IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Genome- and Transcriptome-Wide Analysis of the Major Allergen Genes in Asian White Birch Pollen

Maxwell Shanghao Lu

Haidian Foreign Language Academy, 20 Xingshikou Rd, Haidian District, Beijing, 100195, China; maxwell6624@outlook.com

ABSTRACT: Allergies are immune system diseases affecting a large population in the world. Allergen identification is important for accurate diagnosis and treatment of diseases. It has been shown that Bet v 1 family proteins are the major allergens in European white birch pollen. However, little is known about the allergens in Asian white birch pollen, one of the main allergenic sources in East Asia. Here, I report the identification of 19 *Bet p 1* genes in the Asian white birch genome. They have two exons and one intron and are distributed on three chromosomes. The deduced proteins are highly divergent. All of them contain the Bet v 1 domain with hydrophobic ligand-binding sites and a glycine-rich loop and can form a unique cavity in 3D structures. Phylogenetic analysis shows that 13 *Bet p 1* proteins cluster with the major allergens from the plant family Betulaceae, whereas the remaining six cluster with those from other families. Transcriptome-wide analysis shows that six *Bet p 1* genes are highly expressed in pollen. The results identify and characterize 19 *Bet p 1* genes, six of which appear to encode the major pollen allergen. It provides insight into the major allergens in Asian white birch pollen.

KEYWORDS: Biology, Plant Biology, Allergen, Pollen, Asian White Birch.

Introduction

Allergic diseases, such as asthma, allergic rhinitis, urticaria, eczema, and anaphylactic shock, are a category of globally prevalent chronic disorders caused by abnormal immune system responses, called hypersensitivity reactions. Substances that can provoke an allergic reaction in allergic patients are termed allergens.[1] At the molecular level, allergens are defined as the molecules that induce and bind specific IgE antibodies.[1,2] The sources of allergens are quite diversified, including pollen, foods, house dust mites, chemicals, drugs, and so on. Among them, tree pollen is one of the main allergenic sources and the most important respiratory allergen source.[2]

In the temperate areas of the Northern hemisphere, pollen from Fagales trees, such as European white birch (*Betula verrucosa/B. pendula*) and Asian white birch (*B. platyphylla*), is one of the main causes of hay fever and allergic asthma in spring and winter.[2–4] Gene cloning, recombinant protein production, and clinical studies show that Bet v 1 family proteins are the major allergens in the pollen of these trees.[4,5] This family is named after Bet v 1, the first major birch pollen allergen identified. For nomenclature, members of the family are usually named based on the three letters of the genus and the first letter of the species where the proteins are found.[6] So far, 140 Bet v 1 family proteins from various plant species, such as *B. verrucosa, Corylus avellana*, and *Quercus alba*, have been officially confirmed to be allergens and included in the WHO/IUIS allergen nomenclature database.[6] Among them, 27 proteins are from European white birch, a tree species widely distributed from Europe to central Siberia and Northwest of China. However, there is no such protein allergen from Asian white birch in the database.

Asian white birch is a deciduous tree species belonging to the family Betulaceae in the order Fagales and widely distributed in China, Russia, Mongolia, North Korea, and Japan.[7] Its pollen is one of the main allergenic sources in East Asia and the far east of Russia.[8,9] Compared with the allergens in European white birch that have been intensively studied, the information on Asian white birch pollen allergens is very limited.[8,9] No gene sequence coding for Asian white birch pollen allergens has been identified and characterized. To understand the allergens in Asian white birch, the gene family encoding Bet v 1-like proteins has been identified in the Asian white birch genome. The deduced proteins are comparatively analyzed with European white birch Bet v 1 proteins. Transcriptome-wide analysis of gene expression in Asian white birch pollen is also performed. The study will provide useful information for further investigation of the major allergens in Asian white birch pollen.

Methods

Genome-Wide Search:

The known Bet v 1 protein sequences are downloaded from the WHO/IUIS allergen nomenclature database.[6] The whole genome assembly of *Betula platyphylla* Sukaczev published by Chen *et al.* in 2021 is downloaded from JGI Data Portal (https://data.jgi.doe.gov/refine-download/phytozome?organism=Bplatyphylla&expanded=679&phytozome_version=14).[7] The BLAST software used for genome-wide search of the major allergen genes is downloaded from the NCBI website (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/).[10] Genome-wide search of known Bet v 1 protein sequences against the genome of *B. platyphylla* is carried out using the tBLASTn algorithm.[6,7,10] An e-value cutoff of 1×10^{-5} is applied.

DOI: 10.36838/IJHSR814.24

Gene Model Prediction:

The corresponding DNA sequences are extracted from the *B. platyphylla* genome assembly using the Fasta Extraction function of TBtools-II.[7,11] Gene model prediction is carried out through alignment of the retrieved genomic DNA sequences with Bet v 1-like protein sequences from other plant species using the BLASTx algorithm with the default parameters (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Comparison of the predicted gene models with the results from whole genome annotation is performed using the BLASTn algorithm with the default parameters.[7,10]

Chromosomal and Subcellular Localization:

The location of a gene on the chromosome is determined through comparison analysis of gene sequence with the whole genome assembly of *B. platyphylla* and visualized using the Show Genes on Chromosome function of TBtools-II.[7,11] Protein subcellular localization is predicted using the WoLF PSORT tool by selecting the organism type to be plant (https://www.genscript.com/wolf-psort.html?src=leftbar).

Sequence Feature Analysis:

The length of the coding sequence, the length of the deduced protein, the molecular weight, and the theoretical isoelectric points are determined using the ProtParam tool (https://web.expasy.org/protparam/). The conserved domain, hydrophobic ligand binding site, and glycine-rich loop are analyzed by searching the Conserved Domain Database using the CD-Search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Gene structures are determined by comparing the ORF with the *B. platyphylla* genome assembly.[7] The alignment of the amino acid sequence is carried out using the ClustalW function of MEGA12 with the default parameters.[12] Protein sequence identities are analyzed using the BLASTp algorithm with the default parameters (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The "align two or more sequences" function is selected.

Protein 3D Structure Modeling and Phylogenetic Tree Construction:

Protein 3D structures are computed by the SWISS-MODEL server homology modeling pipeline (https://swissmodel.expasy.org/interactive).[13] For phylogenetic tree construction, protein sequences are aligned using ClustalW in MEGA12 with the default parameters. Then, phylogenetic analysis is performed using the neighbor-joining method with a bootstrap value of 1000 replicates.[12]

Transcriptome Analysis:

B. platyphylla pollen transcriptomic data is downloaded from the National Genomics Data Center (https://ngdc.cncb.ac.cn/bioproject/browse/insdc/PRJNA994611).[14] The data is analyzed, and Transcripts Per Million (TPM) is calculated using the Salmon software (v1.10.3).[15] TPM in log scale is visualized using the HeatMap Illustrator function of TBtools-II.[11]

Results and Discussion

Genome-Wide Identification of 19 Major Allergen Genes in Betula platyphylla:

In order to identify the major allergen genes in Asian white birch, tBLASTn analysis of known Bet v 1 protein sequences against the genome of *B. platyphylla* is performed.[6,7,10] Subsequent gene model prediction results in the identification of 19 putative major allergen genes in *B. platyphylla* (Table 1). Based on the nomenclature of allergens, the identified genes are designated as *Bet p 1.0101–Bet p 1.0103, Bet p 1.0201, Bet p 1.0202, Bet p 1.0301, Bet p 1.0401–Bet p 1.0403, Bet p 1.0501–Bet p 1.0503, Bet p 1.0601, Bet p 1.0701, Bet p 1.0801, Bet p 1.0901, Bet p 1.1001, Bet p 1.1101,* and *Bet p 1.1201*, respectively.[2] In the names, *Bet* is the first three letters of the genus *Betula*, *p* is the first letter of the species *platyphylla*, and the number is given based on its relationship with known *Bet v 1* genes.[6] Gene model comparative analysis shows that five of the gene models identified in this study, including *Bet p 1.0101, Bet p 1.0202, Bet p 1.0401, Bet p 1.0801,* and *Bet p 1.1101*, are identical to *BPChr01G13767, BPChr05G14788, BPChr01G13827, BPChr02G00364,* and *BPChr01G13654* predicted through whole genome annotation.[7] In addition, a portion of *Bet p 1.0201* is annotated as *BPChr05G14860*. The other 13 gene models have not been identified previously.

Table 1: Features of *Bet p 1* genes and their deduced proteins. The length of the deduced protein, the molecular weight, the theoretical isoelectric points, and the subcellular localization are similar among Bet p 1 proteins.

Gene name	Chromosome localization	Length of coding sequence (bp)	Length of protein (AA)	Molecular weight (kDa)	Theoretical isoelectric points (p*I*)	Predicted subcellular localization
Bet p 1.0101	Chr01	480	160	17.56	5.59	Cytoplasmic
Bet p 1.0102	Chr01	480	160	17.55	5.59	Cytoplasmic
Bet p 1.0103	Chr01	480	160	17.55	5.59	Cytoplasmic
Bet p 1.0201	Chr05	480	160	17.61	5.97	Cytoplasmic
Bet p 1.0202	Chr05	480	160	17.52	5.74	Cytoplasmic
Bet p 1.0301	Chr01	480	160	17.42	5.74	Cytoplasmic
Bet p 1.0401	Chr01	480	160	17.64	4.89	Chloroplast
Bet p 1.0402	Chr01	480	160	17.81	5.67	Cytoplasmic
Bet p 1.0403	Chr01	438	146	16.03	5.40	Cytoplasmic
Bet p 1.0501	Chr01	480	160	17.51	5.83	Cytoplasmic
Bet p 1.0502	Chr01	480	160	17.41	5.83	Cytoplasmic
Bet p 1.0503	Chr01	480	160	17.43	5.96	Cytoplasmic
Bet p 1.0601	Chr01	480	160	17.51	6.43	Cytoplasmic
Bet p 1.0701	Chr01	483	161	17.60	4.78	Cytoplasmic
Bet p 1.0801	Chr02	474	158	17.43	5.36	Cytoplasmic
Bet p 1.0901	Chr01	486	162	18.13	5.58	Cytoplasmic
Bet p 1.1001	Chr01	480	160	17.23	5.22	Cytoplasmic
Bet p 1.1101	Chr01	483	161	17.54	4.44	Cytoplasmic
Bet p 1.1201	Chr01	483	161	18.12	9.12	Cytoplasmic

Characteristics of Bet p 1 Genes:

Chromosome localization analysis shows that the 19 identified *Bet p 1* genes are distributed on three of the 14 *B. platyphylla* chromosomes, including Chr01, Chr02, and Chr05 (Figure 1).[7] The distribution is uneven, with 16 on Chr01, one on Chr02, and two on Chr05. The 16 *Bet p 1* genes on Chr01 are clustered together, indicating they are generated through tandem duplication and subsequent mutation.[16] Gene structure analysis shows that all of the identified *Bet p 1* genes contain two exons and one intron (Figure 2). The length of the coding sequence varies from 438 to 486 base pairs (bp), with

the majority being 480 bp (Table 1). These features are similar to *Bet v 1* genes from other plant species.[17,18]

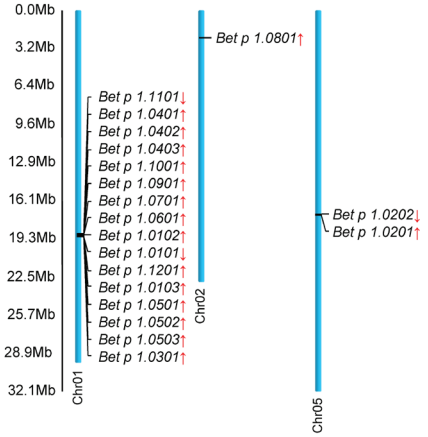

Figure 1: Chromosome localization of *Bet p 1* genes. Nineteen *Bet p 1* genes are unevenly distributed on three chromosomes.

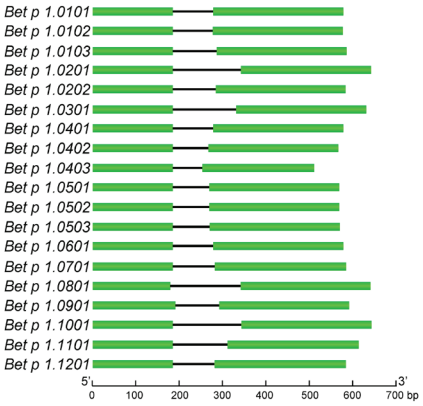

Figure 2: Structures of *Bet p 1* genes. Green boxes indicate exons. Black lines represent introns. Gene structure analysis shows that all of the 19 *Bet p 1* genes have two exons and one intron.

Sequence Features of the Deduced Bet p 1 Protein:

Sequence feature analysis shows that the length of the deduced Bet p 1 proteins varies from 146 to 162 amino acid residuals (AA) with the majority to be 160 AA, the molecular weight varies from 16.03 to 18.13 kDa with the majority to be about 17.00 kDa, and the theoretical isoelectric points varies from 4.44 to 9.12 with the majority to be 5.50 to 6.00 (Table 1). It suggests that the majority of Bet p 1 proteins are small and slightly acidic. These features are consistent with the Bet v 1 family proteins from other plant species.[2,19,20] Computational subcellular localization analysis predicts that 18 of the 19 Bet p 1 proteins are cytoplasmic (Table 1), which is a common feature of Bet v 1 family proteins.[20]

Sequence identity analysis shows that the identity among Bet p 1 proteins varies from 98% to 34% (Table 2). More than half of the data shown in Table 2 is less than 60%. Even for the Bet p 1 proteins clustering on Chr01, some of them share very low identity. For instance, the identity between Bet p 1.0901 and Bet p 1.1101 is only 34%, which is the lowest identity among Bet p 1 proteins. Low sequence identity is also found among other Bet v 1 family proteins.[18] It suggests that Bet v 1 family proteins are highly divergent.

Table 2: Percentages of sequence identity among Bet p 1 proteins. The data highlight the divergence of the Bet v 1 family proteins.

Protein No.*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-																		
2	98	-																	
3	94	94	-																
4	86	86	88	-															
5	87	87	88	98	-														
6	71	71	74	74	74	-													
7	47	47	46	48	48	51	-												
8	47	47	48	50	51	52	63	-											
9	43	43	45	47	48	50	58	87	-										
10	86	86	86	88	89	75	47	50	46	-									
11	86	86	86	89	89	74	47	51	47	96	-								
12	85	85	86	88	89	73	46	50	46	96	98	-							
13	71	71	71	71	71	78	47	48	47	72	73	71	-						
14	65	65	66	69	69	66	53	53	50	68	70	69	65	-					
15	62	61	62	65	66	64	47	51	49	64	64	63	61	61	-				
16	56	56	54	54	56	60	45	45	43	56	56	56	62	56	57	-			
17	47	47	48	48	50	49	48	52	48	50	49	48	51	52	48	53	-		
18	41	41	40	39	41	39	38	42	40	41	41	40	37	39	36	34	39	-	
19	69	69	69	69	69	67	44	44	40	72	69	71	63	51	49	46	43	38	-

* 1. Bet p 1.0101; 2. Bet p 1.0102; 3. Bet p 1.0103; 4. Bet p 1.0201; 5. Bet p 1.0202; 6. Bet p 1.0301; 7. Bet p 1.0401; 8. Bet p 1.0402; 9. Bet p 1.0403; 10. Bet p 1.0501; 11. Bet p 1.0502; 12. Bet p 1.0503; 13. Bet p 1.0601; 14. Bet p 1.0701; 15. Bet p 1.0801; 16. Bet p 1.0901; 17. Bet p 1.1001; 18. Bet p 1.1101; 19. Bet p 1.1201.

Previous studies show that the known Bet v 1 family proteins with allergenicity contain highly conserved Bet v 1 domains.[2] To determine whether the identified Bet p 1 proteins contain such a domain, BLASTp analysis of the 19 identified Bet p 1 proteins against the Conserved Domain Database is performed. The results show that all of the identified Bet p 1 proteins contain the Bet v 1 domain (Figure 3). In the domain, up to 35 hydrophobic ligand-binding sites and a glycine-rich loop consisting of nine amino acid residues are identified (Figure 3). The hydrophobic ligand-binding sites can bind various physiological ligands, such as flavones, sterols, cytokinins, and fatty acids, whereas the glycine-rich loop can serve as an IgE epitope.[21,22] The existence of a glycine-rich loop in all of the identified Bet p 1 proteins indicates their allergenicity.

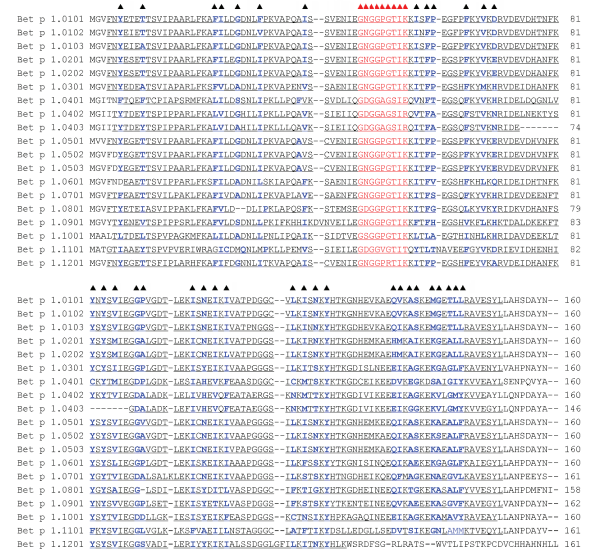

Figure 3: Sequence alignment of Bet p 1 proteins. The conserved Bet v 1 domain is underlined. The hydrophobic ligand binding site is indicated by black solid triangles and blue letters. The glycine-rich loop is indicated by red solid triangles and red letters. All of the 19 Bet p 1 proteins contain the Bet v 1 domain with up to 35 hydrophobic ligand-binding sites and a glycine-rich loop.

DOI: 10.36838/IJHSR814.24

Characterization of the 3D Structure of Bet p 1 Proteins:

Typical Bet v 1 family proteins share a structure with seven-stranded antiparallel β-sheets and three α-helices arranged as β-α_2-β_6-α.[2,23,24] They form a unique cavity for binding ligands.[2] To analyze whether the identified Bet p 1 proteins can form such a cavity, their 3D structures are computed using the SWISS-MODEL server homology modeling pipeline.[13] As shown in Figure 4, all of the 19 Bet p 1 proteins can form a cavity. The cavities formed by 17 of the 19 Bet p 1 proteins are similar to those of Bet v 1.0101. However, those formed by the other two, including Bet p 1.0403 and Bet p 1.1201, are slightly different. Bet p 1.0403 lacks the 5th β-sheet. Its 4th β-sheet is shorter than that of Bet v 1.0101 and other Bet p 1 proteins. For Bet p 1.1201, its 3 ' α-helix is short compared with Bet v 1.0101 and other Bet p 1 proteins (Figure 4). It indicates that the capability of Bet p 1.0403 and Bet p 1.1201 in binding ligands is different from that of Bet v 1.0101 and other Bet p 1 proteins. In addition, the glycine-rich loop connecting the 2nd and the 3rd β-sheets is located on the surface of the 3D structure of all proteins analyzed (Figure 4).[24] The result is consistent with the function of the glycine-rich loop as an IgE epitope.[21,22]

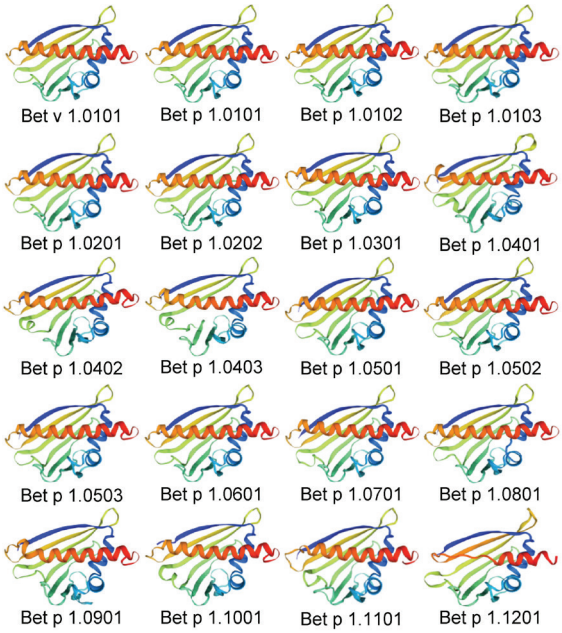

Figure 4: Predicted 3D structures of Bet v 1.0101 and Bet p 1 proteins. The results show that all of the proteins analyzed can form a cavity.

Phylogenetic Relationships of Bet p 1 Proteins:

It was observed that multiple sequences of Bet v 1 family proteins from a plant family formed a monophyletic group.[12,23] To investigate whether the observation can be applied to Bet p 1 proteins, a neighbor-joining tree is constructed for 19 Bet p 1 proteins and 48 known Bet v 1 family proteins from other plant species. As shown in Figure 5, 13 of the Bet p 1 proteins, including Bet p 1.0101–Bet p 1.0103, Bet p 1.0201, Bet p 1.0202, Bet p 1.0301, Bet p 1.0501–Bet p 1.0503, Bet p 1.0601, Bet p 1.0701, Bet p 1.0801, and Bet p 1.1201, cluster with proteins from the plant family Betulaceae. The results for these proteins are consistent with previous observations.[12,23] However, the other six proteins, including Bet p 1.0401–Bet p 1.0403, Bet p 1.0901, Bet p 1.1001, and Bet p 1.1101, cluster with proteins from other plant families, such as Actinidiaceae, Apiaceae, Anacardiaceae, and Fabaceae (Figure 5). It suggests the complex phylogenetic relationships of Bet p 1 proteins.

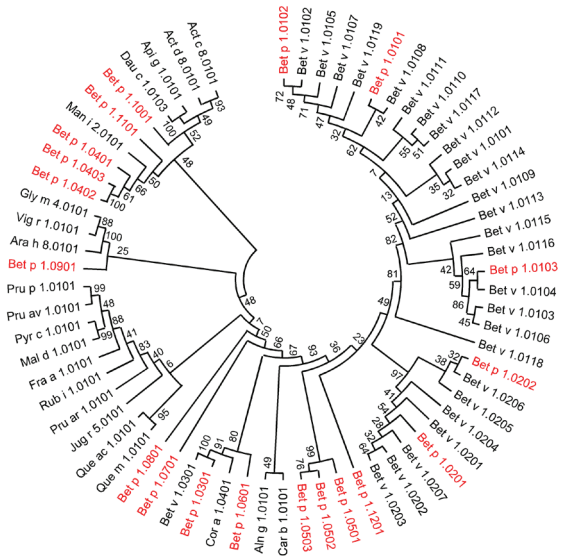

Figure 5: The neighbor-joining tree of Bet p 1 proteins and known Bet v 1 family proteins from other plant species. The percentage of replicate trees is shown next to the branches. Bet p 1 proteins are shown in red. Act c, gold kiwi fruit (*Actinidia chinensis*); Act d, green kiwi fruit (*Actinidia deliciosa*); Aln g, alder (*Alnus glutinosa*); Api g, celery (*Apium graveolens*); Ara h, peanut (*Arachis hypogaea*); Bet p, Asian white birch (*B. platyphylla*); Bet v, European white birch (*B. verrucosa*); Car b, hornbeam (*Carpinus betulus*); Cor a, hazelnut (*Corylus avellana*); Dau c, carrot (*Daucus carota*); Fra a, strawberry (*Fragaria ananassa*); Gly m, soybean (*Glycine max*); Jug r, English walnut (*Juglans regia*); Mal d, apple (*Malus domestica*); Man i, mango (*Mangifera indica*); Pru ar, apricot (*Prunus armeniaca*); Pru av, sweet cherry (*Prunus avium*); Pru p, peach (*Prunus persica*); Pyr c, pear (*Pyrus communis*); Que ac, sawtooth oak (*Quercus acutissima*); Que m, Mongolian oak (*Quercus mongolica*); Rub i, red raspberry (*Rubus idaeus*); Vig r, mung bean (*Vigna radiata*). The tree shows that the Bet v 1 family proteins from a plant family form a polyphyletic group.

Transcriptome Analysis of Bet p 1 Gene Expression in Pollen:

In order to further identify the major allergens in Asian white birch pollen, the expression of *Bet p 1 genes* in *B. platyphylla* pollen is carried out using the transcriptomic data published previously.[14] The data set includes three biological replicates, and a pollen sample for each replicate is collected from three *B. platyphylla* plants.[14] As shown in Figure 6, the expression patterns can be divided into two groups. *Six Bet p 1* genes, including *Bet p 1.0101–Bet p 1.0103, Bet p 1.0201, Bet p 1.0202*, and *Bet p 1.0301*, cluster in a group with relatively higher expression levels than other *Bet p 1* genes. Among them, *Bet p 1.0101* has the highest expression level, followed by *Bet p 1.0102, Bet p 1.0202, Bet p 1.0201, Bet p 1.0103,* and *Bet p 1.0301* (Figure 6). All of them encode a protein with typical allergen features, such as 160 amino acid residues in length, a theoretical pI within 5–6, a molecular weight of about 17 kDa, containing the Bet v 1 domain, and being able to form a 3D structure similar to the well-known Bet v 1 protein. In addition, their deduced proteins show close phylogenetic relationships with the major allergens identified in European white birch pollen (Figure 5). These Bet p 1 genes seem to be the major allergen genes in Asian white birch pollen.

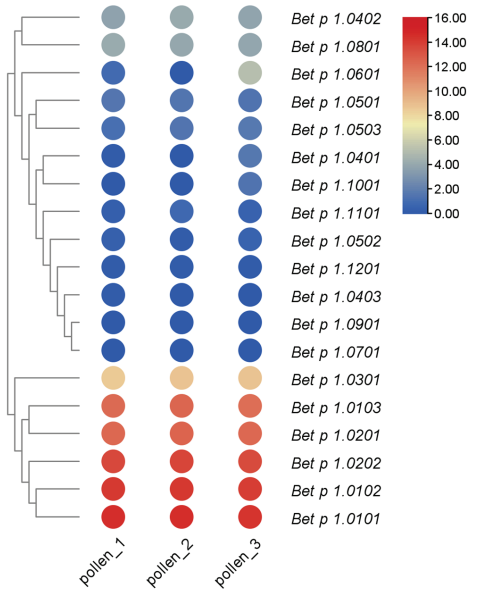

Figure 6: Heatmap of *Bet p 1* gene expression in *B. platyphylla* pollen. Relative gene expression level is analyzed using the transcriptomic data published previously and quantified by TPM on a log scale.[14] Pollen_1, pollen_2 and pollen_3 indicate three biological replicates. *Bet p 1* genes are clustered based on the similarity of gene expression profiles. The color scale represents standardized expression levels with red for high and blue for low. Gene expression analysis shows that six *Bet p 1* genes, including *Bet p 1.0101–Bet p 1.0103, Bet p 1.0201, Bet p 1.0202*, and *Bet p 1.0301*, are highly expressed in pollen.

Conclusion

Allergies, a category of immune system diseases affecting a large population in the world, are a significant public health burden. Identification of allergens that can trigger severe immune responses in hypersensitive individuals is important for accurate diagnosis and subsequent treatment of allergic diseases. Despite extensive efforts having been made in this area, current knowledge of allergens is still limited. For instance, no allergen gene has been identified and characterized from Asian white birch pollen, one of the main allergenic sources in East Asia in spring. Through a comprehensive approach, integrating genome-wide identification, gene sequence characterization, protein feature analysis, conserved domain identification, 3D structure computation, phylogenetic tree construction, and transcriptome-wide gene expression analysis, a total of 19 *Bet p 1* genes are identified and characterized. Six of them are found to be the major allergen genes in Asian white birch pollen.

Among the 19 *Bet p 1* genes, 13 gene models have not been identified previously through whole genome annotation, and one is partially annotated. Chromosome localization analysis identifies a genomic region with 16 *Bet p 1* genes clustered. All of the 19 *Bet p 1* genes encode proteins containing the conserved Bet v 1 domain with up to 35 hydrophobic ligand-binding sites and a glycine-rich loop that can serve as an IgE epitope. 3D structure computation shows that 17 of the 19 deduced Bet p 1 proteins share a structure similar to the typical Bet v 1 family proteins, whereas the other two are slightly different. In addition, different from previous observations that show Bet v 1 family proteins from a plant family forming a monophyletic group,[12,23] the results from this study suggest that 13 of the Bet p 1 proteins cluster with proteins from the plant family Betulaceae, whereas the other six cluster with proteins from other plant families. One possibility causing the difference is that previous studies are not based on a whole-genome scale.

This study identifies six *Bet p 1* genes as putative major allergen genes in Asian white birch pollen. They are highly expressed in pollen and encode a protein with typical allergen features. The results do not exclude the allergenicity of the other 13 Bet p 1 proteins. On the contrary, they could also be allergens, since all of them contain the glycine-rich loop located on the surface of protein 3D structures. The tissues in which these *Bet p 1* genes are expressed remain to be identified. Further experimental tests will confirm the allergenicity of Bet p 1 proteins and will be beneficial to the diagnosis and treatment of allergic diseases caused by Asian white birch.

Acknowledgments

I highly appreciate my family for their research advice and financial support. I would also like to thank my mentor, Dr. Xu, for her continuous encouragement throughout the research.

References

1. Aalberse, R. C. Structural Biology of Allergens. *J. Allergy Clin. Immunol.* **2000**,*106* (2), 228–238. DOI:10.1067/mai.2000.108434.
2. Dramburg, S.; Hilger, C.; Santos, A. F.; de Las Vecillas, L.; Aalberse, R. C.; Acevedo, N.; Aglas, L.; Altmann, F.; Arruda, K. L.; Asero, R.; Ballmer-Weber, B.; Barber, D.; Beyer, K.; Biedermann, T.; Bilo, M. B.; Blank, S.; Bosshard, P. P.; Breiteneder, H.; Brough, H. A.; Bublin, M.; Campbell, D.; Caraballo, L.; Caubet, J. C.; Celi, G.; Chapman, M. D.; Chruszcz, M.; Custovic, A.; Czolk, R.; Davies, J.; Douladiris, N.; Eberlein, B.; Ebisawa, M.; Ehlers, A.; Eigenmann, P.; Gadermaier, G.; Giovannini, M.; Gomez, F.; Grohman, R.; Guillet, C.; Hafner, C.; Hamilton, R. G.; Hauser, M.; Hawranek, T.; Hoffmann, H. J.; Holzhauser, T.; Iizuka, T.; Jacquet, A.; Jakob, T.; Janssen-Weets, B.; Jappe, U.; Jutel, M.; Kalic, T.; Kamath, S.; Kespohl, S.; Kleine-Tebbe, J.; Knol, E.; Knulst, A.; Konradsen, J. R.; Korošec, P.; Kuehn, A.; Lack, G.; Le, T. M.; Lopata, A.; Luengo, O.; Mäkelä, M.; Marra, A. M.; Mills, C.; Morisset, M.; Muraro, A.; Nowak-Wegrzyn, A.; Nugraha, R.; Ollert, M.; Palosuo, K.; Pastorello, E. A.; Patil, S. U.; Platts-Mills, T.; Pomés, A.; Poncet, P.; Potapova, E.; Poulsen, L. K.; Radauer, C.; Radulovic, S.; Raulf, M.; Rougé, P.; Sastre, J.; Sato, S.; Scala, E.; Schmid, J. M.; Schmid-Grendelmeier, P.; Schrama, D.; Sénéchal, H.; Traidl-Hoffmann, C.; Valverde-Monge, M.; van Hage, M.; van Ree, R.; Verhoeckx, K.; Vieths, S.; Wickman, M.; Zakzuk, J.; Matricardi, P. M.; Hoffmann-Sommergruber, K. EAACI Molecular Allergology User's Guide 2.0. *Pediatr. Allergy Immunol.* **2023**, *34* (Suppl. 28), e13854. doi: 10.1111/pai.13854.
3. Schenk, M. F.; Cordewener, J. H.; America, A. H.; Peters, J.; Smulders, M. J.; Gilissen, L. J. Proteomic Analysis of the Major Birch Allergen Bet v 1 Predicts Allergenicity for 15 Birch Species. *J. Proteomics* **2011**, *74* (8), 1290–300. DOI: 10.1016/j.jprot.2011.03.021.
4. Breiteneder, H.; Pettenburger, K.; Bito, A.; Valenta, R.; Kraft, D.; Rumpold, H.; Scheiner, O.; Breitenbach, M. The Gene Coding for the Major Birch Pollen Allergen Betv1 is Highly Homologous to a Pea Disease Resistance Response Gene. *EMBO J.* **1989**, *8* (7), 1935–1938. DOI: 10.1002/j.1460-2075.1989.tb03597.x.

DOI: 10.36838/IJHSR814.24

5. Breiteneder, H.; Hassfeld, W.; Pettenburger, K.; E Jarolim, Breitenbach, M.; Rumpold, H.; Kraft, D.; Scheiner, O. Isolation and Characterization of Messenger RNA from Male Inflorescences and Pollen of the White Birch (*Betula verrucosa*). *Int. Arch. Allergy Appl. Immunol.* **1988**, *87* (1), 19–24. DOI: 10.1159/000234643.
6. Chan, S. K.; Pomés, A.; Hilger, C.; Davies, J. M.; Mueller, G.; Kuehn, A.; Lopata, A. L.; Gadermaier, G.; van Hage, M.; Raulf, M.; Goodman, R. E.; on behalf of the WHO IUIS Allergen Nomenclature Sub-Committee. Keeping Allergen Names Clear and Defined. *Front. Immunol.* **2019**, *10*: 2600. DOI: 10.3389/fimmu.2019.02600.
7. Chen, S.; Wang, Y.; Yu, L.; Zheng, T.; Wang, S.; Yue, Z.; Jiang, J.; Kumari, S.; Zheng, C.; Tang, H.; Li, J.; Li, Y.; Chen, J.; Zhang, W.; Kuang, H.; Robertson, J. S.; Zhao, P. X.; Li, H.; Shu, S.; Yordanov, Y. S.; Huang, H.; Goodstein, D. M.; Gai, Y.; Qi, Q.; Min, J.; Xu, C.; Wang, S.; Qu, G. Z.; Paterson, A. H.; Sankoff, D.; Wei, H.; Liu, G.; Yang, C. Genome Sequence and Evolution of *Betula platyphylla*. *Hortic. Res.* **2021**, *8* (1), 37. DOI: 10.1038/s41438-021-00481-7.
8. Xu, Y.; Gepp, B.; Lengger, N.; Yin, J.; Breiteneder, H. Identification of Probable Pectinesterase as a Major Allergen of Pollen of the Asian White Birch (Betula platyphylla) in Northern China. *Asian Pac. J. Allergy Immunol*. **2023**, Online Ahead of Print. DOI: 10.12932/AP-100722-1409.
9. Ning, X.; Kuang, Y.; Zhao, S.; Hou, W.; Yang, G.; Zhu, X.; Liu, R.; Huang, J. Design of an Optimally-Diagnostic Skin Test for Diagnosis of Sensitivity to Eight Allergens: A First-in-Human Study of Dose Escalation and Simultaneous Administration in Chinese Subjects. *J. Asthma Allergy* **2020**, *13*, 471–481. DOI: 10.2147/JAA.S276720.
10. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17): 3389–3402. DOI: 10.1093/nar/25.17.3389.
11. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y.; Xia, R. TBtools-II: A "One for All, All for One" Bioinformatics Platform for Biological Big-Data Mining. *Mol. Plant* **2023**, *16* (11), 1733–1742. DOI: 10.1016/j.molp.2023.09.010.
12. Kumar, S.; Stecher, G.; Suleski, M.; Sanderford, M.; Sharma, S.; Tamura, K. MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Mol. Biol. Evol.* **2024**, *41* (12), msae263. DOI: 10.1093/molbev/msae263.
13. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. DOI: 10.1093/nar/gky427.
14. Zhang, J.; Shi, J.; Zeng, K.; Cai, M.; Lan, X. Transcriptomic Landscape of Staminate Catkins Development during Overwintering Process in *Betula platyphylla*. *Front. Plant Sci.* **2024**, *14*, 1249122. DOI: 10.3389/fpls.2023.1249122.
15. Sahraeian, S. M. E.; Mohiyuddin, M.; Sebra, R.; Tilgner, H.; Afshar, P. T.; Au, K. F.; Asadi, N. B.; Gerstein, M. B.; Wong, W. H.; Snyder, M. P.; Schadt, E.; Lam, H. Y. K. Gaining Comprehensive Biological Insight Into the Transcriptome by Performing A Broad-Spectrum RNA-Seq Analysis. *Nat. Commun.* **2017**, *8* (1), 59. DOI: 10.1038/s41467-017-00050-4.
16. Cannon, S. B.; Mitra, A.; Baumgarten, A.; Young, N. D.; May, G. The Roles of Segmental and Tandem Gene Duplication in the Evolution of Large Gene Families in *Arabidopsis thaliana*. *BMC Plant Biol.* **2004**, *4*, 10. DOI: 10.1186/1471-2229-4-10.
17. Hoffmann-Sommergruber, K.; Vanek-Krebitz, M.; Radauer, C.; Wen, J.; Ferreira, F.; Scheiner, O.; Breiteneder, H. Genomic Characterization of Members of the Bet v 1 family: Genes Coding for Allergens and Pathogenesis-Related Proteins Share Intron Positions. *Gene* **1997**, *197* (1-2), 91–100. DOI: 10.1016/s0378-1119(97)00246-1.
18. Wen, J.; Vanek-Krebitz, M.; Hoffmann-Sommergruber, K.; Scheiner, O.; Breiteneder, H. The Potential of Betv1 Homologues, A Nuclear Multigene Family, as Phylogenetic Markers in Flowering Plants. *Mol. Phylogenet. Evol.* **1997**, *8* (3), 317–333. DOI: 10.1006/mpev.1997.0447.
19. Franz-Oberdorf, K.; Eberlein, B.; Edelmann, K.; Hücherig, S.; Besbes, F.; Darsow, U.; Ring, J.; Schwab, W. Fra a 1.02 Is the Most Potent Isoform of the Bet v 1-like Allergen in Strawberry Fruit. *J. Agric. Food Chem.* **2016**, *64* (18), 3688–3696. DOI: 10.1021/acs.jafc.6b00488.
20. Breiteneder, H.; Kraft, D. The History and Science of the Major Birch Pollen Allergen Bet v 1. *Biomolecules* **2023**, *13* (7), 1151. DOI: 10.3390/biom13071151.
21. Kundu, S.; Roy, D. Structural Study of Biologically Significant Ligands with Major Birch Pollen Allergen Betv1 by Docking and Molecular Dynamics Simulation. *Bioinformation* **2010**, *4* (7), 326–330. DOI: 10.6026/97320630004326.
22. Schirmer, T.; Hoffimann-Sommergrube, K.; Susani, M.; Breiteneder, H.; Marković-Housley, Z. Crystal Structure of the Major Celery Allergen Api g 1: Molecular Analysis of Cross-Reactivity. *J. Mol. Biol.* **2005**, *351* (5), 1101–1109. DOI: 10.1016/j.jmb.2005.06.054.
23. Radauer, C.; Lackner, P.; Breiteneder, H. The Bet v 1 Fold: An Ancient, Versatile Scaffold for Binding of Large, Hydrophobic Ligands. *BMC Evol Biol.* **2008**, *8*, 286. DOI: 10.1186/1471-2148-8-286.
24. Gajhede, M.; Osmark, P.; Poulsen, F. M.; Ipsen, H.; Larsen, J. N.; Joost van Neerven, R. J.; Schou, C.; Løwenstein, H.; Spangfort, M. D. X-ray and NMR Structure of Bet v 1, the Origin of Birch Pollen Allergy. *Nat. Struct. Biol.* **1996**, *3* (12), 1040–1045. DOI: 10.1038/nsb1296-1040.

Author

Maxwell Shanghao Lu is a high school junior at Haidian Foreign Language Academy. He is enthusiastic about biology, environmental science, and public health. In the future, he aims to investigate the interplay among organisms, environments, and public health through an interdisciplinary lens, ultimately seeking innovative solutions for harmonious human-environmental coexistence.