



COMPARE 2023 Update Documentation

The COMPARE Team is pleased to present the release of the COMPARE 2023 database, released on 01/26/2023 (www.comparedatabase.org).

1. COMPARE 2023: General Overview

1.1. Database entries

The COMPARE 2023 database consists of **2631 protein sequences**. Overall updates to the 2023 COMPARE database include:

- **171 unique additions** to the COMPARE 2022 database.
- The **removal of three entries** from COMPARE 2022 (see table below).

Therefore, COMPARE 2023 = 2463 entries (COMPARE 2022) - 3 removal from COMPARE 2022 + 171 new unique sequences = 2631 entries.

All decisions result from the COMPARE Peer-Review Panel (PRP)'s review of sequence candidates with record dates ranging between May 15, 2021 – May 15, 2022, and associated literature. As in previous years, the candidate sequences were sourced from [NCBI Protein](#), [UniProt](#), and [IUIS](#) databases, as well as a targeted literature search, for the same time window. Sequences were not sourced from [AllergenOnline](#), as done in previous years, because there was no update to AllergenOnline in 2022.

Retired Sequence (Removed from COMPARE 2022)	Replacement Sequence (<i>new or updated sequence</i> in COMPARE 2023)	Justification
P84527.1	(removed)	Duplicate: P84527.1 is fully contained in P84527.2
COMPARE171	(removed)	Sequence does not have at least 10 contiguous, non-X amino acids which is a criterion to be included in COMPARE

**B3EWE5.3***(removed)*

Sequence is only 9 amino acids long. The minimum sequence length accepted for COMPARE is 10 amino acids.

COMPARE 2023 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 **15** allergen entries as part of a task to harmonize the annotation information. Fourteen of the 15 sequences below were reassigned with a “COMPARE #” accession as the past COMPARE accession was not an accession number found in a public sequence database (e.g., NCBI Protein, UniProt or PDB).

PAST COMPARE Accession	NEW COMPARE Accession	NEW COMPARE Description
AHF71021.1A	COMPARE0339	defensin, partial from AHF71021.1
AHF71022.1A	COMPARE0340	defensin, partial from AHF71022.1
AHF71023.1A	COMPARE0341	defensin, partial from AHF71023.1
AHF71024.1A	COMPARE0342	defensin, partial from AHF71024.1
AHF71024.1B	COMPARE0343	defensin, partial from AHF71024.1
BAM22586.1A	COMPARE0346	vitellogenin, partial from BAM22586.1
BAM22586.1B	COMPARE0347	vitellogenin, partial from BAM22586.1
KC700036.1A	COMPARE0348	defensin, partial from AHF71025.1
KC700036.1B	COMPARE0349	defensin, partial from AHF71025.1
Translation from accession AK068307.1	COMPARE0350	unknown function, translation from AK068307.1
Translation from accession AK119653.1	COMPARE0351	unknown function, translation from AK119653.1
Translation from accession AK242260.1	COMPARE0352	unknown function, translation from AK242260.1
P13080	P13080.1	alpha-glucosidase

1.3. Allergen “IUIS Name” field curation (past entries)

The “**IUIS Name**” field was annotated for **635** allergen entries which previously had no IUIS Name listed in COMPARE 2022. This task was done as part of a retroactive review of entries added before 2021, to label allergens with the systematic allergen nomenclature recognized by the [WHO/IUIS Allergen Nomenclature Sub-committee](#). To complete this task, sequences from www.allergen.org were searched against the COMPARE 2022 database. COMPARE sequences



of similar function and length with >70% identity to a sequence in allergen.org from the same species were assigned a corresponding “IUIS Name” based on the expert panel review and recognition that the sequence belonged to the same allergen (or one of its isoforms). In addition, **three** COMPARE entries with existing IUIS Names assignments were updated and are shown in the table below.

COMPARE Accession	IUIS Name in COMPARE2022	Corrected IUIS Name in COMPARE2023
P00785.4	Act c 1	Act d 1
P18153.2	Aed al 2	Aed a 2
CAA76841.1	Can f 4	Can f 3

NEW! Links from COMPARE to the IUIS Database have been included on the Allergen Record page for those entries with annotated IUIS Names.

Allergen

ADK66280.1

Accession	ADK66280.1
Species	Litopanaeus vannamei (Penaeus vannamei)
Common Name	white shrimp
Description	fatty acid-binding protein
IUIS Name	Lit v 13 
Gi	
Length	136
Year Adopted	2023
Sequence	>ADK66280.1 Lit v 13; fatty acid-binding protein [Litopanaeus vannamei (Penaeus vannamei)] MAKIEGKFKMSESNFDEFMKALGVGLVWRKMGNAATPTVEITKDGDTYTKTTTTFTKTEIKFKLGEFEETTADGRVVKSTITLDGNKLVHKQVGDKEKKEKDSLELREFTDDKMLMECKVDDVVK
Parent Accession	

Related Articles

Pubmed	Authors	Title	Journal
34695231	Munera, Marlon, Martinez, Dalgys, Wortmann, Judith, Zakzuk, Josefina, Keller, Walter, Caraballo, Luis and Puerta, Leonardo	Structural and allergenic properties of the fatty acid binding protein from shrimp Litopanaeus vannamei	Allergy. 2022 2022 May (Epub 2021 Oct); 77(5):1534-1544.

Note: Not all IUIS nomenclature assigned by the COMPARE PRP is reflected in allergen.org (i.e., IUIS Database, from the WHO/IUIS Allergen Nomenclature Sub-committee).



1.4. Allergen “description” field curation (past entries)

The PRP continued to examine the “Description” field of past entries and **updated 1700+ entries** to:

- 1) clean-up irrelevant wording from automated text imported from the source databases (e.g., NCBI Protein),
- 2) generate meaningful description lines to the eyes of allergy experts and allergen database users (considering the metadata available from NCBI Protein, UniProt, IUIS, as well as information from the associated literature, related to functional, biochemical, or other types of relevant protein characterization information).

Example for accession A2V735.1

- *Previous description field was:* “RecName: Full=Tropomyosin; AltName: Full=Tropomyosin, slow-tonic isoform; Short=Tm-Chio-tonic”
- *Now updated to:* “tropomyosin”

1.5. Enhanced visibility for Mass Spectrometry (MS) fragments: Parent Accession added to allergen record description line

New! The “Parent Accession” has now been included in the allergen description field so that the user can quickly identify all the MS sequences corresponding to a known full length protein sequence (even though the full-length protein sequence is not an entry in the database itself). An example is shown in the figure below. Q9FSH9.1 is the Parent Accession to five MS sequences in the database.



COMPARE Database

TEST COMPARE 2023 DB Release Date: 12/01/2022

SEARCH:

Run COMPASS
(COMPare Analysis of
Sequences with Software)

[Article Lookup](#) | [2022 FastA \(PDF\)](#) | [2023 FastA \(Text\)](#) |

[2022_Download Data](#) | [Documentation](#)

Search Term: Q9FSH9

Showing 5 total record(s) found.

Species	Common Name	Description	IUIS Name	Accession	Length	Year Adopted	
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1		COMPARE030	21	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1		COMPARE034	17	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1		COMPARE036	15	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1		COMPARE041	11	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial from Q9FSH9.1		COMPARE043	13	2020H_MS	VIEW

Showing 5 total record(s) found.



NOTE: The above-mentioned tasks (1.2-1.5) were performed during the annual review by the [COMPARE Peer Review Panel](#), an independent panel of allergen experts from the public sector.

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 2023's release, COMPASS is also updated to screen against COMPARE 2023 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.

Support COMPARE!

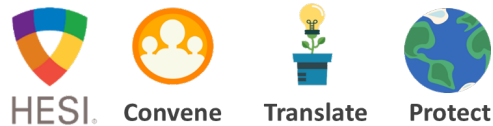
Is COMPARE useful as a resource and do you like its commitment to continuous improvement? If so, support COMPARE! We have other ideas to continue improving this resource and making it as comprehensive and thorough as possible. The COMPARE database is a collaborative HESI program. The annual update of 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.



www.hesiglobal.org