Gene 574 (2015) 317–324

Contents lists available at ScienceDirect

Gene

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

Research paper

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

Laura Berardi [a], Andrea Battisti [a], Enrico Negrisolo [b,*]

[a] Department of Agronomy Food Natural Resources Animals and Environment (DAFNAE), University of Padova, Agripolis, 35020 Legnaro, PD, Italy
[b] Department of Comparative Biomedicine and Food Science (BCA), University of Padova, Agripolis, 35020 Legnaro, PD, Italy

ARTICLE INFO

Article history:
Received 22 January 2015
Received in revised form 16 June 2015
Accepted 10 August 2015
Available online 11 August 2015

Keywords:
Allergenic protein Tha p 2
Processionary moths
Secondary structure prediction
Molecular evolution
Thaumetopoea
Thaumeotopoeinae

ABSTRACT

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

1. Introduction

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

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

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

Setae, like the insect integument, are composed of a chitin skeleton with a matrix of proteins. While the role of chitin is still under investigation, the presence of several proteins with allergic activity against humans has been demonstrated in *Thaumetopoea* (Lamy et al., 1985, 1986; Rodríguez-Mahillo et al., 2012). Three proteins with allergenic activity have been identified for processionary moths. The first protein was named thaumetopoein and is estimated to comprise two subunits with molecular weights of 13 and 15 kDa (Lamy et al., 1986). Unfortunately, thaumetopoein was not sequenced after its discovery. Successively, a second protein named Tha p 1 with an estimated molecular weight of 15 kDa was isolated by Moneo et al. (2003), who published a polypeptide encompassing 18 residues and successively deposited the full-length sequence into GenBank (ADK47876). Larsson and Backlund (2009) showed that the Tha p 1 polypeptide was related to chemosensory proteins found in insects. Finally, a third protein named Tha p 2, unrelated to Tha p 1, was identified in *Thaumetopoea pityocampa*, and the cDNA encoding the complete polypeptide was

Abbreviations: Tha p 2, Tha p 2 protein; ORF, Open Reading Frame; SNPs, single nucleotide polymorphisms; BT, nonparametric bootstrap; mat-Tha p 2, final mature Tha p 2 protein.

* Corresponding author.
E-mail addresses: laura.berardi@studenti.unipd.it (L. Berardi), andrea.battisti@unipd.it (A. Battisti), enrico.negrisolo@unipd.it (E. Negrisolo).

http://dx.doi.org/10.1016/j.gene.2015.08.019
0378-1119/

amplified and sequenced (Rodríguez-Mahillo et al., 2012). The acronym Tha p 2 used to identify the last protein is unfortunate because it induces confusion with a large group of completely unrelated proteins named for the possession of the THAP domain (see Roussigne et al., 2003). Notwithstanding this potential problem, we do not suggest a name change for this protein, given that the stability of allergen nomenclature is a high priority (see www.allergen.org).

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

Most Thaumetopeinae moths live in countries that are difficult to explore due to geopolitical reasons. Furthermore, even for species present in countries without political turmoil, sampling can be difficult due to the nocturnal behavior of the adults and to the lack of conspicuous larval assemblages in the field. The genus *Thaumetopoea* is the only Thaumetopoeinae taxon with a distribution centered in the Western Palaearctic zone (i.e., Europe, Middle East and North Africa). Furthermore, the larvae of this genus have a gregarious nesting behavior that makes them conspicuous and relatively easily sampled in the field (Battisti et al., 2011).

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

2. Materials and methods

2.1. Taxon sampling

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

2.2. DNA extraction

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

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

For cloning purposes (see below), the DNA was concentrated by low-speed centrifugation using Amicon® Ultra-0.5 Centrifugation Filter Devices.

2.3. PCR amplification and sequencing of Tha p 2 gene

Initially, the amplification of the *Tha p 2* gene was performed using the primers TP2 DF, 5′-GTC CCG CAA CTA AGT GAG AAA GC-3′ (forward) and TP2 DR 5′-GGT CCT TGT TCG GCC TAG TAA-3′ (reverse) (Rodríguez-Mahillo et al., 2012). The PCR assay was optimized for a 20 μl volume. The reaction contained 5× PCR buffer, 25 mM $MgCl_2$, 25 mM dNTPs, 10 μM forward primer and 10 μM reverse primer, 5 u/μl *Taq* polymerase, H_2O and 2 μl of DNA. The PCR program consisted of a hot-start at 96 °C for 5′, followed by 35 cycles of 96 °C for 45″, 52 °C for 30″, 72 °C for 30″ and a final extension at 72 °C for 5′.

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

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

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

The DNA portions located upstream and downstream of the published *Tha p 2* sequence were amplified using the GenomeWalker kit

Table 1
List of the species analyzed in the present work, with indication of the collection data.

Family	Species	Site collection	Latitude	Longitude	Date	Legit
Bombycidae	*Bombyx mori*	Padova, Italy	–	–	11/2012	A. Saviane
Lymantridae	*Euproctis chrysorrhoea*	Valencia, Spain	39° 50′ N	1° 17′ E	01/2013	E. Frago
Lymantriidae	*Lymantria dispar*	Sassari, Sardegna, Italy	40° 44′ N	8° 36′ E	11/2012	P. Luciano
Notodontidae	*Anaphe panda*	Kakamega Forest, Western Kenia	0° 10′ N	34°47′ E	11/2012	N. Mbahin
Notodontidae	*Ochrogaster lunifer*	Kenmore (Queensland), Australia	27° 30′ S	152° 56′ E	25/02/2005	M.P. Zalucki
Notodontidae	*Thaumetopoea bonjeani*	Forest of Chélia (Khenchela), Algeria	35° 22′ N	6° 46′ E	09/04/2006	M. Zamoum
Notodontidae	*Thaumetopoea herculeana*	Cabo home, Spain	41° 51′ N	7° 21′ E	04/04/2009	L. Berardi
Notodontidae	*Thaumetopoea ispartaensis*	Kapidagi, Senirkent, Isparta, Turkey	37° 59′ N	30° 36′ E	05/2006	M. Avci
Notodontidae	*Thaumetopoea libanotica*	Tannourine et Tahta, Lebanon	34° 12′ N	35° 55′ E	03/06/2006	N. Nemer, C. Lahousl
Notodontidae	*Thaumetopoea pinivora*	Gotland, Sweden	56° 18′ N	18° 13′ E	15/04/2006	S. Larsson
Notodontidae	*Thaumetopoea pityocampa*	Moggio, Udine, Italy	46° 24′ N	13° 12′ E	12/03/2003	A. Battisti
Notodontidae	*Thaumetopoea pityocampa* ENA (Eastern-Northern Africa)	Bizerte, Tunisia	37° 02′ N	9° 42′ E	11/2000	M. El Habib Ben Jamâa
Notodontidae	*Thaumetopoea processionea*	Caprino Veronese, Verona, Italy	45° 34′ N	10° 46′ E	21/5/2007	M. Faccoli, M. Zampini
Notodontidae	*Thaumetopoea solitaria*	Gamla (Golan heights), Israel	32° 59′ N	35° 42′ E	13/3/2005	A. Battisti
Notodontidae	*Thaumetopoea wilkinsoni*	Aladag, Turkey	37° 33′ N	35° 22′ E	28/3/2004	A. Battisti

–, laboratory rearing.