

Chapter 15

Allergen Database

Yuyi Huang and Ailin Tao

Abstract Over the past ten years, there has been steady growth in the identification of allergen sequences. Different research institutions have built allergen databases for the integration and screening of allergens, thus serving scientific research and medical applications. However, as fewer and fewer new allergens are being identified, it is time for the establishment of standards with the goal of perfecting allergen databases so that they can serve not only as a allergen query platform that allows more effective use of allergen information, but also be an excellent database, i.e., be comprehensive, non-redundant, stringent, accurate, and able to analyze allergen functions.

Keywords Allergen · Bioinformatics · Database · Data mining · Protein · Health

15.1 Introduction

With the advances in the technology and due to the economic effects, allergens have drawn more and more interest because allergic diseases greatly affect human health (Sicherer and Leung 2014, 2015). Allergy is a rapid onset, multisystem hypersensitivity reaction that may be caused by both immunological and nonimmunological mechanisms. Most allergic reactions are immunoglobulin (IgE) mediated (Kirkbright and Brown 2012), show clear genetic predispositions and individual differences, but generally will neither destroy histocytes nor cause tissue damage (Monitto et al. 2010; de Wang 2011). A substance that can induce an allergic

Y. Huang · A. Tao (✉)

Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology,
The State Key Clinical Specialty in Allergy, The State Key Laboratory of Respiratory Disease,
The Second Affiliated Hospital of Guangzhou Medical University, 250# Changgang Road
East, Guangzhou 510260, Guangdong Province, People's Republic of China
e-mail: taoailin@gzhmu.edu.cn

Y. Huang
e-mail: chvipdata@126.com

reaction is called allergen and is an essential component of allergy (Soares-Weiser et al. 2014). According to the particular pathway that causes the allergic reaction, there are five types of allergens: (1) food allergens, like milk, egg, fish, beef, mutton, seafood, animal fat, foreign proteins, alcohol, sesame oil, green onion, ginger, garlic, and some vegetables, fruits, etc. (Meng juan et al. 2014); (2) inhalation allergens, such as pollen, catkins, dust, mites, animal dander, lampblack, paint, exhaust, coal gas, cigarettes, cold air, fog, etc. (Diamant et al. 2013); (3) contact allergens, like cold, heat, ultraviolet light, radiation, cosmetics, shampoo, detergent, hair dye, soap, chemical fiber products, plastic, metal jewelry, bacteria, mildew, viruses, parasites, etc. (Goossens 2015); (4) drug allergens, like drugs, antimicrobial and anti-inflammatory drugs, penicillin, streptomycin, xenogeneic serum, etc. (5) self-allergens, i.e., autoantigens, such as mental stress, working stress, microbial infections, ionizing radiation, burns, and other biological, physical, and chemical factors including trauma or infection can cause changes in the structure or composition of autoantigens that results in the formation of allergens (Hopp et al. 2014).

Identifying and collecting all known allergens and creating an allergen database will provide a convenient query platform for medical researchers, doctors, and the general public that will provide complete information about known allergens. At present, several original and relatively complete international allergen databases have been created.

The existing allergen databases include: (1) the WHO/IUIS database (Radauer et al. 2014) (World Health Organization and International Union of Immunological Societies) that was created by the World Health Organization and International Union of Immunological Societies in 1984. It focuses on the systematic nomenclature of allergens according to their characteristics, structure, function, molecular biologic features, and bioinformatics data. The WHO/IUIS database only includes officially certified allergens. Newly submitted allergens are investigated by an Executive Committee, and the process of adding new allergens to this database requires extensive evidence as demonstrated in published papers and their validity must be further verified. Only when an allergen meets all requirements will it be formally named and included in the allergen database; (2) the Allergome database (Mari and Scala 2006), an online website, was created to provide relevant information about all aspects of an allergen. It includes allergens that cause IgE-dependent allergic diseases such as asthma, allergic dermatitis, allergic conjunctivitis, allergic rhinitis, and urticaria. Moreover, because allergy researchers desire that all allergenic molecules be supported by explicit documented evidence, every allergen in the Allergome database describes the allergen source (e.g., animal or plant, dandruff, fruit, pollen, seeds, spores, venom, or the human body) and its route of exposure (e.g., contact, ingestion, inhalation, or injection). The relevant information about each allergen is attached in a specific page. In addition, all allergens in the WHO/IUIS database have been incorporated into the Allergome database, including their relevant information and systematic nomenclature. The screening protocol of the search function in the Allergome database is the same as the protocol used by the WHO/IUIS database, thus, all data in the Allergome database also come from officially certified allergens. In

addition to all the allergens included in the WHO/IUIS database, the Allergome database also contains a number of allergens that have had their immunological mechanism, clinical data, and other documented evidence thoroughly investigated. The Allergome database devotes itself to serving scientists working in allergy and immunology research. Its web interface and comprehensive data also provide a very convenient and efficient way for the clinician to obtain knowledge of allergens; (3) the AllergenOnline database is a platform that provides a list of manually examined allergens that are searched using certified serial numbers. This database focuses on assessing the safety of a protein that has undergone gene mutation or other food processing methods. Proteins have been assessed by additional experiments such as the serum fixation test, stimulation of basophils for histamine release, and cross-reactivity *in vivo*. This database is updated annually and provides a simple but efficient tool for safety assessment; (4) the ALLERGENIA database aims to promote the sharing of knowledge. All resources in this database are open to all researchers as well as to the public without reservation. Just a click on the “Browse Database” button displays all contents in the database, including accession number of an allergen and information about whether the allergen has been included in the UniProt protein database, the Allergome database or IUIS international allergens database, related information regarding IgE reactivity, and any published support. In addition, with just a click users can link to the interface of the data source, where more detailed content can be acquired. In order to provide the most availability, nonredundancy, astringency, accuracy, and functional analysis for researchers and the public, every sequence in the ALLERGENIA database has undergone strict double blind verification. The ALLERGENIA database is a tool for scientific research that promises accuracy and efficiency. It makes it possible for researchers to focus on the problem itself instead of wasting time searching for the relevant data. Moreover, ALLERGENIA, a public knowledge database, provides a feasible way for our institute, which is based in Guangzhou, and is dedicated to popularizing experimental science, to guide the public away from pseudoscience and lead them to reliable knowledge. Other influential allergen databases are named in Table 15.1.

15.2 Overview of Databases

Never perfect, every database has its own merits and drawbacks. The WHO/IUIS database is an internationally recognized database and includes officially certified allergens. There is no doubt about its authority. Therefore, many scientific researchers choose it as their certified allergen standard. Furthermore, nearly all other allergen databases are founded on this database. For example, the search function in the Allergome database is identical with that in the WHO/IUIS database. But since the WHO/IUIS database is very careful in investigating new allergens and/or due to inadequate funding for the necessary experiments, the process of adding new allergens is very slow. At present, the creation of genetically

Table 15.1 Allergen databases

Name	Website	# Allergens	Organization
Defra	http://allergen.fera.defra.gov.uk	640	Food and Envir. Res Agency
AllergenOnline	http://www.allergenonline.com	1706	University of Nebraska
SDAP (Ivanciuc et al. 2003)	http://fermi.utmb.edu/SDAP/index.html	1526	Univ. of Texas Medical Branch
WHO/IUIS (Radauer et al. 2014)	http://www.allergen.org	771	WHO/IUIS Allergen Nomenclature Sub-committee
Allallergy	http://allallergy.net	6329	Food and Preservative Allergy and Intolerance Database
Allergome (Mari and Scala 2006)	http://www.allergome.org	2915	Allergy Data Laboratories (ADL)
ADFS (Nakamura et al. 2005)	http://allergen.nihs.go.jp/ADFS	1777	NIHS, Japan
AllerMatch (Fiers et al. 2004)	http://www.allermatch.org	1080	Wageningen University
UniProtKB/Swiss-Prot (Boutet et al. 2007)	http://www.uniprot.org	435	EMBL-EBI, PIR
ALLERGENIA	http://ALLERGENIA.gzhmu.edu.cn	2150	GdPKL of Allergy and Immunology

modified products is growing explosively and suspect allergens caused by severe environmental pollution have increased significantly. Therefore, new protein information is very limited in this database.

Allergome, a database that is based on the allergens in the WHO/IUIS allergen database, takes advantage of the searching capabilities of the Internet and can automatically find biological molecules that are IgE sensitive or induce IgE-mediated diseases. Thanks to new computer-based programs, this database is regularly updated and its contents are plentiful. However, results obtained from computer programs inevitably contain some errors. Allergen databases, due to their importance to personal health and safety, should be more rigorous than other types of databases. Therefore, the judgment rule that “no omissions are allowed even at the expense of false judgment” is too inclusive.

The AllergenOnline database is a balance of the above two databases, that is to say, it includes all known allergens and ensures their careful assessment. Information about an allergen is regularly collected from the NCBI protein database and then investigated by a panel. Most importantly, a unified reviewing procedure is in place that all experts abide to when they independently investigate any particular protein. Finally, the panel votes to decide whether an allergen should be included. In this way, the accuracy as well as the growth ability of a database can be ensured. However, the database depends too strongly on preliminary artificial screening, and various defects, such as invalid sequences, repetitive sequences,

and error sequences, can exist in different databases. A more obvious problem is that, huge amounts of sequences with high similarity point to the same allergen, providing several difficulties for the application of data.

In conclusion, an excellent database should be: (1) **comprehensive**, meaning it should include all known sequences of all allergens so that it can function as and meet the requirements of a database. Only when the database contains sequences from comprehensive resources will the users choose it as a regular tool; (2) **non-redundant**, meaning it does not include a large number of redundant sequences i.e., the same sequences with different names, the same sequences with different accession numbers, or integrated sequences and their fragments coexisting in the database. Using redundant sequences to expand the database is unwise. It can mislead users to possibly repeat positive or negative results. Importantly, it can weaken the accuracy and credibility of the software program that is utilized by some users to search the allergen database; (3) **stringent**, meaning that the allergen database should be a useful tool to solve practical problems and not just be an archive of allergens. Similarity of some sequences can reach 97 % and even as high as 99 %. These sequences have limited usefulness for practical applications. Time that is needed to acquire valid information from the database should be reduced. Repeat sequences should be avoided. In other words, the database should provide a practical tool for users to fully solve problems; (4) **accurate**, meaning of all its data should be investigated repeatedly and confirmed by experiment history. This is the most important standard and is at the core of a reliable database. If a database cannot promise its accuracy, any other information will be useless. This is particularly crucial for a medical database, because information regarding the sequences it contains will directly affect personal safety and medical decisions. The right of security is the most significant right for our citizens. Therefore, in order to embody the social responsibility required of a medical database, accuracy of the database is paramount; and (5) **analyzing performance**, meaning the operability and practicability of the database conveniently allows for the acquisition of comprehensive information by the user. Without this, a database will just look like a gold mine. The users, as miners, must spend most of their time excavating data. And this will be even more difficult for the general public who may not have basic computer knowledge. A functional analysis that is too complicated results in a database that looks like a super-excavator ahead of its time, which is hard to manipulate. In other words, the database itself should do the data mining and offer the gold directly to the users. This will reduce the costs in time and money and maximize the benefits.

15.2.1 WHO/IUIS

The WHO/IUIS allergy database (Radauer et al. 2014), an internationally recognized database, contains the officially confirmed allergens. Many institutions use this database as the gold standard. Because of its authority, almost all other allergen databases are based on it. But there are some irregular entries and invalid data in this database, which impacts its usefulness. By 2014, this database contained

a total of 794 allergen indexes, 133 of which only had titles or were invalid. Of 1028 sequences, 941 sequences are marked with an UNIPROT accession number (906 actual effective sequences) and 56 sequences are labeled with NCBI accession number rather than UNIPROT accession number (13 actual effective sequences). After removing the repetitive sequences from the 919 actual effective ones, 911 sequences remain.

15.2.2 Allergome

The Allergome database (Mari and Scala 2006) is based on the WHO/IUIS database. Using published literature and the computer, this database automatically finds biological molecules that react with IgE or induce IgE-mediated disease. Because a computer program automatically searches new literature, this database is updated regularly, thus, its content is rich but it also has more invalid sequences. This database, updated at 10:15 +1 GMT on October 9, 2014, contains a total of 3885 protein allergens sequences and 4561 displayable login entries. As a large number of single entries each correspond to multiple sequences, 6306 sequences can be downloaded through these entries. After removing duplicate sequences, there are 5916 remaining sequences. When 99 % of the similar sequences are removed, 4630 sequences remain and when 97 % of similar sequences are removed, 3738 sequences remain.

15.2.3 AllergenOnline

The AllergenOnline database (Goodman et al. 2014) is updated annually and has a unified approval process that ensures its validity and expandability. According to version 14 updated on January 20, 2014, a total of 1706 allergens are indexed and this includes 1666 nonredundant sequences. After removing 99 % of similar sequences, there are 1404 remaining sequences and after removing 97 % of similar sequences, there are 1112 remaining sequences.

15.3 Introduction to the ALLERGENIA Database

15.3.1 Background

An increasing number of protein molecules have been identified as either allergens or nonallergens during the past 20 years (Sicherer and Leung 2014, 2015). Currently, there is no consensus among different Web-based databases because they have individually gathered different numbers of so-called “authentic”

allergens and, in some cases, have even incorrectly included some nonallergens as “allergens.” Collecting, organizing, and displaying veritable data that is reported in the scientific literature and having them interactively serve to draw unanimous conclusions is becoming a major concern of many investigators who rely on this knowledge to evaluate the allergenicity of their starting materials.

The allergen database ALLERGENIA has obvious advantages with accessibility, nonredundancy, astringency, accuracy, and functional analysis.

15.3.2 General Information

The ALLERGENIA database was developed by the Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, the State Key Laboratory of Respiratory Disease, the Second Affiliated Hospital of Guangzhou Medical University. At this time, FARRP (allergenonline), IUIS, and Allergome, etc., are the most widely used allergen website databases. But these websites have faults and/or limitations. Our ALLERGENIA database substantially outperforms these comprehensive allergen databases and by far exceeds the more narrow databases such as the InformAll Allergenic Food Database. The following four important aspects are featured in the ALLERGENIA database: (1) the highest number veritable allergens are included with the authenticity of each allergen sequence individually verified; (2) no false allergens (nonallergens) are included; (3) only nonredundant allergenic sequences are listed; and (4) the interactive interface is user friendly for a wide range of applications. After extensive and thorough investigation, we gathered 2108 authentic allergenic protein sequences and eliminated any redundant sequences.

15.3.3 BLAST Search

Gene modification as a technology has its advantages and disadvantages. On one hand, it can amplify genes of excellent agronomic traits in a crop to help resolve the conflict between resource shortages, environmental deterioration, and population growth; on the other hand, it may bring novel genes into crops that may also be allergens, thus affecting allergic people. Therefore, it is necessary to evaluate the allergenicity of candidate genes before gene modification and employ a “brave exploration but cautious promotion” approach for the breeding of gene-modified crops. At present, all genetically modified crops that have passed through safety assessment are considered to be safe (Wal 2015). All newly expressed proteins in recombinant-DNA plants that could be present in the final food should be assessed for their potential to cause allergic reactions. Genes derived from known allergenic sources should be assumed to encode an allergen unless scientific evidence demonstrates otherwise. Several bacteria, such as *Staphylococcus aureus*, *Aspergillus*

niger, *Aspergillus fumigatus*, etc., are important allergens (FAO/WHO 2003, 2008), but instead are being neglected by some experts. All gene transformations should comply with the Codex Principle in order to verify the safety of the transgenic food using a programmed process of assessments before commercialization (FAO/WHO 2003). Production and utilization should be suspended immediately if any allergic cases related to a GMO food are found. Transgenic foods should never be rushed to market or be imported from abroad without definite information in regard to allergenic potential to the people of the (importing) countries.

According to the Codex Alimentarius Commission Guidelines (FAO/WHO 2008), if a gene fragment comes from some known allergenic source, it should be considered as an allergen. That is to say, we should assume that this type of gene fragment is harmful before conducting experiments to exclude this assumption. When testing allergenicity for new GMOs, we need to compare the results of allergenicity assessment on a case-by-case basis. If the coding amino acid sequence of one candidate transgene is highly identical to that of a mite allergen, while the amino acid sequence of another candidate gene only resembles that of a peanut allergen, the former has more risk potential than the latter. Furthermore, if the amino acid sequence of a candidate gene only has large identity with the allergen from rice, it will not pose any safety issues to the people feeding on rice.

Therefore, we developed an allergenicity evaluation standardization system, which is divided into three levels. The first standard is called Rice Level, i.e., the acceptable lowest limit or least hazardous level. This standard determines that a new GMO is safe only when the allergenicity of its transformed gene(s) and products is similar to or lower than that of rice, even when the transformed genes and their encoding amino acid sequences are similar to the most potent allergens derived from rice amino acid sequences. The second standard is called Peanut/Soybean Level, i.e., a hazard level that has warning exclusion criteria. This standard deems that a GMO product may be hazardous to the humans when the allergenicity of the newly introduced protein encoded by the transformed gene is similar to the most potent allergens derived from peanut/soybean, especially if the sequence(s) is highly similar to the most potent allergens from peanut/soybean on contiguous and identical fragment(s). This standard states that the transforming candidate genes should have allergenicity assessment and modification. The third standard is called Mite/Olive level, i.e., the highest allergenicity level and candidate for mandatory exclusion. This standard determines that these GMO products would be hazardous to humans, and that the encoding gene should not be transformed when the allergenicity of the encoded protein by the future-transformed gene is similar to the most potent allergens derived from mite/olive (pollen), especially if the sequence(s) is highly similar to the most potent allergens from mite/olive (pollen) on contiguous and identical fragment(s). Here we propose a three-level standardization system for allergenicity evaluation as follows (Fig. 15.1).

Based on the above considerations, the ALLERGENIA database provides various search and comparison methods so as to offer customers comprehensive

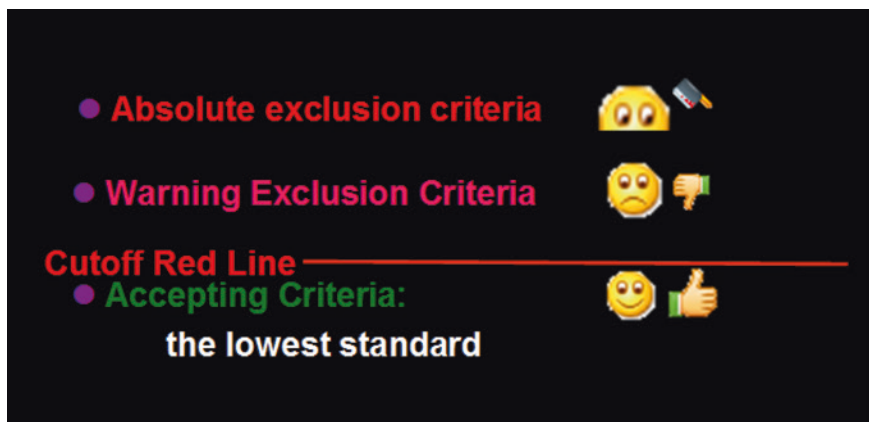


Fig. 15.1 A system for the standardization of allergen evaluation divided into three levels

results. Customers can draw their own conclusions using their own standards if desired or carry out subsequent verification work. This database provides two ways of BLAST, i.e., routine BLAST and BLAST based on FAO/WHO rules. The routine BLAST, this method does not include short identical fragments in the results because its goal is to check whether there sequences that are identical or similar among the entries listed in the queried database. Figure 15.2 shows an example of a routine BLAST. The BLAST method based on FAO/WHO rules searches for identity with contiguous amino acids or a minimum of 35 % sequence identity over a window of 80 amino acids with an allergen as defined by FAO/WHO rules. Through this special search, results obtained from the ALLERGENIA website display different identity levels in five columns: Column 1, a minimum of 35 % sequence identities over a window of 80 amino acids of each query protein with an allergen; Column 2, identities less than 35 % over sliding windows of 80 amino acid segments of each query protein, but with at least one segment identity of eight contiguous amino acids; Column 3, identities less than 35 % for sliding windows of 80 amino acid segments of each query protein, but with a sequence of only seven exact match amino acids; Column 4, identities less than 35 % for sliding windows of 80 amino acid segments of each query protein, but only with six exactly matching amino acids; Column 5, identities less than 35% for sliding windows of 80 amino acid segments of each query protein, but with no more than five exactly matching amino acids. This design was chosen based on a recent study demonstrating that shared identity with only five contiguous amino acids can also lead to cross-reactivity between nonhomologous allergens. Figure 15.3 shows an example of the BLAST comparison method using FAO/WHO rules.

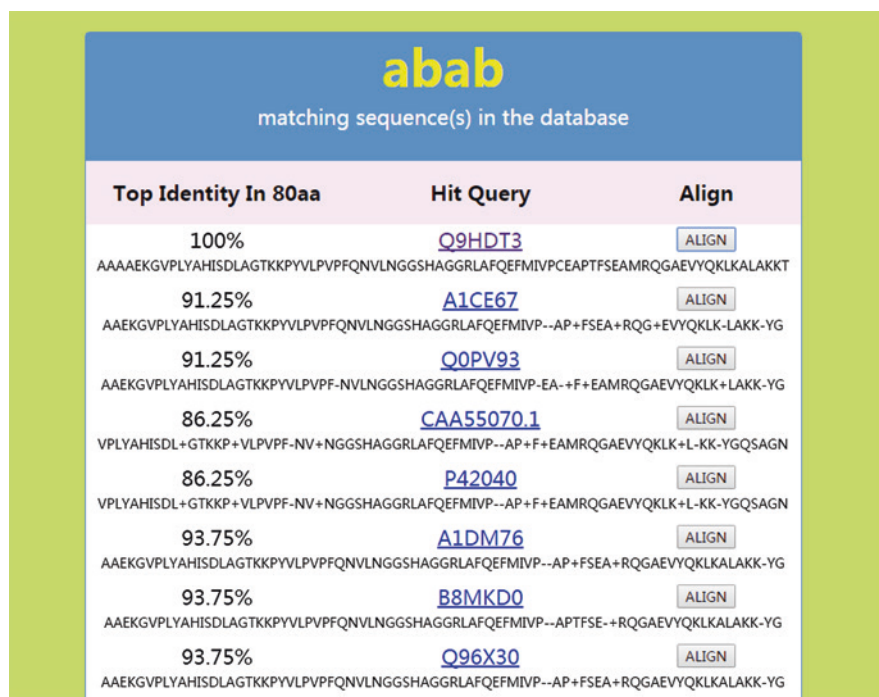


Fig. 15.2 Partial results of a sequence “abab” demonstrating a routine BLAST for

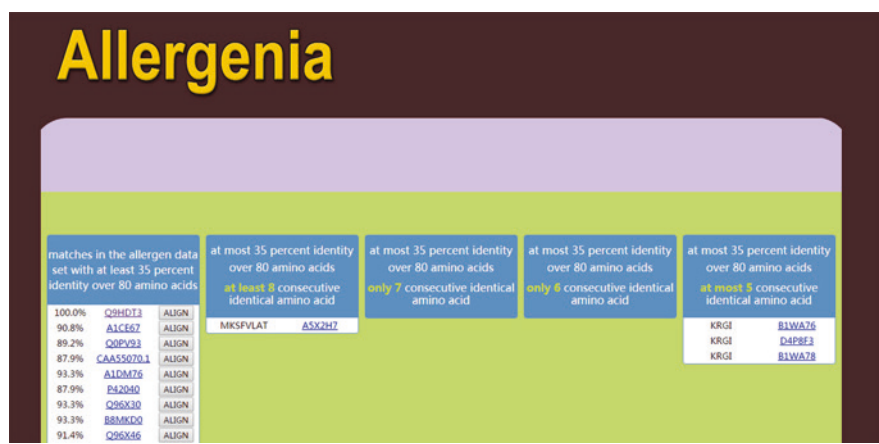


Fig. 15.3 An example a sequence “abab” demonstrating the BLAST comparison method using FAO/WHO rules and the ALLERGENIA database

15.3.4 Browse Database

When browsing the ALLERGENIA database, the page displayed shows all the allergen data gathered in the database for any particular allergen. The “ID” column shows the protein accession number in UniProt or in NCBI format. The allergen facts column includes five items: “UniProt,” “Allergome,” “IUIS,” “IgE,” and “Reference.” For each item, “y” indicates that the objective protein(s) is/are collected as allergens in certain databases, such as UniProt, Allergome, IUIS, or that the protein possesses IgE-binding ability or that other references definitely regard the protein as an allergen, whereas “n” indicates that the corresponding database does not identify this protein as an allergen, or that the protein lacks proof of IgE-binding or other supporting references. It is important that all allergens display complete information in each of the five categories. Faults and/or limitations relating to the existing websites can be clearly viewed, thus allowing researchers to rely on the most authentic allergen database.

15.3.5 Condition Search

Condition searches in the ALLERGENIA database include five options: “UniProt,” “Allergome,” “IUIS,” “IgE,” or “Reference,” which all contain allergen facts. In each module, when the “Existent” condition is selected, the allergens for which there are known facts are retrieved and displayed, whereas when the “Non-Existent” condition is selected, allergens that lack individual facts are retrieved and displayed. These searches allow the allergen facts to be viewed in a three-dimensional way and also show the differences among different databases for each individual allergen.

References

- Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-prot. *Methods Mol Biol.* 2007;406:89–112.
- de Wang Y. Genetic predisposition for atopy and allergic rhinitis in the Singapore Chinese population. *Asia Pac Allergy.* 2011;1(3):152–6.
- Diamant Z, Gauvreau GM, Cockcroft DW, Boulet LP, Sterk PJ, de Jongh FH, Dahlen B. Inhaled allergen bronchoprovocation tests. *J Allergy Clin Immunol.* 2013;132(5):1045–55.
- FAO/WHO. Guideline for the conduct of food safety assessment of foods produced using recombinant DNA microorganisms (CAC/GL 46-2003). Foods derived from modern biotechnology (Second ed.). 2003. p. 35–57.
- FAO/WHO. Guideline for the conduct of food safety assessment of foods derived from recombinant DNA animals (CAC/GL 68-2008). Foods derived from modern biotechnology (Second ed.). 2008. p. 57–76.
- Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP, van Ham RC. Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO codex alimentarius guidelines. *BMC Bioinform.* 2004;5:133.

- Goodman R, van Ree R, Vieths S, Ferreira F, Ebisawa M, Sampson HA, et al. Criteria used to categorise proteins as allergens for inclusion in allergenonline.org: a curated database for risk assessment. *Clin Transl Allergy*. 2014;4(Suppl 2): P12.
- Goossens A. New cosmetic contact allergens. *Cosmetics*. 2015;2(1):22–32.
- Hopp AK, Rupp A, Lukacs-Kornek V. Self-antigen presentation by dendritic cells in autoimmunity. *Front Immunol*. 2014;5:55.
- Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res*. 2003;31(1):359–62.
- Kirkbright SJ, Brown SG. Anaphylaxis–recognition and management. *Aust Fam Physician*. 2012;41(6):366–70.
- Mari A, Scala E. Allergome: a unifying platform. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M*. 2006;95:29–39.
- Meng juan W, Jiang hua L, Lin yu G, Jia jie L, Zhi gang W. Food allergen labeling management: a review of the European experience and its inspiration for China. *Food Sci*. 2014;35(1):261–5.
- Monitto CL, Hamilton RG, Levey E, Jedlicka AE, Dziedzic A, Gearhart JP, Boyadjiev SA, Brown RH. Genetic predisposition to natural rubber latex allergy differs between health care workers and high-risk patients. *Anesth Analg*. 2010;110(5):1310–7.
- Nakamura R, Teshima R, Takagi K, Sawada J. Development of Allergen Database for Food Safety (ADFS): an integrated database to search allergens and predict allergenicity. *Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku*. 2005;(123):32–6.
- Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, Pomes A, Raulf-Heimsoth M, Rozynek P, Thomas WR, et al. Update of the WHO/IUIS allergen nomenclature database based on analysis of allergen sequences. *Allergy*. 2014;69(4):413–9.
- Sicherer SH, Leung DY. Advances in allergic skin disease, anaphylaxis, and hypersensitivity reactions to foods, drugs, and insects in 2013. *J Allergy Clin Immunol*. 2014;133(2):324–34.
- Sicherer SH, Leung DY. Advances in allergic skin disease, anaphylaxis, and hypersensitivity reactions to foods, drugs, and insects in 2014. *J Allergy Clin Immunol*. 2015;135(2):357–67.
- Soares-Weiser K, Takwoingi Y, Panesar SS, Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Halken S, Poulsen L, van Ree R, et al. The diagnosis of food allergy: a systematic review and meta-analysis. *Allergy*. 2014;69(1):76–86.
- Wal JM. 8-Assessing and managing allergenicity of genetically modified (GM) foods. *Handbook of Food Allergen Detection and Control*. 2015: 161–78.

Author's Biography



Yuyi Huang is an Assistant Researcher working at the Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, the State Key Laboratory of Respiratory Disease, and the State Key Clinical Specialty in Allergy of the Second Affiliated Hospital of Guangzhou Medical University. Email: chvipdata@126.com.

Dr. Huang received his Master's Degree in Immunology and Molecular Biology from Guangzhou Medical University in 2010 and is currently working at Guangzhou Medical University. His experience is in alleviation of the allergenic potential of allergens with a main focus on allergen bioinformatics. He has developed an allergen database ALLERGENIA and an allergen predicting software SORTALLER.



Ailin Tao is a Professor at Guangzhou Medical University, Director of Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, Principal Investigator of the State Key Laboratory of Respiratory Disease, Deputy Director of the State Key Clinical Specialty in Allergy of the Second Affiliated Hospital of Guangzhou Medical University, Member of the State Committee for Transgenic Safety Assessment, Standing Committee Member of Allergy Branch of Guangdong Medical Association, Member of Guangdong Provincial Committee for Transgenic Safety Assessment and Master Tutor.

Professor Ailin TAO earned his doctorate degree from the State Key Laboratory of Crop Genetic Improvement of Huazhong Agricultural University in 2002, followed by a post-doctoral training at Postdoctoral Station of Basic Medicine in Shantou University Medical College, majoring in allergen proteins. His most recent research has been on allergy bioinformatics, allergy and clinical immunology and disease models such as allergic asthma, allergic rhinitis, infection and inflammation induced by allergy, inflammatory and protracted diseases caused by antigens or superantigens. He has gained experience in the field of allergology including the mechanisms of immune tolerance, allergy triggering factors, and chronic inflammation pathways and allergenicity evaluation and modification for food and drugs. He proposed some new concepts including “Representative Major Allergens,” “Allergenicity Attenuation” of immunotoxin and allergens, “broad-spectrum immunomodulator” as well as the theoretical hypothesis of “Balanced Stimulation by Whole Antigens.” Professor TAO’s laboratory focuses on the diagnosis of allergic disease and the medical evaluation of food and drug allergenicity and its modification. Professor TAO has now constructed a system for the prediction, quantitative assessment, and simultaneous modification of epitope allergenicity, which has been applied to more than 20 allergens, and he also developed a bioinformatics software program for allergen epitope prediction, SORTALLER (<http://sortaller.gzhmu.edu.cn>), which performed significantly better than other existing software, reaching a perfect balance of high specificity (98.4 %) and sensitivity (98.6 %) for discriminating allergenic proteins from several independent datasets of protein sequences of diverse sources. Furthermore, this program has a Matthews correlation coefficient as high as 0.970, a fast running speed and can rapidly predict a set of amino acid sequences with a single click. The software has been frequently used by researchers from many institutions in China and over 30 countries worldwide, thus becoming the number one allergen epitope prediction software program. Professor TAO has set up an allergen database ALLERGENIA (<http://ALLERGENIA.gzhmu.edu.cn>) that has several advantages over other databases such as a wide selection of nonredundant allergens, excellent stringency and accuracy, and friendly and usable analytical functions.