Agency for Science, Technology and Research
SINGAPORE

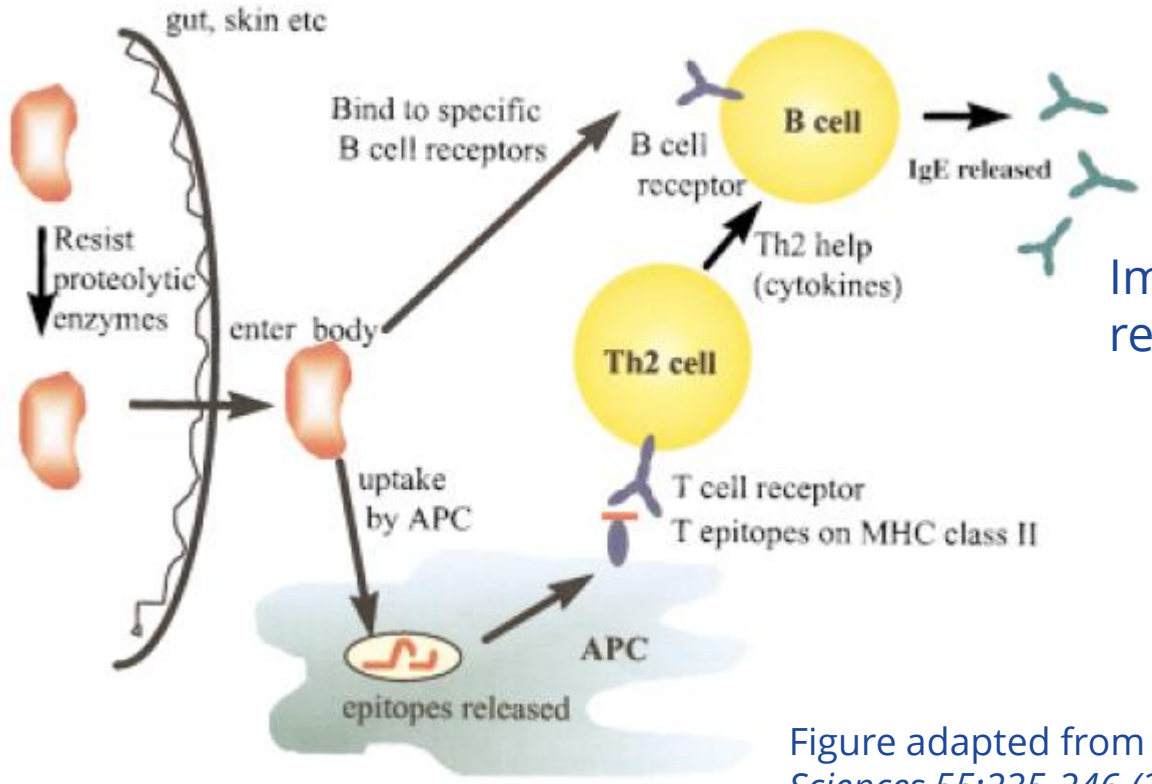
AllerCatPro - Protein Allergenicity Prediction with 3D Structure Features

- Bioinformatics Institute (BII)
- Co-developed with Krutz *et al.* (P&G) as academic free tool
- Gluten work with TCCC

Dr. Sebastian Maurer-Stroh
Executive Director, BII

The aim – biophysics inspired in silico model of 3D allergen recognition

Allergenic protein



Immune response

Senior toxicology advisors:
Ian Kimber, Bruni Bloemeke and Frank Gerberick

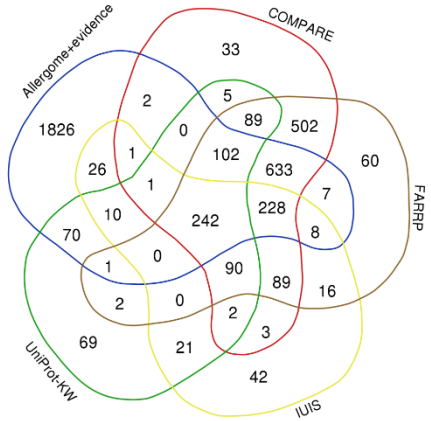
Figure adapted from *Toxicological Sciences 55;235-246 (2000)*

POWERING DISCOVERIES

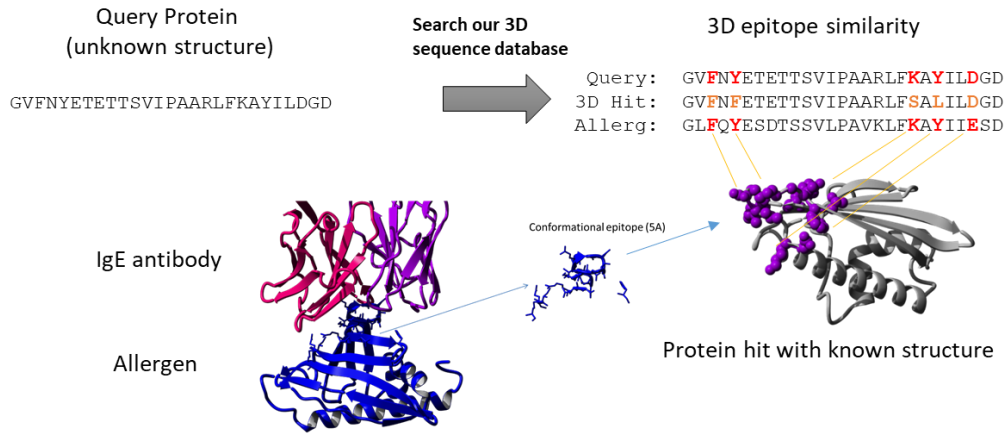
AllerCatPro v1

Maurer-Stroh S, Krutz NL, Kern PS, Gunalan V, Nguyen MN, Limviphuvadh V, Eisenhaber F, Gerberick GF. **AllerCatPro-prediction of protein allergenicity potential from the protein sequence.** *Bioinformatics.* 2019 Sep 1;35(17):3020-3027.
Cited 40 times in <2 years

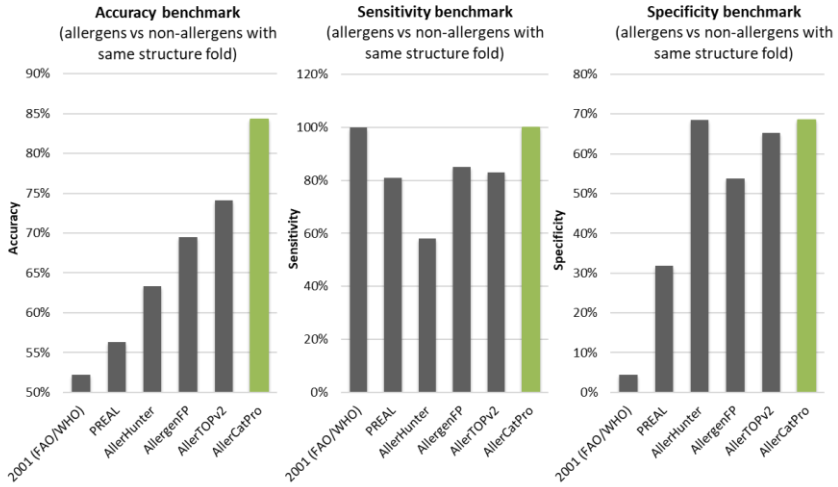
Krutz NL et al., Kimber I, Maurer-Stroh S, Gerberick GF. **Determination of the relative allergenic potency of proteins: hurdles and opportunities.** *Crit Rev Toxicol.* 2020 Jul;50(6):521-530.



Combined database of known allergens



3D surface structure on top of linear sequence window similarity



Highest accuracy on difficult benchmark, compared to Codex rules both have ~100% sensitivity to find known allergens but AllerCatPro has 37-fold higher specificity (less false positives)

POWERING DISCOVERIES

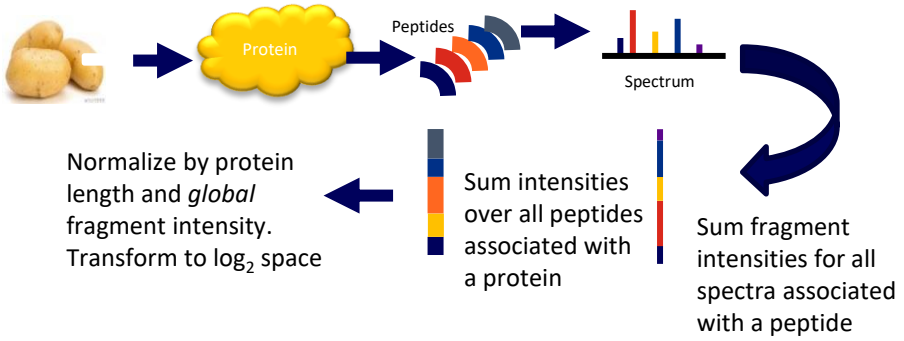
Application 1: New plant product allergenicity screen

1. *Use label-free proteomics to identify most abundant proteins in food or plant material for new products*

P&G

Jason Winget, Nora Krutz

(Frank Gerberick)

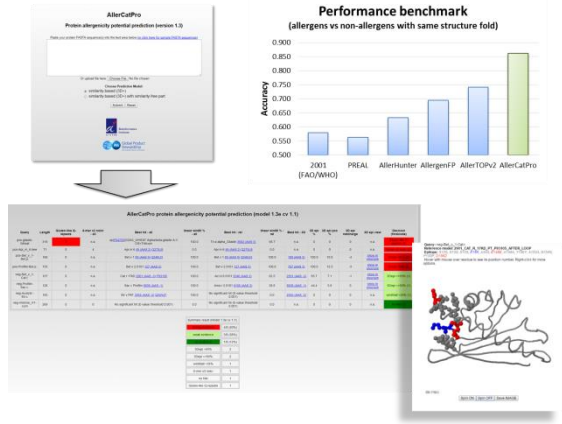


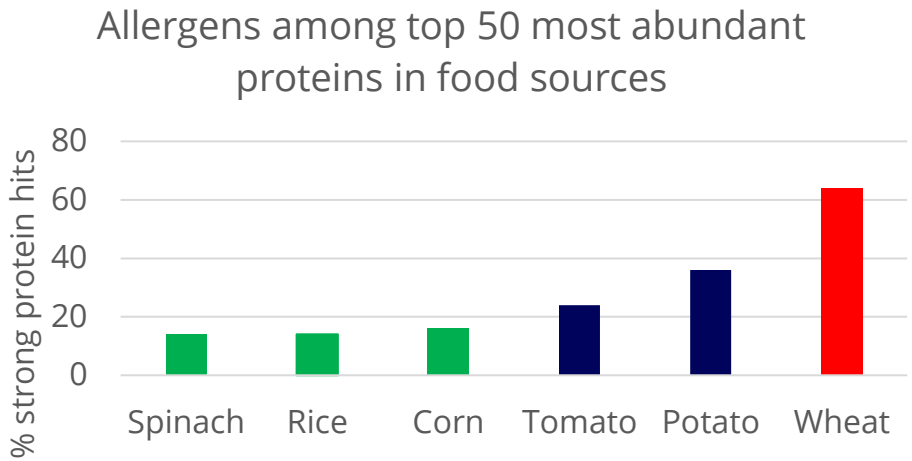
Krutz NL, Winget J, Ryan CA, Wimalasena R, Maurer-Stroh S, Dearman RJ, Kimber I, Gerberick GF. **Proteomic and Bioinformatic Analyses for the Identification of Proteins With Low Allergenic Potential for Hazard Assessment.** *Toxicol Sci.* 2019 Jul 1;170(1):210-222.

2. *Options: Evaluate top 50 protein hits (or above critical concentration)* ***with in silico tool AllerCatPro***

Bioinformatics Institute

BII





POWERING DISCOVERIES

Application 2: Safety assessment for novel food protein

RESEARCH ARTICLE

Soy Leghemoglobin

Molecular Nutrition & Food Research

www.mnf-journal.com

Evaluating Potential Risks of Food Allergy and Toxicity of Soy Leghemoglobin Expressed in *Pichia pastoris*

*Yuan Jin, Xiaoyun He, Kwame Andoh-Kumi, Rachel Z. Fraser, Mei Lu, and Richard E. Goodman**

Scope: The Soybean (*Glycine max*) leghemoglobin c2 (LegHb) gene was introduced into *Pichia pastoris* yeast for sustainable production of a heme-carrying protein, for organoleptic use in plant-based meat. The potential allergenicity and toxicity of LegHb and 17 *Pichia* host-proteins each representing ≥1% of total protein in production batches are evaluated by literature review, bioinformatics sequence comparisons to known allergens or toxins, and in vitro pepsin digestion.

1. Introduction

Hemoglobins (Hbs) are ubiquitous iron binding proteins in nature, present in bacteria, fungi, higher plants, and animals.[1] Consumption of these proteins serves as an efficient source of bioavailable iron, which is required for oxygen transport, respiration, and

Novel food protein: e.g. burger patty based on soy leghemoglobin produced in *Pichia*

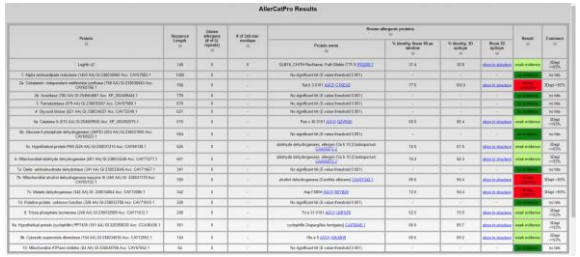
Singapore Food Agency

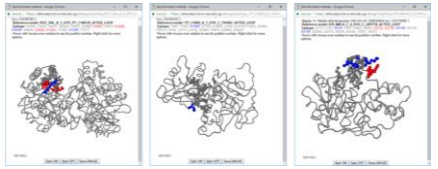
State-of-the-art:

11 of 18 tested proteins have linear window matches to allergens. All further evaluated by literature search…

AllerCatPro:

Only **3** of 18 tested proteins have 3D window matches to allergens. Shortlist for further focussed searches…





POWERING DISCOVERIES

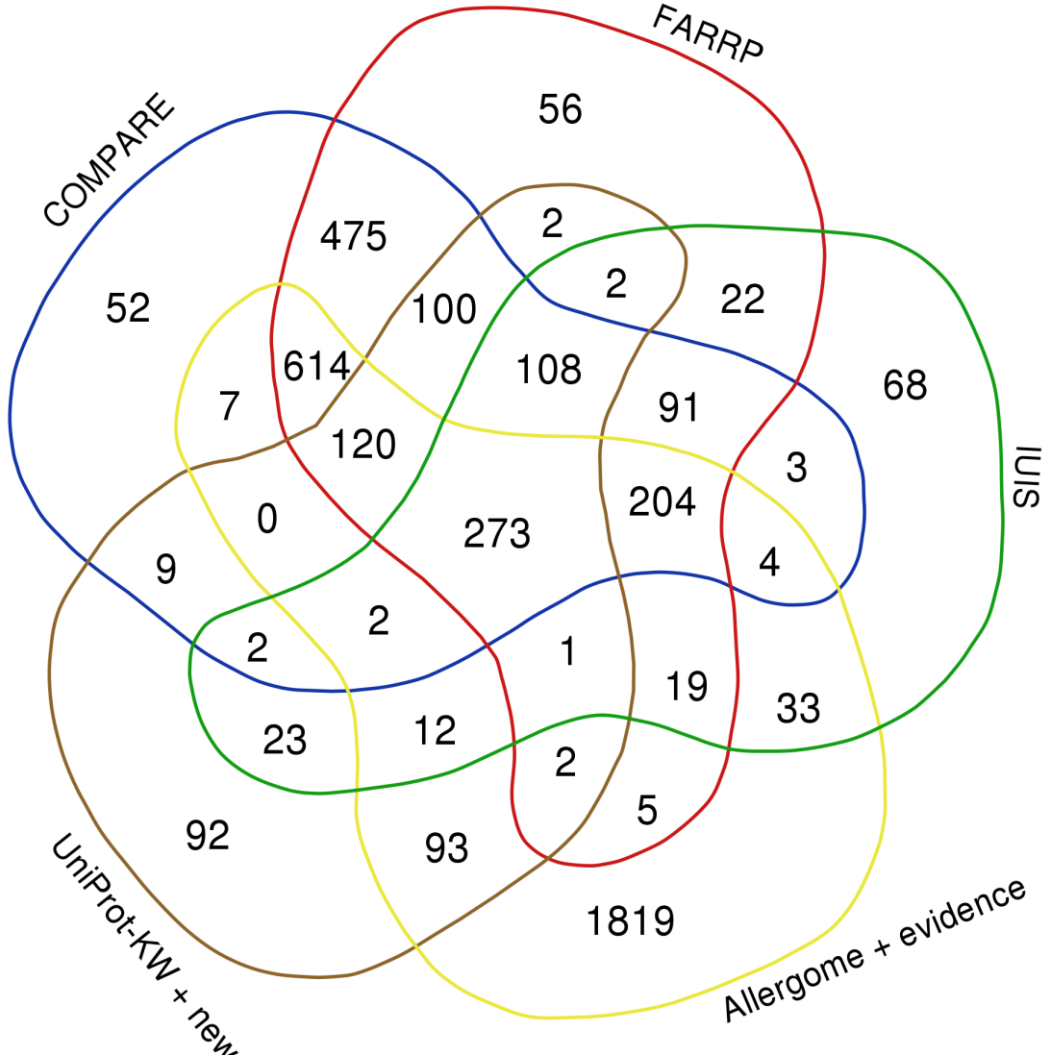
Updates and extensions for upcoming v2.0

1. update of the underlying data set
2. structure-guided approach to detect Celiac disease peptides
3. increased cross-links in the output
4. inclusion of experimental epitopes from IEDB both in a linear and 3D context for selected families
5. application to cross-reactivity between insect and shellfish allergens

POWERING DISCOVERIES

Updates and extensions for upcoming v2.0

1. update of the underlying data set, 4180=>4313



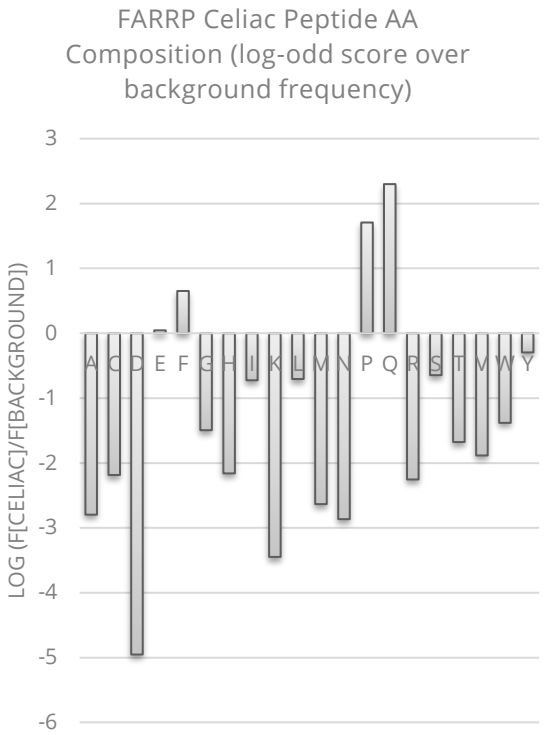
New: 144 Removed: 11

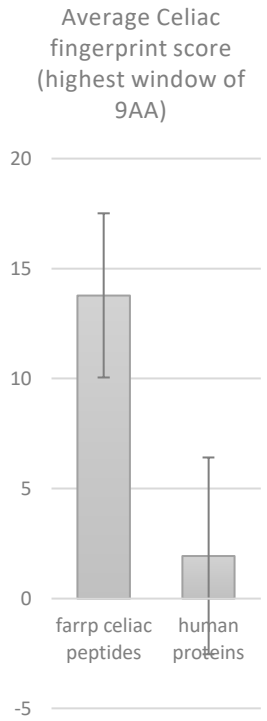
Database	Number of sequences
Allergome	45
IUIS	40
FARRP	17
UniprotKB + literature	24
Compare	18
Total	144

Updates and extensions for upcoming v2.0

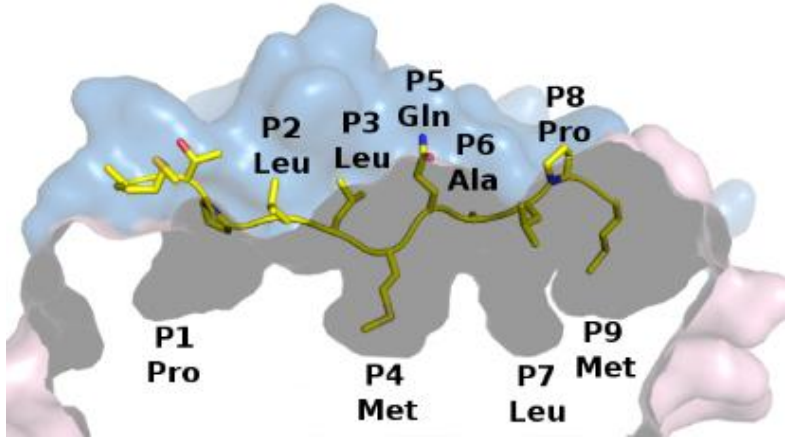
2. structure-guided approach to detect Celiac disease peptides

Old version – 9AA fingerprint





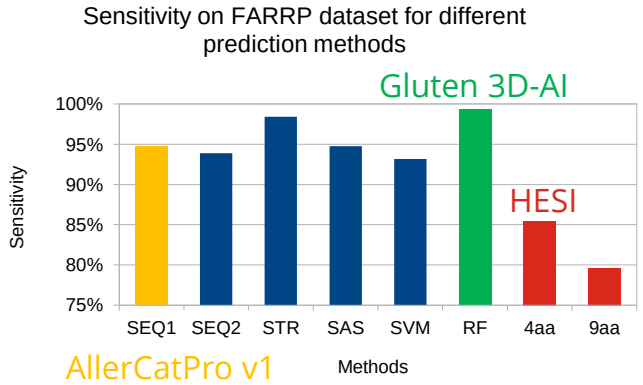
Gluten 3D-AI extension (new)

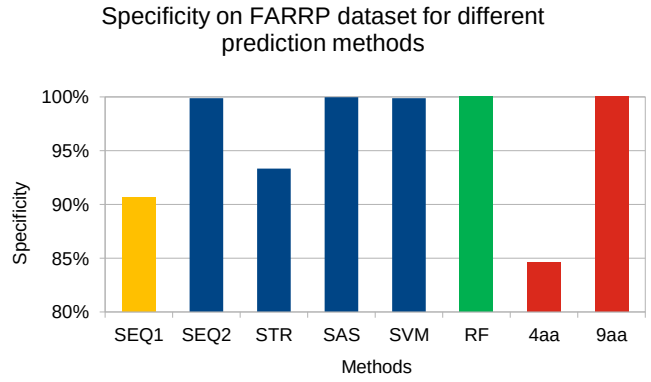


MHC DQ 2.5

Work with TCCC

AI/ML ⇩





POWERING DISCOVERIES

Updates and extensions for upcoming v2.0

3. increased cross-links in the output *(Format will still be simplified!)*

AllerCatPro Results

Protein	Sequence Length	Gluten allergens (# of Q-repeats)	# of 3x6-mer overlaps	Known allergenic proteins																Result	Comment
				# hits	Best hit protein name	Species	Allergen info	UniProt	SUPFAM	Pfam	InterPro	IgE prevalence	Cross reactivity	Sequence homology	% identity, linear 80 aa window	% identity, 3D epitope	Show 3D epitope	% identity, 3D IEDB	Show 3D IEDB	# IEDB	# IEDB-2M		
pos-Profilin-Bet.p.	133	0	-	234	Bet v 2.0101	Betula pendula (European white birch) (Betula verrucosa).	AG(5)	P25816	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	247	-	-	100.0	100.0	show in structure	100.0	show in structure	16	23	strong evidence	3Depi >93%
neg-Profilin-Sac.c.	126	0	-	233	Ama r 2	Amaranthus retroflexus (Redroot amaranth) (Redroot pigweed).	AG(2)	C3W2Q7	SSF55770	PF00235	-	12	-	Che a 2, Cro s 2, Cuc m 2, Hev b 8, Sal k 4	38.8	44.4	show in structure	60.0	show in structure	0	0	weak evidence	3Depi <=93%

List of all hits (not just best)

Hit	Protein Name	Species	AllerCatPro ID	UniProt	SUPFAM	Pfam	InterPro	% identity
1	Bet v 2.0101	Betula pendula (European white birch) (Betula verrucosa).	1887	P25816	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	100.0
2	f|pdb|1CQA|A Chain A, SEQRES	-	-	-	-	-	-	96.2
3	Bet v 2	Betula pendula (European white birch) (Betula verrucosa).	212	A4K9Z8	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	96.2
4	Cor a 2	Corylus avellana (European hazel) (Corylus maxima).	299	A4KA45	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	93.2
5	Cor a 2	Corylus avellana (European hazel) (Corylus maxima).	300	A4KA44	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	91.0
6	Cor a 2	Corylus avellana (European hazel) (Corylus maxima).	302	A4KA39	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	91.0
7	Cor a 2	Corylus avellana (European hazel) (Corylus maxima).	301	A4KA40	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	91.0
8	Ric c 8	Ricinus communis (Castor bean).	1410	B9RKF4	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	89.5
9	Gly m 3	Glycine max (Soybean) (Glycine hispida).	477	I1K602	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	86.5
10	Mer a 1.0101	Mercurialis annua (Annual mercury).	1908	O49894	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	85.0
11	Ole e 2	Olea europaea (Common olive).	835	A4GD52	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	85.8

Cross-links to multiple databases, for protein family (domains), more info in Allergome etc.

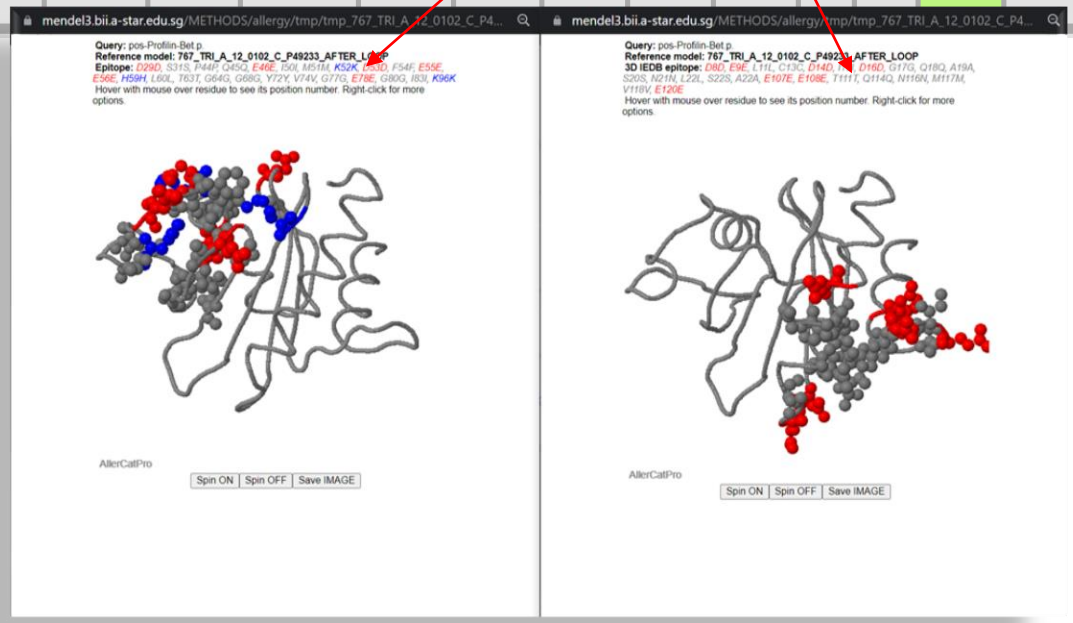
POWERING DISCOVERIES

Updates and extensions for upcoming v2.0

4. inclusion of experimental epitopes from IEDB

AllerCatPro Results

Protein	Sequence Length	Gluten allergens (# of Q-repeats)	# of 3x6-mer overlaps	Known allergenic proteins																	Result	Comment
				# hits	Best hit protein name	Species	Allergen info	UniProt	SUPFAM	Pfam	InterPro	IgE prevalence	Cross reactivity	Sequence homology	% identity, linear 80 aa window	% identity, 3D epitope	Show 3D epitope	% identity, 3D IEDB	Show 3D IEDB	# IEDB	# IEDB-2M		
pos-Profilin-Bet.p.	133	0	-	234	Bet v 2.0101	Betula pendula (European white birch) (Betula verrucosa).	AG(5)	P25816	SSF55770	PF00235	IPR005455 IPR036140 IPR027310	247	-	-	100.0	100.0	show in structure	100.0	show in structure	16	23	strong evidence	3Depi >93%
neg-Profilin-Sac.c.	126	0	-	233	Ama r 2	Amaranthus retroflexus (Redroot amaranth) (Redroot pigweed).	AG(2)	C3W2Q7	SSF55770	PF00235	-	12	-	Che a 2, Cro s 2, Cuc m 2, Hev b 8, Sal k 4	38.8	44.4	show in structure	60.0	show in structure	0	0	weak evidence	3Depi <=93%



For protein families with IEDB info, epitope similarity match over best known B cell epitope with 3D and linear match reported

POWERING DISCOVERIES

Updates and extensions for upcoming v2.0

5. cross-reactivity between insect and shellfish allergens

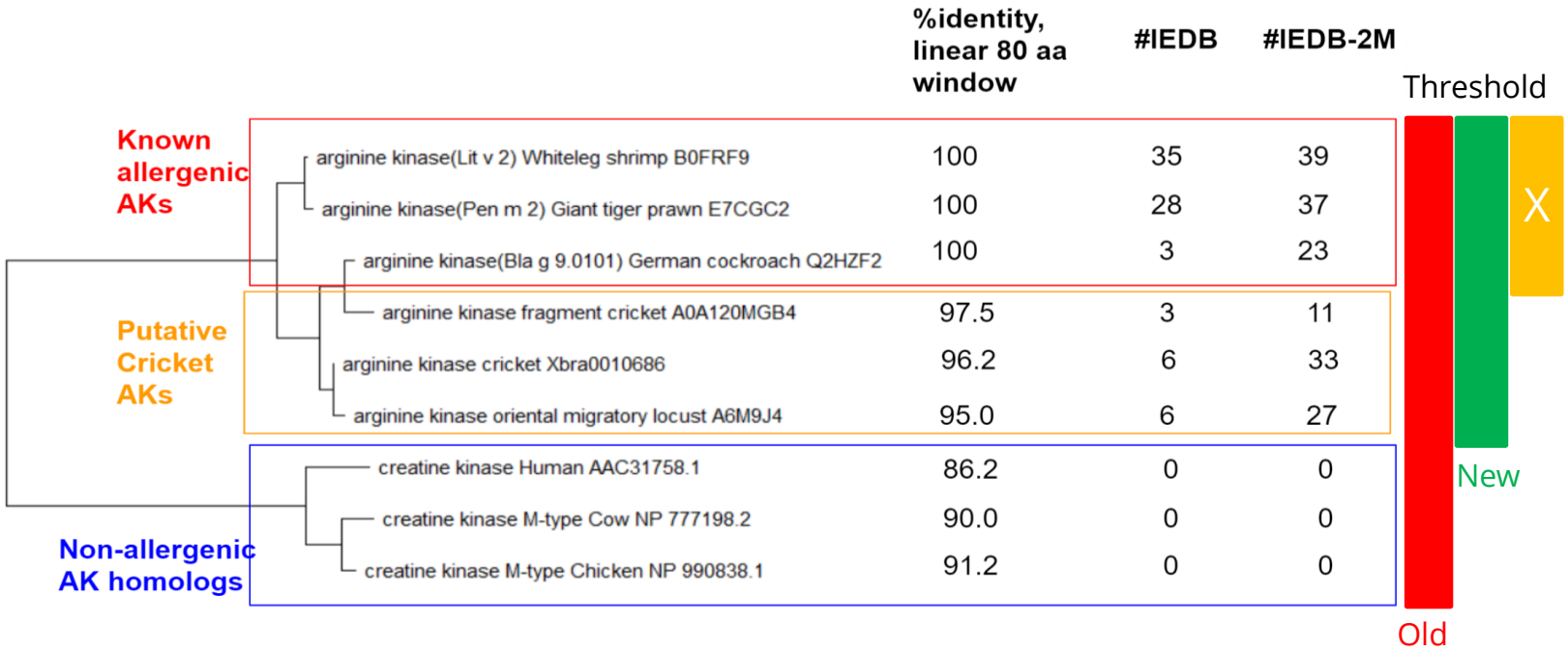
Food Chem. 2021 Jun 30;348:129110. doi: 10.1016/j.foodchem.2021.129110. Epub 2021 Jan 19.

Protein extraction protocols for optimal proteome measurement and arginine kinase quantitation from cricket Acheta domesticus for food safety assessment

Utpal Bose [1], James A Broadbent [1], Angéla Juhász [2], Shaymaviswanathan Karnaneedi [3], Elecia B Johnston [3], Sally Stockwell [1], Keren Byrne [1], Vachiranee Limviphuvadh [4], Sebastian Maurer-Stroh [5], Andreas L Lopata [3], Michelle L Colgrave [6]

Improvement by adding family-specific shellfish allergen epitope score and clinical data guided threshold

Arginine kinase (AK) phylogenetic tree



POWERING DISCOVERIES

POWERING DISCOVERIES

Agency for Science, Technology and Research
SINGAPORE

THANK YOU

www.a-star.edu.sg