



COMPARE 2022 Update Documentation

The COMPARE Team is pleased to present the release of the COMPARE 2022 database, released on 01/26/2022 (www.comparedatabase.org).

1. COMPARE 2022: General Overview

1.1. Database entries

The COMPARE 2022 database consists of 2463 protein sequences. Overall updates to the 2022 COMPARE include:

- 118 unique additions to the COMPARE 2021 database.
- The removal and replacement of the entries listed in the table on page 2.
- The revision of entry with accession “COMPARE012”, based on errors found in the existing sequence which contained Os instead of Qs as well as an asterisk at the N-terminal sequence.

- “COMPARE012” was previously:

```
MDAIAKKKMRKAMEIEKDNALDRADAAEEKVRQMTDKLERIEEELRDTQKKMMQTENDLDKA  
QEDLSTANSNLEEKEKKYQEAEEVAALNRRMTLLEELERAEEERLKLATDKLEEATHTAD  
ESERVRKVMENRSFQDEERANTYESQLKEAOMLAEADRKYDEVARKLAMVEADLERAEER  
AEAGENKIVELEELRYDGNLKSLEYSEEKALQREDSYEEQIRTVSARLKEAETRAEFAE  
RSVQKLQKEVDRLEDELVHEKERYKSISELDLLQELSGY*
```

- Now revised to:

```
MDAIAKKKMRKAMEIEKDNALDRADAAEEKVRQMTDKLERIEEELRDTQKKMMQTENDLDKA  
QEDLSTANSNLEEKEKKVQEAEEVAALNRRMTLLEELERAEEERLKLATDKLEEATHTAD  
ESERVRKVMENRSFQDEERANTVESQLKEAOMLAEADRKYDEVARKLAMVEADLERAEER  
AEAGENKIVELEELRVDGNLKSLEVSEEKALQREDSYEEQIRTVSARLKEAETRAEFAE  
RSVQKLQKEVDRLEDELVHEKERYKSISELDLLQELSGY
```

Therefore, COMPARE 2022 = 2348 entries (COMPARE 2021) - 4 removals from COMPARE 2021 + 1 updated + 118 new unique sequences = 2463 entries.



All decisions result from the COMPARE Peer-Review Panel (PRP)'s review of sequence candidates with record dates ranging between May 15, 2020 – May 15, 2021 and associated literature. As in previous years, the candidate sequences were sourced from NCBI Protein, UniProt, AllergenOnline (version 21) and IUIS databases, as well as a targeted literature search, for the same time window.

Retired Sequence (Removed from COMPARE 2021)	Replacement Sequence (new or updated sequence in COMPARE 2022)	Justification
COMPARE008	(removed)	This is an artificial sequence of Ara h 15.0101 for recombinant production where the hydrophobic domain was exchanged by a histidine region. Remove from database.
COMPARE104	C0HLQ2 (new)	Duplicate: C0HLQ2 is the full-length sequence of the shorter version COMPARE104; otherwise 100% match, same species.
COMPARE151	COMPARE150 (replaced by)	Duplicate: COMPARE150 is longer than COMPARE151; otherwise 100% match, same sequence.
COMPARE007	ABZ81046.1 (updated)	Paper reviewed again and the sequence in Figure 2 is depicted by ABZ81046.1. COMPARE007 had an extra P at the N-terminus.

COMPARE 2022 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. Allergen “Parent Accession” field curation (past entries)

The “**Parent accession**” field was updated for **28** allergen entries as part of a retroactive review of entries added in 2017-2020. This field applies particularly to smaller fragments, derived from mass-spectrometry studies, and is intended to connect the fragment to the full protein from which it is derived (when indicated in the literature associated or via sequence similarity). A “Parent Accession” number will be indicated in that field when applicable and when available (not all entries will have one). **The parent accession is NOT an entry in the database** and merely provided as additional metadata for users’ reference. In total 105 entries were reviewed. A “Parent Accession” shares 100% sequence identity to the allergen entry, is the same species and can be found in an official protein sequence database (UniProt/GenBank/PDB). New “Parent Accession” assignments are shown in the table below.



COMPARE Accession	NEW Parent Accession	COMPARE Accession	NEW Parent Accession
COMPARE147, COMPARE148	5ZGM_A, 5ZGM_B	COMPARE172	PODKH7
COMPARE129, COMPARE131, COMPARE132, COMPARE133, COMPARE134, COMPARE136	ABD16200.1	COMPARE044, COMPARE046, COMPARE048, COMPARE051, COMPARE053	Q42369
KC700036.1A, KC700036.1B	AHF71025.1	COMPARE030, COMPARE034, COMPARE036, COMPARE041, COMPARE043	Q9FSH9
COMPARE162	EAL84189.1	COMPARE058	Q948Z4.1
COMPARE128	G1uh28	COMPARE022	QFG58557.1
COMPARE002	KAF8044512.1	QIJ32297.1	Blank (MN956521 removed)
COMPARE163	O60022		

NEW! The “Parent Accession” can now be searched in the SEARCH window on the main page of the COMPARE database (<http://db.comparedatabase.org/>) 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 is the Parent Accession to five MS sequences in the database.



ABOUT DATABASE PROCESS DEVELOPMENT TIMELINE EVENTS FAQ CONTACT US

COMPARE Database

COMPARE 2022 DB Release Date: 01/07/2022

[Article Lookup](#) | [2022 FastA \(PDF\)](#) | [2022 FastA \(Text\)](#) | [Download Data](#) | [Documentation](#)

SEARCH:

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		COMPARE030	21	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial		COMPARE034	17	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial		COMPARE036	15	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial		COMPARE041	11	2020H_MS	VIEW
Lupinus albus	white lupine	7S globulin, vicilin, gamma-conglutin, partial		COMPARE043	13	2020H_MS	VIEW

Showing 5 total record(s) found.



The Parent Accession field is shown in the individual Allergen Record (below) but can still be searched in the main SEARCH window (above).

Allergen

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

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

The “**IUIS Name**” field was assigned for **501** allergen entries which previously had no IUIS Name designation in COMPARE 2021. This task was done as part of a retroactive review of entries added before 2021, to clearly label allergens that have been given an official designation by the [WHO/IUIS Allergen Nomenclature Sub-committee](#), with their respective code name. To complete this task, sequences from www.allergen.org were searched against the COMPARE 2021 database. Sequences with 100% identity from the same species with no previous IUIS designation in COMPARE 2021 were assigned their corresponding “IUIS Name”. In addition, **11** existing IUIS Names were updated and are shown in the table below:



COMPARE Accession	IUIS Name in COMPARE2021	Corrected IUIS Name in COMPARE2022
CAB38044.1	Hel a s 13	Hel a s 1
ANZ22901.1	Amb a 13	Amb a 12
ACV85695.1	Car i 1	Cari p 1
AEX34122.1	Pen c 26	Pen cr 26
AAP37412.1	Pol a 4	Pol d 4
AAU11502.1	Try p 13	Tyr p 13
AAG23840.1	Se s i 4	Ses i 4
AAC36740.1	Ma l d 2	Mal d 2
ACM89179.1	Lit v 4.	Lit v 4
ABI98020.1	Lit v 2.	Lit v 2
AAA16792.1	Hev b 10.	Hev b 10

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

During the review of the past entries for “IUIS Names” and “Parent Accessions” assignments, PRP also examined the “Description” field of these entries and updated 577 entries to:

- 1) clean-up irrelevant wording from automated text imported from the source databases (eg, NCBI Protein);
- 2) generate meaningful description lines to the eyes of allergy experts and allergen database users (taking into account 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 P85206.1

- *Previous description field was:* “RecName: Full=Non-specific lipid-transfer protein 2; Short=LTP2; AltName: Allergen=Act d 10.02”
- *Now updated to:* “lipid transfer protein”
- *IUIS Name updated to:* “Act d 10” (COMPARE does not store isoform designations)

NOTE: The three above mentioned tasks were performed during the annual PRP review by COMPARE’s independent panel of allergen experts from the public sector.





2. 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). As of 2020 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 COMPASS' "About" page.

As part of COMPARE 2022's release, COMPASS is now also updated to screen against COMPARE 2022.

3. 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 PRP if a more in-depth response is required.

4. 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.

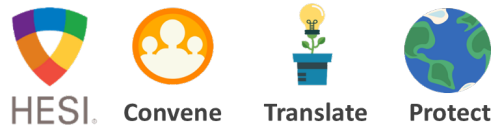
The COMPARE database is a collaborative HESI program. The annual update of the database is a resource intensive process that involves many more partners and collaborators, rolling on a steady annual cycle schedule. The execution of the program relies on the contribution of scientific expertise as well as 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 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