UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF LAND, ENVIRONMENT AGRICULTURE AND FORESTRY

DEPARTMENT OF AGRONOMY, FOOD, NATURAL RESOURCES, ANIMALS AND ENVIRONMENT

MASTER THESIS IN

FOREST AND ENVIRONMENTAL SCIENCES

PROTEINS ASSOCIATED WITH THE URTICATING SETAE OF THE PINE PROCESSIONARY MOTH *Thaumetopoea pityocampa* (Denis & Schiffermüller 1775)

Supervisor:

Prof. Andrea Battisti

Co-supervisor:

Dr. Laura Berardi

Candidate:

Ettore Mitali

Enrollment n. 1060875

Academic Year

2014 – 2015

To my mother Vanda
who, alone, has raised two children thanks to her infinite love.
Despite some difficulties, she has demonstrated that the affection in all its expression, is the winning key to the achievement of many objectives, including this work.

A mia madre Vanda
che, da sola, ha cresciuto due figli grazie al suo amore infinito.
Nonostante alcune difficoltà, ha dimostrato che l'affetto, in ogni sua espressione, è la chiave vincente per il raggiungimento di molti obiettivi, incluso il presente lavoro.

Contents

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 foresters working in infested pine stands, who are exposed to high levels of setae, but also in persons non-occupationally exposed to processionary larvae, such as persons living near infested areas and visitors. 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 we analyzed the protein content through the mass spectrometer and bioinformatics analyses. And a total of 182 urticating and non-urticating proteins were obtained. We confirm that the setae of *Th. pityocampa* contain many proteins, in addition, we add information about the type, quality, and quantity of the proteins associated with the setae.

Riassunto

Indagine sulle proteine associate alle setole urticanti della processionaria del pino *Thaumetopoea pityocampa*

Le larve di processionaria pino (*Thaumetopoea pityocampa*) producono delle setole urticanti utilizzate per la difesa nei confronti di alcuni nemici naturali quali i vertebrati predatori. Il contatto con le setole nell'uomo e in altri animali induce dermatiti, specialmente nelle parti del corpo maggiormente esposte. Solitamente le persone a maggiore rischio sono gli operatori forestali durante la loro attività in pinete infestate sia montane sia mediterranee. Tuttavia anche i soggetti non professionalmente esposti a larve di processionaria, ad esempio gli agricoltori che vivono nelle vicinanze di aree infestate e i visitatori occasionali, hanno possibilità di contatto con le setole, poiché esse si disperdono facilmente nell'ambiente grazie al vento. Recenti studi hanno dimostrato la presenza di un complesso meccanismo urticante dove certe proteine presenti nel setole possono diventare attivatori di risposte immunitarie. Poiché un data set completo di tutte le proteine non è ancora disponibile in letteratura, in questo lavoro sono stati sviluppati due diversi protocolli di estrazione di proteine e analizzati accuratamente i risultati ottenuti tramite spettrometria di massa, che ci hanno permesso di identificare un totale di 182 proteine urticanti e non-urticanti. I risultati confermano che le setole di *Th. pityocampa* contengono abbondanti quantità di proteine, di cui una parte significativa appartiene a quelle riconosciute dal sistema immunitario umano.

1. Introduction

1.1 The processionary moths

The processionary moths (Lepidoptera, Notodontidae, Thaumetopoeinae) (Fig. 1) are organisms of interest and subject of research. These insects cause severe irritation in humans and other animals, and is demonstrate that are a damage also for the vegetation. During their outbreaks they inhibit the development of trees by defoliating them and this require concrete actions in the forestry context. However, since larvae carry urticating hairs, concrete actions are especially required in the urban context, where there is a high exposure and therefore the risk factors are higher. In addition, they are very good indicators of climate change.

Figure 1. Pine processionary larvae *Thaumetopoea pityocampa* (photo Anna Nicholas).

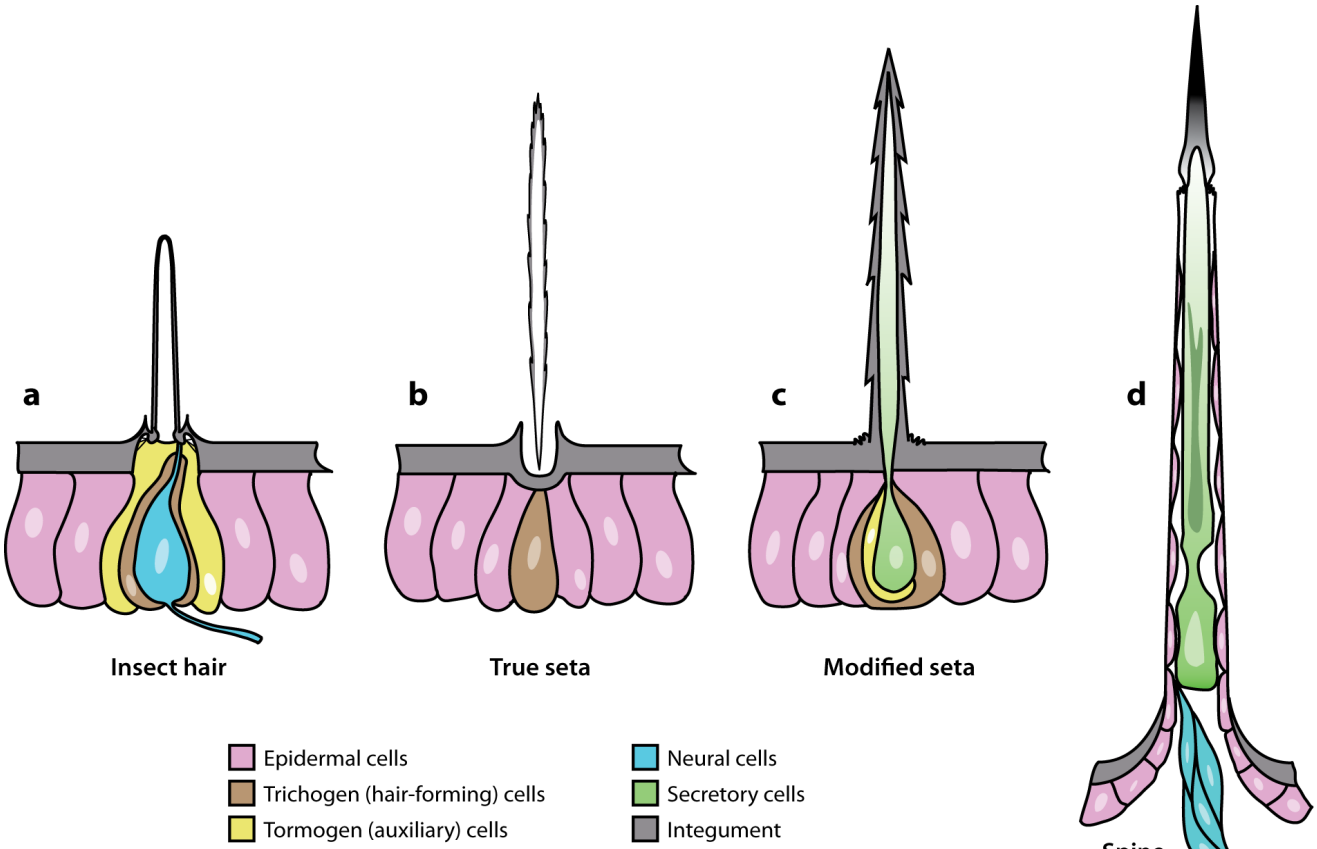
1.2 The urticating setae of the processionary moths

Urticating setae are stiff hair like structures placed on the dorsal part of the abdomen of the insect integument and are considered a defense against vertebrate and invertebrate predators. Incidentally, these setae are also a serious threat to human health when they get in contact with the skin or other parts of the body (Battisti et al.2011).

To be precise, processionary larvae carry only “true setae”, which differ from a second category of urticating setae called “modified setae” carried on other Lepidoptera (Battisti

et al. 2011) (Fig. 2). Nevertheless the nature of these two setae is very different from other defensive hairs, such as spines (Fig. 2). Spines 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 setae can be released easily into the environment, creating health concerns.

In general, Urticating setae are formed by at least two cells [trichogen (or hair-forming cell) and tormogen (or auxiliary cell)] embedded in the epidermal cells and is connected to one or more neurons for the transmission of sensorial information (Fig. 2) (Battisti et al. 2011). Urticating setae, like the insect integument, are built up by a chitin skeleton with a matrix of proteins and are covered by layers of tannin-bound lipoproteins, wax, and mucopolysaccharides (Battisti et al. 2011).



Battisti A, et al. 2011.
Annu. Rev. Entomol. 56:203–20

Figure 2. Schematic representation of (a) an insect hair, (b) a true seta, (c) a modified seta, and (d) a spine.

1.2.1 True setae description

Urticating setae are modified by the loss of the neural connection and the detachment of the proximal end of the hair from the integument (Battisti et al. 2011) (Fig. 2). The sharp basal end of each seta is loosely inserted into a socket and it can be easily removed with any kind of mechanical stimulation.

The setae are short (generally 50–600 µm long, 2–8 µm in diameter), have barbs along the shaft and can easily enter in to the skin at the proximal end, helped by the barbs (Petrucco Toffolo et al. 2014).

In the processionary moths the urticating setae appear during larval development, in particular in *Thaumetopoea pityocampa* from the third to the fifth instars. Early lepidopteran instars are without urticating hairs, although the cellular apparatus that produces them is present.

The setae are packed close together on the dorsal part of the abdomen in specific areas called mirrors (Battisti et al. 2011) (Fig. 5 A). The mirrors increase in number and size as the larva molts, and the number of setae increases as well (Battisti et al. 2011).

The density of setae can be very high. In mature larvae (5^{th} instar) a full set of mirrors consists of several hundred thousand setae, e.g., 630,000 in *Thaumetopoea processionea* and up to 1,000,000 in *Thaumetopoea pityocampa* (60,000 setae/mm^2) (Petrucco Toffolo et al. 2014).

In processionary moths, the larval exuvia left after the molt may carry the old setae that were not dispersed during the previous larval instar (Battisti et al. 2011). Similarly, exuviae left inside the cocoon at pupation are covered with numerous old setae.

All studied species of the Thaumetopoeinae are known to carry urticating setae, either as larva (genus *Thaumetopoea*) or adult (e.g. the African genus *Anaphe* and the Australian *Ochrogaster*) (Battisti et al. 2011).

The urticating setae are a distinct feature of the larval stage of processionary (Thaumetopoeinae, Notodontidae) and tussock moths (Lymantriidae) and occur in the adult stage of a few species [e.g., the African *Anaphe* spp. (Notodontidae) and *Euproctis* spp. (Lymantriidae)]. A few species of Saturniidae in South America and Zygaenidae in Australia carry setae only as adults. True setae are similar to urticating setae released by some tarantula spiders from America of the family Theraphosidae (Table 1).

Table 1. Urticating hair types in relation to taxonomy, distribution, life-history traits, and previous classification types [modified from Battisti et al. (2011)].

Hair type	Group	Family	Distribution
Seta	Spiders	Theraphosidae	America
	Lepidoptera larvae	Lymantriidae	Cosmopolitan
		Notodontidae	
	Lepidoptera moths	Lymantriidae	Cosmopolitan
		Notodontidae	Africa
		Saturniidae	America
		Zygaenidae	Australia
Modified seta	Lepidoptera larvae	Zygaenidae	Cosmopolitan
		Limacodidae	Cosmopolitan
		Nolidae	Australia
		Arctiidae	Cosmopolitan
		Anthelidae	Indo-Australia
		Eupterotidae	Africa, Australia
		Lasiocampidae	Cosmopolitan
		Lymantriidae	Cosmopolitan
Spine	Lepidoptera larvae	Saturniidae	America
		Noctuidae	Cosmopolitan
		Megalopygidae	America
		Limacodidae	Cosmopolitan
		Nymphalidae	

1.2.2 Ecological role of true setae

The ecological role of setae in protection from predators can be discussed in relation to what is known about the defense mechanism. Moneo et al. (2015) point out that the 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. As setae disperse as a cloud around the colony their function could be to keep away predators. In this case, the larger and denser is the cloud, the stronger is the protection; the diversity of seta size may extend such a barrier much farther, with a direct benefit for the colony (Battisti et al. 2011). Other prey of vertebrate predators could indirectly benefit from the protection, and, thus, competition among insect herbivores may increase but the large investment in urticating setae made by these species of processionary moths indicates that the benefits from extended protection are higher than the costs possibly imposed by competition (Petrucco Toffolo et al. 2014). Anyhow the mechanism needs to be elucidated with appropriate experiments.

1.2.3 Size and dispersion dynamics of true setae

The release mechanism of setae by the larvae was firstly explored by Démolin (1963), who showed that the larvae may actively open the integument mirrors when disturbed (Figure 4). In *Th. pityocampa*, the mirrors (Fig. 3) are kept folded under normal conditions and only the distal end of the setae is visible. When disturbed, the larva opens the mirror (Fig. 5 A), releasing the setae, a process that is further facilitated by the action of a few normal hairs that are mixed with the setae in the mirror. Once in the air, the setae can be carried by the wind far from the source (Battisti et al. 2011).

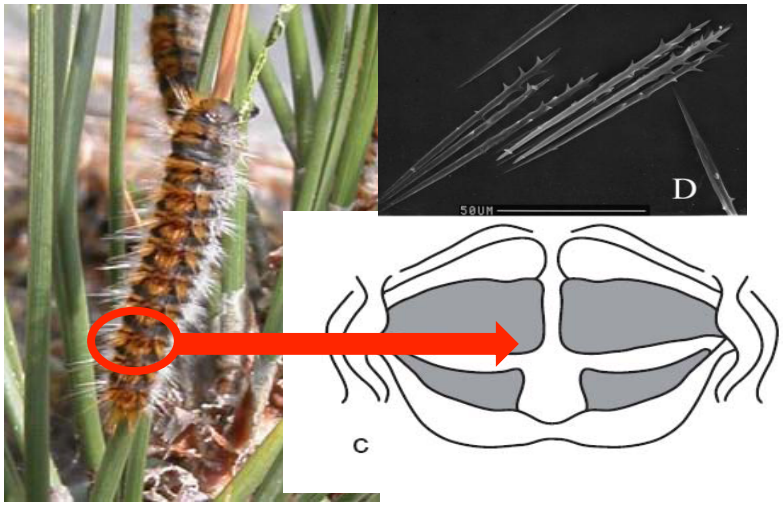


Figure 3. Mirror of a pine processionary larva (from Moneo et al. 2015).

The dynamic properties of urticating setae of the pine processionary moth *Thaumetopoea pityocampa*, the northern pine processionary moth *Thaumetopoea pinivora* and the oak processionary moth *Thaumetopoea processionea*, have been well described in Petrucco Toffolo et al. 2014. The results showed a wide variation in seta length (Fig. 5 B). In the case of *Th. pityocampa*, the longest (680 μm) were approximately 14 times longer than the shortest (50 μm), whereas in *Th. pinivora* (47– 492 μm) and in *Th. processionea* (56 –351 μm) the same ratios were equal to 10 and 6 times, respectively. The short and long setae are intermixed throughout the mirror.

The hypothetical horizontal distance traveled for a seta released at 20 m height in a day with a wind velocity of 2 m/s is 6.5 km for the short setae and 2.4 km for the long setae. The distribution of the length of *Th. pinivora* and the corresponding dispersion distances are 21 and 7.4 km. In *Th. processionea*, the distribution of length resulting in a dispersion of 8 km for a release at 20 m of height and a wind velocity of 2 m/s. It must be mentioned that the velocities and distances given above are for the mean aerodynamic diameter. Because the velocity is inversely proportional to the square aerodynamic diameter, the smaller setae will spread much further. In the studied species of *Thaumetopoea*, the general shape of the seta is the same (Moneo et al. 2015).

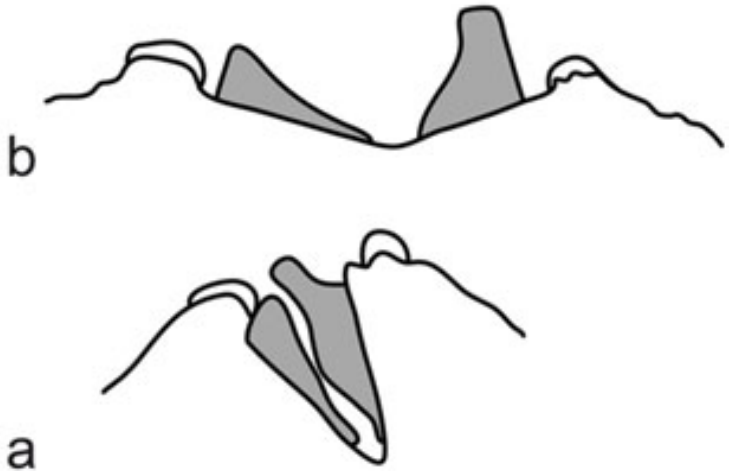
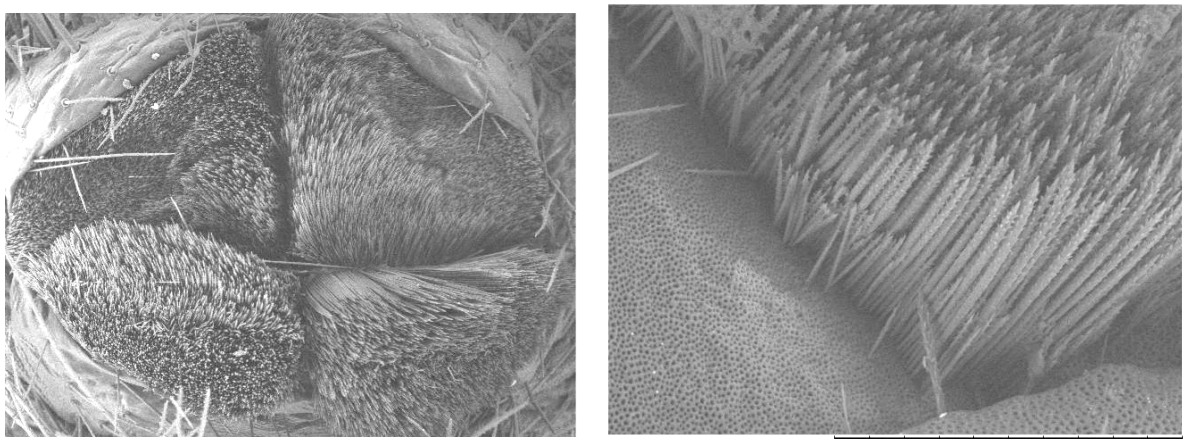


Figure 4. In *Th. pityocampa*, the mirrors are kept folded under normal conditions and only the distal end of the setae is visible (a). When disturbed, the larva opens the mirror, releasing the setae (b) (from, Moneo et al. 2015).



A **B**

Figure 5. Open mirror with urticating setae (A). Long and short setae cohabiting in the same mirror (B) of *Th. pityocampa* (from Petrucco Toffolo et al. 2014).

1.2.4 Medical and veterinary impact of urticating setae

The defoliating processionary moths release abundant quantities of true setae, increasing the contact risk with humans, pets, livestock and wildlife.

The data shown in the previous sub-chapter 1.2.3 demonstrate that setae can be dispersed kilometers away from the origin a fact that now explains why some sensitized subjects experience symptoms without a direct contact with larvae. The setae can persist in the environment for a long time and in silk shelters used by larvae, are reactive for at least one year (Battisti et al. 2011).

Setae can also remain active for long periods in the soil where larvae have pupated, in collection material, and in contaminated clothes, although no precise estimates are available (Battisti et al. 2011).

Thus, humans and other animals can be exposed to setae long after the active insect has disappeared. The urticating setae can induced lesion after penetration in to the skin and probably enzymes causing an additional injury that contributes to an increase of the inflammation observed in individuals who were in contact with larvae (Moneo et al. 2015), induce skin lesions such as urticaria or dermatitis, rhinitis, conjunctivitis, ocular lesions and rarely respiratory symptoms, dyspnea or even anaphylactic shock (Battisti et al. 2011).

Urticating hairs can also affect pets, livestock and wildlife, ingestion of caterpillars, , may have dramatic consequences, such as tongue necrosis in dogs, in addition, setae coming into contact with tissues of the mouth, pharynx, or intestine may provoke severe symptoms and sometimes have life-threatening consequences (Battisti et al. 2011).

These reactions are attributable to a combination of non-allergic and allergic factors. The use of molecular biology has made possible the study of some allergens present in the setae therefore, setae must be considered as a source of allergens and not only as producers of irritant or toxic reactions (Moneo et al. 2015).

The sensitizing capacity of moth allergens is clearly demonstrated with the help of epidemiological studies (Moneo et al. 2015). Frequent contact seems to be the most relevant factor for sensitization and occupationally exposed workers should be carefully checked for sensitization in order to avoid further exposure to the allergens (Moneo et al. 2015).

1.3 Proteins associated with the urticating setae of the pine processionary moth

1.3.1 Thaumetopoein protein (Lamy et al. 1985)

The first study on the proteins associated to urticating setae of *Th. pityocampa* was published in by Lamy et al. (1983) but they mentioned the thaumetopoein protein in the 1985 (Lamy et al. 1985). They described a 28 kDa dimeric protein, exclusive to the setae, called thaumetopoein and formed by two subunits, one of 13kDa and the other of 15 kDa. Thaumetopoein action was proved in guinea pigs where induced mast cell degranulation by a non-immune mechanism (Lamy et al. 1985). Several years later, the same scientific group described a homologue of thaumetopoein in the setae of the oak processionary larvae (Lamy et al. 1988). This protein exhibited the same urticating effect as thaumetopoein in the guinea pigs skin.

1.3.2 Tha p 1 (Moneo et al. 2003)

The protein Tha p 1 was described for the first time in Moneo et al. 2003. In this paper, more than ten different proteins of the extract were able to bind patients IgE, being the most frequently detected a protein of around 15 kDa. 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 terminal sequence GETYSDKYDTIDVNEVLQ for Tha p 1 was obtained, but, at that time, no similarities with other proteins were found using the web interface BLAST of the USA National Centre for Biotechnology Information (NCBI) and so supposed to be an allergen.

Several years later, the complete sequencing of the silkworm *Bombyx mori* genome led to classify Tha p 1 as a chemosensory protein (Larsson & Backlund,2009) .

Despite the high homology between the chemosensory proteins of *Th. pityocampa* and *B. mori*, patients sensitized to the pine processionary larva did not recognize any protein of a silkworm whole body crude extract.

1.3.3 Tha p 2 (Rodriguez-Mahillo et al. 2012)

Tha p 2 was discovered by Rodriguez-Mahillo et al. 2012 from *Th. pityocampa* setae extracts. Setae extracts were characterized by gel staining and immunoblot, with sera from patients with immediate reactions and positive prick test reactions, as well as a rabbit antiserum raised against setae. The most relevant allergen was analysed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS), and its sequence was deduced from an expressed sequence tag bank.

It has not similarity with Tha p 1 and it may correspond to the thaumetopein described in 1985, unfortunately no information about the amino acid composition of thaumetopoein is available, so it was named Tha p 2. This protein detected is a major caterpillar allergen (Rodriguez-Mahillo et al. 2012) and so they confirmed that the penetration of the setae from the pine processionary larvae delivers their allergenic content in addition to causing mechanical or toxic injury. Moneo et al. (2015) noted also that Tha p 2 showed similarity in the carboxy terminal region to a hypothetical protein of the pea aphid *Acyrthosiphon pisum*.

1.3.4 Tha p 3 (Moneo et al. 2015)

Moneo et al. (2015) described a new protein purified by reverse phase HPLC named Tha p 3. The amino end of the low molecular weight setae allergen has been sequenced (LAVETPEPISSN) and some other internal sequences have been obtained by the novo sequencing and MALDI-MS: EKDVHEWTGANWK, DVHEWTGANWK VHVEWKGDN, the K of the last peptide can also be the amino acid Q. None of these sequences had similarities with any other described protein (Moneo et al. 2015).

1.4 Transcriptomics and the first pine processionary moth reference transcriptome

1.4.1 Transcriptome definition and transcriptomics aims

The transcriptome is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA transcribed in one cell or a population of cells. It differs from the exome in that it includes only those RNA molecules found in a specified cell population, and usually includes the amount or concentration of each RNA molecule in addition to the molecular identities. The term can be applied to the total set of transcripts in a given organism, or to the specific subset of transcripts present in a particular cell type.

Unlike the genome, which is roughly fixed for a given cell line (excluding mutations), the transcriptome can vary with external environmental conditions. Because it includes all mRNA transcripts in the cell, the transcriptome reflects the genes that are being actively expressed at any given time, with the exception of mRNA degradation phenomena such as transcriptional attenuation.

The study of transcriptomics, also referred to as expression profiling, examines the expression level of mRNAs in a given cell population, often using high-throughput techniques. The use of next-generation sequencing technology to study the transcriptome is known as RNA-Seq (Want et al. 2009).

1.4.2 The first reference transcriptome of the pine processionary moth

The reference transcriptome was provided by INRA Montpellier, obtained by the combination of 454 and Sanger techniques and used for phenological studies in populations of *Th. pityocampa* (data not published).

1.4.3 Relation to the proteome

The transcriptome can be seen as a precursor for the proteome, that is, the entire set of proteins expressed by a genome. However, the analysis of relative mRNA expression levels can be complicated by the fact that relatively small changes in mRNA expression can produce large changes in the total amount of the corresponding protein present in the cell.

The number of protein molecules synthesized using a given mRNA molecule as a template is highly dependent on translation-initiation features of the mRNA sequence; in particular, the ability of the translation initiation sequence is a key determinant in the recruiting of ribosomes for protein translation.

1.5 Objectives

This project focus on the proteins associated with the urticating setae of the pine processionary moth *Thaumetopoea pityocampa* (Denis & Schiffermüller 1775).

The aim of this thesis is to establish a clear data set of the proteins extracted from the urticating setae of the pine processionary larvae, thus doing these steps:

1) Match the correspondences between the peptides obtained by mass spectrometry results in the reference transcriptome;

2) Identification of the protein given by the reference transcriptome through the Protein BLAST program implemented in the NCBI database;

3) Analyses of the BLAST output.

2. Materials and methods

2.1 Summary of the preparatory work

The information about the proteins and peptides included in the urticating setae of *Th. pityocampa* were obtained starting from a preparatory work conducted in a larger project and consisting of:

1) setae collection from 4th and 5th instar larvae of *Th. pityocampa*;

2) protein extraction from the setae;

3) protein in situ digestion and sodium dodecylsulfate-polyacrilamyde gel electrophoresis of the protein content;

4) mass spectrometry (LC-MS/MS) and peptides matching against the transcriptome dataset;

5) peptides mapping against the Transcriptome and protein sequences detection;

6) protein identification through BLASTp and development of data sets.

The Fig. 6 illustrates the steps for a better comprehension.

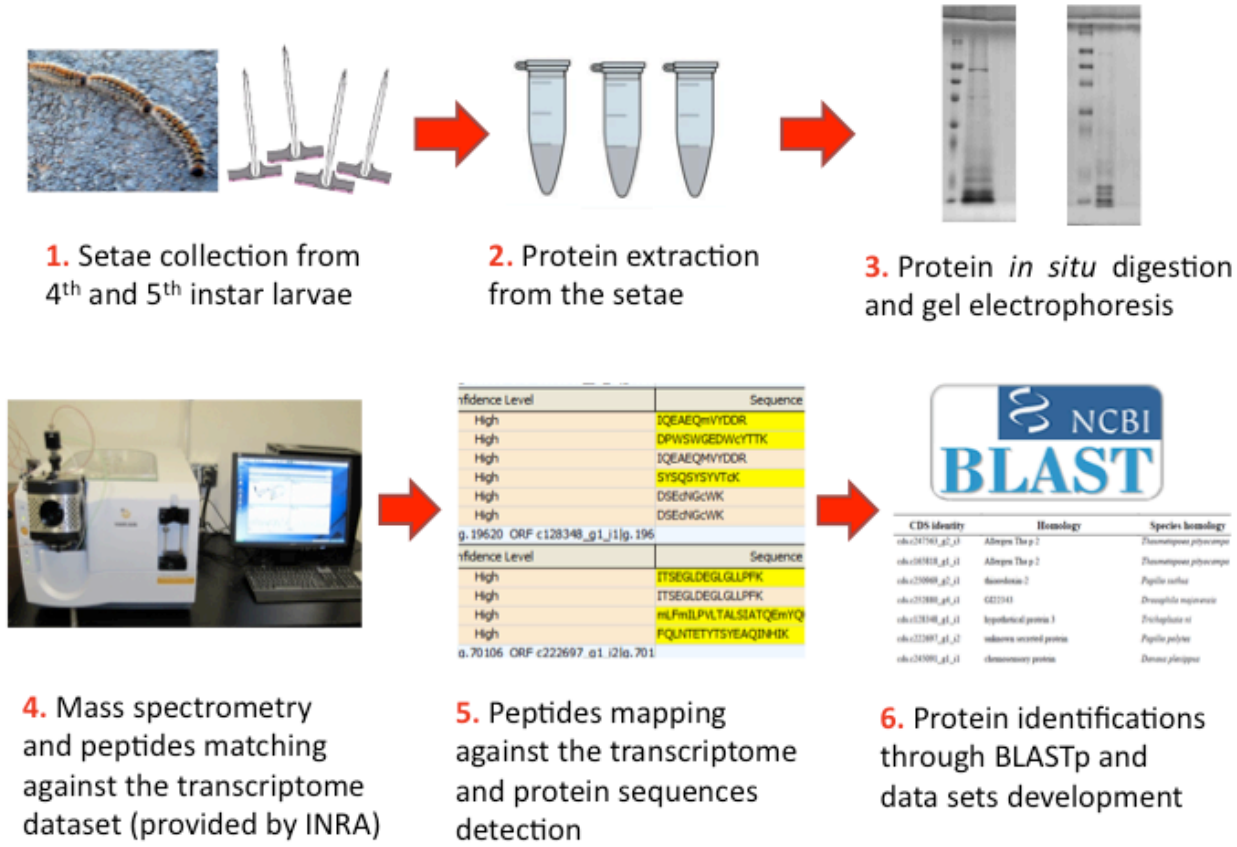


Figure 6. Steps performed for the information retrieval of the urticating setae's protein

content.

In the following paragraphs a detailed description of the steps is presented.

2.2 Protein extraction from urticating setae and gel electrophoresis

The extraction of proteins from the urticating setae was performed with two different protocols. In both there was a mechanical breaking of the setae and the extracts obtained were placed in polyacrylamide gels. However, the second protocol called the Hepes method, has provided a further step in liquid nitrogen for the breaking of the setae and the extracts were subsequently placed in acetone, thus resulting a more fragmented extracts. The extraction protocol allows to extract a good quality of protein from urticating setae.

The extracts obtained from the two different extractions were used for SDS-PAGE (Sodium dodecylsulfate polyacrilamyde gel electrophoresis) using three homemade 13% gels. In the first gel were run the supernatant named SNA and two different pellet (PA, PB), the supernatant obtained from the second method was run in two 13% gels using two different supernatant concentration named 1 and 2. Proteins were visualized by Coomassie staining. The clear bands and the smear obtained from both extraction methods were excised, successively digested and analyzed by the mass spectrometer.

2.3 Analyses of the peptides extracted

The mass spectrometer sequenced every peptide present in each gel bands and compared these peptides with the reference transcriptome. This procedure was implemented localizing the name of the coding DNA sequence (CDS) present in the transcriptome already converted entirely in amino acids. All the CDS found with their respective peptides were put in different spreadsheets. Each spreadsheet corresponds to a specific band of the gel (Fig. 7).

The resulting peptides from the mass spectrometer were about 10-20 amino acid long and were individually selected by eliminating the identical copies (Fig. 7).

Every CDS found by the mass spectrometer was sought with a simple text search in the transcriptome, in order to localize the corresponding protein sequence. Then all the matched CDS and protein sequences were stored in a separated text file. So all protein sequences extracted were localized with at least two independent peptides thus with a high degree of confidence. The list of proteins was exported in spreadsheet data

sets further filtering and cleaning of the same peptides.

Furthermore it was made a test, in order to confirm if the peptides detected by the mass spectrometer were effectively part of the localized protein sequence of the reference transcriptome. All the protein sequences mapped in this way were stored in a dataset for the subsequent identification through BLAST.

	A	B	C
28		High	SGcHVSFGcHK
29		High	SYSQSYSYVQcTQDSEcDGcWK
30		High	SGcHVSFGcHK
31	cds.c165818_g1_i1	m.32531 c165818_g1_i1|g.32531 ORF c165818_g1_i1|g.325	267,39
32		Confidence Level	Sequence
33		High	IQEAEQmVYDDR
34		High	DPWSWGEDWcYTTK
35		High	IQEAEQMVYDDR
36		High	SYSQSYSYVTcK
37		High	DSEcNGcWK
38		High	DSEcNGcWK
39	cds.c128348_g1_i1	m.19620 c128348_g1_i1|g.19620 ORF c128348_g1_i1|g.196	169,11
40		Confidence Level	Sequence
41		High	ITSEGLDEGLGLLPFK
42		High	ITSEGLDEGLGLLPFK
43		High	mLFmILPVLTALSIATQEmYQK
44		High	FQLNTETYTSYEAQINHIK
45	cds.c222697_g1_i2	m.70106 c222697_g1_i2|g.70106 ORF c222697_g1_i2|g.701	149,74
46		Confidence Level	Sequence
47		High	cAccPAcVSYLNEGVAcK
48		High	ELGETPSAIcR
49		High	QNPcTSPVAK
50	cds.c250969_g2_i1	m.182076 c250969_g2_i1|g.182076 ORF c250969_g2_i1|g.1	40,13
51		Confidence Level	Sequence
52		Medium	KIEEFSGANVDK
53	cds.c244184_g2_i1	m.132949 c244184_g2_i1|g.132949 ORF c244184_g2_i1|g.1	39,38
54		Confidence Level	Sequence
55		High	FADYTEEER

Figure 7. A spreadsheet output from the mass spectrometer. The column A represents the CDS (coding DNA sequence) name of the corresponding peptides in the column C. The peptides highlighted are individually selected in order to eliminate the other identical copies.

2.4 Bioinformatics analyses

The data set of protein sequences localized in the transcriptome was analyzed through BLASTp (Basic Local Alignment Search Tool) provided by NCBI web site (Altschul et al. 1990) and was used for identifying any possible homologous protein. The BLASTp algorithm did a matching between the protein sequences localized in the transcriptome and those contained in the universal database of the NCBI by selecting the proteins with the highest identity value.

All the protein identifications were stored in a proteomic data set and grouped by similar family. Only the bands of the SNA and PA were totally analyzed and was searched

only the Tha p 1 and Tha p 2 in the other bands.

In the Hepes method (second gel) instead all the supernatant bands were analyzed. A dataset with all BLAST proteins, their description, the BLAST species, % of coverage, and number of peptide mapped was also created, together with a data set of proteins grouped by family.

The BLAST output gave also the species which correspond to the protein identified, thus some considerations about the species taxa or the urticating trait were done according to the bibliography.

3. Results

3.1 Protein Extraction From Urticating Setae and Gel Electrophoresis

In the gel linked to the first method of extraction (Fig. 8 A) the supernatant named SNA and two different pellets (PA, PB) were run. This gel revealed the presence of a clear band between 60 and 100 kDa, then it detected several weak bands at different molecular weights. Only the peptides obtained from the SNA and PA bands of the first gel were analyzed. In the other gel bands, only the presence of Tha p 1 and Tha p 2, and a protein identified in the second gel called in this thesis Tha p 2 bis, were searched.

The second method (Fig. 8 B) detected a clear band at a high molecular weight close to 100 kDa and others between 8 kDa and 12 kDa. The supernatant obtained from the second method (Fig. 8B) was run in two 13% gels using two different supernatant concentration named 1 and 2 and all the peptides present in the bands were analyzed.

In the first gel, Tha p 1 and the urticating Tha p 2 were detected only in the band at 12 kDa, while in the second gel Tha p 2 was identified in all gel bands (see chapter below) and Tha p 1 only at 8 and 12 kDa bands.

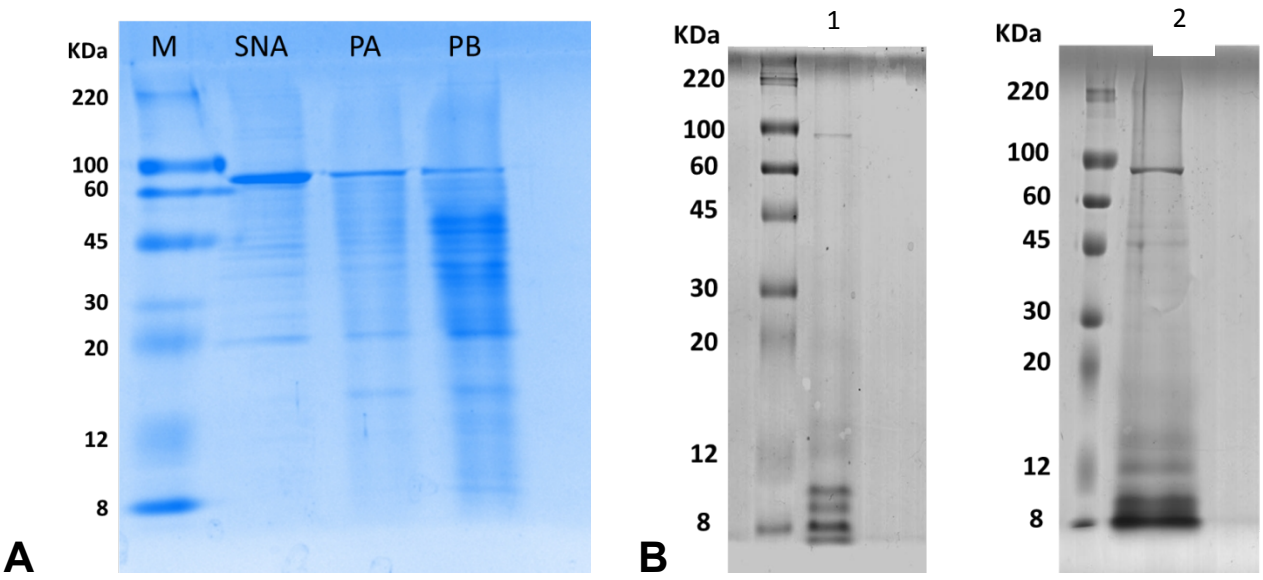


Figure 8. The results of the gel electrophoresis. The left gel represents the first extraction (A) with the supernatant named SNA and two different pellets (PA, PB). The right gels represent the second extraction (B), the supernatant obtained from the second method was run in two 13% gels using two different supernatant concentration named 1 and 2.

3.2 Bioinformatics analyses

The mass spectrometer-based protein identification against the transcriptome and BLASTp detected 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 second gel. Notably, four peptides in each gel band and five in only one band, coded for Tha p 2 (Fig. 9). All the five peptides of Tha p 2 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.

Figure 9. Alignment of the sequence of Tha p 2 and peptides obtained by LC-MS/MS. The Pep 4 was identify only one time.

A total of 115 proteins were identified in the total running gel. The full list of the identified proteins by BLASTp are provided in the Table 3. Only one protein did not match in BLAST database.

Concerning the protein families, those with a frequency higher than 10 were selected (Fig. 10): 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.

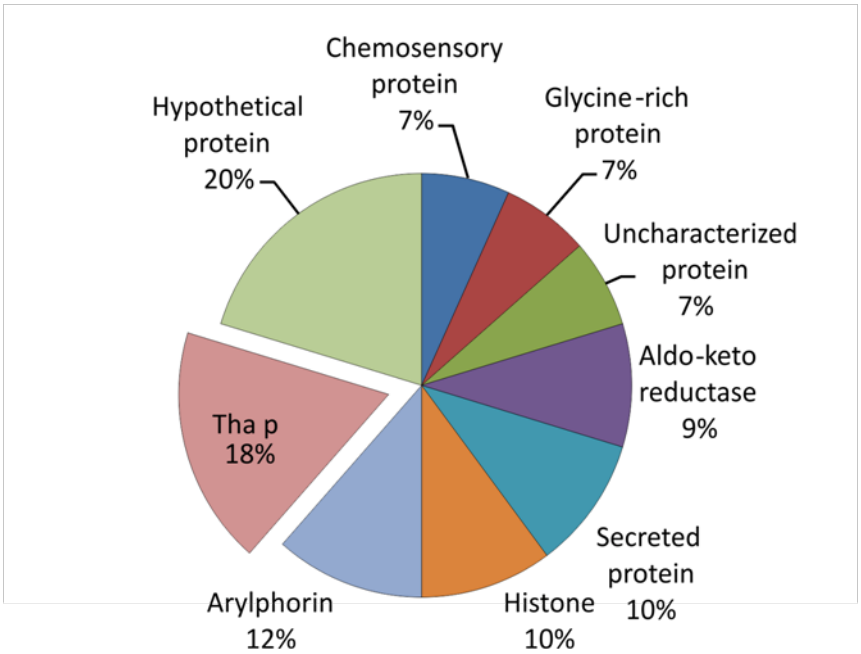


Figure 10. Graphic representation of the frequency of protein families. Only the protein families with a frequency value ≥ 10 were used.

Table 3. List of the proteins identified by BLASTp in the second gel (all the bands analyzed) ordered by the percentage of coverage and by homology (alphabetically). Proteins related to the urticating power are in bold.

CDS identity	Homology	Species homology	% Coverage	No. peptides
cds.c145193_g1_i1	abnormal wing disc protein	*Antheraea pernyi*	100%	2
cds.c254766_g5_i2	actin-2, partial	*Wuchereria bancrofti*	100%	1
cds.c111339_g1_i1	actin-4	*Bombyx mori*	100%	10
cds.c243108_g5_i1	actin, clone 205-like isoformX1	*Apis mellifera*	100%	13
cds.c232174_g1_i1	adenylyltransferase and sulfurtransferase MOCS3 isoform X1	*Nasonia vitripennis*	100%	1
cds.c250034_g2_i1	**Allergen Tha p 1**	***Thaumetopoea pityocampa***	**100%**	**3**
cds.c247563_g2_i3	**Allergen Tha p 2**	***Thaumetopoea pityocampa***	**100%**	**5**
cds.c237572_g1_i1	alpha-N-acetylgalactosaminidase precursor	*Bombyx mori*	100%	1
cds.c243842_g1_i1	apolipophorin III	*Trichoplusia ni*	100%	9
cds.c245915_g1_i5	catalase	*Spodoptera exigua*	100%	9
cds.c253993_g4_i1	chemosensory protein 3 precursor	*Bombyx mori*	100%	1
cds.c249531_g1_i1	chemosensory proteins	*Dendrolimus kikuchii*	100%	1
cds.c245890_g4_i2	chitin binding peritrophin-A	*Papilio xuthus*	100%	2
cds.c229624_g1_i4	chitin deacetylase	*Mamestra brassicae*	100%	1
cds.c214759_g1_i2	coatomer protein complex subunit delta	*Bombyx mori*	100%	1
cds.c227993_g2_i1	cysteine-rich venom protein ENH1-like	*Bombyx mori*	100%	1
cds.c227993_g1_i3	cysteine-rich venom protein ENH1-like	*Bombyx mori*	100%	7
cds.c182227_g2_i1	DEHA2F04796p	*Debaryomyces hansenii*	100%	1
cds.c250126_g4_i1	diapause bioclock protein	*Bombyx mori*	100%	8
cds.c232852_g1_i2	dihydropteridine reductase	*Papilio xuthus*	100%	1
cds.c247998_g1_i2	egalitarian	*Danaus plexippus*	100%	1
cds.c246705_g3_i1	elongation factor	*Nasonia vitripennis*	100%	2
cds.c218396_g2_i1	enolase	*Spodoptera litura*	100%	9
cds.c248159_g11_i1	glyceraldehyde-3-phosphate dehydrogenase	*Spodoptera frugiperda*	100%	3
cds.c146340_g1_i1	glycogen phosphorylase	*Microplitis demolitor*	100%	3
cds.c190487_g1_i3	heat shock protein	*Helicoverpa armigera*	100%	5
cds.c246942_g1_i1	heat shock protein 70	*Spodoptera litura*	100%	2
cds.c255264_g5_i5	hexamerine	*Helicoverpa armigera*	100%	15
cds.c247952_g2_i1	histone H2A-like protein 2	*Bombyx mori*	100%	2
cds.c247482_g4_i2	Histone H2B	*Camponotus floridanus*	100%	2
cds.c358177_g1_i1	hypothetical protein	*Vittaforma corneae*	100%	1
cds.c251838_g1_i5	imaginal disc growth factor-like protein	*Mamestra brassicae*	100%	1
cds.c238074_g1_i2	imaginal disc growth factor-like protein	*Mamestra brassicae*	100%	12
cds.c235010_g1_i1	juvenile hormone binding protein	*Heliothis virescens*	100%	1
cds.c155109_g1_i3	Lin-9-like protein	*Harpegnathos saltator*	100%	1
cds.c17057_g1_i1	malate dehydrogenase, mitochondrial	*Nasonia vitripennis*	100%	1
cds.c248400_g2_i1	mesencephalic astrocyte-derived neurotrophic factor homolog	*Bombyx mori*	100%	1

CDS identity	Homology	Species homology	% Coverage	No. peptides
cds.c255169_g1_i9	methionine-rich storage protein 2	*Manduca sexta*	100%	4
cds.c219735_g1_i1	muscular protein 20	*Bombyx mori*	100%	1
cds.c245136_g2_i2	odorant binding protein 6	*Spodoptera exigua*	100%	1
cds.c162537_g1_i1	peptidyl-prolyl cis-trans isomerase	*Papilio xuthus*	100%	2
cds.c243221_g1_i1	peptidyl-prolyl cis-trans isomerase	*Papilio polytes*	100%	5
cds.c253495_g4_i1	phosphatidylethanolamine binding protein isoform 2	*Bombyx mori*	100%	6
cds.c243750_g2_i1	phosphatidylethanolamine-binding protein	*Bombyx mori*	100%	3
cds.c248468_g2_i4	phosphoglycerate kinase-like	*Bombyx mori*	100%	1
cds.c230591_g3_i2	polyubiquitin-B-like isoform	*Coptotermes formosanus*	100%	2
cds.c255616_g2_i2	proline-rich receptor-like protein kinase	*Bombyx mori*	100%	1
cds.c247424_g3_i3	protein disulfide isomerase	*Papilio xuthus*	100%	5
cds.c9578_g1_i1	protein phosphatase 1 regulatory subunit 21 isoform	*Nasonia vitripennis*	100%	1
cds.c12924_g1_i1	Ribosomal 40S protein S3a-like	*Musca domestica*	100%	1
cds.c230317_g1_i1	Ribosomal acidic 60S protein P2-like	*Bombus impatiens*	100%	2
cds.c254029_g1_i2	serine protease	*Papilio xuthus*	100%	1
cds.c305352_g1_i1	serine protease inhibitor	*Bombyx mori*	100%	1
cds.c244668_g2_i3	serine proteinase-like protein 1	*Helicoverpa armigera*	100%	1
cds.c237352_g1_i3	serine proteinase-like protein precursor	*Bombyx mori*	100%	5
cds.c215532_g1_i3	serine/threonine-protein phosphatase	*Apis mellifera*	100%	1
cds.c232764_g1_i5	similar to CG9796	*Papilio xuthus*	100%	1
cds.c250933_g2_i1	storage protein 1	*Plutella xylostella*	100%	6
cds.c246112_g4_i1	superoxide dismutase	*Danaus plexippus*	100%	6
cds.c324063_g1_i1	translation initiation factor	*Cerapachys biroi*	100%	1
cds.c252682_g3_i6	triosephosphate isomerase	*Helicoverpa armigera*	100%	1
cds.c355329_g1_i1	triosephosphate isomerase	*Microplitis demolitor*	100%	2
cds.c244766_g3_i2	tubulin gamma-1 chain	*Spodoptera exigua*	100%	1
cds.c334690_g1_i1	uncharacterized protein	*Microplitis demolitor*	100%	1
cds.c252733_g3_i1	uncharacterized protein	*Bombyx mori*	100%	3
cds.c244127_g2_i1	uncharacterized protein	*Bombyx mori*	100%	6
cds.c251831_g2_i2	vitellogenin	*Helicoverpa armigera*	100%	2
cds.c236757_g1_i2	voltage-dependent anion-selective channel-like isoform X4	*Bombyx mori*	100%	1
cds.c243108_g8_i1	actin	*Onthophagus nigriventris*	99%	1
cds.c248008_g1_i1	aldo-keto reductase	*Agrotis ipsilon*	99%	15
cds.c254783_g3_i8	alpha-actinin, sarcomeric-like	*Bombyx mori*	99%	3
cds.c18274_g1_i1	binding FK506- protein precursor	*Bombyx mori*	99%	1
cds.c248232_g3_i2	chymotrypsin inhibitor	*Bombyx mori*	99%	1
cds.c248232_g2_i3	chymotrypsin inhibitor CI-8A	*Bombyx mori*	99%	6
cds.c223924_g1_i1	cysteine proteinase inhibitor precursor	*Manduca sexta*	99%	1
cds.c254857_g2_i1	heat shock protein	*Spodoptera litura*	99%	4
cds.c251040_g1_i1	heat shock protein	*Spodoptera litura*	99%	3
cds.c254374_g3_i3	hemolin	*Helicoverpa zea*	99%	10

CDS identity	Homology	Species homology	% Coverage	No. peptides
cds.c249761_g5_i1	hypothetical protein	*Danaus plexippus*	99%	8
cds.c249761_g6_i2	hypothetical protein	*Danaus plexippus*	99%	8
cds.c252693_g6_i3	hypothetical protein	*Danaus plexippus*	99%	1
cds.c234087_g1_i3	protocadherin-16-like PREDICTED:	*Bombyx mori*	99%	1
cds.c254039_g4_i10	serine proteinase inhibitor-1A	*Mamestra brassicae*	99%	2
cds.c265904_g1_i1	sorting nexin-13-like isoform	*Nasonia vitripennis*	99%	1
cds.c162397_g1_i1	syntrophin	*Aedes aegypti*	99%	1
cds.c246154_g1_i1	tetratricopeptide repeat protein	*Bombyx mori*	99%	1
cds.c241261_g1_i1	unknown similar to AMEV109	*Mythimna separata*	99%	13
cds.c223078_g1_i1	yellow-c	*Heliconius melpomene*	99%	13
cds.c215365_g1_i1	yellow-d	*Bombyx mori*	99%	11
cds.c248008_g2_i1	aldo-keto reductase	*Agrotis ipsilon*	98%	12
cds.c251564_g7_i2	allantoinase	*Danaus plexippus*	98%	2
cds.c255429_g2_i4	arylphorin type 2	*Cerura vinula*	98%	36
cds.c242366_g1_i2	calreticulin	*Papilio xuthus*	98%	3
cds.c252458_g5_i2	chemosensory protein 7 precursor	*Bombyx mori*	98%	3
cds.c252167_g1_i1	glucose-6-phosphate isomerase	*Spodoptera exigua*	98%	1
cds.c255264_g5_i10	hexamerine	*Helicoverpa armigera*	98%	8
cds.c209834_g1_i1	hypothetical protein	*Danaus plexippus*	98%	2
cds.c242998_g2_i1	mitochondrial aldehyde dehydrogenase	*Danaus plexippus*	98%	1
cds.c230658_g2_i3	multiple coagulation factor deficiency protein	*Bombyx mori*	98%	1
cds.c250390_g1_i3	probable citrate synthase 1, mitochondrial-like	*Bombyx mori*	98%	1
cds.c254772_g4_i5	secreted protein unknown	*Papilio xuthus*	98%	3
cds.c207263_g1_i3	TBC1 domain family member	*Bombyx mori*	98%	1
cds.c254599_g10_i4	uncharacterized protein	*Bombyx mori*	98%	5
cds.c245014_g3_i1	yellow-c	*Heliconius erato*	98%	1
cds.c253205_g1_i3	arylphorin	*Cerura vinula*	97%	33
cds.c257062_g1_i1	hypothetical protein	*Bombus terrestris*	97%	1
cds.c262202_g1_i1	interference hedgehog-like	*Bombus terrestris*	97%	1
cds.c255169_g1_i3	methionine-rich storage protein 2	*Manduca sexta*	97%	6
cds.c225502_g1_i1	myosin-I heavy chain-like	*Bombyx mori*	97%	1
cds.c255122_g2_i2	saposin	*Papilio polytes*	97%	2
cds.c235717_g1_i3	SCO-spondin-like	*Bombyx mori*	97%	6
cds.c239913_g1_i2	uncharacterized protein LOC101742613 isoform	*Bombyx mori*	97%	5
cds.c217974_g1_i3	unknown secreted protein	*Papilio xuthus*	97%	3
cds.c237613_g1_i1	WAP four-disulfide core domain protein 2	*Papilio xuthus*	97%	1
cds.c237613_g2_i3	WAP four-disulfide core domain protein 2 precursor	*Papilio xuthus*	97%	4
cds.c244264_g3_i1	aldo-keto reductase family 1	*Bombyx mori*	96%	8
cds.c244184_g2_i1	BCP inhibitor precursor	*Bombyx mori*	96%	1
cds.c241205_g3_i2	odorant-binding protein 4	*Chilo suppressalis*	96%	1
cds.c231482_g1_i1	cytochrome c oxidase subunit 4 isoform 1, mitochondrial	*Nasonia vitripennis*	95%	1

CDS identity	Homology	Species homology	% Coverage	No. peptides
cds.c181512_g1_i1	histone H2A-like	*Diaphorina citri*	95%	1
cds.c234385_g1_i1	juvenile hormone binding protein	*Bombyx mori*	95%	2
cds.c248262_g3_i7	mitochondrial aldehyde dehydrogenase	*Danaus plexippus*	95%	3
cds.c245445_g2_i8	unknown protein	*Helicoverpa armigera*	95%	8
cds.c248540_g1_i1	unknown secreted protein	*Papilio polytes*	95%	7
cds.c162794_g1_i1	uncharacterized protein	*Apis dorsata*	94%	1
cds.c245091_g1_i1	chemosensory protein	*Danaus plexippus*	93%	5
cds.c250954_g3_i4	hypothetical protein	*Danaus plexippus*	93%	1
cds.c252077_g1_i7	Moesin A	*Spodoptera frugiperda*	93%	2
cds.c253640_g7_i4	nucleoplasmin isoform 2	*Danaus plexippus*	93%	1
cds.c255429_g2_i1	aryphorin type 2	*Cerura vinula*	92%	1
cds.c243980_g12_i4	hypothetical protein	*Heliothis virescens ascovirus*	92%	1
cds.c225190_g1_i1	neurogenic protein mastermind-like	*Bombyx mori*	92%	1
cds.c185079_g1_i1	REPAT32	*Spodoptera littoralis*	92%	2
cds.c252383_g1_i2	slowmo	*Bombyx mori*	92%	1
cds.c248008_g1_i2	aldo-keto reductase, partial	*Agrotis ipsilon*	91%	1
cds.c193417_g1_i3	alpha-actinin, sarcomeric-like isoform X1	*Bombyx mori*	91%	1
cds.c123959_g1_i1	pterin-4-alpha-carbinolamine dehydratase-like	*Bombyx mori*	91%	1
cds.c99479_g1_i1	unknown secreted protein	*Papilio xuthus*	91%	2
cds.c128348_g1_i1	hypothetical protein 3	*Trichoplusia ni*	90%	6
cds.c243797_g4_i2	juvenile hormone binding protein an-0921 precursor	*Bombyx mori*	90%	1
cds.c255660_g2_i10	unconventional myosin-XV-like	*Bombyx mori*	88%	1
cds.c238976_g1_i2	apolipophorin III	*Trichoplusia ni*	87%	1
cds.c160507_g1_i1	cuticular protein	*Danaus plexippus*	87%	3
cds.c243533_g8_i1	glycerophosphodiester phosphodiesterase	*Pseudomonas aeruginosa*	87%	1
cds.c236041_g1_i5	histone H2A-like	*Nasonia vitripennis*	87%	1
cds.c250475_g7_i1	ML-domain containing secreted protein-like protein	*Antheraea yamamai*	87%	2
cds.c391427_g1_i1	peptide chain release factor (probable)	*Nasonia vitripennis*	87%	1
cds.c242810_g1_i4	protein takeout-like	*Bombyx mori*	87%	1
cds.c239062_g1_i2	uncharacterized protein	*Nasonia vitripennis*	87%	1
cds.c165775_g1_i1	Histone H4	*Macaca mulatta*	85%	1
cds.c234385_g1_i2	juvenile hormone binding protein an-0921 precursor	*Bombyx mori*	85%	9
cds.c222697_g1_i2	unknown secreted protein	*Papilio polytes*	85%	3
cds.c165818_g1_i1	**Allergen Tha p 2**	***Thaumetopoea pityocampa***	**84%**	**5**
cds.c356727_g1_i1	cuticular protein glycine-rich 10 precursor	*Bombyx mori*	84%	4
cds.c245309_g1_i2	promoting protein	*Danaus plexippus*	84%	1
cds.c249497_g1_i4	lysozyme	*Helicoverpa armigera*	82%	1
cds.c225627_g1_i1	uncharacterized protein LOC101739120	*Bombyx mori*	82%	1
cds.c247161_g1_i2	hemolymph proteinase 8	*Manduca sexta*	81%	1
cds.c250969_g2_i1	thioredoxin-2	*Papilio xuthus*	80%	3
cds.c256696_g1_i1	thioredoxin-2	*Papilio xuthus*	80%	2

CDS identity	Homology	Species homology	% Coverage	No. peptides
cds.c235263_g1_i1	putative leucine-rich	*Bombyx mori*	79%	1
cds.c180932_g2_i1	hypothetical protein SINV_02932	*Solenopsis invicta*	77%	1
cds.c234296_g2_i2	hypothetical protein	*Danaus plexippus*	72%	2
cds.c240809_g3_i7	hypothetical protein 31	*Lonomia obliqua*	72%	1
cds.c109578_g1_i1	hypothetical protein	*Tetrapisispora phaffii*	70%	1
cds.c22118_g1_i1	Neurobeachin	*Cerapachys biroi*	70%	1
cds.c246506_g2_i12	acyl-CoA binding protein 1	*Sesamia inferens*	69%	2
cds.c245279_g1_i4	aldo-keto reductase, partial	*Agrotis ipsilon*	69%	5
cds.c147708_g1_i1	probable GPI-anchored adhesin-like protein	*Acyrthosiphon pisum*	67%	1
cds.c247894_g2_i4	glycine-rich protein	*Bombyx mori*	65%	2
cds.c231404_g2_i2	nucleobindin-2-like isoform	*Bombyx mori*	64%	2
cds.c239631_g5_i3	Y+L amino acid transporter 2 isoform X1	*Nasonia vitripennis*	64%	1
cds.c254703_g1_i1	uncharacterized protein	*Bombyx mori*	58%	1
cds.c255774_g1_i2	hypothetical protein	*Danaus plexippus*	48%	1
cds.c245626_g1_i2	hypothetical protein	*Lonomia obliqua*	46%	6
cds.c227972_g2_i1	histone H1	*Oreta rosea*	44%	3
cds.c250370_g3_i2	helicase	*Lactobacillus apodemi*	39%	1
cds.c252804_g1_i2	GI22710	*Drosophila mojavensis*	35%	1
cds.c252880_g4_i1	GI22343	*Drosophila mojavensis*	34%	7
cds.c252247_g1_i3	uncharacterized protein	*Nasonia vitripennis*	29%	1
cds.c255864_g3_i1	No match	No match	No match	1

The list of the proteins identified by BLASTp in the First gel is shown in Table 4.

Table 4. List of the proteins identified by BLASTp in the first gel (only SNA and PA gel bands analyzed, Tha p proteins were not found in this gel but they were found in the PB gel).

CDS identity	Homology	*Species homology*	% Coverage	No. peptides
cds.c111339_g1_i1	actin-4	*Bombyx mori*	100%	4
cds.c248037_g10_i7	acyl-coa dehydrogenase	*Papilio xuthus*	100%	1
cds.c214977_g1_i3	coiled-coil domain-containing protein 112-like PREDICTED:	*Bombyx mori*	100%	1
cds.c190487_g1_i3	heat shock protein	*Helicoverpa armigera*	100%	2
cds.c230131_g1_i1	hemocyte protein-glutamine gamma-glutamyltransferase-like PREDICTED:	*Bombyx mori*	100%	1
cds.c255264_g5_i5	hexamerine	*Helicoverpa armigera*	100%	11
cds.c250522_g2_i1	iron regulatory protein 1	*Manduca sexta*	100%	2
cds.c219735_g1_i1	muscular protein 20	*Bombyx mori*	100%	1
cds.c253495_g4_i1	phosphatidylethanolamine binding protein isoform 2	*Bombyx mori*	100%	1
cds.c255315_g2_i3	retinal dehydrogenase 1-like PREDICTED:	*Bombyx mori*	100%	1
cds.c250933_g2_i1	storage protein 1	*Plutella xylostella*	100%	4
cds.c251286_g4_i5	transitional endoplasmic reticulum ATPase TER94	*Danaus plexippus*	100%	2
cds.c254429_g5_i1	actin	*Zygaena filipendulae*	99%	16
cds.c241485_g1_i3	aldo-keto reductase	*Agrotis ipsilon*	99%	2
cds.c241568_g1_i1	aminopeptidase 110 kDa	*Heliothis virescens*	99%	3
cds.c220370_g1_i1	aminopeptidase N precursor	*Bombyx mori*	99%	2
cds.c256158_g1_i3	apolipophorin precursor protein	*Bombyx mori*	99%	86
cds.c159683_g1_i1	arginine kinase	*Papilio xuthus*	99%	5
cds.c241646_g1_i1	arylphorin	*Cerura vinula*	99%	39
cds.c251716_g3_i10	beta-galactosidase	*Bombyx mori*	99%	1
cds.c252077_g1_i7	Chain A, Moesin From Spodoptera Frugiperda At 2.1 Angstroms Resolution	*Spodoptera frugiperda*	99%	2
cds.c246759_g2_i4	dipeptidyl peptidase 3-like isoform X1 PREDICTED:	*Bombyx mori*	99%	5
cds.c252207_g3_i1-1	ecdysteroid-inducible angiotensin-converting enzyme-related gene product precursor	*Bombyx mori*	99%	2
cds.c17659_g1_i1	elongation factor 1 alpha	*Cryptomeigenia sp. JOS-2*	99%	3
cds.c253664_g1_i1	gelsolin, cytoplasmic-like PREDICTED:	*Bombyx mori*	99%	2
cds.c245489_g2_i3	GL12416	*Drosophila persimilis*	99%	6
cds.c252588_g2_i1	glycogen phosphorylase	*Spodoptera exigua*	99%	3
cds.c254857_g2_i1	heat shock cognate 70 protein	*Sesamia inferens*	99%	12
cds.c190487_g1_i3	heat shock protein	*Helicoverpa armigera*	99%	16
cds.c246254_g6_i1	heat shock protein 90	*Spodoptera litura*	99%	7
cds.c255264_g5_i5	hexamerine	*Helicoverpa armigera*	99%	11
cds.c249503_g8_i1	histone H3.3-like PREDICTED:	*Oryzias latipes*	99%	2
cds.c232726_g1_i2	hypothetical protein	*Antheraea yamamai*	99%	3
cds.c250176_g7_i1	malate dehydrogenase, mitochondrial-like PREDICTED:	*Bombyx mori*	99%	2
cds.c247030_g2_i1	methionine-rich storage protein 2	*Manduca sexta*	99%	8
cds.c256255_g1_i5	muscle myosin heavy chain	*Papilio xuthus*	99%	5
cds.c219735_g1_i1	muscular protein 20	*Bombyx mori*	99%	5
cds.c244306_g3_i4	peroxisomal multifunctional enzyme	*Agrotis segetum*	99%	11
cds.c253495_g4_i1	phosphatidylethanolamine binding protein isoform 2	*Bombyx mori*	99%	4

CDS identity	Homology	*Species homology*	% Coverage	No. peptides
cds.c238188_g1_i1	prophenoloxidase	*Spodoptera exigua*	99%	13
cds.c247424_g1_i1	protein disulfide isomerase	*Papilio xuthus*	99%	4
cds.c239519_g6_i1	putative Aminopeptidase N precursor	*Danaus plexippus*	99%	1
cds.c248697_g1_i10	putative annexin IX-C	*Manduca sexta*	99%	3
cds.c255315_g2_i1	retinal dehydrogenase 1-like PREDICTED:	*Bombyx mori*	99%	4
cds.c254039_g4_i10	serine proteinase inhibitor-1A	*Mamestra brassicae*	99%	2
cds.c249374_g7_i1	transketolase	*Bombyx mori*	99%	10
cds.c255429_g2_i4	aryphorin type 2	*Cerura vinula*	98%	38
cds.c245484_g2_i6	fasciclin-3-like PREDICTED:	*Bombyx mori*	98%	4
cds.c248312_g1_i1	transferrin	*Spodoptera litura*	98%	8
cds.c253205_g1_i3	arylphorin	*Cerura vinula*	97%	36
cds.c255429_g2_i4	aryphorin type 2	*Cerura vinula*	97%	52
cds.c201500_g1_i1	chymotrypsin-2 PREDICTED:	*Nasonia vitripennis*	97%	1
cds.c250160_g1_i7	venom serine carboxypeptidase-like PREDICTED:	*Bombyx mori*	97%	2
cds.c247067_g4_i1	aminopeptidase N3 precursor	*Bombyx mori*	96%	3
cds.c237930_g5_i2	DNA topoisomerase 2 PREDICTED:	*Nasonia vitripennis*	96%	1
cds.c252105_g4_i1	lamin-C-like PREDICTED:	*Bombyx mori*	96%	7
cds.c248262_g3_i7	mitochondrial aldehyde dehydrogenase	*Danaus plexippus*	95%	3
cds.c252600_g3_i7	prophenoloxidase-1	*Heliothis virescens*	95%	6
cds.c159993_g1_i1	ubiquitin thioesterase OTU1-like PREDICTED:	*Megachile rotundata*	95%	1
cds.c162794_g1_i1	uncharacterized protein LOC102679586 isoform X1 PREDICTED:	*Apis dorsata*	94%	1
cds.c229624_g1_i4	uncharacterized protein LOC101740647 PREDICTED:	*Bombyx mori*	93%	2
cds.c252207_g3_i1-2	angiotensin converting enzyme	*Papilio polytes*	92%	1
cds.c243842_g1_i1	apolipophorin III	*Trichoplusia ni*	90%	3
cds.c246984_g5_i1	integrin beta1	*Bombyx mori*	87%	1
cds.c220036_g1_i1	carboxylesterase clade H, member 1 precursor	*Bombyx mori*	85%	9
cds.c245903_g1_i6	prolyl endopeptidase-like PREDICTED:	*Bombyx mori*	84%	1
cds.c280961_g1_i1	laminin subunit beta-2 PREDICTED:	*Astyanax mexicanus*	82%	1
cds.c227972_g2_i1	Histone H1	*Oreta rosea*	43%	3

A few proteins were found to occur in both gels although a full comparison is not possible because the complete sequencing is available only for the second gel.

A full list of the taxa found in the second gel is available in Table 6. Most of them are insects (83%) and among these the Lepidoptera are the 62%. Within the Lepidoptera, the group to which the processionary moths belong (Noctuidea) represents the 54%. In addition to the processionary moth there are other two species of urticating Lepidoptera which share some of the proteins found in the setae.

Table 6. Full list of the taxa matching the proteins found in the second gel, according to a taxonomic alphabetical order.

Species	Insect/Non Insect	Group/Order	Family/Type
Onthophagus nigriventris	Insect	Coleoptera	Scarabaeidae
Aedes aegypti	Insect	Diptera	Culicidae
Drosophila mojavensis	Insect	Diptera	Drosophilidae
Musca domestica	Insect	Diptera	Muscidae
Diaphorina citri	Insect	Homoptera	Psyllidae
Cerapachys biroi	Insect	Hymenoptera	Formicidae
Acyrthosiphon pisum	Insect	Hymenoptera	Aphididae
Apis dorsata	Insect	Hymenoptera	Apidae
Apis mellifera	Insect	Hymenoptera	Apidae
Bombus impatiens	Insect	Hymenoptera	Apidae
Bombus terrestris	Insect	Hymenoptera	Apidae
Microplitis demolitor	Insect	Hymenoptera	Braconidae
Camponotus floridanus	Insect	Hymenoptera	Formicidae
Solenopsis invicta	Insect	Hymenoptera	Formicidae
Nasonia vitripennis	Insect	Hymenoptera	Pteromalidae
Harpegnathos saltator	Insect	Hymenoptera	Formicidae
Coptotermes formosanus	Insect	Isoptera	Rhinotermitidae
Bombyx mori	Insect	Lepidoptera	Bombycidae
Chilo suppressalis	Insect	Lepidoptera	Crambidae
*Dendrolimus kikuchii**	Insect	Lepidoptera	Lasiocampidae
Oreta rosea	Insect	Lepidoptera	Drepaninae
Agrotis ipsilon	Insect	Lepidoptera	Noctuidae
Heliconius erato	Insect	Lepidoptera	Noctuidae
Heliconius melpomene	Insect	Lepidoptera	Noctuidae
Helicoverpa armigera	Insect	Lepidoptera	Noctuidae
Helicoverpa zea	Insect	Lepidoptera	Noctuidae
Heliothis virescens	Insect	Lepidoptera	Noctuidae
Mamestra brassicae	Insect	Lepidoptera	Noctuidae
Sesamia inferens	Insect	Lepidoptera	Noctuidae
Spodoptera exigua	Insect	Lepidoptera	Noctuidae
Spodoptera frugiperda	Insect	Lepidoptera	Noctuidae
Spodoptera littoralis	Insect	Lepidoptera	Noctuidae
Spodoptera litura	Insect	Lepidoptera	Noctuidae
Trichoplusia ni	Insect	Lepidoptera	Noctuidae
Cerura vinula	Insect	Lepidoptera	Notodontidae
Thaumetopoea pityocampa	Insect	Lepidoptera	Notodontidae
Danaus plexippus	Insect	Lepidoptera	Nymphalidae
Papilio polytes	Insect	Lepidoptera	Papilionidae
Papilio xuthus	Insect	Lepidoptera	Papilionidae

Species	Insect/Non Insect	Group/Order	Family/Type
Papilio xuthus	Insect	Lepidoptera	Papilionidae
Plutella xylostella	Insect	Lepidoptera	Plutellidae
Antheraea pernyi	Insect	Lepidoptera	Saturniidae
Antheraea yamamai	Insect	Lepidoptera	Saturniidae
*Lonomia obliqua**	Insect	Lepidoptera	Saturniidae
Manduca sexta	Insect	Lepidoptera	Sphingidae
Pseudomonas aeruginosa VRFPA04	Non Insect	Bacteria	Generalist
Lactobacillus apodemi	Non Insect	Bacteria	Generalist
Macaca mulatta	Non Insect	Mammalia	Monkey
Vittaforma corneae ATCC 50505	Non Insect	Microsporidia	Insect pathogen
Wuchereria bancrofti	Non Insect	Nematoda	Human pathogen
Heliothis virescens ascovirus 3e	Non Insect	Virus	Insect pathogen
Mythimna separata entomopoxvirus 'L'	Non Insect	Virus	Insect pathogen
Tetrapisispora phaffii CBS 4417	Non Insect	Yeast	Generalist
Debaryomyces hansenii CBS767	Non Insect	Yeast	Generalist

* urticating species

4. Discussion

In this thesis project we confirm that the urticating setae of *Th. pityocampa* contain proteins, some of which are recognized by Ig-E of forestry workers presenting skin reactions when exposed to the larvae of this insect in infested areas (Berardi 2015). 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 extractions of the setae allowed the precipitation of a large amount of protein material. Since the 37% of the urticating setae is chitin (Berardi 2015), this brings us to think that most of the rest is mainly proteins.

The second method of extraction allowed to detect the presence of a high number of proteins (182). 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 occurrence in the setae of several proteins typical of the cell metabolism (such as arylphorin and histone proteins) that are very similar to those of other insects, and especially Lepidoptera, may support the hypothesis of the sink function. Interestingly, two proteins were found to be similar to those expressed by two other urticating Lepidoptera (a chemosensory protein of *Dendrolimus* with 100% of coverage and an hypothetical protein of *Lonomia* with 46% of coverage) although the mechanism of reaction induced by these species is different, consisting in modified seta and spine, respectively (Battisti et al. 2011). This aspect clearly needs to be further explored.

The presence of Tha p 2 or its parts in all bands obtained with the second 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, Berardi 2015).

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.

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 people 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.

The awareness of a great protein content in the urticating setae may open the way for better understanding the risks to which humans and domestic mammalians are exposed. In the case of processionary moths, such a risk is very high due to the wide distribution of host plants in both urban and forest areas (EFSA 2009). This is emphasized by a continuous release of setae from the soil or from tents, because of their high persistence. The information provided in the present thesis project may offer an opportunity to explore the importance of setal characteristics from the medical and veterinary points of view and it may be useful to pest managers and decision-makers in planning the control operations of these forest and urban tree pests.

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. Finally, we will also look at the occurrence of similar types of responses in the animals that are considered the natural target of the setae, like the insectivorous birds.

5. 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.

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

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.

Berardi L. (2015). Genetic and proteomic approach to the urticating system of processionary moths (Thaumetopoeinae, Lepidoptera). PhD thesis, Doctorate School of Crop Science, Curriculum Crop Protection, Cycle XXVII, University of Padova.

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

Démolin G. (1963). Les 'miroirs' urticants de la processionnaire du pin (*Thaumetopoea pityocampa* Schiff.). *Revue de Zoologie Agricole et Appliquée* **10–12**: 107–14.

EFSA, European Food Safety Authority (2009). Scientific opinion of the Panel on Plant Health on a pest 7 risk analysis on *Thaumetopoea processionea* L., the oak processionary moth, prepared by the UK and extension of its scope to the EU territory. *The EFSA Journal* **1195**: 1–64.

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*) Stimulate Proliferation of Human Blood Lymphocytes In Vitro. *PLoS ONE* 9(12): e113977.

Lamy M., Vincendeau P., Ducombs G., Pastureaud M. H. (1983). Irritating substance extracted from the *Thaumetopoea pityocampa* caterpillar: mechanism of action. *Experientia*, **39**: 299.

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. Série III, Sciences de la vie*, **301**: 173–176.

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

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

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. (2015). Medical and Veterinary Impact of the Urticating Processionary Larvae. In: *Processionary moths and climate change: an update* (ed. A. Roques). Springer-Quae, Dordrecht: 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.

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.

Wang Z, Gerstein M. & Snyder M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics* **10**(1): 57-63.

Web tools consulted

BLAST
http://blast.ncbi.nlm.nih.gov/Blast.cgi

Clustal W2
http://www.ebi.ac.uk/Tools/msa/clustalw2/