2014

Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences

C. Radauer
Medical University of Vienna, Austria

A. Nandy
Allergopharma GmbH & Co. KG, Reinbek, Germany

F. Ferreira
University of Salzburg, Austria

Richard E. Goodman
University of Nebraska-Lincoln, rgoodman2@unl.edu

J. N. Larsen
ALK A/S, Hørsholm, Denmark

See next page for additional authors

Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY

Allergy

REVIEW ARTICLE

Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences

C. Radauer[1,*], A. Nandy[2,*], F. Ferreira[3,*], R. E. Goodman[4,*], J. N. Larsen[5,*], J. Lidholm[6,*], A. Pomés[7,*], M. Raulf-Heimsoth[8,*], P. Rozynek[8], W. R. Thomas[9,*] & H. Breiteneder[1,*]

[1]Department of Pathophysiology and Allergy Research, Medical University of Vienna, Vienna, Austria; [2]Research and Development, Allergopharma GmbH & Co. KG, Reinbek, Germany; [3]Christian Doppler Laboratory for Allergy Diagnosis and Therapy, University of Salzburg, Salzburg, Austria; [4]Food Allergy Research & Resource Program, University of Nebraska – Lincoln, Lincoln, NE, USA; [5]ALK A/S, Hørsholm, Denmark; [6]Thermo Fisher Scientific, Uppsala, Sweden; [7]Basic Research, Indoor Biotechnologies Inc., Charlottesville, VA, USA; [8]Institute of Prevention and Occupational Medicine of the German Social Accident Insurance, Ruhr-University Bochum (IPA), Bochum, Germany; [9]TVW Telethon Institute for Child Health Research, University of Western Australia, West Perth, WA, Australia

To cite this article: Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, Pomés A, Raulf-Heimsoth M, Rozynek P, Thomas WR, Breiteneder H. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy* 2014; **69**: 413–419.

Keywords
allergen nomenclature; allergens; bioinformatics; database; protein families.

Correspondence
Heimo Breiteneder, Department of Pathophysiology and Allergy Research, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria.
Tel.: +43 1 40400 5102
Fax: +43 1 40400 5130
E-mail: heimo.breiteneder@meduniwien.ac.at

*All of these authors belong to the WHO/IUIS Allergen Nomenclature Sub-Committee.

Accepted for publication 12 November 2013

DOI:10.1111/all.12348

Edited by: Reto Crameri

Abstract

The IUIS Allergen Nomenclature Sub-Committee, under the auspices of the World Health Organization and the International Union of Immunological Societies, maintains the systematic nomenclature of allergenic proteins and publishes a database of approved allergen names on its Web site, www.allergen.org. In this paper, we summarize updates of allergen names approved at the meetings of the committee in 2011 through 2013. These changes reflect recent progress in identification, cloning, and sequencing of allergens. The goals of this update were to increase consistency in the classification of allergens, isoallergens, and variants and in the incorporation of the evolutionary classification of proteins into allergen nomenclature, while keeping changes of established names to a minimum in the interest of continuity. Allergens for which names have been updated include respiratory allergens from birch and ragweed pollen, midge larvae, and horse dander; food allergens from peanut, cow's milk, and tomato; and cereal grain allergens. The IUIS Allergen Nomenclature Sub-Committee encourages researchers to use these updated allergen names in future publications.

The official nomenclature of allergenic proteins is based on the Linnaean binominal nomenclature identifying genus and species of all organisms and was first published in 1986 (1) and revised in 1994 (2–6). The allergen nomenclature is maintained by the IUIS Allergen Nomenclature Sub-Committee under the auspices of the World Health Organization (WHO) and the International Union of Immunological Societies (IUIS). The committee maintains the database of approved allergen names (www.allergen.org), which has developed from a plain text list to a fully functional, searchable database. In order to maintain a consistent allergen nomenclature that complies with the guidelines established by the subcommittee, researchers are required to submit newly described allergens to the Allergen Nomenclature Sub-Committee before submitting their manuscript to a journal for consideration for publication. Submissions are kept confidential by the subcommittee, and no specific information other than the name of the new allergen will be disclosed on the Web site before publication. The submission form is available at www.allergen.org.

Allergen names are composed of an abbreviation of the scientific name of its source (genus: 3–4 letters; species: 1–2 letters) and an Arabic numeral, for example Der p 1 for the first allergen to be described from the house dust mite *Dermatophagoides pteronyssinus*. Originally, new allergens were assigned consecutive numbers. During the past decades, the increase in sequence data together with advances in bioinformatics made it possible to classify allergens into protein families whose members are evolutionary related, have

similar sequences and structures, and are, in some cases, also cross-reactive (7–9). Hence, homologous allergens within a taxonomic order or family are now assigned corresponding numbers whenever possible, in order to reflect evolutionary relationships between allergens from different sources. For instance, in the rose family (Rosaceae), the numbers 1–5 are assigned to Bet v 1-related proteins (e.g., Mal d 1), thaumatin-like proteins (Mal d 2), nonspecific lipid-transfer proteins (Pru p 3), profilins (Mal d 4), and isoflavone reductases (Pyr c 5), respectively. These families of homologous allergens are frequently referred to as groups, although these designations are not part of the official allergen nomenclature. Examples are the group 1 mite allergens (e.g., Der p 1, Blo t 1, Eur m 1), which are papain-like cysteine proteases, and the group 1 grass pollen allergens (e.g., Phl p 1, Lol p1, Cyn d 1), which are β-expansins. Nevertheless, established numbers are usually not changed to avoid inconsistencies with allergen names used in previous publications or in allergen-based products used in clinical practice. Bet v 1-related allergens from legumes, for instance, are named Ara h 8 in peanut, Gly m 4 in soybean, and Vig r 1 in mung bean.

Different closely related molecular species of an allergen are named by four digits following the period after the main allergen number. The first two digits designate isoallergens, which are defined as allergens from a single species with similar molecular masses, similar biochemical functions, and sequence identities >67%. The third and fourth digits distinguish different variants of an isoallergen, which are defined as proteins with more than 90% sequence identity. Variants with different nucleotide sequences encoding identical amino acid sequences do not receive individual designations. Both the 67% and the 90% identity thresholds represent arbitrary limits and serve merely as guidelines. Appropriate allergen designations are assigned on a case-by-case basis.

Table 1 Updated nomenclature of Bet v 1 isoallergens and variants. Boldface: updated allergen designations

Previous name	New name	UniProt	Other names	Comment
Bet v 1.0101	Bet v 1.0101	P15494	Bet v 1a	
Bet v 1.0102	**Bet v 1.0101**	P15494	Bet v 1 clone 224	Identical to Bet v 1.0101
Bet v 1.0103	**Bet v 1.0101**	P15494	Bet v 1 clone 2230	Identical to Bet v 1.0101
Bet v 1.0201	Bet v 1.0201	P45431	Bet v 1b	
Bet v 1.0301	**Bet v 1.0202**	P43176	Bet v 1c	
Bet v 1.0401	**Bet v 1.0102**	P43177	Bet v 1d	
Bet v 1.0402	**Bet v 1.0102**	P43177	Bet v 1h	Identical to 'Bet v 1.0401'
Bet v 1.0501	**Bet v 1.0103**	P43178	Bet v 1e	
Bet v 1.0601	**Bet v 1.0104**	P43179	Bet v 1f	
Bet v 1.0602	**Bet v 1.0104**	P43179	Bet v 1i	Identical to 'Bet v 1.0601'
Bet v 1.0701	**Bet v 1.0105**	P43180	Bet v 1g	
Bet v 1.0801	**Bet v 1.0106**	P43183	Bet v 1j	
Bet v 1.0901	**Bet v 1.0203**	P43184	Bet v 1k	
Bet v 1.1001	**Bet v 1.0107**	P43185	Bet v 1l	
Bet v 1.1101	(Deleted)	Q39417	Bet v 1-Sc1	Pathogen-induced expression
Bet v 1.1201	(Deleted)	Q39420	Bet v 1-Sc2	Pathogen-induced expression
Bet v 1.1301	(Deleted)	Q39415	Bet v 1-Sc3	Pathogen-induced expression
Bet v 1.1401	**Bet v 1.0204**	P43186	Bet v 1m	
Bet v 1.1402	**Bet v 1.0204**	P43186	Bet v 1n	Identical to 'Bet v 1.1401'
Bet v 1.1501	(Deleted)	Q42499	BVGC11	Genomic sequence
Bet v 1.1502	(Deleted)	Q42499	BVGC31	Genomic sequence
Bet v 1.1601	(Deleted)	Q39425	BVGC181	Genomic sequence
Bet v 1.1701	(Deleted)	Q39426	BVGC21	Genomic sequence
Bet v 1.1801	(Deleted)	Q39427	BVGC25	Genomic sequence
Bet v 1.1901	(Deleted)	Q39428	BVGC34	Genomic sequence
Bet v 1.2001	(Deleted)	Q39429	BVGC45	Genomic sequence
Bet v 1.2101	(Deleted)	Q39430	BVGC63	Genomic sequence
Bet v 1.2201	(Deleted)	Q39431	BVGC681	Genomic sequence
Bet v 1.2301	(Deleted)	O23754	BVGC70	Genomic sequence
Bet v 1.2401	**Bet v 1.0108**	Q96365	Bet v 1 clone 167	
Bet v 1.2501	**Bet v 1.0109**	Q96366	Bet v 1 clone 184	
Bet v 1.2601	**Bet v 1.0110**	Q96367	Bet v 1 clone 2225	
Bet v 1.2701	**Bet v 1.0111**	Q96368	Bet v 1 clone 2226	
Bet v 1.2801	**Bet v 1.0112**	P15494 variant F63L	Bet v 1 clone 2227	
Bet v 1.2901	**Bet v 1.0113**	Q96370	Bet v 1 clone 2229	
Bet v 1.3001	**Bet v 1.0114**	Q96371	Bet v 1 clone 2301	

Researchers are encouraged to use full isoallergen and variant designations in order to unambiguously identify the allergens they work with. The importance of correct isoallergen/variant designations is highlighted by examples of highly different IgE-binding and T-cell activating properties of closely related isoallergens of Bet v 1 from birch pollen (10) and Der p 2 from house dust mite (11).

Updates of allergen designations

The 1994 revision of the allergen nomenclature represented the first introduction of bioinformatics into the allergen nomenclature, and sequence information became mandatory for the inclusion of new allergens. During the last three decades, bioinformatics has developed at an ever increasing speed and a very large amount of sequence data related to allergens has been generated. Many allergens recorded in the IUIS allergen database were originally submitted with partial sequences or even without associated sequence data. In most cases, full sequences have later become available, which in some cases has led to inconsistencies concerning the numbers assigned to allergen names. Therefore, the IUIS Allergen Nomenclature Sub-Committee aimed to screen the database for such entries and to correct them based on sequences and data from the literature.

The database was manually searched for entries with conspicuous features such as missing sequence data, biochemical names similar to those of other allergens from the same source, or inconsistent allergen numbers compared with homologous allergens from the same taxonomic family. Allergen sequences were analyzed by pairwise and multiple sequence alignments. We found several types of incorrect allergen designations in the database and changed them according to the guidelines described above. A list of updated allergen designations is shown in Tables 1–4. These changes were approved by the IUIS Allergen Nomenclature Sub-Committee at its meetings in the years 2011–2013.

Table 2 Updated nomenclature of pectate lyase allergens from short ragweed (*Ambrosia artemisiifolia*) pollen. Boldface: updated allergen designations

Previous name	New name	UniProt
Amb a 1.0101	Amb a 1.0101	P27759
Amb a 1.0201	Amb a 1.0201	P27760
Amb a 1.0202	Amb a 1.0202	E1XUL3
Amb a 1.0301	Amb a 1.0301	P27761
Amb a 1.0302	Amb a 1.0302	P27761 (variant L48Y)
Amb a 1.0303	Amb a 1.0303	P27761 (variant H392R)
Amb a 1.0304	Amb a 1.0304	E1XUL4
Amb a 1.0305	Amb a 1.0305	E1XUL5
Amb a 1.0401	Amb a 1.0401	P28744
Amb a 1.0402	Amb a 1.0402	E1XUL9
Amb a 2.0101	**Amb a 1.0501**	P27762
Amb a 2.0102	**Amb a 1.0502**	E1XUM1

Updated nomenclature of Bet v 1 isoallergens and variants

The major birch pollen allergen, Bet v 1, comprises a large number of isoallergens and variants, of which 36 were recorded in the IUIS allergen database (Table 1). A review of these entries revealed several problems: First, 13 entries referred to gene or cDNA sequences whose expression in pollen had not been demonstrated (12, 13). The sequences of Bet v 1.15–Bet v 1.23 were obtained from genomic DNA without determining their expression profiles (12). Bet v 1.11–Bet v 1.13 were identified in pathogen-infected cell cultures and leaves, while no expression in pollen was shown (13). Hence, the entries Bet v 1.11–Bet v 1.13 and Bet v 1.15–Bet v 1.23 were deleted from the database.

Within the remaining 23 entries, four sets of isoallergens had identical protein sequences but different nucleotide sequences (Bet v 1.0101/02/03, Bet v 1.0401/02, Bet v 1.0601/02, Bet v 1.1401/02). In these cases, only the first

Table 3 Updated nomenclature of *Chironomus thummi thummi* hemoglobin allergens. Boldface: updated allergen designations

Previous name	New name	UniProt	Other names
Chi t 1.0101	Chi t 1.0101	P02229	Hemoglobin component III
Chi t 1.0201	Chi t 1.0201	P02230	Hemoglobin component IV
Chi t 2.0101	Chi t 2.0101	P02221	Hemoglobin component I
Chi t 2.0102	Chi t 2.0102	P02221 (variant A113T)	Hemoglobin component IA
Chi t 3.0101	Chi t 3.0101	P02222	Hemoglobin component II-beta
Chi t 4.0101	Chi t 4.0101	P02231	Hemoglobin component IIIA
Chi t 5.0101	**Chi t 3.0201**	P02224	Hemoglobin component VI
Chi t 6.0101	**Chi t 3.0301**	P02226	Hemoglobin component VIIA
Chi t 6.0201	**Chi t 3.0401**	P02223	Hemoglobin component IX
Chi t 7*	**Chi t 3.0501**	P12548	Hemoglobin component VIIB-3
Chi t 7*	**Chi t 3.0601**	P84296	Hemoglobin component VIIB-4
Chi t 7*	**Chi t 3.0701**	P84298	Hemoglobin component VIIB-5/9
Chi t 7*	**Chi t 3.0702**	P12549	Hemoglobin component VIIB-6
Chi t 7*	**Chi t 3.0801**	P12550	Hemoglobin component VIIB-7
Chi t 8.0101	**Chi t 3.0901**	P02227	Hemoglobin component VIII
Chi t 9.0101	Chi t 9.0101	P02228	Hemoglobin component X

*The Uniprot entry P02225, previously listed in the IUIS allergen database for Chi t 7, was demerged into 7 entries, 5 from *C. thummi thummi* and 2 from *C. thummi piger*.

Table 4 Other updated allergen designations

Source	Previous name	New name	Uniprot acc. no.	Biochemical name
High sequence identity to another allergen from the same source				
Arachis hypogaea (peanut)	Ara h 4.0101	Ara h 3.0201	Q9SQH7	11S globulin; legumin; glycinin
High sequence identity to a homologous allergen from another source from the same taxonomic family				
Secale cereale (rye)	Sec c 1.0101	Sec c 38.0101	Q9S8H2	Dimeric α-amylase/trypsin inhibitor
Hordeum vulgare (rye)	Hor v 21.0101	Hor v 20.0101	P80198	γ-Hordein
Duplicate entries				
Hordeum vulgare (barley)	Hor v 1.0101	Hor v 15.0101	P16968	Monomeric α-amylase inhibitor BMAI-1
Equus caballus	Equ c 5.0101	Equ c 4.0101	P82615	Latherin
Different proteins merged into a single allergen name				
Bos domesticus (cattle)	Bos d 8	Bos d 8	–	Whole casein fraction
		Bos d 9.0101	P02662	αS1-Casein
		Bos d 10.0101	P02663	αS2-Casein
		Bos d 11.0101	P02666	β-Casein
		Bos d 12.0101	P02668	κ-Casein
Update of botanical nomenclature				
Solanum lycopersicum (previously	Lyc e 1.0101	Sola l 1.0101	Q93YG7	Profilin
Lycopersicon esculentum; tomato)	Lyc e 2.0101	Sola l 2.0101	Q547Q0	β-Fructofuranosidase
	Lyc e 2.0201	Sola l 2.0201	Q8RVW4	β-Fructofuranosidase
	Lyc e 3.0101	Sola l 3.0101	P93224	Nonspecific lipid-transfer protein
	Lyc e 4.0101	Sola l 4.0101	O49881	Pathogenesis-related protein PR-10

named isoallergen designation was kept in the database and the multiple nucleotide sequence accession numbers coding for identical amino acid sequences were associated with these records. The remaining 18 sequences were grouped into two isoallergens: Bet v 1.01 with 14 variants and Bet v 1.02 with four variants. Different variants of each isoallergen showed 91–99% sequence identity, whereas identity between Bet v 1.01 and Bet v 1.02 sequences was 84–89% (Fig. 1A).

Different allergen numbers assigned to closely related allergens from the same source

The first two identified allergens from short ragweed (*Ambrosia artemisiifolia*) pollen were originally termed antigen E and antigen K (14, 15) and later renamed to Amb a 1 and Amb a 2 with the establishment of the IUIS nomenclature system (1). Both allergens belong to the pectate lyase family and show considerable IgE cross-reactivity (16). A sequence alignment yielded 61–70% sequence identity between the ten Amb a 1 variants (corresponding to four isoallergens) and the two Amb a 2 variants (Fig. 1B). Hence, Amb a 2.0101 and Amb a 2.0102 were renamed to Amb a 1.0501 and Amb a 1.0502, as two variants of a fifth Amb a 1 isoallergen (Table 2).

Freeze-dried larvae of the nonbiting midge *Chironomus thummi thummi* used as fish feed can elicit respiratory allergic reactions to its hemoglobin, which is composed of several distinct proteins originally named CTT I to CTT IX (17) and later designated *Chi t* I (18), following the original allergen nomenclature guidelines (1). As a consequence of the heterogeneous composition of *C. thummi thummi* hemoglobin, the Allergen Nomenclature Sub-Committee assigned a separate allergen name to each hemoglobin component in 2003, creating the allergen designations Chi t 1–Chi t 9. A multiple sequence alignment of these allergens showed sequence identities ranging from 28% to 99%. Particularly, isoallergens of Chi t 5–Chi t 8 showed 51–63% identity to Chi t 3.0101 (Fig. 1C). Despite having sequence identities below the 67% threshold, their identical molecular weights and biochemical functions justified the consolidation of these isoallergens into the single allergen name Chi t 3 (Table 3). The names of Chi t 1, Chi t 2, Chi t 4, and Chi t 9 remained unchanged. Hence, the number of *C. thummi thummi* hemoglobin allergens was reduced from 9 to 5.

Ara h 3 and Ara h 4 from peanut (*Arachis hypogaea*) are major allergens belonging to the 11S seed storage globulin family. Both were submitted to the Allergen Nomenclature Sub-Committee almost simultaneously with partial sequences and different molecular masses (19, 20). Subsequent molecular cloning revealed that their full-length sequences shared 91% identity, by far exceeding the 67% identity threshold for naming isoallergens. Consequently, Ara h 4.0101 was renamed to Ara h 3.0201 (Table 4).

A

	Bet v 1.0101	Bet v 1.0114	Bet v 1.0112	Bet v 1.0109	Bet v 1.0113	Bet v 1.0102	Bet v 1.0107	Bet v 1.0105	Bet v 1.0108	Bet v 1.0110	Bet v 1.0111	Bet v 1.0103	Bet v 1.0104	Bet v 1.0106	Bet v 1.0202	Bet v 1.0203	Bet v 1.0201	Bet v 1.0204
Bet v 1.0101	100	99	99	99	99	96	94	95	98	98	97	95	96	96	88	88	89	89
Bet v 1.0114	99	100	99	99	98	95	94	94	97	97	96	94	95	96	87	88	88	89
Bet v 1.0112	99	99	100	99	98	95	94	94	97	97	98	95	95	96	88	88	89	89
Bet v 1.0109	99	99	99	100	99	96	95	96	98	97	96	96	96	97	88	88	89	89
Bet v 1.0113	99	98	98	99	100	96	95	95	97	96	96	94	95	96	88	88	89	89
Bet v 1.0102	96	95	95	96	96	100	98	98	98	97	96	92	94	93	86	86	86	88
Bet v 1.0107	94	94	94	95	95	98	100	96	97	96	95	92	93	93	84	85	86	86
Bet v 1.0105	95	94	94	96	95	98	96	100	98	96	96	91	93	93	88	88	88	88
Bet v 1.0108	98	97	97	98	97	98	97	98	100	99	98	94	94	95	86	86	87	88
Bet v 1.0110	98	97	97	97	96	97	96	96	99	100	98	93	93	94	85	86	86	87
Bet v 1.0111	97	96	98	96	96	96	95	96	98	98	100	93	93	93	85	86	86	87
Bet v 1.0103	95	94	95	96	94	92	92	91	94	93	93	100	98	99	86	87	88	88
Bet v 1.0104	96	95	95	96	95	94	93	93	94	93	93	98	100	99	86	87	87	88
Bet v 1.0106	96	96	96	97	96	93	93	93	95	94	93	99	99	100	86	87	88	88
Bet v 1.0202	88	87	88	88	88	86	84	88	86	85	85	86	86	86	100	99	99	98
Bet v 1.0203	88	88	88	88	88	86	85	88	86	86	86	87	87	87	99	100	99	99
Bet v 1.0201	89	88	89	89	89	86	86	88	87	86	86	88	87	88	99	99	100	98
Bet v 1.0204	89	89	89	89	89	88	86	88	88	87	87	88	88	88	98	99	98	100

B

	Amb a 1.0101	Amb a 1.0201	Amb a 1.0202	Amb a 1.0301	Amb a 1.0302	Amb a 1.0303	Amb a 1.0304	Amb a 1.0305	Amb a 1.0401	Amb a 1.0402	Amb a 1.0501	Amb a 1.0502
Amb a 1.0101	100	75	75	75	76	75	76	76	82	82	66	66
Amb a 1.0201	75	100	100	87	87	86	87	87	73	73	69	69
Amb a 1.0202	75	100	100	87	87	87	87	87	73	73	69	69
Amb a 1.0301	75	87	87	100	100	100	99	99	72	72	70	70
Amb a 1.0302	76	87	87	100	100	99	99	99	72	72	70	70
Amb a 1.0303	75	86	87	100	99	100	100	99	72	72	70	70
Amb a 1.0304	76	87	87	99	99	100	100	100	72	72	70	70
Amb a 1.0305	76	87	87	99	99	99	100	100	72	72	70	70
Amb a 1.0401	82	73	73	72	72	72	72	72	100	99	62	61
Amb a 1.0402	82	73	73	72	72	72	72	72	99	100	61	61
Amb a 1.0501	66	69	69	70	70	70	70	70	62	61	100	99
Amb a 1.0502	66	69	69	70	70	70	70	70	61	61	99	100

C

	Chi t 3.0301	Chi t 3.0401	Chi t 3.0101	Chi t 3.0701	Chi t 3.0702	Chi t 3.0501	Chi t 3.0601	Chi t 3.0801	Chi t 3.0901	Chi t 3.0201	Chi t 1.0101	Chi t 1.0201	Chi t 2.0101	Chi t 2.0102	Chi t 9.0101	Chi t 4.0101
Chi t 3.0301	100	68	63	62	62	63	63	61	59	54	49	49	46	46	45	33
Chi t 3.0401	68	100	63	55	55	59	61	57	58	51	46	44	45	45	46	31
Chi t 3.0101	63	63	100	55	55	56	57	57	57	51	49	44	45	45	49	29
Chi t 3.0701	62	55	55	100	99	94	88	79	52	55	42	43	43	43	48	32
Chi t 3.0702	62	55	55	99	100	94	88	79	52	56	42	43	43	43	48	32
Chi t 3.0501	63	59	56	94	94	100	91	81	55	58	42	43	42	42	48	30
Chi t 3.0601	63	61	57	88	88	91	100	81	56	57	43	43	42	42	50	32
Chi t 3.0801	61	57	57	79	79	81	81	100	59	58	45	42	46	46	49	33
Chi t 3.0901	59	58	57	52	52	55	56	59	100	54	45	46	52	52	48	32
Chi t 3.0201	54	51	51	55	56	58	57	58	54	100	51	49	47	47	46	34
Chi t 1.0101	49	46	49	42	42	42	43	45	45	51	100	85	47	47	43	33
Chi t 1.0201	49	44	44	43	43	43	43	42	46	49	85	100	45	45	38	30
Chi t 2.0101	46	45	45	43	43	42	42	46	52	47	47	45	100	99	41	28
Chi t 2.0102	46	45	45	43	43	42	42	46	52	47	47	45	99	100	41	28
Chi t 9.0101	45	46	49	48	48	48	50	49	48	46	43	38	41	41	100	33
Chi t 4.0101	33	31	29	32	32	30	32	33	32	34	33	30	28	28	33	100

Figure 1 Percentage amino acid sequence identity matrices of Bet v 1 isoallergens and variants (A), Amb a 1 isoallergens and variants (B), and *C. thummi thummi* hemoglobin allergens (C). The values were calculated based on multiple sequence alignments of the amino acid sequences. Updated allergen names according to Tables 1–4 are used. Shadings in A and B indicate >90% identity. Values >50% (light gray) and >95% (dark gray) are shaded in C.

These three examples from ragweed, nonbiting midges, and peanut illustrate that the 67% identity threshold serves only as a reference guideline, whereas the nature of the allergen (biochemical function, molecular structure, molecular weight), in addition to the degree of identity with homologous allergens, takes equal priority for the assignment of allergen names.

Inconsistent allergen numbers regarding protein family memberships

Sec c 1 from rye (*Secale cereale*), a major baker's asthma allergen, is a dimeric bifunctional inhibitor of proteases and α-amylases (21). However, within the grass family (Poaceae), the allergen number 1 is reserved for β-expansin pollen allergens. The only other dimeric α-amylase/protease inhibitor in the IUIS allergen database is Tri a 28 from wheat (*Triticum aestivum*). The sequences of the N-terminal fragment of Sec c 1 and the C-terminal fragment of Tri a 28 overlap by only 16 residues with 38% identity, whereas Tri a 28 is 100% identical to a different dimeric amylase inhibitor from rye (UniProt: C3VWW4). Hence, Sec c 1 showed no close relationship to any other cereal allergens and was renamed to Sec c 38, receiving the next available number within the Poaceae family (Table 4).

Hor v 21 (γ-hordein) from barley (*Hordeum vulgare*) is a member of the γ-prolamin subfamily, which comprises also γ-secalins from rye and γ-gliadins from wheat. The only other γ-prolamin in the allergen database is Sec c 20 with two isoallergens: Sec c 20.0101 (γ-70 secalin) and Sec c 20.0201 (γ-35 secalin). As the allergen number 21 is already reserved for α-/β-prolamins such as Tri a 21, for the α-/β-gliadin from wheat, Hor v 21 was renamed to Hor v 20 (Table 4).

Duplicate database entries

Previously, the monomeric α-amylase/protease inhibitor BMAI-1 from barley had been assigned the designations Hor v 1 and Hor v 15. As BMAI-1 is homologous to Tri a 15 from wheat (44% sequence identity), the designation Hor v 15 was maintained and Hor v 1 was deleted (Table 4).

The horse (*Equus caballus*) dander allergens Equ c 4 and Equ c 5 were originally submitted with partial sequence data and molecular masses of 18.7 and 16.7 kDa (22). After the full sequence of horse latherin became available, the partial sequences of both Equ c 4 and Equ c 5 matched this protein. The original description of Equ c 4 and Equ c 5 most likely referred to the glycosylated and nonglycosylated forms of latherin. Thus, the entry Equ c 5 was deleted (Table 4).

Updated nomenclature of casein components

The allergen Bos d 8, casein from cow's (*Bos domesticus*) milk, refers to a mixture of several dissimilar proteins. The casein fraction of milk proteins contains components that belong to two unrelated protein families, one family comprising αS1-, αS2-, and β-caseins, while κ-caseins constitute the other family (23). Even within the α-/β-casein family, sequence identities are below 15%. Hence, the entry Bos d 8 was demerged into four separate allergens: Bos d 9.0101 (αS1-casein), Bos d 10.0101 (αS2-casein), Bos d 11.0101 (β-casein), and Bos d 12.0101 (κ-casein; Table 4). The name Bos d 8, which is widely established and has been used in numerous publications and names of commercial diagnostic tests, was kept and designates the whole casein fraction. This example illustrates that the subcommittee takes into consideration not only taxonomic and other scientific aspects but also practical aspects of continuity and public acceptance in its management of the allergen nomenclature.

Adjustment of allergen names to reflect updated taxonomy

In addition to the changes described above, based on sequence similarities and protein family memberships, the IUIS Allergen Nomenclature Sub-Committee changed the designations of tomato allergens from Lyc e 1–Lyc e 4 to Sola l 1–Sola l 4 in order to reflect the establishment of *Solanum lycopersicum* instead of *Lycopersicon esculentum* as the official scientific name of the tomato (24) (Table 4).

Concluding remarks

The WHO/IUIS Allergen Nomenclature Database is continuously updated and supplemented not only with newly submitted allergens, but also with data of already published allergens. The IUIS Allergen Nomenclature Sub-Committee encourages users to notify the committee of missing or inconsistent records in the database and thereby aid in providing a reliable and up-to-date resource of unambiguous allergen names and isoallergen/isoform sequence information for the scientific community. Updates and error reports may be sent to the subcommittee either by using the submission form to be downloaded from www.allergen.org or by directly contacting one of the committee members, whose contact details are published at www.allergen.org.

Acknowledgments

The authors thank all members of the WHO/IUIS Allergen Nomenclature Sub-Committee for fruitful discussions and John Wise (Food Allergy Research & Resource Program, University of Nebraska – Lincoln, Lincoln, NE, USA) for designing and maintaining the WHO/IUIS Allergen Nomenclature Database.

Author contributions

C. Radauer drafted the manuscript and performed the greater part of the sequence analyses. A. Nandy contributed to sequence analysis of Bet v 1 and Amb a 1 isoallergens. M. Raulf-Heimsoth and P. Rozynek performed sequence analysis of *Chironomus* allergens. H. Breiteneder, F. Ferreira, R. Goodman, J. N. Larsen, J. Lidholm, A. Pomés, and W. R. Thomas contributed to discussions of several groups of allergens. All authors critically read and approved the manuscript.

Conflicts of interest

All authors have no conflicts of interest to declare.

References

1. Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TA. Allergen nomenclature. *Bull World Health Organ* 1986;**64**: 767–774.
2. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen nomenclature. *Int Arch Allergy Immunol* 1994;**105**:224–233.
3. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen nomenclature. *Bull World Health Organ* 1994;**72**:797–806.
4. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen nomenclature. *Clin Exp Allergy* 1995;**25**: 27–37.
5. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen nomenclature. *Allergy* 1995;**50**:765–774.
6. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W. Allergen Nomenclature. *J Allergy Clin Immunol* 1995;**96**:5–14.
7. Chapman MD, Pomes A, Breiteneder H, Ferreira F. Nomenclature and structural biology of allergens. *J Allergy Clin Immunol* 2007;**119**:414–420.
8. Ferreira F, Hawranek T, Gruber P, Wopfner N, Mari A. Allergic cross-reactivity: from gene to the clinic. *Allergy* 2004;**59**:243–267.
9. Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol* 2008;**121**:847–852.
10. Ferreira F, Hirtenlehner K, Jilek A, Godnik-Cvar J, Breiteneder H, Grimm R et al. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: potential use of hypoallergenic isoforms for immunotherapy. *J Exp Med* 1996;**183**:599–609.
11. Hales BJ, Hazell LA, Smith W, Thomas WR. Genetic variation of Der p 2 allergens: effects on T cell responses and immunoglobulin E binding. *Clin Exp Allergy* 2002;**32**:1461–1467.
12. Hoffmann-Sommergruber K, Vanek-Krebitz M, Radauer C, Wen J, Ferreira F, Scheiner O et al. Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions. *Gene* 1997;**197**: 91–100.
13. Swoboda I, Scheiner O, Heberle-Bors E, Vicente O. cDNA cloning and characterization of 3 genes in the Bet v 1 gene family that encode pathogenesis-related proteins. *Plant, Cell Environ* 1995;**18**:865–874.
14. King TP, Norman PS, Lichtenstein LM. Isolation and characterization of allergens from ragweed pollen. IV. *Biochemistry* 1967;**6**:1992–2000.
15. Adolphson C, Goodfriend L, Gleich GJ. Reactivity of ragweed allergens with IgE antibodies. Analyses by leukocyte histamine release and the radioallergosorbent test and determination of cross-reactivity. *J Allergy Clin Immunol* 1978;**62**:197–210.
16. Wopfner N, Gadermaier G, Egger M, Asero R, Ebner C, Jahn-Schmid B et al. The spectrum of allergens in ragweed and mugwort pollen. *Int Arch Allergy Immunol* 2005;**138**:337–346.
17. Baur X, Dewair M, Fruhmann G, Aschauer H, Pfletschinger J, Braunitzer G. Hypersensitivity to chironomids (non-biting midges): localization of the antigenic determinants within certain polypeptide sequences of hemoglobins (erythrocruorins) of Chironomus thummi thummi (Diptera). *J Allergy Clin Immunol* 1982;**69**:66–76.
18. Mazur G, Baur X, Liebers V. Hypersensitivity to hemoglobins of the Diptera family Chironomidae: structural and functional studies of their immunogenic/allergenic sites. *Monogr Allergy* 1990;**28**:121–137.
19. Kleber-Janke T, Crameri R, Appenzeller U, Schlaak M, Becker WM. Selective cloning of peanut allergens, including profilin and 2S albumins, by phage display technology. *Int Arch Allergy Immunol* 1999; **119**:265–274.
20. Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW et al. Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *J Clin Invest* 1999;**103**:535–542.
21. Garcia-Casado G, Armentia A, Sanchez-Monge R, Sanchez LM, Lopez-Otin C, Salcedo G. A major baker's asthma allergen from rye flour is considerably more active than its barley counterpart. *FEBS Lett* 1995;**364**:36–40.
22. Goubran Botros H, Poncet P, Rabillon J, Fontaine T, Laval JM, David B. Biochemical characterization and surfactant properties of horse allergens. *Eur J Biochem* 2001;**268**:3126–3136.
23. Ginger MR, Grigor MR. Comparative aspects of milk caseins. *Comp Biochem Physiol B Biochem Mol Biol* 1999;**124**:133–145.
24. Spooner DM, Anderson GJ, Jansen RK. Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). *Am J Botany* 1993;**80**:676–688.