

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) User Guide

Quick Notes: Use [Mozilla Firefox](#) for optimal performance and **PLEASE DO NOT submit more than 10 Level 1 queries at a time**. Wait until they run to completion prior to submitting more. SeqAPASS v1.0 NCBI data was downloaded January 8th, 2016.

Table of Contents

Background	page 2
Accessing SeqAPASS	page 3-4
<i>Returning Users</i> (page 3)	
<i>First Time Users</i> (page 4)	
SeqAPASS Home Tab	page 5
Request SeqAPASS Run Tab	page 6-11
<i>Query “By Species”</i> (page 7)	
<i>Query “By Accession”</i> (page 10)	
SeqAPASS Run Status	page 12-13
View SeqAPASS Reports	page 14
Level 1: Primary Amino Acid Sequence Alignment	page 15-17
Susceptibility Cutoff Box for Level 1	page 18-21
<i>No Ortholog Candidate</i> (page 20)	
Level 2: Functional Domains Alignment	page 22-24
View Level 2 Data Page	page 25-26
Susceptibility Cutoff Box for Level 2	page 27-30
<i>No Ortholog Candidate</i> (page 29)	
Level 3: Individual Amino Acid Residue Alignment	page 31-35
View Level 3 Individual Amino Acid Query and Data Page	page 36-37
Moving Between Level 1, Level 2, and Level 3 Data Pages	page 38
Search, View, and Download Data Tables	page 39-40
Log out	page 41
Pop-up Messages	page 42-44
Detailed Documentation	page 45-53

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Background

The SeqAPASS tool has been developed to predict across species relative intrinsic susceptibility to chemicals with known molecular targets (e.g., pharmaceuticals, pesticides) as well as evaluate conservation of molecular targets from high-throughput screening assays (i.e., U.S. Environmental Protection Agency ToxCast Program) and molecular initiating events (MIEs) and early key events in the adverse outcome framework, as a means to extrapolate such knowledge across species. The term “relative” is used because it is recognized that molecular target similarity is one consideration, though an important one, for making predictions of susceptibility to a chemical. Other important considerations for susceptibility that are not evaluated using the SeqAPASS methodology include how well a chemical is absorbed, distributed, metabolized, and eliminated, life stage, and other life history traits. Also, “relative” indicates that the determination of sequence similarity between proteins is based on comparison to a single protein sequence for a specific species. Additionally, we describe “intrinsic susceptibility” as the vulnerability (or lack thereof) of an organism to chemical perturbation due to its inherent biological composition.

Cross-species comparisons of proteins can be conducted through examination of sequence and structural information, depending on how well the protein has been characterized and what is known about a chemical protein interaction. SeqAPASS allows the user to assess various levels of protein sequence detail across species including comparisons of primary amino acid sequence (including ortholog detection), functional domain(s), and individual amino acid residue positions. Each level requires a greater understanding of the protein and its interaction with a chemical of interest (or similar ligand). Because human and veterinary drugs, as well as pesticides, are designed to act specifically on well characterized molecular targets, these chemical classes have proven useful for demonstrating the utility of the SeqAPASS tool and its application to various hazard assessment/research scenarios.

The pertinent information necessary to begin a SeqAPASS query includes: the identification of a single (or multiple) query species and a query protein, which would be the molecular target(s) of interest (e.g., receptor or enzyme).

The SeqAPASS algorithms mine, collect, and collate information from the National Center for Biotechnology information (NCBI) protein database, conserved domains database, taxonomy database, and strategically utilizes the Stand-Alone Basic Local Alignment Search Tool for proteins (BLASTp) and the Constraint-based Multiple Alignment Tool (COBALT). NCBI data was downloaded January 8th, 2016 for SeqAPASS v1.0.

<http://www.ncbi.nlm.nih.gov/protein/>

<http://www.ncbi.nlm.nih.gov/cdd/>

<http://www.ncbi.nlm.nih.gov/taxonomy/>

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download

http://www.st-va.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi?

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide
Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Accessing SeqAPASS

For optimal SeqAPASS performance use **Mozilla Firefox**

Access SeqAPASS using the following URL: <https://www.seqapass.epa.gov/seqapass/>

Returning Users

Enter Username and Password

The screenshot shows the SeqAPASS login interface. At the top, a blue header contains the text "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)". Below this, a white bar displays "Log In to SeqAPASS" on the left and "Version 1.0" on the right. The main content area is light blue and features a white login box with a grey header "Welcome to SeqAPASS". Inside the box, there are two input fields: "Username" and "Password", followed by a blue "Login" button. Below the button, there is a note: "For optimal SeqAPASS performance use Mozilla Firefox" and a link: "Want an account? To request an account, [click here](#)". At the bottom right of the page, there are two links: "[About SeqAPASS](#)" and "[Report a problem](#)".

Note: If user enters incorrect login information the following message will be displayed

This screenshot shows the same SeqAPASS login interface as above, but with an error message displayed. A yellow warning box with a triangle icon is positioned above the "Username" and "Password" fields. The message reads: "Login Error Invalid credentials (Contact: LaLone.Carlie@epa.gov)". The rest of the page, including the "Login" button and footer links, remains the same as in the previous screenshot.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

First time users

To request a username and password to access the SeqAPASS tool, select “[click here](#)” below the login and a pop-up email will be presented. Send an email to LaLone.Carlie@epa.gov requesting a password. A reply email will be delivered to you with your temporary password.

On the Log in screen the user will enter the provided Login information:

Username: Email address

Password: Temporary password

Upon receiving your temporary password, login to SeqAPASS as described above. Click on “Settings” Tab and Change the temporary password to a user defined password. This is completed by first entering the password from the reply email as the “Current Password” and then typing a new password and re-entering the new password. Click “Change Password.” User will then use the new password to login.



The screenshot displays the SeqAPASS web application interface. At the top, a blue header contains the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" and a "Log out" link. Below the header is a navigation menu with tabs for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "Settings" tab is currently selected. Below the navigation menu, a status bar shows "Welcome to SeqAPASS", "Version 1.0", and "Logged in as: lalone.carlie@epa.gov". The main content area features a "Change Password" form with three input fields: "Current Password", "Enter New Password", and "Re-Enter New Password". A blue "Change Password" button is positioned to the right of the input fields.

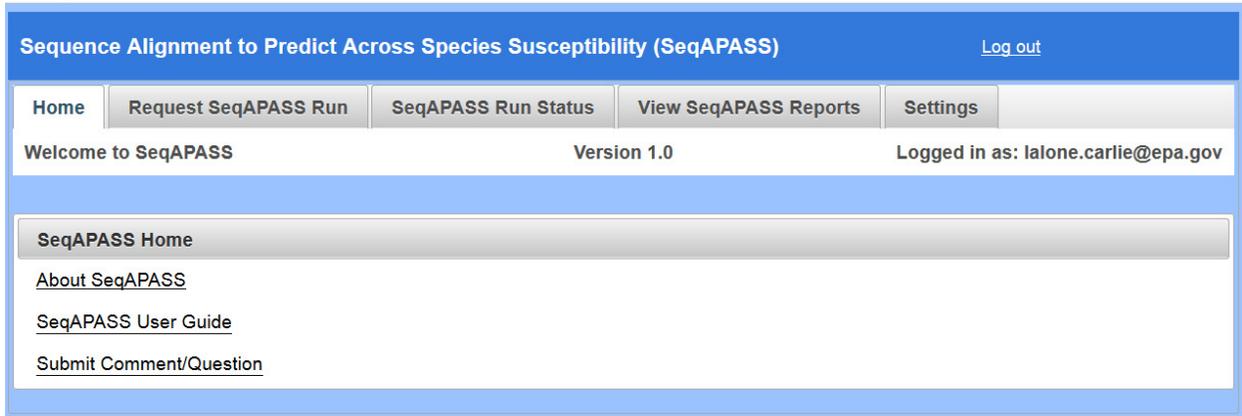
Note: Users can change their password at any time in the “Settings” tab.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

SeqAPASS Home Tab

The “Home” tab indicates who is logged in to the tool (right-hand of the screen) and contains links to obtain information about the SeqAPASS tool (About SeqAPASS), including contact information for support and references to published articles describing the SeqAPASS tool and its applications. Other relevant references to databases and tools are also referenced. A link to the SeqAPASS User Guide can also be found on this page. To Submit a Comment/Question click on the “Submit Comment/Question” link to email the developer. “Log out” icon in upper right-hand corner of screen can be clicked at any time to log out.



The screenshot displays the SeqAPASS web application interface. At the top, a blue header bar contains the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" on the left and a "Log out" link on the right. Below the header is a navigation menu with five tabs: "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "Home" tab is currently selected. Below the navigation menu, a status bar displays "Welcome to SeqAPASS" on the left, "Version 1.0" in the center, and "Logged in as: lalone.carlie@epa.gov" on the right. The main content area features a "SeqAPASS Home" section with three links: "About SeqAPASS", "SeqAPASS User Guide", and "Submit Comment/Question".

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Request SeqAPASS Run Tab

Clicking the “Request SeqAPASS Run” tab opens a page to enter the query information necessary for a SeqAPASS run. Each section of the “Request SeqAPASS Run” will be described below:

The screenshot shows the 'Request SeqAPASS Run' tab in the SeqAPASS application. The page title is 'Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)' and the user is logged in as 'lalone.carlie@epa.gov'. The navigation menu includes 'Home', 'Request SeqAPASS Run', 'SeqAPASS Run Status', 'View SeqAPASS Reports', and 'Settings'. The main content area is divided into three sections: 'Compare Primary Amino Acid Sequences', 'Query Species Selection', and 'SeqAPASS Submission'. The 'Compare Primary Amino Acid Sequences' section has radio buttons for 'By Species' (selected) and 'By Accession'. The 'Query Species Selection' section has a 'Query Species Search' field, an 'Add Query Species' button with a link to 'http://www.ncbi.nlm.nih.gov/taxonomy', and a 'Query Species' list box. The 'Query Protein Selection' section has a 'Query Protein Search' field, a 'Filter Protein' button with a link to 'http://www.ncbi.nlm.nih.gov/protein', a 'Query Proteins' list box, and an 'Add Selected Protein(s)' button. The 'SeqAPASS Submission' section has a 'Final Query Protein(s)' list box, 'Remove Selected Protein(s)' and 'Remove All Proteins' buttons, and 'Request Run' and 'Clear' buttons.

Select Search

There are two options for entering query information: “By Species” or “By Accession” (See radio buttons to the right of “Select Search”). Selecting “By Species” will allow the user to enter text and select from a dropdown list of species and then select a protein from any sequence available for that species in the NCBI protein database. Selecting “By Accession” allows the user to enter a NCBI protein accession.

This close-up screenshot shows the 'Select Search' section of the SeqAPASS interface. It features a 'Select Search:' label followed by two radio buttons: 'By Species' (which is selected) and 'By Accession'.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Query “By Species”

Type the name of the query species of interest in the “Query Species Search” text box. The species common name, scientific name, or Taxid (ID number derived from the NCBI taxonomy database) may be typed into the search bar. This is the species you would like to compare all other species to. The search bar has an autocomplete function and will generate a list of species with their corresponding Taxid. The autocomplete function queries the database in the order of “starts with” then “contains” when text is typed into the search bar. If an integer is typed in the search bar the autocomplete function queries the database in the order of “Taxid”, “starts with”, then “contains.”

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

Request Level 1 SeqAPASS Run Version 1.0 Logged in as: lalone.carlie@epa.gov

Select Search: By Species By Accession

Query Species Selection

Query Species Search: Homo sa

Add Query Species

Query Species:

- Homo sapiens (Taxid:9606)
- Homo sapiens Linnaeus, 1758 (Taxid:9606)
- Trichuris sp. ex Homo sapiens JP-2011 (Taxid:1035824)
- Homo sapiens x Mus musculus hybrid cell line (Taxid:1131344)
- Homo sapiens/Mus musculus xenograft (Taxid:1383439)
- Homo sapiens/Rattus norvegicus xenograft (Taxid:1573476)

Note: The user can also use the NCBI taxonomy database to identify query species using the NCBI link on the right-hand side of the “Add Query Species” button.

Select species of interest by clicking on the name in the drop-down box. Once species is selected, click “Add Query Species” button. This advances the species of interest to the “Query Species” box and fills the “Query Proteins” box with all available protein sequences for that species from the NCBI protein database (although the box only displays the top 200 proteins/species). The protein list includes the protein NCBI accession, protein name, and species scientific name.

Query Species Selection

Query Species Search:

Add Query Species <http://www.ncbi.nlm.nih.gov/taxonomy>

Query Species: Homo sapiens (Taxid:9606)

Query Protein Selection

Query Protein Search:

Filter Protein <http://www.ncbi.nlm.nih.gov/protein>

Query Proteins:

- [NP_005711.1]actin-related protein 2/3 complex subunit 1B
- [XP_006715888.1]PREDICTED: actin-related protein 2/3 complex subunit 1B isoform
- [XP_006715889.1]PREDICTED: actin-related protein 2/3 complex subunit 1B isoform
- [O15143.3]RecName: Full=Actin-related protein 2/3 complex subunit 1B; AltName: I
- [AAB64189.1]p41-Arc

Add Selected Protein(s)

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

To filter the query protein list, type the query protein name or partial name in the “Query Protein Search” box and click the “Filter Protein” button. This action will filter the protein list in the “Query Proteins” box to only display proteins that contain the user defined text. Proteins will be listed in alphabetical order based on NCBI accession Example: typing “estrogen” retrieves all proteins that contain the word “estrogen” in the protein name (the user can scroll to identify protein of interest).

Query Protein Selection

Query Protein Search:

<http://www.ncbi.nlm.nih.gov/protein>

Query Proteins: [A8MWY0.2]RecName: Full=UPF0577 protein KIAA1324-like; AltName: Full=Estroge
[AAA36523.1]estrogen sulfotransferase
[AAA52399.1]estrogen receptor
[AAA52402.1]estrogen receptor, partial
[AAA58461.1]estrogen receptor-related protein

Note: To explore details associated with a protein of interest, click the “Search the NCBI Protein Database” link to the right of the “Filter Protein” button to open NCBI proteins database (See “SeqAPASS Documentation” Section of user guide for details about searching for query proteins using NCBI database).

Highlight the protein or proteins of interest (Ctrl left click to select multiple proteins) in the “Query Proteins” box and click “Add Selected Protein(s)” button. This moves the protein(s) of interest to the “Final Query Protein(s)” box. To remove proteins from the “Final Query Protein(s)” box highlight those to be removed and click the “Remove Selected Protein(s)” button. Select “Remove All Proteins” to discard all proteins from “Final Query Protein(s)” box. The clear button removes all information previously entered on the “Request SeqAPASS Run” page.

Query Protein Selection

Query Protein Search:

<http://www.ncbi.nlm.nih.gov/protein>

Query Proteins: [NP_001278170.1]estrogen receptor isoform 3
[NP_001278641.1]estrogen receptor beta isoform 2
[NP_001278652.1]estrogen receptor beta isoform 2
[NP_001428.1]estrogen receptor beta isoform 1
[NP_001496.1]G-protein coupled estrogen receptor 1

SeqAPASS Submission

Final Query Protein(s) [NP_001278170.1]estrogen receptor isoform 3
[NP_001278652.1]estrogen receptor beta isoform 2
[NP_001496.1]G-protein coupled estrogen receptor 1

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Once user identifies protein(s) to be queried, select “Request Run.” A message will briefly appear in upper right-hand corner of the screen for 10 seconds to alert the user of the request status.

Multiple proteins can be added to the final list for multiple SeqAPASS runs. If another query species is desired, then move to the top to select the next species. Follow the process described above for selecting the proteins associated with this species. The proteins that get filled in the “Query Proteins” box will always be associated with highlighted species from the “Query Species” box.

Note: To avoid longer wait times for the completion of a run, in the current version of SeqAPASS, **PLEASE do not request more than 10 query proteins at a time.**

The screenshot displays the SeqAPASS web interface. At the top right, a grey notification box contains the text "Success Submitted submitted". Below this, the "Query Protein Selection" section features a search input field labeled "Query Protein Search:", a blue "Filter Protein" button, and a purple link "Search the NCBI Protein Database". A large empty text area labeled "Query Proteins:" is positioned below. A blue "Add Selected Protein(s)" button is located at the bottom of this section. The "SeqAPASS Submission" section below it contains a large empty text area labeled "Final Query Proteins(s)". At the bottom of this section are three buttons: "Remove Selected Protein(s)", "Remove All Proteins", and "Request Run". A "Clear" button is also present at the very bottom of the interface.

Note: A user may check the progress of the run by clicking on the “SeqAPASS Run Status” tab. (See SeqAPASS Run Status” section of the user guide for more information)

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Query “By Accession”

The screenshot displays two main sections of the SeqAPASS query interface:

- Query Species Selection:** Features a text input for "Query Species Search", an "Add Query Species" button, and a list of selected species including "Homo sapiens (Taxid:9606)" and "Bos taurus (Taxid:9913)".
- Query Protein Selection:** Includes a "Query Protein Search" input, a "Filter Protein" button, and a link to "Search the NCBI Protein Database". Below this is a scrollable list of protein entries such as "[NP_776588.1]ras-related C3 botulinum toxin substrate 1 precursor [Bos taurus]" and "[AAAF00714.1]GTPase [Bos taurus]". An "Add Selected Protein(s)" button is located at the bottom of this section.

Users familiar with the NCBI database can utilize NCBI protein accessions to query the SeqAPASS tool. This is done by selecting the “By Accession” radio button to the right of the “Select Search” text on the “Request SeqAPASS Run” page.

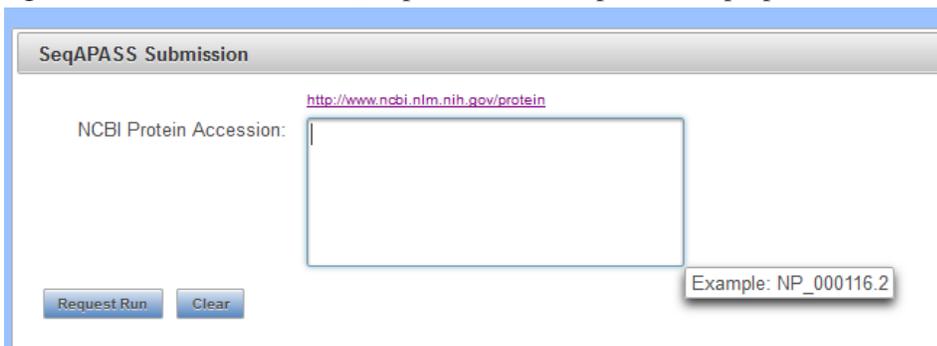
The screenshot shows the main navigation page of SeqAPASS. At the top, the title "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" is displayed alongside a "Log out" link. A navigation menu contains buttons for "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". Below the menu, the page indicates "Request Level 1 SeqAPASS Run", "Version 1.0", and "Logged in as: lalone.carlie@epa.gov". The "Select Search:" section shows two radio buttons: "By Species" (unselected) and "By Accession" (selected).

Upon selecting the “By Accession” radio button, a new query page will be displayed. Type the NCBI protein accession (e.g., NP_000116.2) for the protein of interest (this Accession comes from the NCBI protein database; See “SeqAPASS Documentation” for details) in the “NCBI Protein Accession” box. If desired, more than one NCBI Accession may be entered into the “NCBI Protein Accession” box by clicking the enter key after each additional NCBI Accession entry.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

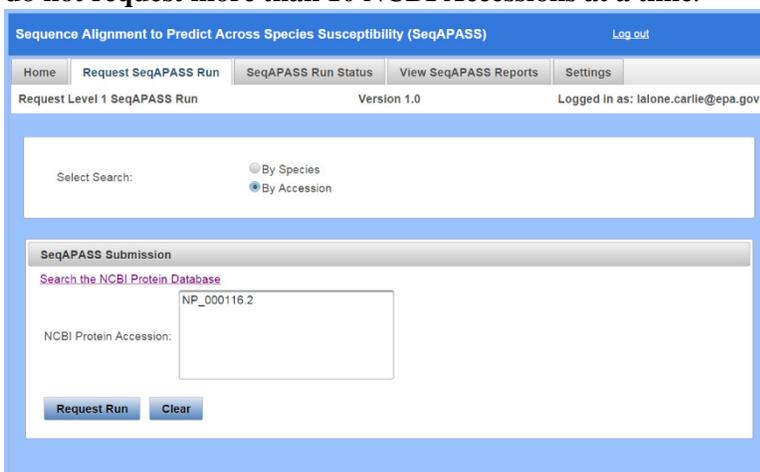
Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon clicking on in the “NCBI Protein Accession” text box, a pop-up message will appear in the lower right-hand side of the text box, to provide an example for the proper format of Accessions to be entered.



The screenshot shows the "SeqAPASS Submission" form. At the top, there is a link: <http://www.ncbi.nlm.nih.gov/protein>. Below it is the "NCBI Protein Accession:" label followed by a text input box. To the right of the input box, a pop-up message displays "Example: NP_000116.2". At the bottom of the form, there are two buttons: "Request Run" and "Clear".

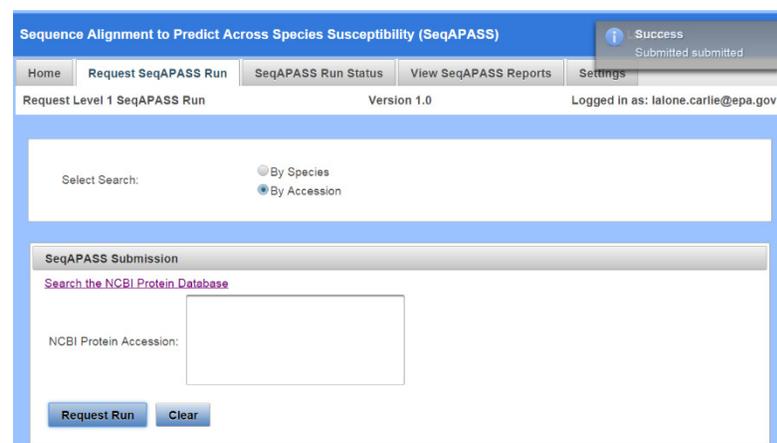
Note: To avoid longer wait times for the completion of a run, in the current version of SeqAPASS, **please do not request more than 10 NCBI Accessions at a time.**



The screenshot shows the main interface of "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)". The top navigation bar includes "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". Below the navigation bar, there are radio buttons for "Select Search:" with options "By Species" and "By Accession". The "By Accession" option is selected. Below this is the "SeqAPASS Submission" form, which includes a link "Search the NCBI Protein Database", a text input box containing "NP_000116.2", and "Request Run" and "Clear" buttons. The top right corner has a "Log out" link. The bottom status bar shows "Request Level 1 SeqAPASS Run", "Version 1.0", and "Logged in as: lalone.carlie@epa.gov".

Clicking the “Clear” button will clear the “NCBI Protein Accession” text box.

After the NCBI accession(s) of interest have been typed in the “NCBI Protein Accession” box, click the “Request Run” button. To remove proteins from the “Accession Area” box click the “Clear” button. A message will briefly appear in the upper right-hand corner of the screen to alert the user of their run request status.



The screenshot shows the main interface of "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" after a successful submission. A "Success" message box is visible in the upper right corner, stating "Submitted submitted". The rest of the interface is identical to the previous screenshot, showing the search options and the submission form.

Note: All NCBI Accessions can include the version number (one digit after the decimal place, e.g., NP_000116.2). Otherwise, the most recent version of the accession will be queried automatically.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

SeqAPASS Run Status

Level 1 SeqAPASS (primary amino acid sequence comparisons) status is displayed as the default. The Accession in the column “Level 1 Query Accession” is that Selected and queried by the user. For a query to finish it must display “complete” in the BLASTp column, 100% in the “Common Domains” column, and 100% in the “Ortholog Candidate” column. The “Common Domains” column displays the % completion for running Reverse Position Specific (RPS)-BLAST (Default e-value of ≤ 0.01) on the Accessions from the Level 1 Full Report. RPS-BLAST, and therefore “Common Domains” status, will take the longest to complete. The “Ortholog Candidate” column displays the % completion for running a reciprocal best hit BLAST evaluation for each hit sequence. The status for the “BLASTp” column is described as “started,” “analyzing,” or “complete.” If the user’s successfully submitted query has entered the run queue, the position of the submitted query in the queue will be indicated in the column (e.g., 2nd in queue). The “Common Domains” and “Ortholog Candidate” columns will also describe the position of the user’s submitted query in the run queue. Once the run has begun processing, the % the run has completed for RPS-BLAST or reciprocal best hit BLAST, respectively, will be displayed. Please see example below:

SeqAPASS Run Id	User	Level 1 Query Accession	BLASTp	Common Domains	Ortholog Candidate	Start Date	Date Completed	SeqAPASS Run Duration
222	lalone.carlie@epa.gov	AAB64189.1	complete	100%	100%	2016 05 23 15:11:27	2016 05 23 16:05:55	54 minute(s) 28 second(s)
221	lalone.carlie@epa.gov	EAL24207.1	complete	100%	100%	2016 05 23 14:09:27	2016 05 23 14:27:23	17 minute(s) 56 second(s)
220	lalone.carlie@epa.gov	P30941.2	complete	100%	100%	2016 05 23 13:31:34	2016 05 23 13:32:29	55 seconds
219	lalone.carlie@epa.gov	BAF36277.1	complete	100%	100%	2016 05 23 10:54:52	2016 05 23 11:11:29	16 minute(s) 37 second(s)
219	lalone.carlie@epa.gov	BAF36278.1	complete	100%	100%	2016 05 23 10:54:52	2016 05 23 11:12:48	17 minute(s) 56 second(s)
217	lalone.carlie@epa.gov	CAAB0601.1	complete	100%	100%	2016 05 23 10:52:43	2016 05 23 10:55:37	2 minute(s) 54 second(s)
217	lalone.carlie@epa.gov	TVDGYF	complete	100%	100%	2016 05 23 10:52:43	2016 05 23 11:18:28	25 minute(s) 45 second(s)
216	lalone.carlie@epa.gov	AAQ94041.1	complete	100%	100%	2016 05 23 08:13:54	2016 05 23 09:20:29	1 hour(s) 6 minute(s) 35 second(s)
215	lalone.carlie@epa.gov	AGN04081.1	complete	100%	100%	2016 05 23 08:12:05	2016 05 23 09:19:54	1 hour(s) 7 minute(s) 49 second(s)
214	lalone.carlie@epa.gov	AAT95191.1	complete	100%	100%	2016 05 23 08:10:57	2016 05 23 09:20:13	1 hour(s) 9 minute(s) 16 second(s)

The user can view the status of requested SeqAPASS runs. Each Run is assigned a unique “SeqAPASS Run Id.” A “Run” is considered a query that was requested either individually or as a batch. The user can view run start and end dates/times, and the duration of the run. (See “Search, View, and Download Data Tables” section of user guide for more information).

The user is also able to view the status of Level 2 (Functional domain(s)) and Level 3 (individual amino acid residue alignments).

View Level 2 Status by selecting the radio button. Also, while viewing the page, the user can click the “Refresh Data” button to refresh the data. “Level 1 Query Accession” column displays the NCBI accession selected and queried by the user. The ‘NCBI Accession’ displays the query sequence identified in the first row of the Level 1 data (top hit: See detailed documentation) and selected for domain comparison in Level 2. Please see below:

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

SeqAPASS Run Status

Version 1.0

Logged in as: lalone.carlie@epa.gov

Level 1 Status
 Level 2 Status
 Level 3 Status

Refresh Data

SeqaPASS Level 2 Run Status

Search: Enter keyword

SeqAPASS Run Id	User	Level 1 Query Accession	NCBI Accession	Domain Type	BLASTp	Start Date	Date Completed	SeqAPASS Run Duration
567	lalone.carlie@epa.gov	EAL24207.1	NP_112601.3	Peptidase_S10	complete	2016 05 23 14:31:01	2016 05 23 14:31:12	11 seconds
568	lalone.carlie@epa.gov	NP_008848.1	NP_008848.1	Nuc_recep-AF1	complete	2016 05 23 13:43:12	2016 05 23 13:43:18	6 seconds
565	lalone.carlie@epa.gov	NP_008848.1	NP_008848.1	NR_DBD_RXR	complete	2016 05 23 13:42:58	2016 05 23 13:43:26	28 seconds
564	lalone.carlie@epa.gov	NP_008848.1	NP_008848.1	NR_LBD_RXR_like	complete	2016 05 23 13:42:31	2016 05 23 13:43:26	55 seconds
563	lalone.carlie@epa.gov	P30941.2	P30941.2	STI	complete	2016 05 23 13:34:44	2016 05 23 13:34:46	2 seconds
562	lalone.carlie@epa.gov	BAF36277.1	BAF36276.1	PHA03087	complete	2016 05 23 13:26:54	2016 05 23 13:26:21	27 seconds
561	lalone.carlie@epa.gov	BAF36277.1	BAF36276.1	7tm_1	complete	2016 05 23 13:25:50	2016 05 23 13:26:21	31 seconds
560	lalone.carlie@epa.gov	BAF36277.1	BAF36276.1	7TM_GPCR_Srx	complete	2016 05 23 13:12:59	2016 05 23 13:13:34	35 seconds
579	Saari.Travis@epa.gov	O95551	NP_057698.2	L1-EN	complete	2016 05 23 12:14:33	2016 05 23 12:14:40	7 seconds
578	lalone.carlie@epa.gov	P21397	NP_000231.1	COG2081	complete	2016 05 23 10:45:09	2016 05 23 10:45:33	24 seconds

(1 of 46) Download Table

Top of Page

View Level 3 Status by selecting the radio button. “Level 1 Query Accession” column displays the NCBI accession selected and queried by the user. The “Job Name” is the user defined name chosen to describe the Level 3 alignment. Also, while viewing the page, the user can click the “Refresh Data” button to refresh the data. Please see below:

SeqAPASS Run Status

Version 1.0

Logged in as: lalone.carlie@epa.gov

Level 1 Status
 Level 2 Status
 Level 3 Status

Refresh Data

SeqaPASS Level 3 Run Status

Search: Enter keyword

SeqAPASS Run Id	User	Job Name	Level 1 Query Accession	Template Accession	COBALT	Start Date	Date Completed	SeqAPASS Run Duration
80	lalone.carlie@epa.gov	test	BAF36277.1	BAF36277.1	complete	2016 05 23 15:14:34	2016 05 23 15:14:35	1 seconds
79	lalone.carlie@epa.gov	test1	EAL24207.1	NP_112601.3	complete	2016 05 23 14:34:33	2016 05 23 14:34:35	2 seconds
78	lalone.carlie@epa.gov	test again	NP_008848.1	NP_008848.1	complete	2016 05 23 13:47:28	2016 05 23 13:47:30	2 seconds
77	lalone.carlie@epa.gov	FASTA test	BAF36277.1	gij117165969 obj BAF36277.1	complete	2016 05 23 13:25:41	2016 05 23 13:25:43	2 seconds
76	Saari.Travis@epa.gov	5-23-16_test1	O95551	NP_057698.2	complete	2016 05 23 12:17:17	2016 05 23 12:17:19	2 seconds
75	lalone.carlie@epa.gov	test12	BAD90546.1	BAD90546.1	complete	2016 05 23 10:34:10	2016 05 23 10:34:12	2 seconds
74	Saari.Travis@epa.gov	5-20-16_test3	P07550.3	XP_003829101.1	complete	2016 05 20 13:37:49	2016 05 20 13:37:52	3 seconds
74	Saari.Travis@epa.gov	5-20-16_test3	P07550	XP_003829101.1	complete	2016 05 20 13:37:49	2016 05 20 13:37:52	3 seconds
72	Saari.Travis@epa.gov	5-20-16_test1	P35638	AAB27103.1	complete	2016 05 20 13:31:56	2016 05 20 13:31:58	2 seconds
70	lalone.carlie@epa.gov	t1	NP_189372.1	NP_189372.1	complete	2016 05 20 12:30:42	2016 05 20 12:30:44	2 seconds

(1 of 7) Download Table

Top of Page

To return to previous tabs click on “Home,” “Request SeqAPASS Run,” or “SeqAPASS Run Status” tabs.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

View SeqAPASS Reports Tab

The “View SeqAPASS Reports” tab provides a table of completed SeqAPASS runs. The completed runs, by default, are listed in the order in which they were completed, with the most recent runs at the top. The table includes information for each run, such as SeqAPASS Run ID (unique for every run regardless of if it is the same protein/species combination ran twice), NCBI Accession, Query Protein Name, taxonomy information for the query species, and the date/time of run completion.

Also, while viewing the page, the user can click the “Refresh Available Reports” button to refresh the table with additional completed runs. Partial protein sequences are highlighted in yellow as illustrated in the example below. (See “Search, View, and Download Data Tables” section of user guide for more information).

To select a completed run to view Level 1 data, select the corresponding radio button in the first column of the table and click “Request Selected Report.” This will open the Level 1 page to view the Level 1 data and to set up queries for Level 2 and Level 3.

Note: The user **MUST** select a radio button **PRIOR** to clicking “Request Selected Report.” If the user fails to select a radio button and clicks “Request Selected Report” a Spinning Wheel will appear and disappear and no completed run will be opened. Further, **there is no pop-up message** indicating that the user did not select a radio button.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

Home Request SeqAPASS Run SeqAPASS Run Status **View SeqAPASS Reports** Settings

SeqAPASS Reports Version 1.0 Logged in as: lalone.carlie@epa.gov

Partial Protein Sequence

Available Reports

Search:

SeqAPASS Run Id	Level 1 Query Accession	Query Protein Name	NCBI Taxonomy ID	Query Species Name	Query Common Name	Taxonomy	SeqAPASS	
<input type="radio"/>	182	ABL64075	ovarian aromatase	8167	Perca flavescens	yellow perch	Actinopteri	2016 05
<input type="radio"/>	181	XP_003831055.1	PREDICTED: cyclic AMP-dependent transcription factor AT	9597	Pan paniscus	pygmy chimpanzee	Mammalia	2016 05
<input type="radio"/>	180	XP_009423541.1	PREDICTED: cyclic AMP-dependent transcription factor AT	9597	Pan paniscus	pygmy chimpanzee	Mammalia	2016 05
<input type="radio"/>	179	NP_001094639.1	long-chain fatty acid transport protein 6	9913	Bos taurus	cattle	Mammalia	2016 05
<input type="radio"/>	177	AAI49775.1	SLC27A6 protein	9913	Bos taurus	cattle	Mammalia	2016 05
<input type="radio"/>	169	NP_001038735.2	finTRIM family, member 14	7955	Danio rerio	zebrafish	Actinopteri	2016 05
<input checked="" type="radio"/>	167	NP_000116.2	estrogen receptor isoform 1	9606	Homo sapiens	human	Mammalia	2016 05
<input type="radio"/>	166	AAH00827.1	NCAPG protein	9606	Homo sapiens	human	Mammalia	2016 05
<input type="radio"/>	165	KTG36042.1	hypothetical protein cypCar_00042704, partial	7962	Cyprinus carpio	common carp	Actinopteri	2016 05
<input type="radio"/>	165	AEA02456.1	hypothetical protein, partial	552617	Carpodacus vinnaceus	vinnaceous rosefinch	Aves	2016 05

(10 of 112)

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 1: Primary Amino Acid Sequence Alignment

From the “View SeqAPASS Reports” tab, upon selecting a radio button and clicking “Request Selected Report” the Level 1 data will be displayed.

The “Level 1 Query Protein Information” box contains the SeqAPASS Run ID, Query Accession, Ortholog Count (# of hits identified as ortholog candidates to the query species protein sequence), NCBI Data (displays the date that NCBI databases were downloaded and incorporated into the SeqAPASS database), Query Species, and Query Protein. Other information in this box will be described below.

The default table is the “Primary Report”, which includes query protein information in the first row below the column titles (Please see “Documentation Section” for details on how the top row query sequence is determined from the BLASTp output), followed by hit proteins whose sequences aligned with the query protein. The hit proteins are ordered from the highest to lowest percent similarity (Maximum percent similarity =100%). For each hit protein, species information is provided including the “Protein Count” which indicates the number of protein records per species in the NCBI protein database, taxonomic information, and species names. Also included are the NCBI protein accession, protein name, BLASTP bitscore (describes overall quality of the alignment, See BLASTp tutorials), and percent similarity ($[\text{hit bitscore}/\text{query bitscore}] * 100$). If the hit protein has been identified as an ortholog candidate (using

The screenshot displays the SeqAPASS web interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. The current page is 'View SeqAPASS Reports'. Below the navigation, there's a 'Main' tab and a 'Level 1' sub-tab. The 'Level 1 Query Protein Information' section shows details for a query protein: 'estrogen receptor bottom 1' from 'Homo sapiens'. It includes a 'Susceptibility Cut-off' plot, 'Level 2 Query Domain' options, and 'Level 3 Query Amino Acid Residue(s)' options. Below this, there are radio buttons for 'Primary Report', 'Full Report', and 'Partial Hit Protein Sequences'. The 'Level 1 Data - Primary' section shows a table with columns for NCBI Accession, Protein Count, Species, Taxonomic Group, Gene Name, Common Name, Protein Name, BLASTP Bitscore, Ortholog Candidate, Ortholog Count, Cut-off, Percent Similarity, and Susceptibility Prediction.

NCBI Accession	Protein Count	Species	Taxonomic Group	Gene Name	Common Name	Protein Name	BLASTP Bitscore	Ortholog Candidate	Ortholog Count	Cut-off	Percent Similarity	Susceptibility Prediction
NP_001116	89432	Homo sapiens	Mammalia	estrogen receptor	human	estrogen receptor bottom 1	1241.9	Y	247	31.2	100.0	Y
AAU02886.1	80175	synthetic construct		estrogen receptor	primer	estrogen receptor 1	1241.9	Y	247	31.2	100.0	Y
AAU01317.1	2	Mammalian expression vector pCIS-2.1		estrogen receptor	other sequences	green fluorescent protein-estrogen receptor gene fusion	1234.6	N	247	31.2	99.4	Y
NP_003119994.1	65063	Homo sapiens	Mammalia	estrogen receptor	chimpanzee	estrogen receptor bottom 1	1229.6	Y	247	31.2	99.0	Y
U07051.1	47156	Homo sapiens	Mammalia	estrogen receptor	human	estrogen receptor bottom 1	1129.0	Y	247	31.2	96.6	Y

reciprocal best hit blast method), it will be noted with a “Y” for yes or if not an ortholog candidate, a “N”, for no. If the hit protein is predicted to be susceptible according to the susceptibility cut-off criteria, that will also be noted with a “Y” for yes or alternatively an “N” for no. The date the analysis was completed is also identified. The data also includes a column describing the number of ortholog candidates identified using the reciprocal best hit BLAST method. The susceptibility cut-off is also listed in a column. The cut-off is determined through identifying local minimums in the density plot of the percent similarity values for the primary report data set and evaluation of ortholog candidates. Links out to the NCBI Protein Database and NCBI Taxonomy Database (specific to the data row) are embedded in the Level 1 data table

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

for “NCBI Accession,” “Species Tax ID,” “Scientific Name,” and “Protein Name” columns. (See “Search, View, and Download Data Tables” section of user guide for more information).

Default highlights identify partial protein sequences, sequences with a bitscore higher than the query sequence and therefore percent similarity greater than 100% (commonly synthetic constructs), and when zero ortholog candidates are identified (in this case a user should consider a different query sequence). Please see “Susceptibility Cutoff Box for Level 1” section of user guide for details when no orthologs are detected.

Level 1 Data - Primary

Search: Enter keyword

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein
NP_000119.2	894432	9508	Mammalia	Homo sapiens	human	estrogen.re
AAK42995.1	80175	32630	synthetic construct	synthetic construct	primer	estroge
AAD47137.1	2	88432	Mammalian expression vector pCI-nG	Mammalian expression vector pCI-nGL1-HEC	other sequences	green fluorescent protein-
XP_003311596.1	63068	9598	Mammalia	Pan troglodytes	chimpanzee	PREDICTED_estro
XP_003811544.1	47189	9597	Mammalia	Pan paniscus	pygmy chimpanzee	PREDICTED_estro
XP_002817538.1	44232	9601	Mammalia	Pongo abelii	Sumatran orangutan	PREDICTED_estro
XP_005552208.1	89797	9541	Mammalia	Macaca fascicularis	crab-eating macaque	PREDICTED_estro
XP_011799152.1	11	336983	Mammalia	Colobus anqolensis palliatus	primates	PREDICTED_estro
NP_001159059.1	55290	9555	Mammalia	Papio anubis	olive baboon	estrou
XP_011852190.1	219	9588	Mammalia	Mandrillus leucophaeus	drill	PREDICTED_estro

(1 of 81) 1 2 3 4 5 6 7 8 9 10 10 Download Table

Top of Page

The user can select the “Full Report” on the “Level 1” page, which includes the same information as the “Primary Report” and additional information pertaining to the alignment of the protein sequence using BLASTp. Additional information includes the number of amino acid residues in the sequence (Hit Length), the number of exact matching amino acids between the hit and query sequence (Identity), the number of exact and similar matches in amino acids between the hit and the query sequence (Positives), the expect value (E-value) describing the number of different alignments expected to occur in the database search by chance, and the conserved domain count. The conserved domain count identifies all domains associated with the query protein in the NCBI conserved domains database (Specific hits, Non-specific hits, Superfamilies, and Multi-domains; See NCBI conserved domains database for details). SeqAPASS algorithms record the query sequence coverage of each curated domain and compares that coverage to that of the hit sequence. If the hit sequence covers the curated domain greater than or equal to the query sequence, then the domain is considered a common domain between the hit and query. The number of common domains comparing each hit sequence to the query sequence are summed and reported. This column displays “0” when the hit protein and query protein do not have any common domains. (See “Search, View, and Download Data Tables” section of user guide for more information).

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 1 Data - Full

Search:

Hit Length	Identity	Positives	E-value	BLASTp Bitscore	Ortholog Candidate	Ortholog Count	Cut-off	Common Domain Count	Percent Similarity	Susceptibility Prediction	Analysis Completed
595	595	595	0.000E0	1241.9	Y	247	31.2	78	100.0	Y	2016 03 11 14:55:31
596	595	595	0.000E0	1241.9	Y	247	31.2	78	100.0	Y	2016 03 11 14:55:31
866	591	592	0.000E0	1234.5	N	247	31.2	76	99.4	Y	2016 03 11 14:55:31
595	590	592	0.000E0	1229.5	Y	247	31.2	75	99.0	Y	2016 03 11 14:55:31
595	589	592	0.000E0	1228.0	Y	247	31.2	75	98.9	Y	2016 03 11 14:55:31
595	589	591	0.000E0	1227.6	Y	247	31.2	75	98.9	Y	2016 03 11 14:55:31
595	588	592	0.000E0	1227.2	Y	247	31.2	75	98.8	Y	2016 03 11 14:55:31
595	586	592	0.000E0	1225.7	Y	247	31.2	73	98.7	Y	2016 03 11 14:55:31
595	587	592	0.000E0	1225.7	Y	247	31.2	75	98.7	Y	2016 03 11 14:55:31
595	587	591	0.000E0	1223.8	Y	247	31.2	75	98.5	Y	2016 03 11 14:55:31

(1 of 81) [Download Table](#)

Top of Page

Note: The “Default Settings” box indicates that the “Primary Report” is only displaying data for hits whose E-value are ≤ 0.01 and have been identified to have ≥ 1 domain in common with the query sequence. Further, the Taxonomic Group column is set to identify and report the “Class” from the NCBI Taxonomy Database. However, if Class is not identified in the NCBI Taxonomic Hierarchy associated with the hit accession, then the algorithm will report the next available Taxonomic Group moving from Class to Order, to Family, to Genus. Finally, the susceptibility predictions are set by using species read-across. (Please view Documentation Section of the User Guide for details on Read-Across settings). Briefly, Species Read-across is used to set the susceptibility prediction, where all ortholog candidates are Susceptible = Y; all species listed above the susceptibility cut-off are Susceptible = Y; all species below the cut-off from the same taxonomic group of one or more species above the cut-off are Susceptible = Y; and those below the cut-off that are not ortholog candidates and do not belong to a taxonomic group above the cut-off are Susceptible = N. The default information can be quickly viewed in the “Default Settings” box on the top of the Level 1 page.

Default Settings

E-value: 0.01

Sorted by Taxonomic Group: Class

Common Domains: 1

Species Read-Across: Yes

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Susceptibility Cutoff Box for Level 1

The susceptibility prediction is determined by identifying ortholog candidates, sequences above a defined susceptibility cutoff, or by identifying those species below the susceptibility cut-off from an organism class above the susceptibility cutoff. The default susceptibility cut-off is set by plotting the distribution of percent similarities calculated for each hit protein. From this plot, the critical points are identified and the local minimums and maximums reported. Using the ortholog candidate data, a susceptibility cut-off is automatically determined by identifying the first ortholog candidate at an equal or higher percent similarity than the first local minimum percent similarity. The user can view this graph by clicking the “View Cutoff” button in the “Susceptibility Cut-off” box. Radio buttons located to the right of the graphical display indicate which Cut-off has been applied for the evaluation of susceptibility in the report. These radio buttons can be selected to change the cut-off in the table to the 2nd local minimum, where the 2nd local minimum is identified in the density plot and the first ortholog candidate at an equal or higher percent similarity than the second local minimum percent similarity is used to set the cut-off. Or the user can define the local minimum by clicking on the “User Defined” radio button. Alternatively, the user can view the closely examine the density plot and manipulate the cut-off by clicking the “View Cutoff” button.



Upon clicking “View Cutoff” button, a new page is displayed with a drop-down that allows the user to set the susceptibility cut-off using the first local minimum and the identified ortholog candidate, the second local minimum and the identified ortholog candidate, or by the “User defined cut-off” (where the user selects the cutoff). To update the cut-off in the Level 1 data report and/or return to the Level 1 page, click “Update Cut-off” button.

Note: The user should have direct empirical evidence that species above the user defined cutoff are susceptible via the protein of interest, or that the species below the user defined cutoff are not susceptible.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide
 Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon selecting the User defined cut-off from the dropdown, the Enter Cut-off text box becomes active and the user can enter a number 1-100.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)
[Log out](#)

Level 1 Susceptibility Cut-off

Local minimums are identified and susceptibility cut-off is set based on % similarity of next ortholog candidate. Use update cut-off button to go back to Level 1 data.

SeqAPASS ID: 4 Query Accession: [NP_000116.2](#) Ortholog Count: 247 NCBI Data: 02/01/2015

Query Species: Homo sapiens

Query Protein: estrogen receptor isoform 1

Select Cut-off: Default: Identify 1st local minimum and find next ortholog candidate Enter Cut-off:

Default: Identify 1st local minimum and find next ortholog candidate
 2 minimum: Identify 2nd local minimum and find next ortholog candidate
 User defined cut-off

Density Plot

Cut-off Based on Ortholog Candidates

Cut-off #	Susceptibility Cut-off
1	31.2
2	59.6

All potential susceptibility cut-offs generated by the data distribution and ortholog candidate identification are reported in the table with columns “Cut-off #” and “Susceptibility Cut-off”. The user can use these numbers to define a cut-off if empirical evidence suggests that the “Default” or “2 minimum” are not supported.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Note: In the case that 0 orthologs are identified from the hit data, the cutoff will be set by the local minimums only, therefore the susceptibility prediction will NOT take into account ortholog candidates. It is recommended that the user identify a different query sequence for the susceptibility predictions. Here, the susceptibility predictions will be highlighted in dark pink in the Level 1 data table to indicate that 0 orthologs were detected and the susceptibility cutoff was determined from plotting the distribution of percent similarities and identifying the local minimums.

The screenshot shows the SeqAPASS web interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. Below these, it says "SeqAPASS Reports" and "Version 1.0". The user is logged in as "lalone.carlie@epa.gov".

The main content area is titled "Level 1 Query Protein Information". It displays the following information:

- SeqAPASS ID: 19 | Query Accession: [CAA74340.1](#) | Ortholog Count: 0 | NCBI Date: 02/01/2015
- Query Species: Bubalus bubalis
- Query Protein: Insulin receptor

There are three main panels for analysis:

- Susceptibility Cut-off:** Includes a graph showing a distribution curve with a local minimum. Radio buttons for "Default", "Second Local Minimum", and "User Defined" are present. A "View Cutoff" button is at the bottom. Below the graph are "Default Settings": Evalue: 0.01, Sorted by Taxonomic Group/Class, Common Domains: 1, Species Read-Across: Yes.
- Level 2 Query Domain:** Includes a "Functional Domains" section with a "Select Domain" dropdown and a "Request Domain Run" button. Below is a "View Level 2 Data" section with a "Choose Domain to View" dropdown and a "View Level 2 Data" button.
- Level 3 Query Amino Acid Residue(s):** Includes a "Select Template Sequence" field, a "Request Residue Run" button, and a "View Level 3 Data" section with a "Choose Query to View" dropdown and a "View Level 3 Data" button.

At the bottom left, there are checkboxes for "Primary Report" (selected) and "Full Report". To the right, there are checkboxes for "Partial Hit Protein Sequence", "Include Orthologs", and "Include Species with Missing Domains".

The "Level 1 Data - Primary" table is shown below, with a search bar. The table has the following columns: NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Solotho Name, Common Name, Protein Name, BLASTp BitScore, Ortholog Candidate, Ortholog Count, Cut-off, Percent Similarity, and Susceptibility Prediction. The table contains three rows of data, all of which are highlighted in dark pink, indicating that 0 orthologs were detected for each.

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Solotho Name	Common Name	Protein Name	BLASTp BitScore	Ortholog Candidate	Ortholog Count	Cut-off	Percent Similarity	Susceptibility Prediction
CAA74340.1	44215	89482	Mammalia	Bubalus bubalis	water buffalo	insulin receptor	228.1	Y	0	49.0	100.0	Y
XP_010828668.1	35841	43148	Mammalia	Bos taurus	domestic ungulate	PREDICTED: insulin receptor domain XI	222.6	N	0	49.0	98.9	Y
XP_005298772.1	130778	9913	Mammalia	Bos taurus	cow	PREDICTED: insulin receptor domain XI	222.6	N	0	49.0	98.9	Y

No Orthologs Detected

If no orthologs are detected from reciprocal best hit blast analysis “Ortholog Count” will be “0” at the top of the “Level 1 Query Protein Information” page. Further, dark pink will highlight the entire query/hits table indicating that the “Susceptibility Prediction” columns were determined from the local minimums identified in the “View Cutoff” density plot, without consideration of orthologs.

This screenshot shows the "Level 1 Query Protein Information" page. It displays the following information:

- Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.
- SeqAPASS ID: 19 | Query Accession: [CAA74340.1](#) | Ortholog Count: 0 | NCBI Date: 02/01/2015
- Query Species: Bubalus bubalis
- Query Protein: insulin receptor

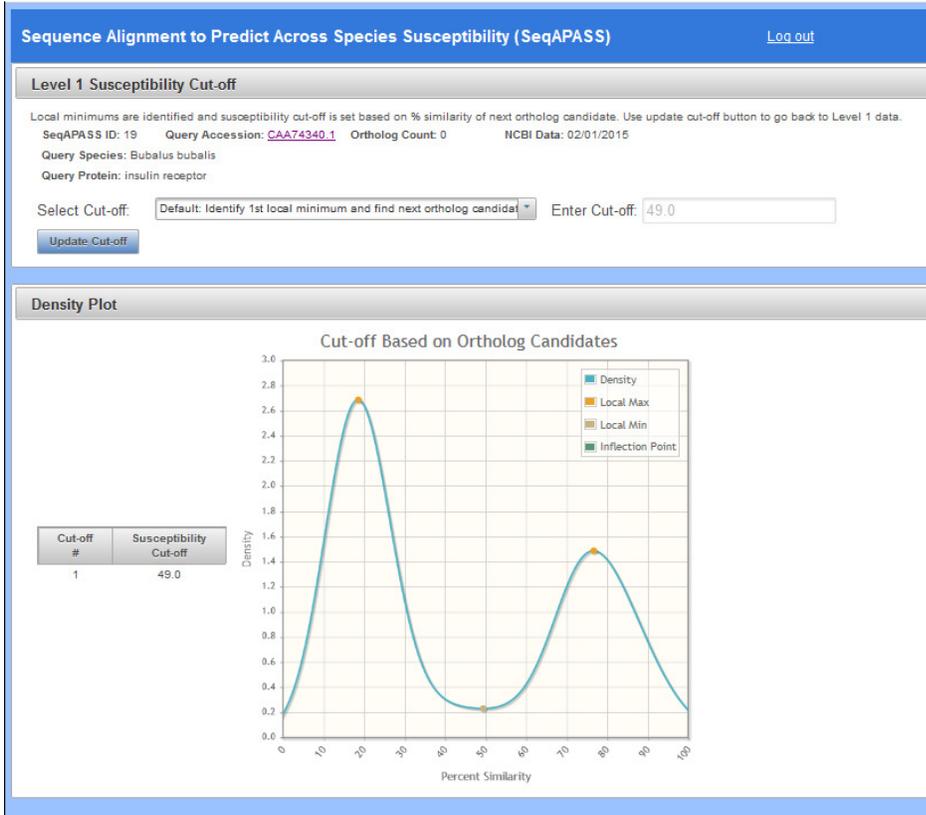
The table below shows the hit data, with all rows highlighted in dark pink to indicate that 0 orthologs were detected.

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Solotho Name	Common Name	Protein Name	BLASTp BitScore	Ortholog Candidate	Ortholog Count	Cut-off	Percent Similarity	Susceptibility Prediction
CAA74340.1	44215	89482	Mammalia	Bubalus bubalis	water buffalo	insulin receptor	228.1	Y	0	49.0	100.0	Y
XP_010828668.1	35841	43148	Mammalia	Bos taurus	domestic ungulate	PREDICTED: insulin receptor domain XI	222.6	N	0	49.0	98.9	Y
XP_005298772.1	130778	9913	Mammalia	Bos taurus	cow	PREDICTED: insulin receptor domain XI	222.6	N	0	49.0	98.9	Y

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

By clicking on the “View Cutoff” button when no orthologs are detected, the “Cut-off #” and “Susceptibility Cut-off” columns will report only the local minimum values.



From the “Level 1” page the user can return to the list of completed SeqAPASS runs by clicking the “Main” button on the upper left-hand side of the “Level 1 Query Protein Information” page.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

SeqAPASS Reports Version 1.0 Logged in as: lalone.carlie@epa.gov

Main Level 1

Level 1 Query Protein Information

Hit proteins are identified for the following query protein. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 19 Query Accession: [CAA74340.1](#) Ortholog Count: 0 NCBI Data: 02/01/2015

Query Species: Bubalus bubalis
Query Protein: insulin receptor

Susceptibility Cut-off

Default
 Second Local Minimum
 User Defined

Level 2

Level 2 Query Domain

[NCBI Conserved Domain Database](#)

Functional Domains

Level 3

Level 3 Query Amino Acid Residue(s)

[NCBI Protein Database](#)

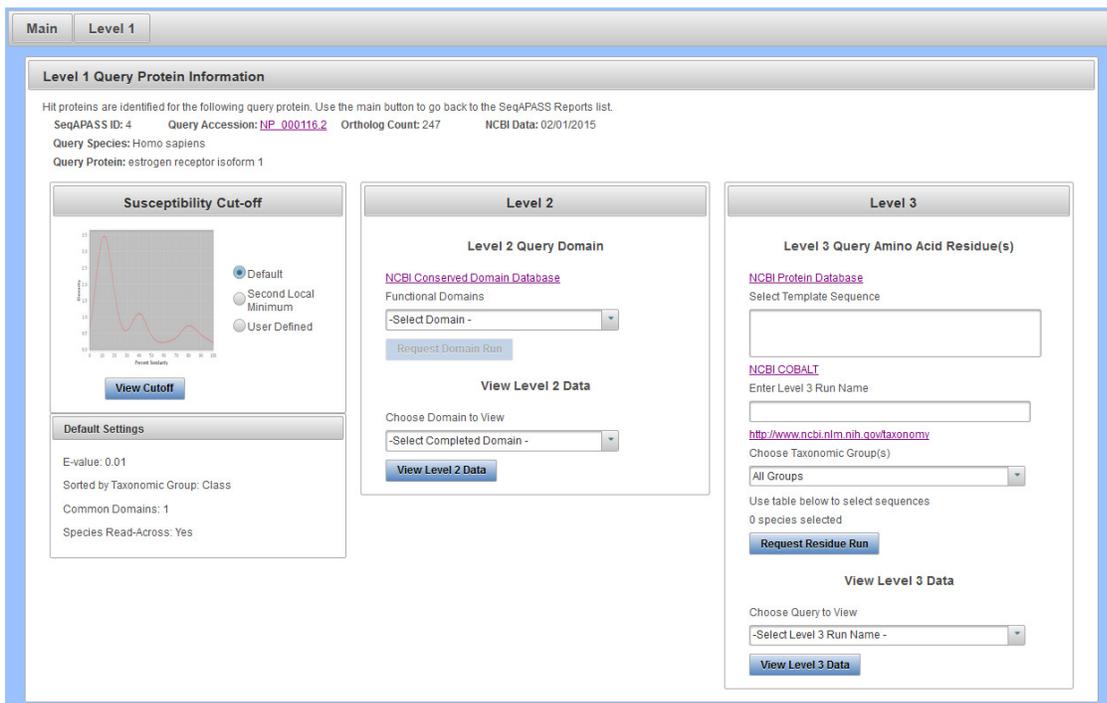
Select Template Sequence

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

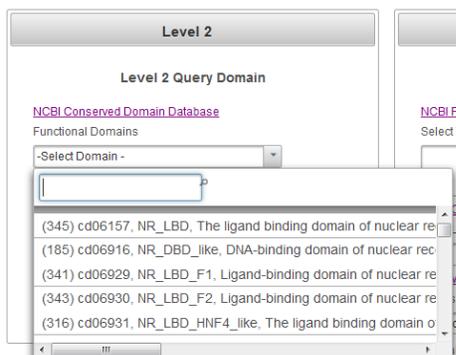
Level 2: Functional Domain(s) Alignment

In the “View SeqAPASS Reports” tab, on the “Level 1 Query Protein Information” page, there is a “Level 2” box for comparing hit domains to the query domain. In the “Level 2” box, there is a link out to the “NCBI Conserved Domain Database” for the query protein of interest. Below this link the user will find a drop-down containing functional domains associated with the query sequence for comparison across species.



In the drop-down box (below the words “Functional Domains”) the user will find all domains associated with the query protein listed in the NCBI Conserved Domains Database. To compare a domain from the query protein to domains of the hit proteins, the user will use the drop-down to highlight a domain and click the “Request Domain Run” button.

Note: Domains in the drop-down are listed with the first amino acid residue position that aligns with the NCBI curated domain in parenthesis, followed by the NCBI domain Accession, domain name, and description.



Note: The user can also use the text box on the top of the drop-down to search the “Functional Domain” list in the drop-down.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

It is recommended that the user click on the “NCBI Conserved Domains Database” link to identify which domains are “Specific hits” in the NCBI Conserved Domains Database. On the NCBI page, the user can scroll over the graphical representation of the domains associated with the query sequence to highlight and identify the Accession associated with domain “Specific hits.” The example below shows the user mousing over the NR_LBD_ER domain.

The screenshot shows the NCBI Conserved Domains Database interface. At the top, there is a navigation bar with links for HOME, SEARCH, GUIDE, NewSearch, Structure Home, 3D Macromolecular Structures, Conserved Domains, Pubchem, and BioSystems. The main content area is titled "Conserved domains on [gi|62821794|ref|NP_000116.2]" and shows the "estrogen receptor isoform 1 [Homo sapiens]". A graphical summary shows the query sequence with various domains highlighted. A list of domain hits is shown, with the NR_LBD_ER domain selected. A detailed view of the NR_LBD_ER domain is shown, including its description and a sequence alignment with the query sequence.

List of domain hits

Name	Accession	Description
NR_LBD_ER	cd06949	Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen); The ligand binding domain (LBD) of Estrogen receptor (ER); Estrogen receptor, a member of nuclear receptor superfamily, is activated by the hormone estrogen. Estrogen regulates many physiological processes including reproduction, bone integrity, cardiovascular health, and behavior. The main mechanism of action of the estrogen receptor element of target genes upon activation by estrogen and then recruiting coactivator proteins which are responsible for the transcriptional activation of target genes. ER has a central well conserved DNA binding domain (DBD), a variant binding domain (LBD). The C-terminal LBD also contains AF-2 activation motif, the dimerization motif, and part of the nuclear localization region. Estrogen receptor has been linked to aging, cancer, obesity and other diseases.

Sequence Alignment:

Pssm-ID: 132747 Cd Length: 235 Bit Score: 426.07 E-value: 1.46e-146

```

gi 62821794 310 LTADQMSALLDAEPPILYSEYDPTTRPFSEASMGLLTNLADRELVRHMINKAKRVPGFVDLTLHDQVHLLSAAWLEILMI 389
Cdd:cd06949 1 LSAEQLSALLEAEPPHYSEYDPTTRPFTEASIMMLLTNLADRELVRHMINKAKRTPGFVDLTLHDQVHLLSAAWLEILMI 80

gi 62821794 390 GLVWRSMHPGKLLFAPNLLDRNQSKVEGMVEIFDMLLATSSRFMMNLQGEFVCLKSIILLNSGVITFPLSILKSL 469
Cdd:cd06949 81 GLVWRSMHPGKLLFAPDILLDRNQSKVEGMVEIFDMLLATSSRFRELQREEVCLKSIILLNSGVITFPLSILKSL 157

gi 62821794 470 EEKDHIRVLDKIIDTLIHLMAKGLTLQQQHQLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLVDLLEMLDAH 547

```

After identifying the domain(s) of interest and the corresponding starting residue and domain Accession, the user can return to the SeqAPASS tool, scroll to the domain of interest in the drop-down. If that domain has not been previously run by the user the “Request Domain Run” button will become active and the user can click it to submit the domain query.

Level 2

Level 2 Query Domain

[NCBI Conserved Domain Database](#)

Functional Domains

(310) cd06949, NR_LBD_ER, Ligand bindir

View Level 2 Data

Choose Domain to View

-Select Completed Domain-

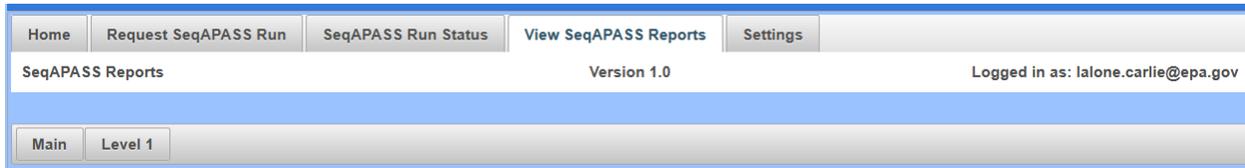
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

When user clicks the “Request Domain Run” button, the following message will appear if the runs has been submitted successfully.

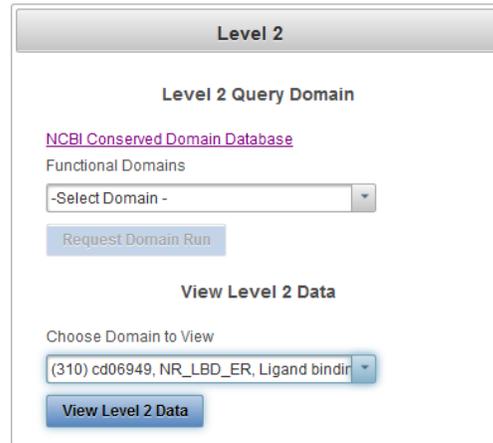
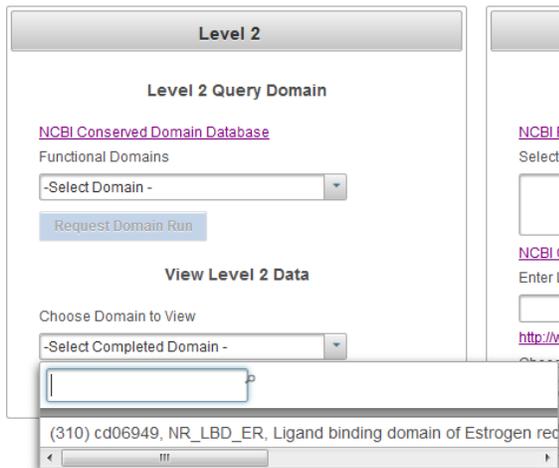


When sequence comparisons have completed for the selected functional domain, the domain will be present in the drop-down in the View Level 2 Data area. The **drop-down is not automatically populated** with the completed domain run. The **user must click on the “Level 1” button to update the page** for the newly completed domain to present itself in the Choose Domain to View drop-down.



To view a completed Level 2 Domain, highlight the domain of interest in the drop-down box and click the “View Domains” button. This will bring the user to the “Level 2” data page for the selected query protein/domain.

Note: The user can also use the text box on the top of the drop-down to search the “Completed Domain” list.



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

View Level 2 Data Page

The “Level 2 Query Domain Information” box contains the SeqAPASS Run ID, Query Accession, Ortholog Count (# of hits identified as ortholog candidates to the query species protein sequence), NCBI Data (displays the date that NCBI databases were downloaded and incorporated into the SeqAPASS database), Query Species, Query Domain (with link out to NCBI domain page), Query Protein name.

The screenshot displays the SeqAPASS web interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. Below these, the page title is "SeqAPASS Reports" and the version is "Version 1.0". The user is logged in as "lalone.carlie@epa.gov".

The main content area has tabs for "Main", "Level 1", and "Level 2". The "Level 2 Query Domain Information" section is active. It contains the following information:

- Hit domains are identified for the following query domain. Use the main button to go back to the SeqAPASS Reports list.
- SeqAPASS ID: 4
- Query Accession: [NP_000116.2](#)
- Ortholog Count: 247
- NCBI Data: 02/01/2015
- Query Species: Homo sapiens
- Query Domain: (310) [cd06949](#), NR_LBD_ER, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)
- Query Protein: estrogen receptor isoform 1

Below this information is a "Susceptibility Cut-off" section. It features a line graph showing Percent Similarity on the y-axis (0.0 to 2.0) and Percent Similarity on the x-axis (0 to 100). The graph shows a peak at approximately 10% similarity. To the right of the graph are three radio button options: "Default" (selected), "Second Local Minimum", and "User Defined". Below the graph is a "View Cutoff" button.

To the right of the graph is a "Default Settings" section with the following options:

- E-value: All
- Sorted by Taxonomic Group: Class
- Species Read-Across: Yes

The default “Level 2” table is the “Primary Report”, which includes query domain information in the first row below the column titles, followed by hit domains whose sequences aligned with the selected query domain. The hit domains are ordered from the highest to lowest percent similarity (Maximum percent similarity =100%). For each hit domain, species information is provided, including the “Protein Count” which indicates the number of protein records per species in the NCBI protein database, taxonomic information, and species names. Also included are the NCBI accession for the query protein, query protein name, Domain Type, BLASTP bitscore (describes overall quality of the alignment, See NCBI BLASTp tutorials), and Domain percent similarity ($[\text{hit bitscore}/\text{query bitscore}] * 100$). If the hit protein has been identified as an ortholog candidate (using reciprocal best hit BLAST method), it will be noted with a “Y” for yes or if not an ortholog candidate, a “N”, for no. A prediction of susceptibility is displayed based on the susceptibility cut-off, identified with a “Y” for yes or an “N” for no. The date/time the analysis was completed is also identified. (See “Search, View, and Download Data Tables” section of user guide for more information).

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Default highlights identify partial protein sequences, sequences with a bitscore higher than the query domain and therefore percent similarity greater than 100% (commonly synthetic constructs), and when zero ortholog candidates are identified (in this case a user should consider a different query sequence).

Partial Hit Protein Sequence
 Percent Similarity > 100%
 Susceptible = Y, Ortholog Count = 0

Level 2 Data - Primary

Search:

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name
NP_000116.2	894432	9606	Mammalia	Homo sapiens	human	estrogen receptor
BAL03259.1	2	899029	Moss transformation vector pPGX6	Moss transformation vector pPGX6	other sequences	XVE fusion p
AA42995.1	80175	32630	synthetic construct	synthetic construct	primer	estrogen rec
XP_003255939.1	30842	61853	Mammalia	Nomascus leucogenys	northern white-cheeked gibbon	PREDICTED: estrogen r
XP_008005788.1	62077	60711	Mammalia	Chlorocebus sabaeus	green monkey	PREDICTED: estrogen r
XP_005552208.1	89797	9541	Mammalia	Macaca fascicularis	crab-eating macaque	PREDICTED: estrogen r
XP_002817538.1	44232	9601	Mammalia	Pongo abelii	Sumatran orangutan	PREDICTED: estrogen r
NP_001158059.1	55290	9555	Mammalia	Papio anubis	olive baboon	estrogen rec
XP_003811544.1	47189	9597	Mammalia	Pan paniscus	pygmy chimpanzee	PREDICTED: estrogen r
XP_011922084.1	633	9531	Mammalia	Cerocebus atys	sooty mangabey	PREDICTED: estrogen r

(1 of 75) [Download Table](#)

[Top of Page](#)

The user can select the “Full Report” on the “Level 2” data page, which includes the same information as the “Primary Report” and additional information pertaining to the alignment of the protein sequence using BLASTp and domain information. Additional information includes the NCBI PSSM ID, NCBI Domain ID, Domain Name, number of amino acid residues in the sequence (Hit Length), the number of exact matching amino acids between the hit and query sequence (Identity), the number of exact and similar (similar side-chain substitutions) matches in amino acids between the hit and the query sequence (Positives), and the expect value (Evalue) describing the number of different alignments expected to occur in the database search by chance. (See “Search, View, and Download Data Tables” section of user guide for more information).

Primary Report
 Full Report
 Partial Hit Protein Sequence
 Percent Similarity > 100%
 Susceptible = Y, Ortholog Count = 0

Level 2 Data - Full

Search:

	NCBI PSSM ID	NCBI Domain Id	Domain Name	Hit Length	Identity	Positive	Evalue	BLASTp Bitscore	Ortholog Candidate	Cut-off	Percent Similarity	Susceptibility Prediction
orm_1	132747	cd06949	NR_LBD_ER	238	238	238	1.078E-177	487.3	Y	42.8	100.0	
n	132747	cd06949	NR_LBD_ER	238	238	238	1.078E-177	487.3	N	42.8	100.0	
_1	132747	cd06949	NR_LBD_ER	238	238	238	1.078E-177	487.3	Y	42.8	100.0	
tor isoform X3	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
tor isoform X2	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
tor isoform X1	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
tor isoform X2	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
if	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
tor isoform X2	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	Y	42.8	99.6	
tor isoform X1	132747	cd06949	NR_LBD_ER	238	237	238	6.557E-177	485.3	N	42.8	99.6	

(1 of 75) [Download Table](#)

[Top of Page](#)

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Susceptibility Cutoff Box for Level 2

The susceptibility prediction is set by identifying ortholog candidates, sequences above a defined susceptibility cutoff, or by identifying those species below the susceptibility cut-off from an organism class above the susceptibility cutoff. The default susceptibility cut-off is set by plotting the distribution of percent similarities calculated for each hit protein. From this plot, the critical points are identified and the local minimums and maximums reported. Using the ortholog candidate data, a susceptibility cut-off is automatically determined by identifying the first ortholog candidate at an equal or higher percent similarity than the first local minimum percent similarity. The user can view this graph by clicking the “View Cutoff” button in the “Susceptibility Cut-off” box. Radio buttons located to the right of the graphical display indicate which Cut-off has been applied for the evaluation of susceptibility in the report. These radio buttons can be selected to change the cut-off in the table to the 2nd local minimum, where the 2nd local minimum is identified in the density plot and the first ortholog candidate at an equal or higher percent similarity than the second local minimum percent similarity is used to set the cut-off. Or the user can define the local minimum by clicking on the “User Defined” radio button. Alternatively, the user can view the closely examine the density plot and manipulate the cut-off by clicking the “View Cutoff” button.

Level 2 Query Domain Information

Hit domains are identified for the following query domain. Use the main button to go back to the SeqAPASS Reports list.

SeqAPASS ID: 4 Query Accession: [NP_000116.2](#) Ortholog Count: 247 NCBI Data: 02/01/2015

Query Species: Homo sapiens

Query Domain: (310) [cd06949](#), NR_LBD_ER,, Ligand binding domain of Estrogen receptor, which are activated by the hormone 17beta-estradiol (estrogen)

Query Protein: estrogen receptor isoform 1

Susceptibility Cut-off

Default
 Second Local Minimum
 User Defined

[View Cutoff](#)

Upon clicking “View Cutoff” button, a new page is displayed with a drop-down that allows the user to set the susceptibility cut-off using the first local minimum and the identified ortholog candidate, the second local minimum and the identified ortholog candidate, or by the “User defined cut-off” (where the user selects the cutoff). To update the cut-off in the Level 1 data report and/or return to the Level 1 page, click “Update Cut-off” button.

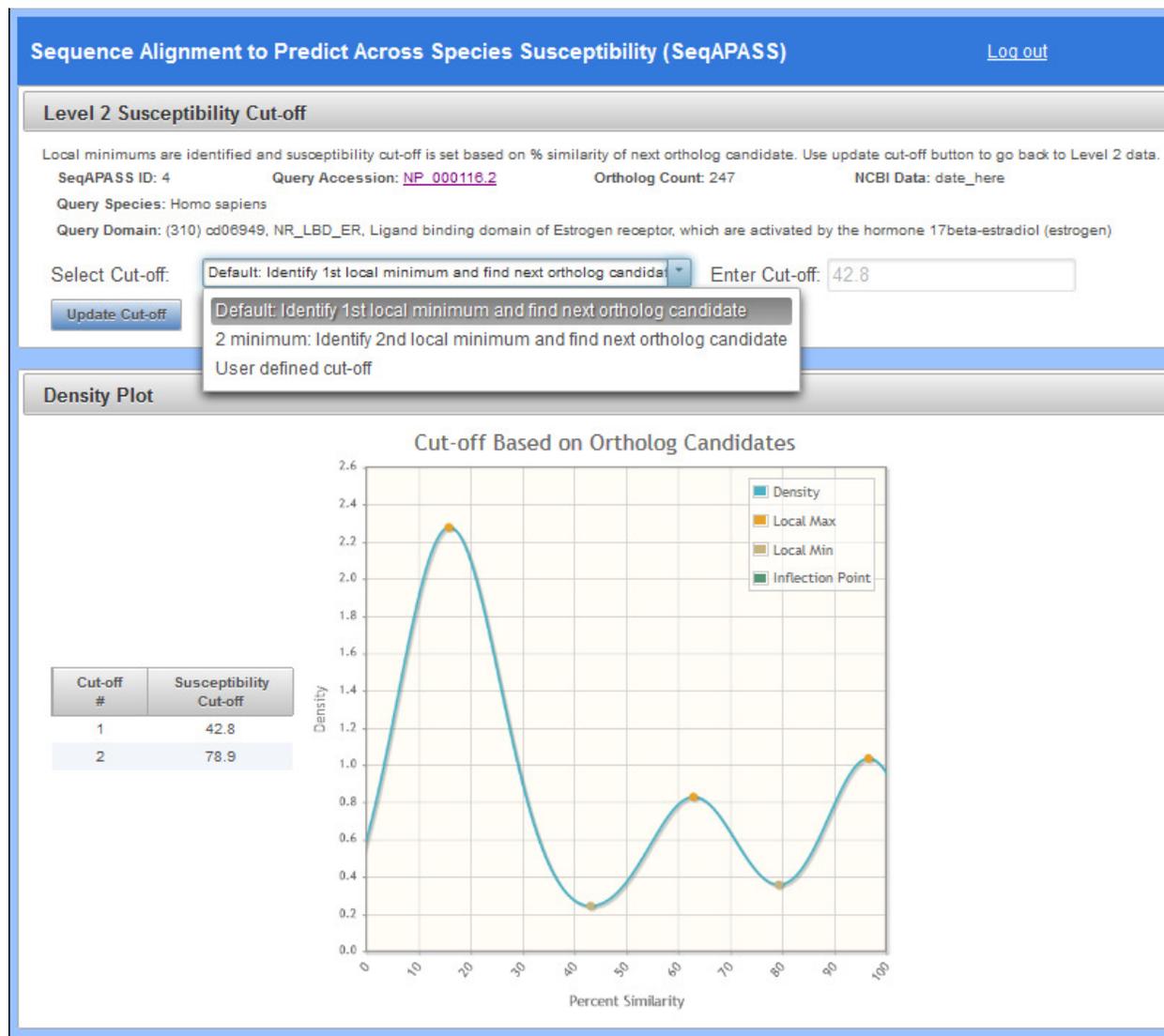
Note: The user should have direct empirical evidence that species above the user defined cutoff are susceptible via the protein of interest, or that the species below the user defined cutoff are not susceptible.

Upon selecting the User defined cut-off from the dropdown, the Enter Cut-off text box becomes active and the user can enter a number 1-100.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Note: In the case that 0 orthologs are identified, the cutoff will be set by the local minimums only, therefore the susceptibility prediction will NOT take into account ortholog candidates. **It is recommended that the user identify a different query sequence for the susceptibility predictions.** Here, the susceptibility predictions will be highlighted in dark pink to indicate that 0 orthologs were detected and the susceptibility cutoff was determined from plotting the distribution of percent similarities and identifying the local minimum



All potential susceptibility cut-offs generated by the data distribution and ortholog candidate identification are reported in the table with columns “Cut-off #” and “Susceptibility Cut-off”. The user can use these numbers to define a cut-off if empirical evidence suggests that the “Default” or “2 minimum” are not supported.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Note: In the case that 0 orthologs are identified from the hit data, the cutoff will be set by the local minimums only, therefore the susceptibility prediction will NOT take into account ortholog candidates. It is recommended that the user identify a different query sequence for the susceptibility predictions. Here, the susceptibility predictions will be highlighted in dark pink in the Level 1 data table to indicate that 0 orthologs were detected and the susceptibility cutoff was determined from plotting the distribution of percent similarities and identifying the local minimums.

The screenshot displays the SeqAPASS web interface. At the top, there is a navigation bar with buttons for Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. Below this is a header section with 'SeqAPASS Reports', 'Version 1.0', and 'Logged in as: lalone.carlie@epa.gov'. The main content area is divided into tabs for Main, Level 1, and Level 2. The Level 2 tab is active, showing 'Level 2 Query Domain Information'. This section includes details about the query domain (SeqAPASS ID: 19, Query Accession: CAA74340.1, Ortholog Count: 0, NCBI Data: 02/01/2015) and the query protein (insulin receptor). A 'Susceptibility Cut-off' plot is shown, with a 'View Cutoff' button. Below the plot are 'Default Settings' for E-value, sorting, and species read-across. A filter section allows users to select report types (Primary, Full) and filter criteria (Partial Hit Protein Sequence, Percent Similarity > 100%, Susceptible = Y, Ortholog Count = 0). The 'Level 2 Data - Primary' table is displayed below, showing a search bar and a table of results. The table has columns for NCBI Accession, Protein Count, Species Tax. ID, Taxonomic Group, Scientific Name, Common Name, Protein Name, and Domain Name. The first three rows are highlighted in dark pink, indicating 0 orthologs detected.

NCBI Accession	Protein Count	Species Tax. ID	Taxonomic Group	Scientific Name	Common Name	Protein Name	Domain Name
CAA74340.1	44215	89482	Mammalia	Bubalus bubalis	water buffalo	insulin receptor	PTKc_InsR
XP_010826805.1	35541	43346	Mammalia	Bison bison bison	even-toed ungulates	PREDICTED_insulin_receptor_isoform X1	PTKc_InsR
XP_008742383.1	25818	9713	Mammalia	Leptonychotes weddellii	Weddell seal	PREDICTED_insulin_receptor	PTKc_InsR

No Orthologs Detected

If no orthologs are detected from reciprocal best hit blast analysis “Ortholog Count” will be “0” at the top of the “Level 1 Query Protein Information” page. Further, dark pink will highlight the entire query/hits table indicating that the “Susceptibility Prediction” columns were determined from the local minimums identified in the “View Cutoff” density plot, without consideration of orthologs.

This screenshot shows the 'Level 2 Query Domain Information' section of the SeqAPASS interface. It provides the same query details as the previous screenshot: SeqAPASS ID: 19, Query Accession: CAA74340.1, Ortholog Count: 0, NCBI Data: 02/01/2015, Query Species: Bubalus bubalis, Query Domain: (1) cd05061, PTKc_InsR, Catalytic domain of the Protein Tyrosine Kinase, Insulin Receptor, and Query Protein: insulin receptor.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

By clicking on the “View Cutoff” button when no orthologs are detected, the “Cut-off #” and “Susceptibility Cut-off” columns will report only the local minimum values.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

Level 2 Susceptibility Cut-off

Local minimums are identified and susceptibility cut-off is set based on % similarity of next ortholog candidate. Use update cut-off button to go back to Level 2 data.

SeqAPASS ID: 19 Query Accession: [CAA74340.1](#) Ortholog Count: 0 NCBI Data: date_here

Query Species: Bubalus bubalis

Query Domain: (1) cd05061, PTKc_InsR, Catalytic domain of the Protein Tyrosine Kinase, Insulin Receptor

Select Cut-off: Enter Cut-off:

Density Plot

Cut-off Based on Ortholog Candidates

Cut-off #	Susceptibility Cut-off
1	64.8
2	92.7

The user can return to the “Level 2” data page by clicking the “Update Cut-off” button.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 3: Individual Amino Acid Residue Alignment

In the “View SeqAPASS Reports” tab, on the “Level 1 Query Protein Information” page, there is a “Level 3” box for setting up the query for comparing individual amino acid residues to a template sequence. It is anticipated that the template sequence and residues that selected to align will be derived from the published literature in most cases. Publications evaluating homology models, pesticide field resistance, or utilizing site-directed mutagenesis are a few examples of the types of studies that may contain such information to guide a Level 3 SeqAPASS evaluation.

The screenshot displays the SeqAPASS web application interface. At the top, there is a navigation bar with links for Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. Below this is a sub-navigation bar with Main, Level 1, and Level 2 tabs. The main content area is titled "Level 1 Query Protein Information" and contains the following sections:

- Susceptibility Cut-off:** A line graph showing a peak at position 10. Below the graph are radio buttons for "Default" (selected), "Second Local Minimum", and "User Defined". A "View Cutoff" button is present.
- Level 2:** A section for "Level 2 Query Domain" with a link to "NCBI Conserved Domain Database". It includes a dropdown menu for "Functional Domains" (currently showing "-Select Domain -") and a "Request Domain Run" button. Below this is a "View Level 2 Data" section with a "Choose Domain to View" dropdown (showing "-Select Completed Domain -") and a "View Level 2 Data" button.
- Level 3:** A section for "Level 3 Query Amino Acid Residue(s)" with a link to "NCBI Protein Database" and a text box for "Select Template Sequence". Below this is a link to "NCBI COBALT" and a text box for "Enter Level 3 Run Name". A link to "http://www.ncbi.nlm.nih.gov/taxonomy" is provided for "Choose Taxonomic Group(s)", with a dropdown menu set to "All Groups". A note says "Use table below to select sequences" and "0 species selected". A "Request Residue Run" button is present. Below this is a "View Level 3 Data" section with a "Choose Query to View" dropdown (showing "-Select Level 3 Run Name -") and a "View Level 3 Data" button.

In the “Level 3” box, there is a link out to the “NCBI Protein Database” for identifying the template sequence of interest. Below this link the user will find a text box where the user can enter an NCBI Protein Accession with the version number (e.g., NP_000116.2) or a FASTA formatted sequence (e.g., < >gi|62821794|ref|NP_000116.2| estrogen receptor isoform 1 [Homo sapiens]
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAAYEFNA
AAAANA
QVYGTGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPFLQPHGQQVPYYLENE
PSGYT
VREAGPPAFYRPNSDNRQGRERLASTNDKGSMAKESAKETRYCAVCNDYASGYHYGVWSC
EGCKAFFK
RSIQGHNDYMCNATNQCTIDKNRRKSCQACRLRKYEVGMMKGGIRKDRRGGRMLKHKRQRD
DGEGRGEV
GSAGDMRAANLWPSPLMIKRSKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMG
LLTNLA
DRELVHMINWAKRVPGFVDLTLHDQV).

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon clicking on in the “Select Template Sequence” text box, a pop-up message will appear to provide examples for the proper format of Accessions or FASTA files to be entered.

The screenshot displays two main panels: "Level 2" and "Level 3".

Level 2 Panel: Titled "Level 2 Query Domain", it features a link to "NCBI Conserved Domain Database" and a section for "Functional Domains". A dropdown menu shows "(345) cd06157, NR_LBD, The ligand bindi". A "Request Domain Run" button is present, along with a "View Level 2 Data" button.

Level 3 Panel: Titled "Level 3 Query Amino Acid Residue(s)", it includes a link to "NCBI Protein Database" and a "Select Template Sequence" text box. Below this is a link to "NCBI COBALT" and an "Enter Level 3 Run Name" text box. A link to "http://www.ncbi.nlm.nih.gov/taxonomy" is provided for "Choose Taxonomic Group(s)", with a dropdown menu set to "All Groups". A note says "Use table below to select sequences" and "0 species selected". A "Request Residue Run" button is located below. At the bottom, there is a "View Level 3 Data" section with a "Choose Query to View" dropdown menu set to "-Select Level 3 Run Name -" and a "View Level 3 Data" button.

Pop-up Window: A white box with a grey border contains the text: "-Enter NCBI Protein Accession OR FASTA Sequence- Examples: NP_000116.2 OR >Sequence description in first line MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVY".

Below the text box where the user selects the template sequence, is a link to NCBI COBALT (Constraint-based Multiple Protein Alignment Tool). The NCBI COBALT allows the user to align multiple sequences and is the alignment tool that SeqAPASS utilizes to set up the query of individual amino acid residues across species.

Note: The user does not need to use the COBALT link to run a Level 3 evaluation, however the link is available in case the user chooses to further evaluate or compare multiple potential template sequences.

Under the text “Enter Level 3 Run Name,” there is a text box where the user can enter a user defined name for the run. The user may only enter letters or integers as text for the name. The user defined name will appear in the “View Level 3 Data” dropdown upon completion of the Level 3 sequence alignment.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

To complete the set-up for a Level 3 query the user must select which sequences to compare to the identified template sequence. Listed in the Choose Taxonomic Group(s) drop-down are all Taxonomic Groups that were identified as hits in the “Level 1” primary amino acid sequence alignment data. Because COBALT is used to align all sequences that are selected, it is recommended that the user selectively identify sequences from the hit table below to align. For example, selecting sequences with low similarity to the template sequence along with sequences sharing high similarity to the template sequence can skew the alignment because COBALT is trying to align all of the sequences together. It is recommended that the user select sequences by first selecting a taxonomic group from the “Choose Taxonomic Group(s) drop-down. The user can also use the NCBI taxonomy link to type in the name of the “Taxonomic Groups” found in the drop-down to look up which species fall in that group.

The screenshot shows the 'Level 3' query interface. At the top, it says 'Level 3' in a grey box. Below that is the title 'Level 3 Query Amino Acid Residue(s)'. There are three main sections: 1) 'NCBI Protein Database' with a 'Select Template Sequence' field containing 'NP_000116.2'. 2) 'NCBI COBALT' with an 'Enter Level 3 Run Name' field containing 'Bony fish'. 3) 'Choose Taxonomic Group(s)' with a dropdown menu currently set to 'All Groups'. This dropdown menu is open, showing a list of taxonomic groups: 'All Groups', 'Actinopteri' (highlighted in blue), 'Amphibia', 'Anthozoa', 'Appendicularia', 'Arachnida', and 'Ascidacea'. At the bottom of the dropdown is a 'View Level 3 Data' button.

By choosing a group from the drop-down menu, the “Level 1” table below will be filtered by the selected Taxonomic Group (see column “Taxonomic Group” in Level 1 data table). When a “Taxonomic group is selected from the drop-down, it can take up to a few seconds for the Level 1 data table to filter completely, depending on the size of the table. The user can then examine each hit protein in the Level 1 table and select those that they would like to compare to the template sequence. To select sequences/species from the filtered Level 1 data table, the user will select the check boxes in the first column of the table. Although it is not typically recommended, the user may also select the header check box in the first column to select all sequences/species in the filtered table.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Note: The user can also type the “Taxonomic Group” of interest in the text search box at the top of the drop-down for quick filtering.

Below is an example where the user selected the “Taxonomic Group” Actinopteri from the drop-down and then selected individual sequences/species to align with the template sequence. The number of selected species will be shown in the text above the “Request Residue Run” button.

Susceptibility Cut-off

Level 2 Query Domain

Level 3 Query Amino Acid Residue(s)

Level 1 Data - Primary

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name	BLAST Bitscore
<input checked="" type="checkbox"/> RAQ82853.1	109	512342	Actinopteri	<i>Atractosteus tropicus</i>	tropical gar	estrogen receptor alpha	648.0
<input checked="" type="checkbox"/> XP_005625908.1	18441	7818	Actinopteri	<i>Lepisosteus oculatus</i>	spotted gar	PREDICTED: estrogen receptor-like	641.3
<input checked="" type="checkbox"/> RAQ82850.1	139	111304	Actinopteri	<i>Acipenser schrenkii</i>	Amur sturgeon	estrogen receptor alpha1	595.0
<input checked="" type="checkbox"/> MKF16865.1	28959	215358	Actinopteri	<i>Larimichthys crocea</i>	large yellow croaker	Estrogen receptor	537.7

(See “Search, View, and Download Data Tables” section of user guide for more information)

The user can choose to align sequences/species from multiple taxonomic groups with the template sequence, by going back to the “Choose Taxonomic Group” drop-down and selecting another group, which filters the Level 1 table based on the group selected, and then the user can select additional species from the newly filtered table. As before, the number of selected species can be tracked in the text above the “Request Residue Run” button that reads “X species selected.”

When the user has selected all sequences they would like to align, then click the “Request Residue Run” button. Upon successful submission of a Level 3 query the user will see the following pop-up message. If submission is unsuccessful, a message will appear describing the reason for the unsuccessful submission.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) [Log out](#)

Home Request SeqAPASS Run SeqAPASS Run Status View SeqAPASS Reports Settings

SeqAPASS Reports Version 1.0 Logged in as: lalone.carlie@epa.gov

Level 3 Run Requested Status queued

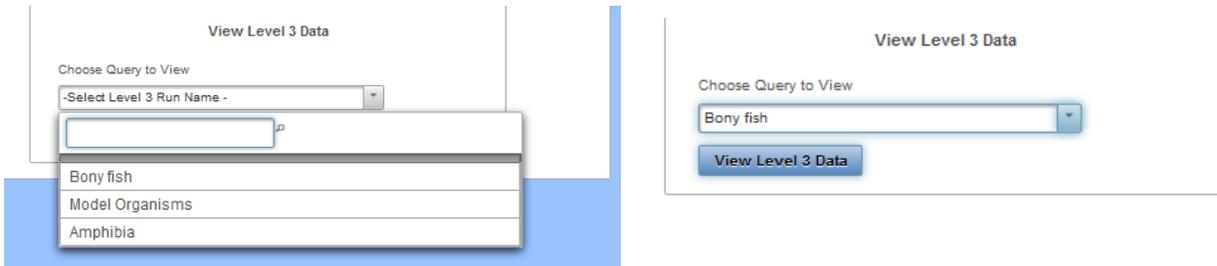
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

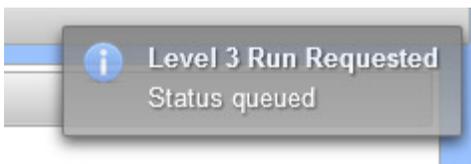
To update the View Level 3 Data, “Choose Query to View” drop-down menu with the completed Level 3 alignments, the user can click on the top left “Level 1” button.



Additionally, the user can check the status of the Level 3 run by clicking the “SeqAPASS Run Status” tab and the radio button for “Level 3 Status.” Typically Level 3 alignments complete in a few seconds. When the Level 3 query completes and the Level 1 page has been updated, the user defined Level 3 Run Name will be available in the “Choose Query to View” drop-down menu. After selecting the desired Run Name from the drop-down, click “View Level 3 Data” button to view the aligned sequences and set up the individual amino acid residue alignments with the selected sequences/species.



Upon a successful Level 3 query submission a pop-up message will be displayed as follows in the upper right-hand side of the screen:



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

View Level 3 Individual Amino Acid Query and Data Page

Clicking the “View Level 3 Data” button, the Level 3 data page opens. The “Level 3 Template Protein Information” box contains the SeqAPASS Run ID, Query Accession (with link out to NCBI), Ortholog Count (# of hits identified as ortholog candidates to the query species protein sequence), NCBI Data (displays the date that NCBI databases were downloaded and incorporated into the SeqAPASS database), Level 3 Run Name (defined by user), Template Species (Entered by user in Level 3 query), Template Protein, and Query Residues (this field is populated with residues upon selection and successful table update).

The screenshot shows the SeqAPASS web interface. At the top, there are navigation tabs: Home, Request SeqAPASS Run, SeqAPASS Run Status, View SeqAPASS Reports, and Settings. The current page is 'View SeqAPASS Reports', with the user logged in as 'lalone.carlie@epa.gov'. Below the navigation is a sub-menu with 'Main', 'Level 1', and 'Level 3' tabs. The 'Level 3' tab is selected, leading to the 'Level 3 Template Protein Information' section. This section contains the following information: SeqAPASS ID: 4, Query Accession: [NP_000116.2](#), Ortholog Count: 247, NCBI Data: 02/01/2015, Level 3 Run Name: Bony fish, Template Species: Homo sapiens, Template Protein: estrogen receptor isoform 1, and Query Residues: No Residues Selected. Below this is a 'Select Amino Acid Residues' section with a list of amino acid single-letter abbreviations (1M, 2T, 3M, 4T, 5L, 6H, 7T, 8K) and an 'Update Report' button. The 'Level 3 Data' section below features a search bar and a table of results. The table has columns for NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Scientific Name, Common Name, Protein Name, and Analysis Date. The table is sorted by percent similarity, with the highest similarity at the top.

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Scientific Name	Common Name	Protein Name	Analysis Date
NP_000116.2	894432	9606	Mammalia	Homo sapiens	human	estrogen receptor isoform 1	2016 03 1
BAG82653.1	109	512342	Actinopteri	Atractosteus tropicus	tropical gar	estrogen receptor alpha	2016 03 1
BAG82650.1	139	111304	Actinopteri	Acipenser schrenckii	Amur sturgeon	estrogen receptor alpha1	2016 03 1
KKF16605.1	28959	216368	Actinopteri	Larimichthys crocea	large yellow croaker	Estrogen receptor	2016 03 1
BAT68973.1	683	7937	Actinopteri	Anguilla japonica	Japanese eel	estrogen receptor 1	2016 03 1

The user can view the “Level 3” data page, which includes the NCBI Accession, Protein Count, Taxonomic information, Protein Name, and date/time the Level 3 run completed. The data table remains in order of percent similarity, with those sequences having the highest percent similarity to the original Level 1 query sequence, on the top, to those with the lowest percent similarity on the bottom. (See “Search, View, and Download Data Tables” section of user guide for more information).

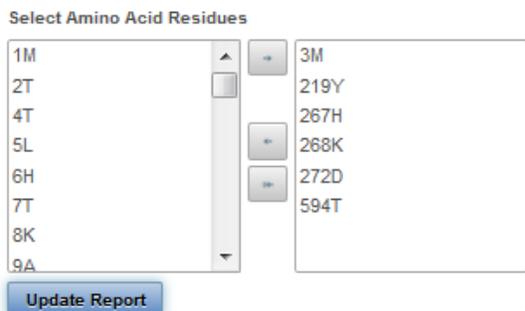
To obtain individual amino acid residue alignment data in the Level 3 data table, the user must use the shuttle in the “Level 3 Template Protein Information box to select positions and amino acid residues from the chosen template sequence to align with the sequences/species that were selected by taxonomic group. Single letter abbreviations are used for the amino acid sequences.

- G: Glycine A: Alanine S: Serine T: Threonine C: Cysteine V: Valine
- L: Leucine I: Isoleucine M: Methionine P: Proline F: Phenylalanine
- Y: Tyrosine W: Tryptophan D: Aspartic Acid E: Glutamic Acid
- N: Asparagine Q: Glutamine H: Histidine K: Lysine R: Arginine

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

The user can select one residue at a time by clicking and highlighting the residue of interest and then clicking the top right arrow shuttle button to move the residue to the right-hand box for inclusion in the alignment. Each time a residue is added to the right-hand box, the left-hand box resets itself to the 1st residue. Or the user can select multiple residues at the same time by holding the Ctrl button, clicking on residues, and then clicking the top right arrow shuttle button to move the residues to the right-hand box. The user can choose to remove selected residues by using the left arrow button to clear one at a time or the double left arrow button to remove all selected residues at once. When residues of interest (likely defined from the literature as described above) have been selected, click the “Update Report” button, which then updates the Level 3 Data table with the individual residue alignment data.



The individual amino acid residue alignment data will then be updated on the right most columns of the Level 3 Data table. The user **can submit a maximum of 50 individual amino acid residues** from the template sequence to compare to the other selected sequences. The individual amino acid residues will be listed in numerical order starting with the 1st position in the template sequence to the last position in the template sequence. The columns will be titled Position 1, Amino Acid 1, Match 1, Position 2 Amino Acid 2, Match 2, Position 3, Amino Acid 3, Match 3..... The template sequence will always be in the top row of the Level 3 Data table followed by the previously selected sequences. Further, the residues selected in the shuttle will also be displayed in the top row corresponding to the template sequence. Each Position and Amino Acid in the following rows are those corresponding to the Protein Accession identified in that row align with the template sequence. The Match X, describes whether the amino acid residue is an exact match to the template, “Y,” for yes, or not an exact match to the template, “N,” for no. The user can evaluate this data to understand how well conserved an amino acid residue is across species or in a species of interest to add an additional line of evidence to support (or question) susceptibility predictions.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Moving Between Level 1, Level 2, and Level 3 Data Pages

As a user chooses to view Level 1, Level 2, or Level 3 data in the “View SeqAPASS Reports” tab, new buttons become available for allowing the user to move between Levels of an analysis. Please see snapshot below.



The user can use the “Main” button to return to the list of completed Level 1 runs and select a different query accession to view. The “Level 1” button brings the user to the Level 1 data page, where the user can set up queries for Level 2 and Level 3, as well as select the button to view Level 2 and Level 3 data pages. Open Level 1, Level 2, and Level 3 pages remain open until the user selects a different run to view on the “Main” page. Moving between tabs, such as “Home,” Request SeqAPASS Run,” and “SeqAPASS Run Status”, does not close the Level 1, Level 2, or Level 3 pages that have been opened.

Note: If the user logs out of the SeqAPASS tool, upon logging back in, the data will reset to default settings. Therefore, the View SeqAPASS Reports tab will not display the “Main,” “Level 1,” “Level 2,” or “Level 3” buttons, until a query is chosen and Level 2 and Level 3 pages are opened.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

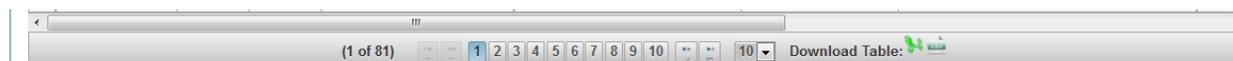
Search, View, and Download Data Tables

The user can use the “Search” box to enter text to search the table. Further, the user can use the arrow buttons and page numbers on the bottom of the screen to view all data and the drop-down to expand the table to 10, 20, or 50 rows. There are also left and right scroll bars at the bottom of the tables to allow the user to view all columns of the table.

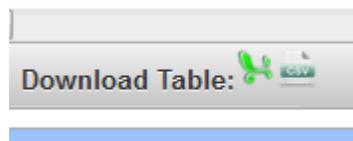
Search using text box on top of tables:



Options for viewing data:



All data tables in the SeqAPASS tool can be downloaded as Excel or csv files. The icons for downloading the files are present on the bottom right-hand side of all tables. Click the icon to download data.



Upon selecting a .xls file, the user can choose to save or open the file. Each file is appropriately named by Level of the SeqAPASS evaluation and report type.

The screenshot shows the SeqAPASS interface with a table of protein data and a file download dialog box. The table has columns for NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, and Species. The dialog box is titled "Opening SeqAPASS_Level1_Primary_Report.xls" and offers options to "Open with Microsoft Excel (default)" or "Save File".

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Species	Other Sequences	Notes
NP_000116.2	894432	9606	Mammalia	Homo sapiens	human	estrogen receptor isoform 1
AA42895.1	80175	32630	synthetic construct	synthetic construct	primer	estrogen receptor 1
AAD47137.1	2	88432	Mammalian expression vector pCI-nC	Mammalian expression vector pCI-nGL1-HE	other sequences	green fluorescent protein-estrogen receptor alpha fusion
XP_003311696.1	63068	9598	Mammalia	Pan troglodytes	chimpanzee	PREDICTED: estrogen receptor isoform X2
XP_003811644.1	47189	9597	Mammalia	Pan paniscus	pygmy chimpanzee	PREDICTED: estrogen receptor isoform X2
XP_002817538.1	44232	9601	Mammalia	Pongo abelii	Sumatan orangutan	PREDICTED: estrogen receptor isoform X2
XP_005552208.1	89797	9541	Mammalia	Macaca fascicularis	crab-eating macaque	PREDICTED: estrogen receptor isoform X1
XP_011799162.1	11	326983	Mammalia	Colobus angolensis palliatus	primates	PREDICTED: estrogen receptor isoform X2
NP_001168059.1	56290	9555	Mammalia	Papio anubis	olive baboon	estrogen receptor
XP_011852190.1	219	9568	Mammalia	Mandrillus leucophaeus	drill	PREDICTED: estrogen receptor isoform X2

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Upon selecting a .csv file, the user can choose to save or open the file. Each file is appropriately named by Level of the SeqAPASS evaluation and report type.

The screenshot displays the SeqAPASS web interface. At the top, a status box indicates "Species Read-Across: Yes". Below this, there are filter options for report types: "Primary Report" (selected) and "Full Report". Filter criteria include "Partial Hit Protein Sequence", "Percent Similarity > 100%", and "Susceptible = Y, Ortholog Count = 0".

A modal dialog box titled "Opening SeqAPASS_Level1_Primary_Report.csv" is open, showing the file name and its location. It asks "What should Firefox do with this file?" with options: "Open with Microsoft Excel (default)", "Save File", and "Do this automatically for files like this from now on." The "Open with" option is selected.

The main content area shows "Level 1 Data - Primary" with a search bar and a table of results. The table has columns for NCBI Accession, Protein Count, Species Tax ID, Taxonomic Group, Species, and a description. The table contains 10 rows of data, including entries for human, synthetic construct, chimpanzee, and various primate species.

NCBI Accession	Protein Count	Species Tax ID	Taxonomic Group	Species	Description
NP_000116.2	894432	9506	Mammalia	Homo sapiens	human estrogen receptor isoform 1
AAU42996.1	80175	32630	synthetic construct	synthetic construct	primer estrogen receptor 1
AAD47137.1	2	88432	Mammalian expression vector pCI-n	Mammalian expression vector pCI-nGL1-HE	other sequences green fluorescent protein-estrogen receptor alpha fusion
XP_003311696.1	63068	9598	Mammalia	Pan troglodytes	chimpanzee PREDICTED: estrogen receptor isoform X2
XP_003811544.1	47189	9597	Mammalia	Pan paniscus	pygmy chimpanzee PREDICTED: estrogen receptor isoform X2
XP_002817538.1	44232	9601	Mammalia	Pongo abelii	Sumatran orangutan PREDICTED: estrogen receptor isoform X2
XP_005552208.1	89797	9541	Mammalia	Macaca fascicularis	crab-eating macaque PREDICTED: estrogen receptor isoform X1
XP_011799152.1	11	336983	Mammalia	Colobus angolensis palliatus	primates PREDICTED: estrogen receptor isoform X2
NP_001158059.1	55290	9555	Mammalia	Papio anubis	olive baboon estrogen receptor
XP_011852190.1	219	9568	Mammalia	Mandrillus leucophaeus	drill PREDICTED: estrogen receptor isoform X2

At the bottom of the table, there is a pagination control showing "(1 of 81)" and a "Download Table" button.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Log out

The user can log out from any page in SeqAPASS, by clicking the “Log out” link on the upper right-hand side of the page. If a user clicks Log out and then Logs back in, all settings will be set back to default. User can log out at any time by clicking the “Log out” link on the upper right-hand side. Any successfully submitted queries that were requested prior to logging out will continue running and when completed, will be available to the user in the “View SeqAPASS Reports” tab.



The screenshot shows the SeqAPASS web application interface. At the top, there is a blue header bar with the text "Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)" on the left and a "Log out" link on the right. Below the header is a navigation menu with five tabs: "Home", "Request SeqAPASS Run", "SeqAPASS Run Status", "View SeqAPASS Reports", and "Settings". The "Home" tab is currently selected. Below the navigation menu, there is a footer area with three pieces of information: "Welcome to SeqAPASS" on the left, "Version 1.0" in the center, and "Logged in as: lalone.carlie@epa.gov" on the right.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

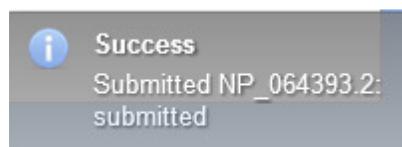
Pop-up Messages

The Spinning Wheel pop-up is used as an indicator to alert the user that an action is taking place, where the interface of the SeqAPASS tool is contacting the backend database. For example upon clicking the “SeqAPASS Run Status” tab, “Refresh Data” button, “View Level 2 Data” button, or “View Level 3 Data” button the Spinning Wheel will pop-up and disappear from the screen. There are multiple other instances where the spinning wheel is used as an indicator to the user that an action is occurring.

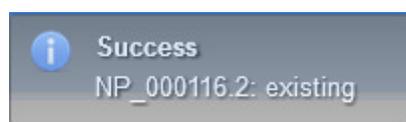


Pop-up messages are meant to guide the user to submit the correct information for a query, inform the user of a successful or failed query submission, or otherwise inform the user of an error. All pop-up messages will appear for 10 seconds on the upper right-hand side of the screen, and then disappear. If the user would like to close the message before the 10 seconds is up, click on the message and an “x” will appear of the upper right hand corner of the message box. Click the x to close the message.

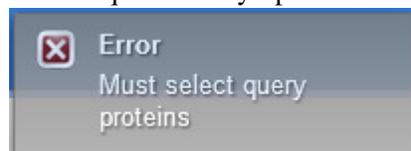
In the “Request SeqAPASS Run” tab, Compare Primary Amino Acid Sequences “By Species” page, a successful Level 1 query submission will display a pop-up message indicating that the query has been submitted to the run queue or if “existing” message appears indicating that the accession has been ran previously either by a user and is available to view.



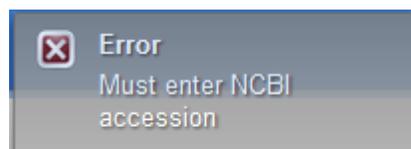
OR



User did not select any query proteins from the “Request SeqAPASS Run” tab, Compare Primary Amino Acid Sequences “By Species” or “By Accession” page, and clicked “Request Run” button.



OR

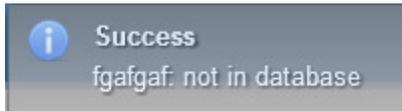


User non-sense text (or any text that is not an NCBI accession) into the CBI Protein Accession” text box for submitting a Level 1 query in the “Request SeqAPASS Run” tab, Compare Primary Amino Acid

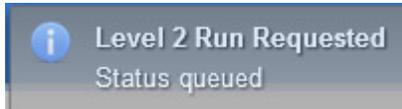
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

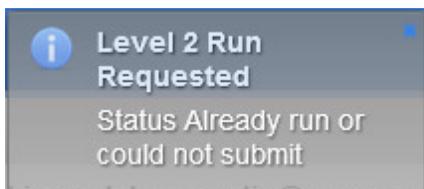
Sequences “By Accession” page, and clicked “Request Run” button. The message below indicates that the Accession entered is not in the SeqAPASS database.



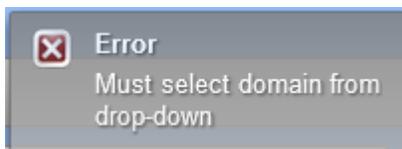
In the “View SeqAPASS Reports” tab, Level 1 page, if a user clicks “View Level 2 Data,” a successful Level 2 query submission will display a pop-up message indicating that the query has entered the run queue.



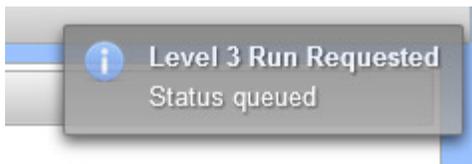
In the “View SeqAPASS Reports” tab, Level 1 page, if a user selects a domain that has already been submitted (but not completed) and clicks “Request Domain Run” a successful Level 2 query submission will display a pop-up message indicating that the query has entered the run queue



In the “View SeqAPASS Reports” tab, Level 1 page, if a user clicks “View Level 2 Data” without selecting a domain to view from the drop-down, the message below will pop-up to indicate that the user must select a domain.



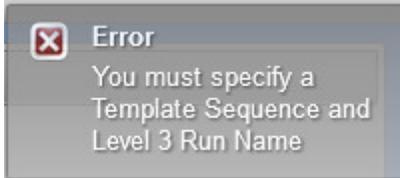
In the “View SeqAPASS Reports” tab, Level 1 page, a successful Level 3 query submission will display a pop-up message indicating that the query has entered the run queue.



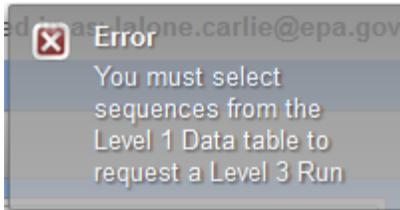
In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to type a user defined Level 3 Run Name, the message below will pop-up to indicate that the user must do so.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

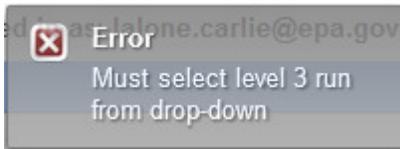
Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov



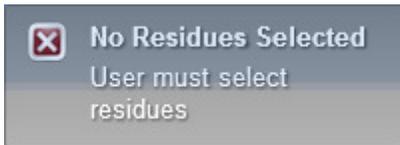
In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to select species from the Level 1 Data table to be compared with the template sequence, the message below will pop-up.



In the “View SeqAPASS Reports” tab, Level 1 page, if a user fails to select a Level 3 Run Name from the Choose Query to View drop-down and clicks the “View Level 3 Date” button, the message below will pop-up.



In the “View SeqAPASS Reports” tab, “Level 3 Template Protein Information” data page, if a user fails to select amino acid residues using the “Select Amino Acid Residues” shuttle and clicks the “View Level 3 Date” button, the message below will pop-up.



Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

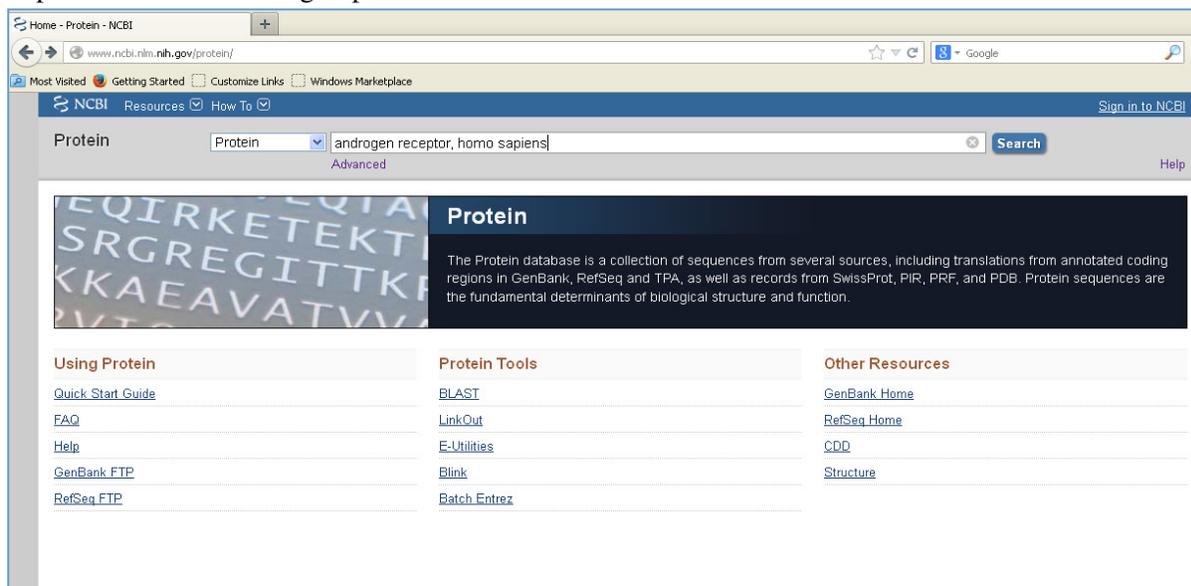
Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS) Documentation

Query Species: The selection of the query species for a SeqAPASS analysis is dependent upon the question the user is addressing. For example, the query species can be the target species (i.e., human or companion animal in the case of drugs; or insect, plant, fungus, or pest in the case of pesticides) or, depending on the application of the susceptibility prediction, the query species may be a species known or hypothesized to be sensitive to a chemical via acting on the protein molecular target of interest. There may be instances where a protein for the species of interest has not been sequenced, in this case it may serve the users purpose to identify another taxonomically related species from the same organism Class, Order, Family, or Genus as a surrogate query species. In certain cases, when there is interest in the susceptibility of a particular species (e.g., honey bee) and in the case that there are numerous potential target species (e.g., neonicotinoids are intended to cause mortality in a number of pest insects) the species of particular concern may serve as the query species.

Query Protein: SeqAPASS can be queried with any protein sequence available in the NCBI protein GenBank database, by protein name, or NCBI Accession. It is suggested that the user of SeqAPASS examines their query protein and species in the NCBI protein database prior to submitting a run to SeqAPASS (use NCBI link on query page). It is not uncommon for a protein of a specific species to be represented by more than one sequence. In such cases there are some guiding principles for identification of the best sequence available for the SeqAPASS run.

- **General guidelines:** These guidelines describe best practices for identifying the most useful sequence for a species susceptibility prediction in SeqAPASS, however, in some cases, limited sequence information is available and therefore less desirable sequences may be used. It is up to the user of SeqAPASS to recognize the quality and limitations of the sequence chosen for the SeqAPASS run. The information about a particular protein can be found on the Protein page in the NCBI database.

<http://www.ncbi.nlm.nih.gov/protein/>



The screenshot shows the NCBI Protein database search results page. The search query is "androgen receptor, homo sapiens". The page displays a search bar with the query, a "Search" button, and a "Help" link. Below the search bar, there is a "Protein" section with a description: "The Protein database is a collection of sequences from several sources, including translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB. Protein sequences are the fundamental determinants of biological structure and function." Below this, there are three columns of links: "Using Protein" (Quick Start Guide, FAQ, Help, GenBank FTP, RefSeq FTP), "Protein Tools" (BLAST, LinkOut, E-Utilities, Blink, Batch Entrez), and "Other Resources" (GenBank Home, RefSeq Home, CDD, Structure).

Search for a protein of interest using protein name and/or species of interest: For the example above, multiple hit proteins were identified.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

NCBI Resources How To Sign in to NCBI

Protein Protein androgen receptor, homo sapiens Search

Save search Advanced Help

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Manage Filters

Results: 1 to 20 of 540

1. [RecName: Full=Androgen receptor; AltName: Full=Dihydrotestosterone receptor; AltName: Full=Nuclear receptor subfamily 3 group C member 4](#)
919 aa protein
Accession: P10275.2 GI: 113830
[GenPept](#) [FASTA](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

2. [androgen receptor \[Homo sapiens\]](#)
917 aa protein
Accession: AAA51772.1 GI: 178882
[GenPept](#) [FASTA](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

3. [androgen receptor, partial \[Homo sapiens\]](#)
2 aa protein
Accession: AAD14959.1 GI: 4262811
[GenPept](#) [FASTA](#) [Graphics](#)

4. [androgen-receptor \[Homo sapiens\]](#)
906 aa protein
Accession: AAA51780.1 GI: 179034
[GenPept](#) [FASTA](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

5. [androgen receptor \[Homo sapiens\]](#)
917 aa protein
Accession: AAA51771.1 GI: 178872
[GenPept](#) [FASTA](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

6. [androgen receptor \[Homo sapiens\]](#)
917 aa protein
Accession: AAA51771.1 GI: 178872
[GenPept](#) [FASTA](#) [Graphics](#) [Related Sequences](#) [Identical Proteins](#)

Species: Animals, Fungi, Bacteria, More...
Enzyme types: Ligases, Oxidoreductases
Source databases: DDBJ, EMBL, GenBank, PDB, PIR, RefSeq, UniProtKB / Swiss-Prot
Sequence length: Custom range...
Molecular weight: Custom range...
Release date: Custom range...
Revision date: Custom range...

Filters: Manage Filters
Top Organisms: Tree
Homo sapiens (531)
Aspergillus niger (4)
Chlorococcus aethiops (1)
Cardiobacterium valvarum F0432 (1)
Streptococcus pneumoniae MNZ41 (1)
All other taxa (2)
More...

Find related data
Database: Select
Find items

Search details
androgen receptor[All Fields] AND ("Homo sapiens"[Organism] OR homo sapiens[All Fields])
Search See more...

Recent activity
Turn Off Clear
androgen receptor, homo sapiens (540)

Select one of the proteins by clicking on the link shown above to see detailed information about the protein

NCBI Resources How To Sign in to NCBI

Protein Protein Search

Advanced Help

Display Settings: GenPept Send to:

androgen receptor [Homo sapiens]

GenBank: AAA51771.1
[FASTA](#) [Graphics](#)

Go to:

LOCUS AAA51771 917 aa linear PRI 31-OCT-1994
DEFINITION androgen receptor [Homo sapiens].
ACCESSION AAA51771
VERSION AAA51771.1 GI:178872
DBSOURCE locus HUMARA accession [M21748.1](#)
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE
1 (residues 1 to 917)
AUTHORS Tilley,W.D., Marcelli,M., Wilson,J.D. and McPhaul,M.J.
TITLE Characterization and expression of a cDNA encoding the human androgen receptor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 86 (1), 327-331 (1989)
PUBMED [2911578](#)
REFERENCE
2 (sites)
AUTHORS Marcelli,M., Tilley,W.D., Wilson,C.M., Griffin,J.E., Wilson,J.D. and McPhaul,M.J.
TITLE Definition of the human androgen receptor gene structure permits the identification of mutations that cause androgen resistance: premature termination of the receptor protein at amino acid residue 588 causes complete androgen resistance
JOURNAL Mol. Endocrinol. 4 (8), 1105-1116 (1990)
PUBMED [2293020](#)
COMMENT [2] sites; androgen resistant mutation.
Draft entry and computer-readable sequence [1] kindly submitted by M.J. McPhaul, 09-DEC-1988.
Method: conceptual translation.
FEATURES
source Location/Qualifiers
1..917
/organism="Homo sapiens"

Change region shown
Customize view

Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features
Find in this Sequence

Protein 3D Structure
Targeting The Binding Function 3 (bG) Site Of The Human Androgen Receptor PDB: 4HLW
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 2.5 Å
See all 54 structures...

Articles about the AR gene
Repression of cell proliferation and androgen receptor activity in prostat [Anticancer Res. 2013]
TALEN-engineered AR gene rearrangements reveal endocrin [Proc Natl Acad Sci U S A. 2013]
Androgen receptor (AR) positive vs negative roles in prostate cancer cell d [Cancer Treat Rev. 2014]
See all...

Identical proteins for AAA51771.1

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Guiding principles: On the NCBI protein page, rows to examine include: “DEFINITION,” “REFERENCES,” “COMMENTS,” and “FEATURES.” The information provided in these rows can aid a SeqAPASS user in the identification of an ideal query sequence for SeqAPASS.

It is desirable to:

- a. Use accessions with the following prefix: NP_
- b. avoid use of protein sequences labeled “partial,” “PREDICTED,” “PROVISIONAL,” “INFERRED,” or “hypothetical”
- c. avoid using those labeled “TPA” (Third Party Annotation), however if TPA is all that is available “TPA: experimental” would be preferred over “TPA: inferential”
- d. Look at the date associated with the protein in the “LOCUS” row of the detailed protein page. A more recent date can have the most up-to-date annotation of the protein. Under the “DBSOURCE” row of the detailed protein page other accessions associated with past protein sequences can be viewed. Many times if the “xrefs” row is heavily populated and has the most recent annotation update date, it is likely to be the best sequence to use as a query sequence in SeqAPASS.
- d. Short sequences should be avoided when possible as query sequences. Many times if one selects the protein from the protein output derived from the NCBI protein database query, they will find that the short sequence is actually a partial sequence described in the “DEFINITION” row of the Protein page.
- e. Unless there is reason for doing so (based on the question the user is trying to address), splice-variants labeled in “FEATURES” rows of the Protein page as “alternatively spliced” would be less desirable
- f. It is important to check the references associated with the selected query protein. In some cases, certain sequences are associated with sensitivity to a given chemical. This can be particularly useful when predicting susceptibility to pesticides, where certain strains of insects are produced to be readily sensitive or insensitive to a chemical.
- g. A secondary check of the sequence used in the SeqAPASS run would be to look at the output derived and see whether ortholog candidates were detected. Ideally a preferential sequence would have more ortholog candidates identified.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Example: Androgen receptor, Bos taurus

NCBI Resources How To Sign in to NCBI

Protein Protein Search Help

Advanced

Display Settings: GenPept Send to:

Change region shown

Customize view

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Targeting The Binding Function 3 (bF3) Site Of The Human Androgen Receptor

PDB: 4HJM

Source: Homo sapiens

Method: X-Ray Diffraction

Resolution: 2.5 Å

See all 54 structures...

Articles about the AR gene

Repression of cell proliferation and androgen receptor activity in prx [Anticancer Res. 2013]

TALEN-engineered AR gene rearrangements reveal endo [Proc Natl Acad Sci U S A. 2013]

Androgen receptor (AR) positive vs negative roles in prostate car [Cancer Treat Rev. 2014]

See all...

Identical proteins for AA051771.1

androgen receptor [Homo sapiens] [AA051772]

See all...

Pathways for the AR gene

Integrated Breast Cancer Pathway

SIDS Susceptibility Pathways

Nuclear Receptors

See all...

Reference sequence information

RefSeq genomic sequence

See the genomic reference sequence for the AR gene (NG_009014.2).

RefSeq protein isoforms

See 4 reference sequence protein isoforms for the AR gene.

More about the AR gene

The androgen receptor gene is more than 90 kb long and codes for a protein that has 3 major functional domains: the N-terminal domain, DNA-binding domain, and LBD...

Also Known As: RP11-383C12.1, AIS, DHT...

Homologs of the AR gene

The AR gene is conserved in Rhesus monkey, dog, cow, mouse, rat, and chicken.

LinkOut to external resources

A selection of literature about the proteins [GoPubMed Proteins]

Transcript/Protein Information [PANTHER Classification System]

biochemicals [ExactAntigen/Labome]

antibody review [ExactAntigen/Labome]

others [ExactAntigen/Labome]

antibody [ExactAntigen/Labome]

cDNA clone [ExactAntigen/Labome]

protein and peptide [ExactAntigen/Labome]

ELISA and assay kit [ExactAntigen/Labome]

LOCUS AA051771 317 aa linear PRI 01-OCT-1994

DEFINITION androgen receptor [Homo sapiens].

ACCESSION AA051771

VERSION AA051771.1 GI:174872

RESOURCE locus HOMOLOC accession H01748.1

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (residues 1 to 317)

AUTHORS Tilley, M. D., Harcell, M. J., Wilson, J. D. and McPhaul, M. J.

TITLE Characterization and expression of a cDNA encoding the human androgen receptor

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 66 (1), 327-331 (1989)

PMID 2911578

REFERENCE 2 (sites)

AUTHORS Harcell, M. J., Tilley, M. D., Wilson, J. D., Griffin, J. E., Wilson, J. D. and McPhaul, M. J.

TITLE Definition of the human androgen receptor gene structure permits the identification of mutations that cause androgen resistance: premature termination of the receptor protein at amino acid residue 588 causes complete androgen resistance

JOURNAL Mol. Endocrinol. 4 (8), 1105-1116 (1990)

PMID 2193948

COMMENT [2] sites: androgen resistant mutation. Draft entry and computer-readable sequence [1] kindly submitted by M. J. McPhaul, 09-DEC-1988. Method: conceptual translation. Location/Qualifiers

source

1..317

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="Xq11.2-q12"

/sex="male"

/tissue_type="prostate"

1..317

/product="androgen receptor"

6..446

/region_name="Androgen_recept"

/note="Androgen receptor; p1am#2166"

/db_xref="CID:111037"

552..622

/region_name="DBD_AR"

/note="DNA-binding domain of androgen receptor (AR) is composed of two 14-type zinc fingers; c00173"

/db_xref="CID:143547"

order(557,560,574,577,593,599,609,612)

/site_type="other"

/note="zinc binding site [ion binding]"

/db_xref="CID:143547"

order(586..589,578,579..579,582..583,591,606..607,610,613)

/site_type="DNA binding"

/note="DNA binding site [nucleotide binding]"

/db_xref="CID:143547"

order(592..596,598..600,605,608)

/site_type="other"

/note="dimer interface [polypeptide binding]"

/db_xref="CID:143547"

670..915

/region_name="LBD_AR"

/note="Ligand binding domain of the nuclear receptor androgen receptor, ligand activated transcription regulator; c007073"

/db_xref="CID:122758"

order(699,702..703,705..706,709,729..740,743..744,747,750,762,778,785,871,875)

/site_type="other"

/note="Ligand binding site [chemical binding]"

/db_xref="CID:122758"

order(711,714,718,724,728,732,736,891..892,895..896)

/site_type="other"

/note="coactivator recognition site [polypeptide binding]"

/db_xref="CID:122758"

1..317

/gene="AR"

/coded_by="M21748.1.163..2316"

/db_xref="Gene:600-120-556"

ORIGIN

1 mevqlglgrv ypppdktyr gqgnlgrv revipggr hpeaaasapp gas111lqg

61 gggggggggg ggggggggts prggggggg dgsqahrrg pt-gylvldee gqgpgsal

121 nchpergcpv ppaavaakh glpqlpapp dehdraapst llllgtfpe ltrcaalkd

181 llrartnml lggggggvss ggggggare rrgggsdhd rylggstss dshelckav

241 svnmglgvea lehlrpgql rdnmyapll gppavpqp eaclaeqgl llddsgktr

301 edkaeyptk gytgkleg elgcgaaa gsgtlelp tilykqal deaayqrd

361 yynplalag pppppppph harikterpl dyaawaaa acrygdla lhgaaagp

421 sgpaaas subelitate gilycpqgg gggggggggg gggggggggg tagavapyy

481 tpppglqg eddtepdhw ypggvsvrp yppctvks wggmdyys pygdrleta

541 rdlpdyfpp qpecllc gbaagchyg altpclevl hkaaeqk ylaarndec

601 ldrtdmcp rlylcyra gmlgarklk hnglkqg qaarvcpv eecqhtov

661 hieytcgp flrvtaiep gvcaqhdn qdifaall: sinelgerl vhwvmskal

721 pgrnlhvdv qnavqyrom glvvtanqr stnvmml ytapdlvne ymbkmys

781 qcvamhlgv efgolqtpg eflnkal11 fsiipvdgk ngtfdelm ryikeldrii

841 ackerlptc rryfqlckl ldsvapiare lhgtdflll khwvsvdp emaeiivg

901 vpkilgqvk pythq

//

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

To evaluate sequences, change settings for “Conservation Setting” from “2 Bits” to “Identity”

The screenshot shows the COBALT web interface. At the top, there's a navigation bar with 'Home', 'Recent Results', and 'Help'. Below that, a title bar reads '- Cobalt RID EMV7SF1X211 (7 seqs)'. A yellow warning box states: 'All queries form only one cluster. No domain information was used for generating constraints. Decreasing maximum in-cluster distance or turning off query clustering option may improve results.' Below the warning, there's a 'Descriptions' section with a table of sequence accessions and descriptions. The table has columns for 'Accession', 'Description', and 'Links'. The 'Accession' column contains entries like P10275.2, AAA51772.1, AAA51780.1, AAA51771.1, AAA51729.1, AAD45921.1, and AAA51886.1. The 'Description' column contains details about androgen receptors. The 'Links' column contains icons for UniGene, GEO, Gene, Structure, and Map Viewer. Below the table, there's an 'Alignments' section with a 'View Format' dropdown set to 'Compact' and a 'Conservation Setting' dropdown set to 'Identity'. The alignment view shows a sequence alignment with columns for each accession and a 'Conservation Setting' dropdown menu that is open, showing options: 1 Bit, 2 Bits, 3 Bits, 4 Bits, and Identity. The 'Identity' option is selected.

Look for differences in the sequence (e.g., conserved residues, gaps) and start by eliminating sequences that have gaps.

i. If after the suggested evaluations of the proteins are performed and questions remain as to which sequence would be best to run in SeqAPASS, run all relevant sequences in SeqAPASS for the evaluation. The individual residue differences between commonly named sequences will become most important when evaluating residues known to be important for binding the chemical or activating the protein (Level 3 SeqAPASS analysis). After completing the SeqAPASS run, select the data that has the greatest number of ortholog candidates for your evaluation of conservation and further predictions of cross species susceptibility.

Depending on the protein of interest, multiple subunits may be associated with a protein. In this case, all relevant subunits can be queried using SeqAPASS.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 1 Calculated Percent Similarity

The SeqAPASS algorithms submit the query to NCBI's standalone BLASTp (using default settings, including BLOSUM-62 matrix), which aligns the query protein with all proteins available in the NCBI protein database and provides a variety of metrics associated with each pairwise alignment between the query and hit sequences. SeqAPASS selectively captures output from BLASTp, including one sequence per species with the highest bit score.

Detailed descriptions of metrics derived from BLASTp (e.g., BLASTP Bitscore, Evaluate, Positives, Identity, Hit length) can be found in:

The NCBI Handbook: (<http://www.ncbi.nlm.nih.gov/books/NBK21106/>);

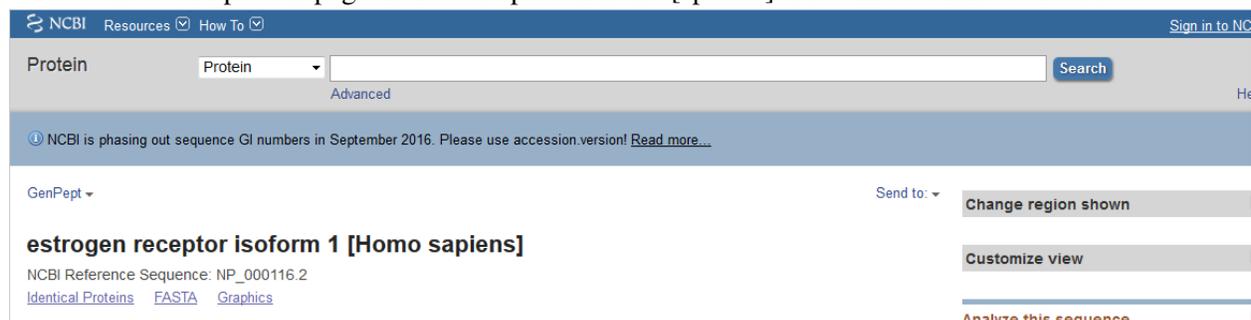
BLAST® Help: (<http://www.ncbi.nlm.nih.gov/books/NBK62051/>)

And

NCBI Glossary Field Guide: (<http://www.ncbi.nlm.nih.gov/Class/FieldGuide/glossary.html>)

The top row of the Level 1 data corresponds to the canonical protein from the user entered query species/protein. Therefore, the accession in the first row may not match the accession the user entered. The logic used to determine the top row query sequence from the BLASTp output proceeds as follows: first identify the top 10 hit sequences that are from the query species submitted by the user, if multiple sequences exist, choose the top hit with the highest Hsp_identity and if sequences are still tied, choose the top hit with the highest Hsp_bit-score and if sequences are still tied, choose the sequence with the lowest Hsp_evalue and if sequences are still tied, choose the sequence with the highest Hit_len, and if still tied This strategy was employed to aid the user in identifying an appropriate sequence for the query sequence. For each sequence queried, the Level 1, top row query species is used to determine the maximum bitscore for the analysis, which is derived from aligning the query sequence to itself using BLASTp. To calculate percent similarity, the bitscore for each hit sequence is normalized to the maximum bit score and then multiplied by 100.

Note: Many NCBI protein accessions represent multiple identical protein sequences in the BLASTp output. This is due to BLASTp querying and presenting data from the non-redundant protein database. Sometimes the identical sequences are from different species. If the top row species identified in the Level 1 data is not the same as the species entered by the user or associated with the original user defined query accession, it is likely that the top row sequence is identical to the user entered query sequence/species. This can be checked by following the link for the top row “NCBI Accession” in the table to the NCBI protein page. Below the protein name [species] title will be a link to “Identical



The screenshot shows the NCBI protein page for 'estrogen receptor isoform 1 [Homo sapiens]'. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' links, and a 'Sign in to NCBI' link. Below this is a search bar with a dropdown menu set to 'Protein' and a 'Search' button. A blue banner below the search bar contains the message: 'NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! Read more...'. The main content area shows the protein name 'estrogen receptor isoform 1 [Homo sapiens]' in bold, followed by the NCBI Reference Sequence: NP_000116.2. There are links for 'Identical Proteins', 'FASTA', and 'Graphics'. On the right side, there are buttons for 'Change region shown' and 'Customize view'. At the bottom right, there is a link that says 'Analyze this sequence'.

Proteins.”

Click the “Identical Proteins” link and look for a sequence in the list from the user defined query species. If a protein for the user defined query species is present in the list, it is identical to the top row sequence in the Level 1 data table and the user can substitute the Accession information for the desired query species/sequence in the first row, moving the initial top row sequence to the second row.

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Note: It is anticipated that in a future version of the SeqAPASS application the BLASTP query and hit accessions will be parsed to identify all the species/accessions from the identical proteins. Therefore, our susceptibility predictions are likely to improve in instances where an accession represents multiple species, which is particularly important for our susceptibility predictions if the species is from a different taxonomic group than the identical hit protein that was identified in BLAST.

Common Domain Count:

Reversed Position Specific BLAST (RPS BLAST) is used to compare each query and hit sequence to conserved domains defined in NCBI's Conserved Domain Database. A hit domain is considered in common with the query domain if it contains the same domain accession as the query and it aligns with the NCBI curated domain with the same or greater amino acid residue coverage than the query sequence.

Ortholog Candidate Identification

Ortholog sequences are those that have diverged from a speciation event and therefore are more likely to maintain similar function. SeqAPASS uses reciprocal best hit (RBH) BLAST for ortholog detection by automatically comparing each hit protein to all protein sequences available for the query species and if the original query protein is identified to be the best match to the hit or maintain the same bitscore, then the hit sequence would be considered an ortholog candidate. The sequence is indicated an Ortholog Candidate or not with a yes (Y) or no (N) in the column.

Susceptibility cut-off:

The susceptibility cut-offs listed on the “Level 1 (and Level 2) Susceptibility Cut-off” page are determined by plotting the % similarity data from the “Primary Report” and identifying the local minimums in the data. The default cut-off is determined by taking the 1st local minimum and moving up in percent similarity until the next ortholog candidate is found. The susceptibility cut-off displayed in the list is the percent similarity of the identified ortholog candidate.

Criteria for Susceptibility Prediction (including Species Read-across):

All sequences identified above the susceptibility cut-off are predicted to be susceptible; therefore Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but identified as an Ortholog Candidate =Y, for “yes,” then the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but belongs to any organism class found above the susceptibility cut-off the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”. This criteria allows susceptibility predictions to be made across taxonomic groups based on the likelihood that the sequences above the cut-off are better matches to the query.

If the hit sequence is below the susceptibility cut-off and not identified as an ortholog candidate (Ortholog Candidate =N, for “no,”) and does not belong to any organism class found above the susceptibility cut-off, the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

Note that the “Primary Report” may yield different Susceptibility Predictions than the “Full Report,” as the predictions are based on the data in the different reports. The Primary Report is filtered to only display E-value ≤ 0.01 and Common Domain Count ≥ 1 .

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): User Guide

Updated 5/25/16; Contact Carlie LaLone with Questions: LaLone.Carlie@epa.gov

Level 2 Calculated Percent Similarity

Data obtained from the Level 1 RPS BLAST evaluation is used to assign sequence ranges that aligned with a user selected domain (from the NCBI CDD database) to each accessions from the Level 1 Full report. BLASTp is then used to align the query domain range to each hit domain range. The percent similarity is calculated based on the bit scores from the BLASTp alignment of the domain regions. For each sequence queried, the Level 2, top row query species is used to determine the maximum bitscore for the analysis, which is derived from aligning the query sequence to itself using BLASTp. To calculate percent similarity, the bitscore for each hit sequence is normalized to the maximum bit score and then multiplied by 100.

Susceptibility cut-off (same method as used in Level 1):

The susceptibility cut-offs listed on the “Level 2 Susceptibility Cut-off” page are determined by plotting the % similarity data from the “Primary Report” and identifying the local minimums in the data. The default cut-off is determined by taking the 1st local minimum and moving up in percent similarity until the next ortholog candidate is found. The susceptibility cut-off displayed in the list is the percent similarity of the identified ortholog candidate.

Criteria for Susceptibility Prediction (including Species Read-across):

All sequences identified above the susceptibility cut-off are predicted to be susceptible; therefore Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but identified as an Ortholog Candidate =Y, for “yes,” then the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”

If the hit sequence is below the susceptibility cut-off, but belongs to any organism class found above the susceptibility cut-off the hit is predicted to be susceptible; therefore, Susceptibility Prediction = Y for “yes”. This criteria allows susceptibility predictions to be made across taxonomic groups based on the likelihood that the sequences above the cut-off are better matches to the query.

If the hit sequence is below the susceptibility cut-off and not identified as an ortholog candidate (Ortholog Candidate =N, for “no,”) and does not belong to any organism class found above the susceptibility cut-off, the hit is predicted to not be susceptible; therefore, Susceptibility Prediction = N for “no”

Level 3 Sequence Alignments:

COBALT is used to align all user selected sequences (from Level 1 hits) with a user defined template sequence. Because COBALT algorithms align all sequences, it is recommended that the user align the template sequence with sequences that are most similar to one another. As a means to capture the most similar sequences from the SeqAPASS data it is recommended that the user filter the Level 1 data by taxonomic group and step through the Level 1 data pages one by one while selecting sequences. It is recommended that the user look at the name of the sequence and exclude ‘partial’ sequences when possible. Requesting a query from one taxonomic group at a time, breaks the data down in manageable alignments.

Selecting Amino Acid Residues to Align:

The user may select up to 50 amino acid residues to compare across selected species in Level 3.