



## **COMPARE 2021 Update Documentation**

The COMPARE Team is pleased to present the release of the COMPARE 2021 database, released on 01/29/2021 ([www.comparedatabase.org](http://www.comparedatabase.org)).

### **1. COMPARE 2021: General Overview**

#### **1.1. Database entries**

The COMPARE 2021 database consists of 2348 protein sequences. Overall updates to the 2021 COMPARE include:

- 104 unique additions to the COMPARE 2020 database.
- The removal and replacement of the entries listed in the table on page 2.
- The revision of entry with accession “MANUAL3”, based on the review of new literature, a Proline residue was added to the C-Terminal sequence
  - “MANUAL3” was previously:  
EIASQIAQEDQSTCEVSKGDFKTFDRMSFTCSFNKSCNVVVAQDCTEHPKFIITTRKVDHQSLSR  
EVHINTSSANITICPAADSSLLVTCNKESVLSDSGVSEYEKDNKIYKNGKTIVIVEAPIHGLKKNVNF  
GEILKVTVASWMRGKTCGVCNNDREKHNELLMPNHKLAHSCSAFVHSHWVLLLEETCSGGCKLQ  
RRYVKLNRNPTIDGEESTCYSVDPVLKCMKDCTPIEKTSVKVGFHCFPKATAVSLLEWQRSSDK  
KSASEDVVESVDADIDCTCTGDSCS
  - Now revised to:  
PEIASQIAQEDQSTCEVSKGDFKTFDRMSFTCSFNKSCNVVVAQDCTEHPKFIITTRKVDHQSLSR  
REVHINTSSANITICPAADSSLLVTCNKESVLSDSGVSEYEKDNKIYKNGKTIVIVEAPIHGLKKNVNF  
DGEILKVTVASWMRGKTCGVCNNDREKHNELLMPNHKLAHSCSAFVHSHWVLLLEETCSGGCKL  
QRRYVKLNRNPTIDGEESTCYSVDPVLKCMKDCTPIEKTSVKVGFHCFPKATAVSLLEWQRSSD  
KKSASEDVVESVDADIDCTCTGDSCS

Therefore, COMPARE 2021 = 2248 entries (COMPARE 2020) - 4 removals from COMPARE 2020, + 104 new unique sequences = 2348 entries.

All decisions result from the COMPARE Peer-Review Panel (PRP)'s review of sequence candidates with record dates ranging between May 2019- May 2020 and associated literature. As in previous years, the candidate sequences were sourced from the NCBI-nr, UniProt, Allergen online (version 20) and IUIS databases, as well as a targeted literature search, for the same time window.



| Retired Sequence<br>(Removed from<br>COMPARE 2020) | Replacement Sequence<br>(new or updated sequence<br>in COMPARE 2021) | Justification   |
|--|--|---|
| 3EHK_A   | Accession CAA55009.1<br>(new)  | Duplicate (CAA55009.1 is longer than 3EHK_A, otherwise 100% match, same species).   |
| Q9FSG7.1   | COMPARE245<br>(new)  | Duplicate (COMPARE245 is 2 aa longer than Q9FSG7.1; otherwise 100% match, same species)   |
| COMPARE099   | MANUAL3<br>(updated)   | Duplicate: with the revision of “MANUAL3” detailed on page 1, COMPARE099 is 100% duplicate.   |
| AAB37406.1   | COMPARE186 and<br>COMPARE187<br>(new)                                | The sequence linked to accession AAB37406.1 (29 aa sequence of Rye) was not in the associated literature (PMID 11260160).<br><br>Instead, the article supports inclusion of a shorter fragment of 23 aa from Rye and a 24 aa fragment of Barley. The two corresponding fragments (COMPARE186 and COMPARE187) are thus added to COMPARE. |

COMPARE 2021 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](#); as well as many user-friendly features implemented with COMPARE 2019.

## 1.2. Database format

In addition to the regular annual update process, **new fields** were added in this 2021 update of the database, as part of COMPARE’s continuous improvement strategies:

- A “**IUIS Name**” field was added in the main database table, to clearly label allergens that have been given an official designation by the [WHO/IUIS Allergen Nomenclature Subcommittee](#), with their respective code name (see [section 1.3](#) for further details).
- A “**Parent accession**” field was added in the details view page of entries. 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). 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.

- Note that in this update, parent accessions are provided only to the new 2021 entries (when applicable and available in the literature associated); a retroactive review of entries added in 2017-2020 (since inception of COMPARE) is being planned to add the information for those past entries as part of next annual update.

### 1.3. Allergen “description” field curation (past entries)

As part of the new IUIS field inclusion, the full content of the COMPARE database needed to be screened and sequences compared to the IUIS Database. To complete this task, each Accession in COMPARE was considered in the context of the IUIS database, by hand, and then reviewed by the independent Peer-Review Panel for curation. Considering the size of the database, the task was broken down in two parts. **This 2021 update includes the outcome from the review of the first part: 621 pre-existing entries for which an official designation was identified in IUIS.** These 621 entries have been updated to reflect the accurate, official, IUIS designation, as listed in the [WHO/IUIS Allergen Nomenclature Sub-committee](#) database, entered under the new “IUIS Name” field in the database. **Note that only these curated 621 entries have content in the “IUIS Name” field in COMPARE 2021; all other COMPARE entries remain as in previous years until further curation.**

The second part of the task will take place during next review cycle for the remaining of the database.

During the review of the 621 entries, PRP also examined the “Description field” and updated it to:

- 1) clean-up irrelevant wording from automated text imported from the source databases (eg, NCBI-nr);

Example for accession P86687.1:

- *Previous description field was:* “RecName: Full=Hyaluronidase; Short=Hya; AltName: Full=Hyaluronoglucosaminidase”
  - *Now updated to:* hyaluronidase
- 2) generate meaningful description lines to the eyes of allergy experts and allergen database users (taking into account the metadata available from NCBI-nr, UniProt, IUIS, as well as information from the associated literature, related to functional, biochemical or other type of relevant protein characterization information).

We anticipate that the 2022 version of COMPARE will be fully curated for allergen descriptions and IUIS designations.



#### 1.4. News rules and decisions adopted this year by the Peer Review Panel

(applied only to the 104 new sequences added in 2021 and the 621 curated entries mentioned in section 1.3)

- The use of the term “putative” is discontinued. IUIS designations are used only when the approved sequence is a 100% match, full or partial, (same species) with a known IUIS entry.
- IUIS isoforms names: IUIS isoforms names will no longer be used for the purpose of COMPARE; only the main IUIS allergen name.
  - Example: isoform “Ole e 1.0101” will be labeled as “Ole e 1”.

These rules will be applied during the second part of the Allergen “description” field curation (described in section 1.3), as well as to new additions in future annual updates.

#### 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). To access the tool, go [www.comparedatabase.org](http://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.

**\*NEW\*** - In July 2020, COMPASS was upgraded to offer a visualization option to view results in a color-coded graphic display.

As part of COMPARE 2021’s release, COMPASS is now also updated to screen against COMPARE 2021.

#### 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](mailto: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 that combines programmatic support from the Joint Institute for Food Safety and Nutrition (JIFSAN), [www.jifsan.umd.edu](http://www.jifsan.umd.edu). 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](mailto: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](http://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](http://www.hesiglobal.org)