medicina

MDPI

Article

Group-1 Grass Pollen Allergens with Near-Identical Sequences Identified in Species of Subtropical Grasses Commonly Found in Southeast Asia

Sirirat Aud-in [1,2,3,†], Koravit Somkid [4,†] and Wisuwat Songnuan [2,5,*]

1 M.Sc. Programme in Plant Sciences, Faculty of Graduate Studies, Mahidol University, Nakhon Pathom 73170, Thailand; sirirat.aud@student.mahidol.edu
2 Department of Plant Science, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
3 Department of Pharmaceutical Botany, Faculty of Pharmacy, Bangkok 10400, Thailand
4 Toxicology graduate programme, Faculty of Science, Mahidol University, Bangkok 10400, Thailand; koravit.sok@gmail.com
5 Systems Biology of Diseases Research Unit, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
* Correspondence: wisuwat.son@mahidol.edu; Tel.: +66-2201-5930
† These authors contributed equally to this work.

Received: 24 December 2018; Accepted: 20 May 2019; Published: 22 May 2019

Abstract: *Background and objectives:* Group-1 grass allergens or beta-expansins (EXPBs) are major allergens from pollen of all grass species. Previous studies showed that they are highly conserved (64–85%) in Pooideae species, which are found mostly in the temperate regions. However, the information about group-1 allergens from common grass species in subtropical areas is still lacking. This study aimed to assess the sequence diversity of group-1 grass pollen allergens in subtropical areas, especially in Southeast Asia. *Materials and Methods:* Group-1 allergens were cloned from pollen of eight grass species using a single set of primers. Sequences were analyzed and IgE and IgG_4 binding regions were compared to the previously reported epitopes in homologous EXPBs. The phylogenetic analysis was used to assess the relationship between sequences of these species and previously characterized EXPBs. Moreover, three-dimensional structure of the EXPB was modeled based on homology to Zea m 1. *Results:* Sequences from eight grass species were nearly identical. It is conceivable that the primers used for cDNA amplification detected the same isoform in different species. In fact, the deduced amino acid sequences shared 97.79–100% identity with each other and 15/819 polymorphic nucleotide positions were identified. The predicted structure showed that the IgE and IgG_4 epitopes and polymorphic residues were located in both domains 1 and 2. The dendrogram presents clustering of class A EXPBs into four groups corresponding to the grass subfamilies. *Conclusions*: This study identified the allergens with near-identical sequences from different grass species. This isoform could be the major cross-reacting allergenic protein from commonly found grass species.

Keywords: group-1 grass allergens; allergic rhinitis; pollen; subtropical grasses; beta-expansin

1. Introduction

Group-1 grass allergens have been defined as major allergens of grass pollen on the basis of high prevalence and potency. The group-1 allergens caused IgE reactivity in more than 90% of grass pollen allergic patients [1]. These allergens have been identified in all grass species, unlike the group-5 allergens that are restricted to grasses of the Pooideae subfamily. In the tropical/subtropical areas where the Panicoideae and Chloridoideae grasses predominate, and temperate trees such as birch and beech are absent, group-1 grass allergens become the most relevant allergens for pollen allergy sufferers.

Group-1 allergens have been reported from several grass species. Phl p 1 from *Phleum pratense* (timothy grass) is among the most extensively studied allergens from the Pooideae grasses. Phl p 1 has high IgE reactivity among grass pollen allergic patients and it has been suggested that Phl p 1 together with Phl p 5 and profiling could be sufficient for a grass pollen allergy diagnosis in the temperate regions [2,3]. In subtropical areas, Cyn d 1, from *Cynodon dactylon* (Bermuda grass) is well-characterized, along with Pas n 1 from *Paspalum notatum* (bahia grass), Lol p 1 from *Lolium perenne* (ryegrass), Zea m 1 from *Zea mays* (maize), and Sor h 1 from *Sorghum halepense* (Johnson grass).

The expansin superfamily contains a large number of 31–35 kDa glycoproteins divided into the following four families based on sequence analysis: alpha-expansin (EXPA), beta-expansin (EXPB), expansin-like A (EXLA), and expansin-like B (EXLB) [4]. All group-1 allergens are classified as a subclass of EXPBs. Expansins are involved in extension and loosening of the extracellular matrix in the plant cell wall [5]. Their biological functions involve pollen tube growth through the female flower. The structure study of Zea m 1 showed that it is composed of two domains: The N-terminal domain 1 that resembles the catalytic domain of family 45 glycoside hydrolases (GH45) and domain 2 that consists of β-sheets with 36% sequence identity to group 2 and 3 grass pollen allergens [6].

Previous studies mostly focused on the characterization of group-1 grass allergens from temperate and subtropical grasses widely distributed in Europe or North America. The group-1 grass allergens are highly conserved, sharing 60–70% sequence identity [7]. Conservation of group-1 grass allergens within each grass subfamily can be even higher, as in the case of Phl p 1, the group-1 allergen of timothy grass that shared 85–95% identity at the amino acid level with EXPBs from other Pooideae species. This high sequence identity leads to high cross-reactivity among grass species [8]. However, information about other grass species is still limited. In subtropical and tropical regions where hundreds of grass species are present and the pollen season is not well-defined, identifying the primary source of grass pollen sensitization is a complicated task. Skin-prick testing with extracts from pollen of all common grass species without other supporting information is not an ideal approach. To help circumvent this problem, it was hypothesized that the sequence identity of grass group-1 allergens or EXPBs from common grass species could help in predicting the degree of cross-reactivity with known allergenic grass pollen species.

This study expands the understanding of sequence diversity of group-1 grass allergens from pollen of subtropical grasses frequently found in Southeast Asia. This is among the first studies to compare beta-expansin sequences obtained from multiple grass species using a single set of primers. We found that these primers yielded PCR products in eight out of ten selected species. Furthermore, all sequences from these eight species were nearly identical. The IgE and IgG_4 binding epitopes and protein structure features were predicted. Clustering analysis was performed to determine the relationship between the sequences obtained in this study and previously reported group-1 allergens. This information is crucial in predicting the contribution of the isoform to the overall allergenicity of the pollen from each grass species, which may lead to the improvement of diagnosis and allergy immunotherapy for grass pollen allergy in subtropical areas, especially in Southeast Asia.

2. Materials and Methods

2.1. Plant Materials

Ten grass species were collected from several sites around Bangkok and metropolitan areas (Table 1). All grass species were identified based on the Key to Flora of North America (Wipff and Thompson, no date) and voucher specimens were preserved by plant taxonomy specialists at the Department of Plant Science, Faculty of Science, Mahidol University, Bangkok, Thailand.

Table 1. Common grass species in Thailand selected for this study.

Scientific Name	Common Name	Reported Allergenicity (Patient Group)	Near Identical EXPB (This Study)	GPS Coordinates
Zoysia matrella (L.) Merr.	Manila grass	No report	Zoy m EXPB	14.079971, 99.702437
Polytrias indica (Houtt.) Veldkamp	Batiki bluegrass	No report	Pol i EXPB	13.850821, 100.527323
Bothriochloa pertusa (L.) A.Camus	Hurricane grass	No report	Bot p EXPB	14.113808, 99.150451
Sorghum halepense (L.) Pers.	Johnson grass	77% (48 GP- allergic), 21% (100 AR with + SPT to common inhalant allergens), 10% (419 AR) [9–11]	Sor h EXPB	14.047468, 99.780618
Saccharum spontaneum L.	Canne sauvage	No report	Sac s EXPB	13.921569, 100.490671
Zea mays L.	Maize, corn	6.25% (48 patients suspected to have nasobronchial allergy) [12]	Zea m EXPB	14.071024, 99.729950
Melinis repens (Willd.) Zizka	Natal grass	No report	Mel r EXPB	14.129449, 99.161192
Eriochloa procera (Retz.) C.E.Hubb.	Cup grass	No report	Eri p EXPB	13.850821, 100.527323
Urochloa mutica (Forssk.) T.Q.Nguyen	Para grass	53.2% (2,383 AR with + SPT to pollen) [13]	-	13.740637, 99.920501
Cynodon dactylon (L.) Pers.	Bermuda grass	84% (48 GP-allergic and AR), 2.1% (419 AR), 12.5% (48 GP- allergic and AR), 54.9% (2,383 AR), 61% (54 Perennial AR), 20.5% (200 As), 52% (133 GP-allergic with + IDST) [9,11–16]	-	14.075435, 99.698798

GP: grass pollen, AR: allergic rhinitis, As: asthma, SPT: skin prick test, IDST: intradermal skin test, +: positive result.

Grass inflorescences were collected from natural sites. To avoid contamination, inflorescences were gathered from areas with few or no other species in close proximity and only a single grass species was processed during each period. Inflorescences were arranged in vessels and allowed to naturally release pollen for one day in a semi-closed area. Pollen grains along with other plant parts were gathered and passed through sieves to obtain pollen grains with high purity. The purity was accessed under a light microscope. Only the pollen samples with a purity of >95% were used. The percentage of pollen purity was calculated as follows: (no. grass pollen grains/[no. grass pollen grains + no. other contaminants]) × 100 = x%. The pollen grains were stored at −80 °C until used.

2.2. RNA Extraction and cDNA Synthesis

For RNA extraction, 100 mg of pollen grains kept at −80 °C were ground to fine powder with liquid nitrogen. Total RNA was isolated using 1 mL TRIzol™ reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. The RNA quality was assessed using agarose gel electrophoresis, and the RNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). Reverse transcription was performed using an iScript™ Select cDNA synthesis kit (Bio-Rad, Hercules, CA, USA).

2.3. RACE PCR Cloning

The rapid amplification of cDNA end (RACE) PCR was conducted using beta-expansin specific primers and a cDNA template from *Cynodon dactylon*, *Zoysia matrella*, *Sorghum halepense*, and *Bothriochloa pertusa*. The 5′ and 3′ RACE cDNA from 1 µg of total RNA was separately synthesized using a SMARTScribe reverse transcriptase (Clontech Laboratories, Mountain View, CA, USA). Universal primer A mix and gene-specific primers were used for amplification: KS-CD1/2 (F: 5′-gattacgccaagcttgccgggccccaacatcactgcaacctac-3′, R: 5′-gattacgccaagcttttggacttgtagacggtgtcgggcttcc-3′) for *C. dactylon* and KS-ZM1/2 (F: 5′-gattacgcc aagcttggggtgcggcaacgagcccatc-3′, R: 5′-gattacgccaagcttaagtttggcgccgctctcgctggt-3′) for *Z. matrella*, *S. halepense*, and *B. pertusa*. The amplification was conducted using touchdown PCR as described in the kit manual.

2.4. Cloning and Sequencing

The beta-expansin (EXPB) cDNAs were amplified by PCR using Vivantis® Taq DNA polymerase (Vivantis, Subang Jaya, Malaysia). The specific primers KS-4.2/p3 (F: 5′-cacatcacattacacagcaggagaaag-3′, R: 5′-ctctaccgacttgtgtgcg-3′) were designed from highly conserved regions of the RACE PCR products. These primers were used for PCR amplification. The amplicons were analyzed by agarose gel electrophoresis and purified by a QIAquick® gel extraction kit (Qiagen, Redwood City, CA, USA). For TA cloning, dATPs were added into the purified PCR products for elongation of the A-tail. The resulting fragments were ligated into pGEM®-T easy vector (Promega, Medison, WI, USA). Transformation into competent *E. coli* DH5α cells was performed using the heat-shock method. For each species, twenty positive clones were picked and colony PCR was performed to select clones with insertions. Selected clones (3–4 clones/species) were cultured overnight and plasmids were extracted from precipitated cells using QIAprep® spin miniprep kit (Qiagen, Germantown, MD, USA). Sanger sequencing was conducted based on the plasmid template by a commercial laboratory (Macrogen, Korea) using M13 primers.

2.5. Sequence Analysis

The nucleotide sequences from forward and reverse sequencing were checked for quality and accuracy based on the electropherograms using BioEdit software [17]. Only sequences of good quality (read length > 1000 bases) were included in the analysis. All sequences were trimmed to begin at the start codon and end at the stop codon prior to the sequence analysis. No insertion/deletion was observed in the consensus sequences. For each species, the clone with the highest percent identity to the consensus sequence was chosen for the intra-species sequence comparison. The nucleotide sequences were subjected to BLASTn homology search using a nucleotide collection (nr/nt) database with Megablast (optimized for highly similar sequences). Deduced amino acid sequences were obtained using the ExPaSy translate tool using standard parameters. Multiple sequence alignments of nucleotide and amino acid sequences were constructed and analyzed using BioEdit. The percent identity was calculated using the Clustal Omega program [18]. The program parameters were set as default: ClustalW with character counts (clustal_num), dealign input sequences (no), mBed-like clustering guide-tree (yes), mBed-like clustering iteration (yes), number of combined iterations (default 0), maximum guide tree iterations (default), maximum HMM iterations (default), and order (align). The percent identity matrix was created using Clustal 2.1 in Clustal Omega program.

2.6. Protein Structure Prediction

The IgE and IgG4 binding epitopes corresponding to Cyn d 1 epitopes were based on previous studies [19–22]. The conserved catalytic sites of family 45 glycoside hydrolases (GH45) enzymes were also predicted based on sequence homology to Zea m 1 [6]. The three-dimensional structure of EXPB was constructed using a template from an automated protein homology-modeling server (SWISS-MODEL, Basel, Switzerland) [23]. The Zea m 1 structure used as a template was extracted from PDB protein databank. The image was generated using the PyMOL molecular graphics system (DeLano Scientific, Palo Alto, CA, USA).

2.7. Phylogenetic Analysis

EXPB sequences of eight grass species in this study were compared to other species from the allergen database reported by the WHO/IUIS. A phylogenetic tree was constructed by MEGA7 [24] using the neighbor-joining method [25]. The phylogeny was tested using the bootstrap method with 100 replicates. The evolutionary distances were computed using the Poisson correction method [26].

3. Results

We assessed the diversity of group-1 grass allergen or beta-expansin (EXPB) sequences from grass species commonly found in subtropical areas such as Thailand. The ten selected species are shown in Table 1. Because the genomes of these grasses were not available, the consensus sequence from RACE-PCR was used for designing the beta-expansin specific primers. One primer pair was used to successfully clone EXPB from eight out of ten selected species. Several clones were obtained from each species (Supplementary Materials Table S1) and only the sequences of good quality were used for further analysis. Not all of the sequences from the same species were identical (percent identity ranged from 95.24–100%). The sequences with the highest percent identity to the consensus sequence was used for inter-species sequence analysis. The amplicons from all eight species had an identical length of 819 base pairs. The full-length encoding regions were translated to 272 amino acids with an approximated molecular weight of 30 kDa, similar to other previously reported EXPBs.

Multiple-sequence alignment showed that the EXPBs obtained from these eight species were remarkably similar, even though no two sequences were identical. The percent identity of nucleotide and amino acid sequences is presented in Table 2. At the nucleotide level, the highest percent identity was 99.88% between Pol i EXPB and Bot p EXPB. Only 1/819 nucleotide was different between these two species. The lowest percent identity (98.78%) was found between Bot p EXPB and Eri p EXPB with ten variations. In total, 15 polymorphic positions were identified. The deduced amino acid sequences of EXPB from the eight species shared 97.79–100% identity with each other. Notably, the deduced amino acid sequence of Zoy m EXPB had a 100% match with that of Pol i EXPB.

Table 2. Percent identity of nucleotide and deduced amino acid sequences of EXPB from eight common grasses.

Nucleotide ↓ / Amino Acid →	1	2	3	4	5	6	7	8	
1	**100.00**	100.00	99.63	99.63	99.26	98.90	98.53	98.53	1. Zoy m EXPB
2	99.76	**100.00**	99.63	99.63	99.26	98.90	98.53	98.53	2. Pol i EXPB
3	99.39	99.63	**100.00**	99.26	98.90	98.53	98.16	98.16	3. Bot p EXPB
4	99.63	99.88	99.51	**100.00**	98.90	98.53	98.16	98.16	4. Sor h EXPB
5	99.63	99.39	99.02	99.27	**100.00**	99.63	98.53	99.26	5. Sac s EXPB
6	99.63	99.39	99.02	99.27	99.76	**100.00**	98.16	99.63	6. Mel r EXPB
7	99.51	99.27	98.90	99.15	99.39	99.39	**100.00**	97.79	7. Zea m EXPB
8	99.39	99.15	98.78	99.02	99.51	99.76	99.15	**100.00**	8. Eri p EXPB

Identity analysis of nucleotide and deduced amino acid sequences was performed by Clustal Omega [18]. The columns indicate nucleotide sequence identity, and the rows indicate deduced amino acid sequence identity. The numbers 1–8 represent eight common grasses.

Alignment of deduced amino acid sequences (Figure 1) showed six variations between Zea m EXPB and Eri p EXPB sequences, leading to the lowest percent identity at 97.79%. In all eight sequences, only nine polymorphic residues were identified (at positions 13, 25, 45, 79, 100, 122, 154, 179, and 225). All sequences contained an identical predicted signal peptide (26 aa) at the N-terminus. The residues that corresponded to the catalytic sites of the family 45 glycoside hydrolases (GH45) enzymes at T52, Y54, C86, H133, and D135 were completely conserved in all sequences. On the basis of the previously characterized human IgE and IgG_4 binding epitopes of Cyn d 1 [19–22], the corresponding epitopes were found in all species. Three changes were observed within these predicted IgE and IgG_4 binding epitopes: S100Y, H179Q, and I225T. The experimental exchangeability for these residues were 0.173, 0.396, and 0.198, respectively [27], suggesting that the H179Q may have had less effect than S100Y and I225T.

```
Zoy m EXPB        MASSSARRQAQLAAVAVLLSAMVGSALCEIGDKPGPNITATYNEEWQDAKATFYGSNPRGAAPDDDGGACGYKNVDKAPF 80
Pol i EXPB        ................................................................................ 80
Bot p EXPB        ........................R.......................................................... 80
Sor h EXPB        ................................................................................ 80
Sac s EXPB        ...............................................K................................ 80
Mel r EXPB        ............T...............................K................................ 80
Zea m EXPB        ...............................................K................................L. 80
Eri p EXPB        ............T...............................K................................ 80
Clustal Consensus ************:*********** ******************:********************************* * 78
                                   IgE/IgG4                          IgE
Zoy m EXPB        FGMTGCGNEPIFKDGLGCGSCFEIKCKEPAECSDKPVLIRITDKNYEHIAAYHFDLSGKAFGSMAKQGQEDNLRKAGELT 160
Pol i EXPB        ................................................................................ 160
Bot p EXPB        ................................................................................ 160
Sor h EXPB        ...................Y............................................................ 160
Sac s EXPB        ...............................................A................................ 160
Mel r EXPB        ...............................................A................................ 160
Zea m EXPB        ...............................................................................C...... 160
Eri p EXPB        ...............................................A................................ 160
Clustal Consensus ******************* ********************:********************************* ****** 156
                                IgG4          IgE/IgG4        IgE/IgG4              IgG4
Zoy m EXPB        LQFRRVKCEYPSKTKITFHVEKGCNDNYLALLVEYAAGDGDIVAVDIKPKGSDEFLPMKPSWGAIWRIDPKKPLKGPFTV 240
Pol i EXPB        ................................................................................ 240
Bot p EXPB        ................................................................................ 240
Sor h EXPB        ................................................................................ 240
Sac s EXPB        ................................................................................ 240
Mel r EXPB        ................................................................................ 240
Zea m EXPB        ...............................................................T................ 240
Eri p EXPB        ..................Q............................................................ 240
Clustal Consensus ******************:********************************************** *************** 235

Zoy m EXPB        RLTSESGAKLVQEDVIPADWKPNTAYTSNLQY 272
Pol i EXPB        ................................ 272
Bot p EXPB        ................................ 272
Sor h EXPB        ................................ 272
Sac s EXPB        ................................ 272
Mel r EXPB        ................................ 272
Zea m EXPB        ................................ 272
Eri p EXPB        ................................ 272
Clustal Consensus ********************************* 267
```

Figure 1. Multiple sequence alignment of deduced amino acid sequences of EXPB from eight common grasses. Sequence alignment was performed using the BioEdit program [17]. Dots represent amino acids that were identical to Zoy m EXPB. Bold residues indicate signal peptide. Rectangular frames indicate catalytic sites identified in GH45 enzymes corresponding residues of EXPB1 (Zea m 1) and EXPB in this study. The IgE and IgG_4 binding epitopes were marked by underlines [6,19–21]. The nucleotide sequences obtained in this study were submitted to GenBank with accession numbers as follows: Zoy m EXPB: MK393185, Pol I EXPB: MK393186, Bot p EXPB: MK393187, Sor h EXPB: MK393188, Sac s EXPB: MK393189, Zea m EXPB: MK393190, Mel r EXPB: MK393191, Eri p EXPB: MK393192.

To gain better insights into the observed polymorphisms, the Zoy m EXPB sequence was used for homology modeling. The crystal structure of beta-expansin protein Zea m 1, with 62.61% identity to the Zoy m EXPB sequence, was selected from PDB as a template model [6]. Figure 2 shows the three-dimensional model structure of Zoy m EXPB. Similar to Zea m 1 structure, Zoy m EXPB contains two domains. Domain 1 consists of a six-stranded β-barrel, short loops, and two α-helices. This domain contains the predicted catalytic site of GH45 enzymes. Domain 2 is composed of eight β-sheets, connected to Domain 1 by a short linker. Five of seven polymorphic residues are found in Domain 1, one in β-barrel, one in α-helix, and three in loop regions. The other two polymorphic residues in Domain 2 are located in the β-sheet region. The predicted IgE/IgG_4 binding epitopes are located in both domains and are mostly exposed on the protein surface.

To assess the relationship of EXPBs obtained from the eight species in this study and previously reported species, a phylogenetic tree was constructed using the neighbor-joining method. Eighteen previously characterized group-1 grass allergens with high (>57%) identity to sequences obtained in this study were retrieved from the GenBank database. The resulting dendrogram (Figure 3) shows that EXPBs could be divided into four subgroups largely corresponding to the grass subfamilies: (I) and (II) Panicoideae and Chlorodoideae subfamilies, (III) Pooideae subfamily, and (IV) Erhartoideae subfamily. Eight allergens from this study were clustered together in subgroup I, with Cyn d 1 (all isoforms) and Uro m 1.0101 as the most closely related sequences. Subgroup II was composed of Sor h 1 and Uro m 1. Beta-expansins from Pooideae grasses were clustered into subgroup III and Ory s 1 was separated into subgroup IV.

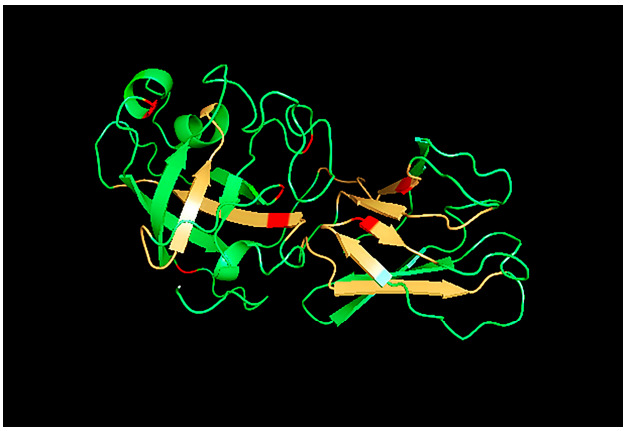

Figure 2. Predicted structure of Zoy m EXPB. The human IgE and IgG_4 binding sites of Cyn d 1 were shown in the three-dimensional structure of EXPB from *Zoysia matrella* (orange). The red residues indicate polymorphisms in eight grass species sequences. The known structure template was crystal structure of EXPB1 (Zea m 1) [6] from an automated protein homology-modeling server (SWISS-MODEL) [23]. The image was generated using the PyMOL molecular graphics system (DeLano Scientific).

4. Discussion

On the basis of the previously reported EXPBs, the expected percent identity between sequences from different grass species is in the range of 64–85% [28]. Surprisingly, this study identified a near-identical (97.79–100%) isoform of EXPBs from eight out of ten selected grass species. Furthermore, this is the first study to report two identical EXPBs from pollen of different grass species.

One possible reason that this near-identical isoform was not characterized previously is that most studies focused on characterization of isoallergens from a single grass species such as Sor h 1 from Johnson grass (2 isoforms) [29], Lol p 1 from rye grass (3 isoforms) [30,31], Hol l 1 from velvet grass (2 isoforms) [32,33], Phl p 1 from timothy grass (2 isoforms) [2,34], and Cyn d 1 from Bermuda grass (5 isoforms) [19,21,35]. The highest sequence identity between isoforms isolated from the same species was found with Cyn d 1 (86.4–99.6%) [18,21]. Because these species were investigated independently, the isoforms were obtained using different conditions. In particular, the PCR primers were designed from different sequences. Hence, the resulting products might not be derived from genes with the highest percent identity. In this study, a common primer pair was used to obtain the PCR products from several species, allowing amplification of the cDNA from the most similar orthologous genes.

On the basis of the clustering analysis, Cyn d 1.0101 (accession no. AAB50734.2), a major allergen from Bermuda grass, is the previously reported allergen most closely related (81.15–81.97%) to the near-identical isoform found in this study. This Cyn d 1 isoform has been shown to have a high frequency of IgE reactivity in grass pollen allergic patients and cross-reacted with Phl p 1 [36,37].

It is likely that the near-identical isoform will have similar IgE reactivity to the Cyn d 1.0101, although a few polymorphisms found within the predicted IgE and IgG_4 epitopes could affect its allergenic potential [21,22].

The near-identical EXPB isoform in this study was cloned from grasses across two grass subfamilies: Chloridoideae (Zoy m EXPB) and Panicoideae (seven other EXPBs), suggesting that this isoform could be considerably prevalent in the grass family rather than limited to closely related taxa. Due to its high percent identity, this isoform could be the major cross-reacting allergenic protein between pollen of commonly found grass species. Nonetheless, the contribution of this isoform to the total allergenic potency of the pollen should be further investigated, since it also depends on the expression level and accessibility of this protein in the context of other EXPBs and other major and minor allergens.

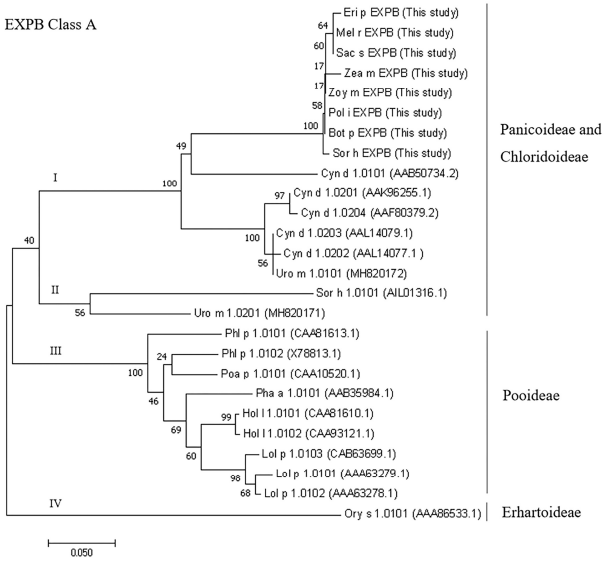

Figure 3. The phylogenetic tree of protein sequences from this study and other grass group 1 pollen allergens. The evolutionary history was constructed in MEGA7 [24] using the neighbor-joining method [25]. The stability of the tree was supported by a bootstrap test with 1000 replicates. The evolutionary distances were computed using the Poisson correction method [26]. Allergen sequences were obtained from the Allergen Nomenclature database and the homologous protein sequences from Blastp analysis. The GenBank accession numbers are indicated after the allergen names.

This study provides supporting evidence that allergen isoforms from different species can have sequences that are more similar than (or identical to) isoforms within the same species. This situation warrants further discussion of the current IUIS Allergen Nomenclature Sub-Committee guideline suggesting that isoallergens are allergens from a single species with >67% sequence identities, and variants of an isoallergen are defined as proteins with >90% sequence identity [38]. Perhaps an additional term such as "orthoallergen" or "homoallergen" should be used to designate identical or near-identical allergens identified from different source species. As more genomic and proteomic data become available, the identification of these identical allergens would be increasingly simple and widespread in the near future.

5. Conclusions

This study expands the understanding of sequence diversity of group-1 grass pollen allergens from subtropical areas. A group-1 allergen isoform was identified from different grass species with high sequence identity. This isoform could be the major cross-reacting allergenic protein between these species. Further investigation of IgE binding of this isoform, especially in comparison with the existing isoforms of beta-expansin, should provide critical information for diagnosis and allergen-specific immunotherapy for subtropical grass pollen allergy.

Supplementary Materials: The following are available online at http://www.mdpi.com/1010-660X/55/5/193/s1, Table S1: Nucleotide sequences from cDNA clones of eight grass species.

Author Contributions: Conceptualization: W.S.; Methodology, S.A. and K.S.; Software, S.A.; Validation, W.S., S.A., and K.S.; Formal analysis, W.S.; Investigation, W.S., S.A., and K.S.; Resources, K.S.; Data curation, S.A.; Writing—original draft preparation, W.S. and S.A.; Writing—review and editing, W.S.; Visualization, W.S.; Supervision, W.S.; Project administration, W.S.; Funding acquisition, W.S.

Funding: This research was funded by the Thailand Research Fund, grant number TRG5780182.

Acknowledgments: The authors gratefully acknowledge S. Dhammachat and W. Yolwong for technical helps, and Thomas N. Stewart for proofreading of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Hrabina, M.; Peltre, G.; Van Ree, R.; Moingeon, P. Grass pollen allergens. *Clin. Exp. Allergy Rev.* **2008**, *8*, 7–11.
2. Valenta, R.; Vrtala, S.; Ebner, C.; Kraft, D.; Scheiner, O. Diagnosis of grass pollen allergy with recombinant timothy grass (*Phleum pratense*) pollen allergens. *Int. Arch. Allergy Immunol.* **1992**, *97*, 287–294.
3. Sastre, J. Molecular diagnosis in allergy. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* **2010**, *40*, 1442–1460.
4. Kende, H.; Bradford, K.; Brummell, D.; Cho, H.-T.; Cosgrove, D.; Fleming, A.; Gehring, C.; Lee, Y.; McQueen-Mason, S.; Rose, J.; et al. *Nomenclature for Members of the Expansin Superfamily of Genes and Proteins*; Kluwer Academic Publishers: Alphen aan den Rijn, The Netherlands, 2004; Volume 55, pp. 311–314.
5. Cosgrove, D.J. Loosening of plant cell walls by expansins. *Nature* **2000**, *407*, 321–326.
6. Yennawar, N.H.; Li, L.-C.; Dudzinski, D.M.; Tabuchi, A.; Cosgrove, D.J. Crystal structure and activities of EXPB1 (Zea m 1), a β-expansin and group-1 pollen allergen from maize. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14664–14671.
7. Cosgrove, D.J.; Bedinger, P.; Durachko, D.M. Group I allergens of grass pollen as cell wall-loosening agents. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6559–6564.
8. Gangl, K.; Niederberger, V.; Valenta, R.; Nandy, A. *Marker Allergens and Panallergens in Tree and Grass Pollen Allergy*; Springer International Publishing: Cham, Switzerland, 2015; Volume 24, pp. 158–169.
9. Morrison, I.C. The changing face of pollen allergy in Queensland. *Med. J. Aust.* **1984**, *141* (Suppl. 5), S12–S13.
10. Pumhirun, P.; Towiwat, P.; Mahakit, P. Aeroallergen sensitivity of Thai patients with allergic rhinitis. *Asian Pac. J. Allergy Immunol.* **1997**, *15*, 183–185.
11. Liang, K.L.; Su, M.C.; Shiao, J.Y.; Wu, S.H.; Li, Y.H.; Jiang, R.S. Role of pollen allergy in Taiwanese patients with allergic rhinitis. *J. Formos. Med Assoc.* **2010**, *109*, 879–885.
12. Prasad, R.; Verma, S.K.; Dua, R.; Kant, S.; Kushwaha, R.A.; Agarwal, S.P. A study of skin sensitivity to various allergens by skin prick test in patients of nasobronchial allergy. *Lung India Off. Organ Indian Chest Soc.* **2009**, *26*, 70–73.
13. Bunnag, C. Airborne pollens in Bangkok, Thailand: An update. *Asian Rhinol. J.* **2016**, *3*, 47–51.
14. Sorensen, H.; Ashoor, A.A.; Maglad, S. Perennial rhinitis in Saudi Arabia. A prospective study. *Ann Allergy* **1986**, *56*, 76–80.
15. Sam, C.K.; Kesavan, P.; Liam, C.K.; Soon, S.C.; Lim, A.L.; Ong, E.K. A study of pollen prevalence in relation to pollen allergy in Malaysian asthmatics. *Asian Pac. J. Allergy Immunol.* **1998**, *16*, 1–4.
16. Singh, A.B.; Shahi, S. Aeroallergens in clinical practice of allergy in India—ARIA Asia Pacific Workshop report. *Asian Pac. J. Allergy Immunol.* **2008**, *26*, 245–256.
17. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
18. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y.M.; Buso, N.; Cowley, A.P.; Lopez, R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41*, W597–W600.
19. Smith, P.M.; Suphioglu, C.; Griffith, I.J.; Theriault, K.; Knox, R.B.; Singh, M.B. Cloning and expression in yeast *Pichia pastoris* of a biologically active form of Cyn d 1, the major allergen of Bermuda grass pollen. *J. Allergy Clin. Immunol.* **1996**, *98*, 331–343.
20. Chang, Z.N.; Peng, H.J.; Lee, W.C.; Chen, T.S.; Chua, K.Y.; Tsai, L.C.; Chi, C.W.; Han, S.H. Sequence polymorphism of the group 1 allergen of Bermuda grass pollen. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* **1999**, *29*, 488–496.
21. Au, L.C.; Lin, S.T.; Peng, H.J.; Liang, C.C.; Lee, S.S.; Liao, C.D.; Chang, Z.N. Molecular cloning and sequence analysis of full-length cDNAs encoding new group of Cyn d 1 isoallergens. *Allergy* **2002**, *57*, 215–220.
22. Yuan, H.C.; Wu, K.G.; Chen, C.J.; Su, S.N.; Shen, H.D.; Chen, Y.J.; Peng, H.J. Mapping of IgE and IgG4 antibody-binding epitopes in Cyn d 1, the major allergen of Bermuda grass pollen. *Int. Arch. Allergy Immunol.* **2012**, *157*, 125–135.
23. Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–3385.
24. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874.
25. Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

26. Zuckerkandl, E.; Pauling, L. Molecules as documents of evolutionary history. *J. Theor. Biol.* **1965**, *8*, 357–366.
27. Yampolsky, L.Y.; Stoltzfus, A. The Exchangeability of Amino Acids in Proteins. *Genetics* **2005**, *170*, 1459–1472.
28. Davies, J.M.; Dang, T.D.; Voskamp, A.; Drew, A.C.; Biondo, M.; Phung, M.; Upham, J.W.; Rolland, J.M.; O'Hehir, R.E. Functional immunoglobulin E cross-reactivity between Pas n 1 of Bahia grass pollen and other group 1 grass pollen allergens. *Clin. Exp. Allergy* **2011**, *41*, 281–291.
29. Campbell, B.C.; Gilding, E.K.; Timbrell, V.; Guru, P.; Loo, D.; Zennaro, D.; Mari, A.; Solley, G.; Hill, M.M.; Godwin, I.D.; et al. Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. *J. Allergy Clin. Immunol.* **2015**, *135*, 133–142.
30. Perez, M.; Ishioka, G.Y.; Walker, L.E.; Chesnut, R.W. cDNA cloning and immunological characterization of the rye grass allergen Lol p I. *J. Biol. Chem.* **1990**, *265*, 16210–16215.
31. Griffith, I.J.; Smith, P.M.; Pollock, J.; Theerakulpisut, P.; Avjioglu, A.; Davies, S.; Hough, T.; Singh, M.B.; Simpson, R.J.; Ward, L.D.; et al. Cloning sequencing of Lol pI, the major allergenic protein of rye-grass pollen. *Febs Lett.* **1991**, *279*, 210–215.
32. Jones, D.H.A.; Christopher, F.; Franklin, H.; Thomas, C.M. Molecular analysis of the operon which encodes the RNA polymerase sigma factor O54 of Escherichia coli. *Microbiology* **1994**, *140*, 1035–1043.
33. Schramm, G.; Petersen, A.; Bufe, A.; Schlaak, M.; Becker, W.M. Identification and characterization of the major allergens of velvet grass (*Holcus lanatus*), Hol l 1 and Hol l 5. *Int. Arch. Allergy Immunol.* **1996**, *110*, 354–363.
34. Petersen, A.; Becker, W.M.; Moll, H.; Blumke, M.; Schlaak, M. Studies on the carbohydrate moieties of the timothy grass pollen allergen Phl p I. *Electrophoresis* **1995**, *16*, 869–875.
35. Suphioglu, C.; Singh, M.B.; Simpson, R.J.; Ward, L.D.; Knox, R.B. Identification of canary grass (Phalaris aquatica) pollen allergens by immunoblotting: IgE and IgG antibody-binding studies. *Allergy* **1993**, *48*, 273–281.
36. Johansen, N.; Weber, R.W.; Ipsen, H.; Barber, D.; Broge, L.; Hejl, C. Extensive IgE cross-reactivity towards the Pooideae grasses substantiated for a large number of grass-pollen-sensitized subjects. *Int. Arch. Allergy Immunol.* **2009**, *150*, 325–334.
37. Davies, J.M. Grass pollen allergens globally: The contribution of subtropical grasses to burden of allergic respiratory diseases. *Clin. Exp. Allergy* **2014**, *44*, 790–801.
38. Radauer, C.; Nandy, A.; Ferreira, F.; Goodman, R.E.; Larsen, J.N.; Lidholm, J.; Pomes, A.; Raulf-Heimsoth, M.; Rozynek, P.; Thomas, W.R.; et al. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy* **2014**, *69*, 413–419.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Supplemental Table S1 Nucleotide sequences from cDNA clones of eight grass species.

Grass species	No. of clones	Seq. identity among clones*	Homologous Cyn d 1 sequence[#]
Zo. matrella	4	99.76-100.0%	Cyn d 1.0203; AF177380.1 (85%)
P. indica	3	99.88-100.00%	Cyn d 1.0203; AF177380.1 (85%)
B. pertusa	4	99.33-99.51%	Cyn d 1.0203; AF177380.1 (85%)
S. halepense	4	99.20-99-30%	Cyn d 1.0203; AF177380.1 (86%)
S. spontaneum	2	100%	Cyn d 1.0203; AF177380.1 (85%)
Ze. mays	4	95.24-99.02%	Cyn d 1.0203; AF177380.1 (83%)
E. procera	4	97.43-100.00%	Cyn d 1.0203; AF177380.1 (85%)
M. repens	3	99.39-99.51%	Cyn d 1.0202; AF177378.1 (85%)

*, Sequence identity among clones was analyzed using Clustal Omega. [#]; Homologous sequences were analyzed using nucleotide Blastn analysis with more than 80**%** sequence identity**.** Percentages of sequence identities are presented in parentheses**.**