



COMPARE 2019 Update Documentation

The COMPARE Team is pleased to present the release of the COMPARE 2019 database.

The COMPARE 2019 database consists of 2081 allergen entries.

There were 50 new unique entries, approved by the COMPARE Peer-Review Panel (PRP), for those sequence candidates dated between May 2017- May 2018.

Overall updates to the 2019 COMPARE include the 50 unique additions to the COMPARE 2018 database, along with the following 6 partial sequence replacement(s) or new record replacements:

- 5 older sequences, fully contained in new entries (exact matches, same source organism) changed for this update. *See table-1 for details.*
- 1 sequence with new record version in NCBI (differing in one aa): AHY24648.2 replaces AHY24648.1

There was one removal of a sequence that has neither been replaced or retired according to PRP procedures:

- Removal of the “GFP-like protein Akane” from Coral (BAW32535), per PRP decision, entered in COMPARE 2018 (see “*COMPARE PRP Statement*” document in the “Documentation” hyperlink of the database page).

**Therefore, COMPARE 2019 =
= 2038 entries (COMPARE 2018) - 6 replacements, - 1 removal, + 50 new unique = 2081 entries.**

Table-1: List of new sequences added to COMPARE 2019, replacing older versions (shorter sequences fully contained in the corresponding newly added sequences; exact matches, same source organism).

Removed sequence (from COMPARE 2018)	Replaced by sequence (added in COMPARE 2019):
P39673.1	BBD75204
BAA04558.1	BBD75204
CAI84642.1	AAV39514.1
ABZ81046.1	COMPARE007
ADD63684.1	ANW82807.1

Other new features in the sequence information of the *50 newly added allergens* in this update are described in the paragraphs below (sections A and B), as well as new “transparency” information (section C) and user-friendly features (section D), available in the online database page itself.



A. “COMPARE #” Accessions

Some new entries added in the COMPARE 2019 update were sequences pulled directly from the literature and are not represented in any public protein database or repository at the time of processing, therefore lacking a public accession #. In those instances, a unique “COMPARE ID number” has been attributed in the following format: COMPARExyz, where “xyz” represent digits.

There are 14 entries in this format: COMPARE001 to COMPARE014.

- A **COMPARE ID** is used only when a public accession # is not available and is listed in the field where the accession# would be entered in the database or in the FastA file.

Example 1:

```
>accession|COMPARE001|Bla g 9.0101 arginine kinase [Blattella germanica]
```

- COMPARE IDs will be replaced by the corresponding public accession # on the next COMPARE annual update following the date at which the accession # has become available.
- A COMPARE ID # is unique and will not be re-assigned, even after “discontinuity”, when replaced by the corresponding public accession #.

B. Allergen “description” fields

For the first time this year, the *description* line for each new entry¹ has been manually curated (in coordination with COMPARE’s Bioinformatics collaborator and PRP), taking into account the metadata available from GenBank, UniProt, IUIS, as well as information from the associated literature (functional, biochemical or other type of relevant protein characterization information).

The reason for this initiative is that in previous COMPARE versions, the “Definition” line was extracted from GenBank records and used “*verbatim*” as a descriptor of an allergen (in the database and in the FASTA file). The reality is that GenBank definition lines are often generated automatically, they include IT coding terms unrelated to protein characterization, and are rarely related to an allergen feature. The COMPARE team saw the potential of having qualified human intervention participating in the review process and build of the database, as an opportunity to generate meaningful description lines to the eyes of allergy experts and allergen database users.

The following step-wise priority approach was adopted in the selection of the terms to describe an allergen:

- 1) **As a first priority, the official IUIS name was adopted and listed as the first terms in the description line,** when one was available in the [IUIS Allergen Nomenclature](#)

¹ Older entries retain the original method used in previous COMPARE versions, for the time being.



[Database](#), for a specific sequence. Additional descriptor terms may follow the IUIS designation.

Example 2:

>accession|ARQ16437.1|**Art an 7.0101 galactose oxidase** [*Artemisia annua*]

- 2) **If an official IUIS name was not available at the time of review, other terms will initiate the description line.** The terms selected pertain to the protein function or biochemical characterization, identified in the Genbank or Uniprot sequence records, or in the supporting literature, and were adopted by PRP.

Example 3:

>accession|ABV55108.1|**Catalase 2** [*Musa acuminata* AAA Group]

- 3) **Precaution to avoid IUIS nomenclature mis-use: use of “putative” to indicate allergen name conforming to the IUIS standard nomenclature, but not officially attributed by IUIS at the time of review:**

Often, allergens in the literature reviewed were attributed a name (in the article) that follows the IUIS standard nomenclature but has not yet been designated officially by IUIS. In those instances, if the sequence is not listed in the IUIS allergen nomenclature database, the term “putative” was added before the allergen name, to make the distinction with *confirmed* official IUIS designations [*i.e.*, entries with description as in paragraph B. 1), above].

Example 4:

>accession|AUH28179.1|**Pollen allergen, Putative Que m 1** [*Quercus mongolica*]

C. Documentation and Transparency

The HESI COMPARE database program is committed to transparency. For that reason, two new “upgrades” are available in COMPARE 2019:

- A new “Documentation” hyperlink is now available in the database page. Under this link, users can find various informational documents for download (including the present Documentation file).
- Decisions and comments from the reviewers recorded during the PRP review process will be made available via a downloadable “Transparency Document” in the database page (under “Documentation”), starting with the release of COMPARE 2019.

D. COMPARE Database user-friendly features

COMPARE 2019 has a fresh new interface, made possible thanks to the tools’ development expertise of collaborators at the Joint Institute for Food Safety and Applied Nutrition ([JIFSAN](#)).

Improvements in the new database format include:

- An **updated referencing system** for the supporting literature, using PubMed IDs.



- Individual records **displaying automatically the list of all related articles, with full reference information** (authors, title, journal and PMID, for each reference associated with a sequence).
- **Clickable metadata** (hyperlinks in the sequence **accession #s** and article **PubMed IDs #s** will redirect the user to the original sources – sequence records in NCBI or UniProt and article abstracts in PubMed).
- **Amino acid sequence data included in each individual record** (in addition to being listed in a FastA format in a downloadable file, as in previous years).

The HESI COMPARE database program is committed to transparency and open dialog. Individuals or organizations are invited to submit their 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.



www.comparedatabase.org