Research Project:
A Preliminary Study to Elucidate Genes involved in Causing Soy allergies

Primary Investigator:
Mary Ann Hazuga

Program Project:
Identifying the Cause of Soy Allergies

Collaborative Investigators:
Gloria Xue
Specific Aims

Soy allergies have, in recent date, grown to become a significant problem around the world. The Food and Agriculture Organization of the United Nations has listed soy as one of the eight most significant food allergens.\(^1\) Even with this knowledge, very few studies have been done to identify the underlying cause of these allergies. This proposal is a part of a larger project that is seeking to find the causes of soy allergies. My colleague, Gloria Xue, will be studying the molecular compounds in soy that cause an allergic reaction, which has already been identified as an area of need for understanding soy allergies.\(^2\) For my part, this specific proposal focuses on human genetics and their role in causing soy allergies. Some studies have suggested that there are several possible genetic causes of allergies, many corresponding to the immune pathway.\(^3,4,5\) Unfortunately, no work has been done to try to identify the possible genetic causes of soy allergies in particular. If the underlying genetic causes of soy allergies were known, further study could be performed to better understand the pathway that causes the allergic reaction and the reasons for the difference between those with and without soy allergies.

Aim 1 will seek to find the specific single nucleotide polymorphisms (SNPs) that are possibly linked to soy allergies. This will be done using a small, preliminary genome-wide association study (GWAS) comparing individuals above age 18 with and without soy allergies. This will give many possible SNPs associated with soy allergy and provide a starting point for studying the genetic basis of the allergic reaction. Based on previous studies, we predict between 5 and 30 statistically significant SNPs will be identified.

Aim 2 will seek to understand the role of the SNPs identified and where more research still needs to be done. Once the GWAS has been performed, a literature study will be performed on the genes correlating to the statistically significant SNPs to see which have identified functions and which still need to be studied. We hypothesize that most genes identified in the GWAS will be related to the immune pathway.

The proposed studies will provide greater knowledge on the genetic basis of soy allergies in humans, which is entirely lacking in the current literature. Such a study will identify SNPs that may be connected to the allergy and their possible functions within humans. This will provide a basis to understanding the pathway that causes soy allergies. A few future studies would be warranted on the basis of this study. This project is just going to be a preliminary study, with a
small sample size, to help determine whether a large scale GWAS would be beneficial in studying the genetic causes of soy allergies. If successful, a future GWAS would have to be performed with more subjects to get a higher sensitivity. Also, any gene identified by the GWAS that has not yet been studied would need further scrutiny to determine its role in human health.

**Background**

This proposal is one part of a larger study seeking to understand the causes of soy allergies in humans. This portion focuses on the human genome specifically to identify the cause of this reaction in humans, while the other portion will look at the allergy from the perspective of the soybean, seeking the molecules that cause soybeans to initiate an allergic response. Finding the causes of soybean allergies could have a huge impact on worldwide use of this product. With soy being listed as one of the top eight allergens\(^1\) and the prevalence of soy allergies being about 1.5%,\(^6\) this is a real issue to many people globally. Especially because soy is finding its way into a lot of modern foods, including cereals, chili, and meats.\(^7\) Soy use is on the rise and people with allergies everywhere are feeling the impact.

An allergy is a hypersensitivity reaction initiated by immunologic mechanisms.\(^8\) More specifically, this happens through a combination of various proteins and cells that either fail to suppress the immune response or activate the response. Oral tolerance is how the body suppresses an immune response to food antigens.\(^9\) There are several systems in your body that aid this oral tolerance, including T cells, antigen presenting cells, and dendritic cells.\(^9\) Elimination or alteration of the controls for oral tolerance can cause an immune response. If this happens, T helper type 2 cells (Th2) will secrete interleukin (IL)-4 and IL-13.\(^10\) This promotes B cells to produce immunoglobulin E (IgE), which binds to receptors on mast cells.\(^10\) Once IgE has been bound to mast cells, the introduction of the food antigen will cause the mast cells to secrete histamine, leukotrienes, and prostaglandins, causing an allergic reaction.\(^10\)

As one can see, there are several steps involved in the pathway for an allergic reaction. Since the pathway for an allergic response is complex,\(^11\) it stands to reason that more than one gene mutation could lead to this result.\(^12\) Due to this, genome wide association studies (GWAS) have been used to study allergies. A GWAS involves the genomes of many subjects being analyzed for single nucleotide polymorphisms (SNPs).\(^13\) When doing a GWAS between a case and control population, statistically significant SNPs are normally in or near genes that have an
effect in causing the case phenotype.\textsuperscript{14} This method allows researchers to find many genes involved in the disease pathway at once.\textsuperscript{15} Due to the lack of knowledge about the genetic basis of soy allergies, this approach will allow for us to broadly identify many possible causal genes that elicit further study.

Thus far, there have been studies done on the genetic causes of allergies in general,\textsuperscript{3,4,5} but none on the genetic causes of soy allergies in particular. The studies mentioned above took into account food allergies, non-food allergies, asthma, and eczema, so their results may not be directly related to the genes responsible in soy allergies. Since there is no current knowledge about the genetic basis of soy allergies, this study will fill a gap in the research regarding this specific allergy. It will serve as a basis for future study on the mechanisms that cause soy allergy and the complex pathway involved with the response.

**Preliminary Research**

As mentioned in the background, there have been a few GWAS studies performed on allergies, although none have been specific to soy. This section will briefly overview the focus of three genome wide association studies related to allergies and their results. Although these studies were much less specific than the one proposed, they were able to successfully identify SNPs associated with allergies, so their success lends credit to the type of research that this study will seek to perform.

In a study performed by Hinds et al. in 2013, 53,862 individuals with pollen, dust mite or cat allergies were screened.\textsuperscript{3} The GWAS identified 16 possible susceptibility loci.\textsuperscript{3} Many of these had already been associated with asthma, suggesting a possible common pathway for these reactions.\textsuperscript{3} Many of these were also involved in the immune pathway, supporting the assumption that allergic reaction is caused by a dysfunction in the immune pathway.\textsuperscript{3}

The study performed by Bønnelykke et al. was a large scale study looking at allergic sensitization for common inhalant and food allergies.\textsuperscript{4} Around 10,000 members in the case study were compared to about 20,000 in the control group.\textsuperscript{4} From this, 10 susceptibility loci were identified, which, according to Bønnelykke, could account for 25\% of the cause of allergic sensitization.\textsuperscript{4}
Another study, performed by Ramasamy et al., focused on allergic rhinitis and grass sensitization specifically. They did a GWAS on approximately 4,000 case and 9,000 control for allergic rhinitis and approximately 2,300 case and 10,000 control for grass sensitization. This study identified three possible susceptibility loci from these comparisons. The result is much less than the other two studies, but the range of allergies tested was also much narrower. This could mean that either this GWAS was not able to capture some susceptibility loci, or that different allergies may be caused by SNPs in different areas.

**Research Design and Methods**

*Design*

The first part of this proposal involves performing a small GWAS to find SNPs that might contribute to causing soy allergies as well as to evaluate whether a large-scale GWAS would be worthwhile to conduct. The first step in this study will be to find appropriate and willing participants from which to collect samples. We hope to collect samples from 200 case and 200 control participants for analysis. A proposed total of 400 participants would allow us to collect enough data to decide whether studying soy allergies through GWAS is valuable, while still keeping the amount of needed participation relatively low. For these participants to be qualified to take part in the study, they must be over 18 years of age and the case participants must be allergic to soy. The control patients would ideally have no allergies.

Once participants have been identified, cheek swabs will be taken from each participant to collect the needed DNA for the experiment. These samples will then need to be purified and amplified for analysis. The samples will be analyzed using the affymetrix Axiom microarray plates and technology. From there, the results will be statistically analyzed using the software Haploveiw to find the SNPs that reach statistical significance. We predict we will find between 5 and 30 SNPs that will reach significance. However, our sample size is limited since this a preliminary experiment, so it may be more difficult to recognize statistically significant SNPs.

Once these SNPs have been identified, an extensive literature search will be performed on the genes associated with them to identify their function. This will provide us with more information on the reasons soy allergies occur and what pathway is affected to cause the reaction. We predict that most SNPs identified in the GWAS will be found to function in the
immune pathway, since allergic reactions are caused by an immune response to an otherwise harmless antigen.\textsuperscript{17}

\textit{Approach}

There were many considerations that had to be made when first designing the experiment. We decided that we would use a GWAS because it will allow us to find the SNPs that cause soy allergies without having a prior hypothesis as to what those genes are.\textsuperscript{18} This will allow us to study the entire genome for SNPs that may not be where we expect, broadening the potential range of results. When collecting candidates for the study, there are a few reasons we wish to find subjects 18 years of age and older. First, about half of all children with soy allergies will grow out of them before this age.\textsuperscript{19} Second, this will allow our subjects to consent to the trial for themselves, as they are legal adults. When it comes to the technology needed to analyze the genomes for the study, we chose the Affymetrix Axiom microarray plates because they provide reasonable sensitivity and coverage at a fairly low cost.\textsuperscript{20} Since this is just a preliminary study, we do not need to use the most high-powered microarray plates on the market. Their competitor, Illumina, provides plates that have a higher sensitivity, but they are more expensive and analyzing the results is more time consuming,\textsuperscript{20} making them impractical for this kind of study. Finally, we are choosing to use the Haploview software to analyze the results of the study because it is user-friendly and provides marker quality statistics, LD information, haplotype blocks, population haplotype frequencies and single marker association statistics.\textsuperscript{21}

\textit{Expected Results and Potential Problems}

From the small scale GWAS this study seeks to perform, we are expecting to find between 5 and 30 SNPs that reach statistical significance. When doing the literature search, we expect that the genes related to these SNPs will be found mostly in the immune pathway. While this is the ideal outcome for this study, there are several problems that could come up. First, we may not be able to find enough participants for the study, even though we are keeping the sample size small. The prevalence of soy allergies is about 1.5%, including children.\textsuperscript{6} Finding enough people 18 and older who are allergic to soy and willing to participate could therefore propose a challenge. These participants would ideally have the same gender, race, and age ratio from case to control to minimize error.\textsuperscript{22} Also, this study is proposing an extremely small sample size, as it is a preliminary project. Due to this, we expect it will be more difficult to discern which SNPs
are actually significant and which seem significant due to random chance.\textsuperscript{23} Unfortunately, this is the nature of small studies and is why, if successful, another, larger GWAS would need to be performed. This will give us a higher sensitivity and more confidence in the results. Another potential problem with GWAS is that not all SNPs identified may be near known genes, or they may be near multiple genes.\textsuperscript{16, 24, 25} In these cases, it is difficult to discern which gene the SNP is affecting. In such cases, expression quantitative trait loci (eQTLs) can be performed using methods like eQTL electrical diagrams (eQED) to associate the SNPs with the gene they most likely interact with.\textsuperscript{25}

\textit{Full Project}

This proposal is a part of a larger project that is seeking to find the causes of soy allergies. My colleague, Gloria Xue, is looking at the soybean itself, seeking to discern what in soy causes an allergic reaction. My project is focused more on the human aspect, looking at potential genetic causes of soy sensitization and allergies. These two parts come together to form a more complete picture of soy allergies, from both the perspective of human reactions to soy and the perspective of the soy product itself. With these results, we will expand current understanding of how soy allergies work, what causes them, and how we could potentially prevent cases of allergic reaction.
References


