Health and Environmental
Sciences Institute

740 15th Street, NW
Suite 600
Washington, DC 20005
USA

1.202.854.8404 office
1.202.659.3859 fax
www.hesiglobal.org

COMPARE
Comprehensive Protein Allergen Resource

COMPARE 2024 Update Documentation

The COMPARE Team is pleased to present the release of the COMPARE 2024 database, released on 01/25/2024 (www.comparedatabase.org).

A new version was updated on 02/02/2024. This version corrected a duplicate header line in the fasta file. No sequences were added/removed or changed between the two releases.

1. COMPARE 2024: General Overview

1.1. Database entries

The COMPARE 2024 database consists of **2748 sequence entries**. Overall updates to the 2024 COMPARE database include:

- **151 unique additions** to the COMPARE 2023 database.
  - 35 full-length proteins
  - 116 mass spectrometry (MS) additions
- The **removal of 34 entries** from COMPARE 2023 (see table below).

Therefore, COMPARE 2024 = 2361 entries (COMPARE 2023) - 34 entries (removal from COMPARE 2023) + 151 new unique entries = 2748 sequence entries.

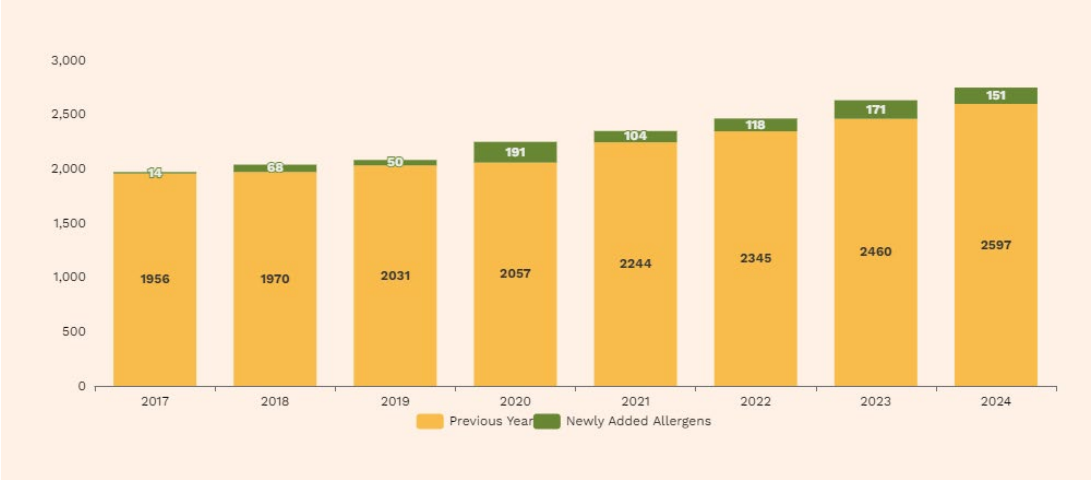


Figure 1: Constant and consistent growth of sequence entries in COMPARE. In 2020 a historical data review was undertaken, and MS sequences were included in the database.

All decisions result from the COMPARE Peer-Review Panel (PRP)'s review of sequence candidates with record dates ranging between May 15, 2022 – May 15, 2023, and associated literature. As in previous years, the candidate sequences were sourced from NCBI Protein, UniProt, AllergenOnline, and IUIS databases, as well as a targeted literature search, for the same time window.

Retired Sequence (Removed from COMPARE 2023)	Replacement Sequence *(new or updated sequence in COMPARE 2024)*	Justification
COMPARE167	ALQ56981.1	Duplicate: COMPARE167 is fully contained in ALQ56981.1
COMPARE227	*(removed)*	Sequence does not contain 10 consecutive amino acids (criteria for inclusion)
COMPARE00433	Q9SPL3.1	Duplicate: COMPARE00433 is fully contained in Q9SPL3.1
COMPARE00438	Q9SPL3.1	Duplicate: COMPARE00438 is fully contained in Q9SPL3.1
COMPARE00442	Q9SPL3.1	Duplicate: COMPARE00442 is fully contained in Q9SPL3.1
COMPARE00443	Q9SPL3.1	Duplicate: COMPARE00443 is fully contained in Q9SPL3.1
COMPARE00445	Q9SPL3.1	Duplicate: COMPARE00445 is fully contained in Q9SPL3.1
COMPARE00446	Q9SPL3.1	Duplicate: COMPARE00446 is fully contained in Q9SPL3.1
COMPARE00447	Q9SPL3.1	Duplicate: COMPARE00447 is fully contained in Q9SPL3.1
COMPARE00448	Q9SPL3.1	Duplicate: COMPARE00448 is fully contained in Q9SPL3.1
COMPARE00449	Q9SPL3.1	Duplicate: COMPARE00449 is fully contained in Q9SPL3.1
COMPARE00455	Q9SPL3.1	Duplicate: COMPARE00455 is fully contained in Q9SPL3.1
COMPARE00458	Q9SPL3.1	Duplicate: COMPARE00458 is fully contained in Q9SPL3.1
COMPARE00460	Q9SPL3.1	Duplicate: COMPARE00460 is fully contained in Q9SPL3.1
COMPARE00461	Q9SPL3.1	Duplicate: COMPARE00461 is fully contained in Q9SPL3.1
COMPARE00462	Q9SPL3.1	Duplicate: COMPARE00462 is fully contained in Q9SPL3.1

COMPARE00464	Q9SPL3.1	Duplicate: COMPARE00464 is fully contained in Q9SPL3.1
COMPARE00436	Q9SPL4.1	Duplicate: COMPARE00436 is fully contained in Q9SPL4.1
COMPARE00437	Q9SPL4.1	Duplicate: COMPARE00437 is fully contained in Q9SPL4.1
COMPARE00439	Q9SPL4.1	Duplicate: COMPARE00439 is fully contained in Q9SPL4.1
COMPARE00451	Q9SPL4.1	Duplicate: COMPARE00451 is fully contained in Q9SPL4.1
COMPARE00452	Q9SPL4.1	Duplicate: COMPARE00452 is fully contained in Q9SPL4.1
COMPARE00454	Q9SPL4.1	Duplicate: COMPARE00454 is fully contained in Q9SPL4.1
COMPARE00456	Q9SPL4.1	Duplicate: COMPARE00456 is fully contained in Q9SPL4.1
COMPARE00327	QDO73345.1	Duplicate: COMPARE00327 is fully contained in QDO73345.1
ABF21078.1	AHM25029.1	Duplicate: ABF21078.1 is fully contained in AHM25029.1
COMPARE105	C0HLQ2	Duplicate: COMPARE105 is fully contained in C0HLQ2
COMPARE106	C0HLQ2	Duplicate: COMPARE106 is fully contained in C0HLQ2
COMPARE00422	XP_042518524.1	Duplicate: COMPARE00422 is fully contained in XP_042518524.1
COMPARE00423	XP_042518524.1	Duplicate: COMPARE00423 is fully contained in XP_042518524.1
COMPARE00424	XP_042518524.1	Duplicate: COMPARE00424 is fully contained in XP_042518524.1
COMPARE00425	XP_042518524.1	Duplicate: COMPARE00425 is fully contained in XP_042518524.1
COMPARE00428	XP_042518524.1	Duplicate: COMPARE00428 is fully contained in XP_042518524.1
BAF47264.1	CBY17558.1	Duplicate: Identical to CBY17558.1

COMPARE 2024 has otherwise followed specifications set in 2019 - e.g., use of “COMPARE #” Accessions when no other public accession number is known for a specific sequence (see paragraph A in “COMPARE 2019 Documentation” file); information sharing via documentation and transparency files, available in the database page under the Documentation tab.

1.2. Accession number updates (past entries)

The **"Accession"** field was updated for **3** allergen entries.

PAST COMPARE Accession	*NEW* COMPARE Accession	COMPARE Description
MANUAL2	**COMPARE247**	cysteine protease
MANUAL3	**COMPARE248**	vitellogenin, partial
COMPARE022	**QFG58557.1**	7S globulin, vicilin, beta-conglutin

1.3. Description updates (past entries)

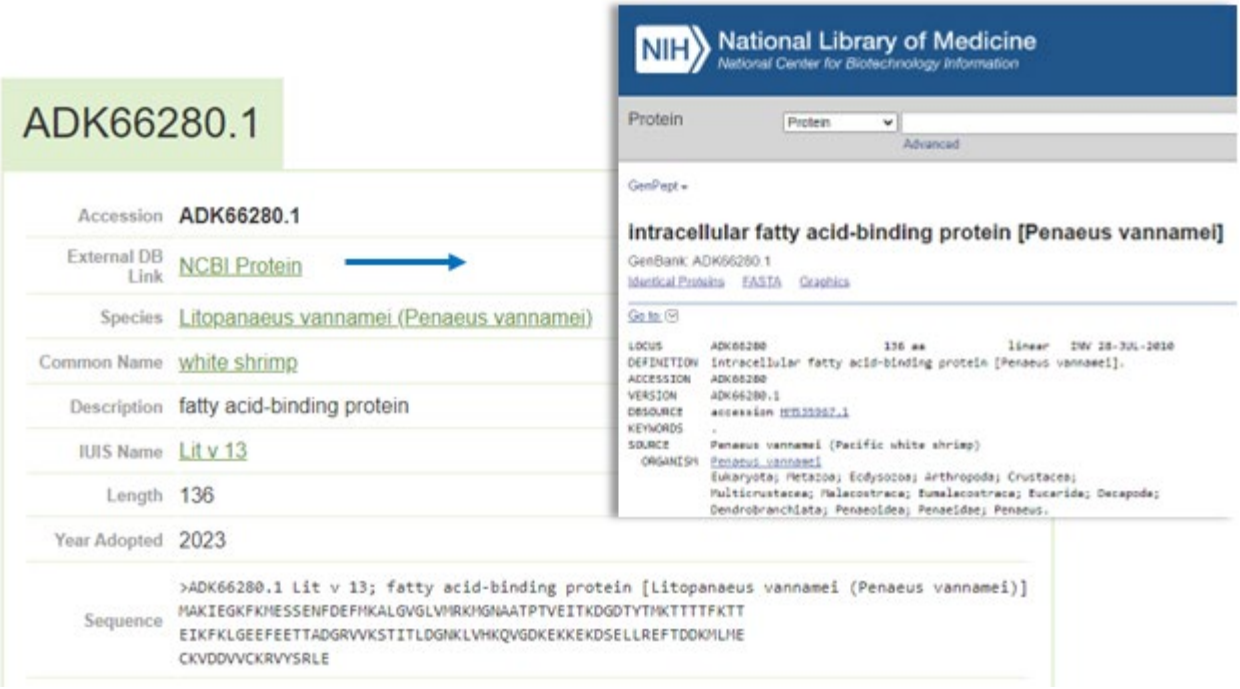
The "**Description**" field was updated for 13 entries.

COMPARE Accession	*PAST* COMPARE Description	*NEW* COMPARE Description
2MC9_A	peptidyl-prolyl isomerase	cyclophilin/peptidyl-prolyl isomerase
CAB44442.1	peptidyl-prolyl isomerase	cyclophilin/peptidyl-prolyl isomerase
CAC84116.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
CAI78448.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
AAP35065.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
AEY79726.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
ALM24136.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
AVV30163.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
CAA09884.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
P81531.2	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
XP_025675300.1	cyclophilin	cyclophilin/peptidyl-prolyl isomerase
ANQ43386.1	chitinase	glyceraldehyde-3-phosphate-dehydrogenase
CAA45085.1	alpha-amylase, partial	alpha-amylase inhibitor, partial

1.4. Other updates to past entries

Accession	Old IUIS Name	New IUIS Name
ABM53752.1	Tyr p 13	None
Accession(s)	**Old Species Name**	**New Species Name**
P80273.2, P33556.1, P80274.1	Vitis sp.	Vitis vinifera
AKV72166.1	Acacia farnesiana	Acacia farnesiana (Vachellia farnesiana)
AIV43662.1	Vachellia farnesiana	Acacia farnesiana (Vachellia farnesiana)
BAH10157.1	Venerupis philippinarum	Ruditapes philippinarum (Venerupis philippinarum)
Accession	**Old Sequence**	**New Sequence**
COMPARE086	IEIESFYEGDOFSETLTR	IEIESFYEGDDFSETLTR

1.5. New! Links to NCBI Protein database have been included on the allergen record page.



1.6. New! A list of biochemical names in COMPARE has been included on the Documentation page.

In 2021, the Peer Review Panel reviewed the description lines of all COMPARE entries to clean-up irrelevant wording from automated text imported from the source databases in order to generate meaningful description lines to the eyes of allergy experts. This list is now included in the Documentation tab here.

1.7. **Enhanced visibility for Mass Spectrometry (MS) fragments: Visualization of MS fragments mapped to their Parent Accession has been incorporated.**

New! A visual representation of MS sequences mapped to their known full length parent sequence (even though the full-length parent protein sequence is not an entry in the database itself) has been included as part of our commitment to continuous improvement. An example is shown in Figure 2 below. COMPARE030 maps to parent sequence Q9FSH9.1. The graphic demonstrates where the peptide sequence is located on the parent and which other peptide sequences in COMPARE are associated with that parent sequence. This helps the user identify the portion of the parent sequence that has published IgE binding evidence.

COMPARE
Comprehensive Protein Allergen Resource

ABOUT DATABASE PROCESS DEVELO

Allergen

COMPARE030

Accession	**COMPARE030**
External DB Link	
Species	Lupinus albus
Common Name	white lupine
Description	7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1
IUIS Name	
Length	21
Year Adopted	2020H_MS
Sequence	>COMPARE030 7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1 [Lupinus albus] AGIALGTHQLEENLVVFDLAR
Parent Accession	Q9FSH9.1

Related Sequences

Peptide sequences mapped to Parent Accession

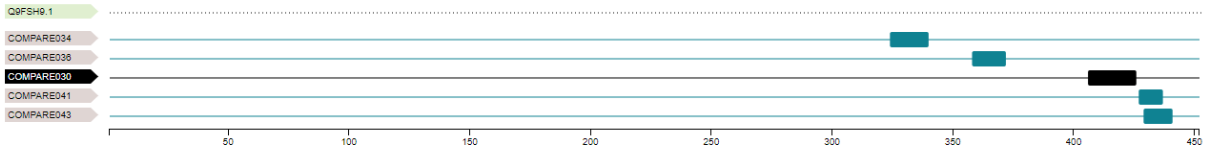


Figure 2: New visualization in COMPARE2024 for records with known Parent Accessions is now available. Allergen record page for COMPARE030.

REMINDER: bioinformatics companion tool, COMPASS (COMPare Analysis of Sequences with Software)

As of June 2019, the COMPARE database is equipped with its companion tool, COMPASS (COMPare Analysis of Sequences with Software), as a built-in feature. COMPASS is a comparative sequence search software, incorporating the open source FASTA software package (FASTA v36). With this tool, COMPARE users can conduct website-based, real-time use of the COMPARE database to identify similarities between a protein sequence of interest and COMPARE's allergens via amino acid sequence alignments (between two or more amino acid sequences). COMPASS offers a visualization option to view results in a color-coded graphic display. To access the tool, go www.comparedatabase.org, click on the "Database" tab and click on the green button "Run COMPASS." For detailed information, instructions on how to use and supporting references, visit the COMPASS' "About" page.

As part of COMPARE 2024's release, COMPASS is also updated to screen against COMPARE 2024 sequences.

Your Feedback is Appreciated - Contact Us

The HESI COMPARE database program is committed to transparency and open dialog. Individuals or organizations are invited to submit feedback, questions, or inquiries via the "Contact us" portal in the COMPARE database website, or email to comparedatabase@hesiglobal.org. HESI staff will respond if the information is readily available or will relay the inquiries to the Peer Review Panel if a more in-depth response is required.

Licensing

When using the database or referring to it in a publication, please cite the COMPARE publication:

> van Ree et al., 2021. The COMPARE Database: A Public Resource for Allergen Identification, Adapted for Continuous Improvement. *Frontiers in Allergy*. https://doi.org/10.3389/falgy.2021.700533.

Support COMPARE!

Is COMPARE useful as a resource and do you like its commitment to continuous improvement? If so, support COMPARE! We have other ideas to continue improving this resource and making it as comprehensive and thorough as possible. The COMPARE database is a collaborative HESI program. The annual update of the database is a resource intensive process that involves many more partners and collaborators, rolling on a steady annual cycle schedule. The execution of the program relies on scientific expertise and in-kind and direct financial support from both public sector and private sector scientific organizations to maintain this free, public resource. If you would like to learn more about how you or your organization can contribute, please contact us at comparedatabase@hesiglobal.org.

Your support is tax deductible in the US: HESI is a non-profit, 501(c)(3) organization committed to generating science for a safer, more sustainable world. Financial support to HESI's scientific programs is considered a tax-deductible charitable donation in the United States. Your support to our mission, through funding and participation, makes our scientific collaborations and outreach possible, and helps improve both human and environmental health across the globe.

We look forward to hearing from you!

About HESI (www.hesiglobal.org): The Health and Environmental Sciences Institute (HESI) is a non-profit institution whose mission is to collaboratively identify and help resolve global health and environmental challenges through the engagement of scientists from academia, government, industry, NGOs, and other strategic partners. Since 1989, HESI has provided the framework for scientists from public and private sectors to meaningfully collaborate in developing science for a safer, more sustainable world.

HESI Convene Translate Protect

www.hesiglobal.org