Bioinformatic Studies on the Group 2 Allergens of *Dermatophagoides farinae* from China

Yubao Cui¹,², Jianglong Peng¹, Peng Zhou², Ming Peng² and Shiyun Qian¹

**SUMMARY** The present study aimed to characterize the group 2 allergens of the house dust mite *Dermatophagoides farinae* (Der f 2) from Hainan Island, a tropical region in Southeastern China. We cloned and sequenced cDNA coding for Der f 2 and found an additional region of 87 base pairs (bp) (from +77 to +163 bp) in our strain that was absent in the reference sequence (GenBank AB195580) used for primer design. However, the BLAST analysis identified the same sequence in strains reported from Reinbek, Germany, and Guangzhou, China. A phylogenetic tree was constructed using the Der f 2 nucleotide sequences from different regions or countries and showed that the Hainan sequence clustered with the strains from Reinbek and Guangzhou. Analysis of the translated amino acid sequence suggests that the encoded peptide is hydrophobic and extracellular with a cleavage site between the 17th and 18th amino acid residues and contains a strong trans-membrane helix from the 6th amino acid to the 24th amino acid, indicating a MD-2-related lipid recognition domain in this protein. Furthermore, the secondary structure of the pro-protein consists of 16.57% alpha helix, 32.57% extended strand and 50.86% random coil. In brief, we obtained a gene coding for Der f 2 and predicted the molecular characteristics of this protein using bioinformatics tools. Our analysis identified that this gene showed several significant differences to those reported previously.

House dust mites are among the most common sources of indoor allergens worldwide. They have been found to produce over 30 types of allergens associated with asthma, atopic dermatitis, rhinitis and other allergic diseases. Specific immunotherapy with mite extracts has been considered by some to be the only effective approach to provide long-lasting relief of allergic symptoms to house dust mites. Unfortunately, these extracts contain non-allergenic proteins and lack effective concentration of important allergens, both of which may be associated with clinical side effects of mite extracts.¹ In contrast, recombinant allergens and their mixtures have shown to be effective and will likely be important for immunotherapy for dust mite-specific IgE-mediated hypersensitivity.²

Of the allergens produced by house dust mites, i.e. groups 1 and 2, were considered to be responsible for the high levels of IgE and Th2 cytokines in 80% of the allergic patients.³ However, there is increasing evidence showing that both groups of allergens were highly polymorphic.⁴⁻⁷ Therefore, it is very important to understand the molecular, physical

From the ¹Department of Parasitology, Hainan Medical College, Chengpi Road, Haikou 571101, the People’s Republic of China, ²State Key Laboratory of Tropical Crop Biotechnology, Institute of Tropical Science, China Academy of Tropical Agricultural Science, Haikou 571101, the People’s Republic of China. Correspondence: Yu-Bao Cui E-mail: 6801609@163.com
and chemical properties of the diverse forms of the allergens before they can be effectively used in clinical diagnosis and immunotherapy. To characterize house dust mites’ allergens from Hainan Island, a tropical region in Southeastern China, we cloned and sequenced the cDNA coding for Der f 2 in the house dust mite Dermatophagoides farinae. Surprisingly, the sequenced results showed an insertion of 87 bp that was not present in the reference sequence (GenBank AB195580). Here, we report our experimental results and the bioinformatics analysis of this sequence.

MATERIALS AND METHODS

D. farinae culture and isolation of adult mites

To isolate house dust mites, we first obtained house dust from the floors of rice and flour shops in Haikou City, Hainan Island and isolated house dust mites under the stereomicroscope. All of the mites potentially belonging to D. farinae were put into a culture chamber. After about two months, the whole culture in the chamber was taken to select for D. farinae by stereomicroscope. Subsequently, mites presumptively identified as D. farinae were purified by cultivating them in small culture chambers. Two months later, representative mites were taken from the chambers and the species were identified. When the mites found to be D. farinae, all of the mites in the chamber were treated to be D. farinae. These mites were then raised at a relative humidity of 75% and temperature of 25°C on a large scale. A culture medium comprising of yeast, wheat flour and rice was used for the large-scale cultivation. Following our previous protocols, the adults were isolated as follows. The whole culture was taken and placed on glass plates, and after thirty minutes, the culture medium was removed manually. The adult mites, larval mites and some aliquots of the culture medium were collected in small ceramic cups using a small writing brush for the mite isolation. Under a lamp, the adults moved rapidly to the basal part of the ceramic cup and away from the larvae. The adults were removed with a writing brush under a microscope, and about 600 mites were collected for total RNA isolation.

Total RNA isolation

Total RNA of the mites was isolated using an RNA isolator (TaKaRa Biotech Co. Ltd, Code No.D312) according to the manufacturer’s instructions. Briefly, about 600 alive adult mites were rapidly frozen with liquid nitrogen and 1 ml of RNA isolator (RNAiso) was added. The mites in RNAiso were homogenized using a PowerGen 125 Tissue Homogenizer (Fisher Scientific) starting at 1,800 x g and gradually going up to approximately 7,200 x g for 30-60 seconds at room temperature. After homogenization, all contents in the tissue homogenizer were removed to an Eppendorf tube and 0.2 ml chloroform was added. After precipitation and air-drying, the RNA pellet was dissolved in RNase-free water and stored at -80°C.

Preparation of cDNA and polymerase chain reaction (PCR)

For PCR amplification of the cDNA coding for Der f 2, two primers were designed and synthesized based on the GenBank sequence AB195580. The forward primer sequence was (5’GGATCCATGATTCCAAAAATCTTGCTCCTT-C3’) and the reverse was (5’CTCGAGTTAATCACGGATTCTAACCCATGGG3’). These two primers contained a BamHI site and an XhoI site at their 5’ end (underlined), respectively. Firstly, reverse transcription (RT) was performed using the total RNA isolated from the mites with a 3’-Full RACE Core Set Ver.2.0 (TaKaRa Biotech Co. Ltd, Code No.D314) in the PCR Thermal Cycler Dice (TaKaRa Biotech Co. Ltd., Code TP600) for 30 minutes. Each 10 μl of reaction mixture for RT contained 3 μl of total RNA, 1x RNA PCR Buffer, 1 mM of dNTP mixture, 5 mM of MgCl2, 1.25 U of M-MLV Reverse Transcriptase XL, 10 U of RNase inhibitor, 10 pmol each of the forward and reverse primers. Secondly, the RT product was used as template for PCR in the same Thermal Cycler Dice with an initial incubation at 94°C for 3 minutes, followed by 30 cycles at 94°C for 30 seconds, at 55°C for 30 seconds and 72°C for 1 minute. The reaction mixture contained RT products (2 μl), 2x GC Buffer (5 mM Mg2+ Plus, 25 μl), 5 U/μl of LA Taq (0.5 μl), 1x cDNA dilution buffer (8 μl), 20 pmol of the primers F and R (1 μl each), and 12.5 μl of dH2O. The PCR amplification products were ligated with the plasmid pMD19-T (TaKaRa Biotech Co. Ltd, Code No.D104) and then transformed into E. coli competent cells JM109 (TaKaRa Biotech Co. Ltd., Code No.D9052). Positive clones were selected by the blue/white coloration on Luria-Bertani (LB)
plates containing 100 μg/ml ampicillin. The presence of the inserted DNA was verified by restriction analysis and DNA sequencing with the BcaBEST M13-47 and BcaBEST RV-M primers on a MegaBACE 1000 DNA sequencer. Digested by BamH I and Xho I, the cloned DNA fragments were electrophoretically analyzed on an agarose gel.

**Nucleotide sequence analysis, alignment and phylogeny**

The obtained sequence was edited to remove vector and restriction sites, and a complete open reading frame (ORF) was obtained using the ORF finder in the NCBI (National Center for Biotechnology Information) Website. By BLASTn searches on NCBI, sequences coding for Der f 2 deposited from other geographic regions were aligned using the VECTOR NTI 9.0 software (IBI, New Haven, CT, USA), and a phylogenetic tree was constructed using MEGA3.0 software. The Der f 2 cDNA sequence and the reference sequence (GenBank AB195580) were aligned by ClustalW 1.83 (http://www.ebi.ac.uk/clustalw/index.html).

**Deduced amino acid sequence and the bioinformatics analysis of the pro-protein**

The amino acid sequence of Der f 2 was deduced by Translate Tools (http://www.expasy.org). Its potential subcellular localization and hydrophobicity were analyzed using the TopPred tools. The SignalP-NN was predicted with the EUK networks method by the SignalP 3.0 software in the SMART Server (http://www.cbs.dtu.dk/services/SignalP/).

The transmembrane region was predicted by the TMpred in ISREC Server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html). Its function was predicted using the Superfamily 1.69 software in the HMM library, and its secondary structure was generated by GOR4.

**RESULTS**

**Cloning and sequence of Der f 2 cDNA from Hainan, China**

Based on the nucleotide sequence of Der f 2 in the GenBank (AB195580), primers were designed to amplify the cDNA fragments by PCR using total RNA isolated from adult mites. By agarose electrophoresis, a DNA band was observed. The recovered PCR products were linked to the plasmid pMD19-T and transformed into the JM109 E. coli competent cells. Positive colonies were identified by the blue/white screening and the recombinant plasmids were extracted and analyzed by restriction enzyme digests using BamH I and Xho I. Two plasmids containing the targeted cDNA fragment were identified and confirmed by DNA sequencing with the primers of BcaBEST Primer M13-47 and BcaBEST Primer RV-M. According to the ORF Finder in the NCBI Web server, an ORF was found from the start codon

```
AB195580

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>110</td>
<td>BamH I</td>
</tr>
<tr>
<td>150</td>
<td>200</td>
<td>Xho I</td>
</tr>
<tr>
<td>205</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>255</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>405</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>455</td>
<td>500</td>
<td></td>
</tr>
</tbody>
</table>
```

**Fig. 1** Comparison of the Hainan Der f 2 cDNA sequence with the reference sequence (AB195580).
ATG to the stop codon TGA with a length of 528 bp. Unexpectedly, there was an additional 87 bp (position 77-163) in our cloned fragment that were absent in the reference sequence (AB195580) (Fig. 1). These two sequences shared 83% nucleotide identity.

**Nucleotide sequence alignment and phylogeny for group 2 allergens reported from different regions or countries**

The sequencing results were queried by BLAST in the NCBI web server and homologous sequences coding for the group 2 allergens from the same species but reported from different geographic regions were chosen and analyzed. Among the analyzed sequences, two GenBank accessions AF3469-05 and Emb AJ682838 reported from China and Germany, respectively, contained the same 87 base pair insertion (Fig. 2). Tables 1 and 2 show the sequence similarity for Der f 2 sequences reported from different regions. After alignment by VECTOR NTI 9.0 software and manual adjusts, a phylogenetic tree was constructed using the MEGA 3.0 software (Fig. 3).

**Fig. 2** Homology and polymorphism of group 2 allergen cDNA sequences reported from different regions or countries.
Table 1  Nucleotide sequence similarity among Der f 2 reported from different regions

<table>
<thead>
<tr>
<th>District and countries</th>
<th>Accession code</th>
<th>Identity</th>
<th>E value (10,000)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guangzhou, China</td>
<td>AF346905</td>
<td>100%</td>
<td>2.3e-206</td>
<td>Hao et al. (2003)</td>
</tr>
<tr>
<td>Nakhon Pathom, Thailand</td>
<td>DQ185511</td>
<td>96%</td>
<td>5e-180</td>
<td>Piboonpocanan et al. (2006)</td>
</tr>
<tr>
<td>Reinbek, Germany</td>
<td>AJ862838</td>
<td>97.7%</td>
<td>3.6e-203</td>
<td>Nandy et al. (2003)</td>
</tr>
<tr>
<td>Seoul Korea</td>
<td>AY066008</td>
<td>98.1%</td>
<td>1.4e-139</td>
<td>Jin et al. (2003)</td>
</tr>
<tr>
<td>Singapore</td>
<td>AY283288</td>
<td>98.9%</td>
<td>8.7e-142</td>
<td>Augus et al. (2003)</td>
</tr>
<tr>
<td>Tokyo, Japan</td>
<td>D10448</td>
<td>83.5%</td>
<td>4.2e-146</td>
<td>Yuuki et al. (1991)</td>
</tr>
<tr>
<td>Shenzhen, China</td>
<td>EF139432</td>
<td>83.5%</td>
<td>3.5e-146</td>
<td>Liu et al. (2006)</td>
</tr>
</tbody>
</table>

By BLASTn searches on NCBI, sequences coding for Der f 2 deposited from other geographic regions were chosen. The identity and E value between our Der f 2 sequence and each of those sequences were computed by LALIGN. (http://www.ch.embnet.org/software/LALIGN_form.html)

Table 2  Similarity for Der f 2 reported from different districts or countries by the VECTOR NTI 9.0 software

<table>
<thead>
<tr>
<th></th>
<th>Guangzhou</th>
<th>Hainan</th>
<th>Reinbek</th>
<th>Nakhon Pathom</th>
<th>Seoul</th>
<th>Shenzhen</th>
<th>Tokyo</th>
<th>Singapore</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guangzhou</td>
<td>100</td>
<td>90</td>
<td>91</td>
<td>68</td>
<td>69</td>
<td>73</td>
<td>78</td>
<td>71</td>
</tr>
<tr>
<td>Hainan</td>
<td>100</td>
<td>87</td>
<td>73</td>
<td>74</td>
<td>78</td>
<td>68</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Reinbek</td>
<td>100</td>
<td>90</td>
<td>97</td>
<td>90</td>
<td>76</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakhon Pathom</td>
<td>100</td>
<td>62</td>
<td>62</td>
<td>65</td>
<td>71</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seoul</td>
<td>100</td>
<td>90</td>
<td>90</td>
<td>76</td>
<td>94</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shenzhen</td>
<td>100</td>
<td>86</td>
<td>96</td>
<td></td>
<td></td>
<td>82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tokyo</td>
<td>100</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Deduced amino acid sequence and bioinformatics studies

Fig. 4 shows the deduced amino acid sequence of 175 residues. The protein encoded a hydrophobic and signature extracellular region with a loop length of 127 residues and 15 Lys and Arg according to the TopPred results (http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html). The most likely cleavage site of the pro-protein was between positions 17 and 18 by the SignalP 3.0 software in SMART Server (http://www.cbs.dtu.dk/services/SignalP/). One strong transmembrane helix was found from position 6 to position 24 with a total score of 1300 as predicted by the TMPred in ISREC Server, suggesting this protein probably anchors on a cytoplasmic membrane and may function as an adaptor molecule that may play a role in cell signaling. InterProscan results (http://www.ebi.ac.uk/InterProScan) revealed that there were two possible domains in this protein: one homologous to the E1 protein, an epididymal secretory protein (from pos. 55 to 172) and the other homologous to the E set superfamily (from pos. 22 to 170). Both of these two domains were indexed in the ML family (MD-2-related lipid recognition domain). With regard to its secondary structure, GOR4 predicted that 16.57% (29) of the protein was in alpha helices, 32.57% (57) as extended strands, and 50.86% (89) as random coils (Fig. 5).

DISCUSSION

It has long been recognized that recombinant allergens are essential for clinical diagnosis of house dust allergies and hold a great potential for
therapy against house dust mite allergy. Previous reports were largely restricted to the screening of cDNA libraries for a gene fragment coding for an active protein. As a result, such obtained proteins would not contain the pro-peptide sequence. Subsequent experiments showed that the purified allergens expressed in \textit{E. coli} had only 50\% of the IgE-binding activity of the natural molecule.\(^9\),\(^{10}\) In contrast, Best et al.\(^{11}\) reported that recombinant forms of Der f 1 (rDer f 1) containing the pro-peptide sequence expressed in \textit{Pichia pastoris} had a similar IgE binding activity as native Der f 1 isolated directly from mites. This result suggested that the pre-pro sequence was essential for the secretion of the mature form of the protein and both the alpha factor signal peptide and the pro-enzyme region were efficiently processed during secretion. Therefore, the production of recombinant allergens should include the pro-protein sequence.

In this study, a complete CDS coding for

\begin{verbatim}
1    ATG ATT TCC AAA ATC TTG TGC CTT TCA TTG TTG GTA GCA GCC GTT  45
  1    Met Ile Ser Lys Ile Leu Cys Leu Ser Leu Leu Val Ala Ala Val  15
  46   GTT GCC GAT CAA GTC GAT GTT AAA GAT TGT GGT AAG TTT GTT TGT  90
  16   Val Ala Asp Gln Val Asp Val Lys Asp Cys Gly Lys Phe Val Cys  30
  91   GTG ATA CAT TTT TTT TCT TTT CAT TTA TTT ACT AAA ACC AAC AAT 135
  31   Val Ile His Phe Phe Ser Phe His Leu Asn Thr Thr His Asn      45
 136  TTT CTT TTT CTT GTT TAT ATT CAT ATA GCC AAC AAT GAA ATC AAA 180
  46   Phe Leu Phe Leu Val Tyr Ile His Ile Ala Asn Glu Ile Lys      60
 181  AAA GGA ATG GTC GAT GCT TGC CAT GGT TCT GAT CAC TGC ATC ATC 225
  61   Lys Val Met Val Asp Gly Cys His Gly Ser Asp Pro Cys Ile Ile  75
 226  CAT CGT GGT AAA CCA TCC ACT TTG GAA GCC TTA TTC GAT GCC AAC 270
  76   His Arg Gly Pro Phe Thr Leu Glu Ala Leu Phe Asp Ala Asn      90
 271  CAA AAC ACT AAA ACC GCT AAA ATT GCA CCA AAA TCT GAT CTC GAT 315
  91   Gin Asn Thr Lys Thr Ala Lys Ile Glu Ile Asp Ser Leu Asn      105
 316  GGT CTT GAA ATT CAT GGT GCC GAT GCT GAC ACC CAT TGG CAT CAT 360
 106  Gly Leu Glu Ile Asp Val Pro Gly Ile Asp Thr Asn Ala Cys His    120
 361  TTT ATG AAA TGC TTG GTT AAA GGT CAA CAA TAT GAT ATC AAA 405
 121   Phe Met Lys Cys Pro Leu Val Lys Gly Gin Gin Tyr Asp Lys      135
 406  TAT ACA TGG AAT TTG CCG AAA ATT GCA CCA AAA TCT GAA AAC ATC 450
 136  Tyr Thr Trp Asn Val Pro Lys Ile Ala Pro Lys Ser Glu Asn Val    150
 451  GTG TGT ACA GTC AAA CTT ATC GGT GAT CAT GGT TTG GCT TGG 495
 151  Val Val Thr Val Lys Leu Ile Gly Asp Asn Gly Val Leu Ala Cys   165
 496  GCT ATT GGT ACC CAT GGT AAA ATC CGT GAT TAA 528
 166   Ala Ile Ala Thr His Gly Lys Ile Arg Asp End
\end{verbatim}

\textbf{Fig. 3} A phylogenetic tree for group 2 house dust mite allergens reported from different regions or countries.

\textbf{Fig. 4} The deduced amino acid sequence of the Hainan Der f 2 allergen.
Der f 2 was selected as the reference sequence for the design of primers to amplify house dust mite allergen from adult \textit{D. farinae}. Unexpectedly, an additional sequence was found in our results. After BLAST analysis against the GenBank database in the NCBI, only four sequences, i.e. GenBank accession numbers AJ862838, AJ862837, AJ862836 and AF346905, contained the extra 87 base pairs, and the other cDNA coding for Der f 2 did not contain this extra region. Our analysis identified sequence polymorphisms among the group 2 allergens. The polymorphisms were found corresponding to nucleotide positions 76, 88, 111 and 125 as reported by Yuuki \textit{et al.} When the cDNA sequences coding for Der f 2 from different geographic regions were analyzed, the Hainan strain was clustered with two sequences, one from Reinbek, Germany and the other from Guangzhou, China. Hainan Island is a tropical region in Southeastern China. If not by airplanes, everyone coming to Hainan Island from mainland China has to go through Guangzhou province first. Therefore, the close genetic relationship of \textit{D. farinae} between Hainan and Guangzhou may have originated from the migrations brought about by hosts' migration and traveling. However, at present, the closeness of the relationship between mites from Hainan and that from Reinbek, Germany remains unknown. We suppose that mites from Hainan, China and Reinbek, Germany, belong to the same sub-species and the mites from Nakhon Pathom, Seoul, Tokyo and Singapore probably belong to another sub-species. The different sub-species of mites may travel from one country to another with international trading, consequently resulting in their particular geographic distribution. Furthermore, a highly conserved region was found between +240 and +555 bp among the Der f 2 sequences from all the indicated regions or countries via sequence comparison (Fig. 3), indicating that this region may be critical for their functions. Moreover, the polymorphisms among Der f 2 cDNA sequences, especially the deletion mutation from +77 to 163 bp, may reflect their functional difference in cell biology.

So far, the biochemical function of the group 2 allergens remains largely unknown and is still debated. Some scholars insisted that the group 2 allergens could be the products of epididymal secretions. Other scholars showed that the tertiary structure of the group 2 allergens was similar to the third and fourth domains of the transglutaminase coagulation factor VIII. However, our analysis suggested that group 2 allergens might have a transmembrane domain and a MD-2-related lipid recognition domain.

In summary, we cloned and sequenced the group 2 allergen from \textit{Dermatophagoides farinae} isolated in Hainan, China. We found that this gene contained an insertion of 87 base pairs not seen in the reference sequence. Our analysis showed that the strain from Hainan was more similar to those from Reinbek, Germany and Guangzhou, China than to strains from other parts of East Asia.

**ACKNOWLEDGEMENTS**

We are grateful to Dr. Wang Bo, Faculty of Health Sciences, University of Cape Town, South

---

**Fig. 5** The secondary structure of the Hainan Der f 2 allergen. Alpha helix (hh): 29 is 16.57%; extended strand (ee): 57 is 32.57%; and random coil (cc): 89 is 50.86%.
Africa and Dr. XU Jianping, Department of Biology, McMaster University, Canada, for patiently modifying our manuscript. This work was supported by the National Sciences Foundation of China (Grant No. 30060166), the Natural Science Fund of Hainan Province (Grant No. 2005-80556), and the Hainan Provincial Health Department (Grant No. 2005-6).

REFERENCES

16. Best EA, Stedman KE, Bozic CM, Hunter SW, Vailes L, Chapman MD, McCall CA, Medermott MJ. A recombinant group 1 house dust mite allergen, rDer f 1, with biological activities similar to those of the native allergen. Protein Exp Purif 2000; 20: 462-71.
17. Yuuki T, Okumura Y, Okudaira H. Genomic organization and polymorphism of the major house dust mite allergen Der f 1, with biological activities similar to those of the native allergen. Protein Exp Purif 2000; 20: 462-71.