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 2026 Update Documentation

The COMPARE Team is pleased to present the release of the COMPARE 2026 database, released on 01/27/2026 (www.comparedatabase.org).

1. COMPARE 2026: General Overview

1.1. Database entries

The COMPARE 2026 database consists of **2,881 sequence entries**. Overall updates to the 2026 COMPARE database include:

- **62 unique additions** to the COMPARE 2025 database.
- **17 sequences removed** from the COMPARE 2025 database

Therefore, COMPARE 2026 = 2,836 entries (COMPARE 2025) – 17 entries + 62 new unique entries = 2,881 sequence entries.

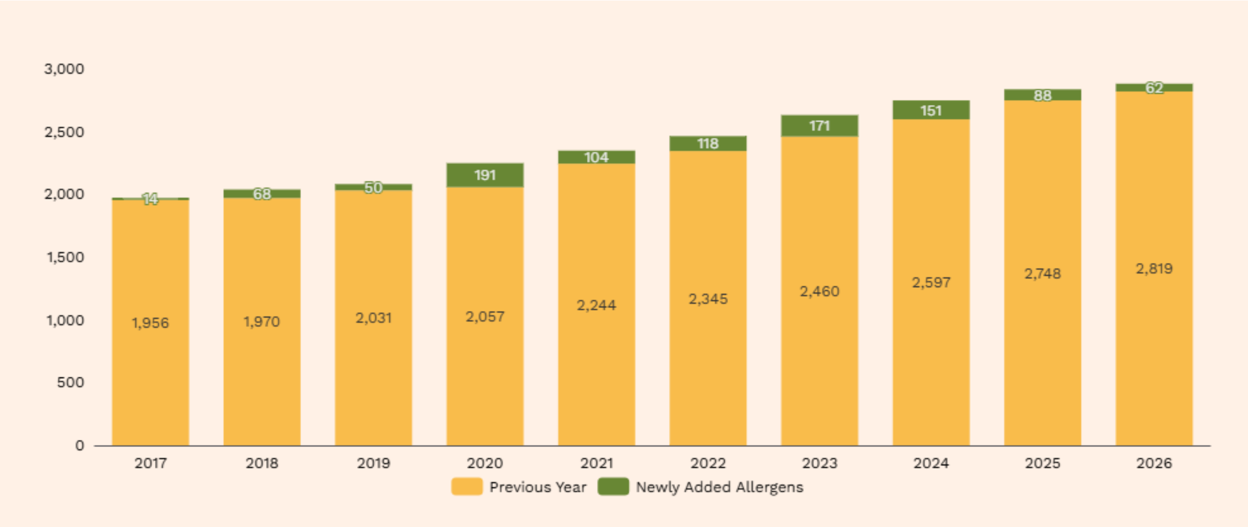

Figure 1: Constant and consistent growth of sequence entries in COMPARE. In 2020 a historical data review was undertaken, and MS sequences started to be included in the database.

All decisions result from the COMPARE Peer-Review Panel (PRP)'s review of sequence candidates with record dates ranging between **May 15, 2024 – May 15, 2025**, 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.

COMPARE 2026 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. Updates to past entries

QC checks conducted to verify the interoperability of sequences between COMPARE and NCBI Protein identified instances of partial sequences in the COMPARE database. To ensure consistency, the accessions in COMPARE were updated to reflect this.

Sequences Removed:

***PAST* COMPARE Accession**	**COMPARE Description**	**Reason**
Q4W1F6	Zea m 25; thioredoxin [Zea mays], 122aa	Removed due to article (PMID: 16364168) lacking appropriate negative controls. Initially included from AOL v16.
Q4W1F8	expansin [Triticum aestivum], 274aa	Removed due to article (PMID: 16364168) lacking appropriate negative controls. Initially included from AOL v16.
Q4W1G2	unknown function [Triticum aestivum], 112aa	Removed due to article (PMID: 16364168) lacking appropriate negative controls. Initially included from AOL v16.
CAI64397.1	unknown function [Triticum aestivum], 137aa	Removed due to lack of appropriate negative controls (PMIDs: 16364168) and irrelevant content in article (PMIDs: 16461145). Initially included from AOL v16.
CAI64396.1	serine carboxypeptidase [Triticum aestivum], 260aa	Removed due to lack of appropriate negative controls (PMIDs: 16364168) and irrelevant content in article (PMIDs: 16461145). Initially included from AOL v16.
CAI64398.1	Tri a 44; lipid transfer protein [Triticum aestivum], 94aa	Removed due to lack of appropriate negative controls (PMIDs: 16364168) and irrelevant content in article (PMIDs: 16461145). Initially included from AOL v16. A longer Tri a 44 already exists: AKJ77990.1

P08819.2	serine carboxypeptidase [Triticum aestivum], 444aa	Removed due to lack of appropriate negative controls (PMIDs: 16364168) and irrelevant content in article (PMIDs: 16461145). Initially included from AOL v16.
COMPARE00930	oleosin, partial from AET74077.1	Fully contained in COMPARE00931
COMPARE00961	11S globulin, cupin, partial	Fully contained in COMPARE00950
COMPARE00952	11S globulin, cupin, partial	Fully contained in COMPARE00950
COMPARE00613	tropomyosin, partial from UTA91547.1 [Alectryonella plicatula (Crassostrea plicatula)]	Fully contained in WES12325.1 (sequence is 100% identical to UTA91547.1)
COMPARE00614		
COMPARE00615		
COMPARE00616		
COMPARE00617		
COMPARE00618		
COMPARE00619		

Accession updates:

PAST **COMPARE Accession**	***NEW*** **COMPARE Accession**	**COMPARE Description**	**Reason**
P02845	**COMPARE01161**	vitellogenin, partial from P02845 [Gallus gallus]	Sequence in COMPARE is not the full-length P02845.
XP_015646887.1	**COMPARE0338**	casein kinase, translation from AK070271.1 [Oryza sativa], 333aa	A new entry has been created to replace the rice allergen sequence XP_015646887.1, based on the corrected translation of cDNA sequence AK070271.1. There was one mismatch between XP_015646887.1 and the translated AK070271.1 sequence.
AAV39514.1	**CAI84642.1**	Tri a 28; alpha-amylase inhibitor [Triticum aestivum] 119aa	AAV39514.1 Removed due to article (PMID: 16364168) lacking appropriate negative controls. Initially included from AOL v16. CAI84642.1 was previously retired due to its inclusion, therefore CAI84642.1 was unretired.

Description updates:

Accession	Old Description	New Description
AAN76862.1	11S globulin, glycinin, cupin	11S globulin, cupin
AKJ77990.1	Endosperm transfer cell-specific PR60	lipid transfer protein
BAG93480.1	unknown function [Oryza sativa]	glycosyl hydrolase 13 family
CAA60533.1	unknown function [Glycine soja]	11S globulin, cupin
ABI32184.1	unknown function [Fagopyrum tataricum]	11S globulin, cupin
Q01882.2	unknown function [Oryza sativa]	alpha-amylase inhibitor
Q01883.2	unknown function [Oryza sativa]	alpha-amylase inhibitor

BAC20650.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAC19997.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07711.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07713.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA01998.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07710.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA01996.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAC20657.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07712.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
AAB99797.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07773.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07772.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
BAA07774.1	unknown function [Oryza sativa]	alpha-amylase inhibitor
AAN76862.1	Ana o 2, 11S globulin, glycinin, cupin [Anacardium occidentale]	Ana o 2, 11S globulin, cupin
BAJ04354.1	unknown function [Cryptomeria japonica]	aspartate protease
AAD13530.2	Bla g 1, microvilli-like protein with unknown function [Blattella germanica]	Bla g 1, nitrile-specifier protein
AAD13531.1	Bla g 1, microvilli-like protein with unknown function [Blattella germanica]	Bla g 1,nitrile-specifier protein
ACY01951.1	unknown function [Cupressus arizonica (Hesperocyparis arizonica)]	calcium-binding protein
AAS93674.1	unknown function [Sarcoptes scabiei]	cysteine protease
CAD23374.1	unknown function [Trichophyton schoenleinii]	dipeptidyl peptidase
CAD23611.1	unknown function [Trichophyton benhamiae (Arthroderma benhamiae)]	dipeptidyl peptidase
BAG88472.1	unknown function [Oryza sativa]	germin-like
BAG86826.1	unknown function [Oryza sativa]	glycosyl hydrolase
BBE74942.1	unknown function [Cryptomeria japonica]	glycosyl hydrolase 5 family
COMPARE0351	unknown function, translation from AK119653.1 [Oryza sativa]	heat shock protein 70
CAA09883.1	Mala s 5, unknown function [Malassezia sympodialis]	Mala s 5, thioredoxin
CCU99457.1	Mala s 5, unknown function [Malassezia sympodialis]	Mala s 5, thioredoxin
ABG35122.1	Der f 22, unknown function [Dermatophagoides farinae]	Der f 22, NPC2-like
AAQ54603.1	unknown function [Glycyphagus domesticus]	NPC2-like
2209273A	unknown function [Zea mays]	Ole e 1-like
AAD46493.1	unknown function [Strongyloides stercoralis]	pathogenesis related protein, PR-1
BAB21489.1	unknown function [Betula platyphylla]	pathogenesis related protein, PR-10, Bet v 1-like
BAB21491.1	unknown function [Betula platyphylla]	pathogenesis related protein, PR-10, Bet v 1-like
BAB21490.1	unknown function [Betula platyphylla]	pathogenesis related protein, PR-10, Bet v 1-like
AAB25850.1	unknown function, partial [Betula sp.]	pathogenesis related protein, PR-10, Bet v 1-like
AAB25851.1	unknown function, partial [Betula sp.]	pathogenesis related protein, PR-10, Bet v 1-like
CAJ28931.1	unknown function [Vespula germanica]	phospholipase A1
BAG95020.1	unknown function [Oryza sativa]	protein kinase
BAA05540.1	unknown function [Bacillus sp.]	serine protease
ABV55106.1	Ani s 9, unknown function [Anisakis simplex]	Ani s 9, SXP/RAL-2 family protein
AAB65434.1	Sol i 3, unknown function [Solenopsis invicta]	Sol i 3, unknown function, antigen 5
P35779.2	Sol r 3, unknown function [Solenopsis richteri]	Sol r 3, unknown function, antigen 5

IUIS Name updates:

Accession & Description	Old IUIS Name	New IUIS Name	Reason
P00785.4 actinidin [Actinidia chinensis]	Act d 1	No IUIS Name	Act d 1 would refer to species Actinidia deliciosa, which this is not. Act c 1 does not exist, therefor leave unassigned.

Sequence updates:

Accession	Reason
P29022	For interoperability with NCBI Protein databases
Old Sequence	**New Sequence**
MANAPRILALGLLALLCAAAGPAAAQNCGCQPNFCCSKFGYCGTTDAYCGDG CQSGPCRSGGGGGGGGGGGGGGSGGANVANVVTDAFFNGIKNQAGSGCEG KNFYTRSAFLSAVNAYPGFAHGGTEVEGKREIAAFFAHVTHETGHFCYISEINKS NAYCDASNRQWPCAAGQKYYGRGPLQISWNYNYGPAGRDIGFNGLADPNRV AQDAVIAFKTALWFWMNNVHGVMPQGFGATIRAINGALECNGNNPAQMN ARVGYYKQYCQQLRVDPGPNLIC	MANAPRILALGLLALLCAAAGPAAAQNCGCQPNFCCSKFGYCGTTDAYCGDG CQSGPCRSGGGGGGGGGGGGGGSGGANVANVVTDAFFNGIKNQAGSGCEG KNFYTRSAFLSAVNAYPGFAHGGTEVEGKREIAAFFAHVTHETGHFCYISEINKS NAYCDASNRQWPCAAGQKYYGRGPLQISWNYNYGPAGRDIGFNGLADPNRV AQDAVIAFKTALWFWMNNVHRVMPQGFGATIRAINGALECNGNNPAQMN ARVGYYKQYCQQLRVDPGPNLTC

1.3 NEW Commercial Data Download Fee

To sustain this valuable resource, we are requiring industry, NGOs and consultants to pay a nominal fee when downloading the database (sequence fasta file + metadata file). This feature remains free for academic and government institutions. Your support is crucial to keeping the online version of COMPARE freely accessible. As always, the web version of the database and its online tool remain publicly accessible and free for all to use at https://comparedatabase.org.

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

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 2026's release, COMPASS is also updated to screen against COMPARE 2026 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.

Citing the database

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 Global (www.hesiglobal.org): The Health and Environmental Sciences Institute (HESI Global) 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