COMPARE

Comprehensive Protein Allergen Resource

The COMPARE Database: a public COMprehensive Protein Allergen REsource for protein allergenicity assessment

Lars K. Poulsen
Professor of Basic Allergology
Allergy Clinic, Copenhagen University Hospital at Gentofte
Copenhagen, Denmark

Health and Environmental Sciences Institute (HESI)
Protein Allergens, Toxins and Bioinformatics (PATB) Committee

HESI PATB Committee

Objective

- Establish weight-of-evidence approach
- Develop new science
- Communicate scientific findings and best practices

Activity

Assessment

COMPARE Allergen Dbase

Protein Toxins Evaluation

Research

Digestibility Study

Allergen Rebuild Study

Matrix Study

GARD Assay Study

Immunogenicity Study

Outreach and Training

Workshops, e.g. *today* and : "Non-IgE mediated reactions to foods" Workshop. 2016. Rome, Italy.

Presentations at international meetings

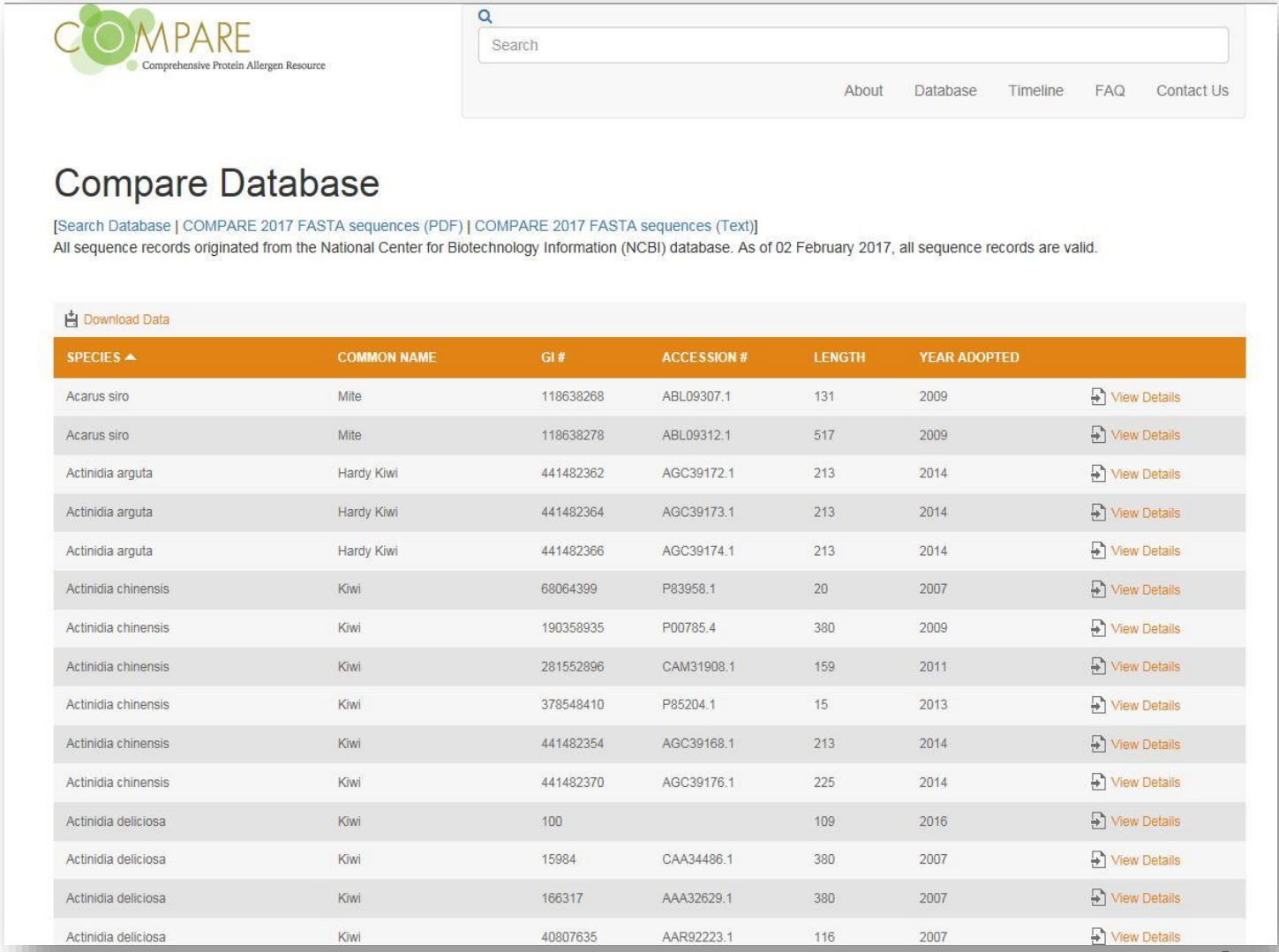
Publications

HESI COMPARE Database:

www.comparedatabase.org

COMPARE
Comprehensive Protein Allergen Resource

- Comprehensive and up to date collection of peer reviewed allergen protein sequences.
- Custom process development to identify new allergens initiated in 2016.
- First release (Feb. 2017) - "COMPARE 2017": 1970 allergens.
  - Includes the allergens listed in AOL v.16 and 14 *new* allergens identified though bioinformatics approaches.
- COMPARE 2018 (Feb. 2018):
  - Added 68 allergens to COMPARE 2017 (2038 allergens total).
- Updated annually



Compare Database

[Search Database | COMPARE 2017 FASTA sequences (PDF) | COMPARE 2017 FASTA sequences (Text)]
All sequence records originated from the National Center for Biotechnology Information (NCBI) database. As of 02 February 2017, all sequence records are valid.

Download Data

SPECIES	COMMON NAME	GI #	ACCESSION #	LENGTH	YEAR ADOPTED	
Acarus siro	Mite	118638268	ABL09307.1	131	2009	View Details
Acarus siro	Mite	118638278	ABL09312.1	517	2009	View Details
Actinidia arguta	Hardy Kiwi	441482362	AGC39172.1	213	2014	View Details
Actinidia arguta	Hardy Kiwi	441482364	AGC39173.1	213	2014	View Details
Actinidia arguta	Hardy Kiwi	441482366	AGC39174.1	213	2014	View Details
Actinidia chinensis	Kiwi	68064399	P83958.1	20	2007	View Details
Actinidia chinensis	Kiwi	190358935	P00785.4	380	2009	View Details
Actinidia chinensis	Kiwi	281552896	CAM31908.1	159	2011	View Details
Actinidia chinensis	Kiwi	378548410	P85204.1	15	2013	View Details
Actinidia chinensis	Kiwi	441482354	AGC39168.1	213	2014	View Details
Actinidia chinensis	Kiwi	441482370	AGC39176.1	225	2014	View Details
Actinidia deliciosa	Kiwi	100		109	2016	View Details
Actinidia deliciosa	Kiwi	15984	CAA34486.1	380	2007	View Details
Actinidia deliciosa	Kiwi	166317	AAA32629.1	380	2007	View Details
Actinidia deliciosa	Kiwi	40807635	AAR92223.1	116	2007	View Details

About the Process

- Automated bioinformatics sorting algorithm.
- Scientific information and publications related to individual identified sequences.
- Independent Peer Review Panel of recognized allergy experts, public sector.
- A quality control and documentation.
- A public release & annual update.

✓ *Publicly accessible, transparent, rigorous and reliable resource.*

COMPARE
Comprehensive Protein Allergen Resource

Why a new allergen database?

There are already multiple allergen databases available, with different features and purposes…

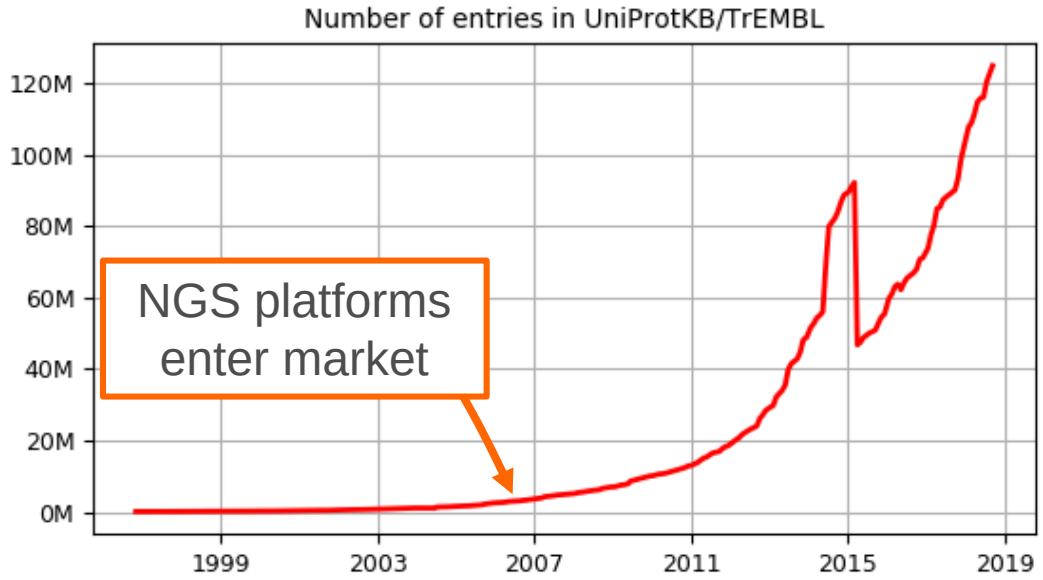
E.g., Allergome, AllergenOnline (AOL), IUIS, AllFam, SDAP…

Sircar et al. 2014. Allergen databases. Methods Mol Biol. 1184:165-81.
Radauer. 2017. Navigating through the Jungle of Allergens: Features and Applications of Allergen Databases. Int Arch Allergy Immunol. 2017;173:1-11.

… but in recent years the widespread use of genomic sequencing technology brought new challenges….

Number of sequences to be filtered has grown exponentially.

- **~ 125 Million sequence entries** (UniProtKB/TrEMBL)



- **Most of them are "*predicted*"**

Protein existence (PE)	Entries	%
1: Evidence at protein level	144459	0.12%
2: Evidence at transcript level	1162753	0.93%
3: Inferred from homology	30704463	24.60%
4: Predicted	92785433	74.35%
5: Uncertain	0	0.00%

Source: https://www.ebi.ac.uk/uniprot/TrEMBLstats

The COMPARE process addresses this growth by implementing a cutting-edge and high-throughput bioinformatic pipeline to identify a *meaningful subset* of these *millions* sequences, for scientific review.

COMPARE: Bringing together the required strategic, technical and logistical considerations

- ✓Tri-partite by design
- ✓Strategic partnerships: JIFSAN, Academia & Government
- ✓HESI infrastructure support and network (independent non-profit, program coordination, IT, data management, science-focused finance, communications).
- ✓Long-term stability for all of the above

RATIONALE FOR ALLERGEN DATABASE at HESI

(1) Novel components due to HESI public-private partnership.

HESI PATB COMPARE

P-P Science Advisory Team

- ✓ Multiple stakeholders engagement.

Independent Peer Review Panel (PRP)

Allergy Science Expertise

- ✓ Decisions based on peer-review

Bioinformatics services

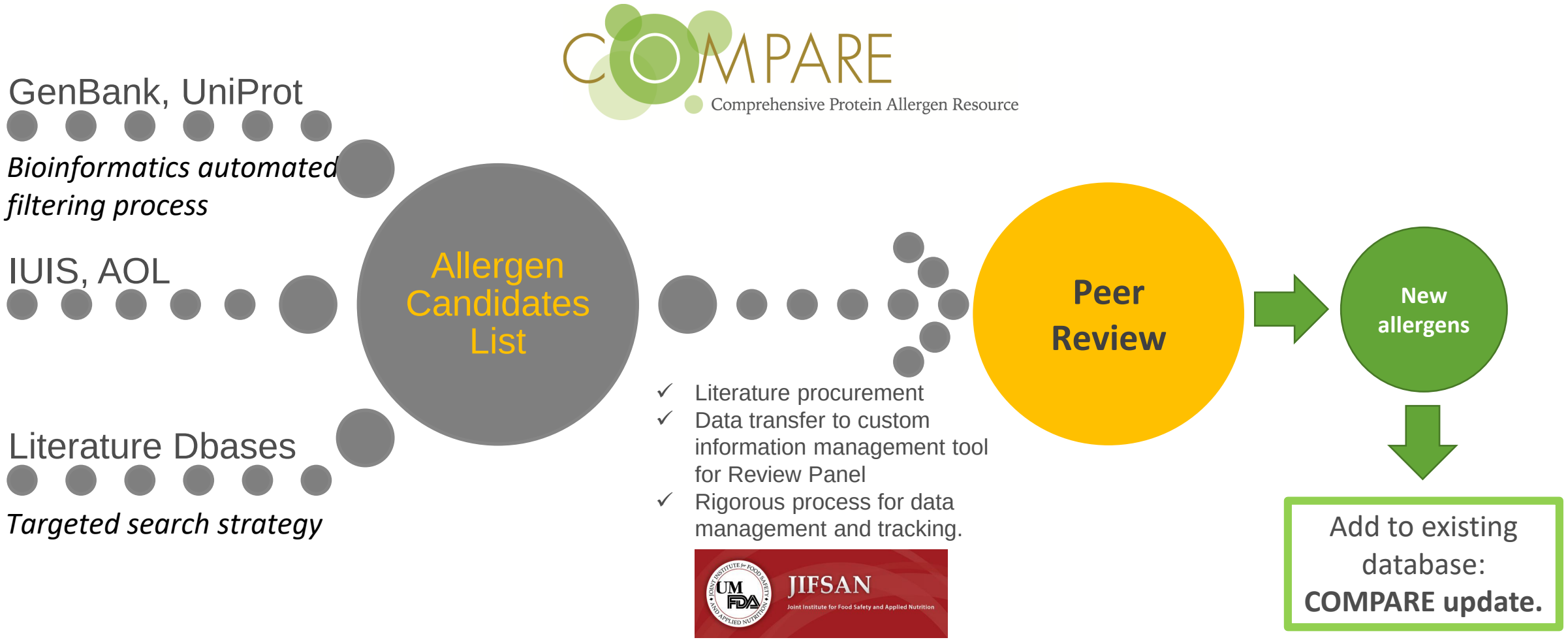
- ✓ Custom algorithm development

- ✓ **Independent coordinating body:**
  - Scientific program management
  - Budget and contracts management with external professional services.

- ✓ Rigorous process for data management and tracking (custom information management tool).
- ✓ Literature procurement

RATIONALE FOR ALLERGEN DATABASE at HESI

(2) Modular process allows multiple sourcing of allergen candidates.



Goal: enhance confidence in the efficiency and quality of allergens identified.

RATIONALE FOR ALLERGEN DATABASE at HESI
(3) Transparency & detailed documentation.

What has been done:

- COMPARE overview brochure: at initiation of the project (Q1 2016).
- Detailed process development described in the COMPARE website: within 6 months of the database public release (COMPARE' 2017). http://comparedatabase.org/process-development/
- Open dialogue w/ stakeholders (http://comparedatabase.org/contact-us/)

In preparation:

- Peer-review publication

COMPARE
Comprehensive Protein Allergen Resource

COMPREHENSIVE PROTEIN ALLERGEN RESOURCE

Program Overview – February 2016

ABOUT HESI

HESI, the Health and Environmental Sciences Institute, is a non-profit, global scientific organization dedicated to developing science for a safer, more sustainable world.

HESI's mission is to collaboratively identify and help to resolve global health and environmental challenges through engagement of scientists from academia, government, industry, NGOs, and other strategic partners.

HESI

OVERVIEW

The COMPARE Database, to be made publicly available in January 2017, will be a transparent resource for identification of protein sequences that are known or putative allergens. The COMPARE database will meet needs for allergy safety assessment via an annual updating process that 1) captures new listings of allergen sequences, 2) filters out non-allergen sequences, 3) identifies published literature linked to the identified potential allergen sequences, and 4) verifies that the newly identified sequences have clinical, published evidence of allergenicity based on standardized criteria.

PROGRAM STRUCTURE

The COMPARE database is a collaborative HESI program that combines programmatic support from the Joint Institute for Food Safety and Nutrition – www.jifsan.org at the University of Maryland - with financial support from private sector partners. Additional in-kind support from multiple academic and government partners is engaged throughout the process. A public-private HESI collaborative team will publicly document all database design and search algorithm decisions. An exclusively public-sector expert panel will review the output from a high-throughput bioinformatic pipeline against peer-reviewed publications. The expert panel will then select sequences for inclusion in the database and document their rationale for inclusion.

TIMELINE

The COMPARE database will be released to the public by January, 2017. To enable this timing, a steering team will convene in early 2016 to ensure timelines, process, and resources are aligned. Sequence searching will be conducted in second quarter 2016 and peer-review in third quarter 2016 before final assembly and release.

Overall program and components designed to enable long term, stable support.

COMPARE
Comprehensive Protein Allergen Resource

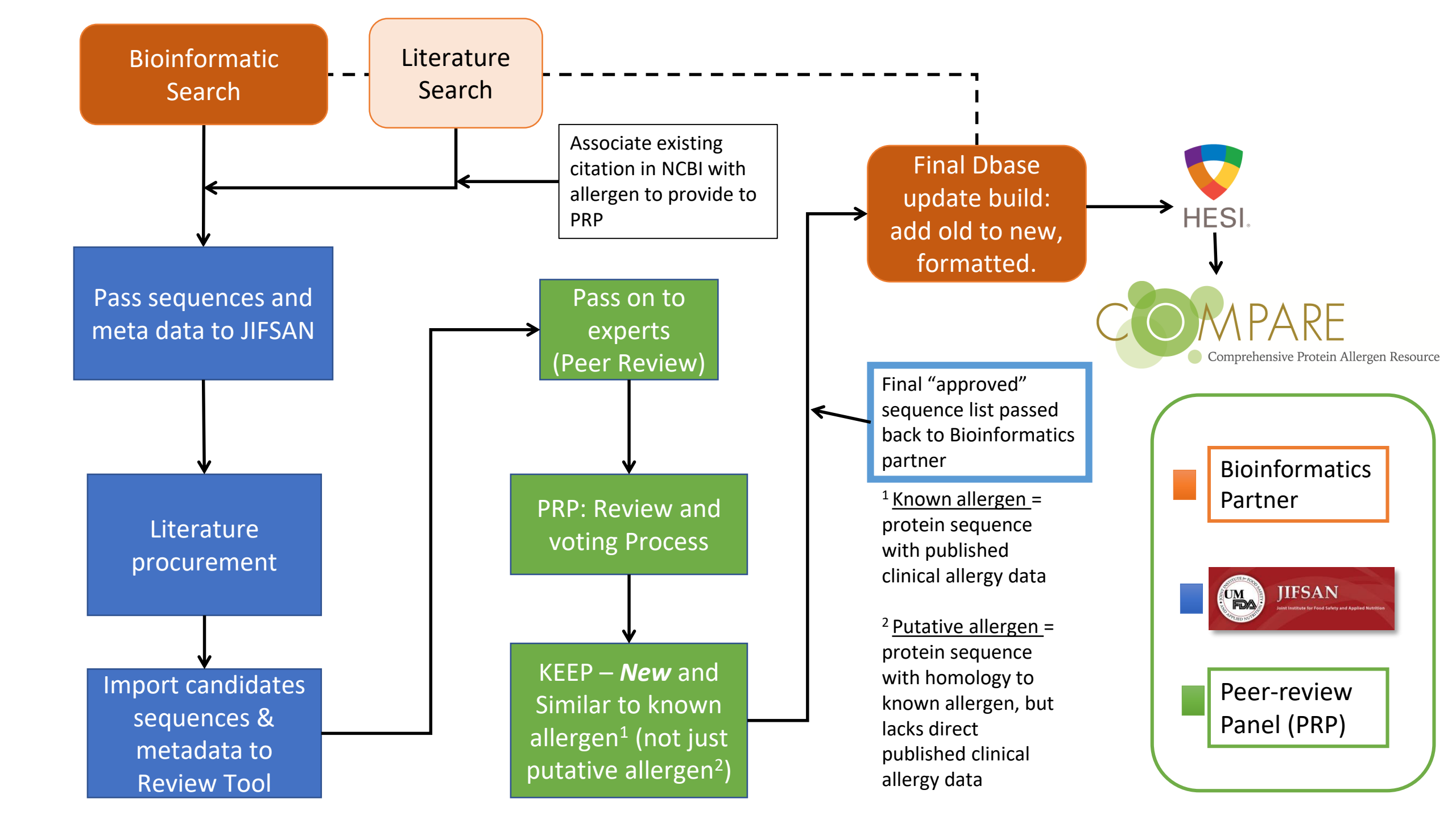
HESI.

Advisory Team	Bioinformatics and Data Communication	Program Coordination	Expert Peer Review Panel	Participating Organizations

• 20 Public/Private stakeholders.
• General program oversight, QC, user feedback.
• Expert knowledge
• No influence on decisions regarding sequence inclusion or exclusion.
	• **Bioinformatics partner**: sequences screening; deduplication, metadata retrieval and QC.
	• **University of Maryland / JIFSAN:**
	  • Literature search and procurement.
	  • Data sharing.
	  • Review Tool development and maintenance.
		• Overall program management.
		• Coordination between partners, PRP assembly and management.
		• Database 'home' and maintenance.
			• Invited public sector allergy experts.
			• Define inclusion /exclusion criteria.
			• Review of sequences and references / publications.
			• Provide final decisions on new sequences to include in database
				Public Sector:
				• University of Amsterdam
				• University of Maryland / JIFSAN
				• U.S. EPA
				• U.S. FDA
				
				Private Sector:
				• BASF
				• Bayer CropScience
				• DowDuPont Pioneer
				• KWS
				• Syngenta
				• Vilmorin & Cie

JIFSAN
Joint Institute for Food Safety and Applied Nutrition

JIFSAN: Joint Institute for Food Safety And Nutrition (UMD/FDA)

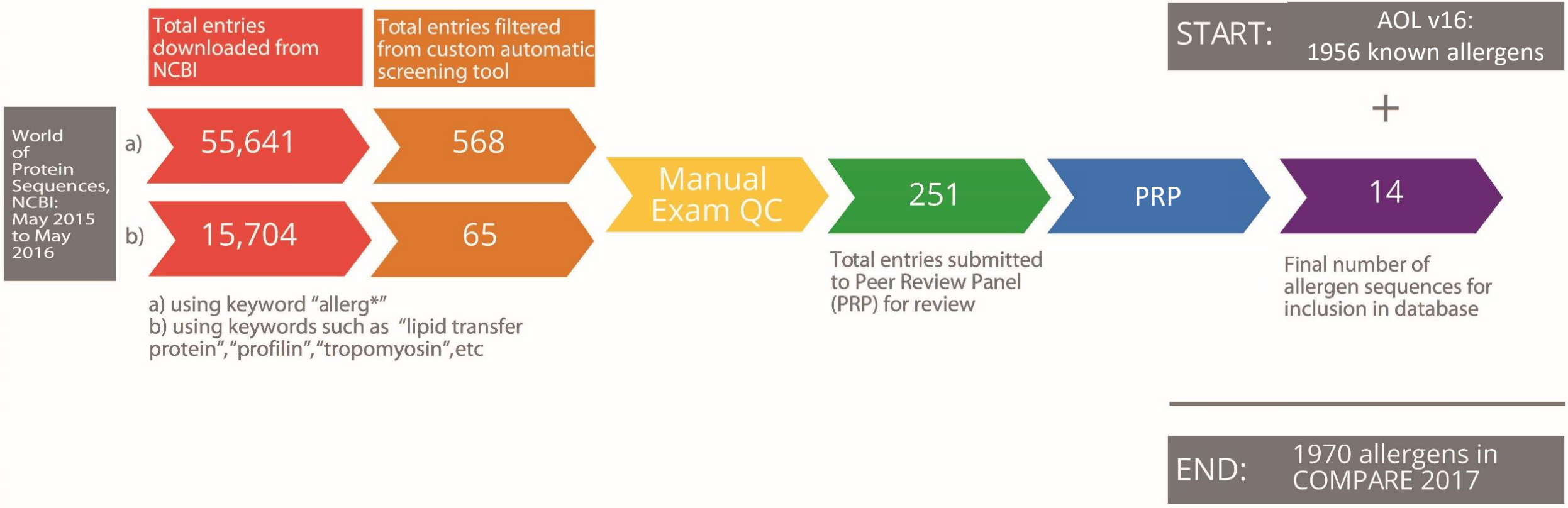

Bioinformatic Search
Literature Search
Associate existing citation in NCBI with allergen to provide to PRP
Pass sequences and meta data to JIFSAN
Literature procurement
Import candidates sequences & metadata to Review Tool
Pass on to experts (Peer Review)
PRP: Review and voting Process
KEEP – New and Similar to known allergen[1] (not just putative allergen[2])
Final Dbase update build: add old to new, formatted.
HESI
COMPARE
Comprehensive Protein Allergen Resource
Final "approved" sequence list passed back to Bioinformatics partner
[1] Known allergen = protein sequence with published clinical allergy data
[2] Putative allergen = protein sequence with homology to known allergen, but lacks direct published clinical allergy data
Bioinformatics Partner
JIFSAN
Joint Institute for Food Safety and Applied Nutrition
Peer-review Panel (PRP)

HESI

COMPARE

Comprehensive Protein Allergen Resource

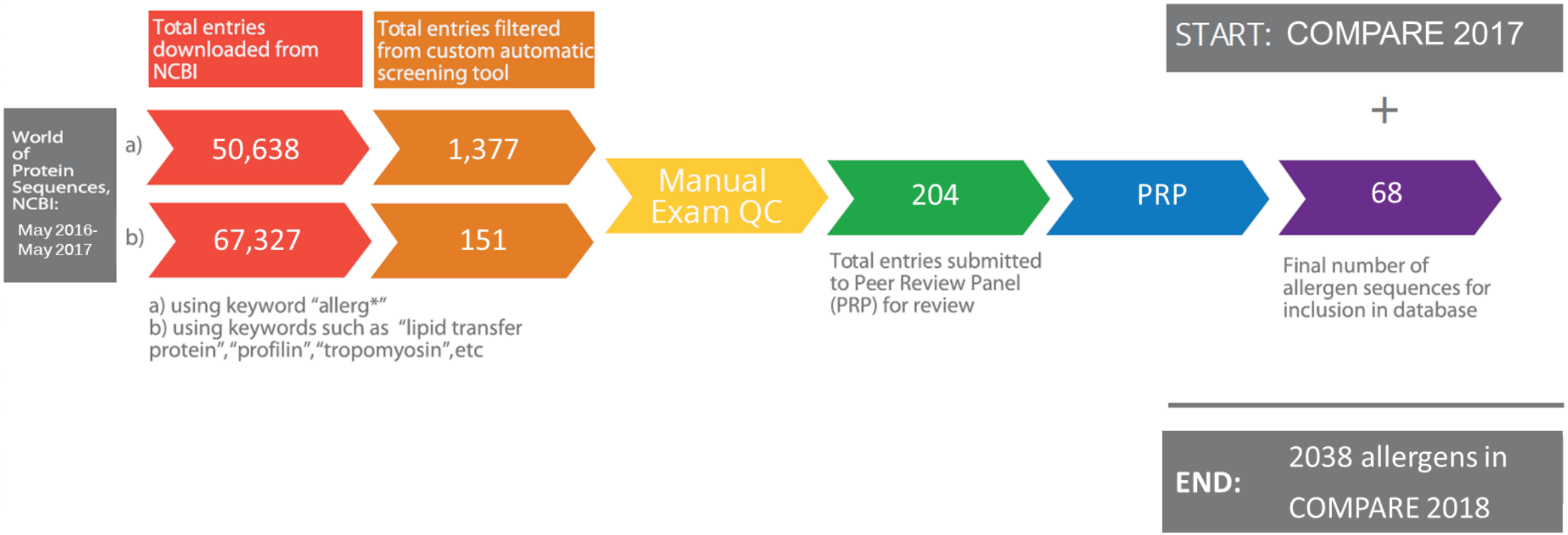
COMPARE **2017**



HESI

COMPARE
Comprehensive Protein Allergen Resource

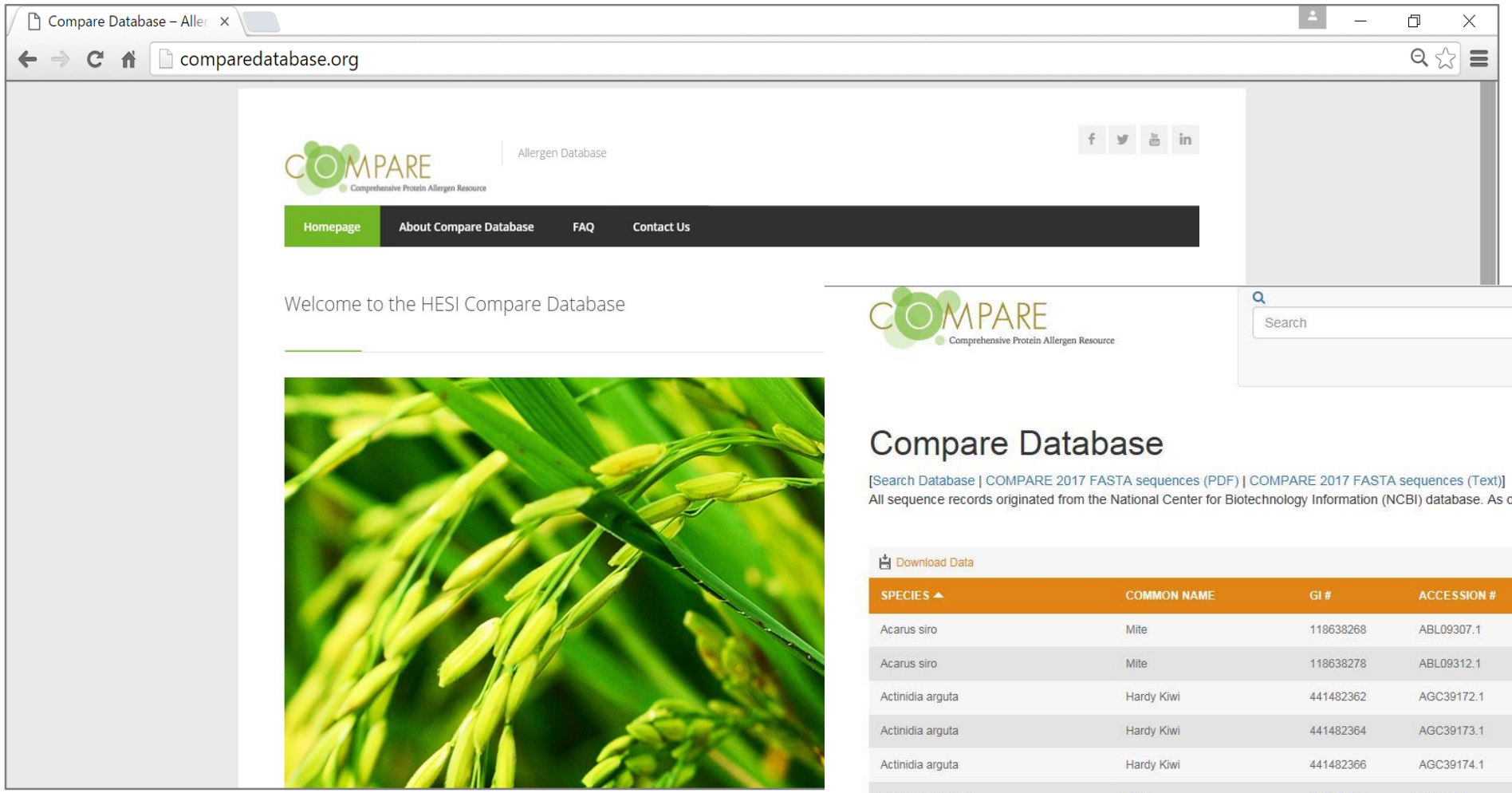
COMPARE 2018



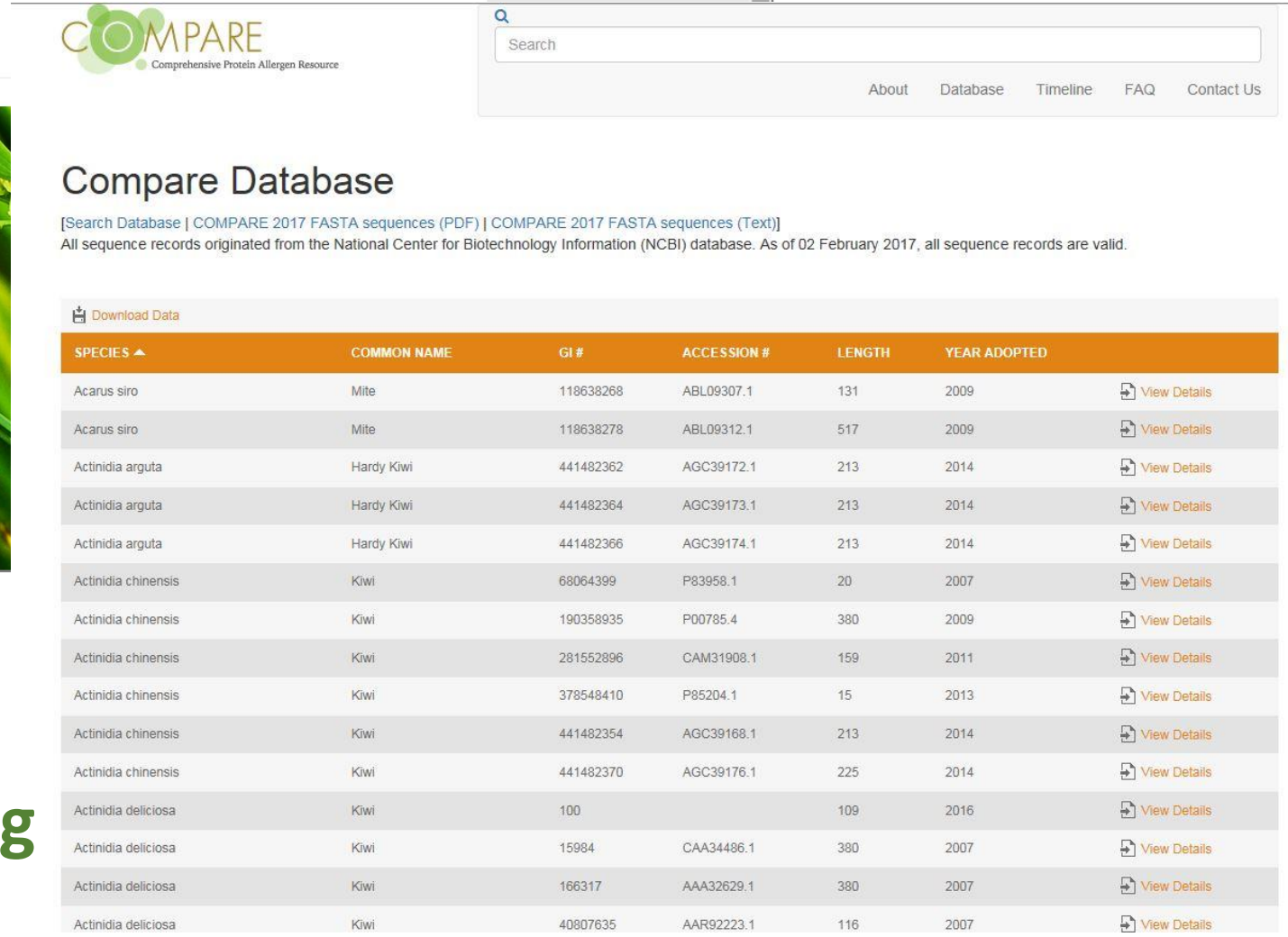
"COMPARE 2018" numbers and current candidates numbers

COMPARE 2018	Reviewed Allergens	Reviewed Articles	**Approved Allergens**
Bioinformatics screenings (NCBI)	93	186	27
Literature Search	-	37	29
AOL 17	22	35	12
TOTAL	115	258	**68**

COMPARE 2019 - preview			
TOTAL	181	137	? (on going)



HESI



SPECIES	COMMON NAME	GI #	ACCESSION #	LENGTH	YEAR ADOPTED	
Acarus siro	Mite	118638268	ABL09307.1	131	2009	View Details
Acarus siro	Mite	118638278	ABL09312.1	517	2009	View Details
Actinidia arguta	Hardy Kiwi	441482362	AGC39172.1	213	2014	View Details
Actinidia arguta	Hardy Kiwi	441482364	AGC39173.1	213	2014	View Details
Actinidia arguta	Hardy Kiwi	441482366	AGC39174.1	213	2014	View Details
Actinidia chinensis	Kiwi	68064399	P83958.1	20	2007	View Details
Actinidia chinensis	Kiwi	190358935	P00785.4	380	2009	View Details
Actinidia chinensis	Kiwi	281552896	CAM31908.1	159	2011	View Details
Actinidia chinensis	Kiwi	378548410	P85204.1	15	2013	View Details
Actinidia chinensis	Kiwi	441482354	AGC39168.1	213	2014	View Details
Actinidia chinensis	Kiwi	441482370	AGC39176.1	225	2014	View Details
Actinidia deliciosa	Kiwi	100		109	2016	View Details
Actinidia deliciosa	Kiwi	15984	CAA34486.1	380	2007	View Details
Actinidia deliciosa	Kiwi	166317	AAA32629.1	380	2007	View Details
Actinidia deliciosa	Kiwi	40807635	AAR92223.1	116	2007	View Details

Overview of www.comparedatabase.org

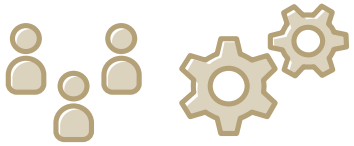
SPECIES	COMMON NAME	GI # ▲	ACCESSION #	LENGTH	YEAR ADOPTED	
Lupinus albus	Lupine (Field)	0075121065	Q6EBC1.1	0533	2017	View Details
Beta vulgaris	Beet	0205829383	P85984.1	0023	2017	View Details
Actinidia deliciosa	Kiwifruit	0906848987	4X9U_A	0189	2017	View Details
Amaranthus retroflexus	Pigweed	0914410010	AKV72168.1	0168	2017	View Details
Anisakis simplex	Nematode	0957554293	BAT62430.1	0217	2017	View Details
Lupinus angustifolius	Narrowleaf Lupine	0980951518	F5B8V9.1	0611	2017	View Details
Lupinus angustifolius	Narrowleaf Lupine	0980951548	F5B8W0.1	0603	2017	View Details
Lupinus angustifolius	Narrowleaf Lupine	0980951550	F5B8W1.1	0580	2017	View Details
Lupinus angustifolius	Narrowleaf Lupine	0980951555	F5B8W2.1	0590	2017	View Details
Lupinus angustifolius	Narrowleaf Lupine	0980951561	F5B8W3.1	0637	2017	View Details
Lupinus angustifolius	Narrowleaf Lupine	0980951568	F5B8W5.1	0605	2017	View Details
Arachis hypogaea	Peanut	1018736824	B3EWP3.1	0071	2017	View Details
Arachis hypogaea	Peanut	1018736830	B3EWP4.1	0079	2017	View Details
Arachis hypogaea	Peanut	1018736837	C0HJZ1.1	0072	2017	View Details

Records 1-14 of 14

COMPARE 2016-2018 and beyond

2017-18 Focus: Adding to Rigor & Effectiveness in Process; Adding New Features

2016

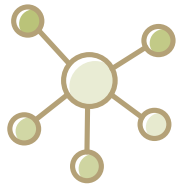


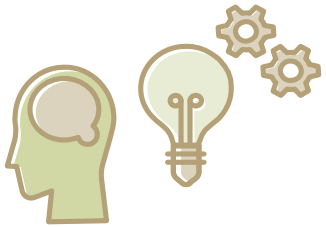
- Steering team formed
- Bioinfo process v 1.0
- PRP set up
- Literature procurement (JIFSAN).

2017

- Bioinfo process v 2.0
- Review Tool Development (JIFSAN) **New!** ;
- Piloting: lit search, IUIS, AOL. **New!**

2018





Identified new collaborators for:

- Bioinformatics screening (to meet the needs of the program);
- Literature search (in parallel of Bioinfo screening).

Process efficiency improvements:

- For capture of sequences & metadata, deduplication & QC;
- Consolidation of multiple sources

New components development and activities:

- COMPARE "historic rebuild"
- FASTA Search Tool;
- Reviewers decisions online (transparency);
- Online Dbase functional improvements.

Translate

Current efforts

HESI.

Developing new features in COMPARE Website:

- FASTA search tool
- Increased transparency: making reviewers decision & comments publicly accessible
- Database improvements: more user friendly (making it easier to retrieve references)

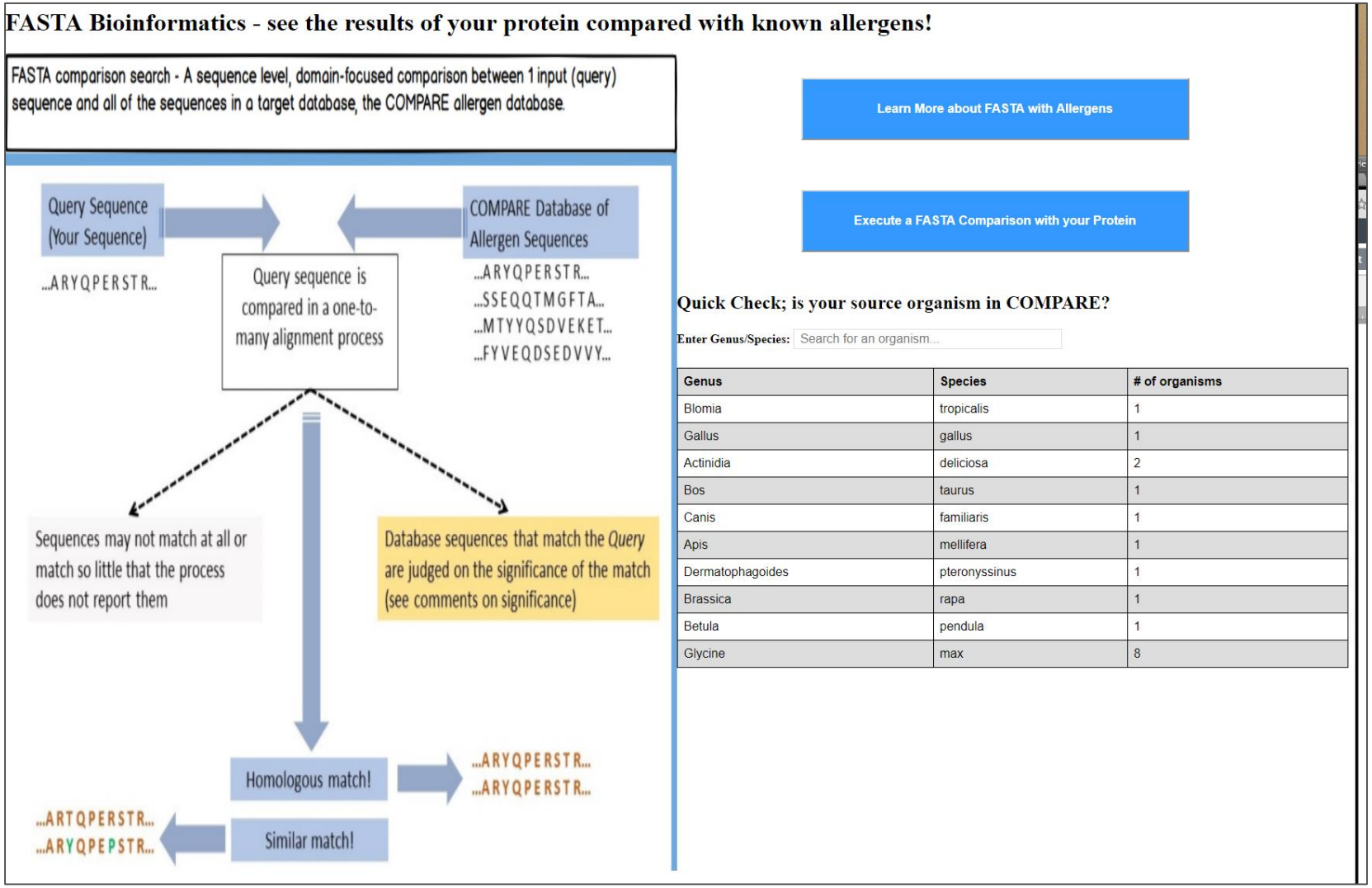
To be available with the "COMPARE 2019" release

Applying the current process to historic data (2016 data and previous years)

- Bioinformatic screenings from 2016 and previous years;
- Reviewing quality of "older" sequences and updating.

a) FASTA Bioinformatics Search

COMPARE
Comprehensive Protein Allergen Resource



Genus	Species	# of organisms
Blomia	tropicalis	1
Gallus	gallus	1
Actinidia	deliciosa	2
Bos	taurus	1
Canis	familiaris	1
Apis	mellifera	1
Dermatophagoides	pteronyssinus	1
Brassica	rapa	1
Betula	pendula	1
Glycine	max	8

http://fasta.test.jifsan.org/

b) Database improvements

COMPARE Comprehensive Protein Allergen

COMPARE Database

[Search Database | COMPARE 2018 FASTA sequences (PDF) | COMPARE 2018 FASTA sequences (Text)]

All sequence records originated from the National Center for Biotechnology Information (NCBI) database.

Species	Sus scrofa	Common Name	Wild Boar
Definition	serum albumin precursor	GI #	
Accession #	NP_001005208	Length	0607
References	20742, 20743, 20739	Year Adopted	2018

BACK

Record 43 of 2038

To retrieve full references, please copy a reference ID number from the table above and paste in the field bellow. Please enter only one number at a time. Multiple references for a same database entry can be retrieved by clicking the "search again" button and copying a new reference ID number.

Search Again

PUBMED	AUTHORS	TITLE	JOURNAL	ID
16364171	Donnay C; Barderas R; Kopferschmitt-Kubler MC; Pauli G; de Blay F	Sensitization to pig albumin and gamma-globulin responsible for occupational respiratory allergy.	Allergy. 2006 Jan;61(1):143-4	20743

Records 1-1 of 1

1) Clickable Metadata?

https://www.ncbi.nlm.nih.gov/protein/52353352

http://www.uniprot.org/uniprot/?query=NP_001005208&sort=score

Click on Species searched COMPARE?

2) Sequence data added

3) Automatically show all related articles. 4) Make Pubmed ID clickable,
5) Allow alternative links, 6) Pull other COMPARE allergens associated with this article

Translate

Who will use the COMPARE Database?

HESI.

Who?

- Product developers providing safety information on novel proteins.
- Regulatory agencies responsible for food and feed safety assessments.
- Clinicians, medical personnel and caregivers in the allergy field.
- The Public, who may be interested in identifying sources of allergens.

How?

- Bioinformatics tools such as the FASTA algorithm are utilized to evaluate the degree of similarity between novel proteins introduced into a crop and known allergens – COMPARE allows these bioinformatic comparisons.
- Researching types of allergens, sequence information and literature associated.
- Staying updated with new allergens identified each year (all users).

Access to a transparent and consensus-based allergen database is a key aspect for supporting public safety.

Acknowledgements

HESI PATB Committee

HESI PATB COMPARE Steering Team:

- Katie Bailey, Syngenta
- Laurent Beuf, Vilmorin et Cie
- Christal Bowman, formerly, Bayer
- Supratim Choudhuri, U.S. FDA
- Eva Gietl, KWS
- Kevin Glenn, formerly, Monsanto
- Rod Herman, DowDupont
- Norma Houston, DowDupont
- Emir Islamovic, BASF
- John Kough, U.S. EPA
- Greg Ladics, Dupont
- Scott McClain, Syngenta
- Kyle McKillop, Univ. of Maryland / JIFSAN
- Henry Mirsky, Dupont
- Clare Narrod, Univ. of Maryland / JIFSAN
- Esmeralda Posada C, BASF
- Andre Silvanovich, Bayer
- Ping Song, DowDupont
- Ronald van Ree, Academic Medical Center, University of Amsterdam

Peer Review Panel Experts

- Cecilia Berin, PhD, Icahn School of Medicine at Mount Sinai, NY, USA
- Karin Hoffman-Sommergruber, PhD, Medical University of Vienna, Austria
- Lars Poulsen, PhD, Copenhagen University Hospital, Denmark
- Suzanne Teuber, MD, University of California, Davis, CA, USA
- Ronald van Ree, PhD, Academic Medical Center, University of Amsterdam, The Netherlands

HESI Staff:

Syril Pettit, Exec. Dir.
Lucilia Mouriès, Sc. Prog. Mngr.
Lauren Peel, Associate Sc. Prog. Mngr
Oscar Bermudez, Sc. Prog. Associate
Brianna Farr, Sc. Prog. Associate

COMPARE

Comprehensive Protein Allergen Resource

To Learn more:

- **PATB events and publications:** http://hesiglobal.org/protein-allergens-toxins-and-bioinformatics-committee-patb/
- **COMPARE:** www.comparedatabase.org
- **Questions:** contact the HESI COMPARE Team at: compare@hesiglobal.org

HESI

Convene Translate Protect

www.hesiglobal.org