and diversify the research field of insect-plant interactions.

5-A. ZUSAMMENFASSUNG

Während der langen evolutionären Geschichte haben pflanzenfressende Insekten verschiedene Strategien entwickelt, um mit einer Vielzahl von schädlichen Verbindungen aus Wirtspflanzen umzugehen. Enzymatische Entgiftung ist eine der wichtigsten Mechanismen, um solche pflanzlichen Allelochemikalien zu neutralisieren, abzubauen oder zu modifizieren. Neben vielen anderen Phase-II-Stoffwechsel-Enzymen, ist UDP-Glycosyltransferase (UGT) an der Glucose-Konjugation von kleinen hydrophoben Molekülen beteiligt und unterstützt hierdurch die Ausscheidung oder die Verwertung im Körper. Daher wird angenommen, dass die Insekten-UGTs wichtige Rollen, vor allem bei der Entgiftung von Xenobiotika, spielen. Andere physiologische Aufgaben wie Sequestrierung, Pigmentierung, Geruchswahrnehmung und endobiotische Regulation wurden ebenfalls in mehreren Insekten nachgewiesen. Allerdings sind Insekten UGTs in Bezug auf die biochemischen Mechanismen, genomische Struktur usw. ein weitestgehend unerschlossenes Feld, vor allem im Vergleich zu anderen Entgiftungsenzymen wie Cytochrom P450 und Glutathion S-Transferasen (GSTs). Die vorliegende Studie ergab mehr als 320 Insekten UGT-Gene aus den Genom-Datenbanken von neun verschiedenen Insektenarten, einschließlich 46 UGTs von *H. armigera*. Eine vergleichende genomische Analyse enthüllte das Ausmaß der UGT Multigenfamilie in Insekten. Phylogenetische Analysen zeigten neben der Konservierung einer bestimmten Unterfamilie quer durch viele Organismen auch eine Abstammungslinien-spezifische Gen-Diversifizierung einiger UGT Unterfamilien in Insekten. UGTs aus zwei Lepidoptera Arten zeigten allgemein hochkonservierte Intron-Exon-Strukturen und genomische Organisation, aber auch eine ausgeprägte Spezies-spezifische Gen-Diversifizierung. Darüber hinaus haben alternatives Spleißen, wie alternative Promotoren, Exon-Skipping und Intron-Retention die Transkript-Vielfalt in *H. armigera* UGTs erhöht. Sequenzabgleiche haben wichtige strukturelle Domänen, wie die N-terminale substratbindende Domäne und die C-terminale UDP-Zucker-bindende Domäne identifiziert. Darüber hinaus wurden weitere wichtige strukturelle Motive, wie Signalpeptide, das zentrale UGT Signatur-Motiv, Transmembrandomänen sowie katalytische Aminosäure-Positionen im Detail identifiziert. Dies ist die erste Genomweite, vergleichende Studie einer Genfamilie von *H. armigera* sowie anderer Insekten.

Capsaicin ist ein Alkaloid, das nur in *Capsicum* spp. (Solanaceae) gefunden wurde, und für die Schärfe von Chili-Schoten verantwortlich ist. Der orientalische Tabakknospenwurm (*Helicoverpa assulta* (Guenée) (Lepidoptera: Noctuidae)), ist eines der wenigen Insekten, das sich erfolgreich von Paprikafrüchten ernähren kann. Die Beziehung zwischen *H. assulta* und Chili ist ein gutes Modellsystem, um die Wirtspflanzen-Spezialisierung aufgrund ihrer kohärenten Wechselwirkungen zu studieren. In der Tat war *H. assulta* toleranter gegenüber Capsaicin als andere getestete Nachtfalterarten. Unter bestimmten experimentellen Bedingungen war Capsaicin sogar für die

Entwicklung des Spezialisten *H. assulta* von Vorteil. Diese Studie legt außerdem nahe, dass Capsaicin gegen diejenigen Insekten, welche nicht an die Capsicum-Pflanze angepasst sind, als effektive Abwehrsubstanz wirkt. Darüber hinaus wirft die positive Wirkung von Capsaicin auf die Entwicklung der Larven von *H. assulta* Fragen bezüglich der physiologischen Rolle von Capsaicin bei der Anpassung von *H. assulta* an die Wirtspflanze, auf der sie der primäre landwirtschaftliche Schädling ist, auf. Dieser erste Teil der vorliegenden Studie führte demzufolge zur Untersuchung über die Entgiftung von Capsaicin bei Insekten.

Die Identifikation der Metaboliten ist der direkteste Weg, um den Entgiftungsmechanismus aufzuklären. Im Ergebnis der Analysen wurde Capsaicin-Glucosid im Kot von *Helicoverpa armigera*, *H. assulta* und *H. zea,* die mit künstlicher Diät unter Zugabe von Capsaicin ernährt wurden, identifiziert. Die unterschiedlichen Ausscheidungsraten von nativem Capsaicin lässt vermuten, dass *H. assulta* im direkten Vergleich mit den anderen beiden Generalisten eine höhere Kapazität besitzt, um mit Capsaicin umgehen zu können. Allerdings war die Ausscheidung von Capsaicin-Glucosid in *H. armigera* und *H. zea* höher als in *H. assulta*, was darauf schließen lässt, dass die aktiven Entgiftungsmechanismen über Glukose-Konjugation in *H. armigera* und *H. zea* aktiver als in *H. assulta* ist. Die Ergebnisse der UGT-Assays mit Rohenzym-Extrakten und Capsaicin wurden mit *in vivo* Glucosylierungsraten abgeglichen. Diese Ergebnisse zeigen, dass in *H. assulta* Glucosylierung kein wesentlicher Entgiftungsmechanismus für Capsaicin-Metabolite ist. Vergleicht man fünf verschiedene Larvalgewebe (labiale Drüsen, Hoden männlicher Larven, Mitteldarm, Malpighische Gefäße und Fettkörper) der drei Arten bezüglich der Bildung des Capsaicin-Glucosids, so ist diese im Fettkörper aller drei Arten am höchsten. Obwohl die geringere Ausscheidungsrate des nativen Capsaicins in *H. assulta* auf eine generell höhere metabolische Kapazität als bei den anderen beiden Generalisten hinweist, scheint die Glucosylierung *per se* nicht ausreichend zu sein, um die Abnahme von Capsaicin im Spezialisten zu erklären. Dies deutet darauf hin, dass *H. assulta* einen weiteren wichtigen, spezifischen Mechanismus für die Entgiftung von Capsaicin haben könnte.

Entgiftungsenzyme sind, vor allem aus der Perspektive der Wirtspflanzen-Spezialisierung, wichtige Komponenten, um die Interaktion zwischen pflanzenfressenden Insekten und ihren Wirtspflanzen zu verstehen. Diese Entgiftungsenzyme sind gewöhnlich Mitglieder von Genfamilien, die den Forschern zwar einen komplexen, aber interessanten "Fingerabdruck" ihrer evolutionären Geschichte aufzeigen können. Wir argumentieren, dass die UGT-Genfamilie ein weiteres wichtiges Werkzeug in der Erforschung der Entgiftung pflanzlicher Toxine darstellt. Die Studie der Wechselwirkung zwischen *Helicoverpa assulta* und Capsaicin liefert ein gutes Modellsystem, um die kohärente Beziehung zwischen dem Wirtspflanzen-Spezialisten und der schädlichen Verbindung aus dieser Wirtspflanze zu untersuchen. Das primäre Capsaicin-Metabolit wurde als Glucosid im Kot der Larven entdeckt, das für diese Konjugation verantwortliche Enzym wurde untersucht und mit anderen Insektenarten verglichen. Die Untersuchung der UGT-Genfamilie wird zur weiteren Identifizierung eines oder mehrerer Gene führen, die nicht nur an der Capsaicin-Glucosylierung beteiligt sind, sondern

wird auch vielfältige Beispiele für xenobiotische oder endobiotische Glucosylierung in Insekten geben, die wahrscheinlich den Forschungsbereich Insekten-Pflanzen-Interaktionen erweitern und bereichern wird.

6. REFERENCES

Ahmad, S., Brattsten, L.B., Mullin, C.A., Yu, S.J., 1986. Enzymes involved in the metabolism of plant allelochemicals, in: Brattsten, L.B., Ahmad, S. (Eds.), Molecular aspects of insect-plant associations. Plenum Press, New York & London, pp. 73-151.

Ahmad, S.A., Hopkins, T.L., 1992. Phenol β-glucosyltransferase and β-glucosidase activities in the tobacco hornworm larva *Manduca sexta* (L.): Properties and tissue localization. Archives of Insect Biochemistry and Physiology 21, 207-224.

Ahmad, S.A., Hopkins, T.L., 1993a. β-Glucosylation of plant phenolics by phenol β-glucosyltransferase in larval tissues of the tobacco hornworm, *Manduca sexta* (L.). Insect Biochemistry and Molecular Biology 23, 581-589.

Ahmad, S.A., Hopkins, T.L., 1993b. Phenol β-glucosyltransferases in six species of insects: properties and tissue localization. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 104, 515-519.

Ahmad, S.A., Hopkins, T.L., Kramer, K.J., 1996. Tyrosine β-glucosyltransferase in the tobacco hornworm, *Manduca sexta* (L): Properties, tissue localization, and developmental profile. Insect Biochemistry and Molecular Biology 26, 49-57.

Ahn, S.-J., Badenes-Pérez, F.R., Heckel, D.G., 2011. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. Journal of Insect Physiology doi: 10.1016/j.jinsphys.2011.05.015.

Al-Anzi, B., Tracey, W.D., Benzer, S., 2006. Response of *Drosophila* to wasabi is mediated by *painless*, the fly homolog of mammalian TRPA1/ANKTM1. Current Biology 16, 1034-1040.

Appendino, G., 2008. Capsaicin and Capsaicinoids, in: Fattorusso, E., Taglialatela-Scafati, O. (Eds.), Modern Alkaloids. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 73-109.

Aza-González, C., Núñez-Palenius, H., Ochoa-Alejo, N., 2010. Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). Plant Cell Reports, 1-12.

Baek, S., Cho, K., Song, Y.H., Lee, J.H., 2009. Sampling plans for estimating pepper fruit damage levels by Oriental tobacco budworm, *Helicoverpa assulta* (Guenée), in hot pepper fields. Journal of Asia-Pacific Entomology 12, 175-178.

Bateman, M.L., 2006. Impact of plant suitability, biogeography, and ecological factors on associations between the specialist herbivore *Heliothis subflexa* G. (Lepidoptera: Noctuidae) and the species in its host genus, *Physalis* L. (Solanaceae), in west-central Mexico. PhD Thesis, North Carolina State University, Raleigh, , North Carolina.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340, 783-795.

Björkman, C., 1997. A dome-shaped relationship between host plant allelochemical concentration and insect size. Biochemical Systematics and Ecology 25, 521-526.

Bock, K.W., 2003. Vertebrate UDP-glucuronosyltransferases: functional and evolutionary aspects. Biochemical Pharmacology 66, 691-696.

Bowles, D., Isayenkova, J., Lim, E.K., Poppenberger, B., 2005. Glycosyltransferases: managers of small molecules. Current Opinion in Plant Biology 8, 254-263.

Brattsten, L.B., 1988. Enzymic adaptations in leaf-feeding insects to host-plant allelochemicals. Journal of Chemical Ecology 14, 1919-1939.

Brattsten, L.B., 1992. Metabolic defenses against plant allelochemicals, in: Rosenthal, G.A., Berenbaum, M.R. (Eds.), Herbivores: Their interactions with secondary plant metabolites, Second edition, Vol. II: Ecological and evolutionary processes. Academic Press, Inc.; Academic Press Ltd., pp. 175-242.

Brown, M.R., Sieglaff, D.H., Rees, H.H., 2009. Gonadal ecdysteroidogenesis in arthropoda: Occurrence and regulation. Annual Review of Entomology 54, 105-125.

Calabrese, E.J., Blain, R., 2005. The occurrence of hormetic dose responses in the toxicological literature, the hormesis database: an overview. Toxicology and Applied Pharmacology 202, 289-301.

Casmuz, A., Laura Juarez, M., Guillermina Socias, M., Gabriela Murua, M., Prieto, S., Medina, S., Willink, E., Gastaminza, G., 2010. Review of the host plants of fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). Revista de la Sociedad Entomologica Argentina 69, 209-231.

Caterina, M.J., Schumacher, M.A., Tominaga, M., Rosen, T.A., Levine, J.D., Julius, D., 1997. The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature 389, 816-824.

Chanda, S., Bashir, M., Babbar, S., Koganti, A., Bley, K., 2008. *In vitro* hepatic and skin metabolism of capsaicin. Drug Metabolism and Disposition 36, 670-675.

Cho, S., Mitchell, A., Mitter, C., Regier, J., Matthews, M., Robertson, R.O.N., 2008. Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. Systematic Entomology 33, 581-594.

Choi, K.S., Boo, K.S., 1989. Food attractancy of the oriental tobacco budworm, *Heliothis assulta*, larvae, Korean Journal of Applied Entomology 28, 88-92.

Claudianos, C., Ranson, H., Johnson, R.M., Biswas, S., Schuler, M.A., Berenbaum, M.R., Feyereisen, R., Oakeshott, J.G., 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. Insect Molecular Biology 15, 615-636.

Consortium, H.G.S., 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931-949.

Cowles, R.S., Keller, J.E., Miller, J.R., 1989. Pungent spices, ground red pepper, and synthetic capsaicin as onion fly ovipositional deterrents. Journal of Chemical Ecology 15, 719-730.

Daimon, T., Hirayama, C., Kanai, M., Ruike, Y., Meng, Y., Kosegawa, E., Nakamura, M., Tsujimoto, G., Katsuma, S., Shimada, T., 2010. The silkworm Green b locus encodes a quercetin 5-O-glucosyltransferase that produces green cocoons with UV-shielding properties. Proceedings of the National Academy of Sciences of the United States of America 107, 11471-11476.

d'Alencon, E., Sezutsu, H., Legeai, F., Permal, E., Bernard-Samain, S., Gimenez, S., Gagneur, C., Cousserans, F., Shimomura, M., Brun-Barale, A., Flutre, T., Couloux, A., East, P., Gordon, K., Mita, K., Quesneville, H., Fournier, P., Feyereisen, R., 2010. Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. Proceedings of the National Academy of Sciences of the United States of America 107, 7680-7685.

Després, L., David, J.P., Gallet, C., 2007. The evolutionary ecology of insect resistance to plant chemicals. Trends in Ecology & Evolution 22, 298-307.

Díaz, J., Pomar, F., Bernal, A., Merino, F., 2004. Peroxidases and the metabolism of capsaicin in *Capsicum annuum* L. Phytochemistry Reviews 3, 141-157.

Dong, J.-F., Zhang, J.-H., Wang, C.-Z., 2002. Effects of plant allelochemicals on nutritional utilization and detoxication enzyme activities in two *Helicoverpa* species. Acta Entomologica Sinica 45, 296-300.

Edelson, J.V., Duthie, J., Roberts, W., 2002. Toxicity of biorational insecticides: activity against the green peach aphid, *Myzus persicae* (Sulzer). Pest Management Science 58, 255-260.

Ehrlich, P.R., Raven, P.H., 1964. Butterflies and plants: a study in coevolution. Evolution 18, 586-608.

Felsenstein, J., 2004. Inferring phylogenies. Sinauer Associates Inc.

Finney, D.J., 1971. Probit analysis, 3rd ed. ed. Cambridge University Press, Cambridge, UK.

Fitt, G.P., 1989. The ecology of *Heliothis* species in relation to agroecosystems. Annual Review of Entomology 34, 17-53.

Flagan, S.F., Leadbetter, J.R., 2006. Utilization of capsaicin and vanillylamine as growth substrates by *Capsicum* (hot pepper)-associated bacteria. Environmental Microbiology 8, 560-565.

Gong, Z., Son, W., Chung, Y.D., Kim, J., Shin, D.W., McClung, C.A., Lee, Y., Lee, H.W., Chang, D.-J., Kaang, B.-K., Cho, H., Oh, U., Hirsh, J., Kernan, M.J., Kim, C., 2004. Two interdependent TRPV channel subunits, Inactive and Nanchung, mediate hearing in *Drosophila*. The Journal of Neuroscience 24, 9059-9066.

Han, M.-W., Lee, J.-H., Son, J.-S., 1994. Intra- and inter-plant distribution of *Helicoverpa assulta* (Lepidoptera: Noctuidae) eggs in red pepper and tobacco fields, Korean Journal of Applied Entomology 33, 6-11.

Heckel, D.G., 2010. Molecular genetics of insecticide resistance in Lepidoptera, in: Goldsmith, M.R., Marec, F. (Eds.), Molecular biology and genetics of the Lepidoptera. Taylor & Francis, pp. 239-269.

Higashiguchi, F., Nakamura, H., Hayashi, H., Kometani, T., 2006. Purification and structure determination of glucosides of capsaicin and dihydrocapsaicin from various *Capsicum* fruits. Journal of Agricultural and Food Chemistry 54, 5948-5953.

Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee Apis mellifera. Nature 443, 931-949.

Hopkins, T.L., Ahmad, S.A., 1991. Flavonoid wing pigments in grasshoppers. Cellular and Molecular Life Sciences 47, 1089-1091.

Hopkins, T.L., Kramer, K.J., 1992. Insect cuticle sclerotization. Annual Review of Entomology 37, 273-302.

Hori, M., Nakamura, H., Fujii, Y., Suzuki, Y., Matsuda, K., 2011. Chemicals affecting the feeding preference of the Solanaceae-feeding lady beetle *Henosepilachna vigintioctomaculata* (Coleoptera: Coccinellidae). Journal of Applied Entomology 135, 121-131.

Huang, F.F., Chai, C.L., Zhang, Z., Liu, Z.H., Dai, F.Y., Lu, C., Xiang, Z.H., 2008. The UDP-glucosyltransferase multigene family in *Bombyx mori*. BMC Genomics 9, 563.

Jung, J.W., Cho, M.C., Cho, Y.S., 2006. Fruit quality of once-over harvest pepper (*Capsicum annuum*) cultivar 'Saengryeg No. 211' and 'Saengryeg No. 213'. Korean Journal of Horticultural Science & Technology 24, 205-209.

Kometani, T., Tanimoto, H., Nishimura, T., Kanbara, I., Okada, S., 1993. Glucosylation of capsaicin by cell suspension cultures of *Coffea arabica*. Bioscience Biotechnology and Biochemistry 57, 2192-2193.

Larue, E., Welty, C., 2010. Oviposition behavior and larval development of the European corn borer (Lepidoptera: Crambidae) on sweet versus hot peppers. Journal of Entomological Science 45, 353-365.

Lee, H.S., Hieu, T.T., Ahn, Y.J., 2006. Oviposition-stimulating activity of (*E*)-capsaicin identified in *Capsicum annuum* fruit and related compounds towards *Helicoverpa assulta* (Lepidoptera: Noctuidae). Chemoecology 16, 153-157.

Levey, D.J., Tewksbury, J.J., Cipollini, M.L., Carlo, T.A., 2006. A field test of the directed deterrence hypothesis in two species of wild chili. Oecologia 150, 61-68.

Lin LZ, West DP, Cordell G. 1993. NMR assignments of *cis*- and *trans*-capsaicin. Natural Product Letters 3, 5-8.

Liu, Z.D., Li, D.M., Gong, P.Y., Wu, K.J., 2004. Life table studies of the cotton bollworm, *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae), on different host plants. Environmental Entomology 33, 1570-1576.

Luque, T., Okano, K., O'Reilly, D.R., 2002. Characterization of a novel silkworm (*Bombyx mori*) phenol UDP-glucosyltransferase. European Journal of Biochemistry 269, 819-825.

Luque, T., O'Reilly, D.R., 2002. Functional and phylogenetic analyses of a putative *Drosophila melanogaster* UDP-glycosyltransferase gene. Insect Biochemistry and Molecular Biology 32, 1597-1604.

Mackenzie, P.I., Owens, I.S., Burchell, B., Bock, K.W., Bairoch, A., Belanger, A., FournelGigleux, S., Green, M., Hum, D.W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J.R., Ritter, J.K., Schachter, H., Tephly, T.R., Tipton, K.F., Nebert, D.W., 1997. The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7, 255-269.

Madhumathy, A.P., Aivazi, A.-A., Vijayan, V.A., 2007. Larvicidal efficacy of *Capsicum annum* against *Anopheles stephensi* and *Culex quinquefasciatus*. Journal of Vector Borne Diseases 44, 223-226.

Magdalou, J., Fournel-Gigleux, S., Ouzzine, M., 2010. Insights on membrane topology and structure/function of UDP-glucuronosyltransferases. Drug Metabolism Reviews 42, 159-166.

Matsuura, H., Sokabe, T., Kohno, K., Tominaga, M., Kadowaki, T., 2009. Evolutionary conservation and changes in insect TRP channels. BMC Evolutionary Biology 9, 228.

Matthews, M., 1999. Heliothine moths of Australia. CSIRO Publishing, Collingwood.

Mazourek, M., Pujar, A., Borovsky, Y., Paran, I., Mueller, L., Jahn, M.M., 2009. A dynamic interface for capsaicinoid systems biology. Plant Physiology 150, 1806-1821.

Miley, M.J., Zielinska, A.K., Keenan, J.E., Bratton, S.M., Radominska-Pandya, A., Redinbo, M.R., 2007. Crystal Structure of the Cofactor-Binding Domain of the Human Phase II Drug-Metabolism Enzyme UDP-Glucuronosyltransferase 2B7. Journal of Molecular Biology 369, 498-511.

Mitter, C., Poole, R.W., Matthews, M., 1993. Biosystematics of the Heliothinae (Lepidoptera: Noctuidae). Annual Review of Entomology 38, 207-225.

Mizokami, H., Yoshitama, K., 2009. Sequestration and metabolism of host-plant flavonoids by the Pale Grass Blue, Pseudozizeeria maha (Lepidoptera: Lycaenidae). Entomological Science 12, 171-176.

Morello, A., Repetto, Y., 1979. UDP-glucosyltransferase activity of housefly microsomal fraction. Biochemical Journal 177, 809-812.

Nijhout, H., 2003. The control of body size in insects. Developmental Biology 261, 1-9.

Noami, M., Igarashi, K., Kasuya, F., Ohta, M., Kanamori-Kataoka, M., Seto, Y., 2006. Study on the analysis of capsaicin glucuronide in rat urine by liquid chromatography-mass spectrometry after enzymatic hydrolysis. Journal of Health Science 52, 660-665.

Oakeshott, J.G., Johnson, R.M., Berenbaum, M.R., Ranson, H., Cristino, A.S., Claudianos, C., 2010. Metabolic enzymes associated with xenobiotic and chemosensory responses in *Nasonia vitripennis*. Insect Molecular Biology 19, 147-163.

Olszewska, J., 2010. Vanilloid receptors - comparison of structure and functions in mammals and invertebrates. Folia Biologica (Kraków) 58, 1-7.

Osmani, S.A., Bak, S., Møller, B.L., 2009. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70, 325-347.

Radominska-Pandya, A., Bratton, S.M., Redinbo, M.R., Miley, M.J., 2010. The crystal structure of human UDP-glucuronosyltransferase 2B7 C-terminal end is the first mammalian UGT target to be revealed: the significance for human UGTs from both the 1A and 2B families. Drug Metabolism Reviews 42, 133-144.

Ramsey, J.S., Rider, D.S., Walsh, T.K., De Vos, M., Gordon, K.H.J., Ponnala, L., Macmil, S.L., Roe, B.A., Jander, G., 2010. Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. Insect Molecular Biology 19, 155-164.

Real, M.D., Ferré, J., Chapa, F.J., 1991. UDP-glucosyltransferase activity toward exogenous substrates in *Drosophila melanogaster*. Analytical Biochemistry 194, 349-352.

Reilly, C.A., Yost, G.S., 2006. Metabolism of capsaicinoids by P450 enzymes: a review of recent findings on reaction mechanisms, bio-activation, and detoxification processes. Drug Metabolism Reviews 38, 685-706.

Robertson, H.M., Martos, R., Sears, C.R., Todres, E.Z., Walden, K.K.O., Nardi, J.B., 1999. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. Insect Molecular Biology 8, 501-518.

Romano, D., Gandolfi, R., Guglielmetti, S., Molinari, F., 2011. Enzymatic hydrolysis of capsaicins for the production of vanillylamine using ECB deacylase from *Actinoplanes utahensis*. Food Chemistry 124, 1096-1098.

Ross, J., Li, Y., Lim, E.-K., Bowles, D.J., 2001. Higher plant glycosyltransferases. Genome Biology 2, reviews3004.3001–3004.3006.

Ruvolo, P.P., 2003. Intracellular signal transduction pathways activated by ceramide and its metabolites. Pharmacological Research 47, 383-392.

Sasai, H., Ishida, M., Murakami, K., Tadokoro, N., Ishihara, A., Nishida, R., Mori, N., 2009. Species-specific glucosylation of DIMBOA in larvae of the rice armyworm. Bioscience Biotechnology and Biochemistry 73, 1333-1338.

Schoonhoven, L.M., Loon, J.J.A.v., Dicke, M., 2005. Insect-Plant Biology, 2nd ed. Oxford University Press.

Seppen, J., Steenken, E., Lindhout, D., Bosma, P.J., Oude Elferink, R.P.J., 1996. A mutation which disrupts the hydrophobic core of the signal peptide of bilirubin UDP-glucuronosyltransferase, an endoplasmic reticulum membrane protein, causes Crigler-Najjar type IIs. FEBS Letters 390, 294-298.

Shimoda, K., Kwon, S., Utsuki, A., Ohiwa, S., Katsuragi, H., Yonemoto, N., Hamada, H., Hamada, H., 2007. Glycosylation of capsaicin and 8-nordihydrocapsaicin by cultured cells of *Catharanthus roseus*. Phytochemistry 68, 1391-1396.

Slansky, F., Jr., 1992. Allelochemical-nutrient interactions in herbivore nutritional ecology, in: Rosenthal, G.A., Berenbaum, M.R. (Eds.), Herbivores: Their interactions with secondary plant metabolites, Second edition, Vol. II: Ecological and evolutionary processes. Academic Press, Inc.; Academic Press Ltd., pp. 135-174.

Smith, J.N., 1962. Detoxication mechanisms. Annual Review of Entomology 7, 465-480.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, D.C., 1985. Measurement of protein using bicinchoninic acid. Analytical Biochemistry 150, 76-85.

Sprong, H., Kruithof, B., Leijendekker, R., Slot, J.W., van Meer, G., van der Sluijs, P., 1998. UDP-galactose:ceramide galactosyltransferase is a class I integral membrane protein of the endoplasmic reticulum. Journal of Biological Chemistry 273, 25880-25888.

Stipanovic, R.D., Williams, H.J., Smith, L.A., 1986. Cotton terpenoid inhibition of *Heliothis virescens* development, Natural Resistance of Plants to Pests. American Chemical Society, pp. 79-94.

Strode, C., Wondji, C.S., David, J.P., Hawkes, N.J., Lumjuan, N., Nelson, D.R., Drane, D.R., Karunaratne, S., Hemingway, J., Black, W.C., Ranson, H., 2008. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. Insect Biochemistry and Molecular Biology 38, 113-123.

Surh, Y.-J., Kundu, J.K., 2011. Molecular mechanisms of chemoprevention with capsaicinoids from chili peppers, in: Mutanen, M., Pajari, A.-M. (Eds.), Vegetables, Whole Grains, and Their Derivatives in Cancer Prevention. Springer Netherlands, pp. 123-142.

Surh, Y.-J., Lee, S.S., 1995. Capsaicin, a double-edged sword: Toxicity, metabolism, and chemopreventive potential. Life Sciences 56, 1845-1855.

Svoboda, J., Weirich, G., 1995. Sterol metabolism in the tobacco hornworm, *Manduca sexta* - A review. Lipids 30, 263-267.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution (In Press).

Teese, M.G., Campbell, P.M., Scott, C., Gordon, K.H.J., Southon, A., Hovan, D., Robin, C., Russell, R.J., Oakeshott, J.G., 2010. Gene identification and proteomic analysis of the esterases of the cotton bollworm, *Helicoverpa armigera*. Insect Biochemistry and Molecular Biology 40, 1-16.

Tewksbury, J.J., Manchego, C., Haak, D.C., Levey, D.J., 2006. Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. Journal of Chemical Ecology 32, 547-564.

Tewksbury, J.J., Nabhan, G.P., 2001. Seed dispersal: directed deterrence by capsaicin in chillies. Nature 412, 403-404.

Tewksbury, J.J., Reagan, K.M., Machnicki, N.J., Carlo, T.A., Haak, D.C., Peñaloza, A.L.C., Levey, D.J., 2008. Evolutionary ecology of pungency in wild chilies. Proceedings of the National Academy of Sciences of the United States of America 105, 11808-11811.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25, 4876-4882.

Tijet, N., Helvig, C., Feyereisen, R., 2001. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: Annotation, intron-exon organization and phylogeny. Gene 262, 189-198.

Tsubota, T., Shiotsuki, T., 2010. Genomic analysis of carboxyl/cholinesterase genes in the silkworm *Bombyx mori*. BMC Genomics 11, 377.

Waldbauer, G.P., 1968. The consumption and utilization of food by insects. Advances in Insect Physiology. 5, 229-288.

Wang, Q., Hasan, G., Pikielny, C.W., 1999. Preferential expression of biotransformation enzymes in the olfactory organs of *Drosophila melanogaster*, the antennae. Journal of Biological Chemistry 274, 10309-10315.

Wang, X., 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Letters 583, 3303-3309.

Weissenberg, M., Klein, M., Meisner, J., Ascher, K.R.S., 1986. Larval growth inhibition of the spiny bollworm, *Earias insulana*, by some steroidal secondary plant compounds. Entomologia Experimentalis et Applicata 42, 213-217.

Wiesen, B., Krug, E., Fiedler, K., Wray, V., Proksch, P., 1994. Sequestration of host-plant-derived flavonoids by lycaenid butterfly *Polyommatus icarus*. Journal of Chemical Ecology 20, 2523-2538.

Wilkinson, C.F., 1986. Xenobiotic Conjugation in Insects, Xenobiotic Conjugation Chemistry. American Chemical Society, pp. 48-61.

Yang, C.-Y., Jeon, H.-Y., Cho, M.-R., Kim, D.-S., Yiem, M.-S., 2004. Seasonal occurrence of oriental tobacco budworm (Lepidoptera: Noctuidae) male and chemical control at red pepper fields. Korean Journal of Applied Entomology 43, 49-54.

Yu, Q.Y., Lu, C., Li, B., Fang, S.M., Zuo, W.D., Dai, F.Y., Zhang, Z., Xiang, Z.H., 2008. Identification, genomic organization

Zagrobelny, M., Scheibye-Alsing, K., Jensen, N.B., Moller, B.L., Gorodkin, J., Bak, S., 2009. 454 pyrosequencing based transcriptome analysis of *Zygaena filipendulae* with focus on genes involved in biosynthesis of cyanogenic glucosides. BMC Genomics 10, 574.

7. ACKNOWLEDGEMENTS

Looking back my 'PhD tunnel' that I have passed through for the past five years, I realize that I have never been alone. When I lost the way in the dark, there was someone who told me where to go; when I was afraid, there was someone who gave me a hand to walk together. I was lucky to have a chance to be here in the prestigious institute with lots of brilliant and nice people whom I never forget. This work would not be possible without substantial help of such people. In first hand, I would like to thank my PhD supervisor **Prof. David G. Heckel** who fueled me with great insights into science in general, power of thinking for challenging problems, and unbiased but warm attitude of mind toward people. These will be inexhaustible energy for me to continue my research career with in the future.

My special thanks go to the colleagues in the Department of Entomology; I was happy to work with **Dr. Francisco R. Badenes-Pérez** who taught me what insect ecology is, **Dr. Astrid Groot** who showed me how important interactive and communicative skills are in science, **Dr. Susanne Preiss** who has kindly helped me integrate into the Department since beginning, **Dr. Yannick Pauchet** who encouraged me always to go further and **Dr. Heiko Vogel** who helped me with brilliant ideas and immeasurable knowledge during the whole period of my PhD. **Dr. Hugh Loxdale** showed me how I have to live as an entomologist. I deliver my warm-hearted 'Danke' to our secretaries, **Katrin** and **Ramona**, who didn't hesitate to help me whenever I had to solve entangled issues on administration, living and family. I am also grateful to **Regina** and **Jutta** for managing our ecology lab, to **Susi**, **Antje**, **Bianca**, and **Henriette** for helping me of molecular and biochemical experiments, and to **Domenica** and **Steffi** for sequencing and bioinformatic advices.

It was enjoyable to work with **Maripaz** who has a magic tool to turn everything into positive mode in the department. I won't forget talking with **Lisa** in the 'Café Corner' every morning; you told me lots of astonishing stories of diverse insects. I got a great help from **Kirsten** who has left Jena already, but always with me anytime from a distance. I was able to get an apartment in the very beginning with help of **Nadine**. I thank all the other Entor/rin, for their help in daily institute life; **Tobias**, **Brent**, **Sebastian**, **Sher**, **Suyog**, **Roy**, **Charles**, **Jeffery**, **Zhudong**, **Gerhard**, **David N.**, **Ariadne**, **Nicole**, **Hanna**, **Silvia**, **Kathrin**, **Juliette**, **Dalial**, **Melanie**, **Sabine**, **Andrea**, and **Anne**. In addition, I thank my five **HiWis** for assisting my experiments and insect rearing.

I am indebted to **Dr. Aleš Svatoš** and **Dr. Bernd Schneider** for their help to identify capsaicin glucoside. **Dr. Michael Reichelt** was indeed my LC-MS teacher who gave me a hand when I was trying to isolate the metabolite. Without their helps, I wouldn't have a chance to discover the beautiful diversity of UGT gene family.

I would like to specially acknowledge all my friends in the institute who come across on the way to the institute, in the institute corridor, or on my way home every daily life; you were all my teachers, friends, and advisors.

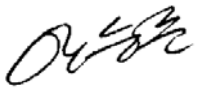
I owe a special gratitude to the IMPRS coordinators; **Elke** for helping me adjust to the initial life in Jena and **Karin** for the administrative helps during my PhD. I am also thankful to the **Max Planck Society** and the **International Max Planck Research School (IMPRS)** for the financial support.

It was a big fun and a kind of 'vitamin' in foreign life to meet Korean friends in the institute; thanks **Young-Joo** and **Sang-Gyu**!

Last but not least, I thank my wife, **Min-Joo**, and my two children, **Chae-Yeong** and **Theo**, for their love, patience and support. Forgive me that I always showed up late at home. Above all, I would like to thank **my parents** who always pray for me and love me in any cases.

8. SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, Seung-Joon Ahn, dass ich die vorliegende Arbeit selbständig und unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.



Seung-Joon Ahn

Jena, am 6. Juli 2011

9. CURRICULUM VITAE

Personal data

Name: Seung-Joon Ahn
Date of birth: Feburary 2, 1976
Place of birth: Ulsan, Korea
Nationality: Korea
Address: Charlottenstr. 23, B201, 07749 Jena, Germany
Marital status: Married, Min-Joo Kim
Daughter, Chae-Yeong Ahn, born in Busan, Korea, at September 28, 2005
Son, Theo Ahn, born in Jena, Germany, at July 3, 2009
E-mail: sahn@ice.mpg.de, massai94@hanmail.net

Education

Ph.D in Ecology, *anticipated 2011*, Friedrich-Schiller-Universität, Biologisch-Pharmazeutische Fakultät / International Max-Planck Research School, Jena, Germany
Dissertation title: Detoxification mechanisms in a host-plant specialist, *Helicoverpa assulta*
Dissertation supervisor: Prof. Dr. David G. Heckel (MPI for Chemical Ecology)
(Second advisor: Prof. Dr. Wolfgang Weisser, FSU)

M.S. in Entomology, 2000, Seoul National University, Korea
Thesis title: Effect of mating and a host plant, *Capsicum annuum*, on sex pheromone production in the Oriental tobacco budworm, *Helicoverpa assulta*
Dissertation supervisor: Prof. Kyung-Saeng Boo

B.S. in Agricultural Biology, 1998, Seoul National University, Korea

Publications

Celorio-Manceraa, M., **Ahn, S.-J.**, Vogel, H. Heckel, D.G. (submitted) Gene expression underlying the hormetic and detrimental effect of a plant secondary metabolite on the generalist lepidopteran Helicoverpa armigera. ***BMC Biology***

Ahn, S.-J., Vogel H., Heckel D.G., (submitted) Comparative analysis of the UDP-glycosyltransferase multigene family in insects ***Insect Biochemistry and Molecular Biology*** (June 30, 2011)

Ahn, S.-J., Badenes-Pérez F. R., Reichelt M., Svatos A., Schneider B., Gershenzon J., Heckel, D. G. (submitted) Metabolic detoxification of capsaicin by UDP-glycosyltransferase in three *Helicoverpa* species. ***Archives of Insect Biochemistry and Physiology*** (June 30, 2011)

Ahn, S.-J., Badenes-Pérez F. R., and Heckel, D. G., 2011. A host-plant specialist, *Helicoverpa assulta*, is more tolerant to capsaicin from *Capsicum annuum* than other noctuid species. ***Journal of Insect Physiology*** doi: 10.1016/j.jinsphys.2011.05.015 (online access)

Ahn, S. J., Yoo, J. H., Lee, H. C., Kim, S. Y., Noh, B. S., Kim, J. H., and Lee, J. K. 2003. Enhanced conversion of sucrose to isomaltulose by a mutant of *Erwinia rhapontici*. Biotechnology Letters 25: 1179-1183

Ahn, S. J., Choi, M. Y., and Boo, K. S. 2002. Mating effect on sex pheromone production of the Oriental tobacco budworm, *Helicoverpa assulta*. Journal of Asia-Pacific Entomology 5: 43-48.

Talk Presentations

Ahn, S.J. Metabolic detoxification of capsaicin: UDP-glucosyltransferase and its gene family in *Helicoverpa* spp. ICE Symposium, MPI for Chemical Ecology, Jena, DE, Sep 2010

Ahn, S.J. More than a spice: The effects of capsaicin from hot pepper on insect development. IMPRS Evaluation Symposium, MPI for Chemical Ecology, Jena, DE, Sep 2007

Ahn, S.J. More than a spice: The effects of capsaicin from hot pepper on insect development. 6th Biannual IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE, Mar 2007

Poster Presentations

Ahn, S.J. Vogel H., Heckel D.G. Molecular evolution of the UDP-glycosyltransferase multigene family in Lepidoptera. SAB Meeting 2010, MPI for Chemical Ecology, Jena, DE, Oct 2010

Ahn, S.J., Vogel H., Heckel D.G. The UDP-glycosyltransferase multigene family in *Helicoverpa armigera*. 4th Annual Arthropod Genomics Symposium, Kansas City, US, Jun 2010

Ahn, S.J., Badenes-Pérez F.R., Heckel D.G. Chemical Ecology of Pungency: Detoxification of Capsaicin in a Host Specialist, *Helicoverpa assulta*. 25th ISCE Meeting, International Society of Chemical Ecology, Neuchâtel, CH, Aug 2009

Ahn S.J., Badenes-Pérez F.R., Heckel D.G. Detoxification of capsaicin in the host-plant specialist *Helicoverpa assulta*. DGaaE-Tagung, Deutsche Gesellschaft für allgemeine und angewandte Entomologie, Göttingen, DE, Mar 2009

Ahn S.J. Chemical ecology of pungency: how a specialist herbivore copes with a noxious compound. 8th IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE, Mar 2009

Ahn S.J., Badenes-Pérez F.R., Heckel D.G. Chemical ecology of pungency: how a specialist herbivore copes with a noxious compound. ICE Symposium, MPI for Chemical Ecology, Jena, DE, Sep 2008

Ahn S.J., Badenes-Pérez F.R., Heckel D.G. Detoxification of Capsaicin in the Host-Plant Specialist *Helicoverpa assulta*. 24th ISCE Meeting, International Society of Chemical Ecology, State College, US, Aug 2008

Ahn S.J. Capsaicin detoxification in the host-plant specialist *Helicoverpa assulta*. 7th IMPRS Symposium, MPI for Chemical Ecology, Dornburg, DE, Feb 2008

Ahn S.J., Badenes-Pérez F.R., Heckel D.G. Capsaicin detoxification in the host-plant specialist *Helicoverpa assulta*. SAB Meeting 2008, MPI for Chemical Ecology, Jena, DE, Jan 2008

Ahn S.J., Heckel D.G. More than a spice: The effects of capsaicin from hot pepper on insect development. 23rd ISCE Meeting, International Society of Chemical Ecology, Jena, DE, Jul 2007

Ahn S.J., Liu Z., Heckel D.G. Detoxification mechanisms in a host-plant specialist, *Helicoverpa assulta*. 5th Biannual IMPRS Symposium, MPI for Chemical Ecology, Jena, DE, Nov 2006

Ahn S.J., Liu Z., Heckel D.G. Detoxification mechanisms in a host-plant specialist, *Helicoverpa assulta*. SAB Meeting 2006, MPI for Chemical Ecology, Jena, DE, Oct 2006

Awards

IMPRS best poster award, The 8th IMPRS Symposium 2009, Jena, Germany

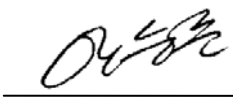
Student poster award, The 25th International Society of Chemical Ecology Annual Meeting 2009, Neuchatel, Switzerland

Manuscript reviewing

Bulletin of Entomological Research (2010)

Training

The 3rd International course - Conservation of Biodiversity in Desert Ecosystem, 2000, Ben-Gurion University of the Negev, Sede Boqer, Israel



Seung-Joon Ahn
Jena, 6 July 2011