Allergen Databases for Food Safety of GMOs and Novel Foods

food allergy research & resource program
farrp

Allergen Databases for Food Safety of GMOs and Novel Foods

Author: Rick Goodman[1]

Affiliation: [1]Food Allergy Research and Resource Program,
Department of Food Science and Technology,
University of Nebraska, Lincoln, NE USA 68588-6207
rgoodman2@unl.edu

Food Safety – Food Allergens

Food allergy has been recognized as an important food safety issue only in the past 50 to 60 years in the United States. Primary risks are for the sub-population of people sensitized and allergic to specific food sources. Food allergy prevalence reportedly increased significantly over 30 years while self-described allergy has grown at a much faster rate. By 1996 allergists, food companies and the US Food and Drug Administration were developing rules and procedures to protect those with specific food allergies. The number of food allergic consumers with severe allergy is still <2% in 2021 as is the prevalence of celiac disease. Peanut, a few species of tree nuts, cow's milk, chicken eggs, crustacean shellfish, finned fish, wheat and soybeans are defined as the big eight allergen sources for the US. In 2021 FDA expanded the list to include sesame seeds for required labeling of packaged foods. Yet only a few proteins in each allergenic source are recognized as the major targets that triggered most reactions. Canada listed mustard as a major allergen. The European Union lists 13 allergenic protein sources with slightly different definitions: cereals containing glutens rather than "wheat", celery (celeriac) mustard and sesame seeds, lupin and molluscs.

There are significant differences in the few studies indicating sources causing allergy in different global populations. Only a few comparative studies in children or adults are widely vetted (Shoemaker et al., 2015; Lyons et al., 2019; Grimshaw et al., 2020; Li et al., 2020). Differences include relatively rare cases in India and China likely due to wider tolerances in some populations for undefined reasons. Social media has led consumers to high rates of self-misdiagnosis of food allergy rather than food intolerances. Intestinal microbiomes and the age of introduction of foods may account for differences in prevalence.

New Food Proteins, Databases and methods

New dietary proteins can be introduced through genetic engineering or introduction of new food sources in diets. Many foods are now altered by industrial processing and combinations of ingredients making food choices complex. Yet the focus in public awareness for thirty years has centered on genetically modified organisms (GMO) with one or a few new proteins in commonly consumed foods. Regulators and developers are doing relevant safety assessments to minimize risks for those already allergic, but many scientists suggest methods beyond simple tests and consumers want validation of safety (Goodman et al., 2008). GMOs offer benefits. Plant diseases can be combatted by transferring a gene from a source, commonly consumed pepper, into bananas to stop bacterial wilt (Jin et al, 2017). New food sources can include algae, specific fungus or another whole organism into a processed food. Yet, considering potential risks of allergy from these requires understanding the source, and the use of genomics, proteomics with bioinformatics (Abdelmotelelb et al., 2021).

> **"There are significant differences in the few studies indicating sources causing allergy in different global populations."**

WHO/IUIS Allergen Nomenclature

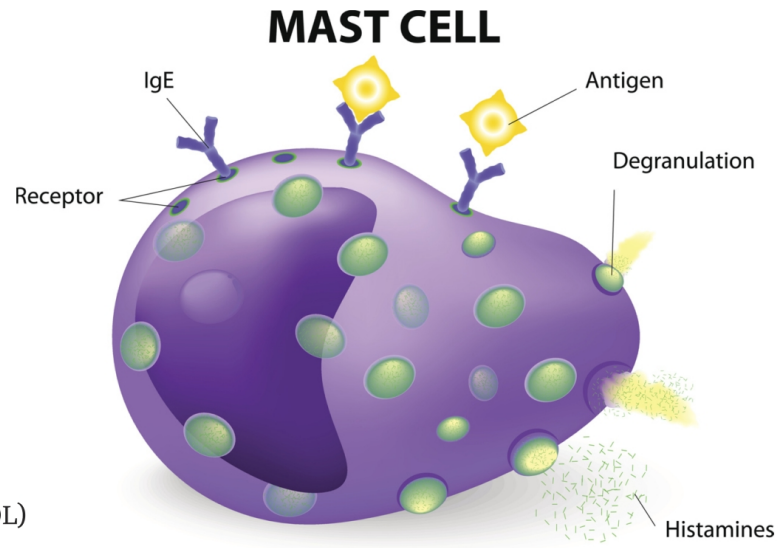
The database was created by David Marsh and a group of allergists and scientists to create a nomenclature system naming proteins that are apparent allergens based on description of the proteins causing immediate allergic responses in the airways, skin or gastrointestinal tract. The allergen sources known then included pollens from grasses and short-ragweed, house dust mite proteins and a few food allergens of codfish and milk (Marsh et al., 1986). The group organized within the International Union of Immunological Societies (IUIS), defined criteria name allergenic proteins from all sources. Names used the first three letters of the genus and the first letter of the species as well as a number representing the order of identification of the allergen within the taxonomic entity. Their 1986 publication listed 56 allergens. The Nomenclature Sub-Committee continued sporadically into the mid-1990's and more consistently since (www.allergen.org). Heimo Breiteneder was Chair of the Committee from 2007-2014. I was elected in 2014 and continue as Chair in 2021. The Sub-Committee is sponsored by three allergy and immunology societies (AAAAI, EAACI and IUIS). The purpose is to provide a standardized naming system for association with proteins identified as binding IgE and possibly causing allergy before first publications. The entries are not necessarily proven allergens. Researchers provide preliminary data of IgE binding using serum donors with described histories of allergy. Biological activity of skin prick tests or basophil activity are described for some providing better evidence. Submissions should include a minimum of five human IgE binding subjects, sometimes fewer. Proteins in this database should be considered presumptive allergens.

> **"Each year newly identified protein sequences are evaluated by the data presented in peer reviewed publications with reviewer comments recorded and maintained in the background."**

Agricultural Biotechnology company databases, pre-2005

As companies developed GMOs, regulators in the United States and elsewhere required an evaluation of new biotech protein in comparison to those of known allergens. History of allergies to the original source of transferred genes are considered and comparison of the amino acid sequence of the new protein to known allergens. Bioinformatic comparisons were done by BLASTP or FASTA of the protein sequences to allergen lists selected from the NCBI Protein database using keyword selections with "allergen" and "allergy". Monsanto Company described searching for homologues by FASTA and then looking for exact matches of eight amino acids to determine possible risks (Astwood and Fuchs, 1996; Metcalfe et al., 1996). The FASTA algorithm was written by Pearson and Lipman (1988). If matches were found or sources were allergenic, serum IgE binding studies should be done using samples from appropriately allergic donors.

AllergenOnline.org 2005 through 2021



Six biotechnology companies agreed to fund the development of a common, public allergen database in the Food Allergy Research and Resource Program (FARRP) at the University of Nebraska-Lincoln. The sponsors recognized the value of an independent group with scientific experts and of allergy. The database is available at www.allergenonline.org (AOL) and the website describes the expert reviewers, criteria and quality of data and a list of publications for each allergen sorted by organism and protein type. The amino acid sequences of a few isoforms are included for common allergens. The history of allergen numbers, species and protein types are described on the History page. The GI numbers and Accession numbers are listed. Access and use of the database are free. Search types include: 1) overall FASTA searches, 2) sliding 80 amino acid FASTA for matches equal to or greater than 35% identity over 80, adjusted to 80AA for shorter matches and 3) Exact 8 amino acid segment matches. The review panel includes experts from professional organizations (AAAAI, EAACI and WAO). Development of the database and criteria are published (Goodman et al., 2016). Each year newly identified protein sequences are evaluated by the data presented in peer reviewed publications with reviewer comments recorded and maintained in the background. Since 2019 the list of allergens shows whether there is merely published IgE binding or additional bioactivity to support inclusion of the protein as an allergen. Some of the 2233 proteins are not important allergens while others cause systemic anaphylaxis or other severe reactions in some allergic subjects. Airway, contact, venom and food allergens are included. In 2012 a Celiac peptide and protein database was added and updated in 2018, with description of criteria and recommended methods.

Allergome

Adriano Mari of Italy introduced the Allergome database (www.allergome.org) in 2003. It has real time information from a variety of sources including publications and protein data. It has been funded by a variety of sources and is cross listed in UniProt. There are interesting tools for searches and information, but it requires patience to review references to understand relevance for allergy. Some of the proteins listed under the “Allergens” tab are unproven as a search for myoglobin lists 9 items out of 7535 in the database, yet a search of scientific literature in PubMed shows only one publication demonstrating possible IgE binding from one allergic subject to bovine myoglobin (Fuentes et al. 2004). That paper did not provide clear evidence of protein identity or allergy for beef allergic subjects.

AllerCatPro

The A*STAR Bioinformatics Institute in Singapore is part of the Agency for Science, Technology and Research public sector R&D agency. AllerCatPro is a bioinformatics tool that has combined allergen sequences from five databases (AllergenOnline listed as FARRP; COMPARE; the WHO/IUIS database, Allergome and the UniProt-KW database) to provide their protein pool. The database provides tools for predicting possible cross-reactivity using sequence, structure and 3D predictions. A search of their database using the bovine myoglobin sequence resulted in a high identity match to Equine myoglobin with 93.8% identity, and 100% predicted 3D epitope concluding with their statement of Strong Evidence of allergenicity. That match is likely due to UniProtKW and Allergome information stems from a publication by the group of Rudolf Valente who used equine myoglobin as a non-allergen backbone to test artificial grass pollen epitope spacing to evaluate activation of basophils (Gieras et al., 2016).

COMPARE

The COMPARE database went online in 2017. It used the 2016 Allergenonline.org (AOL) version as a foundation plus a few entries selected using strategies used by AOL. However, the AOL version 17 had more entries than COMPARE. Until 2019 the two database lists were similar but COMPARE began adding individual short-peptides from LC-MSMS studies that could be used to flag full-length protein sequences. They also use a strict sliding 80mer window without showing matches < 80 AA. Alternatively, AOL calculates a 72 AA segment of peanut allergen Ara h 2 as 90% identity over 80 amino acids. The COMPARE database follows the CODEX Alimentarius statement exactly, but that would not protect peanut allergic people if that 72 AA protein was put into rice since it would likely cause severe allergic reactions as it contains two or three recognized IgE binding epitopes. While the COMPARE and AllergenOnline databases are similar, these differences can be important in ensuring food safety.

AlgPred2.0

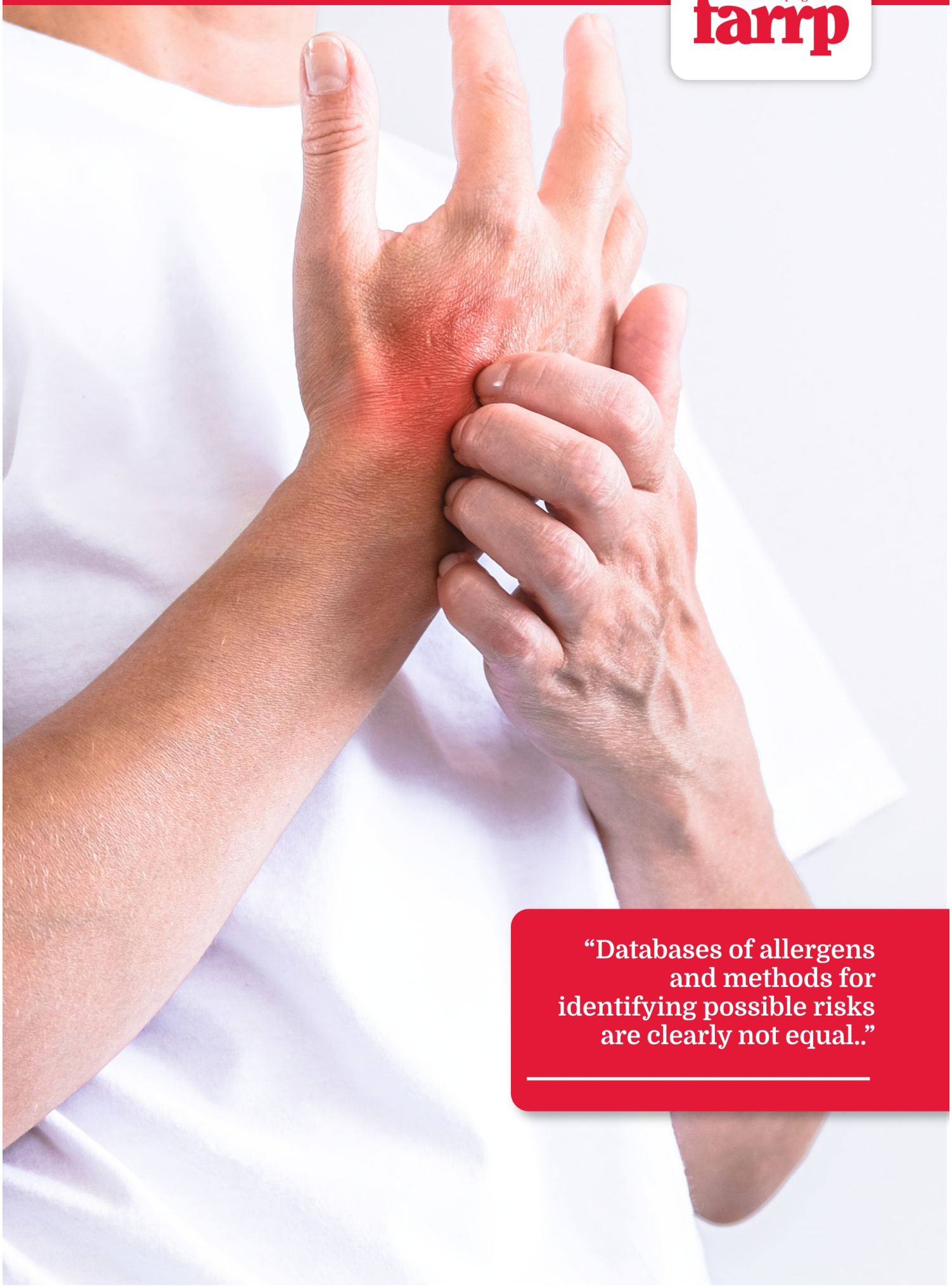
The AlgPred database was published in 2006 and updated in 2021 (Sharma et al., 2021). As described in 2021, they combined 2018 allergens from COMPARE and 2078 from AllergenOnline and allergens and non-allergens from another database, AllerTOP (Dimitrov et al. 2014). AlgPred indicates having used 15,046 IgE epitopes from the IEDB database, which seems in error (below). A search of AlgPred 2 with bovine myoglobin predicts that myoglobin is an allergen by either their AAC based RF or their Hybrid methods. The epitope prediction identified a claimed six amino acid segment (LEKFDK) of bovine myoglobin as an IgE epitope. The Motif scanning with MERCI was negative. Their BLAST results indicate myoglobin is an allergen. Results are conflicting and explanation of IgE epitopes as the IEDB database has fewer IgE epitopes than the claimed 15,046.

IEDB

The NIH-funded Immune Epitope Database (IEDB) was started in 2004 at the La Jolla Institute for Immunology (LJI) by Dr Alessandro Sette. It is freely available online, and catalogs epitopes for humans, non-human primates, rodents, and other vertebrates

food allergy research & resource program

farrp



"Databases of allergens and methods for identifying possible risks are clearly not equal.."

targeted in the context of allergy, infectious disease, autoimmunity, and transplant rejection. The IEDB also provides tools for predicting B cell and T cell epitopes and binding of Major-Histocopatibility receptors. Focusing searches for published B cell epitopes in the literature (IgG, IgE, IgA) recognized in the context of Allergy, the IEDB lists 5239 epitopes, 284 antigens (proteins) and 502 references.

CONCLUSIONS Databases of allergens and methods for identifying possible risks are clearly not equal. Some scientists have called for combining data from all allergen databases as done in AllerCatPro, AlgPred2, AllerTOP and Allergome, which can be inaccurate. Some databases do not claim to have published proof of allergy including the WHO/IUIS Allergen Nomenclature database which names putative allergens. Methods used for sequence matching new proteins to allergens are not equal. Short peptide matches alone are not useful. Overall FASTA or BLAST are often quite useful for identity matches >50% being relatively predictive except for evolutionarily conserved proteins. Bovine myoglobin is 84% identical to human myoglobin, shrimp tropomyosin is 53% identical to human tropomyosin and the human proteins are unlikely to be allergens. Cyclophilins and profilins are also highly conserved (Abdelmoteleb et al., 2021). The use of exact criteria for 80 or more AA is not fully protective for proteins with high identities as shown by Ara h 2. Therefore, ignoring high identity matches >30 AA could be risky for consumers. Focusing on identity matches just above 35% restricts the food industry severely as serum tests are complex and appropriate donors hard to find. Admitting my bias, I suggest www.AllergenOnline.org is the best tool for minimizing risks to both. Users should understand the strengths and limitations of the databases used for these evaluations.

References

Abdelmoteleb M, Zhang C, Furey B, Kozubal M, Griffiths H, Champeaud M, Goodman RE. (2021). Evaluating potential risks of food allergy of novel food sources based on comparison of proteins predicted from genomes and compared to www.AllergenOnline.org. Food Chem Toxicol 147: doi.org/10.1016/j.fct.2020.111888

Astwood JD, Fuchs RL. (1996). Allergenicity of foods derived from transgenic plants. Monographs in Allergy 32:105-120.

CODEX Alimentarius Guideline, Foods derived from modern Biotechnology 2nd Edition. WHO & FAO of the United Nations. CAC/GL 44-2003 and CAC/GL 45-2003 including Annex 1 through Annex 4.

Dimitrov I, Bangov I, Flower DR. (2014). AllerTOP v2—a server for in silico prediction of allergens. J Mol Model 20:22278

EFSA (European Food Safety Authority) (2009). Consolidated presentation of the joint Scientific Opinion of the GMO and BIOHAZ Panels on the “Use of Antibiotic Resistance Genes as Marker Genes in Genetically Modified Plants” EFSA Journal, 1108: 1–8.

Fuentes MM, Palacios R, Garces M, Baballero ML, Moneo I. (2004). Isolation and characterization of a heat-resistant beef allergen: myoglobin. Allergy 59(3):327-331.

Gieras A, Linhart B, Roux KH, Dutta M, Khodoun M, et al, (2016). IgE epitope proximity determines immune complex shape and effector cell activation capacity. J Allergy Clin Immunol 137(5):1557-1565.

Goodman RE, Vieths S, Sampson HA, Hill D, et al. (2008). Allergenicity assessment of genetically modified crops - what makes sense? Nat Biotech 26(1):73-81.

Goodman RE, Ebisawa M, Ferreira F, Sampson HA, van Ree R Vieths S, Baumert JL, Bohle B, Lalithambika S, Wise J, Taylor SL. (2016). AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. Mol Nutr Food Re 60(5):1183-1198.

Grimshaw KEC, Roberts G, Selby A, Reich A, Butiene I, et al., (2020). Risk factors for Hen's egg allergy in Europe: EuroPrevall birth cohort. J Allergy Clin Immunol Pract 8:1341-1348;

Jin Y, Goodman RE, Tetteh AO, Lu M, Tripathi L. (2017). Bioinformatics analysis to assess potential risks of allergenicity and toxicity of HRAP and PFLP proteins in genetically modified bananas resistant to Xanthomonas wilt disease. Food Chem Toxicol 109(Pt 1): 81-89.

Li, J, Ogorodova LM, Mahesh PA, Wang MH, Fedorova OS, Leung TF, et al., (2020). Comparative study of food allergies in children from China, India and Russia, The EuroPrevall-INCO surveys. J Allergy Clin Immunol Pract 8(4):1349-1358.

Lyons SA, Burney PGJ, Ballmer-Weber GK, Fernandez-Rivas M, Barreales L, et al., (2019). Food allergy in adults: substantial variation in prevalence and causative foods across Europe. J Allergy Clin Immunol Pract. 7(6):1920-1928.

Marsh DG, Goodfriend L, King TP, Lowenstein H and Platts-Mills TAE. (1986). Allergen Nomenclature. Bulletin of the World Health Organization 64(5):767-770.

Maurer-Stroh S, Krutz NL, Kern PS G;unalan V, Nguyen MN, Limviphuvadh V, Eisenhaber F, Gergerick GF. (2019). AlleratPro-Prediction of protein allergenicity potential from the protein sequence. Bioinformatics doi: 10.1093/bioinformatics/btz029.

Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL. (1996). Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit Rev Food Sci Nutr 36(S):S165-S186.

Sharma N, Patiyal S, Dhall A, Pande A, Arora C, Raghava GPS. (2021). AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. Brief Bioinformatics 22(4): Doi: 10.1093/bib'bbaa 294.

Shoemaker AA, Sprikkelman AB, Grimshaw KE, Roberts G, Grabenhenrich L, et al., (2015). Incidence and natural history of challenge-proven cow's milk allergy in European children – EuroPrevall birth cohort. Allergy 70:963-972.

FARRP has two primary missions:

Mission 1

Develop and provide the food industry with credible information, expert opinions, tools, and services relating to allergenic foods.

Mission 2

Develop and provide the food and related industries with credible information, expert opinions, tools, and services relating to novel foods and food ingredients including genetically modified products.

FARRP takes a comprehensive approach working with and collaborating with research institutions, governmental authorities, consumer groups, and scientific societies around the globe to share our experience and knowledge to improve the safety of food products for consumers with food allergies and sensitivities.

food allergy research & resource program
farrp