I51
An application to help users identify IgE cross reactivity based on sequence similarity and identity

Rishi Gupta

10th grade

non-RRI

introduction

The main objective of this project is to provide guidance to a user to determine what foods or airborne allergens to avoid based on their history of allergens and protein cross-reactivity. Interactions between antibodies called Immunoglobulin E (IgE) and allergenic proteins trigger allergic reactions. When an individual has produced antibodies against a particular allergenic protein, sometimes the antibodies can fail to distinguish between the original protein and other proteins usually from the same protein family. This phenomenon is called cross-reactivity. So an individual with positive allergy to one substance is at a risk for allergy from other proteins in the same protein family based on the cross reactivity between them. There is a lot of prior research surrounding IgE and cross reactivity but that research is too broad and not easily available in a manner to be used practically. We use results from cross reactivity research and analyze which data and protein family can be reliably used for allergen prediction. This prediction can be further refined for an individual based on their allergen profile. Though clinical skin prick tests(SPTs) can determine the susceptibility of individuals to specific allergens, the proteins that an individual is tested for are limited in scope. Testing individuals for a plethora of proteins has proven to be not very effective. So it is hard to determine if a protein that an individual has not been tested for nor been exposed to, can trigger an allergic reaction. SPT tests have other diagnostic challenges as negative skin tests have a high negative predictive accuracy than positive results and a person with a lower positive score on SPT may have had no experience of allergy to that substance. This app aims to provide an answer to the question of what foods to avoid by combining cross reactivity and SPT data and eliminating some of the constraints with this data by considering real allergic reaction experiences. The cross-reactivity risk score is based on protein sequence similarity and identities between proteins of the same protein family. The allergens are are mapped to their most common source of food, animal and plant based origin and classified as high risk, medium high risk, medium low risk and low risk. This analysis can be used to create avoidance guidelines and plans for immunotherapy.

methods - data acquisition and processing

We selected 3 protein families with the largest set of registered allergens: Profilins, nsLTP and PR-10. For each of these protein families, we generated identity and similarity score using the following steps

1. Acquire globally registered IUIS Genbank accession number from AllFam database, for all proteins in the family
2. Use CLUSTAL Omega to generate multiple sequence alignment of proteins
3. Use SIAS with CLUSTAL data to generate sequence similarity(S) and identities(I) using default parameters
4. Generate pairwise risk score for each protein using the formula Risk score = (I+S)/2
5. Store Risk score index map in mysql DB schema

   "CREATE TABLE IF NOT EXISTS crossreactivity (name VARCHAR(255), origin VARCHAR(255), %s)" % (header)”

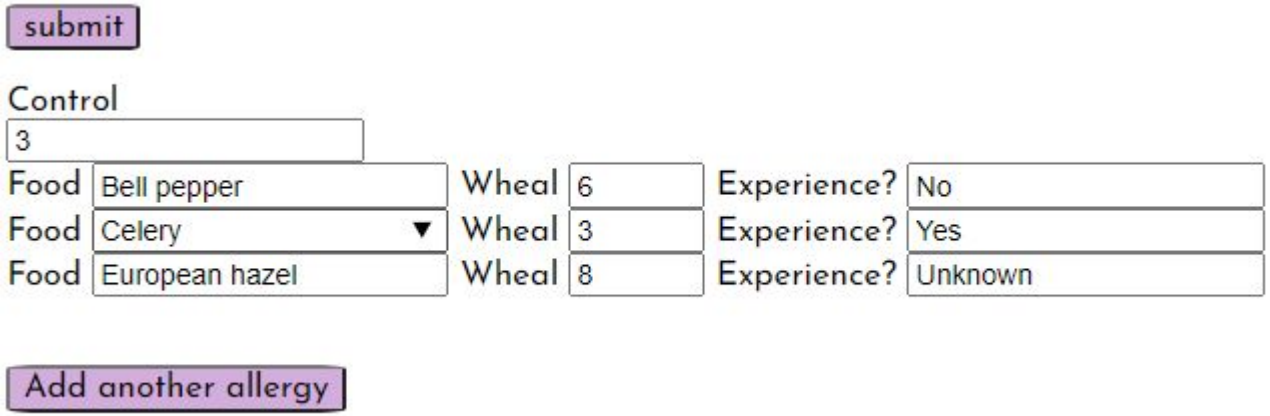
   where header is composed of all protein names in the family

methods - sample data with pairwise risk score

name	origin	P85524	BAA74451	CAA88831	AAB01092	AAQ91847	CAA42646	AAX19889	AHC08073	CAM31908	CAM31909	CAA50327	CAA47366	A
P85524	Kiwi fruit 1	100	25	23.33	24.995	24	24.33	24.665	26.995	26.66	26.33	28.33	26.33	
BAA74451	Mung bean	25	100	24.995	23.22	26.77	27.74	29.03	29.03	27.415	28.06	29.995	28.705	
CAA88831	Celery	23.33	24.995	100	84.415	44.8	44.8	44.15	47.075	52.59	54.215	47.075	45.125	
AAB01092	Carrot	24.995	23.22	84.415	100	41.395	42.085	41.61	42.185	50.31	51.585	44.06	41.815	
AAQ91847	Peanut	24	26.77	44.8	41.395	100	74.84	72.895	49.04	51.905	47.13	50.635	49.995	
CAA42646	Soybean	24.33	27.74	44.8	42.085	74.84	100	79.67	49.675	53.475	50.63	53.475	51.575	
AAX19889	Mung bean 1	24.665	29.03	44.15	41.61	72.895	79.67	100	48.06	49.03	48.38	48.06	46.12	
AHC08073	Tomato	26.995	29.03	47.075	42.185	49.04	49.675	48.06	100	54.71	53.18	49.06	47.795	
CAM31908	Gold kiwi fruit	26.66	27.415	52.59	50.31	51.905	53.475	49.03	54.71	100	74.2	57.86	52.825	
CAM31909	Kiwi fruit	26.33	28.06	54.215	51.585	47.13	50.63	48.38	53.18	74.2	100	55.09	49.995	
CAA50327	European hazel	28.33	29.995	47.075	44.06	50.635	53.475	48.06	49.06	57.86	55.09	100	90.25	
CAA47366	European hornbeam	26.33	28.705	45.125	41.815	49.995	51.575	46.12	47.795	52.825	49.995	90.25	100	
ADK39021	European hop hornbeam	26.33	28.385	46.425	43.435	51.265	52.845	47.415	49.37	55.97	53.18	92.495	96.535	

methods - cross reactivity predictions based on skin prick test(SPT)

- In clinical skin prick tests, a substance that may potentially cause allergies to an individual manifests as a wheal, the size of the wheal relative to a negative control wheal size is used as an indicator of allergic potential. With this tool Users are able to input data about multiple proteins from a skin prick test.
  - Input Wheal size for Control
  - Select individual substances (s) from a list of common name proteins in the protein family nsLTP as it contains one of the largest set of allergenic proteins. For substances that have multiple variants of proteins, we select one variant with a representative risk score
    - Input wheel size and any allergic reactions experienced on exposure to substances that have been tested



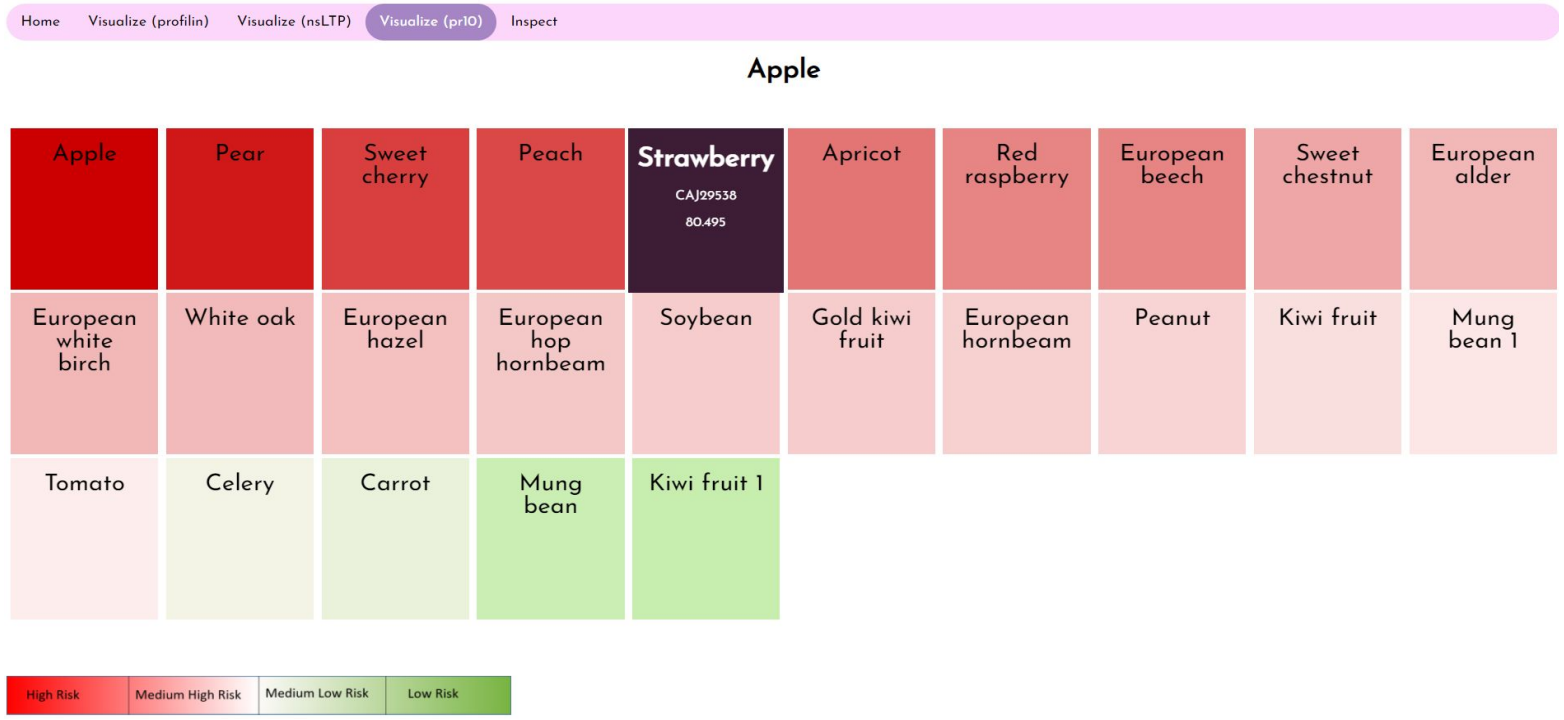
methods - cross reactivity predictions based on skin prick test(SPT)

We use the following heuristics to combine cross-reactivity score with SPT input to determine a risk category for allergens. Food allergens eliciting wheal diameters of at least 3 mm or larger than the negative control are considered positive test results. Negative skin tests have a high negative predictive accuracy, thus usually excluding food allergy to common foods. If there is a history of a no allergic reaction to a food and the skin test is positive, the test may be viewed as non diagnostic

- Filter out substances (s) that an individual is not allergic to.
  - If wheel size of s – control wheel size < 3mm , then that is considered a negative result . Exclude the substance s
  - If exposure by ingestion has not resulted in allergic reaction, then exclude the substance s
- Generate one risk score for each protein by picking max cross-reactivity risk score
  - crossReactivityRiskScore_p = max(crossReactivity_s_p) for each substance s in input list that has not been excluded
- Display predictions about potential allergies to substances as a sorted list based on risk score
  - Show cross-reactivity reaction potential in high, medium-high, medium-low and low risk categories

methods – IgE cross reactivity visualization

- Present a list of common substance names for proteins in the family. User can get more details about computed score and Genbank accession number
- Use php to allow user to select any protein substance in the family
- Use sql queries to retrieve cross reactivity data and display common names from proteins in order of highest potential cross reactivity with the selected protein, based on risk index



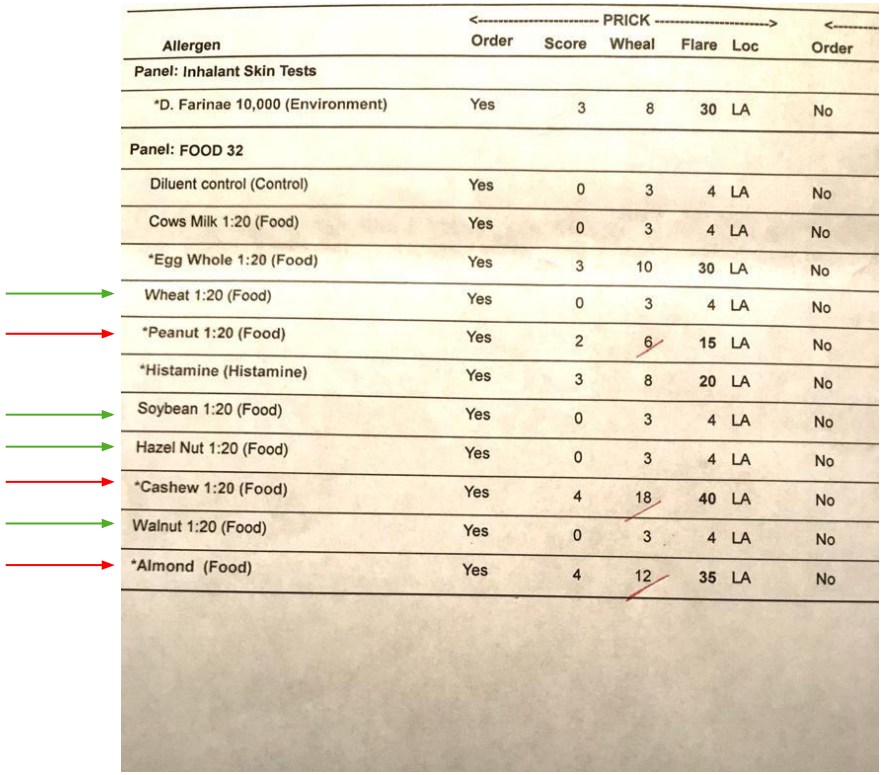
card for strawberry is flipped and is showing the computed score and the genbank accession number

results

Risk category based on SPT and cross-reactivity score from nsLTP

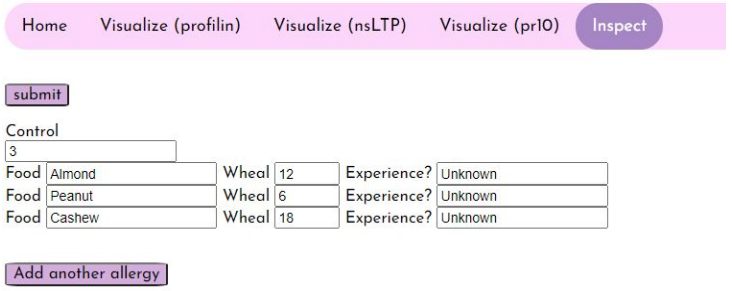
Clinical SPT data from an individual was used to validate the risk category generated by the application. Data was input for a subset of proteins that showed high wheal size compared to negative control. The risk category generated by the app for proteins corresponding to the ones in test panel show similar results. In the result set below, input data from 3 proteins marked by the red arrow with large wheal size generated correctly that the proteins that showed non allergenic in SPT test also show as non-cross reactive in the output generated by the app. So using nsLTP protein family with SPT can be a good way of deciding what proteins to avoid.

Clinical SPT results



Allergen	PRICK Order	Score	Wheal	Flare	Loc	Order
Panel: Inhalant Skin Tests						
*D. Farinae 10,000 (Environment)	Yes	3	8	30	LA	No
Panel: FOOD 32						
Diluent control (Control)	Yes	0	3	4	LA	No
Cows Milk 1:20 (Food)	Yes	0	3	4	LA	No
*Egg Whole 1:20 (Food)	Yes	3	10	30	LA	No
Wheat 1:20 (Food)	Yes	0	3	4	LA	No
*Peanut 1:20 (Food)	Yes	2	6	15	LA	No
*Histamine (Histamine)	Yes	3	8	20	LA	No
Soybean 1:20 (Food)	Yes	0	3	4	LA	No
Hazel Nut 1:20 (Food)	Yes	0	3	4	LA	No
*Cashew 1:20 (Food)	Yes	4	18	40	LA	No
Walnut 1:20 (Food)	Yes	0	3	4	LA	No
*Almond (Food)	Yes	4	12	35	LA	No

Input



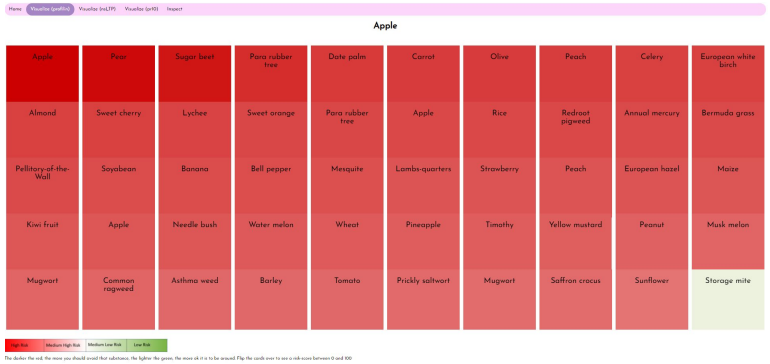
Output

results - contd.

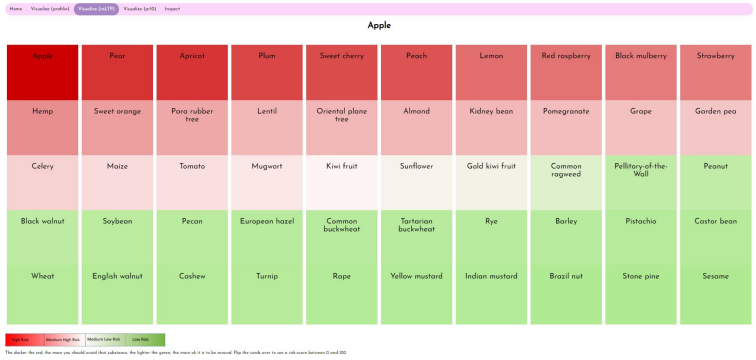
Comparison of cross reactivity across different protein family

This result compares the cross reactivity risk score for the same substance based on the presence of proteins from different protein families. As shown below, the proteins in the profilin family have strong cross-reactivity between almost all protein pairs. So this protein cross reactivity score is not very useful for determining potential for allergic reactions. However both PR10 and nsLTP cross reactivity scores show similar results for the same substance. They also show results inline with the expected relationships of allergies between substances like Almond and sweet cherry or apple and pear. So nsLTP protein family with it's larger size, is a good candidate for predicting potential allergens.

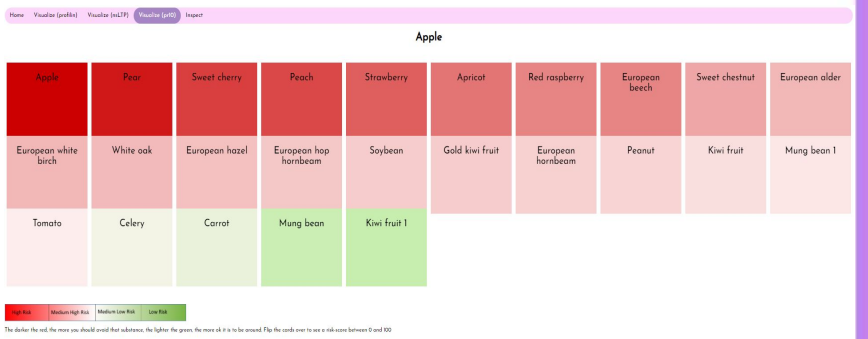
Profilin



nsLTP

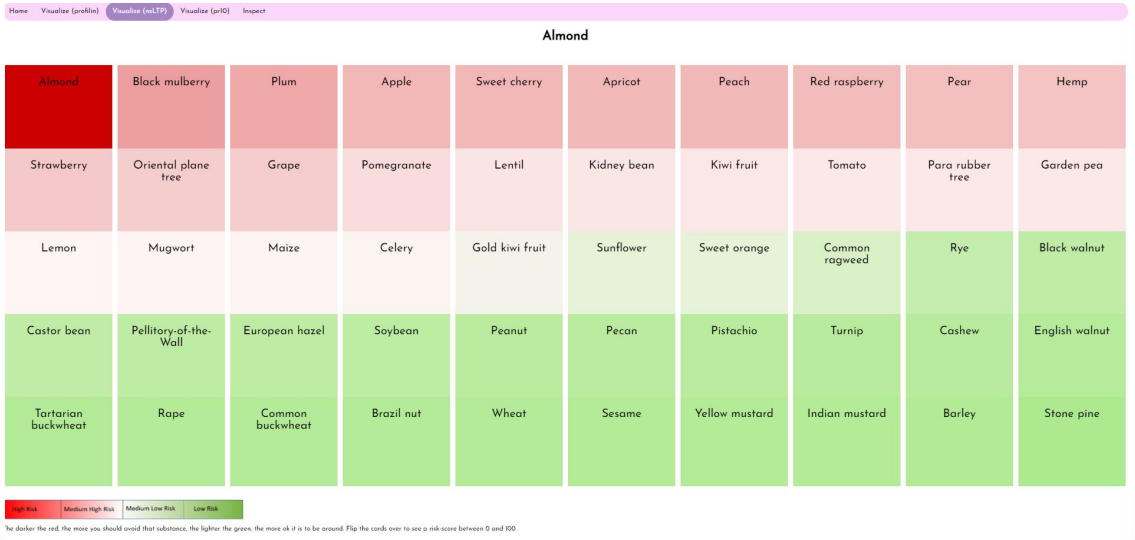


PR-10



Profilin

nsLTP



discussion

Research in cross-reactivity has produced a lot of studies that show how different proteins in substance can cause sensitization to other proteins. However since a substance can have multiple proteins belonging to different protein families with different protein structures and multiple cross reactivity maps, it is hard to use this data for providing guidance about real challenge of what substances to avoid. However if this data can be combined with other sources like an individual's SPT and their allergic experiences then a risk prediction can be made. Once challenge with the prediction is to provide judicious list of substances to avoid as a large list of substances to avoid may be impractical and counter productive. In the initial investigation, cross-reactivity results from multiple protein families for the same substance were used to generate this prediction. It resulted in a list that included almost 80% substances to be avoided. The reason for such a large list was that some proteins may provide a high cross reactivity score with all other protein of the family, so that protein family is not a good candidate to use for generating this guideline. The profilin family is proteins is one such family and we show that using that family does not generate useful guidance. So that family was excluded from the final analysis. Both nsLTP and pr10 proved to be good candidates for this predictive analysis. It was also observed that though using SPT data can refine the results, too large a data set of borderline skin prick test results can render the results useless. This is similar to conclusions from clinical SPT about the diagnostic confidence of low positive results. However combining SPT with data from real allergic experience can increase confidence in the predictions. The analysis of results from this tool was constrained by absence of SPT data from multiple sources. Availability of this data will help in establishing the accuracy of these results with more confidence.

conclusion

We show that using SPT and cross reactivity score to determine risk for allergenic proteins can be a useful tool for determine what foods to avoid and for creating immunotherapy plans. We also show that some protein families are inherently unsuited for food avoidance type usage because the proteins in the family are highly similar leading to a superset of potential allergens. Selection of numerous foods to be eliminated from the diet based on large numbers of poorly selected potential allergens may lead to diets that are difficult for families to follow and may eliminate foods that are clearly tolerated, making adherence to the diet poor. We analyzed various protein families which have a high number of allergens like Profilin, nsLTP and PR-10 and concluded that Profilin cross reactivity is a bad candidate for the reasons discussed above and since cross-reactivity results of PR-10 and nsLTP are consistent with each other and they have a large number of common allergens, using those families for avoidance guideline generates quality results. In addition SPT results need to be judiciously used as positive results are not always diagnostic and negative results are more reliable. Our application gives greater weight to negative results and to real experiences with allergic reaction based on exposure, to increase reliability of risk prediction. Combining SPT data for multiple allergens further increases the reliability of the result. This tool is a step in the right direction for allergies and accessibility. It can warn users about possible allergens and can be especially useful in avoiding serious allergic reactions by avoiding high risk substances until a formal diagnosis can occur. In situations where users experience allergies especially with food and cannot determine the cause, they can use this tool to identify potential allergens. Though our analysis was constrained by the availability of only one dataset for verification, it confirms our basic hypothesis.

references

Chruszcz M, Kapingidza AB, Dolamore C, Kowal K (2018) A robust method for the estimation and visualization of IgE cross-reactivity likelihood between allergens belonging to the same protein family. PLoS ONE 13(11): e0208276. https://doi.org/10.1371/journal.pone.0208276

Rob C. Aalberse, Structural biology of allergens,Journal of Allergy and Clinical Immunology, Volume 106, Issue 2,2000,Pages 228-238,ISSN 0091-6749, https://doi.org/10.1067/mai.2000.108434.

Tham, Elizabeth Huiwen, and Donald Y M Leung. “How Different Parts of the World Provide New Insights Into Food Allergy.” Allergy, asthma & immunology research vol.10,4 (2018): 290-299. doi.org10.4168/aair.2018.10.4.290

Aalberse, Rob C. “Assessment of allergen cross-reactivity.” Clinical and molecular allergy : CMA vol. 5 2. 9 Feb. 2007, doi:10.1186/1476-7961-5-2

Alessandri, C., Ferrara, R., Bernardi, M.L. et al. Molecular approach to a patient’s tailored diagnosis of the oral allergy syndrome. Clin Transl Allergy 10, 22 (2020),https://doi.org/10.1186/s13601-020-00329-8

Van der Valk, J.P.M., Gerth van Wijk, R., Hoorn, E. et al. Measurement and interpretation of skin prick test results. Clin Transl Allergy 6, 8 (2015),https://doi.org/10.1186/s13601-016-0092-0

https://www.sciencedirect.com/topics/medicine-and-dentistry/skin-test