

Full-length article

Cloning, expression, and characterization of pollen allergens from *Humulus scandens* (Lour) Merr and *Ambrosia artemisiifolia* L¹

Ai-lin TAO^{2,3}, Shao-heng HE^{2,4}

²Allergy and Inflammation Research Institute, Shantou University Medical College, Shantou 515031, China; ³The Second Affiliated Hospital of Guangzhou Medical College, Guangzhou 510260, China

Key words

pollen; allergen; profilin; gene expression

¹Project supported by the National Natural Science Foundation of China (No 30470996), the Natural Science Foundation of Guangdong Province (No 034617) and the Li Ka Shing Foundation, Hong Kong, China (No C0200001).

⁴Correspondence to Dr Shao-heng HE.

Phn 86-754-890-0405.

Fax 86-754-890-0192.

E-mail shoahenghe@hotmail.com

Received 2005-05-16

Accepted 2005-06-24

doi: 10.1111/j.1745-7254.2005.00194.x

Abstract

Aim: To clone the pollen allergen genes in *Humulus scandens* (Lour) Merr (LüCao in Chinese) and short ragweed (*Ambrosia artemisiifolia* L) for recombinant allergen production and immunotherapy. **Methods:** The allergen genes were selectively amplified in the weed pollen cDNA pool by using a special PCR profile, with the primers designed by a modeling procedure. Following truncated gene cloning and confirmation of the pollen source, unknown 3'cDNA ends were identified by using the 3'-RACE method. The gene function conferred by the full-length coding region was evaluated by a homologue search in the GenBank database. Recombinant proteins expressed in *Escherichia coli* pET-44 RosettaBlue cells were subsequently characterized by N-terminal end sequencing, IgE binding, and cross-reactivity. **Results:** Three full-length cDNAs were obtained in each weed. Multiple alignment analysis revealed that the deduced amino acid sequences were 83% identical to each other and 56%–90% identical to panallergen profilins from other species. Five recombinant proteins were abundantly expressed in non-fusion forms and were confirmed by using the N-terminal end sequence identity. Sera from patients who were allergic to *A artemisiifolia* reacted not only with rAmb a 8(D03) derived from *A artemisiifolia*, but also with recombinant protein rHum s 1(LCM9) derived from *H scandens*, which confirmed the allergenicity and cross-reactivity of the recombinant proteins from the 2 sources. Comparison of the degenerate primers used for truncated gene cloning with the full-length cDNA demonstrated that alternative nucleotide degeneracy occurred. **Conclusion:** This study demonstrates a useful method for cloning homologous allergen genes across different species, particularly for little-studied species. The recombinant allergens obtained might be useful for the immunotherapeutic treatment of *H scandens* and/or *A artemisiifolia* pollen allergies.

Introduction

Accurate allergy diagnosis and efficient immunotherapy protocols are strongly dependent on the use of standardized extracts or well-characterized allergen mixtures^[1–3]. However, standardization of specific allergens from natural allergen extracts is a huge task^[4] and it is almost impossible to carry out standardization on a large scale for day-to-day clinical practice. Recombinant allergens that have similar

structures, functions and immunoreactivity to their native counterparts could be used to overcome this problem, given that they can be produced in large quantities, leading to more specific, effective and safer immunotherapy treatment^[5].

Currently, gene cloning for novel allergens is mainly based on the construction of a cDNA library, which is time-consuming and expensive^[6–9]. Other methods for isolating unknown allergens involve using the polymerase chain reaction (PCR) with either degenerate primers consisting of a

pool of primers containing most or all of the possible nucleotide sequences encoding a conserved amino acid motif^[10], or consensus primers consisting of a single primer containing the most common nucleotide at each codon position within the motif^[11]. Although these strategies have been successful in isolating closely related sequences, they have generally failed when sequences are more distantly related^[11]. Many programs and algorithms have been developed for sequence comparisons of multiple members of protein families^[12-15]. By using alignment techniques, the first amino acid sequence that emerges, that is, the sequence aligned uppermost in each large and dense cluster in the output tree produced, has provided the most information about amino acid diversity in the cluster. In the current study, we refer to these sequences as “core sequences”, and we designed the degenerate primers based on the core sequences and their cognates. Under special PCR conditions, we expected that the degenerate primers could match all associated sequences in the abovementioned cluster, and thus could be used for cloning new allergen gene members in a distantly related species. We will hereafter refer to these primers as “pan-degenerate primers”. We expected that the process involving the pan-degenerate primers would be more efficient and less expensive than the cDNA library technique, and would have a higher success rate than an ordinary PCR method.

Humulus scandens (Lour) Merr (Moraceae: Cannabaceae; Lücao in Chinese) is a weed that grows in most provinces of China. The plant blossoms in summer and bears a large quantity of pollen. Short ragweed (*Ambrosia artemisiifolia* L; Compositae) is one of the most invasive weeds in the world. It was introduced to China in 1930s, and is now distributed from the Yantze River to the 3 northeastern provinces of China. Nanning, Wuhan, Nanchang-Jiujiang and Shenyang-Tieling-Dandong are the 4 areas in which the weed is most pervasive in China. Pollen from *H scandens* and *A artemisiifolia* have become the main source of pollen allergens in this country. However, little is known of the actual allergens in the pollen of the weeds. In this study, we report on the cloning, expression and characterization of 6 pollen allergen genes from the 2 weeds. Following the non-fusion expression and purification of the 5 encoded proteins, 2 representative proteins of the 5, provisionally named rAmb a 8 (D03) and rHum s 1 (LCM9), were immunologically characterized in detail.

Materials and methods

Sample preparations Wild plants of *H scandens* and *A artemisiifolia* were harvested in Wuhan. Flowers were

bagged, and the pollen collected from each plant was mixed with other pollen collected from the same plant. All pollen collected was immediately submerged in liquid nitrogen until use. Twenty-five patients with *A artemisiifolia* pollen allergy and 10 patients with *H scandens* pollen allergy (as assessed by a positive skin prick test reaction to *A artemisiifolia* and a serum level of *A artemisiifolia*/*H scandens* specific IgE higher than 3.5 kU_A/L) and 13 healthy volunteers were recruited. Blood samples were taken from all participants in the Allergy Clinic, Shenyang North Hospital. All the samples were tested by using the UniCAP Specific IgE Fluoroenzymeimmunoassay (Pharmacia Diagnostics AB, Uppsala, Sweden). After centrifugation at 500×g, 4 °C, for 10 min, serum from each individual was collected and stored at -70 °C until use. The study was approved by the Ethics Committee of Shenyang North Hospital, Shenyang, China, and written informed consent was obtained from all participants.

Primer design The 478 pollen allergens and their related sequences were retrieved by keyword searches in the SWISS-PROT and TrEMBL websites (<http://www.expasy.org>) and were multiple-aligned online by using Clustal W software^[16]. For alignment, the default values were used to obtain the output tree in a phylogram pattern, with the absolute distances labeled. The uppermost allergen sequence in each large and dense cluster in the output tree was taken as the core sequence. There were several of these core sequences found in the output phylogenetic trees. Based on the core sequences, allergen sequences with high homology (with BLAST E values $\leq 10^{-10}$) were selected for further analysis to acquire conservative gene domains. Pan-degenerate primers, which we expected to be successful in all species, were thus designed according to cDNA sequences of the conservative domains chosen.

RT-PCR analysis Frozen pollen samples were ground in a mortar filled with liquid nitrogen, and the total RNA was extracted by using an RNeasy Maxi Kit (Qiagen Valencia, CA, USA) according to the manufacturer's instructions. After quality and concentration assessments, the RNA was either immediately used for RT-PCR analysis or stored at -80 °C until use.

After reverse transcription (2 µg total RNA per sample) was performed by using the ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs, Beverly, MA, USA) according to the manufacturer's instructions, PCR was carried out by using 50 ng of the reverse transcription product in 25 µL of the reaction mixture containing 1.25 U HotstarTaq DNA polymerase (Qiagen), 1× PCR buffer (Qiagen), 0.2 mmol/L dNTP, and 0.4 µmol/L of each primer. Thermal cycling was

performed in a PTC-200 Peltier Thermal Cycler (MJ Research, Watertown, MA, USA) as follows: after an initial step at 95 °C for 15 min to activate the DNA polymerase, 9 touch-down cycles were carried out at 94 °C for 45 s, 68 °C for 30 s and 72 °C for 30 s, with the annealing temperature decreasing by 1 °C per cycle. This was followed by 26 gradient cycles at 94 °C (45 s)→55 °C→61 °C (30 s)→72 °C (30 s). After purification with a QIAquick PCR Purification Kit (Qiagen), the PCR products were cloned into the pGEM-T easy vector (Promega, Madison, WI, USA). The inserted plasmids were transformed into JM109 (Promega). Positive clones were confirmed first by colony PCR with the former PCR primers, and then by sequencing. DNA sequencing was carried out by the BioAsia Biotechnology Co (Shanghai, China) by using the dideoxy chain-terminating method.

Full-length cDNA synthesis and sequencing The clones derived from the degenerate primers did not cover the full-length coding region. Gene-specific primers were therefore synthesized (Table 1), and combined using the RACE process^[17] with RA3, a 3' reverse primer from the GeneRacer Kit (Invitrogen Corporation, Carlsbad, CA, USA). The PCR conditions were similar to those for RT-PCR, with a slight modification in the denaturing temperature and extension time. PCR products were also cloned into the pGEM-T easy vector, and positive clones were confirmed by colony PCR and sequencing. Full-length cDNAs, acquired by deduction of the truncated gene and its 3' end, were translated into amino acid sequences, with their isoelectric point and molecular weight calculated (<http://www.expasy.org>). cDNA and protein homologues, retrieved by using the BLAST program in the NCBI website (<http://www.ncbi.nlm.nih.gov/>), were

aligned by using the CLUSTAL W (1.82) program^[16]. Phylogenetic trees were subsequently generated using MEGA2 software^[18].

Northern blot Northern blots were performed according to the membrane manufacturer's instructions. Briefly, a total of 20 µg RNA was loaded on to a formaldehyde-containing gel. After running the gels at 40 V for 3 h, the RNA bands were electrically transferred onto Nytran SuPerCharge nylon transfer membrane (Schleicher & Schull BioScience, New Hampshire, USA). After UV cross-linking treatment, the membrane was stained with methylene blue [0.02% [w/v] methylene blue in 0.3 mol/L sodium acetate(pH 5.5)] to investigate the transferring efficiency. Following destaining, BrightStar Psoralen-Biotin (Ambion Inc, Autin, TX, USA)-labeled probe D03 was applied for hybridization in ULTRhyb hybridization buffer (Ambion Inc). The membrane was washed twice with a buffer containing 2× SSC (0.3 mol/L NaCl, 0.03 mol/L sodium citrate) and 0.1% sodium dodecylsulfate before the probes were detected by using the Phototope Star Detection Kit (New England Biolabs). Finally, the filter was exposed to Fuji Medical X-ray film at room temperature for 5–8min.

Protein expression and immunoblotting To enable non-fusion protein expression, the S-protein tag and His-tag were removed from the pET-44 vector and the allergen-coding regions were decorated by introducing restriction sites (5' *NdeI* and 3' *PstI*) by PCR using the corresponding primers (Table 1). Subsequent to double restriction digestion and purification, the PCR products were ligated with the pET-44 EK/LIC vector (Novagen, Madison, WI, USA). After transformation of the ligation products into NovaBlue Singles

Table 1. Sequences of primers used in this study. The underlined sequences denote the restriction sites of *NdeI* and *PstI*, respectively. Bracketed alternative nucleotides in primer Sg1P3 represent the degenerate loci.

Primer	Sequence (5' → 3')	Length (base)	Use
Sg1p5	ATGTCGTGGCAGGCGTACGT	20	Degenerate PCR
Sg1p3	CCTCTCAACAA(T/G)CA(C/T)GTTGCATTGTCC	32	Degenerate PCR
PC11	GACCATCTGATGTGCGAAATCGATGG	26	Specific PCR
PA04	GGGAACCAGGTGCTGTTATCAGAGG	25	Specific PCR
PD10	GGAGCTGTGATTCGAGGGAAAAAGG	25	Specific PCR
PD03	CTGCATCAAGAAAACAGGCCAAGC	24	Specific PCR
RA3	CTGCATCAAGAAAACAGGCCAAGC	24	Specific PCR
A5RA4	GGT <u>CATATG</u> TTCGTGGCAGGCGTACGT	26	For construct
B3LA4	AGT <u>CTGCAGT</u> CATTAGAAAACCTGTTCCAGGAG	33	For construct
B3LD3	AT <u>CTGCAGT</u> CATTACATGCCCTGATCGATGAGAT	34	For construct
B3CM9	AGT <u>CTGCAGT</u> CATTAGAGACCCTGATCAATGAGATAATC	39	For construct
B3CS13	AGT <u>CTGCAGT</u> CATTAGAGGTTCTGATCAATAAGGTAATCTC	41	For construct
SProt	CACGATGCGTCCGGCGTAGAG	21	Colony PCR

cells (Novagen), positive clones and in-frame insertions were confirmed, initially by colony PCR with the primer SProt and 3' primers related to each clone, and then by sequencing with the S-Tag 18-mer primer or ColiDown primer (Novagen). Plasmid DNA was extracted with the QIAprep Miniprep Kit (Qiagen), and transformed into *E coli* RosettaBlue (DE3) cells (Novagen) for expression, following the manufacturer's instructions. Cells were grown in 1 L Luria-Bertani (LB)-medium until the OD₆₀₀ value reached 0.6. Target protein expression was induced by the addition of isopropyl-D-thiogalactopyranoside (IPTG) to 1 mmol/L for 4 h at 30 °C before harvest. Finally, the recombinant proteins were purified by affinity chromatography using poly-(L)-proline as the solid phase. The bacterium lysate was applied to the column, followed by washing with 1 mol/L urea and elution with 6 mol/L urea. The elution product was dialyzed stepwise against decreasing amounts of urea (6, 4, 2 mol/L, no urea) in phosphate-buffered saline (PBS; 1.8 mmol/L potassium phosphate, pH 7.4, 0.137 mol/L NaCl) and then dialyzed against double-distilled water, followed by freeze-drying.

Purified rAmb a 8(D03) and rHum s(LCM9) were selected as representatives for immunoblotting. Fifteen micrograms of protein were loaded onto the 15% SDS-PAGE gel, and were then electrically transferred to a polyvinylidene difluoride (PVDF) membrane (Amersco, Solon, OH, USA). Following blocking the membrane overnight with a 4% bovine serum albumin solution at 4 °C, allergic sera (diluted 1:7) were added to each membrane strip for 2 h before the addition of peroxidase conjugated goat anti-human IgE (Sigma-Aldrich, Saint Louis, MO, USA). The blotted band was finally visualized with a diaminobenzidine (DAB) chromogenic substrate solution (Amersco).

N-terminal end sequencing Sequencing of the N-terminal end of recombinant proteins was performed by using the Applied Biosystems Procise 491 Sequencer from Applied Biosystems, according to the manufacturer's instructions.

Results

Homology between different allergen genes The 478 pollen allergens obtained from the SWISS-PROT and TrEMBL websites were clustered into 8 large groups, which consisted of a number of subgroups. Further searching with the core sequences from each subgroup demonstrated that these core sequences matched a large quantity of closely related homologues. For instance, Q9XF40, the first core sequence in group 8, was an apple pollen allergen with extensive homology with other pollen allergens from different species (Figure 1), such as *Poa pratensis*, *Lolium perenne*,

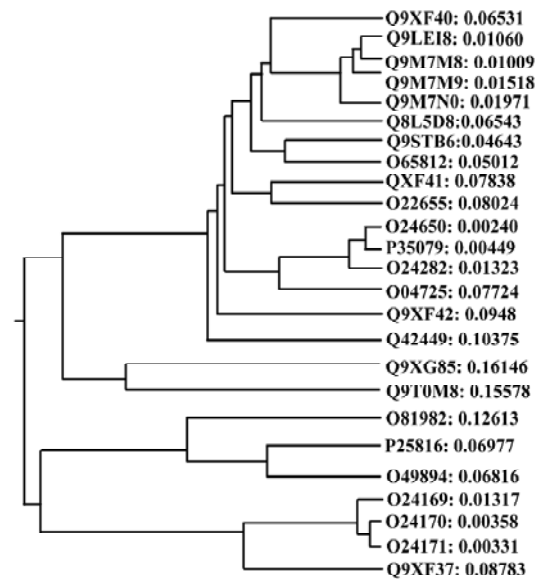


Figure 1. The core subgroup cut off from the large group VIII clustered from 478 allergens obtained from SWISS-PROT. The conserved regions in Q9XF40, O81982 and Q9XF37 were used for pan-degenerate primer (Sg1P5/P3) design. Q9XF40 and O81982 were selected for their clustering priority, whereas Q9XF37 was selected for its amino acid length of 133 aa. All allergens shown were from SWISS-PROT. The numbers on the right-hand sides of each allergen denote absolute distance.

Dactylis glomerata, *Hordeum vulgare*, *Holcus lanatus* (e values $<2e^{-25}$). The conservative domains in the Q9XF40 cDNA sequence (1–20 bp, 343–369 bp) as resolved by multiple sequence alignment of Q9XF40 and its cognates were taken as the Sg1p5/Sg1p3 primer sequence (Table 1). Other degenerate primers were designed following the same principle.

Full-length cDNA generation By combining the gradient PCR with touchdown PCR, 3 gene fragments were obtained from each weed species, and were subsequently inserted into the pGEM-T vector. Sequence analysis showed that the inserts, 5 of 363 bp and 1 of 369 bp, were homologous to the known allergen gene profilin from different species with up to 90% (5 clones except for D03) and 79% (for D03) identity, but with ~30 bp truncated at the 3' ends, suggesting a further requirement of the RACE process for the full-length cDNA. After cloning of the amplificants from the specific PCR with primer pairs PC11/RA3, PM9/RA3, PD10/RA3 and PD03/RA3, a total of 29 positive RACE clones were detected. Subsequent sequencing reactions demonstrated that they all contained the initiation and stop codons, primer sequences and poly (A) tails, reflecting the mRNA source of the clones. Northern blotting with a clone D03 probe further confirmed that the cDNA cloned was derived from the pollen

mRNA (Figure 2). Bridging the 3' end and the corresponding fragment genes produced 6 novel full-length coding genes, among which LCS13, LCM20 and LCM9 were from *H scandens*, whereas A0418, D106 and D03 were from *A artemisiifolia*. They are all available in GenBank under accession Nos AY268422-AY268427.

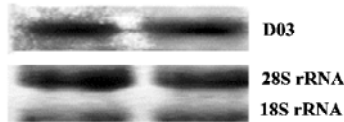


Figure 2. Northern blot results of the total RNA extracted from the *A artemisiifolia* pollen at different times. Total RNA of 20 µg was electrophoresed, blotted onto Nytran SuPerCharge nylon transfer membranes (Schleicher & Schull BioScience) and stained with methylene blue to confirm the transferred RNA quality. After destaining, Psoralen-Biotin labeled probes for D03 were applied for hybridization.

Sequence analysis Sequence analysis revealed that there were several altered nucleotides in the cDNA sequences compared with the degenerate primers used in the fragment-gene cloning process (data not shown), which means that the degeneracy was effectively expanded. The predicted amino acid sequences of the 6 genes were acidic proteins (pI of approximately 5.0) with molecular weights of approximately 14 kDa. Multiple sequence alignment demonstrated that these 6 novel sequences shared a high homology, up to 83%. Moreover, cDNA and protein sequence searches with Blastn and Blastx programs in GenBank showed that the new genes were homologous to profilin genes from more than 10 plant families including Rosaceae, Euphorbiaceae, Phaseoleae, Liliaceae, Poaceae, Urticaceae, Apiaceae, Oleaceae, Solanaceae, Brassicaceae, and Compositae. Alignment of the cognates derived from this search produced a phylogenetic tree, which showed that the 6 novel genes cloned were clearly separated into 3 subgroups, but with high homology (Figure 3).

Of the 6 clones, A0418, D106 and LCM9, 3 closely related clones that encoded 131-amino acid residues, exhibited high homology with apple pollen allergen Q9XF41 (Mal d 4; GenBank accession number: AF129427). LCS13 and LCM20, 2 clones from *H scandens*, had only differences in the cDNA non-coding regions and therefore encoded the same protein, which had 80% and 85% identity with Q8L5D8 from date palm pollen and Q9XF40 from apple pollen, respectively. Clone D03 encoded 133 amino acid residues, and had an identity of 79% and 72% with O81982 (Hela 2)^[19] and CAD12862^[20], 2 profilin proteins from common sunflower and mugwort. Hel a 2 and D03, along with the timothy grass profilins O24282, P35079, O24650, were clustered into one

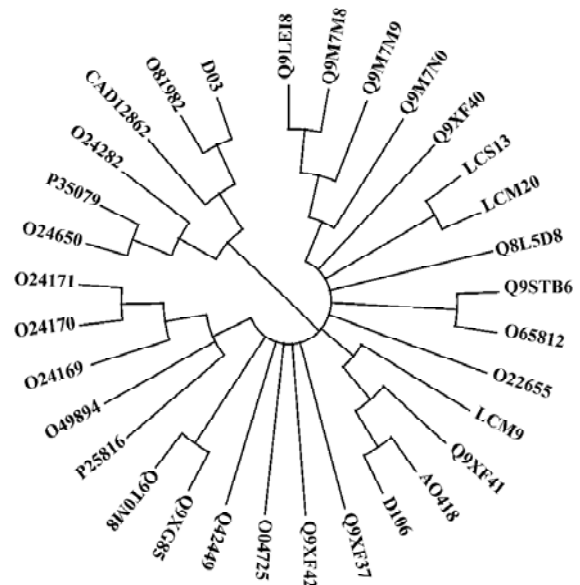


Figure 3. Phylogenetic relationship between the profilin cognates and the 6 novel allergens cloned in the current study as analyzed by using the Mega2 software with amino acid sequences. The amino acid sequences of the 6 new allergens (LCS13, LCM20, LCM9, A0418, D106, D03) are available from GenBank (accession numbers: AAP15198, AAP15199, AAP15200) and SWISS-PROT (accession numbers Q64LH2, Q64LH1, Q64LH0).

subgroup (Figure 3). Furthermore, D03 was quite different from the other 5 clones in 4 regions of its amino acid sequence, which represents an approximately 90% difference between D03 and other profilins (Figure 4).

Protein expression and purification, and immunoblotting After detecting the correct reading frame of the target protein gene in the expression vector, the in-frame plasmids were then successfully transformed into *E coli* RosettaBlue (DE3) cells. Following induction and further incubation at 30 °C for 6 h, the recombinant non-fusion proteins were visualized using SDS-PAGE, and had an apparent molecular mass of ~14 kDa (Figure 5A). The N-terminal sequence of 10 amino acids of rAmb a 8(D03) confirmed that the recombinant protein was the same as that deduced from the cDNA sequence. The yield of purified protein was approximately 8 mg per L of cultured bacteria.

IgE antibodies from 6 of 8 sera (75%) from the *A artemisiifolia* allergic patients reacted with recombinant protein rAmb a 8(D03) in the immunoblot experiments. The bands occurred at approximately 14 kDa (Figure 5B). Other recombinant proteins were also very efficiently expressed and purified in non-fusion form (Figure 5C, 5D). As a representative from *H scandens*, the immunoblotting profile of rHum s 1(LCM9) is shown in Figure 5E. Interestingly, rHum s 1(LCM9) not only

```

LCS13 MSWQAYVDDHLMCEIDG--QHLTAAAIIGHDGSVWAQSSSTFPQFKPEEIAAIMKDFEEPGSLAPTGLHLG 68
LCM20 MSWQAYVDDHLMCEIDG--QHLTAAAIIGHDGSVWAQSSSTFPQFKPEEIAAIMKDFEEPGSLAPTGLHLG 68
A0418 MSWQAYVDDHLMCEIEG--NHLSAAAIIGHDGVVWAQSATFPQVKPEEITGIMNDFNEPGLAPTGLYLGL 68
D106 MSWQAYVDDHLMCEIEG--NHLSAAAIIGHDGVVWAQSATFPQVKPEEITGIMNDFNEPGLAPTGLYLGL 68
LCM9 MSWQAYVDDHLMCEIDG--NHLSAAAIIGHDGSVWAQSAFPPQLKPEEVTGIMNDFNEPGLAPTGLYLGL 68
D03 MSWQAYVDEHLMCDIEGTGHHLTSAAILGHGDTVWAQSSNFPQFKPEEMKGIITEFDQAGTLAPTGMFIA 70
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
GAKYMLQGEQAGAVIRGKKGAGGVTIKKTSQALIGVYDEPMPGQCNMVERLGDYLLDQGM 133
* ****:* * *****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

Figure 4. ClustalX (1.8) alignment of the deduced amino acid sequences of the 6 clones obtained from the pollen of *H. scandens* and *A. artemisiifolia*. The amino acid residues shown represent the specific region in clone D03 that was markedly different from the profilin proteins of other species. The conservative domains in each clone are shown in bold, and the different amino acid residues are in gray.

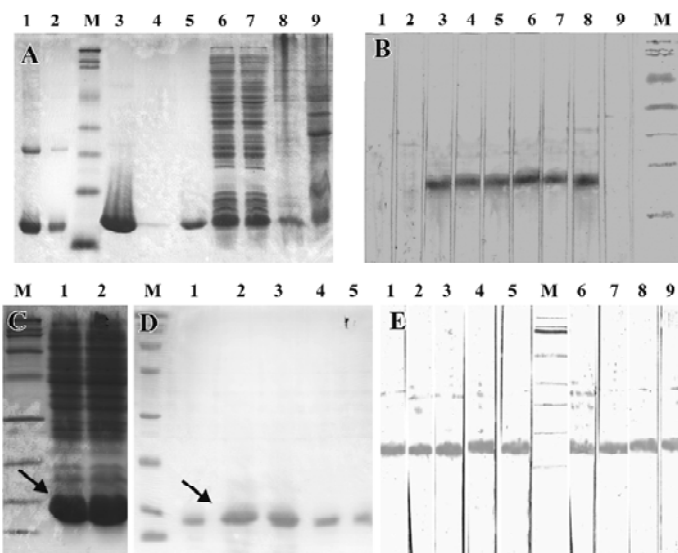


Figure 5. Non-fusion expression, purification and immunological characterization of recombinant proteins. A: 15% SDS-PAGE analysis of the no-fusion expression profile of rAmb a 8 (D03). Lanes 1 and 2, purified target protein in non-reducing loading buffer; Lane M, broad range prestained SDS-PAGE standards (Bio-Rad, catalog No 161-0318), similar to Lane M in all other panels; Lanes 4 and 5, re-purified target protein in non-reducing loading buffer; Lane 6, re-purified target protein in reducing loading buffer; Lanes 7 and 8, cell lysate in loading buffer; Lanes 9 and 10, supernatant and sediment of the cell lysate. B: Western blot of rAmb a 8 (D03). Fifteen micrograms of the purified rAmb a 8 (D03) was loaded into each lane and blotted. Lanes 3–8, sera from patients with allergy to *A. artemisiifolia*; Lanes 1, 2 and 9, healthy control sera. C: Representative presentation of non-fusion protein expression of rHum s 1 (LCS13) (Lane 1) and rHum s 1 (LCM9) (Lane 2) (arrowhead). D: Recombinant proteins purified by affinity chromatography. Lanes 1–5 represent rAmb a 8 (D03), rHum s 1 (LCS13), rHum s 1 (LCM9), rAmb a 8 (D106) and rAmb a 8 (A0418), respectively. E: Western blot of rHum s 1 (LCM9), which was detected in sera from both patients with *H. scandens* allergies (lanes 1–5) and patients with *A. artemisiifolia* allergies (lanes 6–9).

reacted with the sera from 5 patients with *H. scandens* allergy, but also reacted with sera from patients with *A. artemisiifolia* allergy, which suggests that cross-reactivity occurred for rAmb a 8 (D03) and rHum s 1 (LCM9).

Discussion

In the past, most allergen genes have been cloned by cDNA library screening^[7-9], and hundreds of pollen allergen genes from more than 170 species are available in online databases, which can be sorted into several groups. This means that homologous genes exist in many related or unre-

lated species, which provided support for the current method. In this study, we designed pan-degenerate primers and succeeded in cloning pollen allergen cDNAs from different weeds, thus showing that there is a way to clone homologous genes from distantly related species or little-studied species. Furthermore, by using a combination of the gradient PCR process and touchdown steps, the genes with some divergence in their primer regions could be obtained in the weed genome; that is, the degeneracy of the primers was expanded. The method could therefore be used to clone the same allergens in other species or to clone other kinds of allergen genes.

It has been argued that allergenicity cannot be rationalized on the basis of either overall folding pattern or biological function, and thus any protein should be regarded as a potential allergen^[19]. According to alignment studies, the proteins coded by our new clones had more than 60% identity over the full-length peptide with the known allergens, suggesting that the proteins would be allergenic. Thus, D03 and LCM9, the most different variants of the 6 clones obtained, were expressed as recombinant proteins. Through efficient expression in a pET-44 system, ideal solubility of all the recombinant proteins was obtained, which facilitated protein purification through affinity chromatography, and would ensure efficiency in commercial production.

The molecular basis of cross-reactivity has been of primary concern in numerous publications^[19–21]. The relationship between sequence similarity, as obtained by pair-wise alignment, and structural or functional properties has been the goal of much research^[22–24]. Recent studies have confirmed a widely accepted rule-of-thumb that 30% or 35% identity over aligned regions suffices for structural or functional deduction^[25,26]. A remote homologue study^[27] demonstrated that allergens from diverse sources have a common structural motif, namely a groove located inside an α - β motif, which could potentially serve as a ligand-binding site, thus leading to the stimulation of the T cell helper type 2 (Th2) response and a subsequent bias towards the synthesis of IgE. This result expanded cross-reactivity to between microbes and higher plants. In the present study, the clones were different from each other in their cDNA or protein sequence, suggesting open pollination and/or large genetic variation in the weeds. This was also consistent with other studies on ragweed and timothy grass sequence polymorphism^[28,29]. However, multiple sequence analysis demonstrated that the newly cloned genes had up to 83% homology with each other and 60%–90% identity with previously described allergen profilins from latex, food, and pollens from various organisms, which can produce extensive cross-reactivity^[30,31]. This result demonstrated that proteins from *H. scandens* and *A. artemisiifolia* would have cross-reactivity. In a previous study we also inferred that the proteins coded by clone D106 would have cross-reactive properties for food- and/or pollen-sensitive allergic individuals^[32,33]. All of these results imply that the newly cloned genes would function the same way as pan-allergen profilins. Based on the issues described herein and the results obtained in the current study, we assume that cross-reactivity would occur so widely that atopic individuals would be surrounded by a dynamic network of allergens, and would be predisposed towards allergic diseases.

References

- 1 van Ree R. Analytic aspects of the standardization of allergenic extracts. *Allergy* 1997; 52: 795–805.
- 2 Durham SR, Walker SM, Eva-Maria V, Ikilar J, O'Brien F, Noble W, *et al.* Long-term clinical efficacy of grass-pollen immunotherapy. *N Engl J Med* 1999; 341: 468–75.
- 3 van Hage-Hamsten M, Johansson E, Roquet A, Peterson C, Andersson M, Greiff L, *et al.* Nasal challenges with recombinant derivatives of the major birch pollen allergen. *Clin Exp Allergy* 2002; 32: 1448–53.
- 4 Valenta R, Kraft D. Recombinant allergen molecules: tools to study effector cell activation. *Immunol Rev* 2001; 179: 119–27.
- 5 Villalba M, Batanero E, Monsalve RI, Gonzalez de la Pena MA, Lahoz C, Rodriguez R. Cloning and expression of Ole e I, the major allergen from olive tree pollen. Polymorphism analysis and tissue specificity. *J Biol Chem* 1994; 269: 15 217–22.
- 6 Schramm G, Bufe A, Petersen A, Schlaak M, Becker WM. Molecular and immunological characterization of group V allergen isoforms from velvet grass pollen (*Holcus lanatus*). *Eur J Biochem* 1998; 52: 200–6.
- 7 Kawamoto S, Fujimura T, Nishida M, Tanaka T, Aki T, Masubuchi M, *et al.* Molecular cloning and characterization of a new Japanese cedar pollen allergen homologous to plant isoflavone reductase family. *Clin Exp Allergy* 2002; 32: 1064–70.
- 8 Scholl I, Wiedermann U, Forster-Waldl E, Ganglberger E, Baier K, Boltz-Nitulescu G, *et al.* Phage-displayed Bet mim 1, a mimotope of the major birch pollen allergen Bet v 1, induces B cell responses to the natural antigen using bystander T cell help. *Clin Exp Allergy* 2002; 32: 1583–8.
- 9 Yu CJ, Lin YF, Chiang BL, Chow LP. Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. *J Immunol* 2003; 170: 445–53.
- 10 Lüttkopf D, Müller U, Skov PS, Ballmer-Weber BK, Wüthrich B, Skamstrup Hansen K, *et al.* Comparison of four variants of a major allergen in hazelnut (*Corylus avellana*) Cor a 1.04 with the major hazel pollen allergen Cor a 1.01. *Mol Immunol* 2001; 38: 515–25.
- 11 Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* 1998; 26: 1628–35.
- 12 Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988; 16: 10881–90.
- 13 Heger A, Holm L. Sensitive pattern discovery with 'fuzzy' alignments of distantly related proteins. *Bioinformatics* 2003; 19: 1130–7.
- 14 Plewniak F, Bianchetti L, Brelivet Y, Carles A, Chalmel F, Lecomte O, *et al.* PipeAlign: a new toolkit for protein family analysis. *Nucleic Acids Res* 2003; 31: 3829–32.
- 15 Bennett SP, Lu L, Brutlag DL. 3MATRIX and 3MOTIF: a protein structure visualization system for conserved sequence motifs. *Nucleic Acids Res* 2003; 31: 3328–32.
- 16 Higgins D, Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; 22: 4673–80.

- 17 Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 1988; 85: 8998–9002.
- 18 Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001; 17: 1244–5.
- 19 Mirza O, Henriksen A, Ipsen H, Larsen JN, Wissenbach M, Spangfort MD, *et al*. Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *J Immunol* 2000; 165: 331–8.
- 20 Vieths S, Scheurer S, Ballmer-Weber B. Current understanding of cross-reactivity of food allergens and pollen. *Ann NY Acad Sci* 2002; 964: 47–68.
- 21 Fluckiger S, Fijten H, Whitley P, Blaser K, Crameri R. Cyclophilins, a new family of cross-reactive allergens. *Eur J Immunol* 2002; 32: 10–7.
- 22 Bolten E, Schliep A, Schneckener S, Schomburg D, Schrader R. Clustering protein sequences: structure prediction by transitive homology. *Bioinformatics* 2001; 17: 935–41.
- 23 Pearson WR. Identifying distantly related protein sequences. *Comput Appl Biosci* 1997; 13: 325–32.
- 24 Brenner SE, Chothia C, Hubbard TJ. Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships. *Proc Natl Acad Sci USA* 1998; 95: 6073–8.
- 25 Rost B. Twilight zone of protein sequence alignments. *Protein Eng* 1999; 12: 85–94.
- 26 Hileman RE, Silvanovich A, Goodman RE, Rice EA, Holleschak G, Astwood JD, *et al*. Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int Arch Allergy Immunol* 2002; 128: 280–91.
- 27 Furmonaviciene R, Shakib F. The molecular basis of allergenicity: comparative analysis of the three dimensional structures of diverse allergens reveals a common structural motif. *J Clin Pathol Mol Pathol* 2001; 54: 155–9.
- 28 Rafnar T, Griffith IJ, Kuo MC, Bond JF, Rogers BL, Klapper DG. Cloning of Amb a I (Antigen E), the major allergen family of short ragweed pollen. *J Biol Chem* 1991; 266: 1229–36.
- 29 Asturias JA, Arilla MC, Bartolome B, Martinez J, Martinez A, Palacios R. Sequence polymorphism and structural analysis of timothy grass pollen profilin allergen (Phl p11). *Biochim Biophys Acta* 1997; 1352: 253–7.
- 30 Federov AA, Ball T, Mahoney NM, Valenta R, Almo SC. The molecular basis for allergen cross-reactivity: crystal structure and IgE-epitope mapping of birch pollen profilin. *Structure* 1997; 5: 33–45.
- 31 Yagami T. Allergies to cross-reactive plant proteins. Latex-fruit syndrome is comparable with pollen-food allergy syndrome. *Int Arch Allergy Immunol* 2002; 128: 271–9.
- 32 Tao AL, He SH. Bridging PCR and partially overlapping primers for novel allergen gene cloning and expression insert decoration. *World J Gastroenterol* 2004; 10: 2103–8.