

Received Date : 17-Jun-2016

Revised Date : 12-Nov-2016

Accepted Date : 18-Nov-2016

Article type : Original Article: Experimental Allergy and Immunology

**Lip b 1 is a novel allergenic protein isolated from the booklouse,
*Liposcelis bostrychophila***

O. Ishibashi¹, K. Sakuragi¹, Y. Fukutomi², Y. Kawakami³, Y. Kamata⁴, M. Sakurai¹, S. Nakayama¹, H. Uchiyama⁵, H. Kobayashi⁵, H. Kojima⁶ & T. Inui¹

¹Department of Applied Life Sciences, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai, Japan; ²Clinical Research Center for Allergy and Rheumatology, Sagamihara National Hospital, Sagamihara, Japan;

³Laboratory of Environmental Science, FCG Research Institute Incorporation, Tokyo, Japan; ⁴Department of Veterinary Medicine, Iwate University, Morioka, Japan; ⁵NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo, Japan; ⁶Department of Agriculture, Tokyo University of Agriculture, Atsugi, Japan

Short Title: A novel allergen from booklice

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/all.13091

This article is protected by copyright. All rights reserved.

Keywords: booklouse, *Liposcelis bostrychophila*

Correspondence: Takashi Inui, PhD, Laboratory of Biological Macromolecules,
Department of Applied Life Sciences, Graduate School of Life and Environmental
Sciences, Osaka Prefecture University, Sakai, Japan.

Tel.: +81 72 254 9473

Fax: +81 72 254 9474

E-mail: inuit@bioinfo.osakafu-u.ac.jp

Abstract

Background: Booklice, belonging to the order Psocoptera, are small household insect pests that are distributed worldwide. *Liposcelis bostrychophila* (*L. bostrychophila*), a common home-inhabiting species of booklouse, infests old books, sheets of paper and stored food. Recent entomological and serological studies demonstrated that *L. bostrychophila* accounted for the majority of detectable insects in house dust and could be a potent inducer of respiratory allergy. Our recent proteomic analysis identified a potent allergenic protein from *L. bostrychophila*, designated Lip b 1, and determined its partial amino acid sequences.

Methods: Cloning of cDNAs for Lip b 1 was performed by large-scale transcriptome analysis (RNA-seq) and subsequent reverse transcription polymerase chain reaction. The full-length amino acid sequences deduced from Lip b 1 cDNAs were bioinformatically

analyzed. The recombinant proteins of glutathione-S-transferase (GST)-fused Lip b 1 were analyzed by Western blot and enzyme-linked immunosorbent assay.

Results: Lip b 1 cDNAs encoding two types of 254-amino acid proteins were cloned. The clones shared 87% identity, and the deduced molecular weights and isoelectric points were consistent with those determined in our previous study. The two types of Lip b 1 proteins in the GST-fused form were similarly reactive with sera from allergic patients sensitized with *L. bostrychophila*.

Conclusions: Lip b 1 is a novel protein possibly causing booklouse allergy.

Introduction

Booklice, alternatively called psocids, are household insect pests belonging to the order Psocoptera. Members of Psocoptera are distributed in geographical regions ranging from tropical to subarctic zones. Booklice are brownish, soft-bodied insects that are 1.0–1.3-mm long when fully grown. Although considered pests, booklice do not bite humans or animals, transmit diseases, or damage food or fabric. However, booklice grow very rapidly under certain conditions, and can be annoying when present in large numbers (1). Thus, booklice have been considered unsanitary or unpleasant insect pests for decades, causing severe economic losses worldwide due to infestation of stored foods (2).

Liposcelis bostrychophila (*L. bostrychophila*) (Figure S1), considered the most widespread species of booklice, frequently infests stored food (3-7). In addition, a recent entomological study in Japan demonstrated that booklice account for the majority of detectable insects in house dust, presenting at a frequency of 90%~ in dust samples (8).

The frequent detection of booklice in Beijing (9) and Madrid (10) has also been reported.

We previously demonstrated that booklouse-specific IgE antibodies were detected in sera from 22% of Japanese asthmatic patients who participated in our study, with one third showing specific sensitization to booklice (11). A study from India also showed that 20% of patients with a nasobronchial allergy exhibits a strong skin sensitivity to *Psocoptera* spp (12), and a case of a 33 year-old woman with non-occupational allergic asthma to *Liposcelis* spp infesting her house was also reported in France (13). These studies strongly suggest that *L. bostrychophila* detected in house dust could be a potent indoor allergen. We previously identified a 26-kDa protein, designated Lip b 1, as a possible allergenic protein from *L. bostrychophila* (11) and determined its partial amino acid sequences. No proteins containing these partial sequences were found in a database, suggesting that Lip b 1 is a novel allergen. In addition, the analysis of IgE inhibition using various arthropod allergens as inhibitors revealed that IgE reactivity to Lip b 1 was not inhibited by preincubation with other arthropod allergens, indicating that sensitization to this allergen could serve as a marker of genuine sensitization to *L. bostrychophila*, which is independent of insect panallergy.

In the present study, we performed molecular cloning of cDNAs encoding Lip b 1 using large-scale transcriptome analysis (RNA-seq) and subsequent reverse transcription polymerase chain reaction (RT-PCR). Bioinformatics analyses predicted that Lip b 1 could be a novel transmembrane protein. A nucleotide polymorphism was shown in the open reading frames (ORFs) of the Lip b 1 cDNA clones, generating two types of variants that share 87% identity, designated as Lip b 1.0101 and Lip b 1.0102. Furthermore, we produced the recombinant proteins of Lip b 1 as glutathione-S-transferase (GST)-fusion proteins and demonstrated reactivity to sera from asthma patients specifically sensitized with booklice.

This article is protected by copyright. All rights reserved.

Materials and methods

Sera collection and ethical statement

The sera used in this study were collected from 10 consecutive outpatients with allergic asthma who visited Sagamihara National Hospital (Table S1). Informed consent from each patient was provided as described previously (11). These patients showed positivity in the levels of IgE antibodies to *L. bostrychophila*, as determined by our in-house colorimetric ELISA using cyanogen bromide-activated paper discs coupled with the crude extracts of *L. bostrychophila* (>0.70 units/mL), as well as by intradermal testing using these extracts (Table S1). In addition, our previous immunoblot analysis revealed that IgE antibodies in the sera react with a 26-kDa protein in the *L. bostrychophila* extracts. This study was approved by the Ethics Committee of Sagamihara National Hospital.

RNA isolation

L. bostrychophila were reared as described previously (11) and collected in a mortar containing RNA-iso reagent (Takara-bio, Kusatsu, Japan). After quick freezing by adding liquid nitrogen, the booklice bodies were vigorously disrupted using a muddler. The booklouse extracts were transferred to microtubes and subjected to RNA extraction according to the manufacturer's instructions. The quality of isolated total RNA was evaluated by an RNA integrity number (RIN) determined using BioAnalyzer 2100 (Agilent, Santa Clara, CA, USA), and the integrity of 18S and 28S ribosomal RNAs was analyzed by agarose gel electrophoresis. We obtained an RIN of 8.8 and observed clear bands of 28S and 18S ribosomal RNAs, indicating that the isolated RNA had sufficient quality for transcriptome analysis.

RNA-seq

The RNA-seq of *L. bostrychophila* was performed using the HiSeq 2500 sequencing system (Illumina, Hayward, CA, USA) with the aid of the Dragon Genomics Center of Takara Bio Inc. (Otsu, Japan). Briefly, poly-A-harboring RNAs were enriched from the total RNA sample and used to construct a sequence library with a TruSeq Sample Preparation Kit v2 (Illumina, San Diego, CA, USA) according to the instructions of the Directional mRNA-seq Sample Preparation Rev.A. The library was subjected to quality assessment using Bioanalyzer 2100 and sequenced to produce 100-bp single reads. The reads obtained in this analysis were deposited in the DDBJ Sequence Read Archive (DRA) under the accession number PSUB005884. Adaptor sequences and poor quality nucleotides around the 3'-termini were trimmed from the raw reads using Cutadapt software (ver. 1.2.1) (<http://code.google.com/p/cutadapt/>) (14) and Sickle software (<https://github.com/najoshi/sickle>), and were assembled *de novo* into contigs, continuous sequences resulting from the reassembly of the short reads, by using Trinity software (15). ORFs in the contigs were predicted using TransDecoder software (<http://transdecoder.sourceforge.net>). Reads per kilobase of exon per million fragments (RPKM) values were determined using Tophat-Cufflinks (16,17).

De novo sequencing of *L. bostrychophila* genome

The genomic DNA of *L. bostrychophila* was isolated using an Animal and Fungi DNA Preparation Kit (Jena Bioscience, Jena, Germany). Genomic libraries were constructed using the KAPA Hyper Prep Kit (Kapa Biosystems, Woburn, MA, USA) without amplification and TruSeq Synthetic Long-Read DNA Library Prep (Illumina, San Diego, CA, USA). These libraries were sequenced using the Illumina HiSeq 2500 system. The

raw sequence data were filtered for high quality reads. The hybrid assembly of short and long reads was performed using SPAdes 3.6.2 (<http://bioinf.spbau.ru/spades>) for genome contigs.

Cloning of Lip b 1-encoding cDNA

cDNAs spanning the putative coding sequence of Lip b 1 were amplified from *L. bostrychophila* total RNA template using RT-PCR with PrimeScript reverse-transcriptase (Takara-Bio) and oligo-dT primers. Subsequently, PCR was performed using KOD-plus (Toyobo, Osaka, Japan), a high-fidelity thermostable DNA polymerase, and primers designed to recognize sequences outside the ORF (5'-TCAGTTGAAACCGAGTAAAG-3' and 5'-AATCAAGTTCATTTTCATTC-3'). The RT-PCR products were then cloned into pGEM-T Easy vector (Promega, Madison, WI, USA) and sequenced using dye-terminators and the ABI3130 Genetic Analyzer (Thermo Fisher, Waltham, MA, USA).

Production of recombinant GST-fused Lip b 1 protein

ORFs coding for the two types of Lip b 1 were amplified by PCR using KOD-plus and the following primers: 5'-AAGGGATCCATGGCAGCCTTCAAATTCAT-3' and 5'-TTTGAATTCTTACATTCCTTGAACGCAGG-3' (for Lip b 1.0101), or 5'-AAGGGATCCATGGCTGCCATCAAATTCAT-3' and 5'-TTAGAATTCTTACATTGCTGCAACGCAGG-3' (for Lip b 1.0102). These primer sequences were designed based on the predicted ORF sequences found in the two types of Lip b 1 clones. The PCR products were then cloned into pGEX-4T-2 between *Bam*H I and

EcoR I sites to allow production of the recombinant protein N-terminal GST-fused Lip b 1 (GST-Lip b 1). GST-Lip b 1 was expressed in *E. coli* strain BL21 (DE3) cultured at 37°C in the presence of lactose. *E. coli* cells were then harvested and homogenized by sonication. Expressed GST-Lip b 1 proteins formed insoluble inclusion bodies; therefore, they were dissolved with phosphate buffer (60 mM Na₂HPO₄ · 7H₂O, 40 mM NaH₂PO₄, pH 7) containing 2% Amisoft (Ajinomoto, Tokyo, Japan), a peptide-based detergent. The solubilized inclusion bodies were then applied to a Glutathione Sepharose 4B column (GE Healthcare, Little Chalfont, UK) to purify GST-Lip b 1. The purified proteins were concentrated by ultrafiltration and analyzed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) to evaluate its purity.

Western blotting

L. bostrychophila were collected in a mortar containing RIPA buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.1% SDS, 1% sodium deoxycholate). After quick freezing with liquid nitrogen, the booklice bodies were vigorously disrupted using a muddler. This process was followed by centrifugation, and the supernatant was used as a crude extract of *L. bostrychophila*. Purified GST-Lip b 1 proteins and the booklouse extract were separated by 15% SDS-PAGE and transferred to a 0.2-μm polyvinylidene fluoride (PVDF) membrane (Merck Millipore, Guyancourt, France). After blocking for 16 h at room temperature in 3% skim milk, the membrane was incubated with diluted (1:50) sera from patients with allergic asthma for 16 h at 4°C. After washing with TBST, the blots were incubated with a biotin-labeled antibody against human IgE (KPL, Gaithersburg, MD, USA) and HRP-conjugated streptavidin (Thermo Fisher). The

immunoreactivities were visualized using Immobilon Western Chemiluminescent HRP Substrate (Millipore) and a LAS4000 imaging device (GE Healthcare).

Enzyme-linked immunosorbent assay (ELISA)

Colorimetric ELISA was carried out to evaluate the reactivity of the recombinant GST-Lip b 1 proteins to serum IgE from *L. bostrychophila*-sensitized patients, as described previously (18). Briefly, purified GST-Lip b 1 protein or crude lysates of *L. bostrychophila* were immobilized in wells of 96-well ELISA plates at 4°C for 16 h. After blocking with 1% skim milk in phosphate-buffered saline-Tween 20 buffer (PBST), the wells were treated with 1:50-diluted sera from *L. bostrychophila*-sensitized allergic patients or non-allergic controls at 37°C for 1 h. After washing with PBST, the wells were treated with 1:5,000-diluted biotin-labeled affinity-purified anti-human IgE antibody (KPL) followed by 1:10,000-diluted horseradish peroxidase-conjugated streptavidin (Thermo-Fisher) at 37°C for 1 h. After washing with PBST, TMB substrate (Becton Dickinson, Franklin Lakes, NJ, USA) was added to the wells. Serum IgE reacted with GST-Lip b 1 was detected by measuring absorbance at 450 nm.

Bioinformatics

General nucleotide/amino acid analyses including sequence assembly, homology search and multiple alignment were performed using Genetyx integrated sequence analysis software (ver. 13; Genetyx Corporation, Tokyo, Japan). Transmembrane regions of proteins were predicted *in silico* using SOSUI (<http://harrier.nagahama-i-bio.ac.jp/sosui/>), PredictProtein

(<https://www.predictprotein.org/>), TMPred

(http://www.ch.embnet.org/software/TMPRED_form.html) and MEMSAT

(<http://bioinf.cs.ucl.ac.uk/psipred/>). RNA-seq data of *L. bostrychophila* that were registered by other researchers were obtained from the Short Read Archives database of the National Center for Biotechnology Information and analyzed for generating contigs encoding Lip b 1.

Results

***De novo* sequence assembly of transcriptome and identification of the Lip b 1-encoding contig**

The RNA-seq analysis of *L. bostrychophila* produced 168,211,029 raw reads, of which 156,128,195 were passed through the filtering steps to remove poor quality reads. *De novo* assembly of these reads using Trinity software produced 85,817 contigs, and the subsequent analysis using TransDecoder predicted 35,802 ORFs in the contigs (Table S2). Homology search using Blast+ revealed that a contig designated Comp53139 encoded a protein exactly matching a partial amino acid sequence of Lip b 1, *i.e.*, VLHGSIDAHKAD, which was determined previously in our proteomics study (11).

Amino acid sequences deduced from the ORF of Comp53139 showed that this contig encodes a 254-amino acid protein with a molecular weight of 26,616 (Figure S2), which is consistent with the results of our previous study (11). The RPKM value of Comp53139 was calculated to be 6,400, which corresponds to the 18th highest expression among the 85,817 contigs (Table S3). This result indicated that Lip b 1 is expressed at an extremely high level in *L. bostrychophila*.

Molecular cloning of cDNA for Lip b 1.

A cDNA encompassing the ORF of Lip b 1 was amplified by RT-PCR and sequenced directly. Some overlapped nucleotides were found in our sequence data (Figure S3), indicating the presence of multiple Lip b 1 sequences. Thus, the RT-PCR product was cloned into the plasmid vector, and 88 independent clones were isolated and sequenced. As expected, sequencing analysis revealed that these clones could be classified into nine types based on their ORF sequences, of which type 1 (55 clones) and type 6 (24 clones) accounted for a major proportion (Table S4). Amino acid sequences deduced from the two major types of clones shared a high similarity (87%) (Figure 1). Thus, proteins encoded by type 6 and type 1 clones are considered variants, and designated Lip b 1.0101 and Lip b 1.0102, respectively. The remaining seven clone types accounted for a minor proportion (1 or 2 clones for each), and had only a few nucleotide mismatches with the ORF for Lip b 1.0101 or Lip b 1.0102 (Figure S4 and Table S4). Therefore, we consider that these minor clones resulted from errors in PCR amplification or from posttranscriptional modification, which is supported by the result of whole genome sequencing of *L. bostrychophila*, as described below. Our subsequent analyses focused on Lip b 1.0101 and Lip b 1.0102.

The comprehensive transcriptome data of *L. bostrychophila* were also registered by other researchers in the Sequence Read Archive (SRA) database, a public database that provides short reads as DNA sequencing data generated by high-throughput sequencing (accession #ERR73018 and #SRR665441). Using the data from #ERR73018, two types of Lip b 1 ORFs with a single nucleotide difference were identified, both of which encode the same protein equivalent to Lip b 1.0101 (Figure S5). Using the data from #SRR665441,

two types of Lip b 1 ORFs were identified, from which the deduced amino acid sequences are 93% identical and exhibit 79–87% similarities to Lip b 1.0101 and Lip b 1.0102 (Figure S5). These results suggest that the Lip b 1 gene is highly polymorphic.

Whole genome sequencing of *L. bostrychophila*

Using the data of the whole genome sequencing of *L. bostrychophila*, two distinct contigs that encompassed the sequences of Lip b 1.0101 and Lip b 1.0102 were produced, which span 4 and 33 kbp, respectively. No other contigs contained sequences identical or highly similar to Lip b 1, raising the possibility that both variants of Lip b 1 exist in *L. bostrychophila*. The exon-intron structures of Lip b 1.0101 and Lip b 1.0102, including the sequence of their splicing donor/acceptor sites, were demonstrated to be highly similar (Table S5). These results raise the possibility that Lip b 1.0101 and Lip b 1.0102 could be allelic variants of a single gene rather than products from distinct genes.

***In silico* characterization of Lip b 1 proteins**

The deduced amino acid sequences of Lip b 1.0101 and Lip b 1.0102 were both composed of 254 amino acids (Figure 1). The molecular weights of Lip b 1.0101 and Lip b 1.0102 were calculated to be 26,695 and 26,582, respectively. The theoretical isoelectric points (pIs) of Lip b 1.0101 and Lip b 1.0102 were calculated to be 4.09 and 4.15, respectively. These data were consistent with our previous study (11), which demonstrated that the molecular weight and pI of Lip b 1 are 26 kDa and 3.7, respectively. No identical or highly homologous proteins to Lip b 1 were found by BLAST search, indicating that Lip b 1 is a novel protein. On the other hand, the BLAST search detected PHUM604520, a

hypothetical protein of *Pediculus humanus corporis*, with an E value of 1E-12 (Figure S6). Notably, the length of this protein (255 amino acids) was close to that of Lip b 1 (254 amino acids), although the sequence identity of PHUM604520 to Lip b 1.0101 and Lip b 1.0102 was low (21%) (Figure S6).

Since both Lip b 1.0101 and Lip b 1.0102 are composed predominantly of hydrophobic amino acids (151 and 152 out of 254 amino acids, respectively), we addressed the possibility that Lip b 1 is a membrane protein using four representative tools for membrane protein prediction *in silico*. As summarized in Table 1, both Lip b 1.0101 and Lip b 1.0102 were predicted as membrane proteins. Notably, the N-terminal portions predicted to be transmembrane regions contained typical signal sequences, as predicted by the signal sequence prediction tool Signal IP (data not shown); thus it is most likely that these portions function as leader sequences to sort the Lip b 1 protein. These results suggest that Lip b 1 is a possible transmembrane protein, although its topology and cellular localization are as yet uncertain.

Production of recombinant GST-fused Lip b 1 proteins

To verify that Lip b 1 is a protein reactive to sera from *L. bostrychophila*-sensitized patients, recombinant Lip b 1.0101 and Lip b 1.0102 were produced as N-terminal GST-fused proteins using an *E. coli* expression system. Recombinant GST-Lip b 1 proteins were successfully expressed as 54-kDa proteins (Figure 2A), but they formed insoluble inclusion bodies (Figure 2B), reflecting the *in silico* prediction that Lip b 1 is possibly a transmembrane protein. The expression of GST-Lip b 1 proteins as a soluble form was unsuccessful despite our efforts; therefore, the inclusion bodies were solubilized using a mild peptide-based detergent. The solubilized inclusion bodies containing GST-Lip b 1 remained partly soluble even after the removal of the detergent by dialysis;

thus, we applied the soluble fraction to a Glutathione-Sepharose 4B column to purify Lip b 1 proteins. This purification step was followed by on-column digestion with thrombin to release Lip b 1 proteins from GST. However, Lip b 1 was not eluted from the column by this procedure, possibly due to the inaccessibility of thrombin to its cleavage site. Thus, we eluted GST-Lip b 1 proteins from the column with a glutathione-containing buffer. Using affinity chromatography, Lip b 1.0101 and Lip b 1.0102, each fused to GST, were purified to the extent that they exhibited dominant bands on SDS-PAGE analysis (Figure 2C). As these GST-fused proteins were resistant against thrombin digestion even in liquid reaction, we used GST-Lip b 1 proteins for further experiments.

Seroreactivity of recombinant GST-Lip b 1

The seroreactivity of the purified GST-Lip b 1 proteins were analyzed by ELISA and Western blot analysis. IgE-ELISA revealed that the sera from 10 booklouse-sensitized patients showed high IgE reactivity to the crude extract of *L. bostrychophila*, compared to non-allergic serum. Further, the sera of 9 of 10 patients showed high seroreactivity to both GST-Lip b 1.0101 and Lip b 1.0102, compared to non-allergic serum (Figure 3A). Although the serum of one patient (no. 4) showed conversely low seroreactivity to the GST-Lip b 1 proteins compared to non-allergic serum, it also showed pronouncedly low seroreactivity to GST. Thus, the practical seroreactivity of this serum to Lip b 1 is considered to be high. Consistent results were obtained by IgE-Western blotting using the serum from representative patients (no. 5 and no. 10), and IgE reacted with the GST-Lip b 1 proteins immobilized to PVDF membranes (Figure 3B). These results suggest that both Lip b 1.0101 and Lip b 1.0102 are novel proteins that can cause booklouse allergy.

Discussion

Our previous study demonstrated that many of our Japanese allergic asthma patients were specifically sensitized to household booklice (11). Further, we identified a possible allergenic protein from *L. bostrychophila*, designated Lip b 1, and demonstrated that it was a novel allergen by determining its partial peptide sequences (11). Allergenicity of booklice in India was also reported (12). Considering that the pests commonly populate warm and humid places worldwide (19), it is likely that booklice cause allergosis in other countries as well.

In this study, we successfully determined the full-length primary structure of Lip b 1 protein by RNA-seq and subsequent cDNA cloning. We also demonstrated that two variants of Lip b 1 exist in *L. bostrychophila*, and that both are 254-amino acid proteins sharing 87% similarity. Using whole genome sequencing analysis, we produced two distinct contigs encompassing the sequences of these variants. These contigs were highly homologous throughout their sequences, and their exon-intron structures were nearly identical (Table S5). Thus, we speculate that the two contigs are derived from allelic variants of a single gene, although further analyses are necessary to verify this hypothesis. Homology search revealed a hypothetical protein of *Pediculus humanus corporis*, PHUM604520. This result is consistent with that of a recent transcriptome study, which demonstrated that 62% of the annotated *L. bostrychophila* genes exhibited the most significant homology with *Pediculus humanus corporis* genes (20). However, the identity between Lip b 1 and PHUM604520 is only 21%, indicating that Lip b 1 is a novel protein with unknown function.

Although the mechanism underlying this nucleotide polymorphism is unknown, it is likely that the parthenogenesis (a natural form of asexual reproduction) of *L. bostrychophila* (21) is associated with genetic diversity, as reported previously (22). Consistent with this, Mikac *et al.* (23) demonstrated the genetic diversity of *L. bostrychophila* from different geographic sources. Thus, it is conceivable that the allergenicity of *L. bostrychophila* shows geographic differences due to differences in the epitope sequences of Lip b 1.

Bioinformatics analysis suggested that Lip b 1 could be a possible transmembrane protein, although their transmembrane regions and topology require further clarification. Thus, Lip b 1-presenting cells released from booklice could be allergenic; however, it remains to be clarified how these cells are released from booklice bodies to enable respiratory intake.

Using the GST-fused recombinant proteins, it was verified that both Lip b 1.0101 and Lip b 1.0102 were serologically reactive to the patients' IgEs. Representative tools for allergen prediction, such as Allerdicator, PREAL, and AllerTOP, predicted that both Lip b 1.0101 and Lip b 1.0102 are non-allergens (data not shown), indicating that Lip b 1 is a novel allergen with unique IgE epitopes. It should be noted that cockroach and mite GSTs are known as relevant allergens (24,25). Recently, a lack of significant cross-reactivity among GSTs from cockroaches, mites and *Ascaris* was shown in allergic patients from temperate areas (26). Conversely, it was reported that GST from the helminth *Wuchereria bancrofti* was cross-reactive to cockroach and mite GST allergens (27). Thus, some partial cross-reactivity among GSTs from different species is possible, and the generation of recombinant Lip b 1 proteins without any tags or those fused with non-allergenic tags, if any, will be necessary for diagnosis.

This study demonstrated that Lip b 1 is a novel allergen that has possible transmembrane region(s). This is the first report describing the complete primary structure of an allergenic protein from booklice. Although mites and fungi have been studied as major household allergens, we strongly suggest that sensitization to booklice should be also evaluated in allergen testing.

Acknowledgments

We thank Mr. M. Nakatsuji (Osaka Prefecture University) and Dr. M. Watanabe (National Institute of Health Sciences) for their helpful technical suggestions. This study was supported in part by grants 25242046 (to TI), 25660231 (to TI), and 24791017 (to YF) from the Japan Society for the Promotion of Science, and by a Cooperative Research Grant of the Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture.

References

1. Turner BD. Psocids as a nuisance problem in the UK. *Pesticide Outlook* 1998;**9**:27-30.
2. Turner BD. *Liposcelis bostrychophila* (Psocoptera; Liposcelididae), a stored food pest in the UK. *Int J Pest Manag* 1994;**40**:179-190.
3. McFarlane JA. Damage to milled rice by psocids. *Trop Stored Prod Inf* 1982;**44**:3-10.

- Accepted Article
4. Rees DP, Walker AJ. The effect of temperature and relative humidity on population growth of three *Liposcelis* species (Psocoptera: Liposcelidae) infesting stored products in tropical countries. *Bull Entomol Res* 1990;**80**:353-358.
 5. Kučerova Z. Weight losses of wheat grains caused by psocid infestation (*Liposcelis bostrychophila*: Liposcelididae: Psocoptera). *Plant Prot Sci* 2002;**38**:103-107.
 6. Athanassiou CG, Arthur FH, Throne JE. Efficacy of grain protectants against four psocid species on maize, rice and wheat. *Pest Manag. Sci* 2009;**65**:1140-1146.
 7. Nayak MK, Collins PJ., Pavic H., Kopittke RA. Inhibition of egg development by phosphine in the cosmopolitan pest of stored products *Liposcelis bostrychophila* (Psocoptera: Liposcelididae). *Pest Manag Sci* 2003;**59**:1191-1196.
 8. Kawakami Y, Hashimoto T, Fukutomi Y, Taniguchi M, Saito A, Akiyama K. A survey on the distribution of booklice, other allergenic arthropods, and fungi in houses in Tokyo. *Urban Pest Manag* 2014;**14**:65-77.
 9. Sun JL, Shen L, Chen J, Yu JM, Yin J. Mite and booklouse fauna from vacuumed dust samples from Beijing. *Allergy Asthma Immunol Res* 2014;**6**:257-262.
 10. Baz A, Monserrat VJ. Distribution of domestic Psocoptera in Madrid apartments. *Med Vet Entomol* 1999;**13**:259-264.
 11. Fukutomi Y, Kawakami Y, Taniguchi M, Saito A, Fukuda A, Yasueda H et al. Allergenicity and cross-reactivity of booklice (*Liposcelis bostrychophila*): A common household insect pest in Japan. *Int Arch Allergy Immunol* 2012;**157**:339-348.
 12. Patil MP, Niphadkar PV, Bapat MM. Psocoptera spp. (book louse): A new major household allergen in Mumbai. *Ann Allergy Asthma Immunol* 2001;**87**:151-155.

- Accepted Article
13. Perotin JM, Scherer V, Leduc V, Bouchet F, Deslee G, Lavaud F. Allergic asthma to psocids, a new indoor allergen of ecological building materials. *Allergy* 2011;**66**:1252-1259.
 14. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;**1**:10-12.
 15. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc* 2013;**8**:1494–1512.
 16. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;**7**:562-578.
 17. Pollier J, Rombauts S, Goossens A. Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol Biol* 2013;**1011**:305-315.
 18. Yasueda H, Saito A, Sakaguchi M, Ide T, Saito S, Taniguchi Y et al. Identification and characterization of a group 2 conifer pollen allergen from *Chamaecyparis obtusa*, a homologue of Cry j 2 from *Cryptomeria japonica*. *Clin Exp Allergy* 2000;**30**:546-550.
 19. Wang JJ, Tsai JH, Zhao ZM, Li LS. Development and reproduction of the psocid *Liposcelis bostrychophila* (Psocoptera: Liposcelididae) as a function of temperature. *Annal Entomol Soc Am* 2000;**93**:261-270.
 20. Dou W, Shen GM, Niu JZ, Ding TB, Wei DD, Wang JJ. Mining genes involved in insecticide resistance of *Liposcelis bostrychophila* Badonnel by transcriptome and expression profile analysis. *PLoS One* 2013;**8**:e79878.

21. Yusuf M, Turner, BD. Characterisation of Wolbachia-like bacteria isolated from the parthenogenetic stored product pest psocid *Liposcelis bostrychophila* (Badonnel) (Psocoptera: Insecta). *J Stor Prod Res* 2004;**36**:169–175.
22. Mark Welch D, Meselson M. Evidence for the evolution of *Bdelloid rotifers* without sexual reproduction or genetic exchange. *Science* 2000;**288**:1211-1215.
23. Mikac KM, Clarke GM. Tracing the geographic origin of the cosmopolitan parthenogenetic insect pest *Liposcelis bostrychophila* (Psocoptera: Liposcelididae). *Bull Entomol Res* 2006;**96**:523-530.
24. Arruda LK, Vailes LD, Platts-Mills TA, Hayden ML, Chapman MD. Induction of IgE antibody responses by glutathione S-transferase from the German cockroach (*Blattella germanica*). *J Biol Chem* 1997;**272**:20907-20912.
25. O'Neill GM, Donovan GR, Baldo BA. Glutathione S-transferase a major allergen of the house dust mite, *Dermatophagoides pteronyssinus*. *Immunol Lett* 1995;**48**:103-107.
26. Mueller GA, Pedersen LC, Glesner J, Edwards LL, Zakzuk J, London RE et al. Analysis of glutathione S-transferase allergen cross-reactivity in a North American population: Relevance for molecular diagnosis. *J Allergy Clin Immunol* 2015;**136**:1369-1377.
27. Santiago HC, LeeVan E, Bennuru S, Ribeiro-Gomes F, Mueller E, Wilson M et al. Molecular mimicry between cockroach and helminth glutathione S-transferases promotes cross-reactivity and cross-sensitization. *J Allergy Clin Immunol* 2012;**130**:248-256.

Figure Legends

Figure 1 Nucleotide sequences of the open reading frame (ORF) for Lip b 1.0101 and Lip b 1.0102, the deduced amino acid (AA) sequences and their homologies

Nucleotide sequences of the ORF in Lip b 1.0101 (A) and Lip b 1.0102 (B) cDNAs.

Deduced amino acid sequences are shown below the nucleotide sequences. An asterisk represents a stop codon. The nucleotide sequences of the ORF in Lip b 1.0101 and Lip b 1.0102 are registered under accession numbers LC177783 and LC177784, respectively. (C) Homology of the AA sequences between Lip b 1.0101 and Lip b 1.0102. Boxed amino acids represent the partial AA sequence determined previously in our proteomic study.

Figure 2 Expression and purification of recombinant glutathione-S-transferase (GST)-fused Lip b 1 protein

(A) The crude extracts of *E. coli* cells harboring the empty vector (lanes 1, 4), GST-Lip b 1.0101 (lanes 2, 5) and GST-Lip b 1.0102 (lanes 3, 6) vectors were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Lanes 1–3; *E. coli* cells cultured in the presence of lactose. Lanes 4–6; *E. coli* cells cultured in the absence of lactose. (B) The whole extracts (lane 1, 4), soluble fractions (lane 2, 5), and insoluble fractions (lane 3, 6) of *E. coli* cells expressing GST-Lip b 1.0101 (lane 1–3) and GST-Lip b 1.0102 (lane 4–6) were analyzed by SDS-PAGE. (C) Purified GST-Lip b 1.0101 (lane 1) and GST-Lip b 1.0102 (lane 2) were analyzed by SDS-PAGE. Weak bands at a position around 27 kDa (indicated by an asterisk) are considered to correspond to GST tag cleaved off from the GST-fused Lip b 1 proteins because the purified recombinant GST migrates

to the same position (data not shown). Cleavage of the GST-fused proteins seemed to take place continuously even during the subsequent gel-filtration chromatography, so that it could not be removed.

Figure 3 Seroreactivity of recombinant GST-fused Lip b 1 protein

(A) The seroreactivities of whole booklouse (BL) extracts (5 μ g), recombinant GST-Lip b 1.0101, GST-Lip b 1.0102 and GST (0.5 μ g) were analyzed using sera from BL-sensitized patients (no. 1–10) and non-allergic serum (NA) by enzyme-linked immunosorbent assay (ELISA). The data are expressed as the means $\pm$ SD. (B) Whole BL extracts (30 μ g), recombinant GST-Lip b 1.0101, GST-Lip b 1.0102 and GST (0.8 μ g) were subjected to SDS-PAGE and transferred to PVDF membranes. The two sets of protein blots were treated with serum from two representative BL-sensitized patients, no. 5 (upper) and no. 10 (lower) or non-allergic serum, and subjected to Western blotting to detect seroreactivity in parallel. An arrow indicates the position of the GST-Lip b 1 proteins. An arrowhead indicates the position of native Lip b 1. The faint bands below GST-Lip b 1 may represent signals from truncated GST-Lip b 1 proteins that are generated in the process of purification possibly due to their instability.

Figure S1 Dorsal view of *Liposcelis bostrychophila*

Figure S2 Nucleotide sequence and deduced AA sequence of the contig Comp53139

The position of the last AA in each line is indicated on the right. An asterisk represents a stop codon at the end of the ORF. Shaded nucleotides represent in-frame stop codons upstream of the start codon, which excludes the possibility that a 5'-extended ORF exists. Boxed AAs represent the partial AA sequence determined in our previous study (11).

Figure S3 Overlapping nucleotides in the Lip b 1 cDNA sequence chart

A representative sequence chart analysis of the polymerase chain reaction (PCR) product encompassing the Lip b 1 ORF. Red circles represent overlapping nucleotides, below which two different colors of peaks are overlapping (nucleotides indicated below the circles are either of them). A green line represents the nucleotide sequences that encode the partial AA sequences (shown below the line) of Lip b 1 identified previously (11).

Figure S4 Multiple alignment of the nucleotide sequences of ORFs in Lip b 1 cDNA clones

A total of 88 clones isolated and analyzed for sequencing were categorized into nine types (types 1 to 9) based on their ORF sequences. Numbers shown in parentheses represent numbers of each clone type.

Figure S5 Multiple alignment of the deduced AA sequences of Lip b 1

Figure S6 Identity between Lip b 1 and Phum604520

The homology search using Blastp detected a “hit” with Phum604520, a hypothetical protein of *Pediculus humanus corporis*. The pairwise alignment between Lip b 1.0101 and Phum604520 (A), and that between Lip b 1.0102 and Phum604520 (B), are shown.

Table 1 *In silico* characterization of Lip b 1 using transmembrane (TM) protein prediction tools

Tool	Lip b 1.0101		Lip b 1.0102	
	Number of TM regions	Position of TM regions	Number of TM regions	Position of TM regions
SOSUI	4	1-23 40-62 65-87 199-221	4	1-23 34-56 65-87 199-221
PredictProtein	3	1-18 32-49 207-224	3	1-18 31-48 199-216
TMPred	4	1-17 196-215 237-254	3	1-17 64-81 196-215 237-254
MEMSAT	1	199-215	1	195-211

Ishibashi, O. *et al.* Table 1

A

```

atggcagccttcaaatcctcctcctcgccttctcgccttcagcgtgtctcaaactact
M A A F K F I L L A F L A F S V S Q T T 20
gaggccaacgttgtagcgcaaccaagagcctttgacataattgacatcggttaaaaaagtt
E A N V V A Q P R A F D I I D I V K K V 40
atccaaatcggttgaaattgtcgctcgaagctgttaacgggtgcagttgaaccaatcgataaa
I Q I V E I V V E A V N G A V E P I D K 60
attttgcaagaacttattgctttacttctcgtgatgtagctaaaacgggtgactgacgtt
I L Q E L I A L L P A D V A K T V T A V 80
gtcgatgcctcagacaagcaattgccgatgaaaatgtcagaatcgaccacatcattgaa
V D A L R Q A I A D E N V R I D H I I E 100
gtcctcaaaaagctttggatgatctcttggctatcgacccatgttaccaacccaagct
V L N K A L D D L L A I D P C Y Q P Q A 120
gatgccatcaaagccgttcttgacacagctttaagcgggtattgacaaagtccctccacggt
D A I K A V L D T A L S G I D K V L H G 140
tccatcgatgctcacaagccgatatcgataacgttatcgctggattttcacaagacttg
S I D A H K A D I D N V I A G F S Q D L 160
gaagacctcagaacctttatgacactcaactccccgccgctgtgtgtttgcttgaccca
E D L R N L Y D T Q L P A A V V C L T P 180
ggtaatgctgattcctgcacctgcttgatggtgtcaagaaaacgtagtaaacgggtgt
G N A D S C T C L D G V K E T V V N G V 200
gtgagccttgctgccaacttcgttcttcaccttactgccgcttctgatgttcttaaagtt
V S L A A N F V L H L T A A S D V L K V 220
gttgctcctgaggttatcgacgggtgccacccccgctcgtcaacgatggtcttgcagctgct
V V P E V I D G A T P V V N D G L A A A 240
ggtcctctcattgacaacgtgtgcgctgcgttcaaggaatgtaa
G P L I D N V C A C V Q G M * 254

```

B

```

atgggtgccatcaaattcctcctcctcgccttctcgccttcagcgtgtctcaaactact
M A A I K F I L I A F L A F S V S Q T T 20
gaggccaacgttgtagcgcaaccaaaaagcctttgacataattgcaattgttaacaagtt
E A N V V A Q P K A F D I I A I V K Q V 40
atcgacattgttcgaattgtcgctcaaagcgggttaacgatgcagttccagatatcgataaa
I D I V R I V V K A V N D A V P D I D K 60
attttgcaacaacttgttcttacttctcgtgatgtagctgcaacagtgactacagtc
I L Q Q L V A L L P S D V A A T V T T V 80
ctcgatgccatcaacaagcaattaccgatgaaaatgccagaatcgaccacatcattcaa
L D A I K Q A I T D E N A R I D H I I Q 100
gtcctcgaaaaagctttggatgatctcttggctatcgacccatgttaccaacccaagct
V L E K A M D D L L A I D P C Y Q P Q A 120
gatgccattaaagccgttcttgacaaagctttaagcgggtattgacggagtccctccacggt
D A I K A V L D K A L S G I D G V L H G 140
tccatcgatgctcacaagccgatatcgataacgttatcgctggattttcacaagacttg
S I D A H K A D I D N V I A G F S Q D L 160
gaagacctcagaacctttatgacactcaactccccgccgctgtgtgtttgcttgaccca
E D L R N L Y D T Q L P A A V V C L T P 180
ggtaatgctgattcctgcacctgcttgatggtgtcaagaaaacgtagtaaacgggtgt
G N A D S C T C L D G V K E T V V N G V 200
gtgagccttgctgccaacttcgttcttcaccttactgccgcttctgatgttcttaaagtt
V S L A A N F V L H L T A A S D V L K V 220
gttgctcctgaggttatcgacgggtgccacccccgctcgtcaacgatggtcttgcagctgct
V V P E V I D G A T P V V N D G L A A A 240
ggtcctctcattgacaacgtgtgcacctgcgttgagcaatgtaa
G P L I D N V C T C V A A M * 254

```

C

```

Lip b 1.0101 1 MAARKFILDAFLAFSVSQTEANVVAQPRAFDIIIVKQVIDIVRVVAVNDAVPEIDK 60
Lip b 1.0102 1 MAARKFILDAFLAFSVSQTEANVVAQPRAFDITIVKQVIDIVRVVAVNDAVPEIDK 60

Lip b 1.0101 61 ILQLHALLPNDVARTVTAVVDALRQAIIDENVRIDHIIIVLKAADDLLAIDPCYQPPA 120
Lip b 1.0102 61 ILQLHALLPNDVARTVTAVVDALRQAIIDENVRIDHIIIVLKAADDLLAIDPCYQPPA 120

Lip b 1.0101 121 DAIRAVLDALSGIDSVLHGSIDAHKADIDNVIAGFSQDLEDLRNLYDTQLPAAVVCLTP 180
Lip b 1.0102 121 DAIRAVLDALSGIDSVLHGSIDAHKADIDNVIAGFSQDLEDLRNLYDTQLPAAVVCLTP 180

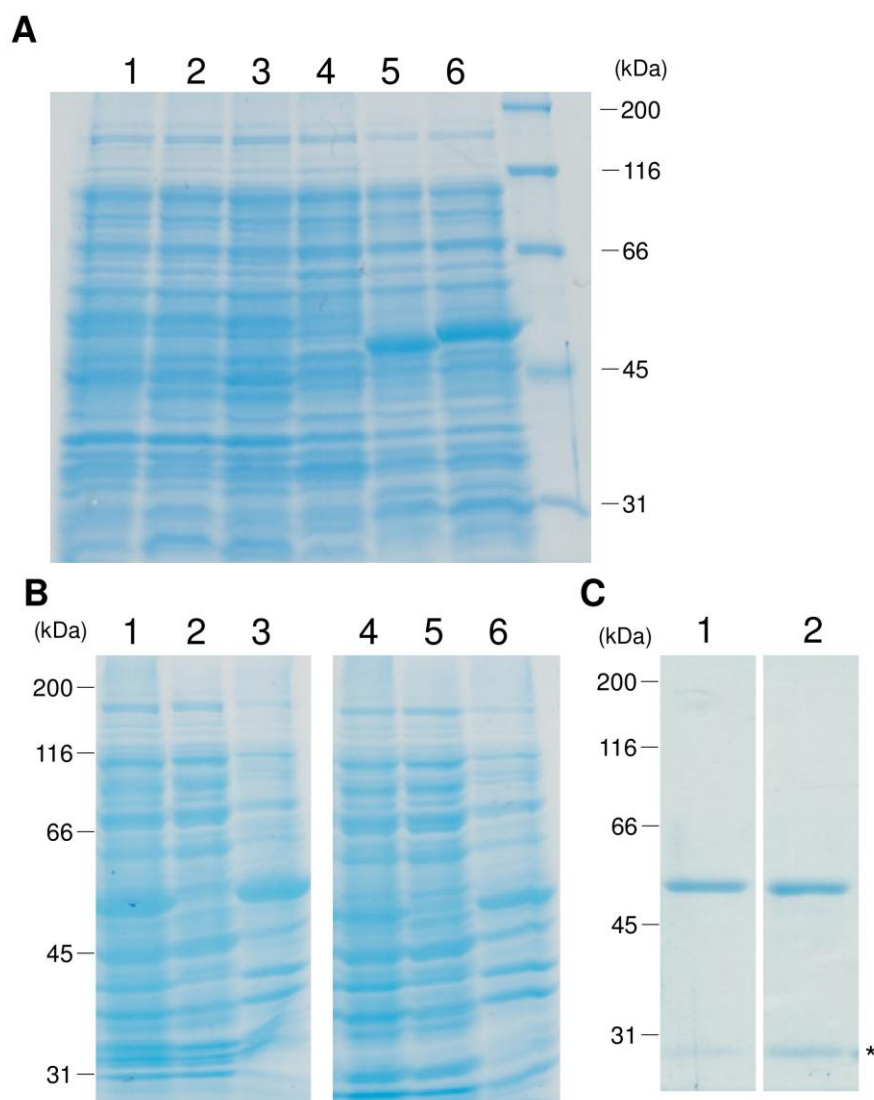
Lip b 1.0101 181 GNADSCCLDGVKETVVNGVVSIAANFVHLTAASDVLKVVVPEVIDGATPVVNDGLAAA 240
Lip b 1.0102 181 GNADSCCLDGVKETVVNGVVSIAANFVHLTAASDVLKVVVPEVINGSTPTVVNDGLAAA 240

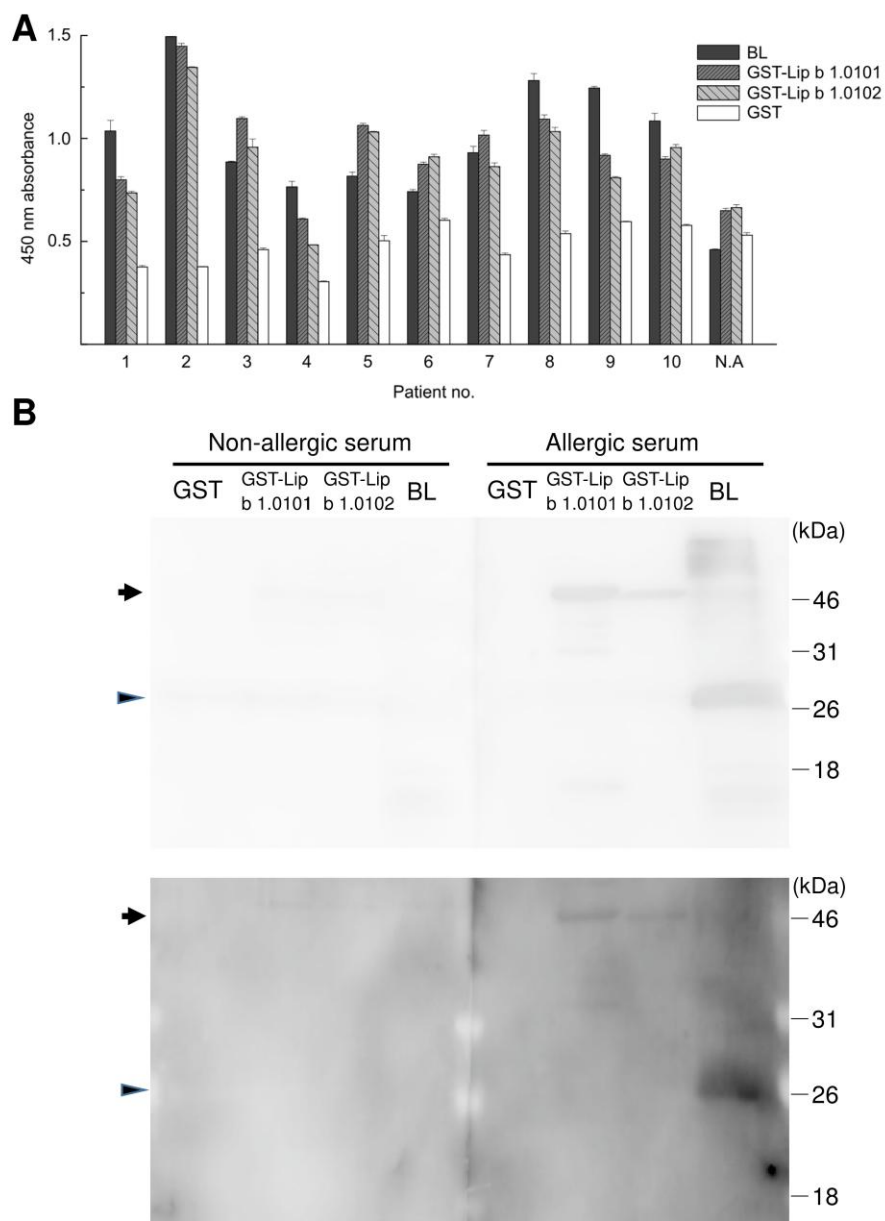
Lip b 1.0101 241 GPLIDNVCAVCQGM 254
Lip b 1.0102 241 GPLIDNVCAVCVAM 254

```

Ishibashi, O. *et al.*

Figure 1

Ishibashi, O. *et al.* Figure 2



Ishibashi, O. *et al.* Figure 3