Predictive Analysis of Allergens in Novel Foods by Using Advanced Proteomic and Bioinformatic Tools

Juan de Dios Alché[1,2], María López-Pedrouso[3], José M. Lorenzo[4], Ramón Moreira[5] and Daniel Franco[5,*]

[1] *Plant Reproductive Biology and Advanced Microscopy Laboratory (PReBAIL). Estación Experimental del Zaidín (EEZ), CSIC, Granada. Spain*

[2] *University Institute of Research on Olive and Olive Oils (INUO), Jaén. Spain*

[3] *Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela, Santiago de Compostela, 15872 A Coruña, Spain*

[4] *Centro Tecnolóxico da Carne de Galicia, Rúa Galicia Nº 4, Parque Tecnológico de Galicia, San Cibrao das Viñas, 32900 Ourense, Spain.*

[5] *Department of Chemical Engineering, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain.*

** author for correspondence: daniel.franco.ruiz@usc.es*

BACKGROUND & OBJECTIVES

One of the most significant challenges in bringing novel foods to market is ensuring food safety. New food scenarios drive the need to detect novel allergens that must be identified and quantified for proper labeling. Allergenic reactions are primarily caused by proteins that are abundant in foods, typically of low molecular mass, glycosylated, water-soluble, and highly stable to proteolysis. The most relevant plant and animal food allergens, such as lipid transfer proteins, profilins, seed storage proteins, lactoglobulins, caseins, tropomyosins, and parvalbumins from fruits, vegetables, nuts, milk, eggs, shellfish, and fish, have been studied. As a massive preliminary screening, proteomic methods should be employed to search for potential allergens. This work focuses on proteomic and bioinformatic tools for food researchers to identify allergens in novel foods.

ALLERGENS IN TRADITIONAL FOOD

Table 1. Several protein allergens in food. For the WHO/IUIS nomenclature, the allergens are named according to the species source of food.

Food	Protein Name	Species	Allergen
Milk	Caseins	*Bos taurus*	
	α S1-casein (23.6 kDa)		Bos d 9
	α S2-casein (25.2 kDa)		Bos d 10
	β -casein (24 kDa)		Bos d 11
	κ-casein (19 kDa)		Bos d 12
	β-lactoglobulin (18.3 kDa) α-lactalbumin (14.2 kDa)		Bos d 5
	Serum albumin (66.3 kDa)		Bos d 4
	Immunoglobulin (160 kDa)		Bos d 6
			Bos d 7
Eggs	Ovomucoid (28 kDa)	*Gallus domesticus*	Gal d 1
	Ovalbumin (44 kDa)		Gal d 2
	Ovotransferrin (78 kDa)		Gal d 3
	Lysozyme (14 kDa)		Gal d 3
	α-livetin (69 kDa)		Gal d 5
	YGP42 (35 kDa)		Gal d 6
Fish	Parvalbumin	*Gadius callarias* (Baltic cod)	Gad p 2
	α-parvalbumin (13 kDa)		Gad p 1
	β-parvalbumin (11.6 kDa)		
Shellfish	Tropomyosin (34 kDa)	*Metapenaeus ensis* (Shrimp)	Met e 1
Peanuts/tree nuts	7 S seed storage globulin, vicilins (64 kDa)	*Arachis hypogaea*	Ara h 1
	2 S albumin (17 kDa)		Ara h 2, Ara h 6, Ara h 7
	Nonspecific lipid transfer proteins		Ara h 9, Ara h 16, Ara h 17
	Oleosins		Ara h 10, Ara h 11, Ara h 14, Ara h 15
			Ara h 12, Ara h 13
	Defensins		Ara h 5
	Profilins		Ara h 8
	Plant pathogenesis-related proteins PR-10		
Soy	7 S seed storage globulin, β-conglycinin	*Glycine max*	Gly m 5
	11 S seed storage globulin, glycinin		Gly m 6
Wheat	α-amylase inhibitor (13 kDa)	*Triticum aestivum*	Tri a 28
	Gamma gliadin (88 kDa)		Tri a 20
	Elongation factor 1		Tri a 45
Sesame	2 S albumins	*Sesamum indicum*	Ses i 1, Ses i 2
	7 S vicilin-type globulin (45 kDa)		Ses i 3.
	Oleosins		Ses i 4, Ses i 5
	11 S globulin, legumins		Ses i 6, Ses i 7
	Profilin		Ses i 8

ALLERGENS IN NOVEL FOODS

Figure 1. Novel protein sources of plant, algal, fungal and insect origins are being researched by the food industry.

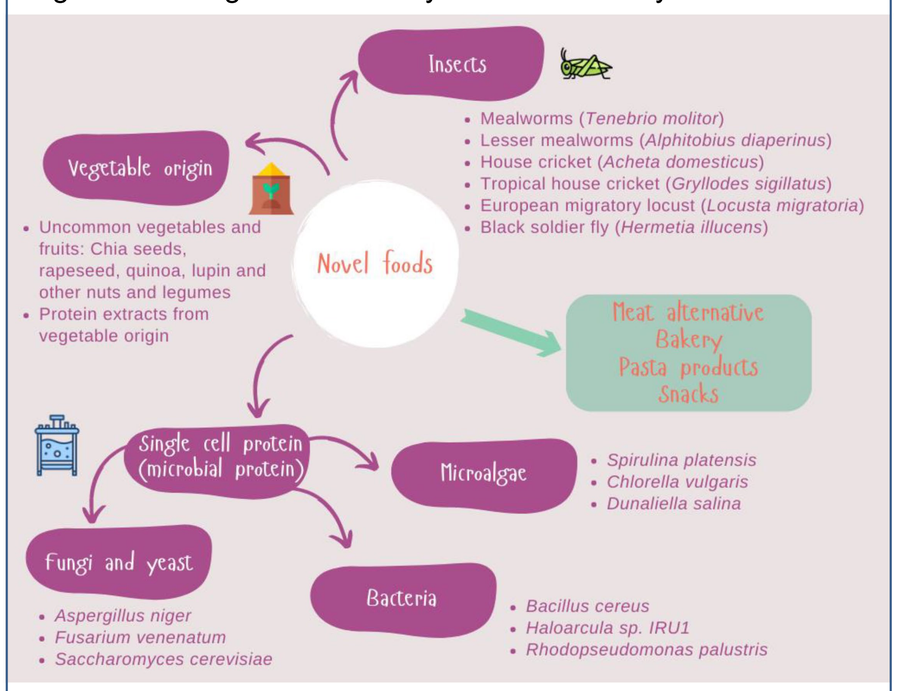


Table 2. Presence of allergens in novel foods based on microalgae and insects.

Food	Protein Name	Species
Microalgae	C-phycocyanin Thioredoxins Superoxide dismutase Glyceraldehyde-3-phosphate dehydrogenase Triosephosphate isomerase	*Microalgae spirulina (A. platensis)*
Microalgae	viz. calmodulin Fructose-bisphosphate aldolase	*Microalgae chlorella (C. vulgaris)*
Insects	Tropomyosin, myosin, actin, troponin C (muscle proteins) Tubulin (cellular proteins) Hemocyanin, defensin (circulating proteins) Arginine kinase, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), triosephosphate isomerase, α-amylase, trypsin, phospholipase A, hyaluronidase (enzymes)	

PROTEOMIC APPROACHES TO IDENTIFY ALLERGENS IN NOVEL FOODS

Table 3. Relevant examples of the recent use of proteomic approaches for the detection of allergens in novel foods.

	Novel Food	Bioinformatic Tool	Goal/Main Achievements
VEGETABLES	Bread wheat spelt and rye	Database of Allergen Families-AllFam AllergenOnline Allergome	Comparison of allergenicity in cereal products
	Cashews	BLASTP Search against AllergenOnline sequence	Analysis of allergen stability under heat treatment
	Goji berries	AlgPred software hybrid approach	Identification of 11 IgE-binding proteins
	Macadamia nut	AllergenOnline Immune Epitope Database Analysis Resource (IEDB)	Analysis of homology and linear epitope similarities to known allergens
	Medicago sativa	COMPARE allergen database	Identification of three allergenic protein families
	Lentil (Lens culinaris)	Blast2GO—Functional Annotation and Genomics	Quantification of major allergen proteins
	White- and red-fleshed pitaya seeds	AllermatchTM webtool Algpred 2.0 AllerCatPro web server	Identification of five potential allergens
SEA WEEDS	Spirulina and chlorella microalgae	AllergenOnline	Six proteins exhibit significant homology with food allergens
INSECTS	Cricket	Allermatch TM webtool AlgPred 2.0 ABCPred Bepipred	Description of the impact of processing on allergenic reactivity of insect proteins.
	Cricket Acheta domesticus	Database of Allergen Families-AllFam Allergen nomenclatura (WHO/IUIS) CLC Genomics Workbench 20.0.4. AllerCatPro web server	Identification of 20 putative allergens
	Lesser mealworms, black soldier flies and their protein hydrolysate	AllermatchTM webtool	Identification of potential allergens by similarity to known allergens
PARASITES	*Anisakis simplex*, *Pseudoterranova decipiens*, and *Contracaecum osculatum*	Blast2GO—Functional Annotation and Genomics AllergenOnline AllerTOP web server ver. 2.0 PREAL web server	Prediction of 53 probable allergens in three species

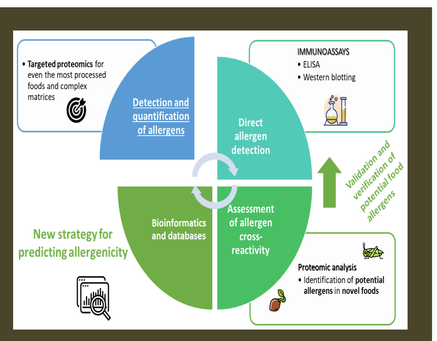
BIOINFORMATIC TOOLS

Table 2. Bioinformatic software tools most used for allergen analysis.

Name	Link (Website)	Description
Allergen nomenclature	http://www.allergen.org (accessed on 12 February 2023)	Official site for the systematic allergen nomenclature provided by the World Health Organization and International Union of Immunological Societies (WHO/IUIS)
AllerBase	http://bioinfo.unipune.ac.in/AllerBase/Home.html (accessed on 12 February 2023)	Database of allergens detected as IgE-binding epitopes, IgE antibodies and cross reactivity. Allergen data such as experimental information on its allergenic activity and food source is compiled, resulting in a curated database.
AllerCatPro	https://allercatpro.bii.a-star.edu.sg/ (accessed on 12 February 2023)	Provides protein allergenicity potential prediction based on the similarity of amino acid sequence and 3D protein structure
AllergenOnline	http://www.allergenonline.org (accessed on 12 February 2023)	Provides sequence database of allergens to identify proteins and assess the potential risk of allergenic cross-reactivity. This database offers 2233 peer-reviewed sequences from 912 taxonomic protein groups (February 2021)
Allergome	http://www.allergome.org (accessed on 12 February 2023)	A website with detailed information on Allergenic Molecules (Allergens) causing an IgE-mediated (allergic, atopic) disease (anaphylaxis, asthma, atopic dermatitis, conjunctivitis, rhinitis, urticaria).
Comprehensive protein allergen resource (COMPARE allergen database)	https://comparedatabase.org/ (accessed on 12 February 2023	A database comprised of protein sequences of known allergens
Database of Allergen Families-AllFam	http://www.meduniwien.ac.at/allfam/ (accessed on 12 February 2023)	Comprises a resource for classifying allergens into protein families as well as biochemical properties and allergology significance
Immune Epitope Database and analysis resource (IEDB)	https://www.iedb.org (accessed on 12 February 2023)	Provides experimental data on antibody and T-cell epitopes to identify allergens and to assist in the prediction and analysis of allergenicity
Structural Database of Allergenic Proteins (SDAP)	https://fermi.utmb.edu (accessed on 12 February 2023)	Tool for testing the FAO/WHO allergenicity rules in new proteins and investigating cross reactivity, also offering information about protein sequence and structure

TAKE-HOME MESSAGE

- Proteomic approaches using advanced MS will continue providing relevant information in food safety.
- Detection, identification and quantification of known allergens in complex matrices and highly processed food have already been developed, and targeted MS allows monitoring of them during food processing.
- Identification of novel protein allergens in insects, seaweeds, microalgae or other non-common vegetable foods is one of the most important challenges over the next few years.
- Bioinformatic tools and curated allergen databases will enable the prediction of potential allergens, wich should be validated subsequently.
- This information could be used to improve the design and safety of food products by novel devices.
- Advanced technologies, including biosensors, could identify specific interactions between receptors and allergens, enabling us to address the challenges of food safety monitoring.



López-Pedrouso, M.; Lorenzo, J.M.; Alché, J.d.D.; Moreira, R.; Franco, D. Advanced Proteomic and Bioinformatic Tools for Predictive Analysis of Allergens in Novel Foods. **Biology 2023, 12, 714.** https://doi.org/10.3390/Biology12050714