External Scientific Report

APPROVED: 30 May 2024
doi: 10.2903/sp.efsa.2024.EN-8840

Novel strategies for predicting allergenicity: development of a ranking method and screening tools to assess the allergy risk of innovative proteins

OC/EFSA/GMO/2021/04

E.N. Clare Mills[1], Federica Orsenigo[1], Diana Salgado[1],

Paul M. Finglas[2], Siân Astley[2]

[1] School of Biosciences, Faculty of Medicine and Health, The University of Surrey, Guildford, GU2 7XH

[2] EuroFIR AISBL, 40 rue Washington, 1050, Brussels, Belgium

Abstract

To protect individuals who already have or are at risk of developing immune-mediated adverse reactions to food, novel foods and genetically modified organisms (GMOs) undergo an allergenicity risk assessment. There are shortcomings in this process that could be improved through use of well-defined clinically relevant allergen molecules with different allergenic potential. The objective of this project was to develop novel strategies for predicting allergenicity of innovative/novel proteins that address this issue. We undertook a systematic review of allergen molecules in foods listed on Annex II of the Food Information for Consumers Regulation together with additional foods known to cause IgE-mediated food allergies in at least one European region with a prevalence of 0.5%. Around 750 in-scope papers were quality assessed to allow clinical relevance of allergen molecules to be ranked. The best characterised clinically relevant allergens were identified in peanut, hazelnut, cow's milk, fish and crustacean shellfish with data lacking for allergens from foods such as pecan, Macadamia, lupin and melon. Furthermore, an assessment of *in silico* tools allergenicity prediction found that, whilst many were able to correctly predict allergenicity, none were able to provide an output that could be linked to the clinical relevance. Building on these outcomes an approach for allergenicity risk assessment has been developed that brings together elements of exposure assessment, combining *in silico*, *in vitro*, and *in vivo* methods. Tools for assessment of risks of cross-reactive allergies are more mature and only require refinement to improve the outputs to inform the allergenicity risk assessment process. However, as mechanisms underlying development of food allergy are not fully elucidated, and remain a matter of ongoing research, prediction of *de novo* sensitisation is uncertain.

Key words: Food, allergen, clinical relevance, systematic review, bioinformatic tools, allergenicity risk assessment

Question number: EFSA-Q-2024-00292
Correspondence: nif@efsa.europa.eu

Acknowledgements: We would like to thank the scientific officers Estefania Noriega Fernandez and Antonio Fernandez Dumont for their critical review of this work. The authors thank those who participated in the stakeholder workshops (06.12.2022, 09.01.2023).

Suggested citation: Mills, E.N. C., Orsenigo, F., Salgado, D., Finglas, P. M., Astley, S. 2024. Novel strategies for predicting allergenicity: development of a ranking method and screening tools to assess the allergy risk of innovative proteins. EFSA supporting publication 2024:EN-8840. 124 pp. doi:10.2903/sp.efsa.2024.EN-8840

ISSN: 2397-8325

Summary

Immune-mediated adverse reactions to food are the result of immune dysregulation and results in conditions such as immediate IgE-mediated hypersensitivity reactions the T-cell mediated intolerance to gluten known as coeliac's disease. To protect susceptible individuals who already have, or are at risk of developing such conditions, novel foods and genetically modified organisms (GMOs) undergo an allergenicity risk assessment. This is undertaken in an integrative fashion combining bioinformatic approaches, comparing new proteins with existing allergens together with experimental data on digestibility and reactivity with serum IgE from patients with relevant types of allergies. These assessments could be improved through use of well-defined clinically relevant allergen molecules with different allergenic potential.

The current project has sought address these issues by of developing novel strategies for predicting allergenicity of innovative/novel proteins by:

(1) Developing a ranking method for proteins with different allergenic potential according to their clinical relevance and screen existing tools to assess allergenicity risk of innovative/ novel proteins for use in subsequent activities.
(2) Investigating potential *in silico* tools and follow up actions (*in vitro* and/or *in vivo* methods) needed for an improved allergenicity assessment.
(3) Developing a novel approach for allergenicity assessment of innovative/novel proteins by integrating *in silico*, *in vitro* and *in vivo* methods through implementation of the final ranking strategy of known allergens

A systematic review was undertaken for using a protocol that include a Population-Outcome (PO) approach, linked to a Population, Exposure, Comparator, Outcome (PECO) approach. This was used to identify relevant literature relating to the allergenic molecules found in foods listed on Annex II of the Food Information for Consumers Regulation together with additional foods known to cause IgE-mediated food allergies in at least one European region with a prevalence of 0.5%, such as kiwi, apple, peach, tomato, banana, melon, carrot, lentil and buckwheat. A total of 752 papers were then ranked regarding the type of patient population and the quality of their food allergy diagnosis from whom biological samples (serum or immune cells) were taken and used to define the allergenicity of a protein molecule. Secondly allergen quality and the test method used to define allergenicity was assessed. Other aspects such as the size of the patient population, and whether it included cases and controls, number of reports and geographic distribution were also considered in the ranking. Issues were identified with the quality of data in many publications with poor descriptions of patient populations and a lack of data on allergen characterisation, with some publications describing allergen sequences without including any sequence accession. High quality patient populations were found in papers where allergens were being considered for use in component resolved diagnosis but there is a lack of transparency regarding the quality of the allergen components used in commercial diagnostic tests used in such studies. The best characterised clinically relevant allergens were identified in peanut, hazelnut, cow's milk, fish and crustacean shellfish. Data were lacking regarding allergens from foods such as pecan, Macadamia, lupin and melon. Clinical relevance was assessed and ranged from very high

(such as the peanut allergen Ara h 2) where the PO assessment scored highly (e.g. included unselected study population(s) with food allergy diagnosed using oral food challenge) and the PECO assessment score was also very high, with studies having a case-control design where sensitisation was linked to clinical allergy with analysis such as receiver operating curve analysis being undertaken across many study centres. In contrast many allergens were scored as being of low or very low clinical relevance, such as the thaumatin-like protein precursor identified in melon of which there is one report in one patient.

Most allergen molecules in the systematic review were aligned and correctly predicted using AllergenONline and COMPARE using the classical FAO-WHO FASTA methodology, likely reflecting their inclusion in the AllergenONline and COMPASS databases. Several novel bioinformatic tools were also assessed, including AllergenFP, AlgPred and AllerCatPro. Sometimes tools provided alignments with proteins which were difficult to interpret with regards their relevance to allergenicity risk assessment whilst others provided a good range of outputs that also sought to address issues such as 3D structure assessment and IgE-epitope analysis. None of the tools were able to provide an output that could be linked to the clinical relevance score, and many suffered from identification of both false positive and false negative allergens.

Building on the outcomes of the systematic review and assessment of *in silico* tools an approach for allergenicity risk assessment has been identified which brings together elements of exposure assessment, alongside assessment of the risks posed by alternative protein ingredients of either causing cross-reactive reactions in the existing allergic population or initiating new food allergies. The exposure elements and the assessment of risks of cross-reactive allergies where *in silico* and *in vitro* test methods are more mature and only require refinement to improve the outputs to inform the allergenicity risk assessment process. However, as the mechanisms underlying the development of food allergy are not fully elucidated and remain a matter of ongoing research inevitably the prediction of *de novo* sensitisation is also uncertain.

The assessment undertaken in this systematic review and the ranking for clinical relevance will help the developers of *in silico* methods to refine their outputs and improve ease of use. However, it has identified shortcomings in the quality of many publications relating to the clinical relevance of allergens. There is a need to revisit approaches, such as those developed in EuroPrevall, to ensure effective studies are published with good quality data on patient populations linked to high quality allergen molecule characterisation and effective test methodology. There is also a need to develop approaches to the validation of *in silico* tools that employ both allergenic and non-allergenic comparators. Filling such gaps will improve the quality of risk assessment relating to risks of causing cross-reactive allergic reactions in the existing population. Prediction of *de novo* sensitisation will require new approaches and building on other bodies of research, such vaccinology where the capacity of substances to initiate antibody responses may add value in developing and validating *in silico* and *in vitro* tests.

Table of contents

1 Introduction

Food allergy is an adverse reaction to foods that arises when the immune system mounts an aberrant response to food components, particularly proteins. It can manifest in conditions such as immediate IgE-mediated hypersensitivity reactions and coeliac's disease, an enteropathy triggered by consumption of gluten that is mediated by T-cells (Johansson *et al.*, 2001). To protect existing allergic populations and prevent development of new food allergies, novel foods and genetically modified organisms (GMOs) undergo risk assessment in terms of both IgE- and non-IgE-mediated adverse reactions (EFSA Panel on Genetically Modified Organisms *et al.*, 2017). This process is complex, in part because dietary proteins can only initiate an allergic response in susceptible individuals (i.e., being allergenic is not a property that resides solely in the protein). Consequently, food allergy challenges many of the classical paradigms associated with novel foods risk assessment since protein molecules responsible for triggering allergic reactions are innocuous for the majority.

In silico methods have proven useful in identifying proteins that have the potential to cause cross-reactive food allergies – proteins in one food or substance (e.g., latex) share characteristics with those in another – building on multiple sequence alignments to identify levels of homology between novel proteins and allergens (Poulsen, 2004). Homologies occur between proteins from foods that originate from closely related species and have been especially well described for tree nut allergies. Thus, the concordance of walnut and pecan nut allergies, like that of pistachio and cashew nut, is very high and reflects the close phylogenetic relationships between these tree nut species, the underlying extensive sequence similarity of their protein components, and shared IgE-epitopes (Brough *et al.*, 2020, Nesbit *et al.*, 2020). Thus, the basis for risk assessment to identify potential allergenicity risk posed to existing allergic populations is well founded. However, the risk assessment process is much less certain in predicting which food proteins are likely to give rise to new food allergies, often termed *de novo* sensitisation. In part, this is because there is a lack of effective predictive animal models and those that are available have widely acknowledged limitations. Such shortcomings are compounded by incomplete understanding of mechanisms whereby individuals become allergic.

There are concerns that the original CODEX Alimentarius recommendations regarding *in silico* analysis, which were developed to assess the allergenic potential of newly expressed proteins in GMOs, overpredict allergenicity especially when applied to novel ingredients (Abdelmoteleb *et al.*, 2021). The latter are composed of a mixture of proteins from an organism and, instead of analysing a single protein sequence, the CODEX bioinformatic assessment limits of >35% identity over 80 amino acids are applied to genome- or proteome-level information. However, such simplistic application of CODEX rules runs the risk of mis-identifying potential allergens. For example, tropomyosins are well characterised crustacean shell-fish allergens, but only tropomyosins from invertebrate species are allergenic (Jenkins *et al.*, 2007). Application of CODEX rules finds >35% identity between tropomyosins of insect (e.g., drosophila), mammalian (e.g., human, bovine), and Piscean origins (Abdelmoteleb *et al.*, 2021). Such spurious findings show the limitations of this *in silico* methodology and indicate that other contextual information (e.g., phylogeny) needs to be taken into account to improve utility and prevent unnecessary requirements for targeted serum screening, which currently would be required to confirm potential cross-reactive allergenicity (EFSA Panel on Genetically Modified Organisms, 2010).

More sophisticated *in silico* approaches have also been developed over the last 15 years to help support this aspect of allergenicity risk assessment, some of which have sought to take into account information that goes beyond simply the amino acid sequence, such as three-dimensional structure [e.g. AllerCatPro (Maurer-Stroh *et al.*, 2019)]. Others have applied machine learning to bring together physicochemical and biochemical properties to help identify potential allergens (Westerhout *et al.*, 2019). Likewise, a list of allergens has been compiled for soybean to assess the impact of transgenesis on endogenous allergenicity (Selb *et al.*, 2017). However, it is not clear what the added value of these tools is within the step-by-step risk assessment process and how they might be used alongside experimental data.

One experimental methodology that has become approaches embedded in CODEX recommendations for allergenicity risk assessment are digestibility tests, such as pepsin resistance (Astwood *et al.*, 1996b, Thomas *et al.*, 2004). Although early correlations observed between pepsin resistance and allergenic potential (Astwood *et al.*, 1996b) have not been confirmed fully for IgE-mediated food allergies (Fu *et al.*, 2002), nevertheless, the fact that digestion reduces sensitisation potential means linking resistance to digestion with immunological readouts is important (Bøgh and Madsen, 2016). Indeed, resistance of gluten proteins has been shown to have a role in the pathogenesis of coeliac disease (Shan *et al.*, 2005) and forms part of the approach developed for inclusion of non-IgE-mediated food allergies in allergenicity risk assessment (EFSA Panel on Genetically Modified Organisms *et al.*, 2017). Further elaboration of *in vitro* digestibility tests has been undertaken by EFSA (Torcello-Gomez *et al.*, 2020a, Torcello-Gomez *et al.*, 2020b) and there is now a pressing need to understand how to integrate digestion data alongside *in silico* data to define persistent digestion fragments that are of concern in a manner that also takes account of peptide length (EFSA Panel on Genetically Modified Organisms *et al.*, 2017, Fernandez *et al.*, 2019).

Advances have also been made in *in vitro* methodology to determine whether IgE responses *per se* are indictive of allergenic activity (Patil *et al.*, 2020) using IgE-epitope and effector cell assays, rather than simple IgE-measurements, which promise to improve readouts by making them more clinically relevant (Bahri *et al.*, 2018, Suprun *et al.*, 2019). It is, therefore, timely to consider how such new methods can be integrated into the wider allergenicity risk assessment toolbox, particularly given the poor performance of animal models and the drive to reduce reliance on animals in experiments. Alternatives include immune cell models (especially the innate immune system) linked with models of the gut barrier epithelium, exploiting *ex-vivo* cell cultures that both reduce the need for animal experimentation and have the potential to provide improved read-outs (Lozano-Ojalvo *et al.*, 2019). These approaches are already being applied for the assessment of novel protein ingredients (Smits *et al.*, 2021a, Smits *et al.*, 2021b), so understanding how such immune-based assays can contribute to improved allergenicity risk assessment is critical.

Underpinning any assessment of *in silico* and alternative experimental approaches for assessing the allergenic potential of novel proteins is the need for well-defined clinically relevant allergen molecules with different allergenic potential, either as purified proteins or in curated allergen sequences databases. A host of proteins have been identified as having the capacity to bind IgE, but experimental data underpinning that identification is of highly variable quality. For example, an almond 60S ribosomal protein (*Prunus dulcis*) has been characterised as IgE-binding (Abolhassani and Roux, 2009), based simply on an immune-dot blot of protein produced using a cDNA expression library from pooled sera in a poorly

described patient population sensitised to almond. Similarly, the profilin allergen from peanut, Ara h 5, also identified using cDNA expression library screening, is not found in peanut seed using proteomics approaches and hence is not as such a relevant food allergen (Johnson *et al.*, 2016). This contrasts with the level of detail and data quality available for the peanut allergens Ara h 2 and Ara h 6, where the importance of the post-translational modification of hydroxy-proline for IgE-binding (Bernard *et al.*, 2015). Furthermore, the clinical significance of Ara h 2 and Ara h 6 has been identified, as a marker of clinical allergy to peanut, in many patient populations in contrast to the birch pollen homologue, Ara h 8, which is associated with tolerance (Nicolaou *et al.*, 2011, Ballmer-Weber *et al.*, 2015, Asarnoj *et al.*, 2012).

Many databases define allergens as those proteins that can trigger an immune-mediated adverse reaction and, notably, those that bind IgE. A recent critical evaluation of allergen sequence databases highlighted issues of data curation and updating as well as financing (Radauer and Breiteneder, 2019) and, whilst some databases, such as the WHO/IUIS Allergen Nomenclature database (www.allergen.org), have clear and strict rules on data quality required to designate a protein as an allergen for inclusion (Sudharson *et al.*, 2021), this is not the case for many others (Radauer and Breiteneder, 2019). There has also been an emphasis on identification of allergen molecules, but little attention has been given to characterisation of proteins from foods that could be identified as potential "hypoallergens". Identifying allergenic comparators has the potential to provide a much-needed benchmark against which the allergenic potential of novel proteins can be evaluated. Despite the importance of identifying clinically relevant allergen sequence sets to support assessment of *in silico* and experimental approaches for allergenicity risk assessment, even highly curated allergen sequence databases, such as WHO/IUIS and allergen-online (www.allergenonline.org), do not identify which allergens are the most clinically relevant. Thus, currently, the lack of a curated database of allergens with differing allergenic potentials is hampering development of improved *in silico* and *in vitro* methods for allergenicity risk assessment.

1.1 Background and terms of reference as provided by the requestor

This contract/grant was awarded by EFSA to:

Contractor/Beneficiary: EuroFIR AISBL and subcontractors the Universities of Manchester and Surrey

Contract/Grant title: Novel strategies for predicting allergenicity: development of a ranking method and screening tools to assess the allergy risk of innovative proteins

Contract/Grant number: OC/EFSA/GMO/2021/04

1.2 Interpretation of the Terms of Reference

The overall aim of the project is to develop novel strategies for predicting allergenicity of innovative/novel proteins. It will be met by delivering the following objectives:

Objective 1: Develop a ranking method for proteins with different allergenic potential according to their clinical relevance and screen existing tools to assess allergenicity risk of innovative/ novel proteins for use in subsequent activities.

This aim will be delivered by collecting, quality-assessing, and integrating information about the clinical relevance of food allergens using a systematic review framework. A suitable ranking approach will be proposed based on clinical relevance and other information useful for risk assessment (e.g., prevalence of sensitisation, allergenic potency [as indicated by minimum eliciting doses] and reported severity of reactions [e.g., derived from anaphylaxis registry data]). This will provide solutions for ranking foods and allergens built on their clinical, and hence public health, relevance.

Existing tools (*in silico*, *in vitro* and *in vivo*) will be reviewed and screened for use in further activities.

Objective 2: Investigate potential in silico tools and follow up actions (*in vitro* and/or *in vivo* methods) needed for an improved allergenicity assessment.

In consultation with EFSA, the most appropriate ranking strategy for proteins with different allergenicities among those identified under Task 1 will be identified and used to develop a master list of foods/molecules in rank order.

The effectiveness of selected in silico tools, used alone and in combination, for identifying potential allergenic risks will be assessed by comparing outputs from these tools (i.e., allergenicity ranking) against the master list. Any shortcomings with the tools will be identified and, where appropriate, adaptations made or recommended to improve future performance.

Where an allergen/food is flagged as being of potential allergenic importance by in silico analysis options for follow-up actions will be identified, regarding choice of *in vitro* and *in vivo* methods to be used, to further improve allergenicity risk assessment.

Objective 3: Develop a novel approach for allergenicity assessment of innovative/novel proteins by integrating *in silico*, *in vitro* and *in vivo* methods through implementation of the final ranking strategy of known allergens.

The integrated strategy developed in Objective 2 will be applied to explore and review the capacity of Objective 3 methods to rank foods/molecules for allergenic potential.

Based on the outcomes of this exercise, optimal combination(s) of methodologies will be identified that offer the greatest assurance in identifying the allergenic potential of a novel protein together with gaps and needs for future development of the method(s). This will form the basis of a strategy for development of an integrated approach linking in silico assessment with follow-up experimental *in vitro* and *in vivo* tests for allergenicity risk assessment in the future.

2 Methodologies

2.1 Systematic review methodology

Searches were performed in MEDLINE (OVID) and Scopus. The list of search terms was compiled in English including common names and synonyms based on SNOMED alt Labels, as well as common names in French, and Spanish; a list in Japanese was also developed using both characters and romaji. This was complemented with backwards-and-forwards searching to ensure all relevant publications were retrieved. Studies were excluded if the full text was not available; studies that did not describe the IgE-binding molecules or unrelated to the questions; book chapters, non-peer reviewed case reports, editorial materials which are expressing the opinion of the editor or publisher, meetings, conferences, seminars, workshops, congress, symposiums, patents, proceeding papers; review articles; animal studies and therapeutic studies.

The approach taken includes a Population-Outcome (PO) approach to address the primary research question "What scientific knowledge (evidence) is there that clinical manifestation(s) of IgE-mediated allergic reaction(s) are caused by ingestion of a food?". This aspect assesses the quality of diagnosis of the patient population used to characterize an allergen molecule (Figure 1).

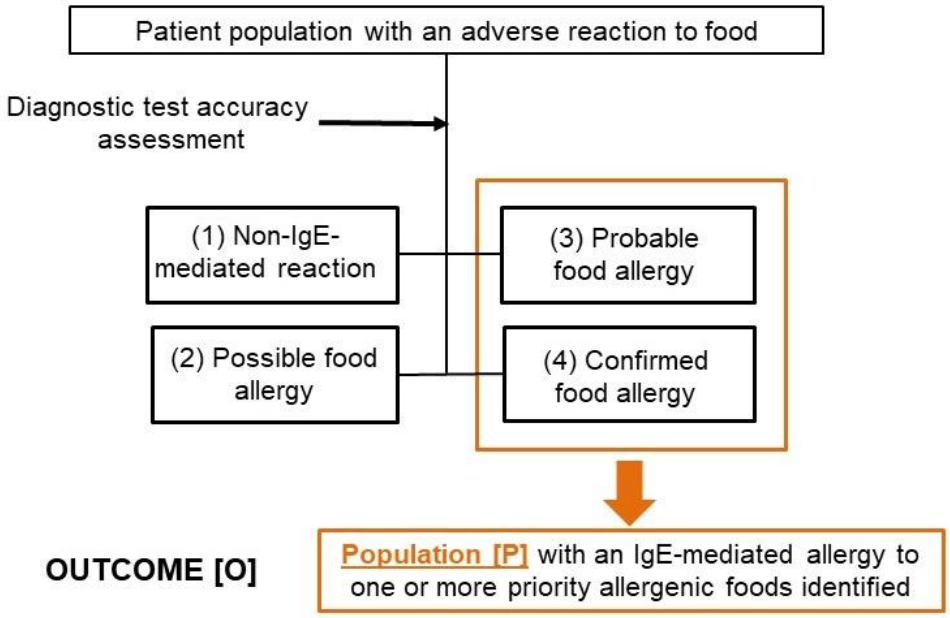

FIGURE 1: Framework for addressing the primary research question using a Population-Outcome (PO) approach.

The PO approach (Figure 1) will be used to answer the primary question where P represents the population evaluated for an IgE-mediated allergy to food and the outcome (O) or condition of interest, in this case whether an individual has an IgE-mediated allergy to one of the selected foods. The outcome will be graded for quality of diagnosis (test accuracy), based on principals described in EAACI Food Allergy Guidelines (Muraro *et al.*, 2014, Soares-Weiser *et al.*, 2014) and criteria proposed by Bjorksten *et al.* (Bjorksten *et al.*, 2008). The quality assessment builds on the following clinical definition of an individual having an IgE-mediated food allergy (Grabenhenrich *et al.*, 2017, FAO-WHO, 2022a) where they must have

Symptoms including any of the following:

- Skin: Itching (pruritus) or tingling (paraesthesia) in the mouth, lips, ears or throat; Swelling of the eyes, lips, or mouth; Nettle sting like rash or itchy skin, or red rash (urticarial rash, flush, erythema); angioedema;
- Alimentary tract: blisters of the oral mucosa; dysphagia; hoarseness or swelling of throat; diarrhoea (other than food poisoning); vomiting (other than food poisoning); stomach cramps; nausea; bloating;
- Respiratory tract: a runny, stuffy nose, or sneezing; red, sore, or running eyes; cough, wheeze, chest tightness, or breathlessness (dyspnoea); laryngeal oedema; dysphonia; reduced peak expiratory flow/drop in FEV1; silence (in lung auscultation); cough;
- Cardiovascular/neurological: Headache; anxiety; tiredness; fainting or dizziness; hypotension/drop of blood pressure; change in consciousness; seizures; change in heart rate/tachycardia; uterine cramps.
- Symptom onset occurring within 2h of consuming an offending food;
- Evidence of sensitisation to food established through skin prick testing and/or serum specific IgE testing.

The grading and risk of bias assessment was included both regards the source of the allergic population and the quality of the diagnosis of an IgE-mediated food allergy. The allergic patient population was graded as follows:

1. Study population or nested case control studies in single study centres
2. Surveys of out-patient clinic patients across multiple study centres
3. Surveys of out-patient clinic patients in a single study centre
4. Case reports

This grading reflects the validity of different study designs to deliver unbiased data with which the primary question can be addressed with grade 1 being the highest quality population to address the primary question. A geographic centre is defined as any location within a 50 mile/80 Km radius of another.

Risk of bias was considered to arise from how closely the study population represents the (food allergic) population. Sources of bias for outpatient clinics result from bias in on-demand healthcare referral systems that disadvantage low socio-economic (SES) groups, those from black and minority ethnic groups or indigenous peoples, and sex and gender biases, where more women than men seek healthcare support, but symptoms are more likely to be negated.

These biases are reduced in unselected study populations, although these too are subject to biases arising from response rates. Similarly, bias from missing data might arise from lack of funding for high quality studies in an unselected study population or for developing outpatient clinic studies with higher numbers or spanning geographic centres. The approach described below to estimating risk of bias is based on the study design used in the EuroPrevall cohorts (Kummeling *et al.*, 2009, Keil *et al.*, 2010, Fernandez-Rivas *et al.*, 2015).

Thus, risk of bias estimates for the population are:

VERY HIGH risk of bias: Case reports and outpatient clinic studies describing <10 patients (single or multicentre).

MEDIUM risk of bias: outpatient clinic studies with at least 100 patients from a single geographic centre

MEDIUM-LOW risk of bias: outpatient clinic studies with at least 100 patients from multiple geographic centres

LOW risk of bias: Unselected study populations e.g., birth cohorts and nested case-control studies appropriately powered.

Studies were then graded for according to the following diagnostic outcome based on as the approach of Bjorksten *et al.* (2008) and (Lyons *et al.*, 2019, Lyons *et al.*, 2020). Grading reflects the quality of diagnosis (test accuracy), i.e., robustness of the outcome for addressing the primary research question and will be as follows:

- Challenge confirmed food allergy: gold standard diagnosis of IgE-mediated food allergy where a clinician confirmed food allergy has been further confirmed by oral food challenge (double blind placebo controlled [DBPCFC] or open).
- Clinician confirmed food allergy: a clinician has diagnosed a patient based on reported symptoms associated with consumption of a particular food which are typical of an IgE-mediated food allergy, symptom onset within 2 hours of contact with food and evidence of sensitization to the same food (either a positive skin prick test (a mean wheal diameter ≥3mm compared to the negative control) or a positive serum specific IgE (≥0.35kU/L) to the same food)
- Probable food allergy: where self-reported food allergy is combined with evidence of sensitization to the same food in the form of a positive skin prick test (a mean wheal diameter ≥3mm compared to the negative control) or positive serum specific IgE (≥0.35kU/L) to the same food.individuals with evidence of sensitisation to selected foods and a convincing history of a reactions to those same foods within two hours of consumption.
- Possible food allergy: self-reported food allergy with symptoms consistent with an IgE-mediated food allergy occurring within 2h of consuming the problem food.

Studies of populations with confirmed food allergy will be ranked higher than those with probable food allergy, the lowest ranking given to those with possible food allergy

In this aspect, risks of bias arise from:

HIGH risk of bias where clinical history and evidence of sensitisation are not linked

MEDIUM risk of bias: linking clinical history to sensitisation (probable food allergy) but there is still a risk of bias since clinical history relies on patient recall and access to healthcare.

LOW risk of bias: evidence of past anaphylaxis or a positive oral food challenge (open, single or a double-blind placebo-controlled food challenge).

Only papers describing patient populations as having either a probable IgE mediated food allergy or confirmed food allergy (through clinician diagnosis, oral food challenge or evidence of severe reactions such as anaphylaxis) ware taken forward into the modified Population-Exposure-Comparator-Outcome (PECO) analysis (Figure 2).

The PECO approach addresses the secondary research question "Which food protein molecules are recognised by serum-IgE from individuals allergic to foods (identified by addressing the primary question) and are responsible for causing an IgE-mediated adverse reaction to those foods?"

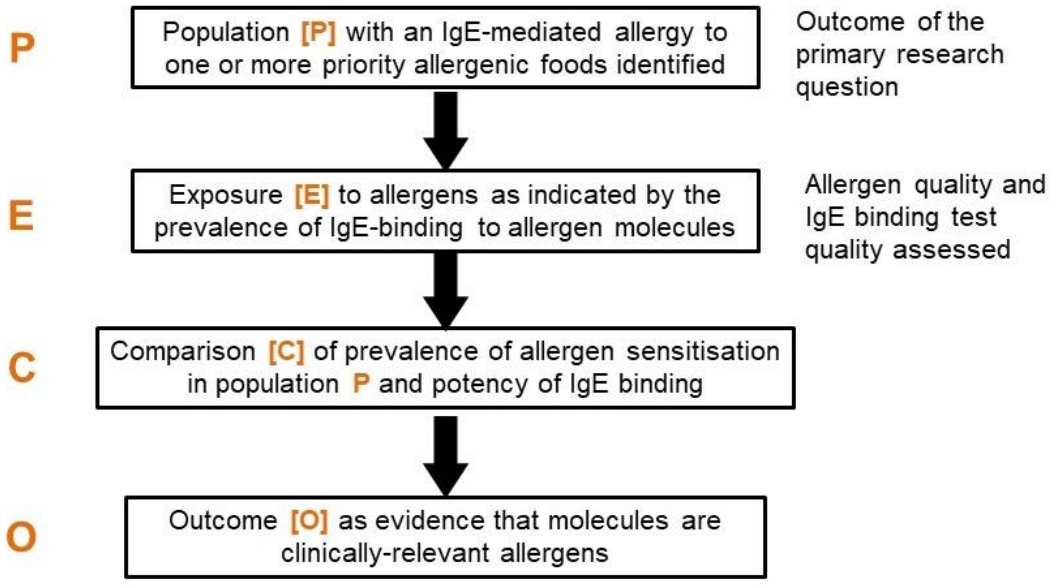

FIGURE 2: Framework for addressing the secondary research question using a modified Population-Exposure-Comparator-Outcome (PECO) approach (based on that of Javed et al., 2017).

Prior to assessing exposure [E], two tests of accuracy were applied, one related to the quality of allergen preparations and the other methodology used to assess IgE binding

Allergen (food protein) preparation and quality characteristics: Food protein preparations can be crude allergen extracts, native purified proteins, or recombinant proteins from food as consumed. The grading reflects the quality of allergenic food proteins used for analysis including their relationship with the food source with the highest quality rank being:

1. Well-characterised purified native allergen (sequence confirmation including N-terminal sequence and mass data) from the food as consumed.
2. Recombinant allergen with confirmed sequence, folding and aggregation information, and protein-level evidence of expression in foods as consumed.
3. Native allergen with no sequence information.
4. Recombinant allergen without folding and/or aggregation confirmation, or peptides corresponding to segments of the allergen sequence, and protein-level evidence of expression in foods as consumed.
5. Partial purified allergen from foods as consumed.
6. Crude extract from foods as consumed.
7. Purified protein, recombinant protein or extracts, but no protein-level evidence of expression or presence in the food as consumed.

Biases resulting from missing data might arise from lack of funding for high quality studies employing well characterised allergens. Clinical studies of IgE reactivity often lack details on biochemical characterisation of allergen molecules used for analysis and vice versa. They were assessed as follows:

HIGH risk of bias: lack of data demonstrating allergens are expressed or present in the food as consumed (e.g., present in root but not in leaves that are typically eaten).

MEDIUM-HIGH risk of bias: allergens have not been authenticated with respect to sequence or folding.

MEDIUM-LOW risk of bias: Purified native allergens or recombinant allergens for which at least molecular masses have been determined by, for example, SDS-PAGE; synthetic peptides used that, whilst retaining parts of the primary sequences, lack post-translational modification or tertiary structures attributes of intact native proteins.

LOW risk of bias: native proteins with a confirmed structural information

Quality assessment of the test used to determine IgE binding: Different types of (diagnostic) tests can be used to define whether a particular protein is an allergen that can induce IgE-mediated reaction(s), with *in vivo* assessments graded higher (1 or 2) than *in vitro* tests using biological samples from patients with a relevant food allergy (graded 3-6). Specifically:

1. *In vivo* challenge test in a confirmed food allergic individual.
2. Skin prick test in a confirmed food allergic individual.
3. Effector cell activation (e.g., basophil histamine release) using either cells or serum from confirmed food allergic individual.
4. IgE-immunoassay using serum samples from confirmed food allergic individual.
5. IgE-dot blotting with a purified protein or immunoblotting following separation of allergen from a confirmed food allergic individual.
6. Dot blotting using allergen extracts and serum samples from confirmed food allergic individual.

It is known that sensitisation to certain types of allergen molecule varies across Europe with the prevalence of sensitisation to Bet v 1 homologues being higher in northern Europe where birch trees are fond, whilst sensitisation to lipid transfer proteins (LTPs) is more common in the Mediterranean area (Fernandez-Rivas *et al.*, 2006, Datema *et al.*, 2015b, Lyons *et al.*, 2021a, Vereda *et al.*, 2011). Consequently, the risk of bias in serological analysis is dependent on both the number of study subjects and their geographic location, with a minimum number of patient sera based on that used for IUIS allergen designation (Pomés *et al.*, 2018) [n=5].

Biases may also result from differences in test methodology. Therefore, the risk of bias will always be lower in studies where multiple test methods are applied. Biases from missing data might arise from lack of funding for high quality studies using proper sampling for biological and technical replicates, control sera, and complementary test methods.

HIGH risk of bias: poor technical replication or low sample numbers (≤5 subjects), or serum pools used, lack of quantitative data, lack of control sera from healthy non-atopic or atopic controls[1].

MEDIUM-HIGH risk of bias: good technical replication but sera from a small study population in only one or multiple centres (≥5-10) used and may lack of control sera from healthy non-atopic subjects or atopic controls.

MEDIUM risk of bias: good technical replication, control sera (atopic and non-atopic control sera) used and sera from a small study population (≥10<20) from either a single or multiple centres.

MEDIUM-LOW risk of bias: good levels of technical replication, control sera (atopic and non-atopic control sera), sera from individuals from single centre (n= ≥20).

LOW risk of bias: good levels of technical replication, control sera (atopic and non-atopic control sera), sera from individuals used, large numbers from multiple centres (n=≥20).

[1] Individuals who are allergic to, for example, pollen or house dust mite but not foods or individuals without a **CONFIRMED** food allergy

Following the execution of the search strategy records were retrieved and curated in individual food specific EndNote libraries (https://doi.org/10.5281/zenodo.11400523). Based on the search terms all references will be uploaded into the relevant EndNote library and duplicate copies of the same article removed using a combination of automatic (using software such as Distiller) and manual screening.

Data were then independently extracted from each of the included references onto a customised data extraction sheet in Excel and checked by two team members. Extracted information included factors such as study design; study population demographic characteristics; study methodology; the quality of allergen preparations used (e.g. well-characterised allergen molecules, recombinant allergen, crude extract etc), methodology used to assess IgE binding (e.g. oral food challenge, SPT, western blotting, dot blotting etc) including (where data quality is sufficient) measures of potency such as IgE-binding capacity; outcomes and prevalence of sensitisation measurement to different allergen molecules in the population; information for assessment of the risk of bias. Tabulation was used to make a summary for exploring the differences and similarities between the studies. The extracted data from the included studies will be quality assessed and used for evidence synthesis.

2.2 *In silico* tools assessment

The performance of different types of bioinformatic tool developed for allergenicity risk assessment was assessed by interrogating them using a set of defined allergens, and non-allergenic sequences. The tools included classical sequence identity searching tools and other types of alignment free methods as summarized in Table 1 below. The tools using these approaches were systematically assessed with regards their ability to correctly classify the allergenic and non-allergenic sequences.

Algorithm Name	Functionality
Classical sequence identity searching tools	
FASTA	A heuristic methods FASTA searches for identical stretches of sequence ≥8 amino acids in length between a query protein and sequences in a database. It is slower and more sensitive than BLAST as it tolerates gaps in aligned sequences. It generates Expect values (E) which indicate the significance of a sequence similarity score and decreases exponentially as the Score (S) of a match increases. It is also a function of the database size used for searching. Pearson, WR; Lipman, DJ (1988). "Improved tools for biological sequence comparison". Proceedings of the National Academy of Sciences of the United States of America 85 (8): 2444–8. doi:10.1073/pnas.85.8.2444. PMID 3162770. Bibcode: 1988PNAS...85.2444P. http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez&artid=280013 https://www.ebi.ac.uk/Tools/sss/fasta/
Alignment-free methods	
Auto cross covariance (ACC)	ACC is a protein sequence mining method where protein sequences are transformed into uniform equal-length vectors. It has been applied to quantitative structure-activity relationships (QSAR) studies of peptides with different lengths and provides a motif searching method (Wold *et al.*, 1993).
Support vector machine (SVM)	This ML classifier has been used alone (AlgPred 2.0) and in combination with BLAST/FASTA searching defined by CODEX (Allerhunter). Allerhunter is no longer available, but an ML algorithm is available in AlgPred 2.0 although it is not clear what the algorithm underpinning the tool is.
Random Forest (RF)	A RF classifier comprises a large number of individual decision trees that operate as an ensemble (a forest). Each individual tree in the random forest provides a class prediction and the class with the most votes becomes our model's prediction
K-Nearest Neighbor (kNN)	kNN is a supervised classifier which uses the k-nearest neighbors to vote for each query point; the value of k is an integer value which is defined by the user. As a dataset grows kNN becomes increasingly less efficient to use.
Multi-layer Perceptron (MLP)	MLP is a neural network classifier, and one tool (AlgPred 2.0) uses the sklearn.neural network module. https://scikit-learn.org/stable/

Table 1: Bioinformatic algorithms and approaches used in allergenicity prediction tools

3 Assessment/Results

3.1 Task 1: Develop a ranking method for proteins with different allergenic potential according to their clinical relevance and screening of existing tools to assess the allergenicity risk of innovative/novel proteins for use in subsequent activities

3.1.1 SubTask 1.1 Systematic review protocol to identify clinically relevant allergens

A protocol for the systematic review was developed based on one previously developed for tree nut allergens (Javed *et al.*, 2017) through discussion with expert advisors to the project and two stakeholder discussions meetings. The final protocol has been published (DOI 10.5281/zenodo.8156129) and submitted to PROSPERO (Clare Mills, Federica Orsenigo, Siân Astley, Paul Finglas, Antonio Fernandez Dumont, Simon Hubbard, Jim Warwicker, Angela Simpson, Silvia Bulfone-Paus, Diana Salgado. Systematic review protocol to identify clinically relevant food allergens. PROSPERO 2023 CRD42023422361. In brief, the adaptation was to:

Updating and refining the primary and secondary research questions and the search terms which were revised to include all foods on Annex II of the Food Information for Consumers Regulation (European Parliament, 2011). These included

Plant derived foods: Tree nuts (Almond, Brazil nut, cashew and pistachio, Macadamia nut, walnut and pecan), legumes (lupin, peanut, soybean), oilseeds (sesame, mustard) and vegetables (celery);

Animal derived foods: milk as cow's milk, egg as hen's egg, crustacean and molluscan shellfish, and fish.

Refining the inclusion and exclusion criteria to take account of advances in the field, such as excluding immunotherapy trials which might add further bias in the data collection since trials sometimes have specific inclusion criteria regarding severity of reaction and patterns of sensitisation.

Updating the quality assessment approach by, for example, including elements relating to peptide mapping used for IgE epitope definition.

Including a much more detailed description of risk of bias assessment.

3.1.2 SubTask 1.2 Search execution

Search terms were applied to retrieve articles from Scopus and PubMed by Dr Siân Astley (EuroFIR, BE), on 11-13th April 2023, 18th July 2023, and 7th August 2023.

Search queries based on the nested and Boolean term (EN) were entered using advanced search query strings that did not limit search fields. Filters were not applied. Search results were sense-checked (i.e., food item and food allergy) in the broadest of terms, largely to eliminate human error (e.g., typographical, transposition, misinterpretation, duplication, incorrect or incomplete terms, wrong field, date and time errors, misalignment [failure to adhere to established standards and conventions], formatting errors, or fatigue).

All reference information was downloaded using Scopus and PubMed export functions (Scopus .ris and PubMed .nbib) including author(s), document title, year, EID, source title, volume, issues, pages, citation count, source & document type, publication stage, DOI, open access, bibliographical information, affiliations, serial identifiers, e.g., ISSN, PubMed ID, publisher, editor(s), language of original document, correspondence address, abbreviated source title, abstract & keywords, abstract, author keywords, indexed keywords, funding details, number, acronym, sponsor, funding text, other information, tradenames & manufacturers, accession numbers & chemicals, and conference information, as available.

Files were uploaded to EndNote 21 and papers selected systematically, based on inclusion and exclusion criteria, within EndNote for easy organisation and tracking, after removal of the duplicate records. Titles and abstracts for each article were screened independently by three researchers (Siân Astley, Angelika Mantur-Vierendeel, and Christina-Ariadni Valagkouti, EuroFIR - BE), based on the primary research question.

Full-text copies were retrieved for articles that passed title and abstract screening.

These papers were read by the same individuals, and Population-Outcome (PO) criteria applied rigorously to decide which studies considered "What scientific knowledge (evidence) is there that clinical manifestation(s) of IgE-mediated allergic reaction(s) are caused by ingestion of a food?". This aspect assesses the quality of diagnosis of the patient population used to characterize an allergen molecule (Figure 1). Those papers that passed this scrutiny were passed to University of Surrey – UK for PECO analysis and further backwards-forwards searching to ensure all relevant publications were captured.

3.1.3 SubTask 1.3: Quality ranking of included references

Quality ranking was undertaken by grading papers and risk-of-bias assessment on a food-by-food basis.

Firstly, this was undertaken regarding the patient population and the outcome (clinical diagnosis of food allergy) [PO]. Overall, the food allergic populations used in allergen molecule discovery and characterisation were drawn from outpatient clinics and generally a single centre. The numbers of patients included in early stage, allergen-discovery studies were generally very small and often less than ten. Some of the publications from studies published more than 30 years ago lacked detailed descriptions of the patient panels used (Leung *et al.*, 1994) and even in more recent studies it was unclear if patients were chosen simply on the basis of sensitisation to a food and whether that was linked to symptoms experienced on consumption of that food (Yun *et al.*, 2022). For foods for which IgE-mediated food allergies are uncommon, such as lupin, it is challenging to find sufficient patients for studies and consequently studies often draw on individuals with mixed food allergies to legumes which in well-framed studies then have their allergy to a specific food confirmed (Peeters *et al.*, 2007) It was also often unclear whether patients diagnosis followed clinical guidelines but in some papers it was clear that diagnosis of food allergy was generally performed with a clinical history accompanied by evidence of food-specific sensitisation established by skin prick testing and/or serum specific IgE testing. Depending on the foods involved, many were diagnosed with oral food challenges (open or double blind) or were excluded from food challenge due to their having a history of anaphylaxis. However, there were instances where serum samples were sourced from in-house serum banks or commercial sources where clinical data supporting the description of patients and their allergies were either lacking or only partially described (Lin *et al.*, 2023, DeWitt *et al.*, 2004). In some instances, such serum samples were used to support allergen identification which was then verified using larger patient panels from well described patients (Bauermeister *et al.*, 2011).

As part of the quality assessment (grading) of papers to support the exposure assessment, the test accuracy was assessed the first of which was the quality of allergen preparations used in studies. There is also a great variation in the data available on the quality of allergen preparations used in the discovery and development phase. This was evident in the PECO analysis there with some allergens, especially recombinant allergens, not being well characterised with regards their folding and aggregation state and in some instances, there is no comparison made between the recombinant. There are also instances and natural protein as laid out as a standard for the EuroPrevall allergen library (Hoffmann-Sommergruber *et al.*, 2008b). Indeed, it was apparent in many instances that papers either had excellent descriptions of patient panels and poor descriptions of the allergens used or excellent standards for allergen preparation and characterisation with superficial summaries of patient characteristics. Some excellent papers that really provided a strong evidence base for classifying allergens as being clinically relevant (or not) came from teams of researchers where clinical and basic science was well integrated (Holzhauser *et al.*, 2009, Kabasser *et al.*, 2021).

Publications describing the discovery of allergens often have poor descriptions of both the patient panels used and/or the allergens. One example of a publication providing only very weak evidence of an allergen being clinically relevant is that of the putative almond allergen, Pru d 5, which is a 60S ribosomal protein isolated by screening an expression library. Not natural protein counterpart was prepared (i.e. a lack of data demonstrating the protein is expressed in the food as eaten), the patients who provided the serum samples for allergen identification were not well described, no data were provided on the structure of the

recombinant allergen and IGE binding was only described using dot-blots (Abolhassani and Roux, 2009). Similarly, there are a number of papers using 2D-PAGE IgE immunoblotting to identify the allergen, often probed with small numbers to patient sera since such methodology requires larger volumes of serum with a highly IgE titre (Lu *et al.*, 2018).

There is also a great variation in the data available on the quality of allergen preparations used in the discovery and development phase. Standards for such preparations were first developed in the EuroPrevall project (Hoffmann-Sommergruber *et al.*, 2008a) but these have not been updated, although there are regulations regarding allergen preparations used commercially for diagnosis (Committee for Medicinal Products for Human use, 2008) for registered products. It is to be noted that few products used in food allergy diagnosis have such an approval. Thus, data on folding, aggregations state and confirmation of sequence are often lacking and for recombinant proteins produced directly from genes, data are often lacking showing the presence of the protein in question in the food as eaten.

The interest in allergen molecules has been driven by observations that certain allergen molecules have value in improving diagnosis of food allergy, with the potential to do away with the need for oral food challenges, and/or identify patients at risk of severe reactions to help support patient management and ensure appropriate advice and medication is given to patients post diagnosis. This has given rise to the field of component resolved diagnosis using platform technologies such as the Phadia-Thermofisher ImmunoCAP and ISAC technology and the IMMULITE platform of Siemens amongst others. Often researchers can prepare customised tests, including custom ImmunoCAPs using their own purified allergens which can then be used to gain information on clinical relevance of allergens for diagnosis, as used for almond by Kabasser and team (Kabasser *et al.*, 2021).

The second aspect of test accuracy that was assessed related to the quality of the test methods used to assess a molecule's allergenicity. Many studies are reliant on the availability of commercial reagents and test methods which many be only available for allergens in foods for which the prevalence and severity are high. It is unclear what drives the decision for test method vendors to choose particular molecules to include in such tests, but their availability and cost introduces and additional bias into the analysis. However, their availability means, for those foods for which they are available, larger-scale studies involving larger numbers of patients (often >50) are feasible. They are often applied in CRD studies based on large, population studies which provide the highest level of evidence of clinical relevance. One example of this is the Australian HealthNuts study were sensitisation to the peanut allergen Ara h 2 was assessed in both cases of peanut allergy and control subjects without peanut allergy from the same population (Dang *et al.*, 2012). It also lends itself to complex studies with large patient populations to support identification of combinations of allergens that are important in different geographic regions. (Datema *et al.*, 2015a). Some more recently have also linked to diagnostic tests utilising effector cells such as basophils, cell lines such as LAD-2 and others for foods such as peanut in particular (Santos *et al.*, 2021b, Santos *et al.*, 2014, Bahri *et al.*, 2018). Nevertheless, the PO and PECO approach taken in this review provided sufficient information as to allow the ranking of allergen molecules as to their clinical relevance for all the priority food allergens listed on Annex II of the Food Information for Consumers regulation (add ref). However, metanalysis did not prove possible due to the heterogeneous nature of the papers identified.

A total of 752 papers were graded (Figure 3). Of the foods listed on Annex II of the food information for consumers regulation, peanut had the greatest number of graded papers followed by hazelnut. Other foods with a substantial evidence base were cow's milk, wheat and (collectively) crustacean and molluscan shellfish. However, very small numbers of papers (less than ten each) were reviewed for pecan, Macadamia and lupin. Regarding supplementary foods, fewer publications met the inclusion criteria, with kiwi fruit, apple and peach at a similar level. Evidence to support evaluation of buckwheat allergens was sparse and very little was available for melon.

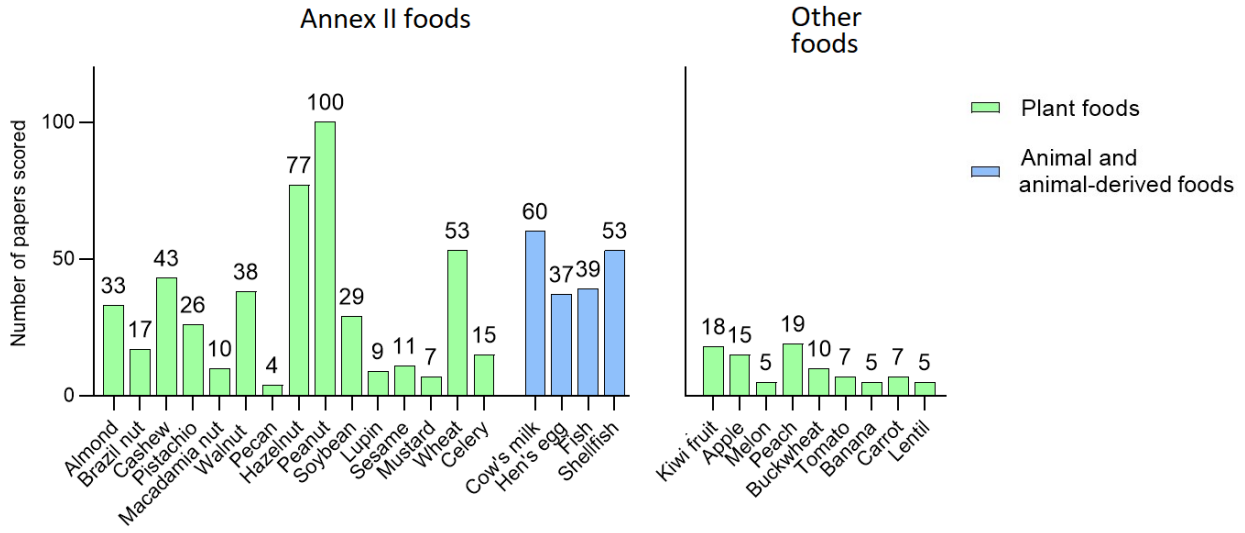

FIGURE 3: summary of numbers of included papers which have been graded on a food-by-food basis. Foods were divided firstly between foods that are mentioned in Annex II from the Regulation (EU) no 1169/2011 of the European Parliament and of the Council on substances or products causing allergies or intolerances, and supplementary foods. Foods are also colour-coded to distinguish from animal and non-animal products.

3.1.4 SubTask 1.4: List of clinically relevant allergens ranked by order of clinical relevance

The allergen molecules identified in the included foods that were quality assessed using the PECO approach were then ranked in order of their clinical relevance (Table 2) and ranged from very low to very high. below

Clinical relevance	Description (at least one of these reasons)
Very low	PO score is low with small numbers of patients, often from case reports PECO score is low and allergen molecules are described for which there is no evidence of they are present in the food as consumed ≤2 papers describing the allergen molecules
Low	PO score is low with small nos of patients from case reports

	PECO score is low >2 papers describing the allergens and patients may only originate from a single study centre
Moderate	PO score is high but patients may only originate from a small number of study centres PECO score is high Studies with a case-control design where significant sensitisation is observed in the control subjects, weakening the link between sensitisation and clinical allergy
High	PO score is high PECO score is high Studies with a case-control design where sensitisation is linked to clinical allergy, and may have allowed analysis such as receiver operating curve analysis. Several papers describing the allergens
Very high	PO is very high including unselected study population(s) PECO score is very high, Studies with a case-control design where sensitisation is linked to clinical allergy, and may have allowed analysis such as receiver operating curve analysis across many study centres Many papers describing the allergen molecules

Table 2: Criteria used to rank the clinical relevance of allergens

They were mapped against the allergens included in the IUIS allergen nomenclature database. Allergens were assigned a clinical relevance from very low (poor quality of allergen preparation and poor quality or small patient population) to very high (high quality allergen preparation and large patient populations with well-defined food allergy).

3.1.5 Plant-derived foods

3.1.5.1 TREE NUTS

Tables 3-8 summarises the ranking of allergens in this food group. In general seed storage protein allergens were generally highly ranked as clinically relevant, followed by the PR10 homologise and LTPs.

Almond: A total of 33 papers were graded and 7 allergens identified (Table 3). One allergen, 60S ribosomal protein, was only described in a single paper from an expression library with only few patients. The lack of characterisation of the allergen molecule, the small number of patients together with a lack of evidence as to the presence of the molecule the food as consumed meant this was given a very low clinical relevance. This was in contrast to the major almond protein, the seed storage protein of almond known as Pru du 6, where there were several papers from different study centres describing its allergenicity, with protein level evidence from the food as eaten including sequence, and in some reports, folding data. These

studies also included larger numbers of patients and demonstrated their high clinical relevance. Interestingly, a protein that is up regulated in almond seeds involved in the synthesis of cyanide that gives almonds their flavour, has been identified as an allergen.

Table 3: Summary of almond allergen quality assessment and ranking
See Suppl Table S3 for references used in the clinical relevance ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
PR10 homologue	Pru du 1	B6CQS9	2	43	Moderate
Profilin	Pru du 4	Q8GSL5	1	18	Low/ Moderate
60S acidic ribosomal protein P2	Pru du 5	Q8H2B9	1	8	Very low
11S seed storage globulin, amandin	Pru du 6	E3SH28, E3SH29	3	326	High
α-hairpinin	Pru du 8	A0A314YX39, P82944	3	73	Moderate
Mandelonitrile lyase 2; Hydroxynitrile lyase	Pru du 10	Q945K2	1	40	Moderate
γ-conglutin 1	Not included in IUIS	P82952	2	21	Moderate

Brazil nut: 17 papers were graded and 2 allergens identified (Table 4). One allergen, the 2S albumin Ber e 1, was described in 7 different papers with 113 patients across 3 study centres. Most papers came from the UK, likely reflecting the higher prevalence of this food allergy in that country. The IUIS Allergen Nomenclature data base also includes the 11S globulin from Brazil nut (Ber e 2) but the citing publication is a meeting abstract and so was excluded in the systematic review.

Table 4: Summary of Brazil nut allergen quality assessment and ranking
See Suppl Table S4 for references used in the clinical relevance ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
2S albumin	Ber e 1	P04403	7	113	High

Cashew and pistachio: These tree nuts were analysed together since they are closely related and allergies to cashew and pistachio show a very high level of concordance (Brough *et al.*, 2020, Nesbit *et al.*, 2020). A total of 43 papers were graded for cashew and 26 for pistachio; three allergens were identified for cashew and four for pistachio (Table 5). Given the quality of papers, and the number of patients and study centres described, all the allergens identified in cashew were classified as being highly clinically relevant. However, level of evidence for clinical relevance was lower, although consistent, across pistachio allergens, and consequently there were classified as being of moderate clinical relevance. In cashew nut, the vicilin-like protein Ana o 1 was described in 11 papers with 331 patients across nine study centres. The legumin-like protein Ana o 2 was described in 13 papers with 430patients across 4 study centres. The 2S albumin Ana o 3 was described in 15 papers with 333 patients across 6 study centres. For pistachio, the 2S albumin Pis v 1 and the 11S globulin Pis v 2 were described in five and 2 papers respectively with 77 and 43 patients. The 7S vicilin Pis v 3 was described in two papers with 35patientswith manganese superoxide dismutase (Pis v 4) was described in three papers with 43 patients. Consequently, the quality of evidence relating to the clinical relevance of pistachio allergens is lower than that for cashew. WHO/IUIS Allergen Nomenclature database also lists another pistachio allergen, Pis v 5, but other than a brief description, but no citation is included in the database entry for the allergen. Although there are papers detailing, for example, the proteomic characterisation of this allergen (Nitride *et al.*, 2013), no publication was identified in the systematic review for this allergen.

Table 5: Summary of cashew and pistachio allergen quality assessment and ranking
See Suppl Table S5 for references used in the clinical relevance ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients	Overall clinical relevance ranking
Cashew					
7S seed storage globulin	Ana o 1	Q8L5L6	11	331	High
11S seed storage globulin	Ana o 2	Q8GZP6	13	430	High
2S albumin	Ana o 3	Q8H2B8	17	758	High
Pistachio					
2S albumin	Pis v 1	B7P072	5	77	Moderate
11S seed storage globulin	Pis v 2	B7P073	2	43	Moderate
7S seed storage globulin	Pis v 3	EF116865	2	35	Moderate
Manganese superoxide dismutase	Pis v 4	B2BDZ8	3	43	Moderate

Macadamia nut: a total of 10 papers was scored and two allergens were identified, both seed storage globulins, which were identified as being of moderate clinical relevance with a single publication (Table 6).

Table 6: Summary of macadamia nut allergen quality assessment and ranking
See Suppl Table S6 for references used in the clinical relevance ranking

IUIS allergen designation	Protein name	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Mac i 1	7S seed storage globulin	Q9SPL3, Q9SPL4	1	27	Moderate
Mac i 2	11S seed storage globulin	C0HLR7	1	27	Moderate

Walnut and pecan: A total of 38 papers was scored for English walnut (Juglans regia) and four papers were scored for pecan (Table 7). For English walnut, four allergens were identified including the 2S albumin, Jug r 1, the 7S vicilin, Jug r 2, the non-specific LTP, Jug r 3 and the 11S globulin Jug r 4. Jug r 4 was identified in 2 papers with 64 participants across 2 study centres, both located in the USA. All the walnut allergens were classified as being highly clinically relevant. There are also entries in the IUIS Allergen nomenclature database for *Juglans nigra* (Black walnut). However, two of these (the 2S albumin allergen designated Jug n 1 and the 7S seed storage globulin designated Jug n 2) are noted in the database as coming from unpublished papers. The third, Jug n 3, is the 11S seed storage globulin, which was scored and given a low clinical relevance score.

For pecan, the 7S vicilin Car i 2 and the 11S globulin Car i 4 were identified. Car i 2 was identified in one paper with 25 participants from a single study centre, while Car i 4 was identified in a single paper with 28 participants from one study centre.

Table 7: Summary of walnut and pecan allergen quality assessment and ranking
See Suppl Table S7 for references used in the clinical relevance ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Black walnut (*Juglans nigra*)					
11S seed storage globulin	Jug n 3	A0A1L6K371	1	25	Low
English Walnut (*Juglans regia*)					
2S albumin	Jug r 1	Q7Y1C2	7	291	High
7S seed storage globulin	Jug r 2	Q7Y1C1	7	466	High
LTP	Jug r 3	C5H617	4	127	High
11S seed storage globulin	Jug r 4	Q2TPW5	4	103	High
Pecan					
7S seed storage globulin	Car i 2	B3STU4	2	25	Moderate
11S seed storage globulin	Car i 4	B5KVH4	1	28	Moderate

Hazelnut: a total of 77 papers was scored, and 8 allergens were identified (Table 8). The Bet v 1 homologue Cor a 1 was identified in 26 papers with 2316 patients across multiple study centres. The profilin Cor a 2 was identified in five papers with 395 participants whilst the LTP allergen, Cor a 8, was identified in 21papers with 11681participants across multiple study centres The 11S globulin, Cor a 9, was identified in 17 papers with 1591participants. The 7S vicilin Cor a 11 was identified in 9 papers with 684participants across. There were few reports describing the oleosins, Cor a 12 and 13, which were identified in 1-2 papers compared the the 2S albumin Cor a 14, which was identified in 12 papers with 1526 participants.

Evidence of clinical relevance across literature was therefore very strong for some of the hazelnut allergens, such as the Bet v 1 homologue Cor a 1, the LTP Cor a 8, the 11S globulin

Cor a 9 and the 2S albumin Cor a 14, which were classified as highly clinically relevance. The 7S vicilin Cor a 11 was classified as being of moderate clinical relevance, while the oleosins Cor a 12 and Cor a 13 as well as the profilin Cor a 2 were classified as being of low clinical relevance.

Table 8: Summary of hazelnut allergen quality assessment and ranking
See Suppl Table S8 for references used in the clinical relevance ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
PR10 homologue	Cor a 1	Q9SWR4	26	2316	High
Profilin	Cor a 2	Q9AXH5	5	395	Low
LTP	Cor a 8	Q9ATH2	21	1681	High
11S seed storage globulin	Cor a 9	A0A0A0P7E3	17	1591	High
7S seed storage globulin	Cor a 11	Q8S4P9	9	685	Moderate
17kDa oleosin	Cor a 12	Q84T21	2	430	Low
14-16kDa oleosin	Cor a 13	Q84T91	1	7	Low
2S albumin	Cor a 14	D0PWG2	12	1526	High

3.1.5.2 LEGUMES

Peanut: a total of 100 papers were graded, and nine allergens were scored (Table 9). The 7S vicilin Ara h 1 appeared in 51papers with 8365participants across more than 40 study centres. The 2S albumin Ara h 2 was described in 52papers with 9158participants across more than 55 study centres. The 11S globulin Ara h 3 appeared in 40 papers with 8218 participants But the profilin, Ara h 5, appeared in only two papers with 97 participants. The 2S albumin Ara h 6 was identified in 19papers with 1584 participants across multiple study centres whilst another 2S albumin variant, Ara h 7, was described in only one paper with 40 participants across 3 study centres. The Bet v 1 homologue Ara h 8 appeared in 30 papers with 7167 participants with the non-specific LTP Ara h 9 identified in 19 papers with 1347participants. Minor allergens included the oleosin Ara h 15, which was described in only one paper with 52 participants across 2 study centres. Ara h 2 was identified as being of very high clinical

relevance, with Ara h 1, 3, 6 and 8 also being identified as of highly clinically relevance due to the very high number of high-quality studies and participants involved in their identification and characterisation. Ara h 9 was identified as being of moderate clinical relevance while Ara h 7 and 15 were identified as being of low clinical relevance.

Table 9: Summary of peanut allergen quality assessment and ranking
See Suppl Table S9 for references used in the clinical relevance ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
7S seed storage globulin	Ara h 1	P43238	51	8365	High
2S albumin	Ara h 2	Q6PSU2	52	9158	Very high
11S see storage globulin	Ara h 3	O82580	40	8218	High
Profilin	Ara h 5	AF059616	2	97	Low
2S albumin	Ara h 6	Q647G9	19	1584	High
2S albumin	Ara h 7	Q9SQH1	1	40	Low
PR10	Ara h 8	B0YIU5	30	7167	High
LTP	Ara h 9	B6CEX8	19	1347	Moderate
Oleosin	Ara h 15	Q647G3	1	82	Low

Soybean: a total of 29 papers were scored, and 11 allergens were identified (Table 10). The hydrophobic protein Gly m 1, the defensin Gly m 2, the profilin Gly m 3, the seed biotinylated protein Gly m 7 and the Bowman-Birk inhibitor Gly m BBI were identified in a single paper with 91 participants in one study centre. The Bet v 1 homologue Gly m 4 was identified in 12 papers with 302 participants across 17 study centres. The 7S vicilin Gly m 5 was identified in 13 papers with 328 participants across 16 study centres whilst the 11S globulin Gly m 6 was identified in 13 papers with 313 participants across 15 study centres. The 2S albumin Gly m 8 was identified in 4 papers with 121 participants across 4 study centres. The Kunitz trypsin inhibitor Gly m KTI was identified in 2 papers with 91 participants across 3 study centres. The

Thiol protease-like protein Gly m Bd 30k was identified in 2 papers with 97 participants across 2 study centres.

Gly m 4, Gly m 5 and Gly m 6 were therefore classified as being of high clinical relevance, while Gly m 8 was classified as being of moderate clinical relevance. The remaining allergens (Gly m 7, Gly m KTI, Gly m BBI and Gly m Bd) were classified as being of low clinical relevance whilst Gly m 1, 2 and 3 were classified as being very low being largely inhalant allergens of little relevance to food.

Table 10: Summary of soybean allergen quality assessment and ranking

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Hydrophobic protein	Gly m 1	Q9S7Z9	1	91	Very Low
Defensin	Gly m 2	C6T0M2a	1	91	Very Low
Profilin	Gly m 3	O65809	1	91	Very Low
PR10	Gly m 4	P26987	12	302	High
7S seed storage globulin	Gly m 5	O22120	13	328	High
11S seed storage globulin	Gly m 6	P04776	13	313	High
Seed biotinylated protein	Gly m 7	C6K8D1	1	91	Low
2S albumin	Gly m 8	P19594	4	121	Moderate
Gly m KTI	Not included in IUIS	P01070	2	91	Low
Gly m BBI	Not included in IUIS	I1L3Q3	1	91	Low
Gly m Bd	Not included in IUIS	O64458	2	97	Low

Lupin: A total of 9 papers were scored, and 8 allergens were identified (Table 11). β-conglutin (Lup an 1) was identified in five papers with 85 participants across five study centres whilst the non-specific LTP (Lup an 3) was identified in one paper with 31 participants across two study centres. γ-conglutin was identified in three papers with 89 participants across three study centres with α- and δ-conglutins being identified in two papers with 40 participants across two study centres. IN addition. the profilin (Lup a 5) was identified in one paper with 31 participants across two study centres, although no sequence was made available in the publication. The 11S globulin Lup 2 was identified in 3 papers with 44 participants across 3 study centres. A vicilin-like protein not included in the WHO/IUIS Allergen nomenclature database was also identified in two different studies, with 49 participants and across two study centres.

On the basis of this assessment the nsLTP Lup an 3 and the profilin were assigned as being of low clinical relevance whilst Lup 2, the 7S vicilin-like protein, the α- and the δ-conglutins were assigned as being of moderate clinical relevance. Lup an 1 and the γ-conglutin were assigned as being clinically highly relevant allergens.

Table 11: Summary of lupin allergen quality assessment and ranking.
See Suppl Table S11 for references used in the clinical relevance ranking.

Protein type	**IUIS allergen designation**	**Sequence accession(s)**	**No of included studies**	**Total no of patients studied**	**Overall clinical relevance ranking**
Narrow leaved lupin (*Lupinus angustifolius*)					
7S seed storage globulin	Lup an 1	Q53HY0	5	85	High
LTP	Lup an 3	A0A4P1RWD8	1	31	Low
α-conglutin	Not included in IUIS	2313076	2	40	Moderate
γ-conglutin	Not included in IUIS	Q42369	2	89	High
δ-conglutin	Not included in IUIS	F5B8W8	3	40	Moderate
White lupin or field lupin (*Lupinus albus*)					
Profilin	Lup a 5	FG090100.1	1	31	Low

Lup 2	Not included in IUIS	85361412	3	44	Moderate
Vicilin-like	Not included in IUIS	89994190	2	49	Moderate

Lentil: Few papers were included for lentil (a total of five) which identified only three allergens which could be ranked two of which were storage proteins and one LTP. Another allergen, an isoallergen of the 7S seed storage globulin (designated Len c 1.0103 by the IUIS allergen nomenclature database), was identified included in a study that passed the PO and PECO analysis, but no sequence nor accession number was provided.

Table 12: Summary of lentil allergen quality assessment and ranking.
See Suppl Table S12 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
γ-vicilin subunit	Len c 1.0101	AJ551424	3	77	Moderate
	Len c 1.0102	AJ551425	1	422	Low
nsLTP1	Len c 3	A0AT29	1	10	Low

3.1.5.3 OTHER SEEDS

Sesame: A total of 11 papers were scored and 6 allergens were identified from three different types of protein, the 2S albumins, the 11S seed storage globulins and oleosins (Table 13).This did not include the 7S seed storage globulin, Ses i 3, which is listed in the IUIS Allergen Nomenclature Database as no study passed the PO and PECO analysis.

Table 13: Summary of sesame (Sesamum indicum) allergen quality assessment and ranking. See Suppl Table S13 for references used in the clinical relevance ranking.

Allergen name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
2S albumin	Ses i 1	Q9AUD1	4	471	High
	Ses i 2	Q9XHP1	3	147	High
Oleosin	Ses i 4	Q9FUJ9	1	35	Moderate
	Ses i 5	Q9XHP2	1	35	Moderate
11S seed storage globulin	Ses i 6	Q9XHP0	1	92	High
	Ses i 7	Q9AUD2	1	92	High

Mustard seed: A total of 7 papers were scored and four allergens were identified (Table 14), all from *Sinapus albus* (yellow mustard) with the 2S albumin and 11S seed storage globulin being ranked as being of the highest clinical relevance. An allergen has been reported from *Brassica juncea* (Indian or oriental mustard) but the publication relating to its characterisation (Monsalve *et al.*, 1993) did not meet the inclusion criterion as using serum samples from patients with either probable or confirmed food allergy.

Table 14: Summary of mustard allergen quality assessment and ranking
See Suppl Table S14 for references used in the clinical relevance ranking.

Protein type	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Yellow mustard (*Sinapus alba*)					
2S albumin	Sin a 1	P15322	2	35	High
11S seed storage globulin	Sin a 2	Q2TLW0	3	53	High
LTP	Sin a 3	E6Y2L9	2	49	Moderate
Profilin	Sin a 4	E6Y2M0	2	49	Low

Wheat: a total of 53 papers were scored, and 15 allergens were identified (Table 15). A number belong the seed storage prolamins which form the gluten fraction of wheat. The most significant of these was the ω-5 gliadin, Tri a 19, which was identified in 31 papers with 1862 participants across 31 study centres. Two other monomeric gliadin types were identified, the γ-gliadin allergen, Tri a 20, which was identified in 12 papers with 471 participants across 12 study centres and the α-gliadin, Tri a 21, which was identified in 10 papers with 186 participants across 10 study centres. Of the polymeric seed storage prolamins, a low molecular weight (LMW) subunit of glutenin, Tri a 36, was identified in 13 papers with 685participants across 9 study centres, together with the high molecular weight (HMW) subunit of glutenin, Tri a 26, which was identified in 9 papers with 503participants across 12 study centres.

Members of the prolamin superfamily were also identified including the non-specific lipid transfer protein 1 (nsLTP-1), Tri a 14, which was identified as an allergen in two papers with 103 participants across two study centres. Several α-amylase inhibitors were also identified including the 0.28 (Tri a 15) inhibitor which was reported in only two papers with four participants across four study centres and the 0.19 inhibitor, Tri a 28, which was reported in two papers with 127 participants across two study centres. Lastly the chloroform-methanol soluble inhibitors known as CM1/CM2 (Tri a 29) and CM3 (Tri a 30) were identified in only a single paper with 110 participants (Tri a 29) and 22 participants (Tri a 30) across three study centres. CM17 (Tri a 40) was also identified as an allergen in three papers with 156 participants across five study centres.

Lastly, several proteins with metabolic function have been identified as allergens in wheat including the Thiol Reductase homologue, Tri a 27, which was identified in two papers with 127 participants across two study centres. A serpin, Tri a 33, was also identified as an allergen

in one paper with 22 participants from only a single 1 study centre, β-amylase also being identified as an allergen (Tri a 17) in only one paper with 110 participants from a single study centre. Lastly, α-purothionin was also identified as an allergen (Tri a 37) in only two papers with 113 participants across two study centres.

Table 15: Summary of wheat allergen quality assessment and ranking See Suppl Table S15 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Profilin	Tri a 12	CAA61945	1	110	Low
nsLTP	Tri a 14	P24296, NLTP1	5	167	High
α-amylase inhibitor	Tri a 15	P01083, AJ223492.1	4	58	Low
β-amylase	Tri a 17	AAP80614, P93594, AMYB	3	173	Low
ω5- gliadin	Tri a 19	Q9FUW7	30	1711	High
γ-gliadin	Tri a 20	P08453, AF234643	12	471	High
α-gliadin	Tri a 21	P04725, Q9M4M6	6	186	Moderate
HMW subunit of glutenin	Tri a 26	ABF14401	9	503	High
27K protein	Tri a 27	BAC76688	1	110	Low
α-amylase inhibitor	Tri a 28	P01084, IAA1	2	58	Low
α-Amylase/trypsin inhibitor CM1	Tri a 29	IAAC1	1	41	Low
Amylase/trypsin inhibitor CM2	Tri a 29	P16850, P16851, P17314, P16159, Q43723	3	64	Low
α-amylase inhibitor CM3	Tri a 30	P17314, IAAC3, AY436554.1	4	181	Low

Serpin	Tri a 33	CAA72273, SPZ1A/B/C & SPZ2A/B	1	151	Low
LMW subunit of glutenin	Tri a 36	JF776367	13	685	High
α-purothionin	Tri a 37	AFQ60540	2	113	High
a-amylase inhibitor CM17	Tri a 40	P16159	6	356	Moderate
0.19 Dimeric a-amylase inhibitor	Tri a 28	AAV39515, Q5UH6	2	151	Low
CM 17 protein precursor	Tri a 40.0101	CAA42453, Q41540	2	151	Low
a-Amylase inhibitor CIII	Tri a 15	0810252A	1	110	Low
Trypsin/a-amylase inhibitor CMX1/CMX3	Not included ; author named as Tri a CMX	IACX1	1	41	Low
Endogenous a-amylase/subtilisin inhibitor	Not included ; author named as Tri a aA_SI	IAAS	1	41	Low
26 kDa endochitinase 1	Not included	CHI1	1	41	Low
Class II chitinase	Not included	Q4Z8L8	1	41	Low
Wheatwin-1	Not included	WHW1	1	41	Low
Wheatwin-2	Not included	WHW2	1	41	Low
Thaumatin-like protein	Not included	Q8S4P7	1	41	Low
b-D-Glucan exohydrolase	Not included	AAM13694	1	110	Low
Peroxidase	Not included ; author named as Tri a Peroxidase 1	AAM88383, Q8LK23	2	151	Low
Tritin (rRNA N-glycosidase)	Not included ; author named as Tri a Tritin	Q07810	1	41	Low

Xylanase inhibitor protein 1	Not included ; author named as Tri a XI	XIP1	1	41	Low

Buckwheat: a total of 10 papers were included in the study and 10 allergens were ranked. (Table 16). Neither the sequences nor the accession numbers for the α-harpin allergen, Fag e 3, and the hevein-like antimicrobial peptide, Fag e 4, were available in the papers included in this study, and therefore these allergens were not scored. However, sequence accessions for both allergens are listed in the relevant IUIS allergen nomenclature database entries. Moreover, the sequences for the allergens Fag e 19kDa and Fag e 9kDa, for which sufficient data were available for quality ranking, were not available in the IUIS nomenclature database. Since the sequences provided in the papers were partial they were not included in subsequent bioinformatic analysis. Other sequences were included in this study but at present date are not available on WHO/IUIS Allergen Nomenclature dataset: Fag t 1 (A9NJG2), Fag e 1 (D87980 and D87982), Fag e 10kD (Q8W3Y9).

Table 16: Summary of peach allergen quality assessment and ranking.
See Suppl Table S16 for references used in the clinical relevance ranking.

Protein name	**IUIS allergen designation**	**Sequence accession(s)**	**No of included studies**	**Total no of patients studied**	**Overall clinical relevance ranking**
Tartary buckwheat (*Fagopyrum tataricum*)					
13S seed storage protein	Not in IUIS but assigned Fag t 1 in Allergome	A9NJG2	1	4	Low
2S albumin	Fag t 2	E9NX73	1	3	Low
Common buckwheat (*Fagopyrum esculentum*)					
13S globulin seed storage protein 1	Not in IUIS but been assigned Fag e 1 by authors	D87980 D87982	1	72	Low
2S albumin	Fag e 2	AY966013	2	62	Moderate

Vicillin-like protein	Fag e 5	Q6QJL1	1	52	Low
BW10KD allergen protein	Not in IUIS but assigned Fag e 10kD by authors	Q8W3Y9	1	16	Low
	Not in IUIS but assigned Fag e 19kDa by authors	Sequence: GDYPLEXCRQKIEH	2	34	Low
	Not in IUIS but assigned Fag e 16kDa by authors	Sequence: RDEGFDLGETQMSSKCMRQVKMNEP	1	63	Low
	Not in IUIS but assigned Fag e 9kDa by authors	Sequence: SDKPQQLLEECRYLXRI	1	34	Low

3.1.5.4 VEGETABLES

Celery: a total of 15 papers were scored and 6 allergens were identified, with the PR10 homologue, known as Api g 1, being the most clinically relevant (Table 17). One allergen, termed Api g 3 by IUIS is a chloroplast chlorophyll a-b binding protein, was not ranked. The paper that is cited in IUIS for this allergen ["Characterization of a chlorophyll ab binding protein from celery as a food allergen Api g 3." Submitted to EMBL/GenBank/DDBJ] was apparently never published and no other supporting publication was identified in the systematic review.

Table 17: Summary of celery allergen quality assessment and ranking
See Suppl Table S17 for references used in the clinical relevance ranking.

Protein type	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
PR10	Api g 1	P49372	8	72	High
nsLTP	Api g 2	E6Y8S8	1	62	Low
Profilin	Api g 4	Q9XF37	2	52	Low
FAD-containing oxidase	Api g 5	P81943	3	16	Low
nsLTP2	Api g 6	P86809	1	34	Low
Defensin-like protein	Api g 7	QUJ17885.1	1	63	Low

Carrot: a total of 7 papers were included in the study and 10 allergens were ranked (Table 18) with again the PR10 homologue, Dau c 1, being one of the most clinically-relevant allergens identified in this food species. Three other allergen sequences were identified In the included studies (Dau c IFR1, Dau c IFR2, Dau c Cyc) but neither the sequence accession number nor the sequence itself were available in the publications.

Table 18: Summary of carrot allergen quality assessment and ranking
See Suppl Table S18 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Relevance
PR-10, Bet v 1 family member	Dau c 1.0104	Z81362, D88388	5	319	Moderate
	Dau c 1.0201	AF456481	4	305	Moderate
	Dau c 1.03	HM064421	2	71	Moderate
	Dau c 1.0501	XP_017220806.1	1	14	Low
	Dau c 1.0601	XP_017215843.1, XP_017218034.1	1	14	Low

Unknown	Dau c 1-like	XP_017220733.1	1	14	Very low
Profilin	Dau c 4	AF456482	3	275	Moderate

3.1.5.5 FRUIT

Kiwi fruit: a total of 18 papers were included in the study and 10 allergens were ranked (Table 19). In the studies, there are two different IUIS names for Actinidain, Act d 1 and Act c 1; this is because the name of the plant was originally designated as *Actinidia deliciosa* but it was later reclassified as a variety of *Actinidia chinensis*.

Act d 3 (P85063) did not contain all residues and it was therefore not possible to include in the analysis. Act d 6, 7 and 8 were cited in publications included in our study and that passed the inclusion criteria, but no sequence nor accession number was provided, therefore they were not included in the assessment of clinical relevance summarised in Table 19.

Table 19: Summary of kiwi fruit allergen quality assessment and ranking.
See Suppl Table S19 for references used in the clinical relevance ranking.

*WHO/IUIS Allergen Nomenclature database indicates "P00785" as accession number for Act d 1, although the sequence is Act c 1.

** The sequence contained residues that could not be assigned in the protein sequencing, and consequently it was not possible to include in the analysis because some tools examined in this work do not allow the submission of sequences with such missing residues.

Protein name	**IUIS allergen designation**	**Sequence accession(s)**	**No of included studies**	**Total no of patients studied**	**Overall clinical relevance ranking**
Actinidia deliciosa					
Cysteine chinase	Act d 1	P00785*	10	1048	Moderate
Thaumatin-like protein	Act d 2	Q5ND92	8	929	Moderate
unknown	Act d 3	P85063**	1	30	Low
Pythocystatin	Act d 4	AAR92223	2	274	Moderate
Kiwellin	Act d 5	P84527	8	664	Moderate
Bet v 1 homologue	Act d 8	AM489568	7	927	Low
Profilin	Act d 9	C0HL99	2	578	Low
nsLTP	Act d 10	P85205	1	74	Low
Bet v 1 homologue	Act d 11	FG437290	3	539	Moderate

11S	Act d 12	C0HJF9	2	60	High
2S albumin	Act d 13	C0HJG0	2	55	Low
Actinidia chinensis					
Actinidain	Act c 1	P00785	1	37	Low
nsLTP	Act c 10	P85204	1	7	Low

Apple: a total of 15 papers were included in the study and three allergens were ranked (Table 20). Mal d 3 was cited in publications included in the study and passed the inclusion criteria with high clinical relevance. However, no sequence nor accession number was provided in the included papers and therefore it was not included in the main ranking Table below.

Table 20: Summary of apple allergen quality assessment and ranking
See Suppl Table S20 for references used in the clinical relevance ranking.

Protein name	**IUIS allergen designation**	**Sequence accession(s)**	**No of included studies**	**Total no of patients studied**	**Overall clinical relevance ranking**
Bet v 1 homologue	Mal d 1	AJ417551	8	602	High
Thaumatin-like protein	Mal d 2	Q9FSG7	5	508	High
nsLTP 1	Mal d 3	AY374225	7	578	High
Profilin	Mal d 4	AJ507457	4	141	Moderate

Melon: five papers were included in the study and three allergens were ranked (Table 21).

Table 21: Summary of melon allergen quality assessment and ranking
See Suppl Table S21 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Cucumisin	Cuc m 1	Q39547	2	57	Moderate
Thaumatin-like protein precursor	Unnamed	38606865	1	1	Very Low
Profilin	Cuc m 2	AJ565931	2	91	Moderate

Peach: 19 papers were included in the study and five allergens were ranked.

Table 22: Summary of peach allergen quality assessment and ranking
See Suppl Table S22 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Bet v 1 homologue	Pru p 1	Q2I6V8	9	698	High
Thaumatin-like protein	Pru p 2	B6CQT7	1	27	Low
nsLTP	Pru p 3	P81402	17	1121	Moderate
Profilin	Pru p 4	Q8GT40	7	690	Moderate
Gibberellin-regulated protein (PF02704)	Pru p 7	P86888	6	540	Moderate

Tomato: seven papers were included in the study and eight allergens were ranked (Table 23) and all but the LTP were of low or very low clinical relevance, with only the LTP allergen, (Sola a 1). Other sequences were also identified as potential allergens but the publications where they were identified did not cite sequence accession numbers or provide the actual sequence. These were the legumin allergen which was cited with a sequence accession TC165005, was not found. Neither the sequence for superoxide dismutase (SOD) nor any accession number was included in the papers that were graded.

Table 23: Summary of tomato allergen quality assessment and ranking
See Suppl Table S23 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Relevance
Beta-fructosidase	Sola l 2	D11350	3	63	Very low
Pectinesterase (PE)	None	P14280	2	15	Very low
Polygalacturonase 2A (PG2A) Sola l PG (Allergome	None	P05117	2	15	Very low
Vicilin	None	AM932874	1	19	Low
nsLTP1	Sola l 3	P93224, P27056	2	62	Low/Moderate
	Sola l 7	P86417	2	22	Low/Moderate
PR10	Sola l 4	KF682291, KF682292	1	68	Low
nsLTP2	Sola l 6	XP_004229753.1	1	62	Low/Moderate

Banana: a total of five papers were included in the study and five allergens were ranked with only one, Mus a 5, being identified as being of moderate clinical relevance.

Table 24: Summary of banana allergen quality assessment and ranking
See Suppl Table S24 for references used in the clinical relevance ranking.

Protein name	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Relevance
Beta 1,3-glucanase	Mus a 5	ADG36438, AAB82772, 83754908, GQ268963	3	84	Very low
Catalase	None	ABV55108.1	1	19	Low
Thaumatin like protein	Mus a 4	gl 88191901	1	51	Moderate
Canary banana					
Unknown	Ba1	EQCGRQAGGALCPGGLCCSQYG	1	15	Low
Unknown	Ba2	EQCGRQAGGALCPGGLCCSQFG	1	15	Low

3.1.6 (2) Animal derived foods

Cow's milk: a total of 60 papers were scored, and six allergens were identified (Table 25) as follows. α-lactalbumin (Bos d 4) was identified in 18 papers with 6339participants. β-lactoglobulin (Bos d 5) was identified in 24 papers with 5901participants. Caseins, identified as a group on WHO/IUIS Allergen Nomenclature database as Bos d 8, were not assigned a score but were scored individually since a key premise of the ranking is to link clinical relevance with specific allergen sequences. α-S1-casein (Bos d 9) was identified in 33 papers with 9605 participants whilst α-S2-casein (Bos d 10) was identified in 26 papers with 9138participants across multiple study centres. β-Casein (Bos d 11) was identified in 30 papers with 9175participants and lastly, κ-casein (Bos d 12) was identified in 26papers with 9014 participants. Based on this analysis the cow's milk allergens Bos d 4, 5, 9 and 11 were therefore classified as being of high clinical relevance, while Bos d 10 and 12 were classified as being of moderate clinical relevance.

Table 25: Summary cow's milk allergen quality assessment and ranking
See Suppl Table S25 for references used in the clinical relevance ranking

Protein type	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
α-lactalbumin	Bos d 4	P00711	18	6339	High
β-lactoglobulin	Bos d 5	P02754	24	5901	High
αS1-casein	Bos d 9	P02662	33	9605	High
αS2-casein	Bos d 10	P02663	26	9138	Moderate
β-casein	Bos d 11	P02666	30	9175	High
κ-casein	Bos d 12	P02668	26	9014	Moderate

Hen's Egg: a total of 37 papers were scored, and six allergens were identified (Table 26). Ovomucoid (Gal d 1) was identified in 30 papers with 7,631participants across 21 study centres whilst ovalbumin (Gal d 2) was identified in 15 papers with 1,1539 participants. Ovotransferrin (Gal d 3) was identified in 7 papers with 668 participants and lysozyme C (Gal d 4) was identified in 5 papers with 668participants. Serum albumin (Gal d 5) was identified as an allergen in five papers with 387 participants whilst Vitellogenin (Gal d 6) was identified in only two papers with 311 participants across two study centres. On the basis of the analysis, Gal d 1 and 2 were classified as being of high clinical relevance, while Gal d 3 and 4 were classified as being of moderate clinical relevance. Gal d 5 and 6 were found to be of low clinical relevance.

Table 26: Summary of egg allergen quality assessment and ranking
See Suppl Table S26 for references used in the clinical relevance ranking.

Allergen type	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Ovomucoid	Gal d 1	P01005	30	7631	Very high
Ovalbumin	Gal d 2	P01012	15	1539	Very high
Ovotransferrin	Gal d 3	P02789	7	796	Moderate
Lysozyme C	Gal d 4	P00698	5	668	Moderate
Serum albumin	Gal d 5	P19121	5	387	Low
Vitellogenin-1	Gal d 6	P87498	2	311	Low

Fish: a total of 39 papers were scored and 45 allergenic sequences were identified (Table 27) and whilst the clinical relevance of β-parvalbumin for any given fish species was often only identified as being of moderate clinical relevanceit was identified in 13 species.

Table 27: Summary of fish allergen quality assessment and ranking. Phylogenetic attributions are as in Fishbase (https://www.fishbase.se/home.php) - See Suppl Table S27 for references used in the clinical relevance ranking.

Fish Species	Protein type	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
Carangaria						
Lates calcarifer (Barramundi)	β-parvalbumin	Lat c 1	KF021278.1	2	45	Moderate
Lates niloticus (Nile carp)	Phosphoglucomutase-1	None	41056111 (proteomic analysis returned Danio rerio (zebra fish) as identification)	1	12	Very Low
	Enolase 3	None	47551317 (proteomic analysis returned Danio rerio (zebra fish) as identification)	1	12	Low
	Creatine kinase isoform 2	None	21694043 (proteomic analysis returned Oreochromis	1	12	Very low

			mossambicus (Tilapia) as identification)			
	Fructose-bisphosphate aldolase A	None	225717412 (proteomic analysis returned Esox Lucius (Northern pike) as identification)	1	12	Low
	Apolipoprotein	None	222354841 (proteomic analysis returned Epinephelus coioides)	1	12	Low
Carrangiformes (Jacks)						
Xiphias gladius (Swordfish)	β-parvalbumin	Xip g 1	FM202668	1	16	Low
Clupeiformes (Herrings)						
Sardinops sagax	β-parvalbumin	Sar sa 1	FM177701	1	10	Low
Stegastes partitus	β-parvalbumin		657565876	1	25	Low
Cyrpriniformes (Carps)						

Cyprinus carpio (carp)		Cyp c 1	Q8UUS3	6	144	High
Gadiformes (Cods)						
Gadus callarias (Baltic cod)	β-parvalbumin	Gad c 1	P02622	2	117	High
Gadus morhua (Atlantic cod)	β-parvalbumin	Gad m 1	14531014	8	213	High
	Aldolase	Gad m 3	P86980	1	62	Moderate
	Enolase	Gad m 4	B63A0L6	1	62	Moderate
	Tropomyosin	None	27127288 (proteomic analysis returned Gadus chalcogrammus)	1	12	Very Low
	Myosin light chain	None	7678762	1	12	Very Low
	Adenylate kinase	None	222088001	1	12	Low
	Creatine kinase muscle isoform 2	None	31322099 (proteomic analysis returned Chaenocephalus kiaceratus)	1	12	Low

	Nucleoside diphosphate kinase B	None	158705974 (proteomic analysis returned Merluccius merluccius)	1	12	Low
	α-enolase	None	213514064 (proteomic analysis returned Salmon salar)	1	12	Very Low
Theragra chalcogramma (Pollack)	β-parvalbumin	The ch 1	AY035587	1	6	Very Low
Pleuronectiformes (Flat fish)						
Lepidorhombus whiffiagonis	β-parvalbumin	Lep w 1	AM9046811	2	22	Low
Salmoniformis (salmons)						
Salmo salar	β-parvalbumin	Sal s 1	Q91482	4	129	High
	β-enolase	Sal s 2	B5DGQ7	3	117	Moderate
	Aldolase	Sal s 3	B5DGM7	1	62	Moderate
	Tropomyosin	Sal s 4	NP_001117128.1	1	43	Very Low

	Creatine kinase	Sal s 7	185133138	1	55	Low
	Triose phosphate isomerase	Sal s 8	ACM09737.1	1	43	Low
	Vitellogenin (from salmon roe)		AB474573	1	5	Very Low
Scrombiformes (Mackerels)						
Thunnus albacares (Yellowfin tuna)	β-parvalbumin	Thu a 1	C6GKU3	1	62	Moderate
	Aldolase	Thu a 3	P86979	1	62	Moderate
	β-enolase	Thu a 4	P86978	1	62	Moderate
Scomber japonicus (Pacific mackerel)	β-parvalbumin	Sco j 1	P59747	3	54	Moderate
	Creatine kinase	Pan h 7	XP_026780620.1	1	77	Low
	Triose phosphate isomerase	Pan h 8	XP_026795867.1	1	77	Moderate
	Pyruvate kinase	Pan h 9	XP_026775867.1	1	77	Very Low

	Lactate dehydrogenase	Pan h 10	XP_026774991.1	1	77	Low
	Glucose-6--phosphate isomerase	Pan h 11	XP_026782721.1	1	77	Very Low
	Glyceraldehyde phosphate dehydrogenase	Pan h 13	XP_026782131.1	1	77	Very Low
Rajiformes						
Dasyatis akajei (red stingray)	β-Parvalbumin	None	N-terminal sequence of natural purified protein returned P02630 from Raja clavata	1	15	Low

Crustacean shellfish: 43 papers were scored, and 36 allergens were identified, which were predominantly tropomyosins (Table 28). The tropomyosin Pen j 1 (GI125995159) from *Penaeus japonicus* was not included in the IUIS allergen nomenclature database but was included in the AllergenOnline database. The myosin light chain, Pro c 5, and the triosephosphate isomerase allergen Pro c 8 from *Procambarus clarkii* (Red swamp crayfish), were listed in the WHO/IUIS Allergen Nomenclature database but were not considered in this report as the associated publications did not pass the PO and PECO analysis.

Molluscan shellfish: a total of 10 papers were scored, and 8 allergens were identified (Table 29).

Table 28: Summary of Crustacean shellfish allergen quality assessment and ranking. Organisms are ordered based on their phylogenetic relationships - See Suppl Table S28 for references used in the clinical relevance ranking.

Food species	Protein type	IUIS allergen designation	Sequence accession(s)	No of included studies	Total no of patients studied	Overall clinical relevance ranking
DECAPODA						
Dendrobranchiata - Penaeidae						
Penaeus aztecus	Tropomyosin	Pen a 1	A1KYZ2.1	9	169	High
	Sarcoplasmic Ca2+ binding protein	None	ROT72773.1	1	15	Low
	Myosin light chain	None	ROT76584.1	1	15	Low
Penaeus monodon	Tropomyosin	Pen m 1	A1KYZ2.1	2	23	Moderate
	Sarcoplasmic Ca2+ binding protein	Pen m 4	XP_037789865.1	2	31	Moderate
	Myosin light chain	Pen m 3	XP_037801777.1	1	15	Low
Penaeus indicus	Tropomyosin	Pen i 1	*Not available*	1	4	Low

Penaeus japonicus	Tropomyosin	Pen j 1	GI125995159	1	17	Low
Litopeneus vannamei	Tropomyosin	Lit v 1	B4YAH6	2	90	High
	Arginine kinase	Lit v 2	Q004B5	2	90	High
	Myosin light chain	Lit v 3	EU449515	3	128	Moderate
	Sarcoplasmic Ca2+ binding protein	Lit v 4	FJ184279	3	90	High
Metapenaeus ensis	Tropomyosin	Met e 1	UO8008	2	9	Low
Pleocyemata – Caridea						
Crangon crangon	Tropomyosin	Cra c 1	FJ457621	1	56	Moderate
	Arginine kinase	Cra c 2	FJ457622	1	56	Moderate
	Myosin light chain	Cra c 4	FJ462739	1	56	Moderate
	Sarcoplasmic Ca binding protein	Cra c 5	FJ462737	1	56	Moderate
	Troponin C	Cra c 6	FJ462740	1	56	Moderate

	Triose phosphate isomerase	Cra c 8	FJ462738	1	56	Moderate
Pandalus borealis	Tropomyosin	Pan b 1	E5BBS3	1	6	Weak
Macrobrachium rosenbergii	Tropomyosin	rMac r 1.0101	D3XNR9	2	23	Moderate
	Hemocyanin	None	Proteomic analysis identified haemocynin with peptides matching against sequences form several different crustacean species	1	13	Low
Pleocyemata - Cambaridae (Crayfish)						
Procambarus clarkia	Tropomyosin	Pro c 1	C0LU07	1	17	Low
	Arginine kinase	Pro c 2	H6VGI2	1	17	Low
Pleocyemata – Brachyura (Crab)						
Charybdis feriata	Tropomyosin	Cha f 1	AF 061783	2	60	Moderate
	Arginine kinase	None	gi|25453078	1	50	Moderate

Portunus pelagicus	Tropomyosin	Pro p 1	gi|119674937, M1H607	1	30	Low
	Arginine kinase	None	gi|25453074	1	30	Low
Scylla paramamosain	Tropomyosin	Scy p 1	A7L5V2	1	24	Moderate
	Myosin light chain	Scy p 3	A0A514C9K9	1	24	Moderate
Scylla olivacea	Tropomyosin	Scy o 1	QHW05411	1	22	Moderate
Scylla serrata	Tropomyosin	None	ABS12233.1	2	32	Moderate
Pleocyemata – Anomura (Hermit crab)						
Paralithodes camtschaticus						
Pleocyemata – Astacidea (Lobster)						
Homarus americanus	Tropomyosin	Hom am 1	O44119-1	1	10	Moderate
Cherax quadricarinatus	Tropomyosin		MZ217128	1	24	Moderate
	Haemocyanin		XP_053627537	1	24	Low
	Arginine kinase	None	XP_053645744.1	1	24	Low

EUPHAUSIACEA (Krill)						
Euphausia superba (krill)	Tropomyosin	Unnamed	gi156712752	2	8	Low

Table 29: Summary of Molluscan shellfish allergen quality assessment and ranking. Organisms are ordered based on their phylogenetic relationships - See Suppl Table S29 for references used in the clinical relevance ranking.

Food species	Protein type	IUIS allergen designation	Sequence accession(s)	No of included studies	No of patients	Overall clinical relevance ranking
GASTROPODA						
Architaenioglossa - Ampullariidae						
Pila polita	Tropomyosin	None	P42636 (proteomics analysis returned *Biomphalaria glabrata*)	1	25	Moderate
	Actin	None	2289975 and Q2LDZ7 (proteomics analysis returned *Heliocidari stuberculata* and *Hirudo medicinalis* respectively)	1	25	Moderate
Cornu aspersum (*Helix aspersa*)	Tropomyosin	Hel asp 1.0101	O97192	1	20	Moderate

BIVALVIA						
Ostreida						
Alectryonella plicatula	Tropomyosin	None	UTA91547.1	1	19	Moderate
Perna canaliculus	Tropomyosin	None	Q9GZ70	1	54	Moderate
	Paramyosin	None	O96064	1	54	Moderate
	Actin	None	Q26065	1	54	Moderate
Crassostrea gigas	Tropomyosin	Cra g 1	AF239173	1	15	Low
Crassostrea angulata	Tropomyosin	Cra a 1	UST29548	1	13	Low
CEPHALOPODA						
Oegopsida and **Octopoda**						
Todarodes pacificus	Tropomyosin	Tod p 1	BAE54431.1	2	9	Moderate
Octopus fangsiao	Arginine kinase	Amp f	JN127374	2	21	Moderate

A final aspect of the review of allergenic molecules was to identify so called "hypoallergens", which are rarely, if at all, ever identified as being able to cause IgE-mediated allergies. The first step in this process was to identify widely consumed foods for which the prevalence of IgE-mediated food allergies is very low.

Two foods were identified, rice and maize. The pan-European EuroPrevall study included maize (corn) as one of its priority foods and showed the prevalence of probable IgE-mediated food allergy varied from 0.05% [0.02-0.48 95%CI] to 0.00% [0.00-0.21 95%CI] in adults (Lyons *et al.*, 2019) and 0.18% (0.02-1.15 95%CI) to 0.00% (0.00-0.35 95%CI) in school age children (Lyons *et al.*, 2020). In the iFAAM school age follow-up of the EuroPrevall birth cohort of 6,069 participants only 5 reported adverse reactions to rice and two to corn, indicating the prevalence of confirmed IgE-mediated food allergy to these foods is very rare (Grabenhenrich *et al.*, 2020). Reflecting this low prevalence there are only a few well characterised allergens that have been identified in these foods (Table 30).

Table 30: Ranking of allergens from "low" allergenic comparators

Food species	**IUIS allergen designation/ protein type**	**Sequence accession(s)**	**No of studies**	**No of patients**	**Overall clinical relevance ranking**	**References**
Maize (*Zea mays*)	Zea m 14.0101 Zea m 14.0102 LTP	P19656-1 P19656-2	2	22	Low	(Pastorello *et al.*, 2009, Pastorello *et al.*, 2000)
	Zea m 8 Endochitinase A, B	P29022	1	7	Very low	(Pastorello *et al.*, 2009)
Rice (*Oryza sativa*)	None LTP		1	3	Very low	(Asero *et al.*, 2007)

Based on these data it was decided to include rice and maize proteins as comparators of very low allergenicity.

The other proteins included were human versions of allergenic proteins such as tropomyosins and parvalbumins. In healthy individuals, the immune system is able to discriminate between self and non-self and does not mount immune responses, including humoral responses, to self-proteins. Classically, the pan-allergens family of tropomyosins have been considered to be exclusively from invertebrates, explaining, in part, why humans are able to mount IgE responses to crustacean shellfish and insect tropomyosins without developing an autoimmune reaction (Jenkins *et al.*, 2007). There is a report of vertebrate tropomyosin from the fish *Oreochromis mossambicus* (Tilapia) being allergenic (Liu *et al.*, 2013). Interestingly, six of

the ten Tilapia allergic patients had been diagnosed with inflammatory bowel disease which has been lined to autoimmune reactions underlying ulcerative colitis and primary sclerosing cholangitis (Sakamaki *et al.*, 2000). Similarly, there is a difference identified between allergenic β-parvalbumins in fish and non-allergenic α-parvalbumins (Jenkins *et al.*, 2007). One notable exception to this is the allergenic α-parvalbumin found in edible frog (Hilger *et al.*, 2002). Very low levels of specific IgE have also been identified towards the α-parvalbumin of salmon which were capable of stimulating effector cell activation although its clinical relevance is likely to be very low, especially compared to the β-parvalbumin (Kalic *et al.*, 2019). On this basis, human homologues of the panallergens tropomyosin and β-parvalbumin were selected to represent another type of "hypoallergen".

3.1.7 SubTask 1.5 Identification and screening of *in silico* tools

Several tools were identified that employed different types of bioinformatics approaches to allergenicity prediction (Table 31) for use in subsequent analysis. Certain tools were no longer available.

Table 31: Bioinformatic tools and platforms for allergenicity prediction

Tool Name	Functionality	Reference
Classical sequence identity searching tools		
Allermatch	An online tool which undertakes FASTA analysis according to the CODEX recommended approaches using a custom sequence database which comprises protein sequences annotated as allergens from The UniProt Protein Knowledgebase The list of allergen nomenclature of the joint World Health Organization and International Union of Immunological Societies The Comprehensive Protein Allergen Resource (COMPARE) Retrieved sequences are compiled in the AllergenDB_original_sequences database which are then automatically processed to remove signal- and pro-peptides. Allermatchtm database as of 04-NOV-2022 contained 2569 polypeptide sequences (2255 UniProt ids, 29 UniProt ids with multiple polypeptide chains, and 314 GenBank RefSeqProtein ids). It is hosted by WFSR (Wageningen University and Research, and Bioscience - Wageningen University and Research) and is regularly updated (last update 05-SEP-2022).	(Fiers *et al.*, 2004) https://www.allermatch.org/index.html
Allergenonline	Undertakes FASTA sequence alignment like the AllermatchTM tool but using the curated AllergenOnline database which is peer reviewed. It is hosted by the Food Allergy Research and Resource Programme at the University of Nebraska at Lincoln, USA.	(Goodman *et al.*, 2016) http://www.allergenonline.org/
COMPASS and COMPARE	COMPASS (COMPare Analysis of Sequences with Software) which undertakes FASTA sequence alignment as AllermatchTM using the COMPARE (Comprehensive Protein Allergen Resource) database. The allergen sequences included in the database undergo a review process prior to inclusion. It is supported through a collaborative effort of the	(van Ree *et al.*, 2021) https://comparedatabase.org/

Tool Name	Functionality	Reference
	HESI (Health and Environmental Sciences Institute) Protein Allergens, Toxins, and Bioinformatics (PATB) Committee.	
AlgPred 2.0	A custom database comprising allergenic and non-allergenic sequences is taken from COMPARE, Allergen Online, AlgPred, AllerTop and Swiss-Prot. It contains 10075 allergens and 10075 non-allergen sequences which were used to develop and validate the machine learning approaches developed for analysis. Sequences can be queries using BLAST or two different machine learning algorithms. It also allows motif and epitope searching using different methods. It is hosted by the Institute of Microbial Technology, Okhla Phase 3, New Delhi, India and is supported by Council of Scientific and Industrial Research (CSIR) and Department of Biotechnology (DBT), Government of India.	(Sharma *et al.*, 2021b) https://webs.iiitd.edu.in/raghava/algpred2/
AllerTOP	The method uses Auto and Cross-Covariance (ACC) analysis where the properties of each amino acid is represented by five E descriptors (Venkatarajan and Braun, 2001) which spans amino acid hydrophobicity, molecular size, helix-forming propensity, relative abundance of amino acids, and β-strand forming propensity. The proteins are then classified by k-nearest neighbor algorithm (kNN,k=1) based on a training set containing 2427 known allergens from different species and 2427 non-allergens. The tool is hosted by Department of Chemistry, Faculty of Pharmacy at the Medical University of Sofia, Bulgaria.	(Dimitrov *et al.*, 2014a) https://www.ddg-pharmfac.net/AllerTOP.
AllergenFP	This algorithm builds on AllerTOP and transforms subsets of allergens and non-allergens into matrices to give each protein a unique binary fingerprint. Tanimoto coefficients are calculated for all protein pairs in the set and used to classify the protein sequences. The tool is hosted by Department of Chemistry, Faculty of Pharmacy at the Medical University of Sofia, Bulgaria.	(Dimitrov *et al.*, 2014b) https://ddg-pharmfac.net/AllergenFP/

Tool Name	Functionality	Reference
AllerCatPro,	AllerCatPro 2.0 uses a manually curated sequence databases comprising 218 allergens and 212 (likely) non-allergens) and a 3D/model database comprising 714 allergens. The tool then first identifies coeliac toxic motifs from the allergen online database. Sequences are subject to a BLASTP search and those identified with positive hits are then evaluated against our 3D model/structure database of known allergens (E-value < 0.001) using a surface homology analysis to assess the similarity of the 3D surface epitope Outcomes are that a proteins is a potential allergens with 'strong evidence' if the sequence identity of 3D surface epitope is >93% (this cutoff is 92% if Gluten-like Q-repeats are found in the query sequence). Otherwise, they are classified as having 'weak evidence' of being an allergen. If no structural hits are identified the linear-window approach is then applied as per the CODEX recommendations. If no hit is found, a hexamer hit approach is taken. The tool is hosted by the Agency for Science, Technology and Research, Singapore (A*STAR).	(Nguyen *et al.*, 2022) https://allercatpro.bii.a-star.edu.sg/
Aller-Hunter	Combines BLAST and a MV approach.	(Muh *et al.*, 2009) No longer available

3.2 Task 2: Investigation of potential in silico tools and follow up actions (*in vitro* and/or *in vivo* methods)

3.2.1 SubTask 2.1: Selection of the most appropriate ranking strategy

AllergenOnline and COMPARE/COMPASS (Goodman *et al.*, 2016, van Ree *et al.*, 2021) allow the submission of one query sequence at a time. Both tools allow the query sequence to be aligned with sequences already present in their databases using either full FASTA method or sliding 80mer window. AllergenOnline database, consisting of 2290 sequences (last version from 25th May 2023), is fully accessible and downloadable in different formats (csv, pdf). COMPARE database is at its 8th iteration and was released on 25th January 2024; it consists of 2748 sequences, and it is fully accessible and downloadable in different formats (csv, pdf, text file) as well. The output of these tools consists of the sequence that were aligned to the query one, as well as percentages of identity and similarity, and e values of the alignments.

AlgPred 2.0 (Sharma *et al.*, 2021a) allows the submission of at least 100 sequences at a time. The tool has 5 different pathways: prediction, IgE epitope mapping, motif scan, BLAST search and design. The "prediction" pathway is design to predict allergenicity of the query sequences using 2 different approaches: an AAC-based Random Forest algorithm and a Machine learning hybrid approach (RF+BLAST+MERCI) respectively. The output is the prediction of allergenicity in binary form (allergen or nonallergen) as well as machine learning scores, depending on the chosen machine learning model. The "IgE epitope mapping" and the "motif scan" pathways scan or map IgE motifs in the query sequence. The former maps IgE epitopes extracted from Immune Epitope Database (IEDB) in a tabular form whilst the latter has two different machine learning approaches (MEME/MAST and MERCI), each of which provides a tabulated output. The outputs of these two pathways are a binary outcome (hit or no hit), and the number of hits that the models found. "BLAST search" allows a BLAST search of the query sequences, and "design" allows the design of non-allergenic proteins from the query sequences, generating possible mutants' peptides with a single mutation which are further predicted using the machine learning model. BLAST search and design pathways and their outputs were not analysed for this report.

AllerTOP and AllergenFP (Dimitrov *et al.*, 2014a, Dimitrov *et al.*, 2014b) allow the submission of 1 query sequence at the time. The database consists of 2427 allergens and 2427 non-allergens and are fully accessible. The output of AllerTOP is a binary prediction of allergenicity; the output of AllergenFP is a binary prediction of allergenicity with the addition of a score, the Tanimoto score, and a link to the most similar sequence that the algorithm identified.

AllerCatPro 2.0 (Maurer-Stroh *et al.*, 2019) allows the submission of 50 sequences at a time. AllerCatPro The query proteins are checked against 714 representatives in a 3D model/structure database of known allergens as well as a dataset of proteins associated with allergenicity (4979 protein in total). The tool workflow is characterised by several decision steps to reach the output. The query proteins are initially checked for gluten-like Q repeats, then against the 3D database, and finally against the 2D database using 80mer sliding window. The outputs is a table with the result for strong, weak or no evidence for allergenicity per protein based on workflow decisions and, in case of a hit, the possibility to view the most similar allergens with detailed results for cross-reactivity, protein information (UniProt/NCBI),

functionality (Pfam, InterPro, SUPFAM), as well as clinical relevance of IgE prevalence (Allergome) and allergen information.

The tools were assessed using curated allergen sequence sets (Table 32) either available online or generated for this purpose, as well as a curated list of nonallergenic sequences from the animal and plant kingdoms.

Table 32: Specifications of the sequence sets used for assessment of bioinformatic tools

Sequence set	Species	Protein sequences by family
Cows' milk (n=87)	Cattle (*Bos taurus*)	αS1-caseins, αS2-caseins, β-caseins, κ-caseins, β-lactoglobulins, α-lactalbumins, serum albumins
Crustacean shellfish (n=57) Crustacean food allergens: n=49 There are 47 sequences retrieved from IUIS; the rest of the sequences were retrieved from AllFam and through a BLASTp search.	Shrimp: White shrimp (*Litopenaeus vannamei*), Black tiger shrimp (*Penaeus monodon*), Brine shrimp (*Artemia franciscana*), North Sea shrimp (*Crangon crangon*), Crayfish (*Archaeopotamobius sibiriensis*), Red swamp crayfish (*Procambarus clarkii*), Greasyback shrimp (*Metapenaeus ensis*), Brown shrimp (*Penaeus aztecus*), Northern red shrimp (*Pandalus borealis*), Western king prawn (*Penaeus latisulcatus* (listed as *Melicertus latisulcatus* in IUIS)), Giant river prawn (*Macrobrachium rosenbergii*), Kuruma prawn (*Penaeus japonicus*), Narrow-clawed crayfish (*Pontastacus leptodactylus*), Crab: Mud crab (*Scylla paramamosain*, *Scylla serrata*), Blue swimmer crab (*Portunus pelagicus*), Crucifix crab (*Charybdis feriatus*), Warrior swimming brown crab (*Callinectes bellicosus*). Chinese mitten crab (*Eriocheir sinensis*). Lobster: American lobster (*Homarus americanus*), Chinese spiny lobster (*Panulirus stimpsoni*), European lobster (*Homarus gammarus*).	Arginine kinases, Fatty acid binding proteins, Glycogen phosphorylase like proteins, Haemocyanins, Myosin light chain 1, Myosin light chain 2, Sarcoplasmic Calcium binding proteins, Troponin 1, Troponin C, Triose phosphate isomerases, Tropomyosins, Ovary-development related proteins, Filamin C.
Molluscan shellfish (n=10)	Portuguese oyster (*Crassostrea angulata*), Brown Garden snail (*Cornu aspersum*), Jade tiger abalone (*Haliotis laevigata* x *Haliotis rubra*), Japanese flying squid (*Todarodes pacificus*),	Argninine kinases, Paramyosins, Tropomyosins,

	Sydney rock oyster (*Saccostrea glomerata*), Veined rapa whelk (*Rapana venosa*)	Sarcoplasmic calcium binding proteins
Tropomyosins (n=16) **All sequences are in IUIS.**	Crustacean shellfish: Shrimp: Northern red shrimp (*Pandalus borealis*), North Sea shrimp (*Crangon crangon*), White shrimp (*Litopenaeus vannamei*), Giant river prawn (*Macrobrachium rosenbergii*), Western king prawn (*Penaeus latisulcatus*) Crab: Crucifix crab (*Charybdis feriatus*), Blue swimmer crab (*Portunus pelagicus).* Lobster: American lobster (*Homarus americanus*), Chinese spiny lobster (*Panulirus stimpsoni*) Molluscan shellfish: Jade tiger abalone (*Haliotis laevigata* x *Haliotis rubra*), Fish: Mozambique tilapia (*Oreochromis mossambicus*) Nematode: Herring worm (*Anisakis simplex*)	Allergenic tropomyosins: 16
Fish (n=34) **n=27 from IUIS** **The rest of the sequences were retrieved from AllFam and through a BLASTp search.**	Yellowfin tuna (*Thunnus albacares*), Atlantic cod (*Gadus morhua*), Common carp (*Cyprinus carpio*), Striped catfish (*Pangasianodon hypophthalmus*), Indian mackerel (*Rastrelliger kanagurta*), Atlantic mackerel (*Scomber scombrus*), Atlantic salmon (*Salmo salar*), Mozambique tilapia (*Oreochromis mossambicus*).	Aldolase A, Beta enolases, Beta parvalbumins, Alpha tropomyosins, Creatine kinases, Triose phosphate isomerases, Pyruvate kinase PKM like proteins, L-lactate dehydrogenases, Glucose-6-phosphate dehydrogenases, Glyceraldehyde-3-phosphate dehydrogenases, Tropomyosins, Collagen alpha, Creatine kinase
Parvalbumins (n=25) **22 parvalbumins are listed in IUIS. The rest of the sequences were**	Grass carp (*Ctenopharyngodon idella*), Northern pike (*Esox lucius*), Yellowback seabream (*Evynnis tumifrons*), Alaska pollock (*Gadus chalcogrammus*), Atlantic cod (*Gadus morhua*), Patagonian grenadier (*Macruronus magellanicus*), Austral hake (*Merluccius australis australis*), Shallow-water Cape hake (*Merluccius capensis*), European hake (*Merluccius merluccius*), Deep-water Cape hake *(Merluccius*	Beta parvalbumins: 25

retrieved from AllFam and through a BLASTp search.	*paradoxus*), Benguela hake (*Merluccius polli*), Salmon (*Salmo salar*), Brook trout (*Salvelinus fontinalis*), Chub mackerel (*Scomber japonicus*), Atlantic mackerel (*Scomber scombrus*), Japanese jack mackerel (*Trachurus japonicus*). Chicken (*Gallus gallus*).	

3.3 SubTask 2.2: Effectiveness of selected *in silico* tools

The effectiveness of *in silico* tools was assessed using allergens that were quality assessed. Results are presented on a food-by-food basis and were analysed using AllergenOnline, COMPARE, AlgPred 2.0, AllergenFP and AllerCatPro.

3.3.1 PLANT FOODS

3.3.1.1 TREE NUTS

All the sequences included were aligned and correctly identified or predicted to be allergenic by all tools analysed for Brazil nut (*Bertholletia excelsa*), Cashew (*Anacardium occidentale*) and pistachio (*Pistachia vera*).

Almond (*Prunus dulcis*): All allergen sequences identified in the systematic review were aligned and predicted correctly by all tools analysed except for the following:

The moderately clinically relevant allergen Pru du 8 (2S albumin) was predicted in AllergenFP as probable non-allergen with the closest protein (Tanimoto score of 0.81) being ATP synthase subunit a from human (*Homo sapiens*).

The moderately clinically relevant allergen Pru du 4 (profilin) identified by AllerCatPro as having weak evidence of allergenicity.

The moderately clinically relevant allergen Pru du 1 (Bet v 1 homologue) was aligned by AllergenOnline and COMPARE with the Pathogenesis-related protein, PR-10, Pru p 1.0201 from peach (Prunus persica) with 96% of identity.

The Mandelonitrile lyase Pru du 10, identified as having moderate clinical relevance, was predicted as probable non-allergen by AllergenFP. The protein with the highest Tanimoto similarity index (0.85) was the Eukaryotic translation initiation factor 4E transporter from human (*Homo sapiens*).

The allergen with low clinical relevance Pru du 5 (60S ribosomal protein) was aligned by AllergenOnline and COMPARE with several microbial 60S ribosomal proteins with 63% of identity (or lower). AllerCatPro assigned it as allergen with strong evidence of allergenicity.

The 7S protein (γ-conglutin 1), identified as having moderate clinical relevance, was identified in AllerCatPro as having weak evidence of allergenicity.

Macadamia (*Macadamia integrifolia*): The 11S globulin Mac i 2 was correctly aligned and identified in AllergenOnline but not in COMPARE, where it was aligned with a partial sequence of the same protein from the same organism with 90.9% of identity. This sequence was also

recognised as nonallergenic by AlgPred with prediction pathway, and by AllergenFP (where the closest protein, with Tanimoto score of 0.79, being ATP-binding cassette sub-family A member 6 from human (*Homo sapiens*). AllerCatPro identified the sequence as weak allergen (weak evidence of allergenicity).

English Walnut (*Juglans regia*) and pecan (*Carya illinoinensis*):The three highly clinically relevant English walnut allergens Jug r 1, 2 and 3 (respectively, 2S albumin, vicilin, and non-specific LTP) and the moderately clinically relevant allergen Jug r 4 (11S globulin) were correctly aligned and identified or predicted to be allergens by all tools analysed.

The moderately clinically relevant allergen from pecan Car i 2 (7S) and the low clinically relevant allergen Car i 4 (11S globulin) were aligned and predicted correctly by all tools analysed.

Hazelnut (*Corylus avellana*): The low clinically relevant allergen Cor a 12 (17kDa oleosin) was aligned and predicted correctly except in AllerTOP/AllergenFP, where it was predicted as probable nonallergen. The protein with the highest Tanimoto similarity index (0.90) was the ATP synthase subunit a from human (Homo sapiens).

All other sequences were aligned and predicted correctly by all tools analysed.

3.3.1.2 LEGUMES

Peanut (*Arachis hypogaea*): The oleosin Ara h 15, identified as low clinically relevant, was predicted by AllergenFP as probable nonallergen. The protein with the highest Tanimoto similarity index (0.83) was the Two pore calcium channel protein 1 from rice (*Oryza sativa* spp. *japonica*).

All other sequences were aligned and predicted correctly by all tools analysed.

Soybean (*Glycine max*): The low clinically relevant allergen Gly m BBI was identified as non-allergenic by AlgPred prediction method, AllergenFP (Tanimoto 0.81), and AllerCatPro (no evidence of allergenicity). The protein with the highest Tanimoto score was Adrenocorticotropic hormone receptor from human (*Homo sapiens*).

The defensin Gly m 2, identified as low clinically relevant, was identified as non-allergenic by AllergenFP (Tanimoto score of 0.78), with the closest protein being ATP synthase subunit a from human (*Homo sapiens*).

All other sequences were aligned and predicted correctly by all tools analysed.

Lupin (*Lupinus angustifolius* and *Lupinus albus*):

The low clinically relevant allergen Lup an 3 (nsLTP) was aligned both from AllergenOnline with a nsLTP from bean (*Phaseolus vulgaris*) and apricot (*Prunus armeniaca*) with 64% of identity, but was aligned correctly in COMPARE.

The γ-conglutin from *Lupinus angustifolius*, identified as highly clinically relevant, was not aligned in AllergenOnline at all, but aligned correctly in Compare. Moreover, AllergenFP identified this allergen as probable non-allergenic, with the closest protein (Tanimoto score of 0.82) being Ankyrin repeat and KH domain-containing protein 1 from human (*Homo sapiens*).

The γ-conglutin from *Lupinus albus*, identified as highly clinically relevant, was aligned in AllergenOnline with 36% of identity with Aca s 4 from flour mite (*Acarus siro*), but correctly aligned in COMPARE. Moreover, AllergenFP identified this allergen as probable non-allergenic, with the closest protein (Tanimoto score of 0.83) being CDKN2A-interacting protein from human (*Homo sapiens*).

The α-conglutin from *Lupinus angustifolius*, identified as moderately clinically relevant, was aligned in AllergenOnline with 56% identity with glycinin from soybean (*Glycine max*), and in COMPARE was aligned with the same protein with 54% of identity. Moreover, AllerCatPro predicted this sequence as allergenic but with weak evidence of allergenicity.

The δ-conglutin from *Lupinus angustifolius*, identified as moderately clinically relevant, was aligned in AllergenOnline with 41% of identity to Ara h 6 from peanut (*Arachis hypogaea*), and with Ara h 2 from peanut in COMPARE with the same percentage.

The 11S globulin from *Lupinus angustifolius*, identified as moderately clinically relevant, was aligned in AllergenOnline with 51% of identity with Ara h 3 from peanut (*Arachis hypogaea*), and in COMPARE with 52% of identity with the same protein.

All other sequences were aligned and predicted correctly by all tools analysed.

Lentil (*Lens culinaris*): All allergens included in the study were aligned and predicted correctly.

3.3.1.3 OTHER SEEDS

Sesame (*Sesanum indicum*): All sequences were aligned and predicted correctly by all tools analysed.

Mustard seed (*Sinapus albus*): All sequences were aligned and predicted correctly by all tools analysed.

Wheat (*Triticum aestivum*): The monomeric α-amylase inhibitor Tri a 15 was aligned in AllergenOnline and COMPARE with putative α-amylase inhibitor 0.28, partial from the same organism with 94.2% of identity.

The β-amylase Tri a 17 had no hit in AllergenOnline but aligned correctly in COMPARE.

The ω5gliadin storage protein Tri a 19 was aligned in AllergenOnline and COMPARE with a partial sequence of the same protein and same organism with 96.8% of identity.

The α-amylase inhibitor 0.53 Tri a 28 was aligned in AllergenOnline and COMPARE with Tri a 28 with 94.4% of identity.

Buckwheat: all allergens included in the study were aligned and predicted correctly, with the following exceptions:

The vicilin-like protein Fag e 5 (Q6QJL1) from common buckwheat (*Fagopyrum esculentum*) was aligned in AllergenOnline with 49.6% of identity with vicilin Ana o 1 from pistachio (*Pistachia vera*). Moreover, the sequence was predicted as non-allergenic in AllergenFP, with the closest protein (Tanimoto score of 0.84) being the Cyclin-dependent kinase 11B from human (*Homo sapiens*).

The Fag e 16kDa was predicted as non-allergen in AllergenFP, with the closest protein (Tanimoto score of 0.68) being UPF0688 protein C1orf174 from human (*Homo sapiens*). Moreover, it was aligned with 100% of identity in COMPARE with Fag e 2.

3.3.1.4 VEGETABLES

Celery (*Apium graveolens*): All sequences were aligned and predicted correctly by all tools analysed with 1 exception. The defensin Api g 7 was predicted by AllergenFP as probable non-allergens with the closest protein (Tanimoto score of 0.82) being the cation transporter HKT2;1 from Oryza sativa (rice).

Carrot (*Daus carrota*): All allergens included in the study were aligned and predicted correctly, with the following exceptions:

The low clinically relevant allergen Bet v 1 homologue Dau c 1.0501 was aligned in AllergenOnline with 58.1% of identity with PRP-like protein from the same organism. It was also recognised by AllerCatPro as allergen with weak evidence of allergenicity.

The low clinically relevant allergen Bet v 1 homologue Dau c 1.0601 (1) was aligned in AllergenOnline with 61.3% of identity with Api g 2 from *Apium graveolens* (Celery). It was also recognised by AllerCatPro as allergen with weak evidence of allergenicity.

The low clinically relevant allergen Bet v 1 homologue Dau c 1.0601 (2) was aligned in AllergenOnline with 56.5% of identity with Api g 2 from Apium graveolens (Celery) and in COMPARE with 56.5 with Api g 1 from Apium graveolens (Celery). It was also recognised by AllerCatPro as allergen with weak evidence of allergenicity.

The low clinically relevant allergen Dau c 1-like was aligned in AllergenOnline and COMPARE with 59.7% of identity with Api g 2 from *Apium graveolens* (Celery). It was also recognised by AllerCatPro as allergen with weak evidence of allergenicity.

3.3.1.5 FRUIT

Apple (*Malus domestica*): All allergens included in the study were aligned and predicted correctly.

Banana (*Musa species*): All allergens included in the study were aligned and predicted correctly, with the following exceptions:

The low clinically relevant catalase was aligned in AllergenOnline with 40.7% with catalase from a fungus (*Penicillium citrinum*). Moreover, it was predicted in AllergenFP as nonallergenic, with the closest protein (Tanimoto score of 0.84) being Catalase isozyme 1 from potato (*Solanum tuberosum*).

The low clinically relevant allergen Ba1 was aligned in AllergenOnline and COMPARE with 100% of identity with Hev b 11, chitinase from rubber tree (*Hevea brasiliensis*).

The low clinically relevant allergen Ba2 was aligned in AllergenOnline and COMPARE with 100% of identity with Pers a 1, chitinase from avocado (*Persea americana*).

The sequences Ba1 and Ba2 were not recognised as sequences by AllergenFP and AllerCatPro.

Kiwi fruit: the moderately clinically relevant allergen Act d 12 was predicted as nonallergenic in AllergenFP, with the closest protein (Tanimoto of 0.63) being Calcium homeostasis endoplasmic reticulum protein from human (*Homo sapiens*).

The moderately clinically relevant allergen Act d 13 was aligned in AllergenOnline with Sin a 1.106 from white mustard (*Sinapis alba*) with 83.3% of identity, and in COMPARE with Pru du 8 with 100% of identity. Moreover, AllerCatPro identified this sequence as nonallergenic (with no evidence of allergenicity).

The low clinically relevant allergen non-specific lipid-transfer protein 1 Act c 10 was aligned in AllergenOnline and COMPARE with Act d 10 with 94.4% of identity.

Melon (*Cucumis melo*):

The low clinically relevant allergen Thaumatin-like protein was aligned in AllergenOnline and COMPARE with 60% of identity with Mal d 2, thaumatin-like from apple (*Malus domestica*). In AllergenFP, it was predicted as nonallergenic, with the closest protein (Tanimoto score of 0.76) being UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 8 from human (*Homo sapiens*).

Peach (*Prunus persica*): All allergens included in the study were aligned and predicted correctly.

Tomato (*Solanum lycopersicum*): All allergens included in the study were aligned and predicted correctly, with the following exceptions:

The low clinically relevant allergen Pectinesterase 1 was predicted as allergen with weak evidence in AllerCatPro and was identified as Act d 7 from *Actinidia deliciosa* (kiwi fruit). Moreover, it was aligned in AllergenOnline with 30.6% of identity with Sal k 1 from Prickly saltwort (*Salsola kali*), and in COMPARE with 52% of identity with pectin methylesterase from Japanese hops (*Humulus japonicus*).

The low clinically relevant allergen Vicilin was predicted as allergen with weak evidence in AllerCatPro and was identified as Jug n 2 from *Juglans nigra* (black walnut). Moreover, it was

aligned in both AllergenOnline and COMPARE with 47% of identity with Jug r 2 from black walnut (*Juglans nigra*).

The low clinically relevant allergen β-fructosidase Sola l 2 (or Lyc e 2) was aligned both in AllergenOnline and COMPARE with 99.2% of identity with the precursor of the same protein from the same organism.

The low clinically relevant allergen Polygalacturonase 2A (PG2A) was aligned in AllergenOnline with 52.2% of identity with ripening-induced polygalacturonase 2 from papaya (*Carica papaya*).

The moderately clinically relevant allergen nsLTP1 was predicted in AllergenFP as nonallergenic, with the closest protein (Tanimoto scores of 0.58) being annexin A7 from human (*Homo sapiens*).

3.3.2 ANIMAL FOODS

3.3.2.1 COWS' MILK

All sequences were aligned and predicted correctly by all tools analysed.

3.3.2.2 HEN'S EGG

All sequences were aligned and predicted correctly by all tools analysed.

3.3.2.3 SHELLFISH

Crustacean ShellfishThe low clinically relevant allergen Myosin light chain from brown shrimp (*Penaeus aztecus*) was assigned in AllerCatPro as allergen with weak evidence of allergenicity.

The moderately clinically relevant allergen actin from New Zealand green-lipped mussel (*Perna canaliculus*) was assigned in AlgPred (prediction pathway) as nonallergenic, and in AllerCatPro as nonallergen, with no evidence of allergenicity. Moreover, it was not aligned in AllergenOnline with any sequence (no hit) and was aligned in COMPARE with the same protein from a species of flowering plant in the bean family Fabaceae (*Delonix regia*) with 96.9% of identity.

The low clinically relevant allergen sarcoplasmic calcium-binding protein beta chain from white shrimp (*Penaeus vannamei*) was aligned in AllergenOnline and COMPARE with same protein from North Sea shrimp (*Crangon crangon*) with 83% of identity.

The low clinically relevant allergen myosin light chain 2 from white shrimp (*Penaeus vannamei*) was aligned in AllergenOnline with same protein from black tiger shrimp (*Penaeus monodon*) with 62.6% of identity, same in COMPARE.

The moderately clinically relevant allergen arginine kinase from crab (*Charybdis feriata*) was aligned in both AllergenOnline and COMPARE with the same protein from green mud crab (*Scylla paramamosain*) with 96.1% of identity.

The moderately clinically relevant allergen arginine kinase from blue swimmer crab (*Portunus pelagicus*) was aligned both in AllergenOnline and COMPARE with the same protein from green mud crab (*Scylla paramamosain*) with 98.6% and 98.3% of identity respectively.

The moderately clinically relevant allergen tropomyosin from red claw crayfish (*Cherax quadricarinatus*) was aligned correctly in AllergenOnline but in COMPARE it was aligned with the same protein from American lobster (*Homarus americanus*) with 99.6% of identity.

The moderately clinically relevant allergen hemocyanin B chain-like from red claw crayfish (*Cherax quadricarinatus*) was aligned both in AllergenOnline and COMPARE with the same protein from black tiger shrimp (*Penaeus monodon*) with 64.6% of identity. Moreover, it was recognised as probable non-allergen in AllergenFP with the closest protein (Tanimoto score of 0.8) being Tyrosine-protein kinase ABL1 from human (*Homo sapiens*).

The moderately clinically relevant allergen arginine kinase Pro c 2.0101 isoform X1 from red claw crayfish (*Cherax quadricarinatus*) was aligned both in AllergenOnline and COMPARE with the same protein from red swamp crayfish (*Procambarus clarkii*) with 96.6% of identity.

The low clinically relevant allergen tropomyosin from fingerprint oyster (*Alectryonella plicatula*) was aligned both in AllergenOnline and COMPARE with the same protein from Pacific oyster (*Crassostrea gigas*) with 99.6% of identity.

The moderately clinically relevant allergen paramyosin from red claw crayfish (*Cherax quadricarinatus*) was aligned both in AllergenOnline and COMPARE with the same protein from disc abalone (*Haliotis discus discus*) with 72.1% of identity.

The low clinically relevant allergen tropomyosin from Pacific oyster (*Crassostrea gigas*) was aligned in AllergenOnline, and in COMPARE it was aligned with the same protein from the same organism with 95.1% of identity.

The low clinically relevant allergen tropomyosin partial sequence from Portugues oyster (Crassostrea angulate) was aligned in both AllergenOnline and in COMPARE with the same protein from Pacific oyster (*Crassostrea gigas*) with 99.6% of identity.

The moderately clinically relevant allergen Tropomyosin Scy o 1 from orange mud crab (*Scylla olivacea*) was aligned in AllergenOnline and COMPARE with 98.6% of identity with tropomyosin from three-spot swimming crab (*Portunus sanguinolentus*).

All other sequences were aligned and predicted correctly by all tools analysed.

3.3.2.4 FISH

The low clinical allergen Phosphoglucomutase-1 from zebrafish (*Danio rerio*) was not aligned in Allergen Online, was aligned in COMPARE with 28.3% of identity with lipid transfer protein from peach (*Prunus persica*). It was also predicted as non-allergen on AllergenFP, with the closest protein (Tanimoto 0.85) being Clathrin coat assembly protein AP180 from human (*Homo sapiens*). It was also predicte by AllerCatPro as nonallergenic (no evidence of allergenicity) and by AlgPred with the prediction pathway.

The low clinical allergen Enolase 3 from zebrafish (*Danio rerio*) was aligned in AllergenOnline with 93.3% of identity with enolase 3-2 from salmon (*Salmo salar*). In COMPARE it was aligned with 98.6% with Cyp c 2, beta-enolase from carp (*Cyprinus carpio*).

The low clinical allergen beta parvalbumin from zebrafish (*Danio rerio*) was aligned in AllergenOnline and COMPARE with 89.9% of identity with parvalbumin from carp (*Cyprinus carpio*).

The low clinically allergen creatine kinase from blackfin icefish (*Chaenocephalus aceratus*) was aligned in AllergenOnline with 88.5% of identity with creatine kinase-2 from salmon (*Salmo salar*). In COMPARE it was aligned it with 44.8% with the arginine kinase Der f 20 from house dust mite (*Dermatophagoides farinae*). AllergenFP predicted it as non-allergen with the closest protein (Tanimoto score of 0.82) being AF4/FMR2 family member 2 from human (*Homo sapiens*).

The low clinically relevant allergen creatine kinase from Mozambique tilapia (*Oreochromis mossambicus*) was aligned in AllergenOnline with 89.2% of identity with creatine kinase, muscle b from Iridescent shark catfish (*Pangasianodon hypophthalmus*). In COMPARE it was aligned with 46.1% with the arginine kinase Der f 20 from house dust mite (*Dermatophagoides farinae*). AllergenFP predicted it as non-allergen with the closest protein (Tanimoto score of 0.83) being Voltage-dependent L-type calcium channel subunit beta-1 from human (*Homo sapiens*).

The moderately clinically relevant allergen Aldolase Gad m 3 from Atlantic cod (*Gadus morhua*) was predicted in AllergenFP as non-allergen with the closest protein (Tanimoto of 0.53) being Adenosine receptor A2a from human (*Homo sapiens*).

The moderately clinically relevant allergen Enolase Gad m 2 from Atlantic cod (*Gadus morhua*) was predicted in AllergenFP as non-allergen with the closest protein (Tanimoto of 0.68) being Alpha-amylase III from rice (*Oryza sativa*). It was also predicted as nonallergenic by AlgPred with the prediction pathway.

The low clinically relevant allergen tropomyosin from Alaska pollock (*Gadus chalcogrammus*) was aligned in AllergenOnline and COMPARE with 96.5% of identity with tropomyosin from Mozambique tilapia (*Oreochromis mossambicus*).

The low clinically relevant allergen myosin (light chain) from Alaska pollock (*Gadus chalcogrammus*) was aligned in AllergenOnline and COMPARE with 63.9% of identity with myosin (light chain) from chicken (*Gallus gallus*). AllergenFP predicted it as non-allergen with the closest protein (Tanimoto of 0.83) being A-kinase anchor protein 12 from human (*Homo sapiens*).

The low clinically relevant allergen apolipoprotein from Orange-spotted grouper (*Epinephelus coioides*) was aligned in AllergenOnline and COMPARE with 22.9% of identity with a house dust mite allergen (*Dermatophagoides pteronyssinus*). AllergenFP predicted it as non-allergen with the closest protein (Tanimoto of 0.85) being apolipoprotein A-I from human (Homo sapiens). It was also predicted as nonallergenic by AllerCatPro (no evidence of allergenicity).

The low clinically relevant allergen adenylate kinase from Orange-spotted grouper (*Epinephelus coioides*) was aligned in AllergenOnline and COMPARE with 33.3% of identity with mite allergen Tyr p 7 (*Tyrophagus putrescentiae*). AllergenFP predicted it as non-allergen with the closest protein (Tanimoto of 0.82) being calcium/calmodulin-dependent serine/threonine-protein kinase 1 from rice (*Oryza sativa* subsp. *japonica*). It was also predicted as nonallergenic by AlgPred with the prediction pathway and by AllerCatPro (no evidence of allergenicity).

The low clinically relevant allergen Nucleoside diphosphate kinase B from European hake (*Merluccius merluccius*) was aligned in AllergenOnline with 31.4% of identity with group 2 allergen Sor h 2.0100 from Johnson grass (*Sorghum halepense*). It was aligned in COMPARE with 64.7% of identity with a partial sequence of nucleoside diphosphate kinase from pecan (*Carya illinoinensis*).

The low clinically relevant allergen β-parvalbumin from Bicolor damselfish (*Stegastes partitus*) was aligned in AllergenOnline and COMPARE with 85.3% of identity with parvalbumin from carp (*Cyprinus carpio*).

The moderately clinically relevant allergen aldolase (Thu a 3) from yellowfin tuna (*Thunnus albacares*) was predicted in AllergenFP as non-allergenic with the closest protein (Tanimoto score of 0.75) being transketolase from rice (*Oryza sativa* subsp. *japonica*)

The moderately clinically relevant allergen β-enolase (Thu a 4) from yellowfin tuna (*Thunnus albacares*) was predicted in AllergenFP as non-allergenic with the closest protein (Tanimoto score of 0.67) being abnormal spindle-like microcephaly-associated protein from human (*Homo sapiens*). It was also predicted as nonallergenic by AlgPred with prediction pathway.

The moderately clinically relevant allergen aldolase Sal s 3 from salmon (*Salmo salar*) was predicted in AllergenFP as non-allergenic with the closest protein (Tanimoto score of 0.89) being fructose-bisphosphate aldolase C from human (*Homo sapiens*).

The moderately clinically relevant allergen Tropomyosin Sal s 4 from salmon (*Salmo salar*) was aligned in COMPARE with 95.4% of identity to tropomyosin Ore m 4 from Mozambique tilapia (*Oreochromis mossambicus*).

The moderately clinically relevant allergen creatine kinase Sal s 7 from salmon (*Salmo salar*) was not aligned in AllergenOnline, and in COMPARE it was aligned it with 90.6% of identity to a partial sequence of actin from a species of flowering plant in the bean family Fabaceae (*Delonix regia*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.98) being Actin, alpha cardiac muscle 1 from human (*Homo sapiens*). It was also predicted as nonallergenic by AlgPred with prediction pathway and by AllerCatPro (no evidence of allergenicity).

The moderately clinically relevant allergen Ttiose phosphate isomerase Sal s 8 from salmon (*Salmo salar*) was aligned in COMPARE with 70% of identity to Der f 25, triosephosphate

isomerase from house dust mite (*Dermatophagoides farina*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.85) being triosephosphate isomerase, cytosolic from rice (*Oryza sativa* subsp. *japonica*).

The low clinically relevant allergens α-enolase from salmon (*Salmo salar*) was aligned in AllergenOnline and COMPARE with 96.8% of identity with enolase Sal s 2 from salmon (*Salmo salar*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.89) being phosphopyruvate hydratase from tomato (Solanum lycopersicum).

The low clinically relevant allergens vitellogenin from salmon (Salmo salar) was aligned in AllergenOnline and COMPARE with 99.5% of identity with vitellogenin from chum salmon (*Oncorhynchus keta*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.76) being acylphosphatase-2 from human (*Homo sapiens*).

The moderately clinically relevant allergen beta parvalbumin from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 87.2% of identity with parvalbumin from carp (*Ctenopharyngodon idella*).

The moderately clinically relevant allergen β-enolase Pan h 2 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in AllergenOnline with 90.6% of identity with enolase 3-2 from salmon (*Salmo salar*). COMPARE aligned it with 92.6% of identity with beta enolase Cyp c 2 from carp (*Cyprinus carpio*)

The moderately clinically relevant allergen aldolase Pan h 3 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 87.2% of identity with aldolase Sal s 3 from salmon (*Salmo salar*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.9) being fructose-bisphosphate aldolase C from human (*Homo sapiens*).

The moderately clinically relevant allergen tropomyosin Pan h 44 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 82.4% of identity with Ore m 4 tropomyosin from Mozambique tilapia (*Oreochromis mossambicus*).

The moderately clinically relevant allergen creatine kinase Pan h 7 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 45.5% of identity with Bomb m 1 arginine kinase from silkworm (*Bombyx mori*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.82) being Brain-enriched guanylate kinase-associated protein from human (*Homo sapiens*).

The moderately clinically relevant allergen triose phosphate isomerase Pan h 8 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 72.7% of identity with Der p 25, triosephosphate isomerase from house dust mite (*Dermatophagoides pteronyssinus*).

The moderately clinically relevant allergen pyruvate dehydrogenase Pan h 9 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 30.4% of identity with Bet v 1, pathogenesis related protein, PR-1 from silver birch (*Betula pendula*). It was also predicted as nonallergenic by AlgPred with prediction pathway.

The moderately clinically relevant allergen lactate dehydrogenase Pan h 10 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 30.2% of identity with Bla g 5, glutathione S-transferase from German cockroach (*Blattella germanica*).

The moderately clinically relevant allergen glucose-6--phosphate isomerase Pan h 11 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE aligned it with 23.1% of identity with Tab y 2, hyaluronidase from a type on insect (*Tabanus yao*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.82) being Bcl-2-like protein 13 from human (*Homo sapiens*). It was also predicted as nonallergenic by AlgPred with prediction pathway.

The moderately clinically relevant allergen glyceraldehyde phosphate dehydrogenase Pan h 13 from iridescent shark catfish (*Pangasianodon hypophthalmus*) was aligned in COMPARE with 79.6% of identity with Per a 13, glyceraldehyde-3-phosphate-dehydrogenase from American cockroach (*Periplaneta americana*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.86) being glyceraldehyde-3-phosphate dehydrogenase 2, cytosolic from rice (*Oryza sativa*).

The low clinically relevant allergen fructose-bisphosphate aldolase A from northern pike (*Esox lucius*) was aligned in AllergenOnline and COMPARE with 94.5% of identity with Sal s 3, aldolase a fructose-bisphosphate 1 from salmon (*Salmo salar*). AllergenFP predicted it as non-allergenic with the closest protein (Tanimoto score of 0.91) being fructose-bisphosphate aldolase A from human (*Homo sapiens*).

The low clinically relevant allergens alpha parvalbumin Raj c 1 from thornback ray (*Raja clavata*) was aligned in AllergenOnline with 62.5% of identity with parvalbumin from Spotless smooth hound (*Mustelus griseus*) and in COMPARE with 56.4% of identity with parvalbumin from Japanese jack mackerel (*Trachurus japonicus*).

All other sequences were aligned and predicted correctly by all tools analysed.

3.4 Sub-task 2.3: Identification of follow-up actions

Almost all food allergens identified to date are proteins, which are the main biopolymer towards which humoral immune responses are directed. Exceptions are molecules which are haptens and only elicit an antibody response because they are attached to a carrier protein. One example of such a molecule is α-galactose, which is involved in the development of allergies to red meat following sensitisation to ticks (Kersh *et al.*, 2023).

On this basis there is a concept of "no protein, no problem" which has led to the exemptions from allergen labelling for food ingredients such as highly refined soybean oils (EFSA, 2007). This approach to exemption was also recommended by the FAO-WHO Ad Hoc group on risk assessment of food allergens (FAO-WHO, 2024). Such considerations can be extended to novel food ingredients that have a very low protein content, since their risk of causing either cross-reactive allergies or *de novo* sensitisation is very low.

For other ingredients containing higher levels of proteins the exposure to the protein component can be mapped using the intended use and P75 consumption levels for an eating occasion. This has previously been applied to converting threshold doses into action levels by the FAO-WHO Ad Hoc expert consultation (FAO-WHO, 2022b), based on the approaches developed in the iFAAM project (Blom *et al.*, 2019, Birot *et al.*, 2018). Risks posed by proteins present at a low level, such some newly expressed proteins in GMOs (e.g. transcription factors) or proteins used as food processing aids, such as ice-structuring protein, will be much

lower than that for proteins being used at high levels in widely consumed foods, such as supplements in bakery products, dairy and meat alternatives.

3.4.1 Role of *in vitro* digestibility assays

Digestibility tests have been developed and used in allergenicity risk assessment to help support the risk assessment process since the development of the pepsin resistance test (Astwood *et al.*, 1996a, Thomas *et al.*, 2004, Codex Alimentarius Commission, 2003, Commission, 2009) which have subsequently been elaborated with the addition of intestinal digestion tests, some of which are more physiologically relevant (Mandalari *et al.*, 2009, Fu *et al.*, 2002). Classical tests employ purified proteases since this allows SDS-PAGE to be used to provide a readout, although a combination of *in vitro* gastric digestion followed by intestinal digestion using bile and pancreatin resulted in more complete digestion of proteins (Akkerdaas *et al.*, 2018) which has also been observed by others using a combination of purified proteins and bile salts in intestinal digestion models (Torcello-Gómez *et al.*, 2020a, Torcello-Gómez *et al.*, 2020b, Wang *et al.*, 2023c). Thus, it is clear biosurfactants, such as bile salts, should be incorporated in such digestion models to avoid overestimating resistance to digestion.

One of the issues that has been of concern is how to identify a clear readout of *in vitro* digestibility tests and what constitutes a persistent fragment (Fernandez *et al.*, 2019). The safety assessment of the ice-structuring protein (ISP) from artic pout utilised both gel-based and mass spectrometry methods of analysis, the value of incorporating data on protein half-life (Baderschneider *et al.*, 2002). This approach was also developed for the EuroPrevall interlaboratory trial of an *in vitro* digestion protocol incorporating both gastric and intestinal phases which highlighted the issues of reproducibly sampling and analysis *in vitro* digestion time courses (Mandalari *et al.*, 2009, Defernez *et al.*, 2010). More recently this approach has been applied to a range of allergenic and non-allergenic proteins following gastric (Wang *et al.*, 2022) and intestinal digestion (Wang *et al.*, 2023b).

These data illustrate the lack of a direct relationship between the rank order of digestibility and allergenic potential since casein, an important cow's milk allergen, was the most digestible protein whilst horse heart cytochrome c, a non-allergenic comparator protein, was highly resistant to intestinal proteolysis. They further point to a weak relationship between resistance to digestion and potential to act as allergens in IgE-mediated food allergy (Bøgh and Madsen, 2016, EFSA Panel on Genetically Modified Organisms *et al.*, 2021). However, stability to digestion is critical to provide information relevant to the exposure assessment, in particular the presentation of immunologically relevant protein and protein fragments to the gut mucosal immune system as part of the weight of evidence approach. It is also known that hydrolysis reduces the allergenicity of foods, with hydrolysates with a size distribution of less than 2.5kDa have a much-reduced capacity to either sensitise or elicit allergic reactions (Bogh *et al.*, 2015, Nutten *et al.*, 2020, Van Hoeyveld *et al.*, 1998). Thus *in vitro* digestibility tests can make an important contribution to understanding exposure of the gut mucosal immune system – and maybe linked to digestibility tests used as part of the nutritional assessment. Indeed, when linked to the level of protein and intended use, rapid degradation of protein supports a low allergenicity risk profile for such foods.

Such digestion conditions and analysis of digestion products should ensure the following:

(1) Non allergenic comparators and standard proteins are included as controls to provide references of different susceptibilities to digestion;
(2) Digestion conditions should model different vulnerable groups such as infants/young children and those taking medication such as antiacids and should include biosurfactants such as bile salts;
(3) To take approaches to calculating properties of proteins, such as half-life and kinetic constants, requires replicate analysis to provide statistical rigour and sufficient time points to be sampled to allow kinetic analysis, such as curve fitting.
(4) Analysis of digestion products should employ methods that are suitable for analysis of intact proteins and large digestion products together with lower molecular weight peptides and take account of larger structures that maybe present as a consequence, for example, of intramolecular disulphide bonding (Mackie *et al.*, 2019, Wang *et al.*, 2023c).
(5) Where densitometric methods are applied in conjunction with gel-based analysis of digestion products requires use of molecular weight markers and protein loading which is within the linear range of the dye used for protein staining.

The remaining framework for identification of follow-up actions falls into two aspects, the possibility that a novel protein could elicit an allergic reaction in someone who already has an IgE-mediated allergy to another food, or that it has the potential to initiate a new IgE response (*de novo* sensitisation).

3.4.2 Proteins that might cause cross-reactive allergies

Step 1A, Phylogenetic analysis: Sequence similarity provides the opportunity to identify whether a protein has the potential to share sufficient sequence, and associated structural, similarity to present potential cross-reactive IgE epitopes. Such sequence similarity – whether at the genetic or protein level – underpins modern, molecular phylogeny (Yang and Rannala, 2012) and has facilitated the development of a new "tree of life" (Hug *et al.*, 2016) although rooting such phylogenetic trees is still a matter of debate amongst evolutionary biologists (Al Jewari and Baldauf, 2023). Thus, for novel food ingredients, a phylogenetic analysis might be a more attractive and simple approach to undertake to gain an initial indication as to potential for cross-reactive allergenicity than attempting to apply bioinformatic approaches developed for single novel proteins developed for allergenicity risk assessment of GMOs. Where phylogenetic differences are large it may obviate the need for further detailed cross-reactivity analysis

In general, *in silico* sequence similarity measures have not been validated as true predictors of *in vivo* clinical cross-reactive allergies. However, close phylogenetic relationships are indicative of concordant clinically related allergies as has been demonstrated for walnut and pecan as well as cashew and pistachio allergies using food challenges (Brough *et al.*, 2020, Nesbit *et al.*, 2020). Such phylogenetic relationships have the potential to provide a route to identifying allergens with a very high likelihood of causing cross-reactive food allergies. It has been applied to identifying allergenicity risks posed by pink peppercorns (Bastiaan-Net *et al.*, 2019, Fong *et al.*, 2019, Too *et al.*, 2019) and has already been applied to understanding the potential allergenicity of insect-derived dietary protein (Verhoeckx *et al.*, 2014).

Thus, an initial step in allergenicity risk assessment could be formulated through an assessment of phylogenetic distance to important allergenic food sources. Identification of comparator allergenic sources would be based on prevalence of allergies in the population, severity of reaction and (where available) potency determined in low dose threshold studies. It could be based on approaches developed in the FAO-WHO Ad Hoc expert consultation on risk assessment of food allergens and the ranking of allergenic and nonallergenic comparator foods developed in the Giant Leaps EU project (GA No 101059632) applied in this report.

Step 2A, Structure and sequence similarity: Clinical cross-reactivity between two closely similar foods results when the IgE response mounted to the repertoire of proteins (allergens) one food structurally resembles those found in a second food. Most antibody epitopes are thought to be discontinuous of conformational, where the epitope is formed by different segments of a proteins primary structure which are brought together topographically because of the folded state of a protein. Thus, cross-reactivity is a function of three-dimensional structural similarity which is in turn drive by primary sequence similarity. A consequence of such structural relatedness between proteins from different food sources is that their IgE binding sites (epitopes) can be largely equivalent and is directly linked to cross-reactivity (Barre *et al.*, 2005, Barre *et al.*, 2007).

Thus, the major, clinically relevant 11S seed storage globulin allergens from walnut and pecan share very high levels of sequence identity of 95.04% of identity between the 11S and 92.71% of identity between the 7S seed storage globulin allergens of walnut and pecan (determined using Clustal OMEGA 2.1). it is interesting to note this level of sequence similarity is consistent with the IUIS definition of an isoallergen (Pomés *et al.*, 2018).This provides a benchmark for the level of sequence identity associated with clinically proven cross-reactive allergy. The corollary of this is the lower threshold for potential cross-reactivity has also been set at 35% and used for many years (Codex Alimentarius Commission, 2003).

For proteins with intermediate sequence identities there is a spectrum of likelihood of causing cross-reactive allergic reactions. This can be illustrated for fish allergy. Concordance of food allergy is also considered to be widespread between different fish species driven by the fact that allergenic β-parvalbumins from different fish species share closely similar three-dimensional structures and consequently present highly conserved IgE epitopes (Kumeta *et al.*, 2017, Moraes *et al.*, 2014). Despite the premise of high levels of cross-reactive allergies between fish species data on clinical reactivity established by oral food challenge to multiple species are sparse. Thus, in one study of 19 children, seven reacted only to cod on food challenge, two only to salmon with six children reacting to cod and salmon and only four to both cod and mackerel (Sørensen *et al.*, 2017). Taking cod fish as the most allergenic fish in the study, the β-parvalbumin Gad m 1.010 from Atlantic cod (Gadus morhua) shares 72.5% sequence identity with the β-parvalbumin Sco s 1.010 from Atlantic mackerel (Scomber scombrus) and 66.1% sequence identity with that from Salmon (Salmo salar), Sal s 1.010. The levels of β-parvalbumins also vary between fish species with 20-30 fold higher levels in fish such as Atlantic cod (Gadus morhua) and whiff (Lepidorhombus whiffiagonis) compared to swordfish (Xiphias gladius) which has been linked to the lower allergenicity of swordfish (Griesmeier *et al.*, 2010).

Structural informatics approaches can provide important complementary information about "hots spots" of conserved surface structures. This can be illustrated by the structural analysis of β-parvalbumins from different species (Moraes *et al.*, 2014). This indicates a close similarity

of 70-77% between different fish species, which is reduced to 44% for human oncomodulin with chicken being intermediate with 44% identity, reflecting their cross-reactivity.

Thus, a second step would explore similarities identified using phylogenetic analysis in more depth, using combinations of sequence and structural similarity to characterize the likelihood that a novel protein would cause a cross-reactive allergy in a vulnerable food allergic group.

Proteins with ≥80% sequence identity to an allergen of high clinical relevance can be considered to have very similar allergenicity to the comparator and is highly likely to cause cross-reactive allergies with that allergen.

Proteins with intermediate sequence homology <80% and ≥35% sequence identity to an allergen of high clinical relevance to clinically relevant allergens would be subject to further testing.

Proteins with ≤ 50% but ≥35% sequence identity to an allergen of high clinical relevance are unlikely to cause a cross-reactive allergy. However, inspection of the aligned sequences should be undertaken to assess whether there is significant conservation of sequence associated with certain domain structures, such as EF-hands found in calcium binding sites and IgE epitopes. This would inform the need for further testing.

Proteins with ≤35% sequence identity to an allergen of high clinical relevance are very unlikely to cause a cross-reactive allergy.

Step 3A, *In vitro* confirmation of cross-reactive potential: For weaker relationships, confirmation of IgE binding using serum panels from food allergic patients is useful but the results of such tests can be misleading as sensitisation is only a risk factor for food allergy. For example, 44% of peanut-allergic patients were found to be sensitized to lupin but only 28% were clinically allergic to lupin (n=8 one study in France)(Moneret-Vautrin *et al.*, 1999), whilst in another study 34% of 47 peanut allergic children in a UK study were sensitized to lupin but only 4% were estimated to be clinically allergic (based on OFC in nine children, one study in the UK)(Shaw *et al.*, 2008). A more recent study legume allergies showed that co-allergies to green pea, lupine, lentil and bean are uncommon (≤16,7%) in a peanut allergic population (n=30), even though patients were sensitized to those foods (Smits *et al.*, 2023). Indeed, misleading cross-reactions in serum IgE testing makes test results complex to interpret. For example, many atopic individuals may be sensitised to house dust mite, where the major clinically relevant allergen is Der p 1, with sensitisation to mite tropomyosin (Der p 10) being considered irrelevant for house dust mite allergy (Huang *et al.*, 2023). IgE to Der p 10 in such patients can cross-react with tropomyosin from crustacean shellfish such as Pen a 1, giving false positive test results. Thus, in routine clinical practice serum IgE-tests and skin prick testing are always interpreted in the context of a good clinical history and profiling of symptoms as well as checking if an individual has recently consumed a food without symptoms (Lyons *et al.*, 2021b).

It may also be that using routine clinical tests, such as the ImmunoCAP, where the antigen is in excess to provide a highly sensitive diagnostic tool, may also overestimate IgE-cross reactivity. Thus, other types of readout, such as effector cell responses, may have more value in characterising the risks novel foods pose of causing an allergic reaction *in vivo*. As per the development of allergic reactions *in vivo*, such assays also require presentation of multiple IgE epitopes to allow cross-linking of cell bound IgE.

Serum IgE binding studies should be undertaken using biological samples from the allergic population at risk which will have been defined in Steps 1 and 2.

Patients providing serum samples should have a well-documented history of reaction to the problem food, evidence of sensitisation to that food in accordance with best clinical practices and, if at all possible, have their food allergy confirmed by oral food challenge (Santos *et al.*, 2023). The patient characteristics need to be appropriately documented with data on symptoms and associated food and inhalant allergies together with skin prick and/or serum specific IgE testing done for diagnostic purposes.

Studies should be sufficiently powered to provide robust data and include serum samples from relevant atopic and non-atopic subjects.

Serum samples will need careful curation to exclude those with high levels of IgE to cross-reactive carbohydrate determinants and may want to exclude those with very high total IgE levels which can pose technical issues in immunoassays.

For immunoassays both direct and inhibition immunoassay formats should be used.

For mediator releases assays, there are a variety of test option employing cells lines (such as the humanised RBL-SX38 cell line expressing the human high affinity IgE receptor (Bucaite *et al.*, 2019), and those using either stripped basophils, a patient's own basophils (Santos *et al.*, 2021a) or cultured mast cell lines (Bahri *et al.*, 2018).

Thus, using serum screening as a third step in checking out signals indicated by sequence similarity is useful, but the selection of patients and the details of the clinical history and diagnosis of IgE-mediated food allergy need to be carefully evaluated to contextualise test results. There would also be added value from using effector cell assays to provide a superior readout to IgE-binding studies.

Ultimately, if concerns remain, undertaking *in vivo* skin testing and food challenge studies which provide data on potency and an indication of severity of reaction would help inform any outcome and subsequent risk management.

3.4.3 Proteins that have sensitisation potential

Novel proteins also have the potential to sensitise vulnerable individuals. It is widely acknowledged that there are limitations in animal models for IgE-mediated allergy and associated conditions such as asthma. Given this and the ethical concerns regarding animal experimentation, alternative approaches need to be applied.

History of use: Understanding any history of sensitization in other contexts of the alternative protein source is valuable such as through occupational and other exposures. Understanding the potential for novel proteins to sensitise humans is useful but has many caveats especially for routes of exposure which do not involve the gastrointestinal tract. For example, wheat flour is acknowledged as being able to cause occupational allergies through inhalation of dust causing a condition known as Baker's asthma. However, individuals with Baker's asthma are sensitised to a different profile of allergens and are generally able to safely consumer wheat-containing foods (Sander *et al.*, 2001, Sander *et al.*, 2015, Armentia *et al.*, 2009). Dermal sensitisation maybe more important, especially in individuals with defects in skin barrier function, resulting from, for example, filaggrin mutations (Kalb *et al.*, 2022). There is good evidence that skin sensitisation, even from exposure to allergenic foods in dust, may play a role in development of food allergy (Brough *et al.*, 2013).

Companion animals also suffer from food allergies (Mueller and Unterer, 2018) and indeed dogs are accepted as being a better model for oral sensitisation to food than small animals (Buchanan and Frick, 2002). Several novel foods, especially protein-rich ingredients, are used in animal feed before being developed as ingredients for human consumption. Evidence of allergenicity (or its absence) from animal feed, especially pet food, may also be useful in assessing the potential for new ingredients to act as *de novo* sensitising agents.

Antigen presentation in *de novo* sensitisation: The predisposition to food allergy, like other atopic diseases, includes genetic factors with genotyping studies having identified a number of predisposing genes, including STAT6, SPINK5, FOXP3, IL-10 and HLA, the latter being particularly prominent. Current knowledge of immune mechanisms involved in sensitisation assumes that the formation of complexes between HLA class II proteins and peptides derived from antigens plays a central role in both T- and B-cell activation, proliferation and differentiation. The ability to process and present a given protein in a host's antigen presenting cells (APCs) is determined by the presence of a set of HLA class II proteins able to form stable peptide-HLA complexes and to present them to naïve Th-cells. If these HLA-types are lacking it is not possible to mount a humoral response to that protein. HLA types have been associated with peanut allergy (Kostara *et al.*, 2020, Hong *et al.*, 2015, Asai *et al.*, 2018, Kanchan *et al.*, 2022a, Kanchan *et al.*, 2022b) together with shrimp, peach and wheat (Noguchi *et al.*, 2019) (Fukunaga *et al.*, 2021). Building on *in silico* approaches developed in vaccine research, it is possible to explore whether a novel protein or protein ingredient carries sequence motifs able to bind to HLA class II molecules, and is a determinant for their immunogenicity, a property required for a protein to be able to sensitise (i.e. initiate an IgE-response). Application of machine learning techniques to this problem is in its infancy but has the potential to provide an in silico tool that could form the first step in an approach to assessing the *de novo* sensitisation potential of a novel protein based on approaches being developed for vaccine design (Wang *et al.*, 2023a, Xu *et al.*, 2022) and are beginning to be applied to allergenicity prediction (Li *et al.*, 2023). The power of such methodology will

increase as the quality of genotyping allergy cohort studies improves and becomes more affordable.

Part of the process involved in antigen presentation and activation of immune cells (including dendritic cells and B-cells) is the uptake of antigen, degradation of antigens through the endosomal pathway, where degradation of proteins provides the peptides which become bound to the HLA class II molecules intracellularly before the HLA-peptide complexes are transported to the cell surface (Adler *et al.*, 2017). The endosomes represent a highly degradative intracellular compartment with a low pH and is host to oxidoreductases which have the capacity to oxidise intramolecular disulphide bonds, together with a a complex mixture of endoproteases which are highly effective at generating peptides of 9-15 amino acids in length which can bind to HLA class II receptors (Bird *et al.*, 2009, Perrin *et al.*, 2019). In recent years it has become evident that the endosomal pathways affect immune responses and *in vitro* endolysosomal assays have been developed and used to map the generation of T-cell epitopes in allergens such as Bet v 1 (Hofer *et al.*, 2017) where lipid interactions have been shown to affect allergen degradation (Soh *et al.*, 2019). However, the added value of such complex digestion assays, over in silico analysis of potential HLA binding sites related to food allergy is unclear at present.

3. Cell-based studies: There are a few cell-based approaches being explored to provide tools that are predictive of *de novo* sensitization. One such approach based on expression profiling of human peripheral blood mononuclear cells did allow differentiation of major clinically relevant allergens from legumes (Smits *et al.*, 2021a). Other more sophisticated co-culture models employing combinations of gut epithelial cells and immune cells involved in antigen presentation and B-cell activation (Zuurveld *et al.*, 2022). However, readouts from such cell-based assays that are related to sensitizing potential of proteins have not been agreed and they require validation using well characterised allergens and non-allergens of differing clinical relevance. There are also issues over stability of human cell lines which may give different results in different laboratories, and they represent only a single repertoire of, for example HLA-types which play an important role in the immune system's ability to discriminate host from foreign protein and mount a humoral immune response. Such gaps mean the technology readiness level of such approaches for use in allergenicity risk assessment is some years away.

In vivo assessment: animal models, including routes of administration (Bloom *et al.*, 2014, Adel-Patient *et al.*, 2012) and studies *in vivo* in humans, (Crevel *et al.*, 2007) have been to confirm lack of immunogenicity, as has been used previously for novel proteins. These approaches, particularly those in humans, provide additional data reassuring a very low likelihood of allergenicity.

3.5 Task 3: Develop a novel approach for the allergenicity assessment of innovative/novel proteins by integrating in silico, *in vitro* and *in vivo* methods through implementation of the final ranking strategy of known allergens

3.5.1 SubTask 3.1: Implementation the final ranking strategy of known allergens

The implementation of the ranking strategy has been considered in the context the allergenicity risk assessments performed for novel foods, using mung bean and insects as examples, together with one on GMO.

3.5.1.1 MUNG BEAN

An example of how the outputs of this study could inform the allergenicity risk assessments is case of mung bean protein. A recent scientific opinion from EFSA journal considered the safety of mung bean as a novel food (EFSA Panel on Nutrition *et al.*, 2021a). The novel ingredient prepared from mung bean comprised88-91% protein with and intended maximum use of 200g/Kg as a protein analogue or substitute for meat, fish or milk (EFSA Panel on Nutrition *et al.*, 2021a). The P75 at a single eating occasion for this type of product is 113g which corresponds to consuming up to 22.6g of the novel ingredient at the maximum intended use level (FAO-WHO, 2022b). The Protein Digestibility Corrected Amino Acid Score (PDCAAS) ranged from 0.635 to 0.580 depending on whether it was raw or cooked which is poorer than other legume ingredients, such as soybean. No *in vitro* digestibility test was reported for the allergenicity risk assessment. These data indicate that there is the potential for the ingredient to pose an allergenic risk.

Potential to cause cross-reactive food allergies

Step 1: Phylogenetic analysis: Mung bean is a legume which like the important legume allergen, peanut, belongs to the subfamily Papilionoideae, but is more closely related to soybean and lupin (Figure 4). Thus, there is the potential for the novel food to pose a risk of reaction to individuals with these legume allergies.

Step 2: Sequence analysis of mung bean proteins: Since several mung bean allergens have been described in the literature these form the focus of the assessment (see Table 33). Their clinical relevance was assessed using the approach described in this report. Since the novel food originated from mung bean seeds the sequence analysis was undertaken focussing on the allergens identified in the seeds, Vig r 2 and Vig r 4.

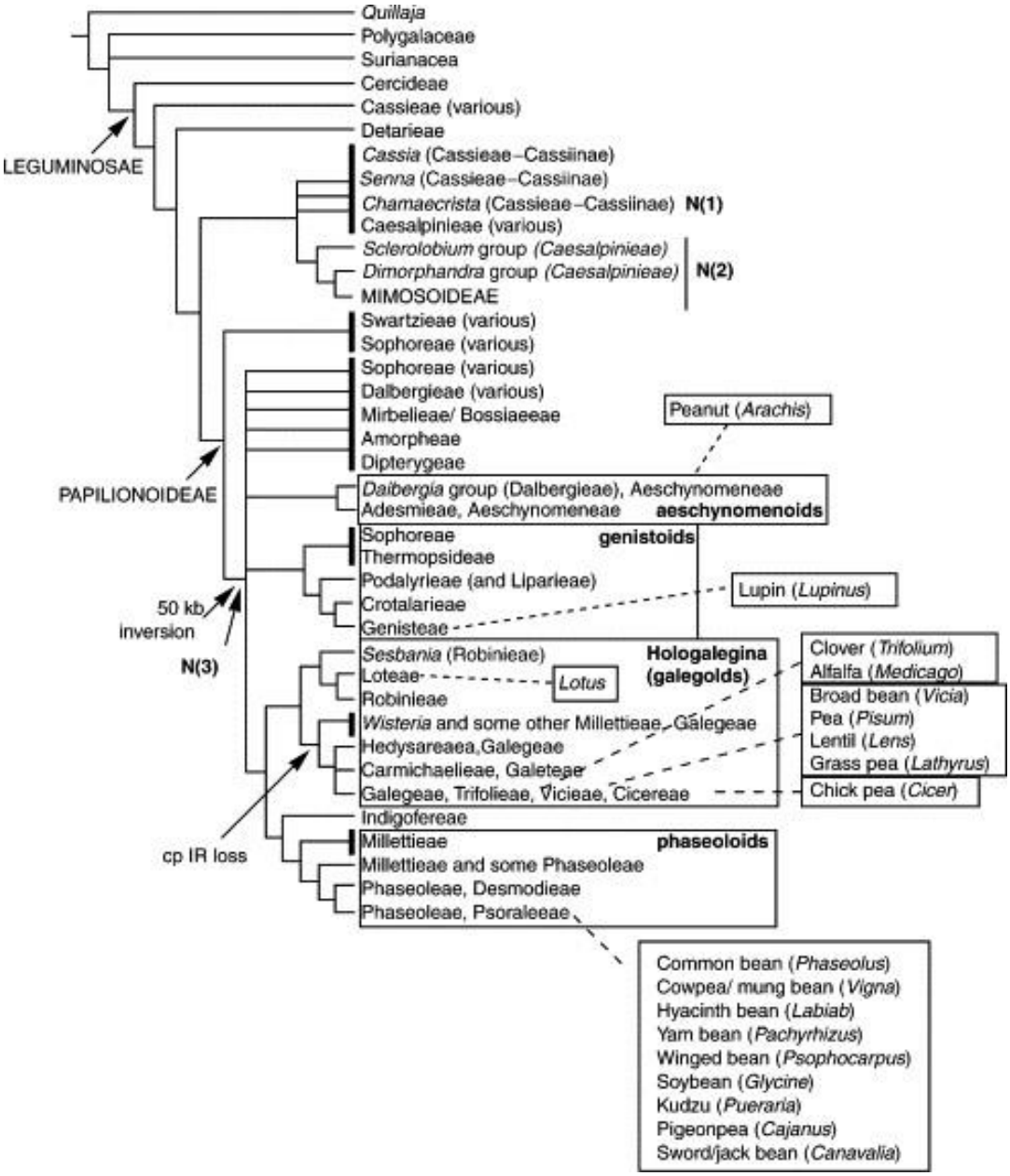

Figure 4: Phylogenetic tree of legumes (Doyle, 2001) Reproduced with permission

Table 33: Mung bean allergens with type of allergen, reference, accession numbers, number of papers and participants found in the literature and the type of test/s used to characterise the allergens. In 2019 Vig r 3 was renamed to Vig r 2.0201, and Vig r 5 was later identified as fragment of Vig r 2 (Misra et al., 2011)

Allergen	Type of allergen	Accession	No of papers	No of participants	Type of test	References
Mung bean seedling (bean sprouts)						
Vig r 1	PR-10, Bet v 1 family member	Q2VU97	1	70	ImmunoCAP with 10 patients reporting symptoms after eating mung bean; all had a positive OFC to soybean.	(Mittag *et al.*, 2005)
Vig r 6	Cytokinin-specific binding protein (CSBP), Bet v 1 family	Q9ZWP8	1	19	Skin prick test, IgE in sera, case history. 32% of pateitns testing positive of whom 5 reproted symptoms with bean sprouts	(Guhsl *et al.*, 2014)
Mung bean seeds						
Vig r 2	8S Globulin vicilin) Vig r 2.0101	Q198W3	1	12	Skin prick test, IgE in sera	(Misra *et al.*, 2011)
	Vig r 2.0201 (formerly Vig r 3)	B1NPN8	1	12	Skin prick test, IgE in sera	(Misra *et al.*, 2011)
Vig r 4	Seed 2S albumin	Q43680	1	12	Skin prick test, IgE in sera	(Misra *et al.*, 2011)

Sequence comparisons were undertaken using BLAST searching against UniProt (The UniProt, 2023), together with FASTA searching of Allergenonline (Goodman *et al.*, 2016) using either the full length of the proteins or a sliding 80 amino acid window (Pearson and Lipman, 1988):

Vig r 2: This 8S globulin, vicilin-like protein was found to have,

- 68.1% (BLAST-UniProt) and 68.0 % (FASTA-Allergenonline) of sequence identity with β-conglycinin α' subunit, Gly m 5, from soybean (Glycine max);

- 53% (BLAST-UniProt) and 52.7% (FASTA-Allergenonline) of sequence identity with the vicilin Pis s 1 from pea (Pisum sativum).

- 49.1% (BLAST-UniProt) and 51.2% (FASTA-Allergenonline) of sequence identity with Ara h 1 P43237.

Vig r 4: the 2S albumin of mung bean, the only hit above the conventional 35% sequence identity threshold was with Pis s Albumin from pea (Pisum sativum), a sequence that is not found in WHO/IUIS Allergen Nomenclature database.

Alignment of the proteins together with the cognate allergens from peanut (Ara h 1 isoforms P43237 and P43238) using Clustal OMEGA (Sievers and Higgins, 2018) showed Vig r 2 had a sequence similarities of 49.6-50.9% with Ara h 1, 62.1-63.9% with Gly m 5 and 52.6-53.7% with Lup an 1. Review of the aligned seuqence shows a high level of conservation in the C-terminal portion of the 7S globulins, the Vig r 2 lacking the N-terminal insert found in Ara h 1 and Gly m 5.

The submission of these sequences to the bioinformatic tools assessed in this report return correct alignments and predictions that they are allergens except for Vig r 4. This protein did not have a hit in AllergenOnline, and AllergenFP identified the sequence as probable non-allergen, with the closest protein (Tanimoto score 0.8) being Beta/gamma crystallin domain-containing protein 1 from human (*Homo sapiens*).

The hits identified were to allergens from peanut (Ara h 1), soybean (Gly m 5) and lupin (Lup an 1) which were all classified as being highly clinically relevant in the systematic review. The prevalence of peanut allergy varies across Europe from 0.00% (0.00-0.28 95%CI) in Iceland (adults and school-age children) to 0.45% (0.05-1.45 95%CI) in adults and 1.1% (0.24-2.71 95%CI) in children, whilst soybean allergy in Europe is very low (Lyons *et al.*, 2019, Lyons *et al.*, 2020). That of lupin is not established but is not highlighted as a significant food allergen in any European epidemiology study.

These data indicate mung bean proteins represent a moderate-low likelihood of causing a cross-reactive allergy in individuals with soybean allergy and lower likelihood of causing cross-reactive allergies in peanut and lupin allergic individuals.

In the absence of structural bioinformatic analysis the next step would be to confirm this ranking using serum IgE-testing.

Step 3: This should be assured by checking the serum IgE-cross-reactivity profile. Although accessing serum samples from peanut allergic individuals is possible it would be more difficult for the much less prevalent allergy to soybean. One publication has described a limited immunoblotting experiment using serum samples form legume allergic subjects although their clinical characteristics were poorly described and it is unclear if they were simply legume

sensitised (Calcinai *et al.*, 2023, Calcinai *et al.*, 2022). One type of *in vivo* clinical test that can also be considered, and would inform the need for oral food challenge, is a prick-to-prick skin test with the novel ingredient. Prick-to-prick tests are used in routine clinical diagnosis and involve making a paste with crushed food before taking a tiny amount on a lance to pricki the skin of a potentially allergic subject. If the patient develops a wheal >3mm in diameter and has negative and positive controls tests using saline and histamine respectively, it indicates the subject is likely sensitised to the food.

The Bet v 1 homologues Vig r 1 and 2 are unlikely to trigger a reaction to the seed. Further assurance as to this low likelihood could be obtained from semi quantitative proteomic profiling or immunoassay analysis to provide protein-level evidence as to the allergens presence in the product. Processing can reduce levels of PR10 homologues as shown, for example, with soybean (Mittag *et al.*, 2004) with serum IgE binding studies/ skin testing in Bet v 1 sensitised individuals as a follow-up action.

That allergens have been identified from the NI indicates it has the capacity for *de novo* sensitisation. However, based on sequence similarity analysis that capacity is likely to be more similar to soybean than peanut suggesting it is a moderate to low risk of allergenicity. It is important to note that populations tend to manifest allergies to foods they eat – for example the prevalence of fish and crustacean shellfish allergy is much higher in European countries where these foods are eaten more extensively (Lyons *et al.*, 2019). Thus, the capacity of a NI to sensitise is likely to reflect extent of use as well as food processing.

3.5.1.2 INSECTS

Insects such as silkworm (*Bombyx mori*), mealworm (*Tenebrio molitor*), cricket (*Acheta domesticus*) and locust (*Locusta migratoria*) have been partially or fully domesticated to be used for their products or as food for animals and, ultimately, for humans. Insects have been well documented as causing allergic reactions, particularly insect stings, Hymenoptera venom allergy being one of the top three causes of anaphylaxis across the world (Stoevesandt *et al.*, 2020) which can be a significant cause of anaphylactic reactions. They can also act as inhalant allergens, with house dust mite allergy being estimated to affect 21% of the European population (Calderón *et al.*, 2015). One novel food which was not granted a marketing authorisation because of concerns of allergenic risks, was bee venom in honey (https://acnfp.food.gov.uk/sites/default/files/mnt/drupal_data/sources/files/multimedia/pdfs/beevenomukopinion.pdf). However, benchmarking allergenic risks from foods, against insects where sensitisation is through stings and inhalation of faeces and insect particles, requires careful interpretation in relation to food allergy.

EFSA have been considering the safety of edible insects including mealworm (*Tenebrio molitor*) (EFSA Panel on Nutrition *et al.*, 2021c), lesser mealworm (*Alphitobius diaperinus larva*)(EFSA Panel on Nutrition *et al.*, 2022b) cricket (*Acheta domesticus*) (EFSA Panel on Nutrition *et al.*, 2021b, EFSA Panel on Nutrition *et al.*, 2022a). Cricket (*Acheta domesticus*) as a partially defatted cricket powder (EFSA Panel on Nutrition *et al.*, 2022a) was selected as another example of how the outputs of this study could inform the allergenicity risk assessment.

Cricket powder comprises 74–78% protein although the EFSA panel review noted the default conversion factor of 6.25 was used for the total protein determination using the Kjeldahl method. This will have likely overestimated the protein content due to the presence of significant amounts of non-protein nitrogen originating from the chitin component. nitrogen determination. The novel ingredient was proposed for use in several food categories at levels ranging from 0.1-5g/100g of final food product. The highest level 5g/Kg was proposed for as a substitute for meat (EFSA Panel on Nutrition *et al.*, 2022a). As for the case of mung bean isolate, the P75 at a single eating occasion for this type of product is 113g which corresponds to consuming up to 5.65g of the novel ingredient at the maximum intended use level (FAO-WHO, 2022b). Data on protein digestibility were proprietary and not disclosed in the EFSA opinion. However, values in the scientific literature for the PDCAAS suggest it can range from around 0.65-1.00 (Stone *et al.*, 2019, Ruggeri *et al.*, 2023). No *in vitro* digestibility test was reported for the allergenicity risk assessment.

These data indicate that there is the potential for the ingredient to pose an allergenic risk.

Risks for cross-reactive allergies:

Step 1 Phylogenetic considerations: Insects, like crustacean seafood species, are Arthropods (Figure 5) and indicates that certain types of insects are closely related to known food allergens, crustacean shellfish, both belonging to the Pancrustacea. However, they are only distantly related to allergenic species, such as house dust mite, which belong to a different, much more distantly related branch of the arthropods, the Chelicerata, alongside ticks and spiders amongst other species. The crustacean species for which the greatest weight of evidence exists as to their allergenicity are shrimp of which one of the best studied is *Peneaus aztecus*, *Paeneus monodon* and *Litopenensis vanameii*. Mites (notably house dust mite species *Dermatophagoides pteronyssinus*, *Dermatophagoides farina* and *Blomia tropicalis*) are a major cause of inhalant allergies triggered by sensitisation to mite faeces. They are members of the Pyroglyphidae from the subclass Acari and are only distantly related to the Pancrustacea to which both Crustaceans and Insects (Heaxapoda) belong (Figure 5).

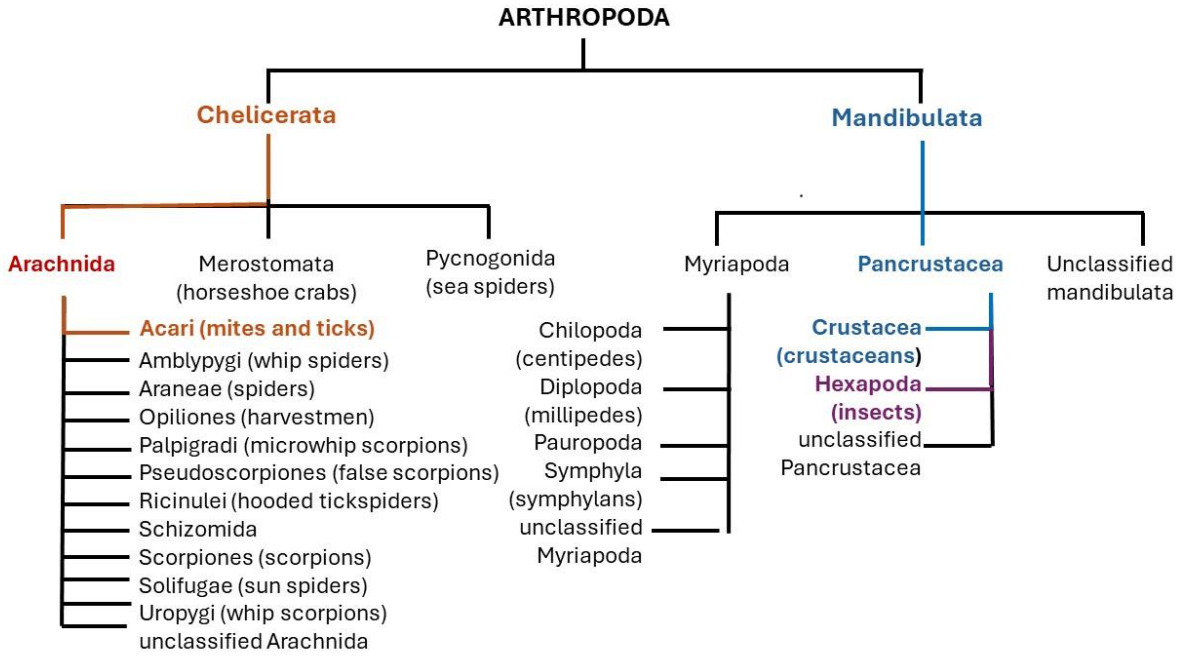

Figure 5 Phylogenetic relationships of the Arthropods

Based on NCBI taxonomy. Branches shown in brown relate to allergenic arthropods by inhalation and those in blue to those by food with insects shown in purple.

Step 2: Sequence analysis of cricket (*Acheta domesticus*) proteins of potential allergenic risk: Based on phylogenetic analysis key allergens identified in the different shrimp species are:

Tropomyosin: Searching UniProt for Acheta domesticus returned two accessions annotated as tropomyosins from the unreviewed database, Tropomyosin 1 A0A4P8D324; Tropomyosin 2 A0A4V1DVH3. Their sequences were aligned using CLUSTAL-Omega, with the tropomyosin's ranked as being of high clinical relevance from *Penaeus aztecus* (Pen a 1 Q3Y8M6), P*enaeus monodon* (Pen m 1, A1KYZ2) and *Litopeneus vannamei* (Lit v 1, B4YAH6). The three tropomyosin sequences from the different shrimp species all aligned with each other with 100% sequence identity, whilst they aligned with the tropomyosin 2 from *Acheta domesticus* with 83.17% sequence identity and the tropomyosin 1 with sequence identity of 67.8%. The alignments of *Acheta domesticus* tropomyosin 1 with the tropomyosin from house dust mite, Der p 10, gave an identity of 82.2%, the tropomyosin 1 having a lower level of sequence identity of 62.9%.

As might be expected based on their close phylogenetic relationship sequence alignments between cricket tropomyosins and those from other insect species gave a high level of sequence identity. Analysis of tropomyosin 1 from *Acheta domesticus* using the AllergenOnline tool gave and alignment with tropomyosin from mealworm (*Tenebrio molitor*) with 83.9% sequence identity, and with tropomyosin from silverfish (*Lepisma saccharinum*) with 81.5% sequence identity. Tropomyosin 2 from *Acheta domesticus* was aligned with

tropomyosin from smokybrown cockroach (*Periplaneta fuliginosa*) and American cockroach (*Periplaneta americana*) with 93.6% of identity and with tropomyosin from German cockroach (*Blattella germanica*) with 93.1% of identity. Alignment with tropomyosins from fish species was lower, at 68% sequence identity.

These data indicate that there is a possibility that *Acheta domesticus* may present a risk of cross-reactive allergic reactions in individuals who are allergic to shrimp species such as *Penaeus aztecus*, *Penaeus monodon* and *Litopeneus vannamei*. Based on concordance of tree nut allergies it was proposed that sequence identities ≥80% would be indicative of concordant allergies. However, such concordance is not observed between house dust mite and crustacean shellfish, likely because dust mite tropomyosin is not considered to be a clinically relevant allergen in house dust mite. There may also be other factors involved in determining clinical reactivity in food allergens that show a high level of sequence similarity to human proteins. This indicated that a straightforward sequence identity cut-off value cannot be assigned but only used as a guide to inform subsequent testing.

Regarding myosin, searching UniProt returned one accession annotated as a fragment of myosin from the unreviewed database, myosin fragment D0UJM7. This sequence was aligned in BLASTp reviewed (Swiss-Prot) with 3 other myosins, from *Aedes aegypti* (Yellowfever mosquito), *Drosophila melanogaster* (Fruit fly), and *Drosophila pseudoobscura pseudoobscura* (Fruit fly), all with 97.2% of identity. This sequence was aligned in BLASTp unreviewed (TrEMBL) with 247 sequences in total. The sequence was aligned with myosin from Homarus americanus (American lobster) with 93% of identity, with putative myosin from *Penaeus vannamei* (Whiteleg shrimp) with 92% of identity, with myosin from *Crassostrea virginica* (Eastern oyster) with 92% of identity, and others.

No analysis could be performed for arginine kinase nor for sarcoplasmic Ca2+ binding protein due to a lack of protein sequences available in UniProt. This lack of data available precludes the use of bioinformatic approaches to assess the allergenic risk assessment with proteins other than tropomyosin.

Step 3: IgE reactivity studies have been performed with *Acheta domesticus* using serum samples from 12 patients with allergy to shrimp and 8 patients with sera purchased through a commercial supplier (De Marchi *et al.*, 2021). The clinical history of patients were not described well. Neither sets of serum samples well described regarding the patient allergies and notably lack data on other associated inhalant allergies especially house dust mite. IgE-reactivity of tropomyosin was confirmed only by immunoblotting and no inhibition ELISAs were performed.

This indicates that there is a possibility for a cross-reactive allergic reaction between *Acheta domesticus* and shrimp species. Wider studies with larger numbers of serum samples from individuals with well documents allergies should be undertaken and including appropriate patient controls (atopic and non-atopic controls). Mediator release assays may also provide a superior readout and are already being calibrated with tropomyosins with confirmed clinical relevance from crustacean shellfish (Pen m 1) and vertebrate tropomyosin from chicken (Klueber *et al.*, 2020). However, further validation of such methodology is required to assess whether it can discriminate between more closely related invertebrate tropomyosins of different clinical relevance, such as Pen m 1 and Der p 10.

Depending on the outcome of the serum study there maybe a need and justification to undertake *in vivo* testing in humans through skin testing and oral food challenge. Oral food challenge studies for mealworm confirmed clinical reactivity in shrimp allergic patients although eliciting doses were much higher, indicating a lower potency (Broekman *et al.*, 2016). Risks of consumption of both insect and crustacean shellfish to dust mite allergic subjects has not been clinically confirmed, indications from epidemiological studies being that dust mite allergy is not a risk factor for allergy to crustacean shellfish.

GMOs expressing phosphomannose isomerase

A third example used for assessing the utility of the outputs of this study relate to assessing the allergenic risks of a newly expressed protein in a GMOs using phosphomannose isomerase (PMI) is a microbial protein which is used as a selectable marker of transformation during the development of GMO crops (Herman *et al.*, 2021) and has been the subject of an EFSA risk assessment in the past (European Food Safety Authority, 2009).

The levels of expression of PMI in the edible tissues (maize kernels) of the GMO crop were below the limits of detection of the ELISA employed which was 0.5 µg PMI protein/g tissue dry weight are not available in the published literature (European Food Safety Authority, 2009). This is of the order of 0.05 mg/Kg level similar to residual protein found in highly refined soybean oil (Rigby *et al.*, 2011) for which an allergenic ingredient labelling exemption has been accepted (European Food Safety Authority, 2007). *In vitro* digestibility tests on the PMI show that it was readily digested (European Food Safety Authority, 2009).

Potential to cause cross-reactive food allergies

Step 1: The PMI used originates from a microbial source – *Escherichia coli*, which is not related to any known food allergen.

Step 2: At the time of the original assessment undertaken by EFSA concerns were raised regarding potential allergenicity of PMI as it was a microbial member of the cupin superfamily (European Food Safety Authority, 2009). Structural bioinformatics assessment showed that it was not sufficiently closely related to allergenic bicupins from major allergenic foods such as peanut (Ara h 1, Ara h 3) to give cause for concern (European Food Safety Authority, 2009). Bioinformatics analysis of the PMI sequence identified an eight amino acid match against the allergenic α-parvalbumin from frog and cod (Herman *et al.*, 2021, European Food Safety Authority, 2009) (Figure 6). However, follow-up targeted serum screening has shown this level of homology to be irrelevant with regards cross-reactive allergies (European Food Safety Authority, 2009).

Based on the combination of expressions level, susceptibility to digestion and sequence analysis this protein has a very low likelihood of being allergenic either to cause cross-reactive allergies or initiate *de novo* sensitisation. This is borne by the lack of reported adverse reactions to GMO crops where the PMI has been used as a marker (Herman *et al.*, 2021).

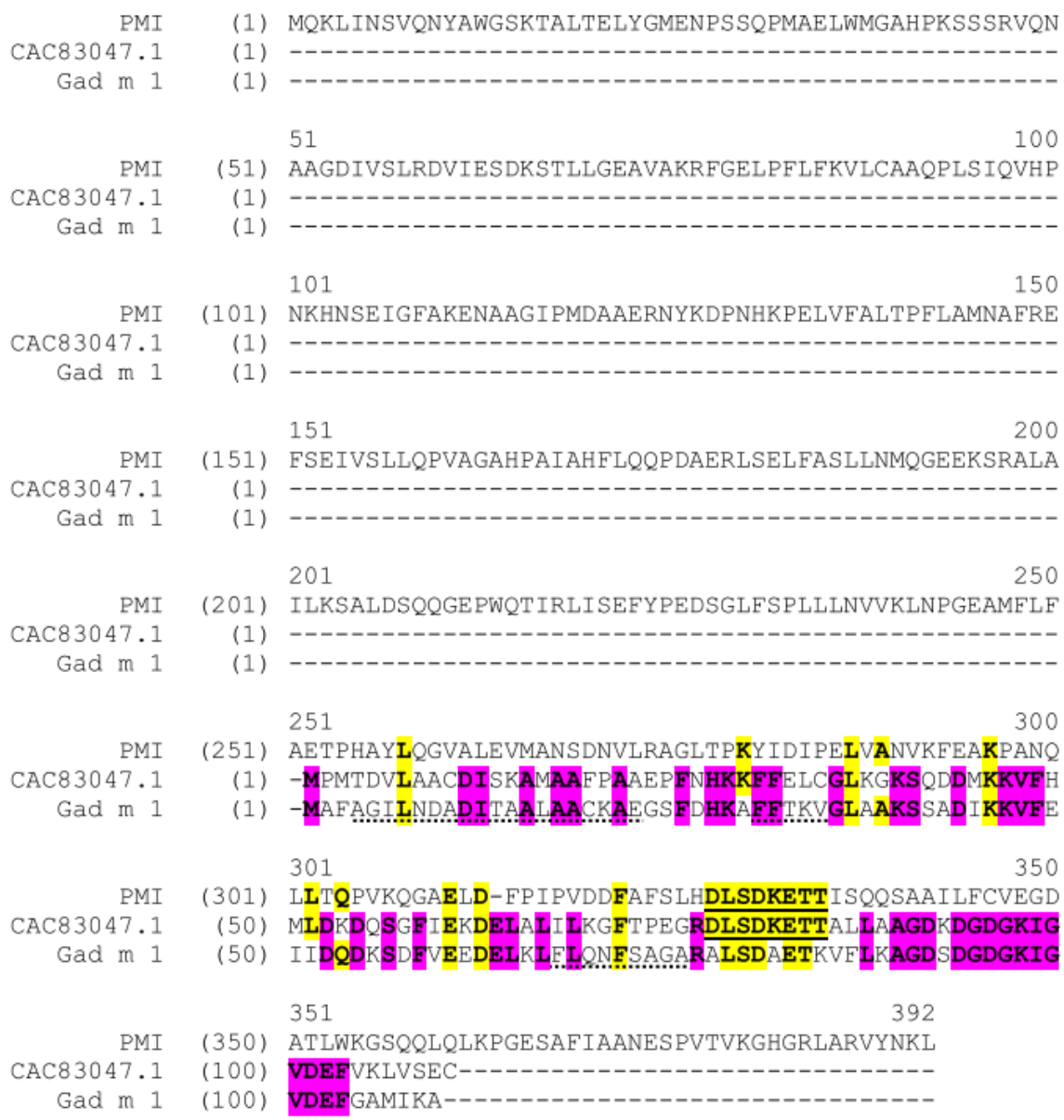

Figure 6: Aligned sequences of PMI and the allergenic parvalbumins from frog (CAC83047.1) and Atlantic cod (Gadus morhua, Gad m 1). Reproduced from (Herman et al., 2021)

3.5.2 SubTask 3.2: Develop the most suitable novel approach for the allergenicity assessment of innovative/novel proteins (in silico, *in vitro*, *in vivo* methods)

Following review of the initial proposed "next steps" described in section 3.2.3 some refinements were identified to further improve the outcomes of the allergenicity risk assessment (Figure 7) as follows:

Considering the level of protein present in the novel ingredient of GMOs

GMOs with very low levels of expression of a transgene could be identified as being of very low risk based on approaches adopted for exempting ingredients, such as highly refined soybean oil, from allergen labelling.

Assessing potential cross-reactivity risks

Phylogenetic analysis should be used to identify the relatedness of a novel proteins source to an existing allergenic food source which causes allergies in at least 0.5% of the European population. This will ensure the public health relevance of the assessment and supporting applicants in undertaking further steps as required by focussing on more accessible allergic patient populations.

Structural homology, in addition to sequence homology, should be considered in order to understand how variation affects the disposition of IgE epitopes and hence the potential for a new protein to display cross-reactive epitopes. Consideration should be paid to ensure the comparator allergens are expressed in the food as eaten and are relevant to the source of the novel protein ingredient.

Confirmation of potential cross-reactivity should be undertaken with patient samples from the allergenic comparator where patients have a confirmed food allergy. Serum binding studies are likely to over-estimate cross-reactivity due to asymptomatic sensitisation, frequently encountered in diagnosis of food allergy. Consideration should be paid to test methods which can address this issue, for example, by taking account of potential protective effects of IgG responses which may block IgE binding. Mediator release assays should be explored in follow-up serum screening as it has the potential to provide a superior readout to simple IgE-binding immunoassays.

In vivo studies, including oral food challenges, are essential for foods that present a significant risk of causing cross-reactive allergies since it will provide data on the potency and severity of likely reactions. Such information is needed by patients and healthcare professionals to inform patient management strategies as to whether the new food should be avoided, or not.

Assessing potential for development of new allergies

Data on the potential for a novel ingredient to cause allergies through exposure routes other than food should be used with caution since there are numerous instances where individuals can be sensitised to food ingredients by inhalation or contact but can safely eat the food.

Novel *in silico* (e.g., HLA binding assessment) and *in vitro* (cell-based) methods to assess the potential for proteins to be immunogenic are being developed and applied in vaccine research. They have great potential to provide insights into *de novo* sensitisation of food but are in their infancy and need validation for this application.

In vivo studies to confirm a lack of immunogenicity, as have been used in the past for novel proteins, provide additional assurance of safety.

Exposure

Data on proposed used of a novel ingredient including the level of protein incorporated in food products and the food categories can all be used to assess intake at a meal occasion. This is useful for the final part of the assessment where data on potential cross-reactivity and *de novo* sensitisation are integrated to provide an outcome as to the likelihood of allergenicity.

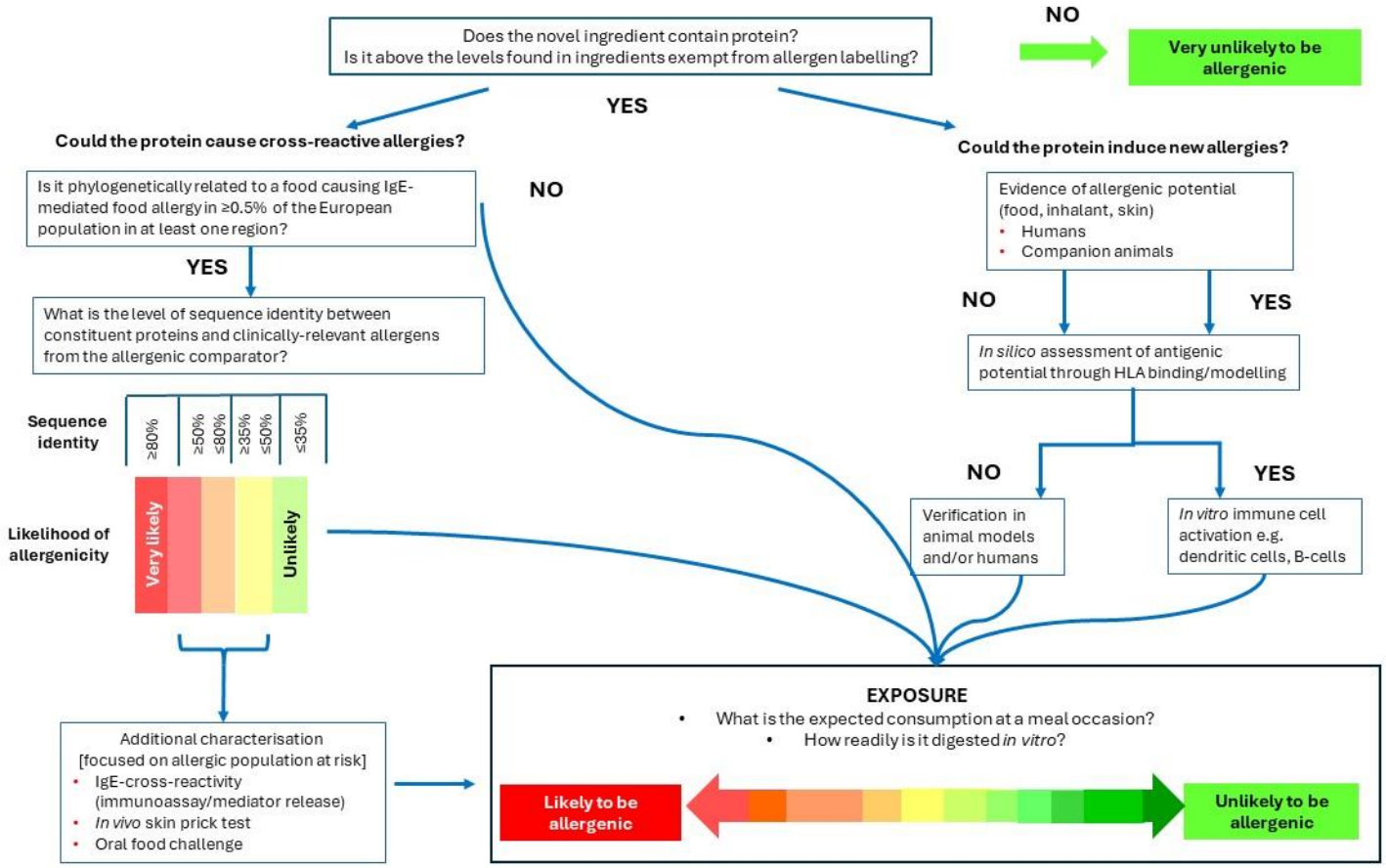

Figure 7: Suggested "next steps" for allergenicity risk assessment of novel proteins

4 Conclusions

Objective 1: Develop a ranking method for proteins with different allergenic potential according to their clinical relevance and screen existing tools to assess allergenicity risk of innovative/ novel proteins for use in subsequent activities.

A method for ranking proteins was developed which included two steps. The first was an analysis of the type of patient population and the quality of their food allergy diagnosis from whom biological samples (serum or immune cells) were taken and used to define the allergenicity of a protein molecule. In a second step analysis of data describing allergen quality and the test methods used to define allergenicity was undertaken. Other aspects such as the size of the patient population, and whether it included cases and controls, number of reports and geographic distribution were also considered in the ranking.

The ranking approach was applied to a total of 752 papers retrieved from the systematic searching which allowed the identification and classification of allergenic protein molecules from all the foods listed on Annex II of the food information for consumers regulation together with additional foods known to cause IgE-mediated food allergies in at least one European region with a prevalence of 0.5%, including fruits such as kiwi, apple and peach, legumes such as lentil and unusual foods, such as buckwheat.

Data were often found to be of poor quality, with poor descriptions of patient populations and a lack of data on allergen characterisation, with some publications describing allergen sequences without including any sequence accession. For example, some authors uploaded allergen sequences to, for example, Genbank for the allergen they characterised but no numbers or detailed information were provided in the publications. This made data difficult to analyse or resulted in the significance of findings being lost. In other cases, authors characterised allergens and uploaded sequences providing enough detail, but no patient data was used to confirm the allergenicity of the new molecules or they used patient sera without providing any further information. This does not guarantee the characterised allergens can cause an allergic reaction.

The best characterised clinically relevant allergens were identified in peanut, hazelnut, cow's milk, fish and crustacean shellfish. However, data were sparse for foods such as pecan, Macadamia, lupin and melon. Data quality was highly variable and many publications, especially relating to "allergen discovery" were of poor quality especially regarding patient panels. High quality patient populations were found in papers where allergens were being considered for use in component resolved diagnosis but there is a lack of transparency regarding the quality of the allergen components used in commercial diagnostic tests used in such studies. There is an urgent need to revisit approaches such as those developed in EuroPrevall to ensure effective studies are published with good quality data on patient populations linked to high quality allergen molecule characterisation and effective test methodology, found in papers such as (Kabasser *et al.*, 2021).

Many researchers rely on data included in databases such as that hosted by the IUIS allergen nomenclature committee (Pomés *et al.*, 2018). However, there are inconsistencies in the database entries and a lack of transparency regarding the submissions made to IUIS for inclusion of a new molecule. Indeed, there are also issues of historic entries where no update

is possible in the light of technological developments. Maintenance of such databases is acknowledged to be an ongoing issue (Radauer and Breiteneder, 2019) and it is clear unless adequate resourcing is provided the issues of inconsistencies will remain.

Objective 2: Investigate potential *in silico* tools and follow up actions (*in vitro* and/or *in vivo* methods) needed for an improved allergenicity assessment.

Most allergens were aligned and correctly predicted using AllergenONline and COMPARE using the full FASTA method or sliding 80mer window likely reflecting their inclusion in the AllergenONline and COMPASS databases. Of the novel bioinformatic tools assessed, AllergenFP often provided alignments with proteins which were difficult to interpret with regards their relevance to allergenicity risk assessment. The analysis of the different bioinformatic tools also highlighted a substantial difference in using the two possible algorithms in AlgPred for motif scanning. All allergens, with 3 exceptions noted so far, were predicted as being non-allergenic protein (with no hit) when using the MERCI algorithm, while they would be predicted as allergenic (with at least 1 hit) if using MEME/MAST. The exceptions were the highly clinically relevant allergens, α-S1-casein from domestic cattle (Bos domesticus), and the ω-5 gliadin Tri a 19 from wheat (Triticum aestivum) together with the β-amylase Tri a 17 from wheat which was classified as being of low clinical relevance. The AllerCatPro tool provided a good range of outputs that also sought to address issues such as 3D structure assessment and IgE-epitope analysis.

However, none of the tools were able to provide an output that could be linked to the clinical relevance score, and many suffered from identification of both false positive and false negative allergens. The assessment undertaken in this systematic review and the ranking for clinical relevance will help the developers of such methods to refine their outputs and improve ease of use. There is also an urgent need to develop approaches to their validation and clear use of both allergenic and non-allergenic comparators.

Objective 3: Develop a novel approach for allergenicity assessment of innovative/novel proteins by integrating *in silico*, *in vitro* and *in vivo* methods through implementation of the final ranking strategy of known allergens.

An approach has been identified which brings together elements of exposure assessment, alongside assessment of the risks posed by alternative protein ingredients of causing cross-reactive reactions in the existing allergic population, as well as causing the development of new food allergies. The latter aspect is more uncertain as the mechanisms underlying the development of food allergy are not fully elucidated and remain a matter of ongoing research. However, building on approaches for vaccine research, it will be possible to develop and validate combination of *in silico* and *in vitro* tests which can inform this part of the risk assessment process. But, they are in their infancy and will require validation before they can truly be incorporated into the allergenicity risk assessment process. This is in contrast to the exposure elements and the assessment of risks of cross-reactive allergies where *in silico* and *in vitro* test methods are better developed and only require refinement to improve the outputs to inform the allergenicity risk assessment process.

5 Recommendations

Recommendation 1: A consensus structure should be developed for reporting the initial characterisation and subsequent use of allergen molecules. A clearer way to present the results which will support their reuse is recommended as follows:

Patient details such as number of patients tested, where patients came from (in case authors were from different parts of the world) and other details such as age, gender, race, etc. need to be clearly described at the beginning of the paper or in a clear supplement. Patient information was constantly found in the results section or under charts or images which made the documents difficult to follow. In some cases, patient information was completely absent or patient status was described as 'allergic' was guaranteed by the authors but without giving details about patients or how they were diagnosed. When sera were purchased from laboratories, patient information was not presented which made data obtained from allergen molecules reacting to these sera difficult to analyse or lost significance. When publishing several papers using the same patients, it is convenient to describe the patients in each publication to present all data together or at least to provide a clear supplement and to clarify if serum samples from the same patients were used in repeated studies.

Allergen details need to be clearly stated early in the papers. When allergens were purchased, some authors did not specify the kind of allergen they were using (e.g., recombinant, native, purified, etc.). In some cases, only extracts were used which did not provide enough information on allergen quality and quantity. There were some papers where other foods were tested in 'parallel' to the main allergen presented. In these kinds of papers not much detail was given about the 'parallel' allergen tested, sometimes only mentioning 'extracts' but not clear whether they were purchased or made by them.

Results obtained need to be presented complete and clear. In the results sections present diagrams or tables with numbers of allergic, sensitised, etc. patients observed clearly presented for ease of reading and future data extraction using automated text mining tools.

Recommendation 2: Funding for the curation and maintenance of allergen sequence databases is an ongoing issue (Radauer and Breiteneder, 2019) and yet access to such curated allergen sequence sets is vital if the *in silico* aspects of allergenicity risk assessment are to improve. The systematic review undertaken within the framework of this current project has developed a curated set of sequences which have a clinical ranking which are essential for future refinement of bioinformatic tools. It is essential that the means be found to ensure these sequences and associated metadata are placed into a searchable repository and downloadable in forms tractable to sequence analysis and with a plan for their ongoing maintenance and updating. Having this undertaken in a manner which is transparent and assures the independence and trustworthiness of the data is essential to maintain consumer trust in allergenicity risk assessment processes.

Recommendation 3: Further development and refinement of existing and new bioinformatic tools for allergenicity prediction.

Analysis of multiple sequences needs to be made easier and quicker. By allowing the submission of multiple sequences at a time, the analysis of datasets will be made more time-effective and less prone to errors due to copying and pasting sequences and results from the analysis.

Output should be in a clear layout and exportable format. Output(s) from the bioinformatic tools, either binary answer to the question or score(s), should be provided in an easier readable format, ready to be downloaded or copied to the relevant documents and/or data backups.

Validation of *in silico* methods. This needs to be undertaken to assess both false positive and negative test results and outputs should be developed which can differentiate allergens of different clinical relevance.

Recommendation 4: Muti-omic methods – linking genomic and transcriptome data with the metabolome and proteome of an organism have made huge advances in the last 10 years. The cost of sequencing has been reduced and computational power has increased allowing such data sets to be analysed in a tractable form – and in some settings, such as molecular medicine – is becoming more routine. Building on initiatives such as the Earth Biogenome project (Lewin *et al.*, 2022) it is critical that reference genomes are developed for key bench mark allergenic food organisms, along with associated transcriptome and proteome data for edible tissues from those organisms. This will help to support safety, including allergenicity, risk assessment and assure rapid deployment of new assessment methodologies in a cost-effective and timely manner (Cattaneo *et al.*, 2023).

Recommendation 5: Further development and refinement of *in vitro* test methods, particularly cell-based methods, for both confirmation of allergenic potential to case cross-reactive allergies and initiate *de novo* sensitisation is urgently needed. Approaches to validation of such test methodology are also needed, which can also support identification of effective readouts which can inform the allergenicity risk assessment process in a meaningful way and can be related to clinical outcomes.

Recommendation 6: Review and refinement of the "next steps" approach is required to gain the input and consensus from different stakeholders, and importantly the clinical community and patient groups, to arrive at a consensus framework. This must provide usable, clinically relevant, outputs from the allergenicity risk assessment process which support implementation of effective risk management decisions and approaches which protect vulnerable allergic consumers in the population.

6 References

ABDELMOTELEB, M., ZHANG, C., FUREY, B., KOZUBAL, M., GRIFFITHS, H., CHAMPEAUD, M. & GOODMAN, R. E. 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**,** 111888.

ABOLHASSANI, M. & ROUX, K. H. 2009. cDNA Cloning, Expression and Characterization of an Allergenic 60s Ribosomal Protein of Almond (Prunus dulcis). *Iranian Journal of Allergy Asthma and Immunology,* 8**,** 77-84.

ADEL-PATIENT, K., NUTTEN, S., BERNARD, H., FRITSCHE, R., AH-LEUNG, S., MEZITI, N., PRIOULT, G., MERCENIER, A. & WAL, J. M. 2012. Immunomodulatory potential of partially hydrolyzed beta-lactoglobulin and large synthetic peptides. *J Agric Food Chem,* 60**,** 10858-66.

ADLER, L. N., JIANG, W., BHAMIDIPATI, K., MILLICAN, M., MACAUBAS, C., HUNG, S. C. & MELLINS, E. D. 2017. The Other Function: Class II-Restricted Antigen Presentation by B Cells. *Front Immunol,* 8**,** 319.

AKKERDAAS, J., TOTIS, M., BARNETT, B., BELL, E., DAVIS, T., EDRINGTON, T., GLENN, K., GRASER, G., HERMAN, R., KNULST, A., LADICS, G., MCCLAIN, S., POULSEN, L. K., RANJAN, R., RASCLE, J. B., SERRANO, H., SPEIJER, D., WANG, R., PEREIRA MOURIÈS, L., CAPT, A. & VAN REE, R. 2018. Protease resistance of food proteins: a mixed picture for predicting allergenicity but a useful tool for assessing exposure. *Clin Transl Allergy,* 8**,** 30.

AL JEWARI, C. & BALDAUF, S. L. 2023. An excavate root for the eukaryote tree of life. *Science Advances,* 9**,** eade4973.

ARMENTIA, A., DÍAZ-PERALES, A., CASTRODEZA, J., DUEÑAS-LAITA, A., PALACIN, A. & FERNÁNDEZ, S. 2009. Why can patients with baker's asthma tolerate wheat flour ingestion? Is wheat pollen allergy relevant? *Allergol Immunopathol (Madr),* 37**,** 203-4.

ASAI, Y., ESLAMI, A., VAN GINKEL, C. D., AKHABIR, L., WAN, M., YIN, D., ELLIS, G., BEN-SHOSHAN, M., MARENHOLZ, I., MARTINO, D., FERREIRA, M. A., ALLEN, K., MAZER, B., DE GROOT, H., DE JONG, N. W., GERTH VAN WIJK, R., DUBOIS, A. E. J., GROSCHE, S., ASHLEY, S., RÜSCHENDORF, F., KALB, B., BEYER, K., NÖTHEN, M. M., LEE, Y. A., CHIN, R., CHEUK, S., HOFFMAN, J., JORGENSEN, E., WITTE, J. S., MELLES, R. B., HONG, X., WANG, X., HUI, J., MUSK, A. W. B., HUNTER, M., JAMES, A. L., KOPPELMAN, G. H., SANDFORD, A. J., CLARKE, A. E. & DALEY, D. 2018. A Canadian genome-wide association study and meta-analysis confirm HLA as a risk factor for peanut allergy independent of asthma. *J Allergy Clin Immunol,* 141**,** 1513-1516.

ASARNOJ, A., NILSSON, C., LIDHOLM, J., GLAUMANN, S., ÖSTBLOM, E., HEDLIN, G., VAN HAGE, M., LILJA, G. & WICKMAN, M. 2012. Peanut component Ara h 8 sensitization and tolerance to peanut. *J Allergy Clin Immunol,* 130**,** 468-72.

ASERO, R., AMATO, S., ALFIERI, B., FOLLONI, S. & MISTRELLO, G. 2007. Rice: another potential cause of food allergy in patients sensitized to lipid transfer protein. *Int Arch Allergy Immunol,* 143**,** 69-74.

ASTWOOD, J. D., LEACH, J. N. & FUCHS, R. L. 1996a. Stability of food allergens to digestion in vitro. *Nat Biotechnol,* 14**,** 1269-73.

ASTWOOD, J. D., LEACH, J. N. & FUCHS, R. L. 1996b. Stability of food allergens to digestion in vitro. *Nature Biotechnology,* 14**,** 1269-1273.

BADERSCHNEIDER, B., CREVEL, R. W., EARL, L. K., LALLJIE, A., SANDERS, D. J. & SANDERS, I. J. 2002. Sequence analysis and resistance to pepsin hydrolysis as part of an assessment of the potential allergenicity of ice structuring protein type III HPLC 12. *Food Chem Toxicol,* 40**,** 965-78.

BAHRI, R., CUSTOVIC, A., KOROSEC, P., TSOUMANI, M., BARRON, M., WU, J., SAYERS, R., WEIMANN, A., RUIZ-GARCIA, M., PATEL, N., ROBB, A., SHAMJI, M. H., FONTANELLA, S., SILAR, M., MILLS, E. N. C., SIMPSON, A., TURNER, P. J. & BULFONE-PAUS, S. 2018. Mast cell activation test in the diagnosis of allergic disease and anaphylaxis. *J Allergy Clin Immunol,* 142**,** 485-496.e16.

BALLMER-WEBER, B. K., LIDHOLM, J., FERNANDEZ-RIVAS, M., SENEVIRATNE, S., HANSCHMANN, K. M., VOGEL, L., BURES, P., FRITSCHE, P., SUMMERS, C., KNULST, A. C., LE, T. M., REIG, I., PAPADOPOULOS, N. G., SINANIOTIS, A., BELOHLAVKOVA, S., POPOV, T., KRALIMARKOVA, T., DE BLAY, F., PUROHIT, A., CLAUSEN, M., KOWALSKI, M. L., ASERO, R., DUBAKIENE, R., BARREALES, L., MILLS, E. N. C., VAN REE, R. & VIETHS, S. 2015. IgE recognition patterns in peanut allergy are age dependent: perspectives of the EuroPrevall study (vol 70, pg 391, 2015). *Allergy,* 70**,** 725-725.

BARRE, A., BORGES, J. P., CULERRIER, R. & ROUGÉ, P. 2005. Homology modelling of the major peanut allergen Ara h 2 and surface mapping of IgE-binding epitopes. *Immunol Lett,* 100**,** 153-8.

BARRE, A., JACQUET, G., SORDET, C., CULERRIER, R. & ROUGÉ, P. 2007. Homology modelling and conformational analysis of IgE-binding epitopes of Ara h 3 and other legumin allergens with a cupin fold from tree nuts. *Mol Immunol,* 44**,** 3243-55.

BASTIAAN-NET, S., REITSMA, M., CORDEWENER, J. H. G., VAN DER VALK, J. P. M., AMERICA, T., DUBOIS, A. E. J., GERTH VAN WIJK, R., SAVELKOUL, H. F. J., DE JONG, N. W. & WICHERS, H. J. 2019. IgE Cross-Reactivity of Cashew Nut Allergens. *Int Arch Allergy Immunol,* 178**,** 19-32.

BAUERMEISTER, K., WANGORSCH, A., GAROFFO, L. P., REUTER, A., CONTI, A., TAYLOR, S. L., LIDHOLM, J., DEWITT, A. M., ENRIQUE, E., VIETHS, S., HOLZHAUSER, T., BALLMER-WEBER, B. & REESE, G. 2011. Generation of a comprehensive panel of crustacean allergens from the North Sea Shrimp Crangon crangon. *Mol Immunol,* 48**,** 1983-92.

BERNARD, H., GUILLON, B., DRUMARE, M. F., PATY, E., DRESKIN, S. C., WAL, J. M., ADEL-PATIENT, K. & HAZEBROUCK, S. 2015. Allergenicity of peanut component Ara h 2: Contribution of conformational versus linear hydroxyproline-containing epitopes. *J Allergy Clin Immunol,* 135**,** 1267-74.e1-8.

BIRD, P. I., TRAPANI, J. A. & VILLADANGOS, J. A. 2009. Endolysosomal proteases and their inhibitors in immunity. *Nature Reviews Immunology,* 9**,** 871-882.

BIROT, S., MADSEN, C. B., KRUIZINGA, A. G., CREPET, A., CHRISTENSEN, T. & BROCKHOFF, P. B. 2018. Food groups for allergen risk assessment: Combining food consumption data from different countries in Europe. *Food Chem Toxicol,* 118**,** 371-381.

BLOM, W. M., REMINGTON, B. C., BAUMERT, J. L., BUCCHINI, L., CREPET, A., CREVEL, R. W. R., MADSEN, C. B., TAYLOR, S. L., HOUBEN, G. F. & KRUIZINGA, A. G. 2019. Sensitivity analysis to derive a food consumption point estimate for deterministic food allergy risk assessment. *Food Chem Toxicol,* 125**,** 413-421.

BLOOM, K. A., HUANG, F. R., BENCHARITIWONG, R., BARDINA, L., ROSS, A., SAMPSON, H. A. & NOWAK-WEGRZYN, A. 2014. Effect of heat treatment on milk and egg proteins allergenicity. *Pediatr Allergy Immunol,* 25**,** 740-6.

BOGH, K. L., BARKHOLT, V. & MADSEN, C. B. 2015. Characterization of the Immunogenicity and Allergenicity of Two Cow's Milk Hydrolysates--A Study in Brown Norway Rats. *Scand J Immunol,* 81**,** 274-83.

BØGH, K. L. & MADSEN, C. B. 2016. Food Allergens: Is There a Correlation between Stability to Digestion and Allergenicity? *Crit Rev Food Sci Nutr,* 56**,** 1545-67.

BROEKMAN, H., VERHOECKX, K. C., DEN HARTOG JAGER, C. F., KRUIZINGA, A. G., PRONK-KLEINJAN, M., REMINGTON, B. C., BRUIJNZEEL-KOOMEN, C. A., HOUBEN, G. F. &

KNULST, A. C. 2016. Majority of shrimp-allergic patients are allergic to mealworm. *J Allergy Clin Immunol,* 137**,** 1261-1263.
BROUGH, H. A., CAUBET, J. C., MAZON, A., HADDAD, D., BERGMANN, M. M., WASSENBERG, J., PANETTA, V., GOURGEY, R., RADULOVIC, S., NIETO, M., SANTOS, A. F., NIETO, A., LACK, G. & EIGENMANN, P. A. 2020. Defining challenge-proven coexistent nut and sesame seed allergy: A prospective multicenter European study. *J Allergy Clin Immunol,* 145**,** 1231-1239.
BROUGH, H. A., SANTOS, A. F., MAKINSON, K., PENAGOS, M., STEPHENS, A. C., DOUIRI, A., FOX, A. T., DU TOIT, G., TURCANU, V. & LACK, G. 2013. Peanut protein in household dust is related to household peanut consumption and is biologically active. *J Allergy Clin Immunol,* 132**,** 630-638.
BUCAITE, G., KANG-PETTINGER, T., MOREIRA, J., GOULD, H. J., JAMES, L. K., SUTTON, B. J. & MCDONNELL, J. M. 2019. Interplay between Affinity and Valency in Effector Cell Degranulation: A Model System with Polcalcin Allergens and Human Patient-Derived IgE Antibodies. *J Immunol,* 203**,** 1693-1700.
BUCHANAN, B. B. & FRICK, O. L. 2002. The dog as a model for food allergy. *Ann N Y Acad Sci,* 964**,** 173-83.
CALCINAI, L., BONOMINI, M. G., LENI, G., FACCINI, A., PUXEDDU, I., GIANNINI, D., PETRELLI, F., PRANDI, B., SFORZA, S. & TEDESCHI, T. 2022. Effectiveness of enzymatic hydrolysis for reducing the allergenic potential of legume by-products. *Scientific Reports,* 12**,** 16902.
CALCINAI, L., PRANDI, B., FACCINI, A., PUXEDDU, I. & TEDESCHI, T. 2023. Molecular characterization and allergenicity assessment of different samples of Mung Bean. *Food Chemistry: X,* 20**,** 100980.
CALDERÓN, M. A., LINNEBERG, A., KLEINE-TEBBE, J., DE BLAY, F., HERNANDEZ FERNANDEZ DE ROJAS, D., VIRCHOW, J. C. & DEMOLY, P. 2015. Respiratory allergy caused by house dust mites: What do we really know? *Journal of Allergy and Clinical Immunology,* 136**,** 38-48.
CATTANEO, I., ASTUTO, M. C., BINAGLIA, M., DEVOS, Y., DORNE, J. L. C. M., FERNANDEZ AGUDO, A., FERNANDEZ DUMONT, A., GARCIA-VELLO, P., KASS, G. E. N., LANZONI, A., LIEM, A. K. D., PANZAREA, M., PARASKEVOPULOS, K., PARRA MORTE, J. M., TARAZONA, J. V. & TERRON, A. 2023. Implementing New Approach Methodologies (NAMs) in food safety assessments: Strategic objectives and actions taken by the European Food Safety Authority. *Trends in Food Science & Technology,* 133**,** 277-290.
CODEX ALIMENTARIUS COMMISSION 2003. Principles for the risk analysis of foods derived from modern biotechnology. FAO.
COMMISSION, C. A. 2009. *Foods derived from modern biotechnology*.
COMMITTEE FOR MEDICINAL PRODUCTS FOR HUMAN USE 2008. Guideline on allergen products: Produciton and quality issues. *In:* EUROPEAN MEDICINES AGENCY (ed.) *EMEA/CHMP/BWP/304831/2007.* London: European Medicines Agency,.
CREVEL, R. W., COOPER, K. J., POULSEN, L. K., HUMMELSHOJ, L., BINDSLEV-JENSEN, C., BURKS, A. W. & SAMPSON, H. A. 2007. Lack of immunogenicity of ice structuring protein type III HPLC12 preparation administered by the oral route to human volunteers. *Food Chem Toxicol,* 45**,** 79-87.
DANG, T. D., TANG, M., CHOO, S., LICCIARDI, P. V., KOPLIN, J. J., MARTIN, P. E., TAN, T., GURRIN, L. C., PONSONBY, A. L., TEY, D., ROBINSON, M., DHARMAGE, S. C. & ALLEN, K. J. 2012. Increasing the accuracy of peanut allergy diagnosis by using Ara h 2. *J Allergy Clin Immunol,* 129**,** 1056-63.
DATEMA, M. R., ZUIDMEER-JONGEJAN, L., ASERO, R., BARREALES, L., BELOHLAVKOVA, S., DE BLAY, F., BURES, P., CLAUSEN, M., DUBAKIENE, R., GISLASON, D., JEDRZEJCZAK-CZECHOWICZ, M., KOWALSKI, M. L., KNULST, A. C., KRALIMARKOVA, T., LE, T. M., LOVEGROVE, A., MARSH, J., PAPADOPOULOS, N. G., POPOV, T., DEL

PRADO, N., PUROHIT, A., REESE, G., REIG, I., SENEVIRATNE, S. L., SINANIOTIS, A., VERSTEEG, S. A., VIETHS, S., ZWINDERMAN, A. H., MILLS, C., LIDHOLM, J., HOFFMANN-SOMMERGRUBER, K., FERNANDEZ-RIVAS, M., BALLMER-WEBER, B. & VAN REE, R. 2015a. Hazelnut allergy across Europe dissected molecularly: A EuroPrevall outpatient clinic survey. *Journal of Allergy and Clinical Immunology,* 136**,** 382-391.

DATEMA, M. R., ZUIDMEER-JONGEJAN, L., ASERO, R., BARREALES, L., BELOHLAVKOVA, S., DE BLAY, F., BURES, P., CLAUSEN, M., DUBAKIENE, R., GISLASON, D., JEDRZEJCZAK-CZECHOWICZ, M., KOWALSKI, M. L., KNULST, A. C., KRALIMARKOVA, T., LE, T. M., LOVEGROVE, A., MARSH, J., PAPADOPOULOS, N. G., POPOV, T., DEL PRADO, N., PUROHIT, A., REESE, G., REIG, I., SENEVIRATNE, S. L., SINANIOTIS, A., VERSTEEG, S. A., VIETHS, S., ZWINDERMAN, A. H., MILLS, C., LIDHOLM, J., HOFFMANN-SOMMERGRUBER, K., FERNANDEZ-RIVAS, M., BALLMER-WEBER, B. & VAN REE, R. 2015b. Hazelnut allergy across Europe dissected molecularly: A EuroPrevall outpatient clinic survey. *J Allergy Clin Immunol,* 136**,** 382-91.

DE MARCHI, L., MAINENTE, F., LEONARDI, M., SCHEURER, S., WANGORSCH, A., MAHLER, V., PILOLLI, R., SORIO, D. & ZOCCATELLI, G. 2021. Allergenicity assessment of the edible cricket Acheta domesticus in terms of thermal and gastrointestinal processing and IgE cross-reactivity with shrimp. *Food Chem,* 359**,** 129878.

DEFERNEZ, M., MANDALARI, G. & MILLS, E. N. 2010. Quantitative assessment of multi-laboratory reproducibility of SDS-PAGE assays: Digestion pattern of beta-casein and beta-lactoglobulin under simulated conditions. *Electrophoresis,* 31**,** 2838-48.

DEWITT, A. M., MATTSSON, L., LAUER, I., REESE, G. & LIDHOLM, J. 2004. Recombinant tropomyosin from Penaeus aztecus (rPen a 1) for measurement of specific immunoglobulin E antibodies relevant in food allergy to crustaceans and other invertebrates. *Mol Nutr Food Res,* 48**,** 370-9.

DIMITROV, I., BANGOV, I., FLOWER, D. R. & DOYTCHINOVA, I. 2014a. AllerTOP v.2--a server for in silico prediction of allergens. *J Mol Model,* 20**,** 2278.

DIMITROV, I., NANEVA, L., DOYTCHINOVA, I. & BANGOV, I. 2014b. AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics,* 30**,** 846-851.

DOYLE, J. J. 2001. Leguminosae. *In:* BRENNER, S. & MILLER, J. H. (eds.) *Encyclopedia of Genetics.* New York: Academic Press.

EFSA 2007. Opinion of the Scientific Panel on Dietetic Products, Nutrition and Allergies on a request from the Commission related to a notification from FEDIOL and IMACE on fully refined soybean oil and fat pursuant to Article 6, paragraph 11 of Directive 2000/13/EC - for permanent exemption from labelling. *EFSA Journal,* 5**,** 570.

EFSA PANEL ON GENETICALLY MODIFIED ORGANISMS 2010. Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. *EFSA Journal,* 8**,** 1700.

EFSA PANEL ON NUTRITION, N. F., FOOD, A., TURCK, D., BOHN, T., CASTENMILLER, J., DE HENAUW, S., HIRSCH-ERNST, K. I., MACIUK, A., MANGELSDORF, I., MCARDLE, H. J., NASKA, A., PELAEZ, C., PENTIEVA, K., SIANI, A., THIES, F., TSABOURI, S., VINCETI, M., CUBADDA, F., FRENZEL, T., HEINONEN, M., MARCHELLI, R., NEUHÄUSER-BERTHOLD, M., POULSEN, M., MARADONA, M. P., SCHLATTER, J. R., VAN LOVEREN, H., AZZOLLINI, D. & KNUTSEN, H. K. 2022a. Safety of partially defatted house cricket (Acheta domesticus) powder as a novel food pursuant to Regulation (EU) 2015/2283. *EFSA Journal,* 20**,** e07258.

EFSA PANEL ON NUTRITION, N. F., FOOD, A., TURCK, D., BOHN, T., CASTENMILLER, J., DE HENAUW, S., HIRSCH-ERNST, K. I., MACIUK, A., MANGELSDORF, I., MCARDLE, H. J., NASKA, A., PELAEZ, C., PENTIEVA, K., SIANI, A., THIES, F., TSABOURI, S., VINCETI, M., CUBADDA, F., FRENZEL, T., HEINONEN, M., MARCHELLI, R., NEUHÄUSER-BERTHOLD, M., POULSEN, M., PRIETO MARADONA, M., SCHLATTER, J. R., VAN LOVEREN, H., VERVERIS, E. & KNUTSEN, H. K. 2022b. Safety of frozen and

freeze-dried formulations of the lesser mealworm (Alphitobius diaperinus larva) as a Novel food pursuant to Regulation (EU) 2015/2283. *EFSA Journal,* 20**,** e07325.

EFSA PANEL ON GENETICALLY MODIFIED ORGANISMS, NAEGELI, H., BIRCH, A. N., CASACUBERTA, J., DE SCHRIJVER, A., GRALAK, MIKOLAJ A., GUERCHE, P., JONES, H., MANACHINI, B., MESSÉAN, A., NIELSEN, E. E., NOGUÉ, F., ROBAGLIA, C., ROSTOKS, N., SWEET, J., TEBBE, C., VISIOLI, F., WAL, J.-M., EIGENMANN, P., EPSTEIN, M., HOFFMANN-SOMMERGRUBER, K., KONING, F., LOVIK, M., MILLS, C., MORENO, F. J., VAN LOVEREN, H., SELB, R. & FERNANDEZ DUMONT, A. 2017. Guidance on allergenicity assessment of genetically modified plants. *EFSA Journal,* 15**,** e04862-n/a.

EFSA PANEL ON GENETICALLY MODIFIED ORGANISMS, NAEGELI, H., BRESSON, J.-L., DALMAY, T., DEWHURST, I. C., EPSTEIN, M. M., FIRBANK, L. G., GUERCHE, P., HEJATKO, J., MORENO, F. J., MULLINS, E., NOGUÉ, F., ROSTOKS, N., SÁNCHEZ SERRANO, J. J., SAVOINI, G., VEROMANN, E., VERONESI, F. & DUMONT, A. F. 2021. Statement on in vitro protein digestibility tests in allergenicity and protein safety assessment of genetically modified plants. *EFSA Journal,* 19**,** e06350.

EFSA PANEL ON NUTRITION, N. F., ALLERGENS, F., TURCK, D., BOHN, T., CASTENMILLER, J., DE HENAUW, S., HIRSCH-ERNST, K. I., MACIUK, A., MANGELSDORF, I., MCARDLE, H. J., NASKA, A., PELAEZ, C., PENTIEVA, K., SIANI, A., THIES, F., TSABOURI, S., VINCETI, M., CUBADDA, F., FRENZEL, T., HEINONEN, M., MARADONA, M. P., MARCHELLI, R., NEUHÄUSER-BERTHOLD, M., POULSEN, M., SCHLATTER, J. R., VAN LOVEREN, H., FERNANDEZ, A. & KNUTSEN, H. K. 2021a. Safety of mung bean protein as a novel food pursuant to Regulation (EU) 2015/2283. *EFSA Journal,* 19**,** e06846.

EFSA PANEL ON NUTRITION, N. F., FOOD, A., TURCK, D., BOHN, T., CASTENMILLER, J., DE HENAUW, S., HIRSCH-ERNST, K. I., MACIUK, A., MANGELSDORF, I., MCARDLE, H. J., NASKA, A., PELAEZ, C., PENTIEVA, K., SIANI, A., THIES, F., TSABOURI, S., VINCETI, M., CUBADDA, F., FRENZEL, T., HEINONEN, M., MARCHELLI, R., NEUHÄUSER-BERTHOLD, M., POULSEN, M., PRIETO MARADONA, M., SCHLATTER, J. R., VAN LOVEREN, H., GOUMPERIS, T. & KNUTSEN, H. K. 2021b. Safety of frozen and dried formulations from whole house crickets (Acheta domesticus) as a Novel food pursuant to Regulation (EU) 2015/2283. *EFSA Journal,* 19**,** e06779.

EFSA PANEL ON NUTRITION, N. F., FOOD, ALLERGENS, TURCK, D., CASTENMILLER, J., DE HENAUW, S., HIRSCH-ERNST, K. I., KEARNEY, J., MACIUK, A., MANGELSDORF, I., MCARDLE, H. J., NASKA, A., PELAEZ, C., PENTIEVA, K., SIANI, A., THIES, F., TSABOURI, S., VINCETI, M., CUBADDA, F., FRENZEL, T., HEINONEN, M., MARCHELLI, R., NEUHÄUSER-BERTHOLD, M., POULSEN, M., PRIETO MARADONA, M., SCHLATTER, J. R., VAN LOVEREN, H., VERVERIS, E. & KNUTSEN, H. K. 2021c. Safety of dried yellow mealworm (Tenebrio molitor larva) as a novel food pursuant to Regulation (EU) 2015/2283. *EFSA Journal,* 19**,** e06343.

EUROPEAN FOOD SAFETY AUTHORITY 2009. Application (Reference EFSA-GMO-UK-2005-11) for the placing on the market of insect-resistant genetically modified maize MIR604 event, for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Syngenta Seeds S.A.S on behalf of Syngenta Crop Protection AG. *EFSA Journal,* 7**,** 1193.

EUROPEAN FOOD SAFETY AUTHORITY, PANEL O. G. M. O. 2007. Opinion of the Scientific Panel on Dietetic Products, Nutrition and Allergies on a request from the Commission related to a notification from FEDIOL and IMACE on fully refined soybean oil and fat pursuant to Article 6, paragraph 11 of Directive 2000/13/EC - for permanent exemption from labelling. *EFSA Journal,* 5**,** 570.

EUROPEAN PARLIAMENT, C. O. T. E. U. 2011. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No

1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004 Text with EEA relevance. *In:* SAFETY, D.-G. F. H. A. F. (ed.).

FAO-WHO 2022a. *Risk Assessment of Food Allergens. Part 1 – Review and validation of Codex Alimentarius priority allergen list through risk assessment. Meeting Report.* , FAO and WHO.

FAO-WHO 2022b. Risk Assessment of Food Allergens. Part 2: Review and establish threshold levels in foods for the priority allergens. Rome, Italy: FAO.

FAO-WHO 2024. Risk assessment of food allergens: part 4: establishing exemptions from mandatory declaration for priority food allergens: meeting report. Rome, Italy: FAO.

FERNANDEZ-RIVAS, M., BARREALES, L., MACKIE, A. R., FRITSCHE, P., VAZQUEZ-CORTES, S., JEDRZEJCZAK-CZECHOWICZ, M., KOWALSKI, M. L., CLAUSEN, M., GISLASON, D., SINANIOTIS, A., KOMPOTI, E., LE, T. M., KNULST, A. C., PUROHIT, A., DE BLAY, F., KRALIMARKOVA, T., POPOV, T., ASERO, R., BELOHLAVKOVA, S., SENEVIRATNE, S. L., DUBAKIENE, R., LIDHOLM, J., HOFFMANN-SOMMERGRUBER, K., BURNEY, P., CREVEL, R., BRILL, M., FERNANDEZ-PEREZ, C., VIETHS, S., MILLS, E. N. C., VAN REE, R. & BALLMER-WEBER, B. K. 2015. The EuroPrevall outpatient clinic study on food allergy: background and methodology. *Allergy,* 70**,** 576-584.

FERNANDEZ-RIVAS, M., BOLHAAR, S., GONZALEZ-MANCEBO, E., ASERO, R., VAN LEEUWEN, A., BOHLE, B., MA, Y., EBNER, C., RIGBY, N., SANCHO, A. I., MILES, S., ZUIDMEER, L., KNULST, A., BREITENEDER, H., MILLS, C., HOFFMANN-SOMMERGRUBER, K. & VAN REE, R. 2006. Apple allergy across Europe: how allergen sensitization profiles determine the clinical expression of allergies to plant foods. *J Allergy Clin Immunol,* 118**,** 481-8.

FERNANDEZ, A., MILLS, E. N. C., KONING, F. & MORENO, F. J. 2019. Safety Assessment of Immune-Mediated Adverse Reactions to Novel Food Proteins. *Trends Biotechnol,* 37**,** 796-800.

FIERS, M. W., KLETER, G. A., NIJLAND, H., PEIJNENBURG, A. A., NAP, J. P. & VAN HAM, R. C. 2004. Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics,* 5**,** 133.

FONG, A. T., DU TOIT, G., VERSTEEG, S. A. & VAN REE, R. 2019. Pink peppercorn: A cross-reactive risk for cashew- and pistachio-allergic patients. *J Allergy Clin Immunol Pract,* 7**,** 724-725.e1.

FU, T. J., ABBOTT, U. R. & HATZOS, C. 2002. Digestibility of food allergens and nonallergenic proteins in simulated gastric fluid and simulated intestinal fluid-a comparative study. *J Agric Food Chem,* 50**,** 7154-60.

FUKUNAGA, K., CHINUKI, Y., HAMADA, Y., FUKUTOMI, Y., SUGIYAMA, A., KISHIKAWA, R., FUKUNAGA, A., ODA, Y., UGAJIN, T., YOKOZEKI, H., HARADA, N., SUEHIRO, M., HIDE, M., NAKAGAWA, Y., NOGUCHI, E., NAKAMURA, M., MATSUNAGA, K., YAGAMI, A., MORITA, E. & MUSHIRODA, T. 2021. Genome-wide association study reveals an association between the HLA-DPB1(∗)02:01:02 allele and wheat-dependent exercise-induced anaphylaxis. *Am J Hum Genet,* 108**,** 1540-1548.

GOODMAN, R. E., EBISAWA, M., FERREIRA, F., SAMPSON, H. A., VAN REE, R., VIETHS, S., BAUMERT, J. L., BOHLE, B., LALITHAMBIKA, S., WISE, J. & TAYLOR, S. L. 2016. AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Mol Nutr Food Res,* 60**,** 1183-98.

GRABENHENRICH, L., TRENDELENBURG, V., BELLACH, J., YUREK, S., REICH, A., FIANDOR, A., RIVERO, D., SIGURDARDOTTIR, S., CLAUSEN, M., PAPADOPOULOS, N. G., XEPAPADAKI, P., SPRIKKELMAN, A. B., DONTJE, B., ROBERTS, G., GRIMSHAW, K., KOWALSKI, M. L., KUROWSKI, M., DUBAKIENE, R., RUDZEVICIENE, O., FERNANDEZ-RIVAS, M., COUCH, P., VERSTEEG, S. A., VAN REE, R., MILLS, C., KEIL,

T. & BEYER, K. 2020. Frequency of food allergy in school-aged children in eight European countries-The EuroPrevall-iFAAM birth cohort. *Allergy,* 75**,** 2294-2308.
GRABENHENRICH, L. B., REICH, A., BELLACH, J., TRENDELENBURG, V., SPRIKKELMAN, A. B., ROBERTS, G., GRIMSHAW, K. E., SIGURDARDOTTIR, S., KOWALSKI, M. L., PAPADOPOULOS, N. G., QUIRCE, S., DUBAKIENE, R., NIGGEMANN, B., FERNANDEZ-RIVAS, M., BALLMER-WEBER, B., VAN REE, R., SCHNADT, S., MILLS, E. N., KEIL, T. & BEYER, K. 2017. A new framework for the documentation and interpretation of oral food challenges in population-based and clinical research. *Allergy,* 72**,** 453-461.
GRIESMEIER, U., VÁZQUEZ-CORTÉS, S., BUBLIN, M., RADAUER, C., MA, Y., BRIZA, P., FERNÁNDEZ-RIVAS, M. & BREITENEDER, H. 2010. Expression levels of parvalbumins determine allergenicity of fish species. *Allergy,* 65**,** 191-8.
GUHSL, E. E., HOFSTETTER, G., HEMMER, W., EBNER, C., VIETHS, S., VOGEL, L., BREITENEDER, H. & RADAUER, C. 2014. Vig r 6, the cytokinin-specific binding protein from mung bean (Vigna radiata) sprouts, cross-reacts with Bet v 1-related allergens and binds IgE from birch pollen allergic patients' sera. *Mol Nutr Food Res,* 58**,** 625-34.
HERMAN, R. A., HOU, Z., MIRSKY, H., NELSON, M. E., MATHESIUS, C. A. & ROPER, J. M. 2021. History of safe exposure and bioinformatic assessment of phosphomannose-isomerase (PMI) for allergenic risk. *Transgenic Res,* 30**,** 201-206.
HILGER, C., GRIGIONI, F., THILL, L., MERTENS, L. & HENTGES, F. 2002. Severe IgE-mediated anaphylaxis following consumption of fried frog legs: definition of alpha-parvalbumin as the allergen in cause. *Allergy,* 57**,** 1053-8.
HOFER, H., WEIDINGER, T., BRIZA, P., ASAM, C., WOLF, M., TWAROCH, T. E., STOLZ, F., NEUBAUER, A., DALL, E., HAMMERL, P., JACQUET, A. & WALLNER, M. 2017. Comparing Proteolytic Fingerprints of Antigen-Presenting Cells during Allergen Processing. *Int J Mol Sci,* 18.
HOFFMANN-SOMMERGRUBER, K., MILLS, E. N. & VIETHS, S. 2008a. Coordinated and standardized production, purification and characterization of natural and recombinant food allergens to establish a food allergen library. *Mol Nutr Food Res,* 52 Suppl 2**,** S159-65.
HOFFMANN-SOMMERGRUBER, K., MILLS, E. N. C. & VIETHS, S. 2008b. Coordinated and standardized production, purification and characterization of natural and recombinant food allergens to establish a food allergen library. *Molecular Nutrition & Food Research,* 52**,** S159-S165.
HOLZHAUSER, T., WACKERMANN, O., BALLMER-WEBER, B. K., BINDSLEV-JENSEN, C., SCIBILIA, J., PERONO-GAROFFO, L., UTSUMI, S., POULSEN, L. K. & VIETHS, S. 2009. Soybean (Glycine max) allergy in Europe: Gly m 5 (beta-conglycinin) and Gly m 6 (glycinin) are potential diagnostic markers for severe allergic reactions to soy. *J Allergy Clin Immunol,* 123**,** 452-8.
HONG, X., HAO, K., LADD-ACOSTA, C., HANSEN, K. D., TSAI, H. J., LIU, X., XU, X., THORNTON, T. A., CARUSO, D., KEET, C. A., SUN, Y., WANG, G., LUO, W., KUMAR, R., FULEIHAN, R., SINGH, A. M., KIM, J. S., STORY, R. E., GUPTA, R. S., GAO, P., CHEN, Z., WALKER, S. O., BARTELL, T. R., BEATY, T. H., FALLIN, M. D., SCHLEIMER, R., HOLT, P. G., NADEAU, K. C., WOOD, R. A., PONGRACIC, J. A., WEEKS, D. E. & WANG, X. 2015. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. *Nat Commun,* 6**,** 6304.
HUANG, H.-J., SARZSINSZKY, E. & VRTALA, S. 2023. House dust mite allergy: The importance of house dust mite allergens for diagnosis and immunotherapy. *Molecular Immunology,* 158**,** 54-67.
HUG, L. A., BAKER, B. J., ANANTHARAMAN, K., BROWN, C. T., PROBST, A. J., CASTELLE, C. J., BUTTERFIELD, C. N., HERNSDORF, A. W., AMANO, Y., ISE, K., SUZUKI, Y., DUDEK, N., RELMAN, D. A., FINSTAD, K. M., AMUNDSON, R., THOMAS, B. C. & BANFIELD, J. F. 2016. A new view of the tree of life. *Nature Microbiology,* 1**,** 16048.

JAVED, B., PADFIELD, P., SPERRIN, M., SIMPSON, A. & MILLS, E. N. C. 2017. A protocol for a systematic review to identify allergenic tree nuts and the molecules responsible for their allergenic properties. *Food Chem Toxicol,* 106**,** 411-416.

JENKINS, J. A., BREITENEDER, H. & MILLS, E. N. 2007. Evolutionary distance from human homologs reflects allergenicity of animal food proteins. *J Allergy Clin Immunol,* 120**,** 1399-405.

JOHANSSON, S. G., HOURIHANE, J. O., BOUSQUET, J., BRUIJNZEEL-KOOMEN, C., DREBORG, S., HAAHTELA, T., KOWALSKI, M. L., MYGIND, N., RING, J., VAN CAUWENBERGE, P., VAN HAGE-HAMSTEN, M. & WÜTHRICH, B. 2001. A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy,* 56**,** 813-24.

JOHNSON, P. E., SAYERS, R. L., GETHINGS, L. A., BALASUNDARAM, A., MARSH, J. T., LANGRIDGE, J. I. & MILLS, E. N. 2016. Quantitative Proteomic Profiling of Peanut Allergens in Food Ingredients Used for Oral Food Challenges. *Anal Chem,* 88**,** 5689-95.

KABASSER, S., HAFNER, C., CHINTHRAJAH, S., SINDHER, S. B., KUMAR, D., KOST, L. E., LONG, A. J., NADEAU, K. C., BREITENEDER, H. & BUBLIN, M. 2021. Identification of Pru du 6 as a potential marker allergen for almond allergy. *Allergy,* 76**,** 1463-1472.

KALB, B., MARENHOLZ, I., JEANRENAUD, A., MEIXNER, L., ARNAU-SOLER, A., ROSILLO-SALAZAR, O. D., GHAURI, A., CIBIN, P., BLÜMCHEN, K., SCHLAGS, R., HANSEN, G., SEIDENBERG, J., KEIL, T., LAU, S., NIGGEMANN, B., BEYER, K. & LEE, Y. A. 2022. Filaggrin loss-of-function mutations are associated with persistence of egg and milk allergy. *J Allergy Clin Immunol,* 150**,** 1125-1134.

KALIC, T., MOREL-CODREANU, F., RADAUER, C., RUETHERS, T., TAKI, A. C., SWOBODA, I., HILGER, C., HOFFMANN-SOMMERGRUBER, K., OLLERT, M., HAFNER, C., LOPATA, A. L., MORISSET, M., BREITENEDER, H. & KUEHN, A. 2019. Patients Allergic to Fish Tolerate Ray Based on the Low Allergenicity of Its Parvalbumin. *J Allergy Clin Immunol Pract,* 7**,** 500-508.e11.

KANCHAN, K., GRINEK, S., BAHNSON, H. T., RUCZINSKI, I., SHANKAR, G., LARSON, D., DU TOIT, G., BARNES, K. C., SAMPSON, H. A., SUAREZ-FARINAS, M., LACK, G., NEPOM, G. T., CEROSALETTI, K. & MATHIAS, R. A. 2022a. HLA alleles and sustained peanut consumption promote IgG4 responses in subjects protected from peanut allergy. *J Clin Invest,* 132.

KANCHAN, K., SHANKAR, G., HUFFAKER, M. F., BAHNSON, H. T., CHINTHRAJAH, R. S., SANDA, S., MANOHAR, M., LING, H., PASCHALL, J. E., TOIT, G. D., RUCZINSKI, I., TOGIAS, A., LACK, G., NADEAU, K. C., JONES, S. M., NEPOM, G. T. & MATHIAS, R. A. 2022b. HLA-associated outcomes in peanut oral immunotherapy trials identify mechanistic and clinical determinants of therapeutic success. *Front Immunol,* 13**,** 941839.

KEIL, T., MCBRIDE, D., GRIMSHAW, K., NIGGEMANN, B., XEPAPADAKI, P., ZANNIKOS, K., SIGURDARDOTTIR, S. T., CLAUSEN, M., RECHE, M., PASCUAL, C., STANCZYK, A. P., KOWALSKI, M. L., DUBAKIENE, R., DRASUTIENE, G., ROBERTS, G., SCHOEMAKER, A. F., SPRIKKELMAN, A. B., FIOCCHI, A., MARTELLI, A., DUFOUR, S., HOURIHANE, J., KULIG, M., WJST, M., YAZDANBAKHSH, M., SZEPFALUSI, Z., VAN REE, R., WILLICH, S. N., WAHN, U., MILLS, E. N. & BEYER, K. 2010. The multinational birth cohort of EuroPrevall: background, aims and methods. *Allergy,* 65**,** 482-90.

KERSH, G. J., SALZER, J., JONES, E. S., BINDER, A. M., ARMSTRONG, P. A., CHOUDHARY, S. K., COMMINS, G. K., AMELIO, C. L., KATO, C. Y., SINGLETON, J., BIGGERSTAFF, B. J., BEARD, C. B., PETERSEN, L. R. & COMMINS, S. P. 2023. Tick bite as a risk factor for alpha-gal-specific immunoglobulin E antibodies and development of alpha-gal syndrome. *Ann Allergy Asthma Immunol,* 130**,** 472-478.

KLUEBER, J., COSTA, J., RANDOW, S., CODREANU-MOREL, F., VERHOECKX, K., BINDSLEV-JENSEN, C., OLLERT, M., HOFFMANN-SOMMERGRUBER, K., MORISSET, M.,

HOLZHAUSER, T. & KUEHN, A. 2020. Homologous tropomyosins from vertebrate and invertebrate: Recombinant calibrator proteins in functional biological assays for tropomyosin allergenicity assessment of novel animal foods. *Clin Exp Allergy,* 50**,** 105-116.

KOSTARA, M., CHONDROU, V., SGOUROU, A., DOUROS, K. & TSABOURI, S. 2020. HLA Polymorphisms and Food Allergy Predisposition. *J Pediatr Genet,* 9**,** 77-86.

KUMETA, H., NAKAYAMA, H. & OGURA, K. 2017. Solution structure of the major fish allergen parvalbumin Sco j 1 derived from the Pacific mackerel. *Sci Rep,* 7**,** 17160.

KUMMELING, I., MILLS, E. N., CLAUSEN, M., DUBAKIENE, R., PEREZ, C. F., FERNANDEZ-RIVAS, M., KNULST, A. C., KOWALSKI, M. L., LIDHOLM, J., LE, T. M., METZLER, C., MUSTAKOV, T., POPOV, T., POTTS, J., VAN REE, R., SAKELLARIOU, A., TONDURY, B., TZANNIS, K. & BURNEY, P. 2009. The EuroPrevall surveys on the prevalence of food allergies in children and adults: background and study methodology. *Allergy,* 64**,** 1493-7.

LEUNG, P. S., CHU, K. H., CHOW, W. K., ANSARI, A., BANDEA, C. I., KWAN, H. S., NAGY, S. M. & GERSHWIN, M. E. 1994. Cloning, expression, and primary structure of Metapenaeus ensis tropomyosin, the major heat-stable shrimp allergen. *J Allergy Clin Immunol,* 94**,** 882-90.

LEWIN, H. A., RICHARDS, S., LIEBERMAN AIDEN, E., ALLENDE, M. L., ARCHIBALD, J. M., BÁLINT, M., BARKER, K. B., BAUMGARTNER, B., BELOV, K., BERTORELLE, G., BLAXTER, M. L., CAI, J., CAPERELLO, N. D., CARLSON, K., CASTILLA-RUBIO, J. C., CHAW, S. M., CHEN, L., CHILDERS, A. K., CODDINGTON, J. A., CONDE, D. A., COROMINAS, M., CRANDALL, K. A., CRAWFORD, A. J., DIPALMA, F., DURBIN, R., EBENEZER, T. E., EDWARDS, S. V., FEDRIGO, O., FLICEK, P., FORMENTI, G., GIBBS, R. A., GILBERT, M. T. P., GOLDSTEIN, M. M., GRAVES, J. M., GREELY, H. T., GRIGORIEV, I. V., HACKETT, K. J., HALL, N., HAUSSLER, D., HELGEN, K. M., HOGG, C. J., ISOBE, S., JAKOBSEN, K. S., JANKE, A., JARVIS, E. D., JOHNSON, W. E., JONES, S. J. M., KARLSSON, E. K., KERSEY, P. J., KIM, J. H., KRESS, W. J., KURAKU, S., LAWNICZAK, M. K. N., LEEBENS-MACK, J. H., LI, X., LINDBLAD-TOH, K., LIU, X., LOPEZ, J. V., MARQUES-BONET, T., MAZARD, S., MAZET, J. A. K., MAZZONI, C. J., MYERS, E. W., O'NEILL, R. J., PAEZ, S., PARK, H., ROBINSON, G. E., ROQUET, C., RYDER, O. A., SABIR, J. S. M., SHAFFER, H. B., SHANK, T. M., SHERKOW, J. S., SOLTIS, P. S., TANG, B., TEDERSOO, L., ULIANO-SILVA, M., WANG, K., WEI, X., WETZER, R., WILSON, J. L., XU, X., YANG, H., YODER, A. D. & ZHANG, G. 2022. The Earth BioGenome Project 2020: Starting the clock. *Proc Natl Acad Sci U S A,* 119.

LI, Y., SACKETT, P. W., NIELSEN, M. & BARRA, C. 2023. NetAllergen, a random forest model integrating MHC-II presentation propensity for improved allergenicity prediction. *Bioinform Adv,* 3**,** vbad151.

LIN, N., CHI, H., NI, L., ZHANG, H. & LIU, Z. 2023. Study on the Sensitization and Antigenic Epitopes of Tropomyosin from Antarctic Krill (Euphausia superba). *J Agric Food Chem,* 71**,** 6445-6457.

LIU, R., HOLCK, A. L., YANG, E., LIU, C. & XUE, W. 2013. Tropomyosin from tilapia (Oreochromis mossambicus) as an allergen. *Clin Exp Allergy,* 43**,** 365-77.

LOZANO-OJALVO, D., BENEDÉ, S., ANTUNES, C. M., BAVARO, S. L., BOUCHAUD, G., COSTA, A., DENERY-PAPINI, S., DÍAZ-PERALES, A., GARRIDO-ARANDIA, M., GAVROVIC-JANKULOVIC, M., HAYEN, S., MARTÍNEZ-BLANCO, M., MOLINA, E., MONACI, L., PIETERS, R. H. H., VILLEMIN, C., WICHERS, H. J., WRÓBLEWSKA, B., WILLEMSEN, L. E. M., ROGGEN, E. L. & VAN BILSEN, J. H. M. 2019. Applying the adverse outcome pathway (AOP) for food sensitization to support in vitro testing strategies. *Trends in Food Science & Technology,* 85**,** 307-319.

LU, M., JIN, Y., CERNY, R., BALLMER-WEBER, B. & GOODMAN, R. E. 2018. Combining 2-DE immunoblots and mass spectrometry to identify putative soybean (Glycine max) allergens. *Food Chem Toxicol,* 116**,** 207-215.

LYONS, S. A., BURNEY, P. G. J., BALLMER-WEBER, B. K., FERNANDEZ-RIVAS, M., BARREALES, L., CLAUSEN, M., DUBAKIENE, R., FERNANDEZ-PEREZ, C., FRITSCHE, P., JEDRZEJCZAK-CZECHOWICZ, M., KOWALSKI, M. L., KRALIMARKOVA, T., KUMMELING, I., MUSTAKOV, T. B., LEBENS, A. F. M., VAN OS-MEDENDORP, H., PAPADOPOULOS, N. G., POPOV, T. A., SAKELLARIOU, A., WELSING, P. M. J., POTTS, J., MILLS, E. N. C., VAN REE, R., KNULST, A. C. & LE, T. M. 2019. Food Allergy in Adults: Substantial Variation in Prevalence and Causative Foods Across Europe. *J Allergy Clin Immunol Pract,* 7**,** 1920-1928.e11.

LYONS, S. A., CLAUSEN, M., KNULST, A. C., BALLMER-WEBER, B. K., FERNANDEZ-RIVAS, M., BARREALES, L., BIELI, C., DUBAKIENE, R., FERNANDEZ-PEREZ, C., JEDRZEJCZAK-CZECHOWICZ, M., KOWALSKI, M. L., KRALIMARKOVA, T., KUMMELING, I., MUSTAKOV, T. B., PAPADOPOULOS, N. G., POPOV, T. A., XEPAPADAKI, P., WELSING, P. M. J., POTTS, J., MILLS, E. N. C., VAN REE, R., BURNEY, P. G. J. & LE, T. M. 2020. Prevalence of Food Sensitization and Food Allergy in Children Across Europe. *J Allergy Clin Immunol Pract,* 8**,** 2736-2746 e9.

LYONS, S. A., DATEMA, M. R., LE, T. M., ASERO, R., BARREALES, L., BELOHLAVKOVA, S., DE BLAY, F., CLAUSEN, M., DUBAKIENE, R., FERNÁNDEZ-PEREZ, C., FRITSCHE, P., GISLASON, D., HOFFMANN-SOMMERGRUBER, K., JEDRZEJCZAK-CZECHOWICZ, M., JONGEJAN, L., KOWALSKI, M. L., KRALIMARKOVA, T. Z., LIDHOLM, J., PAPADOPOULOS, N. G., PONTOPPIDAN, B., POPOV, T. A., PRADO, N. D., PUROHIT, A., REIG, I., SENEVIRATNE, S. L., SINANIOTIS, A., VASSILOPOULOU, E., VERSTEEG, S. A., VIETHS, S., ZWINDERMAN, A. H., WELSING, P. M. J., MILLS, E. N. C., BALLMER-WEBER, B. K., KNULST, A. C., FERNÁNDEZ-RIVAS, M. & VAN REE, R. 2021a. Walnut Allergy Across Europe: Distribution of Allergen Sensitization Patterns and Prediction of Severity. *J Allergy Clin Immunol Pract,* 9**,** 225-235.e10.

LYONS, S. A., KNULST, A. C., BURNEY, P. G. J., FERNANDEZ-RIVAS, M., BALLMER-WEBER, B. K., BARREALES, L., BIELI, C., CLAUSEN, M., DUBAKIENE, R., FERNANDEZ-PEREZ, C., JEDRZEJCZAK-CZECHOWICZ, M., KOWALSKI, M. L., KUMMELING, I., KRALIMARKOVA, T., MUSTAKOV, T. B., VAN OS-MEDENDORP, H., PAPADOPOULOS, N. G., POPOV, T. A., POTTS, J., VERSTEEG, S. A., XEPAPADAKI, P., WELSING, P. M. J., MILLS, C., VAN REE, R. & LE, T. M. 2021b. Predicting food allergy: The value of patient history reinforced. *Allergy,* 76**,** 1454-1462.

MACKIE, A., DUPONT, D., TORCELLO-GÓMEZ, A., JARDIN, J. & DEGLAIRE, A. 2019. Report on EFSA project OC/EFSA/GMO/2017/01 "In vitro protein digestibility" (Allergestion). *EFSA Supporting Publications,* 16**,** 1765E.

MANDALARI, G., ADEL-PATIENT, K., BARKHOLT, V., BARO, C., BENNETT, L., BUBLIN, M., GAIER, S., GRASER, G., LADICS, G. S., MIERZEJEWSKA, D., VASSILOPOULOU, E., VISSERS, Y. M., ZUIDMEER, L., RIGBY, N. M., SALT, L. J., DEFERNEZ, M., MULHOLLAND, F., MACKIE, A. R., WICKHAM, M. S. & MILLS, E. N. 2009. In vitro digestibility of beta-casein and beta-lactoglobulin under simulated human gastric and duodenal conditions: a multi-laboratory evaluation. *Regul Toxicol Pharmacol,* 55**,** 372-81.

MAURER-STROH, S., KRUTZ, N. L., KERN, P. S., GUNALAN, V., NGUYEN, M. N., LIMVIPHUVADH, V., EISENHABER, F. & GERBERICK, G. F. 2019. AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics,* 35**,** 3020-3027.

MISRA, A., KUMAR, R., MISHRA, V., CHAUDHARI, B. P., RAISUDDIN, S., DAS, M. & DWIVEDI, P. D. 2011. Potential allergens of green gram (Vigna radiata L. Millsp) identified as members of cupin superfamily and seed albumin. *Clin Exp Allergy,* 41**,** 1157-68.

MITTAG, D., VIETHS, S., VOGEL, L., BECKER, W. M., RIHS, H. P., HELBLING, A., WÜTHRICH, B. & BALLMER-WEBER, B. K. 2004. Soybean allergy in patients allergic

to birch pollen: clinical investigation and molecular characterization of allergens. *J Allergy Clin Immunol,* 113**,** 148-54.
MITTAG, D., VIETHS, S., VOGEL, L., WAGNER-LOEW, D., STARKE, A., HUNZIKER, P., BECKER, W. M. & BALLMER-WEBER, B. K. 2005. Birch pollen-related food allergy to legumes: identification and characterization of the Bet v 1 homologue in mungbean (Vigna radiata), Vig r 1. *Clin Exp Allergy,* 35**,** 1049-55.
MONERET-VAUTRIN, D. A., GUÉRIN, L., KANNY, G., FLABBEE, J., FRÉMONT, S. & MORISSET, M. 1999. Cross-allergenicity of peanut and lupine: the risk of lupine allergy in patients allergic to peanuts. *J Allergy Clin Immunol,* 104**,** 883-8.
MONSALVE, R. I., GONZALEZ DE LA PEÑA, M. A., MENENDEZ-ARIAS, L., LOPEZ-OTIN, C., VILLALBA, M. & RODRIGUEZ, R. 1993. Characterization of a new oriental-mustard (Brassica juncea) allergen, Bra j IE: detection of an allergenic epitope. *Biochem J,* 293 (Pt 3)**,** 625-32.
MORAES, A. H., ACKERBAUER, D., KOSTADINOVA, M., BUBLIN, M., DE OLIVEIRA, G. A., FERREIRA, F., ALMEIDA, F. C., BREITENEDER, H. & VALENTE, A. P. 2014. Solution and high-pressure NMR studies of the structure, dynamics, and stability of the cross-reactive allergenic cod parvalbumin Gad m 1. *Proteins,* 82**,** 3032-42.
MUELLER, R. S. & UNTERER, S. 2018. Adverse food reactions: Pathogenesis, clinical signs, diagnosis and alternatives to elimination diets. *Vet J,* 236**,** 89-95.
MUH, H. C., TONG, J. C. & TAMMI, M. T. 2009. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. *PLoS One,* 4**,** e5861.
NESBIT, J. B., SCHEIN, C. H., BRAUN, B. A., GIPSON, S. A. Y., CHENG, H., HURLBURT, B. K. & MALEKI, S. J. 2020. Epitopes with similar physicochemical properties contribute to cross reactivity between peanut and tree nuts. *Mol Immunol,* 122**,** 223-231.
NGUYEN, M. N., KRUTZ, N. L., LIMVIPHUVADH, V., LOPATA, A. L., GERBERICK, G. F. & MAURER-STROH, S. 2022. AllerCatPro 2.0: a web server for predicting protein allergenicity potential. *Nucleic Acids Res,* 50**,** W36-w43.
NICOLAOU, N., MURRAY, C., BELGRAVE, D., POORAFSHAR, M., SIMPSON, A. & CUSTOVIC, A. 2011. Quantification of specific IgE to whole peanut extract and peanut components in prediction of peanut allergy. *J Allergy Clin Immunol,* 127**,** 684-5.
NITRIDE, C., MAMONE, G., PICARIELLO, G., MILLS, C., NOCERINO, R., BERNI CANANI, R. & FERRANTI, P. 2013. Proteomic and immunological characterization of a new food allergen from hazelnut (Corylus avellana). *J Proteomics,* 86**,** 16-26.
NOGUCHI, E., AKIYAMA, M., YAGAMI, A., HIROTA, T., OKADA, Y., KATO, Z., KISHIKAWA, R., FUKUTOMI, Y., HIDE, M., MORITA, E., AIHARA, M., HIRAGUN, M., CHINUKI, Y., OKABE, T., ITO, A., ADACHI, A., FUKUNAGA, A., KUBOTA, Y., AOKI, T., AOKI, Y., NISHIOKA, K., ADACHI, T., KANAZAWA, N., MIYAZAWA, H., SAKAI, H., KOZUKA, T., KITAMURA, H., HASHIZUME, H., KANEGANE, C., MASUDA, K., SUGIYAMA, K., TOKUDA, R., FURUTA, J., HIGASHIMOTO, I., KATO, A., SEISHIMA, M., TAJIRI, A., TOMURA, A., TANIGUCHI, H., KOJIMA, H., TANAKA, H., SAKAI, A., MORII, W., NAKAMURA, M., KAMATANI, Y., TAKAHASHI, A., KUBO, M., TAMARI, M., SAITO, H. & MATSUNAGA, K. 2019. HLA-DQ and RBFOX1 as susceptibility genes for an outbreak of hydrolyzed wheat allergy. *J Allergy Clin Immunol,* 144**,** 1354-1363.
NUTTEN, S., MAYNARD, F., JÄRVI, A., RYTZ, A., SIMONS, P. J., HEINE, R. G. & KUSLYS, M. 2020. Peptide size profile and residual immunogenic milk protein or peptide content in extensively hydrolyzed infant formulas. *Allergy,* 75**,** 1446-1449.
PASTORELLO, E. A., FARIOLI, L., PRAVETTONI, V., ISPANO, M., SCIBOLA, E., TRAMBAIOLI, C., GIUFFRIDA, M. G., ANSALONI, R., GODOVAC-ZIMMERMANN, J., CONTI, A., FORTUNATO, D. & ORTOLANI, C. 2000. The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. *J Allergy Clin Immunol,* 106**,** 744-51.

PASTORELLO, E. A., FARIOLI, L., PRAVETTONI, V., SCIBILIA, J., CONTI, A., FORTUNATO, D., BORGONOVO, L., BONOMI, S., PRIMAVESI, L. & BALLMER-WEBER, B. 2009. Maize food allergy: lipid-transfer proteins, endochitinases, and alpha-zein precursor are relevant maize allergens in double-blind placebo-controlled maize-challenge-positive patients. *Anal Bioanal Chem,* 395**,** 93-102.

PATIL, S. U., BUNYAVANICH, S. & BERIN, M. C. 2020. Emerging Food Allergy Biomarkers. *J Allergy Clin Immunol Pract,* 8**,** 2516-2524.

PEARSON, W. R. & LIPMAN, D. J. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A,* 85**,** 2444-8.

PEETERS, K. A., NORDLEE, J. A., PENNINKS, A. H., CHEN, L., GOODMAN, R. E., BRUIJNZEEL-KOOMEN, C. A., HEFLE, S. L., TAYLOR, S. L. & KNULST, A. C. 2007. Lupine allergy: not simply cross-reactivity with peanut or soy. *J Allergy Clin Immunol,* 120**,** 647-53.

PERRIN, P., JONGSMA, M. L., NEEFJES, J. & BERLIN, I. 2019. The labyrinth unfolds: architectural rearrangements of the endolysosomal system in antigen-presenting cells. *Curr Opin Immunol,* 58**,** 1-8.

POMÉS, A., DAVIES, J. M., GADERMAIER, G., HILGER, C., HOLZHAUSER, T., LIDHOLM, J., LOPATA, A. L., MUELLER, G. A., NANDY, A., RADAUER, C., CHAN, S. K., JAPPE, U., KLEINE-TEBBE, J., THOMAS, W. R., CHAPMAN, M. D., VAN HAGE, M., VAN REE, R., VIETHS, S., RAULF, M. & GOODMAN, R. E. 2018. WHO/IUIS Allergen Nomenclature: Providing a common language. *Mol Immunol,* 100**,** 3-13.

POULSEN, L. K. 2004. Allergy assessment of foods or ingredients derived from biotechnology, gene-modified organisms, or novel foods. *Mol Nutr Food Res,* 48**,** 413-23.

RADAUER, C. & BREITENEDER, H. 2019. Allergen databases-A critical evaluation. *Allergy,* 74**,** 2057-2060.

RIGBY, N. M., SANCHO, A. I., SALT, L. J., FOXALL, R., TAYLOR, S., RACZYNSKI, A., COCHRANE, S. A., CREVEL, R. W. & MILLS, E. N. 2011. Quantification and partial characterization of the residual protein in fully and partially refined commercial soybean oils. *J Agric Food Chem,* 59**,** 1752-9.

RUGGERI, M., BIANCHI, E., VIGANI, B., SÁNCHEZ-ESPEJO, R., SPANO, M., TOTARO FILA, C., MANNINA, L., VISERAS, C., ROSSI, S. & SANDRI, G. 2023. Nutritional and Functional Properties of Novel Italian Spray-Dried Cricket Powder. *Antioxidants (Basel),* 12.

SAKAMAKI, S., TAKAYANAGI, N., YOSHIZAKI, N., HAYASHI, S., TAKAYAMA, T., KATO, J., KOGAWA, K., YAMAUCHI, N., TAKEMOTO, N., NOBUOKA, A., AYABE, T., KOHGO, Y. & NIITSU, Y. 2000. Autoantibodies against the specific epitope of human tropomyosin(s) detected by a peptide based enzyme immunoassay in sera of patients with ulcerative colitis show antibody dependent cell mediated cytotoxicity against HLA-DPw9 transfected L cells. *Gut,* 47**,** 236-41.

SANDER, I., FLAGGE, A., MERGET, R., HALDER, T. M., MEYER, H. E. & BAUR, X. 2001. Identification of wheat flour allergens by means of 2-dimensional immunoblotting. *J Allergy Clin Immunol,* 107**,** 907-13.

SANDER, I., RIHS, H. P., DOEKES, G., QUIRCE, S., KROP, E., ROZYNEK, P., VAN KAMPEN, V., MERGET, R., MEURER, U., BRÜNING, T. & RAULF, M. 2015. Component-resolved diagnosis of baker's allergy based on specific IgE to recombinant wheat flour proteins. *J Allergy Clin Immunol,* 135**,** 1529-37.

SANTOS, A. F., ALPAN, O. & HOFFMANN, H. J. 2021a. Basophil activation test: Mechanisms and considerations for use in clinical trials and clinical practice. *Allergy,* 76**,** 2420-2432.

SANTOS, A. F., BERGMANN, M., BROUGH, H. A., COUTO-FRANCISCO, N., KWOK, M., PANETTA, V., HADDAD, D., LACK, G., EIGENMANN, P. & CAUBET, J. C. 2021b.

Basophil Activation Test Reduces Oral Food Challenges to Nuts and Sesame. *J Allergy Clin Immunol Pract,* 9**,** 2016-2027.e6.

SANTOS, A. F., DOUIRI, A., BÉCARES, N., WU, S. Y., STEPHENS, A., RADULOVIC, S., CHAN, S. M., FOX, A. T., DU TOIT, G., TURCANU, V. & LACK, G. 2014. Basophil activation test discriminates between allergy and tolerance in peanut-sensitized children. *J Allergy Clin Immunol,* 134**,** 645-52.

SANTOS, A. F., RIGGIONI, C., AGACHE, I., AKDIS, C. A., AKDIS, M., ALVAREZ-PEREA, A., ALVARO-LOZANO, M., BALLMER-WEBER, B., BARNI, S., BEYER, K., BINDSLEV-JENSEN, C., BROUGH, H. A., BUYUKTIRYAKI, B., CHU, D., DEL GIACCO, S., DUNN-GALVIN, A., EBERLEIN, B., EBISAWA, M., EIGENMANN, P., EIWEGGER, T., FEENEY, M., FERNANDEZ-RIVAS, M., FISHER, H. R., FLEISCHER, D. M., GIOVANNINI, M., GRAY, C., HOFFMANN-SOMMERGRUBER, K., HALKEN, S., HOURIHANE, J. O., JONES, C. J., JUTEL, M., KNOL, E., KONSTANTINOU, G. N., LACK, G., LAU, S., MARQUES MEJIAS, A., MARCHISOTTO, M. J., MEYER, R., MORTZ, C. G., MOYA, B., MURARO, A., NILSSON, C., LOPES DE OLIVEIRA, L. C., O'MAHONY, L., PAPADOPOULOS, N. G., PERRETT, K., PETERS, R. L., PODESTA, M., POULSEN, L. K., ROBERTS, G., SAMPSON, H. A., SCHWARZE, J., SMITH, P., THAM, E. H., UNTERSMAYR, E., VAN REE, R., VENTER, C., VICKERY, B. P., VLIEG-BOERSTRA, B., WERFEL, T., WORM, M., DU TOIT, G. & SKYPALA, I. 2023. EAACI guidelines on the diagnosis of IgE-mediated food allergy. *Allergy,* 78**,** 3057-3076.

SELB, R., WAL, J. M., MORENO, F. J., LOVIK, M., MILLS, C., HOFFMANN-SOMMERGRUBER, K. & FERNANDEZ, A. 2017. Assessment of endogenous allergenicity of genetically modified plants exemplified by soybean - Where do we stand? *Food Chem Toxicol,* 101**,** 139-148.

SHAN, L., QIAO, S.-W., ARENTZ-HANSEN, H., MOLBERG, Ø., GRAY, G. M., SOLLID, L. M. & KHOSLA, C. 2005. Identification and Analysis of Multivalent Proteolytically Resistant Peptides from Gluten: Implications for Celiac Sprue. *Journal of Proteome Research,* 4**,** 1732-1741.

SHARMA, N., PATIYAL, S., DHALL, A., PANDE, A., ARORA, C. & RAGHAVA, G. P. S. 2021a. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Briefings in Bioinformatics,* 22.

SHARMA, N., PATIYAL, S., DHALL, A., PANDE, A., ARORA, C. & RAGHAVA, G. P. S. 2021b. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. *Brief Bioinform,* 22.

SHAW, J., ROBERTS, G., GRIMSHAW, K., WHITE, S. & HOURIHANE, J. 2008. Short communication: Lupin allergy in peanut-allergic children and teenagers. *Allergy,* 63**,** 370-373.

SIEVERS, F. & HIGGINS, D. G. 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science,* 27**,** 135-145.

SMITS, M., MEIJERINK, M., LE, T. M., KNULST, A., DE JONG, A., CASPERS, M. P. M., LIMA, E. S., BABÉ, L., LADICS, G., MCCLAIN, S., HOUBEN, G. & VERHOECKX, K. 2021a. Predicting the allergenicity of legume proteins using a PBMC gene expression assay. *BMC Immunol,* 22**,** 27.

SMITS, M., VERHOECKX, K., KNULST, A., WELSING, P., DE JONG, A., GASPARI, M., EHLERS, A., VERHOEFF, P., HOUBEN, G. & LE, T. M. 2023. Co-sensitization between legumes is frequently seen, but variable and not always clinically relevant. *Front Allergy,* 4**,** 1115022.

SMITS, M., VERHOECKX, K., KNULST, A., WELSING, P., DE JONG, A., HOUBEN, G. & LE, T. M. 2021b. Ranking of 10 legumes according to the prevalence of sensitization as a parameter to characterize allergenic proteins. *Toxicol Rep,* 8**,** 767-773.

SOH, W. T., AGLAS, L., MUELLER, G. A., GILLES, S., WEISS, R., SCHEIBLHOFER, S., HUBER, S., SCHEIDT, T., THOMPSON, P. M., BRIZA, P., LONDON, R. E., TRAIDL-HOFFMANN, C., CABRELE, C., BRANDSTETTER, H. & FERREIRA, F. 2019. Multiple

roles of Bet v 1 ligands in allergen stabilization and modulation of endosomal protease activity. *Allergy,* 74**,** 2382-2393.
SØRENSEN, M., KUEHN, A., MILLS, E. N. C., COSTELLO, C. A., OLLERT, M., SMÅBREKKE, L., PRIMICERIO, R., WICKMAN, M. & KLINGENBERG, C. 2017. Cross-reactivity in fish allergy: A double-blind, placebo-controlled food-challenge trial. *J Allergy Clin Immunol,* 140**,** 1170-1172.
STOEVESANDT, J., STURM, G. J., BONADONNA, P., OUDE ELBERINK, J. N. G. & TRAUTMANN, A. 2020. Risk factors and indicators of severe systemic insect sting reactions. *Allergy,* 75**,** 535-545.
STONE, A. K., TANAKA, T. & NICKERSON, M. T. 2019. Protein quality and physicochemical properties of commercial cricket and mealworm powders. *J Food Sci Technol,* 56**,** 3355-3363.
SUDHARSON, S., KALIC, T., HAFNER, C. & BREITENEDER, H. 2021. Newly defined allergens in the WHO/IUIS Allergen Nomenclature Database during 01/2019-03/2021. *Allergy*.
SUPRUN, M., GETTS, R., RAGHUNATHAN, R., GRISHINA, G., WITMER, M., GIMENEZ, G., SAMPSON, H. A. & SUÁREZ-FARIÑAS, M. 2019. Novel Bead-Based Epitope Assay is a sensitive and reliable tool for profiling epitope-specific antibody repertoire in food allergy. *Sci Rep,* 9**,** 18425.
THE UNIPROT, C. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research,* 51**,** D523-D531.
THOMAS, K., AALBERS, M., BANNON, G. A., BARTELS, M., DEARMAN, R. J., ESDAILE, D. J., FU, T. J., GLATT, C. M., HADFIELD, N., HATZOS, C., HEFLE, S. L., HEYLINGS, J. R., GOODMAN, R. E., HENRY, B., HEROUET, C., HOLSAPPLE, M., LADICS, G. S., LANDRY, T. D., MACINTOSH, S. C., RICE, E. A., PRIVALLE, L. S., STEINER, H. Y., TESHIMA, R., VAN REE, R., WOOLHISER, M. & ZAWODNY, J. 2004. A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. *Regul Toxicol Pharmacol,* 39**,** 87-98.
TOO, J. J. Y., SHEK, L. P. & RAJAKULENDRAN, M. 2019. Cross-reactivity of pink peppercorn in cashew and pistachio allergic individuals. *Asia Pac Allergy,* 9**,** e25.
TORCELLO-GOMEZ, A., DUPONT, D., JARDIN, J., BRIARD-BION, V., DEGLAIRE, A., RISSE, K., MECHOULAN, E. & MACKIE, A. 2020a. Human gastrointestinal conditions affectin vitrodigestibility of peanut and bread proteins. *Food & Function,* 11**,** 6921-6932.
TORCELLO-GÓMEZ, A., DUPONT, D., JARDIN, J., BRIARD-BION, V., DEGLAIRE, A., RISSE, K., MECHOULAN, E. & MACKIE, A. 2020a. Human gastrointestinal conditions affect in vitro digestibility of peanut and bread proteins. *Food Funct,* 11**,** 6921-6932.
TORCELLO-GÓMEZ, A., DUPONT, D., JARDIN, J., BRIARD-BION, V., DEGLAIRE, A., RISSE, K., MECHOULAN, E. & MACKIE, A. 2020b. The pattern of peptides released from dairy and egg proteins is highly dependent on the simulated digestion scenario. *Food Funct,* 11**,** 5240-5256.
TORCELLO-GOMEZ, A., DUPONT, D., JARDIN, J., BRIARD-BION, V., DEGLAIRE, A., RISSE, K., MECHOULAN, E. & MACKIE, A. R. 2020b. The pattern of peptides released from dairy and egg proteins is highly dependent on the simulated digestion scenario. *Food & Function,* 11**,** 5240-5256.
VAN HOEYVELD, E. M., ESCALONA-MONGE, M., DE SWERT, L. F. & STEVENS, E. A. 1998. Allergenic and antigenic activity of peptide fragments in a whey hydrolysate formula. *Clin Exp Allergy,* 28**,** 1131-7.
VAN REE, R., SAPITER BALLERDA, D., BERIN, M. C., BEUF, L., CHANG, A., GADERMAIER, G., GUEVERA, P. A., HOFFMANN-SOMMERGRUBER, K., ISLAMOVIC, E., KOSKI, L., KOUGH, J., LADICS, G. S., MCCLAIN, S., MCKILLOP, K. A., MITCHELL-RYAN, S., NARROD, C. A., PEREIRA MOURIÈS, L., PETTIT, S., POULSEN, L. K., SILVANOVICH, A., SONG, P., TEUBER, S. S. & BOWMAN, C. 2021. The COMPARE Database: A Public Resource for Allergen Identification, Adapted for Continuous Improvement. *Front Allergy,* 2**,** 700533.

VENKATARAJAN, M. S. & BRAUN, W. 2001. New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical–chemical properties. *Molecular modeling annual,* 7**,** 445-453.
VEREDA, A., VAN HAGE, M., AHLSTEDT, S., IBAÑEZ, M. D., CUESTA-HERRANZ, J., VAN ODIJK, J., WICKMAN, M. & SAMPSON, H. A. 2011. Peanut allergy: Clinical and immunologic differences among patients from 3 different geographic regions. *J Allergy Clin Immunol,* 127**,** 603-7.
VERHOECKX, K. C., VAN BROEKHOVEN, S., DEN HARTOG-JAGER, C. F., GASPARI, M., DE JONG, G. A., WICHERS, H. J., VAN HOFFEN, E., HOUBEN, G. F. & KNULST, A. C. 2014. House dust mite (Der p 10) and crustacean allergic patients may react to food containing Yellow mealworm proteins. *Food Chem Toxicol,* 65**,** 364-73.
WANG, G., WU, T., NING, W., DIAO, K., SUN, X., WANG, J., WU, C., CHEN, J., XU, D. & LIU, X. S. 2023a. TLimmuno2: predicting MHC class II antigen immunogenicity through transfer learning. *Brief Bioinform,* 24.
WANG, K., CREVEL, R. W. R. & MILLS, E. N. C. 2022. Assessing protein digestibility in allergenicity risk assessment: A comparison of in silico and high throughput in vitro gastric digestion assays. *Food Chem Toxicol,* 167**,** 113273.
WANG, K., CREVEL, R. W. R. & MILLS, E. N. C. 2023b. An in vitro protocol to characterise the resistance of food proteins to intestinal digestion. *Food Chem Toxicol,* 173**,** 113590.
WANG, K., GALI-MOYA, J., RUANO-ZARAGOZA, M., CAIN, K., D'AURIA, G., DALY, M., BARRAN, P., CREVEL, R. & MILLS, E. N. C. 2023c. Bile salts enhance the susceptibility of the peach allergenic lipid transfer protein, Pru p 3, to in vitro gastrointestinal proteolysis. *Sci Rep,* 13**,** 15155.
WESTERHOUT, J., KRONE, T., SNIPPE, A., BABÉ, L., MCCLAIN, S., LADICS, G. S., HOUBEN, G. F. & VERHOECKX, K. C. 2019. Allergenicity prediction of novel and modified proteins: Not a mission impossible! Development of a Random Forest allergenicity prediction model. *Regul Toxicol Pharmacol,* 107**,** 104422.
WOLD, S., JONSSON, J., SJÖRSTRÖM, M., SANDBERG, M. & RÄNNAR, S. 1993. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. *Analytica Chimica Acta,* 277**,** 239-253.
XU, S., WANG, X. & FEI, C. 2022. A Highly Effective System for Predicting MHC-II Epitopes With Immunogenicity. *Front Oncol,* 12**,** 888556.
YANG, Z. & RANNALA, B. 2012. Molecular phylogenetics: principles and practice. *Nat Rev Genet,* 13**,** 303-14.
YUN, X., LI, M. S., CHEN, Y., HUAN, F., CAO, M. J., LAI, D., CHEN, G. X. & LIU, G. M. 2022. Characterization, Epitope Identification, and Cross-reactivity Analysis of Tropomyosin: An Important Allergen of Crassostrea angulata. *J Agric Food Chem,* 70**,** 9201-9213.
ZUURVELD, M., DÍAZ, C. B., REDEGELD, F., FOLKERTS, G., GARSSEN, J., VAN'T LAND, B. & WILLEMSEN, L. E. M. 2022. An advanced in vitro human mucosal immune model to predict food sensitizing allergenicity risk: A proof of concept using ovalbumin as model allergen. *Front Immunol,* 13**,** 1073034.