Neurobiology, Physiology, Biochemistry

Proteome Analysis of Urticating Setae From *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae)

L. Berardi,[1] M. Pivato,[1] G. Arrigoni,[2,3] E. Mitali,[1] A. R. Trentin,[1] M. Olivieri,[4] C. Kerdelhué,[5] F. Dorkeld,[5] S. Nidelet,[5,6] E. Dubois,[6] A. Battisti,[1,7] and A. Masi[1]

[1]Department of Agronomy Food Natural Resources Animals and Environment (DAFNAE), University of Padova, Viale dell'Universita 16, 35020 Legnaro, Padova, Italy (laura.berardi2@gmail.com; micaela.pivato@gmail.com; ettore.mitali@gmail.com; annarita.trentin@unipd.it; andrea.battisti@unipd.it; antonio.masi@unipd.it), [2]Proteomics Center, University of Padova and Azienda Ospedaliera di Padova, via G. Orus 2/B, 35129 Padova, Padova, Italy (giorgio.arrigoni@unipd.it), [3]Department of Biomedical Sciences, University of Padova, via U. Bassi 58/B, 35131 Padova, Padova, Italy, [4]Occupational Medicine, University Hospital of Verona, Piazzale L.A. Scuro 10, 37134 Verona, Verona, Italy (mario.olivieri@ospedaleuniverona.it), [5]INRA, UMR CBGP (INRA/CIRAD/IRD/Montpellier Supagro), 755 avenue du Campus Agropolis, CS30016, F-34988 Montferrier-sur-Lez Cedex, Montferrier-sur-Lez, France (carole.kerdelhue@inra.fr; franck.dorkeld@inra.fr; sabine.nidelet@inra.fr), [6]MGX-Montpellier GenomiX, Institut de Génomique Fonctionnelle, 141 rue de la cardonille, 34094 Montpellier Cedex 05, Montpellier, France (emeric.-dubois@mgx.cnrs.fr), and [7]Corresponding author, e-mail: andrea.battisti@unipd.it

Abstract

Thaumetopoea pityocampa (Denis & Schiffermüller) (Lepidoptera: Notodontidae) is harmful to conifer trees because of defoliation and to public health because of the release of urticating setae from the caterpillars. Contact with setae by humans and domestic animals induces dermatitis, usually localized to the exposed areas. Recent studies demonstrated the presence of a complex urticating mechanism where proteins present in the setae may play a role as activators of immune responses. Yet, limited information is available at present about the proteins occurring in the setae of *T. pityocampa*. Using a refined method for protein extraction from the setae, and a combination of liquid chromatography tandem-mass spectrometry (LC-MS/MS), de novo assembly of transcriptomic data, and sequence similarity searches, an extensive data set of 353 proteins was obtained. These were further categorized by molecular function, biological process, and cellular location. All the 353 proteins identified were found to match through BLAST search with at least one Lepidoptera sequence available in databases. We found the previously known allergens Tha p 1 and Tha p 2 described from *T. pityocampa*, as well as enzymes involved in chitin biosynthesis, one of the principal components of the setae, and serine proteases that were responsible for inflammatory and allergic reactions in other urticating Lepidoptera. This new proteomic database may allow for a better understanding of the complexity of allergenic reactions due to *T. pityocampa* and to other Lepidoptera sharing similar defense systems.

Key words: caterpillar, hair, allergen, transcriptome

Thaumetopoea pityocampa (Denis & Schiffermüller, 1775) (Lepidoptera: Notodontidae), a common pest of conifer trees, is also a threat to public health in southern Europe, the Middle East, and northern Africa (Roques 2015). Caterpillars feed on needles of different species of pine and cedar and use special hairs (setae) to protect themselves against predators; when disturbed, they release the setae that can be blown by the wind over several kilometers, and persist in the environment for a long time (Battisti et al. 2017). The nature of the setae is thus very different from other defensive hairs described from caterpillars of some Lepidoptera, which are part of the integument and require direct contact with the caterpillar to cause the reaction. However, the urticating setae of *T. pityocampa*

are similar to those produced by other Lepidoptera (e.g., the brown tail moths *Euproctis* spp. and the *Hylesia* spp. moths) and some American spiders (Theraphosidae) (Battisti et al. 2011; Fig. 1).

The seta penetration in the skin and the mucosa is guided by hook-like structures, and it results in strong and painful reactions in humans (Fig. 1). Domestic and farm animals can be also affected, either through direct contact with caterpillars or their tents, or by ingesting fodder contaminated with setae (Battisti et al. 2017). In Europe, the incidence in humans in outbreak areas was as high as 18%, and 60% of veterinary practitioners in France had experience of reaction to setae (Moneo et al. 2015). Among the forestry workers of Verona (Italy), chainsaw operators were
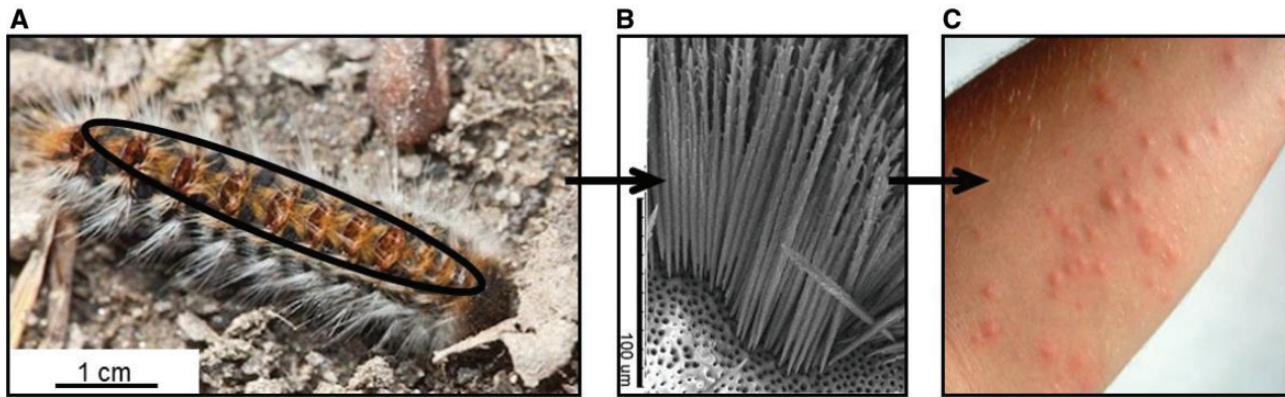
Fig. 1. *Thaumetopoea pityocampa*: (**A**) caterpillar L5 instar, the oval includes the integument dorsal areas of the abdomen where the urticating setae occur; (**B**) close-up of urticating setae, with a part of them removed from their sockets; (**C**) typical reaction on human skin.

identified as being most at risk of skin and respiratory disorders, suggesting the need for a thorough medical surveillance of this occupational allergy (Olivieri et al. 2012). In Australia, where the setae produced by *Ochrogaster lunifer* Herrich-Schäffer are associated with equine amnionitis and fetal loss on horse farms (Perkins et al. 2016), the cost of a similar moth species to equine health has been estimated at AU\$27–43 million per annum (Carrick et al. 2014).

Proteins associated with the urticating setae of Lepidoptera have been described by Bleumink et al. (1982) for the caterpillars of *Euproctis* spp. (Lepidoptera: Erebidae) and by Lundberg et al. (2007) for the female moths of *Hylesia metabus* (Cramer) (Lepidoptera: Saturniidae). In both cases, an enzymatic activity (serine protease) seems to occur and appears to be responsible for the typical reaction associated with the contact of human skin with the setae. The presence of several serine proteases has been confirmed in recent proteomic studies by Cabrera et al. (2016, 2017) in the setae isolated from the female moths of *H. metabus*, which are putatively thought to protect the eggs and the young caterpillars from vertebrate predation. This family of proteases represents 65% of the total proteins extracted from the setae of *H. metabus*.

Proteins associated with setae of *T. pityocampa* were first described by Lamy et al. (1985, 1986). Setae were immersed in liquid nitrogen and then crushed in a saline solution. In total, 16 proteins were detected, and one of them, a 28 kDa protein composed of two subunits of 13 and 15 kDa, was recognized as an antigen and called thaumetopoein. Later, Moneo et al. (2003) obtained a whole caterpillar extract after maceration in a saline solution, agitation for 24 h, and precipitation of the supernatant in ethanol, from which they were able to identify a 15 kDa protein named Tha p 1 because of its putative allergenic potential. Rodríguez-Mahillo et al. (2012) extracted proteins from setae immersed in PBS, followed by sonication on ice. This method allowed the identification of about 70 proteins, seven of which were recognized by sera of exposed persons. One of them (14 kDa) was found to be a major allergen different from Tha p 1 and was then called Tha p 2. Cultured human lymphocytes of persons previously exposed to a closely related species of *Thaumetopoea pinivora* (Treitschke), proliferated when in contact with setae or their extracts, indicating the occurrence of a cell-mediated immune response (Holm et al. 2014). It has also been speculated that chitin, considered to be associated with proteins in the integument (Andersen et al. 1994), may have an adjuvant role in immune reactions (Battisti et al. 2011). Using a protocol adapted from Zhang and Zhu (2006), it was possible to show that chitin makes up to 37% of the urticating setae of *T. pityocampa* (A.B.,

unpublished data). This value suggested that a considerable part of the remaining percentage might be composed of proteins.

Genomic and transcriptomic data are valuable resources to finely analyze the sequences of putatively important proteins, and to search for homologies in model species or in other well-studied organisms. Concerning *T. pityocampa*, a partial transcriptome based on 454 sequencing technology was recently released (Gschloessl et al. 2014). Yet, it was acknowledged that the corresponding data were incomplete and that many transcripts were probably missing, in particular the ones with low levels of expression, or for which expression was restricted to specific organs and tissues and thus possibly undersampled. More, the corresponding resource was obtained from a peculiar population of *T. pityocampa* showing a shifted phenology and occurring only in Portugal. Obtaining an extensive transcriptomic resource from different life stages and sequencing using a high-throughput technology is thus now required.

The aims of this study were to de novo assemble and release a high-quality specific transcriptome and to make a survey of the protein composition of the urticating setae of the caterpillars of *T. pityocampa*. A proteomic characterization of the setae would help to identify the putative allergen(s) and elucidate the mechanisms of the allergic reaction.

Materials and Methods

Transcriptome Analysis

We produced a de novo transcriptome assembly for *T. pityocampa*, obtained from RNA of two *T. pityocampa* populations from northern Italy (Cimolais, 12° 27′ E, 46° 19′ N; Tregnago, 11° 09′ E, 45° 30′ N) reared under laboratory conditions (Berardi et al. 2015b). Note that one of these populations was the same as for the proteomic analysis of the setae (see below). Caterpillars, pupae, and adults were sampled during the rearing and preserved in RNAlater (Thermo Fisher Scientific, Pittsburgh, CA) at −80 °C. A sample of eight caterpillars from different instars, four pupae, and four adults was used for the RNA extraction using a Trizol extraction procedure. RNA quality was checked through migration on an agarose gel, and RNA purity was evaluated by NanoDrop analysis. Whenever the 260/280 OD ratio was <1.7, samples were purified using the Qiagen RNeasy mini kit. RNA concentrations were estimated using the Qubit procedure (Quant-it RNA assay kit, Thermo Fisher Scientific, www.thermofisher.com, accessed 10 June 2017). Fifteen libraries were constructed using the TruSeq stranded mRNA sample prep kit (Illumina, ref. RS-122-2101, San Diego, CA), according to the manufacturer's instructions. Briefly, poly-A RNAs were purified

using oligo-d(T) magnetic beads. The selected RNAs were fragmented and reverse transcribed using random hexamers, SuperScript II (Life Technologies, ref. 18064-014, Thermo Fisher Scientific) and Actinomycin D. Double-stranded cDNAs were adenylated at their 3′ ends and ligation was then performed using Illumina's indexed adapters. Ligated cDNAs were finally amplified by PCR. Libraries were validated using a DNA1000 chip (Agilent, ref. 5067-1504, Santa Clara, CA) on an Agilent Bioanalyzer and quantified by qPCR. The 15 libraries were sequenced using 3 lanes of an Illumina HiSeq2000, using the paired end protocol (2 by 100 nt) at the Montpellier GenomiX (MGX) sequencing core facility. Image analysis and base calling were performed using the HiSeq Control Software and Real-Time Analysis component provided by Illumina. Demultiplexing was performed using Illumina's sequencing analysis software (CASAVA 1.8.2). The quality of the data was assessed using FastQC and FastQ Screen from the Babraham Institute, and the Illumina software SAV (Sequence Analysis Viewer).

Reads produced by high-throughput sequencing must be preprocessed to retain only high-quality reads used in reconstruction. The raw sequences were trimmed using Trimmomatic (Bolger et al. 2014) using the following parameters: ILLUMINACLIP: adaptors_file.fa: 2: 40: 15; HEADCROP: 12; SLIDINGWINDOW: 4: 15 and MINLEN: 30 combined with PRINSEQ-lite (Schmieder and Edwards 2011) to eliminate polyA tails. The cleaned reads, obtained after the quality filtering step described above, were used for de novo transcript assembly using Trinity (Grabherr et al. 2011), with the normalization option on and default kmer value. The completeness of the transcriptome assembly was assessed using the Core Eukaryotic Genes Mapping Approach (CEGMA, Parra et al. 2007). After alignment with hidden Markov models, we thereby obtained the percentage of core genes that were found complete or partial in the dataset (i.e., transcript reconstruction). To further evaluate the quality of the assembly, we used the following metrics: number of transcripts assembled, total length of the assembly (in bp), mean transcript length, N50 value, and number of transcripts longer than 1 kb. We mapped the filtered reads back to assembled transcripts, and we also used the number of such reads (RBMT) as a quality assessment estimator. We used TransDecoder, provided by the Trinity package, to identify potential coding regions from the assembled transcripts. TransDecoder identifies candidate protein-coding regions based on nucleotide composition, open reading frame (ORF) length, and Pfam domain content.

Setae Collection

Twenty tents containing colonies of *T. pityocampa* were collected in March 2013 from *Pinus nigra* trees in Tregnago (the same site from which individuals were collected for the transcriptome analysis described above). Each caterpillar (either in the fourth or fifth instar) was removed from the tent in a vented hood and put in an Eppendorf tube at −20°C. The setae were then removed from thawing caterpillars with forceps using a stereomicroscope at 40× magnification, inside a hood, with special attention to avoid wounds that could contaminate the setae with body fluids. If such a contamination occurred, the material was excluded from subsequent analysis. Because one caterpillar carries about 1 million setae (Petrucco-Toffolo et al. 2014), by collecting all the setae in one caterpillar from each colony, it was estimated that the sample consisted of about 20 million urticating setae. Setae were collected into Eppendorf tubes with 100 µl of acetone, which allows a quick detachment of the setae from the forceps. Setae were frozen in liquid nitrogen and stored in Eppendorf tubes at −20°C for a maximum of 1 wk before protein extraction (Fig. 2).
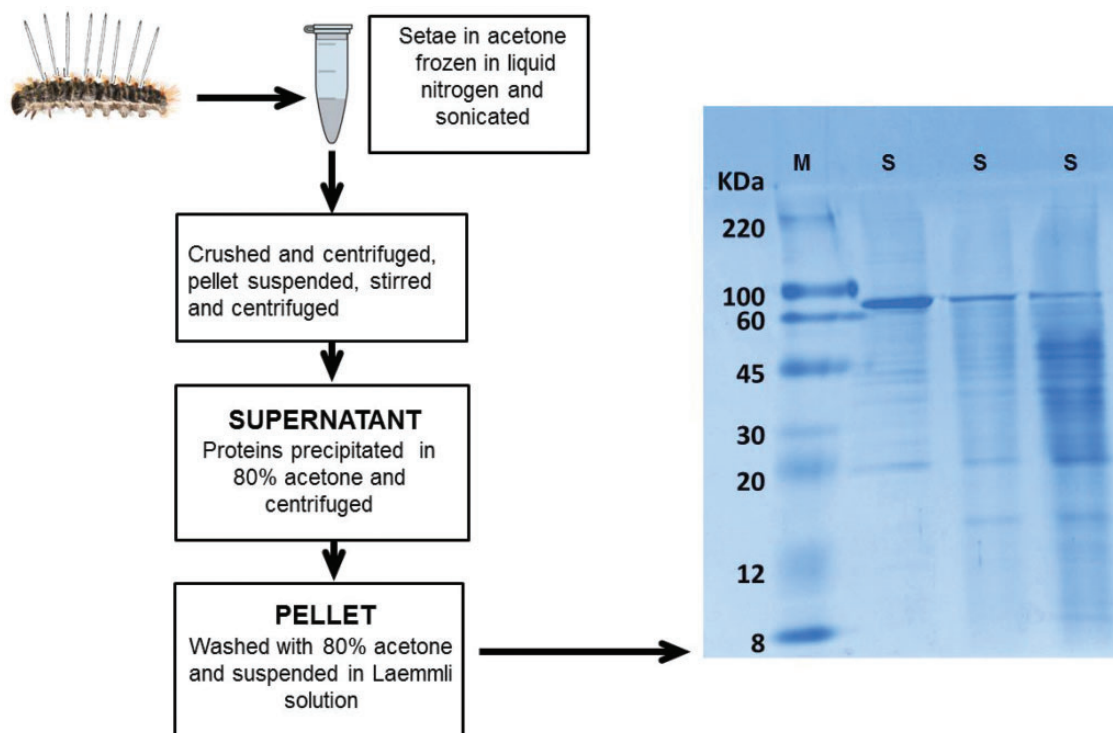
Protein Extraction and Digestion

Setae in Eppendorf tubes with acetone were frozen in liquid nitrogen and sonicated on ice three times, 3 min each at 18 W, then manually crushed and centrifuged for 2 min at 28,672 g-force. Following acetone removal, the pellet was put in 1.5 ml extraction buffer (50 mM Hepes pH 8, 1% Triton X-100, 1 M NaCl, and 1 mM PMSF/Benzamidine). The extract was then stirred for 4 h at room temperature to allow protein solubilization, and centrifuged for 20 min at 21,952 g-force at 4°C. One part of the clear supernatant was transferred in another tube and four parts of ice-cold acetone were added (final 80% acetone concentration); following overnight protein precipitation at −20°C, samples were centrifuged again at 16,128 g-force for 10 min. The supernatant was discarded, while the pellet was washed with 80% cold acetone and 20% water and suspended in Laemmli (1970) solution. Protein content was measured according to the Bradford method (Kruger 1994). Electrophoresis was carried out on a 13% acrylamide gel, and proteins were visualized by Coomassie staining (Fig. 2). The lanes were cut into bands, which were then placed in separate Eppendorf tubes, washed several times with 50 mM ammonium bicarbonate pH 8.0 and dried under vacuum after a short wash with 100% acetonitrile. Cysteines were reduced with 10 mM freshly prepared dithiothreitol in 50 mM ammonium bicarbonate for 1 h at 56°C, and alkylated with 55 mM iodoacetamide in 50 mM ammonium bicarbonate for 45 min at room temperature in the dark. Gel pieces were then washed with alternate steps of ammonium bicarbonate and acetonitrile, and finally dried under vacuum. Proteins were digested in situ with sequencing grade modified trypsin (Promega, Madison, WI) at 37°C overnight (12.5 ng/µl trypsin in 50 mM ammonium bicarbonate). Peptides were extracted with three steps of 50% acetonitrile/0.1% formic acid. Extracted peptides were dried under vacuum and stored at −20°C until liquid chromatography tandem-mass spectrometry (LC-MS/MS) analysis was performed.

LC-MS/MS Analysis

The extract was suspended in H_2O/0.1% formic acid and analyzed by LC-MS/MS (Tolin et al. 2012). The MS analyses were conducted with a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) coupled online with a nano-HPLC Ultimate 3000 (Dionex—Thermo Fisher Scientific). Samples were loaded onto a 10-cm chromatographic column packed into a pico-frit (75 µm id, 15 µm tip, New Objectives) with C18 material (ReproSil, 300 A°, 3 µm). Peptides were separated at a flow rate of 250 nl/min using a linear gradient of acetonitrile/0.1% formic acid from 3% to 50% in 30 min. Spray voltage was set to 1.3 kV with an ion source capillary temperature of 200°C. The instrument was operated in data-dependent mode with a top-ten acquisition method (a full scan at 60,000 resolutions on the Orbitrap followed by MS/MS scans of the 10 most intense ions acquired in the linear trap). 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, United Kingdom). Raw LC-MS/MS spectra were searched against the transcriptome database obtained for *T. pityocampa* and concatenated with a database of the most common protein contaminants found in proteomics experiments. Enzyme specificity was set to trypsin with one missed cleavage, and peptide and fragment tolerance was set to 10 ppm and 0.6 Da, respectively. Carbamidomethylation of cysteine residues and methionine oxidation were set as fixed and variable modification, respectively. Based on the search against the corresponding randomized database, the algorithm Percolator was used to assess false discovery rates (FDR):

Fig. 2. Experimental flow chart for protein extraction from *T. pityocampa* urticating setae and SDS-PAGE separation. All bands of three samples (S) in stained gels were cut in bands which were then subjected to in-gel trypsin digestion followed by LC-MS/MS analysis. M—marker.

only proteins identified with a q value <0.01 (99% confidence) and at least two unique peptides were considered as positive hits. Data were prefiltered to exclude MS/MS spectra containing <5 peaks or with a total ion count <50.

Bioinformatics Analyses

Blast2GO 3.1 (Conesa et al. 2005) was used to analyze gene ontology (GO at UniProtKB, www.uniprot.org, accessed 10 June 2017) terms after BLAST, mapping, and annotation of the amino acid sequences. The BLAST analyses against the NCBI BioSystems database permitted the identification of possible homologous proteins. Protein sequences found in the LC-MS/MS analysis were converted in FASTA format and analyzed (blastp) with the threshold E-value cutoff at 1.0E-3 and number of BLAST hits, 4. Mapping and annotation were performed with the default values, and after that, protein sequences were assigned into three standard classifications: molecular function, biological process, and cellular component, and summarized according to GO (Gene Ontology) criteria. The metabolic pathways were analyzed with KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa et al. 2012) implemented in the Blast2GO program.

Results

For each library (i.e., each development stage), from 12 to 37 million clean read pairs were obtained after filtering. The total number of quality filtered reads was 532,900,683. The de novo transcriptome assembly obtained using Trinity consisted of 233,098 transcripts that were clustered into 188,610 potentially different genes (i.e., unigenes). In particular, the high number of core genes identified with the CEGMA procedure (94.35% when considering only complete genes, and 98.39% when considering partial and complete genes), and the high proportion for RBMT (75%) suggested that the reconstruction was rather robust. Transcripts reconstructed with Trinity were processed using TransDecoder, which retrieved 41,933 coding sequences and 24,992 complete peptides.

The protein electrophoresis revealed the presence of many bands, suggesting a relative complex protein composition even if the proteome was only obtained from a fraction of larval tissues, i.e., setae (Fig. 2). The LC-MS/MS analysis and database search against the reference transcriptome identified a total of 353 proteins. All 353 proteins identified were found to match through BLAST search with at least one Lepidoptera sequence available in the NCBI database. Among the 353 identified proteins, 120 showed a high degree of identity to *Bombyx mori* (L.) proteins, 34 to *Papilio xuthus* L., 32 to *Danaus plexippus* (L.), and 11 to *Papilio polytes* L. proteins. The remaining proteins were associated with other Lepidoptera (Supp. Table 1 [online only]). The BLAST analyses identified the proteins Tha p 1 and Tha p 2 of *T. pityocampa* (Supp. Table 1 [online only]).

Based on GO functional annotations and categories, 224 of the 353 identified proteins were annotated with at least one common GO. The cellular location graph indicated that the majority of urticating setae proteins belonged to the categories of cell, cell part, organelle, macromolecular complex, and membrane (Fig. 3A). A small percentage corresponded to membrane proteins and extracellular regions. Within the classification of biological process, a large number of proteins were assigned to metabolic and cellular or single-organism process (Fig. 3B). The Blast2GO analyses detected two principal functions: catalytic and binding role (Fig. 3C). Looking deeper into the KEGG pathway analyses, enzymes were mapped onto different metabolic pathways, showing that particular proteins were likely to be involved in gluconeogenesis and glutathione metabolism, like glucose-6-phosphate isomerase (5.3.1.9) or
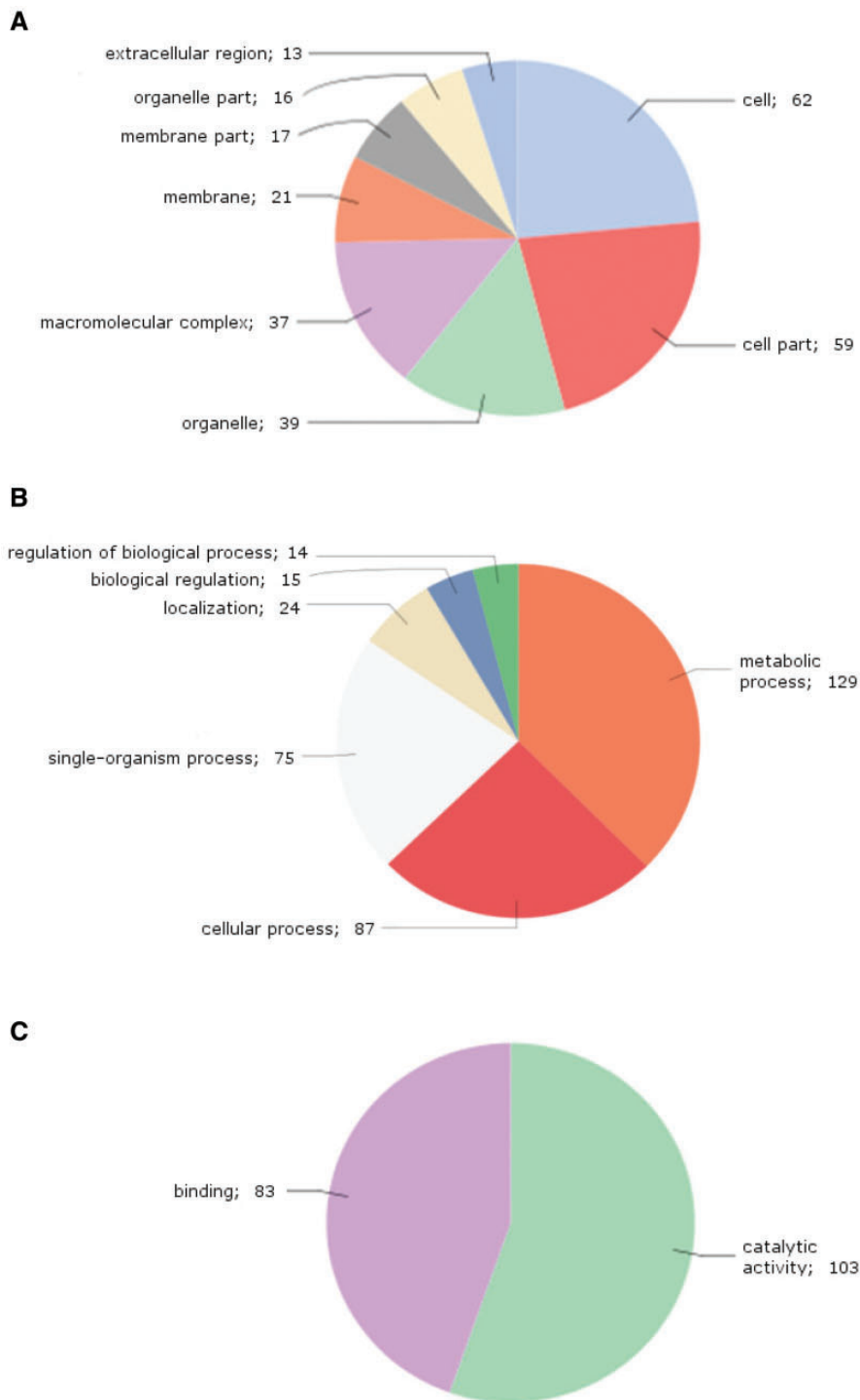
Fig. 3. Proteins as cellular component **(A)**, biological process **(B)**, and molecular function **(C)**. Numbers indicate the proteins within each category (see text for details).

glyceraldehyde-3-phosphate dehydrogenase (1.2.1.12). Members of the chitin production pathway were significantly overrepresented, as well as several serine proteases.

Discussion

We here provide a comprehensive transcriptome assembly for Italian populations of *T. pityocampa*. As the expressed transcripts were obtained from all developmental stages (from larva to adult, through most caterpillar instars and pupae), and deep sequenced using Illumina technology, we obtained a very valuable resource to analyze the resulting proteomic data. All quality assessments suggested that the transcriptomic resource produced here was of good quality and reached a high level of completeness, as suggested by the high proportion of core genes identified through the CEGMA analysis.

Because of the production of this transcriptomic database of *T. pityocampa*, new information about the quality and number of proteins associated with the setae was obtained. The large number of identifications is striking, if compared with the previous work from Rodríguez-Mahillo et al. (2012), where authors demonstrated that the setae contain a complex mixture of about 70 proteins and showed that one of them, Tha p 2, is allergenic in humans. This protein can be now considered as a major component of the seta proteome in *T. pityocampa*, as judged by the high score and the high number of peptide spectrum matches (PSMs) obtained from the LC-MS/MS analyses (see Supp. Table 1 [online only]). It is a glycine-, serine-, and cysteine-rich protein highly conserved among processionary moths (Berardi et al. 2015a). Another protein, called Tha p 1, was previously described as an allergen (Moneo et al. 2003), although it was obtained from an extract of the whole body, and it was not possible to know whether or not it was a setae protein. A later matching of a partial sequence of this protein indicated a possible similarity with chemosensory proteins of other insect species (Larsson and Backlund 2009; Moneo et al. 2015), although our finding associates it with setae, which do not have sensory functions (Battisti et al. 2011). The isolation of this putative allergen and its sequencing would be required to clarify the nature of Tha p 1.

The presence of a larger number of intracellular rather than extracellular proteins may have a dual explanation. First, cytoplasmic proteins of the epidermal cells producing the urticating setae are likely to be stored in the setae as remnants of this biosynthetic process. It is noteworthy that setae formation represents a major metabolic effort, leading to a large number of setae that have to be formed in a short time during each molt (up to 1,000,000 in a mature caterpillar of the *T. pityocampa*, Petrucco-Toffolo et al. 2014, and up to 2,000,000 in the bag-shelter moth *O. lunifer*, Perkins et al. 2016). Alternatively, it could be that hemolymph proteins can pass through membranes and enter the epidermis and, in turn, into the setae.

Because serine proteases have been found to have an important role in mediating the reactions of exposed humans and guinea pigs to the urticating setae of the caterpillars of *Euproctis* spp. (Bleumink et al. 1982) and of the female moths of *H. metabus* (Lundberg et al. 2007; Cabrera et al. 2016, 2017), it would be interesting to assess their role in the *T. pityocampa* system as well. Serine proteases do occur in the proteome of this species and show some similarity with proteins occurring in *Bombyx*, *Helicoverpa*, and *Mamestra* (Supp. Table 1 [online only]). Interestingly, these proteins are also present in the defensive droplets of Zygaenid caterpillars (Pentzold et al. 2016).

This full proteomic analysis of setae, highlighted by a comprehensive and specific transcriptomic resource, can be considered as a reference protein data set, and it will be of great importance to compare it with future studies on other species. It is likely that among the full list of proteins shown here, there are a few more that are directly responsible for immune reactions, as evidenced by Moneo et al. (2015). Enzymes involved in the chitin synthesis pathway were found in the setae. The chitin role in allergenic reaction is not clear yet, although it has been shown to stimulate in vitro human T-lymphocyte proliferation (Holm et al. 2014) and act as an adjuvant in immune reactions (Battisti et al. 2011). This new proteomic resource may enable a better understanding of the complexity of allergenic reactions due to the *T. pityocampa* and to other Lepidoptera sharing similar defense systems. In the future, there is the need to correlate allergenic reactions to the identified proteins and to extend the extraction of the urticating setae to other species of Lepidoptera. Genomic and transcriptomic resources could be produced for these other species, as next-generation sequencing technologies now allow to relatively easily obtain high-quality resources.

Reference Cited

Andersen, S. O., P. Hojrup, and P. Roepstorff. 1994. Insect cuticular proteins. Ins. Biochem. Mol. Biol. 25: 153–176.

Battisti, A., G. Holm, B. Fagrell, and S. Larsson. 2011. Urticating hairs in arthropods: their nature and medical significance. Annu. Rev. Entomol. 56: 203–220.

Battisti, A., S. Larsson, and A. Roques. 2017. Processionary moths and associated urtication risk: global-change driven effects. Annu. Rev. Entomol. 62: 323–342.

Berardi, L., A. Battisti, and E. Negrisolo. 2015a. The allergenic protein Tha p 2 of processionary moths of the genus *Thaumetopoea* (Thaumetopoeinae, Notodontidae, Lepidoptera): characterization and evolution. Gene 574: 317–324.

Berardi, L., M. Branco, M. R. Paiva, H. Santos, and A. Battisti. 2015b. Development time plasticity of the pine processionary moth (*Thaumetopoea pityocampa*) populations under laboratory conditions. Entomologia 3: 273.

Bleumink, E., M.C.J.M. de Jong, F. Kawamoto, G. T. Meyer, A. J. Kloosterhuis and I. J. Slijper-Pal. 1982. Protease activities in the spicule venom of *Euproctis* caterpillars. Toxicon 20: 607–613.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Cabrera, G., V. Salazar, R. Montesino, Y. Támbara, W. B. Struwe, E. Leon, D. J. Harvey, A. Lesur, M. Rincón, B. Domon, et al. 2016. Structural characterization and biological implications of sulfated N-glycans in a serine protease from the neotropical moth *Hylesia metabus*, (Cramer [1775]) (Lepidoptera: Saturniidae). Glycobiology 26: 230–250.

Cabrera, G., U. Lundberg, A. Rodríguez-Ulloa, M. Herrera, W. Machado, M. Portela, S. Palomares, L. A. Espinosa, Y. Ramos, R. Durán, et al. 2017. Protein content of the *Hylesia metabus* egg nest setae (Cramer [1775]) (Lepidoptera: Saturniidae) and its association with the parental investment for the reproductive success and lepidopterism. J. Proteomics 150: 183–200.

Carrick, J. B., N. R. Perkins, and M. P. Zalucki. 2014. Causes of abortion in Australia (2005–2012)—proportion of cases due to equine amnionitis and fetal loss (EAFL). J. Equine Vet. Sci. 34: 212–214.

Conesa, A., S. Götz, J. M. García-Gomez, J. Terol, M. Talón, M. Robles. 2005. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Gschloessl, B., H. Vogel, C. Burban, D. Heckel, R. Streiff, and C. Kerdelhué. 2014. Comparative analysis of two phenologically divergent populations of the pine processionary moth (*Thaumetopoea pityocampa*) by de novo transcriptome sequencing. Insect Biochem. Mol. Biol. 46: 31–42.

Holm, G., M. Andersson, M. Ekberg, B. Fagrell, J. Sjöberg, M. Bottai, and M. Björkholm. 2014. Setae from larvae of the northern processionary moth (*Thaumetopoea pinivora*, TP) stimulate proliferation of human blood lymphocytes in vitro. PLoS One 9: e113977.

Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40: D109–D114.

Kruger, N. J. 1994. The Bradford method for protein quantization. Methods Mol. Biol. 32: 9–15.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685.

Lamy, M., M. H. Pastureaud, and G. Ducombs. 1985. Thaumetopoein, an urticating protein of the processional setae of the caterpillar (*Thaumetopoea pityocampa* Schiff) (Lepidoptera, Thaumetopoeidae). C. R. Acad. Sci. III 301: 173–176.

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

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

Lundberg, U., V. Salazar, M. Tovar, and J. Rodríguez. 2007. Isolation and partial characterization of proteins with vasodegenerative and proinflammatory properties from the egg nests of *Hylesia metabus* (Lepidoptera: Saturniidae). J. Med. Entomol. 44: 440–449.

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

Moneo, I., A. Battisti, B. Dufour, J. C. Garcia-Ortiz, M. Gonzalez-Muñoz, F. Moutou, P. Paolucci, E. Petrucco Toffolo, J. Rivière, A.-I. Rodríguez-Mahillo, et al. 2015. Medical and veterinary impact of the urticating processionary larvae, pp. 359–410. *In* A. Roques (ed.), Processionary moths and climate change: an update. Springer-Quae, Dordrecht, NL.

Olivieri, M., G. L. Lazzarini, I. Goio, S. Schinella, L. Romeo, and L. Perbellini. 2012. Occupational risk by the pine processionary moth *Thaumetopoea pityocampa* in the forestry workers of Verona. G. Ital. Med. Lav. Ergon. 34: 420–422.

Parra, G., K. Bradnam, and I. Korf. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061–1067.

Pentzold, S., M. Zagrobelny, B. Khakimov, S. B. Engelsen, H. Clausen, B. L. Petersen, J. Borch, B. L. Møller, and S. Bak. 2016. Lepidopteran defence droplets—a composite physical and chemical weapon against potential predators. Sci. Rep. 6: 22407.

Perkins, L. E., M. P. Zalucki, N. R. Perkins, A. J. Cawdell-Smith, and B. W. Cribb. 2016. Urticating setae of *Ochrogaster lunifer*, an Australian processionary caterpillar of veterinary importance. Med. Vet. Entomol. 30: 241–245.

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

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

Roques, A. 2015. Processionary moths and climate change: an update. Springer-Quae, Dordrecht NL.

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864.

Tolin, S., G. Pasini, B. Simonato, F. Mainente, and G. Arrigoni. 2012. Analysis of commercial wines by LC-MS/MS reveals the presence of residual milk and egg white allergens. Food Control 28: 321–326.

Zhang, J., and K. Y. Zhu. 2006. Characterization of a chitin synthase cDNA and its increased mRNA level associated with decreased chitin synthesis in *Anopheles quadrimaculatus* exposed to diflubenzuron. Insect Biochem. Mol. Biol. 36: 712–725.