UNIVERSITA' DEGLI STUDI DI PADOVA

Department of Agronomy, Food, Natural Resaurces, Animals and the Environment

(DAFNAE)

DOCTORATE SCHOOL OF CROP SCIENCE

CURRICULUM CROP PROTECTION

CYCLE XXVII

Genetic and proteomic approach to the urticating system of processionary moths (Thaumetopoeinae, Lepidoptera)

Director of the school: Prof. Antonio Berti

Supervisor: Prof. Andrea Battisti

Co-supervisor: Prof. Enrico Negrisolo

PhD student: Laura Berardi

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Padova, January 2015 Laura Berardi

A copy of the thesis will be available at *http://paduaresearch.cab.unipd.it/*

Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

Padova, Gennaio 2015 Laura Berardi

Una copia della tesi sarà disponibile presso *http://paduaresearch.cab.unipd.it/*

Tables of content

Riassunto

Le larve delle processionarie europee e mediterranee (genere *Thaumetopoea*) producono delle setole urticanti dal terzo stadio larvale al quinto sulla parte dorsale dell'addome, in specifiche aree denominate "specchi". Le setole sono facilmente removibili dal tegumento e sono importanti per la difesa dell'insetto nei confronti di vertebrati predatori. L'impatto delle setole urticanti sull'uomo è stato ripetutamente descritto e consiste prevalentemente nella manifestazione da contatto con la complicazione dell'insorgenza di risposte allergiche. Il meccanismo di azione è complesso e riguarda sia una lesione meccanica causata dalla penetrazione della setola sia il rilascio di proteine con potere antigenico. Risulta quindi importante esplorare i geni potenzialmente coinvolti nel meccanismo di difesa, al fine di identificare le proteine associate con le setole e chiarire la loro espressione durante lo sviluppo larvale. Inoltre, il genere comprende diverse specie e sembra opportuno esplorare la variabilità dei geni coinvolti a livello intra- e interspecifico.

Nell'introduzione della tesi presento i diversi tipi di setole urticanti degli artropodi, focalizzandomi su quelle che vengono denominate "setole urticanti" caratteristiche delle processionarie, gli aspetti medici correlati, l'epidemiologia e i geni correlati alle proteine antigeniche. Uno dei principali obiettivi del mio lavoro è quello di caratterizzare tutte le proteine, urticanti e non, presenti all'interno delle setole e provare a identificare altre proteine antigeniche, oltre alla già nota Tha p 2. Altri obiettivi sono quelli di concentrarsi sul profilo di espressione della proteina Tha p 1 (una proteina isolata da larve di processionaria del pino e riconosciuta come antigene da persone esposte all'insetto, ma successivamente identificata come appartenente al gruppo delle proteine chemosensoriali) e Tha p 2 in tutti gli stadi vitali di *Thaumetopoea pityocampa*, e la caratterizzazione del gene *Tha p 2* in tutte le specie del genere *Thaumetopoea* disponibili e anche in altre della sottofamiglia delle Thaumetopoeainae, per studiare l'evoluzione della proteina in questo gruppo.

Nel primo studio ho provato due protocolli di estrazione proteica per creare un data set completo di tutte le proteine, sia urticanti che non, presenti nelle setole. Possibili proteine antigeniche sono state riconosciute utilizzando i sieri di persone precedentemente esposte e con reazioni acute. É stata ottenuta un'elevata quantità di proteine che ha permesso di confermare che le setole urticanti di *Th. pityocampa* contengono proteine, alcune delle quali sono riconosciute da Ig-E di persone precedentemente esposte a larve di questo insetto. Inoltre ho potuto ottenere informazioni riguardo la qualità e la quantità delle proteine associate alle setole.

Nel secondo studio mi sono concentrata sull'espressione dei geni associati alle proteine urticanti Tha p 2 e Tha p 1 in tutti gli stadi larvali di due popolazioni italiane e per gli ultimi stadi e le uova di due popolazioni portoghesi. Ho potuto confermare l'espressione del gene *Tha p 1* in tutti gli stadi di *Th. pityocampa* di tutte le popolazioni, mentre il gene *Tha p 2* è espresso solo negli stadi larvali dove le setole sono prodotte (dal terzo al quinto).

Nel terzo studio, ho sequenziato il gene *Tha p 2* in tutte le specie della sottofamiglia Thaumetopoeinae disponibili, concentrandomi su membri del genere *Thaumetopoea*, così come su specie non affini, per capire le proprietà chimico-fisiche della proteina, la natura dei geni e la loro storia evolutiva. Questo studio ha permesso di identificare due diverse isoforme del gene *Tha p 2* in tutte le specie, che può essere interpretato come un risultato di eterozigosi del singolo gene. L'unica eccezione è rappresentata da una specie (*Thaumetopoea wilkinsoni*) nella quale sono state individuate 20 diverse isoforme in un unico campione; ciò porta a pensare che possano esistere molteplici copie del gene.

I capitoli supplementari includono due parti alle quali ho lavorato durante il periodo della tesi, per ottenere materiale utilizzabile durante gli altri esperimenti. La prima parte riguarda il confronto del tempo di sviluppo di quattro popolazioni di processionaria del pino mantenute in condizioni controllate di laboratorio. Le quattro popolazioni, caratterizzate da diversa fenologia, mantengono un ciclo vitale annuale anche in condizioni favorevoli a un rapido sviluppo larvale, soprattutto grazie alla flessibilità della durata dello stadio pupale. La seconda parte dell'appendice è una descrizione della biologia di *Thaumetopoea herculeana* campionata in Spagna e studiata per l'analisi del gene *Tha p 2*.

Nel complesso, la tesi approfondisce la conoscenza del sistema urticante delle processionarie, puntando a individuare gli ulteriori passi necessari per chiarire il complesso meccanismo associato allo sviluppo di reazioni cutanee negli esseri umani e, possibilmente, in vertebrati predatori che sono il bersaglio naturale della setole. Inoltre, le analisi rivelano che il sistema urticante e i geni associati sono ben conservati nel gruppo e potrebbero essere un fattore importante nella storia evolutiva in questo e in altri gruppi di artropodi che condividono meccanismi di difesa simili.

Summary

Larvae of the *Thaumetopoea* genus produce urticating setae in the third and later instars on the dorsal part of the abdomen in specific areas called mirrors. These setae are easily removed from the integument and are considered to be important for defense against vertebrate predators. Impacts on humans are well described and consist mainly of dermatitis due to contact with the setae and allergic responses. As the mechanism of action in the human skin involves both the mechanical injury of the penetration and the release of proteins with antigenic power, it is important to explore the genes putatively involved in the defense mechanism of the *Thaumetopoea* species, in order to identify the proteins associated with the setae and elucidate their expression pathway during the larval development. In addition, as the genus includes several species, the variability occurring within and among species should be also explored.

In the introduction of the thesis I present the different type of urticating hairs of the arthropods with a focus on the true setae of the processionary moths, the medical aspects related to them, the epidemiology, and the genes encoding for the antigenic proteins. One of the main aim of my work was to characterize all the proteins, urticating and not, occurring in the setae and try to extract other proteins similar to the main one described so far, Tha p 2. Other goals were to focalize on the expression profile of Tha p 1 (a chemosensory protein of *Th. pityocampa*) and Tha p 2 in every life stage of *Thaumetopoea pityocampa* and, at the end, to characterize the orthologous counterparts of Tha p 2 in all the *Thaumetopoea* species available, as well as in other species of Thaumetopoeinae, to better understand the evolution of the protein in this group.

In the first study I tested two different protein extraction protocols of different strength, in order to create a complete data set of all proteins, urticating and not-urticating, occurring in the setae. Candidate antigenic proteins were recognised by using the sera of persons previously exposed to the agent and showing acute reactions. A high quantity of protein was obtained, confirming that the urticating setae of *Th. pityocampa* contain proteins, some of which were recognized by Ig-E of persons previously exposed to the larvae of this insect. In addition, I added information about the type, quality and quantity of the proteins associated with the setae.

In the second study I investigated the expression values of the urticating protein Tha p 2 and also of a chemosensory protein Tha p 1, for all life stages of two Italian populations and for the last development stages and eggs of two Portuguese populations, in order to better understand the expression of the gene *Tha p 2*. I confirmed the expression of *Tha p 1* gene in

all development stages of *Th. pityocampa* of all populations, instead the gene *Tha p 2* was typically expressed only in those larval instars when the setae are produced, i.e. from the third to the fifth.

In the third study, I sequenced the *Tha p 2* gene in all Thaumetopoeinae species available, with a particular focus on members of the genus *Thaumetopoea*, as well as unrelated moth species, for better understanding the chemical-physical properties of the proteins and the nature of the encoding genes, as well as their evolutionary history. This study permitted to identify two isoforms of *Tha p 2* in all species that can be interpreted as the result of heterozygosity in the single gene. The only exception is represented by one species (*Thaumetopoea wilkinsoni)*, in which 20 different isoforms occur in a single specimen, leading to the conclusion that at least in the species multiple copies of *Tha p 2* exist.

The Supplementary chapters includes two parts which I developed during the thesis in order to obtain the material used in the analyses. The first concerns the comparison of the developmental time of four populations of the pine processionary moth under controlled laboratory conditions. The four populations, which are characterized by different life history phenology in the field, maintained an annual life cycle also under favorable laboratory conditions, mainly because of a flexible duration of the pupal stage. The second part of the Supplementary chapters is a description of the life history of a species (*Thaumetopoea herculeana*) collected and studied in Spain for the analysis of the *Tha p 2* gene..

On the whole, the thesis is deepening the knowledge on the urticating system of the processionary moths, pointing at identifying the further steps required for the clarification of the complex mechanism associated with the development of reactions in humans and possibly in the vertebrate predators which are the natural target of the setae. In addition, the analyses reveal that the urticating seta system, and the genes associated, are well conserved in the group and could be a major factor affecting their evolution, as well as that of other groups of arthropods which share similar defense mechanisms.

Chapter 1

Introduction

Urticating Hairs in Arthropods

The arthropods own different types of hairs with several functions (Witz 1990). The hairs play a rule of protection against invertebrate and vertebrate in several taxa, like lepidopterans, spiders, coleopterans, and millipedes (Battisti et al. 2011). A role of protection is insured by the urticating hairs characterize by at least two cells, hair-forming cells, called trichogen cell also named hair-forming cell and tormogen also called auxiliary cell. Those cells are connected to the neurons for transmission of sensorial information (Battisti et al. 2011) (Fig 1). In literature the urticating hairs are divided in three categories: true setae, modified setae and spines depicted in Fig.1.

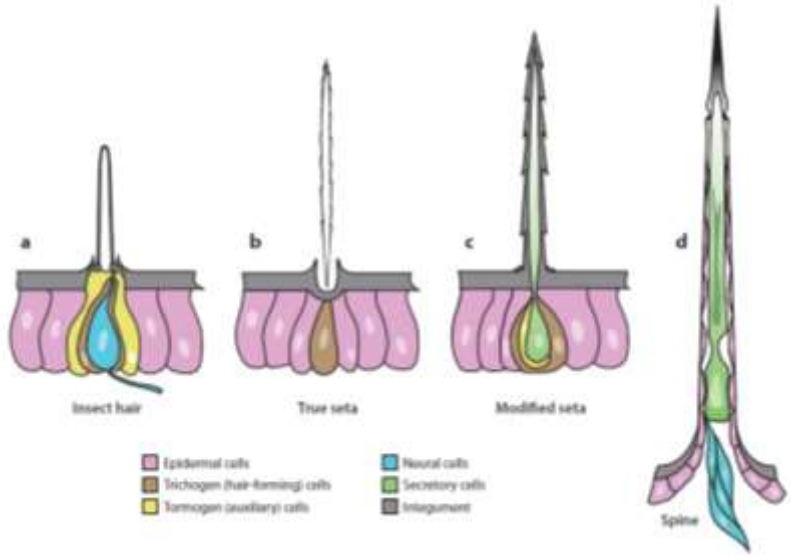


Figure 1. Schematic representation of (a) an insect hair, (b) a true seta, (c) a modified seta, and (d) a spine. Note the different scale between (a), (b), (c), and (d) by comparing the epidermal cell size (Battisti et al. 2011).

True setae

True setae are the particular hairs that occur in Lepidoptera larvae and adults (Gilmer 1925) (Kawamoto and Kumada 1984) and in some spiders (Cooke et al 1972). In particular they are a feature of the larval stage of processionary moths (Thaumetopoeinae, Notodontidae) and adult stage of Notodontidae (Rothschild et al. 1970), Lymantridae, a few species of Saturniidae in South America (Werno 1991) and Zygaenidae in Australia (Tarmann 2004). The early lepidopteran larval instars are without urticating setae, although the cellular apparatus that produces them is present; in particular in the processionary moths, setae emerge in the third stage and increase during every molts. At

the same time other species like the lymantriid *Euproctis* and the Thaumetopoeinae *Ochrogaster lunife*r and *Anaphe panda* use the urticating setae to protect also the eggs. The setae have lost the neuronal connection and they have the proximal end detached from the integument and inserted into a socket (Lamy et al. 1982); they can easily be removed with any kind of mechanical stimulation in case of danger. Usually, the setae are short (100-500 µm long, 3-7 µm in diameter), have barbs along the shaft and occur in special parts of the insect body called mirrors (Foelix et al. 2009) (Démolin 1963) that increase in number and size as the larva molts. When the larva is undisturbed, the mirrors are closed and from outside it is possible to see only the distal part of the setae located on the outer border of the mirror. The density of setae can be very high, from 60,000 setae/mm^2 in the Lepidoptera larvae to 10,000-12,000 setae/mm^2 in spiders (Lamy et al. 1982). Spider setae are situated on the abdomen and start to develop in the third instars, that is the time when the spider molts to the adult stage (Battisti et al. 2011).

Modified setae

Modified setae occur especially on larvae of Lymantriidae (Deml and Dettner 1995), Lasiocampidae, Arctiidae (Kawamodo and Kumada 1984), Anthelidae (Balit et al. 2004) and other small families. Modified setae, have a low density on different part of the body and can easily broken off from the integument; they are longer than true setae (up to 1 mm), are rather stiff and have barbs of various size along the shaft. The mature larvae have the capacity to incorporate the setae into the cocoon as an external protection in case of risk (Balit et al. 2004).

Spines

Spines are typically present in Lepidoptera Megalopygidae, Limacodidae and Saturniidae (Gilmer 1925) and rarely in other lepidopteran families. Respect the other setae, have a complex structure, are stiff, filled with a secretion and respond to mechanical stimulation. Their size may be vary and the diameter is larger than 3-7 µm. Spines can be alone or combined with modified setae.

Medical Aspects

The hairs are well-known to cause a number of diseases to humans and other warm-blooded animals (Arlian 2002, Mullen 2009). The problems are correlated to the arthropod density and the risk is for humans and animals. Modified setae and spines are generally present at a much lower density on the insect body compared with true setae and are not released into the air. The urticant capacity of these hairs is known from the antiquity (Roques and Battisti 2014), although the first descriptions correspond to Reaumur (1736) and Fabre (1900). The Lepidoptera in particular are known to cause 'lepidopterism', which indicates every kind of pathological condition produced by them. Lepidopterism may be further divided into two categories: *lepidopterism* in a narrow sense refers to syndromes caused by adult moths and butterflies, whereas *erucism* refers to reactions resulting from contact with urticating hairs of larvae or pupae (Delgado 1978; Kawamoto and Kumada, 1984). The use of these terms, however, is problematic because *erucism* and *lepidopterism* refer to different life stages of the insects (larvae and adult, respectively) rather than to reactions in humans. The distinction of these syndromes, based on local and general reaction is complicated because true setae can be released by the insects, either larvae or adults, and distributed in the environment by air. For this reason these definitions are of limited value when the effects of true setae on humans are described in clinical practice. Because modified setae and spines are normally not released into the air, they are more likely associated with reactions from direct contact (Battisti et al. 2011). Urticating hairs can also affect domestic animals (Mullen 2009), the ingestion of caterpillars or only urticating hairs, may have dramatic consequences like tongue necrosis in dogs (Jans 2008) (Moneo et al. 2014). Most of the information about the release of true setae in Lepidoptera are related to *Thaumetopoea spp*. (Démolin 1963, Lamy 1982).

Processionary moths (Lepidoptera: Notodontidae, Thaumetopoeinae)

The Thaumetopoeinae clade comprises 111 species in 20 genera occurring in Africa, Madagascar, Europe, India and Australia (Kiriakoff 1970, Schintlmeister 2013). Most of the species studied in this thesis are distributed in Europe except the African *Anaphe panda* and the Australian *Ochrogaster lunifer*. The genus *Thaumetopoea* includes a high number of species distributed mainly in the Mediterranean region. The larvae feed on trees and shrubs such as Pinaceae (pine, cedar), Anacardiaceae (pistachio, sumac) and

Cistaceae, with the exception of one species feeding on Fagaceae (oak) (Simonato et al. 2013). The species considered in this studies, especially in Chapter 4, are listed below.

Table 1. List of the Thaumetopoeinae species analysed in this thesis

Family	**Species**	
Notodontidae	*Anaphe panda*	Biosduval
Notodontidae	*Ochrogaster lunifer*	Herrich-Shäffer
Notodontidae	*Thaumetopoea bonjeani*	Powell,1922
Notodontidae	*Thaumetopoea herculeana*	Rambur 1840
Notodontidae	*Thaumetopoea ispartaensis*	Doganlar and Avici 2001
Notodontidae	*Thaumetopoea libanotica*	Kiriakoff and Talhouk
Notodontidae	*Thaumetopoea pinivora*	Gotland, Sweden
Notodontidae	*Thaumetopoea pityocampa*	Denis and Schiffermüller
Notodontidae	*Thaumetopoea pityocampa ENA* *(Eastern-Northern Africa)*	
Notodontidae	*Thaumetopoea processionea*	Linnaeus, 1758
Notodontidae	*Thaumetopoea solitaria*	Freyer, 1838
Notodontidae	*Thaumetopoea wilkinsoni*	Tams, 1925

Phylogenetic studies linked to ecological and life history traits identified three main clades of processionary moth. Clade A includes *Th. herculeana*, *Th. processionea* and *Th. Solitaria*, clade B *Th. pityocampa* ENA, *Th. pityocampa* and *Th. Wilkinsoni*, clade C *Th. bonjeani*, *Th. ispartensis*, *Th. libanotica* and *Th. pinivora.* All species feeding on Pinaceae belong to clades B and C, while all species feeding on Angiosperms are grouped in clade A. The clades B and C showed different adaptations, the first with a switch of larval feeding to cold season and the second with a retraction to high altitude and latitude and a development cycle extended over two years (Fig. 2). The clade A has a fast spring development of larvae (Simonato et al. 2013).

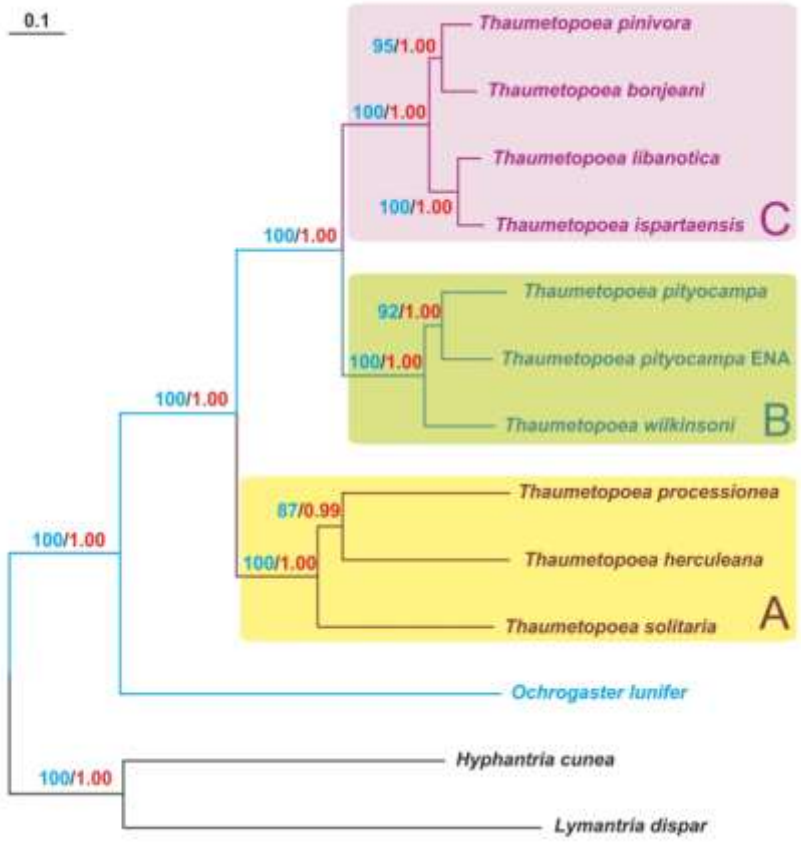


Figure 2. Phylogenetic scheme of some processionary moth species described in Simonato et al. 2013

Urticating system in processionary moths

Setae production and morphology was first studied in the pioneering work of Fabre (1900), continued by Démolin (1963), Lamy et al. (1982) and Novak et al. (1987) (Moneo et al. 2014). *Thaumetopoea* larvae use the setae to protect themselves against predators; when are disturbed, the larvae opens the mirror releasing the setae (Fig. 4) that are carried by the wind away from the tent and persist in the environment for long time; for this reason their nature is very different from those of other types of defensive hairs occurring in Lepidoptera, which are part of the integument and require contact with the larva to cause the reaction (e.g. the larvae of Saturniidae, Megalopygidae and Limacodidae), while they are similar to urticating setae released by some spiders from America (Theraphosidae) (Battisti et al. 2011). The urticating setae are renewed at each molt since the 3rd larval instar (Fig. 3) and the larval exuvia may carry the setae that were not dispersed during the previous instar. All studied species of the Thaumetopoeinae are known to carry urticating setae, either as a larva (genus *Thaumetopoea*) or as an adult,

which use them to protect eggs and larvae of the next generation (e.g. *Anaphe* from Africa and *Ochrogaster* from Australia) (Lamy et al. 1984, Floater 1998). The only exception was *Th. herculeana*, that was not considered urticant in literature. During my thesis I had the possibility to better explore the life history of this species (see below in the Supplementary chapters) and I found that also *Th. herculeana* presents the same urticating system of the other *Thaumetopoea*, but not very efficient because the mirrors do not open upon disturbance and the setae are not released. When applied to the human skin of a sensitized person, however, they produce the same reactions caused by the setae of the other *Thaumetopoea*.

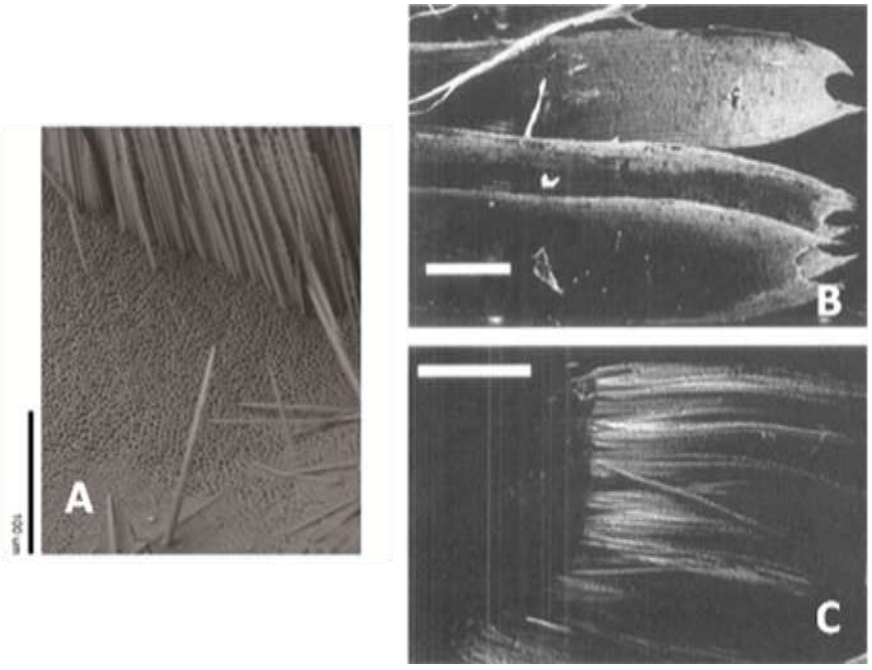


Figure 3. Scanning of electron photographs of the urticating setae in *Thaumetopoea pityocampa* (A) and scales of *Ochrogaster lunifer* (B,C). The holes in A indicate the sockets where the sharp proximal end of the setae is inserted. The photo B of the scales represents the apices of flat scales and other photo (C), section of anal tuft that showing base of flat and filamentous scales (Floater 1998).

The mirrors (Fig. 4) increase with the larval instar and are actively open when the larva is disturbed (Démolin 1988). Studies of the seta system of processionary moths, showed a wide variation in seta length like in case of *Th.pityocampa* the longest (680 µm) were approximately 14 times longer than the shortest (50 µm) (Petrucco Toffolo et al. 2014).

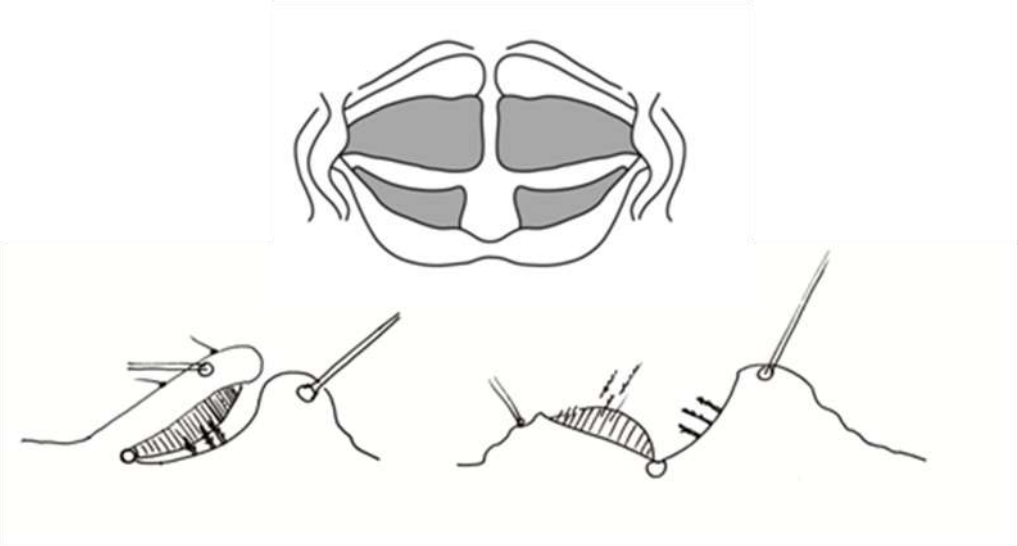


Figure 4. Schematic drawing of the mirrors of *Th. pityocampa* and scheme of the opening of the setae field and release of the urticating setae in *Th. bonjeani* (modified from Démolin 1988).

Epidemiological studies

Th. pityocampa larvae induce skin lesions such as urticaria or dermatitis, conjunctivitis and rarely respiratory symptoms or even anaphylactic shock. Is clear that the mechanical lesion induced by the penetration of the setae in the skin can be responsible for the urticaria that is always present, but the setae may contain toxins and proteins which may contribute to the inflammation (Ducombs et al. 1979). The first symptoms and signs of uricaria include itching, local swelling, red discoloration (erythema) (Hossler 2009) and allergic respiratory response, in rare case anaphylactic reactions (Gottschling 2007). The mild reaction after mild symptoms disappeared in three weeks. Often, the reactions occur in isolated persons; than can be related to exposure in areas of heavily infested tree or in areas close from the infested trees, but with a strong wind (Moneo et al. 2014). Most studies described isolated cases of persons, such as one of *Th. processionea* (6% of 1,025 persons living in radius of 500m from infested trees) (Maier et al. 2003) one of *Th. pinivora* (18% of 4,300 persons living in an infested area of about 3,500 ha) (Holm et al. 2009). Two cases of *Th. pityocampa* confirm that these reactions are common in endemic areas. In the first case 9% of 653 children living in a rural area had reactions, in the second case 12% in rural areas, 10 % in semi-urban areas and 4 % in urban areas of 1,224 adults (Vega et al. 2003). The data shows that peripheral urban areas with nearby pine/oak trees are also areas with a high incidence of *Thaumetopoea* reactions (Moneo et al. 2014).

The urticating mechanism

Setae are build up by chitin, a polysaccharide biopolymer composed of N-acetyl-β-D-glucosamine and proteins, and are covered by a layer of tannin-bound lipoprotein, wax, and mucopolysaccharides (Chapman 1998). The percentage of chitin present in the urticating setae of *Th. pityocampa* in a population of Northern Italy was as high as 36.8% (unpublished data), bringing us to think that the rest is mainly protein. The chitin action is not clear yet, although it can stimulate alone in vitro human T-lymphocyte proliferation (Holm et al. 2014). A much stronger proliferation, however, is induced by setae of *Th. pinivora* in persons previously exposed to the setae, indicating that setae contain molecules which may start cell-mediated immune response (Holm et al. 2014). In the hypothetical scheme proposed by Battisti et al. (2011), the penetration of the seta into the skin induces the macrophages to produce chitinases together with proinflammatory cytokines and other inflammatory and immunoregulatory mediators that break down chitin and start complex interactions with cells of the immune system like T-lymphocytes (Lee et al. 2008). Chitinase break down the chitin skeleton of the setae and chitin fragments, proteins and other antigenic components are released. Setae proteins are taken up and processed by antigen cells (APCs) and presented to lymphocytes for a specific immune response. Chitin fragments promote inflammation and further proliferation of lymphocytes (Fig. 5).

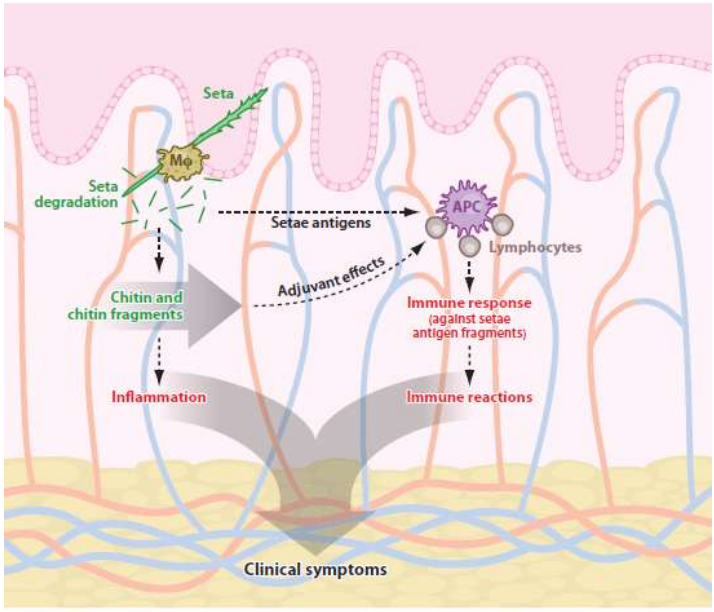


Figure 5. Tentative role of setae as inducers of inflammation and immune reaction in the skin (from Battisti et al. 2011) (see text for an explanation).

The most likely function of the setae seems to be the protection against the vertebrate predators of these insects (Battisti et al. 2011). Indirect evidence comes from the observation that generalist predators avoid the larvae of the pine processionary moth *Thaumetopoea pityocampa,* with the exception of few bird species (cuckoo, hoopoe, tits) which developed foraging techniques to avoid the ingestion of the setae (Barbaro and Battisti, 2011).The reaction correlated to the contact with the setae appears to be quite complex. While the activity of chitin is still under investigation, the presence of a protein with allergic activity has been studied and demonstrated in *Thaumetopoea pityocampa.* Three different antigens reactive with IgE antibodies of previously exposed persons have been described in both larval and setae extracts. The first work was done by Lamy et al. (1986), who described a 28 kDa dimeric protein called thaumetopoein extracted from the setae. This protein is formed by two subunits, one of 13 kDa and the other of 15 kDa. Two years later, the same scientific group described a homologue of thaumetopoein in the setae of the oak processionary larva (*Th. processionea*) (Lamy et al. 1988). This protein exhibited the same urticating effect as thaumetopoein in the guinea pig skin. Later, Moneo et al. (2003) purified a 15 kDa protein from larvae extracts, named it Tha p 1 and identified it as a major allergenas it was recognized by IgE antibodies from previously exposed persons. This protein was purified by ethanol fractionation by differential precipitation of a whole larval extract followed by separation by a reversed-phase high performance liquid chromatography (RP-HPLC). The amino acid terminal sequence GETYSDKYDTIDVNEVLQ for Tha p 1 was obtained and at that time no similarities with other proteins were found using the web interface BLAST of the USA National Centre for Biotechnology Information (NCBI). Several years later, however, the complete sequencing of the silkworm *Bombyx mori* genome led to classify Tha p 1 as a chemosensory protein (Larsson and Backlund 2009), similar to those found in this species (Picimbon et al. 2000). Mature Tha p 1 mRNA (Fig. 6) could be sequenced isolating the larva RNA using the traditional method of TRIzol and performing a retrotranslation followed by a polymerase chain reaction (RT-PCR).

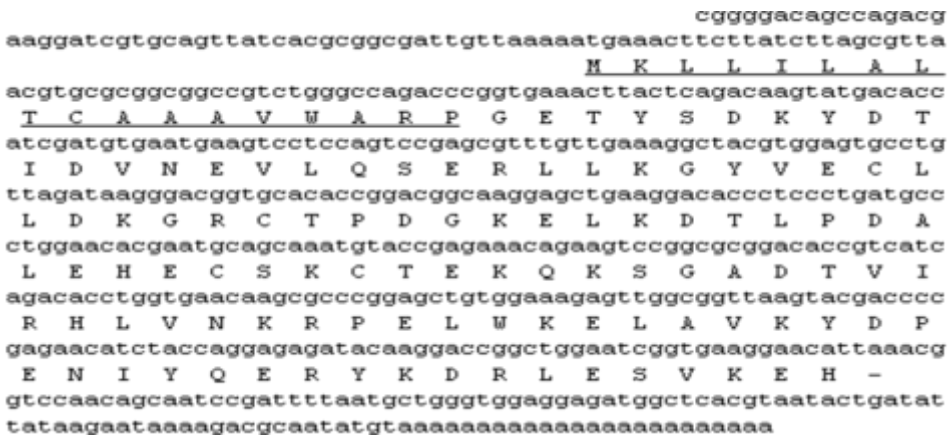


Figure 6. Complete sequence of Tha p 1 mRNA and its translation to amino acids. Stop codon is identified as a dash. Signal peptide is underlined (European Nucleotide Archive accession number HE962022) (from Moneo et al. 2014).

Tha p 1 was modelled upon the crystalline structure of the chemosensory protein 10 from *Bombyx mori*, using the web interface Fist approach mode of Swiss-model (Moneo et al. 2014). Despite the high homology between the chemosensory proteins of *Th. pityocampa* and *B. mori*, person sensitized to the pine processionary larva did not recognize any protein of a silkworm whole body crude extract (unpublished data by Moneo). The second protein Tha p 2 was extracted from the setae by Rodriguez-Mahillo et al. (2012) and characterized as a 14 kDa protein using biochemical and molecular techniques (Fig. 7). It has no similarity with Tha p 1 and it may correspond to the thaumetopein described in 1985, also extracted from the setae, but unfortunately no information about the sequence of this allergen was provided. It is interesting to note that Tha p 2 showed similarity in the carboxyl terminal region to a hypothetical protein of *Acyrthosiphon pisum*, the pea aphid. This fact suggested that both proteins could be members of an unknown family of insect proteins and that more allergens of this family could be found in the future.

Figure 7. Tha p 2, a protein of 14 kDa isolated from the setae of *Th. pityocampa)* (Rodriguez-Mahillo et al. 2012).

Objectives and content of the thesis

Most of my work has concerned Tha p 2 in a comparative study aiming at identify and characterize the orthologous counterparts in all the *Thaumetopoea* moths, as well as in other species of Thaumetopoeinae, and understand the evolution of the protein and the nature of encoding genes (Chapter 4). In this perspective, a part of the study focalized on the expression profile of Tha p 1 and Tha p 2 in every life stage of *Thaumetopoea pityocampa* (Chapter 3).This study was conduct with technique of NGS (Next-Generation Sequencing) that revolutionized genomics research by bringing the sequencing of entire genomes or transcriptome like in our case. NGS generate a large number of reads but in higher speed, lower cost, and small instrument size. In this case, the transcriptome research increased the speed and accuracy of discovering new gene and their expression, as well as determining how gene transcription variation are regulated in different life stages of the insects. NGS offers the opportunity to generate genome data sets (Wang et al. 2009). In particular, RNA-seq is a recently developed large-scale genome wide approach that has been applied successfully to gene discovery and expression profiling, and to the study of functional, comparative and evolutionary genomics. RNA-seq make reference to a transcriptome produced by methods of NGS, which ensure a good coverage of transcripts detection (Wang et al. 2009, Oshlack et al. 2010). The transcriptome involve all types of ribonucleic acids (RNAs), including the protein coding messenger ribonucleic acid (mRNA) and the non-coding ribonucleic acid (ncRNA) such as ribosomal RNAs (rRNA), transfer RNAs (tRNA), and the small nuclear RNAs (snRNA). These RNAs may be differentially expressed according to the tissue, the stage of

development and the physiological condition being accessed (Wang et al. 2009, Anders and Huber 2010). The RNA-seq approaches has emerged as a new approach for the transcriptome complexity. A transcriptome consists in taking a photograph from a specific time in a cell, highlighting only the condition at that short period of time. After that bioinformatics analysis is also an important step and includes the use of computational tools that guarantee the processing of large volumes of data generated by NGS (Gavery and Roberts 2012, Guo et al. 2013). Another aim of this study was the characterization of new possible urticating proteins involved in the defense mechanism of the *Thaumetopoea* species (Chapter 2). Tha p 2 protein showed a IgE and IgG recognition patterns of 18 *Th. pityocampa* sensitized person from whole body and setae extracts (in Rodriguez-Mahillo 2012 and Moneo et al. 2014, Fig. 8). Several allergens present in the setae extract were detected, although the IgG recognition patterns of the whole body and the setae extracts were much weaker then the IgE. Our studies focalized on the characterization of all the proteins, urticating and not, occurring in the setae.

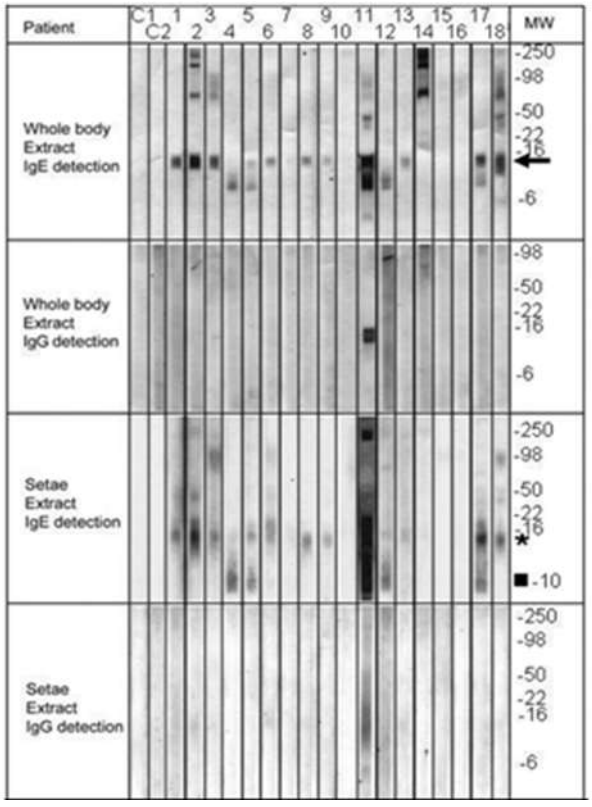


Figure 8. IgE and IgG detection of proteins of *Th. pityocampa* whole body and setae. Individual sensitized person (1-18) and healthy donors (C1-C2) sera were tested. Tha p 1 is marked with an arrow while Tha p 2 is marked with an asterisk (Moneo et al. 2014). The black square indicates another protein which is recognized by sera, parlty sequenced and called Tha p 3 by Moneo et al. (2014)

References

Anders S., Huber W., (2010). Differential expression analysis for sequence count data. *Genome Biology* 11:106.

Arlian L.G., (2002). Arthropod allergens and human health. *Annual Review of Entomology* 47:395–433.

Balit C.R., Geary M.J., Russell R.C., Isbister G.K., (2004). Clinical effects of exposure to the Whitestemmed gum moth (*Chelepteryx collesi*). *Emergency Medicine Australasia* 16:74–81.

Battisti A., Holm G., Fagrell B., Larsson S., (2011). Urticating hairs in arthropods: Their nature and medical significance. *Annual Review of Entomology* 56:203–220.

Chapman R.F., (1998). *The Insects. Structure and Function*. Cambridge, UK: Cambridge Univ. Press pp.770

Cooke J.A., Roth V.D., Miller F.H., (1972). The urticating hairs of theraphosid spiders. *American Museum Novitates* 2498:1–43.

Deml R., Dettner K., (1995). 'Balloon hairs' of gypsy moth larvae (Lep., Lymantriidae): morphology and comparative chemistry. *Comparative Biochemistry and Physiology* 112:673–81.

Delgado Q.A., (1978). Venoms of Lepidoptera. In: *Arthropod Venoms*, ed. S Bettini, pp. 555–611. Berlin: 29. An application of Springer.

Démolin G., (1963). Les 'miroirs' urticants de la processionnaire du pin (*Thaumetopoea pityocampa* Schiff.). Agricultural Zoology Reviews 10–12:107–14.

Démolin G., (1988). Intensification de la protection phytosanitaire des foréts. Algérie 1986-1987. La processionarie du cèdre: *Thaumetopoea bonjeani*. *FAO, Rapport scientifique et rapport iconographique*, Roma, Italy, p.21

Ducombs G., Lamy M., Bergaud J.J, Tamisier J.M., Gervais C., Texier L., (1979). La chenille processionaire (*Thaumetopoea pityiocamapa*, Shiff, Lepidoptéra) et l'homme. Etude morphologique de l'appareil urticant. Enquete épidémiologique, *Annales de dermatologique et de véneréologie*, 106, 769-778.

Fabre J.H., (1900). Souvenirs Entomologiques, E´ tudes sur l'Instinct et les Moeurs des Insectes. Paris: *Delagrave*.

Floater G.J., (1998). Tuft scales and egg protection in *Ochrogaster lunifer* Herrich-Schäffer (Lepidoptera: Thaumetopoeidae). *Australian Journal of Entomology* 37:34–39.

Foelix R., Rast B., Erb B., (2009). Palpal urticating hairs in the tarantula Ephebopus: fine structure and mechanism of release. *Journal of Arachnology* 37:292–98.

Gavery M.R., Roberts S.B., (2012). Characterizing short read sequencing for gene discovery and RNA-Seq analysis in Crassostrea gigas. *Comparative Biochemistry and Physiology*. Part D 7, 94–99.

Gilmer P.M., (1925). A comparative study of the poison apparatus of certain lepidopterous larvae. *Annals of the Entomologcal. Society of America.* 18:203–39.

Gottschling S., Meyer S., Dill-Mueller D., Wurm D., Gortner L., (2007). Outbreak report of airborne caterpillar dermatitis in a kindergarten. *Dermatology* 215:5–9.

Guo Y., Li J., Li C.I., Shyr Y., Samuels D.C., (2013). MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis. *Bioinformatics*, 29:1210–1211.

Holm G, Sjoberg J, Ekstrand C, Bjorkholm M, Granath F, Ekbom A., Larsson S., (2009). Tallprocessionsspinnare stort halsoproblem pa sodra Gotland. *Lakartidningen* 106:1891–94.

Holm G., Andersson M., Ekberg M., Fagrell B., Sjöberg J., Bottai M., Björkholm M., (2014). Setae from Larvae of the Northern Processionary Moth (*Thaumetopoea pinivora, TP*) Stimulate Proliferation of Human Blood Lymphocytes In Vitro. *PLoS ONE* 9(12): e113977.

Hossler E.W., (2009). Caterpillars and moths. *Dermatology and Therapy* 22:353–66.

Kawamoto F., Kumada N., (1984). Biology and venoms of Lepidoptera. In: *Handbook of Natural Toxins*. Vol. 2. Insect Poisons, Allergens, and other Invertebrate Venoms, ed. AT Tu, pp. 291–330. New York: Dekker.

Kiriakoff S.G., (1970). Lepidoptera Familia Thaumetopoeidae. In: Wytsman, P. (Ed.), *Genera Insectorum*. Anvers Belgium: SPRL Mercurius. pp. 1–54.

Jans H.W.A., Franssen A.E.M., (2008). The urticating hairs of the oak processionary caterpillar (*Thaumetopoea processionea* L.), a possible problem for animals? *Tijdschr Diergeneesk* 133:424–29.

Lamy M., Ducombs G., Pastureaud M.H., Vincendeau P., (1982). Productions t´egumentaires de la processionnaire du pin *(Thaumetopoea pityocampa Schiff.)* (L´epidopt`eres). Appareil urticant et appareil de ponte. *Bullettin de la Société Zoologique de France* 107:515–29.

Lamy M., Pastureaud M.H., Novak F., Ducombs G., (1984). Papillons urticants d'Afrique et d'Am´erique du sud (g. *Anaphae* et g. *Hylesia*): contribution du microscope ´electronique `a balayage `a l'´etude de leur appareil urticant et `a leur mode l'action. *Bullettin de la Société Zoologique de France* 109:163–77.

Lamy M., Novak F., Duboscq, M.F., Ducombs G., Maleville J., (1988). The oak Processionary caterpillars (Thaumetopoea processionea L) and man -Urticating apparatus and mechanism of action. *Annales de Dermatologique et de Venereologique* 115: 1023-1032.

Larsson S, Backlund A., (2009). Regarding the putative identity of moth (Thaumetopoea spp.) allergen. *Allergy* 64: 493.

Lee C.G., Da Silva C.A., Lee J.Y., Hartl D., Elias J.A., (2008). Chitin regulation of immune responses: an old molecule with new roles. *Current Opinion in Immunology* 20: 684–89.

Maier H, Spiegel W, Kinaciyan T, Krehan H, Cabaj A, (2003). The oak processionary caterpillar as the cause of an epidemic airborne disease: survey and analysis. *British Journal of Dermatology* 149: 990–997.

Mullen G.R., (2009). Moths and butterflies (Lepidoptera). In: *Medical and Veterinary Entomology*, ed. GR Mullen, LA Durden, pp. 353–70. Amsterdam: Elsevier.

Moneo I., Battisti A., Dufour B., García-Ortiz J.C., González-Munoz M., Moutou F., Paolucci P., Petrucco Toffolo E., Rivière J., Rodriguez-Mahillo A.I., Roques A., Roques L., Vega J.M., Vega J., (2014). Medical and Veterinary Impact of the Urticating Processionary Larvae. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp: 359-410.

Moneo I., Vega J.M., Caballero M.L., Vega J., Alday E., (2003). Isolation and characterization of Thap 1, a major allergen from the pine processionary caterpillar *Thaumetopoea pityocampa*. *Allergy* 58: 34–37.

Novak F., Lamy M., (1987). Etude ultrastructurale de la glande urticante de la chenille processionnaire du pin, Thaumetopoea pityocampa Schiff (Lepidoptera: Thaumetopoeidae). *International Journal of Insect Morphology and Embriology* 16: 263–70.

Oshlack A., Robinson M., Young M., (2010). From RNA-seq reads to differential expression results. *Genome Biology* 11*:* 220. doi: 10.1186/gb-2010-11-12-220.

Petrucco Toffolo, E., Zovi, D., Perin, C., Paolucci, P., Roques, A., Battisti, A., Horvath, H. (2014). Size and dispersion of urticating setae in three species of processionary moths. *Integrative Zoology* 9: 320–327.

Picimbon J.F., Dietrich K., Angeli S., Scaloni A., Krieger J., Breer H., Pelosi P., (2000). Purification and molecular cloning of chemosensory protein from *Bombyx mori*. *Archives of Insect Biochemistry and Physiology*, 44: 120-192.

Reamur R.A.F., (1736). Memoire pour Servir a l'Histoire des Insectes. Paris: *Imprimerie Royale*

Rodríguez-Mahillo A.I., González-Muñoz M., Vega J.M., López J.A., Yart A., Kerdelhué C., Camafeita E., Garcia Ortiz J.C., Vogel H., Toffolo E.P., Zovi D., Battisti A., Roques A., Moneo I., (2012). Setae from the pine processionary moth (Thaumetopoea pityocampa) contain several relevant allergens. *Contact Dermatitis* 67: 367–374.

Roques A., Battisti A., (2014). Introduction. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp: 1-13.

Rothschild M., Reichstein T., Lane N.J., Parsons J., Prince W, Swales SW., (1970). Toxic Lepidoptera. *Toxicon* 8: 293–99.

Schintlmeister A., (2013). *World Catalogue of Insects 11 Notodontidae & Oenosandridae (Lepidoptera)*. Boston: EJ Brill. p.605.

Simonato M., Battisti A., Kerdelhué C., Burban C., Lopez-Vaamonde C., et al., (2013). Host and phenology shifts in the evolution of the social moth genus *Thaumetopoea*. *PloSOne* 8(2) e57192.

Tarmann G.M., (2004). ZygaenidMoths of Australia: A Revision of the Australian Zygaenidae, *Australia: CSIRO* Vol. 9..

Vega M.L., Vega J., Vega J.M., Moneo I., Sanchez E., Miranda A. (2003). Cutaneous reactions to pine processionary caterpillar (*Thaumetopoea pityocampa*) in pediatric population. Pediatric. *Immunology and Allergy* 14:482–86.

Wang Z., Gerstein M., Snyder M., (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57-53.

Werno J., Lamy M., (1991). Urticating hairs of "brown-tail" moth (Euproctis chrysorrhoea L.) (Lepidoptera): preliminary studies in urban and laboratory. *Comptes Rendus de l'Académies des Sciences III* Vie 312:455–59.

Witz B.W., (1990). Antipredator mechanism in arthropods: a twenty year literature survey. *The Florida Entomologist* 73: 71-99.

Chapter 2

Identification of proteins from urticating setae of the pine processionary moth and human recognition

Authors:

Berardi L., Pivato M., Mitali E., Francescato E., Pantera B., Trentin A.R., Masi A., Olivieri M., Battisti A.

I contributed to all experimental parts, data analysis and chapter writing

Paper in preparation for submission.

Abstract

The larvae of the pine processionary moth produce urticating setae which are likely used for protection against vertebrate predators. Contact with urticating setae by humans and animals induces dermatitis, usually located in the exposed areas. Reactions are common in pine wood area workers, who are exposed to high levels of setae, but also in persons non-occupationally exposed to processionary larvae, such as farmers living near infested areas. Recent studies demonstrated the presence of a complex urticating mechanism where the proteins present in the urticating setae may play a role as activators of immune responses. A complete data set of all proteins, , occurring in the setae is not available. In this work, two different protein extraction protocols of different strength were tested and a total of 182 urticating and not-urticating proteins were obtained. We confirm that the setae of *Th. pityocampa* contain many proteins, some of which are recognized by Ig-E of persons manifesting dermatitis when exposed to the larvae of this insect. In addition, we add information about the type, quality and quantity of the proteins associated with the setae.

Introduction

Pine processionary moth (*Thaumetopoea pityocampa*) larvae are a medical and veterinary problem in Mediterranean countries, Southern Europe, Middle East and Northern Africa because of the reactions caused to humans and warm-blooded animals (Battisti et al. 2011). Reactions are common in pine wood area workers, who are exposed to high levels of setae, but also in persons non-occupationally exposed to processionary larvae, because of the setae are dispersed by the wind on long distance (Fenk et al. 2007, Petrucco Toffolo et al. 2014). Contact dermatitis and urticaria are the most frequent symptoms related to the exposure to this insect (Vega et al. 1999).

In addition to the mechanical effect of the seta penetration in the skin, an IgE-mediated mechanism has been demonstrated as inducer of clinical symptoms (Moneo et al. 2014). Lamy et al. (1986) were the first to describe proteins, and their potential antigenic action, of *Th. pityocampa* listing a complex mixture of 16 proteins extracted from urticating setae. Among them, a 28 kDa protein not associated to carbohydrates or lipids was identified and called thaumetopoein, further separated in two subunits of 13 and 15 kDa. This protein induced mast cell degranulation by a non-immune mechanism (Lamy et al. 1985). Two years later, the same group described a homologue of

thaumetopoein in the setae of the oak processionary larva *Th. processionea* (Lamy et al. 1988) with the same effect. Additionally, a 45 kDa protein extracted from pine processionary setae gave also a strong reaction with IgE (Werno et al. 1993). Later, several IgE-binding proteins putatively involved in allergic reactions were described by Moneo et al. (2003), using a crude extract of whole larvae in the last stage (L5). They identified a protein named Tha p 1, with molecular mass of 15 kDa as a major allergen, but could not compare its amino acid sequence with thaumetopoein because it had not been sequenced. The amino acid terminal sequence GETYSDKYDTIDVNEVLQ for Tha p 1 was obtained but the full sequence was published only in 2014 (Moneo et al. 2014). In the while, the Tha p 1 was found to belong to the chemosensory protein family by Larsson and Backlund (2009), by comparing it with the complete sequencing of the silkworm *Bombyx mori* genome, where there are similar proteins (Picimbon et al. 2009).

Another protein with antigenic role against sera of previously exposed persons was described by Rodriguez-Mahillo et al. (2012) from a seta extract. This protein has been sequenced and called Tha p 2, with molecular mass around 14 kDa; in addition, these authors demonstrated that the setae contain a complex mixture of about 70 proteins, some of those recognised by human sera. Very recently, Holm et al. (2014) have shown that cultivated human lymphocytes of previously exposed persons proliferate when in contact with setae and setae extracts of the closely related species *Thaumetopoea pinivora*, indicating the occurrence of a cell-mediated immune response. As proteins of the setae are bound to chitin, these authors also explored the possibility that chitin per se was able to induce lymphocyte proliferation in vitro and the result was positive, although not different between persons previously exposed or not.

The aim of this work is to establish a clear data set of the proteins extracted from the urticating setae of the pine processionary larvae, by testing different protein extraction methods and creating a large profile of proteins to be used in future work on allergic reactions.

Materials and Methods

Setae collection

Larval colonies of *Th. pityocampa* were collected in March 2013 from *Pinus nigra* trees in Tregnago (Verona, Italy). Each individual of the larval instar L4 and L5 was taken out from the tent in a hood and put in an Eppendorf tube at -20°C. Later the setae were

removed from thawing larvae with forceps under stereomicroscope with x40 magnification, inside a hood. About 2,000,000 of urticating setae were collected from 20 larvae.

Protein extraction from the setae

According to the method developed by Olivieri et al. (2013), the setae were put in an Eppendorf with 100 µl of H_2O solution and proteins were extracted in ice using an homogenizer, with two cycles of 30 seconds at maximum speed. The sample was then centrifuged at 12,000 xg for 30 minutes and the supernatant stored at -80°C until used. The supernatant was then quantified by the DC Protein assay reagents. We also used a variant of this method (called here Hepes), consisting in using acetone instead of water, which allows a quick detachment of the setae from the forceps, and the following procedure. The setae were frozen in liquid nitrogen, broken by sonication, then centrifuged for 2 minutes at 16,000 rpm and crushed in 1.5 ml extraction buffer made up by 50 mM Hepes pH 8, 1% Triton X-100 1M NaCl, 1mM PMSF/Benzamidin. The preparation was then stirred for 4 hours at room temperature and centrifuged for 20 minutes at 14,000 rpm at 4°C. The clear supernatant was taken and the proteins were precipitated in 100% acetone at -20°C over night and centrifuged again at 12,000 rpm for 10 minutes. The supernatant was eliminated while the pellet was washed with acetone 80% and resuspended in Laemmli solution. The protein content of the solution was measured according to the Bradford method (Kruger 1994) and by the use of spectrophotometers.

Protein in situ digestion

The extracts obtained from the water method was used for SDS-PAGE (Sodium dodecylsulfate polyacrilamyde gel electrophoresis) using RunBlue Tris-MOPS 4-20% precast minigels according to manufacturer's instructions, while those obtained with the Hepes method were loaded into a homemade 13% gel . Proteins were visualized by Coomassie staining. The bands were excised and washed several times with 50 mM TEAB (triethylammonium bicarbonate) and dried in the vacuum after a short acetonitrile wash. Cysteines were reduced with 10 mM dithiothreitol (in 50 mM TEAB) for 1 hour at 56°C, and alkylated with 55 mM iodoacetamide (in 50 mM TEAB) for 45 minutes at room temperature in the dark. Gel pieces were then washed with alternate steps of TEAB

and acetonitrile, and dried. Proteins were digested in situ with sequencing grade modified trypsin (Promega, Madison, WI, USA) at 37°C overnight (12.5 ng/µl trypsin in 50 mM TEAB). Peptides were extracted with three steps of 50% acetonitrile in water. One µg of each sample was withdrawn to check digestion efficiency using LC-MS/MS analysis, and the remaining peptide solution was dried in the vacuum.

LC-MS/MS analysis and database search, and protein quantification

The extract obtained from the Hepes method was suspended in H_2O/0.1% formic acid and analyzed by LC-MS/MS. The MS analyses were conducted with a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Pittsburgh, CA, USA) coupled online with a nano-HPLC Ultimate 3000 (Dionex - Thermo Fisher Scientific). Samples were loaded onto a homemade 10 cm chromatographic column packed into a pico-frit (75 mm id, 10 mm tip, New Objectives) with C18 material (ReproSil, 300 A°, 3 mm). The LC separation and mass spectrometer instrumental settings used for the analyses were the same as those described in Tolin et al. (2013) and the method was as described by Köcher et al. (2009). The raw LC-MS/MS files were analyzed using Proteome Discoverer 1.4 (Thermo Fisher Scientific), connected to a Mascot Search Engine server (Matrix Science, London, UK). The spectra were searched against a transcriptome protein database provided by Centre de Biologie et Gestion de Populations (INRA Montpellier, France). Enzyme specificity was set to trypsin with two missed cleavages, and peptide and fragment tolerance was set to 10 ppm and 0.6 Da, respectively. Based on the search against the corresponding randomized database, false discovery rates (FDR) of 5% were calculated by the Proteome Discoverer. The data were pre-filtered to exclude MS/MS spectra containing less than 5 peaks or with a total ion count below 50.

Bioinformatics analyses

All proteins extracted by the Hepes method were identified with at least two independent peptides with a high degree of confidence. The list of protein was exported in Excel further filtering and cleaning of the same peptides. BLASTp (Basic Local Alignment Search Tool) provided by NCBI web site (Altschul et al., 1990) was used for indentifying any possible homologous protein, by selecting the proteins with the highest identity value. Study of function of proteins was provided by UniProt (Bairoch et al. 2005). A dataset with all BLAST proteins, their description, the BLAST species, % of coverage, and

number of peptide mapped was also created, together with a dataset of proteins with the same function.

Protein recognition by human sera

Proteins obtained by the extraction in water were transferred onto PVDF (polyvinylidene difluoride membrane) using a Gel Casting Kit (GE Healthcare Hoefer^{TM}) for 1 hour at 25V and 400mA. Blots were incubated with a pool of sera of 5 forestry workers. These persons were suffering from pruritic skin reactions, especially between February and April, when the larvae are more active and they were working in areas with pine trees infested by *Th. pityocampa*. The sera were diluted 1:2 in PBS (Phosphate Buffered Saline) containing 3% bovine serum albumin and 0.05% Tween for 16 hours at 4°C; then washed and probed with secondary antibodies (anti-Human IgE) conjugated with horseradish peroxidase for 1 hour at 25°C. After washing, the membrane was developed with the enhanced chemiluminescence kit ECL Immun-star HRP (BioRad). Chemiluminescence was detected using the ChemiDoc imaging system (BioRad). Quantity One software (BioRad) was used to perform the quantitative analysis of the bands.

The protein product of the Hepes extraction was used for a dot blot screening of 16 persons taken randomly from three different group of persons sera for a representative number of persons different for symptoms and exposition, grouped as follows: n.8 sera from persons showing clinical symptoms after repeated exposure to processionary larvae, n.5 sera of persons with occasional exposure to processionary larvae, and n.3 sera from atopic patients without any symptom related to larvae exposure. A total of 3µl of supernatant were put in a nitrocellulose membrane and the non-specific sites were saturated by soaking in 5% BSA and washed in TBS-T for 1 hour at room temperature . Two washing steps, each of 5 minutes, with Tris 10 mM NaCl 140 mM pH 7.4 0.05% Tween (TBS-T) were carried out. Then the sera (primary antibody) were incubated for 2 hours, washed 4 times with TBS-T, incubated with secondary antibody anti-IgE diluted (conjugated to HPR S A-9667) 1:10,000 for 1 hour at room temperature, washed 3 times with TBS-T and visualized with Bio-Rad.

Results

Sodium dodecylsulfate-polyacrilamyde gel electrophoresis (SDS-PAGE) of proteins

The extraction with sterile water (first method) revealed the presence of a clear band at 75 kDa and of several weak bands at different molecular weight as shown in Fig. 1. The SDS-PAGE of the Hepes method also revealed the presence of many proteins. A clear band was detected at high molecular weight, ~100 kDa and others between 8 kDa and 12 kDa. The clear bands and the smear obtained from both extraction methods were excised and successively digested.

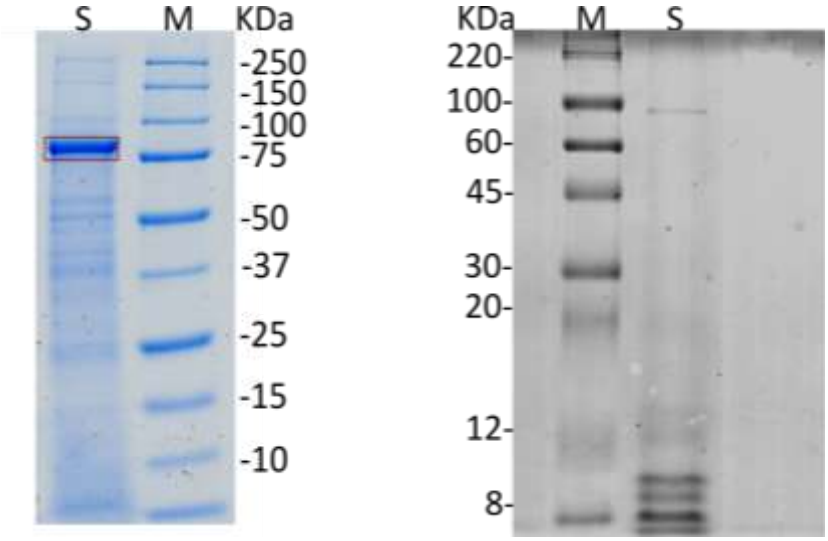


Figure 1. SDS-PAGE of protein extraction. The left gel represents the water extraction while the right one the Hepes method. Numbers refers to the molecular weight and the letters to the marker (M) and supernatant (S).

The MS-based identification of proteins

The results of the LC-MS/MS analyses against the *Th. pityocampa* transcriptome identified 182 proteins and 92 different protein families. The urticating protein Tha p 2 or parts of it were identified in all the bands of the Hepes method. Notably, four peptides in each gel bands and 5 in only one band, coded for Tha p 2, that mapped a total of 86 amino acids of the protein (Fig. 2). All 5 peptides were coded in a band of 8 kDa. In addition, another protein similar to Tha p 2, with a sequence coverage equal to 84% in BLASTp, was also identified in all gel bands. In total, only two proteins showed a higher abundance, being present in all the bands (Tha p 2, and the protein highly similar to Tha p 2), while the Tha p 1 protein was identified only in two gel bands localized at ~12 kDa. Only 115 protein were indentify for one time in the total running gel, those that remain

appears more than one time. The full list of the identified proteins by BLASTp and their putative biological functions identified by UniProt are provided in the Annex Table at the end of the chapter. Only one protein did not match in BLAST database (see the Annex Table at the end of the chapter). Concerning the protein families distinguished, those with a frequency protein in the gel, higher than 10 times were selected (Fig. 3): hypothetical proteins (30), urticating proteins Tha p (26), arylphorin proteins (17), histone proteins (15) secreted proteins (15), aldo-keto reductase proteins (14), uncharacterized proteins (10), glycine rich proteins (10), chemosensory proteins (10). The other proteins exhibited a frequency lower than 10.

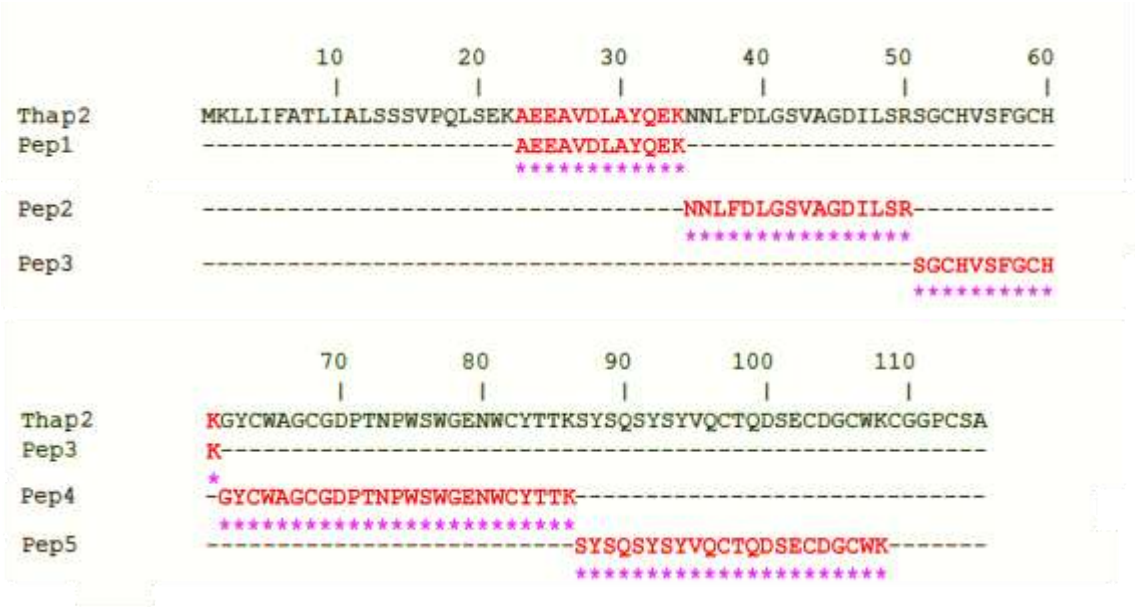


Figure 2. Alignment of the sequence of Tha p 2 and peptides obtained by LC-MS/MS. The Pep 4 was identify only one time.

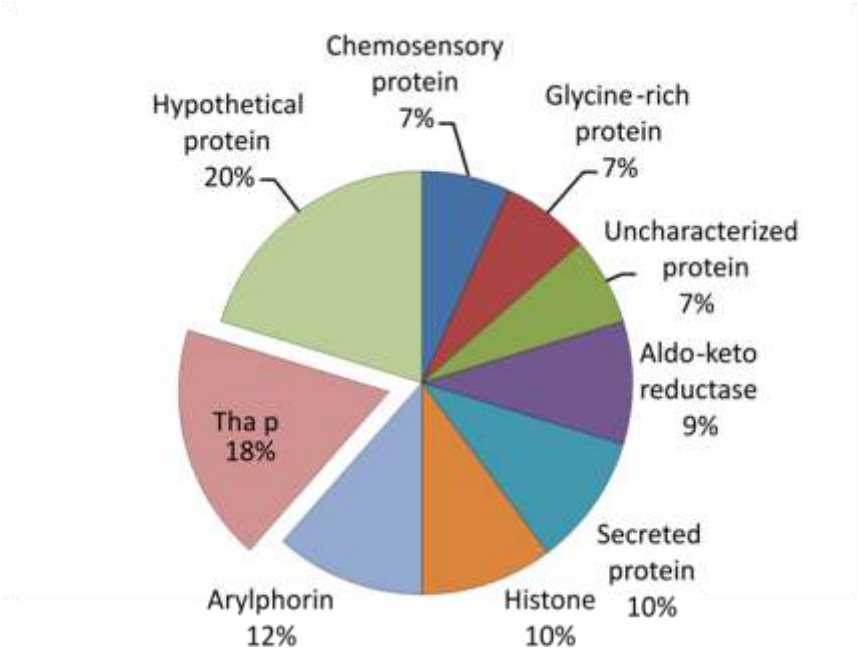


Figure 3. Graphic representation of the protein families. Only the protein families with a frequency value ≥ 10 were used.

Protein recognition by human sera

The recognition of the proteins obtained from the first extraction method by human sera is shown in Fig. 4. At least two bands reacted with the pool of the 5 allergic sera tested. The molecular weight of these bands was 75 kDa and 45 kDa. The recognition of the proteins from the Hepes extraction is shown by the dot blot test in Fig. 4. The positive response of the sera was determined by the visualization of a black dot at the centre of the nitrocellulose paper. The test identified 10 persons with strong response, 3 with a medium level of response, and 3 without recognition. All the persons with a frequent exposure and symptoms answered positively to the antigen, and also three out of five occasionally exposed persons, although without symptoms, had a positive response. Only one person (serum number 13) had a positive recognition without any declared symptom and exposure.

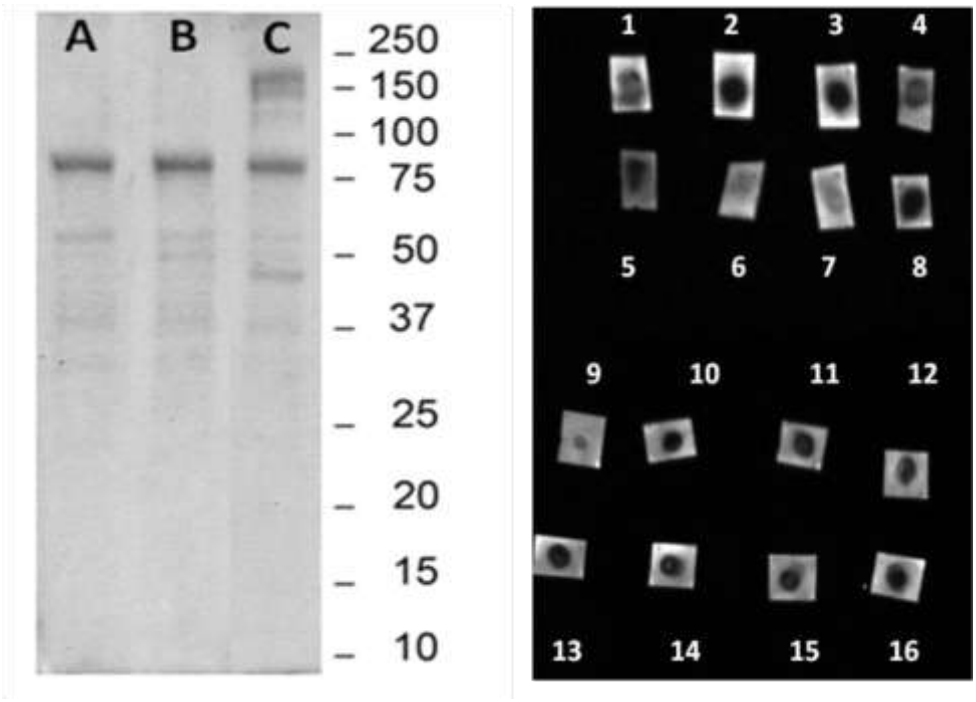


Figure 4. Western blot (left) and dot blot (right).Western blot anti-IgE of the a pool of sera of 5 persons previously exposed to setae, obtained with proteins from the water extraction method. Dot blot anti-IgE of the sera of 16 persons obtained with proteins from the Hepes extraction method. Persons 2, 3, 8, 10-16: positive recognition; persons 1, 4, 5: low recognition; persons 6, 7, 9: no recognition.

Tabale 1. Response of the sera persons of dot blot. Were indicate the number identity of persons, the response P (positive) N (negative), the presence (yes) or absence (no) of symptoms and frequency of exposition during the life.

Persons	Response	Symptoms	Type of Exposure
1	P (low recognition)	yes	occasional
2	P	yes	frequent
3	P	yes	frequent
4	P (low recognition)	no	occasional
5	P (low recognition)	yes	occasional
6	N	no	absent
7	N	yes	frequent
8	P	yes	frequent
9	N	no	absent
10	P	yes	frequent
11	P	yes	frequent
12	P	no	occasional
13	P	no	absent
14	P	yes	frequent
15	P	no	occasional
16	P	yes	frequent

Discussion

In this paper we confirm that the urticating setae of *Th. pityocampa* contain proteins, some of which are recognized by Ig-E of forestry workers presenting intense skin reactions when exposed in the infested area of the larvae of this insect. In addition, we contribute new information about the quality and quantity of the proteins associated with the setae, thanks to the availability of a transcriptome protein database of *Th. pityocampa* provided by Centre de Biologie et Gestion de Populations (INRA Montpellier, France).

The extraction of the setae with the variant of the method proposed by Olivieri et al. (2013) allowed the precipitation of a large amount of protein material. Since the 37% of the urticating setae are made by chitin (unpublished data), this brings us to think that most of the rest is mainly proteins. A technique that permits a good extraction of all proteins, however, is still missing.

Extraction methods

In the first study on the setae (Lamy et al. 1986), proteins were extracted from setae and from cuticles with the same method, consisting of an immersion in liquid nitrogen followed by crushing in a saline solution. A total of 16 proteins were detected in the setae extract while 25 proteins were detected in the cuticle. One protein of 28 kDA, composed by two sub-units of 13 and 15 kDa, was found only in the setae and was recognized as an antigen called thaumetopoein. Later, Moneo et al. (2003) focused only on whole larva extract, which was obtained after maceration in a saline solution, agitation for 24 hours and precipitation of the supernatant in ethanol. With this method they were able to obtain a number of proteins among which one of 15 kDa which was named Tha p 1 because of its allergenic power. Fuentes Aparicio et al. (2004) also started from the mature larvae, which were ground in liquid nitrogen and then extracted by magnetic stirring in agitation in phosphate-buffered saline (PBS). This method permitted to detect proteins of molecular mass ranging between 14 and 107 KDa, with several of them being recognized as antigens. They also found that some bands appeared only in absence of β-mercaptoethanol (non reducing conditions). Two years later, Fuentes Aparicio et al. (2006) prepared same the extract with the mature larvae and confirmed the previous results.

Rodriguez-Mahillo et al. (2012) compared the crude larval extract obtained by Moneo et al. (2003) with a setae extract, resulting from mixing the setae in a phosphate-buffered saline (PBS) followed by sonication on ice. The method allowed the discovery

of about 70 proteins from the setae extract, 7 of which were recognized by sera of exposed persons. One of them was found to be a major allergen different from Tha p 1 and was then called Tha p 2.

These considerations and our results suggest that a large amount of protein can be obtained from the setae when a preliminary step in liquid nitrogen is followed by crushing in saline solution or sonication, indicating that setae need to be broken in order to release an additional amount of proteins. Harsher extraction conditions, such as the use of high amounts of detergents rather than just sterile water, released further amounts of proteins from the setae. The Hepes extraction method allowed to detect the presence of a high number of proteins (181) not identified as allergens. This may suggest that during their growth, the urticating setae may become a sink for proteins occurring in the cytoplasm of the forming epidermal cells and perhaps also for blood proteins, which can easily enter the epidermis (Chapman 1988)). The sink function can be explained by the huge number of setae which have to be formed in a short time during each molt (up to 1,000,000 in a 5^{th} instar larva) (Petrucco Toffolo et al. 2014) and the urgent need for the chitin fibers to be embedded with proteins (Andersen et al. 1994), to provide the setae with a rigid structure that is functional to skin penetration (Battisti et al. 2011).

The presence of Tha p 2 or its parts in all bands obtained with the Hepes extraction method and its overall frequency in the protein profile (18%) confirms the reliability of the extraction method and leads to think that there are 86 amino-acids divided in peptides of different length that remain embedded in other proteins of different molecular weight. The amino-acids probably process the urticating element or characterize the component sequence of a Tha p protein family with a common evolutionary origin from a hypothetical ancestral gene (Andersen et al. 1994). The presence of Tha p protein family could also be hypothesized based on the detection of a protein with a sequence similar to Tha p 2, although the allergenic role of this second protein needs to be demonstrated.

Sera response

Overall we can conclude that the positive dot blot of the sera of persons previously in contact with urticating setae, although at different level of exposure, prove the occurrence of allergenic protein components of the setae extract. The positive response of the persons without any symptom but with exposure could be associated to the deficiency of their immune system, while the case of the person (serum number 13) who responded

without previous record of exposure should be further explored for confirmation with a more detailed analysis. The positive response of sera of exposed persons to a protein of about 75 kDa and 45 kDa, obtained with the first extraction method (water) could be explained either by the milder nature of the extraction or by the specificity of the insect/human population tested. The sera of sensitized persons used by Rodriguez-Mahillo et al. (2012) came from Spain and recognized especially an allergen of 14 kDa obtained from the setae of an Italian population of processionary moth, close to the one used in this work. Those sera in any case detected also several high molecular weight proteins, and one of them could correspond to the 75 kDa or 45 kDa proteins detected in our work where Italian persons were tested. It seems thus that the detection of proteins at different molecular weight as allergens may depend on the strength of the extraction method more than to the origin of either the insects or the humans, although more work has to be done in this regard.

Concluding remarks

In addition, the detection of fragments of Tha p 2 in proteins of different molecular weight could open a new way to understand the immunologic component in the complex mechanism of reaction to setae in humans. In this perspective, also the chitin component is a recognition element for tissue infiltration by innate cells implicated in allergic and immunity (Reese et al. 2007). The chitin action is not clear yet, although it can stimulate alone in vitro human T-lymphocyte proliferation (Holm et al. 2014). A much stronger proliferation, however, is induced by setae of *Th. pinivora* in persons previously exposed to the setae, indicating that setae contain molecules which may start cell-mediated immune response (Holm et al. 2014). A previous study (Rodriguez-Mahillo et al. 2012) demonstrated that the specific IgG was not found in the majority of sensitized persons and suggested that processionary larvae induce a predominantly IgE-mediated response in humans. This fact led us to believe that after skin penetration, allergens could be delivered to the immune system in a dual way: a fast release of allergens present on the outside of the setae and a slow or very slow release that must occur after degradation of the setae by chitinases (Moneo et al. 2014). In conclusion, the setae are considered a source of allergens and the risk for humans and animals is very high; they constitute a serious hazard, but the components, the quantity, the function and the real urticating protein family or components must be further investigated.

In the future, we aim to extend the extraction of the urticating setae to other species of processionary moths and try to determine if the protein extracted from urticating setae are the same. We aim also at repeating the immunologic tests with other persons who were exposed or not to the larvae to better understand the specificity of the response to the urticating setae. Finally, we will also look at the occurrence of similar types of responses in the animals which are considered the natural target of the setae, like the insectivorous birds.

Acknowledgments

We warmly acknowledge all the people that have provided help for this work and especially Franck Dorkeld, Bernhard Gschloessl, Carole Kerdelhué for allowing us to work at the transcriptome of *Th. pityocampa* and to Simone Vincenzi of Padova University for the analysis of the chitin component of the setae. This work was achieved thanks to a grant of the University of Padova.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman D.J., (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Andersen S.O., Hojrup P., Roepstorff P., (1994). Insect cuticular proteins. *Insect Biochemical* and *Molecular Biology* 25: 153-176.

Bairoch A., Apweiler R., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovar C., Redaschi N., Yeh L.,S.,L., (2005). The Universal Protein Resource (UniProt). *Nucleic Acids Research.;33(Database Issue)*: D154-D159.

Chapman R.F., (1998). *The Insects. Structure and Function*. Cambridge, UK: Cambridge Univ. Press pp.770.

Fuentes Aparicio V., de Barrio Fernàndez M., Rubio Sotés M., Rodrìguez Pardes A., Martinez Molero M.I., Zapatero Remon L., Bartolomé Zavala B., 2004. Non-occupational allergy caused by the Pine Processionary Caterpillars (*Thaumetopoea pityocampa*). *Allergol et Immunopathol* 32: 69-75.

Fuentes Aparicio V., Ramon Zapatero L., Martinez Molero M.I., Alonso Lebreros E.A., Baitia Mazuecos J.M., Zavala B.B., 2006. Allergy to pine processionary

caterpillars (*Thauetopoea pityocampa)* in children. Allergol et Immunopathol 34: 59-63.

Holm G., Andersson M., Ekberg M., Fagrell B., Sjöberg J., Bottai M., Björkholm M., (2014). Setae from Larvae of the Northern Processionary Moth (*Thaumetopoea pinivora, TP*) Stimulate Proliferation of Human Blood Lymphocytes In Vitro. *PLoS ONE* 9(12): e113977.

Köcher T., Pichler P., Schutzbier M., Stingl C., Kaul A., Teucher N., Hasenfuss G., Penninger J.M., Mechtler K., (2009). High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a mew spectrometric method combining the benefits of all. *Journal of Proteome Research* 8: 4743-4752.

Kruger N. J., (1994). The Bradford method for protein quantitation. *Methods in Molecular Biology* 32: 9–15.

Lamy M., Pastureaud M.H., Ducombs G., (1985). Thaumetopoein, an urticating protein of the processionary setae of the caterpillar (*Thaumetopoea pityocampa* Schiff) (Lepidoptera, Thaumetopoeidae). *Comptes Rendus l'Académie des Sciences, Series III*, 301: 173–176.

Lamy M., Pastureaud M.H., Novak, F., Ducombs G., Vincendeau P., Maleville J., Texier L., (1986). Thaumetopoein: an urticating protein from the setae and integument of the pine processionary caterpillar (*Thaumetopoea pityocampa* Schiff., Lepidoptera, Thaumetopoeidae). *Toxicon* 24: 347–356.

Lamy M., Novak F., Duboscq M.F., Ducombs G., Maleville J., (1988). The oak Processionary caterpillars (*Thaumetopoea processionea* L) and man – Urticating apparatus and mechanism of action. *Annales de Dermatologique et de Venereologique*, 115: 1023-1032.

Moneo I., Vega J.M., Caballero M.L., Vega J., Alday E., (2003). Isolation and characterization of Tha p 1, a major allergen from the pine processionary caterpillar *Thaumetopoea pityocampa*. *Allergy* 58: 34–37.

Moneo I., Battisti A., Dufour B., García-Ortiz J.C., González-Munoz M., Moutou F., Paolucci P., Petrucco Toffolo E., Rivière J., Rodriguez-Mahillo A.I., Roques A., Roques L., Vega J.M., Vega J., (2014). Medical and Veterinary Impact of the Urticating Processionary Larvae. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp: 359-410.

Olivieri M., Pantera B., Bruno G., Berardi L., Labardi D., Biscardo C.A., Battisti A., Francescato E., (2013). Identification of two novel allergens from the pine processionary caterpillar Thaumetopoea pityocampa. *European Academy of Allergy and Clinical Immunology (EAACI)*. Congress 22-26 June.

Petrucco-Toffolo, E., Zovi, D., Perin, C., Paolucci, P., Roques, A., Battisti, A., Horvath, H. (2014). Size and dispersion of urticating setae in three species of processionary moths. *Integrative Zoology* 9: 320–327.

Picimbon J.F., Dietrich K., Angeli S., Scaloni A., Krieger J., Breer H., Pelosi P., (2000). Purification and molecular cloning of chemosensory protein from *Bombyx mori*. *Archives of Insect Biochemistry and Physiology* 44:120-192.

Reese T. A., Liang H. E., Tager A., M Luster A. D., Van Rooijen N., Voehringer D., Locksley R. M., (2007). Chitin induces tissue accumulation of innate immune cells associated with allergy. *Nature.*447: 92-96.

Rodríguez-Mahillo A.I., González-Muñoz M., Vega J.M., López J.A., Yart A., Kerdelhué C., Camafeita E., Garcia Ortiz J.C., Vogel H., Toffolo E.P., Zovi D., Battisti A., Roques A., Moneo I., (2012). Setae from the pine processionary moth (*Thaumetopoea pityocampa*) contain several relevant allergens. *Contact Dermatitis*. 67: 367–374.

Rodríguez-Mahillo A. I., González-Muñoz M., Vega J.M., Roques A. & Moneo I., (2014). Human immune responses to *Thaumetopoea pityocampa*. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp. 367-379.

Santos-Magadàn S., Gonzàlez de Olano D., Bartolomè-Zavala B., Trujillo-Trujillo M., Melèndez-Baltanàs A., Gonzàlez Mancebo E., (2009). Adverse reactions to the processionary caterpillar: irritant or allergic mechanism. *Contact Dermatitis* 60: 109-110.

Tolin S., Arrigoni G., Moscatiello R., Masi A., Navazio L., Sablok G., Squartini A., 2013. Quantitative analysis of the naringenin-inducible proteome in *Rhizobium leguminosarum* by isobaric tagging and mass spectrometry. *Proteomics*, 12: 1961-1972.

Vega J.M., Moneo I., Armentia A., Fernàndez A., Vega J., De la Fuente R., Sanchez P., Sanchis M.E., (1999). Allergy to the pine processionary caterpillars (*Thaumetopoea pityocampa*). *Clinical and Experimental Allergy* 29: 1418-1423.

Werno J., Lamy M., Vincendeau P., (1993). Caterpillars hairs as allergens. *Lancet* 342:

Annex Tabele. Full list of the proteins identified by BLASTp and their putative biological functions identified by UniProt.

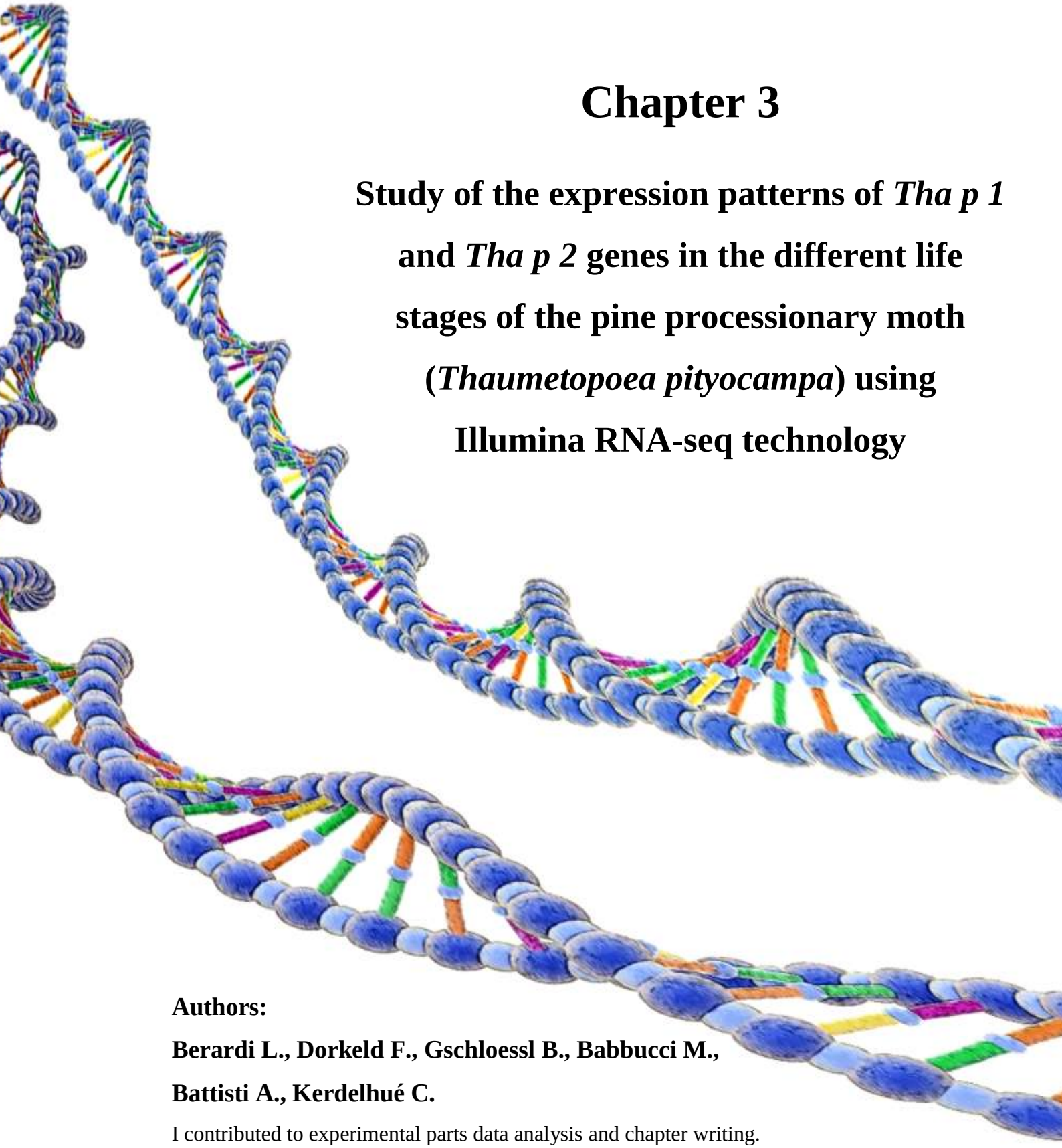
CDS identity	Homology	Species homology	% Coverage	N. peptide
cds.c247563_g2_i3	Allergen Tha p 2	*Thaumetopoea pityocampa*	100%	5
cds.c165818_g1_i1	Allergen Tha p 2	*Thaumetopoea pityocampa*	84%	5
cds.c250969_g2_i1	thioredoxin-2	*Papilio xuthus*	80%	3
cds.c252880_g4_i1	GI22343	*Drosophila mojavensis*	34%	7
cds.c128348_g1_i1	hypothetical protein 3	*Trichoplusia ni*	90%	6
cds.c222697_g1_i2	unknown secreted protein	*Papilio polytes*	85%	3
cds.c245091_g1_i1	chemosensory protein	*Danaus plexippus*	93%	5
cds.c237613_g1_i1	WAP four-disulfide core domain protein 2	*Papilio xuthus*	97%	1
cds.c22118_g1_i1	Neurobeachin	*Cerapachys biroi*	70%	1
cds.c256696_g1_i1	thioredoxin-2	*Papilio xuthus*	80%	2
cds.c244184_g2_i1	BCP inhibitor precursor	*Bombyx mori*	96%	1
cds.c246506_g2_i12	acyl-CoA binding protein 1	*Sesamia inferens*	69%	2
cds.c234296_g2_i2	hypothetical protein	*Danaus plexippus*	72%	2
cds.c247894_g2_i4	glycine-rich protein	*Bombyx mori*	65%	2
cds.c356727_g1_i1	cuticular protein glycine-rich 10 precursor	*Bombyx mori*	84%	4
cds.c230591_g3_i2	polyubiquitin-B-like isoform	*Coptotermes formosanus*	100%	2
cds.c255122_g2_i2	saposin	*Papilio polytes*	97%	2
cds.c255774_g1_i2	hypothetical protein	*Danaus plexippus*	48%	1
cds.c217974_g1_i3	unknown secreted protein	*Papilio xuthus*	97%	3
cds.c253205_g1_i3	arylphorin	*Cerura vinula*	97%	33
cds.c255429_g2_i4	arylphorin type 2	*Cerura vinula*	98%	36
cds.c255264_g5_i5	hexamerine	*Helicoverpa armigera*	100%	15
cds.c248008_g1_i1	aldo-keto reductase	*Agrotis ipsilon*	99%	15
cds.c111339_g1_i1	actin-4	*Bombyx mori*	100%	10
cds.c245915_g1_i5	catalase	*Spodoptera exigua*	100%	9
cds.c227972_g2_i1	histone H1	*Oreta rosea*	44%	3
cds.c248262_g3_i7	mitochondrial aldehyde dehydrogenase	*Danaus plexippus*	95%	3
cds.c250933_g2_i1	storage protein 1	*Plutella xylostella*	100%	6
cds.c246942_g1_i1	heat shock protein 70	*Spodoptera litura*	100%	2
cds.c146340_g1_i1	glycogen phosphorylase	*Microplitis demolitor*	100%	3
cds.c227993_g2_i1	cysteine-rich venom protein ENH1-like	*Bombyx mori*	100%	1
cds.c182227_g2_i1	DEHA2F04796p	*Debaryomyces hansenii*	100%	1
cds.c254783_g3_i8	alpha-actinin, sarcomeric-like	*Bombyx mori*	99%	3
cds.c238976_g1_i2	apolipophorin III	*Trichoplusia ni*	87%	1
cds.c255169_g1_i9	methionine-rich storage protein 2	*Manduca sexta*	100%	4
cds.c234385_g1_i1	juvenile hormone binding protein	*Bombyx mori*	95%	2
cds.c218396_g2_i1	enolase	*Spodoptera litura*	100%	9
cds.c252682_g3_i6	triosephosphate isomerase	*Helicoverpa armigera*	100%	1

CDS identity	Homology	Species homology	% Coverage	N. peptide
cds.c162537_g1_i1	peptidyl-prolyl cis-trans isomerase	*Papilio xuthus*	100%	2
cds.c324063_g1_i1	translation initiation factor	*Cerapachys biroi*	100%	1
cds.c239062_g1_i2	uncharacterized protein	*Nasonia vitripennis*	87%	1
cds.c246705_g3_i1	elongation factor	*Nasonia vitripennis*	100%	2
cds.c252804_g1_i2	GI22710	*Drosophila mojavensis*	35%	1
cds.c265904_g1_i1	sorting nexin-13-like isoform	*Nasonia vitripennis*	99%	1
cds.c237572_g1_i1	alpha-N-acetylgalactosaminidase precursor	*Bombyx mori*	100%	1
cds.c12924_g1_i1	40S ribosomal protein S3a-like	*Musca domestica*	100%	1
cds.c155109_g1_i3	Lin-9-like protein	*Harpegnathos saltator*	100%	1
cds.c255660_g2_i10	unconventional myosin-XV-like	*Bombyx mori*	88%	1
cds.c251838_g1_i5	imaginal disc growth factor-like protein	*Mamestra brassicae*	100%	1
cds.c250126_g4_i1	diapause bioclock protein	*Bombyx mori*	100%	8
cds.c257062_g1_i1	hypothetical protein	*Bombus terrestris*	97%	1
cds.c334690_g1_i1	uncharacterized protein	*Microplitis demolitor*	100%	1
cds.c162397_g1_i1	syntrophin	*Aedes aegypti*	99%	1
cds.c248008_g2_i1	aldo-keto reductase	*Agrotis ipsilon*	98%	12
cds.c249761_g5_i1	hypothetical protein	*Danaus plexippus*	99%	8
cds.c249761_g6_i2	hypothetical protein	*Danaus plexippus*	99%	8
cds.c255264_g5_i10	hexamerine	*Helicoverpa armigera*	98%	8
cds.c190487_g1_i3	heat shock protein	*Helicoverpa armigera*	100%	5
cds.c255169_g1_i3	methionine-rich storage protein 2	*Manduca sexta*	97%	6
cds.c254857_g2_i1	heat shock protein	*Spodoptera litura*	99%	4
cds.c231404_g2_i2	nucleobindin-2-like isoform	*Bombyx mori*	64%	2
cds.c252693_g6_i3	hypothetical protein	*Danaus plexippus*	99%	1
cds.c252077_g1_i7	Moesin A	*Spodoptera frugiperda*	93%	2
cds.c214759_g1_i2	coatomer protein complex subunit delta	*Bombyx mori*	100%	1
cds.c245626_g1_i2	hypothetical protein	*Lonomia obliqua*	46%	6
cds.c207263_g1_i3	TBC1 domain family member	*Bombyx mori*	98%	1
cds.c223078_g1_i1	yellow-c	*Heliconius melpomene*	99%	13
cds.c162794_g1_i1	uncharacterized protein	*Apis dorsata*	94%	1
cds.c250954_g3_i4	hypothetical protein	*Danaus plexippus*	93%	1
cds.c9578_g1_i1	protein phosphatase 1 regulatory subunit 21 isoform	*Nasonia vitripennis*	100%	1
cds.c225190_g1_i1	neurogenic protein mastermind-like	*Bombyx mori*	92%	1
cds.c243108_g5_i1	actin, clone 205-like isoformX1	*Apis mellifera*	100%	13
cds.c241261_g1_i1	unknown similar to AMEV109	*Mythimna separata*	99%	13
cds.c238074_g1_i2	imaginal disc growth factor-like protein	*Mamestra brassicae*	100%	12
cds.c215365_g1_i1	yellow-d	*Bombyx mori*	99%	11
cds.c254374_g3_i3	hemolin	*Helicoverpa zea*	99%	10
cds.c237352_g1_i3	serine proteinase-like protein precursor	*Bombyx mori*	100%	5

CDS identity	Homology	Species homology	% Coverage	N. peptide
cds.c247424_g3_i3	protein disulfide isomerase	*Papilio xuthus*	100%	5
cds.c245279_g1_i4	aldo-keto reductase, partial	*Agrotis ipsilon*	69%	5
cds.c254039_g4_i10	serine proteinase inhibitor-1A	*Mamestra brassicae*	99%	2
cds.c251564_g7_i2	allantoinase	*Danaus plexippus*	98%	2
cds.c242998_g2_i1	mitochondrial aldehyde dehydrogenase	*Danaus plexippus*	98%	1
cds.c242366_g1_i2	calreticulin	*Papilio xuthus*	98%	3
cds.c193417_g1_i3	alpha-actinin, sarcomeric-like isoform X1	*Bombyx mori*	91%	1
cds.c252167_g1_i1	glucose-6-phosphate isomerase	*Spodoptera exigua*	98%	1
cds.c244264_g3_i1	aldo-keto reductase family 1	*Bombyx mori*	96%	8
cds.c248468_g2_i4	phosphoglycerate kinase-like	*Bombyx mori*	100%	1
cds.c248232_g3_i2	chymotrypsin inhibitor	*Bombyx mori*	99%	1
cds.c252247_g1_i3	uncharacterized protein	*Nasonia vitripennis*	29%	1
cds.c305352_g1_i1	serine protease inhibitor	*Bombyx mori*	100%	1
cds.c246154_g1_i1	tetratricopeptide repeat protein	*Bombyx mori*	99%	1
cds.c250370_g3_i2	helicase	*Lactobacillus apodemi*	39%	1
cds.c235263_g1_i1	putative leucine-rich	*Bombyx mori*	79%	1
cds.c229624_g1_i4	chitin deacetylase	*Mamestra brassicae*	100%	1
cds.c243533_g8_i1	glycerophosphodiester phosphodiesterase	*Pseudomonas aeruginosa*	87%	1
cds.c215532_g1_i3	serine/threonine-protein phosphatase	*Apis mellifera*	100%	1
cds.c225502_g1_i1	myosin-I heavy chain-like	*Bombyx mori*	97%	1
cds.c244766_g3_i2	tubulin gamma-1 chain	*Spodoptera exigua*	100%	1
cds.c244668_g2_i3	serine proteinase-like protein 1	*Helicoverpa armigera*	100%	1
cds.c235717_g1_i3	SCO-spondin-like	*Bombyx mori*	97%	6
cds.c147708_g1_i1	probable GPI-anchored adhesin-like protein	*Acyrthosiphon pisum*	67%	1
cds.c239631_g5_i3	Y+L amino acid transporter 2 isoform X1	*Nasonia vitripennis*	64%	1
cds.c250390_g1_i3	probable citrate synthase 1, mitochondrial-like	*Bombyx mori*	98%	1
cds.c234385_g1_i2	juvenile hormone binding protein an-0921 precursor	*Bombyx mori*	85%	9
cds.c245445_g2_i8	unknown protein	*Helicoverpa armigera*	95%	8
cds.c239913_g1_i2	uncharacterized protein LOC101742613 isoform	*Bombyx mori*	97%	5
cds.c248232_g2_i3	chymotrypsin inhibitor CI-8A	*Bombyx mori*	99%	6
cds.c227993_g1_i3	cysteine-rich venom protein ENH1-like	*Bombyx mori*	100%	7
cds.c254599_g10_i4	uncharacterized protein	*Bombyx mori*	98%	5
cds.c243842_g1_i1	apolipophorin III	*Trichoplusia ni*	100%	9
cds.c253495_g4_i1	phosphatidylethanolamine binding protein isoform 2	*Bombyx mori*	100%	6
cds.c248159_g11_i1	glyceraldehyde-3-phosphate dehydrogenase	*Spodoptera frugiperda*	100%	3
cds.c160507_g1_i1	cuticular protein	*Danaus plexippus*	87%	3
cds.c243750_g2_i1	phosphatidylethanolamine-binding protein	*Bombyx mori*	100%	3
cds.c251040_g1_i1	heat shock protein	*Spodoptera litura*	99%	3

CDS identity	Homology	Species homology	% Coverage	N. peptide
cds.c252733_g3_i1	uncharacterized protein	*Bombyx mori*	100%	3
cds.c230658_g2_i3	multiple coagulation factor deficiency protein	*Bombyx mori*	98%	1
cds.c251831_g2_i2	vitellogenin	*Helicoverpa armigera*	100%	2
cds.c236041_g1_i5	histone H2A-like	*Nasonia vitripennis*	87%	1
cds.c245890_g4_i2	chitin binding peritrophin-A	*Papilio xuthus*	100%	2
cds.c248540_g1_i1	unknown secreted protein	*Papilio polytes*	95%	7
cds.c250475_g7_i1	ML-domain containing secreted protein-like protein	*Antheraea yamamai*	87%	2
cds.c18274_g1_i1	binding FK506- protein precursor	*Bombyx mori*	99%	1
cds.c255616_g2_i2	proline-rich receptor-like protein kinase	*Bombyx mori*	100%	1
cds.c17057_g1_i1	malate dehydrogenase, mitochondrial	*Nasonia vitripennis*	100%	1
cds.c235010_g1_i1	juvenile hormone binding protein	*Heliothis virescens*	100%	1
cds.c232852_g1_i2	dihydropteridine reductase	*Papilio xuthus*	100%	1
cds.c355329_g1_i1	triosephosphate isomerase	*Microplitis demolitor*	100%	2
cds.c247998_g1_i2	egalitarian	*Danaus plexippus*	100%	1
cds.c242810_g1_i4	protein takeout-like	*Bombyx mori*	87%	1
cds.c391427_g1_i1	peptide chain release factor (probable)	*Nasonia vitripennis*	87%	1
cds.c254029_g1_i2	serine protease	*Papilio xuthus*	100%	1
cds.c236757_g1_i2	voltage-dependent anion-selective channel-like isoform X4	*Bombyx mori*	100%	1
cds.c243797_g4_i2	juvenile hormone binding protein an-0921 precursor	*Bombyx mori*	90%	1
cds.c253640_g7_i4	nucleoplasmin isoform 2	*Danaus plexippus*	93%	1
cds.c247161_g1_i2	hemolymph proteinase 8	*Manduca sexta*	81%	1
cds.c262202_g1_i1	interference hedgehog-like	*Bombus terrestris*	97%	1
cds.c245014_g3_i1	yellow-c	*Heliconius erato*	98%	1
cds.c243221_g1_i1	peptidyl-prolyl cis-trans isomerase	*Papilio polytes*	100%	5
cds.c244127_g2_i1	uncharacterized protein	*Bombyx mori*	100%	6
cds.c246112_g4_i1	superoxide dismutase	*Danaus plexippus*	100%	6
cds.c254772_g4_i5	secreted protein unknown	*Papilio xuthus*	98%	3
cds.c145193_g1_i1	abnormal wing disc protein	*Antheraea pernyi*	100%	2
cds.c255864_g3_i1	No match	No match	No match	1
cds.c245309_g1_i2	promoting protein	*Danaus plexippus*	84%	1
cds.c219735_g1_i1	muscular protein 20	*Bombyx mori*	100%	1
cds.c180932_g2_i1	hypothetical protein SINV_02932	*Solenopsis invicta*	77%	1
cds.c247482_g4_i2	Histone H2B	*Camponotus floridanus*	100%	2
cds.c232764_g1_i5	similar to CG9796	*Papilio xuthus*	100%	1
cds.c240809_g3_i7	hypothetical protein 31	*Lonomia obliqua*	72%	1
cds.c248400_g2_i1	mesencephalic astrocyte-derived neurotrophic factor homolog	*Bombyx mori*	100%	1
cds.c99479_g1_i1	unknown secreted protein	*Papilio xuthus*	91%	2
cds.c255429_g2_i1	aryphorin type 2	*Cerura vinula*	92%	1

CDS identity	Homology	Species homology	% Coverage	N. peptide
cds.c232174_g1_i1	adenylyltransferase and sulfurtransferase MOCS3 isoform X1	*Nasonia vitripennis*	100%	1
cds.c254703_g1_i1	uncharacterized protein	*Bombyx mori*	58%	1
cds.c250034_g2_i1	Tha p 1 allergen	*Thaumetopoea pityocampa*	100%	3
cds.c247952_g2_i1	histone H2A-like protein 2	*Bombyx mori*	100%	2
cds.c241205_g3_i2	odorant-binding protein 4	*Chilo suppressalis*	96%	1
cds.c249497_g1_i4	lysozyme	*Helicoverpa armigera*	82%	1
cds.c230317_g1_i1	60S acidic ribosomal protein P2-like	*Bombus impatiens*	100%	2
cds.c225627_g1_i1	uncharacterized protein LOC101739120	*Bombyx mori*	82%	1
cds.c254766_g5_i2	actin-2, partial	*Wuchereria bancrofti*	100%	1
cds.c237613_g2_i3	WAP four-disulfide core domain protein 2 precursor	*Papilio xuthus*	97%	4
cds.c252458_g5_i2	chemosensory protein 7 precursor	*Bombyx mori*	98%	3
cds.c245136_g2_i2	odorant binding protein 6	*Spodoptera exigua*	100%	1
cds.c165775_g1_i1	Histone H4	*Macaca mulatta*	85%	1
cds.c243980_g12_i4	hypothetical protein	*Heliothis virescens ascovirus*	92%	1
cds.c109578_g1_i1	hypothetical protein	*Tetrapisispora phaffii*	70%	1
cds.c252383_g1_i2	slowmo	*Bombyx mori*	92%	1
cds.c234087_g1_i3	protocadherin-16-like PREDICTED:	*Bombyx mori*	99%	1
cds.c248008_g1_i2	aldo-keto reductase, partial	*Agrotis ipsilon*	91%	1
cds.c231482_g1_i1	cytochrome c oxidase subunit 4 isoform 1, mitochondrial	*Nasonia vitripennis*	95%	1
cds.c185079_g1_i1	REPAT32	*Spodoptera littoralis*	92%	2
cds.c223924_g1_i1	cysteine proteinase inhibitor precursor	*Manduca sexta*	99%	1
cds.c253993_g4_i1	chemosensory protein 3 precursor	*Bombyx mori*	100%	1
cds.c181512_g1_i1	histone H2A-like	*Diaphorina citri*	95%	1
cds.c123959_g1_i1	pterin-4-alpha-carbinolamine dehydratase-like	*Bombyx mori*	91%	1
cds.c249531_g1_i1	chemosensory proteins	*Dendrolimus kikuchii*	100%	1
cds.c243108_g8_i1	actin	*Onthophagus nigriventris*	99%	1
cds.c209834_g1_i1	hypothetical protein	*Danaus plexippus*	98%	2
cds.c358177_g1_i1	hypothetical protein	*Vittaforma corneae*	100%	1



Chapter 3

Study of the expression patterns of *Tha p 1* and *Tha p 2* genes in the different life stages of the pine processionary moth (*Thaumetopoea pityocampa*) using Illumina RNA-seq technology

Authors:

Berardi L., Dorkeld F., Gschloessl B., Babbucci M., Battisti A., Kerdelhué C.

I contributed to experimental parts data analysis and chapter writing.

Paper in preparation for submission.

Abstract

Larvae of the pine processionary moth, *Thaumetopoea pityocampa*, produce urticating setae from the third to the fifth larval instars. By penetrating the skin, they can cause inflammatory symptoms in man and other vertebrates. These setae convey an urticating component previously identified as a protein named Tha p 2. In the present study, the expression profiles of the gene coding for this urticating protein Tha p 2 and the gene corresponding to a chemosensory protein previously considered as allergenic, Tha p 1, were investigated throughout the insect development. All life stages of two Italian populations and the last development stages and eggs of two Portuguese populations were considered, in order to better understand the expression patterns of both genes in the different development stages. An NGS approach (Illumina HiSeq sequencing) was used for sequencing the cDNA obtained for each stage. The obtained reads were mapped against the reference sequences of both genes using two different bioinformatics methods. We confirmed the expression of *Tha p 1* gene in all life stages of *Th. pityocampa* for both populations, indicating that it is not strictly associated with the urticating setae, while the gene *Tha p 2* was expressed only in larval stages where the setae production occur.

Introduction

All studied species of the Thaumetopoeinae are known to carry urticating setae, either as a larva (genus *Thaumetopoea*) or as an adult, which uses them to protect eggs and larvae of the next generation (e.g. *Anaphe* from Africa and *Ochrogaster* from Australia) (Lamy et al. 1984, Floater 1998).

Urticating setae provide an efficient defense system for the colony but not for the individual, as the symptoms appear with a delay of time, when the larva has already been killed (Battisti et al. 2011). As setae disperse as a cloud around the colony (Fenk et al. 2007), with high concentration of short setae up to 6 km during the day and 12 km during the night, their function could be to keep away predators (Moneo et al. 2014). Incidentally, these setae are a serious threat to human health and the reactions are common in pine wood area workers, who are exposed to high levels of setae, but also in persons non-occupationally exposed to processionary larvae, because the setae are dispersed by the wind over long distances (Fenk et al. 2007, Petrucco Toffolo et al. 2014). Contact dermatitis and urticaria are the most frequent symptoms related to the exposure to this insect (Vega et al. 1999).

Each larva bear integument areas in the abdominal tergites, called mirrors, that enclose the setae and are actively open when the larva is disturbed (Démolin 1963). The setae are short and thin (100-500 µm long and 3-7 µm in diameter) and like the insect integument are built up by chitinUrticating setae are absent from the early larval stages of *Thaumetopoea pityocampa*, although the cellular apparatus that produces them is present. Urticating setae appear in the third stage and their number increases after each molt (Lamy 1990). The urticating setae found in L3 to L5 in larvae are not present in moths, which possess only scales which vary in size and form (Lamy 1986). The urticating reaction in human and animals is correlated to specific allergens. Until now only tree proteins were identified in *Th. pityocampa*, namely thaumetopoein, Tha p 1 (Moneo et al. 2003) and Tha p 2 (Rodriguez-Mahillo et al. 2012). The first protein, named thaumetopoein, is supposed to be composed of two subunits having respectively a molecular weight of 13 and 15 kDa (Lamy et al. 1986). Unfortunately the thaumetopoein was not sequenced after its discovery, and the corresponding gene is unknown. A second protein named Tha p 1 with an estimated molecular weight of 15 kDa was later isolated by Moneo et al. (2003) who published a polypeptide encompassing 18 residues and deposited the full length sequence in GenBank (accession number: HE962022) Three years later, Larsson and Backlund (2009) showed that Tha p 1 polypeptide was related to chemosensory proteins similar to one found in *Bombyx mori*. Finally, a third protein named Tha p 2, unrelated to Tha p 1, was identified and the cDNA encoding the complete polypeptide was amplified and sequenced (Rodriguez-Mahillo et al. 2012). The sensitizing capacity of moth allergens is clearly demonstrated with the help of epidemiological studies.

The presence or absence of the protein, and the expression levels of the corresponding *Tha p 1* and *Tha p 2* genes in the different life stages (including the instars without urticating setae), is still unknown. An NGS approach (Hiseq Illumina sequencing of mRNAs) was used to sequence the expressed genes for the different development stages. RNA-seq is a recently developed large-scale genome-wide approach that has been applied successfully to gene discovery and expression profiling, and to the study of functional, comparative and evolutionary genomics in non-model organisms for which limited previous information existed (Lu et al. 2013). In this study, RNA-seq data were analyzed from every life stages to screen the patterns of expression of the two genes of interest mentioned above. In particular, we wanted to determine whether these genes were

expressed only during the larval stages bearing urticating setae, or if they were also detected in eggs, early larval stages, pupae and moths. In this work we used four different populations of *Th. pityocampa*, two from Italy characterized by early and late adult emergence in the field, and two from Portugal characterized by contrasted life cycles (Santos et al. 2007, 2011).

Material and methods

Sampling

Four *Th. pityocampa* populations were sampled to develop Illumina sequencing of cDNAs. We chose two populations from the Italian Alps characterized by early and late adult emergence as well two populations from Portugal characterized by summer and winter feeding of the larvae. The Italian populations were also used for a development study of *Th. pityocampa* in laboratory conditions (see Supplementary chapters below). The eggs batches from the Italian early population were collected on the needles of *Pinus nigra* in the south eastern Alps (Cimolais, 12°27'E 46°19'N) (n = 12) in July 2012, while those of the late population were collected on the needles of the same pine species in the last week of August 2012 in the southern Prealps (Cinto Euganeo, 11°39'E 45°16'N and Tregnago, 11°09'E 45°30'N). The Italian "early" and "late" larvae were reared in separate boxes under laboratory conditions. Some individuals were put in RNAlater or liquid nitrogen after each molt (see the Supporting Chapter1 below).The pupae were collected few days after pupations started ("early pupae"), then by the and of the pupation stage ("late pupae") and the adults after the emergence.

For the Portugal populations, the samples corresponded to eggs, L3, and all stages from L5 to adults. All samples of summer and winter populations were collected from the maritime pine *Pinus pinaster* in the National Forest of Leiria (Portugal) (39º47' N 8º 58' W). For the winter population eggs, L3 and L5 were sampled from the field. A part of L5 larvae were immediately frozen at -80 °C, while the others were kept in the laboratory at the Institute Superior de Agronomia of Lisbon until pupation and adult emergence. Similarly to the sampling of Italian populations, samples were taken and put in liquid nitrogen few days after pupations started ("early pupae"), then by the end of the pupation stage ("late pupae") and immediately after adult emergence. Concerning summer population, eggs, L3, L5 and early pupae were sampled in the field, and a part of pupae

were kept under laboratory condition until adult emergence (to collect "late pupae" and adults). All sample were put in liquid nitrogen and deep frozen at -80 C° immediately after each collection. Samples were sent in dry ice to the Centre de Biologie pour la Gestion des Populations (CBGP, INRA Montpellier, France) to proceed with RNA extraction.

RNA extraction

Total RNA was isolated from a pool of five samples for each stage, including males and females. RNA was isolated using TRIzol Reagent (Invitrogen) according to the manufacturer's protocol. The RNA was separated with chloroform, precipitated with isopropanol at room temperature and washed with ethanol 75% RNAse and DNAse free. The RNA extracted was resuspended in 50 µl of ultrapure water. RNA quantity and quality were estimated using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and gel electrophoresis. Equal amounts of total RNA were subsequently pooled for each stage and each population, and the RNA samples were sent in dry ice to the Montpellier GenomiX (MGX) NGS platform, c/o Institut de Génomique Fonctionnelle (Montpellier, France) for the sequencing and the preliminary bioinformatics treatments.

Construction and sequencing of RNA libraries

Library construction and Illumina sequencing was done at the MGX platform. The TruSeq Stranded Sample mRNA Sample Preparation kit was used to synthesize a strand of cDNA and construct the libraries. Each stage and population sample was barcoded using a different Illumina index. Quality control of library preparation was done using Agilent DNA 1000 kit as well as qPCR. The librairies were sequenced on 4 lanes of an Illumina Hi-Seq 2000, using the SBS (Sequence by Synthesis) technique and 100 pb, paired-end sequencing. The Next-generation sequencing technologies are being exploited not only to analyze static genomes, but also transcriptomes with rapidly evolving technologies (Marguerat and Bähler 2010).

Bioinformatics analyses, mapping and differential expression analysis

The image of the analyses was realized with HiSeq Control Software (HCS) Illumina, that permitted to identify the position of the clusters, the intensity and the noise. The quality

of the sequencing results was checked by FastQC. All reads that successfully passed the quality controls were mapped against the reference sequences of genes *Tha p 1* (accession number: HE962022) and *Tha p 2* (accession number of the protein: P86360.1) of *Thaumetopoea pityocampa* with Tophat2 (v. 2.0.12) with a default parameters and the BAM file obtained, were visualize with IGV genome browser. After running Tophat2, the resulting alignments files are provided to Cufflinks (v.2.2.1) for assemble the mapped reads in order to retrieve the exon-intron structure of the genes (Trapnell et al. 2012). These assemblies are then merged together, using the Cuffmerge utility, which is included with the Cufflinks package. Then, we merged the GTF files of each cufflinks assembly to one common GTF annotation; this merged assembly provides a uniform basis for calculating and compare genes expression in each stage. The reads and the merged assembly are fed to Cuffdiff, which calculated expression levels and tests the statistical significance of observed changes between two conditions (developmental stages). To be able to compare gene expression for two conditions, Cuffdiff normalizes read counts of several runs (conditions and replicates) for a specific transcript. This normalization is done by calculating the FPKM (fragments per kilobase of transcript per million mapped fragments) value per condition. Hence FPKM allows normalization of read counts by transcript length. FPKM calculation is applied for comparison of two conditions and also for comparing replicates of the same condition. These results are reported as a set of text files and can be displayed in the plotting environment with CummeRbund. The analyses were also validate with CLC Bio Workbench 7 (default parameters) where the reads were mapped against the sequences of genes *Tha p 1* and *Tha p 2*. The Total count (TC) normalization method (Dillies et al. 2012) was applied only to the Tha p 1 (Tables 1-2) because of the number of reads mapping with Tha p 2 was too low.

Results

The RNA was successfully extracted from all the samples except from "late pupae" of the Italian late population (LP). Twenty-seven types of samples, which represented different developmental stages of *Th. Pityocampa,* were sequenced. Approximately 80 million Illumina sequence reads originating from each sample were mapped against the two studied genes. Both approach methods used gave the same results. As regards the Italian populations, the results demonstrate that all stages of each population express *Tha p 1* gene; the numbers of mapped reads were highest in the adult stage for both

populations. It was also high in L5 and late pupae in the early population (Tab.1). Conversely, the *Tha p 2* gene was only expressed in the L4 sample. The values obtained for L3 and late pupae in the early population, and for adults in the late population, were too low to be considered. More precisely, in the early population, 128 reads were mapped on *Tha p 2* for L4 (over a total of 26026052 reads), while in the L3, late pupae and adults only 4, 5 and 21 reads were mapped, respectively. No expression of *Tha p 2* was found in the L1, L2 and early pupal stages. In the case of late population, the L4 stage mapped 1707 reads over a total of 63776998, moth 12 and L2 only 3 reads; the L1, L3, L5 and pre-molt did not express the gene. In both populations *Tha p 2* is not expressed in the L3 and L5 stages. As regards the Portuguese populations, as already said, *Tha p 1* is expressed in all the stages, with a higher number of mapping reads in early pupae and moth in the summer population, but the value remain high also in the other stages, like for winter population where the higher value of mapped reads were visualized on L5 and early pupae (Table 2). *Tha p 2* show a high expression level in L3 and especially in L5 stages of both Portuguese population. In the summer population, L5 mapped 2709 over a total of 81320594 reads and in the winter population 6017 on a total of 82234506. The gene is not expressed in the eggs. The value of other stages are too low to be significant.

Table 1. Mapped reads for *Tha p 1* and *Tha p 2* genes, Total Reads for each stage and Total Count (TC) only for the gene *Tha p 1*, in the early population (EP) and late population (LP),

EP				
	Total reads for stage	**Mapped reads Tha p 1**	**TC Tha p 1**	**Mapped reads Tha p 2**
L1	$73x10^6$	$10x10^3$	7.00	0
L2	$77\ x10^6$	$8x10^3$	5.47	0
L3	$77\ x10^6$	$9x10^3$	5.63	4
L4	$26\ x10^6$	$4x10^3$	8.58	128
L5	$93\ x10^6$	$76\ x10^3$	3.90	0
early pupae	$80\ x10^6$	$3\ x10^3$	2.02	0
late pupae	$103x10^6$	$52\ x10^3$	2.44	5
moth	$64\ x10^6$	$215\ x10^3$	1.60	21
LP				
L1	$96\ x10^6$	$12\ x10^3$	8.44	4
L2	$5\ x10^6$	$4\ x10^3$	5.24	3
L3	$46\ x10^6$	$4\ x10^3$	6.78	1
L4	$63\ x10^6$	$21\ x10^3$	2.22	1707
L5	$55\ x10^6$	$30\ x10^3$	3.60	1
early pupae	$105\ x10^6$	$9\ x10^3$	5.84	1
late pupae	missing	missing	missing	missing
moth	$77\ x10^6$	$294\ x10^3$	2.46	12

Table 2 Mapped reads for *Tha p 1* and *Tha p 2* genes, Total Reads for each stage and Total Count (TC) only for the gene *Tha p 1*, in the summer population (SP) and winter population (WP).

SP				
	Total reads for stage	**Mapped reads Tha p 1**	**TC Tha p 1**	**Mapped reads Tha p 2**
eggs	$67\ x10^6$	$4\ x10^3$	5.09	0
L3	$71\ x10^6$	$16\ x10^3$	1.74	250
L5	$81\ x10^6$	$23\ x10^3$	2.15	2709
early pupae	$104\ x10^6$	$166\ x10^3$	1.17	1
late pupae	$106\ x10^6$	$5\ x10^3$	4.02	4
moth	$80\ x10^6$	$38\ x10^3$	3.56	3
WP				
eggs	$73\ x10^6$	$4\ x10^3$	3.12	0
L3	$62\ x10^6$	$20\ x10^3$	1.61	162
L5	$82\ x10^6$	$39\ x10^3$	2.37	6017
early pupae	$80\ x10^6$	$182\ x10^3$	1.10	2
late pupae	$104\ x10^6$	$24\ x10^3$	1.16	4
moth	$103\ x10^6$	$22\ x10^3$	1.06	6

Discussion

The preliminary results obtained from this study demonstrate that the *Tha p 1* gene is expressed in all developmental stages of *Th. pityocampa*, yet with a higher expression level in the last stages (L5 and moth in Italian populations; L5, pupae and moth in the Portuguese populations). Concerning *Tha p 2*, the number of mapped reads is high only in the L4 stages and low in the moth in the Italian populations while it is not expressed in the other stages, including the L5 larvae, even though this stage bears urticating setae; in contrast to Italian populations, the Portuguese ones express the gene in L3 and L5 while the L4 were not tested.

The presence of *Tha p 1* gene in all stages is consistent with the identification by Larsson and Backlund (2009) as a chemosensory protein, although its role as an allergen has been demonstrated starting from a crude larval extract (Moneo et al. 2003). Based on its occurrence in all life stages of the insect, it is unlikely that it is associated exclusively with the urticating setae. The occurrence of Tha p 1 in the setae has been shown in Chapter 2, although at a very low concentration. The confirmation of the antigenic role of Tha p 1 should be carried out with setae extracts (Moneo et al. 2014). The situation is more simple for *Tha p 2* because it is largely expressed only in the larval stages when the urticating setae are produced, although not in all of them. The reasons why the expression was partly inconsistent among larval instars and populations should looked for in the timing of the extraction and the preservation of the samples. Both Italian population are characterized by a high expression only in the L4, while a high expression level was also expected in the L5 larvae. This can be due to the fact that the sampled individuals were reared under laboratory conditions in Italy, unlike the Portuguese populations that were sampled in the field and might produce urticating proteins at higher rates because of predation pressure. The low expression of *Tha p 2* gene during the moth stage can bring us to hypothesize the presence of Tha p 2 also in the scales of the anal tuft of the female moth, where it should be looked for in future work. Interestingly, other species of processionary moths which bear the urticating setae at the moth stage (e.g. the African *Anaphe* and the Australian *Ochrogaster*) present them in the anal tuft and use them to protect the eggs and the larvae hatched from them (Lamy et al. 1984, Floater 1998). Our results confirmed the absence of the urticating protein during the first larval instars, including the eggs, confirming the situation that was also found in other Thaumetopoea species (Battisti et al. 2011). This can be explained by the short duration of the first two

larval instars L1 and L2 (around 14 days each, see Supplementary Chapters), which may not require the presence of a protection system. In those instars the larvae are also smaller and not as conspicuous as in the other larval instars, when they occur in very visible tents on the most exposed parts of the trees. The long development times of the later instars L3, L4 and L5 (from one to two months each under natural conditions) expose the larvae to a high predation risk by bird like cuckoo, hoopoe and tits, which developed foraging techniques to avoid the ingestion of the setae (Barbaro and Battisti, 2011).

In the future we aim to repeat RNA extraction from L5 Italian sample for a real-time PCR approach to verify and eventually confirm the effective absence of Tha p 2. A similar approach on adults would allow to test the hypothesis that Tha p 2 is expressed in female abdominal scales and involved in egg protection. The expression patterns and levels of *Tha p 1* and *Tha p 2* genes in other *Thaumetopoea* species should also be analysed. Overall it should be noted that the results must be considered as a preliminary and need to be confirmed by further analyses.

Acknowledgements

We want greatly acknowledge the Instituto Superior de Agronomia, University of Lisbon, Portugal, for providing Portugal sample and Laure Sauné for the RNA extractions (Centre de Biologie et Gestion de Populations INRA Montpellier, France).

References

Barbaro L., Battisti A., (2011). Birds as predators of the pine processionary moth (Lepidoptera: Notodontidae). *Biological Control* 56: 107-114.

Battisti A., Holm G., Fagrell B., Larsson S., (2011). Urticating hairs in arthropods: Their nature and medical significance. *Annual Review of Entomology* 56: 203-220.

Démolin G., (1963). Les 'miroirs' urticants de la processionnaire du pin (Thaumetopoea pityocampa Schiff.). *Agricultural Zoology Reviews* 10–12: 107–14.

Dillies M.A., Rau A., Aubert J., Hennequet-Antier C., Jeanmougin M., Servant N., Keime C., Marot G., Castel D., Estelle J., Guernec G., Jagla B., Jouneau L., Laloë D., Le Gall C., Schaëffer B., Le Crom S., Guedj M., Jaffrézic F., (2012). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics* 14,6: 671-683.

Fenk L., Vogel B., Horvath H., (2007). Dispersion of the bio-aerosol produced by the Oak Processionary moth. *Aerobiologia* 23: 79-87.

Floater G.J., (1998). Tuft scales and egg protection in Ochrogaster lunifer Herrich-Schäffer (Lepidoptera: Thaumetopoeidae). *Australian Journal of Entomology* 37: 34–39.

Lamy, M., Pastureaud, M.H., Novak, F., Ducombs, G., Vincendeau, P., Maleville, J., Texier, L., (1986). Thaumetopoein: an urticating protein from the hairs and integument of the pine processionary caterpillar (Thaumetopoea pityocampa Schiff., Lepidoptera, Thaumetopoeidae). *Toxicon* 24: 347–356.

Lamy M., Novak F., Duboscq, M.F., Ducombs G., Maleville J., (1988). The oak Processionary caterpillars (Thaumetopoea processionea L) and man -Urticating apparatus and mechanism of action. *Annales de Dermatologique et de Venereologique* 115: 1023-1032.

Lamy M., (1990). Contact dermatitis (erucism) produced by processionary caterpillars (genus Thaumetopoea). *Journal Apply of Entomology* 110: 425–37.

Larsson S., Backlund A., (2009). Regarding the putative identity of moth (Thaumetopoea spp.) allergen. *Allergy* 64: 493.

Lu X., Li J., Yang J., Liu X., Ma J., (2013). De novo transcriptome of the desert beetle Microdera punctipennis (Coleoptera: Tenebrionidae) using illumina RNA-seq technology. *Molecular Biology Reports* 41, 11: 7293-7303.

Marguerat S., Bähler J., (2010). RNA-seq: from technology to biology. *Cellular and Molecular Life Science* 67: 569-579.

Moneo I., Vega J.M., Caballero M.L., Vega J., Alday E., (2003). Isolation and characterization of Tha p 1, a major allergen from the pine processionary caterpillar Thaumetopoea pityocampa. *Allergy* 58: 34–37.

Moneo I., Battisti A., Dufour B., García-Ortiz J.C., González-Munoz M., Moutou F., Paolucci P., Petrucco Toffolo E., Rivière J., Rodriguez-Mahillo A.I., Roques A., Roques L., Vega J.M., Vega J., (2014). Medical and Veterinary Impact of the Urticating Processionary Larvae. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp: 359-410.

Petrucco-Toffolo E., Zovi D., Perin C., Paolucci P., Roques A., Battisti A., Horvath H., (2014). Size and dispersion of urticating setae in three species of processionary moths. *Integrative Zoology* 9: 320–327.

Rodríguez-Mahillo A.I., González-Muñoz M., Vega J.M., López J.A., Yart A., Kerdelhué C., Camafeit E., García Ortiz J.C., Vogel H., Toffolo E.P., Zovi D., Battisti A., Roques A., Moneo I., (2012). Setae from the pine processionary moth (Thaumetopoea pityocampa) contain several relevant allergens. *Contact Dermatitis* 67: 367–374.

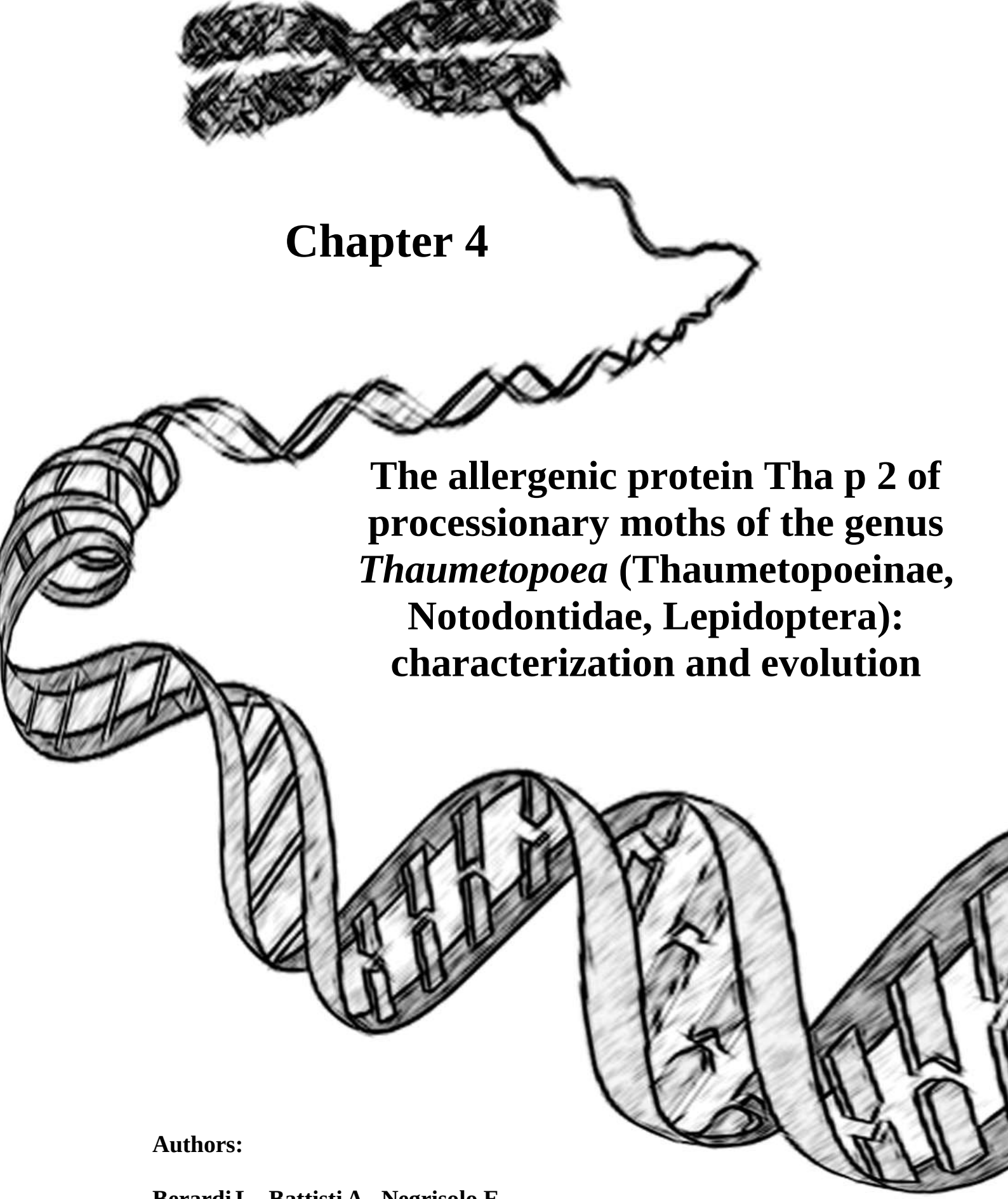
Santos H., Rousselet J., Magnoux E., Paiva M.R., Branco M., Kerdelhué C., (2007). Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society of London Series* B 274 (1612): 935-941.

Santos H., Burban C., Rousselet J., Rossi J.P., Branco M., Kerdelhué C., (2011) Incipient allochronic speciation in the pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera, Notodontidae). *Journal of Evolutionary Biology* 24 (1): 146-158.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L., Pachter L., (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 7,3: 562-578.

Vega J.M., Moneo I., Armentia A., Fernàndez A., Vega J., De la Fuente R., Sanchez P., Sanchis M.E., (1999). Allergy to the pine processionary caterpillars (*Thaumetopoea pityocampa*). *Clinical and Experimental Allergy* 29: 1418-1423.

Wang Z., Gerstein M., Snyder M., (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57–63.



Chapter 4

The allergenic protein Tha p 2 of processionary moths of the genus *Thaumetopoea* (Thaumetopoeinae, Notodontidae, Lepidoptera): characterization and evolution

Authors:

Berardi L., Battisti A., Negrisolo E.

I contributed to experimental parts data analysis and chapter writing.
Manuscript submitted.

Abstract

The allergenic Tha p 2 protein has been extracted recently from urticating setae of the pine processionary moth *Thaumetopoea pityocampa*. In the present paper we test for the occurrence of this protein in other Thaumetopoeinae, with a particular focus on members of the genus *Thaumetopoea*, as well as unrelated moth species, in order to better understand the chemical-physical properties of the protein and the nature of encoding genes, as well as their evolutionary history. Tha p 2 is encoded by the intronless gene *Tha p 2* that is restricted to the processionary moths (Thaumetopoeinae, Notodontidae, Lepidoptera). Most of the species present two isoforms of Tha p 2 that can be interpreted as the result of heterozygosity in the single gene. The only exception is represented by *Thaumetopoea wilkinsoni*, in which 20 different isoforms occur in a single specimen, leading to the conclusion that at least in the species multiple copies of *Tha p 2* exist. The Tha p 2 is a glycine and cysteine rich protein very well conserved among processionary moths. The predicted secondary structures indicate presence of 3 α-helices and six β-barrels. Finally, the evolution of the gene and more markedly of the protein is characterized by negative selection.

Introduction

The larvae of the processionary moths belonging to the genus *Thaumetopoea* (Thaumetopoeinae, Notodontidae; Lepidoptera) present a specialized urticating apparatus, occurring from the third instar onwards on the dorsal part of the abdomen, in specific areas called mirrors (Démolin, 1963). Conversely, the adults of other processionary moths genera, like the African *Anaphe* and the Australian *Ochrogaster*, produce the setae in the anal tuft of female moths (Floater, 1998; Schabel, 2006). The setae are thought to protect the insects from mammalian and avian predators, although the nature of the reactions caused by them has been thoroughly studied only in humans (Battisti et al. 2011).

The setae are short and thin (100-500 µm long, 3-7 µm in diameter), have barbs along the shaft and are modified by the loss of the neuronal connection and the detachment of the proximal end of the hair from the integument (Battisti et al. 2011; Petrucco Toffolo et al. 2014). The base of each seta is inserted into a socket and can easily be removed with any kind of mechanical stimulation (Lamy et al. 1982).

Accidental contact with urticating setae induces symptoms like cutaneous lesion or respiratory responses even anaphylactic shock in humans or tongue necrosis in dogs (Maier et al. 2003), often referred to syndromes such as erucism and lepidopterism. The use of these terms is not clear because erucism and lepidopterism refer to different life stages of the insects (larvae and adults, respectively) rather than to human reactions (Battisti et al. 2011).

Setae, like the insect integument, are build up by a chitin skeleton with a matrix of proteins. While the role of chitin is still under investigation, the presence of several proteins with allergic activity against humans has been demonstrated in *Thaumetopoea* (Lamy et al. 1985; Lamy et al. 1986; Rodriguez-Mahillo et al. 2012). Three proteins with allergenic activity have been identified for processionary moths. The first protein was named thaumetopoein and estimated to be made by two subunits having respectively a molecular weight of 13 and 15 kDa (Lamy et al. 1986). Unfortunately the thaumetopoein was not sequenced after its discovery. Successively, a second protein named Tha p 1 with an estimated molecular weight of 15 kDa was isolated by Moneo et al. (2003) who published a polypeptide encompassing 18 residues and successively deposited the full length sequence in GenBank (ADK47876). Larsson and Backlund (2009) showed that Tha p 1 polypeptide was related to chemosensory proteins found in insects. Finally, a third protein named Tha p 2, unrelated to Tha p 1, was identified in *Thaumetopoea pityocampa* and the cDNA encoding the complete polypeptide was amplified and sequenced (Rodriguez-Mahillo et al. 2012). The acronym Tha p 2 used to identify the last protein is unfortunate because it induces confusion with a large group of totally unrelated proteins named by the possession of the THAP domain (see Roussigne et al., 2003). Notwithstanding this potential problem we did not suggest a change of name for this protein here, provided that the stability of allergens nomenclature is a high priority (see www.allergen.org).

The *Thaumetopoea* genus is included in the subfamily Thaumetopoeinae, which is a well-defined subclade of Notodontidae containing 101 species split in 20 genera distributed in five different geographical regions: Africa, Madagascar, Europe, India, and Australia (Kiriakoff, 1970; Miller, 1991; Schintlmeister, 2013; Zahiri et al. 2011; Zahiri et al. 2013).

Most of Thaumetopeinae moths live in countries that are difficult to be explored due to geopolitical reasons. Furthermore, even for species present in countries without

political turmoil the sampling can be difficult due to the nocturnal behavior of the adults and to the fact that the larvae do not form conspicuous assemblages in the field. The genus *Thaumetopoea*, is the only Thaumetopoeinae taxon having a distribution centered in the Western Palaearctic zone (i.e. Europe, Middle East and North Africa). Furthermore, the larvae of this genus have a gregarious nesting behavior that makes them conspicuous and relatively easy to be sampled in the field (Battisti et al. 2011).

This paper deals with the characterization of the *Tha p 2* gene and Tha p 2 protein in the species of Thaumetopoeinae with a particular focus on species of the genus *Thaumetopoea* and the study of their evolution.

Materials and Methods

Taxon sampling

Larvae of twelve species of processionary moth were studied in present work. Furthermore the Lepidoptera *Euproctis chrysorrhoea*, that presents in the larval stage a similar type of urticating setae, *Lymantria dispar* and *Bombyx mori* were sampled to verify if they encoded in their genome the *Tha p 2* gene. All specimens were collected in ethanol 70% and stored at -20°C. Details on the sampling of studied taxa are provided in Table 1. As mentioned in the Introduction the Thaumetopoeinae are distributed in five different geographical regions: Africa, Madagascar, Europe, India, and Australia (Kiriakoff, 1970; Miller, 1991). Most of the species live in countries that are difficult to access due to several types of geopolitical problems. Thus the collection of samples is a very daunting task. In the case of *Thaumetopoea* the situation is much more favorable for European species while it still remains critical for Middle East and North African moths. For taxa belonging to this latter group we had just a single larva available. As a consequence for all species only one specimen was used to extract DNA and obtained the Tha p 2 sequences in order to ensure an homogenous treatment of the results (see below). The *Thaumetopoea* genus includes 12 species formally described plus a new species identified on a molecular basis (i.e. *Thaumetopoea pityocampa* ENA (Eastern-Northern Africa)) (Simonato et al. 2013). We were able to study 10 of the 13 taxa contained in the genus (see Table 1).

DNA extraction

Total DNA was extracted from single specimens through a salting-out protocol (Patwary et al., 1994) or alternatively performed using the ZR Genomic DNA- Tissue MidiPrep (Zymo Research).

The quality of the extracted DNA was assessed loading 70 ng/µl in a 0.7% agarose gel in 1X TAE, along with DNA size markers such as a 1-kb ladder. DNA was allowed to run for 40 min under a voltage of 100V. The gels were visualized in the Bio-Rad Gel DOC system (Bio-Rad Laboratories, Inc., USA). The DNA concentration was determined by NanoDrop (NanoDrop Thermo Scientific USA) and Qubit (Invitrogen, USA, using the high sensitive DNA quantification kit).

For cloning purposes (see below) DNA was concentrated by low-speed centrifugation using Amicon® Ultra- 0,5 Centrifugation Filter Devices.

PCR amplification and sequencing of Tha p 2 gene

Initially the amplification of *Tha p 2* gene was done using the primers TP2 DF, 5'-GTC CCG CAA CTA AGT GAG AAA GC-3' (forward) and TP2 DR 5'-GGT CCT TGT TCG GCC TAG TAA-3' (reverse) (Rodriguez-Mahillo et al., 2012). PCR assay was optimized for a 20 µl volume. Reaction contained 5X PCR buffer, 25mM $MgCl_2$, 25mM dNTP's, 10 mM primer forward and 10 mM primer reverse, 5u/µl *Taq* polymerase, H_2O and 2 µl of DNA. The PCR program consisted of 35 cycles at 96°C for 5', 96°C for 45'', 52°C for 30'', 72°C for 30'' and 72°C for 5'.

The PCR products exhibiting a single band in an electrophoresis gel (1% agarose), were purified with ExoSAP and sequenced according to Sanger method at the BMR Genomics service (Padova Italy) on automated DNA sequencers with the primers used for PCR amplification.

Further primers were designed for each species in order to complete the lacking portions of the gene.

Sequencing of the DNA portions 5'-upstream and 3'-downstream of the Tha p 2 gene and SNPs

The DNA portions located upstream and downstream to the published *Tha p 2* sequence were amplified using the GenomeWalker kit (Clontech, Basingstoke, UK). The obtained products were cloned with the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA)

and transformed into *E. coli*. Positive clones were PCR amplified and Sanger sequenced using the universal primers T7 (forward) and M13 (reverse). In order to identify the correct association among detected SNPs (single nucleotide polymorphisms, see results section), a cloning strategy was also applied to PCR products produced through a high fidelity polymerase reaction.

Table 1. List of the species analysed in present wok, with indication of the collection data.

Family	Species	Site collection	Latitude	Longitude	Date	Legit
Bombycidae	*Bombyx mori*	Padova, Italy	–	–	11/2012	A.Saviane
Lymantridae	*Euproctis chrysorrhoea*	Valencia, Spain	39° 50' N1°17' E	01/2013	E.Frago	
Lymantriidae	*Lymantria dispar*	Sassari, Sardegna, Italy	40° 44' N	8° 36' E	11/2012	P. Luciano
Notodontidae	*Anaphe panda*	Kakamega Forest, Western Kenia	0° 10’ N	34°47’E	11/2012	N. Mbahin
Notodontidae	*Ochrogaster lunifer*	Kenmore (Queensland), Australia	27° 30' S	152° 56' E	25/02/2005	M.P. Zalucki
Notodontidae	*Thaumetopoea bonjeani*	Forest of Chélia (Khenchela), Algeria	35° 22' N	6° 46'E	09/04/2006	M. Zamoum
Notodontidae	*Thaumetopoea herculeana*	Cabo home, Spain	41° 51' N	7° 21′ E	04/04/2009	L. Berardi
Notodontidae	*Thaumetopoea ispartaensis*	Kapidagi, Senirkent, Isparta, Turkey	37° 59’N	30° 36’ E	05/2006	M. Avci
Notodontidae	*Thaumetopoea libanotica*	Tannourine et Tahta, Lebanon	34° 12' N	35° 55' E	03/06/2006	N.Nemer, C.Lahousl
Notodontidae	*Thaumetopoea pinivora*	Gotland, Sweden	56° 18' N	18° 13' E	15/04/2006	S. Larsson
Notodontidae	*Thaumetopoea pityocampa*	Moggio , Udine, Italy	46° 24' N	13° 12' E	12/03/2003	A. Battisti
Notodontidae	*Thaumetopoea pityocampa* ENA (Eastern-Northern Africa)	Bizerte, Tunisia	37° 02' N	9° 42' E	11/2000	M. El Habib Ben Jamâa
Notodontidae	*Thaumetopoea processionea*	Caprino Veronese, Verona, Italy	45° 34' N	10° 46' E	21/5/2007	M. Faccoli, M. Zampini
Notodontidae	*Thaumetopoea solitaria*	Gamla (Golan heights), Israel	32° 59' N	35° 42' E	13/3/2005	A. Battisti
Notodontidae	*Thaumetopoea wilkinsoni*	Aladag, Turkey	37° 33' N	35° 22' E	28/3/2004	A. Battisti

–, laboratory rearing

Assembly of sequences and annotation

Single chromatograms were analyzed and edited with Chromas Lite program (Technelysium, South Brisbane QLD, Australia). The consensus sequences were produced using the SeqMan program implemented in the software package DNAStar (DNAStar, Madison, WI). The identification of the correct frame and the translation of the DNA into the corresponding Open Reading Frame (ORF) were done with the Transeq program available at the EBI web server (Mc William et al. 2013).

Bioinformatic analyses

The newly determined sequences were compared with the GenBank data bases (both nucleotide and protein) through the BLAST package BLAST available at the NCBI web site (Altschul et al. 1990) in order to identify any possible homologous counterpart.

The chemico-physical properties of Tha p 2 proteins were computed with the ProtParam and ProtScale programs available at the ExPASy web server (Gasteiger et al. 2005). The Logo of the protein multiple alignment was created with the WebLogo tool (Crooks et al., 2004). Protein secondary structures were predicted with PSIPRED (McGuffin et al. 2000), JPRED3 (Cole et al. 2008), and the 1D Protein Structure Prediction Server (Homaeian et al. 2007) packages. Putative disulfide bonds were identified with the DISULFIND server (Ceroni et al. 2006) and the DIpro program implemented in the SCRATCH server (Cheng et al. 2005). Signal peptide cleavage was predicted with SignalP 4.1 Server (Petersen et al. 2011). Search for conserved motifs was performed with the ScanProsite tool available at the ExPASy web server (Sigrist et al. 2013).

Multiple alignment of the DNA/proteins sequences were performed with the MUSCLE (Edgar, 2004) program implemented in MEGA5.2 (Tamura et al. 2005).

The phylogenetic analyses for nucleotides were carried out using maximum likelihood (ML) analysis performed with the RaxML 7.2.6 (Stamatakis 2006) program as implemented in raxmlGUI 0.93 (Silvestro et al., 2011). The molecular evolutionary model was GTR + G (Yang, 2014). Nonparametric bootstrap (BT) test was performed to assess the robustness of ML tree topology (1,000 replicates) (Felsenstein 1985). Types of selection acting on single codon were assessed with the algorithm FUBAR implemented in the datamonkey server (Murrell et al. 2013).

Results

Gene structure and taxonomic distribution

All Thaumetopoeinae species proved to encode the *Tha p 2* gene in their genome. Conversely it was not possible to amplify *Tha p 2* in the other tested Lepidoptera. This latter result could be due to the non-suitability of the used primers rather than to a true absence of the gene. Thus a thoroughly blastp search using the translated ORF was performed against the GenBank data base. This analysis failed to identify any, even distantly related, putative homologous outside the Thaumetopoeinae clade.

Tha p 2 gene contained an intronless continuous coding portion in the species where it was possible to obtain the complete sequence. The sequences upstream or downstream to the ORFs were determined for these species: *Thaumetopoea herculeana, Th. ispartaensis, Th. libanotica, Th. processionea, Th. pinivora*). For other taxa it was not possible to obtained these portions due to technical problems related to the lower quality of the DNA samples. The 5' untranslated portion exhibited a variable size with an in frame stop codon located upstream to the ATG for a minimum of 84 bp (*Th. libanotica* I,II) to a maximum of at least 114 bp (*Th. herculeana*). In this latter case the stop was not reached. BLAST searches (both blastn and blastp) using as query the 5' and 3' portions outside the ORF (maximum 640 bp in 5'; maximum 340 bp in 3) did not provide evidence of homologous sequences in GenBank.

The complete ORF was always 348 bp long included the stop codon (invariably TAA) and coded for a polypeptide 115 amino acids long.

Most of the species had two different sequences encoding always for slightly different proteins. *Th. wilkinsoni* was a notable exception presenting 20 distinct forms, confirmed through the sequencing of multiple independent clones, that clustered in two distinct groups in the phylogenetic analyses (see below). In this latter case 13 ORF differing for at least one amino acids were detected. One of these ORFs was encoded by 6 different DNA sequences and a second ORF was encoded by 3 distinct DNA sequences (Table S1). The *Th. wilkinsoni* set of DNA sequences was obtained from the sequencing of a single specimen, thus our results clearly supported the presence of multiple distinct copies of the *Tha p 2* gene in *Th. wilkinsoni*.

Protein chemico-physical properties

As pointed out above the complete ORFs spanned 115 amino acids. The bioinformatic analysis performed with SignalP 4.1 identified the first 15 residues as constituting the signal peptide for all analyzed complete ORFs. Thus the final mature protein should be 110 residues long (hereafter named mat-Tha p 2). Chemico-physical properties computed for both Tha p 2 and mat-Tha p 2 are provided in Fig. 1. Tha p 2 resulted S,G,C,L-rich (amino acids listed according to their abundance). Indeed the four residues accounted for a minimum of 35.65% in *Th. pinivora* I to a maximum of 40% of *Th. wilkinsoni* XV. More in general polar amino-acids dominated over other categories i.e. basic, acid and hydrophobic residues (Fig. 1). This result was corroborated by the hydropathy profile (Fig. 1).

A similar pattern was observed in mat-Tha p 2, where however L dropped to a decidedly lower position in the amino acid composition (Fig. 1). The presence of a minimum of 9 Cys residues (*Th. libanotica* II) in the Tha p 2 opened up the possibility for numerous intra- and inter-chain S-S bridges (see below).

The estimated average molecular weights for Tha p 2 and mat-Tha p 2 were 12,630 ± 69 Da and 11,049 ± 69 Da, respectively. Both Tha p 2 and mat-Tha p 2 exhibited acid isoelectric points (see Fig. 1).

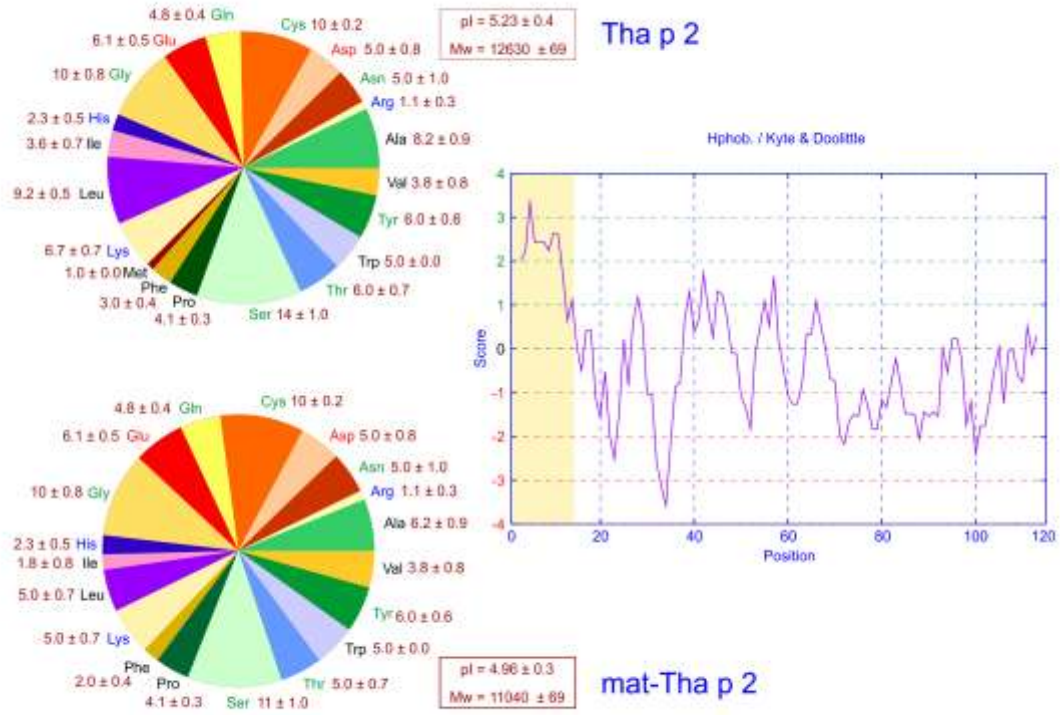


Figure 1. Chemical-physical properties of Tha p 2 protein. On the left, amino acid composition. Value located near the amino acid (listed according to the three letter IUPAC code) indicate the average number of residues followed by the standard deviation. Mw, average molecular weight (Dalton) ± standard deviation; pI, average Isoelectric point ± standard deviation. On the right Hydrophobicity diagram (Kythe and Doolittle, 1992) of Tha p 2 protein.

Secondary structure prediction and motifs

A multiple alignment of Tha p 2s (see below Fig. 4) was produced to evaluate the level of conservation among various sequences and used to generate the WebLogo vignette for this protein (Fig. 2). The average p-distance among Tha p 2s was 0.08 ± 0.04 with the maximum value being 0.19 (e.g. *Th. herculeana* vs *Th. pityocampa* ENA II) and the minimum values being 0.01 (e.g. *Th. libanotica* I vs *Th. ispartaensis* I). Thus the protein proved to be very conserved as well corroborated by the WebLogo vignette (Fig. 2).

A consensus of the predicted secondary structures obtained for the protein is provided in Fig. 2. It must be noted here that the length of both α-helices and β-sheets depicted in Fig. 2 represents the maximum extent obtained by various methods. Minor differences were also observed among the various sequences analyzed.

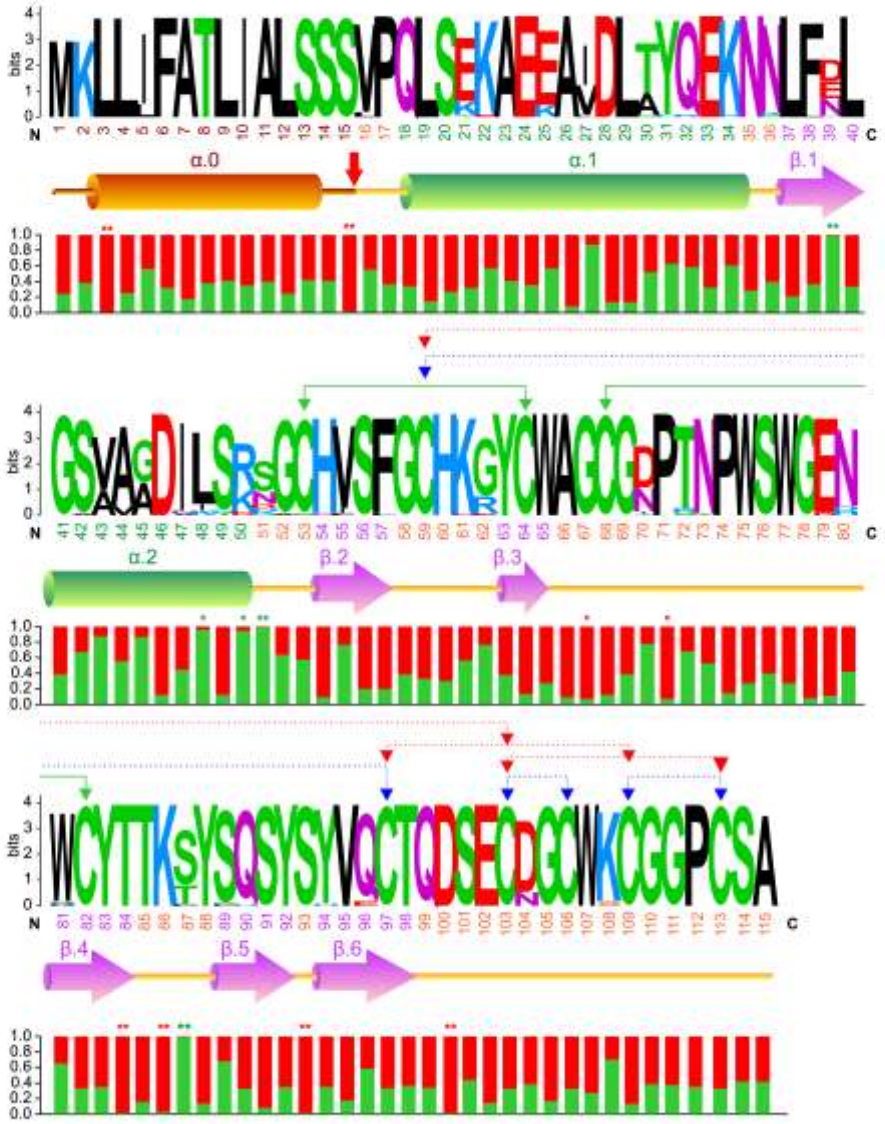


Figure 2. Prediction of secondary structure, Logo vignette of Tha p 2 and type of selection acting on the Tha p 2 gene. The Logo vignette was created with the WebLogo tool. The secondary structure is a consensus of the predictions obtained with PSIPRED, JPRED3 and 1D Protein Structure Prediction Server packages. Red arrow identifies the placement of the signal peptide cleavage. Secondary structure elements: α0-α2, predicted α-helices; β1-β6 predicted β-barrels. S-S bridges connected by a continuous green line were predicted by both DISULFIND and DIpro programs; S-S bridges linked by a dotted blue line were predicted only by DISULFIND while those connected by a red dotted line were identified solely by in red DIpro program. Histograms show the relative contributions of negative and positive selection acting on single codons of *Tha p 2* as calculated with FUBAR. Red bar, negative selection component; green bar, positive selection component. **, posterior probability >95%; *, posterior probability >90%.

An initial α-helix (α.0) encompassed most of the putative signal peptide. The first third of the mat-Tha p 2 included two α-helices (α.1, α.2) interspersed with a short β-sheet not present in all species. The α.1 helix exhibited a high concentration of acidic residues (E, D, Q). The remaining two third of mat-Tha p 2 secondary structure was characterized by five short β-sheets connected by stretches of random coiled residues. Predictions of disulfide bonds identified five intra-chain bridges that were only partly congruent (C53-C64; C68-C82;) or differed more or less markedly (C59-C97; C103-C106; C109-C113; DISULFIND), (C59-C106; C97-C109; C103-C113; DIpro). Thus these predictions must be considered tentative. Furthermore, it remained to be tested if some of the 10 Cys present in the Tha p 2 were actually involved in inter-chains bonds.

Finally the PROSITE scan identified only conserved motifs with high probability of occurrence in proteins (Fig. S1). Thus, the effective role of these motifs for Tha p 2 needs to be evaluated against most compelling evidence.

Evolution of Tha p 2 gene and protein

A maximum likelihood phylogenetic analysis was performed on a multiple alignment including the complete and incomplete ORFs of *Tha p 2* gene. The obtained tree is depicted in Fig. 3. Provided that no reliable outgroup sequences exist for *Tha p 2* (see above), the tree was arbitrarily rooted and four main clades (A-D), each receiving BT ≥ 86%, were identified. These clades were further corroborated by the sharing of unique molecular signatures present in the encoded protein (see below). Clade A, further split in the sub-clades A1 and A2, contained the 20 sequences determined for *Th. wilkinsoni* plus those obtained from *A. panda* (A1) and *O. lunifer* (A2). Subclade A1 received BT support, that however lacked for sub-clade A2. Clades B and C received very high BT supports. Clade B contained the sequences obtained for the *Th. pityocampa* complex. Clade C contained the sequences obtained for *Th. solitaria*, *Th. herculeana* and *Th. processionea*. Finally, clade D encompassed *Tha p 2* derived from *Th. pinivora*, *Th. ispartaensis*, *Th. libanotica*, and *Th. bonjeani*. Comparison of branch lengths showed that the substitution rate was variable among different sequences, with the higher frequencies observed for sequences of clades C and D. Changes of amino acids along the multiple alignment were tracked with the Mesquite program using as reference the tree presented in Fig. 3.

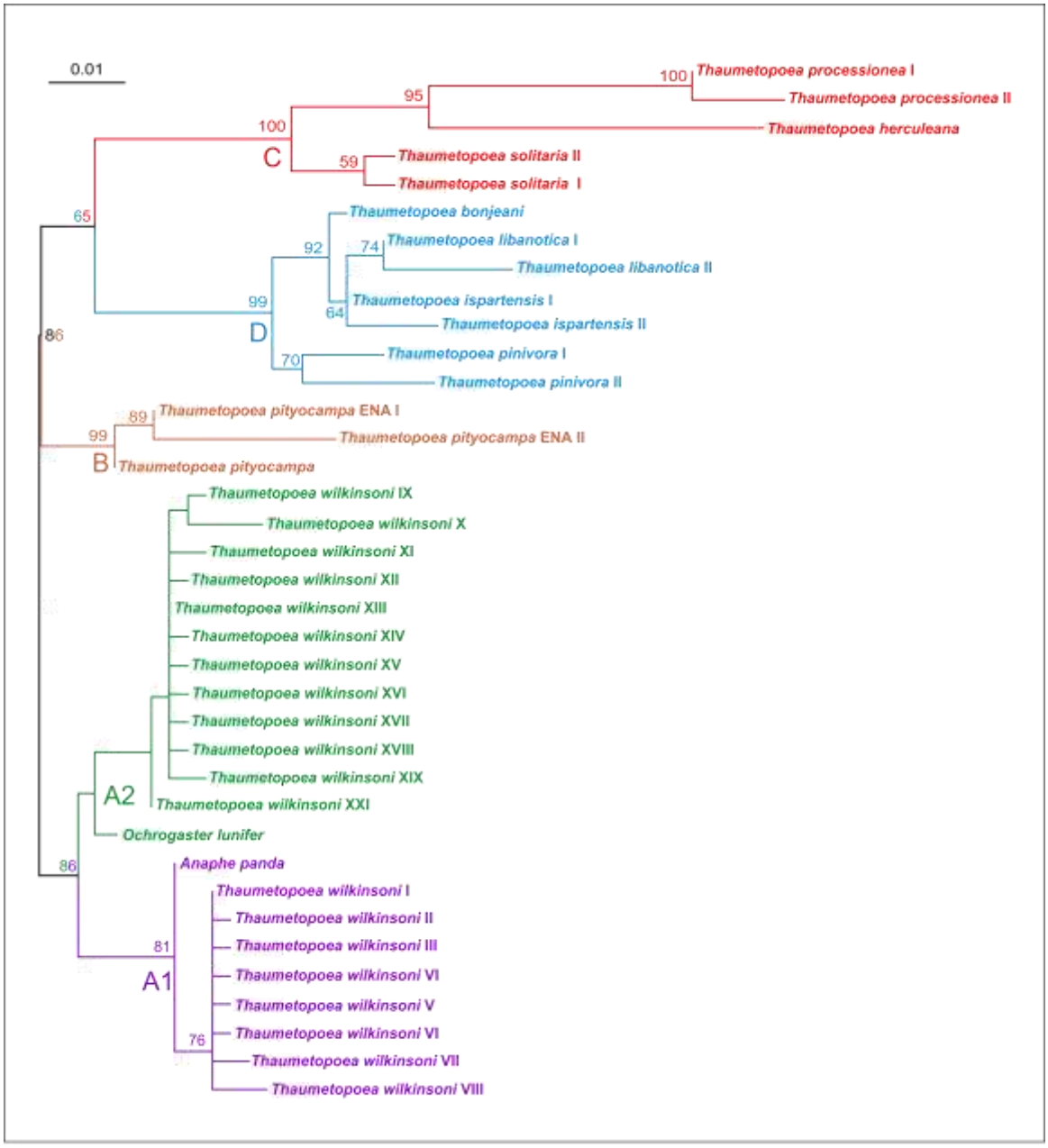


Figure 3. Phylogenetic analysis of Tha p 2 gene. The ML tree (-ln = 1389.189388) was produced with RaxML program by applying the GTR+G model to the Tha p 2 multiple alignment. Numbers indicate bootstrap values (after 1,000 replicates) expressed in percentage. Capital letters label nodes leading to major clades. The scale bar represents 0.01 substitutions/site.

The positions of the Tha p 2 multiple alignment containing these molecular signatures are presented in Fig. 4. The inspection of Fig. 4 shows that variable positions are distributed along the whole length of the alignment with a higher concentration in the first part.

Molecular signatures can be restricted to a single sequence (e.g. *Th. ispartensis* II in position 90 His(H) vs Q(Gln)) or characterized groups. Representative examples will be briefly described below while the complete list of changes is provided in the Supporting Table S2. The sequences of clade C exhibit a peculiar Lys (K) residue in position 25 that in all other species is occupied by a Glu (E). Analogously the sequences of clade A1 are characterized in position 50 by a Lys (K) instead of the widespread Ser (S). The same residue appeared independently also in *Th. solitaria* I-II, while in *Th. processionea* I-II the change produced an Asp (D) residue. Finally clades C and D share a common Asn (N) in position 70 while in other groups Asp (D) is the standard. Further changes occurred independently in *Th. pytiocampa* ENA II (Ile, I) and *Th. ispartensis* II (Tyr, Y).

We tested also the type of selection that acted on the codons of *Tha p 2* gene. The analysis conducted with FUBAR showed that a limited number of codons were subject to positive selection with very strong (codon: 39, 51, 87) or strong (codon: 48, 50) statistical corroboration (Fig. 2). This aspect was further supported by the high variation of the encoded amino acids. Conversely codons 3, 15, 84, 86, 93 and 100 resulted to be under strong negative selection. Also codons 67 and 71 appeared to have experienced a negative selection but in this case the statistical support was less conclusive. The conservation observed at the amino acid level (Fig. 4) empirically supported the view that most of the protein experienced at negative selection in its evolution.

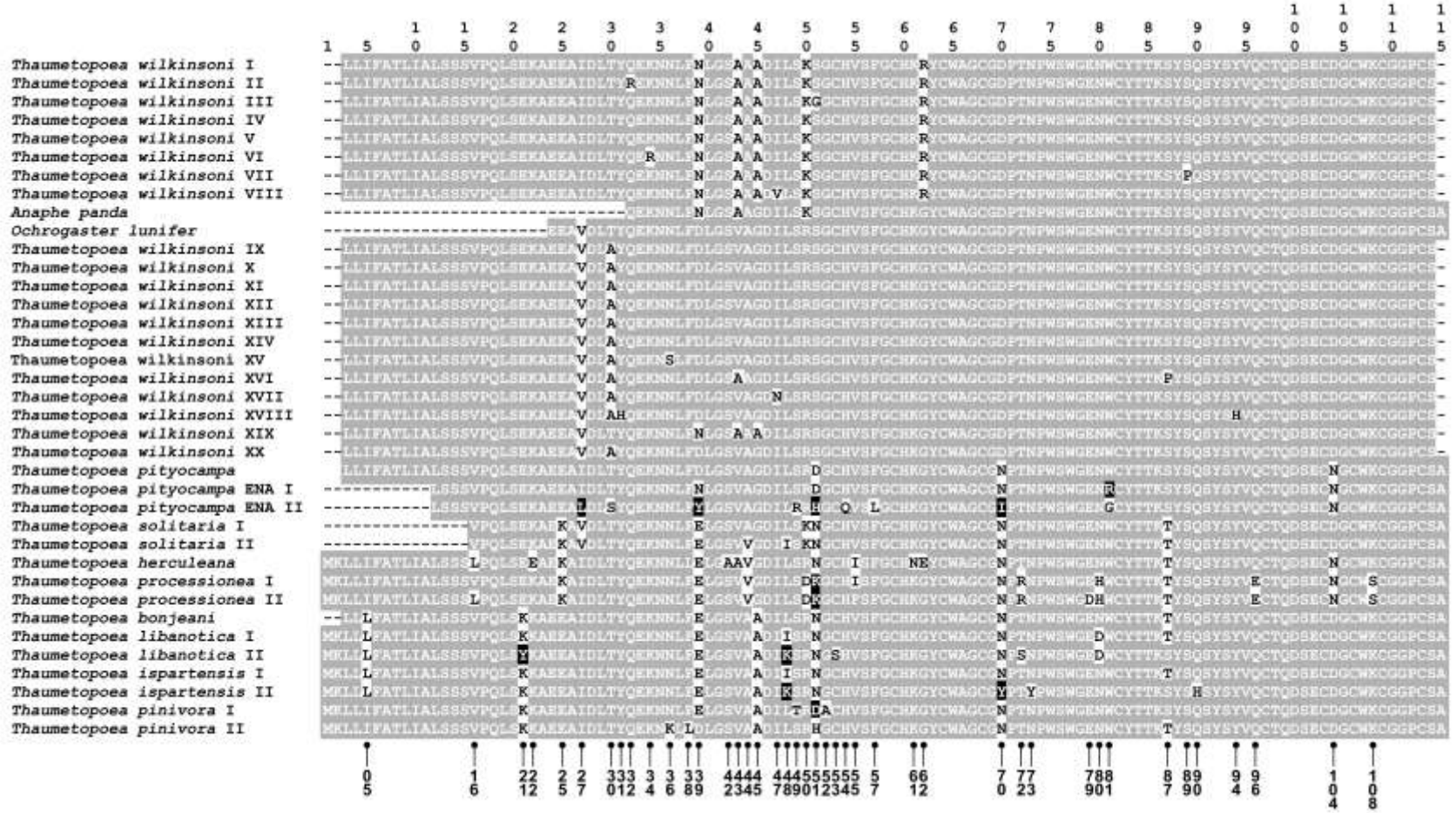


Figure 4. Multiple alignment of Tha p 2 proteins. Conserved amino acid residues are depicted with a gray background. The amino acid characterizing the different clades are presented black colored or with a black background. Positions not fully conserved are marked and numbered below the alignment.

Discussion

We determined from a minimum of one to 20 distinct different DNA sequences of *Tha p 2* gene in the various species of Thaumetopoeinae (Fig. 3). Provided that sequences were obtained for all species from a single specimen (see 2.1. Taxon sampling), the easiest interpretation of these findings is that *Tha p 2* is a single copy gene with different level of heterozygosity in various taxa, with a unique major exception represented by *Th. wilkinsoni,* where we have to postulate the existence of 10 genes provided that we obtained 20 distinct forms. Expansion was certainly favored by the short length of the coding sequence and by the fact that *Tha p 2* gene is intronless. However, we cannot exclude the possibility that two copies, very similar or identical, exist also in other species. The availability of a whole genome for one or more Thaumetopoeinae species will allow to definitely settle this point.

The occurrence of a multigene expansion of *Tha p 2* in *Th. wilkinsoni* is in agreement with the general behavior of cuticular proteins (Cornman 2009).

We were able to predict a secondary structure for the Tha p 2 protein that needs to be corroborated through crystallographic studies that were outside the goals of this paper. Among the various biochemical feature of Tha p 2 is the massive presence (9-10) of Cys residues located along in the C-half of the mat-Tha p 2. The richness of Cys opens up the possibility for several intra- and inter-chain S-S bonds that are well worthy to be explored in further studies in order to understand how Tha p 2 is related with other cuticular proteins present in the larva setae.

The *Tha p 2* gene was identified and sequenced for all tested Thaumetopoeinae species. Conversely we failed to amplify it in other moths. This could have been just the effect of unsuitable primers. Thus we performed a BLASTP search against GenBank that harbors a huge amount of sequences obtained from insect taxa including complete genomes (e.g. *Apis mellifera*, *Drosophila melanogaster, Anopheles gambiae*). This search failed to identify any possible homologous counterpart, even distantly related.

All these findings point in favor of the uniqueness of the *Tha p 2* and to the possibility that it represents a taxon-specific gene restricted to Thaumetopoeinae. We cannot rule out completely the possibility that indeed the gene is present in other insect species. However, if this latter case is true we must hypothesize that other taxa having this gene have not yet been sequenced or that alternatively their "*Tha p 2*" gene encodes a polypeptide so diverging that cannot be associated to Tha p 2 through BLASTp search.

The level of conservation among various Tha p 2s sequenced in present work is very high, thus it appears contradictory to imagine a so rapid level of divergence among the representatives of the same protein.

Tha p 2 protein has been initially isolated from setae of *Th. pityocampa* (Rodríguez-Mahillo et al., 2012) thus it is reasonable to include it within a novel family of cuticular proteins. The fact that it is restricted to a single taxon is in agreement with previous knowledge on cuticular protein that in several cases are taxon-specific (e.g. Willis, 2010). Tha p 2 is Gly-rich, a feature shared with other cuticular proteins (Willis 2010), while its richness in Cys remains peculiar. It is not yet known where Tha p 2 is synthesized.

A limited number of codons was identified with statistical corroboration to have experienced the action of negative selection. Even less numerous resulted the codons under positive selection. All these findings suggest that at the DNA level the large part of the coding portion evolved under a neutral selection. However, a closer inspection of FUBAR results (Fig. 2) shows that for the majority of codons the negative selection component (red bar) prevails over the positive selection component (green bar) more or less clearly (e.g. codon 74, 77, 78). This result suggests that a mild to robust effect of purifying selection shaped the evolution of *Tha p 2* even at the gene level. This aspect is more evident when we consider the level of conservation occurring at the amino acid level (Figs.2,4) where 64% of the positions are constant and 16% of positions are singletons, i.e. the amino acid changes only in one sequence (e.g. positions 79 and 89 of Fig. 2).

The branching off in the *Tha p 2* phylogenetic tree is largely in agreement with the species phylogeny with a major exception represented by *Th. wilkinsoni* that in the species tree is sister taxon of the *Th. pityocampa* complex (Simonato et al. 2013; unpublished data).

The absence of credible putative orthologous counterparts for *Tha p 2* gene makes problematic to identify its origin. Possibly it started as the duplication of an existing gene (Zhang 2003) followed by a very rapid diversification that made impossible to recognize the initial point. Irrespectively to its origin, the *Tha p 2* gene and particularly the encoded protein is very well conserved among the various Thaumetopoeinae species. The typical and widespread reactions caused by Thaumetopoeinae setae in humans and domestic animals (Moneo et al. 2014), and the indirect evidence of their action against birds

(Barbaro and Battisti 2011), strongly support the idea that Tha p 2 plays a pivotal role in the defensive system of the larvae against predatory vertebrates. Indeed the communal behavior of the urticating larvae of the *Thaumetopoea* genus makes them a very conspicuous target for predators like insectivorous birds and mammals (Battisti et al. 2011). Interestingly, the protection of the larvae of the other two genera considered in this study, *Anaphe* and *Ochrogaster*, seem to be indirect as the setae are produced by the female moth and become incorporated in the tent spun by the larvae. Thus we can hypothesize that once appeared, Tha p 2 was selected to maximize its effectiveness against these predators. Apparently this goal was achieved very early in the protein evolution provided that the amino acid variation among different taxa is very limited. As Tha p 2 is one of several proteins associated with the setae, it would be necessary to clarify the nature and action of the other proteins in order to fully understand their role in the defense system.

Acknowledgments

We warmly acknowledge all the people that have provided help for this work and especially Mustafa Avci, Stig Larsson, Pietro Luciano, Norbert Mbahin, Zvi Mendel, Nabil Nemer, Juan Pino, Alessio Saviane, Myron Zalucki, and Mohamed Zamoum for providing insect material and M. Babbucci, A. Basso, R, Franch, M. Simonato for useful suggestions in the laboratory work. This work was funded by the French National Research Agency in the framework of project ANR No. 07BDIV 013 “URTICLIM" and by the University of Padova grant 2010 – C91J10000320001.

Disclosure

The authors have no conflicts of interests to report, including specific financial interests and relationships and affiliations.

References

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J., (1990). Basic local alignment search tool. *Journal of Molecular Biology*. 215: 403–410.

Barbaro L., Battisti A., (2011). Birds as predators of the pine processionary moth (Lepidoptera: Notodontidae). *Biological Control* 56:107–114.

Battisti A., Holm G., Fagrell B., Larsson S., (2011). Urticating hairs in arthropods: their nature and medical significance. *Annual Review of Entomology* 56: 203–220.

Ceroni A., Passerini A., Vullo A., Frasconi P., (2006). DISULFIND: a Disulfide Bonding State and Cysteine Connectivity Prediction Server. *Nucleic Acids Research* 34 (Web Server issue), W177–W181.

Cheng J., Randall A., Sweredoski M., Baldi P., (2005). SCRATCH: a Protein Structure and Structural Feature Prediction Server. *Nucleic Acids Research* 33 (Web Server Issue), W72-W76.

Cole, C., Barber, J.D., Barton, G.J., (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* 36: W197–W201.

Cornman R.S., (2009). Molecular evolution of *Drosophila* cuticular protein genes. *PLoS One*, 4(12): e8345. doi:10.1371/journal.pone.0008345.

Crooks G.E., Hon G., Chandonia, J.M., Brenner S.E., (2004). WebLogo: a sequence logo generator. *Genome Research*. 14 :1188–1190.

Démolin G., (1963). Les 'miroirs' urticants de la processionnaire du pin (*Thaumetopoea pityocampa* Schiff.). *Agricultural Zoological Reviews*. 10–12: 107–114.

Edgar R.C., (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research.* 32: 1792–1797.

Felsenstein J., (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

Floater G.J., (1998). Tuft scales and egg protection in *Ochrogaster lunifer* Herrich-Schäffer (Lepidoptera: Thaumetopoeidae). *Australian Journal of Entomology*. 37: 34–39.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., (2005). Protein identification and analysis tools on the ExPASy Server. In: Walker J.M., editor. *The Proteomics Protocols Handbook.* New York: Humana Press pp. 571–607.

Homaeian L., Kurgan L.A., Cios K.J., Ruan J., Chen K., (2007). Prediction of protein secondary structure content for the twilight zone sequences. *Proteins* 69: 486–498.

Kiriakoff S.G., (1970). Lepidoptera Familia Thaumetopoeidae. In: Wytsman P, editor. *Genera Insectorum*. Anvers Belgium: SPRL Mercurius. pp. 1–54.

Kyte J., Doolittle R.F., (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* 157: 105–132.

Lamy M., Ducombs G., Pastureaud M.H., Vincendeau P., (1982). Cuticular productions of *Thaumetopoea pityocampa* Schiff. Lepidoptera: urticant and laying apparatus. *Bulletin de la Socété zoologique de France* 07: 515–529.

Lamy M., Pastureaud M.H., Ducombs G., (1985). Thaumetopoein, an urticating protein of the processionary hairs of the caterpillar (*Thaumetopoea pityocampa* Schiff) (Lepidoptera, Thaumetopoeidae). *Comptes Rendus de l'Académie des Sciences, Series III* 301, 173–176.

Lamy M., Pastureaud M.H., Novak F., Ducombs G., Vincendeau P., Maleville J., Texier L., (1986). Thaumetopoein: an urticating protein from the hairs and integument of the pine processionary caterpillar (*Thaumetopoea pityocampa* Schiff., Lepidoptera, Thaumetopoeidae). *Toxicon* 24: 347–356.

Larsson S., Backlund A., (2009). Regarding the putative identity of a moth (*Thaumetopoea* spp.) allergen. *Allergy* 64: 493.

Maier H., Spiegel W., Kinaciyan T., Krehan H., Cabaj A., (2003). The oak processionary caterpillar as the cause of an epidemic airborne disease: survey and analysis. *British Journal of Dermatology* 149: 990–997.

McGuffin L.J., Bryson K., Jones D.T.(2000). The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404–405.

McWilliam H., Li W., Uludag M., Squizzato S., Park Y.M., Buso N., Cowley A.P., Lopez R., (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* 41 (Web Server issue), W597–W600.

Miller J.S., (1991). Cladistics and classification of the Notodontidae (Lepidoptera: Noctuoidea) based on larval and adult morphology. *Bulletin of the American Museum Natural History* 204, 1–230.

Moneo I., Vega J.M., Caballero M.L., Vega J., Alday E., (2003). Isolation and characterization of Tha p 1, a major allergen from the pine processionary caterpillar *Thaumetopoea pityocampa*. *Allergy* 58, 34–37.

Moneo I., Battisti A., Dufour B., García-Ortiz J.C., González-Munoz M., Moutou F., Paolucci P., Petrucco Toffolo E., Rivière J., Rodriguez-Mahillo A.I., Roques A., Roques L., Vega J.M., Vega J., (2014). Medical and Veterinary Impact of the Urticating Processionary Larvae. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp: 359-410.

Murrell B., Moola S., Mabona A., Weighill T., Sheward D., Kosakovsky-Pond S.L., Scheffler K., 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology of Evolution* 30, 1196–1205.

Patwary M.U., Kenchington E.L., Bird C.J., Zouros E., (1994). The use of random amplified polymorphic DNA markers in genetic studies of the sea scallop *Placopecten magellanicus* (Gmelin, 1791). *Journal of shellfish research* 13, 547–553.

Petersen T.N., Brunak S., von Heijne G., Nielsen H. 2011., SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8, 785–786.

Petrucco Toffolo E., Zovi D., Perin C., Paolucci P., Roques A., Battisti A., Horvath H., (2014). Size and dispersion of urticating setae in three species of processionary moths. Integrative Zoology 9, 320–327.

Rodríguez-Mahillo A.I., González-Muñoz M., Vega J.M., López J.A., Yart A., Kerdelhué C., Camafeit E., García Ortiz J.C., Vogel H., Toffolo E.P., Zovi D., Battisti A., Roques A., Moneo I., (2012). Setae from the pine processionary moth (*Thaumetopoea pityocampa*) contain several relevant allergens. *Contact Dermatitis* 67, 367–374.

Roussigne M., Kossida S., Lavigne A.C., Clouaire T., Ecochard, V., (2003). The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends *Trends* in *Biochemical Science* 28, 66–69.

Schabel H.G., (2006). *Forest entomology in East Africa: forest insects of Tanzania*. Dordrecht: Springer. 328 p.

Schintlmeister A., (2013). *World Catalogue of Insects 11 Notodontidae & Oenosandridae (Lepidoptera)*. Boston: EJ Brill. 605 p.

Sigrist C.J.A., de Castro E., Cerutti L., Cuche B.A., Hulo N., (2013). New and continuing developments at PROSITE. *Nucleic Acids Research* 41, D344-D347.

Silvestro D., Michalak I., (2012). RaxmlGUI: a graphical front-end for RAxML. *Organism Diverity & Evolution* 13, 335–337.

Simonato M., Battisti A., Kerdelhué C., Burban C., Lopez-Vaamonde C., Pivotto I., Salvato P., Negrisolo E., (2013). Host and phenology shifts in the evolution of the social moth genus *Thaumetopoea*. *PloSOne* 8(2) e57192. doi:10.1371/journal.pone.0057192.

Stamatakis A., (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.

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* 28, 2731–2739.

Willis J.H., (2010). Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochemistry and Molecular Biology* 40, 189–204.

Yang Z., (2014). *Molecular evolution: a statistical approach.* Oxford (UK): Oxford University Press. p.492.

Zahiri R., Kitching I. J., Lafontaine J. D., Mutanen M., Kaila L., Holloway J.D., Wahlberg N., (2011). A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera). *Zoolical Scripta* 40, 158–173.

Zahiri R., Lafontaine D., Schmidt C., Holloway J.D., Kitching I. J., Mutanen M., Wahlberg N., (2013). Relationships among the basal lineages of Noctuidae (Lepidoptera, Noctuoidea) based on eight gene regions. *Zoological Scripta* 42, 488–507.

Zhang J., (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 18, 292–298.

Table 1. List of the species analysed in present wok, with indication of the collection data.

Family	Species	Site collection	Latitude	Longitude	Date	Legit
Bombycidae	*Bombyx mori*	Padova, Italy	–	–	11/2012	A.Saviane
Lymantridae	*Euproctis chrysorrhoea*	Valencia, Spain	39° 50' N1°17' E	01/2013	E.Frago	
Lymantriidae	*Lymantria dispar*	Sassari, Sardegna, Italy	40° 44' N	8° 36' E	11/2012	P. Luciano
Notodontidae	*Anaphe panda*	Kakamega Forest, Western Kenia	0° 10’ N	34°47’E	11/2012	N. Mbahin
Notodontidae	*Ochrogaster lunifer*	Kenmore (Queensland), Australia	27° 30' S	152° 56' E	25/02/2005	M.P. Zalucki
Notodontidae	*Thaumetopoea bonjeani*	Forest of Chélia (Khenchela), Algeria	35° 22' N	6° 46'E	09/04/2006	M. Zamoum
Notodontidae	*Thaumetopoea herculeana*	Cabo home, Spain	41° 51' N	7° 21′ E	04/04/2009	L. Berardi
Notodontidae	*Thaumetopoea ispartaensis*	Kapidagi, Senirkent, Isparta, Turkey	37° 59’N	30° 36’ E	05/2006	M. Avci
Notodontidae	*Thaumetopoea libanotica*	Tannourine et Tahta, Lebanon	34° 12' N	35° 55' E	03/06/2006	N.Nemer, C.Lahousl
Notodontidae	*Thaumetopoea pinivora*	Gotland, Sweden	56° 18' N	18° 13' E	15/04/2006	S. Larsson
Notodontidae	*Thaumetopoea pityocampa*	Moggio , Udine, Italy	46° 24' N	13° 12' E	12/03/2003	A. Battisti
Notodontidae	*Thaumetopoea pityocampa* ENA (Eastern-Northern Africa)	Bizerte, Tunisia	37° 02' N	9° 42' E	11/2000	M. El Habib Ben Jamâa
Notodontidae	*Thaumetopoea processionea*	Caprino Veronese, Verona, Italy	45° 34' N	10° 46' E	21/5/2007	M. Faccoli, M. Zampini
Notodontidae	*Thaumetopoea solitaria*	Gamla (Golan heights), Israel	32° 59' N	35° 42' E	13/3/2005	A. Battisti
Notodontidae	*Thaumetopoea wilkinsoni*	Aladag, Turkey	37° 33' N	35° 22' E	28/3/2004	A. Battisti

–, laboratory rearing

Table S1. List of the Thaumetopoeinae species (Lepidoptera Notodontidae) and accession numbers of the new *Tha p 2* sequences.

Species	**Sequence name**	**Accession**
Anaphe panda	*Anaphe panda*	LM994946
Ochrogaster lunifer	*Ochrogaster lunifer*	LM994966
Thaumetopoea bonjeani	*Thaumetopoea bonjeani*	LM994944
Thaumetopoea herculeana	*Thaumetopoea herculeana*	LM994937
Thaumetopoea ispartaensis	*Thaumetopoea ispartaensis* I	LM994941
	Thaumetopoea ispartaensis II	LM994942
Thaumetopoea libanotica	*Thaumetopoea libanotica* I	LM994943
	Thaumetopoea libanotica II	LN558421
Thaumetopoea pinivora	*Thaumetopoea pinivora* I	LM994967
	Thaumetopoea pinivora II	LM994968
Thaumetopoea pityocampa	*Thaumetopoea pityocampa*	LM994969
Thaumetopoea pityocampa ENA (Eastern-Northern Africa)	*Thaumetopoea pityocampa* I (genotype ENA)	LM994970
	Thaumetopoea pityocampa II (genotype ENA)	LN558420
Thaumetopoea processionea	*Thaumetopoea processionea* I	LM994938
	Thaumetopoea processionea II	LM994939
Thaumetopoea solitaria	*Thaumetopoea solitaria* I	LM994940
	Thaumetopoea solitaria II	LN558419
Thaumetopoea wilkinsoni	*Thaumetopoea wilkinsoni* I	LM994945
	Thaumetopoea wilkinsoni II	LM994947
	Thaumetopoea wilkinsoni III	LM994948
	Thaumetopoea wilkinsoni IV	LM994949
	Thaumetopoea wilkinsoni V	LM994950
	Thaumetopoea wilkinsoni VI	LM994951
	Thaumetopoea wilkinsoni VII	LM994952
	Thaumetopoea wilkinsoni VIII	LM994953
	Thaumetopoea wilkinsoni IX	LM994954
	Thaumetopoea wilkinsoni X	LM994955
	Thaumetopoea wilkinsoni XI	LM994956
	Thaumetopoea wilkinsoni XII	LM994957
	Thaumetopoea wilkinsoni XIII	LM994958
	Thaumetopoea wilkinsoni XIV	LM994959
	Thaumetopoea wilkinsoni XV	LM994960
	Thaumetopoea wilkinsoni XVI	LM994961
	Thaumetopoea wilkinsoni XVII	LM994962
	Thaumetopoea wilkinsoni XVIII	LM994963
	Thaumetopoea wilkinsoni XIX	LM994964
	Thaumetopoea wilkinsoni XX	LM994965

Table S2. Amino acids changes in the Tha p 2 multiple alignment and molecular signatures of clades.

N	SEQUENCES	POS	CHANGE	N	SEQUENCES	POS	CHANGE
1	Clade D except *T. pinivora* I,II	5	L vs I	1	*T. processionea* I, II	50	D vs R
2	*T. herculeana*; *T. processionea* II	16	L vs V	1	Clade A1[g] + Clade A2	51	S (G) vs D, H, K, N
1	Clade D[a]	21	K vs E	1	[g] *T. wilkinsoni III*	51	G vs S
1	[a] *T. libanotica* II	21	Y vs K	1	Clade C[h] + Clade D[i]	51	N vs D, S
1	*T. herculeana*	22	E vs K	1	[h] *T. processionea* I, II	51	K vs N
1	Clade C	25	K vs E	1	[i] *T. pinovora* I	51	D vs H, N
2	Clade A2; *T. solitaria* I,II	27	V vs I	1	[i] *T. pinivora* II	51	H vs D, N
1	*T. pityocampa* ENA II	27	L vs I	1	Clade B[j]	51	D (H) vs N (D,H,K)
1	*T. pityocampa* ENA II	30	S vs T	1	[j] *T. pityocampa* ENA II	51	H vs D
1	Clade A2[b] except *O.lunifer*	30	A vs T	1	*T. pinivora* I	52	A vs G
1	[b] *T. wilkinsoni* XVIII	30	T vs A	1	*T. libanotica* II	53	S vs C
1	*T. wilkinsoni XVII*	31	H vs Y	1	*T. pityocampa* ENA II	54	Q vs H
1	T. *wilkinsoni XVII*	32	R vs Q	1	*T. herculeana* + *T. processionea* I, II	55	I vs V
1	*T.wilkinsoni VI*	34	R vs K	1	*T. pityocampa* ENA II	57	L vs F
1	*T. pinivora* II	36	K vs N	1	*T. herculeana*	61	N vs K
1	*T. wilkinsoni XVI*	36	S vs N	1	*T. herculeana*	62	E vs G
2	*T. herculeana*; *T. pinivora* II	38	L vs F	1	Clade A1 excluded *A.panda*	62	R vs G
1	Clade C + Clade D[c]	39	E vs D, N, Y	1	Clade A1 + Clade A2	70	D vs N
1	[c] *T. pinivora* II	39	D vs E	1	Clade B[k] + Clade C + Clade D[l]	70	N vs D
1	Clade A2[d]	39	D vs N	1	[k] *T. pityocampa* ENA II	70	I vs N
1	[d] *T. wilkinsoni XVIII*. **N ancestral**	39	N vs D	1	[l] *T. ispartensis II*	70	Y vs N
1	*T. pityocampa*. **N ancestral**	39	D vs N	1	*T. processionea* I, II	72	R vs T
1	*T. pityocampa ENA* II. **N ancestral**	39	Y vs N	1	*T. libanotica* II	72	S vs T
3	*T. pityocampa* ENA I; Clade A1; *T. wilkinsoni XVIII*. **D ancestral**	39	N vs D	1	*T. ispartensis* II	73	Y vs N
1	*T. pityocampa* ENA II. **D ancestral**	39	Y vs D	1	*T. processionea* II	79	D vs E
1	*T. herculeana*	42	A vs S	1	*T. processionea* I, II	80	H vs N
3	*T. herculeana*; Clade A1; *T. wilkinsoni XVIII* + *T. wilkinsoni XIX*	43	A vs V	1	*T. libanotica* I, II	80	D vs N
1	Clade C[e]	44	V vs A	1	*T. pityocampa* ENA I	81	R vs W
1	[e] *T. solitaria* 1	44	A vs V	1	*T. pityocampa* ENA II	81	G vs W
3	Clade D; Clade A1 except *A.panda*; *T. wilkinsoni XVIII*	45	A vs G	1	Clade C + Clade D[m]	87	T vs S
1	*T. wilkinsoni II*	47	V vs I	3	[m] *T. pinivora* 1; *T. ispartensis* II; *T. libanotica* II	87	S vs T
1	*T. wilkinsoni XVI*	47	N vs I	1	*T. wilkinsoni XIX*	87	P vs S
2	*T. solitaria* II; *T. ispartensis* I to *T. libanotica* I[f]	48	I vs L	1	*T. wilkinsoni I*	89	P vs S
1	[f] *T. libanotica* II	48	K vs I	1	*T. ispartensis* II	90	H vs Q
1	*T.ispartensis* II	48	K vs L	1	*T. wilkinsoni XVII*	94	H vs Y
1	*T. pityocampa* ENA II	49	R vs S	1	*T. processionea* I, II	96	E vs Q
1	*T. pinivora* I	49	T vs S	2	Clade B; *T. herculeana* + *T. processionea* I, II	104	N vs D
2	*T. solitaria* I,II; Clade A1	50	K vs R	1	*T. processionea* I, II	108	S vs K

N, number of changes; POS, position in the alignment; Change, type of change.
[a-m], these superscript letters refer to taxa, listed immediately below the reference group, departing from the general behavior observed for a specific clade.

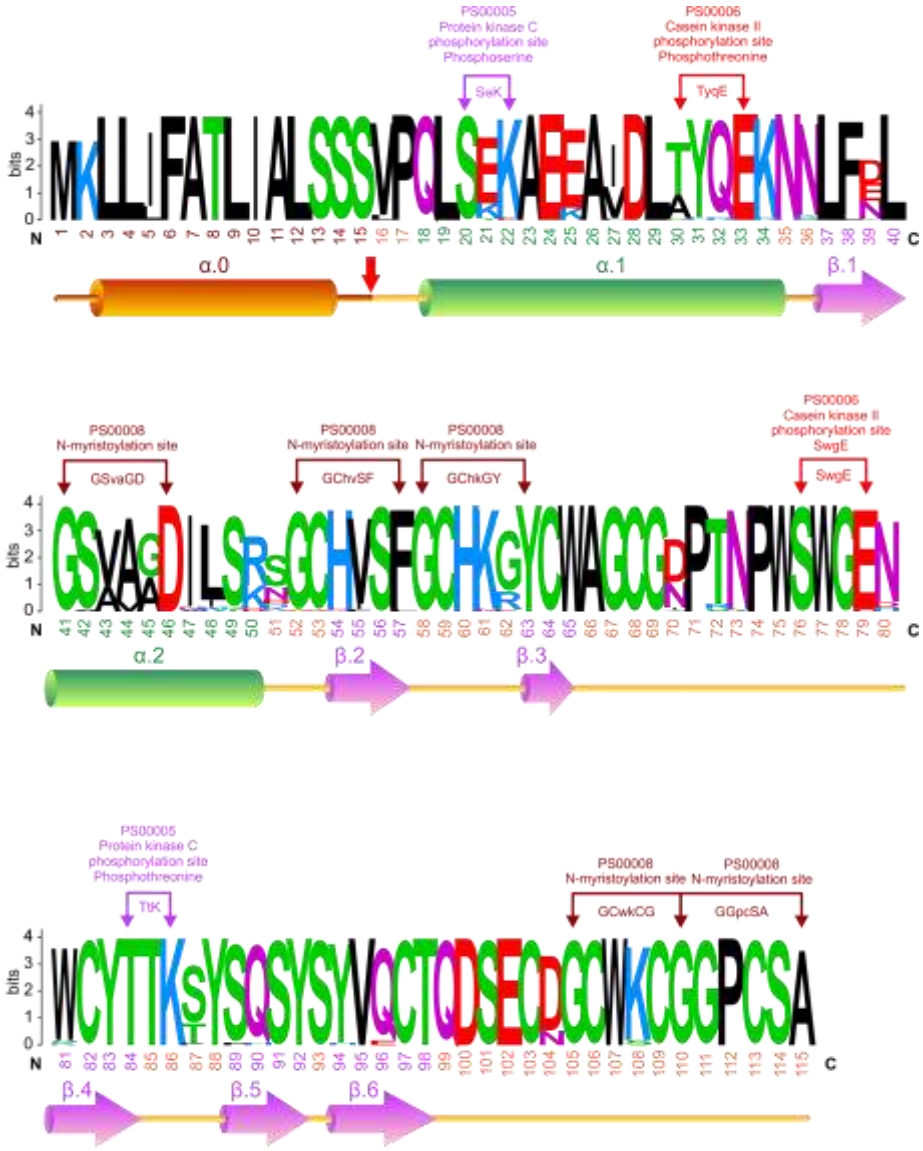


Figure S1. Distribution in Tha p 2s of conserved motifs with high probability of occurrence in proteins. Motifs were identified with the the ScanProsite tool available at the Expasy web server. The Logo vignette was created with the WebLogo tool. The secondary structure is a consensus of the predictions obtained with PSIPRED, JPRED3 and 1D Protein Structure Prediction Server packages. Red arrow identifies the placement of the signal peptide cleavage. Secondary structure elements: α0-α2, predicted α-helices; β1-β6 predicted β-barrels.

Conclusion

There are two general conclusions that can be taken from the results obtained; the first concerns the proteomics and immunological aspects related to the urticating system while the second is oriented to clarify the biological aspects of the processionary moths of the genus *Thaumetopoea* and of other Thaumetopoeinae.

Proteomics and immunological aspects

My studies and results suggest that a large amount of proteins can be obtained from the urticating setae with a harsher extraction method that permitted to strongly break the setae. As many of the extracted proteins were not identified as allergens, I hypothesize that during the growth of the urticating setae they may become a sink for proteins occurring in the cytoplasm of the forming epidermal cells and perhaps also for blood proteins, which can easily enter the epidermis. The occurrence of allergenic proteins is confirmed by the positive blots of the sera of persons previously in contact with urticating setae, although at different level of exposure. In my experiment I was not able to reproduce the same results of Rodrigez-Mahillo (2012) for Tha p 2 protein, but component peptides of the same sequence were found in bands of different molecular weight, allowing to think to a family of urticating proteins, some of them still to identify. Also, the proteins of different molecular weight could open a new way to understand the immunologic component in the complex mechanism of reaction of the setae in humans. In addition to the allergic role of Tha p 2, the role of the chitin bound to these proteins should be explored, as it can be recognized as an allergen for tissue infiltration by innate cells implicated in allergic and immunity (Reese et al. 2007). The chitin action is not clear yet, although it can stimulate alone in vitro human T-lymphocyte proliferation, although at lower intensity compared with the urticating setae of *Th. pinivora* (Holm et al. 2014). In the latter study, a much stronger proliferation was induced in persons previously exposed to the setae, indicating that setae contain molecules which may start cell-mediated immune response (Holm et al. 2014). This fact may suggest that after skin penetration, allergens can be delivered to the immune system in a dual way: a fast release of allergens present on the outside of the setae, and a slow or very slow release that must occur after degradation of the setae by chitinases (Moneo et al. 2014).

In conclusion, the setae are considered a source of allergens and the risk for humans and animals is very high; they constitute a serious hazard, but the total

components, the quantity, the function and the real urticating proteins must be further investigated. This will give the possibility to create a diagnostic test not only for humans but also for animals. There is the need to repeat the immunologic tests with other persons who were exposed or not to the larvae to better understand the specificity of the response to the urticating setae. Finally, the occurrence of similar types of responses in the animals which are considered the natural target of the setae, like the insectivorous birds, should be explored.

Biology of Thaumetopoeinae and their urticating system

The characterization of *Tha p 2* gene and the study of its evolution, permitted to know that it is an intronless gene restricted to the processionary moths (Thaumetopoeinae, Notodontidae, Lepidoptera). Most of the studied species present two isoforms of Tha p 2 that can be interpreted as the result of heterozygosity in the single gene. The only exception is represented by *Thaumetopoea wilkinsoni*, in which 20 different isoforms occur in a single specimen, leading to the conclusion that at least in the species multiple copies of *Tha p 2* exist. The Tha p 2 is a glycine and cysteine rich protein very well conserved among processionary moths, included *Anaphe panda* and *Ochrogaster lunifer* that presents the urticating setae only in the moth.

The results of all this works permitted to clarify the real function of urticating setae as protective agents against predators, produced for this reason during the longest larval instars which are more exposed, and not during the first larval instars, as shown in the Chapter 3 in the study of the expression of the gene. The NGS approach (Illumina RNA-seq) used for sequencing the reads and the subsequent bioinformatics analyses, confirmed the expression of *Tha p 1* gene in all life stages of *Th. pityocampa* of both populations used for the experiment, indicating that it is not strictly associated with the urticating setae, while the gene *Tha p 2* was expressed only in larval stages where the setae production occur.

In addition, the analyses reveal that the urticating setae system and the gene associated, are well conserved in the Thaumetopoeinae group and could be a major factor affecting their evolution, as well as that of other groups of arthropods like spiders which share similar defense mechanisms.

A new objective following this results could be the extension of the extraction of the urticating setae to other species of processionary moths, to determine if the proteins

extracted from urticating setae are the same. It could be also of interest trying to understand the genetic expression of the urticating proteins in species like *Th. herculeana*, that was not considered dangerous in literature, but during my thesis I found that also *Th. herculeana* presents the same urticating system of the other *Thaumetopoea*, but not very efficient, because the mirrors do not open upon disturbance and the setae are not released. When applied to the human skin of a sensitized person, however, they produce the same reactions caused by the setae of the other *Thaumetopoea*. In this case, it will be interesting to understand the real expression of the gene of urticating protein, because of the *Thap 2* gene is also present. Finally, the study of the urticating system of processionary moths will require further development, including the analyses of other genes with specific urticating functions not only in Thaumetopoeinae but also in other groups of arthropods.

References

Holm G., Andersson M., Ekberg M., Fagrell B., Sjöberg J., Bottai M., Björkholm M., (2014). Setae from Larvae of the Northern Processionary Moth (*Thaumetopoea pinivora, TP*) Stimulate Proliferation of Human Blood Lymphocytes In Vitro. *PLoS ONE* 9(12): e113977.

Moneo I., Battisti A., Dufour B., García-Ortiz J.C., González-Munoz M., Moutou F., Paolucci P., Petrucco Toffolo E., Rivière J., Rodriguez-Mahillo A.I., Roques A., Roques L., Vega J.M., Vega J., (2014). Medical and Veterinary Impact of the Urticating Processionary Larvae. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp: 359-410.

Reese T. A., Liang H. E., Tager A., M Luster A. D., Van Rooijen N., Voehringer D., Locksley R. M., (2007). Chitin induces tissue accumulation of innate immune cells associated with allergy. *Nature.*447: 92-96.

Rodríguez-Mahillo A.I., González-Muñoz M., Vega J.M., López J.A., Yart A., Kerdelhué C., Camafeita E., Garcia Ortiz J.C., Vogel H., Toffolo E.P., Zovi D., Battisti A., Roques A., Moneo I., (2012). Setae from the pine processionary moth (Thaumetopoea pityocampa) contain several relevant allergens. *Contact Dermatitis* 67:367–374.

Supplementary Chapters

Chapter S1

Development time of *Thaumetopoea pityocampa* populations under laboratory conditions

---

Autors:

Berardi L., Branco M., Paiva M.-R., Santos H., Battisti A.

I contributed to experimental part, data analyses and paper writing.

Paper in preparation for submission.

Abstract

The pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera, Notodontidae) is a univoltine defoliator that is active over a wide range of latitude and elevation, being largely influenced by temperature variations especially during larval development. To compare developmental time in the field with laboratory rearing under controlled conditions, four populations characterized by different life history phenologies were selected and reared from egg to adult. The selection concerned two populations from the Italian Alps characterized by early and late adult emergence, as well as two populations from Portugal characterized, respectively, by summer and winter feeding of the larvae. The rearing was started from egg stage and maintained in the laboratory at 20-25°C under natural light in aerated boxes. The duration of each larval instar was calculated as the time taken by 50% of the larvae to molt to the next instar. In spite of the different geographic origins and asynchronous time of the larval development, all populations maintained an annual life cycle as well as a phenology very similar to that observed in the field. This outcome was possible due to a flexible duration of the pupal stage, indicating that this is the time when the mechanism maintaining the univoltine life cycle is working.

Introduction

Regulating phenology to be active and reproduce on the most favorable season is a common evolutionary mechanism in animals and plants (Tauber et al. 1986; Danks 1987). Many insect species, and phytophagous insects in particular, are univoltine and semelparous. The single reproductive event occurs normally in the season which allows optimizing survival for the offspring. Seasonality thus appears as a strong factor maintaining the univoltine cycle, and insects use environmental variables to track the right time of activity (Danks 1987; Saunders 1982). For herbivore insects, the univoltine life cycle is often explained by a precise synchronization with a particular developmental phase of the host plant. For example, several early-feeding herbivores of temperate forest ecosystems depend on the ingestion of young foliage, which is available during a restricted time window in the spring, while less specialized feeders are less constrained in time and can develop over longer periods or alternatively become multivoltine (Dajoz 1980). Evergreen coniferous host plants offer a good opportunity to test this hypothesis because they offer a relatively stable food supply over time, i.e. mature leaves. For example, in the diprionid sawflies, some species have a fixed univoltine cycle whereas

others may have univoltine cycles interchanged with multivoltine ones (Larsson et al. 1993). The latter species, when reared in the laboratory under optimal conditions of temperature and photoperiod, may develop continuous generations as in the case of *Diprion pini* (Eichhorn 1976).

Shifts in phenology, however, may occur in any life stage depending on changes of environmental variables. Recent studies show that the effects of day length on the induction and termination of diapause may be modified by rising temperatures (Saunders 1982; Battisti 2004). This happens because photoperiod is not affected by climate change whereas temperature is, especially in temperate and boreal zones and in winter more than in summer (Hodkinson 2005). Evidence may be obtained by demonstrating that conditions exceeding the current threshold of species survival will impose prohibitive mortality or preclude the completion of the life cycle.

The winter pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera, Notodontidae), provides a classic example of an insect feeding on mature foliage that does not need to be synchronized with the spring leaf flush. In addition, the larvae feed during winter, usually while in the third to fifth instar, and respond positively even to a slight increase of temperature (Battisti et al. 2005). As *Th. pityocampa* is univoltine, favorable conditions could promote faster development and thus the possibility to develop continuously. We know for example that embryonic development takes from 20 to 45 days, depending on the temperature recorded in the egg masses, which relates to the adsorption of solar radiation by the layers of protective scales (Milani 1990). Larvae are night feeders and gregarious, with a development time in the field varying between four and nine months (Battisti et al. 2014), depending on the winter conditions. The timing of adult emergence is generally inversely proportional to the length of the larval development, i.e. populations with early adults show the longest larval development, and vice versa (Battisti et al. 2014). Plasticity of phenology is typical among populations of this species, and it has been considered as an adaptive strategy to wide latitudinal and elevational gradients over witch its range is extended. Such a plasticity consists of a complementary duration of larval and pupal stages, for example long feeding period and short pupal time in the cooler areas, and viceversa (Huchon and Démolin 1970).

Interestingly, a population with shifted phenology has recently been found in Portugal, characterized by larval feeding in summer instead of winter. In this population, the adults emerge in late spring (from April to June) instead of in the summer (August-

September), as locally observed for the typical populations. The shifted population is genetically differentiated from the sympatric winter pine processionary moth (Santos et al. 2007; Santos et al. 2011a). Summer and winter populations co-occur in the same area, but are at great extent isolated by different reproductive periods (Santos et al. 2011a).

Considering the large phenotypic plasticity in the duration of the *Th. pityocampa* different stages of development, it would be important to determine how the insect develops under constant laboratory conditions. However, in the literature only two data sets are available. The first one relates to a population of the winter feeding *Th. pityocampa* from Kassandra peninsula in northern Greece (Devkota and Schmidt 1990). This population was taken at the egg stage to a laboratory in Germany (Hannover) and raised at 17-20°C. The emergence of the new adults was observed exactly one year later, although the larval development took between 86 and 109 days only. The second concerns to another Greek population (Kalogria, Schmidt 1989), which provided similar results. Although the pine processionary moth was reared by many entomologists under controlled conditions, no other data have been published about the total duration of larval development at constant temperature. One reason is probably linked to the difficulty to rear larvae which carry urticating hairs, implying the availability of dedicated rearing chambers (Battisti et al. 2011).

We therefore decided to collect data related to the duration of insect development under controlled conditions of four distinct populations of *Th. pityocampa*, two characterized by early versus late emergence in the field, from the Alps (Italy), and two characterized by summer versuswinter emergence, from Portugal. This approach allowed us to compare, among populations with apparent differentiation in phenology, the development time and mortality for each life stages. The general aim was to obtain background information to compare the result with the field development time and to explore the basis for adaptation to different climatic conditions.

Materials and methods

Two populations previously studied for variation of phenology (early and late) and population genetics were identified in north-eastern Italy (Simonato et al. 2012). They are located along an east-to-west gradient of egg size (Zovi et al. 2008), where the eastern populations, which also have early phenology, have about 10% larger eggs than western populations, which have late phenology. Therefore, there is a trade-off between egg number and egg size (Zovi et al. 2008). The eggs batches laid by moths of early population were collected on the needles of *Pinus nigra* in the south eastern Alps (Cimolais, 12°27'E 46°19'N, 650 m above sea level) (n = 12) in July 2012, while those of the late population were collected on the needles of the same pine species in the last week of August 2012 in the southern Prealps (Cinto Euganeo, 11°39'E 45°16'N, 200 m; Tregnago, 11°09'E 45°30'N, 450 m) (n = 16). Egg batches of both Portuguese populations (summer and winter population) were obtained by experimental mating conducted in the laboratory from 2008 to 2012, with the aim to analyze the phenology of both populations (Branco et al., in preparation). Late instar larvae or pupae of both populations were previously collected from maritime pine *Pinus pinaster* in the National Forest of Leiria (Portugal) (39º47' N 8º 58' W, 50 m). Larvae were allowed to pupate in the laboratory in boxes filled with sand. Pupae were separated by population and sex and kept at room temperature until emergence. Eggs batches were obtained for five consecutive years were n= 12, 6, 46, 46, 52 for the summer population and n = 4, 34, 45, 60, 26 for the winter population. Eggs batches were kept then in separate vials at room temperature (20-24ºC) until larvae hatched.

Larval rearing

Each egg batch was put in a plastic vial to wait for hatching in the laboratory under controlled conditions of temperature (20-25°C) and natural photoperiod, which was short day-length for the winter feeding populations and long day-length for the summer feeding population. After hatching, the larvae were transferred in aerated boxes (size cm 15x10x5?) with freshly collected pine twigs carrying needles. Then the egg batch was cleaned from the scales and the number of eggs and hatching holes were counted. Colonies originating from different egg batches were reared inseparate boxes. For the Italian populations, we decided to use *Pinus sylvestris* instead of *P. nigra* because the needles are shorter and easier to handle in the rearing boxes, while the larvae perform

equally well on the two hosts (Stastny et al. 2006). For the Portuguese populations, larvae were fed with fresh maritime pine *Pinus pinaster* needles, the main pine tree in Leiria pine forest. Fresh twigs were provided daily during the feeding period by gently removing the colony from the old twig, as well as the feaces and the feeding debris. On this occasion the developmental stage of the larvae was checked at the individual level, as well as the mortality. For the two Portuguese populations, mortality was recorded only at colony level. Since the third instar, the above checks were conducted under the hood because of the occurrence of urticating setae (Battisti et al. 2011). As the larvae reached the end of their development, the late fifth instar, they were transferred into a box (size cm 25x15x 5) with 4 cm of sterilized sand on the bottom, to offer a suitable site for pupation. Food was also added *ad libitum*. After about five weeks from pupation, the cocoons were sorted out from the sand and the pupae were extracted. Each pupa was then kept individually in a plastic vial in darkness under the same temperature condition. For the two Portuguese populations, pupae were kept in separate boxes, one for each family. The pupae were inspected once a week until the beginning of the emergence of the moths, when they were inspected daily. The mortality of the larvae in each instar of each colony was calculated as the total number of dead larvae in a given instar subtracted by the number of larvae that entered that instar. It was not possible to assess the cause of the death because symptoms were not specific to any factor known. The mean duration of each larval instar was calculated as the time until half the larvae completed the instar, excluding those that died during each specific period. In Portuguese populations the mortality at family level was evaluated during five years. Data was analyzed by logistic regression considering the proportion of dead families a dependent variable and considered the factors population and larval stage. The SD (standard deviation) and t-test were calculated for the development and the hatched eggs of each population, and a χ^2 test was used to test for population mortality.

Results

Development

The larvae of the Italian populations hatched in August in the early population, and in September in the late one. As expected, the egg batches from the early populations had fewer eggs per batch (225.2, SD 50.0, n 12) than those of the late population (251.7, SD 35.4, n 16), although the difference was not significant (t-test = 1.24, p = 0.23), and the number of hatched eggs was similar between the populations (57.2, SD 19.9, n 12, and 65.1, SD 36.3, n 16 for early and late population, respectively) (t-test = 0.57, p = 0.58). The larval development time did not differ significantly between the two populations (t test = 0.51, p = 0.69) (Table 1 and Fig. 1). There were differences between the populations in the duration of the individual instars, with the second shorter in the early and the third longer in the late population. All the surviving larvae moved into the sand for pupation between December and January (different from colonies). From early population were obtained 5 pupae and 7 from late, with an average of 1 pupa per colony. The pupal stage was shorter in the early population compared to the late one. The adults of the early population started to emerge at the beginning of June and those from the late population one month later.

Table 1. Duration of larval and pupal development of populations of *Thaumetopoea pityocampa* under laboratory conditions. Data from literature are added ([+]).

Population	**Larval development (days±SD)**	**Larval T°**	**Pupal development (days±SD)**	**Pupal T°**
Italy-early emergence	100.7±3.25	21.5	223±5.2	21.5
Italy-late emergence	102.5±4.42	21.5	229±0.0	21.5
Portugal-summer population	88.6±9.4	24.9	258±30.0	22.0
Portugal-winter population	100.4±3.9	22.6	234±27.1	22.0
[+]Greece-Kassandra Devkota and Schmidt (1990)	93.3±6.73	18.5	241	n.a.
[+]Greece-Kalogria Schmidt (1989)	>100	21.0	248	n.a.

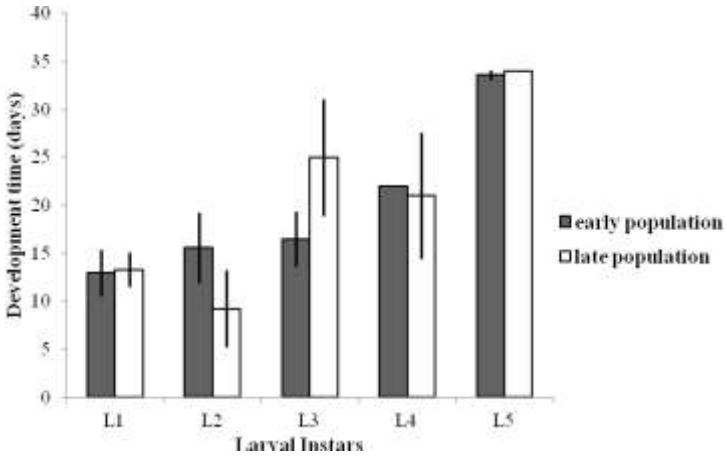


Figure 1. Duration of larval development of each instar of early and late populations of *Thaumetopoea pityocampa* from the Italian Alps. Bars indicate standard deviation. Numbers of colonies are given in the text.

From the 5-year monitoring in Portuguese populations, larval hatch occurred from May 8th to June 2nd for summer population and from August 3rd to September 10th for winter population. Thus, development began about three months earlier in the population with shift phenology than in the typical population. The larval development time of the early population was significantly shorter than that of the late population ($F_{1,53}$ = 39.071, $p < 0.001$) (Table 1 and Fig. 2). There were also differences between the populations in the duration of the individual instars: all the larval stages had shorter development in the summer population (Fig. 2). All the surviving larvae moved into the sand for pupation, in the summer population during the month of August and in the winter population between the second week of November and the end of December. In the following year, Emergences occurred from the first week of May to the end of June for summer population, and from the end of July to mid-August for winter population. Therefore the pupal stage was significantly longer in the summer population than in the winter population ($F_{1,36}$= 14.349, p=0.001) (Table 1) .

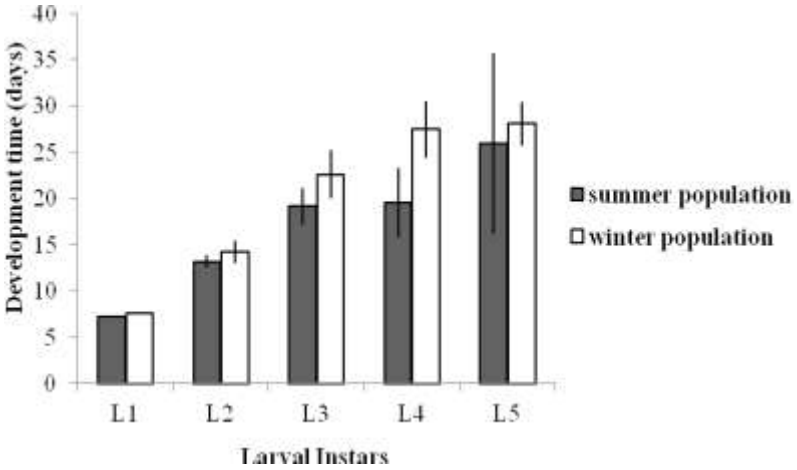


Figure 2. Duration of larval development of each instar of summer and winter populations of *Thaumetopoea pityocampa* from Portugal. Bars indicate standard deviation. Numbers of colonies are given in the text.

Mortality

Larval mortality was associated with the arrest of feeding and the desiccation of the larvae. Symptoms clearly associated with microbial infections were not observed. The colonies which reached the adult stage were 4 out of 12 and 4 out of 16 for the early and late Italian populations, respectively. The individual larval mortality was rather high for young larvae and then progressively declined toward the fourth and fifth instars (Fig. 3), with no significant differences between the populations (test $\chi^2 = 4.3$, $p = 0.63$). Colony disappearance because of total mortality varied strongly along the larval development, without a clear pattern. However, in the colonies which survived until the end, the overall individual mortality was only 5.6% for early and 11.3% for late population. Individual pupal mortality pooled among colonies was 50% for both populations.

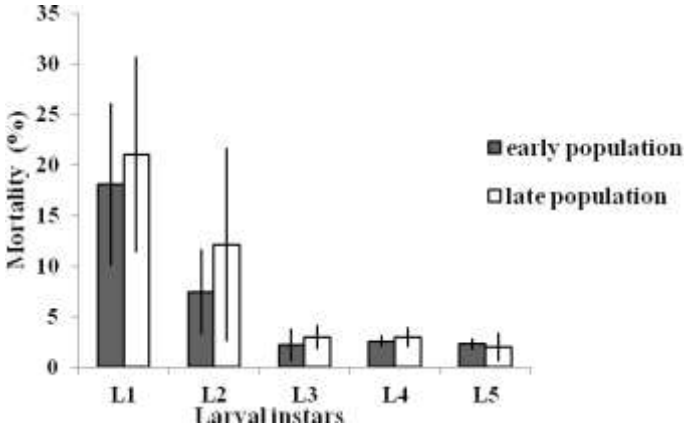


Figure 3. Mortality of larvae of early and late Italian populations of *Thaumetopoea pityocampa*. Bars indicate standard deviation (number of starting colonies: 12 for early and 16 for late populations).

In Portuguese populations the mortality was estimated at colony level only (Table 2). It was higher for the three first instars in the summer population while it was rather similar among larval instars in the winter population. No significant differences between populations were found (Wald $\chi^2_1 = 3.597$; $p = 0.058$), but colony mortality varied significantly with the larval instar (Wald $\chi^2_4 = 10.402$; $p = 0.034$). Overall, younger instars were the most affected ones in both populations. Pupal mortality at colony level was significantly higher for winter than for summer population ($F_{1,41}= 5.889$, $p < 0.01$) (Table 2).

Table 2. Colony mortality percentage (%) of larvae and pupae of the Portuguese populations of *Thaumetopoea pityocampa*. Mean is calculated for five years.

Instars	summer population (mean±SD)	winter population (mean±SD)
L1	32 ± 3.6	38 ± 3.4
L2	32 ± 4.4	38 ± 4.3
L3	42 ± 5.6	30 ± 5.2
L4	13 ± 5.1	24 ± 5.7
L5	18 ± 6.1	45 ± 7.7
Pupae	41 ± 6.0	69 ± 10.2

Discussion

The most striking result of this work is that the pine processionary moth is keeping its annual life cycle even under conditions conducive to continuous development. Raising larvae at room temperature of 20-25ºC, either under short or long day-length, resulted in a developmental time variable between 88 and 102 days only, while larval development under field conditions usually takes between 180 and 300 days for the winter feeding populations and around 90 days for the summer feeding populations (Battisti et al. 2014). The maintenance of the annual life cycle is therefore made possible by a flexible duration of the pupal stage, which under laboratory conditions was much longer than in the field for the winter feeding populations. In addition, the duration of the pupal stage can be much longer in the field in the case of prolonged diapause, while such an extension has never been reported under laboratory conditions in this species. Prolonged diapause is supposed to have evolved owing to enhancing geometric, rather than arithmetic mean survival of the population in fluctuating environments (Menu and Roebuck 2000). A constant environment in the laboratory does not let insects confront variability, thereby reducing possible risk, which in turn might ensure them of suitable environment in the current year. In addition, manipulative process of methodical laboratory experiment may disturb the energy status of insects encouraging them to flip their coin or emergence in the progressing year.

In the winter populations, the duration of the larval period has been shown to be strictly dependent on temperature, as the larvae are continuously feeding as long as the night temperature is above 0°C and the previous day temperature higher than 9°C inside the tent (Battisti et al. 2005), and they may cope with extended periods of starvation and repeated freezing across the winter (Hoch et al. 2009). Interestingly, the switching to the summer feeding observed recently in Portugal (Santos et al. 2007; Santos et al. 2011a) has resulted in developmental times of larvae and pupae which resemble those obtained in the laboratory at 20-25 °C for the winter populations, although both larvae and pupae development occur in different seasons. It seems that adult emergence time is regulated by factors which are relatively independent from temperature, while the duration of both adult and egg stages does not show major variation across all the populations (Battisti et al. 2014).

The laboratory rearing data confirm the general pattern described by Démolin (1969) and based on the collation of a large number of observations done across both

elevational and latitudinal gradients. According to this author, the warmer is the winter the shorter is the larval development, and the longer the pupal stage. Such a mechanism is basically dependent on the variability of the adult emergence, which can extend over several months depending on the location, as it has been confirmed in the survey of Pimentel et al. (2010). The larval development time of the six populations of *Th. pityocampa* for which data under laboratory conditions are available shows little variation at room temperature in spite of different geographic origins. Still, slight variations observed in the laboratory may be partially justified by differences on temperature. Although the correlation was not significant, developmental time tended to decrease with increasing temperature ($r = -0.27$, $p = 0.61$), consistent with the expected effect of temperature on larval development. In the field, the elevation and latitude drastically influenced the development time: at colder sites (high elevation or high latitude) the development time is longer. These results are concordant with the general predictions of temperatures-based models of insect development (Chapman 1998).

In particular, the L3, L4, and L5 instars are longer than the first two instars. The higher biomass of the later instars may justify their longer development time, as larvae attain 22 fold more biomass on L5 compare to that of L2 (Branco et al. 2008). For all larval stages, the development times on the summer population were shorter than those of the Portuguese winter population. This result might be justified by slight differences on temperature on the two periods (Table 2). Otherwise, we can hypothesize that the higher egg size found in the summer population compared to that of the winter population (Santos et al. 2013) may have an adaptive significance allowing lower mortality and shorter development time (Zovi et al. 2008). Nevertheless, for the early and late Italian populations such trend was not evident (Fig. 1). Variation of quantity or quality of the diet could also have effect on larval development; as low growth rates are associated with poor diets (Chapman 1988). This does not seem to be the case of the studied populations as they were raised on their preferred host in the field and regularly provided with fresh pine twigs.

The pupal development time shows much higher variation which does not seem to be explained by temperature. This physiological process is probably influenced by interaction of environmental, hormonal, and genetic signals (Tauber et al. 1986; Danks 1987, Denlinger 2002, Denlinger et al. 2005) that allow maintaining of the univoltine cycle with the right emergence time of the adults, like in the field, for each specific

population. A recent study demonstrated that adults under laboratory conditions emerged one month earlier than in the field (Zhang et al.1998), although the experiment was started with mature larvae ready to pupate and this fact may have disrupted the expected phenology. In the present study, starting rearing from eggs, we observed a good synchronization of the emergences times with those expected for the respective populations in the field.

The mortality percentage of *Th. pityocampa* larvae clearly decreased with increase of the larval stage. It was not possible to assess the cause of death because symptoms were not specific. Presumably it was caused by diseases or drying out, as the limitation of space in the boxes does not allow the colony to move from one branch to another as they do in nature. In the field, the principal causes of mortality are correlated to agents like viruses, bacteria, and fungi. The pupae in the soil are mostly attacked by fungi that are a key factor in causing decline in populations, especially when the stands are closed and the soil moisture is high (Markalas 1989, Battisti et al. 1998). However the individual larval mortality of the Italian populations of *Th. pityocampa* reflects the typical survivorship trend of decreasing with larval growth (Speight et al. 1999). In the field the young larvae of each population are affected by a strong effect of mortality, showing a direct influence from the daunting task of establishing themselves on a food plant (Zalucki et al. 2002). In laboratory, young larvae were found to have higher mortality when exposed to disturbances such as thermal stress (Santos et al, 2011b), possibly because their behavior was negatively affected by the absence of the tent (Battisti et al. 2014).

Development times under laboratory conditions are also available for other species of *Thaumetopoea*. The larvae of *Th. processionea* of a population from Romania were maintained at 20-22 °C and developed in 76-79 days, instead of 96-100 days observed in the field (Dissescu et al. 1968 in Battisti et al. 2014), although the annual life cycle was maintained. The larvae of *Th. jordana* developed faster in the laboratory than in the field and much faster at high temperature, taking 70 days at 20°C, 48 days at 25°C, and 40 days at 30°C (Furth and Halperin 1979), but also in this case there was no anticipation of the adult emergence. Studies done with a Spanish population of *Th. pinivora* (Galapagar, Madrid, Spain) (Montoya and Robredo 1972) indicate that development time is the same as observed in the field, although the extended diapause of the pupal stage is missing. All these data indicate that the processionary moths are well synchronized with the seasonality of their habitats despite thermal variations larvae might

experience, and this may result from a long co-evolution which has defined the development patterns. Yet, it is surprising that the pattern can be completely reversed, as in the case of the summer processionary moth of Portugal, while univoltinism is maintained. As this event is considered to be recent, as far as the genetic distance tells us (Santos et al. 2011 and 2013), we may expect a large flexibility in the responses to changed environmental conditions. The effect of temperature on larval development is quite clear and there is no more action to take in this concern, while the mechanisms regulating the pupal development need to be thoroughly considered as they could be the key to understand the phenology regulation in this group. In other words, the matter of diapause in this species needs to be taken into thorough consideration. We can first identify two different kinds of pupal diapause phenotypes, a) the obligatory diapause, of flexible duration, during which the "development time buffer", or synchronization of emergences, seems to happen, b) the extended diapauses, that rather seems to be involved in mitigating environmental stochasticity, but might also be related to synchronization in some cases (see the particular populations where extended diapause is mandatory, as normal pupal time would be too short for allowing the complete metamorphosis in the same year). In both cases though, the termination signal might be the same, thus ensuring synchronized emergences

Acknowledgements

We warmly thank Salman Habibur Rahman and Mathieu Laparie for useful comments on an earlier version of the manuscript.

References

Aimi A., Zocca A., Minerbi S., Hellrigi K., Gatto P., Battisti A., (2006). The outbreak of the pine processionary moth in Venosta/Vinschgau: ecological and economic aspects. *Forest observer* Vol.2/3 pp. 69-80.

Bale J.S., Masters G.J., Hodkinson I.D., Awmack C., Bezemer T.M., Brown V.K., Butterfield J., Buse A., Coulson J.C., Farrar J., Good J.E.G., Harrington R., Hartley S., Jones T.H., Lindroth R.L., Press M.C., Symrnioudis I., Watt A.D.,Whittaker J.B., (2002). Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biology* 8: 1-16.

Barbaro L., Battisti A., (2011). Birds as predators of the pine processionary moth (Lepidoptera: Notodontidae). *Biological Control* 56: 107-114.

Battisti A., (2004). Forests and climate change: lessons from insects. *iForest* 1: 17-24.

Battisti A., Holm G., Fagrell B., Larsson S., (2011). Urticating hairs in arthropods: Their nature and medical significance. *Annual Review of Entomology* 56: 203-220.

Battisti A., Longo R., Tiberi R., Triggiani O., (1998). Results and perspective in the use of *Bacillus thuringiensis* Berl. var. *kurstaki* and other pathogens against *Thaumetopoea pityocampa* (Den. et Schiff.) in Italy (Lep., Thaumetopoeidae). *Anzeiger für Schadlingskde, Pflanzenschutz, Umweltschutz* 71: 72-76

Battisti A., Stastny, M., Netherer, S., Robinet, C., Schopf, A., Roques, A., Larsson, S., (2005). Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecological Applications* 15: 2084–2096.

Battisti A., Avci M., Avtzis D., Ben Jamaa M., Berardi L., Berretima W., Branco M., Chakali G., El Alaoui M., Frérot B., Hodar J., Ionescu-Malancus I., Ipekdal K., Larsson S., Manole T., Mendel Z., Mirchev P., Nemer N., Paiva M-R., Pino J., Protasov A., Rahim N., Rousselet J., Santos H., Sauvard D., Schopf A., Simonato M., Yart A., Zamoum M., (2014). Natural history of the processionary moths (*Thaumetopoea* spp.): new insights in relation to climate change. *In: Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht, pp. 15-80.

Berggren A., Bjorkman C., Bylund H., Ayres M.P., (2009). The distribution and abundance of animal populations in a climate of uncertainty. *Oikos* 118: 1121-1126.

Branco M., Santos M., Calvão T., Telfer G., Paiva M. R., (2008). Arthropod diversity sheltered in Thaumetopoea pityocampa (Lepidoptera: Notodontidae) larval nests. *Insect Conservation and Diversity*, 1(4): 215-221.

Chapman R. F., (1998). *The insects. Structure and function*. Cambridge University Press, Cambridge, UK.

Danks H.V., (1987). Insect Dormancy: An Ecological Perspective. Ottawa: *Biological Survey of Canada (Terrestrial Arthropods)* pp. 439.

Dajoz R., (1980). Ecologie des insectes forestiers. *Bordas*, Paris, France.

Démolin G., (1969). Bioecologia de la procesionaria del pino *Thaumetopoea pityocampa* Schiff. Incidencia de los factores climaticos. *Boletin del Servicio de Plagas Forestales* 12: 9-24.

Denlinger D.L., Yocum G.D., Rinehart J.P.. (2005). Hormonal control of diapause. In Gilbert, K. Iatrou, S.S. Gill (Eds.), *Comprehensive Molecular Insect Science*, vol. 3 Elsevier Press, Amsterdam pp. 615–650.

Denlinger, D. L. (2002). Regulation of diapause. *Annual Review of Entomology* 47: 93-122.

Devkota B. and Schmidt G.H. (1990). Larval development of *Thaumetopoea pityocampa* (Den. & Schiff) (Lep., Thaumetopoeidae) from Greece as influenced by different host plants under laboratory conditions. *Journal Apply of Entomology* 109: 321-330.

Douma-Petridou, E., (1989). European *Thaumetopoea* species (Lep., Thaumetopoeidae): characteristics and life-cycles. In: Schmidt, G.H. (Ed.), Proc. *Thaumetopoea* Simp., Neustadt/RBGE., Department of Zoology–Entomology, University of Hannover (FRG), pp. 12–19.

Eichhorn O. (1976). Permanent culturing of *Diprion pini* L. with regard to photoperiodic conditions. *Anzeiger für Schadlingskde., Pflanzenschutz, Umweltschutz* 49: 38-41.

Fitzgerald T.D., Panades-Blas X., (2003). Mid-winter foraging of colonies of the pine processionary caterpillar *Thaumetopoea pityocampa* Sciff. (Thaumetopoeidae).The Lepidopterist Society 57: 161-167.

Furth D.G. and Halperin J. (1979). Observations of the phenology and biogeography of *Thaumetopoea jordana* (Staudinger) (Lepidoptera: Thaumetopoeidae). *Israel Jurnal of Entomology* 13: 1-11

Hodkinson I.D. (2005). Terrestrial insects along elevation gradients : species and community responses to altitude. *Biological Review* 80: 489-513.

Huchon H., Démolin G., (1970). La bioécologie de la processionaire du pin. Dispersion potentielle, dispersion actuelle. *Revue Forestière Francaise*, nombre spèciale « La lutte biologique en foret », pp 220-234.

Larsson, S.; Björkman, C., Kidd N. A. C., (1993). Outbreaks in diprionid sawflies: why some species and not others? In: *Sawfly life history adaptations to woody plants* (eds Wagner, M. R.;Raffa, K. F.) pp. 453-483.

Markalas S., (1989). Influence of soil moisture on the mortality, fecundity and diapause of the pine processionary moth (*Thaumetopoea pityocampa* Schiff.). *Journal of Apply Entomology* 107: 211-215

Menu F., Roebuck J. P. (2000). Bet-hedging diapause strategy in stochastic environment. *The American Naturalist* 155: 234–724.

Milani N., (1990). The temperature of the egg masses of *Thaumetopea pityocampa* (Den. & Schiff.) (Lepidoptera, Thaumetopeidae). *Redia* Vol. 73 (1): 149-161.

Montoya R., Robredo F., (1972). Thaumetopoea pinivora, Tr. "La processionaria de verano" *Boletin Estacion Central Ecologia* 1: 43-56.

Netherer S., Schopf A., (2010). Potential effects of climate change on insect herbivores in European forests-General aspects and pine processionary moth as specific example. *Forest Ecology and Management* 259: 831-838.

OEPP/EPPO (2004). Diagnostic protocols for regulated pests *Thaumetopoea pityocampa*. *Bulletin OEPP/EPPO* 34: 155-157.

Pérez-Contreras T., Soler J.J., (2004). Egg parasitoides select for large clutch sizes and covering layers in pine processionary moths (*Thaumetopoea pityocampa*). *Annales Zoologici Fennici* 41: 587-579.

Powell J.A., (1974). "Occurrence of prolonged diapause in ethmiid moths." *Pan-Pacific Entomologist* 50: 220-225.

Santos H., Burban C., Rousselet J., Rossi J.-P., Branco M., Kerdelhué C., (2011a). Incipient allochronic speciation in the pine processionary moth (*Thaumetopoea pityocampa*, Lepidoptera, Notodontidae). *Journal of Evolutionary Biology* 24*:* 146–158.

Santos H., Paiva M.R., Tavares C., Kerdelhue C., Branco M., (2011b). Temperature niche shift observed in a Lepidoptera population under allochronic divergence. *Journal of Evolutionary Biology* 24: 1897–1905.

Santos H.M., Paiva M.R., Rocha S., Kerdelhuè C., Branco M., (2013). Phenotypic divergence in reproductive traits of a moth population experiencing a phonological shift. *Ecology and Evolution* 3: 5098-5108.

Santos H., Rousselet J., Magnoux E., Paiva M.R., Branco M., Kerdelhué C., (2007). Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceeding of the Royal Society London* 274: 935–941.

Saunders D.S., (1982). *Insect Clocks*. Oxford: Pergamon.

Schmidt G.H., (1989). Life cycle of *Thaumetopea* species distributed in different regions of Europe, north Africa and near east. *Proc. Thaumetopoea-Symp*., Neustad/Rbge, pp.20-34.

Simonato M., Battisti A., Zovi D., Medina R.F., (2012). Testing for host-associated differentiation in two egg parasitoids of a forest herbivore. *Entomologia Experimentalis et Applicata* 145: 124-133.

Speight M.R., Hunter M.D., Watt A.D., (1999). *Ecology of Insects: Concepts and Applications.* Blackwell Science Ltd, Oxford.

Stastny M., Battisti A., Petrucco-Toffolo E., Schlyter F., Larsson S., (2006). Host-plant use in the range expansion of the pine processionary moth, *Thaumetopoea pityocampa*. *Ecological Entomology* 31: 481–490.

Tauber M. J., Tauber C. A., Masaki S., O'Riordan V. B., Burnell A. M., (1989). Seasonal Adaptation of Insects (Oxford Univ. Press, New York, 1986); *Comparative Biochemistry and Physiology* 92B: 233.

Zalucki M.P., Clarke A.R., Malcom S.B., (2002). Ecology and behaviour of first instar larval Lepidoptera. *Annual Review of Entomology* 47: 361-93.

Zhang Q.H., Paiva M.R. (1998). Female calling behavior and mal e response to the sex pheromone in *Thaumetopoea pityocampa* (Den. et Schiff.) (Lep. Thaumetopoeidae). *Journal of Applied Entomology* 122: 353-360.

Zovi D., Stastny M., Battisti A., Larsson S., (2008). Ecological costs on local adaptation of an insect herbivore imposed by host plants and enemies. *Ecology* 89: 1388–1398.

Chapter S2

Contribution published in a book'

Alain Roques
Editor

Processionary Moths and Climate Change: An Update

Springer

Editor

Alain Roques
Zoologie Forestie`re
INRA, UR 0633
Orleans, France

ISBN 978-94-017-9339-1 ISBN 978-94-017-9340-7 (eBook)
DOI 10.1007/978-94-017-9340-7
Springer Dordrecht Heidelberg New York London
Library of Congress Control Number: 2014951874
Editions Quæ, R10, 78026 Versailles cedex, France www.quae.com

Natural History of the Processionary Moths (Thaumetopoea spp.): New Insights in Relation to Climate Change

Andrea Battisti, Mustafa Avcı, Dimitrios N. Avtzis, Mohamed L. Ben Jamaa, Laura Berardi, Wahiba Berretima, Manuela Branco, Gahdab Chakali, Moulay Ahmed El Alaoui El Fels, Brigitte Fre´rot, Jose´ A. Ho´dar, Irina Ionescu-Ma˘la˘ncus¸, Kahraman I ˙ pekdal, Stig Larsson, Traian Manole, Zvi Mendel, Nicolas Meurisse, Plamen Mirchev, Nabil Nemer, Maria-Rosa Paiva, Juan Pino, Alex Protasov, Noureddine Rahim, Je´roˆme Rousselet, Helena Santos, Daniel Sauvard, Axel Schopf, Mauro Simonato, Annie Yart, and Mohamed Zamoum

A. Battisti (*) • L. Berardi • M. Simonato
Department DAFNAE-Entomology, University of Padua, Agripolis,
35020 Legnaro, Italy
e-mail: andrea.battisti@unipd.it; laura.berardi@studenti.unipd.it; mauro.simonato@unipd.it

M. Avcı
Orman Faku¨ltesi, Su¨leyman Demirel U¨ niversitesi, Isparta, Turkey
e-mail: mustafaavci@sdu.edu.tr

D.N. Avtzis
Forest Research Institute, Hellenic Agricultural Organization “Demeter”, Vassilika,
Thessaloniki, Greece
e-mail: dimitrios.avtzis@fri.gr

M.L.B. Jamaa
Gestion et Valorisation des Ressources Forestie`res, Institut National de Recherche en Ge´nie Rural, Eaux et Foreˆts (INRGREF), Ariana, Tunisia
e-mail: benjamaa.lahbib@iresa.agrinet.tn

W. Berretima • G. Chakali • N. Rahim
De´partement de Zoologie Agricole et Forestie`re, Ecole Nationale Supe´rieure Agronomique,
El Harrach, Algiers, Algeria
e-mail: wahibaberr@yahoo.fr; chakali_gahdab@yahoo.fr; rahim.noureddine@hotmail.com

M. Branco
Centro de Estudos Florestais (CEF), Instituto Superior de Agronomia (ISA),
University of Lisbon, Lisbon, Portugal
e-mail: mrbranco@isa.ulisboa.pt

A. Roques (ed.), Processionary Moths and Climate Change: An Update,
DOI 10.1007/978-94-017-9340-7_2

Natural History of Thaumetopoea herculeana

Mauro Simonato, Laura Berardi, Andrea Battisti, and Juan Pino

Host Plants and General Distribution

Thaumetopoea herculeana is associated with open areas covered with low natural vegetation of truly Mediterranean type of the Iberian Peninsula and on the southern rim of the Mediterranean basin until the Near East. The larvae feed on Geraniaceae (Erodium moschatum and E. arborescens), and Cistaceae (Helianthemum vulgare, H. croceum and Cystus salviaefolius), which are not crop plants.

Life Cycle

There is large uncertainty about the life cycle of this species. According to Go´mez-Bustillo (1979), the moth's flight period extends from June in the northern and central Spain to September/October in Andalusia. Agenjo (1941) reported moth catches in May, August, September, and October. Larvae are reported to be present from October to April (Gomez de Aizpura 1986). During our small survey in coastal NW Spain, larvae of different instars (second to fifth) were found in April and gave origin to adults emerged in July under laboratory conditions, while in the field moths have been observed between July and September (unpublished data). It is thus possible that the life cycle depends on local weather patterns and host plant phenology. Moths lay eggs on low plants, grouping them in cylindrical batches. Larvae show a gregarious behaviour as they stay in groups of about 20–30 individuals piled up on the ground, forming a small cluster without silk (Fig. 2.4). A silk protection is reported just in the northern and central part of Spain, but not in Andalusia (Go´mez-Bustillo 1978). In the sunny days they move in processions with a triangular shape, separating to feed on the low vegetation around, and regrouping again after it. They usually feed during the night and in the morning.

Larvae have a bluish grey appearance, and are covered by hairs of the same colour during the first two instars; in the next larval instars yellow-green tufts interspersed with grey appear on the back until the last moulting, probably to mimic the surrounding vegetation (for example Ulex europaeus). Larvae pupate in the ground at a low depth between March and April in the south of Spain. Larvae were not reported to be irritant (Agenjo 1941) although they carry urticating setae as in the other species of the genus (unpublished data).

Natural Enemies

The larvae are parasitized by unknown tachinids and preyed by the carabid beetle Macrocarabus lusitanicus (Pino, unpublished data from north-western Spain).

Population Dynamics

No information is available.

Relationships with Climate Change

No information is available

Climate Warming and Past and Present Distribution of the Processionary Moths (Thaumetopoea spp.) in Europe, Asia Minor and North Africa

Alain Roques, Je´roˆme Rousselet, Mustafa Avcı, Dimitrios N. Avtzis, Andrea Basso, Andrea Battisti, Mohamed Lahbib Ben Jamaa, Atia Bensidi, Laura Berardi, Wahiba Berretima, Manuela Branco, Gahdab Chakali, Ejup C¸ ota, Mirza Dautbasˇic´, Horst Delb, Moulay Ahmed El Alaoui El Fels, Sarı¨d El Mercht, Mhamed El Mokhefi, Beat Forster, Jacques Garcia, Georgi Georgiev, Milka M. Glavendekic´, Francis Goussard, Paula Halbig, Lars Henke, Rodolfo Hernan´dez, Jose´ A. Ho´dar, Kahraman I ˙ pekdal, Maja Jurc, Dietrich Klimetzek, Mathieu Laparie, Stig Larsson,
Eduardo Mateus, Dinka Matosˇevic´, Franz Meier, Zvi Mendel, Nicolas Meurisse, Ljubodrag Mihajlovic´, Plamen Mirchev, Sterja Nasceski, Cynthia Nussbaumer, Maria-Rosa Paiva, Irena Papazova, Juan Pino, Jan Podlesnik, Jean Poirot, Alex Protasov, Noureddine Rahim, Gerardo San´chez Penˇ a, Helena Santos, Daniel Sauvard, Axel Schopf, Mauro Simonato, Georgi Tsankov, Eiko Wagenhoff, Annie Yart, Regino Zamora, Mohamed Zamoum, and Christelle Robinet

A. Roques (*) • J. Rousselet • J. Garcia • F. Goussard • M. Laparie •
D. Sauvard • A. Yart • C. Robinet
INRA, UR0633 URZF (Unite´ de Recherche Zoologie Forestie`re), Orle´ans, France
e-mail: alain.roques@orleans.inra.fr; Jerome.Rousselet@orleans.inra.fr;
jacques.garcia@orleans.inra.fr; francis.goussard@orleans.inra.fr; mathieu.laparie@orleans.inra.fr;
daniel.sauvard@orleans.inra.fr; annie.yart@orleans.inra.fr; christelle.robinet@orleans.inra.fr

M. Avcı
Orman Faku¨ltesi, Su¨leyman Demirel U¨ niversitesi, Isparta, Turkey
e-mail: mustafaavci@sdu.edu.tr

D.N. Avtzis
Forest Research Institute, Hellenic Agricultural Organization “Demeter”, Vassilika,
Thessaloniki, Greece
e-mail: dimitrios.avtzis@fri.gr

A. Basso • A. Battisti • L. Berardi • M. Simonato
Department DAFNAE-Entomology, University of Padua, Agripolis, 35020 Legnaro, Italy
e-mail: andrea.basso@studenti.unipd.it; andrea.battisti@unipd.it; laura.berardi@studenti.
unipd.it; mauro.simonato@unipd.it

A. Roques (ed.), Processionary Moths and Climate Change: An Update,
DOI 10.1007/978-94-017-9340-7_3

Thaumetopoea herculeana

Mauro Simonato, Laura Berardi, Juan Pino, and Andrea Battisti

The moth has an almost continuous distribution in the Iberian peninsula, Spain and Canary Islands and Portugal (Agenjo 1941; Go´mez-Bustillo 1979; Gomez de Aizpurua 1986; Bacallado Ara´nega and Herna´ndez Pacheco 1990) (Fig. 3.41). Isolated findings exist from Morocco, Algeria, Tunisia, Libya Cyrenaica, and Palestine, although most of the material is old (Agenjo 1941) and would need
reconfirmation.

Fig. 3.41 Distribution map of Thaumetopoea herculeana in the Iberian Peninsula according to Go´mez-Bustillo (1979)

Acknowledgments

During this PhD experience I had the pleasure to meet and work with several person and all of them, contributed to my professional and moral growth, and is now a pleasure for me to thank them all.

Firstly I would like to thank my supervisors Prof. Andrea Battisti and Prof. Enrico Negrisolo, who assisted and supported me for the three years of this experience. A particular thanks are due to my tutors in France, Carole Kerdelué, for very kind hospitality and for the help in my works, Frank Dorkeld, Bernhard Gschloessl and my special office colleague Julie Pisano.

Special thanks go to my friendly and kindly office colleagues: Davide, Diego and Giovanni and to my friends of department: Andrea, Diego, Edoardo, Ewelina, Fernanda, Giacomo C., Giacomo S., Ines, Isabel, Lorenzo T., Mauro, Micaela, Paola, Riccardo, Salman. I want also to thank my colleagues of the department of BCA, Eleonora, Lisa, Massimiliano, Massimo, Marianna, Nadia, Rafaella, Roberta, Sara, Serena. Thanks also to the entomology staff at the department, Gabriella Frigimelica, Luca Mazzon, Lorenzo Marini and Nicola Mori and Professors Carlo Duso, Giuseppina Pellizzari, and technicians Patrizia Dall'Ara, Paolo Paolucci.

I have to say a special "thank you" to Nicola, my family for the moral support and all my friends.